Additional File 3: taxa, gene contexts and multiple-sequence alignments of all motifs supplementary for: Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea and their metagenomes

Zasha Weinberg, Joy X. Wang, Jarrod Bogue, Jingying Yang,
Keith Corbino, Ryan H. Moy, Ronald R. Breaker

Note: the presentation of this supplementary data on novel RNA motifs follows the pattern of our presentation of previously found conserved RNA motifs [14].

Contents

1 6S-flavo **16**
1.1 Taxa . . . 16
1.2 Notes . . . 17
1.3 Gene contexts . . . 17
1.4 Conserved domains . . . 19
1.5 Multiple-sequence alignment . . . 19

2 *aceE* **26**
2.1 Taxa . . . 26
2.2 Gene contexts . . . 27
2.3 Conserved domains . . . 27
2.4 Multiple-sequence alignment . . . 28

3 Acido-1 **30**
3.1 Taxa . . . 30
3.2 Gene contexts . . . 30
3.3 Conserved domains . . . 30
3.4 Multiple-sequence alignment . . . 31

4 Acido-Lenti-1 **32**
4.1 Taxa . . . 32
4.2 Gene contexts . . . 32
4.3 Conserved domains . . . 34
4.4 Multiple-sequence alignment . . . 35

5 Actino-pnp **39**
5.1 Taxa . . . 39
5.2 Gene contexts . . . 39
5.3 Conserved domains . . . 42
5.4 Multiple-sequence alignment . . . 42

6 AdoCbl-variant **49**
6.1 Taxa . . . 49
6.2 Notes . . . 49
6.3 Gene contexts . . . 49
6.4 Conserved domains . . . 53
6.5 Multiple-sequence alignment . . . 53

7 *asd* **64**
7.1 Taxa . . . 64
7.2 Gene contexts . . . 65
7.3 Conserved domains . . . 66
7.4 Multiple-sequence alignment . . . 67

8 *atoC* **71**
8.1 Taxa . . . 71
8.2 Gene contexts . . . 71
8.3 Conserved domains . . . 72
8.4 Multiple-sequence alignment . . . 72

9 Bacillaceae-1 **74**
9.1 Taxa . . . 74
9.2 Gene contexts . . . 74
9.3 Conserved domains . . . 77
9.4 Multiple-sequence alignment . . . 79

10 ***Bacillus*-plasmid** **86**
10.1 Taxa . . . 86
10.2 Gene contexts . . . 86
10.3 Conserved domains . . . 87
10.4 Multiple-sequence alignment . . . 88

11 Bacteroid-*trp* **91**
11.1 Taxa . . . 91
11.2 Gene contexts . . . 92
11.3 Conserved domains . . . 93
11.4 Multiple-sequence alignment . . . 94

12 Bacteroidales-1 **98**
12.1 Taxa . . . 98
12.2 Gene contexts . . . 98
12.3 Conserved domains . . . 99
12.4 Multiple-sequence alignment . . . 100

13 *Bacteroides*-1 **104**
13.1 Taxa . . . 104
13.2 Gene contexts . . . 105
13.3 Conserved domains . . . 106
13.4 Multiple-sequence alignment . . . 107

14 *Bacteroides*-2 **111**
14.1 Taxa . . . 111
14.2 Gene contexts . . . 112
14.3 Conserved domains . . . 113
14.4 Multiple-sequence alignment . . . 114

15 Burkholderiales-1 **119**
15.1 Taxa . . . 119
15.2 Gene contexts . . . 119
15.3 Conserved domains . . . 121
15.4 Multiple-sequence alignment . . . 124

16 c4 antisense RNA **130**
16.1 Taxa . . . 130
16.2 Gene contexts . . . 134
16.3 Conserved domains . . . 146
16.4 Multiple-sequence alignment . . . 149

17 c4-a1b1 **182**
17.1 Taxa . . . 182
17.2 Gene contexts . . . 183
17.3 Conserved domains . . . 185
17.4 Multiple-sequence alignment . . . 186

18 Chlorobi-1 **190**
18.1 Taxa . . . 190
18.2 Notes . . . 191
18.3 Gene contexts . . . 191
18.4 Conserved domains . . . 191
18.5 Multiple-sequence alignment . . . 192

19 Chlorobi-RRM **194**
19.1 Taxa . . . 194
19.2 Gene contexts . . . 194
19.3 Conserved domains . . . 195
19.4 Multiple-sequence alignment . . . 195

20 Chloroflexi-1 **197**
20.1 Taxa . . . 197
20.2 Gene contexts . . . 197
20.3 Conserved domains . . . 198
20.4 Multiple-sequence alignment . . . 198

21 Clostridiales-1 **199**
21.1 Taxa . . . 199
21.2 Gene contexts . . . 199
21.3 Conserved domains . . . 203
21.4 Multiple-sequence alignment . . . 206

22 COG2252 **224**
22.1 Taxa . . . 224
22.2 Gene contexts . . . 224
22.3 Conserved domains . . . 224
22.4 Multiple-sequence alignment . . . 225

23 *Collinsella*-1 **226**
23.1 Taxa . . . 226
23.2 Gene contexts . . . 227
23.3 Conserved domains . . . 227
23.4 Multiple-sequence alignment . . . 227

24 *crcB* **228**
24.1 Taxa . . . 228
24.2 Notes . . . 234
24.3 Gene contexts . . . 234
24.4 Conserved domains . . . 244
24.5 Multiple-sequence alignment . . . 245

25 Cyano-1 **272**
25.1 Taxa . . . 272
25.2 Gene contexts . . . 273
25.3 Conserved domains . . . 278
25.4 Multiple-sequence alignment . . . 280

26 Cyano-2 **299**
26.1 Taxa . . . 299
26.2 Gene contexts . . . 299
26.3 Conserved domains . . . 301
26.4 Multiple-sequence alignment . . . 301

27 Desulfotalea-1 **307**
27.1 Taxa . . . 307
27.2 Gene contexts . . . 307
27.3 Conserved domains . . . 312
27.4 Multiple-sequence alignment . . . 315

28 Dictyoglomi-1 **330**
28.1 Taxa . . . 330
28.2 Gene contexts . . . 331
28.3 Conserved domains . . . 331
28.4 Multiple-sequence alignment . . . 331

29 Downstream-peptide **332**
29.1 Taxa . . . 332
29.2 Notes . . . 333
29.3 Gene contexts . . . 333
29.4 Conserved domains . . . 346
29.5 Multiple-sequence alignment . . . 347

30 ***epsC*** **391**
30.1 Taxa . . . 391
30.2 Gene contexts . . . 392
30.3 Conserved domains . . . 393
30.4 Multiple-sequence alignment . . . 394

31 ***fixA*** **397**
31.1 Taxa . . . 397
31.2 Gene contexts . . . 397
31.3 Conserved domains . . . 398
31.4 Multiple-sequence alignment . . . 398

32 Flavo-1 **400**
32.1 Taxa . . . 400
32.2 Gene contexts . . . 400
32.3 Conserved domains . . . 405
32.4 Multiple-sequence alignment . . . 409

33 *flg*-Rhizobiales **429**
33.1 Taxa . . . 429
33.2 Gene contexts . . . 430
33.3 Conserved domains . . . 430
33.4 Multiple-sequence alignment . . . 431

34 *flpD* **433**
34.1 Taxa . . . 433
34.2 Gene contexts . . . 433
34.3 Conserved domains . . . 433
34.4 Multiple-sequence alignment . . . 434

35 *gabT* **435**
35.1 Taxa . . . 435
35.2 Gene contexts . . . 436
35.3 Conserved domains . . . 436
35.4 Multiple-sequence alignment . . . 437

36 Gamma-*cis*-1 **439**
36.1 Taxa . . . 439
36.2 Notes . . . 440
36.3 Gene contexts . . . 440
36.4 Conserved domains . . . 441
36.5 Multiple-sequence alignment . . . 442

37 *glnA* **446**
37.1 Taxa . . . 446
37.2 Gene contexts . . . 446
37.3 Conserved domains . . . 467
37.4 Multiple-sequence alignment . . . 469

38 GUCCY-hairpin **557**
38.1 Taxa . . . 557
38.2 Gene contexts . . . 557
38.3 Conserved domains . . . 570
38.4 Multiple-sequence alignment . . . 579

39 Gut-1 **611**
39.1 Taxa . . . 611
39.2 Gene contexts . . . 611
39.3 Conserved domains . . . 612
39.4 Multiple-sequence alignment . . . 613

40 *gyrA* **617**
40.1 Taxa . . . 617
40.2 Gene contexts . . . 618
40.3 Conserved domains . . . 619
40.4 Multiple-sequence alignment . . . 619

41 *hopC* **621**
41.1 Taxa . . . 621
41.2 Gene contexts . . . 622
41.3 Conserved domains . . . 622
41.4 Multiple-sequence alignment . . . 622

42 *icd* **623**
42.1 Taxa . . . 623
42.2 Gene contexts . . . 624
42.3 Conserved domains . . . 625
42.4 Multiple-sequence alignment . . . 625

43 JUMPstart **627**
43.1 Taxa . . . 627
43.2 Gene contexts . . . 629
43.3 Conserved domains . . . 637
43.4 Multiple-sequence alignment . . . 639

44 L17 downstream element **646**
44.1 Taxa . . . 646
44.2 Gene contexts . . . 648
44.3 Conserved domains . . . 650
44.4 Multiple-sequence alignment . . . 650

45 *lactis*-**plasmid** **655**
45.1 Taxa . . . 655
45.2 Gene contexts . . . 655
45.3 Conserved domains . . . 656
45.4 Multiple-sequence alignment . . . 656

46 Lacto-*int* **658**
46.1 Taxa . . . 658
46.2 Gene contexts . . . 659
46.3 Conserved domains . . . 660
46.4 Multiple-sequence alignment . . . 660

47 Lacto-*rpoB* **665**
47.1 Taxa . . . 665
47.2 Gene contexts . . . 666
47.3 Conserved domains . . . 667
47.4 Multiple-sequence alignment . . . 668

48 Lacto-*usp* **671**
48.1 Taxa . . . 671
48.2 Gene contexts . . . 671
48.3 Conserved domains . . . 672
48.4 Multiple-sequence alignment . . . 672

49 Leu/phe leader **673**
49.1 Taxa . . . 673
49.2 Notes . . . 673
49.3 Gene contexts . . . 674
49.4 Conserved domains . . . 674
49.5 Multiple-sequence alignment . . . 674

50 *livK* **676**
50.1 Taxa . . . 676
50.2 Gene contexts . . . 676
50.3 Conserved domains . . . 677
50.4 Multiple-sequence alignment . . . 677

51 Lnt **679**
51.1 Taxa . . . 679
51.2 Gene contexts . . . 679
51.3 Conserved domains . . . 680
51.4 Multiple-sequence alignment . . . 680

52 *manA* **681**
52.1 Taxa . . . 681
52.2 Gene contexts . . . 682
52.3 Conserved domains . . . 686
52.4 Multiple-sequence alignment . . . 687

53 *Methylobacterium*-**1** **705**
53.1 Taxa . . . 705
53.2 Gene contexts . . . 705
53.3 Conserved domains . . . 706
53.4 Multiple-sequence alignment . . . 707

54 Moco-II **711**
54.1 Taxa . . . 711
54.2 Gene contexts . . . 712
54.3 Conserved domains . . . 712
54.4 Multiple-sequence alignment . . . 713

55 *mraW* **714**
55.1 Taxa . . . 714
55.2 Gene contexts . . . 715
55.3 Conserved domains . . . 716
55.4 Multiple-sequence alignment . . . 717

56 *msiK* **721**
56.1 Taxa . . . 721
56.2 Gene contexts . . . 722
56.3 Conserved domains . . . 724
56.4 Multiple-sequence alignment . . . 725

57 *Nitrosococcus*-**1** **732**
57.1 Taxa . . . 732
57.2 Gene contexts . . . 732
57.3 Conserved domains . . . 733
57.4 Multiple-sequence alignment . . . 733

58 *nuoG* **734**
58.1 Taxa . . . 734
58.2 Gene contexts . . . 736
58.3 Conserved domains . . . 739
58.4 Multiple-sequence alignment . . . 740

59 Ocean-V **743**
59.1 Taxa . . . 743
59.2 Gene contexts . . . 743
59.3 Conserved domains . . . 743
59.4 Multiple-sequence alignment . . . 744

60 Ocean-VI **745**
60.1 Taxa . . . 745
60.2 Gene contexts . . . 745
60.3 Conserved domains . . . 751
60.4 Multiple-sequence alignment . . . 752

61 *pan* **761**
61.1 Taxa . . . 761
61.2 Gene contexts . . . 763
61.3 Conserved domains . . . 766
61.4 Multiple-sequence alignment . . . 767

62 *Pedo*-**repair** **774**
62.1 Taxa . . . 774
62.2 Gene contexts . . . 774
62.3 Conserved domains . . . 774
62.4 Multiple-sequence alignment . . . 774

63 *pfl* **776**
63.1 Taxa . . . 776
63.2 Notes . . . 779
63.3 Gene contexts . . . 779
63.4 Conserved domains . . . 784
63.5 Multiple-sequence alignment . . . 785

64 *pheA* **806**
64.1 Taxa . . . 806
64.2 Gene contexts . . . 807
64.3 Conserved domains . . . 808
64.4 Multiple-sequence alignment . . . 808

65 PhotoRC-I **810**
65.1 Taxa . . . 810
65.2 Gene contexts . . . 811
65.3 Conserved domains . . . 812
65.4 Multiple-sequence alignment . . . 813

66 PhotoRC-II **816**
66.1 Taxa . . . 816
66.2 Gene contexts . . . 817
66.3 Conserved domains . . . 826
66.4 Multiple-sequence alignment . . . 827

67 *Polynucleobacter*-**1** **859**
67.1 Taxa . . . 859
67.2 Gene contexts . . . 859
67.3 Conserved domains . . . 860
67.4 Multiple-sequence alignment . . . 860

68 *potC* **862**
68.1 Taxa . . . 862
68.2 Gene contexts . . . 863
68.3 Conserved domains . . . 865
68.4 Multiple-sequence alignment . . . 866

69 *psaA* **876**
69.1 Taxa . . . 876
69.2 Gene contexts . . . 876
69.3 Conserved domains . . . 877
69.4 Multiple-sequence alignment . . . 877

70 *psbNH* **882**
70.1 Taxa . . . 882
70.2 Gene contexts . . . 882
70.3 Conserved domains . . . 883
70.4 Multiple-sequence alignment . . . 883

71 *Pseudomon*-**1** **887**
71.1 Taxa . . . 887
71.2 Gene contexts . . . 887
71.3 Conserved domains . . . 888
71.4 Multiple-sequence alignment . . . 888

72 *Pseudomon*-**2** **890**
72.1 Taxa . . . 890
72.2 Gene contexts . . . 891
72.3 Conserved domains . . . 891
72.4 Multiple-sequence alignment . . . 892

73 *Pseudomon*-**GGDEF** **894**
73.1 Taxa . . . 894
73.2 Gene contexts . . . 894
73.3 Conserved domains . . . 894
73.4 Multiple-sequence alignment . . . 895

74 *Pseudomon-groES* **896**
74.1 Taxa . . . 896
74.2 Gene contexts . . . 897
74.3 Conserved domains . . . 897
74.4 Multiple-sequence alignment . . . 897

75 *Pseudomon*-**Rho** **900**
75.1 Taxa . . . 900
75.2 Gene contexts . . . 900
75.3 Conserved domains . . . 901
75.4 Multiple-sequence alignment . . . 901

76 *Pyrobac*-**1** **903**
76.1 Taxa . . . 903
76.2 Gene contexts . . . 904
76.3 Conserved domains . . . 904
76.4 Multiple-sequence alignment . . . 904

77 *Pyrobac*-**HINT** **906**
77.1 Taxa . . . 906
77.2 Gene contexts . . . 906
77.3 Conserved domains . . . 906
77.4 Multiple-sequence alignment . . . 906

78 *radC* **907**
78.1 Taxa . . . 907
78.2 Gene contexts . . . 908
78.3 Conserved domains . . . 912
78.4 Multiple-sequence alignment . . . 912

79 Rhizobiales-1 **926**
79.1 Taxa . . . 926
79.2 Gene contexts . . . 927
79.3 Conserved domains . . . 942
79.4 Multiple-sequence alignment . . . 954

80 Rhizobiales-2 **1001**
80.1 Taxa . . . 1001
80.2 Gene contexts . . . 1001
80.3 Conserved domains . . . 1002
80.4 Multiple-sequence alignment . . . 1002

81 Rhodopirellula-1 **1004**
81.1 Taxa . . . 1004
81.2 Gene contexts . . . 1004
81.3 Conserved domains . . . 1008
81.4 Multiple-sequence alignment . . . 1009

82 *rmf* **1020**
82.1 Taxa . . . 1020
82.2 Gene contexts . . . 1020
82.3 Conserved domains . . . 1021
82.4 Multiple-sequence alignment . . . 1021

83 *rne*-**II** **1023**
83.1 Taxa . . . 1023
83.2 Gene contexts . . . 1023
83.3 Conserved domains . . . 1024
83.4 Multiple-sequence alignment . . . 1024

84 SAM-Chlorobi **1028**
84.1 Taxa . . . 1028
84.2 Gene contexts . . . 1028
84.3 Conserved domains . . . 1029
84.4 Multiple-sequence alignment . . . 1029

85 SAM-I-IV-variant **1030**
85.1 Taxa . . . 1030
85.2 Notes . . . 1031
85.3 Gene contexts . . . 1031
85.4 Conserved domains . . . 1041
85.5 Multiple-sequence alignment . . . 1043

86 SAM-II long loops **1083**
86.1 Taxa . . . 1083
86.2 Gene contexts . . . 1084
86.3 Conserved domains . . . 1088
86.4 Multiple-sequence alignment . . . 1088

87 SAM/SAH riboswitch **1100**
87.1 Taxa . . . 1100
87.2 Gene contexts . . . 1101
87.3 Conserved domains . . . 1103
87.4 Multiple-sequence alignment . . . 1103

88 *sanguinis*-**hairpin** **1107**
88.1 Taxa . . . 1107
88.2 Gene contexts . . . 1107
88.3 Conserved domains . . . 1108
88.4 Multiple-sequence alignment . . . 1108

89 *sbcD* **1109**
89.1 Taxa . . . 1109
89.2 Gene contexts . . . 1110
89.3 Conserved domains . . . 1110
89.4 Multiple-sequence alignment . . . 1110

90 ScRE **1112**
90.1 Taxa . . . 1112
90.2 Gene contexts . . . 1112
90.3 Conserved domains . . . 1114
90.4 Multiple-sequence alignment . . . 1114

91 Soil-1 **1117**
91.1 Taxa . . . 1117
91.2 Gene contexts . . . 1118
91.3 Conserved domains . . . 1118
91.4 Multiple-sequence alignment . . . 1118

92 *Solibacter*-**1** **1120**
92.1 Taxa . . . 1120
92.2 Gene contexts . . . 1120
92.3 Conserved domains . . . 1121
92.4 Multiple-sequence alignment . . . 1122

93 STAXI **1126**
93.1 Taxa . . . 1126
93.2 Gene contexts . . . 1127
93.3 Conserved domains . . . 1129
93.4 Multiple-sequence alignment . . . 1129

94 *sucA*-**II** **1134**
94.1 Taxa . . . 1134
94.2 Gene contexts . . . 1134
94.3 Conserved domains . . . 1135
94.4 Multiple-sequence alignment . . . 1136

95 *sucC* **1138**
95.1 Taxa . . . 1138
95.2 Notes . . . 1139
95.3 Gene contexts . . . 1139
95.4 Conserved domains . . . 1140
95.5 Multiple-sequence alignment . . . 1140

96 **Termite-*flg*** **1143**
96.1 Taxa . . . 1143
96.2 Gene contexts . . . 1143
96.3 Conserved domains . . . 1144
96.4 Multiple-sequence alignment . . . 1144

97 **Termite-*leu*** **1146**
97.1 Taxa . . . 1146
97.2 Notes . . . 1146
97.3 Gene contexts . . . 1146
97.4 Conserved domains . . . 1147
97.5 Multiple-sequence alignment . . . 1147

98 ***traJ*-II** **1149**
98.1 Taxa . . . 1149
98.2 Gene contexts . . . 1150
98.3 Conserved domains . . . 1152
98.4 Multiple-sequence alignment . . . 1153

99 **Transposase-resistance** **1157**
99.1 Taxa . . . 1157
99.2 Gene contexts . . . 1158
99.3 Conserved domains . . . 1162
99.4 Multiple-sequence alignment . . . 1163

100 **TwoAYGGAY** **1175**
100.1 Taxa . . . 1175
100.2 Gene contexts . . . 1176
100.3 Conserved domains . . . 1180
100.4 Multiple-sequence alignment . . . 1182

101 *wcaG* **1196**
101.1 Taxa . . . 1196
101.2 Gene contexts . . . 1197
101.3 Conserved domains . . . 1199
101.4 Multiple-sequence alignment . . . 1200

102 Whalefall-1 **1209**
102.1 Taxa . . . 1209
102.2 Gene contexts . . . 1209
102.3 Conserved domains . . . 1210
102.4 Multiple-sequence alignment . . . 1210

103 *yjdF* **1212**
103.1 Taxa . . . 1212
103.2 Notes . . . 1214
103.3 Gene contexts . . . 1214
103.4 Conserved domains . . . 1216
103.5 Multiple-sequence alignment . . . 1217

104 *ykkC*-III **1225**
104.1 Taxa . . . 1225
104.2 Notes . . . 1226
104.3 Gene contexts . . . 1226
104.4 Conserved domains . . . 1228
104.5 Multiple-sequence alignment . . . 1228

1 6S-flavo

1.1 Taxa

The taxonomy of each organism containing a putative 6S-flavo RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual 6S-flavo RNA in Sections 1.3 and 1.5.

abbrev. of hits	taxonomy of species
ueu-1-1	Bacteria Bacteroidetes environmental samples unidentified eubacterium SCB49
Fba-1-1	Bacteria Bacteroidetes *Flavobacteria bacterium* BBFL7
Fba-2-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales *Flavobacteria bacterium* BAL38
Cat-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Croceibacter atlanticus* HTCC2559
Ddo-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Dokdonia donghaensis* MED134
Fjo-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Flavobacterium johnsoniae* UW101
Fps-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Flavobacterium psychrophilum* JIP02/86
Lbl-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Leeuwenhoekiella blandensis* MED217
Pir-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Polaribacter irgensii* 23-P
Psp-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Polaribacter sp.* MED152
Pto-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Psychroflexus torquis* ATCC 700755
Rbi-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Robiginitalea biformata* HTCC2501
Fba-3-1	Bacteria Bacteroidetes Flavobacteria *Flavobacteriales bacterium* HTCC2170
Psp-2-1	Bacteria Bacteroidetes Sphingobacteria Sphingobacteriales Sphingobacteriaceae *Pedobacter sp.* BAL39
Gka-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Geobacillus kaustophilus* HTA426
Dps-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfobacterales Desulfobulbaceae *Desulfotalea psychrophila* LSv54
env-1 to env-76	environmental samples

1.2 Notes

Superscript numbers are used to annotate any 6S-flavo RNA that has special characteristics (described below). These numeric annotations will be used in Sections 1.3 and 1.5. The meaning of the numbers are as follows.

[1] overlaps known 6S

1.3 Gene contexts

Each 6S-flavo RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one 6S-flavo RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 1.4. The accession number of the sequence containing each 6S-flavo RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 1.1. Superscript numbers refer to annotations listed in Section 1.2

abbrev.	Seq. accession		5′ at	3′ at	genes
Gka-1-1	NC_006510.1	+	1316460	1316575	HNHc (cd00085)RVT (pfam00078)→ **RNA**→ COG1418 (COG1418)→
[1]Dps-1-1	NC_006138.1	-	3243257	3243154	hypo→ **RNA**→ COG1418 (COG1418)→ TyrS (COG0162)→
Psp-2-1	NZ_ABCM01000031.1	-	12076	11976	DUF710 (pfam05164)→ **RNA**→ COG1094 (COG1094)COG1418 (COG1418)→ hypo→ DUF962 (pfam06127)→
env-1	JCVI_SCAF_1101668538900	+	204	329	**RNA**→ ? (GOS6223)→ COG1418 (COG1418)? (GOS3546)→
env-2	JCVI_SCAF_1096627580171	-	788	663	←? (GOS819) **RNA**→ ? (GOS6223)→
Fba-2-1	NZ_AAXX01000003.1	-	68580	68477	DUF710 (pfam05164)→ **RNA**→ COG1418 (COG1418)→
Fjo-1-1	NC_009441.1	+	128358	128466	DUF710 (pfam05164)→ **RNA**→ COG1418 (COG1418)→
Fps-1-1	NC_009613.1	+	1319608	1319714	hypo→ **RNA**→ COG1418 (COG1418)→
Ddo-1-1	NZ_AAMZ01000005.1	+	144676	144798	hypo→ **RNA**→ COG1418 (COG1418)→
Cat-1-1	NZ_AAMP01000001.1	+	1492344	1492454	hypo→ **RNA**→ COG1418 (COG1418)→ COG3956 (COG3956)→ hypo→
env-3	JCVI_SCAF_1101667205009	+	99	201	**RNA**→
env-4	JCVI_SCAF_1096626999289	+	468	571	? (GOS14297403)→ **RNA**→ ←GalK (COG0153)? (GOS427)
env-5	JCVI_SCAF_1101668678955	+	1063	1165	**RNA**→
env-6	JCVI_SCAF_1096628195137	+	483	586	? (GOS14297403)→ **RNA**→ ? (GOS399)→
env-7	JCVI_SCAF_1096628404094	-	1203	1100	? (GOS14297403)→ **RNA**→ ←? (GOS427)
env-8	JCVI_SCAF_1096627264411	+	1413	1515	? (GOS14297403)→ **RNA**→ ←? (GOS427)
env-9	JCVI_SCAF_1096627821315	+	9	112	**RNA**→
env-10	JCVI_SCAF_1101668383300	-	699	596	? (GOS14297403)→ **RNA**→ ←GalK (COG0153)? (GOS427)
env-11	JCVI_SCAF_1096627338603	+	565	668	? (GOS14297403)→ **RNA**→ ←GalK (COG0153)? (GOS427)
env-12	JCVI_SCAF_1101668386025	+	733	835	**RNA**→
env-13	JCVI_SCAF_1096627171768	+	137	240	**RNA**→ ←GalK (COG0153)? (GOS427)
env-14	JCVI_SCAF_1096627054710	-	1014	911	? (GOS14297403)→ **RNA**→ ←GalK (COG0153)? (GOS427)
env-15	JCVI_SCAF_1101668193208	+	744	847	? (GOS14297403)→ **RNA**→ ←GalK (COG0153)? (GOS427)
env-16	JCVI_SCAF_1096627335608	+	1071	1178	**RNA**→
env-17	JCVI_SCAF_1101668380480	-	494	389	? (GOS14229324)→ **RNA**→ ? (GOS6223)→
env-18	JCVI_SCAF_1096627117432	-	1990	1886	? (GOS14229324)→ **RNA**→ COG1418 (COG1418)? (GOS3546)→
env-19	JCVI_SCAF_1096627377246	-	1148	1045	**RNA**→ ? (GOS6223)→
env-20	JCVI_SCAF_1101668293410	-	1017	911	? (GOS11001734)→ **RNA**→ ←MIP (cd00333)? (GOS38)
env-21	JCVI_SCAF_1096627074548	+	918	1023	? (GOS13021950)→ **RNA**→ ←MIP (cd00333)? (GOS38)

env-22	JCVI_SCAF_1096627999429	+	1628	1734	**RNA**→
Psp-1-1	NZ_AANA01000001.1	-	543738	543632	hypo→ **RNA**→ COG1418 (COG1418)→
env-23	JCVI_SCAF_1101668090991	-	586	483	? (GOS13021950)→ **RNA**→ ? (GOS6223)→
env-24	JCVI_SCAF_1101668418675	-	371	268	**RNA**→
env-25	JCVI_SCAF_1101668709204	+	19	122	**RNA**→
env-26	JCVI_SCAF_1096627858119	-	1234	1130	DUF710 (pfam05164)? (GOS13021950)→ **RNA**→ ←? (GOS427)
env-27	JCVI_SCAF_1101668394845	+	499	603	DUF710 (pfam05164)? (GOS13021950)→ **RNA**→ ←? (GOS427)
env-28	JCVI_SCAF_1101668404018	+	803	909	**RNA**→
env-29	JCVI_SCAF_1096627168612	-	1095	989	DUF710 (pfam05164)? (GOS13021950)→ **RNA**→ ? (GOS3546)→
env-30	JCVI_SCAF_1101667286162	-	859	756	**RNA**→
Pir-1-1	NZ_AAOG01000003.1	-	188107	187981	hypo→ **RNA**→ COG1418 (COG1418)→
env-31	JCVI_SCAF_1096626676035	-	150	43	**RNA**→
env-32	JCVI_SCAF_1101668679737	+	757	863	? (GOS11001734)→ **RNA**→ ? (GOS6223)→
env-33	JCVI_SCAF_1101668337420	+	157	263	**RNA**→ COG1418 (COG1418)? (GOS3546)→
env-34	JCVI_SCAF_1096627384620	+	3535	3641	? (GOS11001734)→ **RNA**→ COG1418 (COG1418)? (GOS3546)→
Lbl-1-1	NZ_AANC01000006.1	-	17405	17297	hypo→ **RNA**→ COG1418 (COG1418)→
env-35	JCVI_SCAF_1096627496052	+	600	708	? (GOS13021950)→ **RNA**→ ? (GOS3546)→
env-36	JCVI_SCAF_1101668019466	+	1316	1423	**RNA**→
env-37	JCVI_SCAF_1096627276620	-	2625	2518	DUF710 (pfam05164)? (GOS11001734)→ **RNA**→ ←? (GOS8166)
Fba-3-1	NZ_AAOC01000004.1	+	42714	42822	hypo→ **RNA**→ COG1418 (COG1418)→
Rbi-1-1	NZ_AAOI01000003.1	+	345742	345849	hypo→ **RNA**→ COG1418 (COG1418)→
ueu-1-1	NZ_ABCO01000011.1	+	67885	68005	hypo→ **RNA**→ COG1418 (COG1418)→
Pto-1-1	NZ_AAPR01000008.1	+	9568	9681	DUF710 (pfam05164)→ **RNA**→ COG1418 (COG1418)→
env-38	JCVI_SCAF_1096627866490	-	199	92	**RNA**→
env-39	JCVI_SCAF_1096627081943	-	715	607	? (GOS11226237)→ **RNA**→ ? (GOS6223)→
env-40	JCVI_SCAF_1101668406192	-	337	229	**RNA**→
env-41	JCVI_SCAF_1096627372169	+	1137	1244	? (GOS13021950)→ **RNA**→ ? (GOS3546)→
env-42	JCVI_SCAF_1101668084529	-	593	486	? (GOS13021950)→ **RNA**→ ? (GOS3546)→
env-43	JCVI_SCAF_1101668684262	-	708	600	? (GOS11001734)→ **RNA**→ ←MIP (cd00333)? (GOS38)
env-44	JCVI_SCAF_1096627447502	-	1390	1281	**RNA**→ ? (GOS6223)→
env-45	JCVI_SCAF_1096626860696	+	794	902	? (GOS11001734)→ **RNA**→ ←MIP (cd00333)? (GOS38)
env-46	JCVI_SCAF_1096627318269	-	2667	2558	? (GOS11001734)→ **RNA**→ ←MIP (cd00333)? (GOS38)
env-47	JCVI_SCAF_1101668259956	-	1026	918	? (GOS11001734)→ **RNA**→ ←MIP (cd00333)? (GOS38)
env-48	JCVI_SCAF_1101668072492	+	15	123	**RNA**→
env-49	JCVI_SCAF_1096626909345	+	172	280	**RNA**→ COG1418 (COG1418)? (GOS3546)→
env-50	JCVI_SCAF_1101667353886	+	33	141	**RNA**→
env-51	JCVI_SCAF_1096627999061	+	1137	1245	? (GOS13021950)→ **RNA**→ ←MIP (cd00333)? (GOS38)
env-52	JCVI_SCAF_1096626292370	-	517	409	? (GOS13021950)→ **RNA**→ ←MIP (cd00333)? (GOS38)
env-53	JCVI_SCAF_1096628393098	+	669	776	? (GOS11001734)→ **RNA**→ COG1418 (COG1418)? (GOS6906)→
env-54	JCVI_SCAF_1096628290523	-	503	396	? (GOS11001734)→ **RNA**→ ? (GOS3546)→
env-55	JCVI_SCAF_1101668445732	-	630	524	**RNA**→ ←? (GOS8279)
env-56	JCVI_SCAF_1096626973177	-	2237	2131	? (GOS11440210)→ **RNA**→ ←? (GOS8279)
env-57	JCVI_SCAF_1096627033120	-	201	95	**RNA**→
env-58	JCVI_SCAF_1096627273815	+	981	1087	? (GOS11440210)→ **RNA**→ ←? (GOS8279)
env-59	JCVI_SCAF_1101668174794	+	977	1083	? (GOS11440210)→ **RNA**→ ←? (GOS8279)
env-60	JCVI_SCAF_1096627172476	-	756	650	? (GOS11440210)→ **RNA**→ ←? (GOS8279)
env-61	JCVI_SCAF_1101668074176	-	372	266	**RNA**→
env-62	JCVI_SCAF_1096626853002	+	1470	1576	? (GOS11001734)→ **RNA**→ ←? (GOS8279)
env-63	JCVI_SCAF_1096627219240	+	1647	1751	? (GOS14483285)→ **RNA**→ ←? (GOS8279)
env-64	JCVI_SCAF_1096628003170	+	183	287	**RNA**→ ←? (GOS8279)
env-65	JCVI_SCAF_1096627276151	-	547	443	? (GOS14483285)→ **RNA**→ ←? (GOS8279)
env-66	JCVI_SCAF_1101667344488	-	491	387	? (GOS14483285)→ **RNA**→ ←? (GOS8279)

env-67	JCVI_SCAF_1096627107045	-	1479	1375	? (GOS14483285)→ **RNA**→ ←? (GOS8279)
env-68	JCVI_SCAF_1096627107851	+	1008	1112	? (GOS14483285)→ **RNA**→ ←? (GOS8279)
env-69	JCVI_SCAF_1101668544692	-	1224	1118	**RNA**→ ←? (GOS8279)
env-70	JCVI_SCAF_1101667566873	-	455	349	? (GOS14578054)→ **RNA**→ ←? (GOS8279)
env-71	JCVI_SCAF_1096627260888	+	18194	18300	? (GOS14578054)→ **RNA**→ ←? (GOS8279)
env-72	JCVI_SCAF_1096627328300	+	3516	3625	**RNA**→
env-73	JCVI_SCAF_1096628305661	+	1761	1870	**RNA**→
Fba-1-1	NZ_AAPD01000002.1	-	349281	349171	hypo→ **RNA**→ COG1418 (COG1418)→
env-74	JCVI_SCAF_1101668047041	-	1267	1140	**RNA**→ COG1094 (COG1094)? (GOS3546)→
[1]env-75	JCVI_SCAF_1096627350108	-	3515	3405	? (GOS13021950)→ **RNA**→ COG1418 (COG1418)? (GOS3546)→
env-76	JCVI_SCAF_1096627431273	-	1609	1478	? (GOS13021950)→ **RNA**→ ? (GOS6223)→ COG1418 (COG1418)? (GOS3546)→

1.4 Conserved domains

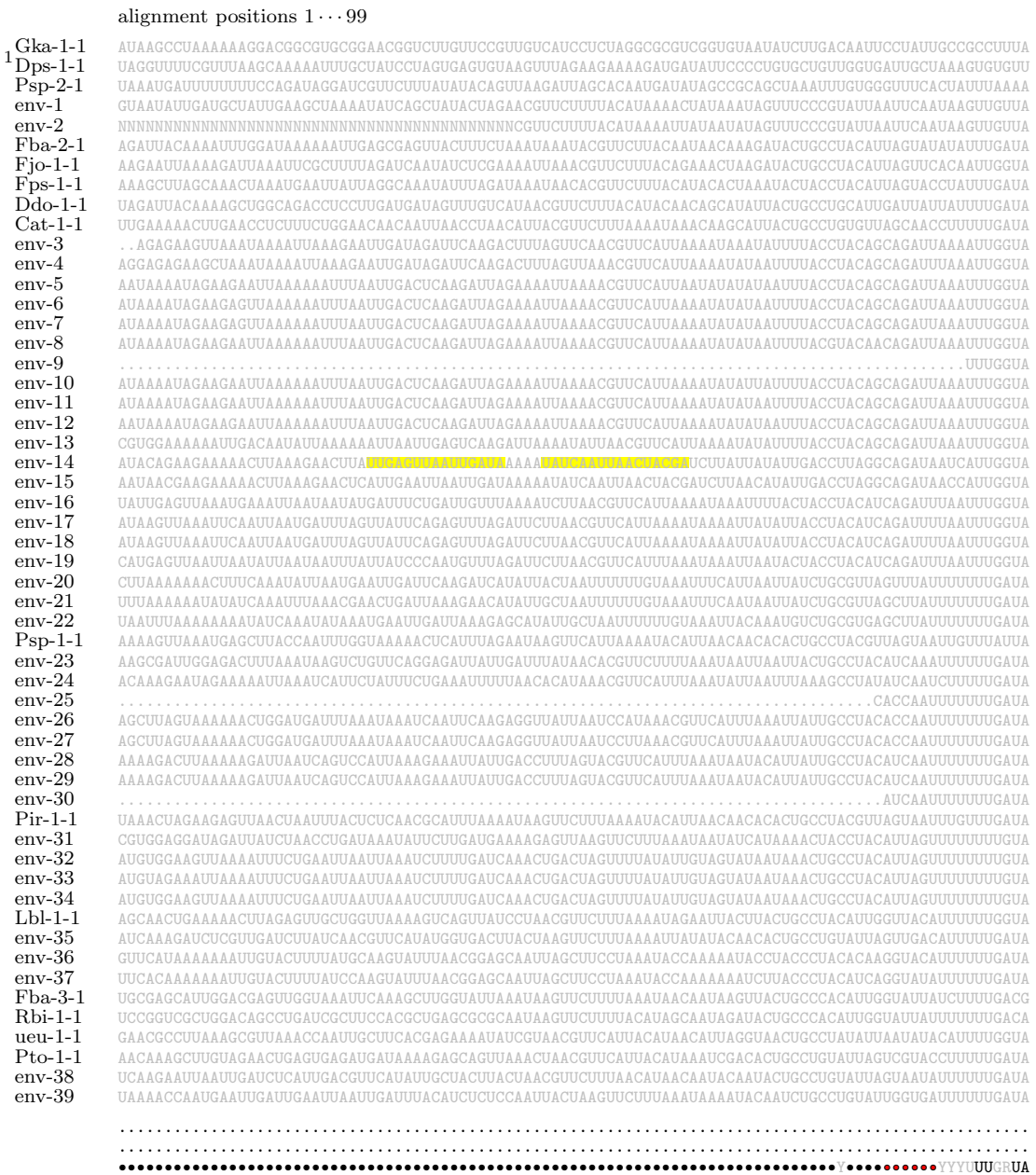
Conserved domains found in protein-coding genes listed in Section 1.3 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one 6S-flavo RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

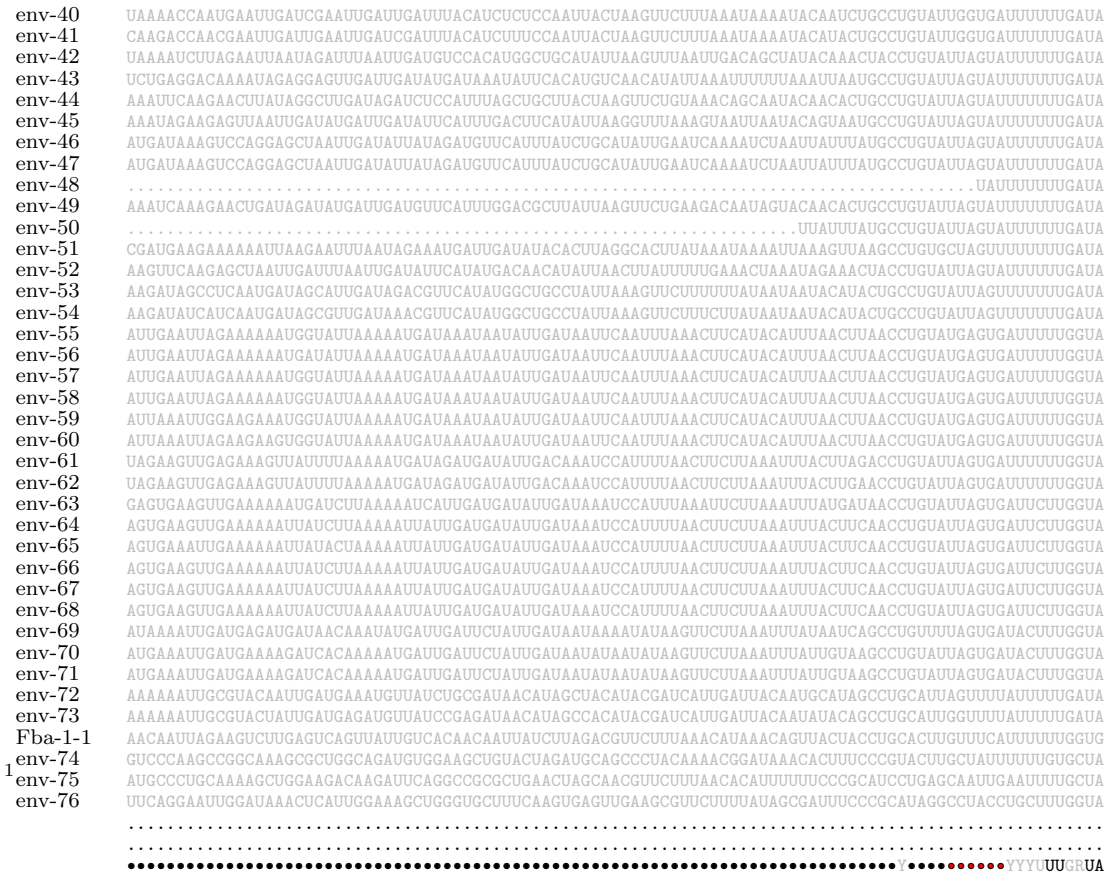
u cd00085 (1) HNH nucleases; HNH endonuclease signature which is found in viral, prokaryotic, and eukaryotic proteins.
d cd00333 (8) Major intrinsic protein (MIP) superfamily.
d COG0153 (6) Galactokinase [Carbohydrate transport and metabolism]
D COG0162 (1) Tyrosyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
D COG1094 (2) Predicted RNA-binding protein (contains KH domains) [General function prediction only]
D COG1418 (24) Predicted HD superfamily hydrolase [General function prediction only]
D COG3956 (1) Protein containing tetrapyrrole methyltransferase domain and MazG-like (predicted pyrophosphatase) domain [General function prediction only]
u GOS11001734 (11) ?
u GOS11226237 (1) ?
u GOS11440210 (4) ?
u GOS13021950 (12) ?
u GOS14229324 (2) ?
u GOS14297403 (8) ?
u GOS14483285 (5) ?
u GOS14578054 (2) ?
D GOS3546 (13) ?
d GOS38 (8) ?
D GOS399 (1) ?
d GOS427 (10) ?
D GOS6223 (9) ?
D GOS6906 (1) ?
d GOS8166 (1) ?
u GOS819 (1) ?
d GOS8279 (15) ?
u pfam00078 (1) Reverse transcriptase (RNA-dependent DNA polymerase).
u pfam05164 (8) Family of unknown function (DUF710).
D pfam06127 (1) Protein of unknown function (DUF962).

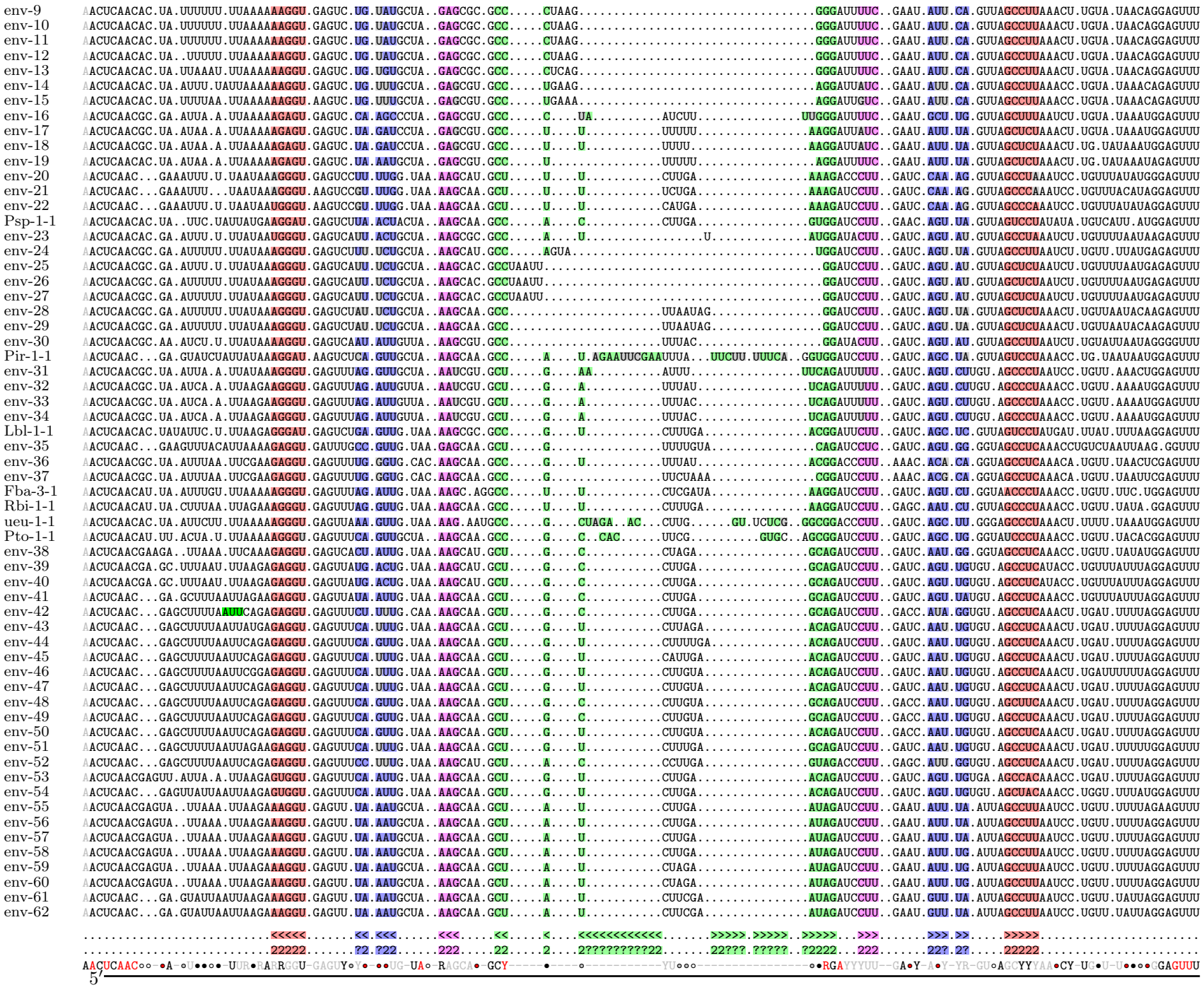
1.5 Multiple-sequence alignment

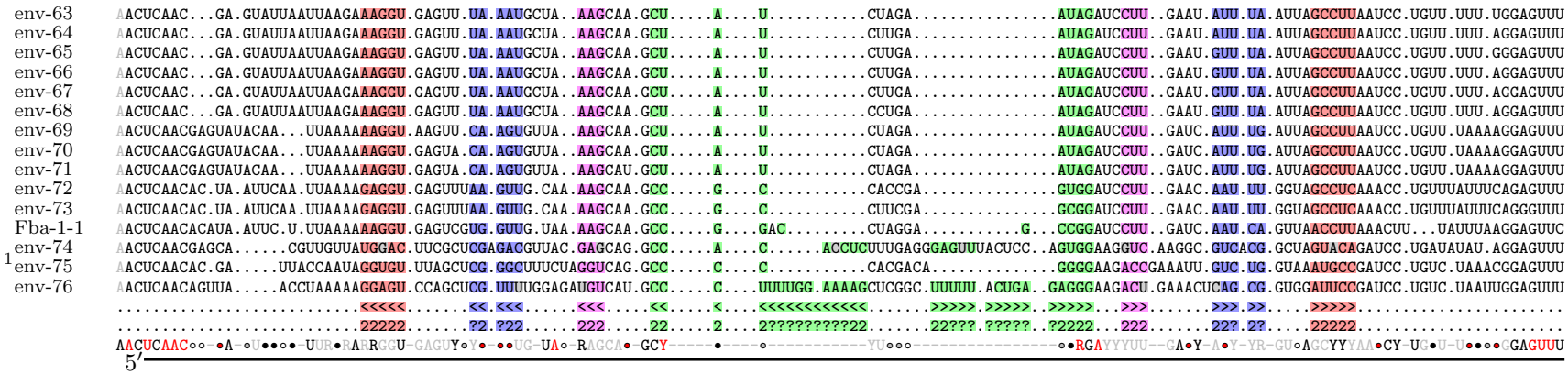
Each 6S-flavo RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 1.1. Superscript numbers refer to annotations listed in Section 1.2. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The 6S-flavo RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" =

"A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

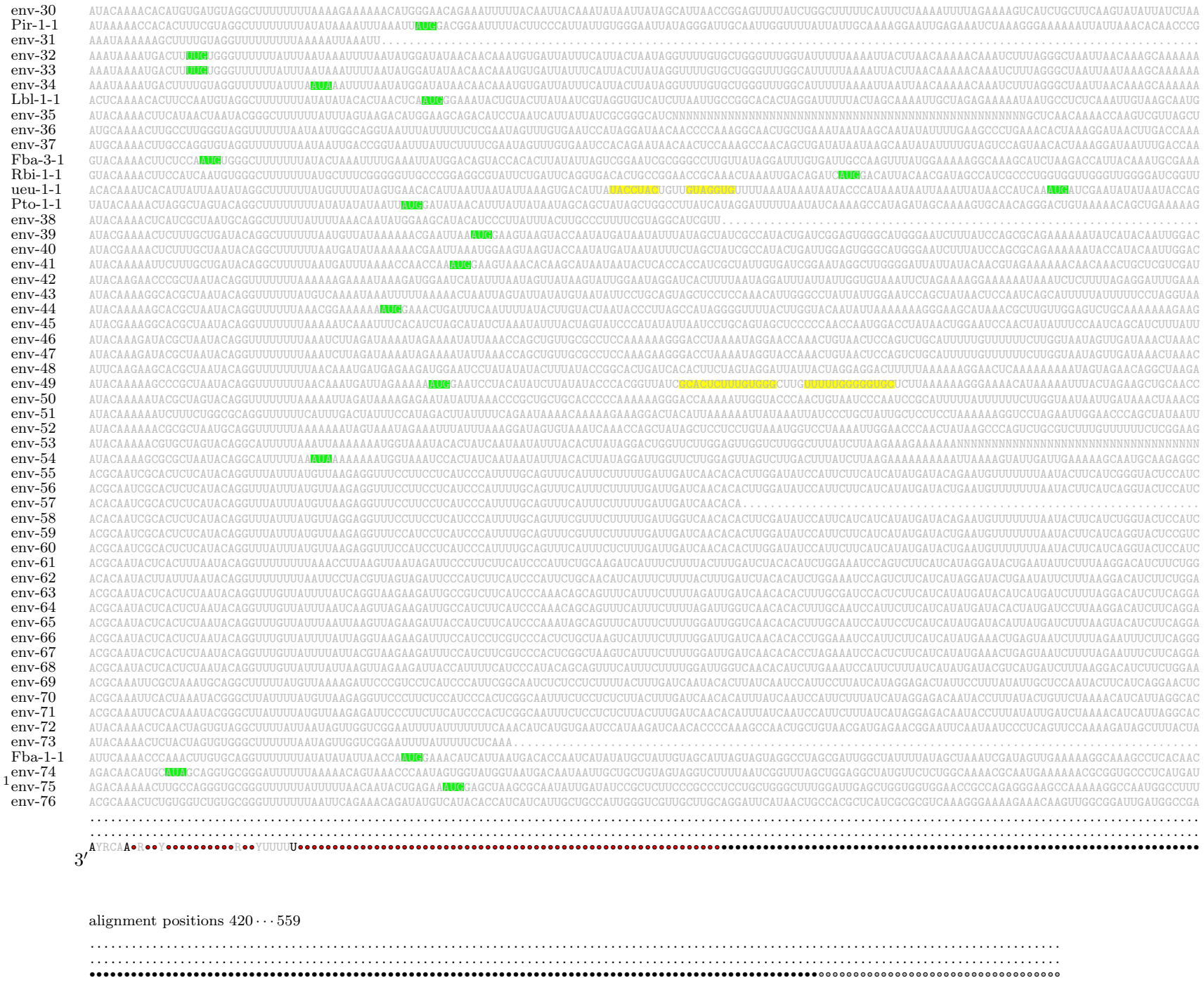



env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
Fba-1-1
env-74
env-75
env-76
alignment positions 100 · · · 259
Gka-1-1
Dps-1-1
Psp-2-1
env-1
env-2
Fba-2-1
Fjo-1-1
Fps-1-1
Ddo-1-1
Cat-1-1
env-3
env-4
env-5
env-6
env-7
env-8
5′


env-9
env-10
env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
Psp-1-1
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
Pir-1-1
env-31
env-32
env-33
env-34
Lbl-1-1
env-35
env-36
env-37
Fba-3-1
Rbi-1-1
ueu-1-1
Pto-1-1
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
5′



alignment positions 260 · · · 419


env-30
Pir-1-1
env-31
env-32
env-33
env-34
Lbl-1-1
env-35
env-36
env-37
Fba-3-1
Rbi-1-1
ueu-1-1
Pto-1-1
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
Fba-1-1
env-74
env-75
env-76
1
3′
alignment positions 420 · · · 559

Gka-1-1
Dps-1-1
Psp-2-1
env-1
env-2
Fba-2-1
Fjo-1-1
Fps-1-1
Ddo-1-1
Cat-1-1
env-3
env-4
env-5
env-6
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
Psp-1-1
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
Pir-1-1
env-31
env-32
env-33
env-34
Lbl-1-1
env-35
env-36
env-37
Fba-3-1
Rbi-1-1
ueu-1-1
Pto-1-1
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47

2 *aceE*

2.1 Taxa

The taxonomy of each organism containing a putative *aceE* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *aceE* RNA in Sections 2.2 and 2.4.

abbrev. of hits	taxonomy of species
Avi-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Azotobacter vinelandii*
Pae-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* 2192
Pae-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* C3719
Pae-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PA7
Pae-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PACS2
Pae-5-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PAO1
Pae-6-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* UCBPP-PA14
Pen-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas entomophila* L48
Pfl-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf-5
Pfl-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf0-1
Pme-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas mendocina ymp*
Ppu-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* F1
Ppu-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* GB-1
Ppu-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* KT2440
Ppu-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* W619
Pst-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas stutzeri* A1501

Psy-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. phaseolicola* 1448A
Psy-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000
env-1	environmental sample

2.2 Gene contexts

Each *aceE* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *aceE* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 2.3. The accession number of the sequence containing each *aceE* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 2.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Avi-1-1	NZ_AAAU03000012.1	-	97225	97146	**RNA**→ TPP_E1_EcPDC_like (cd02017)→ AceF (COG0508)Biotin_lipoyl (pfam00364)→
Pae-3-1	NC_009656.1	+	5928905	5928986	**RNA**→ COG3958 (COG3958)TPP_E1_EcPDC_like (cd02017)→
Pae-5-1	NC_002516.2	+	5636028	5636109	**RNA**→ COG3958 (COG3958)TPP_E1_EcPDC_like (cd02017)→
Pae-2-1	NZ_AAKV01000094.1	-	140786	140705	**RNA**→ COG3958 (COG3958)TPP_E1_EcPDC_like (cd02017)→
Pae-4-1	NZ_AAQW01000001.1	+	5844740	5844821	**RNA**→ COG3958 (COG3958)TPP_E1_EcPDC_like (cd02017)→
Pae-1-1	NZ_AAKW01000064.1	-	169964	169883	**RNA**→ COG3958 (COG3958)TPP_E1_EcPDC_like (cd02017)→
Pae-6-1	NC_008463.1	+	5907425	5907506	**RNA**→ COG3958 (COG3958)TPP_E1_EcPDC_like (cd02017)→
Pme-1-1	NC_009439.1	-	633058	632979	**RNA**→ COG3958 (COG3958)TPP_E1_EcPDC_like (cd02017)→ AceF (COG0508)Biotin_lipoyl (pfam00364)→
Pst-1-1	NC_009434.1	+	4168863	4168945	**RNA**→ TPP_E1_EcPDC_like (cd02017)→ AceF (COG0508)Biotin_lipoyl (pfam00364)→
env-1	soil_Contig41332	+	1133	1211	**RNA**→
Pfl-2-1	NC_007492.1	-	536189	536111	**RNA**→ AceE (COG2609)→ AceF (COG0508)Biotin_lipoyl (pfam00364)→
Pfl-1-1	NC_004129.6	-	593917	593840	**RNA**→ AceE (COG2609)→ AceF (COG0508)Biotin_lipoyl (pfam00364)→
Psy-2-1	NC_007005.1	-	579052	578973	**RNA**→ AceE (COG2609)→
Psy-3-1	NC_004578.1	+	5686568	5686647	**RNA**→ AceE (COG2609)→
Psy-1-1	NC_005773.3	-	593005	592926	**RNA**→ AceE (COG2609)→
Ppu-4-1	NZ_AAVY01000020.1	+	111124	111201	**RNA**→ ←hypo (cons)
Pen-1-1	NC_008027.1	+	5464828	5464905	**RNA**→ AceE (COG2609)→ AceF (COG0508)Biotin_lipoyl (pfam00364)→
Ppu-1-1	NC_009512.1	-	420581	420504	**RNA**→ AceE (COG2609)→ AceF (COG0508)Biotin_lipoyl (pfam00364)→
Ppu-3-1	NC_002947.3	-	410786	410709	**RNA**→ AceE (COG2609)→ AceF (COG0508)Biotin_lipoyl (pfam00364)→
Ppu-2-1	NZ_AAXR01000011.1	+	22482	22559	**RNA**→ ←hypo (cons)

2.3 Conserved domains

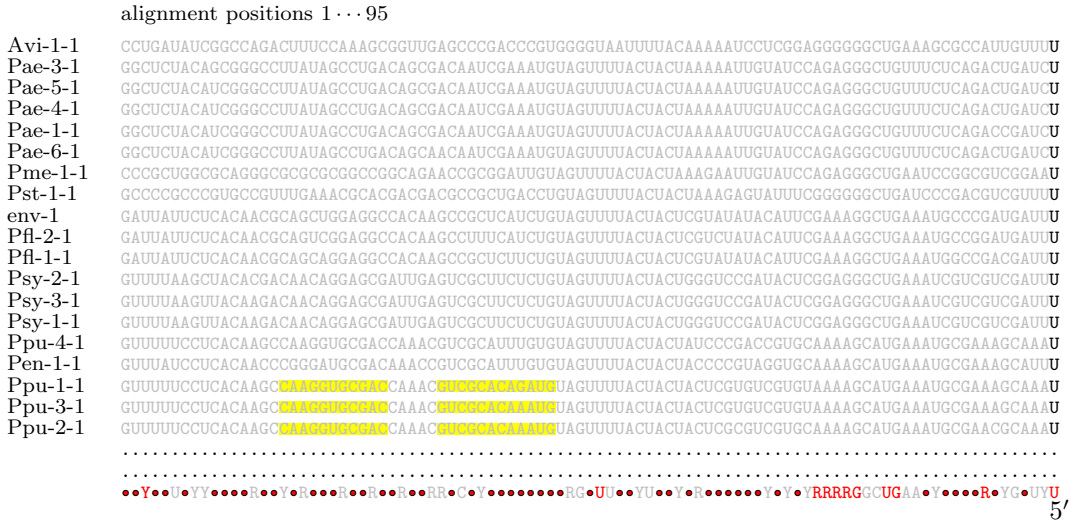
Conserved domains found in protein-coding genes listed in Section 2.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *aceE* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

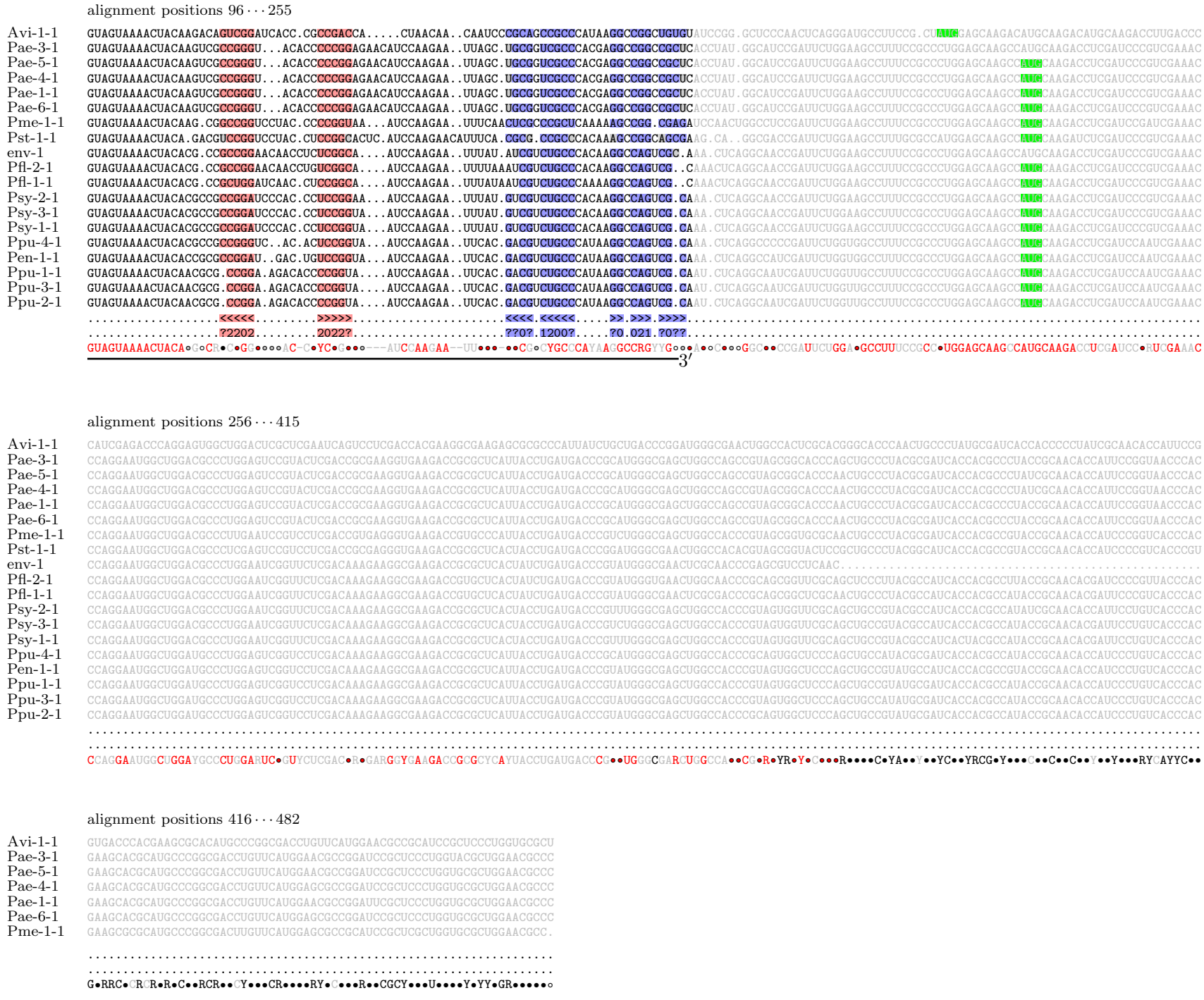
cd02017 (9) Thiamine pyrophosphate (TPP) family, E1 of E.
COG0508 (8) Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component, and related enzymes [Energy production and conversion]
COG2609 (8) Pyruvate dehydrogenase complex, dehydrogenase (E1) component [Energy production and conversion]
COG3958 (7) Transketolase, C-terminal subunit [Carbohydrate transport and metabolism]
pfam00364 (8) Biotin-requiring enzyme.

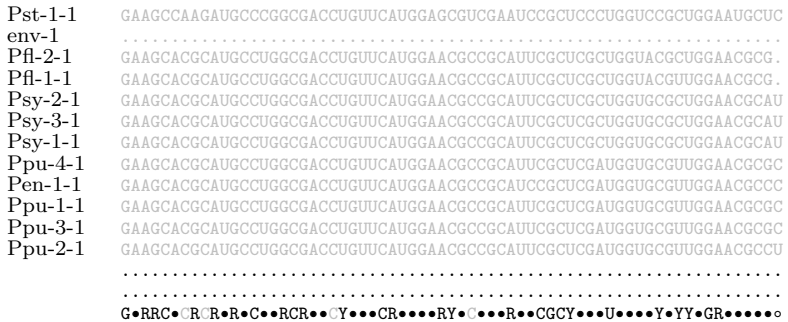
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Pae-2-1

2.4 Multiple-sequence alignment

Each *aceE* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 2.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *aceE* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.




alignment positions 96 · · · 255
Avi-1-1
Pae-3-1
Pae-5-1
Pae-4-1
Pae-1-1
Pae-6-1
Pme-1-1
Pst-1-1
env-1
Pfl-2-1
Pfl-1-1
Psy-2-1
Psy-3-1
Psy-1-1
Ppu-4-1
Pen-1-1
Ppu-1-1
Ppu-3-1
Ppu-2-1
3′
alignment positions 256 · · · 415
Avi-1-1
Pae-3-1
Pae-5-1
Pae-4-1
Pae-1-1
Pae-6-1
Pme-1-1
Pst-1-1
env-1
Pfl-2-1
Pfl-1-1
Psy-2-1
Psy-3-1
Psy-1-1
Ppu-4-1
Pen-1-1
Ppu-1-1
Ppu-3-1
Ppu-2-1
alignment positions 416 · · · 482
Avi-1-1
Pae-3-1
Pae-5-1
Pae-4-1
Pae-1-1
Pae-6-1
Pme-1-1



3 Acido-1

3.1 Taxa

The taxonomy of each organism containing a putative Acido-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Acido-1 RNA in Sections 3.2 and 3.4.

abbrev. of hits	taxonomy of species
Aba-1-1	Bacteria Acidobacteria Acidobacteriales Acidobacteriaceae *Acidobacteria bacterium* Ellin345
Sus-1-1 to Sus-1-3	Bacteria Acidobacteria Solibacteres Solibacterales Solibacteraceae *Solibacter usitatus* Ellin6076

3.2 Gene contexts

Each Acido-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Acido-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 3.3. The accession number of the sequence containing each Acido-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 3.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Aba-1-1	NC_008009.1	-	235844	235785	AcnA_IRP_Swivel (cd01580)AcnA_IRP (cd01586)→ **RNA**→ SUL1 (COG0659)→
Sus-1-1	NC_008536.1	+	3413964	3414030	hypo→ **RNA**→ ←MecR1 (COG4219)
Sus-1-2	NC_008536.1	+	3480719	3480784	hypo→ **RNA**→ TPR (cd00189)→
Sus-1-3	NC_008536.1	+	3481695	3481760	TPR (cd00189)→ **RNA**→ type IV pilus assembly PilZ→

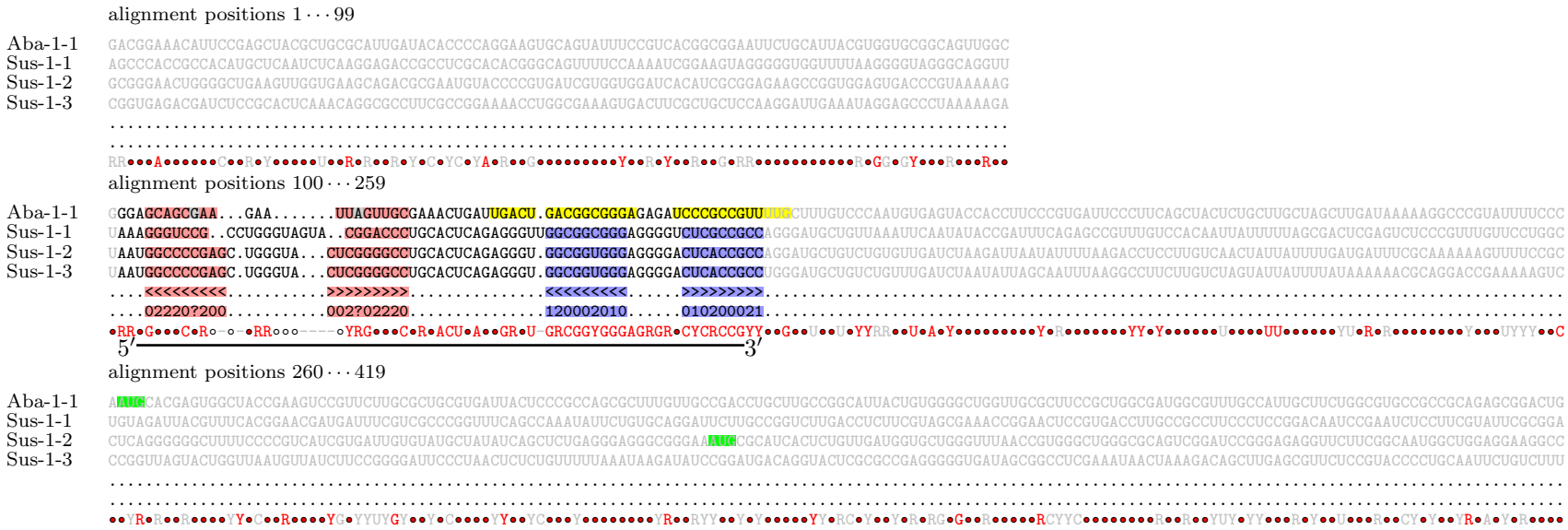
3.3 Conserved domains

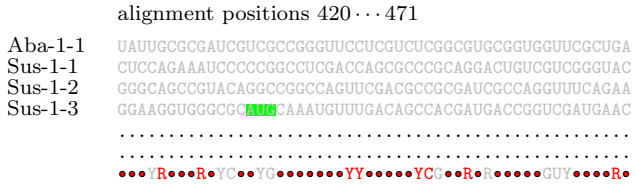
Conserved domains found in protein-coding genes listed in Section 3.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Acido-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

uD cd00189 (2) Tetratricopeptide repeat domain; typically contains 34 amino acids [WLF]-X(2)-[LIM]-[GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-X(2)-[ASL]-X(4)-[PKE]
u cd01580 (1) Aconitase A swivel domain.
u cd01586 (1) Aconitase A catalytic domain.
D COG0659 (1) Sulfate permease and related transporters (MFS superfamily) [Inorganic ion transport and metabolism]
d COG4219 (1) Antirepressor regulating drug resistance, predicted signal transduction N-terminal membrane component [Transcription / Signal transduction mechanisms]

3.4 Multiple-sequence alignment

Each Acido-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 3.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Acido-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.





4 Acido-Lenti-1

4.1 Taxa

The taxonomy of each organism containing a putative Acido-Lenti-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Acido-Lenti-1 RNA in Sections 4.2 and 4.4.

abbrev. of hits	**taxonomy of species**
Aba-1-1 to Aba-1-11	Bacteria Acidobacteria Acidobacteriales Acidobacteriaceae *Acidobacteria bacterium* Ellin345
Lar-1-1 to Lar-1-56	Bacteria Lentisphaerae Lentisphaerales Lentisphaeraceae *Lentisphaera araneosa* HTCC2155
Rba-1-1 to Rba-1-2	Bacteria Planctomycetes Planctomycetacia Planctomycetales Planctomycetaceae *Rhodopirellula baltica* SH 1
Gur-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Geobacteraceae *Geobacter uraniireducens* Rf4
env-1 to env-19	environmental samples

4.2 Gene contexts

Each Acido-Lenti-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Acido-Lenti-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 4.3. The accession number of the sequence containing each Acido-Lenti-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 4.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Lar-1-1	NZ_ABCK01000026.1	+	37913	37958	hypo→ **RNA**→ **RNA**→ ←PDZ_serine_protease (cd00987)
Lar-1-2	NZ_ABCK01000026.1	+	37713	37758	(shown above)
Lar-1-3	NZ_ABCK01000029.1	-	8929	8884	←AslA (COG3119) **RNA**→ ←hypo
Aba-1-1	NC_008009.1	-	1267162	1267117	←hypo **RNA**→ ←SpoU (COG0566)
Lar-1-4	NZ_ABCK01000007.1	-	4868	4823	COG1073 (COG1073)COG1647 (COG1647)→ **RNA**→ ←rve (pfam00665)
Lar-1-5	NZ_ABCK01000002.1	-	270653	270608	←AslA (COG3119) **RNA**→ **RNA**→ **RNA**→ ←COG1413 (COG1413)COG2133 (COG2133)
Lar-1-6	NZ_ABCK01000002.1	-	270461	270416	(shown above)
Lar-1-7	NZ_ABCK01000002.1	-	270845	270800	(shown above)
Lar-1-8	NZ_ABCK01000002.1	+	237476	237521	hypo→ **RNA**→ **RNA**→ AslA (COG3119)→
Lar-1-9	NZ_ABCK01000002.1	+	237916	237961	(shown above)
Lar-1-10	NZ_ABCK01000002.1	+	217312	217357	hypo→ **RNA**→ COG2133 (COG2133)DUF1080 (pfam06439)→
Lar-1-11	NZ_ABCK01000002.1	+	210449	210494	PulG (COG2165)→ **RNA**→ ←RVT (pfam00078)
Lar-1-12	NZ_ABCK01000002.1	+	212552	212597	←hypo **RNA**→ hypo→
Lar-1-13	NZ_ABCK01000005.1	-	275464	275419	RVT (pfam00078)→ **RNA**→ ←AslA (COG3119)
Lar-1-14	NZ_ABCK01000001.1	-	419681	419636	hypo→ **RNA**→ hypo→
Lar-1-15	NZ_ABCK01000012.1	+	53466	53511	EFh (cd00051)→ **RNA**→ hypo→
Lar-1-16	NZ_ABCK01000012.1	-	104879	104834	←hypo **RNA**→ ←AslA (COG3119)
Lar-1-17	NZ_ABCK01000012.1	-	106309	106264	←hypo **RNA**→ ←hypo
Lar-1-18	NZ_ABCK01000010.1	+	121151	121196	hypo→ **RNA**→ **RNA**→ **RNA**→ hypo→
Lar-1-19	NZ_ABCK01000010.1	+	120473	120518	(shown above)
Lar-1-20	NZ_ABCK01000010.1	+	120955	121000	(shown above)
Lar-1-21	NZ_ABCK01000006.1	+	14314	14359	←AslA (COG3119) **RNA**→ **RNA**→ **RNA**→ ←RpoE (COG1595)
Lar-1-22	NZ_ABCK01000006.1	+	14204	14249	(shown above)
Lar-1-23	NZ_ABCK01000006.1	+	14425	14470	(shown above)
Lar-1-24	NZ_ABCK01000001.1	-	416780	416735	COG1943 (COG1943)→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ ←COG3618 (COG3618)
Lar-1-25	NZ_ABCK01000001.1	-	416706	416661	(shown above)
Lar-1-29	NZ_ABCK01000001.1	-	416632	416587	(shown above)
Lar-1-31	NZ_ABCK01000001.1	-	416558	416513	(shown above)
Lar-1-26	NZ_ABCK01000001.1	-	431760	431715	←hypo **RNA**→ **RNA**→ ←vWFA (cd00198)
Lar-1-27	NZ_ABCK01000001.1	-	431565	431520	(shown above)
Lar-1-28	NZ_ABCK01000080.1	+	409	454	hypo→ **RNA**→ hypo→
Lar-1-30	NZ_ABCK01000001.1	-	417523	417478	RVT (pfam00078)→ **RNA**→ COG1943 (COG1943)→
Lar-1-32	NZ_ABCK01000006.1	+	103015	103060	←hypo **RNA**→ hypo→
Lar-1-33	NZ_ABCK01000006.1	+	85512	85557	COG0400 (COG0400)→ **RNA**→ **RNA**→ ←AslA (COG3119)
Lar-1-34	NZ_ABCK01000006.1	+	85251	85296	(shown above)
Lar-1-35	NZ_ABCK01000006.1	+	107180	107225	RVT (pfam00078)→ **RNA**→ ←AslA (COG3119)
Lar-1-36	NZ_ABCK01000016.1	+	16051	16096	←probable cytochrome c precursor **RNA**→ **RNA**→ ←COG3828 (COG3828)
Lar-1-37	NZ_ABCK01000016.1	+	16291	16336	(shown above)
Lar-1-38	NZ_ABCK01000007.1	-	84866	84821	←COG1413 (COG1413)COG2133 (COG2133) **RNA**→ **RNA**→ ←tRNA_Me_trans (pfam03054)
Lar-1-40	NZ_ABCK01000007.1	-	85099	85054	(shown above)
Lar-1-39	NZ_ABCK01000006.1	+	115823	115868	←COG3533 (COG3533) **RNA**→ hypo→
Lar-1-41	NZ_ABCK01000004.1	+	95064	95109	AslA (COG3119)→ **RNA**→ ←Metallophos (pfam00149)
Lar-1-42	NZ_ABCK01000001.1	-	246374	246329	←AslA (COG3119) **RNA**→ **RNA**→ ←RVT (pfam00078)
Lar-1-47	NZ_ABCK01000001.1	-	246515	246470	(shown above)
Lar-1-43	NZ_ABCK01000020.1	+	22700	22745	VanZ (pfam04892)→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ WcaJ (COG2148)→
Lar-1-44	NZ_ABCK01000020.1	+	21790	21835	(shown above)
Lar-1-45	NZ_ABCK01000020.1	+	22585	22630	(shown above)
Lar-1-46	NZ_ABCK01000020.1	+	22470	22515	(shown above)
Lar-1-48	NZ_ABCK01000006.1	+	178011	178056	ThiF_MoeB_HesA_family (cd00757)→ **RNA**→ ←hypo
Lar-1-49	NZ_ABCK01000029.1	+	14774	14819	RVT (pfam00078)→ **RNA**→ ←hypo
Lar-1-50	NZ_ABCK01000001.1	+	226211	226256	LamGL (smart00560)→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ hypo→
Lar-1-51	NZ_ABCK01000001.1	+	226073	226118	(shown above)
Lar-1-52	NZ_ABCK01000001.1	+	226142	226187	(shown above)
Lar-1-53	NZ_ABCK01000001.1	+	225878	225923	(shown above)
env-1	JCVI_SCAF_1096628177727	+	654	698	**RNA**→

4.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 4.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Acido-Lenti-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

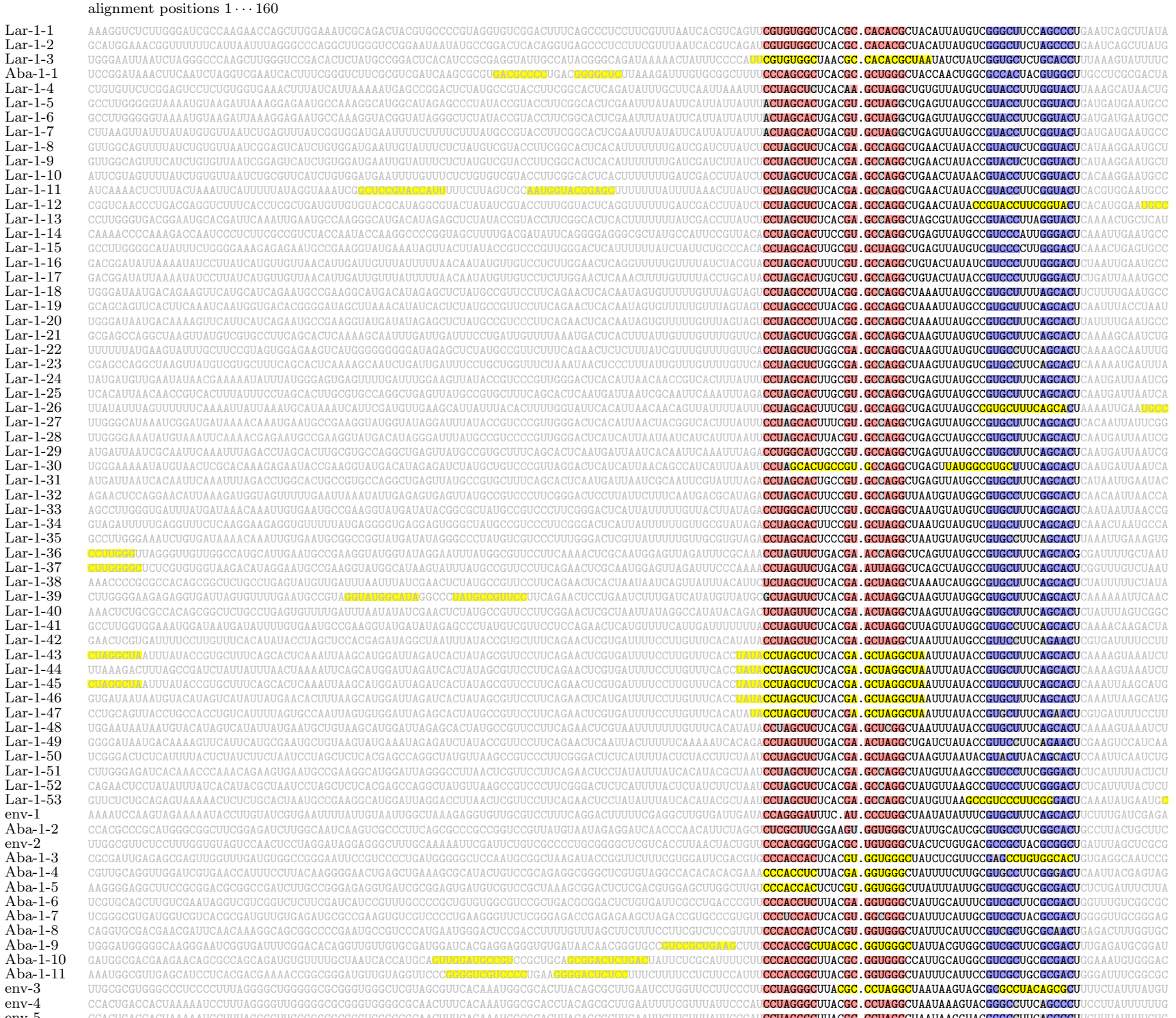
u cd00051 (1) EF-hand, calcium binding motif; A diverse superfamily of calcium sensors and calcium signal modulators; most examples in this alignment model have 2 active canonical EF hands.
d cd00189 (1) Tetratricopeptide repeat domain; typically contains 34 amino acids [WLF]-X(2)-[LIM]-[GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-X(2)-[ASL]-X(4)-[PKE]
d cd00198 (1) Von Willebrand factor type A (vWA) domain was originally found in the blood coagulation protein von Willebrand factor (vWF).
u cd00757 (1) ThiF_MoeB_HesA.
d cd00987 (1) PDZ domain of tryspin-like serine proteases, such as DegP/HtrA, which are oligomeric proteins involved in heat-shock response, chaperone function, and apoptosis.
u d cd02953 (2) DsbD gamma family; DsbD gamma is the C-terminal periplasmic domain of the bacterial protein DsbD.
D COG0021 (1) Transketolase [Carbohydrate transport and metabolism]
u COG0034 (1) Glutamine phosphoribosylpyrophosphate amidotransferase [Nucleotide transport and metabolism]
d COG0046 (1) Phosphoribosylformylglycinamidine (FGAM) synthase, synthetase domain [Nucleotide transport and metabolism]
d COG0150 (1) Phosphoribosylaminoimidazole (AIR) synthetase [Nucleotide transport and metabolism]
d COG0235 (1) Ribulose-5-phosphate 4-epimerase and related epimerases and aldolases [Carbohydrate transport and metabolism]
u COG0312 (1) Predicted Zn-dependent proteases and their inactivated homologs [General function prediction only]
u COG0352 (1) Thiamine monophosphate synthase [Coenzyme metabolism]
u COG0400 (1) Predicted esterase [General function prediction only]
u COG0529 (1) Adenylylsulfate kinase and related kinases [Inorganic ion transport and metabolism]
d COG0566 (1) rRNA methylases [Translation, ribosomal structure and biogenesis]
u COG0677 (1) UDP-N-acetyl-D-mannosaminuronate dehydrogenase [Cell envelope biogenesis, outer membrane]
d COG1004 (1) Predicted UDP-glucose 6-dehydrogenase [Cell envelope biogenesis, outer membrane]
u COG1073 (1) Hydrolases of the alpha/beta superfamily [General function prediction only]
u d COG1413 (2) FOG: HEAT repeat [Energy production and conversion]
u COG1506 (1) Dipeptidyl aminopeptidases/acylaminoacyl-peptidases [Amino acid transport and metabolism]
d COG1595 (1) DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog [Transcription]
d COG1620 (1) L-lactate permease [Energy production and conversion]
u COG1647 (1) Esterase/lipase [General function prediction only]
uD COG1943 (2) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
u COG2085 (1) Predicted dinucleotide-binding enzymes [General function prediction only]
uDd COG2133 (3) Glucose/sorbosone dehydrogenases [Carbohydrate transport and metabolism]
D COG2148 (1) Sugar transferases involved in lipopolysaccharide synthesis [Cell envelope biogenesis, outer membrane]
u COG2165 (1) Type II secretory pathway, pseudopilin PulG [Cell motility and secretion / Intracellular trafficking and secretion]
u COG2814 (1) Arabinose efflux permease [Carbohydrate transport and metabolism]
d COG2877 (1) 3-deoxy-D-manno-octulosonic acid (KDO) 8-phosphate synthase [Cell envelope biogenesis, outer membrane]
uDd COG3119 (10) Arylsulfatase A and related enzymes [Inorganic ion transport and metabolism]
d COG3205 (1) Predicted membrane protein [Function unknown]
u COG3533 (1) Uncharacterized protein conserved in bacteria [Function unknown]
d COG3550 (1) Uncharacterized protein related to capsule biosynthesis enzymes [General function prediction only]
d COG3577 (1) Predicted aspartyl protease [General function prediction only]
d COG3618 (1) Predicted metal-dependent hydrolase of the TIM-barrel fold [General function prediction only]
u COG3666 (1) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
u d COG3828 (2) Uncharacterized protein conserved in bacteria [Function unknown]
u COG4972 (1) Tfp pilus assembly protein, ATPase PilM [Cell motility and secretion / Intracellular trafficking and secretion]
u GOS1134 (1) ?
d GOS11523 (1) ?
u GOS11714918 (2) ?
u d GOS1212 (4) ?
d GOS1254 (1) ?
D GOS13384922 (1) ?
d GOS2400 (1) ?
u GOS4150 (1) ?
d NOG30274 (1)
u d pfam00078 (6) Reverse transcriptase (RNA-dependent DNA polymerase).
d pfam00149 (1) Calcineurin-like phosphoesterase.
u pfam00551 (1) Formyl transferase.
d pfam00665 (1) Integrase core domain.

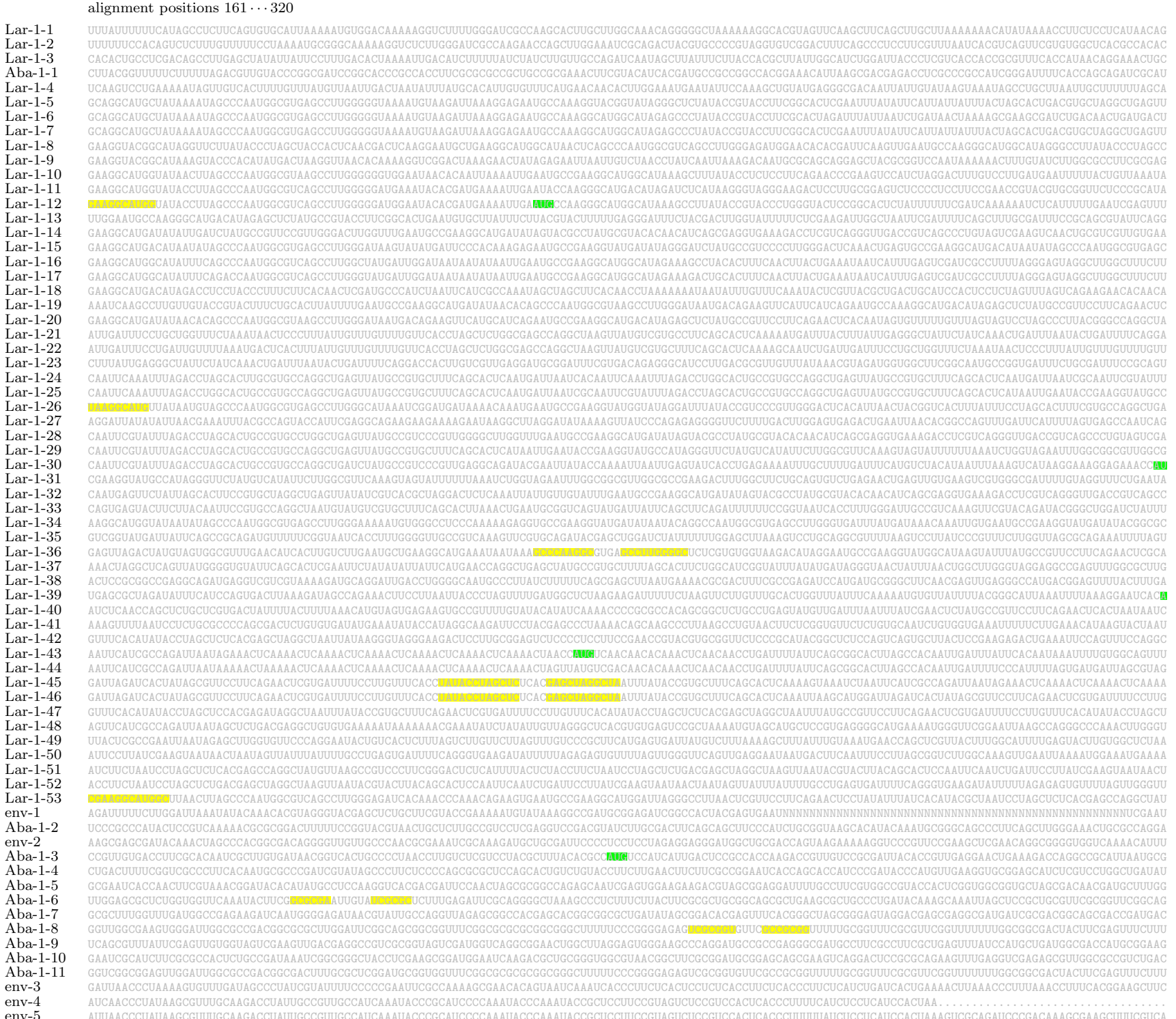
d pfam01507 (1) Phosphoadenosine phosphosulfate reductase family.
d pfam03054 (1) tRNA methyl transferase.
d pfam03720 (2) UDP-glucose/GDP-mannose dehydrogenase family, UDP binding domain.
u pfam04892 (1) VanZ like family.
D pfam06439 (1) Domain of Unknown Function (DUF1080).
u smart00560 (1) LamG-like jellyroll fold domain;

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: env-8, env-9

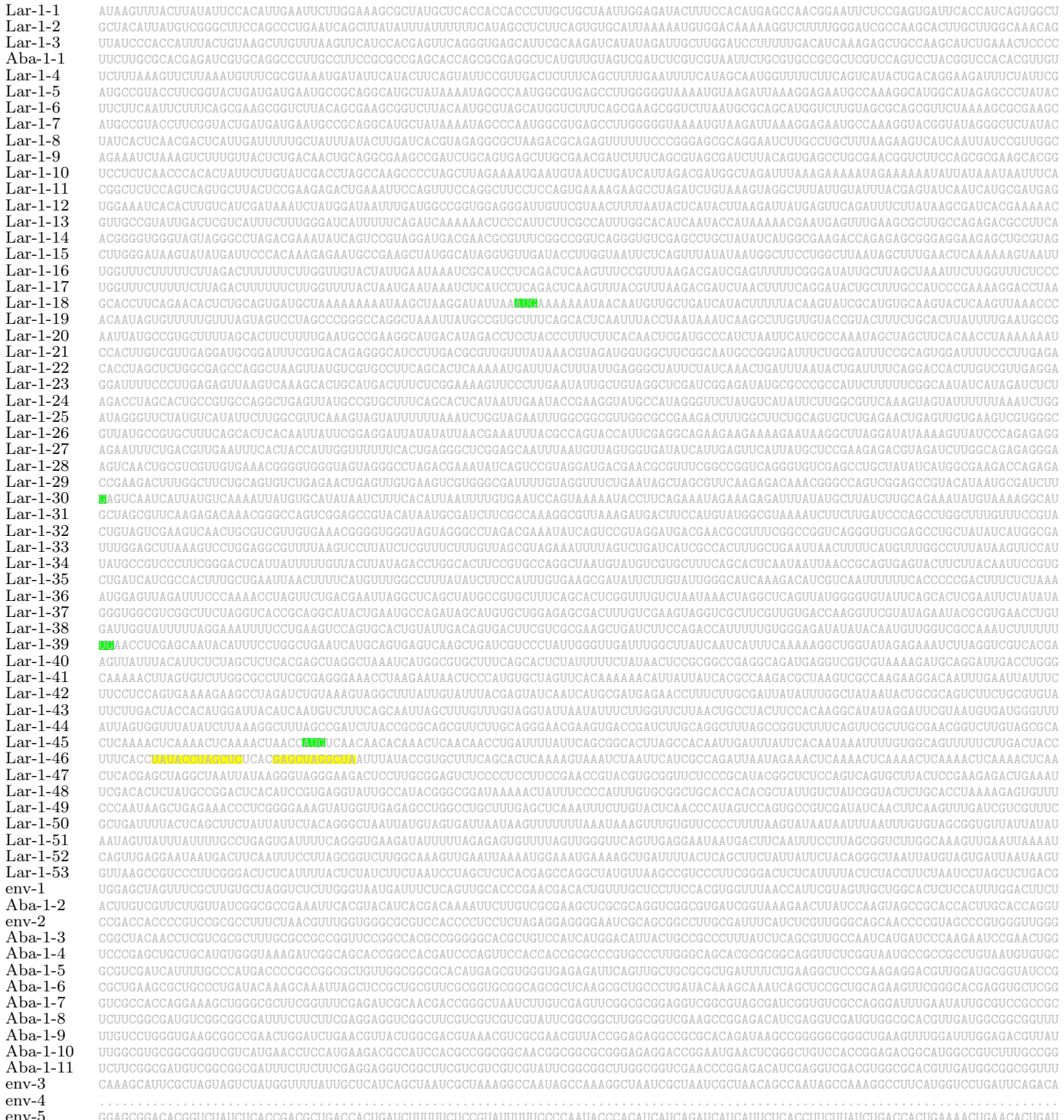
4.4 Multiple-sequence alignment

Each Acido-Lenti-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 4.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Acido-Lenti-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.


alignment positions 1 · · · 160
Lar-1-1
Lar-1-2
Lar-1-3
Aba-1-1
Lar-1-4
Lar-1-5
Lar-1-6
Lar-1-7
Lar-1-8
Lar-1-9
Lar-1-10
Lar-1-11
Lar-1-12
Lar-1-13
Lar-1-14
Lar-1-15
Lar-1-16
Lar-1-17
Lar-1-18
Lar-1-19
Lar-1-20
Lar-1-21
Lar-1-22
Lar-1-23
Lar-1-24
Lar-1-25
Lar-1-26
Lar-1-27
Lar-1-28
Lar-1-29
Lar-1-30
Lar-1-31
Lar-1-32
Lar-1-33
Lar-1-34
Lar-1-35
Lar-1-36
Lar-1-37
Lar-1-38
Lar-1-39
Lar-1-40
Lar-1-41
Lar-1-42
Lar-1-43
Lar-1-44
Lar-1-45
Lar-1-46
Lar-1-47
Lar-1-48
Lar-1-49
Lar-1-50
Lar-1-51
Lar-1-52
Lar-1-53
env-1
Aba-1-2
env-2
Aba-1-3
Aba-1-4
Aba-1-5
Aba-1-6
Aba-1-7
Aba-1-8
Aba-1-9
Aba-1-10
Aba-1-11
env-3
env-4
env-5


alignment positions 161 · · · 320
Lar-1-1
Lar-1-2
Lar-1-3
Aba-1-1
Lar-1-4
Lar-1-5
Lar-1-6
Lar-1-7
Lar-1-8
Lar-1-9
Lar-1-10
Lar-1-11
Lar-1-12
Lar-1-13
Lar-1-14
Lar-1-15
Lar-1-16
Lar-1-17
Lar-1-18
Lar-1-19
Lar-1-20
Lar-1-21
Lar-1-22
Lar-1-23
Lar-1-24
Lar-1-25
Lar-1-26
Lar-1-27
Lar-1-28
Lar-1-29
Lar-1-30
Lar-1-31
Lar-1-32
Lar-1-33
Lar-1-34
Lar-1-35
Lar-1-36
Lar-1-37
Lar-1-38
Lar-1-39
Lar-1-40
Lar-1-41
Lar-1-42
Lar-1-43
Lar-1-44
Lar-1-45
Lar-1-46
Lar-1-47
Lar-1-48
Lar-1-49
Lar-1-50
Lar-1-51
Lar-1-52
Lar-1-53
env-1
Aba-1-2
env-2
Aba-1-3
Aba-1-4
Aba-1-5
Aba-1-6
Aba-1-7
Aba-1-8
Aba-1-9
Aba-1-10
Aba-1-11
env-3
env-4
env-5

alignment positions 321 · · · 447



5 Actino-pnp

5.1 Taxa

The taxonomy of each organism containing a putative Actino-pnp RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Actino-pnp RNA in Sections 5.2 and 5.4.

abbrev. of hits	taxonomy of species
Aod-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Actinomycineae Actinomycetaceae *Actinomyces odontolyticus* ATCC 17982
Mgi-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium gilvum* PYR-GCK
Mle-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium leprae* TN
Msm-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium smegmatis str.* MC2 155
Msp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium sp.* JLS
Msp-2-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium sp.* KMS
Msp-3-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium sp.* MCS
Mul-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium ulcerans* Agy99
Mva-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium vanbaalenii* PYR-1
Mav-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumavium complex (MAC) *Mycobacterium avium* 104
Mav-2-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumavium complex (MAC) *Mycobacterium avium subs*
Mbo-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium bovis* AF2122
Mbo-2-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium bovis* BCG st
Mtu-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium tuberculosis* C
Mtu-2-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium tuberculosis* C
Mtu-3-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium tuberculosis* H
Mtu-4-1 to Mtu-4-2	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium tuberculosis* H
Mtu-5-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium tuberculosis* H
Mtu-6-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium tuberculosis s*
Nfa-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Nocardiaceae *Nocardia farcinica* IFM 10152
Rsp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Nocardiaceae *Rhodococcus sp.* RHA1
Ace-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Frankineae Acidothermaceae *Acidothermus cellulolyticus* 11B
Fal-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Frankineae Frankiaceae *Frankia alni* ACN14a
Fsp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Frankineae Frankiaceae *Frankia sp.* CcI3
Fsp-2-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Frankineae Frankiaceae *Frankia sp.* EAN1pec
Kra-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Frankineae Kineosporiaceae *Kineococcus radiotolerans* SRS30216
Bli-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Brevibacteriaceae *Brevibacterium linens* BL2
Cmi-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Microbacteriaceae *Clavibacter michiganensis subsp. michiganensis* NCPPB 382
Lxy-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Microbacteriaceae *Leifsonia xyli subsp. xyli str.* CTCB07
Aau-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Micrococcaceae *Arthrobacter aurescens* TC1
Asp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Micrococcaceae *Arthrobacter sp.* FB24
Sar-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micromonosporineae Micromonosporaceae *Salinispora arenicola* CNS205
Str-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micromonosporineae Micromonosporaceae *Salinispora tropica* CNB-440
Nsp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Propionibacterineae Nocardioidaceae *Nocardioides sp.* JS614
Pac-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Propionibacterineae Propionibacteriaceae *Propionibacterium acnes* KPA171202
Ser-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Pseudonocardineae Pseudonocardiaceae *Saccharopolyspora erythraea* NRRL 2338
Sav-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Streptomycineae Streptomycetaceae *Streptomyces avermitilis* MA-4680
Sco-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Streptomycineae Streptomycetaceae *Streptomyces coelicolor* A3(2)
Tfu-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Streptosporangineae Nocardiopsaceae *Thermobifida fusca* YX
env-1 to env-60	environmental samples

5.2 Gene contexts

Each Actino-pnp RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Actino-pnp RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 5.3. The accession number of the sequence containing each Actino-pnp RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 5.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	JCVI_SCAF_1096628304755	-	1856	1796	**RNA**→
env-2	JCVI_SCAF_1096627653623	+	419	480	**RNA**→ ? (GOS2596)→ RNase_PH (pfam01138)PNPase (pfam03726)? (GOS2596)→
env-3	JCVI_SCAF_1101668161176	+	1157	1215	**RNA**→
env-4	JCVI_SCAF_1101668134089	-	962	904	**RNA**→ RNase_PH (pfam01138)? (GOS2596)→
Fsp-1-1	NC_007777.1	-	4254111	4254042	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ DapB (COG0289)→
Fsp-2-1	NZ_AAII01000067.1	-	33426	33356	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ DapB (COG0289)→
Fal-1-1	NC_008278.1	-	6238131	6238064	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ DapB (COG0289)→ DnaK (COG0443)→ hypo→ hypo→
Ace-1-1	NC_008578.1	-	1705830	1705770	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ DapB (COG0289)→ COG1233 (COG1233)→
Nsp-1-1	NC_008699.1	-	3377420	3377345	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→
env-5	JCVI_SCAF_1096627678551	-	471	413	**RNA**→ RNase_PH (pfam01138)? (GOS2596)→
Aau-1-1	NC_008711.1	+	1715342	1715414	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ hypo→
Kra-1-1	NC_009664.1	+	1567631	1567701	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ DapB (COG0289)→ conserved hypothetical secreted protein→
Aod-1-1	NZ_AAYI02000004.1	-	1711891	1711833	**RNA**→ Pnp (COG1185)→
Tfu-1-1	NC_007333.1	+	925836	925906	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→
Rsp-1-1	NC_008268.1	+	7150693	7150751	**RNA**→ Pnp (COG1185)→
Pac-1-1	NC_006085.1	-	1592080	1592022	**RNA**→ Pnp (COG1185)→
Nfa-1-1	NC_006361.1	-	4102358	4102295	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→
env-6	JCVI_SCAF_1101668152286	-	1046	983	**RNA**→ RNase_PH (pfam01138)? (GOS2596)→
env-7	JCVI_SCAF_1101668623901	-	761	687	**RNA**→ RNase_PH (pfam01138)RNase_PH_C (pfam03725)? (GOS2596)→
env-8	JCVI_SCAF_1096627457080	-	848	776	**RNA**→ RNase_PH (pfam01138)? (GOS2596)→ ? (GOS2596)→
env-9	JCVI_SCAF_1101668138734	-	842	771	**RNA**→ RNase_PH (pfam01138)? (GOS2596)→
env-10	JCVI_SCAF_1101668248408	-	1068	995	**RNA**→ Pnp (COG1185)? (GOS2596)→
env-11	JCVI_SCAF_1101668243999	-	995	922	**RNA**→ RNase_PH (pfam01138)? (GOS2596)→
env-12	JCVI_SCAF_1096626897378	-	1600	1521	**RNA**→ Pnp (COG1185)? (GOS2596)→
env-13	JCVI_SCAF_1101668100121	-	648	584	**RNA**→ RNase_PH (pfam01138)? (GOS2596)→
env-14	JCVI_SCAF_1101668160289	-	798	727	**RNA**→ RNase_PH (pfam01138)RNase_PH_C (pfam03725)? (GOS2596)→
env-15	JCVI_SCAF_1101668116854	-	1286	1224	**RNA**→ Pnp (COG1185)? (GOS2596)→
env-16	JCVI_SCAF_1101668148428	-	765	703	**RNA**→ RNase_PH (pfam01138)? (GOS2596)→
env-17	JCVI_SCAF_1101668122106	-	1127	1062	**RNA**→ Pnp (COG1185)? (GOS2596)→
env-18	JCVI_SCAF_1101668118458	-	673	600	**RNA**→ RNase_PH (pfam01138)? (GOS2596)→
env-19	JCVI_SCAF_1096626901761	+	1019	1083	**RNA**→ Pnp (COG1185)? (GOS2596)→ PqqL (COG0612)? (GOS289)→
env-20	JCVI_SCAF_1101668114308	-	1382	1311	**RNA**→ Pnp (COG1185)? (GOS2596)→
env-21	JCVI_SCAF_1096627654472	-	1030	960	**RNA**→ RNase_PH (pfam01138)? (GOS2596)→ RNase_PH (pfam01138)? (GOS2596)→
env-22	JCVI_SCAF_1101667162011	-	110	50	**RNA**→
env-23	JCVI_SCAF_1101668104815	+	53	112	**RNA**→
env-24	JCVI_SCAF_1101668107913	-	827	768	**RNA**→ RNase_PH (pfam01138)? (GOS2596)→
env-25	JCVI_SCAF_1096627446429	+	920	978	**RNA**→
env-26	JCVI_SCAF_1096626895935	+	2004	2064	**RNA**→ ←? (GOS11987980)
env-27	JCVI_SCAF_1096627298438	-	941	881	**RNA**→ ←? (GOS11987980)
env-28	JCVI_SCAF_1101668121219	+	759	819	**RNA**→ ←? (GOS11987980)
env-29	JCVI_SCAF_1096626198158	+	767	823	**RNA**→

env-30	JCVI_SCAF_1096627650139	-	515	457	**RNA**→ RNase_PH (pfam01138)? (GOS2596)→
env-31	JCVI_SCAF_1101668247038	+	1346	1402	**RNA**→
env-32	JCVI_SCAF_1096627006102	-	1734	1676	**RNA**→ Pnp (COG1185)? (GOS2596)→
env-33	JCVI_SCAF_1101667164550	-	365	309	**RNA**→ ? (GOS2596)→
env-34	JCVI_SCAF_1101668245159	-	532	468	**RNA**→ RNase_PH (pfam01138)? (GOS2596)→
env-35	JCVI_SCAF_1096627382872	-	2493	2429	**RNA**→ Pnp (COG1185)? (GOS2596)→
env-36	JCVI_SCAF_1096627672664	-	306	242	**RNA**→
env-37	JCVI_SCAF_1101668233903	-	759	700	**RNA**→ RNase_PH (pfam01138)? (GOS2596)→
env-38	JCVI_SCAF_1096627005160	-	593	531	**RNA**→ RNase_PH (pfam01138)? (GOS2596)→
env-39	JCVI_SCAF_1096627015732	-	327	265	**RNA**→
env-40	JCVI_SCAF_1096627667722	+	80	142	**RNA**→
env-41	JCVI_SCAF_1101668120237	-	602	542	**RNA**→ RNase_PH (pfam01138)? (GOS2596)→
env-42	JCVI_SCAF_1101668151658	+	624	697	**RNA**→ RNase_PH (pfam01138)? (GOS2596)→
env-43	JCVI_SCAF_1101668143468	+	1190	1249	**RNA**→
env-44	JCVI_SCAF_1096627298000	+	870	946	**RNA**→ Pnp (COG1185)? (GOS2596)→
env-45	JCVI_SCAF_1101668115969	-	1250	1184	**RNA**→ Pnp (COG1185)? (GOS2596)→
env-46	JCVI_SCAF_1101668253034	+	836	893	**RNA**→ RNase_PH (pfam01138)? (GOS2596)→
env-47	JCVI_SCAF_1101668115647	+	1399	1470	**RNA**→
env-48	JCVI_SCAF_1101668141007	-	840	769	**RNA**→ RNase_PH (pfam01138)? (GOS2596)→
Bli-1-1	NZ_AAGP01000012.1	-	38273	38199	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→
Asp-1-1	NC_008541.1	+	1593772	1593848	**RNA**→ Pnp (COG1185)→
Sar-1-1	NZ_AAWA01000012.1	+	89375	89448	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→
Str-1-1	NC_009380.1	+	1548720	1548793	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→
env-49	JCVI_SCAF_1096627357583	-	353	293	**RNA**→
Sco-1-1	NC_003888.3	+	6259624	6259683	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ DapB (COG0289)→ hypo→
Sav-1-1	NC_003155.3	-	3089217	3089156	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ DapB (COG0289)→ hypo→
Ser-1-1	NC_009142.1	-	6641639	6641570	**RNA**→ Pnp (COG1185)→
Mav-2-1	NC_002944.2	-	3223271	3223201	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ NPD (pfam03060)→
Mav-1-1	NC_008595.1	-	3804947	3804877	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→
Mle-1-1	NC_002677.1	+	1013687	1013757	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→
Mtu-3-1	NC_009565.1	-	3104733	3104663	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ NPD (pfam03060)→
Mbo-2-1	NC_008769.1	-	3046226	3046156	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ NPD (pfam03060)→
Mtu-4-1	NC_009525.1	-	3104746	3104676	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ NPD (pfam03060)→
Mtu-2-1	NC_002755.2	-	3087501	3087431	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ NPD (pfam03060)→
Mtu-5-1	NC_000962.2	-	3092778	3092708	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ NPD (pfam03060)→
Mtu-1-1	NZ_AAKR01000096.1	-	9117	9047	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ NPD (pfam03060)→
Mtu-6-1	NZ_AASN01000042.1	-	38563	38493	**RNA**→
Mtu-4-2	NZ_AAYK01000241.1	-	32936	32866	**RNA**→
Mbo-1-1	NC_002945.3	-	3049327	3049257	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ NPD (pfam03060)→
Msp-3-1	NC_008146.1	+	2248687	2248748	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ PenP (COG2367)→ COG3012 (COG3012)→
Msp-1-1	NC_009077.1	+	2189966	2190027	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ PenP (COG2367)→ COG3012 (COG3012)→
Msp-2-1	NC_008705.1	+	2267671	2267732	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ PenP (COG2367)→ COG3012 (COG3012)→
Mul-1-1	NC_008611.1	+	2397606	2397677	**RNA**→ Pnp (COG1185)→
Msm-1-1	NC_008596.1	+	2735853	2735921	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ PenP (COG2367)→
Mva-1-1	NC_008726.1	+	2507941	2507999	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ PenP (COG2367)→
Mgi-1-1	NC_009338.1	-	4277016	4276958	**RNA**→ Pnp (COG1185)→ PqqL (COG0612)→ PenP (COG2367)→
Lxy-1-1	NC_006087.1	+	918955	919031	**RNA**→ Pnp (COG1185)→
env-50	JCVI_SCAF_1096626904161	-	355	298	**RNA**→
env-51	JCVI_SCAF_1096626894801	-	1400	1343	**RNA**→ Pnp (COG1185)? (GOS2596)→
env-52	JCVI_SCAF_1101667060140	-	382	325	**RNA**→ ? (GOS2596)→
env-53	JCVI_SCAF_1101668772299	+	1225	1288	**RNA**→
env-54	JCVI_SCAF_1096627358323	-	174	111	**RNA**→

env-55	JCVI_SCAF_1096627364885	+	1283	1346	**RNA**→ Pnp (COG1185)? (GOS2596)→
Cmi-1-1	NC_009480.1	-	2312598	2312542	**RNA**→ Pnp (COG1185)→
env-56	JCVI_SCAF_1096627387932	-	1167	1111	**RNA**→ Pnp (COG1185)? (GOS2596)→
env-57	JCVI_SCAF_1096627357648	+	1000	1056	**RNA**→ Pnp (COG1185)? (GOS2596)→ PNPase_KH (cd02393)RNase_PH_C (pfam03725)? (GOS2596)→
env-58	JCVI_SCAF_1101668632187	+	995	1051	**RNA**→ RNase_PH (pfam01138)? (GOS2596)→
env-59	JCVI_SCAF_1096626905106	+	1093	1151	**RNA**→ Pnp (COG1185)? (GOS2596)→
env-60	JCVI_SCAF_1096627298538	-	918	863	**RNA**→ RNase_PH (pfam01138)? (GOS2596)→

5.3 Conserved domains

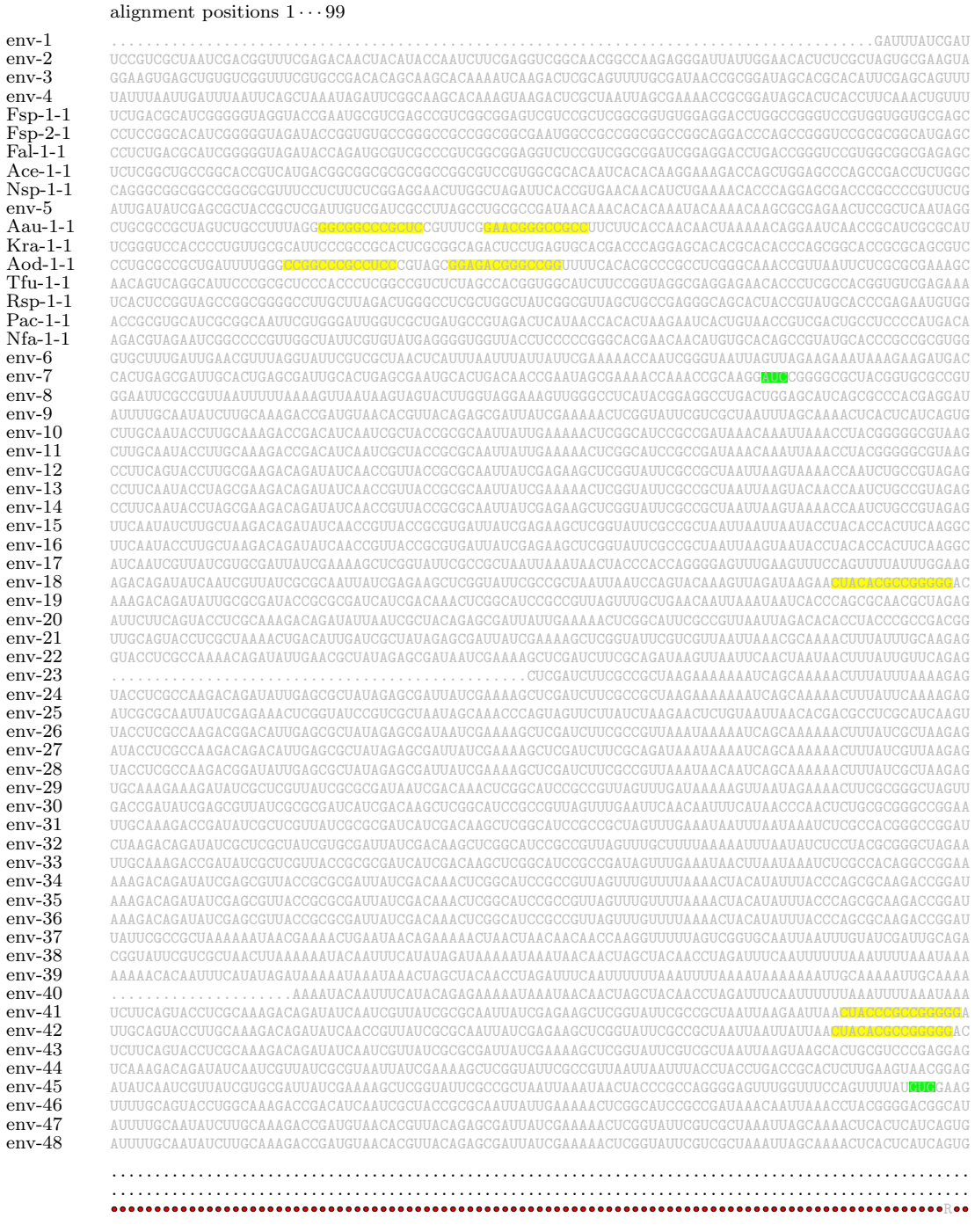
Conserved domains found in protein-coding genes listed in Section 5.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Actino-pnp RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

cd02393 (1) Polynucleotide phosphorylase (PNPase) K homology RNA-binding domain (KH).
COG0289 (7) Dihydrodipicolinate reductase [Amino acid transport and metabolism]
COG0443 (1) Molecular chaperone [Posttranslational modification, protein turnover, chaperones]
COG0612 (31) Predicted Zn-dependent peptidases [General function prediction only]
COG1185 (53) Polyribonucleotide nucleotidyltransferase (polynucleotide phosphorylase) [Translation, ribosomal structure and biogenesis]
COG1233 (1) Phytoene dehydrogenase and related proteins [Secondary metabolites biosynthesis, transport, and catabolism]
COG2367 (6) Beta-lactamase class A [Defense mechanisms]
COG3012 (3) Uncharacterized protein conserved in bacteria [Function unknown]
GOS11987980 (3) ?
GOS2596 (45) ?
GOS289 (1) ?
pfam01138 (25) 3' exoribonuclease family, domain 1.
pfam03060 (8) 2-nitropropane dioxygenase.
pfam03725 (3) 3' exoribonuclease family, domain 2.
pfam03726 (1) Polyribonucleotide nucleotidyltransferase, RNA binding domain.

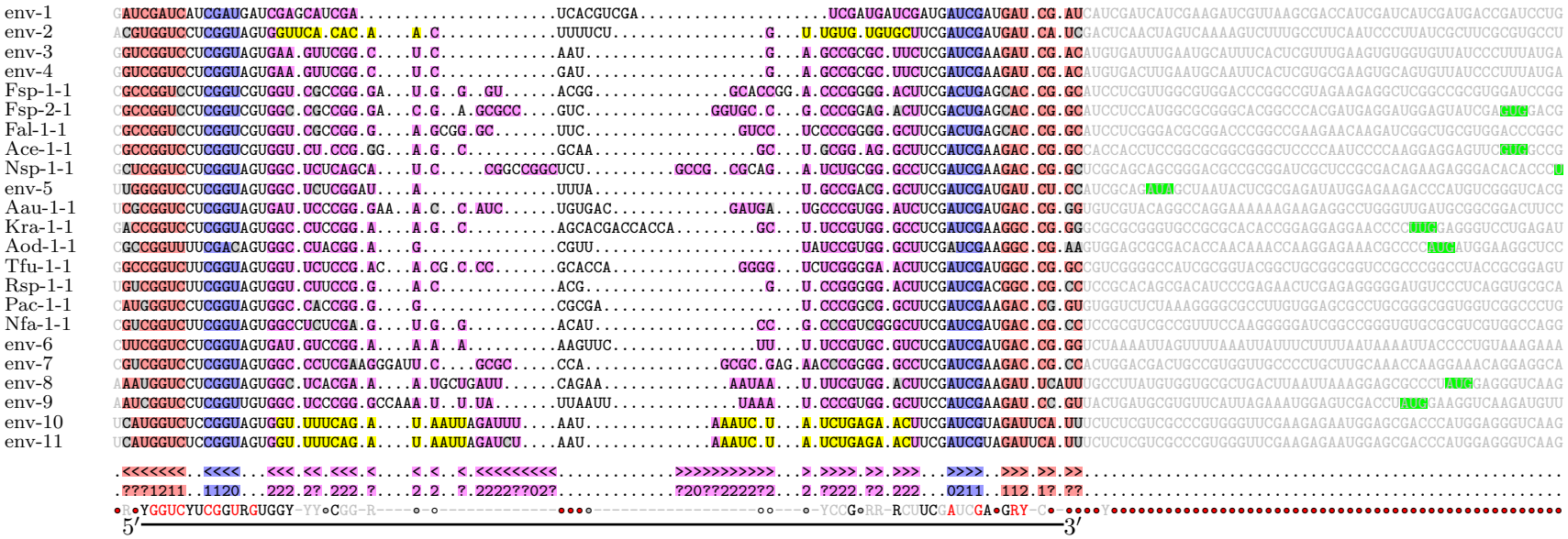
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Mbo-2-1, Msp-1-1, Msp-2-1, Mtu-1-1, Mtu-2-1, Mtu-4-1, Mtu-4-2, Mtu-5-1, Mtu-6-1

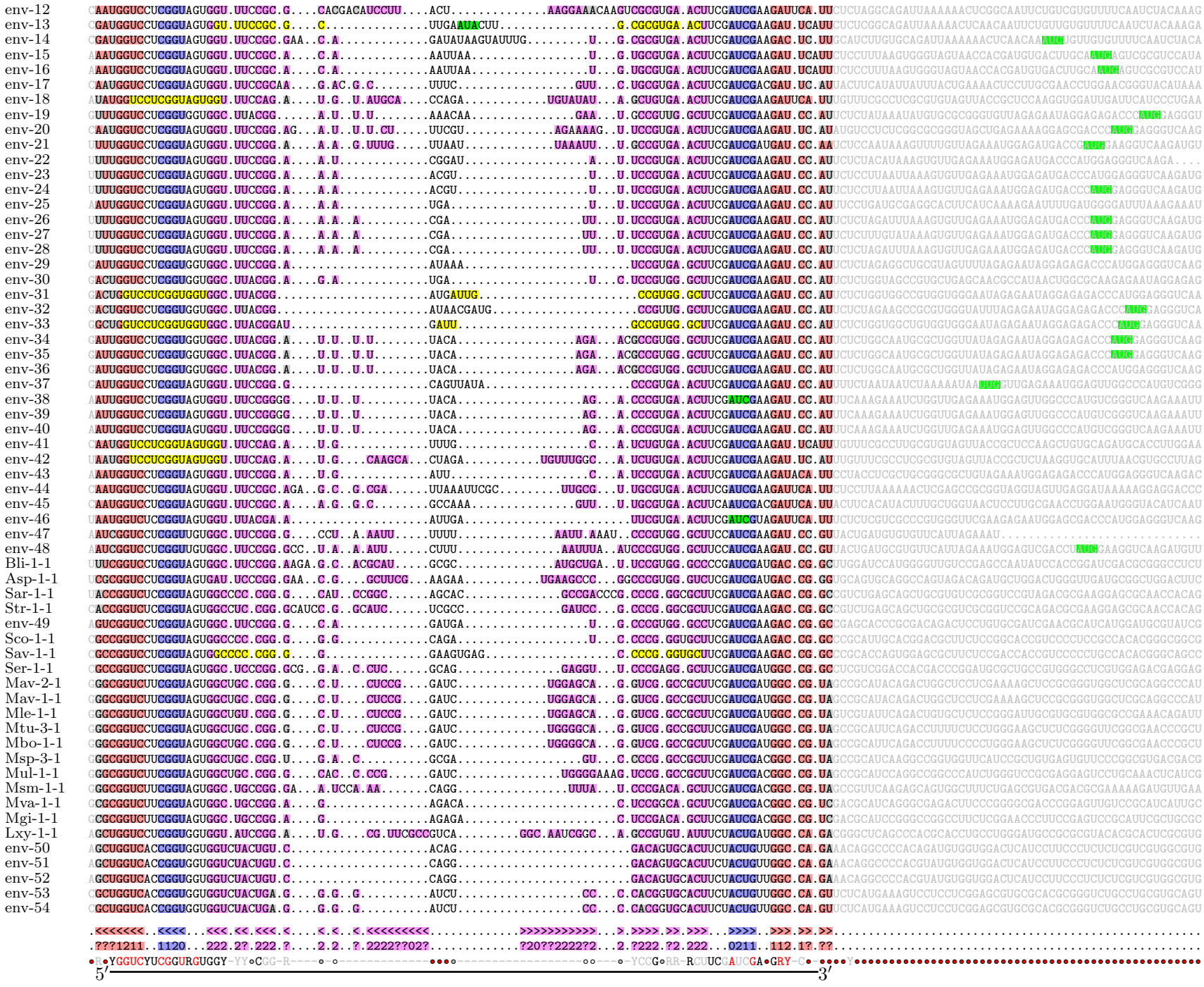
5.4 Multiple-sequence alignment

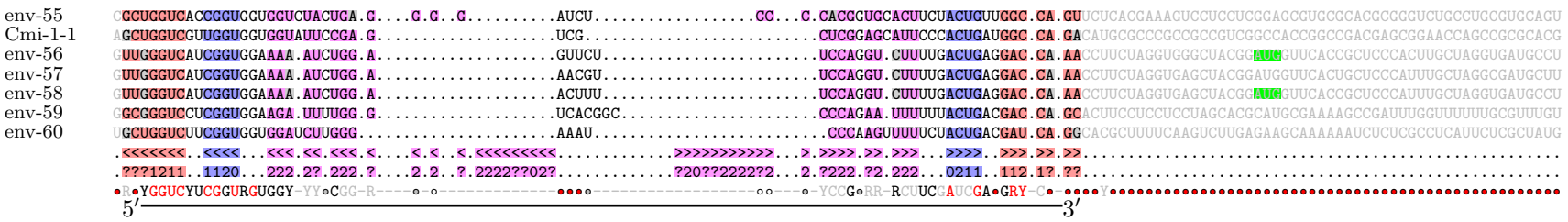
Each Actino-pnp RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 5.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Actino-pnp RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.


alignment positions 1 · · · 99
env-1
env-2
env-3
env-4
Fsp-1-1
Fsp-2-1
Fal-1-1
Ace-1-1
Nsp-1-1
env-5
Aau-1-1
Kra-1-1
Aod-1-1
Tfu-1-1
Rsp-1-1
Pac-1-1
Nfa-1-1
env-6
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48

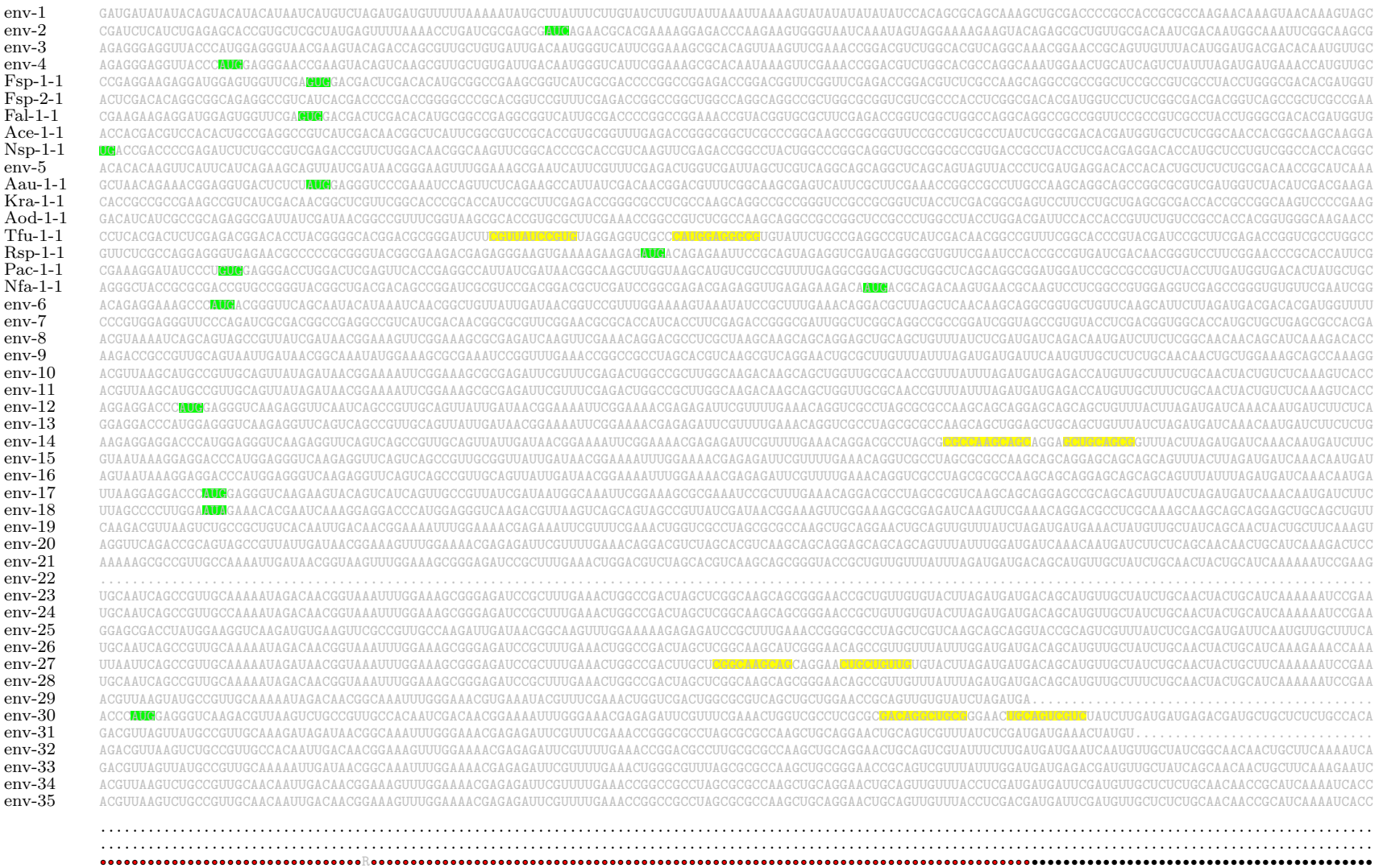
alignment positions 100 · · · 259

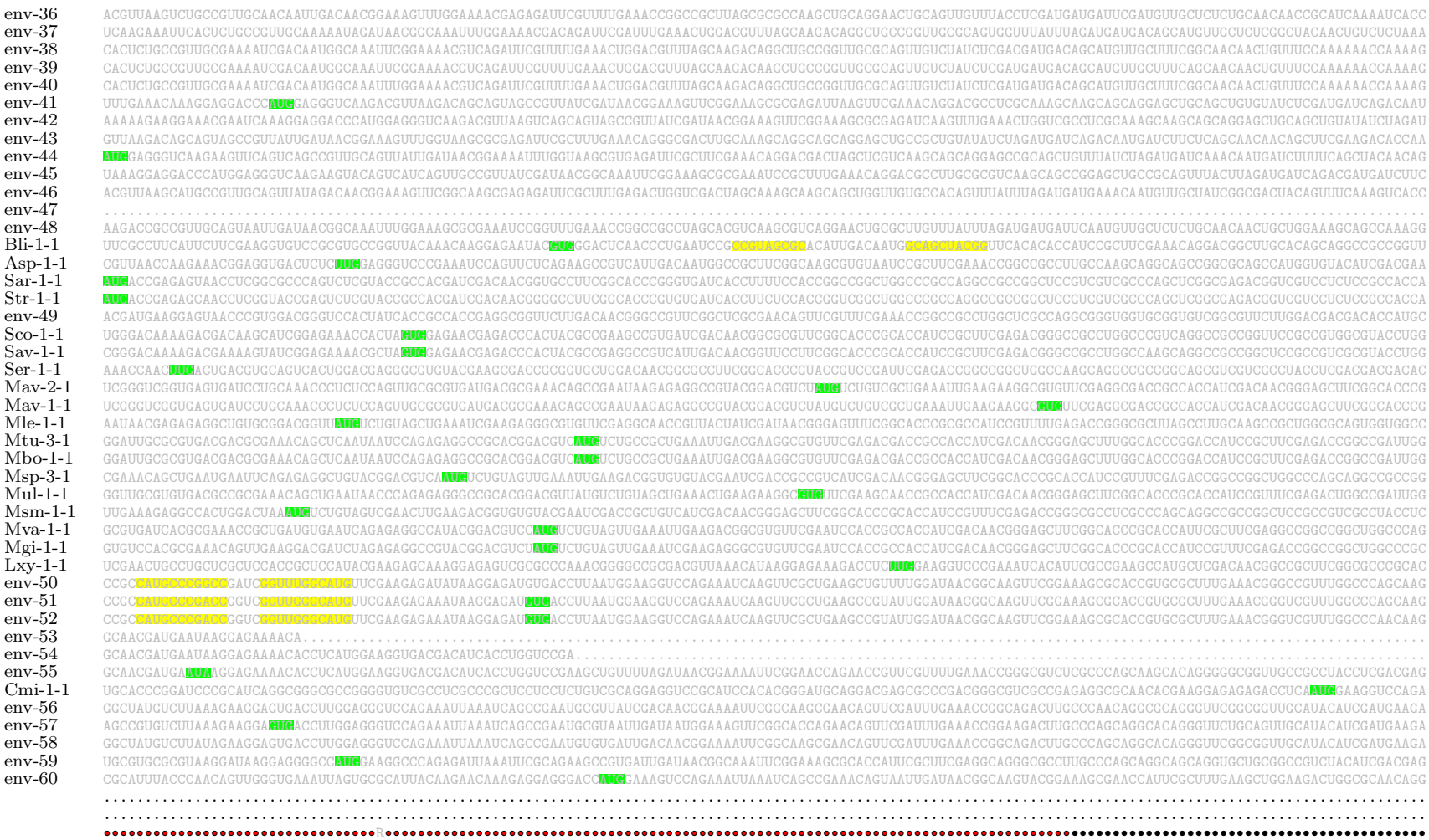


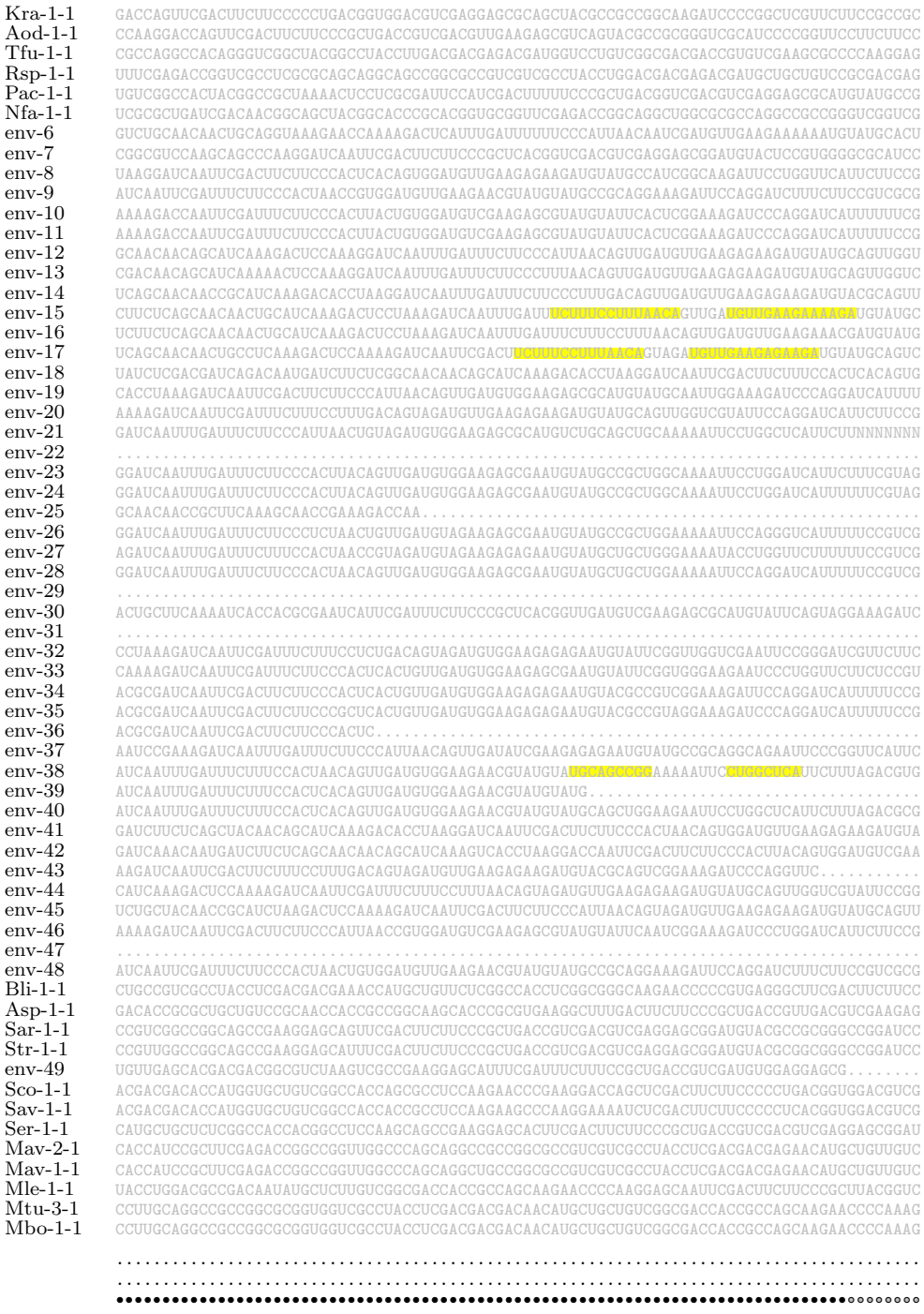

env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
Bli-1-1
Asp-1-1
Sar-1-1
Str-1-1
env-49
Sco-1-1
Sav-1-1
Ser-1-1
Mav-2-1
Mav-1-1
Mle-1-1
Mtu-3-1
Mbo-1-1
Msp-3-1
Mul-1-1
Msm-1-1
Mva-1-1
Mgi-1-1
Lxy-1-1
env-50
env-51
env-52
env-53
env-54
5′
3′



alignment positions 260 · · · 419




env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
Bli-1-1
Asp-1-1
Sar-1-1
Str-1-1
env-49
Sco-1-1
Sav-1-1
Ser-1-1
Mav-2-1
Mav-1-1
Mle-1-1
Mtu-3-1
Mbo-1-1
Msp-3-1
Mul-1-1
Msm-1-1
Mva-1-1
Mgi-1-1
Lxy-1-1
env-50
env-51
env-52
env-53
env-54
env-55
Cmi-1-1
env-56
env-57
env-58
env-59
env-60
alignment positions 420 · · · 506
env-1
env-2
env-3
env-4
Fsp-1-1
Fsp-2-1
Fal-1-1
Ace-1-1
Nsp-1-1
env-5
Aau-1-1


Kra-1-1
Aod-1-1
Tfu-1-1
Rsp-1-1
Pac-1-1
Nfa-1-1
env-6
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
Bli-1-1
Asp-1-1
Sar-1-1
Str-1-1
env-49
Sco-1-1
Sav-1-1
Ser-1-1
Mav-2-1
Mav-1-1
Mle-1-1
Mtu-3-1
Mbo-1-1

6 AdoCbl-variant

6.1 Taxa

The taxonomy of each organism containing a putative AdoCbl-variant RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual AdoCbl-variant RNA in Sections 6.3 and 6.5.

abbrev. of hits	**taxonomy of species**
env-1 to env-144	environmental samples

6.2 Notes

Superscript numbers are used to annotate any AdoCbl-variant RNA that has special characteristics (described below). These numeric annotations will be used in Sections 6.3 and 6.5. The meaning of the numbers are as follows.

[1] Overlaps AdoCbl riboswitch sequence that was already detected as such, although the original sequence is not aligned well, and therefore probably fits the variant motif better.

[2] Near predicted rRNA.

6.3 Gene contexts

Each AdoCbl-variant RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one AdoCbl-variant RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 6.4. The accession number of the sequence containing each AdoCbl-variant RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 6.1. Superscript numbers refer to annotations listed in Section 6.2

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	JCVI_SCAF_1101668560197	+	868	993	**RNA**→
env-2	JCVI_SCAF_1096628287862	-	1414	1296	**RNA**→ ? (GOS4369)→ ? (GOS4665)→ Thy1 (pfam02511)? (GOS5954)→
env-3	JCVI_SCAF_1101668518363	-	1418	1300	**RNA**→ BtuB (COG4206)? (GOS1720)→
env-4	JCVI_SCAF_1101668460391	-	384	266	**RNA**→ ? (GOS1720)→
env-5	JCVI_SCAF_1101668181603	+	993	1106	**RNA**→ ? (GOS1720)→
env-6	JCVI_SCAF_1096628002478	+	1267	1385	**RNA**→ ? (GOS496)→
env-7	JCVI_SCAF_1096626931125	+	391	507	**RNA**→ BtuB (COG4206)? (GOS1720)→
env-8	JCVI_SCAF_1096627830159	+	1185	1300	**RNA**→ ? (GOS4369)→
env-9	JCVI_SCAF_1096627408381	+	141	256	**RNA**→ ? (GOS4369)→ ? (GOS4665)→
env-10	JCVI_SCAF_1096628254439	+	357	495	**RNA**→ ? (GOS1720)→
env-11	ABEF01043191.1	-	931	798	**RNA**→ hypo→
env-12	JCVI_SCAF_1101668133785	-	328	197	**RNA**→
env-13	ABEF01008614.1	+	1	139	**RNA**→
env-14	JCVI_SCAF_1101668593526	+	801	929	**RNA**→ ? (GOS1720)→
env-15	JCVI_SCAF_1101668492791	+	611	749	**RNA**→ ? (GOS920)→
env-16	JCVI_SCAF_1096626110926	+	473	604	**RNA**→ ? (GOS11983570)→
env-17	JCVI_SCAF_1101668495477	+	564	702	**RNA**→ ? (GOS1720)→
env-18	JCVI_SCAF_1096627476376	+	351	479	**RNA**→ ? (GOS9454)→
env-19	JCVI_SCAF_1101668174084	-	1891	1775	**RNA**→ ? (GOS4369)→ ? (GOS4665)→
env-20	JCVI_SCAF_1096627290449	+	2935	3064	**RNA**→
env-21	JCVI_SCAF_1096626307088	-	364	236	**RNA**→ ←? (GOS11790770)
env-22	ABEF01048062.1	+	1373	1515	**RNA**→ hypo→
env-23	JCVI_SCAF_1101668472248	+	19	147	**RNA**→ BtuB (COG4206)? (GOS1720)→
env-24	JCVI_SCAF_1096628149541	-	147	19	**RNA**→
env-25	JCVI_SCAF_1096627100734	-	553	425	**RNA**→ ? (GOS1720)→
env-26	JCVI_SCAF_1096627885039	+	1487	1630	**RNA**→
env-27	JCVI_SCAF_1101668176557	+	771	916	**RNA**→ ? (GOS4369)→
env-28	JCVI_SCAF_1096627126031	-	1664	1524	**RNA**→ BtuB (COG4206)? (GOS1720)→
env-29	JCVI_SCAF_1096628223806	+	1694	1835	**RNA**→
env-30	JCVI_SCAF_1101668002322	+	1360	1487	**RNA**→
env-31	JCVI_SCAF_1096627253662	+	2472	2603	**RNA**→ BtuB (COG4206)? (GOS1720)→
env-32	JCVI_SCAF_1096626931276	-	1019	889	**RNA**→ ? (GOS1720)→
env-33	JCVI_SCAF_1096627339561	+	642	772	**RNA**→ ? (GOS1720)→
env-34	JCVI_SCAF_1101668048187	-	1134	1004	**RNA**→ BtuB (COG4206)? (GOS1720)→
env-35	JCVI_SCAF_1096627699890	-	422	290	**RNA**→ ←? (GOS11714930)
env-36	JCVI_SCAF_1096626643666	+	1300	1432	**RNA**→ ←? (GOS11714930)
env-37	JCVI_SCAF_1096627719709	-	1443	1311	**RNA**→ ? (GOS4369)→
env-38	JCVI_SCAF_1096628259790	+	1208	1350	**RNA**→ ? (GOS4369)→
env-39	JCVI_SCAF_1096626962232	-	1959	1818	**RNA**→ ? (GOS920)→ TonB_dep_Rec (pfam00593)? (GOS4665)→
env-40	JCVI_SCAF_1096627991926	-	141	13	**RNA**→
env-41	JCVI_SCAF_1096627270303	-	1972	1838	**RNA**→ ? (GOS920)→ BtuB (COG4206)? (GOS4665)→

env-42	JCVI_SCAF_1096627854245	-	1394	1264	**RNA**→ ←? (GOS11790770)
env-43	JCVI_SCAF_1101668501944	+	36	167	**RNA**→ ligand_gated_channel (cd01347)? (GOS1720)→
env-44	JCVI_SCAF_1096627301998	+	1198	1329	**RNA**→ ? (GOS1720)→
[2]env-45	JCVI_SCAF_1096628403513	-	1599	1469	**RNA**→ ? (GOS4369)→
env-46	JCVI_SCAF_1096627684153	+	893	1024	**RNA**→ ? (GOS920)→
env-47	JCVI_SCAF_1096627808730	-	815	683	**RNA**→ ? (GOS4369)→
env-48	JCVI_SCAF_1101668713690	-	379	251	**RNA**→ ←? (GOS11790770)
env-49	JCVI_SCAF_1096627256224	+	234	364	**RNA**→ BtuB (COG4206)? (GOS1720)→
env-50	JCVI_SCAF_1096627698506	+	1315	1453	**RNA**→
env-51	JCVI_SCAF_1101668212159	-	162	24	**RNA**→
env-52	JCVI_SCAF_1096627993344	-	1348	1220	**RNA**→ ? (GOS4369)→ BtuB (COG4206)? (GOS1720)→
env-53	JCVI_SCAF_1096627699886	-	468	339	**RNA**→ ? (GOS1720)→
env-54	JCVI_SCAF_1101668502439	-	1211	1082	**RNA**→ BtuB (COG4206)? (GOS1720)→
env-55	JCVI_SCAF_1096628156220	-	1539	1410	**RNA**→ ? (GOS4369)→ ? (GOS9454)→ ? (GOS11245591)→
env-56	JCVI_SCAF_1096627830576	-	1617	1488	**RNA**→ ? (GOS4369)→ BtuB (COG4206)? (GOS1720)→
env-57	JCVI_SCAF_1101667038075	+	579	710	**RNA**→
env-58	JCVI_SCAF_1096627982330	-	566	436	**RNA**→ ? (GOS4369)→
env-59	JCVI_SCAF_1096626313314	+	459	590	**RNA**→ ? (GOS11094971)→
env-60	JCVI_SCAF_1096627886130	+	747	877	**RNA**→ ? (GOS11622442)→
env-61	JCVI_SCAF_1096628272437	+	1440	1571	**RNA**→
env-62	JCVI_SCAF_1096628236348	-	418	295	**RNA**→ ? (GOS1720)→
env-63	JCVI_SCAF_1096628165183	+	63	187	**RNA**→ ? (GOS4369)→
env-64	JCVI_SCAF_1101667474120	+	18	144	**RNA**→ ? (GOS4369)→
env-65	JCVI_SCAF_1096627032404	-	590	462	**RNA**→ ? (GOS920)→
env-66	JCVI_SCAF_1096628065383	+	343	471	**RNA**→ ? (GOS4369)→
env-67	JCVI_SCAF_1096627581275	-	1700	1572	**RNA**→ ? (GOS920)→ TonB_dep_Rec (pfam00593)? (GOS4665)→
env-68	JCVI_SCAF_1096626922281	-	1134	1007	**RNA**→ ? (GOS1720)→
env-69	JCVI_SCAF_1096628084481	+	445	586	**RNA**→ ? (GOS4369)→
env-70	JCVI_SCAF_1101668356142	-	614	484	**RNA**→ ? (GOS4369)→
env-71	JCVI_SCAF_1096626133132	-	131	1	**RNA**→
env-72	JCVI_SCAF_1096626933891	-	1319	1189	**RNA**→ BtuB (COG4206)? (GOS496)→
env-73	JCVI_SCAF_1096627001005	+	640	770	**RNA**→ ? (GOS1720)→
env-74	JCVI_SCAF_1096628223711	+	1507	1637	**RNA**→ ←? (GOS11081972)
env-75	JCVI_SCAF_1096627431207	-	1457	1326	**RNA**→ ? (GOS1720)→ TonB_dep_Rec (pfam00593)? (GOS1720)→
env-76	JCVI_SCAF_1096628197778	-	1231	1100	**RNA**→ ? (GOS1720)→
env-77	JCVI_SCAF_1096627578146	+	73	202	**RNA**→ ? (GOS4369)→ ? (GOS4665)→
env-78	JCVI_SCAF_1101668011934	+	125	250	**RNA**→ ? (GOS920)→ ? (GOS496)→
env-79	JCVI_SCAF_1096627867925	+	240	367	**RNA**→ ? (GOS4369)→ ? (GOS1720)→
env-80	JCVI_SCAF_1101668659621	+	480	607	**RNA**→ ? (GOS4369)→
env-81	JCVI_SCAF_1101668182195	+	1359	1486	**RNA**→
env-82	JCVI_SCAF_1101668012817	-	471	342	**RNA**→ ? (GOS1720)→
env-83	JCVI_SCAF_1096627570490	+	1297	1426	**RNA**→ ? (GOS920)→
env-84	JCVI_SCAF_1101668471885	-	643	515	**RNA**→ ←? (GOS11790770)
env-85	JCVI_SCAF_1101668459580	+	711	851	**RNA**→ ? (GOS920)→
env-86	JCVI_SCAF_1096626582089	+	673	814	**RNA**→
env-87	JCVI_SCAF_1096627925949	+	429	560	**RNA**→ ? (GOS1720)→ BtuB (COG4206)? (GOS496)→
env-88	JCVI_SCAF_1096626980542	-	1015	884	**RNA**→ ? (GOS1720)→
env-89	JCVI_SCAF_1096626944243	+	159	290	**RNA**→ BtuB (COG4206)? (GOS1720)→
env-90	JCVI_SCAF_1096627815144	-	269	138	**RNA**→
env-91	JCVI_SCAF_1096628098054	+	83	213	**RNA**→ ←? (GOS11081972)
env-92	JCVI_SCAF_1101667521383	-	602	472	**RNA**→ ? (GOS4369)→
env-93	JCVI_SCAF_1096628238228	-	371	246	**RNA**→ ? (GOS4369)→

env-94	JCVI_SCAF_1096627993984	-	1344	1214	**RNA**→ ? (GOS1720)→ ? (GOS1720)→
env-95	JCVI_SCAF_1096627324500	-	188	58	**RNA**→
env-96	JCVI_SCAF_1101668283861	+	141	271	**RNA**→ BtuB (COG4206)? (GOS1720)→
env-97	JCVI_SCAF_1096627835374	-	1102	977	**RNA**→ BtuB (COG4206)? (GOS1720)→
env-98	JCVI_SCAF_1096627226916	-	899	771	**RNA**→ ? (GOS1720)→
env-99	JCVI_SCAF_1096628254797	-	1323	1194	**RNA**→ ? (GOS920)→ ? (GOS4665)→
env-100	JCVI_SCAF_1101667554535	-	534	405	**RNA**→ ←? (GOS11790770)
env-101	JCVI_SCAF_1101667253413	+	477	606	**RNA**→ ←? (GOS11790770)
env-102	JCVI_SCAF_1101667515587	+	475	603	**RNA**→ ? (GOS4369)→
env-103	JCVI_SCAF_1096627126564	+	969	1097	**RNA**→ ? (GOS920)→
env-104	JCVI_SCAF_1096627441792	+	1339	1467	**RNA**→ ? (GOS4369)→
env-105	JCVI_SCAF_1096627799570	+	634	764	**RNA**→ ? (GOS1720)→
env-106	JCVI_SCAF_1096628243783	+	51	179	**RNA**→ ? (GOS1720)→ ? (GOS4665)→
env-107	JCVI_SCAF_1096627898401	+	1404	1546	**RNA**→
env-108	JCVI_SCAF_1101668677996	-	164	34	**RNA**→
env-109	JCVI_SCAF_1101667429719	+	358	489	**RNA**→ ? (GOS1720)→
env-110	JCVI_SCAF_1101668547925	-	1002	875	**RNA**→ ? (GOS1720)→
env-111	JCVI_SCAF_1101667533745	+	172	299	**RNA**→ ? (GOS4369)→
[1]env-112	JCVI_SCAF_1101667573135	+	469	599	**RNA**→ ? (GOS1720)→
env-113	JCVI_SCAF_1096628056260	+	1623	1753	**RNA**→
env-114	JCVI_SCAF_1096627507316	-	520	390	**RNA**→ ? (GOS11094971)→
env-115	JCVI_SCAF_1096626884209	+	524	655	**RNA**→ ? (GOS1720)→
env-116	JCVI_SCAF_1101668488469	-	924	793	**RNA**→ ? (GOS496)→
env-117	JCVI_SCAF_1096628393525	+	1318	1448	**RNA**→ ←? (GOS11081972)
env-118	JCVI_SCAF_1096628194870	-	1569	1438	**RNA**→ ? (GOS4369)→ AsnB (cd00712)? (GOS729)→
env-119	JCVI_SCAF_1101668691878	+	589	719	**RNA**→ ? (GOS4369)→
env-120	JCVI_SCAF_1096628389621	-	777	647	**RNA**→ ? (GOS1720)→
env-121	JCVI_SCAF_1101668181310	+	918	1047	**RNA**→ ? (GOS920)→
env-122	JCVI_SCAF_1096626342623	+	28	158	**RNA**→ ? (GOS1720)→
env-123	JCVI_SCAF_1096628133600	-	1261	1131	**RNA**→ ? (GOS920)→
env-124	JCVI_SCAF_1096627559529	-	1746	1616	**RNA**→ ? (GOS4369)→ ? (GOS4665)→
env-125	JCVI_SCAF_1096626626183	+	425	555	**RNA**→ ? (GOS1720)→
env-126	JCVI_SCAF_1096628274095	+	1673	1803	**RNA**→
env-127	JCVI_SCAF_1096627184957	+	1942	2072	**RNA**→
env-128	JCVI_SCAF_1096627599646	+	1443	1573	**RNA**→
env-129	JCVI_SCAF_1101668433054	-	312	182	**RNA**→
env-130	JCVI_SCAF_1096628072499	+	592	722	**RNA**→ ? (GOS1720)→
env-131	JCVI_SCAF_1096627825342	-	570	441	**RNA**→ ? (GOS4369)→
env-132	JCVI_SCAF_1096627794834	-	1956	1826	**RNA**→ ? (GOS1720)→ FepA (COG4771)? (GOS4665)→
env-133	JCVI_SCAF_1096628049855	-	1223	1093	**RNA**→ ? (GOS9454)→ 5S_rRNA→
env-134	JCVI_SCAF_1096628121198	-	484	354	**RNA**→ ←? (GOS11081972)
env-135	JCVI_SCAF_1101667419775	+	622	753	**RNA**→ ? (GOS11983570)→
env-136	JCVI_SCAF_1096628119255	+	580	711	**RNA**→ ? (GOS4665)→
env-137	JCVI_SCAF_1096628284445	-	760	634	**RNA**→ ? (GOS4369)→
env-138	JCVI_SCAF_1096627115721	+	242	372	**RNA**→ BtuB (COG4206)? (GOS1720)→
env-139	JCVI_SCAF_1101668527442	-	1012	862	**RNA**→ ? (GOS1720)→
env-140	JCVI_SCAF_1101668519723	+	854	994	**RNA**→
env-141	JCVI_SCAF_1096626860966	+	57	194	**RNA**→ BtuB (COG4206)? (GOS496)→
env-142	JCVI_SCAF_1096628185343	-	860	742	**RNA**→ ? (GOS1720)→
env-143	JCVI_SCAF_1096627388192	-	1248	1134	**RNA**→ BtuB (COG4206)? (GOS1720)→
env-144	JCVI_SCAF_1101668310910	+	603	724	**RNA**→ ? (GOS1720)→

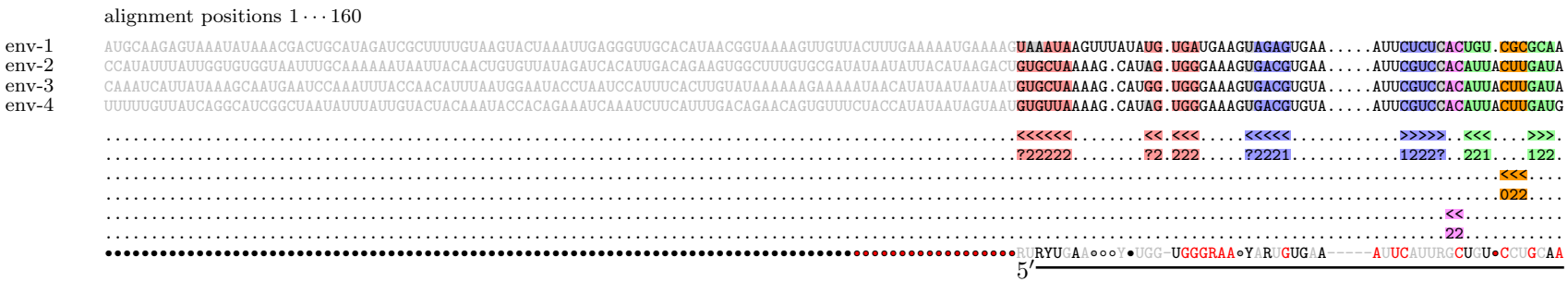
6.4 Conserved domains

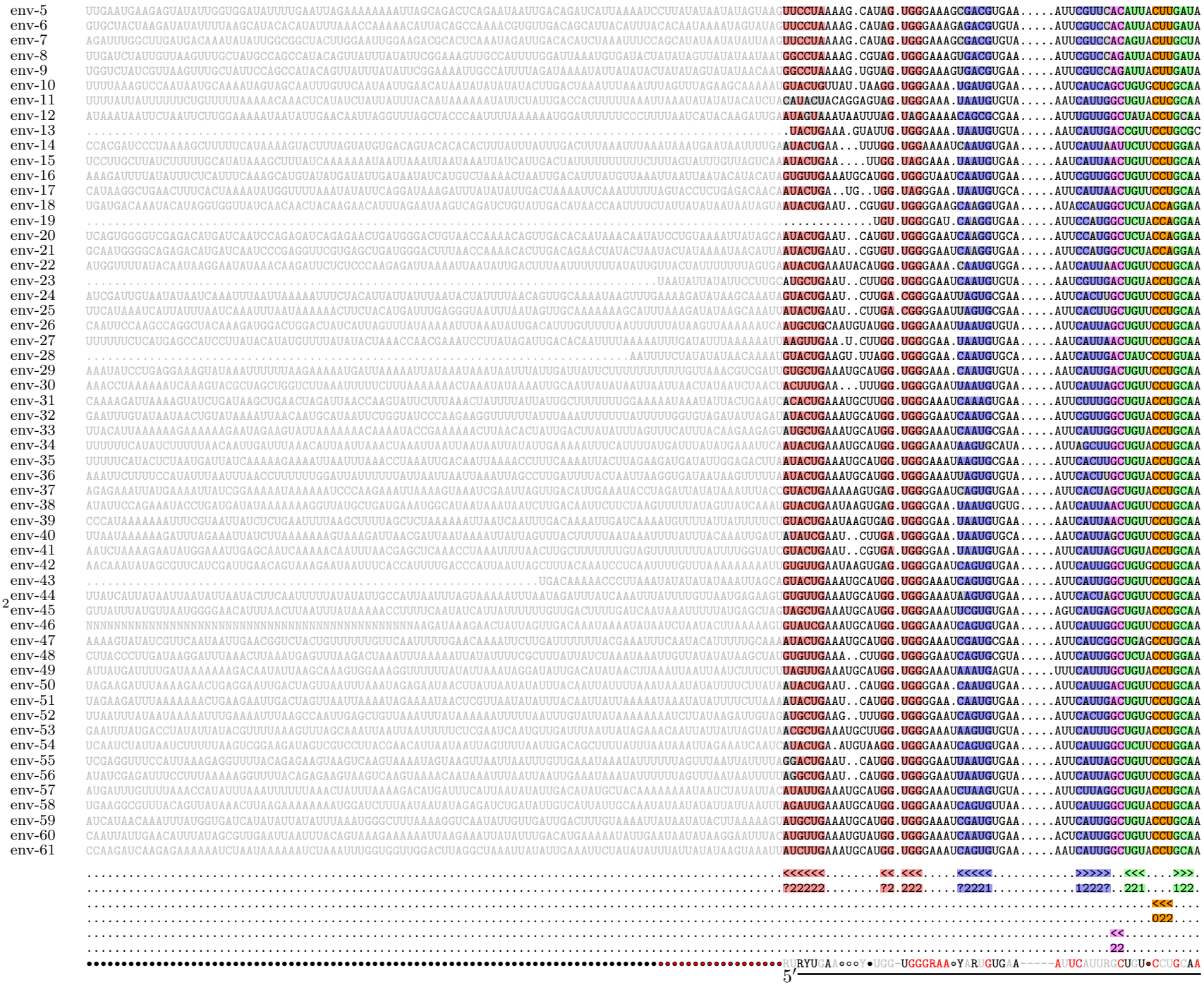
Conserved domains found in protein-coding genes listed in Section 6.3 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one AdoCbl-variant RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

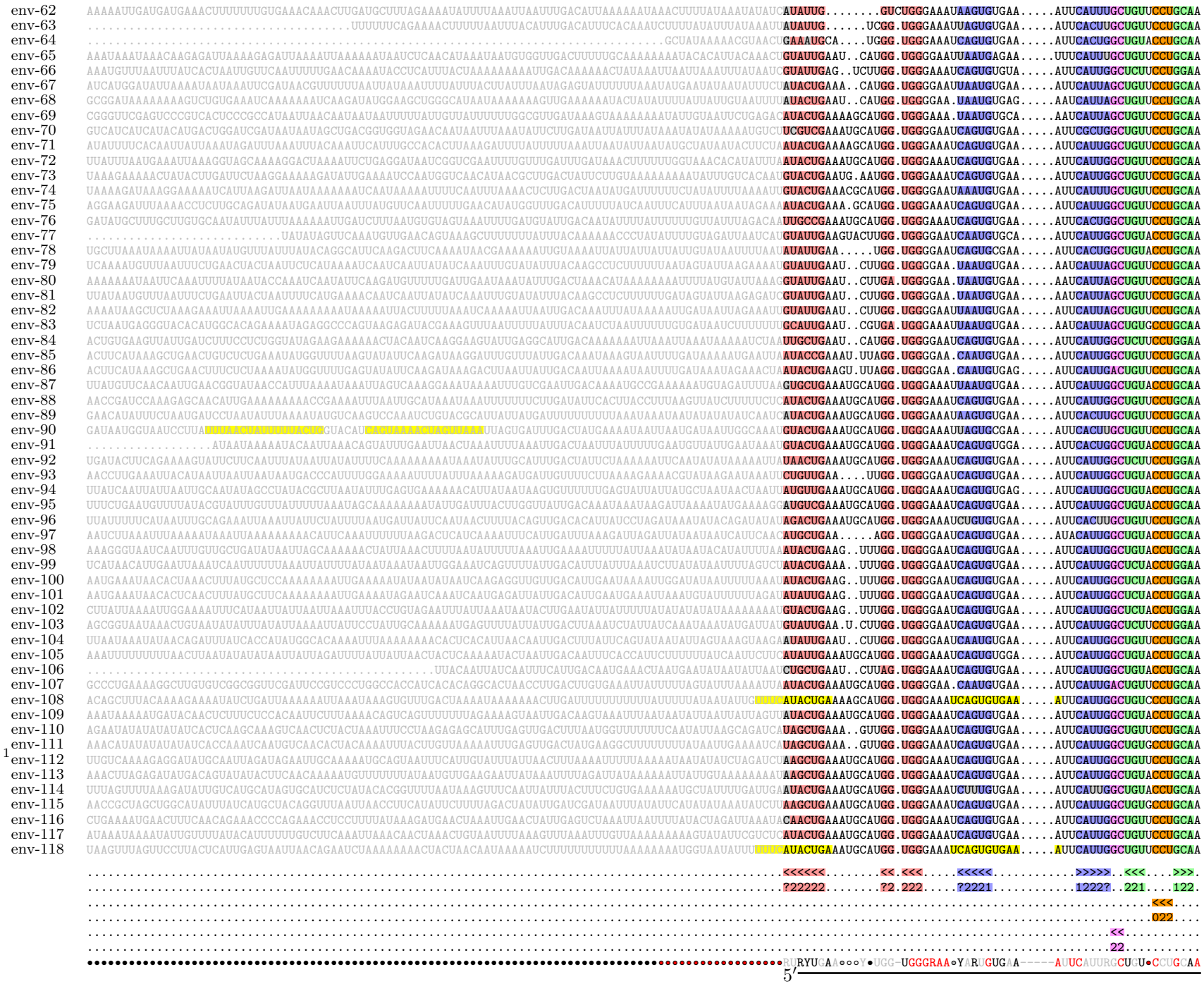
cd00712 (1) Glutamine amidotransferases class-II (GATase) asparagine synthase_B type.
cd01347 (1) TonB dependent/Ligand-Gated channels are created by a monomeric 22 strand (22,24) anti-parallel beta-barrel.
COG4206 (19) Outer membrane cobalamin receptor protein [Coenzyme metabolism]
COG4771 (1) Outer membrane receptor for ferrienterochelin and colicins [Inorganic ion transport and metabolism]
GOS11081972 (4) ?
GOS11094971 (2) ?
GOS11245591 (1) ?
GOS11622442 (1) ?
GOS11714930 (2) ?
GOS11790770 (6) ?
GOS11983570 (2) ?
GOS1720 (53) ?
GOS4369 (31) ?
GOS4665 (12) ?
GOS496 (6) ?
GOS5954 (1) ?
GOS729 (1) ?
GOS920 (13) ?
GOS9454 (3) ?
pfam00593 (3) TonB dependent receptor.
pfam02511 (1) Thymidylate synthase complementing protein.

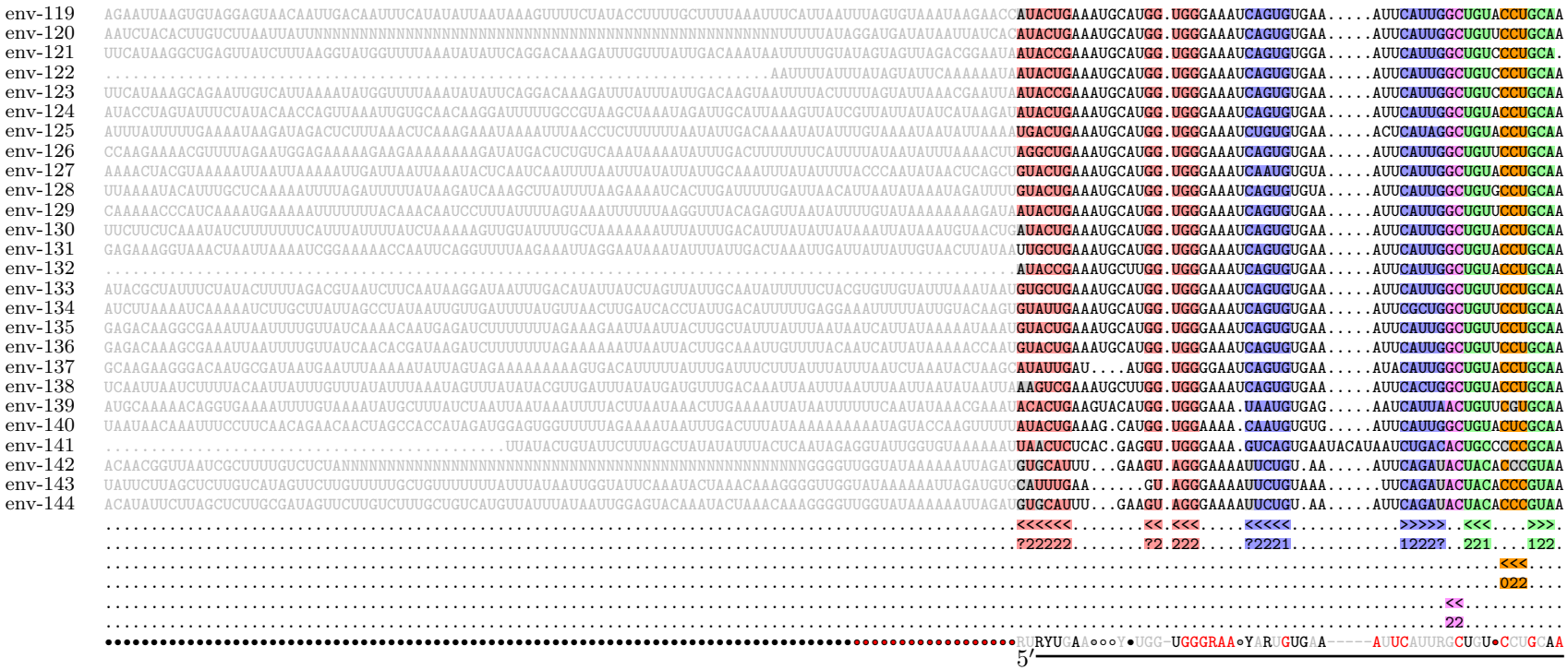
6.5 Multiple-sequence alignment

Each AdoCbl-variant RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 6.1. Superscript numbers refer to annotations listed in Section 6.2. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The AdoCbl-variant RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

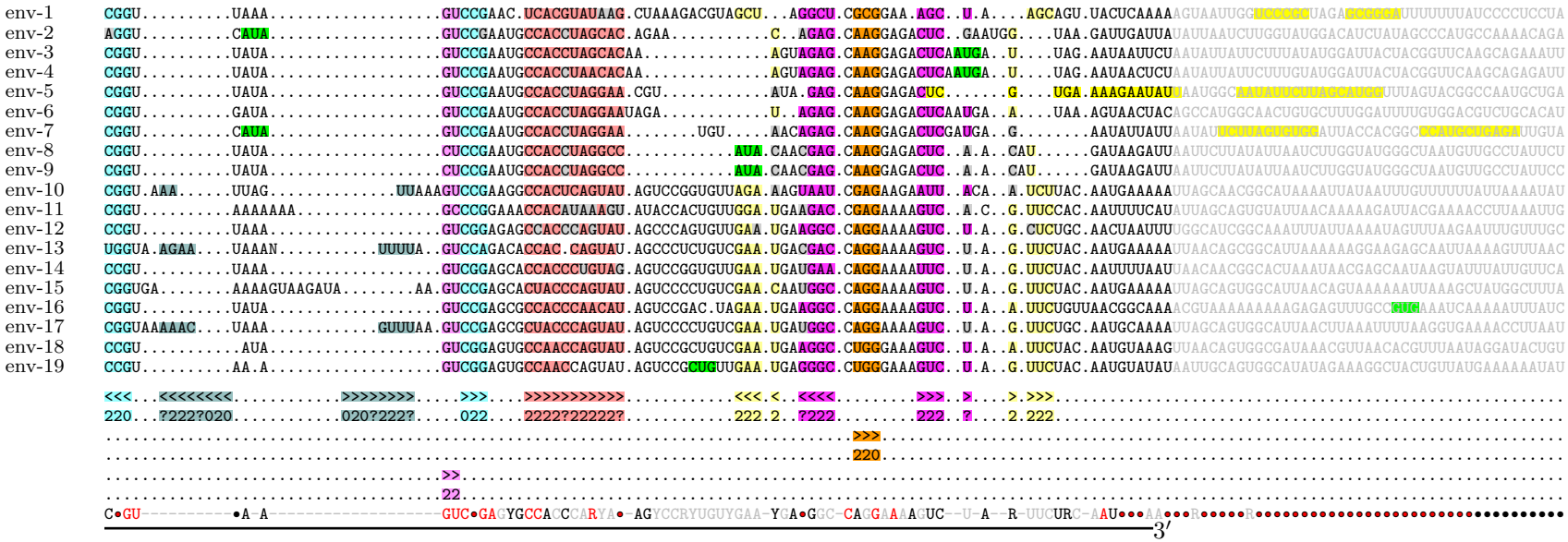


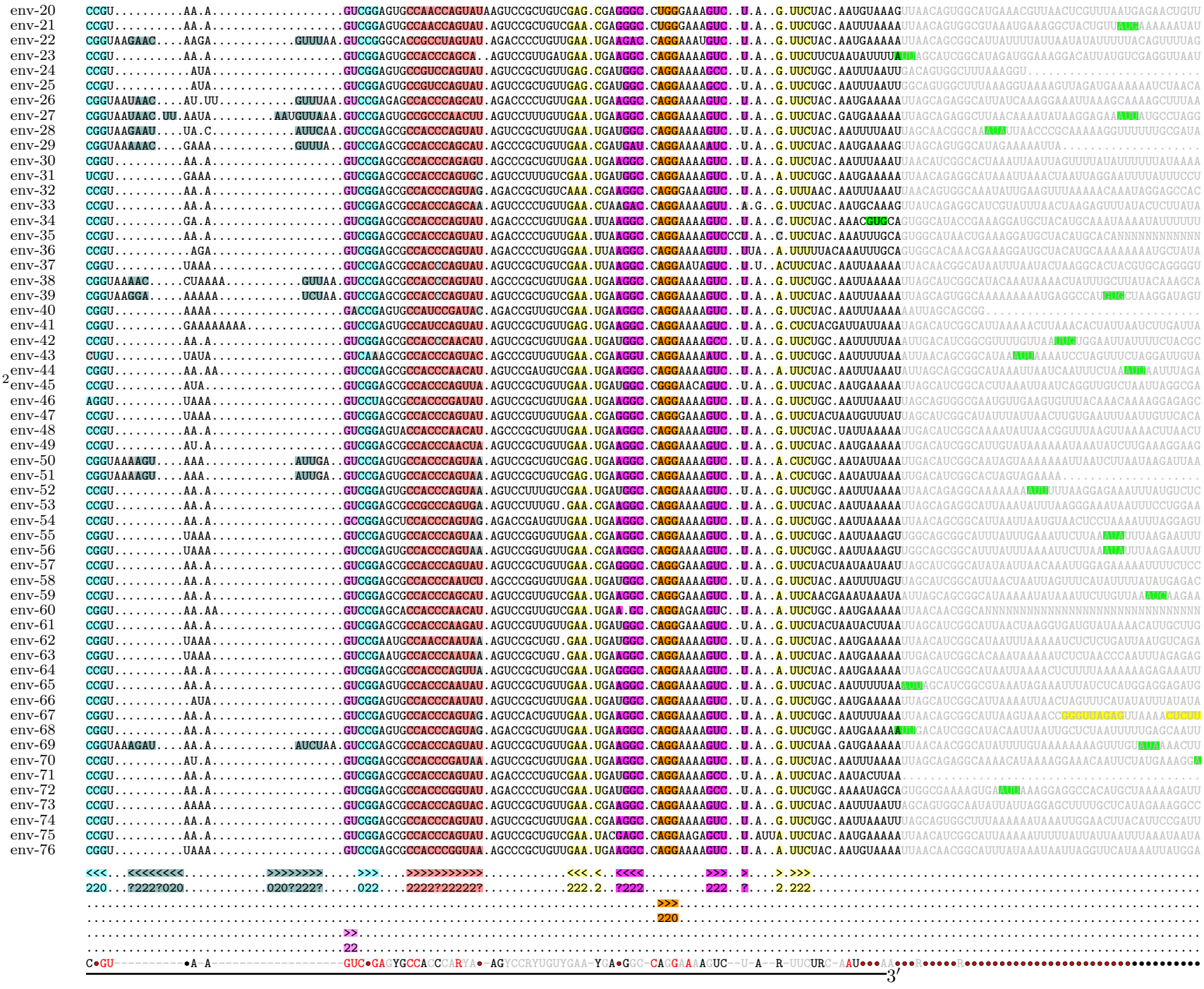

env-5
env-6
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
2env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
5′


env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
5′

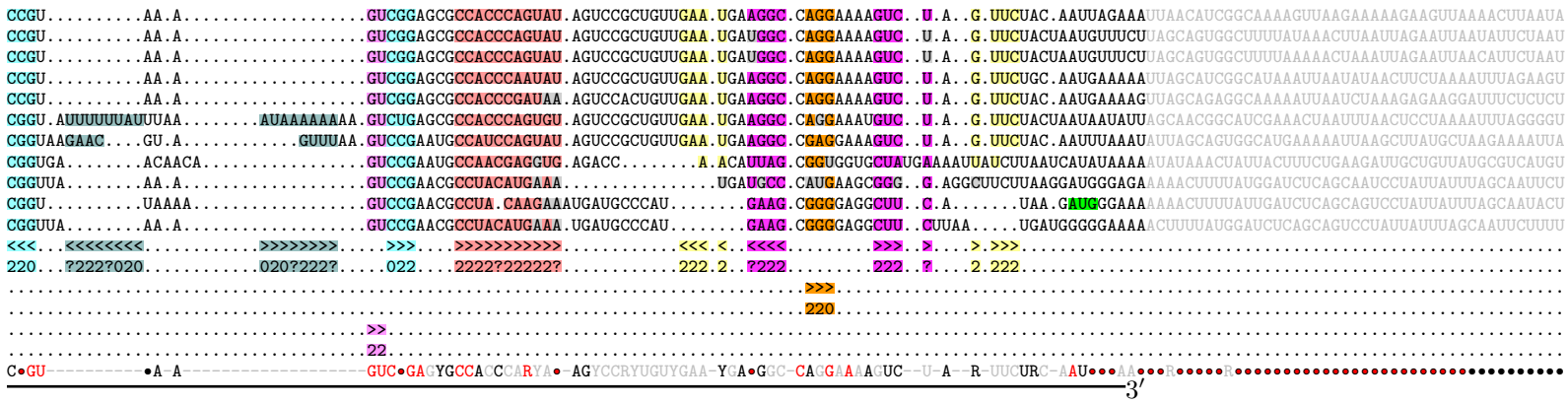


alignment positions 161 · · · 320

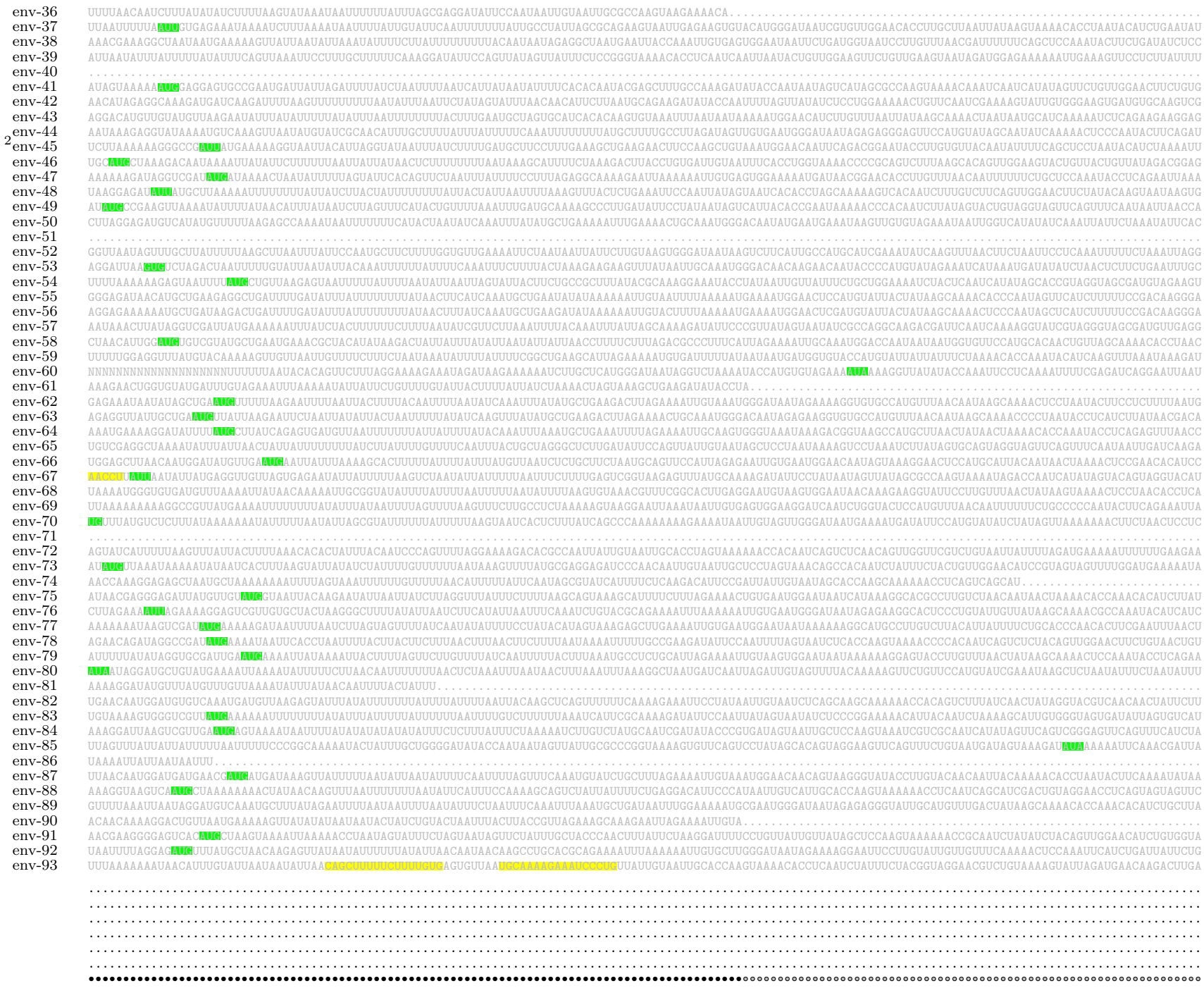


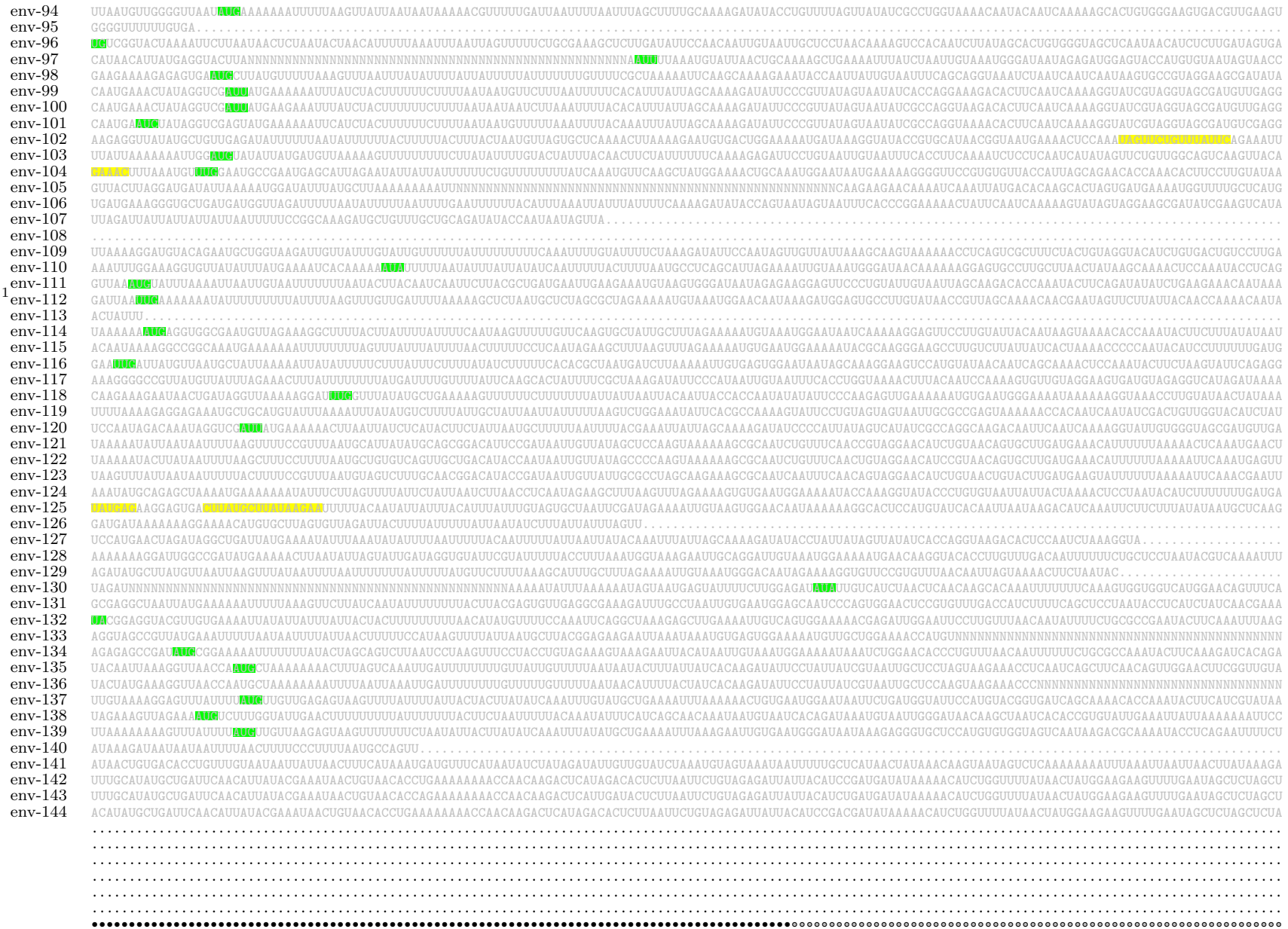

env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
2env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
3′

env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
1
3′



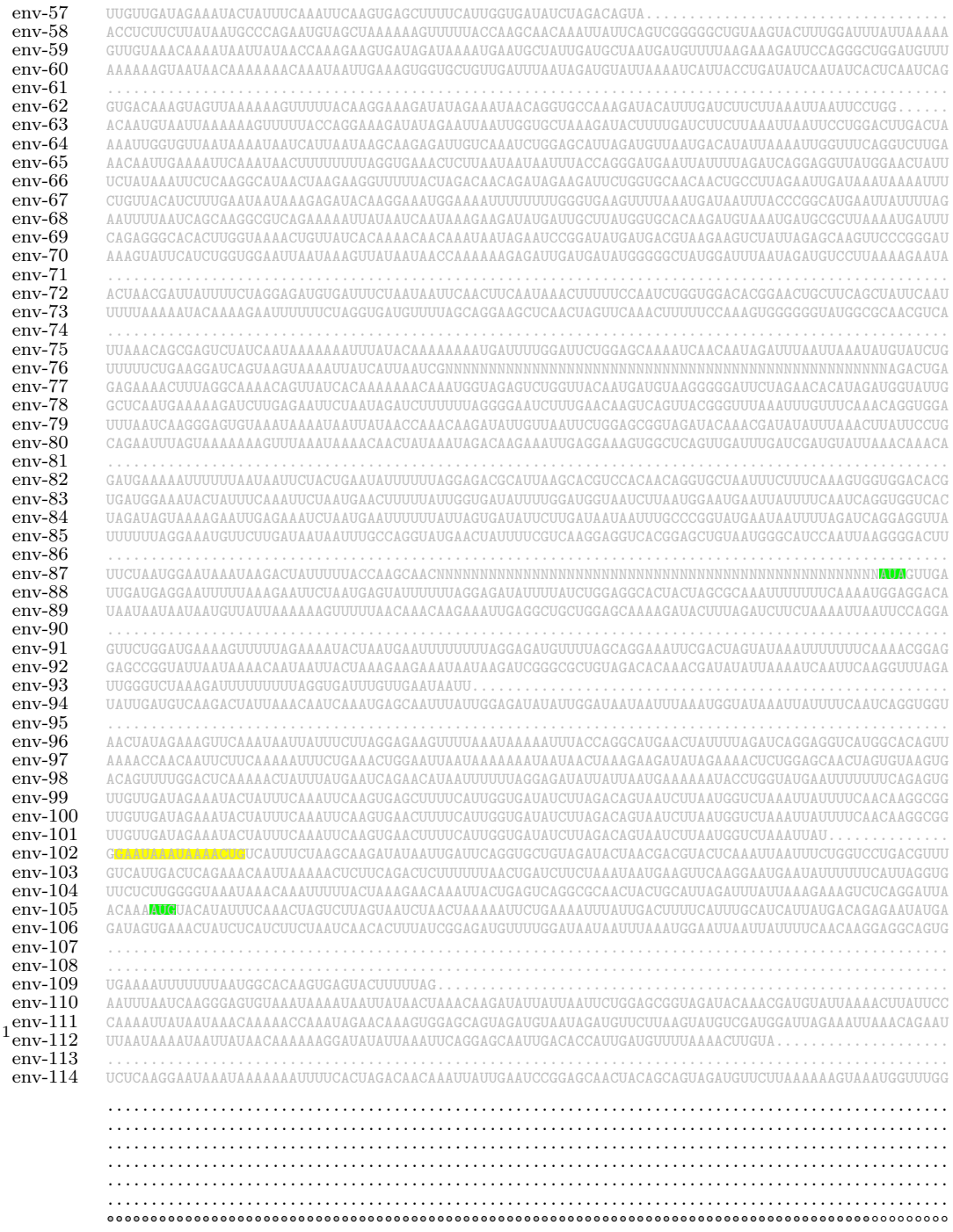
alignment positions 321 · · · 480


env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93


env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-135
env-136
env-137
env-138
env-139
env-140
env-141
env-142
env-143
env-144

alignment positions 481 · · · 577

```
env-1   CAUCAGUUCAAGAUGAAGACACAGAUGAGCUAGUCCAGGGCACAGCAUCUACAUUAUUUUCCCCAACGCCAAUAGCAACCACUUCUUCACUUGGAUA
env-2   ACGAUUGCUGUGCAAUCUGGACCUACAGGACAACAGACUUCUGUGUUCAUGAGAGGAACCAACAGCAAUCAUACUCUGGUAUUGUUAAAUGGCAUAC
env-3   AAGAAAAUAACUUAAUUGAUUUACCUAAAGUUUUAAAAUUUAUAAAUGGCGUUAACAUAACACAGUCAGGCUCGACUGGACAGCAAACAUCUGUGUU
env-4   AAAAGAAUAAUUUAAUUGACUUACCCAAAGUUUUAAAUUUUGUUAAUGGCAUUAAUAUAACA.....................................
env-5   CAAGAAAUAGAAGCAAAUAACAUCAAGGAUCUAAAAUCUGUAAUGGCCCACGUCAACAAUGUGAUGGCUGUUCAGUCAGGUCCCACAGGACAGCAAA
env-6   GGCCUCGAGGACAAACUGUUUCGACGUUCACUAGGGGCACAAAUUCUAAUCACACUCUUUUCAUGAUCAACGGAUCACCGAUCACAGACCAUUCAAC
env-7   AUAAUCUCACCGAUUUAACAAAAGUUCUAAAAUUUGUAAACAAUGUGAAUGUUGUUCAGUCGGGACCAACCGGACAGCAGUCGUCGGUCUUUAUGCG
env-8   ACUAUGUUCAAGGCCUAGACAUAACUCAAUCAGGUCCGACAGGUCAACAAUCAUCUUUAUUUUUAAGAGGCACAAACUCUAAUCACACAUUAGUAUU
env-9   ACUAUGUUCAAGGUUUAGACAUAACUCAAUCAGGUCCUACAGGUCAACAAUCAUCGGUGUUUUUAAGAGGCACUAAUUCUAAUCAUACAUUAGUAUU
env-10  UACAGUUUUGAAAGUAUUAAAAAAAUAACGAUUACGAGAAAACAAAUUGAAC.....................................
env-11  UACUGUUACAGUUAUUGAUGUGGAUAAAAUUAACGAGUCCUCACAUGUUUCUUUAGCACAAAUUAUUGACGACAGCUCGACAAGCACAAACAUGUUU
env-12  .....................................
env-13  .....................................
env-14  GAAGGAUCAGUUAAUAAAAUUGUUAUUAA.....................................
env-15  ACUGUUUUUGAUGAAAAAUUUUUUAAAAAUUCAAAUAAAUUAUUUUUAGGAGAUGUUUUAAAAAACAAUUCAACCAGUAUUAAUUUUUUUCAAAAUG
env-16  AAACGAAUUUUUUUUAGGAGAUUUG.....................................
env-17  GUACUGGAUGAAAAUUUUUUUAAAAAUACUAAUCACUUAUUUUUAGGGGACGUUUUAAAAGGAAAUUCAAACAGUGUAAAUUUCUUUCAAAGUGGCG
env-18  CCAUUGUUGGCUAUGUUUGUGUCGGAUCUGUAUUUGGGGUUCCACUCCAGUAUGUUUUGGGUUUACAUUAGUAUCUUGUUAGGCACAACAUUGAGUC
env-19  AAAUACAAAAGUACAAUCUCAUAGACUUACCUAAGGUAUUGAACUAUGUGCAAGGAUUAGACAUCACACAGUCUGGACCAACAGGACAACAGAGUUC
env-20  .....................................
env-21  AAAUACAAAAGUACAAUCUCAUAGACUUACCU.....................................
env-22  CAGUUUAUGAUAAAAGUUACAUCGAAGGUACAAAUGGCUUUUUCUUAGGAGAUGUUUUGGGAAACGGGACUACUAGUUUUAACUAUUUUCAAACAGG
env-23  AUAAAUAAAAAUCAAAUAAUUGAAUCUGGAGCUAUUGAUUUAAUAGACGUACUUAAUACAGUUCCUAAUAUAAACAUUACCCAAAGCGGUCCUAGGG
env-24  .....................................
env-25  GAAGGUGUUCUUGGAAAAACAGUUAUUACAAAAAAACAAAUUAUCGAGUCUGGCUAUGAAGACGUGAGAAGUGUUUUAGAAUAUAUAGCUGGCGUAG
env-26  .....................................
env-27  UUAACAAAACUAUAAUUACAAAAAAACAAAUCGAAGAAAGUGGUGCAGUAGAUUUAAUUGACGUCUUAGAAUCUGUGCCCGAUAUUAAUAUCACUCA
env-28  GAUUGGUAUUAAUAAACAAAUUAUUACAAAACAAGAUAUAAUUGAUUCAGGUGCUAUAGACUCAAAUGAUAUUCUAAACUUAGUAUCUGGACUUGAU
env-29  .....................................
env-30  .....................................
env-31  AUUCAUCAGAAUUGUCUAAAAGUGGUAUUUAUAAAACUACGAUUUCAAAAAAAGAAAUUGACGAAAGUGGAGCAAUUGAUUUAAUCGAUAUUUUACA
env-32  CUAUAAGAAAUUCAAAAGAAAUUUUUCUUGGCGCAAUAAUUGAUCAAGAGUCUGGAAGCACAAAUACUUUCCAAGAUGGUGGACACGGCACUAACAU
env-33  UAUUCCAAAUCUGGUAUUAACAAAAUCAUAAUAAACAAAAAGCAAAUUGAAGAAAGUGGUGCAGUAGAUCUAAUUGACGUGUUAAAAACUAUUCCAG
env-34  AUUUUUUAGGAGAUAUCCUCUCUAAUAAUUCAACUUCAAUAAAUUUUUUUCAAUCUGGAGGACAUGGUACAGCUUCAGCAAUUCAAUUAAGGGGAUU
env-35  GGGAAAGUUUGAUAAUCAAUUUAAUGAUAGUUUGGAAAUUGCCGGAACGGUCAGAUUACGUGAUACUUAUCUUCAGUAUGAUAAAG...........
env-36  .....................................
env-37  UCAUCAGAUGGGGUUAAUAAAAUAGUUAUAACCAAGCAAGAUAUGAUAAAUGCAGGCGCCUCAGAUAUCAAUGAUGUUCUUAAGCUAAUACCCGGAU
env-38  CAGGAAAGUUUAGGUAAAAUUGUAAUUACCAAACAACAAAUGAUUAAUUCAGGCUAUCAAGAUGUAAGGAGUGUUCUAGAGAAUGUACCUGGUGUAG
env-39  ACGUUGGCUGAUGUUAUUAAUGAUAACUCAACUAGCACAAAUUUUUUUCAAGCUGGAGGUAUUGGGAGUAAUACGGGCAUUCAACUUAGGGGCUUAG
env-40  .....................................
env-41  ACCACUCUAAAUAAUAAAGAAAUUCAGGAGGCUGGGGAAUUUUUUUUAGCAGAUGUUUUAAAUGAUAGUUUACCAGGAAUGAAUUAUUUUAGAUCGG
env-42  UUGAUAAAAAAAUUUUAGAAAAUUCAAACCAAUUCUUUAUCGGAGAUAUUUUAGAUAAUAAUUUAAAUGGAAUAAAUUAUUUUCAAUCAGGUGGAUA
env-43  UAAUUAAAACAGUAAUAACCAAGCAAGAUAUAGAAAAUUCUGGAUAUACUAAUUUAACUGAUAUUCUAAAAACAAUUAGUGGGCUCAAUGCUUAUCA
env-44  GUUUCUGAAAAAUCUAUAAAUAAAACAGUAAUAAAUAGAGAAGAUAUUGAAAAGAGUGGAGCAAUAGAUGUCGUAGAUUUAUUAAAAUAUUUAGAUG
2env-45 UCGGAGGGAAGUCUAGGAAAAACAAUCAUUACCAAAAAACAAUUAAUUGAGUCUGGUUAUGAUGACGUUAGAAGUGUUCUUGAAAAUGAAAUAGGGG
env-46  AAACAAUCAGAAAUUCUAAUGAAAAUUUUCUUGGUGCAAUCAUCGACCAAAAAUCAGGAAGCACUAAUACUUUUCAGGAUGGAGGCCAUGGUACGAA
env-47  UGAAGGUUCACUUAACAAAGUGAUUAUUACAAAAAAACAGAUGGUGGACUCAGGGUAUGAGGAUGUCAGAGGUGUUUUAGAGUAUGUAGCUGGUUUA
env-48  AAAAGGAAAUUGAAAAAUUCUGGCGAGUUUUUUCUUGGUGAUAUACU.....................................
env-49  AGAAACUAUUUCCAAUUCUAAUAAUUUUUUUUUAGGUGAAACUCUUGAUAAAAAUUUACCAGGAAUGAAUUAUUUUAGAUCAGGCGGUUACGGAACC
env-50  AAUUUGGAAUUAACAAAGUUAU.....................................
env-51  .....................................
env-52  AAUAAACAAAACAGUAAUAACUAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAUCAACGAUCAAUCAACACU
env-53  AAAAAGGGGGAUUAAUAAAGUUGUUAUUUCAAAAAAACAAAUUGAUGAAAUGGGAGCUAUUGAUUUAAUUGAUGUACUGGCACAAGUGCCAGACAUU
env-54  UAUAGAUUCUGAAGCUUUAGAAAAAUCAGAACAUAAUUUUUUAGGGGAUAUACUUGAUAACGCUAUGCCUGGUGCGAAUUAUUUUCAAAGUGGGGGG
env-55  AUAAAUAAAAUAGUUAUAAAUAAAGAGGAUAUUAACAAUUUCUGGCGCAUUAGAUACAAAUGAUGUACUUAAGUAUAUUUCGGGUUUGGAUGUCUUUC
env-56  AUAAACAAAAUAGUUAUAAAUAAAGAAGAUAUUAACAAUUCUGGCGCAUUAGAUACAAAUGAUGUACUUAAGUAUAUUUCGGGUUUGGAUGUCUUUC

.....
.....
.....
.....
.....
.....
°°°
```


env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114

7 *asd*

7.1 Taxa

The taxonomy of each organism containing a putative *asd* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *asd* RNA in Sections 7.2 and 7.4.

abbrev. of hits	taxonomy of species
Lla-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. cremoris* MG1363
Lla-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. cremoris* SK11
Lla-3-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. lactis* Il1403
Sag-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 18RS21
Sag-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 2603V/R
Sag-3-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 515
Sag-4-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* A909
Sag-5-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* CJB111
Sag-6-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* COH1
Sag-7-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* H36B
Sag-8-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* NEM316
Smu-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus mutans* UA159
Spn-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* D39
Spn-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* R6
Spn-3-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP11-BS70
Spn-4-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP14-BS69
Spn-5-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP18-BS74
Spn-6-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP19-BS75
Spn-7-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP23-BS72
Spn-8-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP3-BS71
Spn-9-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP6-BS73
Spn-10-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP9-BS68
Spn-11-1 to Spn-11-2	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* TIGR4
Spy-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* M1 GAS
Spy-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* M49 591
Spy-3-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10270
Spy-4-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10394
Spy-5-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10750
Spy-6-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS2096
Spy-7-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS315
Spy-8-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS5005
Spy-9-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS6180
Spy-10-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS8232
Spy-11-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS9429
Spy-12-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* SSI-1
Spy-13-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes str.* Manfredo
Ssa-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus sanguinis* SK36
Ssu-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus suis* 05ZYH33
Ssu-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus suis* 89/1591
Ssu-3-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus suis* 98HAH33
Sth-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* CNRZ1066
Sth-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* LMD-9
Sth-3-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* LMG 18311

7.2 Gene contexts

Each *asd* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *asd* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 7.3. The accession number of the sequence containing each *asd* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 7.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Lla-3-1	NC_002662.1	-	1668839	1668682	hypo→ **RNA**→ MIP (cd00333)→ Asd (COG0136)→ COG1051 (COG1051)→
Lla-1-1	NC_009004.1	+	839782	839941	hypo→ **RNA**→ MIP (cd00333)→ Asd (COG0136)→ COG1051 (COG1051)→ COG1051 (COG1051)→
Lla-2-1	NC_008527.1	-	1641549	1641391	hypo→ **RNA**→ MIP (cd00333)→ Asd (COG0136)→ COG1051 (COG1051)→
Spy-10-1	NC_003485.1	-	972072	971922	Cls (COG1502)→ **RNA**→ ATP-cone (pfam03477)→
Spy-1-1	NC_002737.1	-	996147	995997	Cls (COG1502)→ **RNA**→ ATP-cone (pfam03477)→
Spy-4-1	NC_006086.1	-	919140	918990	Cls (COG1502)→ **RNA**→ ATP-cone (pfam03477)→
Spy-9-1	NC_007296.1	-	916998	916848	Cls (COG1502)→ **RNA**→ ATP-cone (pfam03477)→
Spy-5-1	NC_008024.1	-	1020618	1020468	Cls (COG1502)→ **RNA**→ ATP-cone (pfam03477)→
Spy-2-1	NZ_AAFV01000017.1	-	782	632	Cls (COG1502)→ **RNA**→ ATP-cone (pfam03477)→
Spy-8-1	NC_007297.1	-	914543	914393	Cls (COG1502)→ **RNA**→ ATP-cone (pfam03477)→
Spy-11-1	NC_008021.1	-	986220	986070	Cls (COG1502)→ **RNA**→ ATP-cone (pfam03477)→
Spy-6-1	NC_008023.1	-	946984	946834	Cls (COG1502)→ **RNA**→ ATP-cone (pfam03477)→
Spy-13-1	NC_009332.1	+	878900	879050	Cls (COG1502)→ **RNA**→ ATP-cone (pfam03477)→
Spy-12-1	NC_004606.1	-	1038560	1038410	Cls (COG1502)→ **RNA**→ ATP-cone (pfam03477)→
Spy-7-1	NC_004070.1	-	908040	907890	Cls (COG1502)→ **RNA**→ ATP-cone (pfam03477)→
Spy-3-1	NC_008022.1	-	989456	989306	Cls (COG1502)→ **RNA**→ ATP-cone (pfam03477)→
Smu-1-1	NC_004350.1	+	935481	935617	Cls (COG1502)→ **RNA**→ Asd (COG0136)→ DHDPS (cd00950)→
Ssu-2-1	NZ_AAFA02000005.1	+	26347	26496	hypo (cons)→ **RNA**→ ←PflA (COG1180)
Ssu-1-1	NC_009442.1	+	684117	684266	hypo→ **RNA**→ ←PflA (COG1180)
Ssu-3-1	NC_009443.1	+	683851	684000	hypo→ **RNA**→ ←PflA (COG1180)
Sth-3-1	NC_006448.1	-	1153119	1152969	COG4912 (COG4912)→ **RNA**→ Asd (COG0136)→
Sth-2-1	NC_008532.1	-	1185932	1185782	COG4912 (COG4912)→ **RNA**→ Asd (COG0136)→
Sth-1-1	NC_006449.1	-	1150990	1150840	COG4912 (COG4912)→ **RNA**→ Asd (COG0136)→
Ssa-1-1	NC_009009.1	-	1221575	1221437	RimL (COG1670)→ **RNA**→ Asd (COG0136)→ DHDPS (cd00950)→ hypo→
Sag-4-1	NC_007432.1	-	1119613	1119477	Cls (COG1502)→ **RNA**→ fibrinogen-binding protein, putative→
Sag-7-1	NZ_AAJS01000010.1	+	44065	44201	Cls (COG1502)→ **RNA**→
Sag-6-1	NZ_AAJR01000023.1	+	31271	31407	Cls (COG1502)→ **RNA**→ fibrinogen-binding protein→
Sag-5-1	NZ_AAJQ01000001.1	+	73469	73605	Cls (COG1502)→ **RNA**→ Unknown→
Sag-2-1	NC_004116.1	-	1063292	1063156	Cls (COG1502)→ **RNA**→ ←hypo
Sag-8-1	NC_004368.1	-	1133913	1133777	Cls (COG1502)→ **RNA**→ fbsA→
Sag-1-1	NZ_AAJO01000112.1	+	4936	5072	Cls (COG1502)→ **RNA**→ fibrinogen receptor FbsA→
Sag-3-1	NZ_AAJP01000010.1	+	35943	36079	Cls (COG1502)→ **RNA**→ fibrinogen-binding protein→
Spn-9-1	NZ_ABAA01000002.1	+	167881	168023	tRNA-Thr→ **RNA**→ Asd (COG0136)→ DHDPS (cd00950)→
Spn-5-1	NZ_ABAE01000003.1	-	127675	127533	tRNA-Thr→ **RNA**→ Asd (COG0136)→ DHDPS (cd00950)→
Spn-11-1	NC_003028.2	+	956780	956922	tRNA-Thr→ **RNA**→ Asd (COG0136)→ DHDPS (cd00950)→
Spn-11-2	NZ_AAGY02000049.1	+	10416	10558	COG4708 (COG4708)→ **RNA**→ Asd (COG0136)→ DHDPS (cd00950)→
Spn-1-1	NC_008533.1	+	912569	912711	tRNA-Thr→ **RNA**→ Asd (COG0136)→ DHDPS (cd00950)→
Spn-2-1	NC_003098.1	+	905038	905180	tRNA-Thr→ **RNA**→ Asd (COG0136)→ DHDPS (cd00950)→
Spn-8-1	NZ_AAZZ01000006.1	+	11563	11705	tRNA-Thr→ **RNA**→ Asd (COG0136)→ DHDPS (cd00950)→
Spn-7-1	NZ_ABAG01000003.1	+	110887	111029	tRNA-Thr→ **RNA**→ Asd (COG0136)→ DHDPS (cd00950)→
Spn-6-1	NZ_ABAF01000013.1	+	125	268	hypo→ **RNA**→ Asd (COG0136)→ DHDPS (cd00950)→
Spn-10-1	NZ_ABAB01000021.1	-	32130	31987	tRNA-Thr→ **RNA**→ Asd (COG0136)→ DHDPS (cd00950)→
Spn-3-1	NZ_ABAC01000013.1	+	112	255	hypo→ **RNA**→ Asd (COG0136)→ DHDPS (cd00950)→
Spn-4-1	NZ_ABAD01000016.1	+	11693	11836	tRNA-Thr→ **RNA**→ Asd (COG0136)→ DHDPS (cd00950)→

7.3 Conserved domains

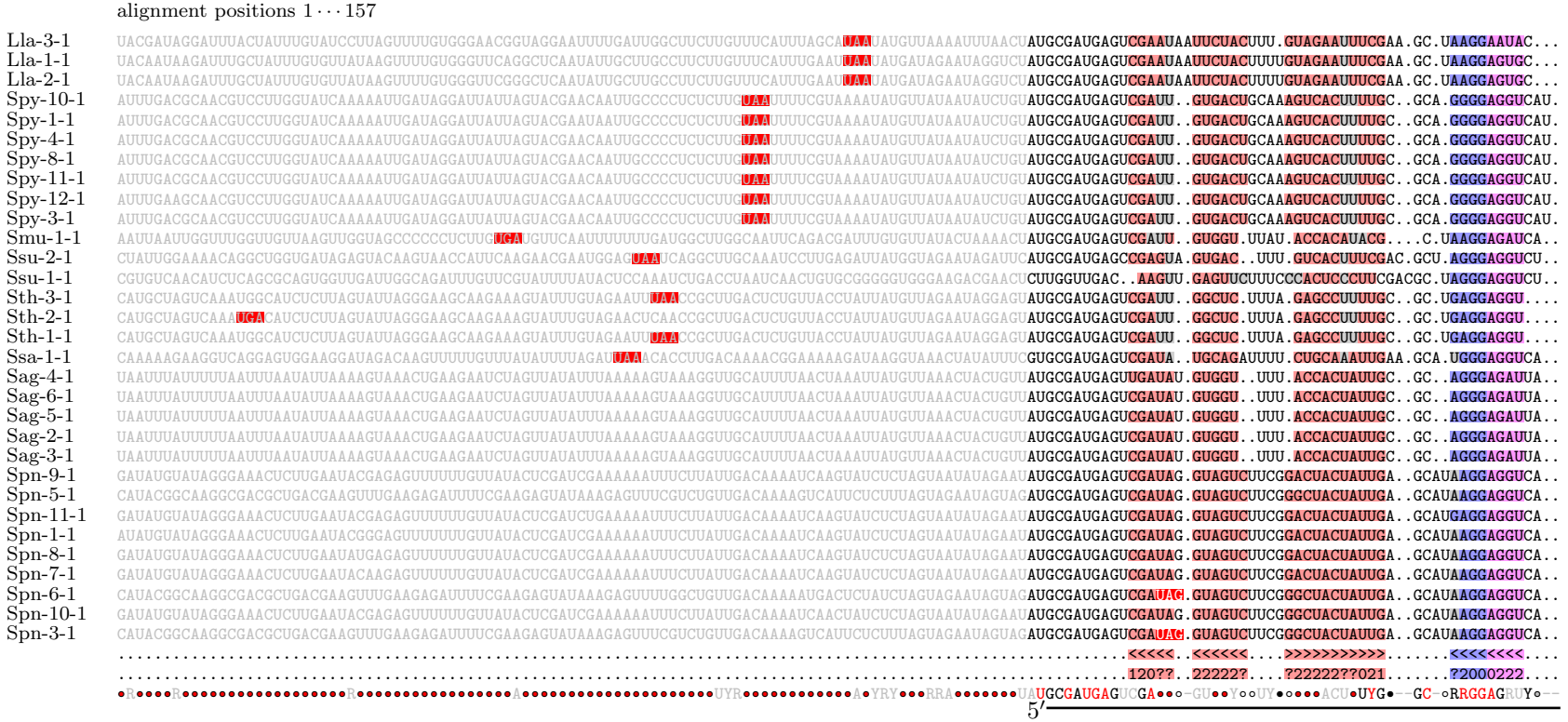
Conserved domains found in protein-coding genes listed in Section 7.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *asd* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

cd00333 (3) Major intrinsic protein (MIP) superfamily.
cd00950 (14) Dihydrodipicolinate synthase (DHDPS) is a key enzyme in lysine biosynthesis.
COG0136 (20) Aspartate-semialdehyde dehydrogenase [Amino acid transport and metabolism]
COG1051 (4) ADP-ribose pyrophosphatase [Nucleotide transport and metabolism]
COG1180 (3) Pyruvate-formate lyase-activating enzyme [Posttranslational modification, protein turnover, chaperones]
COG1502 (22) Phosphatidylserine/phosphatidylglycerophosphate/cardiolipin synthases and related enzymes [Lipid metabolism]
COG1670 (1) Acetyltransferases, including N-acetylases of ribosomal proteins [Translation, ribosomal structure and biogenesis]
COG4708 (1) Predicted membrane protein [Function unknown]
COG4912 (3) Predicted DNA alkylation repair enzyme [DNA replication, recombination, and repair]
pfam03477 (13) ATP cone domain.

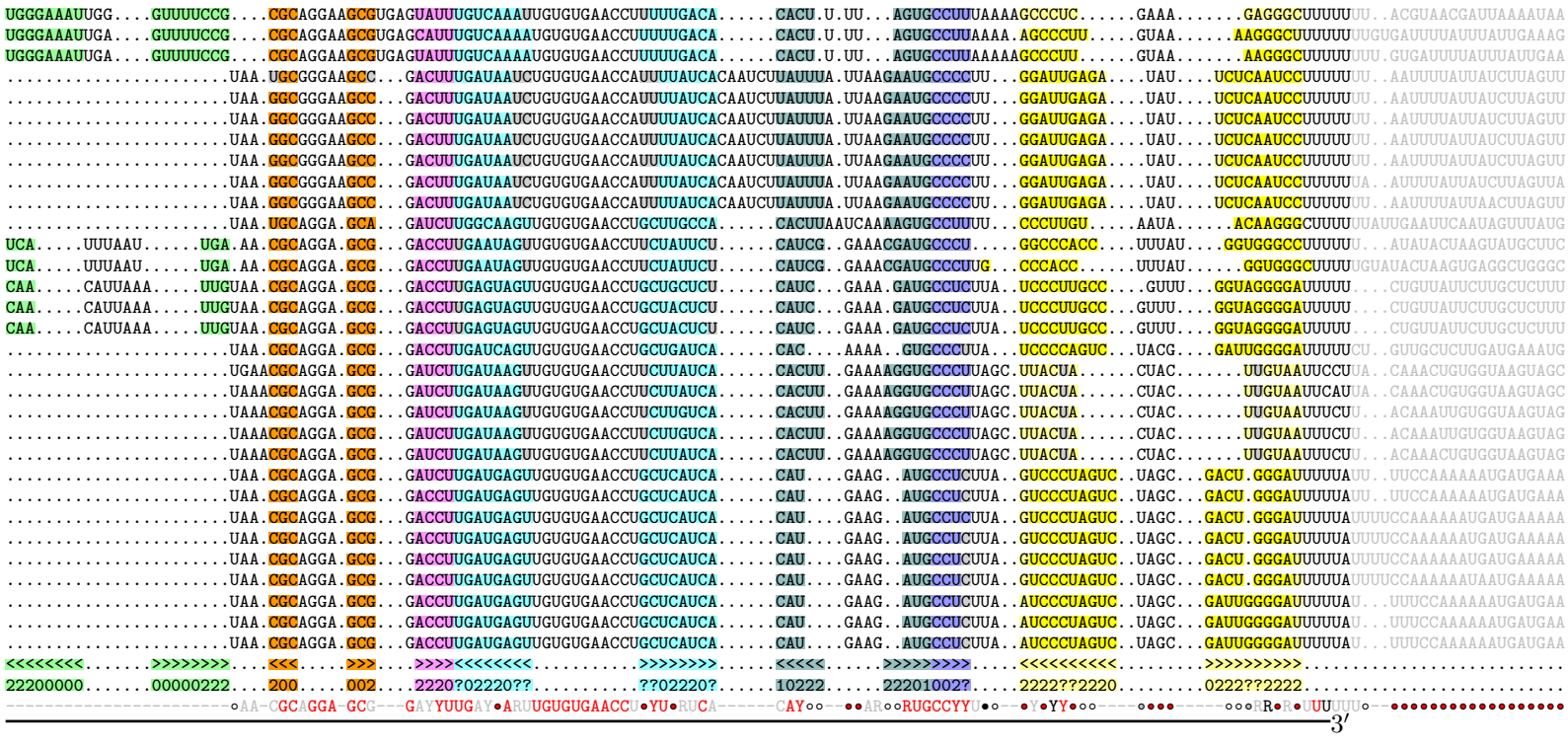
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Sag-1-1, Sag-7-1, Sag-8-1, Spn-11-2, Spn-2-1, Spn-4-1, Spy-13-1, Spy-2-1, Spy-5-1, Spy-6-1, Spy-7-1, Spy-9-1, Ssu-3-1

7.4 Multiple-sequence alignment

Each *asd* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 7.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *asd* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.


alignment positions 1 · · · 157
Lla-3-1
Lla-1-1
Lla-2-1
Spy-10-1
Spy-1-1
Spy-4-1
Spy-8-1
Spy-11-1
Spy-12-1
Spy-3-1
Smu-1-1
Ssu-2-1
Ssu-1-1
Sth-3-1
Sth-2-1
Sth-1-1
Ssa-1-1
Sag-4-1
Sag-6-1
Sag-5-1
Sag-2-1
Sag-3-1
Spn-9-1
Spn-5-1
Spn-11-1
Spn-1-1
Spn-8-1
Spn-7-1
Spn-6-1
Spn-10-1
Spn-3-1
5′

alignment positions 158 · · · 317



alignment positions 318 · · · 477

8 *atoC*

8.1 Taxa

The taxonomy of each organism containing a putative *atoC* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *atoC* RNA in Sections 8.2 and 8.4.

abbrev. of hits	**taxonomy of species**
Gbe-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Geobacteraceae *Geobacter bemidjiensis* Bem
Glo-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Geobacteraceae *Geobacter lovleyi* SZ
Gme-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Geobacteraceae *Geobacter metallireducens* GS-15
Gsp-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Geobacteraceae *Geobacter sp.* FRC-32
Gsu-1-1 to Gsu-1-2	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Geobacteraceae *Geobacter sulfurreducens* PCA
Gur-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Geobacteraceae *Geobacter uraniireducens* Rf4
Ppr-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Pelobacteraceae *Pelobacter propionicus* DSM 2379

8.2 Gene contexts

Each *atoC* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *atoC* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 8.3. The accession number of the sequence containing each *atoC* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 8.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Gsu-1-1	NC_002939.4	+	1810322	1810409	**RNA**→ AtoC (COG2204)→
Glo-1-1	NZ_AAVG01000035.1	+	27997	28086	**RNA**→ HPPK (cd00483)→ COG3350 (COG3350)→ Transaldolase_FSA (cd00956)→
					REC (cd00156)GSPII_E_N (pfam05157)→
Ppr-1-1	NC_008609.1	-	292999	292905	**RNA**→ FolK (COG0801)→ COG3350 (COG3350)→ Transaldolase_FSA (cd00956)→
Gbe-1-1	NZ_AAZA01000001.1	+	121278	121368	**RNA**→ COG1146 (COG1146)RocR (COG3829)Fe_hyd_lg_C (pfam02906)→
Gsu-1-2	NC_002939.4	+	3694623	3694713	**RNA**→ COG1144 (COG1144)RocR (COG3829)Fe_hyd_lg_C (pfam02906)FeS (pfam04060)→ hypo→
Gme-1-1	NC_007517.1	+	67665	67752	**RNA**→ COG1146 (COG1146)RocR (COG3829)Fe_hyd_lg_C (pfam02906)FeS (pfam04060)→
Gsp-1-1	NZ_AASH01000003.1	-	37660	37569	**RNA**→ COG1146 (COG1146)FhlA (COG3604)Fe_hyd_lg_C (pfam02906)FeS (pfam04060)→
Gur-1-1	NC_009483.1	-	4839082	4838991	**RNA**→ COG1144 (COG1144)AtoC (COG2204)Fe_hyd_lg_C (pfam02906)FeS (pfam04060)→

8.3 Conserved domains

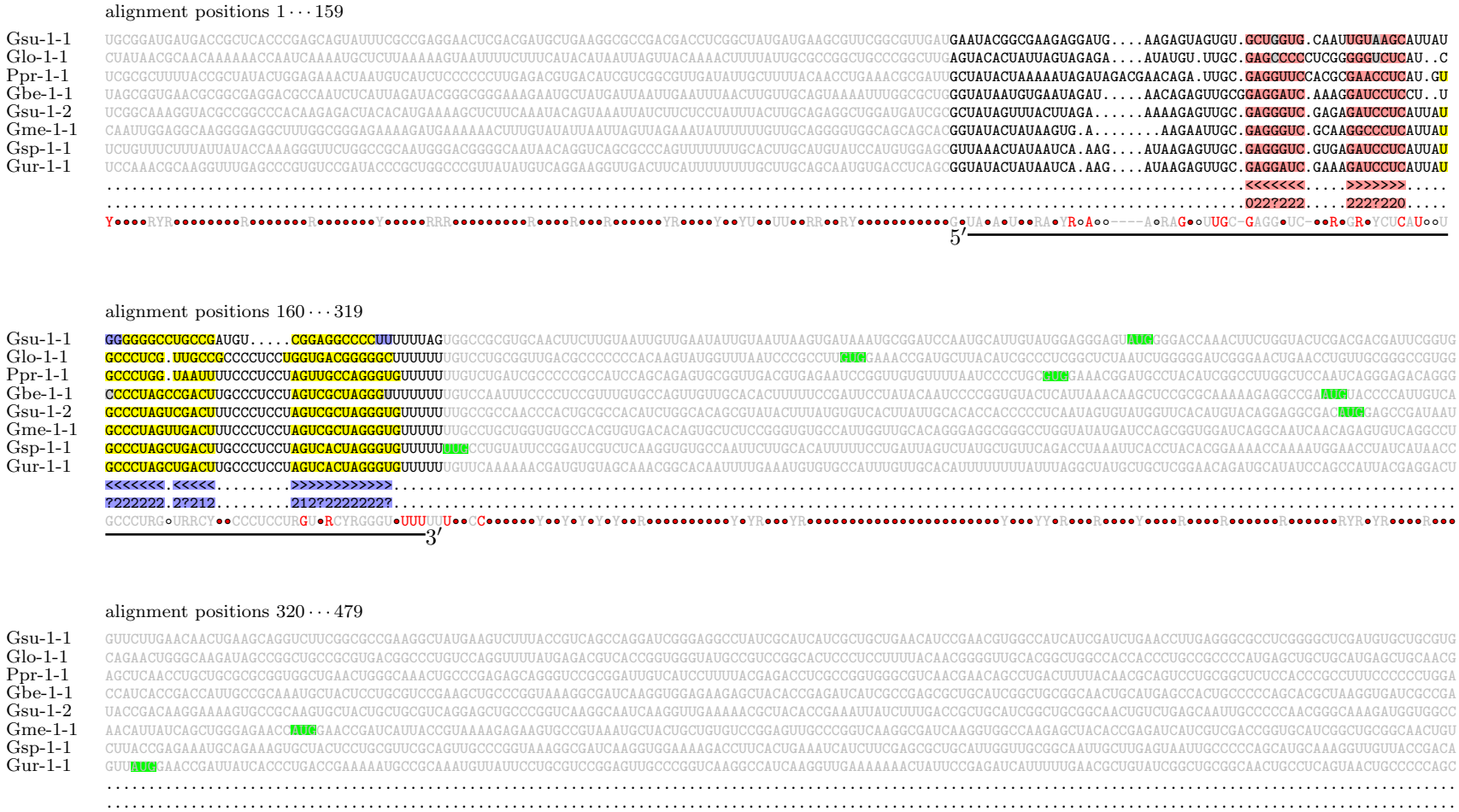
Conserved domains found in protein-coding genes listed in Section 8.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *atoC* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

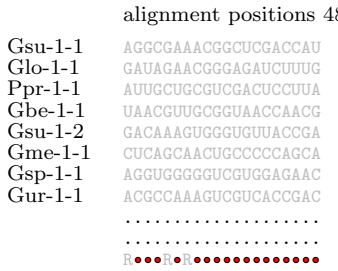
cd00156 (1) Signal receiver domain; originally thought to be unique to bacteria (CheY, OmpR, NtrC, and PhoB), now recently identified in eukaroytes ETR1 Arabidopsis thaliana; this domain receives the signal from the sensor partner in a two-component systems; contains a phosphoacceptor site that is phosphorylated by histidine kinase homologs; usually found N-terminal to a DNA binding effector domain; forms homodimers
cd00483 (1) 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase (HPPK).
cd00956 (2) Transaldolase-like fructose-6-phosphate aldolases (FSA) found in bacteria and archaea, which are member of the MipB/TalC subfamily of class I aldolases.
COG0801 (1) 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase [Coenzyme metabolism]
COG1144 (2) Pyruvate:ferredoxin oxidoreductase and related 2-oxoacid:ferredoxin oxidoreductases, delta subunit [Energy production and conversion]
COG1146 (3) Ferredoxin [Energy production and conversion]
COG2204 (2) Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains [Signal transduction mechanisms]
COG3350 (2) Uncharacterized conserved protein [Function unknown]
COG3604 (1) Transcriptional regulator containing GAF, AAA-type ATPase, and DNA binding domains [Transcription / Signal transduction mechanisms]
COG3829 (3) Transcriptional regulator containing PAS, AAA-type ATPase, and DNA-binding domains [Transcription / Signal transduction mechanisms]
pfam02906 (5) Iron only hydrogenase large subunit, C-terminal domain.
pfam04060 (4) Putative Fe-S cluster.
pfam05157 (1) GSPII_E N-terminal domain.

8.4 Multiple-sequence alignment

Each *atoC* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 8.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *atoC* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently mis-

annotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.





9 Bacillaceae-1

9.1 Taxa

The taxonomy of each organism containing a putative Bacillaceae-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Bacillaceae-1 RNA in Sections 9.2 and 9.4.

abbrev. of hits	**taxonomy of species**
Bam-1-1 to Bam-1-2	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus amyloliquefaciens* FZB42
Bcl-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus clausii* KSM-K16
Bco-1-1 to Bco-1-6	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus coahuilensis m*4-4
Bha-1-1 to Bha-1-4	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus halodurans* C-125
Bli-1-1 to Bli-1-46	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus licheniformis* ATCC 14580 (DSM 13)
Bpu-1-1 to Bpu-1-2	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus pumilus* ATCC 7061
Bpu-2-1 to Bpu-2-6	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus pumilus* SAFR-032
Bsp-1-1 to Bsp-1-35	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus sp.* NRRL B-14911
Bsu-1-1 to Bsu-1-16	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus subtilis subsp. subtilis str.* 168
Bsu-2-1 to Bsu-2-8	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus subtilis subsp. subtilis str.* JH642
Bsu-3-1 to Bsu-3-8	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus subtilis subsp. subtilis str.* NCIB 3610
Bsu-4-1 to Bsu-4-8	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus subtilis subsp. subtilis str.* SMY
Gka-1-1 to Gka-1-2	Bacteria Firmicutes Bacillales Bacillaceae *Geobacillus kaustophilus* HTA426
Oih-1-1 to Oih-1-5	Bacteria Firmicutes Bacillales Bacillaceae *Oceanobacillus iheyensis* HTE831

9.2 Gene contexts

Each Bacillaceae-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Bacillaceae-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 9.3. The accession number of the sequence containing each Bacillaceae-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 9.1.

abbrev.	**Seq. accession**		**5′ at**	**3′ at**	**genes**
Bli-1-1	NC_006270.3	+	1210076	1210143	hypo→ **RNA**→ ←DegV (COG1307)

Bli-1-2	NC_006322.1	+	1209927	1209994	hypo→ **RNA**→ ←DegV (COG1307)
Bli-1-3	NC_006270.3	+	3243766	3243832	←hypo **RNA**→ ←Ndh (COG1252)
Bli-1-4	NC_006322.1	+	3243618	3243684	←YuiA **RNA**→ ←Ndh (COG1252)
Bli-1-5	NC_006270.3	+	800841	800907	MOSC (pfam03473)3-alpha (pfam03475)→ **RNA**→ ←hypo
Bli-1-6	NC_006322.1	+	800647	800713	MOSC (pfam03473)3-alpha (pfam03475)→ **RNA**→ ←YflJ
Bli-1-7	NC_006270.3	+	712173	712240	CypX (COG2124)→ **RNA**→ hypo→
Bli-1-8	NC_006322.1	+	711979	712046	CypX (COG2124)→ **RNA**→ ANTAR (pfam03861)→
Bsu-1-1	NC_000964.2	-	3773345	3773279	hypo→ **RNA**→ FdhD-NarQ (pfam02634)→
Bsu-1-2	NZ_ABQK01000005.1	-	2754870	2754804	**RNA**→
Bsu-3-1	NZ_ABQL01000005.1	-	2754915	2754849	**RNA**→
Bsu-2-1	NZ_ABQM01000009.1	-	598692	598626	**RNA**→
Bsu-4-1	NZ_ABQN01000009.1	-	598702	598636	**RNA**→
Bsu-1-3	NC_000964.2	+	318564	318631	Sugar_tr (pfam00083)→ **RNA**→ ←AraJ (COG2814)
Bsu-1-4	NZ_ABQK01000002.1	+	148540	148607	**RNA**→
Bsu-3-2	NZ_ABQL01000002.1	+	124330	124397	**RNA**→
Bsu-2-2	NZ_ABQM01000002.1	+	148540	148607	**RNA**→
Bsu-4-2	NZ_ABQN01000004.1	+	92219	92286	**RNA**→
Bsu-1-5	NC_000964.2	+	514393	514460	COG3428 (COG3428)→ **RNA**→ ←Rhomboid (pfam01694)
Bsu-1-6	NZ_ABQK01000004.1	+	53417	53484	**RNA**→
Bsu-3-3	NZ_ABQL01000004.1	+	53417	53484	**RNA**→
Bsu-2-3	NZ_ABQM01000006.1	+	22575	22642	**RNA**→
Bsu-4-3	NZ_ABQN01000006.1	+	53419	53486	**RNA**→
Gka-1-1	NC_006510.1	+	32949	33016	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Gka-1-2	NC_006510.1	-	2946074	2946006	16S ribosomal RNA→ **RNA**→ 23S ribosomal RNA→
Bpu-2-1	NC_009848.1	+	138237	138304	←Polysacc_deac_1 (pfam01522) **RNA**→ 16S ribosomal RNA→
Bpu-2-2	NC_009848.1	+	905312	905382	hypo→ **RNA**→ 16S ribosomal RNA→
Bpu-2-3	NC_009848.1	-	2738313	2738246	←COG3859 (COG3859) **RNA**→ 16S ribosomal RNA→
Bpu-2-4	NC_009848.1	+	9503	9570	GyrA (COG0188)→ **RNA**→ 16S ribosomal RNA→
Bpu-2-5	NC_009848.1	+	531946	532013	tRNA-Gly→ **RNA**→ 16S ribosomal RNA→
Bpu-1-1	NZ_ABRX01000002.1	-	133	66	←COG3859 (COG3859) **RNA**→
Bpu-1-2	NZ_ABRX01000010.1	-	918	851	conserved protein YgxA→ **RNA**→
Bpu-2-6	NC_009848.1	+	555169	555236	BofA (pfam07441)→ **RNA**→ 16S ribosomal RNA→
Bha-1-1	NC_002570.2	+	561399	561462	16S ribosomal RNA→ **RNA**→ 23S ribosomal RNA→
Bha-1-2	NC_002570.2	+	52868	52931	16S ribosomal RNA→ **RNA**→ 23S ribosomal RNA→
Bha-1-3	NC_002570.2	+	119801	119864	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Bha-1-4	NC_002570.2	+	1020202	1020265	16S ribosomal RNA→ **RNA**→ 23S ribosomal RNA→
Bli-1-9	NC_006270.3	+	3080604	3080669	←MetK (COG0192) **RNA**→ PEPCK_ATP (cd00484)PckA (COG1866)→
Bli-1-10	NC_006322.1	+	3080457	3080522	←MetK (COG0192) **RNA**→ PEPCK_ATP (cd00484)PckA (COG1866)→
Bsu-1-7	NC_000964.2	+	35162	35228	ribosomal RNA-23S→ **RNA**→ ribosomal RNA-5S→
Bsu-1-8	NZ_ABQK01000001.1	+	35162	35228	**RNA**→
Bsu-3-4	NZ_ABQL01000001.1	+	35162	35228	**RNA**→
Bsu-2-4	NZ_ABQM01000001.1	+	35162	35228	**RNA**→
Bsu-4-4	NZ_ABQN01000001.1	+	35162	35228	**RNA**→
Bli-1-11	NC_006270.3	+	163014	163080	23S ribosomal RNA→ **RNA**→ 5S ribosomal RNA→
Bli-1-12	NC_006322.1	+	162818	162884	23S ribosomal RNA→ **RNA**→ 5S ribosomal RNA→
Bli-1-13	NC_006270.3	-	3118594	3118528	23S ribosomal RNA→ **RNA**→ 5S ribosomal RNA→
Bli-1-14	NC_006322.1	-	3118447	3118381	23S ribosomal RNA→ **RNA**→ 5S ribosomal RNA→
Bli-1-15	NC_006270.3	+	616707	616773	23S ribosomal RNA→ **RNA**→ 5S ribosomal RNA→
Bli-1-16	NC_006322.1	+	616513	616579	23S ribosomal RNA→ **RNA**→ 5S ribosomal RNA→
Bli-1-17	NC_006270.3	+	14778	14844	23S ribosomal RNA→ **RNA**→ 5S ribosomal RNA→
Bli-1-18	NC_006322.1	+	14582	14648	23S ribosomal RNA→ **RNA**→ 5S ribosomal RNA→
Bli-1-19	NC_006270.3	+	100058	100124	23S ribosomal RNA→ **RNA**→ 5S ribosomal RNA→

Bli-1-20	NC_006322.1	+	99862	99928	23S ribosomal RNA→ **RNA**→ 5S ribosomal RNA→
Bli-1-21	NC_006270.3	+	39477	39543	23S ribosomal RNA→ **RNA**→ 5S ribosomal RNA→
Bli-1-22	NC_006322.1	+	39281	39347	23S ribosomal RNA→ **RNA**→ 5S ribosomal RNA→
Bli-1-23	NC_006270.3	-	2767403	2767336	←MGS1 (COG2256) **RNA**→ **RNA**→ **RNA**→ COG1959 (COG1959)→
Bli-1-25	NC_006270.3	-	2767338	2767272	(shown above)
Bli-1-27	NC_006270.3	-	2767468	2767401	(shown above)
Bli-1-24	NC_006322.1	-	2767256	2767189	←MGS1 (COG2256) **RNA**→ **RNA**→ **RNA**→ COG1959 (COG1959)→
Bli-1-26	NC_006322.1	-	2767191	2767125	(shown above)
Bli-1-28	NC_006322.1	-	2767321	2767254	(shown above)
Bcl-1-1	NC_006582.1	+	1440503	1440568	←hypo **RNA**→ ribosomal RNA-16S→
Oih-1-1	NC_004193.1	-	423489	423422	hypo→ **RNA**→ ←DUF72 (pfam01904)
Oih-1-2	NC_004193.1	-	3404930	3404863	COG3333 (COG3333)→ **RNA**→ G_glu_transpept (pfam01019)→
Oih-1-3	NC_004193.1	+	2452262	2452329	←hypo **RNA**→ ←TlyC (COG1253)
Oih-1-4	NC_004193.1	-	2350186	2350119	23S ribosomal RNA→ **RNA**→ 5S ribosomal RNA→
Oih-1-5	NC_004193.1	+	675233	675300	23S ribosomal RNA→ **RNA**→ 5S ribosomal RNA→
Bam-1-1	NC_009725.1	+	92818	92883	16S ribosomal RNA→ **RNA**→ 23S ribosomal RNA→
Bam-1-2	NC_009725.1	+	98703	98768	16S ribosomal RNA→ **RNA**→ 23S ribosomal RNA→
Bsu-1-9	NC_000964.2	-	634985	634919	←ribosomal RNA-16S **RNA**→ ←tRNA-Gly
Bsu-1-10	NZ_ABQK01000004.1	-	174009	173943	**RNA**→
Bsu-3-5	NZ_ABQL01000004.1	-	174009	173943	**RNA**→
Bsu-2-5	NZ_ABQM01000006.1	-	143167	143101	**RNA**→
Bsu-4-5	NZ_ABQN01000006.1	-	174011	173945	**RNA**→
Bsu-2-6	NZ_ABQM01000009.1	+	3026	3092	**RNA**→
Bsu-1-11	NC_000964.2	+	3177677	3177743	←ribosomal RNA-16S **RNA**→ COG3859 (COG3859)→
Bsu-3-6	NZ_ABQL01000005.1	+	2159247	2159313	**RNA**→
Bsu-1-12	NC_000964.2	-	9813	9747	←ribosomal RNA-16S **RNA**→ ←GyrA (COG0188)DNA_topoisoIV (pfam00521)DNA_gyraseA_C (pfam03989)
Bsu-1-13	NZ_ABQK01000001.1	-	9813	9747	**RNA**→
Bsu-3-7	NZ_ABQL01000001.1	-	9813	9747	**RNA**→
Bsu-2-7	NZ_ABQM01000001.1	-	9813	9747	**RNA**→
Bsu-4-6	NZ_ABQN01000001.1	-	9813	9747	**RNA**→
Bsu-1-14	NZ_ABQK01000005.1	+	2159201	2159267	**RNA**→
Bsu-1-15	NC_000964.2	-	946035	945969	←ribosomal RNA-16S **RNA**→ ←hypo
Bsu-1-16	NZ_ABQK01000004.1	-	485060	484994	**RNA**→
Bsu-3-8	NZ_ABQL01000004.1	-	485060	484994	**RNA**→
Bsu-2-8	NZ_ABQM01000006.1	-	454224	454158	**RNA**→
Bsu-4-7	NZ_ABQN01000006.1	-	485065	484999	**RNA**→
Bsu-4-8	NZ_ABQN01000009.1	+	3033	3099	**RNA**→
Bli-1-29	NC_006270.3	-	920927	920861	←16S ribosomal RNA **RNA**→ ←hypo
Bli-1-30	NC_006322.1	-	920778	920712	←16S ribosomal RNA **RNA**→ ←YgxA
Bli-1-31	NC_006270.3	+	3123307	3123373	←16S ribosomal RNA **RNA**→ ←PurR (COG1609)
Bli-1-32	NC_006322.1	+	3123159	3123225	←16S ribosomal RNA **RNA**→ ←PurR (COG1609)
Bli-1-33	NC_006270.3	-	9902	9836	←16S ribosomal RNA **RNA**→ ←GyrA (COG0188)DNA_topoisoIV (pfam00521)DNA_gyraseA_C (pfam03989)
Bli-1-34	NC_006322.1	-	9706	9640	←16S ribosomal RNA **RNA**→ ←GyrA (COG0188)DNA_topoisoIV (pfam00521)DNA_gyraseA_C (pfam03989)
Bli-1-35	NC_006270.3	-	611992	611926	←16S ribosomal RNA **RNA**→ ←tRNA-Gly
Bli-1-36	NC_006322.1	-	611798	611732	←16S ribosomal RNA **RNA**→ ←tRNA-Gly
Bli-1-37	NC_006270.3	-	34600	34534	←16S ribosomal RNA **RNA**→ ←BofA (pfam07441)
Bli-1-38	NC_006322.1	-	34404	34338	←16S ribosomal RNA **RNA**→ ←BofA (pfam07441)NOG18015 (NOG18015)
Bli-1-39	NC_006270.3	+	925587	925653	23S ribosomal RNA→ **RNA**→ 5S ribosomal RNA→
Bli-1-40	NC_006322.1	+	925438	925504	23S ribosomal RNA→ **RNA**→ 5S ribosomal RNA→
Bsp-1-1	NZ_AAOX01000001.1	-	32371	32305	←COG1971 (COG1971) **RNA**→ ←SUA5 (COG0009)SUA5 (pfam03481)
Bsp-1-2	NZ_AAOX01000004.1	+	108232	108299	←L-lactate dehydrogenase **RNA**→ ←hypo
Bsp-1-3	NZ_AAOX01000018.1	-	15801	15737	hypo→ **RNA**→ ←TrkG (COG0168)

Bsp-1-4	NZ_AAOX01000026.1	+	47929	47995	AmpS (COG2309)→ **RNA**→ ←MurC (COG0773)
Bsp-1-5	NZ_AAOX01000016.1	-	93298	93232	group-specific protein→ **RNA**→ **RNA**→ ←IDI-2_FMN (cd02811)
Bsp-1-35	NZ_AAOX01000016.1	-	93112	93046	(shown above)
Bsp-1-6	NZ_AAOX01000033.1	-	19237	19171	←GlyRS_alpha_core (cd00733) **RNA**→ ←RecO (COG1381)
Bsp-1-7	NZ_AAOX01000015.1	-	33444	33377	←COG5401 (COG5401) **RNA**→ ←RecQ (COG0514)
Bsp-1-8	NZ_AAOX01000019.1	-	82190	82124	←Ala_racemase (cd00430) **RNA**→ ←LolA (COG2834)
Bsp-1-9	NZ_AAOX01000026.1	+	29664	29729	←TyrS (COG0162) **RNA**→ MrcB (COG0744)→
Bsp-1-10	NZ_AAOX01000011.1	-	44591	44525	BetT (COG1292)→ **RNA**→ ←PaaX (COG3327)
Bsp-1-11	NZ_AAOX01000014.1	-	110839	110773	←hypo **RNA**→ ←hypo
Bsp-1-12	NZ_AAOX01000014.1	-	110652	110586	←hypo **RNA**→ ←hypo
Bsp-1-13	NZ_AAOX01000041.1	-	673	607	←hypo **RNA**→ ←hypo
Bsp-1-14	NZ_AAOX01000017.1	+	99694	99760	AcoB (COG0022)→ **RNA**→ ←dihydrolipoamide acetyltransferase
Bsp-1-15	NZ_AAOX01000006.1	-	139439	139373	SplA→ **RNA**→ ←hypo
Bsp-1-16	NZ_AAOX01000027.1	-	25525	25460	hypo→ **RNA**→ ←MutY (COG1194)
Bsp-1-17	NZ_AAOX01000017.1	+	72416	72482	stage III sporulation protein AH→ **RNA**→ ←MA (smart00283)
Bsp-1-18	NZ_AAOX01000009.1	-	76759	76693	DUF520 (pfam04461)→ **RNA**→ ←RimL (COG1670)
Bsp-1-19	NZ_AAOX01000001.1	-	33520	33454	←Wzb (COG0394) **RNA**→ ←hypo
Bsp-1-20	NZ_AAOX01000070.1	+	3409	3475	←BglG (COG3711) **RNA**→ ←COG2252 (COG2252)
Bsp-1-21	NZ_AAOX01000004.1	+	227007	227073	hypo→ **RNA**→ ←VanW (pfam04294)
Bsp-1-22	NZ_AAOX01000016.1	-	19511	19445	←hypo **RNA**→ ←COG4918 (COG4918)
Bsp-1-23	NZ_AAOX01000050.1	-	9503	9437	←COG3881 (COG3881) **RNA**→ ←hypo
Bsp-1-24	NZ_AAOX01000022.1	-	85956	85890	COG1988 (COG1988)→ **RNA**→ ←hypo
Bsp-1-25	NZ_AAOX01000020.1	-	52596	52530	←COG0742 (COG0742) **RNA**→ ←hypo
Bsp-1-26	NZ_AAOX01000014.1	+	92801	92867	RHOD_Lact_B (cd01523)GloB (COG0491)→ **RNA**→ DUF81 (pfam01925)→
Bsp-1-27	NZ_AAOX01000032.1	-	33782	33716	hypo→ **RNA**→ ←COG1511 (COG1511)
Bsp-1-28	NZ_AAOX01000004.1	+	240854	240920	←Lig (COG0272) **RNA**→ ←UvrD (COG0210)
Bsp-1-29	NZ_AAOX01000020.1	-	14455	14389	LdcC (COG1982)→ **RNA**→ hypo→
Bsp-1-30	NZ_AAOX01000016.1	-	10629	10563	OYE_YqiM_FMN (cd02932)→ **RNA**→ ←hypo
Bsp-1-31	NZ_AAOX01000009.1	-	145067	145001	←hypo **RNA**→ ←hypo
Bsp-1-32	NZ_AAOX01000009.1	-	75104	75038	FabG (COG1028)→ **RNA**→ ←PurR (COG1609)
Bsp-1-33	NZ_AAOX01000004.1	+	234474	234540	←PutP (COG0591) **RNA**→ ←Aldedh (pfam00171)
Bsp-1-34	NZ_AAOX01000048.1	+	12	78	**RNA**→ ←Ribosomal_S18 (pfam01084)
Bli-1-41	NC_006270.3	+	58035	58101	SASP (pfam00269)→ **RNA**→ IspE (COG1947)→
Bli-1-42	NC_006322.1	+	57839	57905	SASP (pfam00269)NOG16872 (NOG16872)→ **RNA**→ IspE (COG1947)→
Bli-1-43	NC_006270.3	+	176604	176667	tRNA-Glu→ **RNA**→ tRNA-Val→
Bli-1-44	NC_006322.1	+	176409	176472	tRNA-Glu→ **RNA**→ tRNA-Val→
Bli-1-45	NC_006270.3	+	926189	926254	tRNA-Glu→ **RNA**→ tRNA-Val→
Bli-1-46	NC_006322.1	+	926040	926105	tRNA-Glu→ **RNA**→ tRNA-Val→
Bco-1-1	NZ_ABFU01000013.2	+	842	907	16S ribosomal RNA→ **RNA**→ 23S ribosomal RNA→
Bco-1-2	NZ_ABFU01000131.1	+	643	708	**RNA**→ tRNA-Ile→
Bco-1-3	NZ_ABFU01000134.1	-	1560	1495	16S ribosomal RNA→ **RNA**→ tRNA-Ile→
Bco-1-4	NZ_ABFU01000133.1	+	2038	2104	tRNA-Gly→ **RNA**→ 16S ribosomal RNA→
Bco-1-5	NZ_ABFU01000135.1	+	2040	2106	tRNA-Ala→ **RNA**→ 16S ribosomal RNA→
Bco-1-6	NZ_ABFU01000002.2	-	740	674	hypo→ **RNA**→ 16S ribosomal RNA→

9.3 Conserved domains

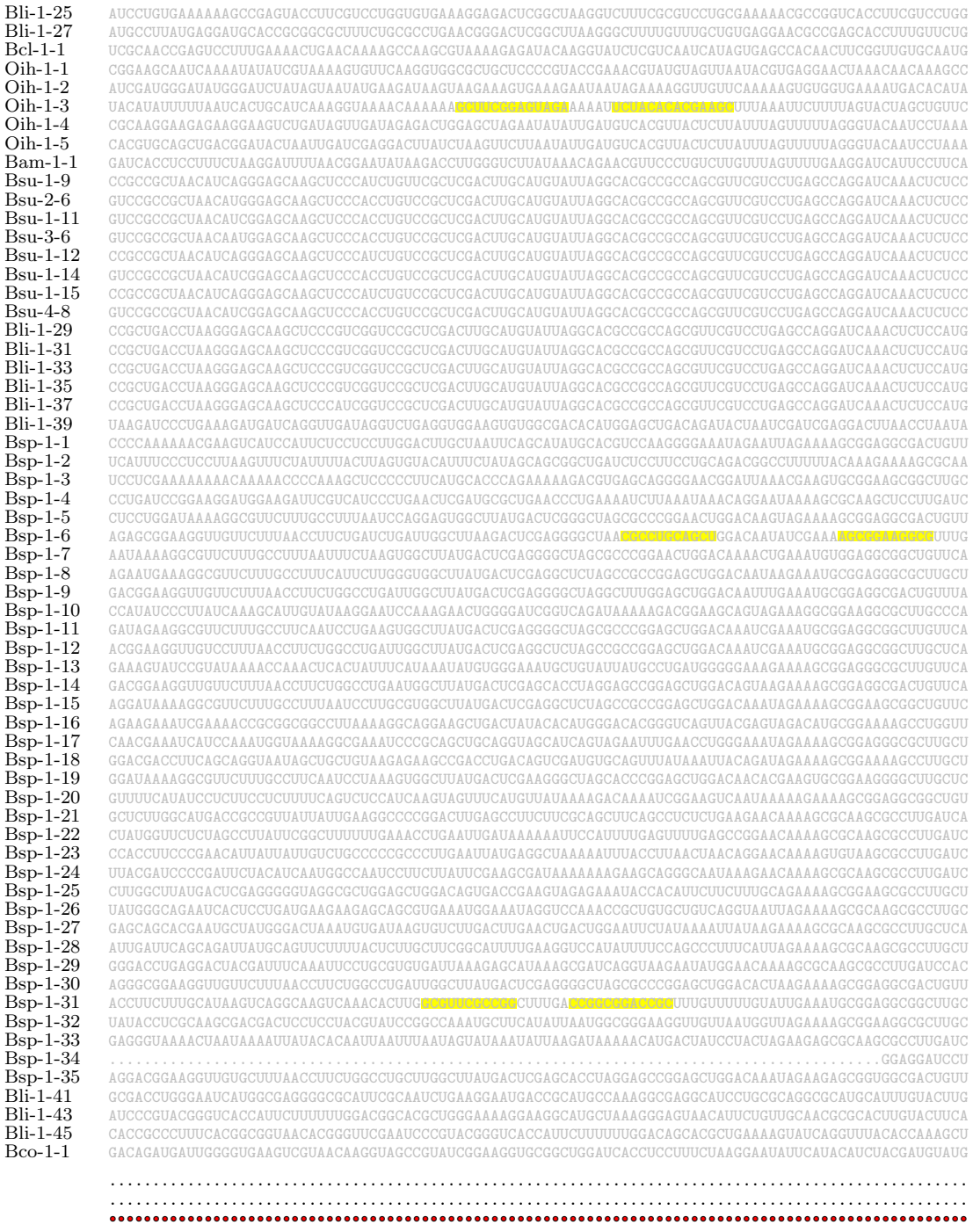
Conserved domains found in protein-coding genes listed in Section 9.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Bacillaceae-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

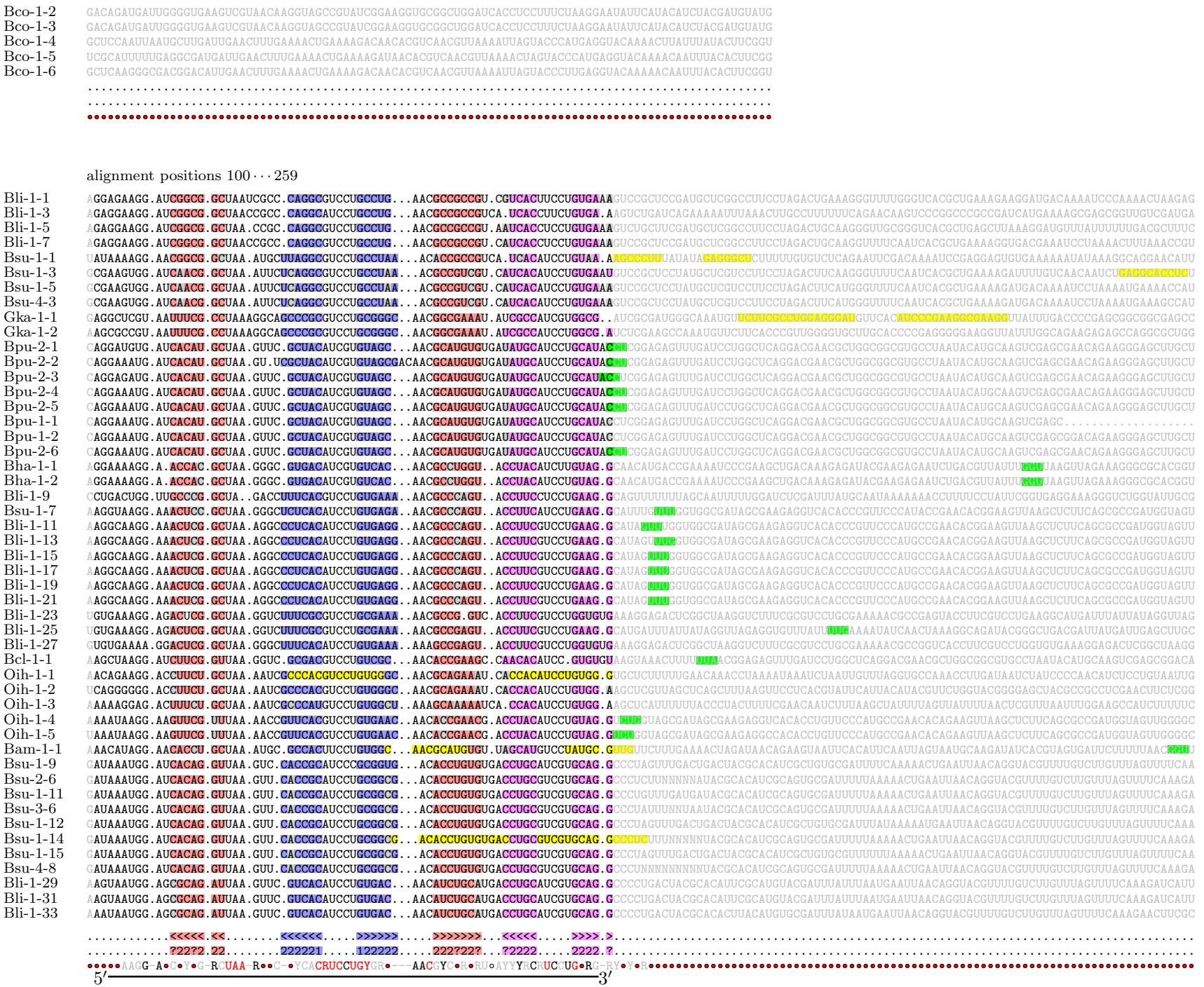
u cd00430 (1) Alanine racemase.
D cd00484 (2) Phosphoenolpyruvate carboxykinase (PEPCK), a critical gluconeogenic enzyme, catalyzes the first committed step in the diversion of tricarboxylic acid cycle intermediates toward gluconeogenesis.
u cd00733 (1) Class II Glycyl-tRNA synthetase (GlyRS) alpha subunit core catalytic domain.
u cd01523 (1) Member of the Rhodanese Homology Domain superfamily.
d cd02811 (1) Isopentenyl Diphosphate:dimethylallyl diphosphate isomerase type 2 (IDI-2) FMN-binding domain.
u cd02932 (1) Old yellow enzyme (OYE) YqjM-like FMN binding domain.
d COG0009 (1) Putative translation factor (SUA5) [Translation, ribosomal structure and biogenesis]
u COG0022 (1) Pyruvate/2-oxoglutarate dehydrogenase complex, dehydrogenase (E1) component, eukaryotic type, beta subunit [Energy production and conversion]
u COG0162 (1) Tyrosyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
d COG0168 (1) Trk-type K+ transport systems, membrane components [Inorganic ion transport and metabolism]
u d COG0188 (4) Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), A subunit [DNA replication, recombination, and repair]
u COG0192 (2) S-adenosylmethionine synthetase [Coenzyme metabolism]
d COG0210 (1) Superfamily I DNA and RNA helicases [DNA replication, recombination, and repair]
u COG0272 (1) NAD-dependent DNA ligase (contains BRCT domain type II) [DNA replication, recombination, and repair]
u COG0394 (1) Protein-tyrosine-phosphatase [Signal transduction mechanisms]
u COG0491 (1) Zn-dependent hydrolases, including glyoxylases [General function prediction only]
d COG0514 (1) Superfamily II DNA helicase [DNA replication, recombination, and repair]
u COG0591 (1) Na+/proline symporter [Amino acid transport and metabolism / General function prediction only]
u COG0742 (1) N6-adenine-specific methylase [DNA replication, recombination, and repair]
D COG0744 (1) Membrane carboxypeptidase (penicillin-binding protein) [Cell envelope biogenesis, outer membrane]
d COG0773 (1) UDP-N-acetylmuramate-alanine ligase [Cell envelope biogenesis, outer membrane]
u COG1028 (1) Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases) [Secondary metabolites biosynthesis, transport, and catabolism / General function prediction only]
d COG1194 (1) A/G-specific DNA glycosylase [DNA replication, recombination, and repair]
d COG1252 (2) NADH dehydrogenase, FAD-containing subunit [Energy production and conversion]
d COG1253 (1) Hemolysins and related proteins containing CBS domains [General function prediction only]
u COG1292 (1) Choline-glycine betaine transporter [Cell envelope biogenesis, outer membrane]
d COG1307 (2) Uncharacterized protein conserved in bacteria [Function unknown]
d COG1381 (1) Recombinational DNA repair protein (RecF pathway) [DNA replication, recombination, and repair]
d COG1511 (1) Predicted membrane protein [Function unknown]
d COG1609 (3) Transcriptional regulators [Transcription]
d COG1670 (1) Acetyltransferases, including N-acetylases of ribosomal proteins [Translation, ribosomal structure and biogenesis]
D COG1866 (2) Phosphoenolpyruvate carboxykinase (ATP) [Energy production and conversion]
D COG1947 (2) 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate synthase [Lipid metabolism]
D COG1959 (2) Predicted transcriptional regulator [Transcription]
u COG1971 (1) Predicted membrane protein [Function unknown]
u COG1982 (1) Arginine/lysine/ornithine decarboxylases [Amino acid transport and metabolism]
u COG1988 (1) Predicted membrane-bound metal-dependent hydrolases [General function prediction only]
u COG2124 (2) Cytochrome P450 [Secondary metabolites biosynthesis, transport, and catabolism]
d COG2252 (1) Permeases [General function prediction only]
u COG2256 (2) ATPase related to the helicase subunit of the Holliday junction resolvase [DNA replication, recombination, and repair]
u COG2309 (1) Leucyl aminopeptidase (aminopeptidase T) [Amino acid transport and metabolism]
d COG2814 (1) Arabinose efflux permease [Carbohydrate transport and metabolism]
d COG2834 (1) Outer membrane lipoprotein-sorting protein [Cell envelope biogenesis, outer membrane]
d COG3327 (1) Phenylacetic acid-responsive transcriptional repressor [Transcription]
u COG3333 (1) Uncharacterized protein conserved in bacteria [Function unknown]
u COG3428 (1) Predicted membrane protein [Function unknown]
u COG3711 (1) Transcriptional antiterminator [Transcription]
uD COG3859 (3) Predicted membrane protein [Function unknown]
u COG3881 (1) Uncharacterized protein conserved in bacteria [Function unknown]
d COG4918 (1) Uncharacterized protein conserved in bacteria [Function unknown]
u COG5401 (1) Spore germination protein [General function prediction only]
u NOG16872 (1)
d NOG18015 (1)
u pfam00083 (1) Sugar (and other) transporter.
d pfam00171 (1) Aldehyde dehydrogenase family.
u pfam00269 (2) Small, acid-soluble spore proteins, alpha/beta type.
d pfam00521 (3) DNA gyrase/topoisomerase IV, subunit A.
D pfam01019 (1) Gamma-glutamyltranspeptidase.
d pfam01084 (1) Ribosomal protein S18.
u pfam01522 (1) Polysaccharide deacetylase.
d pfam01694 (1) Rhomboid family.
d pfam01904 (1) Protein of unknown function DUF72.
D pfam01925 (1) Domain of unknown function DUF81.
D pfam02634 (1) FdhD/NarQ family.
u pfam03473 (2) MOSC domain.
u pfam03475 (2) 3-alpha domain.
d pfam03481 (1) SUA5 domain.
D pfam03861 (1) ANTAR domain.
d pfam03989 (3) DNA gyrase C-terminal domain, beta-propeller.
d pfam04294 (1) VanW like protein.
u pfam04461 (1) Protein of unknown function (DUF520).
u d pfam07441 (3) SigmaK-factor processing regulatory protein BofA.
d smart00283 (1) Methyl-accepting chemotaxis-like domains (chemotaxis sensory transducer).

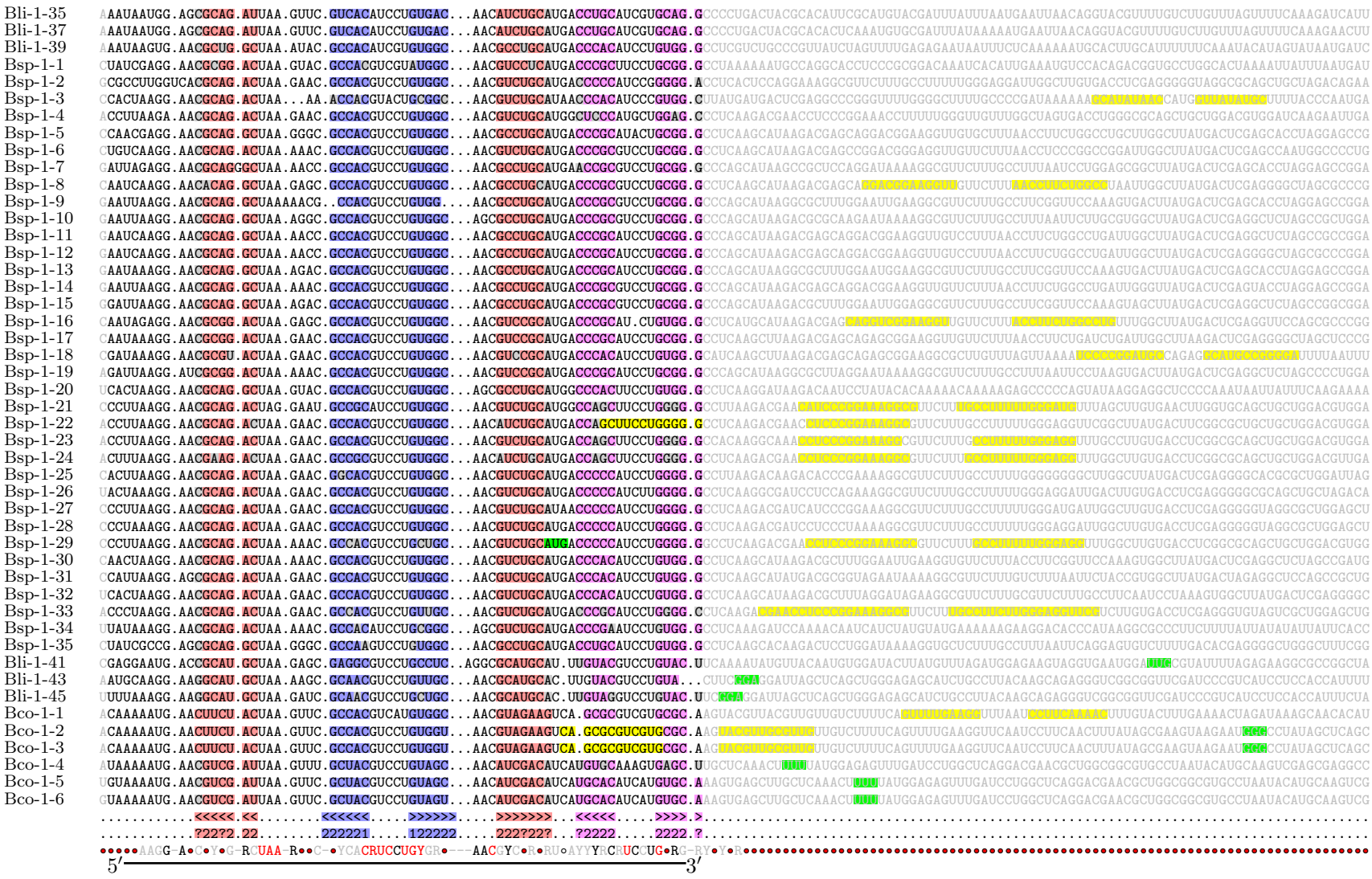
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Bam-1-2, Bha-1-3, Bha-1-4, Bli-1-10, Bli-1-12, Bli-1-14, Bli-1-16, Bli-1-18, Bli-1-2, Bli-1-20, Bli-1-22, Bli-1-24, Bli-1-26, Bli-1-28, Bli-1-30, Bli-1-32, Bli-1-34, Bli-1-36, Bli-1-38, Bli-1-4, Bli-1-40, Bli-1-42, Bli-1-44, Bli-1-46, Bli-1-6, Bli-1-8, Bsu-1-10, Bsu-1-13, Bsu-1-16, Bsu-1-2, Bsu-1-4, Bsu-1-6, Bsu-1-8, Bsu-2-1, Bsu-2-2, Bsu-2-3, Bsu-2-4, Bsu-2-5, Bsu-2-7, Bsu-2-8, Bsu-3-1, Bsu-3-2, Bsu-3-3, Bsu-3-4, Bsu-3-5, Bsu-3-7, Bsu-3-8, Bsu-4-1, Bsu-4-2, Bsu-4-4, Bsu-4-5, Bsu-4-6, Bsu-4-7

9.4 Multiple-sequence alignment

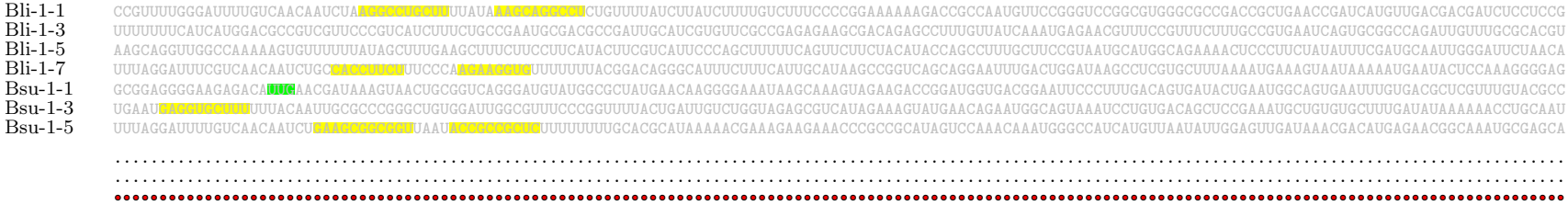
Each Bacillaceae-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 9.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Bacillaceae-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

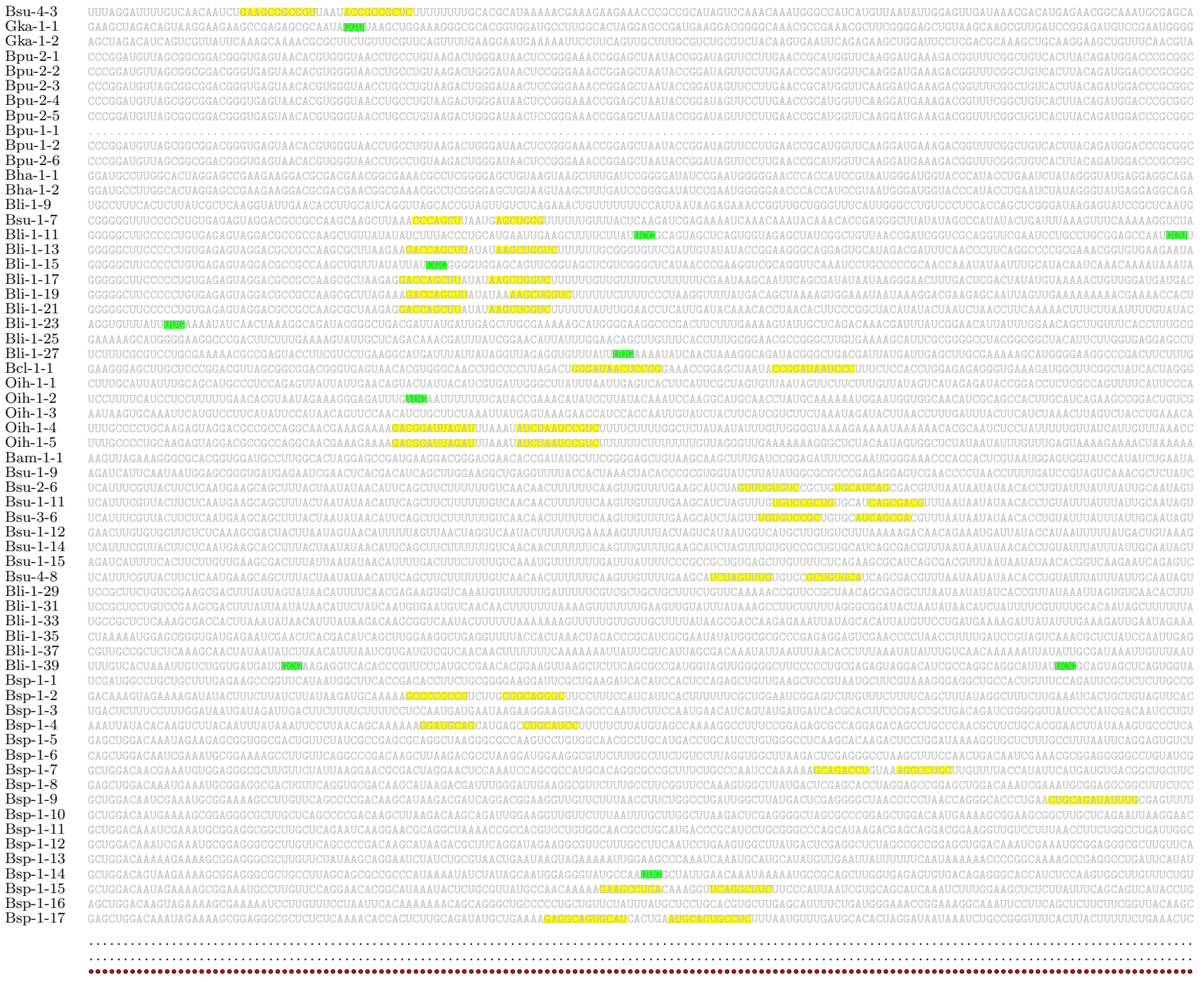

Bli-1-25
Bli-1-27
Bcl-1-1
Oih-1-1
Oih-1-2
Oih-1-3
Oih-1-4
Oih-1-5
Bam-1-1
Bsu-1-9
Bsu-2-6
Bsu-1-11
Bsu-3-6
Bsu-1-12
Bsu-1-14
Bsu-1-15
Bsu-4-8
Bli-1-29
Bli-1-31
Bli-1-33
Bli-1-35
Bli-1-37
Bli-1-39
Bsp-1-1
Bsp-1-2
Bsp-1-3
Bsp-1-4
Bsp-1-5
Bsp-1-6
Bsp-1-7
Bsp-1-8
Bsp-1-9
Bsp-1-10
Bsp-1-11
Bsp-1-12
Bsp-1-13
Bsp-1-14
Bsp-1-15
Bsp-1-16
Bsp-1-17
Bsp-1-18
Bsp-1-19
Bsp-1-20
Bsp-1-21
Bsp-1-22
Bsp-1-23
Bsp-1-24
Bsp-1-25
Bsp-1-26
Bsp-1-27
Bsp-1-28
Bsp-1-29
Bsp-1-30
Bsp-1-31
Bsp-1-32
Bsp-1-33
Bsp-1-34
Bsp-1-35
Bli-1-41
Bli-1-43
Bli-1-45
Bco-1-1


Bco-1-2
Bco-1-3
Bco-1-4
Bco-1-5
Bco-1-6
alignment positions 100 · · · 259
Bli-1-1
Bli-1-3
Bli-1-5
Bli-1-7
Bsu-1-1
Bsu-1-3
Bsu-1-5
Bsu-4-3
Gka-1-1
Gka-1-2
Bpu-2-1
Bpu-2-2
Bpu-2-3
Bpu-2-4
Bpu-2-5
Bpu-1-1
Bpu-1-2
Bpu-2-6
Bha-1-1
Bha-1-2
Bli-1-9
Bsu-1-7
Bli-1-11
Bli-1-13
Bli-1-15
Bli-1-17
Bli-1-19
Bli-1-21
Bli-1-23
Bli-1-25
Bli-1-27
Bcl-1-1
Oih-1-1
Oih-1-2
Oih-1-3
Oih-1-4
Oih-1-5
Bam-1-1
Bsu-1-9
Bsu-2-6
Bsu-1-11
Bsu-3-6
Bsu-1-12
Bsu-1-14
Bsu-1-15
Bsu-4-8
Bli-1-29
Bli-1-31
Bli-1-33
5′
3′

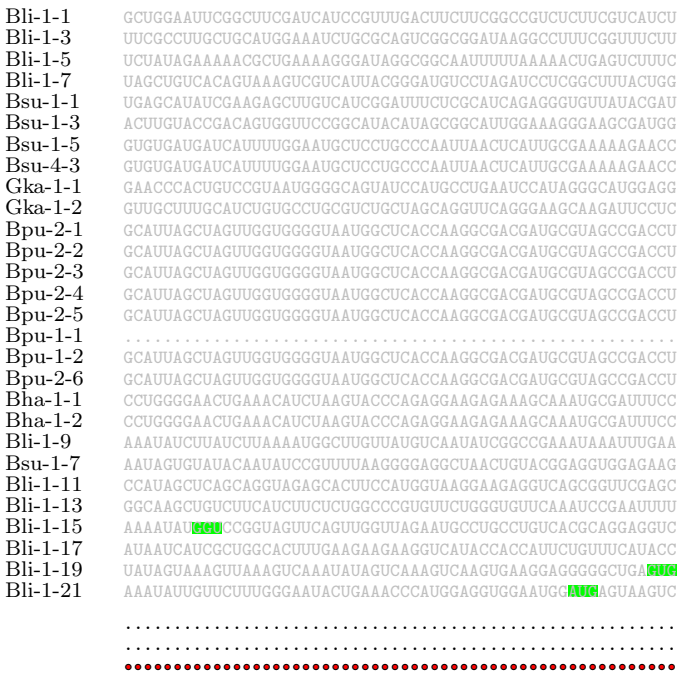


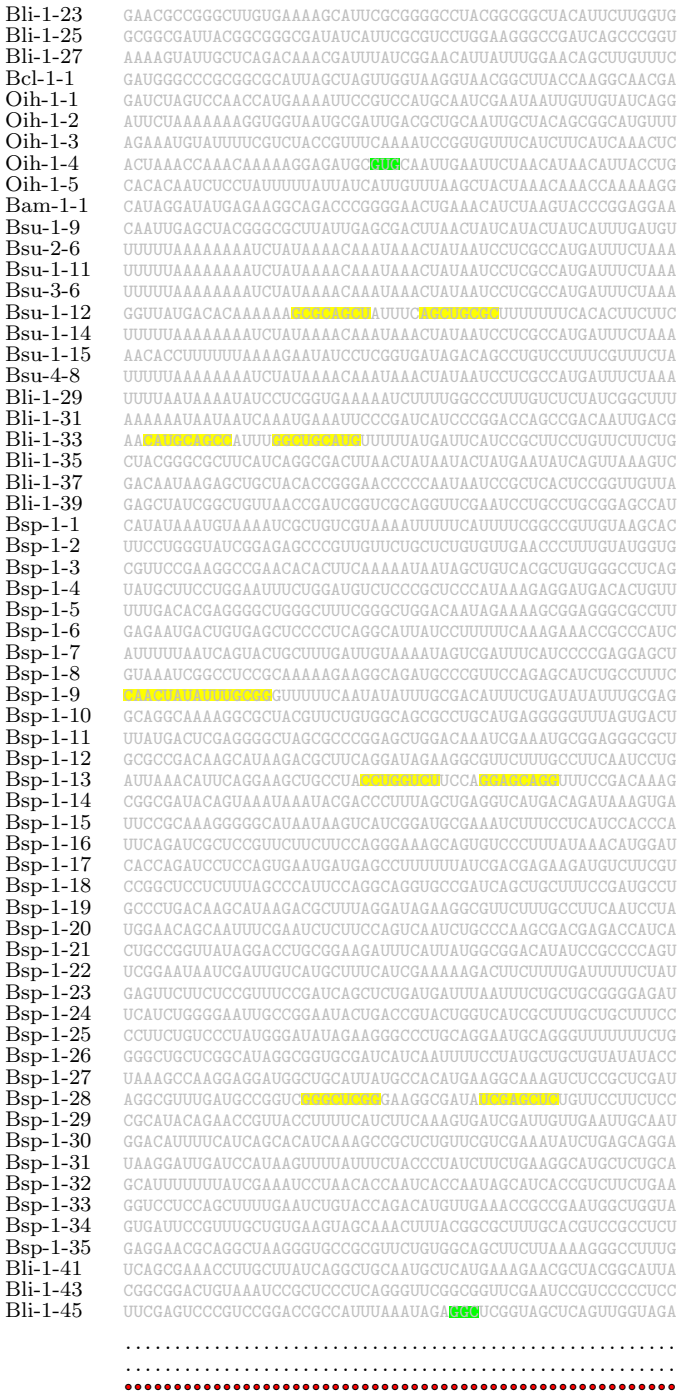
alignment positions 260 · · · 419

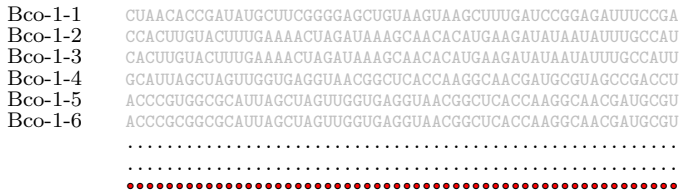



Bsu-4-3
Gka-1-1
Gka-1-2
Bpu-2-1
Bpu-2-2
Bpu-2-3
Bpu-2-4
Bpu-2-5
Bpu-1-1
Bpu-1-2
Bpu-2-6
Bha-1-1
Bha-1-2
Bli-1-9
Bsu-1-7
Bli-1-11
Bli-1-13
Bli-1-15
Bli-1-17
Bli-1-19
Bli-1-21
Bli-1-23
Bli-1-25
Bli-1-27
Bcl-1-1
Oih-1-1
Oih-1-2
Oih-1-3
Oih-1-4
Oih-1-5
Bam-1-1
Bsu-1-9
Bsu-2-6
Bsu-1-11
Bsu-3-6
Bsu-1-12
Bsu-1-14
Bsu-1-15
Bsu-4-8
Bli-1-29
Bli-1-31
Bli-1-33
Bli-1-35
Bli-1-37
Bli-1-39
Bsp-1-1
Bsp-1-2
Bsp-1-3
Bsp-1-4
Bsp-1-5
Bsp-1-6
Bsp-1-7
Bsp-1-8
Bsp-1-9
Bsp-1-10
Bsp-1-11
Bsp-1-12
Bsp-1-13
Bsp-1-14
Bsp-1-15
Bsp-1-16
Bsp-1-17

alignment positions $420 \cdots 475$




Bli-1-23
Bli-1-25
Bli-1-27
Bcl-1-1
Oih-1-1
Oih-1-2
Oih-1-3
Oih-1-4
Oih-1-5
Bam-1-1
Bsu-1-9
Bsu-2-6
Bsu-1-11
Bsu-3-6
Bsu-1-12
Bsu-1-14
Bsu-1-15
Bsu-4-8
Bli-1-29
Bli-1-31
Bli-1-33
Bli-1-35
Bli-1-37
Bli-1-39
Bsp-1-1
Bsp-1-2
Bsp-1-3
Bsp-1-4
Bsp-1-5
Bsp-1-6
Bsp-1-7
Bsp-1-8
Bsp-1-9
Bsp-1-10
Bsp-1-11
Bsp-1-12
Bsp-1-13
Bsp-1-14
Bsp-1-15
Bsp-1-16
Bsp-1-17
Bsp-1-18
Bsp-1-19
Bsp-1-20
Bsp-1-21
Bsp-1-22
Bsp-1-23
Bsp-1-24
Bsp-1-25
Bsp-1-26
Bsp-1-27
Bsp-1-28
Bsp-1-29
Bsp-1-30
Bsp-1-31
Bsp-1-32
Bsp-1-33
Bsp-1-34
Bsp-1-35
Bli-1-41
Bli-1-43
Bli-1-45



10 *Bacillus*-plasmid

10.1 Taxa

The taxonomy of each organism containing a putative *Bacillus*-plasmid RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *Bacillus*-plasmid RNA in Sections 10.2 and 10.4.

abbrev. of hits	**taxonomy of species**
Bpu-1-1 to Bpu-1-2	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus pumilus*
Bsu-1-1 to Bsu-1-6	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus subtilis*
Bsu-2-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus subtilis subsp. subtilis str.* 168
Bce-1-1 to Bce-1-2	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* 03BB108
Bce-2-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* E33L
Bce-4-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus subsp. cytotoxis* NVH 391-98
Bce-3-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* W
Bmy-1-1 to Bmy-1-4	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus mycoides*
Bth-1-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus thuringiensis*
Bth-2-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus thuringiensis serovar israelensis*
Bth-3-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus thuringiensis serovar israelensis* ATCC 35646
Lbr-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus brevis* ATCC 367
Lca-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus casei* ATCC 334
Lcu-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus curvatus*
Lhe-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus helveticus subsp. jugurti*
Lpl-1-1 to Lpl-1-3	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus plantarum*
Lsa-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus sakei*
Pac-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Pediococcus acidilactici*
env-1 to env-2	environmental samples

10.2 Gene contexts

Each *Bacillus*-plasmid RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *Bacillus*-plasmid RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 10.3. The accession number of the sequence containing each *Bacillus*-plasmid RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 10.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Lpl-1-1	NC_003894.1	-	3083	3036	hypo→ **RNA**→ ←hypo
Lpl-1-2	NC_004944.1	-	2644	2584	hypo→ **RNA**→ Rep_3 (pfam01051)→
Lbr-1-1	NC_008499.1	-	332	272	hypo→ **RNA**→
Pac-1-1	NC_004832.1	+	264	324	**RNA**→ Rep_3 (pfam01051)→
Lcu-1-1	NC_003320.1	+	3074	3134	hypo→ **RNA**→ Rep_3 (pfam01051)→
Lsa-1-1	NC_004942.1	+	149	209	**RNA**→ Rep_3 (pfam01051)→
Lca-1-1	NC_008526.1	-	360213	360153	hypo→ **RNA**→ UPF0236 (pfam06782)→
env-1	BABE01000260.1	+	3320	3379	Mob_Pre (pfam01076)→ **RNA**→ Rep_trans (pfam02486)→
env-2	BABE01002635.1	+	459	518	hypo→ **RNA**→
Lpl-1-3	NC_002123.1	+	1928	1989	Rep_1 (pfam01446)→ **RNA**→
Bce-1-1	NZ_ABDM01000007.1	+	167738	167800	**RNA**→
Bmy-1-1	NC_006870.1	+	3291	3366	Peptidase_S26 (pfam00461)→ **RNA**→ ←hypo
Bce-1-2	NZ_ABDM01000060.1	+	278	352	**RNA**→
Bmy-1-2	NC_006869.1	+	120	192	**RNA**→ Mob_Pre (pfam01076)→
Bth-1-1	NC_001446.1	+	6441	6496	hypo→ **RNA**→ Mob_Pre (pfam01076)→
Bth-3-1	NZ_AAJM01000209.1	-	3338	3283	←hypo **RNA**→ Mob_Pre (pfam01076)→
Bsu-1-1	NC_001764.1	+	5114	5169	←HTH_YyaN (cd01109) **RNA**→ Rap40→
Bpu-1-1	NC_004932.1	+	376	432	**RNA**→ TPR (cd00189)→
Bpu-1-2	NC_001858.1	+	104	159	**RNA**→ TPR (cd00189)→
Bsu-1-2	NC_002075.1	+	6989	7044	hypo→ **RNA**→ hypo→
Bsu-1-3	NC_001765.1	+	5173	5228	←SoxR (COG0789) **RNA**→
Bsu-1-4	NC_006825.1	+	5969	6024	←SoxR (COG0789) **RNA**→
Bsu-1-5	NC_007956.1	+	5968	6023	←SoxR (COG0789) **RNA**→
Bsu-1-6	NC_001766.1	+	6354	6409	hypo→ **RNA**→ Rap60→
Bth-2-1	NC_002091.1	+	5075	5141	collagen-like protein→ **RNA**→
Bsu-2-1	NC_000964.2	+	535441	535494	COG2946 (COG2946)→ **RNA**→ hypo→
Bce-2-1	NC_007104.1	+	1986	2052	hypo→ **RNA**→ hypo→
Bce-4-1	NC_009674.1	-	2388712	2388646	←Tra8 (COG2826) **RNA**→ Mob_Pre (pfam01076)→
Bmy-1-3	NC_005703.1	+	594	660	**RNA**→ Mob_Pre (pfam01076)→
Bmy-1-4	NC_005704.1	+	3044	3110	←putative Rep protein **RNA**→
Bce-3-1	NZ_ABCZ01000046.1	-	1174	1108	**RNA**→
Lhe-1-1	NC_001379.1	+	65	124	**RNA**→ RepA_N (pfam06970)→

10.3 Conserved domains

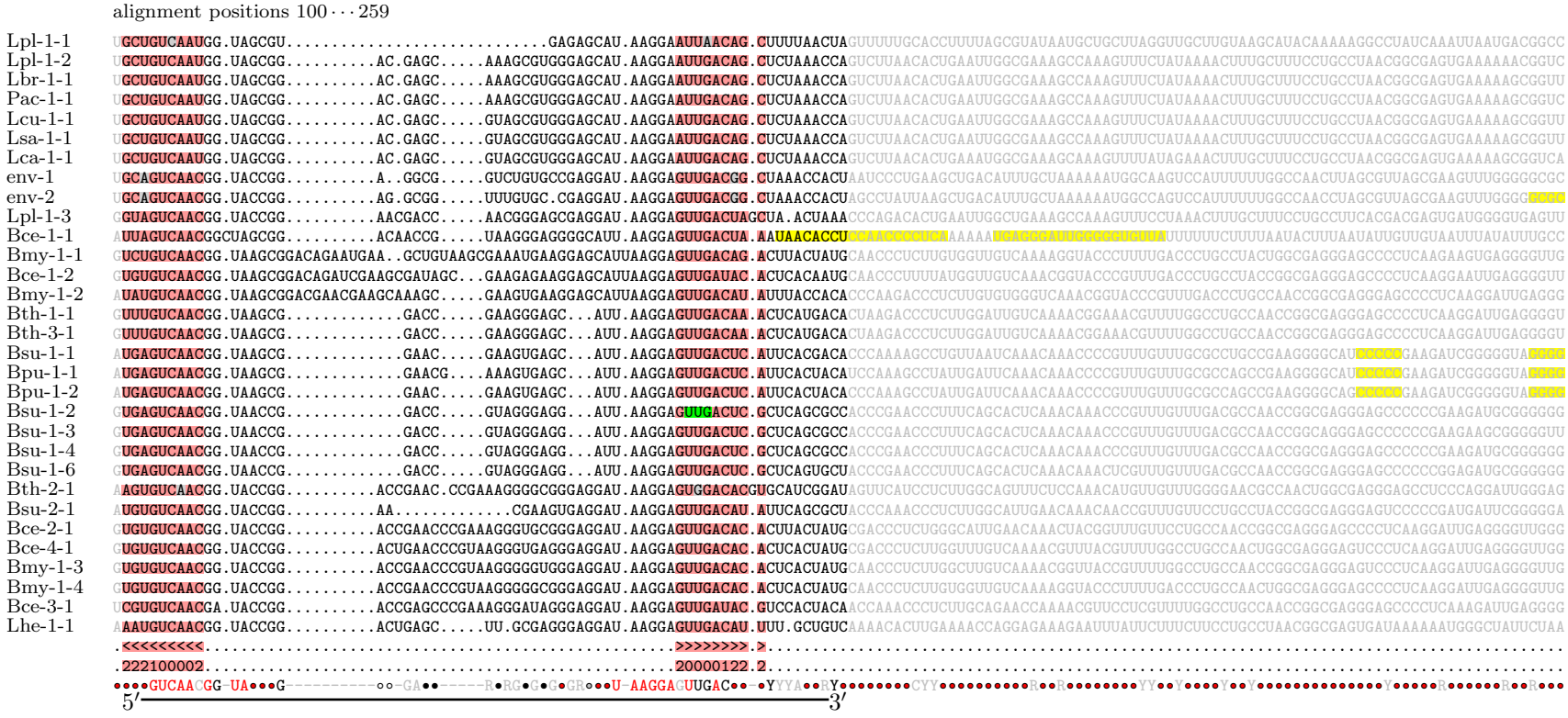
Conserved domains found in protein-coding genes listed in Section 10.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *Bacillus*-plasmid RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

D cd00189 (2) Tetratricopeptide repeat domain; typically contains 34 amino acids [WLF]-X(2)-[LIM]-[GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-X(2)-[ASL]-X(4)-[PKE]
u cd01109 (1) Helix-turn-helix transcription regulator YyaN.
u COG0789 (3) Predicted transcriptional regulators [Transcription]
u COG2826 (1) Transposase and inactivated derivatives, IS30 family [DNA replication, recombination, and repair]
u COG2946 (1) Putative phage replication protein RstA [DNA replication, recombination, and repair]
u pfam00461 (1) Signal peptidase I.
D pfam01051 (4) Initiator Replication protein.
uD pfam01076 (6) Plasmid recombination enzyme.
u pfam01446 (1) Replication protein.
D pfam02486 (1) Replication initiation factor.
D pfam06782 (1) Uncharacterised protein family (UPF0236).
D pfam06970 (1) Replication initiator protein A (RepA) N-terminus.

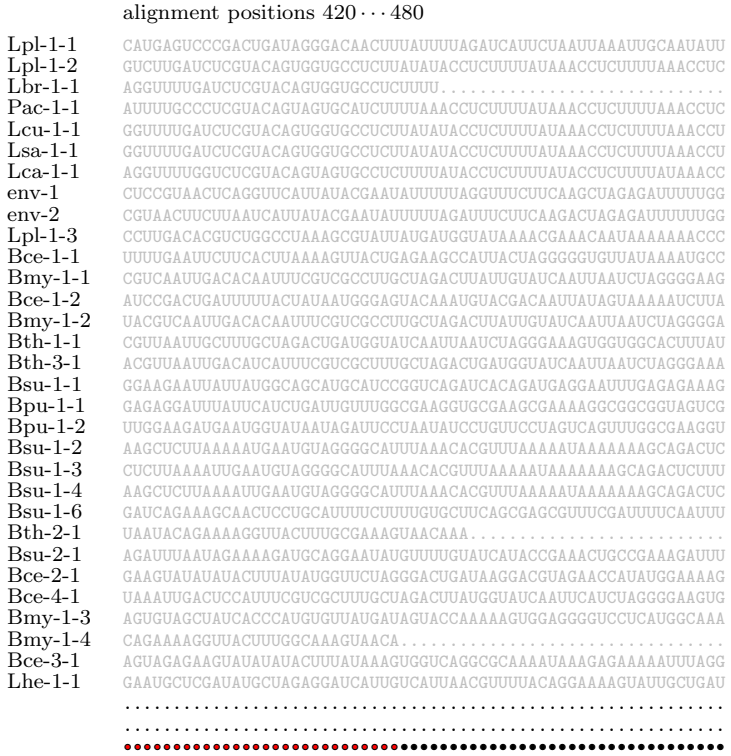
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Bsu-1-5

10.4 Multiple-sequence alignment

Each *Bacillus*-plasmid RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 10.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *Bacillus*-plasmid RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.


alignment positions 100 · · · 259
Lpl-1-1
Lpl-1-2
Lbr-1-1
Pac-1-1
Lcu-1-1
Lsa-1-1
Lca-1-1
env-1
env-2
Lpl-1-3
Bce-1-1
Bmy-1-1
Bce-1-2
Bmy-1-2
Bth-1-1
Bth-3-1
Bsu-1-1
Bpu-1-1
Bpu-1-2
Bsu-1-2
Bsu-1-3
Bsu-1-4
Bsu-1-6
Bth-2-1
Bsu-2-1
Bce-2-1
Bce-4-1
Bmy-1-3
Bmy-1-4
Bce-3-1
Lhe-1-1
5′
3′

alignment positions 260 · · · 419



11 Bacteroid-*trp*

11.1 Taxa

The taxonomy of each organism containing a putative Bacteroid-*trp* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Bacteroid-*trp* RNA in Sections 11.2 and 11.4.

abbrev. of hits	taxonomy of species
Bca-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides caccae* ATCC 43185
Bfr-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides fragilis* NCTC 9343
Bfr-2-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides fragilis* YCH46
Bov-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides ovatus* ATCC 8483
Bth-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides thetaiotaomicron* VPI-5482
Bun-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides uniformis* ATCC 8492
Bvu-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides vulgatus* ATCC 8482
Pdi-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Porphyromonadaceae *Parabacteroides distasonis* ATCC 8503
Pme-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Porphyromonadaceae *Parabacteroides merdae* ATCC 43184
ueu-1-1	Bacteria Bacteroidetes environmental samples unidentified eubacterium SCB49
Fba-1-1	Bacteria Bacteroidetes *Flavobacteria bacterium* BBFL7
Fba-2-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales *Flavobacteria bacterium* BAL38
Cat-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Croceibacter atlanticus* HTCC2559
Fjo-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Flavobacterium johnsoniae* UW101

Fps-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Flavobacterium psychrophilum* JIP02/86
Gfo-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Gramella forsetii* KT0803
Lbl-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Leeuwenhoekiella blandensis* MED217
Pir-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Polaribacter irgensii* 23-P
Psp-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Polaribacter sp.* MED152
Pto-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Psychroflexus torquis* ATCC 700755
Rbi-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Robiginitalea biformata* HTCC2501
Fba-3-1	Bacteria Bacteroidetes Flavobacteria *Flavobacteriales bacterium* HTCC2170
Asp-1-1	Bacteria Bacteroidetes Sphingobacteria Sphingobacteriales Cyclobacteriaceae *Algoriphagus sp.* PR1
Chu-1-1	Bacteria Bacteroidetes Sphingobacteria Sphingobacteriales Flexibacteraceae *Cytophaga hutchinsonii* ATCC 33406
Mma-1-1	Bacteria Bacteroidetes Sphingobacteria Sphingobacteriales Flexibacteraceae *Microscilla marina* ATCC 23134
Psp-2-1	Bacteria Bacteroidetes Sphingobacteria Sphingobacteriales Sphingobacteriaceae *Pedobacter sp.* BAL39
env-1 to env-21	environmental samples

11.2 Gene contexts

Each Bacteroid-*trp* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Bacteroid-*trp* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 11.3. The accession number of the sequence containing each Bacteroid-*trp* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 11.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Psp-2-1	NZ_ABCM01000008.1	+	235022	235164	**RNA**→ TrpE (COG0147)→ PabA (COG0512)→ indole-3-glycerolphosphate synthase→
Asp-1-1	NZ_AAXU01000004.1	+	467000	467130	**RNA**→ TrpE (COG0147)→ Ribosomal_S23p (pfam05635)→ PabA (COG0512)→ TrpD (COG0547)→ IGPS (cd00331)→ TrpF (COG0135)→ TrpB (COG0133)→
Rbi-1-1	NZ_AAOI01000005.1	+	7798	7916	**RNA**→ TrpE (COG0147)→ PabA (COG0512)→ TrpD (COG0547)→ TrpC (COG0134)→ TrpF (COG0135)→ TrpB (COG0133)→ trp_synt_A (cd00346)→
Mma-1-1	NZ_AAWS01000005.1	-	77275	77138	**RNA**→ TrpE (COG0147)→
Cat-1-1	NZ_AAMP01000001.1	-	1429786	1429665	**RNA**→ TrpE (COG0147)→ PabA (COG0512)→ TrpD (COG0547)→ IGPS (cd00331)→ TrpF (COG0135)→ TrpB (COG0133)→
env-1	JCVI_SCAF_1101668064648	+	1028	1145	**RNA**→ ? (GOS681)→
env-2	JCVI_SCAF_1096627072666	+	1417	1534	**RNA**→ TrpE (COG0147)? (GOS681)→
env-3	JCVI_SCAF_1101668438486	-	379	262	**RNA**→
Fba-1-1	NZ_AAPD01000009.1	+	45681	45788	**RNA**→ TrpE (COG0147)→ hypo→ PabA (COG0512)→ TrpD (COG0547)→ IGPS (cd00331)→ TrpF (COG0135)→ TrpB (COG0133)→ TrpA (COG0159)→
Chu-1-1	NC_008255.1	+	3762321	3762447	**RNA**→ TrpE (COG0147)→ PabA (COG0512)→
Pir-1-1	NZ_AAOG01000004.1	-	65950	65830	**RNA**→ TrpE (COG0147)→
Psp-1-1	NZ_AANA01000002.1	+	232794	232908	**RNA**→ TrpE (COG0147)→ PabA (COG0512)→ TrpD (COG0547)→ hypo→ IGPS (cd00331)→ Mug (COG3663)→ TrpF (COG0135)→ TrpB (COG0133)→ TrpA (COG0159)→
Lbl-1-1	NZ_AANC01000011.1	+	62850	62973	**RNA**→ TrpE (COG0147)→ PabA (COG0512)→
env-4	JCVI_SCAF_1096628041415	-	689	564	**RNA**→ Anth_synt_I_N (pfam04715)? (GOS681)→
env-5	BABB01000545.1	-	1746	1604	**RNA**→ TrpB (COG0133)→ hypo→
env-6	BAAU01027801.1	-	741	599	**RNA**→ TrpB (COG0133)→
env-7	BAAY01005429.1	-	1262	1120	**RNA**→ TrpB (COG0133)→ hypo→
Bvu-1-1	NC_009614.1	+	5029321	5029462	**RNA**→ TrpB (COG0133)→ TrpE (COG0147)→ PabA (COG0512)→ TrpD (COG0547)→ TrpC (COG0134)→ TrpF (COG0135)→ trp_synt_A (cd00346)→
env-8	BAAU01000768.1	+	936	1077	**RNA**→ TrpB (COG0133)→ hypo→
env-9	BAAV01013420.1	+	190	331	**RNA**→ TrpB (COG0133)→

env-10	BAAX01007295.1	-	1158	1017	**RNA**→ TrpB (COG0133)→
env-11	BABG01003022.1	+	1510	1651	**RNA**→
env-12	BABB01001584.1	-	2460	2322	**RNA**→ TrpB (COG0133)→ TrpE (COG0147)→
Bun-1-1	NZ_AAYH02000042.1	-	71718	71577	**RNA**→
env-13	BABA01003390.1	+	487	628	**RNA**→ TrpB (COG0133)→ hypo→
env-14	BABG01015768.1	+	769	915	**RNA**→
Bfr-2-1	NC_006347.1	+	3050356	3050502	**RNA**→ TrpB (COG0133)→
env-15	BAAU01006342.1	-	1825	1679	**RNA**→ TrpB (COG0133)→ hypo→
Bfr-1-1	NC_003228.3	+	3113186	3113332	**RNA**→ TrpB (COG0133)→
env-16	BAAZ01029283.1	-	197	51	**RNA**→
Bca-1-1	NZ_AAVM02000001.1	-	509313	509169	**RNA**→ TrpB (COG0133)→ TrpE (COG0147)→ PabA (COG0512)→ TrpD (COG0547)→ IGPS (cd00331)→ TrpF (COG0135)→ TrpA (COG0159)→
Bov-1-1	NZ_AAXF02000052.1	+	120432	120576	**RNA**→
Bth-1-1	NC_004663.1	-	657806	657662	**RNA**→ TrpB (COG0133)→ TrpE (COG0147)→ PabA (COG0512)→ TrpD (COG0547)→ IGPS (cd00331)→ TrpF (COG0135)→ TrpA (COG0159)→
Pme-1-1	NZ_AAXE02000021.1	+	155182	155313	**RNA**→ TrpB (COG0133)→ TrpE (COG0147)→ PabA (COG0512)→ TrpD (COG0547)→ IGPS (cd00331)→ TrpF (COG0135)→ TrpA (COG0159)→
env-17	BABD01002850.1	-	1611	1480	**RNA**→ TrpB (COG0133)→
env-18	BABB01001973.1	-	2743	2604	**RNA**→ TrpB (COG0133)→ TrpE (COG0147)→
Pdi-1-1	NC_009615.1	-	693568	693428	**RNA**→ TrpB (COG0133)→ TrpE (COG0147)→ PabA (COG0512)→
ueu-1-1	NZ_ABCO01000006.1	-	190508	190390	**RNA**→ TrpE (COG0147)→ PabA (COG0512)→ TrpD (COG0547)→ TrpC (COG0134)→ TrpF (COG0135)→ TrpB (COG0133)→ TrpA (COG0159)→
env-19	JCVI_SCAF_1096627859334	+	530	635	**RNA**→ ? (GOS681)→ TrpE (COG0147)? (GOS681)→
Pto-1-1	NZ_AAPR01000002.1	+	281378	281492	**RNA**→ TrpE (COG0147)→ PabA (COG0512)→ TrpD (COG0547)→ IGPS (cd00331)→ TrpF (COG0135)→ TrpB (COG0133)→ TrpA (COG0159)→
Fba-2-1	NZ_AAXX01000005.1	-	34839	34711	**RNA**→ TrpE (COG0147)→
Fjo-1-1	NC_009441.1	+	5912046	5912172	**RNA**→ TrpE (COG0147)→ hypo→ PabA (COG0512)→
Fps-1-1	NC_009613.1	-	621441	621313	**RNA**→ TrpE (COG0147)→
env-20	JCVI_SCAF_1101668030442	+	227	342	**RNA**→ TrpE (COG0147)? (GOS681)→ TrpE (COG0147)? (GOS681)→
Fba-3-1	NZ_AAOC01000005.1	+	43902	44018	**RNA**→ TrpE (COG0147)→ PabA (COG0512)→ TrpD (COG0547)→ TrpC (COG0134)→ TrpF (COG0135)→ TrpB (COG0133)→ TrpA (COG0159)→ hypo→ ArgE (COG0624)→
env-21	wf3_AHAI20157_g2	-	837	716	**RNA**→ TrpE (COG0147)→
Gfo-1-1	NC_008571.1	+	2785164	2785282	**RNA**→ TrpE (COG0147)→ PabA (COG0512)→

11.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 11.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Bacteroid-*trp* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

cd00331 (8) Indole-3-glycerol phosphate synthase (IGPS); an enzyme in the tryptophan biosynthetic pathway, catalyzing the ring closure reaction of 1-(o-carboxyphenylamino)-1-deoxyribulose-5-phosphate (CdRP) to indole-3-glycerol phosphate (IGP), accompanied by the release of carbon dioxide and water.
cd00346 (2) Tryptophan synthase alpha chain; Tryptophan synthase is a bifunctional pyridoxal-5'-phosphate (PLP)-dependent tetrameric (2 alpha and 2 beta subunits) enzyme that catalyzes the last 2 steps of the L-tryptophan synthesis, the condensation of indole-3-glycerol phosphate (IPG) and L-serine to L-tryptophan and water.
COG0133 (26) Tryptophan synthase beta chain [Amino acid transport and metabolism]
COG0134 (4) Indole-3-glycerol phosphate synthase [Amino acid transport and metabolism]
COG0135 (12) Phosphoribosylanthranilate isomerase [Amino acid transport and metabolism]
COG0147 (29) Anthranilate/para-aminobenzoate synthases component I [Amino acid transport and metabolism / Coenzyme metabolism]
COG0159 (8) Tryptophan synthase alpha chain [Amino acid transport and metabolism]
COG0512 (18) Anthranilate/para-aminobenzoate synthases component II [Amino acid transport and metabolism / Coenzyme metabolism]
COG0547 (12) Anthranilate phosphoribosyltransferase [Amino acid transport and metabolism]
COG0624 (1) Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase and related deacylases [Amino acid transport and metabolism]
COG3663 (1) G:T/U mismatch-specific DNA glycosylase [DNA replication, recombination, and

repair]
GOS681 (7) ?

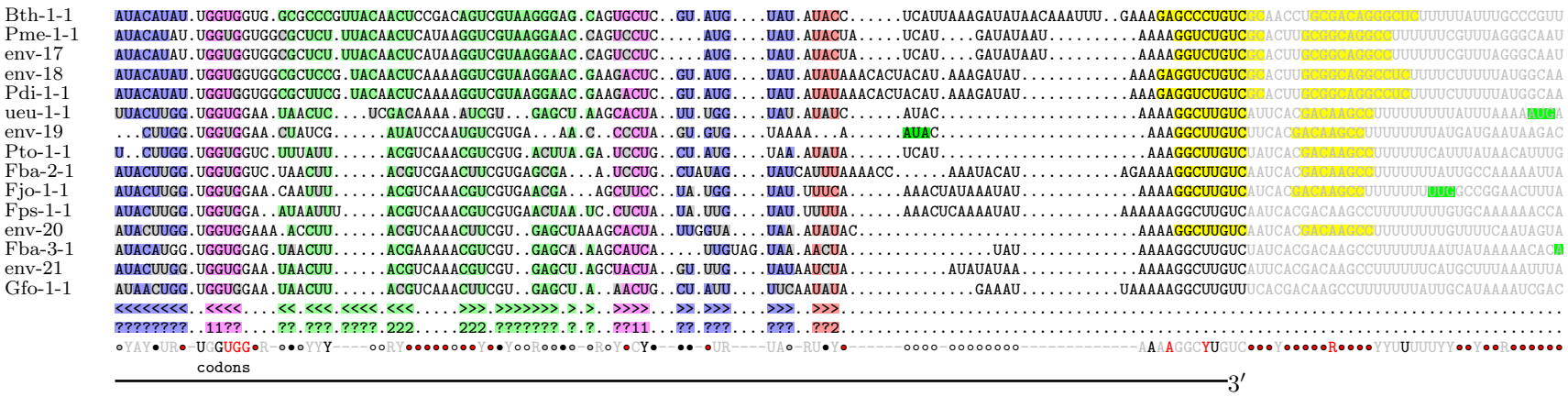
pfam04715 (1) Anthranilate synthase component I, N terminal region.
pfam05635 (1) S23 ribosomal protein.

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: env-9

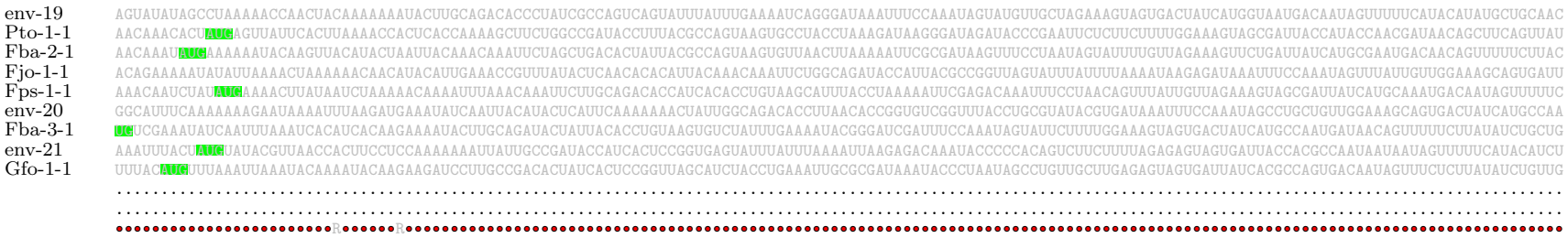
11.4 Multiple-sequence alignment

Each Bacteroid-*trp* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 11.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Bacteroid-*trp* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

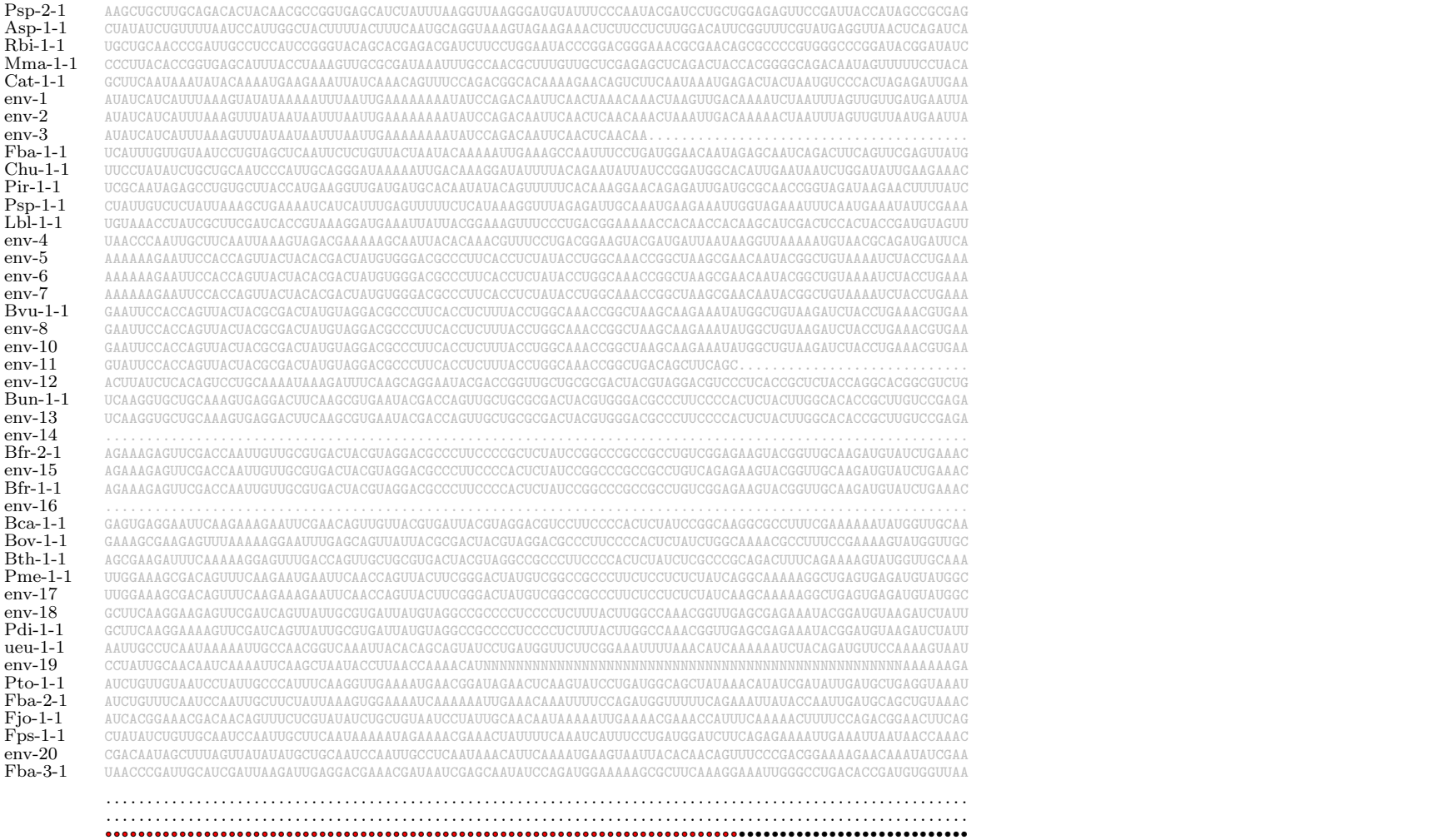
env-15
Bfr-1-1
env-16
Bca-1-1
Bov-1-1
Bth-1-1
Pme-1-1
env-17
env-18
Pdi-1-1
ueu-1-1
env-19
Pto-1-1
Fba-2-1
Fjo-1-1
Fps-1-1
env-20
Fba-3-1
env-21
Gfo-1-1
5′
alignment positions 145 · · · 304
Psp-2-1
Asp-1-1
Rbi-1-1
Mma-1-1
Cat-1-1
env-1
env-2
env-3
Fba-1-1
Chu-1-1
Pir-1-1
Psp-1-1
Lbl-1-1
env-4
env-5
env-6
env-7
Bvu-1-1
env-8
env-10
env-11
env-12
Bun-1-1
env-13
env-14
Bfr-2-1
env-15
Bfr-1-1
env-16
Bca-1-1
Bov-1-1
codons
3′

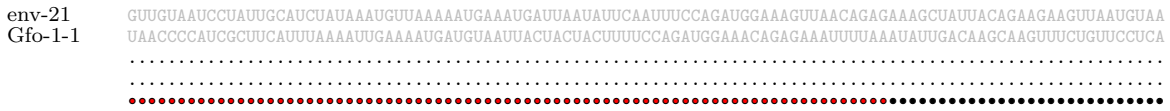


alignment positions 305 · · · 464



alignment positions 465 · · · 569





12 Bacteroidales-1

12.1 Taxa

The taxonomy of each organism containing a putative Bacteroidales-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Bacteroidales-1 RNA in Sections 12.2 and 12.4.

abbrev. of hits	taxonomy of species
Bca-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides caccae* ATCC 43185
Bco-1-1 to Bco-1-2	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides coprocola* DSM 17136
Bdo-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides dorei* DSM 17855
Beg-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides eggerthii* DSM 20697
Bfi-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides finegoldii* DSM 17565
Bfr-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides fragilis* NCTC 9343
Bfr-2-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides fragilis* YCH46
Bin-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides intestinalis* DSM 17393
Bov-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides ovatus* ATCC 8483
Bpl-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides plebeius* DSM 17135
Bst-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides stercoris* ATCC 43183
Bth-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides thetaiotaomicron* VPI-5482
Bun-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides uniformis* ATCC 8492
Bvu-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides vulgatus* ATCC 8482
Pdi-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Porphyromonadaceae *Parabacteroides distasonis* ATCC 8503
Pjo-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Porphyromonadaceae *Parabacteroides johnsonii* DSM 18315
Pme-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Porphyromonadaceae *Parabacteroides merdae* ATCC 43184
Pgi-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Porphyromonadaceae *Porphyromonas gingivalis* ATCC 33277
Pgi-2-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Porphyromonadaceae *Porphyromonas gingivalis* W83
env-1 to env-27	environmental samples

12.2 Gene contexts

Each Bacteroidales-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Bacteroidales-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 12.3. The accession number of the sequence containing each Bacteroidales-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 12.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	BABD01027605.1	-	926	751	←hypo **RNA**→ RRM (smart00360)→
env-2	BAAY01005444.1	-	242	57	**RNA**→
Pgi-1-1	NC_010729.1	-	146751	146582	←MMT1 (COG0053) **RNA**→ ←hypo
Pgi-2-1	NC_002950.2	-	2190292	2190123	←MMT1 (COG0053) **RNA**→ ←hypo
env-3	BAAX01000158.1	-	508	314	HisB (COG0241)IGPD (pfam00475)→ **RNA**→ Lum_binding (pfam00677)→
env-4	BAAX01010365.1	+	733	913	ThrA (COG0460)→ **RNA**→
env-5	BAAX01004754.1	+	420	598	**RNA**→ RibC (COG0307)→
Pdi-1-1	NC_009615.1	+	788079	788256	Ribosomal_L20 (pfam00453)→ **RNA**→ RRM (smart00360)→
env-6	BABA01004120.1	+	1238	1415	Ribosomal_L20 (pfam00453)→ **RNA**→ RRM (smart00360)→
env-7	BABB01008998.1	+	97	274	**RNA**→ RRM (smart00360)→
Pjo-1-1	NZ_ABYH01000068.1	+	291	468	**RNA**→
Pme-1-1	NZ_AAXE02000021.1	-	125990	125813	hypo→ **RNA**→ hypo→
env-8	BABD01027604.1	+	632	810	Ribosomal_L20 (pfam00453)→ **RNA**→
env-9	BABD01016942.1	+	749	930	Ribosomal_L20 (pfam00453)→ **RNA**→
Bpl-1-1	NZ_ABQC02000016.1	+	77157	77332	Ribosomal_L20 (pfam00453)→ **RNA**→ COG2706 (COG2706)→
Bco-1-1	NZ_ABIY02000096.1	-	615	440	RVT (pfam00078)→ **RNA**→ hypo→
Bco-1-2	NZ_ABIY02000127.1	-	14645	14470	COG3344 (COG3344)→ **RNA**→ hypo→
env-10	BABD01000999.1	+	665	840	Ribosomal_L20 (pfam00453)→ **RNA**→ COG1144 (COG1144)→
env-11	BABG01004849.1	+	451	627	**RNA**→ FldA (COG0716)→
Bvu-1-1	NC_009614.1	-	4991729	4991553	Ribosomal_L20 (pfam00453)→ **RNA**→ FldA (COG0716)NuoI (COG1143)→
env-12	BAAU01004407.1	-	862	686	Ribosomal_L20 (pfam00453)→ **RNA**→ FldA (COG0716)→
Bdo-1-1	NZ_ABWZ01000013.1	+	484277	484453	**RNA**→
env-13	BAAX01034787.1	-	345	169	Ribosomal_L20 (pfam00453)→ **RNA**→
env-14	BAAY01001699.1	+	365	541	hypo→ **RNA**→ FldA (COG0716)→
env-15	BABB01000835.1	-	8418	8242	hypo→ **RNA**→ FldA (COG0716)→
env-16	BABA01025825.1	-	358	177	**RNA**→ hypo→
Bca-1-1	NZ_AAVM02000004.1	+	418226	418403	←hypo **RNA**→ FldA (COG0716)NuoI (COG1143)→
Bfi-1-1	NZ_ABXI01000084.1	+	5207	5384	**RNA**→
Bth-1-1	NC_004663.1	+	512928	513105	Ribosomal_L20 (pfam00453)→ **RNA**→ NuoI (COG1143)→
Bov-1-1	NZ_AAXF02000045.1	+	146907	147084	Ribosomal_L20 (pfam00453)→ **RNA**→ NapF (COG1145)→
Bfr-1-1	NC_003228.3	+	1976767	1976943	Ribosomal_L20 (pfam00453)→ **RNA**→ FldA (COG0716)NuoI (COG1143)→
Bfr-2-1	NC_006347.1	+	1960011	1960187	Ribosomal_L20 (pfam00453)→ **RNA**→ FldA (COG0716)NuoI (COG1143)→
env-17	BABB01013913.1	+	442	618	Ribosomal_L20 (pfam00453)→ **RNA**→
Bin-1-1	NZ_ABJL02000008.1	-	781990	781813	Ribosomal_L20 (pfam00453)→ **RNA**→ ←PaaK (COG1541)
env-18	BAAZ01000710.1	+	2540	2717	Ribosomal_L20 (pfam00453)→ **RNA**→
env-19	BABA01003411.1	+	651	828	Ribosomal_L20 (pfam00453)→ **RNA**→ ←hypo
env-20	BAAY01021114.1	+	550	727	Ribosomal_L20 (pfam00453)→ **RNA**→
Bun-1-1	NZ_AAYH02000043.1	+	16479	16656	Ribosomal_L20 (pfam00453)→ **RNA**→ ←PaaK (COG1541)
env-21	BAAX01004326.1	-	2389	2212	**RNA**→ ←PaaK (COG1541)
env-22	BABA01001035.1	-	1522	1345	Ribosomal_L20 (pfam00453)→ **RNA**→ ←PaaK (COG1541)
env-23	BAAZ01019253.1	-	445	268	Ribosomal_L20 (pfam00453)→ **RNA**→
env-24	BABG01009248.1	+	549	726	Ribosomal_L20 (pfam00453)→ **RNA**→
env-25	BABD01001370.1	-	2435	2258	**RNA**→ ←PaaK (COG1541)
env-26	BAAU01008698.1	-	740	563	Ribosomal_L20 (pfam00453)→ **RNA**→ ←hypo
env-27	BABB01001469.1	+	28	205	**RNA**→ ←hypo
Bst-1-1	NZ_ABFZ02000019.1	+	401830	402007	Ribosomal_L20 (pfam00453)→ **RNA**→ hypo→
Beg-1-1	NZ_ABVO01000025.1	+	18262	18439	**RNA**→

12.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 12.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Bacteroidales-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

u COG0053 (2) Predicted Co/Zn/Cd cation transporters [Inorganic ion transport and metabolism]
u COG0241 (1) Histidinol phosphatase and related phosphatases [Amino acid transport and metabolism]
D COG0307 (1) Riboflavin synthase alpha chain [Coenzyme metabolism]
u COG0460 (1) Homoserine dehydrogenase [Amino acid transport and metabolism]
D COG0716 (8) Flavodoxins [Energy production and conversion]
D COG1143 (5) Formate hydrogenlyase subunit 6/NADH:ubiquinone oxidoreductase 23 kD subunit (chain I) [Energy production and conversion]
D COG1144 (1) Pyruvate:ferredoxin oxidoreductase and related 2-oxoacid:ferredoxin oxidoreductases, delta subunit [Energy production and conversion]
D COG1145 (1) Ferredoxin [Energy production and conversion]
d COG1541 (5) Coenzyme F390 synthetase [Coenzyme metabolism]
D COG2706 (1) 3-carboxymuconate cyclase [Carbohydrate transport and metabolism]
u COG3344 (1) Retron-type reverse transcriptase [DNA replication, recombination, and repair]
u pfam00078 (1) Reverse transcriptase (RNA-dependent DNA polymerase).
u pfam00453 (24) Ribosomal protein L20.
u pfam00475 (1) Imidazoleglycerol-phosphate dehydratase.
D pfam00677 (1) Lumazine binding domain.
D smart00360 (4) RNA recognition motif;

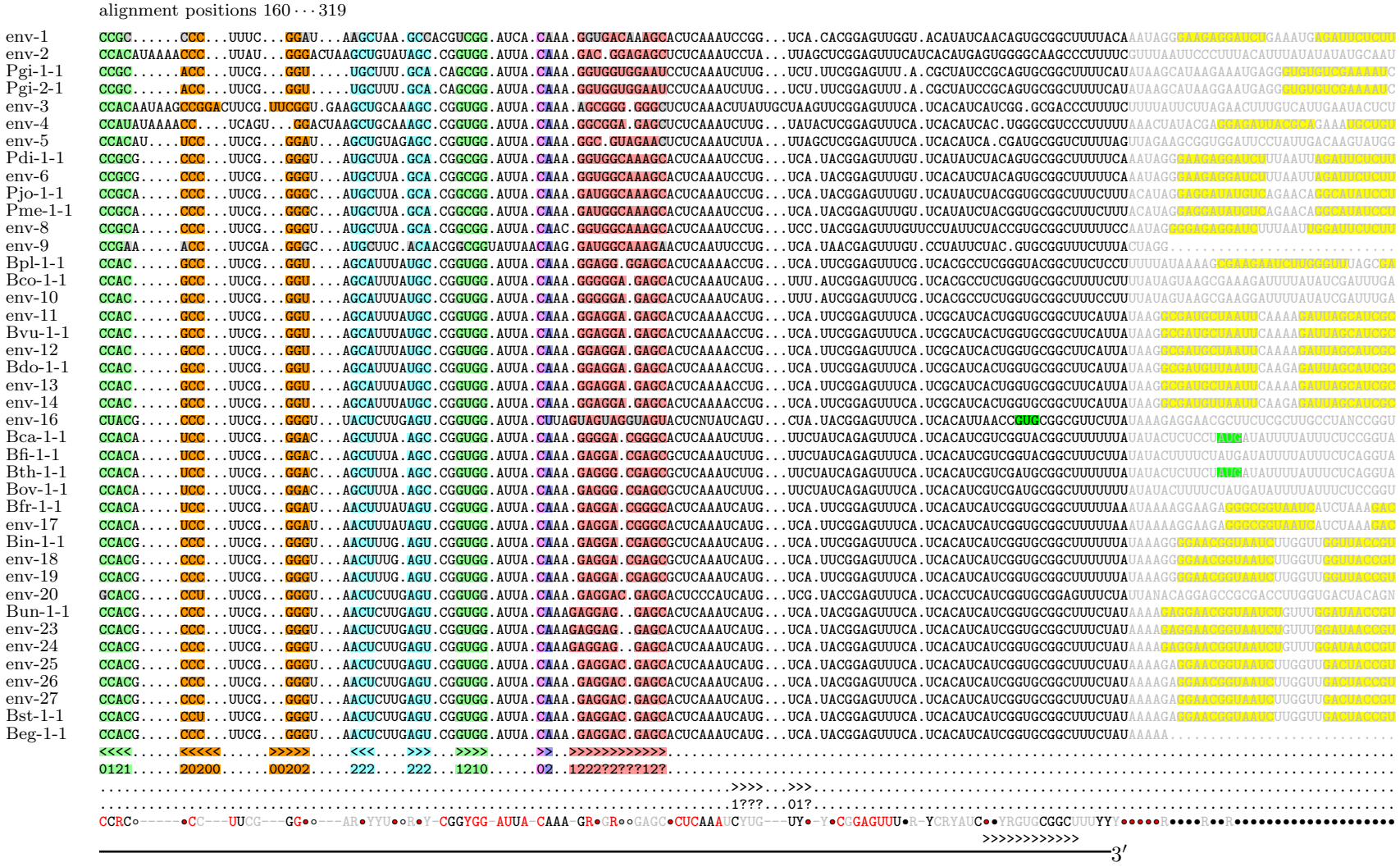
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Bco-1-2, Bfr-2-1, env-15, env-21, env-22, env-7

12.4 Multiple-sequence alignment

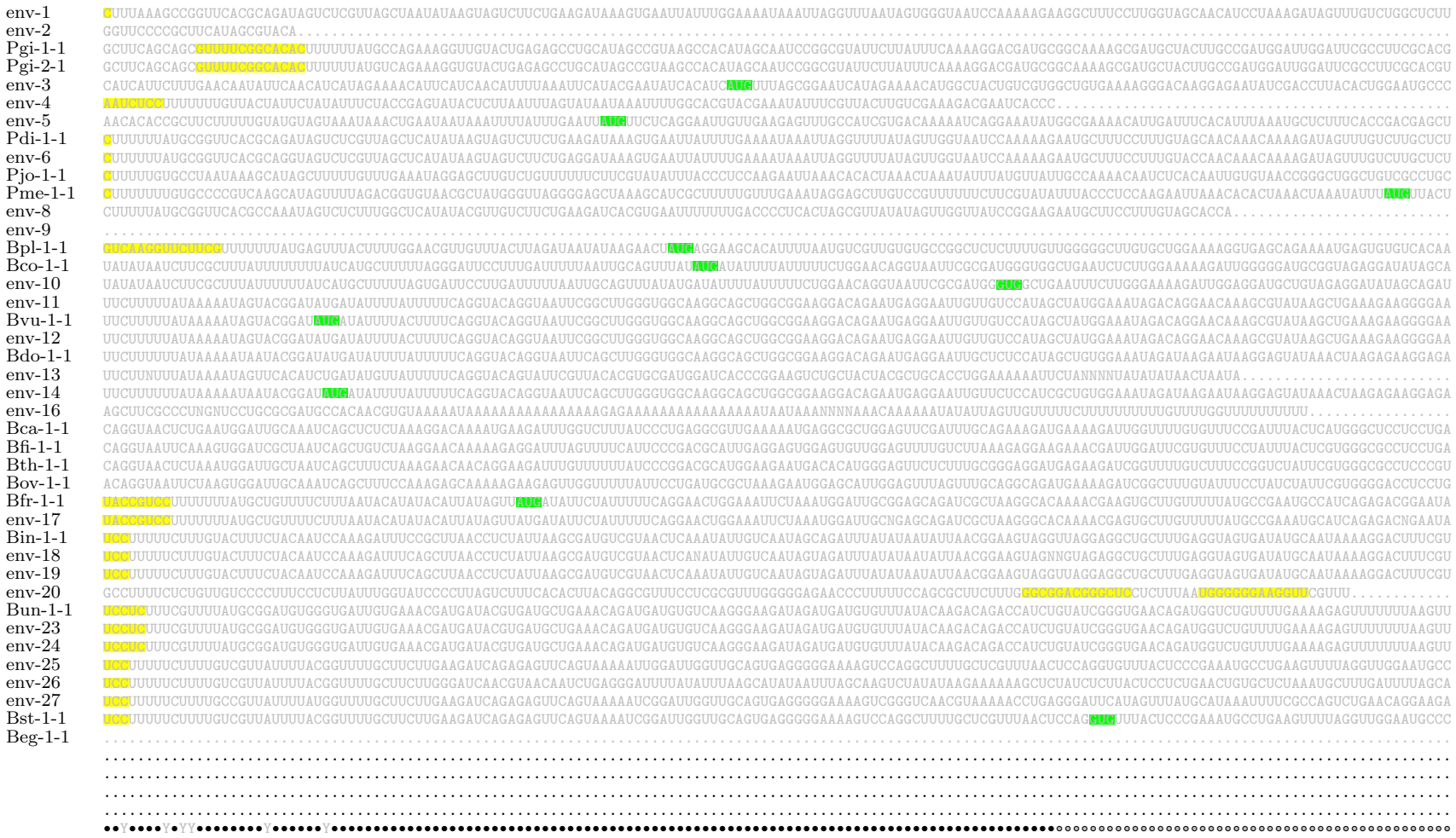
Each Bacteroidales-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 12.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Bacteroidales-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

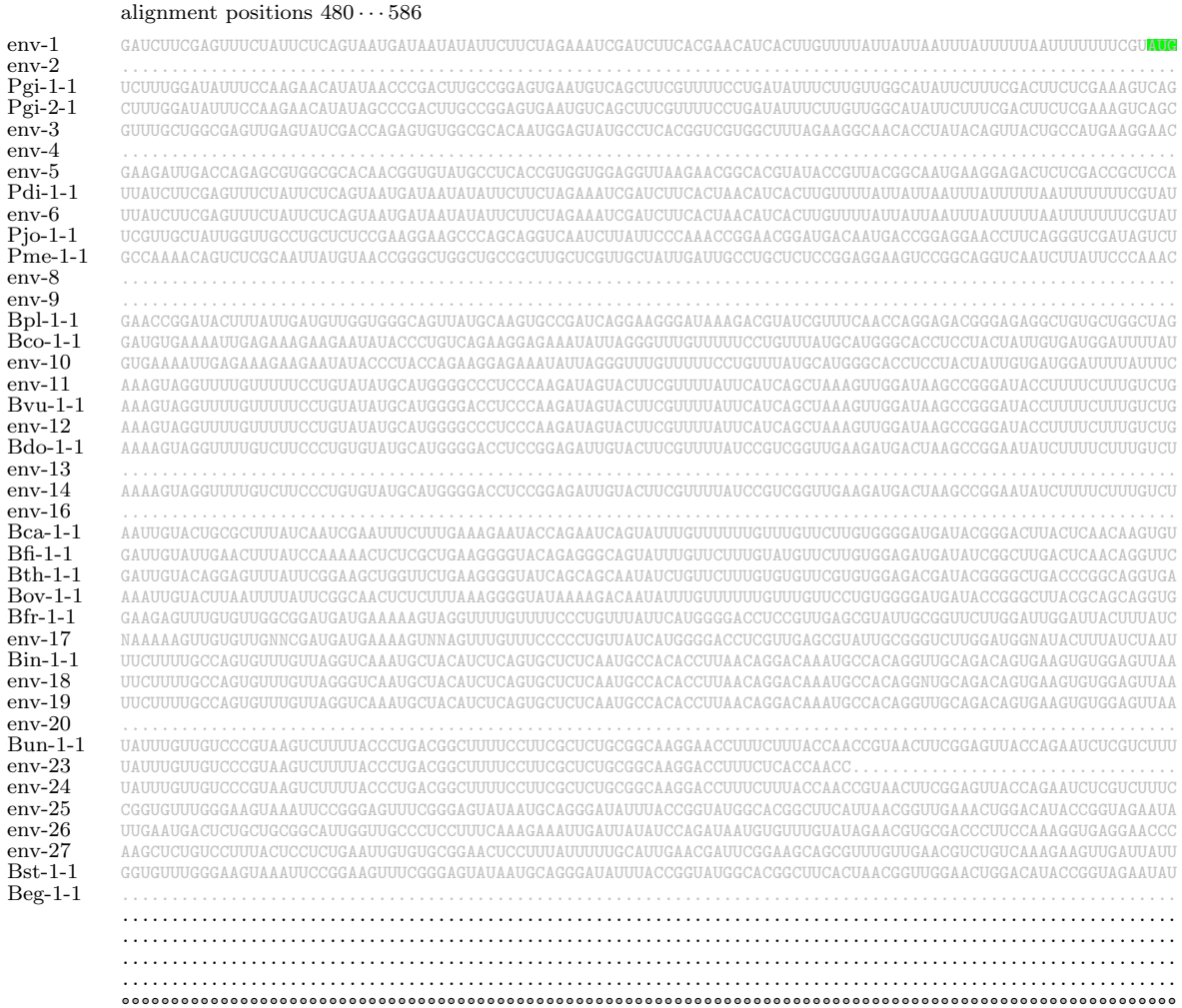
(Features specific to this RNA motif) Canonical *Bacteroidetes* promoters are detected, and we have labeled their -33 motif ("TTTG"), -7 motif ("TAnnTTTG") and rough transcription start site ("TSS").

alignment positions 1 · · · 159
env-1
env-2
Pgi-1-1
Pgi-2-1
env-3
env-4
env-5
Pdi-1-1
env-6
Pjo-1-1
Pme-1-1
env-8
env-9
Bpl-1-1
Bco-1-1
env-10
env-11
Bvu-1-1
env-12
Bdo-1-1
env-13
env-14
env-16
Bca-1-1
Bfi-1-1
Bth-1-1
Bov-1-1
Bfr-1-1
env-17
Bin-1-1
env-18
env-19
env-20
Bun-1-1
env-23
env-24
env-25
env-26
env-27
Bst-1-1
Beg-1-1
TTTG
TAnnTTTG
TSS
5′


alignment positions 160 · · · 319
env-1
env-2
Pgi-1-1
Pgi-2-1
env-3
env-4
env-5
Pdi-1-1
env-6
Pjo-1-1
Pme-1-1
env-8
env-9
Bpl-1-1
Bco-1-1
env-10
env-11
Bvu-1-1
env-12
Bdo-1-1
env-13
env-14
env-16
Bca-1-1
Bfi-1-1
Bth-1-1
Bov-1-1
Bfr-1-1
env-17
Bin-1-1
env-18
env-19
env-20
Bun-1-1
env-23
env-24
env-25
env-26
env-27
Bst-1-1
Beg-1-1
3′

alignment positions 320 · · · 479





13 *Bacteroides*-1

13.1 Taxa

The taxonomy of each organism containing a putative *Bacteroides*-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *Bacteroides*-1 RNA in Sections 13.2 and 13.4.

abbrev. of hits	taxonomy of species
Bca-1-1 to Bca-1-3	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides caccae* ATCC 43185
Bco-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides coprocola* DSM 17136
Bdo-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides dorei* DSM 17855
Beg-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides eggerthii* DSM 20697
Bfr-1-1 to Bfr-1-2	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides fragilis* NCTC 9343
Bfr-2-1 to Bfr-2-2	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides fragilis* YCH46
Bin-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides intestinalis* DSM 17393
Bov-1-1 to Bov-1-3	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides ovatus* ATCC 8483
Bst-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides stercoris* ATCC 43183
Bth-1-1 to Bth-1-3	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides thetaiotaomicron* VPI-5482
Bun-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides uniformis* ATCC 8492
Bvu-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides vulgatus* ATCC 8482
env-1 to env-20	environmental samples

13.2 Gene contexts

Each *Bacteroides*-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *Bacteroides*-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 13.3. The accession number of the sequence containing each *Bacteroides*-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 13.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Bco-1-1	NZ_ABIY02000058.1	+	8754	8822	GalE (COG1087)→ **RNA**→ EriC (COG0038)→
env-1	BABD01000872.1	-	1354	1286	GalE (COG1087)→ **RNA**→ EriC (COG0038)→
Bth-1-1	NC_004663.1	-	2410752	2410667	←hypo **RNA**→ **RNA**→ ←hypo
Bth-1-2	NC_004663.1	-	2410987	2410907	(shown above)
env-2	BAAV01015564.1	+	659	744	Amidase_3 (pfam01520)→ **RNA**→ **RNA**→ hypo→
env-3	BAAV01015564.1	+	424	504	(shown above)
Bca-1-1	NZ_AAVM02000011.1	-	62850	62766	COG3292 (COG3292)HTH_ARAC (smart00342)→ **RNA**→ Cons_hypoth698 (pfam03601)→
Bov-1-1	NZ_AAXF02000043.1	+	119625	119709	AmiC (COG0860)→ **RNA**→ **RNA**→ ←COG4799 (COG4799)
Bov-1-2	NZ_AAXF02000043.1	+	119356	119436	(shown above)
Bca-1-2	NZ_AAVM02000011.1	-	71894	71814	AmiC (COG0860)→ **RNA**→ ←hypo
Bth-1-3	NC_004663.1	+	768695	768775	GalE (COG1087)→ **RNA**→ ←IspE (COG1947)
Bca-1-3	NZ_AAVM02000001.1	-	1013664	1013584	GalE (COG1087)→ **RNA**→ ←IspE (COG1947)
env-4	BABG01016634.1	-	466	385	hypo→ **RNA**→ ←hypo
Bov-1-3	NZ_AAXF02000048.1	+	20925	21006	GalE (COG1087)→ **RNA**→ COG4680 (COG4680)→
env-5	BABB01010445.1	-	346	262	←hypo **RNA**→
Bfr-1-1	NC_003228.3	+	3358325	3358409	hypo→ **RNA**→ ←COG3871 (COG3871)
env-6	BABA01000991.1	+	1142	1226	hypo→ **RNA**→ ←IclR (COG1414)
Bfr-2-1	NC_006347.1	+	3456921	3457005	←hypo **RNA**→ ←COG3871 (COG3871)
env-7	BAAZ01000465.1	+	2353	2434	GalE (COG1087)→ **RNA**→ ←IclR (COG1414)
Bin-1-1	NZ_ABJL02000008.1	-	538328	538247	GalE (COG1087)→ **RNA**→ ←IclR (COG1414)
env-8	BABD01020940.1	+	232	313	hypo→ **RNA**→
env-9	BABG01021432.1	+	589	670	GalE (COG1087)→ **RNA**→
Bun-1-1	NZ_AAYH02000045.1	-	87516	87435	GalE (COG1087)→ **RNA**→ hypo→
env-10	BAAU01022191.1	-	354	273	GalE (COG1087)→ **RNA**→
env-11	BABD01004454.1	+	837	918	GalE (COG1087)→ **RNA**→
env-12	BABB01007489.1	-	586	505	hypo→ **RNA**→ hypo→
Bst-1-1	NZ_ABFZ02000018.1	+	247284	247365	GalE (COG1087)→ **RNA**→ AtpD (COG0055)→
env-13	BAAU01022945.1	+	559	640	GalE (COG1087)→ **RNA**→
Bfr-1-2	NC_003228.3	-	3042077	3041996	GalE (COG1087)→ **RNA**→ ←putative lipoprotein
Bfr-2-2	NC_006347.1	-	2980257	2980176	GalE (COG1087)→ **RNA**→ ←hypo
env-14	BAAY01000403.1	+	207	288	**RNA**→ AtpD (COG0055)→
Beg-1-1	NZ_ABVO01000039.1	+	26199	26280	**RNA**→
env-15	BAAX01008592.1	-	313	232	hypo→ **RNA**→
env-16	BABB01000399.1	-	1462	1380	**RNA**→ hypo→
env-17	BAAU01002748.1	-	674	588	GalE (COG1087)→ **RNA**→ hypo→
Bdo-1-1	NZ_ABWZ01000015.1	+	81732	81818	**RNA**→
env-18	BAAV01002879.1	-	654	568	GalE (COG1087)→ **RNA**→ hypo→
Bvu-1-1	NC_009614.1	-	4788592	4788506	GalE (COG1087)→ **RNA**→ hypo→
env-19	BABA01008930.1	-	1160	1074	hypo→ **RNA**→ hypo→
env-20	BABG01024910.1	+	369	455	hypo→ **RNA**→ hypo→

13.3 Conserved domains

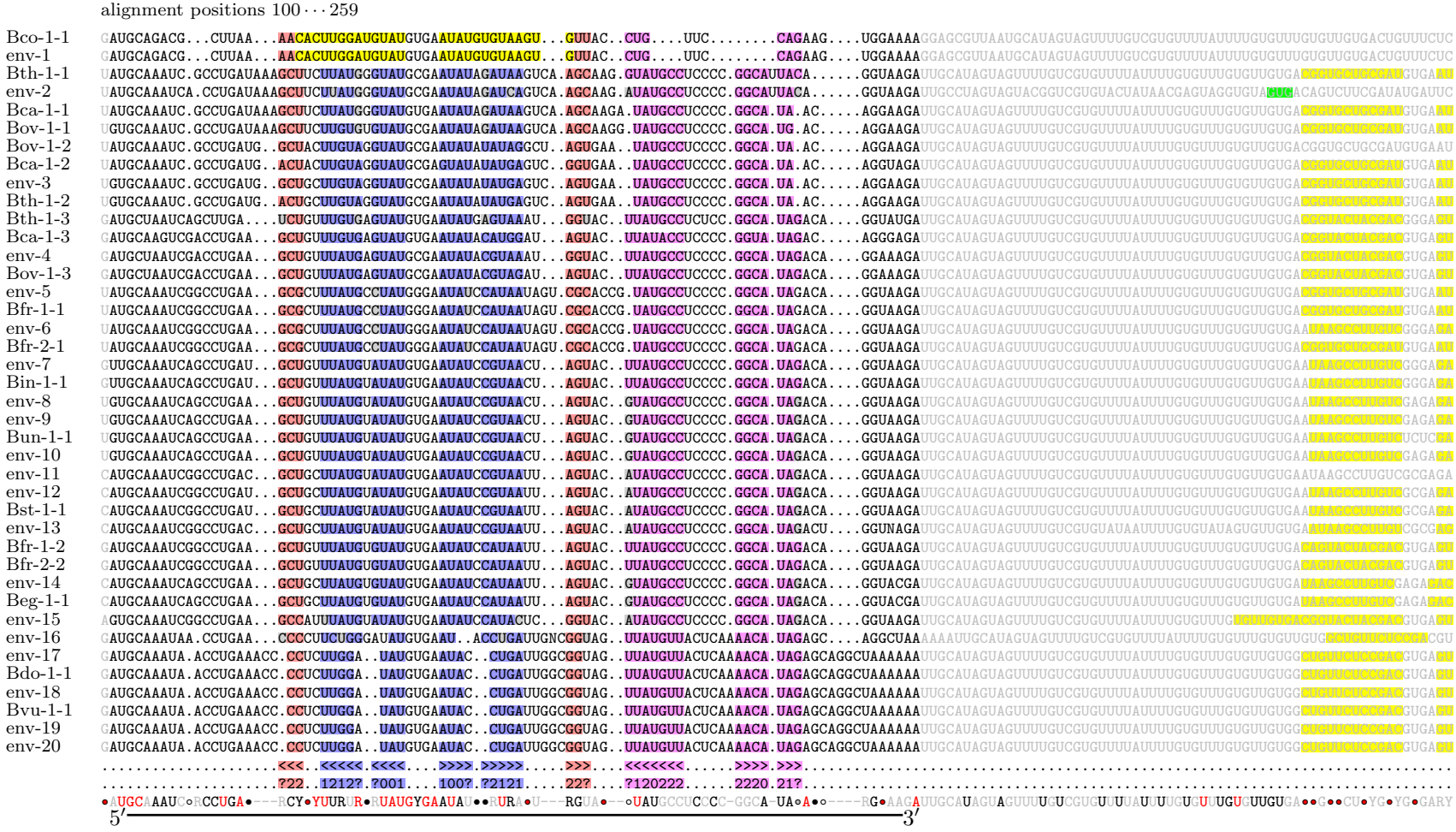
Conserved domains found in protein-coding genes listed in Section 13.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *Bacteroides*-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

D COG0038 (2) Chloride channel protein EriC [Inorganic ion transport and metabolism]
D COG0055 (2) F0F1-type ATP synthase, beta subunit [Energy production and conversion]
u COG0860 (2) N-acetylmuramoyl-L-alanine amidase [Cell envelope biogenesis, outer membrane]
u COG1087 (18) UDP-glucose 4-epimerase [Cell envelope biogenesis, outer membrane]
d COG1414 (3) Transcriptional regulator [Transcription]
d COG1947 (2) 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate synthase [Lipid metabolism]
u COG3292 (1) Predicted periplasmic ligand-binding sensor domain [Signal transduction mechanisms]
d COG3871 (2) Uncharacterized stress protein (general stress protein 26) [General function prediction only]
D COG4680 (1) Uncharacterized protein conserved in bacteria [Function unknown]
d COG4799 (1) Acetyl-CoA carboxylase, carboxyltransferase component (subunits alpha and beta) [Lipid metabolism]
u pfam01520 (1) N-acetylmuramoyl-L-alanine amidase.
D pfam03601 (1) Conserved hypothetical protein 698.
u smart00342 (1) helix_turn_helix, arabinose operon control protein;

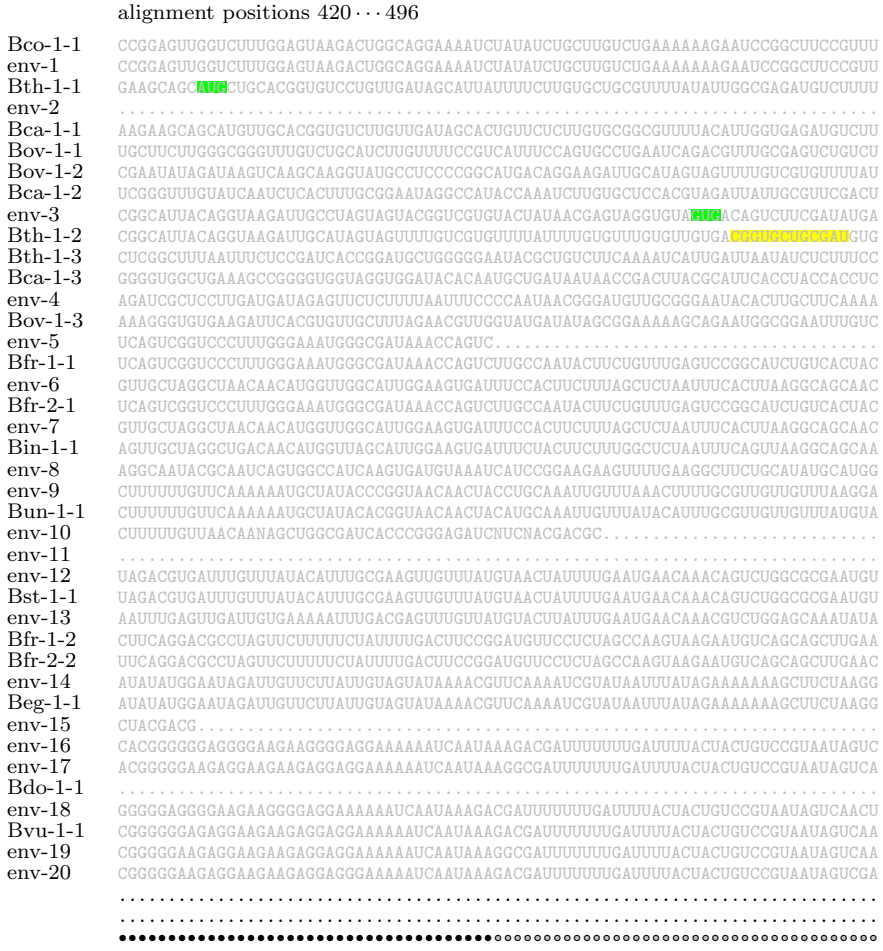
13.4 Multiple-sequence alignment

Each *Bacteroides*-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 13.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *Bacteroides*-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

alignment positions 1 · · · 99


alignment positions 100 · · · 259
Bco-1-1
env-1
Bth-1-1
env-2
Bca-1-1
Bov-1-1
Bov-1-2
Bca-1-2
env-3
Bth-1-2
Bth-1-3
Bca-1-3
env-4
Bov-1-3
env-5
Bfr-1-1
env-6
Bfr-2-1
env-7
Bin-1-1
env-8
env-9
Bun-1-1
env-10
env-11
env-12
Bst-1-1
env-13
Bfr-1-2
Bfr-2-2
env-14
Beg-1-1
env-15
env-16
env-17
Bdo-1-1
env-18
Bvu-1-1
env-19
env-20
5′
3′

alignment positions 260 · · · 419



14 *Bacteroides*-2

14.1 Taxa

The taxonomy of each organism containing a putative *Bacteroides*-2 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *Bacteroides*-2 RNA in Sections 14.2 and 14.4.

abbrev. of hits	**taxonomy of species**
Bca-1-1 to Bca-1-9	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides capillosus* ATCC 29799
env-1 to env-57	environmental samples

14.2 Gene contexts

Each *Bacteroides*-2 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *Bacteroides*-2 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 14.3. The accession number of the sequence containing each *Bacteroides*-2 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 14.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	BABD01006177.1	+	420	476	←hypo **RNA**→ ←OmpR (COG0745)
env-2	BABB01012787.1	-	923	862	**RNA**→ ←Metalloenzyme (pfam01676)
env-3	BABD01026837.1	+	161	220	**RNA**→ **RNA**→ ←Mur_ligase_C (pfam02875)
env-4	BABD01026837.1	+	346	405	(shown above)
env-5	BABA01023012.1	-	391	329	←hypo **RNA**→ **RNA**→ ←hypo
env-8	BABA01023012.1	-	188	126	(shown above)
env-6	BABD01023490.1	+	297	359	←hypo **RNA**→ **RNA**→ ←S1 (smart00316)
env-9	BABD01023490.1	+	500	562	(shown above)
env-7	BABB01003135.1	-	807	746	←hypo **RNA**→ ←S1 (smart00316)
env-10	BABB01003677.1	+	906	956	NtpB (COG1156)→ **RNA**→ ATP-synt_D (pfam01813)→
env-11	BABB01017599.1	+	202	262	**RNA**→ ←hypo
env-12	BABD01023181.1	+	345	406	hypo→ **RNA**→ GTP_EFTU (pfam00009)→
env-13	BAAY01006515.1	-	627	567	ModA (COG0725)→ **RNA**→ ModC (COG4149)→
env-14	BABG01020158.1	-	574	517	←zf-CHC2 (pfam01807) **RNA**→ ←hypo
env-15	BABB01010706.1	+	437	494	←hypo **RNA**→
env-16	BABD01029485.1	+	372	431	**RNA**→ GIYc (smart00465)→
env-17	BAAX01015758.1	+	773	834	←hypo **RNA**→ ←OmpR (COG0745)
env-18	BAAV01027752.1	+	834	892	←HAMP (smart00304) **RNA**→
env-19	BAAZ01007067.1	+	359	417	←hypo **RNA**→ ←OmpR (COG0745)
env-20	BAAY01006259.1	-	378	317	hypo→ **RNA**→ **RNA**→ ←hypo
env-24	BAAY01006259.1	-	593	532	(shown above)
env-21	BABD01002881.1	-	1028	971	←Acetyltransf_1 (pfam00583) **RNA**→ ←IlvH (COG0440)
env-22	BABA01027753.1	+	861	922	hypo→ **RNA**→
env-23	BABB01009884.1	-	396	339	**RNA**→ ←YqgFc (smart00732)
Bca-1-1	NZ_AAXG02000041.1	-	20454	20396	hypo→ **RNA**→ ←Upp (COG0035)
env-25	BABA01007724.1	-	485	426	←hypo **RNA**→ ←hypo
env-26	BABB01011836.1	-	318	260	Leu tRNA→ **RNA**→
Bca-1-2	NZ_AAXG02000019.1	-	42448	42391	HIS2 (COG1387)→ **RNA**→ hypo→
env-27	BAAX01027037.1	+	205	262	←hypo **RNA**→ ←hypo
env-28	BAAX01027037.1	+	435	492	←hypo **RNA**→ ←hypo
env-29	BABA01007680.1	-	711	650	←Enolase_N (pfam03952) **RNA**→ ←Metalloenzyme (pfam01676)
env-30	BABD01023182.1	+	19	79	**RNA**→ ←EFG_C (pfam00679)LepA_C (pfam06421)
env-31	BAAZ01005461.1	-	884	827	←Rnc (COG0571) **RNA**→ ←Rnc (COG0571)
env-32	BABD01023870.1	+	89	150	←hypo **RNA**→ ←DHHA1 (pfam02272)
env-33	BABD01024680.1	-	99	41	hypo→ **RNA**→
env-34	BAAZ01009952.1	+	406	522	COG1937 (COG1937)→ **RNA**→ HMA (cd00371)→
Bca-1-3	NZ_AAXG02000006.1	+	58384	58443	hypo→ **RNA**→ ←rubredoxin_SM (cd00729)Rubrerythrin_like (cd01046)
env-35	BAAZ01029945.1	-	420	361	←hypo **RNA**→ ←hypo

env-36	BAAZ01015734.1	-	284	229	←hypo **RNA**→
Bca-1-4	NZ_AAXG02000006.1	+	93579	93636	hypo→ **RNA**→ ZntA (COG2217)CopZ (COG2608)→
env-37	BAAZ01029758.1	+	489	548	Ile tRNA→ **RNA**→
env-38	BAAV01030571.1	+	354	413	←RbfA (COG0858) **RNA**→
env-39	BAAX01027502.1	+	395	456	hypo→ **RNA**→ ←SIS (pfam01380)
env-40	BAAY01013091.1	+	1353	1413	barnase (cd00933)→ **RNA**→ barnase (cd00933)→
env-42	BAAY01022414.1	-	154	97	hypo→ **RNA**→
env-41	BAAY01022414.1	-	366	309	(shown above)
Bca-1-5	NZ_AAXG02000011.1	+	55260	55319	←hypo **RNA**→ ←AlaS (COG0013)tRNA-synt_2c (pfam01411)DHHA1 (pfam02272)
Bca-1-6	NZ_AAXG02000010.1	-	39737	39678	hypo→ **RNA**→ hypo→
env-43	BAAZ01022445.1	-	602	542	**RNA**→ hypo→
env-44	BAAY01009092.1	-	436	376	DRTGG (pfam07085)→ **RNA**→ **RNA**→ POLIIIAc (smart00481)→
env-49	BAAY01009092.1	-	642	582	(shown above)
env-45	BABA01009539.1	+	34	94	**RNA**→ POLIIIAc (smart00481)→
env-46	BAAZ01022446.1	-	519	459	**RNA**→ **RNA**→ ←Dtyr_deacylase (cd00563)
env-47	BAAZ01022446.1	-	308	248	(shown above)
env-48	BABB01011814.1	+	103	163	**RNA**→ ←Dtyr_deacylase (cd00563)
env-50	BAAX01006424.1	+	687	744	←MazF (COG2337) **RNA**→ ←Ala_racemase (cd00430)
env-51	BABA01022941.1	+	154	213	**RNA**→ ←EF_TS (pfam00889)
env-52	BABE01019904.1	+	317	376	hypo→ **RNA**→ hypo→
env-53	BABA01013052.1	-	751	690	←hypo **RNA**→ **RNA**→ ←hypo
env-54	BABA01013052.1	-	535	474	(shown above)
Bca-1-7	NZ_AAXG02000047.1	-	151807	151748	←hypo **RNA**→ ←EFP (pfam01132)
Bca-1-8	NZ_AAXG02000045.1	-	57489	57431	hypo→ **RNA**→ ←PheT (COG0072)tRNA_bind_bactPheRS (cd02796)
Bca-1-9	NZ_AAXG02000047.1	-	23814	23754	YtcJ_like (cd01300)→ **RNA**→ SdaA (COG1760)→
env-55	BABD01004318.1	+	1157	1216	←hypo **RNA**→ hypo→
env-56	BABG01015003.1	-	206	147	**RNA**→
env-57	BABG01015003.1	-	458	399	**RNA**→

14.3 Conserved domains

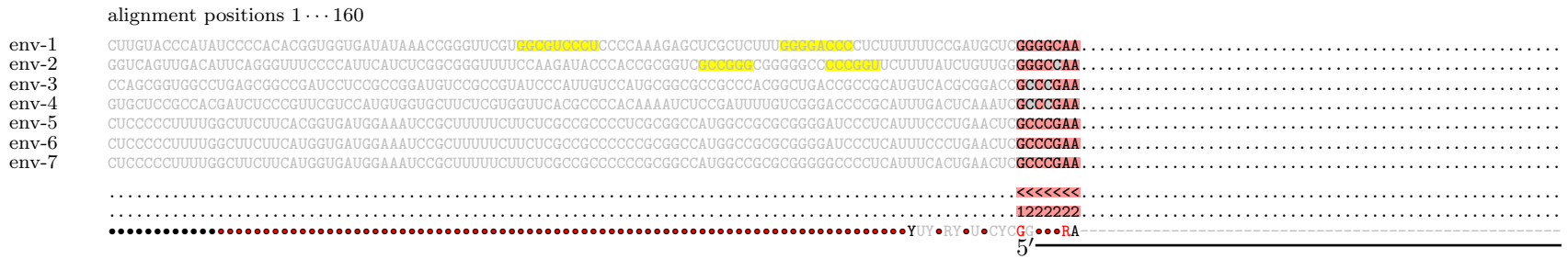
Conserved domains found in protein-coding genes listed in Section 14.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *Bacteroides*-2 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

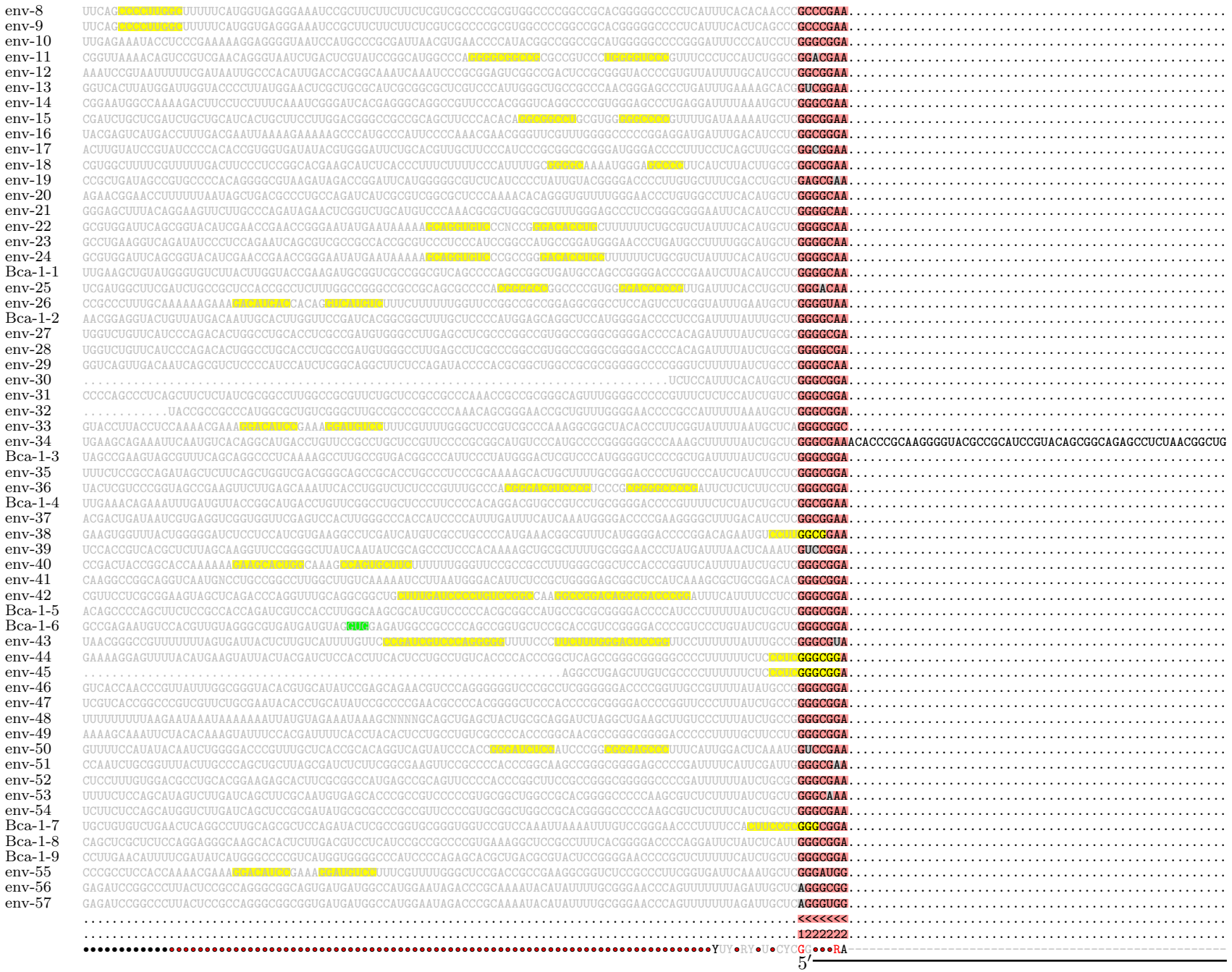
D cd00371 (1) Heavy-metal-associated domain (HMA) is a conserved domain of approximately 30 amino acid residues found in a number of proteins that transport or detoxify heavy metals, for example, the CPx-type heavy metal ATPases and copper chaperones.
d cd00430 (1) Alanine racemase.
d cd00563 (2) D-Tyrosyl-tRNAtyr deacylases; a class of tRNA-dependent hydrolases which are capable of hydrolyzing the ester bond of D-Tyrosyl-tRNA reducing the level of cellular D-Tyrosine while recycling the peptidyl-tRNA; found in bacteria and in eukaryotes but not in archea; beta barrel-like fold structure; forms homodimers in which two surface cavities serve as the active site for tRNA binding
d cd00729 (1) Rubredoxin, Small Modular nonheme iron binding domain containing a [Fe(SCys)4]
uD cd00933 (2) Barnase, a member of the family of homologous microbial ribonucleases, catalyses the cleavage of single-stranded RNA via a two-step mechanism thought to be similar to that of pancreatic ribonuclease.
d cd01046 (1) Rubrerythrin-like domain,similar to rubrerythrin, a nonheme iron binding domain found in many air-sensitive bacteria and archaea, and member of a broad superfamily of ferritin-like diiron-carboxylate proteins.
u cd01300 (1) YtcJ_like metal dependent amidohydrolases.
d cd02796 (1) tRNA-binding-domain-containing prokaryotic phenylalanly tRNA synthetase (PheRS) beta chain.
d COG0013 (1) Alanyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
d COG0035 (1) Uracil phosphoribosyltransferase [Nucleotide transport and metabolism]
d COG0072 (1) Phenylalanyl-tRNA synthetase beta subunit [Translation, ribosomal structure and biogenesis]
d COG0440 (1) Acetolactate synthase, small (regulatory) subunit [Amino acid transport and metabolism]

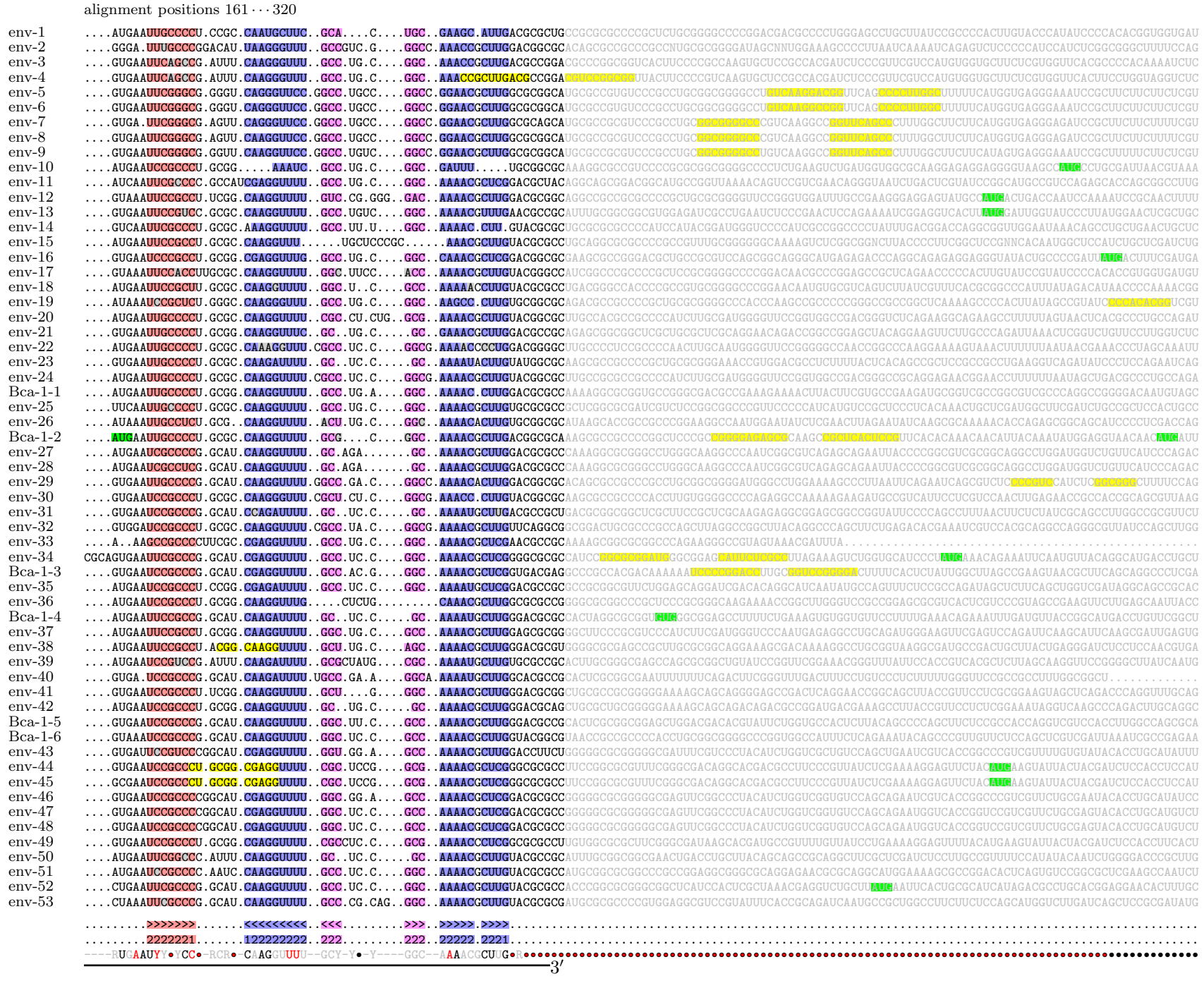
u d COG0571 (2) dsRNA-specific ribonuclease [Transcription]
u COG0725 (1) ABC-type molybdate transport system, periplasmic component [Inorganic ion transport and metabolism]
d COG0745 (3) Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain [Signal transduction mechanisms / Transcription]
u COG0858 (1) Ribosome-binding factor A [Translation, ribosomal structure and biogenesis]
u COG1156 (1) Archaeal/vacuolar-type H+-ATPase subunit B [Energy production and conversion]
u COG1387 (1) Histidinol phosphatase and related hydrolases of the PHP family [Amino acid transport and metabolism / General function prediction only]
D COG1760 (1) L-serine deaminase [Amino acid transport and metabolism]
u COG1937 (1) Uncharacterized protein conserved in bacteria [Function unknown]
D COG2217 (1) Cation transport ATPase [Inorganic ion transport and metabolism]
u COG2337 (1) Growth inhibitor [Signal transduction mechanisms]
D COG2608 (1) Copper chaperone [Inorganic ion transport and metabolism]
D COG4149 (1) ABC-type molybdate transport system, permease component [Inorganic ion transport and metabolism]
D pfam00009 (1) Elongation factor Tu GTP binding domain.
u pfam00583 (1) Acetyltransferase (GNAT) family.
d pfam00679 (1) Elongation factor G C-terminus.
d pfam00889 (1) Elongation factor TS.
d pfam01132 (1) Elongation factor P (EF-P).
d pfam01380 (1) SIS domain.
d pfam01411 (1) tRNA synthetases class II (A).
d pfam01676 (2) Metalloenzyme superfamily.
u pfam01807 (1) CHC2 zinc finger.
D pfam01813 (1) ATP synthase subunit D.
d pfam02272 (2) DHHA1 domain.
d pfam02875 (1) Mur ligase family, glutamate ligase domain.
u pfam03952 (1) Enolase, N-terminal domain.
d pfam06421 (1) GTP-binding protein LepA C-terminus.
u pfam07085 (1) DRTGG domain.
u smart00304 (1) HAMP (Histidine kinases, Adenylyl cyclases, Methyl binding proteins, Phosphatases) domain;
d smart00316 (2) Ribosomal protein S1-like RNA-binding domain;
D smart00465 (1) GIY-YIG type nucleases (URI domain);
D smart00481 (2) DNA polymerase alpha chain like domain; DNA polymerase alpha chain like domain, incl.
d smart00732 (1) Likely ribonuclease with RNase H fold.

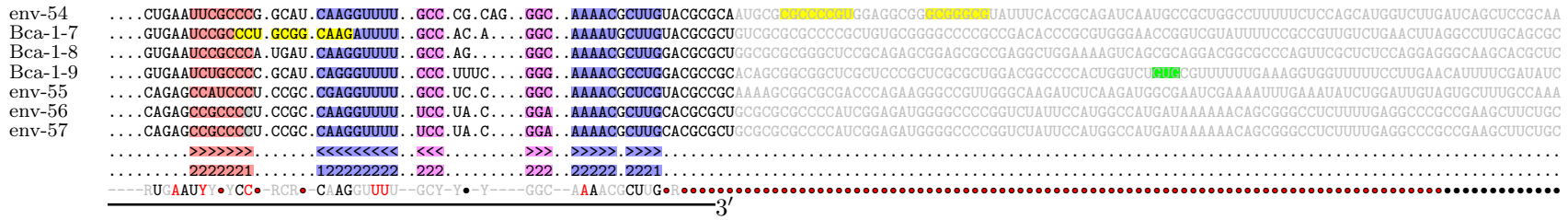
14.4 Multiple-sequence alignment

Each *Bacteroides*-2 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 14.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *Bacteroides*-2 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

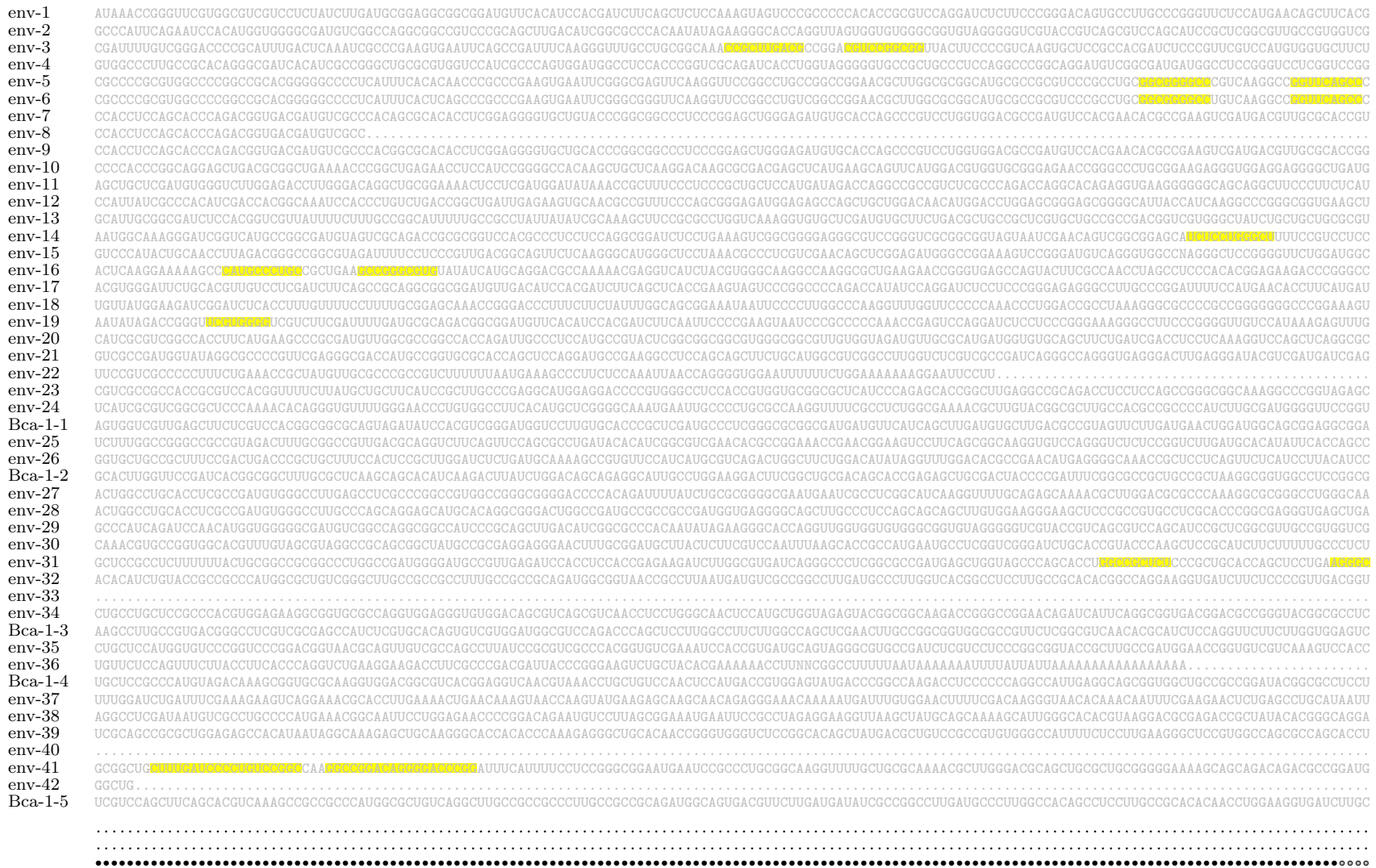


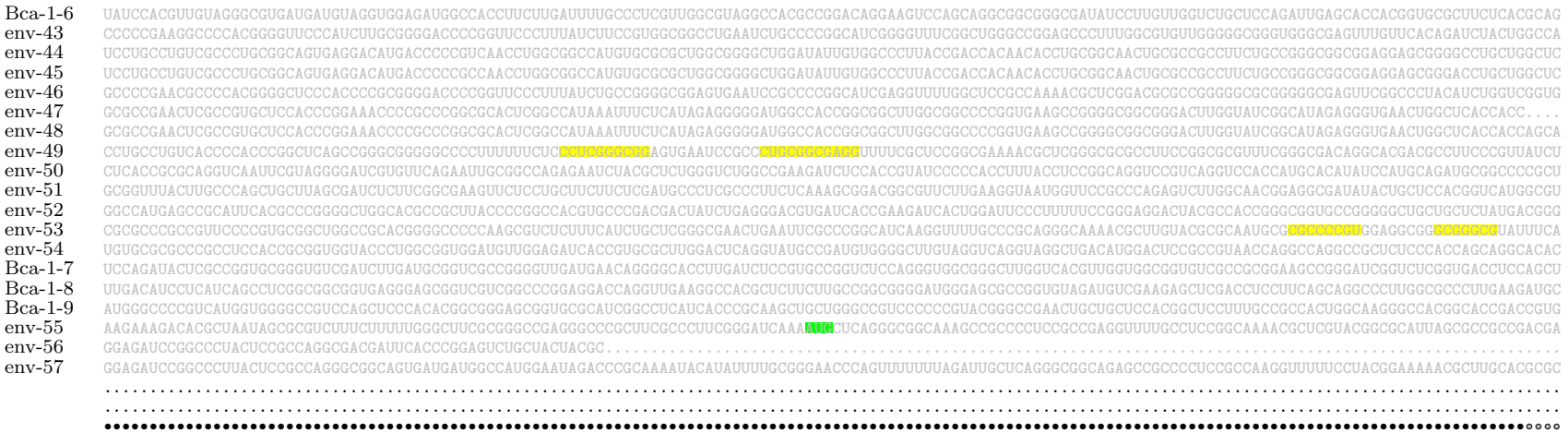

env-8
env-9
env-10
env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
Bca-1-1
env-25
env-26
Bca-1-2
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
Bca-1-3
env-35
env-36
Bca-1-4
env-37
env-38
env-39
env-40
env-41
env-42
Bca-1-5
Bca-1-6
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
Bca-1-7
Bca-1-8
Bca-1-9
env-55
env-56
env-57
5′


alignment positions 161 · · · 320
env-1
env-2
env-3
env-4
env-5
env-6
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
Bca-1-1
env-25
env-26
Bca-1-2
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
Bca-1-3
env-35
env-36
Bca-1-4
env-37
env-38
env-39
env-40
env-41
env-42
Bca-1-5
Bca-1-6
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
3′

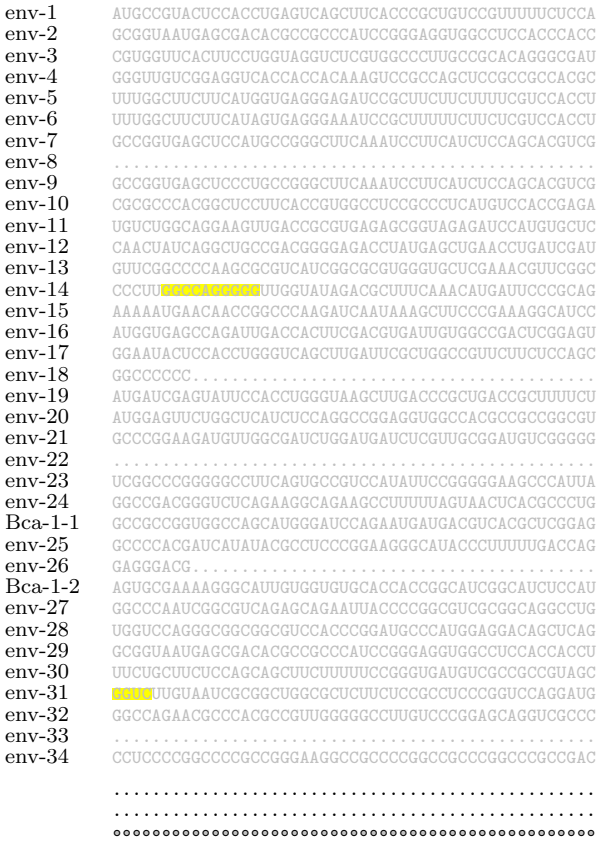


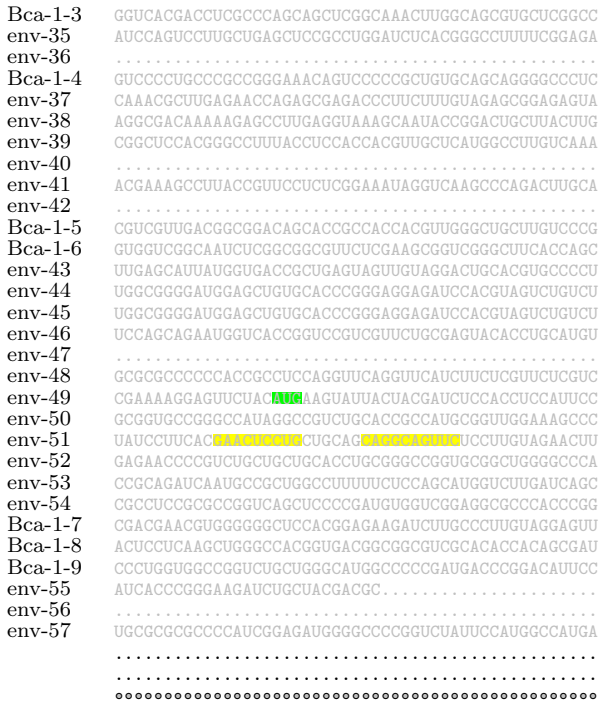
alignment positions 321 · · · 480





alignment positions 481 · · · 529




Bca-1-3 GGUCACGACCUCGCCCAGCAGCUCGGCAAACUUGGCAGCGUGCUCGGCC
env-35 AUCCAGUCCUUGCUGAGCUCCGCCUGGAUCUCACGGGCCUUUUCGGAGA
env-36 ....
Bca-1-4 GUCCCCUGCCCGCCGGGAAACAGUCCCCCGCUGUGCAGCAGGGGCCCUC
env-37 CAAACGCUUGAGAACCAGAGCGAGACCCUUCUUUGUAGAGCGGAGAGUA
env-38 AGGCGACAAAAAGAGCCUUGAGGUAAAGCAAUACCGGACUGCUUACUUG
env-39 CGGCUCCACGGGCCUUUACCUCCACCACGUUGCUCAUGGCCUUGUCAAA
env-40 ....
env-41 ACGAAAGCCUUACCGUUCCUCUCGGAAAUAGGUCAAGCCCAGACUUGCA
env-42 ....
Bca-1-5 CGUCGUUGACGGCGGACAGCACCGCCACCACGUUGGGCUGCUUGUCCCG
Bca-1-6 GUGGUCGGCAAUCUCGGCGGCGUUCUCGAAGCGGUCGGGCUUCACCAGC
env-43 UUGAGCAUUAUGGUGACCGCUGAGUAGUUGUAGGACUGCACGUGCCCCU
env-44 UGGCGGGGAUGGAGCUGUGCACCCGGGAGGAGAUCCACGUAGUCUGUCU
env-45 UGGCGGGGAUGGAGCUGUGCACCCGGGAGGAGAUCCACGUAGUCUGUCU
env-46 UCCAGCAGAAUGGUCACCGGUCCGUCGUUCUGCGAGUACACCUGCAUGU
env-47 ....
env-48 GCGCGCCCCCCACCGCCUCCAGGUUCAGGUUCAUCUUCUCGUUCUCGUC
env-49 CGAAAAGGAGUUCUACAUGAAGUAUUACUACGAUCUCCACCUCCAUUCC
env-50 GCGGUGCCGGGCCAUAGGCCGUCUGCACCGCCAUGCGGUUGGAAAGCCC
env-51 UAUCCUUCACUAACUCCUGCUGCAGCAGGCAGUUCUCCUUGUAGAACUU
env-52 GAGAACCCCGUCUGCUGCUGCACCUGCGGGCCGGUGCGGCUGGGGCCCA
env-53 CCGCAGAUCAAUGCCGCUGGCCUUUUUCUCCAGCAUGGUCUUGAUCAGC
env-54 CGCCUCCGCCGCCGGUCAGCUCCCCGAUGUGGUCGGAGGCGCCCACCCGG
Bca-1-7 CGACGAACGUGGGGGGCUCCACGGAGAAGAUCUUGCCCUUGUAGGAGUU
Bca-1-8 ACUCCUCAAGCUGGGCCACGGUGACGGCGGCGUCGCACACCACAGCGAU
Bca-1-9 CCCUGGUGGCCGGUCUGCUGGGCAUGGCCCCCGAUGACCCGGACAUUCC
env-55 AUCACCCGGGAAGAUCUGCUACGACGC.....................
env-56 ....
env-57 UGCGCGCGCCCCAUCGGAGAUGGGGCCCCGGUCUAUUCCAUGGCCAUGA
.....
.....
°°°

Each Burkholderiales-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Burkholderiales-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 15.3. The accession number of the sequence containing each Burkholderiales-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 15.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	JCVI_SCAF_1096628042959	+	1204	1263	←GD_AH_N (pfam04292)? (GOS3923) **RNA**→ ←? (GOS381)
Rme-1-1	NC_007973.1	+	1410864	1410915	←VapI (COG3093) **RNA**→ ←HyaC (COG1969)
Aav-1-1	NC_008752.1	+	2544313	2544364	←COG2908 (COG2908) **RNA**→ ←cyclophilin_EcCYP_like (cd01920)
Aav-1-2	NC_008752.1	-	3804316	3804265	MA (smart00283)→ **RNA**→ ←TPR (cd00189)COG4976 (COG4976)
Psp-1-1	NC_007948.1	+	2957718	2957769	DppC (COG1173)→ **RNA**→ DppD (COG0444)→
env-2	JCVI_SCAF_1096627195847	+	73	124	**RNA**→ ←Bug (pfam03401)? (GOS6036)
Mpe-1-1	NC_008825.1	-	604396	604345	←FliM (COG1868) **RNA**→ ←FliL (COG1580)
Aav-1-3	NC_008752.1	-	368521	368470	←TufB (COG0050) **RNA**→ ←FusA (COG0480)
env-3	JCVI_SCAF_1096627691218	+	96	147	**RNA**→ ←PaaB (pfam06243)? (GOS13617)
Psp-1-2	NC_007948.1	-	2342985	2342933	←PhoU (COG0704) **RNA**→ ←PstB (COG1117)
Psp-1-3	NC_007948.1	+	609731	609783	←LivH (COG0559) **RNA**→ ←LivG (COG0411)
Psp-1-4	NC_007948.1	+	4969255	4969308	←SucD (COG0074) **RNA**→ ←SucC (COG0045)
Aav-1-4	NC_008752.1	-	1318518	1318468	←RecO (COG1381) **RNA**→ ←Era (COG1159)
env-4	JCVI_SCAF_1096626880474	-	686	635	←CcmA (COG4133)? (GOS1277) **RNA**→
env-5	sludgePhrap_Contig9288	+	699	750	←FcbC (COG0824) **RNA**→ ←Trunc_globin (cd00454)
Aav-1-5	NC_008752.1	-	716470	716399	←NarG (COG5013) **RNA**→ ←NarK (COG2223)
Psp-1-5	NC_007948.1	-	4983586	4983535	←PotA (COG3842) **RNA**→ ←AfuA (COG1840)
Psp-1-6	NC_007948.1	+	260351	260403	←BaeS (COG0642) **RNA**→ ←OmpR (COG0745)
Asp-1-1	NC_008782.1	-	137440	137389	←AcetylCoA_hydro (pfam02550) **RNA**→ ←CaiD (COG1024)
Rfe-1-1	NC_007908.1	-	2308753	2308702	←COG4239 (COG4239) **RNA**→ ←COG4239 (COG4239)
Psp-1-7	NC_007948.1	-	4082160	4082108	←PotC (COG1177) **RNA**→ ←PotB (COG1176)
Psp-1-8	NC_007948.1	+	4172154	4172206	rpsA (cd03322)→ **RNA**→ ←TetR_N (pfam00440)
Psp-1-9	NC_007948.1	-	3565843	3565791	←PDC_hydrolase (cd01311) **RNA**→ ←Bug (pfam03401)
Aav-1-6	NC_008752.1	+	5064250	5064302	←ProA (COG0014) **RNA**→ ←HolA (COG1466)
Aav-1-7	NC_008752.1	+	3868943	3869014	←LivH (COG0559) **RNA**→ ←LivK (COG0683)
Aav-1-8	NC_008752.1	+	1006441	1006492	←PulJ (COG4795) **RNA**→ ←GSPII_IJ (pfam02501)
Aav-1-9	NC_008752.1	-	293325	293273	←hypo **RNA**→ ←Sms (COG1066)
env-6	JCVI_SCAF_1096627900356	+	1222	1274	←? (GOS4375) **RNA**→ ←? (GOS482)
Psp-1-10	NC_007948.1	-	1868152	1868101	←putative transmembrane protein **RNA**→ ←ERG3 (COG3000)
Psp-1-11	NC_007948.1	+	4685970	4686021	←Sun (COG0144) **RNA**→ ←HcaD (COG0446)
Psp-1-12	NC_007948.1	-	2218449	2218398	Bug (pfam03401)→ **RNA**→ fumarate reductase/succinate dehydrogenase flavoprotein-like→
Psp-1-13	NC_007948.1	-	3918470	3918418	P4Hc (smart00702)NOG10085 (NOG10085)→ **RNA**→ MhpC (COG0596)→
Psp-1-14	NC_007948.1	-	3119380	3119329	←CaiD (COG1024) **RNA**→ ←FadB (COG1250)3HCDH (pfam00725)
Psp-1-15	NC_007948.1	-	1086185	1086134	←DUF484 (pfam04340) **RNA**→ **RNA**→ ←DapF (COG0253)
Psp-1-16	NC_007948.1	-	1086353	1086302	(shown above)
Psp-1-17	NC_007948.1	+	4740918	4740970	←RpoN (COG1508) **RNA**→ ←YhbG (COG1137)
Psp-1-18	NC_007948.1	+	4909353	4909403	COG3118 (COG3118)→ **RNA**→ ←SAICAR_synt_Sc (cd01414)
Aav-1-10	NC_008752.1	+	305651	305728	←BaeS (COG0642) **RNA**→ ←OmpR (COG0745)
Psp-1-19	NC_007948.1	-	2444860	2444811	OYE_like_FMN (cd02933)→ **RNA**→ ←Sbp (COG1613)
Psp-1-20	NC_007948.1	-	3818209	3818158	←TrpF (COG0135) **RNA**→ ←PseudoU_synth_EcTruA (cd02570)
Psp-1-21	NC_007948.1	-	4510367	4510315	←SpeE (COG0421)NOG08837 (NOG08837) **RNA**→ ←Pcm (COG2518)

Psp-1-22	NC_007948.1	-	299862	299811	←Bug (pfam03401) **RNA**→ ←AraD (COG0235)
Cte-1-1	NZ_AAUJ01000010.1	-	158321	158270	←TauB (COG1116) **RNA**→ ←TauC (COG0600)
env-7	soil_XZS57751_b1	+	47	98	**RNA**→ ←DUF299 (pfam03618)
Psp-1-23	NC_007948.1	+	2754283	2754334	←Lpd (COG1249) **RNA**→ ←AceF (COG0508)
env-8	JCVI_SCAF_1096627888312	+	254	305	←? (GOS310) **RNA**→ ? (GOS11414712)→
env-9	soil_XZS84935_y1	-	408	358	←hypo **RNA**→ ←ABC-type amino acid transport/signal transduction systems, periplasmic component/domain
Psp-1-24	NC_007948.1	+	2502627	2502678	←COG0436 (COG0436) **RNA**→ ←HTH_1 (pfam00126)LysR_substrate (pfam03466)
Psp-1-25	NC_007948.1	+	3896820	3896871	←Glyoxalase (pfam00903) **RNA**→ ←PutA (COG1012)
env-10	JCVI_SCAF_1101667310215	+	123	174	**RNA**→ ←OmpA (COG2885)? (GOS5249)
Psp-1-26	NC_007948.1	-	4175473	4175422	ArgB (COG0548)→ **RNA**→ OmpR (COG0745)→
env-11	soil_XZS39501_b1	-	355	304	←CoxS (COG2080) **RNA**→ ←Coenzyme F390 synthetase(EC:6.2.1.30)
Psp-1-27	NC_007948.1	-	715115	715064	←hypo **RNA**→ ←hypo
env-12	soil_Contig12405	+	553	604	←CPSase_L_chain (pfam00289) **RNA**→ ←hypo
env-13	soil_Contig7686	+	65	116	**RNA**→ ←8 Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)(EC:1.1.1.100)
Aav-1-11	NC_008752.1	+	445572	445649	COG3556 (COG3556)→ **RNA**→ ←PsiF repeat protein
Asp-1-2	NC_008782.1	+	4080009	4080061	RhtB (COG1280)→ **RNA**→ protein of unknown function DUF6, transmembrane→
Aav-1-12	NC_008752.1	-	1020613	1020537	←DppD (COG0444) **RNA**→ ←DppC (COG1173)
Psp-1-28	NC_007948.1	+	754991	755043	←5-nucleotidase (pfam06189)NOG04570 (NOG04570) **RNA**→ ←GpmB (COG0406)
Aav-1-13	NC_008752.1	+	200376	200453	←COG3415 (COG3415)rve (pfam00665) **RNA**→ ←COG3181 (COG3181)
Aav-1-14	NC_008752.1	+	4536891	4536963	←TesB (COG1946) **RNA**→ ←thiolase (cd00751)
Aav-1-15	NC_008752.1	+	3771060	3771132	COG0354 (COG0354)→ **RNA**→ ←FcbC (COG0824)ElaA (COG2153)
Aav-1-16	NC_008752.1	+	839913	839990	←PaaK (COG1541) **RNA**→ ←LivF (COG0410)
Aav-1-17	NC_008752.1	+	3303992	3304069	←Chorismate_synt (pfam01264) **RNA**→ ←COG0517 (COG0517)
Psp-1-29	NC_007948.1	-	1380999	1380948	←HemN (COG0635) **RNA**→ ←COG0127 (COG0127)
env-14	JCVI_SCAF_1101667585797	+	31	79	? (GOS6292)→ **RNA**→ ? (GOS6292)→
Aav-1-18	NC_008752.1	-	136019	135968	←MinC_C (pfam03775)MinC_N (pfam05209) **RNA**→ ←PssA (COG1183)
Aav-1-19	NC_008752.1	-	4739072	4739021	DUF11 (pfam01345)→ **RNA**→ ←OmpR (COG0745)
Aav-1-20	NC_008752.1	+	289331	289407	IclR (COG1414)→ **RNA**→ ←Gst (COG0625)
Aav-1-21	NC_008752.1	+	3596677	3596754	←COG4174 (COG4174) **RNA**→ ←OppA (COG4166)
Aav-1-22	NC_008752.1	-	2551526	2551449	←PP-ATPase (cd01992) **RNA**→ ←AccA (COG0825)
Aav-1-23	NC_008752.1	+	1025874	1025950	←RpiR (COG1737) **RNA**→ ←G6PD_N (pfam00479)G6PD_C (pfam02781)
Aav-1-24	NC_008752.1	-	4952147	4952098	←COG2872 (COG2872) **RNA**→ ←Qor (COG0604)
Psp-1-30	NC_007948.1	-	3676473	3676421	←HisM (COG0765) **RNA**→ **RNA**→ **RNA**→ ←PBPb (smart00062)
Psp-1-31	NC_007948.1	-	3676570	3676518	(shown above)
Psp-1-32	NC_007948.1	-	3676667	3676615	(shown above)
Mpe-1-2	NC_008825.1	+	3690150	3690202	tRNA-Leu→ **RNA**→ tRNA-Leu→
Mpe-1-3	NC_008825.1	-	3496095	3496043	←COG1189 (COG1189) **RNA**→ ←AdoHcyase (cd00401)
Mpe-1-4	NC_008825.1	-	2782748	2782697	←COG2232 (COG2232) **RNA**→ ←methylene tetrahydromethanopterin dehydrogenase/methylenetetrahydrofolate dehydrogenas
Mpe-1-5	NC_008825.1	-	876440	876389	←PRA-CH (pfam01502) **RNA**→ ←HisF (COG0107)
Mpe-1-6	NC_008825.1	-	509303	509252	←LpxC (pfam03331) **RNA**→ ←FtsZ_type1 (cd02201)
env-15	JCVI_SCAF_1101668241578	-	314	263	←Hmp (COG1018)NapF (COG1145)? (GOS259) **RNA**→ ←? (GOS398)
Psp-1-33	NC_007948.1	-	2521468	2521417	←COG4239 (COG4239) **RNA**→ ←COG4239 (COG4239)
Mpe-1-7	NC_008825.1	-	1032969	1032918	←D-glucarate_dehydratase (cd03323) **RNA**→ ←Bug (pfam03401)
Cte-1-2	NZ_AAUJ01000017.1	-	1110	1059	←CaiC (COG0318) **RNA**→ ←Bug (pfam03401)
env-16	soil_XZS33563_y1	+	529	580	←hypo **RNA**→ ←ABC-type sugar transport system, periplasmic component
env-17	JCVI_SCAF_1096627140474	+	981	1032	←HisKA (cd00082)? (GOS336) **RNA**→ ←OmpR (COG0745)? (GOS4556)
env-18	JCVI_SCAF_1096628011002	+	1578	1631	←WcaG (COG0451)? (GOS4294) **RNA**→

15.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 15.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Burkholderiales-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

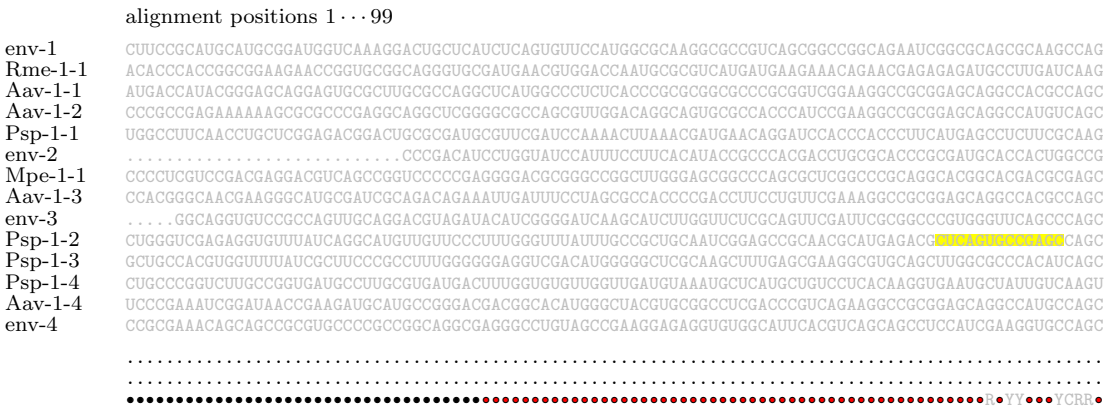
u cd00082 (1) His Kinase A (dimerization/phosphoacceptor) domain; Histidine Kinase A dimers are formed through parallel association of 2 domains creating 4-helix bundles; usually these domains contain a conserved His residue and are activated via trans-autophosphorylation by the catalytic domain of the histidine kinase, they subsequently transfer the phosphoryl group to the Asp acceptor residue of a response regulator protein.
d cd00189 (1) Tetratricopeptide repeat domain; typically contains 34 amino acids [WLF]-X(2)-[LIM]-[GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-X(2)-[ASL]-X(4)-[PKE]
d cd00401 (1) S-adenosyl-L-homocysteine hydrolase (AdoHycase) catalyzes the hydrolysis of S-adenosyl-L-homocysteine (AdoHyc) to form adenosine (Ado) and homocysteine (Hcy).
d cd00454 (1) Truncated hemoglobins (trHbs) are a family of oxygen-binding heme proteins found in cyanobacteria, eubacteria, unicellular eukaryotes, and plants.
d cd00751 (1) Thiolase are ubiquitous enzymes that catalyze the reversible thiolytic cleavage of 3-ketoacyl-CoA into acyl-CoA and acetyl-CoA, a 2-step reaction involving a covalent intermediate formed with a catalytic cysteine.
u cd01311 (1) 2-pyrone-4,6-dicarboxylic acid (PDC) hydrolase hydrolyzes PDC to yield 4-oxalomesaconic acid (OMA) or its tautomer, 4-carboxy-2-hydroxymuconic acid (CHM).
d cd01414 (1) Eukaryotic, prokaryotic and archaeal group of SAICAR synthetases represented by the Saccharomyces cerevisiae (Sc) SAICAR synthetase.
d cd01920 (1) cyclophilin_EcCYP_like: cyclophilin-type A-like peptidylprolyl cis- trans isomerase (PPIase) domain similar to the cytosolic E.
u cd01992 (1) N-terminal domain of predicted ATPase of the PP-loop faimly implicated in cell cycle control [Cell division and chromosome partitioning]
d cd02201 (1) FtsZ is a GTPase that is similar to the eukaryotic tubulins and is essential for cell division in prokaryotes.
d cd02570 (1) PseudoU_synth_EcTruA: Pseudouridine synthase, Escherichia coli TruA like.
u cd02933 (1) Old yellow enzyme (OYE)-like FMN binding domain.
u cd03322 (1) The starvation sensing protein RpsA from E.
u cd03323 (1) D-Glucarate dehydratase (GlucD) catalyzes the dehydration of both D-glucarate and L-idarate to form 5-keto-4-deoxy-D-glucarate (5-KDG) , the initial reaction of the catabolic pathway for (D)-glucarate.
u COG0014 (1) Gamma-glutamyl phosphate reductase [Amino acid transport and metabolism]
d COG0045 (1) Succinyl-CoA synthetase, beta subunit [Energy production and conversion]
u COG0050 (1) GTPases - translation elongation factors [Translation, ribosomal structure and biogenesis]
u COG0074 (1) Succinyl-CoA synthetase, alpha subunit [Energy production and conversion]
d COG0107 (1) Imidazoleglycerol-phosphate synthase [Amino acid transport and metabolism]
d COG0127 (1) Xanthosine triphosphate pyrophosphatase [Nucleotide transport and metabolism]
u COG0135 (1) Phosphoribosylanthranilate isomerase [Amino acid transport and metabolism]
u COG0144 (1) tRNA and rRNA cytosine-C5-methylases [Translation, ribosomal structure and biogenesis]
d COG0235 (1) Ribulose-5-phosphate 4-epimerase and related epimerases and aldolases [Carbohydrate transport and metabolism]
d COG0253 (1) Diaminopimelate epimerase [Amino acid transport and metabolism]
u COG0318 (1) Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II [Lipid metabolism / Secondary metabolites biosynthesis, transport, and catabolism]
u COG0354 (1) Predicted aminomethyltransferase related to GcvT [General function prediction only]
d COG0406 (1) Fructose-2,6-bisphosphatase [Carbohydrate transport and metabolism]
d COG0410 (1) ABC-type branched-chain amino acid transport systems, ATPase component [Amino acid transport and metabolism]
d COG0411 (1) ABC-type branched-chain amino acid transport systems, ATPase component [Amino acid transport and metabolism]
u COG0421 (1) Spermidine synthase [Amino acid transport and metabolism]
u COG0436 (1) Aspartate/tyrosine/aromatic aminotransferase [Amino acid transport and metabolism]
uD COG0444 (2) ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component [Amino acid transport and metabolism / Inorganic ion transport and metabolism]
d COG0446 (1) Uncharacterized NAD(FAD)-dependent dehydrogenases [General function prediction only]
u COG0451 (1) Nucleoside-diphosphate-sugar epimerases [Cell envelope biogenesis, outer membrane / Carbohydrate transport and metabolism]
d COG0480 (1) Translation elongation factors (GTPases) [Translation, ribosomal structure and biogenesis]
d COG0508 (1) Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component, and related enzymes [Energy production and conversion]
d COG0517 (1) FOG: CBS domain [General function prediction only]
u COG0548 (1) Acetylglutamate kinase [Amino acid transport and metabolism]
u COG0559 (2) Branched-chain amino acid ABC-type transport system, permease components [Amino acid transport and metabolism]
D COG0596 (1) Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily) [General function prediction only]
d COG0600 (1) ABC-type nitrate/sulfonate/bicarbonate transport system, permease component [Inorganic ion transport and metabolism]
d COG0604 (1) NADPH:quinone reductase and related Zn-dependent oxidoreductases [Energy production and conversion / General function prediction only]
d COG0625 (1) Glutathione S-transferase [Posttranslational modification, protein turnover, chaperones]
u COG0635 (1) Coproporphyrinogen III oxidase and related Fe-S oxidoreductases [Coenzyme metabolism]
u COG0642 (2) Signal transduction histidine kinase [Signal transduction mechanisms]
d COG0683 (1) ABC-type branched-chain amino acid transport systems, periplasmic component [Amino acid transport and metabolism]
u COG0704 (1) Phosphate uptake regulator [Inorganic ion transport and metabolism]
Dd COG0745 (5) Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain [Signal transduction mechanisms / Transcription]
u COG0765 (1) ABC-type amino acid transport system, permease component [Amino acid

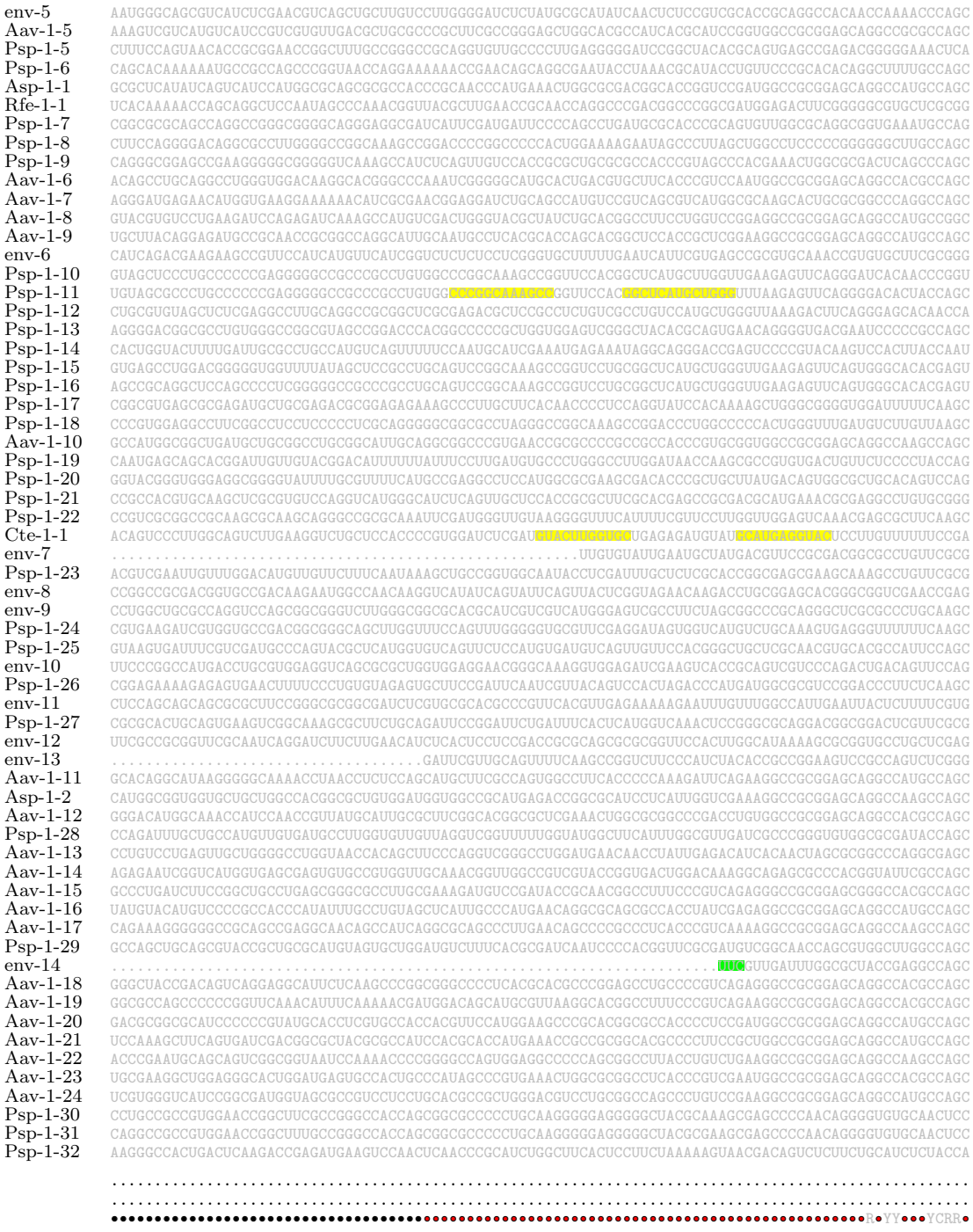
transport and metabolism]
u d COG0824 (2) Predicted thioesterase [General function prediction only]
d COG0825 (1) Acetyl-CoA carboxylase alpha subunit [Lipid metabolism]
d COG1012 (1) NAD-dependent aldehyde dehydrogenases [Energy production and conversion]
u COG1018 (1) Flavodoxin reductases (ferredoxin-NADPH reductases) family 1 [Energy production and conversion]
u d COG1024 (2) Enoyl-CoA hydratase/carnithine racemase [Lipid metabolism]
d COG1066 (1) Predicted ATP-dependent serine protease [Posttranslational modification, protein turnover, chaperones]
u COG1116 (1) ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component [Inorganic ion transport and metabolism]
d COG1117 (1) ABC-type phosphate transport system, ATPase component [Inorganic ion transport and metabolism]
d COG1137 (1) ABC-type (unclassified) transport system, ATPase component [General function prediction only]
u COG1145 (1) Ferredoxin [Energy production and conversion]
d COG1159 (1) GTPase [General function prediction only]
u d COG1173 (2) ABC-type dipeptide/oligopeptide/nickel transport systems, permease components [Amino acid transport and metabolism / Inorganic ion transport and metabolism]
d COG1176 (1) ABC-type spermidine/putrescine transport system, permease component I [Amino acid transport and metabolism]
u COG1177 (1) ABC-type spermidine/putrescine transport system, permease component II [Amino acid transport and metabolism]
d COG1183 (1) Phosphatidylserine synthase [Lipid metabolism]
u COG1189 (1) Predicted rRNA methylase [Translation, ribosomal structure and biogenesis]
u COG1249 (1) Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component, and related enzymes [Energy production and conversion]
d COG1250 (1) 3-hydroxyacyl-CoA dehydrogenase [Lipid metabolism]
u COG1280 (1) Putative threonine efflux protein [Amino acid transport and metabolism]
u COG1381 (1) Recombinational DNA repair protein (RecF pathway) [DNA replication, recombination, and repair]
u COG1414 (1) Transcriptional regulator [Transcription]
d COG1466 (1) DNA polymerase III, delta subunit [DNA replication, recombination, and repair]
u COG1508 (1) DNA-directed RNA polymerase specialized sigma subunit, sigma54 homolog [Transcription]
u COG1541 (1) Coenzyme F390 synthetase [Coenzyme metabolism]
d COG1580 (1) Flagellar basal body-associated protein [Cell motility and secretion]
d COG1613 (1) ABC-type sulfate transport system, periplasmic component [Inorganic ion transport and metabolism]
u COG1737 (1) Transcriptional regulators [Transcription]
d COG1840 (1) ABC-type Fe3+ transport system, periplasmic component [Inorganic ion transport and metabolism]
u COG1868 (1) Flagellar motor switch protein [Cell motility and secretion]
u COG1946 (1) Acyl-CoA thioesterase [Lipid metabolism]
d COG1969 (1) Ni,Fe-hydrogenase I cytochrome b subunit [Energy production and conversion]
u COG2080 (1) Aerobic-type carbon monoxide dehydrogenase, small subunit CoxS/CutS homologs [Energy production and conversion]
d COG2153 (1) Predicted acyltransferase [General function prediction only]
d COG2223 (1) Nitrate/nitrite transporter [Inorganic ion transport and metabolism]
u COG2232 (1) Predicted ATP-dependent carboligase related to biotin carboxylase [General function prediction only]
d COG2518 (1) Protein-L-isoaspartate carboxylmethyltransferase [Posttranslational modification, protein turnover, chaperones]
u COG2872 (1) Predicted metal-dependent hydrolases related to alanyl-tRNA synthetase HxxxH domain [General function prediction only]
d COG2885 (1) Outer membrane protein and related peptidoglycan-associated (lipo)proteins [Cell envelope biogenesis, outer membrane]
u COG2908 (1) Uncharacterized protein conserved in bacteria [Function unknown]
d COG3000 (1) Sterol desaturase [Lipid metabolism]
u COG3093 (1) Plasmid maintenance system antidote protein [General function prediction only]
u COG3118 (1) Thioredoxin domain-containing protein [Posttranslational modification, protein turnover, chaperones]
d COG3181 (1) Uncharacterized protein conserved in bacteria [Function unknown]
u COG3415 (1) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
u COG3556 (1) Predicted membrane protein [Function unknown]
u COG3842 (1) ABC-type spermidine/putrescine transport systems, ATPase components [Amino acid transport and metabolism]
u COG4133 (1) ABC-type transport system involved in cytochrome c biogenesis, ATPase component [Posttranslational modification, protein turnover, chaperones]
d COG4166 (1) ABC-type oligopeptide transport system, periplasmic component [Amino acid transport and metabolism]
u COG4174 (1) ABC-type uncharacterized transport system, permease component [General function prediction only]
u d COG4239 (4) ABC-type uncharacterized transport system, permease component [General function prediction only]
u COG4795 (1) Type II secretory pathway, component PulJ [Intracellular trafficking and secretion]
d COG4976 (1) Predicted methyltransferase (contains TPR repeat) [General function prediction only]
u COG5013 (1) Nitrate reductase alpha subunit [Energy production and conversion]
D GOS11414712 (1) ?
u GOS1277 (1) ?
d GOS13617 (1) ?
u GOS259 (1) ?
u GOS310 (1) ?
u GOS336 (1) ?
d GOS381 (1) ?
u GOS3923 (1) ?
d GOS398 (1) ?
u GOS4294 (1) ?
u GOS4375 (1) ?
d GOS4556 (1) ?
d GOS482 (1) ?
d GOS5249 (1) ?
d GOS6036 (1) ?

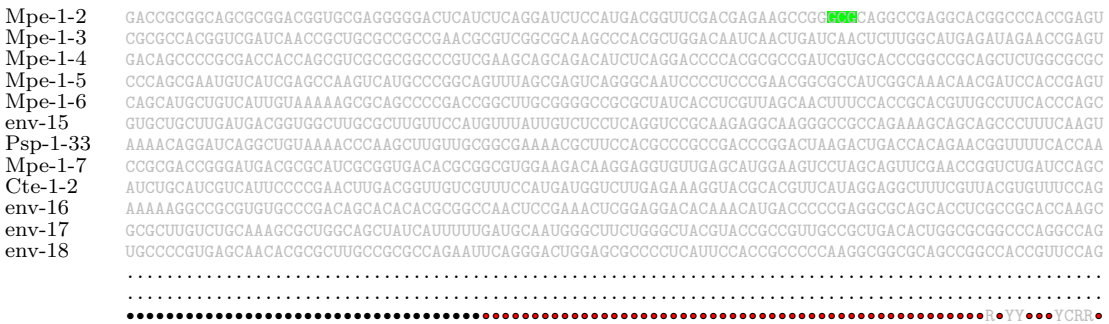
uD GOS6292 (2) ?
u NOG04570 (1)
u NOG08837 (1)
u NOG10085 (1)
d pfam00126 (1) Bacterial regulatory helix-turn-helix protein, lysR family.
u pfam00289 (1) Carbamoyl-phosphate synthase L chain, N-terminal domain.
d pfam00440 (1) Bacterial regulatory proteins, tetR family.
d pfam00479 (1) Glucose-6-phosphate dehydrogenase, NAD binding domain.
u pfam00665 (1) Integrase core domain.
d pfam00725 (1) 3-hydroxyacyl-CoA dehydrogenase, C-terminal domain.
u pfam00903 (1) Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily.
u pfam01264 (1) Chorismate synthase.
u pfam01345 (1) Domain of unknown function DUF11.
u pfam01502 (1) Phosphoribosyl-AMP cyclohydrolase.
d pfam02501 (1) Bacterial type II secretion system protein I/J.
u pfam02550 (1) Acetyl-CoA hydrolase/transferase.
d pfam02781 (1) Glucose-6-phosphate dehydrogenase, C-terminal domain.
u pfam03331 (1) UDP-3-O-acyl N-acetylglycosamine deacetylase.
u d pfam03401 (6) Bordetella uptake gene (bug) product.
d pfam03466 (1) LysR substrate binding domain.
d pfam03618 (1) Domain of unknown function (DUF299).
u pfam03775 (1) Septum formation inhibitor MinC, C-terminal domain.
u pfam04292 (1) D-galactarate dehydratase / Altronate hydrolase, N terminus.
u pfam04340 (1) Protein of unknown function, DUF484.
u pfam05209 (1) Septum formation inhibitor MinC, N-terminal domain.
u pfam06189 (1) 5'-nucleotidase.
d pfam06243 (1) Phenylacetic acid degradation B.
d smart00062 (1) Bacterial periplasmic substrate-binding proteins; bacterial proteins, eukaryotic ones are in PBPe
u smart00283 (1) Methyl-accepting chemotaxis-like domains (chemotaxis sensory transducer).
u smart00702 (1) Prolyl 4-hydroxylase alpha subunit homologues.

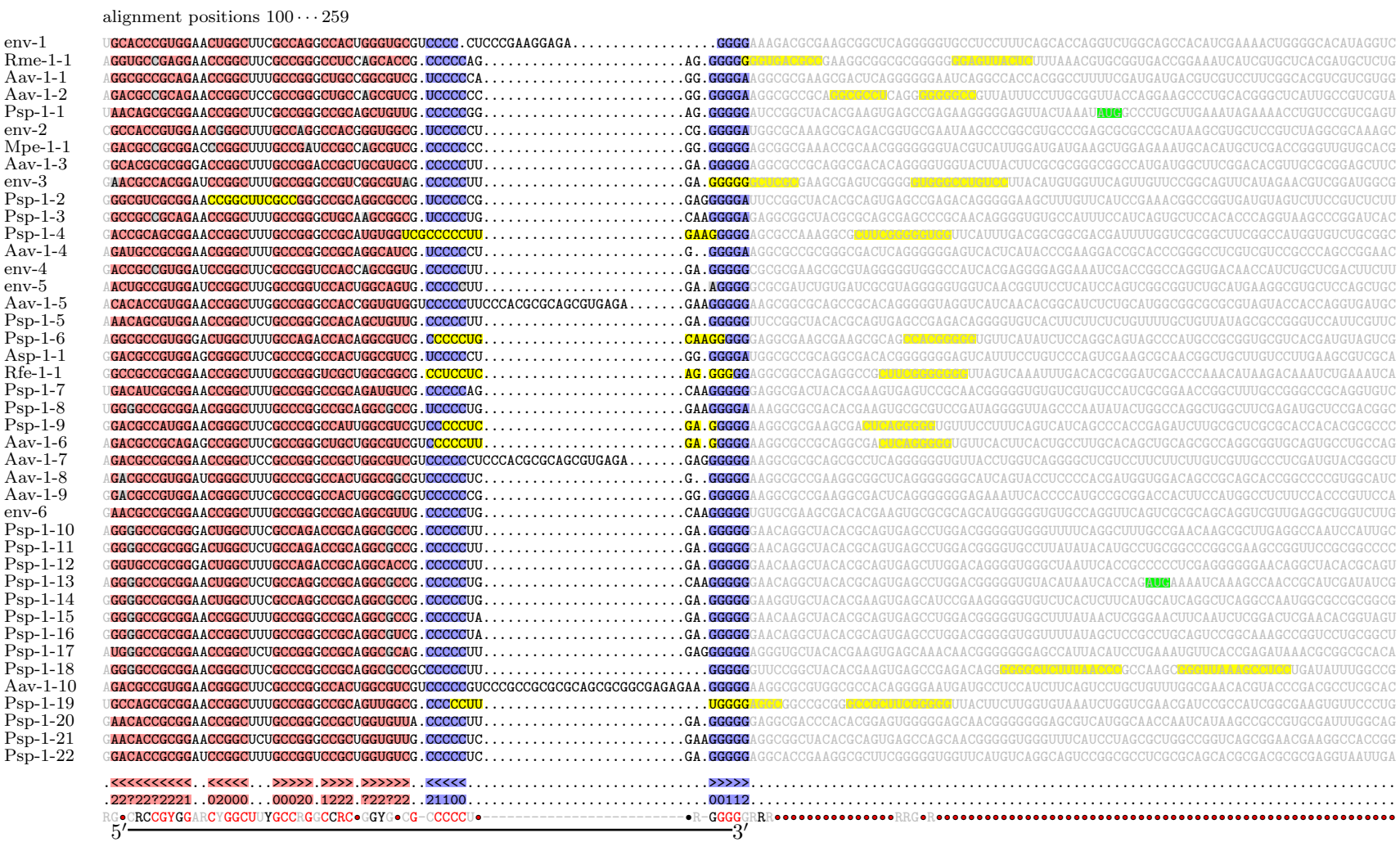
15.4 Multiple-sequence alignment

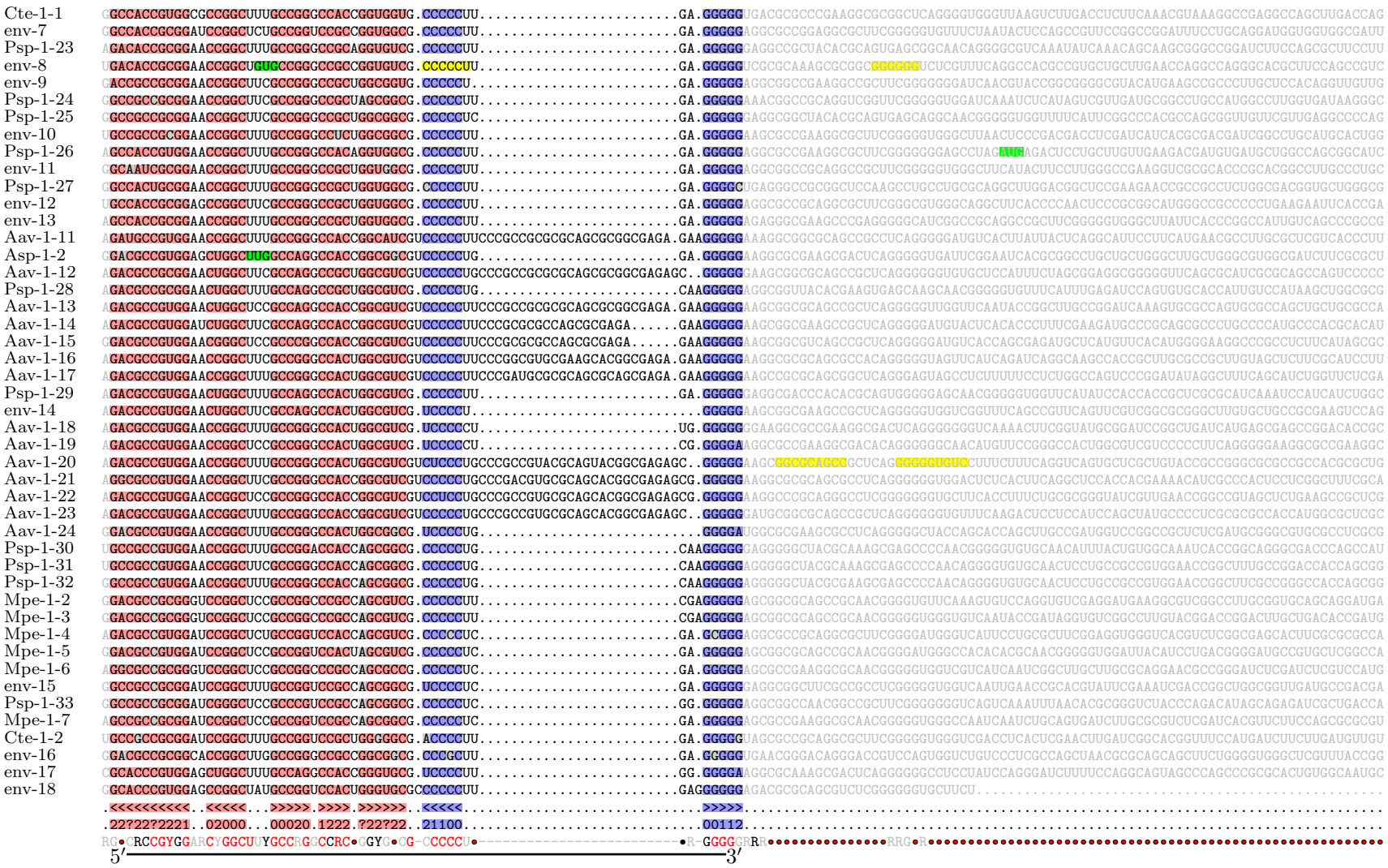
Each Burkholderiales-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 15.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Burkholderiales-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



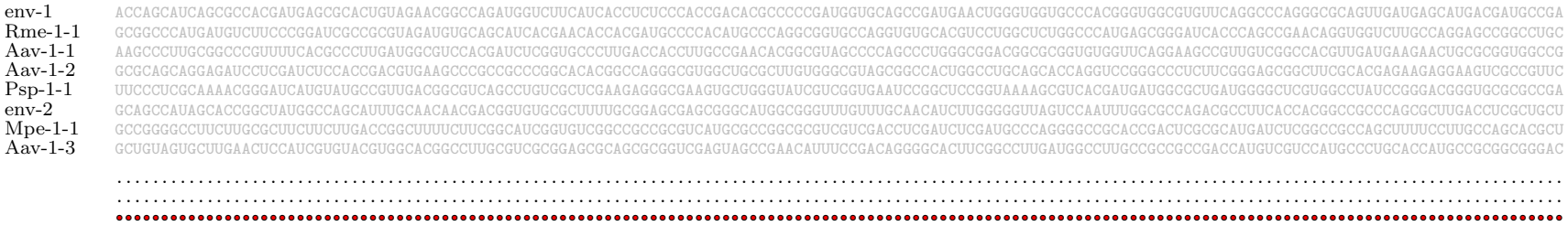

env-5
Aav-1-5
Psp-1-5
Psp-1-6
Asp-1-1
Rfe-1-1
Psp-1-7
Psp-1-8
Psp-1-9
Aav-1-6
Aav-1-7
Aav-1-8
Aav-1-9
env-6
Psp-1-10
Psp-1-11
Psp-1-12
Psp-1-13
Psp-1-14
Psp-1-15
Psp-1-16
Psp-1-17
Psp-1-18
Aav-1-10
Psp-1-19
Psp-1-20
Psp-1-21
Psp-1-22
Cte-1-1
env-7
Psp-1-23
env-8
env-9
Psp-1-24
Psp-1-25
env-10
Psp-1-26
env-11
Psp-1-27
env-12
env-13
Aav-1-11
Asp-1-2
Aav-1-12
Psp-1-28
Aav-1-13
Aav-1-14
Aav-1-15
Aav-1-16
Aav-1-17
Psp-1-29
env-14
Aav-1-18
Aav-1-19
Aav-1-20
Aav-1-21
Aav-1-22
Aav-1-23
Aav-1-24
Psp-1-30
Psp-1-31
Psp-1-32


Mpe-1-2
Mpe-1-3
Mpe-1-4
Mpe-1-5
Mpe-1-6
env-15
Psp-1-33
Mpe-1-7
Cte-1-2
env-16
env-17
env-18


alignment positions 100 · · · 259
env-1
Rme-1-1
Aav-1-1
Aav-1-2
Psp-1-1
env-2
Mpe-1-1
Aav-1-3
env-3
Psp-1-2
Psp-1-3
Psp-1-4
Aav-1-4
env-4
env-5
Aav-1-5
Psp-1-5
Psp-1-6
Asp-1-1
Rfe-1-1
Psp-1-7
Psp-1-8
Psp-1-9
Aav-1-6
Aav-1-7
Aav-1-8
Aav-1-9
env-6
Psp-1-10
Psp-1-11
Psp-1-12
Psp-1-13
Psp-1-14
Psp-1-15
Psp-1-16
Psp-1-17
Psp-1-18
Aav-1-10
Psp-1-19
Psp-1-20
Psp-1-21
Psp-1-22
5′
3′

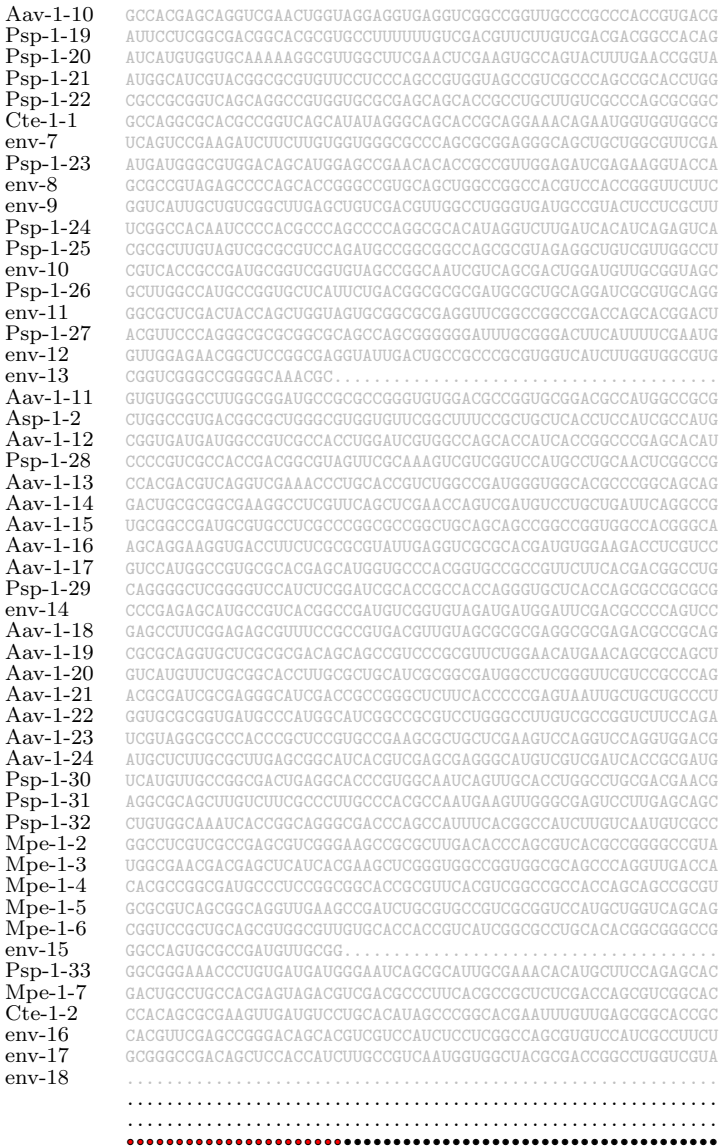


alignment positions 260 · · · 419



env-3
Psp-1-2
Psp-1-3
Psp-1-4
Aav-1-4
env-4
env-5
Aav-1-5
Psp-1-5
Psp-1-6
Asp-1-1
Rfe-1-1
Psp-1-7
Psp-1-8
Psp-1-9
Aav-1-6
Aav-1-7
Aav-1-8
Aav-1-9
env-6
Psp-1-10
Psp-1-11
Psp-1-12
Psp-1-13
Psp-1-14
Psp-1-15
Psp-1-16
Psp-1-17
Psp-1-18
Aav-1-10
Psp-1-19
Psp-1-20
Psp-1-21
Psp-1-22
Cte-1-1
env-7
Psp-1-23
env-8
env-9
Psp-1-24
Psp-1-25
env-10
Psp-1-26
env-11
Psp-1-27
env-12
env-13
Aav-1-11
Asp-1-2
Aav-1-12
Psp-1-28
Aav-1-13
Aav-1-14
Aav-1-15
Aav-1-16
Aav-1-17
Psp-1-29
env-14
Aav-1-18
Aav-1-19
Aav-1-20
Aav-1-21

alignment positions $420 \cdots 479$



16 c4 antisense RNA

16.1 Taxa

The taxonomy of each organism containing a putative c4 antisense RNA RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual c4 antisense RNA RNA in Sections 16.2 and 16.4.

abbrev. of hits	**taxonomy of species**
Bce-1-1 to Bce-1-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* MC0-3
Bce-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* PC184
Bdo-1-1 to Bdo-1-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia dolosa* AUO158
Bmu-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia multivorans* ATCC 17616
Bsp-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia sp.* 383
Bvi-1-1 to Bvi-1-3	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia vietnamiensis* G4
Bma-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* 2002721280
Bma-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* ATCC 10399
Bma-3-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* ATCC 23344
Bma-4-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* FMH
Bma-5-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* GB8 horse 4
Bma-6-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* JHU
Bma-7-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* NCTC 10229
Bma-8-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* NCTC 10247
Bma-9-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* PRL-20
Bma-10-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* SAVP1
Bph-1-1 to Bph-1-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia phytofirmans* PsJN
Bxe-1-1 to Bxe-1-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia xenovorans* LB400
Bok-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia oklahomensis* C6786
Bok-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia oklahomensis* EO147
Bps-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1106a
Bps-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1106b
Bps-3-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 112
Bps-4-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 14
Bps-5-1 to Bps-5-3	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1655
Bps-6-1 to Bps-6-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1710a
Bps-7-1 to Bps-7-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1710b
Bps-8-1 to Bps-8-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 305
Bps-9-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 406e
Bps-10-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 668
Bps-11-1 to Bps-11-3	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 7894
Bps-12-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 9
Bps-13-1 to Bps-13-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 91
Bps-14-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* B7210
Bps-15-1 to Bps-15-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* BCC215
Bps-16-1 to Bps-16-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* K96243
Bps-17-1 to Bps-17-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* Pasteur 52237
Bps-18-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* S13
Bth-1-1 to Bth-1-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia thailandensis* Bt4
Bth-2-1 to Bth-2-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia thailandensis* MSMB43
Bth-3-1 to Bth-3-3	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia thailandensis* TXDOH
Reu-1-1 to Reu-1-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Cupriavidus *Ralstonia eutropha* H16
Reu-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Cupriavidus *Ralstonia eutropha* JMP134
Rpi-1-1 to Rpi-1-6	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Ralstonia pickettii* 12D
Rpi-2-1 to Rpi-2-5	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Ralstonia pickettii* 12J
Rso-1-1 to Rso-1-5	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Ralstonia solanacearum* GMI1000
Rso-2-1 to Rso-2-3	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Ralstonia solanacearum* UW551
Cte-1-1 to Cte-1-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae *Comamonas testosteroni* KF-1
Dac-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae *Delftia acidovorans* SPH-1
Pna-1-1 to Pna-1-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae *Polaromonas naphthalenivorans* CJ2
Cvi-1-1 to Cvi-1-2	Bacteria Proteobacteria Betaproteobacteria Neisseriales Neisseriaceae *Chromobacterium violaceum* ATCC 12472
Nmu-1-1	Bacteria Proteobacteria Betaproteobacteria Nitrosomonadales Nitrosomonadaceae *Nitrosospira multiformis* ATCC 25196
Dps-1-1 to Dps-1-2	Bacteria Proteobacteria Deltaproteobacteria Desulfobacterales Desulfobulbaceae *Desulfotalea psychrophila* LSv54
Asa-1-1	Bacteria Proteobacteria Gammaproteobacteria Aeromonadales Aeromonadaceae *Aeromonas salmonicida subsp. salmonicida* A449
Sde-1-1	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Alteromonadaceae *Saccharophagus degradans* 2-40
Sba-1-1	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella baltica* OS185
Sba-2-1	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella baltica* OS195
Sfr-1-1	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella frigidimarina* NCIMB 400

Dno-1-1 to Dno-1-4	Bacteria Proteobacteria Gammaproteobacteria Cardiobacteriales Cardiobacteriaceae *Dichelobacter nodosus* VCS1703A
Esa-1-1 to Esa-1-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Cronobacter *Enterobacter sakazakii* ATCC BAA-894
Esp-1-1 to Esp-1-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Enterobacter sp.* 638
Eco-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli*
Eco-2-1 to Eco-2-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* 101-1
Eco-3-1 to Eco-3-14	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* 53638
Eco-4-1 to Eco-4-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* APEC O1
Eco-5-1 to Eco-5-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* B
Eco-6-1 to Eco-6-4	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* B171
Eco-7-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* B7A
Eco-8-1 to Eco-8-5	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* E110019
Eco-9-1 to Eco-9-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* E22
Eco-10-1 to Eco-10-4	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* E24377A
Eco-11-1 to Eco-11-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* F11
Eco-12-1 to Eco-12-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 EDL933
Eco-13-1 to Eco-13-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. Sakai
Eco-14-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* SECEC SMS-3-5
Eco-15-1 to Eco-15-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* UTI89
Plu-1-1 to Plu-1-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Photorhabdus luminescens subsp. laumondii* TTO1
Sen-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Salmonella enterica subsp. enterica serovar* Paratyphi A str. ATCC
Sen-2-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Dublin *Salmonella enterica subsp. en*
Sen-3-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Hadar *Salmonella enterica subsp. ent*
Sen-4-1 to Sen-4-5	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Heidelberg *Salmonella enterica subsp.*
Sen-5-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Heidelberg *Salmonella enterica subsp.*
Sen-6-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Kentucky *Salmonella enterica subsp.*
Sen-7-1 to Sen-7-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Newport *Salmonella enterica subsp.* e
Sen-8-1 to Sen-8-4	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Newport *Salmonella enterica subsp.* e
Sen-9-1 to Sen-9-4	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Saintpaul *Salmonella enterica subsp.*
Sen-10-1 to Sen-10-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ent*
Sen-11-1 to Sen-11-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ent*
Sen-12-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Virchow *Salmonella enterica subsp. e*
Sty-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Salmonella typhimurium*
Sty-2-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Salmonella typhimurium* LT2
Spr-1-1 to Spr-1-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Serratia proteamaculans* 568
Sbo-1-1 to Sbo-1-4	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella boydii* BS512
Sbo-2-1 to Sbo-2-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella boydii* Sb227
Sdy-1-1 to Sdy-1-4	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella dysenteriae* 1012
Sdy-2-1 to Sdy-2-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella dysenteriae* Sd197
Sfl-1-1 to Sfl-1-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella flexneri* 2a str. 2457T
Sfl-2-1 to Sfl-2-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella flexneri* 2a str. 301
Sfl-3-1 to Sfl-3-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella flexneri* 5 str. 8401
Sso-1-1 to Sso-1-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella sonnei* Ss046
Sgl-1-1 to Sgl-1-4	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Sodalis glossinidius*
Sgl-2-1 to Sgl-2-18	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Sodalis glossinidius str. 'morsitans'*
Ybe-1-1 to Ybe-1-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia bercovieri* ATCC 43970
Yen-1-1 to Yen-1-4	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia enterocolitica subsp. enterocolitica* 8081
Yfr-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia frederiksenii* ATCC 33641
Yin-1-1 to Yin-1-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia intermedia* ATCC 29909
Ymo-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia mollaretii* ATCC 43969
Ype-1-1 to Ype-1-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* Angola
Ype-2-1 to Ype-2-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* Antiqua
Ype-9-1 to Ype-9-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Antiqua str. B42003004
Ype-10-1 to Ype-10-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Antiqua str. E1979001
Ype-11-1 to Ype-11-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Antiqua str. UG05-0454
Ype-12-1 to Ype-12-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Mediaevalis str. K1973002
Ype-13-1 to Ype-13-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Microtus str. 91001
Ype-14-1 to Ype-14-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Orientalis str. F1991016
Ype-15-1 to Ype-15-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Orientalis str. IP275
Ype-16-1 to Ype-16-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Orientalis str. MG05-1020
Ype-3-1 to Ype-3-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* CA88-4125
Ype-4-1 to Ype-4-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* CO92
Ype-5-1 to Ype-5-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* FV-1
Ype-6-1 to Ype-6-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* KIM
Ype-7-1 to Ype-7-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* Nepal516
Ype-8-1 to Ype-8-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* Pestoides F

Yps-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pseudotuberculosis* IP 31758
Yps-2-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pseudotuberculosis* IP 32953
Msp-1-1 to Msp-1-2	Bacteria Proteobacteria Gammaproteobacteria Oceanospirillales *Marinomonas sp.* MWYL1
Hin-1-1	Bacteria Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae *Haemophilus influenzae* 22.1-21
Hin-2-1	Bacteria Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae *Haemophilus influenzae* 22.4-21
Hin-3-1 to Hin-3-3	Bacteria Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae *Haemophilus influenzae* 3655
Hin-4-1	Bacteria Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae *Haemophilus influenzae* 86-028NP
Hin-5-1 to Hin-5-2	Bacteria Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae *Haemophilus influenzae* PittAA
Hin-6-1	Bacteria Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae *Haemophilus influenzae* PittEE
Hin-7-1 to Hin-7-2	Bacteria Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae *Haemophilus influenzae* PittGG
Hin-8-1	Bacteria Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae *Haemophilus influenzae* PittHH
Hin-9-1	Bacteria Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae *Haemophilus influenzae* PittII
Hin-10-1	Bacteria Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae *Haemophilus influenzae* R2846
Hin-11-1	Bacteria Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae *Haemophilus influenzae* R2866
Hin-12-1	Bacteria Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae *Haemophilus influenzae* R3021
Hso-1-1 to Hso-1-2	Bacteria Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae Histophilus *Haemophilus somnus* 129PT
Hso-2-1	Bacteria Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae Histophilus *Haemophilus somnus* 2336
Mha-1-1 to Mha-1-2	Bacteria Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae *Mannheimia haemolytica* PHL213
Pmu-1-1	Bacteria Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae *Pasteurella multocida subsp. multocida str.* Pm70
Aba-1-1 to Aba-1-24	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Moraxellaceae *Acinetobacter baumannii* ATCC 17978
Asp-1-1 to Asp-1-4	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Moraxellaceae *Acinetobacter sp.* ADP1
Avi-1-1 to Avi-1-3	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Azotobacter vinelandii*
Pae-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* 2192
Pae-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* C3719
Pae-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PA7
Pae-4-1 to Pae-4-2	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PACS2
Pae-5-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PAO1
Pae-6-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* UCBPP-PA14
Pen-1-1 to Pen-1-30	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas entomophila* L48
Pfl-1-1 to Pfl-1-9	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf-5
Pfl-2-1 to Pfl-2-29	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf0-1
Pme-1-1 to Pme-1-2	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas mendocina ymp*
Ppu-1-1 to Ppu-1-10	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* F1
Ppu-2-1 to Ppu-2-5	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* GB-1
Ppu-3-1 to Ppu-3-13	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* KT2440
Ppu-4-1 to Ppu-4-27	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* W619
Pst-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas stutzeri* A1501
Psy-1-1 to Psy-1-6	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. phaseolicola* 1448A
Psy-2-1 to Psy-2-5	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1 to Psy-3-3	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000
Vba-1-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales unclassified Vibrionales *Vibrionales bacterium* SWAT-3
Vch-1-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Vibrio cholerae* MZO-3
Vch-2-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Vibrio cholerae* NCTC 8457
Vch-3-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Vibrio cholerae* O395
Vfi-1-1 to Vfi-1-2	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Vibrio fischeri* ES114
Vsh-1-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Vibrio shilonii* AK1
Vsp-1-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Vibrio splendidus* 12B01
Sma-1-1 to Sma-1-5	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Stenotrophomonas maltophilia* R551-3
Xax-1-1	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas axonopodis pv. citri str.* 306
Xca-1-1	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas campestris pv. campestris str.* ATCC 33913
Xor-1-1	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas oryzae pv. oryzae* KACC10331
Xor-2-1	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas oryzae pv. oryzae* MAFF 311018
Xfa-1-1 to Xfa-1-2	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xylella fastidiosa* 9a5c
Xfa-2-1	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xylella fastidiosa* Ann-1
Xfa-3-1	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xylella fastidiosa* Dixon
Sen-13-1 to Sen-13-2	CVM23701Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar 4,[5],12:i:- *Salmonella ente*
Sen-14-1 to Sen-14-2	GA_MM04042433Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Salmonella enterica subsp. enterica serovar* Javiana str
Sen-15-1 to Sen-15-3	HI_N05-537Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Weltevreden *Salmonella e*
Sen-16-1 to Sen-16-2	SC-B67Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Choleraesuis *Salmonella enter*
Eph-2-1	Viruses dsDNA viruses, no RNA stage Caudovirales Myoviridae *Enterobacteria phage phi*P27
Eph-1-1	Viruses dsDNA viruses, no RNA stage Caudovirales Myoviridae *Enterobacteria phage* SfV
Hph-1-1	Viruses dsDNA viruses, no RNA stage Caudovirales Myoviridae *Haemophilus phage* Aaphi23
Eph-3-1	Viruses dsDNA viruses, no RNA stage Caudovirales Myoviridae P1-like viruses *Enterobacteria phage* P1
Eph-4-1	Viruses dsDNA viruses, no RNA stage Caudovirales Myoviridae P2-like viruses *Enterobacteria phage* P4
Api-1-1	Viruses dsDNA viruses, no RNA stage Caudovirales Podoviridae *Acyrthosiphon pisum secondary endosymbiont phage* 1

Eph-5-1	Viruses dsDNA viruses, no RNA stage Caudovirales Podoviridae *Enterobacteria phage* VT2-Sakai
Eph-6-1	Viruses dsDNA viruses, no RNA stage Caudovirales Podoviridae P22-like viruses *Enterobacteria phage phi*V10
Sph-1-1	Viruses dsDNA viruses, no RNA stage Caudovirales *Salmonella phage* ST64B
Kph-1-1	Viruses dsDNA viruses, no RNA stage Caudovirales Siphoviridae *Klebsiella phage phi*KO2
Pcd-1-1	Viruses dsDNA viruses, no RNA stage Caudovirales Siphoviridae Lambda-like viruses unclassified Lambda-like viruses *Phage cdt*I (Bacteriophage cdtI)
Sco-1-1	Viruses dsDNA viruses, no RNA stage Caudovirales Siphoviridae Lambda-like viruses unclassified Lambda-like viruses *Stx*2 converting phage II (Stx2 converting bacteriophage I
Eph-7-1	Viruses dsDNA viruses, no RNA stage Caudovirales Siphoviridae N15-like viruses *Enterobacteria phage* N15
Yph-1-1	Viruses dsDNA viruses, no RNA stage Caudovirales Siphoviridae *Yersinia phage* PY54
PGi-1-1	Viruses *Phage* Gifsy-1
env-1 to env-52	environmental samples

16.2 Gene contexts

Each c4 antisense RNA RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one c4 antisense RNA RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 16.3. The accession number of the sequence containing each c4 antisense RNA RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 16.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Aba-1-1	NC_009085.1	+	1492888	1492976	EriC (COG0038)→ **RNA**→ putative transcriptional regulator→
Aba-1-2	NC_009085.1	-	2365091	2365008	hypo→ **RNA**→ hypo→
Hin-3-1	NZ_AAZF01000005.1	+	72866	72949	hypo→ **RNA**→ hypo→ hypo→ hypo→ hypo→ hypo→
Hin-5-1	NZ_AAZG01000002.1	+	114094	114177	hypo→ **RNA**→ hypo→ hypo→ hypo→ hypo→ hypo→
Hin-7-1	NC_009567.1	-	760122	760036	hypo→ **RNA**→ hypo→ hypo→ hypo→ COG5519 (COG5519)→
Hso-1-1	NC_008309.1	+	565968	566054	hypo→ **RNA**→ hypo→ hypo→ hypo→
Hso-2-1	NZ_AACJ01000013.1	-	2187	2101	Transposase_2 (pfam01385)Transposase_35 (pfam07282)→ **RNA**→ hypo→
Pmu-1-1	NC_002663.1	+	2002795	2002883	hypo→ **RNA**→ hypo→ hypo→ hypo→ hypo→ COG5519 (COG5519)→
Hin-3-2	NZ_AAZF01000002.1	-	232050	231964	hypo→ **RNA**→ hypo→ hypo→ hypo→ hypo→ hypo→ hypo→ Pox_D5 (pfam03288)→
Hin-2-1	NZ_AAZJ01000012.1	+	9811	9897	hypo→ **RNA**→ hypo→ hypo→ hypo→ hypo→ hypo→ COG5519 (COG5519)→
Mha-1-1	NZ_AASA01000084.1	+	6161	6256	**RNA**→
Mha-1-2	NZ_AASA01000084.1	+	6623	6711	**RNA**→
Pfl-2-1	NC_007492.1	+	3001178	3001267	←hypo **RNA**→ hypo→
env-1	JCVI_SCAF_1096626333345	-	508	421	**RNA**→
Dps-1-1	NC_006139.1	+	46573	46654	←related to TraJ protein **RNA**→ hypo→
Dac-1-1	NZ_AAVD01000003.1	-	489428	489345	Twin-arginine translocation pathway signal→ **RNA**→ hypo→
Sgl-2-1	NC_007712.1	-	4059279	4059192	INT_P4 (cd00801)→ **RNA**→ hypo→
Pfl-2-2	NC_007492.1	-	317345	317260	hypo→ **RNA**→ hypo→
Sbo-1-1	NZ_AAKA01000001.1	-	39540	39453	hypo→ **RNA**→ hypo→
Pna-1-1	NC_008781.1	-	380216	380131	RVT (pfam00078)→ **RNA**→ hypo→
Pna-1-2	NC_008781.1	-	1018953	1018867	hypo→ **RNA**→ hypo→ hypo→ hypo→ COG4643 (COG4643)→ hypo→
Xfa-1-1	NC_002488.3	-	2387378	2387297	←hypo **RNA**→ COG3617 (COG3617)→
Sma-1-1	NZ_AAVZ01000004.1	-	59163	59071	HTH_1 (pfam00126)LysR_substrate (pfam03466)→ **RNA**→ COG4487 (COG4487)→ SIR2-like (cd01406)→ protein of unknown function DUF87→
Dno-1-1	NC_009446.1	-	1080812	1080727	INT_P4 (cd00801)→ **RNA**→ virulence-associated protein VapG3→ VirE (pfam05272)→
Dno-1-2	NC_009446.1	+	824732	824817	DnaC (COG1484)→ **RNA**→ RusA (pfam05866)→ hypo→ hypo→ hypo→
Dno-1-3	NC_009446.1	-	299743	299661	virulence-associated protein VapH→ **RNA**→ virulence-associated protein VapH→ virulence-associated protein VapG1→ VirE (pfam05272)→

Dno-1-4	NC_009446.1	-	292912	292832	←VapI (COG3093) **RNA**→ virulence-associated protein VapG2→
Esp-1-1	NC_009436.1	-	1205701	1205615	FepA (COG4771)→ **RNA**→ hypo→
Plu-1-1	NC_005126.1	+	4085959	4086042	hypo→ **RNA**→ hypo→ Nlp (COG3423)→
Xax-1-1	NC_003921.3	-	1645	1553	hypo→ **RNA**→ hypo→
Aba-1-3	NC_009085.1	-	2570672	2570588	←RimL (COG1670) **RNA**→ putative transcriptional regulator→
Xfa-3-1	NZ_AAAL02000013.1	+	8678	8770	COG3617 (COG3617)→ **RNA**→ COG3617 (COG3617)→ Ssb (COG0629)→
Xfa-2-1	NZ_AAAM03000068.1	+	11383	11475	COG3617 (COG3617)→ **RNA**→ COG3617 (COG3617)→ Ssb (COG0629)→ hypo (cons)→ INT_Lambda_C (cd00800)→
Xfa-1-2	NC_002488.3	-	1588569	1588477	LpxD (COG1044)→ **RNA**→ COG3617 (COG3617)→ Ssb (COG0629)→ hypo→ INT_Lambda_C (cd00800)→
Sma-1-2	NZ_AAVZ01000007.1	+	44241	44329	←hypo **RNA**→ hypo→
env-2	JCVI_SCAF_1096628066130	+	205	288	**RNA**→
env-3	JCVI_SCAF_1096627028318	+	1583	1668	←MoaA (COG2896)? (GOS255) **RNA**→
Sma-1-3	NZ_AAVZ01000011.1	+	25481	25567	←hypo **RNA**→ hypo→
Sma-1-4	NZ_AAVZ01000001.1	+	108170	108255	COG4700 (COG4700)→ **RNA**→ hypo→ hypo→
Plu-1-2	NC_005126.1	+	989054	989139	←hypo **RNA**→ Nlp (COG3423)→
env-4	sludgeOz_Contig11344	+	5032	5114	(,,,)→ **RNA**→ ←(,,,)
Aba-1-4	NC_009085.1	+	871749	871832	Flavin_Reduct (pfam01613)→ **RNA**→ TetR_N (pfam00440)→
Bph-1-1	NZ_AAUH01000032.1	+	20757	20848	COG1981 (COG1981)→ **RNA**→ hypo (cons)→
Bxe-1-1	NC_007951.1	-	4245368	4245279	COG1981 (COG1981)→ **RNA**→ hypo→
env-5	JCVI_SCAF_1096627192337	-	13368	13279	←COG0702 (COG0702)? (GOS4294) **RNA**→ ? (GOS11386653)→
env-6	JCVI_SCAF_1101668744231	-	1049	962	**RNA**→ Nlp (COG3423)? (GOS11363276)→ RimL (COG1670)→
Spr-1-1	NZ_AAUN01000013.1	+	38258	38343	Glyco_hydro_1 (pfam00232)→ **RNA**→ DNA_methylase (pfam00145)→ Vsr (cd00221)→
Pfl-2-3	NC_007492.1	+	523624	523710	hypo→ **RNA**→ hypo→
env-7	JCVI_SCAF_1096628137543	+	1077	1163	? (GOS1771)→ **RNA**→ COG3617 (COG3617)? (GOS11900477)→
Pfl-2-4	NC_007492.1	+	965543	965629	←SBP_bac_5 (pfam00496) **RNA**→ hypo→
Msp-1-1	NC_009654.1	-	652730	652647	←hypo **RNA**→ hypo→ hypo→
Msp-1-2	NC_009654.1	-	619642	619558	←hypo **RNA**→ hypo→
env-8	sludgeOz_Contig11395	-	8077	7988	COG3645 (COG3645)COG3646 (COG3646)→ **RNA**→ (,,,)→ (,,,)→
Aba-1-5	NC_009085.1	+	871633	871716	Flavin_Reduct (pfam01613)→ **RNA**→ TetR_N (pfam00440)→
Pfl-1-1	NC_004129.6	+	5946544	5946629	Acetyltransf_1 (pfam00583)→ **RNA**→ hypo→
Psy-1-1	NC_005773.3	-	1124857	1124771	hypo→ **RNA**→ putative inner membrane protein→
Pae-5-1	NC_002516.2	-	1348513	1348430	←hypo **RNA**→ hypo→
Pae-2-1	NZ_AAKV01000003.1	-	6640	6557	←hypo **RNA**→ hypo→
Pae-6-1	NC_008463.1	+	4286083	4286166	←hypo **RNA**→ hypo→
Pae-1-1	NZ_AAKW01000002.1	-	898	815	←hypo **RNA**→ hypo→
Pae-4-1	NZ_AAQW01000001.1	-	1918648	1918565	←hypo **RNA**→ hypo→
Ppu-4-1	NZ_AAVY01000017.1	+	134946	135034	←Glycos_transf_2 (pfam00535) **RNA**→ hypo→ hypo→ hypo→
Dps-1-2	NC_006138.1	+	446698	446777	←hypo **RNA**→ ←hypo
Asa-1-1	NC_009348.1	+	3150898	3150985	putative phage regulator→ **RNA**→ putative phage regulator→ hypo→
env-9	JCVI_SCAF_1101668263049	-	1534	1447	**RNA**→
Bsp-1-1	NC_007510.1	-	2729355	2729271	←hypo **RNA**→ hypo→ Phage_AlpA (pfam05930)→
Pen-1-1	NC_008027.1	+	2652833	2652926	←RpoE (COG1595) **RNA**→ hypo→
Cvi-1-1	NC_005085.1	+	4158872	4158962	AlpA (COG3311)→ **RNA**→ hypo→ hypo→ hypo→
Sdy-2-1	NC_007606.1	+	1757441	1757528	BsuBI_PstI_RE (pfam06616)NOG08008 (NOG08008)→ **RNA**→ COG3646 (COG3646)→ hypo→
Sbo-1-2	NZ_AAKA01000020.1	-	58210	58123	Transposase_27 (pfam03400)→ **RNA**→ hypo→
env-10	BABG01026403.1	+	226	313	**RNA**→ hypo→ hypo→ hypo→
Cte-1-1	NZ_AAUJ01000008.1	-	30413	30331	COG4748 (COG4748)→ **RNA**→ COG3646 (COG3646)→
Cte-1-2	NZ_AAUJ01000008.1	+	21804	21888	hypo→ **RNA**→ COG3646 (COG3646)→
Sgl-2-2	NC_007712.1	+	2663915	2664000	←hypo **RNA**→ hypo→
Sen-8-1	NZ_ABEW01000010.1	+	115026	115110	**RNA**→
Vfi-1-1	NC_006840.1	-	2244353	2244268	hypo→ **RNA**→ COG3617 (COG3617)→
Vfi-1-2	NC_006840.1	-	2243662	2243579	COG3617 (COG3617)→ **RNA**→ ←SmpB (COG0691)

Sde-1-1	NC_007912.1	+	342124	342208	←Secretion protein HlyD **RNA**→ hypo→
Sba-1-1	NC_009665.1	-	896014	895930	←COG3941 (COG3941) **RNA**→ COG3646 (COG3646)→
Sba-2-1	NZ_AATK01000033.1	-	24005	23923	←COG5283 (COG5283) **RNA**→ COG3646 (COG3646)→
Eco-3-1	NZ_AAKB01000016.1	+	10920	11003	POLAc (smart00482)→ **RNA**→ COG3561 (COG3561)→
Vch-2-1	NZ_AAWD01000002.1	+	57743	57826	←hypo (cons) **RNA**→ COG3617 (COG3617)→
Vch-3-1	NC_009456.1	-	1033991	1033907	←hypo **RNA**→ COG3617 (COG3617)→
Vsp-1-1	NZ_AAMR01000038.1	-	33213	33129	←hypo **RNA**→ COG3646 (COG3646)→
Vch-1-1	NZ_AAUU01000060.1	-	5711	5630	hypo→ **RNA**→ transposase ORF-A, IS-type→
Vba-1-1	NZ_AAZW01000012.1	-	43784	43701	COG3335 (COG3335)→ **RNA**→ KilA-N (pfam04383)→
Sfr-1-1	NC_008345.1	-	1972071	1971987	←hypo **RNA**→ COG3646 (COG3646)→
Vsh-1-1	NZ_ABCH01000029.1	+	41299	41379	←hypo **RNA**→ KilA-N (pfam04383)→
Sgl-1-1	NC_007185.1	+	1478	1562	←RepA protein **RNA**→ HTH_LUXR (smart00421)→
Sgl-2-3	NC_007714.1	-	16189	16105	←replication associated protein RepA1 **RNA**→ HTH_LUXR (smart00421)→
Sgl-1-2	NC_007184.1	+	1478	1562	←putative replication-associated protein **RNA**→ HTH_LUXR (smart00421)→
env-11	BAAW01000646.1	+	4186	4274	MarC (pfam01914)→ **RNA**→ hypo→
Eco-5-1	NZ_AAWW01000074.1	-	5013	4925	←YD repeat protein **RNA**→ hypo→
Sen-11-1	NC_004631.1	+	4519523	4519609	←hypo **RNA**→ hypo→ Nlp (COG3423)→
Sen-10-1	NC_003198.1	+	4536714	4536800	←hypo **RNA**→ hypo→ Nlp (COG3423)→
Sen-12-1	NZ_ABFH01000014.1	+	112132	112219	**RNA**→
Esp-1-2	NC_009436.1	+	4152361	4152449	MarC (pfam01914)→ **RNA**→ hypo→
Sgl-2-4	NC_007712.1	+	3395263	3395347	←hypo **RNA**→ COG3646 (COG3646)→
Sgl-2-5	NC_007712.1	-	1233127	1233043	HTH_LUXR (smart00421)→ **RNA**→ COG3646 (COG3646)→
Sgl-2-6	NC_007712.1	-	2063246	2063161	hypo→ **RNA**→ hypo→ hypo→
Sgl-2-7	NC_007712.1	+	3106295	3106377	hypo→ **RNA**→ hypo→
Sgl-2-8	NC_007712.1	-	1814121	1814038	ArtM (COG4160)→ **RNA**→ COG3617 (COG3617)→
Sdy-1-1	NZ_AAMJ01000005.1	+	122662	122749	←HicB (pfam05534) **RNA**→ COG3646 (COG3646)→
Sbo-1-3	NZ_AAKA01000020.1	+	568	655	**RNA**→ COG3646 (COG3646)→
Eco-10-1	NZ_AAJZ01000001.1	-	4185666	4185579	COG3646 (COG3646)→ **RNA**→ COG3646 (COG3646)→
Sfl-1-1	NC_004741.1	+	1891549	1891636	hypothetical bacteriophage protein→ **RNA**→ hypothetical bacteriophage protein→
Sfl-2-1	NC_004337.1	-	1419385	1419298	COG3646 (COG3646)→ **RNA**→ COG3646 (COG3646)→
Eco-3-2	NZ_AAKB01000057.1	+	7419	7506	COG3646 (COG3646)→ **RNA**→ COG3646 (COG3646)→
Eco-3-3	NZ_AAKB01000097.1	-	235	148	**RNA**→
Eco-3-4	NZ_AAKB01000020.1	+	9022	9109	COG3646 (COG3646)→ **RNA**→ COG3646 (COG3646)→
Sfl-3-1	NC_008258.1	+	771649	771736	COG3646 (COG3646)→ **RNA**→ COG3646 (COG3646)→
Sfl-2-2	NC_004337.1	+	913450	913537	COG3646 (COG3646)→ **RNA**→ COG3646 (COG3646)→
Sfl-1-2	NC_004741.1	+	908250	908337	putative bacteriophage protein→ **RNA**→ putative bacteriophage protein→ hypothetical bacteriophage protein→
Eco-6-1	NZ_AAJX01000004.1	+	60275	60362	COG3646 (COG3646)→ **RNA**→ COG3646 (COG3646)→ hypo→ COG0640: Predicted transcriptional regulators→ hypo→ Dam (pfam05869)→ LexA_DNA_bind (pfam01726)→ RusA (pfam05866)→ KilA-N (pfam04383)→ hypo→
Pcd-1-1	NC_009514.1	+	36578	36665	**RNA**→
Sso-1-1	NC_007384.1	+	1984424	1984511	Transposase_27 (pfam03400)→ **RNA**→ COG3646 (COG3646)→
Sbo-1-4	NZ_AAKA01000012.1	-	22926	22839	COG3646 (COG3646)→ **RNA**→ COG3646 (COG3646)→ hypo→ hypo→
Sfl-3-2	NC_008258.1	+	2768141	2768228	putative bacteriophage protein→ **RNA**→ putative bacteriophage protein→
Eco-3-5	NZ_AAKB01000061.1	+	3762	3849	COG3646 (COG3646)→ **RNA**→ COG3646 (COG3646)→ hypo→ hypo→ hypothetical proteinCould not find sequence in database.→
Eco-13-1	NC_002695.1	+	2598583	2598670	hypo→ **RNA**→ hypo→ putative derepression protein→ hypo→
Eco-12-1	NC_002655.2	+	2673794	2673881	unknown protein encoded by prophage CP-933T→ **RNA**→ unknown protein encoded by prophage CP-933T→ unknown protein encoded by prophage CP-933T→ putative replication protein for prophage CP-933T→ StbA (pfam06406)→ putative stability/partitioning protein encoded within CP-933T→
Eco-3-6	NZ_AAKB01000056.1	+	37	124	**RNA**→
Sen-13-1	NZ_ABAO01000085.1	-	1827	1740	**RNA**→
Sph-1-1	NC_004313.1	+	30105	30192	**RNA**→
Sen-16-1	NC_006905.1	+	1315152	1315239	←HTH_XRE (smart00530)NOG17530 (NOG17530) **RNA**→ COG3646 (COG3646)→ hypo→
Eco-2-1	NZ_AAMK01000006.1	+	59258	59345	COG3646 (COG3646)→ **RNA**→ COG3646 (COG3646)→

env-12	BABF01000198.1	-	474	387	COG3646 (COG3646)→ **RNA**→ COG3646 (COG3646)→
Eco-15-1	NC_007946.1	+	4997122	4997209	COG3646 (COG3646)→ **RNA**→ COG3646 (COG3646)→ hypo→ putative replication protein→ hypo→ Dam (pfam05869)NOG18143 (NOG18143)→ LexA_DNA_bind (pfam01726)→ RusA (pfam05866)→ KilA-N (pfam04383)NOG17190 (NOG17190)→ hypothetical protein YdfU→ putative Q protein→
Eco-4-1	NC_008563.1	+	5012458	5012545	COG3646 (COG3646)→ **RNA**→ COG3646 (COG3646)→
Eco-6-2	NZ_AAJX01000031.1	+	41435	41522	COG3646 (COG3646)→ **RNA**→ COG3646 (COG3646)→ hypo→ hypo→
Eco-6-3	NZ_AAJX01000103.1	-	3179	3092	hypo→ **RNA**→ hypo→ hypo→ hypo→
Eco-2-2	NZ_AAMK01000026.1	+	15471	15558	MarR (COG1846)→ **RNA**→ hypo→ hypo→ hypo→ StbA (pfam06406)→
Eco-8-1	NZ_AAJW01000006.1	+	162043	162130	hypo→ **RNA**→ hypo→ hypo→ hypo→
Sdy-1-2	NZ_AAMJ01000008.1	+	123989	124076	hypo→ **RNA**→ hypo→ hypo→ hypo→
Eco-7-1	NZ_AAJT01000067.1	-	18116	18029	hypo→ **RNA**→ hypo→ hypo→ hypo→ StbA (pfam06406)→ hypo→
Eco-6-4	NZ_AAJX01000034.1	+	25727	25814	hypo→ **RNA**→ hypo→ hypo→ hypo→ StbA (pfam06406)→ hypo→
Eco-9-1	NZ_AAJV01000026.1	+	43700	43787	←COG0043: 3-polyprenyl-4-hydroxybenzoate decarboxylase and related decarboxylases **RNA**→ ←COG0043: 3-polyprenyl-4-hydroxybenzoate decarboxylase and related decarboxylases
Eco-3-7	NZ_AAKB01000017.1	+	49998	50085	COG0583: Transcriptional regulator→ **RNA**→ hypo→ hypo→ hypo→ hypo→
Eco-15-2	NC_007946.1	-	942027	941938	HflC (COG0330)→ **RNA**→ ←hypo
Eco-2-3	NZ_AAMK01000002.1	+	372673	372760	←hypo **RNA**→ ←hypo
Eco-10-2	NZ_AAJZ01000001.1	-	4973644	4973559	hypo→ **RNA**→ hypo→
Spr-1-2	NZ_AAUN01000029.1	+	36100	36187	hypo→ **RNA**→ COG3617 (COG3617)→
Sty-2-1	NC_003197.1	-	2762765	2762682	antiterminator-like protein→ **RNA**→ antirepressor-like protein→
PGi-1-1	NC_010392.1	-	34214	34131	**RNA**→
Sen-9-1	NZ_ABAN01000002.1	+	164	247	**RNA**→
Sen-9-2	NZ_ABAN01000025.1	-	4031	3948	**RNA**→
Sen-14-1	NZ_ABEH01000004.1	-	43689	43606	**RNA**→
Sen-13-2	NZ_ABAO01000067.1	-	2472	2389	**RNA**→
Sen-7-1	NZ_ABEN01000001.1	-	2762799	2762716	**RNA**→
Sen-8-2	NZ_ABEW01000016.1	+	12612	12695	**RNA**→
Sen-15-1	NZ_ABFF01000012.1	+	15475	15558	**RNA**→
Eco-10-3	NZ_AAJZ01000001.1	-	4209920	4209837	hypo→ **RNA**→ ←Caudo_TAP (pfam02413)
Sso-1-2	NC_007384.1	-	2936419	2936336	antitermination protein Q-like protein→ **RNA**→ putative bacteriophage protein→
Ype-7-1	NC_008149.1	+	1777837	1777924	phage antitermination protein→ **RNA**→ COG3617 (COG3617)→
Ype-8-1	NC_009381.1	-	1164990	1164903	phage antitermination protein→ **RNA**→ COG3617 (COG3617)→
Ype-6-1	NC_004088.1	-	2450258	2450171	hypo→ **RNA**→ COG3617 (COG3617)→
Ype-4-1	NC_003143.1	+	2369225	2369312	hypo→ **RNA**→ hypo→ COG3617 (COG3617)→
Ype-13-1	NC_005810.1	+	2151105	2151192	hypo→ **RNA**→ hypo→ COG3617 (COG3617)→
Ype-15-1	NZ_AAOS01000087.1	-	11026	10939	hypo→ **RNA**→ COG3617 (COG3617)→
Ype-5-1	NZ_AAUB01000062.1	+	10162	10249	**RNA**→
Ype-16-1	NZ_AAYS01000003.1	-	187708	187621	**RNA**→
Ype-12-1	NZ_AAYT01000003.1	-	330911	330824	**RNA**→
Ype-9-1	NZ_AAYU01000001.1	+	387329	387416	**RNA**→
Ype-10-1	NZ_AAYV01000003.1	+	195033	195120	**RNA**→
Ype-14-1	NZ_ABAT01000021.1	-	71591	71504	**RNA**→
Ype-3-1	NZ_ABCD01000001.1	-	1281217	1281130	hypothetical phage protein→ **RNA**→ hypothetical phage protein→ COG3617 (COG3617)→
Ype-1-1	NZ_AAKS01000038.1	+	24399	24486	hypo→ **RNA**→ COG3617 (COG3617)→
Eco-3-8	NZ_AAKB01000062.1	+	3801	3885	COG1522: Transcriptional regulators→ **RNA**→ hypo→
Eco-3-9	NZ_AAKB01000008.1	-	102865	102778	COG1763: Molybdopterin-guanine dinucleotide biosynthesis protein→ **RNA**→ COG1763: Molybdopterin-guanine dinucleotide biosynthesis protein→ hypo→
Yen-1-1	NC_008800.1	-	2104229	2104144	hypo→ **RNA**→ Transposase_31 (pfam04754)→
Yen-1-2	NC_008800.1	-	4498927	4498841	←COG3177 (COG3177) **RNA**→ hypo→
Yps-1-1	NC_009708.1	-	1069976	1069890	←HTH_XRE (cd00093) **RNA**→ hypo→
Ype-13-2	NC_005810.1	+	4074558	4074644	←RelE (COG2026) **RNA**→ ←INT_P4 (cd00801)
Yps-2-1	NC_006155.1	+	3687840	3687926	←RelE (COG2026) **RNA**→ ←HP1_INT_C (cd00797)
Ype-2-1	NC_008150.1	-	451213	451127	←RelE (COG2026) **RNA**→ hypo→

Ype-7-2	NC_008149.1	+	3467455	3467541	←RelE (COG2026) **RNA**→ hypo→
Ype-8-2	NC_009381.1	-	711648	711562	←HTH_XRE (cd00093) **RNA**→ hypo→
Ype-6-2	NC_004088.1	+	3599632	3599718	hypo→ **RNA**→ hypo→ hypo→
Ype-4-2	NC_003143.1	+	969600	969686	←RelE (COG2026) **RNA**→ ←INT_P4 (cd00801)
Ype-1-2	NZ_AAKS01000008.1	+	55766	55852	←HTH_XRE (cd00093) **RNA**→ hypo→
Ype-15-2	NZ_AAOS01000023.1	-	12408	12322	←HTH_XRE (cd00093) **RNA**→ hypo→
Ype-5-2	NZ_AAUB01000028.1	-	1859	1773	**RNA**→
Ype-11-1	NZ_AAYR01000002.1	+	390330	390416	**RNA**→
Ype-16-2	NZ_AAYS01000001.1	+	283755	283841	**RNA**→
Ype-12-2	NZ_AAYT01000006.1	+	182777	182863	**RNA**→
Ype-9-2	NZ_AAYU01000005.1	-	204312	204226	**RNA**→
Ype-10-2	NZ_AAYV01000007.1	+	132679	132765	**RNA**→
Ype-14-2	NZ_ABAT01000002.1	-	94342	94256	**RNA**→
Ype-3-2	NZ_ABCD01000004.1	-	31787	31701	←RelE (COG2026) **RNA**→ ←INT_P4 (cd00801)
Yen-1-3	NC_008800.1	-	1858079	1857993	←hypo **RNA**→ ←hypo
Ybe-1-1	NZ_AALC01000034.1	+	29766	29853	Transposase_31 (pfam04754)→ **RNA**→ hypo→
Yin-1-1	NZ_AALF01000002.1	+	162645	162731	←hypo **RNA**→ hypo→
Yen-1-4	NC_008800.1	-	1859900	1859814	hypo→ **RNA**→ Transposase_31 (pfam04754)→
Yin-1-2	NZ_AALF01000002.1	+	164423	164510	hypo→ **RNA**→ hypo→
Sbo-2-1	NC_007613.1	+	1140537	1140623	hypo→ **RNA**→ hypo→ Nlp (COG3423)→
env-13	BAAW01012961.1	-	630	545	**RNA**→ hypo→
Sen-16-2	NC_006905.1	+	1315804	1315890	←HTH_XRE (smart00530)NOG17530 (NOG17530) **RNA**→ COG3646 (COG3646)→ hypo→
Api-1-1	NC_000935.1	-	28158	28072	**RNA**→
Sgl-2-9	NC_007712.1	-	1935487	1935401	←hypo **RNA**→ INT_Lambda_C (cd00800)→
Sdy-1-3	NZ_AAMJ01000022.1	+	3716	3799	hypo→ **RNA**→ hypo→
Eco-3-10	NZ_AAKB01000065.1	+	5191	5278	**RNA**→
Eph-2-1	NC_003356.1	+	9612	9699	**RNA**→
Eco-3-11	NZ_AAKB01000061.1	+	2450	2537	hypo→ **RNA**→ hypo→ hypo→ COG3646 (COG3646)→ hypo→ hypo→
Eco-3-12	NZ_AAKB01000061.1	+	3106	3193	hypothetical proteinCould not find sequence in database.→ hypo→ **RNA**→ hypo→ COG3646 (COG3646)→ hypo→ hypo→
Sdy-1-4	NZ_AAMJ01000006.1	-	84023	83936	hypothetical proteinCould not find sequence in database.→ hypo→ **RNA**→ hypo→ hypo→ hypo→ hypo→ COG0543: 2-polyprenylphenol hydroxylase and related flavodoxin oxidoreductases→ hypo→ hypo→ COG4643 (COG4643)COG5519 (COG5519)→
Eco-8-2	NZ_AAJW01000030.1	-	27742	27658	Dam (pfam05869)→ **RNA**→ COG3561 (COG3561)→ COG3646 (COG3646)ANT (pfam03374)→
Sdy-2-2	NC_007606.1	-	2215306	2215222	COG3561 (COG3561)→ **RNA**→ COG3561 (COG3561)→ COG3646 (COG3646)ANT (pfam03374)→
Eco-10-4	NZ_AAJZ01000001.1	-	4210661	4210579	DUF892 (pfam05974)→ **RNA**→ hypo→
Sfl-2-3	NC_004337.1	+	1400922	1401006	putative bacteriophage protein→ **RNA**→ Transposase_8 (pfam01527)→ rve (pfam00665)→ hypothetical bacteriophage protein→
Sen-10-2	NC_003198.1	-	1915250	1915166	hypo→ **RNA**→ putative bacteriophage protein→
Sdy-2-3	NC_007606.1	-	553748	553664	hypo→ **RNA**→ putative bacteriophage protein→
Eco-9-2	NZ_AAJV01000004.1	-	34572	34487	DUF892 (pfam05974)→ **RNA**→ hypo→
Eph-1-1	NC_003444.1	+	27704	27791	**RNA**→
Eco-13-2	NC_002695.1	+	1263291	1263373	NinE→ **RNA**→ COG3561 (COG3561)→ COG3646 (COG3646)ANT (pfam03374)→ NinG→ hypo→ antitermination protein Q→
Eco-8-3	NZ_AAJW01000017.1	+	44676	44758	Dam (pfam05869)→ **RNA**→ COG3561 (COG3561)→ ANT (pfam03374)→
Eph-5-1	NC_000902.1	+	17279	17361	**RNA**→
Sco-1-1	NC_004914.1	+	49167	49249	**RNA**→
Sen-15-2	NZ_ABFF01000008.1	-	134074	133992	**RNA**→
env-14	BABF01001575.1	+	2753	2840	hypo→ **RNA**→ COG3646 (COG3646)→
Eco-9-3	NZ_AAJV01000109.1	-	1078	991	COG3646 (COG3646)→ **RNA**→ COG3646 (COG3646)→
Sen-2-1	NZ_ABAP01000068.1	-	6237	6150	**RNA**→
Sen-15-3	NZ_ABFF01000019.1	+	53810	53897	**RNA**→
Eco-11-1	NZ_AAJU01000058.1	-	222	135	**RNA**→

env-15	BABF01001366.1	+	2134	2221	**RNA**→
Sen-9-3	NZ_ABAN01000056.1	-	12998	12911	**RNA**→
Eco-4-2	NC_008563.1	+	2745296	2745380	hypo→ **RNA**→ COG3617 (COG3617)→
Eph-6-1	NC_007804.2	-	20363	20278	**RNA**→
env-16	BAAX01012263.1	-	1243	1160	**RNA**→ COG3617 (COG3617)→
Sgl-1-3	NC_007182.1	+	21311	21396	Tpn1 protein→ **RNA**→ hypo→
Sgl-1-4	NC_007183.1	+	21311	21396	Tpn1 protein→ **RNA**→ hypo→
Sgl-2-10	NC_007713.1	+	50694	50779	←Antirestrict (pfam03230) **RNA**→ ←Transposase_31 (pfam04754)
Pfl-2-5	NC_007492.1	+	1502980	1503070	←UbiE (COG2226) **RNA**→ hypo→
Ppu-1-1	NC_009512.1	+	339637	339726	←hypo **RNA**→ hypo→
Pfl-2-6	NC_007492.1	-	5257618	5257533	←AcrA (COG0845)NOG08737 (NOG08737) **RNA**→ hypo→
env-17	sludgeOz_Contig11395	+	4620	4705	FOG(,,,)→ **RNA**→ COG3646 (COG3646)→
env-18	sludgeOz_Contig11626	-	4444	4358	TPR (cd00189)→ **RNA**→ COG3646 (COG3646)→
env-19	sludgeOz_Contig11626	+	29442	29528	←(,,,) **RNA**→ COG3646 (COG3646)→
env-20	sludgeOz_Contig9633	-	3083	2999	**RNA**→
Bth-1-1	NZ_ABBH01000502.1	+	7145	7238	**RNA**→
Bps-11-1	NZ_ABBO01000018.1	-	1738	1645	**RNA**→
Rso-1-1	NC_003295.1	+	1663851	1663938	hypo→ **RNA**→ hypo→ LT_GEWL (cd00254)→
Sgl-2-11	NC_007712.1	-	1527580	1527499	←hypo **RNA**→ COG3646 (COG3646)→ COG3646 (COG3646)NOG18702 (NOG18702)→ hypo→
Bxe-1-2	NC_007951.1	+	915998	916090	COG3415 (COG3415)rve (pfam00665)→ **RNA**→ hypo→
Bph-1-2	NZ_AAUH01000005.1	-	66833	66741	←COG3781 (COG3781) **RNA**→ hypo (cons)→
Ppu-4-2	NZ_AAVY01000026.1	-	50658	50567	GlpC (COG0247)GlcD (COG0277)→ **RNA**→ hypo (cons)→ hypo→
Pen-1-2	NC_008027.1	-	5667972	5667886	←hypo **RNA**→ hypo→ hypo→
Pst-1-1	NC_009434.1	-	813410	813328	hypo→ **RNA**→ hypo→
Sgl-2-12	NC_007712.1	-	2086052	2085968	←hypo **RNA**→ COG3617 (COG3617)→
Eph-7-1	NC_001901.1	-	28388	28301	**RNA**→
env-21	BAAW01005016.1	+	464	551	**RNA**→ hypo→ hypo→
Esa-1-1	NC_009778.1	-	2971361	2971274	Phage_cap_P2 (pfam05125)→ **RNA**→ hypo→
Sen-11-2	NC_004631.1	+	4674209	4674296	phage immunity repressor protein→ **RNA**→ phage immunity repressor protein→ hypo→ hypothetical P4 phage protein→
Sen-10-3	NC_003198.1	+	4691320	4691407	phage immunity repressor protein→ **RNA**→ phage immunity repressor protein→ hypo→ hypothetical P4 phage protein→ COG3378 (COG3378)COG4643 (COG4643)→
Yph-1-1	NC_005069.1	+	23719	23806	**RNA**→
Sgl-2-13	NC_007712.1	+	2901789	2901873	hypo→ **RNA**→ hypo→
Sgl-2-14	NC_007712.1	-	2897417	2897333	hypo→ **RNA**→ RusA (pfam05866)→
Eco-3-13	NZ_AAKB01000013.1	-	22127	22040	Phage_AlpA (pfam05930)→ **RNA**→ hypo→
Sen-7-2	NZ_ABEN01000001.1	+	4718071	4718158	**RNA**→
Esp-1-3	NC_009436.1	+	3381272	3381359	putative CI repressor→ **RNA**→ putative CI repressor→ hypo→ conserved hypothetical P4 phage protein→ COG3378 (COG3378)COG4643 (COG4643)→
Ybe-1-2	NZ_AALC01000015.1	+	69245	69332	hypo→ **RNA**→ hypo→ hypo→ hypo→ COG3378 (COG3378)COG4643 (COG4643)→
Ymo-1-1	NZ_AALD01000007.1	+	16582	16669	hypo→ **RNA**→ hypo→ hypo→ hypo→ COG3378 (COG3378)COG4643 (COG4643)→
Esa-1-2	NC_009778.1	-	596095	596008	AlpA (COG3311)→ **RNA**→ hypo→
Yfr-1-1	NZ_AALE01000013.1	-	3159	3075	hypo→ **RNA**→ hypo→ hypo→ hypo→ COG3378 (COG3378)COG4643 (COG4643)→
Sen-9-4	NZ_ABAN01000037.1	-	23094	23007	**RNA**→
Sen-5-1	NZ_ABEL01000004.1	-	418202	418115	**RNA**→
Sen-4-1	NZ_ABEM01000001.1	-	708764	708677	**RNA**→
Sen-7-3	NZ_ABEN01000001.1	+	2885196	2885283	**RNA**→
Sen-8-3	NZ_ABEW01000002.1	-	653315	653228	**RNA**→
Eco-13-3	NC_002695.1	+	318667	318754	putative CI repressor→ **RNA**→ putative CI repressor→ hypo→ Cnr-like protein→ COG3378 (COG3378)COG4643 (COG4643)→
Eco-12-2	NC_002655.2	+	318668	318755	putative regulator encoded in prophage CP-933I→ **RNA**→ putative regulator encoded in prophage CP-933I→
env-22	BAAX01008057.1	-	541	455	hypo→ **RNA**→ hypo→ hypo→
env-23	BAAW01000581.1	-	4187	4100	AlpA (COG3311)→ **RNA**→ hypo→ hypo→ hypo→ COG3378 (COG3378)COG4643 (COG4643)→

Sen-11-3	NC_004631.1	+	2741095	2741182	hypo→ **RNA**→ hypo→ hypo→ hypo→
Eco-5-2	NZ_AAWW01000009.1	+	31532	31619	putative bacteriophage protein→ **RNA**→ putative bacteriophage protein→ hypo (cons)→ hypo (cons)→
Eco-5-3	NZ_AAWW01000018.1	+	30662	30749	putative bacteriophage protein→ **RNA**→ putative bacteriophage protein→ hypo (cons)→ hypo (cons)→
Eco-14-1	NZ_ABAQ01000005.1	-	134828	134741	**RNA**→
Eco-11-2	NZ_AAJU01000002.1	-	106093	106006	hypo→ **RNA**→ hypo→
Eph-4-1	NC_001609.1	-	8507	8420	**RNA**→
Sbo-2-2	NC_007613.1	-	4221428	4221341	putative bacteriophage protein→ **RNA**→ putative bacteriophage protein→
Eco-8-4	NZ_AAJW01000039.1	+	10939	11026	hypo→ **RNA**→ hypo→
Sgl-2-15	NC_007713.1	+	13096	13181	hypo→ **RNA**→ CitB (COG2197)→
Sty-1-1	NC_005014.1	-	120268	120181	**RNA**→ hypo→
Eco-1-1	NC_007635.1	-	97837	97750	hypo→ **RNA**→ hypo→ HTH_LUXR (smart00421)→
Sen-6-1	NZ_ABAK01000021.1	+	44344	44431	**RNA**→
Sen-4-2	NZ_ABEM01000011.1	-	16405	16318	**RNA**→
Sen-4-3	NZ_ABEM01000012.1	+	9911	9998	**RNA**→
Sen-4-4	NZ_ABEM01000008.1	+	9908	9995	**RNA**→
Sen-4-5	NZ_ABEM01000008.1	+	101276	101363	**RNA**→
Kph-1-1	NC_005857.1	-	32623	32536	**RNA**→
Ype-2-2	NC_008150.1	-	654903	654816	putative prophage protein→ **RNA**→ putative prophage protein→ COG4643 (COG4643)→ repA (cd01125)→
Ype-7-3	NC_008149.1	+	3265707	3265794	prophage protein→ **RNA**→ prophage protein→ COG4643 (COG4643)→ repA (cd01125)→
Ype-8-3	NC_009381.1	+	2954292	2954379	prophage protein→ **RNA**→ prophage protein→ COG4643 (COG4643)→ repA (cd01125)→
Ype-6-3	NC_004088.1	+	3396658	3396745	hypo→ **RNA**→ hypo→ COG4643 (COG4643)→ repA (cd01125)NOG13495 (NOG13495)→
Ype-4-3	NC_003143.1	-	1238167	1238080	putative prophage protein→ **RNA**→ putative prophage protein→ COG4643 (COG4643)→ repA (cd01125)NOG13495 (NOG13495)→
Ype-13-3	NC_005810.1	+	1144762	1144849	putative prophage protein→ **RNA**→ putative prophage protein→ COG4643 (COG4643)→ repA (cd01125)→
Ype-1-3	NZ_AAKS01000002.1	+	71350	71437	hypo→ **RNA**→ hypo→
Ype-15-3	NZ_AAOS01000089.1	+	4167	4254	hypo→ **RNA**→ hypo→ COG4643 (COG4643)→ repA (cd01125)→
Ype-5-3	NZ_AAUB01000092.1	+	7617	7704	**RNA**→
Ype-11-2	NZ_AAYR01000002.1	+	184692	184779	**RNA**→
Ype-16-3	NZ_AAYS01000002.1	+	420406	420493	**RNA**→
Ype-12-3	NZ_AAYT01000002.1	+	577828	577915	**RNA**→
Ype-9-3	NZ_AAYU01000002.1	-	73126	73039	**RNA**→
Ype-10-3	NZ_AAYV01000015.1	-	22777	22690	**RNA**→
Ype-14-3	NZ_ABAT01000001.1	+	556501	556588	**RNA**→
Ype-3-3	NZ_ABCD01000002.1	+	736077	736164	putative prophage protein→ **RNA**→ putative prophage protein→ COG4643 (COG4643)→ repA (cd01125)→
env-24	BABF01000165.1	+	10514	10601	COG3561 (COG3561)→ **RNA**→ hypo→ hypo→ hypo→ hypo→ hypo→ hypo→ hypo→
Sen-1-1	NC_006511.1	+	4467249	4467336	phage immunity repressor protein→ **RNA**→ phage immunity repressor protein→ hypo→
Sen-14-2	NZ_ABEH01000010.1	+	117001	117088	hypothetical P4 phage protein→ COG3378 (COG3378)COG4643 (COG4643)→ **RNA**→
env-25	JCVI_SCAF_1096627954892	+	938	1021	**RNA**→
Ppu-4-3	NZ_AAVY01000013.1	+	86112	86198	←OYE_YqiM_FMN (cd02932) **RNA**→ hypo→
Ppu-2-1	NZ_AAXR01000007.1	+	203630	203715	←OYE_YqiM_FMN (cd02932) **RNA**→ hypo (cons)→
Pen-1-3	NC_008027.1	-	5194676	5194589	←LivK (COG0683) **RNA**→ hypo→
Nmu-1-1	NC_007614.1	-	68166	68083	AlpA (COG3311)NOG19993 (NOG19993)→ **RNA**→ hypo→ hypo→ hypo→
env-26	sludgePhrap_Contig2337	+	52	135	**RNA**→
Avi-1-1	NZ_AAAU03000001.1	-	771062	770978	COG5659 (COG5659)→ **RNA**→ COG3617 (COG3617)→
Pae-4-2	NZ_AAQW01000001.1	+	6009993	6010074	MviM (COG0673)→ **RNA**→ hypo→
Ppu-2-2	NZ_AAXR01000002.1	-	340097	340011	KefB (COG0475)Kch (COG1226)→ **RNA**→ hypo (cons)→
Pen-1-4	NC_008027.1	+	427349	427440	←COG2944 (COG2944) **RNA**→ hypo→
Pen-1-5	NC_008027.1	-	4670418	4670332	←PhnF (COG2188) **RNA**→ hypo→ hypo→
Ppu-4-4	NZ_AAVY01000014.1	-	30948	30861	←AcrA (COG0845) **RNA**→ hypo→ hypo→
Pen-1-6	NC_008027.1	+	1026431	1026518	COG4625 (COG4625)→ **RNA**→ hypo→ hypo→
Pen-1-7	NC_008027.1	+	619442	619529	SGNH_hydrolase_peri1 (cd01825)→ **RNA**→ hypo→ hypo→
Ppu-4-5	NZ_AAVY01000023.1	+	201	288	**RNA**→ hypo→ hypo→

Ppu-3-1	NC_002947.3	+	335774	335862	←hypo **RNA**→ hypo→
Ppu-3-2	NC_002947.3	-	2302915	2302823	hypo→ **RNA**→ hypo→
Ppu-4-6	NZ_AAVY01000003.1	-	8342	8252	OprD (pfam03573)→ **RNA**→ hypo (cons)→ hypo→
Ppu-1-2	NC_009512.1	+	4379908	4379996	←ComEA (COG1555) **RNA**→ hypo→
Ppu-4-7	NZ_AAVY01000059.1	-	821	732	←hypo (cons) **RNA**→ hypo→ hypo→
Ppu-3-3	NC_002947.3	-	2038016	2037928	WecB (COG0381)→ **RNA**→ ←hypo
Ppu-4-8	NZ_AAVY01000023.1	-	18560	18473	hypo→ **RNA**→ repA (cd01125)→
Ppu-4-9	NZ_AAVY01000025.1	+	47012	47100	plasmid pRiA4b ORF-3-like→ **RNA**→ ←CynX (COG2807)
env-27	JCVI_SCAF_1101668291644	-	1191	1104	**RNA**→
Pen-1-8	NC_008027.1	-	3518484	3518397	←hypo **RNA**→ hypo→
Ppu-1-3	NC_009512.1	+	770187	770275	←tRNA-OTHER **RNA**→ hypo→
Ppu-4-10	NZ_AAVY01000026.1	-	49745	49657	hypo→ **RNA**→ hypo (cons)→ hypo→
Ppu-4-11	NZ_AAVY01000021.1	-	79306	79219	←COG1943 (COG1943) **RNA**→ hypo→ hypo→
Pen-1-9	NC_008027.1	-	793129	793043	←hypo **RNA**→ ←hypo
Pen-1-10	NC_008027.1	+	980078	980164	←GGDEF (cd01949) **RNA**→ hypo→ hypo→
Pfl-2-7	NC_007492.1	+	5006731	5006819	COG1721 (COG1721)→ **RNA**→ hypo→
Xor-1-1	NC_006834.1	-	1745312	1745231	TOPRIM_primases (cd01029)NOG06080 (NOG06080)→ **RNA**→ hypo→
Xor-2-1	NC_007705.1	-	1727613	1727532	TOPRIM_primases (cd01029)→ **RNA**→ hypo→ hypo→
env-28	JCVI_SCAF_1096628042224	+	713	793	? (GOS12118414)→ **RNA**→
Xca-1-1	NC_003902.1	+	3556259	3556339	TOPRIM_primases (cd01029)→ **RNA**→ hypo→ hypo→ hypo→ hypo→ hypo→ hypo→
Sma-1-5	NZ_AAVZ01000020.1	-	14481	14397	TonB family protein→ **RNA**→ ←SSL2 (COG1061)
Bdo-1-1	NZ_AAKY01000047.1	+	15904	15993	←hypo **RNA**→ hypo→
Pfl-2-8	NC_007492.1	-	753624	753536	hypo→ **RNA**→ hypo→
Pfl-1-2	NC_004129.6	-	3142100	3142011	←COG3666 (COG3666) **RNA**→ hypo→
Ppu-4-12	NZ_AAVY01000025.1	+	46883	46967	plasmid pRiA4b ORF-3-like→ **RNA**→ ←CynX (COG2807)
Pen-1-11	NC_008027.1	+	28932	29019	SUA5 (COG0009)→ **RNA**→ hypo→
Ppu-3-4	NC_002947.3	+	81574	81663	hypo→ **RNA**→ hypo→
Ppu-2-3	NZ_AAXR01000002.1	-	88895	88806	SUA5 (COG0009)→ **RNA**→ hypo (cons)→
Ppu-4-13	NZ_AAVY01000019.1	+	172	260	**RNA**→ hypo (cons)→
Ppu-1-4	NC_009512.1	+	93169	93257	SUA5 (COG0009)→ **RNA**→ hypo→
env-29	JCVI_SCAF_1096627487547	-	493	405	? (GOS11335753)→ **RNA**→ ? (GOS11335753)→
Psy-3-1	NC_004578.1	-	447442	447360	←ComJ (COG2961) **RNA**→ hypo→
Psy-1-2	NC_005773.3	+	5451361	5451443	←ComJ (COG2961) **RNA**→ ←hypo
Psy-1-3	NC_005773.3	+	4631607	4631691	hypo→ **RNA**→ hypo→
Pfl-2-9	NC_007492.1	-	3357442	3357358	hypo→ **RNA**→ hypo→
Pfl-2-10	NC_007492.1	+	2049523	2049609	ABC_Class2 (cd03227)NOG22115 (NOG22115)→ **RNA**→ hypo→ hypo→
Pfl-2-11	NC_007492.1	+	3662342	3662427	←hypo **RNA**→ hypo→
Pfl-2-12	NC_007492.1	-	2779539	2779454	hypo→ **RNA**→ hypo→
env-30	JCVI_SCAF_1101668291562	+	916	1003	R_hydratase (cd03449)→ **RNA**→ ? (GOS11008091)→
Psy-2-1	NC_007005.1	+	4827916	4828000	hypo→ **RNA**→ hypo→
Pfl-2-13	NC_007492.1	-	2641456	2641372	←hypo **RNA**→ hypo→
Psy-2-2	NC_007005.1	+	5651663	5651744	←ComJ (COG2961) **RNA**→ hypo→
Pfl-2-14	NC_007492.1	-	4317777	4317693	←COG4104 (COG4104) **RNA**→ COG3547 (COG3547)→
Psy-3-2	NC_004578.1	-	6094972	6094888	DnaJ (pfam00226)→ **RNA**→ hypo→
Pfl-2-15	NC_007492.1	-	2739890	2739805	←DUF72 (pfam01904) **RNA**→ hypo→
Avi-1-2	NZ_AAAU03000001.1	+	1076949	1077035	←DUF891 (pfam05973) **RNA**→ hypo→ hypo→ hypo→ TOPRIM_primases (cd01029)→ hypo→
Ppu-2-4	NZ_AAXR01000001.1	-	951433	951347	hypo→ **RNA**→ hypo (cons)→
Pen-1-12	NC_008027.1	-	2115228	2115143	←RhtB (COG1280) **RNA**→ ←ACR_tran (pfam00873)
Pfl-2-16	NC_007492.1	+	3960594	3960682	←COG5606 (COG5606) **RNA**→ hypo→ hypo→
Pfl-2-17	NC_007492.1	+	2273321	2273407	CsgG (COG1462)→ **RNA**→ hypo→ hypo→
Pfl-2-18	NC_007492.1	+	5173901	5173991	Gcd (COG4993)→ **RNA**→ hypo→
Pfl-1-3	NC_004129.6	-	6137446	6137360	←Plasmid_killer (pfam05015) **RNA**→ hypo→

Pen-1-13	NC_008027.1	+	1922676	1922764	COG4178 (COG4178)→ **RNA**→ hypo→
Pfl-2-19	NC_007492.1	-	2414243	2414159	←hypo **RNA**→ hypo→ hypo→
Pfl-2-20	NC_007492.1	-	854931	854844	Tar (COG0840)→ **RNA**→ hypo→
Pfl-2-21	NC_007492.1	-	351669	351583	←COG1724 (COG1724) **RNA**→ hypo→
Pfl-1-4	NC_004129.6	+	1157583	1157668	hypo→ **RNA**→ hypo→
env-31	JCVI_SCAF_1096626143183	+	172	261	**RNA**→
env-32	JCVI_SCAF_1096627151993	-	460	372	DszC (cd01163)? (GOS26)→ **RNA**→ ? (GOS11008091)→
env-33	JCVI_SCAF_1096627846063	+	1373	1459	←PelB (COG3866)? (GOS3774) **RNA**→
env-34	JCVI_SCAF_1096627605558	+	194	280	**RNA**→ ←? (GOS3431)
Pfl-2-22	NC_007492.1	+	5919175	5919260	lyase, putative→ **RNA**→ hypo→ COG1476 (COG1476)→
env-35	JCVI_SCAF_1096627781829	+	501	587	**RNA**→ COG4680 (COG4680)? (GOS13828)→
Pfl-1-5	NC_004129.6	-	1753283	1753198	←filamentous hemagglutinin, intein-containing, putative **RNA**→ hypo→
Psy-1-4	NC_005773.3	+	369048	369143	TauA (COG0715)→ **RNA**→ hypo→
Psy-2-3	NC_007005.1	+	363735	363830	TauA (COG0715)→ **RNA**→ hypo→ hypo→
Pen-1-14	NC_008027.1	-	4962315	4962229	←LipA (COG1075) **RNA**→ hypo→
env-36	JCVI_SCAF_1096627876825	-	932	845	**RNA**→
Pfl-2-23	NC_007492.1	+	2239520	2239607	LysR (COG0583)→ **RNA**→ hypo→
Pfl-2-24	NC_007492.1	-	1747486	1747402	←hypo **RNA**→ hypo→
Pfl-2-25	NC_007492.1	-	1740090	1740006	hypo→ **RNA**→ hypo→
Psy-2-4	NC_007005.1	+	4545144	4545229	←DUF1349 (pfam07081) **RNA**→ COG3617 (COG3617)→ adh_short (pfam00106)→
env-37	JCVI_SCAF_1096627936442	+	472	555	**RNA**→ ? (GOS13412611)→
env-38	JCVI_SCAF_1101667131354	+	39	129	**RNA**→
env-39	JCVI_SCAF_1096628172702	+	1077	1159	**RNA**→
Pfl-1-6	NC_004129.6	+	5731958	5732043	hypo→ **RNA**→ COG5519 (COG5519)TOPRIM_primases (cd01029)→ chromosome segregation SMC protein, putative→
Ppu-3-5	NC_002947.3	+	759603	759688	Acetyltransf_1 (pfam00583)→ **RNA**→ ←GntP_permease (pfam02447)
env-40	JCVI_SCAF_1096628052827	+	1532	1617	? (GOS11069324)→ **RNA**→ ? (GOS11725977)→
Pen-1-15	NC_008027.1	-	791092	791007	tmRNA involved in trans-translation→ **RNA**→ hypo→
Ppu-4-14	NZ_AAVY01000025.1	+	47399	47484	plasmid pRiA4b ORF-3-like→ **RNA**→ ←CynX (COG2807)
Pen-1-16	NC_008027.1	-	4960291	4960206	hypo→ **RNA**→ hypo→
Ppu-4-15	NZ_AAVY01000004.1	+	95972	96057	DTW (pfam03942)→ **RNA**→ ←HlyD (pfam00529)
env-41	JCVI_SCAF_1101668395415	-	172	86	**RNA**→
Psy-3-3	NC_004578.1	+	3480015	3480104	lipoprotein, putative→ **RNA**→ COG3617 (COG3617)→
Pae-3-1	NC_009656.1	-	373488	373402	INT_P4 (cd00801)→ **RNA**→ hypo→ hypo→ transcriptional regulator, AlpA family→ Bbp36→ zf-CHC2 (pfam01807)→ hypo→ hypo→ COG5519 (COG5519)→
Ppu-4-16	NZ_AAVY01000008.1	-	10570	10486	Aes (COG0657)→ **RNA**→ hypo (cons)→
Pfl-1-7	NC_004129.6	-	2447607	2447522	←hypo **RNA**→ hypo→
Pfl-2-26	NC_007492.1	+	2338723	2338805	ArgE (COG0624)→ **RNA**→ hypo→
Pfl-2-27	NC_007492.1	-	3665249	3665164	←Acetyltransf_1 (pfam00583)NOG15906 (NOG15906) **RNA**→ hypo→
env-42	JCVI_SCAF_1101667357642	+	288	376	**RNA**→ ? (GOS14070030)→
env-43	JCVI_SCAF_1101667148170	-	172	86	**RNA**→
Pfl-1-8	NC_004129.6	+	6960256	6960341	AspA (COG1027)→ **RNA**→ hypo→
Avi-1-3	NZ_AAAU03000001.1	-	851854	851768	hypo→ **RNA**→ hypo→ hypo→
Psy-1-5	NC_005773.3	-	5617115	5617032	DnaJ (pfam00226)→ **RNA**→ hypo→
Psy-2-5	NC_007005.1	-	5828644	5828559	DnaJ (pfam00226)→ **RNA**→ hypo→
Pfl-2-28	NC_007492.1	+	6230633	6230715	←hypo **RNA**→ hypo→
env-44	JCVI_SCAF_1096627418547	+	617	700	**RNA**→
Pfl-1-9	NC_004129.6	-	4859973	4859886	hypo→ **RNA**→ hypo→
Ppu-4-17	NZ_AAVY01000010.1	+	155069	155158	hypo→ **RNA**→ hypo (cons)→
Ppu-4-18	NZ_AAVY01000002.1	+	296110	296200	OprD (pfam03573)→ **RNA**→ hypo (cons)→
Pen-1-17	NC_008027.1	+	3966788	3966882	hypo→ **RNA**→ hypo→
Pen-1-18	NC_008027.1	+	3960618	3960708	←NUDIX (pfam00293) **RNA**→ hypo→
Ppu-4-19	NZ_AAVY01000010.1	+	153965	154055	Oligopeptide transporter OPT superfamily→ **RNA**→ hypo (cons)→

Ppu-4-20	NZ_AAVY01000002.1	+	273462	273557	←conserved secreted protein with internal repeats **RNA**→ hypo (cons)→
Pen-1-19	NC_008027.1	+	5582506	5582595	←DUF1234 (pfam06821) **RNA**→ hypo→
Pen-1-20	NC_008027.1	-	910398	910306	←HemA (COG0373) **RNA**→ hypo→
Pen-1-21	NC_008027.1	-	1493171	1493083	←LysR (COG0583) **RNA**→ hypo→
Ppu-3-6	NC_002947.3	+	5390633	5390719	hypo→ **RNA**→ hypo→
Pen-1-22	NC_008027.1	+	3854037	3854123	←hypo **RNA**→ hypo→
Ppu-3-7	NC_002947.3	-	2858046	2857965	hypo→ **RNA**→ yedG domain protein→
Ppu-1-5	NC_009512.1	+	3625062	3625148	hypo→ **RNA**→ hypo→
Ppu-1-6	NC_009512.1	+	4999895	4999981	←PsiE (pfam06146) **RNA**→ hypo→ hypo→
Pen-1-23	NC_008027.1	-	3291941	3291852	hypo→ **RNA**→ hypo→
Pen-1-24	NC_008027.1	+	1180765	1180854	DUF479 (pfam04336)→ **RNA**→ hypo→
Pen-1-25	NC_008027.1	+	4896084	4896172	FecA (COG4772)→ **RNA**→ hypo→
Ppu-4-21	NZ_AAVY01000003.1	+	347373	347461	←ACAD_fadE5 (cd01153) **RNA**→ hypo (cons)→
Ppu-4-22	NZ_AAVY01000026.1	-	47910	47825	hypo (cons)→ **RNA**→ hypo (cons)→
Ppu-4-23	NZ_AAVY01000018.1	+	10366	10453	←lipoprotein, putative **RNA**→ hypo (cons)→
Ppu-1-7	NC_009512.1	-	1439621	1439537	←Rick_17kDa_Anti (pfam05433) **RNA**→ hypo→ hypo→
Ppu-1-8	NC_009512.1	+	1545409	1545496	←Transposase_mut (pfam00872) **RNA**→ hypo→ hypo→
Ppu-3-8	NC_002947.3	-	5149014	5148927	hypo→ **RNA**→ hypo→
Pen-1-26	NC_008027.1	-	2996471	2996385	hypo→ **RNA**→ hypo→ hypo→
Ppu-2-5	NZ_AAXR01000004.1	+	377787	377878	←COG3714 (COG3714) **RNA**→ hypo (cons)→
Ppu-4-24	NZ_AAVY01000008.1	-	10944	10858	Aes (COG0657)→ **RNA**→ hypo (cons)→
Ppu-4-25	NZ_AAVY01000001.1	+	243198	243282	←AbgB (COG1473) **RNA**→ hypo (cons)→
Ppu-3-9	NC_002947.3	-	1995754	1995667	dTDP_sugar_isom (pfam00908)→ **RNA**→ COG1835 (COG1835)→
Pen-1-27	NC_008027.1	+	2592733	2592818	FhuE (COG4773)→ **RNA**→ hypo→ hypo→
Ppu-3-10	NC_002947.3	-	2855914	2855828	←hypo **RNA**→ ←DltE (COG0300)
Ppu-3-11	NC_002947.3	-	1420040	1419951	←Rick_17kDa_Anti (pfam05433) **RNA**→ hypo→
Pen-1-28	NC_008027.1	+	4355591	4355676	←hypo **RNA**→ hypo→ hypo→
Psy-1-6	NC_005773.3	-	2874551	2874467	GAF (pfam01590)→ **RNA**→ COG3617 (COG3617)→
Ppu-4-26	NZ_AAVY01000050.1	-	3054	2970	←hypo **RNA**→ hypo (cons)→ hypo→ hypo (cons)→
Ppu-3-12	NC_002947.3	-	4856630	4856545	←hypo **RNA**→ ←PBPb (smart00062)
Ppu-1-9	NC_009512.1	+	1788565	1788652	hypo→ **RNA**→ hypo→
Ppu-3-13	NC_002947.3	+	5989252	5989335	KefB (COG0475)Kch (COG1226)→ **RNA**→ hypo→
Ppu-1-10	NC_009512.1	+	5795249	5795332	KefB (COG0475)Kch (COG1226)→ **RNA**→ hypo→
Pen-1-29	NC_008027.1	+	5708208	5708293	←VapI (COG3093) **RNA**→ hypo→
env-45	JCVI_SCAF_1096627055672	-	3101	3015	**RNA**→ ? (GOS11222950)→
env-46	JCVI_SCAF_1096628071788	-	497	410	HisM (COG0765)? (GOS4027)→ **RNA**→
Pme-1-1	NC_009439.1	+	3776559	3776646	←hypo **RNA**→ hypo→ hypo→
Pme-1-2	NC_009439.1	-	1667390	1667307	COG2085 (COG2085)→ **RNA**→ hypo→
Rso-2-1	NZ_AAKL01000002.1	-	62829	62742	hypo→ **RNA**→ hypo→ hypo→ DnaG (COG0358)→
Rpi-1-1	NZ_ABDZ01000005.1	-	76964	76878	Ogr_Delta (pfam04606)→ **RNA**→ protein of unknown function DUF1484→ zf-CHC2 (pfam01807)TOPRIM_primases (cd01029)→
env-47	JCVI_SCAF_1096628063291	-	1161	1075	Ogr_Delta (pfam04606)? (GOS14311108)→ **RNA**→
Reu-1-1	NC_008314.1	-	532510	532418	hypo→ **RNA**→ ←hypo
Rso-2-2	NZ_AAKL01000071.1	-	17467	17379	Hypothetical protein→ **RNA**→ Hypothetical Protein→
Reu-2-1	NC_007347.1	+	2423671	2423759	Ogr_Delta (pfam04606)NOG22117 (NOG22117)→ **RNA**→ hypo→ hypo→ DnaG (COG0358)NOG06313 (NOG06313)→
Rso-1-2	NC_003295.1	+	1013264	1013352	Ogr_Delta (pfam04606)NOG22117 (NOG22117)→ **RNA**→ hypo→ hypo→ DnaG (COG0358)NOG06313 (NOG06313)→
Rso-1-3	NC_003295.1	-	3475955	3475871	hypo→ **RNA**→ hypo→
Bth-3-1	NZ_ABBD01000499.1	-	2644	2554	**RNA**→
Bps-8-1	NZ_AAYX01000007.1	+	10966	11055	hypo (cons)→ **RNA**→ hypo (cons)→ hypo (cons)→
Bps-16-1	NC_006350.1	-	1334602	1334513	hypo→ **RNA**→ hypo→
Bps-13-1	NZ_ABBK01000286.1	-	12447	12358	**RNA**→

Bps-11-2	NZ_ABBO01000209.1	+	19106	19197	**RNA**→
Bth-3-2	NZ_ABBD01000806.1	+	5334	5425	**RNA**→
Bth-2-1	NZ_ABBM01000179.1	+	20561	20652	**RNA**→
Bvi-1-1	NC_009256.1	-	463132	463042	hypo→ **RNA**→ hypo→ hypo→ hypo→ hypo→ hypo→
Bth-3-3	NZ_ABBD01000762.1	-	4854	4764	**RNA**→
Bps-16-2	NC_006350.1	-	98311	98220	hypo→ **RNA**→ hypo→ hypo→ hypo→
Bps-10-1	NC_009074.1	-	98244	98153	hypo→ **RNA**→ hypo→ hypo→ hypo→
Bps-4-1	NZ_ABBJ01000023.1	-	9760	9669	**RNA**→
Bps-13-2	NZ_ABBK01000021.1	-	11685	11594	**RNA**→
Bps-7-1	NC_007434.1	-	323597	323506	hypo→ **RNA**→ hypo→
Bps-1-1	NC_009076.1	-	114154	114063	hypo→ **RNA**→ hypo→ hypo→ hypo→
Bps-6-1	NZ_AAHS02000063.1	-	18204	18113	hypo→ **RNA**→ hypo→ hypo→
Bps-17-1	NZ_AAHV02000110.1	-	2933	2842	hypo→ **RNA**→ hypo→ hypo→
Bps-2-1	NZ_AAMB01000054.1	+	10782	10873	hypo→ **RNA**→ hypo→ hypo→
Bps-12-1	NZ_ABBL01000019.1	-	9742	9651	**RNA**→
Bps-14-1	NZ_ABBN01000016.1	-	9766	9675	**RNA**→
Bps-11-3	NZ_ABBO01001513.1	+	7645	7736	**RNA**→
Bps-15-1	NZ_ABBR01000147.1	+	5365	5456	**RNA**→
Bps-8-2	NZ_AAYX01000008.1	+	9285	9376	hypo→ **RNA**→ hypo (cons)→ hypo (cons)→ hypo (cons)→
Bps-5-1	NZ_AAHR02000044.1	-	42873	42782	hypo→ **RNA**→ hypo→ hypo→
Bps-18-1	NZ_AAHW02000026.1	-	10700	10609	hypo→ **RNA**→ hypo→ hypo→
Bok-2-1	NZ_ABBF01000782.1	+	19995	20086	**RNA**→
Bok-1-1	NZ_ABBG01000072.1	+	36101	36192	**RNA**→
Bce-1-1	NZ_AAVA01000001.1	-	268317	268226	←hypo **RNA**→ hypo (cons)→
Bps-5-2	NZ_AAHR02000018.1	-	20965	20874	←PinR (COG1961) **RNA**→ hypo→ hypo→
Bth-1-2	NZ_ABBH01000753.1	+	1638	1728	**RNA**→
Bps-7-2	NC_007434.1	+	1344790	1344879	hypo→ **RNA**→ hypo→ COG4643 (COG4643)COG5519 (COG5519)→
Bps-5-3	NZ_AAHR02000028.1	-	56606	56517	←hypo **RNA**→ hypo→ hypo→ hypo→ hypo→ COG4643 (COG4643)DUF927 (pfam06048)→ COG5519 (COG5519)→
Bps-6-2	NZ_AAHS02000008.1	+	100667	100756	←hypo **RNA**→ hypo→ hypo→ hypo→ hypo→ COG4643 (COG4643)COG5519 (COG5519)→
Bps-17-2	NZ_AAHV02000008.1	-	73985	73896	←hypo **RNA**→ hypo→ hypo→ hypo→ hypo→
Bdo-1-2	NZ_AAKY01000125.1	+	82257	82346	hypo→ **RNA**→ hypo→ hypo→ hypo→ hypo→
Bps-9-1	NZ_AAMM02000028.1	+	49678	49767	←Transposase_8 (pfam01527) **RNA**→ hypo→ hypo→ hypo→ hypo→
Reu-1-2	NC_008314.1	+	273926	274013	CheZ (COG3143)→ **RNA**→ hypo→
Cvi-1-2	NC_005085.1	-	675562	675477	←hypo **RNA**→ hypo→ hypo→ zf-CHC2 (pfam01807)→
Bps-15-2	NZ_ABBR01001020.1	-	7221	7131	**RNA**→
Rpi-2-1	NZ_AAWK01000002.1	+	203495	203583	←hypo (cons) **RNA**→ hypo→
Rpi-1-2	NZ_ABDZ01000003.1	+	406374	406462	←putative cytoplasmic protein **RNA**→ hypo→
env-48	JCVI_SCAF_1096628075889	-	1758	1670	**RNA**→
env-49	sludgePhrap_Contig14725	+	1993	2074	**RNA**→
Bce-1-2	NZ_AAVA01000004.1	+	244809	244900	hypo→ **RNA**→ repA (cd01125)→
Bma-10-1	NC_008785.1	-	443372	443281	←CorB (COG4536) **RNA**→ Transposase_8 (pfam01527)→ rve (pfam00665)→
Bma-8-1	NC_009080.1	+	3220377	3220468	←CorB (COG4536) **RNA**→ Transposase_8 (pfam01527)→ rve (pfam00665)→
Bma-7-1	NC_008836.1	-	1333874	1333783	←CorB (COG4536) **RNA**→ Transposase_8 (pfam01527)→ rve (pfam00665)→
Bma-3-1	NC_006348.1	-	2633521	2633430	←CorB (COG4536) **RNA**→ Transposase_8 (pfam01527)→
Bma-2-1	NZ_AAHN02000004.1	+	44523	44614	←CorB (COG4536) **RNA**→ Transposase_8 (pfam01527)→ rve (pfam00665)→
Bma-5-1	NZ_AAHO01000003.1	-	120218	120127	←CorB (COG4536) **RNA**→ Transposase_8 (pfam01527)→ rve (pfam00665)→
Bma-4-1	NZ_AAIQ02000005.1	-	72568	72477	←CorB (COG4536) **RNA**→ Transposase_8 (pfam01527)→ rve (pfam00665)→
Bma-6-1	NZ_AAIR02000007.1	-	72522	72431	←CorB (COG4536) **RNA**→ Transposase_8 (pfam01527)→ rve (pfam00665)→
Bma-1-1	NZ_AANX02000033.1	+	54491	54582	←CorB (COG4536) **RNA**→ ←rve (pfam00665)
Bma-9-1	NZ_AAZP01000002.1	-	127546	127455	**RNA**→
Bth-2-2	NZ_ABBM01000536.1	-	13915	13824	**RNA**→

Bvi-1-2	NC_009254.1	-	633610	633518	←hypo **RNA**→ hypo→ CysI (COG0155)→ DUF934 (pfam06073)→
env-50	JCVI_SCAF_1096627390482	+	71935	72027	←COG1848 (COG1848)? (GOS13071826) **RNA**→ ←INT_SG1_C (cd01183)? (GOS13019564)
Bce-2-1	NZ_AAKX01000021.1	-	71214	71126	hypo→ **RNA**→ COG0840: Methyl-accepting chemotaxis protein→
Bps-3-1	NZ_ABBP01000530.1	-	16827	16739	**RNA**→
Bmu-1-1	NZ_AAVB01000029.1	+	2859	2950	TrfA (pfam07042)→ **RNA**→ hypo (cons)→
Bvi-1-3	NC_009227.1	-	216892	216802	hypo→ **RNA**→ hypo→ hypo→ hypo→
env-51	JCVI_SCAF_1096627389595	-	53519	53428	? (GOS14593241)→ **RNA**→ COG5642 (COG5642)? (GOS11073445)→ COG5654 (COG5654)? (GOS10365)→
Rso-1-4	NC_003295.1	-	2052344	2052257	hypo→ **RNA**→ hypo→
Rpi-2-2	NZ_AAWK01000002.1	-	1178	1090	←hypo (cons) **RNA**→ hypo→
Rpi-1-3	NZ_ABDZ01000003.1	-	186473	186388	hypo→ **RNA**→ hypo→
Rso-2-3	NZ_AAKL01000003.1	-	1035	948	←Hypothetical Protein **RNA**→ Hypothetical Protein→
Rso-1-5	NC_003295.1	-	3705084	3704998	hypo→ **RNA**→ hypo→ hypo→ hypo→
Rpi-1-4	NZ_ABDZ01000005.1	-	62708	62621	hypo→ **RNA**→ hypo→
Rpi-2-3	NZ_AAWK01000008.1	-	223652	223566	COG2105 (COG2105)→ **RNA**→ hypo→
Rpi-1-5	NZ_ABDZ01000027.1	-	11058	10972	COG2105 (COG2105)→ **RNA**→ hypo→
Rpi-2-4	NZ_AAWK01000002.1	-	155352	155263	←COG2944 (COG2944) **RNA**→ hypo (cons)→
Rpi-1-6	NZ_ABDZ01000003.1	-	358829	358740	←hypo (cons) **RNA**→ hypo (cons)→ hypo (cons)→
Rpi-2-5	NZ_AAWK01000002.1	+	340004	340091	←type I restriction system endonuclease **RNA**→ hypo→
env-52	JCVI_SCAF_1101667386172	+	658	747	**RNA**→
Pen-1-30	NC_008027.1	+	2692820	2692908	COG2863 (COG2863)→ **RNA**→ hypo→
Ppu-4-27	NZ_AAVY01000024.1	+	72189	72273	FhuE (COG4773)→ **RNA**→ hypo (cons)→ hypo→
Pfl-2-29	NC_007492.1	-	4210726	4210638	←COG3950 (COG3950) **RNA**→ hypo→
Sgl-2-16	NC_007712.1	-	1219801	1219714	KilA-N (pfam04383)NOG17190 (NOG17190)→ **RNA**→ DnaC (COG1484)→
Sgl-2-17	NC_007712.1	+	444749	444836	←Transposase_9 (pfam01548)Transposase_20 (pfam02371) **RNA**→ ←Phage_portal (pfam04860)
Eco-8-5	NZ_AAJW01000085.1	-	4756	4672	hypo→ **RNA**→ hypo→ hypo→
Eco-3-14	NZ_AAKB01000008.1	-	102388	102303	hypo→ **RNA**→ hypo→
Sen-3-1	NZ_ABFG01000007.1	+	330034	330119	**RNA**→
Sgl-2-18	NC_007712.1	+	1663464	1663548	←Porin_1 (pfam00267) **RNA**→ putative phage replication protein→
Eph-3-1	NC_005856.1	+	48337	48422	**RNA**→
Sen-8-4	NZ_ABEW01000015.1	-	43415	43331	**RNA**→
Aba-1-6	NC_009085.1	-	3754463	3754380	←HTH_ARAC (smart00342) **RNA**→ putative transcriptional regulator→
Aba-1-7	NC_009085.1	-	1620947	1620860	surface antigen→ **RNA**→ ←LysR (COG0583)
Aba-1-8	NC_009085.1	+	1471058	1471146	HtrB (COG1560)→ **RNA**→ TetR_N (pfam00440)→
Aba-1-9	NC_009085.1	-	3063911	3063825	←hypo **RNA**→ TetR_N (pfam00440)→
Asp-1-1	NC_005966.1	+	2750060	2750143	←hypo **RNA**→ hypo→ COG3152 (COG3152)→
Aba-1-10	NC_009085.1	+	1261816	1261900	NlpA (COG1464)→ **RNA**→ AcrR (COG1309)→
Aba-1-11	NC_009085.1	-	2432140	2432056	←Lrp (COG1522) **RNA**→ putative transcriptional regulator→
Aba-1-12	NC_009085.1	+	1493304	1493388	EriC (COG0038)→ **RNA**→ putative transcriptional regulator→
Aba-1-13	NC_009085.1	-	3419053	3418968	←hypo **RNA**→ hypo→
Aba-1-14	NC_009085.1	+	872109	872195	Flavin_Reduct (pfam01613)→ **RNA**→ TetR_N (pfam00440)→
Aba-1-15	NC_009085.1	+	590336	590422	COG5001 (COG5001)→ **RNA**→ putative transcriptional regulator→
Aba-1-16	NC_009085.1	-	2114540	2114456	COG0385 (COG0385)→ **RNA**→ putative transcriptional regulator→
Aba-1-17	NC_009085.1	-	2055349	2055265	HTH_XRE (cd00093)→ **RNA**→ putative transcriptional regulator→
Aba-1-18	NC_009085.1	+	1581044	1581128	PaaI_thioesterase (cd03443)→ **RNA**→ TetR_N (pfam00440)→
Aba-1-19	NC_009085.1	-	2778777	2778689	MoeA_N (pfam03453)→ **RNA**→ AcrR (COG1309)→
Aba-1-20	NC_009085.1	-	3904932	3904848	←RhtB (COG1280) **RNA**→ ←AgaS (COG2222)
Asp-1-2	NC_005966.1	+	962260	962346	←Acetyltransf_1 (pfam00583) **RNA**→ hypo→
Aba-1-21	NC_009085.1	-	1060224	1060139	Sugar_tr (pfam00083)→ **RNA**→ hypo→
Asp-1-3	NC_005966.1	+	3357993	3358079	←hypo **RNA**→ COG3152 (COG3152)→
Asp-1-4	NC_005966.1	-	920344	920258	←FhaC (COG2831) **RNA**→ ←hypo
Aba-1-22	NC_009085.1	-	98547	98462	←KAS_I_II (cd00834) **RNA**→ ←COG3981 (COG3981)
Aba-1-23	NC_009085.1	+	1469454	1469538	HtrB (COG1560)→ **RNA**→ TetR_N (pfam00440)→

Hin-11-1	NZ_AADP01000001.1	+	917452	917538	COG0419: ATPase involved in DNA repair→ **RNA**→ COG3561 (COG3561)→
Hin-1-1	NZ_AAZD01000004.1	-	76634	76548	hypo→ **RNA**→ COG3561 (COG3561)→
Hin-9-1	NZ_AAZI01000003.1	-	134972	134886	hypo→ **RNA**→ COG3561 (COG3561)→
Hin-7-2	NC_009567.1	+	169172	169257	hypo→ **RNA**→ COG3561 (COG3561)→
Hin-10-1	NZ_AADO01000003.1	+	1933	2018	COG1201: Lhr-like helicases→ **RNA**→ COG3561 (COG3561)→
Hin-12-1	NZ_AAZE01000005.1	+	1255	1340	hypo→ **RNA**→ COG3561 (COG3561)→
Hin-3-3	NZ_AAZF01000003.1	-	38900	38815	hypo→ **RNA**→ COG3561 (COG3561)→
Hin-5-2	NZ_AAZG01000005.1	-	122796	122711	fumarate/nitrate reduction transcriptional regulator→ **RNA**→ COG3561 (COG3561)→
Hin-8-1	NZ_AAZH01000001.1	+	84396	84481	hypo→ **RNA**→ ←hypo
Hin-4-1	NC_007146.1	-	1576614	1576529	hypo→ **RNA**→ COG3561 (COG3561)→
Hin-6-1	NC_009566.1	+	965902	965987	hypo→ **RNA**→ COG3561 (COG3561)→
Hph-1-1	NC_004827.1	-	6364	6280	**RNA**→
Hso-1-2	NC_008309.1	+	1602795	1602876	hypo→ **RNA**→ COG3617 (COG3617)→ Haemophilus-specific protein, uncharacterized→ Haemophilus-specific protein, uncharacterized→ Haemophilus-specific protein, uncharacterized→ RecT (pfam03837)→ hypo→ hypo→ DNA_methylase (pfam00145)→ Haemophilus-specific protein, uncharacterized→ AlpA (COG3311)→
Aba-1-24	NC_009085.1	+	871520	871603	Flavin_Reduct (pfam01613)→ **RNA**→ TetR_N (pfam00440)→

16.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 16.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one c4 antisense RNA RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

cd00093 (5) Helix-turn-helix XRE-family like proteins.
cd00189 (1) Tetratricopeptide repeat domain; typically contains 34 amino acids [WLF]-X(2)-[LIM]-[GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-X(2)-[ASL]-X(4)-[PKE]
cd00221 (1) Very Short Patch Repair (Vsr) Endonuclease.
cd00254 (1) Lytic Transglycosylase (LT) and Goose Egg White Lysozyme (GEWL) domain.
cd00797 (1) Phage HP1 integrase, C-terminal catalytic domain.
cd00800 (3) Lambda integrase, C-terminal catalytic domain.
cd00801 (6) Bacteriophage P4 integrase.
cd00834 (1) Beta-ketoacyl-acyl carrier protein (ACP) synthase (KAS), type I and II.
cd01029 (6) TOPRIM_primases: The topoisomerase-primase (TORPIM) nucleotidyl transferase/hydrolase domain found in the active site regions of bacterial DnaG-type primases and their homologs.
cd01125 (10) Hexameric Replicative Helicase RepA.
cd01153 (1) Putative acyl-CoA dehydrogenase (ACAD).
cd01163 (1) Dibenzothiophene (DBT) desulfurization enzyme C (DszC).
cd01183 (1) INT_SG1, DNA breaking-rejoining enzymes, integrase/recombinases subgroup 1, C-terminal catalytic domain.
cd01406 (1) Sir2-like: Prokaryotic group of uncharacterized Sir2-like proteins which lack certain key catalytic residues and conserved zinc binding cysteines; and are members of the SIR2 superfamily of proteins, silent information regulator 2 (Sir2) enzymes which catalyze NAD+-dependent protein/histone deacetylation.
cd01825 (1) SGNH_peri1; putative periplasmic member of the SGNH-family of hydrolases, a diverse family of lipases and esterases.
cd01949 (1) Diguanylate-cyclase (DGC) or GGDEF domain: Originally named after a conserved residue pattern, and initially described as domain of unknown function 1 (DUF1).
cd02932 (2) Old yellow enzyme (OYE) YqjM-like FMN binding domain.
cd03227 (1) ABC-type Class 2 contains systems involved in cellular processes other than transport.
cd03443 (1) PaaI_thioesterase is a tetrameric acyl-CoA thioesterase with a hot dog fold and one of several proteins responsible for phenylacetic acid (PA) degradation in bacteria.
cd03449 (1) (R)-hydratase [(R)-specific enoyl-CoA hydratase]
COG0009 (3) Putative translation factor (SUA5) [Translation, ribosomal structure and biogenesis]
COG0038 (2) Chloride channel protein EriC [Inorganic ion transport and metabolism]
COG0155 (1) Sulfite reductase, beta subunit (hemoprotein) [Inorganic ion transport and metabolism]
COG0247 (1) Fe-S oxidoreductase [Energy production and conversion]
COG0277 (1) FAD/FMN-containing dehydrogenases [Energy production and conversion]
COG0300 (1) Short-chain dehydrogenases of various substrate specificities [General function prediction only]
COG0330 (1) Membrane protease subunits, stomatin/prohibitin homologs [Posttranslational modification, protein turnover, chaperones]
COG0358 (3) DNA primase (bacterial type) [DNA replication, recombination, and repair]
COG0373 (1) Glutamyl-tRNA reductase [Coenzyme metabolism]
COG0381 (1) UDP-N-acetylglucosamine 2-epimerase [Cell envelope biogenesis, outer membrane]
COG0385 (1) Predicted Na+-dependent transporter [General function prediction only]
COG0475 (3) Kef-type K+ transport systems, membrane components [Inorganic ion transport and metabolism]

COG0583 (3) Transcriptional regulator [Transcription]
COG0624 (1) Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase and related deacylases [Amino acid transport and metabolism]
COG0629 (3) Single-stranded DNA-binding protein [DNA replication, recombination, and repair]
COG0657 (2) Esterase/lipase [Lipid metabolism]
COG0673 (1) Predicted dehydrogenases and related proteins [General function prediction only]
COG0683 (1) ABC-type branched-chain amino acid transport systems, periplasmic component [Amino acid transport and metabolism]
COG0691 (1) tmRNA-binding protein [Posttranslational modification, protein turnover, chaperones]
COG0702 (1) Predicted nucleoside-diphosphate-sugar epimerases [Cell envelope biogenesis, outer membrane / Carbohydrate transport and metabolism]
COG0715 (2) ABC-type nitrate/sulfonate/bicarbonate transport systems, periplasmic components [Inorganic ion transport and metabolism]
COG0765 (1) ABC-type amino acid transport system, permease component [Amino acid transport and metabolism]
COG0840 (1) Methyl-accepting chemotaxis protein [Cell motility and secretion / Signal transduction mechanisms]
COG0845 (2) Membrane-fusion protein [Cell envelope biogenesis, outer membrane]
COG1027 (1) Aspartate ammonia-lyase [Amino acid transport and metabolism]
COG1044 (1) UDP-3-O-[3-hydroxymyristoyl] [3-hydroxymyristoyl]
COG1061 (1) DNA or RNA helicases of superfamily II [Transcription / DNA replication, recombination, and repair]
COG1075 (1) Predicted acetyltransferases and hydrolases with the alpha/beta hydrolase fold [General function prediction only]
COG1226 (3) Kef-type K+ transport systems, predicted NAD-binding component [Inorganic ion transport and metabolism]
COG1280 (2) Putative threonine efflux protein [Amino acid transport and metabolism]
COG1309 (2) Transcriptional regulator [Transcription]
COG1462 (1) Uncharacterized protein involved in formation of curli polymers [Cell envelope biogenesis, outer membrane]
COG1464 (1) ABC-type metal ion transport system, periplasmic component/surface antigen [Inorganic ion transport and metabolism]
COG1473 (1) Metal-dependent amidase/aminoacylase/carboxypeptidase [General function prediction only]
COG1476 (1) Predicted transcriptional regulators [Transcription]
COG1484 (2) DNA replication protein [DNA replication, recombination, and repair]
COG1522 (1) Transcriptional regulators [Transcription]
COG1555 (1) DNA uptake protein and related DNA-binding proteins [DNA replication, recombination, and repair]
COG1560 (2) Lauroyl/myristoyl acyltransferase [Cell envelope biogenesis, outer membrane]
COG1595 (1) DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog [Transcription]
COG1670 (2) Acetyltransferases, including N-acetylases of ribosomal proteins [Translation, ribosomal structure and biogenesis]
COG1721 (1) Uncharacterized conserved protein (some members contain a von Willebrand factor type A (vWA) domain) [General function prediction only]
COG1724 (1) Predicted periplasmic or secreted lipoprotein [Cell motility and secretion]
COG1835 (1) Predicted acyltransferases [Lipid metabolism]
COG1846 (1) Transcriptional regulators [Transcription]
COG1848 (1) Predicted nucleic acid-binding protein, contains PIN domain [General function prediction only]
COG1943 (1) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
COG1961 (1) Site-specific recombinases, DNA invertase Pin homologs [DNA replication, recombination, and repair]
COG1981 (2) Predicted membrane protein [Function unknown]
COG2026 (6) Cytotoxic translational repressor of toxin-antitoxin stability system [Translation, ribosomal structure and biogenesis / Cell division and chromosome partitioning]
COG2085 (1) Predicted dinucleotide-binding enzymes [General function prediction only]
COG2105 (2) Uncharacterized conserved protein [Function unknown]
COG2188 (1) Transcriptional regulators [Transcription]
COG2197 (1) Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain [Signal transduction mechanisms / Transcription]
COG2222 (1) Predicted phosphosugar isomerases [Cell envelope biogenesis, outer membrane]
COG2226 (1) Methylase involved in ubiquinone/menaquinone biosynthesis [Coenzyme metabolism]
COG2807 (3) Cyanate permease [Inorganic ion transport and metabolism]
COG2831 (1) Hemolysin activation/secretion protein [Intracellular trafficking and secretion]
COG2863 (1) Cytochrome c553 [Energy production and conversion]
COG2896 (1) Molybdenum cofactor biosynthesis enzyme [Coenzyme metabolism]
COG2944 (2) Predicted transcriptional regulator [Transcription]
COG2961 (3) Protein involved in catabolism of external DNA [General function prediction only]
COG3093 (2) Plasmid maintenance system antidote protein [General function prediction only]
COG3143 (1) Chemotaxis protein [Cell motility and secretion / Signal transduction mechanisms]
COG3152 (2) Predicted membrane protein [Function unknown]
COG3177 (1) Uncharacterized conserved protein [Function unknown]
COG3311 (5) Predicted transcriptional regulator [Transcription]
COG3335 (1) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
COG3378 (8) Predicted ATPase [General function prediction only]
COG3415 (1) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
COG3423 (6) Predicted transcriptional regulator [Transcription]
COG3547 (1) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
COG3561 (17) Phage anti-repressor protein [Transcription]
COG3617 (29) Prophage antirepressor [Transcription]
COG3645 (1) Uncharacterized phage-encoded protein [Function unknown]
COG3646 (56) Uncharacterized phage-encoded protein [Function unknown]
COG3666 (1) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
COG3714 (1) Predicted membrane protein [Function unknown]
COG3781 (1) Predicted membrane protein [Function unknown]
COG3866 (1) Pectate lyase [Carbohydrate transport and metabolism]
COG3941 (1) Mu-like prophage protein [General function prediction only]
COG3950 (1) Predicted ATP-binding protein involved in virulence [General function prediction only]

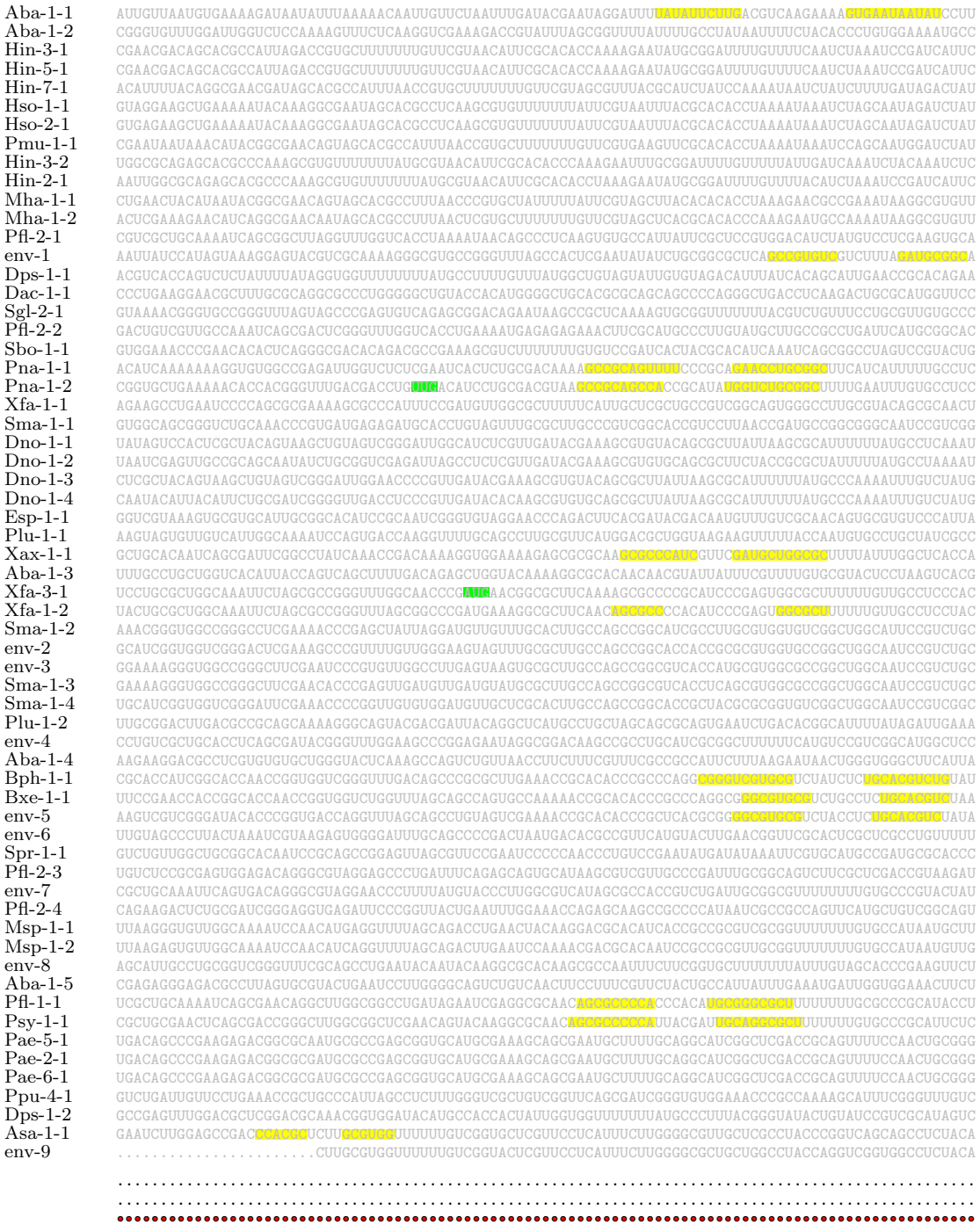
COG3981 (1) Predicted acetyltransferase [General function prediction only]
COG4104 (1) Uncharacterized conserved protein [Function unknown]
COG4160 (1) ABC-type arginine/histidine transport system, permease component [Amino acid transport and metabolism]
COG4178 (1) ABC-type uncharacterized transport system, permease and ATPase components [General function prediction only]
COG4487 (1) Uncharacterized protein conserved in bacteria [Function unknown]
COG4536 (9) Putative Mg2+ and Co2+ transporter CorB [Inorganic ion transport and metabolism]
COG4625 (1) Uncharacterized protein with a C-terminal OMP (outer membrane protein) domain [Function unknown]
COG4643 (21) Uncharacterized protein conserved in bacteria [Function unknown]
COG4680 (1) Uncharacterized protein conserved in bacteria [Function unknown]
COG4700 (1) Uncharacterized protein conserved in bacteria containing a divergent form of TPR repeats [Function unknown]
COG4748 (1) Uncharacterized conserved protein [Function unknown]
COG4771 (1) Outer membrane receptor for ferrienterochelin and colicins [Inorganic ion transport and metabolism]
COG4772 (1) Outer membrane receptor for Fe3+-dicitrate [Inorganic ion transport and metabolism]
COG4773 (2) Outer membrane receptor for ferric coprogen and ferric-rhodotorulic acid [Inorganic ion transport and metabolism]
COG4993 (1) Glucose dehydrogenase [Carbohydrate transport and metabolism]
COG5001 (1) Predicted signal transduction protein containing a membrane domain, an EAL and a GGDEF domain [Signal transduction mechanisms]
COG5283 (1) Phage-related tail protein [Function unknown]
COG5519 (9) Superfamily II helicase and inactivated derivatives [DNA replication, recombination, and repair]
COG5606 (1) Uncharacterized conserved small protein [Function unknown]
COG5642 (1) Uncharacterized conserved protein [Function unknown]
COG5654 (1) Uncharacterized conserved protein [Function unknown]
COG5659 (1) FOG: Transposase [DNA replication, recombination, and repair]
GOS10365 (1) ?
GOS1100891 (2) ?
GOS11069324 (1) ?
GOS11073445 (1) ?
GOS11222950 (1) ?
GOS11335753 (2) ?
GOS11363276 (1) ?
GOS11386653 (1) ?
GOS11725977 (1) ?
GOS11900477 (1) ?
GOS12118414 (1) ?
GOS13019564 (1) ?
GOS13071826 (1) ?
GOS13412611 (1) ?
GOS13828 (1) ?
GOS14070030 (1) ?
GOS14311108 (1) ?
GOS14593241 (1) ?
GOS1771 (1) ?
GOS255 (1) ?
GOS26 (1) ?
GOS3431 (1) ?
GOS3774 (1) ?
GOS4027 (1) ?
GOS4294 (1) ?
NOG06080 (1)
NOG06313 (2)
NOG08008 (1)
NOG08737 (1)
NOG13495 (2)
NOG15906 (1)
NOG17190 (2)
NOG17530 (2)
NOG18143 (1)
NOG18702 (1)
NOG19993 (1)
NOG22115 (1)
NOG22117 (2)
pfam00078 (1) Reverse transcriptase (RNA-dependent DNA polymerase).
pfam00083 (1) Sugar (and other) transporter.
pfam00106 (1) short chain dehydrogenase.
pfam00126 (1) Bacterial regulatory helix-turn-helix protein, lysR family.
pfam00145 (2) C-5 cytosine-specific DNA methylase.
pfam00226 (3) DnaJ domain.
pfam00232 (1) Glycosyl hydrolase family 1.
pfam00267 (1) Gram-negative porin.
pfam00293 (1) NUDIX domain.
pfam00440 (8) Bacterial regulatory proteins, tetR family.
pfam00496 (1) Bacterial extracellular solute-binding proteins, family 5.
pfam00529 (1) HlyD family secretion protein.
pfam00535 (1) Glycosyl transferase.
pfam00583 (4) Acetyltransferase (GNAT) family.
pfam00665 (10) Integrase core domain.
pfam00872 (1) Transposase, Mutator family.
pfam00873 (1) AcrB/AcrD/AcrF family.
pfam00908 (1) dTDP-4-dehydrorhamnose 3,5-epimerase.
pfam01385 (1) Probable transposase.
pfam01527 (10) Transposase.
pfam01548 (1) Transposase.
pfam01590 (1) GAF domain.
pfam01613 (4) Flavin reductase like domain.
pfam01726 (2) LexA DNA binding domain.
pfam01807 (3) CHC2 zinc finger.
pfam01904 (1) Protein of unknown function DUF72.
pfam01914 (2) MarC family integral membrane protein.
pfam02371 (1) Transposase IS116/IS110/IS902 family.
pfam02413 (1) Caudovirales tail fibre assembly protein.
pfam02447 (1) GntP family permease.
pfam03230 (1) Antirestriction protein.
pfam03288 (1) Poxvirus D5 protein-like.

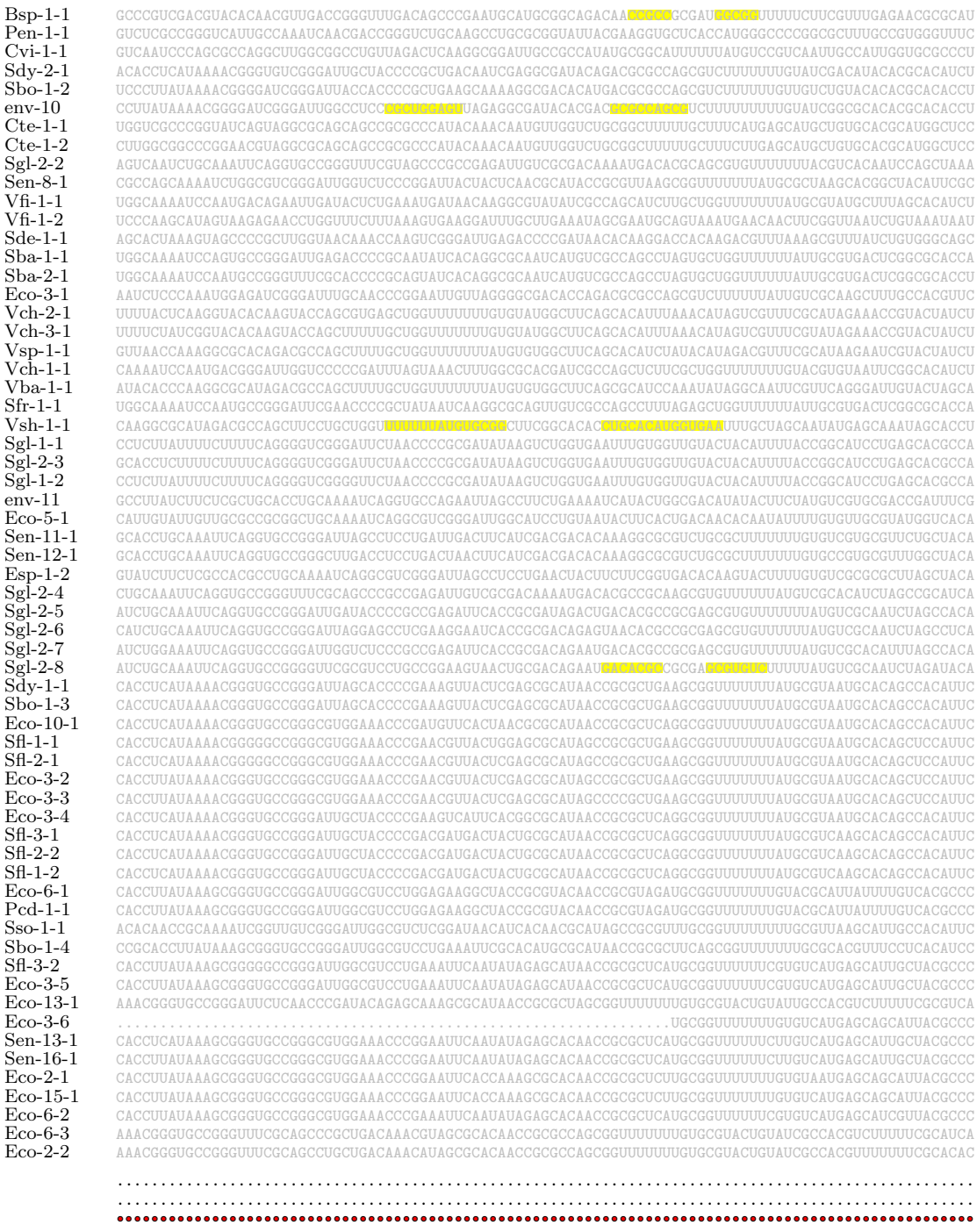
pfam03374 (4) Phage antirepressor protein.
pfam03400 (2) IS1 transposase.
pfam03453 (1) MoeA N-terminal region (domain I and II).
pfam03466 (1) LysR substrate binding domain.
pfam03573 (2) outer membrane porin, OprD family.
pfam03837 (1) RecT family.
pfam03942 (1) DTW domain.
pfam04336 (1) Protein of unknown function, DUF479.
pfam04383 (5) KilA-N domain.
pfam04606 (4) Phage transcriptional activator, Ogr/Delta.
pfam04754 (4) Putative transposase, YhgA-like.
pfam04860 (1) Phage portal protein.
pfam05015 (1) Plasmid maintenance system killer protein.
pfam05125 (1) Phage major capsid protein, P2 family.
pfam05272 (2) Virulence-associated protein E.
pfam05433 (2) Rickettsia 17 kDa surface antigen.
pfam05534 (1) HicB family.
pfam05866 (4) Endodeoxyribonuclease RusA.
pfam05869 (4) DNA N-6-adenine-methyltransferase (Dam).
pfam05930 (2) Prophage CP4-57 regulatory protein (AlpA).
pfam05973 (1) Bacterial protein of unknown function (DUF891).
pfam05974 (2) Protein of unknown function (DUF892).
pfam06048 (1) Domain of unknown function (DUF927).
pfam06073 (1) Bacterial protein of unknown function (DUF934).
pfam06146 (1) Phosphate-starvation-inducible E.
pfam06406 (4) StbA protein.
pfam06616 (1) BsuBI/PstI restriction endonuclease C-terminus.
pfam06821 (1) Protein of unknown function (DUF1234).
pfam07042 (1) TrfA protein.
pfam07081 (1) Protein of unknown function (DUF1349).
pfam07282 (1) Putative transposase DNA-binding domain.
smart00062 (1) Bacterial periplasmic substrate-binding proteins; bacterial proteins, eukaryotic ones are in PBPe
smart00342 (1) helix_turn_helix, arabinose operon control protein;
smart00421 (5) helix_turn_helix, Lux Regulon; lux regulon (activates the bioluminescence operon
smart00482 (1) DNA polymerase A domain;
smart00530 (2) Helix-turn-helix XRE-family like proteins ;

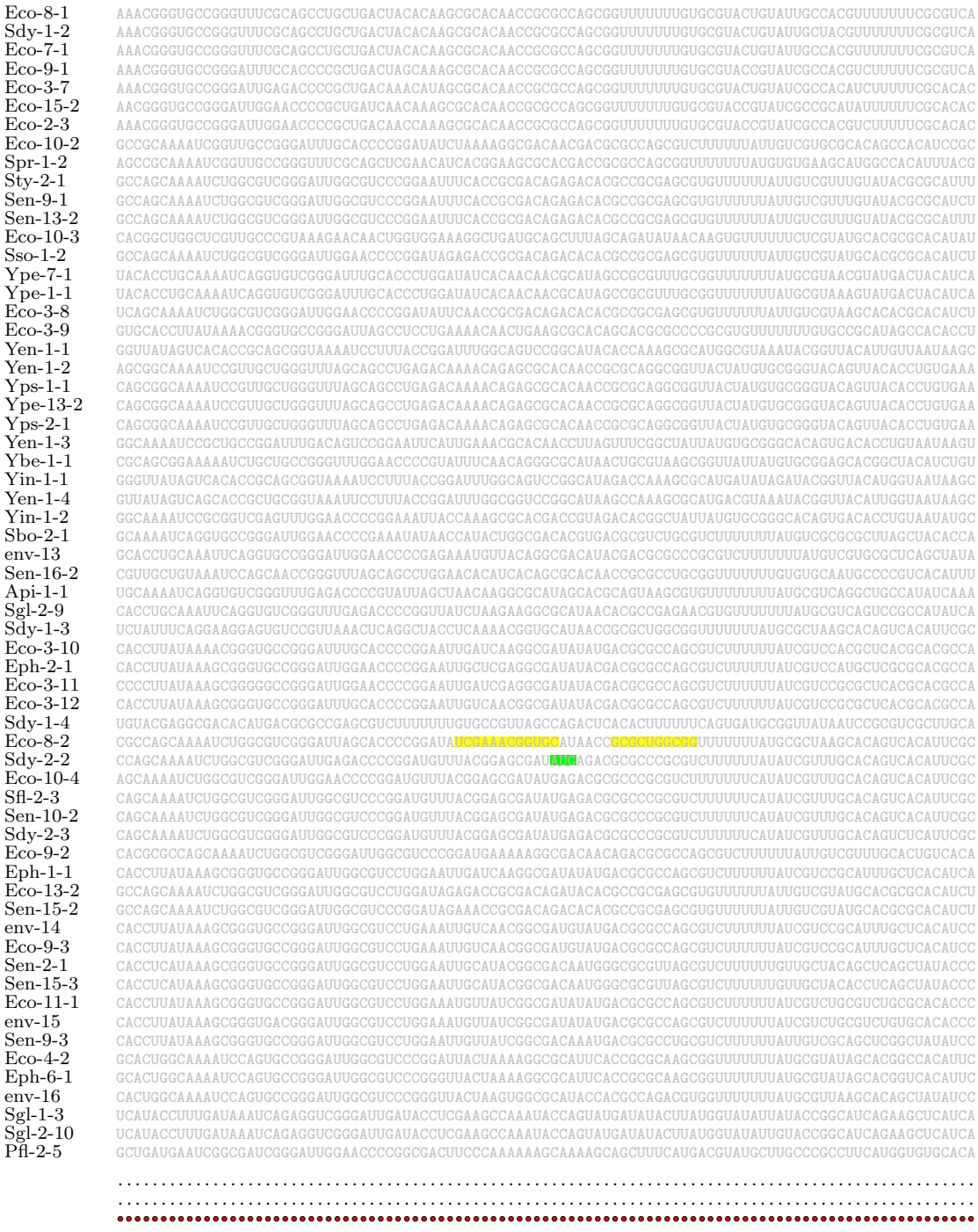
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Bma-1-1, Bma-2-1, Bma-3-1, Bma-4-1, Bma-5-1, Bma-6-1, Bma-7-1, Bma-8-1, Bma-9-1, Bps-1-1, Bps-10-1, Bps-11-3, Bps-12-1, Bps-13-1, Bps-13-2, Bps-14-1, Bps-15-1, Bps-17-1, Bps-17-2, Bps-2-1, Bps-4-1, Bps-5-3, Bps-6-1, Bps-6-2, Eco-12-1, Eco-12-2, Eco-4-1, Eco-5-3, Eco-6-4, Eco-8-3, Eph-5-1, Hin-1-1, Hin-5-2, Hin-9-1, PGi-1-1, Pae-1-1, Pae-4-1, Sco-1-1, Sen-10-1, Sen-10-3, Sen-14-1, Sen-15-1, Sen-4-1, Sen-4-2, Sen-4-3, Sen-4-5, Sen-6-1, Sen-7-1, Sen-8-2, Sen-9-2, Sgl-1-4, Sph-1-1, Xfa-2-1, Xor-2-1, Ype-1-2, Ype-1-3, Ype-10-1, Ype-10-2, Ype-10-3, Ype-11-1, Ype-11-2, Ype-12-1, Ype-12-2, Ype-12-3, Ype-13-1, Ype-13-3, Ype-14-1, Ype-14-2, Ype-14-3, Ype-15-1, Ype-15-2, Ype-15-3, Ype-16-1, Ype-16-2, Ype-16-3, Ype-2-1, Ype-3-1, Ype-3-2, Ype-3-3, Ype-4-1, Ype-4-2, Ype-4-3, Ype-5-1, Ype-5-2, Ype-5-3, Ype-6-1, Ype-6-2, Ype-6-3, Ype-7-2, Ype-7-3, Ype-8-1, Ype-8-2, Ype-8-3, Ype-9-1, Ype-9-2, Ype-9-3, env-12

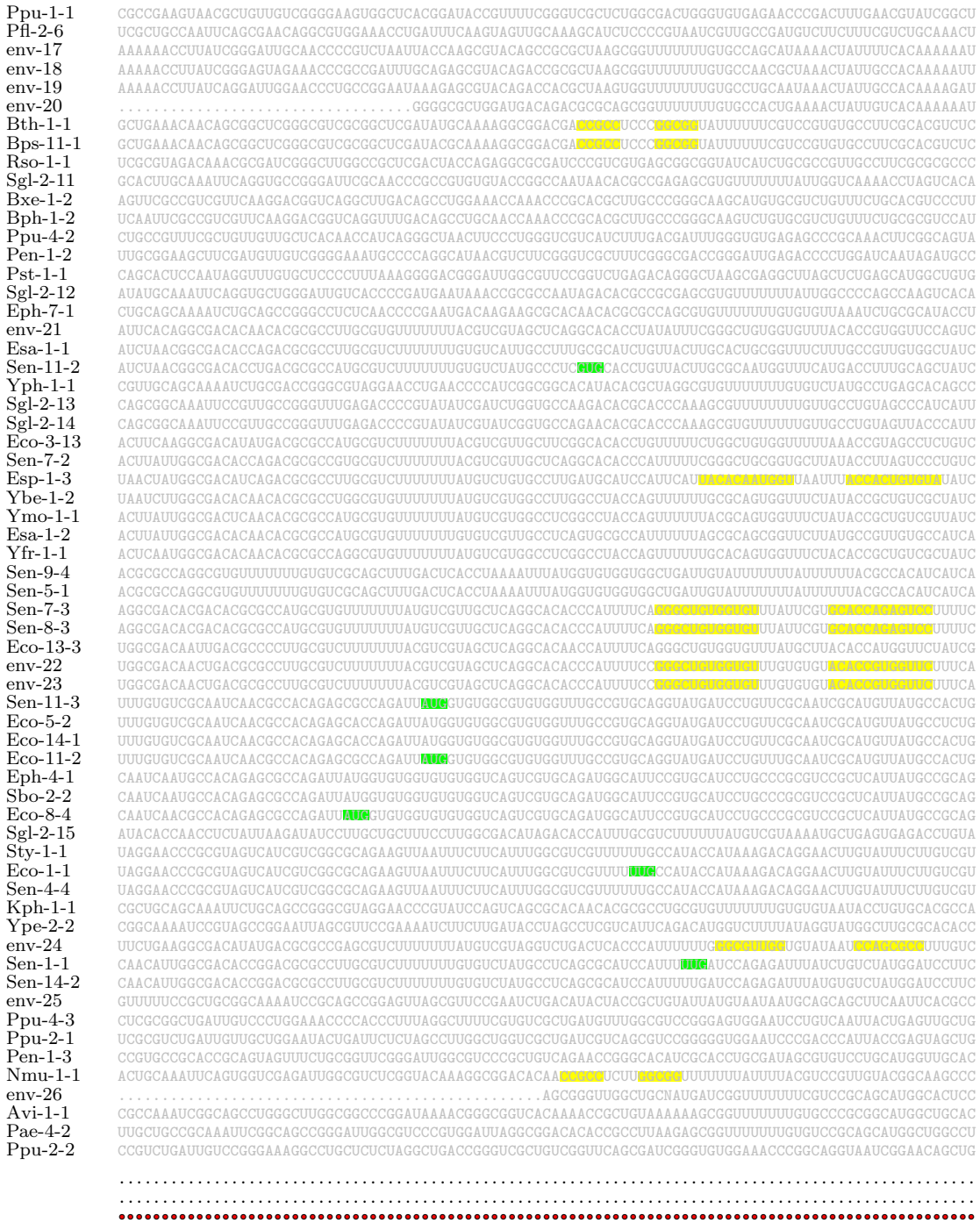
16.4 Multiple-sequence alignment

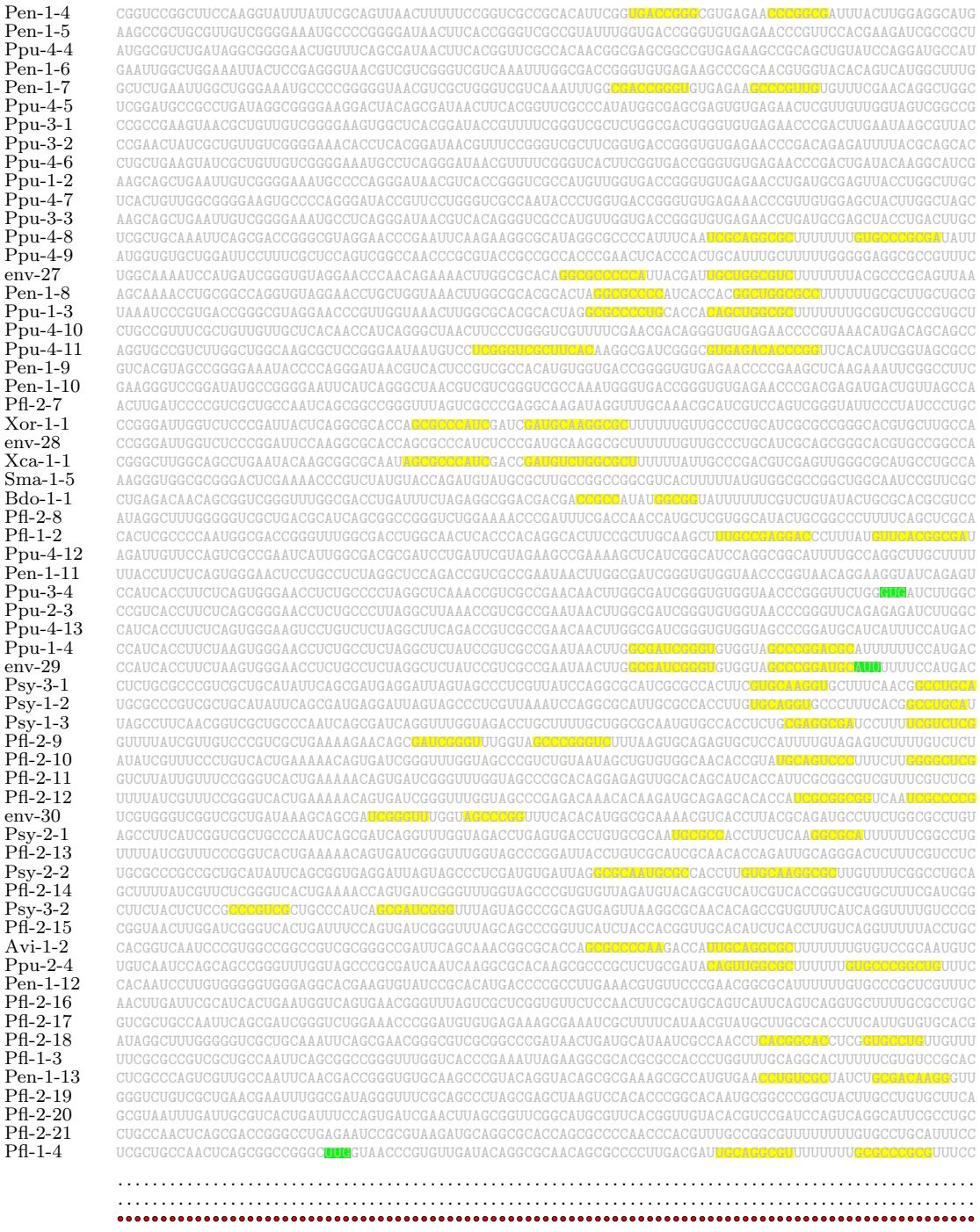
Each c4 antisense RNA RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 16.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The c4 antisense RNA RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

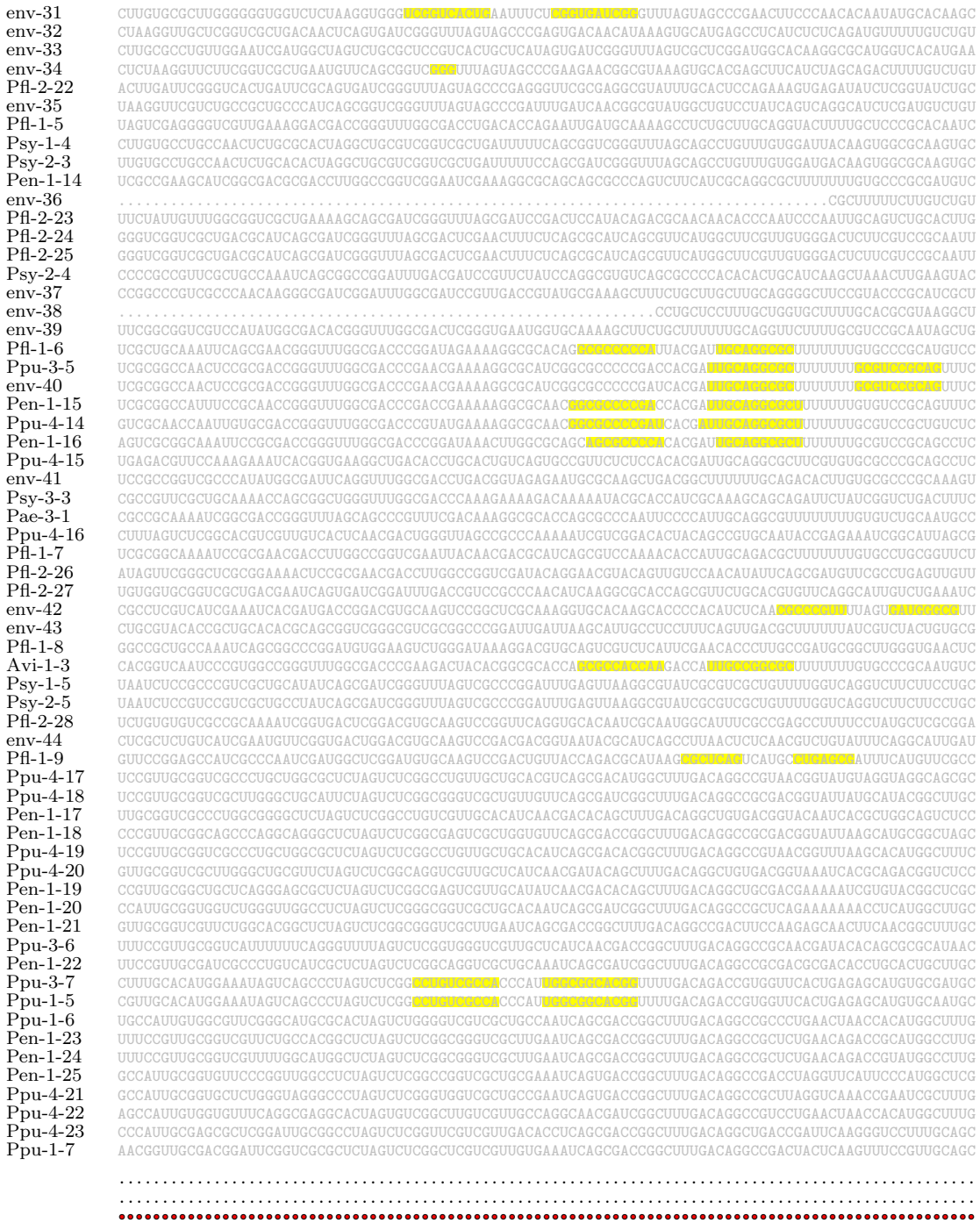

Aba-1-1
Aba-1-2
Hin-3-1
Hin-5-1
Hin-7-1
Hso-1-1
Hso-2-1
Pmu-1-1
Hin-3-2
Hin-2-1
Mha-1-1
Mha-1-2
Pfl-2-1
env-1
Dps-1-1
Dac-1-1
Sgl-2-1
Pfl-2-2
Sbo-1-1
Pna-1-1
Pna-1-2
Xfa-1-1
Sma-1-1
Dno-1-1
Dno-1-2
Dno-1-3
Dno-1-4
Esp-1-1
Plu-1-1
Xax-1-1
Aba-1-3
Xfa-3-1
Xfa-1-2
Sma-1-2
env-2
env-3
Sma-1-3
Sma-1-4
Plu-1-2
env-4
Aba-1-4
Bph-1-1
Bxe-1-1
env-5
env-6
Spr-1-1
Pfl-2-3
env-7
Pfl-2-4
Msp-1-1
Msp-1-2
env-8
Aba-1-5
Pfl-1-1
Psy-1-1
Pae-5-1
Pae-2-1
Pae-6-1
Ppu-4-1
Dps-1-2
Asa-1-1
env-9

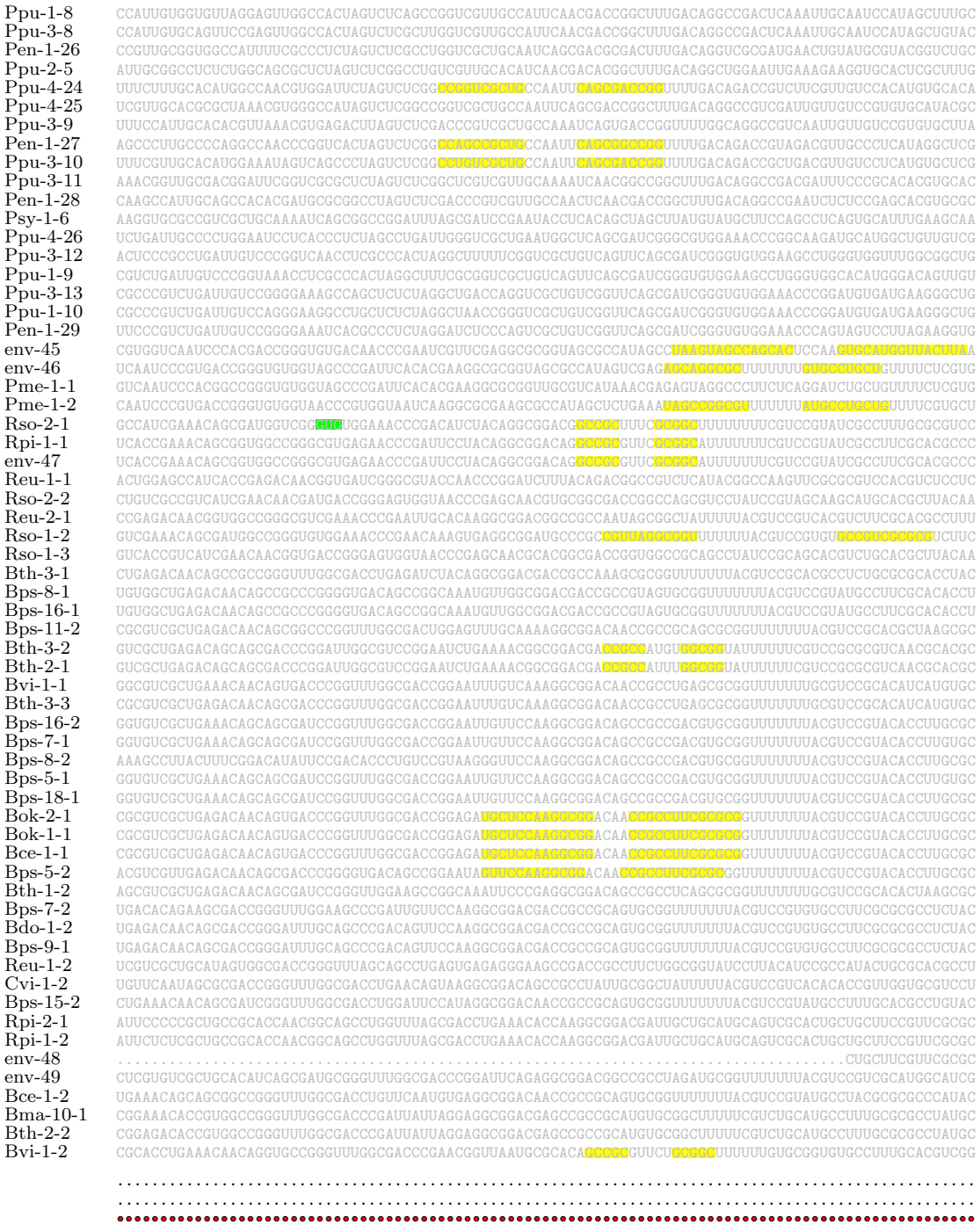

Bsp-1-1
Pen-1-1
Cvi-1-1
Sdy-2-1
Sbo-1-2
env-10
Cte-1-1
Cte-1-2
Sgl-2-2
Sen-8-1
Vfi-1-1
Vfi-1-2
Sde-1-1
Sba-1-1
Sba-2-1
Eco-3-1
Vch-2-1
Vch-3-1
Vsp-1-1
Vch-1-1
Vba-1-1
Sfr-1-1
Vsh-1-1
Sgl-1-1
Sgl-2-3
Sgl-1-2
env-11
Eco-5-1
Sen-11-1
Sen-12-1
Esp-1-2
Sgl-2-4
Sgl-2-5
Sgl-2-6
Sgl-2-7
Sgl-2-8
Sdy-1-1
Sbo-1-3
Eco-10-1
Sfl-1-1
Sfl-2-1
Eco-3-2
Eco-3-3
Eco-3-4
Sfl-3-1
Sfl-2-2
Sfl-1-2
Eco-6-1
Pcd-1-1
Sso-1-1
Sbo-1-4
Sfl-3-2
Eco-3-5
Eco-13-1
Eco-3-6
Sen-13-1
Sen-16-1
Eco-2-1
Eco-15-1
Eco-6-2
Eco-6-3
Eco-2-2

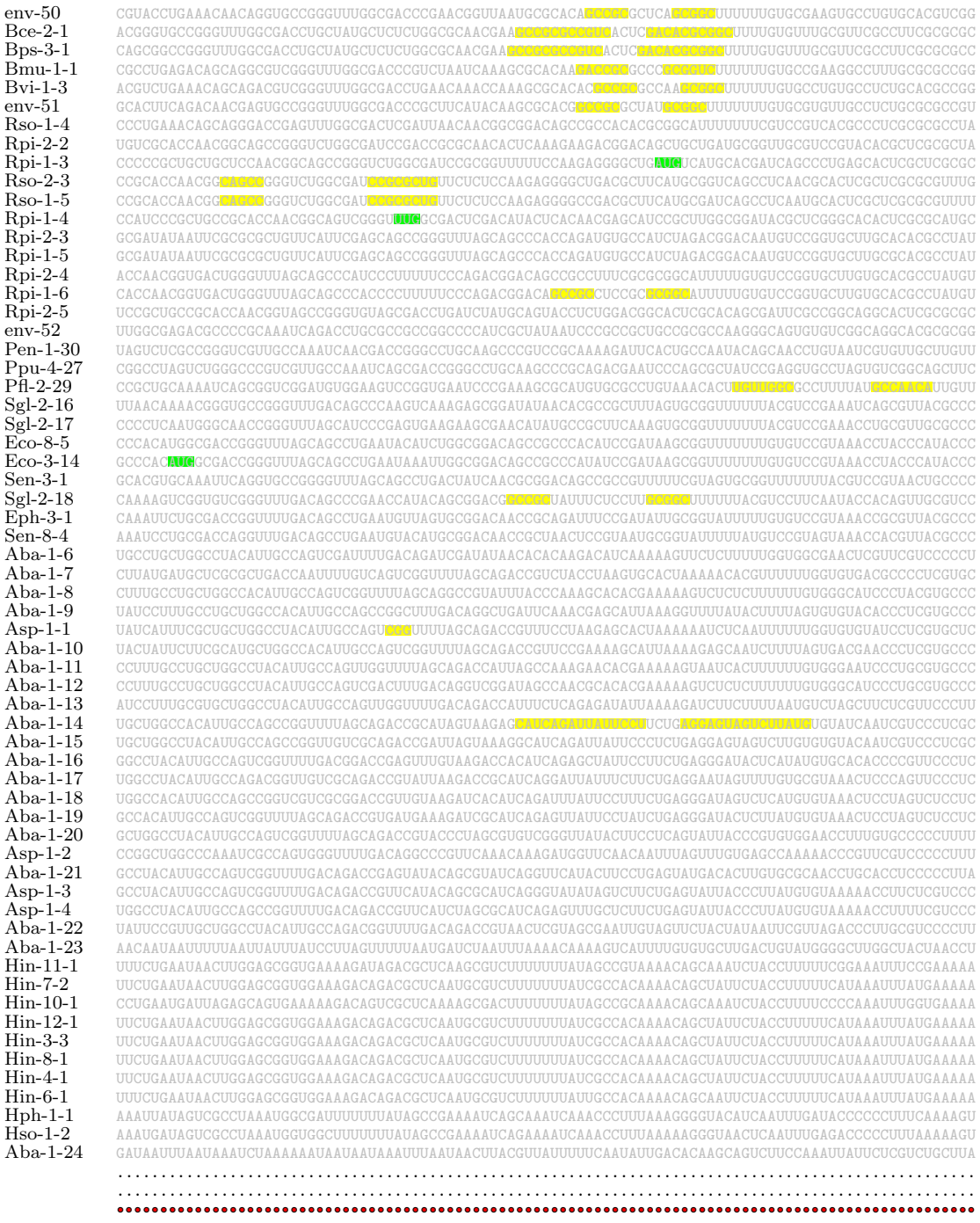

Eco-8-1
Sdy-1-2
Eco-7-1
Eco-9-1
Eco-3-7
Eco-15-2
Eco-2-3
Eco-10-2
Spr-1-2
Sty-2-1
Sen-9-1
Sen-13-2
Eco-10-3
Sso-1-2
Ype-7-1
Ype-1-1
Eco-3-8
Eco-3-9
Yen-1-1
Yen-1-2
Yps-1-1
Ype-13-2
Yps-2-1
Yen-1-3
Ybe-1-1
Yin-1-1
Yen-1-4
Yin-1-2
Sbo-2-1
env-13
Sen-16-2
Api-1-1
Sgl-2-9
Sdy-1-3
Eco-3-10
Eph-2-1
Eco-3-11
Eco-3-12
Sdy-1-4
Eco-8-2
Sdy-2-2
Eco-10-4
Sfl-2-3
Sen-10-2
Sdy-2-3
Eco-9-2
Eph-1-1
Eco-13-2
Sen-15-2
env-14
Eco-9-3
Sen-2-1
Sen-15-3
Eco-11-1
env-15
Sen-9-3
Eco-4-2
Eph-6-1
env-16
Sgl-1-3
Sgl-2-10
Pfl-2-5

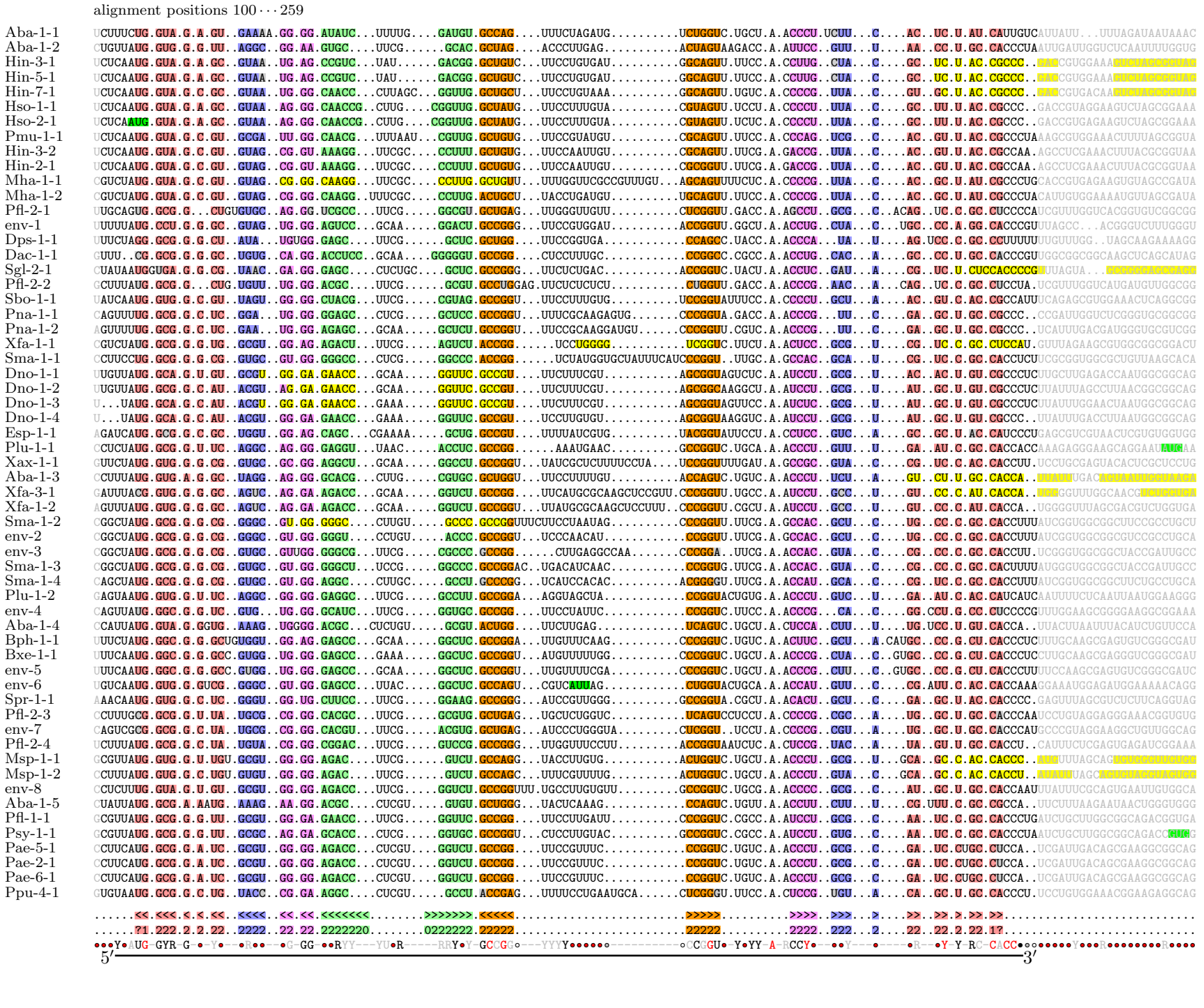

Ppu-1-1
Pfl-2-6
env-17
env-18
env-19
env-20
Bth-1-1
Bps-11-1
Rso-1-1
Sgl-2-11
Bxe-1-2
Bph-1-2
Ppu-4-2
Pen-1-2
Pst-1-1
Sgl-2-12
Eph-7-1
env-21
Esa-1-1
Sen-11-2
Yph-1-1
Sgl-2-13
Sgl-2-14
Eco-3-13
Sen-7-2
Esp-1-3
Ybe-1-2
Ymo-1-1
Esa-1-2
Yfr-1-1
Sen-9-4
Sen-5-1
Sen-7-3
Sen-8-3
Eco-13-3
env-22
env-23
Sen-11-3
Eco-5-2
Eco-14-1
Eco-11-2
Eph-4-1
Sbo-2-2
Eco-8-4
Sgl-2-15
Sty-1-1
Eco-1-1
Sen-4-4
Kph-1-1
Ype-2-2
env-24
Sen-1-1
Sen-14-2
env-25
Ppu-4-3
Ppu-2-1
Pen-1-3
Nmu-1-1
env-26
Avi-1-1
Pae-4-2
Ppu-2-2

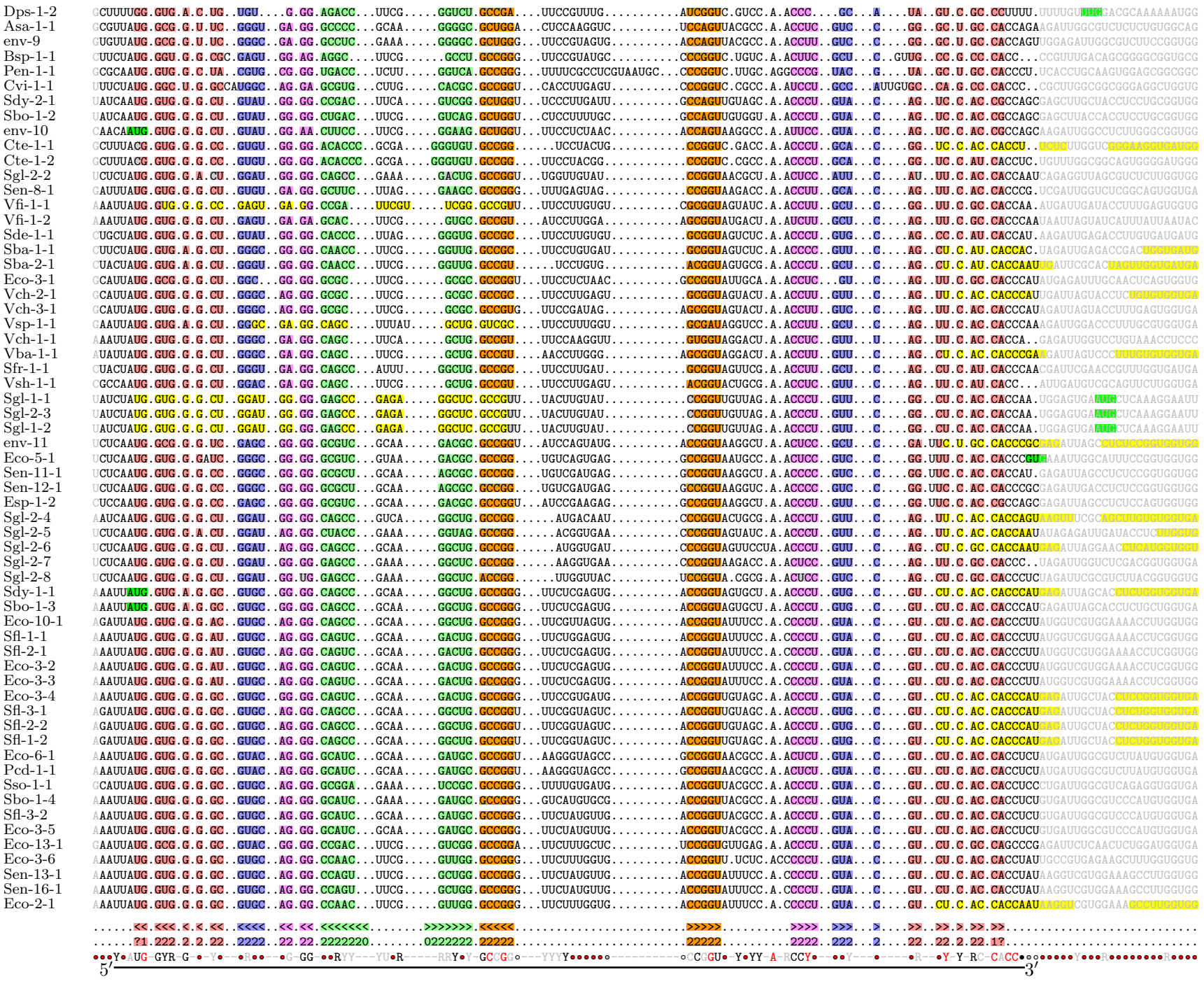

Pen-1-4
Pen-1-5
Ppu-4-4
Pen-1-6
Pen-1-7
Ppu-4-5
Ppu-3-1
Ppu-3-2
Ppu-4-6
Ppu-1-2
Ppu-4-7
Ppu-3-3
Ppu-4-8
Ppu-4-9
env-27
Pen-1-8
Ppu-1-3
Ppu-4-10
Ppu-4-11
Pen-1-9
Pen-1-10
Pfl-2-7
Xor-1-1
env-28
Xca-1-1
Sma-1-5
Bdo-1-1
Pfl-2-8
Pfl-1-2
Ppu-4-12
Pen-1-11
Ppu-3-4
Ppu-2-3
Ppu-4-13
Ppu-1-4
env-29
Psy-3-1
Psy-1-2
Psy-1-3
Pfl-2-9
Pfl-2-10
Pfl-2-11
Pfl-2-12
env-30
Psy-2-1
Pfl-2-13
Psy-2-2
Pfl-2-14
Psy-3-2
Pfl-2-15
Avi-1-2
Ppu-2-4
Pen-1-12
Pfl-2-16
Pfl-2-17
Pfl-2-18
Pfl-1-3
Pen-1-13
Pfl-2-19
Pfl-2-20
Pfl-2-21
Pfl-1-4

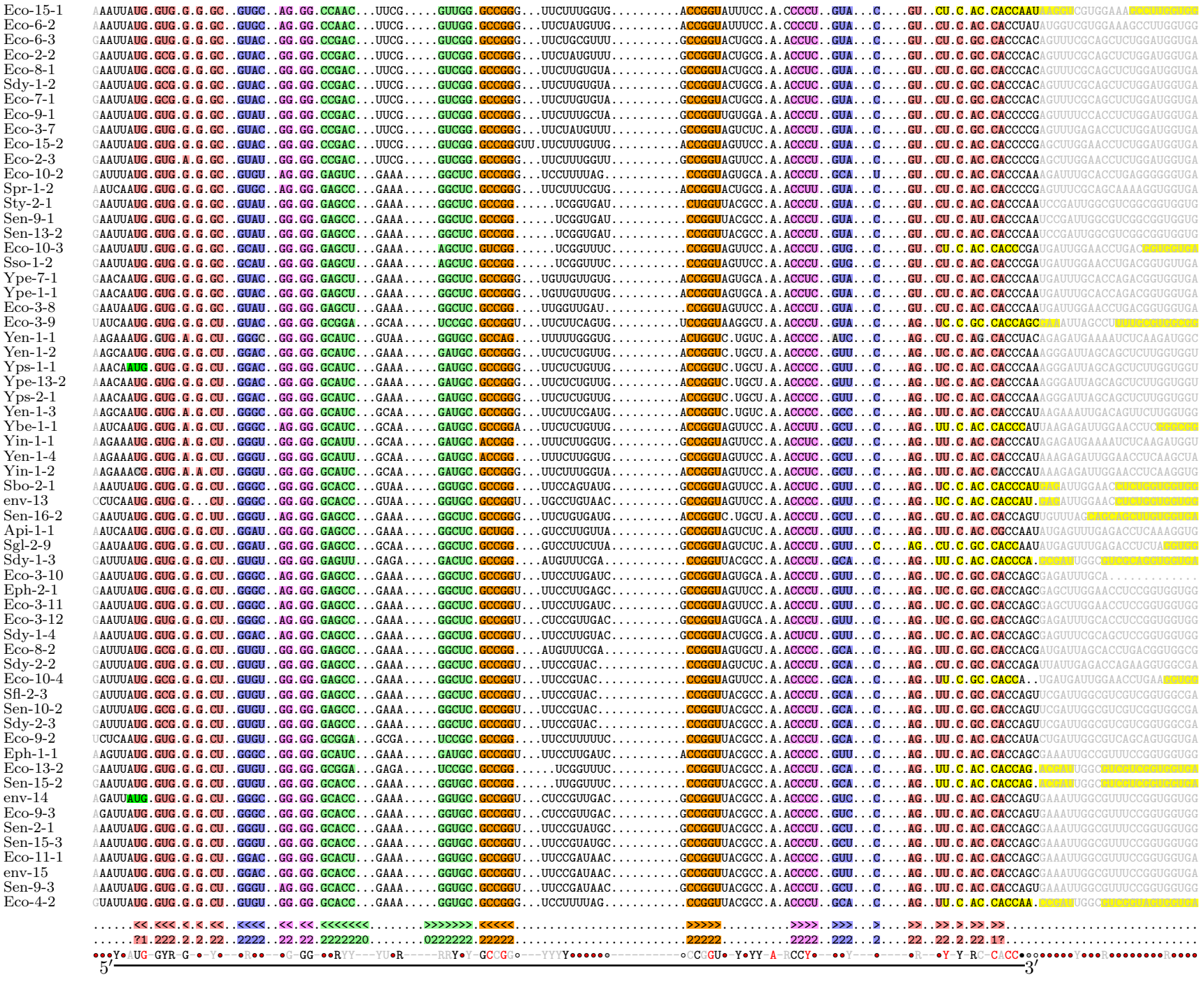

env-31
env-32
env-33
env-34
Pfl-2-22
env-35
Pfl-1-5
Psy-1-4
Psy-2-3
Pen-1-14
env-36
Pfl-2-23
Pfl-2-24
Pfl-2-25
Psy-2-4
env-37
env-38
env-39
Pfl-1-6
Ppu-3-5
env-40
Pen-1-15
Ppu-4-14
Pen-1-16
Ppu-4-15
env-41
Psy-3-3
Pae-3-1
Ppu-4-16
Pfl-1-7
Pfl-2-26
Pfl-2-27
env-42
env-43
Pfl-1-8
Avi-1-3
Psy-1-5
Psy-2-5
Pfl-2-28
env-44
Pfl-1-9
Ppu-4-17
Ppu-4-18
Pen-1-17
Pen-1-18
Ppu-4-19
Ppu-4-20
Pen-1-19
Pen-1-20
Pen-1-21
Ppu-3-6
Pen-1-22
Ppu-3-7
Ppu-1-5
Ppu-1-6
Pen-1-23
Pen-1-24
Pen-1-25
Ppu-4-21
Ppu-4-22
Ppu-4-23
Ppu-1-7

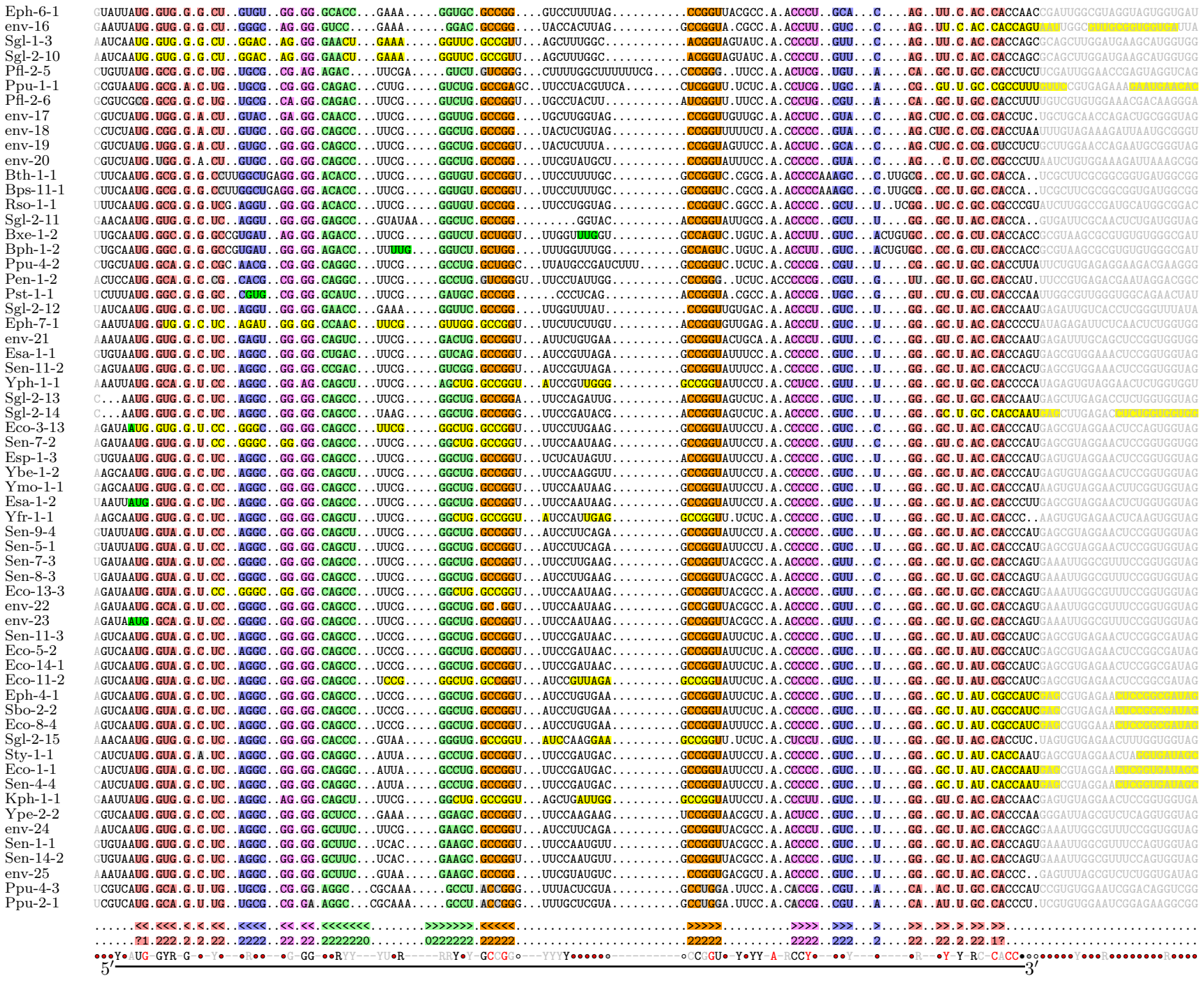

Ppu-1-8
Ppu-3-8
Pen-1-26
Ppu-2-5
Ppu-4-24
Ppu-4-25
Ppu-3-9
Pen-1-27
Ppu-3-10
Ppu-3-11
Pen-1-28
Psy-1-6
Ppu-4-26
Ppu-3-12
Ppu-1-9
Ppu-3-13
Ppu-1-10
Pen-1-29
env-45
env-46
Pme-1-1
Pme-1-2
Rso-2-1
Rpi-1-1
env-47
Reu-1-1
Rso-2-2
Reu-2-1
Rso-1-2
Rso-1-3
Bth-3-1
Bps-8-1
Bps-16-1
Bps-11-2
Bth-3-2
Bth-2-1
Bvi-1-1
Bth-3-3
Bps-16-2
Bps-7-1
Bps-8-2
Bps-5-1
Bps-18-1
Bok-2-1
Bok-1-1
Bce-1-1
Bps-5-2
Bth-1-2
Bps-7-2
Bdo-1-2
Bps-9-1
Reu-1-2
Cvi-1-2
Bps-15-2
Rpi-2-1
Rpi-1-2
env-48
env-49
Bce-1-2
Bma-10-1
Bth-2-2
Bvi-1-2

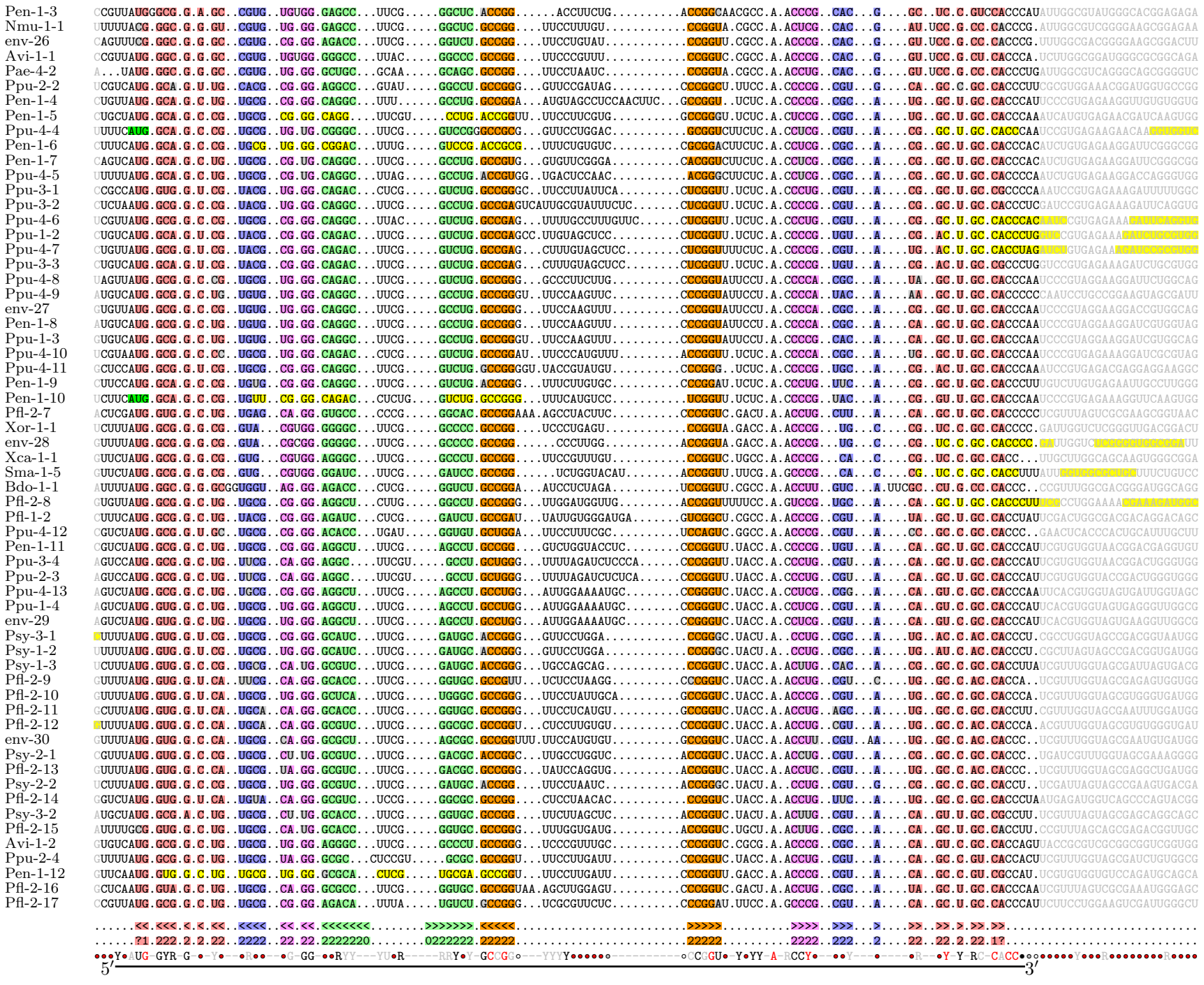

env-50
Bce-2-1
Bps-3-1
Bmu-1-1
Bvi-1-3
env-51
Rso-1-4
Rpi-2-2
Rpi-1-3
Rso-2-3
Rso-1-5
Rpi-1-4
Rpi-2-3
Rpi-1-5
Rpi-2-4
Rpi-1-6
Rpi-2-5
env-52
Pen-1-30
Ppu-4-27
Pfl-2-29
Sgl-2-16
Sgl-2-17
Eco-8-5
Eco-3-14
Sen-3-1
Sgl-2-18
Eph-3-1
Sen-8-4
Aba-1-6
Aba-1-7
Aba-1-8
Aba-1-9
Asp-1-1
Aba-1-10
Aba-1-11
Aba-1-12
Aba-1-13
Aba-1-14
Aba-1-15
Aba-1-16
Aba-1-17
Aba-1-18
Aba-1-19
Aba-1-20
Asp-1-2
Aba-1-21
Asp-1-3
Asp-1-4
Aba-1-22
Aba-1-23
Hin-11-1
Hin-7-2
Hin-10-1
Hin-12-1
Hin-3-3
Hin-8-1
Hin-4-1
Hin-6-1
Hph-1-1
Hso-1-2
Aba-1-24

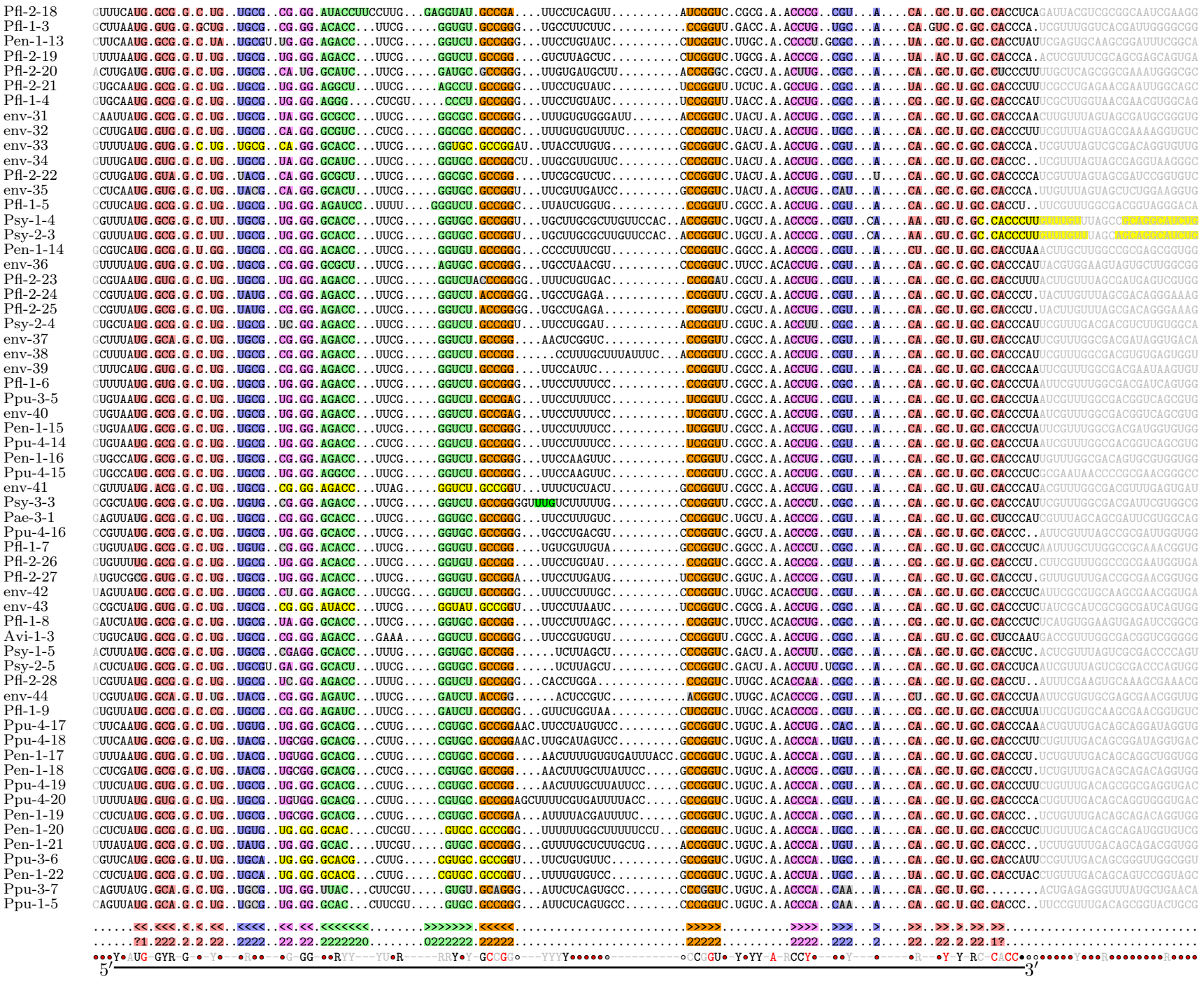

alignment positions 100 · · · 259
Aba-1-1
Aba-1-2
Hin-3-1
Hin-5-1
Hin-7-1
Hso-1-1
Hso-2-1
Pmu-1-1
Hin-3-2
Hin-2-1
Mha-1-1
Mha-1-2
Pfl-2-1
env-1
Dps-1-1
Dac-1-1
Sgl-2-1
Pfl-2-2
Sbo-1-1
Pna-1-1
Pna-1-2
Xfa-1-1
Sma-1-1
Dno-1-1
Dno-1-2
Dno-1-3
Dno-1-4
Esp-1-1
Plu-1-1
Xax-1-1
Aba-1-3
Xfa-3-1
Xfa-1-2
Sma-1-2
env-2
env-3
Sma-1-3
Sma-1-4
Plu-1-2
env-4
Aba-1-4
Bph-1-1
Bxe-1-1
env-5
env-6
Spr-1-1
Pfl-2-3
env-7
Pfl-2-4
Msp-1-1
Msp-1-2
env-8
Aba-1-5
Pfl-1-1
Psy-1-1
Pae-5-1
Pae-2-1
Pae-6-1
Ppu-4-1
5′
3′

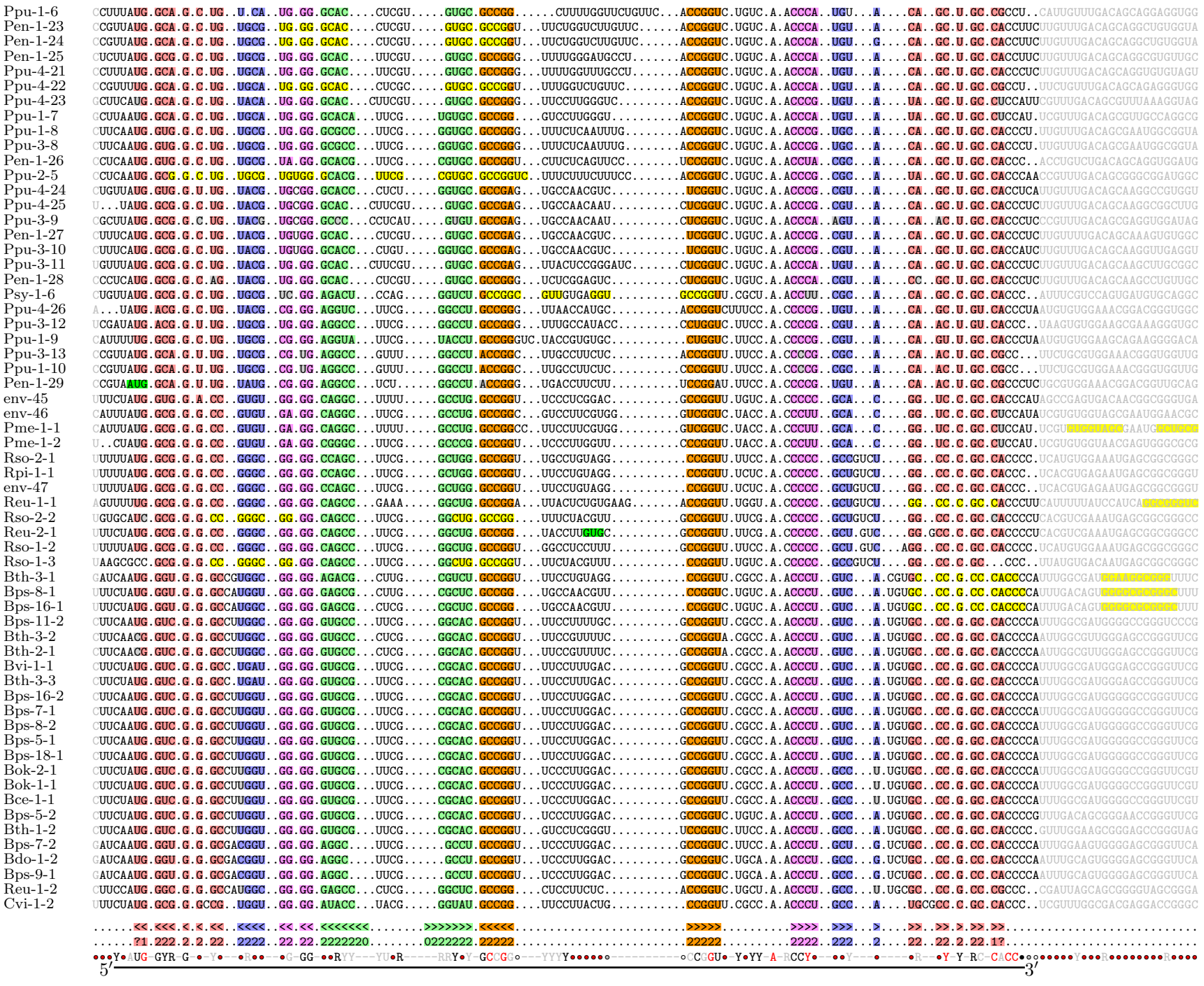

Dps-1-2
Asa-1-1
env-9
Bsp-1-1
Pen-1-1
Cvi-1-1
Sdy-2-1
Sbo-1-2
env-10
Cte-1-1
Cte-1-2
Sgl-2-2
Sen-8-1
Vfi-1-1
Vfi-1-2
Sde-1-1
Sba-1-1
Sba-2-1
Eco-3-1
Vch-2-1
Vch-3-1
Vsp-1-1
Vch-1-1
Vba-1-1
Sfr-1-1
Vsh-1-1
Sgl-1-1
Sgl-2-3
Sgl-1-2
env-11
Eco-5-1
Sen-11-1
Sen-12-1
Esp-1-2
Sgl-2-4
Sgl-2-5
Sgl-2-6
Sgl-2-7
Sgl-2-8
Sdy-1-1
Sbo-1-3
Eco-10-1
Sfl-1-1
Sfl-2-1
Eco-3-2
Eco-3-3
Eco-3-4
Sfl-3-1
Sfl-2-2
Sfl-1-2
Eco-6-1
Pcd-1-1
Sso-1-1
Sbo-1-4
Sfl-3-2
Eco-3-5
Eco-13-1
Eco-3-6
Sen-13-1
Sen-16-1
Eco-2-1
5′
3′

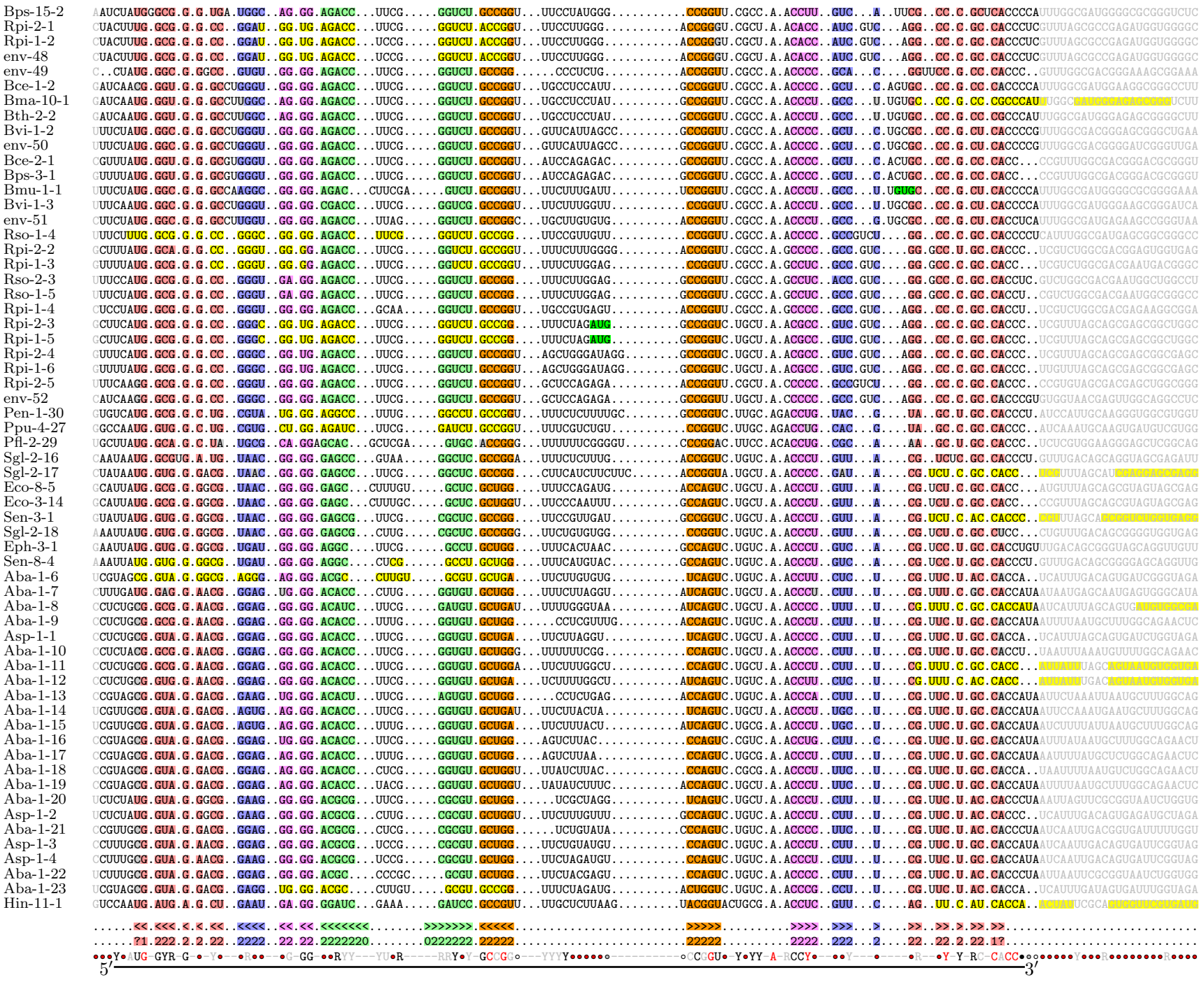

Eco-15-1
Eco-6-2
Eco-6-3
Eco-2-2
Eco-8-1
Sdy-1-2
Eco-7-1
Eco-9-1
Eco-3-7
Eco-15-2
Eco-2-3
Eco-10-2
Spr-1-2
Sty-2-1
Sen-9-1
Sen-13-2
Eco-10-3
Sso-1-2
Ype-7-1
Ype-1-1
Eco-3-8
Eco-3-9
Yen-1-1
Yen-1-2
Yps-1-1
Ype-13-2
Yps-2-1
Yen-1-3
Ybe-1-1
Yin-1-1
Yen-1-4
Yin-1-2
Sbo-2-1
env-13
Sen-16-2
Api-1-1
Sgl-2-9
Sdy-1-3
Eco-3-10
Eph-2-1
Eco-3-11
Eco-3-12
Sdy-1-4
Eco-8-2
Sdy-2-2
Eco-10-4
Sfl-2-3
Sen-10-2
Sdy-2-3
Eco-9-2
Eph-1-1
Eco-13-2
Sen-15-2
env-14
Eco-9-3
Sen-2-1
Sen-15-3
Eco-11-1
env-15
Sen-9-3
Eco-4-2
5′
3′

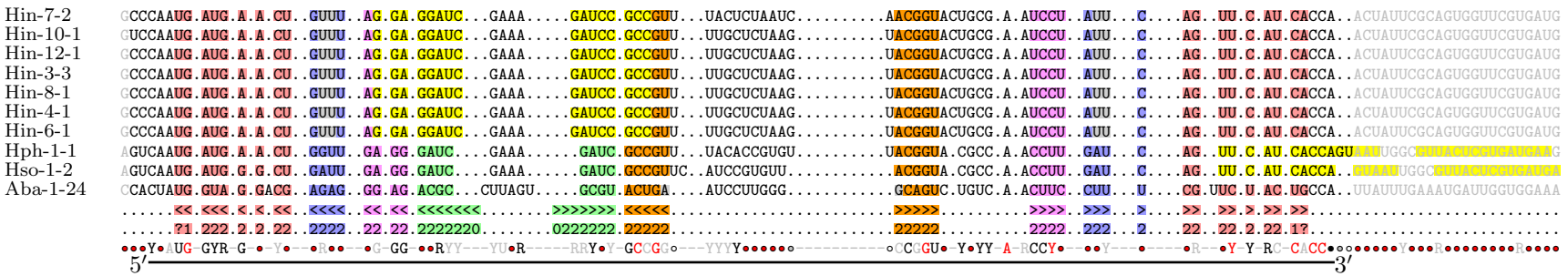

Eph-6-1
env-16
Sgl-1-3
Sgl-2-10
Pfl-2-5
Ppu-1-1
Pfl-2-6
env-17
env-18
env-19
env-20
Bth-1-1
Bps-11-1
Rso-1-1
Sgl-2-11
Bxe-1-2
Bph-1-2
Ppu-4-2
Pen-1-2
Pst-1-1
Sgl-2-12
Eph-7-1
env-21
Esa-1-1
Sen-11-2
Yph-1-1
Sgl-2-13
Sgl-2-14
Eco-3-13
Sen-7-2
Esp-1-3
Ybe-1-2
Ymo-1-1
Esa-1-2
Yfr-1-1
Sen-9-4
Sen-5-1
Sen-7-3
Sen-8-3
Eco-13-3
env-22
env-23
Sen-11-3
Eco-5-2
Eco-14-1
Eco-11-2
Eph-4-1
Sbo-2-2
Eco-8-4
Sgl-2-15
Sty-1-1
Eco-1-1
Sen-4-4
Kph-1-1
Ype-2-2
env-24
Sen-1-1
Sen-14-2
env-25
Ppu-4-3
Ppu-2-1
5′
3′


Pen-1-3
Nmu-1-1
env-26
Avi-1-1
Pae-4-2
Ppu-2-2
Pen-1-4
Pen-1-5
Ppu-4-4
Pen-1-6
Pen-1-7
Ppu-4-5
Ppu-3-1
Ppu-3-2
Ppu-4-6
Ppu-1-2
Ppu-4-7
Ppu-3-3
Ppu-4-8
Ppu-4-9
env-27
Pen-1-8
Ppu-1-3
Ppu-4-10
Ppu-4-11
Pen-1-9
Pen-1-10
Pfl-2-7
Xor-1-1
env-28
Xca-1-1
Sma-1-5
Bdo-1-1
Pfl-2-8
Pfl-1-2
Ppu-4-12
Pen-1-11
Ppu-3-4
Ppu-2-3
Ppu-4-13
Ppu-1-4
env-29
Psy-3-1
Psy-1-2
Psy-1-3
Pfl-2-9
Pfl-2-10
Pfl-2-11
Pfl-2-12
env-30
Psy-2-1
Pfl-2-13
Psy-2-2
Pfl-2-14
Psy-3-2
Pfl-2-15
Avi-1-2
Ppu-2-4
Pen-1-12
Pfl-2-16
Pfl-2-17
5′
3′

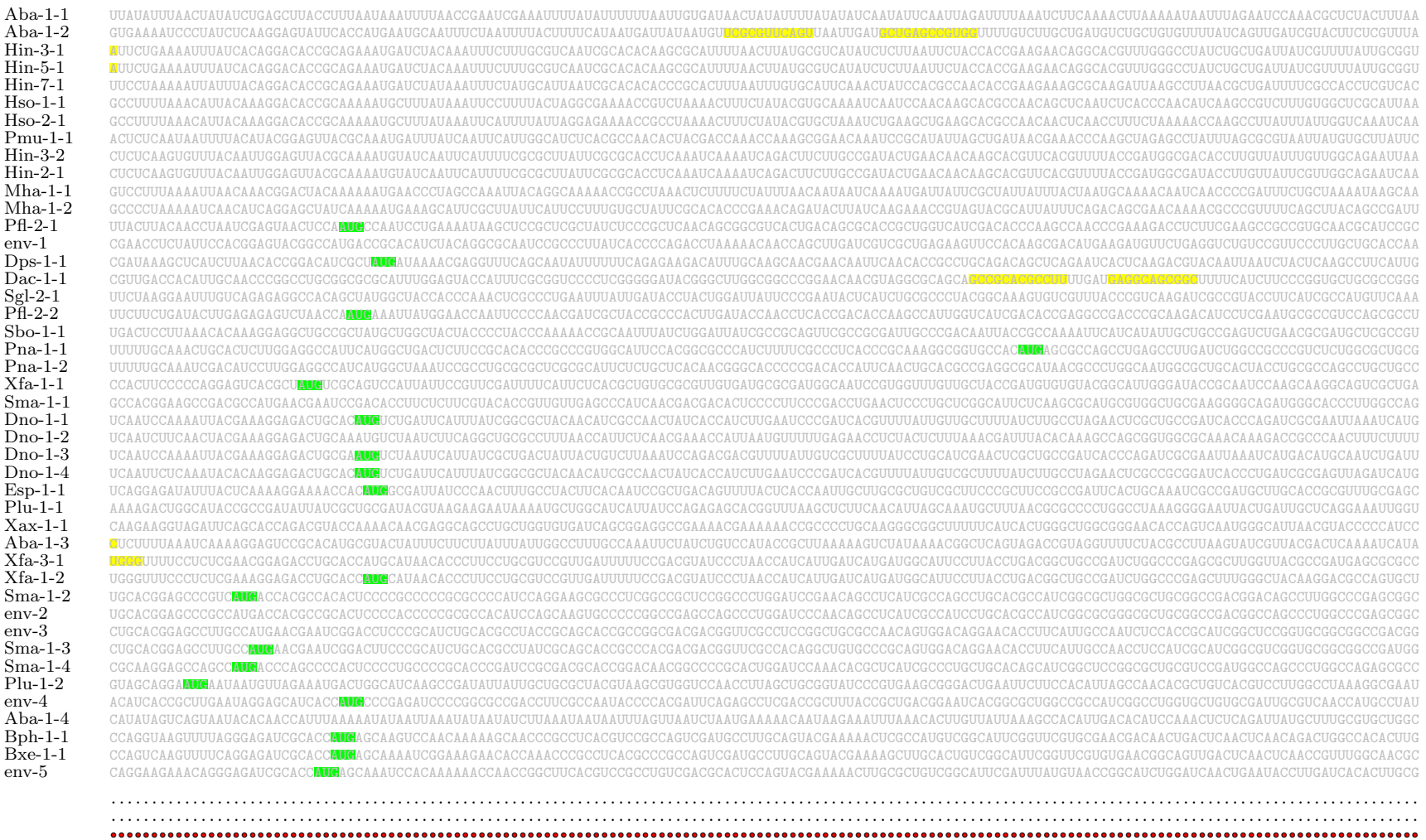

Pfl-2-18
Pfl-1-3
Pen-1-13
Pfl-2-19
Pfl-2-20
Pfl-2-21
Pfl-1-4
env-31
env-32
env-33
env-34
Pfl-2-22
env-35
Pfl-1-5
Psy-1-4
Psy-2-3
Pen-1-14
env-36
Pfl-2-23
Pfl-2-24
Pfl-2-25
Psy-2-4
env-37
env-38
env-39
Pfl-1-6
Ppu-3-5
env-40
Pen-1-15
Ppu-4-14
Pen-1-16
Ppu-4-15
env-41
Psy-3-3
Pae-3-1
Ppu-4-16
Pfl-1-7
Pfl-2-26
Pfl-2-27
env-42
env-43
Pfl-1-8
Avi-1-3
Psy-1-5
Psy-2-5
Pfl-2-28
env-44
Pfl-1-9
Ppu-4-17
Ppu-4-18
Pen-1-17
Pen-1-18
Ppu-4-19
Ppu-4-20
Pen-1-19
Pen-1-20
Pen-1-21
Ppu-3-6
Pen-1-22
Ppu-3-7
Ppu-1-5
5′
3′

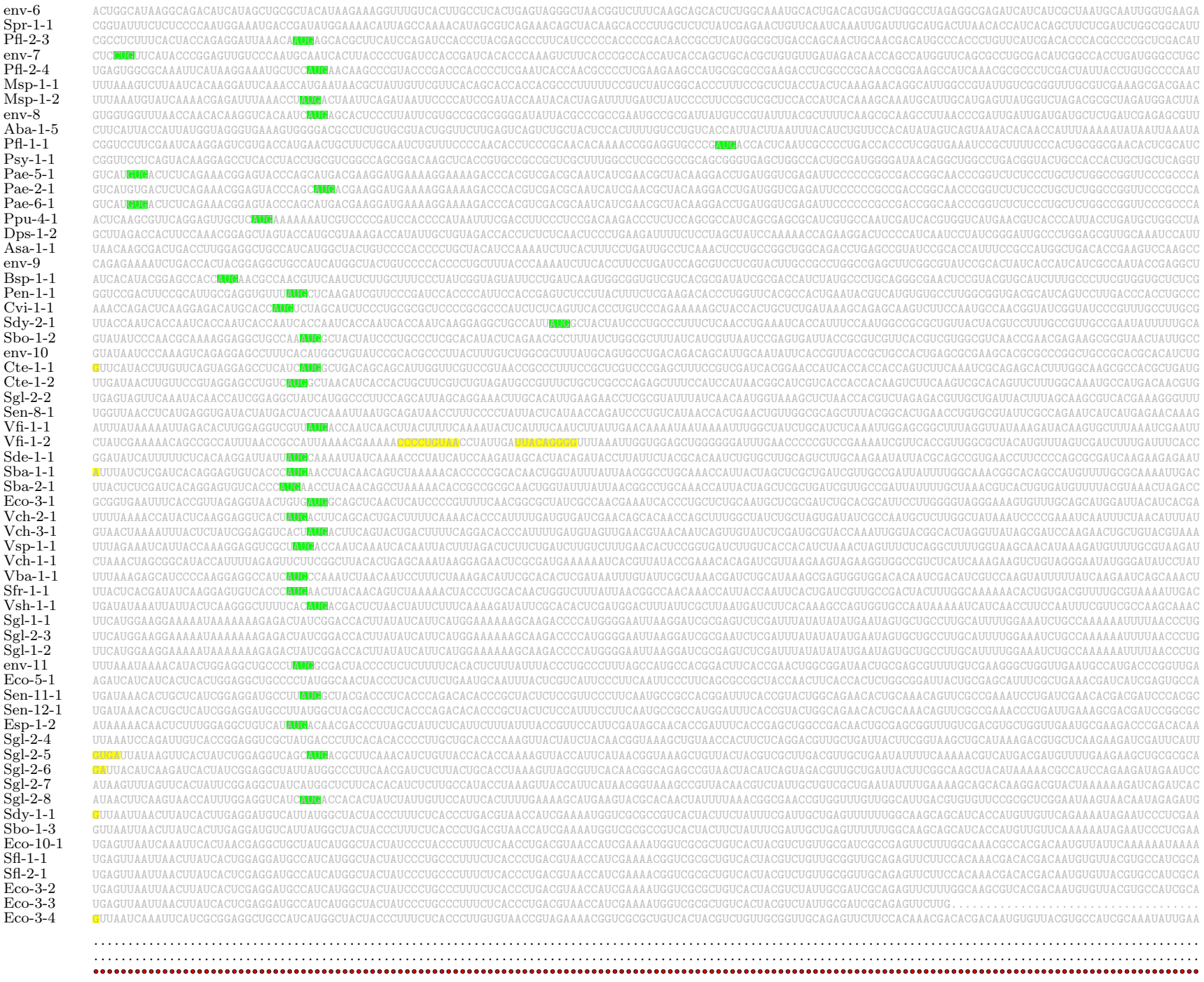

Ppu-1-6
Pen-1-23
Pen-1-24
Pen-1-25
Ppu-4-21
Ppu-4-22
Ppu-4-23
Ppu-1-7
Ppu-1-8
Ppu-3-8
Pen-1-26
Ppu-2-5
Ppu-4-24
Ppu-4-25
Ppu-3-9
Pen-1-27
Ppu-3-10
Ppu-3-11
Pen-1-28
Psy-1-6
Ppu-4-26
Ppu-3-12
Ppu-1-9
Ppu-3-13
Ppu-1-10
Pen-1-29
env-45
env-46
Pme-1-1
Pme-1-2
Rso-2-1
Rpi-1-1
env-47
Reu-1-1
Rso-2-2
Reu-2-1
Rso-1-2
Rso-1-3
Bth-3-1
Bps-8-1
Bps-16-1
Bps-11-2
Bth-3-2
Bth-2-1
Bvi-1-1
Bth-3-3
Bps-16-2
Bps-7-1
Bps-8-2
Bps-5-1
Bps-18-1
Bok-2-1
Bok-1-1
Bce-1-1
Bps-5-2
Bth-1-2
Bps-7-2
Bdo-1-2
Bps-9-1
Reu-1-2
Cvi-1-2
5′
3′

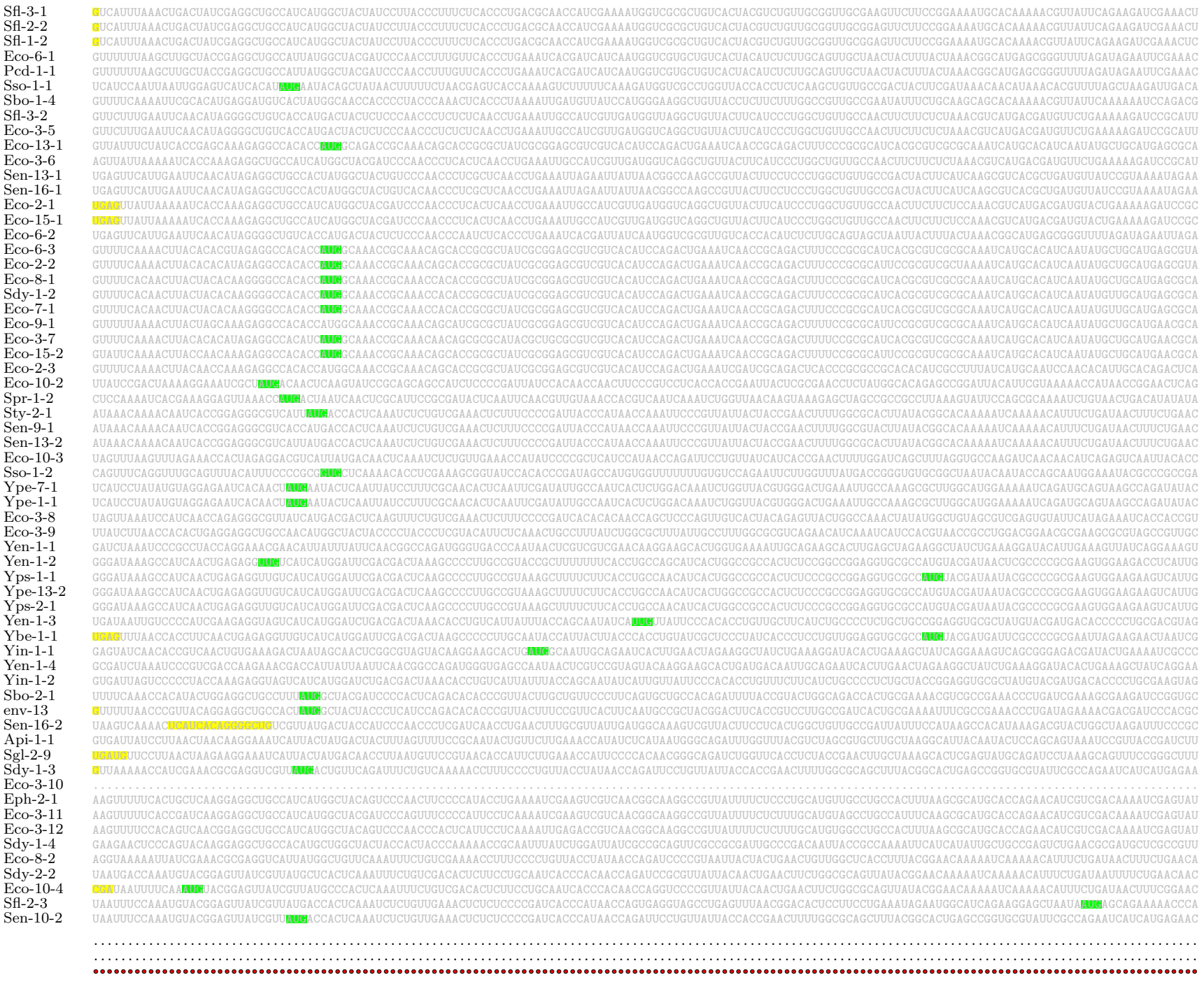

Bps-15-2
Rpi-2-1
Rpi-1-2
env-48
env-49
Bce-1-2
Bma-10-1
Bth-2-2
Bvi-1-2
env-50
Bce-2-1
Bps-3-1
Bmu-1-1
Bvi-1-3
env-51
Rso-1-4
Rpi-2-2
Rpi-1-3
Rso-2-3
Rso-1-5
Rpi-1-4
Rpi-2-3
Rpi-1-5
Rpi-2-4
Rpi-1-6
Rpi-2-5
env-52
Pen-1-30
Ppu-4-27
Pfl-2-29
Sgl-2-16
Sgl-2-17
Eco-8-5
Eco-3-14
Sen-3-1
Sgl-2-18
Eph-3-1
Sen-8-4
Aba-1-6
Aba-1-7
Aba-1-8
Aba-1-9
Asp-1-1
Aba-1-10
Aba-1-11
Aba-1-12
Aba-1-13
Aba-1-14
Aba-1-15
Aba-1-16
Aba-1-17
Aba-1-18
Aba-1-19
Aba-1-20
Asp-1-2
Aba-1-21
Asp-1-3
Asp-1-4
Aba-1-22
Aba-1-23
Hin-11-1
5′
3′

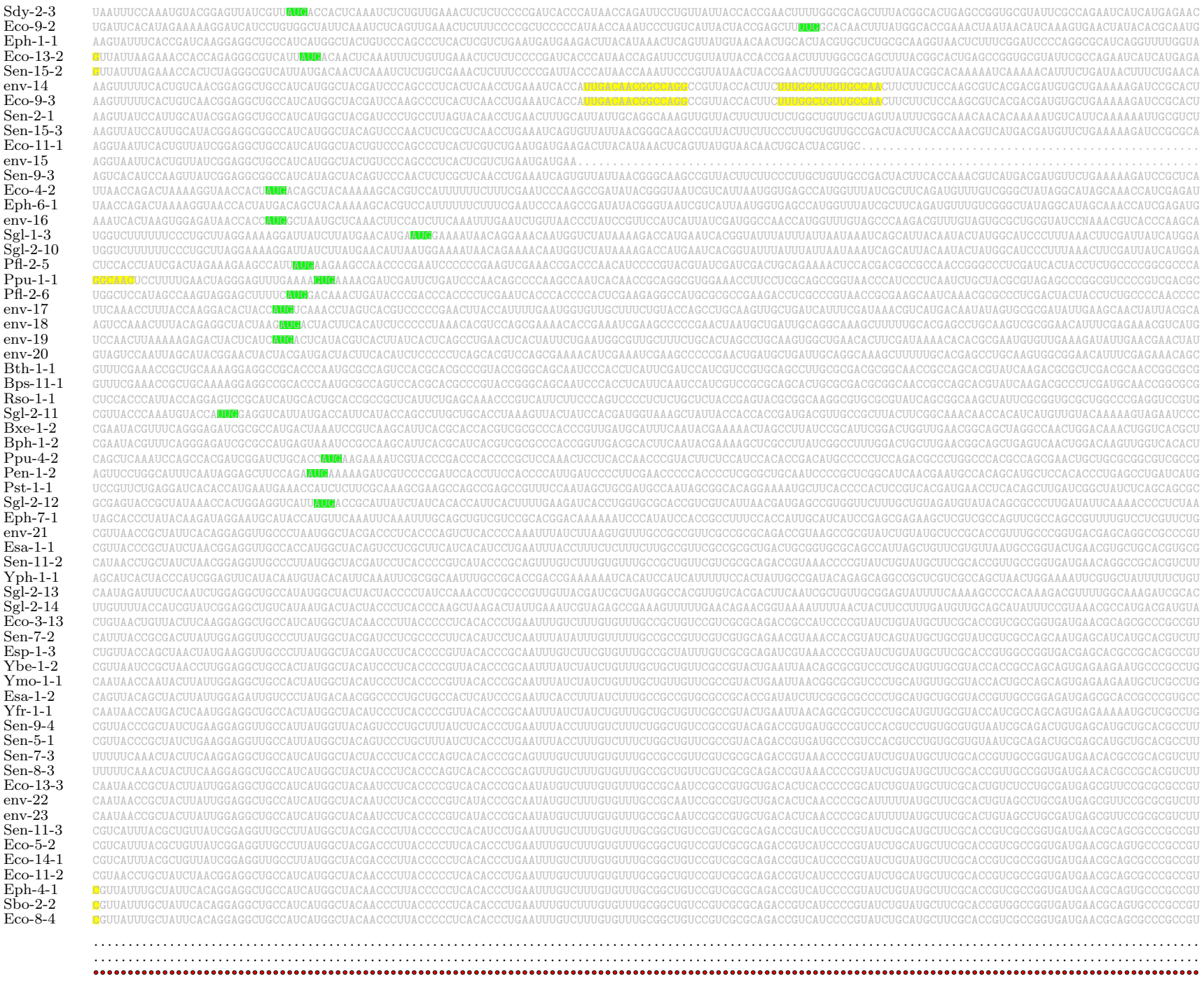


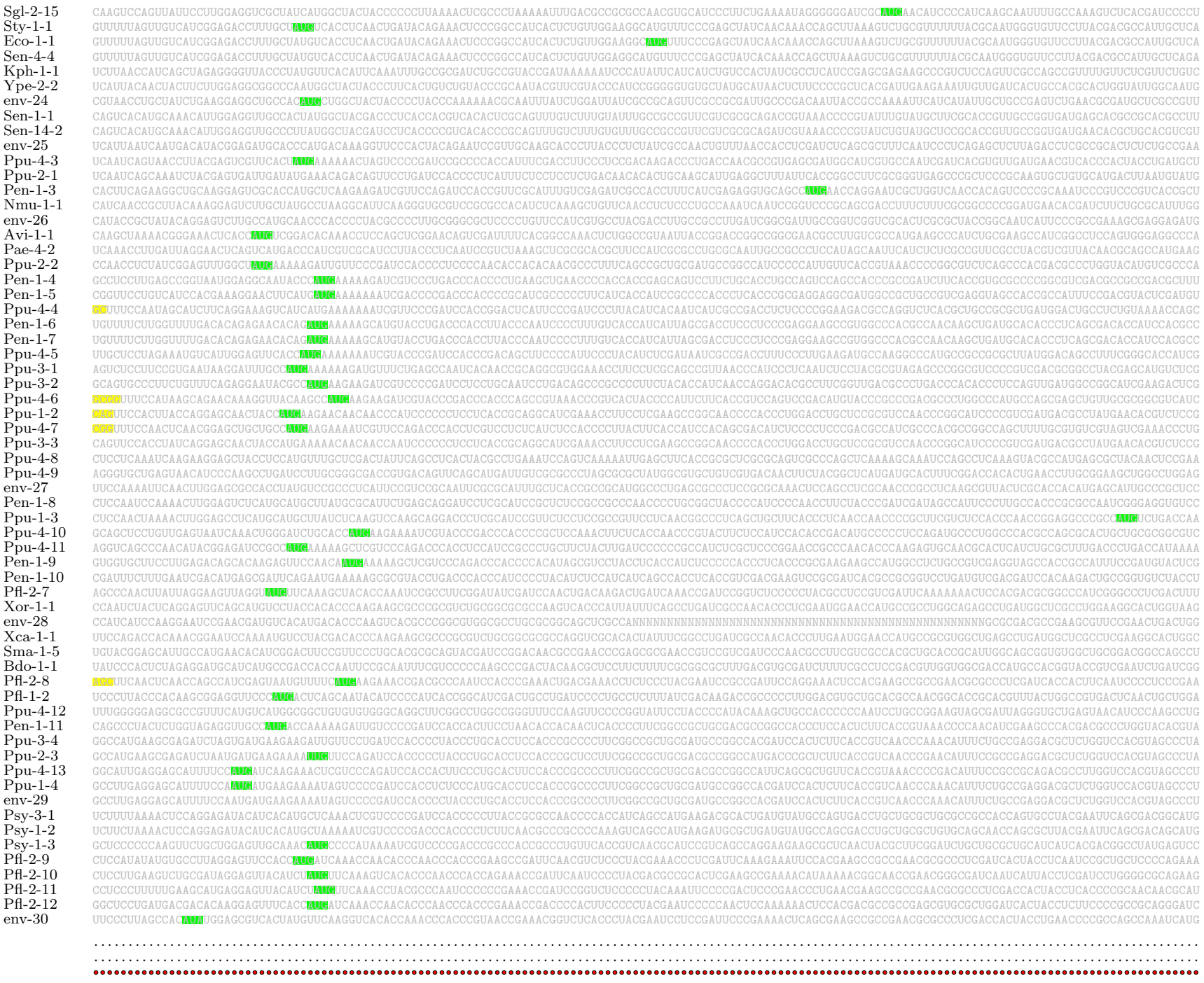
alignment positions 260 · · · 419

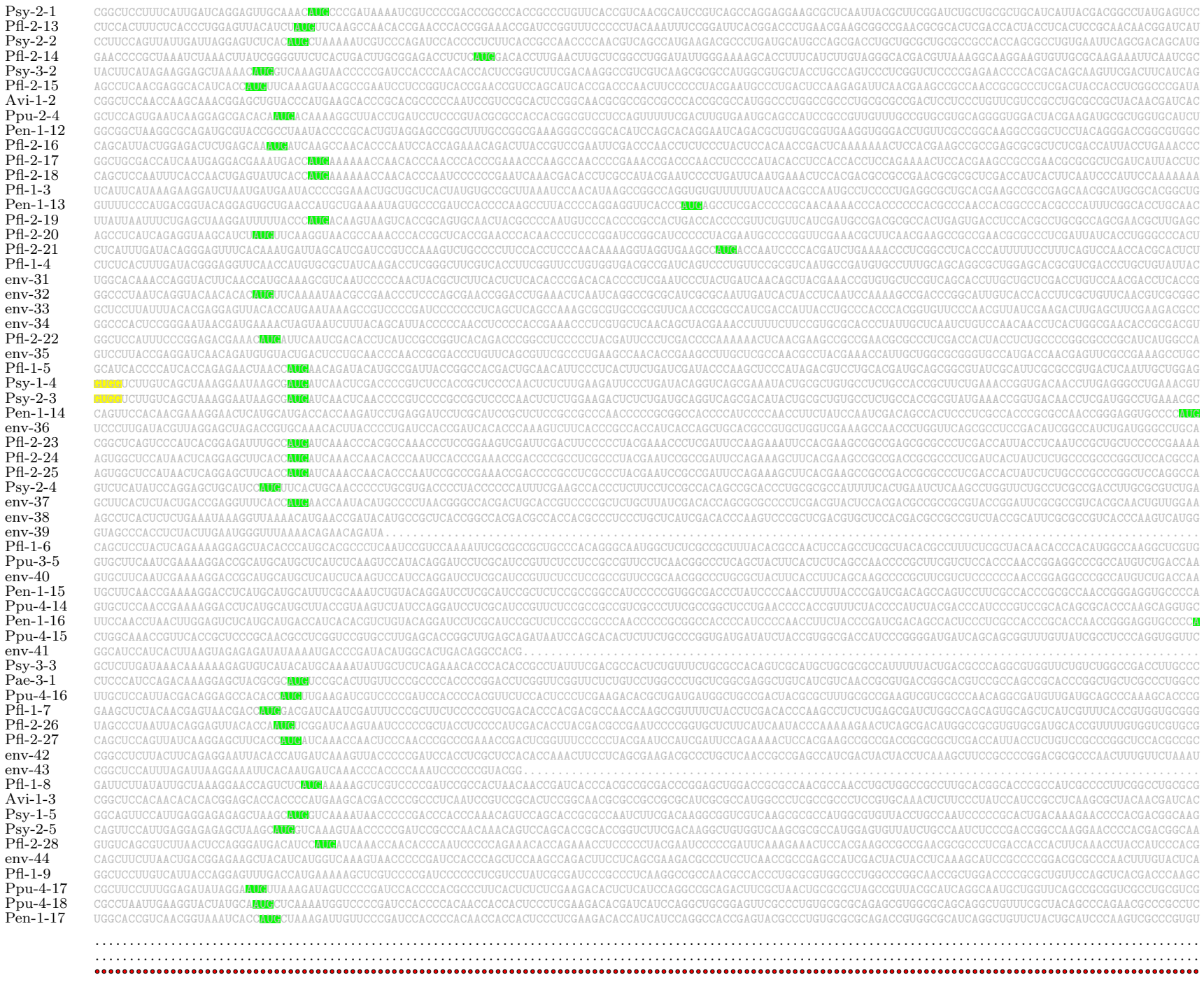


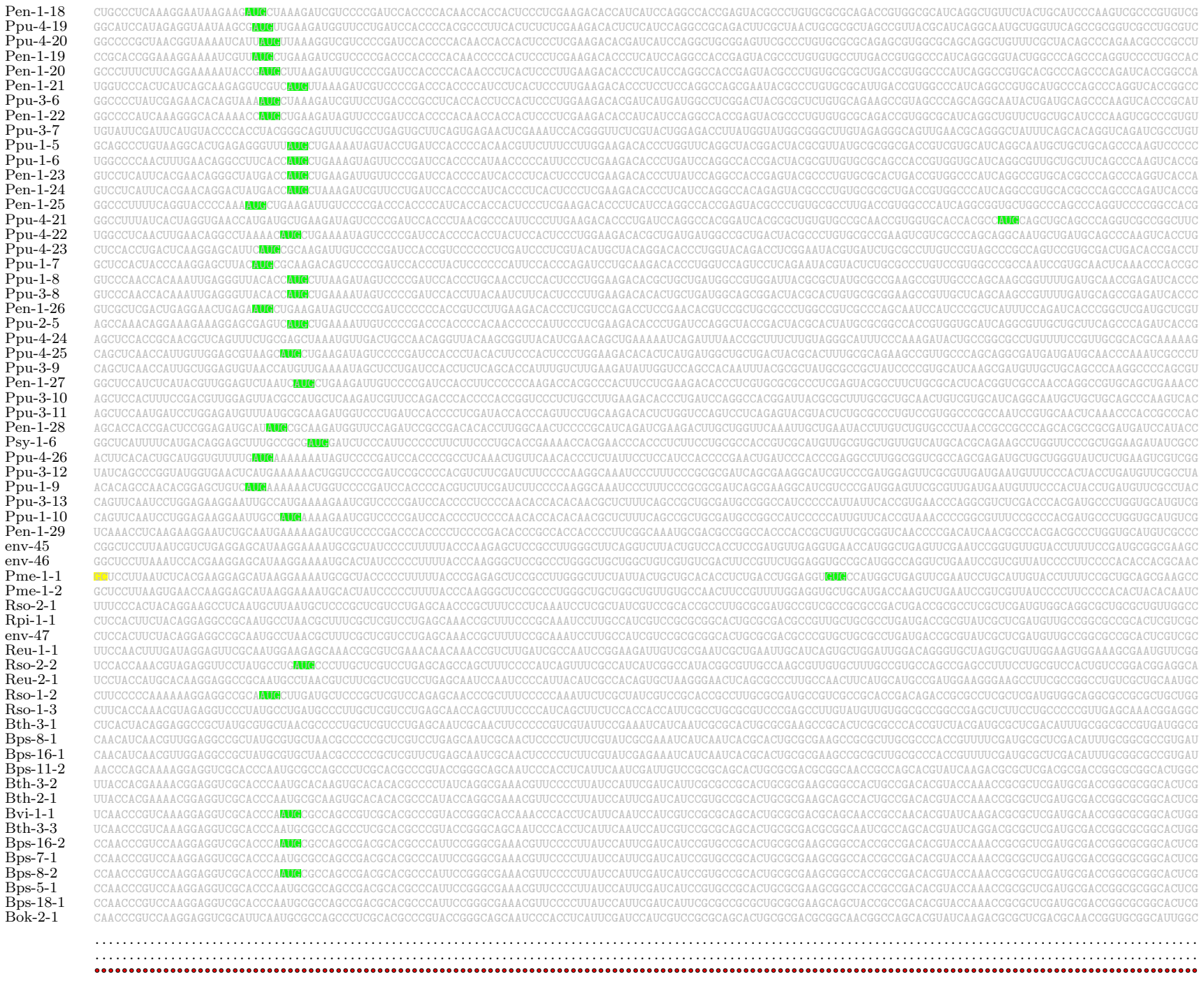

env-6
Spr-1-1
Pfl-2-3
env-7
Pfl-2-4
Msp-1-1
Msp-1-2
env-8
Aba-1-5
Pfl-1-1
Psy-1-1
Pae-5-1
Pae-2-1
Pae-6-1
Ppu-4-1
Dps-1-2
Asa-1-1
env-9
Bsp-1-1
Pen-1-1
Cvi-1-1
Sdy-2-1
Sbo-1-2
env-10
Cte-1-1
Cte-1-2
Sgl-2-2
Sen-8-1
Vfi-1-1
Vfi-1-2
Sde-1-1
Sba-1-1
Sba-2-1
Eco-3-1
Vch-2-1
Vch-3-1
Vsp-1-1
Vch-1-1
Vba-1-1
Sfr-1-1
Vsh-1-1
Sgl-1-1
Sgl-2-3
Sgl-1-2
env-11
Eco-5-1
Sen-11-1
Sen-12-1
Esp-1-2
Sgl-2-4
Sgl-2-5
Sgl-2-6
Sgl-2-7
Sgl-2-8
Sdy-1-1
Sbo-1-3
Eco-10-1
Sfl-1-1
Sfl-2-1
Eco-3-2
Eco-3-3
Eco-3-4

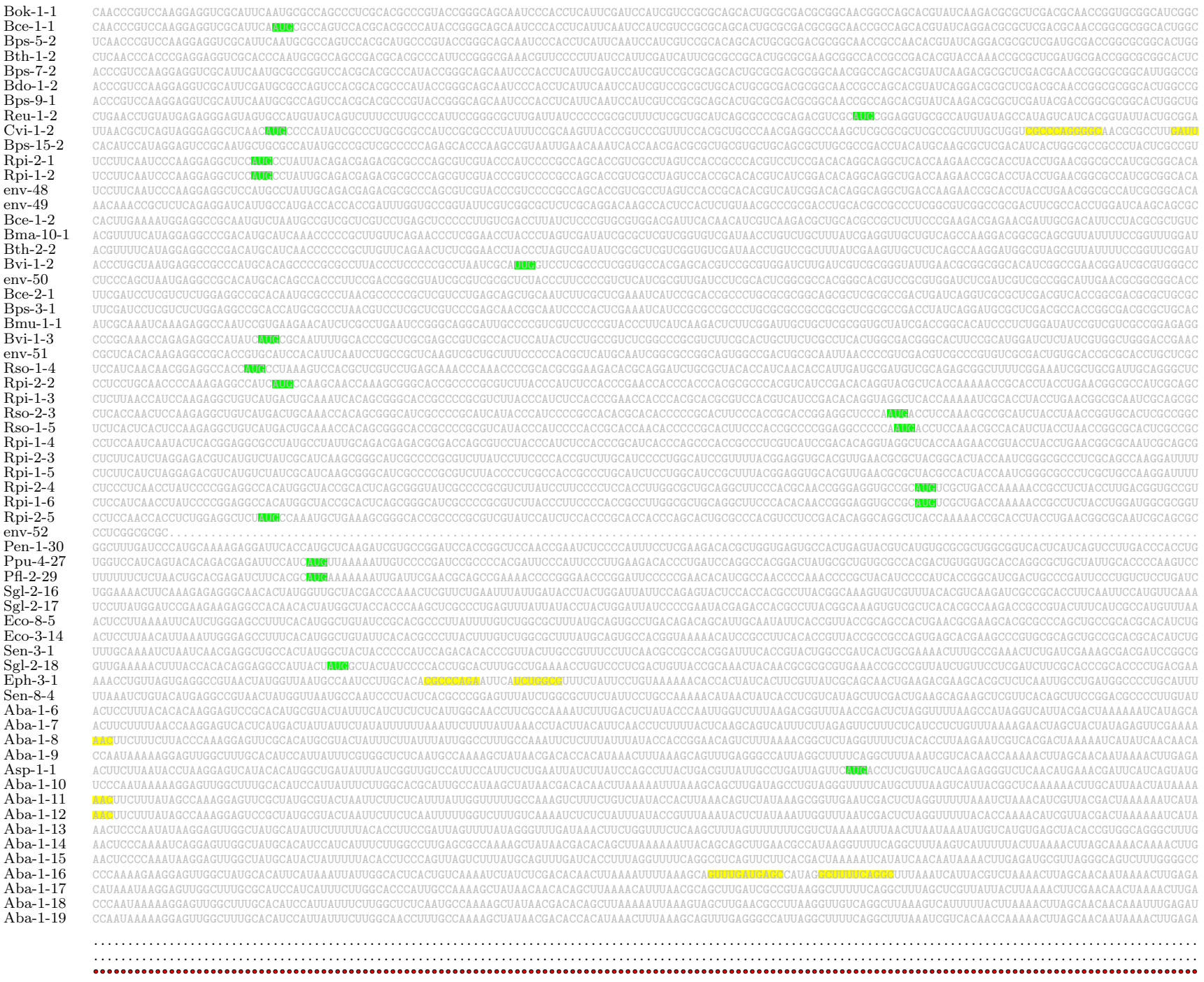

Sfl-3-1
Sfl-2-2
Sfl-1-2
Eco-6-1
Pcd-1-1
Sso-1-1
Sbo-1-4
Sfl-3-2
Eco-3-5
Eco-13-1
Eco-3-6
Sen-13-1
Sen-16-1
Eco-2-1
Eco-15-1
Eco-6-2
Eco-6-3
Eco-2-2
Eco-8-1
Sdy-1-2
Eco-7-1
Eco-9-1
Eco-3-7
Eco-15-2
Eco-2-3
Eco-10-2
Spr-1-2
Sty-2-1
Sen-9-1
Sen-13-2
Eco-10-3
Sso-1-2
Ype-7-1
Ype-1-1
Eco-3-8
Eco-3-9
Yen-1-1
Yen-1-2
Yps-1-1
Ype-13-2
Yps-2-1
Yen-1-3
Ybe-1-1
Yin-1-1
Yen-1-4
Yin-1-2
Sbo-2-1
env-13
Sen-16-2
Api-1-1
Sgl-2-9
Sdy-1-3
Eco-3-10
Eph-2-1
Eco-3-11
Eco-3-12
Sdy-1-4
Eco-8-2
Sdy-2-2
Eco-10-4
Sfl-2-3
Sen-10-2


Sdy-2-3
Eco-9-2
Eph-1-1
Eco-13-2
Sen-15-2
env-14
Eco-9-3
Sen-2-1
Sen-15-3
Eco-11-1
env-15
Sen-9-3
Eco-4-2
Eph-6-1
env-16
Sgl-1-3
Sgl-2-10
Pfl-2-5
Ppu-1-1
Pfl-2-6
env-17
env-18
env-19
env-20
Bth-1-1
Bps-11-1
Rso-1-1
Sgl-2-11
Bxe-1-2
Bph-1-2
Ppu-4-2
Pen-1-2
Pst-1-1
Sgl-2-12
Eph-7-1
env-21
Esa-1-1
Sen-11-2
Yph-1-1
Sgl-2-13
Sgl-2-14
Eco-3-13
Sen-7-2
Esp-1-3
Ybe-1-2
Ymo-1-1
Esa-1-2
Yfr-1-1
Sen-9-4
Sen-5-1
Sen-7-3
Sen-8-3
Eco-13-3
env-22
env-23
Sen-11-3
Eco-5-2
Eco-14-1
Eco-11-2
Eph-4-1
Sbo-2-2
Eco-8-4

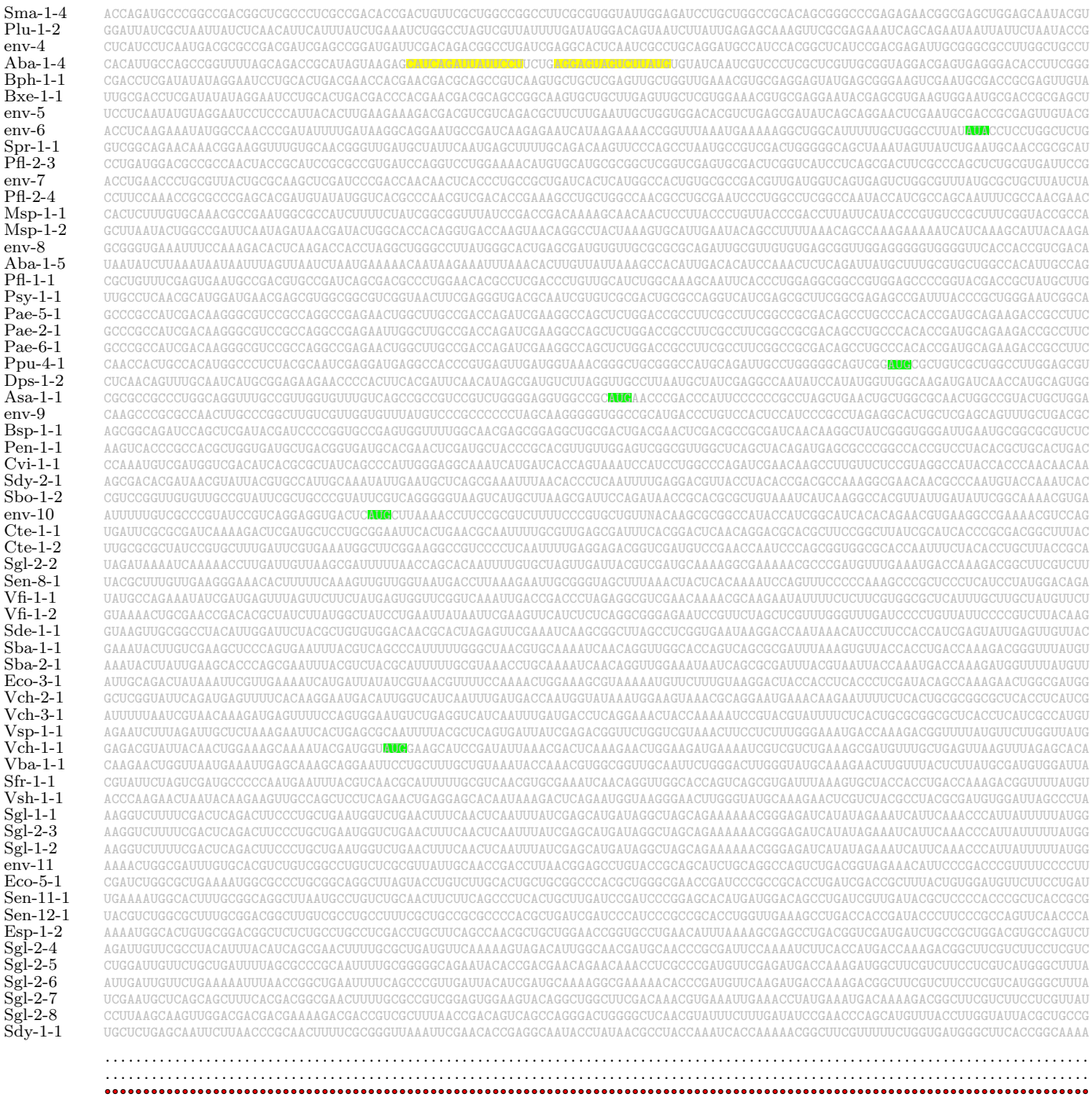

Sgl-2-15
Sty-1-1
Eco-1-1
Sen-4-4
Kph-1-1
Ype-2-2
env-24
Sen-1-1
Sen-14-2
env-25
Ppu-4-3
Ppu-2-1
Pen-1-3
Nmu-1-1
env-26
Avi-1-1
Pae-4-2
Ppu-2-2
Pen-1-4
Pen-1-5
Ppu-4-4
Pen-1-6
Pen-1-7
Ppu-4-5
Ppu-3-1
Ppu-3-2
Ppu-4-6
Ppu-1-2
Ppu-4-7
Ppu-3-3
Ppu-4-8
Ppu-4-9
env-27
Pen-1-8
Ppu-1-3
Ppu-4-10
Ppu-4-11
Pen-1-9
Pen-1-10
Pfl-2-7
Xor-1-1
env-28
Xca-1-1
Sma-1-5
Bdo-1-1
Pfl-2-8
Pfl-1-2
Ppu-4-12
Pen-1-11
Ppu-3-4
Ppu-2-3
Ppu-4-13
Ppu-1-4
env-29
Psy-3-1
Psy-1-2
Psy-1-3
Pfl-2-9
Pfl-2-10
Pfl-2-11
Pfl-2-12
env-30

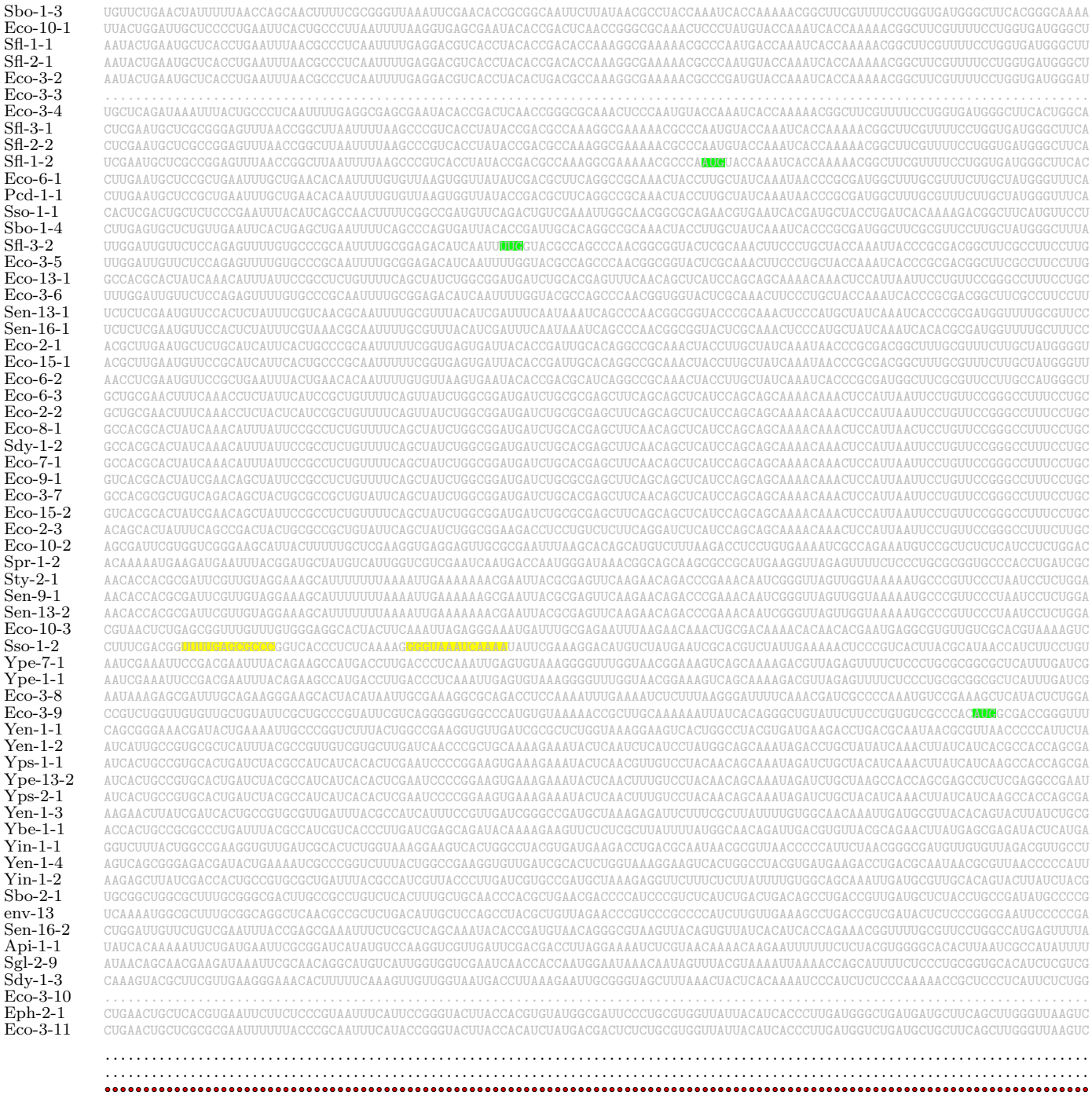

Psy-2-1
Pfl-2-13
Psy-2-2
Pfl-2-14
Psy-3-2
Pfl-2-15
Avi-1-2
Ppu-2-4
Pen-1-12
Pfl-2-16
Pfl-2-17
Pfl-2-18
Pfl-1-3
Pen-1-13
Pfl-2-19
Pfl-2-20
Pfl-2-21
Pfl-1-4
env-31
env-32
env-33
env-34
Pfl-2-22
env-35
Pfl-1-5
Psy-1-4
Psy-2-3
Pen-1-14
env-36
Pfl-2-23
Pfl-2-24
Pfl-2-25
Psy-2-4
env-37
env-38
env-39
Pfl-1-6
Ppu-3-5
env-40
Pen-1-15
Ppu-4-14
Pen-1-16
Ppu-4-15
env-41
Psy-3-3
Pae-3-1
Ppu-4-16
Pfl-1-7
Pfl-2-26
Pfl-2-27
env-42
env-43
Pfl-1-8
Avi-1-3
Psy-1-5
Psy-2-5
Pfl-2-28
env-44
Pfl-1-9
Ppu-4-17
Ppu-4-18
Pen-1-17

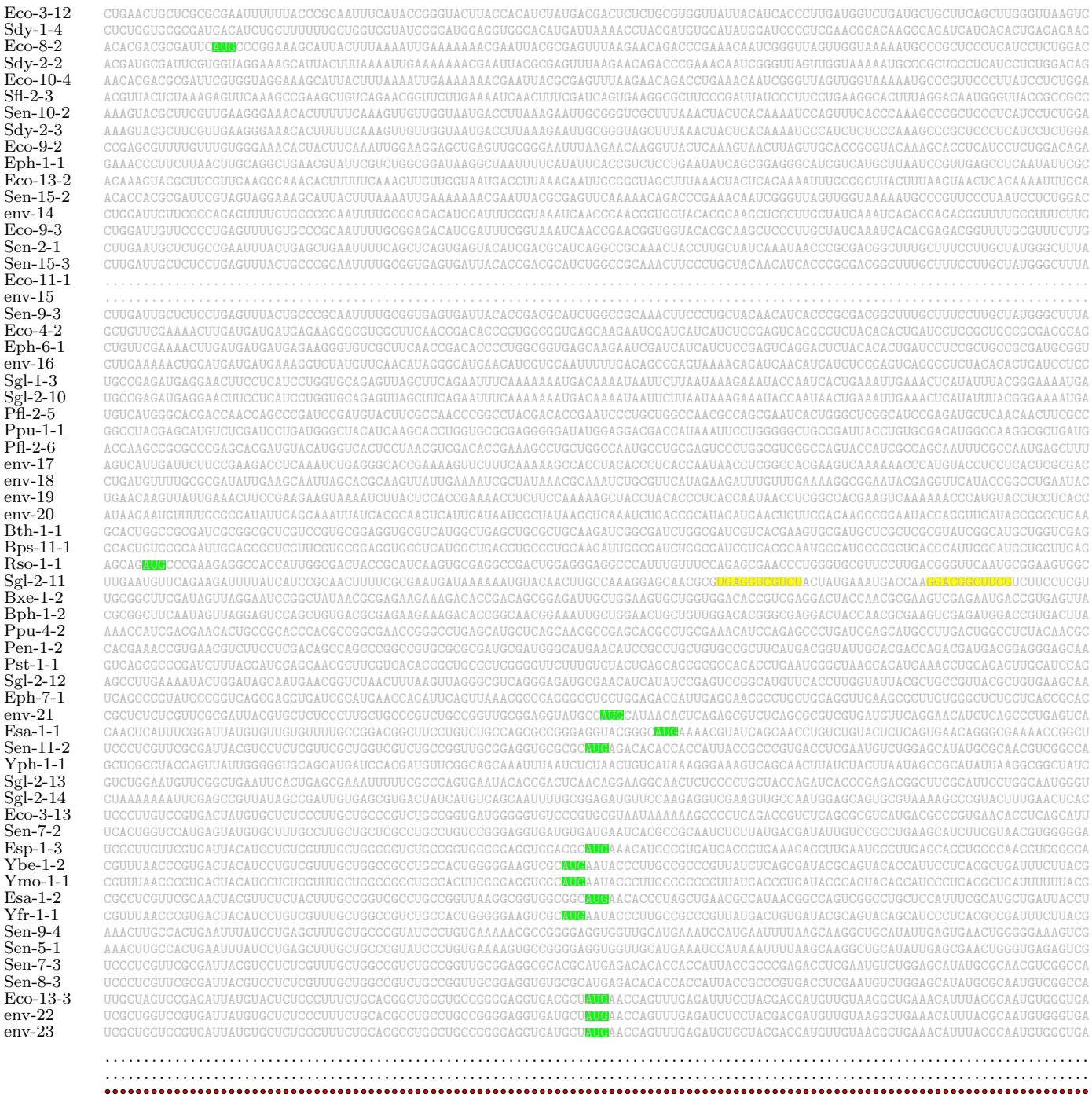

Pen-1-18
Ppu-4-19
Ppu-4-20
Pen-1-19
Pen-1-20
Pen-1-21
Ppu-3-6
Pen-1-22
Ppu-3-7
Ppu-1-5
Ppu-1-6
Pen-1-23
Pen-1-24
Pen-1-25
Ppu-4-21
Ppu-4-22
Ppu-4-23
Ppu-1-7
Ppu-1-8
Ppu-3-8
Pen-1-26
Ppu-2-5
Ppu-4-24
Ppu-4-25
Ppu-3-9
Pen-1-27
Ppu-3-10
Ppu-3-11
Pen-1-28
Psy-1-6
Ppu-4-26
Ppu-3-12
Ppu-1-9
Ppu-3-13
Ppu-1-10
Pen-1-29
env-45
env-46
Pme-1-1
Pme-1-2
Rso-2-1
Rpi-1-1
env-47
Reu-1-1
Rso-2-2
Reu-2-1
Rso-1-2
Rso-1-3
Bth-3-1
Bps-8-1
Bps-16-1
Bps-11-2
Bth-3-2
Bth-2-1
Bvi-1-1
Bth-3-3
Bps-16-2
Bps-7-1
Bps-8-2
Bps-5-1
Bps-18-1
Bok-2-1

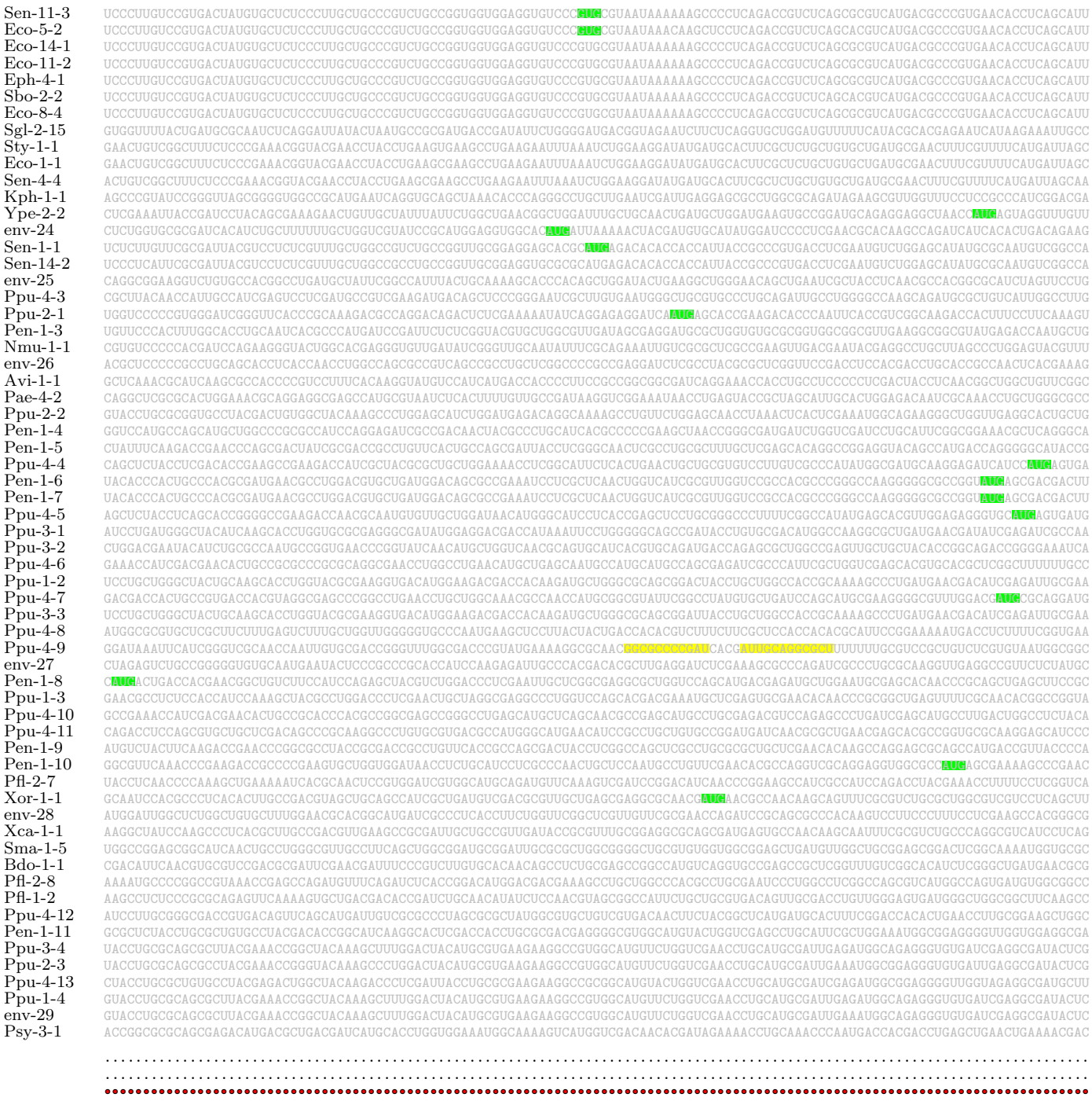

Bok-1-1
Bce-1-1
Bps-5-2
Bth-1-2
Bps-7-2
Bdo-1-2
Bps-9-1
Reu-1-2
Cvi-1-2
Bps-15-2
Rpi-2-1
Rpi-1-2
env-48
env-49
Bce-1-2
Bma-10-1
Bth-2-2
Bvi-1-2
env-50
Bce-2-1
Bps-3-1
Bmu-1-1
Bvi-1-3
env-51
Rso-1-4
Rpi-2-2
Rpi-1-3
Rso-2-3
Rso-1-5
Rpi-1-4
Rpi-2-3
Rpi-1-5
Rpi-2-4
Rpi-1-6
Rpi-2-5
env-52
Pen-1-30
Ppu-4-27
Pfl-2-29
Sgl-2-16
Sgl-2-17
Eco-8-5
Eco-3-14
Sen-3-1
Sgl-2-18
Eph-3-1
Sen-8-4
Aba-1-6
Aba-1-7
Aba-1-8
Aba-1-9
Asp-1-1
Aba-1-10
Aba-1-11
Aba-1-12
Aba-1-13
Aba-1-14
Aba-1-15
Aba-1-16
Aba-1-17
Aba-1-18
Aba-1-19

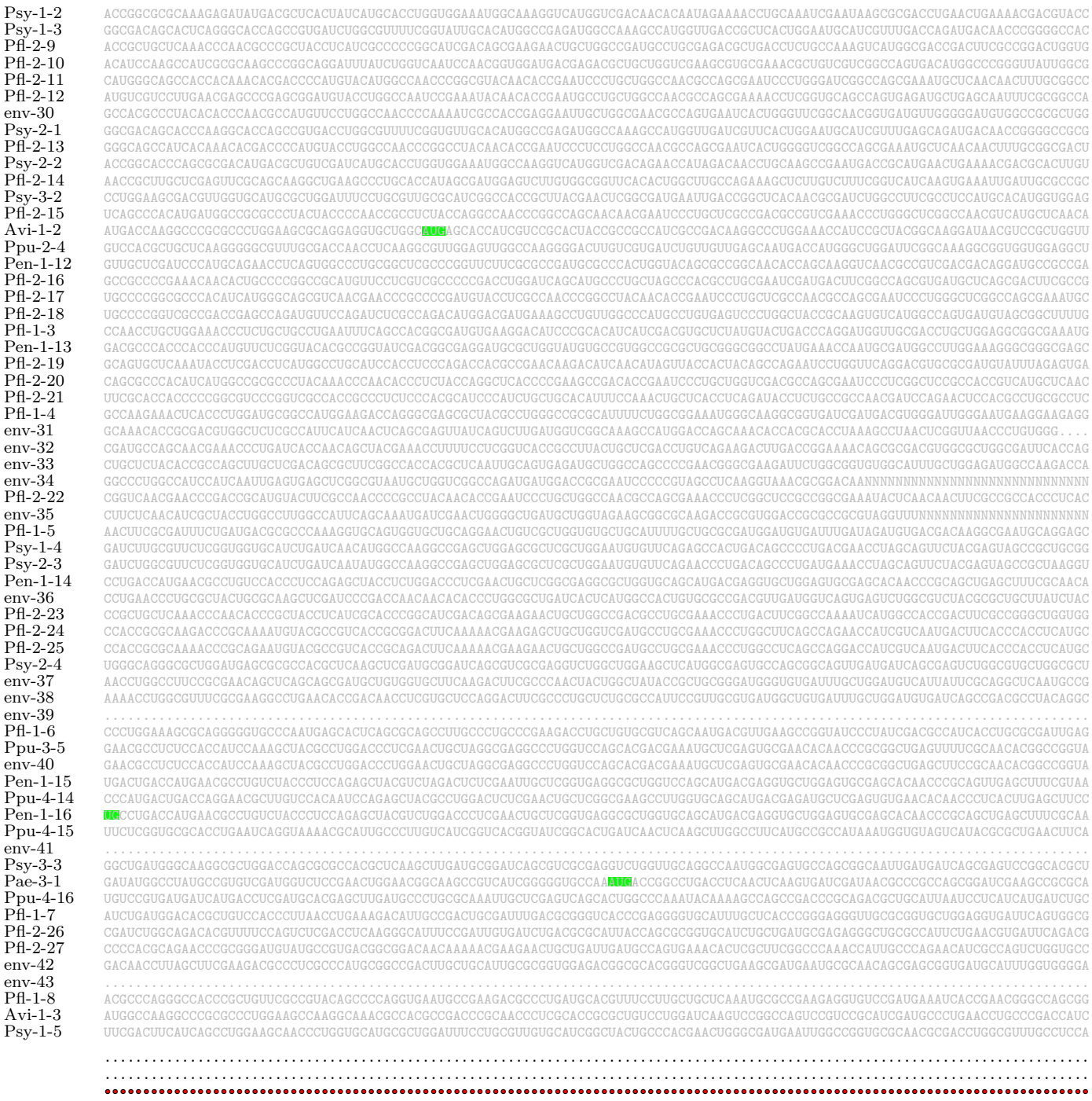
alignment positions $420 \cdots 546$

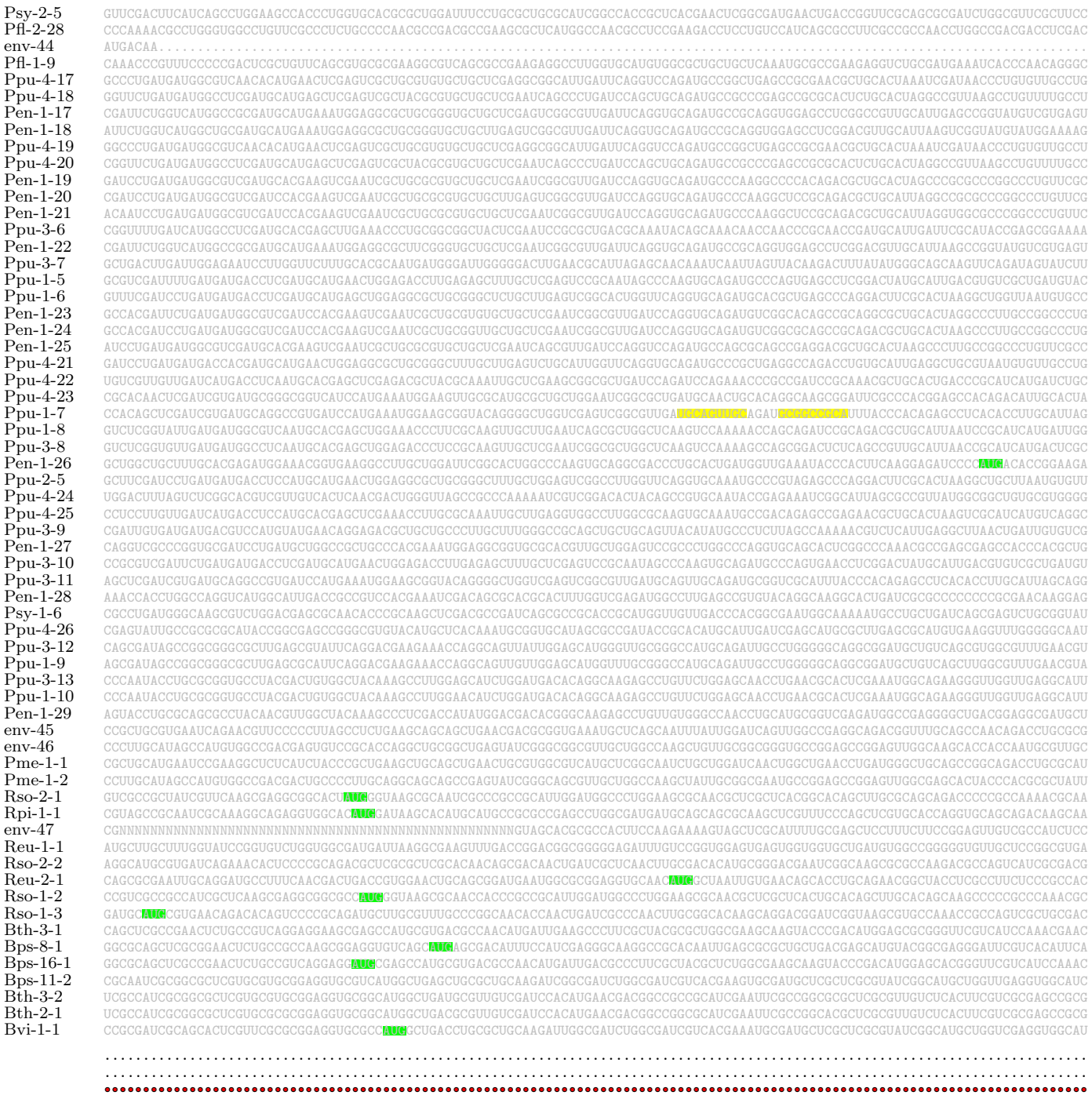

Sma-1-4
Plu-1-2
env-4
Aba-1-4
Bph-1-1
Bxe-1-1
env-5
env-6
Spr-1-1
Pfl-2-3
env-7
Pfl-2-4
Msp-1-1
Msp-1-2
env-8
Aba-1-5
Pfl-1-1
Psy-1-1
Pae-5-1
Pae-2-1
Pae-6-1
Ppu-4-1
Dps-1-2
Asa-1-1
env-9
Bsp-1-1
Pen-1-1
Cvi-1-1
Sdy-2-1
Sbo-1-2
env-10
Cte-1-1
Cte-1-2
Sgl-2-2
Sen-8-1
Vfi-1-1
Vfi-1-2
Sde-1-1
Sba-1-1
Sba-2-1
Eco-3-1
Vch-2-1
Vch-3-1
Vsp-1-1
Vch-1-1
Vba-1-1
Sfr-1-1
Vsh-1-1
Sgl-1-1
Sgl-2-3
Sgl-1-2
env-11
Eco-5-1
Sen-11-1
Sen-12-1
Esp-1-2
Sgl-2-4
Sgl-2-5
Sgl-2-6
Sgl-2-7
Sgl-2-8
Sdy-1-1

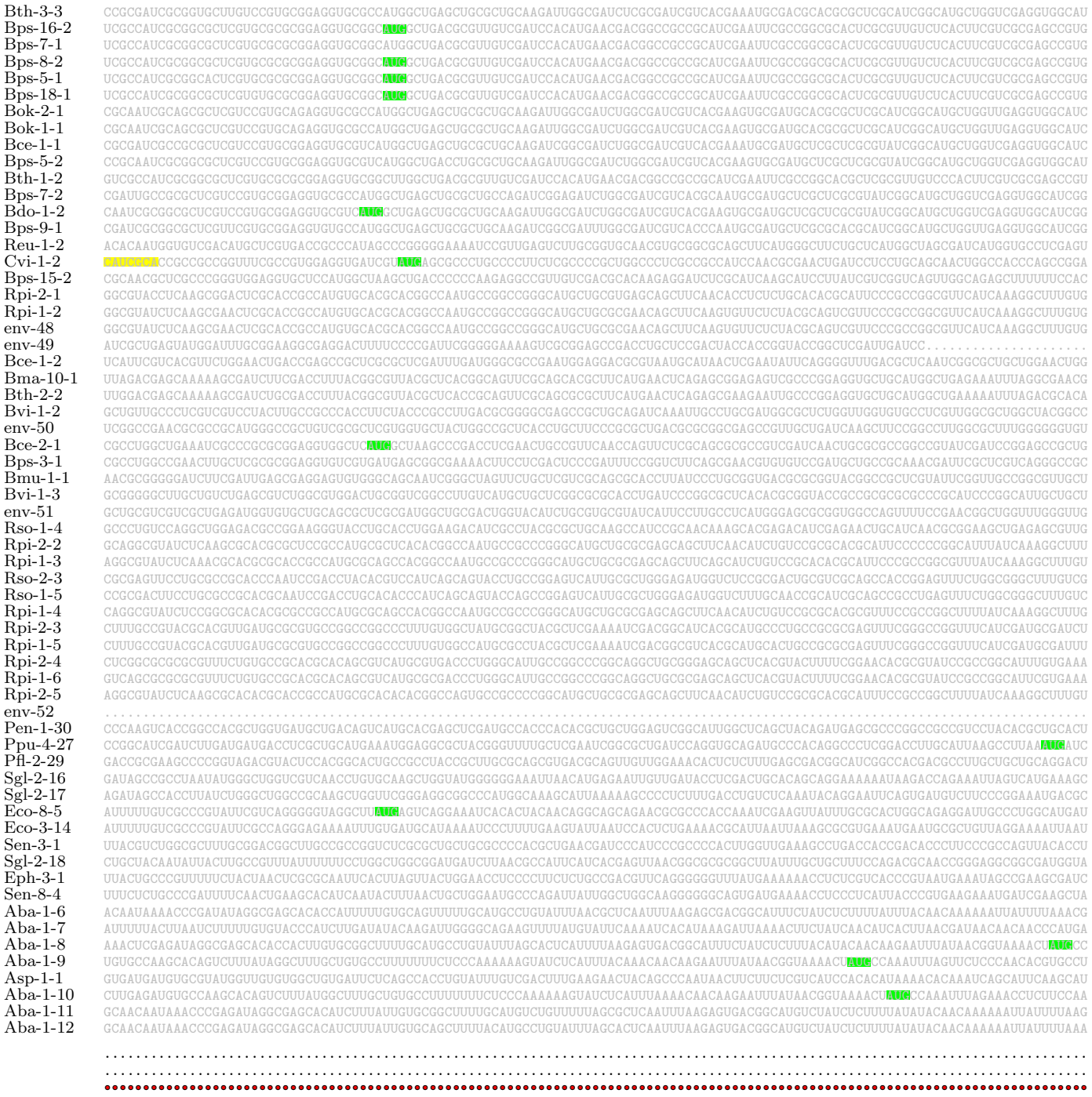

Sbo-1-3
Eco-10-1
Sfl-1-1
Sfl-2-1
Eco-3-2
Eco-3-3
Eco-3-4
Sfl-3-1
Sfl-2-2
Sfl-1-2
Eco-6-1
Pcd-1-1
Sso-1-1
Sbo-1-4
Sfl-3-2
Eco-3-5
Eco-13-1
Eco-3-6
Sen-13-1
Sen-16-1
Eco-2-1
Eco-15-1
Eco-6-2
Eco-6-3
Eco-2-2
Eco-8-1
Sdy-1-2
Eco-7-1
Eco-9-1
Eco-3-7
Eco-15-2
Eco-2-3
Eco-10-2
Spr-1-2
Sty-2-1
Sen-9-1
Sen-13-2
Eco-10-3
Sso-1-2
Ype-7-1
Ype-1-1
Eco-3-8
Eco-3-9
Yen-1-1
Yen-1-2
Yps-1-1
Ype-13-2
Yps-2-1
Yen-1-3
Ybe-1-1
Yin-1-1
Yen-1-4
Yin-1-2
Sbo-2-1
env-13
Sen-16-2
Api-1-1
Sgl-2-9
Sdy-1-3
Eco-3-10
Eph-2-1
Eco-3-11

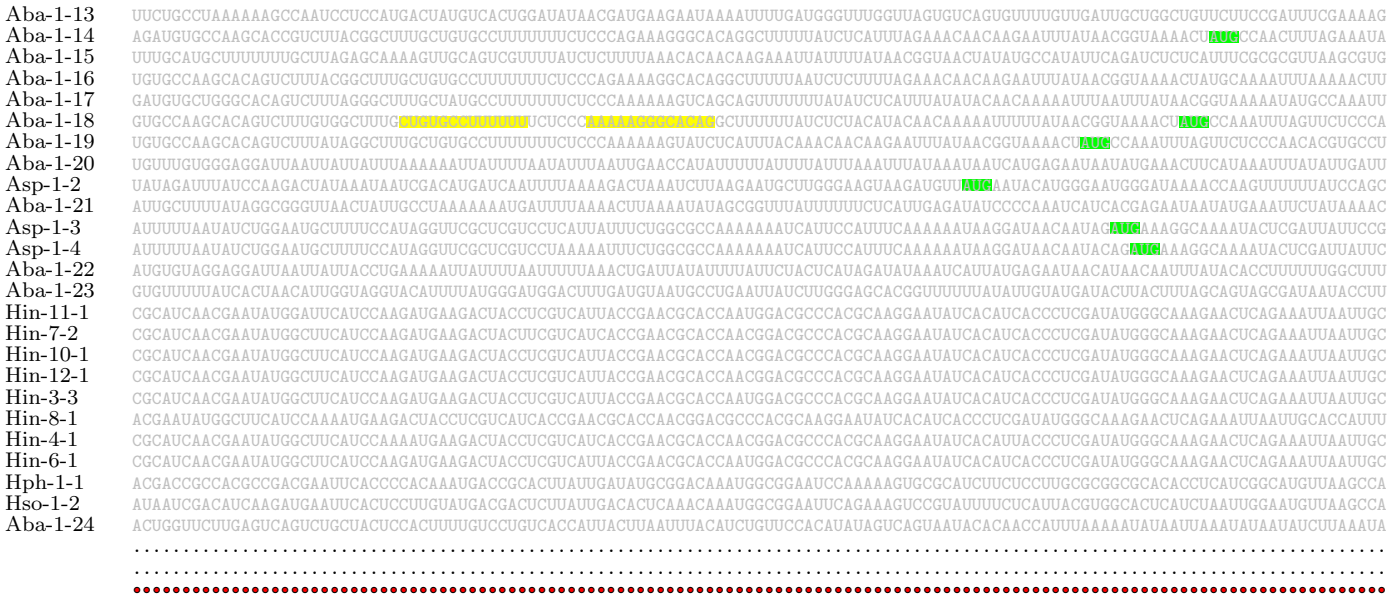

Eco-3-12
Sdy-1-4
Eco-8-2
Sdy-2-2
Eco-10-4
Sfl-2-3
Sen-10-2
Sdy-2-3
Eco-9-2
Eph-1-1
Eco-13-2
Sen-15-2
env-14
Eco-9-3
Sen-2-1
Sen-15-3
Eco-11-1
env-15
Sen-9-3
Eco-4-2
Eph-6-1
env-16
Sgl-1-3
Sgl-2-10
Pfl-2-5
Ppu-1-1
Pfl-2-6
env-17
env-18
env-19
env-20
Bth-1-1
Bps-11-1
Rso-1-1
Sgl-2-11
Bxe-1-2
Bph-1-2
Ppu-4-2
Pen-1-2
Pst-1-1
Sgl-2-12
Eph-7-1
env-21
Esa-1-1
Sen-11-2
Yph-1-1
Sgl-2-13
Sgl-2-14
Eco-3-13
Sen-7-2
Esp-1-3
Ybe-1-2
Ymo-1-1
Esa-1-2
Yfr-1-1
Sen-9-4
Sen-5-1
Sen-7-3
Sen-8-3
Eco-13-3
env-22
env-23


Sen-11-3
Eco-5-2
Eco-14-1
Eco-11-2
Eph-4-1
Sbo-2-2
Eco-8-4
Sgl-2-15
Sty-1-1
Eco-1-1
Sen-4-4
Kph-1-1
Ype-2-2
env-24
Sen-1-1
Sen-14-2
env-25
Ppu-4-3
Ppu-2-1
Pen-1-3
Nmu-1-1
env-26
Avi-1-1
Pae-4-2
Ppu-2-2
Pen-1-4
Pen-1-5
Ppu-4-4
Pen-1-6
Pen-1-7
Ppu-4-5
Ppu-3-1
Ppu-3-2
Ppu-4-6
Ppu-1-2
Ppu-4-7
Ppu-3-3
Ppu-4-8
Ppu-4-9
env-27
Pen-1-8
Ppu-1-3
Ppu-4-10
Ppu-4-11
Pen-1-9
Pen-1-10
Pfl-2-7
Xor-1-1
env-28
Xca-1-1
Sma-1-5
Bdo-1-1
Pfl-2-8
Pfl-1-2
Ppu-4-12
Pen-1-11
Ppu-3-4
Ppu-2-3
Ppu-4-13
Ppu-1-4
env-29
Psy-3-1


Psy-1-2
Psy-1-3
Pfl-2-9
Pfl-2-10
Pfl-2-11
Pfl-2-12
env-30
Psy-2-1
Pfl-2-13
Psy-2-2
Pfl-2-14
Psy-3-2
Pfl-2-15
Avi-1-2
Ppu-2-4
Pen-1-12
Pfl-2-16
Pfl-2-17
Pfl-2-18
Pfl-1-3
Pen-1-13
Pfl-2-19
Pfl-2-20
Pfl-2-21
Pfl-1-4
env-31
env-32
env-33
env-34
Pfl-2-22
env-35
Pfl-1-5
Psy-1-4
Psy-2-3
Pen-1-14
env-36
Pfl-2-23
Pfl-2-24
Pfl-2-25
Psy-2-4
env-37
env-38
env-39
Pfl-1-6
Ppu-3-5
env-40
Pen-1-15
Ppu-4-14
Pen-1-16
Ppu-4-15
env-41
Psy-3-3
Pae-3-1
Ppu-4-16
Pfl-1-7
Pfl-2-26
Pfl-2-27
env-42
env-43
Pfl-1-8
Avi-1-3
Psy-1-5


Psy-2-5
Pfl-2-28
env-44
Pfl-1-9
Ppu-4-17
Ppu-4-18
Pen-1-17
Pen-1-18
Ppu-4-19
Ppu-4-20
Pen-1-19
Pen-1-20
Pen-1-21
Ppu-3-6
Pen-1-22
Ppu-3-7
Ppu-1-5
Ppu-1-6
Pen-1-23
Pen-1-24
Pen-1-25
Ppu-4-21
Ppu-4-22
Ppu-4-23
Ppu-1-7
Ppu-1-8
Ppu-3-8
Pen-1-26
Ppu-2-5
Ppu-4-24
Ppu-4-25
Ppu-3-9
Pen-1-27
Ppu-3-10
Ppu-3-11
Pen-1-28
Psy-1-6
Ppu-4-26
Ppu-3-12
Ppu-1-9
Ppu-3-13
Ppu-1-10
Pen-1-29
env-45
env-46
Pme-1-1
Pme-1-2
Rso-2-1
Rpi-1-1
env-47
Reu-1-1
Rso-2-2
Reu-2-1
Rso-1-2
Rso-1-3
Bth-3-1
Bps-8-1
Bps-16-1
Bps-11-2
Bth-3-2
Bth-2-1
Bvi-1-1


Bth-3-3
Bps-16-2
Bps-7-1
Bps-8-2
Bps-5-1
Bps-18-1
Bok-2-1
Bok-1-1
Bce-1-1
Bps-5-2
Bth-1-2
Bps-7-2
Bdo-1-2
Bps-9-1
Reu-1-2
Cvi-1-2
Bps-15-2
Rpi-2-1
Rpi-1-2
env-48
env-49
Bce-1-2
Bma-10-1
Bth-2-2
Bvi-1-2
env-50
Bce-2-1
Bps-3-1
Bmu-1-1
Bvi-1-3
env-51
Rso-1-4
Rpi-2-2
Rpi-1-3
Rso-2-3
Rso-1-5
Rpi-1-4
Rpi-2-3
Rpi-1-5
Rpi-2-4
Rpi-1-6
Rpi-2-5
env-52
Pen-1-30
Ppu-4-27
Pfl-2-29
Sgl-2-16
Sgl-2-17
Eco-8-5
Eco-3-14
Sen-3-1
Sgl-2-18
Eph-3-1
Sen-8-4
Aba-1-6
Aba-1-7
Aba-1-8
Aba-1-9
Asp-1-1
Aba-1-10
Aba-1-11
Aba-1-12



17 c4-a1b1

17.1 Taxa

The taxonomy of each organism containing a putative c4-a1b1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual c4-a1b1 RNA in Sections 17.2 and 17.4.

abbrev. of hits	taxonomy of species
Eco-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* E110019
Eco-2-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. Sakai
Sen-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Newport *Salmonella enterica subsp. en*
Sen-2-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Newport *Salmonella enterica subsp. en*
Sen-3-1 to Sen-3-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Saintpaul *Salmonella enterica subsp. e*
Sty-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Salmonella typhimurium* LT2
Sbo-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella boydii* BS512
Pae-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* 2192
Pae-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* C3719
Pae-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PA7
Pae-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PACS2
Pae-5-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PAO1
Pae-6-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* UCBPP-PA14
Pen-1-1 to Pen-1-12	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas entomophila* L48
Pfl-1-1 to Pfl-1-2	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf-5
Pfl-2-1 to Pfl-2-2	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf0-1
Pme-1-1 to Pme-1-2	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas mendocina ymp*
Ppu-1-1 to Ppu-1-4	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* F1
Ppu-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* GB-1
Ppu-3-1 to Ppu-3-10	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* KT2440
Ppu-4-1 to Ppu-4-7	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* W619

Psy-1-1 to Psy-1-6	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. phaseolicola* 1448A
Psy-2-1 to Psy-2-5	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1 to Psy-3-3	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000
Sen-4-1	CVM23701Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar 4,[5],12:i:- *Salmonella enter*
Sen-5-1	GA_MM04042433Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Salmonella enterica subsp. enterica serovar* Javiana str
Sen-6-1	HI_N05-537Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Weltevreden *Salmonella en*
Eph-1-1	Viruses dsDNA viruses, no RNA stage Caudovirales Myoviridae P1-like viruses *Enterobacteria phage* P1
Eph-2-1	Viruses dsDNA viruses, no RNA stage Caudovirales Myoviridae P2-like viruses *Enterobacteria phage* P4
Api-1-1	Viruses dsDNA viruses, no RNA stage Caudovirales Podoviridae *Acyrthosiphon pisum secondary endosymbiont phage* 1
Eph-3-1	Viruses dsDNA viruses, no RNA stage Caudovirales Podoviridae *Enterobacteria phage* VT2-Sakai
Eph-4-1	Viruses dsDNA viruses, no RNA stage Caudovirales Podoviridae P22-like viruses *Enterobacteria phage phi*V10
Kph-1-1	Viruses dsDNA viruses, no RNA stage Caudovirales Siphoviridae *Klebsiella phage phi*KO2
Sco-1-1	Viruses dsDNA viruses, no RNA stage Caudovirales Siphoviridae Lambda-like viruses unclassified Lambda-like viruses *Stx*2 converting phage II (Stx2 converting bacteriophage II)
Eph-5-1	Viruses dsDNA viruses, no RNA stage Caudovirales Siphoviridae N15-like viruses *Enterobacteria phage* N15
Yph-1-1	Viruses dsDNA viruses, no RNA stage Caudovirales Siphoviridae *Yersinia phage* PY54
PGi-1-1	Viruses *Phage* Gifsy-1
env-1 to env-6	environmental samples

17.2 Gene contexts

Each c4-a1b1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one c4-a1b1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 17.3. The accession number of the sequence containing each c4-a1b1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 17.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Pme-1-1	NC_009439.1	-	1667511	1667459	COG2085 (COG2085)→ **RNA**→ hypo→
Psy-1-1	NC_005773.3	-	2874654	2874602	GAF (pfam01590)→ **RNA**→ COG3617 (COG3617)→
Pfl-2-1	NC_007492.1	-	4210840	4210788	←COG3950 (COG3950) **RNA**→ hypo→
Ppu-3-1	NC_002947.3	+	81504	81557	SUA5 (COG0009)→ **RNA**→ hypo→
Eph-1-1	NC_005856.1	+	48217	48269	**RNA**→
Pme-1-2	NC_009439.1	+	3776440	3776492	←hypo **RNA**→ hypo→ hypo→
env-1	JCVI_SCAF_1096627055672	-	3216	3164	**RNA**→
env-2	JCVI_SCAF_1096628137543	+	962	1014	? (GOS1771)→ **RNA**→ COG3617 (COG3617)? (GOS11900477)→
Eph-2-1	NC_001609.1	-	8705	8653	**RNA**→
Ppu-4-1	NZ_AAVY01000023.1	-	18674	18622	hypo→ **RNA**→ repA (cd01125)→
Yph-1-1	NC_005069.1	+	23607	23659	**RNA**→
Sbo-1-1	NZ_AAKA01000001.1	-	39678	39626	hypo→ **RNA**→ hypo→
Kph-1-1	NC_005857.1	-	32735	32683	**RNA**→
Eph-5-1	NC_001901.1	-	28502	28449	**RNA**→
Psy-3-1	NC_004578.1	+	3479907	3479959	lipoprotein, putative→ **RNA**→ COG3617 (COG3617)→
Psy-2-1	NC_007005.1	+	4545039	4545091	←DUF1349 (pfam07081) **RNA**→ COG3617 (COG3617)→ adh_short (pfam00106)→
Psy-1-2	NC_005773.3	-	1124972	1124920	hypo→ **RNA**→ putative inner membrane protein→
Pae-6-1	NC_008463.1	+	4285944	4285996	←hypo **RNA**→ hypo→
Pae-1-1	NZ_AAKW01000002.1	-	1037	985	←hypo **RNA**→ hypo→
Pae-5-1	NC_002516.2	-	1348652	1348600	←hypo **RNA**→ hypo→
Pae-2-1	NZ_AAKV01000003.1	-	6779	6727	←hypo **RNA**→ hypo→

Pae-4-1	NZ_AAQW01000001.1	-	1918787	1918735	←hypo **RNA**→ hypo→
Ppu-3-2	NC_002947.3	+	759489	759541	Acetyltransf_1 (pfam00583)→ **RNA**→ ←GntP_permease (pfam02447)
env-3	JCVI_SCAF_1096628052827	+	1418	1470	? (GOS11069324)→ **RNA**→ ? (GOS11725977)→
Pae-3-1	NC_009656.1	-	373603	373551	INT_P4 (cd00801)→ **RNA**→ hypo→ hypo→ transcriptional regulator, AlpA family→ Bbp36→
					zf-CHC2 (pfam01807)→ hypo→ hypo→ COG5519 (COG5519)→
Pfl-1-1	NC_004129.6	+	5946430	5946482	Acetyltransf_1 (pfam00583)→ **RNA**→ hypo→
env-4	JCVI_SCAF_1096627616436	+	788	840	←BRCT (smart00292)? (GOS11076434) **RNA**→ COG3617 (COG3617)→
Pfl-1-2	NC_004129.6	+	5731844	5731896	hypo→ **RNA**→ COG5519 (COG5519)TOPRIM_primases (cd01029)→
					chromosome segregation SMC protein, putative→
Psy-1-3	NC_005773.3	+	4631506	4631559	hypo→ **RNA**→ hypo→
Psy-2-2	NC_007005.1	+	4827814	4827867	hypo→ **RNA**→ hypo→
Pfl-2-2	NC_007492.1	+	2273208	2273260	CsgG (COG1462)→ **RNA**→ hypo→ hypo→
Ppu-1-1	NC_009512.1	+	1788494	1788546	hypo→ **RNA**→ hypo→
Ppu-3-3	NC_002947.3	-	4856696	4856644	←hypo **RNA**→ ←PBPb (smart00062)
Ppu-3-4	NC_002947.3	+	5989185	5989237	KefB (COG0475)Kch (COG1226)→ **RNA**→ hypo→
Ppu-4-2	NZ_AAVY01000017.1	+	134874	134926	←Glycos_transf_2 (pfam00535) **RNA**→ hypo→ hypo→ hypo→
Psy-2-3	NC_007005.1	+	5651557	5651610	←ComJ (COG2961) **RNA**→ hypo→
Psy-3-2	NC_004578.1	-	447545	447492	←ComJ (COG2961) **RNA**→ hypo→
Psy-1-4	NC_005773.3	+	5451255	5451308	←ComJ (COG2961) **RNA**→ ←hypo
Psy-3-3	NC_004578.1	-	6095069	6095018	DnaJ (pfam00226)→ **RNA**→ hypo→
Psy-2-4	NC_007005.1	-	5828745	5828694	DnaJ (pfam00226)→ **RNA**→ hypo→
Psy-1-5	NC_005773.3	-	5617216	5617165	DnaJ (pfam00226)→ **RNA**→ hypo→
Ppu-4-3	NZ_AAVY01000018.1	+	10292	10344	←lipoprotein, putative **RNA**→ hypo (cons)→
Ppu-4-4	NZ_AAVY01000002.1	+	296036	296087	OprD (pfam03573)→ **RNA**→ hypo (cons)→
Pen-1-1	NC_008027.1	+	3960544	3960595	←NUDIX (pfam00293) **RNA**→ hypo→
Pen-1-2	NC_008027.1	+	3853964	3854015	←hypo **RNA**→ hypo→
Pen-1-3	NC_008027.1	-	910473	910421	←HemA (COG0373) **RNA**→ hypo→
Ppu-3-5	NC_002947.3	-	1995826	1995774	dTDP_sugar_isom (pfam00908)→ **RNA**→ COG1835 (COG1835)→
Ppu-1-2	NC_009512.1	+	4999822	4999873	←PsiE (pfam06146) **RNA**→ hypo→ hypo→
Ppu-3-6	NC_002947.3	-	1420112	1420060	←Rick_17kDa_Anti (pfam05433) **RNA**→ hypo→
Ppu-1-3	NC_009512.1	-	1439694	1439642	←Rick_17kDa_Anti (pfam05433) **RNA**→ hypo→ hypo→
Pen-1-4	NC_008027.1	+	4896010	4896062	FecA (COG4772)→ **RNA**→ hypo→
Pen-1-5	NC_008027.1	+	1180693	1180743	DUF479 (pfam04336)→ **RNA**→ hypo→
Pen-1-6	NC_008027.1	-	3292013	3291963	hypo→ **RNA**→ hypo→
Pen-1-7	NC_008027.1	-	1493248	1493198	←LysR (COG0583) **RNA**→ hypo→
Pen-1-8	NC_008027.1	+	2592660	2592712	FhuE (COG4773)→ **RNA**→ hypo→ hypo→
Ppu-3-7	NC_002947.3	-	2855987	2855935	←hypo **RNA**→ ←DltE (COG0300)
Ppu-4-5	NZ_AAVY01000008.1	-	11017	10965	Aes (COG0657)→ **RNA**→ hypo (cons)→
Ppu-4-6	NZ_AAVY01000001.1	+	243123	243175	←AbgB (COG1473) **RNA**→ hypo (cons)→
Ppu-3-8	NC_002947.3	-	5149089	5149038	hypo→ **RNA**→ hypo→
Pen-1-9	NC_008027.1	+	4355520	4355572	←hypo **RNA**→ hypo→ hypo→
Ppu-3-9	NC_002947.3	+	5390561	5390612	hypo→ **RNA**→ hypo→
Ppu-2-1	NZ_AAXR01000004.1	+	377710	377762	←COG3714 (COG3714) **RNA**→ hypo (cons)→
Pen-1-10	NC_008027.1	+	5582431	5582483	←DUF1234 (pfam06821) **RNA**→ hypo→
Ppu-1-4	NC_009512.1	+	3624986	3625038	hypo→ **RNA**→ hypo→
Ppu-3-10	NC_002947.3	-	2858122	2858070	hypo→ **RNA**→ yedG domain protein→
Ppu-4-7	NZ_AAVY01000026.1	-	47983	47932	hypo (cons)→ **RNA**→ hypo (cons)→
Pen-1-11	NC_008027.1	+	2692719	2692771	COG2863 (COG2863)→ **RNA**→ hypo→
Pen-1-12	NC_008027.1	+	2652730	2652782	←RpoE (COG1595) **RNA**→ hypo→
env-5	JCVI_SCAF_1101668291562	+	813	865	R_hydratase (cd03449)→ **RNA**→ ? (GOS11008091)→
Psy-1-6	NC_005773.3	+	368971	369023	TauA (COG0715)→ **RNA**→ hypo→
Psy-2-5	NC_007005.1	+	363657	363710	TauA (COG0715)→ **RNA**→ hypo→ hypo→
Eph-4-1	NC_007804.2	-	20476	20424	**RNA**→

env-6	BAAX01012263.1	-	1357	1305	**RNA**→ COG3617 (COG3617)→
Eco-1-1	NZ_AAJW01000017.1	+	44561	44613	Dam (pfam05869)→ **RNA**→ COG3561 (COG3561)→ ANT (pfam03374)→
Eco-2-1	NC_002695.1	+	1263176	1263228	NinE→ **RNA**→ COG3561 (COG3561)→ COG3646 (COG3646)ANT (pfam03374)→ NinG→ hypo→ antitermination protein Q→
Eph-3-1	NC_000902.1	+	17164	17216	**RNA**→
Sco-1-1	NC_004914.1	+	49052	49104	**RNA**→
Sen-3-1	NZ_ABAN01000025.1	-	4146	4094	**RNA**→
Sen-5-1	NZ_ABEH01000004.1	-	43804	43752	**RNA**→
Sen-4-1	NZ_ABAO01000067.1	-	2587	2535	**RNA**→
Sen-1-1	NZ_ABEN01000001.1	-	2762914	2762862	**RNA**→
Sen-2-1	NZ_ABEW01000016.1	+	12497	12549	**RNA**→
Sen-6-1	NZ_ABFF01000012.1	+	15360	15412	**RNA**→
Sen-3-2	NZ_ABAN01000002.1	+	49	101	**RNA**→
Sty-1-1	NC_003197.1	-	2762880	2762828	antiterminator-like protein→ **RNA**→ antirepressor-like protein→
PGi-1-1	NC_010392.1	-	34329	34277	**RNA**→
Api-1-1	NC_000935.1	-	28276	28224	**RNA**→

17.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 17.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one c4-a1b1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

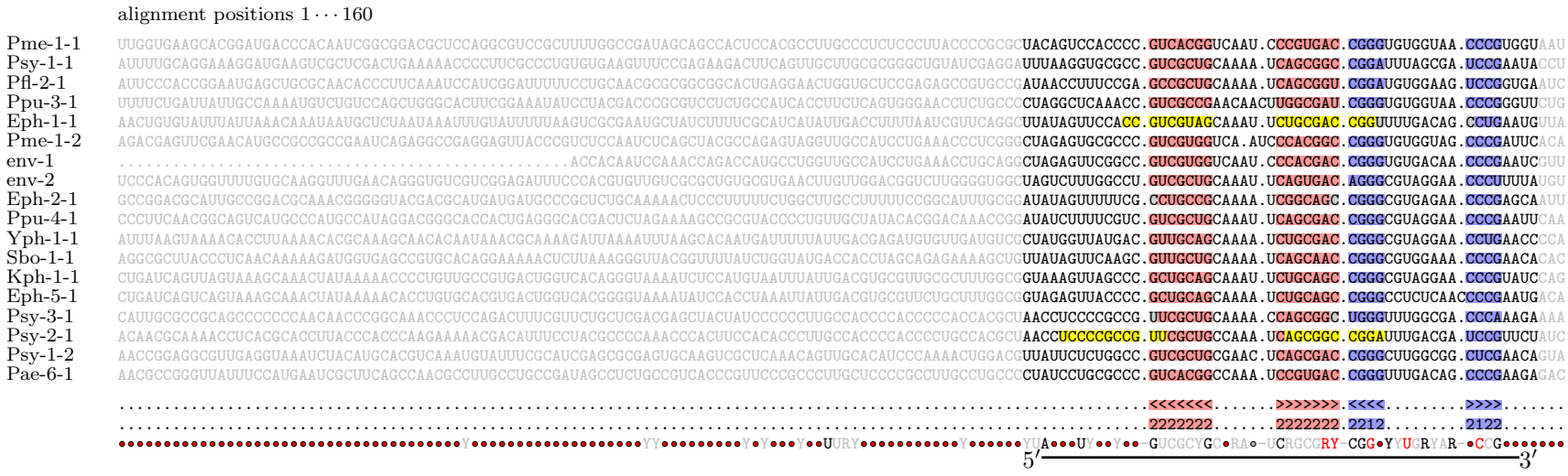
cd00801 (1) Bacteriophage P4 integrase.
cd01029 (1) TOPRIM_primases: The topoisomerase-primase (TORPIM) nucleotidyl transferase/hydrolase domain found in the active site regions of bacterial DnaG-type primases and their homologs.
cd01125 (1) Hexameric Replicative Helicase RepA.
cd03449 (1) (R)-hydratase [(R)-specific enoyl-CoA hydratase]
COG0009 (1) Putative translation factor (SUA5) [Translation, ribosomal structure and biogenesis]
COG0300 (1) Short-chain dehydrogenases of various substrate specificities [General function prediction only]
COG0373 (1) Glutamyl-tRNA reductase [Coenzyme metabolism]
COG0475 (1) Kef-type K+ transport systems, membrane components [Inorganic ion transport and metabolism]
COG0583 (1) Transcriptional regulator [Transcription]
COG0657 (1) Esterase/lipase [Lipid metabolism]
COG0715 (2) ABC-type nitrate/sulfonate/bicarbonate transport systems, periplasmic components [Inorganic ion transport and metabolism]
COG1226 (1) Kef-type K+ transport systems, predicted NAD-binding component [Inorganic ion transport and metabolism]
COG1462 (1) Uncharacterized protein involved in formation of curli polymers [Cell envelope biogenesis, outer membrane]
COG1473 (1) Metal-dependent amidase/aminoacylase/carboxypeptidase [General function prediction only]
COG1595 (1) DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog [Transcription]
COG1835 (1) Predicted acyltransferases [Lipid metabolism]
COG2085 (1) Predicted dinucleotide-binding enzymes [General function prediction only]
COG2863 (1) Cytochrome c553 [Energy production and conversion]
COG2961 (3) Protein involved in catabolism of external DNA [General function prediction only]
COG3561 (2) Phage anti-repressor protein [Transcription]
COG3617 (6) Prophage antirepressor [Transcription]
COG3646 (1) Uncharacterized phage-encoded protein [Function unknown]
COG3714 (1) Predicted membrane protein [Function unknown]
COG3950 (1) Predicted ATP-binding protein involved in virulence [General function prediction only]
COG4772 (1) Outer membrane receptor for Fe3+-dicitrate [Inorganic ion transport and metabolism]
COG4773 (1) Outer membrane receptor for ferric coprogen and ferric-rhodotorulic acid [Inorganic ion transport and metabolism]
COG5519 (2) Superfamily II helicase and inactivated derivatives [DNA replication, recombination, and repair]
GOS11008091 (1) ?
GOS11069324 (1) ?
GOS11076434 (1) ?
GOS11725977 (1) ?
GOS11900477 (1) ?
GOS1771 (1) ?
pfam00106 (1) short chain dehydrogenase.
pfam00226 (3) DnaJ domain.
pfam00293 (1) NUDIX domain.
pfam00535 (1) Glycosyl transferase.

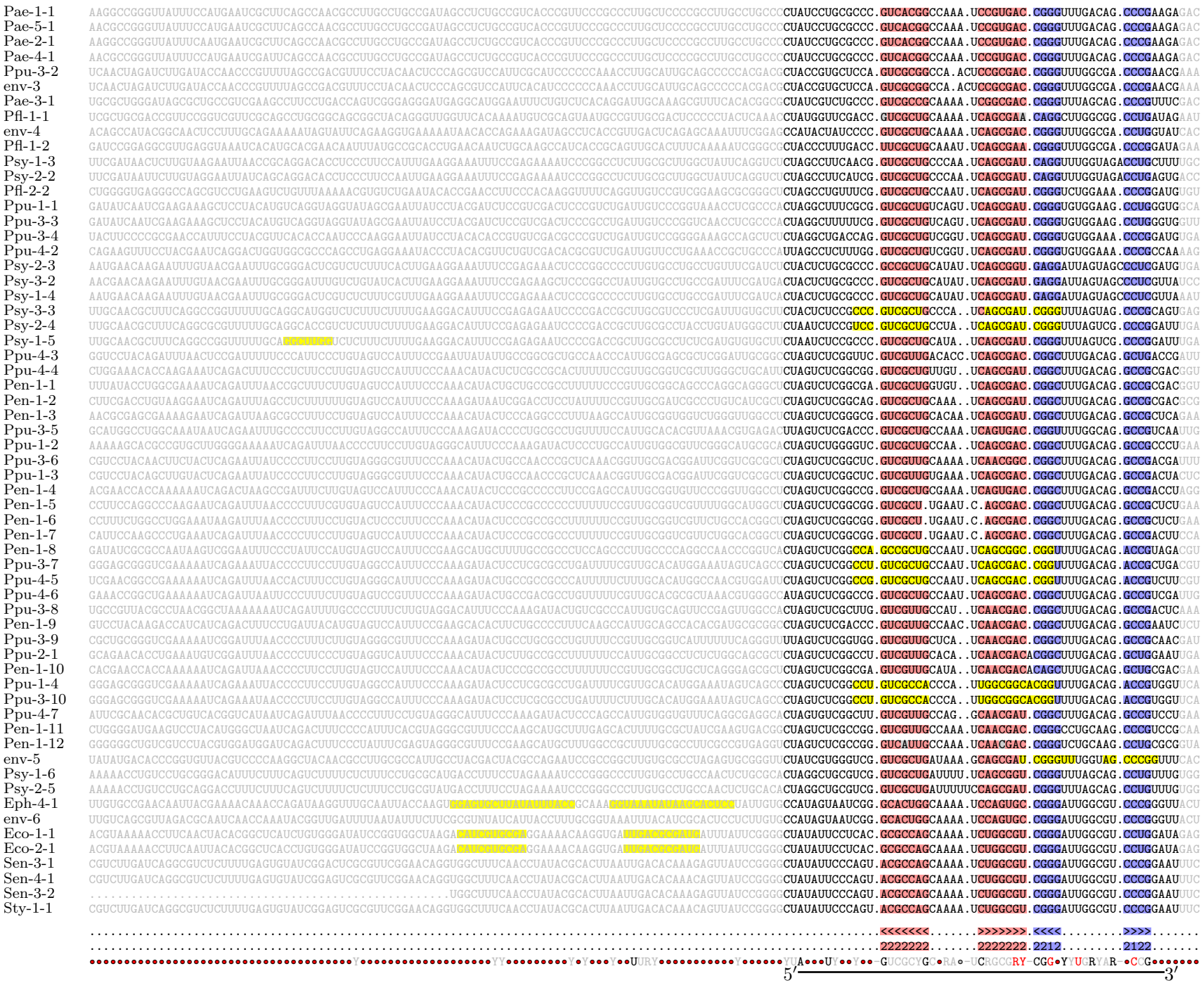
pfam00583 (2) Acetyltransferase (GNAT) family.
pfam00908 (1) dTDP-4-dehydrorhamnose 3,5-epimerase.
pfam01590 (1) GAF domain.
pfam01807 (1) CHC2 zinc finger.
pfam02447 (1) GntP family permease.
pfam03374 (2) Phage antirepressor protein.
pfam03573 (1) outer membrane porin, OprD family.
pfam04336 (1) Protein of unknown function, DUF479.
pfam05433 (2) Rickettsia 17 kDa surface antigen.
pfam05869 (1) DNA N-6-adenine-methyltransferase (Dam).
pfam06146 (1) Phosphate-starvation-inducible E.
pfam06821 (1) Protein of unknown function (DUF1234).
pfam07081 (1) Protein of unknown function (DUF1349).
smart00062 (1) Bacterial periplasmic substrate-binding proteins; bacterial proteins, eukaryotic ones are in PBPe
smart00292 (1) breast cancer carboxy-terminal domain;

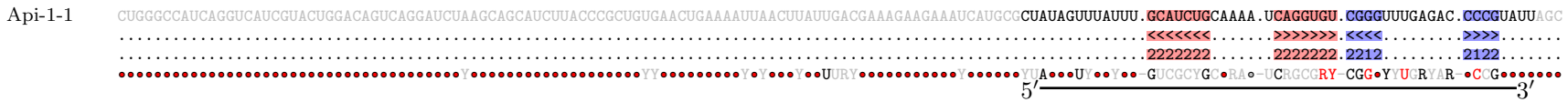
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Eph-3-1, PGi-1-1, Sco-1-1, Sen-1-1, Sen-2-1, Sen-5-1, Sen-6-1

17.4 Multiple-sequence alignment

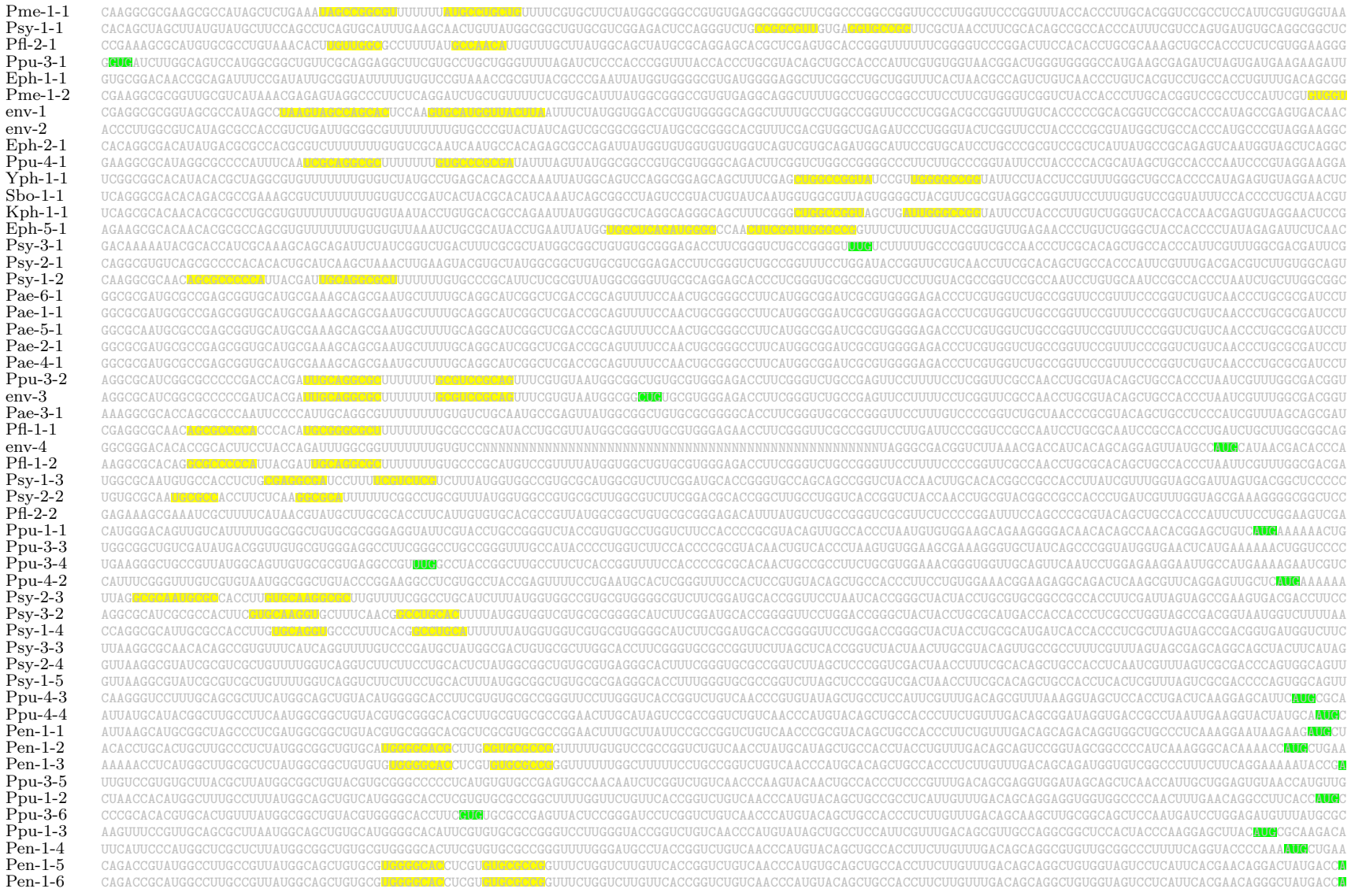
Each c4-a1b1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 17.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The c4-a1b1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



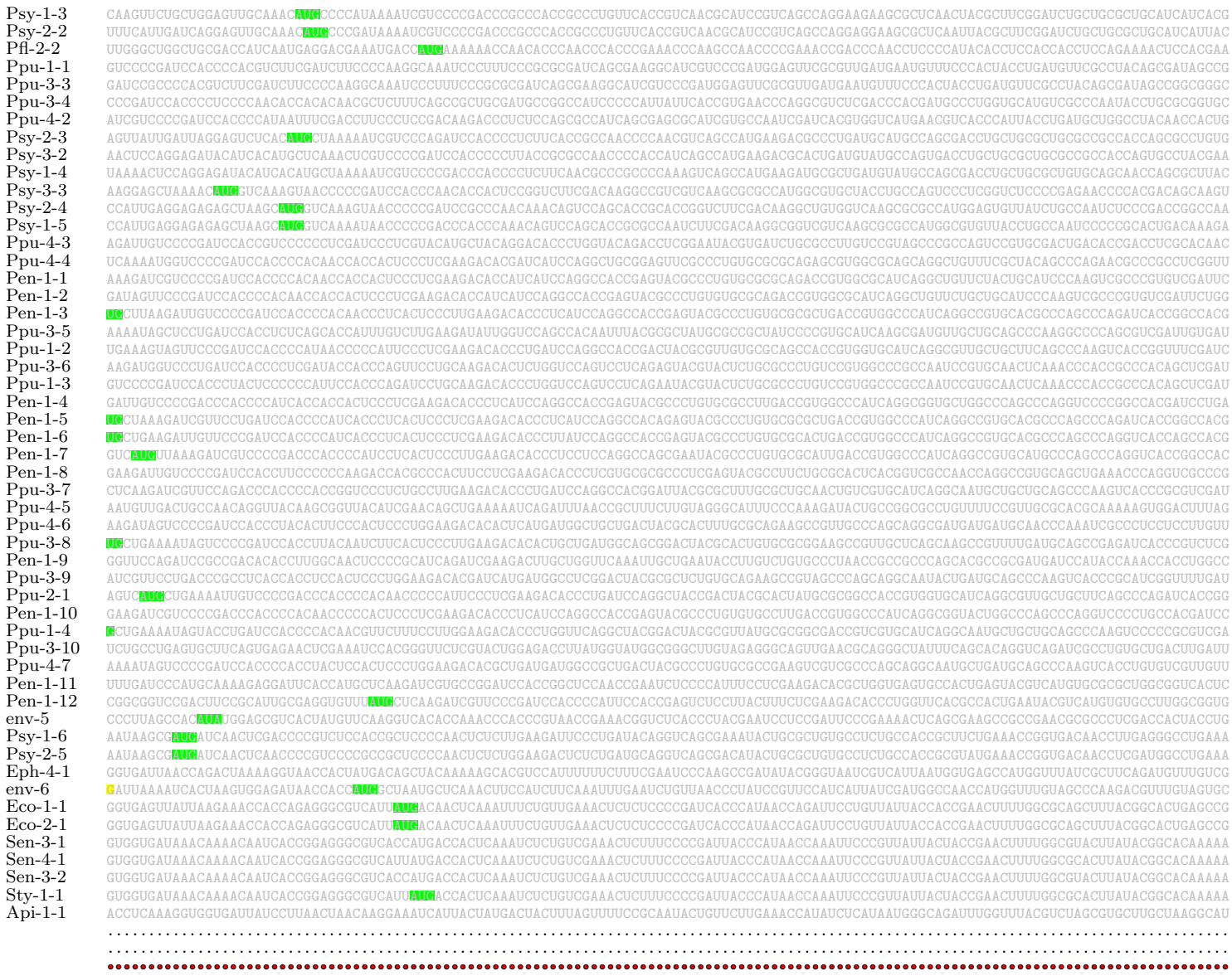

Pae-1-1
Pae-5-1
Pae-2-1
Pae-4-1
Ppu-3-2
env-3
Pae-3-1
Pfl-1-1
env-4
Pfl-1-2
Psy-1-3
Psy-2-2
Pfl-2-2
Ppu-1-1
Ppu-3-3
Ppu-3-4
Ppu-4-2
Psy-2-3
Psy-3-2
Psy-1-4
Psy-3-3
Psy-2-4
Psy-1-5
Ppu-4-3
Ppu-4-4
Pen-1-1
Pen-1-2
Pen-1-3
Ppu-3-5
Ppu-1-2
Ppu-3-6
Ppu-1-3
Pen-1-4
Pen-1-5
Pen-1-6
Pen-1-7
Pen-1-8
Ppu-3-7
Ppu-4-5
Ppu-4-6
Ppu-3-8
Pen-1-9
Ppu-3-9
Ppu-2-1
Pen-1-10
Ppu-1-4
Ppu-3-10
Ppu-4-7
Pen-1-11
Pen-1-12
env-5
Psy-1-6
Psy-2-5
Eph-4-1
env-6
Eco-1-1
Eco-2-1
Sen-3-1
Sen-4-1
Sen-3-2
Sty-1-1
5′
3′



alignment positions 161 · · · 320



alignment positions 321 · · · 457



18 Chlorobi-1

18.1 Taxa

The taxonomy of each organism containing a putative Chlorobi-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Chlorobi-1 RNA in Sections 18.3 and 18.5.

abbrev. of hits	taxonomy of species
Cli-1-1 to Cli-1-3	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium limicola* DSM 245
Cph-1-1 to Cph-1-2	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium phaeobacteroides* BS1
Cph-2-1 to Cph-2-3	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium phaeovibrioides* DSM 265 (Prosthecochloris vibrioformis
Plu-1-1 to Plu-1-3	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Pelodictyon luteolum* DSM 273
Pae-1-1 to Pae-1-4	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae *Prosthecochloris aestuarii* DSM 271

18.2 Notes

Superscript numbers are used to annotate any Chlorobi-1 RNA that has special characteristics (described below). These numeric annotations will be used in Sections 18.3 and 18.5. The meaning of the numbers are as follows.

[1] Accepted with E-value 1.9 in global miniscan.

18.3 Gene contexts

Each Chlorobi-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Chlorobi-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 18.4. The accession number of the sequence containing each Chlorobi-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 18.1. Superscript numbers refer to annotations listed in Section 18.2

abbrev.	Seq. accession		5′ at	3′ at	genes
[1]Pae-1-1	NZ_AAIJ01000001.1	-	66513	66450	←hypo (cons) **RNA**→ LMWPc (cd00115)→
Cph-2-1	NC_009337.1	-	146066	146001	hypo→ **RNA**→ ←Pol_V (cd01700)
Cph-2-2	NC_009337.1	+	970580	970645	hypo→ **RNA**→ chlorosome envelope protein B→
Pae-1-2	NZ_AAIJ01000007.1	-	10387	10320	←AtpD (COG0055) **RNA**→ ←TPR (cd00189)COG2956 (COG2956)COG4783 (COG4783)
Cph-1-1	NZ_AAIC01000009.1	+	56947	57014	←RadC (COG2003) **RNA**→ ←LDH-like_MDH (cd01339)
Cph-1-2	NZ_AAIC01000002.1	-	18795	18730	PAD_porph (pfam04371)→ **RNA**→ ←EXOIII (smart00479)
Pae-1-3	NZ_AAIJ01000012.1	-	17426	17357	PAD_porph (pfam04371)→ **RNA**→ **RNA**→ ←LDH-like_MDH (cd01339)
Pae-1-4	NZ_AAIJ01000012.1	-	17100	17033	(shown above)
Plu-1-1	NC_007512.1	-	1700572	1700507	hypo→ **RNA**→ **RNA**→ hypo→
Plu-1-2	NC_007512.1	-	1700346	1700281	(shown above)
Plu-1-3	NC_007512.1	-	1700120	1700054	hypo→ **RNA**→ ←LDH-like_MDH (cd01339)
Cph-2-3	NC_009337.1	-	1489337	1489270	PAD_porph (pfam04371)→ **RNA**→ ←LDH-like_MDH (cd01339)
Cli-1-1	NZ_AAHJ01000012.1	-	10407	10340	hypo→ **RNA**→ **RNA**→ **RNA**→ hypo→
Cli-1-2	NZ_AAHJ01000012.1	-	10694	10627	(shown above)
Cli-1-3	NZ_AAHJ01000012.1	-	10125	10058	(shown above)

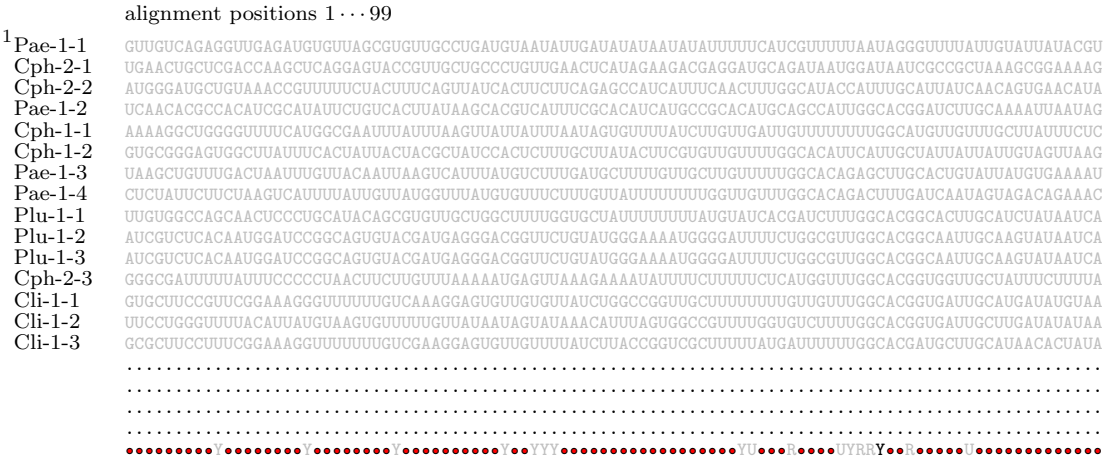
18.4 Conserved domains

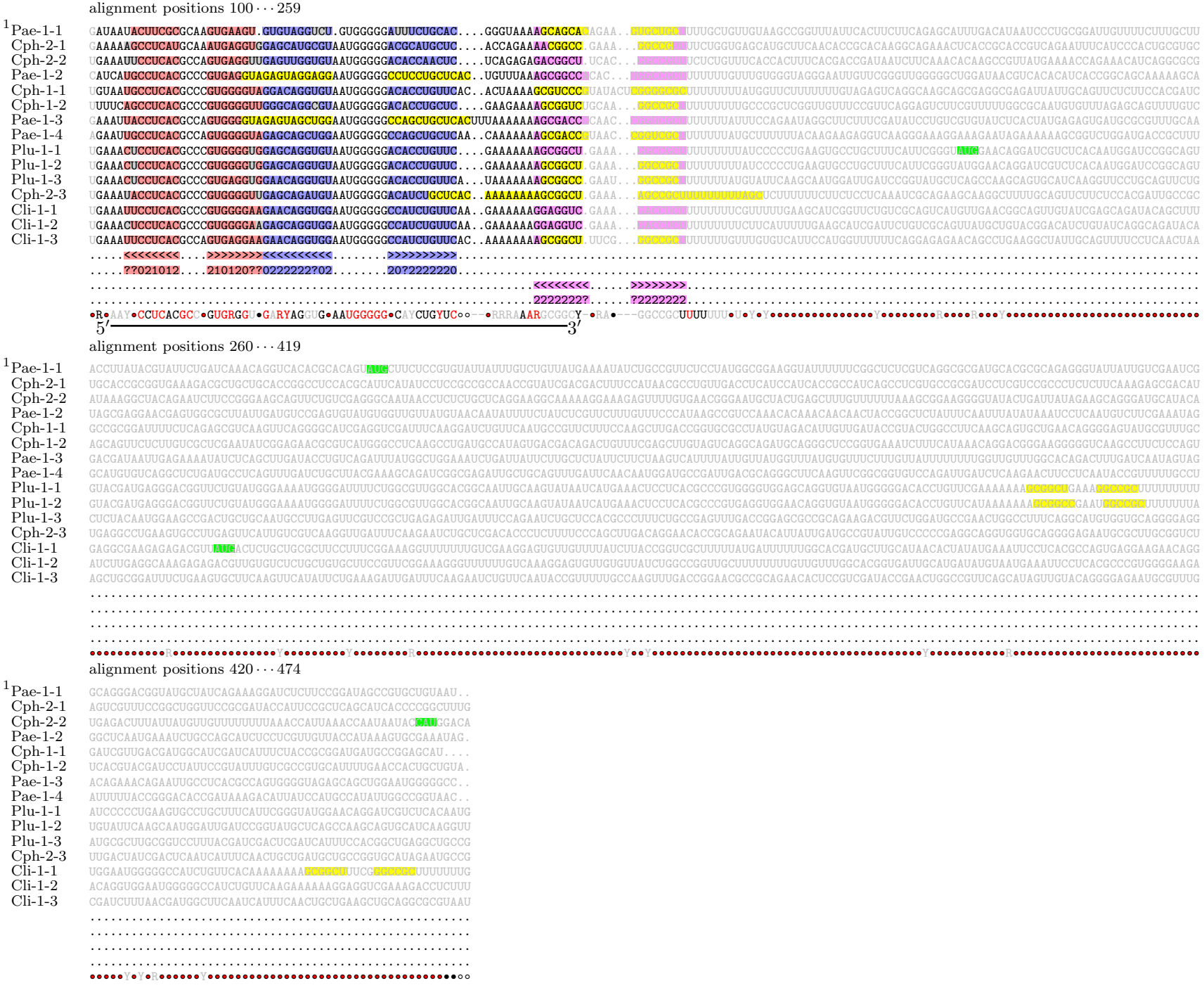
Conserved domains found in protein-coding genes listed in Section 18.3 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Chlorobi-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

D cd00115 (1) Substituted updates: Aug 22, 2001
d cd00189 (1) Tetratricopeptide repeat domain; typically contains 34 amino acids [WLF]-X(2)-[LIM]-[GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-X(2)-[ASL]-X(4)-[PKE]
d cd01339 (4) LDH-like structure and DMH enzymatic activity; member of the family of NAD-dependent 2-hydroxycarboxylate dehydrogenases.
d cd01700 (1) Pol V was discovered in Escherichia coli as Umuc and UmuD proteins induced by UV.
u COG0055 (1) F0F1-type ATP synthase, beta subunit [Energy production and conversion]
u COG2003 (1) DNA repair proteins [DNA replication, recombination, and repair]
d COG2956 (1) Predicted N-acetylglucosaminyl transferase [Carbohydrate transport and metabolism]
d COG4783 (1) Putative Zn-dependent protease, contains TPR repeats [General function prediction only]
u pfam04371 (3) Porphyromonas-type peptidyl-arginine deiminase.
d smart00479 (1) exonuclease domain in DNA-polymerase alpha and epsilon chain, ribonuclease T and other exonucleases

18.5 Multiple-sequence alignment

Each Chlorobi-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 18.1. Superscript numbers refer to annotations listed in Section 18.2. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Chlorobi-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.




alignment positions 100 · · · 259
1
Pae-1-1
Cph-2-1
Cph-2-2
Pae-1-2
Cph-1-1
Cph-1-2
Pae-1-3
Pae-1-4
Plu-1-1
Plu-1-2
Plu-1-3
Cph-2-3
Cli-1-1
Cli-1-2
Cli-1-3
5′
3′
alignment positions 260 · · · 419
1
Pae-1-1
Cph-2-1
Cph-2-2
Pae-1-2
Cph-1-1
Cph-1-2
Pae-1-3
Pae-1-4
Plu-1-1
Plu-1-2
Plu-1-3
Cph-2-3
Cli-1-1
Cli-1-2
Cli-1-3
alignment positions 420 · · · 474
1
Pae-1-1
Cph-2-1
Cph-2-2
Pae-1-2
Cph-1-1
Cph-1-2
Pae-1-3
Pae-1-4
Plu-1-1
Plu-1-2
Plu-1-3
Cph-2-3
Cli-1-1
Cli-1-2
Cli-1-3

19 Chlorobi-RRM

19.1 Taxa

The taxonomy of each organism containing a putative Chlorobi-RRM RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Chlorobi-RRM RNA in Sections 19.2 and 19.4.

abbrev. of hits	taxonomy of species
Cte-1-1 to Cte-1-2	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobaculum *Chlorobium tepidum* TLS
Cch-1-1	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium chlorochromatii* CaD3
Cfe-1-1 to Cfe-1-2	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium ferrooxidans* DSM 13031
Cli-1-1 to Cli-1-2	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium limicola* DSM 245
Cph-1-1 to Cph-1-2	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium phaeobacteroides* BS1
Cph-2-1 to Cph-2-2	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium phaeobacteroides* DSM 266
Cph-3-1 to Cph-3-2	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium phaeovibrioides* DSM 265 (Prosthecochloris vibrioformis
Plu-1-1 to Plu-1-2	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Pelodictyon luteolum* DSM 273
Pph-1-1 to Pph-1-2	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Pelodictyon phaeoclathratiforme* BU-1
Pae-1-1 to Pae-1-2	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae *Prosthecochloris aestuarii* DSM 271
Hch-1-1	Bacteria Proteobacteria Gammaproteobacteria Oceanospirillales Hahellaceae *Hahella chejuensis* KCTC 2396

19.2 Gene contexts

Each Chlorobi-RRM RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Chlorobi-RRM RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 19.3. The accession number of the sequence containing each Chlorobi-RRM RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 19.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Hch-1-1	NC_007645.1	+	455464	455534	**RNA**→ RRM (smart00360)→
Cph-3-1	NC_009337.1	-	1000654	1000582	**RNA**→ RRM (smart00360)→ BaeS (COG0642)→
Cte-1-1	NC_002932.3	-	127933	127857	**RNA**→ RRM (smart00360)→
Pae-1-1	NC_011059.1	-	1344490	1344426	**RNA**→ RRM (smart00360)→
Cli-1-1	NC_010803.1	+	2339792	2339872	**RNA**→ RRM (smart00360)→
Cfe-1-1	NZ_AASE01000008.1	+	61479	61554	**RNA**→ RRM (smart00360)→
Cph-1-1	NC_010831.1	-	1507528	1507452	**RNA**→ RRM (smart00360)→
Cte-1-2	NC_002932.3	+	1802605	1802684	**RNA**→ ←hypo
Cph-3-2	NC_009337.1	-	178328	178253	**RNA**→ RRM (smart00360)→
Cph-1-2	NC_010831.1	-	100324	100249	**RNA**→ RRM (smart00360)→
Plu-1-1	NC_007512.1	-	94199	94127	**RNA**→ RRM (smart00360)→
Pae-1-2	NC_011059.1	-	134270	134199	**RNA**→ RRM (smart00360)→
Cli-1-2	NC_010803.1	-	100693	100622	**RNA**→ RRM (smart00360)→
Cph-2-1	NC_008639.1	+	3066282	3066355	**RNA**→ RRM (smart00360)→
Cch-1-1	NC_007514.1	+	2033621	2033693	**RNA**→ RRM (smart00360)→
Plu-1-2	NC_007512.1	+	1162497	1162569	**RNA**→ RRM (smart00360)→ BaeS (COG0642)→ OmpR (COG0745)→

Cfe-1-2	NZ_AASE01000039.1	+	9214	9286	**RNA**→
Pph-1-1	NC_011060.1	+	1458931	1459003	**RNA**→ RRM (smart00360)→
Pph-1-2	NC_011060.1	+	2187184	2187256	**RNA**→ hypo→ RRM (smart00360)→ hypo→
Cph-2-2	NC_008639.1	+	1452008	1452080	**RNA**→ RRM (smart00360)→

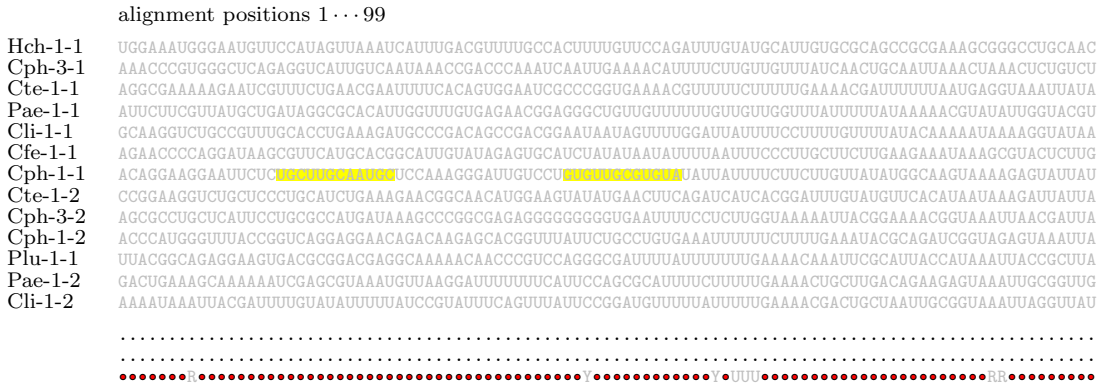
19.3 Conserved domains

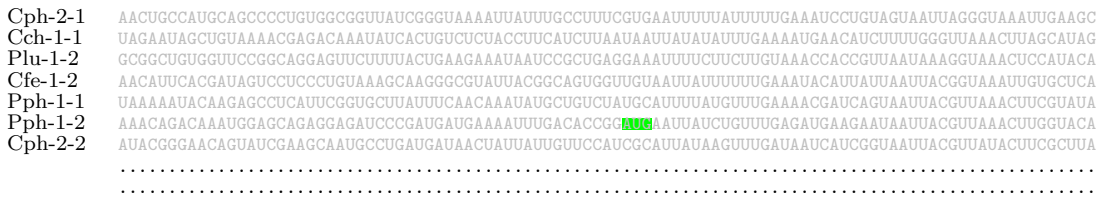
Conserved domains found in protein-coding genes listed in Section 19.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Chlorobi-RRM RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

COG0642 (2) Signal transduction histidine kinase [Signal transduction mechanisms]
COG0745 (1) Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain [Signal transduction mechanisms / Transcription]
smart00360 (18) RNA recognition motif;

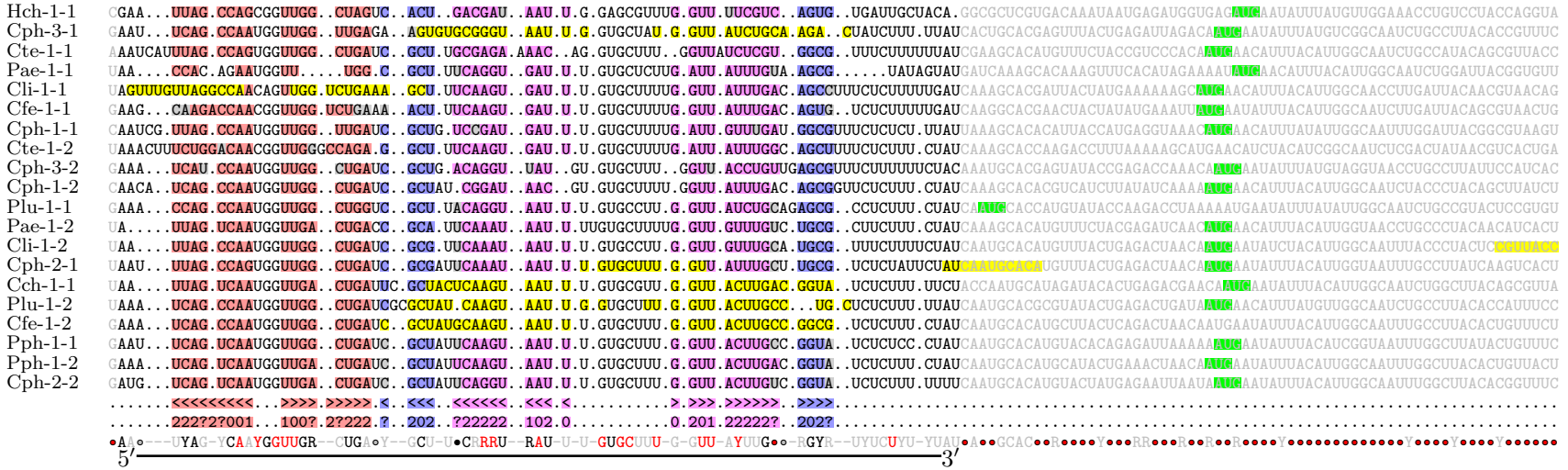
19.4 Multiple-sequence alignment

Each Chlorobi-RRM RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 19.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Chlorobi-RRM RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

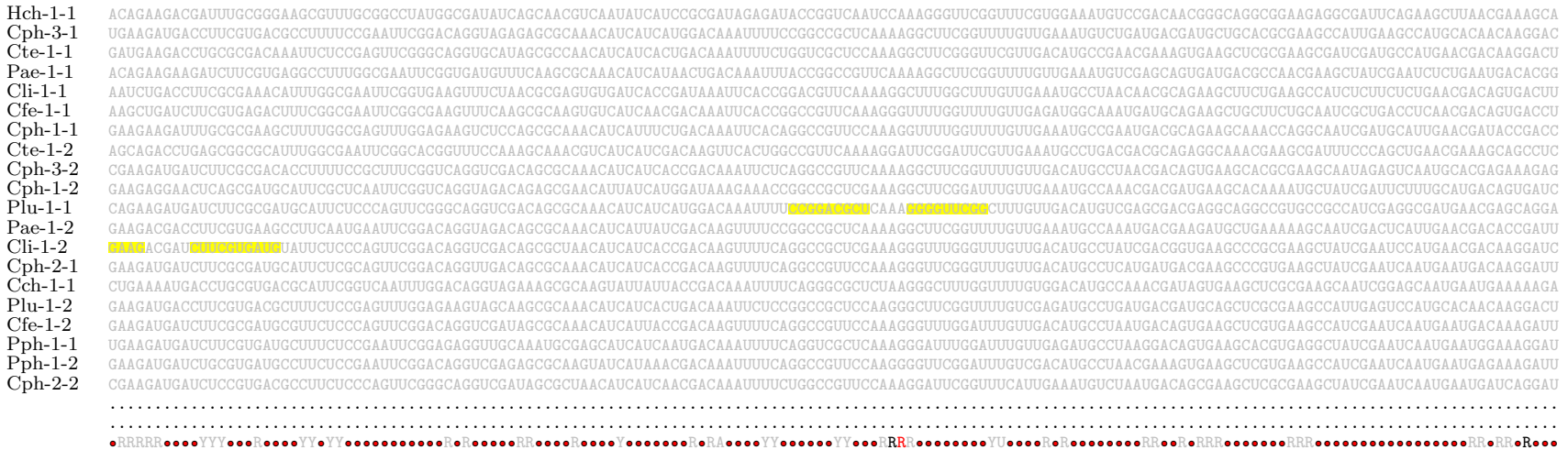


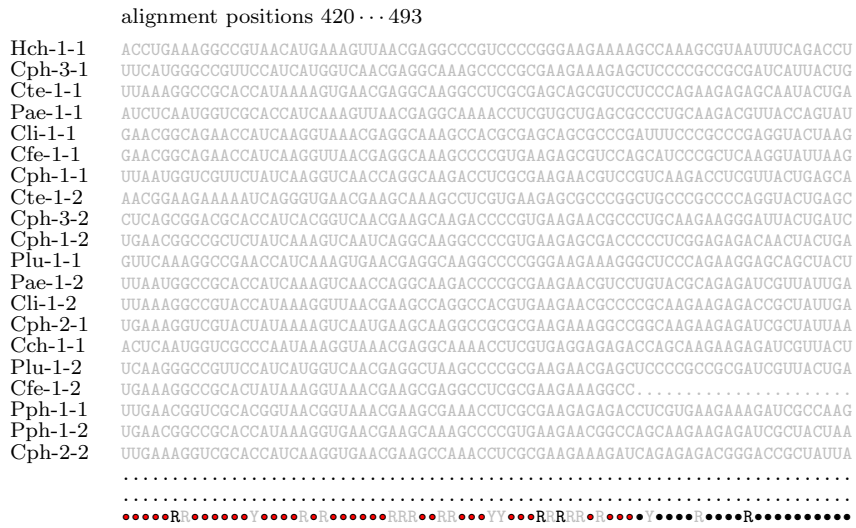


alignment positions 100 · · · 259



alignment positions 260 · · · 419





20 Chloroflexi-1

20.1 Taxa

The taxonomy of each organism containing a putative Chloroflexi-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Chloroflexi-1 RNA in Sections 20.2 and 20.4.

abbrev. of hits	taxonomy of species
Cag-1-1 to Cag-1-3	Bacteria Chloroflexi Chloroflexales Chloroflexaceae *Chloroflexus aggregans* DSM 9485

20.2 Gene contexts

Each Chloroflexi-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Chloroflexi-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 20.3. The accession number of the sequence containing each Chloroflexi-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 20.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Cag-1-1	NZ_AAUI01000002.1	+	27961	28045	hypo→ **RNA**→ hypo→
Cag-1-2	NZ_AAUI01000026.1	+	1295	1378	←RarD (COG2962) **RNA**→ ATS1 (COG5184)→
Cag-1-3	NZ_AAUI01000026.1	+	3992	4075	ATS1 (COG5184)→ **RNA**→ hypo→

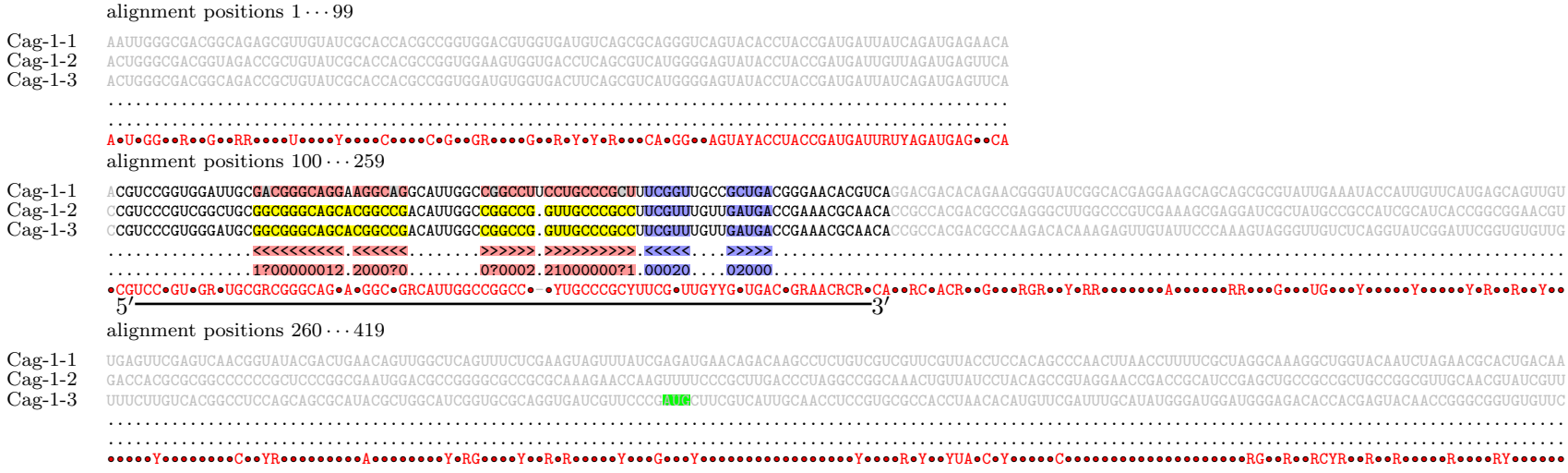
20.3 Conserved domains

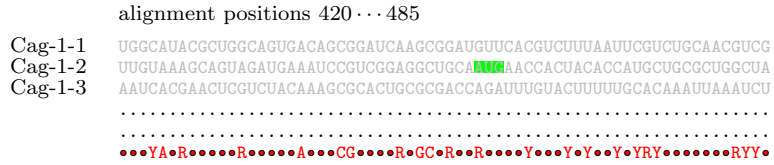
Conserved domains found in protein-coding genes listed in Section 20.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Chloroflexi-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

u COG2962 (1) Predicted permeases [General function prediction only]
uD COG5184 (2) Alpha-tubulin suppressor and related RCC1 domain-containing proteins [Cell division and chromosome partitioning / Cytoskeleton]

20.4 Multiple-sequence alignment

Each Chloroflexi-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 20.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Chloroflexi-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.




alignment positions 420 · · · 485
Cag-1-1 UGGCAUACGCUGGCAGUGACAGCGGAUCAAGCGGAUGUUCACGUCUUUAAUUCGUCUGCAACGUCG
Cag-1-2 UUGUAAAGCAGUAGAUGAAAUCCGUCGGAGGCUGCAAUGAACCACUACACCAUGCUGCGCUGGCUA
Cag-1-3 AAUCACGAACUCGUCUACAAAGCGCACUGCGCGACCAGAUUUGUACUUUUUGCACAAAUUAAAUCU
· · · ·

Eve-1-1	NZ_AAVL02000037.1	+	174692	174858	←hypo **RNA**→ MarR (pfam01047)→
env-9	BAAV01001515.1	-	1917	1759	3HCDH (pfam00725)→ **RNA**→ SCAD/SBCAD (cd01158)→
env-10	BAAV01002929.1	-	1019	863	hypo→ **RNA**→ ←ABC_MsbA_C (cd03251)
env-11	BAAU01003785.1	+	137	308	hypo→ **RNA**→ ←hypo
env-12	BAAY01014504.1	+	21	201	**RNA**→ ←EAL (cd01948)
env-13	BAAU01003080.1	+	9	185	**RNA**→ hypo→
env-14	BAAU01005865.1	-	1412	1225	PurL (COG0047)→ **RNA**→ hypo→
env-15	BABG01004091.1	+	375	562	←hypo **RNA**→ ←GlgA (COG0297)
env-16	BAAV01002036.1	-	1046	870	←hypo **RNA**→ hypo→
env-17	hgutS8_s8_169515	-	1029	842	←HisC (COG0079) **RNA**→ hypo→
env-18	hgutS8_s8_169849	+	467	654	←UPF0029 (pfam01205) **RNA**→ Transgly (pfam00912)→
env-19	BABB01002899.1	-	2303	2116	**RNA**→ ←GlgA (COG0297)
Eve-1-2	NZ_AAVL02000036.1	-	99484	99318	←hypo **RNA**→ ←hypo
env-20	BAAX01023897.1	-	957	790	**RNA**→ ArgC (COG0002)→
Eve-1-3	NZ_AAVL02000035.1	+	139817	139954	←HTH_XRE (cd00093)COG1917 (COG1917) **RNA**→ HIS2 (COG1387)→
Eve-1-4	NZ_AAVL02000031.1	+	48980	49147	hypo→ **RNA**→ hypo→
env-21	BABD01004363.1	-	554	380	hypo→ **RNA**→
Eve-1-5	NZ_AAVL02000027.1	+	21462	21630	GlgB (COG0296)→ **RNA**→ hypo→
env-22	BAAV01016969.1	-	686	523	hypo→ **RNA**→ hypo→
env-23	BAAX01018853.1	+	543	708	←DppD (COG0444) **RNA**→
env-24	BAAU01028694.1	+	160	335	hypo→ **RNA**→ hypo→
env-25	BAAZ01005512.1	-	486	314	hypo→ **RNA**→
Rob-1-1	NZ_AAVO02000012.1	+	105629	105797	hypo→ **RNA**→ MET17 (COG2873)→
Rob-1-2	NZ_AAVO02000013.1	-	48190	48022	IlvA (COG1171)→ **RNA**→ ArgE (COG0624)→
Eve-1-6	NZ_AAVL02000023.1	-	56132	55966	←HrcA (COG1420) **RNA**→ ←DegV (COG1307)
env-26	BABE01004765.1	-	1080	911	**RNA**→ DnaD (pfam04271)→
env-27	BABA01009489.1	+	308	486	GuaB (COG0516)→ **RNA**→ AdcA (cd01017)→
env-28	BAAZ01027624.1	-	869	703	**RNA**→ ←COG2070 (COG2070)
env-29	BAAV01006167.1	+	863	1032	PRA-CH (pfam01502)→ **RNA**→
env-30	BABD01002878.1	-	1413	1244	←hypo **RNA**→ ←nuc_hydro_3 (cd02653)
env-31	BAAY01020707.1	-	871	700	hypo→ **RNA**→ ←MurI (COG0796)
env-32	BABG01005618.1	+	356	523	←hypo **RNA**→ ←RimI (COG0456)
Eve-1-7	NZ_AAVL02000037.1	+	286875	287041	enolase (cd03313)→ **RNA**→ EriC (COG0038)→
env-33	BAAV01006120.1	+	617	787	←Nth (COG0177) **RNA**→ hypo→
env-34	BAAY01008255.1	-	580	413	←DMRL_synthase (pfam00885) **RNA**→ ←RibA (COG0807)
Eve-1-8	NZ_AAVL02000035.1	-	75876	75709	GreA (COG0782)→ **RNA**→ LysU (COG1190)→
Rla-1-1	NZ_ABOU02000027.1	-	54252	54086	←LeuS (COG0495) **RNA**→ ←COG0385 (COG0385)
env-35	BABG01000170.1	-	596	450	←hypo **RNA**→ hypo→
Dfo-1-1	NZ_AAXA02000013.1	+	250273	250440	←hypo **RNA**→ ←hypo
env-36	BABG01013486.1	+	1030	1197	PheS (COG0016)→ **RNA**→
Rla-1-2	NZ_ABOU02000020.1	-	5236	5069	←GGDEF (cd01949) **RNA**→ ←hypo
Rla-1-3	NZ_ABOU02000026.1	-	3951	3784	←HTH_XRE (cd00093) **RNA**→ hypo→
env-37	BABG01029695.1	-	208	75	GRS1 (COG0423)→ **RNA**→
Rla-1-4	NZ_ABOU02000048.1	+	174081	174247	LeuA (COG0119)HTH_ARAC (smart00342)→ **RNA**→ ←hypo
Rla-1-5	NZ_ABOU02000036.1	-	12225	12058	←hypo **RNA**→ ←hypo
Rla-1-6	NZ_ABOU02000034.1	-	13978	13811	←CLECT (cd00037) **RNA**→ ←hypo
Rla-1-7	NZ_ABOU02000050.1	-	2046	1878	←COG1725 (COG1725) **RNA**→ ←CstA (COG1966)
Rla-1-8	NZ_ABOU02000006.1	-	26217	26050	CSD (pfam00313)→ **RNA**→ ←hypo
Rla-1-9	NZ_ABOU02000050.1	-	140438	140271	←FtsX (pfam02687)ABC_MJ0796_Lo1CDE_FtsE (cd03255) **RNA**→ ←hypo
Rla-1-10	NZ_ABOU02000010.1	+	11853	12020	←PP_kinase (pfam02503) **RNA**→ ←COG4667 (COG4667)
Rla-1-11	NZ_ABOU02000050.1	+	30313	30480	←NptA (COG1283) **RNA**→ Relaxase (pfam03432)→
Rla-1-12	NZ_ABOU02000034.1	-	1573	1406	←LGT (pfam01790) **RNA**→ ←COG0012 (COG0012)

Rla-1-13	NZ_ABOU02000032.1	-	102267	102100	←hypo **RNA**→ ←ComFC (COG1040)
Rla-1-14	NZ_ABOU02000050.1	-	152044	151877	hypo→ **RNA**→ ←LysR (COG0583)
Rla-1-15	NZ_ABOU02000036.1	-	26948	26781	←GatC (COG0721) **RNA**→ ←hypo
Rla-1-16	NZ_ABOU02000033.1	+	1730	1897	←COG1032 (COG1032) **RNA**→ hypo→
Rla-1-17	NZ_ABOU02000034.1	+	55546	55713	hypo→ **RNA**→ ←IbpA (COG0071)
Rla-1-18	NZ_ABOU02000028.1	-	138359	138192	←hypo **RNA**→ COG1418 (COG1418)→
Rla-1-19	NZ_ABOU02000049.1	+	14048	14215	←hypo **RNA**→ ←Uup (COG0488)
Rla-1-20	NZ_ABOU02000055.1	+	90388	90555	←hypo **RNA**→ ←hypo
Rla-1-21	NZ_ABOU02000049.1	+	54819	54986	←GalM (COG2017) **RNA**→
Rla-1-22	NZ_ABOU02000035.1	-	71253	71086	COG3481 (COG3481)→ **RNA**→ TrmA (COG2265)→
Rla-1-23	NZ_ABOU02000053.1	+	15696	15863	PulA (COG1523)→ **RNA**→ ←DnaJ (COG0484)
Rla-1-24	NZ_ABOU02000046.1	+	32106	32273	CYK3 (COG5279)→ **RNA**→ CarA (COG0505)→
Rla-1-25	NZ_ABOU02000031.1	+	28256	28423	←ClpA (COG0542) **RNA**→ ←hypo
Rla-1-26	NZ_ABOU02000028.1	-	127880	127713	hypo→ **RNA**→ hypo→
Rla-1-27	NZ_ABOU02000049.1	+	30234	30401	←COG1873 (COG1873) **RNA**→ hypo→
Rla-1-28	NZ_ABOU02000017.1	+	92862	93029	←hypo **RNA**→ HTH_XRE (cd00093)COG0436 (COG0436)→
Rla-1-29	NZ_ABOU02000015.1	+	694	861	←GlpR (COG1349) **RNA**→ ←MscL (pfam01741)
Rla-1-30	NZ_ABOU02000046.1	-	647	480	←PTS_IIA_fru (cd00211)BglG (COG3711) **RNA**→ ←MscL (pfam01741)
Rla-1-31	NZ_ABOU02000037.1	-	8075	7919	COG1434 (COG1434)→ **RNA**→ ←ThiH (COG1060)
env-38	BAAU01002076.1	-	1922	1755	←Beta-lactamase (pfam00144) **RNA**→ ←hypo
env-39	BABG01002268.1	-	1271	1104	GatC (COG0721)→ **RNA**→ Amidase (pfam01425)→
env-40	BABG01000453.1	-	2218	2049	hypo→ **RNA**→ GATase1_GMP_Synthase (cd01742)GuaA (COG0519)→
env-41	BAAZ01002221.1	+	1348	1515	←Sialidase (cd00260) **RNA**→
Rla-1-32	NZ_ABOU02000047.1	-	15912	15745	←FN3 (cd00063)Sialidase (cd00260) **RNA**→ ←hypo
env-42	BAAU01024958.1	+	207	374	hypo→ **RNA**→ hypo→
env-43	BABG01002811.1	-	1092	925	UvrD (COG0210)→ **RNA**→ hypo→
env-44	BABG01000343.1	-	2063	1896	GGDEF (cd01949)→ **RNA**→ Prc (COG0793)→
env-45	BABG01000353.1	+	3907	4074	Iso_dh (pfam00180)→ **RNA**→ hypo→
env-46	BAAU01001385.1	+	1948	2124	PldB (COG2267)→ **RNA**→ PlsC (COG0204)→
Dfo-1-2	NZ_AAXA02000010.1	-	9926	9759	hypo→ **RNA**→ CysE (COG1045)→
env-47	BABG01021235.1	-	944	778	**RNA**→ ←hypo
Rla-1-33	NZ_ABOU02000007.1	+	34416	34583	←AdeC (COG1001) **RNA**→ ←hypo
Rla-1-34	NZ_ABOU02000026.1	+	37629	37796	SDH_alpha (pfam03313)→ **RNA**→ ←hypo
Rla-1-35	NZ_ABOU02000035.1	+	2399	2566	←AcnA (COG1048) **RNA**→ ←hypo
Rla-1-36	NZ_ABOU02000028.1	+	3639	3806	←hypo **RNA**→ ←hypo
Rla-1-37	NZ_ABOU02000023.1	+	2814	2981	HTH_XRE (cd00093)→ **RNA**→ ←hypo
Rla-1-38	NZ_ABOU02000039.1	-	74302	74135	←DUF177 (pfam02620) **RNA**→ ←hypo
Rla-1-39	NZ_ABOU02000036.1	-	901	734	←ARA1 (COG0656) **RNA**→ ←hypo
Rla-1-40	NZ_ABOU02000032.1	-	95004	94837	←UvrA (COG0178) **RNA**→ ←hypo
Rla-1-41	NZ_ABOU02000048.1	-	41861	41694	←HGD-D (pfam06050) **RNA**→ ←hypo
Rla-1-42	NZ_ABOU02000029.1	-	8103	7936	COG3335 (COG3335)COG3415 (COG3415)→ **RNA**→ ←hypo
Rla-1-43	NZ_ABOU02000029.1	-	251	83	←FeoA (pfam04023) **RNA**→ ←hypo
Rla-1-44	NZ_ABOU02000032.1	-	41539	41372	←hypo **RNA**→ ←hypo
Rla-1-45	NZ_ABOU02000042.1	+	30676	30843	hypo→ **RNA**→ ←hypo
Rla-1-46	NZ_ABOU02000052.1	-	12141	11974	Sms (COG1066)→ **RNA**→ hypo→
Rla-1-47	NZ_ABOU02000051.1	-	49001	48834	←Cof (COG0561) **RNA**→ ←hypo
Rla-1-48	NZ_ABOU02000020.1	-	7403	7236	←EutG (COG1454) **RNA**→ ←hypo
Rla-1-49	NZ_ABOU02000034.1	-	8847	8680	←hypo **RNA**→ ←hypo
Rla-1-50	NZ_ABOU02000045.1	-	33335	33167	←GreA (COG0782) **RNA**→ ←hypo
Rla-1-51	NZ_ABOU02000035.1	+	23673	23840	←Cyt_C5_DNA_methylase (cd00315) **RNA**→ ←hypo
Rla-1-52	NZ_ABOU02000028.1	+	88980	89147	←hypo **RNA**→ ←hypo
Rla-1-53	NZ_ABOU02000053.1	+	313	480	**RNA**→ ←hypo

Rla-1-54	NZ_ABOU02000026.1	-	49611	49444	←GlgP (COG0058) **RNA**→ ←hypo
env-48	BABG01006456.1	+	435	583	hypo→ **RNA**→ ←hypo
Rla-1-55	NZ_ABOU02000037.1	-	173	6	←hypo **RNA**→
Rla-1-56	NZ_ABOU02000047.1	-	5367	5200	←PAP2_Aur1_like (cd03386) **RNA**→ ←hypo
Rla-1-57	NZ_ABOU02000056.1	+	160	327	**RNA**→ ←hypo
Rla-1-58	NZ_ABOU02000047.1	-	13649	13478	←hypo **RNA**→ ←hypo
env-49	BABG01003277.1	+	999	1150	LysR_substrate (pfam03466)→ **RNA**→ COG1561 (COG1561)→
env-50	BABA01019112.1	-	596	425	**RNA**→ hypo→
env-51	BAAY01010317.1	+	1179	1351	DctQ (COG1593)→ **RNA**→
env-52	BAAZ01003651.1	-	1125	947	NorM (COG0534)→ **RNA**→ ←Pro tRNA
env-53	hgutS8_s8_162567	-	687	514	**RNA**→ ←hypo
env-54	BAAU01002871.1	-	989	818	←hypo **RNA**→ MarR (pfam01047)→
env-55	BABG01026309.1	-	776	604	hypo→ **RNA**→ **RNA**→ hypo→
env-95	BABG01026309.1	-	567	389	(shown above)
env-56	BAAU01018938.1	-	559	388	LysR_substrate (pfam03466)→ **RNA**→ PseudoU_synth_RluCD_like (cd02869)→
env-57	BABG01019443.1	-	442	271	**RNA**→ hypo→
env-58	BABG01006738.1	-	1128	957	ManC (COG0662)→ **RNA**→ OYE_like_FMN_family (cd02803)→
env-59	BABB01006594.1	-	1215	1043	←hypo **RNA**→ ArcC (COG0549)→
env-60	BAAU01010790.1	-	206	35	←AraC_binding (pfam02311) **RNA**→
env-61	BAAX01004650.1	+	1582	1753	hypo→ **RNA**→ ManX (COG2893)→
env-62	BABB01010898.1	-	183	12	hypo→ **RNA**→
env-63	BAAX01029733.1	-	904	738	hypo→ **RNA**→ hypo→
env-64	BAAX01029732.1	+	509	679	FAD-oxidase_C (pfam02913)→ **RNA**→
env-65	BABG01034003.1	+	219	392	**RNA**→
env-66	BABA01033985.1	-	198	26	**RNA**→ hypo→
env-67	BAAU01012045.1	+	8	179	←hypo **RNA**→ ←COG5263 (COG5263)
env-68	BAAX01011900.1	-	516	344	←pyridoxal_pyridoxamine_kinase (cd01173) **RNA**→ hypo→
env-69	BABA01008401.1	-	371	199	hypo→ **RNA**→ ←DbpA (pfam03880)
env-70	BAAV01014729.1	+	399	571	hypo→ **RNA**→ hypo→
env-71	BABG01028078.1	+	59	230	←hypo **RNA**→ ←FepC (COG1120)
env-72	BAAU01021894.1	-	499	327	**RNA**→ ←hypo
env-73	BAAU01019454.1	+	338	511	ErfK_YbiS_YhnG (pfam03734)→ **RNA**→ hypo→
env-74	BAAU01006425.1	-	310	139	HypD (pfam01924)→ **RNA**→
env-75	BABG01019422.1	+	349	526	←hypo **RNA**→ hypo→
Aca-1-1	NZ_ABAX03000010.1	-	68909	68756	hypo→ **RNA**→ ←hypo
env-76	BAAU01004259.1	-	2023	1862	**RNA**→ RplI (COG0359)→
env-77	BABG01022086.1	-	503	334	**RNA**→
env-78	BABG01022086.1	+	432	604	**RNA**→
env-79	BABG01022086.1	-	338	167	**RNA**→
env-80	BABG01022086.1	+	599	773	**RNA**→
env-81	hgutS8_s8_175420	+	669	843	FepC (COG1120)→ **RNA**→
env-82	BAAU01025016.1	-	257	80	hypo→ **RNA**→
env-83	BAAU01001327.1	+	473	650	←hypo **RNA**→ MelB (COG2211)→
env-84	BABB01016599.1	+	587	763	hypo→ **RNA**→ hypo→
env-85	BABD01030819.1	+	530	707	**RNA**→
env-86	BABA01021654.1	+	738	917	**RNA**→
env-87	BAAU01001711.1	-	3615	3436	**RNA**→ ←hypo
env-88	BAAU01018515.1	+	135	314	**RNA**→
env-89	BABA01021655.1	-	915	740	hypo→ **RNA**→ hypo→
env-90	BABG01001408.1	+	592	769	←hypo **RNA**→ ←hypo
env-91	hgutS8_s8_159931_1	-	2869	2692	←Hit (COG0537) **RNA**→ ←RecR (COG0353)
Csp-2-1	NZ_ABGC03000035.1	+	101575	101752	←Hit (COG0537) **RNA**→ ←RecR (COG0353)

env-92	BABG01003064.1	+	575	752	←HINT_subgroup (cd01277) **RNA**→ ←RecR (COG0353)
env-93	BAAV01013666.1	+	622	801	hypo→ **RNA**→ ←hypo
env-94	BAAX01013747.1	+	612	800	←hypo **RNA**→ hypo→
env-96	BABG01009698.1	+	320	498	←hypo **RNA**→ ←CorC (COG4535)
env-97	BAAU01018414.1	+	172	350	hypo→ **RNA**→ ←hypo
env-98	BABD01009573.1	-	298	120	COG2607 (COG2607)→ **RNA**→
env-99	BABG01004235.1	-	1830	1652	COG2607 (COG2607)→ **RNA**→ ModE (COG2005)→
env-100	hgutS8_s8_158962	+	686	864	←ABC-type sugar transport system, periplasmic comp onent **RNA**→ ←HTH_ARAC (smart00342)
env-101	BAAU01019504.1	+	336	499	hypo→ **RNA**→ hypo→
env-102	BABG01032637.1	-	818	657	**RNA**→ GD_AH_N (pfam04292)→
Ceu-1-2	NZ_ABEY02000008.1	-	42145	41983	OmpR (COG0745)→ **RNA**→ RluA (COG0564)→
env-103	BABB01014983.1	-	514	357	**RNA**→ hypo→
env-104	BAAX01013085.1	-	392	234	←hypo **RNA**→ hypo→
Ceu-1-3	NZ_ABEY02000023.1	-	60158	59997	←hypo **RNA**→ ←tRNA_bind_EcMetRS_like (cd02800)
env-105	hgutS8_s8_159061	+	771	932	←hypo **RNA**→ ←tRNA_bind_EcMetRS_like (cd02800)
env-106	BABG01015553.1	+	646	811	hypo→ **RNA**→ hypo→
env-107	BABA01025769.1	+	311	472	hypo→ **RNA**→ ←hypo
env-108	BAAZ01006092.1	+	658	819	GreA (COG0782)→ **RNA**→ hypo→
env-109	BABA01017275.1	+	327	488	←hypo **RNA**→
env-110	BABG01006513.1	-	1158	997	←YBL036c_PLPDEIII (cd00635) **RNA**→ ←hypo
env-111	BABG01006599.1	-	897	736	**RNA**→ ←hypo
env-112	BAAU01013082.1	-	887	727	GreA (COG0782)→ **RNA**→ ←hypo
env-113	BAAU01013081.1	+	474	635	COG1959 (COG1959)→ **RNA**→ ←hypo
env-114	BABG01034775.1	+	591	754	hypo→ **RNA**→
env-115	BAAU01019179.1	+	574	738	hypo→ **RNA**→ ←hypo
env-116	BAAU01026019.1	-	748	585	**RNA**→ ←HELICc2 (smart00491)
env-117	BABG01018910.1	+	823	983	SAICAR_synt_Sc (cd01414)→ **RNA**→
env-118	BAAU01018222.1	-	959	800	hypo→ **RNA**→ DnrI (COG3629)→
env-119	BAAU01026020.1	+	410	570	hypo→ **RNA**→
Dfo-1-3	NZ_AAXA02000006.1	+	32226	32385	hypo→ **RNA**→ hypo→
env-120	BABG01008432.1	+	584	749	Acm (COG3757)→ **RNA**→
env-121	BABG01026909.1	-	932	754	hypo→ **RNA**→ ←SASP (pfam00269)
Ceu-1-4	NZ_ABEY02000023.1	-	72091	71928	←LacZ (COG3250)Bgal_small_N (pfam02929)Bgal_small_C (pfam02930) **RNA**→ AraC_binding (pfam02311)HTH_ARAC (smart003
env-122	BAAZ01010534.1	+	457	622	COG1959 (COG1959)→ **RNA**→ hypo→
Csp-1-1	NZ_AAYW02000007.1	+	152558	152719	←tRNA-Pro **RNA**→ hypo→
env-123	BABG01021957.1	+	675	835	←hypo **RNA**→
env-124	BABG01021956.1	-	564	404	**RNA**→ hypo→
env-125	BABA01014690.1	+	426	585	←hypo **RNA**→ OmpR (COG0745)→
env-126	BABG01032678.1	-	440	280	hypo→ **RNA**→ hypo→
env-127	BABG01005904.1	+	388	548	hypo→ **RNA**→ hypo→
env-128	BAAZ01018676.1	-	282	125	hypo→ **RNA**→ hypo→

21.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 21.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Clostridiales-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

u cd00037 (1) C-type lectin (CTL) and CTL-like domains; Many of these domains function as calcium-dependent carbohydrate binding modules and are known as carbohydrate-recognition domains (CRDs).
u cd00063 (1) Fibronectin type 3 domain; One of three types of internal repeats found in the plasma protein fibronectin.
uD cd00093 (4) Helix-turn-helix XRE-family like proteins.
u cd00211 (1) PTS_IIA, PTS system, fructose/mannitol specific IIA subunit.
u cd00260 (2) Sialidases or neuraminidases function to bind and hydrolyze terminal sialic acid residues from various glycoconjugates as well as playing roles in pathogenesis, bacterial nutrition and cellular interactions. [X]
u cd00315 (1) Cytosine-C5 specific DNA methylases; Methyl transfer reactions play an important role in many aspects of biology.
D cd00565 (1) ThiS (ThiaminS) is a sulfur carrier protein involved in thiamin biosynthesis in bacteria.
u cd00635 (1) PLP dependent enzymes class III (PLPDE_III).
D cd01017 (1) Metal binding protein AcdA.
D cd01158 (1) Short chain acyl-CoA dehydrogenase (SCAD).
u cd01173 (1) Pyridoxal kinase plays a key role in the synthesis of the active coenzyme pyridoxal-5'-phosphate (PLP), by catalyzing the phosphorylation of the precursor vitamin B6 in the presence of Zn2+ and ATP.
u cd01277 (1) HINT (histidine triad nucleotide-binding protein) subgroup: Members of this CD belong to the superfamily of histidine triad hydrolases that act on alpha-phosphate of ribonucleotides.
u cd01414 (1) Eukaryotic, prokaryotic and archaeal group of SAICAR synthetases represented by the Saccharomyces cerevisiae (Sc) SAICAR synthetase.
D cd01742 (1) Type 1 glutamine amidotransferase (GATase1) domain found in GMP synthetase.
d cd01948 (1) EAL domain.
u cd01949 (2) Diguanylate-cyclase (DGC) or GGDEF domain: Originally named after a conserved residue pattern, and initially described as domain of unknown function 1 (DUF1).
d cd02653 (1) NH_3: A subgroup of nucleoside hydrolases.
d cd02800 (2) tRNA-binding-domain-containing Escherichia coli methionyl-tRNA synthetase (EcMetRS)-like proteins.
D cd02803 (1) Old yellow enzyme (OYE)-like FMN binding domain.
D cd02869 (1) PseudoU_synth_RsuA/RluD: Pseudouridine synthase, RsuA/RluD family.
d cd03251 (1) MsbA is an essential ABC transporter, closely related to eukaryotic MDR proteins.
u cd03255 (1) This family is comprised of MJ0796 ATP-binding cassette, macrolide-specific ABC-type efflux carrier (MacAB), and proteins involved in cell division (FtsE) and release of liporoteins from the cytoplasmic membrane (LolCDE).
u cd03313 (1) Enolase: Enolases are homodimeric enzymes that catalyse the reversible dehydration of 2-phospho-D-glycerate to phosphoenolpyruvate as part of the glycolytic and gluconeogenesis pathways.
u cd03344 (1) GroEL_like type I chaperonin.
u cd03386 (1) PAP2_like proteins, Aur1_like subfamily.
D COG0002 (1) Acetylglutamate semialdehyde dehydrogenase [Amino acid transport and metabolism]
d COG0012 (1) Predicted GTPase, probable translation factor [Translation, ribosomal structure and biogenesis]
u COG0016 (1) Phenylalanyl-tRNA synthetase alpha subunit [Translation, ribosomal structure and biogenesis]
D COG0038 (1) Chloride channel protein EriC [Inorganic ion transport and metabolism]
u COG0047 (1) Phosphoribosylformylglycinamidine (FGAM) synthase, glutamine amidotransferase domain [Nucleotide transport and metabolism]
u COG0058 (1) Glucan phosphorylase [Carbohydrate transport and metabolism]
d COG0071 (1) Molecular chaperone (small heat shock protein) [Posttranslational modification, protein turnover, chaperones]
u COG0079 (1) Histidinol-phosphate/aromatic aminotransferase and cobyric acid decarboxylase [Amino acid transport and metabolism]
u COG0119 (1) Isopropylmalate/homocitrate/citramalate synthases [Amino acid transport and metabolism]
u COG0177 (1) Predicted EndoIII-related endonuclease [DNA replication, recombination, and repair]
u COG0178 (1) Excinuclease ATPase subunit [DNA replication, recombination, and repair]
D COG0204 (1) 1-acyl-sn-glycerol-3-phosphate acyltransferase [Lipid metabolism]
u COG0210 (1) Superfamily I DNA and RNA helicases [DNA replication, recombination, and repair]
u COG0296 (1) 1,4-alpha-glucan branching enzyme [Carbohydrate transport and metabolism]
d COG0297 (2) Glycogen synthase [Carbohydrate transport and metabolism]
d COG0353 (3) Recombinational DNA repair protein (RecF pathway) [DNA replication, recombination, and repair]
D COG0359 (1) Ribosomal protein L9 [Translation, ribosomal structure and biogenesis]
d COG0385 (1) Predicted Na+-dependent transporter [General function prediction only]
u COG0423 (1) Glycyl-tRNA synthetase (class II) [Translation, ribosomal structure and biogenesis]
D COG0436 (1) Aspartate/tyrosine/aromatic aminotransferase [Amino acid transport and metabolism]
u COG0444 (1) ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component [Amino acid transport and metabolism / Inorganic ion transport and metabolism]
d COG0456 (1) Acetyltransferases [General function prediction only]
d COG0484 (1) DnaJ-class molecular chaperone with C-terminal Zn finger domain [Posttranslational modification, protein turnover, chaperones]
d COG0488 (1) ATPase components of ABC transporters with duplicated ATPase domains [General function prediction only]
u COG0495 (1) Leucyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
D COG0505 (1) Carbamoylphosphate synthase small subunit [Amino acid transport and metabolism / Nucleotide transport and metabolism]
u COG0516 (1) IMP dehydrogenase/GMP reductase [Nucleotide transport and metabolism]
D COG0519 (1) GMP synthase, PP-ATPase domain/subunit [Nucleotide transport and metabolism]
u COG0534 (1) Na+-driven multidrug efflux pump [Defense mechanisms]
u COG0537 (2) Diadenosine tetraphosphate (Ap4A) hydrolase and other HIT family hydrolases [Nucleotide transport and metabolism / Carbohydrate transport and metabolism / General function prediction only]
u COG0542 (1) ATPases with chaperone activity, ATP-binding subunit [Posttranslational modification, protein turnover, chaperones]
D COG0549 (1) Carbamate kinase [Amino acid transport and metabolism]
D COG0550 (1) Topoisomerase IA [DNA replication, recombination, and repair]

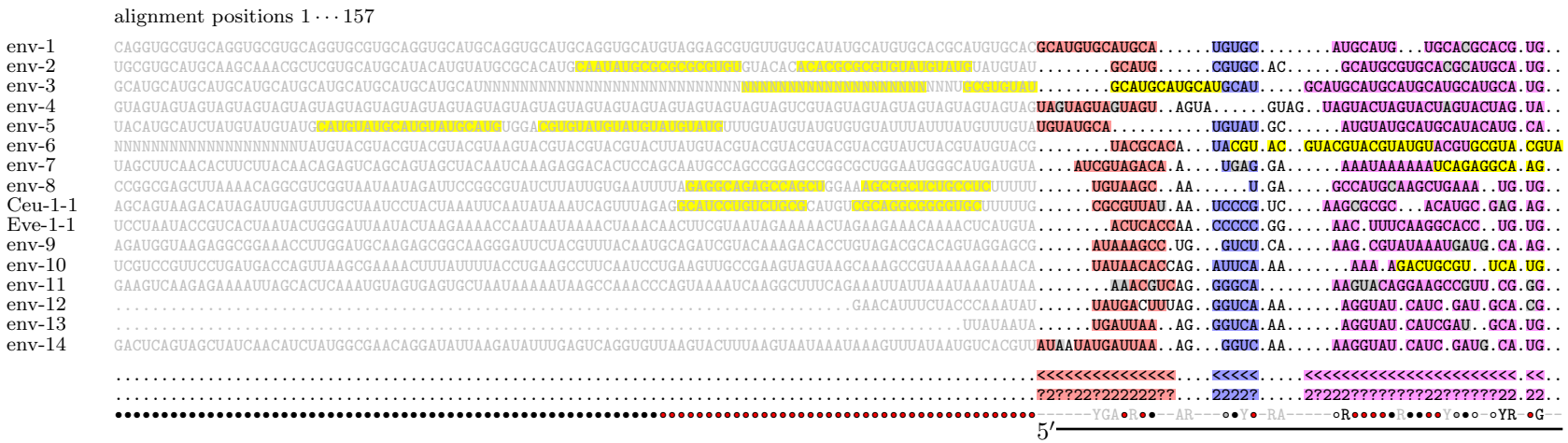
u COG0561 (1) Predicted hydrolases of the HAD superfamily [General function prediction only]
D COG0564 (1) Pseudouridylate synthases, 23S RNA-specific [Translation, ribosomal structure and biogenesis]
d COG0583 (1) Transcriptional regulator [Transcription]
D COG0624 (1) Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase and related deacylases [Amino acid transport and metabolism]
u COG0656 (1) Aldo/keto reductases, related to diketogulonate reductase [General function prediction only]
u COG0662 (1) Mannose-6-phosphate isomerase [Carbohydrate transport and metabolism]
u COG0721 (2) Asp-tRNAAsn/Glu-tRNAGln amidotransferase C subunit [Translation, ribosomal structure and biogenesis]
uD COG0745 (2) Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain [Signal transduction mechanisms / Transcription]
u COG0782 (4) Transcription elongation factor [Transcription]
D COG0793 (1) Periplasmic protease [Cell envelope biogenesis, outer membrane]
d COG0796 (1) Glutamate racemase [Cell envelope biogenesis, outer membrane]
d COG0807 (1) GTP cyclohydrolase II [Coenzyme metabolism]
u COG1001 (1) Adenine deaminase [Nucleotide transport and metabolism]
u COG1032 (1) Fe-S oxidoreductase [Energy production and conversion]
d COG1040 (1) Predicted amidophosphoribosyltransferases [General function prediction only]
D COG1045 (1) Serine acetyltransferase [Amino acid transport and metabolism]
u COG1048 (1) Aconitase A [Energy production and conversion]
d COG1060 (1) Thiamine biosynthesis enzyme ThiH and related uncharacterized enzymes [Coenzyme metabolism / General function prediction only]
u COG1066 (1) Predicted ATP-dependent serine protease [Posttranslational modification, protein turnover, chaperones]
u d COG1120 (2) ABC-type cobalamin/Fe3+-siderophores transport systems, ATPase components [Inorganic ion transport and metabolism / Coenzyme metabolism]
u COG1171 (1) Threonine dehydratase [Amino acid transport and metabolism]
D COG1190 (1) Lysyl-tRNA synthetase (class II) [Translation, ribosomal structure and biogenesis]
u COG1283 (1) Na+/phosphate symporter [Inorganic ion transport and metabolism]
Dd COG1307 (2) Uncharacterized protein conserved in bacteria [Function unknown]
u COG1349 (1) Transcriptional regulators of sugar metabolism [Transcription / Carbohydrate transport and metabolism]
D COG1387 (1) Histidinol phosphatase and related hydrolases of the PHP family [Amino acid transport and metabolism / General function prediction only]
D COG1418 (1) Predicted HD superfamily hydrolase [General function prediction only]
u COG1420 (1) Transcriptional regulator of heat shock gene [Transcription]
u COG1434 (1) Uncharacterized conserved protein [Function unknown]
u COG1454 (1) Alcohol dehydrogenase, class IV [Energy production and conversion]
u COG1523 (1) Type II secretory pathway, pullulanase PulA and related glycosidases [Carbohydrate transport and metabolism]
D COG1561 (1) Uncharacterized stress-induced protein [Function unknown]
u COG1593 (1) TRAP-type C4-dicarboxylate transport system, large permease component [Carbohydrate transport and metabolism]
u COG1725 (1) Predicted transcriptional regulators [Transcription]
u COG1873 (1) Uncharacterized conserved protein [Function unknown]
u COG1917 (1) Uncharacterized conserved protein, contains double-stranded beta-helix domain [Function unknown]
u COG1959 (2) Predicted transcriptional regulator [Transcription]
d COG1966 (1) Carbon starvation protein, predicted membrane protein [Signal transduction mechanisms]
D COG2005 (1) N-terminal domain of molybdenum-binding protein [General function prediction only]
u COG2017 (1) Galactose mutarotase and related enzymes [Carbohydrate transport and metabolism]
d COG2070 (1) Dioxygenases related to 2-nitropropane dioxygenase [General function prediction only]
D COG2211 (1) Na+/melibiose symporter and related transporters [Carbohydrate transport and metabolism]
D COG2265 (1) SAM-dependent methyltransferases related to tRNA (uracil-5-)-methyltransferase [Translation, ribosomal structure and biogenesis]
u COG2267 (1) Lysophospholipase [Lipid metabolism]
u COG2607 (2) Predicted ATPase (AAA+ superfamily) [General function prediction only]
D COG2873 (1) O-acetylhomoserine sulfhydrylase [Amino acid transport and metabolism]
D COG2893 (1) Phosphotransferase system, mannose/fructose-specific component IIA [Carbohydrate transport and metabolism]
u COG3250 (1) Beta-galactosidase/beta-glucuronidase [Carbohydrate transport and metabolism]
u COG3335 (1) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
u COG3415 (1) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
u COG3481 (1) Predicted HD-superfamily hydrolase [General function prediction only]
D COG3629 (1) DNA-binding transcriptional activator of the SARP family [Signal transduction mechanisms]
u COG3711 (1) Transcriptional antiterminator [Transcription]
u COG3757 (1) Lyzozyme M1 (1,4-beta-N-acetylmuramidase) [Cell envelope biogenesis, outer membrane]
u COG3968 (1) Uncharacterized protein related to glutamine synthetase [General function prediction only]
d COG4535 (1) Putative Mg2+ and Co2+ transporter CorC [Inorganic ion transport and metabolism]
d COG4667 (1) Predicted esterase of the alpha-beta hydrolase superfamily [General function prediction only]
d COG5263 (1) FOG: Glucan-binding domain (YG repeat) [General function prediction only]
u COG5279 (1) Uncharacterized protein involved in cytokinesis, contains TGc (transglutaminase/protease-like) domain [Cell division and chromosome partitioning]
u pfam00144 (1) Beta-lactamase.
u pfam00180 (1) Isocitrate/isopropylmalate dehydrogenase.
d pfam00269 (1) Small, acid-soluble spore proteins, alpha/beta type.
u pfam00313 (1) 'Cold-shock' DNA-binding domain.
u pfam00725 (1) 3-hydroxyacyl-CoA dehydrogenase, C-terminal domain.
u pfam00885 (1) 6,7-dimethyl-8-ribityllumazine synthase.
D pfam00912 (1) Transglycosylase.
D pfam01047 (2) MarR family.
u pfam01205 (1) Uncharacterized protein family UPF0029.

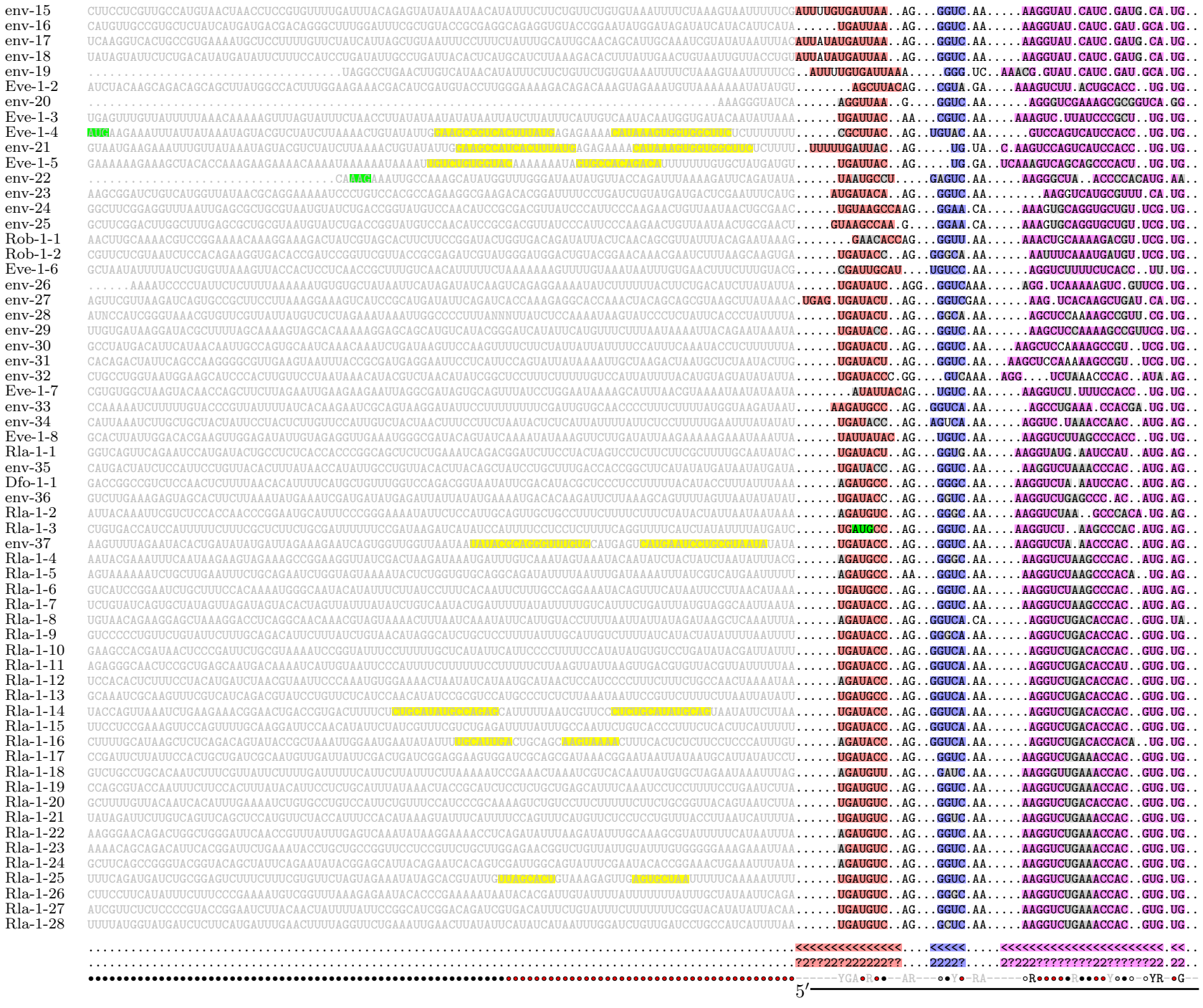
D pfam01425 (1) Amidase.
u pfam01502 (1) Phosphoribosyl-AMP cyclohydrolase.
d pfam01741 (2) Large-conductance mechanosensitive channel, MscL.
u pfam01790 (1) Prolipoprotein diacylglyceryl transferase.
u pfam01924 (1) Hydrogenase formation hypA family.
uD pfam02311 (2) AraC-like ligand binding domain.
u pfam02503 (1) Polyphosphate kinase.
u pfam02620 (1) Uncharacterized ACR, COG1399.
u pfam02687 (1) Predicted permease.
u pfam02913 (1) FAD linked oxidases, C-terminal domain.
u pfam02929 (1) Beta galactosidase small chain, N terminal domain.
u pfam02930 (1) Beta galactosidase small chain, C terminal domain.
u pfam03313 (1) Serine dehydratase alpha chain.
D pfam03432 (1) Relaxase/Mobilisation nuclease domain.
u pfam03466 (2) LysR substrate binding domain.
u pfam03734 (1) ErfK/YbiS/YcfS/YnhG.
d pfam03880 (1) DbpA RNA binding domain.
u pfam04023 (1) FeoA domain.
D pfam04271 (1) DnaD-like domain.
D pfam04292 (1) D-galactarate dehydratase / Altronate hydrolase, N terminus.
u pfam06050 (1) 2-hydroxyglutaryl-CoA dehydratase, D-component. [4Fe-4S]
uDd smart00342 (3) helix_turn_helix, arabinose operon control protein;
d smart00491 (1) helicase superfamily c-terminal domain;

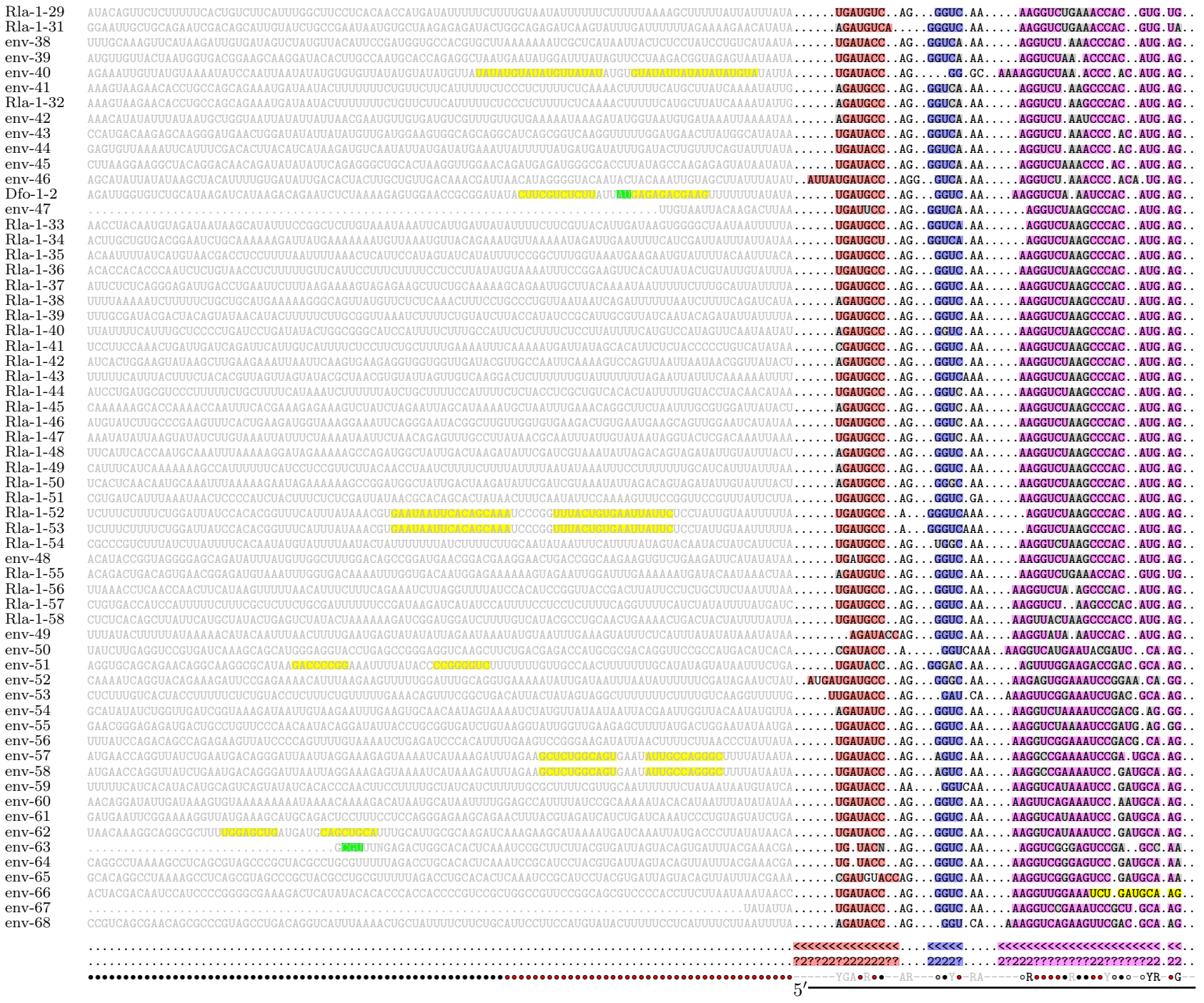
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Rla-1-30

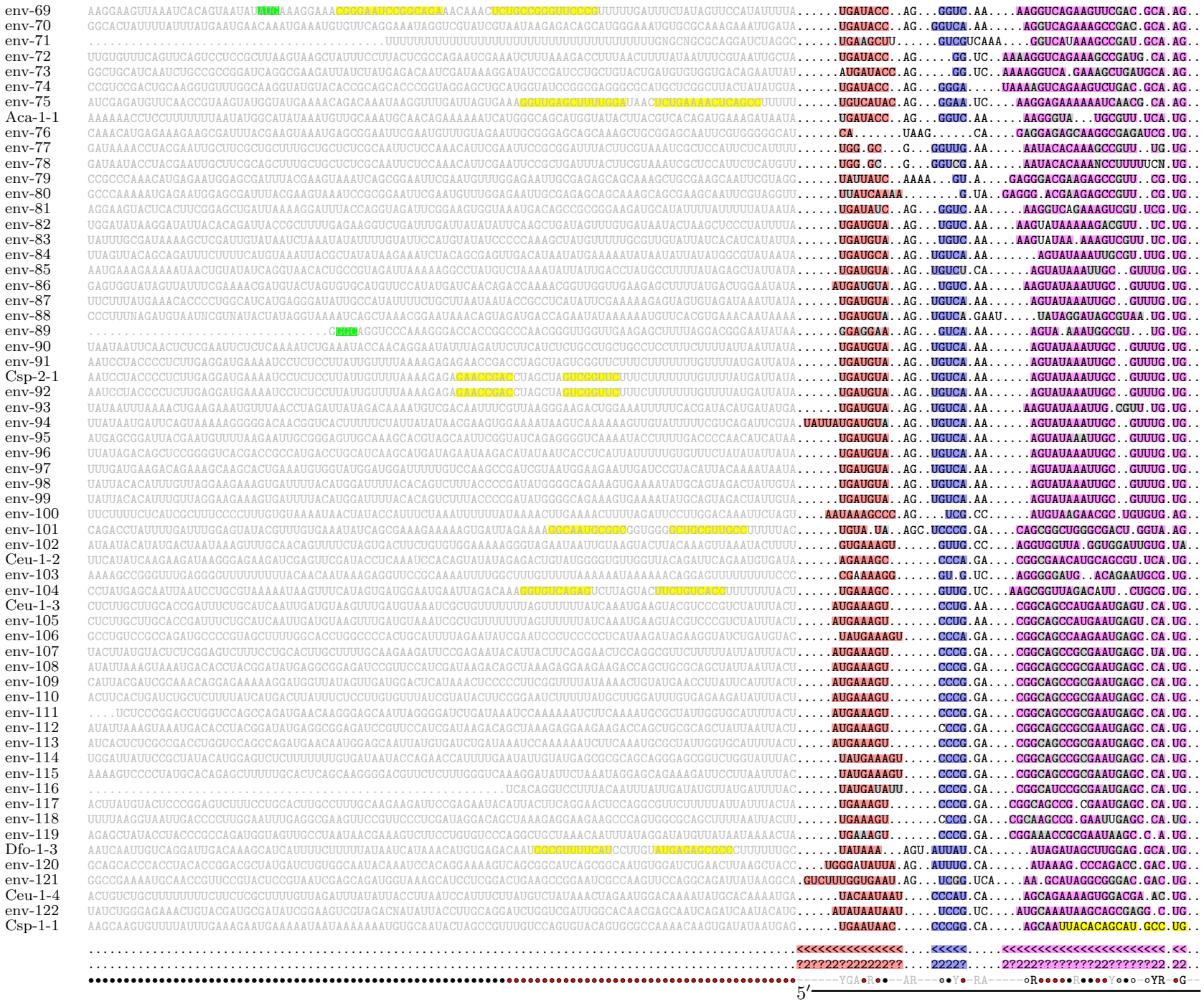
21.4 Multiple-sequence alignment

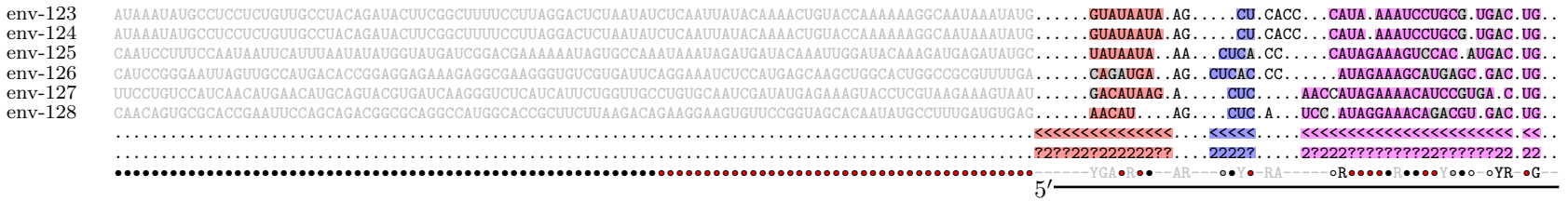
Each Clostridiales-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 21.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Clostridiales-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



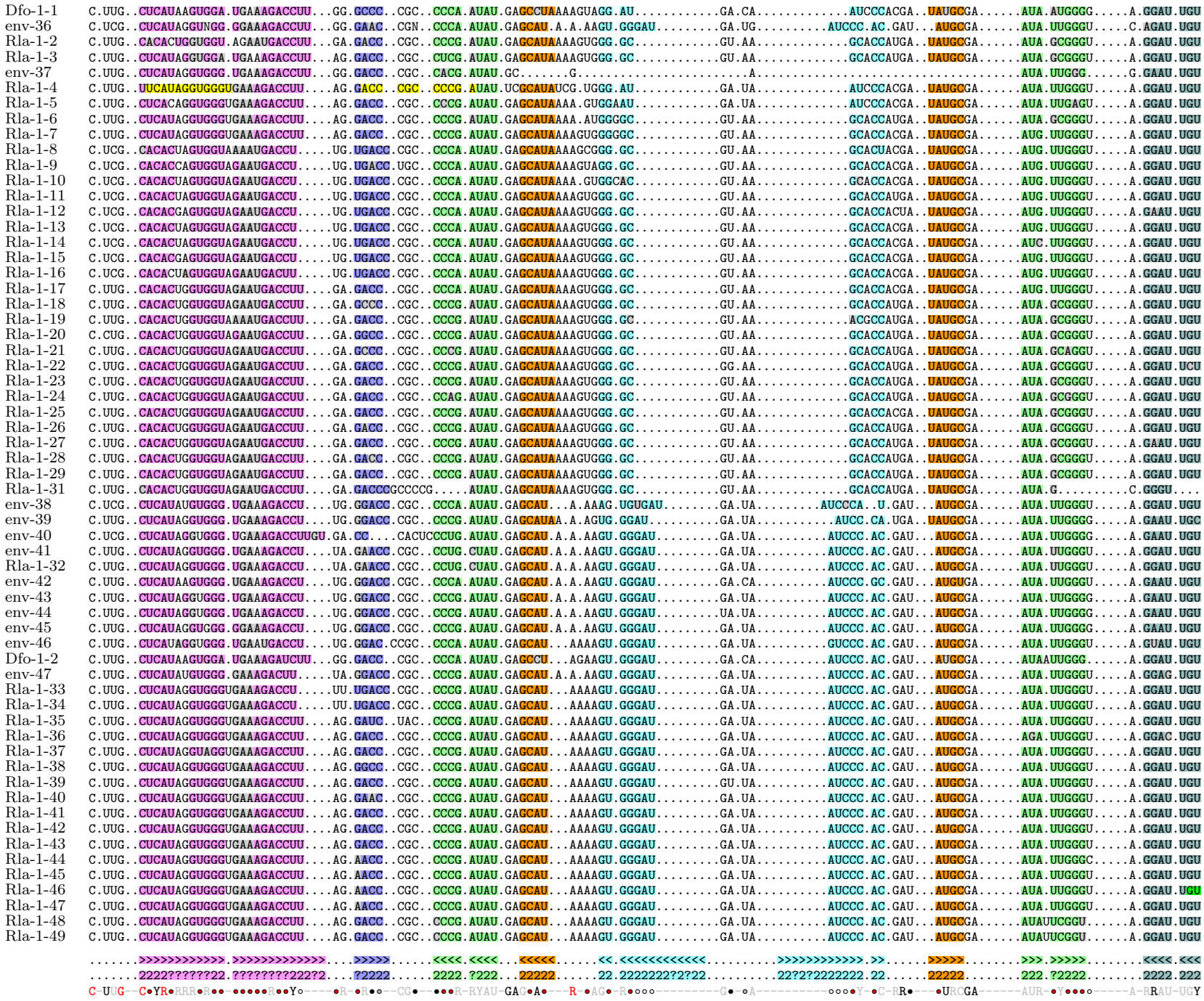

env-15
env-16
env-17
env-18
env-19
Eve-1-2
env-20
Eve-1-3
Eve-1-4
env-21
Eve-1-5
env-22
env-23
env-24
env-25
Rob-1-1
Rob-1-2
Eve-1-6
env-26
env-27
env-28
env-29
env-30
env-31
env-32
Eve-1-7
env-33
env-34
Eve-1-8
Rla-1-1
env-35
Dfo-1-1
env-36
Rla-1-2
Rla-1-3
env-37
Rla-1-4
Rla-1-5
Rla-1-6
Rla-1-7
Rla-1-8
Rla-1-9
Rla-1-10
Rla-1-11
Rla-1-12
Rla-1-13
Rla-1-14
Rla-1-15
Rla-1-16
Rla-1-17
Rla-1-18
Rla-1-19
Rla-1-20
Rla-1-21
Rla-1-22
Rla-1-23
Rla-1-24
Rla-1-25
Rla-1-26
Rla-1-27
Rla-1-28
5′

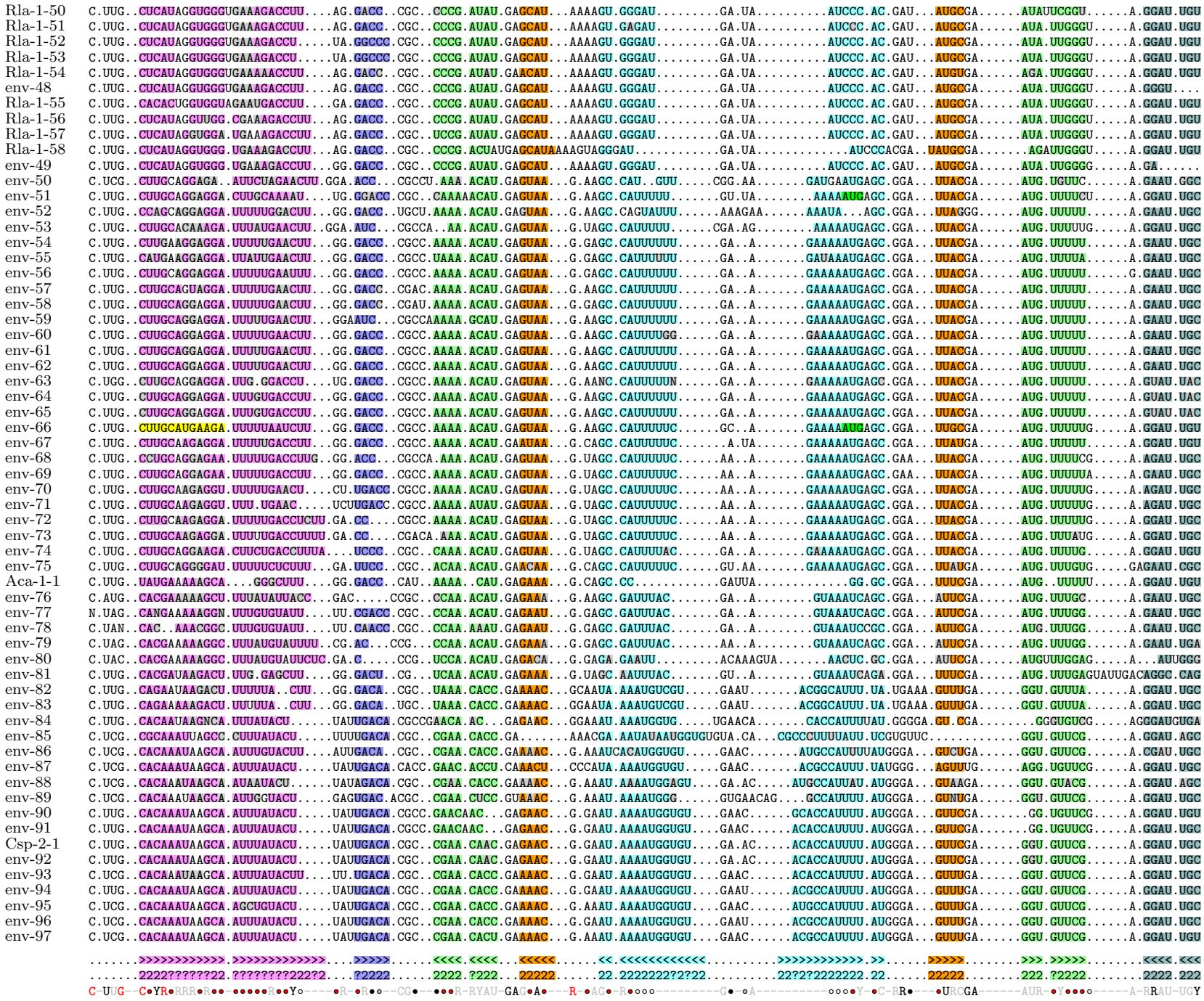

Rla-1-29
Rla-1-31
env-38
env-39
env-40
env-41
Rla-1-32
env-42
env-43
env-44
env-45
env-46
Dfo-1-2
env-47
Rla-1-33
Rla-1-34
Rla-1-35
Rla-1-36
Rla-1-37
Rla-1-38
Rla-1-39
Rla-1-40
Rla-1-41
Rla-1-42
Rla-1-43
Rla-1-44
Rla-1-45
Rla-1-46
Rla-1-47
Rla-1-48
Rla-1-49
Rla-1-50
Rla-1-51
Rla-1-52
Rla-1-53
Rla-1-54
env-48
Rla-1-55
Rla-1-56
Rla-1-57
Rla-1-58
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
5′

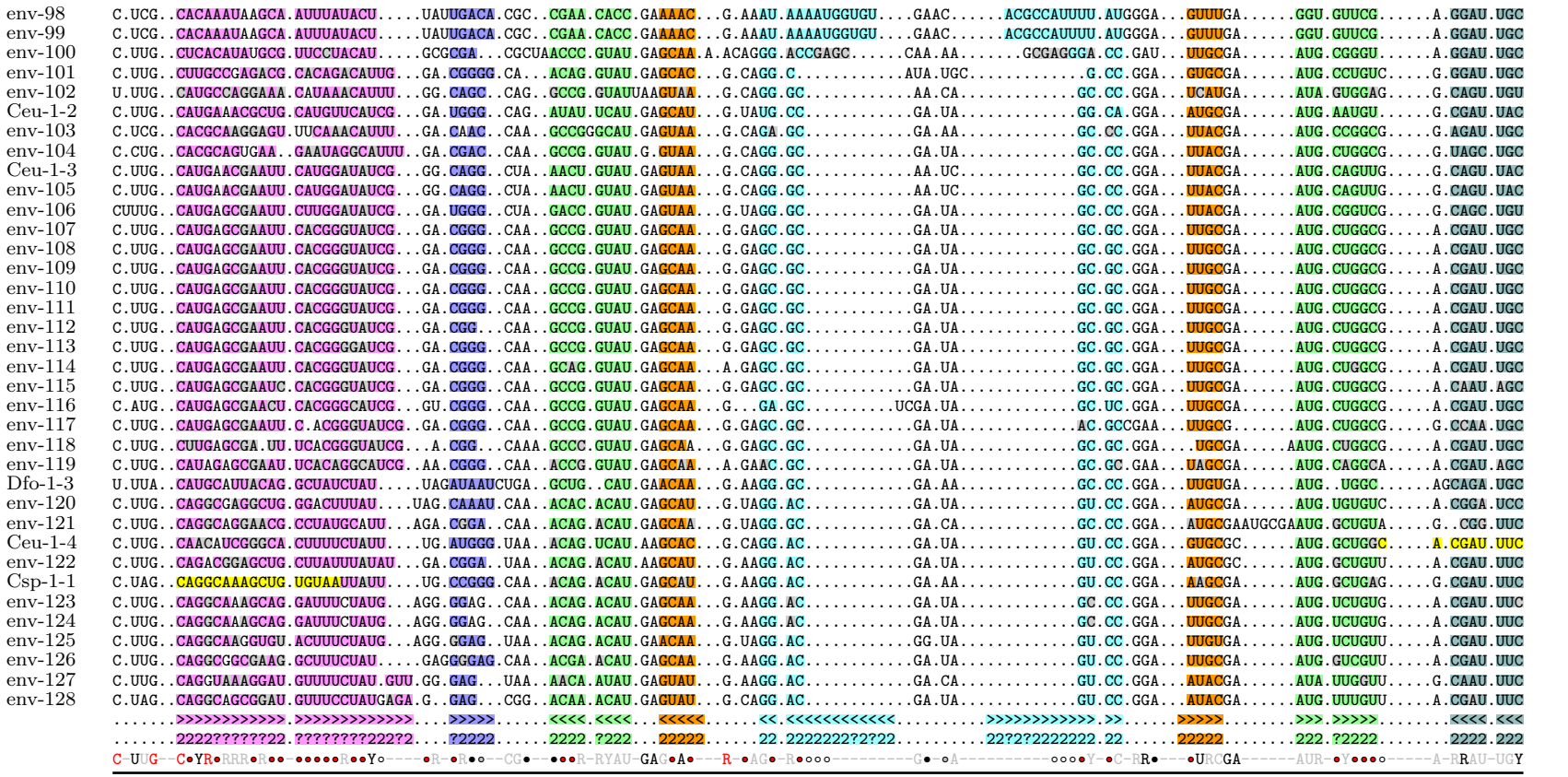

env-69
env-70
env-71
env-72
env-73
env-74
env-75
Aca-1-1
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
Csp-2-1
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
Ceu-1-2
env-103
env-104
Ceu-1-3
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
Dfo-1-3
env-120
env-121
Ceu-1-4
env-122
Csp-1-1
5′

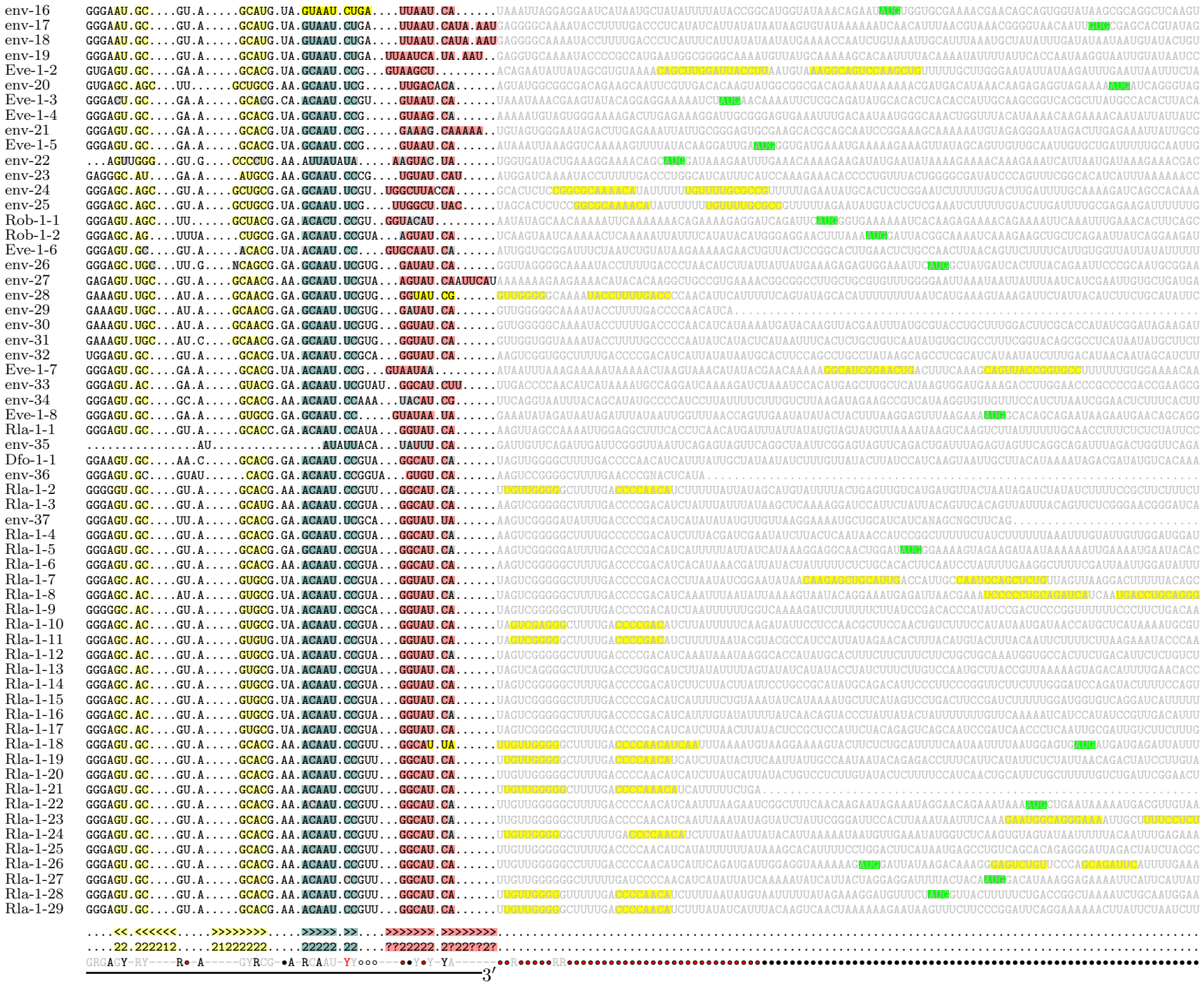


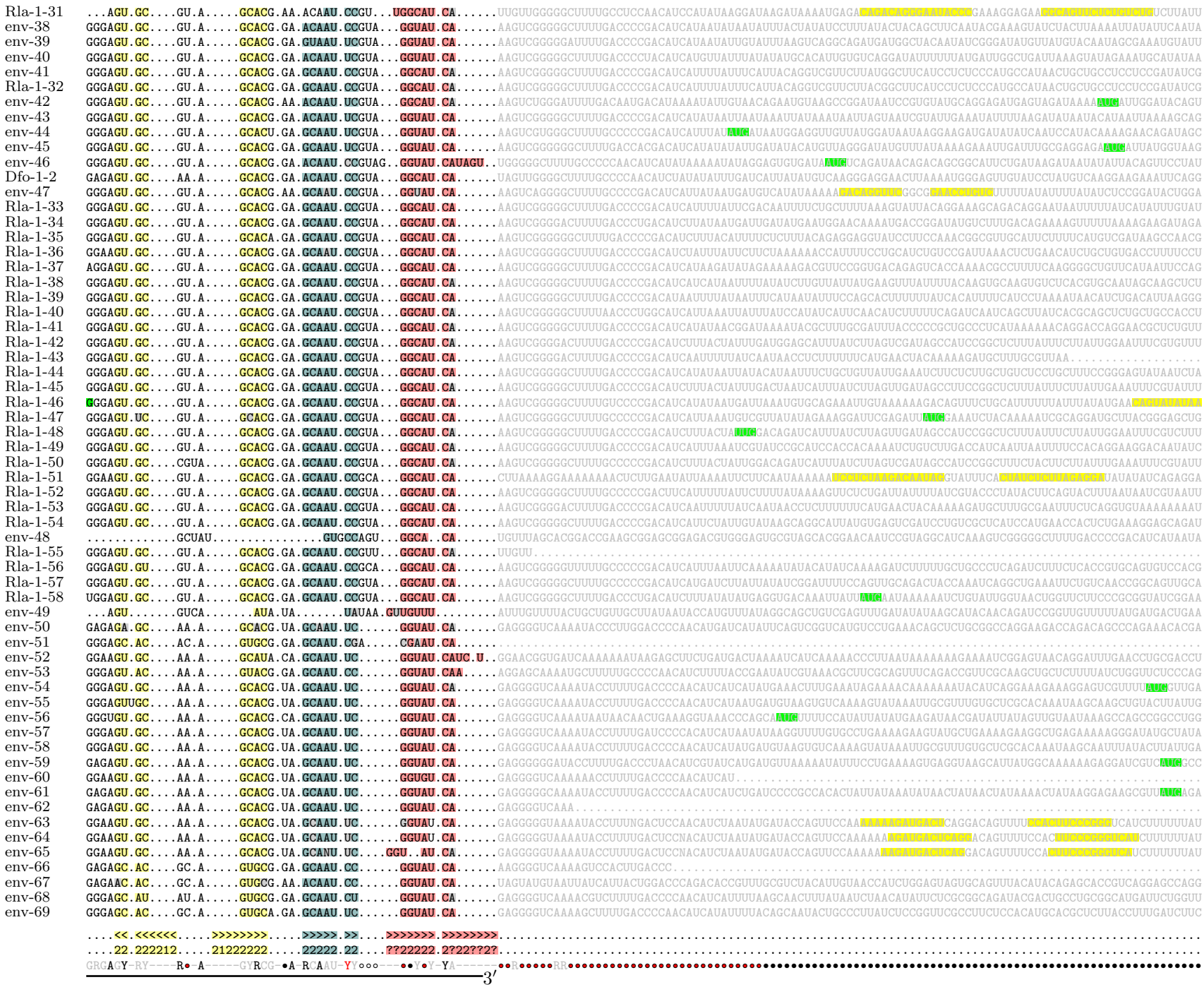
alignment positions 158 · · · 312

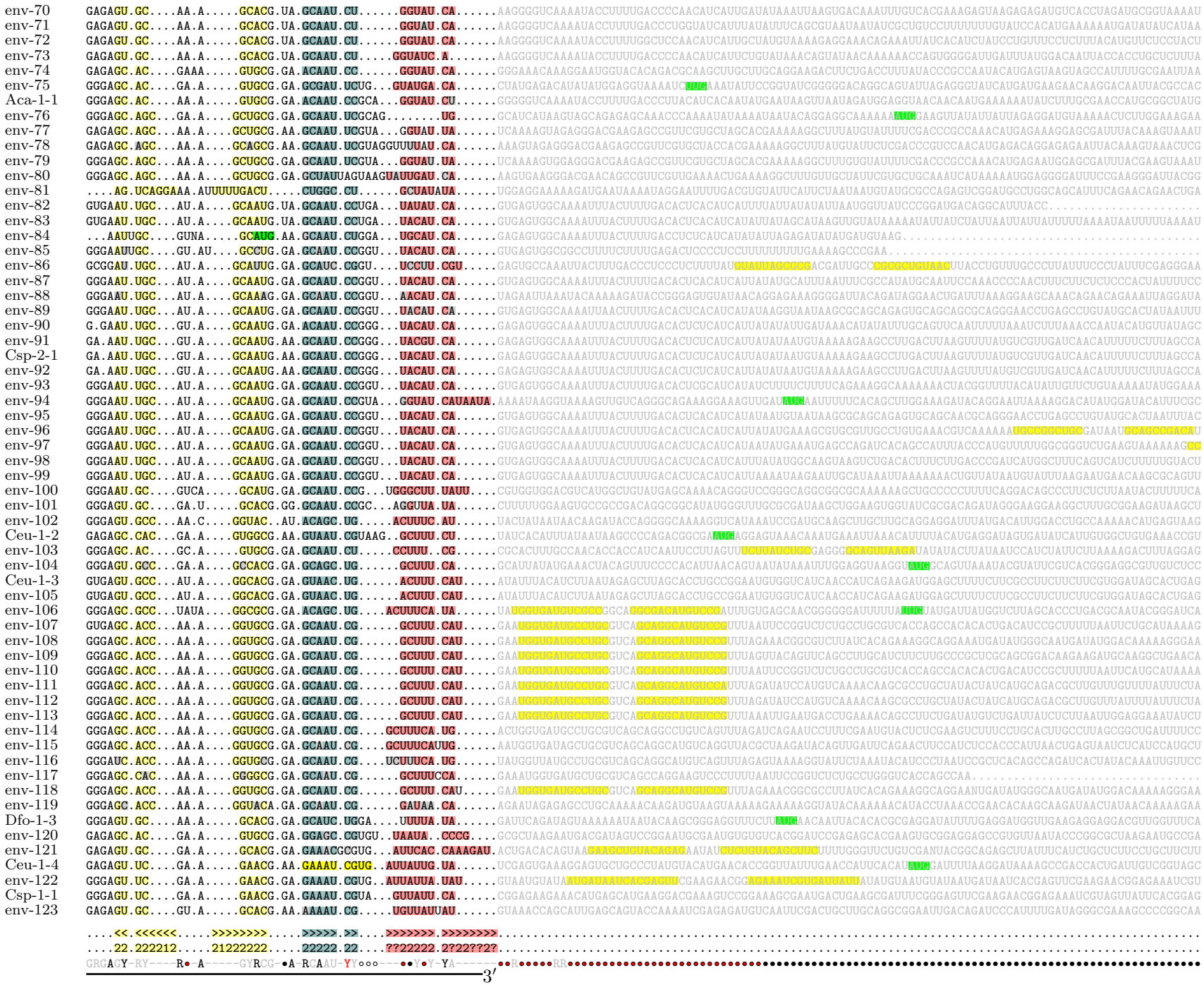

Dfo-1-1
env-36
Rla-1-2
Rla-1-3
env-37
Rla-1-4
Rla-1-5
Rla-1-6
Rla-1-7
Rla-1-8
Rla-1-9
Rla-1-10
Rla-1-11
Rla-1-12
Rla-1-13
Rla-1-14
Rla-1-15
Rla-1-16
Rla-1-17
Rla-1-18
Rla-1-19
Rla-1-20
Rla-1-21
Rla-1-22
Rla-1-23
Rla-1-24
Rla-1-25
Rla-1-26
Rla-1-27
Rla-1-28
Rla-1-29
Rla-1-31
env-38
env-39
env-40
env-41
Rla-1-32
env-42
env-43
env-44
env-45
env-46
Dfo-1-2
env-47
Rla-1-33
Rla-1-34
Rla-1-35
Rla-1-36
Rla-1-37
Rla-1-38
Rla-1-39
Rla-1-40
Rla-1-41
Rla-1-42
Rla-1-43
Rla-1-44
Rla-1-45
Rla-1-46
Rla-1-47
Rla-1-48
Rla-1-49

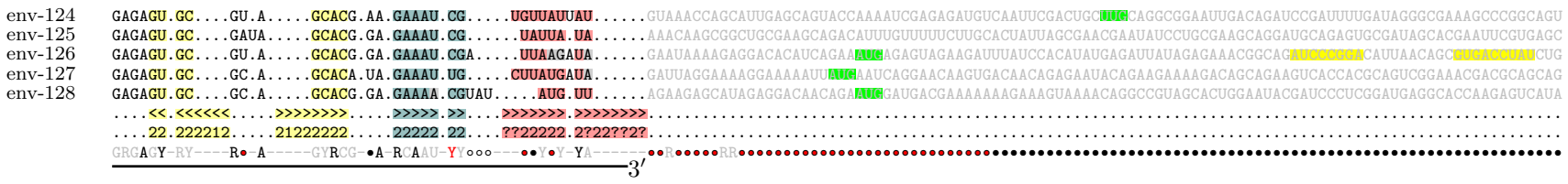

Rla-1-50
Rla-1-51
Rla-1-52
Rla-1-53
Rla-1-54
env-48
Rla-1-55
Rla-1-56
Rla-1-57
Rla-1-58
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
Aca-1-1
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
Csp-2-1
env-92
env-93
env-94
env-95
env-96
env-97


alignment positions 313 · · · 472
3′

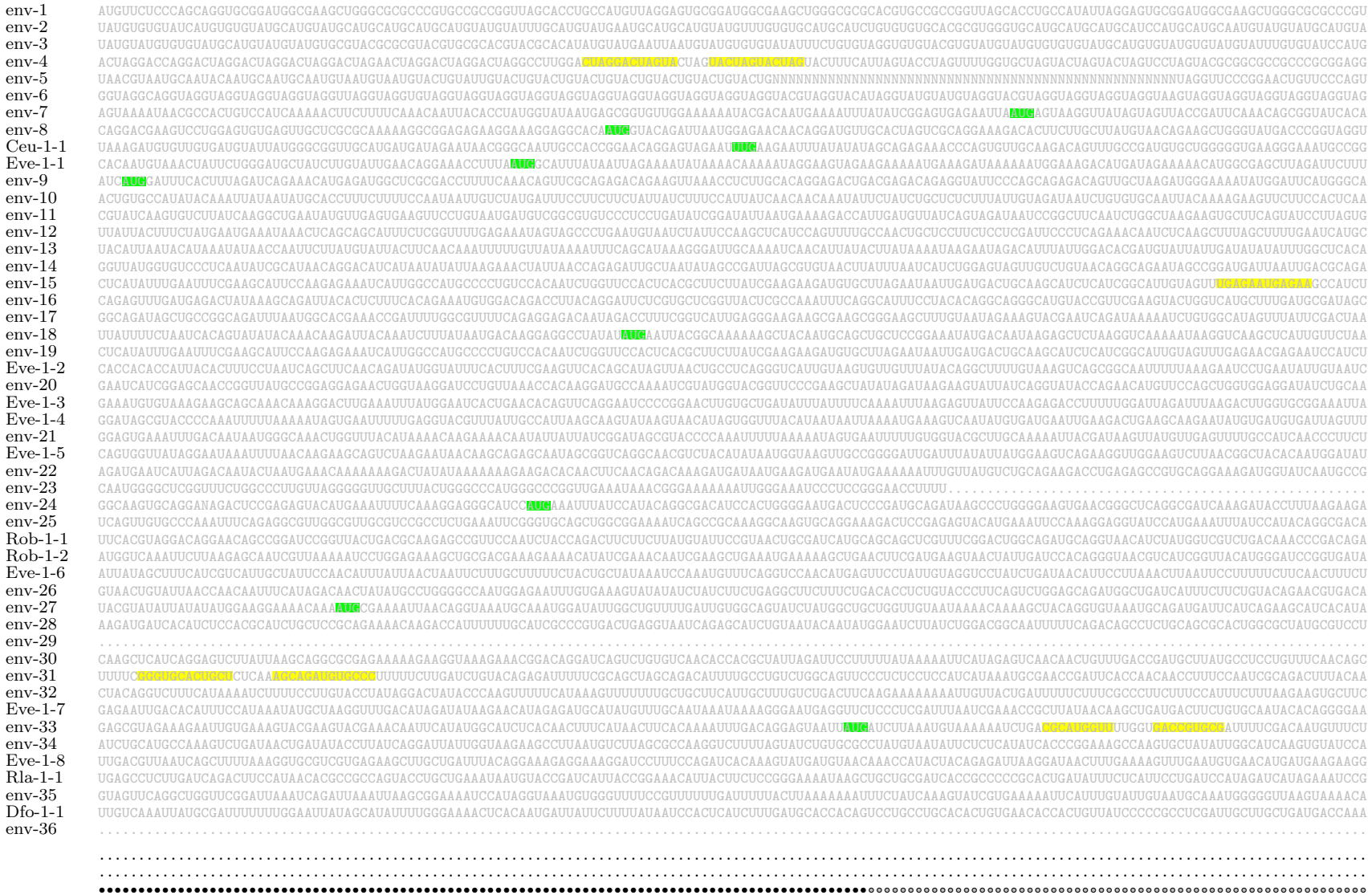

env-16
env-17
env-18
env-19
Eve-1-2
env-20
Eve-1-3
Eve-1-4
env-21
Eve-1-5
env-22
env-23
env-24
env-25
Rob-1-1
Rob-1-2
Eve-1-6
env-26
env-27
env-28
env-29
env-30
env-31
env-32
Eve-1-7
env-33
env-34
Eve-1-8
Rla-1-1
env-35
Dfo-1-1
env-36
Rla-1-2
Rla-1-3
env-37
Rla-1-4
Rla-1-5
Rla-1-6
Rla-1-7
Rla-1-8
Rla-1-9
Rla-1-10
Rla-1-11
Rla-1-12
Rla-1-13
Rla-1-14
Rla-1-15
Rla-1-16
Rla-1-17
Rla-1-18
Rla-1-19
Rla-1-20
Rla-1-21
Rla-1-22
Rla-1-23
Rla-1-24
Rla-1-25
Rla-1-26
Rla-1-27
Rla-1-28
Rla-1-29
3′


Rla-1-31
env-38
env-39
env-40
env-41
Rla-1-32
env-42
env-43
env-44
env-45
env-46
Dfo-1-2
env-47
Rla-1-33
Rla-1-34
Rla-1-35
Rla-1-36
Rla-1-37
Rla-1-38
Rla-1-39
Rla-1-40
Rla-1-41
Rla-1-42
Rla-1-43
Rla-1-44
Rla-1-45
Rla-1-46
Rla-1-47
Rla-1-48
Rla-1-49
Rla-1-50
Rla-1-51
Rla-1-52
Rla-1-53
Rla-1-54
env-48
Rla-1-55
Rla-1-56
Rla-1-57
Rla-1-58
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
3′

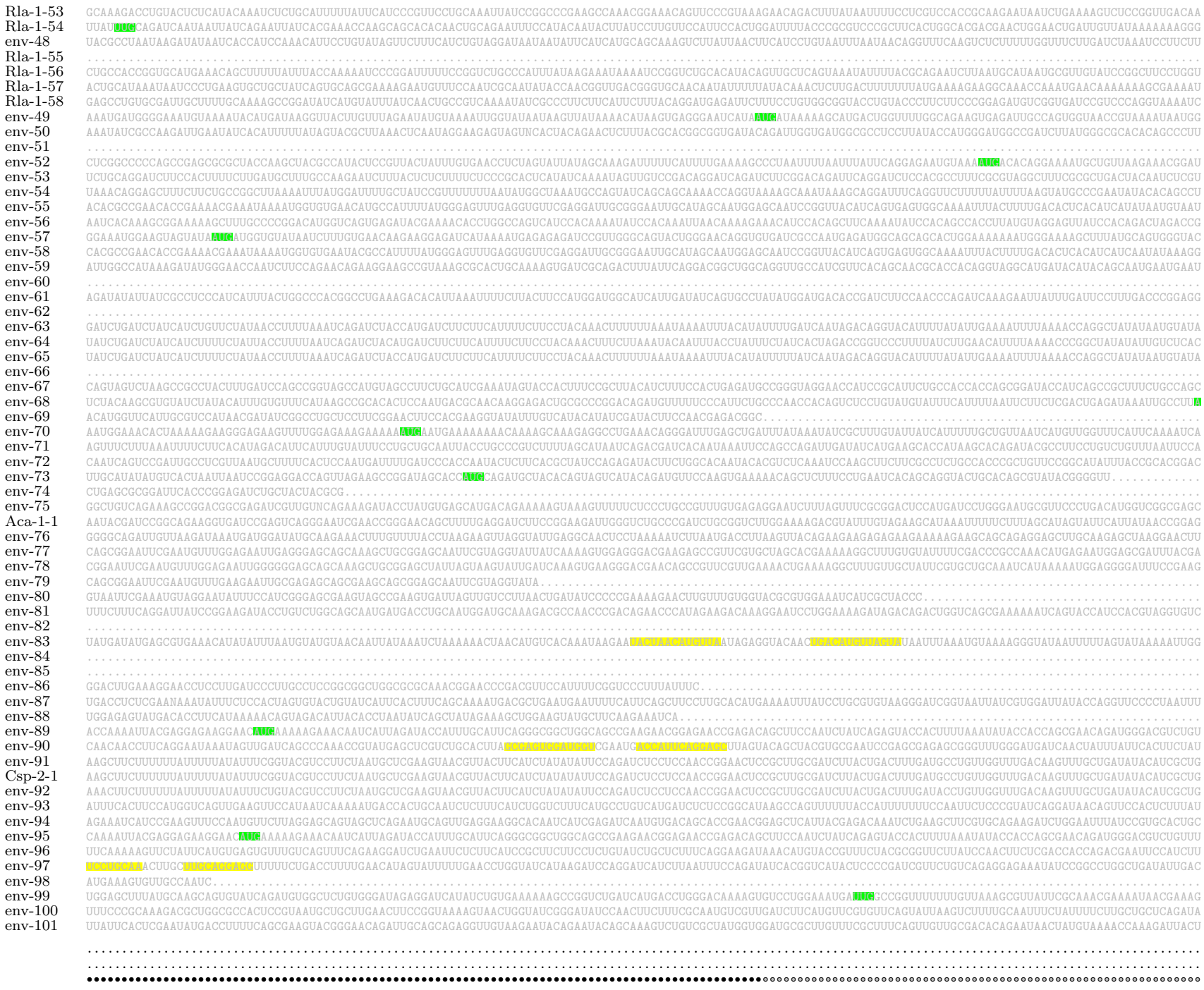

env-70
env-71
env-72
env-73
env-74
env-75
Aca-1-1
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
Csp-2-1
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
Ceu-1-2
env-103
env-104
Ceu-1-3
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
Dfo-1-3
env-120
env-121
Ceu-1-4
env-122
Csp-1-1
env-123
3′



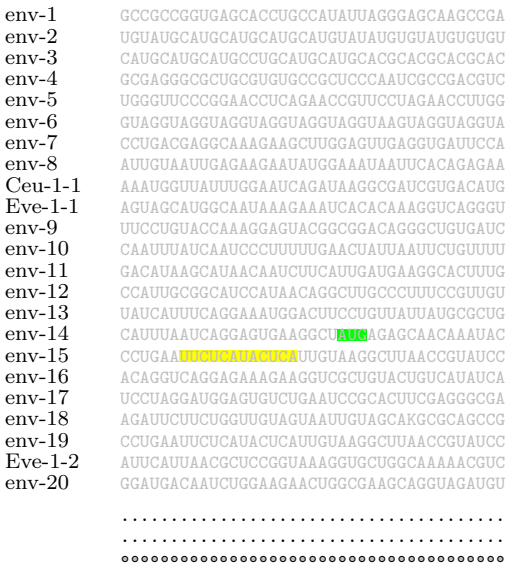
alignment positions 473 · · · 632



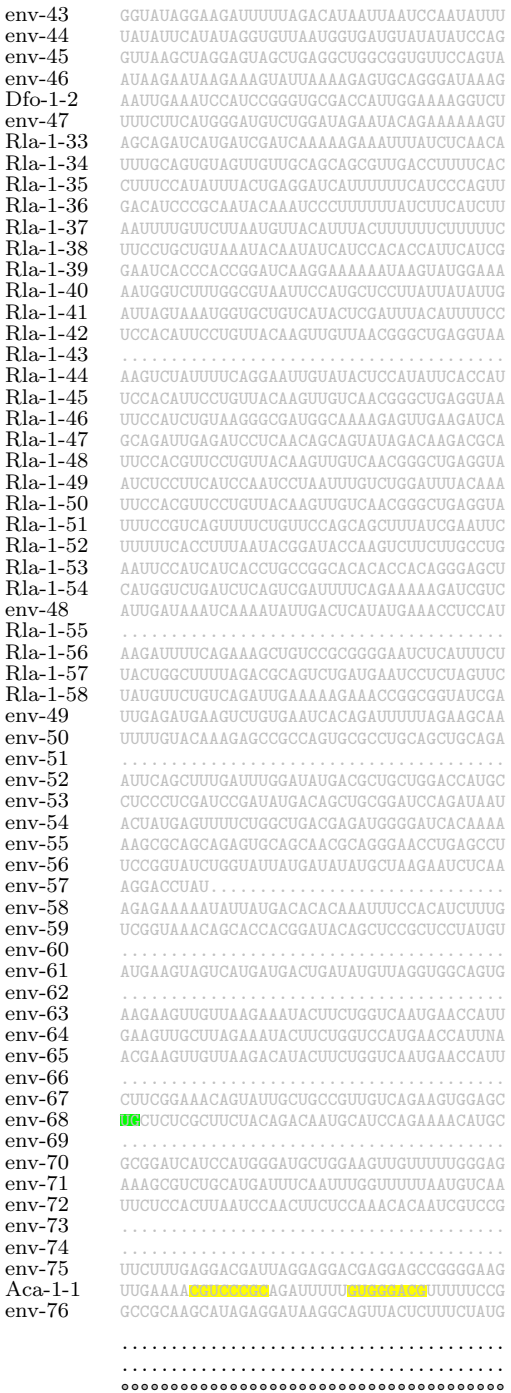
Rla-1-2
Rla-1-3
env-37
Rla-1-4
Rla-1-5
Rla-1-6
Rla-1-7
Rla-1-8
Rla-1-9
Rla-1-10
Rla-1-11
Rla-1-12
Rla-1-13
Rla-1-14
Rla-1-15
Rla-1-16
Rla-1-17
Rla-1-18
Rla-1-19
Rla-1-20
Rla-1-21
Rla-1-22
Rla-1-23
Rla-1-24
Rla-1-25
Rla-1-26
Rla-1-27
Rla-1-28
Rla-1-29
Rla-1-31
env-38
env-39
env-40
env-41
Rla-1-32
env-42
env-43
env-44
env-45
env-46
Dfo-1-2
env-47
Rla-1-33
Rla-1-34
Rla-1-35
Rla-1-36
Rla-1-37
Rla-1-38
Rla-1-39
Rla-1-40
Rla-1-41
Rla-1-42
Rla-1-43
Rla-1-44
Rla-1-45
Rla-1-46
Rla-1-47
Rla-1-48
Rla-1-49
Rla-1-50
Rla-1-51
Rla-1-52


Rla-1-53
Rla-1-54
env-48
Rla-1-55
Rla-1-56
Rla-1-57
Rla-1-58
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
Aca-1-1
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
Csp-2-1
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101

alignment positions 633 · · · 671



```
Eve-1-3     ACAUAAUUGACUACAAUGGUACACUUGACCUUGGUCAGG
Eve-1-4     ACAUAACUUGAAAACGUACCACAAAAUUUGGGAAAAUGC
env-21      ACUUAAGUGACACAUUUUUGUUUUUAUGCUAUAAUACUU
Eve-1-5     GGAAGAAGAACAUAUGCGUGAUGUUAUUGUCCCUGUUAA
env-22      GGAAACAUAUAUGUAUGUACAGAUUGUUUGCAGAGAAGU
env-23      ........................................
env-24      UAAUUGCCCAGGCUGGGAAAAUGCAGGCGGACUGUCUGU
env-25      UCCACUGGGGAAUGACUCCCGAUGCAGAUAAGCCCUGAC
Rob-1-1     AGAUGUUUUUGAAAAAAGAAUUGCAGCUCUUGAAGGUGG
Rob-1-2     AGAUUCUCGCUUUUGAUGCCCAUAUCGAUACGGUUGGUA
Eve-1-6     UUACAAAGUUAUCAAGAUUCUCUGUUCCGUCUGAAUAAA
env-26      UUAUCCGCGCUCUGAAAUACUGGGAGAAAGAGAAACUCC
env-27      AAAUUUCAGUUGUCGCAACGGUAUUUCCCUAUUAUGAUU
env-28      CCAGACAUAAUUGUAUCACUAAAUUUAUUUCUUACAGCU
env-29      ........................................
env-30      UGUGUAAGCAUCAAAACCACUGCAAAGUGAAUGAUCAGC
env-31      GCGGAAAAUGUGUACAGCCAAGAACUACUGCCUUAAUAC
env-32      AGCCACACCUUGUCUUCUAUAUGAUGGAUGCACAGCAAC
Eve-1-7     GUCAGCUUUUUACGCGUUAUGAGCUAAAUUUUAUGCGUU
env-33      UUUCGUCUCCUCAAAGGCGAACAUAUCGGCCUGAUCGGU
env-34      UUCCUUCAUCCUGAAGUGCAUAAGCCUUUAACUUAUUUA
Eve-1-8     AAAGCUUGUUAAAGAUCCUGCAAAGCUUGUUUCAAUUGC
Rla-1-1     GGUUUUCUCAUCCUGCUUUAACAGAACAGAGAUUCCCCA
env-35      UCCAGACAACCGAUAAACUUAUCAUAAAGAAAUAAAAAA
Dfo-1-1     CAGACGUUUCCAGCUUUCCUGAAUUCUCGCUUUCUGCUC
env-36      ........................................
Rla-1-2     GAACAAUGUCCCUUGUAAUCUGCUAGGUGGUCUAUAUUG
Rla-1-3     AUAGUCAUCCUCUCUCUGACUUCCUCCAAUGAUCUCACC
env-37      ........................................
Rla-1-4     AAACGGUCAACGGGCUGAGGUAAAGCGGGCCUGCACCGC
Rla-1-5     GCUUUUUGACGAUGCACAUAAGGCACUGGAUGCUGCCGU
Rla-1-6     UACUGUAGAUCAUAGCAUCAUAACCUGCUGCUUCAAUCG
Rla-1-7     AGAUAAAGGAAGAAAUUACGAACAUAUAAAAACUUCCCG
Rla-1-8     UUAAGAACAGCAGAAGCAUCUGCCCCGUUAGCAAAAGCA
Rla-1-9     AGCAGCUCUUUCAGUGCCGGAGAAAGCAGCCAGGAUUCG
Rla-1-10    GUAAGCCUUUCUGUAAAGCUUCUGCAUUGCAGGAGAAAG
Rla-1-11    GAAAUGUCAAUGUAAAAUAUGCUCCAUGCAAGUCUGCAG
Rla-1-12    CAAAGUCUGUAUGGAUCUUUCCUGCUGCCUGAGGUGCUU
Rla-1-13    CGGGUAUUCCUGAUCCUCAGCAACCGAUCCGUACACACC
Rla-1-14    CGAUCAAAAGAUUGAAAAGUGUUGCCCUGUCUCUGACUC
Rla-1-15    AGUCUUUUGUUCCCGAGACAUCAUACACCCCUGCCGGAA
Rla-1-16    UCUGUGAACCGUGUCAGGUCGGGAACGGAGCAGCACUAA
Rla-1-17    UUCCUCGUCUUUAUCAAGACCCUUAGCUGCAGAUACGGU
Rla-1-18    AUUUUAACUAUAAGUGGUGUAAGGCACUUGGACUUGAUU
Rla-1-19    AACUCCGGUUCAUGACAGAUCAGAAGGAUGCUUCCCCGG
Rla-1-20    GCAUAUCGUGGUGCUUACGCCCCACUUUUAUGCUCAUAU
Rla-1-21    ........................................
Rla-1-22    UCCGUCGGCAGAUCGUGUAGAACCGAAGUGUGCAUUUGC
Rla-1-23    CAUCAAAUCUUCUGAGAGCUUCUUUUGCUUCUGCACUCA
Rla-1-24    UUACAGGAACAAGCAUUGGUUCGACGAAGGAUAUGAUCA
Rla-1-25    AUUUACCAGCAAUCUCGGCACAAAUGCCGGAGUGAUGGC
Rla-1-26    UGCAAAAGCAUUGAAAGAAGCGGUUAAUAAGUAACGAAU
Rla-1-27    UUACUGCGGUUAUGCCGAAAUGGGGGAAGGAGAUGAUAU
Rla-1-28    UGAGCUGUAAAACGGCAGAAGAUCCUGCUCCGACCGGUU
Rla-1-29    ACAAGCUUAUUGAUCGCUUUCAUCAGACAGAACAAAAUA
Rla-1-31    CUGCCGUUUCAGGAGAGGCAUAGUCUGUCAGAUAUUCUU
env-38      AUAAUACUCGGCUUUUUUACAGCAGUCAUACAUAUCCUC
env-39      UGAAAAGUCUGAAGACAAGCUUAACUGCUAUGUGACAUU
env-40      GCAGAUUAAGGCGAUGAACCCUAAAGGAAUUAUCCUUAC
env-41      UCUCAAAAUACAGAUCCCCUGCACCGGCACUUCCACGUU
Rla-1-32    UCUCAAAAUACAGAUCCCCUGCACCGGCACUUCCACGUU
env-42      UUAUAGCCGCUUCGCAGAUAUUUUUAAUUAUUUAUCUUA

            ........................................
            ........................................
            °°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°
```


env-43 GGUAUAGGAAGAUUUUUAGACAUAAUUAAUCCAAUAUUU
env-44 UAUAUUCAUAUAGGUGUUAAUGGUGAUGUAUAUAUCCAG
env-45 GUUAAGCUAGGAGUAGCUGAGGCUGGCGGUGUUCCAGUA
env-46 AUAAGAAUAAGAAAGUAUUAAAAGAGUGCAGGGAUAAAG
Dfo-1-2 AAUUGAAAUCCAUCCGGGUGCGACCAUUGGAAAAGGUCU
env-47 UUUCUUCAUGGGAUGUCUGGAUAGAAUACAGAAAAAAGU
Rla-1-33 AGCAGAUCAUGAUCGAUCAAAAAGAAAUUUAUCUCAACA
Rla-1-34 UUUGCAGUGUAGUUGUUGCAGCAGCGUUGACCUUUUCAC
Rla-1-35 CUUUCCAUAUUUACUGAGGAUCAUUUUUUCAUCCCAGUU
Rla-1-36 GACAUCCCGCAAUACAAAUCCCUUUUUUAUCUUCAUCUU
Rla-1-37 AAUUUUGUUCUUAAUGUUACAUUUACUUUUUUCUUUUUC
Rla-1-38 UUCCUGCUGUAAAUACAAUAUCAUCCACACCAUUCAUCG
Rla-1-39 GAAUCACCCACCGGAUCAAGGAAAAAAUAAGUAUGGAAA
Rla-1-40 AAUGGUCUUUGGCGUAAUUCCAUGCUCCUUAUUAUAUUG
Rla-1-41 AUUAGUAAAUGGUGCUGUCAUACUCGAUUUACAUUUUCC
Rla-1-42 UCCACAUUCCUGUUACAAGUUGUUAACGGGCUGAGGUAA
Rla-1-43 .....
Rla-1-44 AAGUCUAUUUUCAGGAAUUGUAUACUCCAUAUUCACCAU
Rla-1-45 UCCACAUUCCUGUUACAAGUUGUCAACGGGCUGAGGUAA
Rla-1-46 UUCCAUCUGUAAGGGCGAUGGCAAAAGAGUUGAAGAUCA
Rla-1-47 GCAGAUUGAGAUCCUCAACAGCAGUAUAGACAAGACGCA
Rla-1-48 UUCCACGUUCCUGUUACAAGUUGUCAACGGGCUGAGGUA
Rla-1-49 AUCUCCUUCAUCCAAUCCUAAUUUGUCUGGAUUUACAAA
Rla-1-50 UUCCACGUUCCUGUUACAAGUUGUCAACGGGCUGAGGUA
Rla-1-51 UUUCCGUCAGUUUUCUGUUCCAGCAGCUUUAUCGAAUUC
Rla-1-52 UUUUUCACCUUUAAUACGGAUACCAAGUCUUCUUGCCUG
Rla-1-53 AAUUCCAUCAUCACCUGCCGGCACACACCACAGGGAGCU
Rla-1-54 CAUGGUCUGAUCUCAGUCGAUUUUCAGAAAAAGAUCGUC
env-48 AUUGAUAAAUCAAAAUAUUGACUCAUAUGAAACCUCCAU
Rla-1-55 .....
Rla-1-56 AAGAUUUUCAGAAAGCUGUCCGCGGGGAAUCUCAUUUCU
Rla-1-57 UACUGGCUUUUAGACGCAGUCUGAUGAAUCCUCUAGUUC
Rla-1-58 UAUGUUCUGUCAGAUUGAAAAAGAAACCGGCGGUAUCGA
env-49 UUGAGAUGAAGUCUGUGAAUCACAGAUUUUUAGAAGCAA
env-50 UUUUGUACAAAGAGCCGCCAGUGCGCCUGCAGCUGCAGA
env-51 .....
env-52 AUUCAGCUUUGAUUUGGAUAUGACGCUGCUGGACCAUGC
env-53 CUCCCUCGAUCCGAUAUGACAGCUGCGGAUCCAGAUAAU
env-54 ACUAUGAGUUUUCUGGCUGACGAGAUGGGGAUCACAAAA
env-55 AAGCGCAGCAGAGUGCAGCAACGCAGGGAACCUGAGCCU
env-56 UCCGGUAUCUGGUAUUAUGAUAUAUGCUAAGAAUCUCAA
env-57 AGGACCUAU.....
env-58 AGAGAAAAAUAUUAUGACACACAAAUUUCCACAUCUUUG
env-59 UCGGUAAACAGCACCACGGAUACAGCUCCGCUCCUAUGU
env-60 .....
env-61 AUGAAGUAGUCAUGAUGACUGAUAUGUUAGGUGGCAGUG
env-62 .....
env-63 AAGAAGUUGUUAAGAAAUACUUCUGGUCAAUGAACCAUU
env-64 GAAGUUGCUUAGAAAUACUUCUGGUCCAUGAACCAUUNA
env-65 ACGAAGUUGUUAAGACAUACUUCUGGUCAAUGAACCAUU
env-66 .....
env-67 CUUCGGAAACAGUAUUGCUGCCGUUGUCAGAAGUGGAGC
env-68 UCCUCUCGCUUCUACAGACAAUGCAUCCAGAAAACAUGC
env-69 .....
env-70 GCGGAUCAUCCAUGGGAUGCUGGAAGUUGUUUUUGGGAG
env-71 AAAGCGUCUGCAUGAUUUCAAUUUGGUUUUUAAUGUCAA
env-72 UUCUCCACUUAAUCCAACUUCUCCAAACACAAUCGUCCG
env-73 .....
env-74 .....
env-75 UUCUUUGAGGACGAUUAGGAGGACGAGGAGCCGGGGAAG
Aca-1-1 UUGAAAACGUUCCCCCAGAUUUUUGUCGGGACGUUUUUCCG
env-76 GCCGCAAGCAUAGAGGAUAAGGCAGUUACUCUUUCUAUG
.....
.....
°°

```
env-77      AGUAAAUCAGCGGAAUUCGAAUGUUUGAAGAAUUGCGAG
env-78      GGAUUACGGGUAAUUCGAAAUGUAGGAAUAUUUCCAUCG
env-79      ........................................
env-80      ........................................
env-81      UGUCGUCUCCCGGAUGGAAAGCAUGAAUGUCCAAGGAUG
env-82      ........................................
env-83      ACAGGUAAAAAUUGUACGAUAAAUACUGAAUAUAGUAAU
env-84      ........................................
env-85      ........................................
env-86      ........................................
env-87      GUUCGCUCGCCGUCAGGCGUUUGCCGCCAUCAGGCGGCG
env-88      ........................................
env-89      UUGAUCUGGAAGAGAAUGGUUAUUUCUAUACAAGACUGG
env-90      UUAUUCACUUCGAAAUGCUGAUAAUCCUUAGAGCUUCUC
env-91      UUGUCUCACCUUCCAGACUGGAGUUUGUUGCUAUGAUCA
Csp-2-1     UUGUCUCACCUUCCAGACUGGAGUUUGUUGCUAUGAUCA
env-92      UUGUCUCACCUUCCAGACUGGAGUUUGUUGCUAUGAUCA
env-93      CCCAUACUACCUGGUAUGUAUGGAUAUCUUCGCUGCUUU
env-94      AGGAGAUGGGAAAAUAAAAAAUAUAAGAAAAUUUAAAAA
env-95      GAUCUGGAAGAGAAUGGUUAUUUCUAUACAAGACUGGCA
env-96      CAGUGACAACUUCAUCACCCAUUUGAGGUAAACGGUCCA
env-97      UAACAUUUUUCUUGCCCAUGAGGUGCGGUCUUCGUAGUC
env-98      ........................................
env-99      GAGAAUGGGAAUGGGGGAAAGCAGGCUUCACUACGGAAU
env-100     UUGGGUAAAGGAGCAGCCCACCUCCUUCUGAAAGAUCUG
env-101     GAUCAGGGCAUUAUAGACUUUAAGAACGGCAGACAUCCU
env-102     CAUUGUUAAGAUCAAUCAGAAUGAUAACGUUGCCGUUGC
Ceu-1-2     GCCUGAGGCGGCAGCAGCUCUCUCAAAGCAGAGUGAGGA
env-103     NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
env-104     ........................................
Ceu-1-3     AGUUUCUUUGACUUCGGUACUGCAACACACUCUUUAACC
env-105     AGUUUCUUUGACUUCGGUACCGCAACACACU........
env-106     GGAGAACCAGACCGGUUCUCUGAUCACCAAAGAUAAGGA
env-107     GAUACCGGGAAUAAAAGCUCAUAUNAUAUUCUUCCAGUU
env-108     CGGAAAUGGAGAGUAUGCUAAUCCAUCCGGCAGUUUAAG
env-109     GUAUGUUUUUUAUAAUACAACGUUUUACUUGUUAUUCUA
env-110     UGAUACCGGGAAUAAAAGCUCAUAUAAUAUUCUUCCAGU
env-111     ACCAAAAUAUACAUUCUUUUCAUUAUGAUGAAUAUCCUG
env-112     GACGUAUACAAUACCUGAUCUUCAUCACCAUAAACAGUU
env-113     UUACAGCAGUAACCCCUCUCUCCAUAUUUGUCAUGAUCG
env-114     ........................................
env-115     ........................................
env-116     CAUGCAUUACUUUAUUCAUUCCAGGAUAUUGGUAGGCAU
env-117     ........................................
env-118     CGGAAAUGGAGAGUAUGCUAAUCCAUCCGGCAGUUUAAG
env-119     GAAAAAACCACCAAAUAACGACUAUAAAAAGGAAACCAG
Dfo-1-3     GAAUACAUUUGAAAUCUUUCCGUGGAGACCGCAGCAGGG
env-120     UUUAUGGUAAAAUAUCCCAAUCUAUAACAAAAAAUGUGC
env-121     UUACUUGCCAUUUCGUUCACCUCCGUUUGGAUUUUGGUU
Ceu-1-4     UCCGUCCAAAAGAGCUUCAGAAAUAUGAAUACUAUGGGG
env-122     AAAAAUUAUAUAUGGAAUAUUGAUCAUCUGUAUGGUGUU
Csp-1-1     GGCAGACGAUGAGAACGUAACCGCAUAUAUAUUUGGAGU
env-123     ........................................
env-124     UUUCCGCCAUCAGAACAGCCCAGCGGGAGACCAAAGAUG
env-125     AAAUUCGAUCUGGCACUUUUGGAUAUCAUGCUGCCGGGA
env-126     CUUCAGCCUAGAUCCUGC......................
env-127     UGUACGAUACGGACGGCGGAAGUGAAGUGGCAGCCCGGA
env-128     ........................................
            ........................................
            ........................................
            °°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°
```

22 COG2252

22.1 Taxa

The taxonomy of each organism containing a putative COG2252 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual COG2252 RNA in Sections 22.2 and 22.4.

abbrev. of hits	taxonomy of species
Pen-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas entomophila* L48
Pfl-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf-5
Pfl-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf0-1
Ppu-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* F1
Ppu-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* GB-1
Ppu-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* KT2440
Ppu-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* W619
Psy-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. phaseolicola* 1448A
Psy-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000
env-1	environmental sample

22.2 Gene contexts

Each COG2252 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one COG2252 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 22.3. The accession number of the sequence containing each COG2252 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 22.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	JCVI_SCAF_1096627385357	+	1717	1779	**RNA**→ COG2252 (COG2252)tRNA_U5-meth_tr (pfam05958)? (GOS376)→
Pfl-2-1	NC_007492.1	+	5520810	5520872	**RNA**→ COG2252 (COG2252)→ tRNA_U5-meth_tr (pfam05958)→
Pen-1-1	NC_008027.1	+	5083735	5083796	**RNA**→ COG2252 (COG2252)→ tRNA_U5-meth_tr (pfam05958)→ LysR (COG0583)→
Ppu-4-1	NC_010501.1	-	877139	877077	**RNA**→ COG2252 (COG2252)→ tRNA_U5-meth_tr (pfam05958)→ LysR (COG0583)→
Ppu-3-1	NC_002947.3	+	5278425	5278487	**RNA**→ COG2252 (COG2252)→ tRNA_U5-meth_tr (pfam05958)→
Ppu-1-1	NC_009512.1	+	5040655	5040717	**RNA**→ COG2252 (COG2252)→ tRNA_U5-meth_tr (pfam05958)→
Ppu-2-1	NC_010322.1	+	5185631	5185693	**RNA**→ COG2252 (COG2252)→ tRNA_U5-meth_tr (pfam05958)→
Psy-2-1	NC_007005.1	+	5111437	5111499	**RNA**→ COG2252 (COG2252)→ tRNA_U5-meth_tr (pfam05958)→
Psy-3-1	NC_004578.1	+	5256253	5256315	**RNA**→ COG2252 (COG2252)→ tRNA_U5-meth_tr (pfam05958)→
Psy-1-1	NC_005773.3	+	4965899	4965961	**RNA**→ COG2252 (COG2252)→ tRNA_U5-meth_tr (pfam05958)→
Pfl-1-1	NC_004129.6	+	6139548	6139611	**RNA**→ COG2252 (COG2252)→ tRNA_U5-meth_tr (pfam05958)→

22.3 Conserved domains

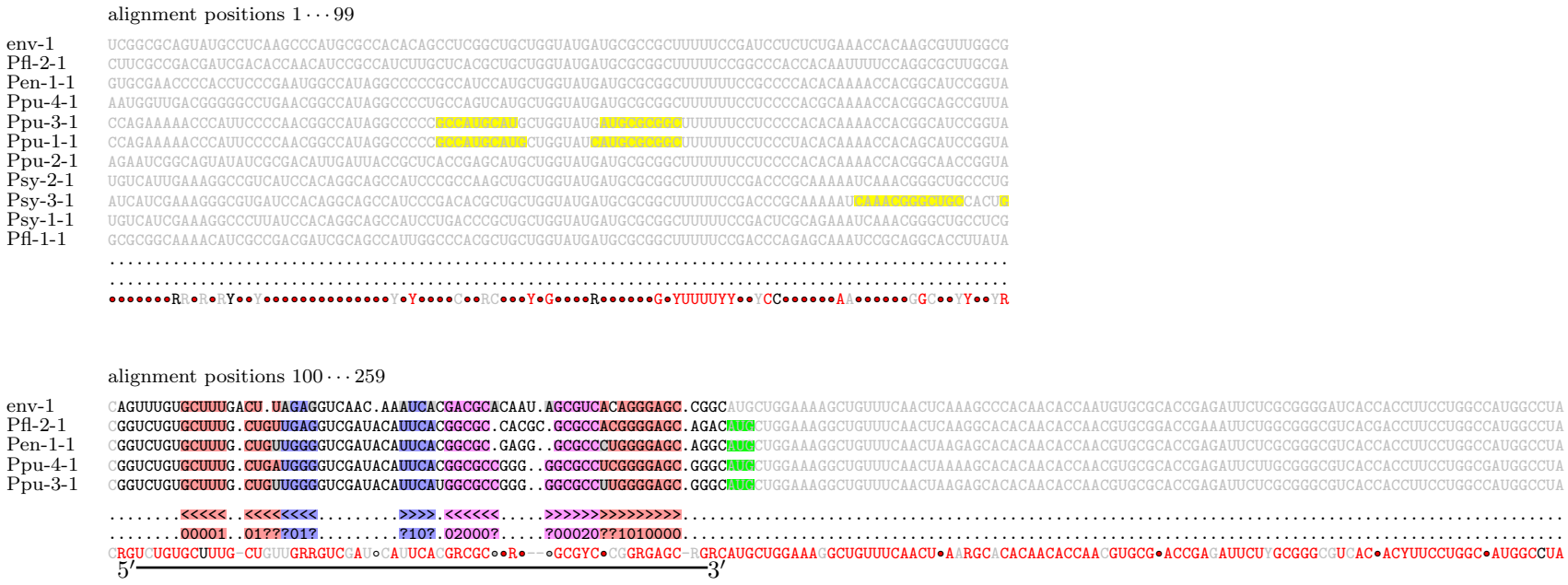
Conserved domains found in protein-coding genes listed in Section 22.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one COG2252 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

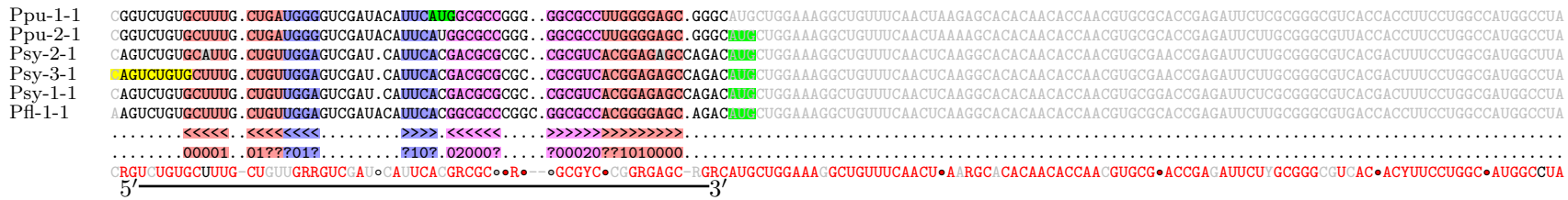
COG0583 (2) Transcriptional regulator [Transcription]
COG2252 (11) Permeases [General function prediction only]
GOS376 (1) ?

pfam05958 (11) tRNA (Uracil-5-)-methyltransferase.

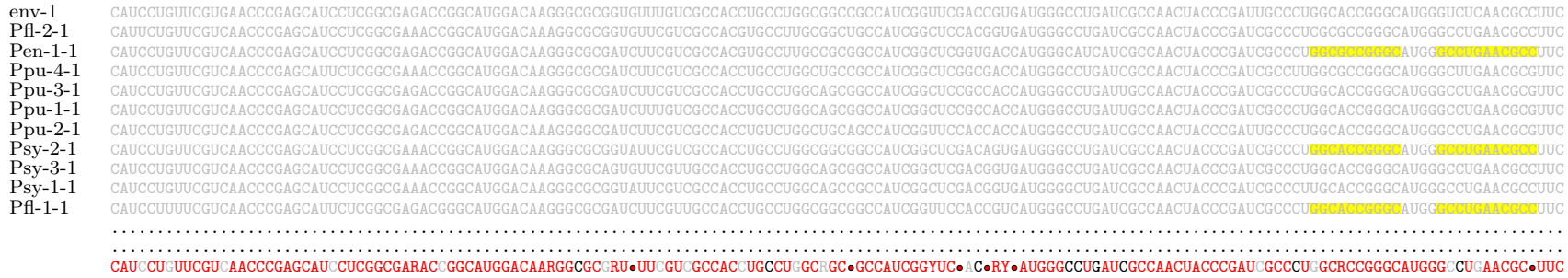
22.4 Multiple-sequence alignment

Each COG2252 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 22.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The COG2252 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.





alignment positions 260 · · · 419



alignment positions 420 · · · 467

23 *Collinsella*-1

23.1 Taxa

The taxonomy of each organism containing a putative *Collinsella*-1 RNA is listed, with abbreviations identifying each hit (e.g., “Eco-1-1” and “Eco-1-2” might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *Collinsella*-1 RNA in Sections 23.2 and 23.4.

abbrev. of hits	**taxonomy of species**
Cae-1-1	Bacteria Actinobacteria Coriobacteridae Coriobacteriales Coriobacterineae Coriobacteriaceae *Collinsella aerofaciens* ATCC 25986
uba-1-1	Bacteria environmental samples uncultured bacterium
env-1 to env-4	environmental samples

23.2 Gene contexts

Each *Collinsella*-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *Collinsella*-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 23.3. The accession number of the sequence containing each *Collinsella*-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 23.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
uba-1-1	NC_008154.1	+	4029	4216	←COG4186 (COG4186) **RNA**→ putative replication protein→
env-1	hgutS7_s7_166479	-	545	363	hypo→ **RNA**→ ←hypo
Cae-1-1	NZ_AAVN02000005.1	+	150266	150458	hypo→ **RNA**→ ←ADP_ribosyl_GH (pfam03747)
env-2	BAAV01020360.1	+	623	820	hypo→ **RNA**→
env-3	hgutS8_s8_179322_2	-	1637	1441	hypo→ **RNA**→ ←DUF1016 (pfam06250)
env-4	hgutS7_s7_179322_2	-	721	525	**RNA**→ ←Uncharacterized conserved protein

23.3 Conserved domains

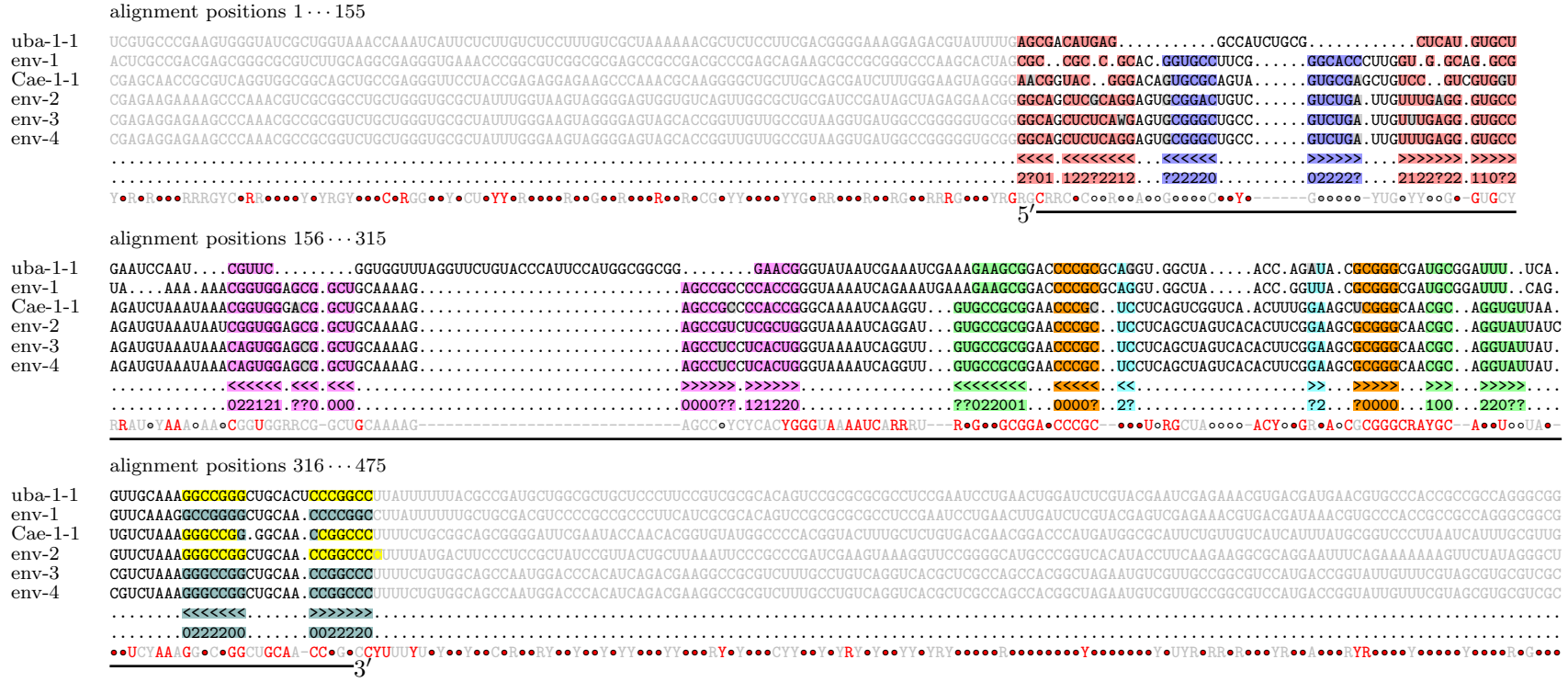
Conserved domains found in protein-coding genes listed in Section 23.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *Collinsella*-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

`u` COG4186 (1) Predicted phosphoesterase or phosphohydrolase [General function prediction only]

`d` pfam03747 (1) ADP-ribosylglycohydrolase.
`d` pfam06250 (1) Protein of unknown function (DUF1016).

23.4 Multiple-sequence alignment

Each *Collinsella*-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 23.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *Collinsella*-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



24 *crcB*

24.1 Taxa

The taxonomy of each organism containing a putative *crcB* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *crcB* RNA in Sections 24.3 and 24.5.

abbrev. of hits	**taxonomy of species**
Mse-1-1	Archaea Crenarchaeota Thermoprotei Sulfolobales Sulfolobaceae *Metallosphaera sedula* DSM 5348
Sso-1-1	Archaea Crenarchaeota Thermoprotei Sulfolobales Sulfolobaceae *Sulfolobus solfataricus* P2
Abo-1-1 to Abo-1-4	Archaea Euryarchaeota *Aciduliprofundum boonei* T469
Afu-1-1	Archaea Euryarchaeota Archaeoglobi Archaeoglobales Archaeoglobaceae *Archaeoglobus fulgidus* DSM 4304
Msm-1-1	Archaea Euryarchaeota Methanobacteria Methanobacteriales Methanobacteriaceae *Methanobrevibacter smithii* ATCC 35061
Msm-2-1	Archaea Euryarchaeota Methanobacteria Methanobacteriales Methanobacteriaceae *Methanobrevibacter smithii* DSM 2374
Msm-3-1	Archaea Euryarchaeota Methanobacteria Methanobacteriales Methanobacteriaceae *Methanobrevibacter smithii* DSM 2375
Mst-1-1	Archaea Euryarchaeota Methanobacteria Methanobacteriales Methanobacteriaceae *Methanosphaera stadtmanae* DSM 3091
Mth-1-1	Archaea Euryarchaeota Methanobacteria Methanobacteriales Methanobacteriaceae *Methanothermobacter thermautotrophicus str.* Delta H
CMe-1-1 to CMe-1-2	Archaea Euryarchaeota Methanomicrobia Methanomicrobiales Generaincertae sedis *Candidatus* Methanoregula boonei 6A8
Mhu-1-1 to Mhu-1-3	Archaea Euryarchaeota Methanomicrobia Methanomicrobiales Methanospirillaceae *Methanospirillum hungatei* JF-1
Mac-1-1	Archaea Euryarchaeota Methanomicrobia Methanosarcinales Methanosarcinaceae *Methanosarcina acetivorans* C2A
Mba-1-1	Archaea Euryarchaeota Methanomicrobia Methanosarcinales Methanosarcinaceae *Methanosarcina barkeri str.* Fusaro
Pab-1-1	Archaea Euryarchaeota Thermococci Thermococcales Thermococcaceae *Pyrococcus abyssi* GE5
Pfu-1-1	Archaea Euryarchaeota Thermococci Thermococcales Thermococcaceae *Pyrococcus furiosus* DSM 3638
Pho-1-1	Archaea Euryarchaeota Thermococci Thermococcales Thermococcaceae *Pyrococcus horikoshii* OT3
Tba-1-1	Archaea Euryarchaeota Thermococci Thermococcales Thermococcaceae *Thermococcus barophilus* MP
Tko-1-1	Archaea Euryarchaeota Thermococci Thermococcales Thermococcaceae *Thermococcus kodakarensis* KOD1
Ton-1-1	Archaea Euryarchaeota Thermococci Thermococcales Thermococcaceae *Thermococcus onnurineus* NA1
Tsp-1-1	Archaea Euryarchaeota Thermococci Thermococcales Thermococcaceae *Thermococcus sp.* AM4
Pto-1-1	Archaea Euryarchaeota Thermoplasmata Thermoplasmatales Picrophilaceae *Picrophilus torridus* DSM 9790
Tvo-1-1 to Tvo-1-2	Archaea Euryarchaeota Thermoplasmata Thermoplasmatales Thermoplasmataceae *Thermoplasma volcanium* GSS1
Aba-1-1	Bacteria Acidobacteria Acidobacteriales Acidobacteriaceae *Acidobacteria bacterium* Ellin345
Mab-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium abscessus*
Mma-1-1 to Mma-1-3	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium marinum* M
Mul-1-1 to Mul-1-2	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium ulcerans* Agy99
Nfa-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Nocardiaceae *Nocardia farcinica* IFM 10152
Ace-1-1 to Ace-1-2	Bacteria Actinobacteria Actinobacteridae Actinomycetales Frankineae Acidothermaceae *Acidothermus cellulolyticus* 11B
Lxy-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Microbacteriaceae *Leifsonia xyli subsp. xyli str.* CTCB07
Bad-1-1	Bacteria Actinobacteria Actinobacteridae Bifidobacteriales Bifidobacteriaceae *Bifidobacterium adolescentis* ATCC 15703
Bad-2-1	Bacteria Actinobacteria Actinobacteridae Bifidobacteriales Bifidobacteriaceae *Bifidobacterium adolescentis* L2-32
Ban-1-1	Bacteria Actinobacteria Actinobacteridae Bifidobacteriales Bifidobacteriaceae *Bifidobacterium angulatum* DSM 20098
Blo-1-1 to Blo-1-2	Bacteria Actinobacteria Actinobacteridae Bifidobacteriales Bifidobacteriaceae *Bifidobacterium longum* DJO10A
Blo-2-1	Bacteria Actinobacteria Actinobacteridae Bifidobacteriales Bifidobacteriaceae *Bifidobacterium longum* NCC2705
Blo-3-1	Bacteria Actinobacteria Actinobacteridae Bifidobacteriales Bifidobacteriaceae *Bifidobacterium longum subsp. infantis* CCUG 52486
Rxy-1-1	Bacteria Actinobacteria Rubrobacteridae Rubrobacterales Rubrobacterineae Rubrobacteraceae *Rubrobacter xylanophilus* DSM 9941
Hsp-1-1 to Hsp-1-2	Bacteria Aquificae Aquificales Aquificaceae *Hydrogenobaculum sp.* Y04AAS1
Bca-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides capillosus* ATCC 29799
Bco-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides coprocola* DSM 17136
Bpl-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides plebeius* DSM 17135
Bth-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides thetaiotaomicron* VPI-5482
Lbl-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Leeuwenhoekiella blandensis* MED217
Psp-1-1 to Psp-1-2	Bacteria Bacteroidetes Sphingobacteria Sphingobacteriales Sphingobacteriaceae *Pedobacter sp.* BAL39
Cag-1-1	Bacteria Chloroflexi Chloroflexales Chloroflexaceae *Chloroflexus aggregans* DSM 9485
Cau-1-1	Bacteria Chloroflexi Chloroflexales Chloroflexaceae *Chloroflexus aurantiacus* J-10-fl
Csp-1-1	Bacteria Chloroflexi Chloroflexales Chloroflexaceae *Chloroflexus sp.* Y-400-fl
Rsp-1-1	Bacteria Chloroflexi Chloroflexales Chloroflexaceae *Roseiflexus sp.* RS-1
Sel-1-1	Bacteria Cyanobacteria Chroococcales *Synechococcus elongatus* PCC 6301
Sel-2-1	Bacteria Cyanobacteria Chroococcales *Synechococcus elongatus* PCC 7942
Ssp-1-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* JA-2-3B'a(2-13)
Ssp-2-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* JA-3-3Ab
Ssp-3-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* PCC 7002
Gvi-1-1	Bacteria Cyanobacteria Gloeobacteria Gloeobacterales *Gloeobacter violaceus* PCC 7421
Ava-1-1	Bacteria Cyanobacteria Nostocales Nostocaceae *Anabaena variabilis* ATCC 29413 (Anabaena flos-aquae UTEX 1444)
Nsp-1-1	Bacteria Cyanobacteria Nostocales Nostocaceae *Nostoc sp.* PCC 7120
Dge-1-1	Bacteria Deinococcus-Thermus Deinococci Deinococcales Deinococcaceae *Deinococcus geothermalis* DSM 11300
Taq-1-1	Bacteria Deinococcus-Thermus Deinococci Thermales Thermaceae *Thermus aquaticus* Y51MC23
Dth-1-1	Bacteria Dictyoglomi Dictyoglomales Dictyoglomaceae *Dictyoglomus thermophilum* H-6-12
Ban-3-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* 'Ames Ancestor'

Ban-4-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* A0174
Ban-5-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* A0193
Ban-6-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* A0389
Ban-7-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* A0442
Ban-8-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* A0465
Ban-9-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* A0488
Ban-10-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* A1055
Ban-11-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* A2012
Ban-12-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Ames
Ban-13-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Australia 94
Ban-14-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* CNEVA-9066
Ban-15-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Kruger B
Ban-16-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Sterne
Ban-17-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Vollum
Ban-18-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Western North America USA6153
Ban-2-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis* Tsiankovskii-I
Bce-1-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* 03BB108
Bce-2-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* AH1134
Bce-3-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* AH187
Bce-4-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* AH820
Bce-5-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* ATCC 10987
Bce-6-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* ATCC 14579
Bce-7-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* B4264
Bce-8-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* E33L
Bce-9-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* G9241
Bce-10-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* G9842
Bce-11-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* H3081.97
Bce-12-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* NVH0597-99
Bce-14-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus subsp. cytotoxis* NVH 391-98
Bce-13-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* W
Bth-2-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus thuringiensis serovar israelensis* ATCC 35646
Bth-3-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus thuringiensis serovar konkukian str.* 97-27
Bth-4-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus thuringiensis str.* Al Hakam
Bwe-1-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus weihenstephanensis* KBAB4
Gka-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Geobacillus kaustophilus* HTA426
Gsp-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Geobacillus sp.* G11MC16
Gsp-2-1	Bacteria Firmicutes Bacillales Bacillaceae *Geobacillus sp.* WCH70
Gsp-3-1	Bacteria Firmicutes Bacillales Bacillaceae *Geobacillus sp.* Y412MC10
Gth-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Geobacillus thermodenitrificans* NG80-2
Psp-2-1	Bacteria Firmicutes Bacillales Paenibacillaceae *Paenibacillus sp.* JDR-2
Ame-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Alkaliphilus metalliredigens* QYMF
Aor-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Alkaliphilus oremlandii* OhILAs
Cac-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium acetobutylicum* ATCC 824
Cbo-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium bolteae* ATCC BAA-613
Cbo-2-1 to Cbo-2-2	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* A str. ATCC 19397
Cbo-3-1 to Cbo-3-2	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* A str. ATCC 3502
Cbo-4-1 to Cbo-4-2	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* A str. Hall
Cbo-5-1 to Cbo-5-2	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* A3 str. Loch Maree
Cbo-6-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* B str. Eklund 17B
Cbo-7-1 to Cbo-7-2	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* B1 str. Okra
Cbo-8-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* Bf
Cbo-9-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* E3 str. Alaska E43
Cbo-10-1 to Cbo-10-2	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* F str. Langeland
Cbo-11-1 to Cbo-11-2	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* NCTC 2916
Cbu-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium butyricum* 5521
Cdi-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium difficile* 630
Cdi-2-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium difficile* ATCC 43255
Cdi-3-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium difficile* CIP 107932
Cdi-4-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium difficile* QCD-23m63
Cdi-5-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium difficile* QCD-32g58
Cdi-6-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium difficile* QCD-37x79
Cdi-7-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium difficile* QCD-63q42
Cdi-8-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium difficile* QCD-66c26
Cdi-9-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium difficile* QCD-76w55
Cdi-10-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium difficile* QCD-97b34

Chi-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium hiranonis* DSM 13275
Ckl-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium kluyveri* DSM 555
Cpe-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium perfringens* ATCC 13124
Cpe-2-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium perfringens* B str. ATCC 3626
Cpe-3-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium perfringens* C str. JGS1495
Cpe-4-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium perfringens* CPE str. F4969
Cpe-5-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium perfringens* D str. JGS1721
Cpe-6-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium perfringens* E str. JGS1987
Cpe-7-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium perfringens* NCTC 8239
Cpe-8-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium perfringens* SM101
Cpe-9-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium perfringens str.* 13
Cph-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium phytofermentans* ISDg
Csp-2-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium sp.* SS2/1
Csp-3-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium sporogenes* ATCC 15579
Cth-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium thermocellum* ATCC 27405
Cth-2-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium thermocellum* DSM 4150
Cba-1-1	Bacteria Firmicutes Clostridia *Clostridiales bacterium* 1_7_47_FAA
Ahy-1-1	Bacteria Firmicutes Clostridia Clostridiales ClostridialesFamily XI. Incertae Sedis *Anaerococcus hydrogenalis* DSM 7454
Ala-1-1	Bacteria Firmicutes Clostridia Clostridiales ClostridialesFamily XI. Incertae Sedis *Anaerococcus lactolyticus* ATCC 51172
Fma-1-1	Bacteria Firmicutes Clostridia Clostridiales ClostridialesFamily XI. Incertae Sedis *Finegoldia magna* ATCC 29328
Eve-1-1	Bacteria Firmicutes Clostridia Clostridiales Eubacteriaceae *Eubacterium ventriosum* ATCC 27560
Hmo-1-1	Bacteria Firmicutes Clostridia Clostridiales Heliobacteriaceae *Heliobacterium modesticaldum* Ice1
Ceu-1-1	Bacteria Firmicutes Clostridia Clostridiales Lachnospiraceae *Coprococcus eutactus* ATCC 27759
Dlo-1-1	Bacteria Firmicutes Clostridia Clostridiales Lachnospiraceae *Dorea longicatena* DSM 13814
Rin-1-1	Bacteria Firmicutes Clostridia Clostridiales Lachnospiraceae *Roseburia intestinalis* L1-82
Dha-1-1	Bacteria Firmicutes Clostridia Clostridiales Peptococcaceae *Desulfitobacterium hafniense* DCB-2
Dha-2-1	Bacteria Firmicutes Clostridia Clostridiales Peptococcaceae *Desulfitobacterium hafniense* Y51
Pth-1-1	Bacteria Firmicutes Clostridia Clostridiales Peptococcaceae *Pelotomaculum thermopropionicum* SI
Rob-1-1	Bacteria Firmicutes Clostridia Clostridiales Ruminococcaceae *Ruminococcus obeum* ATCC 29174
Tca-1-1	Bacteria Firmicutes Clostridia Clostridiales Veillonellaceae *Thermosinus carboxydivorans* Nor1
Nth-1-1	Bacteria Firmicutes Clostridia Natranaerobiales Natranaerobiaceae *Natranaerobius thermophilus* JW/NM-WN-LF
Cpa-1-1 to Cpa-1-3	Bacteria Firmicutes Clostridia Thermoanaerobacterales Thermoanaerobacteraceae *Carboxydibrachium pacificum* DSM 12653
Chy-1-1	Bacteria Firmicutes Clostridia Thermoanaerobacterales Thermoanaerobacteraceae *Carboxydothermus hydrogenoformans* Z-2901
Mth-2-1	Bacteria Firmicutes Clostridia Thermoanaerobacterales Thermoanaerobacteraceae Moorella group *Moorella thermoacetica* ATCC 39073
Tps-1-1 to Tps-1-2	Bacteria Firmicutes Clostridia Thermoanaerobacterales Thermoanaerobacteraceae *Thermoanaerobacter pseudethanolicus* ATCC 33223
Tsp-2-1 to Tsp-2-2	Bacteria Firmicutes Clostridia Thermoanaerobacterales Thermoanaerobacteraceae *Thermoanaerobacter sp.* X514
Tte-1-1 to Tte-1-3	Bacteria Firmicutes Clostridia Thermoanaerobacterales Thermoanaerobacteraceae *Thermoanaerobacter tengcongensis* MB4
Csa-1-1 to Csa-1-2	Bacteria Firmicutes Clostridia Thermoanaerobacterales Thermoanaerobacterales Family III. Incertae Sedis *Caldicellulosiruptor saccharolyticus* DSM 8903
Ebi-1-1	Bacteria Firmicutes Erysipelotrichi Erysipelotrichales Erysipelotrichaceae *Eubacterium biforme* DSM 3989
Csp-4-1	Bacteria Firmicutes Lactobacillales Carnobacteriaceae *Carnobacterium sp.* AT7
Efa-1-1	Bacteria Firmicutes Lactobacillales Enterococcaceae *Enterococcus faecalis* OG1RF
Efa-2-1	Bacteria Firmicutes Lactobacillales Enterococcaceae *Enterococcus faecalis* V583
Efa-3-1	Bacteria Firmicutes Lactobacillales Enterococcaceae *Enterococcus faecium* DO
Lac-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus acidophilus* NCFM
Lca-1-1 to Lca-1-2	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus casei* ATCC 334
Lca-2-1 to Lca-2-2	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus casei* BL23
Lga-1-1 to Lga-1-2	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus gasseri* ATCC 33323
Lga-2-1 to Lga-2-2	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus gasseri* MV-22
Lhe-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus helveticus* DPC 4571
Ljo-1-1 to Ljo-1-2	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus johnsonii* NCC 533
Lpa-1-1 to Lpa-1-2	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus paracasei subsp. paracasei* 8700:2
Lpl-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus plantarum* WCFS1
Lrh-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus rhamnosus* HN001
Lla-1-1 to Lla-1-2	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. cremoris* MG1363
Lla-2-1 to Lla-2-2	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. cremoris* SK11
Lla-3-1 to Lla-3-2	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. lactis* Il1403
Tye-1-1	Bacteria Nitrospirae Nitrospirales Nitrospiraceae *Thermodesulfovibrio yellowstonii* DSM 11347
Bja-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Bradyrhizobium japonicum* USDA 110
Nha-1-1 to Nha-1-2	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Nitrobacter hamburgensis* X14
Nsp-2-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Nitrobacter sp.* Nb-311A
Nwi-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Nitrobacter winogradskyi* Nb-255
Oca-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Oligotropha carboxidovorans* OM5
Rpa-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* HaA2
Oan-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Brucellaceae *Ochrobactrum anthropi* ATCC 49188
Mra-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Methylobacteriaceae *Methylobacterium radiotolerans* JCM 2831

Msp-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Methylobacteriaceae *Methylobacterium sp.* 4-46
Acr-1-1 to Acr-1-2	Bacteria Proteobacteria Alphaproteobacteria Rhodospirillales Acetobacteraceae *Acidiphilium cryptum* JF-5
Gdi-1-1 to Gdi-1-4	Bacteria Proteobacteria Alphaproteobacteria Rhodospirillales Acetobacteraceae *Gluconacetobacter diazotrophicus* PAl 5
Gox-1-1 to Gox-1-3	Bacteria Proteobacteria Alphaproteobacteria Rhodospirillales Acetobacteraceae *Gluconobacter oxydans* 621H
Rru-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodospirillales Rhodospirillaceae *Rhodospirillum rubrum* ATCC 11170
Zmo-1-1	Bacteria Proteobacteria Alphaproteobacteria Sphingomonadales Sphingomonadaceae *Zymomonas mobilis subsp. mobilis* ZM4
Bam-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia ambifaria* AMMD
Bam-2-1 to Bam-2-3	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia ambifaria* IOP40-10
Bam-3-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia ambifaria* MC40-6
Bam-4-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia ambifaria* MEX-5
Bce-15-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* AU 1054
Bce-16-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* HI2424
Bce-17-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* J2315
Bce-18-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* MC0-3
Bce-19-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* PC184
Bmu-1-1 to Bmu-1-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia multivorans* ATCC 17616
Bsp-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia sp.* 383
Bvi-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia vietnamiensis* G4
Bgr-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia graminis* C4D1M
Bma-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* ATCC 10399
Bma-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* ATCC 23344
Bma-3-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* FMH
Bma-4-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* GB8 horse 4
Bma-5-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* JHU
Bma-6-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* NCTC 10229
Bma-7-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* NCTC 10247
Bma-8-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* PRL-20
Bma-9-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* SAVP1
Bph-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia phymatum* STM815
Bph-2-1 to Bph-2-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia phytofirmans* PsJN
Bsp-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia sp.* H160
Bub-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia ubonensis* Bu
Bxe-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia xenovorans* LB400
Bok-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia oklahomensis* C6786
Bok-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia oklahomensis* EO147
Bps-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1106a
Bps-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1106b
Bps-3-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 112
Bps-4-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 14
Bps-5-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1655
Bps-6-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1710a
Bps-7-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1710b
Bps-8-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 305
Bps-9-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 406e
Bps-10-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 668
Bps-11-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 7894
Bps-12-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 9
Bps-13-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 91
Bps-14-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* B7210
Bps-15-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* BCC215
Bps-16-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* DM98
Bps-17-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* K96243
Bps-18-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* NCTC 13177
Bps-19-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* Pasteur 52237
Bps-20-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* S13
Bth-5-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia thailandensis* Bt4
Bth-6-1 to Bth-6-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia thailandensis* E264
Bth-7-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia thailandensis* MSMB43
Bth-8-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia thailandensis* TXDOH
Rpi-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Ralstonia pickettii* 12D
Rpi-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Ralstonia pickettii* 12J
Rso-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Ralstonia solanacearum* GMI1000
Rso-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Ralstonia solanacearum* IPO1609
Rso-3-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Ralstonia solanacearum* MolK2
Rso-4-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Ralstonia solanacearum* UW551

Dac-1-1 to Dac-1-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae *Delftia acidovorans* SPH-1
Cvi-1-1	Bacteria Proteobacteria Betaproteobacteria Neisseriales Neisseriaceae *Chromobacterium violaceum* ATCC 12472
Neu-1-1 to Neu-1-3	Bacteria Proteobacteria Betaproteobacteria Nitrosomonadales Nitrosomonadaceae *Nitrosomonas europaea* ATCC 19718
Dol-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfobacterales Desulfobacteraceae *Desulfococcus oleovorans* Hxd3
Dpi-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfovibrionales Desulfovibrionaceae *Desulfovibrio piger* ATCC 29098
Dac-2-1	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Desulfuromonadaceae *Desulfuromonas acetoxidans* DSM 684
Sfu-1-1	Bacteria Proteobacteria Deltaproteobacteria Syntrophobacterales Syntrophobacteraceae *Syntrophobacter fumaroxidans* MPOB
Afe-1-1	Bacteria Proteobacteria Gammaproteobacteria Acidithiobacillales Acidithiobacillaceae *Acidithiobacillus ferrooxidans* ATCC 53993
Cko-1-1 to Cko-1-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Citrobacter koseri* ATCC BAA-895
Eta-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Erwinia tasmaniensis* Et1/99
Kpn-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Klebsiella pneumoniae subsp. pneumoniae* MGH 78578
Pat-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Pectobacterium atrosepticum* SCRI1043
Pca-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Pectobacterium carotovorum subsp. brasiliensis* PBR1692
Pca-2-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Pectobacterium carotovorum subsp. carotovorum* WPP14
Pal-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Providencia alcalifaciens* DSM 30120
Pru-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Providencia rustigianii* DSM 4541
Pst-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Providencia stuartii* ATCC 25827
Spr-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Serratia proteamaculans* 568
Sgl-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Sodalis glossinidius str. 'morsitans'*
Yin-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia intermedia* ATCC 29909
Ymo-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia mollaretii* ATCC 43969
Ype-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* Angola
Ype-2-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* Antiqua
Ype-9-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Antiqua str. B42003004
Ype-10-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Antiqua str. E1979001
Ype-11-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Antiqua str. UG05-0454
Ype-12-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Mediaevalis str. K1973002
Ype-13-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Microtus str. 91001
Ype-14-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Orientalis str. F1991016
Ype-15-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Orientalis str. IP275
Ype-16-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Orientalis str. MG05-1020
Ype-3-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* CA88-4125
Ype-4-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* CO92
Ype-5-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* FV-1
Ype-6-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* KIM
Ype-7-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* Nepal516
Ype-8-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* Pestoides F
Yps-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pseudotuberculosis* IP 31758
Yps-2-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pseudotuberculosis* IP 32953
Yps-3-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pseudotuberculosis* PB1/+
Yps-4-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pseudotuberculosis* YPIII
Csa-2-1	Bacteria Proteobacteria Gammaproteobacteria Oceanospirillales Halomonadaceae *Chromohalobacter salexigens* DSM 3043
Aba-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Moraxellaceae *Acinetobacter baumannii* AB0057
Aba-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Moraxellaceae *Acinetobacter baumannii* ACICU
Aba-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Moraxellaceae *Acinetobacter baumannii* ATCC 17978
Aba-5-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Moraxellaceae *Acinetobacter baumannii* AYE
Aba-6-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Moraxellaceae *Acinetobacter baumannii* SDF
Asp-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Moraxellaceae *Acinetobacter sp.* ADP1
Asp-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Moraxellaceae *Acinetobacter sp.* ATCC 27244
Pae-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* 2192
Pae-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* C3719
Pae-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PA7
Pae-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PACS2
Pae-5-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PAO1
Pae-6-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* UCBPP-PA14
Pen-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas entomophila* L48
Psy-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. maculicola*
Psy-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000
Vba-1-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales unclassified Vibrionales *Vibrionales bacterium* SWAT-3
Asa-1-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Aliivibrio salmonicida* LFI1238
Vfi-1-1 to Vfi-1-2	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae Aliivibrio *Vibrio fischeri* ES114
Vfi-2-1 to Vfi-2-2	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae Aliivibrio *Vibrio fischeri* MJ11
Ppr-1-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Photobacterium profundum* 3TCK
Ppr-2-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Photobacterium profundum* SS9

Psp-3-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Photobacterium sp.* SKA34
Val-1-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Vibrio alginolyticus* 12G01
Van-1-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Vibrio angustum* S14
Vca-1-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Vibrio campbellii* AND4
Vha-1-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Vibrio harveyi* ATCC BAA-1116
Vha-2-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Vibrio harveyi* HY01
Vsh-1-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Vibrio shilonii* AK1
Vsp-1-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Vibrio sp.* MED222
Vsp-2-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Vibrio splendidus* 12B01
Sma-1-1	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Stenotrophomonas maltophilia* K279a
Sma-2-1	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Stenotrophomonas maltophilia* R551-3
Xax-1-1	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas axonopodis pv. citri str.* 306
Xca-1-1	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas campestris pv. campestris str.* 8004
Xca-2-1	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas campestris pv. campestris str.* ATCC 33913
Xca-3-1	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas campestris pv. campestris str.* B100
Xca-4-1	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas campestris pv. vesicatoria str.* 85-10
Xor-1-1	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas oryzae pv. oryzae* KACC10331
Xor-2-1	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas oryzae pv. oryzae* MAFF 311018
Xor-3-1 to Xor-3-2	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas oryzae pv. oryzae* PXO99A
Xor-4-1	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas oryzae pv. oryzicola* BLS256
Mpi-1-1	Bacteria Thermotogae Thermotogales Thermotogaceae *Marinitoga piezophila* KA3
Tle-1-1	Bacteria Thermotogae Thermotogales Thermotogaceae *Thermotoga lettingae* TMO
Tpe-1-1	Bacteria Thermotogae Thermotogales Thermotogaceae *Thermotoga petrophila* RKU-1
Tsp-3-1	Bacteria Thermotogae Thermotogales Thermotogaceae *Thermotoga sp.* RQ2
Cfl-1-1	Bacteria Verrucomicrobia Spartobacteria *Chthoniobacter flavus* Ellin428
Min-1-1 to Min-1-2	Bacteria Verrucomicrobia unclassified Verrucomicrobia *Methylacidiphilum infernorum* V4
bEl-1-1	Bacteria Verrucomicrobia Verrucomicrobiae Verrucomicrobiales Verrucomicrobia subdivision 3 bacterium Ellin514
env-1 to env-45	environmental samples

24.2 Notes

Superscript numbers are used to annotate any *crcB* RNA that has special characteristics (described below). These numeric annotations will be used in Sections 24.3 and 24.5. The meaning of the numbers are as follows.

[1] This RNA is upstream of locus SSO1026.

24.3 Gene contexts

Each *crcB* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *crcB* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 24.4. The accession number of the sequence containing each *crcB* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 24.1. Superscript numbers refer to annotations listed in Section 24.2

abbrev.	Seq. accession		5′ at	3′ at	genes
Archaea Crenarchaeota					
[1]Sso-1-1	NC_002754.1	-	885812	885728	**RNA**→ HyfC (COG0650)→ HyfE (COG4237)→ HyfB (COG0651)→ HycE (COG3261)→ COG3260 (COG3260)→ HyfB (COG0651)→ hypo→

Mse-1-1	NC_009440.1	+	1397941	1398019	**RNA**→ HyfB (COG0651)→ HycE (COG3261)→ COG3260 (COG3260)→ HyfB (COG0651)→ HyfC (COG0650)→ HyfE (COG4237)→
Archaea Euryarchaeota					
Pto-1-1	NC_005877.1	-	847940	847870	**RNA**→ HyfC (COG0650)→ HyfE (COG4237)→ HyfB (COG0651)→ HycE (COG3261)→ COG3260 (COG3260)→ HyfB (COG0651)→
Tvo-1-1	NC_002689.2	+	669722	669784	**RNA**→ MutS (COG0249)→ MutS (COG0249)→
Mhu-1-1	NC_007796.1	+	1236994	1237053	**RNA**→ COG1689 (COG1689)→ hypo→
Tvo-1-2	NC_002689.2	+	1528463	1528534	**RNA**→ HyfC (COG0650)→ HyfE (COG4237)→ HyfB (COG0651)→ HycE (COG3261)→ COG3260 (COG3260)→ HyfB (COG0651)→
CMe-1-1	NC_009712.1	+	914945	915010	**RNA**→ enolase→ COG0218 (COG0218)→
Mhu-1-2	NC_007796.1	-	1244505	1244443	**RNA**→ ←COG1993 (COG1993)
Abo-1-1	NZ_ABSD01000048.1	+	33401	33462	**RNA**→
Abo-1-2	NZ_ABSD01000051.1	-	73291	73230	**RNA**→
Abo-1-3	NZ_ABSD01000048.1	-	169379	169321	**RNA**→
Abo-1-4	NZ_ABSD01000050.1	+	82708	82766	**RNA**→
Mhu-1-3	NC_007796.1	+	1238338	1238398	**RNA**→ Era (COG1159)→
Msm-1-1	NC_009515.1	+	1167738	1167808	**RNA**→ KefB (COG0475)→
Msm-2-1	NZ_ABYV01000006.1	-	24382	24312	**RNA**→
Msm-3-1	NZ_ABYW01000007.1	+	311285	311355	**RNA**→
Mth-1-1	NC_000916.1	-	688929	688867	**RNA**→ NhaP (COG0025)→
CMe-1-2	NC_009712.1	-	914829	914764	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→ COG1993 (COG1993)→
Tko-1-1	NC_006624.1	+	440492	440567	**RNA**→ COG1689 (COG1689)→ CrcB (COG0239)→ COG1993 (COG1993)→ hypo→ Dfp (COG0452)→ COG4087 (COG4087)→ COG1739 (COG1739)→
Ton-1-1	NC_011529.1	+	737518	737592	**RNA**→ COG1689 (COG1689)→ hypo→ hypo→ hypo→ CrcB (COG0239)→ COG1993 (COG1993)→
Tba-1-1	NZ_ABSF01000003.1	+	198437	198512	**RNA**→
Tsp-1-1	NZ_ABXN01000016.1	-	250900	250825	**RNA**→
Pab-1-1	NC_000868.1	+	690134	690214	**RNA**→ COG1689 (COG1689)→ hypo→ hypo→ COG2520 (COG2520)→
Pfu-1-1	NC_003413.1	-	1333263	1333185	**RNA**→ COG1689 (COG1689)→ hypo→
Pho-1-1	NC_000961.1	-	1278365	1278285	**RNA**→ COG1689 (COG1689)→ hypo→ hypo→ COG2520 (COG2520)→
Afu-1-1	NC_000917.1	+	1391410	1391471	**RNA**→ COG1689 (COG1689)→ HTH_ARSR (smart00418)→ DUF432 (pfam04254)→ MS_channel (pfam00924)→ hypo→
Mst-1-1	NC_007681.1	+	1672674	1672743	**RNA**→ KefB (COG0475)→
Mac-1-1	NC_003552.1	+	4831358	4831427	**RNA**→ CrcB (COG0239)→ COG1993 (COG1993)→
Mba-1-1	NC_007355.1	-	4790369	4790299	**RNA**→ CrcB (COG0239)→ COG1993 (COG1993)→ pyrophosphatase (cd00412)→
Bacteria					
Gvi-1-1	NC_005125.1	-	2366225	2366144	**RNA**→ CrcB (COG0239)→
Ssp-2-1	NC_007775.1	-	520035	519935	**RNA**→ USP_Like (cd00293)NhaP (COG0025)→
Ssp-3-1	NC_010474.1	-	84698	84619	**RNA**→ USP_Like (cd00293)NhaP (COG0025)→
Sel-1-1	NC_006576.1	-	2053114	2053041	**RNA**→ USP_Like (cd00293)NhaP (COG0025)→
Sel-2-1	NC_007604.1	+	2263916	2263989	**RNA**→ USP_Like (cd00293)NhaP (COG0025)→
Ssp-1-1	NC_007776.1	-	171311	171206	**RNA**→ USP_Like (cd00293)NhaP (COG0025)→ CrcB (COG0239)→
Ava-1-1	NC_007413.1	+	118149	118230	**RNA**→ USP_Like (cd00293)NhaP (COG0025)→
Nsp-1-1	NC_003272.1	+	2723228	2723309	**RNA**→ USP_Like (cd00293)NhaP (COG0025)→
Dth-1-1	NC_011297.1	+	51886	51949	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Tye-1-1	NC_011296.1	+	1737215	1737284	**RNA**→ KefB (COG0475)→ USP_Like (cd00293)→
Taq-1-1	NZ_ABVK01000004.1	-	157546	157484	**RNA**→ DNA mismatch repair protein MutS domain protein→
Dge-1-1	NC_008010.2	+	504027	504091	**RNA**→ CrcB (COG0239)→ COG1993 (COG1993)→ COG1993 (COG1993)→ hypo→
Aba-1-1	NC_008009.1	-	3066978	3066906	**RNA**→ CrcB (COG0239)→ Acetyltransf_1 (pfam00583)→ GEWL (cd01021)→
Hsp-1-1	NC_011126.1	-	523442	523357	**RNA**→ CrcB (COG0239)→ COG1993 (COG1993)→ COG1993 (COG1993)→ EriC (COG0038)→ Pyruvate_Kinase (cd00288)→
Hsp-1-2	NC_011126.1	+	516815	516883	**RNA**→ hypo→ ISOPREN_C2_like (cd00688)→

Bacteria Actinobacteria

Ban-1-1	NZ_ABYS01000027.1	-	262836	262758	**RNA**→
Lxy-1-1	NC_006087.1	-	68491	68420	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Mab-1-1	NC_010397.1	+	2498027	2498099	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→ Usp (pfam00582)→
Mma-1-1	NC_010612.1	-	5250157	5250088	**RNA**→ NarK (COG2223)→
Mul-1-1	NC_008611.1	-	976465	976395	**RNA**→ hypo→ NarK (COG2223)→
Mma-1-2	NC_010612.1	+	5239035	5239105	**RNA**→ hypo→
Ace-1-1	NC_008578.1	-	804142	804061	**RNA**→ HyfC (COG0650)→ hydrogenase subunit→ HyfB (COG0651)→ HycE (COG3261)Complex1_30kDa (pfam00329)→ COG3260 (COG3260)→ HyfB (COG0651)→
Mul-1-2	NC_008611.1	+	3045327	3045401	**RNA**→ COG1993 (COG1993)DUF190 (pfam02641)→ MalY (COG1168)→
Mma-1-3	NC_010612.1	+	4256142	4256216	**RNA**→ COG1993 (COG1993)DUF190 (pfam02641)→ MalY (COG1168)→
Ace-1-2	NC_008578.1	+	805841	805902	**RNA**→ MutS (COG0249)→ MutS (COG0249)→
Rxy-1-1	NC_008148.1	-	554775	554692	**RNA**→ CRCB (COG0239)→ CrcB (COG0239)→ COG1993 (COG1993)→ hypo→ hypo→ COG1993 (COG1993)→ major facilitator transporter→
Blo-3-1	NZ_ABQQ01000001.1	+	79958	80027	**RNA**→
Blo-2-1	NC_004307.2	+	925740	925809	**RNA**→ CrcB (COG0239)NOG22924 (NOG22924)→ CrcB (COG0239)→ hypo→
Blo-1-1	NC_010816.1	+	969434	969503	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→ GtrA (pfam04138)→
Blo-1-2	NZ_AABM02000006.1	+	7959	8028	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→ GtrA (pfam04138)→
Bad-1-1	NC_008618.1	+	922963	923040	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Bad-2-1	NZ_AAXD02000028.1	-	152193	152116	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Nfa-1-1	NC_006363.1	-	61342	61253	**RNA**→ ←putative transposase

Bacteria Bacteroidetes

Bpl-1-1	NZ_ABQC02000019.1	+	404312	404380	**RNA**→ EriC (COG0038)→
Bco-1-1	NZ_ABIY02000058.1	+	8984	9052	**RNA**→ EriC (COG0038)→
Bth-1-1	NC_004663.1	-	5995735	5995673	**RNA**→ CrcB (COG0239)→ COG1357 (COG1357)→
Bca-1-1	NZ_AAXG02000029.1	+	31858	31932	**RNA**→ Na_H_Exchanger (pfam00999)→
Lbl-1-1	NZ_AANC01000009.1	-	2226	2162	**RNA**→ EriC (COG0038)→
Psp-1-1	NZ_ABCM01000001.1	-	151944	151879	**RNA**→ ←hypo
Psp-1-2	NZ_ABCM01000022.1	+	54798	54866	**RNA**→ PAP2_like_4 (cd03395)→ COG2510 (COG2510)→ enolase (cd03313)→ EriC (COG0038)→

Bacteria Chloroflexi

Rsp-1-1	NC_009523.1	+	3080181	3080241	**RNA**→ CrcB (COG0239)→ COG1993 (COG1993)COG2524 (COG2524)CBS (pfam00571)→
Cau-1-1	NC_010175.1	+	429462	429523	**RNA**→ ABC_MutS_homologs (cd03243)→ ABC_MutS_homologs (cd03243)→
Csp-1-1	NZ_ABPG01000010.1	+	170766	170827	**RNA**→ ABC_MutS_homologs (cd03243)→ ABC_MutS_homologs (cd03243)→
Cag-1-1	NZ_AAUI01000037.1	+	30963	31024	**RNA**→ Acetyltransf_1 (pfam00583)→

Bacteria Firmicutes

Ebi-1-1	NZ_ABYT01000096.1	-	44476	44400	**RNA**→

Bacteria Firmicutes Bacillales

Gsp-2-1	NZ_ABLJ01000009.1	-	43939	43880	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→ COG2966 (COG2966)→ COG3610 (COG3610)→
Gth-1-1	NC_009328.1	-	2739339	2739277	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Gsp-1-1	NZ_ABVH01000009.1	-	150338	150276	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Gka-1-1	NC_006510.1	-	2726409	2726347	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Bth-2-1	NZ_AAJM01000092.1	-	4654	4595	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Bce-10-1	NZ_ABDJ01000008.1	-	214042	213983	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Bwe-1-1	NC_010184.1	+	4869306	4869365	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Bce-13-1	NZ_ABCZ02000002.1	-	119714	119655	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Bce-12-1	NZ_ABDK02000007.1	-	216428	216369	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Ban-16-1	NC_005945.1	+	4825068	4825127	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Ban-3-1	NC_007530.2	+	4823859	4823918	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Ban-12-1	NC_003997.3	+	4823733	4823792	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→

Ban-11-1	NZ_AAAC02000001.1	+	165548	165607	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Ban-14-1	NZ_AAEN01000018.1	-	235474	235415	**RNA**→
Ban-10-1	NZ_AAEO01000027.1	-	235458	235399	**RNA**→
Ban-17-1	NZ_AAEP01000006.1	-	11056	10997	**RNA**→
Ban-15-1	NZ_AAEQ01000042.1	+	19179	19238	**RNA**→
Ban-18-1	NZ_AAER01000034.1	-	116308	116249	**RNA**→
Ban-13-1	NZ_AAES01000020.1	+	17811	17870	**RNA**→
Ban-2-1	NZ_ABDN02000026.1	+	19393	19452	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Ban-9-1	NZ_ABJC01000004.1	+	17851	17910	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Ban-5-1	NZ_ABKF01000039.1	-	21440	21381	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Ban-7-1	NZ_ABKG01000033.1	+	18847	18906	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Ban-6-1	NZ_ABLB01000039.1	+	17548	17607	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Ban-8-1	NZ_ABLH01000033.1	+	18167	18226	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Ban-4-1	NZ_ABLT01000037.1	-	10125	10066	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Bth-3-1	NC_005957.1	+	4836662	4836721	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Bce-9-1	NZ_AAEK01000016.1	-	81808	81749	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Bce-8-1	NC_006274.1	+	4897977	4898036	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Bce-4-1	NZ_AAUE01000005.1	-	214692	214633	**RNA**→ CrcB (COG0239)→
Bce-7-1	NZ_ABDI01000008.1	-	206570	206511	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Bce-3-1	NZ_AAUF01000015.1	-	62117	62058	**RNA**→ CrcB (COG0239)→
Bce-11-1	NZ_ABDL02000014.1	-	72280	72221	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Bce-2-1	NZ_ABDA02000004.1	+	142970	143029	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Bth-4-1	NC_008600.1	+	4850198	4850257	**RNA**→ ←hypo
Bce-1-1	NZ_ABDM02000031.1	-	13426	13367	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Bce-5-1	NC_003909.8	+	4815585	4815644	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Bce-6-1	NC_004722.1	+	4975681	4975742	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Bce-14-1	NC_009674.1	+	3708919	3708978	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Gsp-3-1	NZ_ABRG01000006.1	-	47838	47778	**RNA**→ EriC (COG0038)→
Psp-2-1	NZ_ABKS01000027.1	+	41421	41481	**RNA**→ EriC (COG0038)→

Bacteria Firmicutes Clostridia

Mth-2-1	NC_007644.1	+	544530	544593	**RNA**→ CrcB (COG0239)→ camphor resistance protein CrcB→ COG1993 (COG1993)→ tRNA-Gly→ tRNA-Gly→
Tte-1-1	NC_003869.1	-	1840648	1840588	**RNA**→ KefB (COG0475)→ Usp (pfam00582)→ Era (COG1159)→
Cpa-1-1	NZ_ABXP01000064.1	-	2744	2684	**RNA**→
Cpa-1-2	NZ_ABXP01000064.1	-	2857	2798	**RNA**→
Tte-1-2	NC_003869.1	-	1840761	1840702	**RNA**→ KefB (COG0475)→ Usp (pfam00582)→ Era (COG1159)→
Cac-1-1	NC_003030.1	+	1730397	1730460	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Cdi-5-1	NZ_AAML04000007.1	+	96644	96720	**RNA**→ EriC (COG0038)→ NhaP (COG0025)→
Cdi-8-1	NZ_ABFD01000027.1	+	11837	11913	**RNA**→
Cdi-10-1	NZ_ABHF01000064.1	+	26742	26818	**RNA**→ EriC (COG0038)→ NhaP (COG0025)→
Cdi-7-1	NZ_ABHD01000009.1	-	19917	19841	**RNA**→ EriC (COG0038)→ NhaP (COG0025)→
Cdi-4-1	NZ_ABKL01000011.1	-	19174	19098	**RNA**→ EriC (COG0038)→
Cdi-9-1	NZ_ABHE01000037.1	+	26741	26817	**RNA**→ EriC (COG0038)→ NhaP (COG0025)→
Cdi-3-1	NZ_ABKK01000024.1	-	20543	20467	**RNA**→ EriC (COG0038)→ NhaP (COG0025)→
Cdi-2-1	NZ_ABKJ01000031.1	+	18054	18130	**RNA**→ EriC (COG0038)→ NhaP (COG0025)→
Cdi-6-1	NZ_ABHG01000008.1	-	19941	19865	**RNA**→ EriC (COG0038)→ NhaP (COG0025)→
Cdi-1-1	NC_009089.1	+	782179	782255	**RNA**→ EriC (COG0038)→ NhaP (COG0025)→
Cbu-1-1	NZ_ABDT01000004.2	-	28954	28883	**RNA**→ EriC (COG0038)→
Nth-1-1	NC_010718.1	-	285558	285483	**RNA**→ NhaP (COG0025)→
Aor-1-1	NC_009922.1	+	1242223	1242295	**RNA**→ Na_H_Exchanger (pfam00999)→
Ceu-1-1	NZ_ABEY02000020.1	+	92291	92375	**RNA**→ NhaP (COG0025)→ COG0637 (COG0637)→
Cba-1-1	NZ_ABQR01000024.1	-	52165	52090	**RNA**→

Rob-1-1	NZ_AAVO02000004.1	+	13914	13993	**RNA**→ NhaP (COG0025)→ COG3467 (COG3467)→
Dlo-1-1	NZ_AAXB02000001.1	+	235634	235709	**RNA**→ NhaP (COG0025)→
Ahy-1-1	NZ_ABXA01000043.1	+	134014	134086	**RNA**→
Cph-1-1	NC_010001.1	-	3806199	3806127	**RNA**→ NhaP (COG0025)→ MTHFR (cd00537)→ Acm (COG3757)→
Cbo-1-1	NZ_ABCC02000039.1	+	237860	237932	**RNA**→ NhaP (COG0025)→ CN_hydrolase (pfam00795)→ REC (cd00156)BaeS (COG0642)AtoS (COG2202)→
Fma-1-1	NC_010376.1	-	1641037	1640963	**RNA**→ Na_H_Exchanger (pfam00999)→
Ala-1-1	NZ_ABYO01000015.1	+	56802	56875	**RNA**→
Chi-1-1	NZ_ABWP01000024.1	-	12375	12304	**RNA**→ NhaP (COG0025)→
Rin-1-1	NZ_ABYJ01000311.1	-	4868	4795	**RNA**→
Eve-1-1	NZ_AAVL02000037.1	+	287218	287292	**RNA**→ EriC (COG0038)→
Csp-2-1	NZ_ABGC03000023.1	+	52156	52229	**RNA**→ EriC (COG0038)→
Csa-1-1	NC_009437.1	+	751276	751339	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Csa-1-2	NC_009437.1	+	782565	782643	**RNA**→ hypo→
Tps-1-1	NC_010321.1	-	2271016	2270957	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→ COG2206 (COG2206)→ hypo→ sigmaK-factor processing regulatory BofA→ hypo→
Tsp-2-1	NC_010320.1	+	40525	40583	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Tsp-2-2	NC_010320.1	+	1156972	1157031	**RNA**→ hypo→ hypo→
Tps-1-2	NC_010321.1	+	673881	673940	**RNA**→ hypo→
Cpa-1-3	NZ_ABXP01000064.1	-	9651	9591	**RNA**→
Tte-1-3	NC_003869.1	-	1847557	1847497	**RNA**→ hypo→ hypo→ hypo→ AraJ (COG2814)→
Tca-1-1	NZ_AAWL01000036.1	+	2586	2647	**RNA**→ Era (COG1159)→ KefB (COG0475)→ USP_Like (cd00293)→ COG1993 (COG1993)→
Ame-1-1	NC_009633.1	-	4518368	4518305	**RNA**→ NhaP (COG0025)→ MazG (COG1694)→ COG2846 (COG2846)→
Dha-2-1	NC_007907.1	+	2441450	2441510	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Dha-1-1	NZ_AAAW04000001.1	+	623601	623661	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Ckl-1-1	NC_009706.1	+	304406	304465	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→ COG1993 (COG1993)→
Cbo-5-1	NC_010520.1	-	3494982	3494923	**RNA**→ hypo→ COG3481 (COG3481)→
Cbo-11-1	NZ_ABDO02000002.1	+	363051	363110	**RNA**→ hypo→ COG3481 (COG3481)→
Cbo-3-1	NC_009495.1	-	3416950	3416891	**RNA**→ hypo→
Cbo-2-1	NC_009697.1	-	3392939	3392880	**RNA**→ hypo→
Cbo-4-1	NC_009698.1	-	3290046	3289987	**RNA**→ hypo→
Cbo-7-1	NC_010516.1	-	3485265	3485206	**RNA**→ hypo→ COG3481 (COG3481)→
Cbo-10-1	NC_009699.1	-	3509923	3509864	**RNA**→ hypo→ COG3481 (COG3481)→
Cpe-7-1	NZ_ABDY01000003.1	-	8838	8779	**RNA**→ EriC (COG0038)→
Cpe-1-1	NC_008261.1	+	678979	679038	**RNA**→ EriC (COG0038)→
Cpe-2-1	NZ_ABDV01000012.1	+	79943	80002	**RNA**→ EriC (COG0038)→
Cpe-6-1	NZ_ABDW01000033.1	-	8933	8874	**RNA**→ EriC (COG0038)→
Cpe-5-1	NZ_ABOO01000041.1	-	8519	8460	**RNA**→ EriC (COG0038)→
Cpe-8-1	NC_008262.1	+	656850	656909	**RNA**→ EriC (COG0038)→
Cpe-4-1	NZ_ABDX01000022.1	-	8803	8744	**RNA**→ EriC (COG0038)→
Cpe-3-1	NZ_ABDU01000025.2	+	239794	239853	**RNA**→ EriC (COG0038)→
Cpe-9-1	NC_003366.1	+	731613	731672	**RNA**→ EriC (COG0038)→
Cbo-9-1	NC_010723.1	-	1342444	1342385	**RNA**→ EriC (COG0038)→
Cbo-3-2	NC_009495.1	-	1975371	1975313	**RNA**→ NhaP (COG0025)→
Cbo-2-2	NC_009697.1	-	1904011	1903953	**RNA**→ NhaP (COG0025)→
Cbo-4-2	NC_009698.1	-	1904496	1904438	**RNA**→ NhaP (COG0025)→
Cbo-7-2	NC_010516.1	-	1975401	1975343	**RNA**→ NhaP (COG0025)→
Cbo-10-2	NC_009699.1	-	1991940	1991882	**RNA**→ NhaP (COG0025)→
Cbo-8-1	NZ_ABDP01000020.1	-	8552	8494	**RNA**→ transporter, monovalent cation:proton antiporter-2 (CPA2) family→
Cbo-5-2	NC_010520.1	-	2040077	2040019	**RNA**→ Na_H_Exchanger (pfam00999)→
Cbo-11-2	NZ_ABDO02000006.1	-	211703	211645	**RNA**→ NhaP (COG0025)→
Csp-3-1	NZ_ABKW02000004.1	+	1354900	1354958	**RNA**→ NhaP (COG0025)→
Cbo-6-1	NC_010674.1	+	2720217	2720276	**RNA**→ EriC (COG0038)→
Cth-1-1	NC_009012.1	+	234427	234485	**RNA**→ NhaP (COG0025)→ hypo→

Cth-2-1	NZ_ABVG01000004.1	+	46786	46844	**RNA**→ NhaP (COG0025)→
Hmo-1-1	NC_010337.1	-	1203106	1203045	**RNA**→ IscU (COG0822)→ hypo→ CrcB (COG0239)→ CrcB (COG0239)→ COG1993 (COG1993)→ hypo→
Pth-1-1	NC_009454.1	+	615295	615357	**RNA**→ Era (COG1159)→
Chy-1-1	NC_007503.1	-	1888854	1888795	**RNA**→ KefB (COG0475)→ hypo→ GTP_EFTU (pfam00009)→
Bacteria Firmicutes Lactobacillales					
Efa-2-1	NC_004668.1	-	2384514	2384445	**RNA**→ EriC (COG0038)→ hypo→ IMPase_like (cd01637)→
Efa-1-1	NZ_ABPI01000001.1	-	1998814	1998745	**RNA**→
Lrh-1-1	NZ_ABWJ01000039.1	+	17614	17668	**RNA**→ CrcB (COG0239)→ Integral membrane protein for chromosome condensation→
Lpa-1-1	NZ_ABQV01000049.1	-	12339	12285	**RNA**→
Lca-1-1	NC_008526.1	-	2429871	2429817	**RNA**→ CrcB (COG0239)→ integral membrane protein for chromosome condensation→
Lca-2-1	NC_010999.1	-	2570801	2570747	**RNA**→ CrcB (COG0239)→ Protein crcB homolog→
Csp-4-1	NZ_ABHH01000009.1	-	50257	50182	**RNA**→ Na_H_Exchanger (pfam00999)→ EriC (COG0038)→
Lac-1-1	NC_006814.3	+	965933	965991	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Lhe-1-1	NC_010080.1	+	1061364	1061422	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
Lga-1-1	NC_008530.1	+	283109	283169	**RNA**→ ←hypo
Lga-2-1	NZ_ABWH01000036.1	+	18440	18500	**RNA**→
Ljo-1-1	NC_005362.1	-	1172633	1172571	**RNA**→ enolase (cd03313)→
Lga-1-2	NC_008530.1	-	1279279	1279217	**RNA**→ Cof (COG0561)→ CrcB (COG0239)→ CrcB (COG0239)→
Lga-2-2	NZ_ABWH01000014.1	-	31911	31849	**RNA**→
Ljo-1-2	NC_005362.1	+	798633	798695	**RNA**→ Cof (COG0561)→ CrcB (COG0239)→ CrcB (COG0239)→
Lpl-1-1	NC_004567.1	-	1741235	1741170	**RNA**→ enolase (cd03313)→ MgtA (COG0474)→
Efa-3-1	NZ_AAAK03000106.1	+	3476	3542	**RNA**→ EriC (COG0038)→
Lca-2-2	NC_010999.1	+	2571522	2571586	**RNA**→ Enolase 2 (2-phosphoglycerate dehydratase 2) (2-phospho-D-glycerate hydro-lyase 2)→
Lpa-1-2	NZ_ABQV01000049.1	+	13062	13126	**RNA**→
Lca-1-2	NC_008526.1	+	2430592	2430656	**RNA**→ hypo→
Lla-3-1	NC_002662.1	+	1105212	1105286	**RNA**→ EriC (COG0038)→
Lla-1-1	NC_009004.1	-	1450312	1450238	**RNA**→ EriC (COG0038)→
Lla-2-1	NC_008527.1	+	1111074	1111148	**RNA**→ EriC (COG0038)→
Lla-3-2	NC_002662.1	+	274025	274097	**RNA**→ enolase (cd03313)→
Lla-2-2	NC_008527.1	+	262409	262481	**RNA**→ enolase (cd03313)→
Lla-1-2	NC_009004.1	+	272338	272410	**RNA**→ ←putative secreted protein
Bacteria Proteobacteria Alphaproteobacteria					
Zmo-1-1	NC_006526.1	-	542156	542079	**RNA**→ EriC (COG0038)→
Gox-1-1	NC_006677.1	+	2355903	2355981	**RNA**→ EriC (COG0038)→
Gdi-1-1	NC_010125.1	-	3849897	3849819	**RNA**→ EriC (COG0038)→
Gdi-1-2	NC_011365.1	-	2909225	2909147	**RNA**→ EriC (COG0038)→
Gdi-1-3	NC_010125.1	-	3856660	3856597	**RNA**→ MutS (COG0249)→ ABC_MutS_homologs (cd03243)→
Gdi-1-4	NC_011365.1	-	2915994	2915931	**RNA**→ MutS (COG0249)→ ABC_MutS_homologs (cd03243)→
Nha-1-1	NC_007964.1	-	1049491	1049431	**RNA**→ ←hypo
Gox-1-2	NC_006672.1	-	152206	152146	**RNA**→ UspA (COG0589)→ hypo→
Oan-1-1	NC_009668.1	+	989075	989138	**RNA**→ CrcB (COG0239)→
Gox-1-3	NC_006672.1	+	152384	152449	**RNA**→ CrcB (COG0239)→ COG2510 (COG2510)→ hypo→
Mra-1-1	NC_010505.1	+	4759637	4759698	**RNA**→ CrcB (COG0239)→
Bja-1-1	NC_004463.1	-	2887392	2887315	**RNA**→ UhpC (COG2271)→ CrcB (COG0239)→ CrcB (COG0239)→ COG1993 (COG1993)→ ClpS (pfam02617)→
Oca-1-1	NC_011386.1	+	1408427	1408488	**RNA**→ CrcB (COG0239)→
Nwi-1-1	NC_007406.1	+	166253	166313	**RNA**→ CrcB (COG0239)→
Nha-1-2	NC_007964.1	-	237736	237676	**RNA**→ CrcB (COG0239)→
Nsp-2-1	NZ_AAMY01000002.1	+	145285	145345	**RNA**→ hypo→ CrcB (COG0239)→
Rpa-1-1	NC_007778.1	+	3360939	3360999	**RNA**→ hypo→
Rru-1-1	NC_007643.1	+	293075	293134	**RNA**→ CrcB (COG0239)→ COG1993 (COG1993)→ hypo→ hypo→ COG1051 (COG1051)→

Msp-1-1	NC_010511.1	+	4458865	4458926	**RNA**→ Na+/H+ antiporter→
Acr-1-1	NC_009484.1	-	351714	351653	**RNA**→ CrcB (COG0239)→ COG1993 (COG1993)→
Acr-1-2	NC_009484.1	+	677221	677283	**RNA**→ HyfC (COG0650)→
Bacteria Proteobacteria Betaproteobacteria					
Neu-1-1	NC_004757.1	+	2158790	2158856	**RNA**→ AbrB (COG2002)→ hypo→
Dac-1-1	NC_010002.1	-	1926512	1926446	**RNA**→ EriC (COG0038)→ hypo→ hypo→
Cvi-1-1	NC_005085.1	-	2613318	2613254	**RNA**→ EriC (COG0038)→
Dac-1-2	NC_010002.1	+	3908908	3908971	**RNA**→ CrcB (COG0239)→
Neu-1-2	NC_004757.1	+	1756266	1756341	**RNA**→ VacB (COG0557)→ TonB (COG0810)→ AroE (COG0169)→ Transgly (pfam00912)→
Neu-1-3	NC_004757.1	-	1838968	1838902	**RNA**→ CrcB (COG0239)→ COG1993 (COG1993)NOG18330 (NOG18330)→
Bph-2-1	NC_010681.1	-	4147651	4147580	**RNA**→ EriC (COG0038)→ hypo→
Bxe-1-1	NC_007951.1	-	1757402	1757341	**RNA**→ CRCB (COG0239)→
Bph-2-2	NC_010681.1	-	1800329	1800268	**RNA**→ CrcB (COG0239)→
Bph-1-1	NC_010622.1	+	2012557	2012617	**RNA**→ CrcB (COG0239)→
Bgr-1-1	NZ_ABLD01000041.1	+	16968	17073	**RNA**→ CrcB (COG0239)→ PRC (pfam05239)→
Bsp-2-1	NZ_ABYL01000251.1	-	1837	1776	**RNA**→ CrcB (COG0239)→
Bub-1-1	NZ_ABBE01000034.1	+	547	651	**RNA**→ EriC (COG0038)→
Rpi-2-1	NC_010678.1	+	247334	247411	**RNA**→ EriC (COG0038)→
Rpi-1-1	NZ_ABDZ01000002.1	+	179749	179826	**RNA**→ EriC (COG0038)→
Bmu-1-1	NC_010086.1	-	2066773	2066667	**RNA**→ EriC (COG0038)→
Bmu-1-2	NC_010805.1	+	408531	408637	**RNA**→ EriC (COG0038)→
Rso-1-1	NC_003295.1	-	1941892	1941826	**RNA**→ EriC (COG0038)→
Rso-3-1	NW_002196564.1	+	565243	565309	**RNA**→ EriC (COG0038)→
Rso-2-1	NW_002196568.1	+	1142118	1142184	**RNA**→ EriC (COG0038)→
Rso-4-1	NZ_AAKL01000043.1	-	37416	37350	**RNA**→ EriC (COG0038)→
Bce-19-1	NZ_AAKX01000112.1	-	6305	6201	**RNA**→ EriC (COG0038)→ hypo→
Bce-18-1	NC_010515.1	-	639633	639529	**RNA**→ EriC (COG0038)→ hypo→
Bce-15-1	NC_008061.1	-	1826897	1826793	**RNA**→ EriC (COG0038)→ hypo→
Bce-16-1	NC_008543.1	+	654224	654328	**RNA**→ EriC (COG0038)→ hypo→
Bam-2-1	NZ_ABLC01000125.1	-	14642	14539	**RNA**→ EriC (COG0038)→
Bam-3-1	NC_010552.1	+	643438	643541	**RNA**→ EriC (COG0038)→
Bam-1-1	NC_008391.1	+	2589399	2589502	**RNA**→ EriC (COG0038)→ SK (cd00464)→ hypo→
Bam-4-1	NZ_ABLK01000079.1	+	2858	2961	**RNA**→ hypo→ EriC (COG0038)→
Bth-6-1	NC_007650.1	+	603169	603235	**RNA**→ EriC (COG0038)→
Bth-6-2	NZ_AACX01000033.1	+	18652	18718	**RNA**→
Bth-5-1	NZ_ABBH01000052.1	+	1772	1838	**RNA**→ EriC (COG0038)→
Bps-14-1	NZ_ABBN01001252.1	-	1464	1398	**RNA**→ EriC (COG0038)→
Bps-3-1	NZ_ABBP01001115.1	-	6777	6711	**RNA**→ EriC (COG0038)→
Bps-7-1	NC_007435.1	-	1222860	1222794	**RNA**→ EriC (COG0038)→
Bps-6-1	NZ_AAHS03000001.1	+	321518	321584	**RNA**→ EriC (COG0038)→
Bps-4-1	NZ_ABBJ01001663.1	-	1415	1349	**RNA**→ EriC (COG0038)→
Bps-12-1	NZ_ABBL01001544.1	-	1458	1392	**RNA**→ EriC (COG0038)→
Bps-5-1	NZ_AAHR02000024.1	+	6414	6480	**RNA**→ EriC (COG0038)→
Bok-2-1	NZ_ABBF01000636.1	-	6946	6880	**RNA**→ EriC (COG0038)→
Bps-17-1	NC_006351.1	-	2539005	2538939	**RNA**→ EriC (COG0038)→
Bps-19-1	NZ_AAHV02000023.1	+	4311	4377	**RNA**→ EriC (COG0038)→
Bps-13-1	NZ_ABBK01001514.1	-	879	813	**RNA**→ voltage-gated chloride channel→
Bth-7-1	NZ_ABBM01000878.1	-	872	806	**RNA**→ voltage-gated chloride channel→
Bok-1-1	NZ_ABBG01000437.1	-	4831	4765	**RNA**→ EriC (COG0038)→
Bps-16-1	NZ_ABBI01002081.1	-	1424	1358	**RNA**→ EriC (COG0038)→
Bps-10-1	NC_009075.1	-	2566222	2566156	**RNA**→ EriC (COG0038)→

Bps-9-1	NZ_AAMM02000004.1	+	6108	6174	**RNA**→ EriC (COG0038)→
Bps-11-1	NZ_ABBO01001356.1	-	911	845	**RNA**→ voltage-gated chloride channel→
Bps-15-1	NZ_ABBR01000883.1	-	6870	6804	**RNA**→ ←hypo
Bma-2-1	NC_006349.2	+	224010	224076	**RNA**→ ←hypo
Bma-9-1	NC_008784.1	+	1388196	1388262	**RNA**→ EriC (COG0038)→ hypo→
Bma-7-1	NC_009079.1	+	225172	225238	**RNA**→ EriC (COG0038)→
Bma-6-1	NC_008835.1	+	1653068	1653134	**RNA**→ EriC (COG0038)→ hypo→
Bma-1-1	NZ_AAHN02000008.1	+	103741	103807	**RNA**→ EriC (COG0038)→
Bma-4-1	NZ_AAHO01000134.1	+	6134	6200	**RNA**→ COG0038: Chloride channel protein EriC→
Bps-20-1	NZ_AAHW02000001.1	-	301150	301084	**RNA**→ EriC (COG0038)→
Bma-3-1	NZ_AAIQ02000003.1	+	20585	20651	**RNA**→ EriC (COG0038)→
Bma-5-1	NZ_AAIR02000026.1	+	4415	4481	**RNA**→ EriC (COG0038)→
Bma-8-1	NZ_AAZP01000053.1	-	29998	29932	**RNA**→ ←hypo
Bps-1-1	NC_009078.1	-	2493251	2493185	**RNA**→ EriC (COG0038)→
Bps-2-1	NZ_AAMB02000003.1	+	113680	113746	**RNA**→ ←hypo
Bps-8-1	NZ_AAYX01000002.1	-	364269	364203	**RNA**→ EriC (COG0038)→
Bps-18-1	NZ_ABBQ01000905.1	-	4347	4281	**RNA**→ ←hypo
Bth-8-1	NZ_ABBD01000040.1	+	1808	1874	**RNA**→ EriC (COG0038)→
Bsp-1-1	NC_007511.1	-	2629309	2629206	**RNA**→ EriC (COG0038)→ SK (cd00464)NOG33876 (NOG33876)→
Bce-17-1	NC_011001.1	+	913754	913857	**RNA**→ EriC (COG0038)→ hypo→
Bvi-1-1	NC_009255.1	+	985285	985390	**RNA**→ EriC (COG0038)→
Bam-2-2	NZ_ABLC01000046.1	-	24103	24037	**RNA**→ hypo→
Bam-2-3	NZ_ABLC01000083.1	+	20871	20937	**RNA**→ ←hypo
Bacteria Proteobacteria Deltaproteobacteria					
Sfu-1-1	NC_008554.1	-	3591388	3591322	**RNA**→ TEF1 (COG5256)→
Dpi-1-1	NZ_ABXU01000089.1	+	23581	23642	**RNA**→
Dol-1-1	NC_009943.1	+	3573766	3573832	**RNA**→ CrcB (COG0239)→
Dac-2-1	NZ_AAEW02000001.1	-	156146	156076	**RNA**→ EriC (COG0038)→ NADPH_oxidoreductase_1 (cd02150)→
Bacteria Proteobacteria Gammaproteobacteria					
Xca-1-1	NC_007086.1	+	2656872	2656947	**RNA**→ EriC (COG0038)→
Xca-3-1	NC_010688.1	-	2646740	2646665	**RNA**→ EriC (COG0038)→
Xca-2-1	NC_003902.1	-	2322664	2322589	**RNA**→ EriC (COG0038)→
Xor-1-1	NC_006834.1	+	2682944	2683020	**RNA**→ EriC (COG0038)→
Xor-2-1	NC_007705.1	+	2662419	2662495	**RNA**→ EriC (COG0038)→ hypo→ CrcB (COG0239)→ COG1993 (COG1993)→
Xor-3-1	NC_010717.1	+	2555457	2555533	**RNA**→ EriC (COG0038)→
Xor-3-2	NC_010717.1	+	2767544	2767620	**RNA**→ EriC (COG0038)→
Xor-4-1	NZ_AAQN01000001.1	-	2446869	2446793	**RNA**→ EriC (COG0038)→ hypo→
Ppr-1-1	NZ_AAPH01000002.1	+	303167	303249	**RNA**→ ←hypo
Ppr-2-1	NC_006370.1	-	1717779	1717696	**RNA**→ ←hypo
Psp-3-1	NZ_AAOU01000011.1	-	78540	78473	**RNA**→ IscU (COG0822)→ hypo→
Van-1-1	NZ_AAOJ01000002.1	+	263194	263261	**RNA**→ IscU (COG0822)→ hypo→
Eta-1-1	NC_010694.1	-	2891609	2891537	**RNA**→ CrcB (COG0239)→
Cko-1-1	NC_009792.1	-	1490183	1490109	**RNA**→ Usp (pfam00582)→ enolase (cd03313)→
Vfi-2-1	NC_011184.1	+	34842	34908	**RNA**→ EriC (COG0038)→ hypo→
Vfi-1-1	NC_006840.2	+	35443	35510	**RNA**→ EriC (COG0038)→
Sgl-1-1	NC_007712.1	-	1344619	1344556	**RNA**→ ←putative phage DNA adenine methylase
Csa-2-1	NC_007963.1	+	2902118	2902253	**RNA**→ CrcB (COG0239)→ COG1993 (COG1993)→
Vca-1-1	NZ_ABGR01000008.1	-	95568	95493	**RNA**→ IscU (COG0822)→ hypo→
Vha-2-1	NZ_AAWP01000027.1	+	44793	44868	**RNA**→ IscU (COG0822)→ hypo (cons)→
Vha-1-1	NC_009783.1	+	3337036	3337110	**RNA**→ IscU (COG0822)→ hypo→

Val-1-1	NZ_AAPS01000002.1	-	29498	29423	**RNA**→ IscU (COG0822)→ hypo→
Afe-1-1	NC_011206.1	+	2508708	2508771	**RNA**→ CrcB (COG0239)→ COG1993 (COG1993)→ COG1689 (COG1689)→
Sma-1-1	NC_010943.1	-	1735698	1735635	**RNA**→ CrcB (COG0239)→
Sma-2-1	NC_011071.1	-	1611287	1611224	**RNA**→ CrcB (COG0239)→
Vsp-1-1	NZ_AAND01000003.1	-	123453	123370	**RNA**→ IscU (COG0822)→ hypo→
Vba-1-1	NZ_AAZW01000073.1	-	10424	10341	**RNA**→ IscU (COG0822)→
Vsp-2-1	NZ_AAMR01000022.1	-	13784	13701	**RNA**→ IscU (COG0822)→ hypo→
Vfi-1-2	NC_006841.2	-	1258800	1258735	**RNA**→ IscU (COG0822)→ hypo→
Vfi-2-2	NC_011186.1	-	1365444	1365379	**RNA**→ IscU (COG0822)→ hypo→
Asa-1-1	NC_011313.1	-	1142244	1142180	**RNA**→ IscU (COG0822)→ hypo→
Vsh-1-1	NZ_ABCH01000002.1	+	2875	2939	**RNA**→ IscU (COG0822)→ hypo→
Pca-1-1	NZ_ABVX01000036.1	-	35509	35405	**RNA**→
Pca-2-1	NZ_ABVY01000154.1	-	6708	6604	**RNA**→
Psy-2-1	NC_007005.1	-	1675154	1675082	**RNA**→ EriC (COG0038)→ hypo→ COG1993 (COG1993)→ hypo→
Psy-3-1	NC_004578.1	-	5215709	5215637	**RNA**→ EriC (COG0038)→
Cko-1-2	NC_009792.1	+	1484941	1485003	**RNA**→ MutS (COG0249)→ MutS (COG0249)→
Psy-1-1	NC_002759.1	+	22906	22967	**RNA**→ CrcB (COG0239)→
Pae-3-1	NC_009656.1	-	5109630	5109568	**RNA**→ CrcB (COG0239)→
Pae-6-1	NC_008463.1	-	5078189	5078127	**RNA**→ CrcB (COG0239)→ COG1993 (COG1993)→
Pae-5-1	NC_002516.2	-	4914637	4914575	**RNA**→ CrcB (COG0239)→
Pae-1-1	NZ_AAKW01000074.1	+	187514	187576	**RNA**→ CrcB (COG0239)→ COG1993 (COG1993)→
Pae-2-1	NZ_AAKV01000114.1	+	24659	24721	**RNA**→ CrcB (COG0239)→ COG1993 (COG1993)→
Pae-4-1	NZ_AAQW01000001.1	+	1175247	1175309	**RNA**→ CrcB (COG0239)→ COG1993 (COG1993)→
Pen-1-1	NC_008027.1	-	3719897	3719836	**RNA**→ CrcB (COG0239)→
Xca-4-1	NC_007508.1	+	2350991	2351059	**RNA**→ CrcB (COG0239)→ hypo→
Xax-1-1	NC_003919.1	+	2352963	2353030	**RNA**→ CrcB (COG0239)→
Pst-1-1	NZ_ABJD02000101.1	+	724041	724112	**RNA**→ CrcB (COG0239)→
Asp-2-1	NZ_ABYN01000048.1	+	132548	132614	**RNA**→
Asp-1-1	NC_005966.1	+	406046	406112	**RNA**→ CrcB (COG0239)→
Aba-3-1	NC_010611.1	+	439864	439930	**RNA**→ CrcB (COG0239)→
Aba-4-1	NC_009085.1	+	421777	421843	**RNA**→ CrcB (COG0239)→
Aba-6-1	NC_010400.1	-	2879674	2879608	**RNA**→ CrcB (COG0239)→
Aba-5-1	NC_010410.1	-	3434387	3434321	**RNA**→ CrcB (COG0239)→
Aba-2-1	NZ_ABJM01000001.1	+	502200	502266	**RNA**→
Ymo-1-1	NZ_AALD01000017.1	-	44340	44289	**RNA**→ CrcB (COG0239)→
Yin-1-1	NZ_AALF01000008.1	-	82054	81992	**RNA**→ CrcB (COG0239)→
Yps-4-1	NC_010465.1	+	3331346	3331409	**RNA**→ CrcB (COG0239)→
Ype-13-1	NC_005810.1	+	1200714	1200777	**RNA**→ CrcB (COG0239)→
Ype-1-1	NC_010159.1	+	1924705	1924768	**RNA**→ CrcB (COG0239)→
Ype-2-1	NC_008150.1	+	2792665	2792728	**RNA**→ CrcB (COG0239)→
Ype-7-1	NC_008149.1	-	1244276	1244213	**RNA**→ CrcB (COG0239)→
Ype-8-1	NC_009381.1	+	3010272	3010335	**RNA**→ CrcB (COG0239)→
Ype-6-1	NC_004088.1	-	1314365	1314302	**RNA**→ CrcB (COG0239)→
Ype-4-1	NC_003143.1	-	2918233	2918170	**RNA**→ CrcB (COG0239)→
Ype-15-1	NZ_AAOS02000009.1	-	153684	153621	**RNA**→ CrcB (COG0239)→
Ype-5-1	NZ_AAUB01000309.1	+	2856	2919	**RNA**→ CrcB (COG0239)→
Ype-11-1	NZ_AAYR01000001.1	+	1051950	1052013	**RNA**→ CrcB (COG0239)→
Ype-16-1	NZ_AAYS01000013.1	+	76475	76538	**RNA**→ CrcB (COG0239)→
Ype-12-1	NZ_AAYT01000001.1	-	638230	638167	**RNA**→ CrcB (COG0239)→
Ype-9-1	NZ_AAYU01000004.1	+	393680	393743	**RNA**→ CrcB (COG0239)→
Ype-10-1	NZ_AAYV01000001.1	+	356990	357053	**RNA**→ CrcB (COG0239)→
Ype-14-1	NZ_ABAT01000003.1	-	35794	35731	**RNA**→ CrcB (COG0239)→

Ype-3-1	NZ_ABCD01000001.1	+	732214	732277	**RNA**→ CrcB (COG0239)→
Yps-2-1	NC_006155.1	-	1307162	1307099	**RNA**→ CrcB (COG0239)→
Yps-3-1	NC_010634.1	-	1289674	1289611	**RNA**→ CrcB (COG0239)→
Yps-1-1	NC_009708.1	+	3329098	3329161	**RNA**→ CrcB (COG0239)→
Spr-1-1	NC_009832.1	-	1310091	1310029	**RNA**→ CrcB (COG0239)→
Pru-1-1	NZ_ABXV01000084.1	+	33475	33542	**RNA**→
Pal-1-1	NZ_ABXW01000051.1	+	16406	16474	**RNA**→
Pat-1-1	NC_004547.2	-	1468958	1468853	**RNA**→ CrcB (COG0239)→
Cko-1-3	NC_009792.1	+	1409469	1409541	**RNA**→ PPX1 (COG1227)→ COG5383 (COG5383)→ PnbA (COG2272)→ hypo→
Kpn-1-1	NC_009651.1	+	81552	81622	**RNA**→ CrcB (COG0239)→
Bacteria Thermotogae					
Tpe-1-1	NC_009486.1	+	1794817	1794880	**RNA**→ ABC_MutS_homologs (cd03243)→ ABC_MutS_homologs (cd03243)→
Tsp-3-1	NC_010483.1	+	1853799	1853862	**RNA**→ ABC_MutS_homologs (cd03243)→ ABC_MutS_homologs (cd03243)→
Mpi-1-1	NZ_ABXR01000005.1	-	79537	79474	**RNA**→
Tle-1-1	NC_009828.1	-	982299	982238	**RNA**→ hypo→ COG0714 (COG0714)→ COG1721 (COG1721)→ hypo→
Bacteria Verrucomicrobia					
Cfl-1-1	NZ_ABVL01000003.1	+	298778	298860	**RNA**→ EriC (COG0038)→ hypo→
Min-1-1	NC_010794.1	+	1993796	1993891	**RNA**→ ←hypo
bEl-1-1	NZ_ABOX01000011.1	-	261745	261663	**RNA**→ hypo→ EriC (COG0038)→ UPF0132 (pfam03675)→
Min-1-2	NC_010794.1	+	1995740	1995815	**RNA**→ ←hypo
environmental					
env-1	JCVI_SCAF_1096627186085	-	22480	22408	**RNA**→ Phosphoesterase (pfam04185)? (GOS1882)→
env-2	amd_scaffold_498	-	3365	3303	**RNA**→ MutS (COG0249)→ Mismatch repair ATPase (MutS family)→
env-3	BABD01000872.1	-	1124	1056	**RNA**→ EriC (COG0038)→
env-4	soil_Contig38989	+	707	775	**RNA**→ CrcB (COG0239)→
env-5	soil_Contig18198	+	1168	1236	**RNA**→ hypo→
env-6	BAAV01024370.1	-	684	614	**RNA**→ hypo→
env-7	amd_scaffold_44	+	18357	18431	**RNA**→ ←hypo
env-8	amd_scaffold_352	+	2172	2233	**RNA**→ ←hypo
env-9	BABE01001448.1	-	1961	1891	**RNA**→ hypo→
env-10	BAAV01007000.1	+	1090	1152	**RNA**→
env-11	soil_Contig7507	-	664	577	**RNA**→ CrcB (COG0239)→ COG1993 (COG1993)→
env-12	BABC01000066.1	-	7792	7723	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→ GtrA (pfam04138)→
env-13	BABE01006493.1	+	378	447	**RNA**→ ←hypo
env-14	BABD01002470.1	-	1603	1534	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→ GtrA (pfam04138)→
env-15	BABB01014491.1	+	724	793	**RNA**→
env-16	BABE01000053.1	+	15068	15137	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→ COG2246 (COG2246)→
env-17	BABF01001201.1	-	3795	3726	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→ COG2246 (COG2246)→
env-18	BAAY01001852.1	+	271	348	**RNA**→ CrcB (COG0239)→ CrcB (COG0239)→
env-19	hgutS7_s7_167263	+	2364	2441	**RNA**→ ←hypo
env-20	BAAZ01017612.1	-	714	637	**RNA**→ CrcB (COG0239)→
env-21	amd_scaffold_1090	+	122	193	**RNA**→ HyfC (COG0650)→
env-22	amd_scaffold_653	+	2844	2904	**RNA**→ EmrA (COG1566)→
env-23	amd_scaffold_33	-	7896	7826	**RNA**→ AraJ (COG2814)→
env-24	amd_scaffold_470	+	2269	2327	**RNA**→ MutS (COG0249)→ MutS (COG0249)→
env-25	JCVI_SCAF_1101668140910	-	192	117	**RNA**→
env-26	BAAY01017851.1	-	254	182	**RNA**→ ←hypo
env-27	BABB01015950.1	-	548	449	**RNA**→ hypo→
env-28	BABF01003476.1	+	1904	1985	**RNA**→

env-29	BAAV01013884.1	-	268	196	**RNA**→ ←hypo
env-30	BABG01014606.1	-	762	690	**RNA**→ hypo→
env-31	BAAZ01021286.1	-	777	699	**RNA**→ Na_H_Exchanger (pfam00999)→
env-32	BAAV01021889.1	-	203	131	**RNA**→
env-33	BABG01014343.1	-	547	474	**RNA**→ hypo→
env-34	hgutS8_s8_174661	-	467	395	**RNA**→ Chloride channel protein EriC→
env-35	hgutS8_s8_165414	-	1469	1396	**RNA**→ EriC (COG0038)→
env-36	BABG01003409.1	-	431	358	**RNA**→ hypo→
env-37	soil_XZS55564_b1	-	357	289	**RNA**→ CRCB (COG0239)→
env-38	BABC01003766.1	-	468	402	**RNA**→ hypo→
env-39	BABF01015109.1	+	43	109	**RNA**→ EriC (COG0038)→
env-40	hgutS8_s8_178954_2	+	1879	1949	**RNA**→ KefB (COG0475)→
env-41	hgutS7_s7_178954_3	+	4537	4610	**RNA**→ KefB (COG0475)→
env-42	JCVI_SCAF_1096627184930	-	6761	6699	**RNA**→ CrcB (COG0239)? (GOS11882)→
env-43	JCVI_SCAF_1096627388964	+	899189	899292	**RNA**→ EriC (COG0038)? (GOS1383)→ ? (GOS1352)→ PAP2 (pfam01569)→
env-44	BAAW01000138.1	-	2195	2134	**RNA**→ EriC (COG0038)→
env-45	soil_XZS73310_b1	-	89	25	**RNA**→

24.4 Conserved domains

Conserved domains found in protein-coding genes listed in Section 24.3 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *crcB* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

cd00156 (1) Signal receiver domain; originally thought to be unique to bacteria (CheY, OmpR, NtrC, and PhoB), now recently identified in eukaryotes ETR1 Arabidopsis thaliana; this domain receives the signal from the sensor partner in a two-component systems; contains a phosphoacceptor site that is phosphorylated by histidine kinase homologs; usually found N-terminal to a DNA binding effector domain; forms homodimers

cd00288 (1) Pyruvate kinase (PK): Large allosteric enzyme that regulates glycolysis through binding of the substrate, phosphoenolpyruvate, and one or more allosteric effectors.

cd00293 (9) Usp: Universal stress protein family.

cd00412 (1) Inorganic pyrophosphatase.

cd00464 (2) Shikimate kinase (SK) is the fifth enzyme in the shikimate pathway, a seven-step biosynthetic pathway which converts erythrose-4-phosphate to chorismic acid, found in bacteria, fungi and plants.

cd00537 (1) Methylenetetrahydrofolate reductase (MTHFR).

cd00688 (1) This group contains class II terpene cyclases, protein prenyltransferases beta subunit, two broadly specific proteinase inhibitors alpha2-macroglobulin (alpha (2)-M) and pregnancy zone protein (PZP) and, the C3 C4 and C5 components of vertebrate complement.

cd01021 (1) Goose Egg White Lysozyme domain.

cd01637 (1) Inositol-monophosphatase-like domains.

cd02150 (1) NAD(P)H:flavin oxidoreductase-like family 1.

cd03243 (10) The MutS protein initiates DNA mismatch repair by recognizing mispaired and unpaired bases embedded in duplex DNA and activating endo- and exonucleases to remove the mismatch.

cd03313 (6) Enolase: Enolases are homodimeric enzymes that catalyse the reversible dehydration of 2-phospho-D-glycerate to phosphoenolpyruvate as part of the glycolytic and gluconeogenesis pathways.

cd03395 (1) PAP2_like_4 proteins.

COG0025 (33) NhaP-type Na+/H+ and K+/H+ antiporters [Inorganic ion transport and metabolism]

COG0038 (109) Chloride channel protein EriC [Inorganic ion transport and metabolism]

COG0169 (1) Shikimate 5-dehydrogenase [Amino acid transport and metabolism]

COG0218 (1) Predicted GTPase [General function prediction only]

COG0239 (202) Integral membrane protein possibly involved in chromosome condensation [Cell division and chromosome partitioning]

COG0249 (11) Mismatch repair ATPase (MutS family) [DNA replication, recombination, and repair]

COG0452 (1) Phosphopantothenoylcysteine synthetase/decarboxylase [Coenzyme metabolism]

COG0474 (1) Cation transport ATPase [Inorganic ion transport and metabolism]

COG0475 (9) Kef-type K+ transport systems, membrane components [Inorganic ion transport and metabolism]

COG0557 (1) Exoribonuclease R [Transcription]

COG0561 (2) Predicted hydrolases of the HAD superfamily [General function prediction only]

COG0589 (1) Universal stress protein UspA and related nucleotide-binding proteins [Signal transduction mechanisms]

COG0637 (1) Predicted phosphatase/phosphohexomutase [General function prediction only]

COG0642 (1) Signal transduction histidine kinase [Signal transduction mechanisms]

COG0650 (7) Formate hydrogenlyase subunit 4 [Energy production and conversion]

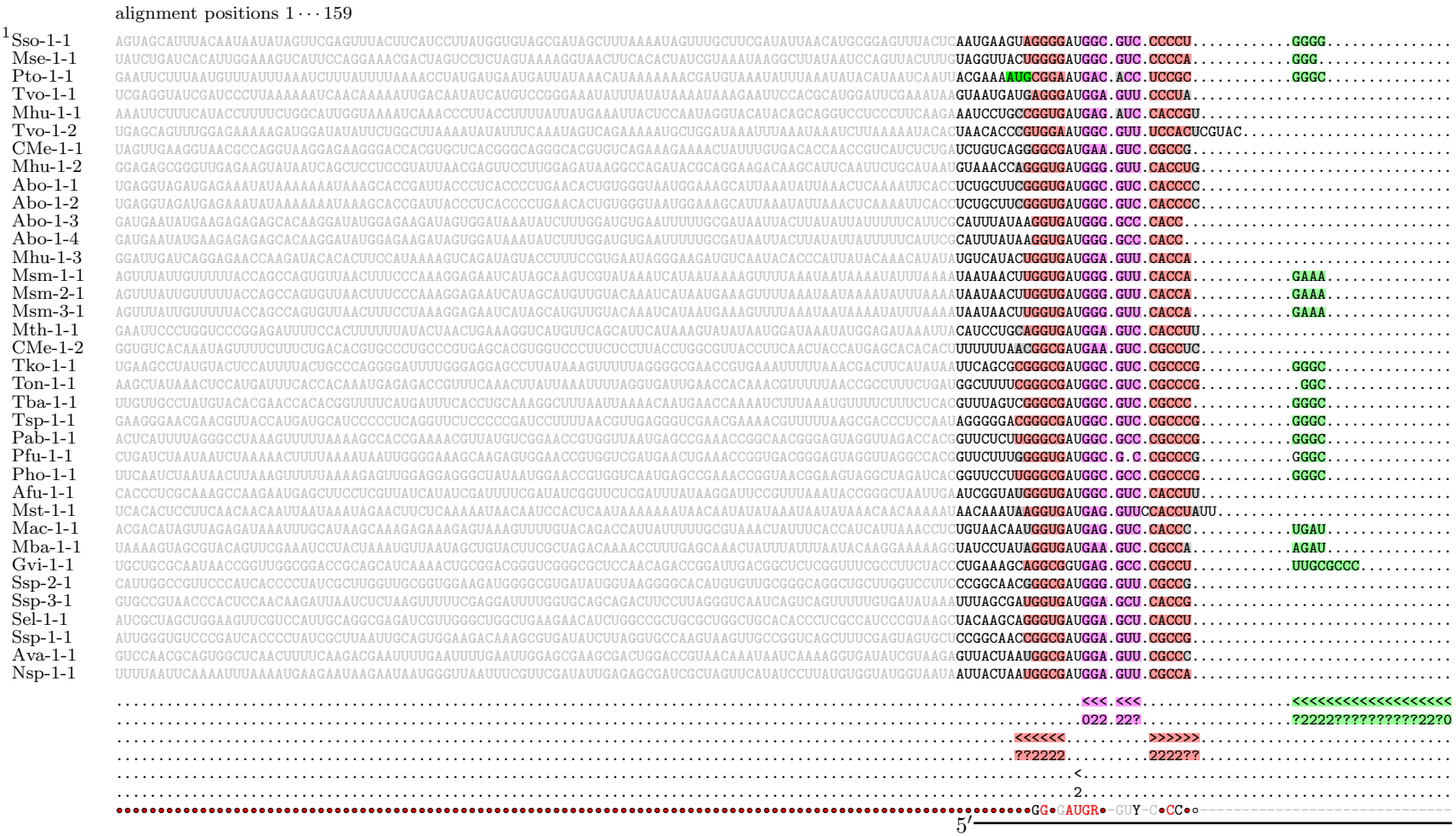
COG0651 (10) Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD

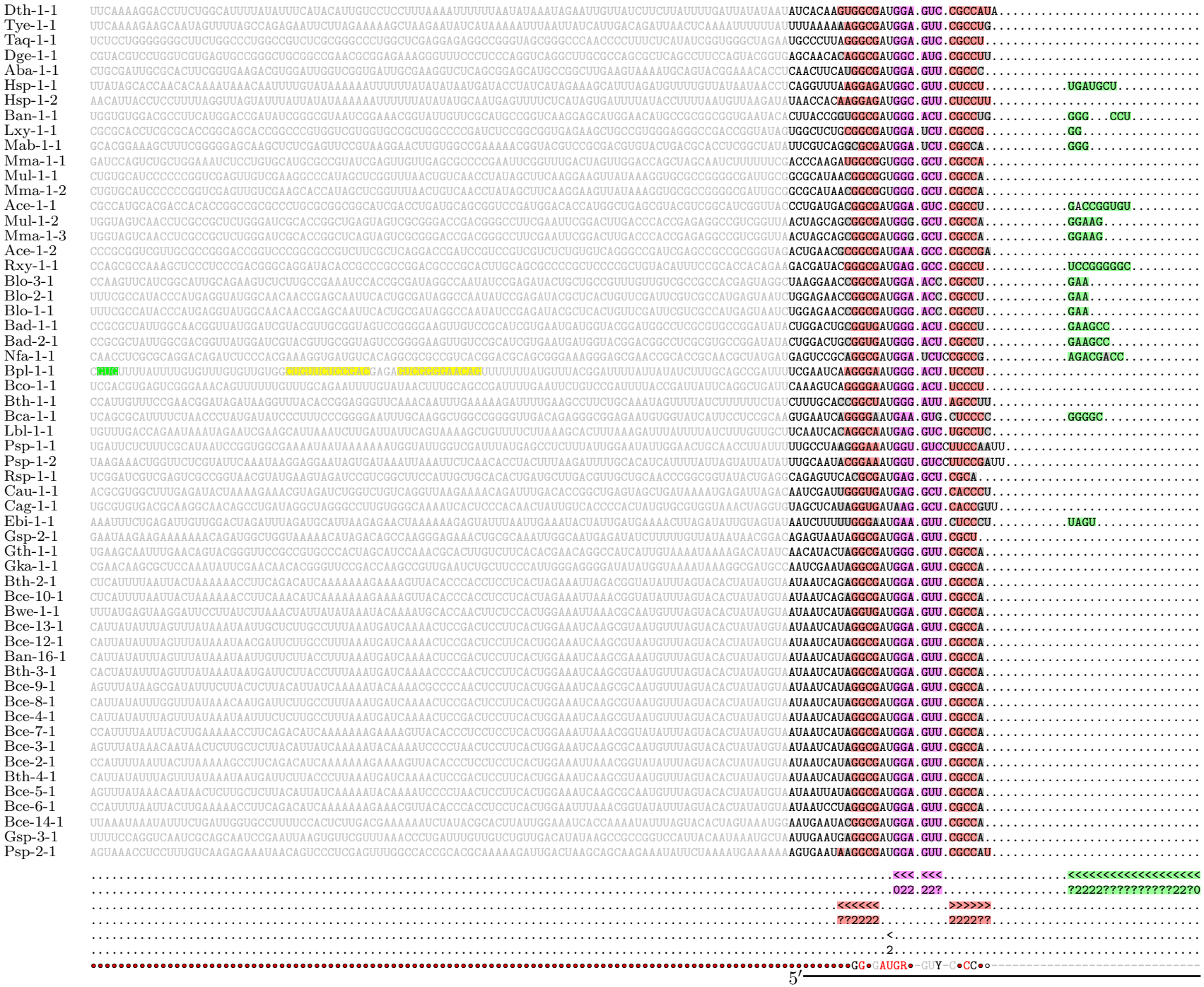
subunit [Energy production and conversion / Inorganic ion transport and metabolism]
COG0714 (1) MoxR-like ATPases [General function prediction only]
COG0810 (1) Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane]
COG0822 (14) NifU homolog involved in Fe-S cluster formation [Energy production and conversion]
COG1051 (1) ADP-ribose pyrophosphatase [Nucleotide transport and metabolism]
COG1159 (5) GTPase [General function prediction only]
COG1168 (2) Bifunctional PLP-dependent enzyme with beta-cystathionase and maltose regulon repressor activities [Amino acid transport and metabolism]
COG1227 (1) Inorganic pyrophosphatase/exopolyphosphatase [Energy production and conversion]
COG1357 (1) Uncharacterized low-complexity proteins [Function unknown]
COG1566 (1) Multidrug resistance efflux pump [Defense mechanisms]
COG1689 (8) Uncharacterized protein conserved in archaea [Function unknown]
COG1694 (1) Predicted pyrophosphatase [General function prediction only]
COG1721 (1) Uncharacterized conserved protein (some members contain a von Willebrand factor type A (vWA) domain) [General function prediction only]
COG1739 (1) Uncharacterized conserved protein [Function unknown]
COG1993 (32) Uncharacterized conserved protein [Function unknown]
COG2002 (1) Regulators of stationary/sporulation gene expression [Transcription]
COG2202 (1) FOG: PAS/PAC domain [Signal transduction mechanisms]
COG2206 (1) HD-GYP domain [Signal transduction mechanisms]
COG2223 (2) Nitrate/nitrite transporter [Inorganic ion transport and metabolism]
COG2246 (2) Predicted membrane protein [Function unknown]
COG2271 (1) Sugar phosphate permease [Carbohydrate transport and metabolism]
COG2272 (1) Carboxylesterase type B [Lipid metabolism]
COG2510 (2) Predicted membrane protein [Function unknown]
COG2520 (2) Predicted methyltransferase [General function prediction only]
COG2524 (1) Predicted transcriptional regulator, contains C-terminal CBS domains [Transcription]
COG2814 (2) Arabinose efflux permease [Carbohydrate transport and metabolism]
COG2846 (1) Regulator of cell morphogenesis and NO signaling [Cell division and chromosome partitioning]
COG2966 (1) Uncharacterized conserved protein [Function unknown]
COG3260 (5) Ni,Fe-hydrogenase III small subunit [Energy production and conversion]
COG3261 (5) Ni,Fe-hydrogenase III large subunit [Energy production and conversion]
COG3467 (1) Predicted flavin-nucleotide-binding protein [General function prediction only]
COG3481 (4) Predicted HD-superfamily hydrolase [General function prediction only]
COG3610 (1) Uncharacterized conserved protein [Function unknown]
COG3757 (1) Lyzozyme M1 (1,4-beta-N-acetylmuramidase) [Cell envelope biogenesis, outer membrane]
COG4087 (1) Soluble P-type ATPase [General function prediction only]
COG4237 (4) Hydrogenase 4 membrane component (E) [Energy production and conversion]
COG5256 (1) Translation elongation factor EF-1alpha (GTPase) [Translation, ribosomal structure and biogenesis]
COG5383 (1) Uncharacterized protein conserved in bacteria [Function unknown]
GOS11882 (1) ?
GOS1352 (1) ?
GOS1383 (1) ?
GOS1882 (1) ?
NOG18330 (1)
NOG22924 (1)
NOG33876 (1)
pfam00009 (1) Elongation factor Tu GTP binding domain.
pfam00329 (1) Respiratory-chain NADH dehydrogenase, 30 Kd subunit.
pfam00571 (1) CBS domain.
pfam00582 (4) Universal stress protein family.
pfam00583 (2) Acetyltransferase (GNAT) family.
pfam00795 (1) Carbon-nitrogen hydrolase.
pfam00912 (1) Transglycosylase.
pfam00924 (1) Mechanosensitive ion channel.
pfam00999 (6) Sodium/hydrogen exchanger family.
pfam01569 (1) PAP2 superfamily.
pfam02617 (1) ATP-dependent Clp protease adaptor protein ClpS.
pfam02641 (2) Uncharacterized ACR, COG1993.
pfam03675 (1) Uncharacterized protein family (UPF0132).
pfam04138 (4) GtrA-like protein.
pfam04185 (1) Phosphoesterase family.
pfam04254 (1) Protein of unknown function (DUF432).
pfam05239 (1) PRC-barrel domain.
smart00418 (1) helix_turn_helix, Arsenical Resistance Operon Repressor;

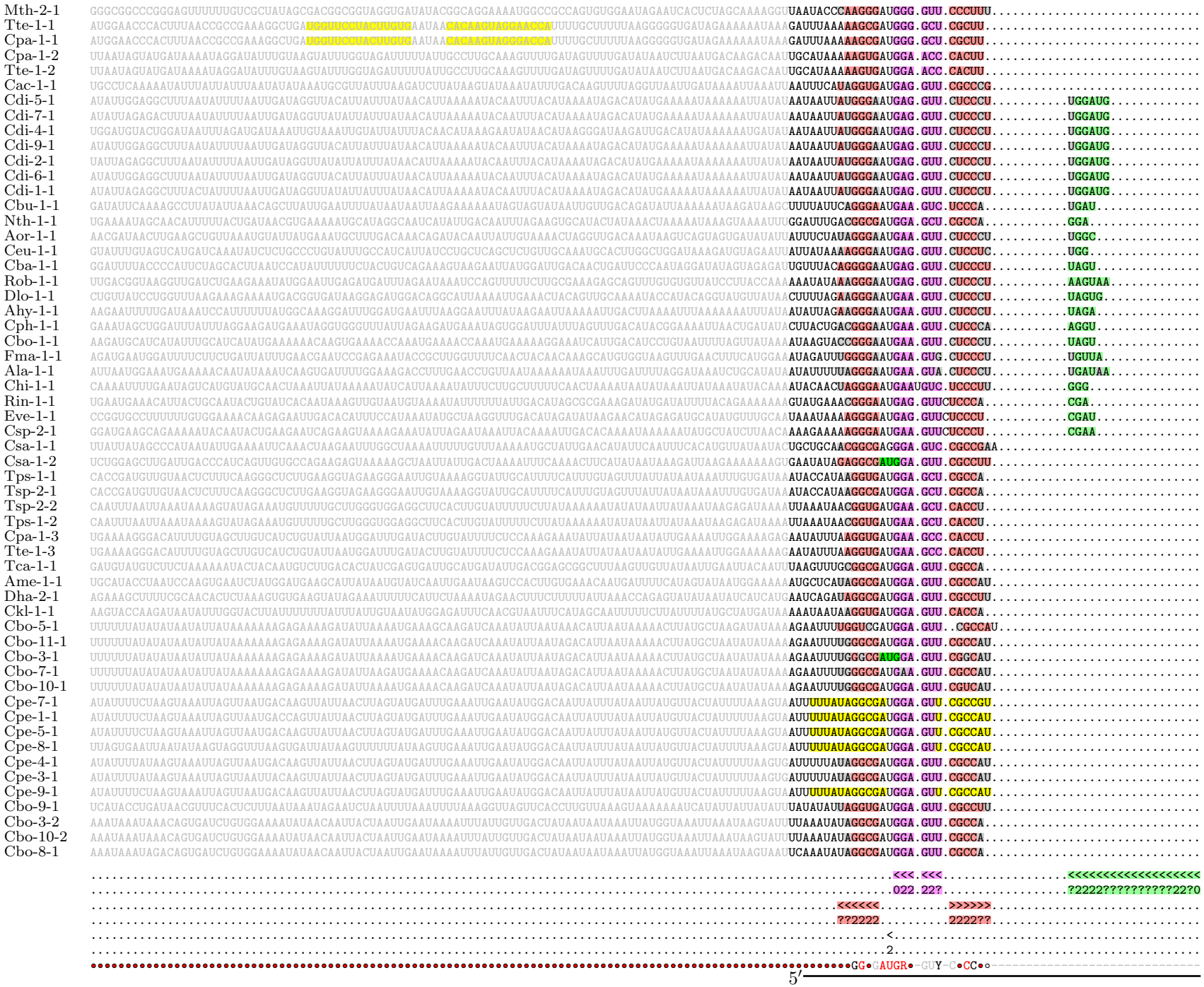
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Aba-2-1, Aba-6-1, Ban-10-1, Ban-11-1, Ban-12-1, Ban-13-1, Ban-14-1, Ban-15-1, Ban-17-1, Ban-18-1, Ban-2-1, Ban-3-1, Ban-4-1, Ban-5-1, Ban-6-1, Ban-7-1, Ban-8-1, Ban-9-1, Bce-1-1, Bce-11-1, Bce-16-1, Blo-1-2, Bma-1-1, Bma-3-1, Bma-4-1, Bma-5-1, Bma-6-1, Bma-7-1, Bma-8-1, Bma-9-1, Bmu-1-2, Bps-11-1, Bps-12-1, Bps-13-1, Bps-15-1, Bps-2-1, Bps-20-1, Bps-3-1, Bps-4-1, Bps-6-1, Bth-5-1, Bth-6-2, Cbo-2-1, Cbo-2-2, Cbo-4-1, Cbo-4-2, Cbo-7-2, Cdi-10-1, Cdi-3-1, Cdi-8-1, Cpe-2-1, Cpe-6-1, Csp-1-1, Cth-2-1, Dha-1-1, Gdi-1-4, Gsp-1-1, Lga-2-1, Lga-2-2, Mpi-1-1, Pae-1-1, Psy-3-1, Rpi-1-1, Rso-4-1, Sel-2-1, Tsp-3-1, Xca-2-1, Xca-3-1, Xor-2-1, Xor-3-1, Xor-3-2, Ype-1-1, Ype-10-1, Ype-11-1, Ype-12-1, Ype-14-1, Ype-15-1, Ype-16-1, Ype-2-1, Ype-3-1, Ype-4-1, Ype-5-1, Ype-6-1, Ype-7-1, Ype-8-1, Ype-9-1, Yps-3-1, env-14, env-18

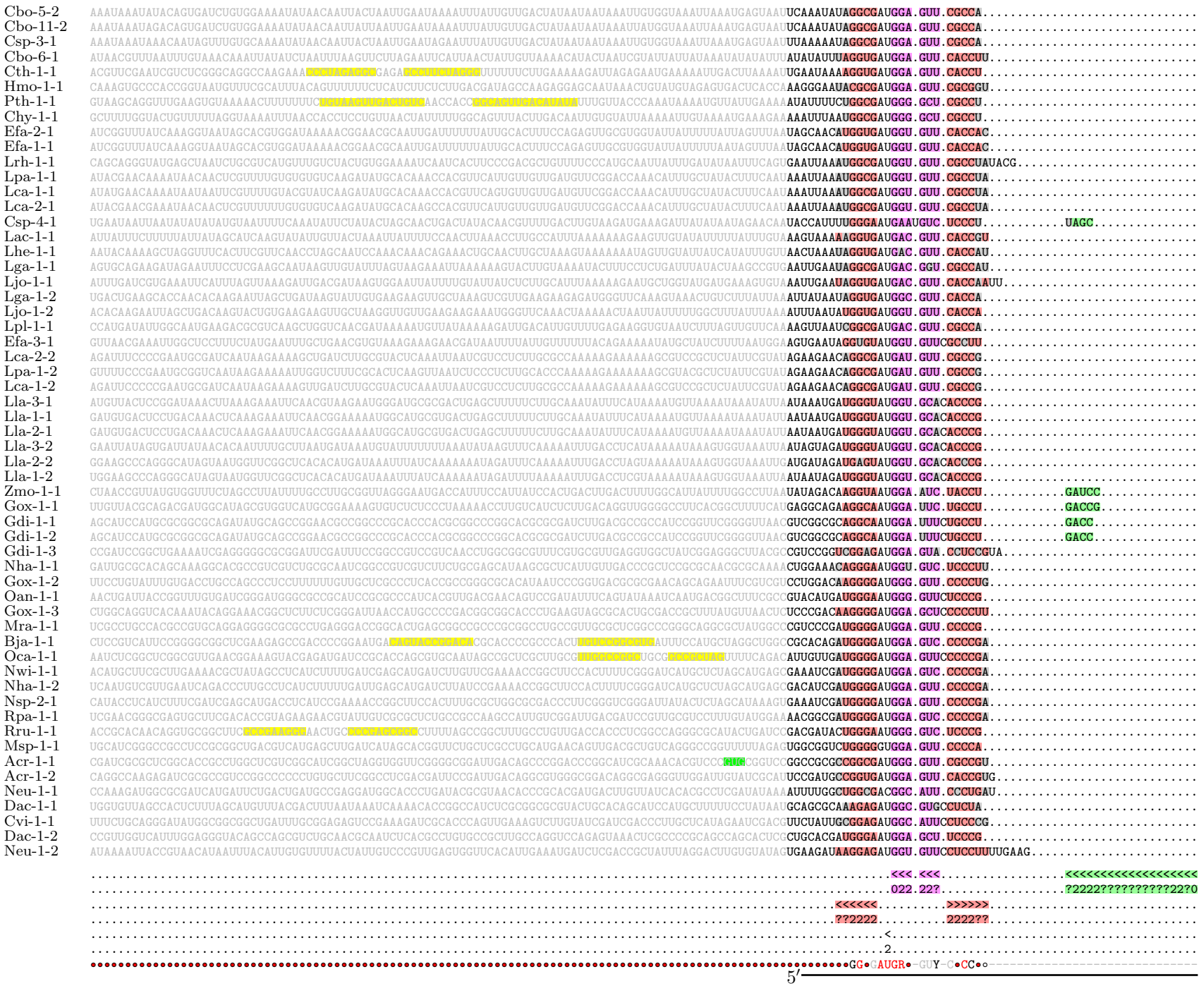
24.5 Multiple-sequence alignment

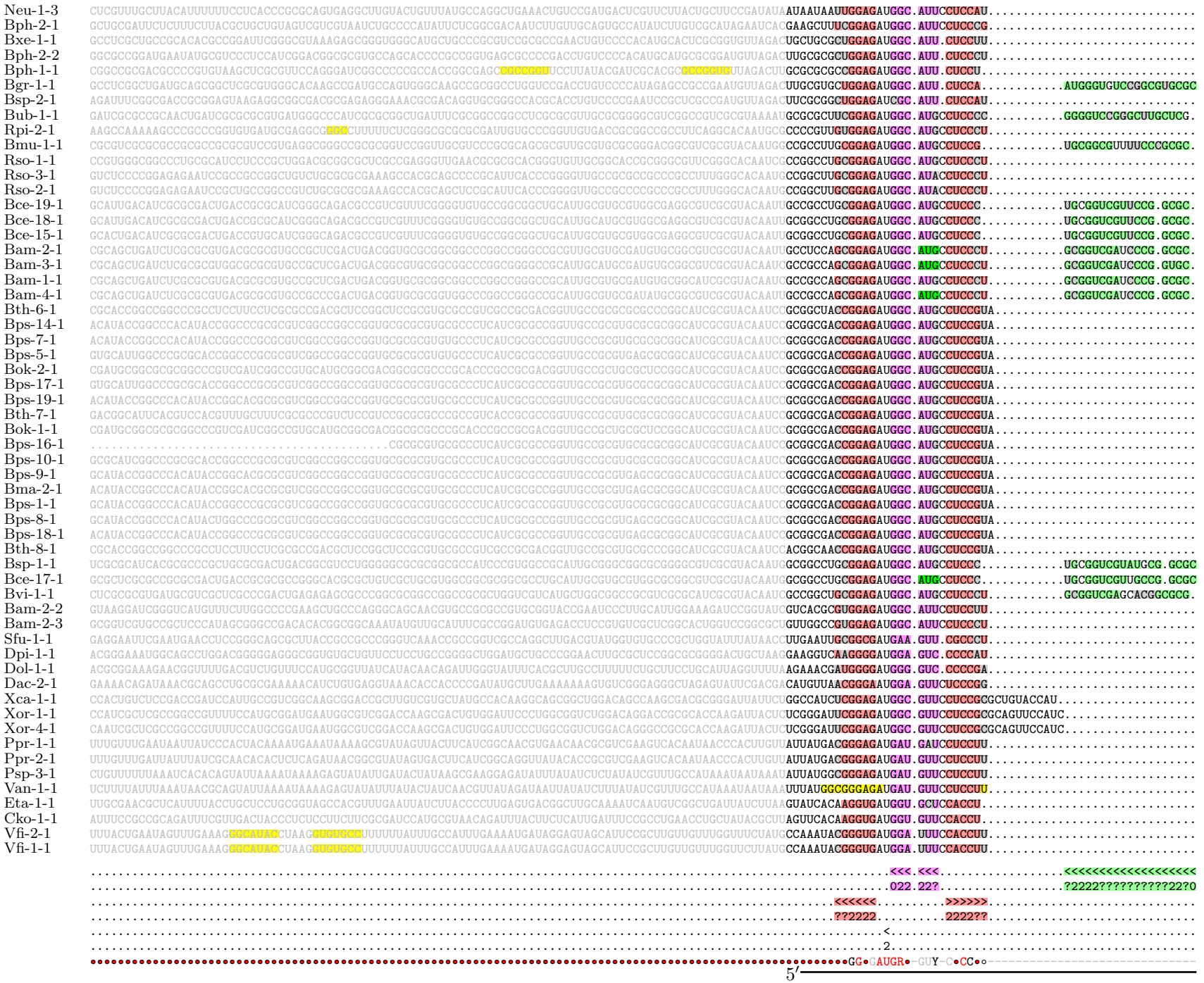
Each *crcB* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 24.1. Superscript numbers refer to annotations listed in Section 24.2. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *crcB* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

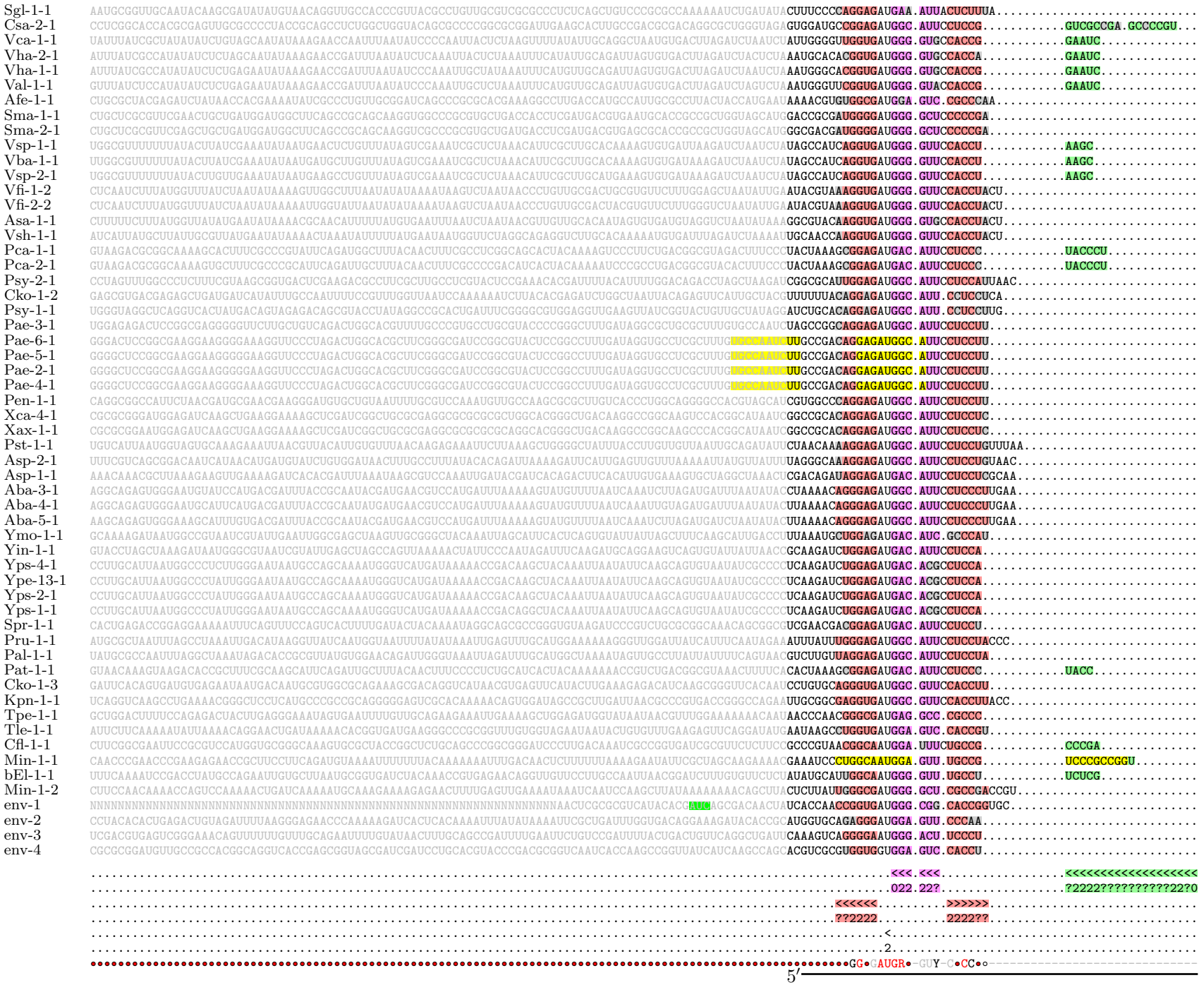


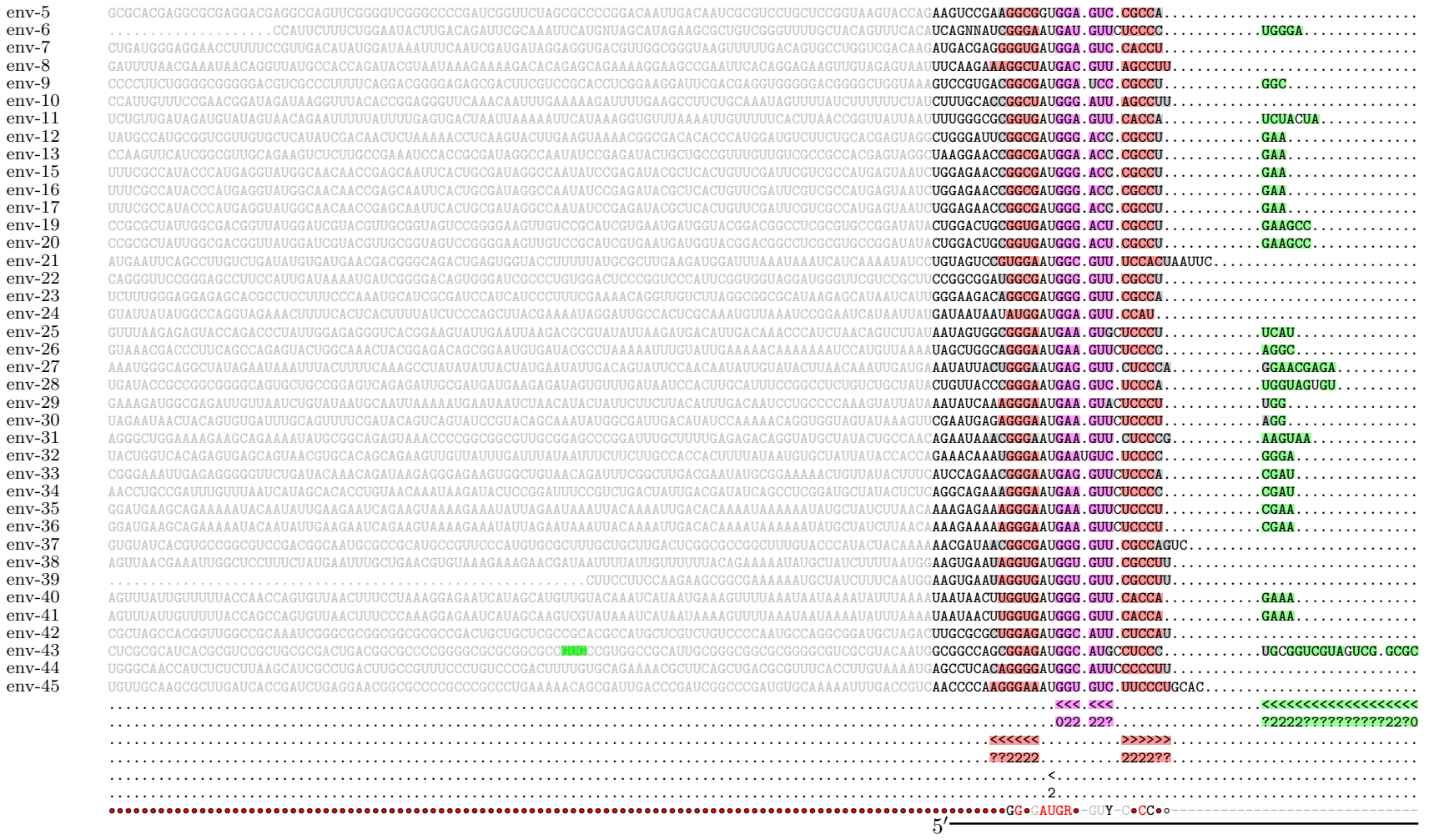

Dth-1-1
Tye-1-1
Taq-1-1
Dge-1-1
Aba-1-1
Hsp-1-1
Hsp-1-2
Ban-1-1
Lxy-1-1
Mab-1-1
Mma-1-1
Mul-1-1
Mma-1-2
Ace-1-1
Mul-1-2
Mma-1-3
Ace-1-2
Rxy-1-1
Blo-3-1
Blo-2-1
Blo-1-1
Bad-1-1
Bad-2-1
Nfa-1-1
Bpl-1-1
Bco-1-1
Bth-1-1
Bca-1-1
Lbl-1-1
Psp-1-1
Psp-1-2
Rsp-1-1
Cau-1-1
Cag-1-1
Ebi-1-1
Gsp-2-1
Gth-1-1
Gka-1-1
Bth-2-1
Bce-10-1
Bwe-1-1
Bce-13-1
Bce-12-1
Ban-16-1
Bth-3-1
Bce-9-1
Bce-8-1
Bce-4-1
Bce-7-1
Bce-3-1
Bce-2-1
Bth-4-1
Bce-5-1
Bce-6-1
Bce-14-1
Gsp-3-1
Psp-2-1
5′

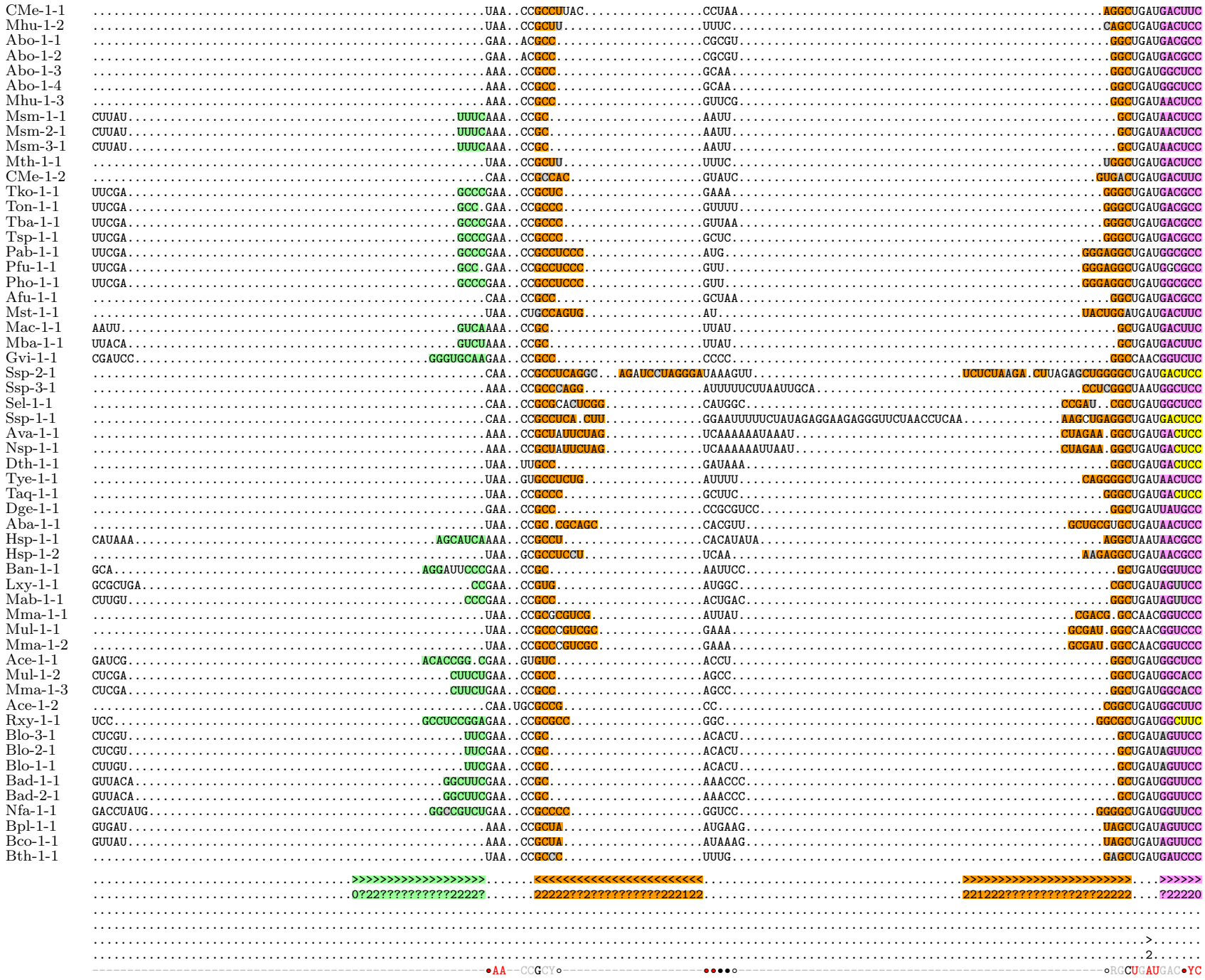

Mth-2-1
Tte-1-1
Cpa-1-1
Cpa-1-2
Tte-1-2
Cac-1-1
Cdi-5-1
Cdi-7-1
Cdi-4-1
Cdi-9-1
Cdi-2-1
Cdi-6-1
Cdi-1-1
Cbu-1-1
Nth-1-1
Aor-1-1
Ceu-1-1
Cba-1-1
Rob-1-1
Dlo-1-1
Ahy-1-1
Cph-1-1
Cbo-1-1
Fma-1-1
Ala-1-1
Chi-1-1
Rin-1-1
Eve-1-1
Csp-2-1
Csa-1-1
Csa-1-2
Tps-1-1
Tsp-2-1
Tsp-2-2
Tps-1-2
Cpa-1-3
Tte-1-3
Tca-1-1
Ame-1-1
Dha-2-1
Ckl-1-1
Cbo-5-1
Cbo-11-1
Cbo-3-1
Cbo-7-1
Cbo-10-1
Cpe-7-1
Cpe-1-1
Cpe-5-1
Cpe-8-1
Cpe-4-1
Cpe-3-1
Cpe-9-1
Cbo-9-1
Cbo-3-2
Cbo-10-2
Cbo-8-1
5′

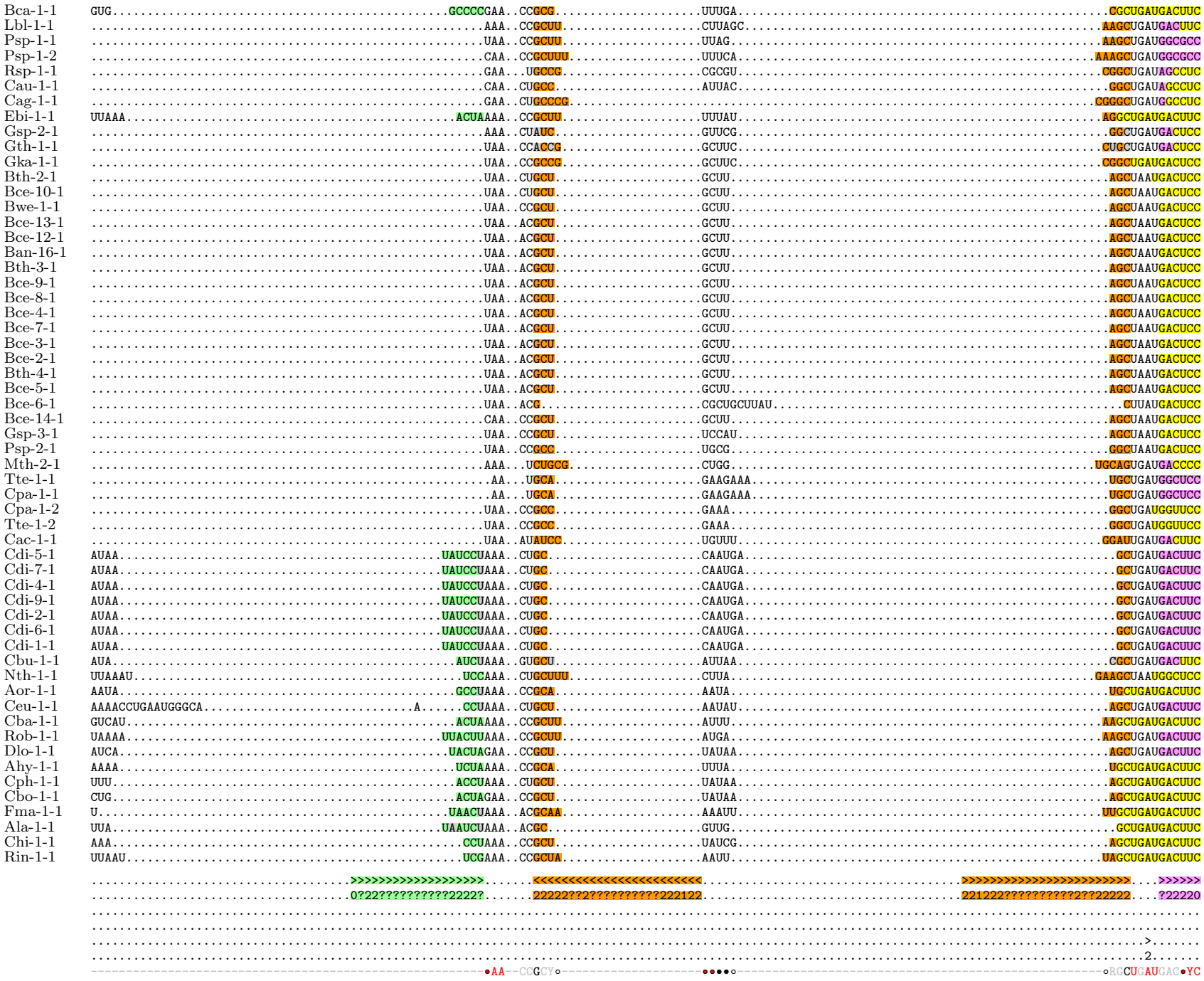

Cbo-5-2
Cbo-11-2
Csp-3-1
Cbo-6-1
Cth-1-1
Hmo-1-1
Pth-1-1
Chy-1-1
Efa-2-1
Efa-1-1
Lrh-1-1
Lpa-1-1
Lca-1-1
Lca-2-1
Csp-4-1
Lac-1-1
Lhe-1-1
Lga-1-1
Ljo-1-1
Lga-1-2
Ljo-1-2
Lpl-1-1
Efa-3-1
Lca-2-2
Lpa-1-2
Lca-1-2
Lla-3-1
Lla-1-1
Lla-2-1
Lla-3-2
Lla-2-2
Lla-1-2
Zmo-1-1
Gox-1-1
Gdi-1-1
Gdi-1-2
Gdi-1-3
Nha-1-1
Gox-1-2
Oan-1-1
Gox-1-3
Mra-1-1
Bja-1-1
Oca-1-1
Nwi-1-1
Nha-1-2
Nsp-2-1
Rpa-1-1
Rru-1-1
Msp-1-1
Acr-1-1
Acr-1-2
Neu-1-1
Dac-1-1
Cvi-1-1
Dac-1-2
Neu-1-2
5′

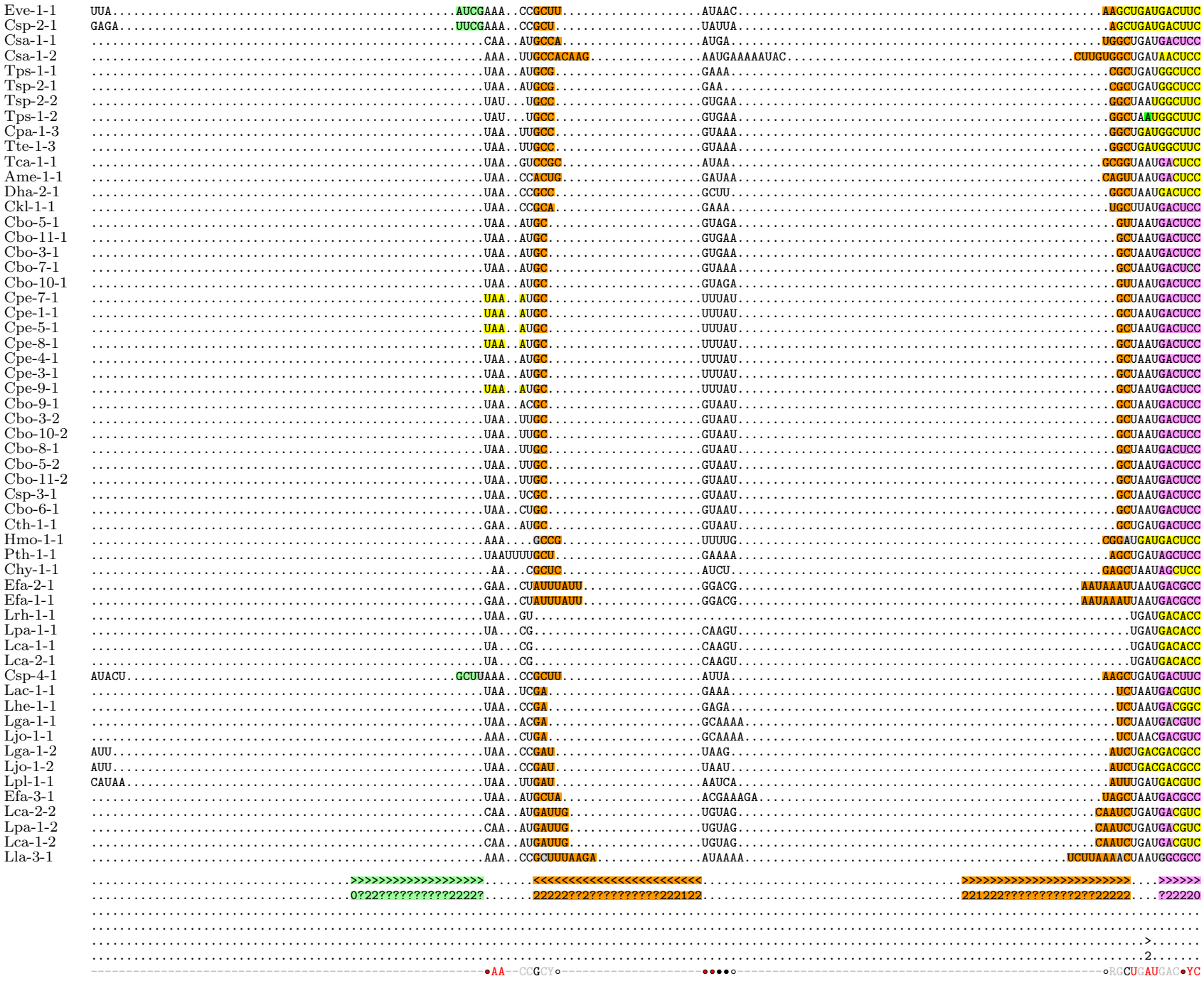

Neu-1-3
Bph-2-1
Bxe-1-1
Bph-2-2
Bph-1-1
Bgr-1-1
Bsp-2-1
Bub-1-1
Rpi-2-1
Bmu-1-1
Rso-1-1
Rso-3-1
Rso-2-1
Bce-19-1
Bce-18-1
Bce-15-1
Bam-2-1
Bam-3-1
Bam-1-1
Bam-4-1
Bth-6-1
Bps-14-1
Bps-7-1
Bps-5-1
Bok-2-1
Bps-17-1
Bps-19-1
Bth-7-1
Bok-1-1
Bps-16-1
Bps-10-1
Bps-9-1
Bma-2-1
Bps-1-1
Bps-8-1
Bps-18-1
Bth-8-1
Bsp-1-1
Bce-17-1
Bvi-1-1
Bam-2-2
Bam-2-3
Sfu-1-1
Dpi-1-1
Dol-1-1
Dac-2-1
Xca-1-1
Xor-1-1
Xor-4-1
Ppr-1-1
Ppr-2-1
Psp-3-1
Van-1-1
Eta-1-1
Cko-1-1
Vfi-2-1
Vfi-1-1
5′

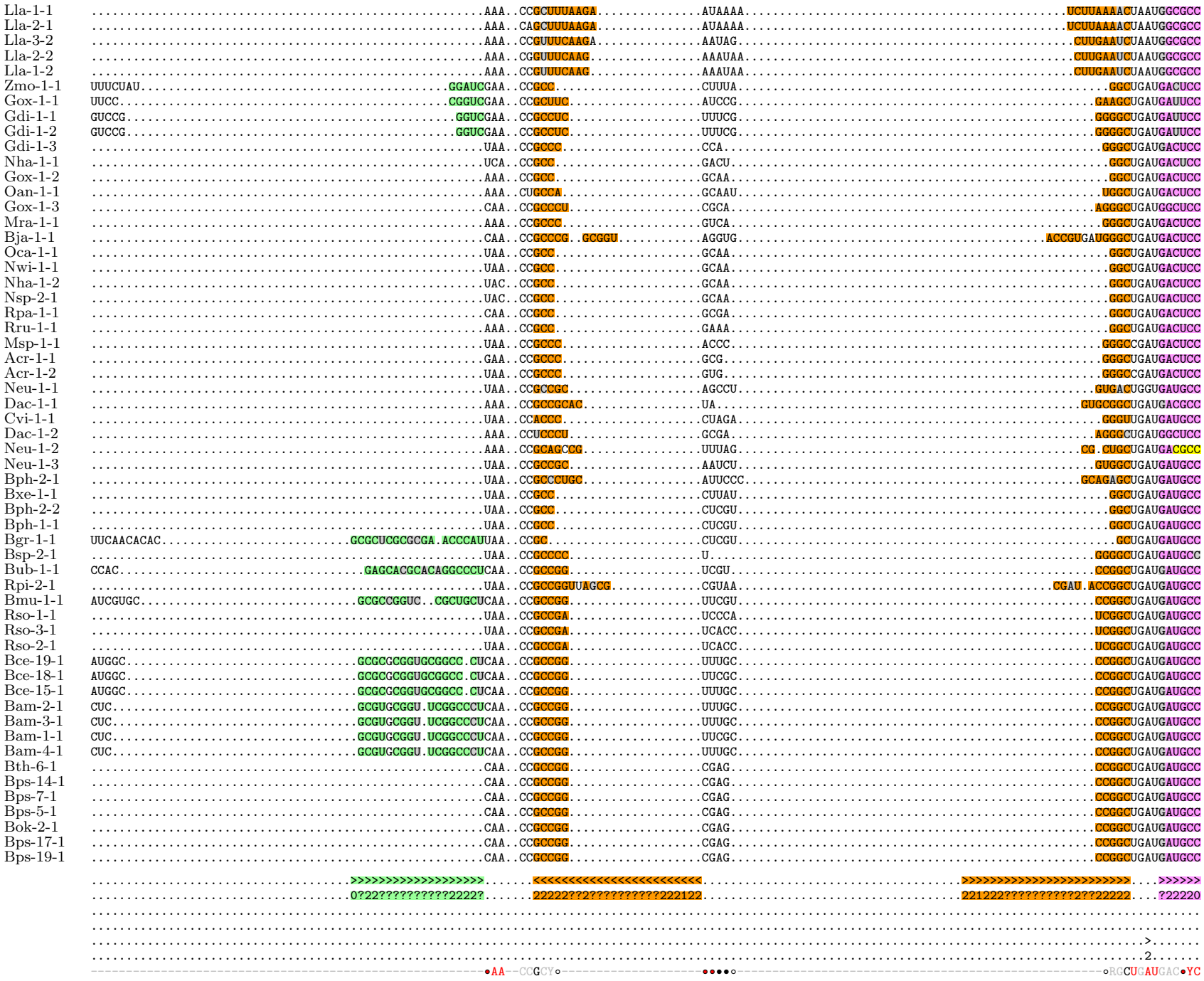

Sgl-1-1
Csa-2-1
Vca-1-1
Vha-2-1
Vha-1-1
Val-1-1
Afe-1-1
Sma-1-1
Sma-2-1
Vsp-1-1
Vba-1-1
Vsp-2-1
Vfi-1-2
Vfi-2-2
Asa-1-1
Vsh-1-1
Pca-1-1
Pca-2-1
Psy-2-1
Cko-1-2
Psy-1-1
Pae-3-1
Pae-6-1
Pae-5-1
Pae-2-1
Pae-4-1
Pen-1-1
Xca-4-1
Xax-1-1
Pst-1-1
Asp-2-1
Asp-1-1
Aba-3-1
Aba-4-1
Aba-5-1
Ymo-1-1
Yin-1-1
Yps-4-1
Ype-13-1
Yps-2-1
Yps-1-1
Spr-1-1
Pru-1-1
Pal-1-1
Pat-1-1
Cko-1-3
Kpn-1-1
Tpe-1-1
Tle-1-1
Cfl-1-1
Min-1-1
bEl-1-1
Min-1-2
env-1
env-2
env-3
env-4
5′

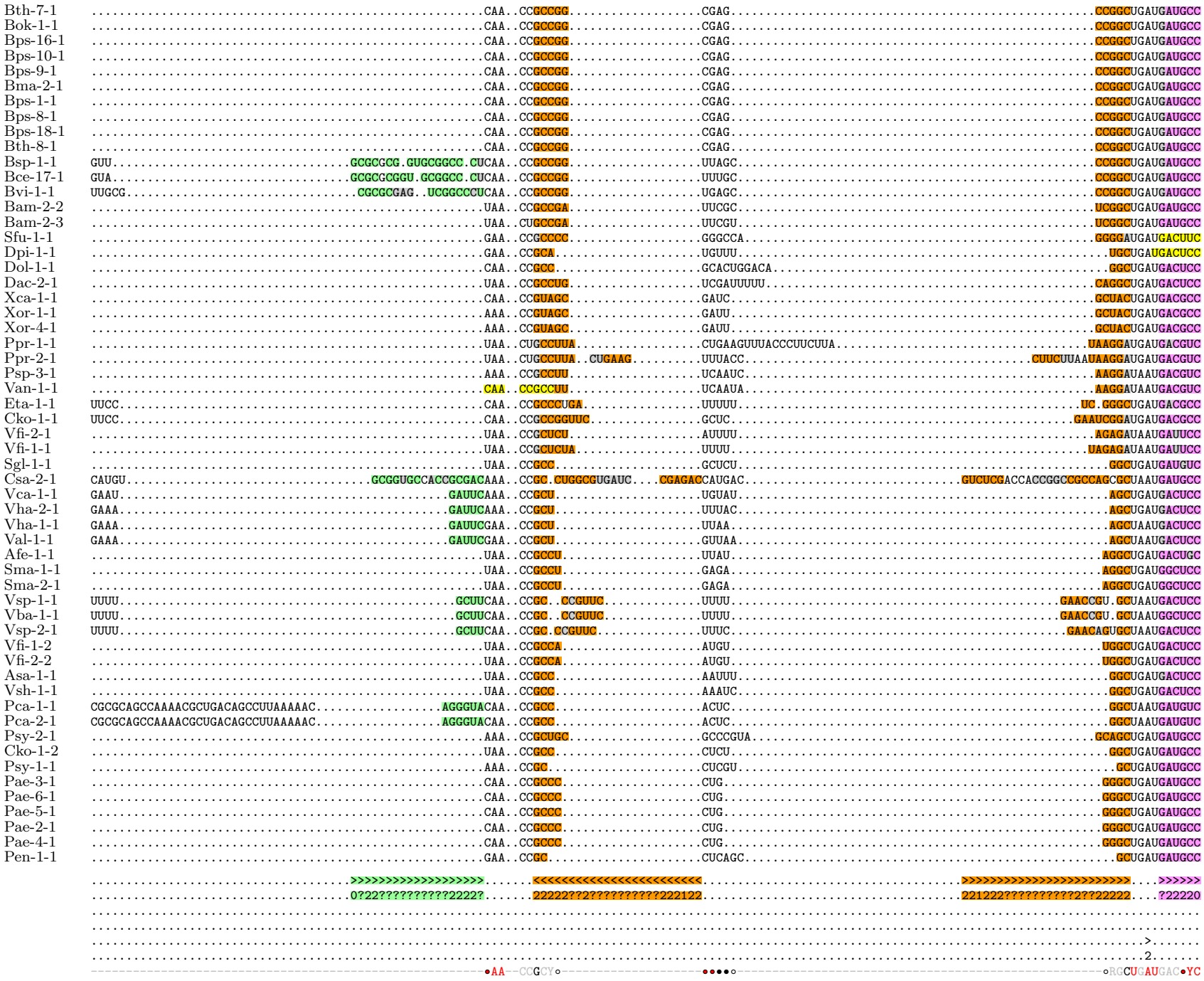

env-5
env-6
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-15
env-16
env-17
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
5′
alignment positions 160 · · · 317
1 Sso-1-1
Mse-1-1
Pto-1-1
Tvo-1-1
Mhu-1-1
Tvo-1-2

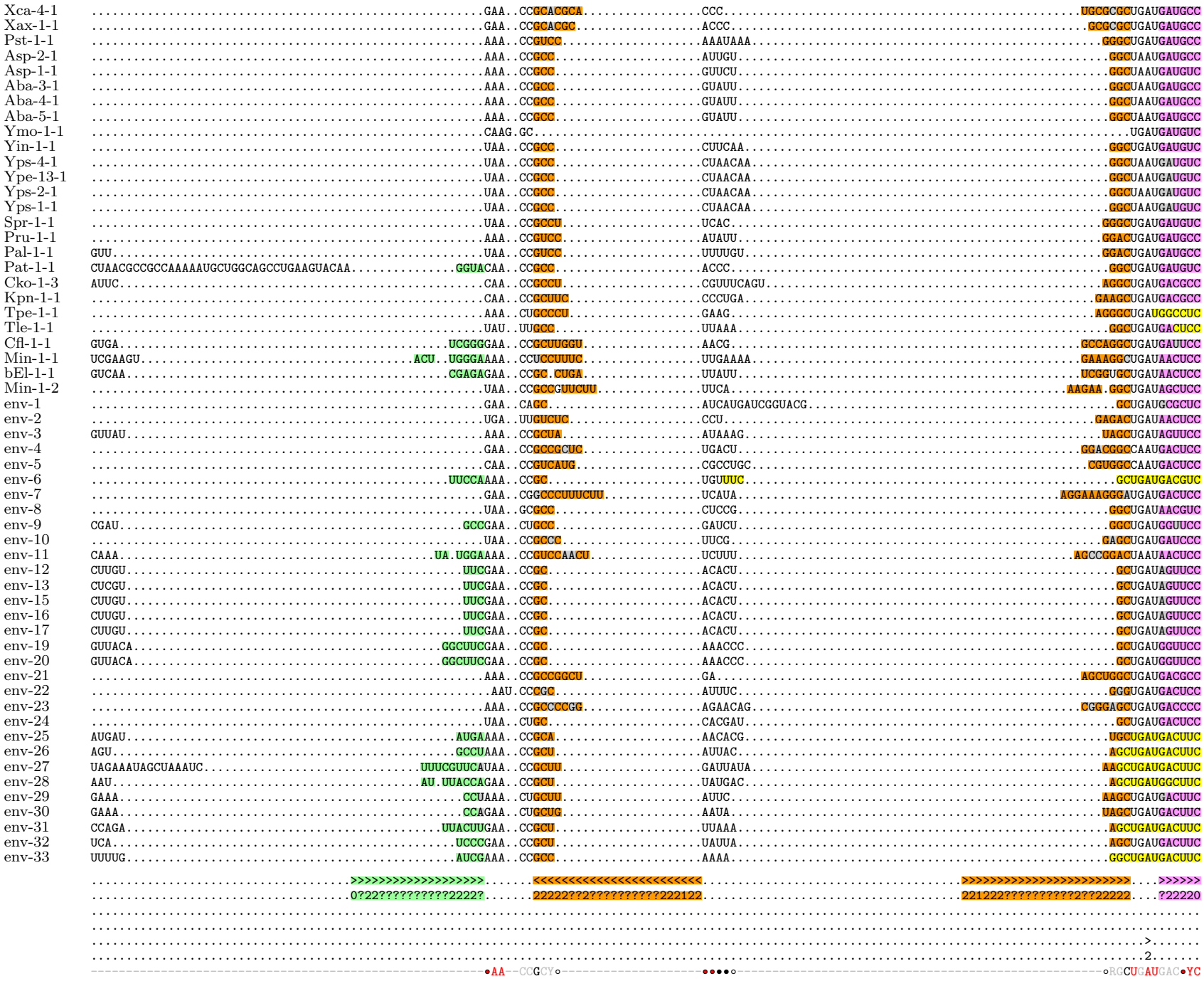

CMe-1-1
Mhu-1-2
Abo-1-1
Abo-1-2
Abo-1-3
Abo-1-4
Mhu-1-3
Msm-1-1
Msm-2-1
Msm-3-1
Mth-1-1
CMe-1-2
Tko-1-1
Ton-1-1
Tba-1-1
Tsp-1-1
Pab-1-1
Pfu-1-1
Pho-1-1
Afu-1-1
Mst-1-1
Mac-1-1
Mba-1-1
Gvi-1-1
Ssp-2-1
Ssp-3-1
Sel-1-1
Ssp-1-1
Ava-1-1
Nsp-1-1
Dth-1-1
Tye-1-1
Taq-1-1
Dge-1-1
Aba-1-1
Hsp-1-1
Hsp-1-2
Ban-1-1
Lxy-1-1
Mab-1-1
Mma-1-1
Mul-1-1
Mma-1-2
Ace-1-1
Mul-1-2
Mma-1-3
Ace-1-2
Rxy-1-1
Blo-3-1
Blo-2-1
Blo-1-1
Bad-1-1
Bad-2-1
Nfa-1-1
Bpl-1-1
Bco-1-1
Bth-1-1

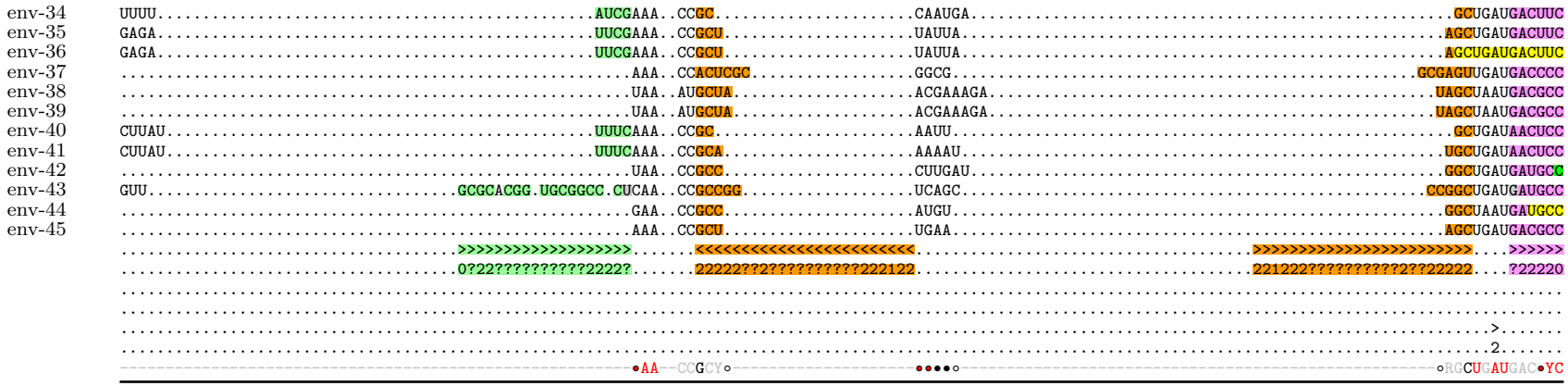

Bca-1-1 GUG.....GCCCCGAA..CCGCG.....UUUGA.....CGCUGAUGACUUC
Lbl-1-1 .....AAA..CCGCUU.....CUUAGC.....AAGCUGAUGACUUC
Psp-1-1 .....UAA..CCGCUU.....UUAG.....AAGCUGAUGGCGCC
Psp-1-2 .....CAA..CCGCUUU.....UUUCA.....AAAGCUGAUGGCGCC
Rsp-1-1 .....GAA...UGCCG.....CGCGU.....CGGCUGAUAGCCUC
Cau-1-1 .....CAA..CUGCC.....AUUAC.....GGCUGAUAGCCUC
Cag-1-1 .....GAA..CUGCCCG.....CGGGCUGAUGGCCUC
Ebi-1-1 UUAAA.....ACUAAAA..CCGCUU.....UUUAU.....AGCCUGAUGACUUC
Gsp-2-1 .....AAA..CUAUC.....GUUCG.....GGCUGAUGACUCC
Gth-1-1 .....UAA..CCACCG.....GCUUC.....CUGCUGAUGACUCC
Gka-1-1 .....UAA..CCGCCG.....GCUUC.....CGGCUGAUGACUCC
Bth-2-1 .....UAA..CUGCU.....GCUU.....AGCUAAUGACUCC
Bce-10-1 .....UAA..CUGCU.....GCUU.....AGCUAAUGACUCC
Bwe-1-1 .....UAA..CCGCU.....GCUU.....AGCUAAUGACUCC
Bce-13-1 .....UAA..ACGCU.....GCUU.....AGCUAAUGACUCC
Bce-12-1 .....UAA..ACGCU.....GCUU.....AGCUAAUGACUCC
Ban-16-1 .....UAA..ACGCU.....GCUU.....AGCUAAUGACUCC
Bth-3-1 .....UAA..ACGCU.....GCUU.....AGCUAAUGACUCC
Bce-9-1 .....UAA..ACGCU.....GCUU.....AGCUAAUGACUCC
Bce-8-1 .....UAA..ACGCU.....GCUU.....AGCUAAUGACUCC
Bce-4-1 .....UAA..ACGCU.....GCUU.....AGCUAAUGACUCC
Bce-7-1 .....UAA..ACGCU.....GCUU.....AGCUAAUGACUCC
Bce-3-1 .....UAA..ACGCU.....GCUU.....AGCUAAUGACUCC
Bce-2-1 .....UAA..ACGCU.....GCUU.....AGCUAAUGACUCC
Bth-4-1 .....UAA..ACGCU.....GCUU.....AGCUAAUGACUCC
Bce-5-1 .....UAA..ACGCU.....GCUU.....AGCUAAUGACUCC
Bce-6-1 .....UAA..ACG.....CGCUGCUUAU.....CUUAUGACUCC
Bce-14-1 .....CAA..CCGCU.....GCUU.....AGCUAAUGACUCC
Gsp-3-1 .....UAA..CCGCU.....UCCAU.....AGCUAAUGACUCC
Psp-2-1 .....UAA..CCGCC.....UGCG.....GGCUAAUGACUCC
Mth-2-1 .....AAA...UCUGCG.....CUGG.....UGCAGUGAUGACCCC
Tte-1-1 .....AA...UGCA.....GAAGAAA.....UGCUGAUGGCUCC
Cpa-1-1 .....AA...UGCA.....GAAGAAA.....UGCUGAUGGCUCC
Cpa-1-2 .....UAA..CCGCC.....GAAA.....GGCUGAUGGUUCC
Tte-1-2 .....UAA..CCGCC.....GAAA.....GGCUGAUGGUUCC
Cac-1-1 .....UAA..AUAUCC.....UGUUU.....GGAUUGAUGACUUC
Cdi-5-1 AUAA.....UAUCCUAAA..CUGC.....CAAUGA.....GCUGAUGACUUC
Cdi-7-1 AUAA.....UAUCCUAAA..CUGC.....CAAUGA.....GCUGAUGACUUC
Cdi-4-1 AUAA.....UAUCCUAAA..CUGC.....CAAUGA.....GCUGAUGACUUC
Cdi-9-1 AUAA.....UAUCCUAAA..CUGC.....CAAUGA.....GCUGAUGACUUC
Cdi-2-1 AUAA.....UAUCCUAAA..CUGC.....CAAUGA.....GCUGAUGACUUC
Cdi-6-1 AUAA.....UAUCCUAAA..CUGC.....CAAUGA.....GCUGAUGACUUC
Cdi-1-1 AUAA.....UAUCCUAAA..CUGC.....CAAUGA.....GCUGAUGACUUC
Cbu-1-1 AUA.....AUCUAAA..GUGCU.....AUUAA.....CGCUGAUGACUUC
Nth-1-1 UUAAAU.....UCCAAA..CUGCUUU.....CUUA.....GAAGCUAAUGGCUCC
Aor-1-1 AAUA.....GCCUAAA..CCGCA.....AAUA.....UGCUGAUGACUUC
Ceu-1-1 AAAACCUGAAUGGGCA.....A.....CCUAAA..CUGCU.....AAUAU.....AGCUGAUGACUUC
Cba-1-1 GUCAU.....ACUAAAA..CCGCUU.....AUUU.....AAGCUGAUGACUUC
Rob-1-1 UAAAA.....UUACUUAAA..CCGCUU.....AUGA.....AAGCUGAUGACUUC
Dlo-1-1 AUCA.....UACUAGAA..CCGCU.....UAUAA.....AGCUGAUGACUUC
Ahy-1-1 AAAA.....UCUAAAA..CCGCA.....UUUA.....UGCUGAUGACUUC
Cph-1-1 UUU.....ACCUAAA..CUGCU.....UAUAA.....AGCUGAUGACUUC
Cbo-1-1 CUG.....ACUAGAA..CCGCU.....UAUAA.....AGCUGAUGACUUC
Fma-1-1 U.....UAACUAAA..ACGCAA.....AAAUU.....UUGCUGAUGACUUC
Ala-1-1 UUA.....UAAUCUAAA..ACGC.....GUUG.....GCUGAUGACUUC
Chi-1-1 AAA.....CCUAAA..CCGCU.....UAUCG.....AGCUGAUGACUUC
Rin-1-1 UUAAU.....UCGAAA..CCGCUA.....AAUU.....UAGCUGAUGACUUC


Eve-1-1
Csp-2-1
Csa-1-1
Csa-1-2
Tps-1-1
Tsp-2-1
Tsp-2-2
Tps-1-2
Cpa-1-3
Tte-1-3
Tca-1-1
Ame-1-1
Dha-2-1
Ckl-1-1
Cbo-5-1
Cbo-11-1
Cbo-3-1
Cbo-7-1
Cbo-10-1
Cpe-7-1
Cpe-1-1
Cpe-5-1
Cpe-8-1
Cpe-4-1
Cpe-3-1
Cpe-9-1
Cbo-9-1
Cbo-3-2
Cbo-10-2
Cbo-8-1
Cbo-5-2
Cbo-11-2
Csp-3-1
Cbo-6-1
Cth-1-1
Hmo-1-1
Pth-1-1
Chy-1-1
Efa-2-1
Efa-1-1
Lrh-1-1
Lpa-1-1
Lca-1-1
Lca-2-1
Csp-4-1
Lac-1-1
Lhe-1-1
Lga-1-1
Ljo-1-1
Lga-1-2
Ljo-1-2
Lpl-1-1
Efa-3-1
Lca-2-2
Lpa-1-2
Lca-1-2
Lla-3-1

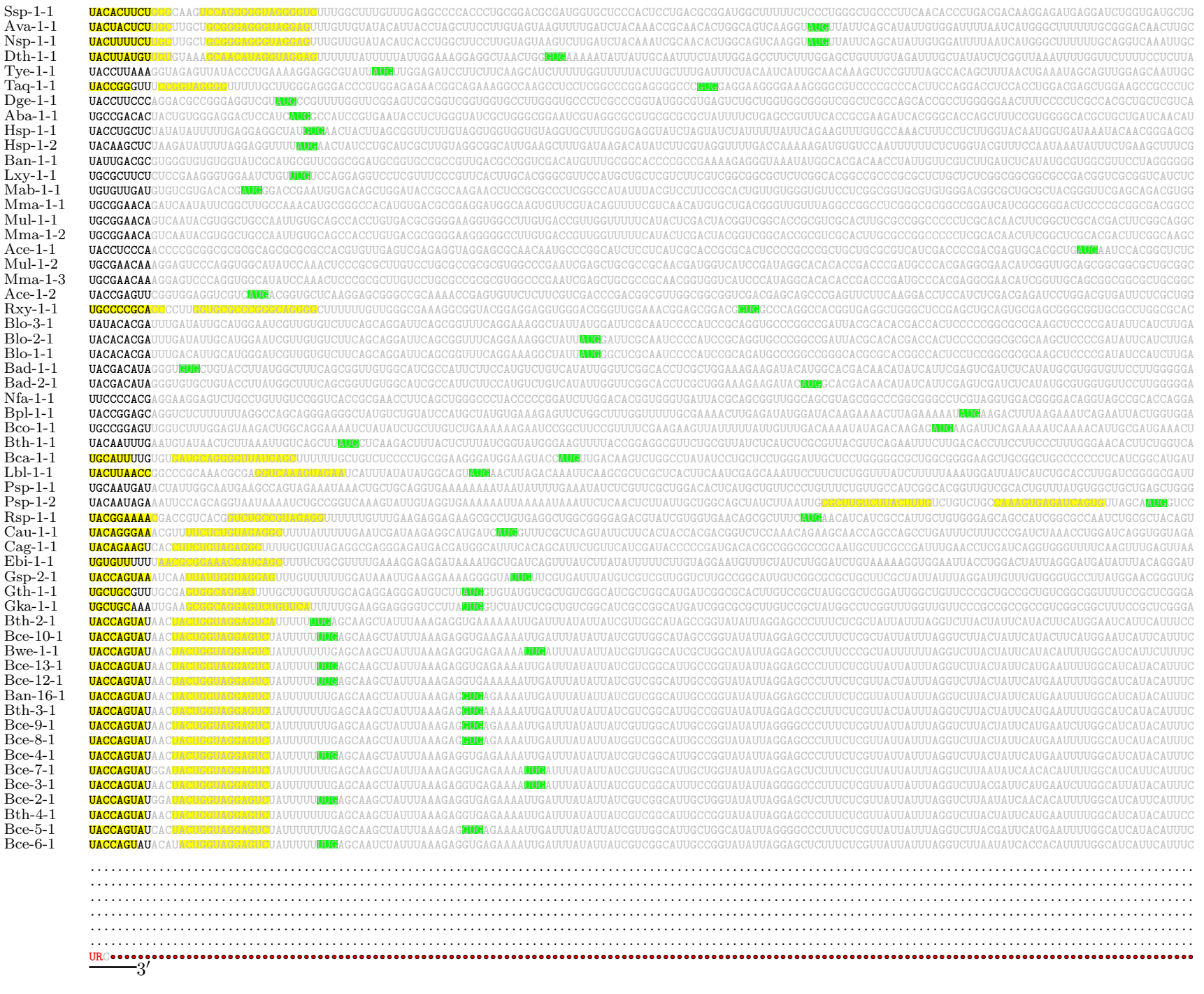

Lla-1-1
Lla-2-1
Lla-3-2
Lla-2-2
Lla-1-2
Zmo-1-1
Gox-1-1
Gdi-1-1
Gdi-1-2
Gdi-1-3
Nha-1-1
Gox-1-2
Oan-1-1
Gox-1-3
Mra-1-1
Bja-1-1
Oca-1-1
Nwi-1-1
Nha-1-2
Nsp-2-1
Rpa-1-1
Rru-1-1
Msp-1-1
Acr-1-1
Acr-1-2
Neu-1-1
Dac-1-1
Cvi-1-1
Dac-1-2
Neu-1-2
Neu-1-3
Bph-2-1
Bxe-1-1
Bph-2-2
Bph-1-1
Bgr-1-1
Bsp-2-1
Bub-1-1
Rpi-2-1
Bmu-1-1
Rso-1-1
Rso-3-1
Rso-2-1
Bce-19-1
Bce-18-1
Bce-15-1
Bam-2-1
Bam-3-1
Bam-1-1
Bam-4-1
Bth-6-1
Bps-14-1
Bps-7-1
Bps-5-1
Bok-2-1
Bps-17-1
Bps-19-1

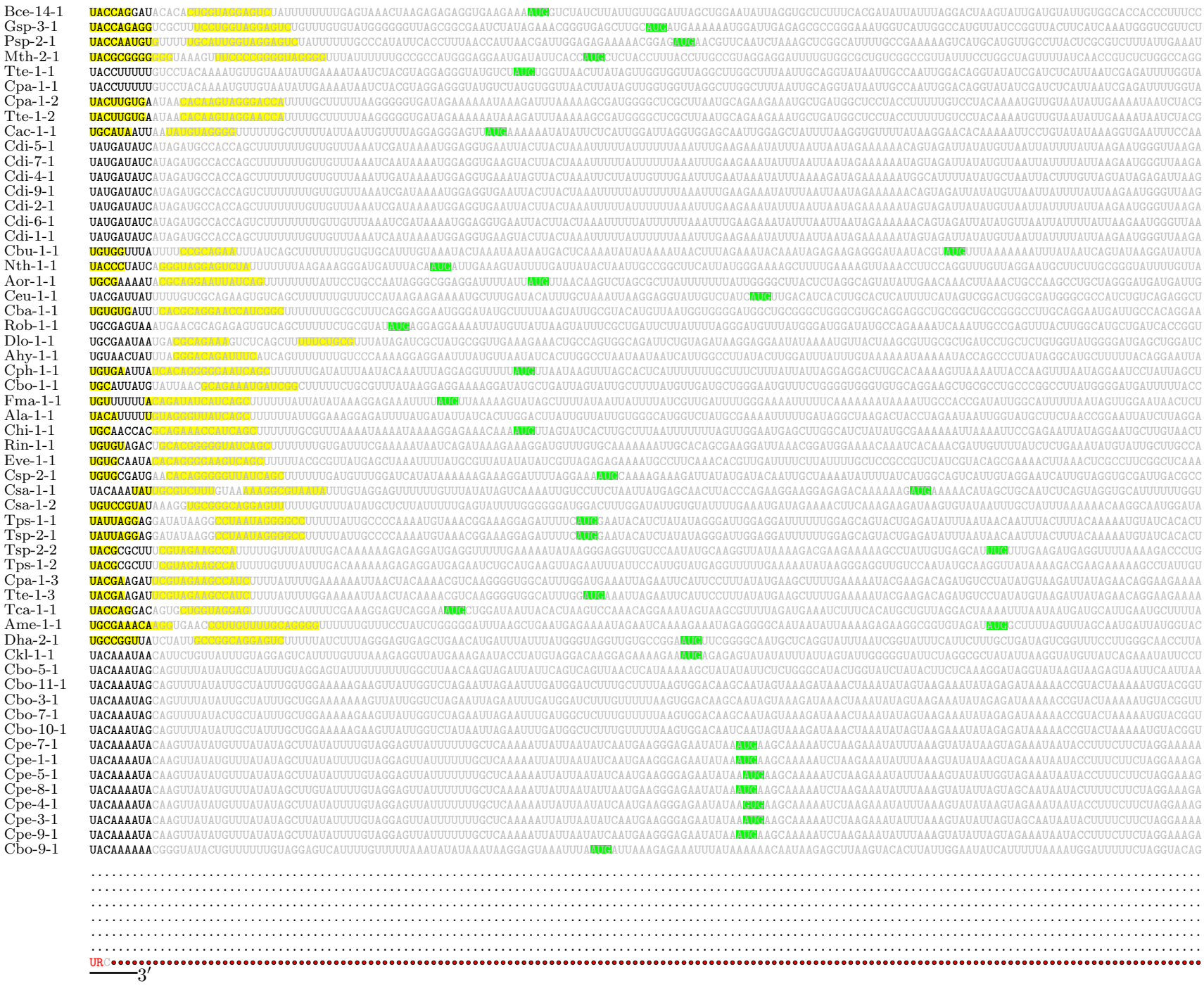

Bth-7-1
Bok-1-1
Bps-16-1
Bps-10-1
Bps-9-1
Bma-2-1
Bps-1-1
Bps-8-1
Bps-18-1
Bth-8-1
Bsp-1-1
Bce-17-1
Bvi-1-1
Bam-2-2
Bam-2-3
Sfu-1-1
Dpi-1-1
Dol-1-1
Dac-2-1
Xca-1-1
Xor-1-1
Xor-4-1
Ppr-1-1
Ppr-2-1
Psp-3-1
Van-1-1
Eta-1-1
Cko-1-1
Vfi-2-1
Vfi-1-1
Sgl-1-1
Csa-2-1
Vca-1-1
Vha-2-1
Vha-1-1
Val-1-1
Afe-1-1
Sma-1-1
Sma-2-1
Vsp-1-1
Vba-1-1
Vsp-2-1
Vfi-1-2
Vfi-2-2
Asa-1-1
Vsh-1-1
Pca-1-1
Pca-2-1
Psy-2-1
Cko-1-2
Psy-1-1
Pae-3-1
Pae-6-1
Pae-5-1
Pae-2-1
Pae-4-1
Pen-1-1

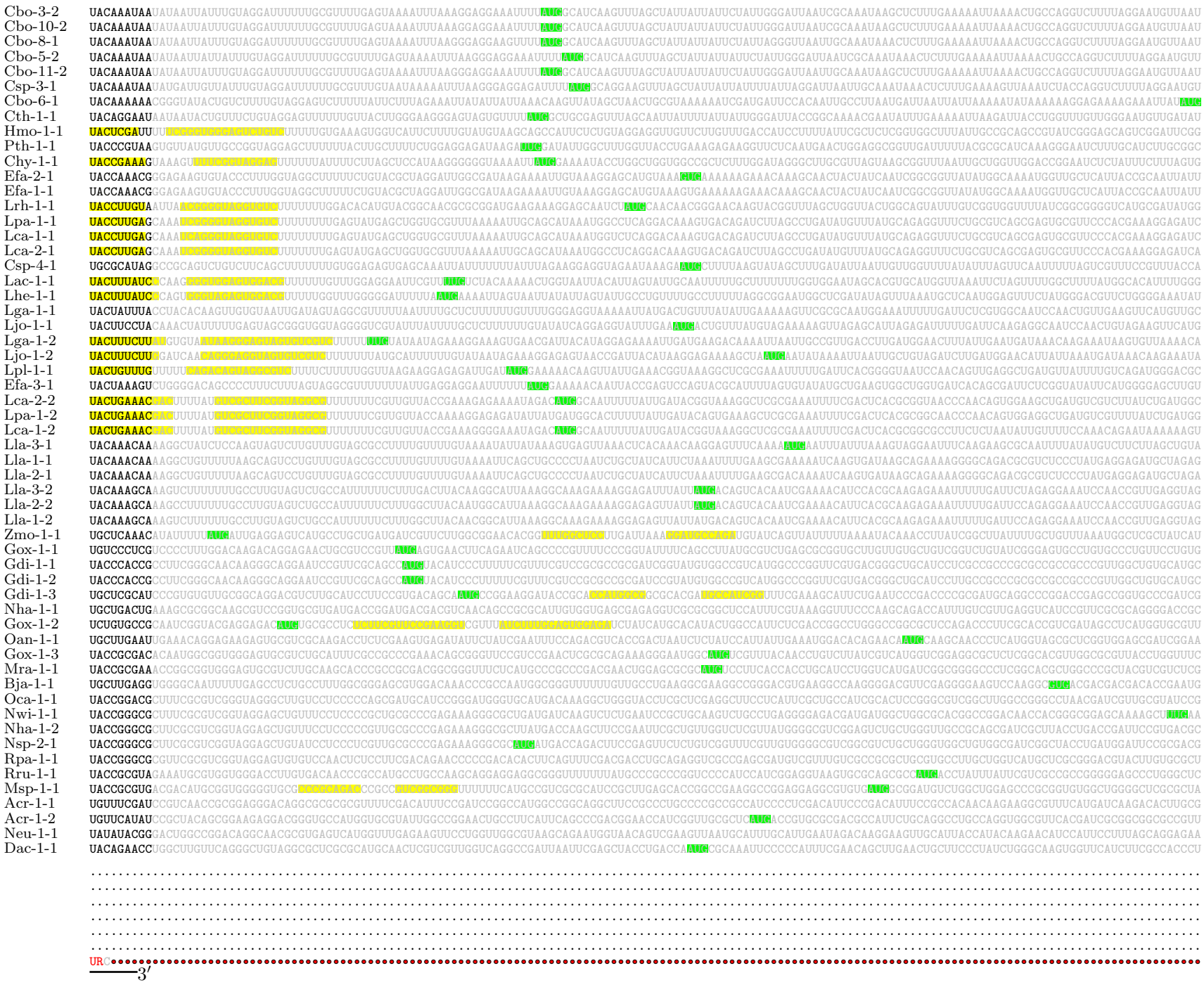

Xca-4-1
Xax-1-1
Pst-1-1
Asp-2-1
Asp-1-1
Aba-3-1
Aba-4-1
Aba-5-1
Ymo-1-1
Yin-1-1
Yps-4-1
Ype-13-1
Yps-2-1
Yps-1-1
Spr-1-1
Pru-1-1
Pal-1-1
Pat-1-1
Cko-1-3
Kpn-1-1
Tpe-1-1
Tle-1-1
Cfl-1-1
Min-1-1
bEl-1-1
Min-1-2
env-1
env-2
env-3
env-4
env-5
env-6
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-15
env-16
env-17
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33

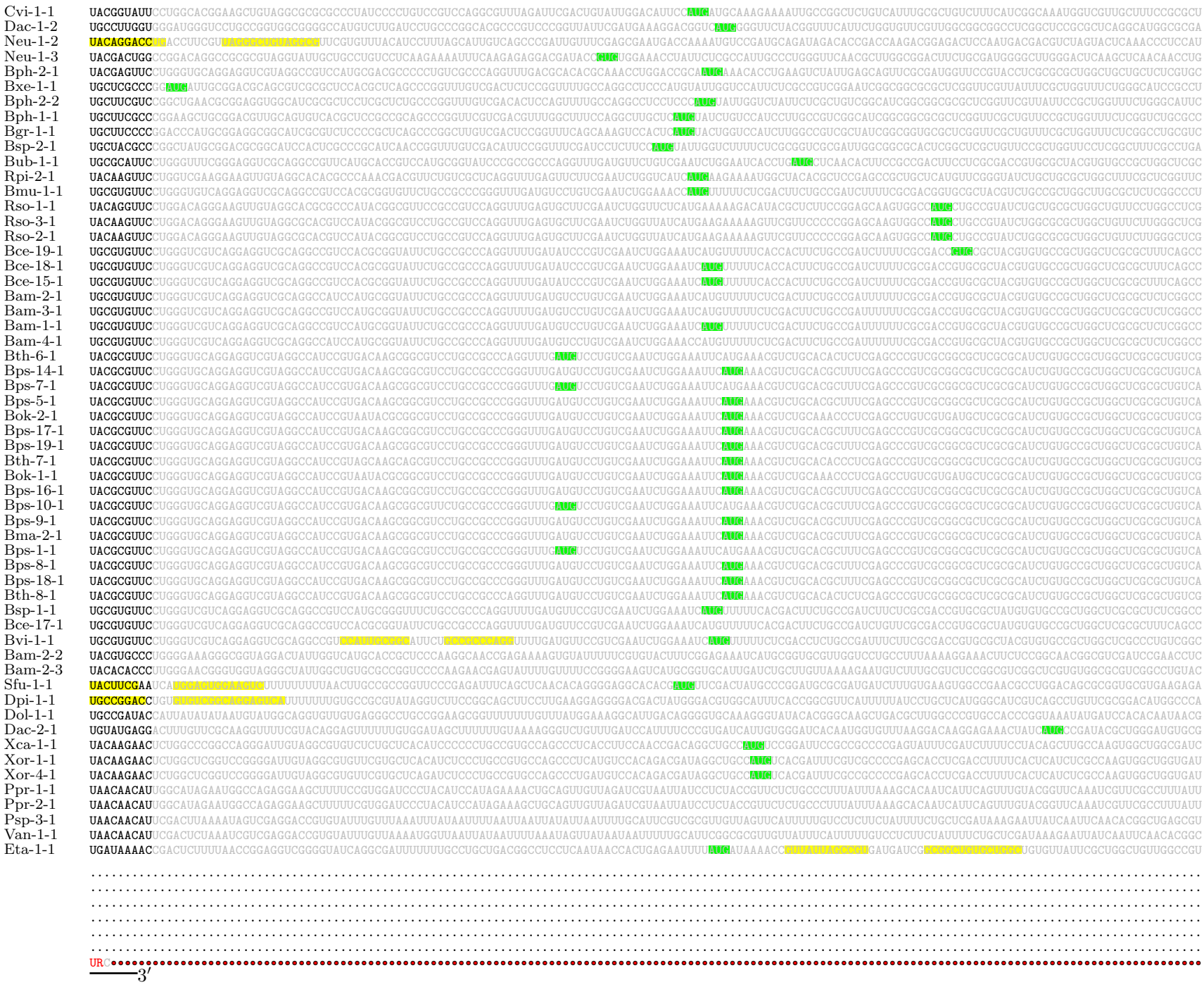


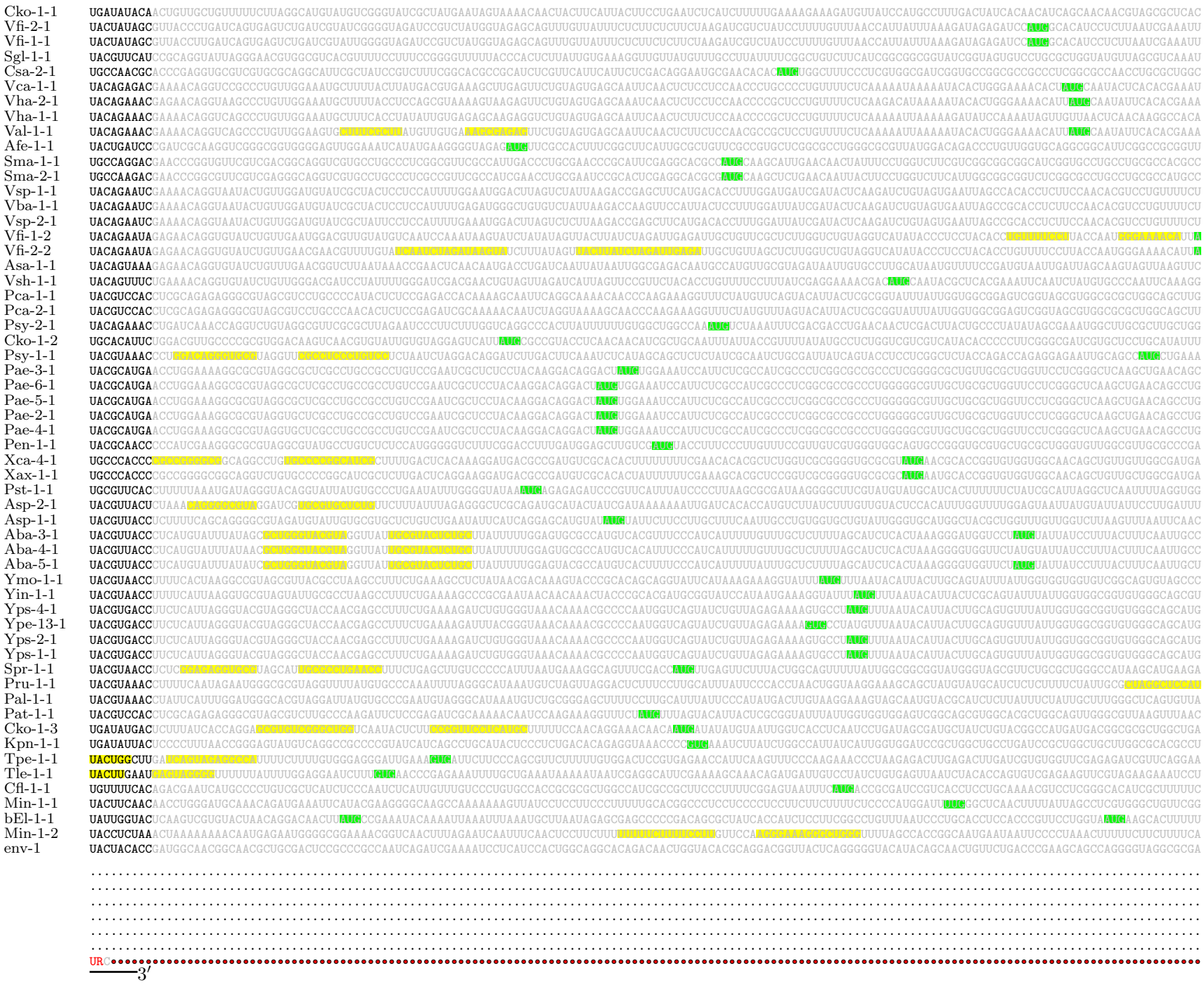
alignment positions 318 · · · 477

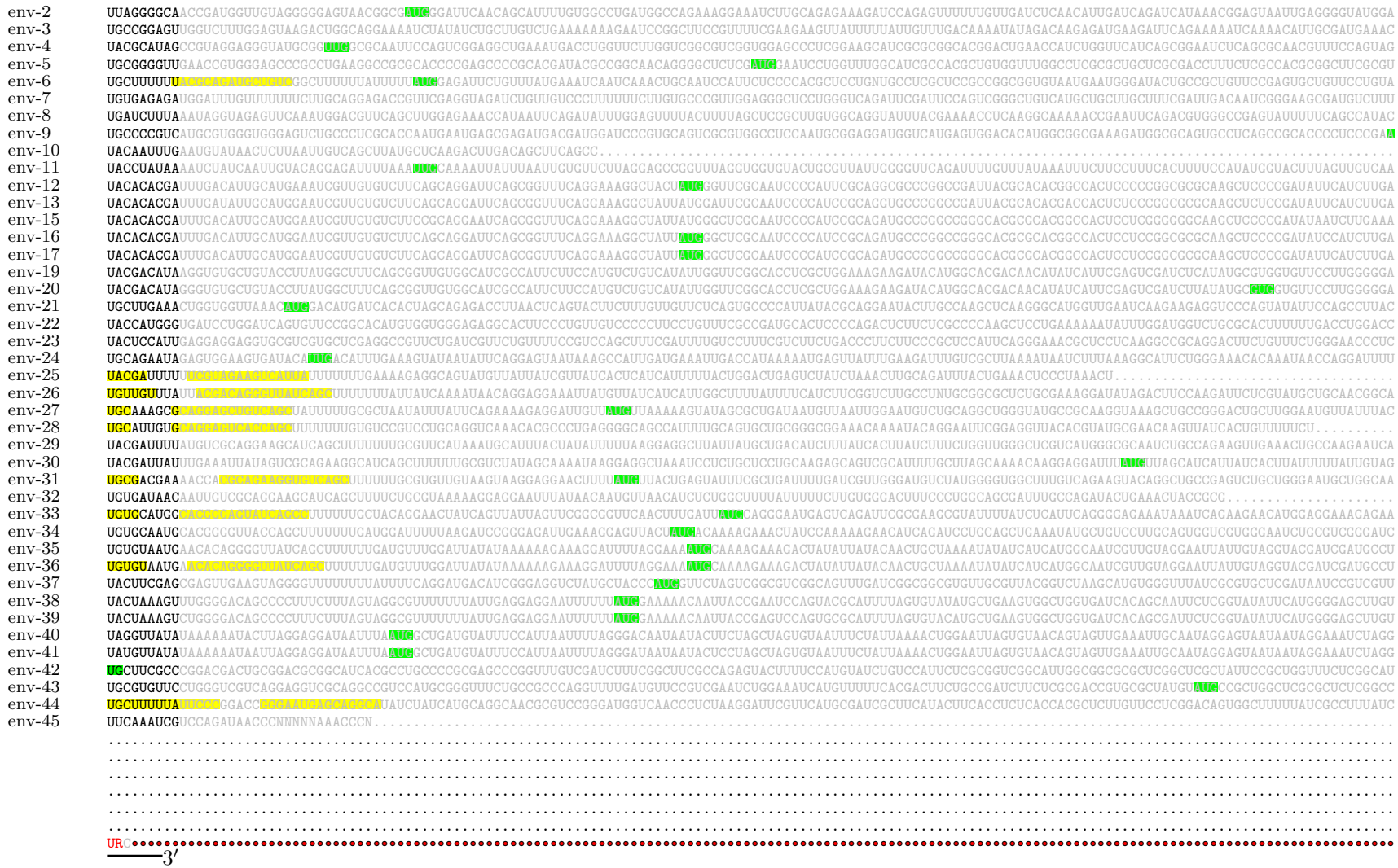

Ssp-1-1
Ava-1-1
Nsp-1-1
Dth-1-1
Tye-1-1
Taq-1-1
Dge-1-1
Aba-1-1
Hsp-1-1
Hsp-1-2
Ban-1-1
Lxy-1-1
Mab-1-1
Mma-1-1
Mul-1-1
Mma-1-2
Ace-1-1
Mul-1-2
Mma-1-3
Ace-1-2
Rxy-1-1
Blo-3-1
Blo-2-1
Blo-1-1
Bad-1-1
Bad-2-1
Nfa-1-1
Bpl-1-1
Bco-1-1
Bth-1-1
Bca-1-1
Lbl-1-1
Psp-1-1
Psp-1-2
Rsp-1-1
Cau-1-1
Cag-1-1
Ebi-1-1
Gsp-2-1
Gth-1-1
Gka-1-1
Bth-2-1
Bce-10-1
Bwe-1-1
Bce-13-1
Bce-12-1
Ban-16-1
Bth-3-1
Bce-9-1
Bce-8-1
Bce-4-1
Bce-7-1
Bce-3-1
Bce-2-1
Bth-4-1
Bce-5-1
Bce-6-1
3′

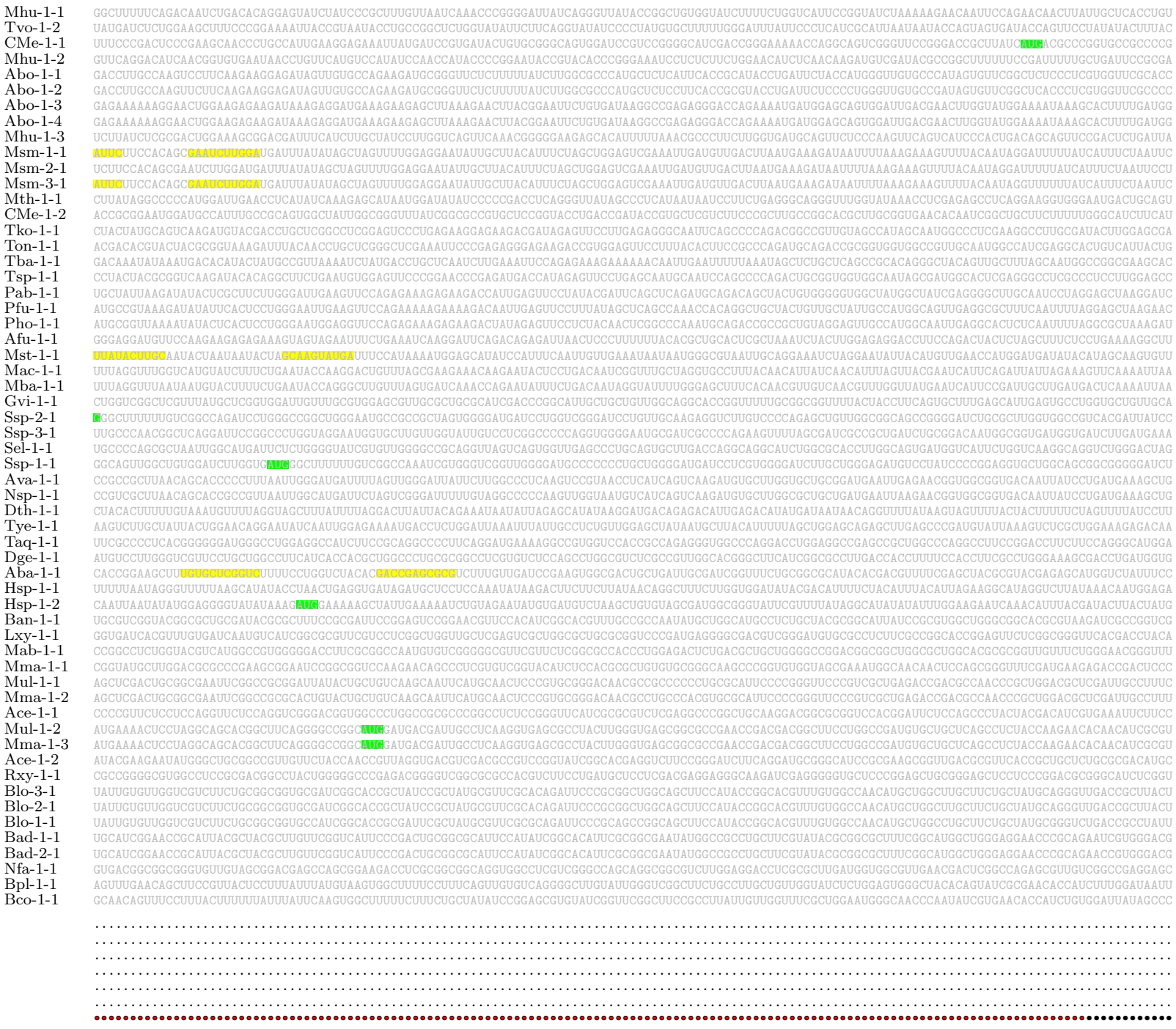

Bce-14-1
Gsp-3-1
Psp-2-1
Mth-2-1
Tte-1-1
Cpa-1-1
Cpa-1-2
Tte-1-2
Cac-1-1
Cdi-5-1
Cdi-7-1
Cdi-4-1
Cdi-9-1
Cdi-2-1
Cdi-6-1
Cdi-1-1
Cbu-1-1
Nth-1-1
Aor-1-1
Ceu-1-1
Cba-1-1
Rob-1-1
Dlo-1-1
Ahy-1-1
Cph-1-1
Cbo-1-1
Fma-1-1
Ala-1-1
Chi-1-1
Rin-1-1
Eve-1-1
Csp-2-1
Csa-1-1
Csa-1-2
Tps-1-1
Tsp-2-1
Tsp-2-2
Tps-1-2
Cpa-1-3
Tte-1-3
Tca-1-1
Ame-1-1
Dha-2-1
Ckl-1-1
Cbo-5-1
Cbo-11-1
Cbo-3-1
Cbo-7-1
Cbo-10-1
Cpe-7-1
Cpe-1-1
Cpe-5-1
Cpe-8-1
Cpe-4-1
Cpe-3-1
Cpe-9-1
Cbo-9-1
3′

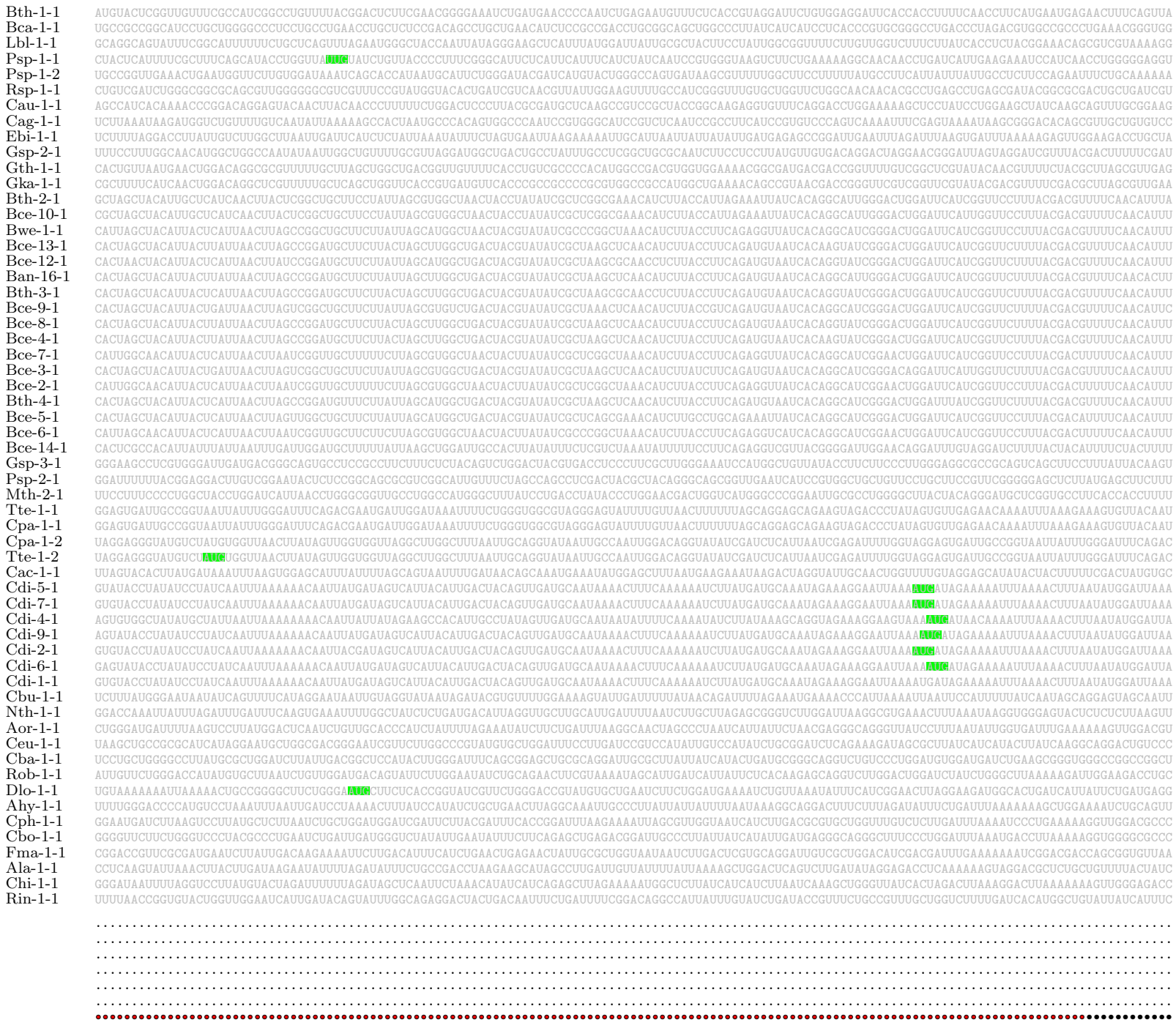

Cbo-3-2
Cbo-10-2
Cbo-8-1
Cbo-5-2
Cbo-11-2
Csp-3-1
Cbo-6-1
Cth-1-1
Hmo-1-1
Pth-1-1
Chy-1-1
Efa-2-1
Efa-1-1
Lrh-1-1
Lpa-1-1
Lca-1-1
Lca-2-1
Csp-4-1
Lac-1-1
Lhe-1-1
Lga-1-1
Ljo-1-1
Lga-1-2
Ljo-1-2
Lpl-1-1
Efa-3-1
Lca-2-2
Lpa-1-2
Lca-1-2
Lla-3-1
Lla-1-1
Lla-2-1
Lla-3-2
Lla-2-2
Lla-1-2
Zmo-1-1
Gox-1-1
Gdi-1-1
Gdi-1-2
Gdi-1-3
Nha-1-1
Gox-1-2
Oan-1-1
Gox-1-3
Mra-1-1
Bja-1-1
Oca-1-1
Nwi-1-1
Nha-1-2
Nsp-2-1
Rpa-1-1
Rru-1-1
Msp-1-1
Acr-1-1
Acr-1-2
Neu-1-1
Dac-1-1
3′


Cvi-1-1
Dac-1-2
Neu-1-2
Neu-1-3
Bph-2-1
Bxe-1-1
Bph-2-2
Bph-1-1
Bgr-1-1
Bsp-2-1
Bub-1-1
Rpi-2-1
Bmu-1-1
Rso-1-1
Rso-3-1
Rso-2-1
Bce-19-1
Bce-18-1
Bce-15-1
Bam-2-1
Bam-3-1
Bam-1-1
Bam-4-1
Bth-6-1
Bps-14-1
Bps-7-1
Bps-5-1
Bok-2-1
Bps-17-1
Bps-19-1
Bth-7-1
Bok-1-1
Bps-16-1
Bps-10-1
Bps-9-1
Bma-2-1
Bps-1-1
Bps-8-1
Bps-18-1
Bth-8-1
Bsp-1-1
Bce-17-1
Bvi-1-1
Bam-2-2
Bam-2-3
Sfu-1-1
Dpi-1-1
Dol-1-1
Dac-2-1
Xca-1-1
Xor-1-1
Xor-4-1
Ppr-1-1
Ppr-2-1
Psp-3-1
Van-1-1
Eta-1-1
3′

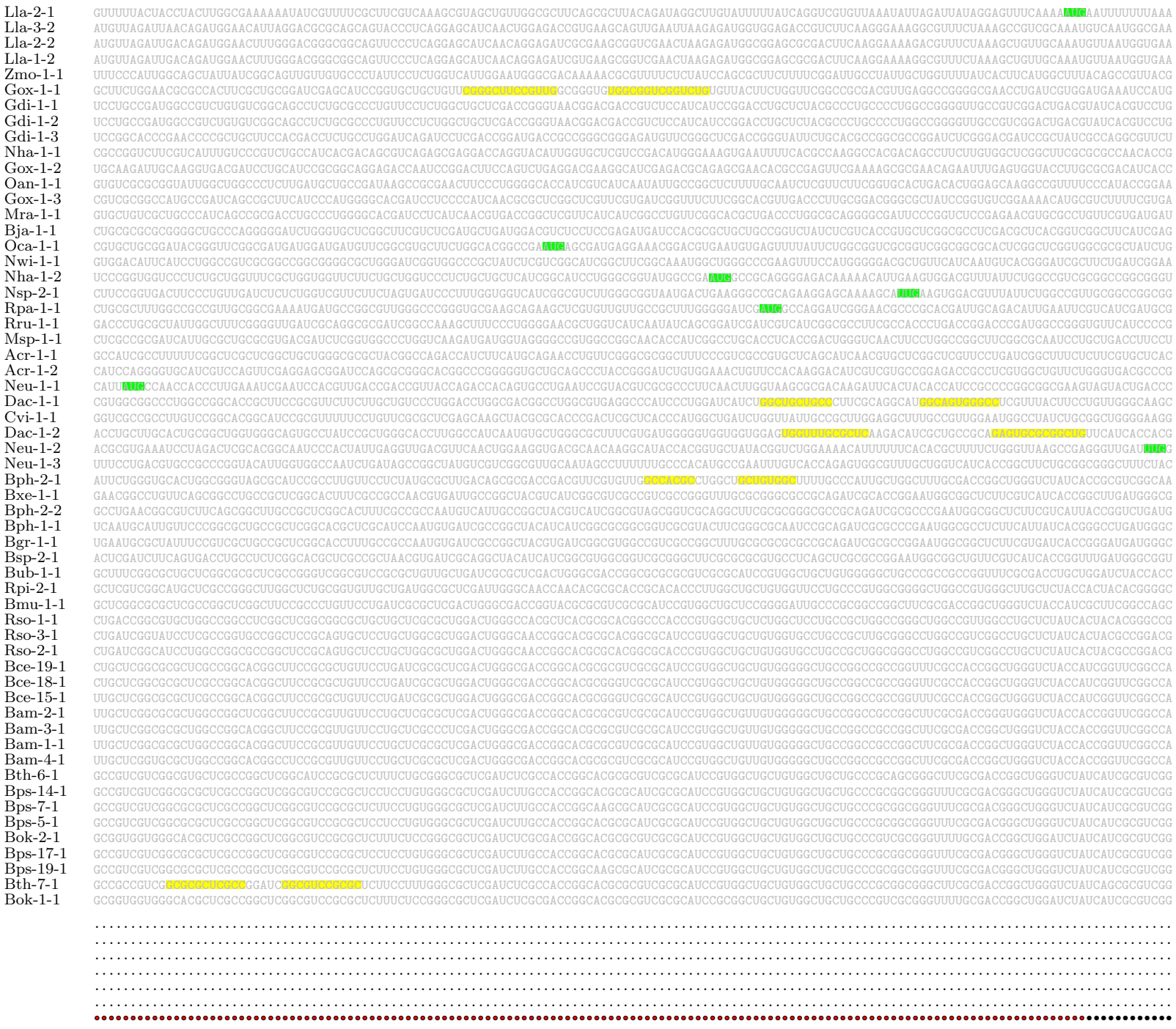

Cko-1-1
Vfi-2-1
Vfi-1-1
Sgl-1-1
Csa-2-1
Vca-1-1
Vha-2-1
Vha-1-1
Val-1-1
Afe-1-1
Sma-1-1
Sma-2-1
Vsp-1-1
Vba-1-1
Vsp-2-1
Vfi-1-2
Vfi-2-2
Asa-1-1
Vsh-1-1
Pca-1-1
Pca-2-1
Psy-2-1
Cko-1-2
Psy-1-1
Pae-3-1
Pae-6-1
Pae-5-1
Pae-2-1
Pae-4-1
Pen-1-1
Xca-4-1
Xax-1-1
Pst-1-1
Asp-2-1
Asp-1-1
Aba-3-1
Aba-4-1
Aba-5-1
Ymo-1-1
Yin-1-1
Yps-4-1
Ype-13-1
Yps-2-1
Yps-1-1
Spr-1-1
Pru-1-1
Pal-1-1
Pat-1-1
Cko-1-3
Kpn-1-1
Tpe-1-1
Tle-1-1
Cfl-1-1
Min-1-1
bEl-1-1
Min-1-2
env-1
3′

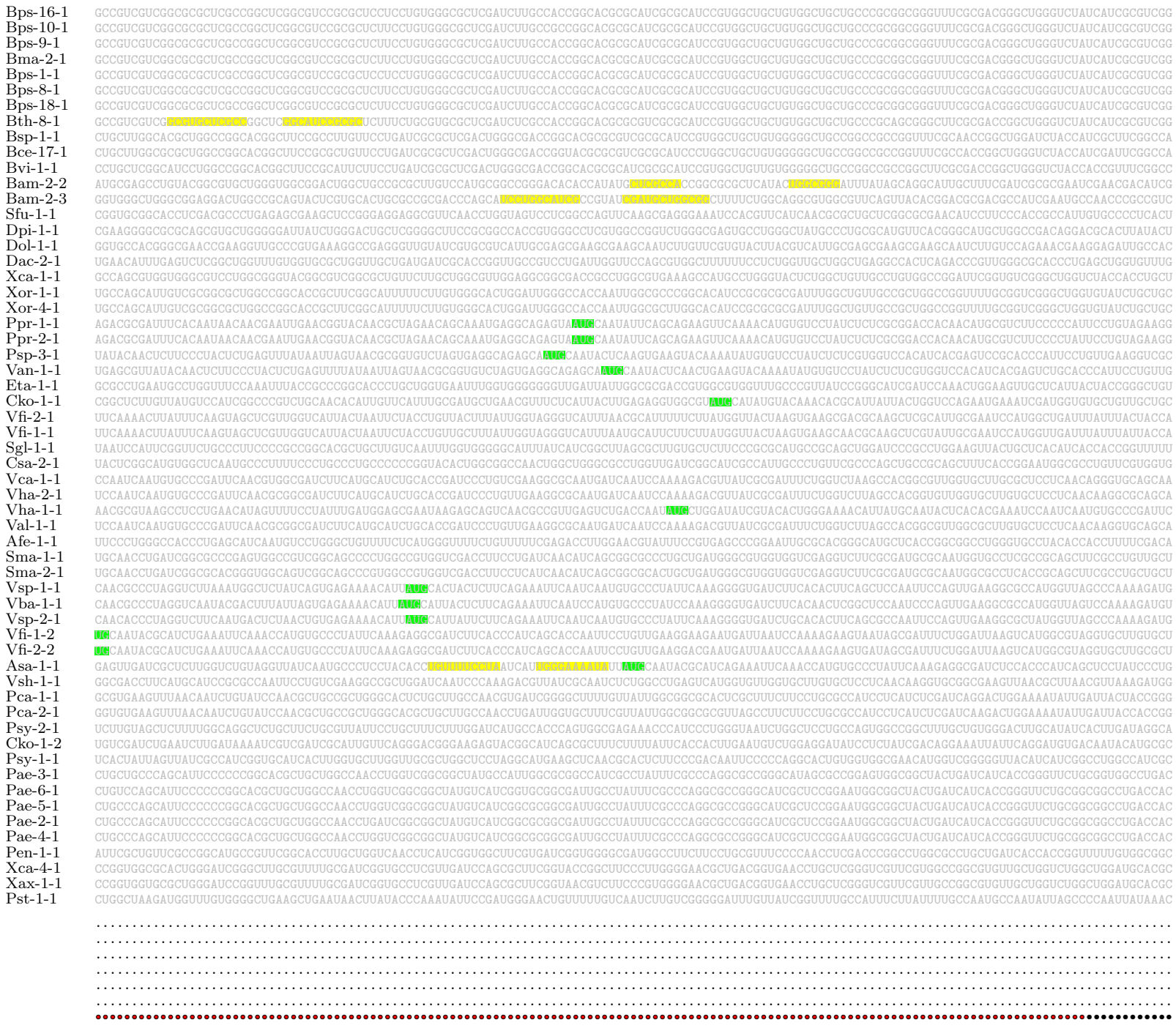

env-2
env-3
env-4
env-5
env-6
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-15
env-16
env-17
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
3′
alignment positions 478 · · · 626
1Sso-1-1
Mse-1-1
Pto-1-1
Tvo-1-1


Mhu-1-1
Tvo-1-2
CMe-1-1
Mhu-1-2
Abo-1-1
Abo-1-2
Abo-1-3
Abo-1-4
Mhu-1-3
Msm-1-1
Msm-2-1
Msm-3-1
Mth-1-1
CMe-1-2
Tko-1-1
Ton-1-1
Tba-1-1
Tsp-1-1
Pab-1-1
Pfu-1-1
Pho-1-1
Afu-1-1
Mst-1-1
Mac-1-1
Mba-1-1
Gvi-1-1
Ssp-2-1
Ssp-3-1
Sel-1-1
Ssp-1-1
Ava-1-1
Nsp-1-1
Dth-1-1
Tye-1-1
Taq-1-1
Dge-1-1
Aba-1-1
Hsp-1-1
Hsp-1-2
Ban-1-1
Lxy-1-1
Mab-1-1
Mma-1-1
Mul-1-1
Mma-1-2
Ace-1-1
Mul-1-2
Mma-1-3
Ace-1-2
Rxy-1-1
Blo-3-1
Blo-2-1
Blo-1-1
Bad-1-1
Bad-2-1
Nfa-1-1
Bpl-1-1
Bco-1-1

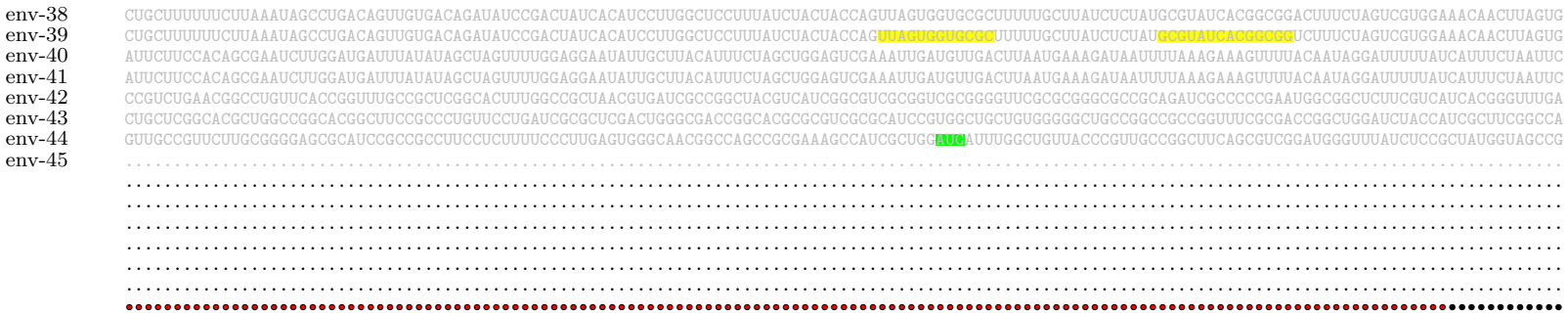

Bth-1-1
Bca-1-1
Lbl-1-1
Psp-1-1
Psp-1-2
Rsp-1-1
Cau-1-1
Cag-1-1
Ebi-1-1
Gsp-2-1
Gth-1-1
Gka-1-1
Bth-2-1
Bce-10-1
Bwe-1-1
Bce-13-1
Bce-12-1
Ban-16-1
Bth-3-1
Bce-9-1
Bce-8-1
Bce-4-1
Bce-7-1
Bce-3-1
Bce-2-1
Bth-4-1
Bce-5-1
Bce-6-1
Bce-14-1
Gsp-3-1
Psp-2-1
Mth-2-1
Tte-1-1
Cpa-1-1
Cpa-1-2
Tte-1-2
Cac-1-1
Cdi-5-1
Cdi-7-1
Cdi-4-1
Cdi-9-1
Cdi-2-1
Cdi-6-1
Cdi-1-1
Cbu-1-1
Nth-1-1
Aor-1-1
Ceu-1-1
Cba-1-1
Rob-1-1
Dlo-1-1
Ahy-1-1
Cph-1-1
Cbo-1-1
Fma-1-1
Ala-1-1
Chi-1-1
Rin-1-1

Eve-1-1
Csp-2-1
Csa-1-1
Csa-1-2
Tps-1-1
Tsp-2-1
Tsp-2-2
Tps-1-2
Cpa-1-3
Tte-1-3
Tca-1-1
Ame-1-1
Dha-2-1
Ckl-1-1
Cbo-5-1
Cbo-11-1
Cbo-3-1
Cbo-7-1
Cbo-10-1
Cpe-7-1
Cpe-1-1
Cpe-5-1
Cpe-8-1
Cpe-4-1
Cpe-3-1
Cpe-9-1
Cbo-9-1
Cbo-3-2
Cbo-10-2
Cbo-8-1
Cbo-5-2
Cbo-11-2
Csp-3-1
Cbo-6-1
Cth-1-1
Hmo-1-1
Pth-1-1
Chy-1-1
Efa-2-1
Efa-1-1
Lrh-1-1
Lpa-1-1
Lca-1-1
Lca-2-1
Csp-4-1
Lac-1-1
Lhe-1-1
Lga-1-1
Ljo-1-1
Lga-1-2
Ljo-1-2
Lpl-1-1
Efa-3-1
Lca-2-2
Lpa-1-2
Lca-1-2
Lla-3-1
Lla-1-1


Lla-2-1
Lla-3-2
Lla-2-2
Lla-1-2
Zmo-1-1
Gox-1-1
Gdi-1-1
Gdi-1-2
Gdi-1-3
Nha-1-1
Gox-1-2
Oan-1-1
Gox-1-3
Mra-1-1
Bja-1-1
Oca-1-1
Nwi-1-1
Nha-1-2
Nsp-2-1
Rpa-1-1
Rru-1-1
Msp-1-1
Acr-1-1
Acr-1-2
Neu-1-1
Dac-1-1
Cvi-1-1
Dac-1-2
Neu-1-2
Neu-1-3
Bph-2-1
Bxe-1-1
Bph-2-2
Bph-1-1
Bgr-1-1
Bsp-2-1
Bub-1-1
Rpi-2-1
Bmu-1-1
Rso-1-1
Rso-3-1
Rso-2-1
Bce-19-1
Bce-18-1
Bce-15-1
Bam-2-1
Bam-3-1
Bam-1-1
Bam-4-1
Bth-6-1
Bps-14-1
Bps-7-1
Bps-5-1
Bok-2-1
Bps-17-1
Bps-19-1
Bth-7-1
Bok-1-1


Bps-16-1
Bps-10-1
Bps-9-1
Bma-2-1
Bps-1-1
Bps-8-1
Bps-18-1
Bth-8-1
Bsp-1-1
Bce-17-1
Bvi-1-1
Bam-2-2
Bam-2-3
Sfu-1-1
Dpi-1-1
Dol-1-1
Dac-2-1
Xca-1-1
Xor-1-1
Xor-4-1
Ppr-1-1
Ppr-2-1
Psp-3-1
Van-1-1
Eta-1-1
Cko-1-1
Vfi-2-1
Vfi-1-1
Sgl-1-1
Csa-2-1
Vca-1-1
Vha-2-1
Vha-1-1
Val-1-1
Afe-1-1
Sma-1-1
Sma-2-1
Vsp-1-1
Vba-1-1
Vsp-2-1
Vfi-1-2
Vfi-2-2
Asa-1-1
Vsh-1-1
Pca-1-1
Pca-2-1
Psy-2-1
Cko-1-2
Psy-1-1
Pae-3-1
Pae-6-1
Pae-5-1
Pae-2-1
Pae-4-1
Pen-1-1
Xca-4-1
Xax-1-1
Pst-1-1

Asp-2-1
Asp-1-1
Aba-3-1
Aba-4-1
Aba-5-1
Ymo-1-1
Yin-1-1
Yps-4-1
Ype-13-1
Yps-2-1
Yps-1-1
Spr-1-1
Pru-1-1
Pal-1-1
Pat-1-1
Cko-1-3
Kpn-1-1
Tpe-1-1
Tle-1-1
Cfl-1-1
Min-1-1
bEl-1-1
Min-1-2
env-1
env-2
env-3
env-4
env-5
env-6
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-15
env-16
env-17
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37



25 Cyano-1

25.1 Taxa

The taxonomy of each organism containing a putative Cyano-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Cyano-1 RNA in Sections 25.2 and 25.4.

abbrev. of hits	**taxonomy of species**
Ama-1-1	Bacteria Cyanobacteria *Acaryochloris marina* MBIC11017
Cwa-1-1 to Cwa-1-2	Bacteria Cyanobacteria Chroococcales *Crocosphaera watsonii* WH 8501 (Synechocystis sp. WH 8501)
Csp-1-1	Bacteria Cyanobacteria Chroococcales *Cyanobium sp.* PCC 7001
Csp-2-1 to Csp-2-2	Bacteria Cyanobacteria Chroococcales *Cyanothece sp.* ATCC 51142
Csp-3-1 to Csp-3-2	Bacteria Cyanobacteria Chroococcales *Cyanothece sp.* CCY 0110
Csp-4-1	Bacteria Cyanobacteria Chroococcales *Cyanothece sp.* PCC 7424
Csp-5-1 to Csp-5-2	Bacteria Cyanobacteria Chroococcales *Cyanothece sp.* PCC 7822
Csp-6-1	Bacteria Cyanobacteria Chroococcales *Cyanothece sp.* PCC 8801
Csp-7-1	Bacteria Cyanobacteria Chroococcales *Cyanothece sp.* PCC 8802
Mae-1-1 to Mae-1-2	Bacteria Cyanobacteria Chroococcales *Microcystis aeruginosa* NIES-843
Sel-1-1	Bacteria Cyanobacteria Chroococcales *Synechococcus elongatus* PCC 6301
Sel-2-1	Bacteria Cyanobacteria Chroococcales *Synechococcus elongatus* PCC 7942
Ssp-1-1 to Ssp-1-2	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* BL107
Ssp-2-1 to Ssp-2-10	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* CC9311
Ssp-3-1 to Ssp-3-5	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* CC9605
Ssp-4-1 to Ssp-4-3	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* CC9902
Ssp-5-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* JA-2-3B'a(2-13)
Ssp-6-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* JA-3-3Ab
Ssp-7-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* RCC307
Ssp-8-1 to Ssp-8-2	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* RS9916
Ssp-9-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* RS9917
Ssp-10-1 to Ssp-10-2	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 5701
Ssp-11-1 to Ssp-11-2	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 7803
Ssp-12-1 to Ssp-12-2	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 7805
Ssp-13-1 to Ssp-13-4	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 8102
Ssp-14-1 to Ssp-14-3	Bacteria Cyanobacteria Chroococcales *Synechocystis sp.* PCC 6803
Tel-1-1 to Tel-1-2	Bacteria Cyanobacteria Chroococcales *Thermosynechococcus elongatus* BP-1
Ava-1-1 to Ava-1-3	Bacteria Cyanobacteria Nostocales Nostocaceae *Anabaena variabilis* ATCC 29413 (Anabaena flos-aquae UTEX 1444)
Npu-1-1	Bacteria Cyanobacteria Nostocales Nostocaceae *Nostoc punctiforme* PCC 73102 (Nostoc punctiforme ATCC 29133)
Nsp-1-1 to Nsp-1-5	Bacteria Cyanobacteria Nostocales Nostocaceae *Nostoc sp.* PCC 7120
Lsp-1-1	Bacteria Cyanobacteria Oscillatoriales *Lyngbya sp.* PCC 8106
Ter-1-1	Bacteria Cyanobacteria Oscillatoriales *Trichodesmium erythraeum* IMS101
Pma-1-1	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* AS9601
Pma-2-1	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9211

Pma-3-1	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9215
Pma-4-1 to Pma-4-2	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9301
Pma-5-1 to Pma-5-2	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9303
Pma-6-1 to Pma-6-2	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9312
Pma-7-1 to Pma-7-2	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9313
Pma-8-1 to Pma-8-2	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9515
Pma-9-1 to Pma-9-4	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* NATL1A
Pma-10-1 to Pma-10-4	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* NATL2A
Pma-11-1 to Pma-11-3	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus subsp. marinus str.* CCMP1375
Pma-12-1 to Pma-12-4	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus subsp. pastoris str.* CCMP1986
env-1 to env-179	environmental samples

25.2 Gene contexts

Each Cyano-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Cyano-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 25.3. The accession number of the sequence containing each Cyano-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 25.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Ssp-14-1	NC_000911.1	-	1558977	1558901	TxlA (cd02950)→ **RNA**→ hypo→
Sel-1-1	NC_006576.1	+	483196	483269	hypo→ **RNA**→ COG0398 (COG0398)→
Sel-2-1	NC_007604.1	-	1137608	1137535	hypo→ **RNA**→ COG0398 (COG0398)→
Ama-1-1	NC_009925.1	-	407448	407375	←tRNA-Pro **RNA**→ ←SRP RNA
Ter-1-1	NC_008312.1	-	4576095	4576023	MhpC (COG0596)→ **RNA**→ ←hypo
Tel-1-1	NC_004113.1	+	599896	599961	hypo→ **RNA**→ COG0606 (COG0606)→
Ava-1-1	NC_007413.1	+	2186368	2186439	tRNA-synt_1d_C (pfam05746)NOG25202 (NOG25202)→ **RNA**→ ←Cof (COG0561)
Nsp-1-1	NC_003272.1	-	4768549	4768478	tRNA-synt_1d_C (pfam05746)NOG25202 (NOG25202)→ **RNA**→ ←Cof (COG0561)
Pma-9-1	NC_008819.1	+	401851	401916	hypo→ **RNA**→ hypo→
Pma-2-1	NC_009976.1	+	1238813	1238882	←hypo **RNA**→ ←beta-lactamase
Pma-11-1	NC_005042.1	+	1298860	1298930	←hypo **RNA**→ ←hypo
Pma-11-2	NC_005042.1	-	636238	636172	←hypo **RNA**→ hypo→
Pma-11-3	NC_005042.1	-	556707	556642	hypo→ **RNA**→ Ribosomal_S21 (pfam01165)→
env-1	JCVI_SCAF_1096628057746	-	1221	1165	? (GOS11385446)→ **RNA**→ ? (GOS11326834)→
Ssp-13-1	NC_005070.1	-	1411144	1411088	high light inducible protein→ **RNA**→ hypo→
Ssp-14-2	NC_000911.1	-	3398354	3398286	hypo→ **RNA**→ COG4449 (COG4449)→
Ssp-8-1	NZ_AAUA01000001.1	+	379823	379889	hypo→ **RNA**→ ←FolC (COG0285)
env-2	JCVI_SCAF_1101668398476	-	679	615	**RNA**→ ←? (GOS12835)
env-3	JCVI_SCAF_1096627574863	+	1687	1755	? (GOS303)→ **RNA**→
Pma-6-1	NC_007577.1	-	1173349	1173281	tRNA-Ser→ **RNA**→ ←fer2 (cd00207)
Pma-8-1	NC_008817.1	+	45910	45978	←hypo **RNA**→ hypo→
env-4	JCVI_SCAF_1096626304682	-	260	197	? (GOS11457573)→ **RNA**→
env-5	JCVI_SCAF_1101668388273	-	763	700	**RNA**→ ←? (GOS12835)
env-6	JCVI_SCAF_1096627816644	-	901	832	**RNA**→ adh_short (pfam00106)? (GOS49)→
env-7	JCVI_SCAF_1096627600585	+	525	594	? (GOS11661723)→ **RNA**→ ←CobQ (COG1492)? (GOS3592)
Pma-12-1	NC_005072.1	-	972180	972111	←hypo **RNA**→ ←inward rectifier potassium channel

env-8	JCVI_SCAF_1096627537559	+	1508	1575	←? (GOS12183470) **RNA**→
env-9	JCVI_SCAF_1096626998670	-	696	627	Ser tRNA→ **RNA**→
env-10	JCVI_SCAF_1096627551526	-	1309	1240	**RNA**→ ←fer2 (cd00207)? (GOS9508)
env-11	JCVI_SCAF_1096627632931	-	234	167	? (GOS11053562)→ **RNA**→
env-12	JCVI_SCAF_1096627726923	+	1602	1671	? (GOS11661723)→ **RNA**→
Pma-12-2	NC_005072.1	+	383387	383453	←hypo **RNA**→ AceF (COG0508)→
env-13	JCVI_SCAF_1096627084234	+	850	917	←? (GOS11973280) **RNA**→ ←Asn tRNA
env-14	JCVI_SCAF_1101668306973	+	1040	1109	←? (GOS11385352) **RNA**→
env-15	JCVI_SCAF_1096627155230	+	115	182	**RNA**→ FpaA (COG0426)Flavin_Reduct (pfam01613)? (GOS395)→
env-16	JCVI_SCAF_1101668274677	-	485	417	? (GOS303)→ **RNA**→
env-17	JCVI_SCAF_1096628045981	+	1113	1182	←PiuC (COG3128)? (GOS1330) **RNA**→
Pma-12-3	NC_005072.1	+	346826	346893	HTH_MARR (smart00347)→ **RNA**→ malic enzyme→
env-18	JCVI_SCAF_1096628123936	+	329	396	**RNA**→ ←? (GOS13019159)
env-19	JCVI_SCAF_1101668304461	+	1115	1182	? (GOS303)→ **RNA**→
env-20	JCVI_SCAF_1096627549615	-	1297	1230	**RNA**→ ←CobQ (COG1492)? (GOS3592)
Pma-12-4	NC_005072.1	+	654509	654576	hypo→ **RNA**→ ←hypo
env-21	JCVI_SCAF_1096626931145	-	667	600	? (GOS303)→ **RNA**→ ? (GOS10697)→
env-22	JCVI_SCAF_1096627548057	+	1354	1423	? (GOS11661723)→ **RNA**→
env-23	JCVI_SCAF_1096628122610	+	1217	1286	Ser tRNA→ **RNA**→ ? (GOS10697)→
env-24	JCVI_SCAF_1096626989098	+	1363	1432	Rhomboid (pfam01694)? (GOS303)→ **RNA**→ ←fer2 (cd00207)? (GOS9508)
env-25	JCVI_SCAF_1096628273739	+	691	760	Rhomboid (pfam01694)? (GOS303)→ **RNA**→ ? (GOS13095788)→
env-26	JCVI_SCAF_1096627553294	-	1215	1145	**RNA**→ ←fer2 (cd00207)? (GOS9508)
env-27	JCVI_SCAF_1101668225625	+	844	913	**RNA**→ ? (GOS10697)→
env-28	JCVI_SCAF_1096626118376	+	371	440	**RNA**→
env-29	JCVI_SCAF_1101667201924	+	344	413	**RNA**→
env-30	JCVI_SCAF_1096626580640	-	77	10	? (GOS11661723)→ **RNA**→
env-31	JCVI_SCAF_1096627767908	-	556	489	Ser tRNA→ **RNA**→
env-32	JCVI_SCAF_1096627269409	-	1686	1617	? (GOS11661723)→ **RNA**→ ? (GOS10697)→
env-33	JCVI_SCAF_1096627581678	+	655	724	←? (GOS7134) **RNA**→ ←? (GOS1694)
env-34	JCVI_SCAF_1096627311661	-	1407	1338	Ser tRNA→ **RNA**→ ? (GOS10697)→
env-35	JCVI_SCAF_1096627591944	-	1486	1417	**RNA**→ SRA (smart00466)? (GOS2778)→
env-36	JCVI_SCAF_1101668221336	-	295	229	←? (GOS11385352) **RNA**→
env-37	JCVI_SCAF_1101667545738	+	700	769	Rhomboid (pfam01694)? (GOS303)→ **RNA**→
env-38	JCVI_SCAF_1096628120170	+	629	696	**RNA**→ ←? (GOS13534)
Pma-4-1	NC_009091.1	-	133115	133048	←hypo **RNA**→ ←tRNA-Asn
Pma-8-2	NC_008817.1	-	977450	977385	←hypo **RNA**→ ←inward rectifier potassium channel
env-39	JCVI_SCAF_1096627722494	+	1168	1235	←? (GOS11953) **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-40	JCVI_SCAF_1096627764602	+	511	579	18S_rRNA→ **RNA**→ ? (GOS15049)→
env-41	JCVI_SCAF_1096626925644	+	987	1054	18S_rRNA→ **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-42	JCVI_SCAF_1096627569082	-	1103	1036	? (GOS11907928)→ **RNA**→ ? (GOS15049)→
env-43	JCVI_SCAF_1096628094227	-	1965	1896	**RNA**→ DUF1499 (pfam07386)? (GOS7536)→
Pma-4-2	NC_009091.1	-	949300	949233	←hypo **RNA**→ DUF1499 (pfam07386)→
env-44	JCVI_SCAF_1096627732672	+	850	917	18S_rRNA→ **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-45	JCVI_SCAF_1096626922070	+	106	173	**RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-46	JCVI_SCAF_1101667367890	+	504	571	**RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-47	JCVI_SCAF_1096626975232	-	2089	2022	**RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-48	JCVI_SCAF_1101668301977	+	598	665	←18S_rRNA **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-49	JCVI_SCAF_1096626117883	+	357	425	18S_rRNA→ **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-50	JCVI_SCAF_1096627546133	+	1102	1169	←18S_rRNA **RNA**→ ? (GOS7536)→
env-51	JCVI_SCAF_1096626991635	-	2734	2667	? (GOS11599787)→ **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-52	JCVI_SCAF_1096627563490	-	954	887	←? (GOS19324) **RNA**→ ? (GOS15049)→
env-53	JCVI_SCAF_1096627520985	-	414	347	**RNA**→ ? (GOS7536)→

env-54	JCVI_SCAF_1096627549356	+	1513	1580	18S_rRNA→ **RNA**→
env-55	JCVI_SCAF_1096627766018	+	1345	1412	18S_rRNA→ **RNA**→ ? (GOS7536)→
env-56	JCVI_SCAF_1101667117829	+	720	787	18S_rRNA→ **RNA**→
env-57	JCVI_SCAF_1101667100941	+	211	278	**RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-58	JCVI_SCAF_1101668200313	+	1340	1407	18S_rRNA→ **RNA**→
env-59	JCVI_SCAF_1096627545220	-	403	336	←18S_rRNA **RNA**→
Pma-3-1	NC_009840.1	-	982248	982181	←hypo **RNA**→ DUF1499 (pfam07386)→
env-60	JCVI_SCAF_1096627622873	-	1456	1389	**RNA**→ COG4446 (COG4446)? (GOS7536)→
env-61	JCVI_SCAF_1096627530590	-	349	282	←? (GOS19324) **RNA**→
env-62	JCVI_SCAF_1101667146941	+	61	128	**RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-63	JCVI_SCAF_1096627062687	+	785	852	18S_rRNA→ **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-64	JCVI_SCAF_1101668311145	-	1502	1435	**RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-65	JCVI_SCAF_1096627001989	-	749	682	←? (GOS19324) **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-66	JCVI_SCAF_1096627705637	-	781	714	←? (GOS11953) **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-67	JCVI_SCAF_1096626999505	+	210	277	**RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-68	JCVI_SCAF_1101667152469	+	849	916	18S_rRNA→ **RNA**→
env-69	JCVI_SCAF_1096628295458	-	641	574	ABC_Metallic_Cations (cd03235)? (GOS159)→ **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-70	JCVI_SCAF_1101668293418	-	927	860	←? (GOS19324) **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-71	JCVI_SCAF_1096628073177	+	535	602	18S_rRNA→ **RNA**→ ? (GOS7536)→
env-72	JCVI_SCAF_1101668290942	-	1389	1322	←18S_rRNA **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-73	JCVI_SCAF_1096627562179	+	256	323	**RNA**→ ←OsmC (pfam02566)? (GOS10947)
env-74	JCVI_SCAF_1101668205915	-	398	331	←? (GOS19324) **RNA**→ ? (GOS7536)→
env-75	JCVI_SCAF_1096628227480	+	1746	1813	18S_rRNA→ **RNA**→
env-76	JCVI_SCAF_1101667224325	-	444	377	←18S_rRNA **RNA**→ ? (GOS7536)→
Pma-6-2	NC_007577.1	-	952144	952077	←hypo **RNA**→ DUF1499 (pfam07386)→
env-77	JCVI_SCAF_1101668197125	-	796	729	←? (GOS19324) **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-78	JCVI_SCAF_1096627061123	-	706	639	←18S_rRNA **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-79	JCVI_SCAF_1096627755302	-	374	307	←? (GOS19324) **RNA**→
env-80	JCVI_SCAF_1096627541626	+	556	623	? (GOS11907928)→ **RNA**→ ? (GOS7536)→
env-81	JCVI_SCAF_1096626485580	+	500	567	18S_rRNA→ **RNA**→
env-82	JCVI_SCAF_1096628286184	+	1714	1781	18S_rRNA→ **RNA**→
env-83	JCVI_SCAF_1096627758008	+	452	519	18S_rRNA→ **RNA**→ ? (GOS15049)→
env-84	JCVI_SCAF_1101668001054	+	132	199	**RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-85	JCVI_SCAF_1096627742934	-	1053	986	←? (GOS19324) **RNA**→
env-86	JCVI_SCAF_1096627739148	+	442	509	18S_rRNA→ **RNA**→ ? (GOS15049)→
env-87	JCVI_SCAF_1096627740395	-	1423	1356	←18S_rRNA **RNA**→ COG4446 (COG4446)? (GOS7536)→
env-88	JCVI_SCAF_1096627055327	-	1621	1554	**RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-89	JCVI_SCAF_1096627713867	+	632	699	18S_rRNA→ **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-90	JCVI_SCAF_1096627048952	+	1057	1124	18S_rRNA→ **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-91	JCVI_SCAF_1101668280001	-	1609	1542	**RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-92	JCVI_SCAF_1096626959190	+	716	783	18S_rRNA→ **RNA**→ COG4446 (COG4446)? (GOS7536)→
env-93	JCVI_SCAF_1096626968620	-	87	20	? (GOS11599787)→ **RNA**→
env-94	JCVI_SCAF_1101667134076	+	589	656	18S_rRNA→ **RNA**→
env-95	JCVI_SCAF_1101667543876	-	270	203	←? (GOS19324) **RNA**→
env-96	JCVI_SCAF_1096628374971	+	1470	1537	←18S_rRNA **RNA**→
env-97	JCVI_SCAF_1101668226838	-	832	765	←? (GOS19324) **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-98	JCVI_SCAF_1096627207376	-	245	178	←? (GOS11953) **RNA**→
env-99	JCVI_SCAF_1096628297173	-	142	75	←18S_rRNA **RNA**→
env-100	JCVI_SCAF_1101668178152	+	203	270	**RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-101	JCVI_SCAF_1101668316227	+	308	375	18S_rRNA→ **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-102	JCVI_SCAF_1096627226399	-	387	320	←? (GOS19324) **RNA**→
env-103	JCVI_SCAF_1096627629356	-	1256	1189	**RNA**→ DUF1499 (pfam07386)? (GOS7536)→

Pma-1-1	NC_008816.1	-	950815	950748	←hypo **RNA**→ DUF1499 (pfam07386)→
env-104	JCVI_SCAF_1096627724050	+	1423	1490	←18S_rRNA **RNA**→
env-105	JCVI_SCAF_1096628046754	-	1116	1049	←? (GOS19324) **RNA**→ ←? (GOS11036104)
env-106	JCVI_SCAF_1096628112325	-	1399	1332	←? (GOS19324) **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-107	JCVI_SCAF_1101668193073	+	1424	1491	18S_rRNA→ **RNA**→
env-108	JCVI_SCAF_1096626961825	-	1162	1095	←18S_rRNA **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-109	JCVI_SCAF_1096627566666	+	1137	1204	←18S_rRNA **RNA**→
env-110	JCVI_SCAF_1096627779711	+	83	150	**RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-111	JCVI_SCAF_1101667078247	-	316	249	18S_rRNA→ **RNA**→
env-112	JCVI_SCAF_1101668187194	+	1267	1334	18S_rRNA→ **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-113	JCVI_SCAF_1101668318877	+	1250	1317	18S_rRNA→ **RNA**→ ? (GOS7536)→
env-114	JCVI_SCAF_1096626137142	-	346	279	? (GOS11599787)→ **RNA**→
env-115	JCVI_SCAF_1096628075878	+	1232	1299	←? (GOS11953) **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-116	JCVI_SCAF_1101667155913	-	190	123	←? (GOS19324) **RNA**→
env-117	JCVI_SCAF_1096627543325	+	1488	1554	18S_rRNA→ **RNA**→
env-118	JCVI_SCAF_1096627526801	-	1092	1025	←? (GOS19324) **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-119	JCVI_SCAF_1096627308669	+	557	624	18S_rRNA→ **RNA**→ DUF1499 (pfam07386)? (GOS7536)→
env-120	JCVI_SCAF_1096628291445	-	132	65	←? (GOS19324) **RNA**→
env-121	JCVI_SCAF_1096626973562	+	2739	2807	18S_rRNA→ **RNA**→
env-122	JCVI_SCAF_1096627800663	-	934	866	←? (GOS787) **RNA**→ ? (GOS10697)→
Pma-10-1	NC_007335.2	-	44898	44834	hypo→ **RNA**→ hypo→
Pma-9-2	NC_008819.1	-	53519	53454	hypo→ **RNA**→ hypo→
Pma-9-3	NC_008819.1	+	1037361	1037424	hypo→ **RNA**→ hypo→
Pma-10-2	NC_007335.2	+	386958	387022	high light inducible protein→ **RNA**→ hypo→
env-123	JCVI_SCAF_1096628087621	+	1053	1117	**RNA**→
env-124	JCVI_SCAF_1096628087621	+	450	514	**RNA**→
Pma-9-4	NC_008819.1	+	397455	397519	COG0398 (COG0398)→ **RNA**→ ←hypo
Pma-10-3	NC_007335.2	+	382482	382546	COG0398 (COG0398)→ **RNA**→ ←hypo
Pma-10-4	NC_007335.2	+	983401	983465	elongation factor Tu domain-containing protein→ **RNA**→ hypo→
env-125	JCVI_SCAF_1096628101613	-	1601	1537	**RNA**→ CobF (COG2243)? (GOS152)→
env-126	JCVI_SCAF_1096628040712	-	336	272	? (GOS13019008)→ **RNA**→
env-127	JCVI_SCAF_1101668539429	-	355	289	COG0398 (COG0398)? (GOS1258)→ **RNA**→
Ssp-3-1	NC_007516.1	-	1424714	1424654	hypo→ **RNA**→ ←hypo
Ssp-2-1	NC_008319.1	-	1487818	1487754	hypo→ **RNA**→ ←hypo
Ssp-2-2	NC_008319.1	+	1138022	1138086	SNase (pfam00565)→ **RNA**→ hypo→
Ssp-2-3	NC_008319.1	+	997952	998016	hypo→ **RNA**→ SNase (pfam00565)→
env-128	JCVI_SCAF_1096626611670	+	1400	1464	**RNA**→
Ssp-2-4	NC_008319.1	-	772842	772778	←hypo **RNA**→ ←hypo
Pma-5-1	NC_008820.1	+	356821	356885	hypo→ **RNA**→ hypo→
Pma-7-1	NC_005071.1	-	1667396	1667332	carbonic anhydrase→ **RNA**→ ←COG3152 (COG3152)
Ssp-2-5	NC_008319.1	+	1026369	1026433	←hypo **RNA**→ ←tRNA-Pro
Ssp-2-6	NC_008319.1	+	785499	785563	COG0412 (COG0412)→ **RNA**→ ←COG3386 (COG3386)
Ssp-2-7	NC_008319.1	+	979499	979563	hypo→ **RNA**→ putative lipoprotein→
Ssp-2-8	NC_008319.1	-	1490305	1490241	hypo→ **RNA**→ ←hypo
Ssp-10-1	NZ_AANO01000001.1	-	236655	236586	hypo→ **RNA**→ ←RplL (COG0222)
Ssp-2-9	NC_008319.1	+	2394765	2394834	hypo→ **RNA**→ ←RplL (COG0222)
Pma-7-2	NC_005071.1	-	2205028	2204960	hypo→ **RNA**→ ←RplL (COG0222)
Pma-5-2	NC_008820.1	-	2447837	2447769	hypo→ **RNA**→ ←hypo
env-129	JCVI_SCAF_1101668324115	+	456	524	? (GOS8033)→ **RNA**→ ? (GOS8033)→
Ssp-11-1	NC_009481.1	+	2191549	2191619	hypo→ **RNA**→ ←RplL (COG0222)
env-130	JCVI_SCAF_1096627998360	+	1284	1354	←RnhA (COG0328)? (GOS3931) **RNA**→ ? (GOS10654)→
Ssp-12-1	NZ_AAOK01000007.1	-	27940	27870	hypo→ **RNA**→ ←RplL (COG0222)

env-131	JCVI_SCAF_1101668230400	+	37	106	**RNA**→ ? (GOS10654)→
env-132	JCVI_SCAF_1096628046072	+	1220	1286	? (GOS10654)→ **RNA**→ ? (GOS8033)→
Ssp-7-1	NC_009482.1	+	2048577	2048642	←RnhA (COG0328) **RNA**→ ←RplL (COG0222)
Ssp-4-1	NC_007513.1	+	2054658	2054726	hypo→ **RNA**→ ←RplL (COG0222)
env-133	JCVI_SCAF_1096628043949	-	364	296	? (GOS13202071)→ **RNA**→ ? (GOS10654)→
env-134	JCVI_SCAF_1101667377116	-	115	47	? (GOS13096573)→ **RNA**→
env-135	JCVI_SCAF_1096627216867	+	473	541	? (GOS13202071)→ **RNA**→ ? (GOS10654)→
env-136	JCVI_SCAF_1096627197861	-	470	402	←? (GOS3931) **RNA**→ ? (GOS10654)→
Ssp-13-2	NC_005070.1	+	2242696	2242764	hypo→ **RNA**→ ←RplL (COG0222)
env-137	JCVI_SCAF_1101667184502	+	226	294	**RNA**→ ? (GOS8033)→
Ssp-3-2	NC_007516.1	+	2274186	2274254	hypo→ **RNA**→ ←Ribosomal_L12 (pfam00542)
env-138	JCVI_SCAF_1096627153447	+	3625	3693	? (GOS13096573)→ **RNA**→
env-139	JCVI_SCAF_1096628072039	+	978	1046	? (GOS2932)→ **RNA**→ ←RplL (COG0222)? (GOS996)
Ssp-10-2	NZ_AANO01000044.1	-	1123	1054	hypo→ **RNA**→ ←RplL (COG0222)
Ssp-1-1	NZ_AATZ01000002.1	-	532918	532850	hypo→ **RNA**→ ←RplL (COG0222)
env-140	JCVI_SCAF_1101668676589	-	444	376	? (GOS13096573)→ **RNA**→ ? (GOS8033)→
env-141	JCVI_SCAF_1101667122093	+	605	673	? (GOS13202071)→ **RNA**→
env-142	JCVI_SCAF_1096627383380	+	89	157	? (GOS8033)→ **RNA**→ ? (GOS8033)→
env-143	JCVI_SCAF_1096627969158	+	114	182	**RNA**→ ←RplJ (COG0244)? (GOS407)
env-144	JCVI_SCAF_1101667335056	+	528	596	? (GOS13202071)→ **RNA**→
env-145	JCVI_SCAF_1096627214382	+	918	986	←? (GOS3931) **RNA**→ ? (GOS10654)→
env-146	JCVI_SCAF_1096627824847	-	469	402	? (GOS13096573)→ **RNA**→ ? (GOS10654)→
Ssp-9-1	NZ_AANP01000001.1	-	504528	504459	hypo→ **RNA**→ ←RplL (COG0222)
env-147	JCVI_SCAF_1096627937755	+	23	92	? (GOS10654)→ **RNA**→ ? (GOS10654)→
env-148	JCVI_SCAF_1101668618928	-	1142	1073	? (GOS10654)→ **RNA**→ ? (GOS10654)→
env-149	JCVI_SCAF_1096627362115	-	1614	1545	? (GOS8033)→ **RNA**→ ? (GOS10654)→
env-150	JCVI_SCAF_1096627084801	+	994	1064	? (GOS13096573)→ **RNA**→ ? (GOS10654)→
env-151	JCVI_SCAF_1096627194150	-	880	813	←RnhA (COG0328)? (GOS3931) **RNA**→ ←RplL (COG0222)? (GOS996)
Ssp-8-2	NZ_AAUA01000003.1	-	154002	153938	←hypo **RNA**→ ←hypo
env-152	JCVI_SCAF_1096627074932	+	890	953	? (GOS11211626)→ **RNA**→ ←? (GOS15927)
env-153	JCVI_SCAF_1096627970216	+	3	65	**RNA**→ ? (GOS14642395)→
env-154	JCVI_SCAF_1096626970485	-	1112	1050	**RNA**→
env-155	JCVI_SCAF_1096628060967	+	1600	1660	ZnMc (cd00203)? (GOS4228)→ **RNA**→
env-156	JCVI_SCAF_1096627192266	+	5787	5848	? (GOS11820130)→ **RNA**→ ? (GOS11820130)→
env-157	JCVI_SCAF_1096628080391	-	572	511	←Ser tRNA **RNA**→ ? (GOS11820130)→
env-158	JCVI_SCAF_1096627183482	-	1345	1283	Soj (COG1192)? (GOS1716)→ **RNA**→
Ssp-13-3	NC_005070.1	+	1131767	1131827	hypo→ **RNA**→ ←hypo
env-159	JCVI_SCAF_1096627190267	-	899	838	? (GOS15836)→ **RNA**→
Ssp-3-3	NC_007516.1	+	534384	534444	hypo→ **RNA**→ GST_C_Zeta (cd03191)GST_N_Zeta (cd03042)→
env-160	JCVI_SCAF_1096627082690	-	719	656	? (GOS13547206)→ **RNA**→ ? (GOS11359600)→
env-161	JCVI_SCAF_1096627963120	-	495	432	? (GOS13547206)→ **RNA**→ ? (GOS11359600)→
env-162	JCVI_SCAF_1096627838359	+	250	313	**RNA**→
Ssp-4-2	NC_007513.1	-	1540528	1540466	←hypo **RNA**→ ←carbamoyl-phosphate synthase L chain
env-163	JCVI_SCAF_1096627282915	+	1522	1585	Pro tRNA→ **RNA**→
env-164	JCVI_SCAF_1096627828942	+	61	124	**RNA**→ AslB (COG0641)? (GOS255)→
env-165	JCVI_SCAF_1096628082578	+	1267	1332	←? (GOS11713070) **RNA**→
env-166	JCVI_SCAF_1096627303375	-	1736	1672	←? (GOS11177499) **RNA**→ ? (GOS13261968)→
env-167	JCVI_SCAF_1101668294323	+	1286	1350	? (GOS11249006)→ **RNA**→ ? (GOS11249006)→
env-168	JCVI_SCAF_1096628074362	+	174	238	? (GOS11249006)→ **RNA**→ ? (GOS11249006)→
env-169	JCVI_SCAF_1096627371490	+	2060	2124	? (GOS11249006)→ **RNA**→ ? (GOS11249006)→
env-170	JCVI_SCAF_1096628059788	+	1566	1630	←? (GOS11760973) **RNA**→
env-171	JCVI_SCAF_1096627516204	-	194	130	? (GOS11249006)→ **RNA**→ ? (GOS11249006)→

env-172	JCVI_SCAF_1101668668167	+	1286	1350	? (GOS11249006)→ **RNA**→ ? (GOS11249006)→
env-173	JCVI_SCAF_1096627369637	+	1187	1251	? (GOS14642395)→ **RNA**→ ? (GOS14642395)→
env-174	JCVI_SCAF_1096627961667	+	502	566	? (GOS14642395)→ **RNA**→ ? (GOS14642395)→
Ssp-3-4	NC_007516.1	+	1389620	1389683	hypo→ **RNA**→ hypo→
Ssp-4-3	NC_007513.1	+	1422400	1422465	←tRNA-Cys **RNA**→ membrane protein-like→
Ssp-1-2	NZ_AATZ01000003.1	-	191809	191743	hypo→ **RNA**→ ←hypo
env-175	JCVI_SCAF_1101668028160	-	650	584	**RNA**→
env-176	JCVI_SCAF_1096627794613	+	1686	1753	? (GOS11634178)→ **RNA**→
Ssp-3-5	NC_007516.1	-	524530	524469	hypo→ **RNA**→ ←hypo
env-177	JCVI_SCAF_1101668743973	-	770	706	? (GOS12184068)→ **RNA**→
env-178	JCVI_SCAF_1096627116428	-	1163	1099	? (GOS11249006)→ **RNA**→ ? (GOS11249006)→
Ssp-13-4	NC_005070.1	+	1127966	1128025	hypo→ **RNA**→ hypo→
env-179	JCVI_SCAF_1096627066253	-	975	909	? (GOS11335739)→ **RNA**→ ←? (GOS11395197)
Csp-1-1	NZ_ABSE01000014.1	+	50433	50504	**RNA**→
Tel-1-2	NC_004113.1	+	1092076	1092149	←tRNA-Pro **RNA**→ COG0742 (COG0742)→
Ssp-14-3	NC_000911.1	+	2730521	2730588	←COG3330 (COG3330) **RNA**→ YafJ (cd01908)→
Nsp-1-2	NC_003272.1	-	3740057	3739982	←REC (cd00156)BaeS (COG0642)AtoS (COG2202)PAC (pfam00785)PAS (smart00091) **RNA**→ **RNA**→ **RNA**→ ←hypo
Nsp-1-3	NC_003272.1	-	3740217	3740143	(shown above)
Nsp-1-4	NC_003272.1	-	3740423	3740349	(shown above)
Ava-1-2	NC_007413.1	+	4756984	4757059	←PAS (cd00130)REC (cd00156)BaeS (COG0642)AtoS (COG2202)PAS (smart00091) **RNA**→ **RNA**→ ←hypo
Ava-1-3	NC_007413.1	+	4756822	4756897	(shown above)
Npu-1-1	NC_010628.1	-	7794370	7794295	AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→ **RNA**→ COG1896 (COG1896)→
Nsp-1-5	NC_003272.1	-	3744859	3744786	AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→ **RNA**→ ←REC (cd00156)BaeS (COG0642)AtoS (COG2202)PAC (pfam(
Mae-1-1	NC_010296.1	-	3189581	3189508	hypo→ **RNA**→ PlsC (COG0204)→
Csp-2-1	NC_010546.1	-	4333822	4333760	←HCP (cd01914) **RNA**→ ←hypo
Cwa-1-1	NZ_AADV02000005.1	+	52634	52696	←DYNc (smart00053) **RNA**→ hypo→
Csp-2-2	NC_010546.1	+	3842870	3842932	Mfd (COG1197)→ **RNA**→ hypo→
Csp-3-1	NZ_AAXW01000004.1	+	185218	185280	←COG3335 (COG3335) **RNA**→ SgcQ (COG0434)→
Lsp-1-1	NZ_AAVU01000009.1	-	97263	97206	SecA (COG0653)→ **RNA**→ ←hypo
Ssp-6-1	NC_007775.1	+	383657	383722	hypo→ **RNA**→ ←Glyoxalase (pfam00903)
Ssp-5-1	NC_007776.1	-	729361	729297	hypo→ **RNA**→ BaeS (COG0642)→
Mae-1-2	NC_010296.1	-	3427434	3427352	←Transposase_2 (pfam01385)Transposase_35 (pfam07282) **RNA**→ GvpP (COG4980)→
Csp-5-1	NZ_ABVE01000004.1	+	406621	406694	CysE (COG1045)→ **RNA**→ COG2334 (COG2334)→
Csp-4-1	NZ_ABOY01000062.1	-	6593	6521	HTH_XRE (cd00093)→ **RNA**→ COG2334 (COG2334)→
Csp-5-2	NZ_ABVE01000004.1	-	807677	807604	PP2Ac (cd00144)→ **RNA**→ DUF159 (pfam02586)→
Cwa-1-2	NZ_AADV02000078.1	-	6931	6871	←hypo **RNA**→ ←Chorismate_synt (pfam01264)
Csp-7-1	NZ_ABVI01000008.1	+	27442	27506	←protein of unknown function DUF1824 **RNA**→ GlyRS_alpha_core (cd00733)→
Csp-6-1	NZ_ABLR01000007.1	+	51654	51718	←Domain of unknown function DUF1824 **RNA**→ GlyRS_alpha_core (cd00733)→
Csp-3-2	NZ_AAXW01000069.1	-	22055	21991	IbpA (COG0071)→ **RNA**→ ←hypo
Ssp-12-2	NZ_AAOK01000003.1	-	24359	24290	←hypo **RNA**→ ←COG4338 (COG4338)
Ssp-11-2	NC_009481.1	-	901908	901843	←hypo **RNA**→ ←COG4338 (COG4338)
Ssp-2-10	NC_008319.1	+	972621	972684	←hypo **RNA**→ ←hypo

25.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 25.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Cyano-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

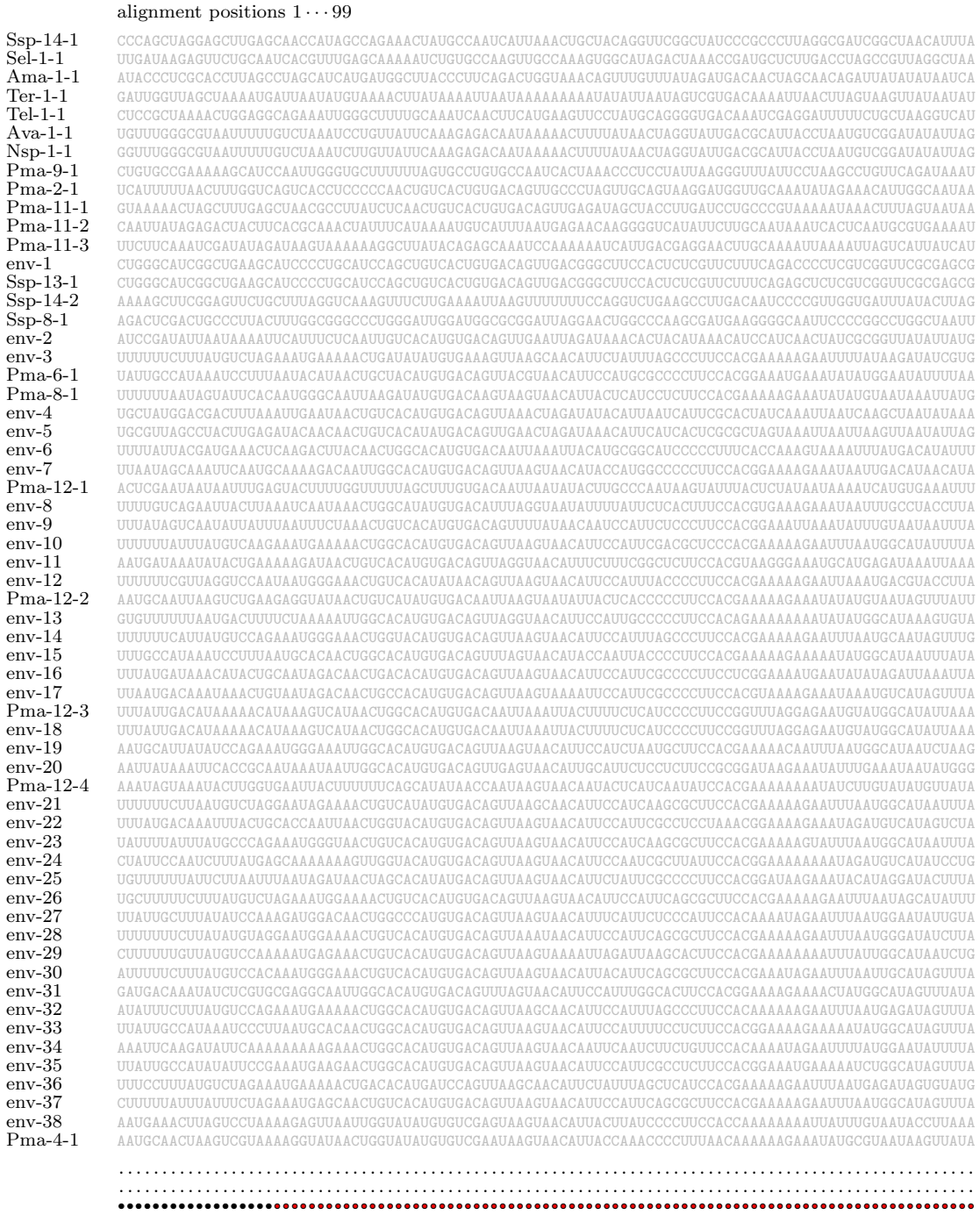
u cd00093 (1) Helix-turn-helix XRE-family like proteins.
u cd00130 (1) PAS domain; PAS motifs appear in archaea, eubacteria and eukarya.
u cd00144 (1) Protein phosphatase 2A homologues, catalytic domain.
u d cd00156 (3) Signal receiver domain; originally thought to be unique to bacteria (CheY, OmpR, NtrC, and PhoB), now recently identified in eukaroytes ETR1 Arabidopsis thaliana; this domain receives the signal from the sensor partner in a two-component systems; contains a phosphoacceptor site that is phosphorylated by histidine kinase homologs; usually found N-terminal to a DNA binding effector domain; forms homodimers
u cd00203 (1) Zinc-dependent metalloprotease; Neutral zinc metallopeptidases.
d cd00207 (4) 2Fe-2S iron-sulfur cluster binding domain.
D cd00733 (2) Class II Glycyl-tRNA synthetase (GlyRS) alpha subunit core catalytic domain.
D cd01908 (1) Glutamine amidotransferases class-II (Gn-AT)_YafJ-type.
u cd01914 (1) Hybrid cluster protein (HCP), formerly known as prismane, is thought to play a role in nitrogen metabolism but its specific function is unknown. [Fe4-S4] cubane cluster similar to that of carbon monoxide dehydrogenase (CODH). The second cluster, referred to as the hybrid cluster, is a hybrid [Fe4-S2-O2]
u cd02950 (1) TRX-like protein A (TxlA) family; TxlA was originally isolated from the cyanobacterium Synechococcus.
D cd03042 (1) GST_N family, Class Zeta subfamily; GSTs are cytosolic dimeric proteins involved in cellular detoxification by catalyzing the conjugation of glutathione (GSH) with a wide range of endogenous and xenobiotic alkylating agents, including carcinogens, therapeutic drugs, environmental toxins and products of oxidative stress.
D cd03191 (1) GST_C family, Class Zeta subfamily; GSTs are cytosolic dimeric proteins involved in cellular detoxification by catalyzing the conjugation of glutathione (GSH) with a wide range of endogenous and xenobiotic alkylating agents, including carcinogens, therapeutic drugs, environmental toxins, and products of oxidative stress.
u cd03235 (1) ABC-type Mn/Zn transport systems, ATPase component; This family is comprised of systems involved in the uptake of various metallic cations such as iron, manganese and zinc.
u COG0071 (1) Molecular chaperone (small heat shock protein) [Posttranslational modification, protein turnover, chaperones]
D COG0204 (1) 1-acyl-sn-glycerol-3-phosphate acyltransferase [Lipid metabolism]
d COG0222 (13) Ribosomal protein L7/L12 [Translation, ribosomal structure and biogenesis]
d COG0244 (1) Ribosomal protein L10 [Translation, ribosomal structure and biogenesis]
d COG0285 (1) Folylpolyglutamate synthase [Coenzyme metabolism]
u COG0328 (3) Ribonuclease HI [DNA replication, recombination, and repair]
uD COG0398 (5) Uncharacterized conserved protein [Function unknown]
u COG0412 (1) Dienelactone hydrolase and related enzymes [Secondary metabolites biosynthesis, transport, and catabolism]
D COG0426 (1) Uncharacterized flavoproteins [Energy production and conversion]
D COG0434 (1) Predicted TIM-barrel enzyme [General function prediction only]
D COG0508 (1) Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component, and related enzymes [Energy production and conversion]
d COG0561 (2) Predicted hydrolases of the HAD superfamily [General function prediction only]
u COG0596 (1) Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily) [General function prediction only]
D COG0606 (1) Predicted ATPase with chaperone activity [Posttranslational modification, protein turnover, chaperones]
D COG0641 (1) Arylsulfatase regulator (Fe-S oxidoreductase) [General function prediction only]
uDd COG0642 (4) Signal transduction histidine kinase [Signal transduction mechanisms]
u COG0653 (1) Preprotein translocase subunit SecA (ATPase, RNA helicase) [Intracellular trafficking and secretion]
D COG0742 (1) N6-adenine-specific methylase [DNA replication, recombination, and repair]
u COG1045 (1) Serine acetyltransferase [Amino acid transport and metabolism]
u COG1192 (1) ATPases involved in chromosome partitioning [Cell division and chromosome partitioning]
u COG1197 (1) Transcription-repair coupling factor (superfamily II helicase) [DNA replication, recombination, and repair / Transcription]
d COG1492 (2) Cobyric acid synthase [Coenzyme metabolism]
D COG1896 (1) Predicted hydrolases of HD superfamily [General function prediction only]
u d COG2202 (3) FOG: PAS/PAC domain [Signal transduction mechanisms]
D COG2243 (1) Precorrin-2 methylase [Coenzyme metabolism]
D COG2334 (2) Putative homoserine kinase type II (protein kinase fold) [General function prediction only]
u COG3128 (1) Uncharacterized iron-regulated protein [Function unknown]
d COG3152 (1) Predicted membrane protein [Function unknown]
u COG3330 (1) Uncharacterized protein conserved in bacteria [Function unknown]
u COG3335 (1) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
d COG3386 (1) Gluconolactonase [Carbohydrate transport and metabolism]
d COG4338 (2) Uncharacterized protein conserved in bacteria [Function unknown]
D COG4446 (3) Uncharacterized protein conserved in bacteria [Function unknown]
D COG4449 (1) Predicted protease of the Abi (CAAX) family [General function prediction only]
D COG4980 (1) Gas vesicle protein [General function prediction only]
uD GOS10654 (14) ?
D GOS10697 (6) ?
d GOS10947 (1) ?
d GOS11036104 (1) ?
u GOS11053562 (1) ?
u GOS11177499 (1) ?
u GOS11211626 (1) ?
uD GOS11249006 (12) ?
D GOS11326834 (1) ?
u GOS11335739 (1) ?
D GOS11359600 (2) ?
u GOS11385352 (2) ?
u GOS11385446 (1) ?
d GOS11395197 (1) ?
u GOS11457573 (1) ?
u GOS11599787 (3) ?
u GOS11634178 (1) ?
u GOS11661723 (5) ?
u GOS11713070 (1) ?
u GOS11760973 (1) ?
uD GOS11820130 (3) ?
u GOS11907928 (2) ?
u GOS11953 (4) ?
u GOS11973280 (1) ?

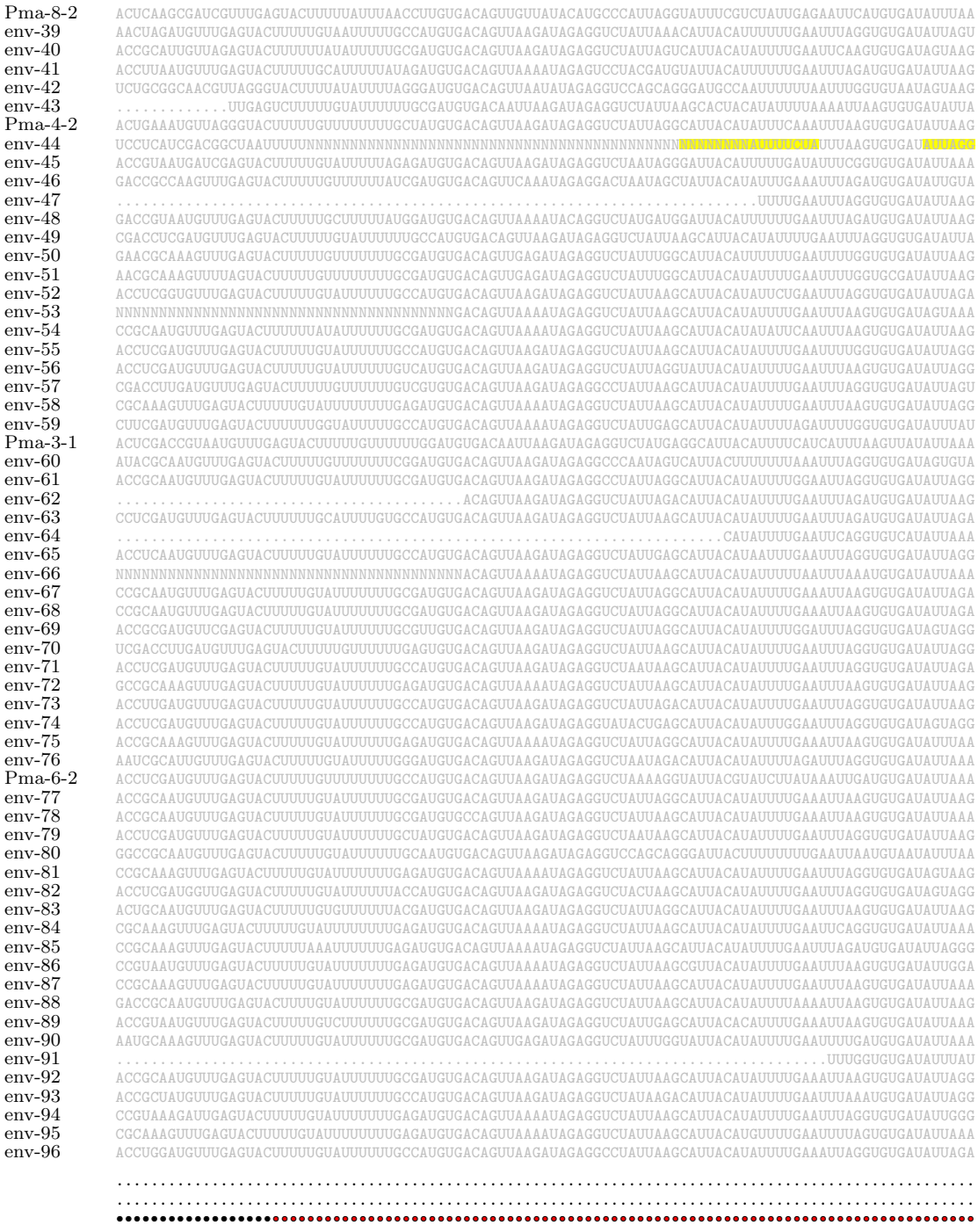
u GOS12183470 (1) ?
u GOS12184068 (1) ?
u GOS1258 (1) ?
d GOS12835 (2) ?
u GOS13019008 (1) ?
d GOS13019159 (1) ?
D GOS13095788 (1) ?
u GOS13096573 (5) ?
u GOS13202071 (4) ?
D GOS13261968 (1) ?
u GOS1330 (1) ?
d GOS13534 (1) ?
u GOS13547206 (2) ?
uD GOS14642395 (5) ?
D GOS15049 (5) ?
D GOS152 (1) ?
u GOS15836 (1) ?
u GOS159 (1) ?
d GOS15927 (1) ?
d GOS1694 (1) ?
u GOS1716 (1) ?
u GOS19324 (16) ?
D GOS255 (1) ?
D GOS2778 (1) ?
u GOS2932 (1) ?
u GOS303 (7) ?
d GOS3592 (2) ?
u GOS3931 (4) ?
D GOS395 (1) ?
d GOS407 (1) ?
u GOS4228 (1) ?
D GOS49 (1) ?
u GOS7134 (1) ?
D GOS7536 (49) ?
u GOS787 (1) ?
uD GOS8033 (8) ?
d GOS9508 (3) ?
d GOS996 (2) ?
u NOG25202 (2)
D pfam00106 (1) short chain dehydrogenase.
d pfam00542 (1) Ribosomal protein L7/L12 C-terminal domain.
uD pfam00565 (2) Staphylococcal nuclease homologue. [SNase numbering in parentheses]
u d pfam00785 (2) PAC motif.
d pfam00903 (1) Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily.
D pfam01165 (1) Ribosomal protein S21.
d pfam01264 (1) Chorismate synthase.
u pfam01385 (1) Probable transposase.
D pfam01613 (1) Flavin reductase like domain.
u pfam01694 (3) Rhomboid family.
u pfam01808 (2) AICARFT/IMPCHase bienzyme.
u pfam02142 (2) MGS-like domain.
d pfam02566 (1) OsmC-like protein.
D pfam02586 (1) Uncharacterised ACR, COG2135.
u pfam05746 (2) DALR anticodon binding domain.
u pfam07282 (1) Putative transposase DNA-binding domain.
D pfam07386 (42) Protein of unknown function (DUF1499).
u smart00053 (1) Dynamin, GTPase; Large GTPases that mediate vesicle trafficking.
u d smart00091 (3) PAS domain; PAS motifs appear in archaea, eubacteria and eukarya.
u smart00347 (1) helix_turn_helix multiple antibiotic resistance protein;
D smart00466 (1) SET and RING finger associated domain.

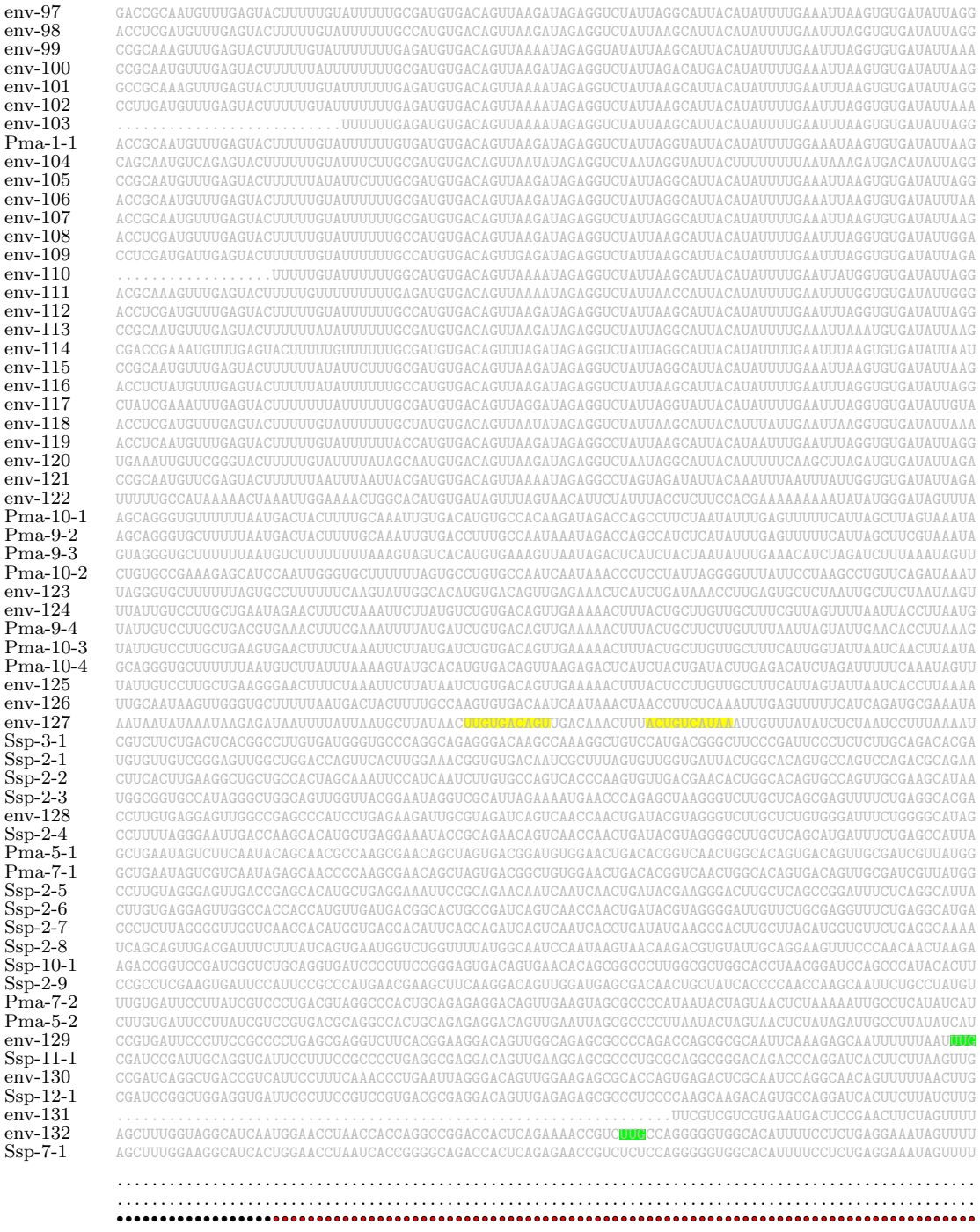
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Sel-2-1

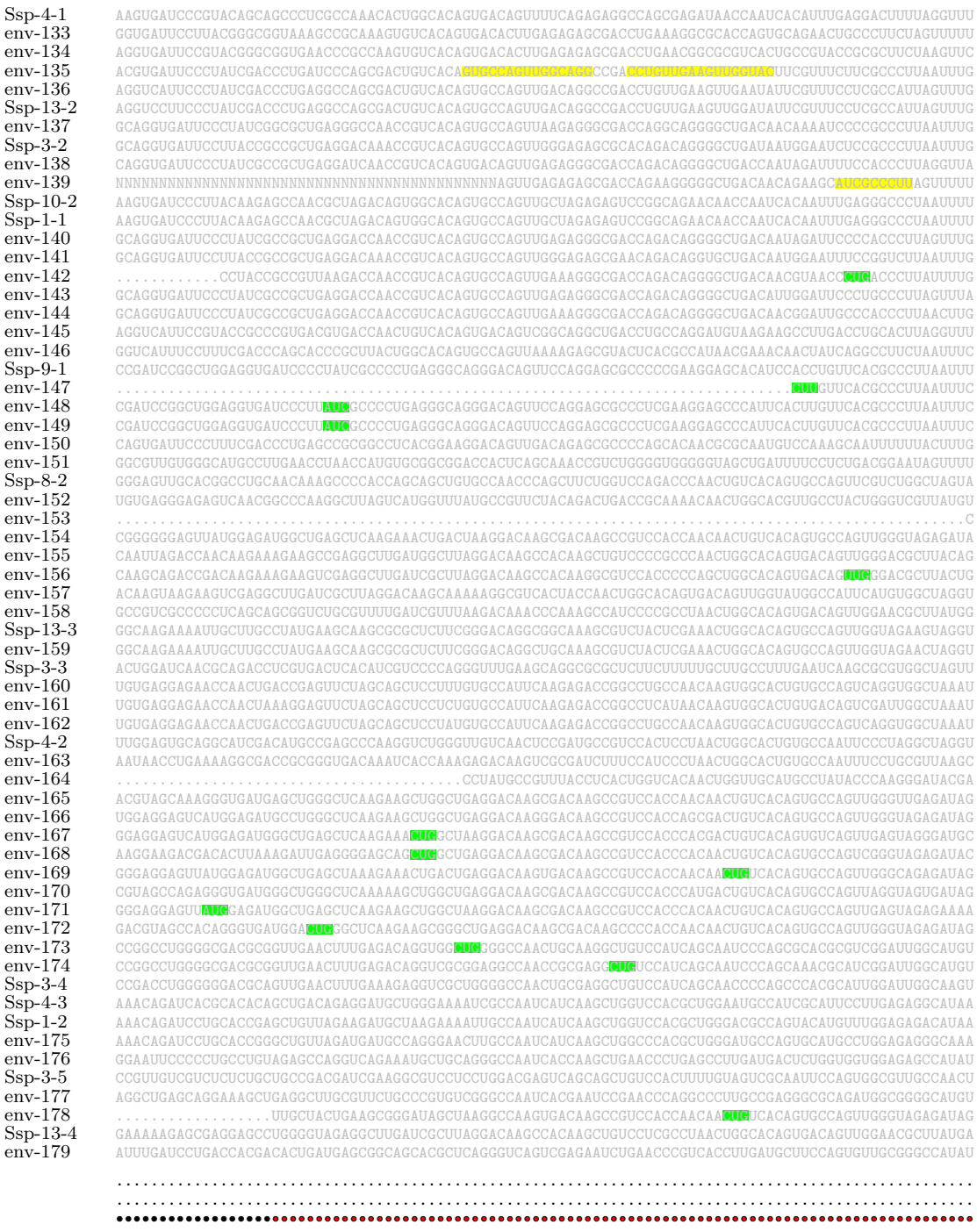
25.4 Multiple-sequence alignment

Each Cyano-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 25.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Cyano-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

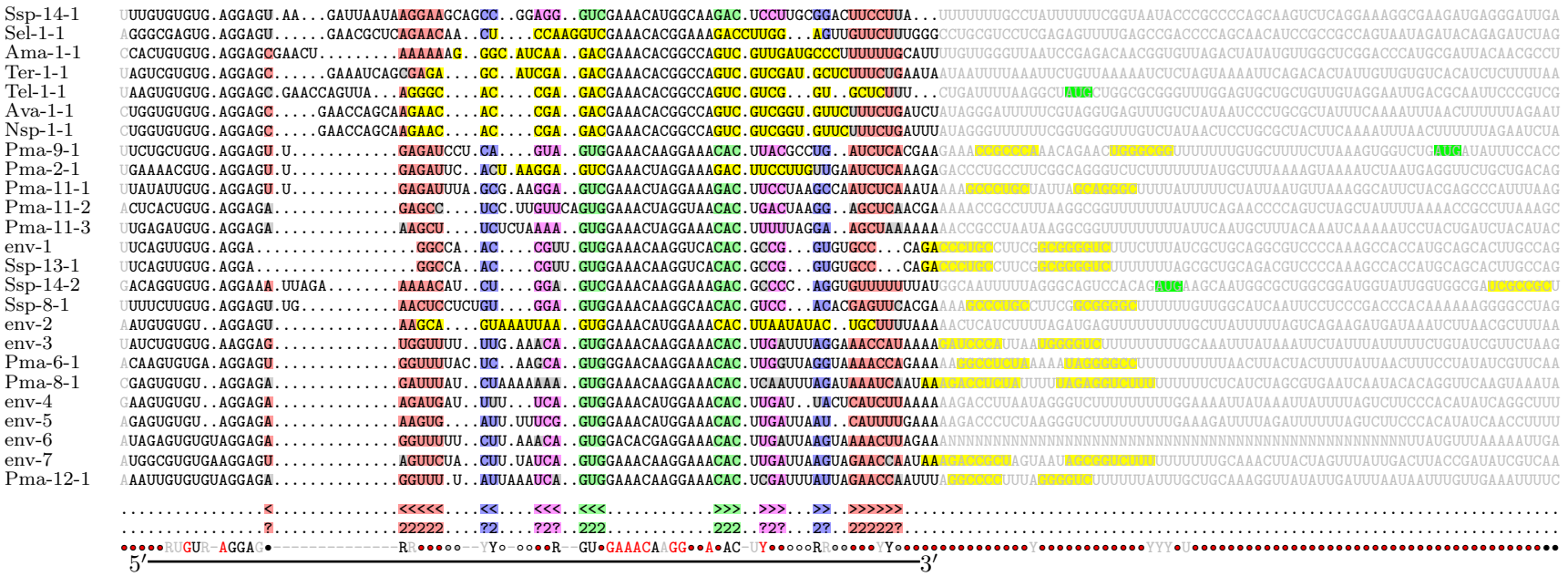

alignment positions 1 · · · 99
Ssp-14-1
Sel-1-1
Ama-1-1
Ter-1-1
Tel-1-1
Ava-1-1
Nsp-1-1
Pma-9-1
Pma-2-1
Pma-11-1
Pma-11-2
Pma-11-3
env-1
Ssp-13-1
Ssp-14-2
Ssp-8-1
env-2
env-3
Pma-6-1
Pma-8-1
env-4
env-5
env-6
env-7
Pma-12-1
env-8
env-9
env-10
env-11
env-12
Pma-12-2
env-13
env-14
env-15
env-16
env-17
Pma-12-3
env-18
env-19
env-20
Pma-12-4
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
Pma-4-1

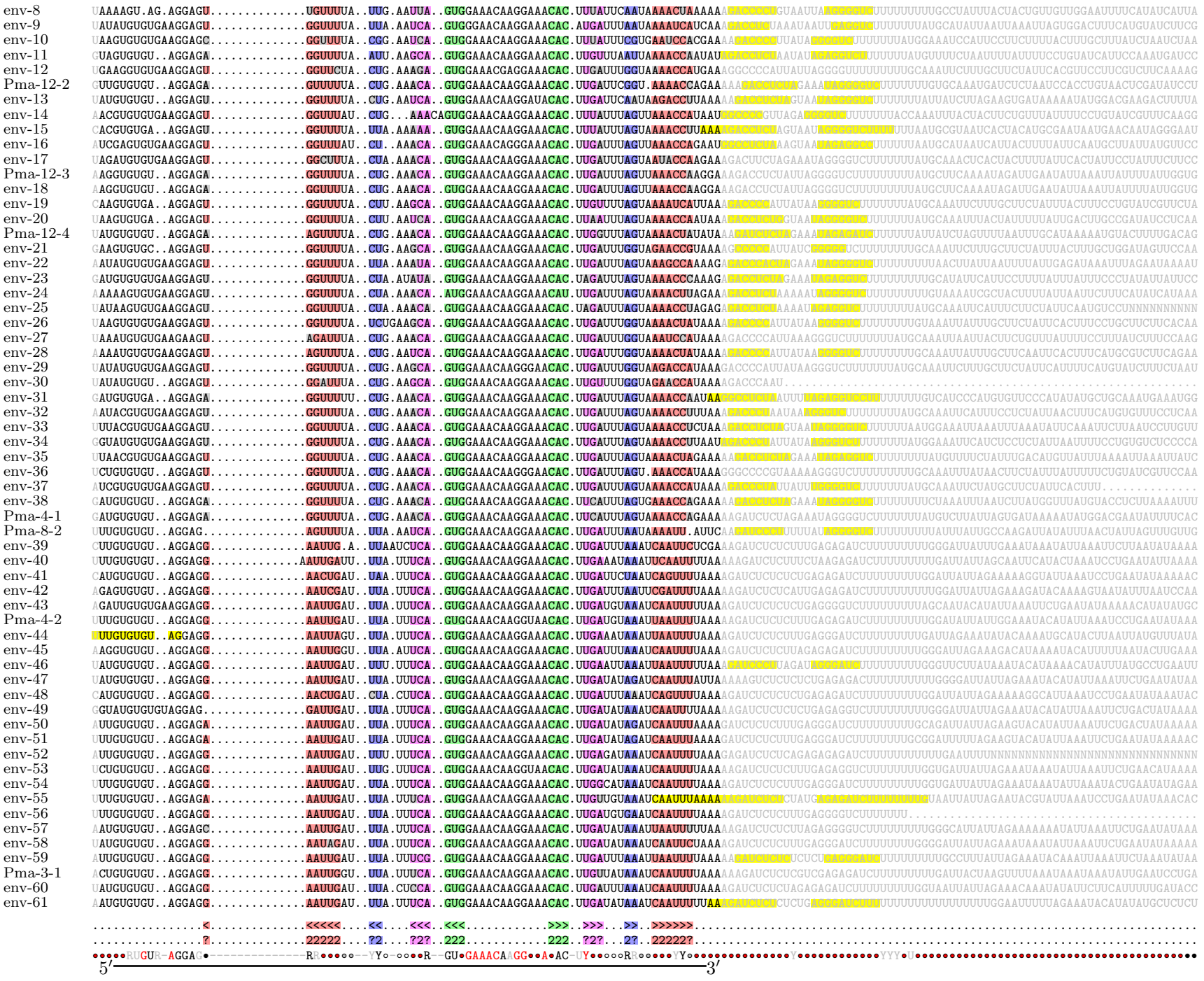

Pma-8-2
env-39
env-40
env-41
env-42
env-43
Pma-4-2
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
Pma-3-1
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
Pma-6-2
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96

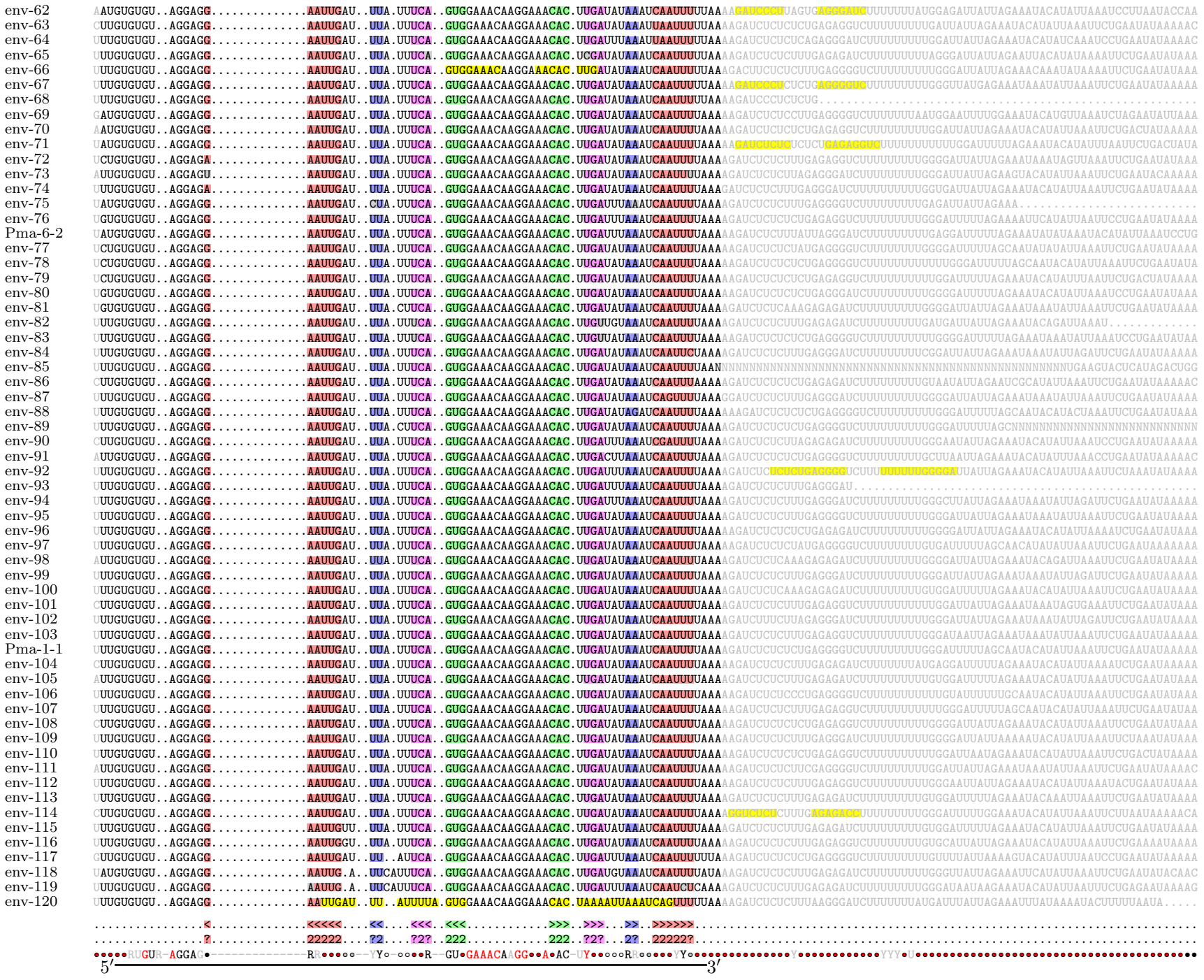

env-97
env-98
env-99
env-100
env-101
env-102
env-103
Pma-1-1
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
Pma-10-1
Pma-9-2
Pma-9-3
Pma-10-2
env-123
env-124
Pma-9-4
Pma-10-3
Pma-10-4
env-125
env-126
env-127
Ssp-3-1
Ssp-2-1
Ssp-2-2
Ssp-2-3
env-128
Ssp-2-4
Pma-5-1
Pma-7-1
Ssp-2-5
Ssp-2-6
Ssp-2-7
Ssp-2-8
Ssp-10-1
Ssp-2-9
Pma-7-2
Pma-5-2
env-129
Ssp-11-1
env-130
Ssp-12-1
env-131
env-132
Ssp-7-1

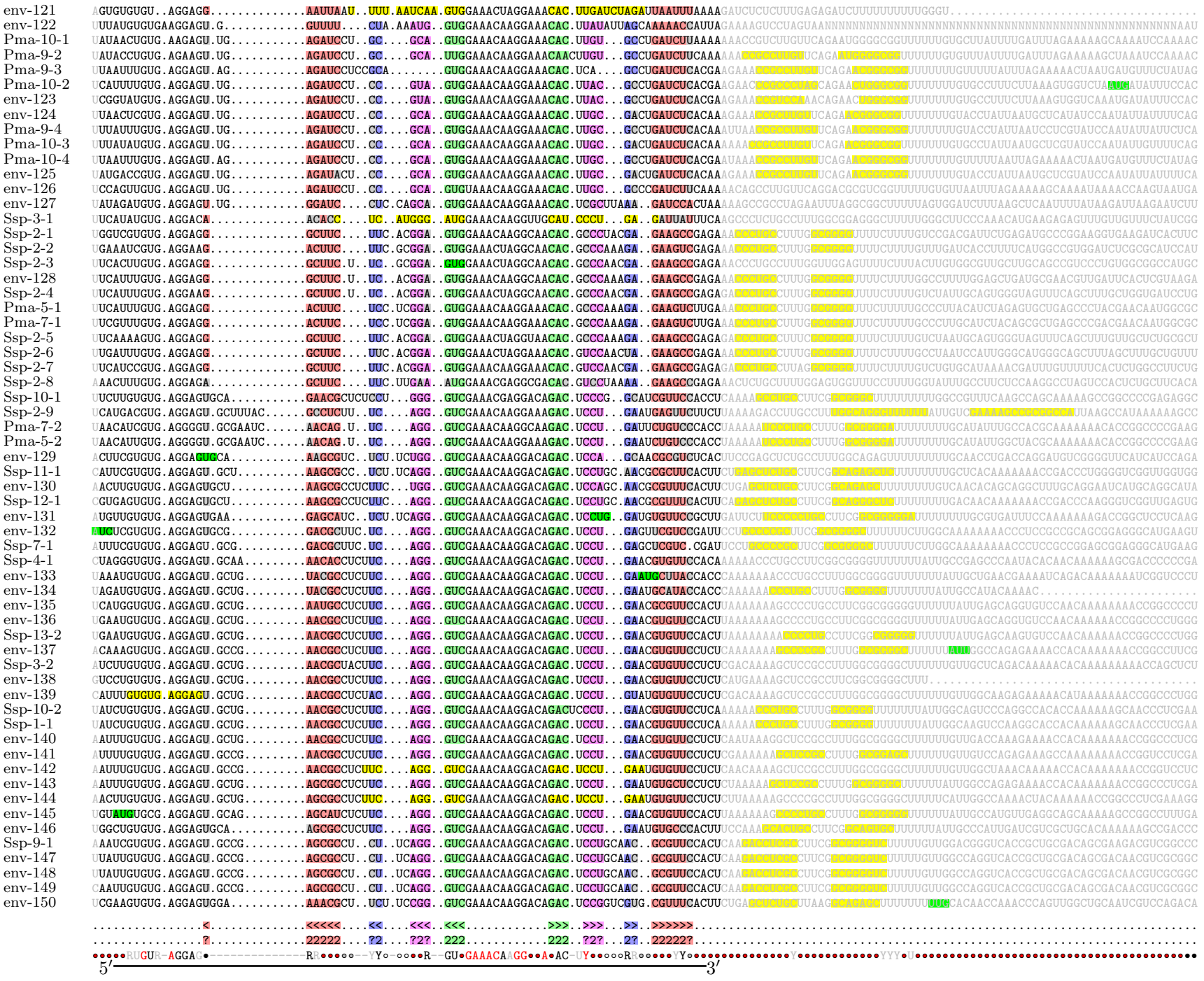

Ssp-4-1
env-133
env-134
env-135
env-136
Ssp-13-2
env-137
Ssp-3-2
env-138
env-139
Ssp-10-2
Ssp-1-1
env-140
env-141
env-142
env-143
env-144
env-145
env-146
Ssp-9-1
env-147
env-148
env-149
env-150
env-151
Ssp-8-2
env-152
env-153
env-154
env-155
env-156
env-157
env-158
Ssp-13-3
env-159
Ssp-3-3
env-160
env-161
env-162
Ssp-4-2
env-163
env-164
env-165
env-166
env-167
env-168
env-169
env-170
env-171
env-172
env-173
env-174
Ssp-3-4
Ssp-4-3
Ssp-1-2
env-175
env-176
Ssp-3-5
env-177
env-178
Ssp-13-4
env-179

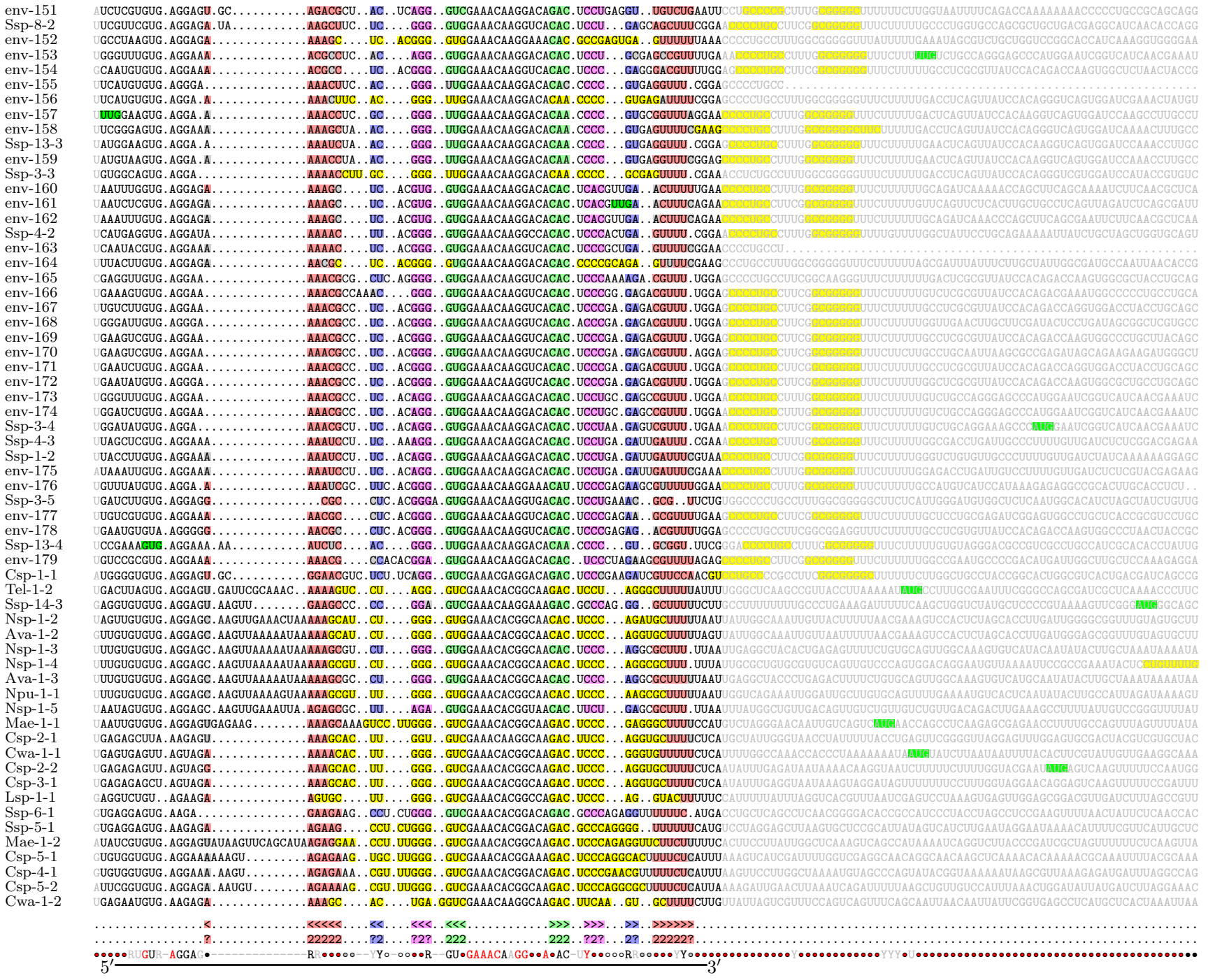
alignment positions $100 \cdots 259$

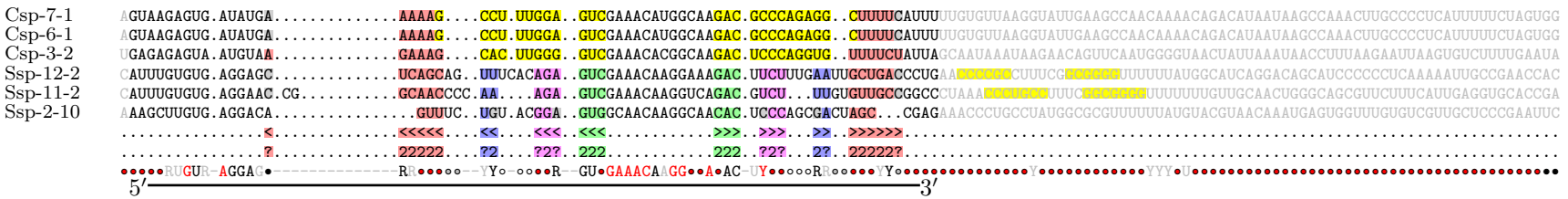



env-8
Pma-12-2
env-13
env-14
env-15
env-16
env-17
Pma-12-3
env-18
env-19
env-20
Pma-12-4
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
Pma-4-1
Pma-8-2
env-39
env-40
env-41
env-42
env-43
Pma-4-2
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
Pma-3-1
env-60
env-61
5′
3′

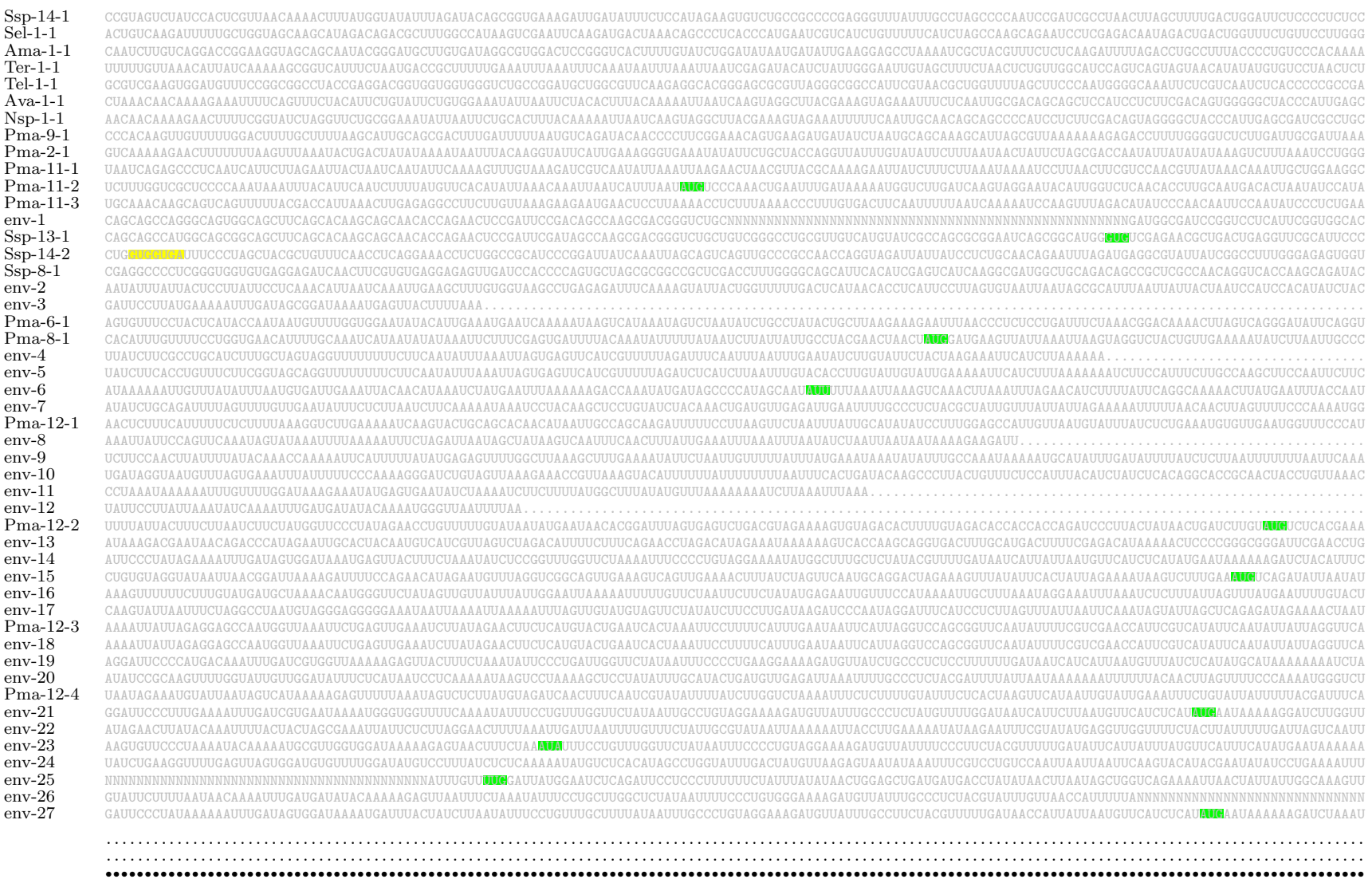

env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
Pma-6-2
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
Pma-1-1
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
5′
3′

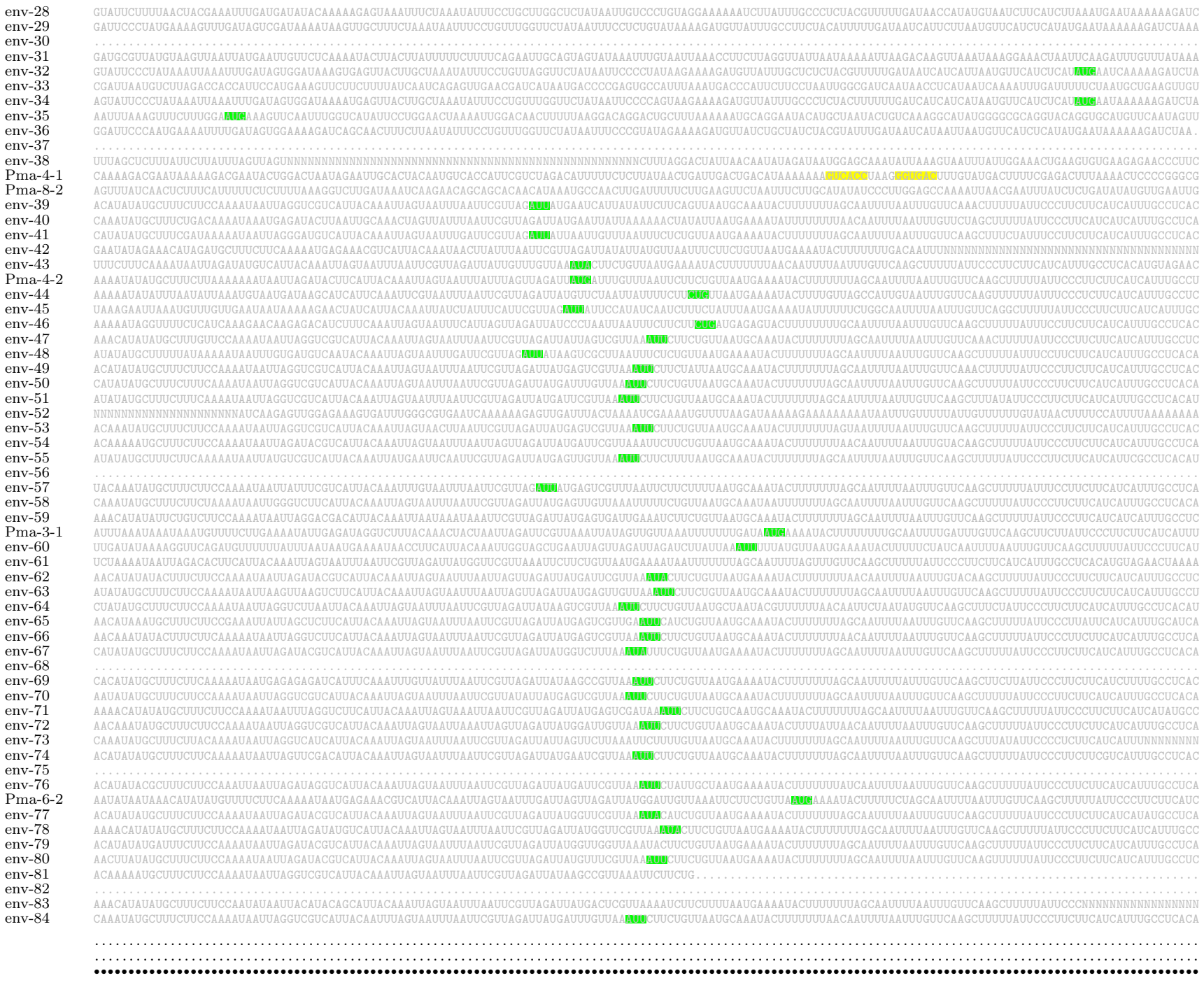

env-121
env-122
Pma-10-1
Pma-9-2
Pma-9-3
Pma-10-2
env-123
env-124
Pma-9-4
Pma-10-3
Pma-10-4
env-125
env-126
env-127
Ssp-3-1
Ssp-2-1
Ssp-2-2
Ssp-2-3
env-128
Ssp-2-4
Pma-5-1
Pma-7-1
Ssp-2-5
Ssp-2-6
Ssp-2-7
Ssp-2-8
Ssp-10-1
Ssp-2-9
Pma-7-2
Pma-5-2
env-129
Ssp-11-1
env-130
Ssp-12-1
env-131
env-132
Ssp-7-1
Ssp-4-1
env-133
env-134
env-135
env-136
Ssp-13-2
env-137
Ssp-3-2
env-138
env-139
Ssp-10-2
Ssp-1-1
env-140
env-141
env-142
env-143
env-144
env-145
env-146
Ssp-9-1
env-147
env-148
env-149
env-150
5′
3′

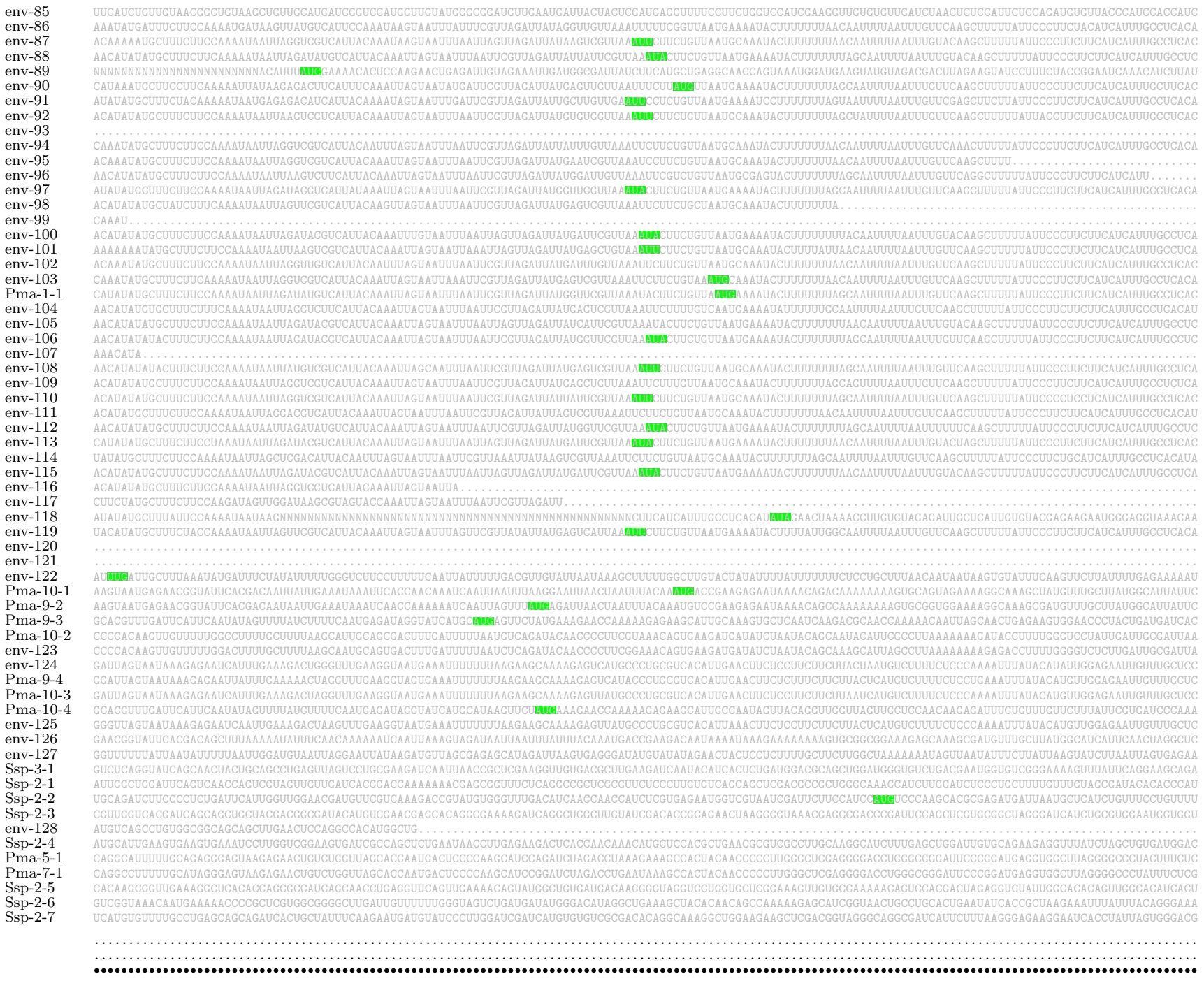

env-151
Ssp-8-2
env-152
env-153
env-154
env-155
env-156
env-157
env-158
Ssp-13-3
env-159
Ssp-3-3
env-160
env-161
env-162
Ssp-4-2
env-163
env-164
env-165
env-166
env-167
env-168
env-169
env-170
env-171
env-172
env-173
env-174
Ssp-3-4
Ssp-4-3
Ssp-1-2
env-175
env-176
Ssp-3-5
env-177
env-178
Ssp-13-4
env-179
Csp-1-1
Tel-1-2
Ssp-14-3
Nsp-1-2
Ava-1-2
Nsp-1-3
Nsp-1-4
Ava-1-3
Npu-1-1
Nsp-1-5
Mae-1-1
Csp-2-1
Cwa-1-1
Csp-2-2
Csp-3-1
Lsp-1-1
Ssp-6-1
Ssp-5-1
Mae-1-2
Csp-5-1
Csp-4-1
Csp-5-2
Cwa-1-2
5′
3′



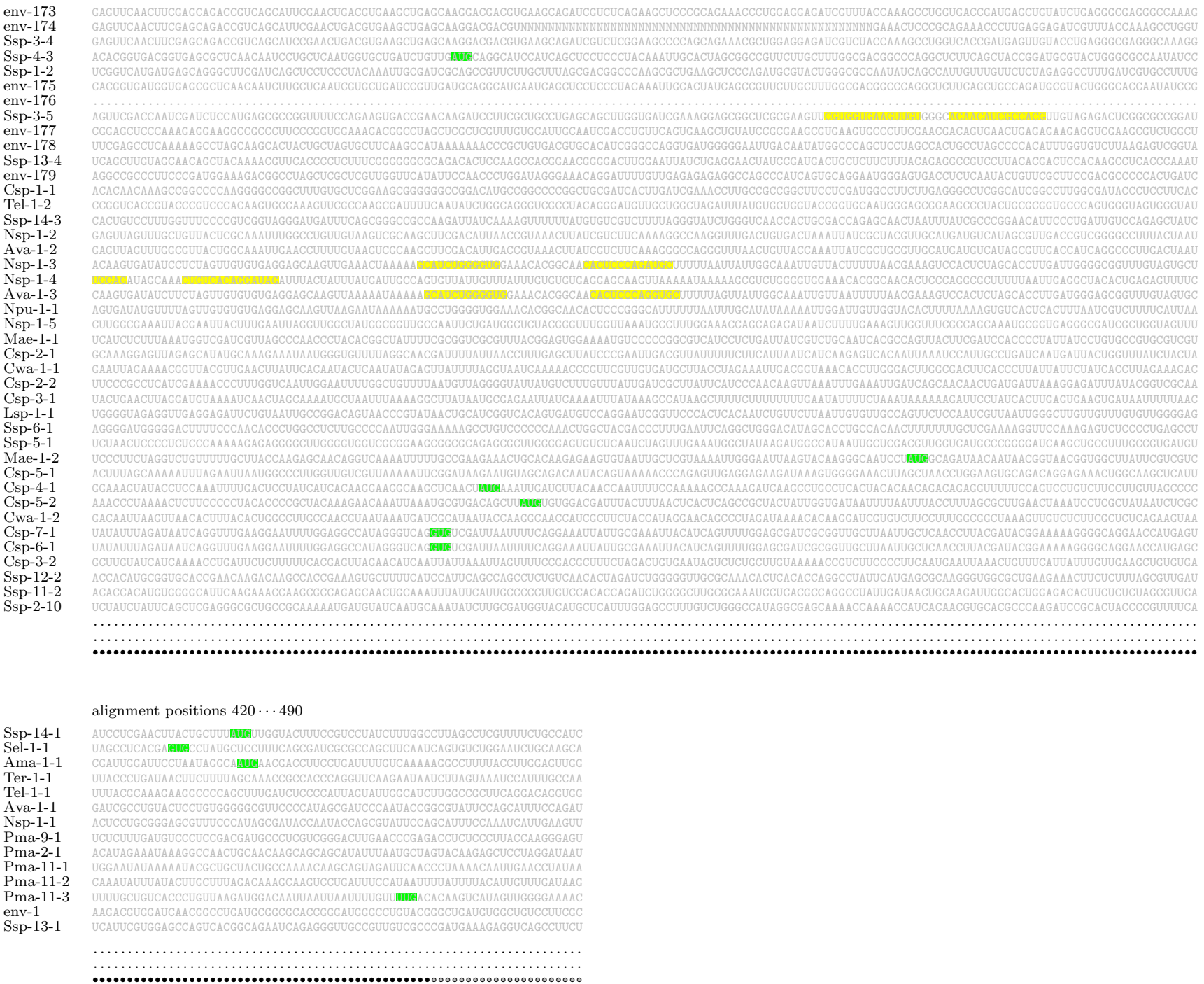
alignment positions 260 · · · 419

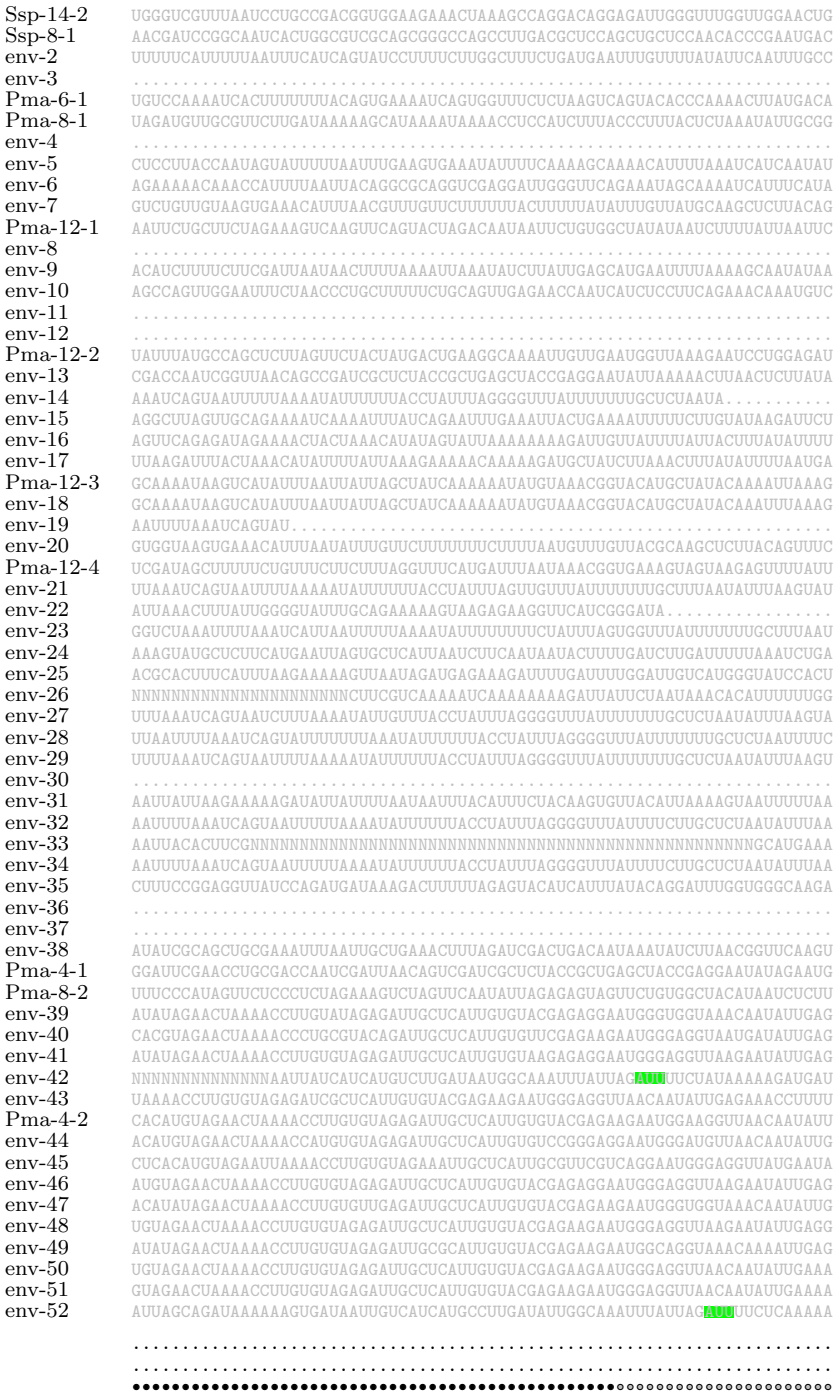


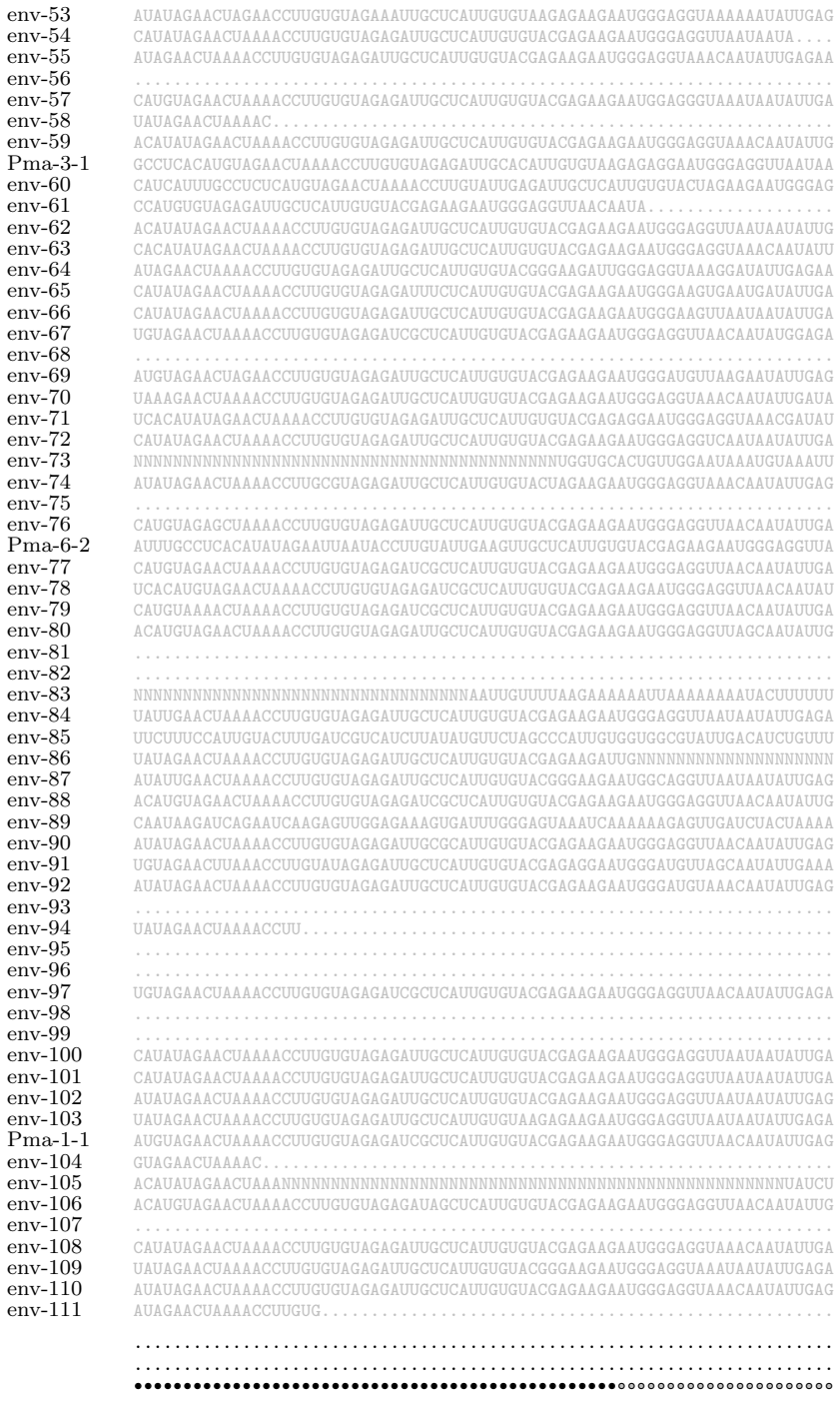

env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
Pma-4-1
Pma-8-2
env-39
env-40
env-41
env-42
env-43
Pma-4-2
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
Pma-3-1
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
Pma-6-2
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84

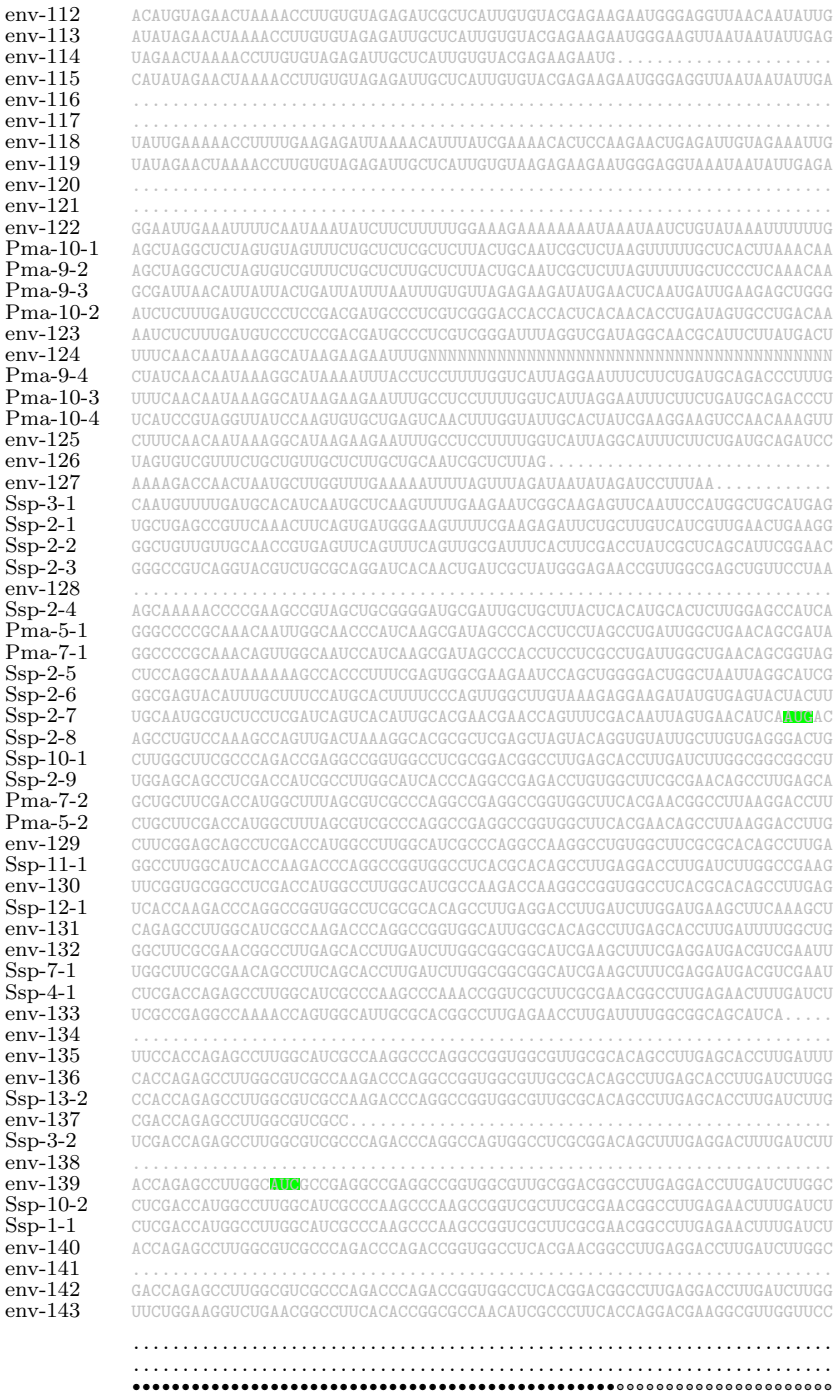

env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
Pma-1-1
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
Pma-10-1
Pma-9-2
Pma-9-3
Pma-10-2
env-123
env-124
Pma-9-4
Pma-10-3
Pma-10-4
env-125
env-126
env-127
Ssp-3-1
Ssp-2-1
Ssp-2-2
Ssp-2-3
env-128
Ssp-2-4
Pma-5-1
Pma-7-1
Ssp-2-5
Ssp-2-6
Ssp-2-7

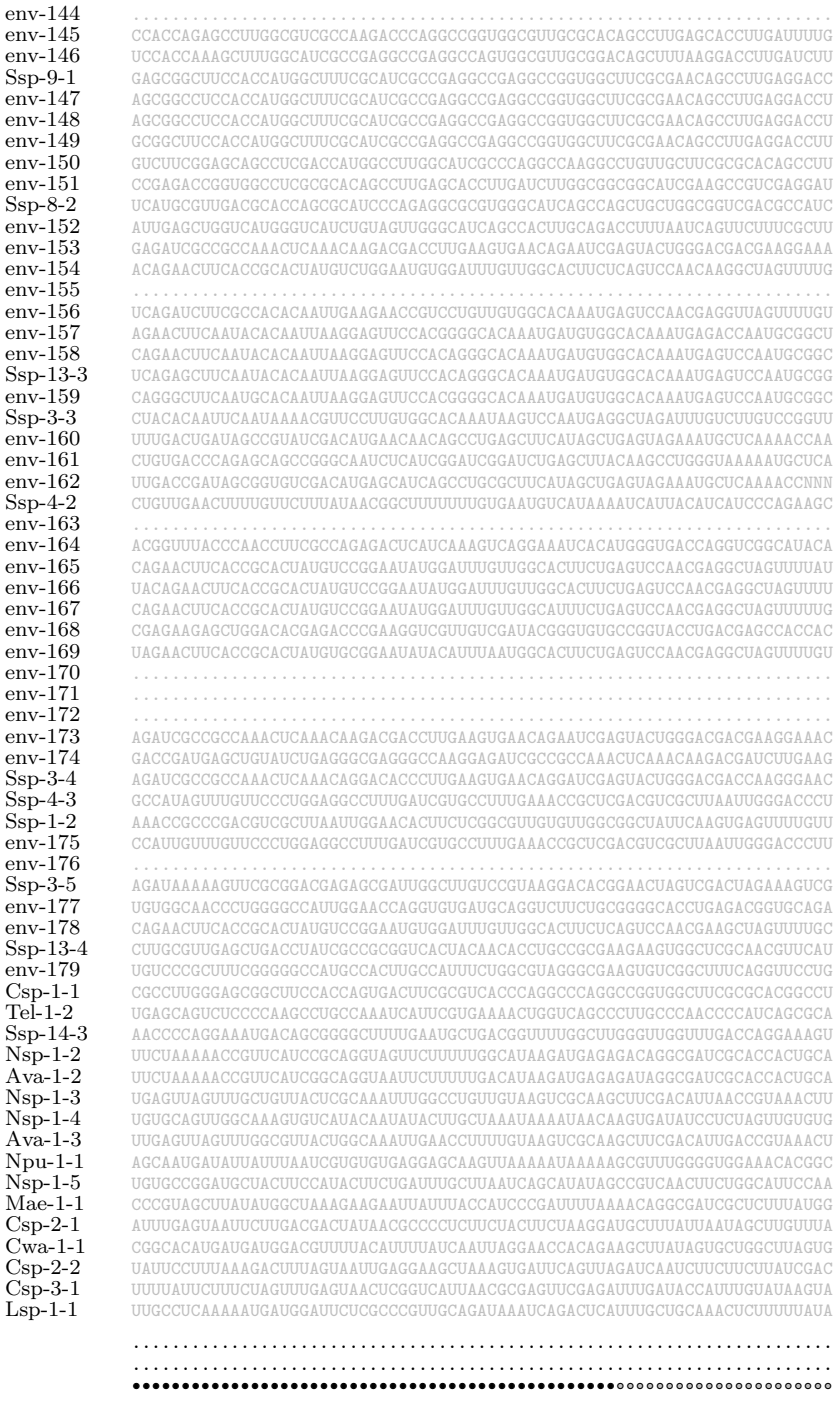
Ssp-2-8
Ssp-10-1
Ssp-2-9
Pma-7-2
Pma-5-2
env-129
Ssp-11-1
env-130
Ssp-12-1
env-131
env-132
Ssp-7-1
Ssp-4-1
env-133
env-134
env-135
env-136
Ssp-13-2
env-137
Ssp-3-2
env-138
env-139
Ssp-10-2
Ssp-1-1
env-140
env-141
env-142
env-143
env-144
env-145
env-146
Ssp-9-1
env-147
env-148
env-149
env-150
env-151
Ssp-8-2
env-152
env-153
env-154
env-155
env-156
env-157
env-158
Ssp-13-3
env-159
Ssp-3-3
env-160
env-161
env-162
Ssp-4-2
env-163
env-164
env-165
env-166
env-167
env-168
env-169
env-170
env-171
env-172

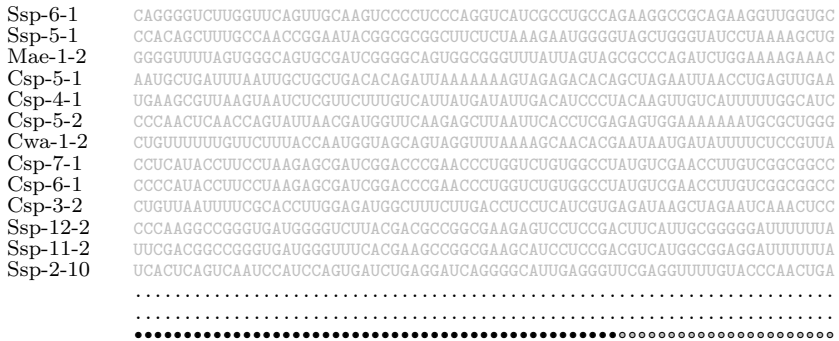

env-173
env-174
Ssp-3-4
Ssp-4-3
Ssp-1-2
env-175
env-176
Ssp-3-5
env-177
env-178
Ssp-13-4
env-179
Csp-1-1
Tel-1-2
Ssp-14-3
Nsp-1-2
Ava-1-2
Nsp-1-3
Nsp-1-4
Ava-1-3
Npu-1-1
Nsp-1-5
Mae-1-1
Csp-2-1
Cwa-1-1
Csp-2-2
Csp-3-1
Lsp-1-1
Ssp-6-1
Ssp-5-1
Mae-1-2
Csp-5-1
Csp-4-1
Csp-5-2
Cwa-1-2
Csp-7-1
Csp-6-1
Csp-3-2
Ssp-12-2
Ssp-11-2
Ssp-2-10
alignment positions 420 · · · 490
Ssp-14-1
Sel-1-1
Ama-1-1
Ter-1-1
Tel-1-1
Ava-1-1
Nsp-1-1
Pma-9-1
Pma-2-1
Pma-11-1
Pma-11-2
Pma-11-3
env-1
Ssp-13-1


Ssp-14-2
Ssp-8-1
env-2
env-3
Pma-6-1
Pma-8-1
env-4
env-5
env-6
env-7
Pma-12-1
env-8
env-9
env-10
env-11
env-12
Pma-12-2
env-13
env-14
env-15
env-16
env-17
Pma-12-3
env-18
env-19
env-20
Pma-12-4
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
Pma-4-1
Pma-8-2
env-39
env-40
env-41
env-42
env-43
Pma-4-2
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52


env-53
env-54
env-55
env-56
env-57
env-58
env-59
Pma-3-1
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
Pma-6-2
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
Pma-1-1
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111


env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
Pma-10-1
Pma-9-2
Pma-9-3
Pma-10-2
env-123
env-124
Pma-9-4
Pma-10-3
Pma-10-4
env-125
env-126
env-127
Ssp-3-1
Ssp-2-1
Ssp-2-2
Ssp-2-3
env-128
Ssp-2-4
Pma-5-1
Pma-7-1
Ssp-2-5
Ssp-2-6
Ssp-2-7
Ssp-2-8
Ssp-10-1
Ssp-2-9
Pma-7-2
Pma-5-2
env-129
Ssp-11-1
env-130
Ssp-12-1
env-131
env-132
Ssp-7-1
Ssp-4-1
env-133
env-134
env-135
env-136
Ssp-13-2
env-137
Ssp-3-2
env-138
env-139
Ssp-10-2
Ssp-1-1
env-140
env-141
env-142
env-143


env-144
env-145
env-146
Ssp-9-1
env-147
env-148
env-149
env-150
env-151
Ssp-8-2
env-152
env-153
env-154
env-155
env-156
env-157
env-158
Ssp-13-3
env-159
Ssp-3-3
env-160
env-161
env-162
Ssp-4-2
env-163
env-164
env-165
env-166
env-167
env-168
env-169
env-170
env-171
env-172
env-173
env-174
Ssp-3-4
Ssp-4-3
Ssp-1-2
env-175
env-176
Ssp-3-5
env-177
env-178
Ssp-13-4
env-179
Csp-1-1
Tel-1-2
Ssp-14-3
Nsp-1-2
Ava-1-2
Nsp-1-3
Nsp-1-4
Ava-1-3
Npu-1-1
Nsp-1-5
Mae-1-1
Csp-2-1
Cwa-1-1
Csp-2-2
Csp-3-1
Lsp-1-1



26 Cyano-2

26.1 Taxa

The taxonomy of each organism containing a putative Cyano-2 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Cyano-2 RNA in Sections 26.2 and 26.4.

abbrev. of hits	taxonomy of species
Ssp-1-1 to Ssp-1-5	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* CC9605
Ssp-2-1 to Ssp-2-5	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 8102
env-1 to env-47	environmental samples

26.2 Gene contexts

Each Cyano-2 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Cyano-2 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 26.3. The accession number of the sequence containing each Cyano-2 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 26.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	JCVI_SCAF_1096626604590	+	264	485	**RNA**→ ←? (GOS13939)
env-2	JCVI_SCAF_1096627196445	+	280	502	**RNA**→ ←? (GOS13939)
env-3	JCVI_SCAF_1101667386531	+	595	818	**RNA**→
Ssp-2-1	NC_005070.1	-	1051140	1050918	hypo→ **RNA**→ ←hypo
env-4	JCVI_SCAF_1096628040145	+	1264	1487	? (GOS11467169)→ **RNA**→ ←? (GOS13939)
env-5	JCVI_SCAF_1101668670117	-	764	566	? (GOS11039316)→ **RNA**→ ? (GOS11026)→
Ssp-1-1	NC_007516.1	+	1429118	1429312	hypo→ **RNA**→ COG1359 (COG1359)NOG29985 (NOG29985)→

env-6	JCVI_SCAF_1096626931592	-	1393	1187	**RNA**→ ? (GOS13122847)→
env-7	JCVI_SCAF_1096627964155	-	1030	832	? (GOS11250293)→ **RNA**→ ? (GOS13122847)→
env-8	JCVI_SCAF_1101668187095	+	1010	1208	? (GOS11250293)→ **RNA**→ ? (GOS12129995)→
env-9	JCVI_SCAF_1101668674017	-	1417	1219	**RNA**→ ? (GOS12129995)→
env-10	JCVI_SCAF_1096626917678	-	525	327	? (GOS11250293)→ **RNA**→ ? (GOS12129995)→
Ssp-1-2	NC_007516.1	+	1587566	1587776	←hypo **RNA**→ ←hypo
env-11	JCVI_SCAF_1101668673415	-	421	205	? (GOS11776927)→ **RNA**→
env-12	JCVI_SCAF_1096627155971	+	2428	2644	? (GOS11776927)→ **RNA**→ ? (GOS13966)→
env-13	JCVI_SCAF_1096627749470	-	426	212	←? (GOS11342131) **RNA**→
env-14	JCVI_SCAF_1101668675632	-	1073	860	**RNA**→
env-15	JCVI_SCAF_1096627965582	+	1119	1333	**RNA**→
env-16	JCVI_SCAF_1101667184623	-	634	419	**RNA**→
env-17	JCVI_SCAF_1096628048551	+	504	716	**RNA**→
Ssp-2-2	NC_005070.1	-	1129611	1129409	←hypo **RNA**→ ←hypo
env-18	JCVI_SCAF_1101667386135	-	214	12	←? (GOS11671314) **RNA**→
env-19	JCVI_SCAF_1096627526224	-	240	37	←? (GOS11944983) **RNA**→
env-20	JCVI_SCAF_1101667223140	+	321	525	**RNA**→
env-21	JCVI_SCAF_1096627961690	-	424	221	←? (GOS11944983) **RNA**→
env-22	JCVI_SCAF_1096626924384	-	2335	2132	**RNA**→ ? (GOS14498412)→
env-23	JCVI_SCAF_1096627515178	+	1041	1244	←DAP_epimerase (pfam01678)? (GOS5775) **RNA**→
env-24	JCVI_SCAF_1101668666528	-	448	245	? (GOS11760973)→ **RNA**→
env-25	JCVI_SCAF_1101668668167	-	629	425	? (GOS11116329)→ **RNA**→ ? (GOS11116329)→
env-26	JCVI_SCAF_1096627302430	+	48	252	? (GOS11116329)→ **RNA**→ ? (GOS11116329)→
env-27	JCVI_SCAF_1096627328368	-	1754	1551	←? (GOS14395810) **RNA**→ ←? (GOS13482625)
env-28	JCVI_SCAF_1096627158393	-	1145	942	**RNA**→
env-29	JCVI_SCAF_1101668672249	+	528	698	**RNA**→
env-30	JCVI_SCAF_1096627369623	-	628	424	? (GOS11116329)→ **RNA**→ ? (GOS11116329)→
Ssp-1-3	NC_007516.1	-	1291073	1290872	hypo→ **RNA**→ TFIIE (smart00531)→
env-31	JCVI_SCAF_1096628082578	-	727	523	**RNA**→ ? (GOS11713070)→
env-32	JCVI_SCAF_1096628257323	-	1683	1475	? (GOS11116329)→ **RNA**→ ? (GOS11116329)→
env-33	JCVI_SCAF_1096627794547	+	1128	1328	**RNA**→
env-34	JCVI_SCAF_1096628362723	+	603	802	**RNA**→
Ssp-1-4	NC_007516.1	-	535123	534921	←GST_C_Zeta (cd03191)GST_N_Zeta (cd03042) **RNA**→ ←hypo
Ssp-1-5	NC_007516.1	+	913565	913767	←hypo **RNA**→ hypo→
env-35	JCVI_SCAF_1096627540547	-	385	184	←DUF427 (pfam04248) **RNA**→
env-36	JCVI_SCAF_1096628046417	+	1622	1825	? (GOS13136162)→ **RNA**→
env-37	JCVI_SCAF_1101667383508	-	525	322	**RNA**→
env-38	JCVI_SCAF_1096627192266	-	6496	6293	**RNA**→ ←? (GOS11820130)
Ssp-2-3	NC_005070.1	-	1848644	1848441	←hypo **RNA**→ hypo→
env-39	JCVI_SCAF_1101667585229	+	234	437	**RNA**→
env-40	JCVI_SCAF_1096626459164	+	6	207	**RNA**→ ←? (GOS12231)
env-41	JCVI_SCAF_1101667380836	-	694	493	**RNA**→ ←? (GOS11820130)
env-42	JCVI_SCAF_1096628075702	+	466	667	**RNA**→ ←? (GOS11047726)
env-43	JCVI_SCAF_1096627183482	+	289	488	**RNA**→ ←Soj (COG1192)? (GOS1716)
Ssp-2-4	NC_005070.1	-	1132490	1132288	hypo→ **RNA**→ ←hypo
Ssp-2-5	NC_005070.1	-	372286	372086	←GAF (pfam01590) **RNA**→ ←COG3019 (COG3019)
env-44	JCVI_SCAF_1096627192266	-	3116	2919	? (GOS11402347)→ **RNA**→ **RNA**→ ←? (GOS13534)
env-45	JCVI_SCAF_1096627192266	-	3745	3548	(shown above)
env-46	JCVI_SCAF_1096628390651	-	1845	1648	**RNA**→
env-47	JCVI_SCAF_1096627798650	-	1673	1473	**RNA**→ ? (GOS13050596)→

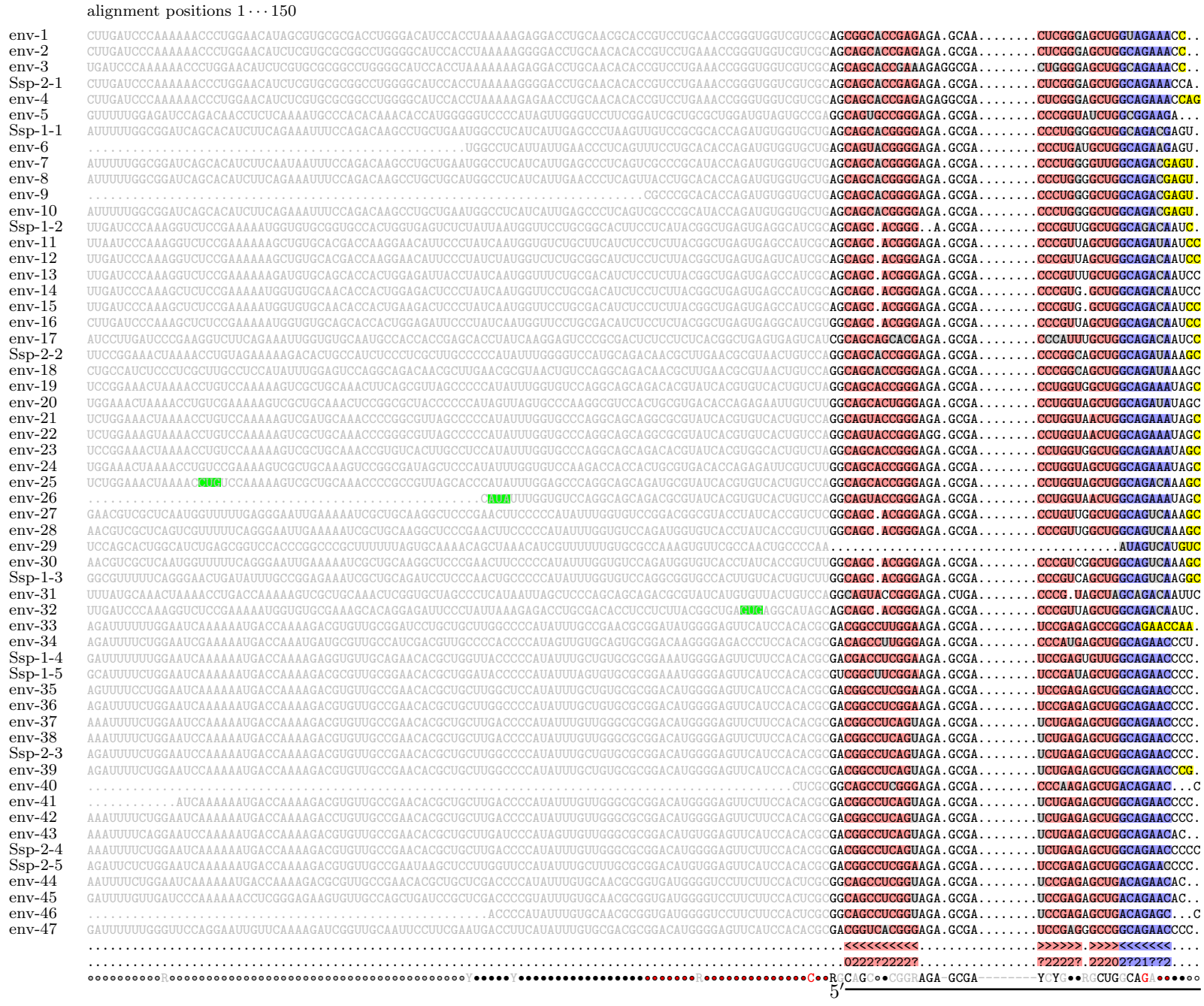
26.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 26.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Cyano-2 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

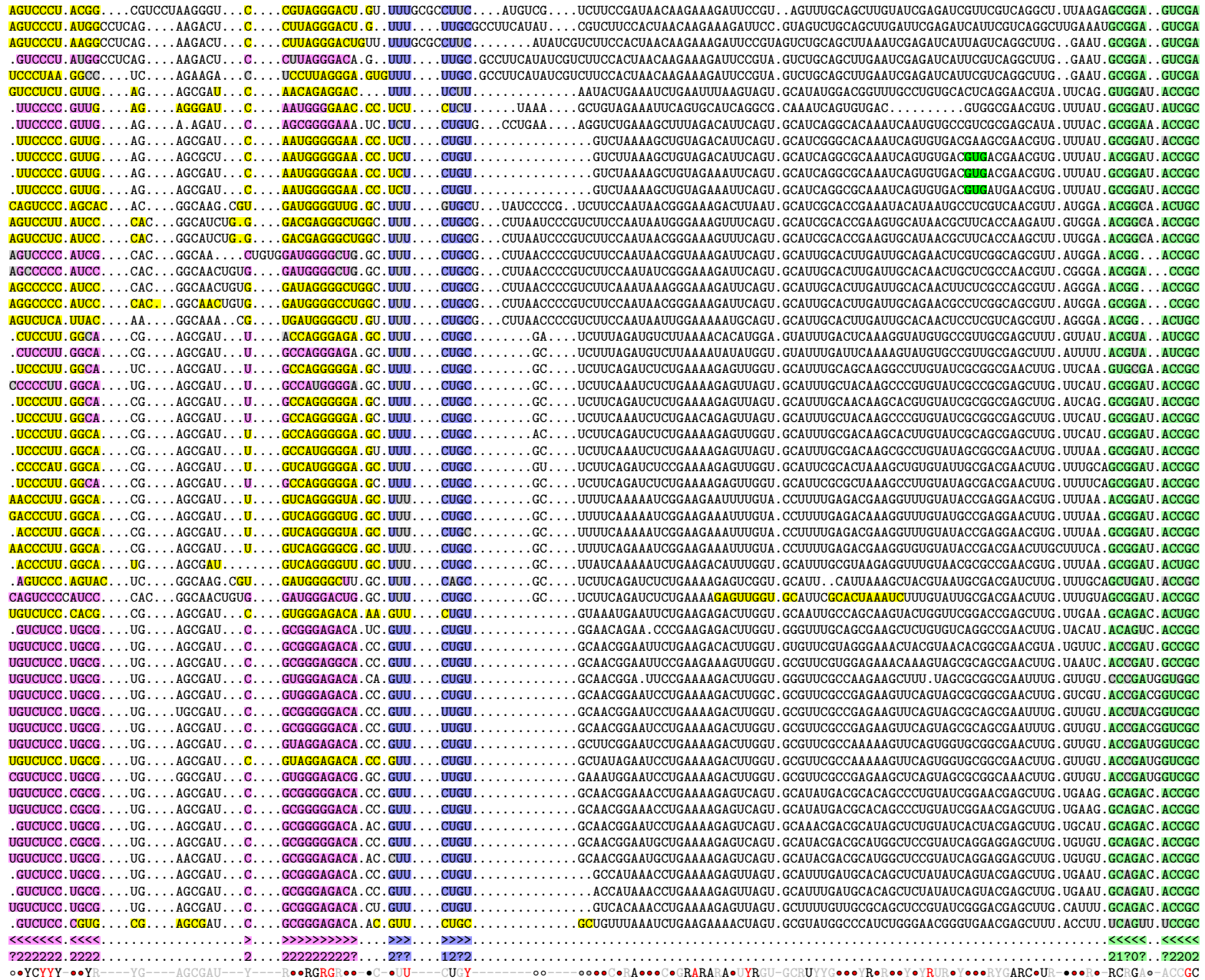
u cd03042 (1) GST_N family, Class Zeta subfamily; GSTs are cytosolic dimeric proteins involved in cellular detoxification by catalyzing the conjugation of glutathione (GSH) with a wide range of endogenous and xenobiotic alkylating agents, including carcinogens, therapeutic drugs, environmental toxins and products of oxidative stress.
u cd03191 (1) GST_C family, Class Zeta subfamily; GSTs are cytosolic dimeric proteins involved in cellular detoxification by catalyzing the conjugation of glutathione (GSH) with a wide range of endogenous and xenobiotic alkylating agents, including carcinogens, therapeutic drugs, environmental toxins, and products of oxidative stress.
d COG1192 (1) ATPases involved in chromosome partitioning [Cell division and chromosome partitioning]
D COG1359 (1) Uncharacterized conserved protein [Function unknown]
d COG3019 (1) Predicted metal-binding protein [General function prediction only]
D GOS11026 (1) ?
u GOS11039316 (1) ?
d GOS11047726 (1) ?
uD GOS11116329 (8) ?
u GOS11250293 (3) ?
u GOS11342131 (1) ?
u GOS11402347 (1) ?
u GOS11467169 (1) ?
u GOS11671314 (1) ?
D GOS11713070 (1) ?
u GOS11760973 (1) ?
u GOS11776927 (2) ?
d GOS11820130 (2) ?
u GOS11944983 (2) ?
D GOS12129995 (3) ?
d GOS12231 (1) ?
D GOS13050596 (1) ?
D GOS13122847 (2) ?
u GOS13136162 (1) ?
d GOS13482625 (1) ?
d GOS13534 (1) ?
d GOS13939 (3) ?
D GOS13966 (1) ?
u GOS14395810 (1) ?
D GOS14498412 (1) ?
d GOS1716 (1) ?
u GOS5775 (1) ?
D NOG29985 (1)
u pfam01590 (1) GAF domain.
u pfam01678 (1) Diaminopimelate epimerase.
u pfam04248 (1) Domain of unknown function (DUF427).
D smart00531 (1) Transcription initiation factor IIE;

26.4 Multiple-sequence alignment

Each Cyano-2 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 26.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Cyano-2 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

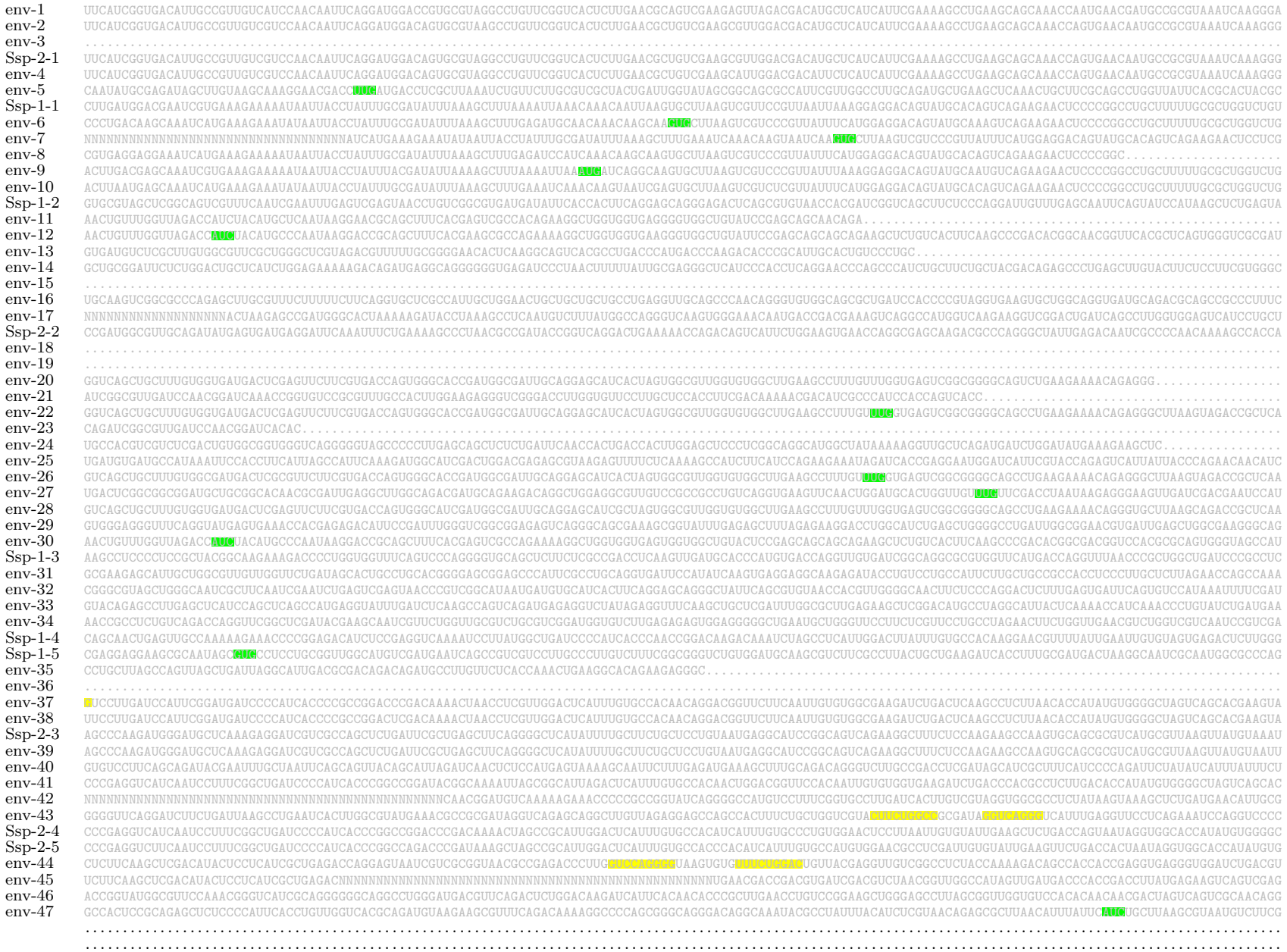

alignment positions 1 · · · 150
env-1
Ssp-2-1
env-4
env-5
Ssp-1-1
env-6
env-7
env-8
env-9
env-10
Ssp-1-2
env-11
env-12
env-13
env-14
env-15
env-16
env-17
Ssp-2-2
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
Ssp-1-3
env-31
env-32
env-33
env-34
Ssp-1-4
Ssp-1-5
env-35
env-36
env-37
env-38
Ssp-2-3
env-39
env-40
env-41
env-42
env-43
Ssp-2-4
Ssp-2-5
env-44
env-45
env-46
env-47
5′

alignment positions 151 · · · 307



alignment positions 308 · · · 467
env-1
Ssp-2-1
env-2
env-3
env-4
env-5
Ssp-1-1
env-6
env-7
env-8
env-9
env-10
Ssp-1-2
env-11
env-12
env-13
env-14
env-15
env-16
env-17
Ssp-2-2
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
Ssp-1-3
env-31
env-32
env-33
env-34
Ssp-1-4
Ssp-1-5
env-35
env-36
env-37
env-38
Ssp-2-3
env-39
env-40
env-41
env-42
env-43
Ssp-2-4
Ssp-2-5
env-44
env-45
env-46
env-47
3′

alignment positions 468 · · · 627



alignment positions 628 · · · 667

```
env-1     AAAAGGGCGCCAAGCACGGCCUUCGGUUACAAAUACUUAC
env-2     AAAAGGGCGCCAAGCACGGCAAUCAAUUUUCAAUGUUUAC
env-3     ........................................
Ssp-2-1   AAAAGGGCGCCAAGCACGGCAAUCAAUUUUCAAUGUUUAC
env-4     AAAAGGGCGCCAAGCACGGCAAUCAAUUUUCAACGUUUAC
env-5     AGUGCUUUAGAGACGAAUACAGGGAGGAAUACGUACCUGG
Ssp-1-1   GAGCAAGGUGACGGACAGGCUCUAUGCGGUUGCGGCAACG
env-6     UGAGCAAGGUUACUGACAAGCUCUAUGCGGUUGCGGCAAC
env-7     GCCUGCUUUUUGCGCUGGUCUGUGAGCAAGGUGACUGACA
env-8     ........................................
env-9     UGAGCAAGGUGACUGACAAGCUCUAUGCGGUUGCGGCGAC
env-10    UGAGCAAAGUUGCUGACAGGCUCUAUGCGGUUGCGGCAAC
Ssp-1-2   ACACAUCUCAUAGGGAUUCCGCCAAUCGAGGUCGUUAUGU
env-11    ........................................
env-12    CAGGACCUACAGCUACGUUCCGCCCCGCCCGCCUGAGCCG
env-13    ........................................
env-14    UCGCCUCACGGGCGUAGCCCACUACGCGAGUGAAUUCGUG
env-15    ........................................
env-16    AUCCCUGCAACUCCCCCUCGUCGAUAAAACUCCACUACUC
env-17    GGCCUUCAGCAAGAUAAUUCUUCCCUUAGUCCUGCUUGGU
Ssp-2-2   GAGACAAUCUCGAGGAGCCGGGUUGUCUCAAAAUGACGUG
env-18    ........................................
env-19    ........................................
env-20    ........................................
env-21    ........................................
env-22    AGUGGCAGGCACAUCUUGAACCUGCAGCCCGAGGGGGUCU
env-23    ........................................
env-24    ........................................
env-25    GCGUAUCGCUUGAUUUGCCUCGAAGUAAAGGAGAUCAUCU
env-26    GUGGCAGGCGCAUCUUGAACCUGCAGCCCGAGGGGGUCUG
env-27    GCCAAUCGCAAACCAGCAAUGCGCUUGCGAGCCUUGCGGC
env-28    GCAGUGAGCGCAUCUUGAGCCUGCAGCCCGAGGGGGUCUG
env-29    CUAAAGAUUGCGUUCCGCCGCACCCGCCUGAGCCGAUGAC
env-30    CAGGACGUACAGCUACGUUCCGCCCCGCCCUCCUGAGCCG
Ssp-1-3   UGGAGCCCCGCAGCGAUAGUUUGCAUUUUUUCUGAGAAUG
env-31    CCCAAGAAGCGCAGGCGAAAUAGCGGUAGGCACCAGGUCU
env-32    AACACUUCUCAUCAGACUUCCUAUAGUUGUGAUCAAUAUG
env-33    GCGCAACUAUUCACCGAUACCGAAUUGGUAAUCCAUUAGU
env-34    AGCGACACGAUUUCCAACCUCUUUCUGUUCGCCCCCAUUU
Ssp-1-4   UGACACCAUAUGUGGGGCUGGUCAGCACGUAGUAGCUGGG
Ssp-1-5   CAGAAAGCCAAGAAGCUCCAUUGCUGAUCAGGCUUCGACU
env-35    ........................................
env-36    ........................................
env-37    GCUGAGCCAUAUUUAGGUUUCCCCCAUCACCCAGCCCGUU
env-38    GCUGAGCCAUAUUUAGGUUUCCCCCAUCACCCAGCCCGUU
Ssp-2-3   AAAAUUUUUUAAUCCAAAUGAAUAGCUGGUUUGCAGCGAU
env-39    AAAUUUUUUUAAAUCCAAAUGAAUAGCUGGUUUGCAGCGA
env-40    UUCAAUGUUUUCAGGUAAGUAAUGCAGCUUUGCCGAAUCU
env-41    GAAGUAGCUGAGCCAUAUUUAGGUUUCCCCCAUCACCCAG
env-42    AGCCACUUCUUCGCGGCAGGUGUUGUAACGCCCGCGGCGA
env-43    GUUCCUUGGCCUGGUGAUCUGCGCCUUGUAGAAGCAAGGU
Ssp-2-4   UGGUCAGUACUUGGCGGCUAAGCCAUAUUUAGGUUUCCCC
Ssp-2-5   GGGCUGGUCAGUACUAGGCGGCUAAGCCAUAUUUAGGUUU
env-44    GCCAACACCUUGAAGCCCGCGUACUCAUCCAUAGGCUUUC
env-45    AAGUGUGAUUUGGAGGAUGGGAGUCCUGUUGUCGGUGUCG
env-46    CGAUCUCAGCUGCAGUUGGUAUCAACGCAGCAAAUGGUUG
env-47    GGCUCUGGUGAUAGCUGGUACCAGGAUGCAGCUGCCGCGC
          ........................................
          ........................................
          °°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°
```

27 Desulfotalea-1

27.1 Taxa

The taxonomy of each organism containing a putative Desulfotalea-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Desulfotalea-1 RNA in Sections 27.2 and 27.4.

abbrev. of hits	taxonomy of species
Dha-1-1	Bacteria Firmicutes Clostridia Clostridiales Peptococcaceae *Desulfitobacterium hafniense* Y51
Dps-1-1 to Dps-1-20	Bacteria Proteobacteria Deltaproteobacteria Desulfobacterales Desulfobulbaceae *Desulfotalea psychrophila* LSv54
Sba-1-1 to Sba-1-8	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella baltica* OS155
Sba-2-1 to Sba-2-10	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella baltica* OS185
Sba-3-1 to Sba-3-10	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella baltica* OS195
Sba-4-1 to Sba-4-6	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella baltica* OS223
Sde-1-1 to Sde-1-10	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella denitrificans* OS217
Sfr-1-1 to Sfr-1-2	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella frigidimarina* NCIMB 400
Sha-1-1 to Sha-1-2	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella halifaxensis* HAW-EB4
Slo-1-1 to Slo-1-9	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella loihica* PV-4
Son-1-1 to Son-1-8	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella oneidensis* MR-1
Spe-1-1 to Spe-1-7	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella pealeana* ATCC 700345
Spu-1-1 to Spu-1-11	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella putrefaciens* 200
Spu-2-1 to Spu-2-10	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella putrefaciens* CN-32
Sse-1-1 to Sse-1-10	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella sediminis* HAW-EB3
Ssp-1-1 to Ssp-1-9	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella sp.* ANA-3
Ssp-2-1 to Ssp-2-13	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella sp.* MR-4
Ssp-3-1 to Ssp-3-9	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella sp.* MR-7
Ssp-4-1 to Ssp-4-8	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella sp.* W3-18-1
Swo-1-1 to Swo-1-15	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella woodyi* ATCC 51908
Spr-1-1 to Spr-1-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Serratia proteamaculans* 568
Osp-1-1 to Osp-1-5	Bacteria Proteobacteria Gammaproteobacteria Oceanospirillales *Oceanobacter sp.* RED65
Epe-1-1 to Epe-1-2	Bacteria Proteobacteria Gammaproteobacteria sulfur-oxidizingsymbionts *Endoriftia persephone* 'Hot96_1+Hot96_2'
env-1 to env-41	environmental samples

27.2 Gene contexts

Each Desulfotalea-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Desulfotalea-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 27.3. The accession number of the sequence containing each Desulfotalea-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 27.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Dps-1-1	NC_006138.1	-	2160819	2160757	←MdaB (COG2249)Flavodoxin_2 (pfam02525) **RNA→** **RNA→** **RNA→** **RNA→** ←5S ribosomal RNA
Dps-1-2	NC_006138.1	-	2159792	2159730	(shown above)
Dps-1-12	NC_006138.1	-	2160915	2160863	(shown above)
Dps-1-14	NC_006138.1	-	2159888	2159836	(shown above)
Dps-1-4	NC_006138.1	+	1458168	1458230	hypo→ **RNA→** **RNA→** **RNA→** hypo→

Dps-1-5	NC_006138.1	+	1458521	1458582	(shown above)
Dps-1-15	NC_006138.1	+	1458072	1458124	(shown above)
Dha-1-1	NC_007907.1	-	3254431	3254370	MGS_CPS_II (cd01424)CarB (COG0458)CPSase_L_D3 (pfam02787)→ **RNA**→ UbiB (COG0543)→
Epe-1-1	NZ_AASF01001432.1	+	2173	2233	←hypo **RNA**→ ←hypo
env-1	JCVI_SCAF_1096626308176	-	830	769	**RNA**→ ←ribokinase_group_A (cd01942)? (GOS4908)
env-2	JCVI_SCAF_1096627348977	+	2300	2361	NTP_transf_2 (pfam01909)? (GOS11052049)→ **RNA**→ ←Ala tRNA
Epe-1-2	NZ_AASF01000563.1	-	448	387	TOPRIM domain protein→ **RNA**→ TOPRIM domain protein→
env-3	wf1_AGZO6492_g2	-	920	859	hypo→ **RNA**→ hypo→
Sba-1-1	NC_009052.1	-	120367	120306	←hypo **RNA**→ ←HTH_XRE (smart00530)
Ssp-3-1	NC_008322.1	-	489145	489084	COG5635 (COG5635)→ **RNA**→ COG3046 (COG3046)→
Ssp-1-1	NC_008577.1	+	3929063	3929124	←hypo **RNA**→ ←COG0412 (COG0412)
Son-1-1	NC_004347.1	-	1172282	1172221	rve (pfam00665)→ **RNA**→ hypo→
Son-1-2	NC_004347.1	+	4341375	4341436	hypo→ **RNA**→ COG3046 (COG3046)→
Sba-4-1	NZ_ABAU01000009.1	-	74558	74497	←hypo (cons) **RNA**→ ←OmpA_membrane (pfam01389)
Ssp-4-1	NC_008750.1	-	620879	620818	Transgly (pfam00912)→ **RNA**→ hypo→
Sba-1-2	NC_009052.1	-	517146	517085	hypo→ **RNA**→ ←COG1853 (COG1853)
Son-1-3	NC_004347.1	-	488661	488600	←hypo **RNA**→ ←NUDIX (pfam00293)
Sba-3-1	NC_009997.1	-	1643237	1643176	COG1556 (COG1556)→ **RNA**→ AAA (pfam00004)→
Sba-4-2	NZ_ABAU01000042.1	+	9975	10036	COG1556 (COG1556)→ **RNA**→ AAA (pfam00004)→
Sba-3-2	NC_009997.1	+	1893207	1893268	hypo→ **RNA**→ ←hypo
Sba-1-3	NC_009052.1	-	1546107	1546046	COG1556 (COG1556)→ **RNA**→ ATPase central domain-containing protein→
Sba-2-1	NC_009665.1	-	1604137	1604076	COG1556 (COG1556)→ **RNA**→ AAA (pfam00004)→
Sse-1-1	NC_009831.1	-	3283522	3283462	←hypo **RNA**→ hypo→
env-4	soil_XZS79699_x1	+	711	771	Imidazolone-5PH (cd01296)→ **RNA**→ ←hypo
env-5	wf1_AGZO16092_g2	+	815	876	**RNA**→
Osp-1-1	NZ_AAQH01000003.1	+	35782	35843	tRNA-Arg→ **RNA**→ hypo→
env-6	JCVI_SCAF_1096627348411	-	5981	5920	←NtrB (COG3852)? (GOS336) **RNA**→ ←Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)
env-7	wf1_Contig1252	+	513	574	Exopolyphosphatase(EC:3.6.1.40)→ **RNA**→ ←Phosphoenolpyruvate carboxykinase (ATP)
env-8	wf1_Contig1312	-	813	752	**RNA**→ ←PEPCK_ATP (cd00484)
env-9	wf1_Contig2174	-	228	167	Valyl-tRNA synthetase→ **RNA**→
env-11	wf1_Contig2174	-	280	219	(shown above)
env-10	wf2_AHAA10370_x1	-	323	263	←hypo **RNA**→
env-12	JCVI_SCAF_1101668564577	-	673	611	**RNA**→
Osp-1-2	NZ_AAQH01000022.1	+	41614	41675	←OraA (COG2137) **RNA**→ ←RecA (pfam00154)
Dps-1-6	NC_006138.1	-	827237	827176	←Nuc_deoxyrib_tr (pfam05014)NOG22301 (NOG22301) **RNA**→ ←5S ribosomal RNA
Dps-1-7	NC_006138.1	-	1424102	1424040	←Tar (COG0840)Cache (pfam02743) **RNA**→ ←AlsT (COG1115)
Dps-1-8	NC_006138.1	+	2324776	2324838	←hypo **RNA**→ ←hypo
Dps-1-9	NC_006138.1	+	1445432	1445494	←hypo **RNA**→ **RNA**→ ←hypo
Dps-1-11	NC_006138.1	+	1445882	1445944	(shown above)
Dps-1-10	NC_006138.1	+	1446411	1446473	←hypo **RNA**→ ←5S ribosomal RNA
Dps-1-13	NC_006138.1	-	812320	812268	←hypo **RNA**→ **RNA**→ ←hypo
Dps-1-3	NC_006138.1	-	812225	812163	(shown above)
Dps-1-16	NC_006138.1	-	2049951	2049888	←COG2968 (COG2968) **RNA**→ ←AccD (COG0777)AccA (COG0825)
Dps-1-17	NC_006138.1	-	2735941	2735879	←hypo **RNA**→ ←hypo
Dps-1-18	NC_006139.1	-	37032	36970	hypo→ **RNA**→ hypo→
Dps-1-19	NC_006138.1	+	2392578	2392640	←Met_dep_hydrolase_A (cd01299) **RNA**→ hypo→
Dps-1-20	NC_006138.1	+	1448411	1448470	hypo→ **RNA**→ hypo→
Osp-1-3	NZ_AAQH01000028.1	-	29628	29567	23S ribosomal RNA→ **RNA**→ 5S ribosomal RNA→
Osp-1-4	NZ_AAQH01000004.1	+	4024	4085	hypo→ **RNA**→ 5S ribosomal RNA→
Osp-1-5	NZ_AAQH01000014.1	-	77617	77556	hypo→ **RNA**→ 5S ribosomal RNA→
env-13	JCVI_SCAF_1096627235929	-	157	96	? (GOS14479710)→ **RNA**→ ? (GOS14479710)→
Spr-1-1	NC_009832.1	-	1300562	1300501	CSD (pfam00313)→ **RNA**→ ←GGDEF (cd01949)

Spr-1-2	NC_009832.1	+	3202354	3202415	←TauA (COG0715) **RNA**→ ←AmiR (COG3707)
Spr-1-3	NC_009832.1	+	2919065	2919126	←hypo **RNA**→ ←acidPPc (smart00014)
Sse-1-2	NC_009831.1	+	711934	711994	hypo→ **RNA**→ AlsT (COG1115)→
Sse-1-3	NC_009831.1	+	3777195	3777255	←UbiG (COG2227) **RNA**→ ProS (COG0442)→
Sse-1-4	NC_009831.1	+	2552830	2552889	hypo→ **RNA**→ GntR (COG1802)→
Slo-1-1	NC_009092.1	+	1060677	1060738	COG3547 (COG3547)→ **RNA**→ ←Ig family protein
Slo-1-2	NC_009092.1	-	2651592	2651531	←hypo **RNA**→ ←Abhydrolase_1 (pfam00561)
Swo-1-1	NC_010506.1	+	774208	774320	←COG1359 (COG1359) **RNA**→ ←PrfB (COG1186)
Slo-1-3	NC_009092.1	+	4021098	4021159	←hypo **RNA**→ SpeB (COG0010)→
Sse-1-5	NC_009831.1	-	5412662	5412549	DUF890 (pfam05971)→ **RNA**→ ←COG1910 (COG1910)
Swo-1-2	NC_010506.1	-	4595466	4595405	hypo→ **RNA**→ ←FBPase (cd00354)
Swo-1-3	NC_010506.1	-	4380115	4380054	hypo→ **RNA**→ **RNA**→ hypo→
Swo-1-14	NC_010506.1	-	4380230	4380170	(shown above)
Swo-1-4	NC_010506.1	-	5910922	5910861	Nucleoside_tra2 (pfam01773)→ **RNA**→ ←DUF1006 (pfam06224)
Swo-1-5	NC_010506.1	+	5927100	5927161	hypo→ **RNA**→ hypo→
Swo-1-6	NC_010506.1	+	4600300	4600361	hypo→ **RNA**→ ←Transposase_17 (pfam01797)
Swo-1-7	NC_010506.1	+	3624166	3624227	←hypo **RNA**→ ←hypo
Slo-1-4	NC_009092.1	-	3688293	3688232	hypo→ **RNA**→ ←rDP_like (cd01301)
Slo-1-5	NC_009092.1	-	1186360	1186299	hypo→ **RNA**→ Pfs (COG0775)→
Slo-1-6	NC_009092.1	-	864553	864492	PAS (cd00130)GGDEF (cd01949)→ **RNA**→ ←hypo
Sse-1-6	NC_009831.1	-	2551176	2551116	←hypo **RNA**→ ←GpmB (COG0406)
Spe-1-1	NC_009901.1	-	3591093	3591033	COG3547 (COG3547)→ **RNA**→ **RNA**→ hypo→
Spe-1-5	NC_009901.1	-	3590797	3590736	(shown above)
Sha-1-1	NC_010334.1	-	2712982	2712922	hypo→ **RNA**→ ←PMSR (pfam01625)
Spe-1-2	NC_009901.1	-	4953283	4953223	AbgB (COG1473)→ **RNA**→ ←hypo
Slo-1-7	NC_009092.1	-	3066234	3066173	hypo→ **RNA**→ ←COG1733 (COG1733)
Spe-1-3	NC_009901.1	+	816712	816773	Bcp (COG1225)→ **RNA**→ ←RIO1 (COG1718)
Sse-1-7	NC_009831.1	+	5439863	5439924	←OmpA_membrane (pfam01389) **RNA**→ ←COG2819 (COG2819)
Sfr-1-1	NC_008345.1	+	2377787	2377900	GreA (COG0782)→ **RNA**→ hypo→
Sfr-1-2	NC_008345.1	+	2432452	2432565	hypo→ **RNA**→ COG4584 (COG4584)→
env-14	JCVI_SCAF_1096627390099	-	160749	160688	? (GOS11588659)→ **RNA**→ ←? (GOS11627891)
Ssp-1-2	NC_008577.1	-	2788875	2788814	hypo→ **RNA**→ ←hypo
env-15	JCVI_SCAF_1096627388991	-	4053	3992	COG4271 (COG4271)? (GOS11522405)→ **RNA**→ COG4711 (COG4711)? (GOS12025990)→
Spu-2-1	NC_009438.1	+	1566542	1566603	hypo→ **RNA**→ ←decaheme cytochrome c
Ssp-4-2	NC_008750.1	-	3101842	3101781	hypo→ **RNA**→ ←decaheme cytochrome c
Spu-1-1	NZ_AAWY01000022.1	+	34792	34853	Transposase_32 (pfam04986)→ **RNA**→ ←decaheme cytochrome c
Slo-1-8	NC_009092.1	-	3928538	3928477	hypo→ **RNA**→ ←DUF481 (pfam04338)
Ssp-2-1	NC_008321.1	+	3905760	3905821	DegQ (COG0265)→ **RNA**→ COG1315 (COG1315)→
Ssp-3-2	NC_008322.1	-	794304	794243	DegQ (COG0265)→ **RNA**→ COG1315 (COG1315)→
Sde-1-1	NC_007954.1	-	334646	334533	←SSL2 (COG1061)COG3886 (COG3886) **RNA**→ ←COG2253 (COG2253)NOG14035 (NOG14035)
Spe-1-4	NC_009901.1	+	3911360	3911421	←hypo **RNA**→ ←hypo
Swo-1-8	NC_010506.1	-	4687364	4687303	←hypo **RNA**→ ←COG4146 (COG4146)
Sha-1-2	NC_010334.1	-	968451	968390	←hypo **RNA**→ ←SdhA (COG1053)
Spe-1-6	NC_009901.1	-	1415607	1415546	COG3547 (COG3547)→ **RNA**→ RecN (COG0497)→
Spe-1-7	NC_009901.1	+	898673	898734	←hypo **RNA**→ Med (COG1744)→
Sde-1-2	NC_007954.1	-	1309545	1309484	←COG2963 (COG2963) **RNA**→ INT_IntI (cd01193)→
Sba-4-3	NZ_ABAU01000046.1	+	18198	18259	Ada (COG0350)HTH_ARAC (smart00342)→ **RNA**→ COG1741 (COG1741)→
Ssp-3-3	NC_008322.1	+	1764328	1764389	UbiE (COG2226)→ **RNA**→ ←XthA (COG0708)
Sde-1-3	NC_007954.1	-	733198	733137	←putative lipoprotein **RNA**→ ←GST_C_Ure2p_like (cd03178)GST_N_Ure2p_like (cd03048)
Sde-1-4	NC_007954.1	-	4054983	4054922	←Transposase_11 (pfam01609)NOG07925 (NOG07925) **RNA**→ ←DadA (COG0665)
Sde-1-5	NC_007954.1	-	2843798	2843687	hypo→ **RNA**→ ←ManC (COG0662)
env-16	JCVI_SCAF_1096627389125	-	209051	208990	? (GOS14347974)→ **RNA**→ ←? (GOS13516489)

Ssp-4-3	NC_008750.1	-	3512493	3512432	hypo→ **RNA**→ DPBB_1 (pfam03330)→
Spu-1-2	NZ_AAWY01000014.1	+	74425	74486	hypo (cons)→ **RNA**→ ←Transposase_20 (pfam02371)
Spu-2-2	NC_009438.1	+	1220337	1220398	hypo→ **RNA**→ RlpA (COG0797)→
Sba-4-4	NZ_ABAU01000018.1	-	43511	43450	protein of unknown function DUF1568→ **RNA**→ ←COG3314 (COG3314)
Sba-3-3	NC_009997.1	-	659413	659352	←HSDR_N (pfam04313) **RNA**→ ←COG3258 (COG3258)
Sba-1-4	NC_009052.1	+	4423420	4423481	←HSDR_N (pfam04313) **RNA**→ ←COG2863 (COG2863)
Sba-2-2	NC_009665.1	-	3818743	3818682	CgeB family protein→ **RNA**→ ←COG3314 (COG3314)
Sde-1-6	NC_007954.1	+	2349930	2349991	hypo→ **RNA**→ ←Pgm (COG0033)
Spu-1-3	NZ_AAWY01000008.1	-	61270	61209	←Transposase_9 (pfam01548)Transposase_20 (pfam02371) **RNA**→ COG3673 (COG3673)→
Spu-1-4	NZ_AAWY01000024.1	-	27004	26943	SEC-C (pfam02810)→ **RNA**→ ←COG3150 (COG3150)
Ssp-4-4	NC_008750.1	-	2122035	2121974	UPF0167 (pfam03691)→ **RNA**→ ←COG3150 (COG3150)
Spu-2-3	NC_009438.1	-	4061889	4061828	hypo→ **RNA**→ ←hypo
Ssp-4-5	NC_008750.1	+	487038	487099	hypo→ **RNA**→ ←hypo
Spu-1-5	NZ_AAWY01000028.1	+	10812	10873	hypo→ **RNA**→ ←hypo (cons)
Sba-3-4	NC_009997.1	-	3353343	3353282	hypo→ **RNA**→ ←XthA (COG0708)
Sba-1-5	NC_009052.1	+	1266290	1266351	hypo→ **RNA**→ DPBB_1 (pfam03330)→
Sba-3-5	NC_009997.1	+	1414133	1414194	Transposase_32 (pfam04986)→ **RNA**→ DPBB_1 (pfam03330)→
Spu-2-4	NC_009438.1	-	3422364	3422303	Acetyltransf_1 (pfam00583)→ **RNA**→ COG3673 (COG3673)→
Spu-2-5	NC_009438.1	+	2495246	2495307	hypo→ **RNA**→ ←COG3150 (COG3150)
Sba-1-6	NC_009052.1	-	4071734	4071673	Ada (COG0350)HTH_ARAC (smart00342)→ **RNA**→ COG1741 (COG1741)→
Sba-3-6	NC_009997.1	+	1062421	1062482	←COG3344 (COG3344) **RNA**→ COG1741 (COG1741)→
Ssp-3-4	NC_008322.1	-	2627196	2627136	COG5293 (COG5293)→ **RNA**→ ←hypo
Son-1-4	NC_004347.1	+	3939830	3939891	←hypo **RNA**→ CsdB (COG0520)→
env-17	JCVI_SCAF_1096627385280	-	42059	41998	RbsB (COG1879)? (GOS5488)→ **RNA**→ predicted protein→
env-18	JCVI_SCAF_1096627195755	+	1607	1668	**RNA**→ ←GlnQ (COG1126)? (GOS2568)
Ssp-1-3	NC_008577.1	-	1067869	1067808	putative lipoprotein→ **RNA**→ ←dipeptidyl-peptidase 7
Ssp-3-5	NC_008322.1	-	1090676	1090615	hypo→ **RNA**→ ←dipeptidyl-peptidase 7
Ssp-2-2	NC_008321.1	-	1066577	1066516	hypo→ **RNA**→ COG3547 (COG3547)→
Ssp-2-3	NC_008321.1	+	3245705	3245766	S_TKc (cd00180)→ **RNA**→ ←DAP2 (COG1506)
env-19	JCVI_SCAF_1096627387715	-	52913	52852	? (GOS418)→ **RNA**→ ←XthA (COG0708)? (GOS6435)
env-20	JCVI_SCAF_1096627208462	-	192427	192366	←INT_Lambda_C (cd00800)? (GOS1694) **RNA**→ ←? (GOS11627891)
Ssp-2-4	NC_008321.1	+	3614162	3614223	hypo→ **RNA**→ COG3547 (COG3547)→
env-21	JCVI_SCAF_1096627191725	-	132397	132336	? (GOS11166260)→ **RNA**→ ←TrkG (COG0168)? (GOS6072)
Sba-4-5	NZ_ABAU01000030.1	-	9236	9175	Transposase_11 (pfam01609)→ **RNA**→ ←protein of unknown function DUF1379
Sde-1-7	NC_007954.1	+	1402477	1402538	Transposase_11 (pfam01609)NOG07925 (NOG07925)→ **RNA**→ ←hypo
Sba-2-3	NC_009665.1	+	1024503	1024564	Ada (COG0350)HTH_ARAC (smart00342)→ **RNA**→ COG1741 (COG1741)→
env-22	JCVI_SCAF_1096627390099	+	205614	205675	GGDEF (cd01949)? (GOS13550016)→ **RNA**→ ←Pro tRNA
Sse-1-8	NC_009831.1	-	2080558	2080497	←hypo **RNA**→ hypo→
Swo-1-9	NC_010506.1	-	5699016	5698955	←OLE1 (COG1398) **RNA**→ ←HTH_XRE (cd00093)
Swo-1-10	NC_010506.1	+	1288851	1288911	hypo→ **RNA**→ ←YcaC_related (cd01012)
Swo-1-11	NC_010506.1	+	567231	567291	hypo→ **RNA**→ Peptidase_M48 (pfam01435)→
Swo-1-12	NC_010506.1	-	1261348	1261288	←ANK (cd00204) **RNA**→ ←hypo
Swo-1-13	NC_010506.1	+	4870073	4870133	hypo→ **RNA**→ ←MdlB (COG1132)
Sse-1-9	NC_009831.1	-	248456	248395	hypo→ **RNA**→ ←Cytochrom_NNT (pfam03264)
Swo-1-15	NC_010506.1	+	1630791	1630852	←hypo **RNA**→ ←Peptidase_M28 (pfam04389)
Sde-1-8	NC_007954.1	-	4160783	4160723	←RecG (COG1200) **RNA**→ ←CitB (COG2197)
Sde-1-9	NC_007954.1	+	4438735	4438848	←UPF0153 (pfam03692) **RNA**→ ←tRNA-Asp
Sde-1-10	NC_007954.1	+	2847542	2847603	←hypo **RNA**→ ←COG3791 (COG3791)
Slo-1-9	NC_009092.1	+	681172	681233	hypo→ **RNA**→ UbiE (COG2226)→
Ssp-2-5	NC_008321.1	-	2558580	2558467	hypo→ **RNA**→ ←hypo
Sba-2-4	NC_009665.1	-	1891830	1891717	←COG3593 (COG3593) **RNA**→ hypo→
Sba-2-5	NC_009665.1	+	4934353	4934413	←hypo **RNA**→ ←UbiG (COG2227)

Sba-1-7	NC_009052.1	-	1261597	1261536	←Transposase_11 (pfam01609) **RNA**→ ←Aldedh (pfam00171)
Ssp-1-4	NC_008577.1	+	1069306	1069367	←putative lipoprotein **RNA**→ OmpA (COG2885)→
Spu-2-6	NC_009438.1	-	1219059	1218998	←hypo **RNA**→ ←Aldedh (pfam00171)
Spu-1-6	NZ_AAWY01000014.1	-	73147	73086	←hypo (cons) **RNA**→ ←aldehyde dehydrogenase
Son-1-5	NC_004347.1	-	2320864	2320751	hypo→ **RNA**→ ←LMWPc (cd00115)
env-23	JCVI_SCAF_1096627388974	-	239576	239515	Amidase (pfam01425)? (GOS672)→ **RNA**→ **RNA**→ COG1741 (COG1741)? (GOS5791)→
env-28	JCVI_SCAF_1096627388974	-	239343	239283	(shown above)
env-24	JCVI_SCAF_1096627388987	-	40406	40345	? (GOS14131473)→ **RNA**→ ←COG3314 (COG3314)? (GOS4580)
env-25	JCVI_SCAF_1096627388963	-	175755	175695	? (GOS12136901)→ **RNA**→ COG1741 (COG1741)? (GOS5791)→
env-26	JCVI_SCAF_1101668744010	-	402	342	? (GOS418)→ **RNA**→
env-27	JCVI_SCAF_1096627388998	-	175188	175128	? (GOS418)→ **RNA**→ ←MTH1175 (cd00851)? (GOS11736581)
Ssp-2-6	NC_008321.1	+	865136	865248	chromosome segregation ATPase→ **RNA**→ COG1741 (COG1741)→
Son-1-6	NC_004347.1	+	923243	923303	hypo→ **RNA**→ COG1741 (COG1741)→
env-29	JCVI_SCAF_1096627388993	-	179856	179743	COG4928 (COG4928)? (GOS11137131)→ **RNA**→ ←ProA (COG0014)? (GOS530)
env-30	JCVI_SCAF_1096627195142	+	9098	9158	? (GOS11452331)→ **RNA**→ ? (GOS11512624)→
env-31	JCVI_SCAF_1096627388986	-	17055	16994	←? (GOS14027118) **RNA**→ ←? (GOS13449409)
env-32	JCVI_SCAF_1096627388993	+	181868	181929	←COG4928 (COG4928)? (GOS11137131) **RNA**→ ←COG2856 (COG2856)HTH_XRE (smart00530)? (GOS1328)
Son-1-7	NC_004347.1	+	2135235	2135348	INT_IntI (cd01193)→ **RNA**→ ←EAL (cd01948)
Spu-2-7	NC_009438.1	+	159326	159387	hypo→ **RNA**→ ←hypo
Spu-1-7	NZ_AAWY01000007.1	-	24167	24106	hypo (cons)→ **RNA**→ ←leucine-rich repeat
Sba-3-7	NC_009997.1	+	826487	826548	HipA (COG3550)→ **RNA**→ hypo→
Sba-3-8	NC_009997.1	+	823256	823317	←putative esterase **RNA**→ HTH_XRE (cd00093)→
Sba-2-6	NC_009665.1	+	790094	790155	HEAT repeat-containing PBS lyase→ **RNA**→ HTH_XRE (cd00093)→
Ssp-1-5	NC_008577.1	-	2692065	2691952	←hypo **RNA**→ ←EAL (cd01948)
Ssp-1-6	NC_008577.1	+	4061579	4061640	←hypo **RNA**→ ←COG3529 (COG3529)
env-33	JCVI_SCAF_1096627390099	-	64909	64848	? (GOS12125594)→ **RNA**→ ←EAL (cd01948)? (GOS2819)
Ssp-4-6	NC_008750.1	-	4046106	4046045	hypo→ **RNA**→ hypo→
Sba-4-6	NZ_ABAU01000021.1	+	30672	30733	←ThrS (COG0441) **RNA**→ ←GcvP (COG0403)GcvP (COG1003)
Sba-1-8	NC_009052.1	-	707263	707202	AcrR (COG1309)→ **RNA**→ ←GcvP (COG0403)GcvP (COG1003)
Son-1-8	NC_004347.1	-	3779844	3779783	←hypo **RNA**→ ←hypo
env-34	JCVI_SCAF_1096627388971	+	117203	117264	COG2968 (COG2968)? (GOS7701)→ **RNA**→ ←? (GOS11679563)
Sba-3-9	NC_009997.1	+	1878992	1879053	←hypo **RNA**→ ←DUF1415 (pfam07209)
Spu-2-8	NC_009438.1	+	4063708	4063821	hypo→ **RNA**→ ←CspR (COG0219)
Ssp-4-7	NC_008750.1	-	485219	485106	←hypo **RNA**→ ←CspR (COG0219)
Spu-1-8	NZ_AAWY01000028.1	-	8977	8917	←hypo **RNA**→ ←CspR (COG0219)
Sba-2-7	NC_009665.1	+	165983	166044	glycosyl transferase family protein→ **RNA**→ ←putative lipoprotein
env-35	JCVI_SCAF_1096627389620	-	111480	111367	←? (GOS12137839) **RNA**→ PerM (COG0628)? (GOS5032)→
Ssp-2-7	NC_008321.1	-	1689515	1689402	hypo→ **RNA**→ ←Tar (COG0840)
Ssp-1-7	NC_008577.1	-	4728916	4728803	hypo→ **RNA**→ ←CspR (COG0219)
env-36	JCVI_SCAF_1101667373750	-	882	769	**RNA**→ ←CspR (COG0219)? (GOS493)
env-37	JCVI_SCAF_1101667400326	-	266	205	**RNA**→
env-38	JCVI_SCAF_1096627390098	-	45674	45561	←? (GOS418) **RNA**→ ←EAL (cd01948)? (GOS2819)
Ssp-2-8	NC_008321.1	+	2617054	2617115	←hypo **RNA**→ ←LMWPc (cd00115)
Ssp-3-6	NC_008322.1	+	2844817	2844930	←hypo **RNA**→ ←hypo
Ssp-3-7	NC_008322.1	+	2630493	2630554	←hypo **RNA**→ ←hypo
Ssp-1-8	NC_008577.1	-	3255357	3255296	←RimI (COG0456) **RNA**→ ←DUF6 (pfam00892)
env-39	JCVI_SCAF_1096627390075	-	63159	63098	←? (GOS14094631) **RNA**→ OmpA (COG2885)? (GOS11593935)→
Ssp-1-9	NC_008577.1	-	734223	734162	←SIR2-like (cd01406) **RNA**→ ←hypo
Ssp-2-9	NC_008321.1	+	2562107	2562168	←hypo **RNA**→ ←hypo
Ssp-2-10	NC_008321.1	-	2742481	2742420	COG3791 (COG3791)→ **RNA**→ ←G_glu_transpept (pfam01019)
Ssp-3-8	NC_008322.1	-	2809537	2809424	hypo→ **RNA**→ ←G_glu_transpept (pfam01019)
Ssp-2-11	NC_008321.1	-	3612319	3612258	←hypo **RNA**→ ←TPR repeat-containing protein

Ssp-3-9	NC_008322.1	-	3683793	3683732	←hypo **RNA**→ ←TPR repeat-containing protein
Ssp-2-12	NC_008321.1	+	2746294	2746407	←hypo **RNA**→ ←ArgE (COG0624)Peptidase_M20 (pfam01546)
Ssp-2-13	NC_008321.1	-	792103	792042	DUF469 (pfam04320)→ **RNA**→ ←MurA (COG0766)
env-40	JCVI_SCAF_1101667406492	+	78	139	**RNA**→ ←CspR (COG0219)? (GOS493)
Sba-2-8	NC_009665.1	-	1186168	1186109	COG4538 (COG4538)→ **RNA**→ ←DUF805 (pfam05656)
Ssp-4-8	NC_008750.1	-	1081336	1081276	hypo→ **RNA**→ MurI (COG0796)→
Spu-1-9	NZ_AAWY01000008.1	+	117088	117148	hypo→ **RNA**→ MurI (COG0796)→
Spu-2-9	NC_009438.1	+	3477323	3477383	hypo→ **RNA**→ MurI (COG0796)→
Sba-2-9	NC_009665.1	-	4215414	4215354	←hypo **RNA**→ ←MetK (COG0192)
Spu-2-10	NC_009438.1	-	4204629	4204568	hypo→ **RNA**→ COG4395 (COG4395)→
Spu-1-10	NZ_AAWY01000009.1	-	89170	89109	hypo (cons)→ **RNA**→ COG4395 (COG4395)→
env-41	JCVI_SCAF_1096627205315	+	46954	47015	? (GOS13258397)→ **RNA**→ Tim44 (pfam04280)? (GOS1170)→
Sba-2-10	NC_009665.1	+	605388	605449	←hypo **RNA**→ ←SAICAR_synt_Sc (cd01414)
Sba-3-10	NC_009997.1	-	5240430	5240369	SH3b (smart00287)→ **RNA**→ ←hypo
Spu-1-11	NZ_AAWY01000001.1	-	239916	239855	hypo→ **RNA**→ UvrD (COG0210)→
Sse-1-10	NC_009831.1	-	4566273	4566212	hypo→ **RNA**→ hypo→

27.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 27.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Desulfotalea-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

Dd cd00093 (3) Helix-turn-helix XRE-family like proteins.
d cd00115 (2) Substituted updates: Aug 22, 2001
u cd00130 (1) PAS domain; PAS motifs appear in archaea, eubacteria and eukarya.
u cd00180 (1) Serine/Threonine protein kinases, catalytic domain.
u cd00204 (1) ankyrin repeats; ankyrin repeats mediate protein-protein interactions in very diverse families of proteins.
d cd00354 (1) Fructose-1,6-bisphosphatase, an enzyme that catalyzes the hydrolysis of fructose-1,6-biphosphate into fructose-6-phosphate and is critical in gluconeogenesis pathway.
d cd00484 (1) Phosphoenolpyruvate carboxykinase (PEPCK), a critical gluconeogenic enzyme, catalyzes the first committed step in the diversion of tricarboxylic acid cycle intermediates toward gluconeogenesis.
u cd00800 (1) Lambda integrase, C-terminal catalytic domain.
d cd00851 (1) This uncharacterized conserved protein belongs to a family of iron-molybdenum cluster-binding proteins that includes NifX, NifB, and NifY, all of which are involved in the synthesis of an iron-molybdenum cofactor (FeMo-co) that binds the active site of the dinitrogenase enzyme.
d cd01012 (1) YcaC related amidohydrolases; E.
uD cd01193 (2) IntI (E2) integrases, site-specific tyrosine recombinases, DNA breaking-rejoining enzymes, N- and C-terminal domains.
u cd01296 (1) Imidazolonepropionase/imidazolone-5-propionate hydrolase (Imidazolone-5PH) catalyzes the third step in the histidine degradation pathway, the hydrolysis of (S)-3-(5-oxo-4,5-dihydro-3H-imidazol-4-yl)propanoate to N-formimidoyl-L-glutamate.
u cd01299 (1) Metallo-dependent hydrolases, subgroup A is part of the superfamily of metallo-dependent hydrolases, a large group of proteins that show conservation in their 3-dimensional fold (TIM barrel) and in details of their active site.
d cd01301 (1) renal dipeptidase (rDP), best studied in mammals and also called membrane or microsomal dipeptidase, is a membrane-bound glycoprotein hydrolyzing dipeptides and is involved in hydrolytic metabolism of penem and carbapenem beta-lactam antibiotics.
u cd01406 (1) Sir2-like: Prokaryotic group of uncharacterized Sir2-like proteins which lack certain key catalytic residues and conserved zinc binding cysteines; and are members of the SIR2 superfamily of proteins, silent information regulator 2 (Sir2) enzymes which catalyze NAD+-dependent protein/histone deacetylation.
d cd01414 (1) Eukaryotic, prokaryotic and archaeal group of SAICAR synthetases represented by the Saccharomyces cerevisiae (Sc) SAICAR synthetase.
u cd01424 (1) Methylglyoxal synthase-like domain from type II glutamine-dependent carbamoyl phosphate synthetase (CSP).
d cd01942 (1) Ribokinase-like subgroup A.
d cd01948 (4) EAL domain.
u d cd01949 (3) Diguanylate-cyclase (DGC) or GGDEF domain: Originally named after a conserved residue pattern, and initially described as domain of unknown function 1 (DUF1).
d cd03048 (1) GST_N family, Ure2p-like subfamily; composed of the Saccharomyces cerevisiae Ure2p and related GSTs. [Ure3], an inactive, self-propagating and infectious amyloid. Ure2p displays a GST fold containing an N-terminal TRX-fold domain and a C-terminal alpha helical domain, with an active site located in a cleft between the two domains. The N-terminal TRX-fold domain is sufficient to induce the [Ure3]
d cd03178 (1) GST_C family, Ure2p-like subfamily; composed of the Saccharomyces cerevisiae Ure2p and related GSTs. [Ure3], an inactive, self-propagating and infectious amyloid. Ure2p displays a GST fold containing an N-terminal thioredoxin-fold domain and a C-terminal alpha

helical domain. The N-terminal thioredoxin-fold domain is sufficient to induce the [Ure3]
D COG0010 (1) Arginase/agmatinase/formimionoglutamate hydrolase, arginase family [Amino acid transport and metabolism]
d COG0014 (1) Gamma-glutamyl phosphate reductase [Amino acid transport and metabolism]
d COG0033 (1) Phosphoglucomutase [Carbohydrate transport and metabolism]
d COG0168 (1) Trk-type K+ transport systems, membrane components [Inorganic ion transport and metabolism]
d COG0192 (1) S-adenosylmethionine synthetase [Coenzyme metabolism]
D COG0210 (1) Superfamily I DNA and RNA helicases [DNA replication, recombination, and repair]
d COG0219 (6) Predicted rRNA methylase (SpoU class) [Translation, ribosomal structure and biogenesis]
u COG0265 (2) Trypsin-like serine proteases, typically periplasmic, contain C-terminal PDZ domain [Posttranslational modification, protein turnover, chaperones]
u COG0350 (3) Methylated DNA-protein cysteine methyltransferase [DNA replication, recombination, and repair]
d COG0403 (2) Glycine cleavage system protein P (pyridoxal-binding), N-terminal domain [Amino acid transport and metabolism]
d COG0406 (1) Fructose-2,6-bisphosphatase [Carbohydrate transport and metabolism]
d COG0412 (1) Dienelactone hydrolase and related enzymes [Secondary metabolites biosynthesis, transport, and catabolism]
u COG0441 (1) Threonyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
D COG0442 (1) Prolyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
u COG0456 (1) Acetyltransferases [General function prediction only]
u COG0458 (1) Carbamoylphosphate synthase large subunit (split gene in MJ) [Amino acid transport and metabolism / Nucleotide transport and metabolism]
D COG0497 (1) ATPase involved in DNA repair [DNA replication, recombination, and repair]
D COG0520 (1) Selenocysteine lyase [Amino acid transport and metabolism]
D COG0543 (1) 2-polyprenylphenol hydroxylase and related flavodoxin oxidoreductases [Coenzyme metabolism / Energy production and conversion]
d COG0624 (1) Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase and related deacylases [Amino acid transport and metabolism]
D COG0628 (1) Predicted permease [General function prediction only]
d COG0662 (1) Mannose-6-phosphate isomerase [Carbohydrate transport and metabolism]
d COG0665 (1) Glycine/D-amino acid oxidases (deaminating) [Amino acid transport and metabolism]
d COG0708 (3) Exonuclease III [DNA replication, recombination, and repair]
u COG0715 (1) ABC-type nitrate/sulfonate/bicarbonate transport systems, periplasmic components [Inorganic ion transport and metabolism]
d COG0766 (1) UDP-N-acetylglucosamine enolpyruvyl transferase [Cell envelope biogenesis, outer membrane]
D COG0775 (1) Nucleoside phosphorylase [Nucleotide transport and metabolism]
d COG0777 (1) Acetyl-CoA carboxylase beta subunit [Lipid metabolism]
u COG0782 (1) Transcription elongation factor [Transcription]
D COG0796 (3) Glutamate racemase [Cell envelope biogenesis, outer membrane]
D COG0797 (1) Lipoproteins [Cell envelope biogenesis, outer membrane]
d COG0825 (1) Acetyl-CoA carboxylase alpha subunit [Lipid metabolism]
u d COG0840 (2) Methyl-accepting chemotaxis protein [Cell motility and secretion / Signal transduction mechanisms]
d COG1003 (2) Glycine cleavage system protein P (pyridoxal-binding), C-terminal domain [Amino acid transport and metabolism]
d COG1053 (1) Succinate dehydrogenase/fumarate reductase, flavoprotein subunit [Energy production and conversion]
u COG1061 (1) DNA or RNA helicases of superfamily II [Transcription / DNA replication, recombination, and repair]
Dd COG1115 (2) Na+/alanine symporter [Amino acid transport and metabolism]
d COG1126 (1) ABC-type polar amino acid transport system, ATPase component [Amino acid transport and metabolism]
d COG1132 (1) ABC-type multidrug transport system, ATPase and permease components [Defense mechanisms]
d COG1186 (1) Protein chain release factor B [Translation, ribosomal structure and biogenesis]
u COG1200 (1) RecG-like helicase [DNA replication, recombination, and repair / Transcription]
u COG1225 (1) Peroxiredoxin [Posttranslational modification, protein turnover, chaperones]
u COG1309 (1) Transcriptional regulator [Transcription]
D COG1315 (2) Predicted polymerase, most proteins contain PALM domain, HD hydrolase domain and Zn-ribbon domain [DNA replication, recombination, and repair]
u COG1359 (1) Uncharacterized conserved protein [Function unknown]
u COG1398 (1) Fatty-acid desaturase [Lipid metabolism]
u COG1473 (1) Metal-dependent amidase/aminoacylase/carboxypeptidase [General function prediction only]
d COG1506 (1) Dipeptidyl aminopeptidases/acylaminoacyl-peptidases [Amino acid transport and metabolism]
u COG1556 (4) Uncharacterized conserved protein [Function unknown]
d COG1718 (1) Serine/threonine protein kinase involved in cell cycle control [Signal transduction mechanisms / Cell division and chromosome partitioning]
d COG1733 (1) Predicted transcriptional regulators [Transcription]
D COG1741 (8) Pirin-related protein [General function prediction only]
D COG1744 (1) Uncharacterized ABC-type transport system, periplasmic component/surface lipoprotein [General function prediction only]
D COG1802 (1) Transcriptional regulators [Transcription]
d COG1853 (1) Conserved protein/domain typically associated with flavoprotein oxygenases, DIM6/NTAB family [General function prediction only]
u COG1879 (1) ABC-type sugar transport system, periplasmic component [Carbohydrate transport and metabolism]
d COG1910 (1) Periplasmic molybdate-binding protein/domain [Inorganic ion transport and metabolism]
u COG2137 (1) Uncharacterized protein conserved in bacteria [General function prediction only]
d COG2197 (1) Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain [Signal transduction mechanisms / Transcription]
uD COG2226 (2) Methylase involved in ubiquinone/menaquinone biosynthesis [Coenzyme metabolism]
u d COG2227 (2) 2-polyprenyl-3-methyl-5-hydroxy-6-metoxy-1,4-benzoquinol methylase [Coenzyme metabolism]
u COG2249 (1) Putative NADPH-quinone reductase (modulator of drug activity B) [General function prediction only]
d COG2253 (1) Uncharacterized conserved protein [Function unknown]
d COG2819 (1) Predicted hydrolase of the alpha/beta superfamily [General function prediction

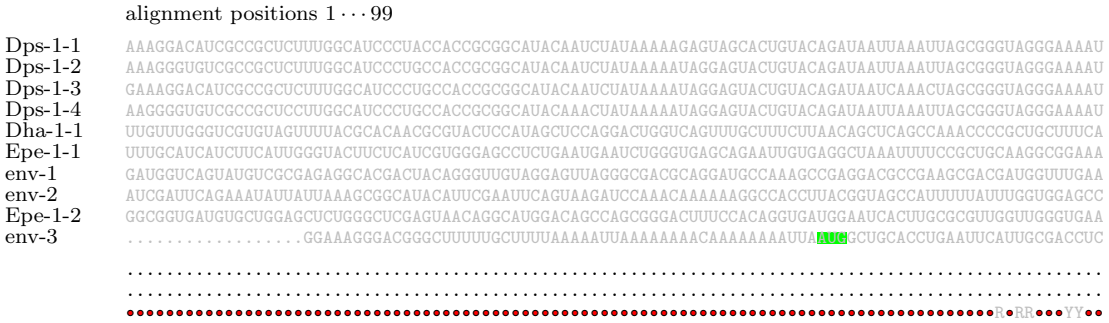
only]
d COG2856 (1) Predicted Zn peptidase [Amino acid transport and metabolism]
d COG2863 (1) Cytochrome c553 [Energy production and conversion]
D COG2885 (2) Outer membrane protein and related peptidoglycan-associated (lipo)proteins [Cell envelope biogenesis, outer membrane]
u COG2963 (1) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
u COG2968 (2) Uncharacterized conserved protein [Function unknown]
D COG3046 (2) Uncharacterized protein related to deoxyribodipyrimidine photolyase [General function prediction only]
d COG3150 (3) Predicted esterase [General function prediction only]
d COG3258 (1) Cytochrome c [Energy production and conversion]
d COG3314 (3) Uncharacterized protein conserved in bacteria [Function unknown]
u COG3344 (1) Retron-type reverse transcriptase [DNA replication, recombination, and repair]
d COG3529 (1) Predicted nucleic-acid-binding protein containing a Zn-ribbon domain [General function prediction only]
uD COG3547 (5) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
u COG3550 (1) Uncharacterized protein related to capsule biosynthesis enzymes [General function prediction only]
u COG3593 (1) Predicted ATP-dependent endonuclease of the OLD family [DNA replication, recombination, and repair]
D COG3673 (2) Uncharacterized conserved protein [Function unknown]
d COG3707 (1) Response regulator with putative antiterminator output domain [Signal transduction mechanisms]
u d COG3791 (2) Uncharacterized conserved protein [Function unknown]
u COG3852 (1) Signal transduction histidine kinase, nitrogen specific [Signal transduction mechanisms]
u COG3886 (1) Predicted HKD family nuclease [DNA replication, recombination, and repair]
d COG4146 (1) Predicted symporter [General function prediction only]
u COG4271 (1) Predicted nucleotide-binding protein containing TIR -like domain [Transcription]
D COG4395 (2) Uncharacterized protein conserved in bacteria [Function unknown]
u COG4538 (1) Uncharacterized conserved protein [Function unknown]
D COG4584 (1) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
D COG4711 (1) Predicted membrane protein [Function unknown]
u COG4928 (2) Predicted P-loop ATPase [General function prediction only]
u COG5293 (1) Uncharacterized protein conserved in bacteria [Function unknown]
u COG5635 (1) Predicted NTPase (NACHT family) [Signal transduction mechanisms]
u GOS11052049 (1) ?
u GOS11137131 (2) ?
u GOS11166260 (1) ?
u GOS11452331 (1) ?
D GOS11512624 (1) ?
u GOS11522405 (1) ?
u GOS11588659 (1) ?
D GOS11593935 (1) ?
d GOS11627891 (2) ?
d GOS11679563 (1) ?
D GOS1170 (1) ?
d GOS11736581 (1) ?
D GOS12025990 (1) ?
u GOS12125594 (1) ?
u GOS12136901 (1) ?
u GOS12137839 (1) ?
u GOS13258397 (1) ?
d GOS1328 (1) ?
d GOS13449409 (1) ?
d GOS13516489 (1) ?
u GOS13550016 (1) ?
u GOS14027118 (1) ?
u GOS14094631 (1) ?
u GOS14131473 (1) ?
u GOS14347974 (1) ?
uD GOS14479710 (2) ?
u GOS1694 (1) ?
d GOS2568 (1) ?
d GOS2819 (2) ?
u GOS336 (1) ?
d GOS3709 (1) ?
u GOS418 (4) ?
d GOS4580 (1) ?
d GOS4908 (1) ?
d GOS493 (2) ?
D GOS5032 (1) ?
d GOS530 (1) ?
u GOS5488 (1) ?
D GOS5791 (2) ?
d GOS6072 (1) ?
d GOS6435 (1) ?
u GOS672 (1) ?
u GOS7701 (1) ?
u NOG07925 (2)
d NOG14035 (1)
u NOG22301 (1)
D pfam00004 (3) ATPase family associated with various cellular activities (AAA).
d pfam00120 (1) Glutamine synthetase, catalytic domain.
d pfam00154 (1) recA bacterial DNA recombination protein.
d pfam00171 (2) Aldehyde dehydrogenase family.
d pfam00293 (1) NUDIX domain.
u pfam00313 (1) 'Cold-shock' DNA-binding domain.
d pfam00561 (1) alpha/beta hydrolase fold.
u pfam00583 (1) Acetyltransferase (GNAT) family.
u pfam00665 (1) Integrase core domain.
d pfam00892 (1) Integral membrane protein DUF6.
u pfam00912 (1) Transglycosylase.
d pfam01019 (2) Gamma-glutamyltranspeptidase.
u d pfam01389 (2) OmpA-like transmembrane domain.
u pfam01425 (1) Amidase.
D pfam01435 (1) Peptidase family M48.
d pfam01546 (1) Peptidase family M20/M25/M40.

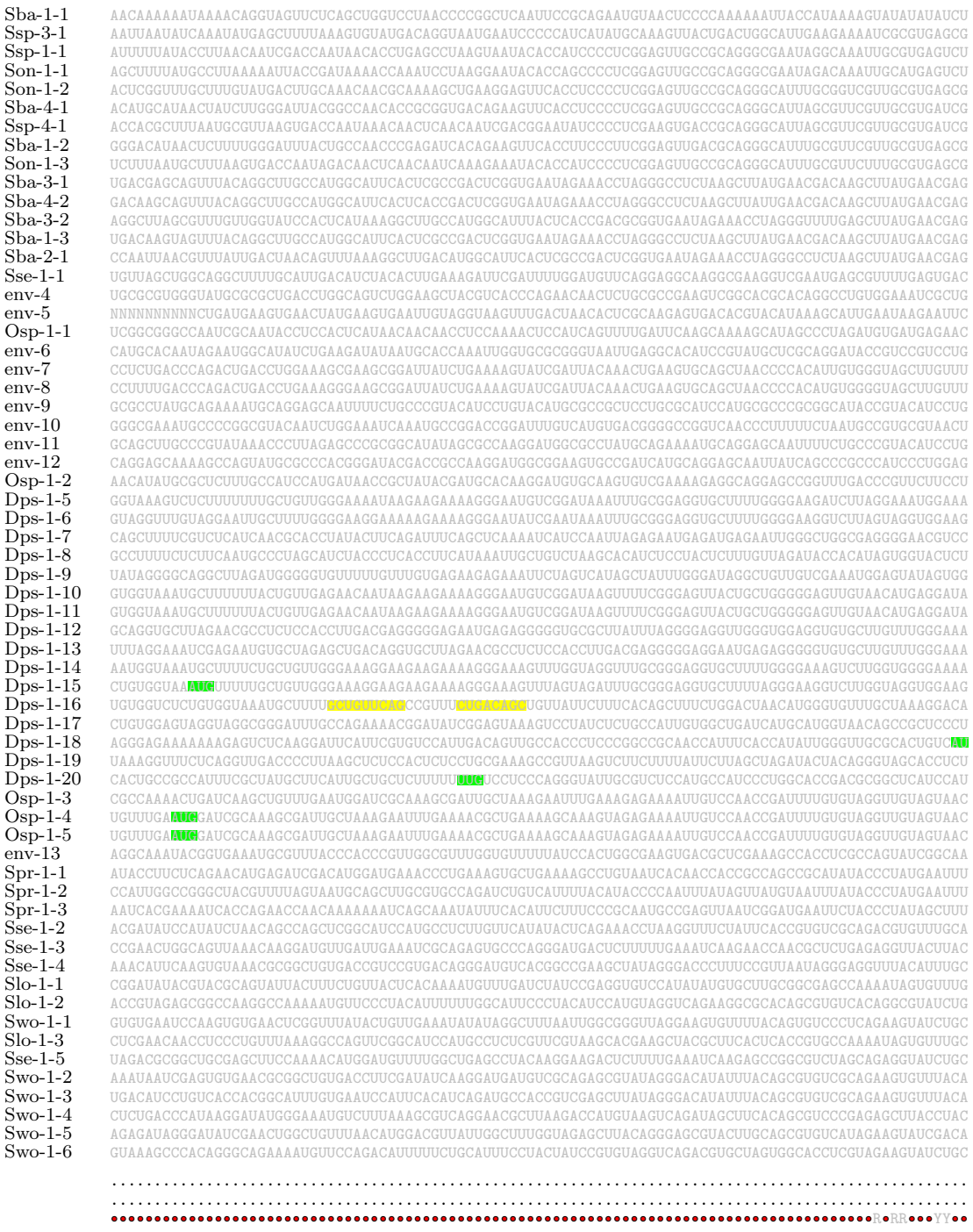
u pfam01548 (1) Transposase.
u pfam01609 (4) Transposase DDE domain.
d pfam01625 (1) Peptide methionine sulfoxide reductase.
u pfam01773 (1) Na+ dependent nucleoside transporter.
d pfam01797 (1) Transposase IS200 like.
u pfam01909 (1) Nucleotidyltransferase domain.
u d pfam02371 (2) Transposase IS116/IS110/IS902 family.
u pfam02525 (1) Flavodoxin-like fold.
u pfam02743 (1) Cache domain.
u pfam02787 (1) Carbamoyl-phosphate synthetase large chain, oligomerisation domain.
u pfam02810 (1) SEC-C motif. [HC]
d pfam03264 (1) NapC/NirT cytochrome c family, N-terminal region.
D pfam03330 (3) Rare lipoprotein A (RlpA)-like double-psi beta-barrel.
u pfam03691 (1) Uncharacterised protein family (UPF0167).
u pfam03692 (1) Uncharacterised protein family (UPF0153).
d pfam03951 (1) Glutamine synthetase, beta-Grasp domain.
D pfam04280 (1) Tim44-like domain.
u pfam04313 (2) Type I restriction enzyme R protein N terminus (HSDR_N).
u pfam04320 (1) Protein with unknown function (DUF469).
d pfam04338 (1) Protein of unknown function, DUF481.
d pfam04389 (1) Peptidase family M28.
u pfam04986 (2) Putative transposase.
u pfam05014 (1) Nucleoside 2-deoxyribosyltransferase.
d pfam05656 (1) Protein of unknown function (DUF805).
u pfam05971 (1) Protein of unknown function (DUF890).
d pfam06224 (1) Protein of unknown function (DUF1006).
d pfam07209 (1) Protein of unknown function (DUF1415).
d smart00014 (1) Acid phosphatase homologues;
u smart00287 (1) Bacterial SH3 domain homologues;
u smart00342 (3) helix_turn_helix, arabinose operon control protein;
d smart00530 (2) Helix-turn-helix XRE-family like proteins ;

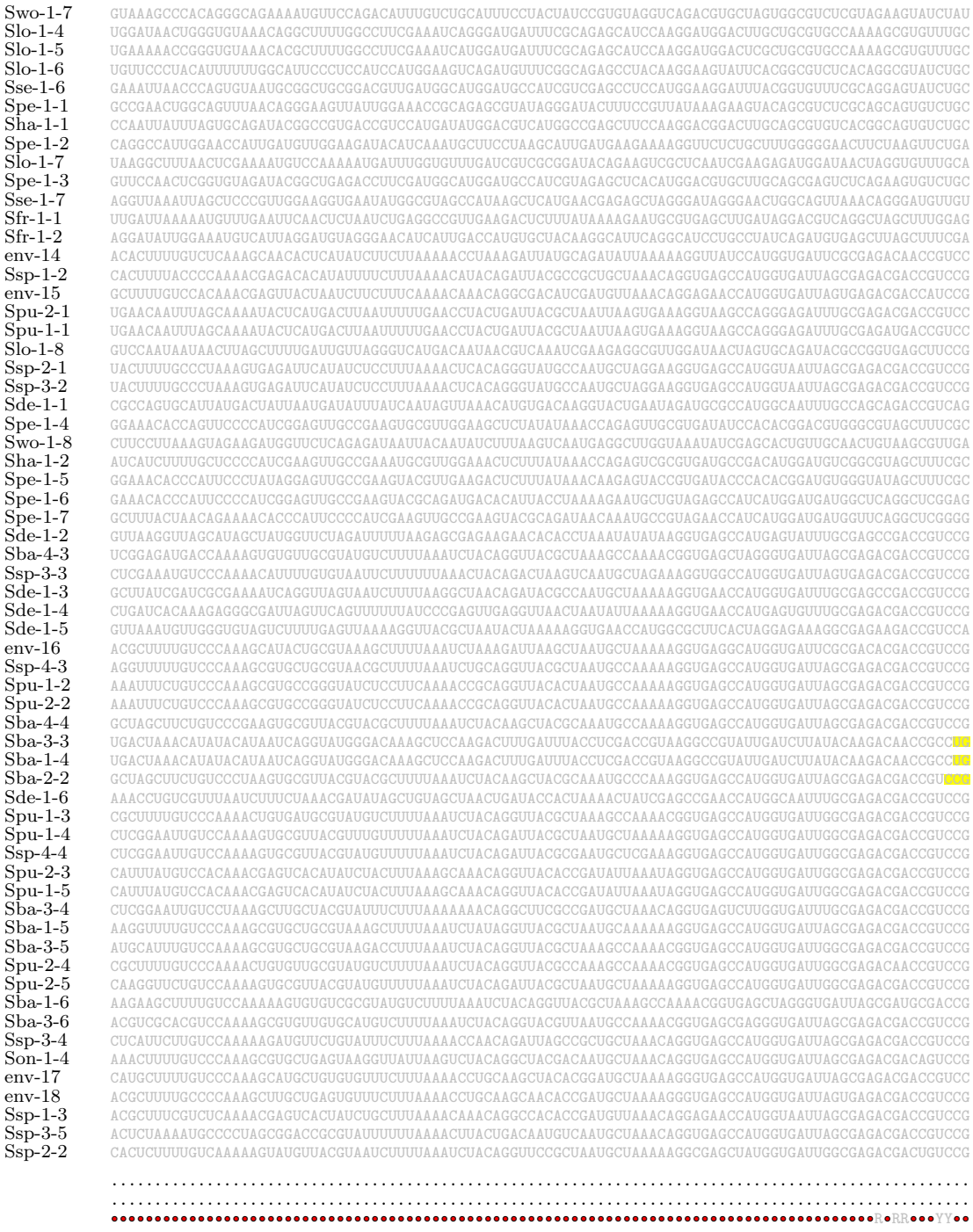
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Spu-1-10, Ssp-4-2, Ssp-4-5, Ssp-4-7

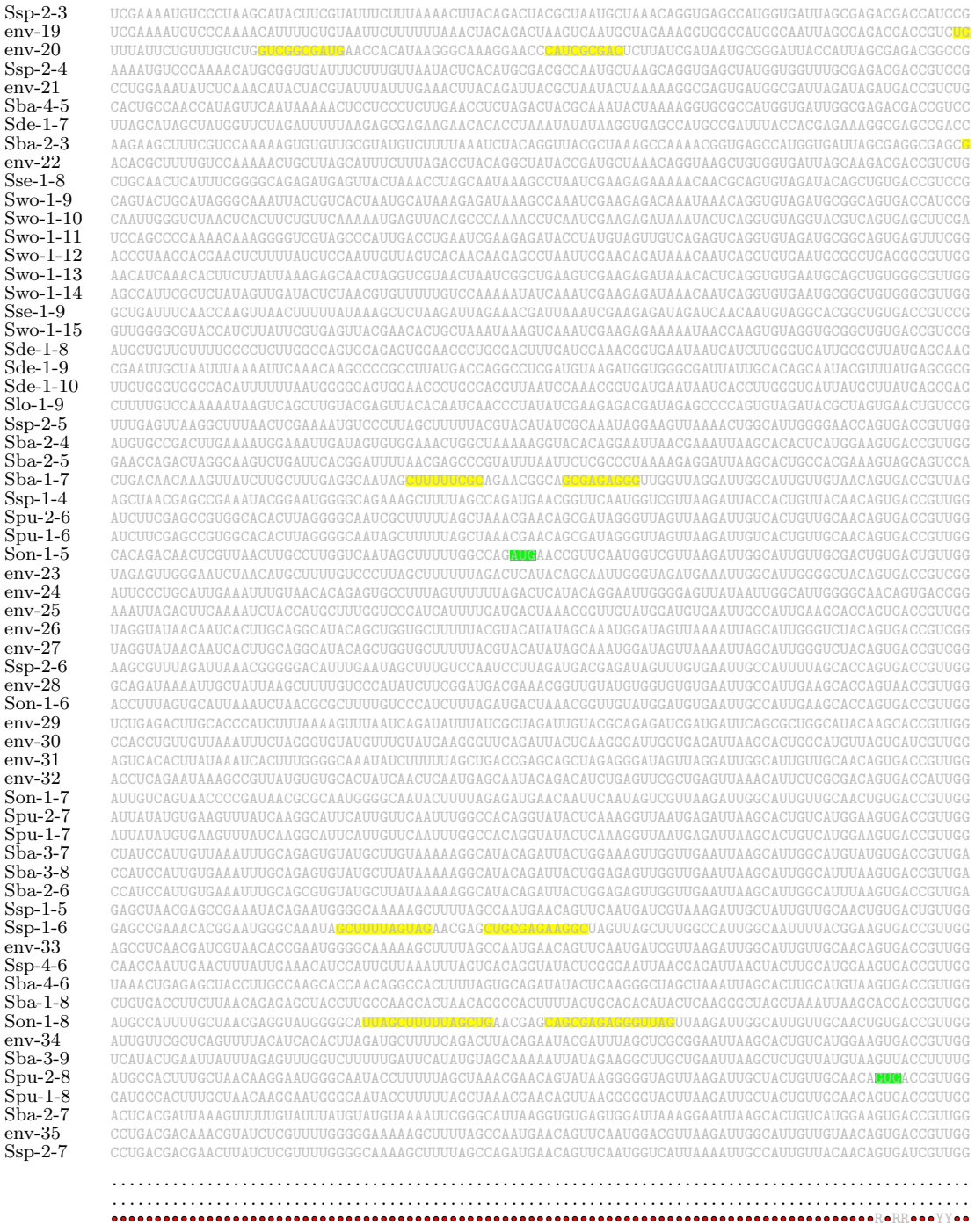
27.4 Multiple-sequence alignment

Each Desulfotalea-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 27.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Desulfotalea-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

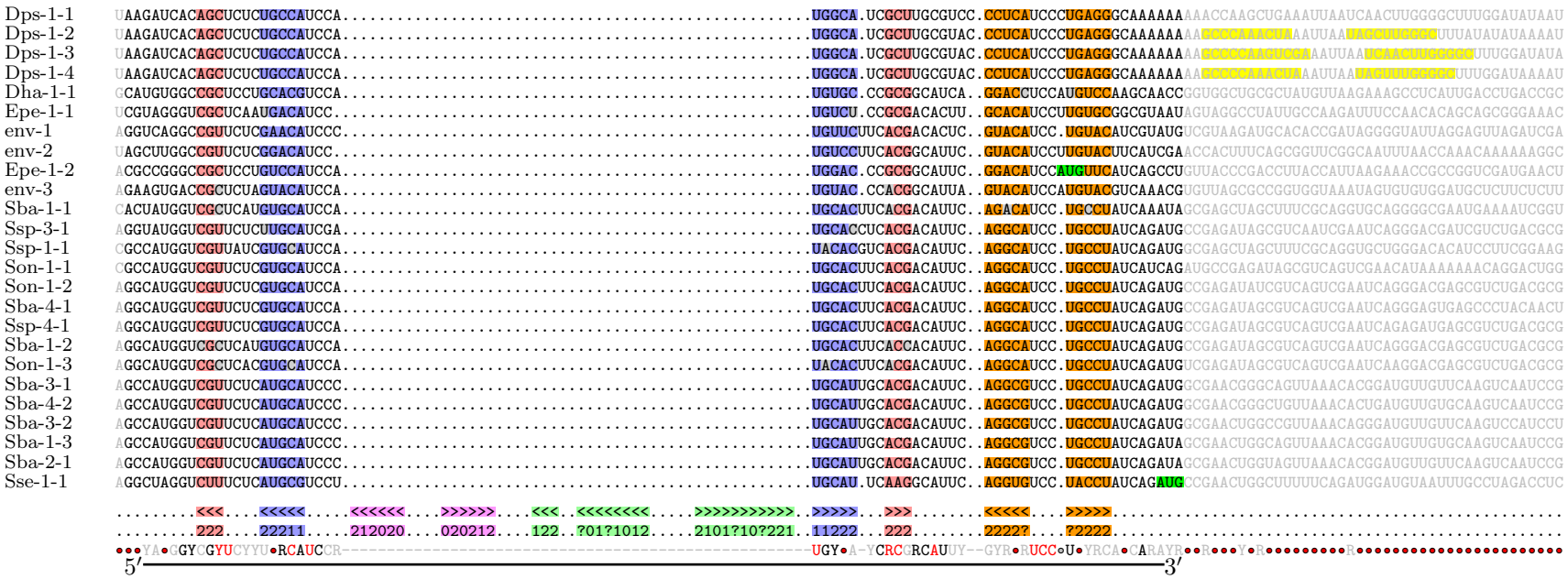


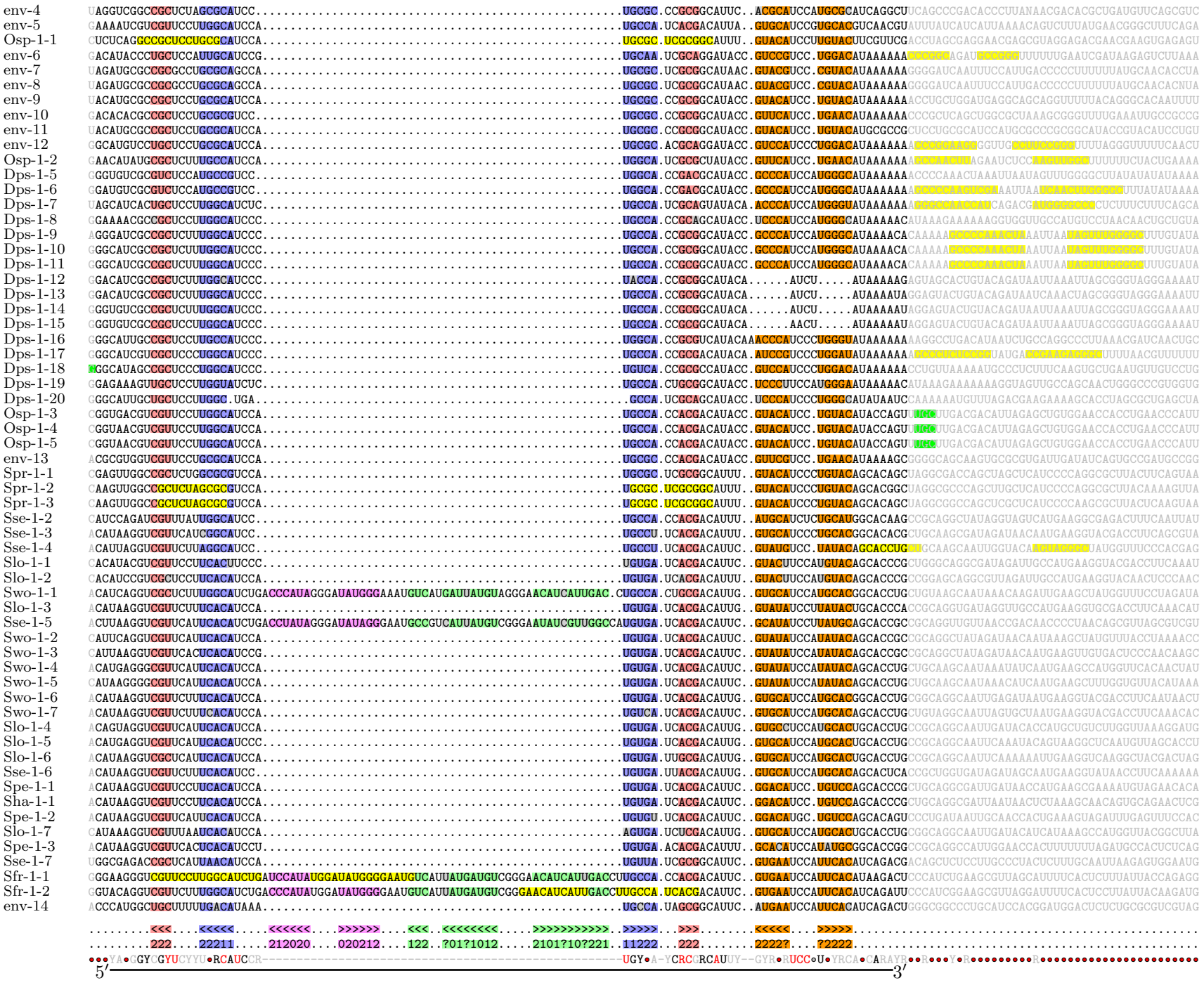

Sba-1-1
Ssp-3-1
Ssp-1-1
Son-1-1
Son-1-2
Sba-4-1
Ssp-4-1
Sba-1-2
Son-1-3
Sba-3-1
Sba-4-2
Sba-3-2
Sba-1-3
Sba-2-1
Sse-1-1
env-4
env-5
Osp-1-1
env-6
env-7
env-8
env-9
env-10
env-11
env-12
Osp-1-2
Dps-1-5
Dps-1-6
Dps-1-7
Dps-1-8
Dps-1-9
Dps-1-10
Dps-1-11
Dps-1-12
Dps-1-13
Dps-1-14
Dps-1-15
Dps-1-16
Dps-1-17
Dps-1-18
Dps-1-19
Dps-1-20
Osp-1-3
Osp-1-4
Osp-1-5
env-13
Spr-1-1
Spr-1-2
Spr-1-3
Sse-1-2
Sse-1-3
Sse-1-4
Slo-1-1
Slo-1-2
Swo-1-1
Slo-1-3
Sse-1-5
Swo-1-2
Swo-1-3
Swo-1-4
Swo-1-5
Swo-1-6

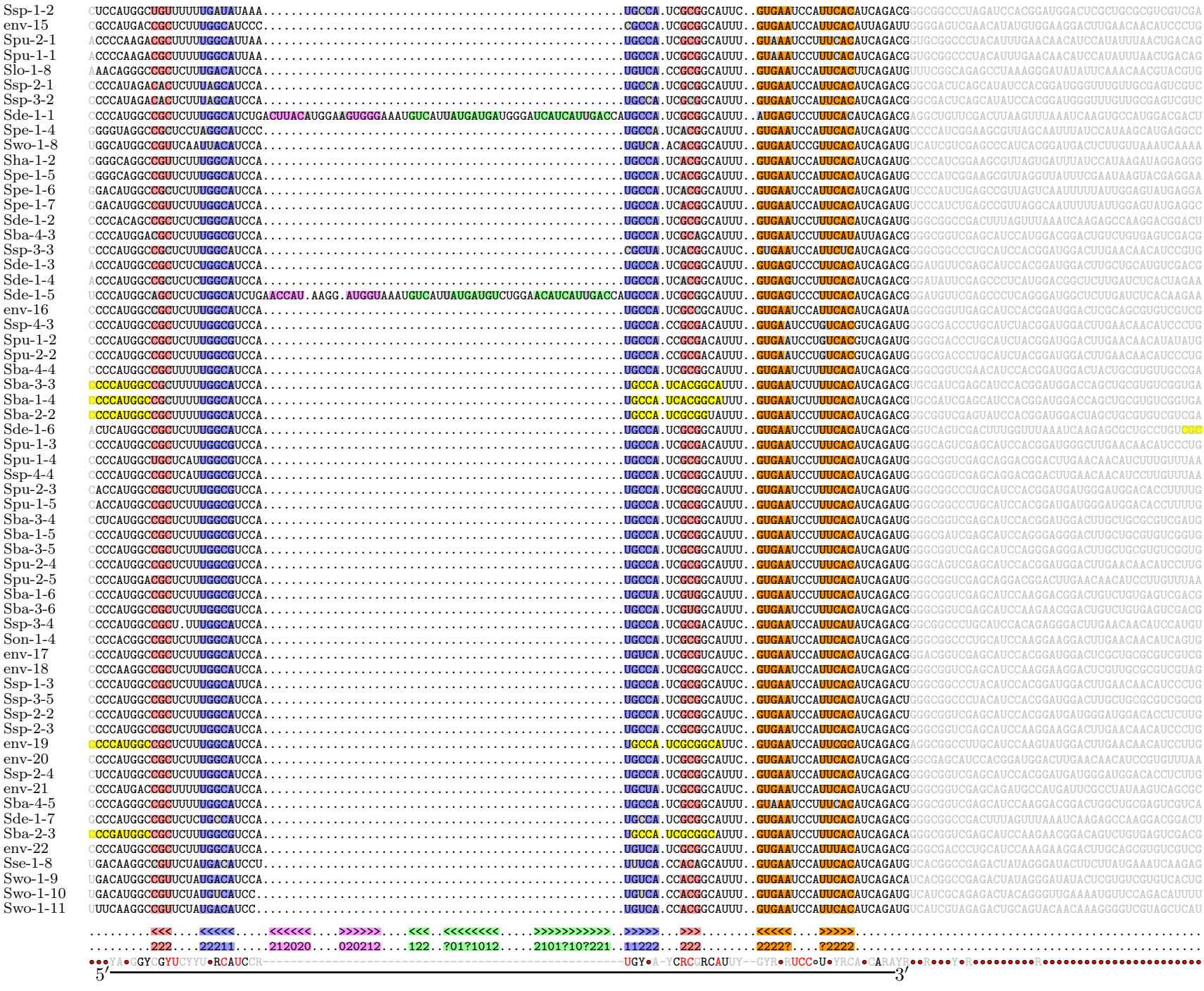

Swo-1-7
Slo-1-4
Slo-1-5
Slo-1-6
Sse-1-6
Spe-1-1
Sha-1-1
Spe-1-2
Slo-1-7
Spe-1-3
Sse-1-7
Sfr-1-1
Sfr-1-2
env-14
Ssp-1-2
env-15
Spu-2-1
Spu-1-1
Slo-1-8
Ssp-2-1
Ssp-3-2
Sde-1-1
Spe-1-4
Swo-1-8
Sha-1-2
Spe-1-5
Spe-1-6
Spe-1-7
Sde-1-2
Sba-4-3
Ssp-3-3
Sde-1-3
Sde-1-4
Sde-1-5
env-16
Ssp-4-3
Spu-1-2
Spu-2-2
Sba-4-4
Sba-3-3
Sba-1-4
Sba-2-2
Sde-1-6
Spu-1-3
Spu-1-4
Ssp-4-4
Spu-2-3
Spu-1-5
Sba-3-4
Sba-1-5
Sba-3-5
Spu-2-4
Spu-2-5
Sba-1-6
Sba-3-6
Ssp-3-4
Son-1-4
env-17
env-18
Ssp-1-3
Ssp-3-5
Ssp-2-2

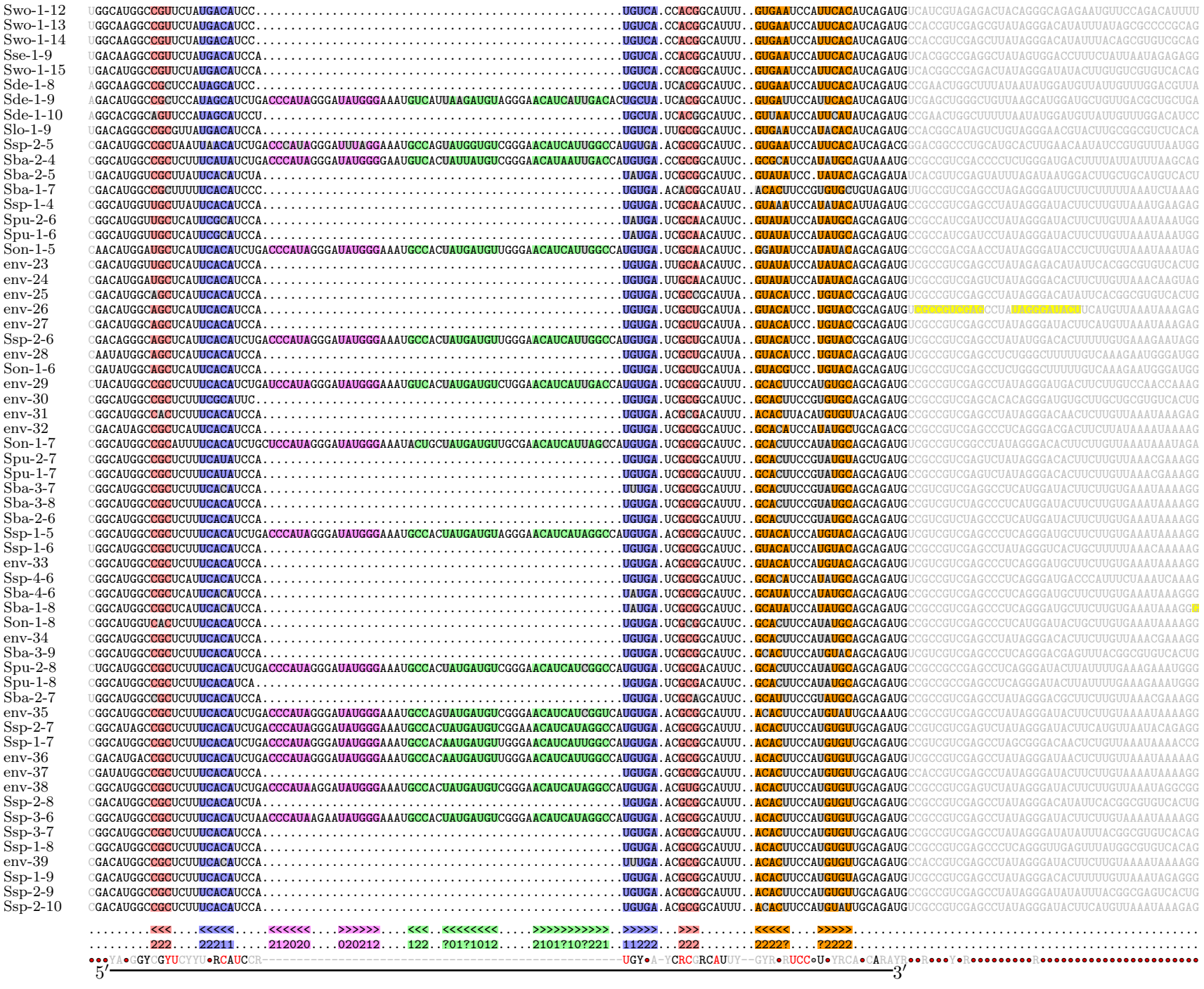

Ssp-2-3
env-19
env-20
Ssp-2-4
env-21
Sba-4-5
Sde-1-7
Sba-2-3
env-22
Sse-1-8
Swo-1-9
Swo-1-10
Swo-1-11
Swo-1-12
Swo-1-13
Swo-1-14
Sse-1-9
Swo-1-15
Sde-1-8
Sde-1-9
Sde-1-10
Slo-1-9
Ssp-2-5
Sba-2-4
Sba-2-5
Sba-1-7
Ssp-1-4
Spu-2-6
Spu-1-6
Son-1-5
env-23
env-24
env-25
env-26
env-27
Ssp-2-6
env-28
Son-1-6
env-29
env-30
env-31
env-32
Son-1-7
Spu-2-7
Spu-1-7
Sba-3-7
Sba-3-8
Sba-2-6
Ssp-1-5
Ssp-1-6
env-33
Ssp-4-6
Sba-4-6
Sba-1-8
Son-1-8
env-34
Sba-3-9
Spu-2-8
Spu-1-8
Sba-2-7
env-35
Ssp-2-7

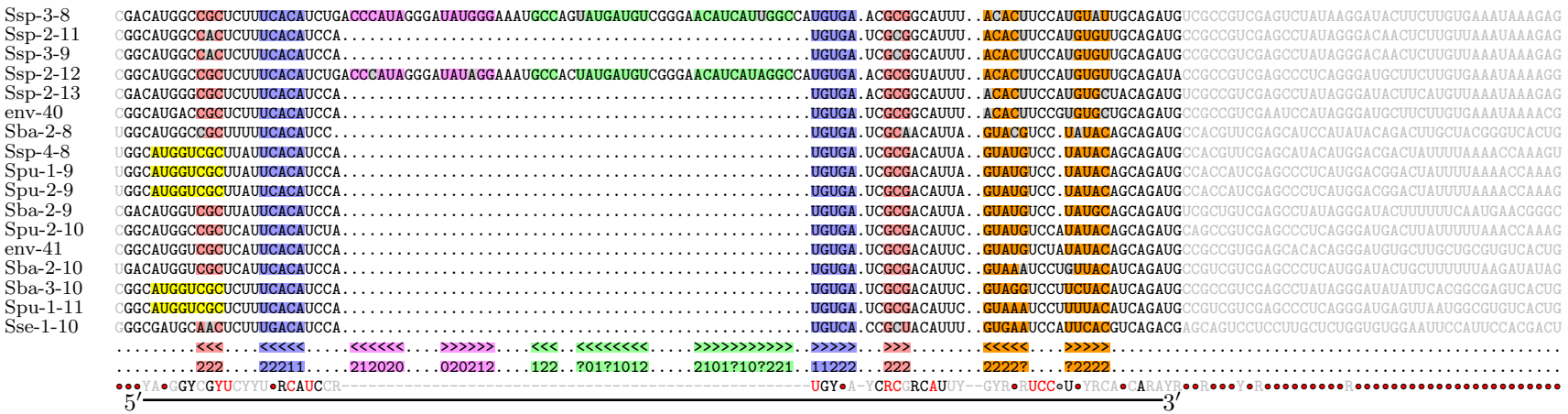
alignment positions 100 · · · 259



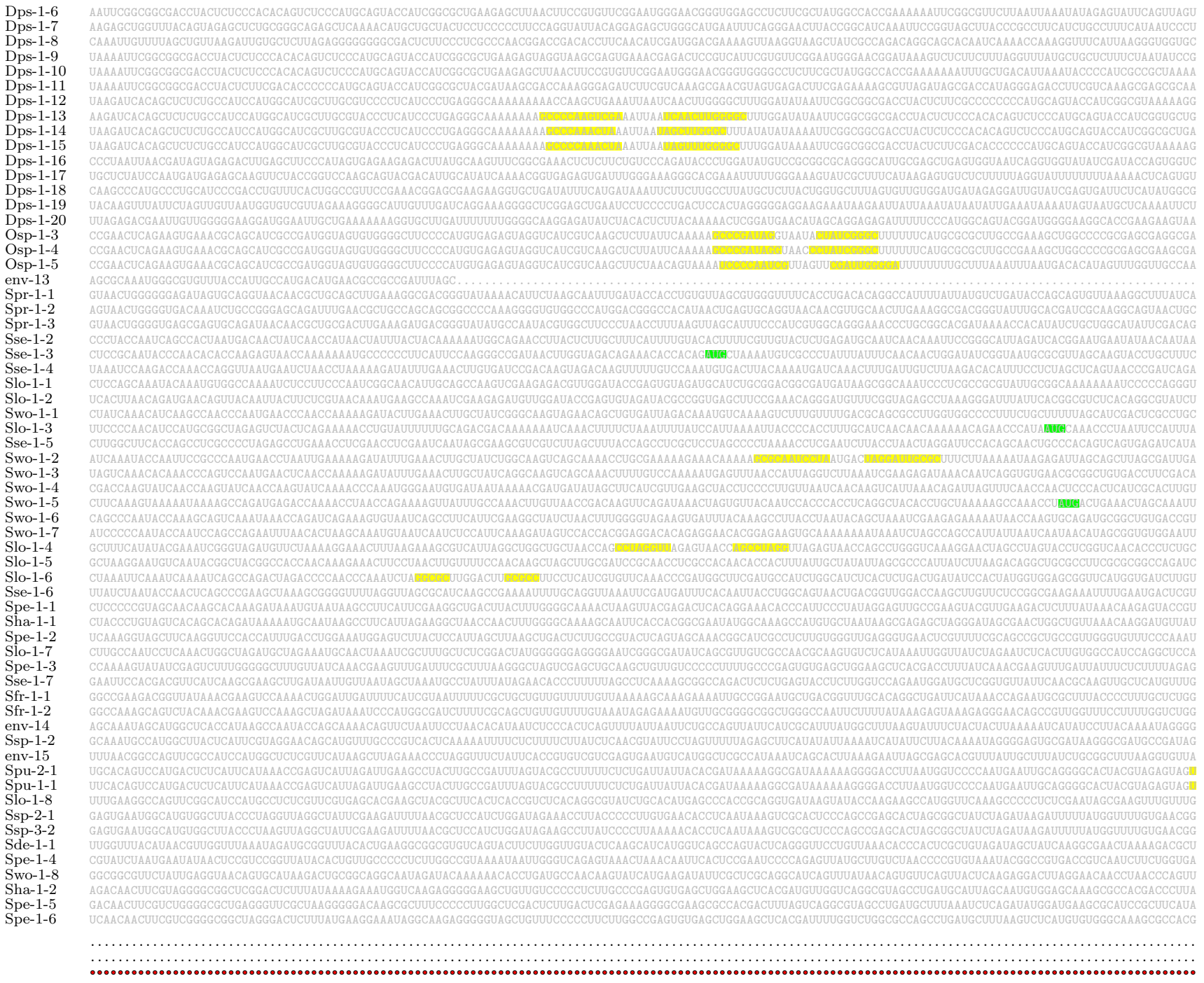

env-4
env-5
Osp-1-1
env-6
env-7
env-8
env-9
env-10
env-11
env-12
Osp-1-2
Dps-1-5
Dps-1-6
Dps-1-7
Dps-1-8
Dps-1-9
Dps-1-10
Dps-1-11
Dps-1-12
Dps-1-13
Dps-1-14
Dps-1-15
Dps-1-16
Dps-1-17
Dps-1-18
Dps-1-19
Dps-1-20
Osp-1-3
Osp-1-4
Osp-1-5
env-13
Spr-1-1
Spr-1-2
Spr-1-3
Sse-1-2
Sse-1-3
Sse-1-4
Slo-1-1
Slo-1-2
Swo-1-1
Slo-1-3
Sse-1-5
Swo-1-2
Swo-1-3
Swo-1-4
Swo-1-5
Swo-1-6
Swo-1-7
Slo-1-4
Slo-1-5
Slo-1-6
Sse-1-6
Spe-1-1
Sha-1-1
Spe-1-2
Slo-1-7
Spe-1-3
Sse-1-7
Sfr-1-1
Sfr-1-2
env-14
5′
3′

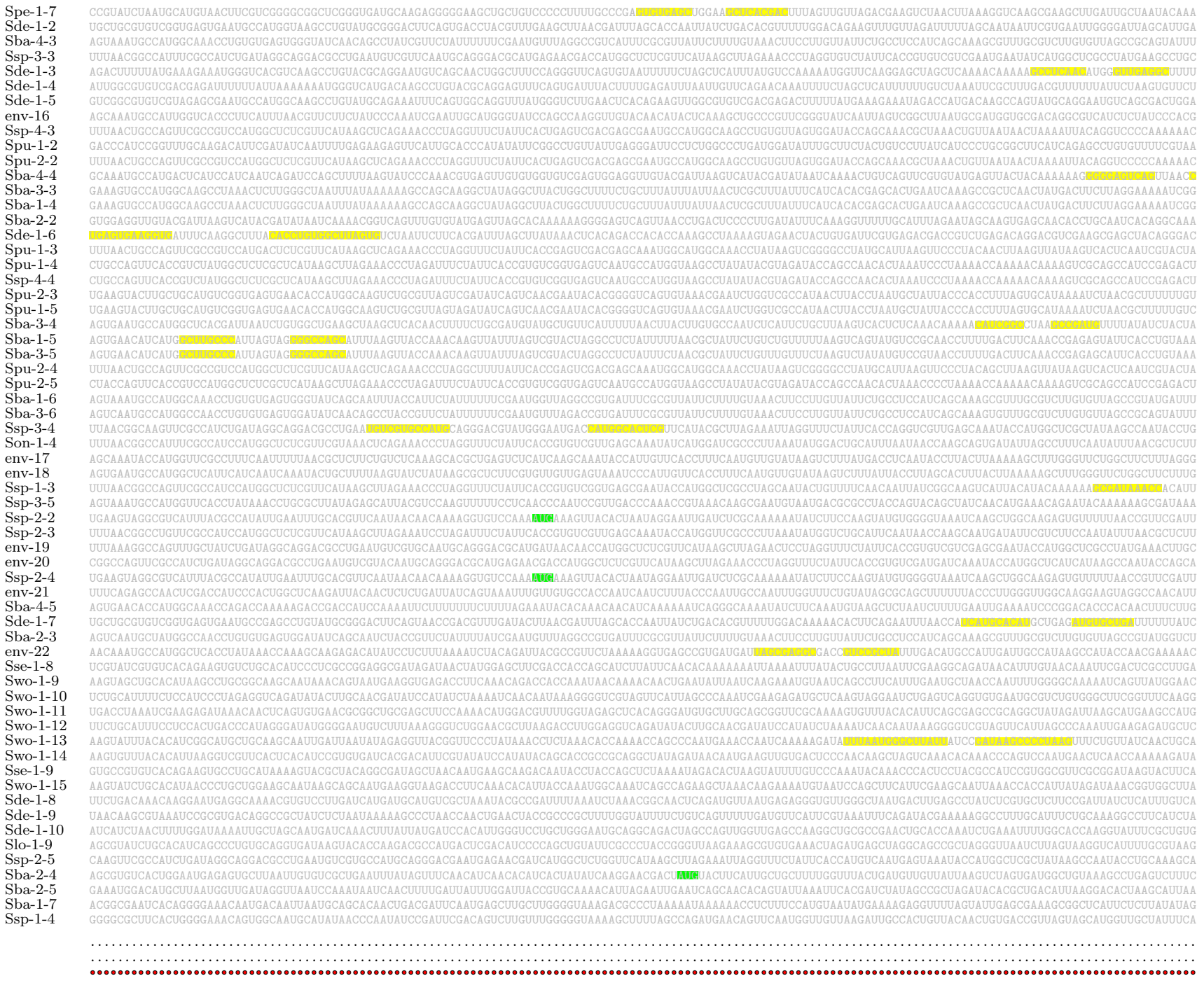

Ssp-1-2
env-15
Spu-2-1
Spu-1-1
Slo-1-8
Ssp-2-1
Ssp-3-2
Sde-1-1
Spe-1-4
Swo-1-8
Sha-1-2
Spe-1-5
Spe-1-6
Spe-1-7
Sde-1-2
Sba-4-3
Ssp-3-3
Sde-1-3
Sde-1-4
Sde-1-5
env-16
Ssp-4-3
Spu-1-2
Spu-2-2
Sba-4-4
Sba-3-3
Sba-1-4
Sba-2-2
Sde-1-6
Spu-1-3
Spu-1-4
Ssp-4-4
Spu-2-3
Spu-1-5
Sba-3-4
Sba-1-5
Sba-3-5
Spu-2-4
Spu-2-5
Sba-1-6
Sba-3-6
Ssp-3-4
Son-1-4
env-17
env-18
Ssp-1-3
Ssp-3-5
Ssp-2-2
Ssp-2-3
env-19
env-20
Ssp-2-4
env-21
Sba-4-5
Sde-1-7
Sba-2-3
env-22
Sse-1-8
Swo-1-9
Swo-1-10
Swo-1-11
5′
3′

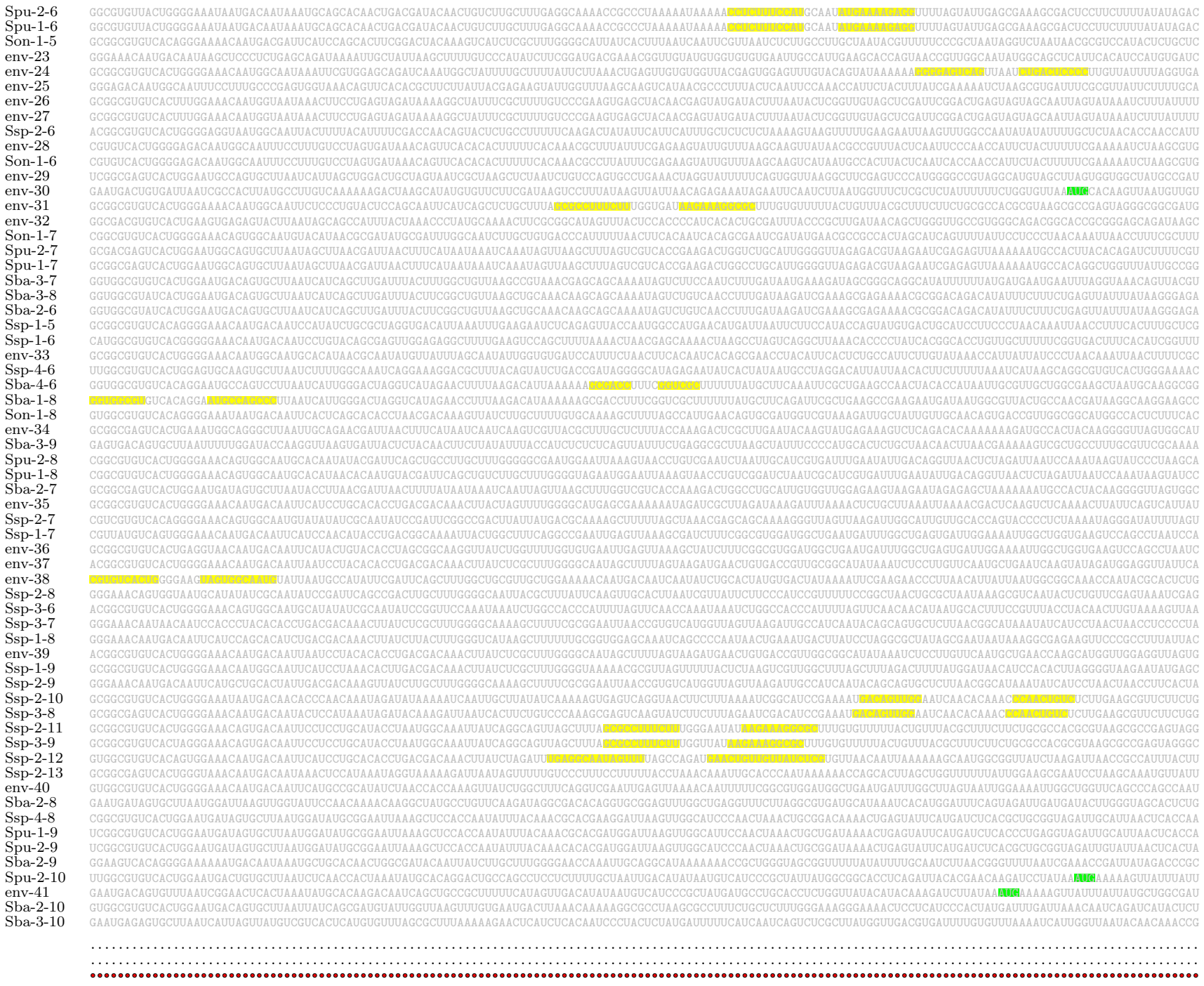

Swo-1-12
Swo-1-13
Swo-1-14
Sse-1-9
Swo-1-15
Sde-1-8
Sde-1-9
Sde-1-10
Slo-1-9
Ssp-2-5
Sba-2-4
Sba-2-5
Sba-1-7
Ssp-1-4
Spu-2-6
Spu-1-6
Son-1-5
env-23
env-24
env-25
env-26
env-27
Ssp-2-6
env-28
Son-1-6
env-29
env-30
env-31
env-32
Son-1-7
Spu-2-7
Spu-1-7
Sba-3-7
Sba-3-8
Sba-2-6
Ssp-1-5
Ssp-1-6
env-33
Ssp-4-6
Sba-4-6
Sba-1-8
Son-1-8
env-34
Sba-3-9
Spu-2-8
Spu-1-8
Sba-2-7
env-35
Ssp-2-7
Ssp-1-7
env-36
env-37
env-38
Ssp-2-8
Ssp-3-6
Ssp-3-7
Ssp-1-8
env-39
Ssp-1-9
Ssp-2-9
Ssp-2-10
5′
3′

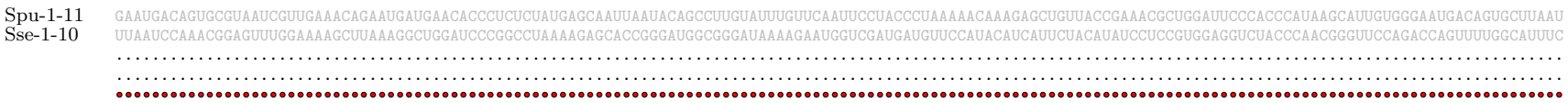


alignment positions 260 · · · 419

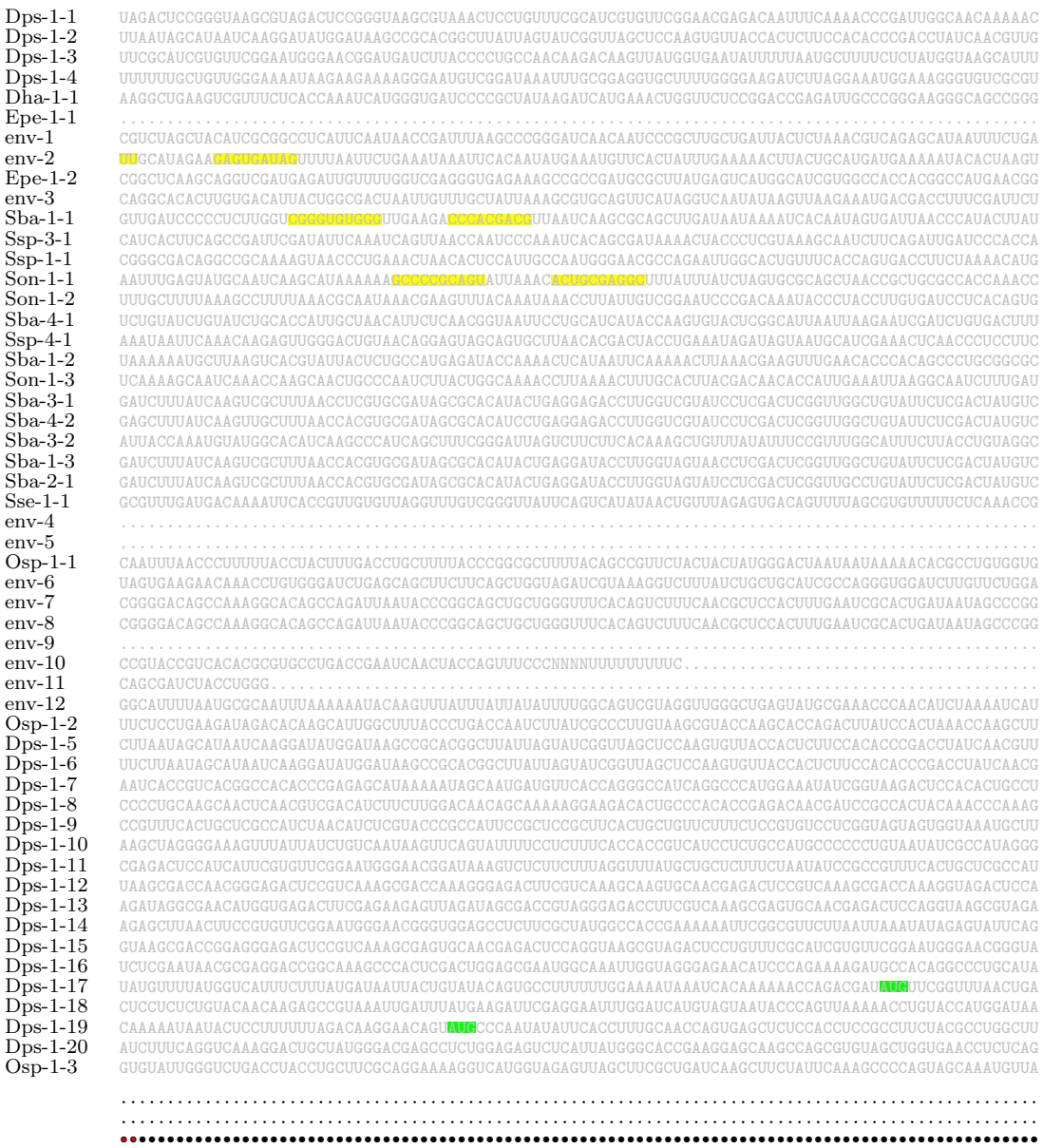

Dps-1-6
Dps-1-7
Dps-1-8
Dps-1-9
Dps-1-10
Dps-1-11
Dps-1-12
Dps-1-13
Dps-1-14
Dps-1-15
Dps-1-16
Dps-1-17
Dps-1-18
Dps-1-19
Dps-1-20
Osp-1-3
Osp-1-4
Osp-1-5
env-13
Spr-1-1
Spr-1-2
Spr-1-3
Sse-1-2
Sse-1-3
Sse-1-4
Slo-1-1
Slo-1-2
Swo-1-1
Slo-1-3
Sse-1-5
Swo-1-2
Swo-1-3
Swo-1-4
Swo-1-5
Swo-1-6
Swo-1-7
Slo-1-4
Slo-1-5
Slo-1-6
Sse-1-6
Spe-1-1
Sha-1-1
Spe-1-2
Slo-1-7
Spe-1-3
Sse-1-7
Sfr-1-1
Sfr-1-2
env-14
Ssp-1-2
env-15
Spu-2-1
Spu-1-1
Slo-1-8
Ssp-2-1
Ssp-3-2
Sde-1-1
Spe-1-4
Swo-1-8
Sha-1-2
Spe-1-5
Spe-1-6

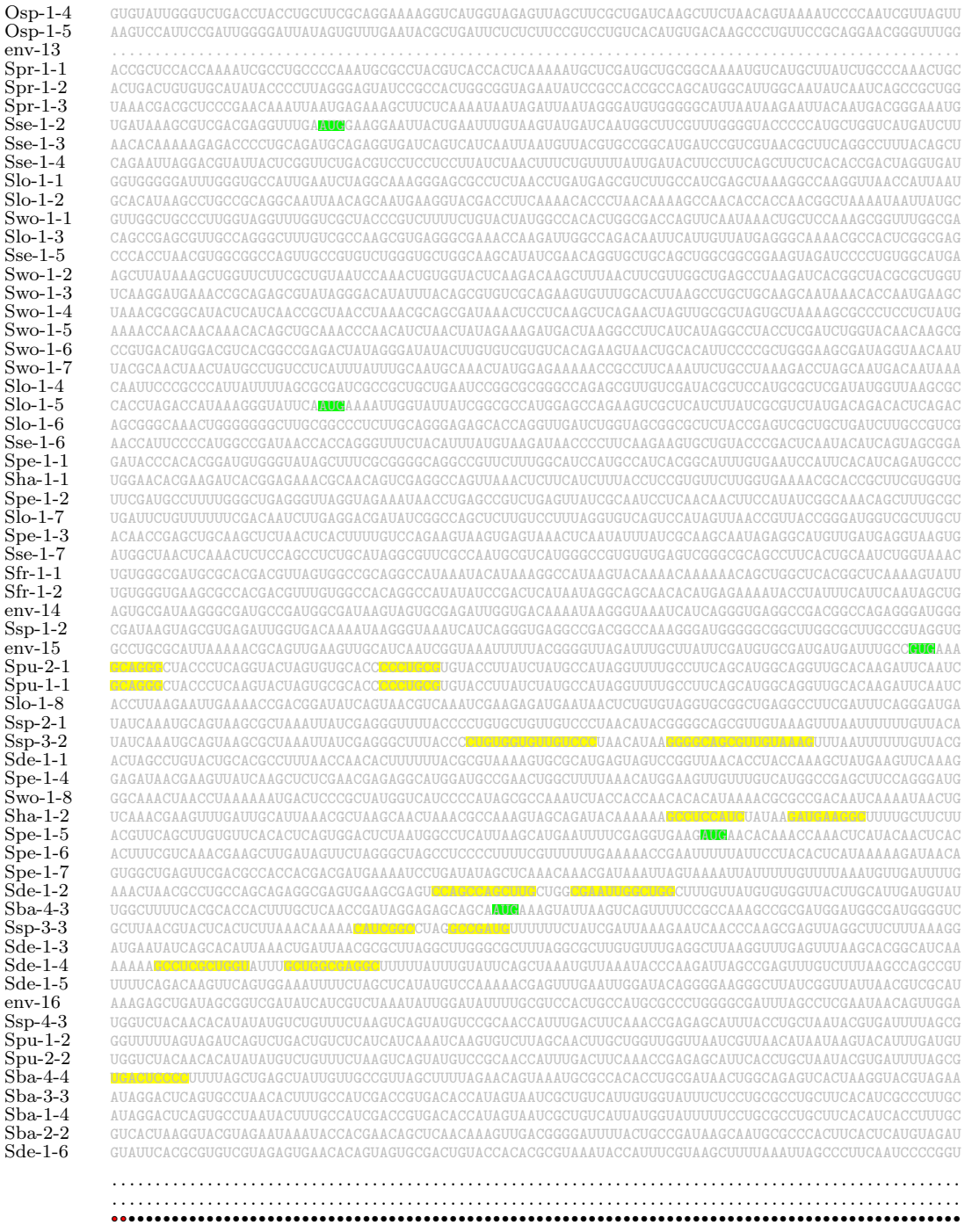

Spe-1-7
Sde-1-2
Sba-4-3
Ssp-3-3
Sde-1-3
Sde-1-4
Sde-1-5
env-16
Ssp-4-3
Spu-1-2
Spu-2-2
Sba-4-4
Sba-3-3
Sba-1-4
Sba-2-2
Sde-1-6
Spu-1-3
Spu-1-4
Ssp-4-4
Spu-2-3
Spu-1-5
Sba-3-4
Sba-1-5
Sba-3-5
Spu-2-4
Spu-2-5
Sba-1-6
Sba-3-6
Ssp-3-4
Son-1-4
env-17
env-18
Ssp-1-3
Ssp-3-5
Ssp-2-2
Ssp-2-3
env-19
env-20
Ssp-2-4
env-21
Sba-4-5
Sde-1-7
Sba-2-3
env-22
Sse-1-8
Swo-1-9
Swo-1-10
Swo-1-11
Swo-1-12
Swo-1-13
Swo-1-14
Sse-1-9
Swo-1-15
Sde-1-8
Sde-1-9
Sde-1-10
Slo-1-9
Ssp-2-5
Sba-2-4
Sba-2-5
Sba-1-7
Ssp-1-4

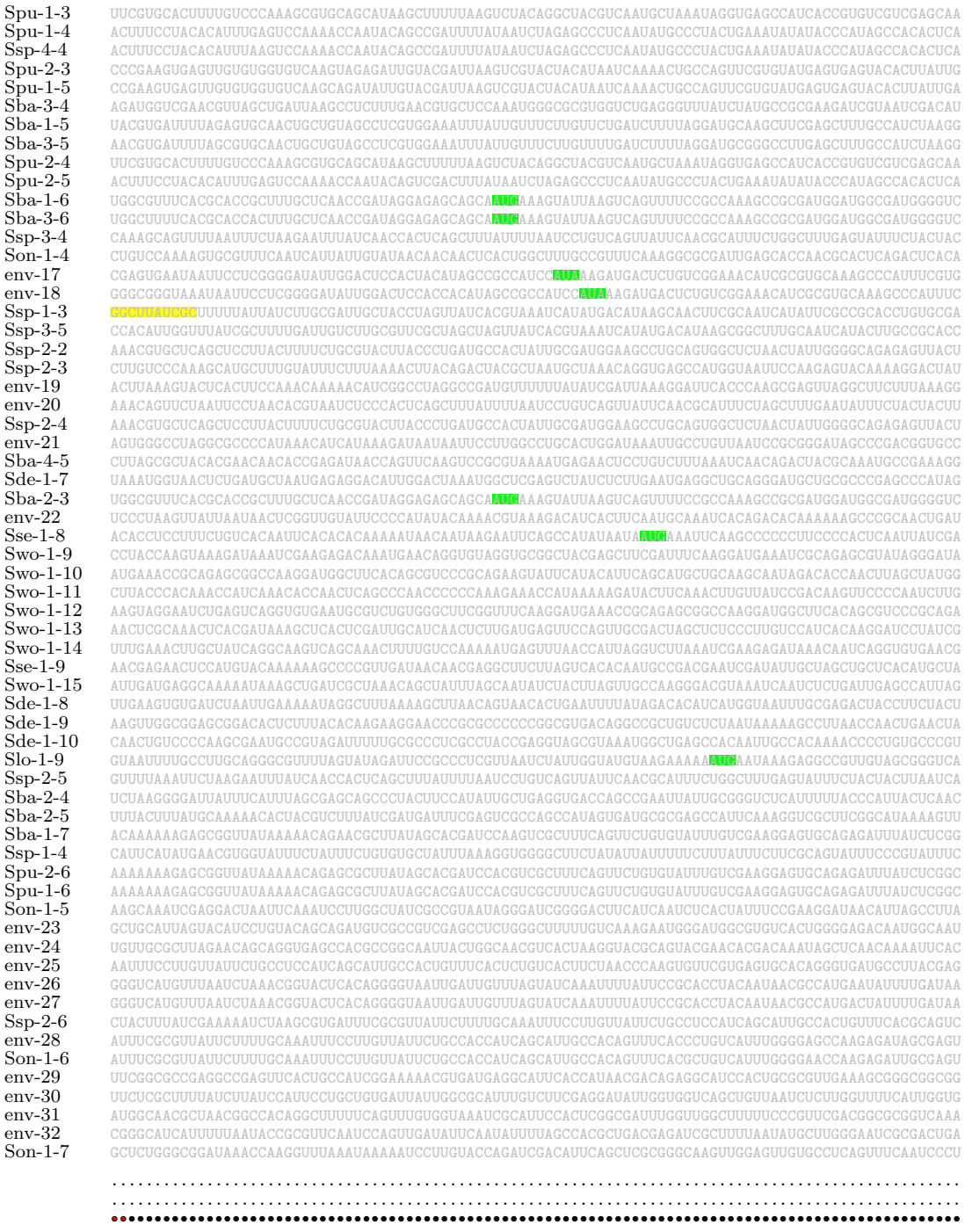

Spu-2-6
Spu-1-6
Son-1-5
env-23
env-24
env-25
env-26
env-27
Ssp-2-6
env-28
Son-1-6
env-29
env-30
env-31
env-32
Son-1-7
Spu-2-7
Spu-1-7
Sba-3-7
Sba-3-8
Sba-2-6
Ssp-1-5
Ssp-1-6
env-33
Ssp-4-6
Sba-4-6
Sba-1-8
Son-1-8
env-34
Sba-3-9
Spu-2-8
Spu-1-8
Sba-2-7
env-35
Spu-2-7
Ssp-1-7
env-36
env-37
env-38
Ssp-2-8
Ssp-3-6
Ssp-3-7
Ssp-1-8
env-39
Ssp-1-9
Ssp-2-9
Ssp-2-10
Ssp-3-8
Ssp-2-11
Ssp-3-9
Ssp-2-12
Ssp-2-13
env-40
Sba-2-8
Ssp-4-8
Spu-1-9
Spu-2-9
Sba-2-9
Spu-2-10
env-41
Sba-2-10
Sba-3-10

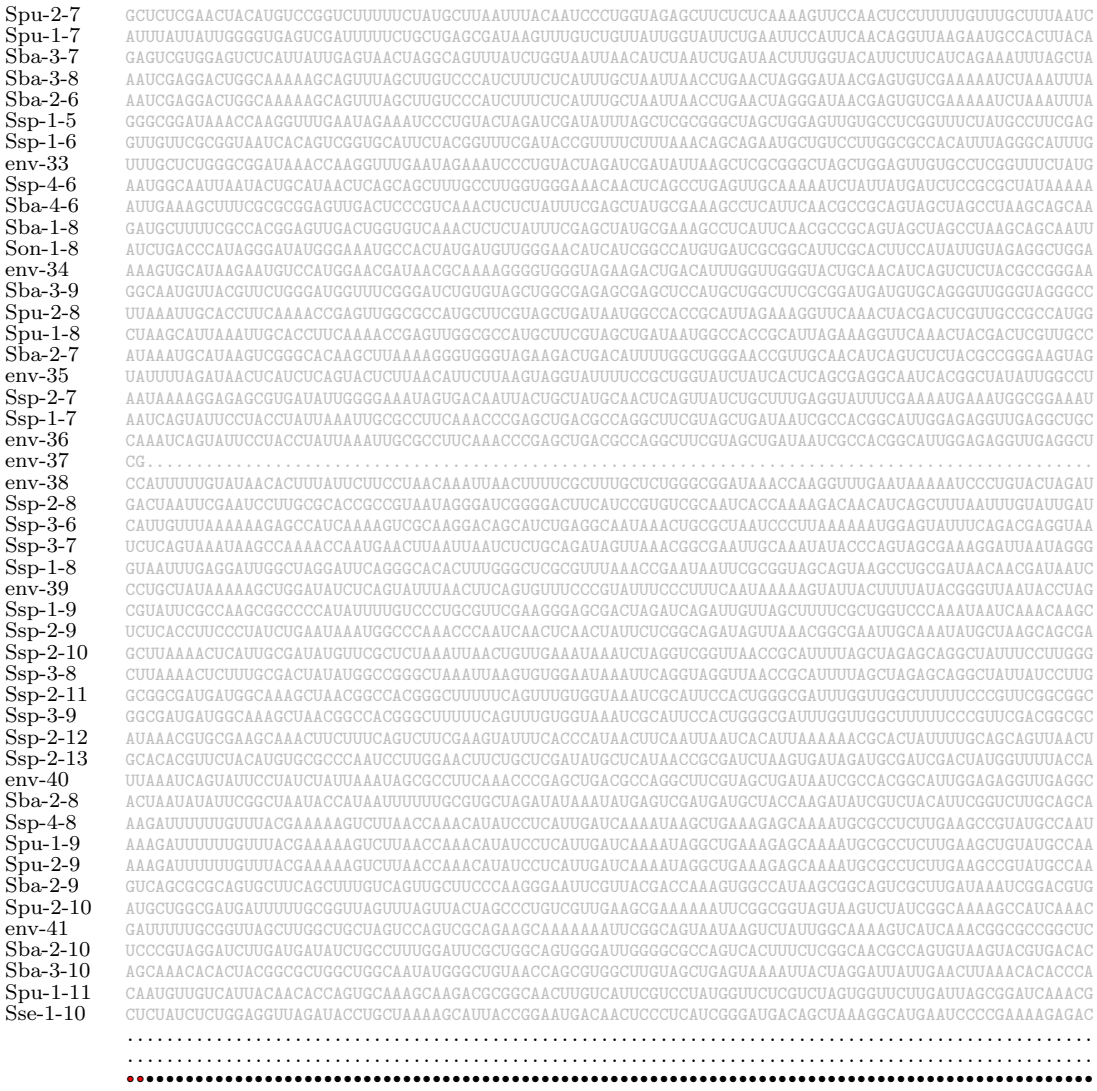


alignment positions 420 · · · 517




Osp-1-4
Osp-1-5
env-13
Spr-1-1
Spr-1-2
Spr-1-3
Sse-1-2
Sse-1-3
Sse-1-4
Slo-1-1
Slo-1-2
Swo-1-1
Slo-1-3
Sse-1-5
Swo-1-2
Swo-1-3
Swo-1-4
Swo-1-5
Swo-1-6
Swo-1-7
Slo-1-4
Slo-1-5
Slo-1-6
Sse-1-6
Spe-1-1
Sha-1-1
Spe-1-2
Slo-1-7
Spe-1-3
Sse-1-7
Sfr-1-1
Sfr-1-2
env-14
Ssp-1-2
env-15
Spu-2-1
Spu-1-1
Slo-1-8
Ssp-2-1
Ssp-3-2
Sde-1-1
Spe-1-4
Swo-1-8
Sha-1-2
Spe-1-5
Spe-1-6
Spe-1-7
Sde-1-2
Sba-4-3
Ssp-3-3
Sde-1-3
Sde-1-4
Sde-1-5
env-16
Ssp-4-3
Spu-1-2
Spu-2-2
Sba-4-4
Sba-3-3
Sba-1-4
Sba-2-2
Sde-1-6


Spu-1-3
Spu-1-4
Ssp-4-4
Spu-2-3
Spu-1-5
Sba-3-4
Sba-1-5
Sba-3-5
Spu-2-4
Spu-2-5
Sba-1-6
Sba-3-6
Ssp-3-4
Son-1-4
env-17
env-18
Ssp-1-3
Ssp-3-5
Ssp-2-2
Ssp-2-3
env-19
env-20
Ssp-2-4
env-21
Sba-4-5
Sde-1-7
Sba-2-3
env-22
Sse-1-8
Swo-1-9
Swo-1-10
Swo-1-11
Swo-1-12
Swo-1-13
Swo-1-14
Sse-1-9
Swo-1-15
Sde-1-8
Sde-1-9
Sde-1-10
Slo-1-9
Ssp-2-5
Sba-2-4
Sba-2-5
Sba-1-7
Ssp-1-4
Spu-2-6
Spu-1-6
Son-1-5
env-23
env-24
env-25
env-26
env-27
Ssp-2-6
env-28
Son-1-6
env-29
env-30
env-31
env-32
Son-1-7



28 Dictyoglomi-1

28.1 Taxa

The taxonomy of each organism containing a putative Dictyoglomi-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Dictyoglomi-1 RNA in Sections 28.2 and 28.4.

abbrev. of hits	taxonomy of species
Dth-1-1 to Dth-1-2	Bacteria Dictyoglomi Dictyoglomales Dictyoglomaceae *Dictyoglomus thermophilum* H-6-12
Dtu-1-1 to Dtu-1-2	Bacteria Dictyoglomi Dictyoglomales Dictyoglomaceae *Dictyoglomus turgidum* DSM 6724

28.2 Gene contexts

Each Dictyoglomi-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Dictyoglomi-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 28.3. The accession number of the sequence containing each Dictyoglomi-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 28.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Dth-1-1	NC_011297.1	-	1770511	1770277	hypo→ **RNA**→ lipoprotein, putative→
Dtu-1-1	NZ_ABYR01000004.1	+	195482	195716	hypo→ **RNA**→ glycoside hydrolase, family 5→
Dtu-1-2	NZ_ABYR01000001.1	-	451862	451627	Lpd (COG1249)→ **RNA**→ hypo→
Dth-1-2	NC_011297.1	+	383960	384195	Lpd (COG1249)→ **RNA**→ HDc (cd00077)→

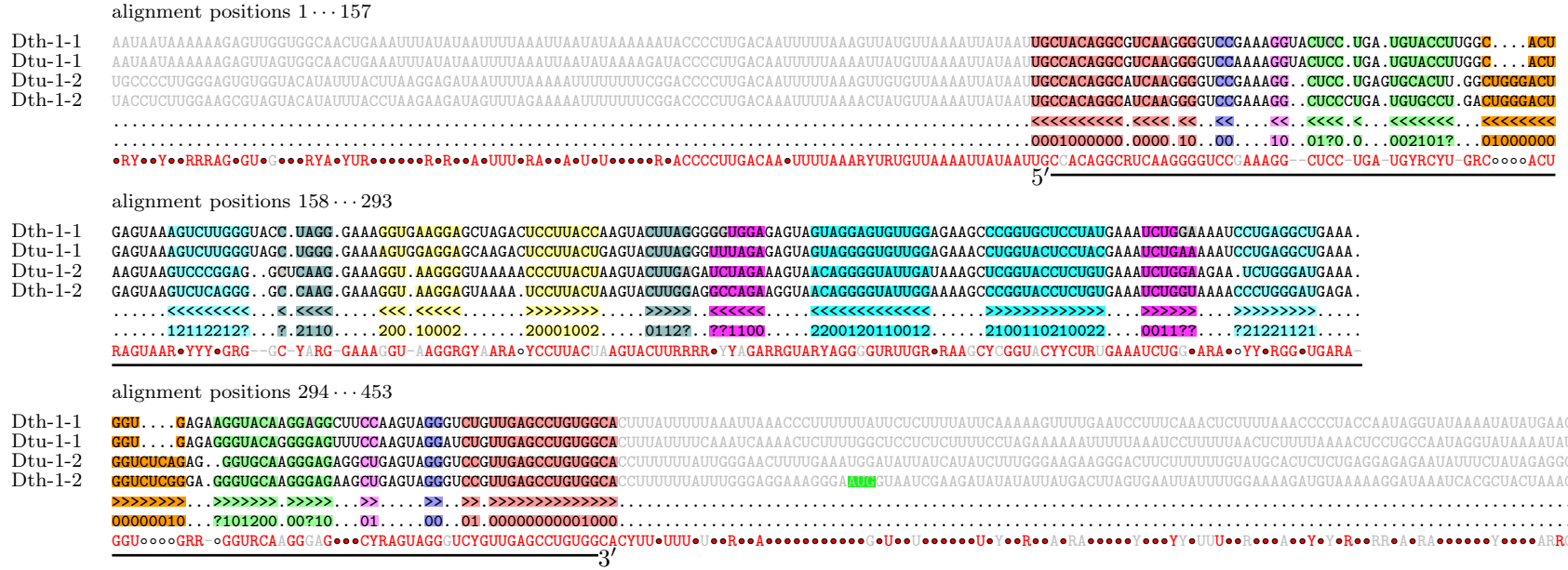
28.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 28.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Dictyoglomi-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

D cd00077 (1) Metal dependent phosphohydrolases with conserved 'HD' motif
u COG1249 (2) Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component, and related enzymes [Energy production and conversion]

28.4 Multiple-sequence alignment

Each Dictyoglomi-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 28.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Dictyoglomi-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



29 Downstream-peptide

29.1 Taxa

The taxonomy of each organism containing a putative Downstream-peptide RNA is listed, with abbreviations identifying each hit (e.g., “Eco-1-1” and “Eco-1-2” might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Downstream-peptide RNA in Sections 29.3 and 29.5.

abbrev. of hits	taxonomy of species
Ssp-1-1 to Ssp-1-7	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* BL107
Ssp-2-1 to Ssp-2-3	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* CC9311

Ssp-3-1 to Ssp-3-8	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* CC9605
Ssp-4-1 to Ssp-4-6	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* CC9902
Ssp-5-1 to Ssp-5-3	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* RCC307
Ssp-6-1 to Ssp-6-6	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* RS9916
Ssp-7-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* RS9917
Ssp-8-1 to Ssp-8-2	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 5701
Ssp-9-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 7803
Ssp-10-1 to Ssp-10-2	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 7805
Ssp-11-1 to Ssp-11-9	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 8102
Pma-1-1 to Pma-1-4	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* AS9601
Pma-2-1 to Pma-2-2	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9211
Pma-3-1 to Pma-3-5	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9301
Pma-4-1 to Pma-4-2	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9303
Pma-5-1 to Pma-5-4	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9312
Pma-6-1 to Pma-6-2	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9313
Pma-7-1 to Pma-7-3	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9515
Pma-8-1 to Pma-8-6	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* NATL1A
Pma-9-1 to Pma-9-6	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* NATL2A
Pma-10-1 to Pma-10-2	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus subsp. marinus str.* CCMP1375
Pma-11-1 to Pma-11-4	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus subsp. pastoris str.* CCMP1986
Pph-1-1	Viruses dsDNA viruses, no RNA stage Caudovirales Podoviridae T7-like viruses *Prochlorococcus phage* P-SSP7
env-1 to env-545	environmental samples

29.2 Notes

Superscript numbers are used to annotate any Downstream-peptide RNA that has special characteristics (described below). These numeric annotations will be used in Sections 29.3 and 29.5. The meaning of the numbers are as follows.

[1] Immediately upstream of the *ykkC* motif [5]. However there is a poly-N region that might correspond to a longer gap.

[2] The experimentally determined transcription start site [47] is shaded in blue. Nucleotides that are expected to anneal to either of the probes used for Northern blotting (oligonucleotides "y6MS" or "PMM3882n" in Supplementary Table S5 of [47] are underlined.

[3] This sequence does not appear to have a peptide encoded in the usual location.

[4] This RNA sequence corresponds to the yfr6 sequence in *Prochlorococcus marinus* MED4 in [47]. The sequence is labeled "MED4_yfr6" in Supplementary Fig. S10(A) of that paper [47]. Note: based on RefSeq data, we refer to this species as "*Prochlorococcus marinus* subsp. *pastoris* str. CCMP1986".

[5] This RNA sequence corresponds to the yfr6 sequence in *Prochlorococcus marinus* SS120 in [47]. The sequence is labeled "SS120_yfr6" in Supplementary Fig. S10(A) of that paper [47]. Note: based on RefSeq data, we refer to this species as "*Prochlorococcus marinus* subsp. *marinus* str. CCMP1375".

[6] This RNA sequence corresponds to the sequence annotated as "PMM3822n" in Supplementary Fig. S10(A) of [47].

[7] Although there is a downstream gene encoding a phage capsid protein, an unannotated short ORF is present that is highly similar to many other peptides downstream of other Downstream-peptide RNAs. This ORF is predicted to encode the peptide MKVTFVYRGVAYTRVIG.

29.3 Gene contexts

Each Downstream-peptide RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Downstream-peptide RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 29.4. The accession number of the sequence containing each Downstream-peptide RNA is given. Accessions beginning with

"NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 29.1. Superscript numbers refer to annotations listed in Section 29.2

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	JCVI_SCAF_1101668289908	+	228	281	**RNA**→ ? (GOS13866)→
env-2	JCVI_SCAF_1101668655546	-	1188	1132	**RNA**→ ? (GOS11953)→
env-3	JCVI_SCAF_1096627994902	+	138	191	**RNA**→ ? (GOS13266240)→
Ssp-3-1	NC_007516.1	-	2424541	2424485	**RNA**→ hypo→
Ssp-6-1	NZ_AAUA01000003.1	+	77229	77287	**RNA**→ hypo→ hypo→
Ssp-2-1	NC_008319.1	-	2112063	2112005	**RNA**→ hypo→
env-4	JCVI_SCAF_1101668046265	-	644	586	**RNA**→ ? (GOS11989004)→
env-5	JCVI_SCAF_1101668409450	+	832	888	**RNA**→
env-6	JCVI_SCAF_1096627064433	+	173	229	**RNA**→ ←? (GOS12142765)
env-7	JCVI_SCAF_1096627826325	-	1624	1548	**RNA**→ ? (GOS8442)→
env-8	JCVI_SCAF_1101668659752	-	427	367	**RNA**→
Ssp-1-1	NZ_AATZ01000001.1	-	53754	53694	**RNA**→ hypo→
Ssp-4-1	NC_007513.1	-	995827	995767	**RNA**→ hypo→ hypo→
env-9	JCVI_SCAF_1101668191067	-	603	543	**RNA**→ ? (GOS11176343)→
Pma-6-1	NC_005071.1	+	1082118	1082175	**RNA**→ COG1917 (COG1917)→
Pma-4-1	NC_008820.1	-	984332	984275	**RNA**→ hypo→ COG1917 (COG1917)→
env-10	JCVI_SCAF_1096627521448	+	1033	1088	**RNA**→ ? (GOS13006886)→
env-11	JCVI_SCAF_1096628226907	+	235	290	**RNA**→ ? (GOS13006886)→
env-12	JCVI_SCAF_1101667546431	-	632	577	**RNA**→ ←? (GOS13233019)
env-13	JCVI_SCAF_1101668743995	-	788	733	**RNA**→ ←? (GOS13233019)
env-14	JCVI_SCAF_1096628072622	-	474	419	**RNA**→ ←? (GOS13233019)
env-15	JCVI_SCAF_1096628119360	+	1608	1663	**RNA**→ ←? (GOS13233019)
env-16	JCVI_SCAF_1096626980065	-	297	242	**RNA**→ ←? (GOS13233019)
env-17	JCVI_SCAF_1096627785235	+	230	284	**RNA**→ ? (GOS1472)→
env-18	JCVI_SCAF_1096628007144	-	1142	1084	**RNA**→ ? (GOS1472)→
env-19	JCVI_SCAF_1101668456928	+	772	830	**RNA**→
env-20	JCVI_SCAF_1096627083963	-	1664	1606	**RNA**→ ? (GOS1472)→
env-21	JCVI_SCAF_1096627068486	-	1474	1416	**RNA**→ ? (GOS1472)→
env-22	JCVI_SCAF_1096627052674	+	1232	1294	**RNA**→ ? (GOS14256)→
Pma-3-1	NC_009091.1	-	1086706	1086644	**RNA**→ hypo→
Pma-5-1	NC_007577.1	-	1134894	1134832	**RNA**→ hypo→
env-23	JCVI_SCAF_1096626999439	+	3201	3258	**RNA**→ ? (GOS13006886)→
env-24	JCVI_SCAF_1096626959601	-	636	579	**RNA**→ ? (GOS15836)→
env-25	JCVI_SCAF_1101667564102	-	638	581	**RNA**→ ←? (GOS11438381)
Pma-3-2	NC_009091.1	+	1058156	1058213	**RNA**→ hypo→
env-26	JCVI_SCAF_1096627520867	-	486	429	**RNA**→ ? (GOS13006886)→
env-27	JCVI_SCAF_1096627569593	+	1791	1848	**RNA**→
env-28	JCVI_SCAF_1096627606732	-	1370	1313	**RNA**→ ? (GOS13006886)→
env-29	JCVI_SCAF_1101667106942	-	267	210	**RNA**→ ←? (GOS13233019)
env-30	JCVI_SCAF_1096627610797	+	746	803	**RNA**→ ABC_PhnC_transporter (cd03256)? (GOS2568)→
env-31	JCVI_SCAF_1096626471656	+	147	204	**RNA**→ ←? (GOS13233019)
env-32	JCVI_SCAF_1096628295520	+	1175	1232	**RNA**→ ←? (GOS13233019)
env-33	JCVI_SCAF_1096626095788	-	554	497	**RNA**→ ←? (GOS13233019)
env-34	JCVI_SCAF_1096627569103	-	458	401	**RNA**→ ←? (GOS13233019)
env-35	JCVI_SCAF_1096626936231	-	1425	1368	**RNA**→ ←? (GOS13233019)
env-36	JCVI_SCAF_1101667552712	-	80	23	**RNA**→

env-37	JCVI_SCAF_1096626960057	+	550	607	**RNA**→ ←? (GOS13233019)
env-38	JCVI_SCAF_1101668218297	-	80	23	**RNA**→
env-39	JCVI_SCAF_1101667400692	-	611	554	**RNA**→ ←? (GOS13233019)
env-40	JCVI_SCAF_1101667536972	+	721	778	**RNA**→ ←? (GOS13233019)
env-41	JCVI_SCAF_1096628275431	-	1538	1481	**RNA**→ ←? (GOS13233019)
env-42	JCVI_SCAF_1096626978838	+	1025	1082	**RNA**→
env-43	JCVI_SCAF_1101667134718	-	320	263	**RNA**→ ←? (GOS13233019)
env-44	JCVI_SCAF_1101667138729	-	545	483	**RNA**→
env-45	JCVI_SCAF_1096626097519	+	676	738	**RNA**→ ? (GOS14256)→
env-46	JCVI_SCAF_1101667571257	+	449	511	**RNA**→
env-47	JCVI_SCAF_1096628259733	-	1036	974	**RNA**→ ? (GOS11904738)→
env-48	JCVI_SCAF_1101667135252	-	367	305	**RNA**→ ? (GOS13006886)→
env-49	JCVI_SCAF_1096627630644	+	511	573	**RNA**→ ? (GOS14256)→
env-50	JCVI_SCAF_1101668305811	+	1046	1108	**RNA**→
env-51	JCVI_SCAF_1101667222200	-	375	313	**RNA**→
env-52	JCVI_SCAF_1101668191088	-	1189	1127	**RNA**→ ? (GOS14256)→
env-53	JCVI_SCAF_1096628243958	+	1430	1492	**RNA**→
env-54	JCVI_SCAF_1096626999314	-	1332	1270	**RNA**→ ? (GOS14256)→
env-55	JCVI_SCAF_1096627618465	+	344	406	**RNA**→ ←? (GOS16291)
env-56	JCVI_SCAF_1101667218927	-	261	199	**RNA**→
env-57	JCVI_SCAF_1101668287475	+	1153	1215	**RNA**→
env-58	JCVI_SCAF_1096627578122	+	550	612	**RNA**→ ? (GOS11904738)→
env-59	JCVI_SCAF_1096628253868	+	436	498	**RNA**→ ←? (GOS13431674)
env-60	JCVI_SCAF_1096627586819	-	1134	1072	**RNA**→ ? (GOS11671983)→ ? (GOS11904738)→
env-61	JCVI_SCAF_1096627606803	+	1182	1244	**RNA**→ ? (GOS14256)→
env-62	JCVI_SCAF_1101668745825	+	1158	1220	**RNA**→
env-63	JCVI_SCAF_1096627518427	+	1011	1073	**RNA**→
env-64	JCVI_SCAF_1101668188755	-	1391	1329	**RNA**→ ? (GOS14256)→
env-65	JCVI_SCAF_1096627026942	+	645	707	**RNA**→ ? (GOS13006886)→
env-66	JCVI_SCAF_1096627708039	-	1351	1289	**RNA**→ ? (GOS13006886)→
env-67	JCVI_SCAF_1101668231065	-	574	512	**RNA**→ ? (GOS13006886)→
env-68	JCVI_SCAF_1101668185167	+	742	804	**RNA**→ ? (GOS14256)→
env-69	JCVI_SCAF_1096626990755	-	2049	1987	**RNA**→ ? (GOS14256)→
env-70	JCVI_SCAF_1096627561662	-	765	703	**RNA**→ ? (GOS14256)→
Pma-1-1	NC_008816.1	-	1075404	1075342	**RNA**→ hypo→
env-71	JCVI_SCAF_1096626968262	+	386	448	**RNA**→ ? (GOS14256)→ ? (GOS11671983)→ ? (GOS11904738)→
env-72	JCVI_SCAF_1096627059895	+	890	952	**RNA**→ ? (GOS14256)→ ? (GOS11671983)→ ? (GOS11904738)→
env-73	JCVI_SCAF_1096627061190	+	904	966	**RNA**→ ? (GOS14256)→
env-74	JCVI_SCAF_1096627058721	-	776	714	**RNA**→ ? (GOS14256)→
Pma-7-1	NC_008817.1	-	378828	378767	**RNA**→ hypo→
Pma-11-1	NC_005072.1	+	817330	817389	**RNA**→ tRNA-Thr→
Pma-8-1	NC_008819.1	-	1135171	1135114	**RNA**→ hypo→ hypo→ hypo→
env-75	JCVI_SCAF_1096628080188	+	934	991	**RNA**→
Pma-9-1	NC_007335.1	-	532841	532784	**RNA**→ hypo→
Ssp-4-2	NC_007513.1	+	937913	937973	**RNA**→ ←possible transcriptional regulator
Ssp-1-2	NZ_AATZ01000001.1	-	347865	347805	**RNA**→ ←possible transcriptional regulator
env-76	JCVI_SCAF_1096627402304	-	1430	1376	**RNA**→ ? (GOS13266240)→
env-77	JCVI_SCAF_1101667261160	-	468	413	**RNA**→ ? (GOS13266240)→
env-78	JCVI_SCAF_1101667265019	-	684	629	**RNA**→ ? (GOS13266240)→
env-79	JCVI_SCAF_1096627282746	+	553	606	**RNA**→ ←Peptidase_C39A (cd02549)FlgJ (COG1705)? (GOS13049017)
env-80	JCVI_SCAF_1096627267455	-	1860	1803	**RNA**→ ←adh_short (pfam00106)? (GOS1303)
Ssp-5-1	NC_009482.1	+	1340859	1340915	**RNA**→ ←hypo

env-81	JCVI_SCAF_1101667018194	-	201	137	**RNA**→
Pma-8-2	NC_008819.1	-	1478715	1478655	**RNA**→ hypo→ hypo→
Pma-9-2	NC_007335.1	-	893813	893753	**RNA**→ protein family PM-16→
env-82	JCVI_SCAF_1101668222309	+	299	361	**RNA**→
env-83	JCVI_SCAF_1096627516743	+	518	580	**RNA**→ ←? (GOS11316857)
env-84	JCVI_SCAF_1101667092473	-	141	79	**RNA**→
env-85	JCVI_SCAF_1101667546597	+	705	767	**RNA**→
env-86	JCVI_SCAF_1096627523083	+	876	938	**RNA**→
env-87	JCVI_SCAF_1096627602968	+	406	468	**RNA**→ ←? (GOS12053)
env-88	JCVI_SCAF_1096627593144	-	406	344	**RNA**→
env-89	JCVI_SCAF_1096627743440	-	977	915	**RNA**→ ←? (GOS12053)
env-90	JCVI_SCAF_1101668304136	-	858	796	**RNA**→
env-91	JCVI_SCAF_1101667229515	-	628	566	**RNA**→ ←? (GOS11943545)
env-92	JCVI_SCAF_1096626934041	+	1076	1138	**RNA**→
env-93	JCVI_SCAF_1096628053320	+	1446	1508	**RNA**→
env-94	JCVI_SCAF_1101668313317	-	588	526	**RNA**→
env-95	JCVI_SCAF_1101668318303	+	10	72	**RNA**→ ←? (GOS12053)
env-96	JCVI_SCAF_1096627622609	-	554	492	**RNA**→
env-97	JCVI_SCAF_1096627523083	+	1090	1152	**RNA**→
env-98	JCVI_SCAF_1096627604214	+	1035	1097	**RNA**→
env-99	JCVI_SCAF_1096627561598	-	1301	1239	**RNA**→ ←? (GOS12053)
env-100	JCVI_SCAF_1096626934990	-	530	468	**RNA**→
env-101	JCVI_SCAF_1096627202871	-	129	67	**RNA**→
env-102	JCVI_SCAF_1096628082332	-	1938	1876	**RNA**→ ? (GOS16662)→
env-103	JCVI_SCAF_1096628238651	+	1406	1468	**RNA**→
env-104	JCVI_SCAF_1096627566151	-	1540	1478	**RNA**→ ←? (GOS12053)
env-105	JCVI_SCAF_1096627002424	-	184	122	**RNA**→
env-106	JCVI_SCAF_1101667225526	+	184	246	**RNA**→
env-107	JCVI_SCAF_1101667385030	+	606	668	**RNA**→
env-108	JCVI_SCAF_1096628249206	+	227	289	**RNA**→ ? (GOS16662)→
env-109	JCVI_SCAF_1101668190215	-	366	304	**RNA**→
env-110	JCVI_SCAF_1096627586348	+	779	841	**RNA**→ ←? (GOS12053)
env-111	JCVI_SCAF_1096627610532	-	783	721	**RNA**→
env-112	JCVI_SCAF_1101668312275	-	366	304	**RNA**→
env-113	JCVI_SCAF_1096627059153	-	1749	1687	**RNA**→ ←? (GOS11316857)
env-114	JCVI_SCAF_1096627723867	+	565	627	**RNA**→ ←? (GOS12053)
Pma-3-3	NC_009091.1	-	1338899	1338837	**RNA**→ ←hypo
env-115	JCVI_SCAF_1096626935261	+	1374	1436	**RNA**→
env-116	JCVI_SCAF_1096627711188	-	701	639	**RNA**→
env-117	JCVI_SCAF_1101667132727	-	816	754	**RNA**→
env-118	JCVI_SCAF_1096627701549	-	140	78	**RNA**→
env-119	JCVI_SCAF_1096628264309	-	566	504	**RNA**→
env-120	JCVI_SCAF_1096626141979	+	527	589	**RNA**→
env-121	JCVI_SCAF_1101668307066	+	547	609	**RNA**→ ←? (GOS12053)
env-122	JCVI_SCAF_1096627584522	+	287	349	**RNA**→ ←? (GOS12053)
env-123	JCVI_SCAF_1096627527906	+	586	648	**RNA**→ ←? (GOS12053)
env-124	JCVI_SCAF_1096628144671	-	926	864	**RNA**→
env-125	JCVI_SCAF_1096626927485	-	1352	1290	**RNA**→ ←? (GOS12053)
env-126	JCVI_SCAF_1101667113543	-	94	32	**RNA**→
env-127	JCVI_SCAF_1096628048852	-	491	429	**RNA**→
env-128	JCVI_SCAF_1096628290390	+	1378	1440	**RNA**→
env-129	JCVI_SCAF_1096627628925	-	1773	1711	**RNA**→ ←? (GOS12053)

env-130	JCVI_SCAF_1096627702601	+	87	149	**RNA**→ ←? (GOS12053)
env-131	JCVI_SCAF_1096628248893	-	724	662	**RNA**→
env-132	JCVI_SCAF_1101667089406	-	501	439	**RNA**→
env-133	JCVI_SCAF_1101667089573	-	443	381	**RNA**→
env-134	JCVI_SCAF_1096626954386	+	64	126	**RNA**→ ←? (GOS12053)
env-135	JCVI_SCAF_1096626967194	-	441	379	**RNA**→
env-136	JCVI_SCAF_1096627633563	-	95	33	**RNA**→
env-137	JCVI_SCAF_1096626957726	-	845	783	**RNA**→ ←? (GOS12053)
env-138	JCVI_SCAF_1096626100743	+	899	961	**RNA**→
Pma-1-2	NC_008816.1	-	1365468	1365406	**RNA**→ ←hypo
env-139	JCVI_SCAF_1096627524822	-	442	380	**RNA**→
env-140	JCVI_SCAF_1101667106312	+	426	488	**RNA**→
env-141	JCVI_SCAF_1096627774175	-	279	217	**RNA**→
Pma-5-2	NC_007577.1	-	1396332	1396270	**RNA**→ ←hypo
env-142	JCVI_SCAF_1096627781874	-	1189	1127	**RNA**→ ←? (GOS12053)
Pma-11-2	NC_005072.1	-	350508	350446	**RNA**→ hypo→
env-143	JCVI_SCAF_1101668213263	+	254	316	**RNA**→ ←? (GOS13140869)
env-144	JCVI_SCAF_1101667200386	+	483	545	**RNA**→
env-145	JCVI_SCAF_1101667354626	-	613	551	**RNA**→
env-146	JCVI_SCAF_1096627589973	+	1256	1323	**RNA**→
env-147	JCVI_SCAF_1096627594591	-	672	610	**RNA**→
Ssp-11-1	NC_005070.1	+	2311575	2311634	**RNA**→ hypo→
env-148	JCVI_SCAF_1096627494697	+	423	475	**RNA**→
env-149	JCVI_SCAF_1101668196604	-	682	630	**RNA**→ ←? (GOS11555940)
env-150	JCVI_SCAF_1096627183319	+	558	613	**RNA**→ ? (GOS1580)→
env-151	JCVI_SCAF_1096627287289	+	1490	1549	**RNA**→ ? (GOS11176343)→
env-152	JCVI_SCAF_1101668477457	+	669	727	**RNA**→
env-153	JCVI_SCAF_1096628280301	-	1487	1429	**RNA**→
Pma-8-3	NC_008819.1	-	1071068	1071010	**RNA**→ hypo→ hypo→
Pma-9-3	NC_007335.1	-	468724	468666	**RNA**→ hypo→ hypo→
env-154	JCVI_SCAF_1096627481112	-	1633	1579	**RNA**→ ←Leu tRNA
Ssp-8-1	NZ_AANO01000053.1	-	1174	1115	**RNA**→ hypo→
Ssp-1-3	NZ_AATZ01000001.1	+	313473	313532	**RNA**→ hypo→
Ssp-4-3	NC_007513.1	-	973366	973307	**RNA**→ hypo→ AmpC (COG1680)→
env-155	JCVI_SCAF_1096627192017	+	1369	1429	**RNA**→
Ssp-11-2	NC_005070.1	+	1344747	1344807	**RNA**→ hypo→
env-156	JCVI_SCAF_1096628065506	-	458	398	**RNA**→ ? (GOS11064110)→
[3]Ssp-10-1	NZ_AAOK01000003.1	-	23492	23438	**RNA**→ hypo→
Ssp-6-2	NZ_AAUA01000003.1	-	375474	375414	**RNA**→ hypo→
Ssp-9-1	NC_009481.1	-	901002	900948	**RNA**→ hypo→ MnhC (COG1006)→ HyfB (COG0651)→ Conserved membrane protein, putative subunit of the Multisubunit Na+/H+ antiporter→ Conserved membrane protein, putative subunit of the Multisubunit Na+/H+ antiporter→ Possible Multisubunit Na+/H+ antiporter, MnhG subunit→ COG1563 (COG1563)→ Possible Multisubunit Na+/H+ antiporter, MnhB subunit→ hypo→ hypo→ Uncharacterized membrane protein→
env-157	JCVI_SCAF_1096627355771	-	1924	1868	**RNA**→ ? (GOS14707434)→
env-158	JCVI_SCAF_1096627357620	+	448	504	**RNA**→ ? (GOS14707434)→
env-159	JCVI_SCAF_1096627945736	-	263	207	**RNA**→ ? (GOS14707434)→
env-160	JCVI_SCAF_1096627948887	-	131	75	**RNA**→
env-161	JCVI_SCAF_1096627146609	-	423	367	**RNA**→ ? (GOS14707434)→
env-162	JCVI_SCAF_1096627948636	+	1770	1826	**RNA**→
Ssp-7-1	NZ_AANP01000002.1	-	393072	393016	**RNA**→ hypo→
env-163	JCVI_SCAF_1096627350436	+	2224	2280	**RNA**→
env-164	JCVI_SCAF_1096627351762	-	1030	974	**RNA**→ ? (GOS14707434)→

env-165	JCVI_SCAF_1101668655416	+	1065	1124	**RNA**→ ? (GOS11064110)→
env-166	JCVI_SCAF_1101668656695	-	1362	1303	**RNA**→ ? (GOS11064110)→
env-167	JCVI_SCAF_1101668659847	+	181	240	**RNA**→ ? (GOS11064110)→ AmpC (COG1680)→
env-168	JCVI_SCAF_1096627970216	+	1593	1652	**RNA**→
env-169	JCVI_SCAF_1101668657394	-	1175	1116	**RNA**→ ? (GOS11064110)→
Ssp-3-2	NC_007516.1	+	1391085	1391144	**RNA**→ hypo→ AmpC (COG1680)→
env-170	JCVI_SCAF_1101668751243	-	264	207	**RNA**→
Ssp-5-2	NC_009482.1	+	1019340	1019397	**RNA**→ hypo→ hypo→
env-171	JCVI_SCAF_1101667587068	-	496	435	**RNA**→ ? (GOS11064110)→
env-172	JCVI_SCAF_1096627187494	+	1803	1864	**RNA**→ ? (GOS11064110)→
Ssp-1-4	NZ_AATZ01000003.1	-	210964	210907	**RNA**→ hypo→
env-173	JCVI_SCAF_1101668260076	-	1148	1063	**RNA**→ ? (GOS1580)→
env-174	JCVI_SCAF_1101667511121	-	544	490	**RNA**→ ? (GOS1580)→
env-175	JCVI_SCAF_1096627531593	+	824	918	**RNA**→ ←? (GOS11555940)
env-176	JCVI_SCAF_1101668039154	+	344	400	**RNA**→ ? (GOS1580)→
env-177	JCVI_SCAF_1096627284860	+	2763	2819	**RNA**→ ? (GOS1580)→ ? (GOS1580)→
env-178	JCVI_SCAF_1101668731450	-	993	941	**RNA**→ ? (GOS1580)→
env-179	JCVI_SCAF_1096628256812	-	827	771	**RNA**→ ? (GOS1580)→
env-180	JCVI_SCAF_1096626607299	-	412	303	**RNA**→
env-181	JCVI_SCAF_1096626947839	-	1119	1065	**RNA**→ ? (GOS13219520)→
env-182	JCVI_SCAF_1096628129588	-	127	73	**RNA**→
env-183	JCVI_SCAF_1101668427640	-	540	488	**RNA**→ ? (GOS11044226)→
env-184	JCVI_SCAF_1096627233442	+	573	627	**RNA**→ ? (GOS11044226)→
env-185	JCVI_SCAF_1096627593981	+	587	641	**RNA**→ ? (GOS1580)→
env-186	JCVI_SCAF_1096626969116	+	975	1029	**RNA**→ ? (GOS11044226)→
env-187	JCVI_SCAF_1096627562133	+	862	916	**RNA**→ ? (GOS1580)→
env-188	JCVI_SCAF_1096627541658	-	1110	1058	**RNA**→ ? (GOS1580)→ ? (GOS8463)→
env-189	JCVI_SCAF_1101668217552	+	1139	1191	**RNA**→ ? (GOS1580)→
env-190	JCVI_SCAF_1096627534126	-	1607	1555	**RNA**→ ? (GOS1580)→
env-191	JCVI_SCAF_1096627306694	+	756	808	**RNA**→ ? (GOS1580)→
env-192	JCVI_SCAF_1096627567771	+	1540	1592	**RNA**→
env-193	JCVI_SCAF_1096627252940	-	653	601	**RNA**→ ? (GOS1580)→
env-194	JCVI_SCAF_1096627547107	-	458	406	**RNA**→ ? (GOS1580)→
env-195	JCVI_SCAF_1101668289755	+	203	255	**RNA**→ ? (GOS1580)→ ? (GOS1580)→
env-196	JCVI_SCAF_1101667222996	-	60	8	**RNA**→
env-197	JCVI_SCAF_1096626974808	+	2130	2191	**RNA**→ ? (GOS1580)→
env-198	JCVI_SCAF_1096627322451	-	613	561	**RNA**→ ? (GOS1580)→
env-199	JCVI_SCAF_1096627056233	-	1492	1438	**RNA**→ ? (GOS1580)→
env-200	JCVI_SCAF_1096627585996	+	983	1037	**RNA**→ ? (GOS1580)→
env-201	JCVI_SCAF_1101667336283	+	106	157	**RNA**→
env-202	JCVI_SCAF_1101668674577	+	852	906	**RNA**→ ? (GOS1580)→
env-203	JCVI_SCAF_1096628029163	-	1139	1085	**RNA**→ ? (GOS1580)→
env-204	JCVI_SCAF_1096627607392	-	1176	1123	**RNA**→ ? (GOS1580)→ ? (GOS891)→
env-205	JCVI_SCAF_1096627960620	-	540	486	**RNA**→ ? (GOS1580)→
env-206	JCVI_SCAF_1096627607836	+	82	139	**RNA**→ ? (GOS11025)→
env-207	JCVI_SCAF_1096627179155	-	1483	1424	**RNA**→ ? (GOS1580)→
env-208	JCVI_SCAF_1096626940869	+	1976	2032	**RNA**→ ? (GOS11025)→
env-209	JCVI_SCAF_1096626957614	+	2035	2090	**RNA**→
env-210	JCVI_SCAF_1096627559437	-	363	310	**RNA**→
env-211	JCVI_SCAF_1096627761896	+	1451	1504	**RNA**→
env-212	JCVI_SCAF_1096627759154	+	1154	1207	**RNA**→ ? (GOS1580)→
env-213	JCVI_SCAF_1101668300360	+	130	183	**RNA**→ ? (GOS1580)→ ? (GOS1580)→ ? (GOS8463)→

env-214	JCVI_SCAF_1096627732172	+	1056	1109	**RNA**→ ? (GOS1580)→
env-215	JCVI_SCAF_1096627056426	+	402	455	**RNA**→ ? (GOS1580)→
env-216	JCVI_SCAF_1096627055347	+	1265	1330	**RNA**→ ? (GOS11614916)→ ? (GOS1580)→
env-217	JCVI_SCAF_1096627386793	-	409	356	**RNA**→
env-218	JCVI_SCAF_1096626185629	+	481	534	**RNA**→ ? (GOS1580)→
env-219	JCVI_SCAF_1096626987464	-	221	168	**RNA**→
env-220	JCVI_SCAF_1096627228425	+	947	1009	**RNA**→ ? (GOS1580)→ ? (GOS8463)→
env-221	JCVI_SCAF_1101667158650	-	99	37	**RNA**→
env-222	JCVI_SCAF_1096627055829	+	3528	3590	**RNA**→ ? (GOS1580)→ ? (GOS1580)→
env-223	JCVI_SCAF_1096627555668	+	432	494	**RNA**→ ? (GOS1580)→
env-224	JCVI_SCAF_1096626154670	-	204	142	**RNA**→
env-225	JCVI_SCAF_1096627059539	+	3638	3700	**RNA**→ ? (GOS1580)→ ? (GOS1580)→
env-226	JCVI_SCAF_1096628032276	+	1276	1338	**RNA**→
env-227	JCVI_SCAF_1101668269141	+	1468	1530	**RNA**→
env-228	JCVI_SCAF_1096627753581	+	109	171	**RNA**→ ? (GOS1580)→ ? (GOS1580)→ ? (GOS8463)→
env-229	JCVI_SCAF_1096627743763	+	1104	1166	**RNA**→ ? (GOS1580)→ ? (GOS1580)→
env-230	JCVI_SCAF_1096627238603	-	980	922	**RNA**→ ? (GOS1580)→
env-231	JCVI_SCAF_1096628163587	+	356	414	**RNA**→ ? (GOS2977)→
env-232	JCVI_SCAF_1101668300923	-	1152	1099	**RNA**→ ? (GOS1580)→
env-233	JCVI_SCAF_1096627000238	+	488	553	**RNA**→ ? (GOS11614916)→ ? (GOS1580)→
env-234	JCVI_SCAF_1096626657250	+	823	888	**RNA**→ ? (GOS1580)→
env-235	JCVI_SCAF_1101667141015	-	506	441	**RNA**→
env-236	JCVI_SCAF_1096626984772	-	127	62	**RNA**→
env-237	JCVI_SCAF_1101667119071	+	688	741	**RNA**→
env-238	JCVI_SCAF_1096626931250	-	899	846	**RNA**→ ? (GOS1580)→
env-239	JCVI_SCAF_1096627385166	+	1694	1747	**RNA**→ ? (GOS1580)→
env-240	JCVI_SCAF_1101668425212	-	327	276	**RNA**→
env-241	JCVI_SCAF_1096626946934	+	2102	2155	**RNA**→ ? (GOS1580)→
env-242	JCVI_SCAF_1101668340711	+	824	876	**RNA**→
env-243	JCVI_SCAF_1096627801357	+	27	79	**RNA**→ ? (GOS1580)→
env-244	JCVI_SCAF_1096627621612	+	1207	1261	**RNA**→ ? (GOS11025)→
env-245	JCVI_SCAF_1096627898310	+	303	357	**RNA**→ ? (GOS1580)→
env-246	JCVI_SCAF_1096626852511	-	1442	1390	**RNA**→ ? (GOS1580)→ ? (GOS1580)→
env-247	JCVI_SCAF_1101668223808	-	1187	1131	**RNA**→ ? (GOS11025)→ ? (GOS11025)→
env-248	JCVI_SCAF_1096626961784	-	682	626	**RNA**→ ? (GOS1580)→
env-249	JCVI_SCAF_1096627570175	+	176	230	**RNA**→ ? (GOS1580)→ ? (GOS1580)→ ? (GOS8463)→
env-250	JCVI_SCAF_1096627046039	-	821	769	**RNA**→ ? (GOS1580)→
env-251	JCVI_SCAF_1096626848371	-	874	820	**RNA**→ ? (GOS1580)→
env-252	JCVI_SCAF_1096627284636	+	1560	1614	**RNA**→ ? (GOS1580)→
env-253	JCVI_SCAF_1101668044215	-	364	310	**RNA**→
env-254	JCVI_SCAF_1096628395592	+	699	753	**RNA**→ ? (GOS12135550)→
env-255	JCVI_SCAF_1096627558875	-	316	262	**RNA**→ ? (GOS12135550)→
env-256	JCVI_SCAF_1096627604031	-	1069	1017	**RNA**→ PstS (COG0226)? (GOS1508)→
env-257	JCVI_SCAF_1096628107425	-	1647	1590	**RNA**→ ? (GOS11025)→
env-258	JCVI_SCAF_1101667104132	+	141	198	**RNA**→ ? (GOS11025)→
[7]Pph-1-1	NC_006882.1	+	19588	19642	**RNA**→ T7-like capsid protein→ T7-like tail tubular protein A→ T7-like tail tubular protein B→ hypo→ hypo→ T7-like internal core protein→ T7-like internal core protein→
env-259	JCVI_SCAF_1096627055919	+	805	859	**RNA**→ ? (GOS1580)→
env-260	JCVI_SCAF_1096626140998	-	774	720	**RNA**→ ? (GOS1580)→
env-261	JCVI_SCAF_1096627550492	+	666	720	**RNA**→ ? (GOS1580)→
env-262	JCVI_SCAF_1096627559086	+	393	449	**RNA**→ ? (GOS11025)→
env-263	JCVI_SCAF_1096627593736	-	1896	1838	**RNA**→ ? (GOS11025)→
env-264	JCVI_SCAF_1101667128598	+	660	716	**RNA**→

env-265	JCVI_SCAF_1101668182812	+	388	443	**RNA**→ ? (GOS11025)→ ? (GOS13219520)→
env-266	JCVI_SCAF_1101668277221	-	595	540	**RNA**→ ? (GOS11025)→
env-267	JCVI_SCAF_1096627584573	+	967	1022	**RNA**→ ? (GOS11025)→
env-268	JCVI_SCAF_1096626977542	+	1102	1156	**RNA**→ ? (GOS1580)→
env-269	JCVI_SCAF_1101667113641	+	738	793	**RNA**→
env-270	JCVI_SCAF_1101667122372	+	396	451	**RNA**→
env-271	JCVI_SCAF_1096626115537	+	791	846	**RNA**→ ? (GOS1580)→
env-272	JCVI_SCAF_1096627577636	+	274	331	**RNA**→ ? (GOS11025)→
env-273	JCVI_SCAF_1096627772361	-	1045	991	**RNA**→ ? (GOS1580)→
env-274	JCVI_SCAF_1096628182170	-	501	447	**RNA**→ ? (GOS1508)→
env-275	JCVI_SCAF_1096626959620	+	1265	1319	**RNA**→
env-276	JCVI_SCAF_1101668192782	-	1395	1340	**RNA**→ ? (GOS1580)→
env-277	JCVI_SCAF_1101667149976	-	478	424	**RNA**→
env-278	JCVI_SCAF_1096627601770	+	648	702	**RNA**→ ? (GOS1580)→
env-279	JCVI_SCAF_1101668186173	+	71	126	**RNA**→ ? (GOS1580)→
env-280	JCVI_SCAF_1096626973886	+	1013	1068	**RNA**→
env-281	JCVI_SCAF_1101667556189	+	683	738	**RNA**→
env-282	JCVI_SCAF_1096626977926	-	509	454	**RNA**→
env-283	JCVI_SCAF_1096626939382	-	1522	1467	**RNA**→ ? (GOS1580)→
env-284	JCVI_SCAF_1096626934751	-	1614	1559	**RNA**→ ? (GOS1580)→
env-285	JCVI_SCAF_1101668001942	+	562	614	**RNA**→ ? (GOS1580)→
env-286	JCVI_SCAF_1096628024144	-	569	515	**RNA**→
env-287	JCVI_SCAF_1096626888574	-	355	301	**RNA**→
env-288	JCVI_SCAF_1096627591618	-	604	550	**RNA**→ ? (GOS1580)→
env-289	JCVI_SCAF_1096628067998	-	826	764	**RNA**→ ←? (GOS15045)
env-290	JCVI_SCAF_1096627698342	+	462	519	**RNA**→ ? (GOS13281280)→
env-291	JCVI_SCAF_1096627057876	+	1518	1576	**RNA**→ ? (GOS13281280)→
env-292	JCVI_SCAF_1096626251546	-	449	392	**RNA**→ ? (GOS13281280)→
env-293	JCVI_SCAF_1096627041152	+	494	551	**RNA**→ ? (GOS13281280)→
env-294	JCVI_SCAF_1101667132427	-	162	105	**RNA**→
env-295	JCVI_SCAF_1101667219126	+	600	658	**RNA**→ ? (GOS13281280)→
env-296	JCVI_SCAF_1101667224811	+	389	447	**RNA**→ ? (GOS13281280)→
Pma-7-2	NC_008817.1	-	426597	426540	**RNA**→ UbiE (COG2226)→ CsdB (COG0520)→
env-297	JCVI_SCAF_1096627851635	-	128	71	**RNA**→
env-298	JCVI_SCAF_1096627810874	+	577	632	**RNA**→ ? (GOS13125777)→ ? (GOS5892)→ CsdB (COG0520)? (GOS1308)→
env-299	JCVI_SCAF_1096628000660	-	1690	1635	**RNA**→ ? (GOS11225560)→ ? (GOS5892)→ ? (GOS5892)→ CsdB (COG0520)? (GOS1308)→
env-300	JCVI_SCAF_1101668324310	-	358	303	**RNA**→ ? (GOS11225560)→
env-301	JCVI_SCAF_1096627845583	-	1519	1465	**RNA**→ ? (GOS11225560)→
env-302	JCVI_SCAF_1101668735551	+	423	478	**RNA**→ UbiE (COG2226)? (GOS5892)→ ? (GOS1308)→
env-303	JCVI_SCAF_1096627815181	+	1183	1238	**RNA**→ CobL (COG2242)? (GOS5892)→
env-304	JCVI_SCAF_1096627787819	-	885	830	**RNA**→ UbiE (COG2226)? (GOS5892)→
env-305	JCVI_SCAF_1096627322195	-	2062	2004	**RNA**→ ? (GOS13125777)→ UbiE (COG2226)? (GOS5892)→ CsdB (COG0520)? (GOS1308)→
env-306	JCVI_SCAF_1096627628617	+	1392	1449	**RNA**→
[2,6]Pma-11-3	NC_005072.1	-	395850	395793	**RNA**→ UbiE (COG2226)NOG07229 (NOG07229)→ CsdB (COG0520)→
env-307	JCVI_SCAF_1096627746746	-	1226	1169	**RNA**→ ? (GOS13125777)→
env-308	JCVI_SCAF_1101667374010	+	685	742	**RNA**→
env-309	JCVI_SCAF_1096628286913	-	206	149	**RNA**→
env-310	JCVI_SCAF_1096628074631	-	1596	1540	**RNA**→ ? (GOS13125777)→ ? (GOS5892)→
env-311	JCVI_SCAF_1096627543797	+	1591	1648	**RNA**→
env-312	JCVI_SCAF_1096627617081	-	1416	1359	**RNA**→ ? (GOS13125777)→ ? (GOS5892)→ ? (GOS5892)→ ? (GOS1308)→
env-313	JCVI_SCAF_1096627764455	+	1436	1493	**RNA**→ ? (GOS13125777)→
env-314	JCVI_SCAF_1096626942351	-	762	705	**RNA**→ ? (GOS13125777)→ CobL (COG2242)? (GOS5892)→

env-315	JCVI_SCAF_1096627060472	-	1610	1553	**RNA**→ ? (GOS13125777)→ ? (GOS5892)→ UbiE (COG2226)? (GOS5892)→ ? (GOS1308)→
env-316	JCVI_SCAF_1101667543856	+	41	98	**RNA**→ ? (GOS13125777)→ CobL (COG2242)? (GOS5892)→
env-317	JCVI_SCAF_1096626988324	+	373	430	**RNA**→ ? (GOS13125777)→ UbiE (COG2226)? (GOS5892)→
env-318	JCVI_SCAF_1101668295396	-	608	551	**RNA**→ ? (GOS13125777)→ ? (GOS5892)→
env-319	JCVI_SCAF_1101668318893	-	252	195	**RNA**→ ? (GOS13125777)→
env-320	JCVI_SCAF_1096627538416	-	1451	1395	**RNA**→ ? (GOS13125777)→ ? (GOS5892)→
env-321	JCVI_SCAF_1101667529909	-	251	195	**RNA**→ ? (GOS11225560)→
env-322	JCVI_SCAF_1096627542196	+	1219	1275	**RNA**→ ? (GOS13125777)→ ? (GOS5892)→
env-323	JCVI_SCAF_1101667086255	-	329	273	**RNA**→ ? (GOS13125777)→
env-324	JCVI_SCAF_1101668213000	+	602	658	**RNA**→ ? (GOS13125777)→ CobL (COG2242)? (GOS5892)→
Pma-5-3	NC_007577.1	-	391792	391735	**RNA**→ hypo→ UbiE (COG2226)NOG07229 (NOG07229)→ CsdB (COG0520)→
env-325	JCVI_SCAF_1096628399786	+	774	830	**RNA**→ UbiE (COG2226)? (GOS5892)→
env-326	JCVI_SCAF_1096627551868	-	192	135	**RNA**→
env-327	JCVI_SCAF_1101668305582	+	603	660	**RNA**→ CobL (COG2242)? (GOS5892)→
env-328	JCVI_SCAF_1101668315801	-	885	828	**RNA**→ ? (GOS11449188)→ UbiE (COG2226)? (GOS5892)→
Pma-1-3	NC_008816.1	-	407110	407052	**RNA**→ hypo→ UbiE (COG2226)→ CsdB (COG0520)→
env-329	JCVI_SCAF_1096627323418	+	1265	1322	**RNA**→ ? (GOS13125777)→ UbiE (COG2226)? (GOS5892)→
env-330	JCVI_SCAF_1096628249556	+	1507	1564	**RNA**→ ? (GOS11225560)→
env-331	JCVI_SCAF_1096626984393	+	1568	1625	**RNA**→ ? (GOS13125777)→ UbiE (COG2226)? (GOS5892)→
env-332	JCVI_SCAF_1096627323349	-	635	578	**RNA**→ ? (GOS13125777)→ ? (GOS5892)→
env-333	JCVI_SCAF_1096627759764	-	1000	943	**RNA**→ UbiE (COG2226)? (GOS5892)→ ? (GOS1308)→
env-334	JCVI_SCAF_1096628046814	+	1471	1528	**RNA**→
env-335	JCVI_SCAF_1096626133318	-	181	124	**RNA**→
env-336	JCVI_SCAF_1101667090484	-	938	882	**RNA**→ ? (GOS13125777)→ UbiE (COG2226)? (GOS5892)→
env-337	JCVI_SCAF_1101667091360	+	789	846	**RNA**→
env-338	JCVI_SCAF_1096627621948	+	693	750	**RNA**→ UbiE (COG2226)? (GOS5892)→
env-339	JCVI_SCAF_1096627743957	+	1022	1079	**RNA**→ ? (GOS13125777)→ SmtA (COG0500)? (GOS5892)→
env-340	JCVI_SCAF_1096627966571	+	841	898	**RNA**→ ? (GOS13125777)→ CobL (COG2242)? (GOS5892)→
env-341	JCVI_SCAF_1096627521107	+	271	328	**RNA**→ ? (GOS13125777)→
env-342	JCVI_SCAF_1096627602717	-	1368	1311	**RNA**→ ? (GOS13125777)→
env-343	JCVI_SCAF_1096627719175	+	571	628	**RNA**→ ? (GOS13125777)→
env-344	JCVI_SCAF_1096627779389	-	140	83	**RNA**→
env-345	JCVI_SCAF_1096628228469	+	1739	1796	**RNA**→
env-346	JCVI_SCAF_1096628398538	+	244	301	**RNA**→ ? (GOS13125777)→
env-347	JCVI_SCAF_1096628400286	+	1587	1644	**RNA**→
env-348	JCVI_SCAF_1101668191233	-	1059	1002	**RNA**→ ? (GOS13125777)→ SmtA (COG0500)? (GOS5892)→
env-349	JCVI_SCAF_1096626977265	-	1271	1214	**RNA**→ ? (GOS13125777)→ UbiE (COG2226)? (GOS5892)→
env-350	JCVI_SCAF_1096628058388	+	316	373	**RNA**→ ? (GOS13125777)→ ? (GOS5892)→
env-351	JCVI_SCAF_1101667158716	-	203	146	**RNA**→
env-352	JCVI_SCAF_1101667151009	-	827	770	**RNA**→ ? (GOS13125777)→ CobL (COG2242)? (GOS5892)→
env-353	JCVI_SCAF_1101667150996	+	569	626	**RNA**→ ? (GOS13125777)→
env-354	JCVI_SCAF_1101667119113	+	97	154	**RNA**→ ? (GOS13125777)→ ? (GOS5892)→
env-355	JCVI_SCAF_1096627626877	+	807	864	**RNA**→ ? (GOS13125777)→ CobL (COG2242)? (GOS5892)→
env-356	JCVI_SCAF_1101667221498	+	553	610	**RNA**→ ? (GOS13125777)→
env-357	JCVI_SCAF_1101667392158	-	821	764	**RNA**→ ? (GOS13125777)→ GCD14 (COG2519)? (GOS5892)→
env-358	JCVI_SCAF_1101667563334	+	590	647	**RNA**→ ? (GOS13125777)→
env-359	JCVI_SCAF_1101668765935	-	325	268	**RNA**→ ? (GOS13125777)→
env-360	JCVI_SCAF_1096627054461	-	1965	1908	**RNA**→ ? (GOS13125777)→ UbiE (COG2226)? (GOS5892)→ CsdB (COG0520)? (GOS1308)→
Pma-3-4	NC_009091.1	-	381532	381475	**RNA**→ hypo→ UbiE (COG2226)→ CsdB (COG0520)→
env-361	JCVI_SCAF_1096627594731	-	1262	1205	**RNA**→ ? (GOS5892)→ CsdB (COG0520)? (GOS1308)→
env-362	JCVI_SCAF_1101667154645	-	181	124	**RNA**→
env-363	JCVI_SCAF_1101668188028	+	572	629	**RNA**→ ? (GOS13125777)→ UbiE (COG2226)? (GOS5892)→

env-364	JCVI_SCAF_1096628400013	-	511	454	**RNA**→ ? (GOS13125777)→ ? (GOS5892)→
env-365	JCVI_SCAF_1096628077091	-	1662	1605	**RNA**→ ? (GOS13125777)→ GCD14 (COG2519)? (GOS5892)→ ? (GOS5892)→ CsdB (COG0520)? (GOS1308)→
env-366	JCVI_SCAF_1096628282187	+	138	195	**RNA**→ ? (GOS13125777)→ ? (GOS5892)→
env-367	JCVI_SCAF_1101668764647	-	326	269	**RNA**→ ? (GOS13125777)→
env-368	JCVI_SCAF_1096627759474	+	1119	1176	**RNA**→ ? (GOS13125777)→ ? (GOS5892)→
env-369	JCVI_SCAF_1096627738470	-	1097	1040	**RNA**→ ? (GOS13125777)→
env-370	JCVI_SCAF_1096628063727	+	1824	1881	**RNA**→
env-371	JCVI_SCAF_1096627849186	-	1416	1359	**RNA**→ ? (GOS13125777)→ CobL (COG2242)? (GOS5892)→
env-372	JCVI_SCAF_1096628051169	+	1230	1287	**RNA**→ ? (GOS13125777)→ ? (GOS5892)→
env-373	JCVI_SCAF_1096628011481	-	607	548	**RNA**→ ? (GOS11176343)→
Ssp-6-3	NZ_AAUA01000001.1	-	589598	589524	**RNA**→ hypo→
Pma-2-1	NZ_AALP01000001.1	+	966992	967054	**RNA**→ hypo→
env-374	JCVI_SCAF_1096628337442	-	578	516	**RNA**→
env-375	JCVI_SCAF_1096627354876	-	1349	1286	**RNA**→ ? (GOS11043201)→
env-376	JCVI_SCAF_1096627139434	-	1393	1330	**RNA**→ ? (GOS11043201)→
env-377	JCVI_SCAF_1096627358842	+	843	906	**RNA**→ ←? (GOS11226869)
env-378	JCVI_SCAF_1096627143055	+	268	331	**RNA**→ ←? (GOS11226869)
env-379	JCVI_SCAF_1096627946152	-	495	432	**RNA**→ ? (GOS11043201)→
env-380	JCVI_SCAF_1096627136584	-	462	399	**RNA**→ ? (GOS11043201)→
env-381	JCVI_SCAF_1096627142281	+	1556	1619	**RNA**→ ←? (GOS11226869)
env-382	JCVI_SCAF_1101668628745	-	1146	1083	**RNA**→ ←? (GOS11226869)
env-383	JCVI_SCAF_1096627935886	-	92	29	**RNA**→
env-384	JCVI_SCAF_1101667323853	-	831	768	**RNA**→ ? (GOS11043201)→
env-385	JCVI_SCAF_1101668642528	-	730	667	**RNA**→ ←? (GOS11226869)
env-386	JCVI_SCAF_1096627951343	-	490	427	**RNA**→ ? (GOS11043201)→
env-387	JCVI_SCAF_1096627146162	+	159	222	**RNA**→ ←? (GOS11226869)
env-388	JCVI_SCAF_1101668607896	-	1362	1299	**RNA**→ ? (GOS11043201)→
env-389	JCVI_SCAF_1096627137105	+	1067	1130	**RNA**→ ←? (GOS11226869)
env-390	JCVI_SCAF_1096627938221	+	1343	1406	**RNA**→ ? (GOS11043201)→
env-391	JCVI_SCAF_1101668623266	+	952	1015	**RNA**→ ? (GOS11043201)→
Ssp-8-2	NZ_AANO01000069.1	+	694	756	**RNA**→
Ssp-1-5	NZ_AATZ01000002.1	-	423708	423646	**RNA**→ hypo→
Ssp-4-4	NC_007513.1	+	2177551	2177613	**RNA**→ hypo→
Pma-4-2	NC_008820.1	+	1495940	1496001	**RNA**→ hypo→
Pma-6-2	NC_005071.1	-	599397	599336	**RNA**→ hypo→
env-392	JCVI_SCAF_1101668169125	-	1167	1108	**RNA**→ ? (GOS11970823)→
env-393	JCVI_SCAF_1096626988366	-	1439	1380	**RNA**→ ? (GOS11970823)→
env-394	JCVI_SCAF_1096628103692	+	94	153	**RNA**→ ? (GOS11970823)→
env-395	JCVI_SCAF_1096626949042	-	520	461	**RNA**→ ? (GOS11970823)→
env-396	JCVI_SCAF_1101668233196	-	76	17	**RNA**→
env-397	JCVI_SCAF_1101667210224	-	608	549	**RNA**→ ? (GOS11970823)→
env-398	JCVI_SCAF_1096627214220	-	1284	1225	**RNA**→ ? (GOS11970823)→
env-399	JCVI_SCAF_1096627519589	-	1214	1155	**RNA**→ ? (GOS11970823)→
env-400	JCVI_SCAF_1096627726654	-	909	848	**RNA**→ ? (GOS11991820)→
env-401	JCVI_SCAF_1096628057766	-	236	176	**RNA**→
env-402	JCVI_SCAF_1096628004134	+	152	214	**RNA**→ ←TauB (COG1116)? (GOS2568)
env-403	JCVI_SCAF_1096626928045	+	901	962	**RNA**→
env-404	JCVI_SCAF_1096628060017	-	637	576	**RNA**→
env-405	JCVI_SCAF_1096627562639	-	1428	1367	**RNA**→ ? (GOS11635207)→
env-406	JCVI_SCAF_1096627565380	+	1213	1274	**RNA**→ ? (GOS5157)→
env-407	JCVI_SCAF_1096627729032	+	1113	1174	**RNA**→
env-408	JCVI_SCAF_1096627775759	+	379	440	**RNA**→

env-409	JCVI_SCAF_1096627549558	-	1346	1285	**RNA**→ ? (GOS10476)→
env-410	JCVI_SCAF_1096627581416	+	807	868	**RNA**→ ←? (GOS5767)
env-411	JCVI_SCAF_1101667227512	-	384	323	**RNA**→
env-412	JCVI_SCAF_1096627721094	+	18	79	**RNA**→ ←? (GOS5767)
env-413	JCVI_SCAF_1096627569833	+	1006	1067	**RNA**→ ? (GOS13254074)→ ? (GOS13612)→
env-414	JCVI_SCAF_1096627778388	-	1620	1559	**RNA**→
env-415	JCVI_SCAF_1096626164536	+	900	961	**RNA**→
env-416	JCVI_SCAF_1096627555107	-	1365	1304	**RNA**→ ←? (GOS13273817)
env-417	JCVI_SCAF_1096628205761	-	184	123	**RNA**→
Pma-5-4	NC_007577.1	-	621613	621552	**RNA**→ tRNA-Ser→
env-418	JCVI_SCAF_1101668188753	-	1332	1271	**RNA**→ ? (GOS10476)→ ? (GOS10476)→ Ser tRNA→
env-419	JCVI_SCAF_1096626925004	+	235	296	**RNA**→
env-420	JCVI_SCAF_1096626967522	+	1418	1479	**RNA**→ ←? (GOS13612)
env-421	JCVI_SCAF_1101668306349	-	432	371	**RNA**→
env-422	JCVI_SCAF_1101668213441	-	129	68	**RNA**→
env-423	JCVI_SCAF_1096627631311	+	568	629	**RNA**→ ? (GOS18783)→
env-424	JCVI_SCAF_1096627736286	-	871	810	**RNA**→
env-425	JCVI_SCAF_1096626993894	+	1872	1933	**RNA**→
env-426	JCVI_SCAF_1096627576850	+	1593	1654	**RNA**→
env-427	JCVI_SCAF_1096627730556	-	1435	1374	**RNA**→ ←? (GOS11437910)
env-428	JCVI_SCAF_1096626636754	-	1456	1395	**RNA**→ ←? (GOS14428175)
env-429	JCVI_SCAF_1096627729564	+	270	331	**RNA**→ ? (GOS12061403)→
env-430	JCVI_SCAF_1101667411987	-	181	120	**RNA**→
env-431	JCVI_SCAF_1096627522154	+	409	470	**RNA**→ ←? (GOS7959)
env-432	JCVI_SCAF_1096627612714	+	733	794	**RNA**→
env-433	JCVI_SCAF_1096626988330	+	1245	1306	**RNA**→
Pma-3-5	NC_009091.1	-	632920	632859	**RNA**→ hypo→
env-434	JCVI_SCAF_1101667228755	+	413	474	**RNA**→
env-435	JCVI_SCAF_1096626937007	-	1463	1402	**RNA**→ Ser tRNA→
env-436	JCVI_SCAF_1096627304666	+	1756	1817	**RNA**→
env-437	JCVI_SCAF_1096627725111	-	105	44	**RNA**→
env-438	JCVI_SCAF_1101668281928	+	21	82	**RNA**→ Ser tRNA→ ? (GOS16289)→ ? (GOS10000)→
env-439	JCVI_SCAF_1096627193686	-	944	883	**RNA**→ Ser tRNA→
env-440	JCVI_SCAF_1096627584545	-	1185	1124	**RNA**→ ←? (GOS11511851)
env-441	JCVI_SCAF_1101668220586	-	360	299	**RNA**→
env-442	JCVI_SCAF_1096627053672	-	741	680	**RNA**→
Pma-1-4	NC_008816.1	-	634086	634025	**RNA**→ hypo→
env-443	JCVI_SCAF_1096628223486	+	833	894	**RNA**→ ? (GOS16289)→ ? (GOS10000)→
env-444	JCVI_SCAF_1096627004245	+	317	378	**RNA**→ Ser tRNA→ ? (GOS16289)→ ? (GOS10000)→ ? (GOS14004)→
env-445	JCVI_SCAF_1096627600056	-	242	181	**RNA**→
env-446	JCVI_SCAF_1096626435080	+	687	748	**RNA**→
env-447	JCVI_SCAF_1096627610126	+	168	229	**RNA**→ ? (GOS13254074)→
env-448	JCVI_SCAF_1096628404081	+	977	1038	**RNA**→
env-449	JCVI_SCAF_1101668228128	+	27	88	**RNA**→ Ser tRNA→ ? (GOS16289)→ ? (GOS10000)→
env-450	JCVI_SCAF_1096627960584	-	130	69	**RNA**→
env-451	JCVI_SCAF_1096626254939	+	48	109	**RNA**→
env-452	JCVI_SCAF_1101668170234	-	949	889	**RNA**→ ←? (GOS5767)
env-453	JCVI_SCAF_1096627589047	-	1816	1756	**RNA**→ ←DiS_P_DiS (pfam06750)? (GOS5767)
env-454	JCVI_SCAF_1096626954136	+	763	824	**RNA**→ Ser tRNA→ ? (GOS16289)→ ? (GOS10000)→ ? (GOS14004)→
env-455	JCVI_SCAF_1096627562283	-	1254	1193	**RNA**→ ? (GOS10682)→
env-456	JCVI_SCAF_1096627626922	-	1405	1344	**RNA**→ ? (GOS13534)→
env-457	JCVI_SCAF_1096626956049	-	1475	1414	**RNA**→ ? (GOS13236868)→

	env-458	JCVI_SCAF_1101668310396	-	208	147	**RNA**→
[2,4]	Pma-11-4	NC_005072.1	-	627956	627894	**RNA**→ ←LIGANc (cd00114)
	env-459	JCVI_SCAF_1096627268048	+	293	355	**RNA**→
	Pma-7-3	NC_008817.1	-	662192	662132	**RNA**→ ←hypo
	env-460	JCVI_SCAF_1096628245343	+	241	303	**RNA**→ ←? (GOS11677616)
	Pma-9-4	NC_007335.1	-	484696	484634	**RNA**→ ←protein family PM-14
	env-461	JCVI_SCAF_1096628298282	+	23	85	**RNA**→ ←? (GOS11677616)
	Pma-8-4	NC_008819.1	-	1087010	1086948	**RNA**→ ←hypo
	Pma-8-5	NC_008819.1	+	1573052	1573109	**RNA**→ ←hypo
	Pma-9-5	NC_007335.1	+	986065	986122	**RNA**→ hypo→
[2,5]	Pma-10-1	NC_005042.1	-	924002	923944	**RNA**→ ←hypo
	env-462	JCVI_SCAF_1096628101613	+	1775	1837	**RNA**→
	env-463	JCVI_SCAF_1096628087621	-	277	215	**RNA**→
	Pma-8-6	NC_008819.1	-	397281	397219	**RNA**→ ←COG0398 (COG0398)
	Pma-9-6	NC_007335.1	-	1674805	1674743	**RNA**→ ←COG0398 (COG0398)
	env-464	JCVI_SCAF_1101668656185	+	767	826	**RNA**→ ? (GOS11047726)→
	env-465	JCVI_SCAF_1096627971008	-	669	610	**RNA**→
	env-466	JCVI_SCAF_1101668663213	+	42	101	**RNA**→ ? (GOS11047726)→
	env-467	JCVI_SCAF_1096627369518	-	700	640	**RNA**→ ? (GOS11047726)→
	env-468	JCVI_SCAF_1101668369421	-	560	501	**RNA**→ ? (GOS11047726)→
	env-469	JCVI_SCAF_1101667331091	-	553	494	**RNA**→ ? (GOS11047726)→
	env-470	JCVI_SCAF_1101668677373	+	758	817	**RNA**→ ? (GOS11047726)→
	env-471	JCVI_SCAF_1101668266469	+	417	476	**RNA**→ ? (GOS11047726)→
	env-472	JCVI_SCAF_1096627963640	-	655	596	**RNA**→ ? (GOS11047726)→
	env-473	JCVI_SCAF_1101668668358	-	1140	1081	**RNA**→ ? (GOS11047726)→
	env-474	JCVI_SCAF_1101668273946	-	966	907	**RNA**→ ? (GOS11047726)→
	env-475	JCVI_SCAF_1096626985270	+	1760	1820	**RNA**→ ? (GOS11047726)→
	env-476	JCVI_SCAF_1096628041581	-	1502	1443	**RNA**→ ? (GOS11047726)→
	env-477	JCVI_SCAF_1096627574193	-	760	702	**RNA**→ ? (GOS11047726)→
	env-478	JCVI_SCAF_1096627058373	-	1834	1776	**RNA**→ ? (GOS11047726)→ ? (GOS12146751)→
	env-479	JCVI_SCAF_1096627153363	-	368	310	**RNA**→ ? (GOS11047726)→
	env-480	JCVI_SCAF_1101668660040	+	1236	1294	**RNA**→ ? (GOS11047726)→
	env-481	JCVI_SCAF_1096628075702	-	1123	1065	**RNA**→ ? (GOS11047726)→
	Ssp-11-3	NC_005070.1	+	1128117	1128176	**RNA**→ hypo→
	Ssp-3-3	NC_007516.1	-	1932912	1932853	**RNA**→ hypo→ hypo→
	env-482	JCVI_SCAF_1101668366051	-	768	707	**RNA**→
	Ssp-3-4	NC_007516.1	+	518912	518973	**RNA**→ hypo→ hypo→ hypo→
	env-483	JCVI_SCAF_1096626920899	-	2281	2220	**RNA**→ ? (GOS11200898)→ ? (GOS14022)→
	env-484	JCVI_SCAF_1101668668457	+	98	159	**RNA**→ ? (GOS11200898)→ ? (GOS14022)→
	env-485	JCVI_SCAF_1096627764109	-	641	580	**RNA**→
	env-486	JCVI_SCAF_1096628317283	+	733	794	**RNA**→
	Ssp-11-4	NC_005070.1	+	1376352	1376413	**RNA**→ UPF0061 (pfam02696)→ hypo→ DUF924 (pfam06041)→ hypo→
	env-487	JCVI_SCAF_1096628039489	+	1120	1181	**RNA**→ ? (GOS12056613)→ ? (GOS3431)→
	Ssp-11-5	NC_005070.1	+	1854939	1854998	**RNA**→ hypo→ hypo→
	env-488	JCVI_SCAF_1101667376581	-	793	734	**RNA**→ ? (GOS11047726)→ ? (GOS13298551)→
	env-489	JCVI_SCAF_1101668217734	+	1196	1255	**RNA**→ ? (GOS11047726)→ ? (GOS13298551)→
	Ssp-2-2	NC_008319.1	-	1642629	1642569	**RNA**→ hypo→
	env-490	JCVI_SCAF_1096627752841	-	465	407	**RNA**→
	env-491	JCVI_SCAF_1096628071987	-	1245	1186	**RNA**→ ? (GOS11047726)→
	Ssp-1-6	NZ_AATZ01000001.1	+	222184	222243	**RNA**→ hypo→
	Ssp-4-5	NC_007513.1	+	1170507	1170566	**RNA**→ hypo→
	env-492	JCVI_SCAF_1096628069533	+	1036	1096	**RNA**→ ? (GOS11971539)→

Ssp-11-6	NC_005070.1	-	376374	376314	**RNA**→ ←hypo
env-493	JCVI_SCAF_1096627183482	-	796	738	**RNA**→
Ssp-2-3	NC_008319.1	+	778927	778986	**RNA**→ ←axoneme-associated protein gasp-180
Ssp-4-6	NC_007513.1	+	1411351	1411410	**RNA**→ hypo→
Ssp-1-7	NZ_AATZ01000003.1	-	202513	202454	**RNA**→ ←hypo
env-494	JCVI_SCAF_1096628057607	-	1455	1397	**RNA**→ ←COG0313 (COG0313)? (GOS1209)
env-495	JCVI_SCAF_1096627155165	-	781	723	**RNA**→ ? (GOS13063693)→
env-496	JCVI_SCAF_1101668666528	+	1129	1187	**RNA**→
env-497	JCVI_SCAF_1096627154348	-	1415	1357	**RNA**→ ←Soj (COG1192)? (GOS1716)
env-498	JCVI_SCAF_1096627371490	+	324	382	**RNA**→ ? (GOS11249006)→
env-499	JCVI_SCAF_1101667380837	+	844	903	**RNA**→
env-500	JCVI_SCAF_1096627807443	+	925	986	**RNA**→ **RNA**→ ←? (GOS11480162)
env-502	JCVI_SCAF_1096627807443	+	1157	1218	(shown above)
env-501	JCVI_SCAF_1096627155081	+	1466	1527	**RNA**→ ←? (GOS11480162)
env-503	JCVI_SCAF_1101667137087	-	904	840	**RNA**→ ? (GOS11047726)→
env-504	JCVI_SCAF_1096627189592	-	1112	1048	**RNA**→ ? (GOS11047726)→
env-505	JCVI_SCAF_1101667386446	+	129	193	**RNA**→ ? (GOS11047726)→
Ssp-11-7	NC_005070.1	+	2354202	2354266	**RNA**→ hypo→
env-506	JCVI_SCAF_1096627206592	-	841	777	**RNA**→ ? (GOS11047726)→
env-507	JCVI_SCAF_1096627618798	+	1297	1360	**RNA**→ ? (GOS11047726)→
env-508	JCVI_SCAF_1096627846736	+	127	189	**RNA**→ ←? (GOS10219)
[3]env-509	JCVI_SCAF_1101668291845	+	816	877	**RNA**→ ←DnaJ (smart00271)? (GOS11309686)
env-510	JCVI_SCAF_1096627083446	-	334	273	**RNA**→
env-511	JCVI_SCAF_1096627066685	-	1513	1452	**RNA**→ ←DnaJ (smart00271)? (GOS11309686)
env-512	JCVI_SCAF_1096627091462	+	1590	1651	**RNA**→
env-513	JCVI_SCAF_1096627817532	+	996	1057	**RNA**→
env-514	JCVI_SCAF_1096627822828	-	1285	1224	**RNA**→ ←? (GOS11831288)
Ssp-11-8	NC_005070.1	+	1382950	1383011	**RNA**→ ←hypo
Ssp-10-2	NZ_AAOK01000003.1	-	2958	2897	**RNA**→ hypo→ hypo→
[1]env-515	JCVI_SCAF_1096628382219	-	972	911	**RNA**→ SpeB (COG0010)? (GOS4555)→
env-516	JCVI_SCAF_1096627201917	+	1869	1930	**RNA**→ ? (GOS14201702)→
env-517	JCVI_SCAF_1096627303926	+	1850	1911	**RNA**→ ←DnaJ (smart00271)? (GOS11309686)
env-518	JCVI_SCAF_1096627819941	-	242	181	**RNA**→
env-519	JCVI_SCAF_1096627523909	+	343	404	**RNA**→ ←? (GOS11309686)
Ssp-3-5	NC_007516.1	-	1025849	1025788	**RNA**→ hypo→
env-520	JCVI_SCAF_1101668663730	+	52	113	**RNA**→ ←DnaJ (smart00271)? (GOS11309686)
env-521	JCVI_SCAF_1096627371929	-	538	477	**RNA**→
Ssp-6-4	NZ_AAUA01000003.1	-	162424	162364	**RNA**→ hypo→
Ssp-5-3	NC_009482.1	-	1727489	1727429	**RNA**→ hypo→ hypo→
env-522	JCVI_SCAF_1096627813562	-	1541	1482	**RNA**→ ? (GOS13240897)→
env-523	JCVI_SCAF_1096627161547	-	1303	1243	**RNA**→ ←? (GOS13513552)
env-524	JCVI_SCAF_1096627384016	+	3324	3384	**RNA**→
env-525	JCVI_SCAF_1096627799922	+	190	250	**RNA**→ ? (GOS13240897)→
env-526	JCVI_SCAF_1096627784172	+	1449	1509	**RNA**→
env-527	JCVI_SCAF_1096627371492	+	725	785	**RNA**→ SNase (pfam00565)? (GOS14580720)→ ? (GOS13240897)→
env-528	JCVI_SCAF_1101668733680	+	417	477	**RNA**→ ? (GOS11047726)→
Ssp-6-5	NZ_AAUA01000001.1	+	260927	260989	**RNA**→ hypo→
Ssp-6-6	NZ_AAUA01000001.1	+	1150407	1150468	**RNA**→ ←hypo
Ssp-3-6	NC_007516.1	+	2439513	2439574	**RNA**→ hypo→ hypo→
env-529	JCVI_SCAF_1096627985059	-	253	192	**RNA**→
env-530	JCVI_SCAF_1096627813825	-	366	305	**RNA**→
env-531	JCVI_SCAF_1096627156145	+	1025	1086	**RNA**→ ? (GOS11277288)→

env-532	JCVI_SCAF_1101668670099	+	448	509	**RNA**→ ? (GOS11277288)→
env-533	JCVI_SCAF_1101667148414	-	398	337	**RNA**→
env-534	JCVI_SCAF_1096627155618	-	1908	1847	**RNA**→ ? (GOS11277288)→
env-535	JCVI_SCAF_1096627153433	+	399	460	**RNA**→ ? (GOS11277288)→
env-536	JCVI_SCAF_1101668670690	-	763	701	**RNA**→ ? (GOS11669244)→ ? (GOS11277288)→
env-537	JCVI_SCAF_1096628046256	+	1319	1379	**RNA**→ ? (GOS11596502)→
env-538	JCVI_SCAF_1096627762300	+	326	386	**RNA**→ ? (GOS11596502)→
Ssp-3-7	NC_007516.1	+	532787	532847	**RNA**→ ←hypo
env-539	JCVI_SCAF_1101668748982	-	69	9	**RNA**→
env-540	JCVI_SCAF_1096627203457	+	1346	1406	**RNA**→ ? (GOS12056613)→
env-541	JCVI_SCAF_1096628046417	-	208	149	**RNA**→
Ssp-11-9	NC_005070.1	+	379176	379235	**RNA**→ hypo→
Pma-2-2	NZ_AALP01000001.1	+	294776	294836	**RNA**→ hypo→
Ssp-3-8	NC_007516.1	+	738974	739032	**RNA**→ hypo→
env-542	JCVI_SCAF_1101668745194	+	555	612	**RNA**→ ←? (GOS17124)
env-543	JCVI_SCAF_1096627194619	+	706	763	**RNA**→ ←? (GOS17124)
env-544	JCVI_SCAF_1096627199205	+	3151	3208	**RNA**→ ←? (GOS17124)
env-545	JCVI_SCAF_1096628278366	+	868	930	**RNA**→ ←? (GOS11309686)
Pma-10-2	NC_005042.1	-	398283	398223	**RNA**→ hypo→

29.4 Conserved domains

Conserved domains found in protein-coding genes listed in Section 29.3 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Downstream-peptide RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

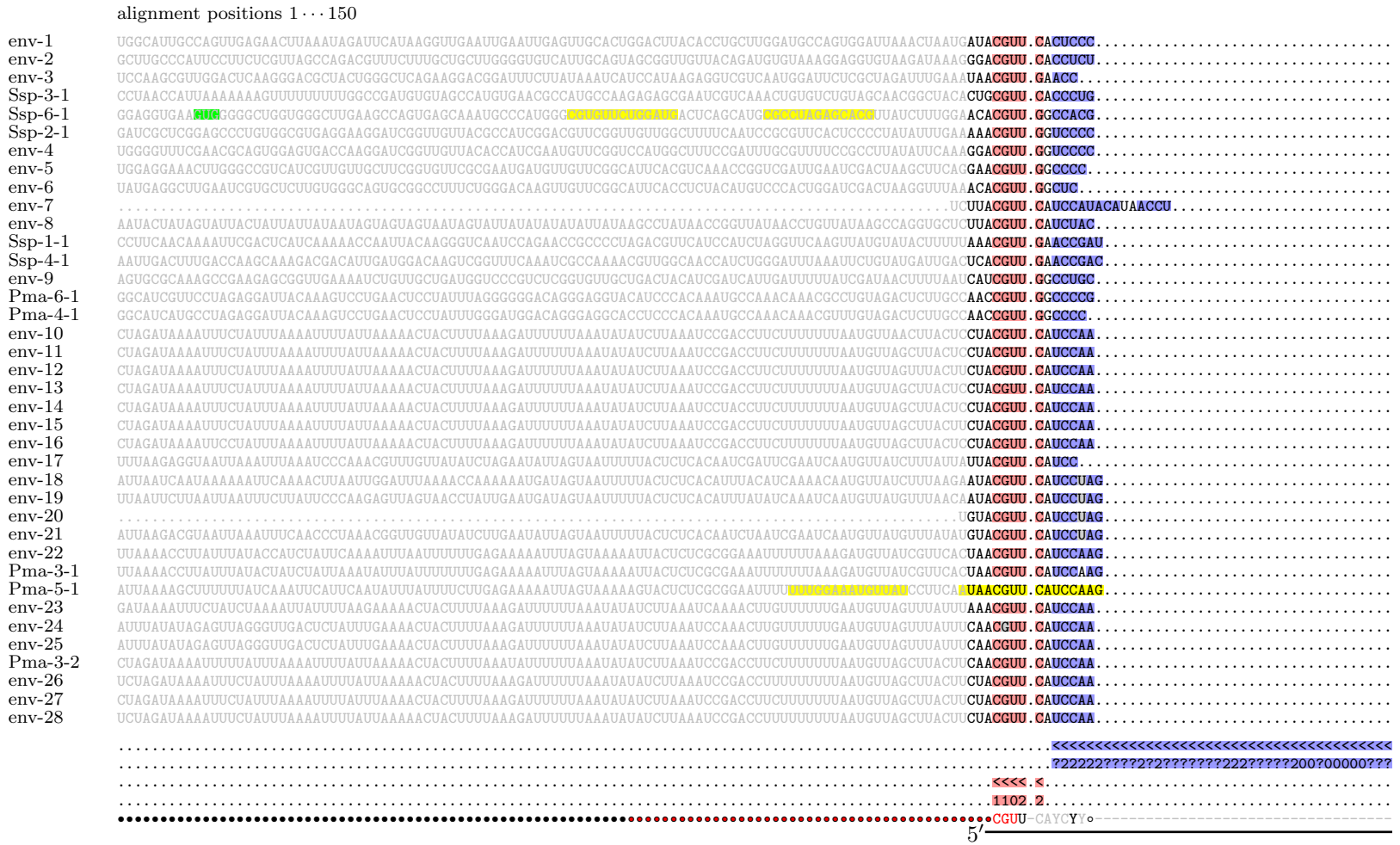
cd00114 (1) NAD+ dependent DNA ligase adenylation domain.
cd02549 (1) A sub-family of peptidase family C39.
cd03256 (1) ABC-type phosphate/phosphonate transport system.
COG0010 (1) Arginase/agmatinase/formimionoglutamate hydrolase, arginase family [Amino acid transport and metabolism]
COG0226 (1) ABC-type phosphate transport system, periplasmic component [Inorganic ion transport and metabolism]
COG0313 (1) Predicted methyltransferases [General function prediction only]
COG0398 (2) Uncharacterized conserved protein [Function unknown]
COG0500 (2) SAM-dependent methyltransferases [Secondary metabolites biosynthesis, transport, and catabolism / General function prediction only]
COG0520 (11) Selenocysteine lyase [Amino acid transport and metabolism]
COG0651 (1) Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit [Energy production and conversion / Inorganic ion transport and metabolism]
COG1006 (1) Multisubunit Na+/H+ antiporter, MnhC subunit [Inorganic ion transport and metabolism]
COG1116 (1) ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component [Inorganic ion transport and metabolism]
COG1192 (1) ATPases involved in chromosome partitioning [Cell division and chromosome partitioning]
COG1563 (1) Predicted subunit of the Multisubunit Na+/H+ antiporter [Inorganic ion transport and metabolism]
COG1680 (3) Beta-lactamase class C and other penicillin binding proteins [Defense mechanisms]
COG1705 (1) Muramidase (flagellum-specific) [Cell motility and secretion / Intracellular trafficking and secretion]
COG1917 (2) Uncharacterized conserved protein, contains double-stranded beta-helix domain [Function unknown]
COG2226 (20) Methylase involved in ubiquinone/menaquinone biosynthesis [Coenzyme metabolism]
COG2242 (9) Precorrin-6B methylase 2 [Coenzyme metabolism]
COG2519 (2) tRNA(1-methyladenosine) methyltransferase and related methyltransferases [Translation, ribosomal structure and biogenesis]
GOS10000 (5) ?
GOS10219 (1) ?
GOS10476 (3) ?
GOS10682 (1) ?
GOS11025 (13) ?
GOS11043201 (9) ?
GOS11044226 (3) ?
GOS11047726 (26) ?
GOS11064110 (7) ?
GOS11176343 (3) ?
GOS11200898 (2) ?
GOS11225560 (5) ?
GOS11226869 (7) ?

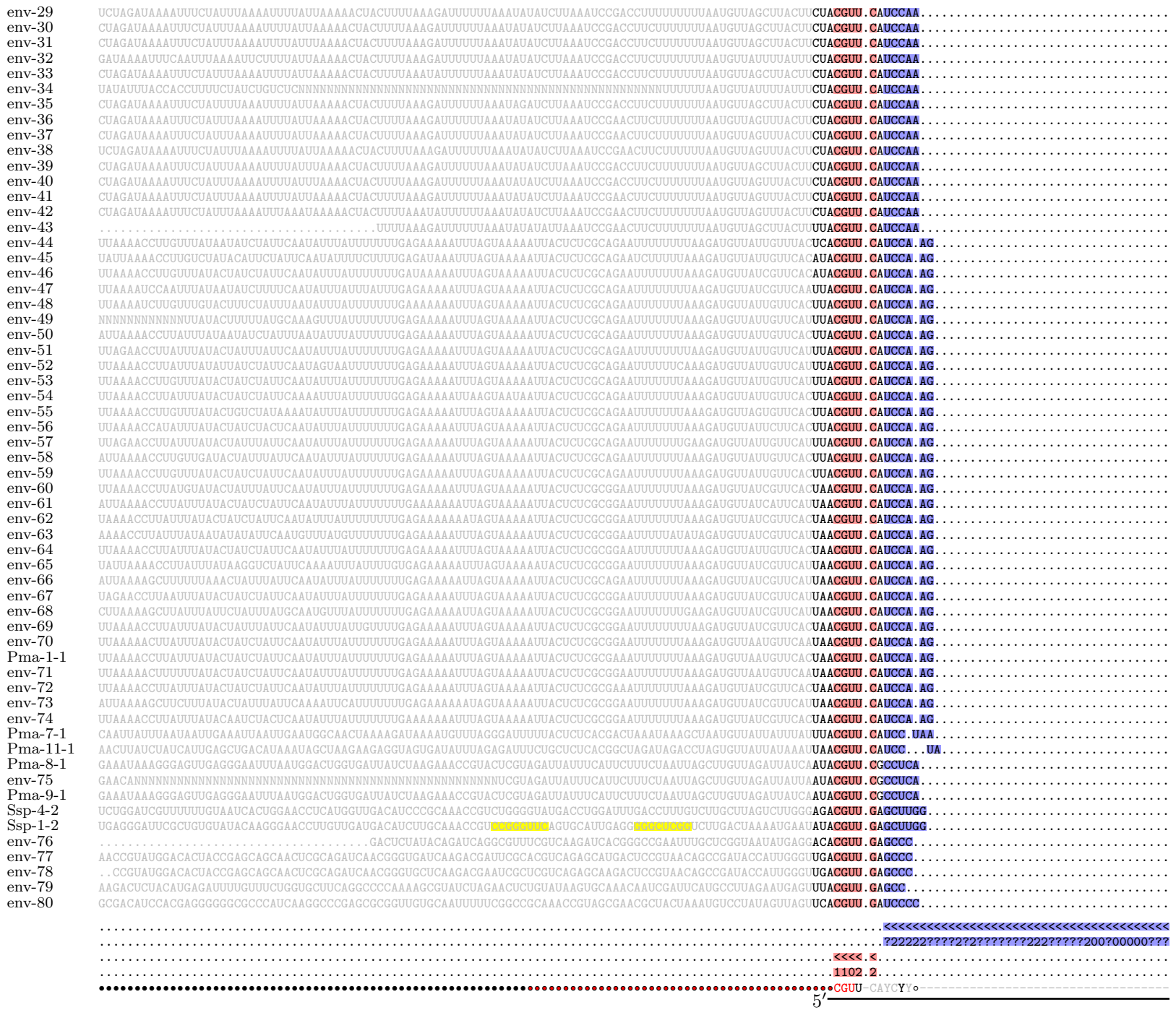
GOS11249006 (1) ?
GOS11277288 (5) ?
GOS11309686 (6) ?
GOS11316857 (2) ?
GOS11437910 (1) ?
GOS11438381 (1) ?
GOS11449188 (1) ?
GOS11480162 (2) ?
GOS11511851 (1) ?
GOS11555940 (2) ?
GOS11596502 (2) ?
GOS11614916 (2) ?
GOS11635207 (1) ?
GOS11669244 (1) ?
GOS11671983 (3) ?
GOS11677616 (2) ?
GOS11831288 (1) ?
GOS11904738 (5) ?
GOS11943545 (1) ?
GOS11953 (1) ?
GOS11970823 (7) ?
GOS11971539 (1) ?
GOS11989004 (1) ?
GOS11991820 (1) ?
GOS12053 (16) ?
GOS12056613 (2) ?
GOS12061403 (1) ?
GOS1209 (1) ?
GOS12135550 (2) ?
GOS12142765 (1) ?
GOS12146751 (1) ?
GOS13006886 (9) ?
GOS1303 (1) ?
GOS13049017 (1) ?
GOS13063693 (1) ?
GOS1308 (10) ?
GOS13125777 (47) ?
GOS13140869 (1) ?
GOS13219520 (2) ?
GOS13233019 (16) ?
GOS13236868 (1) ?
GOS13240897 (3) ?
GOS13254074 (2) ?
GOS13266240 (4) ?
GOS13273817 (1) ?
GOS13281280 (6) ?
GOS13298551 (2) ?
GOS13431674 (1) ?
GOS13513552 (1) ?
GOS13534 (1) ?
GOS13612 (2) ?
GOS13866 (1) ?
GOS14004 (2) ?
GOS14022 (2) ?
GOS14201702 (1) ?
GOS14256 (14) ?
GOS14428175 (1) ?
GOS14580720 (1) ?
GOS14707434 (5) ?
GOS1472 (4) ?
GOS15045 (1) ?
GOS1508 (2) ?
GOS1580 (74) ?
GOS15836 (1) ?
GOS16289 (5) ?
GOS16291 (1) ?
GOS16662 (2) ?
GOS17124 (3) ?
GOS1716 (1) ?
GOS18783 (1) ?
GOS2568 (2) ?
GOS2977 (1) ?
GOS3431 (1) ?
GOS4555 (1) ?
GOS5157 (1) ?
GOS5767 (4) ?
GOS5892 (47) ?
GOS7959 (1) ?
GOS8442 (1) ?
GOS8463 (5) ?
GOS891 (1) ?
NOG07229 (2)
pfam00106 (1) short chain dehydrogenase.
pfam00565 (1) Staphylococcal nuclease homologue. [SNase numbering in parentheses]
pfam02696 (1) Uncharacterized ACR, YdiU/UPF0061 family.
pfam06041 (1) Bacterial protein of unknown function (DUF924).
pfam06750 (1) Bacterial Peptidase A24 N-terminal domain.
smart00271 (4) DnaJ molecular chaperone homology domain;

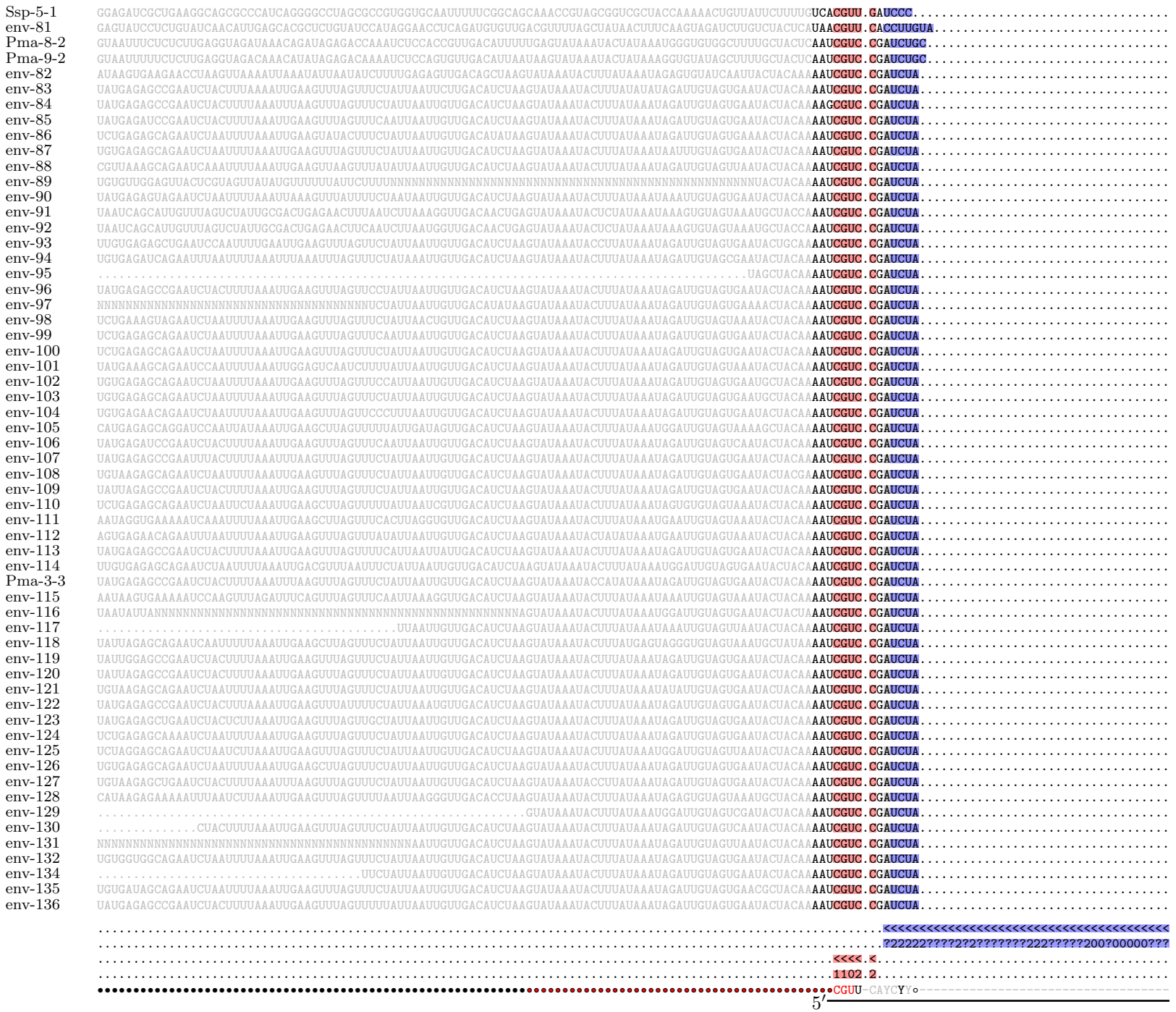
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Ssp-1-3, Ssp-1-5, env-133, env-387

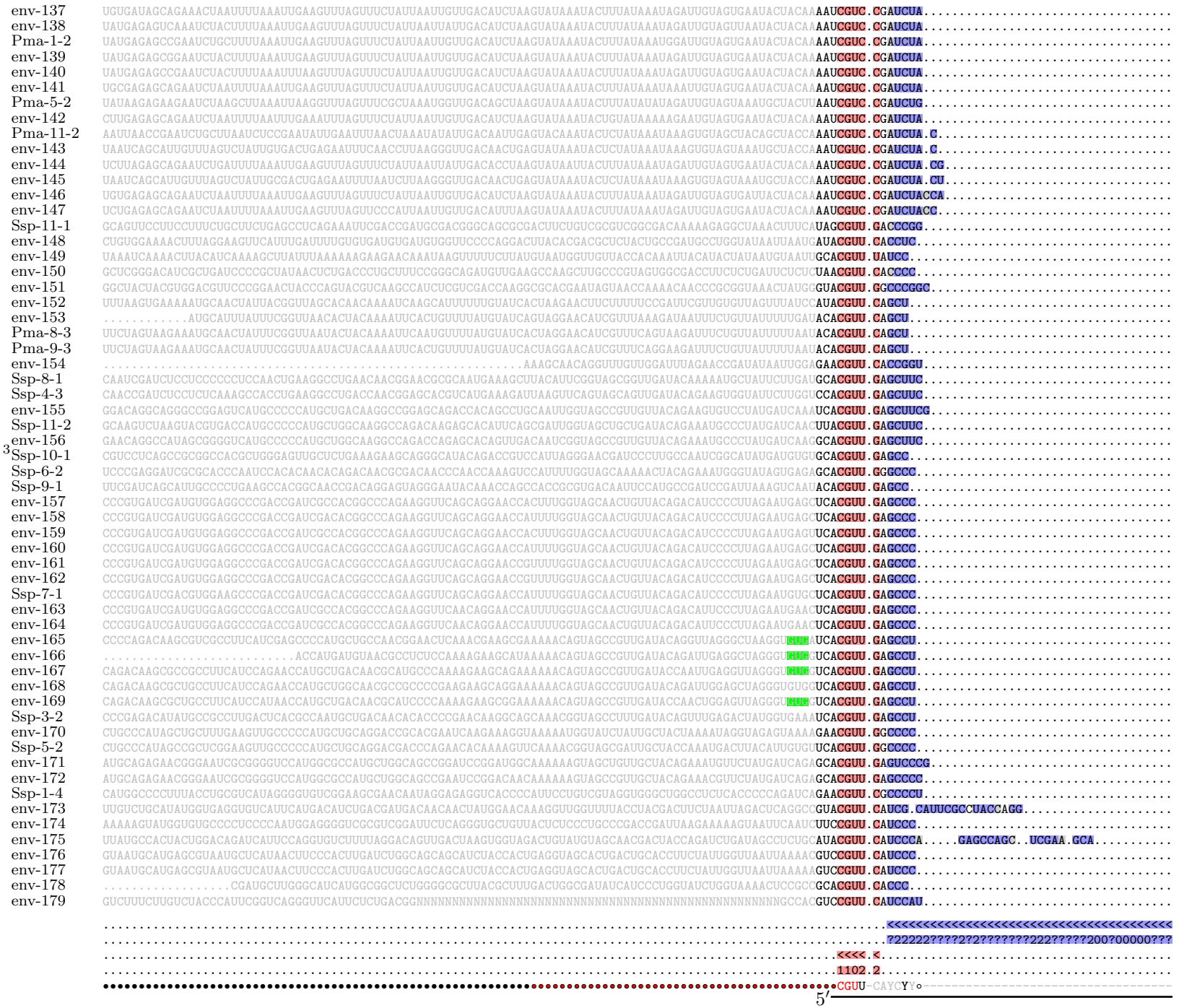
29.5 Multiple-sequence alignment

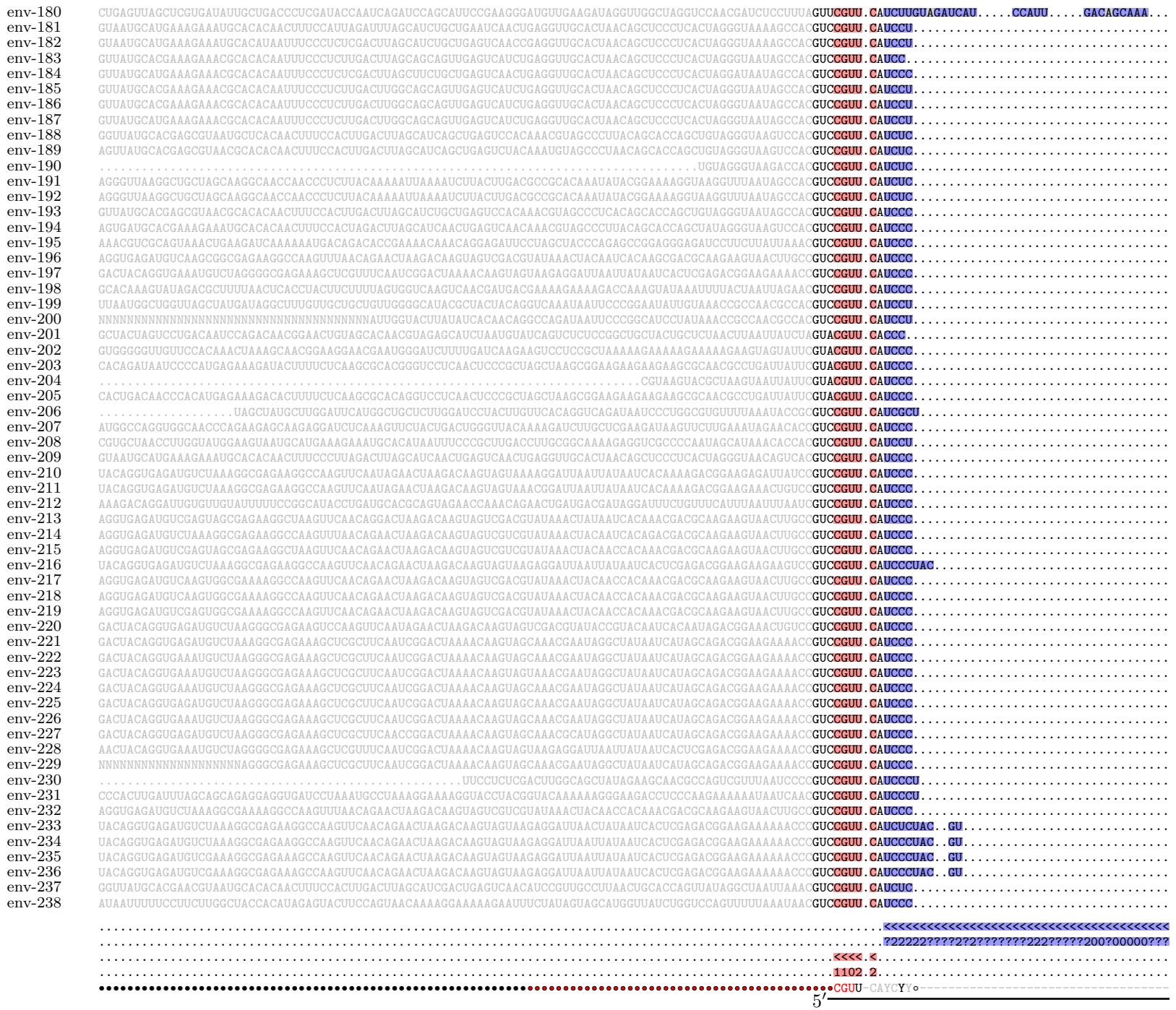
Each Downstream-peptide RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 29.1. Superscript numbers refer to annotations listed in Section 29.2. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Downstream-peptide RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

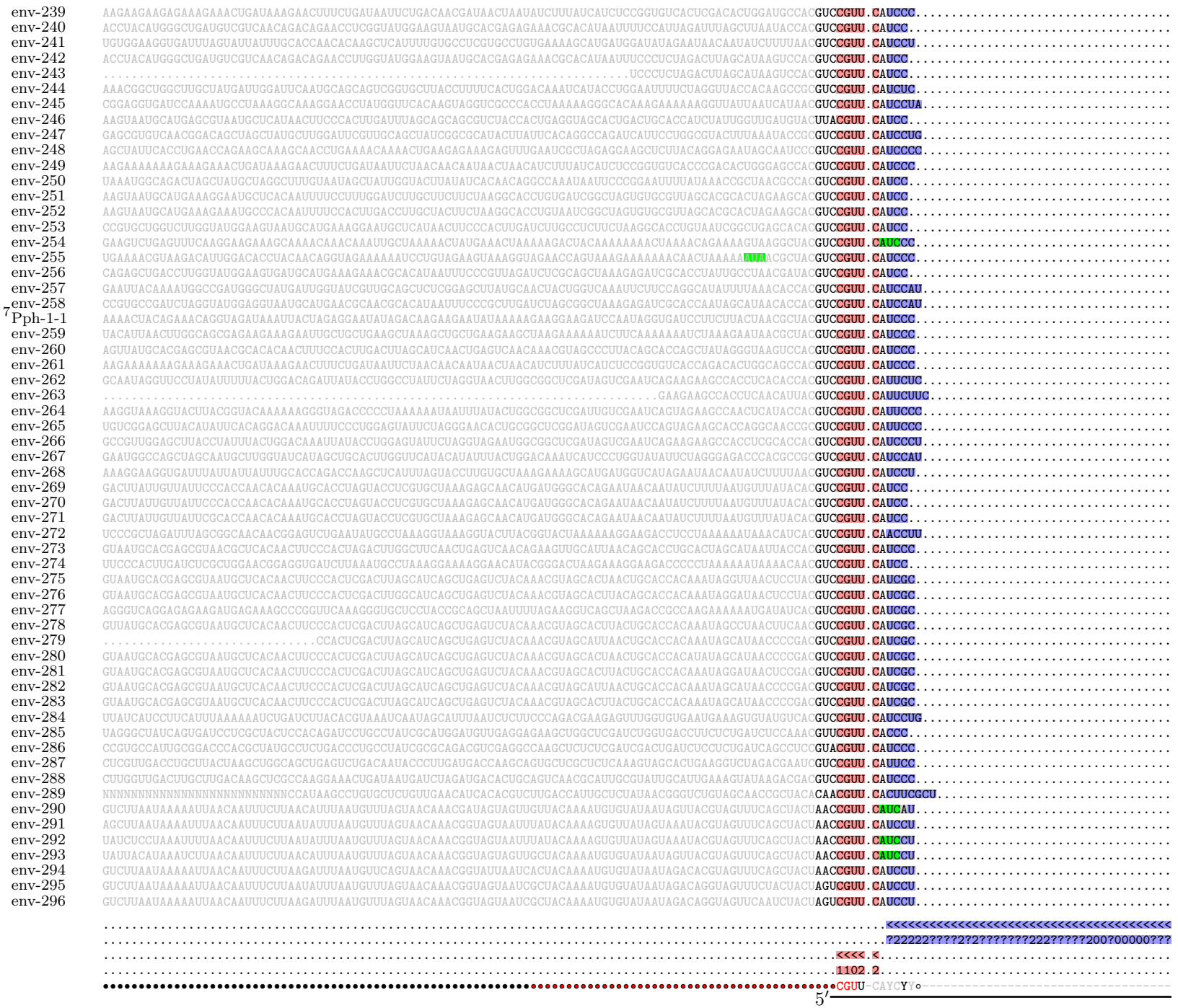


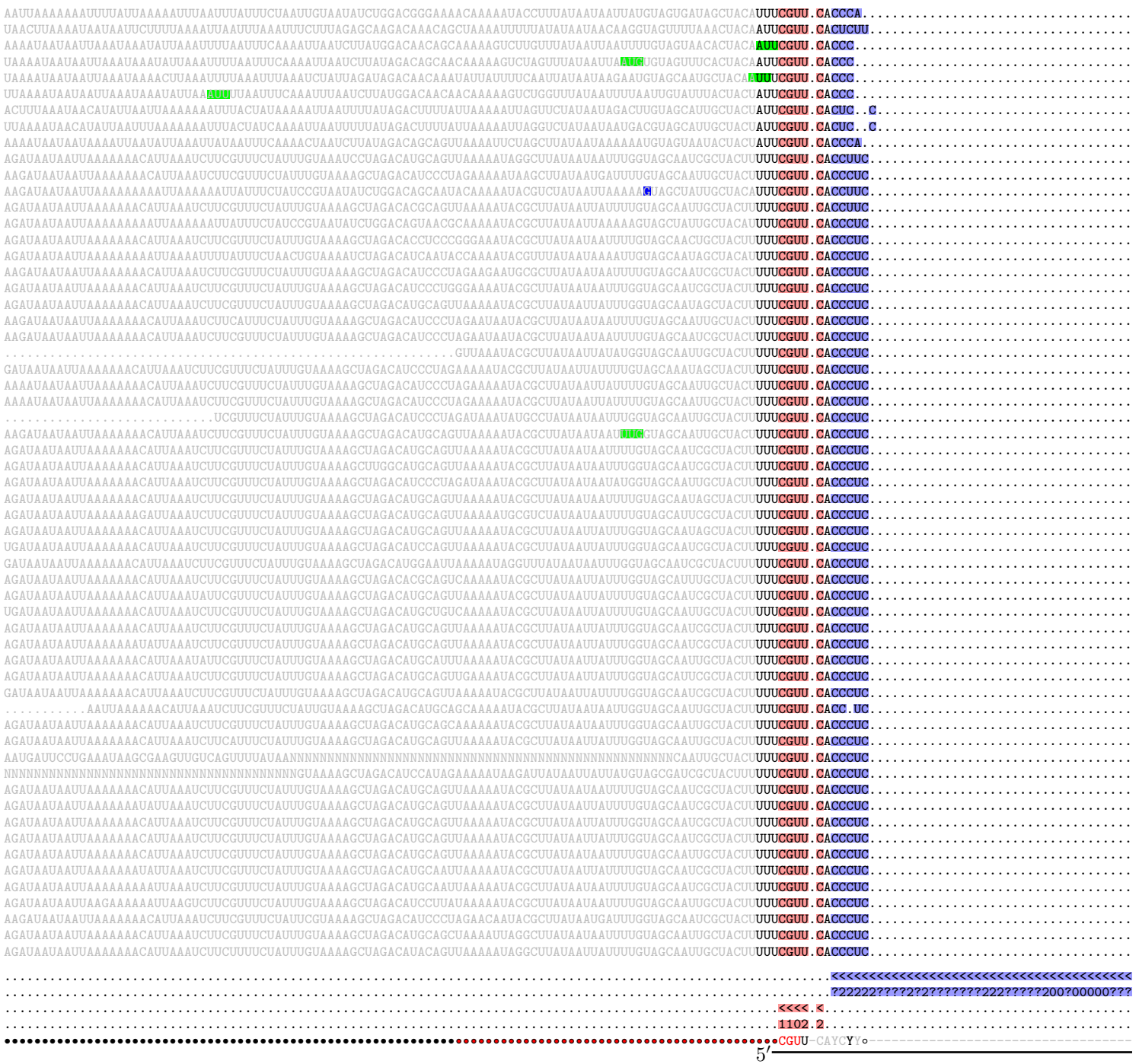

env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
Pma-1-1
env-71
env-72
env-73
env-74
Pma-7-1
Pma-11-1
Pma-8-1
env-75
Pma-9-1
Ssp-4-2
Ssp-1-2
env-76
env-77
env-78
env-79
env-80
5′

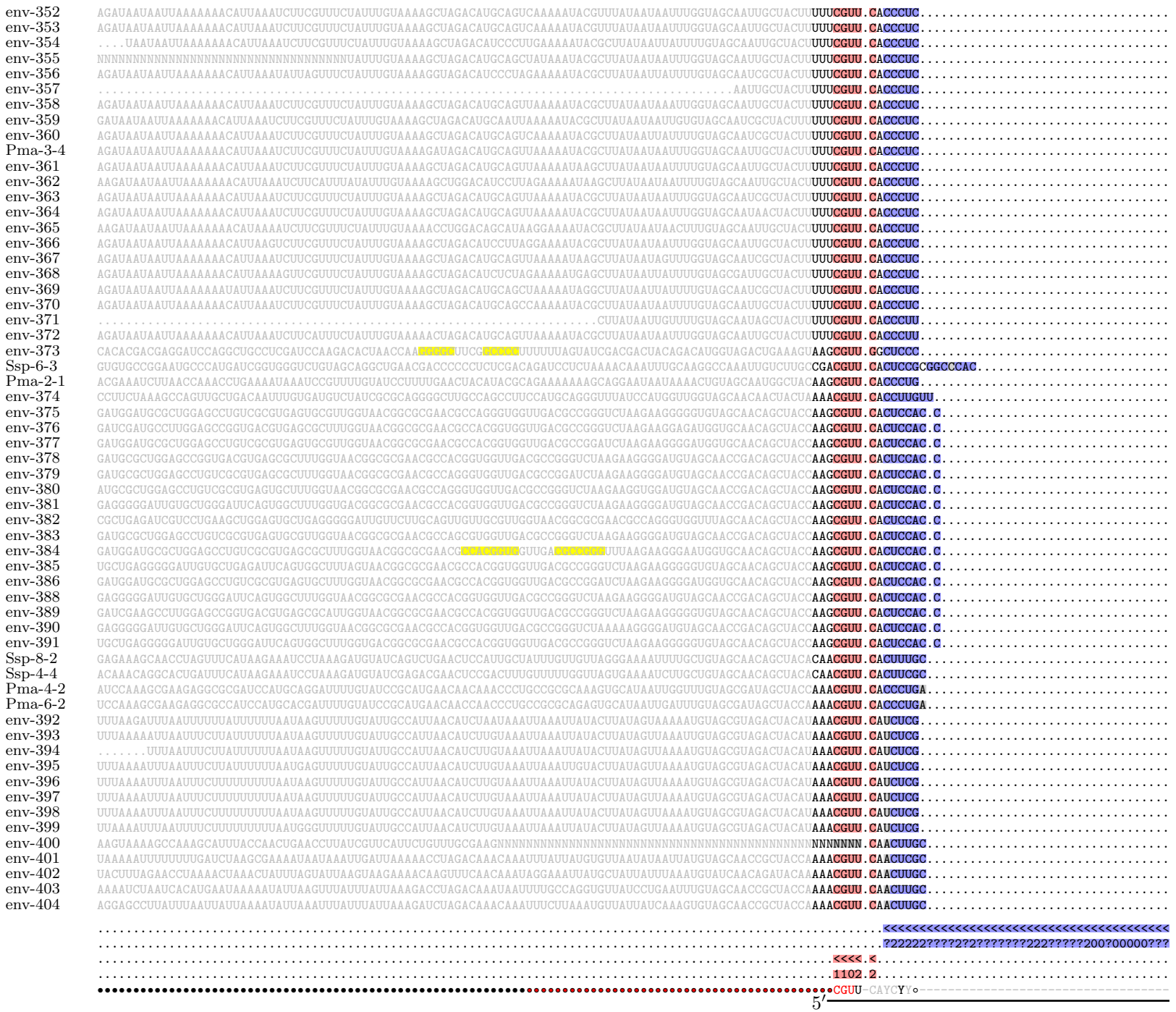

Ssp-5-1
env-81
Pma-8-2
Pma-9-2
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
Pma-3-3
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-134
env-135
env-136
5′

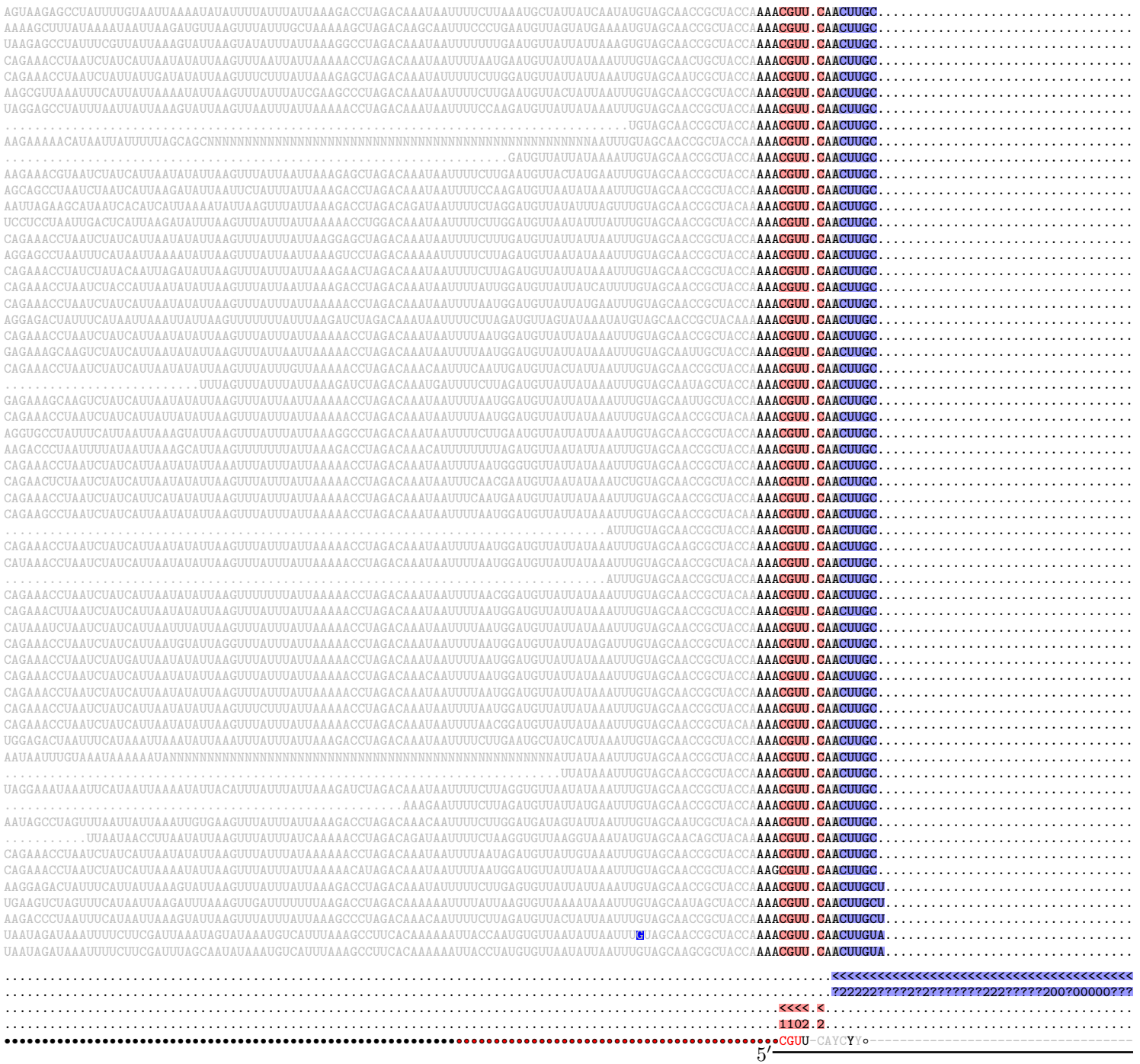

env-137
env-138
Pma-1-2
env-139
env-140
env-141
Pma-5-2
env-142
Pma-11-2
env-143
env-144
env-145
env-146
env-147
Ssp-11-1
env-148
env-149
env-150
env-151
env-152
env-153
Pma-8-3
Pma-9-3
env-154
Ssp-8-1
Ssp-4-3
env-155
Ssp-11-2
env-156
3
Ssp-10-1
Ssp-6-2
Ssp-9-1
env-157
env-158
env-159
env-160
env-161
env-162
Ssp-7-1
env-163
env-164
env-165
env-166
env-167
env-168
env-169
Ssp-3-2
env-170
Ssp-5-2
env-171
env-172
Ssp-1-4
env-173
env-174
env-175
env-176
env-177
env-178
env-179
5′

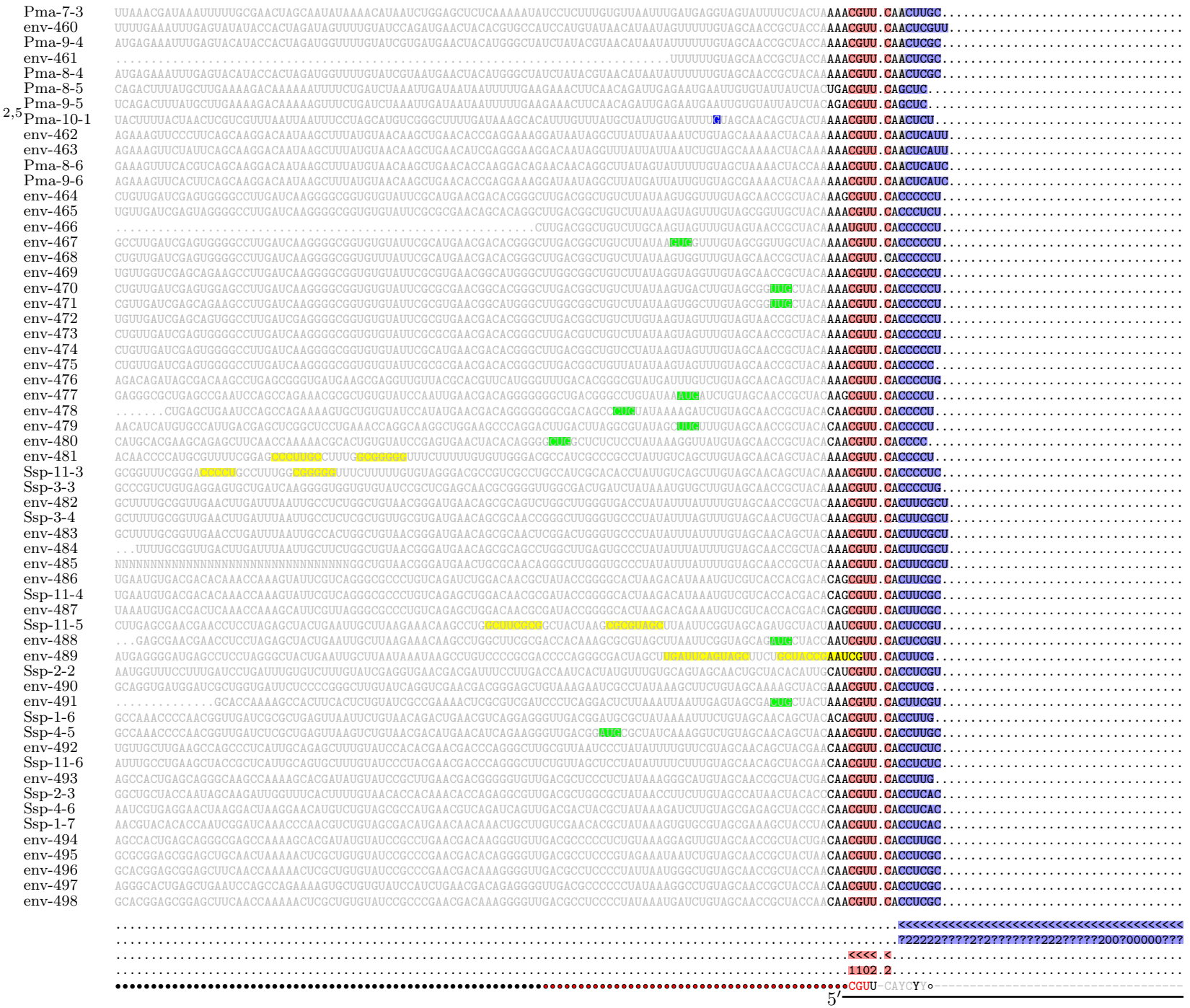

env-180
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
env-190
env-191
env-192
env-193
env-194
env-195
env-196
env-197
env-198
env-199
env-200
env-201
env-202
env-203
env-204
env-205
env-206
env-207
env-208
env-209
env-210
env-211
env-212
env-213
env-214
env-215
env-216
env-217
env-218
env-219
env-220
env-221
env-222
env-223
env-224
env-225
env-226
env-227
env-228
env-229
env-230
env-231
env-232
env-233
env-234
env-235
env-236
env-237
env-238
5′

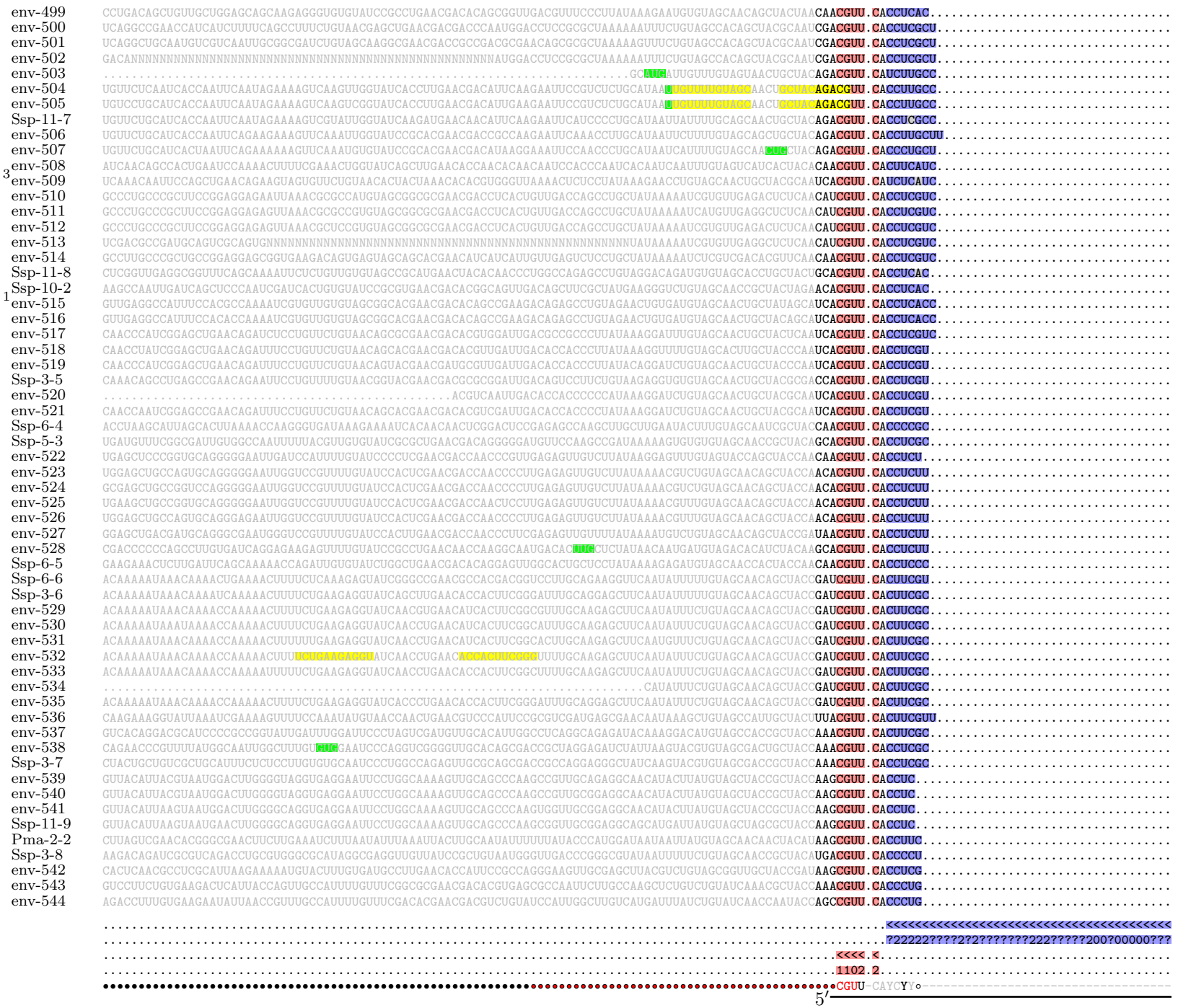

env-239
env-240
env-241
env-242
env-243
env-244
env-245
env-246
env-247
env-248
env-249
env-250
env-251
env-252
env-253
env-254
env-255
env-256
env-257
env-258
7 Pph-1-1
env-259
env-260
env-261
env-262
env-263
env-264
env-265
env-266
env-267
env-268
env-269
env-270
env-271
env-272
env-273
env-274
env-275
env-276
env-277
env-278
env-279
env-280
env-281
env-282
env-283
env-284
env-285
env-286
env-287
env-288
env-289
env-290
env-291
env-292
env-293
env-294
env-295
env-296
5′
CGUU-CAYCYY

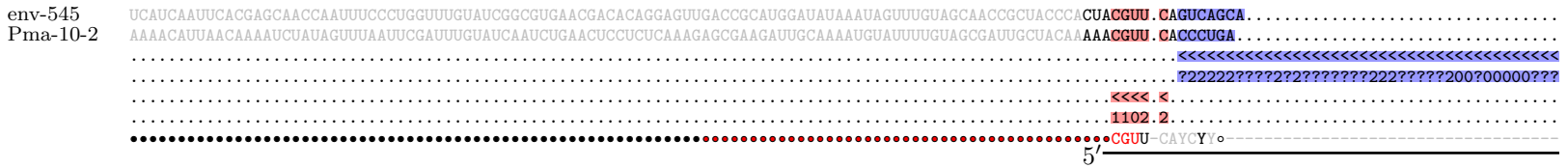

Pma-7-2
env-297
env-298
env-299
env-300
env-301
env-302
env-303
env-304
env-305
env-306
2,6Pma-11-3
env-307
env-308
env-309
env-310
env-311
env-312
env-313
env-314
env-315
env-316
env-317
env-318
env-319
env-320
env-321
env-322
env-323
env-324
Pma-5-3
env-325
env-326
env-327
env-328
Pma-1-3
env-329
env-330
env-331
env-332
env-333
env-334
env-335
env-336
env-337
env-338
env-339
env-340
env-341
env-342
env-343
env-344
env-345
env-346
env-347
env-348
env-349
env-350
env-351
CGUU-CAYCYY
5′


env-352
env-353
env-354
env-355
env-356
env-357
env-358
env-359
env-360
Pma-3-4
env-361
env-362
env-363
env-364
env-365
env-366
env-367
env-368
env-369
env-370
env-371
env-372
env-373
Ssp-6-3
Pma-2-1
env-374
env-375
env-376
env-377
env-378
env-379
env-380
env-381
env-382
env-383
env-384
env-385
env-386
env-388
env-389
env-390
env-391
Ssp-8-2
Ssp-4-4
Pma-4-2
Pma-6-2
env-392
env-393
env-394
env-395
env-396
env-397
env-398
env-399
env-400
env-401
env-402
env-403
env-404
5′

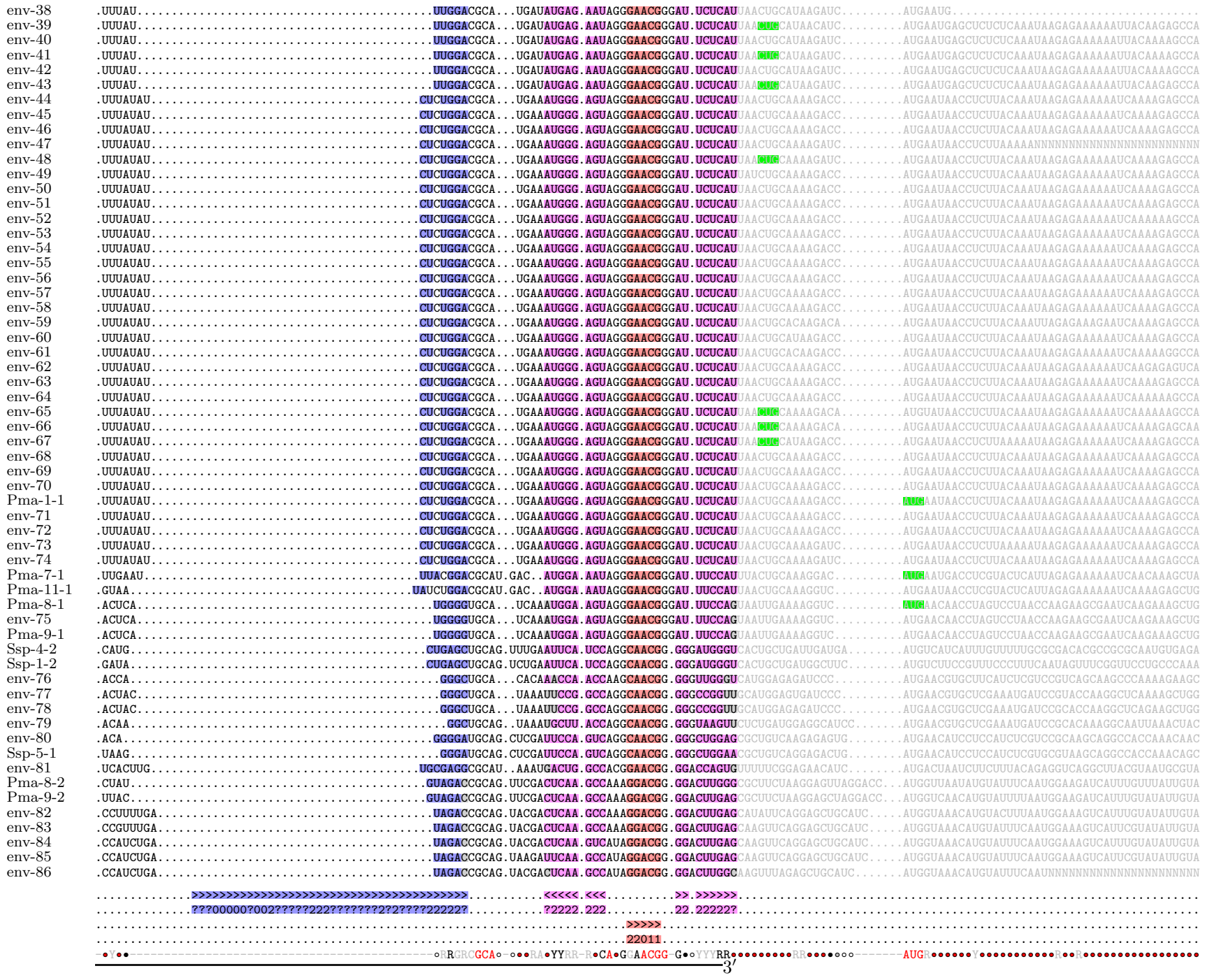

env-405
env-406
env-407
env-408
env-409
env-410
env-411
env-412
env-413
env-414
env-415
env-416
env-417
Pma-5-4
env-418
env-419
env-420
env-421
env-422
env-423
env-424
env-425
env-426
env-427
env-428
env-429
env-430
env-431
env-432
env-433
Pma-3-5
env-434
env-435
env-436
env-437
env-438
env-439
env-440
env-441
env-442
Pma-1-4
env-443
env-444
env-445
env-446
env-447
env-448
env-449
env-450
env-451
env-452
env-453
env-454
env-455
env-456
env-457
env-458
2,4Pma-11-4
env-459
5′

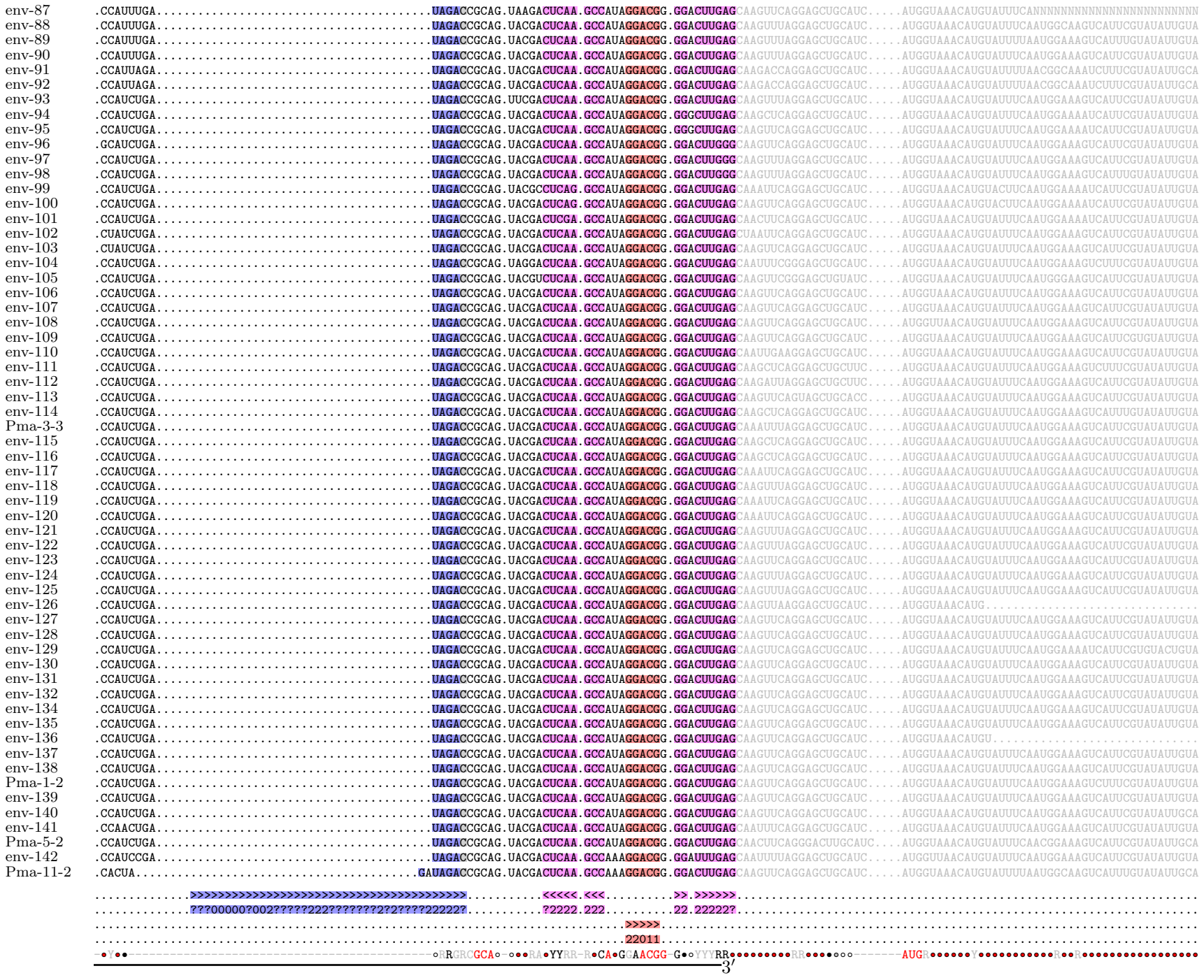

Pma-7-3
env-460
Pma-9-4
env-461
Pma-8-4
Pma-8-5
Pma-9-5
2,5Pma-10-1
env-462
env-463
Pma-8-6
Pma-9-6
env-464
env-465
env-466
env-467
env-468
env-469
env-470
env-471
env-472
env-473
env-474
env-475
env-476
env-477
env-478
env-479
env-480
env-481
Ssp-11-3
Ssp-3-3
env-482
Ssp-3-4
env-483
env-484
env-485
env-486
Ssp-11-4
env-487
Ssp-11-5
env-488
env-489
Ssp-2-2
env-490
env-491
Ssp-1-6
Ssp-4-5
env-492
Ssp-11-6
env-493
Ssp-2-3
Ssp-4-6
Ssp-1-7
env-494
env-495
env-496
env-497
env-498
5′

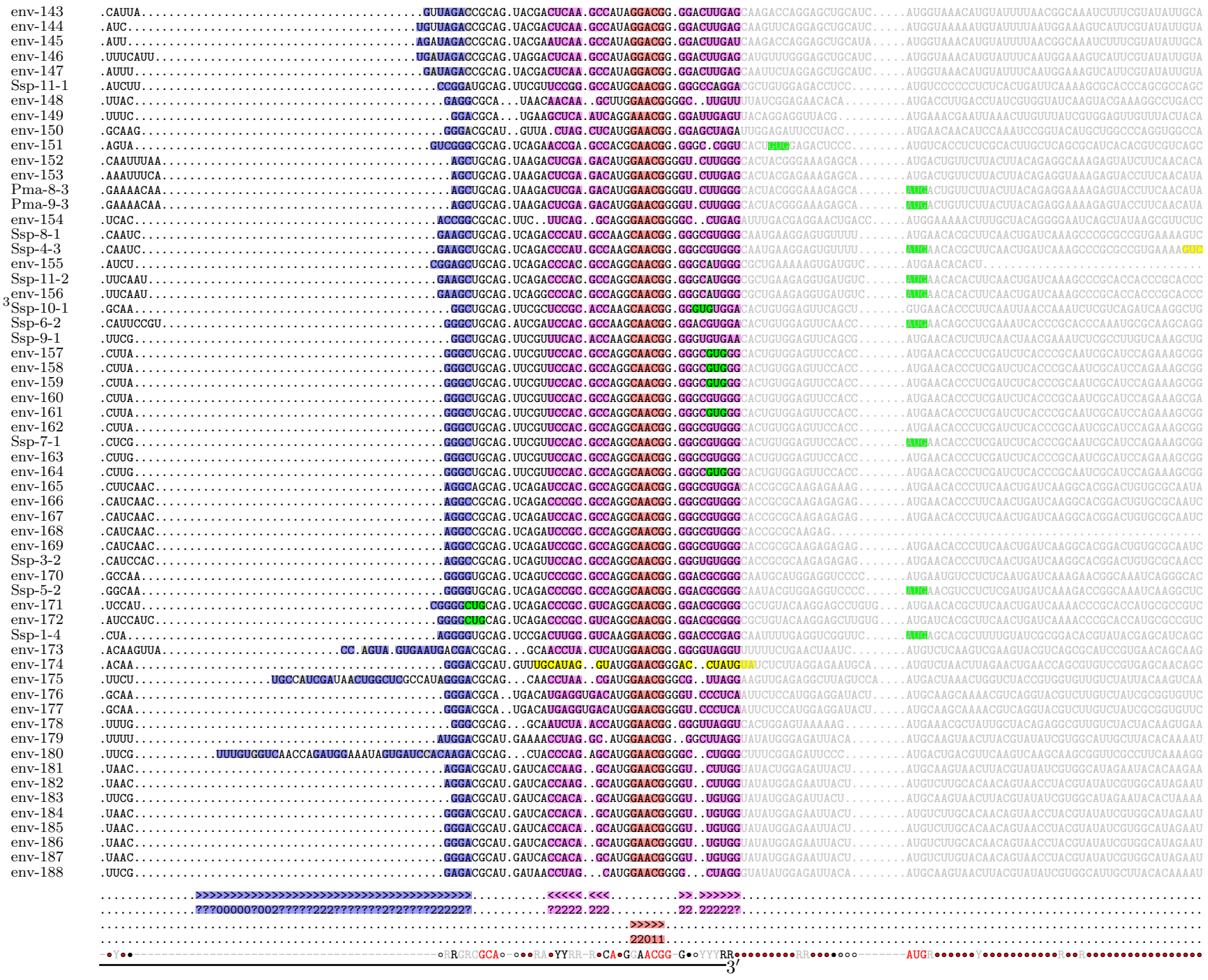

env-499
env-500
env-501
env-502
env-503
env-504
env-505
Ssp-11-7
env-506
env-507
env-508
env-509
env-510
env-511
env-512
env-513
env-514
Ssp-11-8
Ssp-10-2
env-515
env-516
env-517
env-518
env-519
Ssp-3-5
env-520
env-521
Ssp-6-4
Ssp-5-3
env-522
env-523
env-524
env-525
env-526
env-527
env-528
Ssp-6-5
Ssp-6-6
Ssp-3-6
env-529
env-530
env-531
env-532
env-533
env-534
env-535
env-536
env-537
env-538
Ssp-3-7
env-539
env-540
env-541
Ssp-11-9
Pma-2-2
Ssp-3-8
env-542
env-543
env-544
3
1
5′

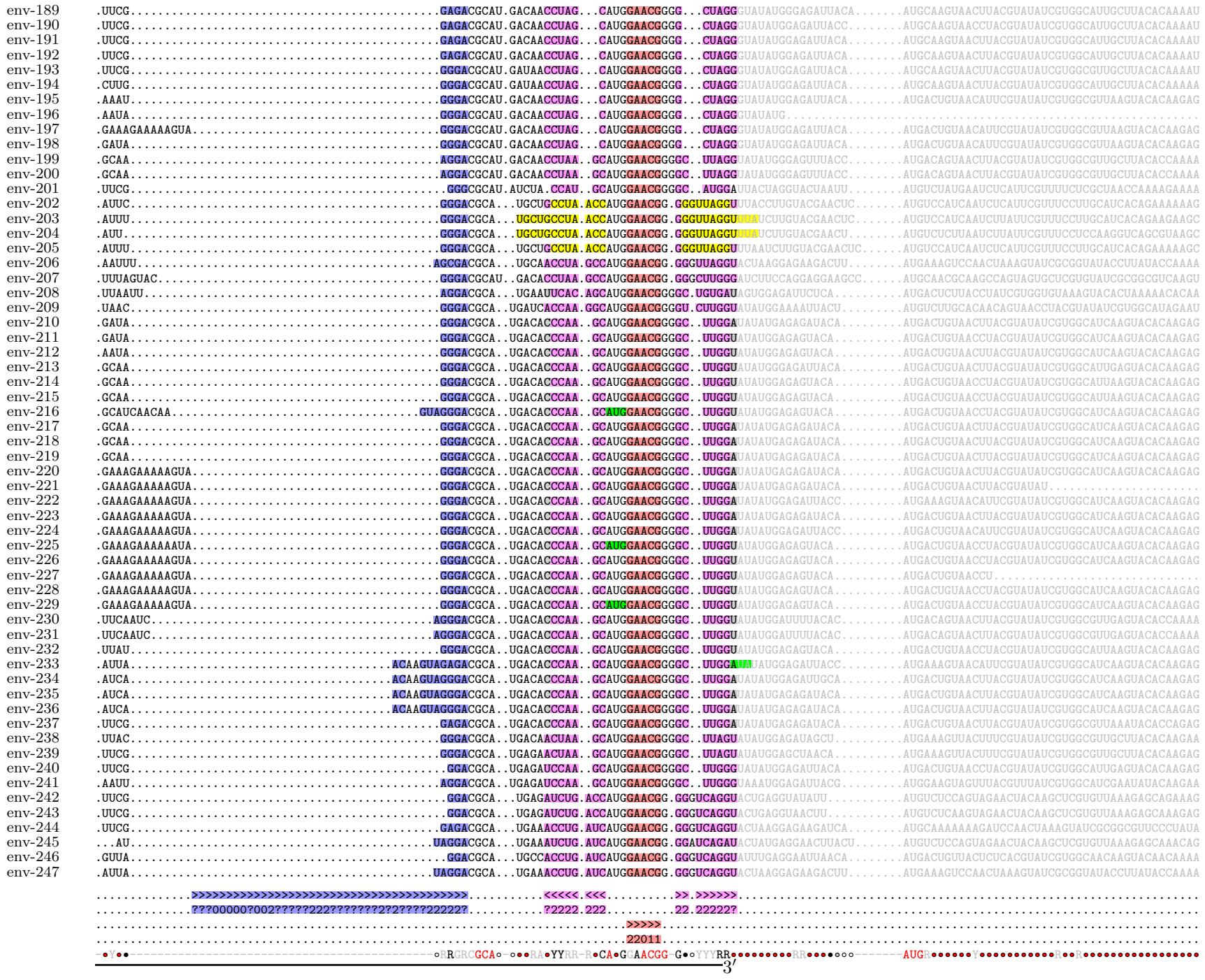


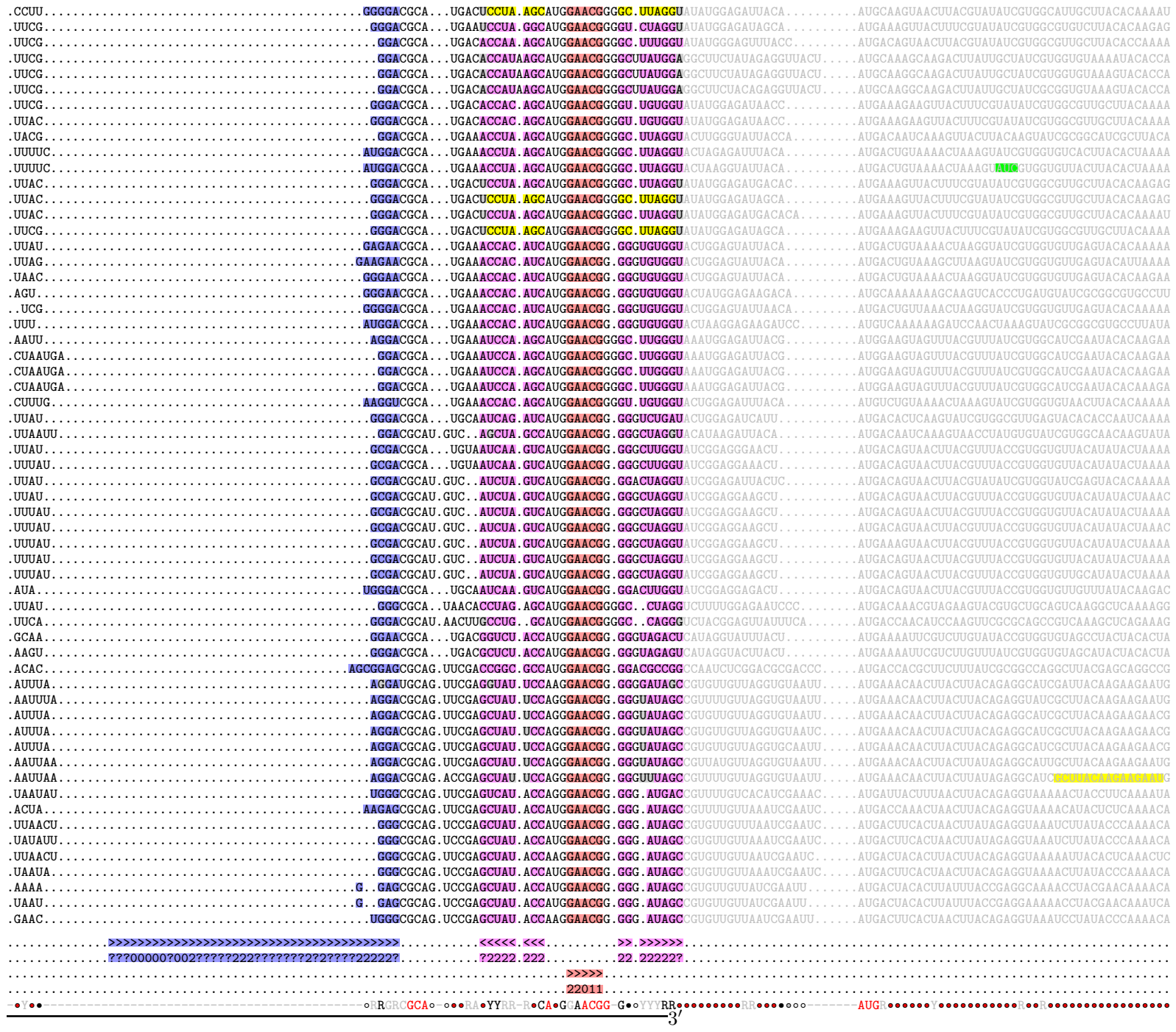
alignment positions 151 · · · 310

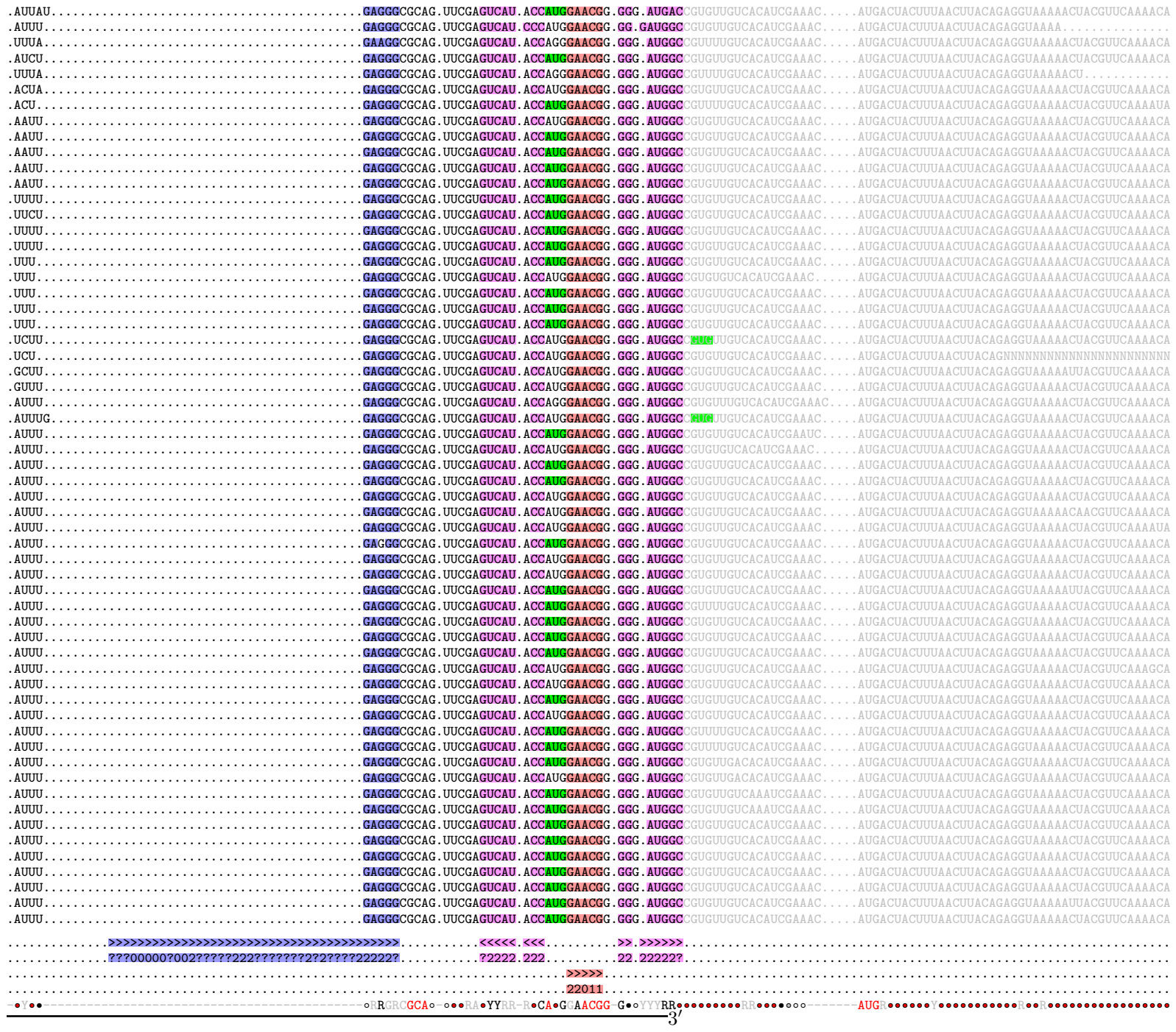

env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
Pma-1-1
env-71
env-72
env-73
env-74
Pma-7-1
Pma-11-1
Pma-8-1
env-75
Pma-9-1
Ssp-4-2
Ssp-1-2
env-76
env-77
env-78
env-79
env-80
Ssp-5-1
env-81
Pma-8-2
Pma-9-2
env-82
env-83
env-84
env-85
env-86
3′

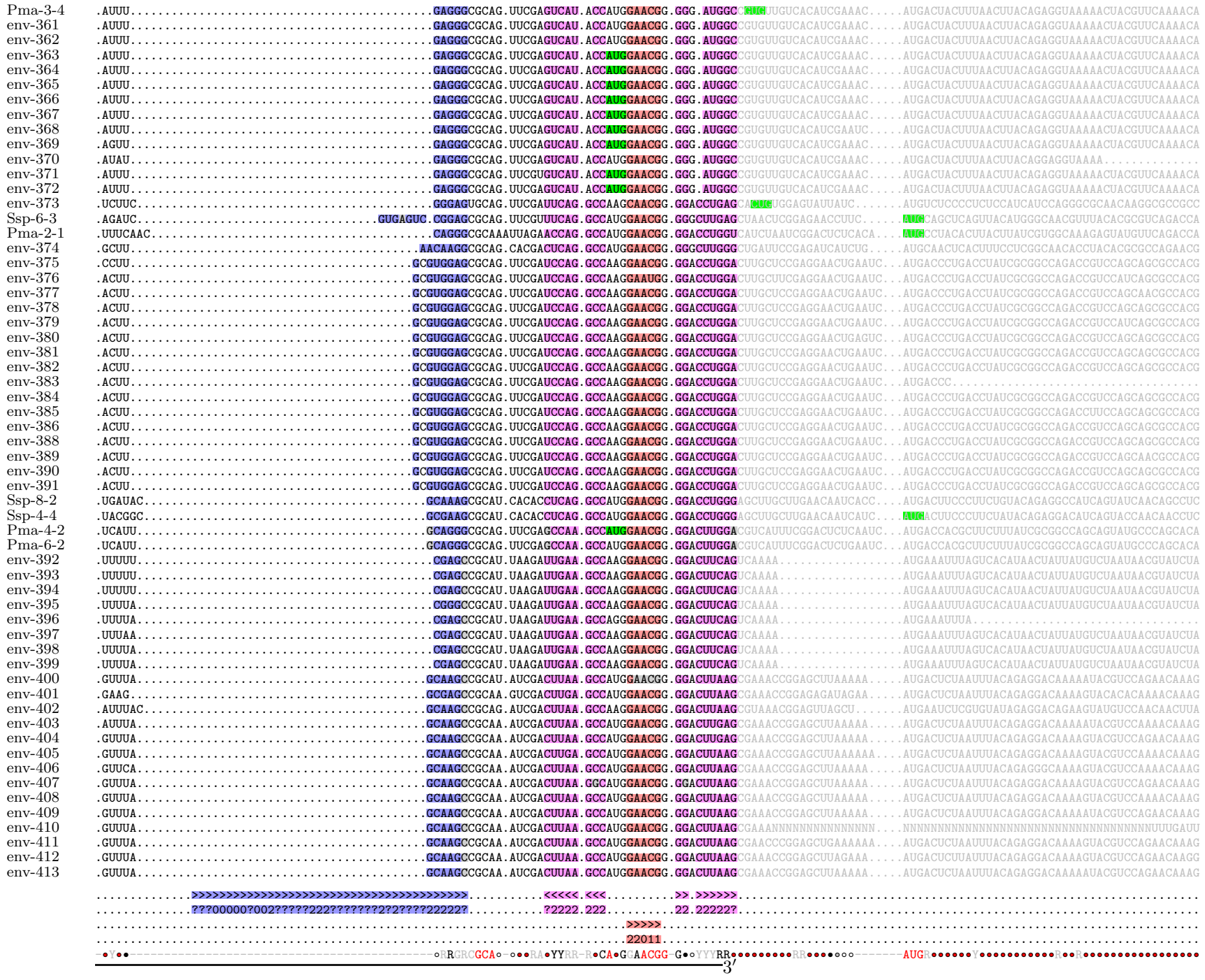

env-87
Pma-3-3
Pma-1-2
Pma-5-2
Pma-11-2
3′

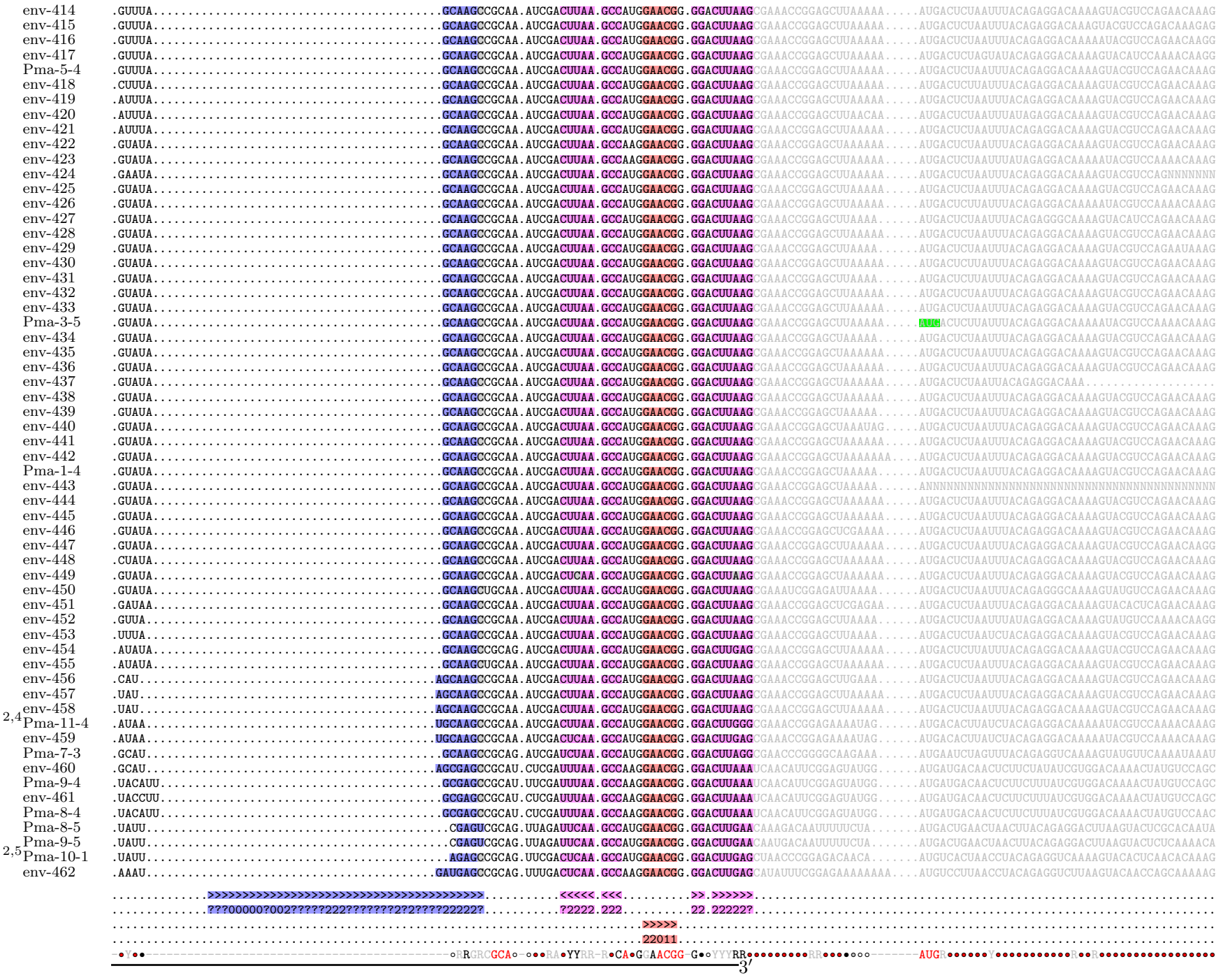

env-143
env-144
env-145
env-146
env-147
Ssp-11-1
env-148
env-149
env-150
env-151
env-152
env-153
Pma-8-3
Pma-9-3
env-154
Ssp-8-1
Ssp-4-3
env-155
Ssp-11-2
env-156
Ssp-10-1
Ssp-6-2
Ssp-9-1
env-157
env-158
env-159
env-160
env-161
env-162
Ssp-7-1
env-163
env-164
env-165
env-166
env-167
env-168
env-169
Ssp-3-2
env-170
Ssp-5-2
env-171
env-172
Ssp-1-4
env-173
env-174
env-175
env-176
env-177
env-178
env-179
env-180
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
3′

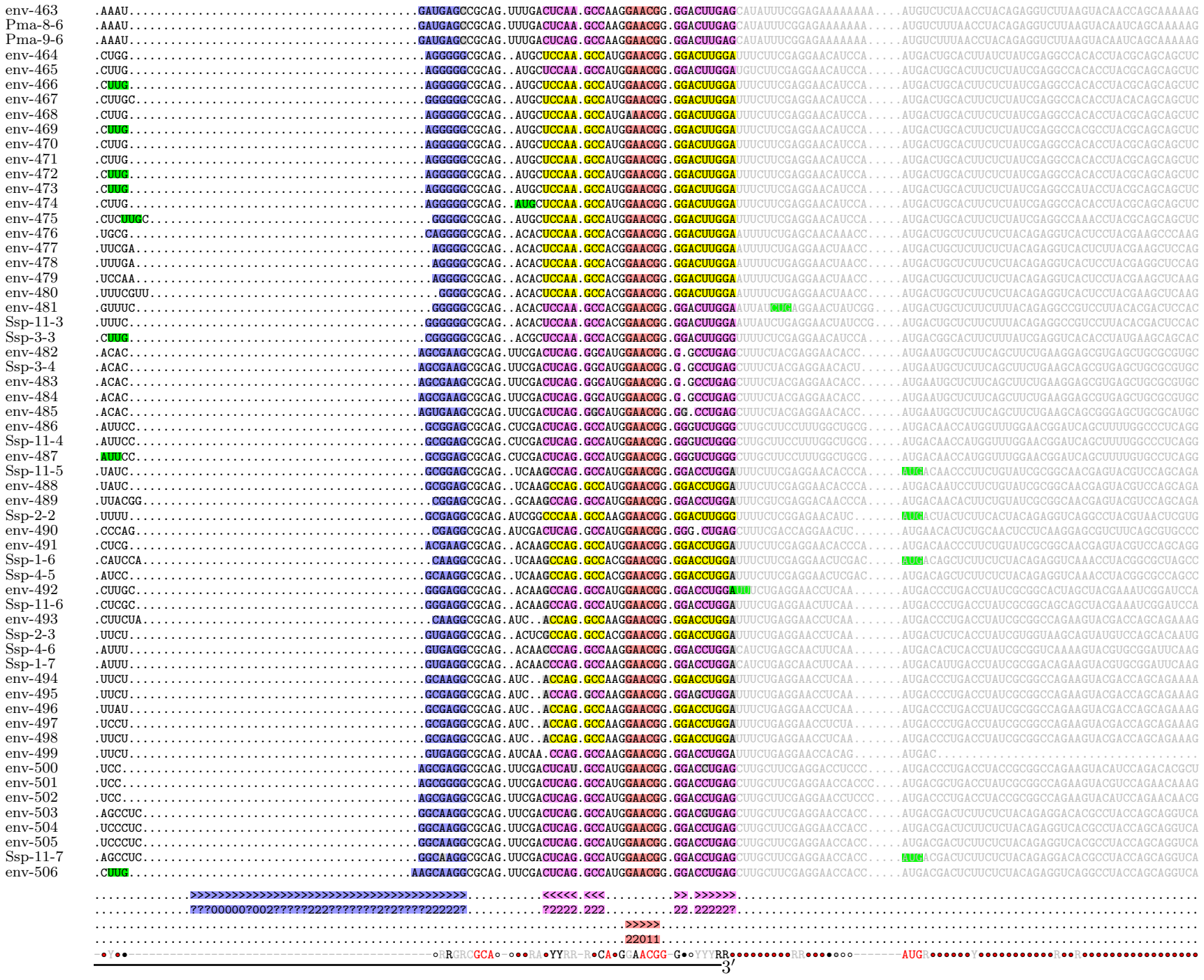

env-189
env-190
env-191
env-192
env-193
env-194
env-195
env-196
env-197
env-198
env-199
env-200
env-201
env-202
env-203
env-204
env-205
env-206
env-207
env-208
env-209
env-210
env-211
env-212
env-213
env-214
env-215
env-216
env-217
env-218
env-219
env-220
env-221
env-222
env-223
env-224
env-225
env-226
env-227
env-228
env-229
env-230
env-231
env-232
env-233
env-234
env-235
env-236
env-237
env-238
env-239
env-240
env-241
env-242
env-243
env-244
env-245
env-246
env-247
3′

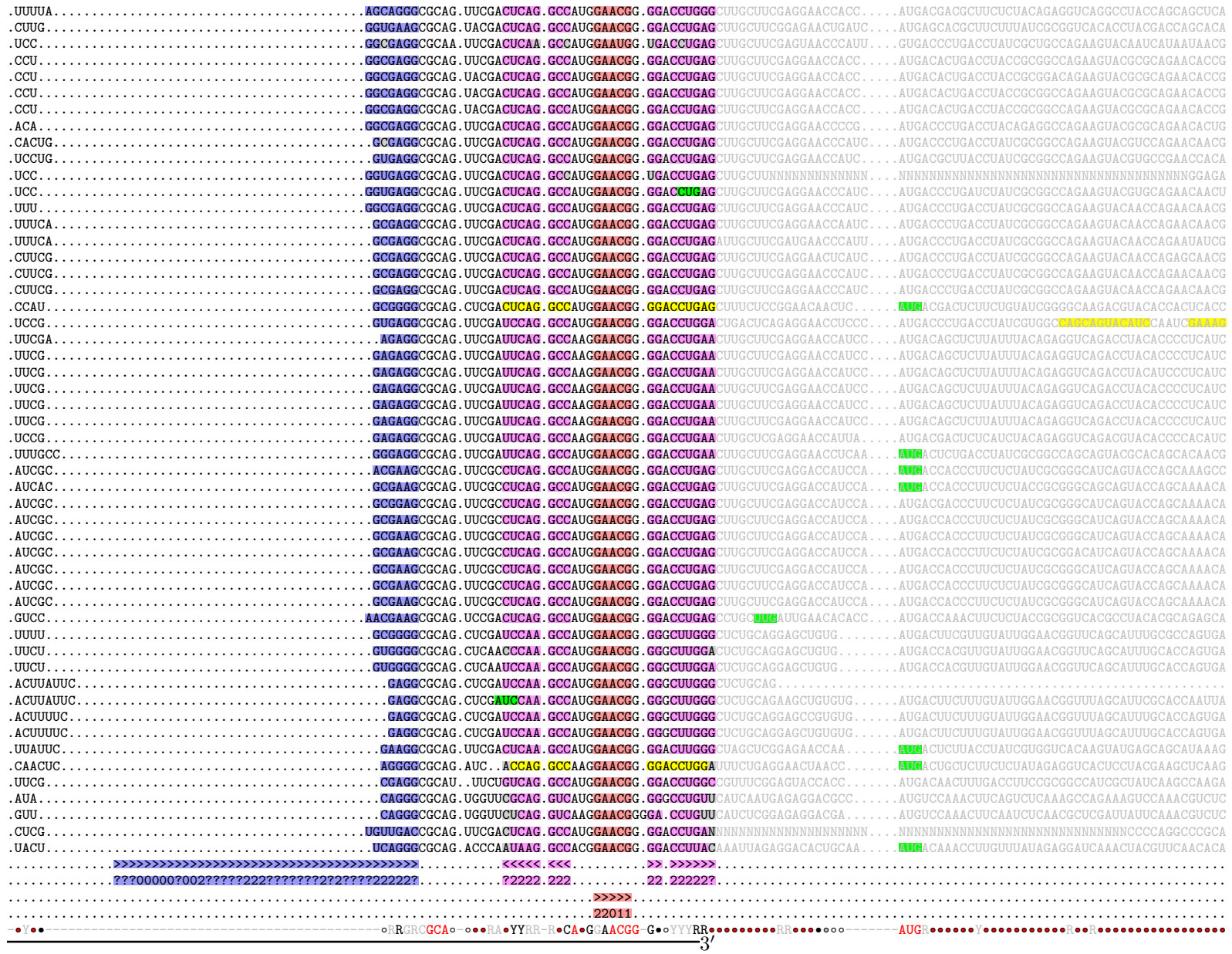

env-248
env-249
env-250
env-251
env-252
env-253
env-254
env-255
env-256
env-257
env-258
7Pph-1-1
env-259
env-260
env-261
env-262
env-263
env-264
env-265
env-266
env-267
env-268
env-269
env-270
env-271
env-272
env-273
env-274
env-275
env-276
env-277
env-278
env-279
env-280
env-281
env-282
env-283
env-284
env-285
env-286
env-287
env-288
env-289
env-290
env-291
env-292
env-293
env-294
env-295
env-296
Pma-7-2
env-297
env-298
env-299
env-300
env-301
env-302
env-303
env-304
3′

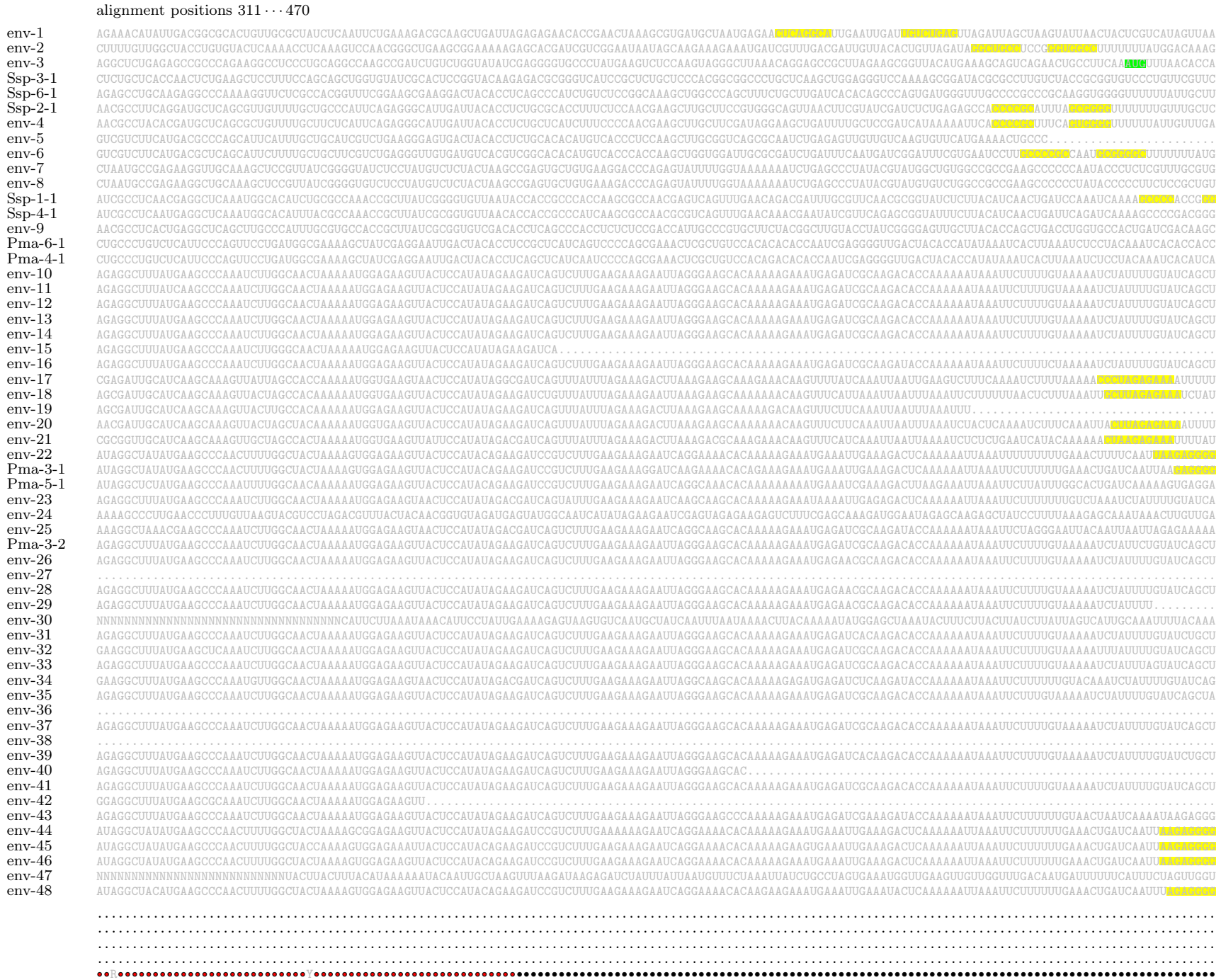

env-305
env-306
2,6Pma-11-3
env-307
env-308
env-309
env-310
env-311
env-312
env-313
env-314
env-315
env-316
env-317
env-318
env-319
env-320
env-321
env-322
env-323
env-324
Pma-5-3
env-325
env-326
env-327
env-328
Pma-1-3
env-329
env-330
env-331
env-332
env-333
env-334
env-335
env-336
env-337
env-338
env-339
env-340
env-341
env-342
env-343
env-344
env-345
env-346
env-347
env-348
env-349
env-350
env-351
env-352
env-353
env-354
env-355
env-356
env-357
env-358
env-359
env-360
3′

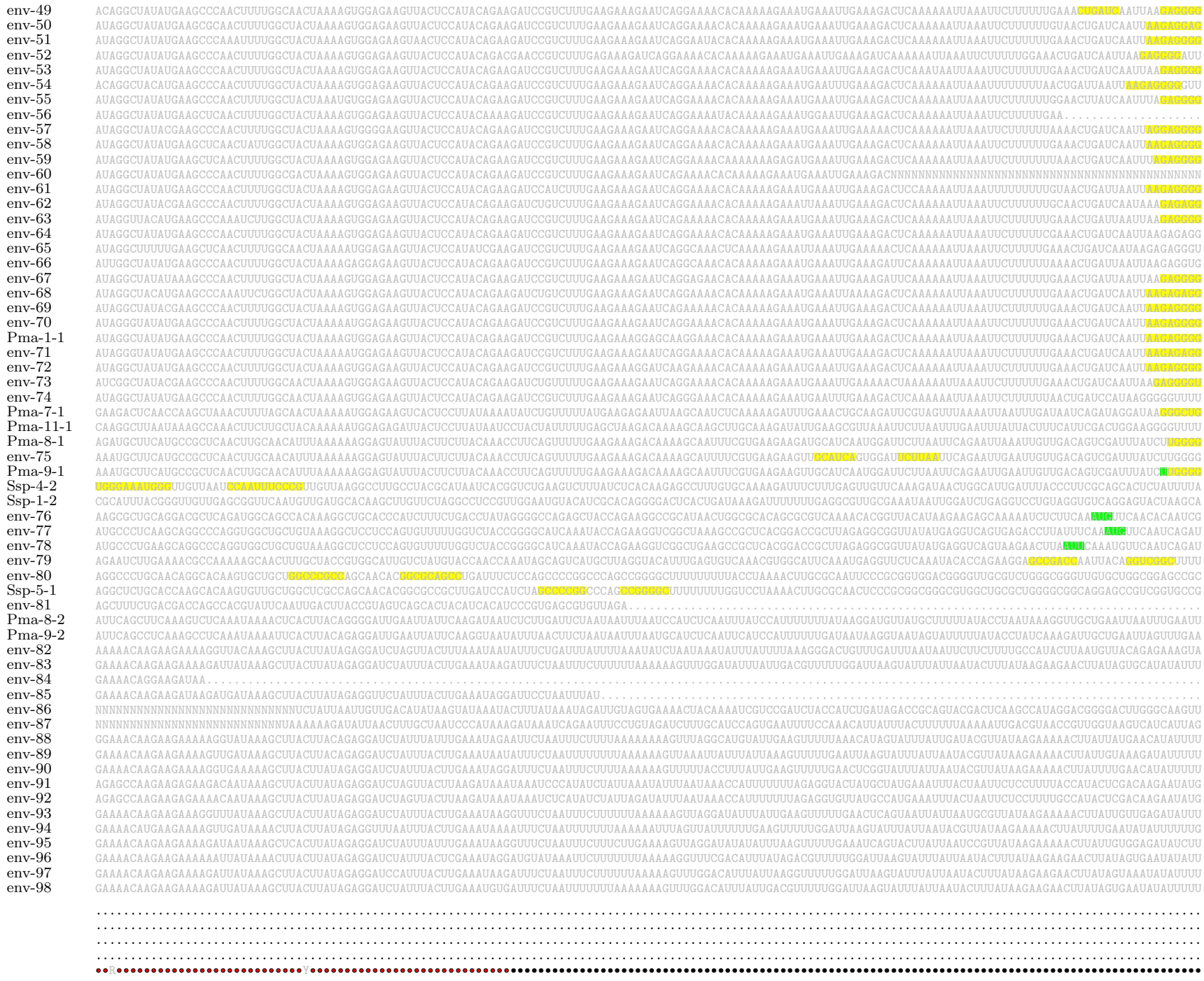

Pma-3-4
env-361
env-362
env-363
env-364
env-365
env-366
env-367
env-368
env-369
env-370
env-371
env-372
env-373
Ssp-6-3
Pma-2-1
env-374
env-375
env-376
env-377
env-378
env-379
env-380
env-381
env-382
env-383
env-384
env-385
env-386
env-388
env-389
env-390
env-391
Ssp-8-2
Ssp-4-4
Pma-4-2
Pma-6-2
env-392
env-393
env-394
env-395
env-396
env-397
env-398
env-399
env-400
env-401
env-402
env-403
env-404
env-405
env-406
env-407
env-408
env-409
env-410
env-411
env-412
env-413
3′

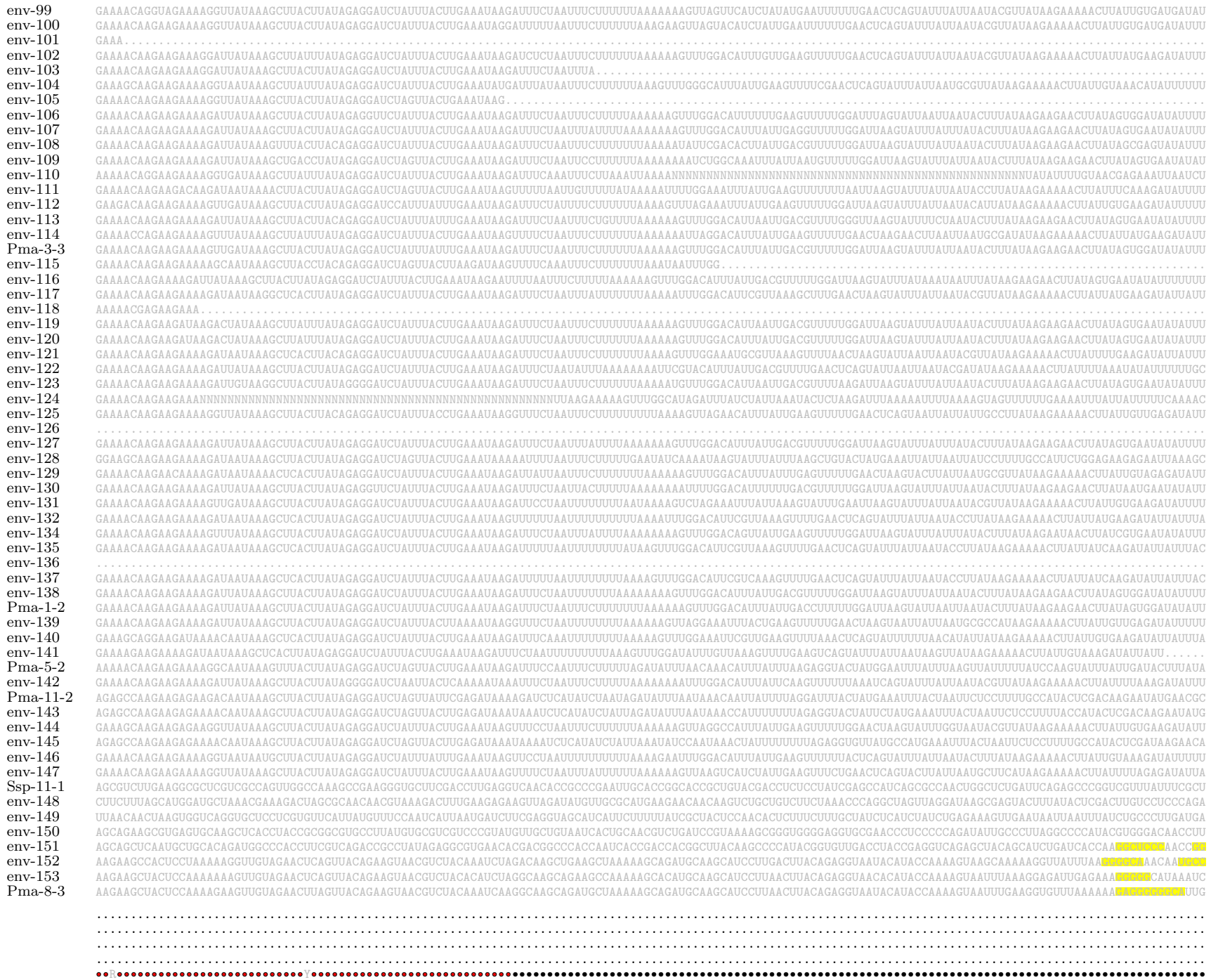

env-414
env-415
env-416
env-417
Pma-5-4
env-418
env-419
env-420
env-421
env-422
env-423
env-424
env-425
env-426
env-427
env-428
env-429
env-430
env-431
env-432
env-433
Pma-3-5
env-434
env-435
env-436
env-437
env-438
env-439
env-440
env-441
env-442
Pma-1-4
env-443
env-444
env-445
env-446
env-447
env-448
env-449
env-450
env-451
env-452
env-453
env-454
env-455
env-456
env-457
2,4
env-458
Pma-11-4
env-459
Pma-7-3
env-460
Pma-9-4
env-461
Pma-8-4
Pma-8-5
Pma-9-5
2,5
Pma-10-1
env-462
3′


env-463
Pma-8-6
Pma-9-6
env-464
env-465
env-466
env-467
env-468
env-469
env-470
env-471
env-472
env-473
env-474
env-475
env-476
env-477
env-478
env-479
env-480
env-481
Ssp-11-3
Ssp-3-3
env-482
Ssp-3-4
env-483
env-484
env-485
env-486
Ssp-11-4
env-487
Ssp-11-5
env-488
env-489
Ssp-2-2
env-490
env-491
Ssp-1-6
Ssp-4-5
env-492
Ssp-11-6
env-493
Ssp-2-3
Ssp-4-6
Ssp-1-7
env-494
env-495
env-496
env-497
env-498
env-499
env-500
env-501
env-502
env-503
env-504
env-505
Ssp-11-7
env-506
3′


env-507
env-508
env-509
env-510
env-511
env-512
env-513
env-514
Ssp-11-8
Ssp-10-2
env-515
env-516
env-517
env-518
env-519
Ssp-3-5
env-520
env-521
Ssp-6-4
Ssp-5-3
env-522
env-523
env-524
env-525
env-526
env-527
env-528
Ssp-6-5
Ssp-6-6
Ssp-3-6
env-529
env-530
env-531
env-532
env-533
env-534
env-535
env-536
env-537
env-538
Ssp-3-7
env-539
env-540
env-541
Ssp-11-9
Pma-2-2
Ssp-3-8
env-542
env-543
env-544
env-545
Pma-10-2
3
1
3′


alignment positions 311 · · · 470
env-1
env-2
env-3
Ssp-3-1
Ssp-6-1
Ssp-2-1
env-4
env-5
env-6
env-7
env-8
Ssp-1-1
Ssp-4-1
env-9
Pma-6-1
Pma-4-1
env-10
env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
Pma-3-1
Pma-5-1
env-23
env-24
env-25
Pma-3-2
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48

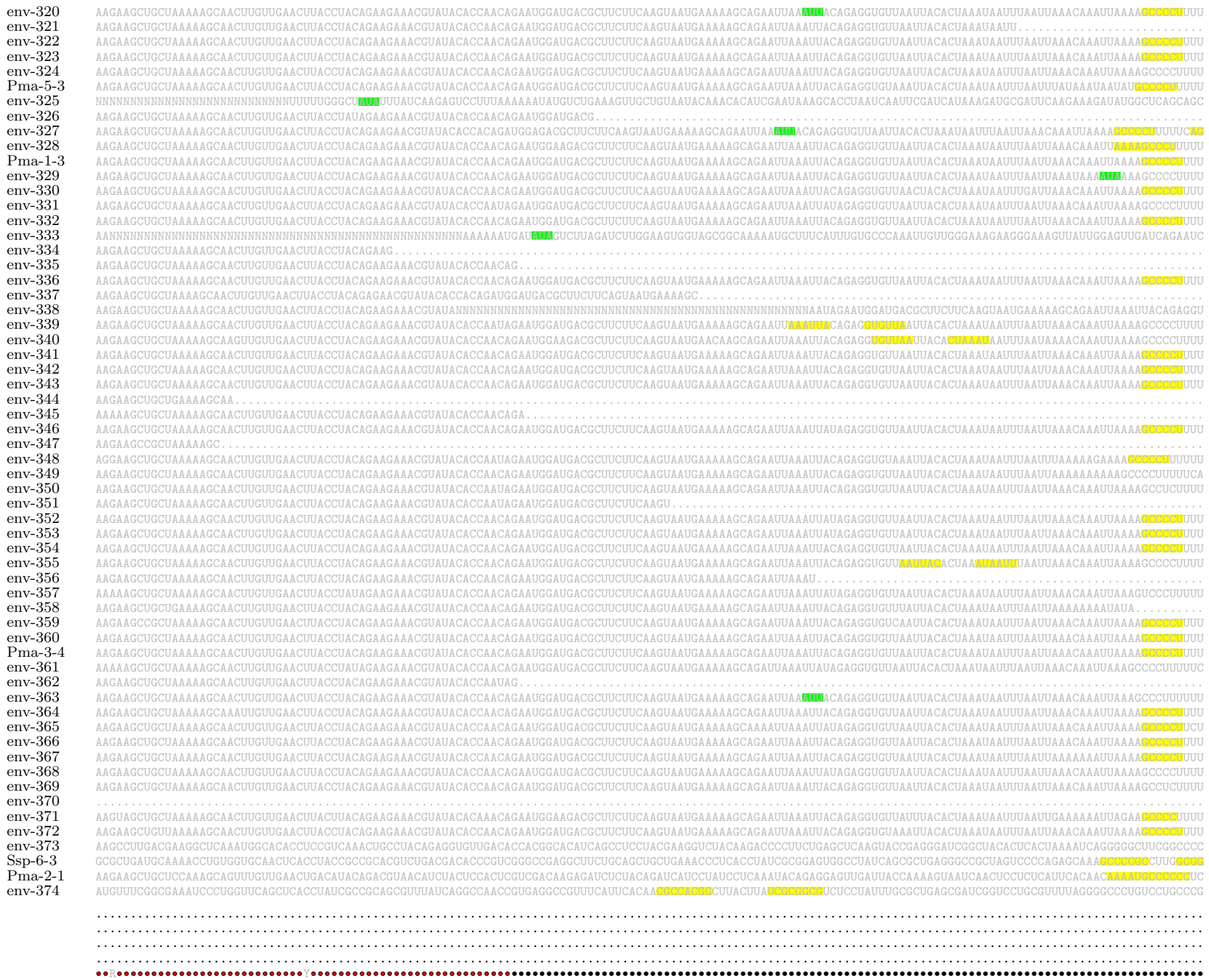

env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
Pma-1-1
env-71
env-72
env-73
env-74
Pma-7-1
Pma-11-1
Pma-8-1
env-75
Pma-9-1
Ssp-4-2
Ssp-1-2
env-76
env-77
env-78
env-79
env-80
Ssp-5-1
env-81
Pma-8-2
Pma-9-2
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98

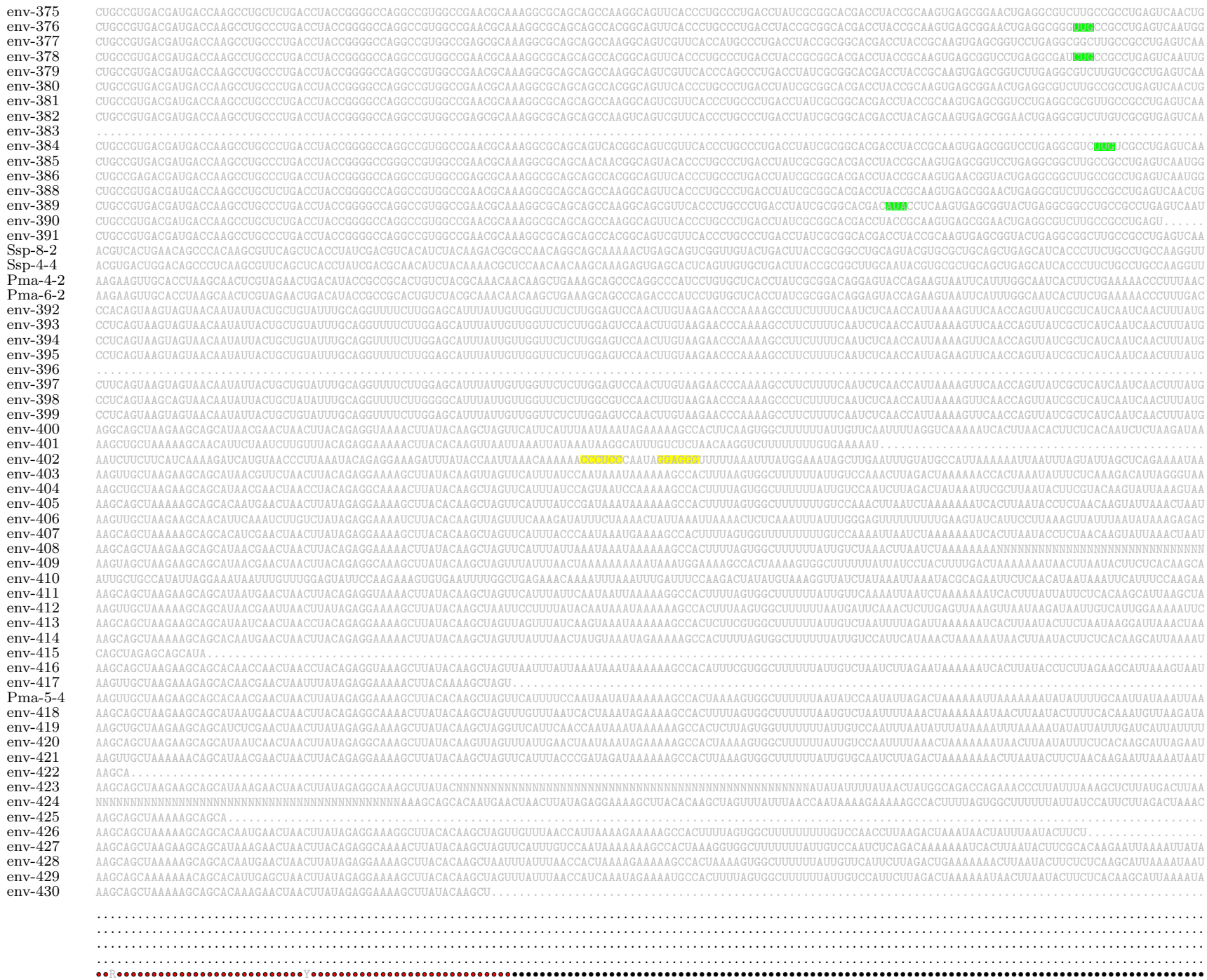

env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
Pma-3-3
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-134
env-135
env-136
env-137
env-138
Pma-1-2
env-139
env-140
env-141
Pma-5-2
env-142
Pma-11-2
env-143
env-144
env-145
env-146
env-147
Ssp-11-1
env-148
env-149
env-150
env-151
env-152
env-153
Pma-8-3

Pma-9-3
env-154
Ssp-8-1
Ssp-4-3
env-155
Ssp-11-2
env-156
Ssp-10-1
Ssp-6-2
Ssp-9-1
env-157
env-158
env-159
env-160
env-161
env-162
Ssp-7-1
env-163
env-164
env-165
env-166
env-167
env-168
env-169
Ssp-3-2
env-170
Ssp-5-2
env-171
env-172
Ssp-1-4
env-173
env-174
env-175
env-176
env-177
env-178
env-179
env-180
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
env-190
env-191
env-192
env-193
env-194
env-195
env-196
env-197
env-198
env-199
env-200
env-201
env-202

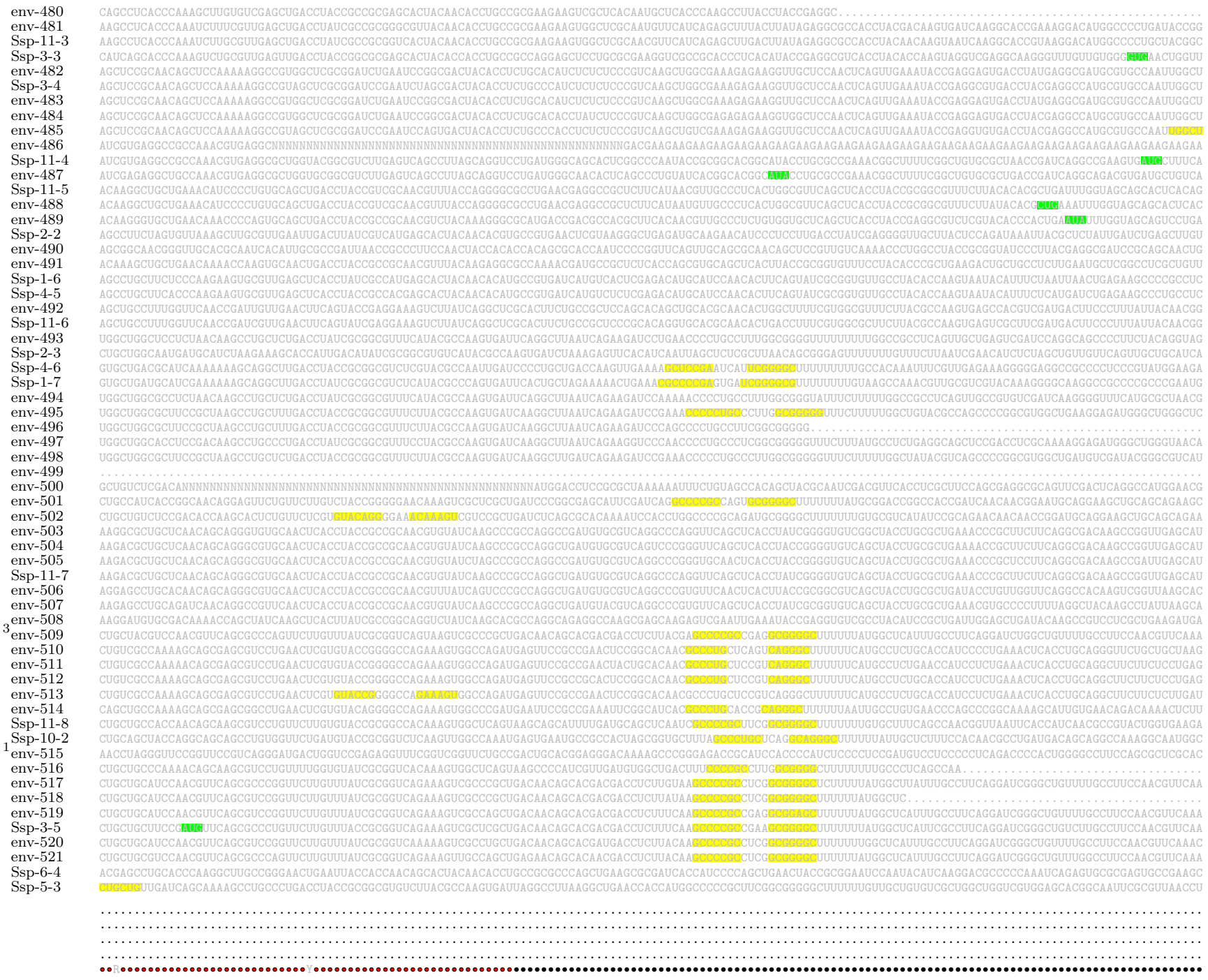
env-203
env-204
env-205
env-206
env-207
env-208
env-209
env-210
env-211
env-212
env-213
env-214
env-215
env-216
env-217
env-218
env-219
env-220
env-221
env-222
env-223
env-224
env-225
env-226
env-227
env-228
env-229
env-230
env-231
env-232
env-233
env-234
env-235
env-236
env-237
env-238
env-239
env-240
env-241
env-242
env-243
env-244
env-245
env-246
env-247
env-248
env-249
env-250
env-251
env-252
env-253
env-254
env-255
env-256
env-257
env-258
7Pph-1-1
env-259
env-260
env-261

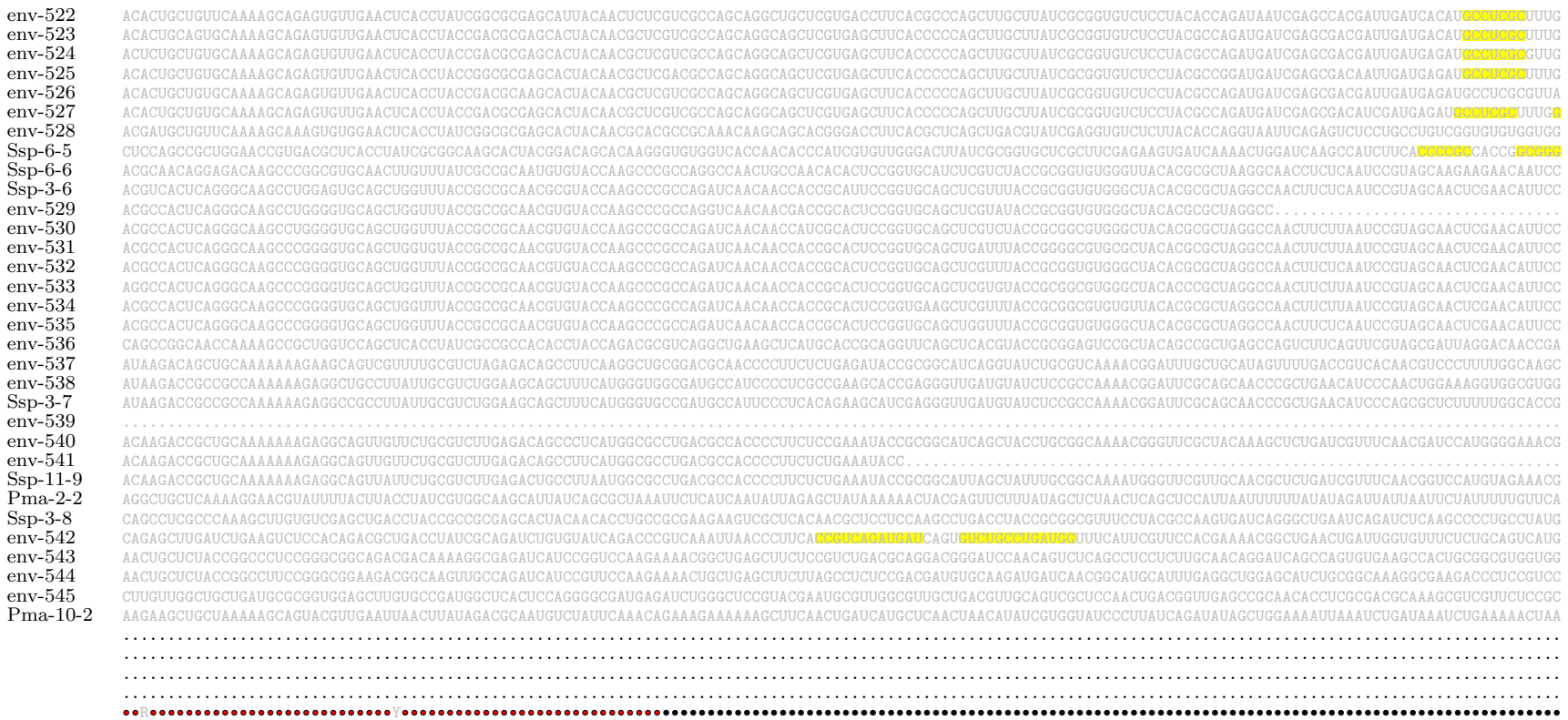
env-262
env-263
env-264
env-265
env-266
env-267
env-268
env-269
env-270
env-271
env-272
env-273
env-274
env-275
env-276
env-277
env-278
env-279
env-280
env-281
env-282
env-283
env-284
env-285
env-286
env-287
env-288
env-289
env-290
env-291
env-292
env-293
env-294
env-295
env-296
Pma-7-2
env-297
env-298
env-299
env-300
env-301
env-302
env-303
env-304
env-305
env-306
2,6Pma-11-3
env-307
env-308
env-309
env-310
env-311
env-312
env-313
env-314
env-315
env-316
env-317
env-318
env-319

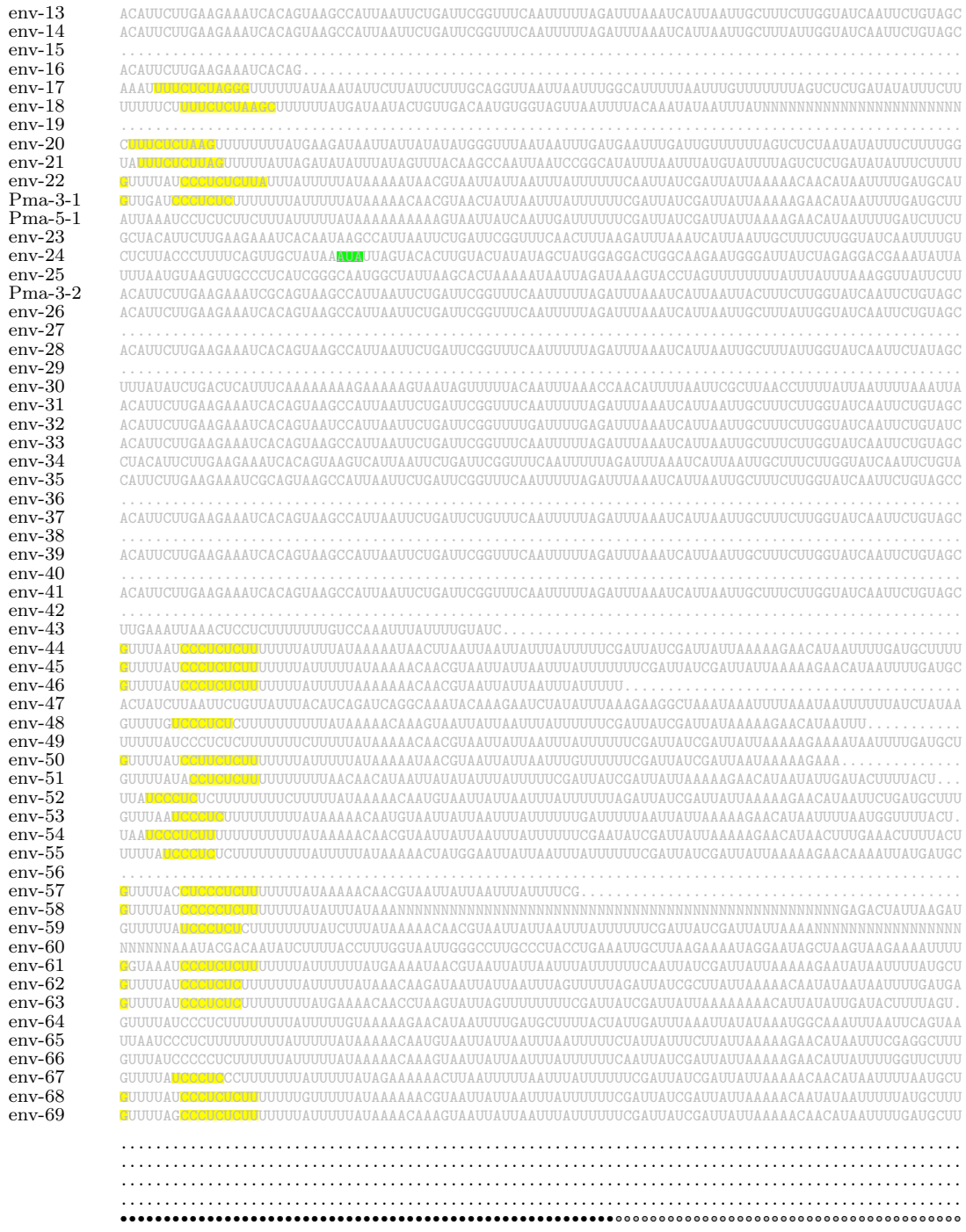

env-320
env-321
env-322
env-323
env-324
Pma-5-3
env-325
env-326
env-327
env-328
Pma-1-3
env-329
env-330
env-331
env-332
env-333
env-334
env-335
env-336
env-337
env-338
env-339
env-340
env-341
env-342
env-343
env-344
env-345
env-346
env-347
env-348
env-349
env-350
env-351
env-352
env-353
env-354
env-355
env-356
env-357
env-358
env-359
env-360
Pma-3-4
env-361
env-362
env-363
env-364
env-365
env-366
env-367
env-368
env-369
env-370
env-371
env-372
env-373
Ssp-6-3
Pma-2-1
env-374

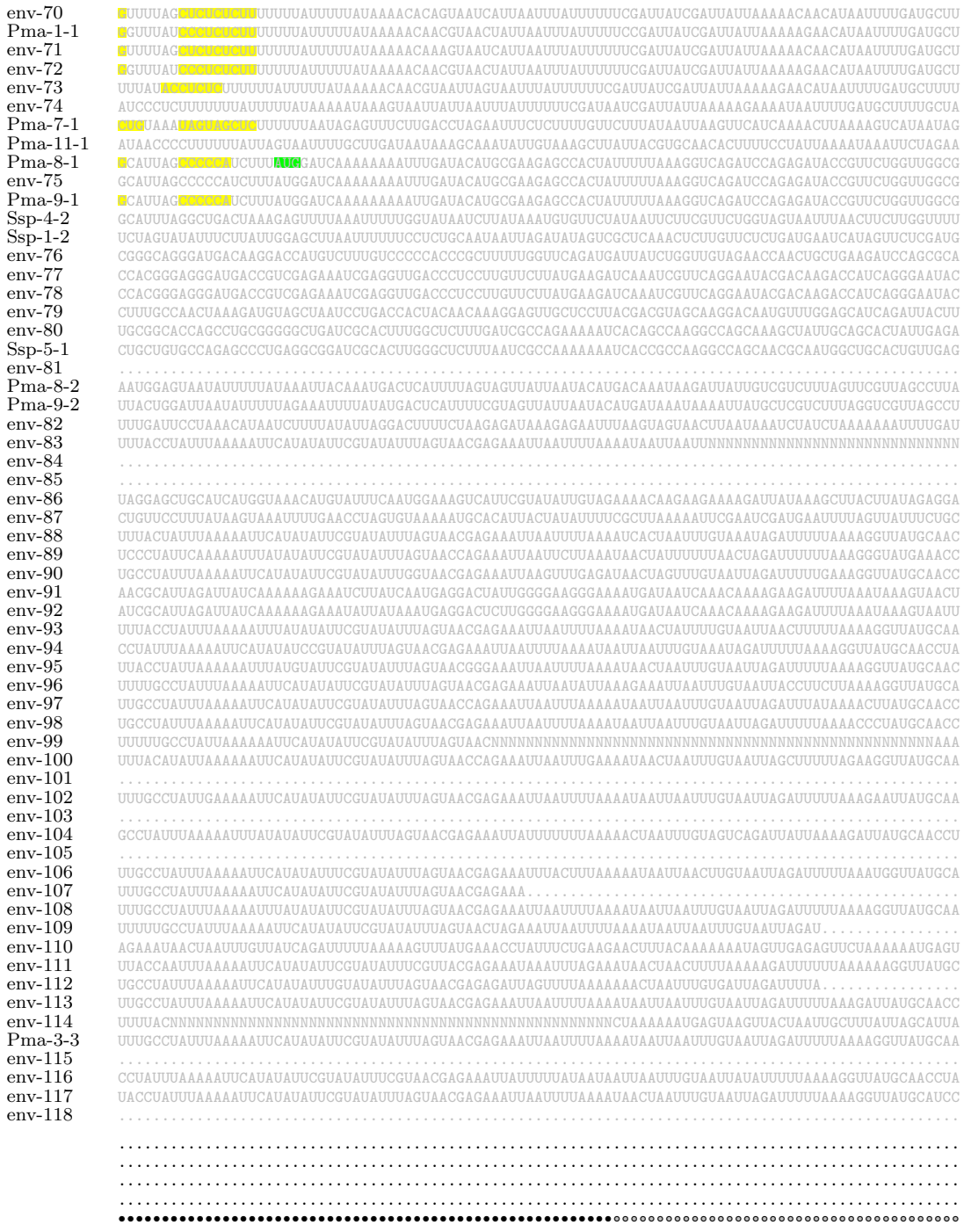

env-375
env-376
env-377
env-378
env-379
env-380
env-381
env-382
env-383
env-384
env-385
env-386
env-388
env-389
env-390
env-391
Ssp-8-2
Ssp-4-4
Pma-4-2
Pma-6-2
env-392
env-393
env-394
env-395
env-396
env-397
env-398
env-399
env-400
env-401
env-402
env-403
env-404
env-405
env-406
env-407
env-408
env-409
env-410
env-411
env-412
env-413
env-414
env-415
env-416
env-417
Pma-5-4
env-418
env-419
env-420
env-421
env-422
env-423
env-424
env-425
env-426
env-427
env-428
env-429
env-430

env-431
env-432
env-433
Pma-3-5
env-434
env-435
env-436
env-437
env-438
env-439
env-440
env-441
env-442
Pma-1-4
env-443
env-444
env-445
env-446
env-447
env-448
env-449
env-450
env-451
env-452
env-453
env-454
env-455
env-456
env-457
env-458
2,4Pma-11-4
env-459
Pma-7-3
env-460
Pma-9-4
env-461
Pma-8-4
Pma-8-5
Pma-9-5
2,5Pma-10-1
env-462
env-463
Pma-8-6
Pma-9-6
env-464
env-465
env-466
env-467
env-468
env-469
env-470
env-471
env-472
env-473
env-474
env-475
env-476
env-477
env-478
env-479

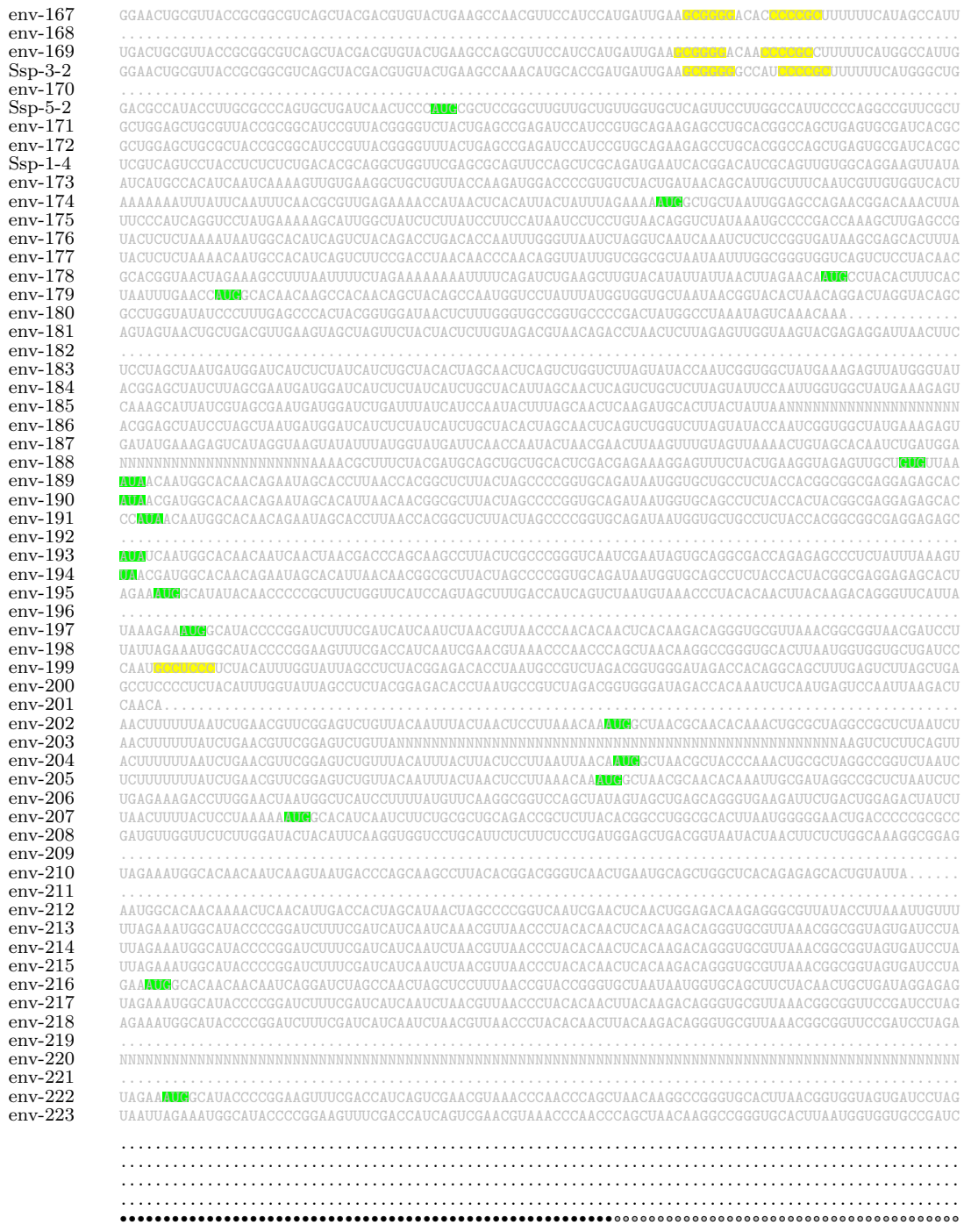

env-480
env-481
Ssp-11-3
Ssp-3-3
env-482
Ssp-3-4
env-483
env-484
env-485
env-486
Ssp-11-4
env-487
Ssp-11-5
env-488
env-489
Ssp-2-2
env-490
env-491
Ssp-1-6
Ssp-4-5
env-492
Ssp-11-6
env-493
Ssp-2-3
Ssp-4-6
Ssp-1-7
env-494
env-495
env-496
env-497
env-498
env-499
env-500
env-501
env-502
env-503
env-504
env-505
Ssp-11-7
env-506
env-507
3 env-508
env-509
env-510
env-511
env-512
env-513
env-514
Ssp-11-8
1 Ssp-10-2
env-515
env-516
env-517
env-518
env-519
Ssp-3-5
env-520
env-521
Ssp-6-4
Ssp-5-3

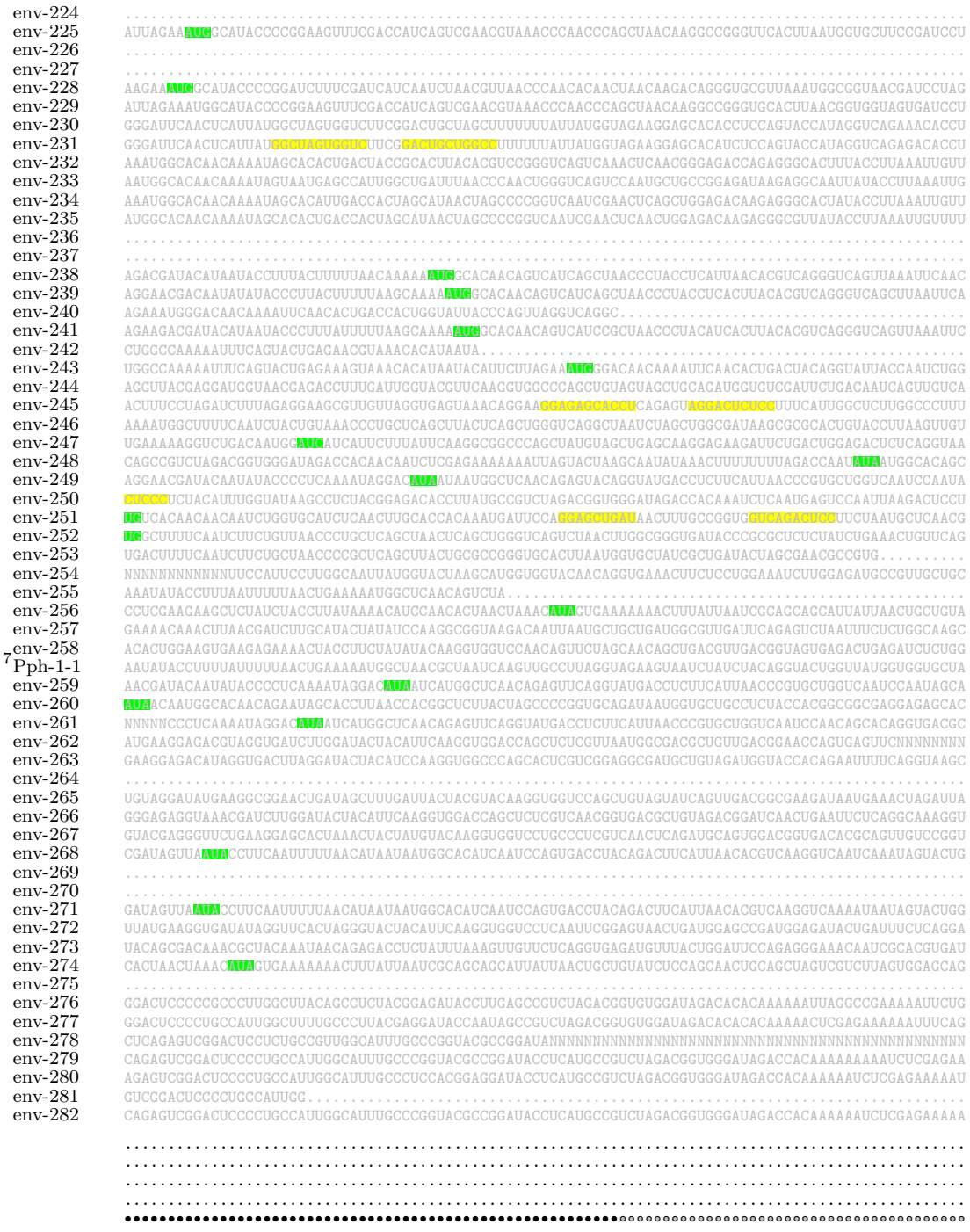

env-522
env-523
env-524
env-525
env-526
env-527
env-528
Ssp-6-5
Ssp-6-6
Ssp-3-6
env-529
env-530
env-531
env-532
env-533
env-534
env-535
env-536
env-537
env-538
Ssp-3-7
env-539
env-540
env-541
Ssp-11-9
Pma-2-2
Ssp-3-8
env-542
env-543
env-544
env-545
Pma-10-2
R
Y
alignment positions 471 · · · 567
env-1
env-2
env-3
Ssp-3-1
Ssp-6-1
Ssp-2-1
env-4
env-5
env-6
env-7
env-8
Ssp-1-1
Ssp-4-1
env-9
Pma-6-1
Pma-4-1
env-10
env-11
env-12

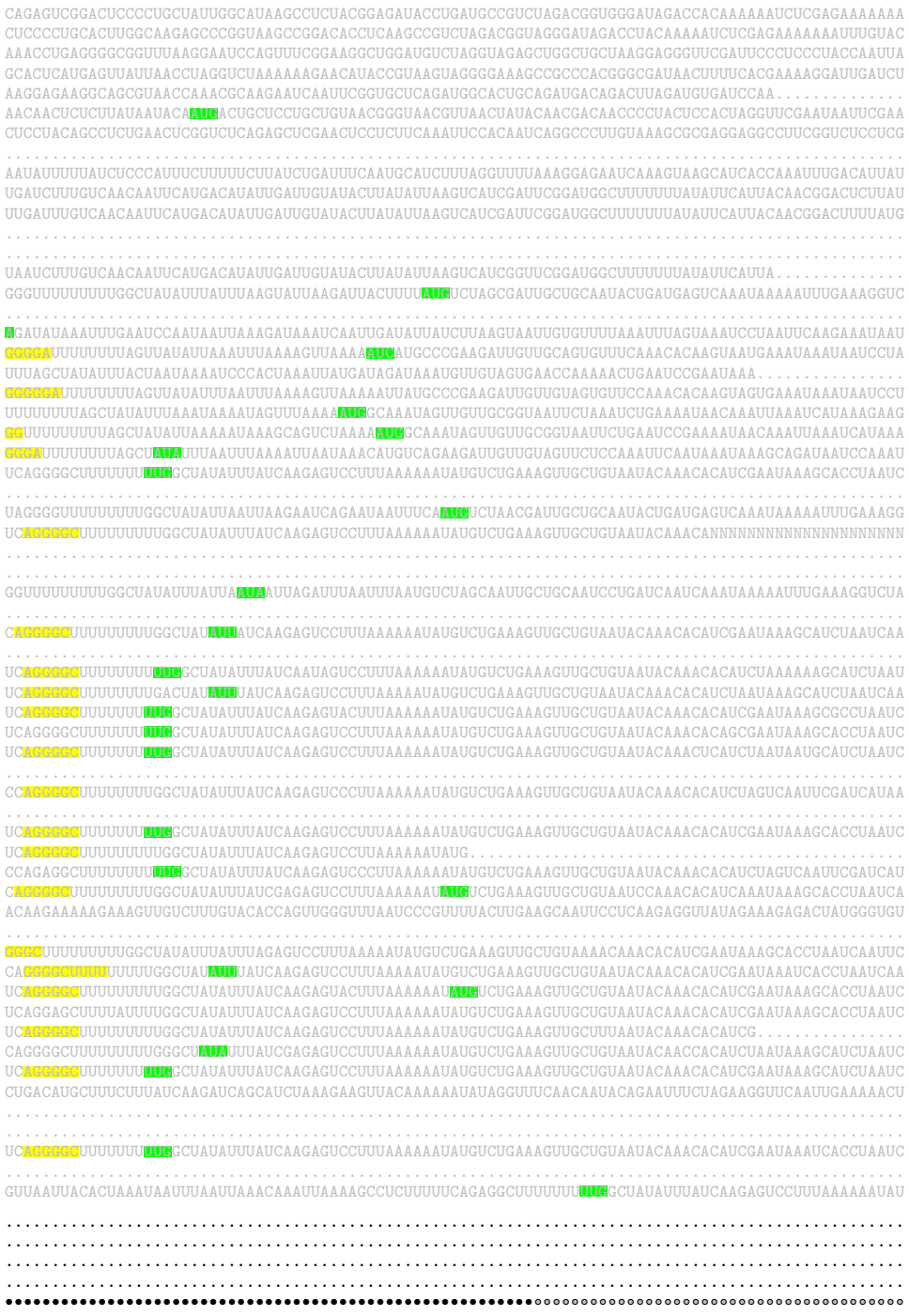

env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
Pma-3-1
Pma-5-1
env-23
env-24
env-25
Pma-3-2
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69

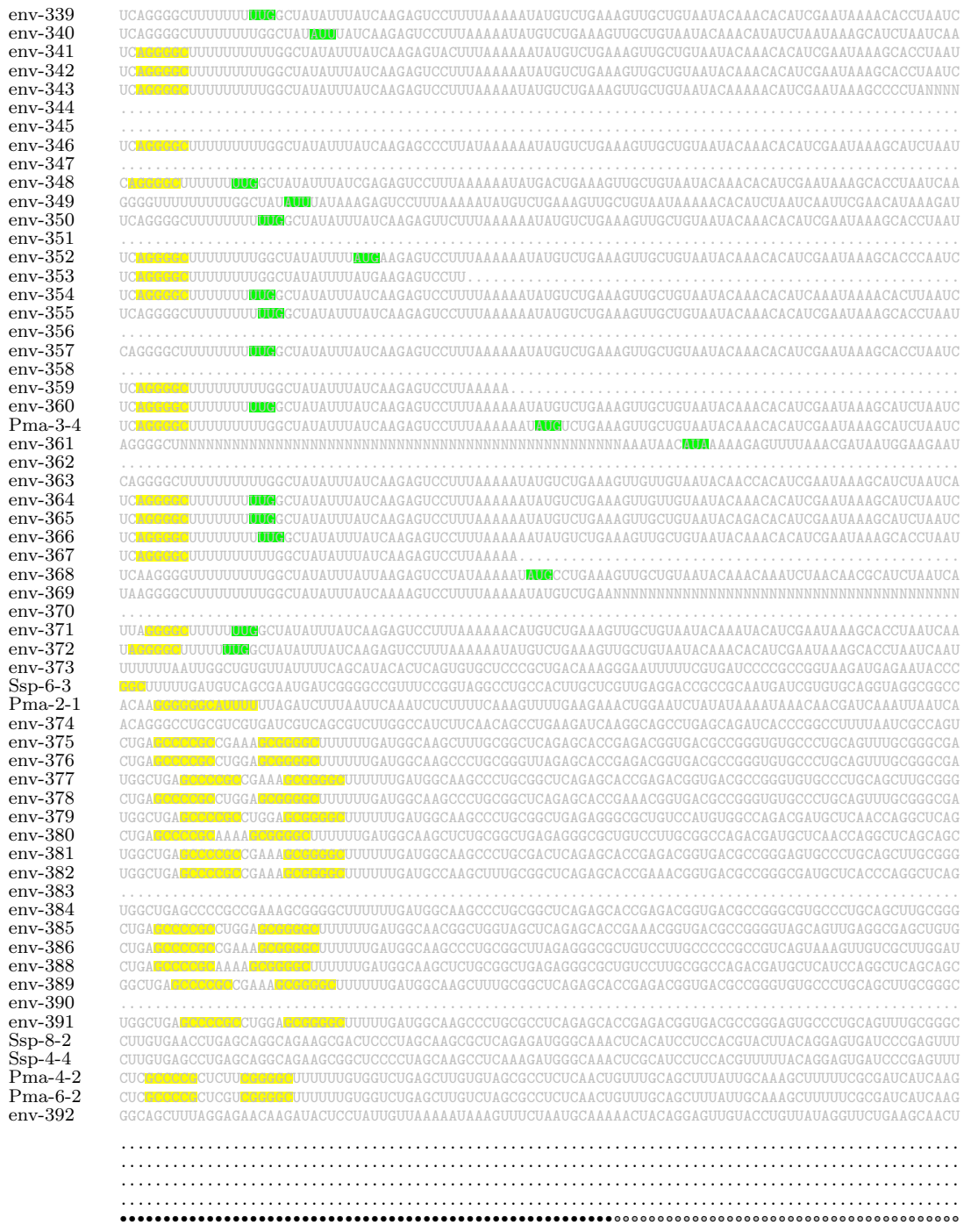

env-70
Pma-1-1
env-71
env-72
env-73
env-74
Pma-7-1
Pma-11-1
Pma-8-1
env-75
Pma-9-1
Ssp-4-2
Ssp-1-2
env-76
env-77
env-78
env-79
env-80
Ssp-5-1
env-81
Pma-8-2
Pma-9-2
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
Pma-3-3
env-115
env-116
env-117
env-118

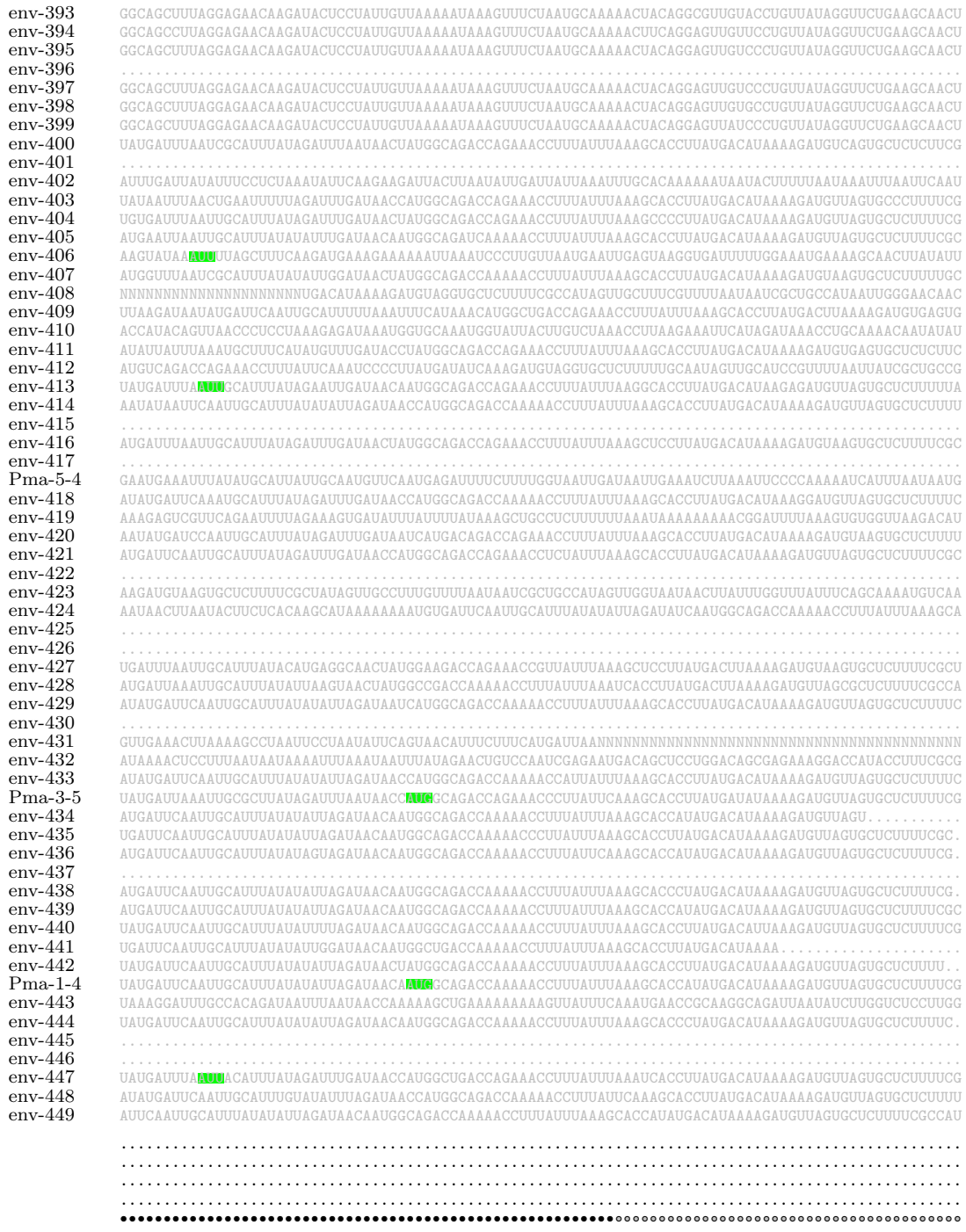
env-119
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-134
env-135
env-136
env-137
env-138
Pma-1-2
env-139
env-140
env-141
Pma-5-2
env-142
Pma-11-2
env-143
env-144
env-145
env-146
env-147
Ssp-11-1
env-148
env-149
env-150
env-151
env-152
env-153
Pma-8-3
Pma-9-3
env-154
Ssp-8-1
Ssp-4-3
env-155
Ssp-11-2
env-156
Ssp-10-1
Ssp-6-2
Ssp-9-1
env-157
env-158
env-159
env-160
env-161
env-162
Ssp-7-1
env-163
env-164
env-165
env-166

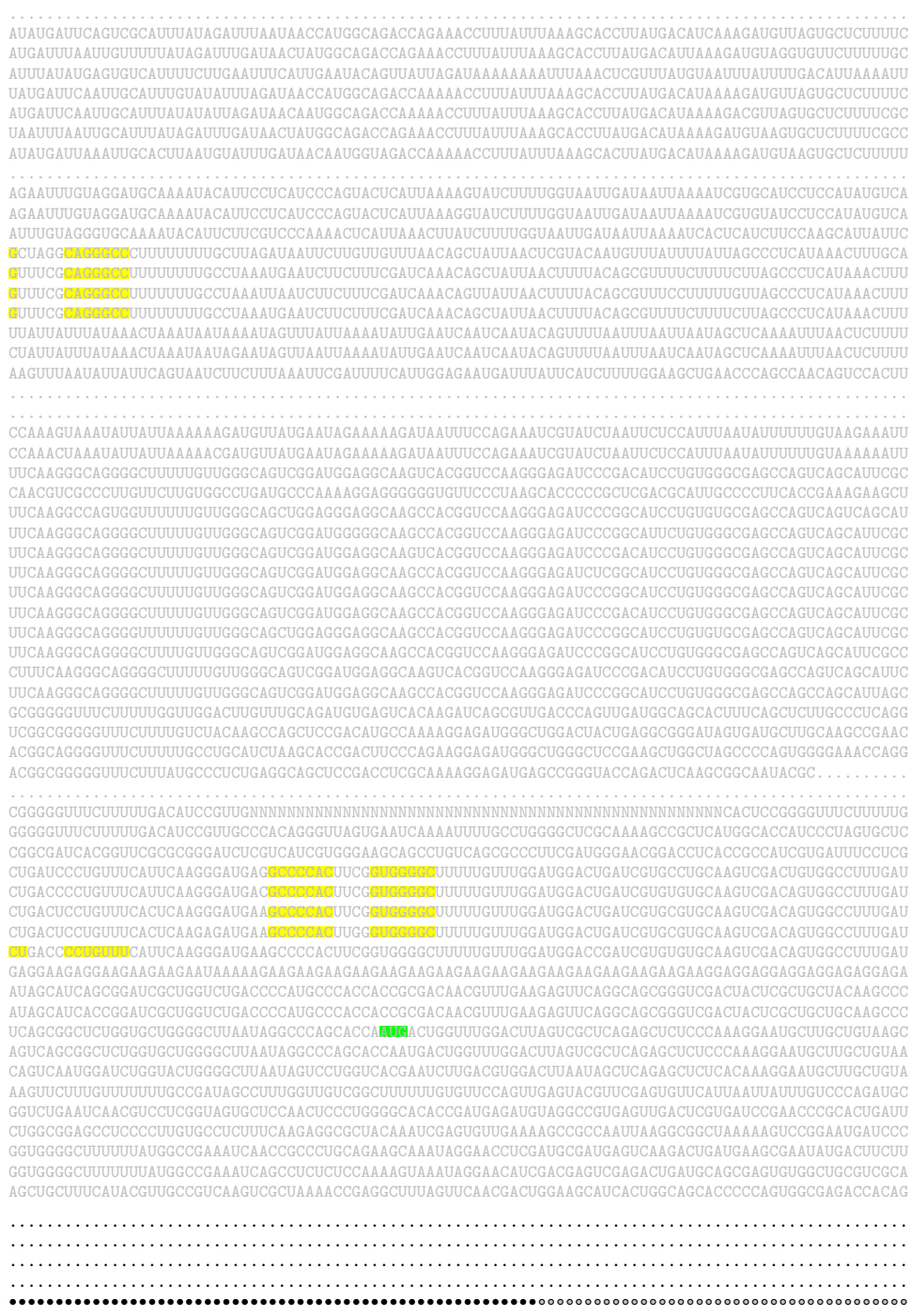

env-167
env-168
env-169
Ssp-3-2
env-170
Ssp-5-2
env-171
env-172
Ssp-1-4
env-173
env-174
env-175
env-176
env-177
env-178
env-179
env-180
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
env-190
env-191
env-192
env-193
env-194
env-195
env-196
env-197
env-198
env-199
env-200
env-201
env-202
env-203
env-204
env-205
env-206
env-207
env-208
env-209
env-210
env-211
env-212
env-213
env-214
env-215
env-216
env-217
env-218
env-219
env-220
env-221
env-222
env-223

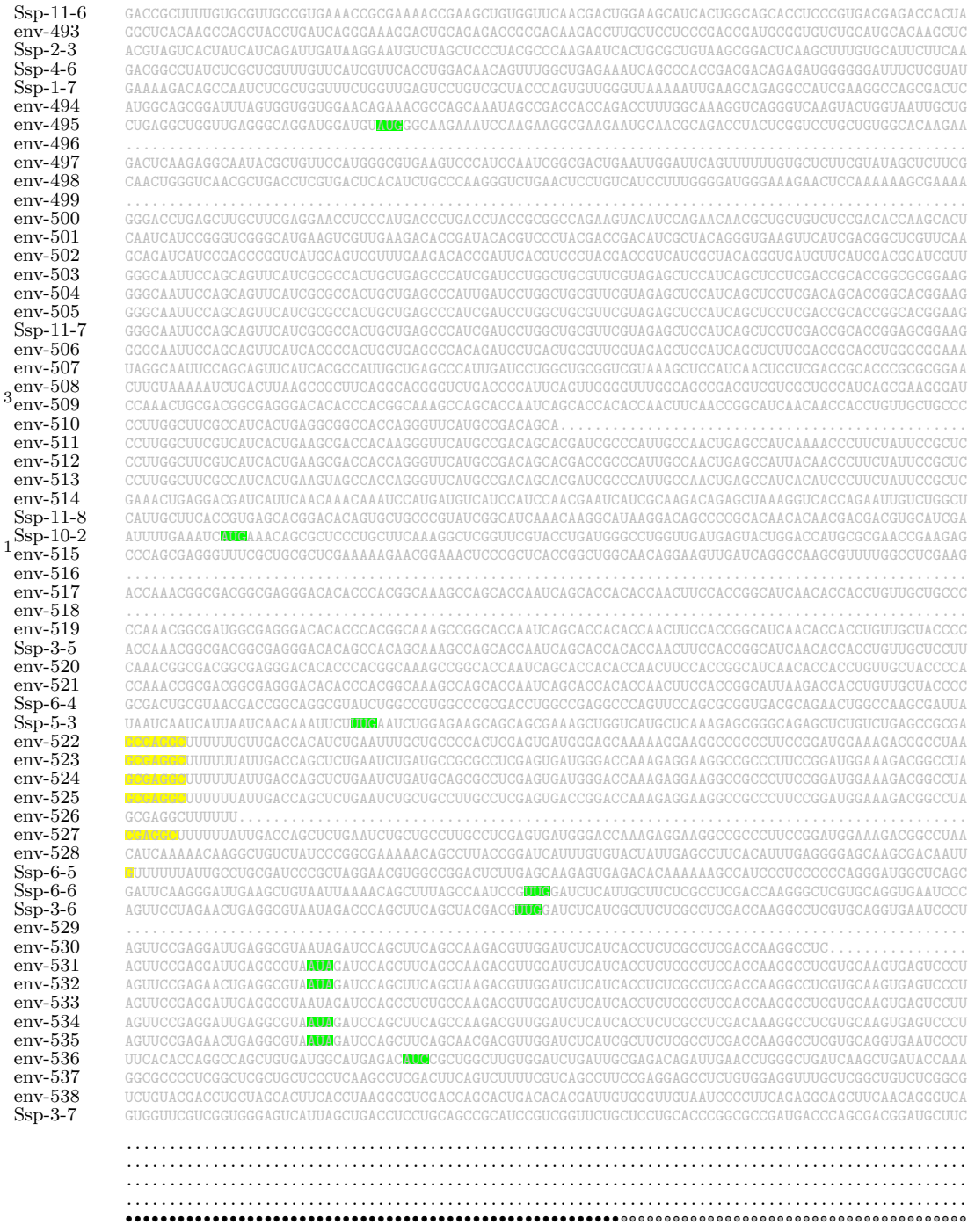

env-224
env-225
env-226
env-227
env-228
env-229
env-230
env-231
env-232
env-233
env-234
env-235
env-236
env-237
env-238
env-239
env-240
env-241
env-242
env-243
env-244
env-245
env-246
env-247
env-248
env-249
env-250
env-251
env-252
env-253
env-254
env-255
env-256
env-257
env-258
7Pph-1-1
env-259
env-260
env-261
env-262
env-263
env-264
env-265
env-266
env-267
env-268
env-269
env-270
env-271
env-272
env-273
env-274
env-275
env-276
env-277
env-278
env-279
env-280
env-281
env-282

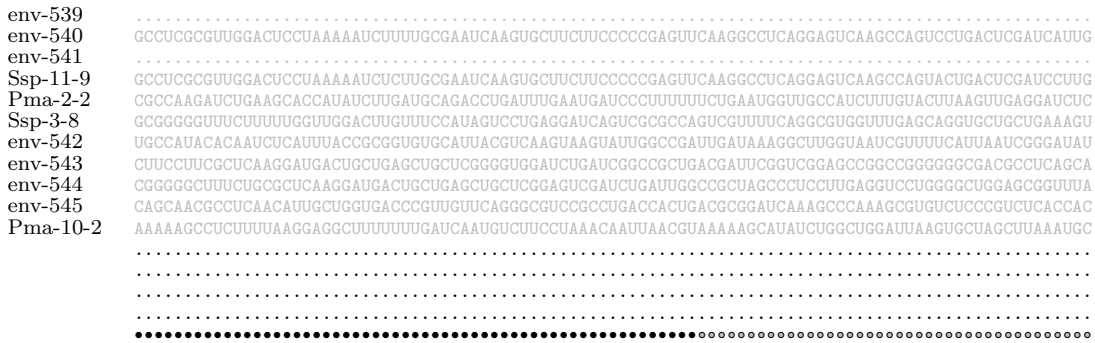

env-283
env-284
env-285
env-286
env-287
env-288
env-289
env-290
env-291
env-292
env-293
env-294
env-295
env-296
Pma-7-2
env-297
env-298
env-299
env-300
env-301
env-302
env-303
env-304
env-305
env-306
2,6Pma-11-3
env-307
env-308
env-309
env-310
env-311
env-312
env-313
env-314
env-315
env-316
env-317
env-318
env-319
env-320
env-321
env-322
env-323
env-324
Pma-5-3
env-325
env-326
env-327
env-328
Pma-1-3
env-329
env-330
env-331
env-332
env-333
env-334
env-335
env-336
env-337
env-338


env-339
env-340
env-341
env-342
env-343
env-344
env-345
env-346
env-347
env-348
env-349
env-350
env-351
env-352
env-353
env-354
env-355
env-356
env-357
env-358
env-359
env-360
Pma-3-4
env-361
env-362
env-363
env-364
env-365
env-366
env-367
env-368
env-369
env-370
env-371
env-372
env-373
Ssp-6-3
Pma-2-1
env-374
env-375
env-376
env-377
env-378
env-379
env-380
env-381
env-382
env-383
env-384
env-385
env-386
env-388
env-389
env-390
env-391
Ssp-8-2
Ssp-4-4
Pma-4-2
Pma-6-2
env-392


env-393
env-394
env-395
env-396
env-397
env-398
env-399
env-400
env-401
env-402
env-403
env-404
env-405
env-406
env-407
env-408
env-409
env-410
env-411
env-412
env-413
env-414
env-415
env-416
env-417
Pma-5-4
env-418
env-419
env-420
env-421
env-422
env-423
env-424
env-425
env-426
env-427
env-428
env-429
env-430
env-431
env-432
env-433
Pma-3-5
env-434
env-435
env-436
env-437
env-438
env-439
env-440
env-441
env-442
Pma-1-4
env-443
env-444
env-445
env-446
env-447
env-448
env-449


env-450
env-451
env-452
env-453
env-454
env-455
env-456
env-457
env-458
2,4 Pma-11-4
env-459
Pma-7-3
env-460
Pma-9-4
env-461
Pma-8-4
Pma-8-5
Pma-9-5
2,5 Pma-10-1
env-462
env-463
Pma-8-6
Pma-9-6
env-464
env-465
env-466
env-467
env-468
env-469
env-470
env-471
env-472
env-473
env-474
env-475
env-476
env-477
env-478
env-479
env-480
env-481
Ssp-11-3
Ssp-3-3
env-482
Ssp-3-4
env-483
env-484
env-485
env-486
Ssp-11-4
env-487
Ssp-11-5
env-488
env-489
Ssp-2-2
env-490
env-491
Ssp-1-6
Ssp-4-5
env-492


Ssp-11-6
env-493
Ssp-2-3
Ssp-4-6
Ssp-1-7
env-494
env-495
env-496
env-497
env-498
env-499
env-500
env-501
env-502
env-503
env-504
env-505
Ssp-11-7
env-506
env-507
3 env-508
env-509
env-510
env-511
env-512
env-513
env-514
Ssp-11-8
1 Ssp-10-2
env-515
env-516
env-517
env-518
env-519
Ssp-3-5
env-520
env-521
Ssp-6-4
Ssp-5-3
env-522
env-523
env-524
env-525
env-526
env-527
env-528
Ssp-6-5
Ssp-6-6
Ssp-3-6
env-529
env-530
env-531
env-532
env-533
env-534
env-535
env-536
env-537
env-538
Ssp-3-7



30 *epsC*

30.1 Taxa

The taxonomy of each organism containing a putative *epsC* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *epsC* RNA in Sections 30.2 and 30.4.

abbrev. of hits	taxonomy of species
Bam-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus amyloliquefaciens* FZB42
Bcl-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus clausii* KSM-K16
Bco-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus coagulans* 36D1
Bco-2-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus coahuilensis m*4-4
Bha-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus halodurans* C-125
Bli-1-1 to Bli-1-2	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus licheniformis* ATCC 14580 (DSM 13)
Bpu-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus pumilus* ATCC 7061
Bpu-2-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus pumilus* SAFR-032
Bse-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus selenitireducens* MLS10
Bsp-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus sp.* NRRL B-14911
Bsu-1-1 to Bsu-1-2	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus subtilis subsp. subtilis str.* 168
Bsu-2-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus subtilis subsp. subtilis str.* JH642
Bsu-3-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus subtilis subsp. subtilis str.* NCIB 3610
Bsu-4-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus subtilis subsp. subtilis str.* SMY
Bce-1-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* 03BB108
Bce-2-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* AH1134
Bce-3-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* AH187
Bce-4-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* ATCC 10987
Bce-5-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* ATCC 14579
Bce-6-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* G9842
Bce-7-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* H3081.97
Bce-8-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* NVH0597-99
Bth-1-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus thuringiensis serovar israelensis* ATCC 35646
Bth-2-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus thuringiensis serovar konkukian str.* 97-27
Bth-3-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus thuringiensis str.* Al Hakam
Bwe-1-1 to Bwe-1-2	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus weihenstephanensis* KBAB4
Gka-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Geobacillus kaustophilus* HTA426
Gsp-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Geobacillus sp.* G11MC16
Gsp-2-1	Bacteria Firmicutes Bacillales Bacillaceae *Geobacillus sp.* WCH70
Gsp-3-1	Bacteria Firmicutes Bacillales Bacillaceae *Geobacillus sp.* Y412MC10
Gth-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Geobacillus thermodenitrificans* NG80-2
Psp-1-1	Bacteria Firmicutes Bacillales Paenibacillaceae *Paenibacillus sp.* JDR-2

30.2 Gene contexts

Each *epsC* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *epsC* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 30.3. The accession number of the sequence containing each *epsC* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 30.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Bse-1-1	NZ_ABHZ01000002.1	+	284014	284130	**RNA**→ COG1086 (COG1086)→ GalU (COG1210)→
Gsp-3-1	NZ_ABRG01000007.1	-	296142	296022	**RNA**→ GalU (COG1210)→ WcaJ (COG2148)→ WcaG (COG0451)→ Gmd (COG1089)→ RfbX (COG2244)→ Glycos_transf_2 (pfam00535)→ Glycos_transf_2 (pfam00535)→ hypo→ hypo→ RfaG (COG0438)→
Bcl-1-1	NC_006582.1	-	3853484	3853372	**RNA**→ GalU (COG1210)→ WcaJ (COG2148)→
Bha-1-1	NC_002570.2	-	3844415	3844295	**RNA**→ COG1086 (COG1086)→ GalU (COG1210)→
Psp-1-1	NZ_ABKS01000003.1	+	262896	263009	**RNA**→ GalU (COG1210)→ WcaJ (COG2148)→ WcaG (COG0451)→ Gmd (COG1089)→ RfaG (COG0438)→ WbbJ (COG0110)→ RfaG (COG0438)→ Glycos_transf_2 (pfam00535)→
Bco-2-1	NZ_ABFU01000061.2	+	6074	6193	**RNA**→ hypo→
Bsp-1-1	NZ_AAOX01000001.1	+	153591	153706	**RNA**→ Polysacc_synt_2 (pfam02719)→
Bce-2-1	NZ_ABDA02000004.1	-	319655	319532	**RNA**→ GalU (COG1210)→ Polysacc_synt_2 (pfam02719)→ WecE (COG0399)→ WcaJ (COG2148)→
Bth-1-1	NZ_AAJM01000088.1	+	3912	4034	**RNA**→ GalU (COG1210)→ WcaJ (COG2148)→ RfbX (COG2244)→
Bce-6-1	NZ_ABDJ01000008.1	+	39213	39337	**RNA**→ GalU (COG1210)→ WcaJ (COG2148)→ COG5360 (COG5360)→ RfaG (COG0438)→ hypo→ MviN (COG0728)→ RfaG (COG0438)→ WecC (COG0677)→
Bce-7-1	NZ_ABDL02000017.1	+	36572	36696	**RNA**→ GalU (COG1210)→ Glycos_transf_2 (pfam00535)→ WcaJ (COG2148)→ Wzy_C (pfam04932)→ RfaG (COG0438)→ RfaG (COG0438)→ Glycos_transf_2 (pfam00535)→
Bwe-1-1	NC_010184.1	-	5040395	5040271	**RNA**→ GalU (COG1210)→ Polysacc_synt_2 (pfam02719)→ WcaJ (COG2148)→ WcaG (COG0451)→ RfaG (COG0438)→ Ugd (COG1004)→ Epimerase (pfam01370)→ hypo→
Bce-3-1	NZ_AAUF01000043.1	-	27981	27857	**RNA**→ GalU (COG1210)→ Polysacc_synt_2 (pfam02719)→ WcaJ (COG2148)→ WcaG (COG0451)→ RfaG (COG0438)→ COG1086 (COG1086)→ WcaG (COG0451)RfbC (COG1898)→ WecB (COG0381)→ RfaG (COG0438)→
Bce-5-1	NC_004722.1	-	5186485	5186361	**RNA**→ GalU (COG1210)→ Polysacc_synt_2 (pfam02719)→ WecE (COG0399)→ AccC (COG0439)→ WcaG (COG0451)→ WcaJ (COG2148)→ Glycos_transf_1 (pfam00534)→ secreted polysaccharide polymerase→ Glycos_transf_2 (pfam00535)→
Bth-3-1	NC_008600.1	-	5036733	5036609	**RNA**→ GalU (COG1210)→ WcaJ (COG2148)→ RfbX (COG2244)→ RfaG (COG0438)→ hypo→ Ugd (COG1004)→ RfaG (COG0438)→ RfaG (COG0438)→ ManC (COG0662)ManC (COG0836)→
Bth-2-1	NC_005957.1	-	5012340	5012216	**RNA**→ GalU (COG1210)→ WcaJ (COG2148)→ COG5360 (COG5360)NOG18105 (NOG18105)→ RfaG (COG0438)→ hypo→ MviN (COG0728)→ RfaG (COG0438)→ WecC (COG0677)→
Bce-4-1	NC_003909.8	-	4986228	4986105	**RNA**→ GalU (COG1210)→ Polysacc_synt_2 (pfam02719)→ WecE (COG0399)→ WcaJ (COG2148)→ Glycos_transf_2 (pfam00535)→ RfaG (COG0438)→
Bce-1-1	NZ_ABDM02000010.1	+	36823	36946	**RNA**→ GalU (COG1210)→ WcaJ (COG2148)→ RfbX (COG2244)→ Wzy_C (pfam04932)→ Glycos_transf_2 (pfam00535)→ Asn_Synthase_B_C (cd01991)→ Glycos_transf_2 (pfam00535)→ Ugd (COG1004)→
Bli-1-1	NC_006270.3	-	3521520	3521404	**RNA**→ Polysacc_synt_2 (pfam02719)→ RfaG (COG0438)→ Glycos_transf_2 (pfam00535)→ RfaG (COG0438)→ putative capsular polysaccharide biosynthesis protein YveQ→ Glycos_transf_2 (pfam00535)→ COG5039 (COG5039)→ Glycos_transf_2 (pfam00535)→ Polysacc_synt (pfam01943)→
Bli-1-2	NC_006322.1	-	3521372	3521256	**RNA**→ Polysacc_synt_2 (pfam02719)→ RfaG (COG0438)→ Glycos_transf_2 (pfam00535)→ RfaG (COG0438)→ YveQ→ Glycos_transf_2 (pfam00535)→ COG5039 (COG5039)→ Glycos_transf_2 (pfam00535)→ Polysacc_synt (pfam01943)→
Bam-1-1	NC_009725.1	-	3285925	3285809	**RNA**→ Polysacc_synt_2 (pfam02719)→ Glycos_transf_1 (pfam00534)→ Glycos_transf_2 (pfam00535)→ RfaG (COG0438)→ EpsG→ Glycos_transf_2 (pfam00535)→ COG5039 (COG5039)→ Glycos_transf_2 (pfam00535)→ RfbX (COG2244)→

Bsu-1-1	NC_000964.2	-	3527482	3527366	**RNA**→ Polysacc_synt_2 (pfam02719)→ Glycos_transf_1 (pfam00534)→ Glycos_transf_2 (pfam00535)→ RfaG (COG0438)→ hypo→ Glycos_transf_2 (pfam00535)→ COG5039 (COG5039)→ Glycos_transf_2 (pfam00535)→ hypo→
Bsu-1-2	NZ_ABQK01000005.1	-	2509005	2508889	**RNA**→
Bsu-3-1	NZ_ABQL01000005.1	-	2509049	2508933	**RNA**→
Bsu-2-1	NZ_ABQM01000009.1	-	352827	352711	**RNA**→
Bsu-4-1	NZ_ABQN01000009.1	-	352835	352719	**RNA**→
Bpu-2-1	NC_009848.1	-	3064659	3064548	**RNA**→ Polysacc_synt_2 (pfam02719)→ Glycos_transf_1 (pfam00534)→ Glycos_transf_2 (pfam00535)→ RfaG (COG0438)→ hypo→ Glycos_transf_2 (pfam00535)→ COG5039 (COG5039)→ RfbX (COG2244)→ WcaJ (COG2148)→
Bpu-1-1	NZ_ABRX01000002.1	-	446041	445929	**RNA**→ Polysacc_synt_2 (pfam02719)→ Glycos_transf_1 (pfam00534)→ Glycos_transf_2 (pfam00535)→
Bce-8-1	NZ_ABDK02000003.1	-	291375	291262	**RNA**→ Polysacc_synt_2 (pfam02719)→ RfaG (COG0438)→ Glycos_transf_1 (pfam00534)→ Glycos_transf_2 (pfam00535)→ capsular polysaccharide biosynthesis protein→ OCH1 (COG3774)→ RfaJ (COG1442)→ Glycos_transf_1 (pfam00534)→ Glycos_transf_2 (pfam00535)→
Bwe-1-2	NC_010184.1	-	3464620	3464507	**RNA**→ Polysacc_synt_2 (pfam02719)→ RfaG (COG0438)→ Glycos_transf_1 (pfam00534)→ Glycos_transf_2 (pfam00535)→ capsular polysaccharide biosynthesis protein→ Glycos_transf_1 (pfam00534)→ Glycos_transf_1 (pfam00534)→ Glycos_transf_1 (pfam00534)→
Gka-1-1	NC_006510.1	-	3368754	3368630	**RNA**→ hypo→ hypo→
Gth-1-1	NC_009328.1	-	3376553	3376429	**RNA**→ GalU (COG1210)→ WcaJ (COG2148)→
Gsp-1-1	NZ_ABVH01000011.1	-	99992	99868	**RNA**→ GalU (COG1210)→
Gsp-2-1	NZ_ABLJ01000003.1	+	87590	87714	**RNA**→ hypo (cons)→
Bco-1-1	NZ_AAWV01000020.1	-	35683	35572	**RNA**→ GalU (COG1210)→ WcaJ (COG2148)→ RfaG (COG0438)→ WcaK (COG2327)→

30.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 30.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *epsC* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

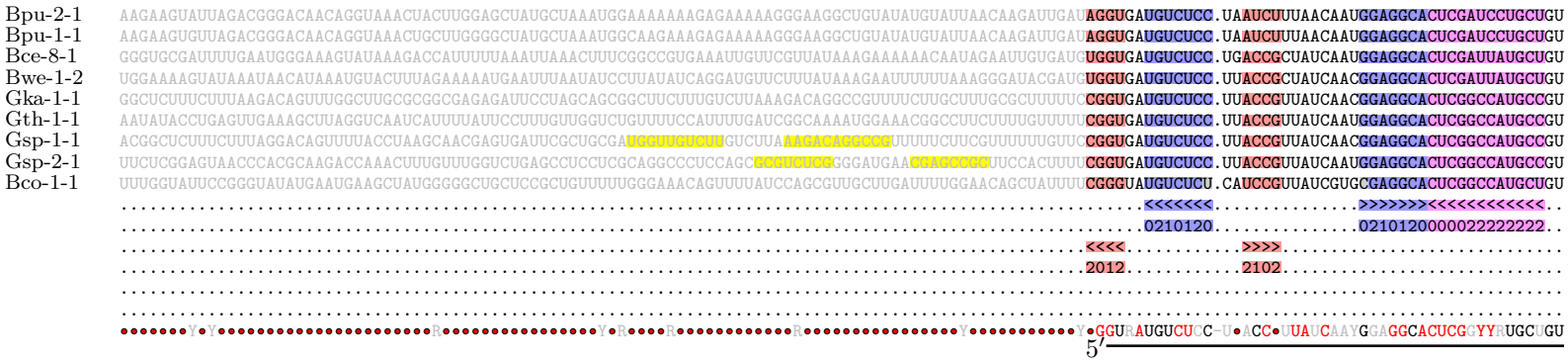
cd01991 (1) The C-terminal domain of Asparagine Synthase B.
COG0110 (1) Acetyltransferase (isoleucine patch superfamily) [General function prediction only]
COG0381 (1) UDP-N-acetylglucosamine 2-epimerase [Cell envelope biogenesis, outer membrane]
COG0399 (3) Predicted pyridoxal phosphate-dependent enzyme apparently involved in regulation of cell wall biogenesis [Cell envelope biogenesis, outer membrane]
COG0438 (26) Glycosyltransferase [Cell envelope biogenesis, outer membrane]
COG0439 (1) Biotin carboxylase [Lipid metabolism]
COG0451 (6) Nucleoside-diphosphate-sugar epimerases [Cell envelope biogenesis, outer membrane / Carbohydrate transport and metabolism]
COG0662 (1) Mannose-6-phosphate isomerase [Carbohydrate transport and metabolism]
COG0677 (2) UDP-N-acetyl-D-mannosaminuronate dehydrogenase [Cell envelope biogenesis, outer membrane]
COG0728 (2) Uncharacterized membrane protein, putative virulence factor [General function prediction only]
COG0836 (1) Mannose-1-phosphate guanylyltransferase [Cell envelope biogenesis, outer membrane]
COG1004 (3) Predicted UDP-glucose 6-dehydrogenase [Cell envelope biogenesis, outer membrane]
COG1086 (3) Predicted nucleoside-diphosphate sugar epimerases [Cell envelope biogenesis, outer membrane / Carbohydrate transport and metabolism]
COG1089 (2) GDP-D-mannose dehydratase [Cell envelope biogenesis, outer membrane]
COG1210 (19) UDP-glucose pyrophosphorylase [Cell envelope biogenesis, outer membrane]
COG1442 (1) Lipopolysaccharide biosynthesis proteins, LPS:glycosyltransferases [Cell envelope biogenesis, outer membrane]
COG1898 (1) dTDP-4-dehydrorhamnose 3,5-epimerase and related enzymes [Cell envelope biogenesis, outer membrane]
COG2148 (17) Sugar transferases involved in lipopolysaccharide synthesis [Cell envelope biogenesis, outer membrane]
COG2244 (6) Membrane protein involved in the export of O-antigen and teichoic acid [General function prediction only]
COG2327 (1) Uncharacterized conserved protein [Function unknown]
COG3774 (1) Mannosyltransferase OCH1 and related enzymes [Cell envelope biogenesis, outer membrane]
COG5039 (5) Exopolysaccharide biosynthesis protein [Carbohydrate transport and metabolism / Cell envelope biogenesis, outer membrane]
COG5360 (2) Uncharacterized protein conserved in bacteria [Function unknown]
NOG18105 (1)
pfam00534 (11) Glycosyl transferases group 1.
pfam00535 (27) Glycosyl transferase.

pfam01370 (1) NAD dependent epimerase/dehydratase family.
pfam01943 (2) Polysaccharide biosynthesis protein.
pfam02719 (14) Polysaccharide biosynthesis protein.
pfam04932 (2) O-Antigen Polymerase.

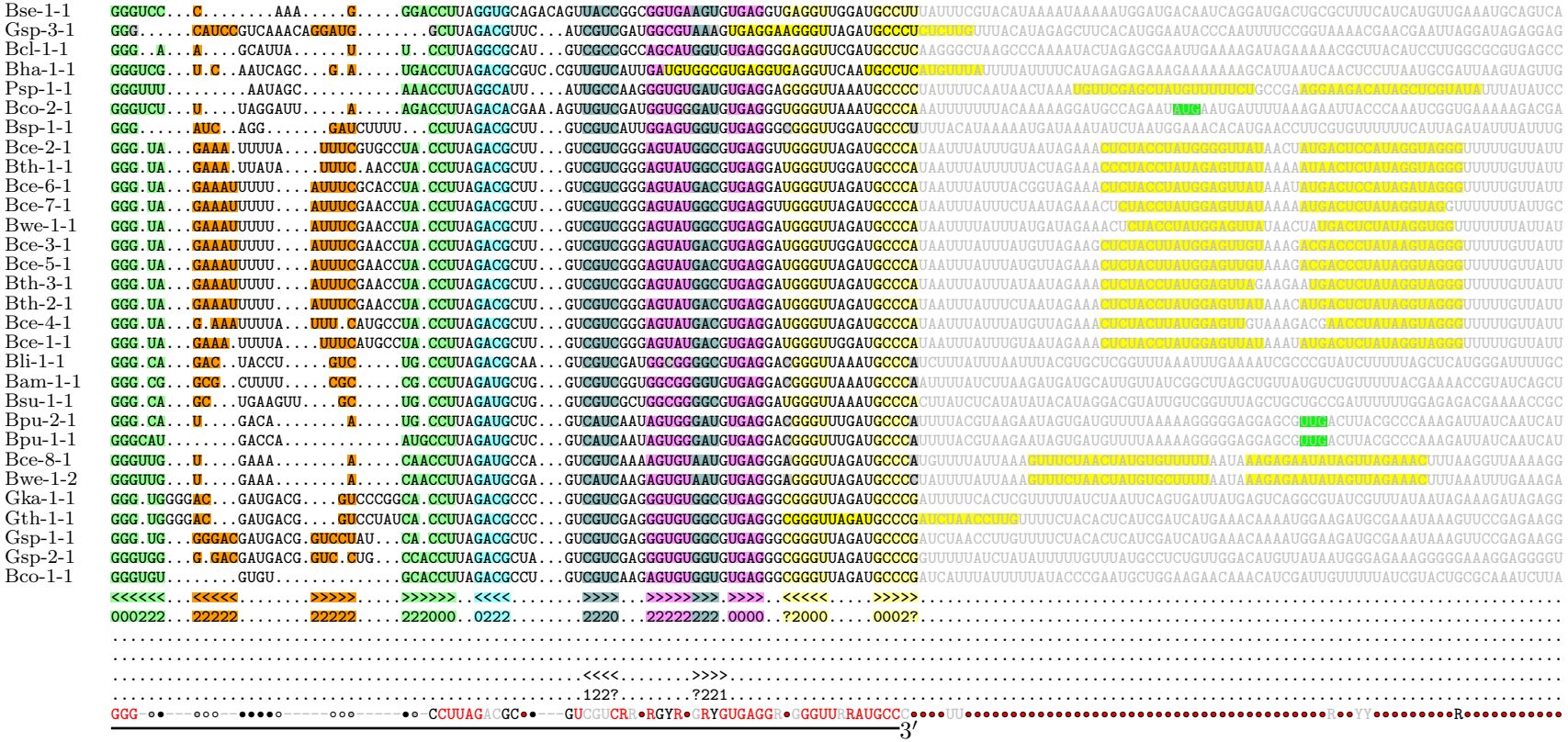
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Bli-1-2, Bsu-1-2, Bsu-2-1, Bsu-3-1, Bsu-4-1

30.4 Multiple-sequence alignment

Each *epsC* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 30.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *epsC* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

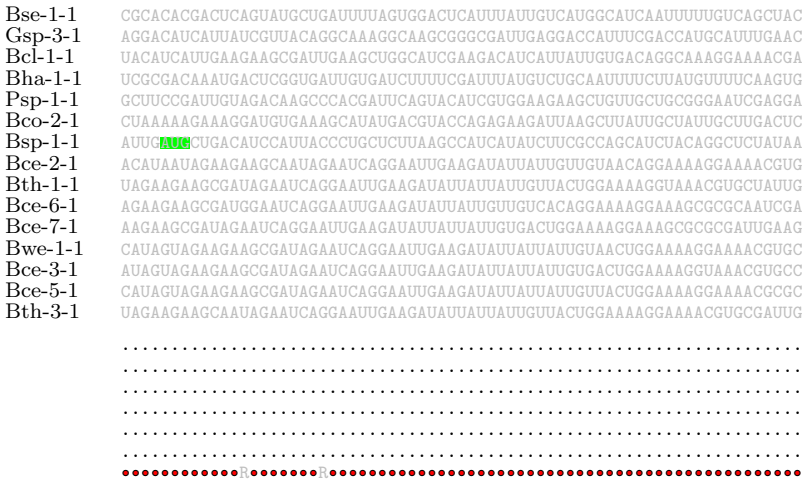


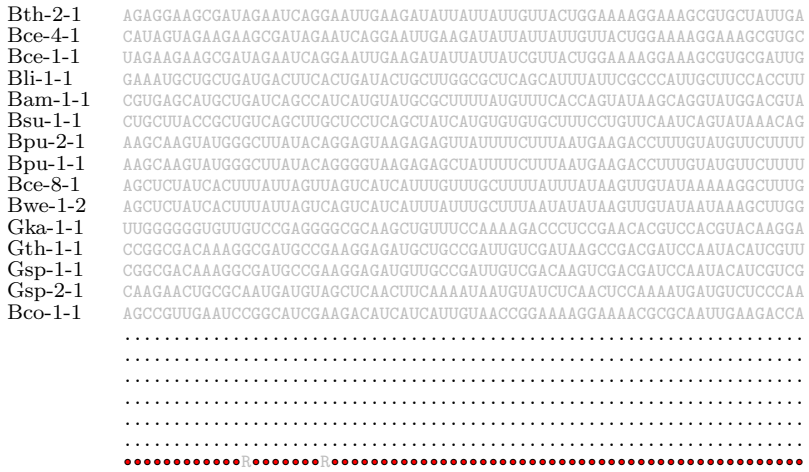
alignment positions 149 · · · 308



alignment positions 309 · · · 468

alignment positions 469 · · · 537





31 *fixA*

31.1 Taxa

The taxonomy of each organism containing a putative *fixA* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *fixA* RNA in Sections 31.2 and 31.4.

abbrev. of hits	taxonomy of species
Pen-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas entomophila* L48
Pfl-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf-5
Pfl-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf0-1
Ppu-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* F1
Ppu-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* GB-1
Ppu-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* KT2440
Ppu-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* W619
Psy-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. phaseolicola* 1448A
Psy-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000

31.2 Gene contexts

Each *fixA* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *fixA* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 31.3. The accession number of the sequence containing each *fixA* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 31.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Pen-1-1	NC_008027.1	-	3942702	3942622	**RNA**→ FixA (COG2086)→ FixB (COG2025)→ HisJ (COG0834)→ lipoprotein→ OmpA (COG2885)→
Ppu-4-1	NZ_AAVY01000007.1	+	129956	130036	**RNA**→ FixA (COG2086)→ FixB (COG2025)→ HisJ (COG0834)→ lipoprotein, putative→ OmpA (COG2885)→ ARO8 (COG1167)→ hypo (cons)→
Ppu-3-1	NC_002947.3	-	4748049	4747969	**RNA**→ FixA (COG2086)→ FixB (COG2025)→ HisJ (COG0834)→ lipoprotein, putative→ OmpA (COG2885)→ ARO8 (COG1167)→ hypo→
Ppu-2-1	NZ_AAXR01000004.1	+	282918	282998	**RNA**→ FixA (COG2086)→ FixB (COG2025)→ HisJ (COG0834)→ lipoprotein, putative→ OmpA (COG2885)→ ARO8 (COG1167)→ hypo (cons)→
Ppu-1-1	NC_009512.1	+	1856103	1856183	**RNA**→ FixA (COG2086)→ FixB (COG2025)→ HisJ (COG0834)→ hypo→ OmpA (COG2885)→ ARO8 (COG1167)→ hypo→
Pfl-2-1	NC_007492.1	-	4658805	4658726	**RNA**→ FixA (COG2086)NOG22109 (NOG22109)→ FixB (COG2025)→
Pfl-1-1	NC_004129.6	+	2020999	2021078	**RNA**→ FixA (COG2086)→ FixB (COG2025)→ PBPb (smart00062)NOG13337 (NOG13337)→ lipoprotein, putative→ OmpA (COG2885)→
Psy-1-1	NC_005773.3	+	2312729	2312819	**RNA**→ FixA (COG2086)→ FixB (COG2025)→
Psy-3-1	NC_004578.1	+	2406749	2406831	**RNA**→ FixA (COG2086)→ FixB (COG2025)→
Psy-2-1	NC_007005.1	+	2316945	2317035	**RNA**→ FixA (COG2086)→ FixB (COG2025)→

31.3 Conserved domains

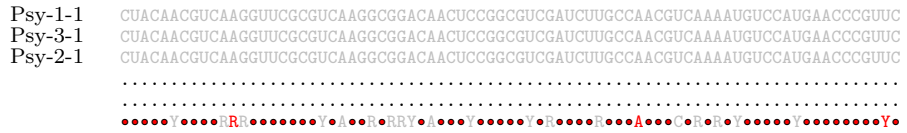
Conserved domains found in protein-coding genes listed in Section 31.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *fixA* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

COG0834 (5) ABC-type amino acid transport/signal transduction systems, periplasmic component/domain [Amino acid transport and metabolism / Signal transduction mechanisms]
COG1167 (4) Transcriptional regulators containing a DNA-binding HTH domain and an aminotransferase domain (MocR family) and their eukaryotic orthologs [Transcription / Amino acid transport and metabolism]
COG2025 (10) Electron transfer flavoprotein, alpha subunit [Energy production and conversion]
COG2086 (10) Electron transfer flavoprotein, beta subunit [Energy production and conversion]
COG2885 (6) Outer membrane protein and related peptidoglycan-associated (lipo)proteins [Cell envelope biogenesis, outer membrane]
NOG13337 (1)
NOG22109 (1)
smart00062 (1) Bacterial periplasmic substrate-binding proteins; bacterial proteins, eukaryotic ones are in PBPe

31.4 Multiple-sequence alignment

Each *fixA* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 31.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *fixA* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

alignment positions 1 · · · 99
Pen-1-1
Ppu-4-1
Ppu-3-1
Ppu-2-1
Ppu-1-1
Pfl-2-1
Pfl-1-1
Psy-1-1
Psy-3-1
Psy-2-1
alignment positions 100 · · · 259
Pen-1-1
Ppu-4-1
Ppu-3-1
Ppu-2-1
Ppu-1-1
Pfl-2-1
Pfl-1-1
Psy-1-1
Psy-3-1
Psy-2-1
5′
3′
alignment positions 260 · · · 419
Pen-1-1
Ppu-4-1
Ppu-3-1
Ppu-2-1
Ppu-1-1
Pfl-2-1
Pfl-1-1
Psy-1-1
Psy-3-1
Psy-2-1
alignment positions 420 · · · 498
Pen-1-1
Ppu-4-1
Ppu-3-1
Ppu-2-1
Ppu-1-1
Pfl-2-1
Pfl-1-1



32 Flavo-1

32.1 Taxa

The taxonomy of each organism containing a putative Flavo-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Flavo-1 RNA in Sections 32.2 and 32.4.

abbrev. of hits	taxonomy of species
Bfr-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides fragilis* NCTC 9343
Bth-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides thetaiotaomicron* VPI-5482
ueu-1-1	Bacteria Bacteroidetes environmental samples unidentified eubacterium SCB49
Fba-1-1 to Fba-1-2	Bacteria Bacteroidetes *Flavobacteria bacterium* BBFL7
Fba-2-1 to Fba-2-8	Bacteria Bacteroidetes Flavobacteria Flavobacteriales *Flavobacteria bacterium* BAL38
Cat-1-1 to Cat-1-3	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Croceibacter atlanticus* HTCC2559
Ddo-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Dokdonia donghaensis* MED134
Fjo-1-1 to Fjo-1-35	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Flavobacterium johnsoniae* UW101
Fps-1-1 to Fps-1-96	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Flavobacterium psychrophilum* JIP02/86
Gfo-1-1 to Gfo-1-4	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Gramella forsetii* KT0803
Lbl-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Leeuwenhoekiella blandensis* MED217
Pir-1-1 to Pir-1-3	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Polaribacter irgensii* 23-P
Psp-1-1 to Psp-1-5	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Polaribacter sp.* MED152
Pto-1-1 to Pto-1-6	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Psychroflexus torquis* ATCC 700755
Chu-1-1	Bacteria Bacteroidetes Sphingobacteria Sphingobacteriales Flexibacteraceae *Cytophaga hutchinsonii* ATCC 33406
Tde-1-1	Bacteria Spirochaetes Spirochaetales Spirochaetaceae *Treponema denticola* ATCC 35405
env-1 to env-70	environmental samples

32.2 Gene contexts

Each Flavo-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Flavo-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 32.3. The accession number of the sequence containing each Flavo-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 32.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	JCVI_SCAF_1101668255295	-	932	865	**RNA**→ Ile_Leu_Val_MetRS_core (cd00668)? (GOS2161)→
Ddo-1-1	NZ_AAMZ01000012.1	+	2669	2763	hypo→ **RNA**→ ←hypo
Pto-1-1	NZ_AAPR01000002.1	-	378782	378671	←hypo **RNA**→ ←MM_CoA_mut_B12_BD (cd02071)ArgK (COG1703)MM_CoA_mutase (pfam01642)

Cat-1-1	NZ_AAMP01000001.1	+	1197511	1197619	←adh_short (pfam00106) **RNA**→ ←thiolase (cd00751)
Fjo-1-1	NC_009441.1	+	5741049	5741154	COG4339 (COG4339)→ **RNA**→ ←hypo
Pto-1-2	NZ_AAPR01000018.1	+	8274	8366	NapH (COG0348)→ **RNA**→ hypo→
Fps-1-1	NC_009613.1	-	2312473	2312389	←Rpe (COG0036) **RNA**→ ←RpoD (COG0568)
Fjo-1-2	NC_009441.1	+	2016697	2016781	←PurT (COG0027) **RNA**→ hypo→
Fjo-1-3	NC_009441.1	-	6000023	5999939	Sialate O-acetylesterase→ **RNA**→ Glyco_hydro_3 (pfam00933)Glyco_hydro_3_C (pfam01915)→
Fps-1-2	NC_009613.1	+	87466	87549	←hypo **RNA**→ ←ProC (COG0345)
Fps-1-3	NC_009613.1	+	727266	727349	Prc (COG0793)→ **RNA**→ Tgt (COG0343)→
Fps-1-4	NC_009613.1	+	774964	775047	GyrA (COG0188)→ **RNA**→ **RNA**→ ←Peptidase_M28 (pfam04389)
Fps-1-17	NC_009613.1	+	774374	774469	(shown above)
Bfr-1-1	NC_006873.1	-	806	733	←COG4227 (COG4227) **RNA**→ ←Resolvase (pfam00239)
Gfo-1-1	NC_008571.1	-	442238	442157	Pnp (COG1185)→ **RNA**→ RpoD (COG0568)→
Bth-1-1	NC_004703.1	-	18733	18657	←hypo **RNA**→ ←hypo
Cat-1-2	NZ_AAMP01000001.1	+	75896	75953	←AlaDh_PNT_C (pfam01262)AlaDh_PNT_N (pfam05222) **RNA**→ ←hypo
Fjo-1-4	NC_009441.1	+	6019588	6019666	←FtsY (COG0552) **RNA**→ ←hypo
Fps-1-5	NC_009613.1	-	1534455	1534380	Rhomboid (pfam01694)→ **RNA**→ ←SpoU (COG0566)
Gfo-1-2	NC_008571.1	+	3524314	3524387	←COG2908 (COG2908) **RNA**→ ←COG0720 (COG0720)
Psp-1-1	NZ_AANA01000001.1	-	1303039	1302957	PseudoU_synth_EcTruA (cd02570)→ **RNA**→ ←SCAD/SBCAD (cd01158)
Fba-2-1	NZ_AAXX01000001.1	-	700803	700709	←hypo **RNA**→ ←RHOD (cd00158)GloB (COG0491)PspE (COG0607)
Pto-1-3	NZ_AAPR01000003.1	+	277159	277307	URO-D (cd00717)→ **RNA**→ ←ArgE (COG0624)
Fba-2-2	NZ_AAXX01000001.1	+	681187	681276	hypo→ **RNA**→ ←hypo
Fjo-1-5	NC_009441.1	-	2219540	2219456	←hypo **RNA**→ LysU (COG1190)→
Tde-1-1	NC_002967.9	+	1611636	1611767	RhsA (COG3209)→ **RNA**→ RhsA (COG3209)→
Fjo-1-6	NC_009441.1	-	5269445	5269368	hypo→ **RNA**→ ←hypo
Pto-1-4	NZ_AAPR01000007.1	+	184845	184929	←AlaDh_PNT_C (pfam01262)AlaDh_PNT_N (pfam05222) **RNA**→ ribonuclease activity regulator protein RraA→
env-2	JCVI_SCAF_1101667376369	-	529	462	**RNA**→ ←ABC_MJ0796_Lo1CDE_FtsE (cd03255)? (GOS2568)
Fps-1-6	NC_009613.1	+	1377002	1377143	←Ribosomal_S21 (pfam01165) **RNA**→ ←LCAD (cd01160)
env-3	JCVI_SCAF_1096627455085	-	253	175	**RNA**→
env-4	JCVI_SCAF_1096627685283	+	1235	1335	←AtoD (COG1788)? (GOS7754) **RNA**→ ←? (GOS377)
Chu-1-1	NC_008255.1	+	1783674	1783753	←rve (pfam00665) **RNA**→ hypo→
env-5	sludgeOz_Contig8743	-	743	661	GpmI (COG0696)→ **RNA**→ ←AtoA (COG2057)
Cat-1-3	NZ_AAMP01000003.1	+	116229	116299	NUC1 (COG1864)→ **RNA**→ hypo→
Fps-1-7	NC_009613.1	+	1248668	1248760	←adh_short (pfam00106) **RNA**→ ←PhoH (pfam02562)
Gfo-1-3	NC_008571.1	+	386266	386363	Endonuclease_DUF559 (cd01038)MetH (COG1410)→ **RNA**→ MTHFR (cd00537)→
env-6	JCVI_SCAF_1101668142625	-	673	598	COG3943 (COG3943)? (GOS6297)→ **RNA**→ ←? (GOS241)
Fps-1-8	NC_009613.1	-	2036769	2036665	hypo→ **RNA**→ RpsA (COG0539)→
env-7	JCVI_SCAF_1096627503475	-	754	679	Transposase_20 (pfam02371)? (GOS4503)→ **RNA**→ COG0820 (COG0820)? (GOS4046)→
Fps-1-9	NC_009613.1	-	1170893	1170815	←PhrB (COG0415) **RNA**→ ←adh_short (pfam00106)
Fps-1-10	NC_009613.1	-	2053796	2053715	AcrA (COG0845)→ **RNA**→ AcrB (COG0841)→
Fps-1-11	NC_009613.1	+	1308986	1309129	Glycos_transf_2 (pfam00535)→ **RNA**→ ManB (COG1109)→
Fba-1-1	NZ_AAPD01000003.1	+	306499	306573	←TlpA_like_DipZ_like (cd03012) **RNA**→ ←ThiE (COG0352)
Fjo-1-7	NC_009441.1	+	4084547	4084621	←hypo **RNA**→ ←DUF808 (pfam05661)
env-8	JCVI_SCAF_1101668138832	+	1068	1161	**RNA**→
env-9	JCVI_SCAF_1101668152124	-	472	396	Cof (COG0561)? (GOS7545)→ **RNA**→ UPF0004 (pfam00919)? (GOS12090)→
env-10	JCVI_SCAF_1101668140290	-	874	798	Cof (COG0561)? (GOS7545)→ **RNA**→ MiaB (COG0621)? (GOS1409)→
env-11	soil_XZS18398_b2	-	606	535	←hypo **RNA**→ Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component, and related
env-12	JCVI_SCAF_1101668161155	+	1191	1266	**RNA**→
env-13	JCVI_SCAF_1096627458992	+	766	841	←LipB (COG0321)? (GOS6621) **RNA**→ tRNA_anti (pfam01336)? (GOS1394)→
Fba-2-3	NZ_AAXX01000001.1	+	83759	83835	←GrpE (cd00446) **RNA**→ ←COG1090 (COG1090)
Pir-1-1	NZ_AAOG01000002.1	-	549859	549739	ArgS (COG0018)→ **RNA**→ Lpd (COG1249)→
Fps-1-12	NC_009613.1	-	499344	499263	hypo→ **RNA**→ FabD (COG0331)→
Fjo-1-8	NC_009441.1	-	1737234	1737153	←MetH (COG0646) **RNA**→ ←TrxB (COG0492)

env-14	wf3_Contig2121	+	888	961	**RNA**→
Fba-2-4	NZ_AAXX01000001.1	+	707496	707593	hypo→ **RNA**→ ←LolE (COG4591)
Fba-2-5	NZ_AAXX01000003.1	-	127063	126966	←hypo **RNA**→ ThrS (COG0441)→
Fjo-1-9	NC_009441.1	-	3138939	3138862	Peptidase_S8 (pfam00082)→ **RNA**→ ←CSD (pfam00313)
Fps-1-13	NC_009613.1	-	2778208	2778065	NUC (smart00477)→ **RNA**→ ←FTSW_RODA_SPOVE (pfam01098)
env-15	sludgeOz_Contig3324	+	965	1040	NuoH (COG1005)→ **RNA**→ NuoI (COG1143)→
env-16	JCVI_SCAF_1096627458069	+	1298	1369	**RNA**→
Fps-1-14	NC_009613.1	+	310503	310574	←COG1721 (COG1721) **RNA**→ ←Thioredoxin (pfam00085)
Fps-1-15	NC_009613.1	+	2214869	2215012	←hypo **RNA**→ ←Urocanase (pfam01175)
Psp-1-2	NZ_AANA01000002.1	+	151471	151560	←Rpe (COG0036) **RNA**→ ←RpoD (COG0568)
Pto-1-5	NZ_AAPR01000031.1	-	20849	20774	←TypA (COG1217) **RNA**→ ←ThiC (pfam01964)
Pir-1-2	NZ_AAOG01000003.1	+	81116	81191	←ThiE (COG0352) **RNA**→ ←ThiC (pfam01964)
Pto-1-6	NZ_AAPR01000031.1	-	29522	29447	hypo→ **RNA**→ TPR (cd00189)→
Fps-1-16	NC_009613.1	-	633751	633658	hypo→ **RNA**→ ligand_gated_channel (cd01347)→
env-17	wf3_AHAI827_g1	-	728	659	**RNA**→ ←Pyruvate_Kinase (cd00288)
Fba-2-6	NZ_AAXX01000002.1	-	108484	108388	COG4123 (COG4123)→ **RNA**→ GCD (cd01151)→
Fjo-1-10	NC_009441.1	-	5524598	5524517	adh_short (pfam00106)→ **RNA**→ Transaldolase_FSA (cd00956)→
Fba-2-7	NZ_AAXX01000001.1	-	665852	665786	hypo→ **RNA**→ FolC (COG0285)→
env-18	JCVI_SCAF_1096626882301	-	1519	1445	**RNA**→ ←? (GOS12690)
Fjo-1-11	NC_009441.1	+	2274636	2274717	←hypo **RNA**→ PotE (COG0531)→
env-19	JCVI_SCAF_1096627456999	-	1317	1229	←Maf_Ham1 (cd00985)? (GOS75) **RNA**→ ? (GOS869)→
env-20	wf3_Contig7634	-	2493	2411	←YqgFc (smart00732) **RNA**→ DapD (COG2171)→
Psp-1-3	NZ_AANA01000001.1	-	2098710	2098633	←TrxB (COG0492) **RNA**→ ←NifU (pfam01106)
Fps-1-18	NC_009613.1	-	1895558	1895473	←TrxB (COG0492) **RNA**→ ←NifU (pfam01106)
env-21	sludgeOz_Contig4957	-	674	596	←Methionine synthase I (cobalamin-dependent), meth yltransferase domain(,,,) **RNA**→ ←(,,,)
Fps-1-19	NC_009613.1	+	1509239	1509334	←CarA (COG0505) **RNA**→ ←RplQ (COG0203)
env-22	wf3_Contig1248	-	354	285	**RNA**→
env-23	JCVI_SCAF_1101668153802	+	1301	1383	**RNA**→
env-24	JCVI_SCAF_1101668158720	-	1312	1229	**RNA**→ ←? (GOS1344)
env-25	wf3_AHAI6559_x1	-	243	146	**RNA**→
Fps-1-20	NC_009613.1	+	1651557	1651638	←DBI_PRT (pfam02277)BluB (cd02145) **RNA**→ ←CobU (cd00544)
env-26	wf3_Contig6579	+	145	224	**RNA**→ Glyco_tran_28_C (pfam04101)→
Fps-1-21	NC_009613.1	+	1035239	1035316	←PqqL (COG0612) **RNA**→ ←Ribosomal_L20 (pfam00453)
ueu-1-1	NZ_ABCO01000011.1	-	43084	43011	←16S ribosomal RNA **RNA**→ ←hypo
env-27	JCVI_SCAF_1101668153165	+	199	289	**RNA**→
env-28	wf3_Contig4932	-	593	515	←COG2908 (COG2908) **RNA**→ ←COG0720 (COG0720)
env-29	wf1_Contig1516	+	616	690	**RNA**→
env-30	wf1_Contig56	+	328	402	**RNA**→ hypo→
env-31	wf3_AHAI17743_g2	+	467	545	hypo→ **RNA**→ Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component, and related
env-32	JCVI_SCAF_1101667193122	+	549	625	**RNA**→
env-33	JCVI_SCAF_1096626901114	-	1414	1334	**RNA**→ ←ArgE (COG0624)? (GOS351)
env-34	sludgeOz_Contig3301	-	517	438	←TrxB (COG0492) **RNA**→ ←NifU (pfam01106)
Fps-1-22	NC_009613.1	+	1105089	1105174	←HisA (COG0106) **RNA**→ ←HisH (COG0118)
env-35	wf3_Contig4156	-	1207	1132	**RNA**→ ←GpsA (COG0240)
Fps-1-23	NC_009613.1	+	1068341	1068426	←LolE (COG4591) **RNA**→ ←AcrA (COG0845)
Fps-1-24	NC_009613.1	+	1132910	1132995	←KdsB (COG1212) **RNA**→ ←RecD (COG0507)
env-36	JCVI_SCAF_1096628075201	+	1093	1177	**RNA**→
env-37	wf3_Contig3210	-	404	333	hypo→ **RNA**→ hypo→
Fps-1-25	NC_009613.1	+	1446748	1446827	←hypo **RNA**→ ←hypo
Fps-1-26	NC_009613.1	+	1526942	1527028	←Ribosomal_S12 (cd03368) **RNA**→ Bac_surface_Ag (pfam01103)→
Fps-1-27	NC_009613.1	+	1244341	1244422	←16S ribosomal RNA **RNA**→ ←hypo
Fps-1-28	NC_009613.1	+	1590815	1590896	←16S ribosomal RNA **RNA**→ ←RpsT (COG0268)

Fps-1-29	NC_009613.1	+	2504159	2504240	←16S ribosomal RNA **RNA**→ ←LolE (COG4591)
Fps-1-30	NC_009613.1	-	1039046	1038965	←16S ribosomal RNA **RNA**→ ThrS (COG0441)TGS (pfam02824)→
Fps-1-31	NC_009613.1	+	1726586	1726667	←16S ribosomal RNA **RNA**→ OmpR (COG0745)→
Fps-1-32	NC_009613.1	-	570785	570704	←16S ribosomal RNA **RNA**→ ←InfB (COG0532)IF2_N (pfam04760)
env-38	wf3_Contig3028	-	464	392	hypo→ **RNA**→ ←Predicted periplasmic solute-binding protein
Fps-1-33	NC_009613.1	+	1383707	1383784	←CysK (COG0031)COG3620 (COG3620) **RNA**→ ←hypo
Fps-1-34	NC_009613.1	+	501025	501102	←PurT (COG0027) **RNA**→ ←COG0436 (COG0436)
Fjo-1-12	NC_009441.1	+	2021125	2021206	←hypo **RNA**→ ←COG0436 (COG0436)
Fps-1-35	NC_009613.1	+	615290	615371	←TrpA (COG0159) **RNA**→ ←TrpB (COG0133)
env-39	JCVI_SCAF_1096628183349	+	1118	1197	? (GOS12106160)→ **RNA**→ Seryl_tRNA_N (pfam02403)? (GOS1816)→
env-40	wf3_Contig6516	-	2138	2052	←hypo **RNA**→ ←TIM (pfam00121)
env-41	wf3_Contig947	-	503	423	←FabG (COG1028) **RNA**→ ←Predicted pyridoxal phosphate-dependent enzyme apparently involved in regulation of cell wall bi
env-42	wf3_Contig6825	-	266	187	**RNA**→
env-43	wf3_Contig1282	-	858	779	←HdrC (COG1150) **RNA**→ ←HdrC (COG1150)
env-44	wf2_AHAA11796_x1	-	323	253	**RNA**→
env-45	wf3_Contig3627	+	198	269	**RNA**→ SerS (COG0172)→
env-46	wf3_Contig5507	+	579	651	←AccD (COG0777) **RNA**→ ←Fructose/tagatose bisphosphate aldolase(EC:4.1.2.13)
Psp-1-4	NZ_AANA01000001.1	+	1083753	1083830	←ribosomal protein S21 **RNA**→ ←LCAD (cd01160)
Gfo-1-4	NC_008571.1	+	2681842	2681951	←AlaDh_PNT_C (pfam01262)AlaDh_PNT_N (pfam05222) **RNA**→ ←hypo
Fba-2-8	NZ_AAXX01000002.1	-	6337	6256	←PutA (COG1012) **RNA**→ EGL-9 (COG3751)→
env-47	JCVI_SCAF_1101668544195	-	55	1	**RNA**→
Fps-1-36	NC_009613.1	+	2110213	2110300	←SpoT (COG0317) **RNA**→ ←SpoU (COG0566)
env-48	wf3_AHAI3907_b1	+	789	858	**RNA**→
Fps-1-37	NC_009613.1	-	1213637	1213560	adh_short (pfam00106)→ **RNA**→ Transaldolase_FSA (cd00956)→
env-49	wf2_Contig4188	+	537	647	hypo→ **RNA**→ Cation/multidrug efflux pump→
Fjo-1-13	NC_009441.1	-	4794620	4794545	←hypo **RNA**→ ←REC (cd00156)
Fjo-1-14	NC_009441.1	+	3128289	3128364	←NhaP (COG0025) **RNA**→ HSP70 (pfam00012)→
Fjo-1-15	NC_009441.1	+	2573472	2573547	←hypo **RNA**→ ←MS_channel (pfam00924)
Fjo-1-16	NC_009441.1	-	5783002	5782927	←FumC (COG0114) **RNA**→ ←ligand_gated_channel (cd01347)
Fjo-1-17	NC_009441.1	+	5193557	5193632	←ManC (COG0662) **RNA**→ AcrR (COG1309)→
Fjo-1-18	NC_009441.1	-	5605119	5605044	←hypo **RNA**→ ←CysI (COG0155)Rubredoxin (pfam00301)
env-50	sludgePhrap_Contig7145	+	120	199	**RNA**→ ←GARS_C (pfam02843)
Fps-1-38	NC_009613.1	-	861460	861379	hypo→ **RNA**→ hypo→
Fjo-1-19	NC_009441.1	-	6017100	6017017	←AmpC (COG1680) **RNA**→ hypo→
env-51	wf2_AHAA10979_x1	-	659	542	**RNA**→ hypo→
Lbl-1-1	NZ_AANC01000009.1	-	127964	127876	←COG3185 (COG3185) **RNA**→ ←HgmA (pfam04209)
Fjo-1-20	NC_009441.1	-	2303534	2303455	←16S ribosomal RNA **RNA**→ ←COG2141 (COG2141)
Fjo-1-21	NC_009441.1	+	30089	30168	←16S ribosomal RNA **RNA**→ ThrS (COG0441)TGS (pfam02824)→
Fjo-1-22	NC_009441.1	+	506429	506508	←16S ribosomal RNA **RNA**→ ←LolE (COG4591)
Fjo-1-23	NC_009441.1	-	3263136	3263057	←16S ribosomal RNA **RNA**→ ←hypo
env-52	JCVI_SCAF_1096628078931	+	1013	1088	**RNA**→ MhpD (COG0179)? (GOS5149)→
Fps-1-39	NC_009613.1	-	1910757	1910674	KdpC (COG2156)→ **RNA**→ hypo→
env-53	JCVI_SCAF_1101668150988	+	919	987	Ribonuc_red_sm (pfam00268)? (GOS1236)→ **RNA**→ ATP-cone (pfam03477)? (GOS2385)→
Fjo-1-24	NC_009441.1	+	2164515	2164593	Era (COG1159)→ **RNA**→ COG1160 (COG1160)→
env-54	sludgeOz_Contig5510	+	235	313	(,,,)→ **RNA**→ (,,,)→
env-55	JCVI_SCAF_1096627682972	-	861	755	←CarA (COG0505)? (GOS4387) **RNA**→ ←RplQ (COG0203)? (GOS1143)
Fps-1-40	NC_009613.1	+	1939445	1939549	←Cmk (COG0283) **RNA**→ ←hypo
env-56	soil_XZS22982_b2	-	448	371	←COG0720 (COG0720) **RNA**→ ←hypo
Fjo-1-25	NC_009441.1	-	2008746	2008669	hypo→ **RNA**→ COG1741 (COG1741)→
Fjo-1-26	NC_009441.1	+	4304111	4304188	hypo→ **RNA**→ hypo→
Fjo-1-27	NC_009441.1	+	143366	143445	hypo→ **RNA**→ hypo→
Fps-1-41	NC_009613.1	-	746192	746081	←COG0637 (COG0637) **RNA**→ RpoE (COG1595)→

Fps-1-42	NC_009613.1	+	2287165	2287240	←HisS (COG0124) **RNA**→ SalY (COG0577)→
env-57	JCVI_SCAF_1096627031671	-	1398	1306	FolE (COG0302)? (GOS8129)→ **RNA**→ CysS (COG0215)? (GOS3464)→
Fps-1-43	NC_009613.1	+	715409	715482	FolE (COG0302)→ **RNA**→ CysS (COG0215)→
env-58	JCVI_SCAF_1101668151107	+	768	839	Toluene_X (pfam03349)? (GOS1685)→ **RNA**→ UPF0004 (pfam00919)? (GOS1409)→
Fps-1-44	NC_009613.1	+	1628589	1628684	←FtsY (COG0552) **RNA**→ ←hypo
Fps-1-45	NC_009613.1	+	253670	253741	COG1215 (COG1215)→ **RNA**→ ←hypo
Fps-1-46	NC_009613.1	-	496580	496509	←GalE (COG1087) **RNA**→ ←WecE (COG0399)
Fps-1-47	NC_009613.1	-	1414124	1414055	←S4 (cd00165)PseudoU_synth_RluF (cd02554) **RNA**→ ←MhpC (COG0596)
Fps-1-48	NC_009613.1	-	2705148	2705079	←LeuS (COG0495) **RNA**→ ←PAP2_like_5 (cd03394)
Fps-1-49	NC_009613.1	-	1920625	1920557	←ValRS_core (cd00817) **RNA**→ hypo→
Fps-1-50	NC_009613.1	+	1406269	1406338	←DapF (COG0253) **RNA**→ DegQ (COG0265)→
Fps-1-51	NC_009613.1	-	2364097	2364027	PaaY (COG0663)→ **RNA**→ ←MurI (COG0796)
Fps-1-52	NC_009613.1	+	790061	790132	←AtoC (COG2204) **RNA**→ ←NatB (COG1668)
Fps-1-53	NC_009613.1	+	2032864	2032933	←ManA (COG1482) **RNA**→ ←Bcp (COG1225)
Fps-1-54	NC_009613.1	-	704714	704643	←COG1216 (COG1216) **RNA**→ Uup (COG0488)→
Fps-1-55	NC_009613.1	-	2367514	2367443	←COG1331 (COG1331) **RNA**→ ←NifU (pfam01106)
Fps-1-56	NC_009613.1	+	285771	285842	←Glyoxalase (pfam00903) **RNA**→ RbfA (COG0858)→
Fps-1-57	NC_009613.1	+	554571	554650	←Ribosomal_L13 (pfam00572) **RNA**→ ←Ribonuc_red_lgN (pfam00317)Ribonuc_red_lgC (pfam02867)ATP-cone (pfam03477)
Fjo-1-28	NC_009441.1	+	4289731	4289814	hypo→ **RNA**→ hypo→
Fps-1-58	NC_009613.1	+	2008367	2008445	←TPR (cd00189) **RNA**→ ←ArgD (COG4992)
Fps-1-59	NC_009613.1	+	609953	610031	←AraJ (COG2814) **RNA**→ ←EmrA (COG1566)
Fjo-1-29	NC_009441.1	-	2288233	2288156	KdpC (COG2156)→ **RNA**→ hypo→
Fjo-1-30	NC_009441.1	+	4306293	4306370	hypo→ **RNA**→ hypo→
Fps-1-60	NC_009613.1	-	77564	77485	←hypo **RNA**→ MetG (COG0143)tRNA_bind_EcMetRS_like (cd02800)→
Fps-1-61	NC_009613.1	-	1728075	1727998	←COG3513 (COG3513) **RNA**→ ←OmpR (COG0745)
Fps-1-62	NC_009613.1	-	1187860	1187781	←MhpD (COG0179) **RNA**→ ←GltS_FMN (cd02808)
Fps-1-63	NC_009613.1	+	304427	304506	hypo→ **RNA**→ ←ThrA (COG0460)
Fps-1-64	NC_009613.1	+	2078980	2079061	←FN3 (cd00063) **RNA**→ ←UshA (COG0737)
env-59	JCVI_SCAF_1096628065663	-	1910	1834	**RNA**→
Fjo-1-31	NC_009441.1	-	3168916	3168833	←hypo **RNA**→ hypo→
Fps-1-65	NC_009613.1	-	1935873	1935792	←DYS1 (COG1899) **RNA**→ ←LysA (COG0019)
Fps-1-66	NC_009613.1	+	1008767	1008846	←PrfB (COG1186) **RNA**→ GLY1 (COG2008)→
Fps-1-67	NC_009613.1	+	563089	563169	←deoxycytidylate_deaminase (cd01286) **RNA**→ ←hypo
Fps-1-68	NC_009613.1	+	2808043	2808124	←dUTPase (pfam00692) **RNA**→ ←RfbX (COG2244)
Fps-1-69	NC_009613.1	-	265405	265326	←RssA (COG1752)Bac_surface_Ag (pfam01103) **RNA**→ ←RssA (COG1752)Bac_surface_Ag (pfam01103)
Fps-1-70	NC_009613.1	+	2333989	2334068	←Gliding motility protein GldB **RNA**→ NAD_synthase (cd00553)→
Fjo-1-32	NC_009441.1	+	1625353	1625432	←ERG12 (COG1577) **RNA**→ ←HMG-CoA_reductase_classII (cd00644)
Fps-1-71	NC_009613.1	-	1550751	1550672	←Dcp (COG0339) **RNA**→ ←PurE (COG0041)
Fps-1-72	NC_009613.1	-	380759	380680	←RsuA (COG1187) **RNA**→ ←UbiA (pfam01040)
Fps-1-73	NC_009613.1	+	2670156	2670237	←KAS_I_II (cd00834) **RNA**→ ←BioA (COG0161)
Fps-1-74	NC_009613.1	-	1246430	1246349	SAICAR_synt_Sc (cd01414)→ **RNA**→ ←FrvX (COG1363)
Fps-1-75	NC_009613.1	+	197892	197971	←RecB (COG1074) **RNA**→ SodA (COG0605)→
Fps-1-76	NC_009613.1	-	689704	689625	←GlcD (COG0277) **RNA**→ ←NarK (COG2223)
env-60	sludgeOz_Contig5854	+	323	402	←CPSase_L_chain (pfam00289) **RNA**→ (,,,)→
Fjo-1-33	NC_009441.1	-	2946722	2946641	SerC (COG1932)→ **RNA**→ SerA (COG0111)→
Fps-1-77	NC_009613.1	-	909081	909000	thiolase (cd00751)→ **RNA**→ NLPC_P60 (pfam00877)→
Fps-1-78	NC_009613.1	+	1193056	1193135	←AcoB (COG0022)TPP_E1_PDC_ADC_BCADC (cd02000) **RNA**→ PcbC (COG3491)→
Fps-1-79	NC_009613.1	-	730679	730599	←AccA (COG0825) **RNA**→ ←hypo
Fps-1-80	NC_009613.1	+	796790	796871	COG3296 (COG3296)→ **RNA**→ PadR (pfam03551)→
Fps-1-81	NC_009613.1	+	679787	679868	←hypo **RNA**→ DUF344 (pfam03976)→
Fps-1-82	NC_009613.1	-	47864	47783	←AtoD (COG1788) **RNA**→ ←MrcA (COG5009)
Fps-1-83	NC_009613.1	-	2118876	2118797	hypo→ **RNA**→ tRNA_m1G_MT (pfam01746)→

Fps-1-84	NC_009613.1	+	1392373	1392452	←LipB (COG0321) **RNA**→ LysU (COG1190)→
Fps-1-85	NC_009613.1	-	2004068	2003989	←THI80 (COG1564) **RNA**→ DUF349 (pfam03993)→
Fps-1-86	NC_009613.1	-	557267	557188	Ribonuc_red_sm (pfam00268)→ **RNA**→ Ribonuc_red_lgN (pfam00317)Ribonuc_red_lgC (pfam02867)ATP-cone (pfam03477)→
Fps-1-87	NC_009613.1	+	229911	229990	←Sfp (COG2091) **RNA**→ AdoHcyase (pfam05221)→
Fps-1-88	NC_009613.1	+	1545638	1545717	←Obg (COG0536) **RNA**→ ←TlyC (COG1253)
Fps-1-89	NC_009613.1	+	102392	102471	←hypo **RNA**→ EGL-9 (COG3751)→
env-61	JCVI_SCAF_1096627697293	+	829	932	? (GOS7413)→ **RNA**→ ←? (GOS16)
Fps-1-90	NC_009613.1	-	1200485	1200417	Ung (COG0692)→ **RNA**→ ←UPF0121 (pfam03661)
env-62	soil_XZS115611_b1	+	675	758	←hypo **RNA**→ (EC:2.1.2.11)→
env-63	wf3_AHAI4584_g1	+	545	610	**RNA**→ hypo→
env-64	JCVI_SCAF_1096628075775	+	208	285	? (GOS8207)→ **RNA**→ Acyl-CoA_dh_N (pfam02771)? (GOS26)→
env-65	JCVI_SCAF_1101668262840	-	874	765	←? (GOS530) **RNA**→ ←3-HAO (pfam06052)? (GOS9899)
Fps-1-91	NC_009613.1	-	2223908	2223831	Era (COG1159)→ **RNA**→ COG1160 (COG1160)→
Fps-1-92	NC_009613.1	-	2174176	2174099	←ligand_gated_channel (cd01347) **RNA**→ PrfA (COG0216)→
Fps-1-93	NC_009613.1	-	2408210	2408133	←GpmI (COG0696) **RNA**→ hypo→
Fjo-1-34	NC_009441.1	+	4125942	4126020	←hypo **RNA**→ ←OmpR (COG0745)
Fjo-1-35	NC_009441.1	-	3049569	3049492	←YXKO-related (cd01171) **RNA**→ GcvT (COG0404)→
env-66	wf3_Contig6008	+	166	237	**RNA**→ ←ABC_UvrA_II (cd03271)
env-67	JCVI_SCAF_1101668256483	+	1228	1320	**RNA**→
env-68	wf3_Contig1018	+	373	444	**RNA**→ ←HATPase_c (smart00387)
env-69	wf3_Contig548	-	205	134	**RNA**→
Psp-1-5	NZ_AANA01000001.1	+	7555	7625	Cmk (COG0283)→ **RNA**→ ←NrdG (COG0602)
Pir-1-3	NZ_AAOG01000001.1	-	717586	717518	←COG3185 (COG3185) **RNA**→ ←HgmA (pfam04209)
Fps-1-94	NC_009613.1	-	970882	970805	hypo→ **RNA**→ COG1611 (COG1611)→
Fps-1-95	NC_009613.1	+	1832426	1832503	←COG2971 (COG2971) **RNA**→ ←GapA (COG0057)
Fps-1-96	NC_009613.1	-	2047222	2047145	PurB (cd01598)→ **RNA**→ BioF (COG0156)→
env-70	wf3_AHAI1047_b1	+	641	722	AcrA (COG0845)→ **RNA**→ ←hypo
Fba-1-2	NZ_AAPD01000009.1	+	100597	100698	AcrA (COG0845)→ **RNA**→ AcrB (COG0841)→

32.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 32.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Flavo-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

cd00063 (1) Fibronectin type 3 domain; One of three types of internal repeats found in the plasma protein fibronectin.
cd00156 (1) Signal receiver domain; originally thought to be unique to bacteria (CheY, OmpR, NtrC, and PhoB), now recently identified in eukaryotes ETR1 Arabidopsis thaliana; this domain receives the signal from the sensor partner in a two-component systems; contains a phosphoacceptor site that is phosphorylated by histidine kinase homologs; usually found N-terminal to a DNA binding effector domain; forms homodimers
cd00158 (1) Rhodanese Homology Domain (RHOD); an alpha beta fold domain found duplicated in the rhodanese protein.
cd00165 (1) S4/Hsp/ tRNA synthetase RNA-binding domain; The domain surface is populated by conserved, charged residues that define a likely RNA-binding site; Found in stress proteins, ribosomal proteins and tRNA synthetases; This may imply a hitherto unrecognized functional similarity between these three protein classes.
cd00189 (2) Tetratricopeptide repeat domain; typically contains 34 amino acids [WLF]-X(2)-[LIM]-[GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-X(2)-[ASL]-X(4)-[PKE]
cd00288 (1) Pyruvate kinase (PK): Large allosteric enzyme that regulates glycolysis through binding of the substrate, phosphoenolpyruvate, and one or more allosteric effectors.
cd00446 (1) GrpE is the adenine nucleotide exchange factor of DnaK (Hsp70)-type ATPases.
cd00537 (1) Methylenetetrahydrofolate reductase (MTHFR).
cd00544 (1) Adenosylcobinamide kinase / adenosylcobinamide phosphate guanyltransferase (CobU).
cd00553 (1) NAD+ synthase is a homodimer, which catalyzes the final step in de novo nicotinamide adenine dinucleotide (NAD+) biosynthesis, an amide transfer from either ammonia or glutamine to nicotinic acid adenine dinucleotide (NaAD).
cd00644 (1) Hydroxymethylglutaryl-coenzyme A (HMG-CoA) reductase (HMGR), class II, prokaryotic enzyme is a homodimer.

cd00668 (1) This is the catalytic core domain of isoleucyl, leucyl, valyl and methioninyl tRNA synthetases.
cd00717 (1) Uroporphyrinogen decarboxylase (URO-D) is a dimeric cytosolic enzyme that decarboxylates the four acetate side chains of uroporphyrinogen III (uro-III) to create coproporphyrinogen III, without requiring any prosthetic groups or cofactors.
cd00751 (2) Thiolase are ubiquitous enzymes that catalyze the reversible thiolytic cleavage of 3-ketoacyl-CoA into acyl-CoA and acetyl-CoA, a 2-step reaction involving a covalent intermediate formed with a catalytic cysteine.
cd00817 (1) This is the catalytic core domain of valine amino-acyl tRNA synthetases (ValRS) .
cd00834 (1) Beta-ketoacyl-acyl carrier protein (ACP) synthase (KAS), type I and II.
cd00956 (2) Transaldolase-like fructose-6-phosphate aldolases (FSA) found in bacteria and archaea, which are member of the MipB/TalC subfamily of class I aldolases.
cd00985 (1) Maf_Ham1.
cd01038 (1) Domain of unknown function, appears to be related to a diverse group of endonucleases.
cd01151 (1) Glutaryl-CoA dehydrogenase (GCD).
cd01158 (1) Short chain acyl-CoA dehydrogenase (SCAD).
cd01160 (2) Long chain acyl-CoA dehydrogenase (LCAD) is acyl-CoA dehydrogenases (ACAD), which is found in the mitochondria of eukaryotes and in some prokaryotes.
cd01171 (1) B.
cd01286 (1) Deoxycytidylate deaminase domain.
cd01347 (3) TonB dependent/Ligand-Gated channels are created by a monomeric 22 strand (22,24) anti-parallel beta-barrel.
cd01414 (1) Eukaryotic, prokaryotic and archaeal group of SAICAR synthetases represented by the Saccharomyces cerevisiae (Sc) SAICAR synthetase.
cd01598 (1) PurB: PurB_like adenylosuccinases (adenylsuccinate lyase, ASL).
cd02000 (1) Thiamine pyrophosphate (TPP) family, E1 of PDC_ADC_BCADC subfamily, TPP-binding module; composed of proteins similar to the E1 components of the human pyruvate dehydrogenase complex (PDC), the acetoin dehydrogenase complex (ADC) and the branched chain alpha-keto acid dehydrogenase/2-oxoisovalerate dehydrogenase complex (BCADC).
cd02071 (1) methylmalonyl CoA mutase B12 binding domain.
cd02145 (1) Subfamily of the nitroreductase family that includes BluB protein in Rhodobacter capsulatus is involved in the conversion of cobinamide to cobalamin in Cobalamin (vitamin B12) biosynthesis.
cd02554 (1) PseudoU_synth_RluF_like: Pseudouridine synthase, Escherichia coli RluF like.
cd02570 (1) PseudoU_synth_EcTruA: Pseudouridine synthase, Escherichia coli TruA like.
cd02800 (1) tRNA-binding-domain-containing Escherichia coli methionyl-tRNA synthetase (EcMetRS)-like proteins.
cd02808 (1) Glutamate synthase (GltS) FMN-binding domain.
cd03012 (1) TlpA-like family, DipZ-like subfamily; composed uncharacterized proteins containing a TlpA-like TRX domain.
cd03255 (1) This family is comprised of MJ0796 ATP-binding cassette, macrolide-specific ABC-type efflux carrier (MacAB), and proteins involved in cell division (FtsE) and release of liporoteins from the cytoplasmic membrane (LolCDE).
cd03271 (1) The excision repair protein UvrA domain II; Nucleotide excision repair in eubacteria is a process that repairs DNA damages by the removal of a 12-13-mer oligonucleotide containing the lesion.
cd03368 (1) S12-like family, 30S ribosomal protein S12 subfamily; S12 is located at the interface of the large and small ribosomal subunits of prokaryotes, chloroplasts and mitochondria, where it plays an important role in both tRNA and ribosomal subunit interactions.
cd03394 (1) PAP2_like_5 proteins.
COG0018 (1) Arginyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
COG0019 (1) Diaminopimelate decarboxylase [Amino acid transport and metabolism]
COG0022 (1) Pyruvate/2-oxoglutarate dehydrogenase complex, dehydrogenase (E1) component, eukaryotic type, beta subunit [Energy production and conversion]
COG0025 (1) NhaP-type Na+/H+ and K+/H+ antiporters [Inorganic ion transport and metabolism]
COG0027 (2) Formate-dependent phosphoribosylglycinamide formyltransferase (GAR transformylase) [Nucleotide transport and metabolism]
COG0031 (1) Cysteine synthase [Amino acid transport and metabolism]
COG0036 (2) Pentose-5-phosphate-3-epimerase [Carbohydrate transport and metabolism]
COG0041 (1) Phosphoribosylcarboxyaminoimidazole (NCAIR) mutase [Nucleotide transport and metabolism]
COG0057 (1) Glyceraldehyde-3-phosphate dehydrogenase/erythrose-4-phosphate dehydrogenase [Carbohydrate transport and metabolism]
COG0106 (1) Phosphoribosylformimino-5-aminoimidazole carboxamide ribonucleotide (ProFAR) isomerase [Amino acid transport and metabolism]
COG0111 (1) Phosphoglycerate dehydrogenase and related dehydrogenases [Amino acid transport and metabolism]
COG0114 (1) Fumarase [Energy production and conversion]
COG0118 (1) Glutamine amidotransferase [Amino acid transport and metabolism]
COG0124 (1) Histidyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
COG0133 (1) Tryptophan synthase beta chain [Amino acid transport and metabolism]
COG0143 (1) Methionyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
COG0155 (1) Sulfite reductase, beta subunit (hemoprotein) [Inorganic ion transport and metabolism]
COG0156 (1) 7-keto-8-aminopelargonate synthetase and related enzymes [Coenzyme metabolism]
COG0159 (1) Tryptophan synthase alpha chain [Amino acid transport and metabolism]
COG0161 (1) Adenosylmethionine-8-amino-7-oxononanoate aminotransferase [Coenzyme metabolism]
COG0172 (1) Seryl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
COG0179 (2) 2-keto-4-pentenoate hydratase/2-oxohepta-3-ene-1,7-dioic acid hydratase (catechol pathway) [Secondary metabolites biosynthesis, transport, and catabolism]
COG0188 (1) Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), A subunit [DNA replication, recombination, and repair]
COG0203 (2) Ribosomal protein L17 [Translation, ribosomal structure and biogenesis]
COG0215 (2) Cysteinyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
COG0216 (1) Protein chain release factor A [Translation, ribosomal structure and biogenesis]
COG0240 (1) Glycerol-3-phosphate dehydrogenase [Energy production and conversion]
COG0253 (1) Diaminopimelate epimerase [Amino acid transport and metabolism]
COG0265 (1) Trypsin-like serine proteases, typically periplasmic, contain C-terminal PDZ domain [Posttranslational modification, protein turnover, chaperones]
COG0268 (1) Ribosomal protein S20 [Translation, ribosomal structure and biogenesis]
COG0277 (1) FAD/FMN-containing dehydrogenases [Energy production and conversion]
COG0283 (2) Cytidylate kinase [Nucleotide transport and metabolism]
COG0285 (1) Folylpolyglutamate synthase [Coenzyme metabolism]

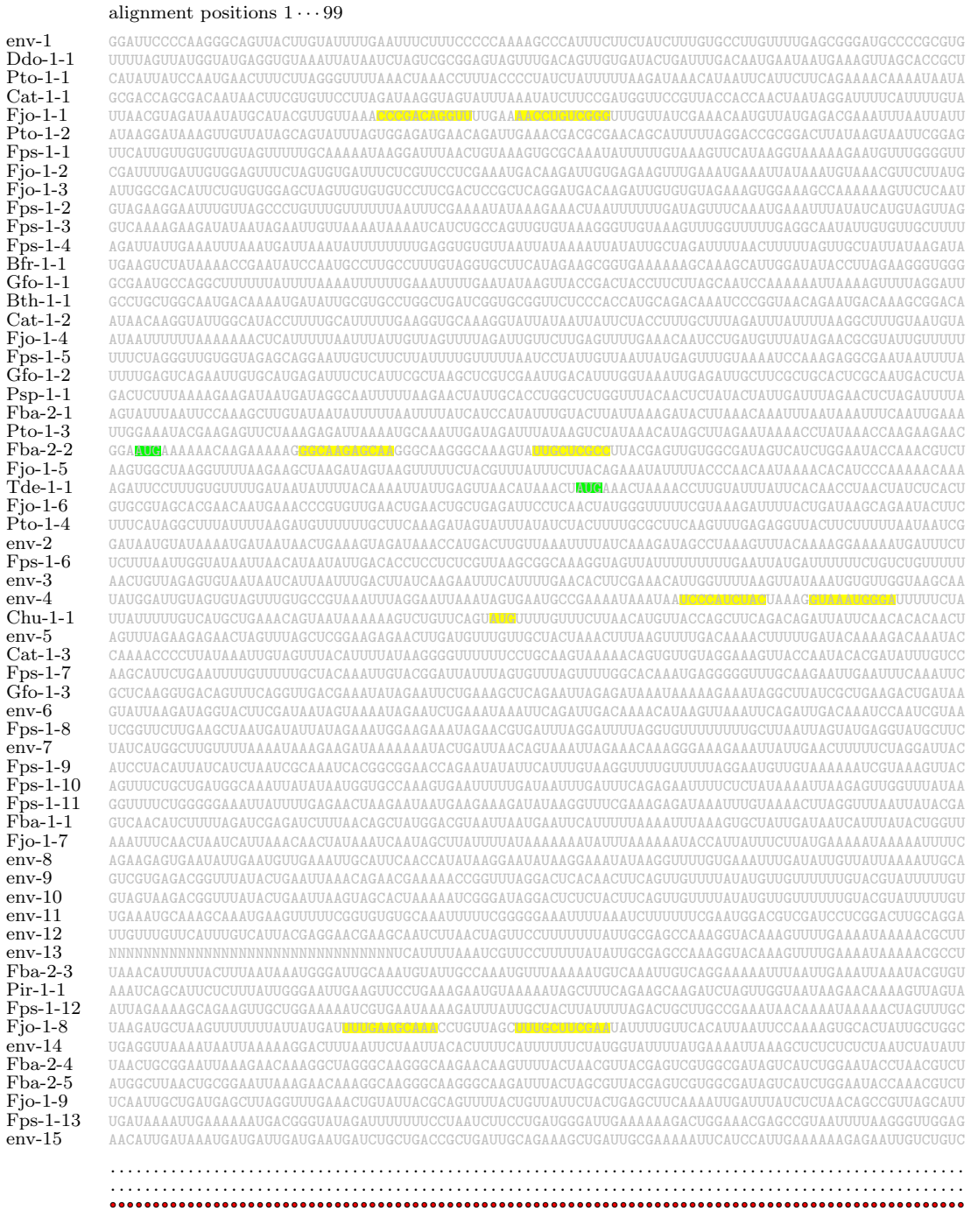
COG0302 (2) GTP cyclohydrolase I [Coenzyme metabolism]
COG0317 (1) Guanosine polyphosphate pyrophosphohydrolases/synthetases [Signal transduction mechanisms / Transcription]
COG0321 (2) Lipoate-protein ligase B [Coenzyme metabolism]
COG0331 (1) (acyl-carrier-protein) S-malonyltransferase [Lipid metabolism]
COG0339 (1) Zn-dependent oligopeptidases [Amino acid transport and metabolism]
COG0343 (1) Queuine/archaeosine tRNA-ribosyltransferase [Translation, ribosomal structure and biogenesis]
COG0345 (1) Pyrroline-5-carboxylate reductase [Amino acid transport and metabolism]
COG0348 (1) Polyferredoxin [Energy production and conversion]
COG0352 (2) Thiamine monophosphate synthase [Coenzyme metabolism]
COG0399 (1) Predicted pyridoxal phosphate-dependent enzyme apparently involved in regulation of cell wall biogenesis [Cell envelope biogenesis, outer membrane]
COG0404 (1) Glycine cleavage system T protein (aminomethyltransferase) [Amino acid transport and metabolism]
COG0415 (1) Deoxyribodipyrimidine photolyase [DNA replication, recombination, and repair]
COG0436 (2) Aspartate/tyrosine/aromatic aminotransferase [Amino acid transport and metabolism]
COG0441 (3) Threonyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
COG0460 (1) Homoserine dehydrogenase [Amino acid transport and metabolism]
COG0488 (1) ATPase components of ABC transporters with duplicated ATPase domains [General function prediction only]
COG0491 (1) Zn-dependent hydrolases, including glyoxylases [General function prediction only]
COG0492 (4) Thioredoxin reductase [Posttranslational modification, protein turnover, chaperones]
COG0495 (1) Leucyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
COG0505 (2) Carbamoylphosphate synthase small subunit [Amino acid transport and metabolism / Nucleotide transport and metabolism]
COG0507 (1) ATP-dependent exoDNAse (exonuclease V), alpha subunit - helicase superfamily I member [DNA replication, recombination, and repair]
COG0531 (1) Amino acid transporters [Amino acid transport and metabolism]
COG0532 (1) Translation initiation factor 2 (IF-2; GTPase) [Translation, ribosomal structure and biogenesis]
COG0536 (1) Predicted GTPase [General function prediction only]
COG0539 (1) Ribosomal protein S1 [Translation, ribosomal structure and biogenesis]
COG0552 (2) Signal recognition particle GTPase [Intracellular trafficking and secretion]
COG0561 (2) Predicted hydrolases of the HAD superfamily [General function prediction only]
COG0566 (2) rRNA methylases [Translation, ribosomal structure and biogenesis]
COG0568 (3) DNA-directed RNA polymerase, sigma subunit (sigma70/sigma32) [Transcription]
COG0577 (1) ABC-type antimicrobial peptide transport system, permease component [Defense mechanisms]
COG0596 (1) Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily) [General function prediction only]
COG0602 (1) Organic radical activating enzymes [Posttranslational modification, protein turnover, chaperones]
COG0605 (1) Superoxide dismutase [Inorganic ion transport and metabolism]
COG0607 (1) Rhodanese-related sulfurtransferase [Inorganic ion transport and metabolism]
COG0612 (1) Predicted Zn-dependent peptidases [General function prediction only]
COG0621 (1) 2-methylthioadenine synthetase [Translation, ribosomal structure and biogenesis]
COG0624 (2) Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase and related deacylases [Amino acid transport and metabolism]
COG0637 (1) Predicted phosphatase/phosphohexomutase [General function prediction only]
COG0646 (1) Methionine synthase I (cobalamin-dependent), methyltransferase domain [Amino acid transport and metabolism]
COG0662 (1) Mannose-6-phosphate isomerase [Carbohydrate transport and metabolism]
COG0663 (1) Carbonic anhydrases/acetyltransferases, isoleucine patch superfamily [General function prediction only]
COG0692 (1) Uracil DNA glycosylase [DNA replication, recombination, and repair]
COG0696 (2) Phosphoglyceromutase [Carbohydrate transport and metabolism]
COG0720 (3) 6-pyruvoyl-tetrahydropterin synthase [Coenzyme metabolism]
COG0737 (1) 5'-nucleotidase/2',3'-cyclic phosphodiesterase and related esterases [Nucleotide transport and metabolism]
COG0745 (3) Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain [Signal transduction mechanisms / Transcription]
COG0777 (1) Acetyl-CoA carboxylase beta subunit [Lipid metabolism]
COG0793 (1) Periplasmic protease [Cell envelope biogenesis, outer membrane]
COG0796 (1) Glutamate racemase [Cell envelope biogenesis, outer membrane]
COG0820 (1) Predicted Fe-S-cluster redox enzyme [General function prediction only]
COG0825 (1) Acetyl-CoA carboxylase alpha subunit [Lipid metabolism]
COG0841 (2) Cation/multidrug efflux pump [Defense mechanisms]
COG0845 (4) Membrane-fusion protein [Cell envelope biogenesis, outer membrane]
COG0858 (1) Ribosome-binding factor A [Translation, ribosomal structure and biogenesis]
COG1005 (1) NADH:ubiquinone oxidoreductase subunit 1 (chain H) [Energy production and conversion]
COG1012 (1) NAD-dependent aldehyde dehydrogenases [Energy production and conversion]
COG1028 (1) Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases) [Secondary metabolites biosynthesis, transport, and catabolism / General function prediction only]
COG1074 (1) ATP-dependent exoDNAse (exonuclease V) beta subunit (contains helicase and exonuclease domains) [DNA replication, recombination, and repair]
COG1087 (1) UDP-glucose 4-epimerase [Cell envelope biogenesis, outer membrane]
COG1090 (1) Predicted nucleoside-diphosphate sugar epimerase [General function prediction only]
COG1109 (1) Phosphomannomutase [Carbohydrate transport and metabolism]
COG1143 (1) Formate hydrogenlyase subunit 6/NADH:ubiquinone oxidoreductase 23 kD subunit (chain I) [Energy production and conversion]
COG1150 (2) Heterodisulfide reductase, subunit C [Energy production and conversion]
COG1159 (2) GTPase [General function prediction only]
COG1160 (2) Predicted GTPases [General function prediction only]
COG1185 (1) Polyribonucleotide nucleotidyltransferase (polynucleotide phosphorylase) [Translation, ribosomal structure and biogenesis]
COG1186 (1) Protein chain release factor B [Translation, ribosomal structure and biogenesis]
COG1187 (1) 16S rRNA uridine-516 pseudouridylate synthase and related pseudouridylate synthases [Translation, ribosomal structure and biogenesis]
COG1190 (2) Lysyl-tRNA synthetase (class II) [Translation, ribosomal structure and biogenesis]
COG1212 (1) CMP-2-keto-3-deoxyoctulosonic acid synthetase [Cell envelope biogenesis, outer membrane]
COG1215 (1) Glycosyltransferases, probably involved in cell wall biogenesis [Cell envelope biogenesis, outer membrane]

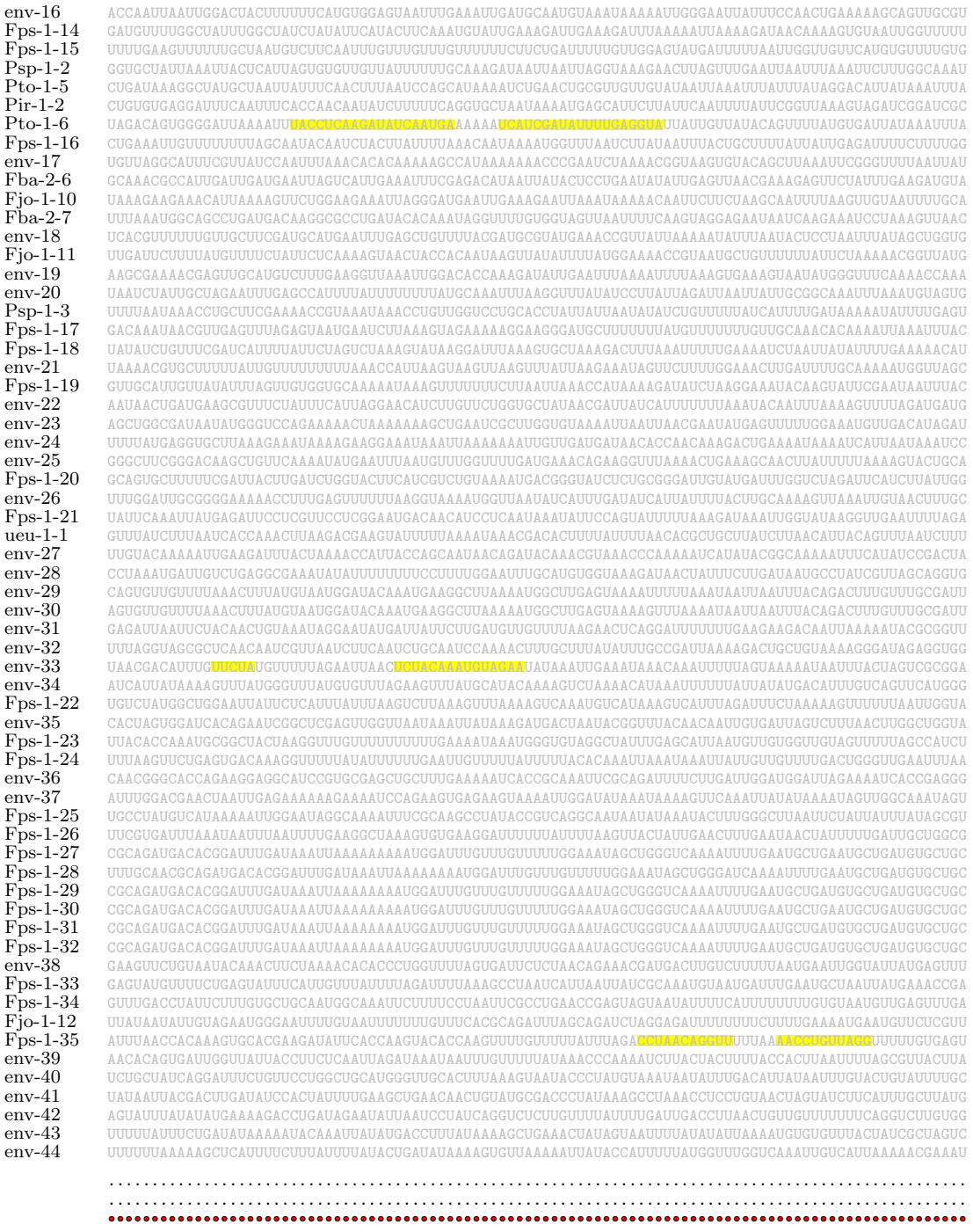
COG1216 (1) Predicted glycosyltransferases [General function prediction only]
COG1217 (1) Predicted membrane GTPase involved in stress response [Signal transduction mechanisms]
COG1225 (1) Peroxiredoxin [Posttranslational modification, protein turnover, chaperones]
COG1249 (1) Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component, and related enzymes [Energy production and conversion]
COG1253 (1) Hemolysins and related proteins containing CBS domains [General function prediction only]
COG1309 (1) Transcriptional regulator [Transcription]
COG1331 (1) Highly conserved protein containing a thioredoxin domain [Posttranslational modification, protein turnover, chaperones]
COG1363 (1) Cellulase M and related proteins [Carbohydrate transport and metabolism]
COG1410 (1) Methionine synthase I, cobalamin-binding domain [Amino acid transport and metabolism]
COG1482 (1) Phosphomannose isomerase [Carbohydrate transport and metabolism]
COG1564 (1) Thiamine pyrophosphokinase [Coenzyme metabolism]
COG1566 (1) Multidrug resistance efflux pump [Defense mechanisms]
COG1577 (1) Mevalonate kinase [Lipid metabolism]
COG1595 (1) DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog [Transcription]
COG1611 (1) Predicted Rossmann fold nucleotide-binding protein [General function prediction only]
COG1668 (1) ABC-type Na+ efflux pump, permease component [Energy production and conversion / Inorganic ion transport and metabolism]
COG1680 (1) Beta-lactamase class C and other penicillin binding proteins [Defense mechanisms]
COG1703 (1) Putative periplasmic protein kinase ArgK and related GTPases of G3E family [Amino acid transport and metabolism]
COG1721 (1) Uncharacterized conserved protein (some members contain a von Willebrand factor type A (vWA) domain) [General function prediction only]
COG1741 (1) Pirin-related protein [General function prediction only]
COG1752 (2) Predicted esterase of the alpha-beta hydrolase superfamily [General function prediction only]
COG1788 (2) Acyl CoA:acetate/3-ketoacid CoA transferase, alpha subunit [Lipid metabolism]
COG1864 (1) DNA/RNA endonuclease G, NUC1 [Nucleotide transport and metabolism]
COG1899 (1) Deoxyhypusine synthase [Posttranslational modification, protein turnover, chaperones]
COG1932 (1) Phosphoserine aminotransferase [Coenzyme metabolism / Amino acid transport and metabolism]
COG2008 (1) Threonine aldolase [Amino acid transport and metabolism]
COG2057 (1) Acyl CoA:acetate/3-ketoacid CoA transferase, beta subunit [Lipid metabolism]
COG2091 (1) Phosphopantetheinyl transferase [Coenzyme metabolism]
COG2141 (1) Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases [Energy production and conversion]
COG2156 (2) K+-transporting ATPase, c chain [Inorganic ion transport and metabolism]
COG2171 (1) Tetrahydrodipicolinate N-succinyltransferase [Amino acid transport and metabolism]
COG2204 (1) Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains [Signal transduction mechanisms]
COG2223 (1) Nitrate/nitrite transporter [Inorganic ion transport and metabolism]
COG2244 (1) Membrane protein involved in the export of O-antigen and teichoic acid [General function prediction only]
COG2814 (1) Arabinose efflux permease [Carbohydrate transport and metabolism]
COG2908 (2) Uncharacterized protein conserved in bacteria [Function unknown]
COG2971 (1) Predicted N-acetylglucosamine kinase [Carbohydrate transport and metabolism]
COG3185 (2) 4-hydroxyphenylpyruvate dioxygenase and related hemolysins [Amino acid transport and metabolism / General function prediction only]
COG3209 (2) Rhs family protein [Cell envelope biogenesis, outer membrane]
COG3296 (1) Uncharacterized protein conserved in bacteria [Function unknown]
COG3491 (1) Isopenicillin N synthase and related dioxygenases [General function prediction only]
COG3513 (1) Uncharacterized protein conserved in bacteria [Function unknown]
COG3620 (1) Predicted transcriptional regulator with C-terminal CBS domains [Transcription]
COG3751 (2) Predicted proline hydroxylase [Posttranslational modification, protein turnover, chaperones]
COG3943 (1) Virulence protein [General function prediction only]
COG4123 (1) Predicted O-methyltransferase [General function prediction only]
COG4227 (1) Antirestriction protein [DNA replication, recombination, and repair]
COG4339 (1) Uncharacterized protein conserved in bacteria [Function unknown]
COG4591 (4) ABC-type transport system, involved in lipoprotein release, permease component [Cell envelope biogenesis, outer membrane]
COG4992 (1) Ornithine/acetylornithine aminotransferase [Amino acid transport and metabolism]
COG5009 (1) Membrane carboxypeptidase/penicillin-binding protein [Cell envelope biogenesis, outer membrane]
GOS1143 (1) ?
GOS12090 (1) ?
GOS12106160 (1) ?
GOS1236 (1) ?
GOS12690 (1) ?
GOS1344 (1) ?
GOS1394 (1) ?
GOS1409 (2) ?
GOS16 (1) ?
GOS1685 (1) ?
GOS1816 (1) ?
GOS2161 (1) ?
GOS2385 (1) ?
GOS241 (1) ?
GOS2568 (1) ?
GOS26 (1) ?
GOS3464 (1) ?
GOS351 (1) ?
GOS377 (1) ?
GOS4046 (1) ?
GOS4387 (1) ?
GOS4503 (1) ?
GOS5149 (1) ?
GOS530 (1) ?
GOS6297 (1) ?
GOS6621 (1) ?

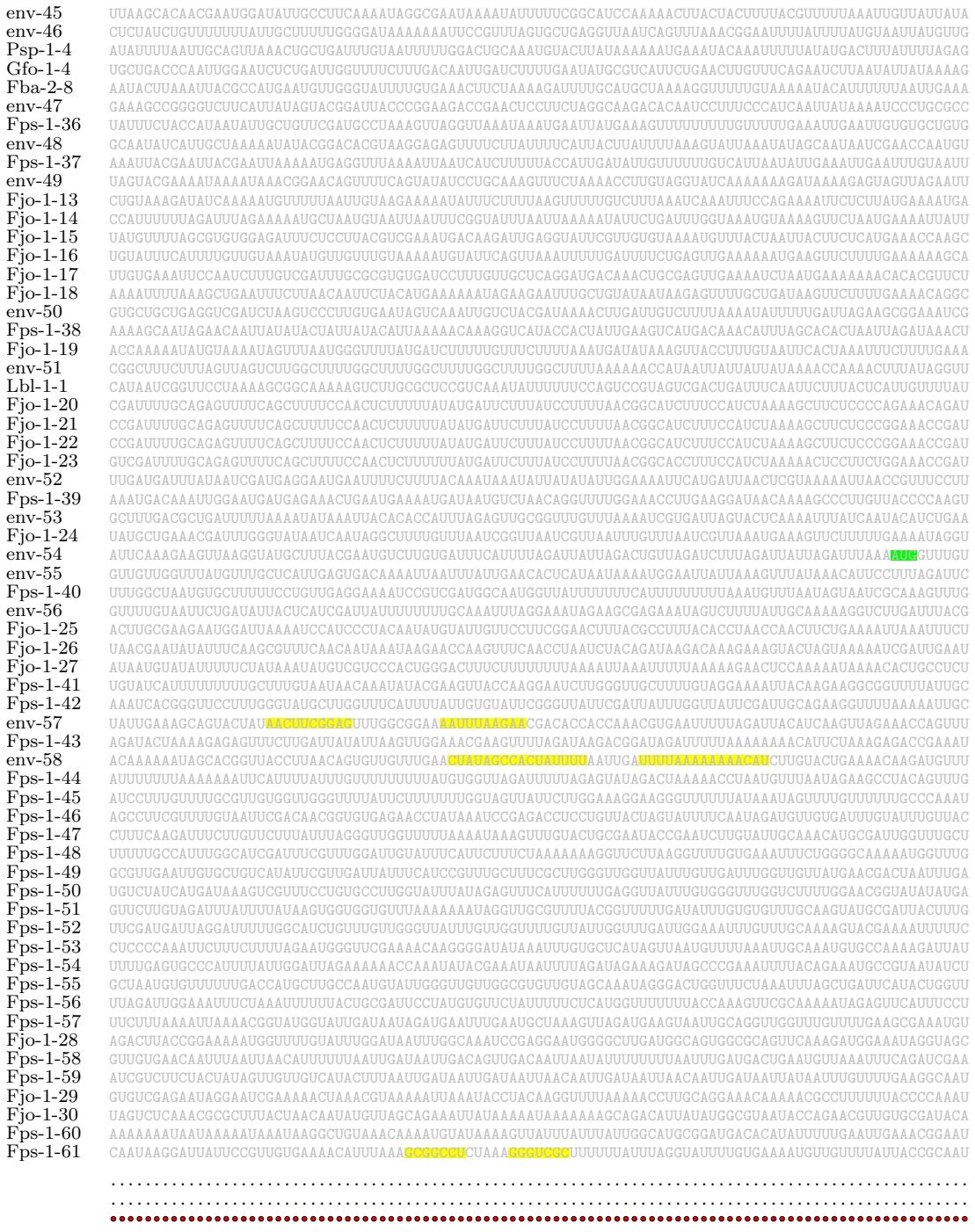
GOS7413 (1) ?
GOS75 (1) ?
GOS7545 (2) ?
GOS7754 (1) ?
GOS8129 (1) ?
GOS8207 (1) ?
GOS869 (1) ?
GOS9899 (1) ?
pfam00012 (1) Hsp70 protein.
pfam00082 (1) Subtilase family.
pfam00085 (1) Thioredoxin.
pfam00106 (5) short chain dehydrogenase.
pfam00121 (1) Triosephosphate isomerase.
pfam00239 (1) Resolvase, N terminal domain.
pfam00268 (2) Ribonucleotide reductase, small chain.
pfam00289 (1) Carbamoyl-phosphate synthase L chain, N-terminal domain.
pfam00301 (1) Rubredoxin.
pfam00313 (1) 'Cold-shock' DNA-binding domain.
pfam00317 (2) Ribonucleotide reductase, all-alpha domain.
pfam00453 (1) Ribosomal protein L20.
pfam00535 (1) Glycosyl transferase.
pfam00572 (1) Ribosomal protein L13.
pfam00665 (1) Integrase core domain.
pfam00692 (1) dUTPase.
pfam00877 (1) NlpC/P60 family.
pfam00903 (1) Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily.
pfam00919 (2) Uncharacterized protein family UPF0004.
pfam00924 (1) Mechanosensitive ion channel.
pfam00933 (1) Glycosyl hydrolase family 3 N terminal domain.
pfam01040 (1) UbiA prenyltransferase family.
pfam01098 (1) Cell cycle protein.
pfam01103 (3) Surface antigen.
pfam01106 (4) NifU-like domain.
pfam01165 (1) Ribosomal protein S21.
pfam01175 (1) Urocanase.
pfam01262 (3) Alanine dehydrogenase/PNT, C-terminal domain.
pfam01336 (1) OB-fold nucleic acid binding domain.
pfam01642 (1) Methylmalonyl-CoA mutase.
pfam01694 (1) Rhomboid family.
pfam01746 (1) tRNA (Guanine-1)-methyltransferase.
pfam01915 (1) Glycosyl hydrolase family 3 C terminal domain.
pfam01964 (2) ThiC family.
pfam02277 (1) Phosphoribosyltransferase.
pfam02371 (1) Transposase IS116/IS110/IS902 family.
pfam02403 (1) Seryl-tRNA synthetase N-terminal domain.
pfam02562 (1) PhoH-like protein.
pfam02771 (1) Acyl-CoA dehydrogenase, N-terminal domain.
pfam02824 (2) TGS domain.
pfam02843 (1) Phosphoribosylglycinamide synthetase, C domain.
pfam02867 (2) Ribonucleotide reductase, barrel domain.
pfam03349 (1) Outer membrane protein transport protein (OMPP1/FadL/TodX).
pfam03477 (3) ATP cone domain.
pfam03551 (1) Transcriptional regulator PadR-like family.
pfam03661 (1) Uncharacterized protein family (UPF0121).
pfam03976 (1) Domain of unknown function (DUF344).
pfam03993 (1) Domain of Unknown Function (DUF349).
pfam04101 (1) Glycosyltransferase family 28 C-terminal domain.
pfam04209 (2) homogentisate 1,2-dioxygenase.
pfam04389 (1) Peptidase family M28.
pfam04760 (1) Translation initiation factor IF-2, N-terminal region.
pfam05221 (1) S-adenosyl-L-homocysteine hydrolase.
pfam05222 (3) Alanine dehydrogenase/PNT, N-terminal domain.
pfam05661 (1) Protein of unknown function (DUF808).
pfam06052 (1) 3-hydroxyanthranilic acid dioxygenase.
smart00387 (1) Histidine kinase-like ATPases; Histidine kinase-, DNA gyrase B-, phytochrome-like ATPases.
smart00477 (1) DNA/RNA non-specific endonuclease; prokaryotic and eukaryotic double- and single-stranded DNA and RNA endonucleases also present in phosphodiesterases
smart00732 (1) Likely ribonuclease with RNase H fold.

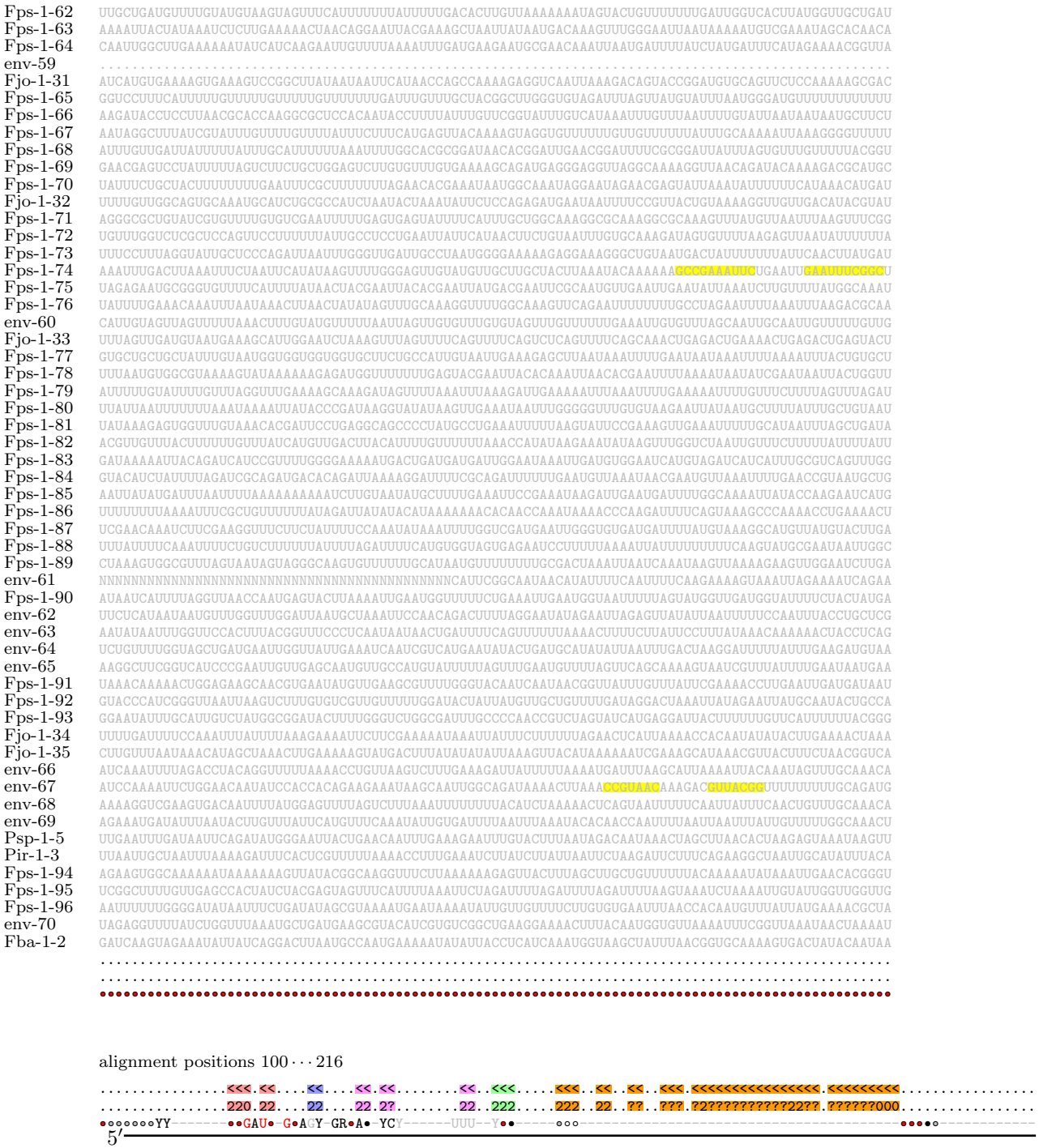
32.4 Multiple-sequence alignment

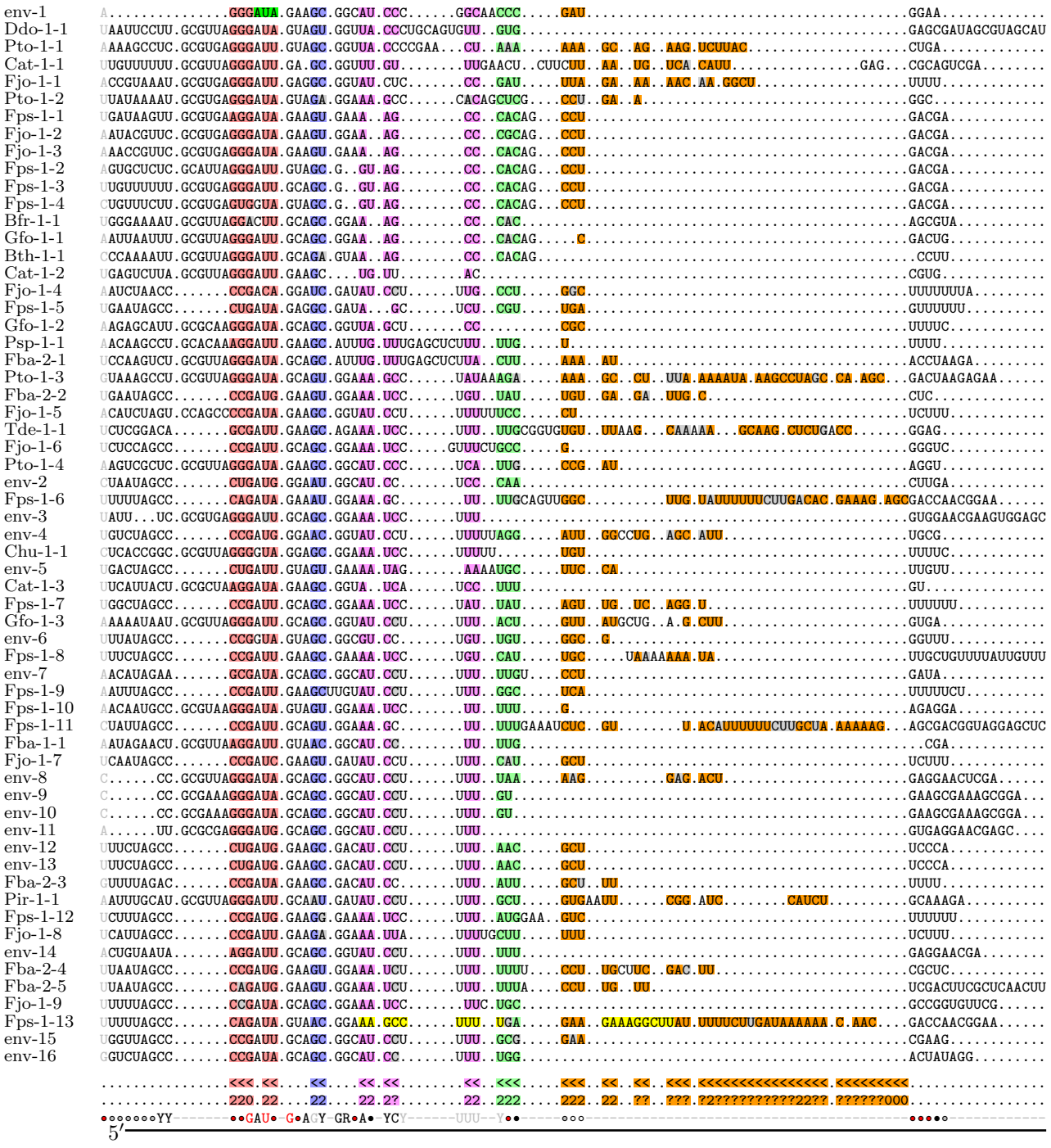
Each Flavo-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 32.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Flavo-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

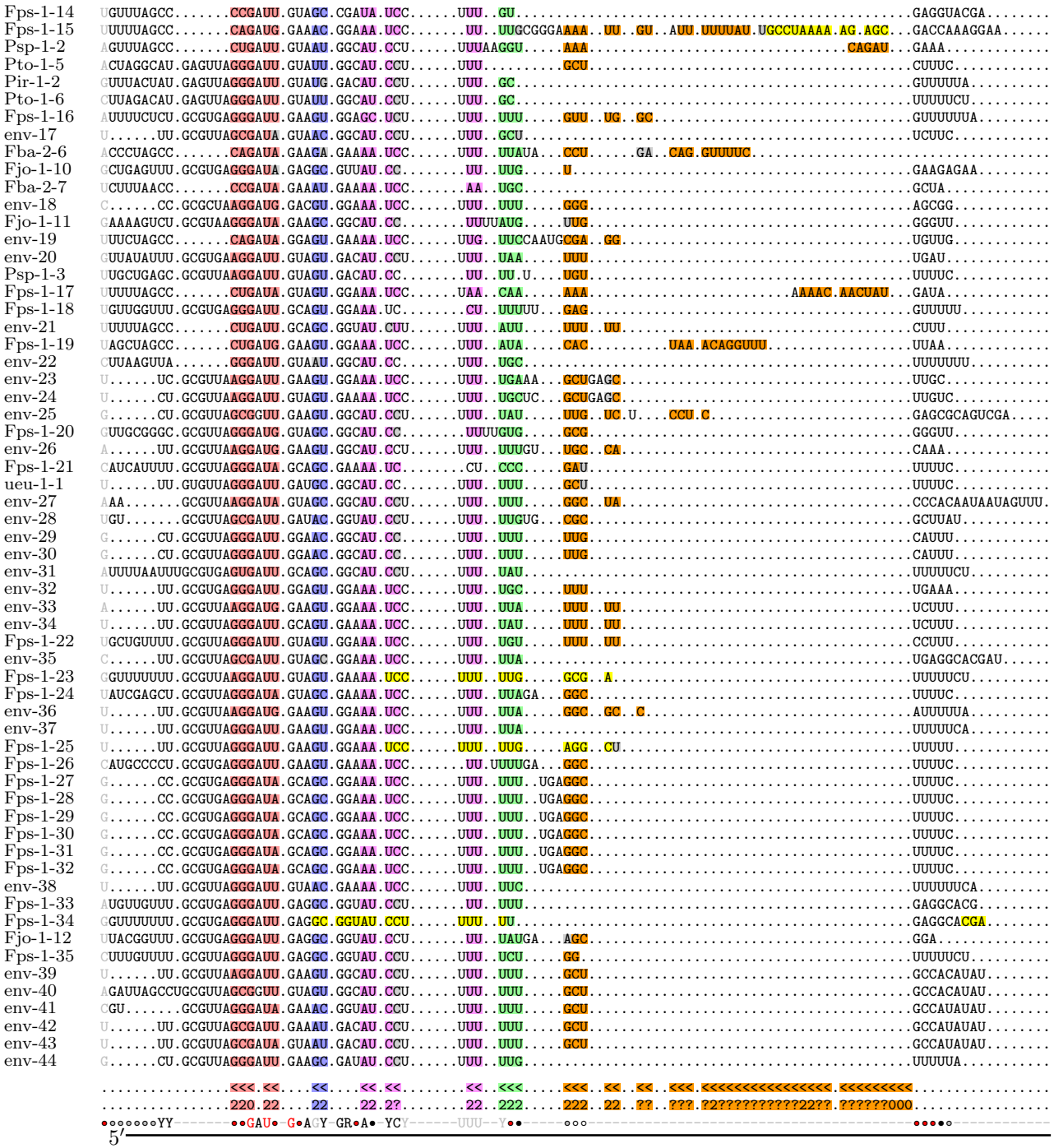

alignment positions 1 · · · 99
env-1
Ddo-1-1
Pto-1-1
Cat-1-1
Fjo-1-1
Pto-1-2
Fps-1-1
Fjo-1-2
Fjo-1-3
Fps-1-2
Fps-1-3
Fps-1-4
Bfr-1-1
Gfo-1-1
Bth-1-1
Cat-1-2
Fjo-1-4
Fps-1-5
Gfo-1-2
Psp-1-1
Fba-2-1
Pto-1-3
Fba-2-2
Fjo-1-5
Tde-1-1
Fjo-1-6
Pto-1-4
env-2
Fps-1-6
env-3
env-4
Chu-1-1
env-5
Cat-1-3
Fps-1-7
Gfo-1-3
env-6
Fps-1-8
env-7
Fps-1-9
Fps-1-10
Fps-1-11
Fba-1-1
Fjo-1-7
env-8
env-9
env-10
env-11
env-12
env-13
Fba-2-3
Pir-1-1
Fps-1-12
Fjo-1-8
env-14
Fba-2-4
Fba-2-5
Fjo-1-9
Fps-1-13
env-15

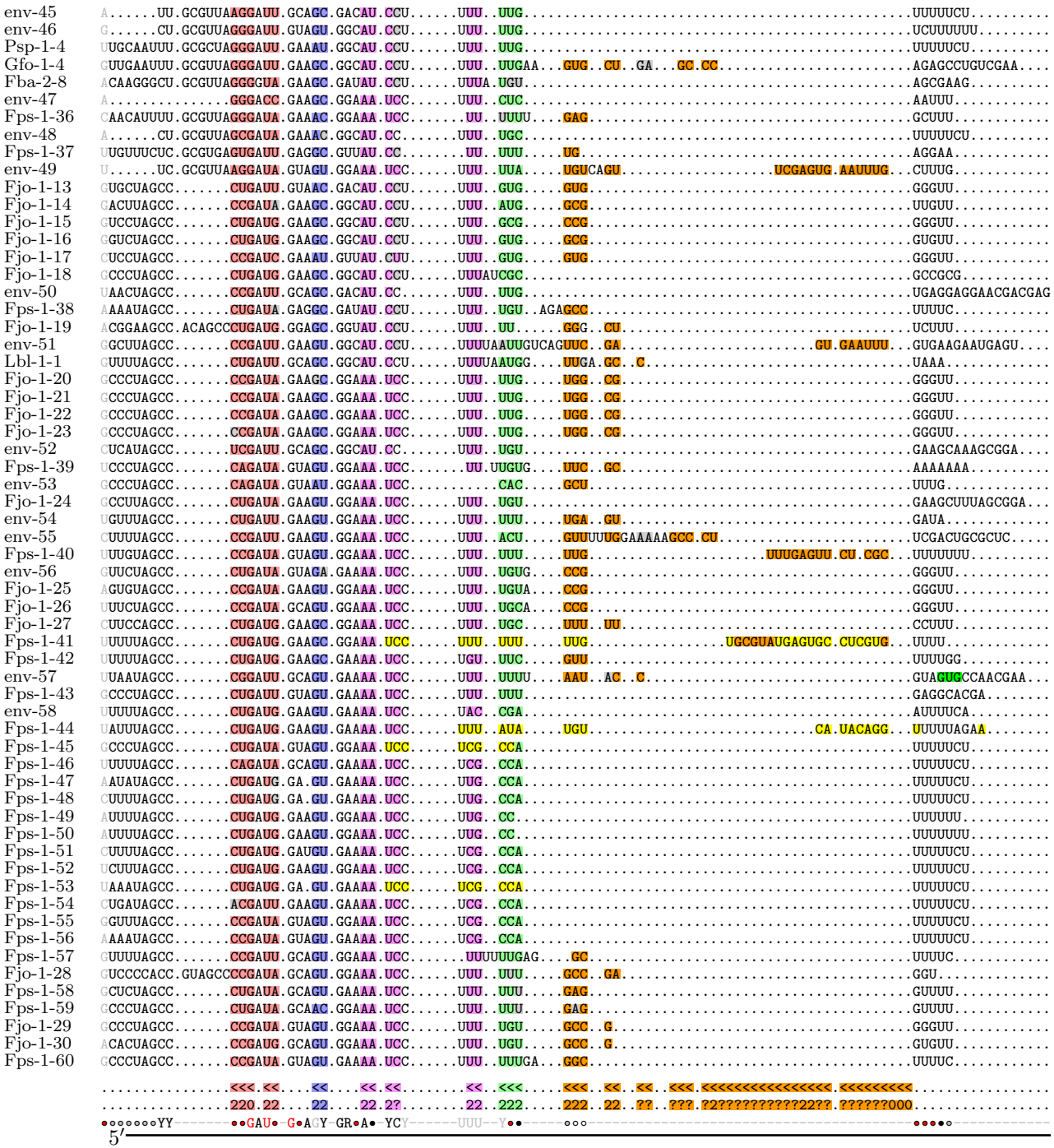

env-16
Fps-1-14
Fps-1-15
Psp-1-2
Pto-1-5
Pir-1-2
Pto-1-6
Fps-1-16
env-17
Fba-2-6
Fjo-1-10
Fba-2-7
env-18
Fjo-1-11
env-19
env-20
Psp-1-3
Fps-1-17
Fps-1-18
env-21
Fps-1-19
env-22
env-23
env-24
env-25
Fps-1-20
env-26
Fps-1-21
ueu-1-1
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
Fps-1-22
env-35
Fps-1-23
Fps-1-24
env-36
env-37
Fps-1-25
Fps-1-26
Fps-1-27
Fps-1-28
Fps-1-29
Fps-1-30
Fps-1-31
Fps-1-32
env-38
Fps-1-33
Fps-1-34
Fjo-1-12
Fps-1-35
env-39
env-40
env-41
env-42
env-43
env-44

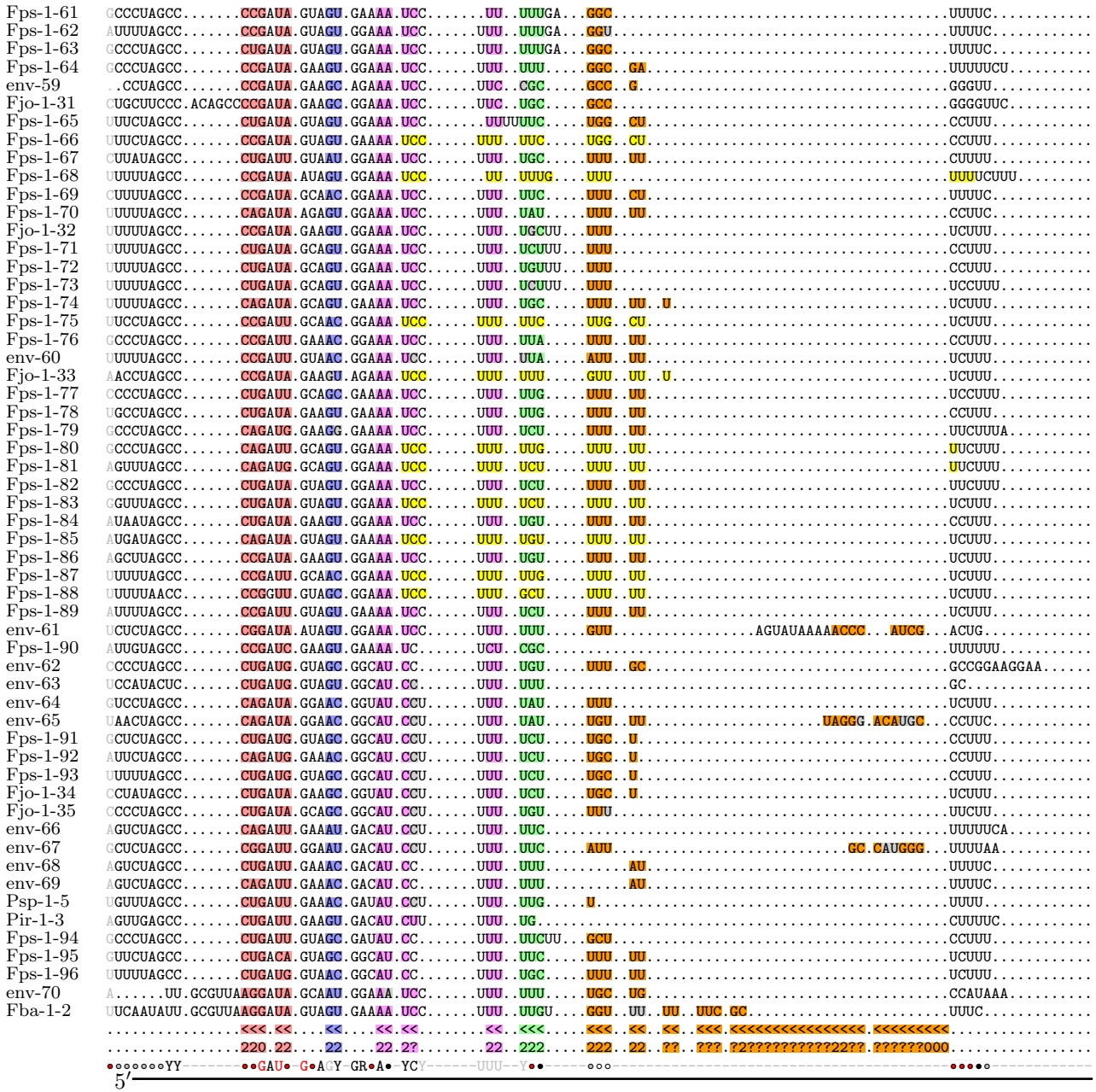

env-45
env-46
Psp-1-4
Gfo-1-4
Fba-2-8
env-47
Fps-1-36
env-48
Fps-1-37
env-49
Fjo-1-13
Fjo-1-14
Fjo-1-15
Fjo-1-16
Fjo-1-17
Fjo-1-18
env-50
Fps-1-38
Fjo-1-19
env-51
Lbl-1-1
Fjo-1-20
Fjo-1-21
Fjo-1-22
Fjo-1-23
env-52
Fps-1-39
env-53
Fjo-1-24
env-54
env-55
Fps-1-40
env-56
Fjo-1-25
Fjo-1-26
Fjo-1-27
Fps-1-41
Fps-1-42
env-57
Fps-1-43
env-58
Fps-1-44
Fps-1-45
Fps-1-46
Fps-1-47
Fps-1-48
Fps-1-49
Fps-1-50
Fps-1-51
Fps-1-52
Fps-1-53
Fps-1-54
Fps-1-55
Fps-1-56
Fps-1-57
Fjo-1-28
Fps-1-58
Fps-1-59
Fjo-1-29
Fjo-1-30
Fps-1-60
Fps-1-61

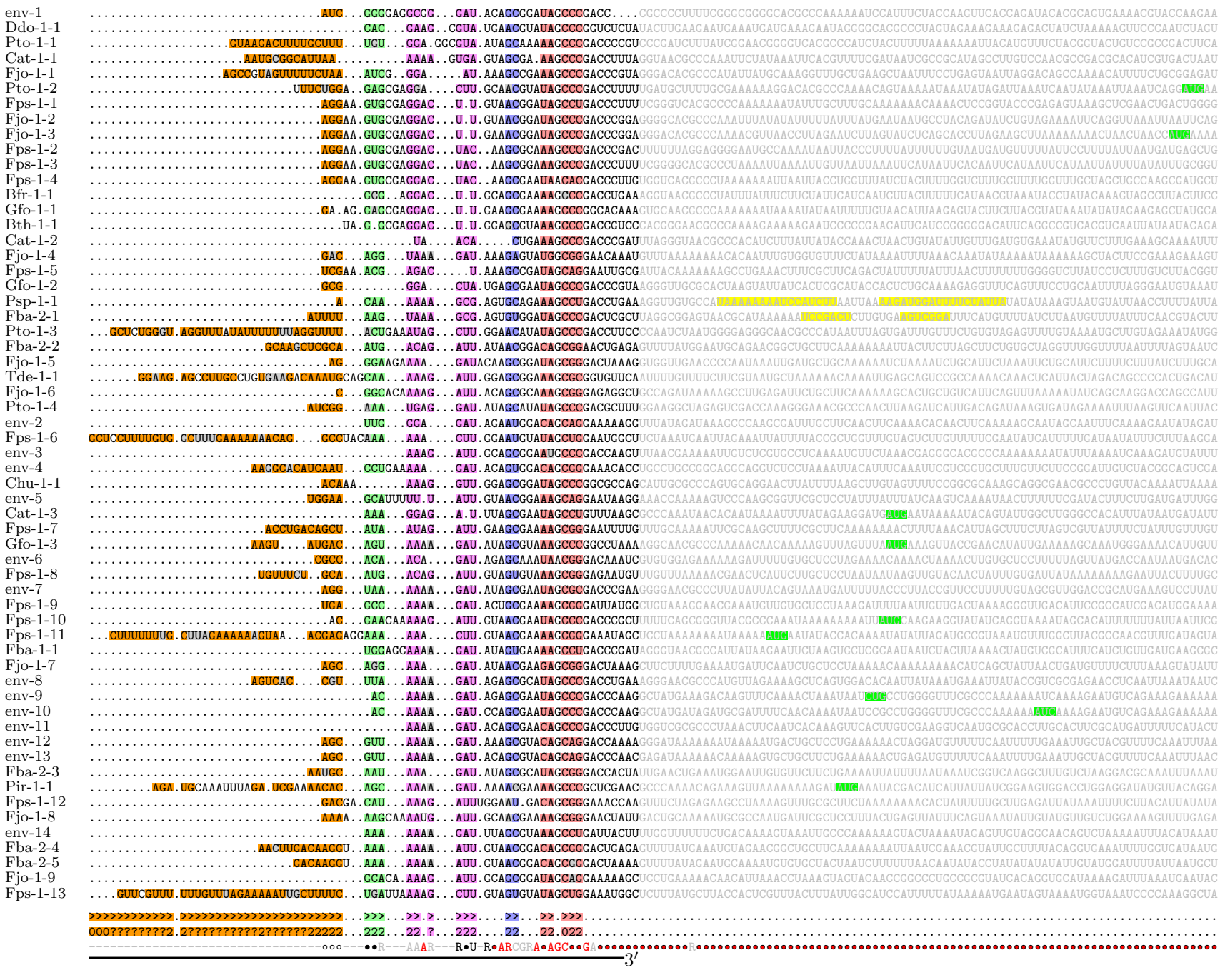

Fps-1-62
Fps-1-63
Fps-1-64
env-59
Fjo-1-31
Fps-1-65
Fps-1-66
Fps-1-67
Fps-1-68
Fps-1-69
Fps-1-70
Fjo-1-32
Fps-1-71
Fps-1-72
Fps-1-73
Fps-1-74
Fps-1-75
Fps-1-76
env-60
Fjo-1-33
Fps-1-77
Fps-1-78
Fps-1-79
Fps-1-80
Fps-1-81
Fps-1-82
Fps-1-83
Fps-1-84
Fps-1-85
Fps-1-86
Fps-1-87
Fps-1-88
Fps-1-89
env-61
Fps-1-90
env-62
env-63
env-64
env-65
Fps-1-91
Fps-1-92
Fps-1-93
Fjo-1-34
Fjo-1-35
env-66
env-67
env-68
env-69
Psp-1-5
Pir-1-3
Fps-1-94
Fps-1-95
Fps-1-96
env-70
Fba-1-2
alignment positions 100 · · · 216
5′

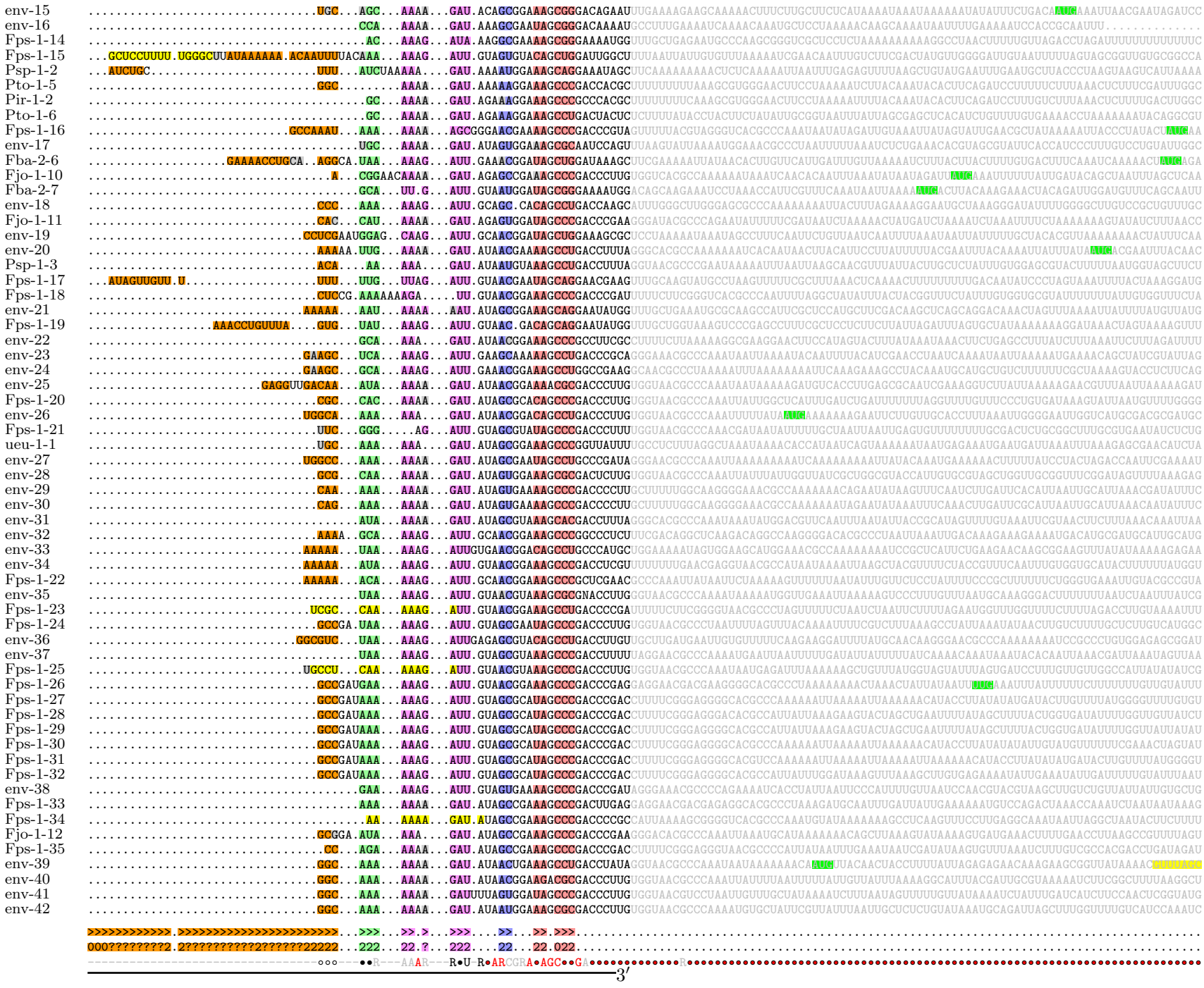

env-1
Ddo-1-1
Pto-1-1
Cat-1-1
Fjo-1-1
Pto-1-2
Fps-1-1
Fjo-1-2
Fjo-1-3
Fps-1-2
Fps-1-3
Fps-1-4
Bfr-1-1
Gfo-1-1
Bth-1-1
Cat-1-2
Fjo-1-4
Fps-1-5
Gfo-1-2
Psp-1-1
Fba-2-1
Pto-1-3
Fba-2-2
Fjo-1-5
Tde-1-1
Fjo-1-6
Pto-1-4
env-2
Fps-1-6
env-3
env-4
Chu-1-1
env-5
Cat-1-3
Fps-1-7
Gfo-1-3
env-6
Fps-1-8
env-7
Fps-1-9
Fps-1-10
Fps-1-11
Fba-1-1
Fjo-1-7
env-8
env-9
env-10
env-11
env-12
env-13
Fba-2-3
Pir-1-1
Fps-1-12
Fjo-1-8
env-14
Fba-2-4
Fba-2-5
Fjo-1-9
Fps-1-13
env-15
env-16
5′

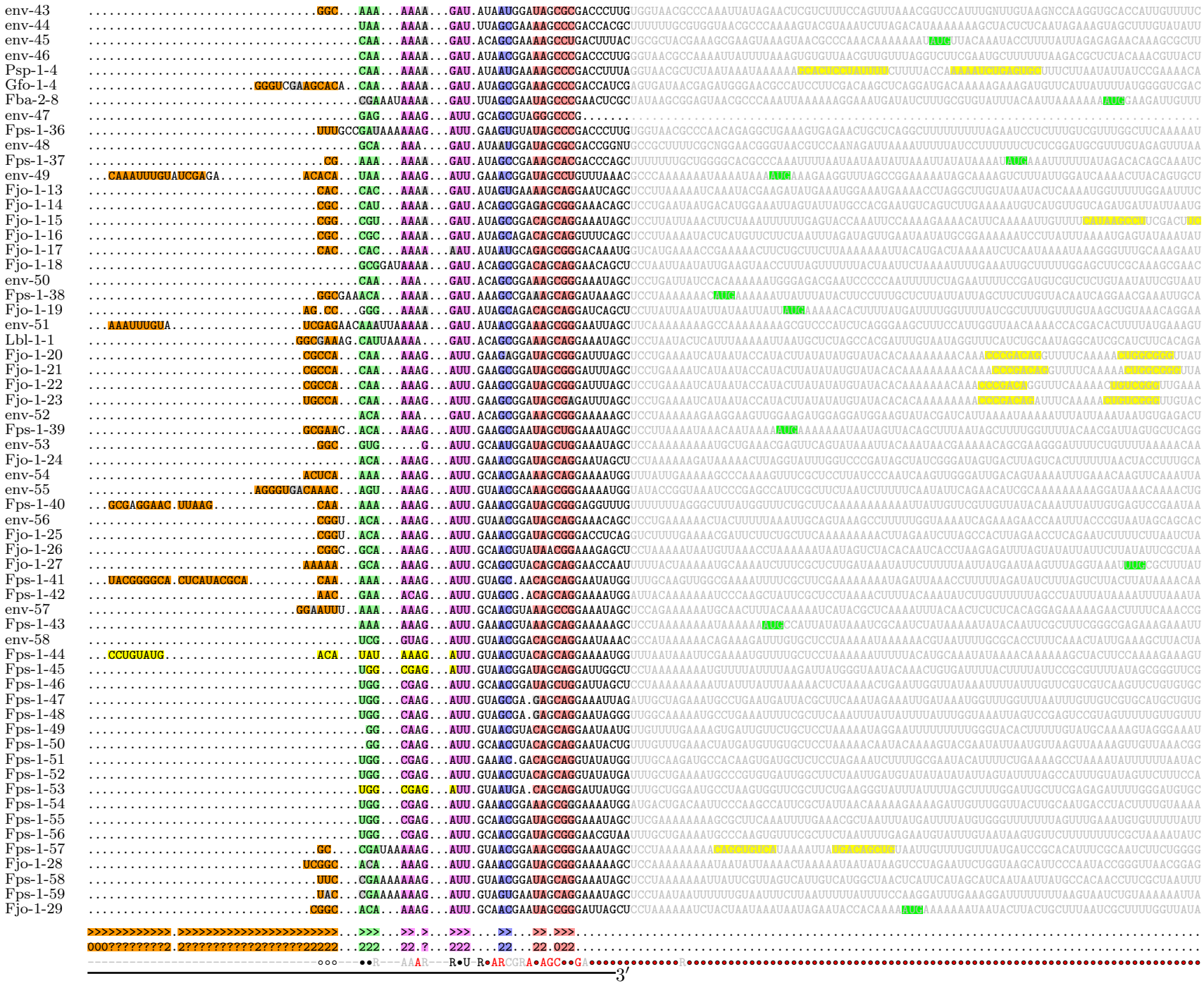

Fps-1-14
Fps-1-15
Psp-1-2
Pto-1-5
Pir-1-2
Pto-1-6
Fps-1-16
env-17
Fba-2-6
Fjo-1-10
Fba-2-7
env-18
Fjo-1-11
env-19
env-20
Psp-1-3
Fps-1-17
Fps-1-18
env-21
Fps-1-19
env-22
env-23
env-24
env-25
Fps-1-20
env-26
Fps-1-21
ueu-1-1
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
Fps-1-22
env-35
Fps-1-23
Fps-1-24
env-36
env-37
Fps-1-25
Fps-1-26
Fps-1-27
Fps-1-28
Fps-1-29
Fps-1-30
Fps-1-31
Fps-1-32
env-38
Fps-1-33
Fps-1-34
Fjo-1-12
Fps-1-35
env-39
env-40
env-41
env-42
env-43
env-44
5′

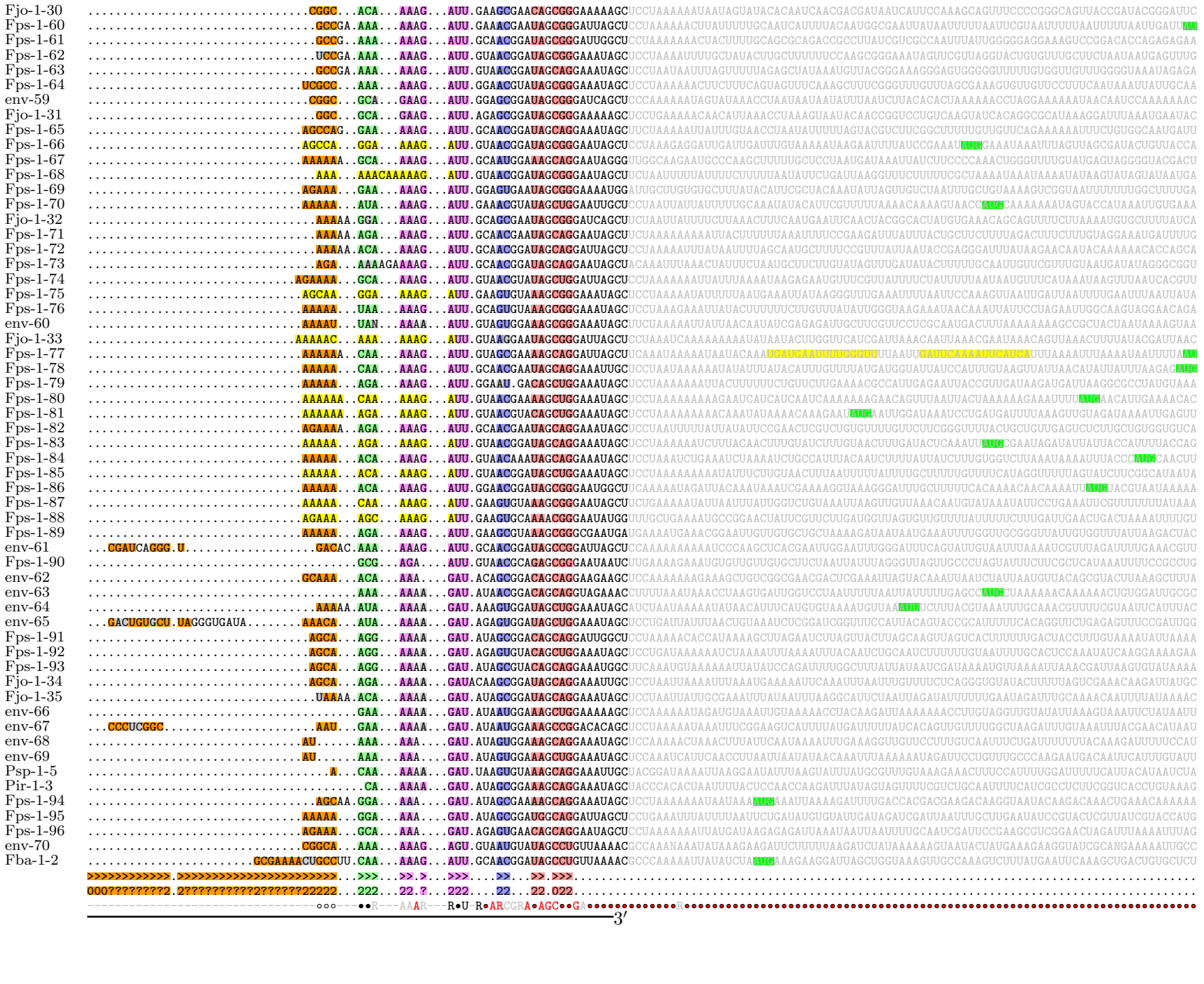

env-45
env-46
Psp-1-4
Gfo-1-4
Fba-2-8
env-47
Fps-1-36
env-48
Fps-1-37
env-49
Fjo-1-13
Fjo-1-14
Fjo-1-15
Fjo-1-16
Fjo-1-17
Fjo-1-18
env-50
Fps-1-38
Fjo-1-19
env-51
Lbl-1-1
Fjo-1-20
Fjo-1-21
Fjo-1-22
Fjo-1-23
env-52
Fps-1-39
env-53
Fjo-1-24
env-54
env-55
Fps-1-40
env-56
Fjo-1-25
Fjo-1-26
Fjo-1-27
Fps-1-41
Fps-1-42
env-57
Fps-1-43
env-58
Fps-1-44
Fps-1-45
Fps-1-46
Fps-1-47
Fps-1-48
Fps-1-49
Fps-1-50
Fps-1-51
Fps-1-52
Fps-1-53
Fps-1-54
Fps-1-55
Fps-1-56
Fps-1-57
Fjo-1-28
Fps-1-58
Fps-1-59
Fjo-1-29
Fjo-1-30
Fps-1-60
5′

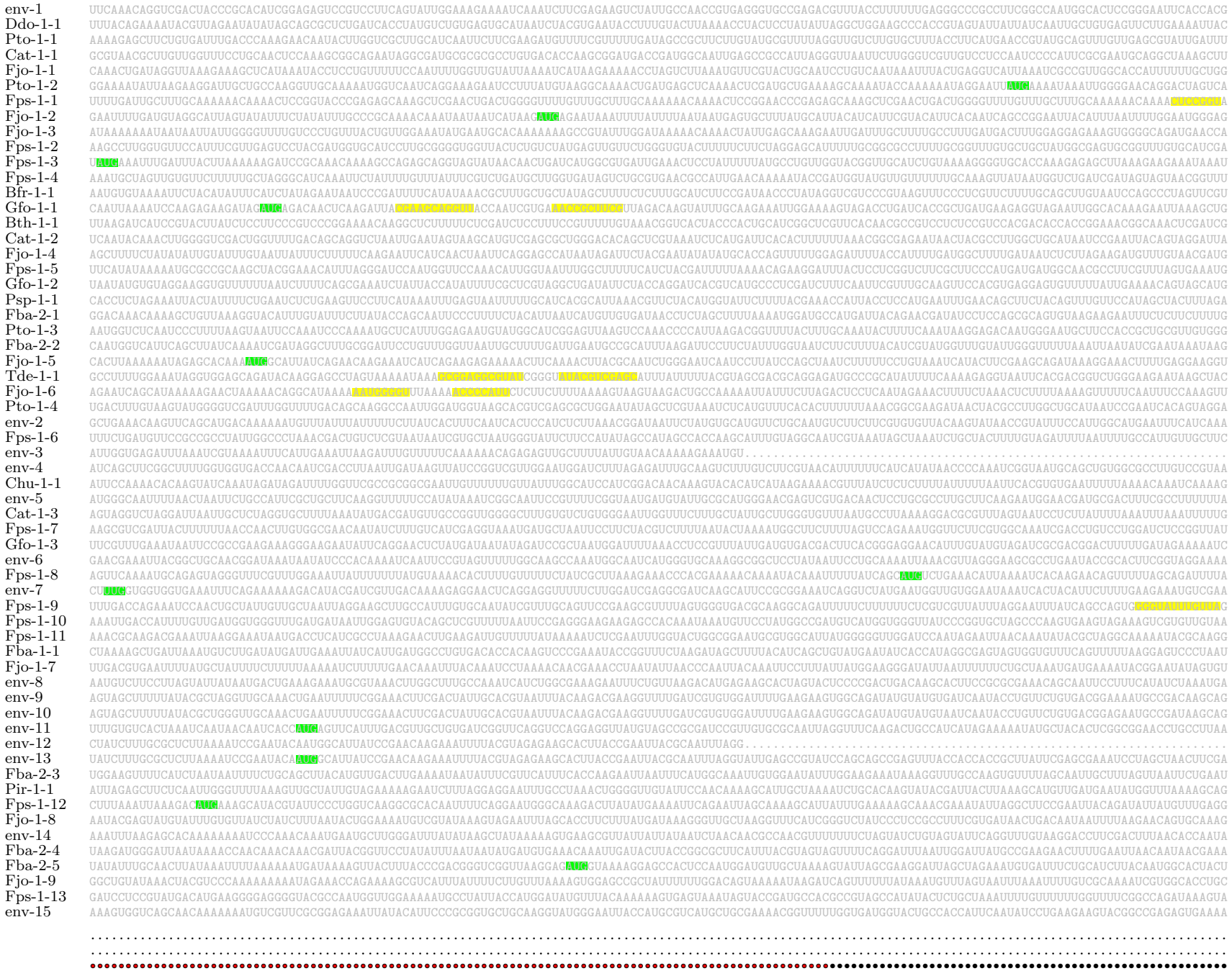

Fps-1-61
Fps-1-62
Fps-1-63
Fps-1-64
env-59
Fjo-1-31
Fps-1-65
Fps-1-66
Fps-1-67
Fps-1-68
Fps-1-69
Fps-1-70
Fjo-1-32
Fps-1-71
Fps-1-72
Fps-1-73
Fps-1-74
Fps-1-75
Fps-1-76
env-60
Fjo-1-33
Fps-1-77
Fps-1-78
Fps-1-79
Fps-1-80
Fps-1-81
Fps-1-82
Fps-1-83
Fps-1-84
Fps-1-85
Fps-1-86
Fps-1-87
Fps-1-88
Fps-1-89
env-61
Fps-1-90
env-62
env-63
env-64
env-65
Fps-1-91
Fps-1-92
Fps-1-93
Fjo-1-34
Fjo-1-35
env-66
env-67
env-68
env-69
Psp-1-5
Pir-1-3
Fps-1-94
Fps-1-95
Fps-1-96
env-70
Fba-1-2
5′

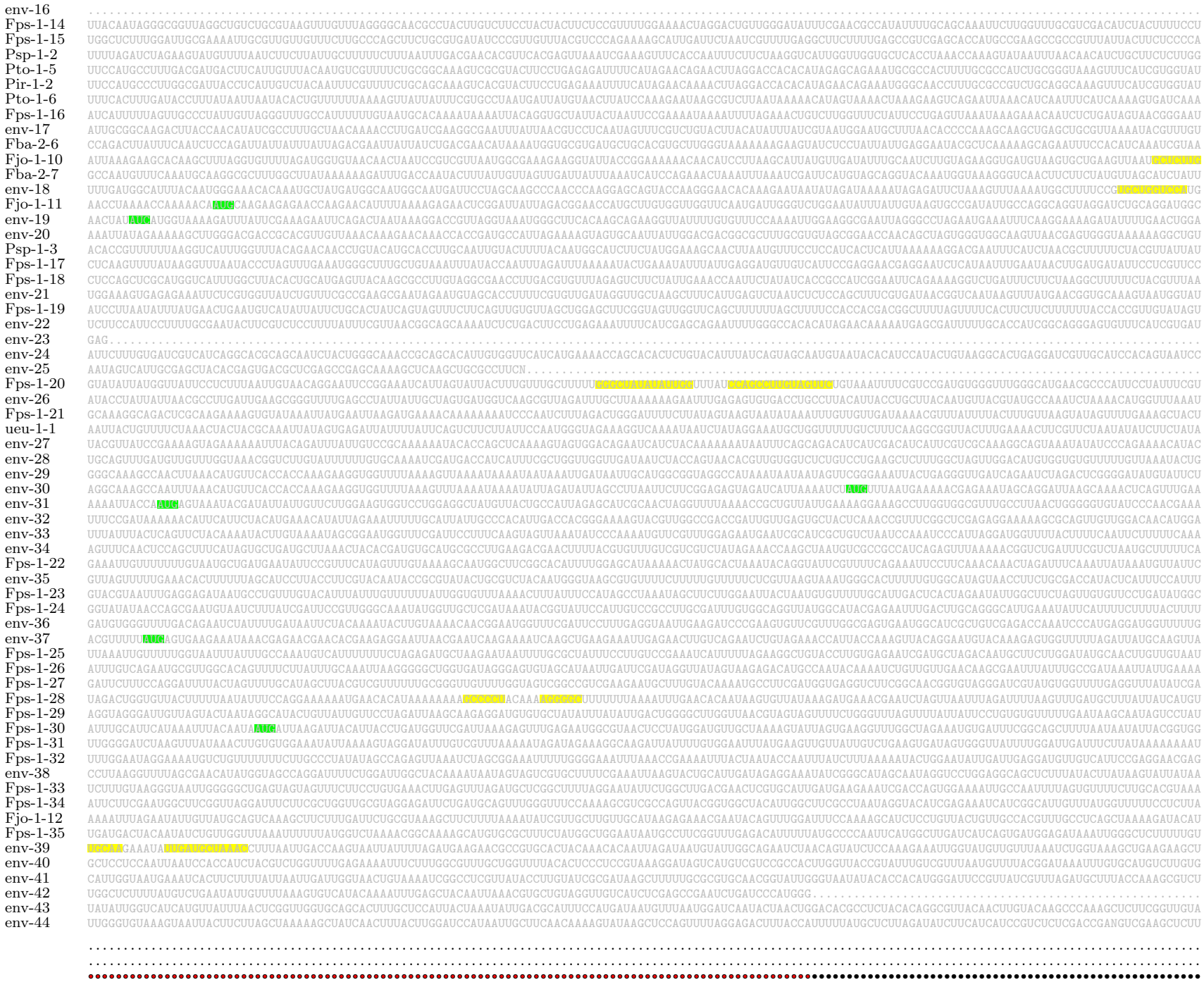

alignment positions 217 · · · 376
env-1
Ddo-1-1
Pto-1-1
Cat-1-1
Fjo-1-1
Pto-1-2
Fps-1-1
Fjo-1-2
Fjo-1-3
Fps-1-2
Fps-1-3
Fps-1-4
Bfr-1-1
Gfo-1-1
Bth-1-1
Cat-1-2
Fjo-1-4
Fps-1-5
Gfo-1-2
Psp-1-1
Fba-2-1
Pto-1-3
Fba-2-2
Fjo-1-5
Tde-1-1
Fjo-1-6
Pto-1-4
env-2
Fps-1-6
env-3
env-4
Chu-1-1
env-5
Cat-1-3
Fps-1-7
Gfo-1-3
env-6
Fps-1-8
env-7
Fps-1-9
Fps-1-10
Fps-1-11
Fba-1-1
Fjo-1-7
env-8
env-9
env-10
env-11
env-12
env-13
Fba-2-3
Pir-1-1
Fps-1-12
Fjo-1-8
env-14
Fba-2-4
Fba-2-5
Fjo-1-9
Fps-1-13
3′

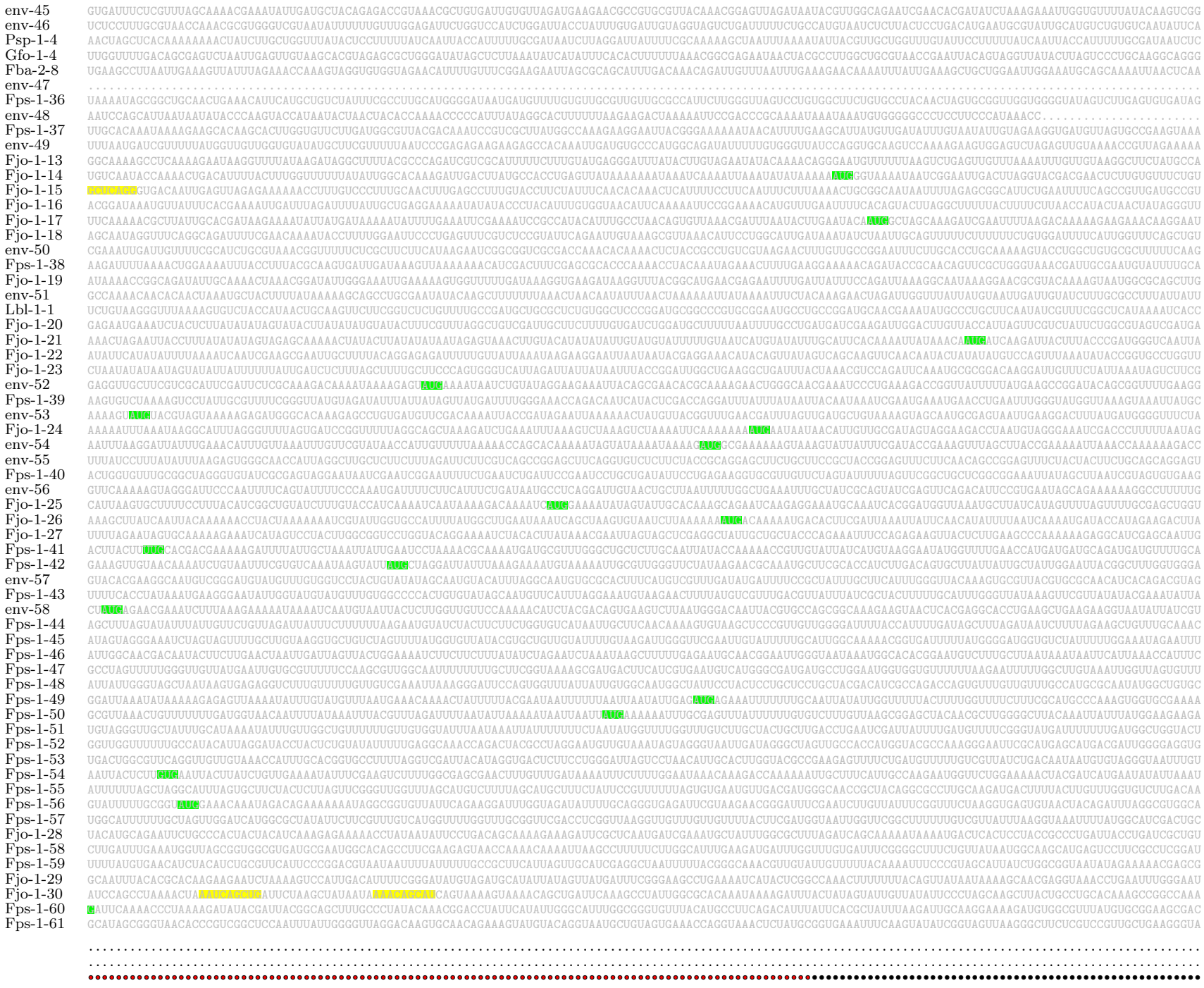

env-15
Fps-1-14
Fps-1-15
Psp-1-2
Pto-1-5
Pir-1-2
Pto-1-6
Fps-1-16
env-17
Fba-2-6
Fjo-1-10
Fba-2-7
env-18
Fjo-1-11
env-19
env-20
Psp-1-3
Fps-1-17
Fps-1-18
env-21
Fps-1-19
env-22
env-23
env-24
env-25
Fps-1-20
env-26
Fps-1-21
ueu-1-1
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
Fps-1-22
env-35
Fps-1-23
Fps-1-24
env-36
env-37
Fps-1-25
Fps-1-26
Fps-1-27
Fps-1-28
Fps-1-29
Fps-1-30
Fps-1-31
Fps-1-32
env-38
Fps-1-33
Fps-1-34
Fjo-1-12
Fps-1-35
env-39
env-40
env-41
env-42
3′

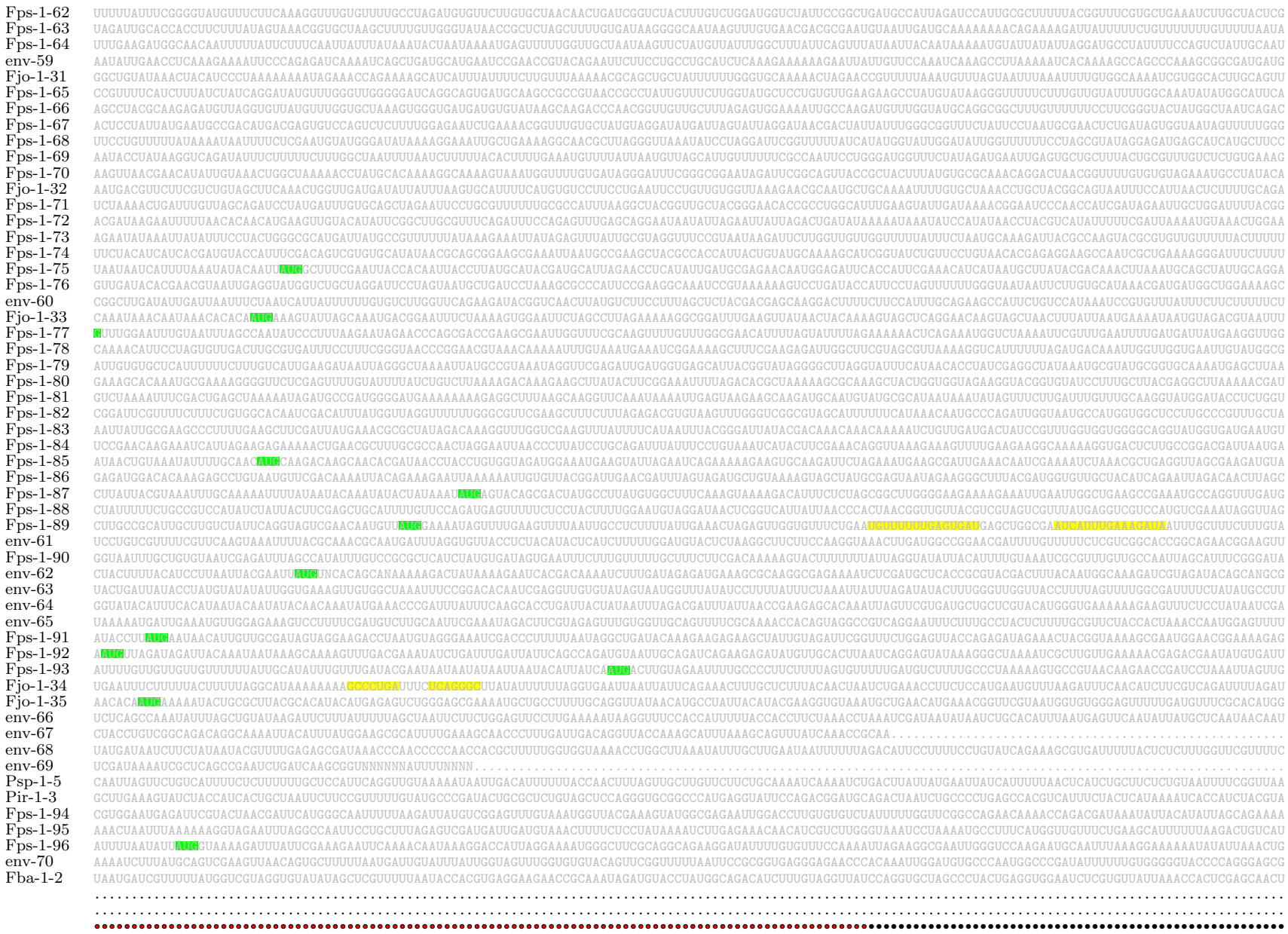

env-43
env-44
env-45
env-46
Psp-1-4
Gfo-1-4
Fba-2-8
env-47
Fps-1-36
env-48
Fps-1-37
env-49
Fjo-1-13
Fjo-1-14
Fjo-1-15
Fjo-1-16
Fjo-1-17
Fjo-1-18
env-50
Fps-1-38
Fjo-1-19
env-51
Lbl-1-1
Fjo-1-20
Fjo-1-21
Fjo-1-22
Fjo-1-23
env-52
Fps-1-39
env-53
Fjo-1-24
env-54
env-55
Fps-1-40
env-56
Fjo-1-25
Fjo-1-26
Fjo-1-27
Fps-1-41
Fps-1-42
env-57
Fps-1-43
env-58
Fps-1-44
Fps-1-45
Fps-1-46
Fps-1-47
Fps-1-48
Fps-1-49
Fps-1-50
Fps-1-51
Fps-1-52
Fps-1-53
Fps-1-54
Fps-1-55
Fps-1-56
Fps-1-57
Fjo-1-28
Fps-1-58
Fps-1-59
Fjo-1-29
3′


Fjo-1-30
Fps-1-60
Fps-1-61
Fps-1-62
Fps-1-63
Fps-1-64
env-59
Fjo-1-31
Fps-1-65
Fps-1-66
Fps-1-67
Fps-1-68
Fps-1-69
Fps-1-70
Fjo-1-32
Fps-1-71
Fps-1-72
Fps-1-73
Fps-1-74
Fps-1-75
Fps-1-76
env-60
Fjo-1-33
Fps-1-77
Fps-1-78
Fps-1-79
Fps-1-80
Fps-1-81
Fps-1-82
Fps-1-83
Fps-1-84
Fps-1-85
Fps-1-86
Fps-1-87
Fps-1-88
Fps-1-89
env-61
Fps-1-90
env-62
env-63
env-64
env-65
Fps-1-91
Fps-1-92
Fps-1-93
Fjo-1-34
Fjo-1-35
env-66
env-67
env-68
env-69
Psp-1-5
Pir-1-3
Fps-1-94
Fps-1-95
Fps-1-96
env-70
Fba-1-2
3′

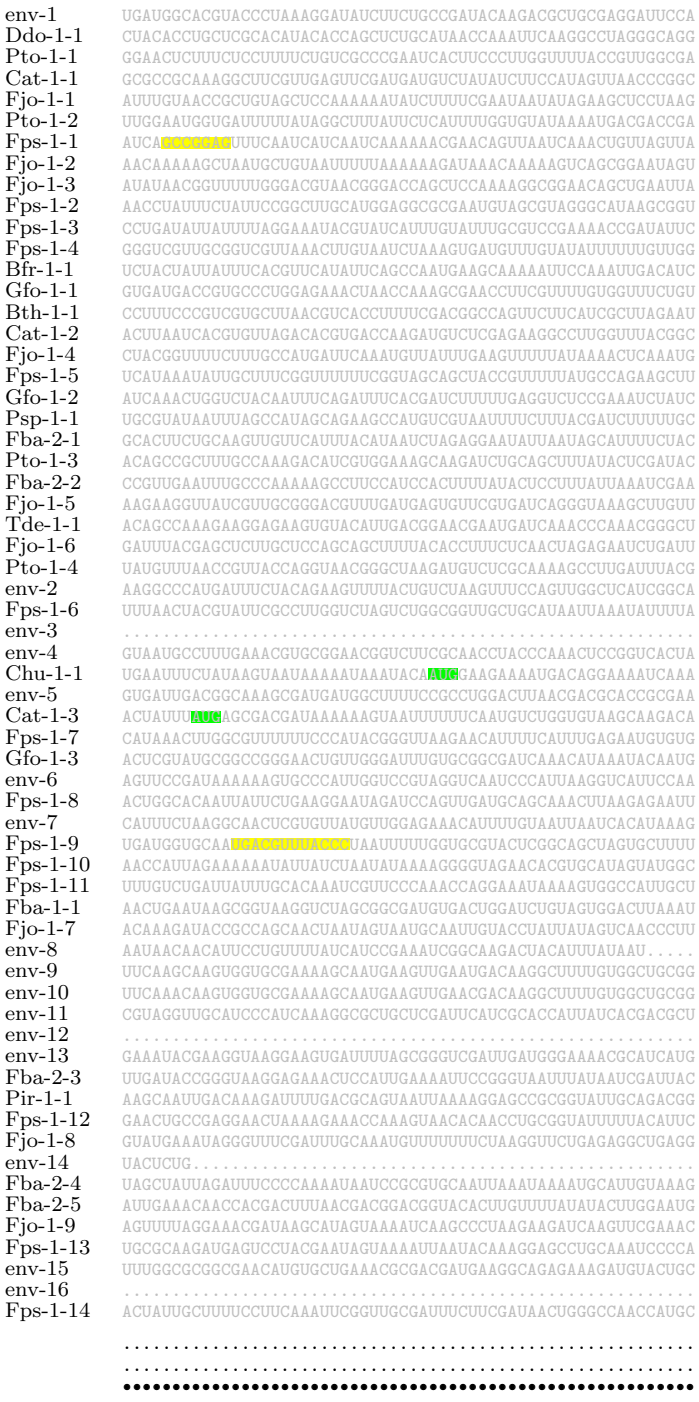

alignment positions 377 · · · 536
env-1
Ddo-1-1
Pto-1-1
Cat-1-1
Fjo-1-1
Pto-1-2
Fps-1-1
Fjo-1-2
Fjo-1-3
Fps-1-2
Fps-1-3
Fps-1-4
Bfr-1-1
Gfo-1-1
Bth-1-1
Cat-1-2
Fjo-1-4
Fps-1-5
Gfo-1-2
Psp-1-1
Fba-2-1
Pto-1-3
Fba-2-2
Fjo-1-5
Tde-1-1
Fjo-1-6
Pto-1-4
env-2
Fps-1-6
env-3
env-4
Chu-1-1
env-5
Cat-1-3
Fps-1-7
Gfo-1-3
env-6
Fps-1-8
env-7
Fps-1-9
Fps-1-10
Fps-1-11
Fba-1-1
Fjo-1-7
env-8
env-9
env-10
env-11
env-12
env-13
Fba-2-3
Pir-1-1
Fps-1-12
Fjo-1-8
env-14
Fba-2-4
Fba-2-5
Fjo-1-9
Fps-1-13
env-15

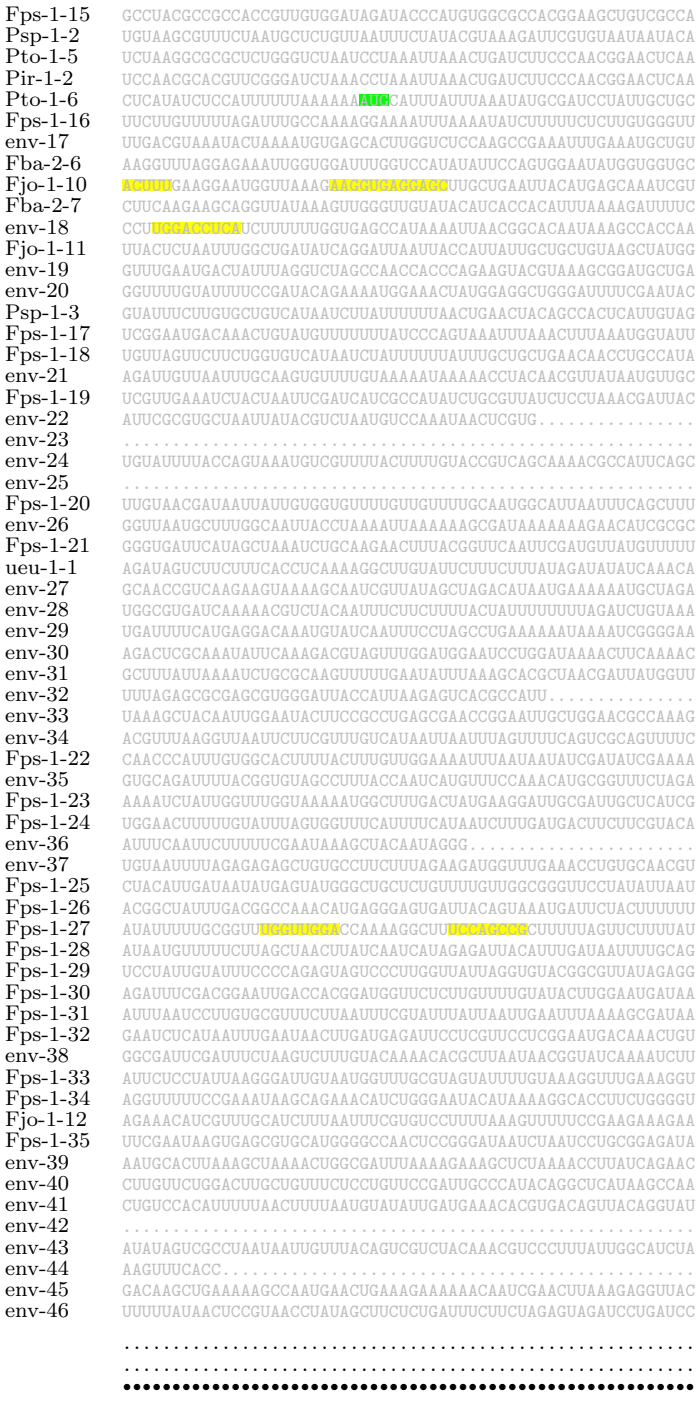

env-16
Fps-1-14
Fps-1-15
Psp-1-2
Pto-1-5
Pir-1-2
Pto-1-6
Fps-1-16
env-17
Fba-2-6
Fjo-1-10
Fba-2-7
env-18
Fjo-1-11
env-19
env-20
Psp-1-3
Fps-1-17
Fps-1-18
env-21
Fps-1-19
env-22
env-23
env-24
env-25
Fps-1-20
env-26
Fps-1-21
ueu-1-1
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
Fps-1-22
env-35
Fps-1-23
Fps-1-24
env-36
env-37
Fps-1-25
Fps-1-26
Fps-1-27
Fps-1-28
Fps-1-29
Fps-1-30
Fps-1-31
Fps-1-32
env-38
Fps-1-33
Fps-1-34
Fjo-1-12
Fps-1-35
env-39
env-40
env-41
env-42
env-43
env-44

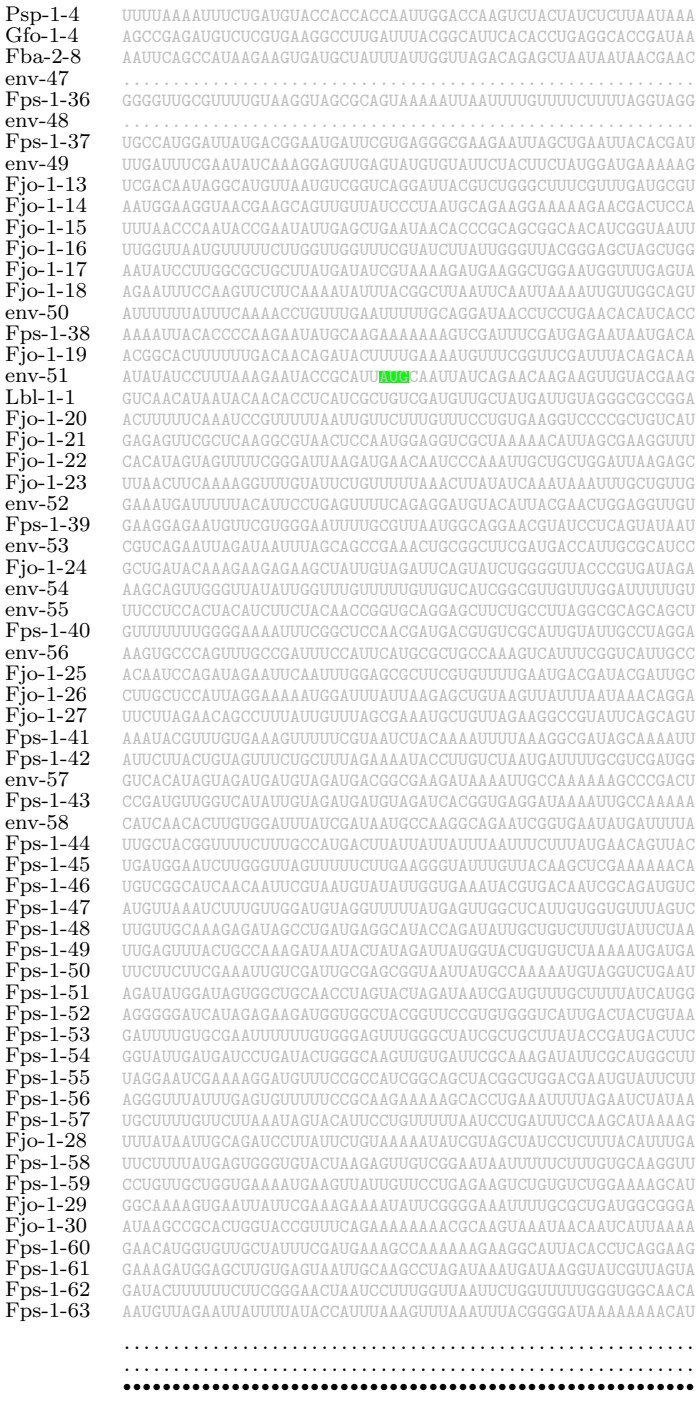

env-45
Fps-1-36
env-48
Fps-1-37
env-49
Fjo-1-13
Fjo-1-14
Fjo-1-15
Fjo-1-16
Fjo-1-17
Fjo-1-18
env-50
Fps-1-38
Fjo-1-19
env-51
Lbl-1-1
Fjo-1-20
Fjo-1-21
Fjo-1-22
Fjo-1-23
env-52
Fps-1-39
env-53
Fjo-1-24
env-54
env-55
Fps-1-40
env-56
Fjo-1-25
Fjo-1-26
Fjo-1-27
Fps-1-41
Fps-1-42
env-57
Fps-1-43
env-58
Fps-1-44
Fps-1-45
Fps-1-46
Fps-1-47
Fps-1-48
Fps-1-49
Fps-1-50
Fps-1-51
Fps-1-52
Fps-1-53
Fps-1-54
Fps-1-55
Fps-1-56
Fps-1-57
Fjo-1-28
Fps-1-58
Fps-1-59
Fjo-1-29
Fjo-1-30
Fps-1-60
Fps-1-61

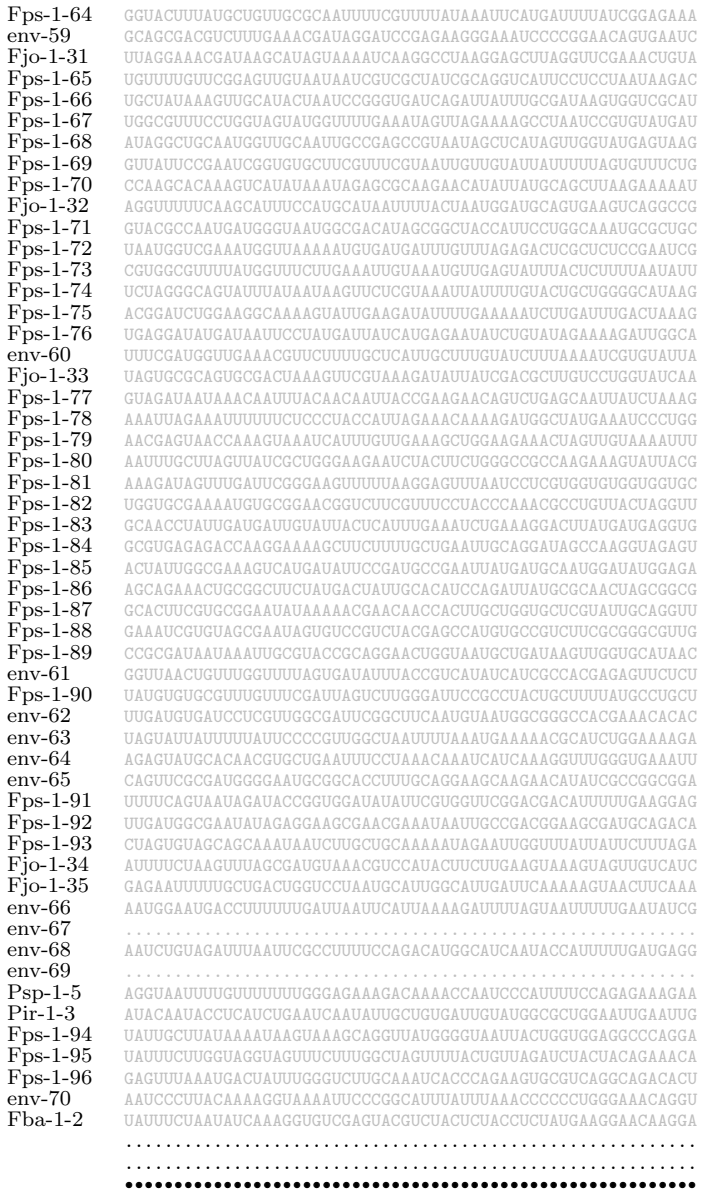


alignment positions 537 · · · 594


env-1
Ddo-1-1
Pto-1-1
Cat-1-1
Fjo-1-1
Pto-1-2
Fps-1-1
Fjo-1-2
Fjo-1-3
Fps-1-2
Fps-1-3
Fps-1-4
Bfr-1-1
Gfo-1-1
Bth-1-1
Cat-1-2
Fjo-1-4
Fps-1-5
Gfo-1-2
Psp-1-1
Fba-2-1
Pto-1-3
Fba-2-2
Fjo-1-5
Tde-1-1
Fjo-1-6
Pto-1-4
env-2
Fps-1-6
env-3
env-4
Chu-1-1
env-5
Cat-1-3
Fps-1-7
Gfo-1-3
env-6
Fps-1-8
env-7
Fps-1-9
Fps-1-10
Fps-1-11
Fba-1-1
Fjo-1-7
env-8
env-9
env-10
env-11
env-12
env-13
Fba-2-3
Pir-1-1
Fps-1-12
Fjo-1-8
env-14
Fba-2-4
Fba-2-5
Fjo-1-9
Fps-1-13
env-15
env-16
Fps-1-14


Fps-1-15
Psp-1-2
Pto-1-5
Pir-1-2
Pto-1-6
Fps-1-16
env-17
Fba-2-6
Fjo-1-10
Fba-2-7
env-18
Fjo-1-11
env-19
env-20
Psp-1-3
Fps-1-17
Fps-1-18
env-21
Fps-1-19
env-22
env-23
env-24
env-25
Fps-1-20
env-26
Fps-1-21
ueu-1-1
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
Fps-1-22
env-35
Fps-1-23
Fps-1-24
env-36
env-37
Fps-1-25
Fps-1-26
Fps-1-27
Fps-1-28
Fps-1-29
Fps-1-30
Fps-1-31
Fps-1-32
env-38
Fps-1-33
Fps-1-34
Fjo-1-12
Fps-1-35
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46


Psp-1-4
Gfo-1-4
Fba-2-8
env-47
Fps-1-36
env-48
Fps-1-37
env-49
Fjo-1-13
Fjo-1-14
Fjo-1-15
Fjo-1-16
Fjo-1-17
Fjo-1-18
env-50
Fps-1-38
Fjo-1-19
env-51
Lbl-1-1
Fjo-1-20
Fjo-1-21
Fjo-1-22
Fjo-1-23
env-52
Fps-1-39
env-53
Fjo-1-24
env-54
env-55
Fps-1-40
env-56
Fjo-1-25
Fjo-1-26
Fjo-1-27
Fps-1-41
Fps-1-42
env-57
Fps-1-43
env-58
Fps-1-44
Fps-1-45
Fps-1-46
Fps-1-47
Fps-1-48
Fps-1-49
Fps-1-50
Fps-1-51
Fps-1-52
Fps-1-53
Fps-1-54
Fps-1-55
Fps-1-56
Fps-1-57
Fjo-1-28
Fps-1-58
Fps-1-59
Fjo-1-29
Fjo-1-30
Fps-1-60
Fps-1-61
Fps-1-62
Fps-1-63


Fps-1-64 GGUACUUUAUGCUGUUGCGCAAUUUUCGUUUUAUAAAUUCAUGAUUUUAUCGGAGAAA
env-59 GCAGCGACGUCUUUGAAACGAUAGGAUCCGAGAAGGGAAAUCCCCGGAACAGUGAAUC
Fjo-1-31 UUAGGAAACGAUAAGCAUAGUAAAAUCAAGGCCUAAGGAGCUUAGGUUCGAAACUGUA
Fps-1-65 UGUUUUGUUCGGAGUUGUAAUAAUCGUCGCUAUCGCAGGUCAUUCCUCCUAAUAAGAC
Fps-1-66 UGCUAUAAAGUUGCAUACUAAUCCGGGUGAUCAGAUUAUUUGCGAUAAGUGGUCGCAU
Fps-1-67 UGGCGUUUCCUGGUAGUAUGGUUUUGAAAUAGUUAGAAAAGCCUAAUCCGUGUAUGAU
Fps-1-68 AUAGGCUGCAAUGGUUGCAAUUGCCGAGCCGUAAUAGCUCAUAGUUGGUAUGAGUAAG
Fps-1-69 GUUAUUCCGAAUCGGUGUGCUUCGUUUCGUAAUUGUUGUAUUAUUUUUAGUGUUUCUG
Fps-1-70 CCAAGCACAAAGUCAUAUAAAUAGAGCGCAAGAACAUAUUAUGCAGCUUAAGAAAAAU
Fjo-1-32 AGGUUUUUCAAGCAUUUCCAUGCAUAAUUUUACUAAUGGAUGCAGUGAAGUCAGGCCG
Fps-1-71 GUACGCCAAUGAUGGGUAAUGGCGACAUAGCGGCUACCAUUCCUGGCAAAUGCGCUGC
Fps-1-72 UAAUGGUCGAAAUGGUUAAAAAUGUGAUGAUUUGUUUAGAGACUCGCUCUCCGAAUCG
Fps-1-73 CGUGGCGUUUUAUGGUUUCUUGAAAUUGUAAAUGUUGAGUAUUUACUCUUUUAAUAUU
Fps-1-74 UCUAGGGCAGUAUUUAUAAUAAGUUCUCGUAAAUUAUUUUGUACUGCUGGGGCAUAAG
Fps-1-75 ACGGAUCUGGAAGGCAAAAGUAUUGAAGAUAUUUUGAAAAAUCUUGAUUUGACUAAAG
Fps-1-76 UGAGGAUAUGAUAAUUCCUAUGAUUAUCAUGAGAAUAUCUGUAUAGAAAAGAUUGGCA
env-60 UUUCGAUGGUUGAAACGUUCUUUUGCUCAUUGCUUUGUAUCUUUAAAAUCGUGUAUUA
Fjo-1-33 UAGUGCGCAGUGCGACUAAAGUUCGUAAAGAUAUUAUCGACGCUUGUCCUGGUAUCAA
Fps-1-77 GUAGAUAAUAAACAAUUUACAACAAUUACCGAAGAACAGUCUGAGCAAUUAUCUAAAG
Fps-1-78 AAAUUAGAAAUUUUUUCUCCCUACCAUUAGAAACAAAAGAUGGCUAUGAAAUCCCUGG
Fps-1-79 AACGAGUAACCAAAGUAAAUCAUUUGUUGAAAGCUGGAAGAAACUAGUUGUAAAAUUU
Fps-1-80 AAUUUGCUUAGUUAUCGCUGGGAAGAAUCUACUUCUGGGCCGCCAAGAAAGUAUUACG
Fps-1-81 AAAGAUAGUUUGAUUCGGGAAGUUUUUAAGGAGUUUAAUCCUCGUGGUGUGGUGGUGC
Fps-1-82 UGGUGCGAAAAUGUGCGGAACGGUCUUCGUUUCCUACCCAAACGCCUGUUACUAGGUU
Fps-1-83 GCAACCUAUUGAUGAUUGUAUUACUCAUUUGAAAUCUGAAAGGACUUAUGAUGAGGUG
Fps-1-84 GCGUGAGAGACCAAGGAAAAGCUUCUUUUGCUGAAUUGCAGGAUAGCCAAGGUAGAGU
Fps-1-85 ACUAUUGGCGAAAGUCAUGAUAUUCCGAUGCCGAAUUAUGAUGCAAUGGAUAUGGAGA
Fps-1-86 AGCAGAAACUGCGGCUUCUAUGACUAUUGCACAUCCAGAUUAUGCGCAACUAGCGGCG
Fps-1-87 GCACUUCGUGCGGAAUAUAAAAACGAACAACCACUUGCUGGUGCUCGUAUUGCAGGUU
Fps-1-88 GAAAUCGUGUAGCGAAUAGUGUCCGUCUACGAGCCAUGUGCCGUCUUCGCGGGCGUUG
Fps-1-89 CCGCGAUAAUAAAUUGCGUACCGCAGGAACUGGUAAUGCUGAUAAGUUGGUGCAUAAC
env-61 GGUUAACUGUUUGGUUUUAGUGAUAUUUACCGUCAUAUCAUCGCCACGAGAGUUCUCU
Fps-1-90 UAUGUGUGCGUUUGUUUCGAUUAGUCUUGGGAUUCCGCCUACUGCUUUUAUGCCUGCU
env-62 UUGAUGUGAUCCUCGUUGGCGAUUCGGCUUCAAUGUAAUGGCGGGCCACGAAACACAC
env-63 UAGUAUUAUUUUUAUUCCCCGUUGGCUAAUUUUAAAUGAAAAACGCAUCUGGAAAAGA
env-64 AGAGUAUGCACAACGUGCUGAAUUUCCUAAACAAAUCAUCAAAGGUUUGGGUGAAAUU
env-65 CAGUUCGCGAUGGGGAAUGCGGCACCUUUGCAGGAAGCAAGAACAUAUCGCCGGCGGA
Fps-1-91 UUUUCAGUAAUAGAUACCGGUGGAUAUAUUCGUGGUUCGGACGACAUUUUUGAAGGAG
Fps-1-92 UUGAUGGCGAAUAUAGAGGAAGCGAACGAAAUAAUUGCCGACGGAAGCGAUGCAGACA
Fps-1-93 CUAGUGUAGCAGCAAAUAAUCUUGCUGCAAAAAUAGAAUUGGUUUAUUAUUCUUUAGA
Fjo-1-34 AUUUUCUAAGUUUAGCGAUGUAAACGUCCAUACUUCUUGAAGUAAAGUAGUUGUCAUC
Fjo-1-35 GAGAAUUUUUGCUGACUGGUCCUAAUGCAUUGGCAUUGAUUCAAAAAGUAACUUCAAA
env-66 AAUGGAAUGACCUUUUUUGAUUAAUUCAUUAAAAGAUUUUAGUAAUUUUUGAAUAUCG
env-67 ....
env-68 AAUCUGUAGAUUUAAUUCGCCUUUUCCAGACAUGGCAUCAAUACCAUUUUUGAUGAGG
env-69 ....
Psp-1-5 AGGUAAUUUUGUUUUUUUGGGAGAAAGACAAAACCAAUCCCAUUUUCCAGAGAAAGAA
Pir-1-3 AUACAAUACCUCAUCUGAAUCAAUAUUGCUGUGAUUGUAUGGCGCUGGAAUUGAAUUG
Fps-1-94 UAUUGCUUAUAAAAUAAGUAAAGCAGGUUAUGGGGUAAUUACUGGUGGAGGCCCAGGA
Fps-1-95 UAUUUCUUGGUAGGUAGUUUCUUUGGCUAGUUUUACUGUUAGAUCUACUACAGAAACA
Fps-1-96 GAGUUUAAAUGACUAUUUGGGUCUUGCAAAUCACCCAGAAGUGCGUCAGGCAGACACU
env-70 AAUCCCUUACAAAAGGUAAAAUUCCCGGCAUUUAUUUAAACCCCCCUGGGAAACAGGU
Fba-1-2 UAUUUCUAAUAUCAAAGGUGUCGAGUACGUCUACUCUACCUCUAUGAAGGAACAAGGA
.....
.....
••••••

The taxonomy of each organism containing a putative *flg*-Rhizobiales RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *flg*-Rhizobiales RNA in Sections 33.2 and 33.4.

abbrev. of hits	taxonomy of species
Bja-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Bradyrhizobium japonicum* USDA 110
Bsp-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Bradyrhizobium sp.* BTAi1
Bsp-2-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Bradyrhizobium sp.* ORS278
Nha-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Nitrobacter hamburgensis* X14
Nsp-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Nitrobacter sp.* Nb-311A
Nwi-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Nitrobacter winogradskyi* Nb-255
Oca-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Oligotropha carboxidovorans* OM5
Rpa-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* BisA53
Rpa-2-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* BisB18
Rpa-3-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* BisB5
Rpa-4-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* CGA009
Rpa-5-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* HaA2
Rpa-6-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* TIE-1
env-1	environmental sample

33.2 Gene contexts

Each *flg*-Rhizobiales RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *flg*-Rhizobiales RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 33.3. The accession number of the sequence containing each *flg*-Rhizobiales RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 33.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	soil_Contig28079	-	649	507	**RNA**→ FlgG (COG4786)→
Bja-1-1	NC_004463.1	+	6396344	6396490	**RNA**→ FlgG (COG4786)→ FlgG (COG4786)→ FlgA (pfam03240)NOG07957 (NOG07957)→ FlgH (COG2063)→
Oca-1-1	NC_011386.1	-	1375948	1375814	**RNA**→ FlgG (COG4786)→ FlgG (COG4786)→ FlgA (COG1261)→ FlgH (COG2063)→
Rpa-2-1	NC_007925.1	-	1644277	1644136	**RNA**→ FlgG (COG4786)→ FlgG (COG4786)→ FlgA (COG1261)NOG07957 (NOG07957)→ FlgH (COG2063)→
Rpa-1-1	NC_008435.1	-	1727246	1727103	**RNA**→ FlgG (COG4786)→ FlgG (COG4786)→ FlgA (COG1261)→ FlgH (COG2063)→
Bsp-1-1	NC_009485.1	+	5726916	5727066	**RNA**→ FlgG (COG4786)→ FlgG (COG4786)→ FlgA (COG1261)→ FlgH (COG2063)→
Bsp-2-1	NC_009445.1	+	5227487	5227637	**RNA**→ FlgG (COG4786)→ FlgG (COG4786)→ FlgA (COG1261)→ FlgH (COG2063)→
Nha-1-1	NC_007964.1	-	1569053	1568914	**RNA**→ FlgG (COG4786)→ FlgG (COG4786)→
Nwi-1-1	NC_007406.1	-	1232441	1232299	**RNA**→ FlgG (COG4786)→ FlgG (COG4786)→ FlgA (COG1261)NOG07957 (NOG07957)→ FlgH (COG2063)→
Nsp-1-1	NZ_AAMY01000006.1	-	139562	139418	**RNA**→ FlgG (COG4786)→ FlgG (COG4786)→ FlgA (pfam03240)→ FlgH (COG2063)→
Rpa-4-1	NC_005296.1	+	4400699	4400837	**RNA**→ FlgG (COG4786)→ FlgG (COG4786)→ FlgA (COG1261)→ FlgH (COG2063)→
Rpa-6-1	NC_011004.1	+	4748751	4748889	**RNA**→ FlgG (COG4786)→ FlgG (COG4786)→ FlgA (COG1261)→ FlgH (COG2063)→
Rpa-5-1	NC_007778.1	+	4321689	4321823	**RNA**→ FlgG (COG4786)→ FlgG (COG4786)→ FlgA (COG1261)→ FlgH (COG2063)→
Rpa-3-1	NC_007958.1	-	1900071	1899939	**RNA**→ FlgG (COG4786)→ FlgG (COG4786)→ FlgA (COG1261)→ FlgH (COG2063)→

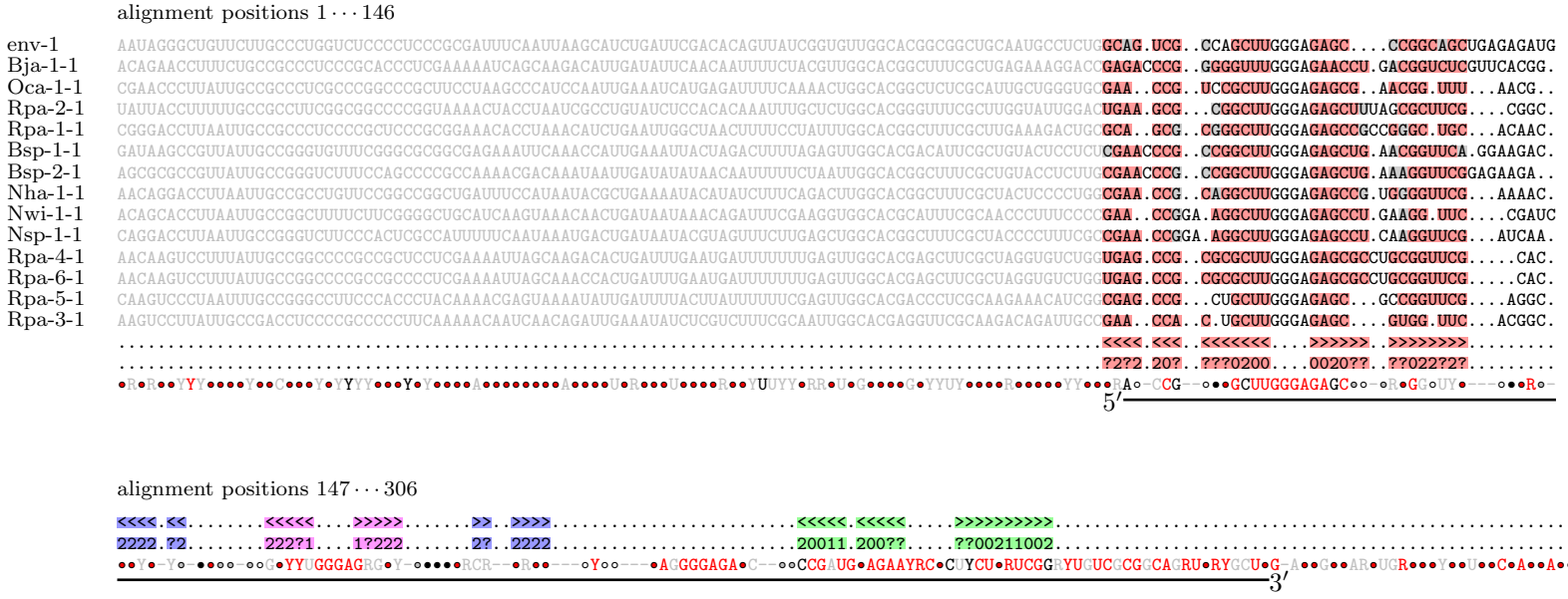
33.3 Conserved domains

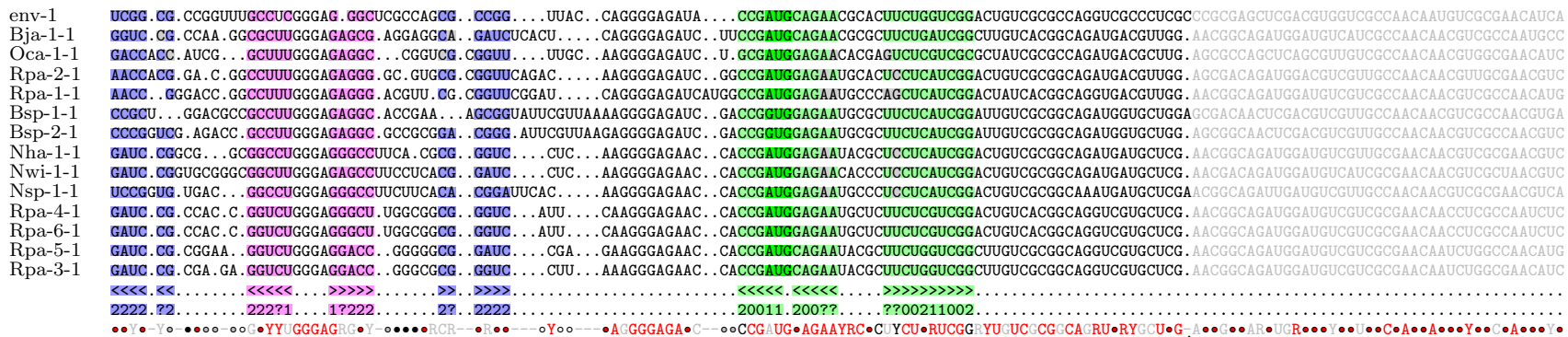
Conserved domains found in protein-coding genes listed in Section 33.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *flg*-Rhizobiales RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

COG1261 (10) Flagellar basal body P-ring biosynthesis protein [Cell motility and secretion / Posttranslational modification, protein turnover, chaperones]
COG2063 (12) Flagellar basal body L-ring protein [Cell motility and secretion]
COG4786 (27) Flagellar basal body rod protein [Cell motility and secretion]
NOG07957 (3)
pfam03240 (2) FlgA family.

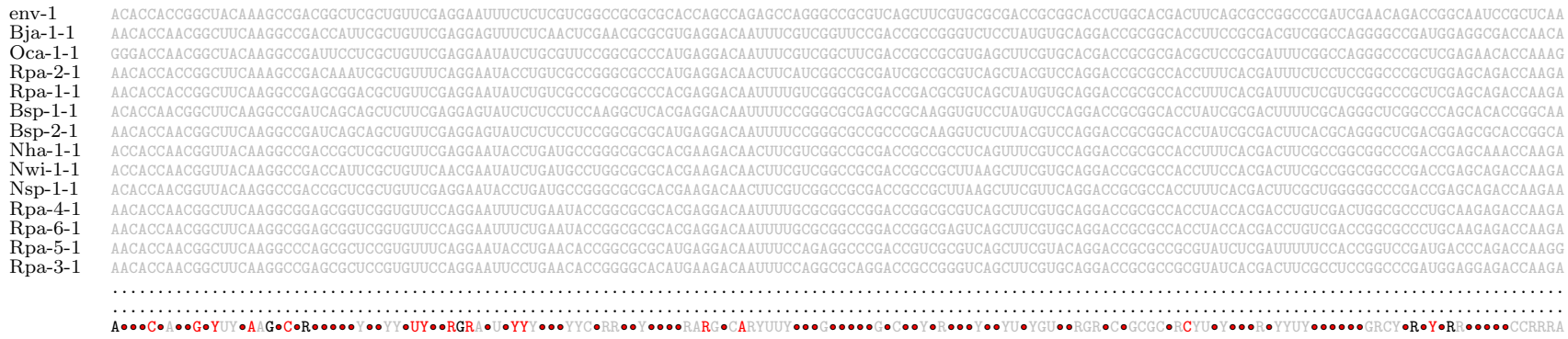
33.4 Multiple-sequence alignment

Each *flg*-Rhizobiales RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 33.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *flg*-Rhizobiales RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

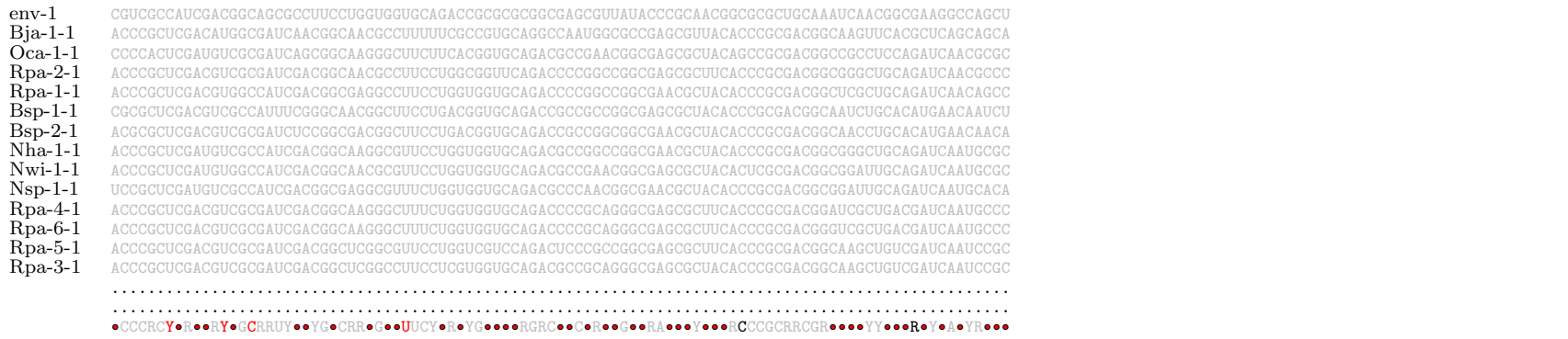




alignment positions 307 · · · 466



alignment positions 467 · · · 565



34 *flpD*

34.1 Taxa

The taxonomy of each organism containing a putative *flpD* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *flpD* RNA in Sections 34.2 and 34.4.

abbrev. of hits	taxonomy of species
Mja-1-1	Archaea Euryarchaeota Methanococci Methanococcales Methanocaldococcaceae *Methanocaldococcus jannaschii* DSM 2661
Mae-1-1	Archaea Euryarchaeota Methanococci Methanococcales Methanococcaceae *Methanococcus aeolicus* Nankai-3
Mma-1-1	Archaea Euryarchaeota Methanococci Methanococcales Methanococcaceae *Methanococcus maripaludis* C5
Mma-2-1	Archaea Euryarchaeota Methanococci Methanococcales Methanococcaceae *Methanococcus maripaludis* C6
Mma-3-1	Archaea Euryarchaeota Methanococci Methanococcales Methanococcaceae *Methanococcus maripaludis* C7
Mma-4-1	Archaea Euryarchaeota Methanococci Methanococcales Methanococcaceae *Methanococcus maripaludis* S2
Mva-1-1	Archaea Euryarchaeota Methanococci Methanococcales Methanococcaceae *Methanococcus vannielii* SB
Mka-1-1	Archaea Euryarchaeota Methanopyri Methanopyrales Methanopyraceae *Methanopyrus kandleri* AV19

34.2 Gene contexts

Each *flpD* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *flpD* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 34.3. The accession number of the sequence containing each *flpD* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 34.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Mka-1-1	NC_003551.1	-	172713	172667	**RNA**→ ←hypo
Mva-1-1	NC_009634.1	-	1013956	1013910	**RNA**→ FlpD (pfam02662)→
Mma-4-1	NC_005791.1	-	1636889	1636843	**RNA**→ FlpD (pfam02662)→ FrhG (COG1941)→ FrhA (COG3259)→ F420 non-reducing hydrogenase subunit→
Mma-1-1	NC_009135.1	+	1662793	1662839	**RNA**→ FlpD (pfam02662)→ FrhG (COG1941)→ FrhA (COG3259)→ F420 non-reducing hydrogenase subunit→ HybA (COG0437)HycB (COG1142)NuoI (COG1143)→ FwdB (COG1029)→
Mma-2-1	NC_009975.1	+	925551	925597	**RNA**→ FlpD (pfam02662)→ FrhG (COG1941)→ FrhA (COG3259)→ F420 non-reducing hydrogenase subunit→ HycB (COG1142)COG1146 (COG1146)→ FwdB (COG1029)→
Mma-3-1	NC_009637.1	-	931385	931339	**RNA**→ FlpD (pfam02662)→ FrhG (COG1941)→ FrhA (COG3259)→ F420 non-reducing hydrogenase subunit→ HycB (COG1142)→ FwdB (COG1029)→
Mja-1-1	NC_000909.1	+	1131452	1131497	**RNA**→ FlpD (pfam02662)→
Mae-1-1	NC_009635.1	+	1282979	1283025	**RNA**→ FlpD (pfam02662)→

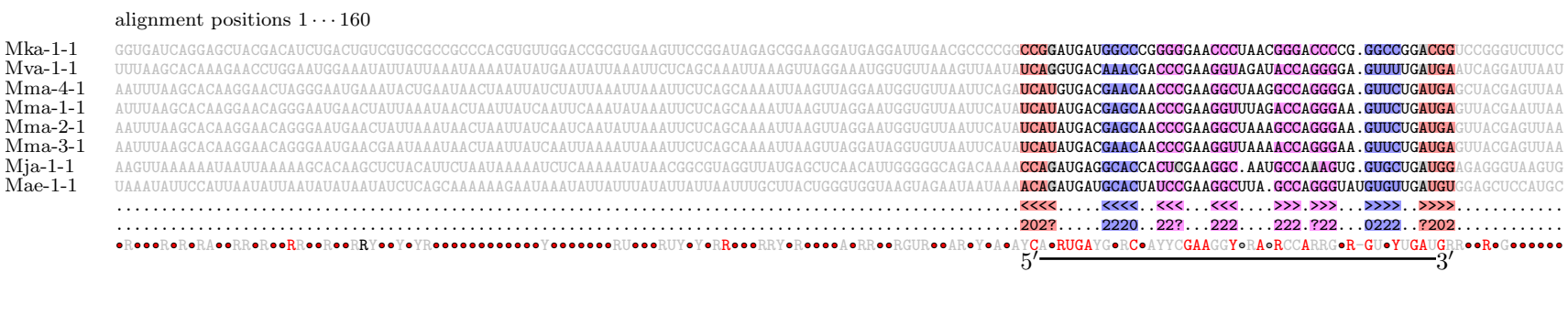
34.3 Conserved domains

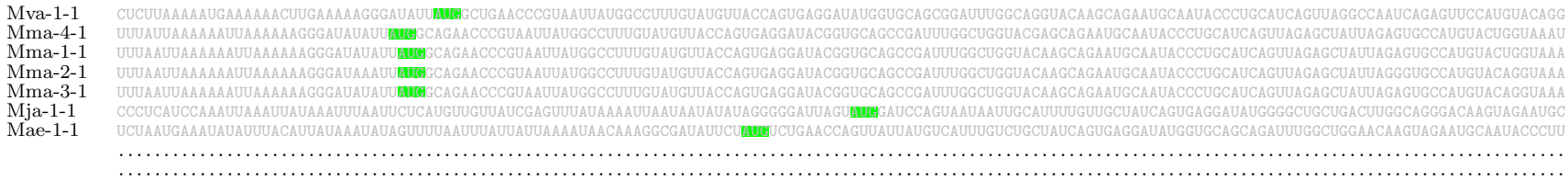
Conserved domains found in protein-coding genes listed in Section 34.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *flpD* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

COG0437 (1) Fe-S-cluster-containing hydrogenase components 1 [Energy production and conversion]
COG1029 (3) Formylmethanofuran dehydrogenase subunit B [Energy production and conversion]
COG1142 (3) Fe-S-cluster-containing hydrogenase components 2 [Energy production and conversion]
COG1143 (1) Formate hydrogenlyase subunit 6/NADH:ubiquinone oxidoreductase 23 kD subunit (chain I) [Energy production and conversion]
COG1146 (1) Ferredoxin [Energy production and conversion]
COG1941 (4) Coenzyme F420-reducing hydrogenase, gamma subunit [Energy production and conversion]
COG3259 (4) Coenzyme F420-reducing hydrogenase, alpha subunit [Energy production and conversion]
pfam02662 (7) Methyl-viologen-reducing hydrogenase, delta subunit.

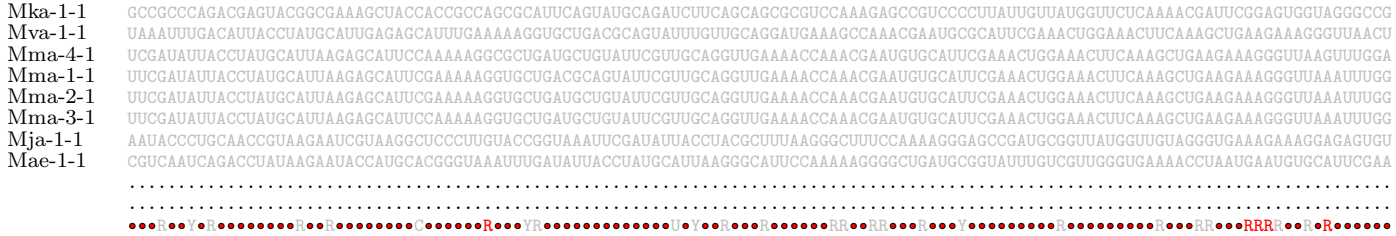
34.4 Multiple-sequence alignment

Each *flpD* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 34.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *flpD* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.





alignment positions 321 · · · 448



35 *gabT*

35.1 Taxa

The taxonomy of each organism containing a putative *gabT* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *gabT* RNA in Sections 35.2 and 35.4.

abbrev. of hits	**taxonomy of species**
Pae-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* 2192
Pae-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* C3719
Pae-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PA7
Pae-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PACS2
Pae-5-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PAO1
Pae-6-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* UCBPP-PA14
Pen-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas entomophila* L48
Pfl-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf-5
Pfl-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf0-1
Pme-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas mendocina ymp*
Ppu-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* F1
Ppu-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* GB-1
Ppu-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* KT2440
Ppu-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* W619
Pst-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas stutzeri* A1501
Psy-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. phaseolicola* 1448A
Psy-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000
env-1 to env-3	environmental samples

35.2 Gene contexts

Each *gabT* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *gabT* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 35.3. The accession number of the sequence containing each *gabT* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 35.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Pme-1-1	NC_009439.1	+	239262	239322	**RNA**→ GabT (COG0160)→
env-1	JCVI_SCAF_1096627187522	-	1330	1269	**RNA**→
Pst-1-1	NC_009434.1	+	847701	847762	**RNA**→ GabT (COG0160)→
Pae-3-1	NC_009656.1	+	354519	354583	**RNA**→ ←hypo
Pae-6-1	NC_008463.1	+	315247	315311	**RNA**→ GabT (COG0160)→
Pae-5-1	NC_002516.2	+	301106	301170	**RNA**→ GabT (COG0160)→
Pae-4-1	NZ_AAQW01000001.1	+	289333	289397	**RNA**→ GabT (COG0160)→
Pae-2-1	NZ_AAKV01000079.1	-	88292	88228	**RNA**→ GabT (COG0160)→
Pae-1-1	NZ_AAKW01000054.1	+	138158	138222	**RNA**→ GabT (COG0160)→
Pfl-2-1	NC_007492.1	+	226838	226898	**RNA**→ GabT (COG0160)→
Psy-1-1	NC_005773.3	-	106095	106036	**RNA**→ GabT (COG0160)→
Psy-2-1	NC_007005.1	-	99873	99814	**RNA**→ GabT (COG0160)→
Psy-3-1	NC_004578.1	+	327286	327345	**RNA**→ GabT (COG0160)→ hypo→ REC (cd00156)COG1639 (COG1639)→
env-2	JCVI_SCAF_1101667132619	-	485	424	**RNA**→ ? (GOS80)→
env-3	JCVI_SCAF_1096627200952	-	477	416	**RNA**→ ? (GOS80)→
Pfl-1-1	NC_004129.6	+	220742	220801	**RNA**→ GabT (COG0160)→
Ppu-2-1	NZ_AAXR01000011.1	-	173376	173316	**RNA**→ ←hypo (cons)
Ppu-4-1	NZ_AAVY01000021.1	-	28238	28177	**RNA**→ GabT (COG0160)→
Ppu-3-1	NC_002947.3	+	264770	264831	**RNA**→ GabT (COG0160)→
Ppu-1-1	NC_009512.1	+	267930	267991	**RNA**→ GabT (COG0160)→
Pen-1-1	NC_008027.1	+	199162	199223	**RNA**→ GabT (COG0160)→

35.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 35.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *gabT* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

cd00156 (1) Signal receiver domain; originally thought to be unique to bacteria (CheY, OmpR, NtrC, and PhoB), now recently identified in eukaryotes ETR1 Arabidopsis thaliana; this domain receives the signal from the sensor partner in a two-component systems; contains a phosphoacceptor site that is phosphorylated by histidine kinase homologs; usually found N-terminal to a DNA binding effector domain; forms homodimers

COG0160 (16) 4-aminobutyrate aminotransferase and related aminotransferases [Amino acid transport and metabolism]

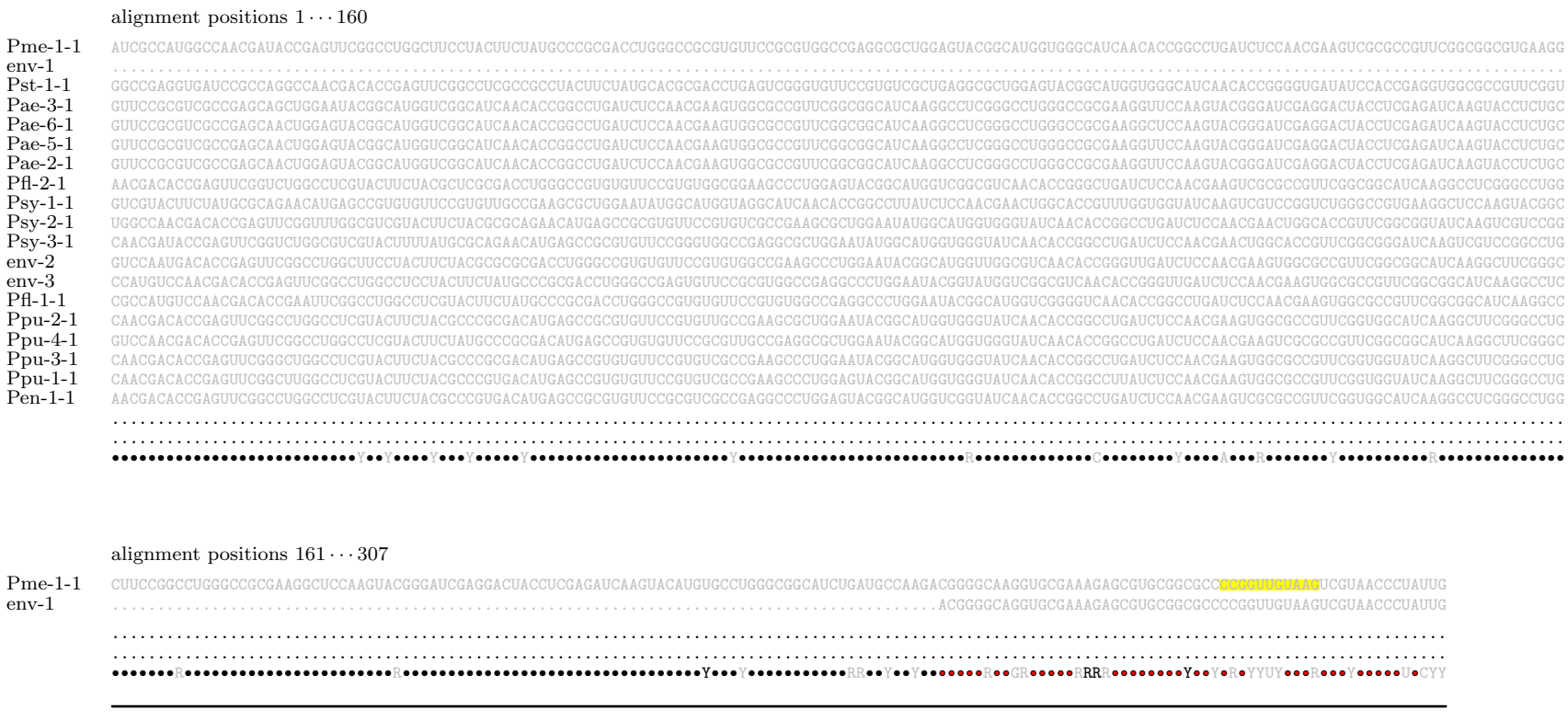
COG1639 (1) Predicted signal transduction protein [Signal transduction mechanisms]

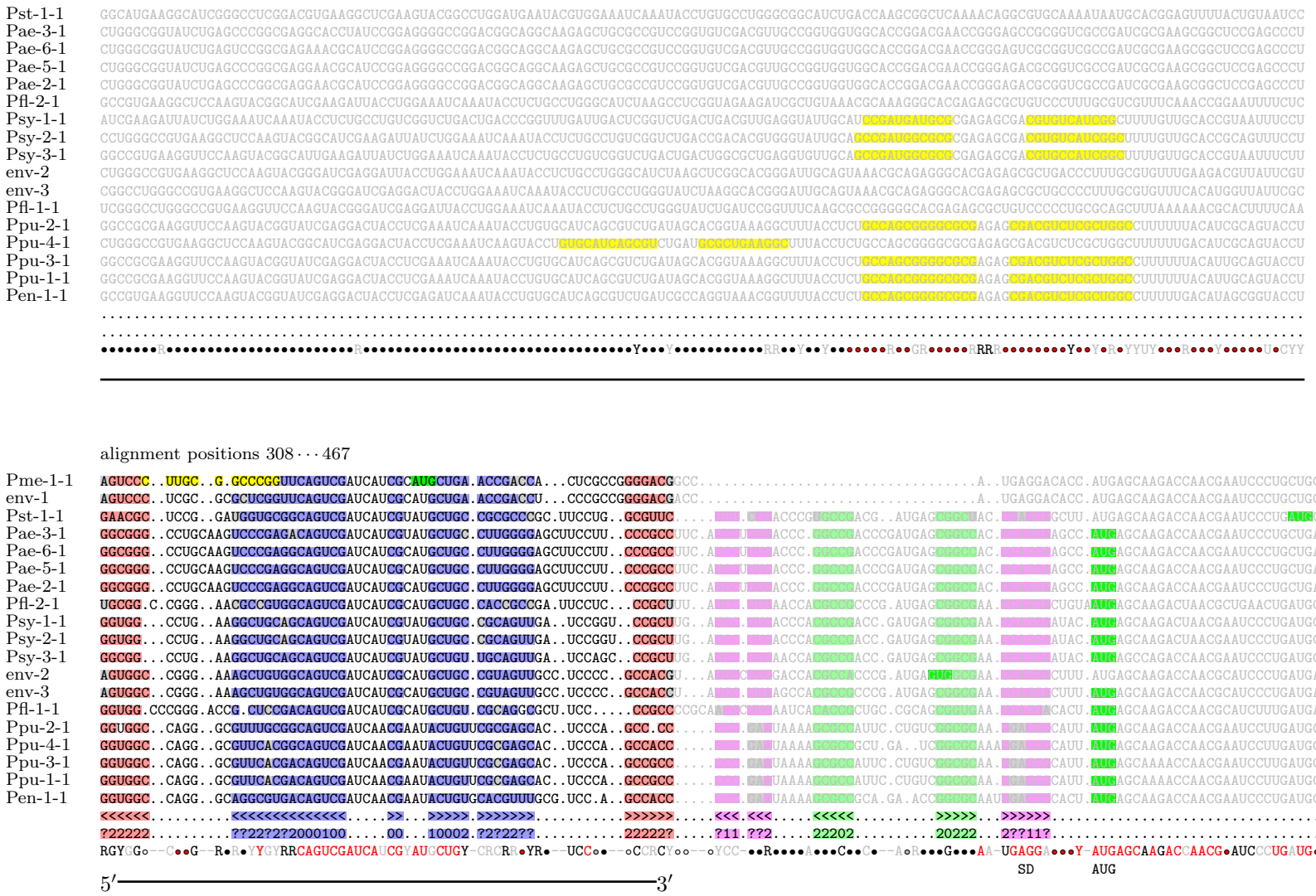
GOS80 (2) ?

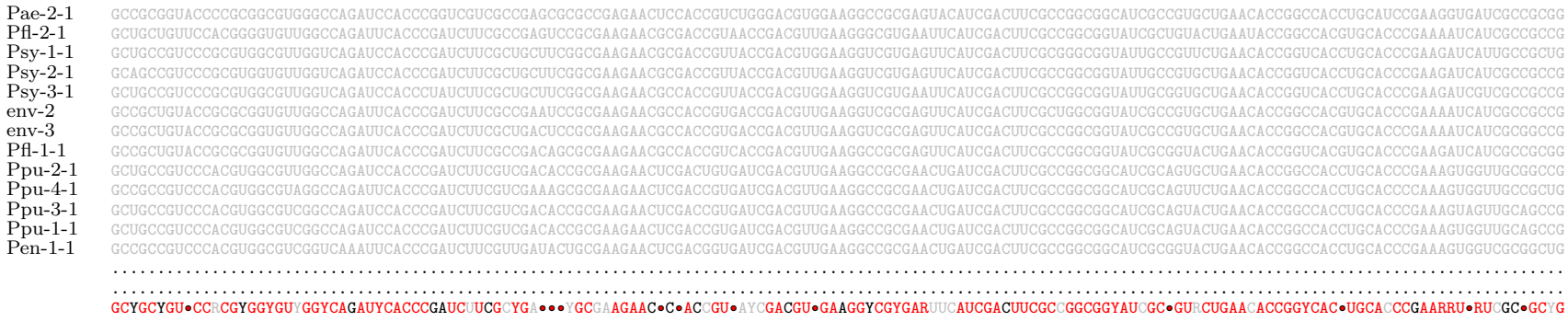
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Pae-1-1, Pae-4-1

35.4 Multiple-sequence alignment

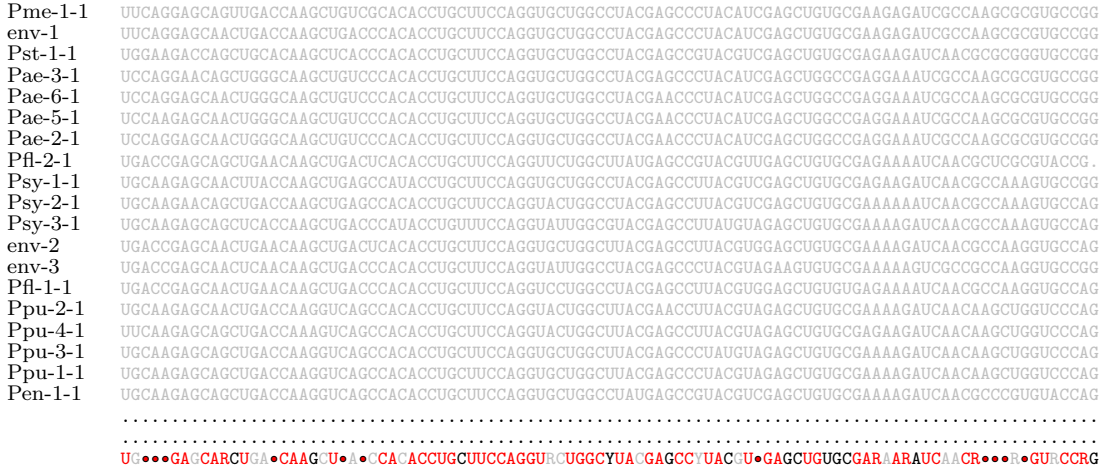
Each *gabT* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 35.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *gabT* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.




Pst-1-1
Pae-3-1
Pae-6-1
Pae-5-1
Pae-2-1
Pfl-2-1
Psy-1-1
Psy-2-1
Psy-3-1
env-2
env-3
Pfl-1-1
Ppu-2-1
Ppu-4-1
Ppu-3-1
Ppu-1-1
Pen-1-1
alignment positions 308 · · · 467
Pme-1-1
env-1
Pst-1-1
Pae-3-1
Pae-6-1
Pae-5-1
Pae-2-1
Pfl-2-1
Psy-1-1
Psy-2-1
Psy-3-1
env-2
env-3
Pfl-1-1
Ppu-2-1
Ppu-4-1
Ppu-3-1
Ppu-1-1
Pen-1-1
SD
AUG
5′
3′
alignment positions 468 · · · 627
Pme-1-1
env-1
Pst-1-1
Pae-3-1
Pae-6-1
Pae-5-1



alignment positions 628 · · · 726



36 Gamma-*cis*-1

36.1 Taxa

The taxonomy of each organism containing a putative Gamma-*cis*-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Gamma-*cis*-1 RNA in Sections 36.3 and 36.5.

abbrev. of hits	**taxonomy of species**
Sde-1-1	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Alteromonadaceae *Saccharophagus degradans* 2-40
Ptu-1-1 to Ptu-1-4	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Pseudoalteromonadaceae *Pseudoalteromonas tunicata* D2
Sam-1-1 to Sam-1-2	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella amazonensis* SB2B
Sba-1-1 to Sba-1-2	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella baltica* OS195
Slo-1-1 to Slo-1-3	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella loihica* PV-4

Spu-1-1 to Spu-1-2	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella putrefaciens* CN-32
Ssp-1-1	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella sp.* ANA-3
Ssp-2-1 to Ssp-2-3	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella sp.* MR-7
Ssp-3-1 to Ssp-3-2	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella sp.* W3-18-1
Cli-1-1	Bacteria Proteobacteria Gammaproteobacteria *Congregibacter litoralis* KT71
Hch-1-1 to Hch-1-2	Bacteria Proteobacteria Gammaproteobacteria Oceanospirillales Hahellaceae *Hahella chejuensis* KCTC 2396
Vch-1-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Vibrio cholerae* B33
env-1 to env-26	environmental samples

36.2 Notes

Superscript numbers are used to annotate any Gamma-*cis*-1 RNA that has special characteristics (described below). These numeric annotations will be used in Sections 36.3 and 36.5. The meaning of the numbers are as follows.

[1] Homology is unclear.

[2] Accepted in local miniscan with E-value $< 10^{-7}$. However, the terminal loops of P2 and P3 are somewhat unclear, particularly the lack of poly-A in the P2 terminal loop.

[3] Has lower G+C than other hits, but not unreasonable. Gene is annotated as probably taurine catabolism dioxygenase, which is related to the glyoxalase (whose conserved domain is more generally a dioxygenase)

36.3 Gene contexts

Each Gamma-*cis*-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Gamma-*cis*-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 36.4. The accession number of the sequence containing each Gamma-*cis*-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 36.1. Superscript numbers refer to annotations listed in Section 36.2

abbrev.	Seq. accession		5′ at	3′ at	genes
Ptu-1-1	NZ_AAOH01000004.1	-	360181	360092	**RNA**→ MhpC (COG0596)→
Ptu-1-2	NZ_AAOH01000004.1	-	495474	495385	**RNA**→ hypo→
Ssp-2-1	NC_008322.1	+	4601816	4601898	**RNA**→ hypo→
Sde-1-1	NC_007912.1	-	2957581	2957495	**RNA**→ hypo→ hypo→
Sba-1-1	NZ_AATK01000025.1	+	23316	23402	**RNA**→ ←hypo (cons)
env-1	JCVI_SCAF_1096627220188	-	1200	1100	**RNA**→ ? (GOS12001879)→
Ssp-1-1	NC_008577.1	+	2692968	2693066	**RNA**→ Acetyltransf_1 (pfam00583)→ hypo→
Hch-1-1	NC_007645.1	-	4662996	4662899	**RNA**→ hypo→
[3]env-2	ABEF01049329.1	-	2351	2254	**RNA**→ TauD (COG2175)→
env-3	wf2_AHAA14201_x1	-	256	159	**RNA**→
env-4	JCVI_SCAF_1096627200416	-	2705	2612	**RNA**→ ? (GOS14514810)→
env-5	JCVI_SCAF_1096628116680	+	1162	1244	**RNA**→
[2]env-6	ABEF01050261.1	-	1023	935	**RNA**→ hypo→
env-7	JCVI_SCAF_1101667304983	+	190	275	**RNA**→ Acetyltransf_1 (pfam00583)? (GOS8607)→
Ssp-2-2	NC_008322.1	-	4631235	4631153	**RNA**→ INT_SG3_C (cd01186)→

Hch-1-2	NC_007645.1	+	1352816	1352907	**RNA**→ ←tRNA-Phe
env-8	wf2_AHAA1455_x1	+	550	639	**RNA**→
env-9	ABEF01000658.1	+	366	455	**RNA**→ Glyoxalase (pfam00903)→
env-10	ABEF01050261.1	-	1871	1780	**RNA**→ hypo→
env-11	ABEF01039699.1	-	420	332	**RNA**→
env-12	ABEF01047328.1	+	54	142	**RNA**→ hypo→
env-13	ABEF01047328.1	+	1124	1212	**RNA**→ hypo→
env-14	JCVI_SCAF_1096627388979	-	18944	18865	**RNA**→ ←LdhA (COG1052)? (GOS1248)
Cli-1-1	NZ_AAOA01000002.1	+	49053	49135	**RNA**→ hypo→ Acetyltransf_1 (pfam00583)→
Spu-1-1	NC_009438.1	-	2521407	2521311	**RNA**→ hypo→ hypo→
Ssp-3-1	NC_008750.1	+	2096938	2097034	**RNA**→ hypo→ hypo→
Ssp-2-3	NC_008322.1	-	4031687	4031590	**RNA**→ hypo→
env-15	JCVI_SCAF_1101668032441	-	1032	935	**RNA**→ ? (GOS11650400)→
Sam-1-1	NC_008700.1	-	1637918	1637821	**RNA**→ hypo→ hypo→ hypo→ hypo→
env-16	JCVI_SCAF_1096627187232	+	3271	3370	**RNA**→ Thiolase_N (pfam00108)? (GOS3111)→
Sba-1-2	NZ_AATK01000060.1	-	19214	19122	**RNA**→ COG3236 (COG3236)→
env-17	JCVI_SCAF_1096627197474	+	9165	9262	**RNA**→ ? (GOS11643631)→
env-18	JCVI_SCAF_1096627187232	+	3732	3829	**RNA**→ Thiolase_N (pfam00108)? (GOS3111)→
env-19	JCVI_SCAF_1096627187685	+	5448	5545	**RNA**→
Vch-1-1	NZ_AAWE01000032.1	+	23779	23876	**RNA**→ HTH_XRE (smart00530)→
env-20	JCVI_SCAF_1096627190631	+	2911	3008	**RNA**→
Ptu-1-3	NZ_AAOH01000001.1	-	366609	366512	**RNA**→ hypo→
Slo-1-1	NC_009092.1	+	1516016	1516108	**RNA**→ hypo→
Ptu-1-4	NZ_AAOH01000004.1	-	365677	365580	**RNA**→ hypo→ hypo→
env-21	JCVI_SCAF_1101667112935	-	586	494	**RNA**→ COG3666 (COG3666)? (GOS1730)→
Slo-1-2	NC_009092.1	-	4272239	4272151	**RNA**→ hypo→
Sam-1-2	NC_008700.1	+	3384963	3385032	**RNA**→ hypo→
Slo-1-3	NC_009092.1	+	1949101	1949186	**RNA**→ COG5006 (COG5006)→
Spu-1-2	NC_009438.1	+	1853321	1853425	**RNA**→ hypo→
Ssp-3-2	NC_008750.1	+	1133692	1133796	**RNA**→ hypo→
env-22	JCVI_SCAF_1101668592552	+	635	723	**RNA**→
env-23	JCVI_SCAF_1101668592552	+	167	255	**RNA**→
env-24	JCVI_SCAF_1096627390075	+	64503	64583	**RNA**→ ? (GOS11304561)→
[1]env-25	ABEF01047328.1	+	2089	2190	**RNA**→
env-26	ABEF01001886.1	+	667	767	**RNA**→ hypo→

36.4 Conserved domains

Conserved domains found in protein-coding genes listed in Section 36.3 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Gamma-*cis*-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

cd01186 (1) INT_SG3, DNA breaking-rejoining enzymes, integrase/recombinases subgroup 3, catalytic domain.
COG0596 (1) Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily) [General function prediction only]
COG1052 (1) Lactate dehydrogenase and related dehydrogenases [Energy production and conversion / Coenzyme metabolism / General function prediction only]
COG2175 (1) Probable taurine catabolism dioxygenase [Secondary metabolites biosynthesis, transport, and catabolism]
COG3236 (1) Uncharacterized protein conserved in bacteria [Function unknown]
COG3666 (1) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
COG5006 (1) Predicted permease, DMT superfamily [General function prediction only]
GOS11304561 (1) ?
GOS11643631 (1) ?

GOS11650400 (1) ?
GOS12001879 (1) ?
GOS1248 (1) ?
GOS14514810 (1) ?
GOS1730 (1) ?
GOS3111 (2) ?

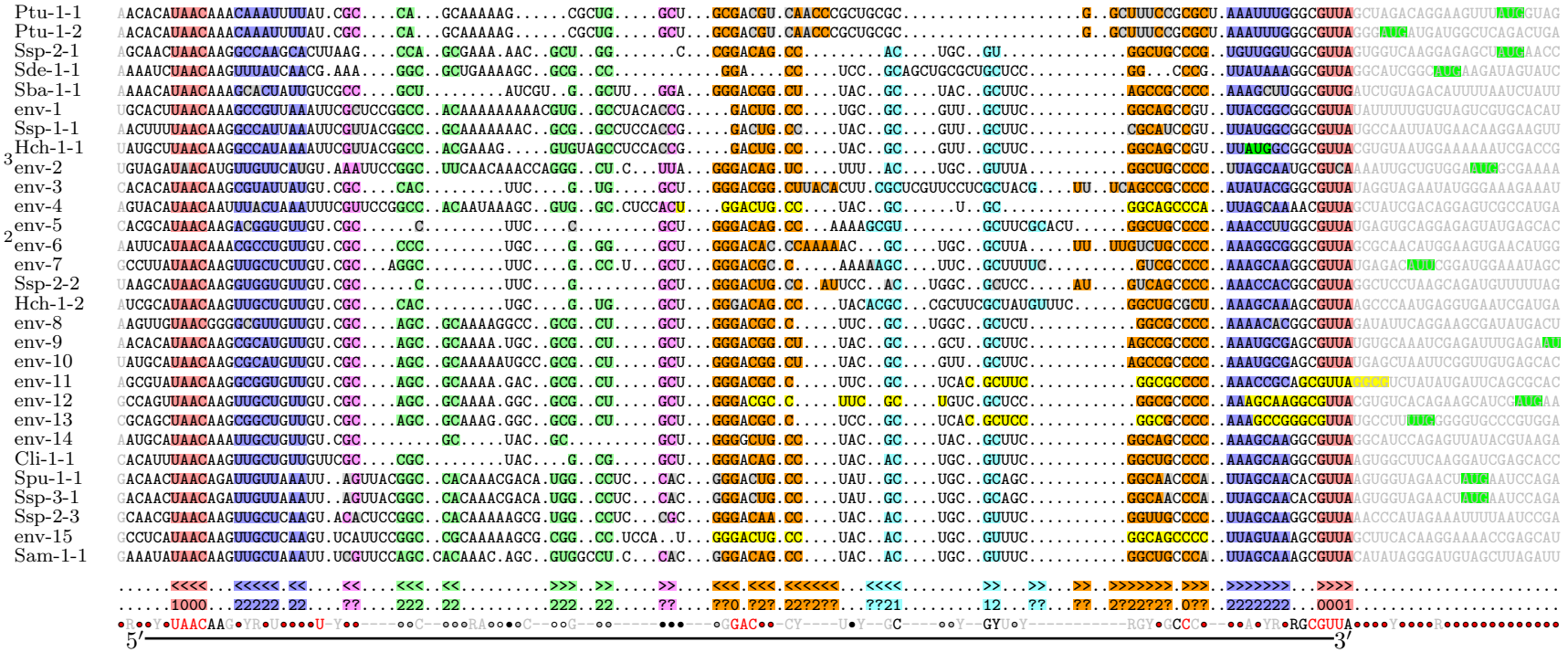
GOS8607 (1) ?
pfam00108 (2) Thiolase, N-terminal domain.
pfam00583 (3) Acetyltransferase (GNAT) family.
pfam00903 (1) Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily.
smart00530 (1) Helix-turn-helix XRE-family like proteins ;

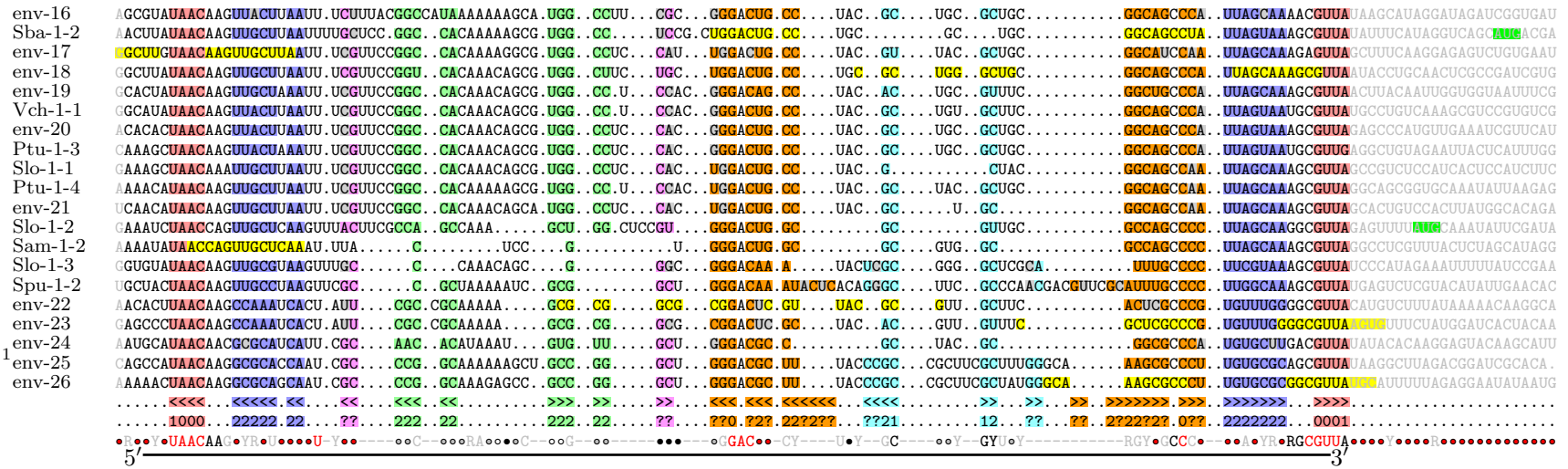
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Ssp-3-2

36.5 Multiple-sequence alignment

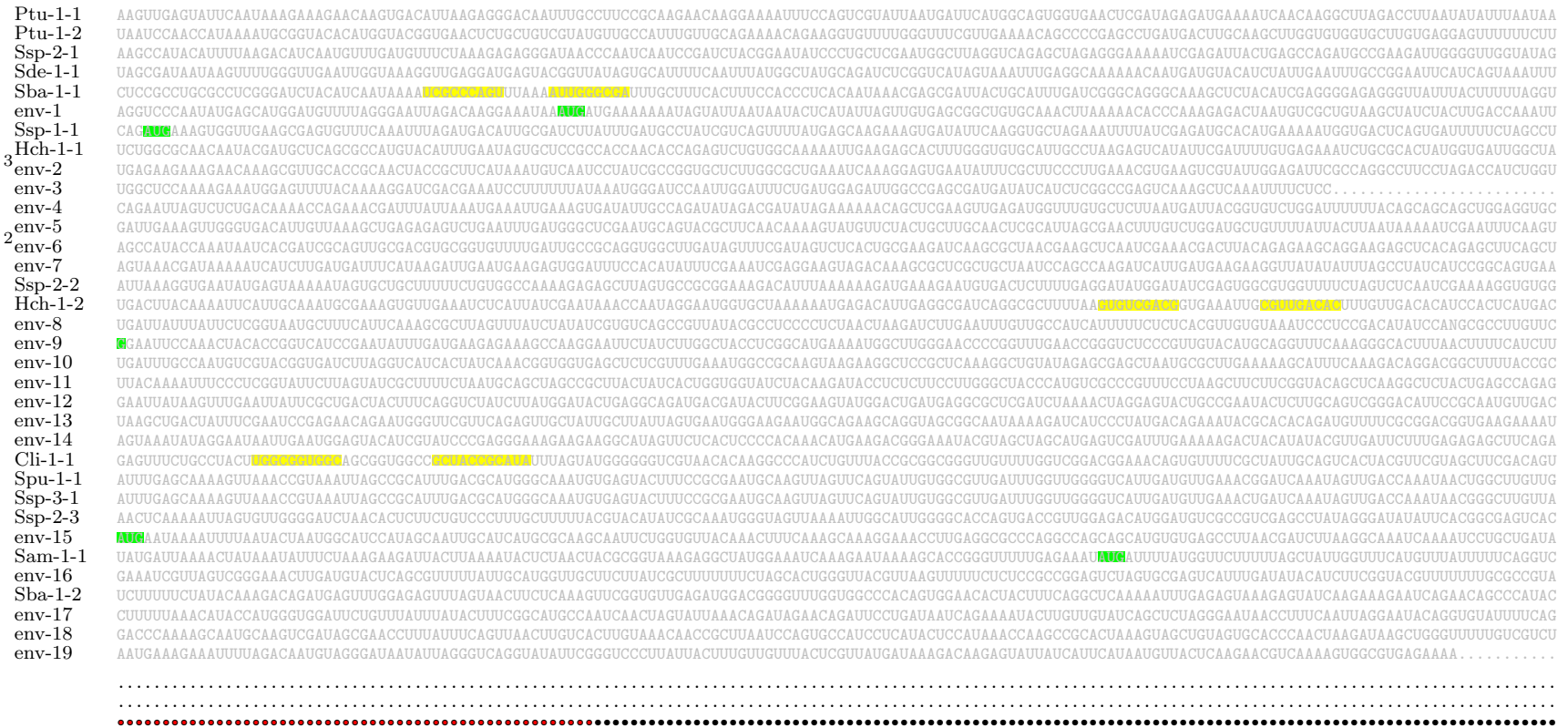
Each Gamma-*cis*-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 36.1. Superscript numbers refer to annotations listed in Section 36.2. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Gamma-*cis*-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

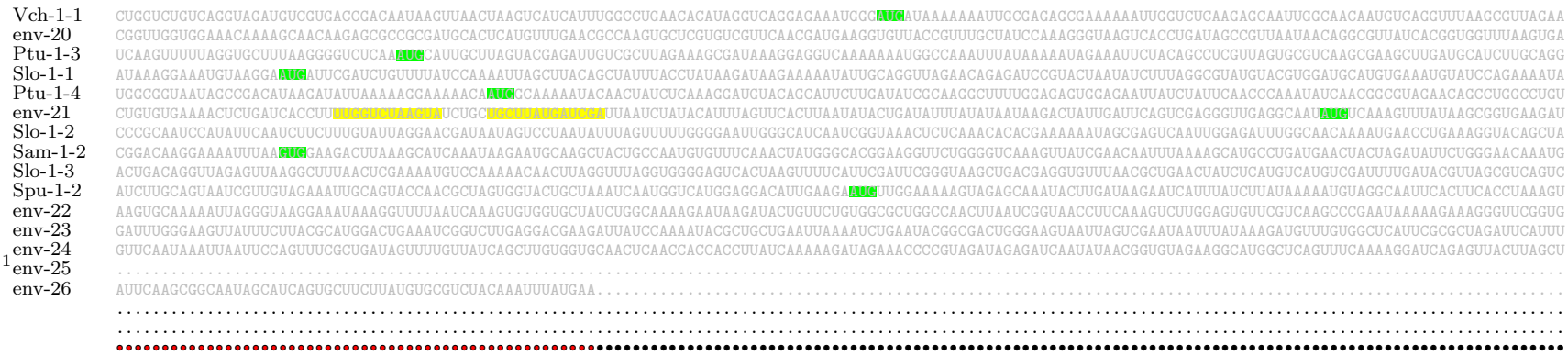
alignment positions 100 · · · 259



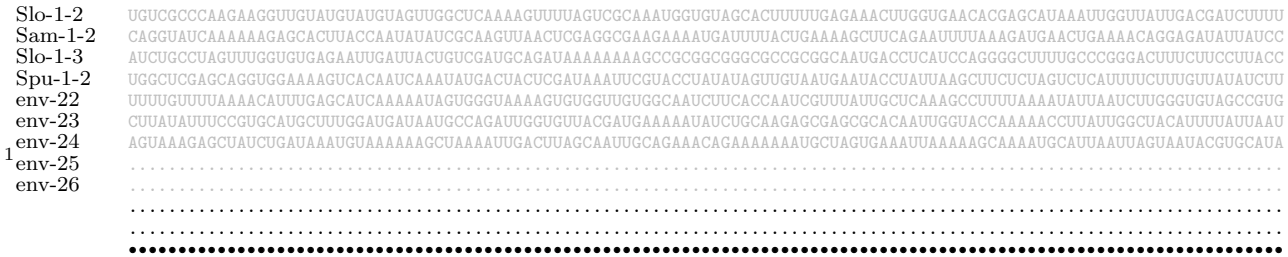


alignment positions 260 · · · 419





alignment positions 420 · · · 536



37 *glnA*

37.1 Taxa

The taxonomy of each organism containing a putative *glnA* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *glnA* RNA in Sections 37.2 and 37.4.

abbrev. of hits	**taxonomy of species**
Ama-1-1 to Ama-1-2	Bacteria Cyanobacteria *Acaryochloris marina* MBIC11017
Cwa-1-1 to Cwa-1-2	Bacteria Cyanobacteria Chroococcales *Crocosphaera watsonii* WH 8501 (Synechocystis sp. WH 8501)
Csp-1-1 to Csp-1-3	Bacteria Cyanobacteria Chroococcales *Cyanothece sp.* ATCC 51142
Csp-2-1	Bacteria Cyanobacteria Chroococcales *Cyanothece sp.* CCY 0110
Csp-3-1 to Csp-3-3	Bacteria Cyanobacteria Chroococcales *Cyanothece sp.* PCC 7424
Csp-4-1 to Csp-4-3	Bacteria Cyanobacteria Chroococcales *Cyanothece sp.* PCC 7425
Csp-5-1 to Csp-5-2	Bacteria Cyanobacteria Chroococcales *Cyanothece sp.* PCC 7822
Csp-6-1	Bacteria Cyanobacteria Chroococcales *Cyanothece sp.* PCC 8801
Csp-7-1	Bacteria Cyanobacteria Chroococcales *Cyanothece sp.* PCC 8802
Sel-1-1	Bacteria Cyanobacteria Chroococcales *Synechococcus elongatus* PCC 6301
Sel-2-1	Bacteria Cyanobacteria Chroococcales *Synechococcus elongatus* PCC 7942
Ssp-1-1 to Ssp-1-2	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* JA-2-3B'a(2-13)
Ssp-2-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* JA-3-3Ab
Ssp-3-1 to Ssp-3-2	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* PCC 7335
Ssp-4-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* RS9916
Ssp-5-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* RS9917
Ssp-6-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 5701
Ssp-7-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 7803
Ssp-8-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 7805
Ssp-9-1	Bacteria Cyanobacteria Chroococcales *Synechocystis sp.* PCC 6803
Tel-1-1 to Tel-1-2	Bacteria Cyanobacteria Chroococcales *Thermosynechococcus elongatus* BP-1
Ama-2-1 to Ama-2-2	Bacteria Cyanobacteria Oscillatoriales *Arthrospira maxima* CS-328
Lsp-1-1	Bacteria Cyanobacteria Oscillatoriales *Lyngbya sp.* PCC 8106
Mch-1-1	Bacteria Cyanobacteria Oscillatoriales *Microcoleus chthonoplastes* PCC 7420
Ter-1-1	Bacteria Cyanobacteria Oscillatoriales *Trichodesmium erythraeum* IMS101
Pma-1-1	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9303
Pma-2-1	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9313
Pph-1-1	Viruses dsDNA viruses, no RNA stage Caudovirales Myoviridae T4-like viruses *Prochlorococcus phage* P-SSM2
Sph-1-1	Viruses dsDNA viruses, no RNA stage Caudovirales Myoviridae T4-like viruses *Synechococcus phage* S-PM2
env-1 to env-997	environmental samples

37.2 Gene contexts

Each *glnA* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *glnA* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 37.3. The accession number of the sequence containing each *glnA* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 37.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	JCVI_SCAF_1096627355794	+	4	62	**RNA**→ ←Dcm (COG0270)? (GOS606)
Ssp-5-1	NZ_AANP01000006.1	-	217608	217533	**RNA**→ possible Extracellular link domain protein→
env-2	JCVI_SCAF_1101667315177	-	642	567	**RNA**→ ? (GOS11377265)→ ? (GOS16501)→
env-3	JCVI_SCAF_1101668630508	-	1057	982	**RNA**→ ? (GOS11377265)→ ? (GOS16501)→ GlnK (COG0347)? (GOS11223)→
env-4	JCVI_SCAF_1096627357780	+	5650	5725	**RNA**→ ? (GOS11377265)→ ? (GOS16501)→ GlnK (COG0347)? (GOS11223)→
env-5	JCVI_SCAF_1096627358964	-	1891	1816	**RNA**→ ? (GOS11377265)→ ? (GOS16501)→ GlnK (COG0347)? (GOS11223)→
env-6	JCVI_SCAF_1096626709985	+	433	508	**RNA**→ ? (GOS11377265)→ ? (GOS16501)→ GlnK (COG0347)? (GOS11223)→
Ssp-6-1	NZ_AANO01000002.1	+	225182	225256	**RNA**→ possible Extracellular link domain protein→ hypo→ GlnK (COG0347)→
Sel-1-1	NC_006576.1	+	1705656	1705712	**RNA**→ ←NC (pfam04970)
Sel-2-1	NC_007604.1	-	2611130	2611074	**RNA**→ hypo→
env-7	JCVI_SCAF_1101667522281	-	633	550	**RNA**→ ? (GOS1720)→
env-8	JCVI_SCAF_1101667254791	+	120	193	**RNA**→ ? (GOS14643050)→
env-9	JCVI_SCAF_1096627077622	-	884	811	**RNA**→ ? (GOS14643050)→ ? (GOS16501)→ GlnK (COG0347)? (GOS11223)→
env-10	JCVI_SCAF_1096627077904	+	318	391	**RNA**→ ? (GOS14643050)→ ? (GOS16501)→ GlnK (COG0347)? (GOS11223)→
env-11	JCVI_SCAF_1096627070116	-	124	51	**RNA**→ ? (GOS14643050)→
Ssp-8-1	NZ_AAOK01000001.1	-	260895	260820	**RNA**→ possible Extracellular link domain protein→ hypo→ GlnK (COG0347)→
Ssp-7-1	NC_009481.1	+	1858904	1858979	**RNA**→ hypo→ hypo→ GlnK (COG0347)→
env-12	JCVI_SCAF_1096627553262	-	529	472	**RNA**→
env-13	JCVI_SCAF_1096627531828	+	156	211	**RNA**→ ? (GOS11194950)→
env-14	JCVI_SCAF_1096628003214	-	1798	1744	**RNA**→ ←? (GOS11195016)
env-15	JCVI_SCAF_1101668228312	+	266	322	**RNA**→ ? (GOS11025)→
env-16	JCVI_SCAF_1096627537457	+	234	287	**RNA**→ PstS (COG0226)? (GOS1508)→
env-17	JCVI_SCAF_1096626928539	-	1398	1344	**RNA**→ ? (GOS15797)→
env-18	JCVI_SCAF_1101668182179	+	921	977	**RNA**→ ? (GOS11194950)→
env-19	JCVI_SCAF_1096627634647	+	45	103	**RNA**→
Ama-1-1	NC_009925.1	+	3254265	3254355	**RNA**→ hypo→
env-20	JCVI_SCAF_1101667194234	-	600	481	**RNA**→
Ama-1-2	NC_009925.1	+	2595141	2595223	**RNA**→ hypo→
env-21	JCVI_SCAF_1096627480451	+	1344	1404	**RNA**→
env-22	JCVI_SCAF_1096626879438	-	2040	1979	**RNA**→ ? (GOS11025)→
env-23	JCVI_SCAF_1096627888893	-	189	142	**RNA**→
env-24	JCVI_SCAF_1096627318946	-	587	536	**RNA**→ ? (GOS13251683)→ ? (GOS11034)→
env-25	JCVI_SCAF_1101668319706	-	502	456	**RNA**→ ? (GOS15286)→ ? (GOS12174553)→
env-26	JCVI_SCAF_1101668391020	+	740	786	**RNA**→
env-27	JCVI_SCAF_1096628009553	+	183	229	**RNA**→ ? (GOS15599)→
env-28	JCVI_SCAF_1096627905606	-	292	246	**RNA**→ ? (GOS12119065)→
env-29	JCVI_SCAF_1096626937891	+	485	533	**RNA**→ Thy1 (pfam02511)? (GOS5954)→
env-30	JCVI_SCAF_1096627342341	+	301	347	**RNA**→ ? (GOS12119065)→ ? (GOS15286)→
env-31	JCVI_SCAF_1101668558910	-	701	658	**RNA**→
env-32	JCVI_SCAF_1096628118909	-	627	579	**RNA**→ ? (GOS15286)→ ? (GOS19514)→
env-33	JCVI_SCAF_1096628275403	-	1797	1694	**RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→

env-34	JCVI_SCAF_1096626544885	-	893	788	**RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-35	JCVI_SCAF_1096627242437	-	1629	1524	**RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-36	JCVI_SCAF_1096628163644	+	1091	1196	**RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-37	JCVI_SCAF_1096626945342	-	1327	1222	**RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-38	JCVI_SCAF_1096627245638	-	2603	2498	**RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
Ter-1-1	NC_008312.1	-	4678876	4678823	**RNA**→ hypo→
env-39	JCVI_SCAF_1101667555549	-	241	109	**RNA**→
env-40	JCVI_SCAF_1096627294440	-	1039	901	**RNA**→ ←SecF (COG0341)? (GOS431)
env-41	JCVI_SCAF_1096626283011	-	503	445	**RNA**→
env-42	JCVI_SCAF_1096627528507	+	1319	1384	**RNA**→ ? (GOS11194950)→
env-43	JCVI_SCAF_1101668519955	-	1530	1468	**RNA**→ ? (GOS46)→ Ammonium_transp (pfam00909)? (GOS46)→
env-44	JCVI_SCAF_1101668755609	+	60	133	**RNA**→ ←? (GOS11672939)
Ama-2-1	NZ_ABYK01000008.1	+	24464	24568	**RNA**→ hypo→
env-45	JCVI_SCAF_1096628394569	+	1565	1646	**RNA**→
env-46	JCVI_SCAF_1096627386197	-	221	140	**RNA**→
env-47	JCVI_SCAF_1096627272733	+	284	409	**RNA**→ ? (GOS11295941)→ UbiG (COG2227)? (GOS721)→ Glyoxalase (pfam00903)? (GOS1339)→ MhpC (COG0596)? (GOS4325)→ adh_short (pfam00106)? (GOS49)→
env-48	JCVI_SCAF_1096627764200	-	1155	1039	**RNA**→ ? (GOS12637)→ ? (GOS9229)→
env-49	JCVI_SCAF_1096627580829	+	475	526	**RNA**→ ? (GOS15286)→ ? (GOS11833740)→
env-50	JCVI_SCAF_1096627258300	-	8684	8627	**RNA**→ ? (GOS8468)→ ? (GOS11251)→ PiuC (COG3128)? (GOS1330)→ ? (GOS7556)→
env-51	JCVI_SCAF_1096628179846	-	1016	959	**RNA**→ TpiA (COG0149)? (GOS6286)→
env-52	JCVI_SCAF_1101667501166	+	433	490	**RNA**→
env-53	JCVI_SCAF_1096627244519	-	1222	1165	**RNA**→ ? (GOS8468)→ ? (GOS11251)→
env-54	JCVI_SCAF_1096627385745	+	2401	2458	**RNA**→ ? (GOS8468)→ ? (GOS11251)→ P4Hc (smart00702)? (GOS1330)→ ? (GOS7556)→ SSL2 (COG1061)? (GOS164)→ ? (GOS1457)→
env-55	JCVI_SCAF_1096628376916	+	1055	1112	**RNA**→
Sph-1-1	NC_006820.1	+	142359	142415	**RNA**→ hypo→ hypo→ hypo→ thymidilate synthase→ hypo→ PhoH→ putative exonuclease→ putative late promoter transcription factor→ hypo→ ssDNA binding protein gp32→
env-56	JCVI_SCAF_1096626958398	+	1278	1327	**RNA**→ ? (GOS15286)→ ? (GOS12056256)→
env-57	JCVI_SCAF_1096627583466	+	1694	1743	**RNA**→
env-58	JCVI_SCAF_1096627979444	-	757	703	**RNA**→ ? (GOS11025)→
env-59	JCVI_SCAF_1096626966619	-	1740	1694	**RNA**→ ? (GOS15286)→
env-60	JCVI_SCAF_1096627706162	+	993	1041	**RNA**→ ? (GOS15286)→ ? (GOS19514)→
env-61	JCVI_SCAF_1101667130020	-	499	451	**RNA**→ ? (GOS15286)→
env-62	JCVI_SCAF_1096627593511	+	1151	1199	**RNA**→ ? (GOS15286)→
env-63	JCVI_SCAF_1096626955717	-	590	542	**RNA**→ ? (GOS15286)→
env-64	JCVI_SCAF_1096627736976	+	1213	1261	**RNA**→
env-65	JCVI_SCAF_1096626929666	+	1354	1402	**RNA**→ ? (GOS15286)→
env-66	JCVI_SCAF_1096626956894	+	787	835	**RNA**→ ? (GOS15286)→
env-67	JCVI_SCAF_1096627778758	+	781	829	**RNA**→ ? (GOS15286)→
env-68	JCVI_SCAF_1101668210897	-	709	661	**RNA**→ ? (GOS15286)→
env-69	JCVI_SCAF_1096627607567	-	642	594	**RNA**→ ? (GOS15286)→
env-70	JCVI_SCAF_1096628277120	-	411	363	**RNA**→ ? (GOS15286)→
env-71	JCVI_SCAF_1096627049209	-	614	566	**RNA**→ ? (GOS15286)→ ? (GOS1330)→
env-72	JCVI_SCAF_1096626962982	-	903	855	**RNA**→ ? (GOS17289)→ ? (GOS5954)→
env-73	JCVI_SCAF_1101668166887	+	288	339	**RNA**→ ? (GOS11194950)→ ? (GOS11025)→ ? (GOS8589)→
env-74	JCVI_SCAF_1101668284766	-	486	438	**RNA**→
env-75	JCVI_SCAF_1101668286815	-	55	7	**RNA**→
env-76	JCVI_SCAF_1096627750024	-	1292	1244	**RNA**→ Thy1 (pfam02511)? (GOS5954)→ ? (GOS13561)→
env-77	JCVI_SCAF_1096627709848	+	973	1021	**RNA**→ ? (GOS15286)→
env-78	JCVI_SCAF_1096627719263	+	1311	1359	**RNA**→ ? (GOS15286)→ ? (GOS19514)→
env-79	JCVI_SCAF_1101667128953	+	534	582	**RNA**→ ? (GOS15286)→
env-80	JCVI_SCAF_1096627725731	-	424	376	**RNA**→ ? (GOS15286)→

env-81	JCVI_SCAF_1096626964267	+	1710	1758	**RNA**→ ? (GOS15286)→
env-82	JCVI_SCAF_1096627035542	+	1008	1056	**RNA**→ ? (GOS15286)→ ? (GOS19514)→
env-83	JCVI_SCAF_1096627766995	+	1259	1307	**RNA**→ ? (GOS15286)→
env-84	JCVI_SCAF_1101667158329	-	845	797	**RNA**→ ? (GOS15286)→ ? (GOS19514)→ ? (GOS7317)→
env-85	JCVI_SCAF_1101668321986	-	562	514	**RNA**→ ? (GOS15286)→ ? (GOS19514)→
env-86	JCVI_SCAF_1096627060460	-	178	130	**RNA**→
env-87	JCVI_SCAF_1101668743799	+	1295	1343	**RNA**→ ? (GOS15286)→
env-88	JCVI_SCAF_1101668291709	-	127	79	**RNA**→
env-89	JCVI_SCAF_1096626242393	+	409	457	**RNA**→ ? (GOS15286)→
env-90	JCVI_SCAF_1096626237175	-	243	195	**RNA**→ ? (GOS15286)→
env-91	JCVI_SCAF_1101667205597	-	71	23	**RNA**→
env-92	JCVI_SCAF_1101668322629	+	527	575	**RNA**→ ? (GOS15286)→
env-93	JCVI_SCAF_1096627282429	+	772	835	**RNA**→ ? (GOS11421997)→
env-94	JCVI_SCAF_1096626861083	+	1235	1297	**RNA**→
env-95	JCVI_SCAF_1096626846484	+	3605	3667	**RNA**→ ? (GOS11421997)→
env-96	JCVI_SCAF_1101668013918	+	1414	1476	**RNA**→
env-97	JCVI_SCAF_1101668716459	-	807	761	**RNA**→ ? (GOS17289)→ ? (GOS9797)→
env-98	JCVI_SCAF_1101668548586	-	1328	1280	**RNA**→ IbpA (COG0071)? (GOS1219)→
env-99	JCVI_SCAF_1101668695858	+	779	827	**RNA**→ ? (GOS15286)→
env-100	JCVI_SCAF_1101668499361	+	814	862	**RNA**→
env-101	JCVI_SCAF_1096627117896	+	1227	1275	**RNA**→ ? (GOS11141882)→
Pph-1-1	NC_006883.1	+	247163	247214	**RNA**→ ←hypo
env-102	JCVI_SCAF_1096627913912	+	1012	1063	**RNA**→ ? (GOS15286)→
env-103	JCVI_SCAF_1096627081105	-	1030	980	**RNA**→ ? (GOS15286)→ ? (GOS17794)→ ? (GOS17289)→ ? (GOS5954)→
env-104	JCVI_SCAF_1101668278476	+	680	730	**RNA**→ ? (GOS17794)→
env-105	JCVI_SCAF_1096627981730	+	1209	1258	**RNA**→ ? (GOS11141882)→
env-106	JCVI_SCAF_1096627738076	-	504	458	**RNA**→ ? (GOS15286)→ ? (GOS17794)→
env-107	JCVI_SCAF_1096627059820	+	125	171	**RNA**→ ? (GOS15286)→ PMM (pfam03332)? (GOS7317)→ ? (GOS12901)→ ? (GOS6163)→
env-108	JCVI_SCAF_1096627029879	-	1165	1120	**RNA**→ ? (GOS15286)→ PMM (pfam03332)? (GOS7317)→
env-109	JCVI_SCAF_1096627090889	-	1754	1709	**RNA**→ ? (GOS15286)→ PMM (pfam03332)? (GOS7317)→ ? (GOS12901)→ ? (GOS1110)→
env-110	JCVI_SCAF_1101668656216	+	1281	1326	**RNA**→ ? (GOS15286)→
env-111	JCVI_SCAF_1101667268174	-	392	347	**RNA**→ ? (GOS15286)→
env-112	JCVI_SCAF_1096627973485	+	715	760	**RNA**→ PMM (pfam03332)? (GOS7317)→
env-113	JCVI_SCAF_1096627170580	-	663	618	**RNA**→ ? (GOS15286)→ ? (GOS7317)→
env-114	JCVI_SCAF_1101667348900	+	191	236	**RNA**→
env-115	JCVI_SCAF_1101668401770	-	724	679	**RNA**→ ? (GOS15286)→ ? (GOS7317)→
env-116	JCVI_SCAF_1101667208281	-	463	412	**RNA**→ ? (GOS11194950)→
env-117	JCVI_SCAF_1101667241724	+	568	624	**RNA**→
env-118	JCVI_SCAF_1096627006028	-	1649	1600	**RNA**→ ? (GOS17794)→ Thy1 (pfam02511)? (GOS5954)→ ? (GOS13561)→ ? (GOS1288)→
env-119	JCVI_SCAF_1101667032616	-	442	388	**RNA**→
env-120	JCVI_SCAF_1096627794682	+	388	442	**RNA**→ **RNA**→ ? (GOS13113871)→
env-122	JCVI_SCAF_1096627794682	+	444	507	(shown above)
Csp-1-1	NC_010546.1	-	2707245	2707170	**RNA**→ hypo→
env-121	JCVI_SCAF_1101667143436	-	436	381	**RNA**→
env-123	JCVI_SCAF_1096627556098	+	1050	1112	**RNA**→ ? (GOS11070785)→
env-124	JCVI_SCAF_1096627041542	-	1034	971	**RNA**→ **RNA**→ ? (GOS13113871)→
env-392	JCVI_SCAF_1096627041542	-	970	918	(shown above)
env-125	JCVI_SCAF_1096626879418	+	1566	1629	**RNA**→ 16S_rRNA,5S_rRNA→
env-126	JCVI_SCAF_1096626966023	+	1977	2040	**RNA**→ **RNA**→ ? (GOS13113871)→
env-400	JCVI_SCAF_1096626966023	+	2041	2093	(shown above)
env-127	JCVI_SCAF_1101667484142	-	480	417	**RNA**→ **RNA**→ ? (GOS13113871)→
env-395	JCVI_SCAF_1101667484142	-	416	364	(shown above)

env-128	JCVI_SCAF_1101668268859	+	100	163	**RNA**→ **RNA**→ ? (GOS13113871)→
env-338	JCVI_SCAF_1101668268859	+	164	216	(shown above)
env-129	JCVI_SCAF_1096627127231	-	586	523	**RNA**→ **RNA**→ ? (GOS13113871)→ 5S_rRNA→ 16S_rRNA,5S_rRNA→
env-394	JCVI_SCAF_1096627127231	-	522	470	(shown above)
env-130	JCVI_SCAF_1096626910112	-	4925	4862	**RNA**→ **RNA**→ ? (GOS13113871)→
env-398	JCVI_SCAF_1096626910112	-	4861	4809	(shown above)
env-131	JCVI_SCAF_1096627124111	+	683	746	**RNA**→ **RNA**→ ? (GOS13113871)→
env-397	JCVI_SCAF_1096627124111	+	747	799	(shown above)
env-132	JCVI_SCAF_1096627080456	+	3686	3749	**RNA**→ **RNA**→ ? (GOS13113871)→
env-399	JCVI_SCAF_1096627080456	+	3750	3802	(shown above)
env-133	JCVI_SCAF_1101668537207	+	46	109	**RNA**→ **RNA**→ ? (GOS13113871)→ 16S_rRNA,5S_rRNA→ 5S_rRNA→ 16S_rRNA,18S_rRNA,5S_rRNA→
env-393	JCVI_SCAF_1101668537207	+	110	162	(shown above)
env-134	JCVI_SCAF_1096627130209	+	1346	1409	**RNA**→ **RNA**→ ? (GOS13113871)→ 16S_rRNA,5S_rRNA→ 5S_rRNA→ 16S_rRNA,18S_rRNA,5S_rRNA→
env-396	JCVI_SCAF_1096627130209	+	1410	1462	(shown above)
env-135	JCVI_SCAF_1101668528491	-	113	50	**RNA**→
env-137	JCVI_SCAF_1101667247305	-	609	552	**RNA**→
env-138	JCVI_SCAF_1096626845051	-	383	329	**RNA**→
env-139	JCVI_SCAF_1101668292618	+	460	513	**RNA**→ ? (GOS11670331)→
env-140	JCVI_SCAF_1096627103299	-	153	99	**RNA**→
env-141	JCVI_SCAF_1096627172087	+	755	808	**RNA**→ ? (GOS11670331)→
env-142	JCVI_SCAF_1096627784782	-	166	116	**RNA**→ ? (GOS11670331)→
env-143	JCVI_SCAF_1101667236253	-	661	611	**RNA**→ ? (GOS11670331)→
env-144	JCVI_SCAF_1096627824195	+	394	444	**RNA**→ ? (GOS11670331)→
env-145	JCVI_SCAF_1096627865163	+	546	596	**RNA**→ ? (GOS11670331)→
env-146	JCVI_SCAF_1101668336149	-	471	421	**RNA**→ ? (GOS11670331)→
env-147	JCVI_SCAF_1096627279180	-	979	851	**RNA**→ ←? (GOS7297)
env-148	JCVI_SCAF_1101667117392	-	208	77	**RNA**→
env-149	JCVI_SCAF_1096627793828	-	1537	1406	**RNA**→ ←? (GOS13261206)
env-150	JCVI_SCAF_1101668358160	-	726	595	**RNA**→ ←? (GOS13261206)
env-151	JCVI_SCAF_1096627975658	-	1192	1062	**RNA**→ ←? (GOS13261206)
env-152	JCVI_SCAF_1096627998782	+	1625	1756	**RNA**→ ←? (GOS13261206)
env-153	JCVI_SCAF_1101668331157	+	411	541	**RNA**→ ? (GOS14064195)→
env-154	JCVI_SCAF_1096627172186	-	786	655	**RNA**→ ←? (GOS13261206)
Csp-3-1	NZ_ABOY01000018.1	-	62085	62009	**RNA**→ hypo→
env-155	JCVI_SCAF_1101668736363	+	60	113	**RNA**→ ? (GOS11531948)→
env-156	JCVI_SCAF_1096627055100	-	1267	1172	**RNA**→ RpsU (COG0828)→
env-158	JCVI_SCAF_1101667078320	+	894	949	**RNA**→
env-160	JCVI_SCAF_1101668053145	-	303	250	**RNA**→
env-161	JCVI_SCAF_1096628270535	+	1028	1075	**RNA**→ ? (GOS9335)→ ? (GOS7261)→
env-162	JCVI_SCAF_1101668388391	+	405	452	**RNA**→ ? (GOS11670331)→
env-163	JCVI_SCAF_1096627056769	-	332	285	**RNA**→ ? (GOS11024366)→
env-164	JCVI_SCAF_1096627155593	-	979	931	**RNA**→ ? (GOS11024366)→
env-165	JCVI_SCAF_1096627294720	-	944	896	**RNA**→ ? (GOS11024366)→
env-166	JCVI_SCAF_1101667031347	-	664	617	**RNA**→ ? (GOS11200173)→
env-167	JCVI_SCAF_1096626871571	-	410	362	**RNA**→ ? (GOS11024366)→
env-168	JCVI_SCAF_1096627571351	+	951	1034	**RNA**→ ←? (GOS11477096)
Csp-2-1	NZ_AAXW01000008.1	-	84668	84587	**RNA**→ riboflavin synthase subunit alpha→
Csp-1-2	NC_010546.1	+	2393902	2393983	**RNA**→ hypo→
env-169	JCVI_SCAF_1096627394014	+	168	217	**RNA**→ ←? (GOS11075491)
env-170	JCVI_SCAF_1096627558708	-	508	458	**RNA**→
env-171	JCVI_SCAF_1096627864484	+	1324	1374	**RNA**→
Csp-4-1	NZ_ABVJ01000012.1	+	104536	104631	**RNA**→ hypo (cons)→

env-172	JCVI_SCAF_1096627701349	-	396	342	**RNA**→
env-173	JCVI_SCAF_1096627326751	-	1424	1369	**RNA**→ **RNA**→ ←? (GOS142)
env-412	JCVI_SCAF_1096627326751	-	1365	1313	(shown above)
env-174	JCVI_SCAF_1096628198335	+	423	483	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→ GyrB (COG0187)? (GOS1203)→
env-621	JCVI_SCAF_1096628198335	+	496	547	(shown above)
env-175	JCVI_SCAF_1096627892500	-	438	378	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-623	JCVI_SCAF_1096627892500	-	365	315	(shown above)
env-176	JCVI_SCAF_1101668218319	+	256	309	**RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-177	JCVI_SCAF_1096627234618	+	1170	1228	**RNA**→ ? (GOS46)→
env-178	JCVI_SCAF_1096628222158	+	1332	1390	**RNA**→ ? (GOS46)→
env-179	JCVI_SCAF_1096627220767	+	419	478	**RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
Cwa-1-1	NZ_AADV02000064.1	-	2221	2140	**RNA**→ ←COG5553 (COG5553)
Csp-4-2	NZ_ABVJ01000011.1	-	7142	7046	**RNA**→ hypo (cons)→
Tel-1-1	NC_004113.1	-	1753204	1753120	**RNA**→ hypo→
env-180	JCVI_SCAF_1101668327369	+	1297	1358	**RNA**→
Ssp-1-1	NC_007776.1	+	2902190	2902288	**RNA**→ hypo→
env-181	JCVI_SCAF_1101668327369	+	1362	1418	**RNA**→
env-182	JCVI_SCAF_1101667247305	-	548	490	**RNA**→
env-183	JCVI_SCAF_1096626283011	-	441	384	**RNA**→
env-184	JCVI_SCAF_1096627533222	+	260	381	**RNA**→ ? (GOS14064195)→
env-185	JCVI_SCAF_1096626467744	+	570	623	**RNA**→ ? (GOS3708)→
env-186	JCVI_SCAF_1096627203859	-	10605	10463	**RNA**→ ? (GOS11295941)→
env-187	JCVI_SCAF_1096627048988	-	308	212	**RNA**→
env-188	JCVI_SCAF_1101668672515	+	48	165	**RNA**→ ? (GOS13271387)→
env-189	JCVI_SCAF_1101668476000	+	502	562	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-647	JCVI_SCAF_1101668476000	+	575	624	(shown above)
env-190	JCVI_SCAF_1096627221839	+	777	836	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-646	JCVI_SCAF_1096627221839	+	849	901	(shown above)
env-191	JCVI_SCAF_1096628257329	+	1503	1561	**RNA**→
env-192	JCVI_SCAF_1101668489971	+	511	569	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-643	JCVI_SCAF_1101668489971	+	582	634	(shown above)
env-193	JCVI_SCAF_1101667572161	+	255	313	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-625	JCVI_SCAF_1101667572161	+	326	379	(shown above)
env-194	JCVI_SCAF_1101667384842	+	56	114	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-645	JCVI_SCAF_1101667384842	+	127	179	(shown above)
env-195	JCVI_SCAF_1096628199405	-	1553	1494	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→ GlpK (COG0554)? (GOS667)→
env-322	JCVI_SCAF_1096628199405	-	1481	1427	(shown above)
env-196	JCVI_SCAF_1096627229208	+	757	816	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-481	JCVI_SCAF_1096627229208	+	829	878	(shown above)
env-197	JCVI_SCAF_1096628222536	-	1168	1110	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)? (GOS3709)→
env-611	JCVI_SCAF_1096628222536	-	1097	1047	(shown above)
env-198	JCVI_SCAF_1096628272537	-	1135	1077	**RNA**→ **RNA**→ ? (GOS3709)→ ? (GOS1203)→
env-618	JCVI_SCAF_1096628272537	-	1064	1014	(shown above)
env-199	JCVI_SCAF_1096628097878	-	315	257	**RNA**→
env-200	JCVI_SCAF_1096628217934	+	1393	1451	**RNA**→
env-201	JCVI_SCAF_1101668543518	+	1122	1182	**RNA**→
env-202	JCVI_SCAF_1096627793293	-	1042	984	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-614	JCVI_SCAF_1096627793293	-	971	921	(shown above)
env-203	JCVI_SCAF_1101667034399	-	276	217	**RNA**→
env-204	JCVI_SCAF_1101668388409	+	868	927	**RNA**→
env-205	JCVI_SCAF_1101668342362	-	418	359	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-620	JCVI_SCAF_1101668342362	-	346	296	(shown above)

env-206	JCVI_SCAF_1096627444611	-	1228	1169	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)? (GOS3709)→
env-971	JCVI_SCAF_1096627444611	-	1156	1103	(shown above)
env-207	JCVI_SCAF_1101668323352	-	608	548	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-972	JCVI_SCAF_1101668323352	-	535	484	(shown above)
env-208	JCVI_SCAF_1101668687189	-	286	227	**RNA**→
env-209	JCVI_SCAF_1096628073689	+	1841	1901	**RNA**→
env-210	JCVI_SCAF_1101668358887	+	1020	1079	**RNA**→
env-211	JCVI_SCAF_1096628216939	+	1496	1556	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-243	JCVI_SCAF_1096628216939	+	1569	1624	(shown above)
env-212	JCVI_SCAF_1101667572934	+	299	359	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-244	JCVI_SCAF_1101667572934	+	372	427	(shown above)
env-213	JCVI_SCAF_1101667511635	+	766	826	**RNA**→
env-214	JCVI_SCAF_1096627689138	+	1414	1462	**RNA**→ ←Leu tRNA
env-215	JCVI_SCAF_1096627481796	+	1182	1230	**RNA**→ ? (GOS11024366)→
Lsp-1-1	NZ_AAVU01000006.1	-	147500	147393	**RNA**→ hypo→
Tel-1-2	NC_004113.1	-	147826	147714	**RNA**→ hypo→
env-216	JCVI_SCAF_1096627769335	+	593	675	**RNA**→
Ssp-1-2	NC_007776.1	-	2191743	2191662	**RNA**→ ←hypo
env-217	JCVI_SCAF_1096627527311	-	62	3	**RNA**→
env-219	JCVI_SCAF_1101668006615	+	943	993	**RNA**→
env-220	JCVI_SCAF_1096626910894	-	2162	2111	**RNA**→ **RNA**→ ←ligand_gated_channel (cd01347)? (GOS1720)
env-427	JCVI_SCAF_1096626910894	-	2107	2054	(shown above)
env-221	JCVI_SCAF_1096624997047	+	699	750	**RNA**→
Csp-1-3	NC_010546.1	+	1043231	1043292	**RNA**→ hypo→ hypo→
Cwa-1-2	NZ_AADV02000026.1	-	14400	14340	**RNA**→ hypo→
env-222	JCVI_SCAF_1096627619435	-	327	277	**RNA**→
env-223	JCVI_SCAF_1096627875590	-	1406	1345	**RNA**→ **RNA**→ ? (GOS46)→ ? (GOS11907)→
env-833	JCVI_SCAF_1096627875590	-	1332	1278	(shown above)
env-224	JCVI_SCAF_1096627875590	-	473	412	**RNA**→ **RNA**→ ? (GOS46)→
env-834	JCVI_SCAF_1096627875590	-	399	345	(shown above)
env-225	JCVI_SCAF_1096628133299	+	938	999	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-835	JCVI_SCAF_1096628133299	+	1012	1066	(shown above)
Csp-4-3	NZ_ABVJ01000008.1	-	134172	134088	**RNA**→ hypo (cons)→
env-226	JCVI_SCAF_1096627454446	-	958	905	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ? (GOS8353)→
env-227	JCVI_SCAF_1096627395352	-	1215	1163	**RNA**→ ? (GOS8468)→ ? (GOS11251)→
env-228	JCVI_SCAF_1101668107327	-	711	659	**RNA**→ ? (GOS11024366)→ AHP1 (COG0678)? (GOS7685)→
env-229	JCVI_SCAF_1096627229247	+	4544	4597	**RNA**→ COG2308 (COG2308)? (GOS3708)→ DUF403 (pfam04168)? (GOS5926)→
env-230	JCVI_SCAF_1101667550414	-	136	83	**RNA**→
env-231	JCVI_SCAF_1096628283615	+	1205	1258	**RNA**→ DUF404 (pfam04169)? (GOS3708)→
env-232	JCVI_SCAF_1096627200717	-	386	334	**RNA**→ ? (GOS3708)→
Ssp-3-1	NZ_ABRV01000025.1	+	6494	6677	**RNA**→
env-233	JCVI_SCAF_1096627274504	+	1001	1059	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-457	JCVI_SCAF_1096627274504	+	1072	1126	(shown above)
env-234	JCVI_SCAF_1096627346029	+	979	1031	**RNA**→
env-235	JCVI_SCAF_1096626883488	+	426	478	**RNA**→ **RNA**→ ←? (GOS7356)
env-417	JCVI_SCAF_1096626883488	+	482	534	(shown above)
env-236	JCVI_SCAF_1096628100585	+	673	738	**RNA**→ **RNA**→ ←PanB (COG0413)? (GOS5559)
env-218	JCVI_SCAF_1096628100585	+	751	808	(shown above)
env-237	JCVI_SCAF_1101668014490	+	1222	1275	**RNA**→
env-238	JCVI_SCAF_1096627861307	-	881	823	**RNA**→ Glu_synth_NTN (pfam04897)? (GOS1994)→
env-239	JCVI_SCAF_1096627222235	-	915	772	**RNA**→ ? (GOS11295941)→
env-240	JCVI_SCAF_1101667422402	-	670	527	**RNA**→ ? (GOS11295941)→

env-245	JCVI_SCAF_1101667457688	-	719	664	**RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-246	JCVI_SCAF_1096626994412	-	2051	1995	**RNA**→ **RNA**→ COG2308 (COG2308)? (GOS3708)→ ? (GOS5926)→
env-830	JCVI_SCAF_1096626994412	-	1984	1930	(shown above)
env-247	JCVI_SCAF_1096628256401	-	1037	982	**RNA**→ **RNA**→ COG2308 (COG2308)? (GOS3708)→
env-860	JCVI_SCAF_1096628256401	-	970	916	(shown above)
env-248	JCVI_SCAF_1101667569811	+	756	811	**RNA**→
env-249	JCVI_SCAF_1096628228853	-	152	97	**RNA**→
env-250	JCVI_SCAF_1096626999319	+	334	389	**RNA**→ **RNA**→ COG2308 (COG2308)? (GOS3708)→
env-771	JCVI_SCAF_1096626999319	+	401	455	(shown above)
env-251	JCVI_SCAF_1101668217991	+	1099	1154	**RNA**→ **RNA**→ ? (GOS3708)→
env-775	JCVI_SCAF_1101668217991	+	1166	1220	(shown above)
env-252	JCVI_SCAF_1096627605943	+	1492	1547	**RNA**→
env-253	JCVI_SCAF_1101667543498	+	412	467	**RNA**→ **RNA**→ ? (GOS3708)→
env-766	JCVI_SCAF_1101667543498	+	479	534	(shown above)
env-254	JCVI_SCAF_1096627519549	-	1259	1204	**RNA**→ **RNA**→ ? (GOS3708)→ COG2308 (COG2308)? (GOS3708)→
env-776	JCVI_SCAF_1096627519549	-	1192	1138	(shown above)
env-255	JCVI_SCAF_1096627631062	-	301	246	**RNA**→
env-256	JCVI_SCAF_1096627704160	+	134	189	**RNA**→ **RNA**→ COG2308 (COG2308)? (GOS3708)→
env-779	JCVI_SCAF_1096627704160	+	201	255	(shown above)
env-257	JCVI_SCAF_1096626085716	-	226	171	**RNA**→
env-258	JCVI_SCAF_1096628063320	-	1442	1387	**RNA**→ **RNA**→ ? (GOS3708)→ XylB (COG1070)? (GOS667)→
env-774	JCVI_SCAF_1096628063320	-	1375	1321	(shown above)
env-259	JCVI_SCAF_1101668675783	+	151	206	**RNA**→ **RNA**→ COG2308 (COG2308)? (GOS3708)→
env-773	JCVI_SCAF_1101668675783	+	218	272	(shown above)
env-260	JCVI_SCAF_1096628250203	+	1038	1093	**RNA**→ **RNA**→ DUF404 (pfam04169)? (GOS3708)→
env-777	JCVI_SCAF_1096628250203	+	1105	1159	(shown above)
env-261	JCVI_SCAF_1096628241071	+	306	361	**RNA**→ **RNA**→ ? (GOS3708)→ DUF407 (pfam04174)? (GOS3708)→ ? (GOS5926)→
env-770	JCVI_SCAF_1096628241071	+	373	427	(shown above)
env-262	JCVI_SCAF_1101668215945	+	583	638	**RNA**→ **RNA**→ COG2308 (COG2308)? (GOS3708)→
env-768	JCVI_SCAF_1101668215945	+	650	704	(shown above)
env-263	JCVI_SCAF_1101667075072	-	442	387	**RNA**→ **RNA**→ ? (GOS3708)→
env-765	JCVI_SCAF_1101667075072	-	375	321	(shown above)
env-264	JCVI_SCAF_1096627211009	+	119	175	**RNA**→ **RNA**→ COG2308 (COG2308)? (GOS3708)→ DUF403 (pfam04168)? (GOS5926)→ ? (GOS7780)→
env-764	JCVI_SCAF_1096627211009	+	188	242	(shown above)
env-265	JCVI_SCAF_1096628054229	+	1441	1497	**RNA**→
env-266	JCVI_SCAF_1096628291340	+	1013	1069	**RNA**→ **RNA**→ DUF404 (pfam04169)? (GOS3708)→
env-769	JCVI_SCAF_1096628291340	+	1082	1137	(shown above)
env-267	JCVI_SCAF_1096627416880	-	187	138	**RNA**→
env-268	JCVI_SCAF_1101668062376	-	888	839	**RNA**→ ? (GOS13256619)→
env-269	JCVI_SCAF_1096627756218	-	386	328	**RNA**→ **RNA**→ ? (GOS13558286)→
env-976	JCVI_SCAF_1096627756218	-	327	273	(shown above)
env-270	JCVI_SCAF_1096626869722	-	2914	2867	**RNA**→ ? (GOS11024366)→
env-271	JCVI_SCAF_1096627290494	+	1348	1400	**RNA**→ ? (GOS11024366)→ ? (GOS17606)→
env-272	JCVI_SCAF_1096626047404	-	761	710	**RNA**→ ? (GOS11024366)→ ? (GOS17606)→
env-273	JCVI_SCAF_1096627759027	+	365	417	**RNA**→ ? (GOS9335)→
env-274	JCVI_SCAF_1096626989394	-	1525	1476	**RNA**→ ? (GOS7096)→ Ammonium_transp (pfam00909)? (GOS46)→
env-275	JCVI_SCAF_1096627002061	+	997	1046	**RNA**→ ? (GOS7096)→
env-276	JCVI_SCAF_1101668320571	-	642	593	**RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-277	JCVI_SCAF_1101668214678	-	810	760	**RNA**→ ? (GOS7096)→
env-278	JCVI_SCAF_1096627335734	-	1227	1177	**RNA**→ ? (GOS7096)→ ? (GOS46)→
env-279	JCVI_SCAF_1096627120494	-	1415	1365	**RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-280	JCVI_SCAF_1096626996647	+	1032	1082	**RNA**→ ? (GOS46)→

env-281	JCVI_SCAF_1101668211792	+	943	992	**RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-282	JCVI_SCAF_1096628091091	-	932	883	**RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-283	JCVI_SCAF_1096628008861	-	402	353	**RNA**→ ? (GOS46)→
env-284	JCVI_SCAF_1101668192843	+	1411	1460	**RNA**→
env-285	JCVI_SCAF_1101668308867	-	463	414	**RNA**→ ? (GOS46)→
env-286	JCVI_SCAF_1101668656990	-	1041	992	**RNA**→ ? (GOS13158272)→
env-287	JCVI_SCAF_1096627819889	+	838	888	**RNA**→ ←? (GOS11850984)
env-288	JCVI_SCAF_1101668076121	-	563	512	**RNA**→ ? (GOS11024366)→
env-289	JCVI_SCAF_1096627566718	+	593	642	**RNA**→ ←? (GOS11991836)
env-290	JCVI_SCAF_1101668022020	+	1588	1637	**RNA**→
env-291	JCVI_SCAF_1101668413655	-	143	87	**RNA**→
env-292	JCVI_SCAF_1101668393179	-	635	579	**RNA**→ **RNA**→ DUF1255 (pfam06865)? (GOS13738)→
env-409	JCVI_SCAF_1101668393179	-	575	523	(shown above)
env-293	JCVI_SCAF_1096626927317	-	1042	987	**RNA**→ ? (GOS7096)→ ? (GOS46)→
env-294	JCVI_SCAF_1096627790420	+	687	736	**RNA**→
env-295	JCVI_SCAF_1101667581639	+	756	811	**RNA**→
env-296	JCVI_SCAF_1096627742567	+	1735	1787	**RNA**→
env-297	JCVI_SCAF_1096627782227	+	200	252	**RNA**→ ←? (GOS751)
env-298	JCVI_SCAF_1101668002618	-	1363	1313	**RNA**→ GltD (COG0493)? (GOS556)→
env-300	JCVI_SCAF_1096626594489	-	634	585	**RNA**→ ? (GOS13211008)→
env-301	JCVI_SCAF_1096627562360	+	1091	1142	**RNA**→ ? (GOS13211008)→
env-302	JCVI_SCAF_1096627516209	-	472	420	**RNA**→
env-303	JCVI_SCAF_1096628132418	-	1122	1072	**RNA**→ ←RNR_1 (cd01679)? (GOS835)
env-304	JCVI_SCAF_1096628387932	-	1057	1006	**RNA**→ ? (GOS11394952)→
Ama-2-2	NZ_ABYK01000024.1	+	20386	20477	**RNA**→ hypo (cons)→
env-305	JCVI_SCAF_1096626523639	+	170	229	**RNA**→ ? (GOS46)→
env-306	JCVI_SCAF_1096627245403	+	995	1052	**RNA**→
env-307	JCVI_SCAF_1096627242765	+	323	380	**RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-308	JCVI_SCAF_1096626352565	+	97	155	**RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-309	JCVI_SCAF_1096627269361	+	2209	2267	**RNA**→ ? (GOS46)→
env-310	JCVI_SCAF_1101668496424	-	1342	1284	**RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-311	JCVI_SCAF_1096627920177	+	1148	1205	**RNA**→ ? (GOS46)→
env-312	JCVI_SCAF_1096627185428	+	174	231	**RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-313	JCVI_SCAF_1096627921092	-	950	893	**RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-314	JCVI_SCAF_1096626975337	+	1416	1466	**RNA**→
env-315	JCVI_SCAF_1096627305184	-	16054	16004	**RNA**→ ? (GOS7096)→ Ammonium_transp (pfam00909)? (GOS46)→
env-316	JCVI_SCAF_1101667025116	-	678	626	**RNA**→
env-317	JCVI_SCAF_1096627283537	+	219	271	**RNA**→ ←GRX_GRXb_1_3_like (cd03418)? (GOS6441)
env-318	JCVI_SCAF_1096627413455	+	688	741	**RNA**→ ←GRX_GRXb_1_3_like (cd03418)? (GOS6441)
env-319	JCVI_SCAF_1101668011470	-	495	442	**RNA**→
env-320	JCVI_SCAF_1096628133333	-	1848	1798	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-622	JCVI_SCAF_1096628133333	-	1785	1735	(shown above)
Mch-1-1	NZ_ABRS01000006.1	+	124108	124201	**RNA**→
env-321	JCVI_SCAF_1101668540739	+	1195	1247	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-464	JCVI_SCAF_1101668540739	+	1260	1316	(shown above)
env-323	JCVI_SCAF_1096627272613	+	1766	1818	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-722	JCVI_SCAF_1096627272613	+	1821	1873	(shown above)
env-324	JCVI_SCAF_1101667113996	+	449	509	**RNA**→
env-325	JCVI_SCAF_1096626977263	+	1498	1558	**RNA**→
env-326	JCVI_SCAF_1101667462135	+	777	837	**RNA**→
env-327	JCVI_SCAF_1096628132198	+	660	720	**RNA**→ **RNA**→ ←? (GOS670)
env-403	JCVI_SCAF_1096628132198	+	721	773	(shown above)

env-328	JCVI_SCAF_1096627233108	+	1994	2054	**RNA**→ **RNA**→ adh_short (pfam00106)? (GOS49)→
env-404	JCVI_SCAF_1096627233108	+	2055	2107	(shown above)
env-329	JCVI_SCAF_1096627286437	-	622	572	**RNA**→ Trp tRNA→
env-330	JCVI_SCAF_1096627044280	+	1485	1532	**RNA**→ ? (GOS11178686)→
env-331	JCVI_SCAF_1096627055670	-	1668	1621	**RNA**→ ? (GOS11178686)→
env-332	JCVI_SCAF_1096627715998	-	1448	1401	**RNA**→ ? (GOS11228318)→ ? (GOS3516)→
env-333	JCVI_SCAF_1096627711419	-	221	174	**RNA**→ ? (GOS11200173)→
env-334	JCVI_SCAF_1096626891626	+	646	693	**RNA**→ ? (GOS11200173)→
env-335	JCVI_SCAF_1101668758342	-	720	674	**RNA**→ ? (GOS11975097)→
env-336	JCVI_SCAF_1096627604588	-	1556	1510	**RNA**→ ? (GOS11975097)→
env-337	JCVI_SCAF_1096627531764	+	796	842	**RNA**→ ←NUDIX (pfam00293)? (GOS10442)
env-339	JCVI_SCAF_1096627654090	-	1202	1155	**RNA**→ ? (GOS17794)→ Thy1 (pfam02511)? (GOS5954)→
env-340	JCVI_SCAF_1096627512910	+	1532	1580	**RNA**→
env-341	JCVI_SCAF_1101668291752	-	885	837	**RNA**→ ←? (GOS18102)
env-342	JCVI_SCAF_1096627157832	+	532	580	**RNA**→ ←? (GOS18102)
env-343	JCVI_SCAF_1096627748003	-	672	624	**RNA**→ ←? (GOS18102)
env-344	JCVI_SCAF_1101668328462	-	1206	1153	**RNA**→ ←? (GOS12151970)
env-345	JCVI_SCAF_1101667158833	-	205	156	**RNA**→
env-346	JCVI_SCAF_1096627336793	+	937	987	**RNA**→ COG2308 (COG2308)? (GOS3708)→
env-347	JCVI_SCAF_1096628369533	+	950	1000	**RNA**→ COG2308 (COG2308)? (GOS3708)→
env-348	JCVI_SCAF_1096627559540	-	874	824	**RNA**→ COG2308 (COG2308)? (GOS3708)→
env-349	JCVI_SCAF_1101668744738	+	1535	1585	**RNA**→ ? (GOS3708)→
env-350	JCVI_SCAF_1101667188030	-	330	280	**RNA**→ ? (GOS3708)→
env-351	JCVI_SCAF_1096627322980	+	250	300	**RNA**→ COG2308 (COG2308)? (GOS3708)→
env-352	JCVI_SCAF_1096627323201	-	354	304	**RNA**→ ? (GOS3708)→
env-353	JCVI_SCAF_1096627053207	-	1530	1480	**RNA**→ COG2308 (COG2308)? (GOS3708)→
env-354	JCVI_SCAF_1096626959977	+	1198	1249	**RNA**→ COG2308 (COG2308)? (GOS3708)→
env-355	JCVI_SCAF_1101667592311	+	517	567	**RNA**→ ? (GOS3708)→
env-357	JCVI_SCAF_1096627972247	-	340	287	**RNA**→
env-358	JCVI_SCAF_1096628022325	-	472	419	**RNA**→
env-360	JCVI_SCAF_1101668702892	-	244	191	**RNA**→
env-361	JCVI_SCAF_1096627384535	-	1195	1142	**RNA**→ ? (GOS17069)→ ? (GOS8954)→
env-362	JCVI_SCAF_1096627130067	+	916	970	**RNA**→ ? (GOS17069)→ ? (GOS8954)→ AA_permease (pfam00324)? (GOS492)→ COG1834 (COG1834)? (GOS595)→
env-363	JCVI_SCAF_1101668056955	-	301	248	**RNA**→ ? (GOS17069)→
env-364	JCVI_SCAF_1096626858292	-	1778	1725	**RNA**→ ? (GOS17069)→
env-365	JCVI_SCAF_1096627413087	-	1421	1368	**RNA**→ ? (GOS17069)→
env-366	JCVI_SCAF_1101667014695	+	286	339	**RNA**→ ? (GOS17069)→
env-367	JCVI_SCAF_1096627487976	+	204	257	**RNA**→ ? (GOS17069)→
env-368	JCVI_SCAF_1096626877146	-	219	166	**RNA**→ ? (GOS17069)→
env-369	JCVI_SCAF_1096627463955	+	1097	1151	**RNA**→ ? (GOS17069)→
env-370	JCVI_SCAF_1101668064939	+	1100	1154	**RNA**→ ? (GOS17069)→
env-371	JCVI_SCAF_1096626909425	+	1626	1680	**RNA**→ ? (GOS17069)→
env-372	JCVI_SCAF_1096626051831	-	894	840	**RNA**→ ? (GOS17069)→
env-373	JCVI_SCAF_1096626879926	+	1291	1345	**RNA**→ ? (GOS17069)→ ? (GOS8954)→
env-374	JCVI_SCAF_1096627370562	+	2127	2180	**RNA**→
env-375	JCVI_SCAF_1101668402120	+	609	662	**RNA**→ ? (GOS17069)→
env-376	JCVI_SCAF_1101668541395	-	988	934	**RNA**→ ? (GOS17069)→ ? (GOS8954)→
env-377	JCVI_SCAF_1096628365560	-	546	496	**RNA**→ DUF404 (pfam04169)? (GOS3708)→
env-378	JCVI_SCAF_1096627615818	-	380	330	**RNA**→ ? (GOS3708)→
env-379	JCVI_SCAF_1101668356959	-	451	399	**RNA**→ **RNA**→ ? (GOS11573861)→ ? (GOS13107772)→
env-955	JCVI_SCAF_1101668356959	-	505	455	(shown above)

env-380	JCVI_SCAF_1101667123518	-	549	491	**RNA**→ ? (GOS1994)→
env-381	JCVI_SCAF_1101667122340	-	200	142	**RNA**→
env-382	JCVI_SCAF_1096628013946	+	195	253	**RNA**→ Glu_synth_NTN (pfam04897)? (GOS1994)→ ? (GOS1994)→
env-383	JCVI_SCAF_1096628396058	+	378	436	**RNA**→ ? (GOS1994)→ Glu_syn_central (pfam04898)? (GOS1994)→
env-384	JCVI_SCAF_1096626999726	-	1001	941	**RNA**→ Glu_synth_NTN (pfam04897)? (GOS1994)→
env-385	JCVI_SCAF_1096628126728	-	859	801	**RNA**→ Glu_synth_NTN (pfam04897)? (GOS1994)→
env-386	JCVI_SCAF_1096627270317	+	936	994	**RNA**→
env-387	JCVI_SCAF_1101667147916	-	652	594	**RNA**→ GltB (COG0067)? (GOS1994)→
env-388	JCVI_SCAF_1101667352379	+	276	335	**RNA**→ ←? (GOS12022033)
env-389	JCVI_SCAF_1101668386454	-	822	764	**RNA**→ Glu_synth_NTN (pfam04897)? (GOS1994)→
env-390	JCVI_SCAF_1096627088737	+	2303	2362	**RNA**→ ←? (GOS12022033)
env-391	JCVI_SCAF_1101667021630	+	125	171	**RNA**→ ? (GOS3516)→
env-401	JCVI_SCAF_1096628130299	-	505	453	**RNA**→
env-405	JCVI_SCAF_1101667113996	+	510	562	**RNA**→
env-406	JCVI_SCAF_1096626977263	+	1559	1611	**RNA**→
env-408	JCVI_SCAF_1096627701349	-	338	286	**RNA**→
env-410	JCVI_SCAF_1096627284410	+	49	101	**RNA**→ ←COG2510 (COG2510)? (GOS183)
env-411	JCVI_SCAF_1101668413655	-	83	30	**RNA**→
env-416	JCVI_SCAF_1096627346029	+	1035	1087	**RNA**→
env-418	JCVI_SCAF_1101668024329	-	742	690	**RNA**→ **RNA**→ ? (GOS11573861)→ ? (GOS13107772)→
env-959	JCVI_SCAF_1101668024329	-	797	746	(shown above)
env-421	JCVI_SCAF_1101667001906	-	253	201	**RNA**→ **RNA**→ ? (GOS11573861)→ ? (GOS13107772)→
env-958	JCVI_SCAF_1101667001906	-	308	257	(shown above)
env-423	JCVI_SCAF_1101668437343	-	212	160	**RNA**→ **RNA**→ ? (GOS11573861)→
env-953	JCVI_SCAF_1101668437343	-	266	216	(shown above)
env-424	JCVI_SCAF_1096626860522	-	137	85	**RNA**→ **RNA**→ ? (GOS11573861)→
env-960	JCVI_SCAF_1096626860522	-	192	141	(shown above)
env-428	JCVI_SCAF_1096624997047	+	754	807	**RNA**→
env-429	JCVI_SCAF_1101667032616	-	384	331	**RNA**→
env-430	JCVI_SCAF_1101668006615	+	997	1050	**RNA**→
env-432	JCVI_SCAF_1096627845062	+	1447	1500	**RNA**→
env-435	JCVI_SCAF_1096627619435	-	276	222	**RNA**→
env-436	JCVI_SCAF_1096627961386	-	1817	1764	**RNA**→ ←? (GOS12151970)
env-437	JCVI_SCAF_1096626240129	+	315	365	**RNA**→ ? (GOS11024366)→
env-438	JCVI_SCAF_1096627844076	-	1577	1528	**RNA**→ ? (GOS11024366)→
env-439	JCVI_SCAF_1096627565517	+	1650	1699	**RNA**→
env-440	JCVI_SCAF_1096628108266	+	1541	1590	**RNA**→
env-441	JCVI_SCAF_1096627507057	+	518	567	**RNA**→ ? (GOS13158272)→
env-442	JCVI_SCAF_1096627061092	+	510	560	**RNA**→ ? (GOS13158272)→
env-443	JCVI_SCAF_1096627550666	-	801	750	**RNA**→ ? (GOS13211008)→
env-444	JCVI_SCAF_1096627299677	+	1134	1183	**RNA**→ ? (GOS11024366)→
env-445	JCVI_SCAF_1096626885721	-	417	369	**RNA**→
env-446	JCVI_SCAF_1096627726691	-	1191	1140	**RNA**→ ←? (GOS8297)
env-447	JCVI_SCAF_1101668687868	+	846	898	**RNA**→ ? (GOS13158272)→ ? (GOS11876226)→
env-448	JCVI_SCAF_1101668525502	+	910	962	**RNA**→ ? (GOS556)→
env-449	JCVI_SCAF_1096627708113	-	1736	1684	**RNA**→ Glu_synth_NTN (pfam04897)? (GOS1994)→ Glu_synth_NTN (pfam04897)? (GOS1994)→
env-450	JCVI_SCAF_1096627462536	-	503	454	**RNA**→ ? (GOS11024366)→
env-451	JCVI_SCAF_1096627231623	-	4403	4350	**RNA**→ ? (GOS13271387)→ ? (GOS15375)→ ? (GOS425)→ ? (GOS13543291)→ ? (GOS13277091)→
env-452	JCVI_SCAF_1101668746662	+	55	108	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-480	JCVI_SCAF_1101668746662	+	121	171	(shown above)
env-453	JCVI_SCAF_1101667545639	+	674	734	**RNA**→
env-454	JCVI_SCAF_1101667572791	-	778	719	**RNA**→ **RNA**→ ? (GOS13524402)→

env-136	JCVI_SCAF_1101667572791	-	662	610	(shown above)
env-455	JCVI_SCAF_1096628142664	+	122	172	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→ FGGY_C (pfam02782)? (GOS667)→
env-479	JCVI_SCAF_1096628142664	+	184	240	(shown above)
env-456	JCVI_SCAF_1096627311201	-	233	180	**RNA**→
env-460	JCVI_SCAF_1096628145455	-	329	269	**RNA**→
env-463	JCVI_SCAF_1096627336041	+	980	1038	**RNA**→
env-473	JCVI_SCAF_1096626899658	-	290	240	**RNA**→ ? (GOS11024366)→
Csp-3-2	NZ_ABOY01000004.1	+	130098	130156	**RNA**→ hypo→
Csp-5-1	NZ_ABVE01000009.1	-	60680	60623	**RNA**→ hypo→
env-475	JCVI_SCAF_1096628054355	+	1444	1498	**RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-476	JCVI_SCAF_1096627800145	+	712	766	**RNA**→ Gln-synt_C (pfam00120)? (GOS3709)→
env-477	JCVI_SCAF_1096627301637	+	101	156	**RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→ Gln-synt_C (pfam00120)? (GOS3709)→
env-478	JCVI_SCAF_1101668463055	-	781	728	**RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-482	JCVI_SCAF_1096628050596	+	170	226	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→ GlpK (COG0554)? (GOS667)→
env-551	JCVI_SCAF_1096628050596	+	238	293	(shown above)
env-483	JCVI_SCAF_1101668291475	-	902	852	**RNA**→ **RNA**→ **RNA**→ GlnK (COG0347)? (GOS11223)→ ? (GOS46)→
env-552	JCVI_SCAF_1101668291475	-	781	728	(shown above)
env-591	JCVI_SCAF_1101668291475	-	850	800	(shown above)
env-484	JCVI_SCAF_1101668446099	+	427	477	**RNA**→ **RNA**→ **RNA**→ ←? (GOS11767479)
env-561	JCVI_SCAF_1101668446099	+	548	601	(shown above)
env-597	JCVI_SCAF_1101668446099	+	479	529	(shown above)
env-485	JCVI_SCAF_1101668338074	+	1280	1330	**RNA**→ **RNA**→ ? (GOS1720)→
env-608	JCVI_SCAF_1101668338074	+	1332	1382	(shown above)
env-486	JCVI_SCAF_1101668396049	+	456	506	**RNA**→ ? (GOS1720)→
env-487	JCVI_SCAF_1096627081033	-	377	327	**RNA**→ ? (GOS1720)→
env-488	JCVI_SCAF_1096627218726	-	505	455	**RNA**→ ? (GOS1720)→
env-489	JCVI_SCAF_1096627333888	-	1125	1075	**RNA**→ **RNA**→ **RNA**→ ? (GOS7096)→ GlnK (COG0347)? (GOS11223)→
env-559	JCVI_SCAF_1096627333888	-	1004	951	(shown above)
env-586	JCVI_SCAF_1096627333888	-	1073	1023	(shown above)
env-490	JCVI_SCAF_1096626668463	+	687	737	**RNA**→ ? (GOS1720)→
env-491	JCVI_SCAF_1101668334768	-	239	189	**RNA**→
env-492	JCVI_SCAF_1101668422867	+	753	803	**RNA**→
env-493	JCVI_SCAF_1096627077036	+	876	926	**RNA**→
env-494	JCVI_SCAF_1096627178110	+	193	243	**RNA**→ **RNA**→ **RNA**→ ? (GOS7096)→
env-553	JCVI_SCAF_1096627178110	+	314	367	(shown above)
env-590	JCVI_SCAF_1096627178110	+	245	295	(shown above)
env-495	JCVI_SCAF_1096627333808	+	55	105	**RNA**→ CirA (COG1629)? (GOS1720)→
env-496	JCVI_SCAF_1096626272304	+	1314	1364	**RNA**→
env-497	JCVI_SCAF_1096628400923	+	180	230	**RNA**→ ? (GOS1720)→
env-498	JCVI_SCAF_1096628400923	+	872	922	**RNA**→ ? (GOS1720)→
env-499	JCVI_SCAF_1096627064304	-	864	814	**RNA**→ **RNA**→ **RNA**→ ←? (GOS508)
env-554	JCVI_SCAF_1096627064304	-	743	690	(shown above)
env-587	JCVI_SCAF_1096627064304	-	812	762	(shown above)
env-500	JCVI_SCAF_1096627174225	-	828	778	**RNA**→ **RNA**→ **RNA**→ ←? (GOS11767479)
env-557	JCVI_SCAF_1096627174225	-	707	654	(shown above)
env-589	JCVI_SCAF_1096627174225	-	776	726	(shown above)
env-501	JCVI_SCAF_1096627258443	-	991	941	**RNA**→ **RNA**→ **RNA**→ ? (GOS7096)→
env-564	JCVI_SCAF_1096627258443	-	870	817	(shown above)
env-593	JCVI_SCAF_1096627258443	-	939	889	(shown above)
env-502	JCVI_SCAF_1096627383906	-	2641	2591	**RNA**→ **RNA**→ **RNA**→ GlnK (COG0347)? (GOS11223)→ Ammonium_transp (pfam00909)? (GOS46)→
env-560	JCVI_SCAF_1096627383906	-	2520	2467	(shown above)
env-594	JCVI_SCAF_1096627383906	-	2589	2539	(shown above)

env-503	JCVI_SCAF_1096627798697	+	474	524	**RNA**→ ? (GOS1720)→
env-504	JCVI_SCAF_1101668388415	+	358	408	**RNA**→ ? (GOS1720)→
env-505	JCVI_SCAF_1096626976286	-	262	212	**RNA**→
env-506	JCVI_SCAF_1096627091101	-	439	389	**RNA**→ ? (GOS1720)→
env-507	JCVI_SCAF_1096627227070	+	1138	1188	**RNA**→ **RNA**→ ? (GOS1720)→
env-607	JCVI_SCAF_1096627227070	+	1190	1239	(shown above)
env-508	JCVI_SCAF_1096627338034	-	287	237	**RNA**→
env-509	JCVI_SCAF_1101667364359	-	763	713	**RNA**→ ? (GOS1720)→
env-510	JCVI_SCAF_1096627232138	+	777	827	**RNA**→ **RNA**→ **RNA**→ ? (GOS7096)→
env-566	JCVI_SCAF_1096627232138	+	898	951	(shown above)
env-599	JCVI_SCAF_1096627232138	+	829	879	(shown above)
env-511	JCVI_SCAF_1096627390365	+	9959	10009	**RNA**→ **RNA**→ **RNA**→ ? (GOS7096)→ Ammonium_transp (pfam00909)? (GOS46)→
env-565	JCVI_SCAF_1096627390365	+	10080	10133	(shown above)
env-598	JCVI_SCAF_1096627390365	+	10011	10061	(shown above)
env-512	JCVI_SCAF_1096627057394	-	530	480	**RNA**→ ? (GOS1720)→
env-513	JCVI_SCAF_1096627267139	-	1443	1393	**RNA**→ CirA (COG1629)? (GOS1720)→
env-514	JCVI_SCAF_1096627242975	+	25	75	**RNA**→ ? (GOS1720)→
env-515	JCVI_SCAF_1096628270050	-	1115	1065	**RNA**→ ? (GOS1720)→ COG2236 (COG2236)? (GOS1256)→
env-516	JCVI_SCAF_1101668439660	-	56	6	**RNA**→
env-517	JCVI_SCAF_1101668175207	-	758	708	**RNA**→ ? (GOS1720)→
env-518	JCVI_SCAF_1096628099339	-	103	51	**RNA**→
env-519	JCVI_SCAF_1096627989985	-	1664	1614	**RNA**→ ←Iap (COG2234)? (GOS508)
env-520	JCVI_SCAF_1096626976286	-	141	91	**RNA**→
env-521	JCVI_SCAF_1096627863875	-	692	642	**RNA**→
env-522	JCVI_SCAF_1096626984467	-	4421	4371	**RNA**→ **RNA**→ ligand_gated_channel (cd01347)? (GOS1720)→ COG2236 (COG2236)? (GOS1256)→ ERG3 (COG3000)? (GOS1535)→
env-603	JCVI_SCAF_1096626984467	-	4369	4319	(shown above)
env-523	JCVI_SCAF_1096627381158	+	6925	6975	**RNA**→
env-524	JCVI_SCAF_1096627390317	-	28329	28279	**RNA**→ **RNA**→ TonB_dep_Rec (pfam00593)? (GOS1720)→
env-604	JCVI_SCAF_1096627390317	-	28277	28227	(shown above)
env-525	JCVI_SCAF_1096626974517	+	2937	2987	**RNA**→
env-526	JCVI_SCAF_1101668718028	+	47	97	**RNA**→ **RNA**→ ligand_gated_channel (cd01347)? (GOS1720)→
env-609	JCVI_SCAF_1101668718028	+	99	149	(shown above)
env-527	JCVI_SCAF_1096627726183	-	1053	1003	**RNA**→ **RNA**→ ? (GOS1720)→
env-601	JCVI_SCAF_1096627726183	-	1001	952	(shown above)
env-528	JCVI_SCAF_1096627848963	-	1730	1680	**RNA**→ **RNA**→ ? (GOS1720)→ ? (GOS1720)→
env-605	JCVI_SCAF_1096627848963	-	1678	1628	(shown above)
env-529	JCVI_SCAF_1101667557540	-	610	560	**RNA**→ **RNA**→ ? (GOS1720)→
env-602	JCVI_SCAF_1101667557540	-	558	508	(shown above)
env-530	JCVI_SCAF_1096627719576	+	259	310	**RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-534	JCVI_SCAF_1096628124606	-	171	120	**RNA**→
env-535	JCVI_SCAF_1096627565604	-	468	417	**RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-536	JCVI_SCAF_1101667158538	+	508	559	**RNA**→
env-541	JCVI_SCAF_1096628148493	-	249	198	**RNA**→
env-543	JCVI_SCAF_1096627216387	-	268	217	**RNA**→
env-544	JCVI_SCAF_1101667446355	-	261	210	**RNA**→
env-549	JCVI_SCAF_1096627297493	+	599	649	**RNA**→ ? (GOS11024366)→
env-550	JCVI_SCAF_1096626904160	+	1227	1277	**RNA**→ ? (GOS11024366)→
env-555	JCVI_SCAF_1096627991035	-	811	758	**RNA**→ ←? (GOS508)
env-556	JCVI_SCAF_1096626968786	+	10464	10517	**RNA**→ ? (GOS7096)→ Ammonium_transp (pfam00909)? (GOS46)→
env-558	JCVI_SCAF_1096627186440	+	34	87	**RNA**→ ? (GOS7096)→ ? (GOS46)→
env-563	JCVI_SCAF_1101668334768	-	118	65	**RNA**→

env-567	JCVI_SCAF_1101668422867	+	874	927	**RNA**→
env-568	JCVI_SCAF_1101667446355	-	333	281	**RNA**→
env-569	JCVI_SCAF_1101667133470	+	235	287	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-545	JCVI_SCAF_1101667133470	+	307	358	(shown above)
env-570	JCVI_SCAF_1101667488496	+	155	207	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-546	JCVI_SCAF_1101667488496	+	227	278	(shown above)
env-571	JCVI_SCAF_1096627224787	+	1276	1328	**RNA**→
env-572	JCVI_SCAF_1096626633360	+	258	310	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-538	JCVI_SCAF_1096626633360	+	330	381	(shown above)
env-573	JCVI_SCAF_1096628136195	+	46	98	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-539	JCVI_SCAF_1096628136195	+	118	169	(shown above)
env-574	JCVI_SCAF_1096626983740	-	1458	1406	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-542	JCVI_SCAF_1096626983740	-	1386	1335	(shown above)
env-575	JCVI_SCAF_1096627056656	-	650	598	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-532	JCVI_SCAF_1096627056656	-	578	527	(shown above)
env-576	JCVI_SCAF_1096628148493	-	321	269	**RNA**→
env-577	JCVI_SCAF_1101668310916	-	654	602	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-531	JCVI_SCAF_1101668310916	-	582	531	(shown above)
env-578	JCVI_SCAF_1101667158538	+	436	488	**RNA**→
env-579	JCVI_SCAF_1096627216387	-	340	288	**RNA**→
env-580	JCVI_SCAF_1096627258043	-	1761	1709	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-548	JCVI_SCAF_1096627258043	-	1689	1638	(shown above)
env-581	JCVI_SCAF_1096627204158	+	479	531	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-537	JCVI_SCAF_1096627204158	+	551	602	(shown above)
env-582	JCVI_SCAF_1096627221663	-	1617	1565	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-540	JCVI_SCAF_1096627221663	-	1545	1494	(shown above)
env-583	JCVI_SCAF_1096627258887	+	56	108	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-533	JCVI_SCAF_1096627258887	+	128	179	(shown above)
env-584	JCVI_SCAF_1096628196398	-	1325	1273	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→ MutY (COG1194)? (GOS3797)→
env-547	JCVI_SCAF_1096628196398	-	1253	1202	(shown above)
env-585	JCVI_SCAF_1096628124606	-	242	190	**RNA**→
env-588	JCVI_SCAF_1096627077036	+	928	978	**RNA**→
env-592	JCVI_SCAF_1101668422867	+	805	855	**RNA**→
env-595	JCVI_SCAF_1101667249050	+	33	83	**RNA**→ **RNA**→ GlnK (COG0347)? (GOS11223)→
env-562	JCVI_SCAF_1101667249050	+	102	155	(shown above)
env-596	JCVI_SCAF_1096626976286	-	210	160	**RNA**→
env-600	JCVI_SCAF_1101668334768	-	187	137	**RNA**→
env-606	JCVI_SCAF_1096626974517	+	2989	3039	**RNA**→
env-610	JCVI_SCAF_1101668543518	+	1195	1245	**RNA**→
env-612	JCVI_SCAF_1096628136788	+	1002	1054	**RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-613	JCVI_SCAF_1096628097878	-	244	194	**RNA**→
env-615	JCVI_SCAF_1101668388409	+	940	990	**RNA**→
env-616	JCVI_SCAF_1096628217934	+	1464	1514	**RNA**→
env-617	JCVI_SCAF_1101668687189	-	214	163	**RNA**→
env-619	JCVI_SCAF_1101667034399	-	204	153	**RNA**→
env-624	JCVI_SCAF_1096628164470	+	9	58	**RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→ DNA_gyraseB_C (pfam00986)TOPRIM_TopoIIA_GyrB (cd03366)? (GOS1203)→ ? (GOS667)→
env-626	JCVI_SCAF_1101668405913	-	100	45	**RNA**→
env-630	JCVI_SCAF_1096627273126	+	3289	3345	**RNA**→
env-633	JCVI_SCAF_1096627221975	+	1200	1255	**RNA**→
env-635	JCVI_SCAF_1101668311191	+	1225	1280	**RNA**→
env-637	JCVI_SCAF_1096627186513	+	16	72	**RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→

env-642	JCVI_SCAF_1096627791220	+	1500	1557	**RNA**→
env-644	JCVI_SCAF_1096628257329	+	1574	1626	**RNA**→
env-649	JCVI_SCAF_1101667249454	+	630	686	**RNA**→
env-650	JCVI_SCAF_1101667385586	-	220	164	**RNA**→
env-651	JCVI_SCAF_1096628127348	+	579	633	**RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-657	JCVI_SCAF_1096626672581	-	967	915	**RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-659	JCVI_SCAF_1096627068607	+	1309	1361	**RNA**→
env-660	JCVI_SCAF_1101668410571	+	978	1030	**RNA**→
env-661	JCVI_SCAF_1101668427728	-	162	110	**RNA**→
env-671	JCVI_SCAF_1101667267724	+	42	94	**RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-675	JCVI_SCAF_1101668706853	+	1347	1399	**RNA**→
env-676	JCVI_SCAF_1096627342462	+	890	942	**RNA**→
env-678	JCVI_SCAF_1096627106760	-	1180	1128	**RNA**→ Gln-synt_N (pfam03951)Gln-synt_C (pfam00120)? (GOS3709)→
env-682	JCVI_SCAF_1096627330902	-	286	234	**RNA**→
env-685	JCVI_SCAF_1096627053252	+	327	379	**RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-686	JCVI_SCAF_1096626950233	+	4828	4880	**RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-687	JCVI_SCAF_1096628121887	-	763	711	**RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-688	JCVI_SCAF_1096627389411	+	14975	15027	**RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-689	JCVI_SCAF_1096626894633	-	2916	2863	**RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-690	JCVI_SCAF_1101667358519	-	216	163	**RNA**→
env-691	JCVI_SCAF_1096628215343	-	1120	1064	**RNA**→ **RNA**→ ←RHOD (cd00158)? (GOS5736)
env-655	JCVI_SCAF_1096628215343	-	1045	993	(shown above)
env-692	JCVI_SCAF_1096627389867	-	1163	1107	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-667	JCVI_SCAF_1096627389867	-	1088	1036	(shown above)
env-693	JCVI_SCAF_1096627247510	+	1664	1720	**RNA**→ **RNA**→ ? (GOS3709)→
env-666	JCVI_SCAF_1096627247510	+	1739	1791	(shown above)
env-694	JCVI_SCAF_1101667459031	+	472	528	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-662	JCVI_SCAF_1101667459031	+	547	599	(shown above)
env-695	JCVI_SCAF_1101667517526	-	103	47	**RNA**→
env-696	JCVI_SCAF_1101667472665	-	456	400	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-663	JCVI_SCAF_1101667472665	-	381	329	(shown above)
env-697	JCVI_SCAF_1101667507305	+	111	167	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-664	JCVI_SCAF_1101667507305	+	186	238	(shown above)
env-698	JCVI_SCAF_1096626924930	-	1339	1283	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-665	JCVI_SCAF_1096626924930	-	1264	1212	(shown above)
env-699	JCVI_SCAF_1096627169202	+	76	132	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→ Ammonium_transp (pfam00909)? (GOS46)→
env-654	JCVI_SCAF_1096627169202	+	151	203	(shown above)
env-700	JCVI_SCAF_1096627390105	-	156571	156515	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→ ? (GOS7096)→ Ammonium_transp (pfam00909)? (GOS46)→
env-653	JCVI_SCAF_1096627390105	-	156496	156444	(shown above)
env-701	JCVI_SCAF_1096627068607	+	1238	1290	**RNA**→
env-702	JCVI_SCAF_1096627325435	+	1142	1194	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-658	JCVI_SCAF_1096627325435	+	1213	1265	(shown above)
env-703	JCVI_SCAF_1101668410571	+	907	959	**RNA**→
env-704	JCVI_SCAF_1096627254817	+	1235	1287	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→ Ammonium_transp (pfam00909)? (GOS46)→
env-668	JCVI_SCAF_1096627254817	+	1306	1358	(shown above)
env-705	JCVI_SCAF_1101667099104	-	733	681	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-669	JCVI_SCAF_1101667099104	-	662	610	(shown above)
env-706	JCVI_SCAF_1096626639935	-	677	625	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-670	JCVI_SCAF_1096626639935	-	606	554	(shown above)
env-707	JCVI_SCAF_1096627726331	-	1151	1099	**RNA**→ **RNA**→ ? (GOS3709)→ Gln-synt_C (pfam00120)? (GOS3709)→

env-683	JCVI_SCAF_1096627726331	-	1080	1028	(shown above)
env-708	JCVI_SCAF_1096627163966	-	1229	1175	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-673	JCVI_SCAF_1096627163966	-	1156	1104	(shown above)
env-709	JCVI_SCAF_1096627383705	+	2430	2484	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-679	JCVI_SCAF_1096627383705	+	2503	2555	(shown above)
env-710	JCVI_SCAF_1101668706853	+	1274	1328	**RNA**→
env-711	JCVI_SCAF_1096627074794	-	394	340	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-672	JCVI_SCAF_1096627074794	-	321	269	(shown above)
env-712	JCVI_SCAF_1096627072524	+	232	286	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-681	JCVI_SCAF_1096627072524	+	305	357	(shown above)
env-713	JCVI_SCAF_1096627342462	+	817	871	**RNA**→
env-714	JCVI_SCAF_1101668423585	+	101	155	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)Gln-synt_C (pfam00120)? (GOS3709)→
env-680	JCVI_SCAF_1101668423585	+	174	226	(shown above)
env-715	JCVI_SCAF_1101668427728	-	235	181	**RNA**→
env-716	JCVI_SCAF_1096627154817	-	933	879	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-656	JCVI_SCAF_1096627154817	-	860	808	(shown above)
env-717	JCVI_SCAF_1096626284530	+	365	419	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-677	JCVI_SCAF_1096626284530	+	438	490	(shown above)
env-718	JCVI_SCAF_1101668371062	-	719	665	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-674	JCVI_SCAF_1101668371062	-	646	594	(shown above)
env-719	JCVI_SCAF_1096627170973	+	1285	1339	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-684	JCVI_SCAF_1096627170973	+	1358	1410	(shown above)
env-720	JCVI_SCAF_1096627330902	-	359	305	**RNA**→
env-721	JCVI_SCAF_1096627170439	+	316	370	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-652	JCVI_SCAF_1096627170439	+	389	441	(shown above)
env-723	JCVI_SCAF_1096627481335	+	923	973	**RNA**→ ? (GOS11024366)→
env-724	JCVI_SCAF_1096627487807	-	481	431	**RNA**→ ? (GOS11024366)→
env-725	JCVI_SCAF_1101668160569	+	988	1038	**RNA**→ ? (GOS11024366)→
env-726	JCVI_SCAF_1096627301919	+	2284	2334	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-727	JCVI_SCAF_1096628289144	+	506	562	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→ ? (GOS3709)→ HATPase_c (pfam02518)? (GOS1203)→
env-242	JCVI_SCAF_1096628289144	+	575	629	(shown above)
env-728	JCVI_SCAF_1096627796989	-	1034	982	**RNA**→ Gln-synt_C (pfam00120)? (GOS3709)→
env-729	JCVI_SCAF_1101668736101	-	364	312	**RNA**→ **RNA**→ ? (GOS3709)→
env-241	JCVI_SCAF_1101668736101	-	299	245	(shown above)
env-730	JCVI_SCAF_1096627105110	-	738	686	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-641	JCVI_SCAF_1096627105110	-	673	617	(shown above)
env-731	JCVI_SCAF_1101667351140	+	314	366	**RNA**→
env-732	JCVI_SCAF_1096627074376	-	1301	1249	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-648	JCVI_SCAF_1096627074376	-	1236	1180	(shown above)
env-733	JCVI_SCAF_1101668455281	+	323	375	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-628	JCVI_SCAF_1101668455281	+	388	442	(shown above)
env-734	JCVI_SCAF_1096628404388	-	1094	1042	**RNA**→ ? (GOS3709)→ ? (GOS1203)→
env-735	JCVI_SCAF_1096628233403	+	1288	1340	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-639	JCVI_SCAF_1096628233403	+	1353	1407	(shown above)
env-736	JCVI_SCAF_1101668663684	+	582	634	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-631	JCVI_SCAF_1101668663684	+	647	701	(shown above)
env-737	JCVI_SCAF_1096627221975	+	1135	1187	**RNA**→
env-738	JCVI_SCAF_1096628061696	+	966	1018	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-632	JCVI_SCAF_1096628061696	+	1031	1086	(shown above)
env-739	JCVI_SCAF_1101668311191	+	1160	1212	**RNA**→
env-740	JCVI_SCAF_1096626914631	+	2396	2448	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-634	JCVI_SCAF_1096626914631	+	2461	2516	(shown above)

env-741	JCVI_SCAF_1096627309681	+	662	714	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-629	JCVI_SCAF_1096627309681	+	727	783	(shown above)
env-742	JCVI_SCAF_1096628161808	+	645	697	**RNA**→ TopoII_Trans_DNA_gyrase (cd00822)? (GOS1203)→
env-743	JCVI_SCAF_1101668405913	-	165	113	**RNA**→
env-744	JCVI_SCAF_1096627273126	+	3224	3276	**RNA**→
env-745	JCVI_SCAF_1101667455822	-	393	341	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-636	JCVI_SCAF_1101667455822	-	328	273	(shown above)
env-746	JCVI_SCAF_1101668413483	+	47	99	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-638	JCVI_SCAF_1101668413483	+	112	169	(shown above)
env-747	JCVI_SCAF_1096627255997	+	362	414	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-627	JCVI_SCAF_1096627255997	+	427	482	(shown above)
env-748	JCVI_SCAF_1096627916950	+	1828	1880	**RNA**→
env-749	JCVI_SCAF_1096627791220	+	1435	1487	**RNA**→
env-750	JCVI_SCAF_1096627848169	-	858	806	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)? (GOS3709)→
env-640	JCVI_SCAF_1096627848169	-	793	737	(shown above)
env-751	JCVI_SCAF_1101668461281	+	113	164	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)Gln-synt_C (pfam00120)? (GOS3709)→
env-469	JCVI_SCAF_1101668461281	+	177	237	(shown above)
env-752	JCVI_SCAF_1101667385586	-	284	233	**RNA**→
env-753	JCVI_SCAF_1096627051225	-	1880	1826	**RNA**→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
env-472	JCVI_SCAF_1096627051225	-	1813	1757	(shown above)
env-754	JCVI_SCAF_1101667246615	+	262	316	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-471	JCVI_SCAF_1101667246615	+	329	384	(shown above)
env-755	JCVI_SCAF_1096627199003	+	1217	1271	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-466	JCVI_SCAF_1096627199003	+	1284	1340	(shown above)
env-756	JCVI_SCAF_1101667394109	+	530	584	**RNA**→ **RNA**→ ? (GOS3709)→
env-468	JCVI_SCAF_1101667394109	+	597	653	(shown above)
env-757	JCVI_SCAF_1096627582676	-	448	394	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-470	JCVI_SCAF_1096627582676	-	381	326	(shown above)
env-758	JCVI_SCAF_1096628160185	-	419	367	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-467	JCVI_SCAF_1096628160185	-	354	296	(shown above)
env-759	JCVI_SCAF_1101667249454	+	565	617	**RNA**→
env-760	JCVI_SCAF_1096627336041	+	916	967	**RNA**→
env-761	JCVI_SCAF_1096628008570	+	319	370	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-465	JCVI_SCAF_1096628008570	+	383	440	(shown above)
Csp-5-2	NZ_ABVE01000002.1	+	596451	596507	**RNA**→ hypo→ hypo→
env-762	JCVI_SCAF_1101668688918	+	1279	1332	**RNA**→
env-767	JCVI_SCAF_1096627197994	-	6317	6263	**RNA**→ COG2308 (COG2308)? (GOS3708)→
env-772	JCVI_SCAF_1096627631062	-	234	180	**RNA**→
env-778	JCVI_SCAF_1096626085716	-	159	105	**RNA**→
env-780	JCVI_SCAF_1096627605943	+	1559	1613	**RNA**→
env-781	JCVI_SCAF_1096628054229	+	1510	1567	**RNA**→
env-782	JCVI_SCAF_1096628228853	-	85	30	**RNA**→
env-783	JCVI_SCAF_1096628111374	-	929	873	**RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-784	JCVI_SCAF_1101667498360	-	203	149	**RNA**→
env-787	JCVI_SCAF_1096627862779	-	241	187	**RNA**→
env-788	JCVI_SCAF_1096626959183	+	50	104	**RNA**→ Ammonium_transp (pfam00909)? (GOS46)→ ? (GOS18681)→
env-795	JCVI_SCAF_1096626483611	+	27	81	**RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-799	JCVI_SCAF_1096627190635	+	977	1031	**RNA**→
env-806	JCVI_SCAF_1096626713916	-	274	220	**RNA**→
env-807	JCVI_SCAF_1096626993499	-	68	14	**RNA**→
env-812	JCVI_SCAF_1096627981427	+	1607	1660	**RNA**→
env-815	JCVI_SCAF_1101667256724	+	614	667	**RNA**→

env-817	JCVI_SCAF_1096628059617	-	189	136	**RNA**→
env-823	JCVI_SCAF_1101668527119	+	8	61	**RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-832	JCVI_SCAF_1096626584800	+	785	838	**RNA**→
env-839	JCVI_SCAF_1096627075773	-	134	79	**RNA**→
env-841	JCVI_SCAF_1096626119229	+	789	844	**RNA**→
env-844	JCVI_SCAF_1101667550644	-	261	206	**RNA**→
env-845	JCVI_SCAF_1101668352403	-	258	203	**RNA**→
env-847	JCVI_SCAF_1101668657873	-	1345	1290	**RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-848	JCVI_SCAF_1096628149509	+	1624	1678	**RNA**→
env-849	JCVI_SCAF_1096626330522	+	568	622	**RNA**→
env-854	JCVI_SCAF_1101668436807	+	59	114	**RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-859	JCVI_SCAF_1101668352707	+	501	555	**RNA**→
env-862	JCVI_SCAF_1101668401086	+	2	55	**RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-863	JCVI_SCAF_1096628118709	+	534	588	**RNA**→ **RNA**→ ? (GOS11907)→ ? (GOS18681)→
env-808	JCVI_SCAF_1096628118709	+	601	655	(shown above)
env-864	JCVI_SCAF_1096626496426	+	527	581	**RNA**→ **RNA**→ ? (GOS46)→
env-763	JCVI_SCAF_1096626496426	+	594	648	(shown above)
env-865	JCVI_SCAF_1096628133447	+	728	782	**RNA**→ ? (GOS5495)→
env-866	JCVI_SCAF_1101668063903	-	1488	1436	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-157	JCVI_SCAF_1101668063903	-	1428	1374	(shown above)
env-867	JCVI_SCAF_1101668053145	-	363	311	**RNA**→
env-868	JCVI_SCAF_1096628206963	+	150	204	**RNA**→ **RNA**→ ? (GOS46)→
env-159	JCVI_SCAF_1096628206963	+	212	268	(shown above)
env-869	JCVI_SCAF_1101667078320	+	833	886	**RNA**→
env-870	JCVI_SCAF_1096628145455	-	404	342	**RNA**→
env-871	JCVI_SCAF_1096628271645	+	1415	1477	**RNA**→ **RNA**→ ? (GOS46)→
env-462	JCVI_SCAF_1096628271645	+	1490	1550	(shown above)
env-872	JCVI_SCAF_1096628206933	-	83	20	**RNA**→
env-873	JCVI_SCAF_1096627231643	+	780	842	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→ ? (GOS11907)→
env-461	JCVI_SCAF_1096627231643	+	855	915	(shown above)
env-874	JCVI_SCAF_1096627250235	+	403	465	**RNA**→ **RNA**→ ? (GOS46)→
env-459	JCVI_SCAF_1096627250235	+	478	538	(shown above)
env-875	JCVI_SCAF_1096627113946	+	1271	1334	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-458	JCVI_SCAF_1096627113946	+	1347	1407	(shown above)
env-876	JCVI_SCAF_1096627075773	-	201	147	**RNA**→
env-877	JCVI_SCAF_1096627862779	-	308	254	**RNA**→
env-878	JCVI_SCAF_1096627306209	+	1858	1912	**RNA**→ **RNA**→ ? (GOS46)→
env-840	JCVI_SCAF_1096627306209	+	1925	1980	(shown above)
env-879	JCVI_SCAF_1101667498360	-	272	216	**RNA**→
env-880	JCVI_SCAF_1096627033251	-	2161	2107	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-846	JCVI_SCAF_1096627033251	-	2094	2039	(shown above)
env-881	JCVI_SCAF_1101668403691	-	993	938	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-852	JCVI_SCAF_1101668403691	-	925	870	(shown above)
env-882	JCVI_SCAF_1101668688918	+	1213	1266	**RNA**→
env-883	JCVI_SCAF_1096627072670	-	1706	1652	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-838	JCVI_SCAF_1096627072670	-	1639	1584	(shown above)
env-884	JCVI_SCAF_1096627129480	-	869	815	**RNA**→ **RNA**→ ? (GOS46)→
env-796	JCVI_SCAF_1096627129480	-	802	747	(shown above)
env-885	JCVI_SCAF_1096627999640	-	486	432	**RNA**→ **RNA**→ ? (GOS46)→
env-829	JCVI_SCAF_1096627999640	-	419	366	(shown above)
env-886	JCVI_SCAF_1101667075817	+	322	377	**RNA**→ **RNA**→ ? (GOS46)→
env-791	JCVI_SCAF_1101667075817	+	390	444	(shown above)

env-887	JCVI_SCAF_1096628385710	-	926	871	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-802	JCVI_SCAF_1096628385710	-	858	804	(shown above)
env-888	JCVI_SCAF_1096627531788	+	955	1010	**RNA**→ **RNA**→ ? (GOS46)→
env-803	JCVI_SCAF_1096627531788	+	1023	1077	(shown above)
env-889	JCVI_SCAF_1096628018808	-	1349	1294	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-801	JCVI_SCAF_1096628018808	-	1281	1227	(shown above)
env-890	JCVI_SCAF_1101668208772	-	1425	1370	**RNA**→ **RNA**→ ? (GOS46)→ Ammonium_transp (pfam00909)? (GOS46)→
env-804	JCVI_SCAF_1101668208772	-	1357	1303	(shown above)
env-891	JCVI_SCAF_1101667531032	+	476	531	**RNA**→ **RNA**→ ? (GOS46)→
env-798	JCVI_SCAF_1101667531032	+	544	598	(shown above)
env-892	JCVI_SCAF_1096627190635	+	909	964	**RNA**→
env-893	JCVI_SCAF_1096628294854	-	1051	996	**RNA**→ ? (GOS46)→
env-894	JCVI_SCAF_1101668756945	-	699	644	**RNA**→ **RNA**→ ? (GOS46)→
env-800	JCVI_SCAF_1101668756945	-	631	577	(shown above)
env-895	JCVI_SCAF_1096626441944	-	488	433	**RNA**→ **RNA**→ ? (GOS46)→
env-797	JCVI_SCAF_1096626441944	-	420	366	(shown above)
env-896	JCVI_SCAF_1101668352403	-	325	271	**RNA**→
env-897	JCVI_SCAF_1096627843270	-	1200	1145	**RNA**→ **RNA**→ ? (GOS46)→ Ammonium_transp (pfam00909)? (GOS46)→
env-827	JCVI_SCAF_1096627843270	-	1132	1077	(shown above)
env-898	JCVI_SCAF_1101668343701	-	587	532	**RNA**→ **RNA**→ ? (GOS46)→
env-828	JCVI_SCAF_1101668343701	-	519	464	(shown above)
env-899	JCVI_SCAF_1096626934800	-	1992	1937	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→ ? (GOS18681)→
env-793	JCVI_SCAF_1096626934800	-	1924	1870	(shown above)
env-900	JCVI_SCAF_1096628265077	+	467	522	**RNA**→ **RNA**→ ? (GOS46)→ ? (GOS11907)→
env-842	JCVI_SCAF_1096628265077	+	535	589	(shown above)
env-901	JCVI_SCAF_1096626119229	+	721	776	**RNA**→
env-902	JCVI_SCAF_1101668708247	+	472	527	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-825	JCVI_SCAF_1101668708247	+	540	595	(shown above)
env-903	JCVI_SCAF_1096628002280	-	1701	1646	**RNA**→ **RNA**→ ? (GOS46)→ Ammonium_transp (pfam00909)? (GOS46)→
env-792	JCVI_SCAF_1096628002280	-	1633	1579	(shown above)
env-904	JCVI_SCAF_1101667517323	+	190	245	**RNA**→ **RNA**→ ? (GOS46)→
env-786	JCVI_SCAF_1101667517323	+	258	312	(shown above)
env-905	JCVI_SCAF_1096627208704	-	1223	1168	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-850	JCVI_SCAF_1096627208704	-	1155	1100	(shown above)
env-906	JCVI_SCAF_1096627051375	-	438	383	**RNA**→ **RNA**→ ? (GOS46)→
env-794	JCVI_SCAF_1096627051375	-	370	316	(shown above)
env-907	JCVI_SCAF_1101667415248	+	839	894	**RNA**→
env-908	JCVI_SCAF_1096627050834	+	232	287	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-856	JCVI_SCAF_1096627050834	+	299	354	(shown above)
env-909	JCVI_SCAF_1096627735853	+	1475	1530	**RNA**→ **RNA**→ ? (GOS46)→
env-789	JCVI_SCAF_1096627735853	+	1543	1597	(shown above)
env-910	JCVI_SCAF_1101667550644	-	328	273	**RNA**→
env-911	JCVI_SCAF_1096628393673	+	1097	1152	**RNA**→ **RNA**→ ? (GOS46)→
env-858	JCVI_SCAF_1096628393673	+	1164	1218	(shown above)
env-912	JCVI_SCAF_1096627576638	-	908	853	**RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-913	JCVI_SCAF_1096626993499	-	2177	2122	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→ ? (GOS46)→ ? (GOS11907)→
env-975	JCVI_SCAF_1096626993499	-	2109	2053	(shown above)
env-914	JCVI_SCAF_1096628097735	+	786	841	**RNA**→ **RNA**→ DsbDgamma (cd02953)? (GOS2888)→ PRX_family (cd02971)? (GOS7685)→
env-855	JCVI_SCAF_1096628097735	+	853	908	(shown above)
env-915	JCVI_SCAF_1096627863652	-	1257	1202	**RNA**→ **RNA**→ ? (GOS46)→ Ammonium_transp (pfam00909)? (GOS46)→
env-843	JCVI_SCAF_1096627863652	-	1190	1135	(shown above)
env-916	JCVI_SCAF_1096628149509	+	1557	1612	**RNA**→

env-917	JCVI_SCAF_1096627589645	+	625	680	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-805	JCVI_SCAF_1096627589645	+	693	748	(shown above)
env-918	JCVI_SCAF_1096626330522	+	501	556	**RNA**→
env-919	JCVI_SCAF_1096626713916	-	342	287	**RNA**→
env-920	JCVI_SCAF_1096628234023	-	771	716	**RNA**→ **RNA**→ ? (GOS46)→
env-853	JCVI_SCAF_1096628234023	-	703	648	(shown above)
env-921	JCVI_SCAF_1096626993499	-	136	81	**RNA**→
env-922	JCVI_SCAF_1101668352707	+	434	488	**RNA**→
env-923	JCVI_SCAF_1096628240258	+	330	384	**RNA**→ **RNA**→ ? (GOS46)→ Ammonium_transp (pfam00909)? (GOS46)→ ? (GOS11907)→
env-851	JCVI_SCAF_1096628240258	+	397	452	(shown above)
env-924	JCVI_SCAF_1096627033251	-	120	65	**RNA**→
env-925	JCVI_SCAF_1101668390717	-	744	690	**RNA**→ **RNA**→ ? (GOS46)→
env-785	JCVI_SCAF_1101668390717	-	677	623	(shown above)
env-926	JCVI_SCAF_1096627842115	+	121	175	**RNA**→ **RNA**→ ? (GOS46)→ Ammonium_transp (pfam00909)? (GOS46)→
env-790	JCVI_SCAF_1096627842115	+	188	242	(shown above)
env-927	JCVI_SCAF_1096627385082	-	1854	1800	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-826	JCVI_SCAF_1096627385082	-	1787	1732	(shown above)
env-928	JCVI_SCAF_1101667384348	+	210	263	**RNA**→ **RNA**→ ? (GOS46)→
env-831	JCVI_SCAF_1101667384348	+	276	329	(shown above)
env-929	JCVI_SCAF_1101668458381	-	1094	1040	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-824	JCVI_SCAF_1101668458381	-	1027	974	(shown above)
env-930	JCVI_SCAF_1101668526306	+	294	346	**RNA**→ **RNA**→ ? (GOS46)→ Ammonium_transp (pfam00909)? (GOS46)→
env-813	JCVI_SCAF_1101668526306	+	359	412	(shown above)
env-931	JCVI_SCAF_1096626886824	-	1270	1217	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-836	JCVI_SCAF_1096626886824	-	1204	1151	(shown above)
env-932	JCVI_SCAF_1101668089644	+	634	689	**RNA**→ **RNA**→ ? (GOS46)→
env-837	JCVI_SCAF_1101668089644	+	702	755	(shown above)
env-933	JCVI_SCAF_1101668172147	+	691	744	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-811	JCVI_SCAF_1101668172147	+	757	810	(shown above)
env-934	JCVI_SCAF_1096627981427	+	1541	1594	**RNA**→
env-935	JCVI_SCAF_1101668490928	+	316	369	**RNA**→ **RNA**→ ? (GOS46)→
env-810	JCVI_SCAF_1101668490928	+	382	436	(shown above)
env-936	JCVI_SCAF_1096626584800	+	719	772	**RNA**→
env-937	JCVI_SCAF_1096628118232	-	1044	991	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-822	JCVI_SCAF_1096628118232	-	978	925	(shown above)
env-938	JCVI_SCAF_1101667404391	-	784	729	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-818	JCVI_SCAF_1101667404391	-	716	663	(shown above)
env-939	JCVI_SCAF_1101667508463	+	405	460	**RNA**→ **RNA**→ ? (GOS46)→
env-821	JCVI_SCAF_1101667508463	+	473	526	(shown above)
env-940	JCVI_SCAF_1096626545992	+	147	202	**RNA**→ **RNA**→ ? (GOS46)→
env-819	JCVI_SCAF_1096626545992	+	215	268	(shown above)
env-941	JCVI_SCAF_1096627867123	-	1524	1471	**RNA**→ **RNA**→ ? (GOS46)→ Ammonium_transp (pfam00909)? (GOS46)→
env-814	JCVI_SCAF_1096627867123	-	1458	1405	(shown above)
env-942	JCVI_SCAF_1101668443020	+	480	533	**RNA**→ **RNA**→ ? (GOS46)→
env-816	JCVI_SCAF_1101668443020	+	546	599	(shown above)
env-943	JCVI_SCAF_1101667256724	+	548	601	**RNA**→
env-944	JCVI_SCAF_1101668674516	-	1312	1259	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-809	JCVI_SCAF_1101668674516	-	1246	1192	(shown above)
env-945	JCVI_SCAF_1096626918183	-	1646	1593	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-820	JCVI_SCAF_1096626918183	-	1580	1526	(shown above)
env-946	JCVI_SCAF_1096628059617	-	255	202	**RNA**→
env-947	JCVI_SCAF_1096626939830	-	446	392	**RNA**→ **RNA**→ ? (GOS46)→

env-861	JCVI_SCAF_1096626939830	-	379	325	(shown above)
env-948	JCVI_SCAF_1101668501247	+	589	639	**RNA**→ **RNA**→ Gln-synt_N (pfam03951)? (GOS3709)→
env-474	JCVI_SCAF_1101668501247	+	653	707	(shown above)
env-949	JCVI_SCAF_1096626883472	+	3602	3653	**RNA**→ **RNA**→ ? (GOS13107772)→
env-420	JCVI_SCAF_1096626883472	+	3657	3709	(shown above)
env-950	JCVI_SCAF_1096626906110	+	1003	1053	**RNA**→ **RNA**→ ? (GOS13107772)→
env-415	JCVI_SCAF_1096626906110	+	1057	1109	(shown above)
env-951	JCVI_SCAF_1096627508400	-	1351	1301	**RNA**→ **RNA**→ ←? (GOS11570888)
env-422	JCVI_SCAF_1096627508400	-	1297	1245	(shown above)
env-952	JCVI_SCAF_1096627501889	-	673	623	**RNA**→ **RNA**→ ? (GOS13107772)→ ? (GOS1322)→
env-426	JCVI_SCAF_1096627501889	-	619	567	(shown above)
env-954	JCVI_SCAF_1101668558217	-	393	343	**RNA**→ **RNA**→ ? (GOS13107772)→
env-407	JCVI_SCAF_1101668558217	-	339	287	(shown above)
env-956	JCVI_SCAF_1096627966170	+	20	70	**RNA**→ **RNA**→ ? (GOS13107772)→
env-414	JCVI_SCAF_1096627966170	+	74	126	(shown above)
env-957	JCVI_SCAF_1101668027335	-	1151	1101	**RNA**→ **RNA**→ ? (GOS13107772)→
env-425	JCVI_SCAF_1101668027335	-	1097	1045	(shown above)
env-961	JCVI_SCAF_1101668255758	+	107	157	**RNA**→ **RNA**→ ? (GOS13107772)→ WcaG (COG0451)? (GOS1322)→ ? (GOS1322)→
env-419	JCVI_SCAF_1101668255758	+	161	213	(shown above)
env-962	JCVI_SCAF_1096626922817	+	246	296	**RNA**→ **RNA**→ ? (GOS13107772)→
env-413	JCVI_SCAF_1096626922817	+	300	352	(shown above)
env-963	JCVI_SCAF_1101668702892	-	302	248	**RNA**→
env-964	JCVI_SCAF_1096628022325	-	530	476	**RNA**→
env-965	JCVI_SCAF_1096627972247	-	398	344	**RNA**→
env-966	JCVI_SCAF_1101667208217	-	603	549	**RNA**→ **RNA**→ ←? (GOS11837156)
env-356	JCVI_SCAF_1101667208217	-	545	492	(shown above)
env-967	JCVI_SCAF_1101668677538	+	255	308	**RNA**→ **RNA**→ ←RHOD (cd00158)? (GOS594)
env-359	JCVI_SCAF_1101668677538	+	312	365	(shown above)
env-968	JCVI_SCAF_1096627307718	-	53	3	**RNA**→
env-969	JCVI_SCAF_1096626921984	-	1062	1010	**RNA**→ **RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-299	JCVI_SCAF_1096626921984	-	998	944	(shown above)
env-970	JCVI_SCAF_1096628278002	-	1514	1460	**RNA**→ **RNA**→ ? (GOS46)→
env-857	JCVI_SCAF_1096628278002	-	1447	1392	(shown above)
env-973	JCVI_SCAF_1096628213122	+	1106	1165	**RNA**→ ? (GOS46)→
env-974	JCVI_SCAF_1096627385483	-	854	804	**RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)? (GOS3709)→
Csp-3-3	NZ_ABOY01000012.1	-	115931	115872	**RNA**→ hypo (cons)→
env-977	JCVI_SCAF_1096627331249	-	936	874	**RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
env-978	JCVI_SCAF_1096627120495	+	573	635	**RNA**→ Ammonium_transp (pfam00909)? (GOS46)→
Ssp-9-1	NC_000911.1	-	458952	458891	**RNA**→ hypo→
Ssp-2-1	NC_007775.1	-	931758	931667	**RNA**→ hypo→ hypo→
env-979	JCVI_SCAF_1101668755125	+	543	595	**RNA**→
Csp-6-1	NZ_ABLR01000009.1	-	86436	86371	**RNA**→ hypo→ hypo→
Csp-7-1	NZ_ABVI01000006.1	+	274338	274403	**RNA**→ hypo→ hypo→
env-980	JCVI_SCAF_1096627748669	-	1133	1082	**RNA**→ ←? (GOS13130752)
env-981	JCVI_SCAF_1096628002649	+	1629	1681	**RNA**→
env-982	JCVI_SCAF_1101668453019	-	747	695	**RNA**→
env-983	JCVI_SCAF_1096627821896	-	284	231	**RNA**→
env-984	JCVI_SCAF_1101668735976	-	934	881	**RNA**→ ←? (GOS13130752)
env-985	JCVI_SCAF_1101668552909	+	335	391	**RNA**→ **RNA**→ ←? (GOS10539)
env-402	JCVI_SCAF_1101668552909	+	395	447	(shown above)
env-986	JCVI_SCAF_1096628307813	+	193	244	**RNA**→ **RNA**→ ←? (GOS7137)
env-434	JCVI_SCAF_1096628307813	+	248	301	(shown above)

env-987	JCVI_SCAF_1096627845062	+	1392	1443	**RNA**→
env-988	JCVI_SCAF_1096628020530	-	922	871	**RNA**→ **RNA**→ ←? (GOS11827493)
env-433	JCVI_SCAF_1096628020530	-	867	814	(shown above)
env-989	JCVI_SCAF_1096627602663	+	810	861	**RNA**→ **RNA**→ ←COG2849 (COG2849)? (GOS7137)
env-431	JCVI_SCAF_1096627602663	+	865	918	(shown above)
Ssp-3-2	NZ_ABRV01000019.1	+	87833	87915	**RNA**→
env-990	JCVI_SCAF_1096628130299	-	563	509	**RNA**→
env-991	JCVI_SCAF_1096628012613	+	1600	1648	**RNA**→ ? (GOS15286)→
env-992	JCVI_SCAF_1101668728438	+	18	66	**RNA**→ Thy1 (pfam02511)? (GOS5954)→
env-993	JCVI_SCAF_1096626951452	+	626	672	**RNA**→ ? (GOS15286)→
env-994	JCVI_SCAF_1101668721758	+	122	168	**RNA**→ ? (GOS15286)→
env-995	JCVI_SCAF_1101668732231	+	692	738	**RNA**→ ? (GOS17289)→ Thy1 (pfam02511)? (GOS5954)→
env-996	JCVI_SCAF_1096627334047	+	1519	1565	**RNA**→
env-997	JCVI_SCAF_1101668717320	+	178	224	**RNA**→ ? (GOS17289)→ MT-A70 (pfam05063)? (GOS9797)→
Pma-1-1	NC_008820.1	-	469994	469919	**RNA**→ hypo→ GlnK (COG0347)→
Pma-2-1	NC_005071.1	+	1564016	1564091	**RNA**→ extracellular link domain-containing protein→ hypo→ GlnK (COG0347)→
Ssp-4-1	NZ_AAUA01000001.1	+	657594	657669	**RNA**→ possible Extracellular link domain→

37.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 37.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *glnA* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

cd00158 (2) Rhodanese Homology Domain (RHOD); an alpha beta fold domain found duplicated in the rhodanese protein.
cd00822 (1) TopoIIA_Trans_DNA_gyrase: Transducer domain, having a ribosomal S5 domain 2-like fold, of the type found in proteins of the type IIA family of DNA topoisomerases similar to the B subunits of E.
cd01347 (3) TonB dependent/Ligand-Gated channels are created by a monomeric 22 strand (22,24) anti-parallel beta-barrel.
cd01679 (1) RNR, class I.
cd02953 (1) DsbD gamma family; DsbD gamma is the C-terminal periplasmic domain of the bacterial protein DsbD.
cd02971 (1) Peroxiredoxin (PRX) family; composed of the different classes of PRXs including many proteins originally known as bacterioferritin comigratory proteins (BCP), based on their electrophoretic mobility before their function was identified.
cd03366 (1) TOPRIM_TopoIIA_GyrB: topoisomerase-primase (TOPRIM) nucleotidyl transferase/hydrolase domain of the type found in proteins of the type IIA family of DNA topoisomerases similar to the Escherichia coli GyrB subunit.
cd03418 (2) Glutaredoxin (GRX) family, GRX bacterial class 1 and 3 (b_1_3)-like subfamily; composed of bacterial GRXs, approximately 10 kDa in size, and proteins containing a GRX or GRX-like domain.
COG0067 (1) Glutamate synthase domain 1 [Amino acid transport and metabolism]
COG0071 (1) Molecular chaperone (small heat shock protein) [Posttranslational modification, protein turnover, chaperones]
COG0149 (1) Triosephosphate isomerase [Carbohydrate transport and metabolism]
COG0187 (1) Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), B subunit [DNA replication, recombination, and repair]
COG0226 (1) ABC-type phosphate transport system, periplasmic component [Inorganic ion transport and metabolism]
COG0270 (1) Site-specific DNA methylase [DNA replication, recombination, and repair]
COG0341 (1) Preprotein translocase subunit SecF [Intracellular trafficking and secretion]
COG0347 (15) Nitrogen regulatory protein PII [Amino acid transport and metabolism]
COG0413 (1) Ketopantoate hydroxymethyltransferase [Coenzyme metabolism]
COG0451 (1) Nucleoside-diphosphate-sugar epimerases [Cell envelope biogenesis, outer membrane / Carbohydrate transport and metabolism]
COG0493 (1) NADPH-dependent glutamate synthase beta chain and related oxidoreductases [Amino acid transport and metabolism / General function prediction only]
COG0554 (2) Glycerol kinase [Energy production and conversion]
COG0596 (1) Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily) [General function prediction only]
COG0678 (3) Peroxiredoxin [Posttranslational modification, protein turnover, chaperones]
COG0828 (1) Ribosomal protein S21 [Translation, ribosomal structure and biogenesis]
COG1061 (1) DNA or RNA helicases of superfamily II [Transcription / DNA replication, recombination, and repair]
COG1070 (1) Sugar (pentulose and hexulose) kinases [Carbohydrate transport and metabolism]
COG1194 (1) A/G-specific DNA glycosylase [DNA replication, recombination, and repair]
COG1629 (2) Outer membrane receptor proteins, mostly Fe transport [Inorganic ion transport and metabolism]
COG1834 (1) N-Dimethylarginine dimethylaminohydrolase [Amino acid transport and

metabolism]
COG2227 (1) 2-polyprenyl-3-methyl-5-hydroxy-6-metoxy-1,4-benzoquinol methylase [Coenzyme metabolism]
COG2234 (1) Predicted aminopeptidases [General function prediction only]
COG2236 (2) Predicted phosphoribosyltransferases [General function prediction only]
COG2308 (16) Uncharacterized conserved protein [Function unknown]
COG2510 (1) Predicted membrane protein [Function unknown]
COG2849 (1) Uncharacterized protein conserved in bacteria [Function unknown]
COG3000 (1) Sterol desaturase [Lipid metabolism]
COG3128 (1) Uncharacterized iron-regulated protein [Function unknown]
COG5553 (1) Predicted metal-dependent enzyme of the double-stranded beta helix superfamily [General function prediction only]
GOS10442 (1) ?
GOS10539 (1) ?
GOS11024366 (20) ?
GOS11025 (4) ?
GOS11034 (1) ?
GOS11070785 (1) ?
GOS11075491 (1) ?
GOS1110 (1) ?
GOS11141882 (2) ?
GOS11178686 (2) ?
GOS11194950 (5) ?
GOS11195016 (1) ?
GOS11200173 (3) ?
GOS11223 (10) ?
GOS11228318 (1) ?
GOS11251 (4) ?
GOS11295941 (4) ?
GOS11377265 (5) ?
GOS11394952 (1) ?
GOS11421997 (2) ?
GOS11477096 (1) ?
GOS11531948 (1) ?
GOS11570888 (1) ?
GOS11573861 (5) ?
GOS11670331 (8) ?
GOS11672939 (1) ?
GOS11767479 (2) ?
GOS11827493 (1) ?
GOS11833740 (1) ?
GOS11837156 (1) ?
GOS11850984 (1) ?
GOS11876226 (1) ?
GOS11907 (6) ?
GOS11975097 (2) ?
GOS11991836 (1) ?
GOS12022033 (2) ?
GOS1203 (6) ?
GOS12056256 (1) ?
GOS12119065 (2) ?
GOS12151970 (2) ?
GOS12174553 (1) ?
GOS1219 (1) ?
GOS1256 (2) ?
GOS12637 (1) ?
GOS1288 (1) ?
GOS12901 (2) ?
GOS13107772 (11) ?
GOS13113871 (11) ?
GOS13130752 (2) ?
GOS13158272 (4) ?
GOS13211008 (3) ?
GOS1322 (3) ?
GOS13251683 (1) ?
GOS13256619 (1) ?
GOS13261206 (5) ?
GOS13271387 (2) ?
GOS13277091 (1) ?
GOS1330 (3) ?
GOS1339 (1) ?
GOS13524402 (1) ?
GOS13543291 (1) ?
GOS13558286 (1) ?
GOS13561 (2) ?
GOS13738 (1) ?
GOS14064195 (2) ?
GOS142 (1) ?
GOS1457 (1) ?
GOS14643050 (4) ?
GOS1508 (1) ?
GOS15286 (44) ?
GOS1535 (1) ?
GOS15375 (1) ?
GOS15599 (1) ?
GOS15797 (1) ?
GOS164 (1) ?
GOS16501 (7) ?
GOS17069 (15) ?
GOS1720 (27) ?
GOS17289 (5) ?
GOS17606 (2) ?
GOS17794 (5) ?
GOS18102 (3) ?
GOS183 (1) ?
GOS18681 (3) ?
GOS19514 (6) ?
GOS1994 (12) ?
GOS2888 (1) ?
GOS3516 (2) ?
GOS3708 (34) ?
GOS3709 (112) ?
GOS3797 (1) ?
GOS425 (1) ?

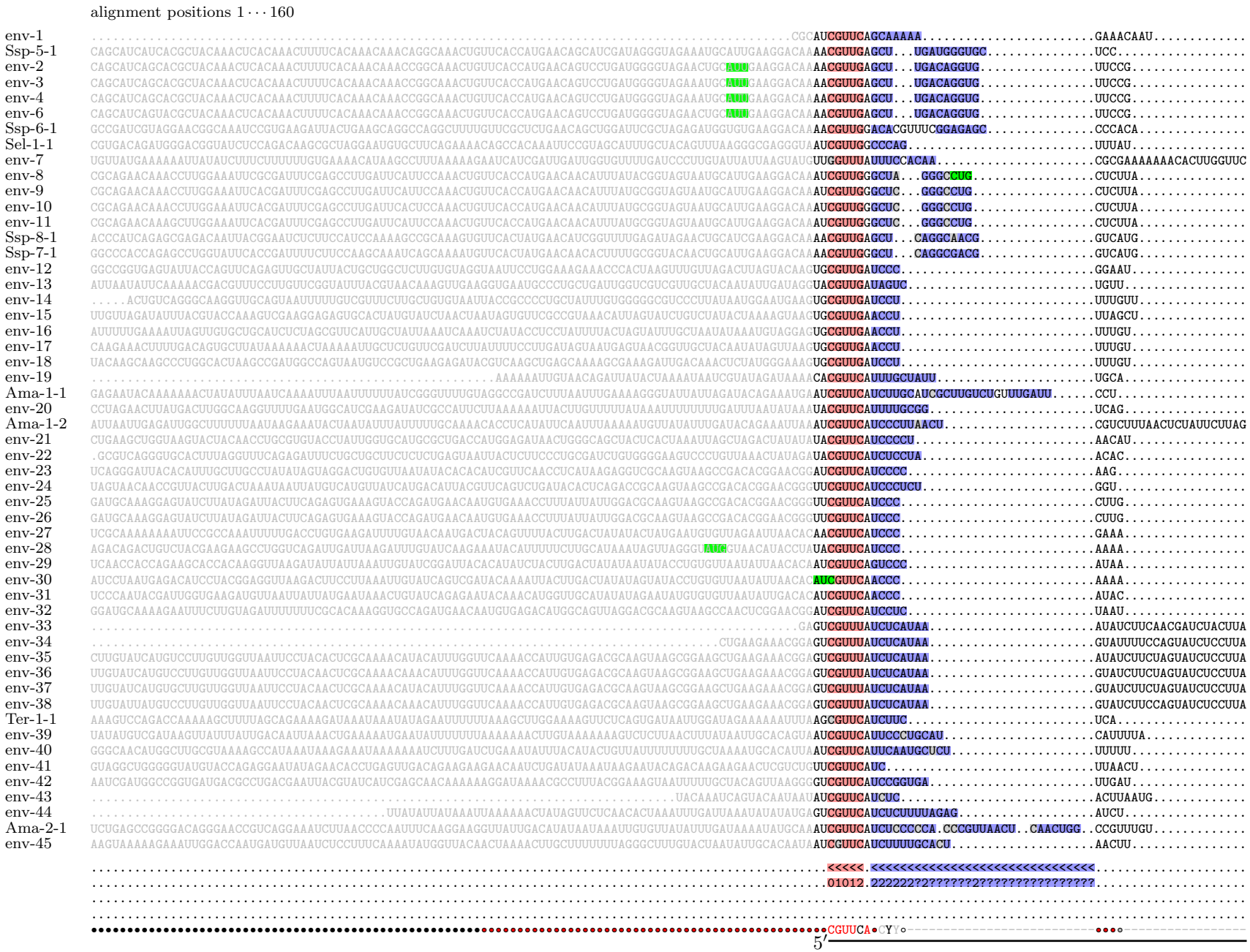
GOS431 (1) ?
GOS4325 (1) ?
GOS46 (117) ?
GOS49 (2) ?
GOS492 (1) ?
GOS508 (3) ?
GOS5495 (1) ?
GOS5559 (1) ?
GOS556 (2) ?
GOS5736 (1) ?
GOS5926 (4) ?
GOS594 (1) ?
GOS595 (1) ?
GOS5954 (8) ?
GOS606 (1) ?
GOS6163 (1) ?
GOS6286 (1) ?
GOS6441 (2) ?
GOS667 (5) ?
GOS670 (1) ?
GOS7096 (14) ?
GOS7137 (2) ?
GOS721 (1) ?
GOS7261 (1) ?
GOS7297 (1) ?
GOS7317 (7) ?
GOS7356 (1) ?
GOS751 (1) ?
GOS7556 (2) ?
GOS7685 (4) ?
GOS7780 (1) ?
GOS8297 (1) ?
GOS835 (1) ?
GOS8353 (1) ?
GOS8468 (4) ?
GOS8589 (1) ?
GOS8954 (4) ?
GOS9229 (1) ?
GOS9335 (2) ?
GOS9797 (2) ?
pfam00106 (2) short chain dehydrogenase.
pfam00120 (50) Glutamine synthetase, catalytic domain.
pfam00293 (1) NUDIX domain.
pfam00324 (1) Amino acid permease.
pfam00593 (1) TonB dependent receptor.
pfam00903 (1) Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily.
pfam00909 (63) Ammonium Transporter Family.
pfam00986 (1) DNA gyrase B subunit, carboxyl terminus.
pfam02511 (6) Thymidylate synthase complementing protein.
pfam02518 (1) Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase.
pfam02782 (1) FGGY family of carbohydrate kinases, C-terminal domain.
pfam03332 (4) Eukaryotic phosphomannomutase.
pfam03951 (98) Glutamine synthetase, beta-Grasp domain.
pfam04168 (2) Bacterial domain of unknown function (DUF403).
pfam04169 (4) Domain of unknown function (DUF404).
pfam04174 (1) Domain of unknown function (DUF407).
pfam04897 (7) Glutamate synthase amidotransferase domain.
pfam04898 (1) Glutamate synthase central domain.
pfam04970 (1) NC domain.
pfam05063 (1) MT-A70.
pfam06865 (1) Protein of unknown function (DUF1255).
smart00702 (1) Prolyl 4-hydroxylase alpha subunit homologues.

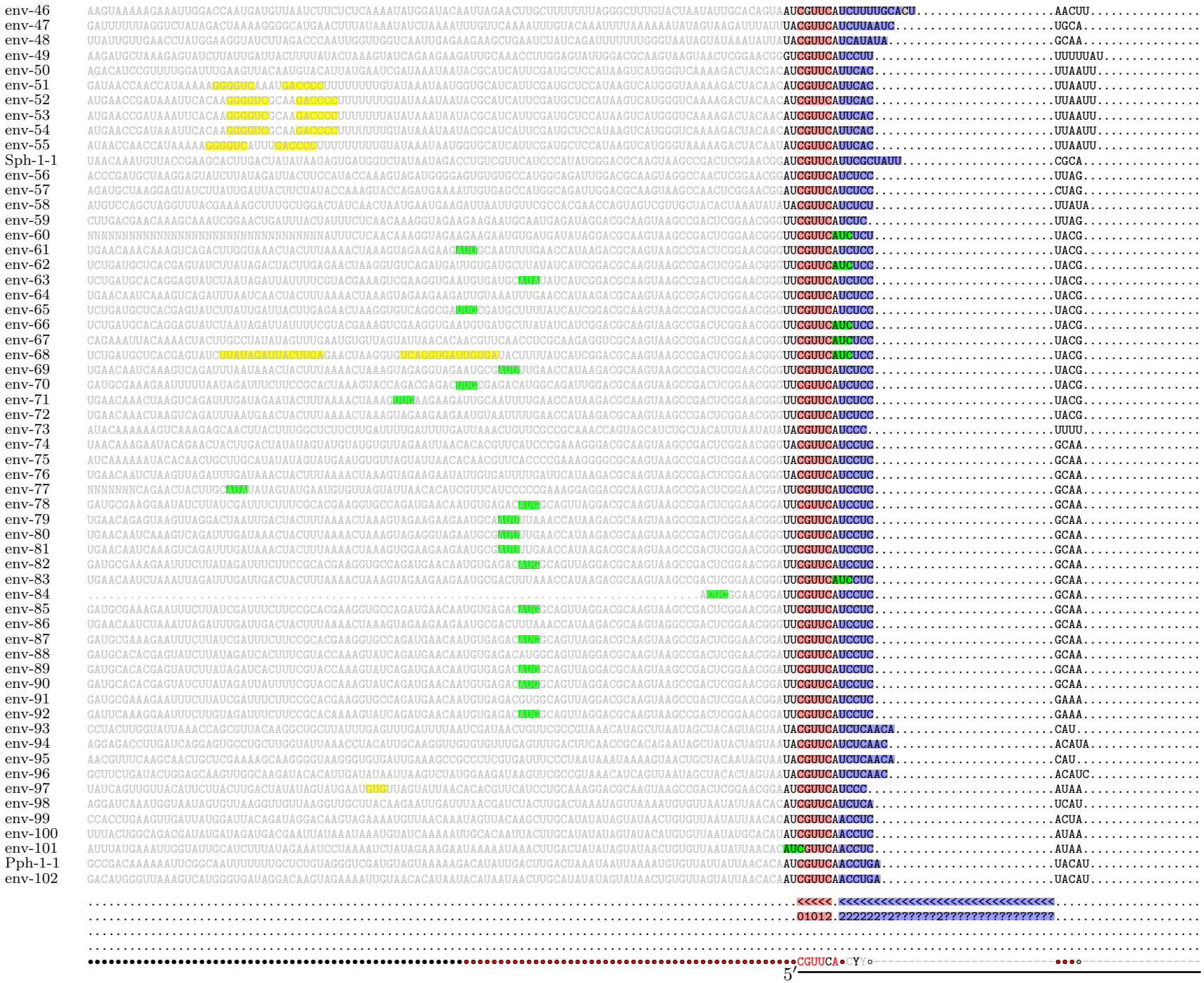
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Sel-2-1, env-224, env-498, env-5, env-540, env-546, env-570, env-664, env-665, env-670, env-697, env-698, env-706

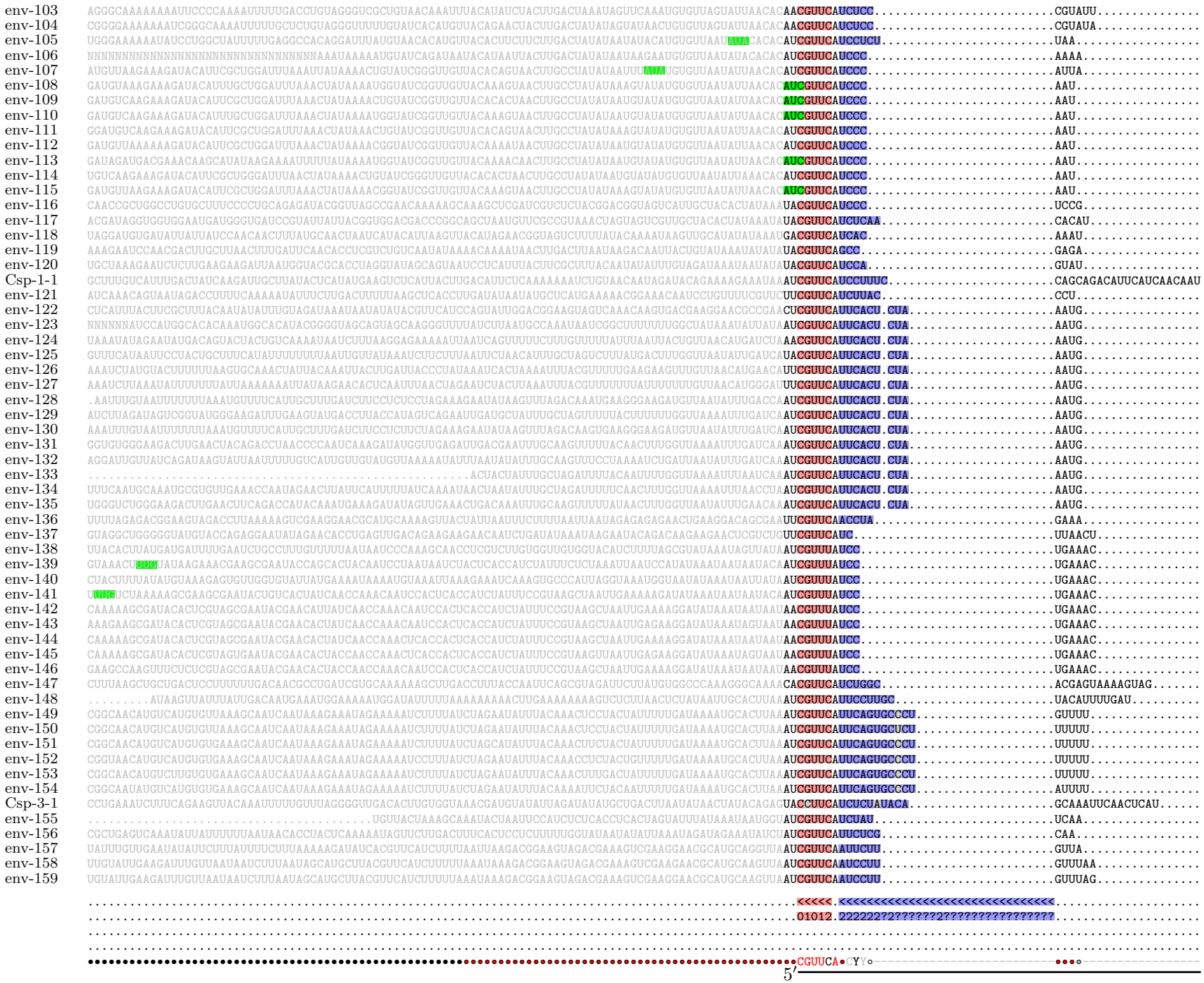
37.4 Multiple-sequence alignment

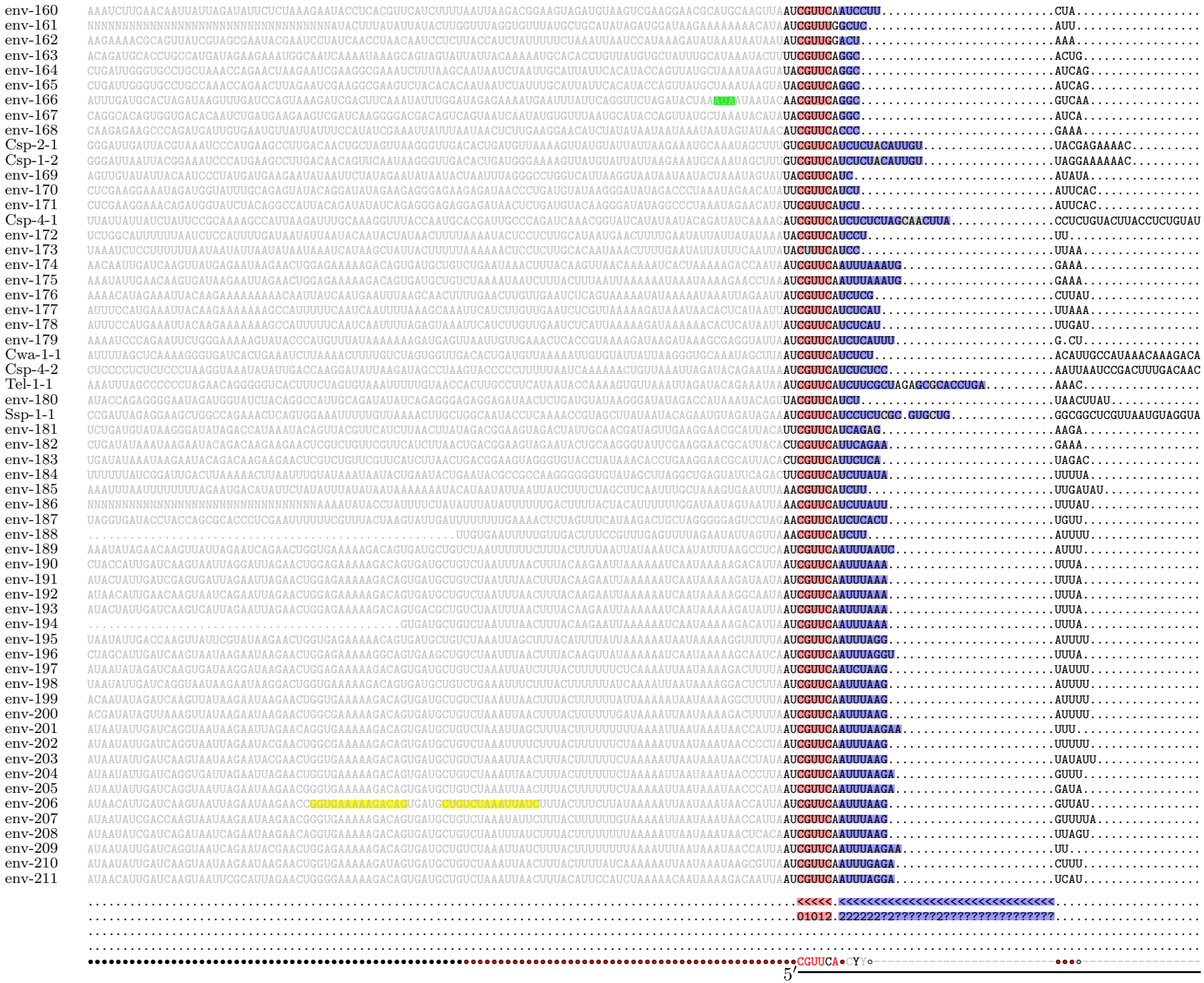
Each *glnA* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 37.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *glnA* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black

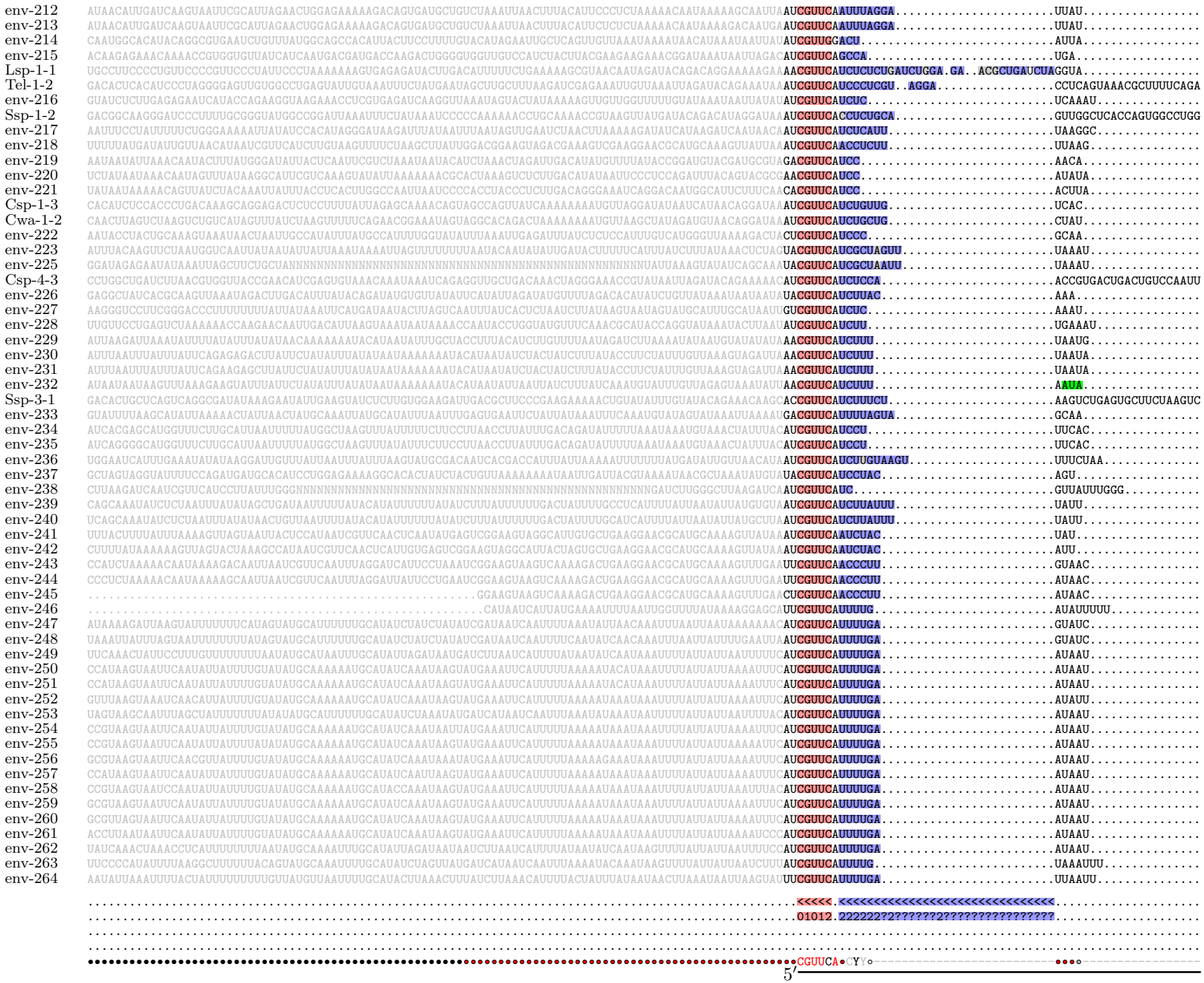
circle (•): 90%, gray circle (⊙): 75%, white circle (○): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

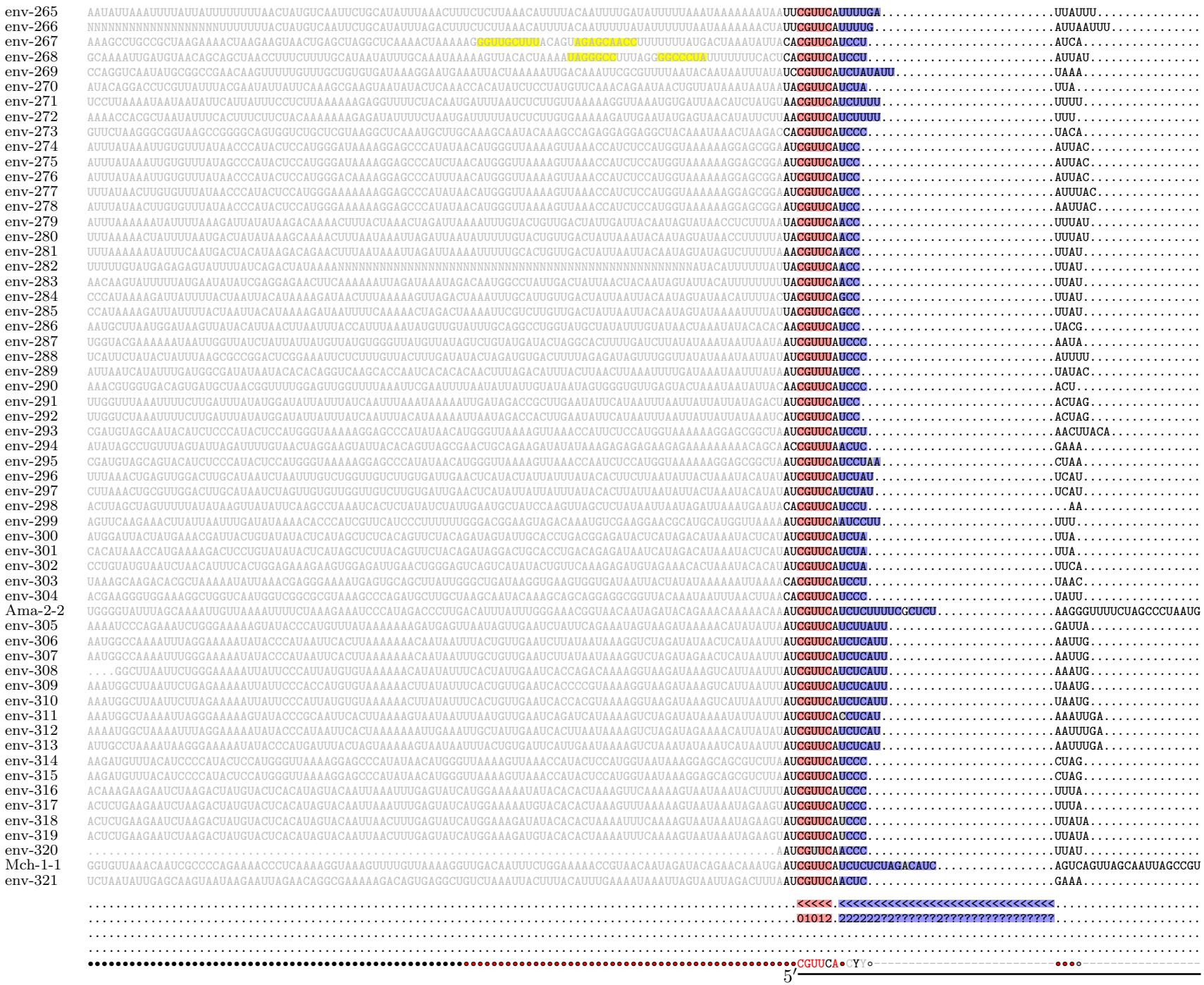


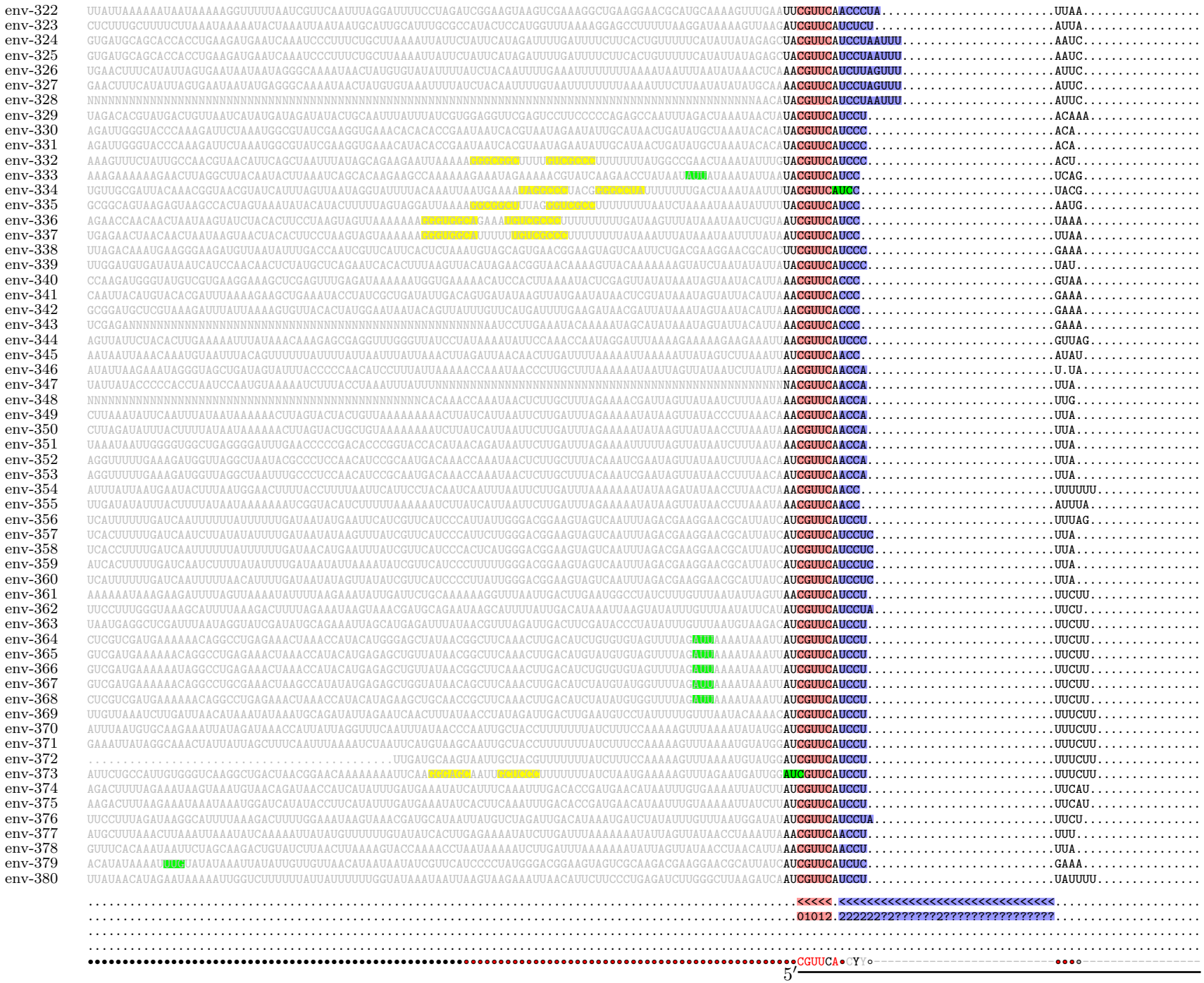

env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
Sph-1-1
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
Pph-1-1
env-102
5′

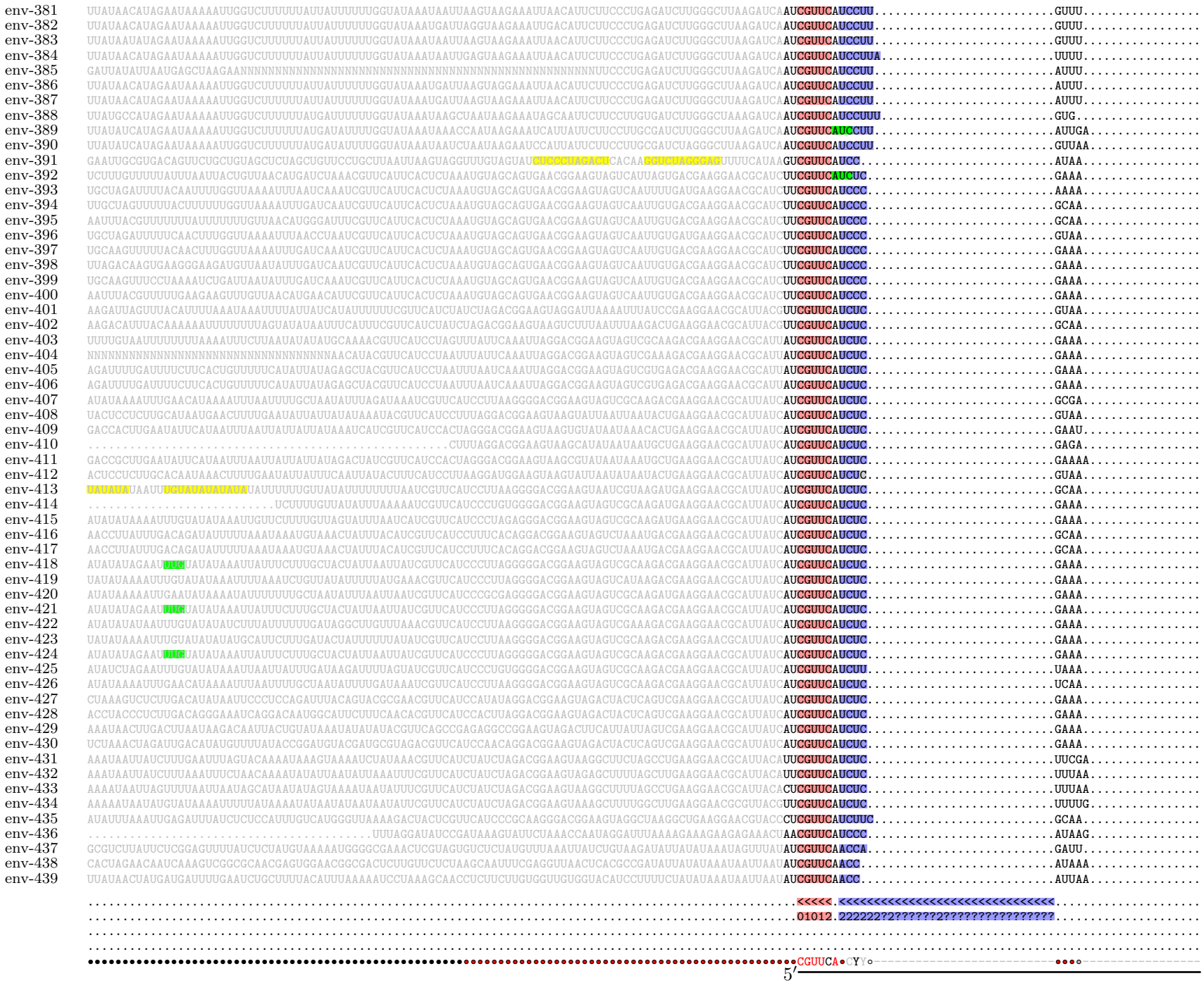

env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
Csp-1-1
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-135
env-136
env-137
env-138
env-139
env-140
env-141
env-142
env-143
env-144
env-145
env-146
env-147
env-148
env-149
env-150
env-151
env-152
env-153
env-154
Csp-3-1
env-155
env-156
env-157
env-158
env-159
5′

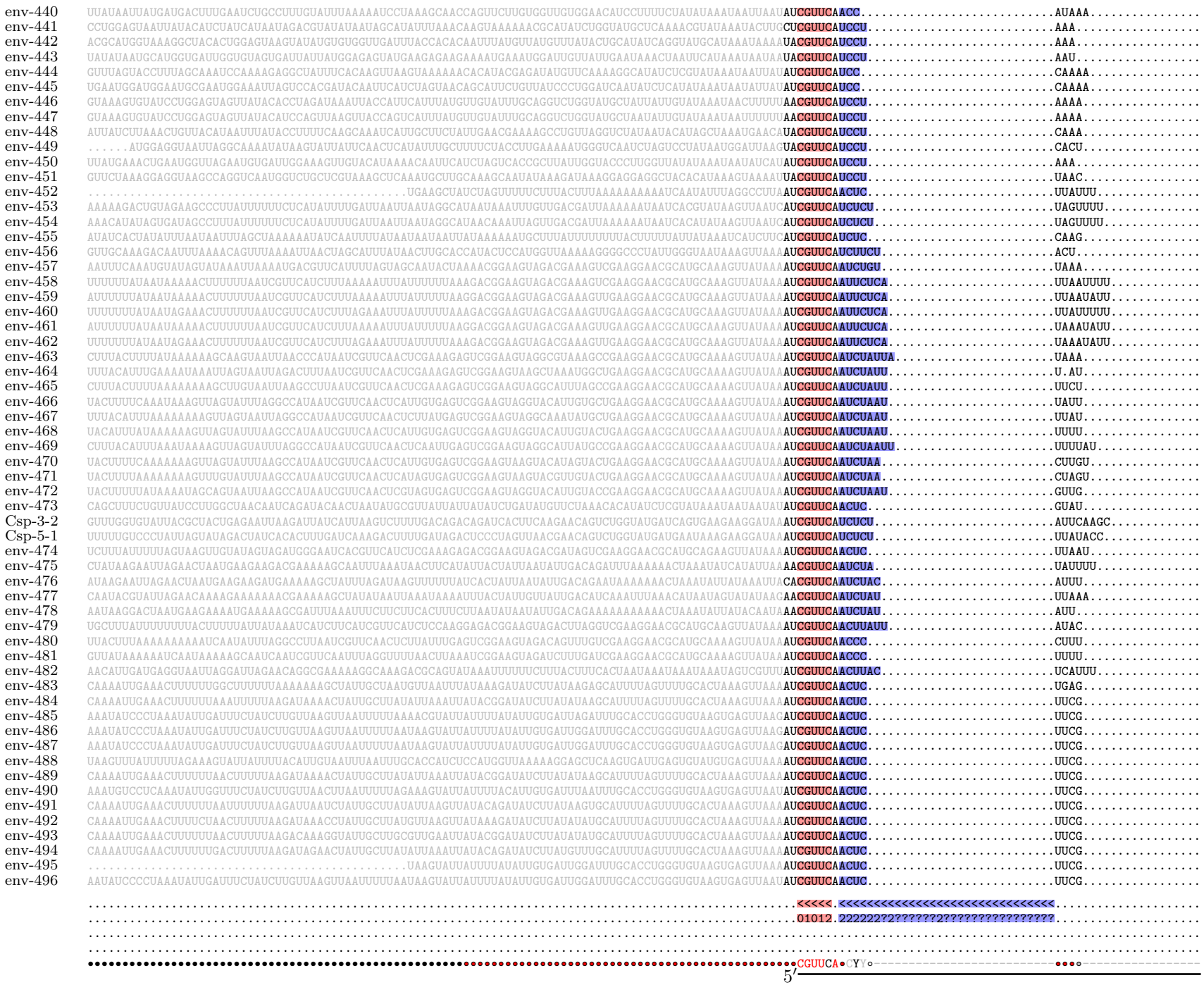

env-160
env-161
env-162
env-163
env-164
env-165
env-166
env-167
env-168
Csp-2-1
Csp-1-2
env-169
env-170
env-171
Csp-4-1
env-172
env-173
env-174
env-175
env-176
env-177
env-178
env-179
Cwa-1-1
Csp-4-2
Tel-1-1
env-180
Ssp-1-1
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
env-190
env-191
env-192
env-193
env-194
env-195
env-196
env-197
env-198
env-199
env-200
env-201
env-202
env-203
env-204
env-205
env-206
env-207
env-208
env-209
env-210
env-211
5′

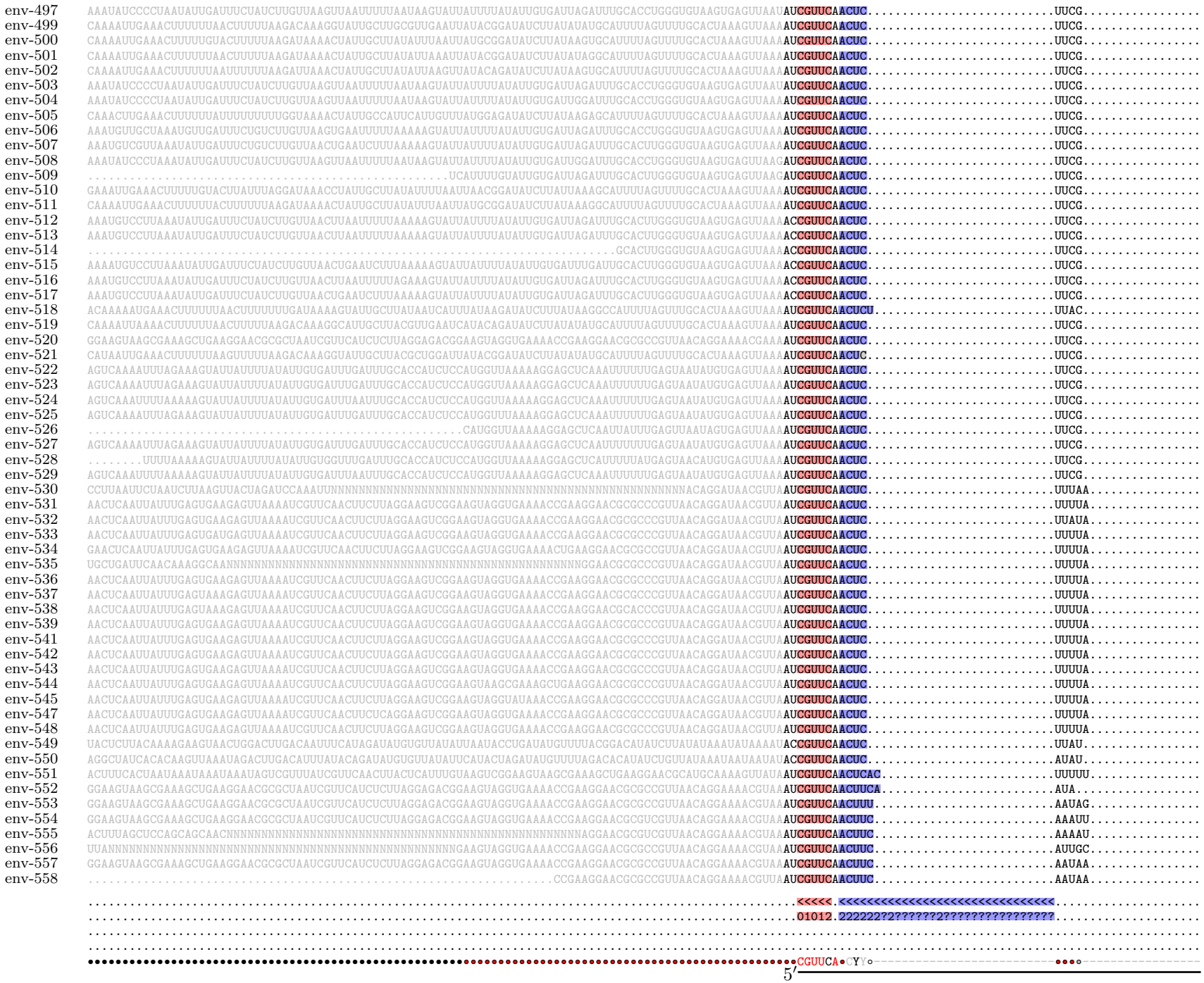

env-212
env-213
env-214
env-215
Lsp-1-1
Tel-1-2
env-216
Ssp-1-2
env-217
env-218
env-219
env-220
env-221
Csp-1-3
Cwa-1-2
env-222
env-223
env-225
Csp-4-3
env-226
env-227
env-228
env-229
env-230
env-231
env-232
Ssp-3-1
env-233
env-234
env-235
env-236
env-237
env-238
env-239
env-240
env-241
env-242
env-243
env-244
env-245
env-246
env-247
env-248
env-249
env-250
env-251
env-252
env-253
env-254
env-255
env-256
env-257
env-258
env-259
env-260
env-261
env-262
env-263
env-264
5′

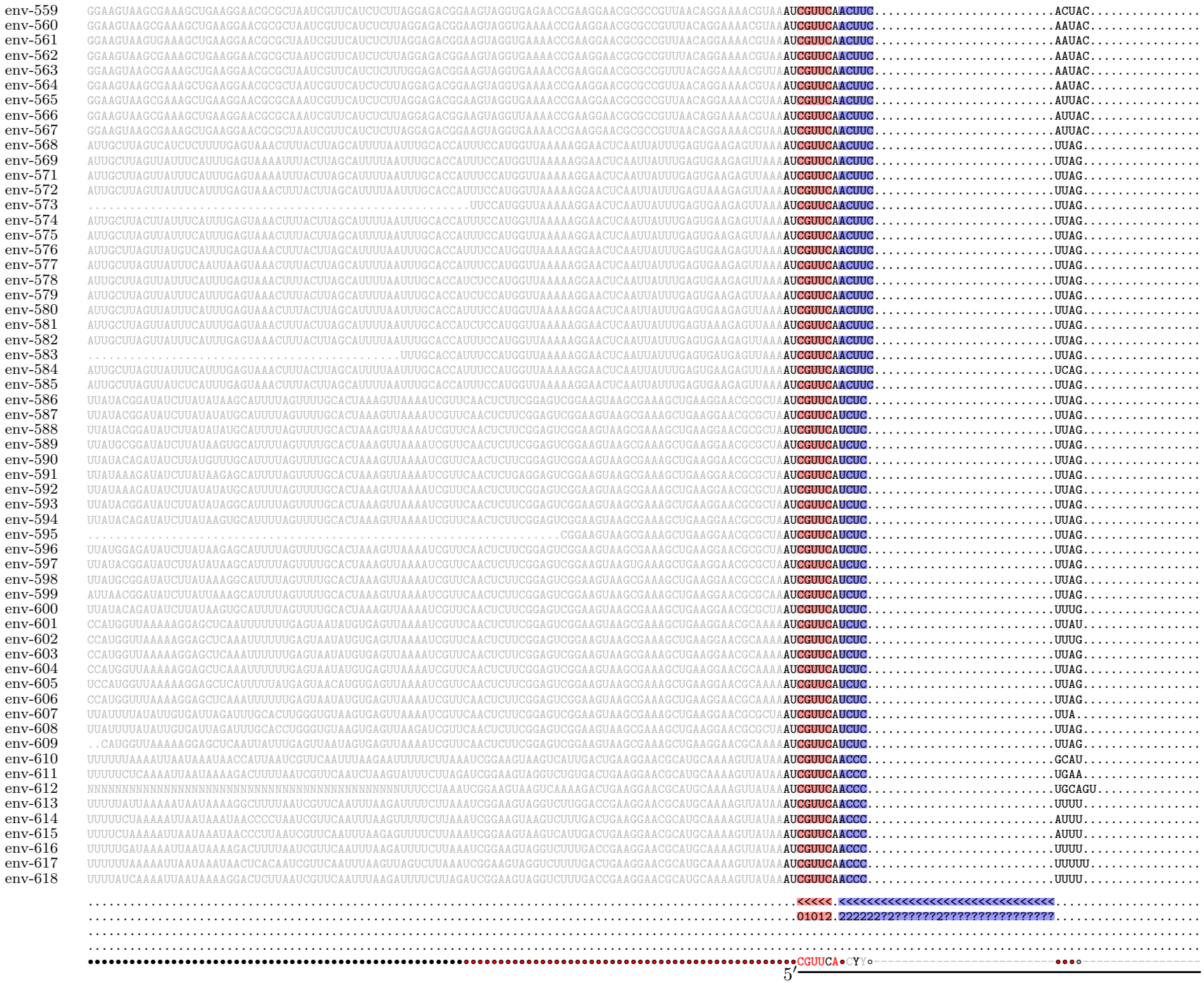

env-265
env-266
env-267
env-268
env-269
env-270
env-271
env-272
env-273
env-274
env-275
env-276
env-277
env-278
env-279
env-280
env-281
env-282
env-283
env-284
env-285
env-286
env-287
env-288
env-289
env-290
env-291
env-292
env-293
env-294
env-295
env-296
env-297
env-298
env-299
env-300
env-301
env-302
env-303
env-304
Ama-2-2
env-305
env-306
env-307
env-308
env-309
env-310
env-311
env-312
env-313
env-314
env-315
env-316
env-317
env-318
env-319
env-320
Mch-1-1
env-321
5′

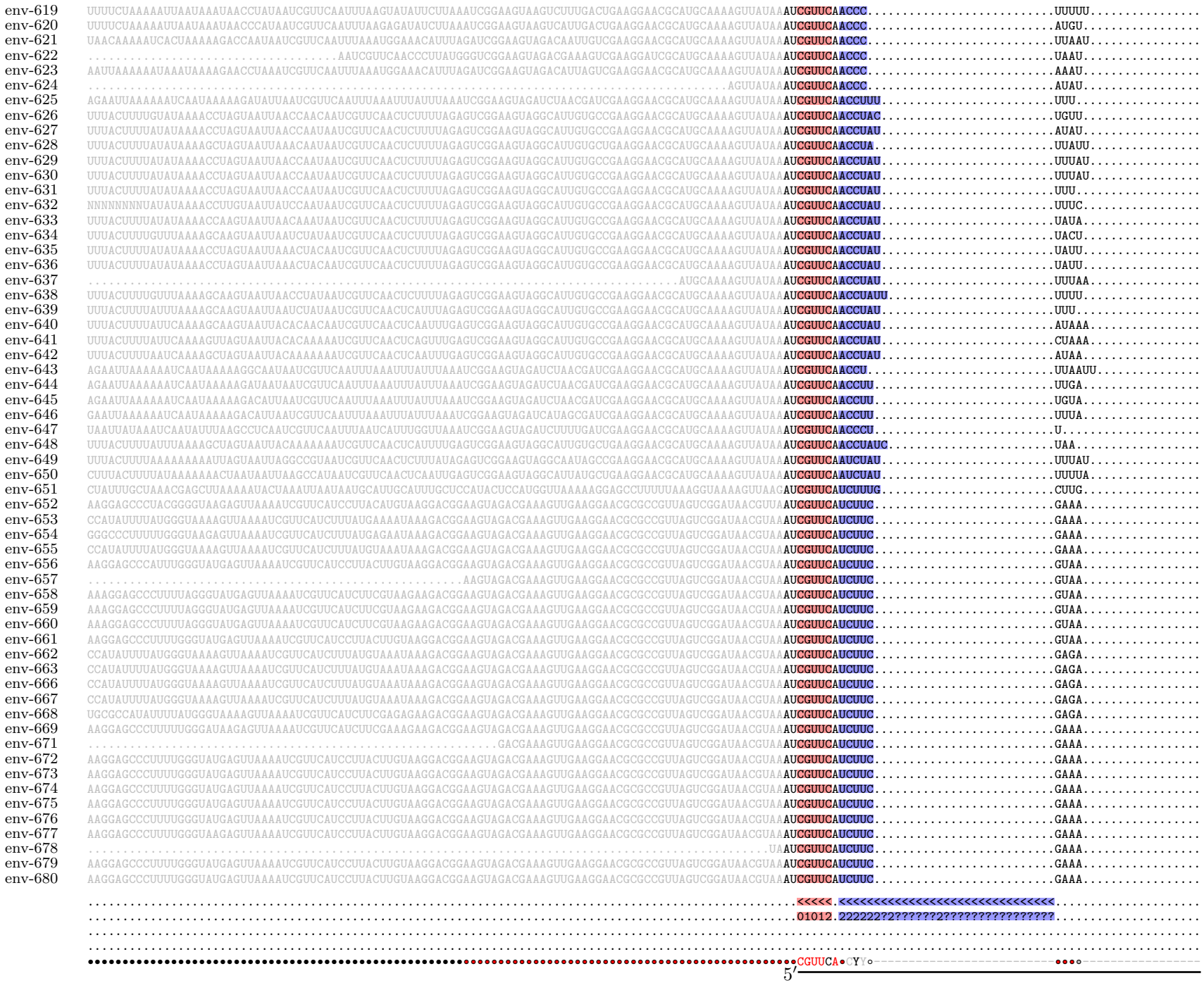

env-322
env-323
env-324
env-325
env-326
env-327
env-328
env-329
env-330
env-331
env-332
env-333
env-334
env-335
env-336
env-337
env-338
env-339
env-340
env-341
env-342
env-343
env-344
env-345
env-346
env-347
env-348
env-349
env-350
env-351
env-352
env-353
env-354
env-355
env-356
env-357
env-358
env-359
env-360
env-361
env-362
env-363
env-364
env-365
env-366
env-367
env-368
env-369
env-370
env-371
env-372
env-373
env-374
env-375
env-376
env-377
env-378
env-379
env-380
5′

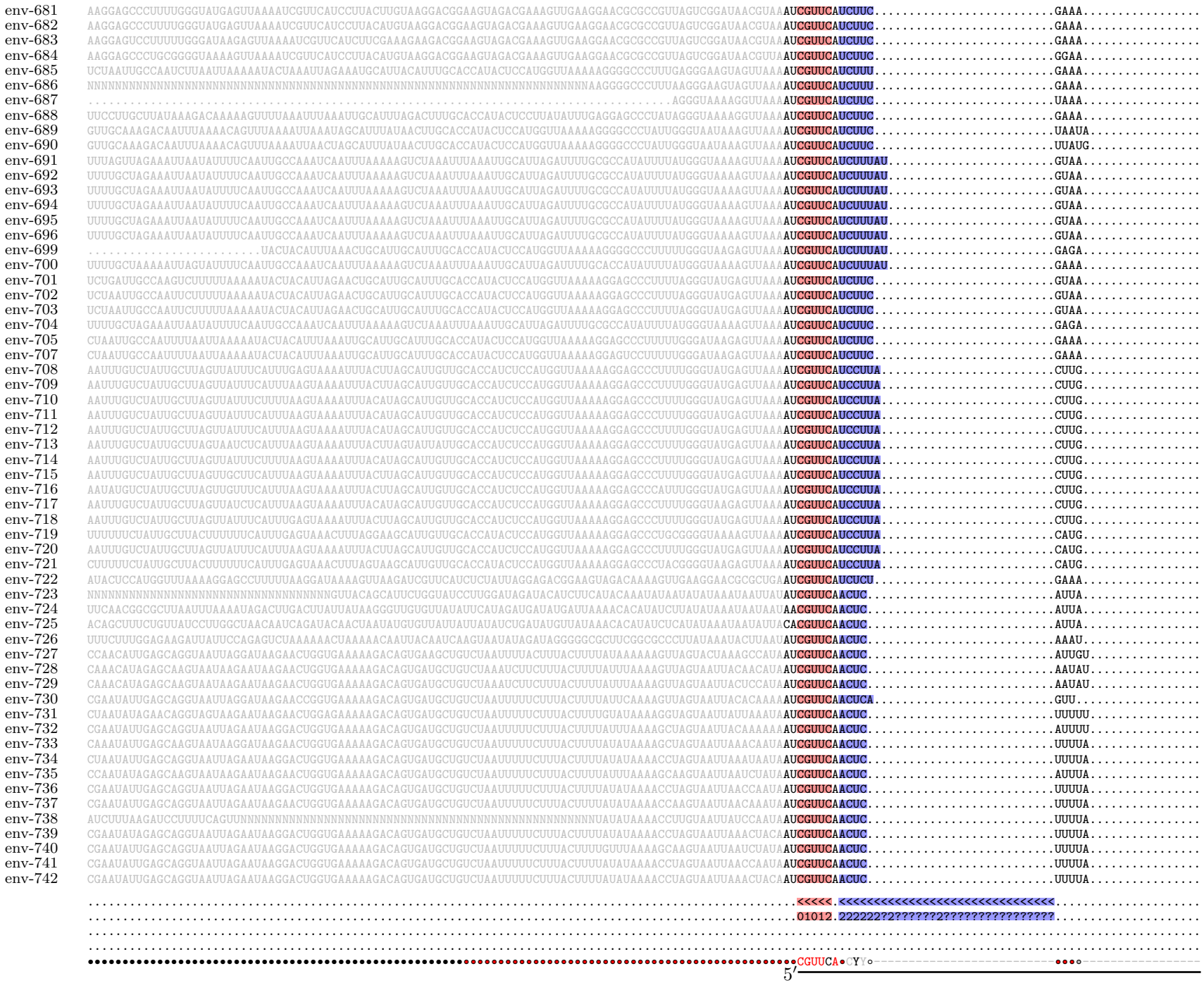

env-381
env-382
env-383
env-384
env-385
env-386
env-387
env-388
env-389
env-390
env-391
env-392
env-393
env-394
env-395
env-396
env-397
env-398
env-399
env-400
env-401
env-402
env-403
env-404
env-405
env-406
env-407
env-408
env-409
env-410
env-411
env-412
env-413
env-414
env-415
env-416
env-417
env-418
env-419
env-420
env-421
env-422
env-423
env-424
env-425
env-426
env-427
env-428
env-429
env-430
env-431
env-432
env-433
env-434
env-435
env-436
env-437
env-438
env-439
5′

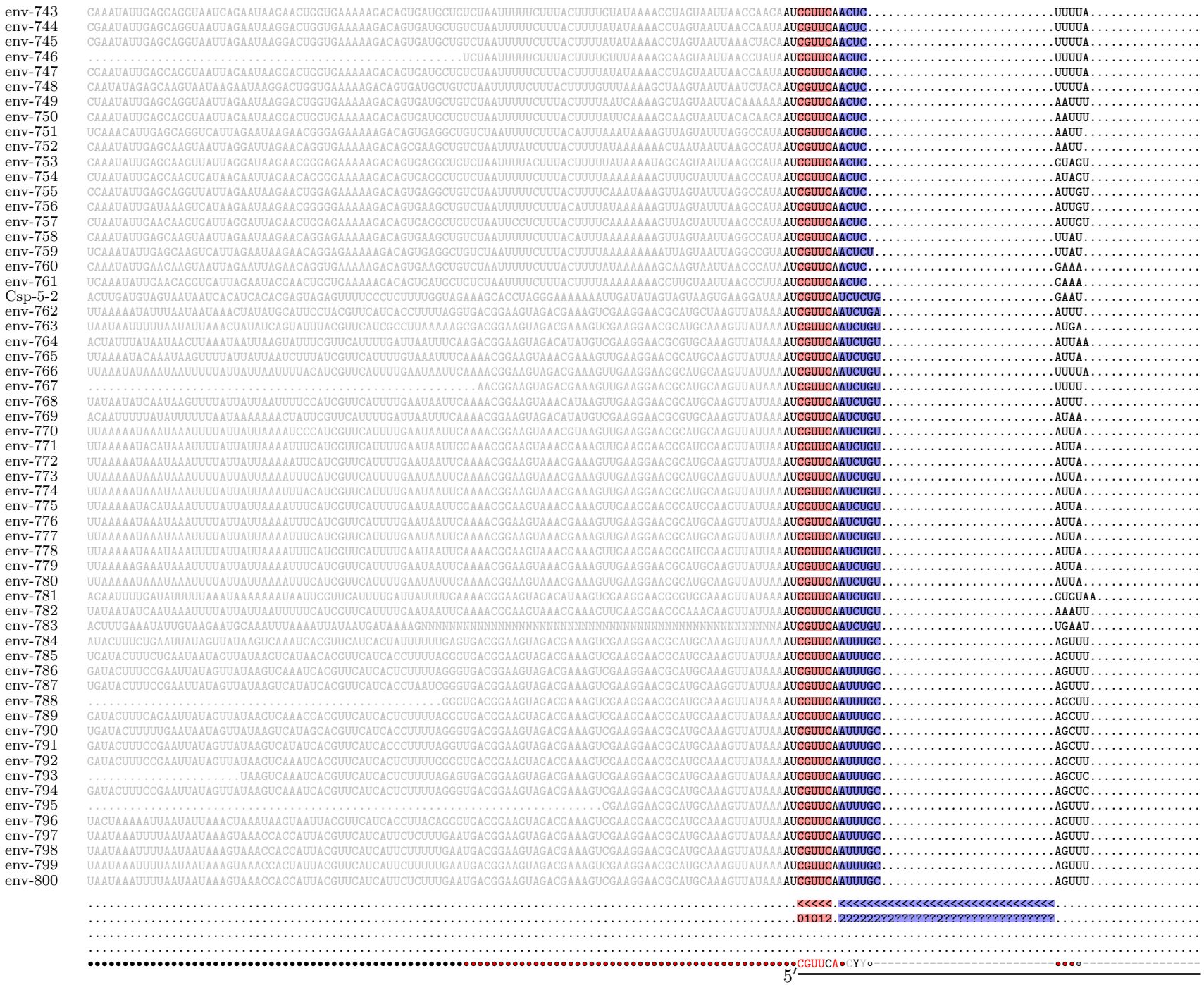

env-440
env-441
env-442
env-443
env-444
env-445
env-446
env-447
env-448
env-449
env-450
env-451
env-452
env-453
env-454
env-455
env-456
env-457
env-458
env-459
env-460
env-461
env-462
env-463
env-464
env-465
env-466
env-467
env-468
env-469
env-470
env-471
env-472
env-473
Csp-3-2
Csp-5-1
env-474
env-475
env-476
env-477
env-478
env-479
env-480
env-481
env-482
env-483
env-484
env-485
env-486
env-487
env-488
env-489
env-490
env-491
env-492
env-493
env-494
env-495
env-496
5′

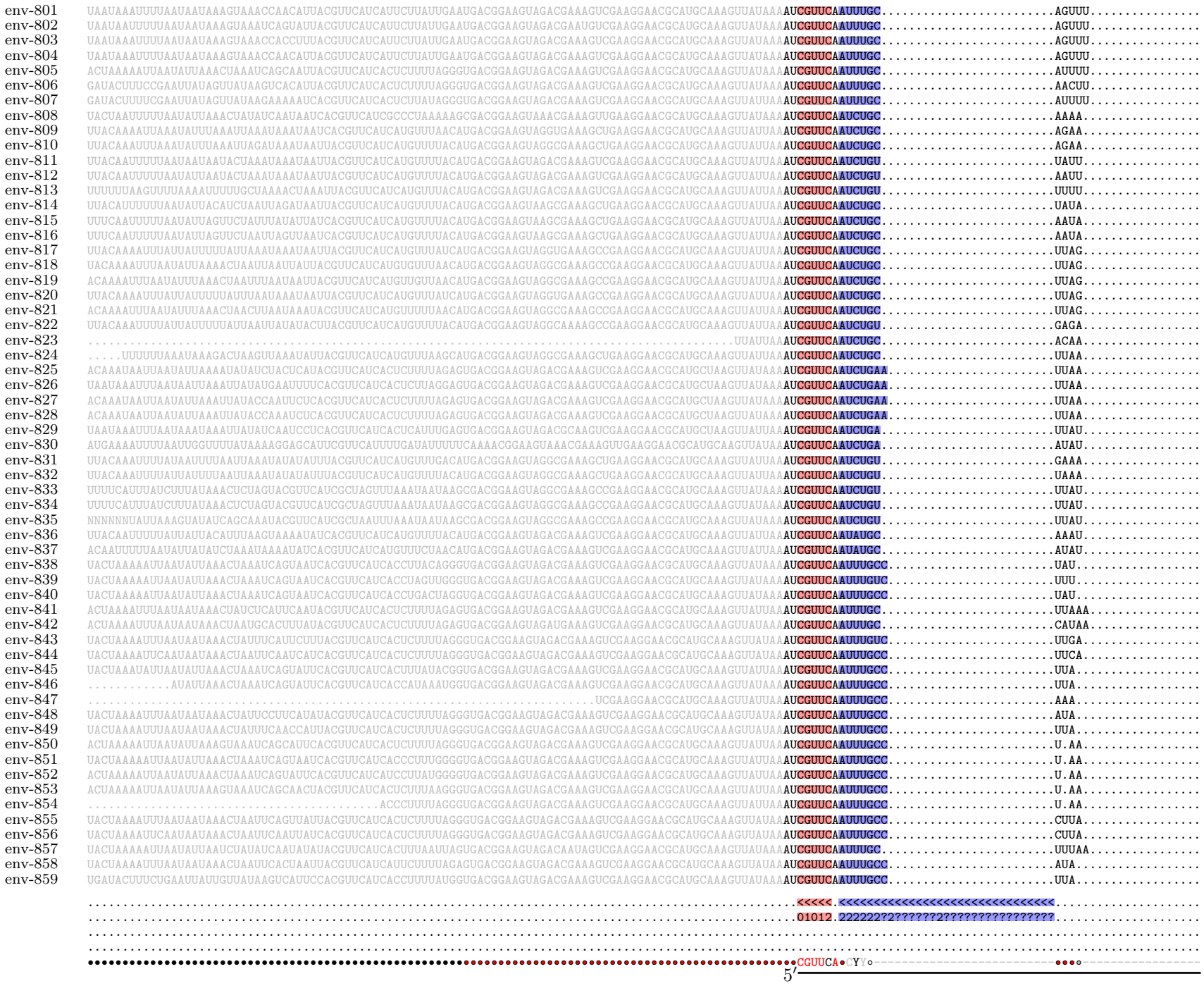

env-497
env-499
env-500
env-501
env-502
env-503
env-504
env-505
env-506
env-507
env-508
env-509
env-510
env-511
env-512
env-513
env-514
env-515
env-516
env-517
env-518
env-519
env-520
env-521
env-522
env-523
env-524
env-525
env-526
env-527
env-528
env-529
env-530
env-531
env-532
env-533
env-534
env-535
env-536
env-537
env-538
env-539
env-541
env-542
env-543
env-544
env-545
env-547
env-548
env-549
env-550
env-551
env-552
env-553
env-554
env-555
env-556
env-557
env-558
5′

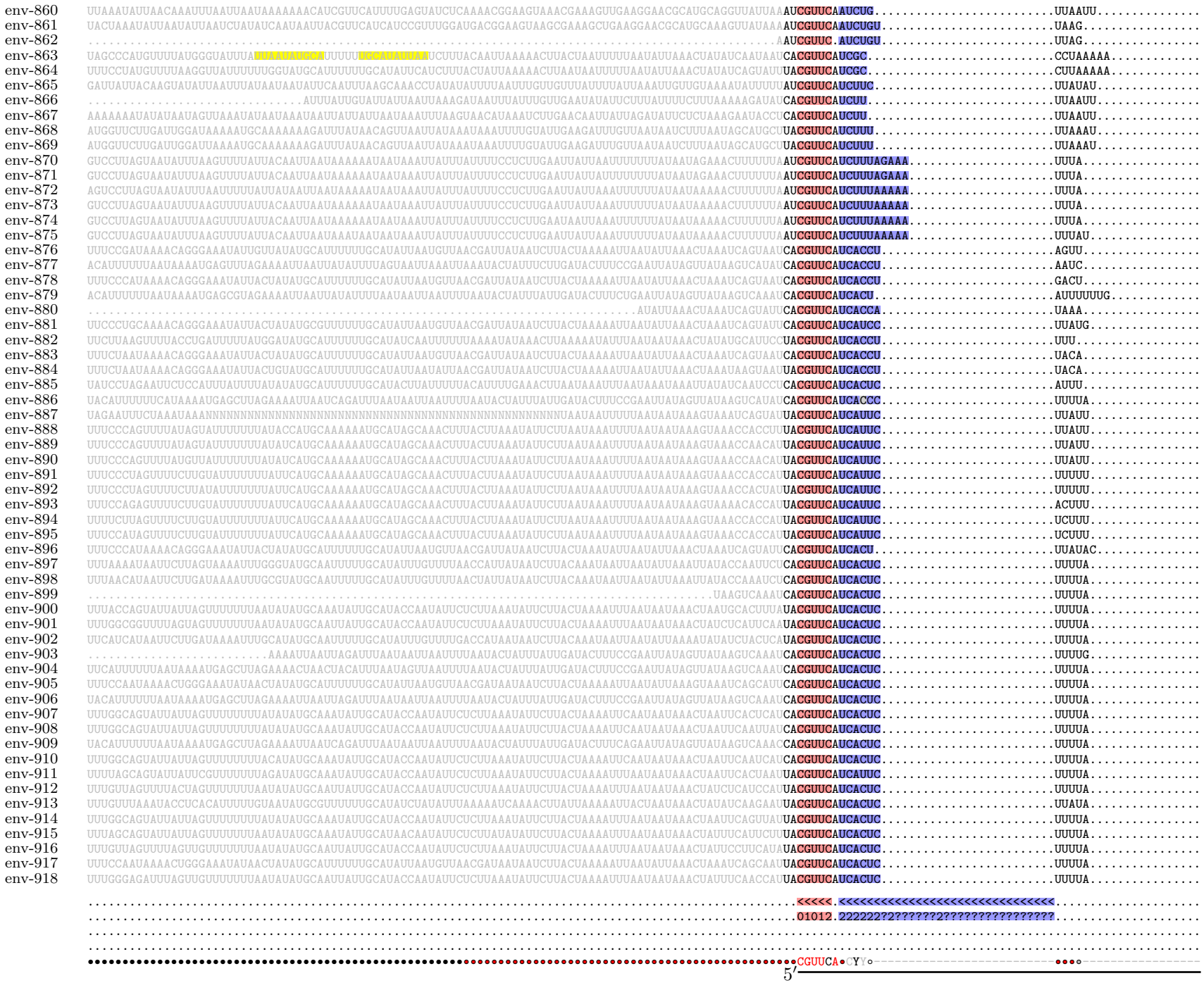

env-559
env-560
env-561
env-562
env-563
env-564
env-565
env-566
env-567
env-568
env-569
env-571
env-572
env-573
env-574
env-575
env-576
env-577
env-578
env-579
env-580
env-581
env-582
env-583
env-584
env-585
env-586
env-587
env-588
env-589
env-590
env-591
env-592
env-593
env-594
env-595
env-596
env-597
env-598
env-599
env-600
env-601
env-602
env-603
env-604
env-605
env-606
env-607
env-608
env-609
env-610
env-611
env-612
env-613
env-614
env-615
env-616
env-617
env-618
5′

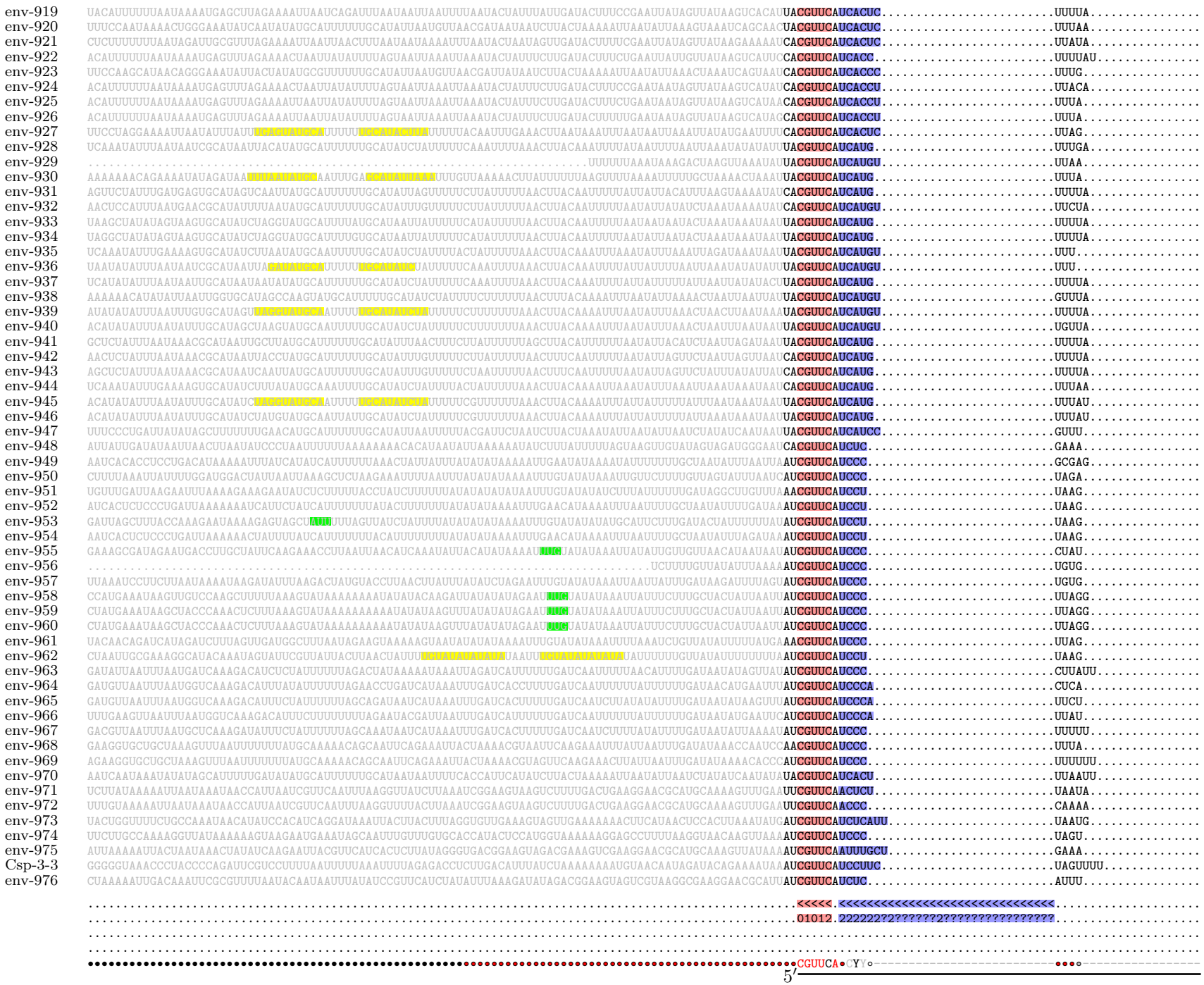

env-619
env-620
env-621
env-622
env-623
env-624
env-625
env-626
env-627
env-628
env-629
env-630
env-631
env-632
env-633
env-634
env-635
env-636
env-637
env-638
env-639
env-640
env-641
env-642
env-643
env-644
env-645
env-646
env-647
env-648
env-649
env-650
env-651
env-652
env-653
env-654
env-655
env-656
env-657
env-658
env-659
env-660
env-661
env-662
env-663
env-666
env-667
env-668
env-669
env-671
env-672
env-673
env-674
env-675
env-676
env-677
env-678
env-679
env-680
5′

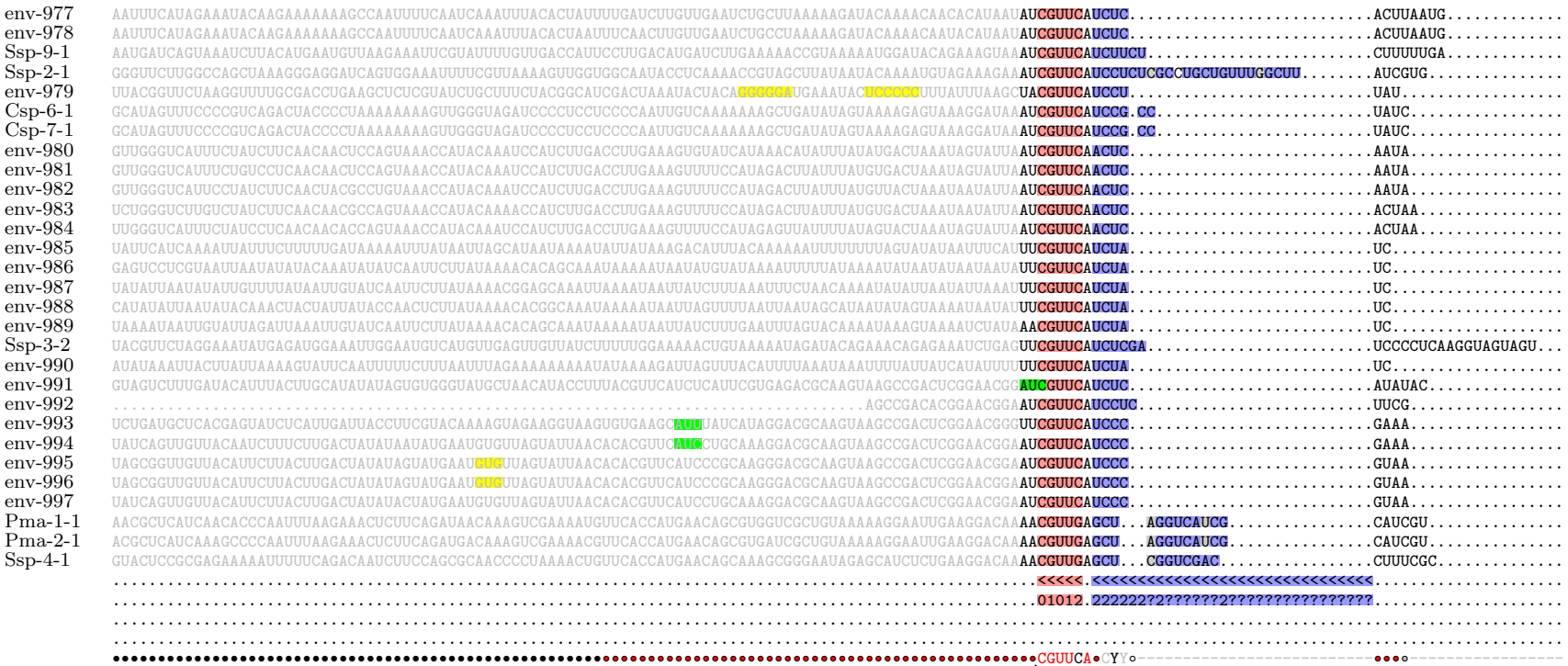

env-681
env-682
env-683
env-684
env-685
env-686
env-687
env-688
env-689
env-690
env-691
env-692
env-693
env-694
env-695
env-696
env-699
env-700
env-701
env-702
env-703
env-704
env-705
env-707
env-708
env-709
env-710
env-711
env-712
env-713
env-714
env-715
env-716
env-717
env-718
env-719
env-720
env-721
env-722
env-723
env-724
env-725
env-726
env-727
env-728
env-729
env-730
env-731
env-732
env-733
env-734
env-735
env-736
env-737
env-738
env-739
env-740
env-741
env-742
5′

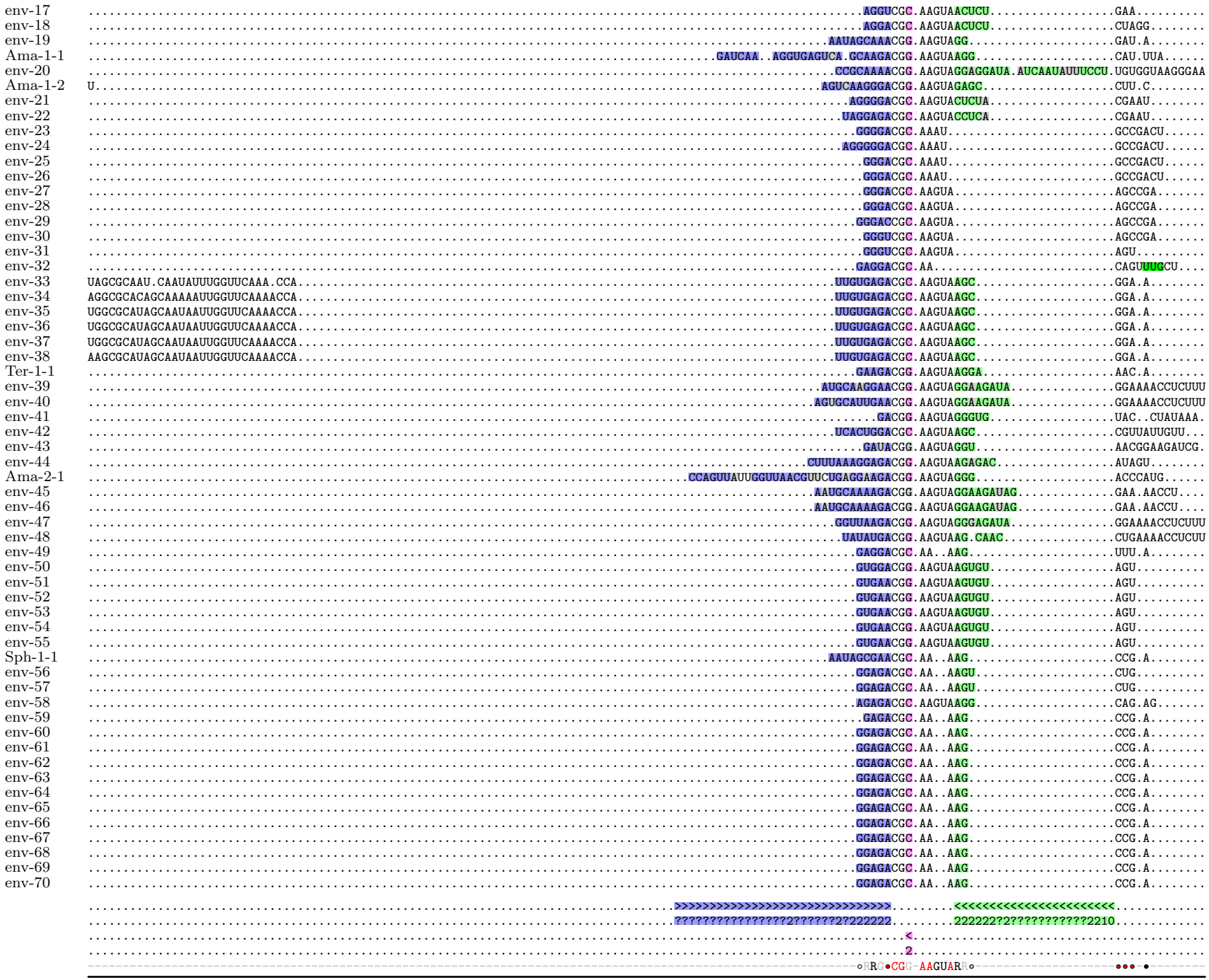

env-743
env-744
env-745
env-746
env-747
env-748
env-749
env-750
env-751
env-752
env-753
env-754
env-755
env-756
env-757
env-758
env-759
env-760
env-761
Csp-5-2
env-762
env-763
env-764
env-765
env-766
env-767
env-768
env-769
env-770
env-771
env-772
env-773
env-774
env-775
env-776
env-777
env-778
env-779
env-780
env-781
env-782
env-783
env-784
env-785
env-786
env-787
env-788
env-789
env-790
env-791
env-792
env-793
env-794
env-795
env-796
env-797
env-798
env-799
env-800
5′

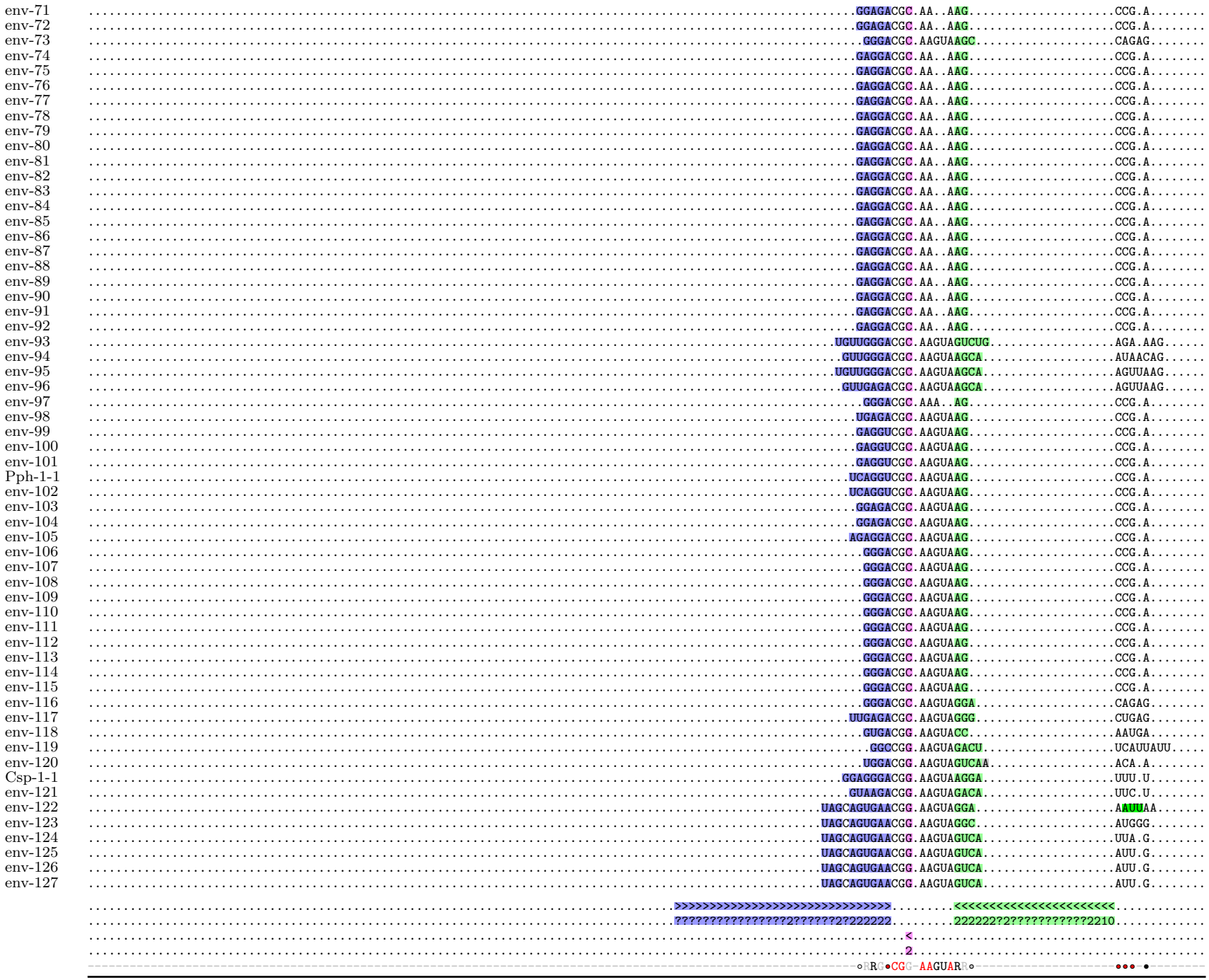

env-801
env-802
env-803
env-804
env-805
env-806
env-807
env-808
env-809
env-810
env-811
env-812
env-813
env-814
env-815
env-816
env-817
env-818
env-819
env-820
env-821
env-822
env-823
env-824
env-825
env-826
env-827
env-828
env-829
env-830
env-831
env-832
env-833
env-834
env-835
env-836
env-837
env-838
env-839
env-840
env-841
env-842
env-843
env-844
env-845
env-846
env-847
env-848
env-849
env-850
env-851
env-852
env-853
env-854
env-855
env-856
env-857
env-858
env-859
5′

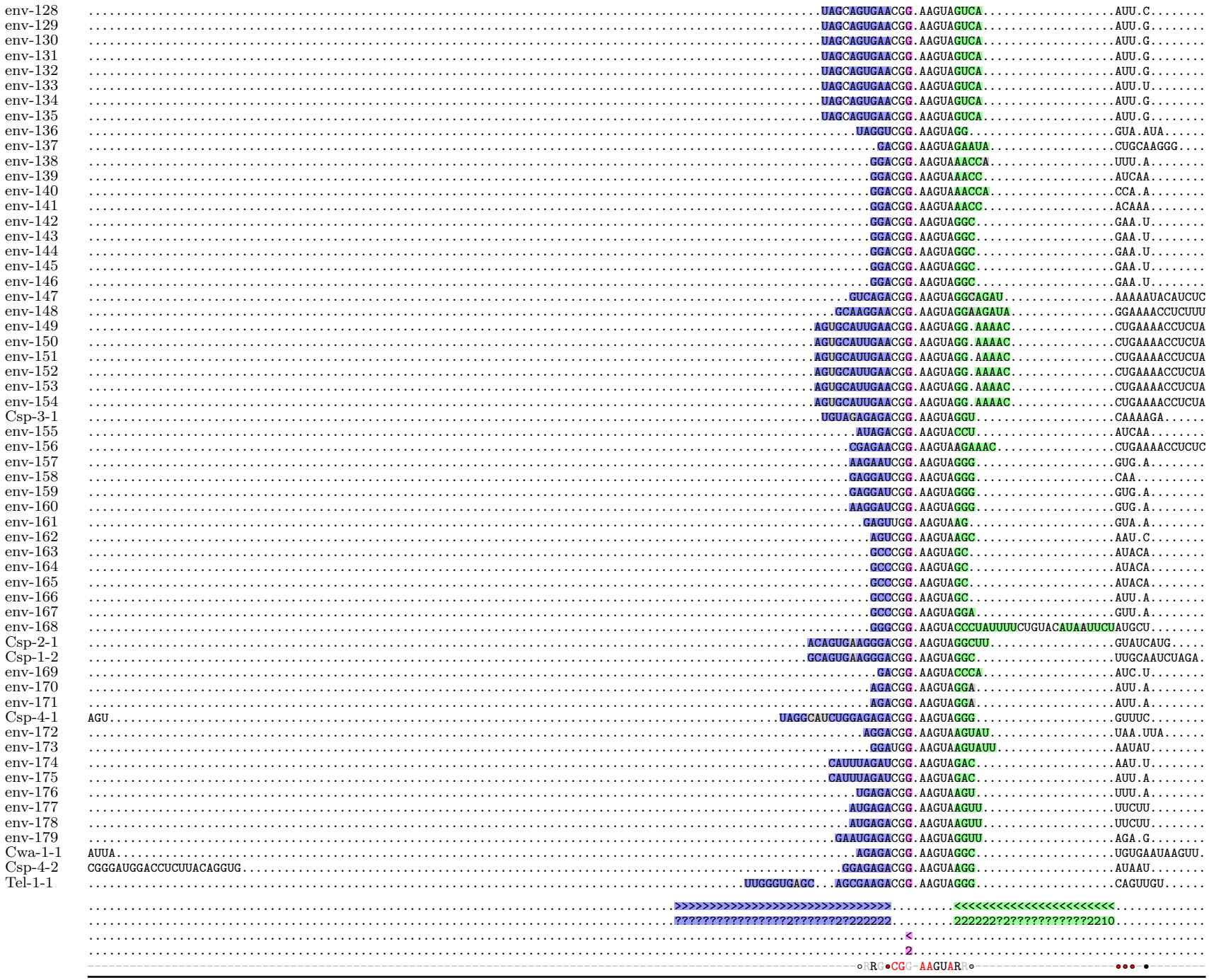

env-860
env-861
env-862
env-863
env-864
env-865
env-866
env-867
env-868
env-869
env-870
env-871
env-872
env-873
env-874
env-875
env-876
env-877
env-878
env-879
env-880
env-881
env-882
env-883
env-884
env-885
env-886
env-887
env-888
env-889
env-890
env-891
env-892
env-893
env-894
env-895
env-896
env-897
env-898
env-899
env-900
env-901
env-902
env-903
env-904
env-905
env-906
env-907
env-908
env-909
env-910
env-911
env-912
env-913
env-914
env-915
env-916
env-917
env-918
5′

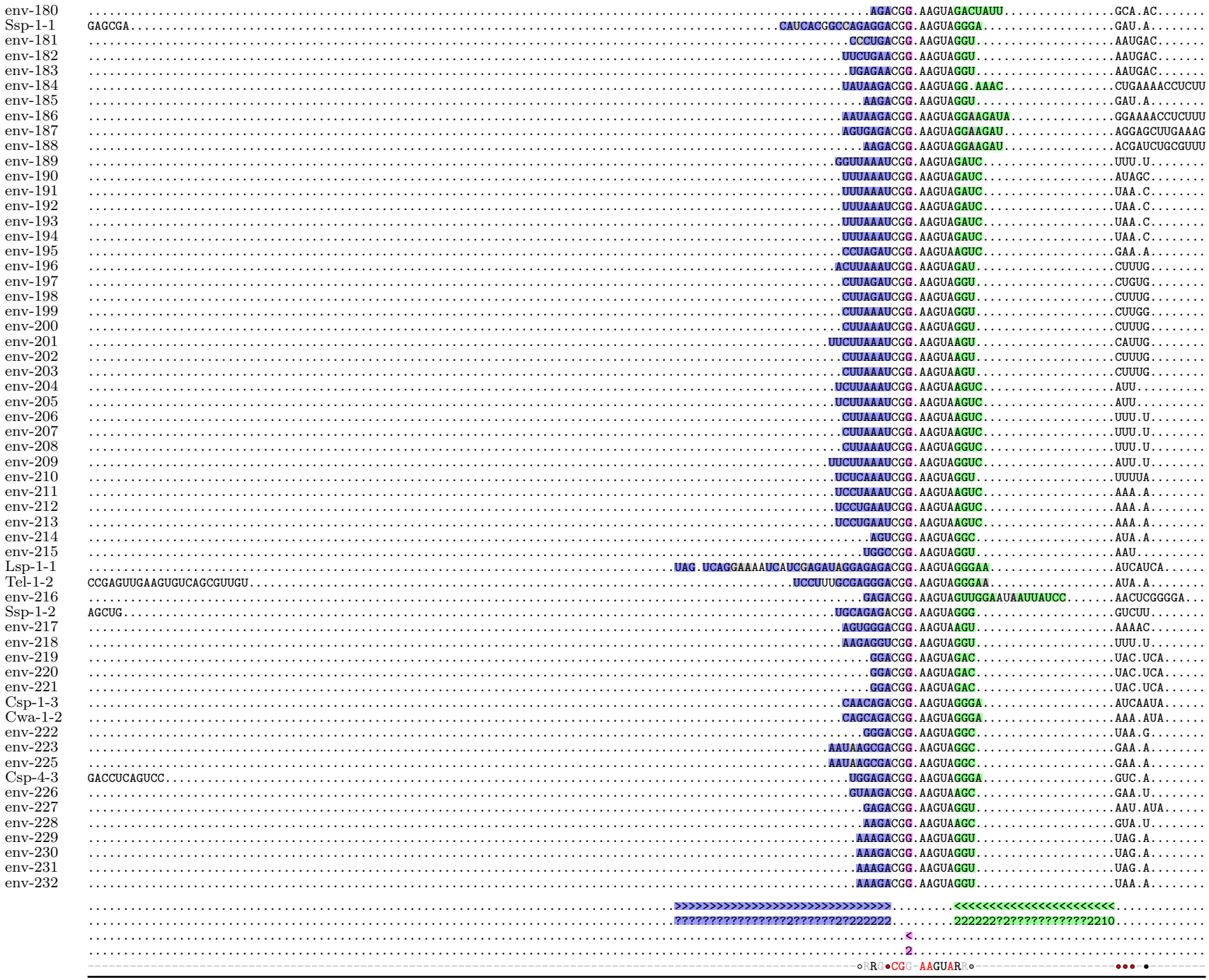

env-919
env-920
env-921
env-922
env-923
env-924
env-925
env-926
env-927
env-928
env-929
env-930
env-931
env-932
env-933
env-934
env-935
env-936
env-937
env-938
env-939
env-940
env-941
env-942
env-943
env-944
env-945
env-946
env-947
env-948
env-949
env-950
env-951
env-952
env-953
env-954
env-955
env-956
env-957
env-958
env-959
env-960
env-961
env-962
env-963
env-964
env-965
env-966
env-967
env-968
env-969
env-970
env-971
env-972
env-973
env-974
env-975
Csp-3-3
env-976
5′

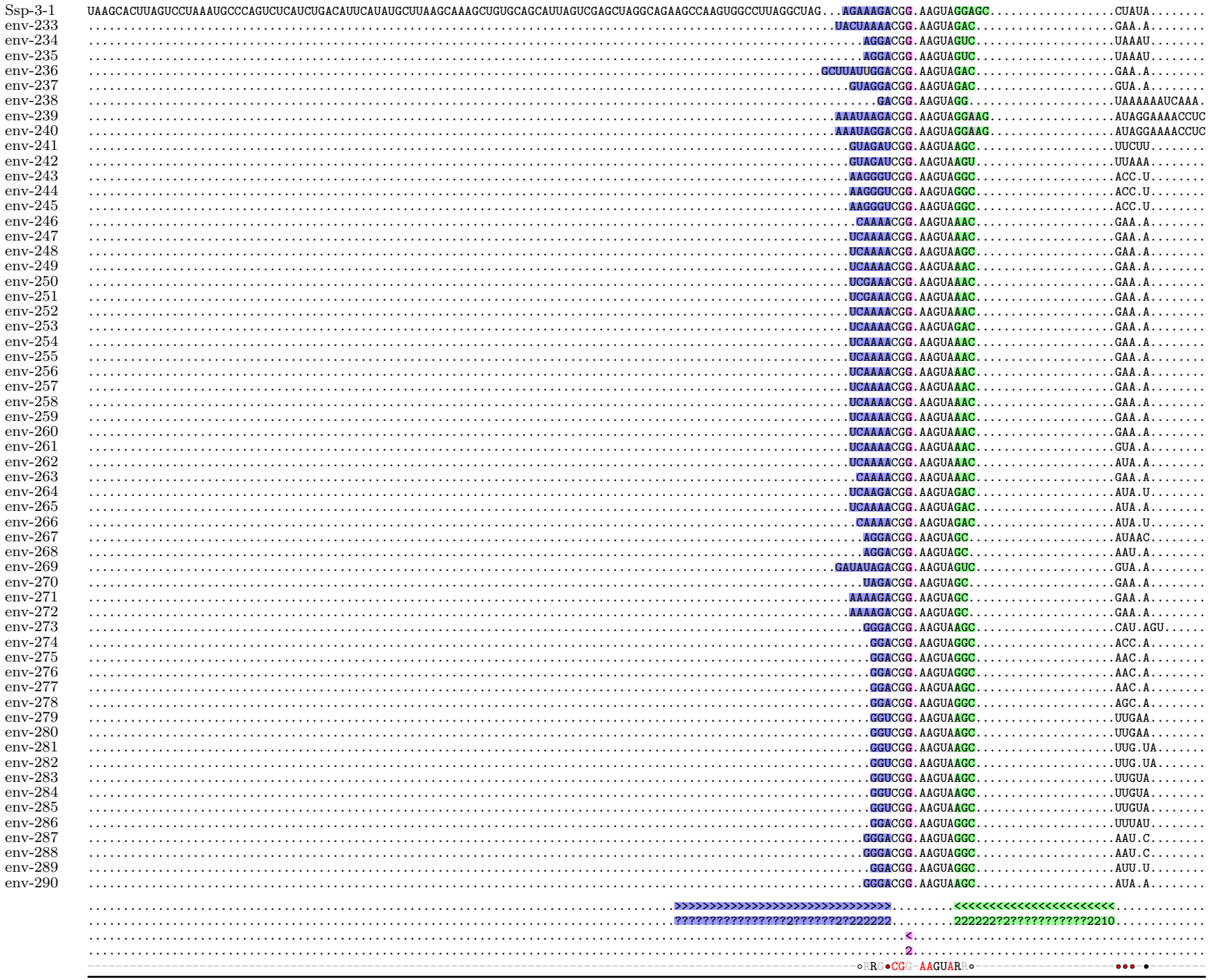

env-977
env-978
Ssp-9-1
Ssp-2-1
env-979
Csp-6-1
Csp-7-1
env-980
env-981
env-982
env-983
env-984
env-985
env-986
env-987
env-988
env-989
Ssp-3-2
env-990
env-991
env-992
env-993
env-994
env-995
env-996
env-997
Pma-1-1
Pma-2-1
Ssp-4-1
5′
alignment positions 161 · · · 320
env-1
Ssp-5-1
env-2
env-3
env-4
env-6
Ssp-6-1
Sel-1-1
env-7
env-8
env-9
env-10
env-11
Ssp-8-1
Ssp-7-1
env-12
env-13
env-14
env-15
env-16

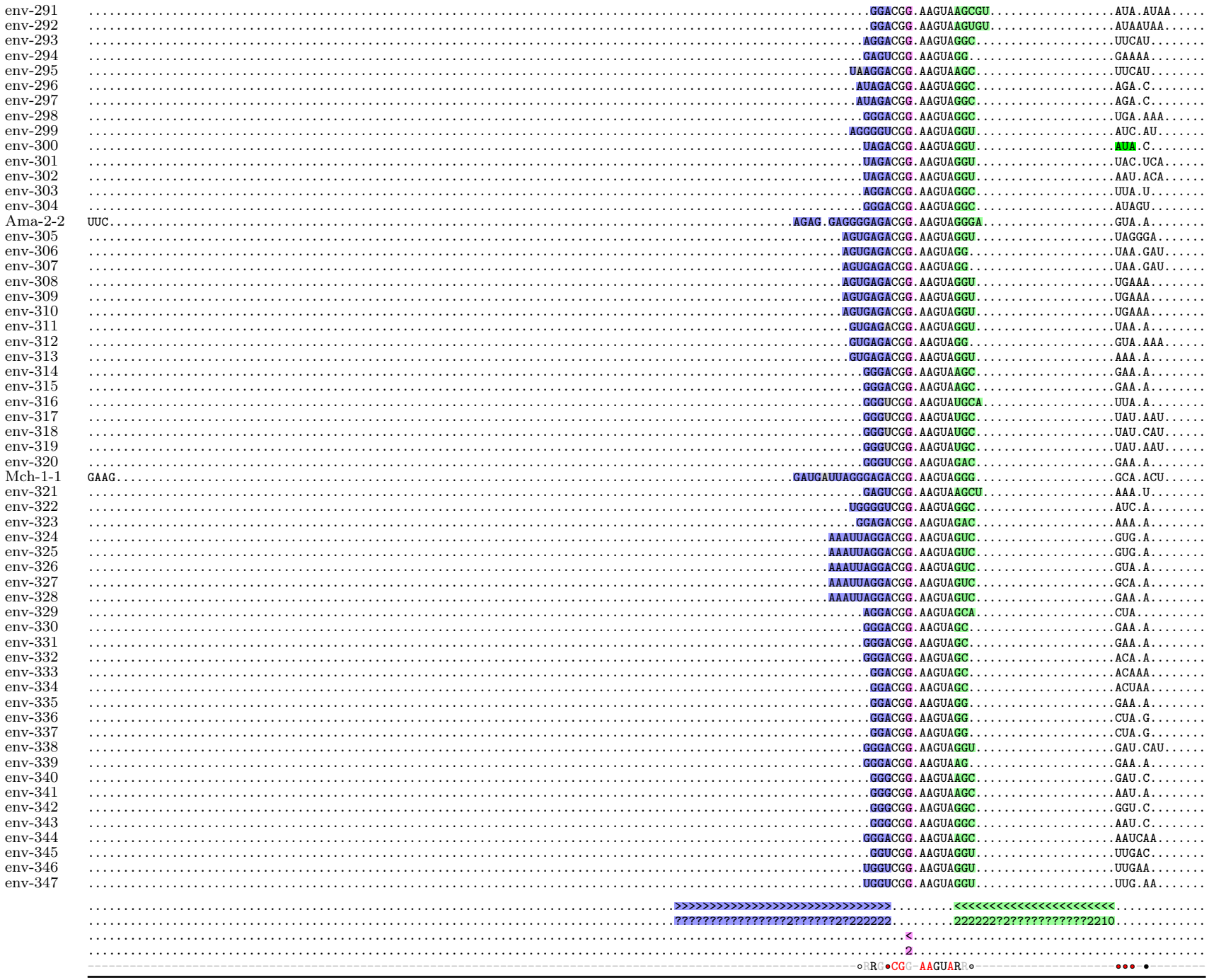

env-17
env-18
env-19
Ama-1-1
env-20
Ama-1-2
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
Ter-1-1
env-39
env-40
env-41
env-42
env-43
env-44
Ama-2-1
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
Sph-1-1
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70

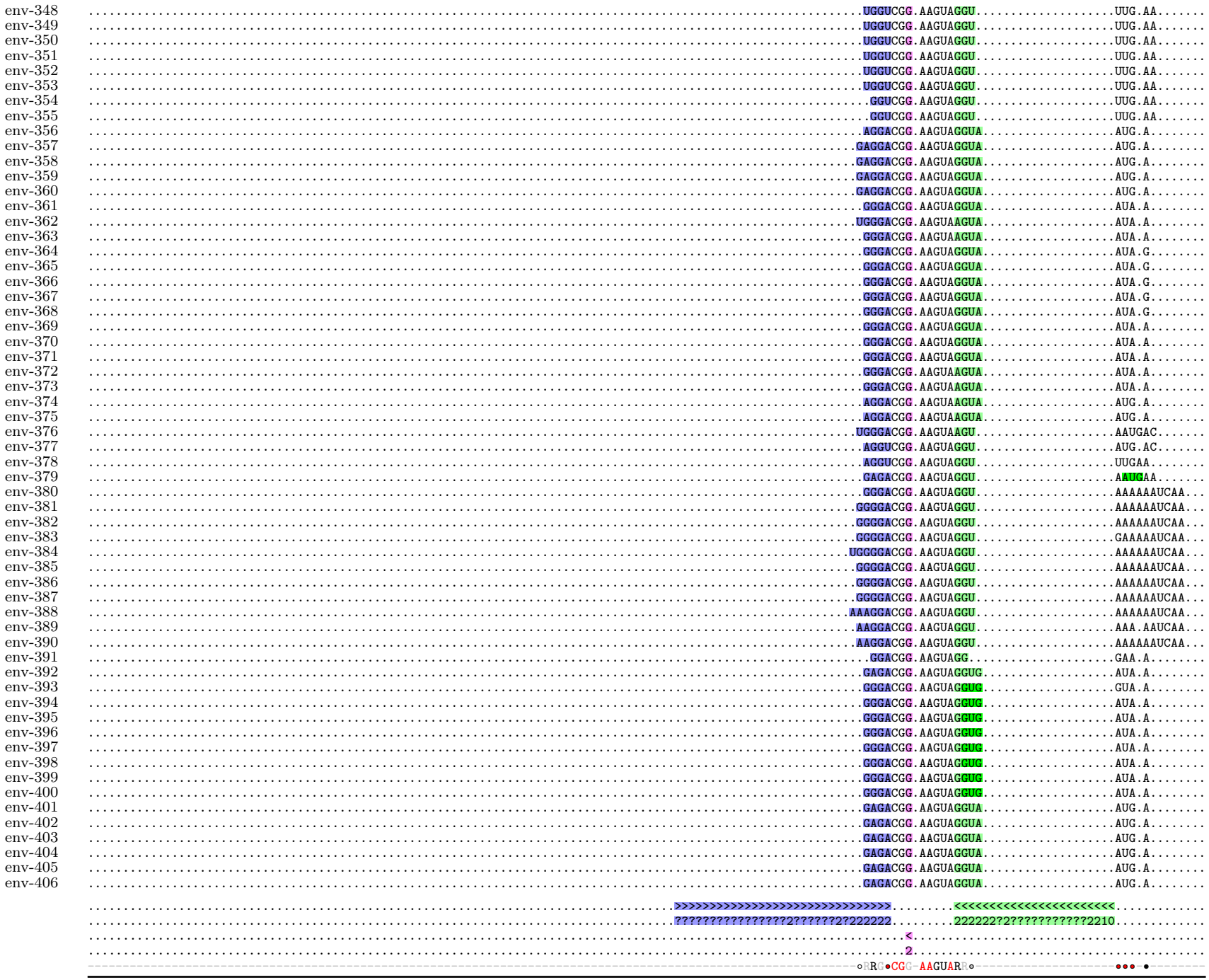

env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
Pph-1-1
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
Csp-1-1
env-121
env-122
env-123
env-124
env-125
env-126
env-127

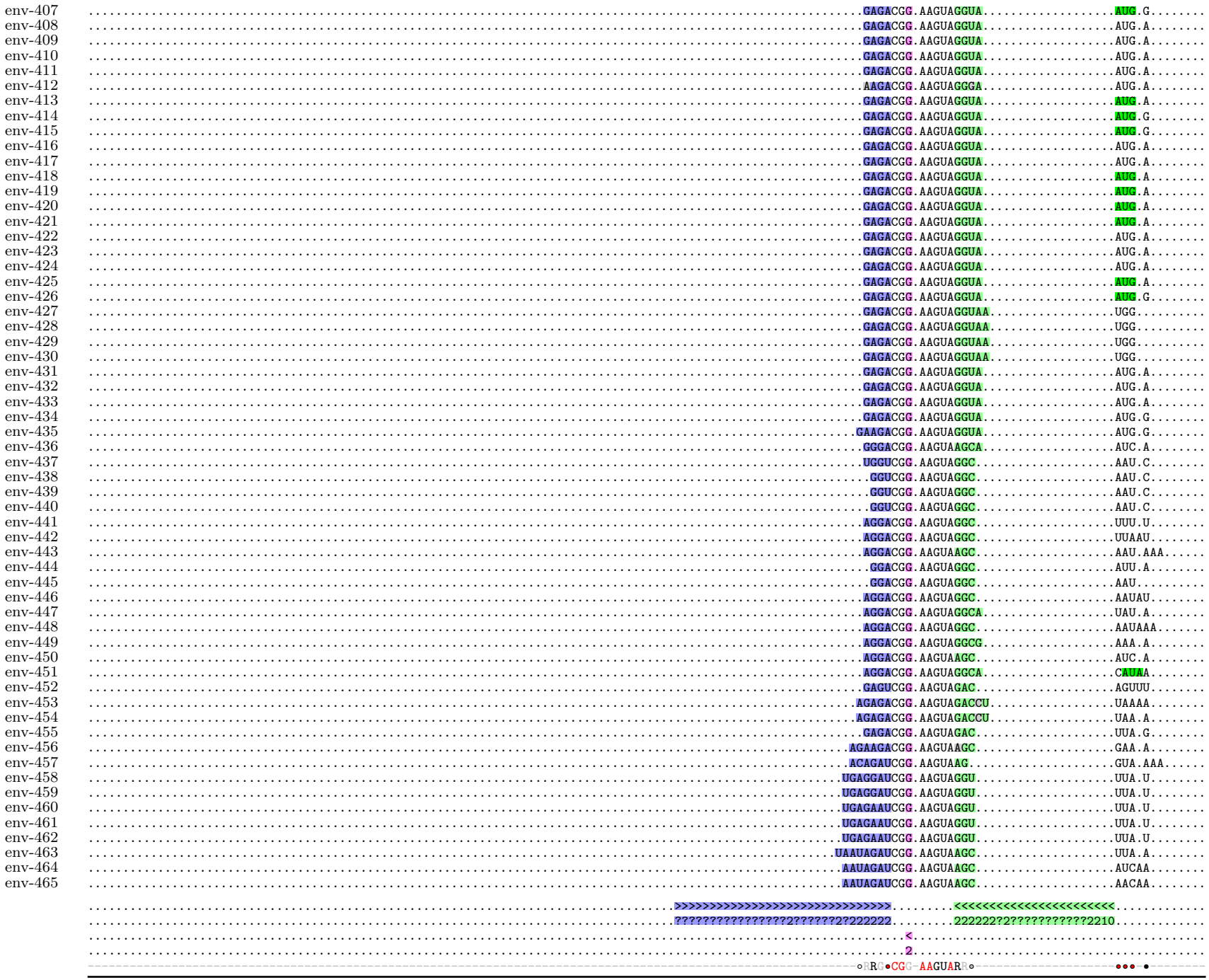

env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-135
env-136
env-137
env-138
env-139
env-140
env-141
env-142
env-143
env-144
env-145
env-146
env-147
env-148
env-149
env-150
env-151
env-152
env-153
env-154
Csp-3-1
env-155
env-156
env-157
env-158
env-159
env-160
env-161
env-162
env-163
env-164
env-165
env-166
env-167
env-168
Csp-2-1
Csp-1-2
env-169
env-170
env-171
Csp-4-1
env-172
env-173
env-174
env-175
env-176
env-177
env-178
env-179
Cwa-1-1
Csp-4-2
Tel-1-1

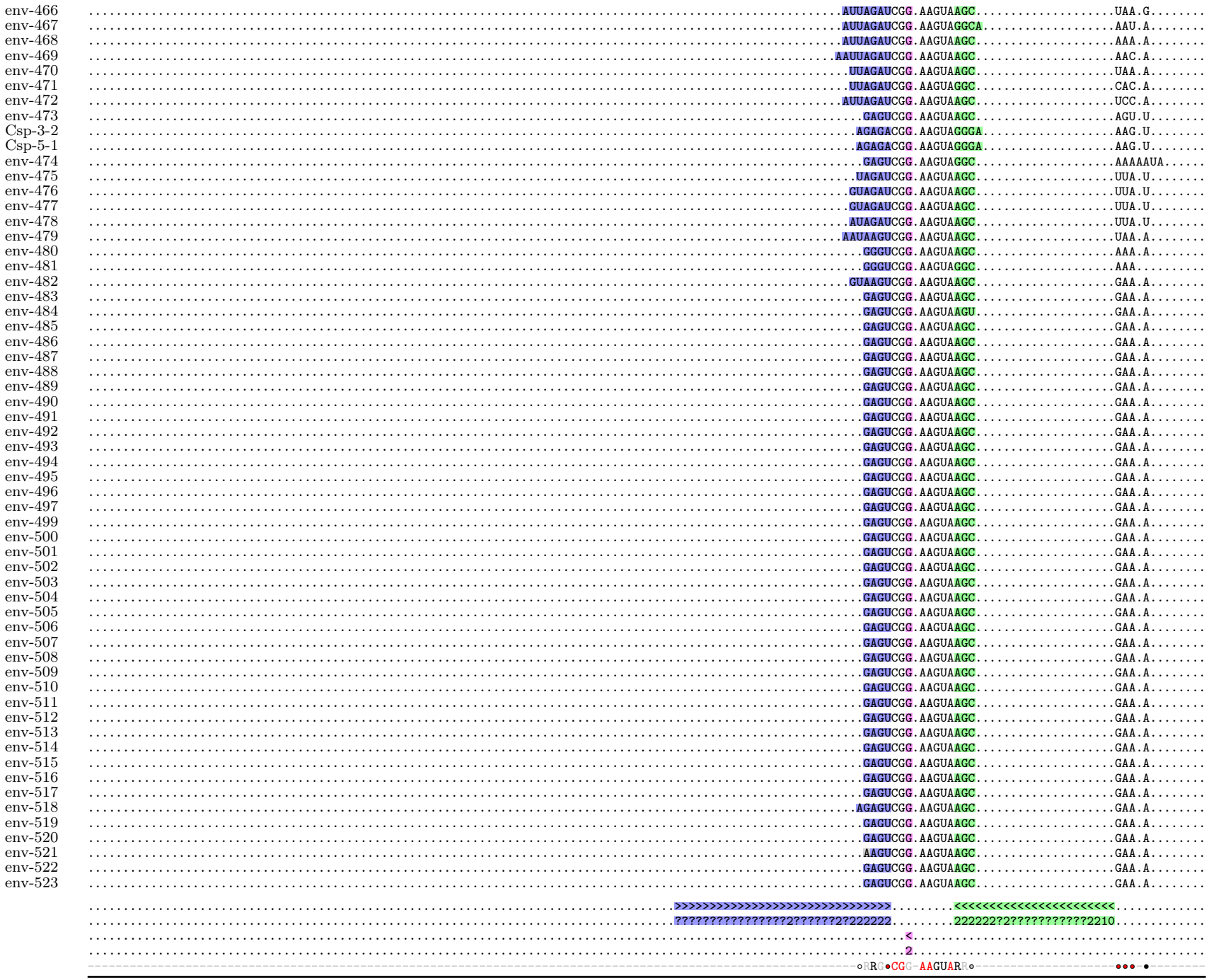

env-180
Ssp-1-1
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
env-190
env-191
env-192
env-193
env-194
env-195
env-196
env-197
env-198
env-199
env-200
env-201
env-202
env-203
env-204
env-205
env-206
env-207
env-208
env-209
env-210
env-211
env-212
env-213
env-214
env-215
Lsp-1-1
Tel-1-2
env-216
Ssp-1-2
env-217
env-218
env-219
env-220
env-221
Csp-1-3
Cwa-1-2
env-222
env-223
env-225
Csp-4-3
env-226
env-227
env-228
env-229
env-230
env-231
env-232

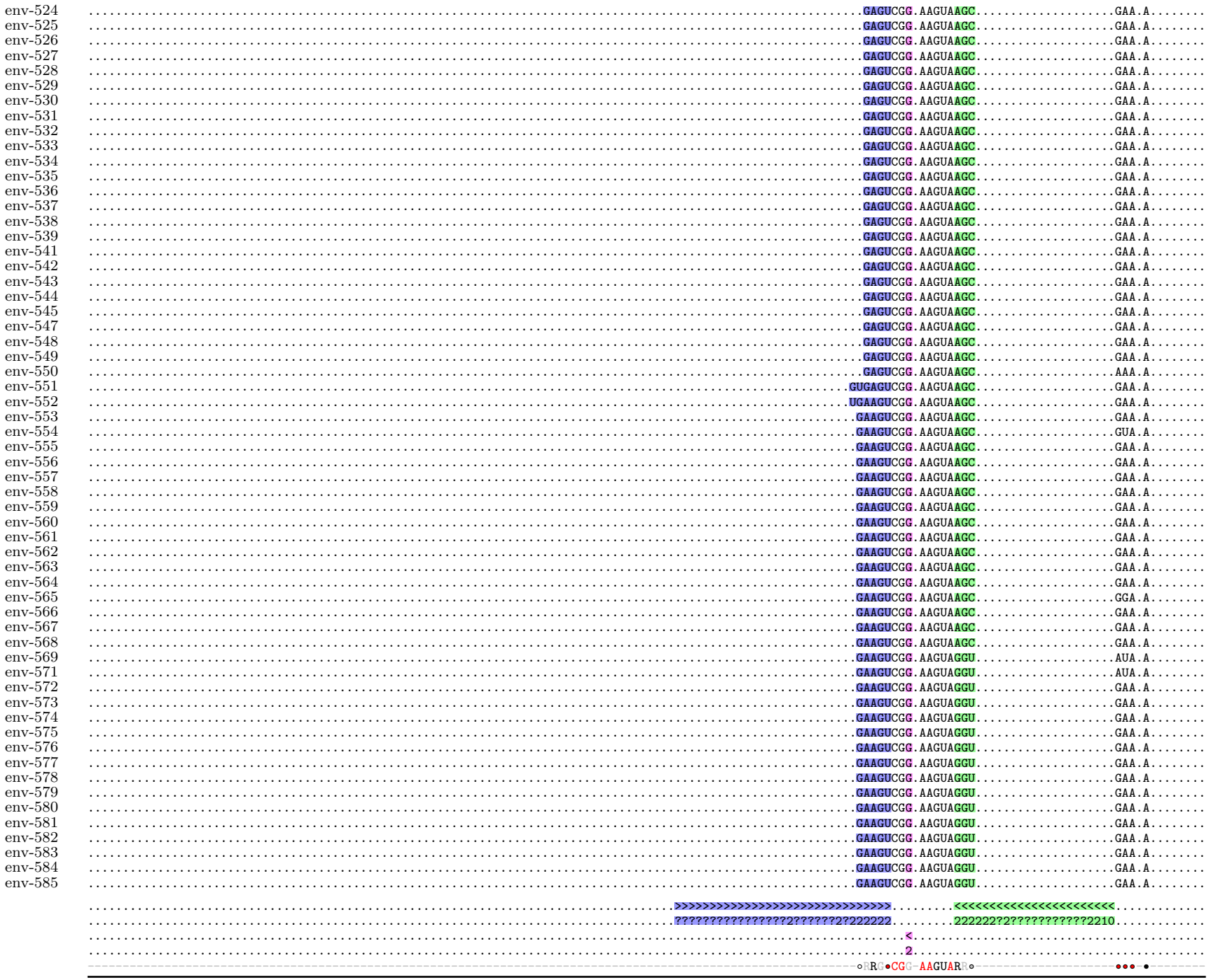

Ssp-3-1 UAAGCACUUAGUCCUAAAUGCCCAGUCUCAUCUGACAUUCAUAUGCUUAAGCAAAGCUGUGCAGCAUUAGUCGAGCUAGGCAGAAGCCAAGUGGCCUUAGGCUAG...AGAAAGACGG.AAGUAGGAGC.....CUAUA
env-233 .....UACUAAAACGG.AAGUAGAC.....GAA.A
env-234 .....AGGACGG.AAGUAGUC.....UAAAU
env-235 .....AGGACGG.AAGUAGUC.....UAAAU
env-236 .....GCUUAUUGGACGG.AAGUAGAC.....GAA.A
env-237 .....GUAGGACGG.AAGUAGAC.....GUA.A
env-238 .....GACGG.AAGUAGG.....UAAAAAAUCAAA.
env-239 .....AAAUAAGACGG.AAGUAGGAAG.....AUAGGAAAACCUC
env-240 .....AAAUAGGACGG.AAGUAGGAAG.....AUAGGAAAACCUC
env-241 .....GUAGAUCGG.AAGUAAGC.....UUCUU
env-242 .....GUAGAUCGG.AAGUAAGU.....UUAAA
env-243 .....AAGGGUCGG.AAGUAGGC.....ACC.U
env-244 .....AAGGGUCGG.AAGUAGGC.....ACC.U
env-245 .....AAGGGUCGG.AAGUAGGC.....ACC.U
env-246 .....CAAAACGG.AAGUAAAC.....GAA.A
env-247 .....UCAAAACGG.AAGUAAAC.....GAA.A
env-248 .....UCAAAACGG.AAGUAAGC.....GAA.A
env-249 .....UCAAAACGG.AAGUAAAC.....GAA.A
env-250 .....UCGAAACGG.AAGUAAAC.....GAA.A
env-251 .....UCGAAACGG.AAGUAAAC.....GAA.A
env-252 .....UCAAAACGG.AAGUAAAC.....GAA.A
env-253 .....UCAAAACGG.AAGUAGAC.....GAA.A
env-254 .....UCAAAACGG.AAGUAAAC.....GAA.A
env-255 .....UCAAAACGG.AAGUAAAC.....GAA.A
env-256 .....UCAAAACGG.AAGUAAAC.....GAA.A
env-257 .....UCAAAACGG.AAGUAAAC.....GAA.A
env-258 .....UCAAAACGG.AAGUAAAC.....GAA.A
env-259 .....UCAAAACGG.AAGUAAAC.....GAA.A
env-260 .....UCAAAACGG.AAGUAAAC.....GAA.A
env-261 .....UCAAAACGG.AAGUAAAC.....GUA.A
env-262 .....UCAAAACGG.AAGUAAAC.....AUA.A
env-263 .....CAAAACGG.AAGUAAAC.....GAA.A
env-264 .....UCAAGACGG.AAGUAGAC.....AUA.U
env-265 .....UCAAAACGG.AAGUAGAC.....AUA.A
env-266 .....CAAAACGG.AAGUAGAC.....AUA.U
env-267 .....AGGACGG.AAGUAGC.....AUAAC
env-268 .....AGGACGG.AAGUAGC.....AAU.A
env-269 .....GAUAUAGACGG.AAGUAGUC.....GUA.A
env-270 .....UAGACGG.AAGUAGC.....GAA.A
env-271 .....AAAAGACGG.AAGUAGC.....GAA.A
env-272 .....AAAAGACGG.AAGUAGC.....GAA.A
env-273 .....GGGACGG.AAGUAAGC.....CAU.AGU
env-274 .....GGACGG.AAGUAGGC.....ACC.A
env-275 .....GGACGG.AAGUAGGC.....AAC.A
env-276 .....GGACGG.AAGUAGGC.....AAC.A
env-277 .....GGACGG.AAGUAAGC.....AAC.A
env-278 .....GGACGG.AAGUAGGC.....AGC.A
env-279 .....GGUCGG.AAGUAAGC.....UUGAA
env-280 .....GGUCGG.AAGUAAGC.....UUGAA
env-281 .....GGUCGG.AAGUAAGC.....UUG.UA
env-282 .....GGUCGG.AAGUAAGC.....UUG.UA
env-283 .....GGUCGG.AAGUAAGC.....UUGUA
env-284 .....GGUCGG.AAGUAAGC.....UUGUA
env-285 .....GGUCGG.AAGUAAGC.....UUGUA
env-286 .....GGACGG.AAGUAGGC.....UUUAU
env-287 .....GGGACGG.AAGUAGGC.....AAU.C
env-288 .....GGGACGG.AAGUAGGC.....AAU.C
env-289 .....GGACGG.AAGUAGGC.....AUU.U
env-290 .....GGGACGG.AAGUAAGC.....AUA.A
>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>.....<<<<<<<<<<<<<<<<<<<<<<<<
????????????????????2???????2?222222.....222222?2?????????????2210
<.
2.
∘RRG•CGG-AAGUARR∘ ●●●-●

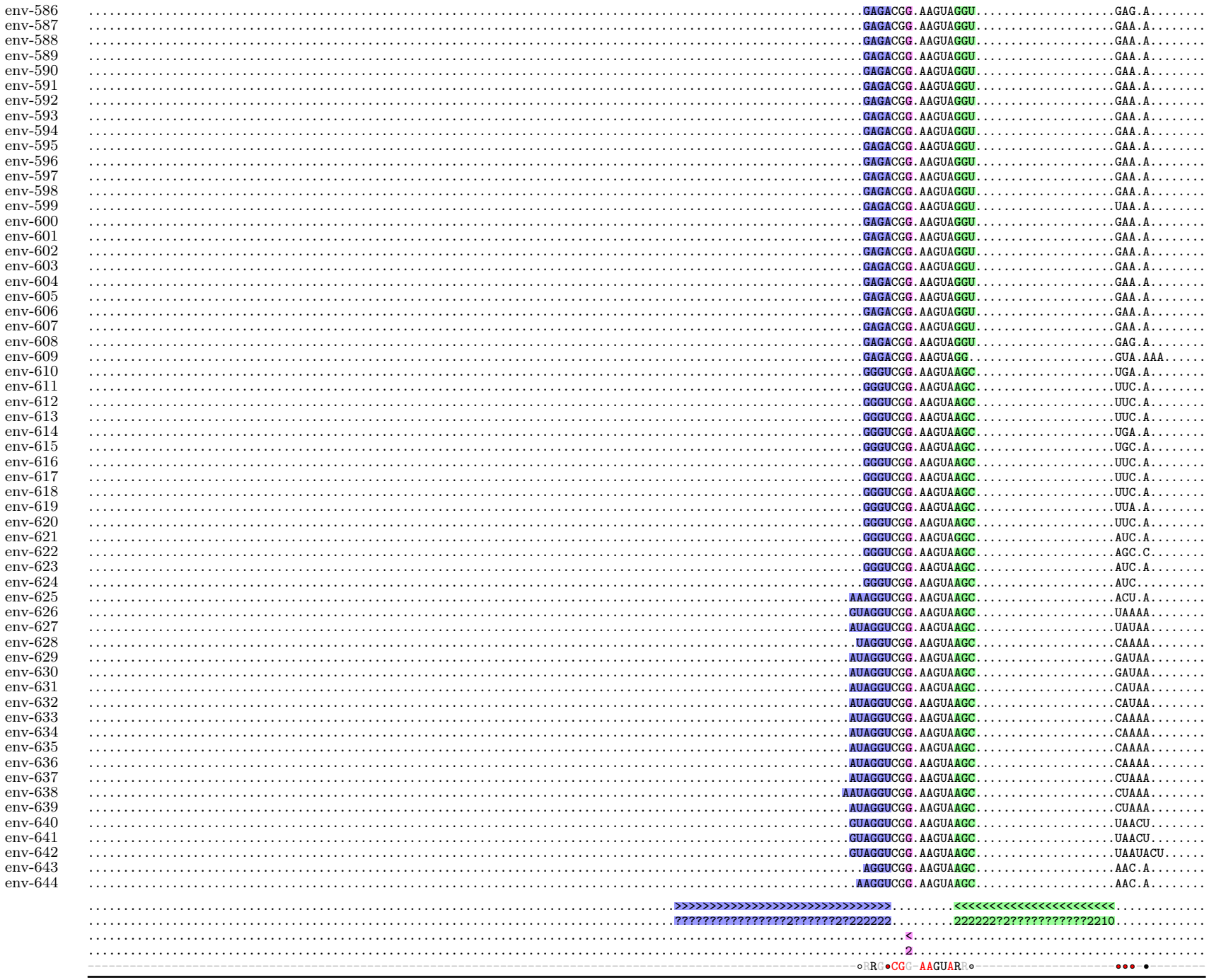

env-291
env-292
env-293
env-294
env-295
env-296
env-297
env-298
env-299
env-300
env-301
env-302
env-303
env-304
Ama-2-2
env-305
env-306
env-307
env-308
env-309
env-310
env-311
env-312
env-313
env-314
env-315
env-316
env-317
env-318
env-319
env-320
Mch-1-1
env-321
env-322
env-323
env-324
env-325
env-326
env-327
env-328
env-329
env-330
env-331
env-332
env-333
env-334
env-335
env-336
env-337
env-338
env-339
env-340
env-341
env-342
env-343
env-344
env-345
env-346
env-347

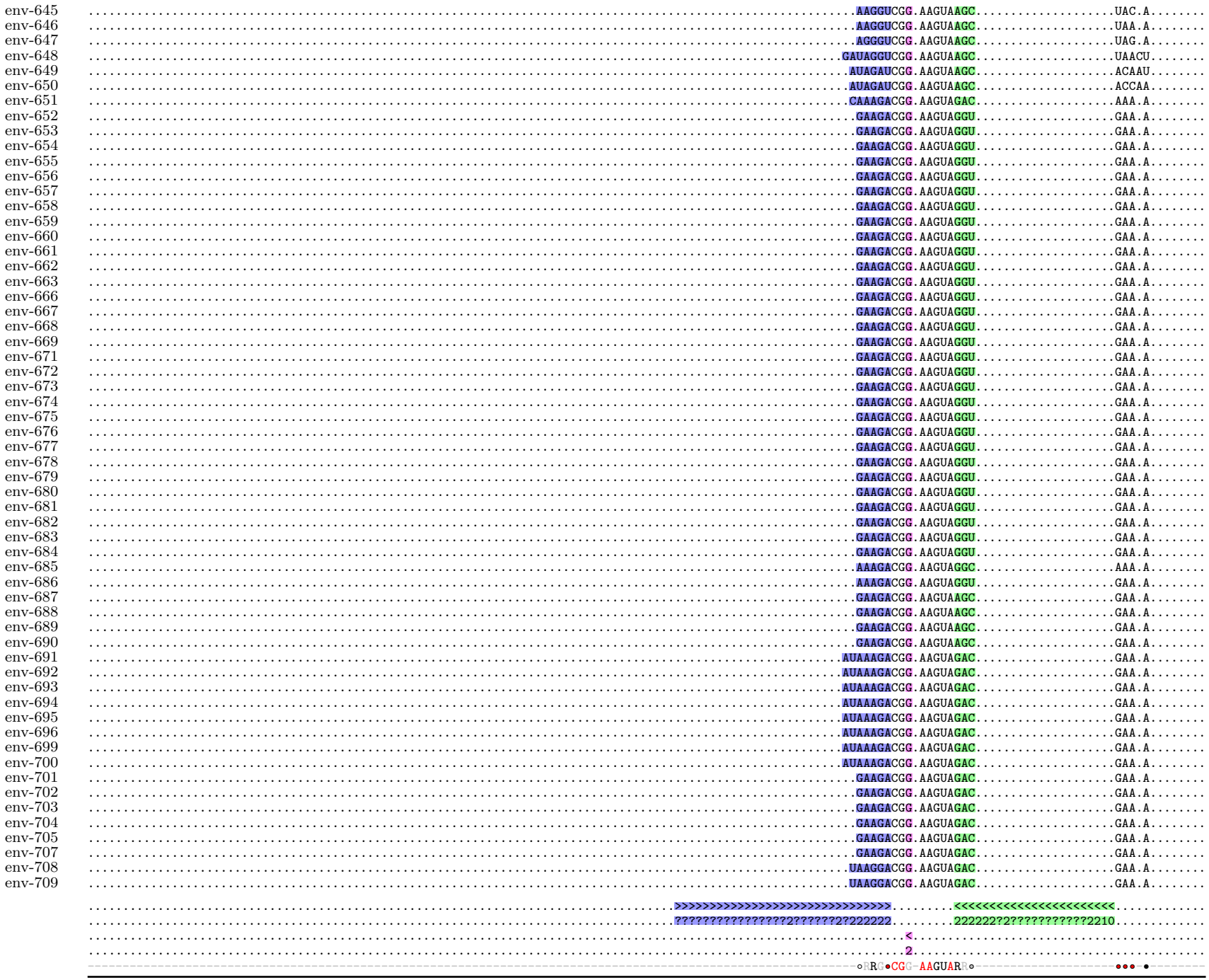

env-348
env-349
env-350
env-351
env-352
env-353
env-354
env-355
env-356
env-357
env-358
env-359
env-360
env-361
env-362
env-363
env-364
env-365
env-366
env-367
env-368
env-369
env-370
env-371
env-372
env-373
env-374
env-375
env-376
env-377
env-378
env-379
env-380
env-381
env-382
env-383
env-384
env-385
env-386
env-387
env-388
env-389
env-390
env-391
env-392
env-393
env-394
env-395
env-396
env-397
env-398
env-399
env-400
env-401
env-402
env-403
env-404
env-405
env-406

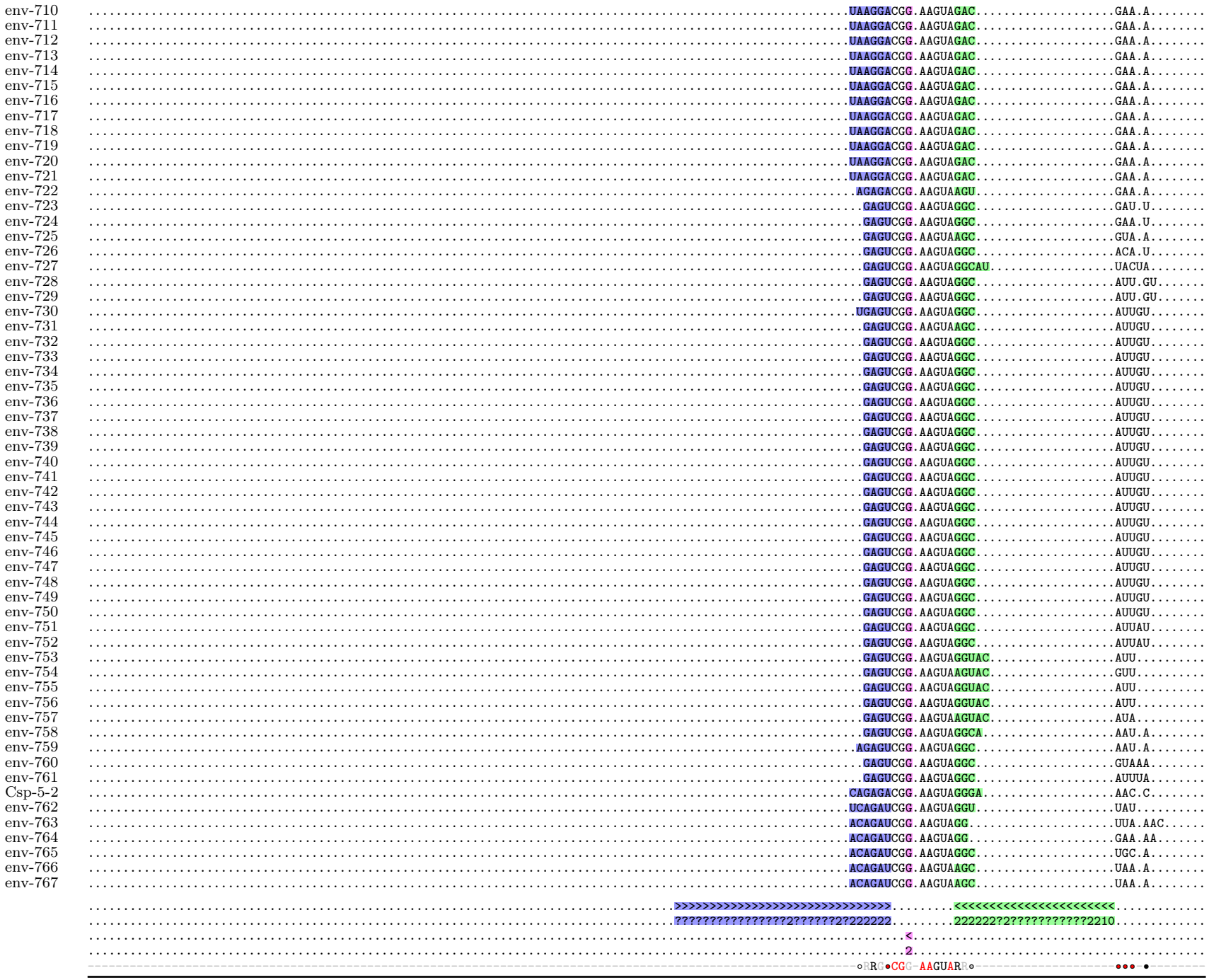

env-407
env-408
env-409
env-410
env-411
env-412
env-413
env-414
env-415
env-416
env-417
env-418
env-419
env-420
env-421
env-422
env-423
env-424
env-425
env-426
env-427
env-428
env-429
env-430
env-431
env-432
env-433
env-434
env-435
env-436
env-437
env-438
env-439
env-440
env-441
env-442
env-443
env-444
env-445
env-446
env-447
env-448
env-449
env-450
env-451
env-452
env-453
env-454
env-455
env-456
env-457
env-458
env-459
env-460
env-461
env-462
env-463
env-464
env-465

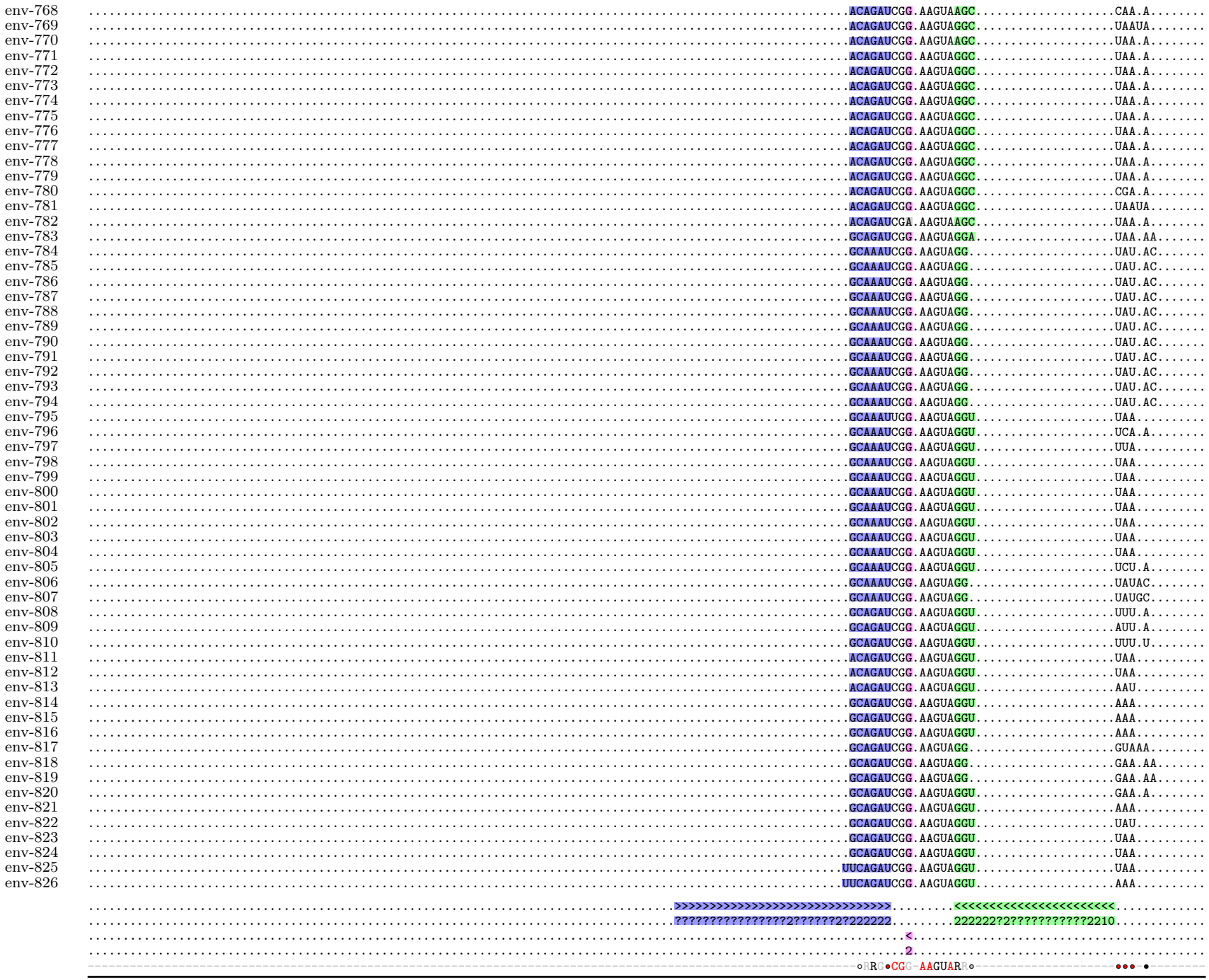

env-466
env-467
env-468
env-469
env-470
env-471
env-472
env-473
Csp-3-2
Csp-5-1
env-474
env-475
env-476
env-477
env-478
env-479
env-480
env-481
env-482
env-483
env-484
env-485
env-486
env-487
env-488
env-489
env-490
env-491
env-492
env-493
env-494
env-495
env-496
env-497
env-499
env-500
env-501
env-502
env-503
env-504
env-505
env-506
env-507
env-508
env-509
env-510
env-511
env-512
env-513
env-514
env-515
env-516
env-517
env-518
env-519
env-520
env-521
env-522
env-523

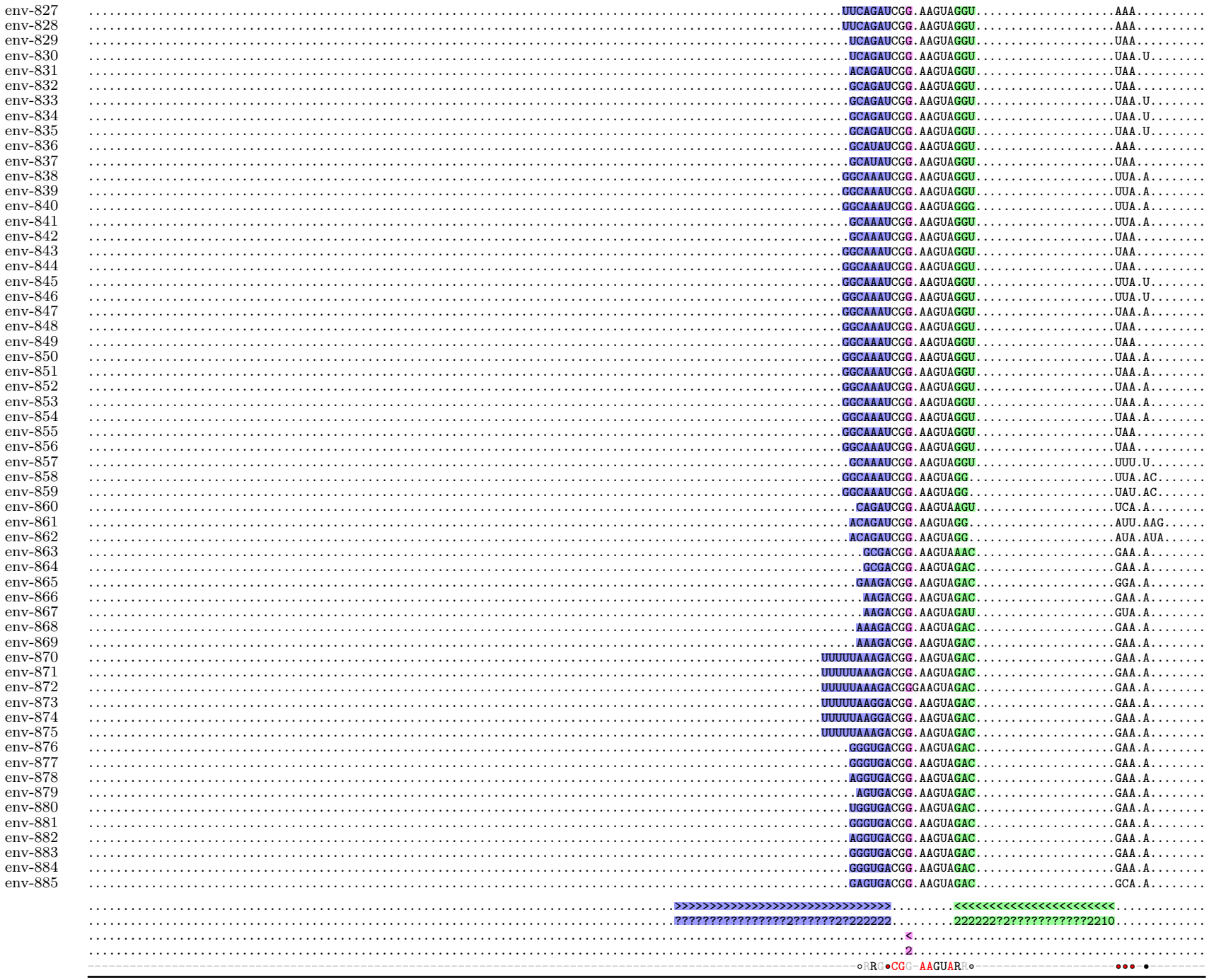

env-524 GAGUCGG.AAGUAAGC GAA.A
env-525 GAGUCGG.AAGUAAGC GAA.A
env-526 GAGUCGG.AAGUAAGC GAA.A
env-527 GAGUCGG.AAGUAAGC GAA.A
env-528 GAGUCGG.AAGUAAGC GAA.A
env-529 GAGUCGG.AAGUAAGC GAA.A
env-530 GAGUCGG.AAGUAAGC GAA.A
env-531 GAGUCGG.AAGUAAGC GAA.A
env-532 GAGUCGG.AAGUAAGC GAA.A
env-533 GAGUCGG.AAGUAAGC GAA.A
env-534 GAGUCGG.AAGUAAGC GAA.A
env-535 GAGUCGG.AAGUAAGC GAA.A
env-536 GAGUCGG.AAGUAAGC GAA.A
env-537 GAGUCGG.AAGUAAGC GAA.A
env-538 GAGUCGG.AAGUAAGC GAA.A
env-539 GAGUCGG.AAGUAAGC GAA.A
env-541 GAGUCGG.AAGUAAGC GAA.A
env-542 GAGUCGG.AAGUAAGC GAA.A
env-543 GAGUCGG.AAGUAAGC GAA.A
env-544 GAGUCGG.AAGUAAGC GAA.A
env-545 GAGUCGG.AAGUAAGC GAA.A
env-547 GAGUCGG.AAGUAAGC GAA.A
env-548 GAGUCGG.AAGUAAGC GAA.A
env-549 GAGUCGG.AAGUAAGC GAA.A
env-550 GAGUCGG.AAGUAAGC AAA.A
env-551 GUGAGUCGG.AAGUAAGC GAA.A
env-552 UGAAGUCGG.AAGUAAGC GAA.A
env-553 GAAGUCGG.AAGUAAGC GAA.A
env-554 GAAGUCGG.AAGUAAGC GUA.A
env-555 GAAGUCGG.AAGUAAGC GAA.A
env-556 GAAGUCGG.AAGUAAGC GAA.A
env-557 GAAGUCGG.AAGUAAGC GAA.A
env-558 GAAGUCGG.AAGUAAGC GAA.A
env-559 GAAGUCGG.AAGUAAGC GAA.A
env-560 GAAGUCGG.AAGUAAGC GAA.A
env-561 GAAGUCGG.AAGUAAGC GAA.A
env-562 GAAGUCGG.AAGUAAGC GAA.A
env-563 GAAGUCGG.AAGUAAGC GAA.A
env-564 GAAGUCGG.AAGUAAGC GAA.A
env-565 GAAGUCGG.AAGUAAGC GGA.A
env-566 GAAGUCGG.AAGUAAGC GAA.A
env-567 GAAGUCGG.AAGUAAGC GAA.A
env-568 GAAGUCGG.AAGUAAGC GAA.A
env-569 GAAGUCGG.AAGUAGGU AUA.A
env-571 GAAGUCGG.AAGUAGGU AUA.A
env-572 GAAGUCGG.AAGUAGGU GAA.A
env-573 GAAGUCGG.AAGUAGGU GAA.A
env-574 GAAGUCGG.AAGUAGGU GAA.A
env-575 GAAGUCGG.AAGUAGGU GAA.A
env-576 GAAGUCGG.AAGUAGGU GAA.A
env-577 GAAGUCGG.AAGUAGGU GAA.A
env-578 GAAGUCGG.AAGUAGGU GAA.A
env-579 GAAGUCGG.AAGUAGGU GAA.A
env-580 GAAGUCGG.AAGUAGGU GAA.A
env-581 GAAGUCGG.AAGUAGGU GAA.A
env-582 GAAGUCGG.AAGUAGGU GAA.A
env-583 GAAGUCGG.AAGUAGGU GAA.A
env-584 GAAGUCGG.AAGUAGGU GAA.A
env-585 GAAGUCGG.AAGUAGGU GAA.A
>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>> <<<<<<<<<<<<<<<<<<<<<<<<
????????????????????2???????2?222222 222222?2?????????????2210
<
2
oRRG•CGG–AAGUARRo •••–•

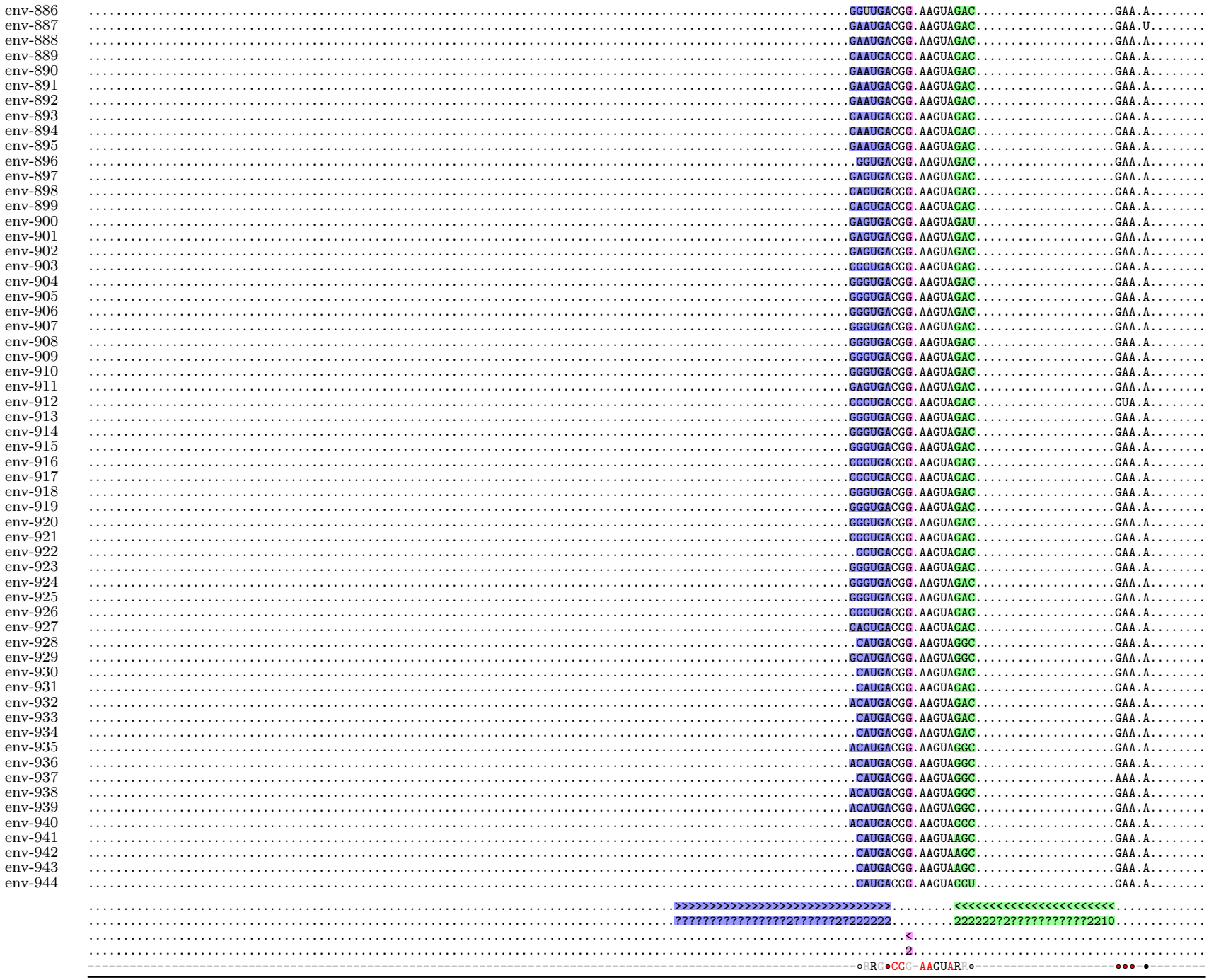

env-586
env-587
env-588
env-589
env-590
env-591
env-592
env-593
env-594
env-595
env-596
env-597
env-598
env-599
env-600
env-601
env-602
env-603
env-604
env-605
env-606
env-607
env-608
env-609
env-610
env-611
env-612
env-613
env-614
env-615
env-616
env-617
env-618
env-619
env-620
env-621
env-622
env-623
env-624
env-625
env-626
env-627
env-628
env-629
env-630
env-631
env-632
env-633
env-634
env-635
env-636
env-637
env-638
env-639
env-640
env-641
env-642
env-643
env-644

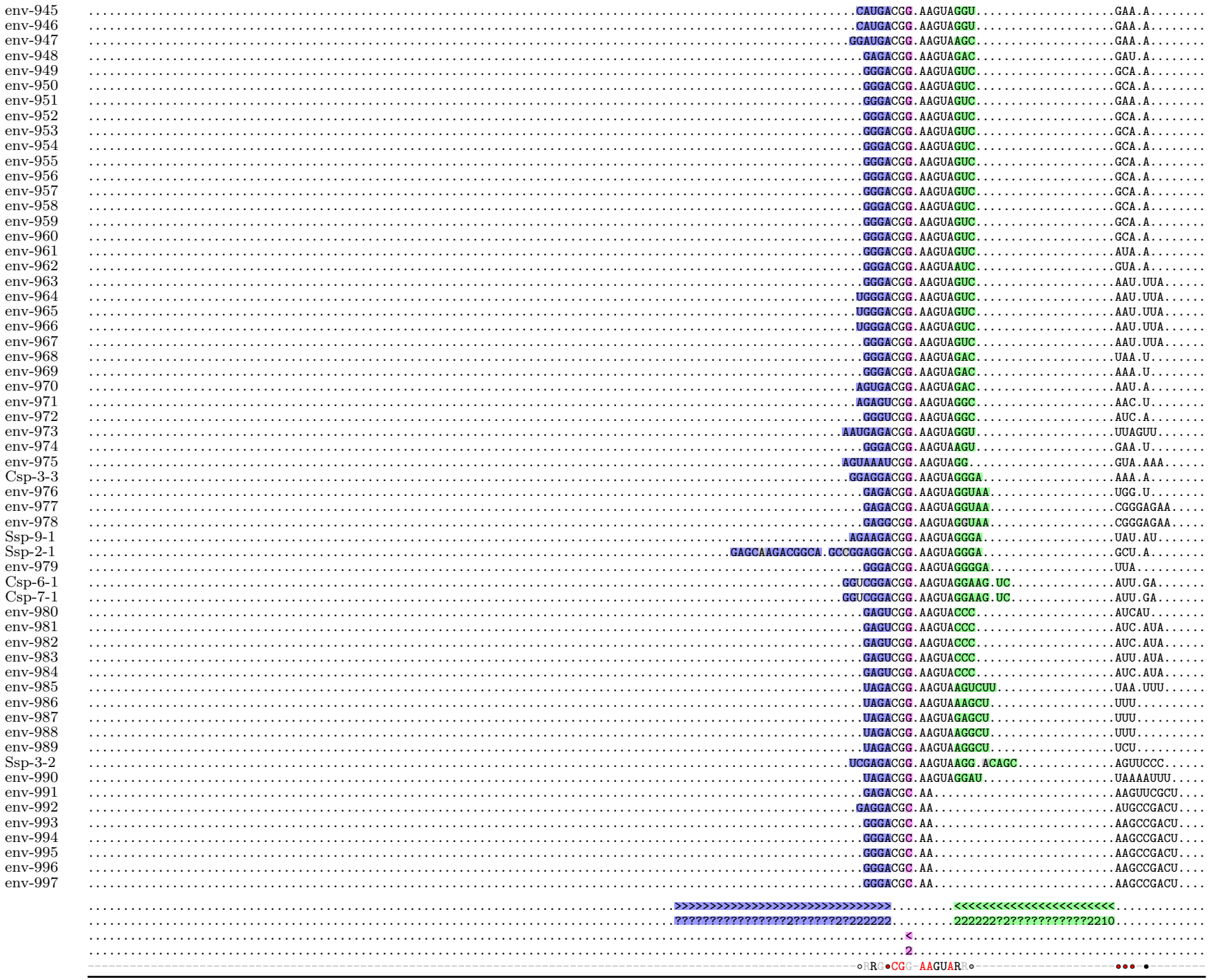

env-645
env-646
env-647
env-648
env-649
env-650
env-651
env-652
env-653
env-654
env-655
env-656
env-657
env-658
env-659
env-660
env-661
env-662
env-663
env-666
env-667
env-668
env-669
env-671
env-672
env-673
env-674
env-675
env-676
env-677
env-678
env-679
env-680
env-681
env-682
env-683
env-684
env-685
env-686
env-687
env-688
env-689
env-690
env-691
env-692
env-693
env-694
env-695
env-696
env-699
env-700
env-701
env-702
env-703
env-704
env-705
env-707
env-708
env-709

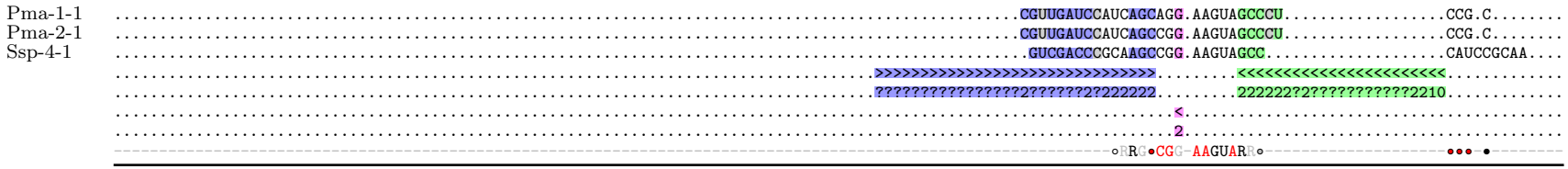

env-710
env-711
env-712
env-713
env-714
env-715
env-716
env-717
env-718
env-719
env-720
env-721
env-722
env-723
env-724
env-725
env-726
env-727
env-728
env-729
env-730
env-731
env-732
env-733
env-734
env-735
env-736
env-737
env-738
env-739
env-740
env-741
env-742
env-743
env-744
env-745
env-746
env-747
env-748
env-749
env-750
env-751
env-752
env-753
env-754
env-755
env-756
env-757
env-758
env-759
env-760
env-761
Csp-5-2
env-762
env-763
env-764
env-765
env-766
env-767


env-768
env-769
env-770
env-771
env-772
env-773
env-774
env-775
env-776
env-777
env-778
env-779
env-780
env-781
env-782
env-783
env-784
env-785
env-786
env-787
env-788
env-789
env-790
env-791
env-792
env-793
env-794
env-795
env-796
env-797
env-798
env-799
env-800
env-801
env-802
env-803
env-804
env-805
env-806
env-807
env-808
env-809
env-810
env-811
env-812
env-813
env-814
env-815
env-816
env-817
env-818
env-819
env-820
env-821
env-822
env-823
env-824
env-825
env-826

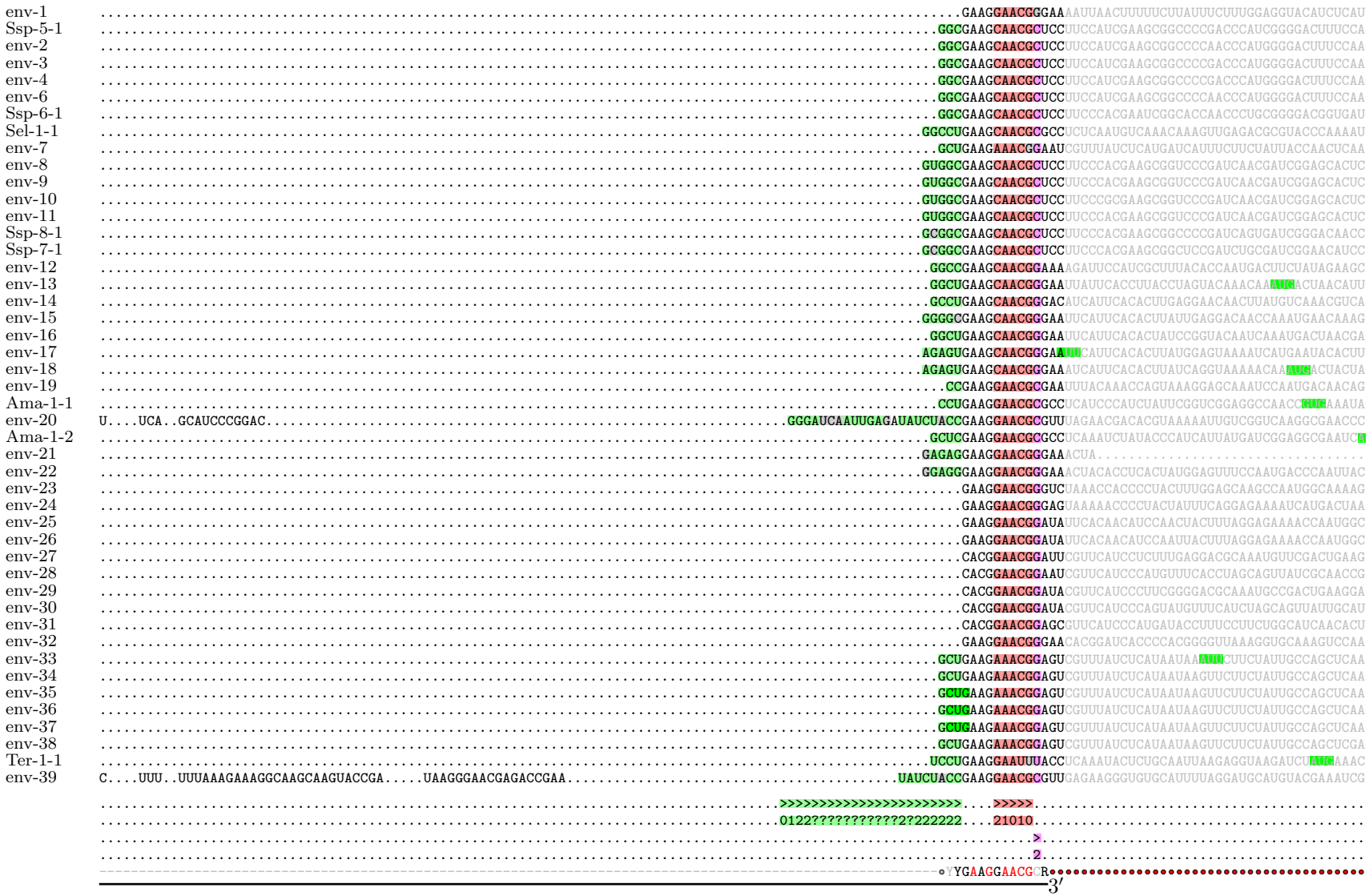

env-827
env-828
env-829
env-830
env-831
env-832
env-833
env-834
env-835
env-836
env-837
env-838
env-839
env-840
env-841
env-842
env-843
env-844
env-845
env-846
env-847
env-848
env-849
env-850
env-851
env-852
env-853
env-854
env-855
env-856
env-857
env-858
env-859
env-860
env-861
env-862
env-863
env-864
env-865
env-866
env-867
env-868
env-869
env-870
env-871
env-872
env-873
env-874
env-875
env-876
env-877
env-878
env-879
env-880
env-881
env-882
env-883
env-884
env-885

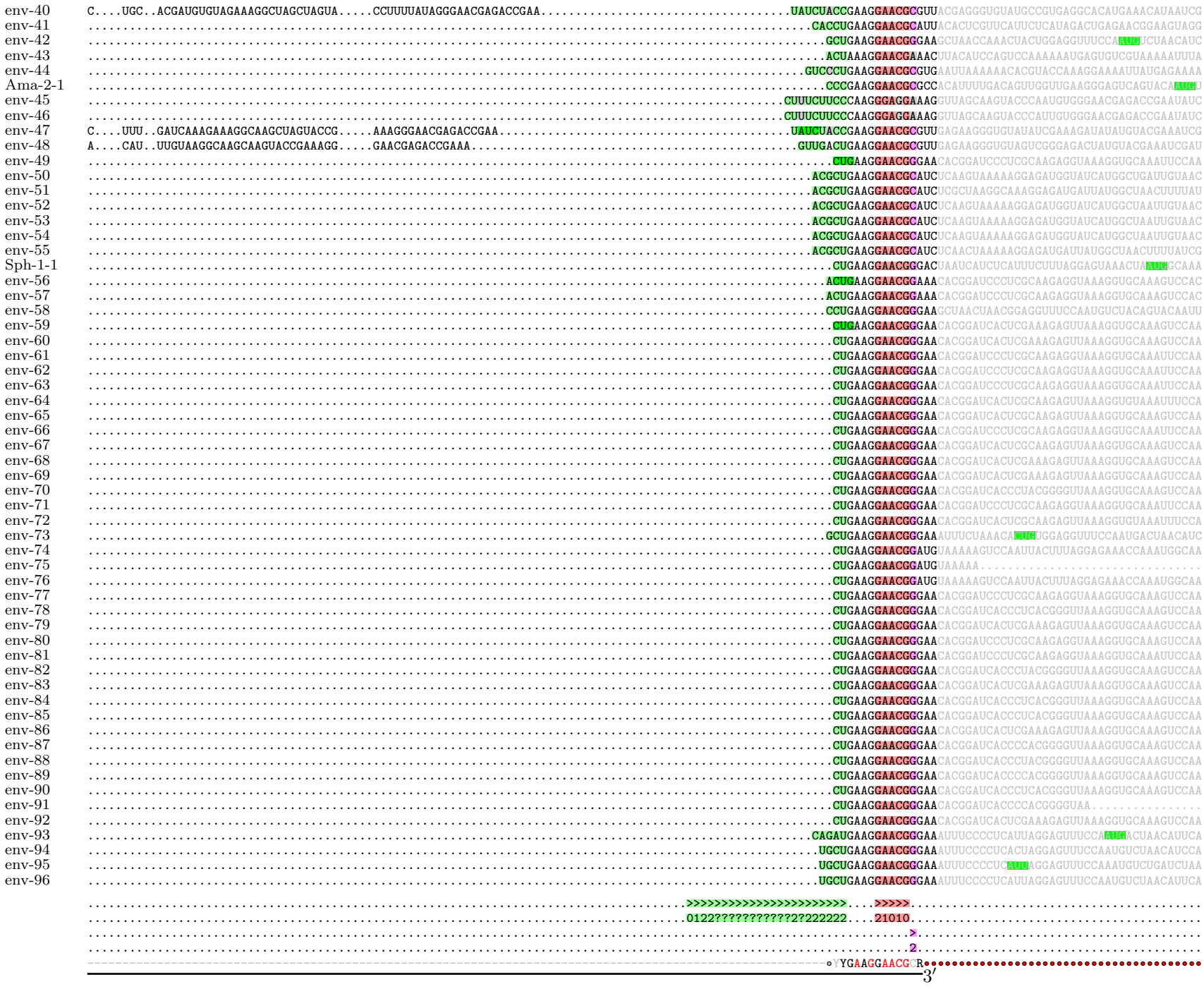

env-886 ..... GGUUGACGG.AAGUAGAC..... GAA.A.....
env-887 ..... GAAUGACGG.AAGUAGAC..... GAA.U.....
env-888 ..... GAAUGACGG.AAGUAGAC..... GAA.A.....
env-889 ..... GAAUGACGG.AAGUAGAC..... GAA.A.....
env-890 ..... GAAUGACGG.AAGUAGAC..... GAA.A.....
env-891 ..... GAAUGACGG.AAGUAGAC..... GAA.A.....
env-892 ..... GAAUGACGG.AAGUAGAC..... GAA.A.....
env-893 ..... GAAUGACGG.AAGUAGAC..... GAA.A.....
env-894 ..... GAAUGACGG.AAGUAGAC..... GAA.A.....
env-895 ..... GAAUGACGG.AAGUAGAC..... GAA.A.....
env-896 ..... GGUGACGG.AAGUAGAC..... GAA.A.....
env-897 ..... GAGUGACGG.AAGUAGAC..... GAA.A.....
env-898 ..... GAGUGACGG.AAGUAGAC..... GAA.A.....
env-899 ..... GAGUGACGG.AAGUAGAC..... GAA.A.....
env-900 ..... GAGUGACGG.AAGUAGAU..... GAA.A.....
env-901 ..... GAGUGACGG.AAGUAGAC..... GAA.A.....
env-902 ..... GAGUGACGG.AAGUAGAC..... GAA.A.....
env-903 ..... GGGUGACGG.AAGUAGAC..... GAA.A.....
env-904 ..... GGGUGACGG.AAGUAGAC..... GAA.A.....
env-905 ..... GGGUGACGG.AAGUAGAC..... GAA.A.....
env-906 ..... GGGUGACGG.AAGUAGAC..... GAA.A.....
env-907 ..... GGGUGACGG.AAGUAGAC..... GAA.A.....
env-908 ..... GGGUGACGG.AAGUAGAC..... GAA.A.....
env-909 ..... GGGUGACGG.AAGUAGAC..... GAA.A.....
env-910 ..... GGGUGACGG.AAGUAGAC..... GAA.A.....
env-911 ..... GAGUGACGG.AAGUAGAC..... GAA.A.....
env-912 ..... GGGUGACGG.AAGUAGAC..... GUA.A.....
env-913 ..... GGGUGACGG.AAGUAGAC..... GAA.A.....
env-914 ..... GGGUGACGG.AAGUAGAC..... GAA.A.....
env-915 ..... GGGUGACGG.AAGUAGAC..... GAA.A.....
env-916 ..... GGGUGACGG.AAGUAGAC..... GAA.A.....
env-917 ..... GGGUGACGG.AAGUAGAC..... GAA.A.....
env-918 ..... GGGUGACGG.AAGUAGAC..... GAA.A.....
env-919 ..... GGGUGACGG.AAGUAGAC..... GAA.A.....
env-920 ..... GGGUGACGG.AAGUAGAC..... GAA.A.....
env-921 ..... GGGUGACGG.AAGUAGAC..... GAA.A.....
env-922 ..... GGUGACGG.AAGUAGAC..... GAA.A.....
env-923 ..... GGGUGACGG.AAGUAGAC..... GAA.A.....
env-924 ..... GGGUGACGG.AAGUAGAC..... GAA.A.....
env-925 ..... GGGUGACGG.AAGUAGAC..... GAA.A.....
env-926 ..... GGGUGACGG.AAGUAGAC..... GAA.A.....
env-927 ..... GAGUGACGG.AAGUAGAC..... GAA.A.....
env-928 ..... CAUGACGG.AAGUAGGC..... GAA.A.....
env-929 ..... GCAUGACGG.AAGUAGGC..... GAA.A.....
env-930 ..... CAUGACGG.AAGUAGAC..... GAA.A.....
env-931 ..... CAUGACGG.AAGUAGAC..... GAA.A.....
env-932 ..... ACAUGACGG.AAGUAGAC..... GAA.A.....
env-933 ..... CAUGACGG.AAGUAGAC..... GAA.A.....
env-934 ..... CAUGACGG.AAGUAGAC..... GAA.A.....
env-935 ..... ACAUGACGG.AAGUAGGC..... GAA.A.....
env-936 ..... ACAUGACGG.AAGUAGGC..... GAA.A.....
env-937 ..... CAUGACGG.AAGUAGGC..... AAA.A.....
env-938 ..... ACAUGACGG.AAGUAGGC..... GAA.A.....
env-939 ..... ACAUGACGG.AAGUAGGC..... GAA.A.....
env-940 ..... ACAUGACGG.AAGUAGGC..... GAA.A.....
env-941 ..... CAUGACGG.AAGUAAGC..... GAA.A.....
env-942 ..... CAUGACGG.AAGUAAGC..... GAA.A.....
env-943 ..... CAUGACGG.AAGUAAGC..... GAA.A.....
env-944 ..... CAUGACGG.AAGUAGGU..... GAA.A.....
..... >>>>

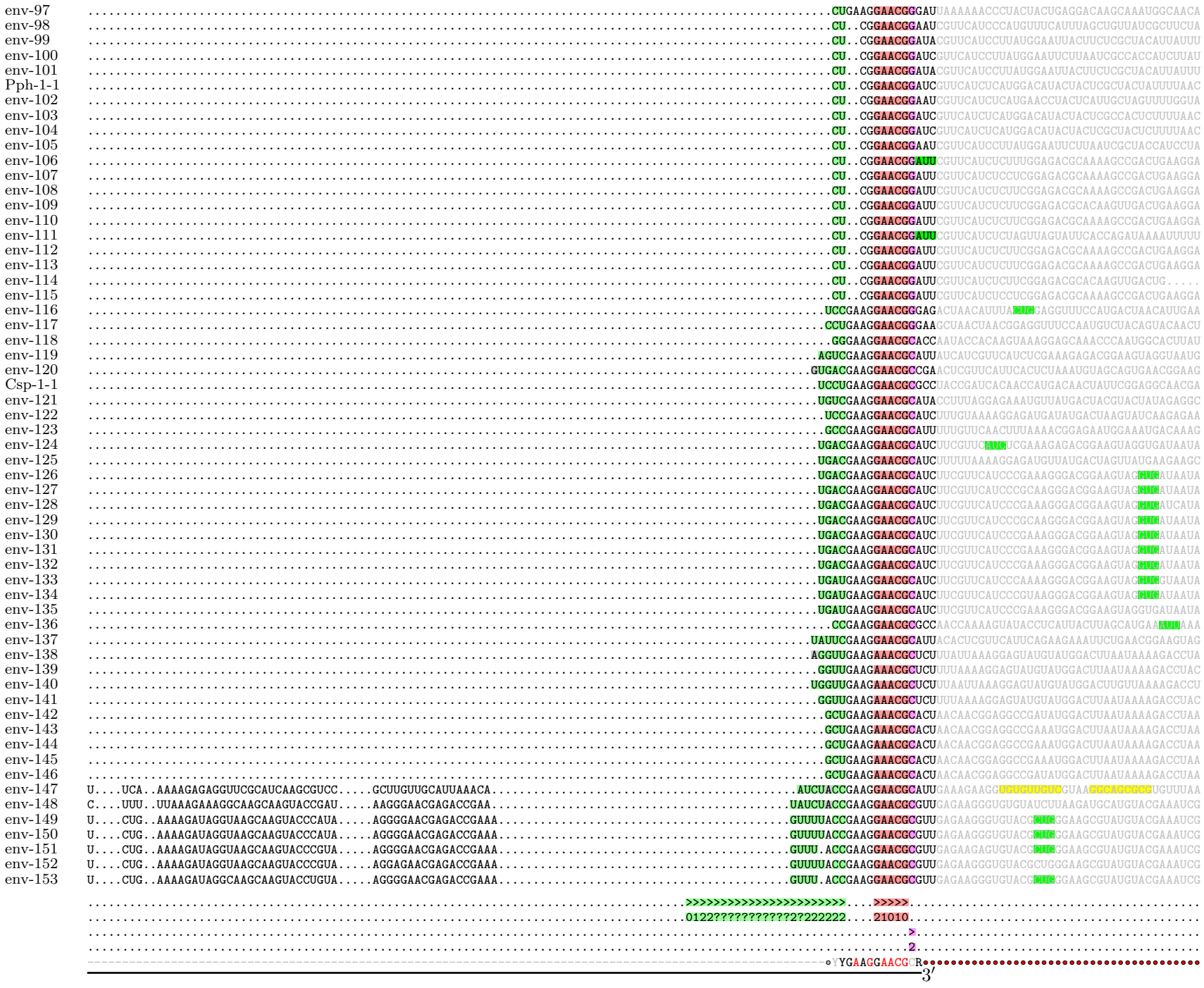

env-945
env-946
env-947
env-948
env-949
env-950
env-951
env-952
env-953
env-954
env-955
env-956
env-957
env-958
env-959
env-960
env-961
env-962
env-963
env-964
env-965
env-966
env-967
env-968
env-969
env-970
env-971
env-972
env-973
env-974
env-975
Csp-3-3
env-976
env-977
env-978
Ssp-9-1
Ssp-2-1
env-979
Csp-6-1
Csp-7-1
env-980
env-981
env-982
env-983
env-984
env-985
env-986
env-987
env-988
env-989
Ssp-3-2
env-990
env-991
env-992
env-993
env-994
env-995
env-996
env-997

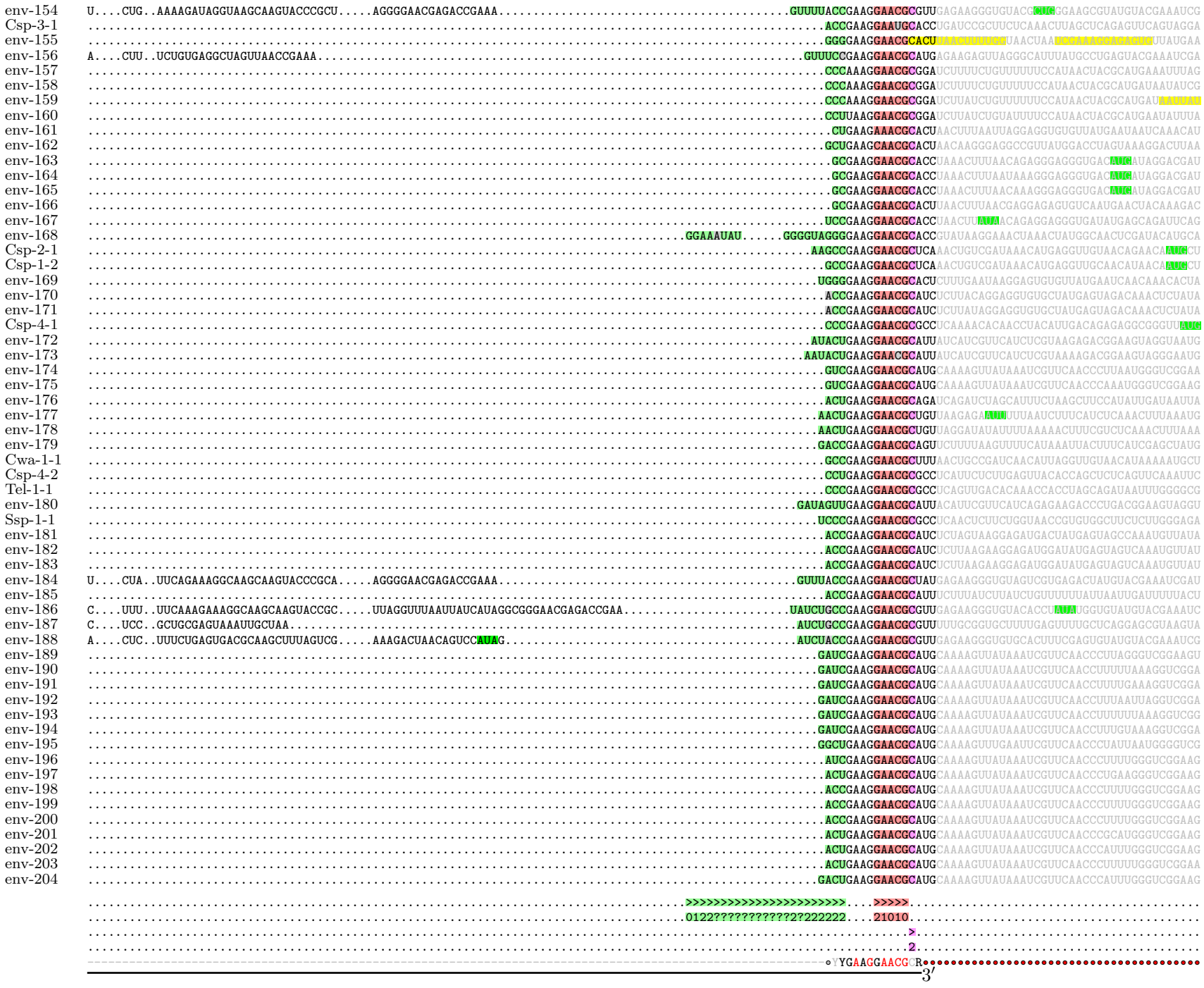


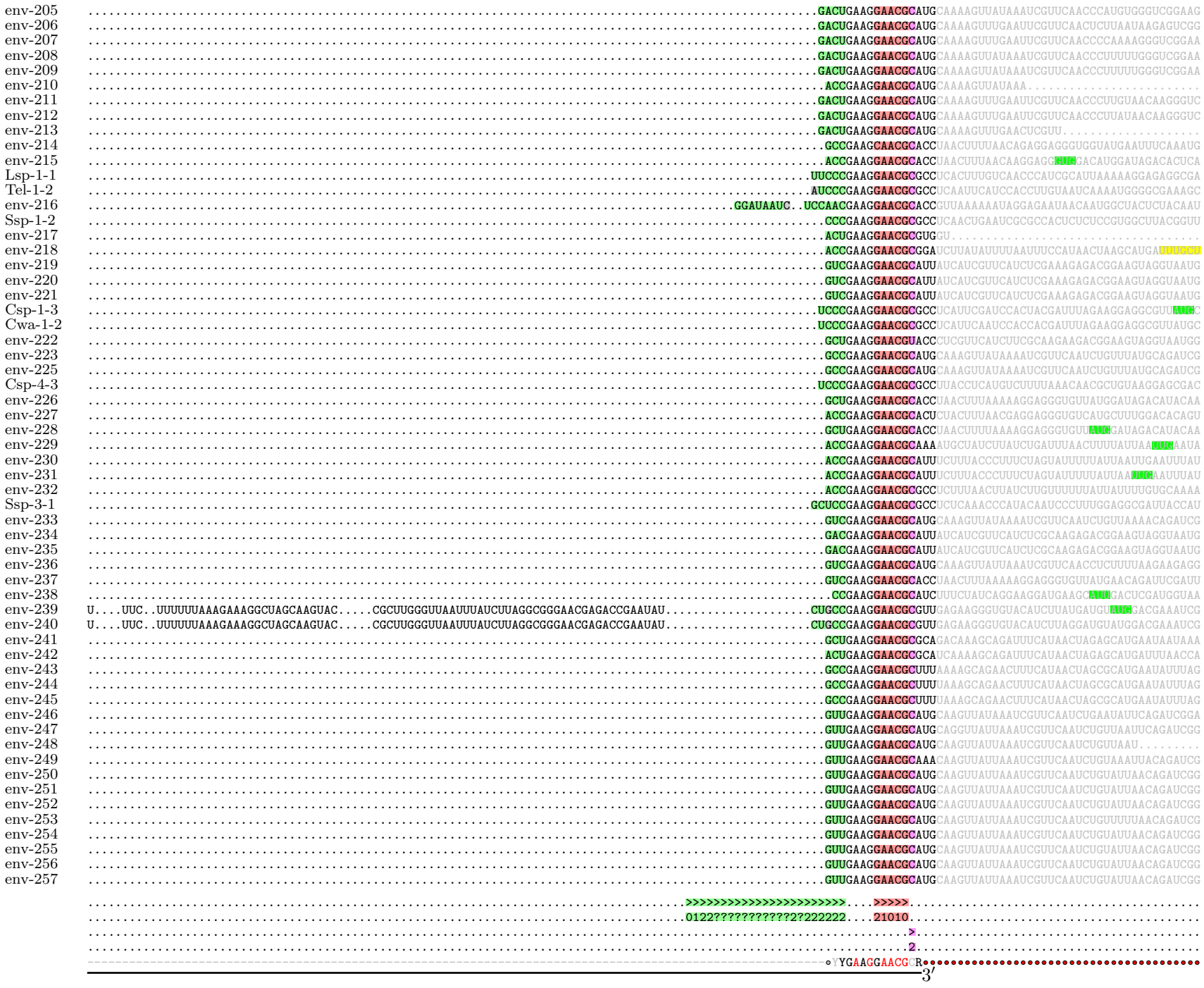
alignment positions 321 · · · 480

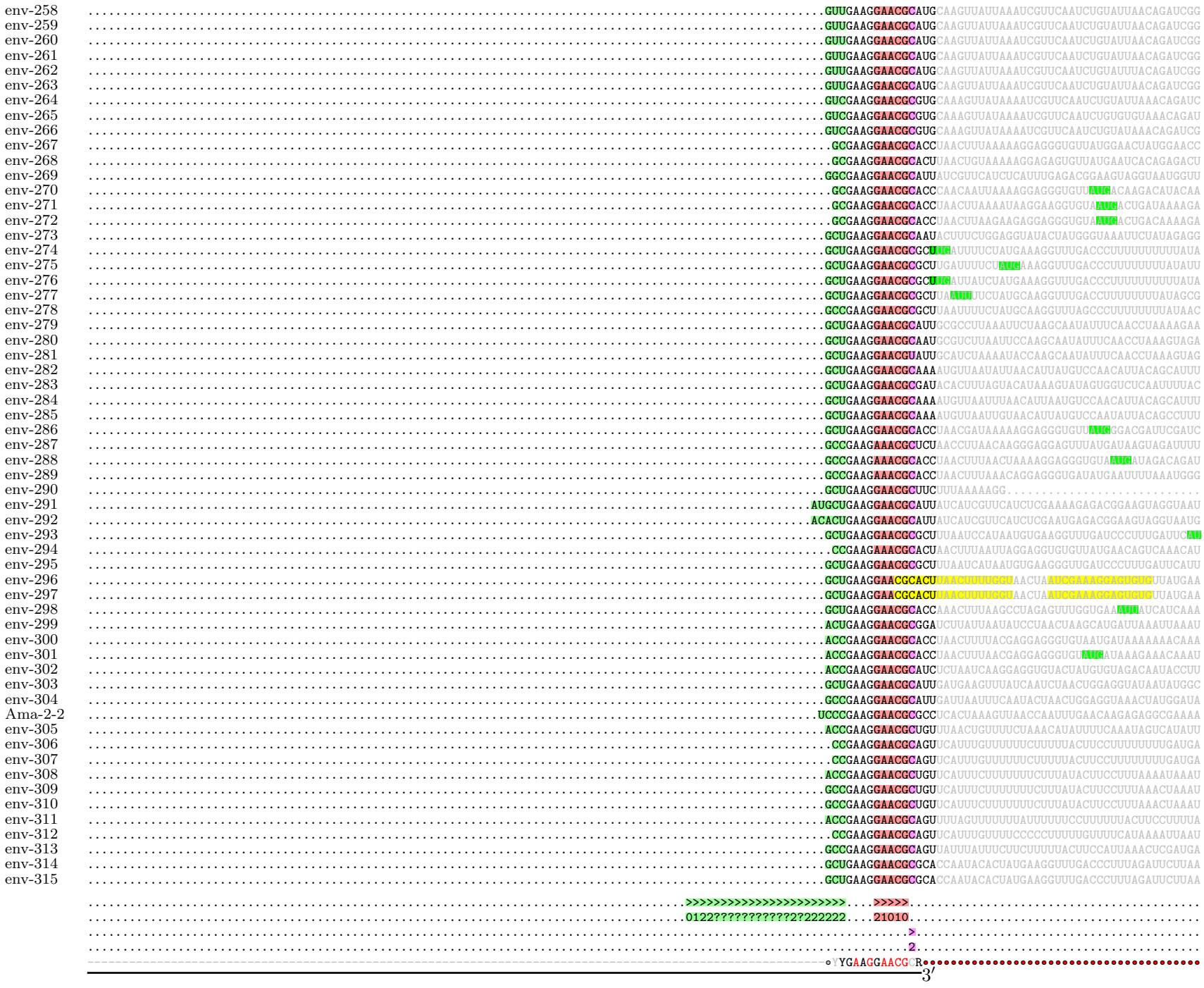


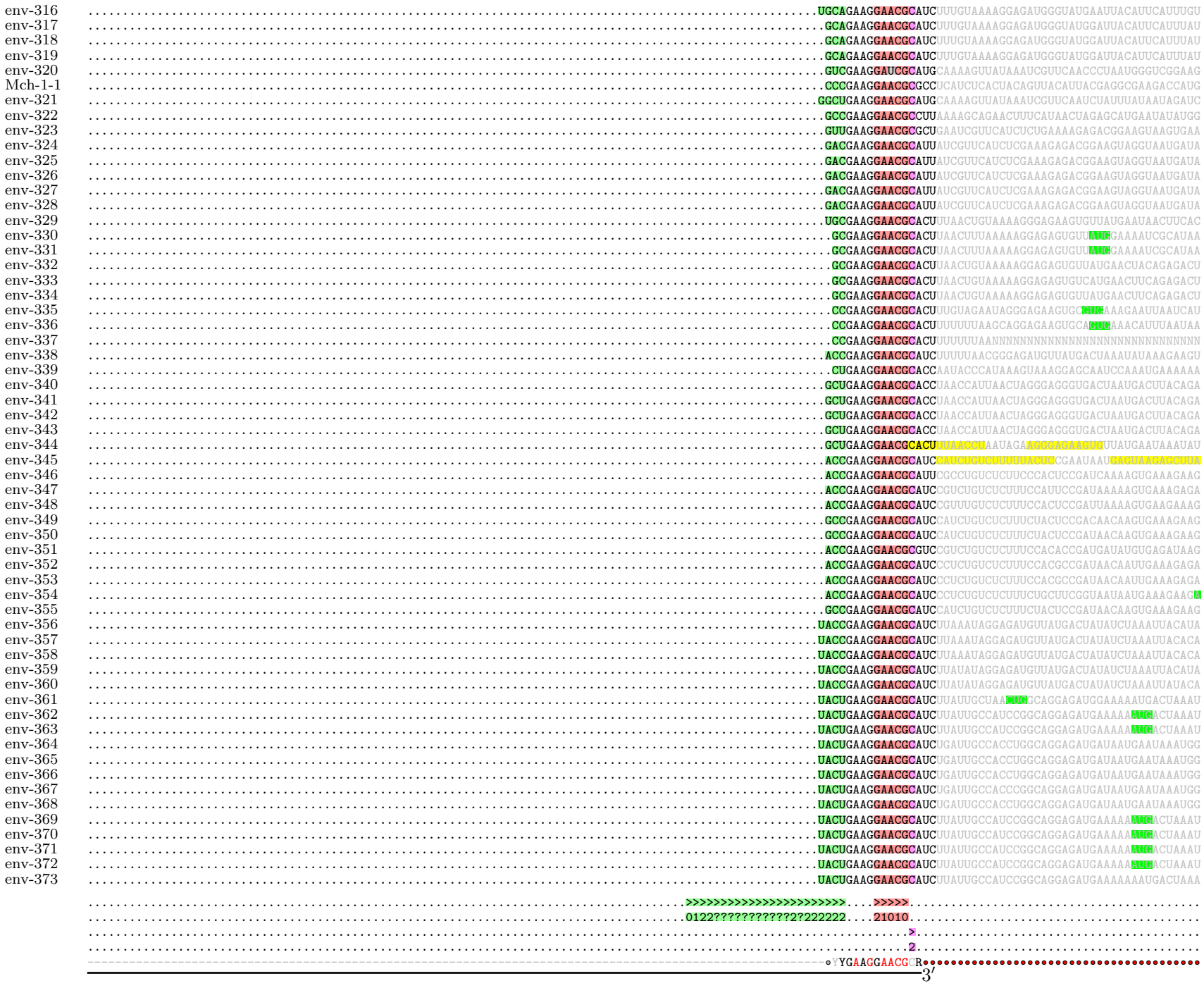

env-40
Ama-2-1
Sph-1-1
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
3′

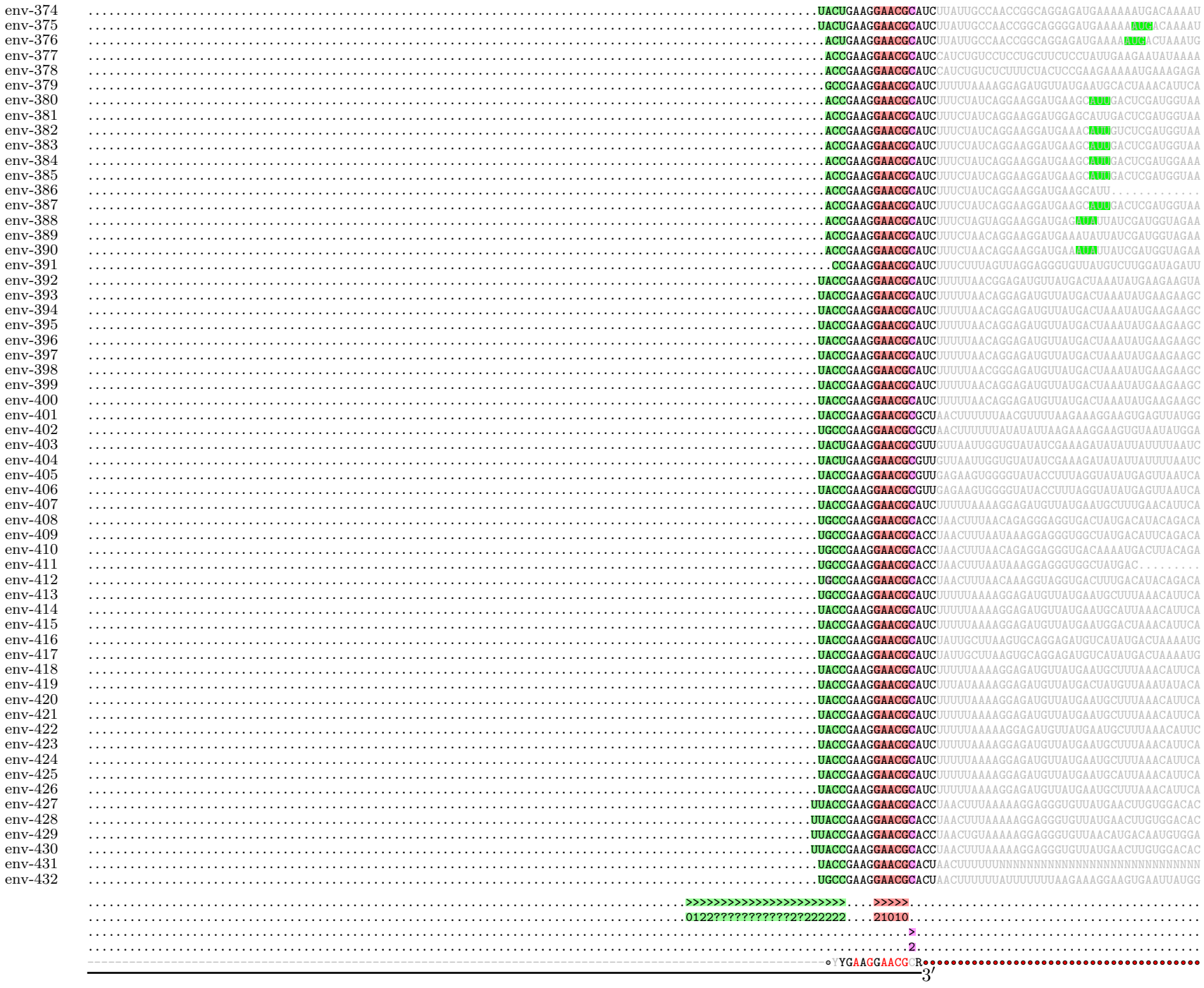

env-97
env-98
env-99
env-100
env-101
Pph-1-1
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
Csp-1-1
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-135
env-136
env-137
env-138
env-139
env-140
env-141
env-142
env-143
env-144
env-145
env-146
env-147
env-148
env-149
env-150
env-151
env-152
env-153
3′

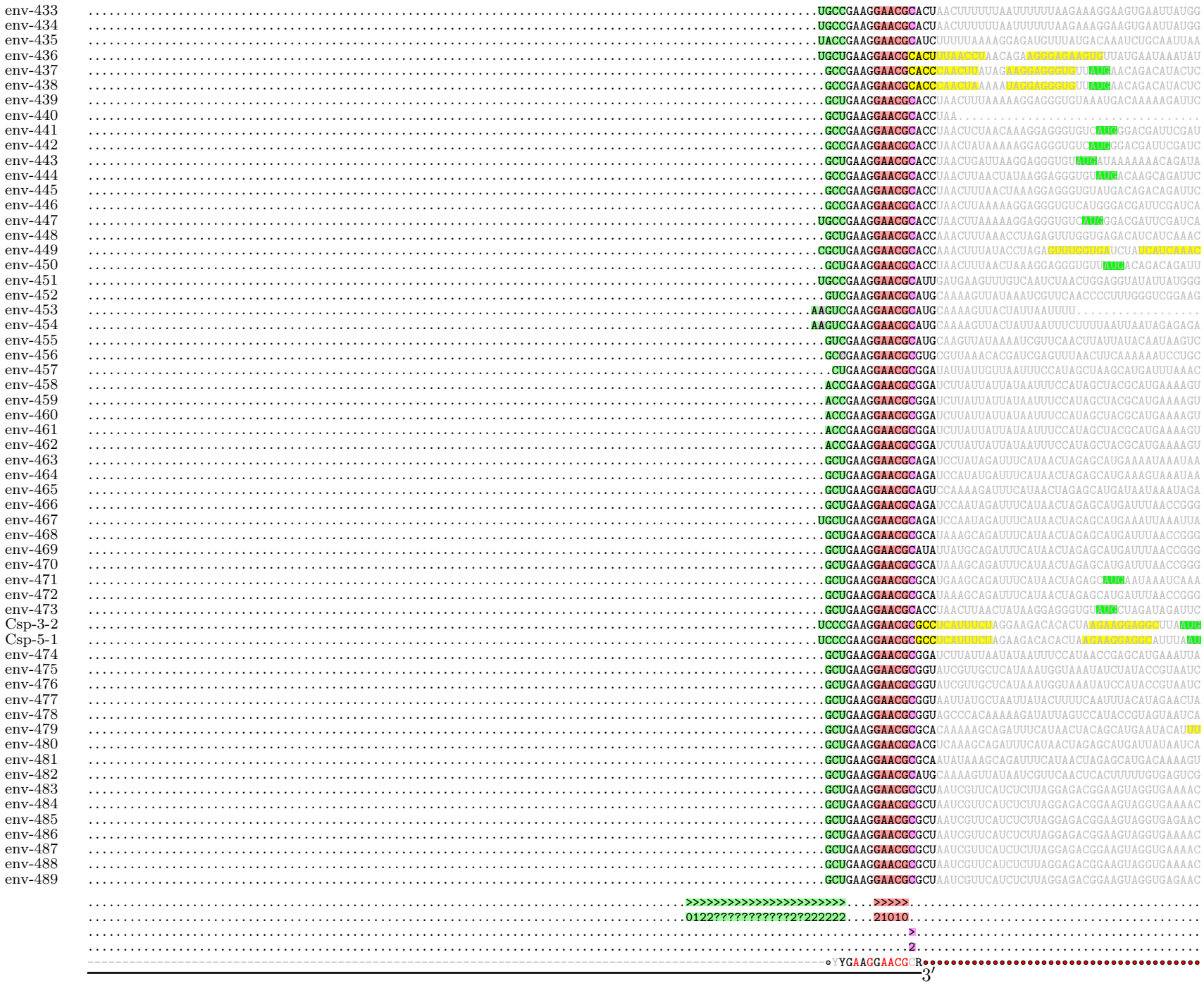

env-154
Csp-3-1
env-155
env-156
env-157
env-158
env-159
env-160
env-161
env-162
env-163
env-164
env-165
env-166
env-167
env-168
Csp-2-1
Csp-1-2
env-169
env-170
env-171
Csp-4-1
env-172
env-173
env-174
env-175
env-176
env-177
env-178
env-179
Cwa-1-1
Csp-4-2
Tel-1-1
env-180
Ssp-1-1
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
env-190
env-191
env-192
env-193
env-194
env-195
env-196
env-197
env-198
env-199
env-200
env-201
env-202
env-203
env-204
3′

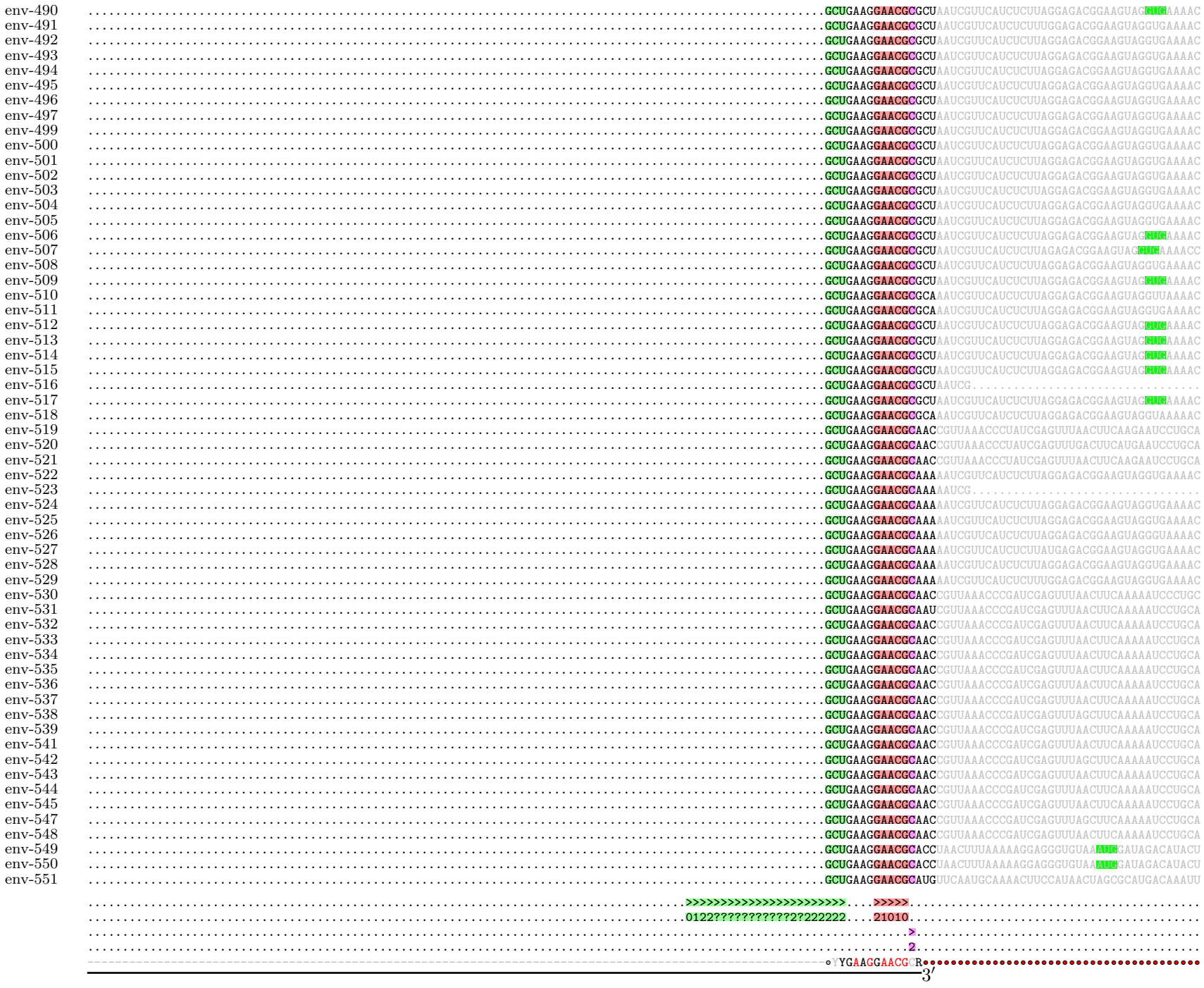

env-205
env-206
env-207
env-208
env-209
env-210
env-211
env-212
env-213
env-214
env-215
Lsp-1-1
Tel-1-2
env-216
Ssp-1-2
env-217
env-218
env-219
env-220
env-221
Csp-1-3
Cwa-1-2
env-222
env-223
env-225
Csp-4-3
env-226
env-227
env-228
env-229
env-230
env-231
env-232
Ssp-3-1
env-233
env-234
env-235
env-236
env-237
env-238
env-239
env-240
env-241
env-242
env-243
env-244
env-245
env-246
env-247
env-248
env-249
env-250
env-251
env-252
env-253
env-254
env-255
env-256
env-257
3′

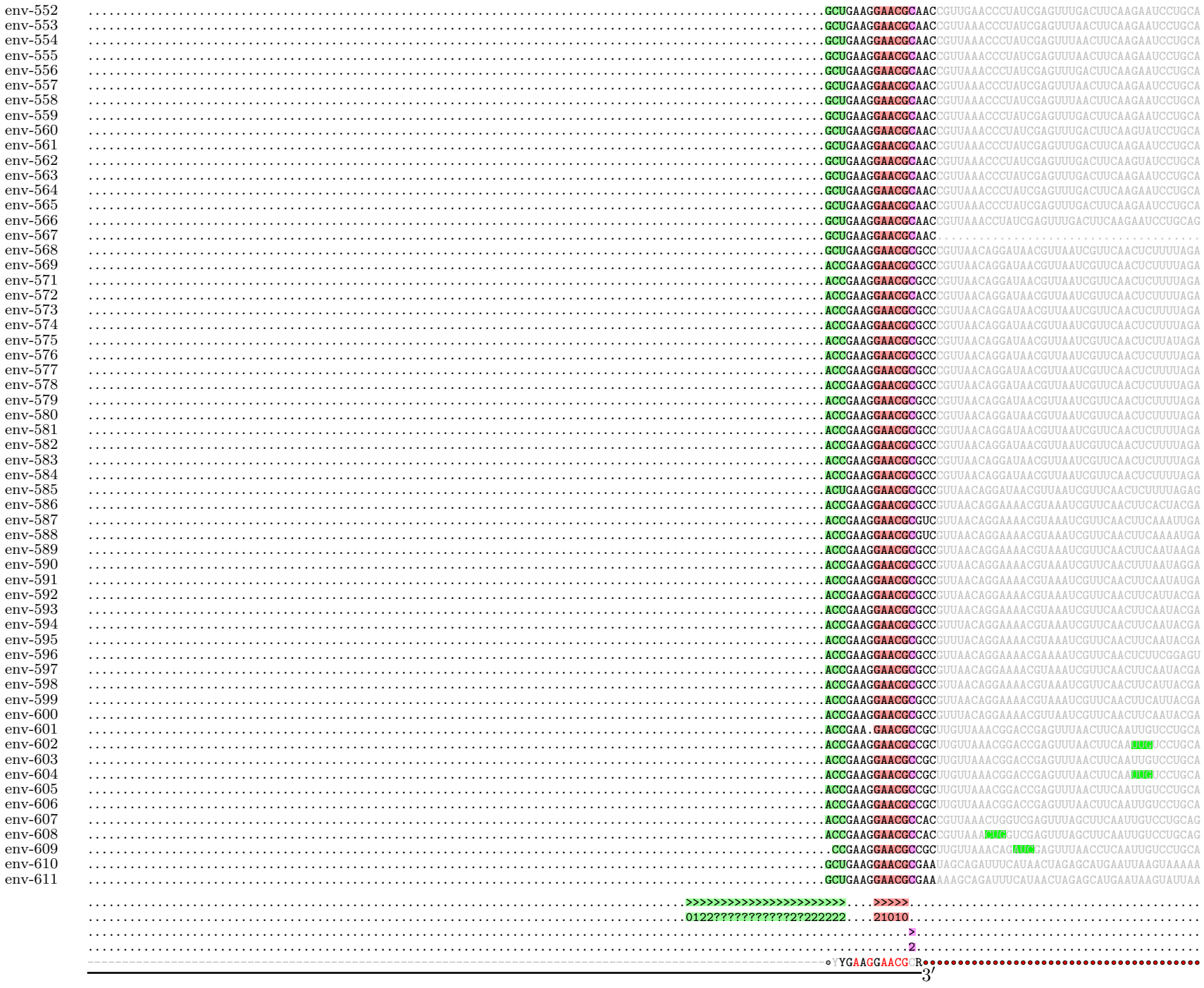

env-258
env-259
env-260
env-261
env-262
env-263
env-264
env-265
env-266
env-267
env-268
env-269
env-270
env-271
env-272
env-273
env-274
env-275
env-276
env-277
env-278
env-279
env-280
env-281
env-282
env-283
env-284
env-285
env-286
env-287
env-288
env-289
env-290
env-291
env-292
env-293
env-294
env-295
env-296
env-297
env-298
env-299
env-300
env-301
env-302
env-303
env-304
Ama-2-2
env-305
env-306
env-307
env-308
env-309
env-310
env-311
env-312
env-313
env-314
env-315
3′

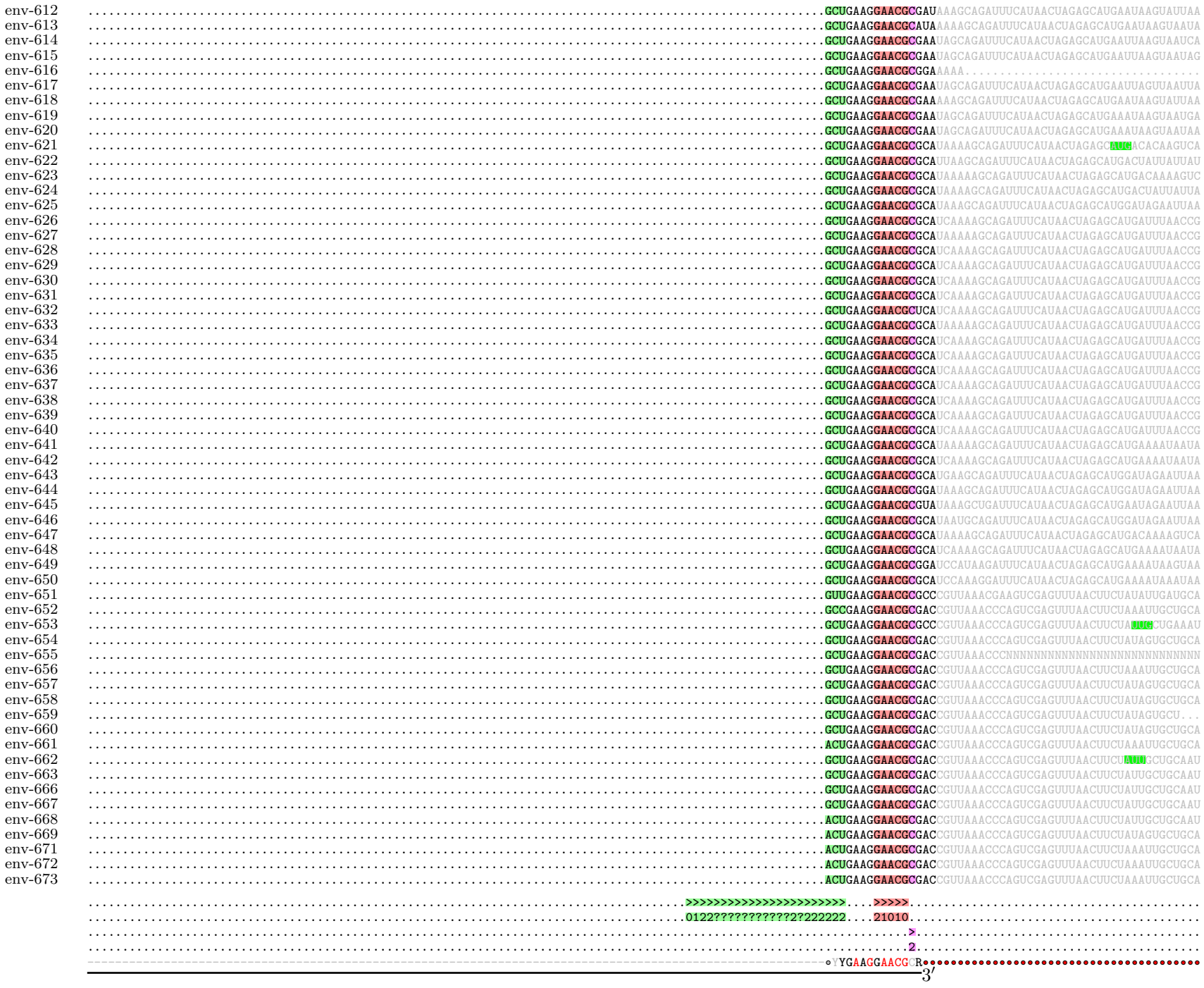

env-316
env-317
env-318
env-319
env-320
Mch-1-1
env-321
env-322
env-323
env-324
env-325
env-326
env-327
env-328
env-329
env-330
env-331
env-332
env-333
env-334
env-335
env-336
env-337
env-338
env-339
env-340
env-341
env-342
env-343
env-344
env-345
env-346
env-347
env-348
env-349
env-350
env-351
env-352
env-353
env-354
env-355
env-356
env-357
env-358
env-359
env-360
env-361
env-362
env-363
env-364
env-365
env-366
env-367
env-368
env-369
env-370
env-371
env-372
env-373
3′

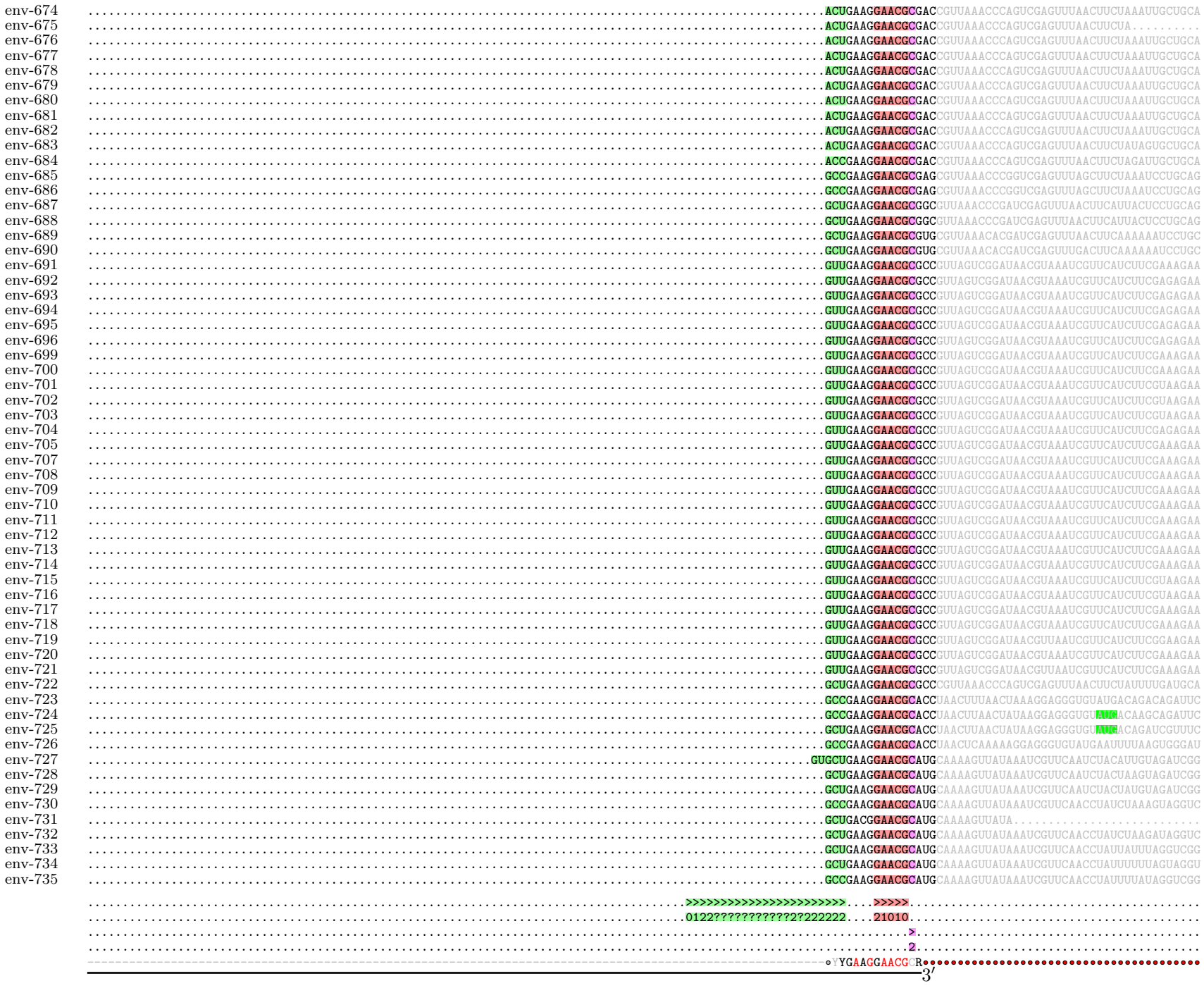

env-374
env-375
env-376
env-377
env-378
env-379
env-380
env-381
env-382
env-383
env-384
env-385
env-386
env-387
env-388
env-389
env-390
env-391
env-392
env-393
env-394
env-395
env-396
env-397
env-398
env-399
env-400
env-401
env-402
env-403
env-404
env-405
env-406
env-407
env-408
env-409
env-410
env-411
env-412
env-413
env-414
env-415
env-416
env-417
env-418
env-419
env-420
env-421
env-422
env-423
env-424
env-425
env-426
env-427
env-428
env-429
env-430
env-431
env-432
3′

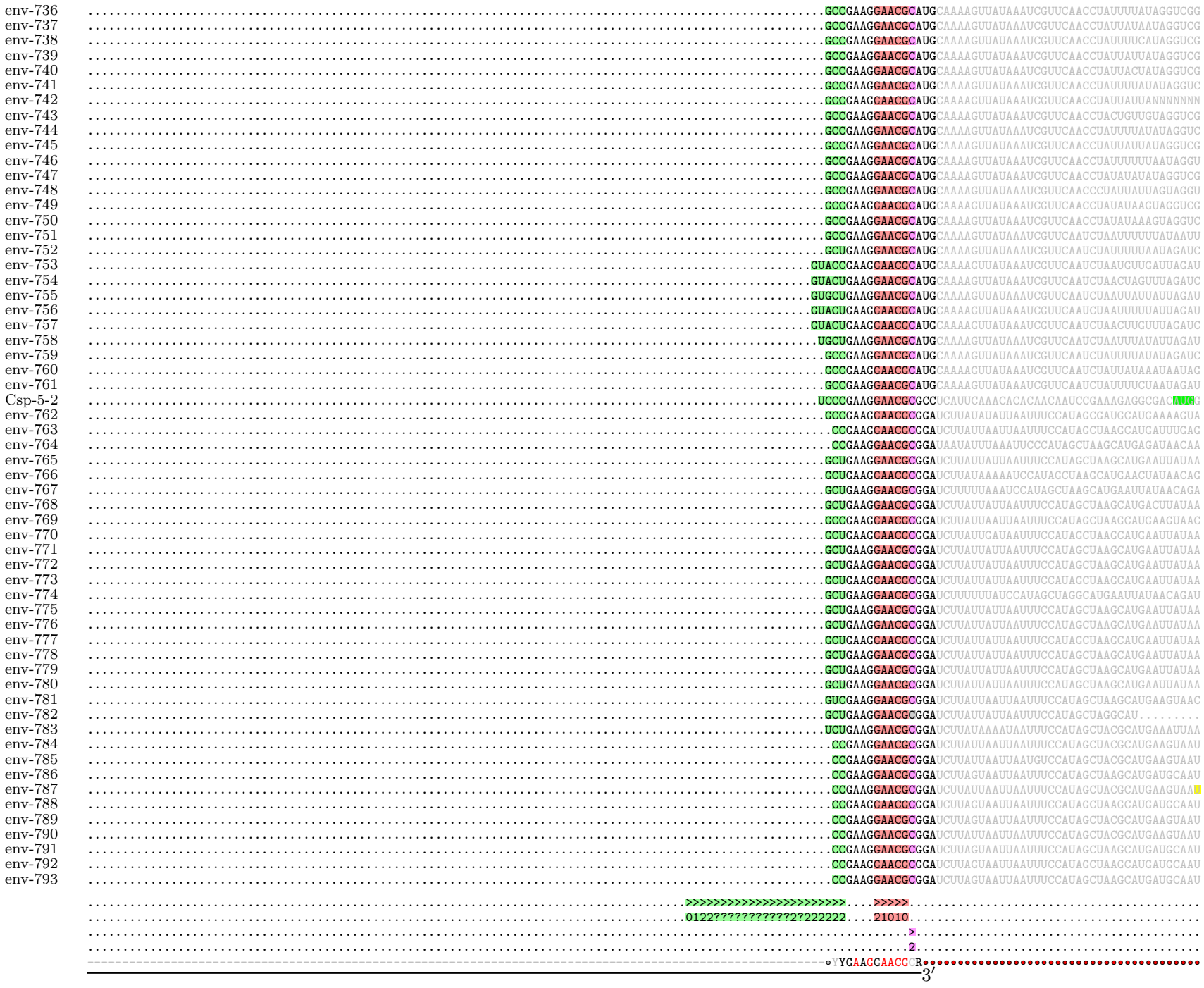

env-433
env-434
env-435
env-436
env-437
env-438
env-439
env-440
env-441
env-442
env-443
env-444
env-445
env-446
env-447
env-448
env-449
env-450
env-451
env-452
env-453
env-454
env-455
env-456
env-457
env-458
env-459
env-460
env-461
env-462
env-463
env-464
env-465
env-466
env-467
env-468
env-469
env-470
env-471
env-472
env-473
Csp-3-2
Csp-5-1
env-474
env-475
env-476
env-477
env-478
env-479
env-480
env-481
env-482
env-483
env-484
env-485
env-486
env-487
env-488
env-489
3′

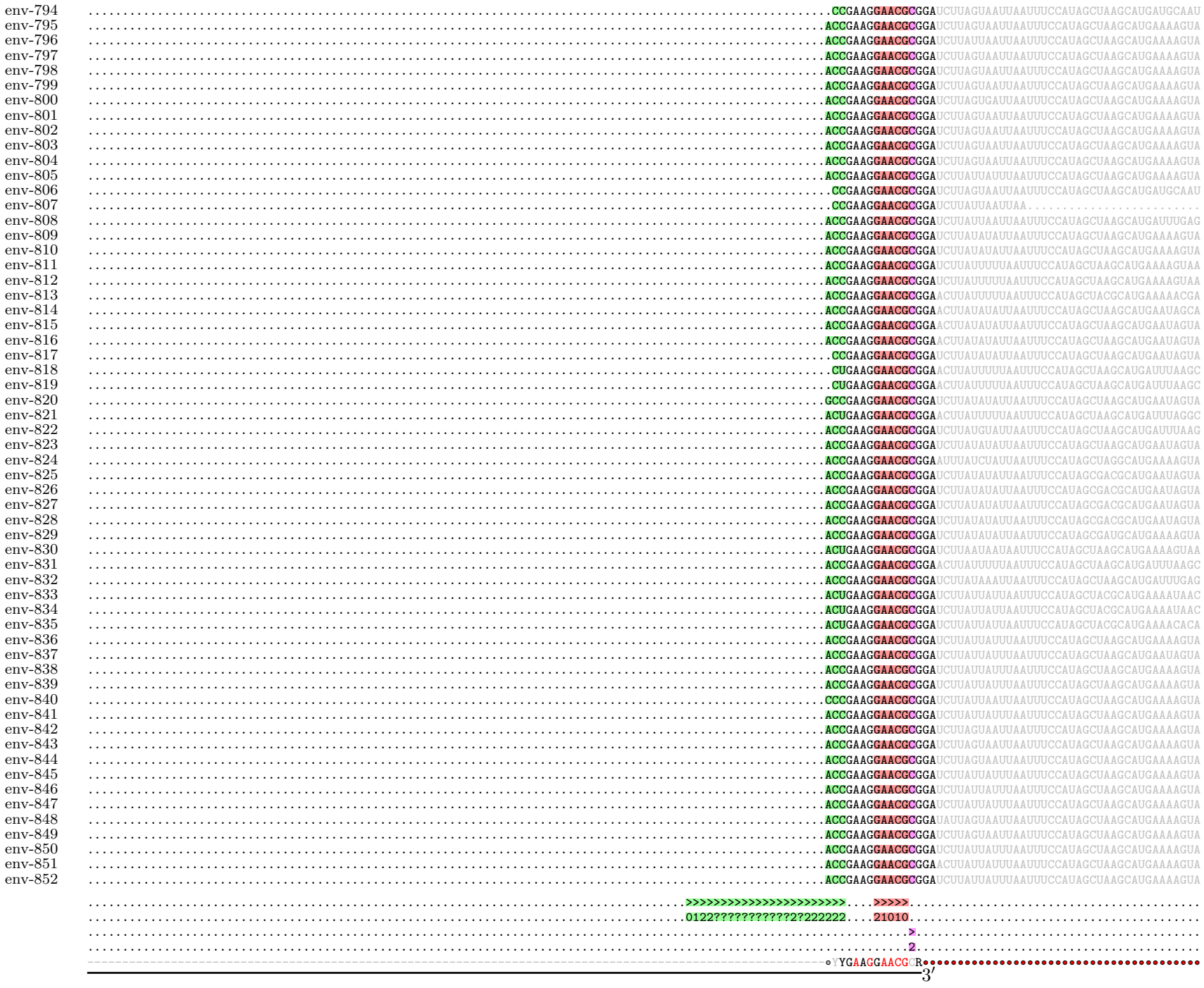

env-490
env-491
env-492
env-493
env-494
env-495
env-496
env-497
env-499
env-500
env-501
env-502
env-503
env-504
env-505
env-506
env-507
env-508
env-509
env-510
env-511
env-512
env-513
env-514
env-515
env-516
env-517
env-518
env-519
env-520
env-521
env-522
env-523
env-524
env-525
env-526
env-527
env-528
env-529
env-530
env-531
env-532
env-533
env-534
env-535
env-536
env-537
env-538
env-539
env-541
env-542
env-543
env-544
env-545
env-547
env-548
env-549
env-550
env-551
3′

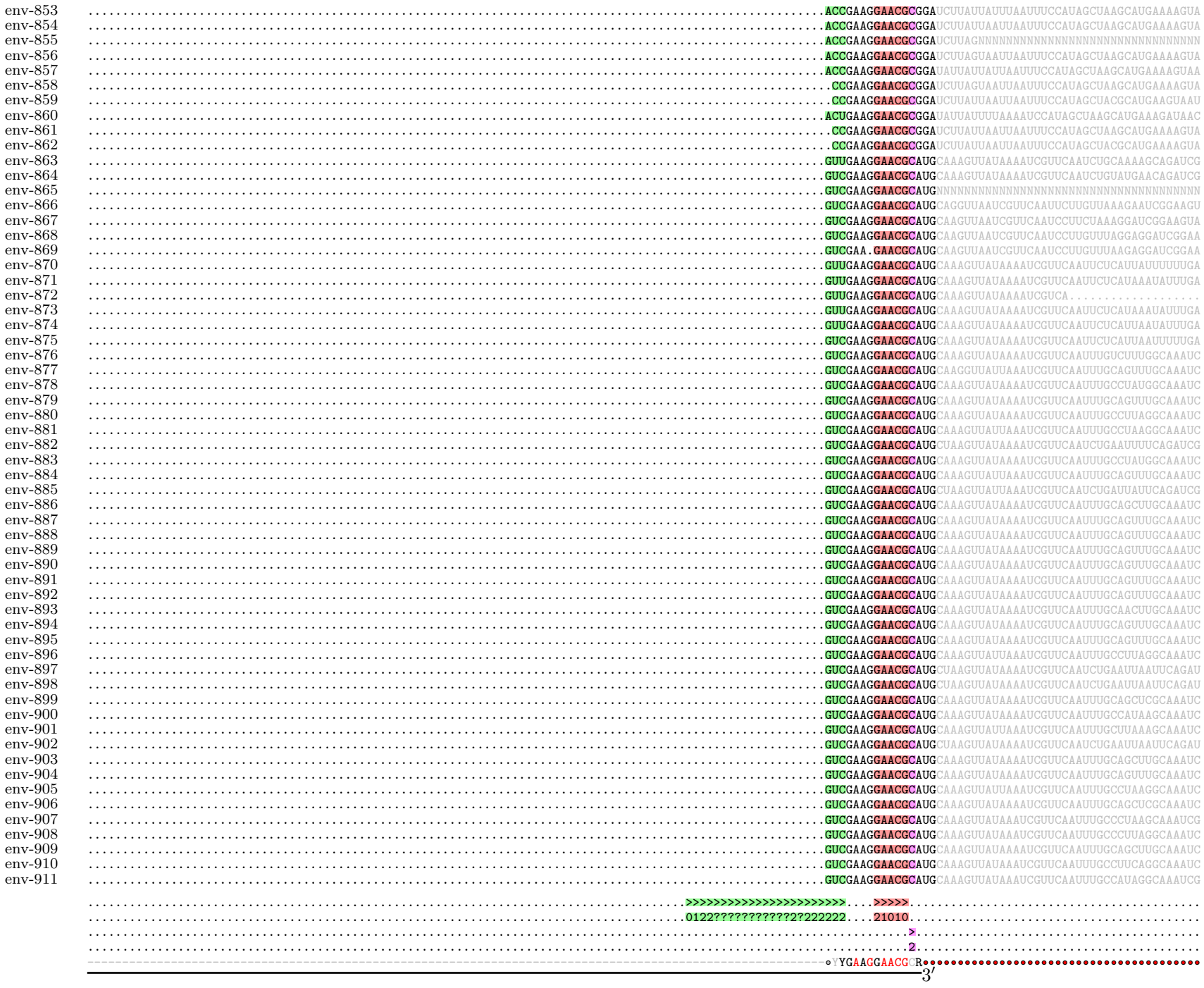

env-552
env-553
env-554
env-555
env-556
env-557
env-558
env-559
env-560
env-561
env-562
env-563
env-564
env-565
env-566
env-567
env-568
env-569
env-571
env-572
env-573
env-574
env-575
env-576
env-577
env-578
env-579
env-580
env-581
env-582
env-583
env-584
env-585
env-586
env-587
env-588
env-589
env-590
env-591
env-592
env-593
env-594
env-595
env-596
env-597
env-598
env-599
env-600
env-601
env-602
env-603
env-604
env-605
env-606
env-607
env-608
env-609
env-610
env-611
3′

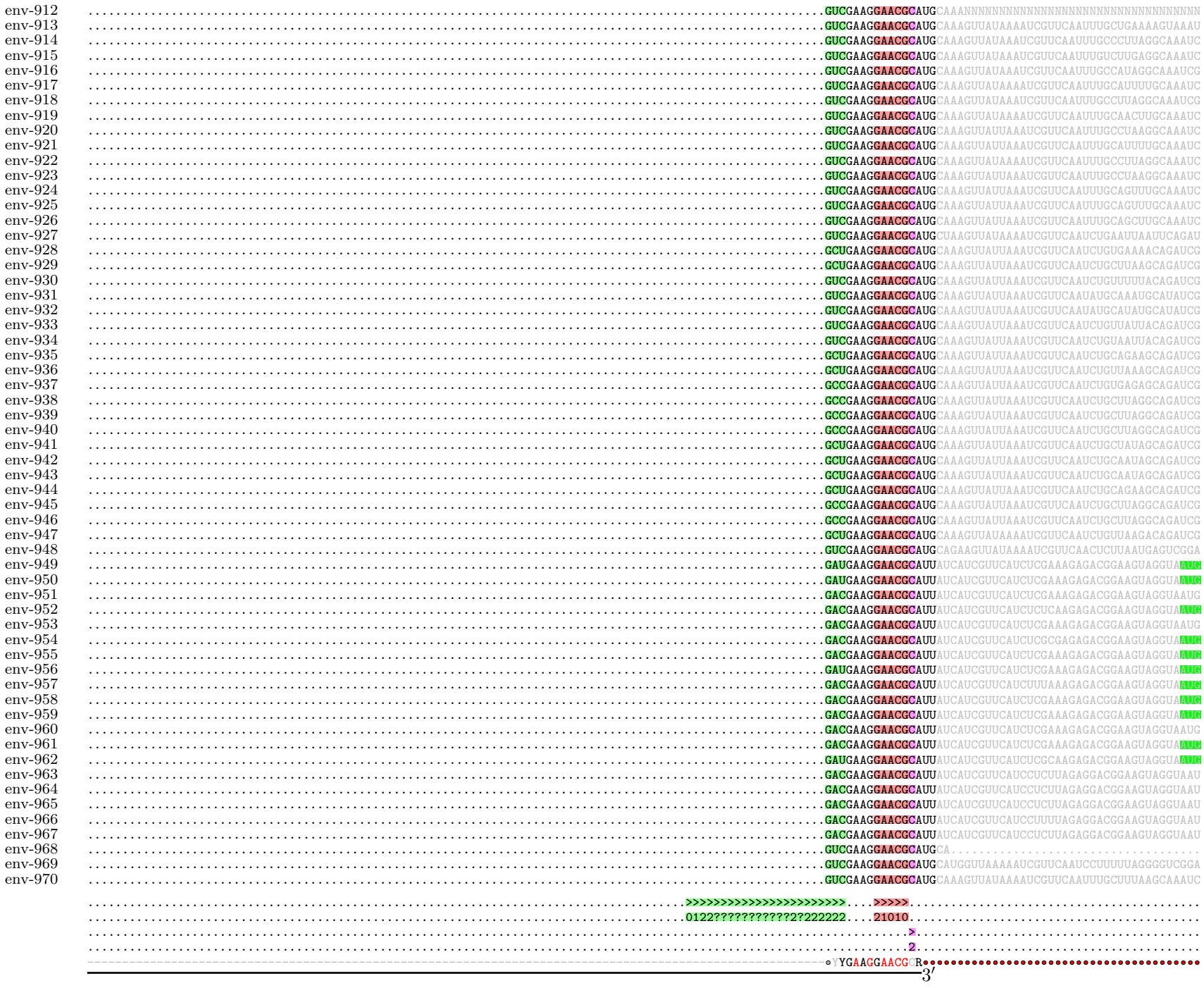

env-612
env-613
env-614
env-615
env-616
env-617
env-618
env-619
env-620
env-621
env-622
env-623
env-624
env-625
env-626
env-627
env-628
env-629
env-630
env-631
env-632
env-633
env-634
env-635
env-636
env-637
env-638
env-639
env-640
env-641
env-642
env-643
env-644
env-645
env-646
env-647
env-648
env-649
env-650
env-651
env-652
env-653
env-654
env-655
env-656
env-657
env-658
env-659
env-660
env-661
env-662
env-663
env-666
env-667
env-668
env-669
env-671
env-672
env-673
3′

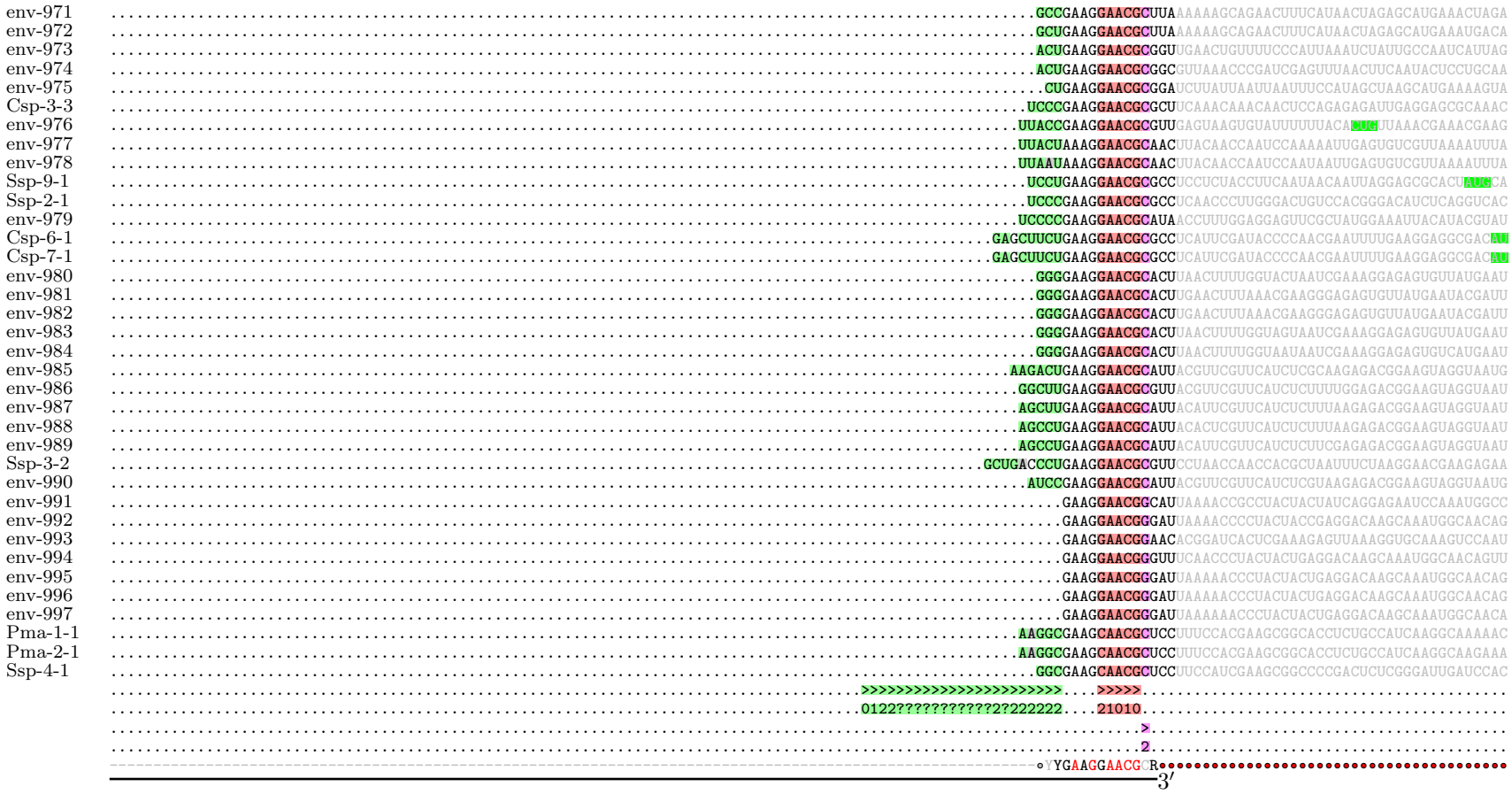

env-674
env-675
env-676
env-677
env-678
env-679
env-680
env-681
env-682
env-683
env-684
env-685
env-686
env-687
env-688
env-689
env-690
env-691
env-692
env-693
env-694
env-695
env-696
env-699
env-700
env-701
env-702
env-703
env-704
env-705
env-707
env-708
env-709
env-710
env-711
env-712
env-713
env-714
env-715
env-716
env-717
env-718
env-719
env-720
env-721
env-722
env-723
env-724
env-725
env-726
env-727
env-728
env-729
env-730
env-731
env-732
env-733
env-734
env-735
3′

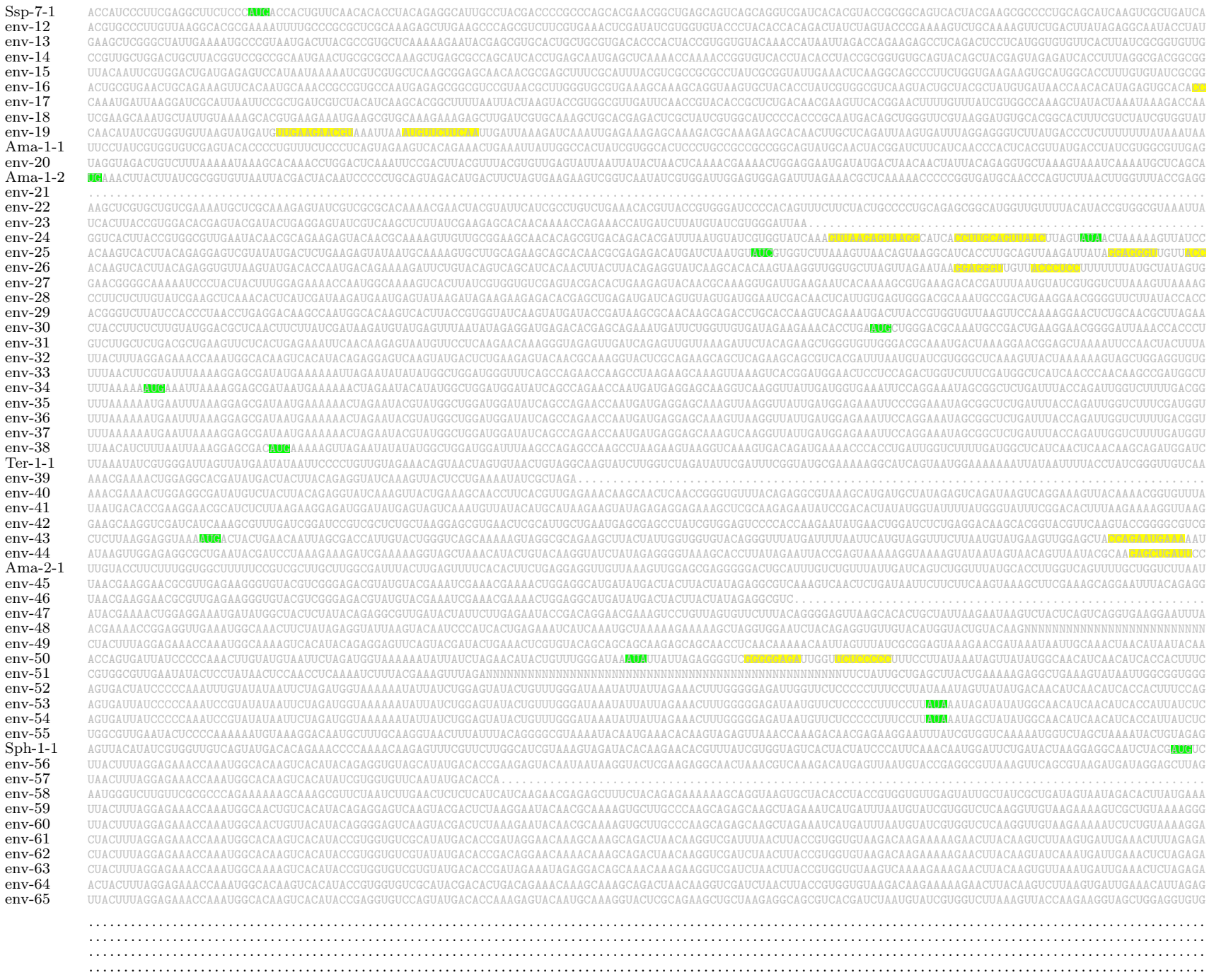

env-736
env-737
env-738
env-739
env-740
env-741
env-742
env-743
env-744
env-745
env-746
env-747
env-748
env-749
env-750
env-751
env-752
env-753
env-754
env-755
env-756
env-757
env-758
env-759
env-760
env-761
Csp-5-2
env-762
env-763
env-764
env-765
env-766
env-767
env-768
env-769
env-770
env-771
env-772
env-773
env-774
env-775
env-776
env-777
env-778
env-779
env-780
env-781
env-782
env-783
env-784
env-785
env-786
env-787
env-788
env-789
env-790
env-791
env-792
env-793
3′

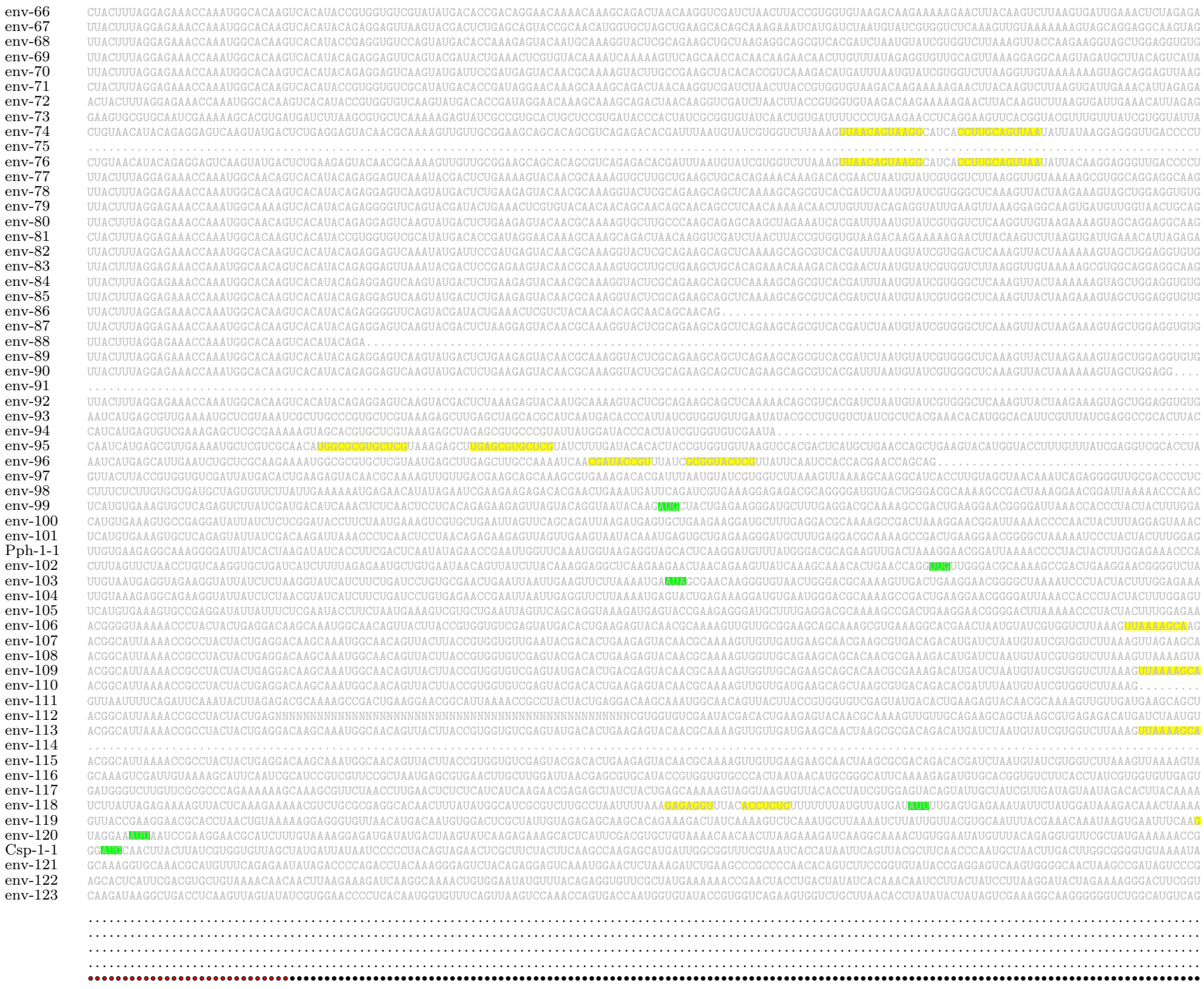

env-794
env-795
env-796
env-797
env-798
env-799
env-800
env-801
env-802
env-803
env-804
env-805
env-806
env-807
env-808
env-809
env-810
env-811
env-812
env-813
env-814
env-815
env-816
env-817
env-818
env-819
env-820
env-821
env-822
env-823
env-824
env-825
env-826
env-827
env-828
env-829
env-830
env-831
env-832
env-833
env-834
env-835
env-836
env-837
env-838
env-839
env-840
env-841
env-842
env-843
env-844
env-845
env-846
env-847
env-848
env-849
env-850
env-851
env-852
3′

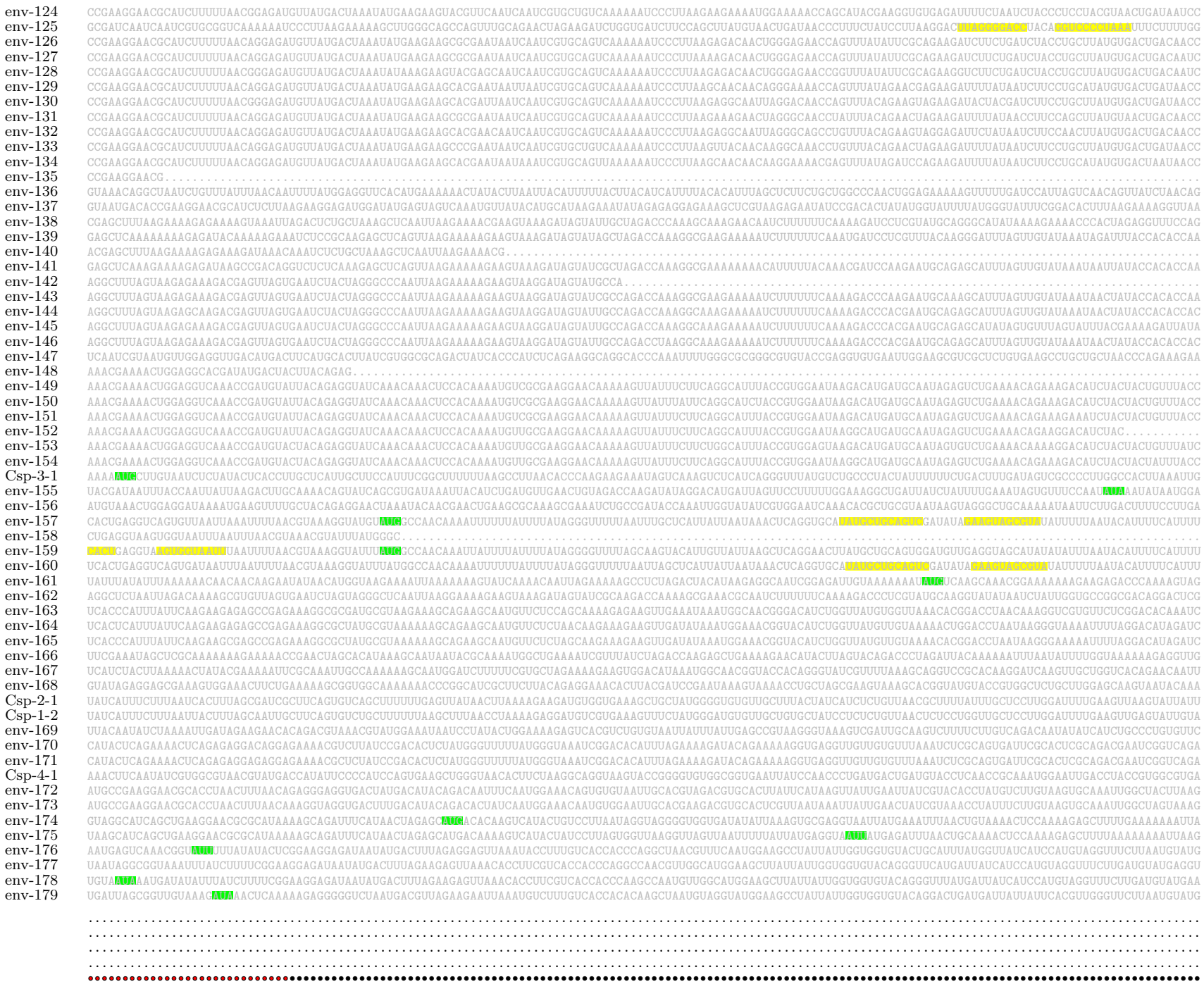

env-853
env-854
env-855
env-856
env-857
env-858
env-859
env-860
env-861
env-862
env-863
env-864
env-865
env-866
env-867
env-868
env-869
env-870
env-871
env-872
env-873
env-874
env-875
env-876
env-877
env-878
env-879
env-880
env-881
env-882
env-883
env-884
env-885
env-886
env-887
env-888
env-889
env-890
env-891
env-892
env-893
env-894
env-895
env-896
env-897
env-898
env-899
env-900
env-901
env-902
env-903
env-904
env-905
env-906
env-907
env-908
env-909
env-910
env-911
3′


env-912
env-913
env-914
env-915
env-916
env-917
env-918
env-919
env-920
env-921
env-922
env-923
env-924
env-925
env-926
env-927
env-928
env-929
env-930
env-931
env-932
env-933
env-934
env-935
env-936
env-937
env-938
env-939
env-940
env-941
env-942
env-943
env-944
env-945
env-946
env-947
env-948
env-949
env-950
env-951
env-952
env-953
env-954
env-955
env-956
env-957
env-958
env-959
env-960
env-961
env-962
env-963
env-964
env-965
env-966
env-967
env-968
env-969
env-970
3′


env-971
env-972
env-973
env-974
env-975
Csp-3-3
env-976
env-977
env-978
Ssp-9-1
Ssp-2-1
env-979
Csp-6-1
Csp-7-1
env-980
env-981
env-982
env-983
env-984
env-985
env-986
env-987
env-988
env-989
Ssp-3-2
env-990
env-991
env-992
env-993
env-994
env-995
env-996
env-997
Pma-1-1
Pma-2-1
Ssp-4-1
3′
alignment positions 481 · · · 640
env-1
Ssp-5-1
env-2
env-3
env-4
env-6
Ssp-6-1
Sel-1-1
env-7
env-8
env-9
env-10
env-11
Ssp-8-1


Ssp-7-1
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
Ama-1-1
env-20
Ama-1-2
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
Ter-1-1
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
Ama-2-1
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
Sph-1-1
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65


env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
Pph-1-1
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
Csp-1-1
env-121
env-122
env-123

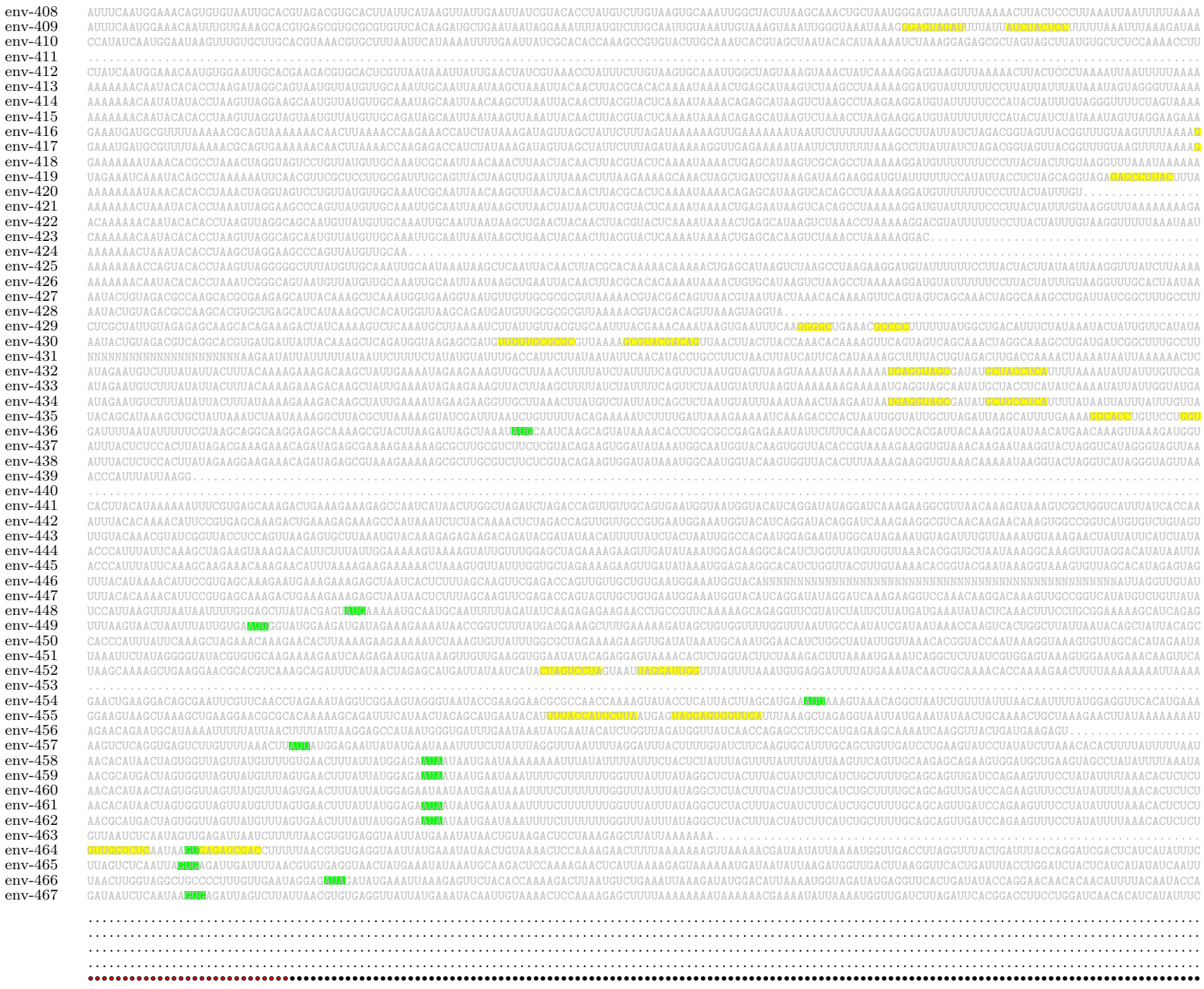

env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-135
env-136
env-137
env-138
env-139
env-140
env-141
env-142
env-143
env-144
env-145
env-146
env-147
env-148
env-149
env-150
env-151
env-152
env-153
env-154
Csp-3-1
env-155
env-156
env-157
env-158
env-159
env-160
env-161
env-162
env-163
env-164
env-165
env-166
env-167
env-168
Csp-2-1
Csp-1-2
env-169
env-170
env-171
Csp-4-1
env-172
env-173
env-174
env-175
env-176
env-177
env-178
env-179

Cwa-1-1
Csp-4-2
Tel-1-1
env-180
Ssp-1-1
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
env-190
env-191
env-192
env-193
env-194
env-195
env-196
env-197
env-198
env-199
env-200
env-201
env-202
env-203
env-204
env-205
env-206
env-207
env-208
env-209
env-210
env-211
env-212
env-213
env-214
env-215
Lsp-1-1
Tel-1-2
env-216
Ssp-1-2
env-217
env-218
env-219
env-220
env-221
Csp-1-3
Cwa-1-2
env-222
env-223
env-225
Csp-4-3
env-226
env-227
env-228
env-229
env-230

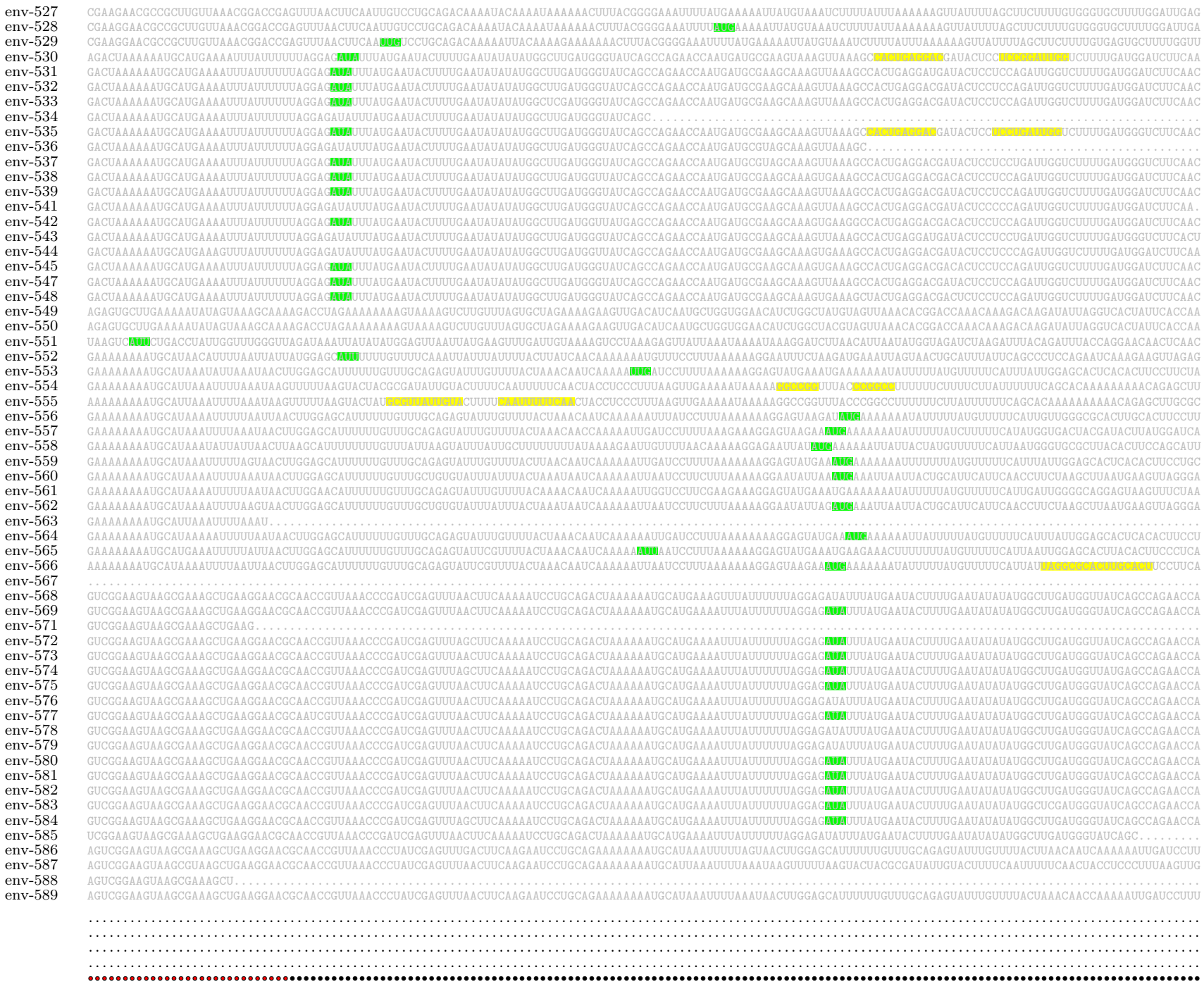
env-231
env-232
Ssp-3-1
env-233
env-234
env-235
env-236
env-237
env-238
env-239
env-240
env-241
env-242
env-243
env-244
env-245
env-246
env-247
env-248
env-249
env-250
env-251
env-252
env-253
env-254
env-255
env-256
env-257
env-258
env-259
env-260
env-261
env-262
env-263
env-264
env-265
env-266
env-267
env-268
env-269
env-270
env-271
env-272
env-273
env-274
env-275
env-276
env-277
env-278
env-279
env-280
env-281
env-282
env-283
env-284
env-285
env-286
env-287
env-288
env-289

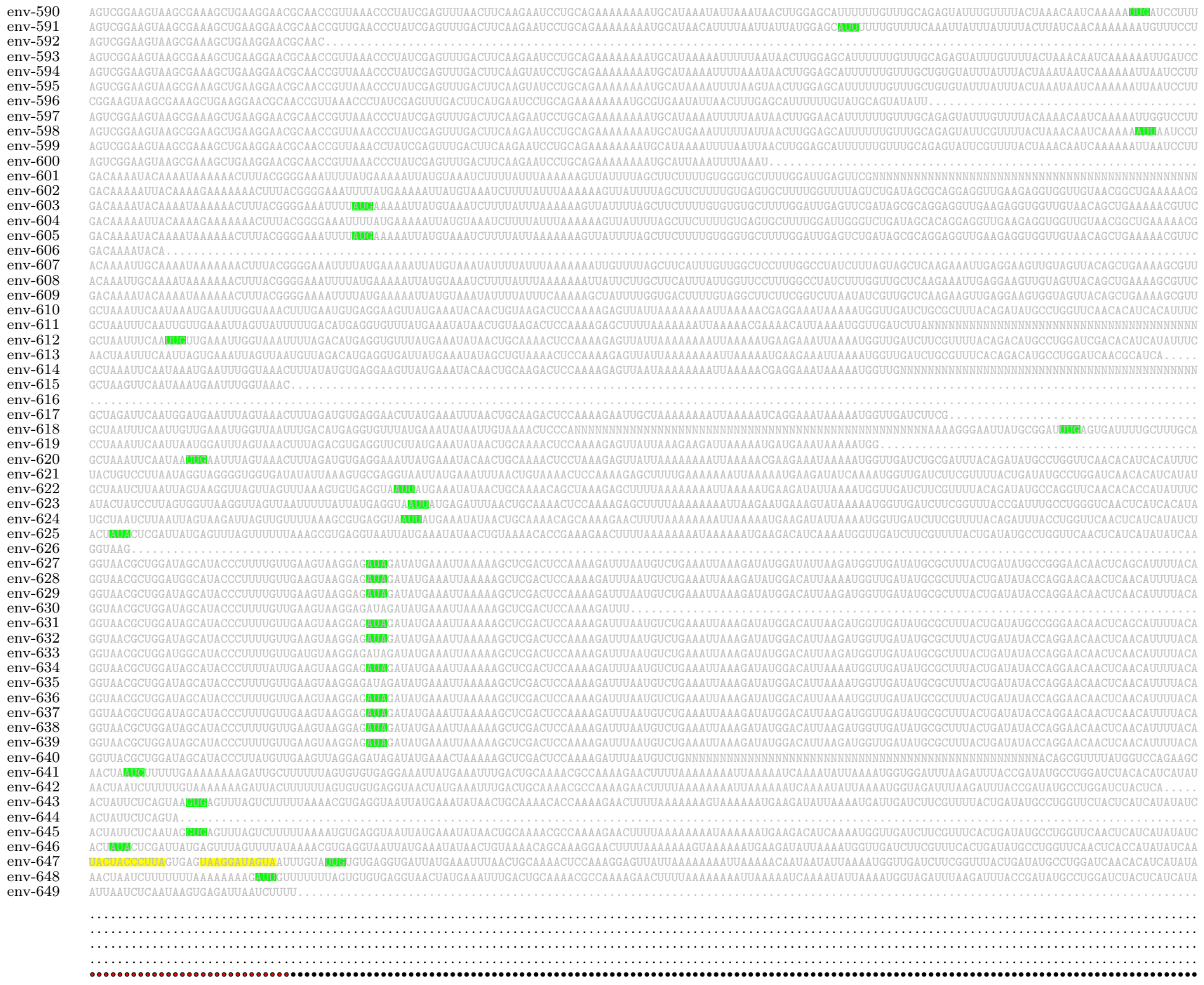
env-290
env-291
env-292
env-293
env-295
env-296
env-297
env-298
env-299
env-300
env-301
env-302
env-303
env-304
Ama-2-2
env-305
env-306
env-307
env-308
env-309
env-310
env-311
env-312
env-313
env-314
env-315
env-316
env-317
env-318
env-319
env-320
Mch-1-1
env-321
env-322
env-323
env-324
env-325
env-326
env-327
env-328
env-329
env-330
env-331
env-332
env-333
env-334
env-335
env-336
env-337
env-338
env-339
env-340
env-341
env-342
env-343
env-344
env-345
env-346
env-347

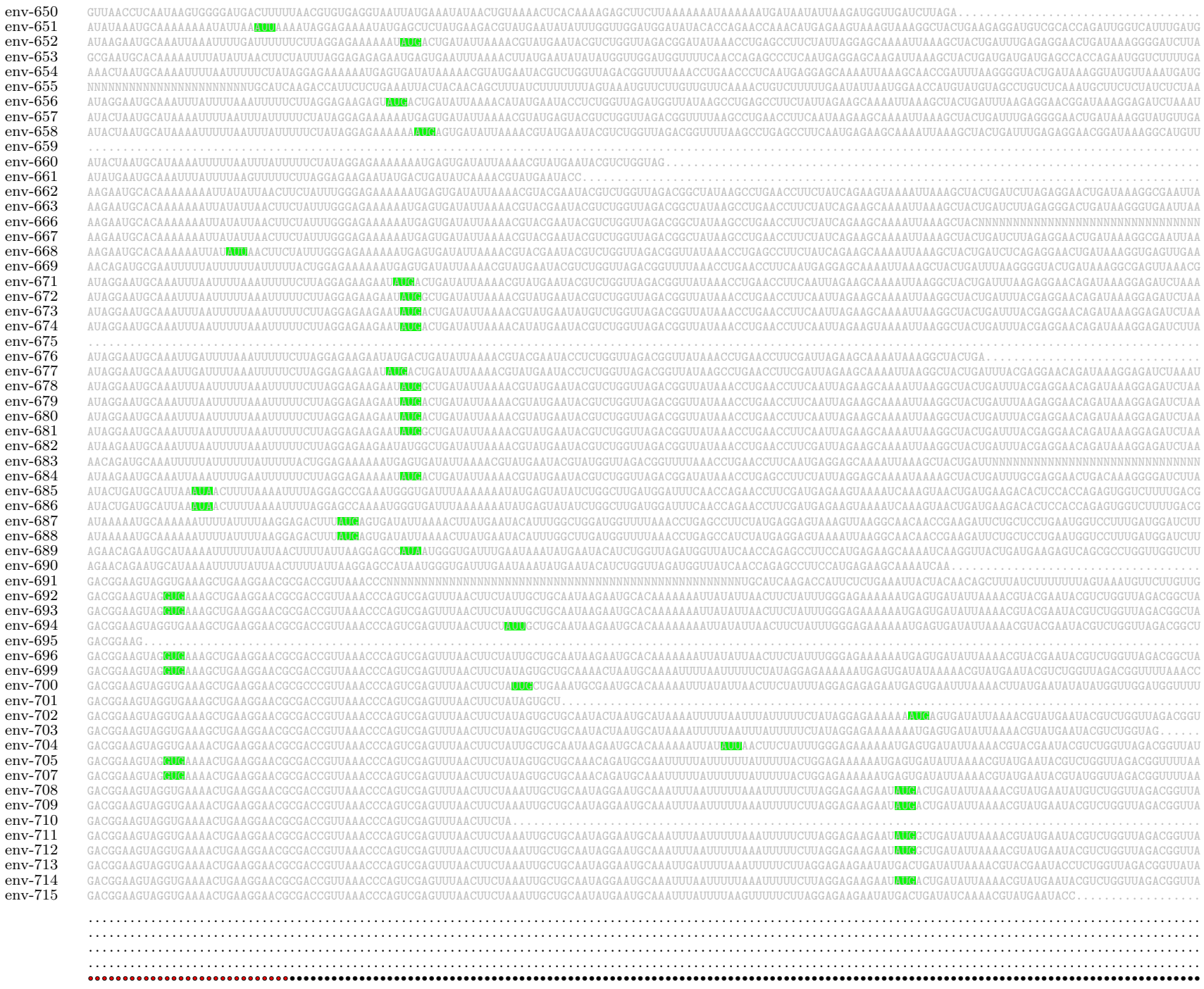
env-348
env-349
env-350
env-351
env-352
env-353
env-354
env-355
env-356
env-357
env-358
env-359
env-360
env-361
env-362
env-363
env-364
env-365
env-366
env-367
env-368
env-369
env-370
env-371
env-372
env-373
env-374
env-375
env-376
env-377
env-378
env-379
env-380
env-381
env-382
env-383
env-384
env-385
env-386
env-387
env-388
env-389
env-390
env-391
env-392
env-393
env-394
env-395
env-396
env-397
env-398
env-399
env-400
env-401
env-402
env-403
env-404
env-405
env-406
env-407


env-408
env-409
env-410
env-411
env-412
env-413
env-414
env-415
env-416
env-417
env-418
env-419
env-420
env-421
env-422
env-423
env-424
env-425
env-426
env-427
env-428
env-429
env-430
env-431
env-432
env-433
env-434
env-435
env-436
env-437
env-438
env-439
env-440
env-441
env-442
env-443
env-444
env-445
env-446
env-447
env-448
env-449
env-450
env-451
env-452
env-453
env-454
env-455
env-456
env-457
env-458
env-459
env-460
env-461
env-462
env-463
env-464
env-465
env-466
env-467

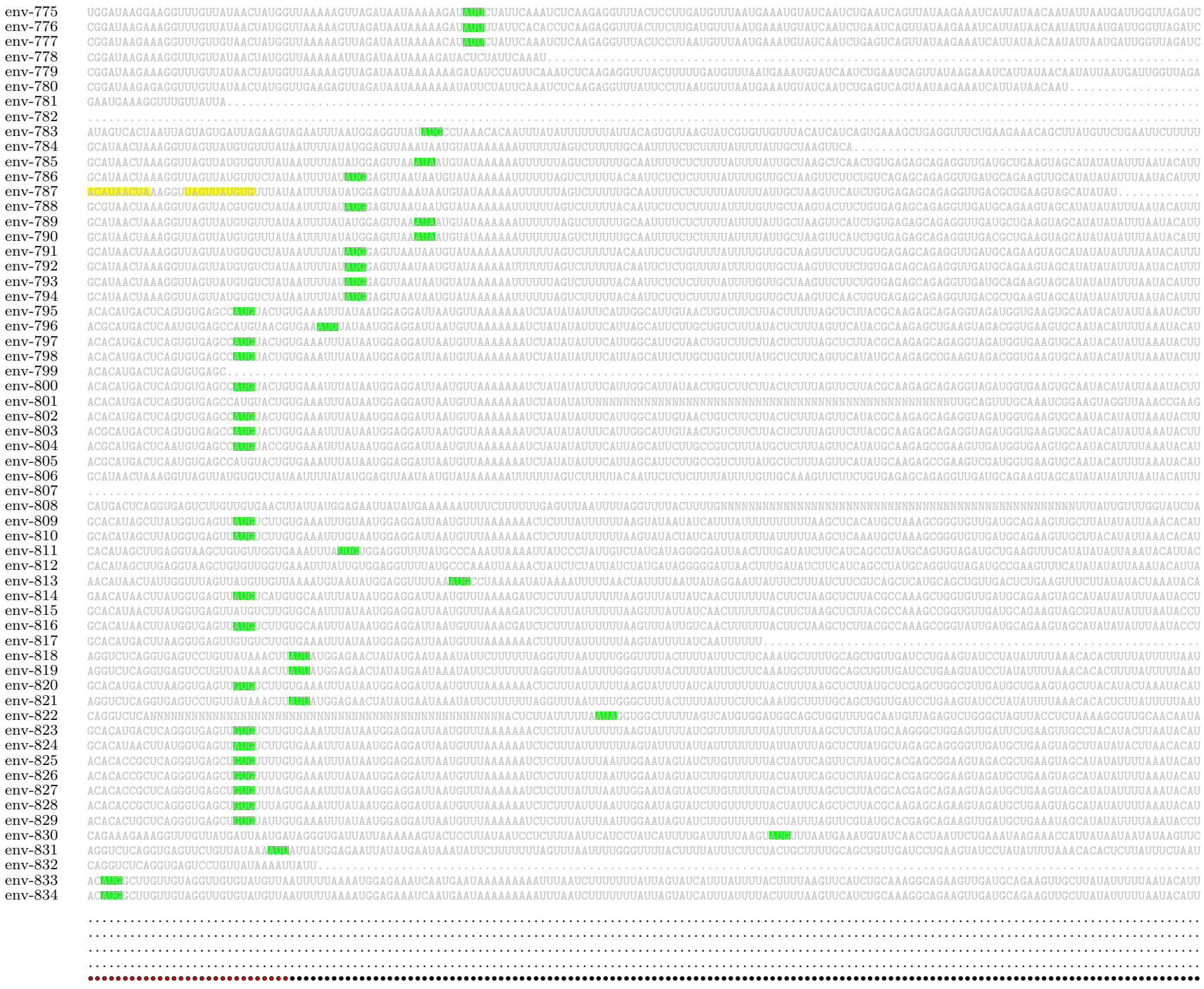
env-468
env-469
env-470
env-471
env-472
env-473
Csp-3-2
Csp-5-1
env-474
env-475
env-476
env-477
env-478
env-479
env-480
env-481
env-482
env-483
env-484
env-485
env-486
env-487
env-488
env-489
env-490
env-491
env-492
env-493
env-494
env-495
env-496
env-497
env-499
env-500
env-501
env-502
env-503
env-504
env-505
env-506
env-507
env-508
env-509
env-510
env-511
env-512
env-513
env-514
env-515
env-516
env-517
env-518
env-519
env-520
env-521
env-522
env-523
env-524
env-525
env-526


env-527
env-528
env-529
env-530
env-531
env-532
env-533
env-534
env-535
env-536
env-537
env-538
env-539
env-541
env-542
env-543
env-544
env-545
env-547
env-548
env-549
env-550
env-551
env-552
env-553
env-554
env-555
env-556
env-557
env-558
env-559
env-560
env-561
env-562
env-563
env-564
env-565
env-566
env-567
env-568
env-569
env-571
env-572
env-573
env-574
env-575
env-576
env-577
env-578
env-579
env-580
env-581
env-582
env-583
env-584
env-585
env-586
env-587
env-588
env-589

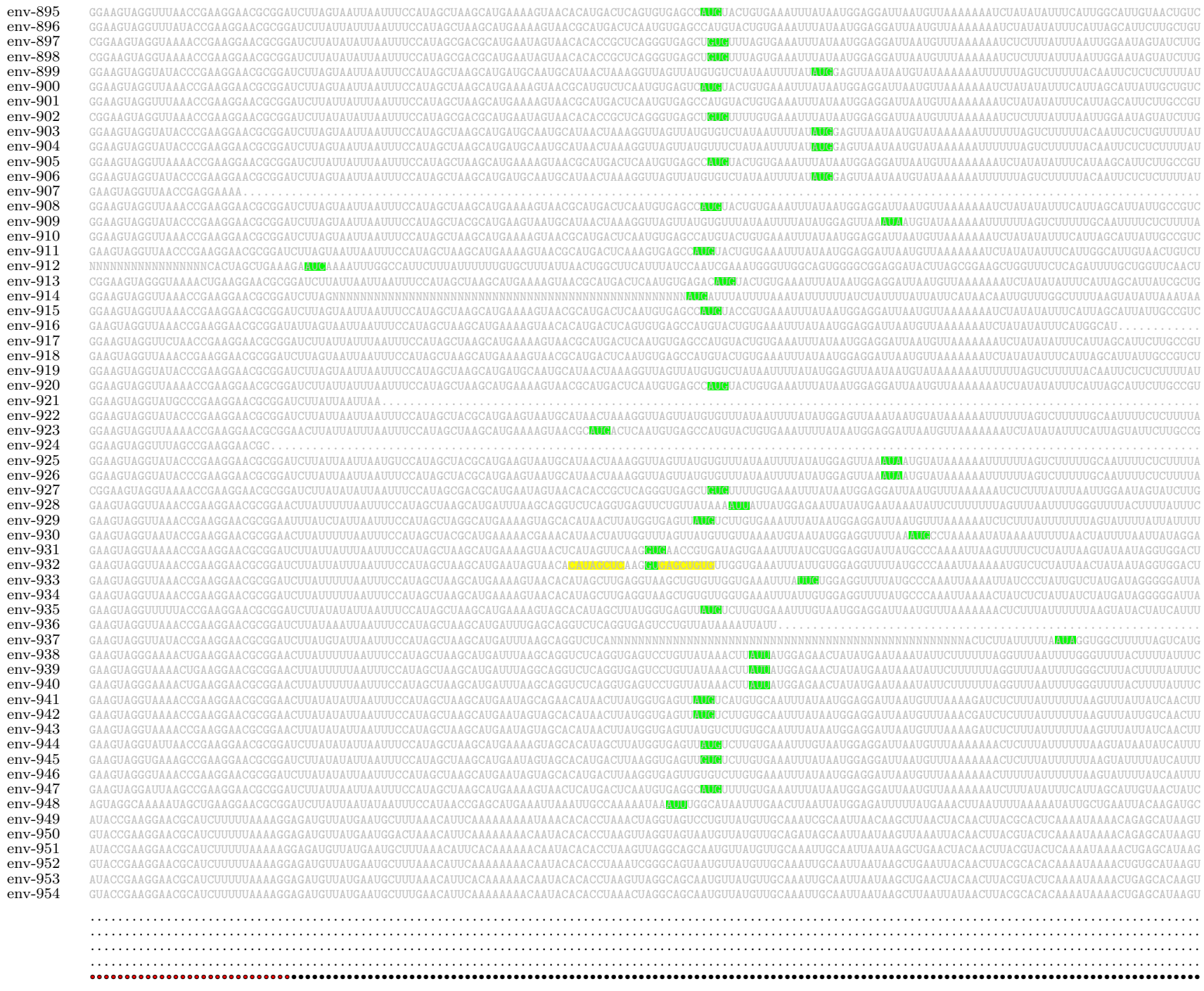

env-590
env-591
env-592
env-593
env-595
env-596
env-597
env-598
env-599
env-600
env-601
env-602
env-603
env-604
env-605
env-606
env-607
env-608
env-609
env-610
env-611
env-612
env-613
env-614
env-615
env-616
env-617
env-618
env-619
env-620
env-621
env-622
env-623
env-624
env-625
env-626
env-627
env-628
env-629
env-630
env-631
env-632
env-633
env-634
env-635
env-636
env-637
env-638
env-639
env-640
env-641
env-642
env-643
env-644
env-645
env-646
env-647
env-648
env-649

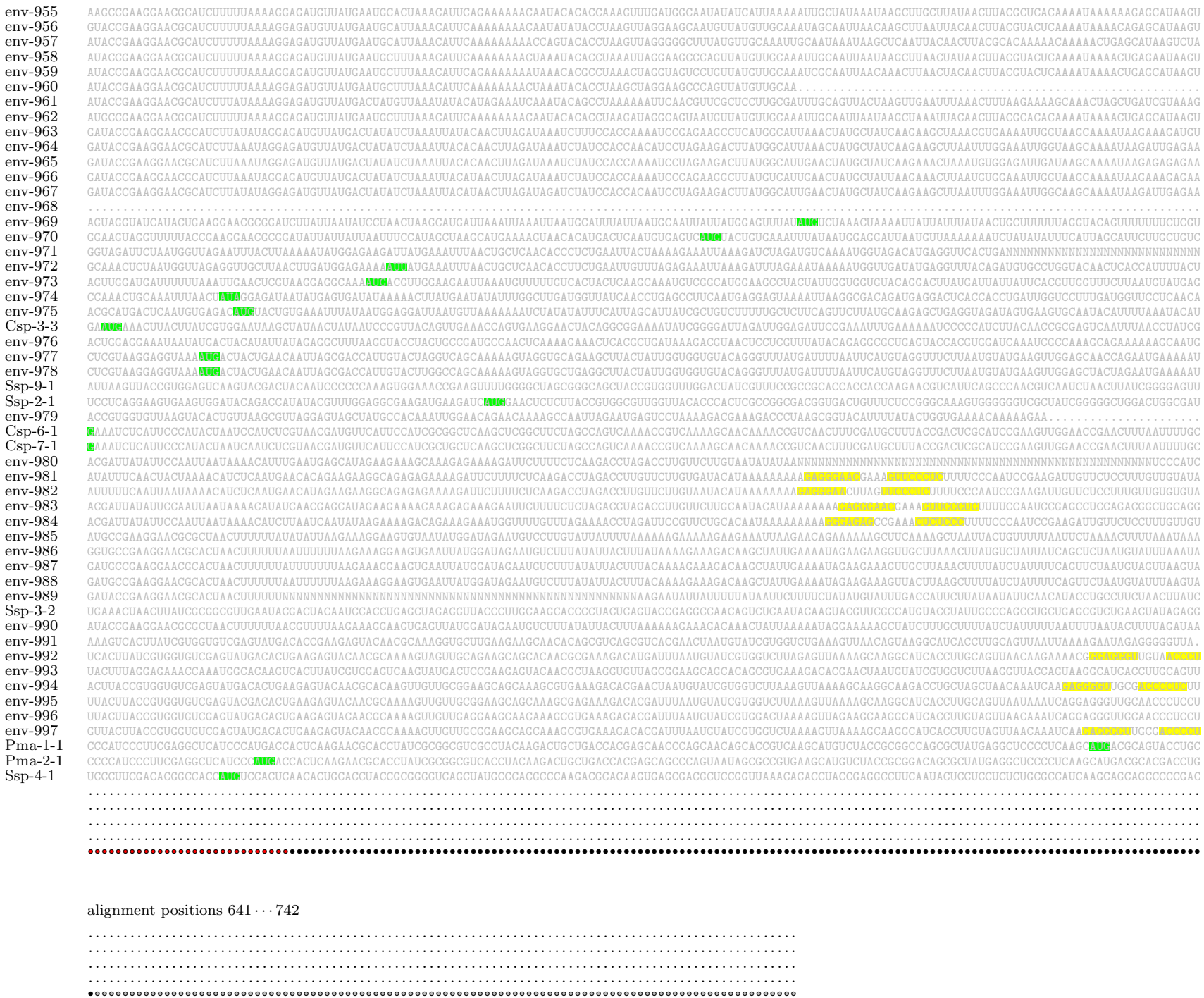

env-650
env-651
env-652
env-653
env-654
env-655
env-656
env-657
env-658
env-659
env-660
env-661
env-662
env-663
env-666
env-667
env-668
env-669
env-671
env-672
env-673
env-674
env-675
env-676
env-677
env-678
env-679
env-680
env-681
env-682
env-683
env-684
env-685
env-686
env-687
env-688
env-689
env-690
env-691
env-692
env-693
env-694
env-695
env-696
env-699
env-700
env-701
env-702
env-703
env-704
env-705
env-707
env-708
env-709
env-710
env-711
env-712
env-713
env-714
env-715

env-716
env-717
env-718
env-719
env-720
env-721
env-722
env-723
env-724
env-725
env-726
env-727
env-728
env-729
env-730
env-731
env-732
env-733
env-734
env-735
env-736
env-737
env-738
env-739
env-740
env-741
env-742
env-743
env-744
env-745
env-746
env-747
env-748
env-749
env-750
env-751
env-752
env-753
env-754
env-755
env-756
env-757
env-758
env-759
env-760
env-761
Csp-5-2
env-762
env-763
env-764
env-765
env-766
env-767
env-768
env-769
env-770
env-771
env-772
env-773
env-774


env-775
env-776
env-777
env-778
env-779
env-780
env-781
env-782
env-783
env-784
env-785
env-786
env-787
env-788
env-789
env-790
env-791
env-792
env-793
env-794
env-795
env-796
env-797
env-798
env-799
env-800
env-801
env-802
env-803
env-804
env-805
env-806
env-807
env-808
env-809
env-810
env-811
env-812
env-813
env-814
env-815
env-816
env-817
env-818
env-819
env-820
env-821
env-822
env-823
env-824
env-825
env-826
env-827
env-828
env-829
env-830
env-831
env-832
env-833
env-834

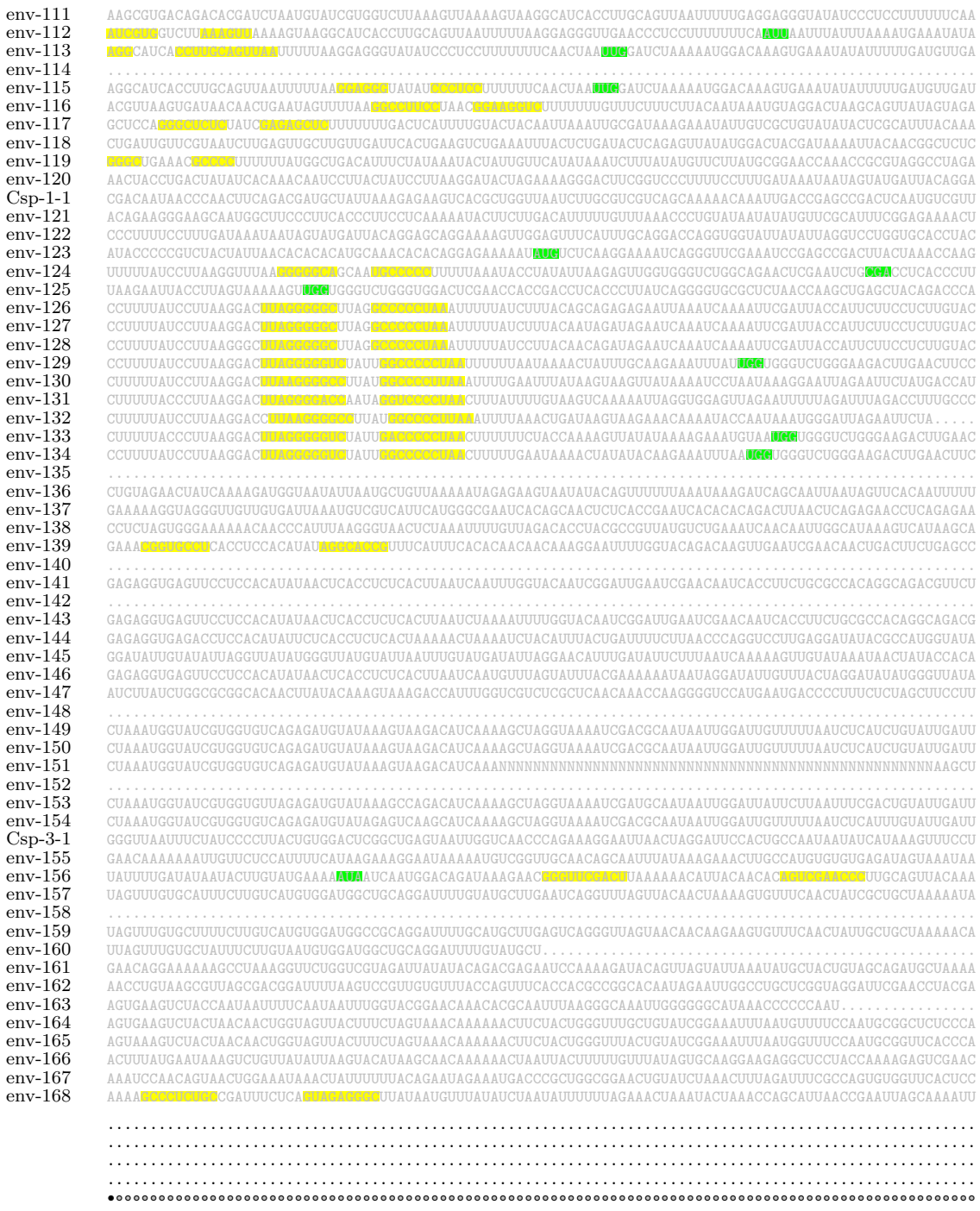
env-835
env-836
env-837
env-838
env-839
env-840
env-841
env-842
env-843
env-844
env-845
env-846
env-847
env-848
env-849
env-850
env-851
env-852
env-853
env-854
env-855
env-856
env-857
env-858
env-859
env-860
env-861
env-862
env-863
env-864
env-865
env-866
env-867
env-868
env-869
env-870
env-871
env-872
env-873
env-874
env-875
env-876
env-877
env-878
env-879
env-880
env-881
env-882
env-883
env-884
env-885
env-886
env-887
env-888
env-889
env-890
env-891
env-892
env-893
env-894

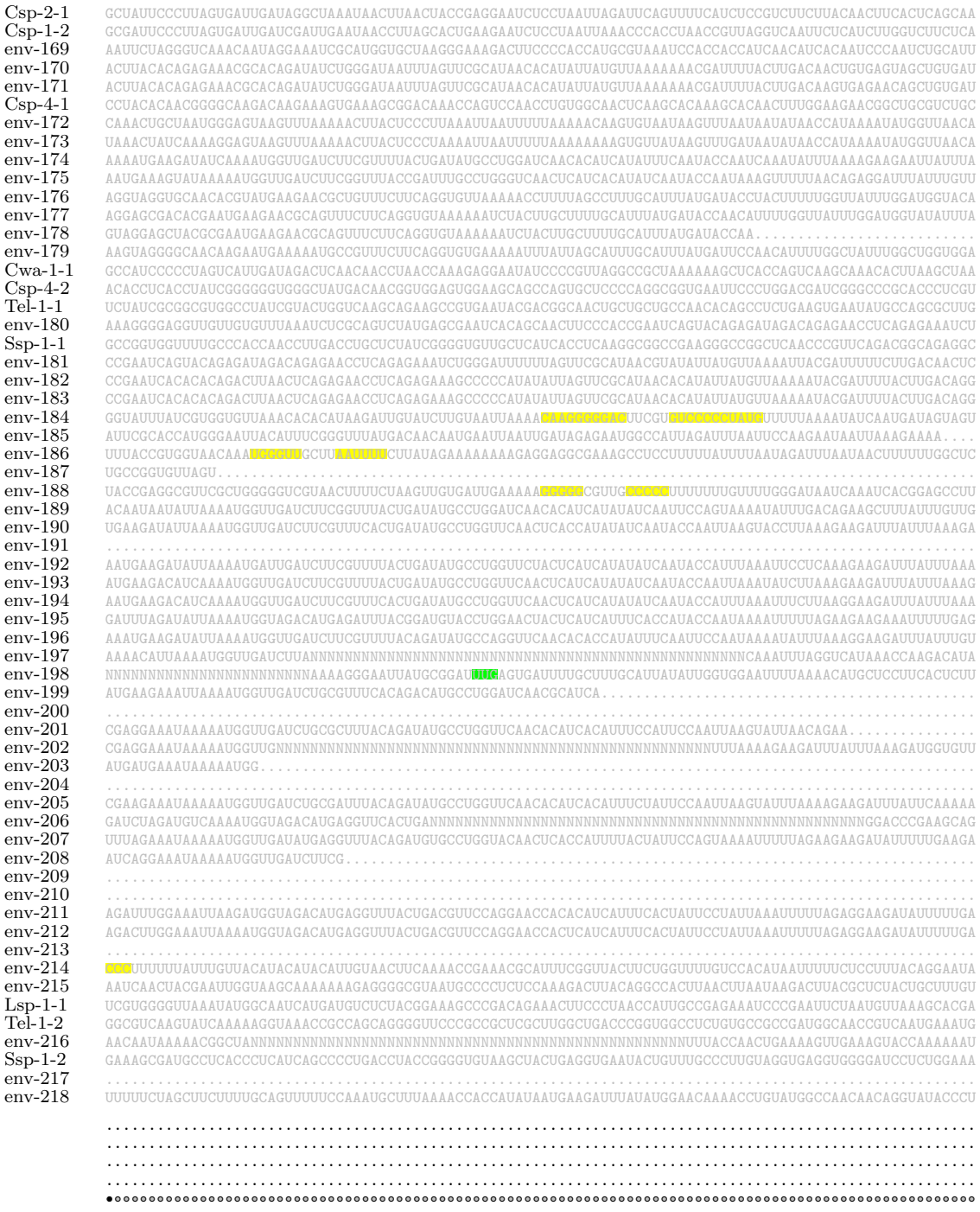

env-895
env-896
env-897
env-898
env-899
env-900
env-901
env-902
env-903
env-904
env-905
env-906
env-907
env-908
env-909
env-910
env-911
env-912
env-913
env-914
env-915
env-916
env-917
env-918
env-919
env-920
env-921
env-922
env-923
env-924
env-925
env-926
env-927
env-928
env-929
env-930
env-931
env-932
env-933
env-934
env-935
env-936
env-937
env-938
env-939
env-940
env-941
env-942
env-943
env-944
env-945
env-946
env-947
env-948
env-949
env-950
env-951
env-952
env-953
env-954

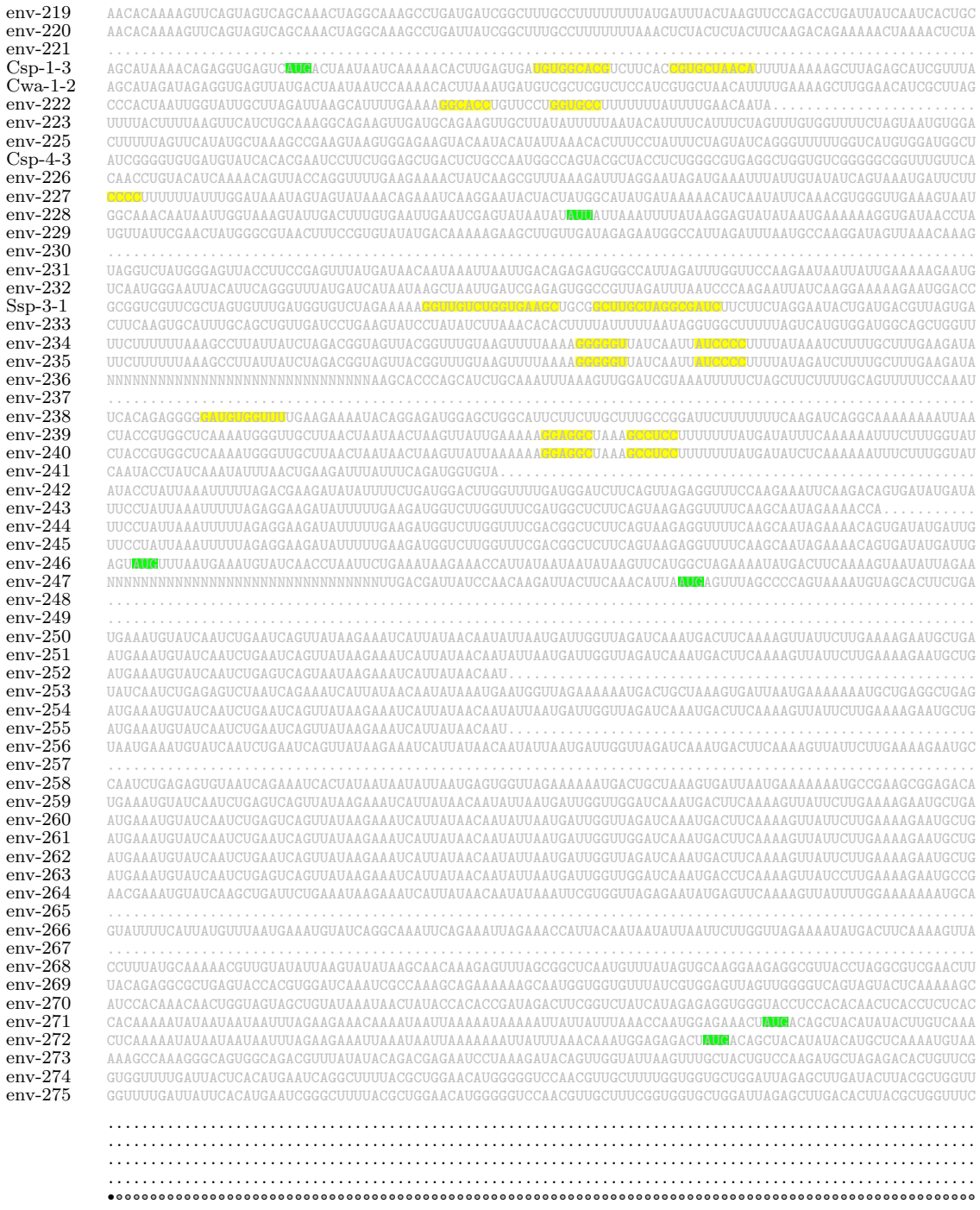

env-955
env-956
env-957
env-958
env-959
env-960
env-961
env-962
env-963
env-964
env-965
env-966
env-967
env-968
env-969
env-970
env-971
env-972
env-973
env-974
env-975
Csp-3-3
env-976
env-977
env-978
Ssp-9-1
Ssp-2-1
env-979
Csp-6-1
Csp-7-1
env-980
env-981
env-982
env-983
env-984
env-985
env-986
env-987
env-988
env-989
Ssp-3-2
env-990
env-991
env-992
env-993
env-994
env-995
env-996
env-997
Pma-1-1
Pma-2-1
Ssp-4-1
alignment positions 641 · · · 742

env-1
Ssp-5-1
env-2
env-3
env-4
env-6
Ssp-6-1
Sel-1-1
env-7
env-8
env-9
env-10
env-11
Ssp-8-1
Ssp-7-1
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
Ama-1-1
env-20
Ama-1-2
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
Ter-1-1
env-39
env-40
env-41
env-42
env-43
env-44
Ama-2-1
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52

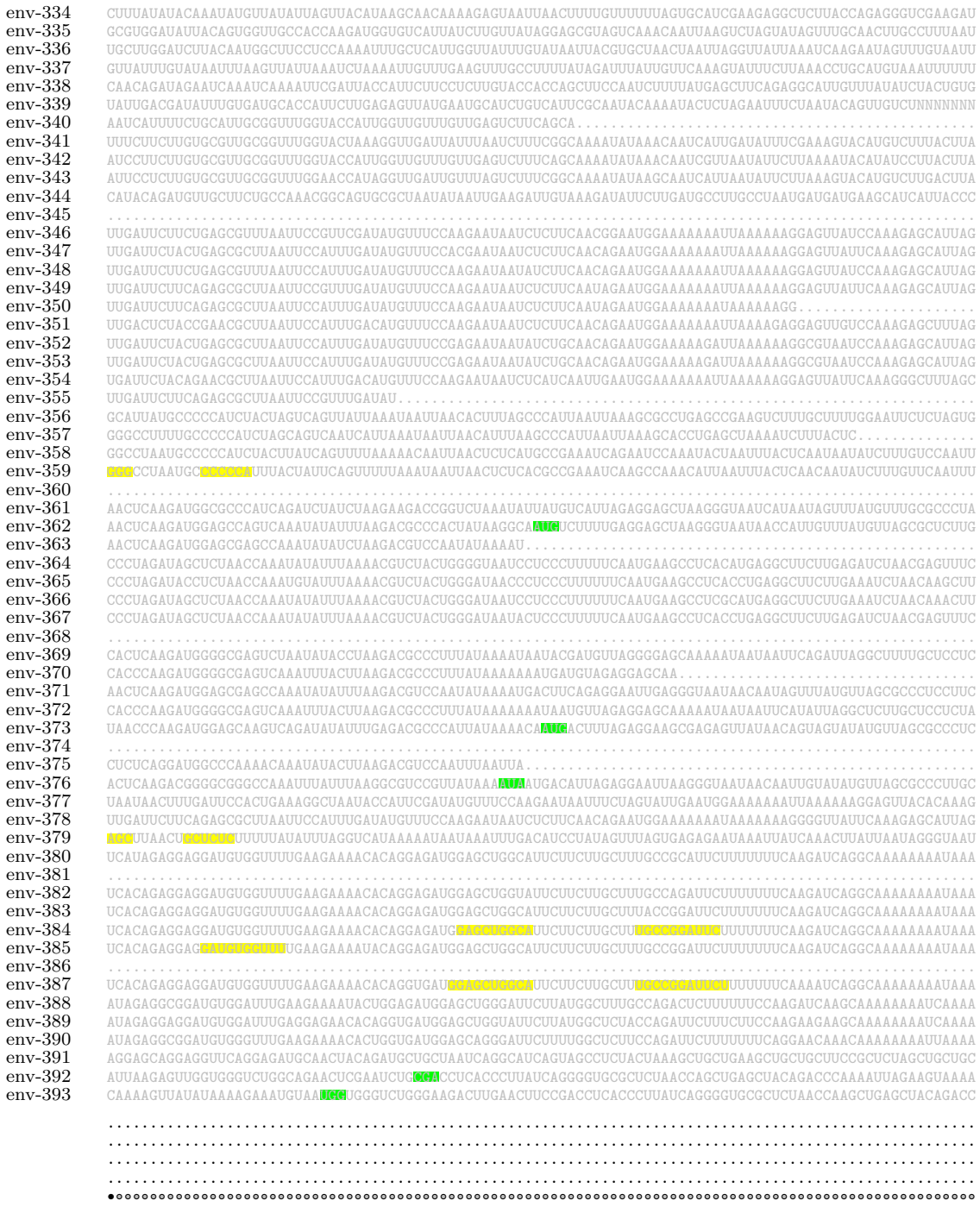
env-53
env-54
env-55
Sph-1-1
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
Pph-1-1
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110


env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
Csp-1-1
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-135
env-136
env-137
env-138
env-139
env-140
env-141
env-142
env-143
env-144
env-145
env-146
env-147
env-148
env-149
env-150
env-151
env-152
env-153
env-154
Csp-3-1
env-155
env-156
env-157
env-158
env-159
env-160
env-161
env-162
env-163
env-164
env-165
env-166
env-167
env-168

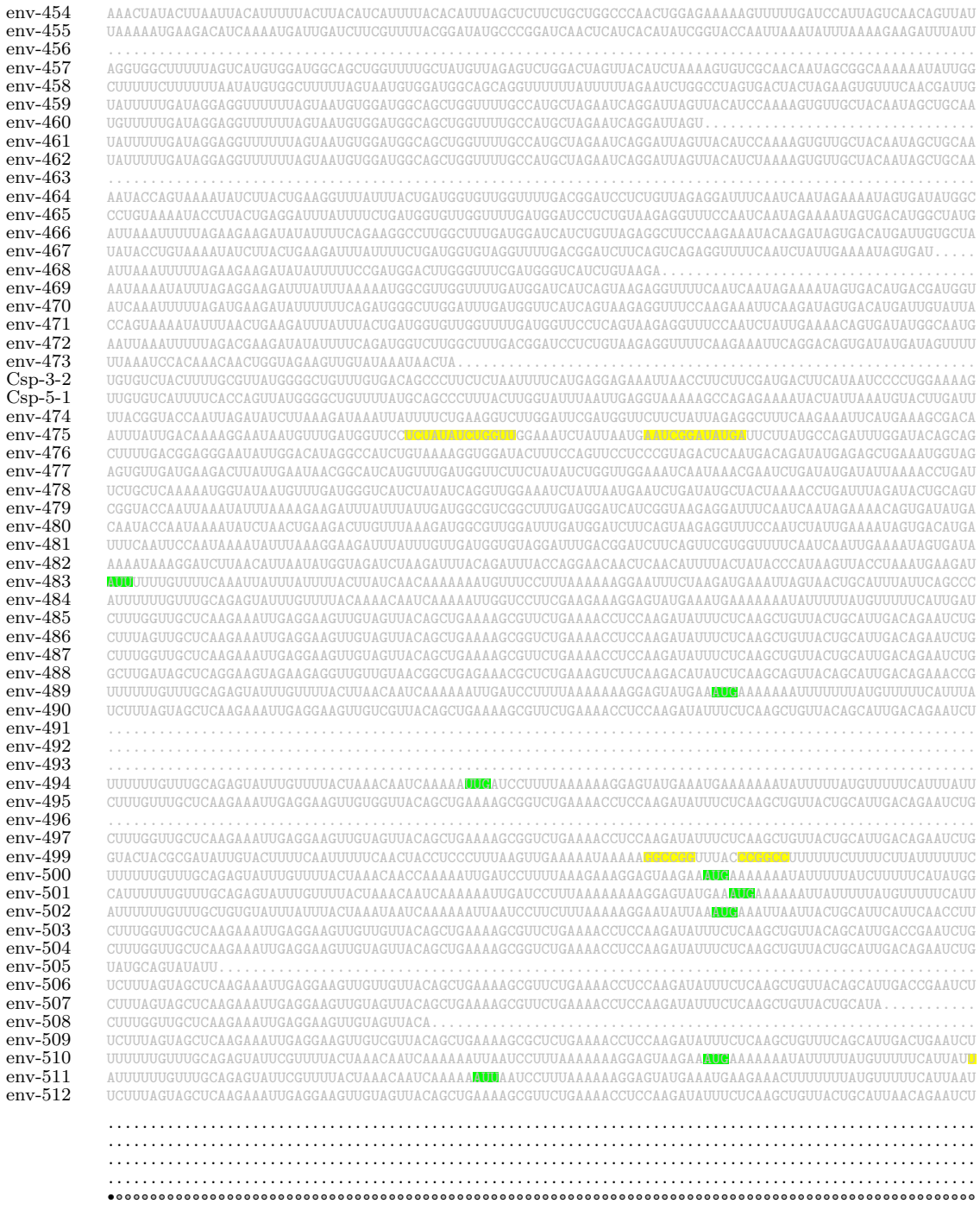

Csp-2-1
Csp-1-2
env-169
env-170
env-171
Csp-4-1
env-172
env-173
env-174
env-175
env-176
env-177
env-178
env-179
Cwa-1-1
Csp-4-2
Tel-1-1
env-180
Ssp-1-1
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
env-190
env-191
env-192
env-193
env-194
env-195
env-196
env-197
env-198
env-199
env-200
env-201
env-202
env-203
env-204
env-205
env-206
env-207
env-208
env-209
env-210
env-211
env-212
env-213
env-214
env-215
Lsp-1-1
Tel-1-2
env-216
Ssp-1-2
env-217
env-218

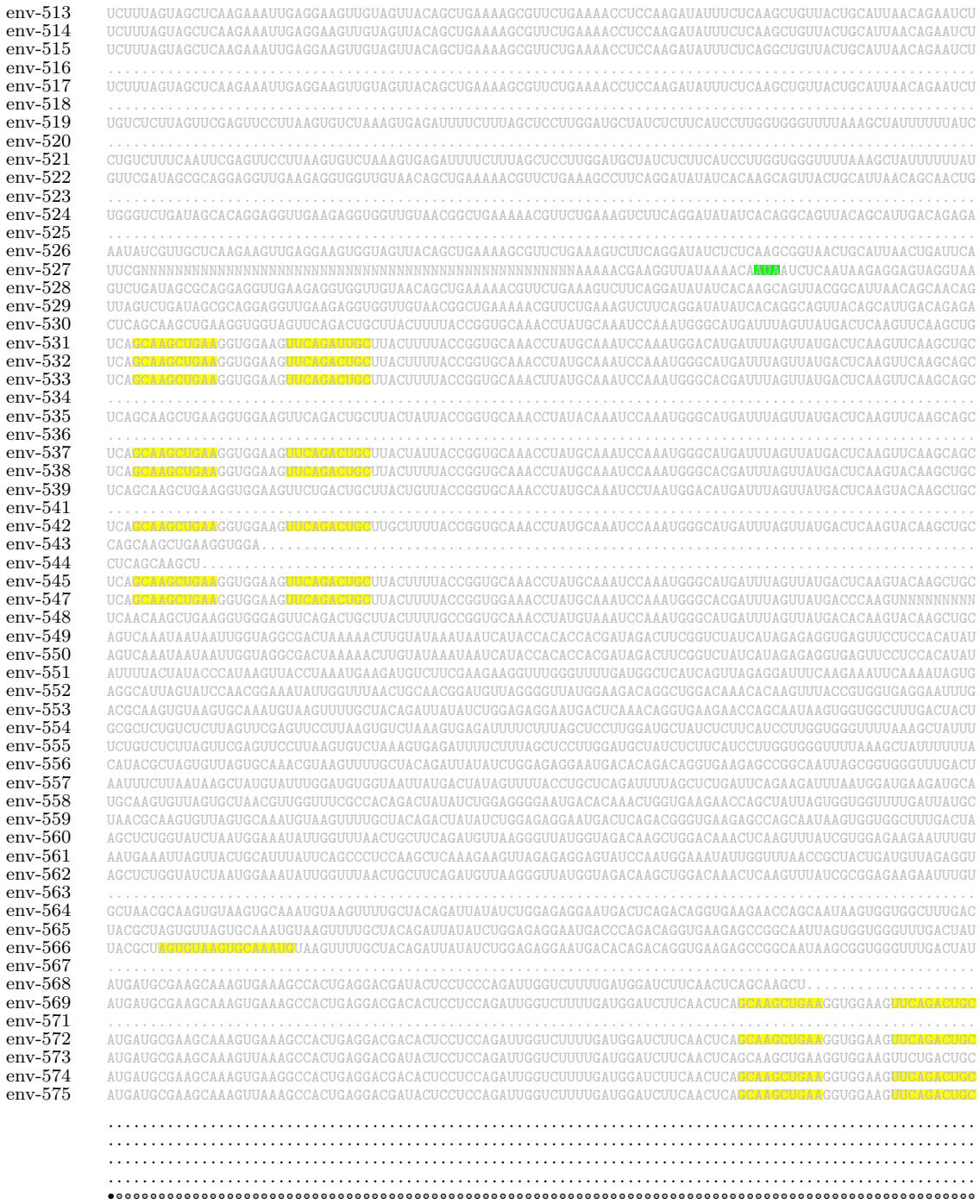

env-219
env-220
env-221
Csp-1-3
Cwa-1-2
env-222
env-223
env-225
Csp-4-3
env-226
env-227
env-228
env-229
env-230
env-231
env-232
Ssp-3-1
env-233
env-234
env-235
env-236
env-237
env-238
env-239
env-240
env-241
env-242
env-243
env-244
env-245
env-246
env-247
env-248
env-249
env-250
env-251
env-252
env-253
env-254
env-255
env-256
env-257
env-258
env-259
env-260
env-261
env-262
env-263
env-264
env-265
env-266
env-267
env-268
env-269
env-270
env-271
env-272
env-273
env-274
env-275

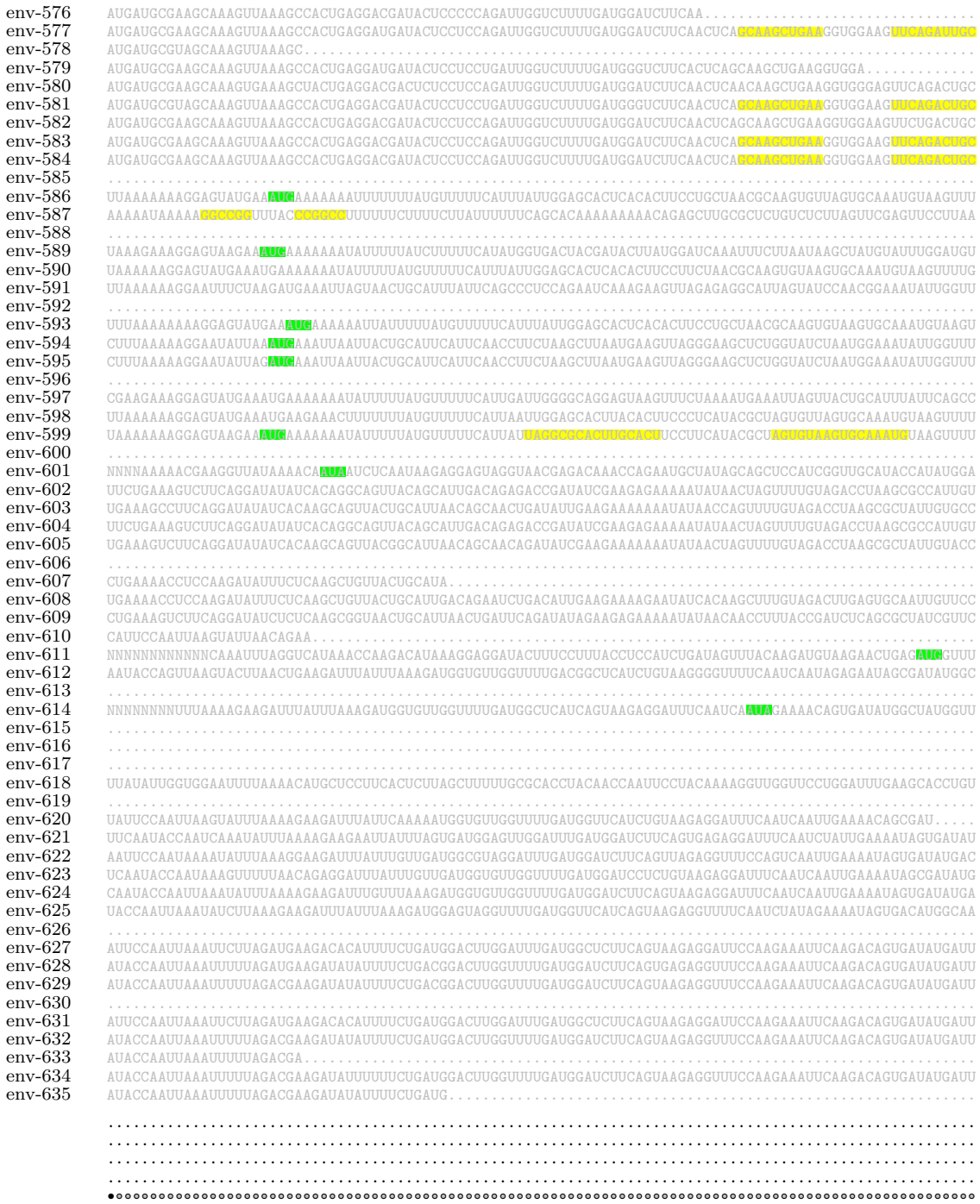
env-276
env-277
env-278
env-279
env-280
env-281
env-282
env-283
env-284
env-285
env-286
env-287
env-288
env-289
env-290
env-291
env-292
env-293
env-294
env-295
env-296
env-297
env-298
env-299
env-300
env-301
env-302
env-303
env-304
Ama-2-2
env-305
env-306
env-307
env-308
env-309
env-310
env-311
env-312
env-313
env-314
env-315
env-316
env-317
env-318
env-319
env-320
Mch-1-1
env-321
env-322
env-323
env-324
env-325
env-326
env-327
env-328
env-329
env-330
env-331
env-332
env-333

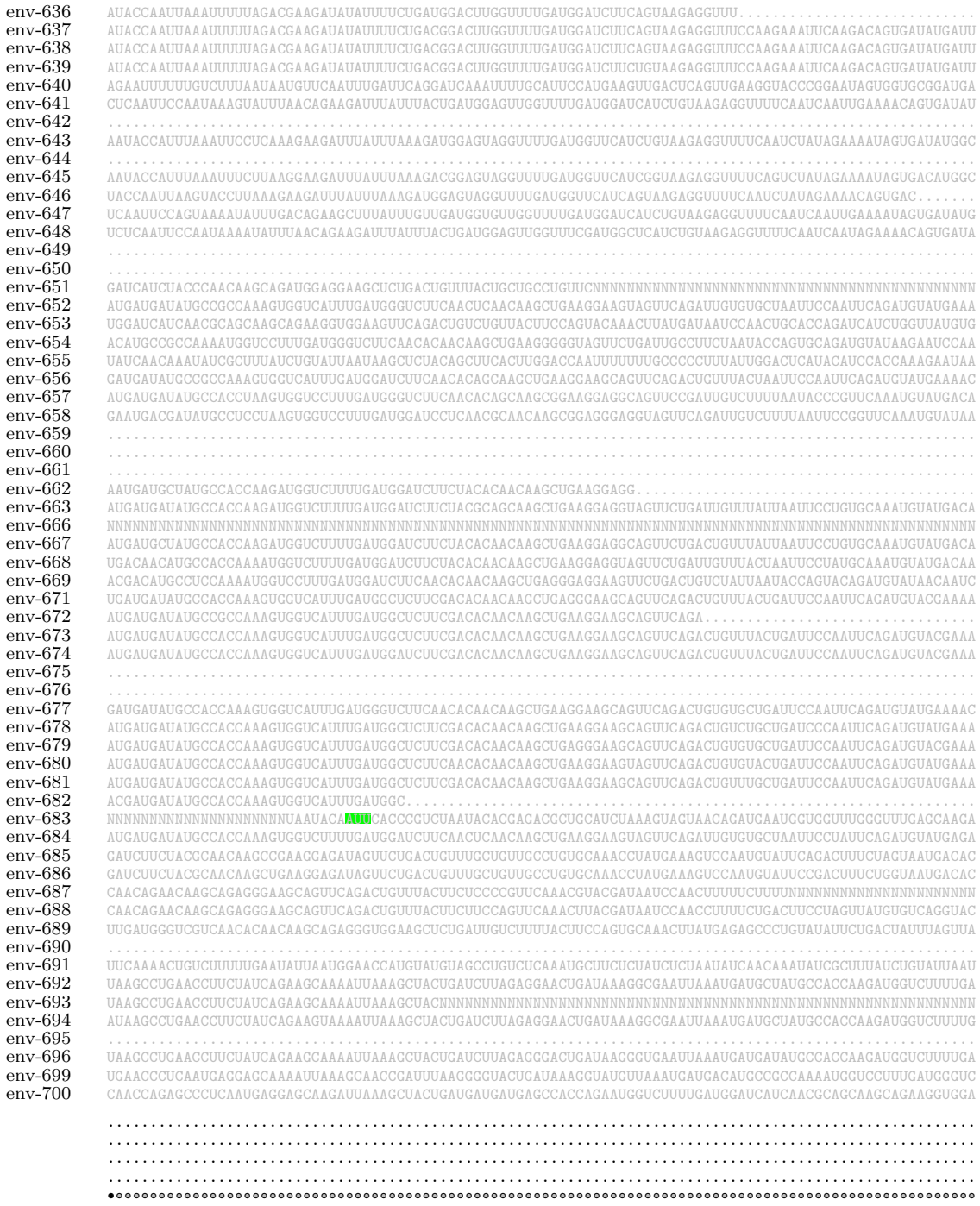

env-334
env-335
env-336
env-337
env-338
env-339
env-340
env-341
env-342
env-343
env-344
env-345
env-346
env-347
env-348
env-349
env-350
env-351
env-352
env-353
env-354
env-355
env-356
env-357
env-358
env-359
env-360
env-361
env-362
env-363
env-364
env-365
env-366
env-367
env-368
env-369
env-370
env-371
env-372
env-373
env-374
env-375
env-376
env-377
env-378
env-379
env-380
env-381
env-382
env-383
env-384
env-385
env-386
env-387
env-388
env-389
env-390
env-391
env-392
env-393

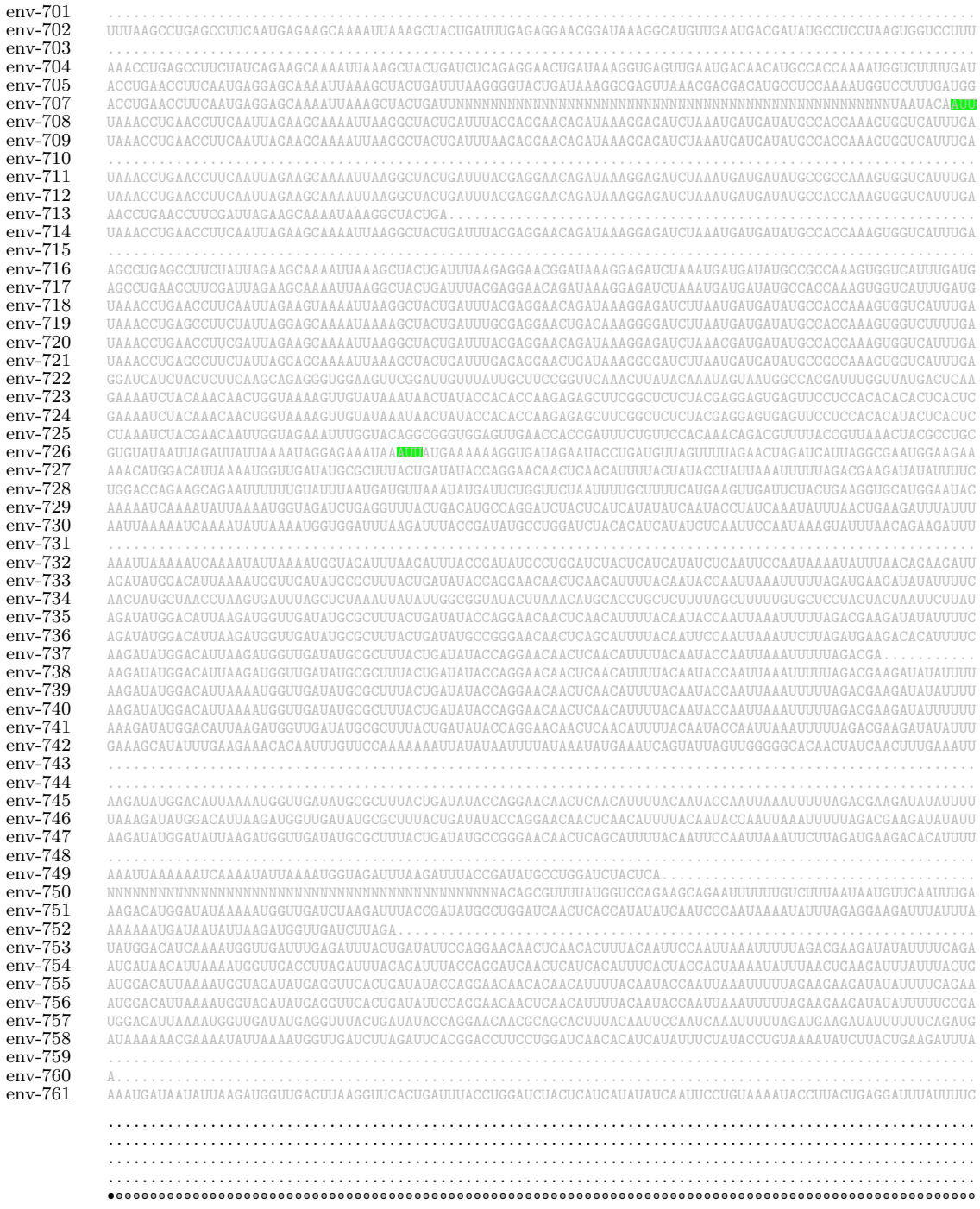
env-394
env-395
env-396
env-397
env-398
env-399
env-400
env-401
env-402
env-403
env-404
env-405
env-406
env-407
env-408
env-409
env-410
env-411
env-412
env-413
env-414
env-415
env-416
env-417
env-418
env-419
env-420
env-421
env-422
env-423
env-424
env-425
env-426
env-427
env-428
env-429
env-430
env-431
env-432
env-433
env-434
env-435
env-436
env-437
env-438
env-439
env-440
env-441
env-442
env-443
env-444
env-445
env-446
env-447
env-448
env-449
env-450
env-451
env-452
env-453

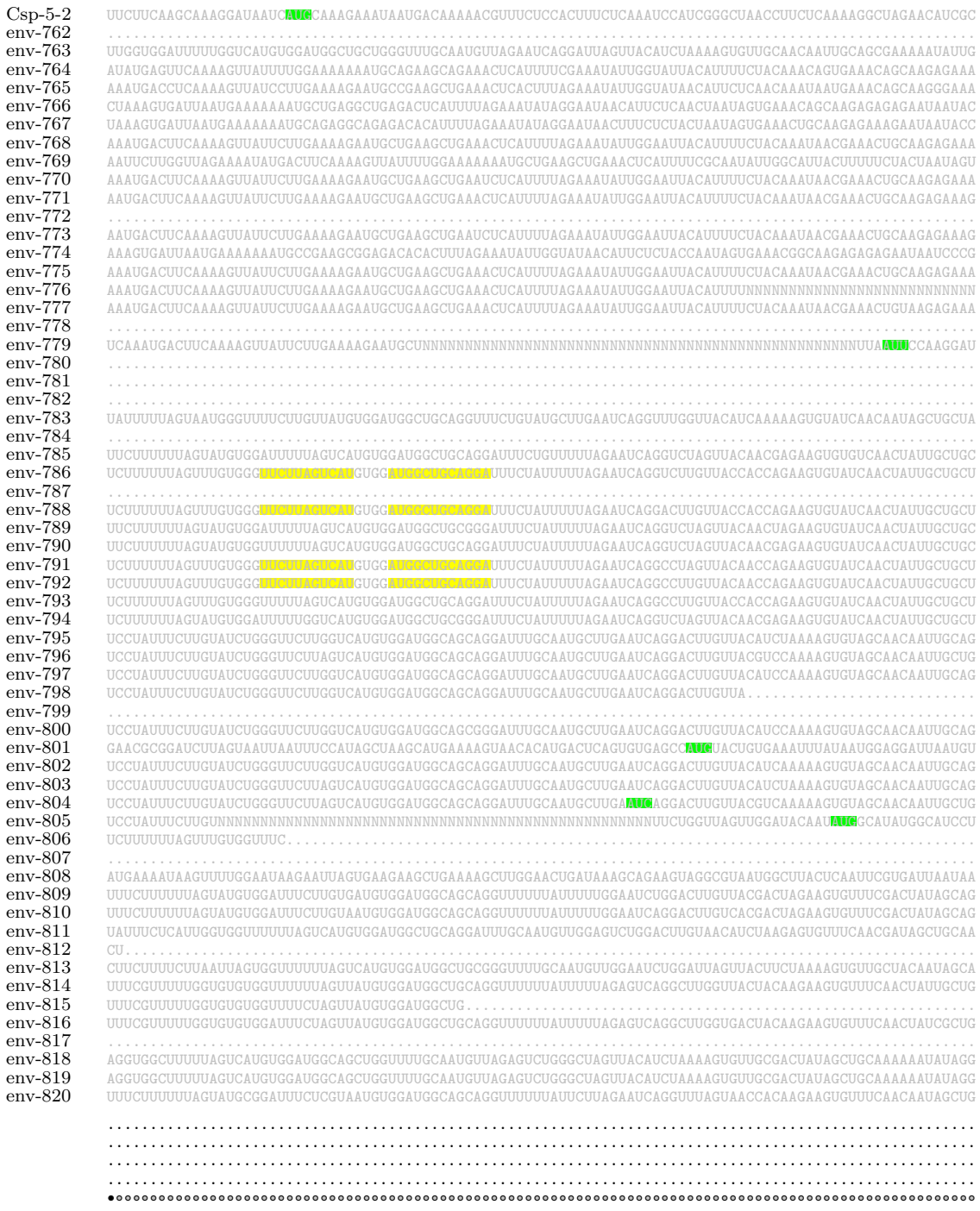

env-454
env-455
env-456
env-457
env-458
env-459
env-460
env-461
env-462
env-463
env-464
env-465
env-466
env-467
env-468
env-469
env-470
env-471
env-472
env-473
Csp-3-2
Csp-5-1
env-474
env-475
env-476
env-477
env-478
env-479
env-480
env-481
env-482
env-483
env-484
env-485
env-486
env-487
env-488
env-489
env-490
env-491
env-492
env-493
env-494
env-495
env-496
env-497
env-499
env-500
env-501
env-502
env-503
env-504
env-505
env-506
env-507
env-508
env-509
env-510
env-511
env-512

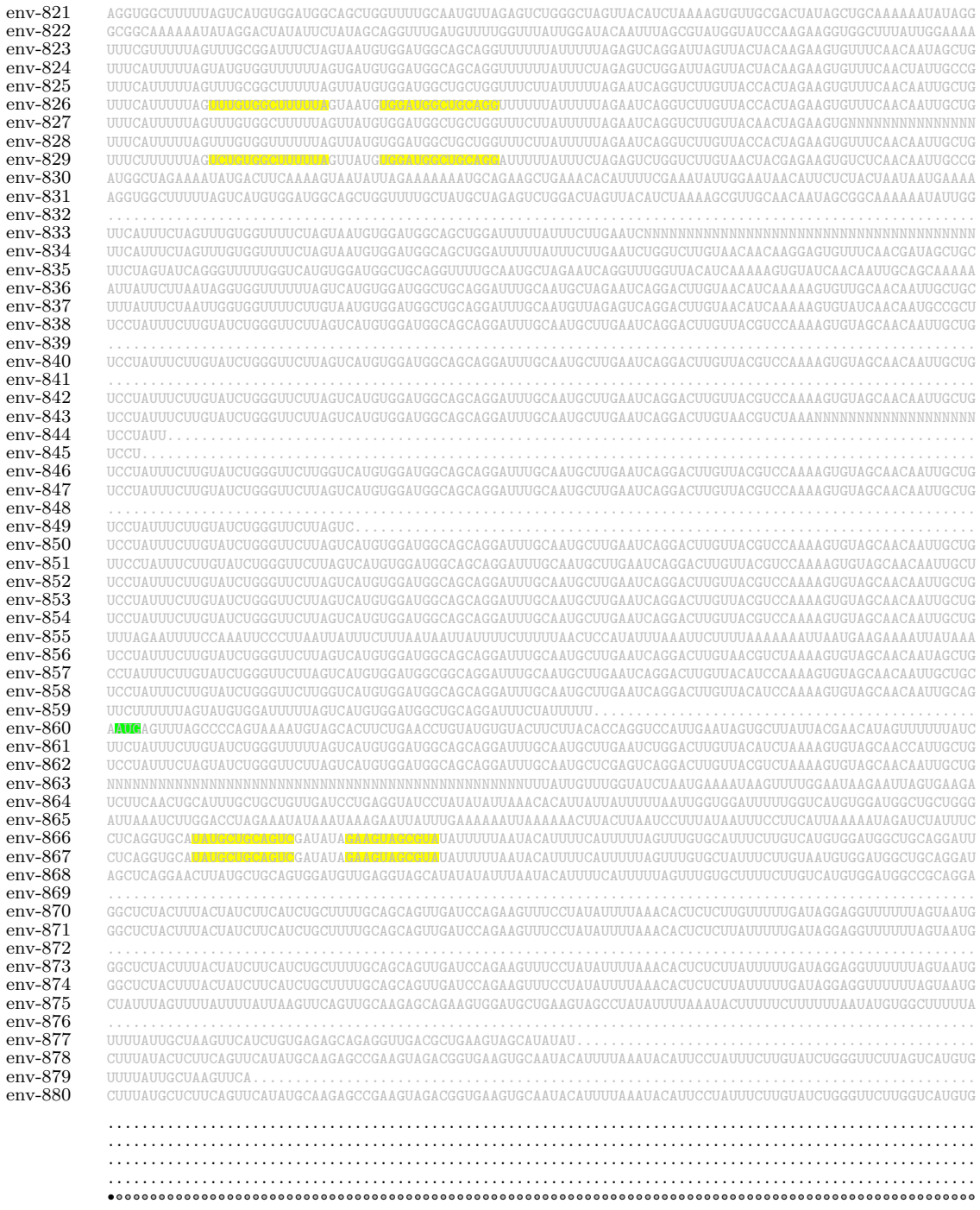

env-513
env-514
env-515
env-516
env-517
env-518
env-519
env-520
env-521
env-522
env-523
env-524
env-525
env-526
env-527
env-528
env-529
env-530
env-531
env-532
env-533
env-534
env-535
env-536
env-537
env-538
env-539
env-541
env-542
env-543
env-544
env-545
env-547
env-548
env-549
env-550
env-551
env-552
env-553
env-554
env-555
env-556
env-557
env-558
env-559
env-560
env-561
env-562
env-563
env-564
env-565
env-566
env-567
env-568
env-569
env-571
env-572
env-573
env-574
env-575

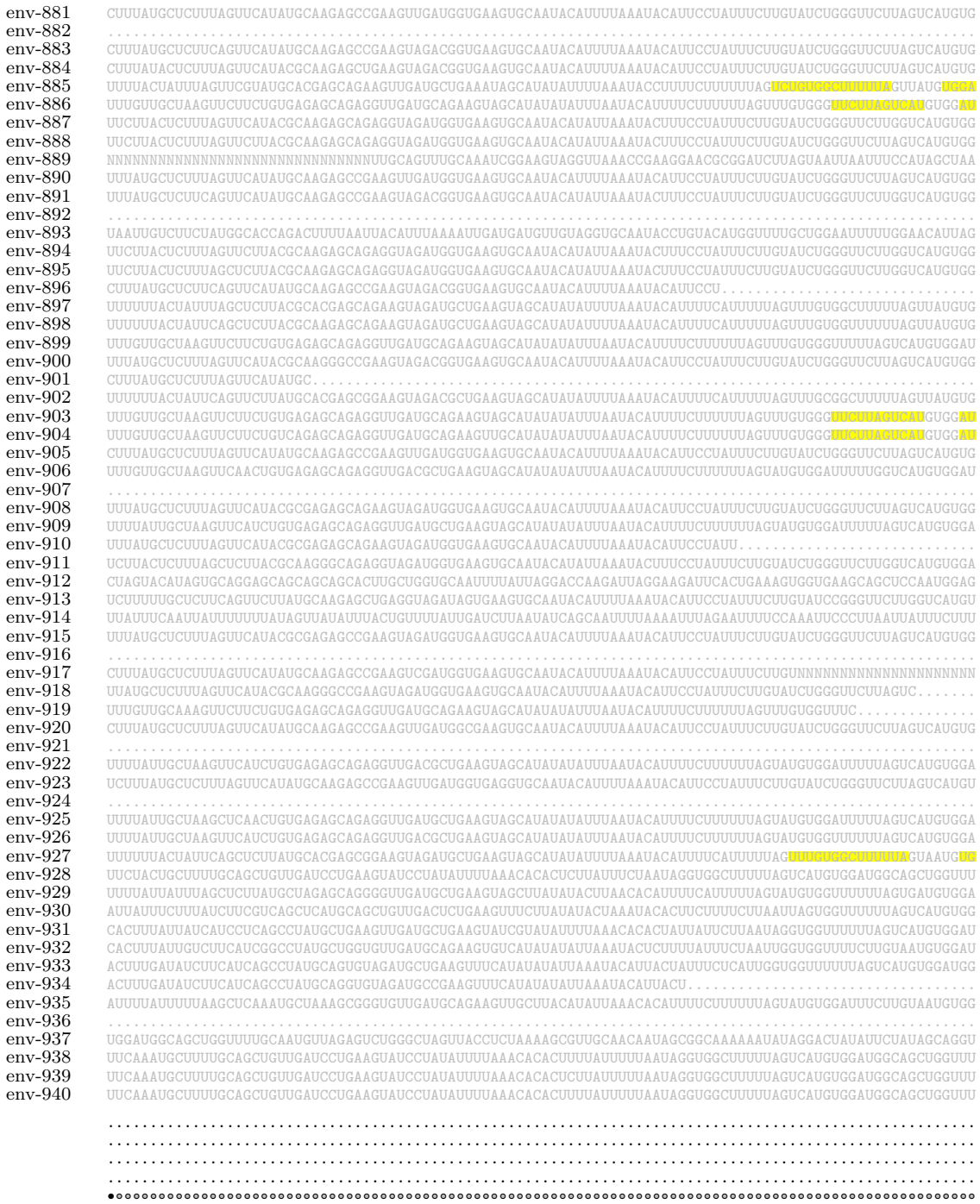

env-576
env-577
env-578
env-579
env-580
env-581
env-582
env-583
env-584
env-585
env-586
env-587
env-588
env-589
env-590
env-591
env-592
env-593
env-594
env-595
env-596
env-597
env-598
env-599
env-600
env-601
env-602
env-603
env-604
env-605
env-606
env-607
env-608
env-609
env-610
env-611
env-612
env-613
env-614
env-615
env-616
env-617
env-618
env-619
env-620
env-621
env-622
env-623
env-624
env-625
env-626
env-627
env-628
env-629
env-630
env-631
env-632
env-633
env-634
env-635

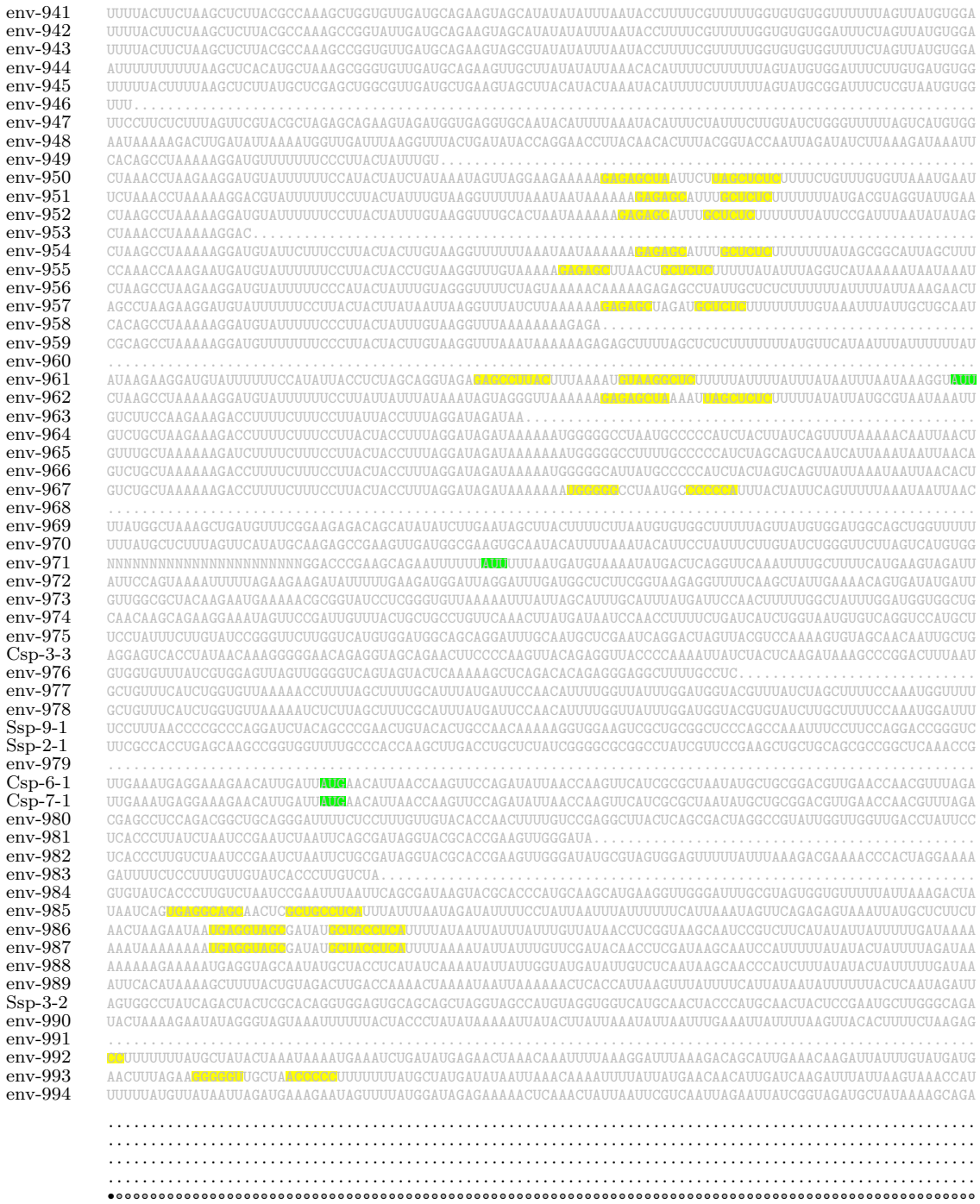

env-636
env-637
env-638
env-639
env-640
env-641
env-642
env-643
env-644
env-645
env-646
env-647
env-648
env-649
env-650
env-651
env-652
env-653
env-654
env-655
env-656
env-657
env-658
env-659
env-660
env-661
env-662
env-663
env-666
env-667
env-668
env-669
env-671
env-672
env-673
env-674
env-675
env-676
env-677
env-678
env-679
env-680
env-681
env-682
env-683
env-684
env-685
env-686
env-687
env-688
env-689
env-690
env-691
env-692
env-693
env-694
env-695
env-696
env-699
env-700

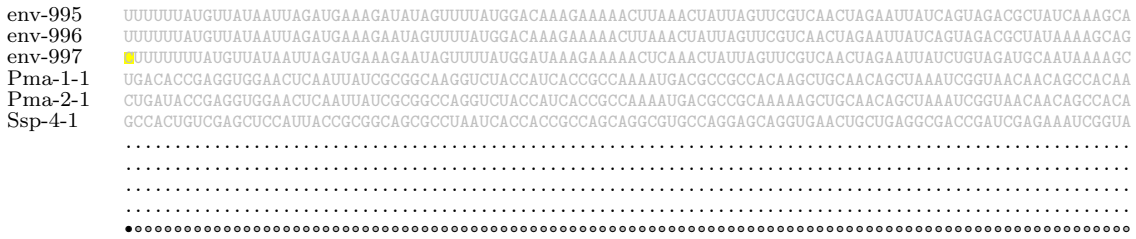

env-701
env-702
env-703
env-704
env-705
env-707
env-708
env-709
env-710
env-711
env-712
env-713
env-714
env-715
env-716
env-717
env-718
env-719
env-720
env-721
env-722
env-723
env-724
env-725
env-726
env-727
env-728
env-729
env-730
env-731
env-732
env-733
env-734
env-735
env-736
env-737
env-738
env-739
env-740
env-741
env-742
env-743
env-744
env-745
env-746
env-747
env-748
env-749
env-750
env-751
env-752
env-753
env-754
env-755
env-756
env-757
env-758
env-759
env-760
env-761


Csp-5-2
env-762
env-763
env-764
env-765
env-766
env-767
env-768
env-769
env-770
env-771
env-772
env-773
env-774
env-775
env-776
env-777
env-778
env-779
env-780
env-781
env-782
env-783
env-784
env-785
env-786
env-787
env-788
env-789
env-790
env-791
env-792
env-793
env-794
env-795
env-796
env-797
env-798
env-799
env-800
env-801
env-802
env-803
env-804
env-805
env-806
env-807
env-808
env-809
env-810
env-811
env-812
env-813
env-814
env-815
env-816
env-817
env-818
env-819
env-820


env-821
env-822
env-823
env-824
env-825
env-826
env-827
env-828
env-829
env-830
env-831
env-832
env-833
env-834
env-835
env-836
env-837
env-838
env-839
env-840
env-841
env-842
env-843
env-844
env-845
env-846
env-847
env-848
env-849
env-850
env-851
env-852
env-853
env-854
env-855
env-856
env-857
env-858
env-859
env-860
env-861
env-862
env-863
env-864
env-865
env-866
env-867
env-868
env-869
env-870
env-871
env-872
env-873
env-874
env-875
env-876
env-877
env-878
env-879
env-880


env-881
env-882
env-883
env-884
env-885
env-886
env-887
env-888
env-889
env-890
env-891
env-892
env-893
env-894
env-895
env-896
env-897
env-898
env-899
env-900
env-901
env-902
env-903
env-904
env-905
env-906
env-907
env-908
env-909
env-910
env-911
env-912
env-913
env-914
env-915
env-916
env-917
env-918
env-919
env-920
env-921
env-922
env-923
env-924
env-925
env-926
env-927
env-928
env-929
env-930
env-931
env-932
env-933
env-934
env-935
env-936
env-937
env-938
env-939
env-940


env-941
env-942
env-943
env-944
env-945
env-946
env-947
env-948
env-949
env-950
env-951
env-952
env-953
env-954
env-955
env-956
env-957
env-958
env-959
env-960
env-961
env-962
env-963
env-964
env-965
env-966
env-967
env-968
env-969
env-970
env-971
env-972
env-973
env-974
env-975
Csp-3-3
env-976
env-977
env-978
Ssp-9-1
Ssp-2-1
env-979
Csp-6-1
Csp-7-1
env-980
env-981
env-982
env-983
env-984
env-985
env-986
env-987
env-988
env-989
Ssp-3-2
env-990
env-991
env-992
env-993
env-994



38 GUCCY-hairpin

38.1 Taxa

The taxonomy of each organism containing a putative GUCCY-hairpin RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual GUCCY-hairpin RNA in Sections 38.2 and 38.4.

abbrev. of hits	taxonomy of species
Fba-1-1	Bacteria Bacteroidetes *Flavobacteria bacterium* BBFL7
Fba-2-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales *Flavobacteria bacterium* BAL38
Cat-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Croceibacter atlanticus* HTCC2559
Gfo-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Gramella forsetii* KT0803
Lbl-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Leeuwenhoekiella blandensis* MED217
Sru-1-1 to Sru-1-3	Bacteria Bacteroidetes Sphingobacteria Sphingobacteriales Rhodothermaceae *Salinibacter ruber* DSM 13855
Cph-1-1 to Cph-1-134	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium phaeobacteroides* BS1
Fpe-1-1 to Fpe-1-8	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Aurantimonadaceae *Fulvimarina pelagi* HTCC2506
Bsp-1-1 to Bsp-1-3	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Bradyrhizobium sp.* BTAi1
Bsp-2-1 to Bsp-2-63	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Bradyrhizobium sp.* ORS278
Nha-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Nitrobacter hamburgensis* X14
Atu-1-1 to Atu-1-6	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Rhizobiaceae Rhizobium/Agrobacterium group *Agrobacterium tumefaciens str.* C58
Oal-1-1 to Oal-1-6	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Hyphomonadaceae *Oceanicaulis alexandrii* HTCC2633
Lag-1-1 to Lag-1-2	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Labrenzia aggregata* IAM 12614
Rsp-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Roseovarius sp.* 217
Mma-1-1 to Mma-1-24	Bacteria Proteobacteria Alphaproteobacteria Rhodospirillales Rhodospirillaceae *Magnetospirillum magnetotacticum* MS-1
Eli-1-1 to Eli-1-6	Bacteria Proteobacteria Alphaproteobacteria Sphingomonadales Erythrobacteraceae *Erythrobacter litoralis* HTCC2594
Nar-1-1 to Nar-1-6	Bacteria Proteobacteria Alphaproteobacteria Sphingomonadales Sphingomonadaceae *Novosphingobium aromaticivorans* DSM 12444
Ssp-1-1 to Ssp-1-4	Bacteria Proteobacteria Alphaproteobacteria Sphingomonadales Sphingomonadaceae *Sphingomonas sp.* SKA58
Swi-1-1 to Swi-1-18	Bacteria Proteobacteria Alphaproteobacteria Sphingomonadales Sphingomonadaceae *Sphingomonas wittichii* RW1
Sal-1-1 to Sal-1-16	Bacteria Proteobacteria Alphaproteobacteria Sphingomonadales Sphingomonadaceae *Sphingopyxis alaskensis* RB2256
Psp-1-1 to Psp-1-40	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae *Polaromonas sp.* JS666
Dol-1-1 to Dol-1-2	Bacteria Proteobacteria Deltaproteobacteria Desulfobacterales Desulfobacteraceae *Desulfococcus oleovorans* Hxd3
Ssp-2-1 to Ssp-2-2	Bacteria Proteobacteria Epsilonproteobacteria *Sulfurovum sp.* NBC37-1
Osp-1-1 to Osp-1-70	Bacteria Proteobacteria Gammaproteobacteria Oceanospirillales *Oceanobacter sp.* RED65
Fno-1-1	Bacteria Proteobacteria Gammaproteobacteria Thiotrichales Francisellaceae *Francisella novicida* FTG
Fno-2-1 to Fno-2-2	Bacteria Proteobacteria Gammaproteobacteria Thiotrichales Francisellaceae *Francisella novicida* GA99-3548
Fno-3-1	Bacteria Proteobacteria Gammaproteobacteria Thiotrichales Francisellaceae *Francisella novicida* GA99-3549
env-1 to env-227	environmental samples

38.2 Gene contexts

Each GUCCY-hairpin RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one GUCCY-hairpin RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 38.3. The accession number of the sequence containing each GUCCY-hairpin RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 38.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Sru-1-1	NC_007677.1	-	2503472	2503438	←COG3677 (COG3677) **RNA**→ pyrophosphatase (cd00412)→
Sru-1-2	NC_007677.1	+	2503378	2503412	←pyrophosphatase (cd00412) **RNA**→ COG3677 (COG3677)→
Sru-1-3	NC_007677.1	+	2631441	2631475	←COG0820 (COG0820) **RNA**→ Histidinol_dh (pfam00815)→
env-1	JCVI_SCAF_1101668563340	-	164	130	**RNA**→
env-2	wf3_AHAI11607_x1	+	492	528	**RNA**→ ←COG2827 (COG2827)
env-3	JCVI_SCAF_1096627189024	+	839	873	S1 (pfam00575)? (GOS501)→ **RNA**→ ? (GOS8865)→
env-4	wf3_AHAI14153_x1	-	367	332	**RNA**→
env-5	wf2_AHAA16149_x1	+	359	394	**RNA**→
env-6	wf2_AHAA9624_x1	-	495	459	Lrp (COG1522)→ **RNA**→ ←GreA (COG0782)
env-7	JCVI_SCAF_1101668606422	-	364	330	RecR (COG0353)? (GOS8597)→ **RNA**→
Cph-1-1	NC_010831.1	-	2054795	2054759	RfbD (COG1091)→ **RNA**→ **RNA**→ **RNA**→ hypo→
Cph-1-4	NC_010831.1	-	2055131	2055095	(shown above)
Cph-1-10	NC_010831.1	-	2054626	2054590	(shown above)
Cph-1-2	NC_010831.1	+	2658989	2659025	←hypo **RNA**→ ←VapI (COG3093)DUF955 (pfam06114)
Cph-1-3	NC_010831.1	-	2265338	2265302	hypo→ **RNA**→ hypo→
Cph-1-5	NC_010831.1	-	1165871	1165835	←COG2503 (COG2503) **RNA**→ hypo→
Cph-1-6	NC_010831.1	-	873662	873626	←hypo **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ KdtA (COG1519)→
Cph-1-11	NC_010831.1	-	874303	874267	(shown above)
Cph-1-12	NC_010831.1	-	873571	873535	(shown above)
Cph-1-17	NC_010831.1	-	874028	873992	(shown above)
Cph-1-23	NC_010831.1	-	873937	873901	(shown above)
Cph-1-7	NC_010831.1	+	2036783	2036819	COG1600 (COG1600)→ **RNA**→ ←AICARFT_IMPCHas (pfam01808)MGS (pfam02142)
Cph-1-8	NC_010831.1	+	2372357	2372393	FolK (COG0801)→ **RNA**→ ←ArsA (COG0003)
Cph-1-9	NC_010831.1	-	2077580	2077544	dTDP_sugar_isom (pfam00908)→ **RNA**→ hypo→
Cph-1-14	NC_010831.1	+	2132099	2132135	←hypo **RNA**→ **RNA**→ **RNA**→ ←COG2827 (COG2827)
Cph-1-13	NC_010831.1	+	2131681	2131717	(shown above)
Cph-1-19	NC_010831.1	+	2131973	2132009	(shown above)
Cph-1-15	NC_010831.1	-	2062776	2062740	GlmS (COG0449)→ **RNA**→ **RNA**→ CysQ (cd01638)→
Cph-1-21	NC_010831.1	-	2062609	2062573	(shown above)
Cph-1-16	NC_010831.1	-	2032719	2032683	COG3349 (COG3349)Rieske (cd03467)→ **RNA**→ ←Beta-lactamase (pfam00144)
Cph-1-18	NC_010831.1	+	1524680	1524716	COG0702 (COG0702)→ **RNA**→ PepN (COG0308)→
Cph-1-20	NC_010831.1	-	2368393	2368357	←Tig (COG0544) **RNA**→ ←PqqL (COG0612)
Cph-1-22	NC_010831.1	-	1714522	1714486	HyaD (COG0680)→ **RNA**→ ←PerM (COG0628)
Cph-1-24	NC_010831.1	+	910818	910854	IspE (COG1947)→ **RNA**→ ←COG0388 (COG0388)
env-8	JCVI_SCAF_1101668633095	+	173	207	**RNA**→ ←Rnd (COG0349)? (GOS1030)
env-9	JCVI_SCAF_1101668590036	+	288	322	**RNA**→ ←COG2368 (COG2368)? (GOS490)
env-10	soil_Contig25495	+	511	545	hypo→ **RNA**→ ←hypo
env-11	JCVI_SCAF_1101668610559	+	1165	1199	**RNA**→ COG2827 (COG2827)? (GOS1465)→
env-12	wf2_AHAA744_b1	+	664	700	Lrp (COG1522)→ **RNA**→
env-13	soil_XZS98530_b1	-	508	474	PurS (COG1828)→ **RNA**→ ←(EC:4.2.3.4)
env-14	JCVI_SCAF_1101668115658	+	878	912	←RhtB (COG1280)? (GOS192) **RNA**→ ←Acetyltransf_1 (pfam00583)? (GOS8607)

env-15	wf3_AHAI20491_b2	-	852	816	**RNA**→
env-16	wf3_AHAI1709_g1	-	422	386	**RNA**→ ←COG3565 (COG3565)
env-17	wf3_AHAI1027_g1	+	221	257	←hypo **RNA**→ ←hypo
env-18	wf2_AHAA9624_x1	+	407	443	GreA (COG0782)→ **RNA**→ ←Lrp (COG1522)
env-19	wf2_Contig447	-	381	345	COG1610 (COG1610)→ **RNA**→
env-20	wf2_AHAA10811_x1	+	242	276	←hypo **RNA**→ ←hypo
env-21	ABEF01009878.1	+	636	670	←hypo **RNA**→ hypo→
Cat-1-1	NZ_AAMP01000001.1	+	1202353	1202387	←CaiD (COG1024) **RNA**→ ←DUF59 (pfam01883)
Fno-2-1	NZ_ABAH01000013.1	-	125524	125490	**RNA**→
env-22	JCVI_SCAF_1101667020378	+	191	225	**RNA**→
env-23	JCVI_SCAF_1101668604990	+	828	862	←Glu tRNA **RNA**→ ? (GOS1264)→
env-24	JCVI_SCAF_1096627479218	+	1060	1094	Glyoxalase (pfam00903)→ **RNA**→ ←? (GOS5834)
env-25	wf2_AHAA2242_g1	+	351	385	Signal peptidase I→ **RNA**→ ←hypo
env-26	wf2_AHAA8239_y1	+	146	180	**RNA**→ ←DapB (COG0289)
env-27	wf3_AHAI18676_g2	+	743	777	←RbfA (COG0858) **RNA**→
env-28	wf2_AHAA6174_x1	+	587	621	←TrpF (COG0135) **RNA**→
env-29	wf3_Contig7131	-	3020	2986	**RNA**→ ←hypo
Nar-1-1	NC_007794.1	-	1239615	1239581	←hypo **RNA**→ hypo→
env-30	JCVI_SCAF_1101668567100	-	1453	1419	**RNA**→ DegQ (COG0265)? (GOS899)→
env-31	JCVI_SCAF_1096627132434	+	591	625	←RpoA (COG0202)? (GOS5342) **RNA**→ ←RpsD (COG0522)? (GOS1133)
env-32	JCVI_SCAF_1096628043581	-	498	464	←COG1611 (COG1611)? (GOS5783) **RNA**→
env-33	JCVI_SCAF_1101668066154	+	1128	1162	DLH (pfam01738)? (GOS6241)→ **RNA**→
Cph-1-25	NC_010831.1	-	507736	507702	SUA5 (COG0009)SUA5 (pfam03481)→ **RNA**→ ←hypo
env-34	soil_XZS49895_b1	-	414	380	hypo→ **RNA**→ ←hypo
Lbl-1-1	NZ_AANC01000002.1	+	392992	393026	COG1416 (COG1416)→ **RNA**→ putative oxidoreductase→
env-35	JCVI_SCAF_1096626614167	+	1257	1291	CsdB (COG0520)? (GOS1308)→ **RNA**→ COG3695 (COG3695)? (GOS5716)→
Osp-1-1	NZ_AAQH01000002.1	+	31948	31982	GalU (COG1210)→ **RNA**→ **RNA**→ **RNA**→ GalE (COG1087)→
Osp-1-34	NZ_AAQH01000002.1	+	31986	32020	(shown above)
Osp-1-37	NZ_AAQH01000002.1	+	31908	31942	(shown above)
env-36	JCVI_SCAF_1101668135384	+	316	350	RL11 (smart00649)? (GOS625)→ **RNA**→ Ribosomal_L1 (pfam00687)? (GOS6270)→
env-37	JCVI_SCAF_1096627950213	+	1388	1422	Transposase_20 (pfam02371)? (GOS4503)→ **RNA**→
env-38	JCVI_SCAF_1096627621690	-	844	810	COG0730 (COG0730)? (GOS644)→ **RNA**→ ←CTP_transf_1 (pfam01148)? (GOS130)
env-39	JCVI_SCAF_1101668553963	-	924	890	? (GOS5950)→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ Cation_ATPase_N (pfam00690)? (GOS168)→
env-40	JCVI_SCAF_1101668553963	-	878	844	(shown above)
env-41	JCVI_SCAF_1101668553963	-	832	798	(shown above)
env-42	JCVI_SCAF_1101668553963	-	786	752	(shown above)
Sal-1-1	NC_008048.1	-	3118317	3118283	hypo→ **RNA**→ ←RlpA (COG0797)
Fba-1-1	NZ_AAPD01000004.1	+	129200	129234	←HsdM (COG0286) **RNA**→ ←hypo (cons)
Osp-1-2	NZ_AAQH01000002.1	-	51511	51477	ArgF (COG0078)→ **RNA**→ COG0436 (COG0436)→
env-43	JCVI_SCAF_1101668528582	+	472	506	**RNA**→ ←COG1521 (COG1521)? (GOS7235)
Nar-1-2	NC_007794.1	+	3277600	3277634	←tRNA-Leu **RNA**→ ←SppA (COG0616)
env-44	wf3_AHAI13901_x1	-	673	639	←RuvB_N (pfam05496) **RNA**→ COG2827 (COG2827)→
env-45	wf3_AHAI13901_x1	+	201	235	**RNA**→ ←COG2827 (COG2827)
Nar-1-3	NC_007794.1	+	3481136	3481170	←IleS (COG0060) **RNA**→ ←RibF (COG0196)
Nar-1-4	NC_007794.1	-	2437858	2437824	NuoF (COG1894)→ **RNA**→ NuoG (COG1034)→
Nar-1-5	NC_007794.1	-	910681	910647	←CaiD (COG1024) **RNA**→ ←CaiD (COG1024)
env-46	wf3_AHAI10348_x1	-	599	565	**RNA**→ fer2 (cd00207)→
env-47	JCVI_SCAF_1096627925836	-	219	185	Transposase_12 (pfam01610)? (GOS4676)→ **RNA**→
env-48	wf1_AGZO18857_g2	-	562	528	**RNA**→ ←hypo
Osp-1-3	NZ_AAQH01000010.1	-	69162	69128	CaiD (COG1024)→ **RNA**→ ←TPMT (pfam05724)
Osp-1-4	NZ_AAQH01000012.1	-	33667	33633	hypo→ **RNA**→ ←USP_Like (cd00293)Usp (pfam00582)
Osp-1-5	NZ_AAQH01000002.1	-	60512	60478	hypo→ **RNA**→ ←LysR (COG0583)

Osp-1-6	NZ_AAQH01000012.1	+	22920	22954	←MalT (COG2909) **RNA**→ ←DUF1329 (pfam07044)
Osp-1-7	NZ_AAQH01000002.1	-	16980	16946	←HisC (COG0079) **RNA**→ ←PheA (COG0077)PheA (COG1605)
Osp-1-8	NZ_AAQH01000006.1	-	24646	24612	←FlgF (COG4787) **RNA**→ ←FlgE (COG1749)FlgG (COG4786)
Osp-1-9	NZ_AAQH01000002.1	-	29169	29135	←YSH1 (COG1236) **RNA**→ ←PGI (pfam00342)
Osp-1-10	NZ_AAQH01000012.1	+	41465	41499	MhpC (COG0596)→ **RNA**→ **RNA**→ ←sugar fermentation stimulation protein A
Osp-1-11	NZ_AAQH01000012.1	+	41289	41323	(shown above)
Osp-1-12	NZ_AAQH01000002.1	-	65811	65777	COG4230 (COG4230)→ **RNA**→ ←COG2827 (COG2827)
Osp-1-13	NZ_AAQH01000004.1	-	37225	37191	FabG (COG1028)→ **RNA**→ **RNA**→ ←FolK (COG0801)
Osp-1-20	NZ_AAQH01000004.1	-	37313	37279	(shown above)
Osp-1-14	NZ_AAQH01000002.1	-	15444	15410	←PheA (COG0077)PheA (COG1605) **RNA**→ **RNA**→ **RNA**→ ←SerC (COG1932)
Osp-1-15	NZ_AAQH01000002.1	-	15355	15321	(shown above)
Osp-1-16	NZ_AAQH01000002.1	-	15533	15499	(shown above)
Osp-1-17	NZ_AAQH01000024.1	+	16710	16744	GST_C_GTT1_like (cd03189)GST_N_GTT1_like (cd03046)→ **RNA**→ ←NIF3 (pfam01784)
Osp-1-18	NZ_AAQH01000008.1	+	41000	41034	←DUF6 (pfam00892) **RNA**→ ←TufB (COG0050)
Osp-1-19	NZ_AAQH01000014.1	+	76982	77016	←hypo **RNA**→ ←5S ribosomal RNA
Osp-1-21	NZ_AAQH01000001.1	+	503096	503130	hypo→ **RNA**→ **RNA**→ ←AcrB (COG0841)
Osp-1-22	NZ_AAQH01000001.1	+	503010	503044	(shown above)
env-49	wf2_AHAA6604_x1	-	422	388	**RNA**→
Osp-1-23	NZ_AAQH01000004.1	+	37024	37058	FolK (COG0801)→ **RNA**→ **RNA**→ **RNA**→ ←FabG (COG1028)
Osp-1-25	NZ_AAQH01000004.1	+	37230	37264	(shown above)
Osp-1-67	NZ_AAQH01000004.1	+	37142	37176	(shown above)
Osp-1-24	NZ_AAQH01000024.1	-	16793	16759	NIF3 (pfam01784)→ **RNA**→ ←GST_C_GTT1_like (cd03189)GST_N_GTT1_like (cd03046)
Osp-1-26	NZ_AAQH01000001.1	+	510187	510221	COG4392 (COG4392)→ **RNA**→ ←EFP (pfam01132)
Osp-1-27	NZ_AAQH01000002.1	-	61459	61425	COG1161 (COG1161)→ **RNA**→ putative nucleoprotein/polynucleotide-associated enzyme→
Osp-1-28	NZ_AAQH01000011.1	+	59903	59937	←PRX_Typ2cys (cd03015) **RNA**→ ←OLE1 (COG1398)
Osp-1-29	NZ_AAQH01000011.1	+	55209	55243	Acs (COG0365)→ **RNA**→ **RNA**→ **RNA**→ ←hypo
Osp-1-30	NZ_AAQH01000011.1	+	55383	55417	(shown above)
Osp-1-31	NZ_AAQH01000011.1	+	55296	55330	(shown above)
env-50	wf2_AHAA6604_x1	-	127	93	**RNA**→
env-51	ABEF01035071.1	+	182	216	hypo→ **RNA**→ DUF861 (pfam05899)→
env-52	JCVI_SCAF_1101668637199	+	31	65	**RNA**→ ←RplI (COG0359)? (GOS1168)
env-53	JCVI_SCAF_1096627473590	+	1475	1509	←? (GOS14402686) **RNA**→
env-54	JCVI_SCAF_1101668550812	+	119	153	**RNA**→ HATPase_c (pfam02518)? (GOS11062282)→
env-55	wf2_Contig82	+	189	223	**RNA**→
env-56	JCVI_SCAF_1101668149540	+	1175	1209	? (GOS5684)→ **RNA**→
env-57	ABEF01014020.1	+	271	305	hypo→ **RNA**→ hypo→
env-58	wf3_Contig7757	+	2649	2684	Gst (COG0625)→ **RNA**→ ←RimI (COG0456)
env-59	wf1_Contig1736	-	215	181	←PseudoU_synth_TruC (cd02563) **RNA**→
env-60	JCVI_SCAF_1101668613633	-	805	771	**RNA**→
env-61	ABDH01006045.1	+	5	39	**RNA**→ ←DUF62 (pfam01887)
env-62	ABDH01008616.1	-	421	387	←Ribosomal_L27 (pfam01016) **RNA**→ ←Ribosomal_L21p (pfam00829)
env-63	wf2_AHAA16979_x1	-	545	511	←LexA_DNA_bind (pfam01726) **RNA**→
Osp-1-32	NZ_AAQH01000001.1	+	416898	416932	DadA (COG0665)→ **RNA**→ hypo→
env-64	wf1_AGZO6705_g2	-	539	505	**RNA**→ ←Predicted ATPase of the PP-loop superfamily implicated in cell cycle control
env-65	wf1_AGZO6705_b2	+	79	113	**RNA**→ radical_SAM_B12_BD (cd02068)→
Osp-1-33	NZ_AAQH01000001.1	+	417582	417616	hypo→ **RNA**→ ←hypo
env-66	wf2_AHAA6604_x1	+	274	309	**RNA**→
env-67	wf2_AHAA6604_x1	+	430	465	**RNA**→
Osp-1-35	NZ_AAQH01000012.1	-	22827	22793	DUF1329 (pfam07044)→ **RNA**→ MalT (COG2909)→
Osp-1-36	NZ_AAQH01000002.1	+	30685	30719	YSH1 (COG1236)→ **RNA**→ **RNA**→ GalU (COG1210)→
Osp-1-56	NZ_AAQH01000002.1	+	30722	30756	(shown above)
Osp-1-38	NZ_AAQH01000006.1	+	52682	52716	FliD (COG1345)→ **RNA**→ ←TPR (cd00189)UbiE (COG2226)

Osp-1-39	NZ_AAQH01000006.1	+	24563	24597	FlgE (COG1749)FlgG (COG4786)→ **RNA**→ **RNA**→ FlgF (COG4787)→
Osp-1-60	NZ_AAQH01000006.1	+	24713	24747	(shown above)
Osp-1-40	NZ_AAQH01000001.1	+	348759	348793	COG2130 (COG2130)→ **RNA**→ ←BioA (COG0161)
Osp-1-41	NZ_AAQH01000011.1	-	55292	55258	hypo→ **RNA**→ **RNA**→ ←Acs (COG0365)
Osp-1-42	NZ_AAQH01000011.1	-	55182	55148	(shown above)
Osp-1-43	NZ_AAQH01000011.1	-	59987	59953	OLE1 (COG1398)→ **RNA**→ PRX_Typ2cys (cd03015)→
Osp-1-44	NZ_AAQH01000012.1	-	41460	41426	sugar fermentation stimulation protein A→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ ←MhpC (COG0596)
Osp-1-45	NZ_AAQH01000012.1	-	41548	41514	(shown above)
Osp-1-46	NZ_AAQH01000012.1	-	41196	41162	(shown above)
Osp-1-47	NZ_AAQH01000012.1	-	41636	41602	(shown above)
Osp-1-53	NZ_AAQH01000012.1	-	41372	41338	(shown above)
Osp-1-55	NZ_AAQH01000012.1	-	41284	41250	(shown above)
Osp-1-48	NZ_AAQH01000002.1	+	15361	15395	SerC (COG1932)→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ PheA (COG0077)PheA (COG1605)→
Osp-1-49	NZ_AAQH01000002.1	+	15450	15484	(shown above)
Osp-1-50	NZ_AAQH01000002.1	+	15539	15573	(shown above)
Osp-1-51	NZ_AAQH01000002.1	+	15628	15662	(shown above)
Osp-1-52	NZ_AAQH01000002.1	+	6166	6200	←GGDEF (cd01949) **RNA**→ ←FabG (COG1028)
Osp-1-54	NZ_AAQH01000002.1	+	16897	16931	PheA (COG0077)PheA (COG1605)→ **RNA**→ HisC (COG0079)→
Osp-1-57	NZ_AAQH01000010.1	+	72036	72070	←ACAD_fadE5 (cd01153) **RNA**→ **RNA**→ COG4977 (COG4977)→
Osp-1-58	NZ_AAQH01000010.1	+	71963	71997	(shown above)
Osp-1-59	NZ_AAQH01000010.1	+	69078	69112	TPMT (pfam05724)→ **RNA**→ ←CaiD (COG1024)
Osp-1-61	NZ_AAQH01000002.1	-	49831	49797	COG0436 (COG0436)→ **RNA**→ ←Wzy_C (pfam04932)
Osp-1-62	NZ_AAQH01000002.1	+	61463	61497	←putative nucleoprotein/polynucleotide-associated enzyme **RNA**→ **RNA**→ ←COG1161 (COG1161)
Osp-1-68	NZ_AAQH01000002.1	+	61376	61410	(shown above)
Osp-1-63	NZ_AAQH01000012.1	+	39688	39722	←SrmB (COG0513) **RNA**→ **RNA**→ MhpC (COG0596)→
Osp-1-65	NZ_AAQH01000012.1	+	39959	39993	(shown above)
Osp-1-64	NZ_AAQH01000004.1	+	4680	4714	5S ribosomal RNA→ **RNA**→ ←hypo
Osp-1-66	NZ_AAQH01000012.1	-	82185	82151	SfcA (COG0281)→ **RNA**→ ←FusA (COG0480)
Psp-1-1	NC_007948.1	-	3322711	3322677	←COG1721 (COG1721) **RNA**→ ←COG0714 (COG0714)
Cph-1-26	NC_010831.1	-	1129954	1129920	←ChlI (COG1239) **RNA**→ ←Lpd (COG1249)
env-68	JCVI_SCAF_1101668032871	+	477	511	ATP-synt (pfam00231)? (GOS1385)→ **RNA**→ ←ManA (COG1482)? (GOS13198391)
env-69	sludgePhrap_Contig7695	-	173	139	←adh_short (pfam00106) **RNA**→
env-70	JCVI_SCAF_1101668557757	-	903	869	AroK (COG0703)? (GOS7695)→ **RNA**→ D-AAT_like (cd01558)? (GOS4366)→
env-71	wf1_AGZO6863_g2	-	833	798	**RNA**→
env-72	JCVI_SCAF_1101668570623	-	849	815	? (GOS1720)→ **RNA**→ SSP411 (cd02955)? (GOS9409)→
env-73	soil_XZS14765_b1	+	515	549	**RNA**→
Eli-1-1	NC_007722.1	+	1690838	1690872	←RpsG (COG0049) **RNA**→ ←Ribosomal_S12 (cd03368)
env-74	JCVI_SCAF_1101668580956	+	31	65	**RNA**→ ←NPD (pfam03060)? (GOS4645)
Eli-1-2	NC_007722.1	+	1557065	1557099	RpsA (COG0539)→ **RNA**→ GloB (COG0491)→
env-75	ABDH01049625.1	+	208	242	←hypo **RNA**→ ←DppB (COG0601)
env-76	ABDH01016308.1	+	529	563	hypo→ **RNA**→ hypo→
env-77	ABDH01052590.1	+	209	243	**RNA**→ ←TPR (cd00189)
env-78	JCVI_SCAF_1096628049322	-	940	906	Thr tRNA→ **RNA**→ Cfa (COG2230)? (GOS4122)→
env-79	ABDH01016308.1	-	600	566	←hypo **RNA**→ ←hypo
env-80	wf2_AHAA15655_x1	+	762	796	**RNA**→
env-81	ABEF01047781.1	-	883	849	COG1160 (COG1160)→ **RNA**→ ←DctP (COG1638)
env-82	ABEF01038460.1	+	1456	1490	Amidohydro_2 (pfam04909)→ **RNA**→ Amidohydro_2 (pfam04909)→
env-83	ABEF01051355.1	+	1415	1449	Transposase_11 (pfam01609)→ **RNA**→ ←hypo
Psp-1-2	NC_007948.1	+	222170	222204	COG4391 (COG4391)→ **RNA**→ RfaF (COG0859)→
env-84	JCVI_SCAF_1101668579874	-	828	794	**RNA**→ ←DHHA1 (pfam02272)? (GOS401)
env-85	JCVI_SCAF_1101668567100	+	1379	1413	←DegQ (COG0265)? (GOS899) **RNA**→
env-86	wf2_AHAA11964_y1	-	335	301	EXOIII (smart00479)→ **RNA**→ ←hypo

env-87	JCVI_SCAF_1101668133812	-	1510	1476	**RNA**→ LemA (pfam04011)? (GOS8926)→
Fba-2-1	NZ_AAXX01000003.1	-	121096	121062	DUF890 (pfam05971)→ **RNA**→ ←Asparaginase_2 (pfam01112)
env-88	JCVI_SCAF_1096627451797	+	32	66	**RNA**→
Ssp-2-1	NC_009663.1	+	296326	296360	Crp (COG0664)→ **RNA**→ ←Lrp (COG1522)
env-89	JCVI_SCAF_1096627812332	+	269	303	? (GOS7235)→ **RNA**→
Fno-2-2	NZ_ABAH01000013.1	-	57292	57258	**RNA**→
Fno-1-1	NZ_ABXZ01000002.1	+	236551	236585	RecD (COG0507)→ **RNA**→ RecD (COG0507)→
Fno-3-1	NZ_AAYF01000010.1	-	54073	54039	**RNA**→
env-90	soil_Contig5138	-	347	313	←NifU_N (pfam01592) **RNA**→ ←Cysteine sulfinate desulfinase/cysteine desulfurase and related enzymes(EC:2.8.1.7)
Psp-1-3	NC_007948.1	+	3322638	3322672	COG0714 (COG0714)→ **RNA**→ COG1721 (COG1721)→
env-91	JCVI_SCAF_1101668155579	+	64	98	**RNA**→ OYE_YqiM_FMN (cd02932)? (GOS207)→
Oal-1-1	NZ_AAMQ01000006.1	-	40179	40145	COG2866 (COG2866)→ **RNA**→ ←hypo
Oal-1-2	NZ_AAMQ01000002.1	+	503835	503869	←hypo **RNA**→ ←hypo
Oal-1-3	NZ_AAMQ01000002.1	+	504245	504279	←hypo **RNA**→ hypo→
Oal-1-4	NZ_AAMQ01000001.1	-	797839	797805	←Aes (COG0657) **RNA**→ ←UvrD (COG0210)
Oal-1-5	NZ_AAMQ01000002.1	+	575832	575866	methyltransferase→ **RNA**→ ←MdlB (COG1132)
env-92	JCVI_SCAF_1101668633095	-	46	12	Rnd (COG0349)? (GOS1030)→ **RNA**→
env-94	JCVI_SCAF_1101668633095	-	329	295	(shown above)
env-93	JCVI_SCAF_1101668584978	-	192	158	←TldD (COG0312)? (GOS3716) **RNA**→ ←TldD (COG0312)? (GOS3716)
Atu-1-1	NC_003063.2	+	459957	459991	COG1123 (COG1123)→ **RNA**→ ←Lrp (COG1522)
Atu-1-2	NC_003063.2	+	1652643	1652677	PurR (COG1609)→ **RNA**→ hypo→
Atu-1-3	NC_003062.2	-	935215	935181	LYS9 (COG1748)→ **RNA**→ ←UPF0061 (pfam02696)
Atu-1-4	NC_003062.2	-	1163912	1163878	←RhtB (COG1280) **RNA**→ ←RibC (COG0307)
Psp-1-4	NC_007948.1	+	3580382	3580416	←GntR (COG1802) **RNA**→ ←HcaE (COG4638)
Psp-1-5	NC_007948.1	+	3577201	3577235	←LivH (COG0559) **RNA**→ ←LivK (COG0683)
env-95	wf2_AHAA7366_x1	-	512	478	**RNA**→ ←hypo
env-96	JCVI_SCAF_1101668329979	-	582	548	**RNA**→ InfA (COG0361)? (GOS1275)→
Osp-1-69	NZ_AAQH01000016.1	+	65469	65503	hypo→ **RNA**→ SLC13_permease (cd01115)TrkA_C (pfam02080)→
Osp-1-70	NZ_AAQH01000008.1	-	41083	41049	TufB (COG0050)→ **RNA**→ DUF6 (pfam00892)→
env-97	JCVI_SCAF_1101668572351	+	205	239	**RNA**→ ACAD_fadE5 (cd01153)? (GOS26)→
env-98	wf2_AHAA16729_y1	+	695	729	WcaG (COG0451)→ **RNA**→
Rsp-1-1	NZ_AAMV01000001.1	-	71792	71758	MeTr (cd00740)→ **RNA**→ hypo→
Psp-1-6	NC_007948.1	-	5038662	5038628	←hypo **RNA**→ ←DTW (pfam03942)
Psp-1-7	NC_007948.1	-	3580471	3580437	HcaE (COG4638)→ **RNA**→ GntR (COG1802)→
Psp-1-8	NC_007948.1	-	1431332	1431298	←LivH (COG0559) **RNA**→ ←LivK (COG0683)
Psp-1-9	NC_007948.1	-	2940061	2940027	Crp (COG0664)→ **RNA**→ ←BON (pfam04972)NOG14632 (NOG14632)
env-99	soil_Contig31790	-	852	818	hypo→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ Exoribonuclease R(EC:3.1.13.1)→
env-100	soil_Contig31790	-	781	747	(shown above)
env-101	soil_Contig31790	-	923	889	(shown above)
env-177	soil_Contig31790	-	689	655	(shown above)
Psp-1-10	NC_007948.1	-	1767879	1767845	PhyH (pfam05721)NOG08005 (NOG08005)→ **RNA**→ COG4091 (COG4091)→
Mma-1-1	NZ_AAAP01003746.1	+	1856	1890	←FBPase_glpX (cd01516) **RNA**→ ←ACH1 (COG0427)
Mma-1-2	NZ_AAAP01003851.1	-	8797	8763	COG5283: Phage-related tail protein→ **RNA**→ ←hypo
env-102	JCVI_SCAF_1101668615421	-	65	31	Tgt (COG0343)? (GOS4174)→ **RNA**→
Psp-1-11	NC_007948.1	-	222246	222212	←RfaF (COG0859) **RNA**→ ←COG4391 (COG4391)
env-103	soil_Contig10267	+	1604	1638	AdoHcyase (cd00401)→ **RNA**→
env-104	JCVI_SCAF_1096628080544	-	1646	1612	**RNA**→ ←PpiB (COG0652)? (GOS0)
env-105	gws2_d4_0169_47	+	31203	31237	FixB (COG2025)→ **RNA**→ 4Fe-4S ferredoxin, iron-sulfur binding:Protein o f unknown function DUF224→
env-106	soil_Contig12548	+	863	897	←hypo **RNA**→ PaaI (COG2050)→
env-107	soil_XZS67761_b1	-	448	414	←COG1569 (COG1569) **RNA**→ ←UPF0020 (pfam01170)
env-108	soil_XZS100995_g1	-	660	626	PDZ_serine_protease (cd00987)→ **RNA**→ ←hypo
Psp-1-12	NC_007948.1	-	578420	578386	XdhA (COG4630)→ **RNA**→ COG0714 (COG0714)→

Psp-1-13	NC_007948.1	+	1431383	1431417	LivK (COG0683)→ **RNA**→ LivH (COG0559)→
Psp-1-14	NC_007948.1	+	1767787	1767821	←COG4091 (COG4091) **RNA**→ ←PhyH (pfam05721)NOG08005 (NOG08005)
env-109	soil_XZS1879_b1	+	187	221	**RNA**→ ←hypo
env-110	soil_Contig35217	-	259	225	←ABC_UvrA (cd03238) **RNA**→ ←hypo
Psp-1-15	NC_007948.1	+	2881782	2881816	Vgb (COG4257)→ **RNA**→ ←hypo
Psp-1-16	NC_007948.1	-	2137193	2137159	TPR (cd00189)→ **RNA**→ GloB (COG0491)→
Psp-1-17	NC_007948.1	+	745712	745746	COG5605 (COG5605)→ **RNA**→ Methyltransf_6 (pfam03737)→
Psp-1-18	NC_007948.1	-	3921049	3921015	GlpC (COG0247)→ **RNA**→ hypo→
Psp-1-19	NC_007948.1	+	2662932	2662966	hypo→ **RNA**→ DUF1223 (pfam06764)→
Psp-1-20	NC_007948.1	-	4356788	4356754	DhlC (COG4147)→ **RNA**→ ←hypo
Psp-1-21	NC_007948.1	-	1409867	1409833	REC (cd00156)→ **RNA**→ ←major facilitator transporter
env-111	soil_Contig31790	+	714	748	←Exoribonuclease R(EC:3.1.13.1) **RNA**→ ←hypo
env-112	JCVI_SCAF_1101668582610	+	133	167	**RNA**→ ←PgsA (COG0558)? (GOS7357)
env-113	gws2_d1_0542_84	-	36163	36129	**RNA**→ COG1556 (COG1556)→
env-114	JCVI_SCAF_1101668559813	+	99	133	**RNA**→ ←NusG (COG0250)? (GOS1240)
env-115	wf2_AHAA16979_x1	+	466	500	**RNA**→ LexA_DNA_bind (pfam01726)→
env-116	soil_XZS2824_g1	+	642	676	←Pyruvate_Kinase (cd00288) **RNA**→
env-117	ABEF01048853.1	-	494	460	hypo→ **RNA**→ ←hypo
env-118	JCVI_SCAF_1101668595544	-	768	734	predicted protein→ **RNA**→
env-120	JCVI_SCAF_1101668051072	-	198	164	←UbiH (COG0654)? (GOS3198) **RNA**→
env-119	JCVI_SCAF_1101668051072	-	395	361	(shown above)
env-121	soil_XZS49895_b1	+	301	335	hypo→ **RNA**→ ←hypo
Ssp-1-1	NZ_AAQG01000019.1	-	66858	66824	hypo→ **RNA**→ ←tRNA_Me_trans (cd01998)
Ssp-1-2	NZ_AAQG01000003.1	+	235744	235778	←AdSS (cd03108) **RNA**→ hypo→
Ssp-1-3	NZ_AAQG01000002.1	-	66272	66238	hypo→ **RNA**→ ←hypo
Sal-1-2	NC_008048.1	+	1366747	1366781	←NuoF (COG1894) **RNA**→ ←NuoE (COG1905)Complex1_24kDa (pfam01257)
Sal-1-3	NC_008048.1	+	3118241	3118275	RlpA (COG0797)→ **RNA**→ ←hypo
env-122	JCVI_SCAF_1101668066154	-	1237	1203	**RNA**→ ←DLH (pfam01738)? (GOS6241)
env-123	JCVI_SCAF_1096627461402	+	623	657	←Ldh_2 (pfam02615)? (GOS5495) **RNA**→ ←? (GOS3923)
Sal-1-4	NC_008048.1	-	1926702	1926668	←SrmB (COG0513) **RNA**→ ←COG2081 (COG2081)
Eli-1-3	NC_007722.1	-	1710434	1710400	hypo→ **RNA**→ hypo→
Eli-1-4	NC_007722.1	-	1947408	1947374	PKCI_related (cd01276)→ **RNA**→ FcbC (COG0824)→
Eli-1-5	NC_007722.1	-	2149312	2149278	ArgE (COG0624)→ **RNA**→ GltP (COG1301)→
env-124	soil_XZS84691_b1	-	110	76	hypo→ **RNA**→
env-125	wf3_AHAI4765_x1	+	593	627	COG2827 (COG2827)→ **RNA**→
Eli-1-6	NC_007722.1	+	1691421	1691455	←Ribosomal_S12 (cd03368) **RNA**→ hypo→
env-126	JCVI_SCAF_1101668067676	+	373	407	**RNA**→ ←NHase_beta (pfam02211)? (GOS9147)
Ssp-1-4	NZ_AAQG01000002.1	-	136926	136892	←hypo **RNA**→ PotE (COG0531)→
env-127	JCVI_SCAF_1096628067911	-	1053	1019	? (GOS530)→ **RNA**→ ←DUF404 (pfam04169)? (GOS2566)
env-128	JCVI_SCAF_1096628066931	-	1216	1182	tRNA_Me_trans (cd01998)? (GOS4865)→ **RNA**→ ThiL (COG0611)? (GOS4853)→
env-129	gws2_d4_0142_71	-	4797	4763	COG2827 (COG2827)→ **RNA**→ KAS_III (cd00830)→
Atu-1-5	NC_003063.2	+	1583228	1583262	DdpA (COG0747)→ **RNA**→ DppB (COG0601)→
Atu-1-6	NC_003063.2	-	1620249	1620215	TerC (COG0861)→ **RNA**→ CitT (COG0471)→
env-130	soil_Contig3739	+	46	80	**RNA**→ ←hypo
env-131	gws2_d1_0068_1	+	30068	30102	adh_short (pfam00106)→ **RNA**→ ←hypo
env-132	gws2_d1_0011_1	+	119604	119638	COG1451 (COG1451)→ **RNA**→ ←CsdB (COG0520)
env-133	gws2_d1_0008_4	+	4709	4743	←hypo **RNA**→ ←hypo
env-134	gws2_d4_0371	+	5558	5592	←SH3b (smart00287) **RNA**→ ←SdhA (COG1053)
env-135	gws2_d4_0178_49	+	15248	15282	←DUF1269 (pfam06897) **RNA**→ ←PLDc (cd00138)
env-136	ABEF01034290.1	-	1640	1606	**RNA**→ ←hypo
Sal-1-5	NC_008048.1	-	555182	555148	←PseudoU_synth_EcTruB (cd02573) **RNA**→ ←PseudoU_synth_EcTruB (cd02573)
Sal-1-6	NC_008048.1	-	2956151	2956117	←Pta (COG0280)SfcA (COG0281) **RNA**→ MutS (COG0249)→

env-137	JCVI_SCAF_1101668565414	+	503	537	←STE14 (COG2020)? (GOS7589) **RNA**→ ←predicted protein
env-138	sludgeOz_Contig8012	-	277	243	←(,,,) **RNA**→
env-139	JCVI_SCAF_1101668579514	+	758	792	←TroR (COG1321)? (GOS11127) **RNA**→ ? (GOS997)→
Dol-1-1	NC_009943.1	-	2493662	2493628	hypo→ **RNA**→ RhsA (COG3209)→
Dol-1-2	NC_009943.1	-	2489677	2489643	hypo→ **RNA**→ RhsA (COG3209)→
env-140	soil_XZS67761_b1	+	369	403	UPF0020 (pfam01170)→ **RNA**→ COG1569 (COG1569)→
env-141	JCVI_SCAF_1101668142573	-	394	360	←? (GOS241) **RNA**→ ←? (GOS1293)
env-142	soil_XZS3001_g1	+	175	209	**RNA**→ ←Aminopeptidase N
env-143	soil_XZS44342_g1	+	732	766	←Mug (COG3663) **RNA**→
env-144	soil_XZS1879_b1	-	163	129	hypo→ **RNA**→
env-145	gws2_d4_0016_47	-	136075	136041	PAS (smart00091)→ **RNA**→ ←Pol_IV_kappa (cd03586)
Cph-1-27	NC_010831.1	+	1330271	1330305	←hypo **RNA**→ PlsC (COG0204)→
env-147	JCVI_SCAF_1101668099018	-	1025	991	? (GOS2296)→ **RNA**→ Transposase_11 (pfam01609)? (GOS4735)→
env-148	JCVI_SCAF_1096627348822	+	994	1028	PheT (COG0072)? (GOS2624)→ **RNA**→
env-149	ABEF01042264.1	-	725	691	NusG (COG0250)→ **RNA**→ hypo→
env-150	ABEF01052502.1	+	548	582	NusG (COG0250)→ **RNA**→ Ribosomal_S23p (pfam05635)→
Nha-1-1	NC_007964.1	+	34941	34977	InfB (COG0532)IF2_N (pfam04760)→ **RNA**→ RbfA (COG0858)→
env-151	wf3_AHAI13919_y1	-	391	355	←ThiE (COG0352) **RNA**→
Cph-1-28	NC_010831.1	+	2173847	2173881	←FtsK (COG1674) **RNA**→ ←GcvT (COG0404)
Cph-1-29	NC_010831.1	-	1892459	1892425	hypo→ **RNA**→ ←FhlA (COG3604)
Cph-1-30	NC_010831.1	-	2164092	2164058	←OraA (COG2137) **RNA**→ COG3347 (COG3347)adh_short (pfam00106)→
Cph-1-31	NC_010831.1	+	1561029	1561063	←COG0670 (COG0670) **RNA**→ ←HslU (COG1220)
env-152	JCVI_SCAF_1101668598461	-	1154	1120	**RNA**→ ←COG2958 (COG2958)? (GOS13103195)
env-153	gws2_d4_0513	-	4294	4260	←hypo **RNA**→ ←hypo
env-154	wf2_AHAA15655_x1	-	520	485	**RNA**→
env-155	JCVI_SCAF_1101668127707	+	1018	1052	? (GOS13034328)→ **RNA**→ ? (GOS2593)→
Swi-1-1	NC_009511.1	+	1682165	1682199	COG2938 (COG2938)→ **RNA**→ Mfd (COG1197)→
Bsp-2-1	NC_009445.1	-	961411	961377	DdpA (COG0747)→ **RNA**→ ←CyaA (COG2114)CHASE2 (pfam05226)
env-156	wf2_AHAA8239_y1	-	237	203	DapB (COG0289)→ **RNA**→
Bsp-2-2	NC_009445.1	-	6143066	6143032	TPR (cd00189)COG4249 (COG4249)→ **RNA**→ ←hypo
Bsp-2-3	NC_009445.1	-	565240	565206	NhaP (COG0025)→ **RNA**→ Plasmid_stabil (pfam05016)→
Bsp-2-4	NC_009445.1	-	7247950	7247916	COG3185 (COG3185)→ **RNA**→ ←CaiB (COG1804)
Bsp-2-5	NC_009445.1	-	6840351	6840317	OmpW (COG3047)→ **RNA**→ ←hypo
Bsp-2-6	NC_009445.1	-	768074	768040	MA (smart00283)→ **RNA**→ ←RHOD_PspE2 (cd01521)
Bsp-2-7	NC_009445.1	+	7134019	7134053	MAPEG (pfam01124)→ **RNA**→ ←LivK (COG0683)
Bsp-2-8	NC_009445.1	+	2311337	2311371	hypo→ **RNA**→ ←ACAD_fadE5 (cd01153)
Bsp-2-9	NC_009445.1	+	898696	898730	Intron_gpII→ **RNA**→ ←COG1032 (COG1032)
Bsp-2-10	NC_009445.1	+	518131	518165	Aamy (smart00642)→ **RNA**→ ←Aamy (smart00642)
Bsp-2-11	NC_009445.1	-	518258	518224	Aamy (smart00642)→ **RNA**→ ←Aamy (smart00642)
Bsp-2-12	NC_009445.1	-	822240	822206	COG1011 (COG1011)→ **RNA**→ DapD (COG2171)→
Bsp-2-13	NC_009445.1	-	4509731	4509697	hypo→ **RNA**→ SCP-x_thiolase (cd00829)→
Bsp-2-14	NC_009445.1	-	2560181	2560147	←cysteine_hydrolases (cd00431) **RNA**→ ←DegQ (COG0265)
Bsp-2-15	NC_009445.1	+	6501288	6501322	LysR (COG0583)→ **RNA**→ ←COG2215 (COG2215)
Bsp-2-16	NC_009445.1	+	767945	767979	RHOD_PspE2 (cd01521)→ **RNA**→ ←MA (smart00283)
Bsp-2-17	NC_009445.1	+	7283189	7283223	PtrB (COG1770)→ **RNA**→ ←Methyltransf_6 (pfam03737)
Bsp-2-18	NC_009445.1	+	1853006	1853040	←hypo **RNA**→ ←DUF989 (pfam06181)
Bsp-2-19	NC_009445.1	+	782094	782128	PEPCK_ATP (cd00484)PckA (COG1866)→ **RNA**→ ←putative thioesterase/thiol ester dehydrase-isomerase
Bsp-2-20	NC_009445.1	+	735886	735920	←TlyC (COG1253) **RNA**→ ←COG3637 (COG3637)OprB (COG3659)
Bsp-2-21	NC_009445.1	+	696813	696847	TPR (cd00189)TadD (COG5010)→ **RNA**→ ←COG2141 (COG2141)
Bsp-2-22	NC_009445.1	-	1205770	1205736	←Beta-lactamase (pfam00144) **RNA**→ DUF1486 (pfam07366)→
Bsp-2-23	NC_009445.1	-	5956817	5956783	←Imp (COG1452) **RNA**→ ←REC (cd00156)BaeS (COG0642)
Bsp-1-1	NC_009485.1	-	7467253	7467219	←CaiC (COG0318) **RNA**→ ←thiolase (cd00751)

Bsp-2-24	NC_009445.1	-	696937	696903	COG2141 (COG2141)→ **RNA**→ ←TPR (cd00189)TadD (COG5010)
Bsp-2-25	NC_009445.1	+	1726928	1726962	hypo→ **RNA**→ ←CrtC (pfam07143)
Bsp-2-26	NC_009445.1	-	604262	604228	Rotamase (pfam00639)→ **RNA**→ ArgJ (pfam01960)→
Bsp-2-27	NC_009445.1	+	3114292	3114326	Qor (COG0604)→ **RNA**→ ←CaiD (COG1024)
Bsp-2-28	NC_009445.1	-	7134147	7134113	LivK (COG0683)→ **RNA**→ ←MAPEG (pfam01124)
Bsp-2-29	NC_009445.1	+	6259364	6259398	COG5465 (COG5465)→ **RNA**→ ProC (COG0345)→
Bsp-2-30	NC_009445.1	-	2534511	2534477	hypo→ **RNA**→ ←LigB (pfam02900)
Bsp-2-31	NC_009445.1	-	386695	386661	COG2968 (COG2968)→ **RNA**→ ←COG2220 (COG2220)
Bsp-2-32	NC_009445.1	+	260585	260619	OmpA (COG2885)→ **RNA**→ **RNA**→ **RNA**→ ←ArgG (COG0137)
Bsp-2-36	NC_009445.1	+	260428	260462	(shown above)
Bsp-2-58	NC_009445.1	+	260271	260305	(shown above)
Bsp-2-33	NC_009445.1	+	6299127	6299161	PqqL (COG0612)→ **RNA**→ PrfB (COG1186)→
Bsp-2-34	NC_009445.1	+	6498655	6498689	SQCY_1 (cd02892)→ **RNA**→ ←DltE (COG0300)
Bsp-2-35	NC_009445.1	-	665886	665852	COG4784 (COG4784)→ **RNA**→ ←CarD_TRCF (pfam02559)
Bsp-2-37	NC_009445.1	-	1991515	1991481	←hypo **RNA**→ ←Intron_gpII
Bsp-2-38	NC_009445.1	-	5790814	5790780	COG3384 (COG3384)→ **RNA**→ UhpC (COG2271)→
Bsp-2-39	NC_009445.1	+	386568	386602	COG2220 (COG2220)→ **RNA**→ ←COG2968 (COG2968)
Bsp-2-40	NC_009445.1	-	7283316	7283282	Methyltransf_6 (pfam03737)→ **RNA**→ ←PtrB (COG1770)
Bsp-2-41	NC_009445.1	-	3114415	3114381	CaiD (COG1024)→ **RNA**→ ←Qor (COG0604)
Bsp-2-42	NC_009445.1	-	93392	93358	←hypo **RNA**→ ←COG0802 (COG0802)COG3178 (COG3178)
Bsp-2-43	NC_009445.1	-	339049	339015	ApbA (COG1893)→ **RNA**→ ←COG2863 (COG2863)
Bsp-2-44	NC_009445.1	-	6299255	6299221	←PrfB (COG1186) **RNA**→ ←PqqL (COG0612)
Bsp-1-2	NC_009485.1	+	2403993	2404027	PaaI_thioesterase (cd03443)→ **RNA**→ ←hypo
Bsp-2-45	NC_009445.1	+	5792442	5792476	←COG3384 (COG3384) **RNA**→ ←COG2259 (COG2259)
Bsp-2-46	NC_009445.1	+	1790717	1790751	hypo→ **RNA**→ PstS (COG0226)→
Bsp-2-47	NC_009445.1	-	1911512	1911478	COG4126 (COG4126)→ **RNA**→ ←COG0388 (COG0388)
Bsp-1-3	NC_009485.1	-	1439127	1439093	COG3449 (COG3449)HTH_ARAC (smart00342)→ **RNA**→ ←COG2215 (COG2215)
Bsp-2-48	NC_009445.1	-	6996174	6996140	LivK (COG0683)→ **RNA**→ Lhr (COG1201)→
Bsp-2-49	NC_009445.1	+	1991388	1991422	Intron_gpII→ **RNA**→ hypo→
Bsp-2-50	NC_009445.1	+	364018	364052	bact_SO_family_Moco (cd02108)→ **RNA**→ ←DUF1228 (pfam06779)
Bsp-2-51	NC_009445.1	-	6259495	6259461	←ProC (COG0345) **RNA**→ ←COG5465 (COG5465)
Bsp-2-52	NC_009445.1	-	1790844	1790810	←PstS (COG0226) **RNA**→ ←hypo
Bsp-2-53	NC_009445.1	-	990972	990938	CaiC (COG0318)→ **RNA**→ COG2945 (COG2945)→
Bsp-2-54	NC_009445.1	+	1891657	1891691	←ModA (COG0725) **RNA**→ ←Amidase (pfam01425)
Bsp-2-55	NC_009445.1	-	992693	992659	thiolase (cd00751)→ **RNA**→ CaiC (COG0318)→
Bsp-2-56	NC_009445.1	+	5956689	5956723	REC (cd00156)BaeS (COG0642)→ **RNA**→ Imp (COG1452)→
Bsp-2-57	NC_009445.1	+	1191774	1191808	hypo→ **RNA**→ ←IMPase (cd01639)
Bsp-2-59	NC_009445.1	-	984807	984773	Amidase (pfam01425)→ **RNA**→ ←AdhC (COG1062)
Bsp-2-60	NC_009445.1	-	6498932	6498898	DltE (COG0300)→ **RNA**→ ←SQCY_1 (cd02892)
Bsp-2-61	NC_009445.1	-	6501416	6501382	COG2215 (COG2215)→ **RNA**→ ←LysR (COG0583)
Bsp-2-62	NC_009445.1	+	4509604	4509638	←SCP-x_thiolase (cd00829) **RNA**→ ←hypo
Bsp-2-63	NC_009445.1	+	2560054	2560088	DegQ (COG0265)→ **RNA**→ cysteine_hydrolases (cd00431)→
env-157	JCVI_SCAF_1101668155579	-	139	105	←OYE_YqiM_FMN (cd02932)? (GOS207) **RNA**→
env-158	gws2_d4_0265_9	-	3476	3442	HD (pfam01966)→ **RNA**→ ←COG1204 (COG1204)
env-159	gws2_d4_0144_58	+	3655	3689	AlaS (COG0013)→ **RNA**→ Histidinol-phosphate/aromatic aminotransferase(EC:2.6.1.9)→
env-160	JCVI_SCAF_1096626899633	-	1004	970	←NTP_transf_2 (pfam01909)? (GOS11233935) **RNA**→ **RNA**→ ←? (GOS987)
env-161	JCVI_SCAF_1096626899633	-	966	932	(shown above)
env-162	wf2_AHAA16729_y1	-	783	749	**RNA**→ ←WcaG (COG0451)
Psp-1-22	NC_007948.1	-	605789	605755	LivK (COG0683)NOG05970 (NOG05970)→ **RNA**→ LivF (COG0410)→
Lag-1-1	NZ_AAUW01000017.1	-	159884	159850	aconitate hydratase→ **RNA**→ **RNA**→ ←COG0780 (COG0780)
Lag-1-2	NZ_AAUW01000017.1	-	160160	160126	(shown above)
Cph-1-32	NC_010831.1	-	2291094	2291060	RfaG (COG0438)→ **RNA**→ **RNA**→ ←hypo

Cph-1-103	NC_010831.1	-	2290875	2290841	(shown above)
env-163	soil_Contig31279	-	455	421	←HINT_subgroup (cd01277) **RNA**→ ←Uncharacterized protein conserved in bacteria
env-164	JCVI_SCAF_1096627479218	-	1148	1114	? (GOS5834)→ **RNA**→ ←Glyoxalase (pfam00903)
env-165	JCVI_SCAF_1101668576954	-	624	590	←RfaE_N (cd02172)? (GOS271) **RNA**→
Cph-1-33	NC_010831.1	-	1200410	1200376	Cfa (COG2230)→ **RNA**→ ←DUF1343 (pfam07075)
env-166	sludgeOz_Contig10742	+	4347	4381	←3Beta_HSD (pfam01073) **RNA**→
env-167	sludgeOz_Contig10742	+	3966	4000	(shown above)
env-169	sludgeOz_Contig10742	+	4220	4254	(shown above)
env-168	sludgeOz_Contig10909	+	3921	3955	SrmB (COG0513)DbpA (pfam03880)→ **RNA**→ ←(,,,)
env-170	sludgeOz_Contig10592	-	1328	1294	(,,,)→ **RNA**→ (,,,)→
env-171	sludgeOz_Contig5748	-	476	442	←SufC (COG0396) **RNA**→ ←ABC-type transport system involved in Fe-S cluste r assembly, permease component(,,,)
Mma-1-3	NZ_AAAP01003863.1	+	42607	42641	hypo→ **RNA**→ ←hypo
Mma-1-4	NZ_AAAP01003877.1	-	50801	50767	hypo→ **RNA**→ ←hypo
Mma-1-5	NZ_AAAP01003622.1	+	2648	2682	←COG0535 (COG0535) **RNA**→ ←trans_reg_C (cd00383)
Mma-1-6	NZ_AAAP01003819.1	+	5196	5230	COG1607 (COG1607)→ **RNA**→ ←Ssb (COG0629)
Mma-1-7	NZ_AAAP01003835.1	+	15085	15119	hypo→ **RNA**→ ←CaiD (COG1024)
Mma-1-8	NZ_AAAP01003834.1	-	1400	1366	MarR (pfam01047)→ **RNA**→ HELICc (cd00079)DEXDc (smart00487)→
Mma-1-9	NZ_AAAP01003813.1	+	12679	12713	←RecO (COG1381) **RNA**→ ←Era (COG1159)
Mma-1-10	NZ_AAAP01003836.1	-	18120	18086	SrmB (COG0513)DbpA (pfam03880)→ **RNA**→ Ffh (COG0541)→
Mma-1-11	NZ_AAAP01003811.1	-	10391	10357	COG1051 (COG1051)→ **RNA**→ ←COG0457: FOG: TPR repeat
Mma-1-12	NZ_AAAP01003742.1	+	1443	1477	←PaaK (COG1541) **RNA**→ ←COG1042 (COG1042)
Mma-1-13	NZ_AAAP01003796.1	+	11087	11121	HTH_XRE (cd00093)COG3800 (COG3800)→ **RNA**→ ←hypo
Mma-1-14	NZ_AAAP01003864.1	-	16684	16650	COG2867 (COG2867)→ **RNA**→ ←RpsA (COG0539)
Mma-1-15	NZ_AAAP01003879.1	+	5146	5180	←LeuA (COG0119) **RNA**→ ←COG4770 (COG4770)
Mma-1-16	NZ_AAAP01003622.1	-	731	697	←Pfs (COG0775) **RNA**→ ←ISOPREN_C2_like (cd00688)
Mma-1-17	NZ_AAAP01003862.1	+	33803	33837	CysG (COG0007)→ **RNA**→ hypo→
Mma-1-18	NZ_AAAP01003877.1	+	50706	50740	hypo→ **RNA**→ ←hypo
Mma-1-19	NZ_AAAP01003803.1	+	16319	16353	←ACAD_fadE5 (cd01153) **RNA**→ ←FAA1 (COG1022)
Mma-1-20	NZ_AAAP01003851.1	+	8702	8736	hypo→ **RNA**→ ←COG5283: Phage-related tail protein
Mma-1-21	NZ_AAAP01003744.1	-	1615	1581	←hypo **RNA**→ ←hypo
Mma-1-22	NZ_AAAP01003746.1	-	1951	1917	ACH1 (COG0427)→ **RNA**→ FBPase_glpX (cd01516)→
Mma-1-23	NZ_AAAP01003848.1	+	6630	6664	COG0494: NTP pyrophosphohydrolases including oxidative damage repair enzymes→ **RNA**→ IbpA (COG0071)→
env-172	JCVI_SCAF_1101668115658	-	972	938	Acetyltransf_1 (pfam00583)? (GOS8607)→ **RNA**→ RhtB (COG1280)? (GOS192)→
Sal-1-7	NC_008048.1	-	738789	738755	COG2827 (COG2827)→ **RNA**→ hypo→
Sal-1-8	NC_008048.1	+	555106	555140	PseudoU_synth_EcTruB (cd02573)→ **RNA**→ PseudoU_synth_EcTruB (cd02573)→
Sal-1-9	NC_008048.1	-	1366823	1366789	NuoE (COG1905)Complex1_24kDa (pfam01257)→ **RNA**→ NuoF (COG1894)→
Psp-1-23	NC_007948.1	+	3920977	3921011	←hypo **RNA**→ ←GlpC (COG0247)
env-173	soil_Contig35217	+	180	214	hypo→ **RNA**→ ABC_UvrA (cd03238)→
Cph-1-34	NC_010831.1	+	1880495	1880529	HisF (COG0107)→ **RNA**→ ←hypo
Cph-1-35	NC_010831.1	+	2031997	2032031	Beta-lactamase (pfam00144)→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**
Cph-1-36	NC_010831.1	+	2032037	2032071	(shown above)
Cph-1-37	NC_010831.1	+	2032157	2032191	(shown above)
Cph-1-38	NC_010831.1	+	2032077	2032111	(shown above)
Cph-1-39	NC_010831.1	+	2032237	2032271	(shown above)
Cph-1-40	NC_010831.1	+	2032317	2032351	(shown above)
Cph-1-41	NC_010831.1	+	2032477	2032511	(shown above)
Cph-1-42	NC_010831.1	+	2032117	2032151	(shown above)
Cph-1-43	NC_010831.1	+	2032397	2032431	(shown above)
Cph-1-44	NC_010831.1	+	2032557	2032591	(shown above)
Cph-1-45	NC_010831.1	+	2032277	2032311	(shown above)
Cph-1-46	NC_010831.1	+	2032517	2032551	(shown above)
Cph-1-47	NC_010831.1	+	2032597	2032631	(shown above)

Cph-1-48	NC_010831.1	+	2032197	2032231	(shown above)
Cph-1-49	NC_010831.1	+	2032357	2032391	(shown above)
Cph-1-50	NC_010831.1	+	2032437	2032471	(shown above)
Cph-1-53	NC_010831.1	+	2032637	2032671	(shown above)
Cph-1-51	NC_010831.1	-	2132357	2132323	COG2827 (COG2827)→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ hypo→
Cph-1-52	NC_010831.1	-	2132055	2132021	(shown above)
Cph-1-123	NC_010831.1	-	2132269	2132235	(shown above)
Cph-1-124	NC_010831.1	-	2131967	2131933	(shown above)
Cph-1-133	NC_010831.1	-	2132181	2132147	(shown above)
Cph-1-55	NC_010831.1	-	2151282	2151248	RpsA (COG0539)→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ ←Era (COG1159)
Cph-1-56	NC_010831.1	-	2151522	2151488	(shown above)
Cph-1-67	NC_010831.1	-	2151243	2151209	(shown above)
Cph-1-86	NC_010831.1	-	2151483	2151449	(shown above)
Cph-1-57	NC_010831.1	+	2120104	2120138	←Rfe (COG0472) **RNA**→ ←putative glycosyl transferase, family 4
Cph-1-58	NC_010831.1	+	2062291	2062325	←CysQ (cd01638) **RNA**→ **RNA**→ **RNA**→ ←GlmS (COG0449)
Cph-1-83	NC_010831.1	+	2062694	2062728	(shown above)
Cph-1-85	NC_010831.1	+	2062527	2062561	(shown above)
Cph-1-59	NC_010831.1	+	2054880	2054914	←hypo **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ ←RfbD (COG1091)
Cph-1-107	NC_010831.1	+	2055048	2055082	(shown above)
Cph-1-109	NC_010831.1	+	2055217	2055251	(shown above)
Cph-1-110	NC_010831.1	+	2054374	2054408	(shown above)
Cph-1-111	NC_010831.1	+	2054712	2054746	(shown above)
Cph-1-112	NC_010831.1	+	2054543	2054577	(shown above)
Cph-1-60	NC_010831.1	+	1275702	1275736	hypo→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ ←Cfa (COG2230)
Cph-1-61	NC_010831.1	+	1275867	1275901	(shown above)
Cph-1-81	NC_010831.1	+	1275537	1275571	(shown above)
Cph-1-82	NC_010831.1	+	1275372	1275406	(shown above)
Cph-1-62	NC_010831.1	+	897795	897829	←hypo **RNA**→ hypo→
Cph-1-63	NC_010831.1	+	1184603	1184637	←hypo **RNA**→ **RNA**→ ←hypo
Cph-1-64	NC_010831.1	+	1184732	1184766	(shown above)
Cph-1-65	NC_010831.1	-	2372439	2372405	ArsA (COG0003)→ **RNA**→ ←FolK (COG0801)
Cph-1-66	NC_010831.1	+	1126968	1127002	←WrbA (COG0655) **RNA**→ ←RluA (COG0564)
Cph-1-69	NC_010831.1	-	1368905	1368871	←GlpC (COG0247)GlcD (COG0277)FAD-oxidase_C (pfam02913) **RNA**→ hypo→
Cph-1-70	NC_010831.1	+	388590	388624	←hypo **RNA**→ ←RfaE_N (cd02172)
Cph-1-71	NC_010831.1	+	1715478	1715512	←HyaD (COG0680) **RNA**→ ←HyaC (COG1969)
Cph-1-72	NC_010831.1	-	1226928	1226894	←hypo **RNA**→ ←hypo
Cph-1-73	NC_010831.1	+	2157826	2157860	hypo→ **RNA**→ hypo→
Cph-1-74	NC_010831.1	-	1459078	1459044	←Blc (COG3040) **RNA**→ Acetyltransf_1 (pfam00583)→
Cph-1-75	NC_010831.1	-	449766	449732	←tRNA-Lys **RNA**→ hypo→
Cph-1-76	NC_010831.1	+	323832	323866	RpoC (COG0086)→ **RNA**→ ←AccC (COG0439)
Cph-1-77	NC_010831.1	+	2441001	2441035	←hypo **RNA**→ ←hypo
Cph-1-78	NC_010831.1	+	2101253	2101287	←hypo **RNA**→ ←hypo
Cph-1-79	NC_010831.1	+	2108062	2108096	←hypo **RNA**→ ←hypo
Cph-1-80	NC_010831.1	-	2490793	2490759	vWA_BatA_type (cd01467)→ **RNA**→ vWA_BatA_type (cd01467)→
Cph-1-84	NC_010831.1	+	2060902	2060936	←hypo **RNA**→ hypo→
Cph-1-87	NC_010831.1	-	2132854	2132820	hypo→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ COG2827 (COG2827)→
Cph-1-54	NC_010831.1	-	2133243	2133209	(shown above)
Cph-1-68	NC_010831.1	-	2133204	2133170	(shown above)
Cph-1-96	NC_010831.1	-	2133092	2133058	(shown above)
Cph-1-88	NC_010831.1	+	939948	939982	COG2361 (COG2361)→ **RNA**→ hypo→
Cph-1-89	NC_010831.1	-	453958	453924	←ETF_beta (cd01714) **RNA**→ FcbT1 (COG4663)→
Cph-1-90	NC_010831.1	-	1102024	1101990	COG1032 (COG1032)→ **RNA**→ UbiE (COG2226)→

Cph-1-91	NC_010831.1	+	487804	487838	Lig (COG0272)→ **RNA**→ **RNA**→ COG1611 (COG1611)→
Cph-1-92	NC_010831.1	+	487895	487929	(shown above)
Cph-1-93	NC_010831.1	+	874037	874071	←KdtA (COG1519) **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ hypo→
Cph-1-94	NC_010831.1	+	874129	874163	(shown above)
Cph-1-95	NC_010831.1	+	873671	873705	(shown above)
Cph-1-97	NC_010831.1	+	873489	873523	(shown above)
Cph-1-98	NC_010831.1	+	874221	874255	(shown above)
Cph-1-99	NC_010831.1	+	873855	873889	(shown above)
Cph-1-100	NC_010831.1	+	873946	873980	(shown above)
Cph-1-101	NC_010831.1	+	873763	873797	(shown above)
Cph-1-102	NC_010831.1	+	873580	873614	(shown above)
Sal-1-10	NC_008048.1	+	1926618	1926652	COG2081 (COG2081)→ **RNA**→ SrmB (COG0513)→
Cph-1-104	NC_010831.1	-	576655	576621	IMPase (cd01639)→ **RNA**→ **RNA**→ **RNA**→ LpxC (pfam03331)FabZ (cd01288)→
Cph-1-105	NC_010831.1	-	576384	576350	(shown above)
Cph-1-106	NC_010831.1	-	576926	576892	(shown above)
Cph-1-108	NC_010831.1	+	2077497	2077531	←hypo **RNA**→ ←dTDP_sugar_isom (pfam00908)
Cph-1-113	NC_010831.1	-	2036866	2036832	AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→ **RNA**→ ←COG1600 (COG1600)
Cph-1-114	NC_010831.1	+	2070731	2070765	←hypo **RNA**→ ←VapC (COG1487)
Cph-1-116	NC_010831.1	+	2127113	2127147	←hypo **RNA**→ **RNA**→ ←hypo
Cph-1-115	NC_010831.1	+	2126932	2126966	(shown above)
Cph-1-117	NC_010831.1	+	2126750	2126784	hypo→ **RNA**→ ←hypo
Cph-1-118	NC_010831.1	-	1524190	1524156	←PepN (COG0308) **RNA**→ **RNA**→ **RNA**→ ←COG0702 (COG0702)
Cph-1-121	NC_010831.1	-	1524444	1524410	(shown above)
Cph-1-122	NC_010831.1	-	1524317	1524283	(shown above)
Cph-1-119	NC_010831.1	+	1714624	1714658	PerM (COG0628)→ **RNA**→ **RNA**→ ←HyaD (COG0680)
Cph-1-120	NC_010831.1	+	1714532	1714566	(shown above)
Cph-1-125	NC_010831.1	+	1908014	1908048	←HypC (COG0298) **RNA**→ ←HypF (COG0068)
Cph-1-126	NC_010831.1	+	2618536	2618570	←DUF452 (pfam04301) **RNA**→ ←BioF (COG0156)
Cph-1-127	NC_010831.1	+	1609652	1609686	←COG1342 (COG1342) **RNA**→ ←C4dic_mal_tran (pfam03595)
Cph-1-128	NC_010831.1	-	976667	976633	←Peptidase_S26 (pfam00461) **RNA**→ ←hypo
Cph-1-129	NC_010831.1	+	468137	468171	RfbX (COG2244)→ **RNA**→ ←hypo
Cph-1-130	NC_010831.1	+	1737348	1737382	FHIT (cd01275)→ **RNA**→ UbiG (COG2227)→
Cph-1-131	NC_010831.1	+	558433	558467	YcaC_related (cd01012)→ **RNA**→ ←LpxK (COG1663)
Cph-1-132	NC_010831.1	+	1614494	1614528	←DBI_PRT (pfam02277) **RNA**→ ←CobU (cd00544)
Cph-1-134	NC_010831.1	+	2137207	2137241	←tRNA-Phe **RNA**→ ←GluRS_core (cd00808)
env-174	JCVI_SCAF_1096627960249	-	102	68	CrcB (COG0239)? (GOS11882)→ **RNA**→
Sal-1-11	NC_008048.1	+	2199820	2199854	←hypo **RNA**→ CopG/Arc/MetJ family transcriptional regulator→
env-175	JCVI_SCAF_1096627461402	-	713	679	? (GOS3923)→ **RNA**→ Ldh_2 (pfam02615)? (GOS5495)→
env-176	wf3_AHAI18676_g2	-	837	803	**RNA**→ RbfA (COG0858)→
env-178	ABEF01048164.1	-	451	417	←Arg tRNA **RNA**→ hypo→
env-179	ABEF01030085.1	-	1496	1462	**RNA**→ ←CoxL (COG1529)
env-180	ABEF01054369.1	-	2911	2877	←ABC_MJ0796_Lo1CDE_FtsE (cd03255) **RNA**→ hypo→
env-181	ABEF01053713.1	+	2290	2324	←hypo **RNA**→ ←Ald_Xan_dh_C2 (pfam02738)
env-182	ABEF01036124.1	-	1108	1074	**RNA**→ Gid (COG1206)→
env-183	ABEF01052997.1	+	617	651	←hypo **RNA**→ COG0637 (COG0637)→
env-184	ABEF01047701.1	-	2485	2451	**RNA**→ ←FadB (COG1250)3HCDH (pfam00725)
env-185	ABEF01049519.1	+	1091	1125	MR_like (cd03316)→ **RNA**→
env-186	JCVI_SCAF_1101668567042	+	73	107	**RNA**→
env-188	ABEF01018597.1	+	691	725	←hypo **RNA**→
env-187	ABEF01018597.1	+	543	577	(shown above)
env-189	ABEF01033210.1	-	316	282	hypo→ **RNA**→ ←hypo
Oal-1-6	NZ_AAMQ01000001.1	-	730002	729968	hypo→ **RNA**→ ←HisF (COG0107)

env-190	JCVI_SCAF_1101668112714	-	622	588	MhpD (COG0179)? (GOS5149)→ **RNA**→ ←? (GOS11112103)
env-192	gws2_d4_0304_2	+	8289	8323	←Qor (COG0604) **RNA**→ ←Sfp (COG2091)
env-193	ABEF01035364.1	-	448	414	COG5573 (COG5573)→ **RNA**→ ←hypo
env-194	ABEF01004283.1	+	412	446	hypo→ **RNA**→ ←hypo
env-195	JCVI_SCAF_1096626899633	+	886	920	? (GOS987)→ **RNA**→ NTP_transf_2 (pfam01909)? (GOS11233935)→
env-196	ABEF01053726.1	+	1583	1617	←DHDPS-like (cd00408) **RNA**→ ←GLY1 (COG2008)
Swi-1-2	NC_009511.1	-	1561222	1561188	PtrB (COG1770)→ **RNA**→ ←CcmB (pfam03379)
Swi-1-3	NC_009511.1	+	2181355	2181389	hypo→ **RNA**→ Sugar_tr (pfam00083)→
Swi-1-4	NC_009511.1	+	5165076	5165110	SMC_N (pfam02463)SMC_C (pfam02483)→ **RNA**→ sialate_O-acetylesterase_like2 (cd01828)→
Swi-1-5	NC_009511.1	+	1465185	1465219	Hit (COG0537)→ **RNA**→ ←COG1213 (COG1213)
Swi-1-6	NC_009511.1	-	3971601	3971567	←EmrE (COG2076) **RNA**→ ←OtsA (COG0380)
Swi-1-7	NC_009511.1	-	2080479	2080445	ligand_gated_channel (cd01347)→ **RNA**→ ←AslA (COG3119)
Swi-1-8	NC_009511.1	+	364033	364067	CaiD (COG1024)→ **RNA**→ ←SCP-x_thiolase (cd00829)
Swi-1-9	NC_009511.1	-	5147495	5147461	hypo→ **RNA**→ hypo→
Swi-1-10	NC_009511.1	-	2661570	2661536	←YtcJ_like (cd01300) **RNA**→ ←MmsB (COG2084)
Swi-1-11	NC_009511.1	-	1086089	1086055	←MhpC (COG0596) **RNA**→ ←1,2-CTD (cd03460)
Swi-1-12	NC_009511.1	-	1067438	1067404	←hypo **RNA**→ ←FkbR2 (cd03451)
Swi-1-13	NC_009511.1	+	4652756	4652790	hypo→ **RNA**→ PnbA (COG2272)→
Mma-1-24	NZ_AAAP01003860.1	-	7829	7795	←FcbC (COG0824) **RNA**→ ←RuvB (COG2255)
Psp-1-24	NC_007948.1	-	4224475	4224441	←AcrR (COG1309) **RNA**→ ←Phasin (pfam05597)NOG13350 (NOG13350)
Psp-1-25	NC_007948.1	+	578327	578361	←COG0714 (COG0714) **RNA**→ ←XdhA (COG4630)
Psp-1-26	NC_007948.1	+	118540	118574	hypo→ **RNA**→ ←UraA (COG2233)
Psp-1-27	NC_007948.1	+	319843	319878	hypo→ **RNA**→ hypo→
Psp-1-28	NC_007948.1	-	451348	451314	MR_like (cd03316)→ **RNA**→ Bug (pfam03401)→
Psp-1-29	NC_007948.1	-	4115638	4115604	COG4941 (COG4941)→ **RNA**→ TetR_N (pfam00440)NOG14032 (NOG14032)→
Psp-1-30	NC_007948.1	+	4356729	4356763	hypo→ **RNA**→ ←DhlC (COG4147)
Psp-1-31	NC_007948.1	-	2168178	2168144	hypo→ **RNA**→ hypo→
Psp-1-32	NC_007948.1	-	4128181	4128147	COG4339 (COG4339)→ **RNA**→ RHOD (cd00158)HTH_ARSR (smart00418)→
Psp-1-33	NC_007948.1	-	631111	631077	HTH_MARR (smart00347)→ **RNA**→ OsmC (COG1764)→
Psp-1-34	NC_007948.1	+	1081359	1081393	←Polysacc_deac_1 (pfam01522) **RNA**→ ←GloB (COG0491)NUDIX (pfam00293)
Psp-1-35	NC_007948.1	-	3662321	3662287	←RluA (COG0564) **RNA**→ ←COG4628 (COG4628)
Psp-1-36	NC_007948.1	+	1026625	1026659	←PotD (COG0687) **RNA**→ ←PotA (COG3842)
Psp-1-37	NC_007948.1	-	1241334	1241300	OsmY (COG2823)→ **RNA**→ IbpA (COG0071)→
Psp-1-38	NC_007948.1	-	3614971	3614937	←PurR (COG1609) **RNA**→ Cellulase (pfam00150)→
Psp-1-39	NC_007948.1	-	2881841	2881807	hypo→ **RNA**→ ←Vgb (COG4257)
Psp-1-40	NC_007948.1	+	2683163	2683197	←adh_short (pfam00106) **RNA**→ ←ElaA (COG2153)
Sal-1-12	NC_008048.1	-	2275314	2275280	HTH_SoxR (cd01110)→ **RNA**→ ←COG1359 (COG1359)
Sal-1-13	NC_008048.1	+	1531307	1531341	IlvH (COG0440)→ **RNA**→ IlvC (COG0059)→
Sal-1-14	NC_008048.1	+	738714	738748	←hypo **RNA**→ ←COG2827 (COG2827)
env-197	JCVI_SCAF_1101668113247	-	103	69	CIMS_C_terminal_like (cd03311)? (GOS5458)→ **RNA**→
env-191	JCVI_SCAF_1101668113247	-	291	257	(shown above)
env-198	gws2_d4_0301_2	+	6425	6459	←SIRT5_Af1_CobB (cd01412) **RNA**→ MurC (COG0773)→
env-199	JCVI_SCAF_1096626896474	+	1463	1497	←SerS (COG0172)? (GOS1816) **RNA**→
env-146	JCVI_SCAF_1096626896474	+	1356	1390	(shown above)
env-200	JCVI_SCAF_1101668588731	+	999	1033	←REC (cd00156) **RNA**→
env-201	wf2_AHAA11964_y1	+	433	467	hypo→ **RNA**→ ←EXOIII (smart00479)
Fpe-1-1	NZ_AATP01000005.1	+	101099	101133	←hypo **RNA**→ **RNA**→ ←hypo
Fpe-1-5	NZ_AATP01000005.1	+	101182	101216	(shown above)
Fpe-1-2	NZ_AATP01000001.1	-	825726	825692	hypo→ **RNA**→ hypo→
env-202	ABEF01034290.1	+	1560	1594	hypo→ **RNA**→
Sal-1-15	NC_008048.1	-	390589	390555	←PQQ_DH (cd00216)COG2863 (COG2863) **RNA**→ ←MetAP1 (cd01086)
Fpe-1-3	NZ_AATP01000001.1	+	654235	654269	GGDEF (cd01949)→ **RNA**→ ←SerA (COG0111)

Fpe-1-4	NZ_AATP01000001.1	+	825512	825546	←hypo **RNA**→ **RNA**→ ←hypo
Fpe-1-8	NZ_AATP01000001.1	+	825598	825632	(shown above)
Fpe-1-6	NZ_AATP01000001.1	+	133030	133064	hypo→ **RNA**→ hypo→
Fpe-1-7	NZ_AATP01000003.1	-	2567	2533	hypo→ **RNA**→ SSL2 (COG1061)→
env-203	soil_XZS64445_x1	+	54	88	**RNA**→ ←PurL (cd02193)AIRS_C (pfam02769)
Sal-1-16	NC_008048.1	+	291797	291831	hypo→ **RNA**→ hypo→
env-204	JCVI_SCAF_1101668105527	+	734	768	? (GOS1138)→ **RNA**→ ←COG5569 (COG5569)? (GOS3488)
env-205	JCVI_SCAF_1096627430754	+	18	52	**RNA**→ ←PyrD (COG0167)? (GOS684)
env-206	JCVI_SCAF_1101668642023	+	107	141	**RNA**→
env-207	JCVI_SCAF_1101668384162	-	939	905	**RNA**→ SBP_bac_5 (pfam00496)? (GOS423)→
env-208	wf2_AHAA7366_x1	+	535	569	hypo→ **RNA**→
env-209	JCVI_SCAF_1101668627300	+	792	826	YqgFc (smart00732)? (GOS10025)→ **RNA**→ tRNA_bind_EcMetRS_like (cd02800)? (GOS1316)→
env-210	wf2_AHAA16149_x1	+	244	279	**RNA**→
env-211	wf2_AHAA14901_x1	-	305	270	RfbB (COG1088)→ **RNA**→ dTDP-glucose pyrophosphorylase(EC:2.7.7.24)→
env-212	wf2_AHAA9369_x1	-	228	193	←hypo **RNA**→
Ssp-2-2	NC_009663.1	-	46842	46808	←Dgt (COG0232) **RNA**→ ←Ribosomal_L20 (pfam00453)
env-213	JCVI_SCAF_1101668631463	-	336	302	RadC (COG2003)? (GOS9318)→ **RNA**→
env-214	wf3_AHAI13901_x1	-	295	261	COG2827 (COG2827)→ **RNA**→
Swi-1-14	NC_009511.1	+	2410609	2410643	←SalY (COG0577) **RNA**→ ←ABC_MJ0796_Lo1CDE_FtsE (cd03255)
Gfo-1-1	NC_008571.1	+	366995	367029	hypo→ **RNA**→ MET17 (COG2873)→
env-215	JCVI_SCAF_1096627148666	+	1720	1754	COG2827 (COG2827)? (GOS1465)→ **RNA**→ TrpB (COG0133)? (GOS1336)→
Nar-1-6	NC_007794.1	+	638841	638875	←COG1835 (COG1835)NOG38593 (NOG38593) **RNA**→ ←AarF (COG0661)
Swi-1-15	NC_009511.1	-	55649	55615	Transgly (pfam00912)→ **RNA**→ ←DUF149 (pfam02575)
Swi-1-16	NC_009511.1	-	3587240	3587206	Lrp (COG1522)→ **RNA**→ ←AcrR (COG1309)
Swi-1-17	NC_009511.1	+	804632	804666	←ligand_gated_channel (cd01347) **RNA**→ ←Lrp (COG1522)
Swi-1-18	NC_009511.1	-	2181454	2181420	←Sugar_tr (pfam00083) **RNA**→ ←hypo
env-216	wf3_AHAI10348_x1	+	505	539	←fer2 (cd00207) **RNA**→
env-217	wf3_Contig6965	+	823	857	←SIRT5_Af1_CobB (cd01412) **RNA**→ DapB (COG0289)→
env-218	JCVI_SCAF_1096628181124	-	1054	1020	**RNA**→
env-219	wf2_AHAA9106_y1	-	408	372	**RNA**→ ←Glycyl-tRNA synthetase, beta subunit(EC:6.1.1.14)
env-220	wf2_Contig2936	+	546	582	hypo→ **RNA**→
env-221	wf2_AHAA14676_x1	+	563	599	hypo→ **RNA**→
env-222	wf2_AHAA16792_x1	+	208	244	**RNA**→ ←Ribonucleotide reductase, alpha subunit
env-223	wf2_Contig2452	-	1022	986	**RNA**→ ←FabI (COG0623)
env-224	wf2_AHAA19774_x1	+	270	306	**RNA**→ ←tRNA delta(2)-isopentenylpyrophosphate transferase(EC:2.5.1.8)
env-225	wf2_Contig4200	+	652	688	DNA modification methylase→ **RNA**→ ←A/G-specific DNA glycosylase
env-226	wf2_Contig2452	+	934	970	FabI (COG0623)→ **RNA**→
env-227	wf2_AHAA19650_x1	-	413	377	Predicted secreted protein→ **RNA**→

38.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 38.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one GUCCY-hairpin RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

D cd00079 (1) Helicase superfamily c-terminal domain; associated with DEXDc-, DEAD-, and DEAH-box proteins, yeast initiation factor 4A, Ski2p, and Hepatitis C virus NS3 helicases; this domain is found in a wide variety of helicases and helicase related proteins; may not be an autonomously folding unit, but an integral part of the helicase; 4 helicase superfamilies at present according to the organization of their signature motifs; all helicases share the ability to unwind nucleic acid duplexes with a distinct directional polarity; they utilize the free energy from nucleoside triphosphate hydrolysis to fuel their translocation along DNA, unwinding the duplex in the process

u cd00093 (1) Helix-turn-helix XRE-family like proteins.
d cd00138 (1) Phospholipase D.
u d cd00156 (4) Signal receiver domain; originally thought to be unique to bacteria (CheY, OmpR, NtrC, and PhoB), now recently identified in eukaroytes ETR1 Arabidopsis thaliana; this domain receives the signal from the sensor partner in a two-component systems; contains a phosphoacceptor site that is phosphorylated by histidine kinase homologs; usually found N-terminal to a DNA binding effector domain; forms homodimers
D cd00158 (1) Rhodanese Homology Domain (RHOD); an alpha beta fold domain found duplicated in the rhodanese protein.
u d cd00189 (6) Tetratricopeptide repeat domain; typically contains 34 amino acids [WLF]-X(2)-[LIM]-[GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-X(2)-[ASL]-X(4)-[PKE]
uD cd00207 (2) 2Fe-2S iron-sulfur cluster binding domain.
u cd00216 (1) Dehydrogenases with pyrrolo-quinoline quinone (PQQ) as cofactor, like ethanol, methanol, and membrane bound glucose dehydrogenases.
u cd00288 (1) Pyruvate kinase (PK): Large allosteric enzyme that regulates glycolysis through binding of the substrate, phosphoenolpyruvate, and one or more allosteric effectors.
d cd00293 (1) Usp: Universal stress protein family.
d cd00383 (1) Effector domain of response regulator.
u cd00401 (1) S-adenosyl-L-homocysteine hydrolase (AdoHycase) catalyzes the hydrolysis of S-adenosyl-L-homocysteine (AdoHyc) to form adenosine (Ado) and homocysteine (Hcy).
u cd00408 (1) Dihydrodipicolinate synthase family.
uD cd00412 (2) Inorganic pyrophosphatase.
uD cd00431 (2) Cysteine hydrolases; This family contains amidohydrolases, like CSHase (N-carbamoylsarcosine amidohydrolase), involved in creatine metabolism and nicotinamidase, converting nicotinamide to nicotinic acid and ammonia in the pyridine nucleotide cycle.
u cd00484 (1) Phosphoenolpyruvate carboxykinase (PEPCK), a critical gluconeogenic enzyme, catalyzes the first committed step in the diversion of tricarboxylic acid cycle intermediates toward gluconeogenesis.
d cd00544 (1) Adenosylcobinamide kinase / adenosylcobinamide phosphate guanyltransferase (CobU).
d cd00688 (1) This group contains class II terpene cyclases, protein prenyltransferases beta subunit, two broadly specific proteinase inhibitors alpha2-macroglobulin (alpha (2)-M) and pregnancy zone protein (PZP) and, the C3 C4 and C5 components of vertebrate complement.
u cd00740 (1) MeTr subgroup of pterin binding enzymes.
u d cd00751 (2) Thiolase are ubiquitous enzymes that catalyze the reversible thiolytic cleavage of 3-ketoacyl-CoA into acyl-CoA and acetyl-CoA, a 2-step reaction involving a covalent intermediate formed with a catalytic cysteine.
d cd00808 (1) Descriminating Glutamyl-tRNA synthetase (GluRS) catalytic core domain .
uDd cd00829 (3) Thiolase domain associated with sterol carrier protein (SCP)-x isoform and related proteins; SCP-2 has multiple roles in intracellular lipid circulation and metabolism.
D cd00830 (1) Ketoacyl-acyl carrier protein synthase III (KASIII) initiates the elongation in type II fatty acid synthase systems.
u cd00987 (1) PDZ domain of tryspin-like serine proteases, such as DegP/HtrA, which are oligomeric proteins involved in heat-shock response, chaperone function, and apoptosis.
u cd01012 (1) YcaC related amidohydrolases; E.
d cd01086 (1) Methionine Aminopeptidase 1.
u cd01110 (1) Helix-turn-helix transcriptional regulator SoxR.
D cd01115 (1) Permease SLC13 (solute carrier 13).
uDd cd01153 (4) Putative acyl-CoA dehydrogenase (ACAD).
u cd01275 (1) FHIT (fragile histidine family): FHIT proteins, related to the HIT family carry a motif HxHxH/Qxx (x, is a hydrophobic amino acid), On the basis of sequence, substrate specificity, structure, evolution and mechanism, HIT proteins are classified into three branches: the Hint branch, which consists of adenosine 5' -monophosphoramide hydrolases, the Fhit branch, that consists of diadenosine polyphosphate hydrolases, and the GalT branch consisting of specific nucloside monophosphate transferases.
u cd01276 (1) Protein Kinase C Interacting protein related (PKCI): PKCI and related proteins belong to the ubiquitous HIT family of hydrolases that act on alpha-phosphates of ribonucleotides.
u cd01277 (1) HINT (histidine triad nucleotide-binding protein) subgroup: Members of this CD belong to the superfamily of histidine triad hydrolases that act on alpha-phosphate of ribonucleotides.
D cd01288 (1) FabZ is a 17kD beta-hydroxyacyl-acyl carrier protein (ACP) dehydratase that primarily catalyzes the dehydration of beta-hydroxyacyl-ACP to trans-2-acyl-ACP, the third step in the elongation phase of the bacterial/ plastid, type II, fatty-acid biosynthesis pathway.
u cd01300 (1) YtcJ_like metal dependent amidohydrolases.
u cd01347 (2) TonB dependent/Ligand-Gated channels are created by a monomeric 22 strand (22,24) anti-parallel beta-barrel.
u cd01412 (2) SIRT5_Af1_CobB: Eukaryotic, archaeal and prokaryotic group (class3) which includes human sirtuin SIRT5, Archaeoglobus fulgidus Sir2-Af1, and E.
uD cd01467 (2) VWA BatA type: Von Willebrand factor type A (vWA) domain was originally found in the blood coagulation protein von Willebrand factor (vWF).
uD cd01516 (2) Bacterial fructose-1,6-bisphosphatase, glpX-encoded.
u d cd01521 (2) Member of the Rhodanese Homology Domain superfamily.
D cd01558 (1) D-Alanine aminotransferase (D-AAT_like): D-amino acid aminotransferase catalyzes transamination between D-amino acids and their respective alpha-keto acids.
uD cd01638 (2) CysQ, a 3'-Phosphoadenosine-5'-phosphosulfate (PAPS) 3'-phosphatase, is a bacterial member of the inositol monophosphatase family.
u d cd01639 (2) IMPase, inositol monophosphatase and related domains.
u cd01714 (1) The electron transfer flavoprotein (ETF) serves as a specific electron acceptor for various mitochondrial dehydrogenases.
D cd01828 (1) sialate_O-acetylesterase_like subfamily of the SGNH-hydrolases, , a diverse family of lipases and esterases.
u cd01949 (2) Diguanylate-cyclase (DGC) or GGDEF domain: Originally named after a conserved residue pattern, and initially described as domain of unknown function 1 (DUF1).
u d cd01998 (2) tRNA methyl transferase.
D cd02068 (1) B12 binding domain_like associated with radical SAM domain.
u cd02108 (1) bacterial subgroup of the sulfite oxidase (SO) family of molybdopterin binding domains.
u d cd02172 (2) RfaE is a protein involved in the biosynthesis of ADP-L-glycero-D-manno-heptose, a precursor for LPS inner core biosynthesis.
d cd02193 (1) Formylglycinamide ribonucleotide amidotransferase (FGAR-AT) catalyzes the ATP-dependent conversion of formylglycinamide ribonucleotide (FGAR) and glutamine to formylglycinamidine ribonucleotide (FGAM), ADP, phosphate, and glutamate in the fourth step of the purine biosynthetic pathway.
u cd02563 (1) tRNA pseudouridine isomerase C: Pseudouridine synthases catalyze the isomer-

ization of specific uridines in an tRNA molecule to pseudouridines (5-ribosyluracil, psi).
uDd cd02573 (4) PseudoU_synth_EcTruB: Pseudouridine synthase, Escherichia coli TruB like.
D cd02800 (1) tRNA-binding-domain-containing Escherichia coli methionyl-tRNA synthetase (EcMetRS)-like proteins.
u d cd02892 (2) Squalene cyclase (SQCY) domain subgroup 1; found in class II terpene cyclases that have an alpha 6 - alpha 6 barrel fold.
uD cd02932 (2) Old yellow enzyme (OYE) YqjM-like FMN binding domain.
D cd02955 (1) TRX domain, SSP411 protein family; members of this family are highly conserved proteins present in eukaryotes, bacteria and archaea, about 600-800 amino acids in length, which contain a TRX domain with a redox active CXXC motif.
uD cd03015 (2) Peroxiredoxin (PRX) family, Typical 2-Cys PRX subfamily; PRXs are thiol-specific antioxidant (TSA) proteins, which confer a protective role in cells through its peroxidase activity by reducing hydrogen peroxide, peroxynitrite, and organic hydroperoxides.
u d cd03046 (2) GST_N family, Saccharomyces cerevisiae GTT1-like subfamily; composed of predominantly uncharacterized proteins with similarity to the S.
u cd03108 (1) Adenylosuccinate synthetase (AdSS) catalyzes the first step in the de novo biosynthesis of AMP.
u d cd03189 (2) GST_C family, Saccharomyces cerevisiae GTT1-like subfamily; composed of predominantly uncharacterized proteins with similarity to the S.
uD cd03238 (2) The excision repair protein UvrA; Nucleotide excision repair in eubacteria is a process that repairs DNA damages by the removal of a 12-13-mer oligonucleotide containing the lesion.
u d cd03255 (2) This family is comprised of MJ0796 ATP-binding cassette, macrolide-specific ABC-type efflux carrier (MacAB), and proteins involved in cell division (FtsE) and release of liporoteins from the cytoplasmic membrane (LolCDE).
u cd03311 (1) CIMS - Cobalamine-independent methonine synthase, or MetE, C-terminal domain_like.
u cd03316 (2) Mandelate racemase (MR)-like subfamily of the enolase superfamily.
u d cd03368 (2) S12-like family, 30S ribosomal protein S12 subfamily; S12 is located at the interface of the large and small ribosomal subunits of prokaryotes, chloroplasts and mitochondria, where it plays an important role in both tRNA and ribosomal subunit interactions.
u cd03443 (1) PaaI_thioesterase is a tetrameric acyl-CoA thioesterase with a hot dog fold and one of several proteins responsible for phenylacetic acid (PA) degradation in bacteria.
d cd03451 (1) FkbR2 is a Streptomyces hygroscopicus protein with a hot dog fold that belongs to a conserved family of proteins found in prokaryotes and archaea but not in eukaryotes.
d cd03460 (1) Catechol 1,2 dioxygenase (1,2-CTD) catalyzes an intradiol cleavage reaction of catechol to form cis,cis-muconate.
u d cd03467 (2) Rieske domain; a [2Fe-2S] cluster binding domain commonly found in Rieske non-heme iron oxygenase (RO) systems such as naphthalene and biphenyl dioxygenases, as well as in plant/cyanobacterial chloroplast b6f and mitochondrial cytochrome bc(1) complexes. The Rieske domain can be divided into two subdomains, with an incomplete six-stranded, antiparallel beta-barrel at one end, and an iron-sulfur cluster binding subdomain at the other. The Rieske iron-sulfur center contains a [2Fe-2S]
d cd03586 (1) Pol_IV_kappa, a member of the Y-family of DNA polymerases.
u d COG0003 (2) Predicted ATPase involved in chromosome partitioning [Cell division and chromosome partitioning]
u COG0007 (1) Uroporphyrinogen-III methylase [Coenzyme metabolism]
u COG0009 (1) Putative translation factor (SUA5) [Translation, ribosomal structure and biogenesis]
u COG0013 (1) Alanyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
u COG0025 (1) NhaP-type Na+/H+ and K+/H+ antiporters [Inorganic ion transport and metabolism]
u COG0049 (1) Ribosomal protein S7 [Translation, ribosomal structure and biogenesis]
u d COG0050 (2) GTPases - translation elongation factors [Translation, ribosomal structure and biogenesis]
D COG0059 (1) Ketol-acid reductoisomerase [Amino acid transport and metabolism / Coenzyme metabolism]
u COG0060 (1) Isoleucyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
d COG0068 (1) Hydrogenase maturation factor [Posttranslational modification, protein turnover, chaperones]
D COG0071 (2) Molecular chaperone (small heat shock protein) [Posttranslational modification, protein turnover, chaperones]
u COG0072 (1) Phenylalanyl-tRNA synthetase beta subunit [Translation, ribosomal structure and biogenesis]
uDd COG0077 (4) Prephenate dehydratase [Amino acid transport and metabolism]
u COG0078 (1) Ornithine carbamoyltransferase [Amino acid transport and metabolism]
uD COG0079 (2) Histidinol-phosphate/aromatic aminotransferase and cobyric acid decarboxylase [Amino acid transport and metabolism]
u COG0086 (1) DNA-directed RNA polymerase, beta' subunit/160 kD subunit [Transcription]
u d COG0107 (2) Imidazoleglycerol-phosphate synthase [Amino acid transport and metabolism]
d COG0111 (1) Phosphoglycerate dehydrogenase and related dehydrogenases [Amino acid transport and metabolism]
u COG0119 (1) Isopropylmalate/homocitrate/citramalate synthases [Amino acid transport and metabolism]
D COG0133 (1) Tryptophan synthase beta chain [Amino acid transport and metabolism]
u COG0135 (1) Phosphoribosylanthranilate isomerase [Amino acid transport and metabolism]
d COG0137 (1) Argininosuccinate synthase [Amino acid transport and metabolism]
d COG0156 (1) 7-keto-8-aminopelargonate synthetase and related enzymes [Coenzyme metabolism]
d COG0161 (1) Adenosylmethionine-8-amino-7-oxononanoate aminotransferase [Coenzyme metabolism]
d COG0167 (1) Dihydroorotate dehydrogenase [Nucleotide transport and metabolism]
u COG0172 (1) Seryl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
u COG0179 (1) 2-keto-4-pentenoate hydratase/2-oxohepta-3-ene-1,7-dioic acid hydratase (catechol pathway) [Secondary metabolites biosynthesis, transport, and catabolism]
d COG0196 (1) FAD synthase [Coenzyme metabolism]
u COG0202 (1) DNA-directed RNA polymerase, alpha subunit/40 kD subunit [Transcription]
D COG0204 (1) 1-acyl-sn-glycerol-3-phosphate acyltransferase [Lipid metabolism]
d COG0210 (1) Superfamily I DNA and RNA helicases [DNA replication, recombination, and repair]
uD COG0226 (2) ABC-type phosphate transport system, periplasmic component [Inorganic ion transport and metabolism]
u COG0232 (1) dGTP triphosphohydrolase [Nucleotide transport and metabolism]
u COG0239 (1) Integral membrane protein possibly involved in chromosome condensation [Cell division and chromosome partitioning]
u d COG0247 (3) Fe-S oxidoreductase [Energy production and conversion]

D COG0249 (1) Mismatch repair ATPase (MutS family) [DNA replication, recombination, and repair]
u d COG0250 (3) Transcription antiterminator [Transcription]
uDd COG0265 (4) Trypsin-like serine proteases, typically periplasmic, contain C-terminal PDZ domain [Posttranslational modification, protein turnover, chaperones]
u COG0272 (1) NAD-dependent DNA ligase (contains BRCT domain type II) [DNA replication, recombination, and repair]
u COG0277 (1) FAD/FMN-containing dehydrogenases [Energy production and conversion]
u COG0280 (1) Phosphotransacetylase [Energy production and conversion]
u COG0281 (2) Malic enzyme [Energy production and conversion]
u COG0286 (1) Type I restriction-modification system methyltransferase subunit [Defense mechanisms]
uDd COG0289 (3) Dihydrodipicolinate reductase [Amino acid transport and metabolism]
u COG0298 (1) Hydrogenase maturation factor [Posttranslational modification, protein turnover, chaperones]
u d COG0300 (2) Short-chain dehydrogenases of various substrate specificities [General function prediction only]
d COG0307 (1) Riboflavin synthase alpha chain [Coenzyme metabolism]
uD COG0308 (2) Aminopeptidase N [Amino acid transport and metabolism]
u d COG0312 (2) Predicted Zn-dependent proteases and their inactivated homologs [General function prediction only]
uD COG0318 (3) Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II [Lipid metabolism / Secondary metabolites biosynthesis, transport, and catabolism]
u COG0343 (1) Queuine/archaeosine tRNA-ribosyltransferase [Translation, ribosomal structure and biogenesis]
uD COG0345 (2) Pyrroline-5-carboxylate reductase [Amino acid transport and metabolism]
u d COG0349 (2) Ribonuclease D [Translation, ribosomal structure and biogenesis]
u COG0352 (1) Thiamine monophosphate synthase [Coenzyme metabolism]
u COG0353 (1) Recombinational DNA repair protein (RecF pathway) [DNA replication, recombination, and repair]
d COG0359 (1) Ribosomal protein L9 [Translation, ribosomal structure and biogenesis]
D COG0361 (1) Translation initiation factor 1 (IF-1) [Translation, ribosomal structure and biogenesis]
u d COG0365 (2) Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases [Lipid metabolism]
d COG0380 (1) Trehalose-6-phosphate synthase [Carbohydrate transport and metabolism]
d COG0388 (2) Predicted amidohydrolase [General function prediction only]
u COG0396 (1) ABC-type transport system involved in Fe-S cluster assembly, ATPase component [Posttranslational modification, protein turnover, chaperones]
d COG0404 (1) Glycine cleavage system T protein (aminomethyltransferase) [Amino acid transport and metabolism]
D COG0410 (1) ABC-type branched-chain amino acid transport systems, ATPase component [Amino acid transport and metabolism]
u d COG0427 (2) Acetyl-CoA hydrolase [Energy production and conversion]
uD COG0436 (2) Aspartate/tyrosine/aromatic aminotransferase [Amino acid transport and metabolism]
u COG0438 (1) Glycosyltransferase [Cell envelope biogenesis, outer membrane]
d COG0439 (1) Biotin carboxylase [Lipid metabolism]
u COG0440 (1) Acetolactate synthase, small (regulatory) subunit [Amino acid transport and metabolism]
u d COG0449 (2) Glucosamine 6-phosphate synthetase, contains amidotransferase and phosphosugar isomerase domains [Cell envelope biogenesis, outer membrane]
u d COG0451 (2) Nucleoside-diphosphate-sugar epimerases [Cell envelope biogenesis, outer membrane / Carbohydrate transport and metabolism]
d COG0456 (1) Acetyltransferases [General function prediction only]
D COG0471 (1) Di- and tricarboxylate transporters [Inorganic ion transport and metabolism]
u COG0472 (1) UDP-N-acetylmuramyl pentapeptide phosphotransferase/UDP-N-acetylglucosamine-1-phosphate transferase [Cell envelope biogenesis, outer membrane]
d COG0480 (1) Translation elongation factors (GTPases) [Translation, ribosomal structure and biogenesis]
Dd COG0491 (3) Zn-dependent hydrolases, including glyoxylases [General function prediction only]
uD COG0507 (2) ATP-dependent exoDNAse (exonuclease V), alpha subunit - helicase superfamily I member [DNA replication, recombination, and repair]
uD COG0513 (5) Superfamily II DNA and RNA helicases [DNA replication, recombination, and repair / Transcription / Translation, ribosomal structure and biogenesis]
u d COG0520 (2) Selenocysteine lyase [Amino acid transport and metabolism]
d COG0522 (1) Ribosomal protein S4 and related proteins [Translation, ribosomal structure and biogenesis]
D COG0531 (1) Amino acid transporters [Amino acid transport and metabolism]
u COG0532 (1) Translation initiation factor 2 (IF-2; GTPase) [Translation, ribosomal structure and biogenesis]
u COG0535 (1) Predicted Fe-S oxidoreductases [General function prediction only]
u COG0537 (1) Diadenosine tetraphosphate (Ap4A) hydrolase and other HIT family hydrolases [Nucleotide transport and metabolism / Carbohydrate transport and metabolism / General function prediction only]
u d COG0539 (3) Ribosomal protein S1 [Translation, ribosomal structure and biogenesis]
D COG0541 (1) Signal recognition particle GTPase [Intracellular trafficking and secretion]
u COG0544 (1) FKBP-type peptidyl-prolyl cis-trans isomerase (trigger factor) [Posttranslational modification, protein turnover, chaperones]
d COG0558 (1) Phosphatidylglycerophosphate synthase [Lipid metabolism]
uD COG0559 (3) Branched-chain amino acid ABC-type transport system, permease components [Amino acid transport and metabolism]
u d COG0564 (2) Pseudouridylate synthases, 23S RNA-specific [Translation, ribosomal structure and biogenesis]
u COG0577 (1) ABC-type antimicrobial peptide transport system, permease component [Defense mechanisms]
u d COG0583 (3) Transcriptional regulator [Transcription]
uDd COG0596 (4) Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily) [General function prediction only]
Dd COG0601 (2) ABC-type dipeptide/oligopeptide/nickel transport systems, permease components [Amino acid transport and metabolism / Inorganic ion transport and metabolism]
u d COG0604 (3) NADPH:quinone reductase and related Zn-dependent oxidoreductases [Energy production and conversion / General function prediction only]
D COG0611 (1) Thiamine monophosphate kinase [Coenzyme metabolism]
u d COG0612 (3) Predicted Zn-dependent peptidases [General function prediction only]
d COG0616 (1) Periplasmic serine proteases (ClpP class) [Posttranslational modification, protein turnover, chaperones / Intracellular trafficking and secretion]

u d COG0623 (2) Enoyl-[acyl-carrier-protein] [acyl-carrier-protein]
u COG0624 (1) Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase and related deacylases [Amino acid transport and metabolism]
u COG0625 (1) Glutathione S-transferase [Posttranslational modification, protein turnover, chaperones]
u d COG0628 (2) Predicted permease [General function prediction only]
d COG0629 (1) Single-stranded DNA-binding protein [DNA replication, recombination, and repair]
D COG0637 (1) Predicted phosphatase/phosphohexomutase [General function prediction only]
u d COG0642 (2) Signal transduction histidine kinase [Signal transduction mechanisms]
d COG0652 (1) Peptidyl-prolyl cis-trans isomerase (rotamase) - cyclophilin family [Posttranslational modification, protein turnover, chaperones]
u COG0654 (1) 2-polyprenyl-6-methoxyphenol hydroxylase and related FAD-dependent oxidoreductases [Coenzyme metabolism / Energy production and conversion]
u COG0655 (1) Multimeric flavodoxin WrbA [General function prediction only]
u COG0657 (1) Esterase/lipase [Lipid metabolism]
d COG0661 (1) Predicted unusual protein kinase [General function prediction only]
u COG0664 (2) cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases [Signal transduction mechanisms]
u COG0665 (1) Glycine/D-amino acid oxidases (deaminating) [Amino acid transport and metabolism]
u COG0670 (1) Integral membrane protein, interacts with FtsH [General function prediction only]
u d COG0680 (3) Ni,Fe-hydrogenase maturation factor [Energy production and conversion]
u d COG0683 (7) ABC-type branched-chain amino acid transport systems, periplasmic component [Amino acid transport and metabolism]
u COG0687 (1) Spermidine/putrescine-binding periplasmic protein [Amino acid transport and metabolism]
u d COG0702 (2) Predicted nucleoside-diphosphate-sugar epimerases [Cell envelope biogenesis, outer membrane / Carbohydrate transport and metabolism]
u COG0703 (1) Shikimate kinase [Amino acid transport and metabolism]
uDd COG0714 (4) MoxR-like ATPases [General function prediction only]
u COG0725 (1) ABC-type molybdate transport system, periplasmic component [Inorganic ion transport and metabolism]
u COG0730 (1) Predicted permeases [General function prediction only]
u COG0747 (2) ABC-type dipeptide transport system, periplasmic component [Amino acid transport and metabolism]
D COG0773 (1) UDP-N-acetylmuramate-alanine ligase [Cell envelope biogenesis, outer membrane]
u COG0775 (1) Nucleoside phosphorylase [Nucleotide transport and metabolism]
d COG0780 (1) Enzyme related to GTP cyclohydrolase I [General function prediction only]
u d COG0782 (2) Transcription elongation factor [Transcription]
u d COG0797 (2) Lipoproteins [Cell envelope biogenesis, outer membrane]
u d COG0801 (4) 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase [Coenzyme metabolism]
d COG0802 (1) Predicted ATPase or kinase [General function prediction only]
u COG0820 (1) Predicted Fe-S-cluster redox enzyme [General function prediction only]
uD COG0824 (2) Predicted thioesterase [General function prediction only]
d COG0841 (1) Cation/multidrug efflux pump [Defense mechanisms]
uD COG0858 (3) Ribosome-binding factor A [Translation, ribosomal structure and biogenesis]
uD COG0859 (2) ADP-heptose:LPS heptosyltransferase [Cell envelope biogenesis, outer membrane]
u COG0861 (1) Membrane protein TerC, possibly involved in tellurium resistance [Inorganic ion transport and metabolism]
u COG1011 (1) Predicted hydrolase (HAD superfamily) [General function prediction only]
d COG1022 (1) Long-chain acyl-CoA synthetases (AMP-forming) [Lipid metabolism]
u d COG1024 (9) Enoyl-CoA hydratase/carnithine racemase [Lipid metabolism]
u d COG1028 (3) Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases) [Secondary metabolites biosynthesis, transport, and catabolism / General function prediction only]
u d COG1032 (2) Fe-S oxidoreductase [Energy production and conversion]
D COG1034 (1) NADH dehydrogenase/NADH:ubiquinone oxidoreductase 75 kD subunit (chain G) [Energy production and conversion]
d COG1042 (1) Acyl-CoA synthetase (NDP forming) [Energy production and conversion]
u COG1051 (1) ADP-ribose pyrophosphatase [Nucleotide transport and metabolism]
d COG1053 (1) Succinate dehydrogenase/fumarate reductase, flavoprotein subunit [Energy production and conversion]
D COG1061 (1) DNA or RNA helicases of superfamily II [Transcription / DNA replication, recombination, and repair]
d COG1062 (1) Zn-dependent alcohol dehydrogenases, class III [Energy production and conversion]
D COG1087 (1) UDP-glucose 4-epimerase [Cell envelope biogenesis, outer membrane]
u COG1088 (1) dTDP-D-glucose 4,6-dehydratase [Cell envelope biogenesis, outer membrane]
u d COG1091 (2) dTDP-4-dehydrorhamnose reductase [Cell envelope biogenesis, outer membrane]
u COG1123 (1) ATPase components of various ABC-type transport systems, contain duplicated ATPase [General function prediction only]
d COG1132 (1) ABC-type multidrug transport system, ATPase and permease components [Defense mechanisms]
d COG1159 (2) GTPase [General function prediction only]
u COG1160 (1) Predicted GTPases [General function prediction only]
u d COG1161 (2) Predicted GTPases [General function prediction only]
uD COG1186 (2) Protein chain release factor B [Translation, ribosomal structure and biogenesis]
D COG1197 (1) Transcription-repair coupling factor (superfamily II helicase) [DNA replication, recombination, and repair / Transcription]
D COG1201 (1) Lhr-like helicases [General function prediction only]
d COG1204 (1) Superfamily II helicase [General function prediction only]
D COG1206 (1) NAD(FAD)-utilizing enzyme possibly involved in translation [Translation, ribosomal structure and biogenesis]
uD COG1210 (2) UDP-glucose pyrophosphorylase [Cell envelope biogenesis, outer membrane]
d COG1213 (1) Predicted sugar nucleotidyltransferases [Cell envelope biogenesis, outer membrane]
d COG1220 (1) ATP-dependent protease HslVU (ClpYQ), ATPase subunit [Posttranslational modification, protein turnover, chaperones]
u COG1236 (2) Predicted exonuclease of the beta-lactamase fold involved in RNA processing [Translation, ribosomal structure and biogenesis]
u COG1239 (1) Mg-chelatase subunit ChlI [Coenzyme metabolism]
d COG1249 (1) Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydro-

genase (E3) component, and related enzymes [Energy production and conversion]
d COG1250 (1) 3-hydroxyacyl-CoA dehydrogenase [Lipid metabolism]
u COG1253 (1) Hemolysins and related proteins containing CBS domains [General function prediction only]
uD COG1280 (3) Putative threonine efflux protein [Amino acid transport and metabolism]
D COG1301 (1) Na+/H+-dicarboxylate symporters [Energy production and conversion]
u d COG1309 (2) Transcriptional regulator [Transcription]
u COG1321 (1) Mn-dependent transcriptional regulator [Transcription]
u COG1342 (1) Predicted DNA-binding proteins [General function prediction only]
u COG1345 (1) Flagellar capping protein [Cell motility and secretion]
d COG1359 (1) Uncharacterized conserved protein [Function unknown]
u COG1381 (1) Recombinational DNA repair protein (RecF pathway) [DNA replication, recombination, and repair]
u d COG1398 (2) Fatty-acid desaturase [Lipid metabolism]
u COG1416 (1) Uncharacterized conserved protein [Function unknown]
u COG1451 (1) Predicted metal-dependent hydrolase [General function prediction only]
uD COG1452 (2) Organic solvent tolerance protein OstA [Cell envelope biogenesis, outer membrane]
d COG1482 (1) Phosphomannose isomerase [Carbohydrate transport and metabolism]
d COG1487 (1) Predicted nucleic acid-binding protein, contains PIN domain [General function prediction only]
uD COG1519 (2) 3-deoxy-D-manno-octulosonic-acid transferase [Cell envelope biogenesis, outer membrane]
d COG1521 (1) Putative transcriptional regulator, homolog of Bvg accessory factor [Transcription]
u d COG1522 (7) Transcriptional regulators [Transcription]
d COG1529 (1) Aerobic-type carbon monoxide dehydrogenase, large subunit CoxL/CutL homologs [Energy production and conversion]
u COG1541 (1) Coenzyme F390 synthetase [Coenzyme metabolism]
D COG1556 (1) Uncharacterized conserved protein [Function unknown]
uD COG1569 (2) Predicted nucleic acid-binding protein, contains PIN domain [General function prediction only]
u d COG1600 (2) Uncharacterized Fe-S protein [Energy production and conversion]
uDd COG1605 (4) Chorismate mutase [Amino acid transport and metabolism]
u COG1607 (1) Acyl-CoA hydrolase [Lipid metabolism]
u COG1609 (2) Transcriptional regulators [Transcription]
u COG1610 (1) Uncharacterized conserved protein [Function unknown]
uD COG1611 (2) Predicted Rossmann fold nucleotide-binding protein [General function prediction only]
d COG1638 (1) TRAP-type C4-dicarboxylate transport system, periplasmic component [Carbohydrate transport and metabolism]
d COG1663 (1) Tetraacyldisaccharide-1-P 4'-kinase [Cell envelope biogenesis, outer membrane]
u COG1674 (1) DNA segregation ATPase FtsK/SpoIIIE and related proteins [Cell division and chromosome partitioning]
uD COG1721 (2) Uncharacterized conserved protein (some members contain a von Willebrand factor type A (vWA) domain) [General function prediction only]
u COG1748 (1) Saccharopine dehydrogenase and related proteins [Amino acid transport and metabolism]
u d COG1749 (2) Flagellar hook protein FlgE [Cell motility and secretion]
D COG1764 (1) Predicted redox protein, regulator of disulfide bond formation [Posttranslational modification, protein turnover, chaperones]
u d COG1770 (3) Protease II [Amino acid transport and metabolism]
uD COG1802 (2) Transcriptional regulators [Transcription]
d COG1804 (1) Predicted acyl-CoA transferases/carnitine dehydratase [Energy production and conversion]
u COG1828 (1) Phosphoribosylformylglycinamidine (FGAM) synthase, PurS component [Nucleotide transport and metabolism]
u COG1835 (1) Predicted acyltransferases [Lipid metabolism]
u COG1866 (1) Phosphoenolpyruvate carboxykinase (ATP) [Energy production and conversion]
u COG1893 (1) Ketopantoate reductase [Coenzyme metabolism]
uD COG1894 (3) NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit [Energy production and conversion]
u d COG1905 (2) NADH:ubiquinone oxidoreductase 24 kD subunit [Energy production and conversion]
u d COG1932 (2) Phosphoserine aminotransferase [Coenzyme metabolism / Amino acid transport and metabolism]
u COG1947 (1) 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate synthase [Lipid metabolism]
d COG1969 (1) Ni,Fe-hydrogenase I cytochrome b subunit [Energy production and conversion]
u COG2003 (1) DNA repair proteins [DNA replication, recombination, and repair]
d COG2008 (1) Threonine aldolase [Amino acid transport and metabolism]
u COG2020 (1) Putative protein-S-isoprenylcysteine methyltransferase [Posttranslational modification, protein turnover, chaperones]
u COG2025 (1) Electron transfer flavoprotein, alpha subunit [Energy production and conversion]
D COG2050 (1) Uncharacterized protein, possibly involved in aromatic compounds catabolism [Secondary metabolites biosynthesis, transport, and catabolism]
u COG2076 (1) Membrane transporters of cations and cationic drugs [Inorganic ion transport and metabolism]
u d COG2081 (2) Predicted flavoproteins [General function prediction only]
d COG2084 (1) 3-hydroxyisobutyrate dehydrogenase and related beta-hydroxyacid dehydrogenases [Lipid metabolism]
d COG2091 (1) Phosphopantetheinyl transferase [Coenzyme metabolism]
d COG2114 (1) Adenylate cyclase, family 3 (some proteins contain HAMP domain) [Signal transduction mechanisms]
u COG2130 (1) Putative NADP-dependent oxidoreductases [General function prediction only]
u COG2137 (1) Uncharacterized protein conserved in bacteria [General function prediction only]
u d COG2141 (2) Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases [Energy production and conversion]
d COG2153 (1) Predicted acyltransferase [General function prediction only]
D COG2171 (1) Tetrahydrodipicolinate N-succinyltransferase [Amino acid transport and metabolism]
u d COG2215 (3) ABC-type uncharacterized transport system, permease component [General function prediction only]
u d COG2220 (2) Predicted Zn-dependent hydrolases of the beta-lactamase fold [General function prediction only]
Dd COG2226 (2) Methylase involved in ubiquinone/menaquinone biosynthesis [Coenzyme metabolism]

D COG2227 (1) 2-polyprenyl-3-methyl-5-hydroxy-6-metoxy-1,4-benzoquinol methylase [Coenzyme metabolism]
uDd COG2230 (3) Cyclopropane fatty acid synthase and related methyltransferases [Cell envelope biogenesis, outer membrane]
d COG2233 (1) Xanthine/uracil permeases [Nucleotide transport and metabolism]
u COG2244 (1) Membrane protein involved in the export of O-antigen and teichoic acid [General function prediction only]
d COG2255 (1) Holliday junction resolvasome, helicase subunit [DNA replication, recombination, and repair]
d COG2259 (1) Predicted membrane protein [Function unknown]
D COG2271 (1) Sugar phosphate permease [Carbohydrate transport and metabolism]
D COG2272 (1) Carboxylesterase type B [Lipid metabolism]
u COG2361 (1) Uncharacterized conserved protein [Function unknown]
d COG2368 (1) Aromatic ring hydroxylase [Secondary metabolites biosynthesis, transport, and catabolism]
u COG2503 (1) Predicted secreted acid phosphatase [General function prediction only]
u COG2823 (1) Predicted periplasmic or secreted lipoprotein [General function prediction only]
uDd COG2827 (14) Predicted endonuclease containing a URI domain [DNA replication, recombination, and repair]
u d COG2863 (2) Cytochrome c553 [Energy production and conversion]
u COG2866 (1) Predicted carboxypeptidase [Amino acid transport and metabolism]
u COG2867 (1) Oligoketide cyclase/lipid transport protein [Lipid metabolism]
D COG2873 (1) O-acetylhomoserine sulfhydrylase [Amino acid transport and metabolism]
u COG2885 (1) Outer membrane protein and related peptidoglycan-associated (lipo)proteins [Cell envelope biogenesis, outer membrane]
uD COG2909 (2) ATP-dependent transcriptional regulator [Transcription]
u COG2938 (1) Uncharacterized conserved protein [Function unknown]
D COG2945 (1) Predicted hydrolase of the alpha/beta superfamily [General function prediction only]
d COG2958 (1) Uncharacterized protein conserved in bacteria [Function unknown]
u d COG2968 (2) Uncharacterized conserved protein [Function unknown]
u COG3040 (1) Bacterial lipocalin [Cell envelope biogenesis, outer membrane]
u COG3047 (1) Outer membrane protein W [Cell envelope biogenesis, outer membrane]
d COG3093 (1) Plasmid maintenance system antidote protein [General function prediction only]
d COG3119 (1) Arylsulfatase A and related enzymes [Inorganic ion transport and metabolism]
d COG3178 (1) Predicted phosphotransferase related to Ser/Thr protein kinases [General function prediction only]
u COG3185 (1) 4-hydroxyphenylpyruvate dioxygenase and related hemolysins [Amino acid transport and metabolism / General function prediction only]
D COG3209 (2) Rhs family protein [Cell envelope biogenesis, outer membrane]
D COG3347 (1) Uncharacterized conserved protein [Function unknown]
u d COG3349 (2) Uncharacterized conserved protein [Function unknown]
u COG3384 (2) Uncharacterized conserved protein [Function unknown]
u COG3449 (1) DNA gyrase inhibitor [DNA replication, recombination, and repair]
d COG3565 (1) Predicted dioxygenase of extradiol dioxygenase family [General function prediction only]
d COG3604 (1) Transcriptional regulator containing GAF, AAA-type ATPase, and DNA binding domains [Transcription / Signal transduction mechanisms]
d COG3637 (1) Opacity protein and related surface antigens [Cell envelope biogenesis, outer membrane]
d COG3659 (1) Carbohydrate-selective porin [Cell envelope biogenesis, outer membrane]
u COG3663 (1) G:T/U mismatch-specific DNA glycosylase [DNA replication, recombination, and repair]
uD COG3677 (2) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
D COG3695 (1) Predicted methylated DNA-protein cysteine methyltransferase [DNA replication, recombination, and repair]
u COG3800 (1) Predicted transcriptional regulator [General function prediction only]
d COG3842 (1) ABC-type spermidine/putrescine transport systems, ATPase components [Amino acid transport and metabolism]
uD COG4091 (2) Predicted homoserine dehydrogenase [Amino acid transport and metabolism]
u COG4126 (1) Hydantoin racemase [Amino acid transport and metabolism]
u d COG4147 (2) Predicted symporter [General function prediction only]
u COG4230 (1) Delta 1-pyrroline-5-carboxylate dehydrogenase [Energy production and conversion]
u COG4249 (1) Uncharacterized protein containing caspase domain [General function prediction only]
u d COG4257 (2) Streptogramin lyase [Defense mechanisms]
u COG4339 (1) Uncharacterized protein conserved in bacteria [Function unknown]
u d COG4391 (2) Uncharacterized protein conserved in bacteria [Function unknown]
u COG4392 (1) Predicted membrane protein [Function unknown]
d COG4628 (1) Uncharacterized conserved protein [Function unknown]
u d COG4630 (2) Xanthine dehydrogenase, iron-sulfur cluster and FAD-binding subunit A [Nucleotide transport and metabolism]
u d COG4638 (2) Phenylpropionate dioxygenase and related ring-hydroxylating dioxygenases, large terminal subunit [Inorganic ion transport and metabolism / General function prediction only]
D COG4663 (1) TRAP-type mannitol/chloroaromatic compound transport system, periplasmic component [Secondary metabolites biosynthesis, transport, and catabolism]
d COG4770 (1) Acetyl/propionyl-CoA carboxylase, alpha subunit [Lipid metabolism]
u COG4784 (1) Putative Zn-dependent protease [General function prediction only]
u d COG4786 (2) Flagellar basal body rod protein [Cell motility and secretion]
uD COG4787 (2) Flagellar basal body rod protein [Cell motility and secretion]
u COG4941 (1) Predicted RNA polymerase sigma factor containing a TPR repeat domain [Transcription]
D COG4977 (1) Transcriptional regulator containing an amidase domain and an AraC-type DNA-binding HTH domain [Transcription]
u d COG5010 (2) Flp pilus assembly protein TadD, contains TPR repeats [Intracellular trafficking and secretion]
u d COG5465 (2) Uncharacterized conserved protein [Function unknown]
d COG5569 (1) Uncharacterized conserved protein [Function unknown]
u COG5573 (1) Predicted nucleic-acid-binding protein, contains PIN domain [General function prediction only]
u COG5605 (1) Predicted small integral membrane protein [Function unknown]
d GOS0 (1) ?
u GOS10025 (1) ?
u d GOS1030 (2) ?
D GOS11062282 (1) ?

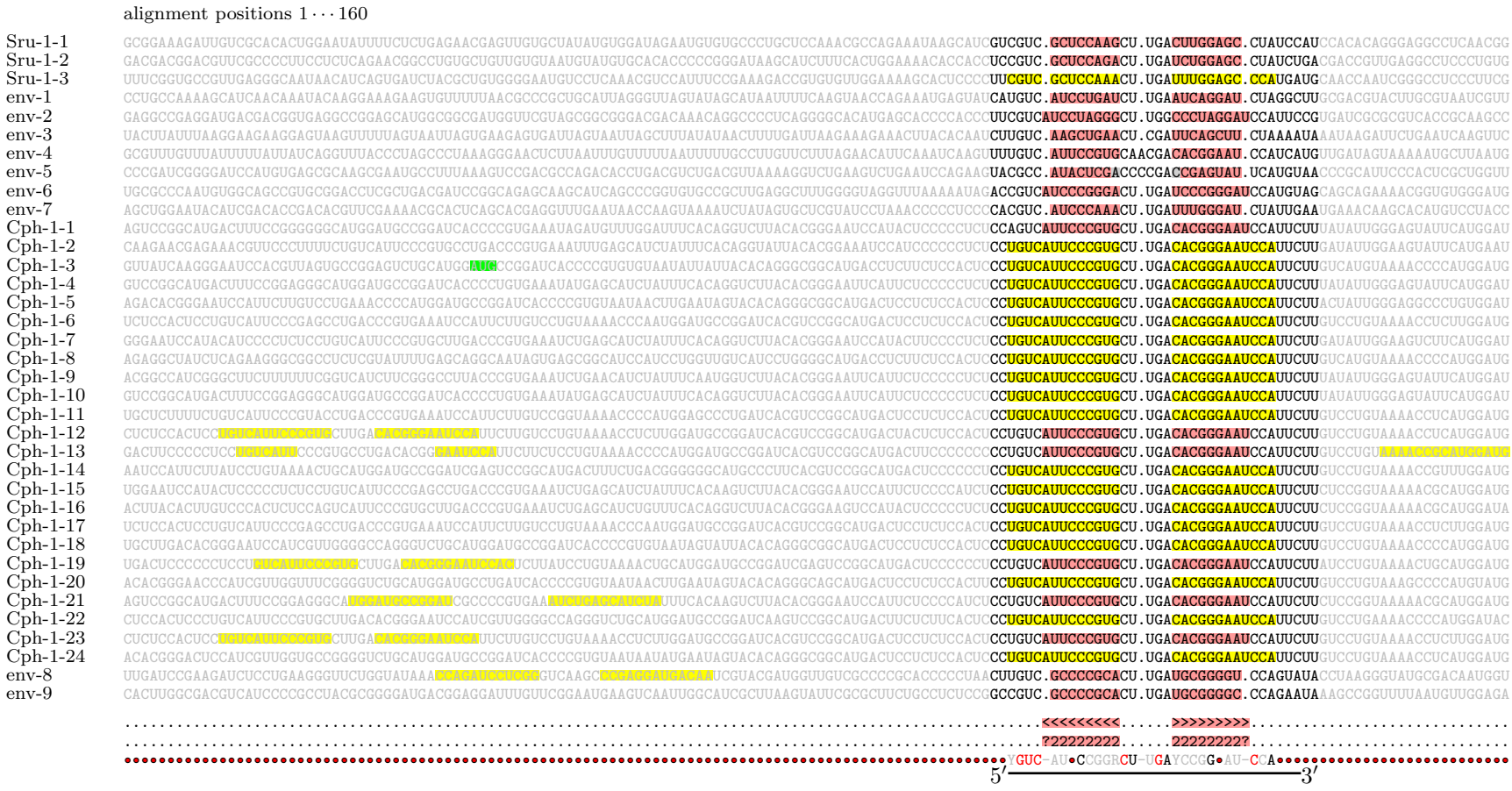
d GOS11112103 (1) ?
u GOS11127 (1) ?
uD GOS11233935 (2) ?
d GOS1133 (1) ?
u GOS1138 (1) ?
d GOS1168 (1) ?
u GOS11882 (1) ?
d GOS1240 (1) ?
D GOS1264 (1) ?
D GOS1275 (1) ?
d GOS1293 (1) ?
d GOS130 (1) ?
u GOS13034328 (1) ?
u GOS1308 (1) ?
d GOS13103195 (1) ?
D GOS1316 (1) ?
d GOS13198391 (1) ?
D GOS1336 (1) ?
u GOS1385 (1) ?
u GOS14402686 (1) ?
uD GOS1465 (2) ?
D GOS168 (1) ?
u GOS1720 (1) ?
u GOS1816 (1) ?
uD GOS192 (2) ?
uD GOS207 (2) ?
u GOS2296 (1) ?
u GOS241 (1) ?
d GOS2566 (1) ?
D GOS2593 (1) ?
D GOS26 (1) ?
u GOS2624 (1) ?
u GOS271 (1) ?
u GOS3198 (1) ?
d GOS3488 (1) ?
u d GOS3716 (2) ?
u d GOS3923 (2) ?
d GOS401 (1) ?
D GOS4122 (1) ?
u GOS4174 (1) ?
D GOS423 (1) ?
D GOS4366 (1) ?
u GOS4503 (1) ?
d GOS4645 (1) ?
u GOS4676 (1) ?
D GOS4735 (1) ?
D GOS4853 (1) ?
u GOS4865 (1) ?
d GOS490 (1) ?
u GOS501 (1) ?
u GOS5149 (1) ?
u GOS530 (1) ?
u GOS5342 (1) ?
u GOS5458 (1) ?
uD GOS5495 (2) ?
u GOS5684 (1) ?
D GOS5716 (1) ?
u GOS5783 (1) ?
u d GOS5834 (2) ?
u GOS5950 (1) ?
u d GOS6241 (2) ?
u GOS625 (1) ?
D GOS6270 (1) ?
u GOS644 (1) ?
d GOS684 (1) ?
u d GOS7235 (2) ?
d GOS7357 (1) ?
u GOS7589 (1) ?
u GOS7695 (1) ?
u GOS8597 (1) ?
u d GOS8607 (2) ?
D GOS8865 (1) ?
D GOS8926 (1) ?
uD GOS899 (2) ?
d GOS9147 (1) ?
u GOS9318 (1) ?
D GOS9409 (1) ?
u d GOS987 (2) ?
D GOS997 (1) ?
u NOG05970 (1)
u d NOG08005 (2)
d NOG13350 (1)
D NOG14032 (1)
d NOG14632 (1)
u NOG38593 (1)
uD pfam00083 (2) Sugar (and other) transporter.
uD pfam00106 (4) short chain dehydrogenase.
u d pfam00144 (3) Beta-lactamase.
D pfam00150 (1) Cellulase (glycosyl hydrolase family 5).
u pfam00231 (1) ATP synthase.
d pfam00293 (1) NUDIX domain.
d pfam00342 (1) Phosphoglucose isomerase.
D pfam00440 (1) Bacterial regulatory proteins, tetR family.
d pfam00453 (1) Ribosomal protein L20.
u pfam00461 (1) Signal peptidase I.
D pfam00496 (1) Bacterial extracellular solute-binding proteins, family 5.
u pfam00575 (1) S1 RNA binding domain.
d pfam00582 (1) Universal stress protein family.
uDd pfam00583 (3) Acetyltransferase (GNAT) family.
u pfam00639 (1) PPIC-type PPIASE domain.
D pfam00687 (1) Ribosomal protein L1p/L10e family.
D pfam00690 (1) Cation transporter/ATPase, N-terminus.
d pfam00725 (1) 3-hydroxyacyl-CoA dehydrogenase, C-terminal domain.
D pfam00815 (1) Histidinol dehydrogenase.

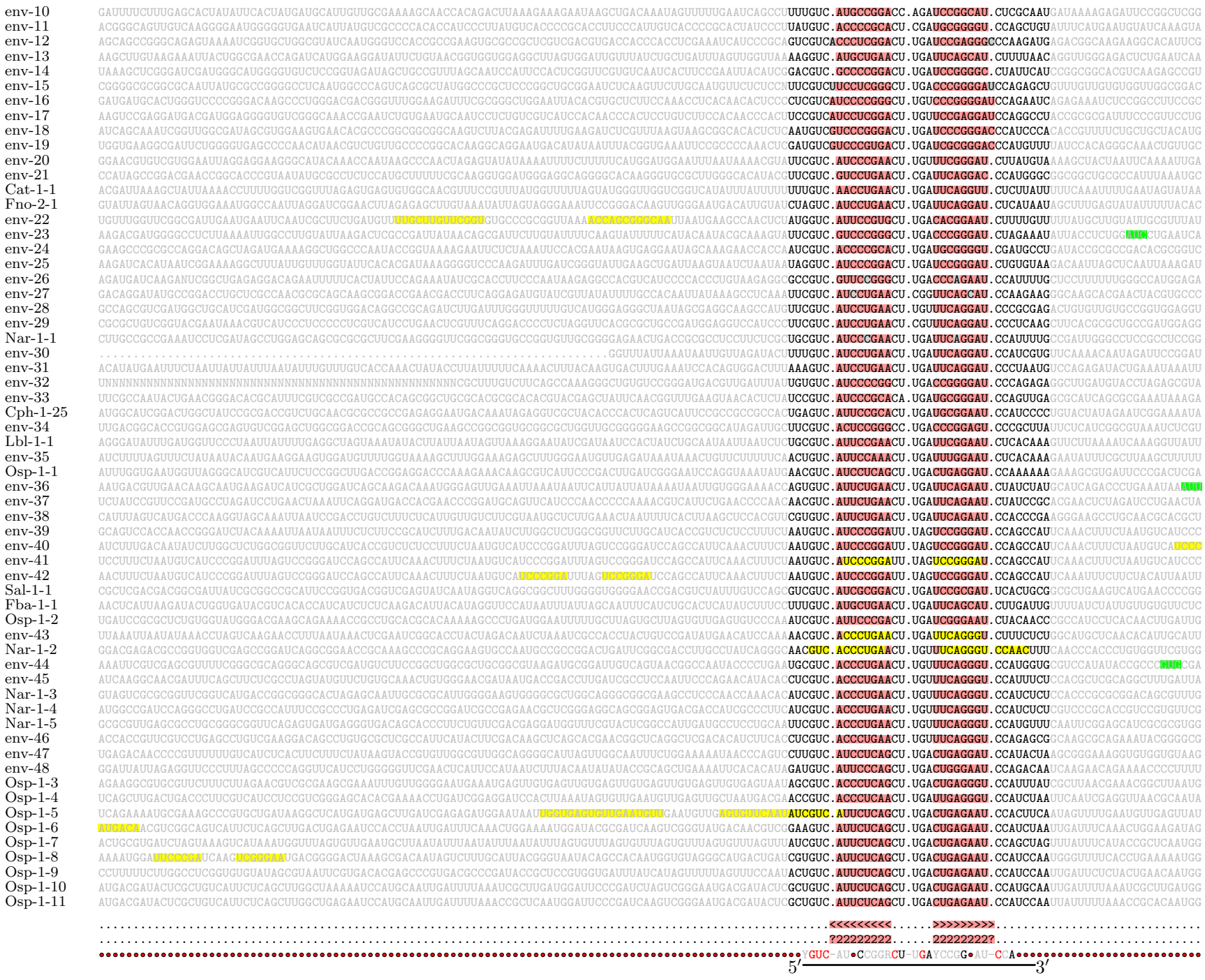
d pfam00829 (1) Ribosomal prokaryotic L21 protein.
uD pfam00892 (2) Integral membrane protein DUF6.
u d pfam00903 (2) Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily.
u d pfam00908 (2) dTDP-4-dehydrorhamnose 3,5-epimerase.
u pfam00912 (1) Transglycosylase.
u pfam01016 (1) Ribosomal L27 protein.
u pfam01047 (1) MarR family.
u pfam01073 (1) 3-beta hydroxysteroid dehydrogenase/isomerase family.
d pfam01112 (1) Asparaginase.
u d pfam01124 (2) MAPEG family.
d pfam01132 (1) Elongation factor P (EF-P).
d pfam01148 (1) Cytidylyltransferase family.
u d pfam01170 (2) Putative RNA methylase family UPF0020.
u d pfam01257 (2) Respiratory-chain NADH dehydrogenase 24 Kd subunit.
u d pfam01425 (2) Amidase.
u pfam01522 (1) Polysaccharide deacetylase.
u pfam01592 (1) NifU-like N terminal domain.
uD pfam01609 (2) Transposase DDE domain.
u pfam01610 (1) Transposase.
uD pfam01726 (2) LexA DNA binding domain.
u d pfam01738 (2) Dienelactone hydrolase family.
u d pfam01784 (2) NIF3 (NGG1p interacting factor 3).
u d pfam01808 (2) AICARFT/IMPCHase bienzyme.
d pfam01883 (1) Domain of unknown function DUF59.
d pfam01887 (1) Protein of unknown function DUF62.
uD pfam01909 (2) Nucleotidyltransferase domain.
D pfam01960 (1) ArgJ family.
u pfam01966 (1) HD domain.
D pfam02080 (1) TrkA-C domain.
u d pfam02142 (2) MGS-like domain.
d pfam02211 (1) Nitrile hydratase beta subunit.
d pfam02272 (1) DHHA1 domain.
u pfam02277 (1) Phosphoribosyltransferase.
u pfam02371 (1) Transposase IS116/IS110/IS902 family.
u pfam02463 (1) RecF/RecN/SMC N terminal domain.
u pfam02483 (1) SMC family, C-terminal domain.
D pfam02518 (1) Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase.
d pfam02559 (1) CarD-like/TRCF domain.
d pfam02575 (1) Uncharacterised BCR, YbaB family COG0718.
uD pfam02615 (2) Malate/L-lactate dehydrogenase.
d pfam02696 (1) Uncharacterized ACR, YdiU/UPF0061 family.
d pfam02738 (1) Aldehyde oxidase and xanthine dehydrogenase, molybdopterin binding domain.
d pfam02769 (1) AIR synthase related protein, C-terminal domain.
d pfam02900 (1) Catalytic LigB subunit of aromatic ring-opening dioxygenase.
u pfam02913 (1) FAD linked oxidases, C-terminal domain.
d pfam03060 (1) 2-nitropropane dioxygenase.
D pfam03331 (1) UDP-3-O-acyl N-acetylglycosamine deacetylase.
d pfam03379 (1) CcmB protein.
D pfam03401 (1) Bordetella uptake gene (bug) product.
u pfam03481 (1) SUA5 domain.
d pfam03595 (1) C4-dicarboxylate transporter/malic acid transport protein.
uDd pfam03737 (3) Demethylmenaquinone methyltransferase.
u pfam03880 (2) DbpA RNA binding domain.
d pfam03942 (1) DTW domain.
D pfam04011 (1) LemA family.
d pfam04169 (1) Domain of unknown function (DUF404).
u pfam04301 (1) Protein of unknown function (DUF452).
u pfam04760 (1) Translation initiation factor IF-2, N-terminal region.
uD pfam04909 (2) Amidohydrolase.
d pfam04932 (1) O-Antigen Polymerase.
d pfam04972 (1) Putative phospholipid-binding domain.
D pfam05016 (1) Plasmid stabilisation system protein.
d pfam05226 (1) CHASE2 domain.
u pfam05496 (1) Holliday junction DNA helicase ruvB N-terminus.
d pfam05597 (1) Poly(hydroxyalcanoate) granule associated protein (phasin).
D pfam05635 (1) S23 ribosomal protein.
u d pfam05721 (2) Phytanoyl-CoA dioxygenase (PhyH).
u d pfam05724 (2) Thiopurine S-methyltransferase (TPMT).
D pfam05899 (1) Protein of unknown function (DUF861).
u pfam05971 (1) Protein of unknown function (DUF890).
d pfam06114 (1) Domain of unknown function (DUF955).
d pfam06181 (1) Protein of unknown function (DUF989).
D pfam06764 (1) Protein of unknown function (DUF1223).
d pfam06779 (1) Protein of unknown function (DUF1228).
u pfam06897 (1) Protein of unknown function (DUF1269).
u d pfam07044 (2) Protein of unknown function (DUF1329).
d pfam07075 (1) Protein of unknown function (DUF1343).
d pfam07143 (1) Hydroxyneurosporene synthase (CrtC).
D pfam07366 (1) Protein of unknown function (DUF1486).
u smart00091 (1) PAS domain; PAS motifs appear in archaea, eubacteria and eukarya.
u d smart00283 (2) Methyl-accepting chemotaxis-like domains (chemotaxis sensory transducer).
u smart00287 (1) Bacterial SH3 domain homologues;
u smart00342 (1) helix_turn_helix, arabinose operon control protein;
u smart00347 (1) helix_turn_helix multiple antibiotic resistance protein;
D smart00418 (1) helix_turn_helix, Arsenical Resistance Operon Repressor;
u d smart00479 (2) exonuclease domain in DNA-polymerase alpha and epsilon chain, ribonuclease T and other exonucleases
D smart00487 (1) DEAD-like helicases superfamily;
u d smart00642 (4) Alpha-amylase domain;
u smart00649 (1) Ribosomal protein L11/L12;
u smart00732 (1) Likely ribonuclease with RNase H fold.

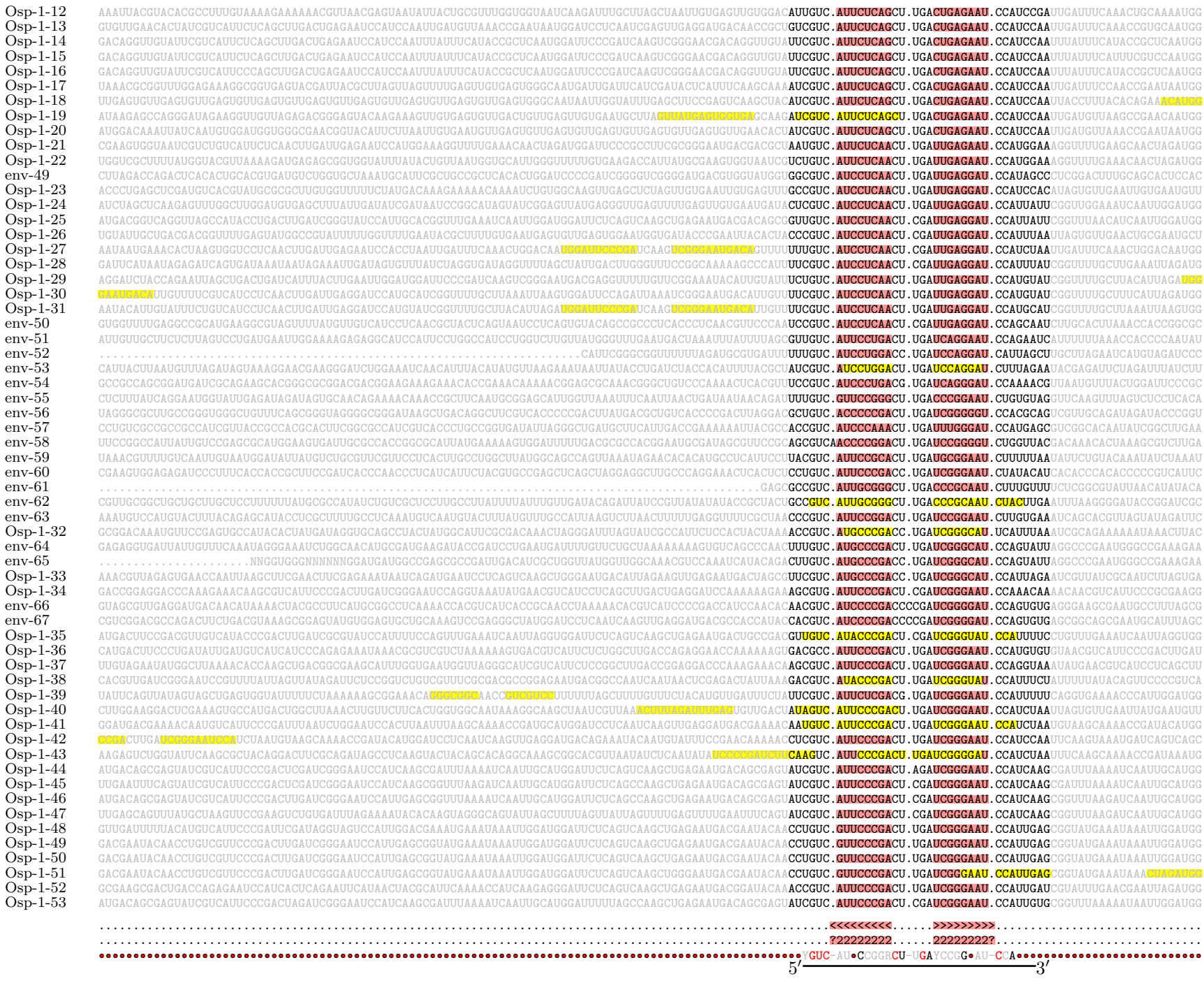
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Cph-1-79

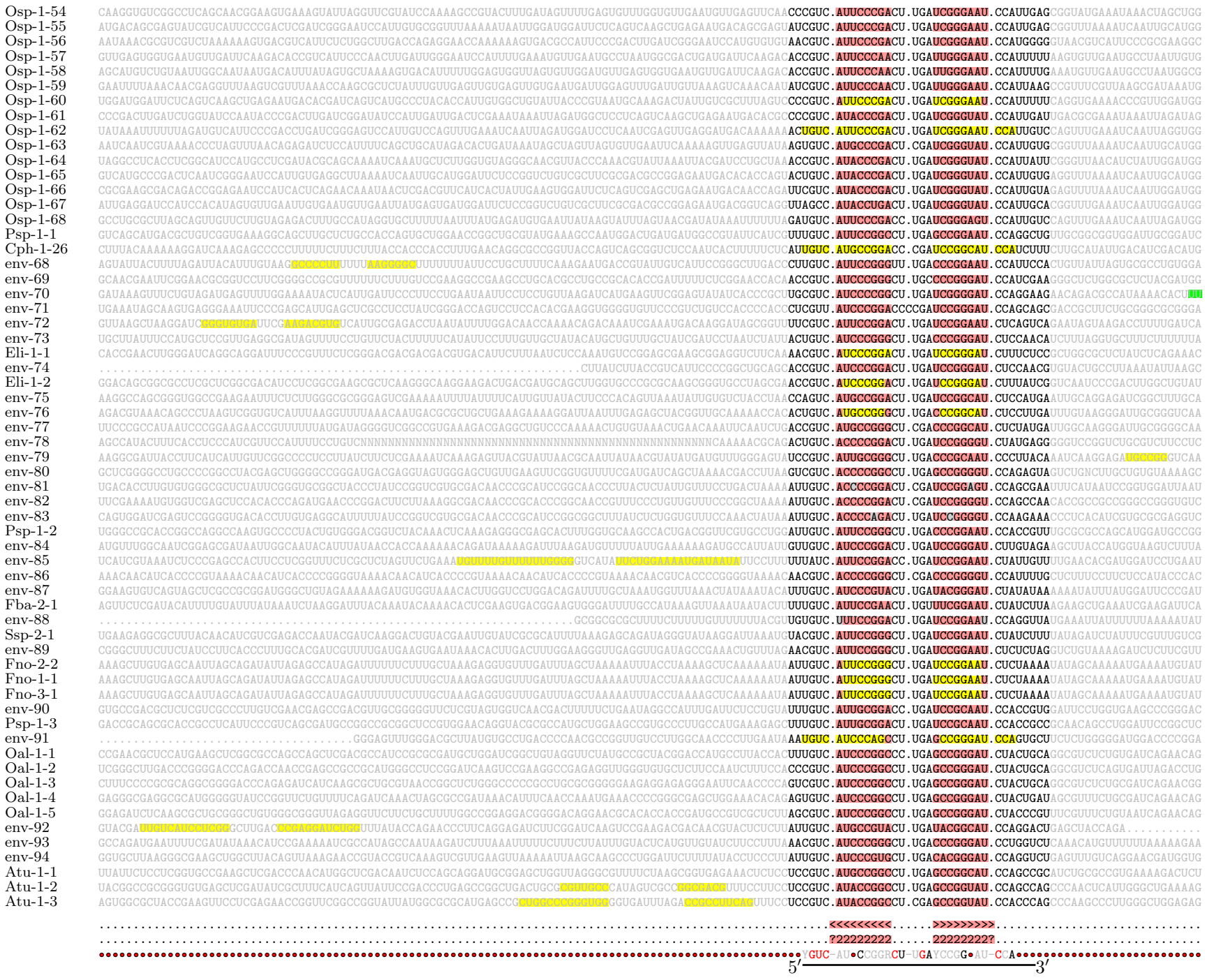
38.4 Multiple-sequence alignment

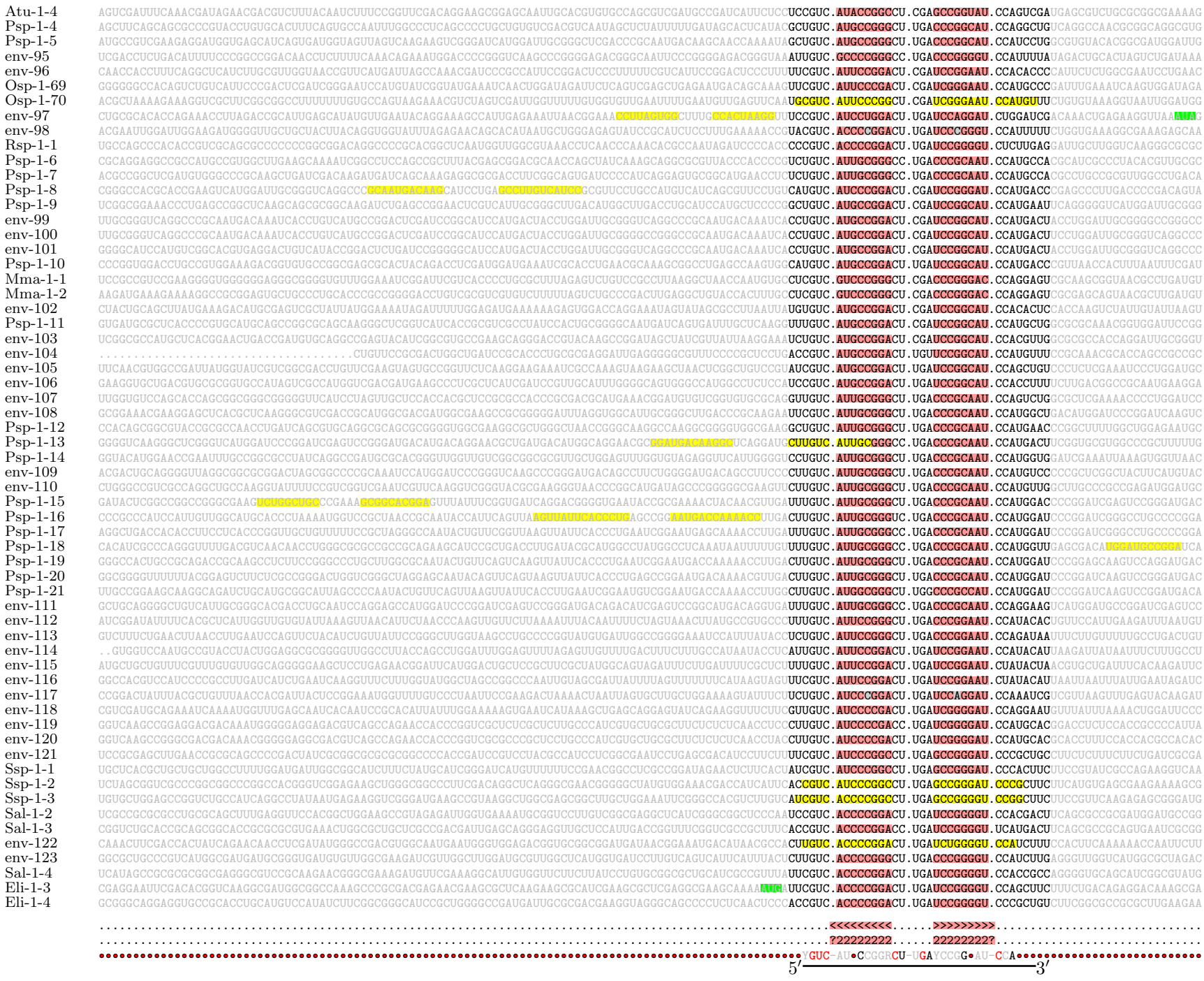
Each GUCCY-hairpin RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 38.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The GUCCY-hairpin RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

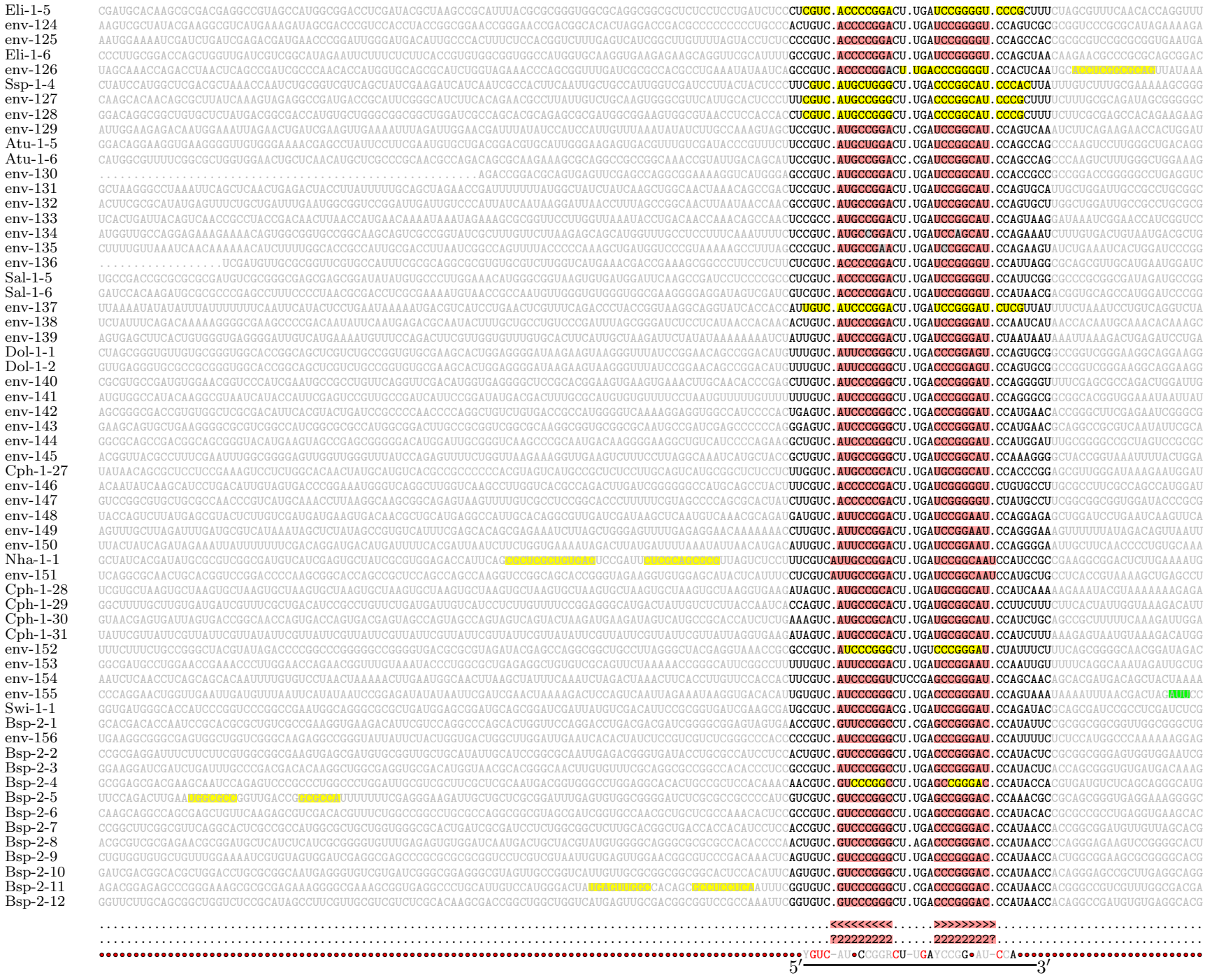


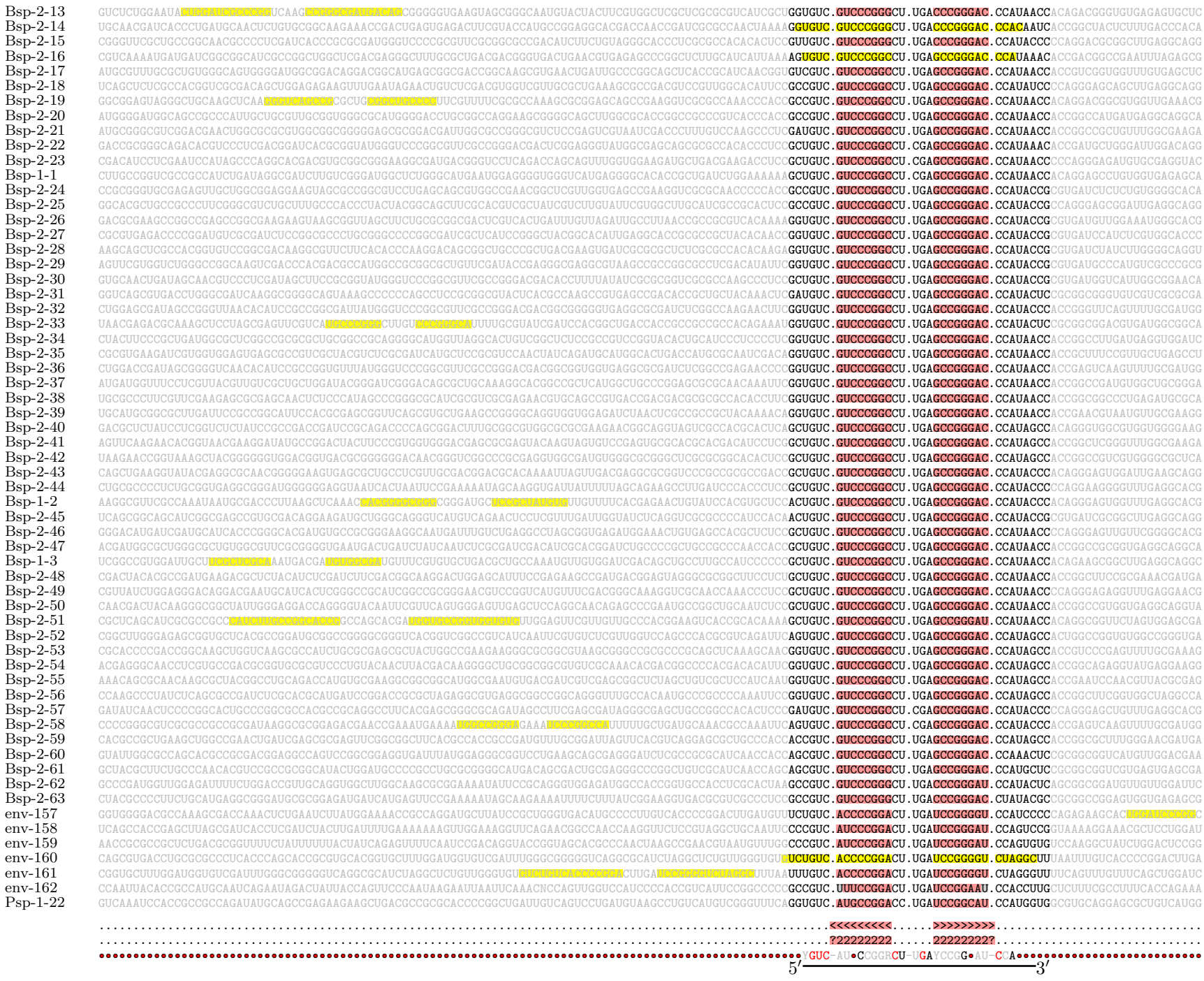

env-10
env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
Cat-1-1
Fno-2-1
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
Nar-1-1
env-30
env-31
env-32
env-33
Cph-1-25
env-34
Lbl-1-1
env-35
Osp-1-1
env-36
env-37
env-38
env-39
env-40
env-41
env-42
Sal-1-1
Fba-1-1
Osp-1-2
env-43
Nar-1-2
env-44
env-45
Nar-1-3
Nar-1-4
Nar-1-5
env-46
env-47
env-48
Osp-1-3
Osp-1-4
Osp-1-5
Osp-1-6
Osp-1-7
Osp-1-8
Osp-1-9
Osp-1-10
Osp-1-11
5′
3′

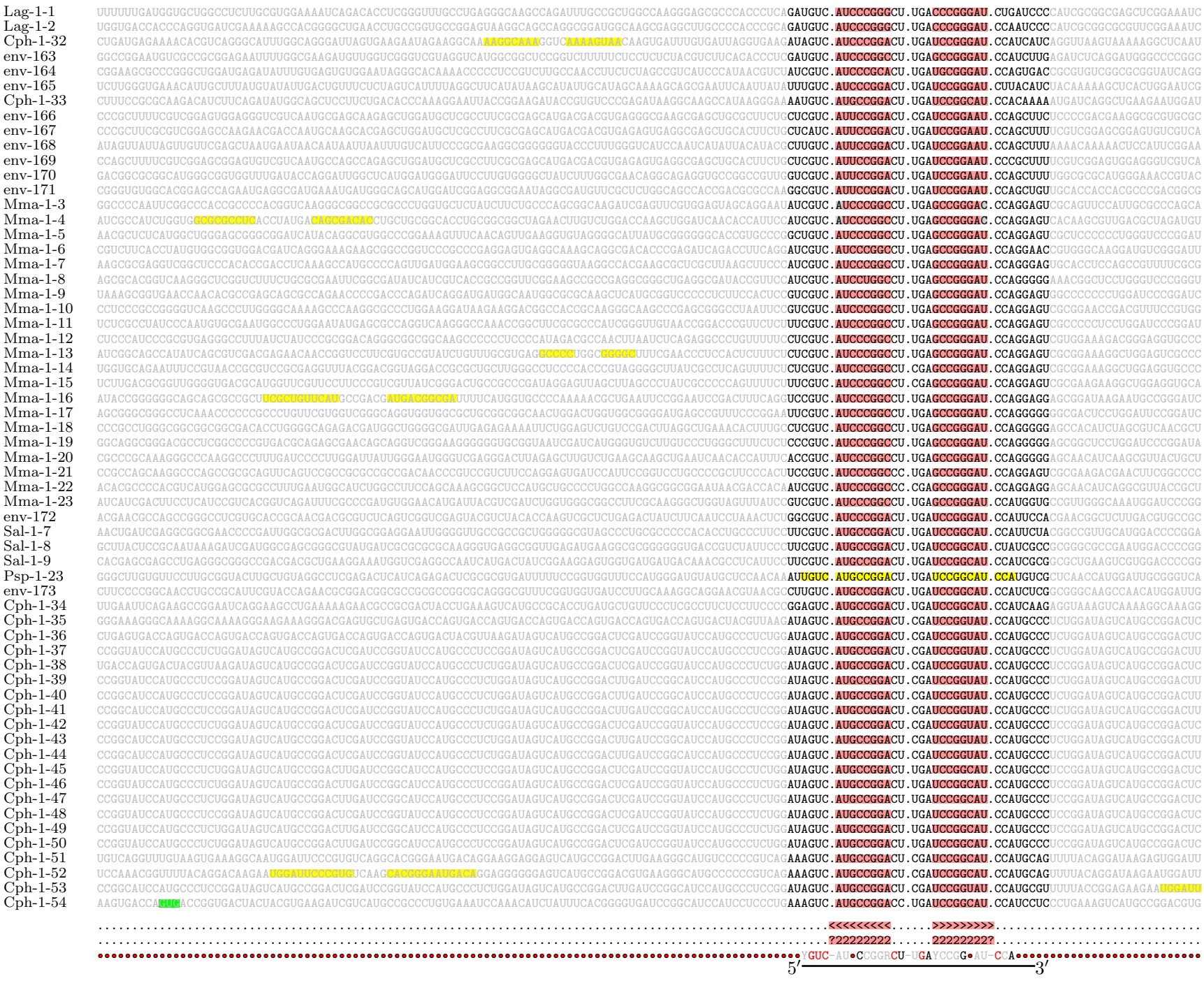

Osp-1-12
Osp-1-13
Osp-1-14
Osp-1-15
Osp-1-16
Osp-1-17
Osp-1-18
Osp-1-19
Osp-1-20
Osp-1-21
Osp-1-22
env-49
Osp-1-23
Osp-1-24
Osp-1-25
Osp-1-26
Osp-1-27
Osp-1-28
Osp-1-29
Osp-1-30
Osp-1-31
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
Osp-1-32
env-64
env-65
Osp-1-33
Osp-1-34
env-66
env-67
Osp-1-35
Osp-1-36
Osp-1-37
Osp-1-38
Osp-1-39
Osp-1-40
Osp-1-41
Osp-1-42
Osp-1-43
Osp-1-44
Osp-1-45
Osp-1-46
Osp-1-47
Osp-1-48
Osp-1-49
Osp-1-50
Osp-1-51
Osp-1-52
Osp-1-53
5′
3′

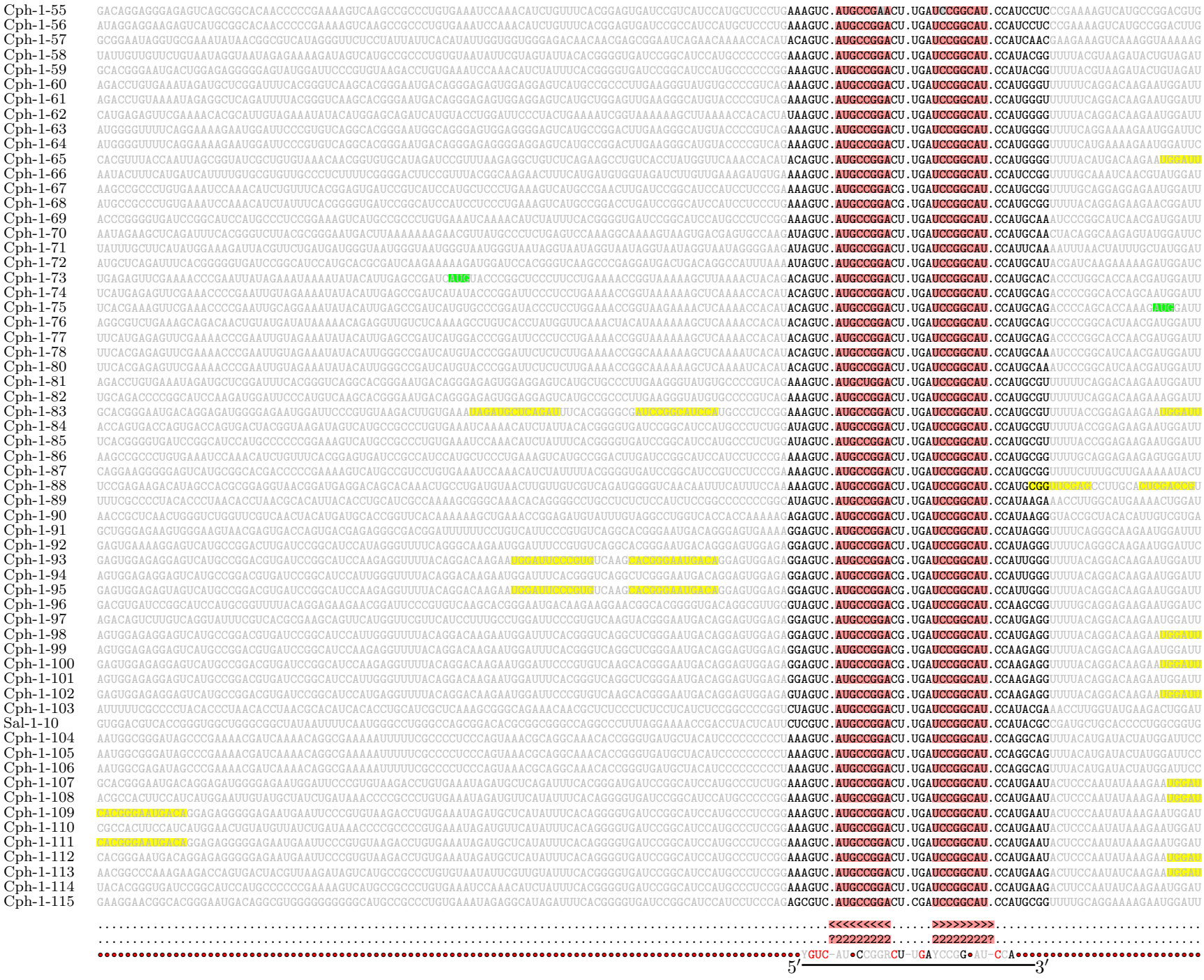

Osp-1-54
Osp-1-55
Osp-1-56
Osp-1-57
Osp-1-58
Osp-1-59
Osp-1-60
Osp-1-61
Osp-1-62
Osp-1-63
Osp-1-64
Osp-1-65
Osp-1-66
Osp-1-67
Osp-1-68
Psp-1-1
Cph-1-26
env-68
env-69
env-70
env-71
env-72
env-73
Eli-1-1
env-74
Eli-1-2
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
Psp-1-2
env-84
env-85
env-86
env-87
Fba-2-1
env-88
Ssp-2-1
env-89
Fno-2-2
Fno-1-1
Fno-3-1
env-90
Psp-1-3
env-91
Oal-1-1
Oal-1-2
Oal-1-3
Oal-1-4
Oal-1-5
env-92
env-93
env-94
Atu-1-1
Atu-1-2
Atu-1-3
5′
3′

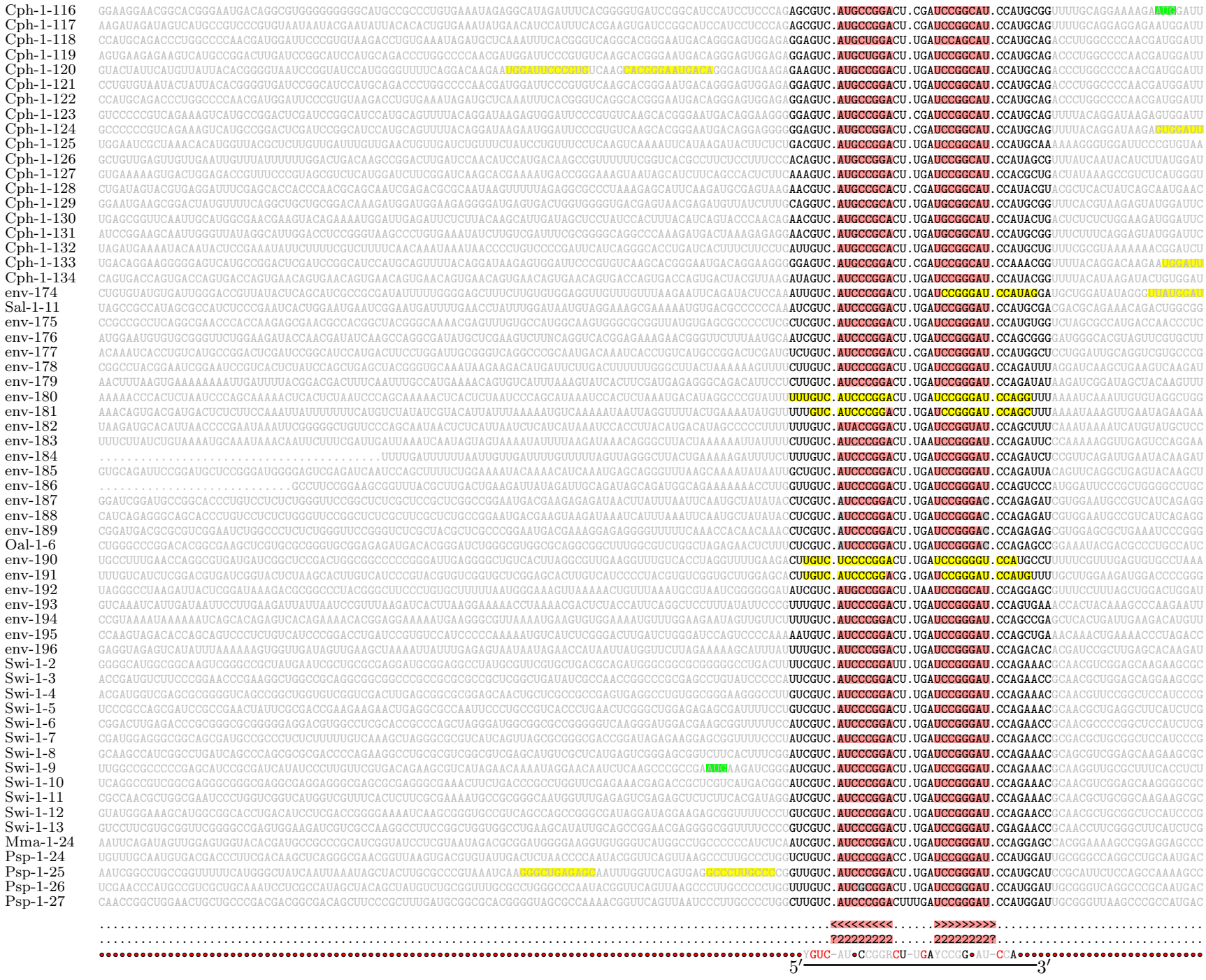

Atu-1-4
Psp-1-4
Psp-1-5
env-95
env-96
Osp-1-69
Osp-1-70
env-97
env-98
Rsp-1-1
Psp-1-6
Psp-1-7
Psp-1-8
Psp-1-9
env-99
env-100
env-101
Psp-1-10
Mma-1-1
Mma-1-2
env-102
Psp-1-11
env-103
env-104
env-105
env-106
env-107
env-108
Psp-1-12
Psp-1-13
Psp-1-14
env-109
env-110
Psp-1-15
Psp-1-16
Psp-1-17
Psp-1-18
Psp-1-19
Psp-1-20
Psp-1-21
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
Ssp-1-1
Ssp-1-2
Ssp-1-3
Sal-1-2
Sal-1-3
env-122
env-123
Sal-1-4
Eli-1-3
Eli-1-4
5′
3′

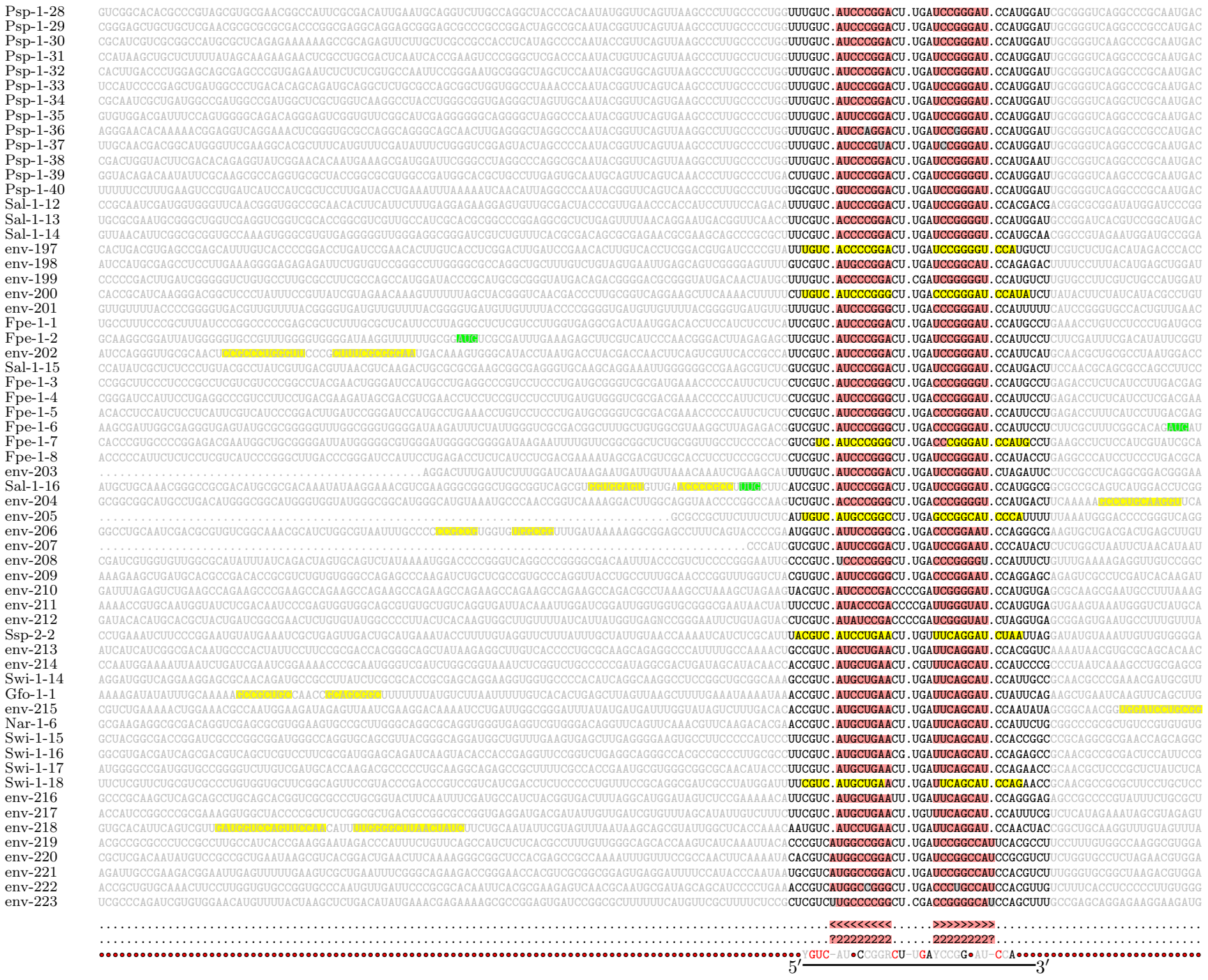

Eli-1-5
env-124
env-125
Eli-1-6
env-126
Ssp-1-4
env-127
env-128
env-129
Atu-1-5
Atu-1-6
env-130
env-131
env-132
env-133
env-134
env-135
env-136
Sal-1-5
Sal-1-6
env-137
env-138
env-139
Dol-1-1
Dol-1-2
env-140
env-141
env-142
env-143
env-144
env-145
Cph-1-27
env-146
env-147
env-148
env-149
env-150
Nha-1-1
env-151
Cph-1-28
Cph-1-29
Cph-1-30
Cph-1-31
env-152
env-153
env-154
env-155
Swi-1-1
Bsp-2-1
env-156
Bsp-2-2
Bsp-2-3
Bsp-2-4
Bsp-2-5
Bsp-2-6
Bsp-2-7
Bsp-2-8
Bsp-2-9
Bsp-2-10
Bsp-2-11
Bsp-2-12
5′
3′

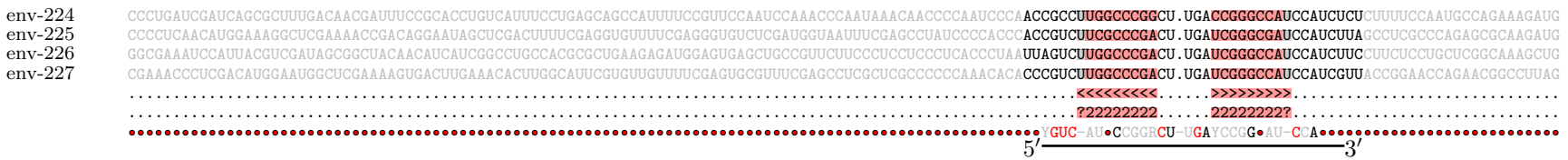

Bsp-2-13
Bsp-2-14
Bsp-2-15
Bsp-2-16
Bsp-2-17
Bsp-2-18
Bsp-2-19
Bsp-2-20
Bsp-2-21
Bsp-2-22
Bsp-2-23
Bsp-1-1
Bsp-2-24
Bsp-2-25
Bsp-2-26
Bsp-2-27
Bsp-2-28
Bsp-2-29
Bsp-2-30
Bsp-2-31
Bsp-2-32
Bsp-2-33
Bsp-2-34
Bsp-2-35
Bsp-2-36
Bsp-2-37
Bsp-2-38
Bsp-2-39
Bsp-2-40
Bsp-2-41
Bsp-2-42
Bsp-2-43
Bsp-2-44
Bsp-1-2
Bsp-2-45
Bsp-2-46
Bsp-2-47
Bsp-1-3
Bsp-2-48
Bsp-2-49
Bsp-2-50
Bsp-2-51
Bsp-2-52
Bsp-2-53
Bsp-2-54
Bsp-2-55
Bsp-2-56
Bsp-2-57
Bsp-2-58
Bsp-2-59
Bsp-2-60
Bsp-2-61
Bsp-2-62
Bsp-2-63
env-157
env-158
env-159
env-160
env-161
env-162
Psp-1-22
5′
3′


Lag-1-1
Lag-1-2
Cph-1-32
env-163
env-164
env-165
Cph-1-33
env-166
env-167
env-168
env-169
env-170
env-171
Mma-1-3
Mma-1-4
Mma-1-5
Mma-1-6
Mma-1-7
Mma-1-8
Mma-1-9
Mma-1-10
Mma-1-11
Mma-1-12
Mma-1-13
Mma-1-14
Mma-1-15
Mma-1-16
Mma-1-17
Mma-1-18
Mma-1-19
Mma-1-20
Mma-1-21
Mma-1-22
Mma-1-23
env-172
Sal-1-7
Sal-1-8
Sal-1-9
Psp-1-23
env-173
Cph-1-34
Cph-1-35
Cph-1-36
Cph-1-37
Cph-1-38
Cph-1-39
Cph-1-40
Cph-1-41
Cph-1-42
Cph-1-43
Cph-1-44
Cph-1-45
Cph-1-46
Cph-1-47
Cph-1-48
Cph-1-49
Cph-1-50
Cph-1-51
Cph-1-52
Cph-1-53
Cph-1-54
5′
3′

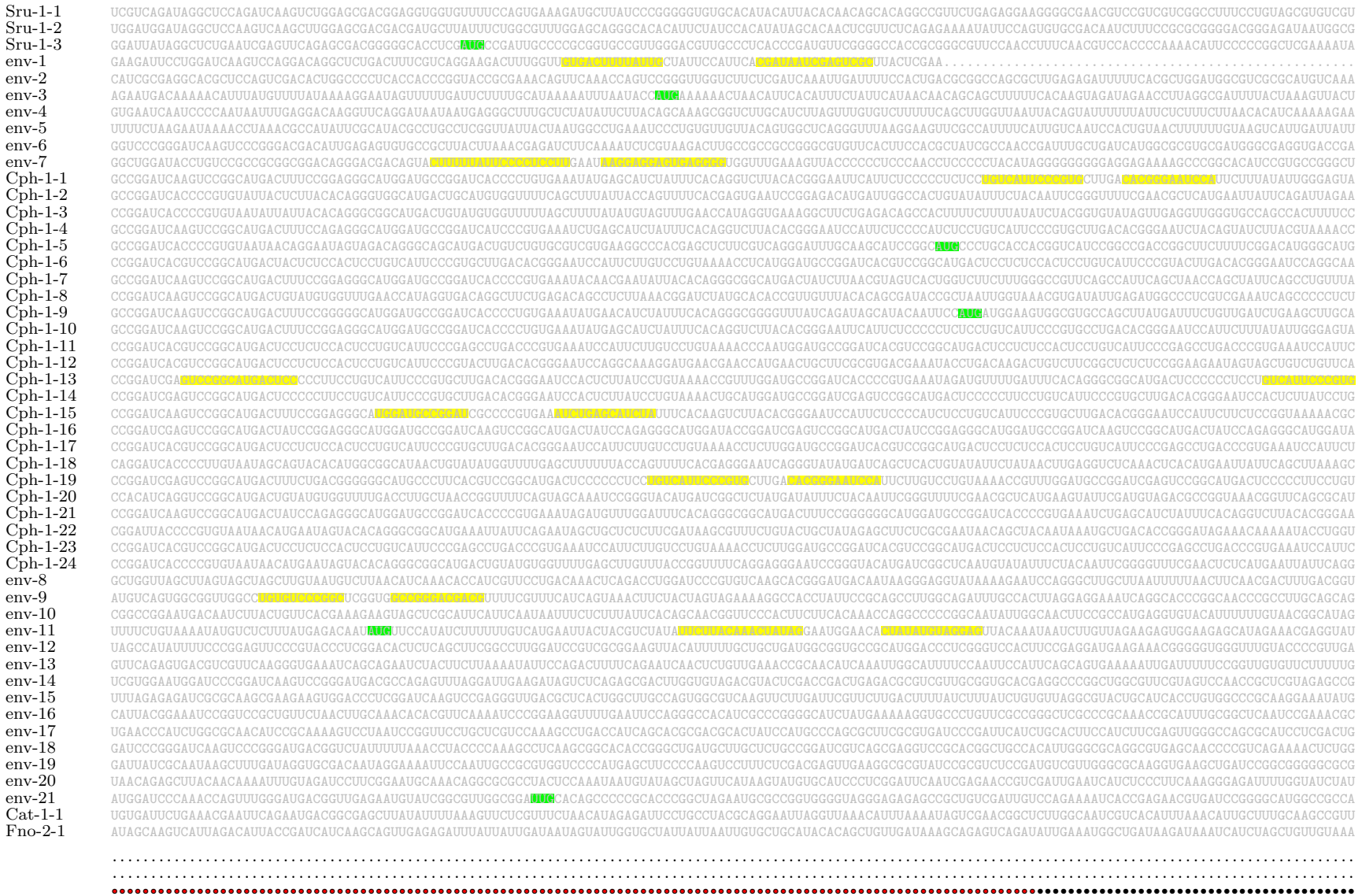

Cph-1-55
Cph-1-56
Cph-1-57
Cph-1-58
Cph-1-59
Cph-1-60
Cph-1-61
Cph-1-62
Cph-1-63
Cph-1-64
Cph-1-65
Cph-1-66
Cph-1-67
Cph-1-68
Cph-1-69
Cph-1-70
Cph-1-71
Cph-1-72
Cph-1-73
Cph-1-74
Cph-1-75
Cph-1-76
Cph-1-77
Cph-1-78
Cph-1-80
Cph-1-81
Cph-1-82
Cph-1-83
Cph-1-84
Cph-1-85
Cph-1-86
Cph-1-87
Cph-1-88
Cph-1-89
Cph-1-90
Cph-1-91
Cph-1-92
Cph-1-93
Cph-1-94
Cph-1-95
Cph-1-96
Cph-1-97
Cph-1-98
Cph-1-99
Cph-1-100
Cph-1-101
Cph-1-102
Cph-1-103
Sal-1-10
Cph-1-104
Cph-1-105
Cph-1-106
Cph-1-107
Cph-1-108
Cph-1-109
Cph-1-110
Cph-1-111
Cph-1-112
Cph-1-113
Cph-1-114
Cph-1-115
5′
3′

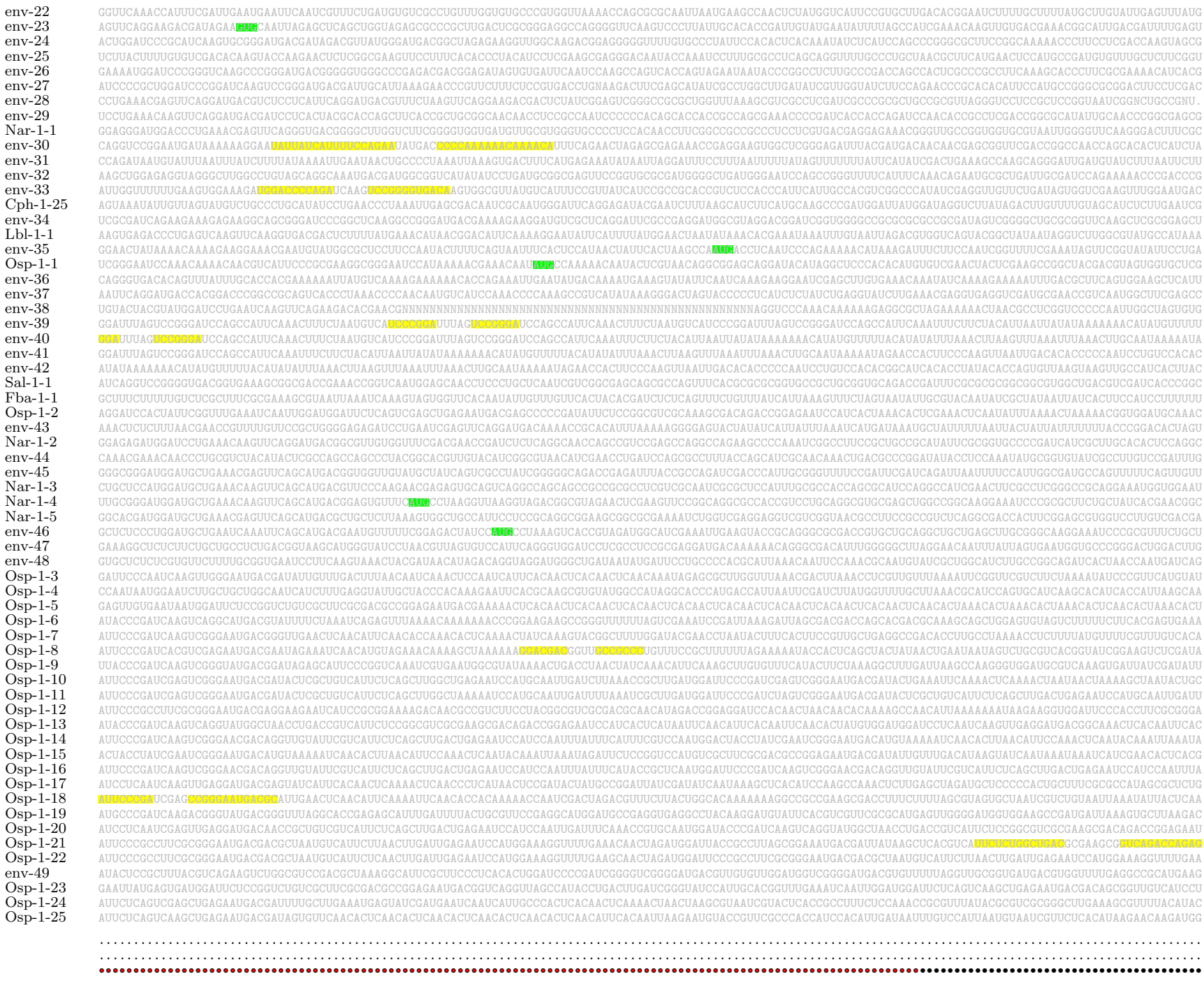

Cph-1-116
Cph-1-117
Cph-1-118
Cph-1-119
Cph-1-120
Cph-1-121
Cph-1-122
Cph-1-123
Cph-1-124
Cph-1-125
Cph-1-126
Cph-1-127
Cph-1-128
Cph-1-129
Cph-1-130
Cph-1-131
Cph-1-132
Cph-1-133
Cph-1-134
env-174
Sal-1-11
env-175
env-176
env-177
env-178
env-179
env-180
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
Oal-1-6
env-190
env-191
env-192
env-193
env-194
env-195
env-196
Swi-1-2
Swi-1-3
Swi-1-4
Swi-1-5
Swi-1-6
Swi-1-7
Swi-1-8
Swi-1-9
Swi-1-10
Swi-1-11
Swi-1-12
Swi-1-13
Mma-1-24
Psp-1-24
Psp-1-25
Psp-1-26
Psp-1-27
5′
3′

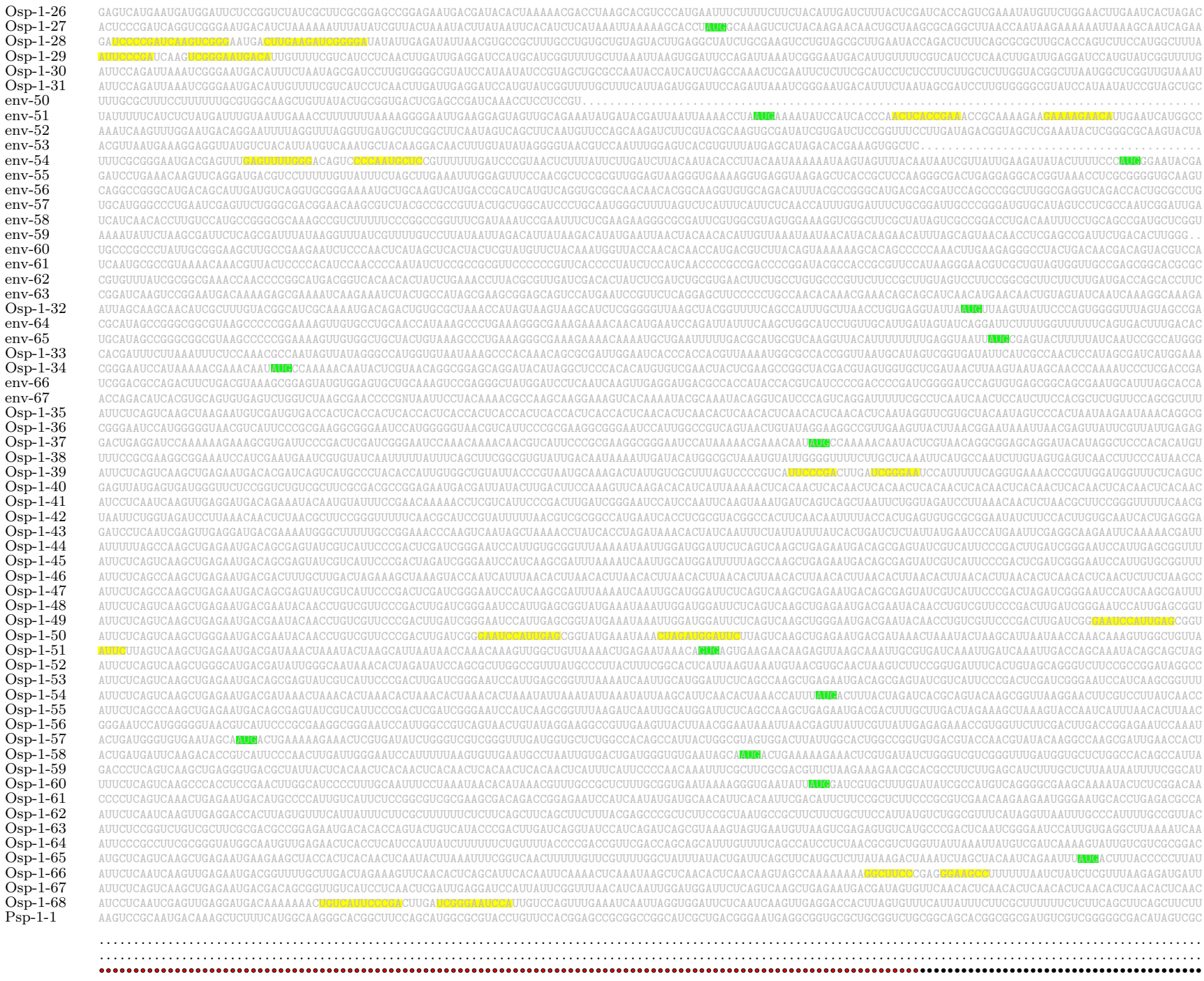

Psp-1-28
Psp-1-29
Psp-1-30
Psp-1-31
Psp-1-32
Psp-1-33
Psp-1-34
Psp-1-35
Psp-1-36
Psp-1-37
Psp-1-38
Psp-1-39
Psp-1-40
Sal-1-12
Sal-1-13
Sal-1-14
env-197
env-198
env-199
env-200
env-201
Fpe-1-1
Fpe-1-2
env-202
Sal-1-15
Fpe-1-3
Fpe-1-4
Fpe-1-5
Fpe-1-6
Fpe-1-7
Fpe-1-8
env-203
Sal-1-16
env-204
env-205
env-206
env-207
env-208
env-209
env-210
env-211
env-212
Ssp-2-2
env-213
env-214
Swi-1-14
Gfo-1-1
env-215
Nar-1-6
Swi-1-15
Swi-1-16
Swi-1-17
Swi-1-18
env-216
env-217
env-218
env-219
env-220
env-221
env-222
env-223
5′
3′



alignment positions 161 · · · 320

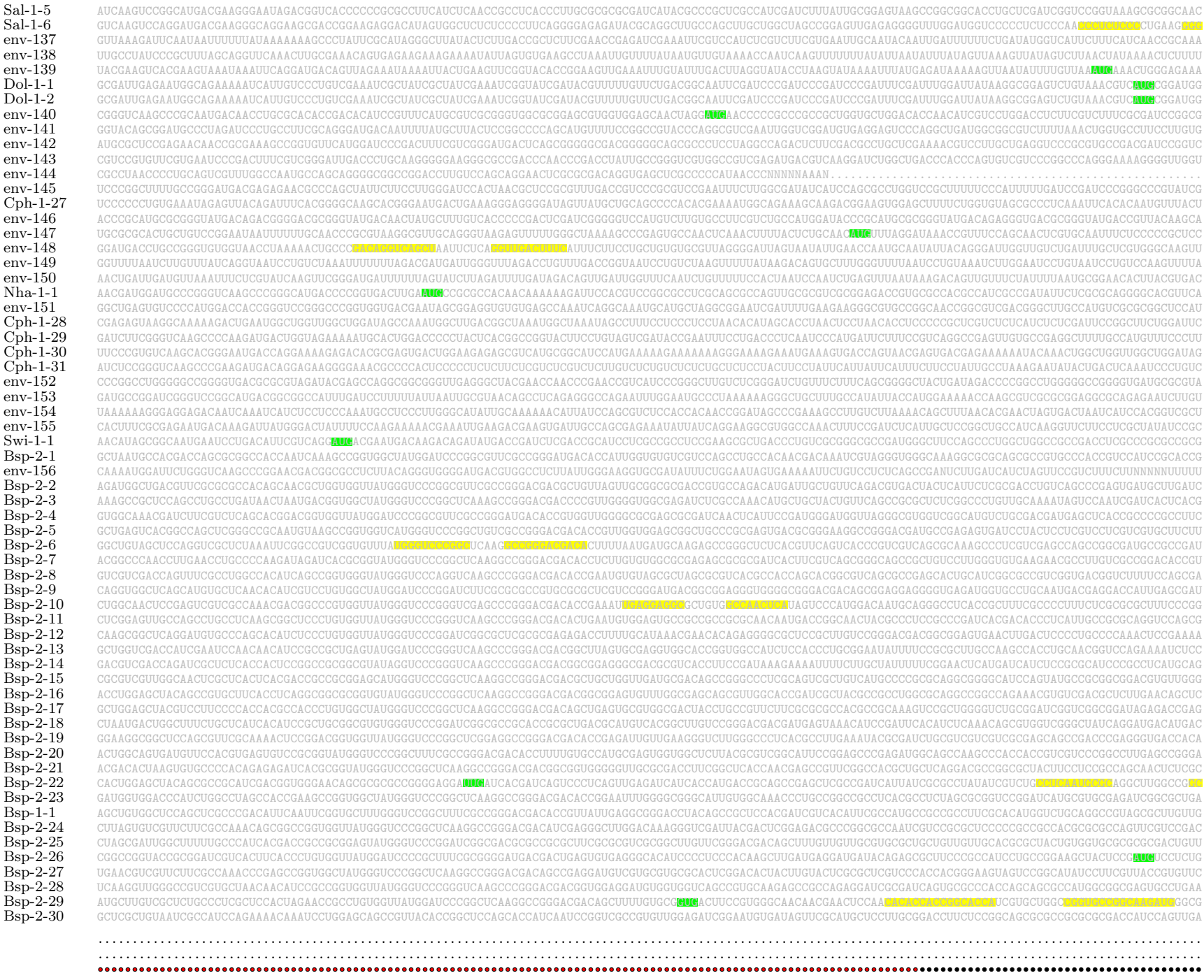


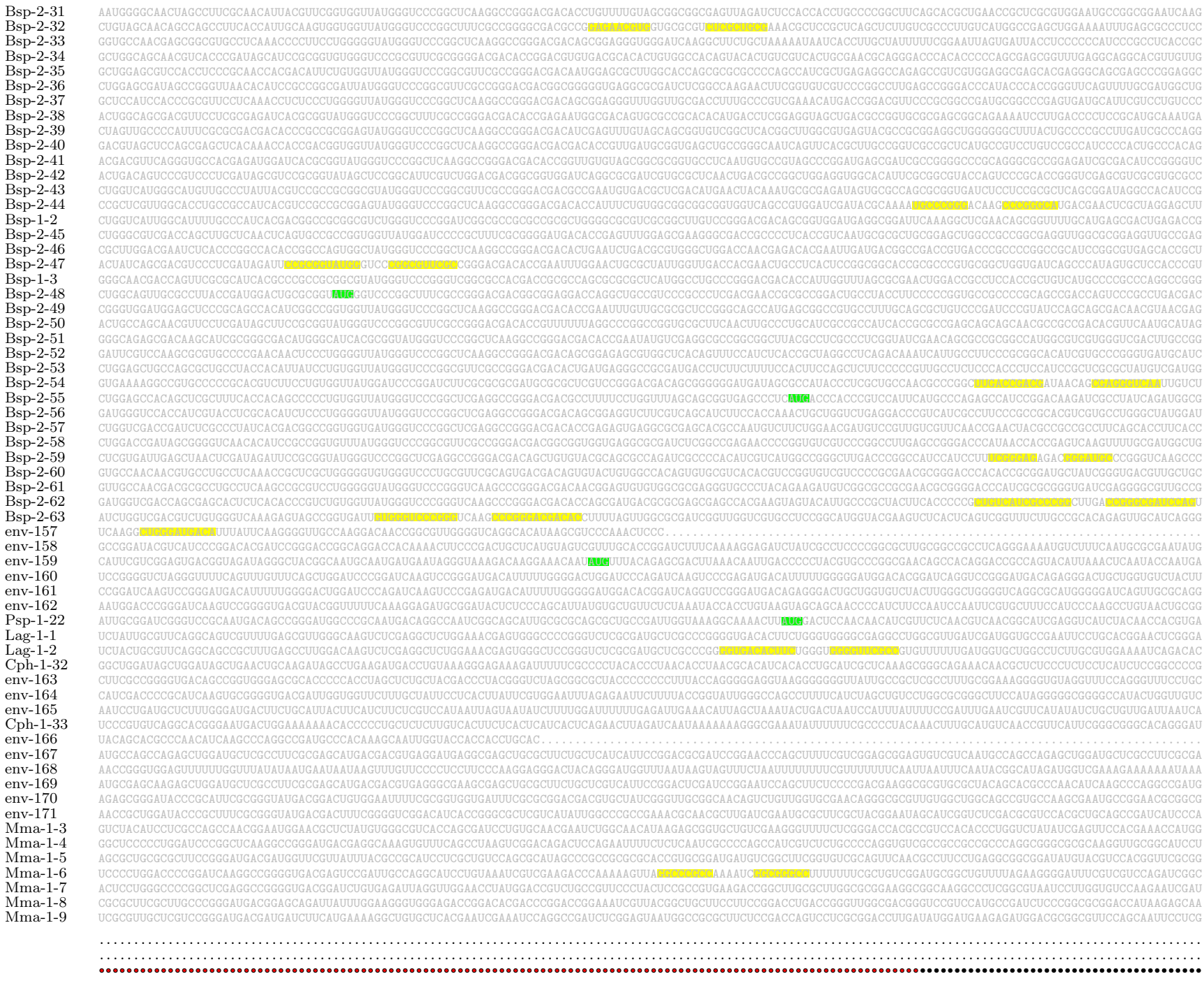

env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
Nar-1-1
env-30
env-31
env-32
env-33
Cph-1-25
env-34
Lbl-1-1
env-35
Osp-1-1
env-36
env-37
env-38
env-39
env-40
env-41
env-42
Sal-1-1
Fba-1-1
Osp-1-2
env-43
Nar-1-2
env-44
env-45
Nar-1-3
Nar-1-4
Nar-1-5
env-46
env-47
env-48
Osp-1-3
Osp-1-4
Osp-1-5
Osp-1-6
Osp-1-7
Osp-1-8
Osp-1-9
Osp-1-10
Osp-1-11
Osp-1-12
Osp-1-13
Osp-1-14
Osp-1-15
Osp-1-16
Osp-1-17
Osp-1-18
Osp-1-19
Osp-1-20
Osp-1-21
Osp-1-22
env-49
Osp-1-23
Osp-1-24
Osp-1-25

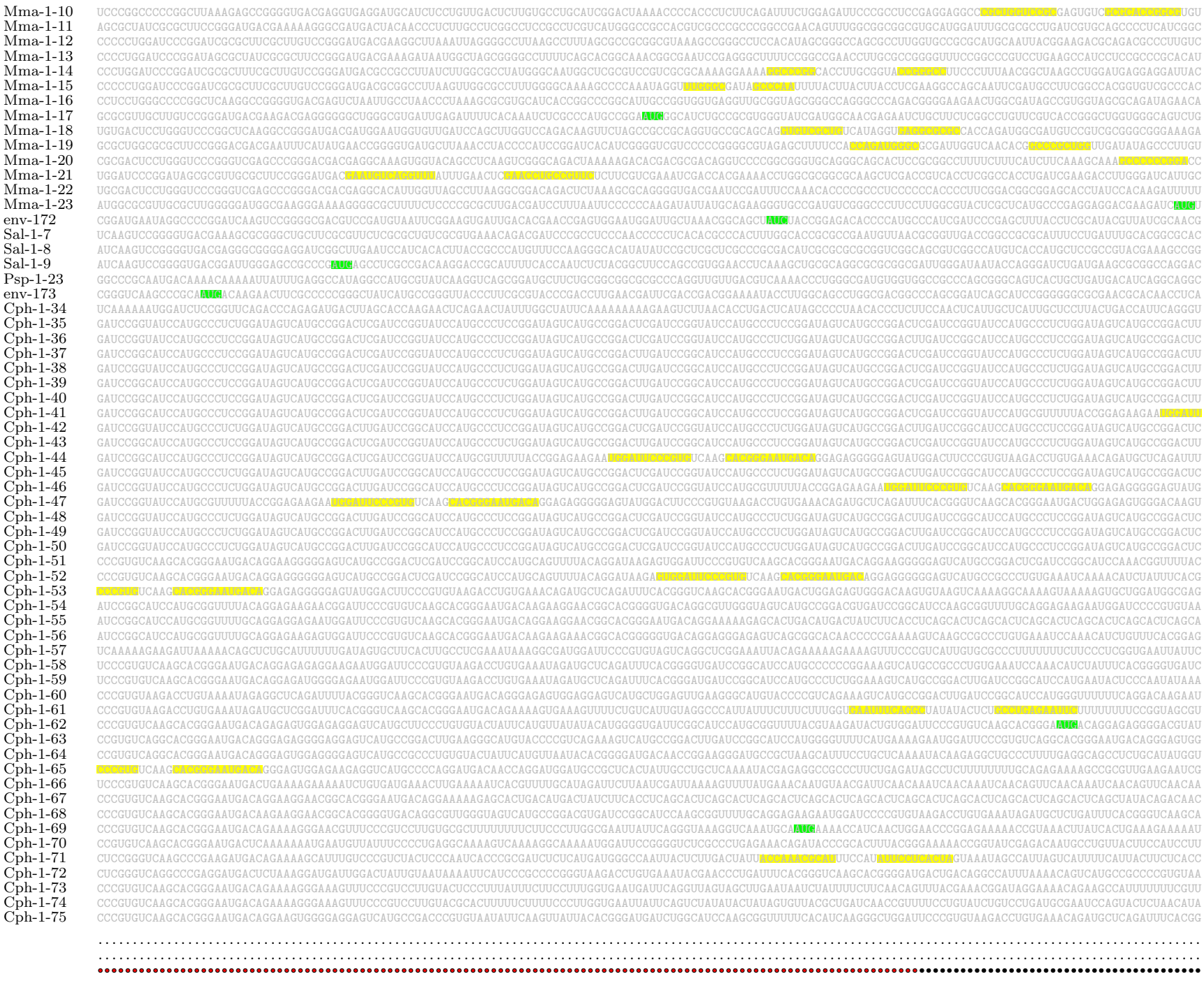

Osp-1-26
Osp-1-27
Osp-1-28
Osp-1-29
Osp-1-30
Osp-1-31
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
Osp-1-32
env-64
env-65
Osp-1-33
Osp-1-34
env-66
env-67
Osp-1-35
Osp-1-36
Osp-1-37
Osp-1-38
Osp-1-39
Osp-1-40
Osp-1-41
Osp-1-42
Osp-1-43
Osp-1-44
Osp-1-45
Osp-1-46
Osp-1-47
Osp-1-48
Osp-1-49
Osp-1-50
Osp-1-51
Osp-1-52
Osp-1-53
Osp-1-54
Osp-1-55
Osp-1-56
Osp-1-57
Osp-1-58
Osp-1-59
Osp-1-60
Osp-1-61
Osp-1-62
Osp-1-63
Osp-1-64
Osp-1-65
Osp-1-66
Osp-1-67
Osp-1-68
Psp-1-1

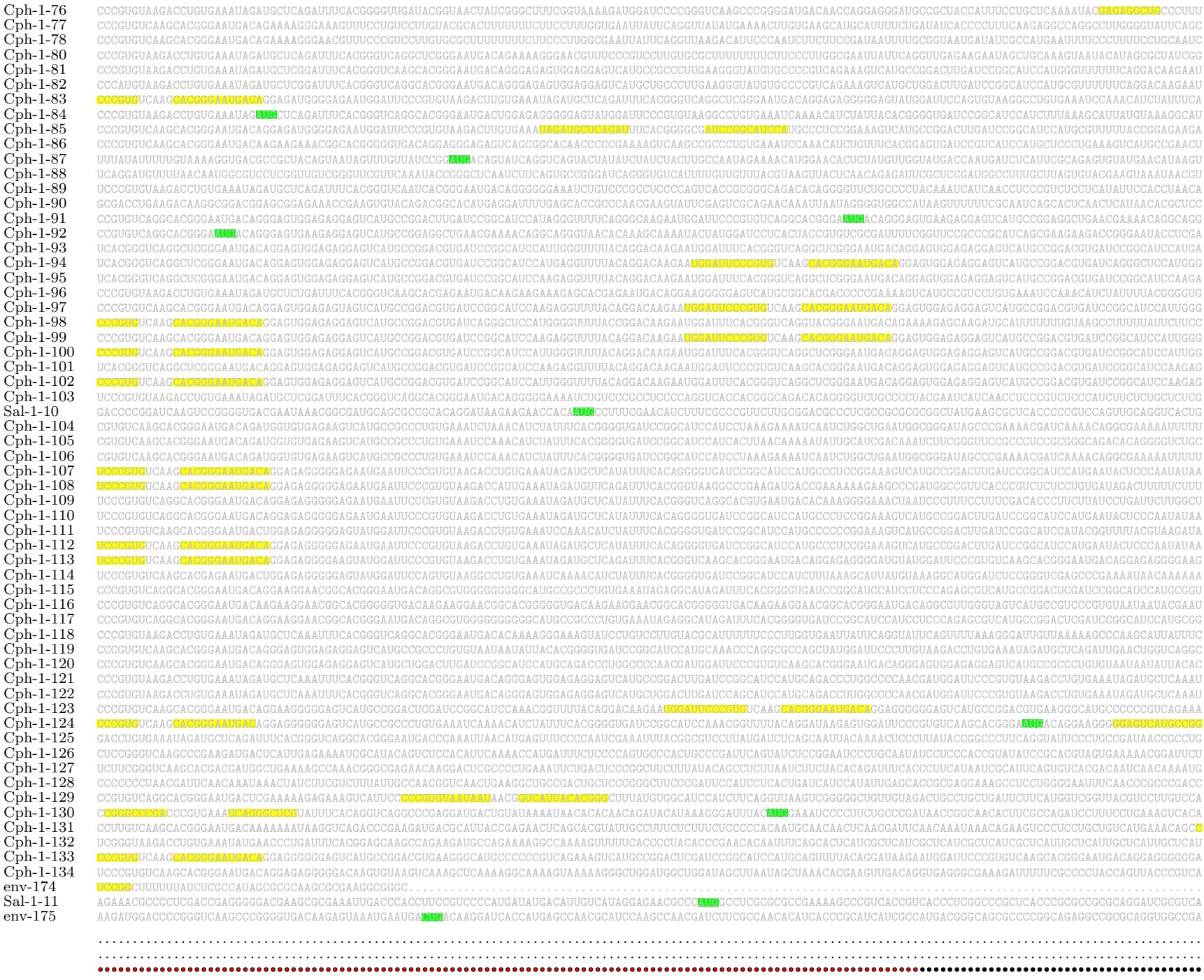
Cph-1-26
env-68
env-69
env-70
env-71
env-72
env-73
Eli-1-1
env-74
Eli-1-2
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
Psp-1-2
env-84
env-85
env-86
env-87
Fba-2-1
env-88
Ssp-2-1
env-89
Fno-2-2
Fno-1-1
Fno-3-1
env-90
Psp-1-3
env-91
Oal-1-1
Oal-1-2
Oal-1-3
Oal-1-4
Oal-1-5
env-92
env-93
env-94
Atu-1-1
Atu-1-2
Atu-1-3
Atu-1-4
Psp-1-4
Psp-1-5
env-95
env-96
Osp-1-69
Osp-1-70
env-97
env-98
Rsp-1-1
Psp-1-6
Psp-1-7
Psp-1-8
Psp-1-9
env-99
env-100
env-101

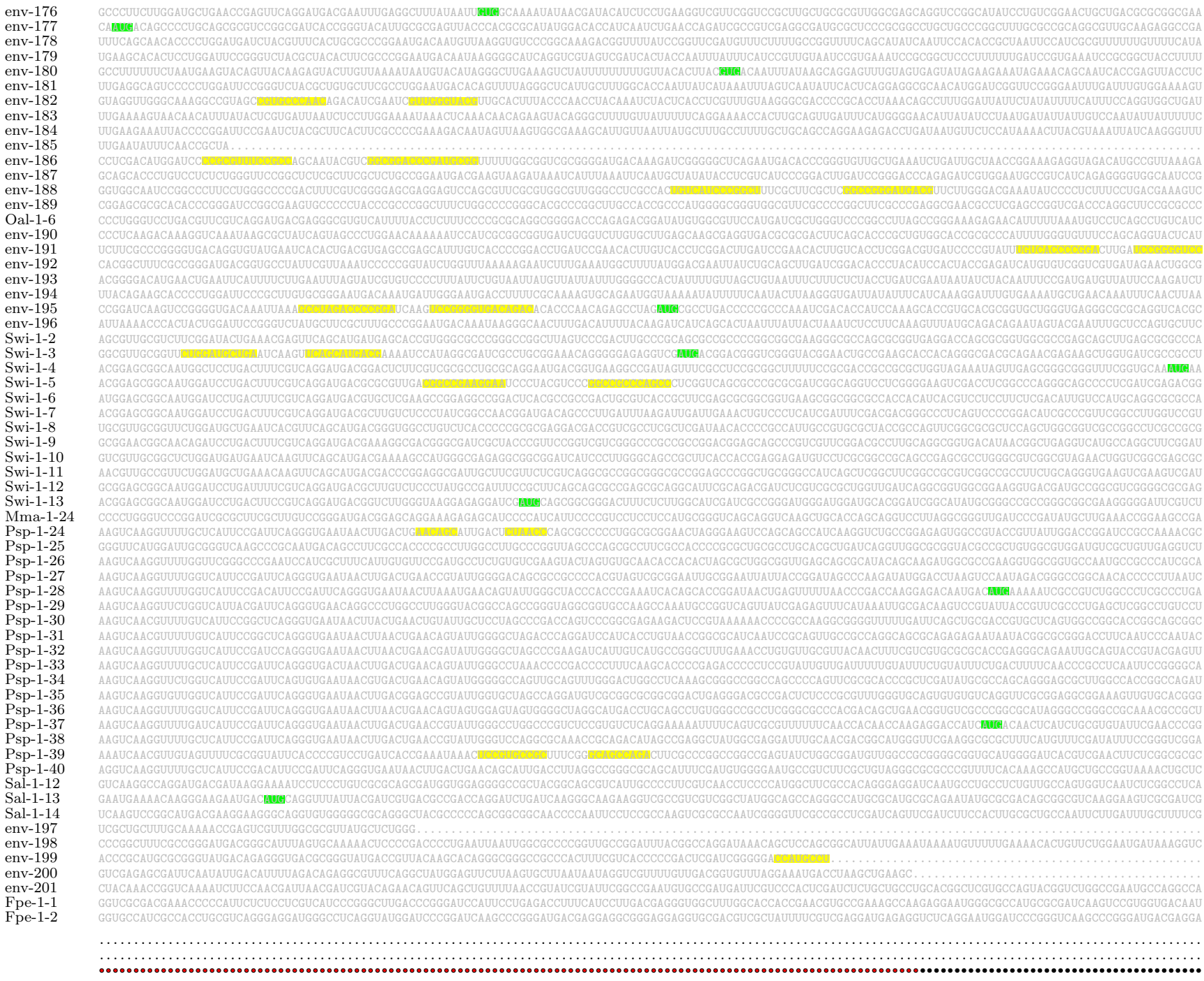
Psp-1-10
Mma-1-1
Mma-1-2
env-102
Psp-1-11
env-103
env-104
env-105
env-106
env-107
env-108
Psp-1-12
Psp-1-13
Psp-1-14
env-109
env-110
Psp-1-15
Psp-1-16
Psp-1-17
Psp-1-18
Psp-1-19
Psp-1-20
Psp-1-21
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
Ssp-1-1
Ssp-1-2
Ssp-1-3
Sal-1-2
Sal-1-3
env-122
env-123
Sal-1-4
Eli-1-3
Eli-1-4
Eli-1-5
env-124
env-125
Eli-1-6
env-126
Ssp-1-4
env-127
env-128
env-129
Atu-1-5
Atu-1-6
env-130
env-131
env-132
env-133
env-134
env-135
env-136


Sal-1-5
Sal-1-6
env-137
env-138
env-139
Dol-1-1
Dol-1-2
env-140
env-141
env-142
env-143
env-144
env-145
Cph-1-27
env-146
env-147
env-148
env-149
env-150
Nha-1-1
env-151
Cph-1-28
Cph-1-29
Cph-1-30
Cph-1-31
env-152
env-153
env-154
env-155
Swi-1-1
Bsp-2-1
env-156
Bsp-2-2
Bsp-2-3
Bsp-2-4
Bsp-2-5
Bsp-2-6
Bsp-2-7
Bsp-2-8
Bsp-2-9
Bsp-2-10
Bsp-2-11
Bsp-2-12
Bsp-2-13
Bsp-2-14
Bsp-2-15
Bsp-2-16
Bsp-2-17
Bsp-2-18
Bsp-2-19
Bsp-2-20
Bsp-2-21
Bsp-2-22
Bsp-2-23
Bsp-1-1
Bsp-2-24
Bsp-2-25
Bsp-2-26
Bsp-2-27
Bsp-2-28
Bsp-2-29
Bsp-2-30

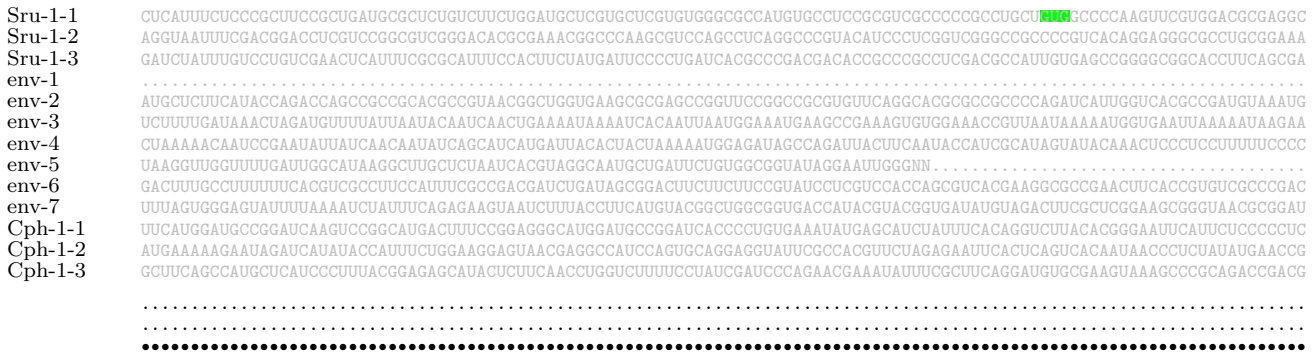

Bsp-2-31
Bsp-2-32
Bsp-2-33
Bsp-2-34
Bsp-2-35
Bsp-2-36
Bsp-2-37
Bsp-2-38
Bsp-2-39
Bsp-2-40
Bsp-2-41
Bsp-2-42
Bsp-2-43
Bsp-2-44
Bsp-1-2
Bsp-2-45
Bsp-2-46
Bsp-2-47
Bsp-1-3
Bsp-2-48
Bsp-2-49
Bsp-2-50
Bsp-2-51
Bsp-2-52
Bsp-2-53
Bsp-2-54
Bsp-2-55
Bsp-2-56
Bsp-2-57
Bsp-2-58
Bsp-2-59
Bsp-2-60
Bsp-2-61
Bsp-2-62
Bsp-2-63
env-157
env-158
env-159
env-160
env-161
env-162
Psp-1-22
Lag-1-1
Lag-1-2
Cph-1-32
env-163
env-164
env-165
Cph-1-33
env-166
env-167
env-168
env-169
env-170
env-171
Mma-1-3
Mma-1-4
Mma-1-5
Mma-1-6
Mma-1-7
Mma-1-8
Mma-1-9


Mma-1-10
Mma-1-11
Mma-1-12
Mma-1-13
Mma-1-14
Mma-1-15
Mma-1-16
Mma-1-17
Mma-1-18
Mma-1-19
Mma-1-20
Mma-1-21
Mma-1-22
Mma-1-23
env-172
Sal-1-7
Sal-1-8
Sal-1-9
Psp-1-23
env-173
Cph-1-34
Cph-1-35
Cph-1-36
Cph-1-37
Cph-1-38
Cph-1-39
Cph-1-40
Cph-1-41
Cph-1-42
Cph-1-43
Cph-1-44
Cph-1-45
Cph-1-46
Cph-1-47
Cph-1-48
Cph-1-49
Cph-1-50
Cph-1-51
Cph-1-52
Cph-1-53
Cph-1-54
Cph-1-55
Cph-1-56
Cph-1-57
Cph-1-58
Cph-1-59
Cph-1-60
Cph-1-61
Cph-1-62
Cph-1-63
Cph-1-64
Cph-1-65
Cph-1-66
Cph-1-67
Cph-1-68
Cph-1-69
Cph-1-70
Cph-1-71
Cph-1-72
Cph-1-73
Cph-1-74
Cph-1-75

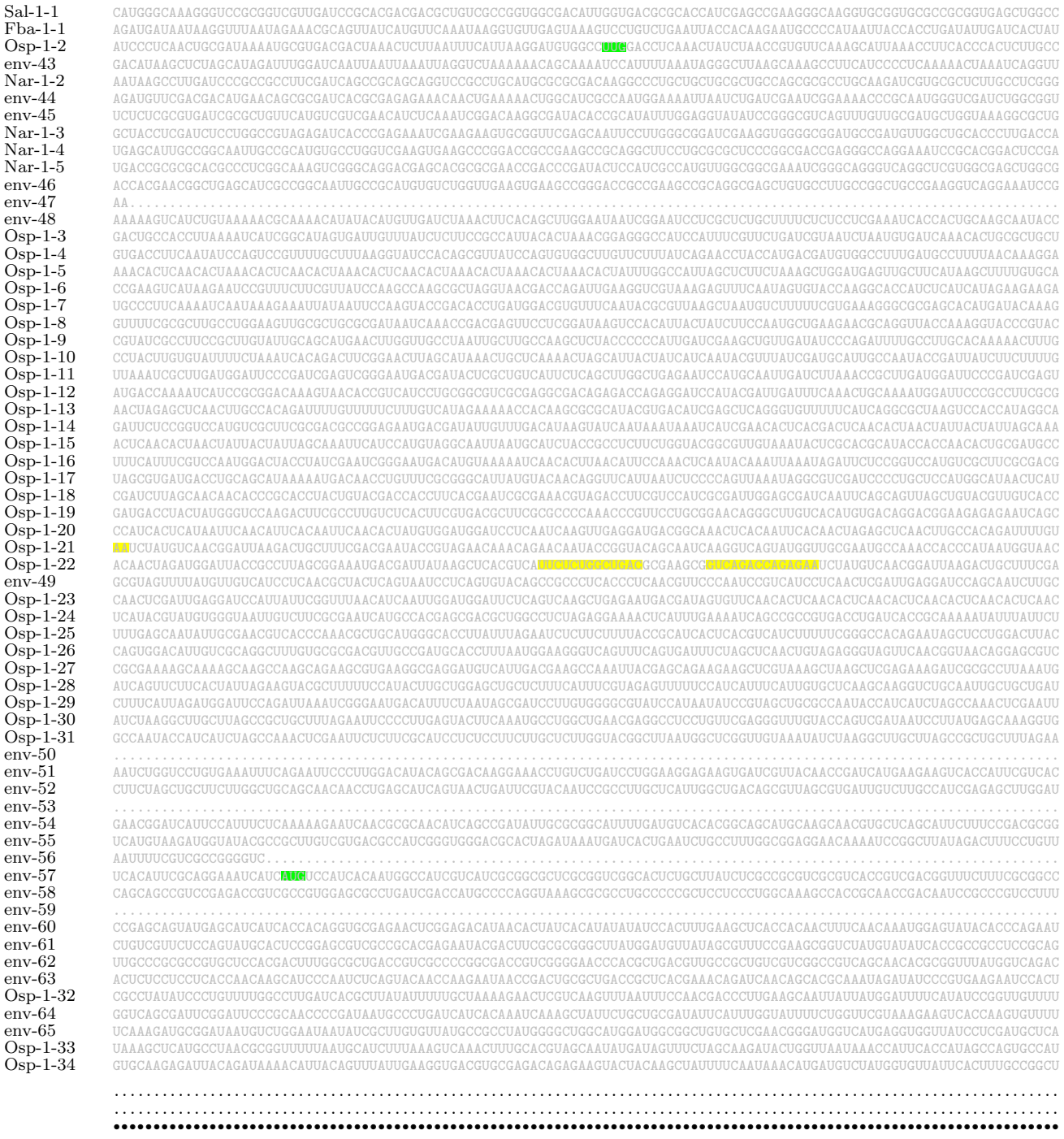

Cph-1-76
Cph-1-77
Cph-1-78
Cph-1-80
Cph-1-81
Cph-1-82
Cph-1-83
Cph-1-84
Cph-1-85
Cph-1-86
Cph-1-87
Cph-1-88
Cph-1-89
Cph-1-90
Cph-1-91
Cph-1-92
Cph-1-93
Cph-1-94
Cph-1-95
Cph-1-96
Cph-1-97
Cph-1-98
Cph-1-99
Cph-1-100
Cph-1-101
Cph-1-102
Cph-1-103
Sal-1-10
Cph-1-104
Cph-1-105
Cph-1-106
Cph-1-107
Cph-1-108
Cph-1-109
Cph-1-110
Cph-1-111
Cph-1-112
Cph-1-113
Cph-1-114
Cph-1-115
Cph-1-116
Cph-1-117
Cph-1-118
Cph-1-119
Cph-1-120
Cph-1-121
Cph-1-122
Cph-1-123
Cph-1-124
Cph-1-125
Cph-1-126
Cph-1-127
Cph-1-128
Cph-1-129
Cph-1-130
Cph-1-131
Cph-1-132
Cph-1-133
Cph-1-134
env-174
Sal-1-11
env-175

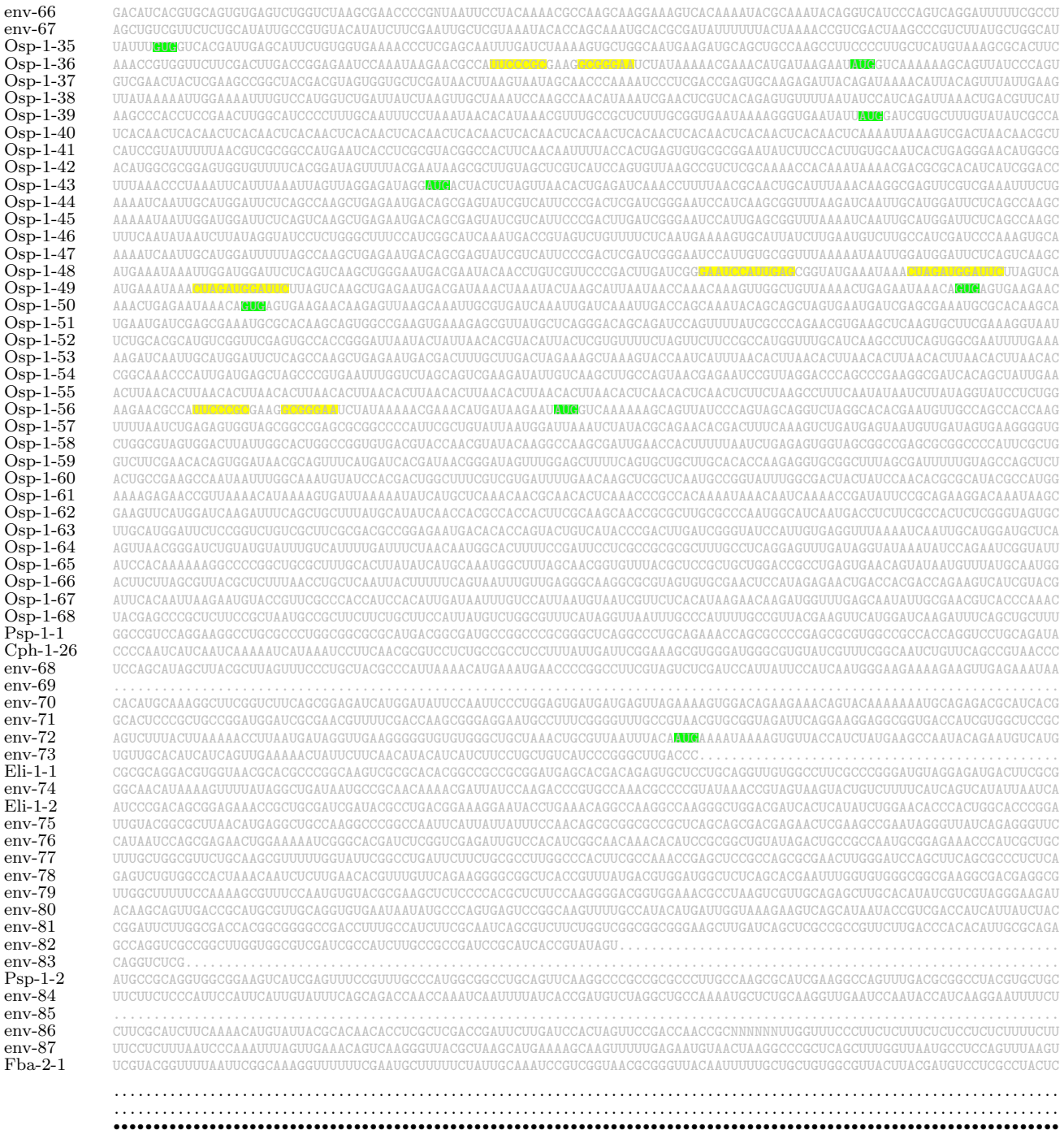

env-176
env-177
env-178
env-179
env-180
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
Oal-1-6
env-190
env-191
env-192
env-193
env-194
env-195
env-196
Swi-1-2
Swi-1-3
Swi-1-4
Swi-1-5
Swi-1-6
Swi-1-7
Swi-1-8
Swi-1-9
Swi-1-10
Swi-1-11
Swi-1-12
Swi-1-13
Mma-1-24
Psp-1-24
Psp-1-25
Psp-1-26
Psp-1-27
Psp-1-28
Psp-1-29
Psp-1-30
Psp-1-31
Psp-1-32
Psp-1-33
Psp-1-34
Psp-1-35
Psp-1-36
Psp-1-37
Psp-1-38
Psp-1-39
Psp-1-40
Sal-1-12
Sal-1-13
Sal-1-14
env-197
env-198
env-199
env-200
env-201
Fpe-1-1
Fpe-1-2

alignment positions 321 · · · 438

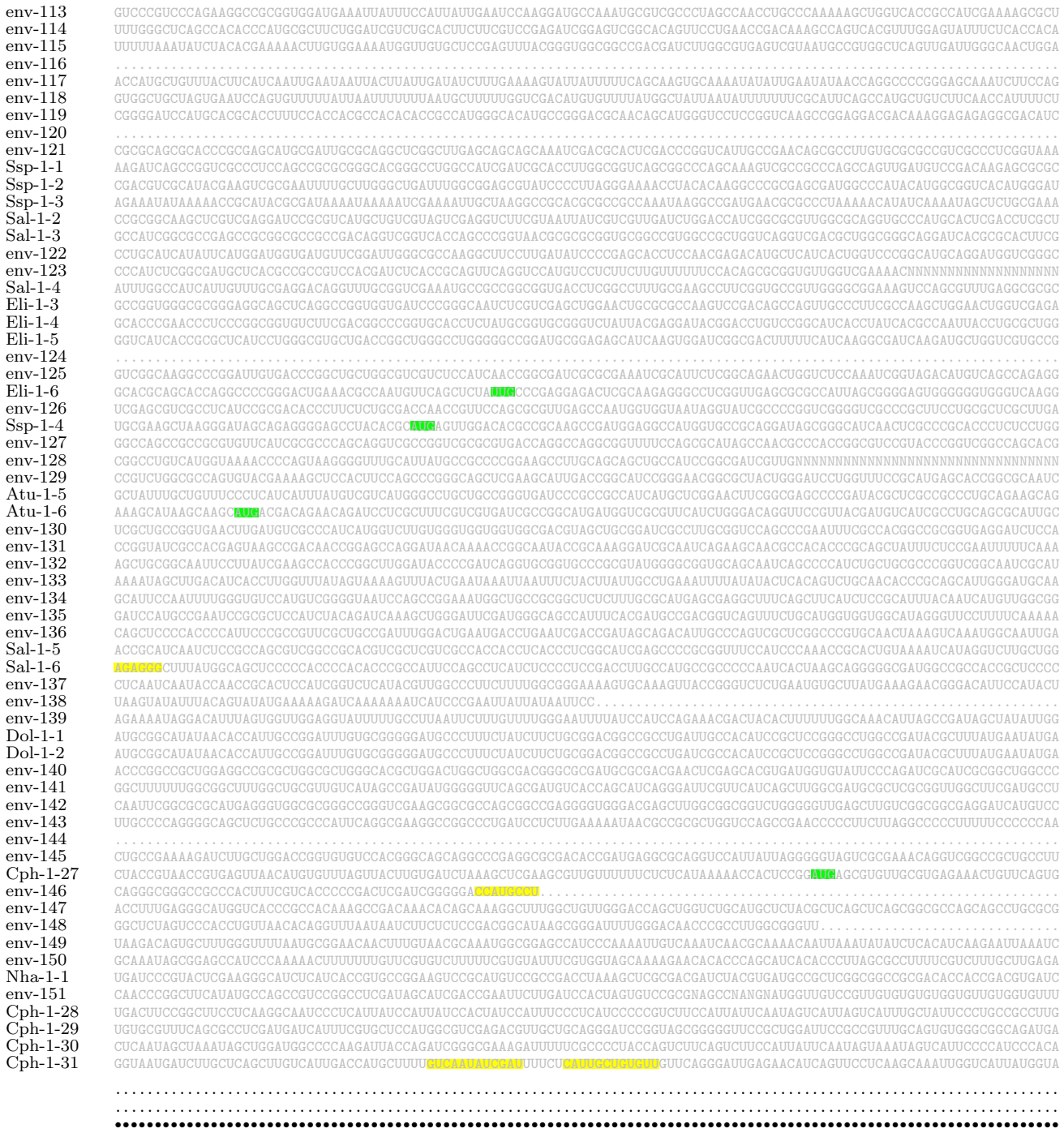


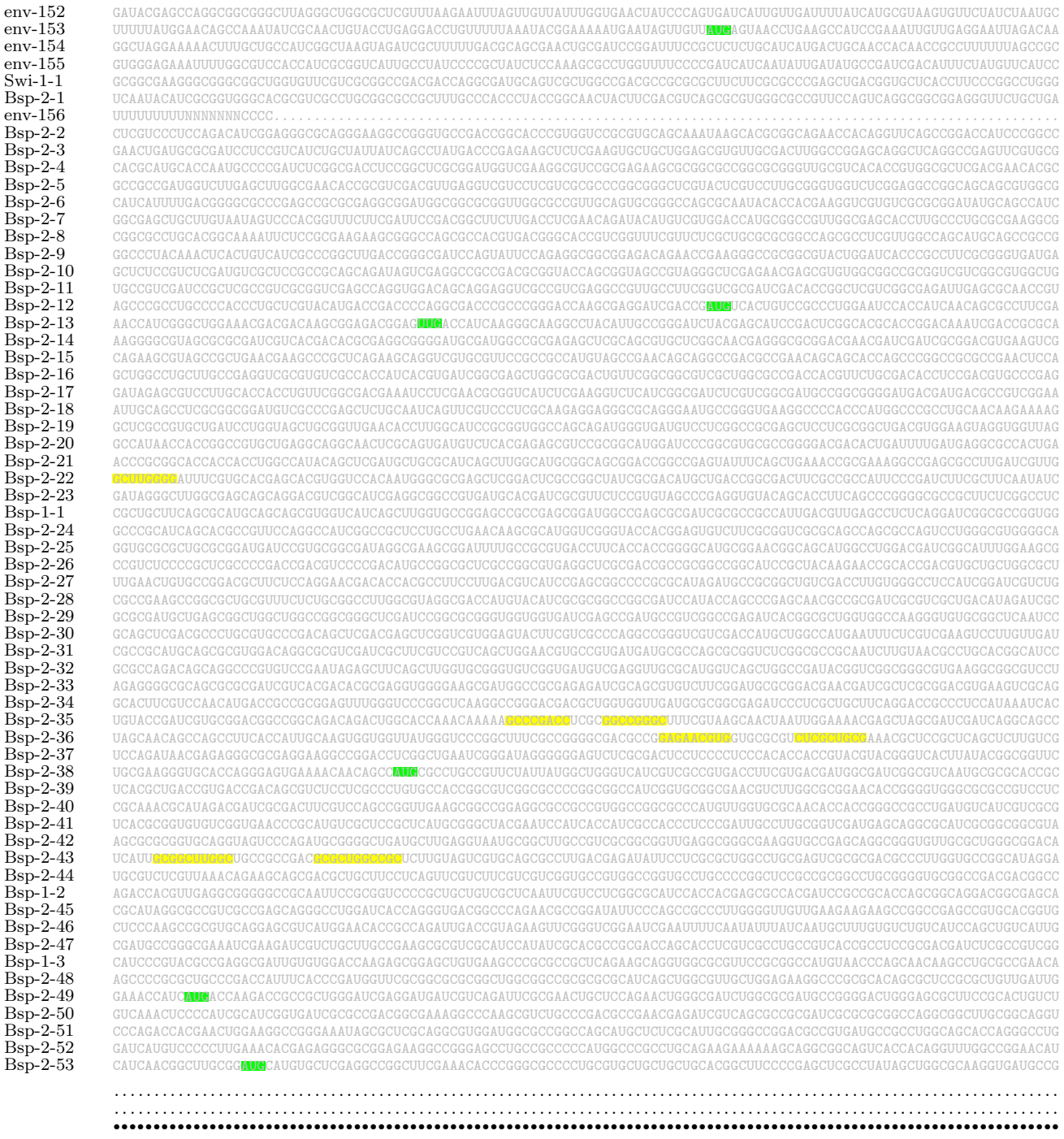
Cph-1-4
Cph-1-5
Cph-1-6
Cph-1-7
Cph-1-8
Cph-1-9
Cph-1-10
Cph-1-11
Cph-1-12
Cph-1-13
Cph-1-14
Cph-1-15
Cph-1-16
Cph-1-17
Cph-1-18
Cph-1-19
Cph-1-20
Cph-1-21
Cph-1-22
Cph-1-23
Cph-1-24
env-8
env-9
env-10
env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
Cat-1-1
Fno-2-1
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
Nar-1-1
env-30
env-31
env-32
env-33
Cph-1-25
env-34
Lbl-1-1
env-35
Osp-1-1
env-36
env-37
env-38
env-39
env-40
env-41
env-42

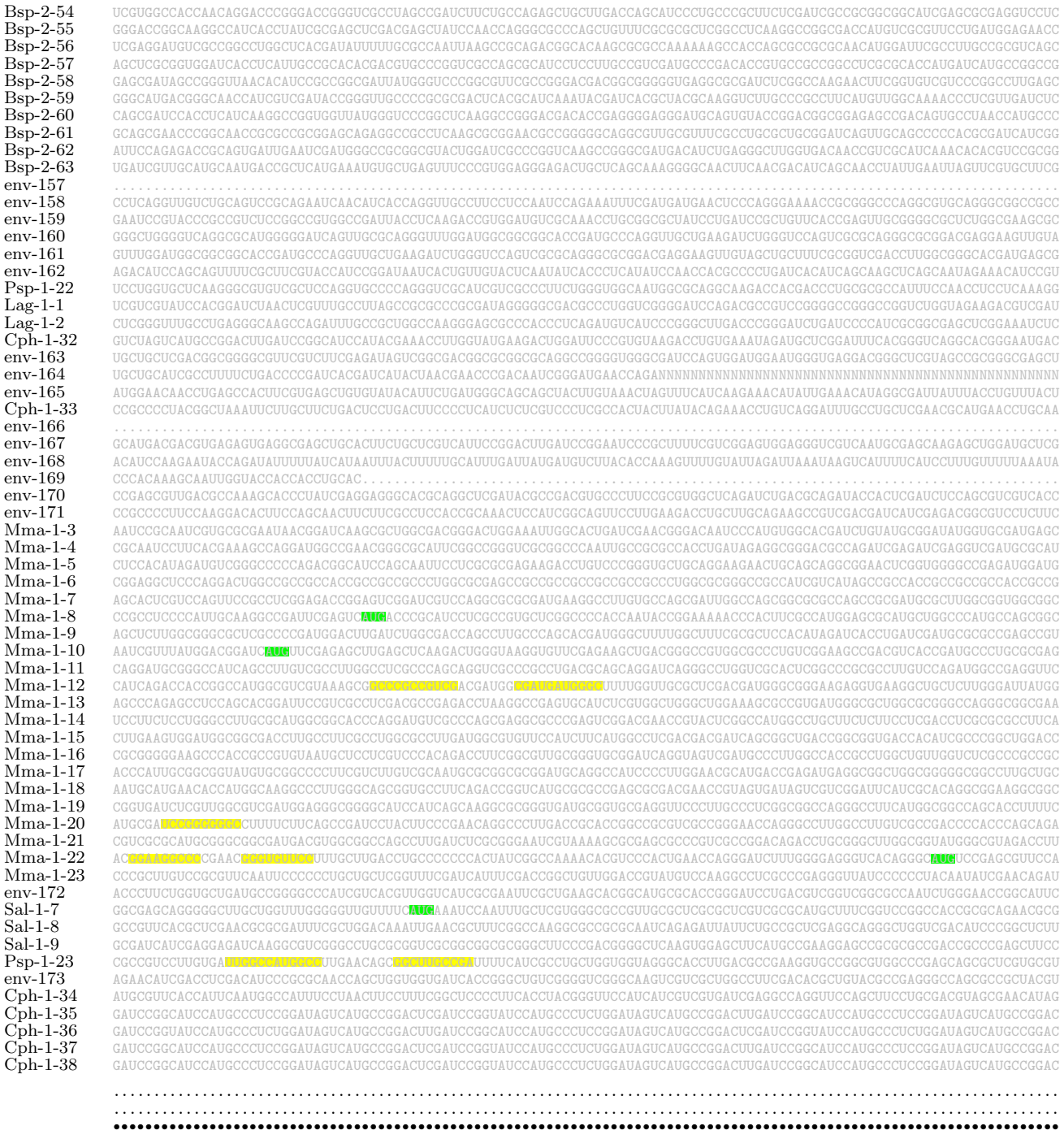

Sal-1-1
Fba-1-1
Osp-1-2
env-43
Nar-1-2
env-44
env-45
Nar-1-3
Nar-1-4
Nar-1-5
env-46
env-47
env-48
Osp-1-3
Osp-1-4
Osp-1-5
Osp-1-6
Osp-1-7
Osp-1-8
Osp-1-9
Osp-1-10
Osp-1-11
Osp-1-12
Osp-1-13
Osp-1-14
Osp-1-15
Osp-1-16
Osp-1-17
Osp-1-18
Osp-1-19
Osp-1-20
Osp-1-21
Osp-1-22
env-49
Osp-1-23
Osp-1-24
Osp-1-25
Osp-1-26
Osp-1-27
Osp-1-28
Osp-1-29
Osp-1-30
Osp-1-31
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
Osp-1-32
env-64
env-65
Osp-1-33
Osp-1-34

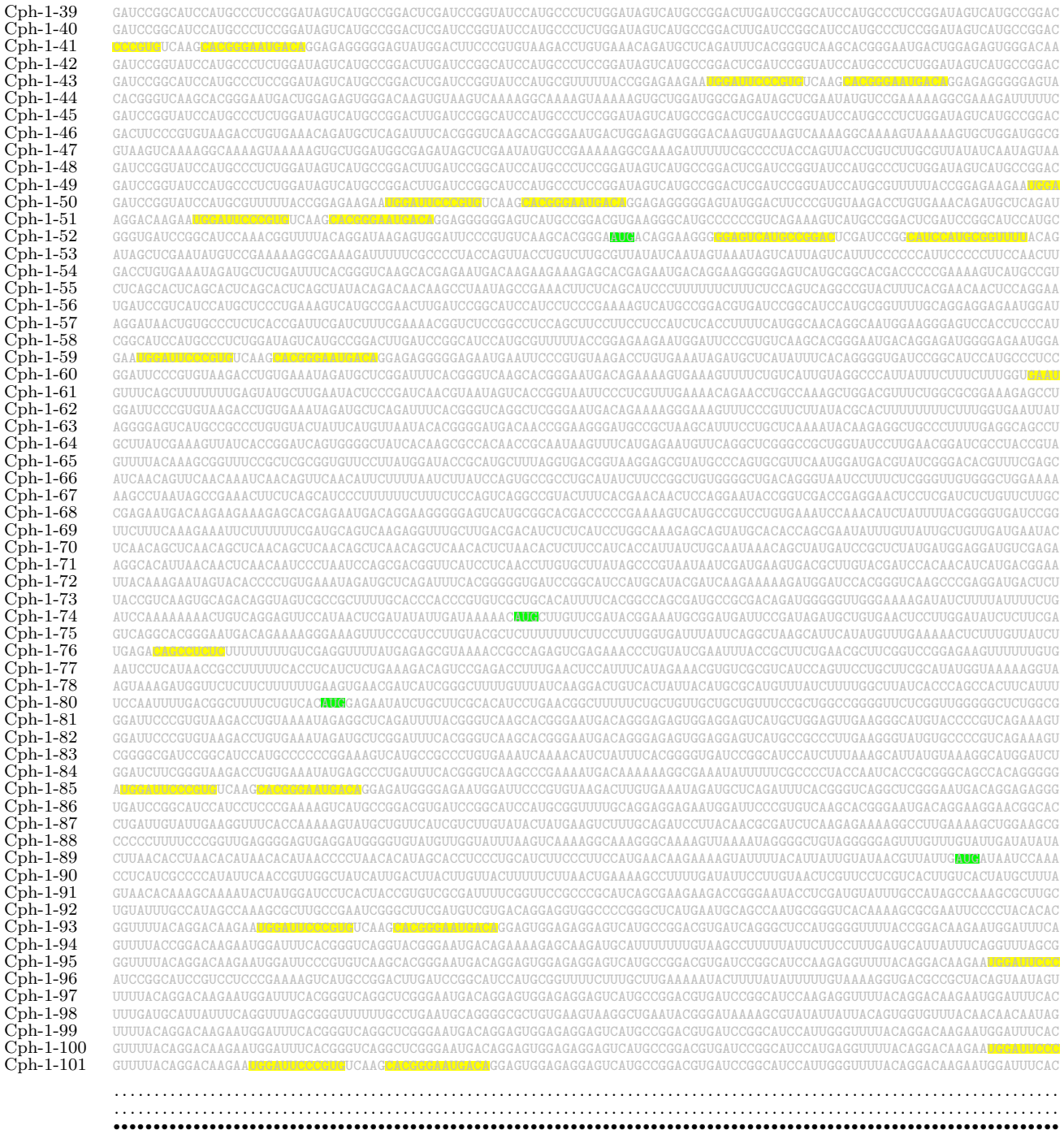

env-66
env-67
Osp-1-35
Osp-1-36
Osp-1-37
Osp-1-38
Osp-1-39
Osp-1-40
Osp-1-41
Osp-1-42
Osp-1-43
Osp-1-44
Osp-1-45
Osp-1-46
Osp-1-47
Osp-1-48
Osp-1-49
Osp-1-50
Osp-1-51
Osp-1-52
Osp-1-53
Osp-1-54
Osp-1-55
Osp-1-56
Osp-1-57
Osp-1-58
Osp-1-59
Osp-1-60
Osp-1-61
Osp-1-62
Osp-1-63
Osp-1-64
Osp-1-65
Osp-1-66
Osp-1-67
Osp-1-68
Psp-1-1
Cph-1-26
env-68
env-69
env-70
env-71
env-72
env-73
Eli-1-1
env-74
Eli-1-2
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
Psp-1-2
env-84
env-85
env-86
env-87
Fba-2-1

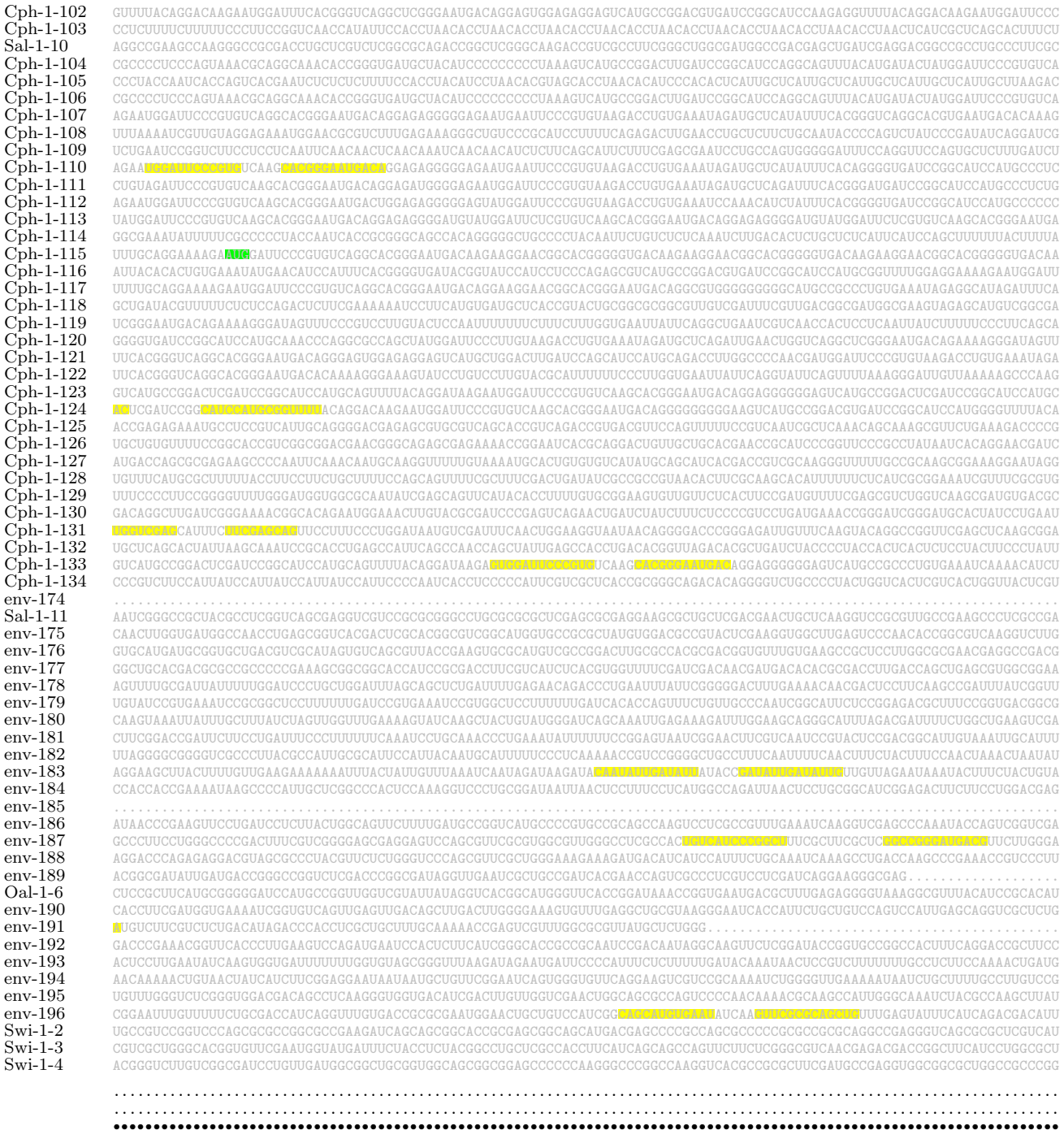
env-88
Ssp-2-1
env-89
Fno-2-2
Fno-1-1
Fno-3-1
env-90
Psp-1-3
env-91
Oal-1-1
Oal-1-2
Oal-1-3
Oal-1-4
Oal-1-5
env-92
env-93
env-94
Atu-1-1
Atu-1-2
Atu-1-3
Atu-1-4
Psp-1-4
Psp-1-5
env-95
env-96
Osp-1-69
Osp-1-70
env-97
env-98
Rsp-1-1
Psp-1-6
Psp-1-7
Psp-1-8
Psp-1-9
env-99
env-100
env-101
Psp-1-10
Mma-1-1
Mma-1-2
env-102
Psp-1-11
env-103
env-104
env-105
env-106
env-107
env-108
Psp-1-12
Psp-1-13
Psp-1-14
env-109
env-110
Psp-1-15
Psp-1-16
Psp-1-17
Psp-1-18
Psp-1-19
Psp-1-20
Psp-1-21
env-111
env-112

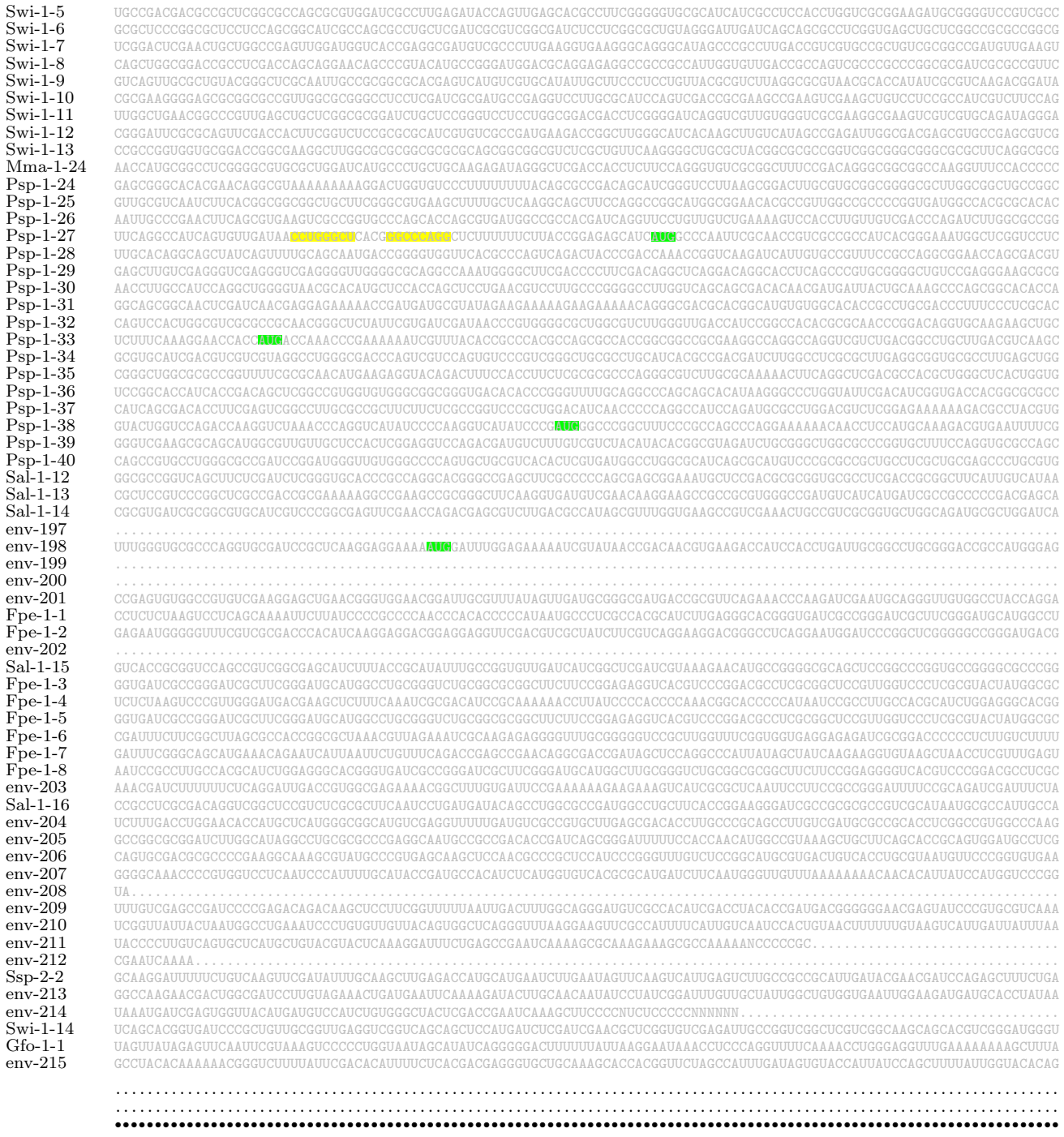

env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
Ssp-1-1
Ssp-1-2
Ssp-1-3
Sal-1-2
Sal-1-3
env-122
env-123
Sal-1-4
Eli-1-3
Eli-1-4
Eli-1-5
env-124
env-125
Eli-1-6
env-126
Ssp-1-4
env-127
env-128
env-129
Atu-1-5
Atu-1-6
env-130
env-131
env-132
env-133
env-134
env-135
env-136
Sal-1-5
Sal-1-6
env-137
env-138
env-139
Dol-1-1
Dol-1-2
env-140
env-141
env-142
env-143
env-144
env-145
Cph-1-27
env-146
env-147
env-148
env-149
env-150
Nha-1-1
env-151
Cph-1-28
Cph-1-29
Cph-1-30
Cph-1-31

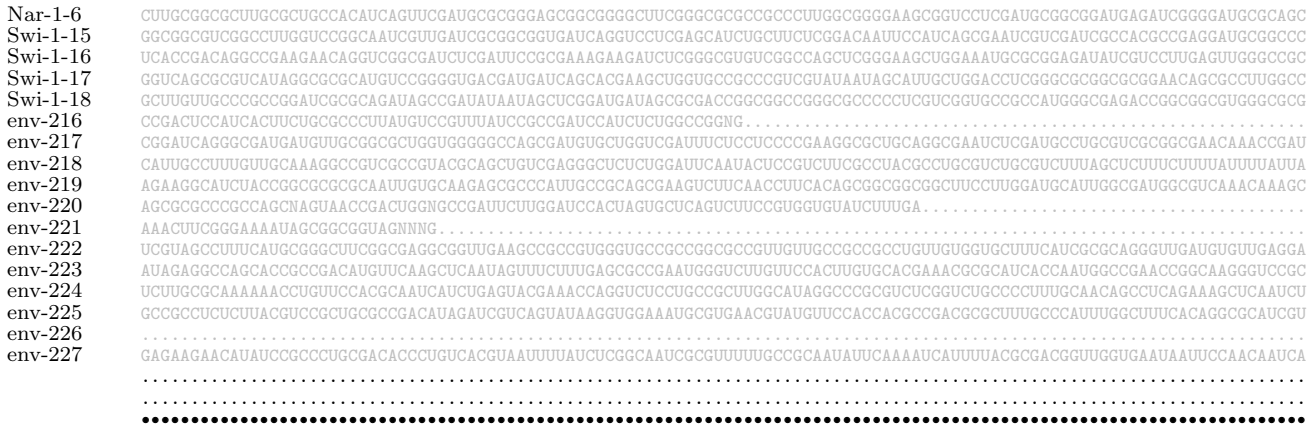

env-152
env-153
env-154
env-155
Swi-1-1
Bsp-2-1
env-156
Bsp-2-2
Bsp-2-3
Bsp-2-4
Bsp-2-5
Bsp-2-6
Bsp-2-7
Bsp-2-8
Bsp-2-9
Bsp-2-10
Bsp-2-11
Bsp-2-12
Bsp-2-13
Bsp-2-14
Bsp-2-15
Bsp-2-16
Bsp-2-17
Bsp-2-18
Bsp-2-19
Bsp-2-20
Bsp-2-21
Bsp-2-22
Bsp-2-23
Bsp-1-1
Bsp-2-24
Bsp-2-25
Bsp-2-26
Bsp-2-27
Bsp-2-28
Bsp-2-29
Bsp-2-30
Bsp-2-31
Bsp-2-32
Bsp-2-33
Bsp-2-34
Bsp-2-35
Bsp-2-36
Bsp-2-37
Bsp-2-38
Bsp-2-39
Bsp-2-40
Bsp-2-41
Bsp-2-42
Bsp-2-43
Bsp-2-44
Bsp-1-2
Bsp-2-45
Bsp-2-46
Bsp-2-47
Bsp-1-3
Bsp-2-48
Bsp-2-49
Bsp-2-50
Bsp-2-51
Bsp-2-52
Bsp-2-53


Bsp-2-54
Bsp-2-55
Bsp-2-56
Bsp-2-57
Bsp-2-58
Bsp-2-59
Bsp-2-60
Bsp-2-61
Bsp-2-62
Bsp-2-63
env-157
env-158
env-159
env-160
env-161
env-162
Psp-1-22
Lag-1-1
Lag-1-2
Cph-1-32
env-163
env-164
env-165
Cph-1-33
env-166
env-167
env-168
env-169
env-170
env-171
Mma-1-3
Mma-1-4
Mma-1-5
Mma-1-6
Mma-1-7
Mma-1-8
Mma-1-9
Mma-1-10
Mma-1-11
Mma-1-12
Mma-1-13
Mma-1-14
Mma-1-15
Mma-1-16
Mma-1-17
Mma-1-18
Mma-1-19
Mma-1-20
Mma-1-21
Mma-1-22
Mma-1-23
env-172
Sal-1-7
Sal-1-8
Sal-1-9
Psp-1-23
env-173
Cph-1-34
Cph-1-35
Cph-1-36
Cph-1-37
Cph-1-38


Cph-1-39
Cph-1-40
Cph-1-41
Cph-1-42
Cph-1-43
Cph-1-44
Cph-1-45
Cph-1-46
Cph-1-47
Cph-1-48
Cph-1-49
Cph-1-50
Cph-1-51
Cph-1-52
Cph-1-53
Cph-1-54
Cph-1-55
Cph-1-56
Cph-1-57
Cph-1-58
Cph-1-59
Cph-1-60
Cph-1-61
Cph-1-62
Cph-1-63
Cph-1-64
Cph-1-65
Cph-1-66
Cph-1-67
Cph-1-68
Cph-1-69
Cph-1-70
Cph-1-71
Cph-1-72
Cph-1-73
Cph-1-74
Cph-1-75
Cph-1-76
Cph-1-77
Cph-1-78
Cph-1-80
Cph-1-81
Cph-1-82
Cph-1-83
Cph-1-84
Cph-1-85
Cph-1-86
Cph-1-87
Cph-1-88
Cph-1-89
Cph-1-90
Cph-1-91
Cph-1-92
Cph-1-93
Cph-1-94
Cph-1-95
Cph-1-96
Cph-1-97
Cph-1-98
Cph-1-99
Cph-1-100
Cph-1-101


Cph-1-102
Cph-1-103
Sal-1-10
Cph-1-104
Cph-1-105
Cph-1-106
Cph-1-107
Cph-1-108
Cph-1-109
Cph-1-110
Cph-1-111
Cph-1-112
Cph-1-113
Cph-1-114
Cph-1-115
Cph-1-116
Cph-1-117
Cph-1-118
Cph-1-119
Cph-1-120
Cph-1-121
Cph-1-122
Cph-1-123
Cph-1-124
Cph-1-125
Cph-1-126
Cph-1-127
Cph-1-128
Cph-1-129
Cph-1-130
Cph-1-131
Cph-1-132
Cph-1-133
Cph-1-134
env-174
Sal-1-11
env-175
env-176
env-177
env-178
env-179
env-180
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
Oal-1-6
env-190
env-191
env-192
env-193
env-194
env-195
env-196
Swi-1-2
Swi-1-3
Swi-1-4


Swi-1-5
Swi-1-6
Swi-1-7
Swi-1-8
Swi-1-9
Swi-1-10
Swi-1-11
Swi-1-12
Swi-1-13
Mma-1-24
Psp-1-24
Psp-1-25
Psp-1-26
Psp-1-27
Psp-1-28
Psp-1-29
Psp-1-30
Psp-1-31
Psp-1-32
Psp-1-33
Psp-1-34
Psp-1-35
Psp-1-36
Psp-1-37
Psp-1-38
Psp-1-39
Psp-1-40
Sal-1-12
Sal-1-13
Sal-1-14
env-197
env-198
env-199
env-200
env-201
Fpe-1-1
Fpe-1-2
env-202
Sal-1-15
Fpe-1-3
Fpe-1-4
Fpe-1-5
Fpe-1-6
Fpe-1-7
Fpe-1-8
env-203
Sal-1-16
env-204
env-205
env-206
env-207
env-208
env-209
env-210
env-211
env-212
Ssp-2-2
env-213
env-214
Swi-1-14
Gfo-1-1
env-215



39 Gut-1

39.1 Taxa

The taxonomy of each organism containing a putative Gut-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Gut-1 RNA in Sections 39.2 and 39.4.

abbrev. of hits	**taxonomy of species**
env-1 to env-43	environmental samples

39.2 Gene contexts

Each Gut-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Gut-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 39.3. The accession number of the sequence containing each Gut-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 39.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	BAAY01013587.1	-	1155	1074	**RNA**→ ←WecE (COG0399)
env-2	hgutS8_s8_175002	+	1511	1592	←hypo **RNA**→ ←WecE (COG0399)
env-3	BABA01021652.1	+	1029	1103	SSB (pfam00436)→ **RNA**→
env-4	BAAZ01025322.1	-	207	137	Arg tRNA→ **RNA**→
env-5	BAAZ01000095.1	+	996	1085	hypo→ **RNA**→ ←Ada (COG0350)
env-6	BAAZ01000132.1	+	11470	11559	hypo→ **RNA**→ ←Ada (COG0350)
env-7	BAAZ01000085.1	-	2840	2766	**RNA**→ CaiC (COG0318)→
env-8	BAAZ01000322.1	+	194	268	hypo→ **RNA**→ ←hypo
env-9	BAAZ01000187.1	-	6811	6737	←bacterial_SERPIN (cd02049) **RNA**→ ←PgsA (COG0558)
env-10	BAAZ01023714.1	-	541	467	**RNA**→ ←hypo
env-11	BAAZ01023713.1	+	569	643	←hypo **RNA**→
env-12	BAAV01002358.1	-	574	499	Tyr tRNA→ **RNA**→
env-39	BAAV01002358.1	-	807	713	(shown above)
env-13	BAAY01000359.1	+	865	937	hypo→ **RNA**→ ←XylB (COG1070)
env-14	BAAV01000249.1	+	935	1007	hypo→ **RNA**→ ←XylB (COG1070)
env-15	hgutS8_s8_175014	-	2866	2792	Icc (COG1409)→ **RNA**→ Transposase, IS256 family→
env-16	BAAZ01000840.1	-	447	370	Glu tRNA→ **RNA**→ hypo→
env-17	BAAZ01000059.1	+	8055	8130	Ser tRNA→ **RNA**→ ←MdlB (COG1132)
env-18	BAAZ01000224.1	-	3117	3043	Leu tRNA→ **RNA**→ hypo→
env-19	BAAV01000831.1	-	176	100	Val tRNA→ **RNA**→
env-20	BAAV01005774.1	-	722	647	Asn tRNA→ **RNA**→ ←hypo
env-21	BAAZ01000130.1	-	2480	2404	Tyr tRNA→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ ←AbgB (COG1473)
env-22	BAAZ01000130.1	-	2240	2164	(shown above)
env-24	BAAZ01000130.1	-	2720	2644	(shown above)
env-26	BAAZ01000130.1	-	2960	2884	(shown above)
env-23	BABA01009772.1	-	249	173	Tyr tRNA→ **RNA**→ **RNA**→ **RNA**→ ←hypo
env-25	BABA01009772.1	-	489	413	(shown above)
env-27	BABA01009772.1	-	729	653	(shown above)
env-28	BAAX01002128.1	+	1051	1127	Leu tRNA→ **RNA**→ ←Peptidase_M20 (pfam01546)
env-29	BAAV01000542.1	-	1906	1830	Pseudo tRNA→ **RNA**→ CsdB (COG0520)→
env-30	BAAY01000031.1	-	2665	2589	Arg tRNA→ **RNA**→ hypo→
env-31	BAAV01000007.1	-	9765	9689	hypo→ **RNA**→ ←COG0535 (COG0535)
env-32	BAAX01002721.1	+	439	515	**RNA**→ ←COG0535 (COG0535)
env-33	BAAY01000011.1	+	11651	11727	hypo→ **RNA**→ ←hypo
env-34	BAAY01000734.1	+	1413	1489	←hypo **RNA**→ GlpR (COG1349)→
env-35	BAAY01000985.1	+	3257	3333	Arg tRNA→ **RNA**→
env-36	BAAY01000139.1	-	5333	5257	Leu tRNA→ **RNA**→ ←Peptidase_M20 (pfam01546)
env-37	BAAV01000478.1	-	3780	3704	Leu tRNA→ **RNA**→ ←Peptidase_M20 (pfam01546)
env-38	BAAU01007736.1	-	751	678	Tyr tRNA→ **RNA**→ Ribosomal_S23p (pfam05635)→
env-40	BAAU01005691.1	-	934	861	**RNA**→ ←COG3760 (COG3760)
env-41	BAAV01001911.1	+	1505	1595	←Ribosomal_S23p (pfam05635) **RNA**→
env-42	BAAV01002358.1	+	1261	1351	←Thr tRNA **RNA**→ hypo→
env-43	hgutS7_s7_168049	+	102	169	**RNA**→ ←hypo

39.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 39.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Gut-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

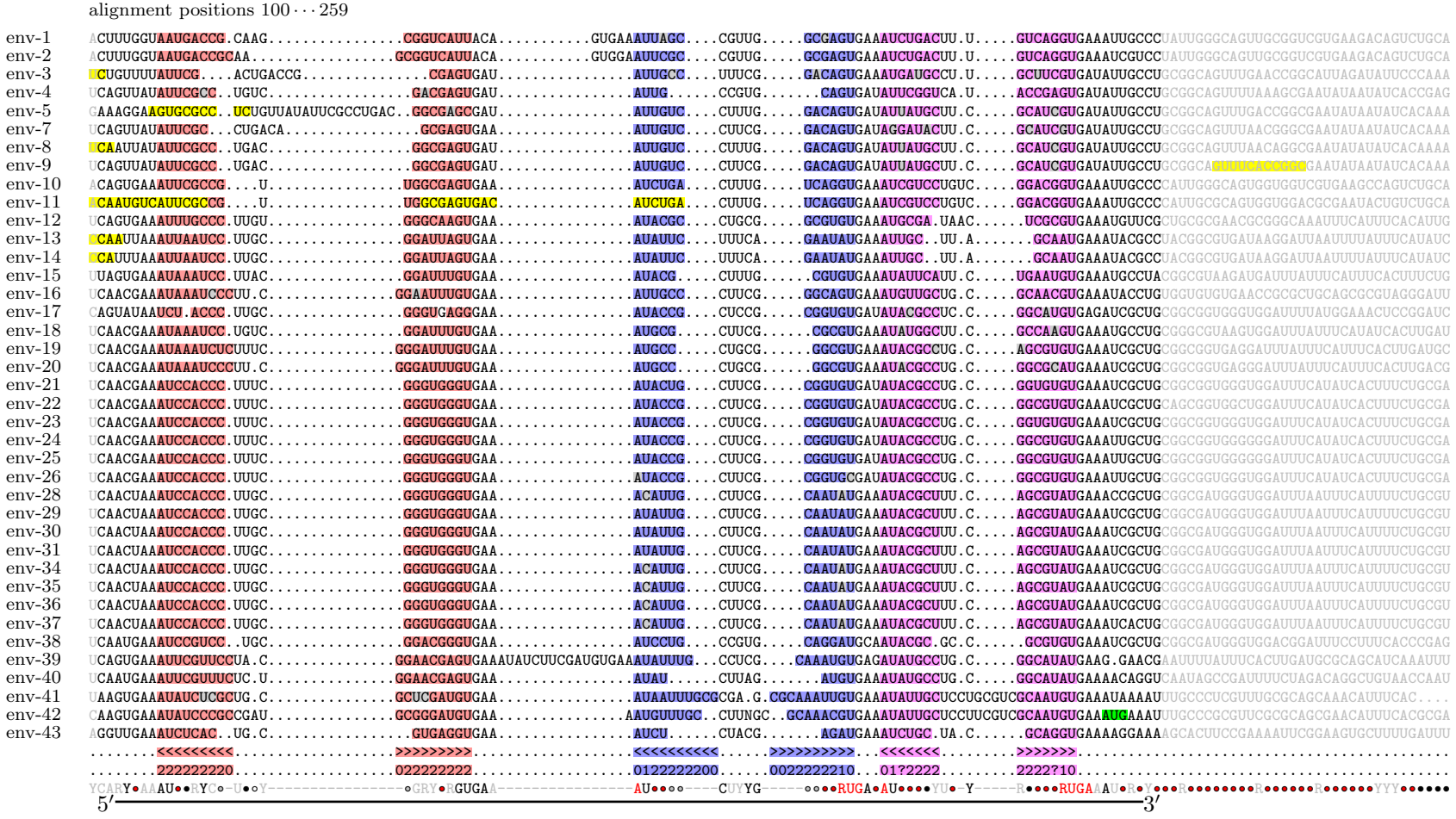
u cd02049 (1) SERine Proteinase INhibitors (serpins), prokaryotic subgroup.
D COG0318 (1) Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II [Lipid metabolism / Secondary metabolites biosynthesis, transport, and catabolism]
d COG0350 (2) Methylated DNA-protein cysteine methyltransferase [DNA replication, recombination, and repair]
d COG0399 (2) Predicted pyridoxal phosphate-dependent enzyme apparently involved in regulation of cell wall biogenesis [Cell envelope biogenesis, outer membrane]
D COG0520 (1) Selenocysteine lyase [Amino acid transport and metabolism]
d COG0535 (2) Predicted Fe-S oxidoreductases [General function prediction only]
d COG0558 (1) Phosphatidylglycerophosphate synthase [Lipid metabolism]
d COG1070 (2) Sugar (pentulose and hexulose) kinases [Carbohydrate transport and metabolism]
d COG1132 (1) ABC-type multidrug transport system, ATPase and permease components [Defense mechanisms]
D COG1349 (1) Transcriptional regulators of sugar metabolism [Transcription / Carbohydrate transport and metabolism]
u COG1409 (1) Predicted phosphohydrolases [General function prediction only]
d COG1473 (1) Metal-dependent amidase/aminoacylase/carboxypeptidase [General function prediction only]
d COG3760 (1) Uncharacterized conserved protein [Function unknown]
u pfam00436 (1) Single-strand binding protein family.
d pfam01546 (3) Peptidase family M20/M25/M40.
uD pfam05635 (2) S23 ribosomal protein.

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: env-27, env-32, env-33, env-6

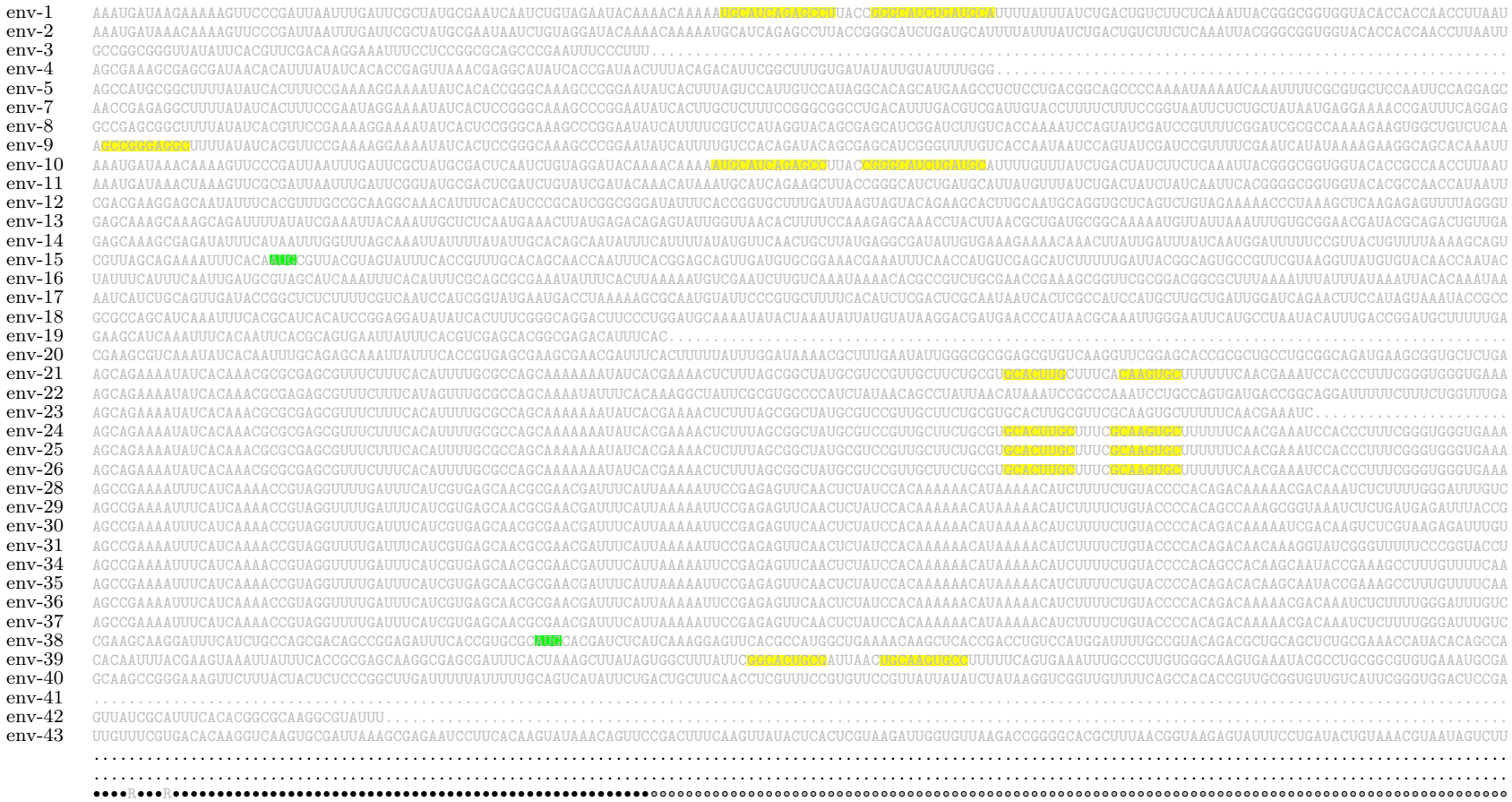
39.4 Multiple-sequence alignment

Each Gut-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 39.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Gut-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

alignment positions 1 · · · 99
env-1
env-2
env-3
env-4
env-5
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-28
env-29
env-30
env-31
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43


alignment positions 100 · · · 259
env-1
env-2
env-3
env-4
env-5
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-28
env-29
env-30
env-31
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
5′
3′

alignment positions 260 · · · 419



40 *gyrA*

40.1 Taxa

The taxonomy of each organism containing a putative *gyrA* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *gyrA* RNA in Sections 40.2 and 40.4.

abbrev.	of hits	taxonomy of species
Pae-1-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* 2192
Pae-2-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* C3719
Pae-3-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PA7
Pae-4-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PACS2
Pae-5-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PAO1
Pae-6-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* UCBPP-PA14

Pen-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas entomophila* L48
Pfl-1-1 to Pfl-1-2	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf-5
Pfl-2-1 to Pfl-2-2	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf0-1
Pme-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas mendocina ymp*
Ppu-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* F1
Ppu-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* GB-1
Ppu-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* KT2440
Ppu-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* W619
Psy-1-1 to Psy-1-2	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. phaseolicola* 1448A
Psy-2-1 to Psy-2-2	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1 to Psy-3-2	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000
env-1 to env-2	environmental samples

40.2 Gene contexts

Each *gyrA* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *gyrA* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 40.3. The accession number of the sequence containing each *gyrA* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 40.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Psy-2-1	NC_007005.1	-	4990467	4990380	**RNA**→ DnaJ (COG0484)→ DapB (COG0289)→
Psy-1-1	NC_005773.3	-	4794981	4794894	**RNA**→ DnaJ (COG0484)→ DapB (COG0289)→
Psy-3-1	NC_004578.1	-	5075507	5075421	**RNA**→ DnaJ (COG0484)→ DapB (COG0289)→
env-1	ABEF01001215.1	+	22	109	**RNA**→
Pfl-1-1	NC_004129.6	+	956664	956750	**RNA**→ DnaJ (COG0484)→ DapB (COG0289)→
Pfl-2-1	NC_007492.1	+	897043	897130	**RNA**→ DnaJ (COG0484)→ DapB (COG0289)→
Pae-3-1	NC_009656.1	+	1991443	1991527	**RNA**→ DNA_topoisoIV (pfam00521)DNA_gyraseA_C (pfam03989)→ SerC (COG1932)→ PheA (COG0077)PheA (COG1605)→ HisC (COG0079)→ AroA (COG0128)TyrA (COG0287)→ Cmk (COG0283)→
Pae-6-1	NC_008463.1	+	2014658	2014742	**RNA**→ DNA_topoisoIV (pfam00521)DNA_gyraseA_C (pfam03989)→ SerC (COG1932)→ PheA (COG0077)PheA (COG1605)→ HisC (COG0079)→ AroA (COG0128)TyrA (COG0287)→ Cmk (COG0283)→
Pae-5-1	NC_002516.2	-	3559294	3559210	**RNA**→ DNA_topoisoIV (pfam00521)DNA_gyraseA_C (pfam03989)→ SerC (COG1932)→ PheA (COG0077)PheA (COG1605)→ HisC (COG0079)→ AroA (COG0128)TyrA (COG0287)→ Cmk (COG0283)→
Pae-2-1	NZ_AAKV01000042.1	-	42244	42160	**RNA**→ DNA_topoisoIV (pfam00521)DNA_gyraseA_C (pfam03989)→ SerC (COG1932)→ PheA (COG0077)PheA (COG1605)→ HisC (COG0079)→ AroA (COG0128)TyrA (COG0287)→ Cmk (COG0283)→
Pae-1-1	NZ_AAKW01000027.1	-	39109	39025	**RNA**→ DNA_topoisoIV (pfam00521)DNA_gyraseA_C (pfam03989)→
Pae-4-1	NZ_AAQW01000001.1	-	4257639	4257555	**RNA**→ DNA_topoisoIV (pfam00521)DNA_gyraseA_C (pfam03989)→ SerC (COG1932)→ PheA (COG0077)PheA (COG1605)→ HisC (COG0079)→ AroA (COG0128)TyrA (COG0287)→ Cmk (COG0283)→
Pme-1-1	NC_009439.1	+	2034231	2034317	**RNA**→ GyrA (COG0188)→
env-2	ABEF01045426.1	+	1407	1493	**RNA**→
Pfl-2-2	NC_007492.1	-	4610446	4610360	**RNA**→ GyrA (COG0188)→ SerC (COG1932)→ PheA (COG0077)PheA (COG1605)→ HisC (COG0079)→ EPSP_synthase (cd01556)TyrA (COG0287)→ Cmk (COG0283)→
Pfl-1-2	NC_004129.6	-	5017949	5017863	**RNA**→ GyrA (COG0188)→ SerC (COG1932)NOG13316 (NOG13316)→ PheA (COG0077)PheA (COG1605)→ HisC (COG0079)→ EPSP_synthase (cd01556)TyrA (COG0287)→ Cmk (COG0283)→

Pen-1-1	NC_008027.1	+	1558985	1559071	**RNA**→ GyrA (COG0188)→
Psy-3-2	NC_004578.1	+	1913388	1913472	**RNA**→ GyrA (COG0188)→
Psy-2-2	NC_007005.1	-	4333410	4333326	**RNA**→ GyrA (COG0188)→
Psy-1-2	NC_005773.3	-	4214931	4214847	**RNA**→ GyrA (COG0188)→
Ppu-3-1	NC_002947.3	+	1970972	1971057	**RNA**→ GyrA (COG0188)→
Ppu-4-1	NZ_AAVY01000017.1	+	71841	71926	**RNA**→ ←hypo
Ppu-1-1	NC_009512.1	-	4452264	4452179	**RNA**→ GyrA (COG0188)→
Ppu-2-1	NZ_AAXR01000003.1	-	426174	426089	**RNA**→ GyrA (COG0188)→

40.3 Conserved domains

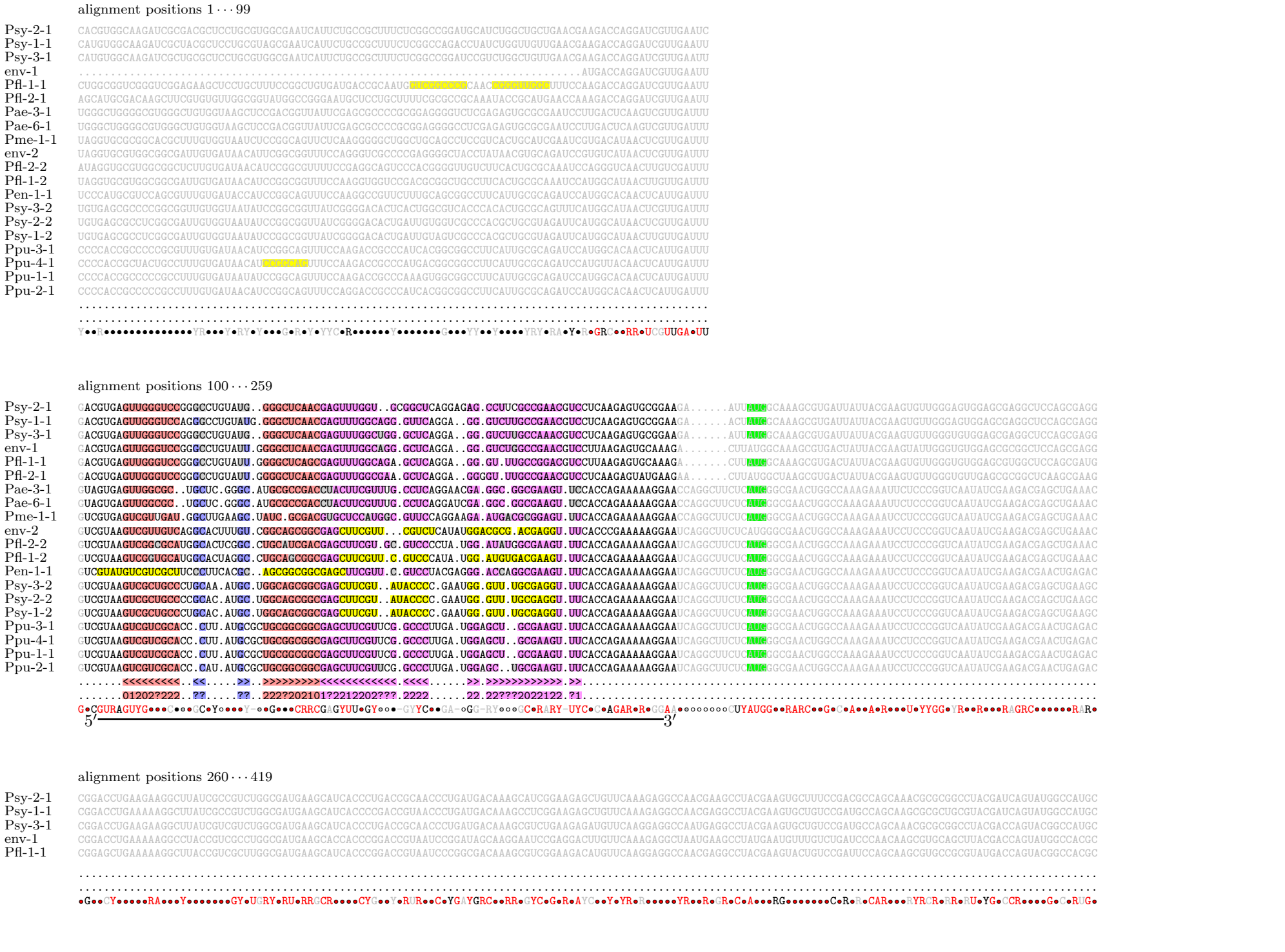
Conserved domains found in protein-coding genes listed in Section 40.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *gyrA* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

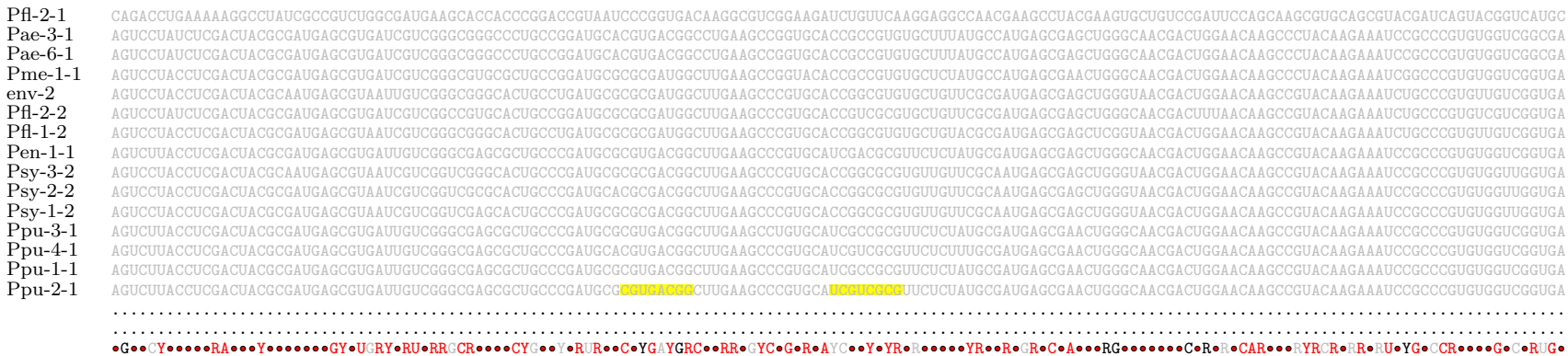
cd01556 (2) EPSP synthase domain.
COG0077 (7) Prephenate dehydratase [Amino acid transport and metabolism]
COG0079 (7) Histidinol-phosphate/aromatic aminotransferase and cobyric acid decarboxylase [Amino acid transport and metabolism]
COG0128 (5) 5-enolpyruvylshikimate-3-phosphate synthase [Amino acid transport and metabolism]
COG0188 (10) Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), A subunit [DNA replication, recombination, and repair]
COG0283 (7) Cytidylate kinase [Nucleotide transport and metabolism]
COG0287 (7) Prephenate dehydrogenase [Amino acid transport and metabolism]
COG0289 (5) Dihydrodipicolinate reductase [Amino acid transport and metabolism]
COG0484 (5) DnaJ-class molecular chaperone with C-terminal Zn finger domain [Posttranslational modification, protein turnover, chaperones]
COG1605 (7) Chorismate mutase [Amino acid transport and metabolism]
COG1932 (7) Phosphoserine aminotransferase [Coenzyme metabolism / Amino acid transport and metabolism]
NOG13316 (1)
pfam00521 (6) DNA gyrase/topoisomerase IV, subunit A.
pfam03989 (6) DNA gyrase C-terminal domain, beta-propeller.

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Pae-1-1, Pae-2-1, Pae-4-1, Pae-5-1

40.4 Multiple-sequence alignment

Each *gyrA* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 40.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *gyrA* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.


alignment positions 1 · · · 99
Psy-2-1
Psy-1-1
Psy-3-1
env-1
Pfl-1-1
Pfl-2-1
Pae-3-1
Pae-6-1
Pme-1-1
env-2
Pfl-2-2
Pfl-1-2
Pen-1-1
Psy-3-2
Psy-2-2
Psy-1-2
Ppu-3-1
Ppu-4-1
Ppu-1-1
Ppu-2-1
alignment positions 100 · · · 259
Psy-2-1
Psy-1-1
Psy-3-1
env-1
Pfl-1-1
Pfl-2-1
Pae-3-1
Pae-6-1
Pme-1-1
env-2
Pfl-2-2
Pfl-1-2
Pen-1-1
Psy-3-2
Psy-2-2
Psy-1-2
Ppu-3-1
Ppu-4-1
Ppu-1-1
Ppu-2-1
5′
3′
alignment positions 260 · · · 419
Psy-2-1
Psy-1-1
Psy-3-1
env-1
Pfl-1-1



41 *hopC*

41.1 Taxa

The taxonomy of each organism containing a putative *hopC* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *hopC* RNA in Sections 41.2 and 41.4.

abbrev.	of hits	taxonomy of species
Hac-1-1		Bacteria Proteobacteria Epsilonproteobacteria Campylobacterales Helicobacteraceae *Helicobacter acinonychis str.* Sheeba
Hpy-1-1		Bacteria Proteobacteria Epsilonproteobacteria Campylobacterales Helicobacteraceae *Helicobacter pylori* 26695
Hpy-2-1		Bacteria Proteobacteria Epsilonproteobacteria Campylobacterales Helicobacteraceae *Helicobacter pylori* HPAG1
Hpy-3-1		Bacteria Proteobacteria Epsilonproteobacteria Campylobacterales Helicobacteraceae *Helicobacter pylori* J99

41.2 Gene contexts

Each *hopC* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *hopC* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 41.3. The accession number of the sequence containing each *hopC* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 41.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Hac-1-1	NC_008229.1	+	1116631	1116693	**RNA**→ outer membrane protein 21→ outer membrane protein 22→
Hpy-3-1	NC_000921.1	+	931454	931513	**RNA**→ Outer membrane protein/porin→ Outer membrane protein/porin→
Hpy-2-1	NC_008086.1	+	931504	931564	**RNA**→ outer membrane protein HopC→ outer membrane protein HopB→
Hpy-1-1	NC_000915.1	+	965107	965166	**RNA**→ outer membrane protein (omp20)→ outer membrane protein (omp21)→

41.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 41.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *hopC* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

41.4 Multiple-sequence alignment

Each *hopC* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 41.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *hopC* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

42 *icd*

42.1 Taxa

The taxonomy of each organism containing a putative *icd* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *icd* RNA in Sections 42.2 and 42.4.

abbrev. of hits	taxonomy of species
Pae-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* 2192
Pae-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* C3719
Pae-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PA7
Pae-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PACS2
Pae-5-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PAO1
Pae-6-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* UCBPP-PA14
Pen-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas entomophila* L48
Pfl-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf-5
Pfl-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf0-1
Pme-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas mendocina ymp*
Ppu-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* F1
Ppu-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* GB-1
Ppu-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* KT2440
Ppu-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* W619
Pst-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas stutzeri* A1501
Psy-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. phaseolicola* 1448A
Psy-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000
env-1 to env-2	environmental samples

42.2 Gene contexts

Each *icd* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *icd* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 42.3. The accession number of the sequence containing each *icd* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 42.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Pae-6-1	NC_008463.1	-	2614922	2614853	**RNA**→ IDH (pfam03971)→ putative NUDIX hydrolase→ tRNA_Me_trans (cd01998)→ COG2915 (COG2915)→ DUF6 (pfam00892)→ PurB (cd01598)→ COG2850 (COG2850)→ ElaA (COG2153)→ hypo→ Secretin_C (pfam00263)Secretin_N (pfam03958)→
Pae-5-1	NC_002516.2	+	2966988	2967057	**RNA**→ IDH (pfam03971)→ hypo→ tRNA_Me_trans (cd01998)→ COG2915 (COG2915)→ DUF6 (pfam00892)→ PurB (cd01598)→ COG2850 (COG2850)→ ElaA (COG2153)→ hypo→ Secretin_C (pfam00263)Secretin_N (pfam03958)NOG17035 (NOG17035)→
Pae-2-1	NZ_AAKV01000029.1	+	67214	67283	**RNA**→ IDH (pfam03971)→ COG0494: NTP pyrophosphohydrolases including oxidative damage repair enzymes→ tRNA_Me_trans (cd01998)COG2915 (COG2915)→ DUF6 (pfam00892)→ PurB (cd01598)→ COG2850 (COG2850)→ ElaA (COG2153)→ hypo→ Secretin_C (pfam00263)Secretin_N (pfam03958)→
Pae-1-1	NZ_AAKW01000021.1	+	182003	182072	**RNA**→ IDH (pfam03971)→ COG0494: NTP pyrophosphohydrolases including oxidative damage repair enzymes→ tRNA_Me_trans (cd01998)→ COG2915 (COG2915)→ DUF6 (pfam00892)→ PurB (cd01598)→ COG2850 (COG2850)→ ElaA (COG2153)→ hypo→ Secretin_C (pfam00263)Secretin_N (pfam03958)→
Pae-4-1	NZ_AAQW01000001.1	+	3627791	3627860	**RNA**→ IDH (pfam03971)→ hypo→ tRNA_Me_trans (cd01998)→ COG2915 (COG2915)→ DUF6 (pfam00892)→ PurB (cd01598)→ COG2850 (COG2850)→ ElaA (COG2153)→ hypo→ Secretin_C (pfam00263)Secretin_N (pfam03958)→

Pae-3-1	NC_009656.1	-	2607165	2607096	**RNA**→ IDH (pfam03971)→ hypo→ tRNA_Me_trans (cd01998)→ COG2915 (COG2915)→ DUF6 (pfam00892)→ PurB (cd01598)→ COG2850 (COG2850)→ ElaA (COG2153)→ hypo→ Secretin_C (pfam00263)Secretin_N (pfam03958)→
Pme-1-1	NC_009439.1	+	2644352	2644430	**RNA**→ IDH (pfam03971)→
Pst-1-1	NC_009434.1	+	2512479	2512556	**RNA**→ IDH (pfam03971)→
Psy-2-1	NC_007005.1	+	3814381	3814449	**RNA**→ IDH (pfam03971)→
Psy-3-1	NC_004578.1	+	3792906	3792974	**RNA**→ IDH (pfam03971)→
Psy-1-1	NC_005773.3	+	3597053	3597121	**RNA**→ IDH (pfam03971)→
Pfl-2-1	NC_007492.1	+	4061823	4061892	**RNA**→ IDH (pfam03971)→ NUDIX (pfam00293)→ tRNA_Me_trans (cd01998)→ COG2915 (COG2915)→ PurB (cd01598)→ COG2850 (COG2850)→ ElaA (COG2153)→
env-1	JCVI_SCAF_1096627837757	-	599	531	**RNA**→ ? (GOS10229)→
Ppu-4-1	NZ_AAVY01000010.1	+	176396	176463	**RNA**→ IDH (pfam03971)→
Pen-1-1	NC_008027.1	-	2316895	2316828	**RNA**→ IDH (pfam03971)→
Ppu-1-1	NC_009512.1	-	2092942	2092875	**RNA**→ IDH (pfam03971)→
Ppu-3-1	NC_002947.3	+	4521981	4522048	**RNA**→ IDH (pfam03971)→
Ppu-2-1	NZ_AAXR01000012.1	-	4896	4829	**RNA**→ IDH (pfam03971)→
env-2	ABEF01050855.1	+	1590	1658	**RNA**→
Pfl-1-1	NC_004129.6	+	4487914	4487981	**RNA**→ IDH (pfam03971)→

42.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 42.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *icd* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

cd01598 (7) PurB: PurB_like adenylosuccinases (adenylsuccinate lyase, ASL).
cd01998 (7) tRNA methyl transferase.
COG2153 (7) Predicted acyltransferase [General function prediction only]
COG2850 (7) Uncharacterized conserved protein [Function unknown]
COG2915 (7) Uncharacterized protein involved in purine metabolism [General function prediction only]
GOS10229 (1) ?
NOG17035 (1)
pfam00263 (6) Bacterial type II and III secretion system protein.
pfam00293 (1) NUDIX domain.
pfam00892 (6) Integral membrane protein DUF6.
pfam03958 (6) Bacterial type II/III secretion system short domain.
pfam03971 (18) Monomeric isocitrate dehydrogenase.

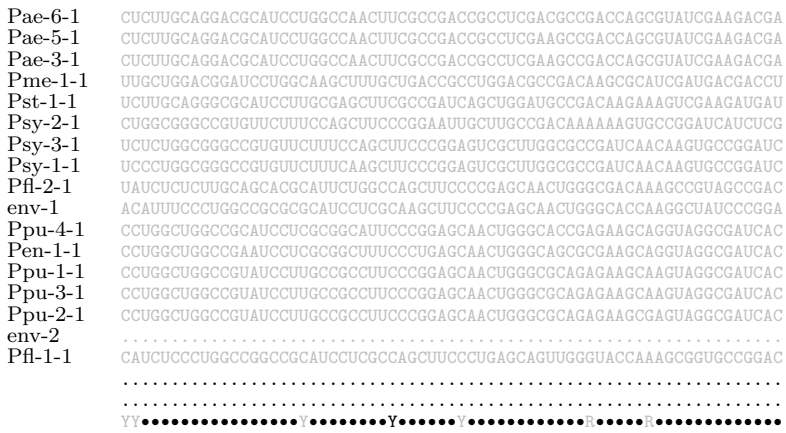
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Pae-1-1, Pae-2-1, Pae-4-1

42.4 Multiple-sequence alignment

Each *icd* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 42.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *icd* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides

for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

alignment positions 420 · · · 486



43 JUMPstart

43.1 Taxa

The taxonomy of each organism containing a putative JUMPstart RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual JUMPstart RNA in Sections 43.2 and 43.4.

abbrev. of hits	**taxonomy of species**
Sen-1-1 to Sen-1-2	AKU_12601Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Salmonella enterica subsp. enterica serovar* Paratyphi A str
Bli-1-1 to Bli-1-2	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus licheniformis* ATCC 14580 (DSM 13)
Ahy-1-1 to Ahy-1-9	Bacteria Proteobacteria Gammaproteobacteria Aeromonadales Aeromonadaceae *Aeromonas hydrophila subsp. hydrophila* ATCC 7966
Asa-1-1 to Asa-1-8	Bacteria Proteobacteria Gammaproteobacteria Aeromonadales Aeromonadaceae *Aeromonas salmonicida subsp. salmonicida* A449
Sam-1-1	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella amazonensis* SB2B
Sfr-1-1 to Sfr-1-2	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella frigidimarina* NCIMB 400
Son-1-1	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella oneidensis* MR-1

Spu-1-1	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella putrefaciens* 200
Ssp-1-1	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella sp.* MR-4
Ssp-2-1	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella sp.* MR-7
Esa-1-1 to Esa-1-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Cronobacter *Enterobacter sakazakii* ATCC BAA-894
Eca-1-1 to Eca-1-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Enterobacter cancerogenus* ATCC 35316
Esp-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Enterobacter sp.* 638
Eal-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia albertii* TW07627
Eco-1-1 to Eco-1-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* 101-1
Eco-2-1 to Eco-2-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* 536
Eco-3-1 to Eco-3-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* 53638
Eco-4-1 to Eco-4-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* APEC O1
Eco-5-1 to Eco-5-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* ATCC 8739
Eco-6-1 to Eco-6-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* B171
Eco-7-1 to Eco-7-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* B7A
Eco-8-1 to Eco-8-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* CFT073
Eco-9-1 to Eco-9-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* E110019
Eco-10-1 to Eco-10-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* E22
Eco-11-1 to Eco-11-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* E24377A
Eco-12-1 to Eco-12-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* F11
Eco-13-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* HS
Eco-14-1 to Eco-14-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 EDL933
Eco-15-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4024
Eco-16-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4042
Eco-17-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4045
Eco-18-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4076
Eco-19-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4113
Eco-20-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4115
Eco-21-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4206
Eco-22-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4401
Eco-23-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4486
Eco-24-1 to Eco-24-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4501
Eco-25-1 to Eco-25-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC508
Eco-26-1 to Eco-26-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC869
Eco-27-1 to Eco-27-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. Sakai
Eco-28-1 to Eco-28-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. TW14588
Eco-29-1 to Eco-29-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* SE11
Eco-30-1 to Eco-30-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* SMS-3-5
Eco-32-1 to Eco-32-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli str.* K12 substr. DH10B
Eco-33-1 to Eco-33-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli str.* K12 substr. MG1655
Eco-34-1 to Eco-34-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli str.* K12 substr. W3110
Eco-31-1 to Eco-31-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* UTI89
Pat-1-1 to Pat-1-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Pectobacterium atrosepticum* SCRI1043
Sen-2-1 to Sen-2-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Salmonella enterica subsp. enterica serovar* Paratyphi A str. ATCC
Sen-3-1 to Sen-3-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Agona *Salmonella enterica subsp. ente*
Sen-4-1 to Sen-4-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Dublin *Salmonella enterica subsp. ent*
Sen-5-1 to Sen-5-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Gallinarum *Salmonella enterica subsp.*
Sen-6-1 to Sen-6-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Hadar *Salmonella enterica subsp. ente*
Sen-7-1 to Sen-7-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Heidelberg *Salmonella enterica subsp.*
Sen-8-1 to Sen-8-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Heidelberg *Salmonella enterica subsp.*
Sen-9-1 to Sen-9-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Kentucky *Salmonella enterica subsp.*
Sen-10-1 to Sen-10-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Kentucky *Salmonella enterica subsp.*
Sen-11-1 to Sen-11-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Newport *Salmonella enterica subsp. e*
Sen-12-1 to Sen-12-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Newport *Salmonella enterica subsp. e*
Sen-13-1 to Sen-13-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Paratyphi B *Salmonella enterica subsp*
Sen-14-1 to Sen-14-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Saintpaul *Salmonella enterica subsp.*
Sen-15-1 to Sen-15-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Saintpaul *Salmonella enterica subsp.*
Sen-16-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*
Sen-17-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*
Sen-18-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*
Sen-19-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*
Sen-20-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*
Sen-21-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*
Sen-22-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*
Sen-23-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*
Sen-24-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*

Sen-25-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*
Sen-26-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*
Sen-27-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*
Sen-28-1 to Sen-28-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Virchow *Salmonella enterica subsp. en*
Sty-1-1 to Sty-1-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Salmonella typhimurium* LT2
Spr-1-1 to Spr-1-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Serratia proteamaculans* 568
Sbo-1-1 to Sbo-1-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella boydii* CDC 3083-94
Sbo-2-1 to Sbo-2-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella boydii* Sb227
Sdy-1-1 to Sdy-1-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella dysenteriae* 1012
Sdy-2-1 to Sdy-2-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella dysenteriae* Sd197
Sfl-1-1 to Sfl-1-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella flexneri* 2a str. 2457T
Sfl-2-1 to Sfl-2-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella flexneri* 2a str. 301
Sfl-3-1 to Sfl-3-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella flexneri* 5 str. 8401
Sso-1-1 to Sso-1-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella sonnei* Ss046
Sgl-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Sodalis glossinidius str. 'morsitans'*
Ybe-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia bercovieri* ATCC 43970
Yen-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia enterocolitica subsp. enterocolitica* 8081
Yfr-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia frederiksenii* ATCC 33641
Yin-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia intermedia* ATCC 29909
Ymo-1-1 to Ymo-1-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia mollaretii* ATCC 43969
Ype-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* Angola
Ype-2-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* Antiqua
Ype-9-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Antiqua str. B42003004
Ype-10-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Antiqua str. E1979001
Ype-11-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Antiqua str. UG05-0454
Ype-12-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Mediaevalis str. K1973002
Ype-13-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Microtus str. 91001
Ype-14-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Orientalis str. F1991016
Ype-15-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Orientalis str. IP275
Ype-16-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Orientalis str. MG05-1020
Ype-3-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* CA88-4125
Ype-4-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* CO92
Ype-5-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* FV-1
Ype-6-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* KIM
Ype-7-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* Nepal516
Ype-8-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* Pestoides F
Yps-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pseudotuberculosis* IP 31758
Yps-2-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pseudotuberculosis* IP 32953
Yps-3-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pseudotuberculosis* PB1/+
Yps-4-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pseudotuberculosis* YPIII
Ppr-1-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Photobacterium profundum* 3TCK
Ppr-2-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Photobacterium profundum* SS9
Van-1-1 to Van-1-2	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Vibrio angustum* S14
Sen-29-1 to Sen-29-4	CVM19633Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Salmonella enterica subsp. enterica serovar* Schwarzengrund str
Sen-30-1 to Sen-30-2	CVM23701Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar 4,[5],12:i:- *Salmonella enter*
Sen-31-1 to Sen-31-2	GA_MM04042433Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Salmonella enterica subsp. enterica serovar* Javiana str
Sen-32-1 to Sen-32-2	HI_N05-537Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Weltevreden *Salmonella en*
Sen-33-1 to Sen-33-2	P125109Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Enteritidis *Salmonella enterica*
Sen-34-1 to Sen-34-2	SC-B67Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Choleraesuis *Salmonella enteri*
Sen-29-1 to Sen-29-4	SL480Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Salmonella enterica subsp. enterica serovar* Schwarzengrund str
env-1 to env-19	environmental samples

43.2 Gene contexts

Each JUMPstart RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one JUMPstart RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 43.3. The accession number of the sequence containing each JUMPstart RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other

accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 43.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	ABEF01026186.1	+	1410	1447	**RNA**→ HcaE (COG4638)→
Bli-1-1	NC_006270.3	-	4124580	4124542	**RNA**→ UgpA (COG1175)→ MalG (COG3833)→ LacA (COG1874)→
Bli-1-2	NC_006322.1	-	4124432	4124394	**RNA**→ UgpA (COG1175)→ MalG (COG3833)→ LacA (COG1874)→
env-2	JCVI_SCAF_1096627929873	-	1579	1544	**RNA**→ ? (GOS3665)→ DltB (COG1696)? (GOS7219)→
Pat-1-1	NC_004547.2	+	1980483	1980520	**RNA**→ DegT_DnrJ_EryC1 (pfam01041)→
Ppr-2-1	NC_006370.1	-	3119947	3119910	**RNA**→ Poly_export (pfam02563)→
Ppr-1-1	NZ_AAPH01000003.1	+	46145	46182	**RNA**→ Poly_export (pfam02563)→
Van-1-1	NZ_AAOJ01000005.1	-	28893	28856	**RNA**→ COG5338 (COG5338)→ Poly_export (pfam02563)→ Mrp (COG0489)GumC (COG3206)→ RfaL (COG3307)→ WcaG (COG0451)→ RfbX (COG2244)→ Glycos_transf_2 (pfam00535)→ Glycos_transf_2 (pfam00535)→ Glycos_transf_2 (pfam00535)→
Asa-1-1	NC_009348.1	+	4144866	4144902	**RNA**→ hypo→
Eal-1-1	NZ_ABKX01000002.1	+	545679	545717	**RNA**→ Glyco_transf_9 (pfam01075)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ Glycos_transf_1 (pfam00534)MGDG_synth (pfam06925)→ RfaJ (COG1442)→ RfaJ (COG1442)→ WaaY (pfam06176)→ lipopolysaccharide core biosynthesis protein RfaZ→ Glycos_transf_1 (pfam00534)→
Sdy-2-1	NC_007606.1	-	3800484	3800446	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)NOG06227 (NOG06227)→ RfaJ (COG1442)→ RfaJ (COG1442)→ WaaY (pfam06176)NOG10262 (NOG10262)→ RfaJ (COG1442)→ Glyco_transf_25 (pfam01755)→ RfaL (COG3307)→
Eco-11-1	NC_009801.1	-	4124151	4124113	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glycos_transf_2 (pfam00535)→ Wzy_C (pfam04932)→
Eco-13-1	NC_009800.1	-	3837796	3837758	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glycos_transf_2 (pfam00535)→ Wzy_C (pfam04932)→
Eco-27-1	NC_002695.1	-	4552074	4552036	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)NOG06227 (NOG06227)→ RfaJ (COG1442)→ WaaY (pfam06176)NOG10262 (NOG10262)→ RfaJ (COG1442)→ Glycos_transf_1 (pfam00534)→
Eco-29-1	NC_011415.1	-	4056405	4056367	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glycos_transf_1 (pfam00534)→
Eco-20-1	NC_011353.1	-	4655178	4655140	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→
Eco-5-1	NC_010468.1	+	83560	83598	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glycos_transf_2 (pfam00535)→ Wzy_C (pfam04932)→
Eco-14-1	NC_002655.2	-	4621057	4621019	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ WaaY (pfam06176)NOG10262 (NOG10262)→ RfaJ (COG1442)→ Glycos_transf_1 (pfam00534)→
Eco-7-1	NZ_AAJT02000039.1	-	11411	11373	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glyco_transf_25 (pfam01755)→ RfaL (COG3307)→
Eco-10-1	NZ_AAJV02000011.1	-	4338	4300	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→
Eco-9-1	NZ_AAJW02000028.1	-	5501	5463	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→
Eco-6-1	NZ_AAJX02000103.1	-	585	547	**RNA**→
Eco-3-1	NZ_AAKB02000001.1	+	2174464	2174502	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glycos_transf_2 (pfam00535)→ Wzy_C (pfam04932)→
Eco-21-1	NZ_ABHK02000001.1	-	1930647	1930609	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glycos_transf_1 (pfam00534)→
Eco-17-1	NZ_ABHL02000001.1	-	1798997	1798959	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→
Eco-16-1	NZ_ABHM02000001.1	-	2918744	2918706	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glycos_transf_1 (pfam00534)→
Eco-19-1	NZ_ABHP01000008.1	+	165958	165996	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→
Eco-18-1	NZ_ABHQ01000003.1	-	355493	355455	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glycos_transf_1 (pfam00534)→
Eco-22-1	NZ_ABHR01000001.1	-	371080	371042	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glycos_transf_1 (pfam00534)→
Eco-23-1	NZ_ABHS01000001.1	-	355145	355107	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glycos_transf_1 (pfam00534)→

Eco-24-1	NZ_ABHT01000001.1	-	354130	354092	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→
Eco-26-1	NZ_ABHU01000002.1	-	373953	373915	**RNA**→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→
Eco-25-1	NZ_ABHW01000001.1	-	355150	355112	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glycos_transf_1 (pfam00534)→
Eco-15-1	NZ_ABJT01000006.1	-	16767	16729	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glycos_transf_1 (pfam00534)→
Eco-28-1	NZ_ABKY01000002.1	-	361134	361096	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glycos_transf_1 (pfam00534)→
Eco-2-1	NC_008253.1	-	3912685	3912647	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glycos_transf_2 (pfam00535)→ RfaL (COG3307)→
Eco-31-1	NC_007946.1	-	4066743	4066705	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glycos_transf_2 (pfam00535)→ RfaL (COG3307)→
Eco-4-1	NC_008563.1	-	4099539	4099501	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glycos_transf_2 (pfam00535)→ RfaL (COG3307)→
Eco-12-1	NZ_AAJU02000002.1	+	147548	147586	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glycos_transf_2 (pfam00535)→ RfaL (COG3307)→
env-3	BABF01000147.1	-	6789	6751	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→
Sdy-1-1	NZ_AAMJ02000001.1	-	141888	141850	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→
Eco-32-1	NC_010473.1	-	3903820	3903782	**RNA**→ Glyco_transf_9 (pfam01075)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ lipopolysaccharide core biosynthesis protein→ Glycos_transf_1 (pfam00534)→ RfaJ (COG1442)→ RfaJ (COG1442)→ WaaY (pfam06176)→ lipopolysaccharide core biosynthesis protein→ RfaF (COG0859)→
Eco-34-1	AC_000091.1	+	3832195	3832233	**RNA**→ Glyco_transf_9 (pfam01075)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ lipopolysaccharide core biosynthesis protein→ Glycos_transf_1 (pfam00534)→ RfaJ (COG1442)→ RfaJ (COG1442)→ WaaY (pfam06176)→ lipopolysaccharide core biosynthesis protein→ RfaF (COG0859)→
Eco-33-1	NC_000913.2	-	3806243	3806205	**RNA**→ Glyco_transf_9 (pfam01075)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)NOG06227 (NOG06227)→ lipopolysaccharide core biosynthesis protein→ Glycos_transf_1 (pfam00534)→ RfaJ (COG1442)→ RfaJ (COG1442)→ WaaY (pfam06176)NOG10262 (NOG10262)→ lipopolysaccharide core biosynthesis protein→ RfaF (COG0859)→
Sfl-3-1	NC_008258.1	+	4000634	4000672	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glycos_transf_1 (pfam00534)→
Sfl-2-1	NC_004337.1	-	3777713	3777675	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)NOG06227 (NOG06227)→ RfaJ (COG1442)→ WaaY (pfam06176)NOG10262 (NOG10262)→ RfaJ (COG1442)→ Glycos_transf_1 (pfam00534)→
Sfl-1-1	NC_004741.1	+	3997504	3997542	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)NOG06227 (NOG06227)→ RfaJ (COG1442)→ WaaY (pfam06176)NOG10262 (NOG10262)→ RfaJ (COG1442)→ Glycos_transf_1 (pfam00534)→
Eco-8-1	NC_004431.1	-	4239213	4239175	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)NOG06227 (NOG06227)→ RfaJ (COG1442)→ RfaJ (COG1442)→ WaaY (pfam06176)NOG10262 (NOG10262)→ RfaJ (COG1442)→ Glycos_transf_2 (pfam00535)→ RfaL (COG3307)→
Sso-1-1	NC_007384.1	+	3959667	3959705	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glycos_transf_2 (pfam00535)→ Wzy_C (pfam04932)→
Eco-30-1	NC_010498.1	-	4046236	4046198	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glycos_transf_2 (pfam00535)→ Wzy_C (pfam04932)→
Sbo-2-1	NC_007613.1	-	3653813	3653775	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)NOG06227 (NOG06227)→ RfaJ (COG1442)→ RfaJ (COG1442)→ WaaY (pfam06176)NOG10262 (NOG10262)→ RfaJ (COG1442)→
Sbo-1-1	NC_010658.1	-	3791037	3790999	**RNA**→ RfaF (COG0859)→ Glycos_transf_1 (pfam00534)→ Kdo (pfam06293)→ RfaJ (COG1442)→ RfaJ (COG1442)→ WaaY (pfam06176)→ RfaJ (COG1442)→ Glycos_transf_2 (pfam00535)→ Wzy_C (pfam04932)→
Ymo-1-1	NZ_AALD01000014.1	+	69626	69663	**RNA**→ Gmd (COG1089)→
Sgl-1-1	NC_007712.1	+	1611345	1611382	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→
Yin-1-1	NZ_AALF01000005.1	-	79831	79794	**RNA**→ WxcM_C (pfam05523)→
Ybe-1-1	NZ_AALC01000045.1	+	35079	35116	**RNA**→
Ymo-1-2	NZ_AALD01000014.1	+	49023	49060	**RNA**→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→
Esa-1-1	NC_009778.1	+	1149747	1149785	**RNA**→ RfbB (COG1088)→ RfbA (COG1209)→
Sbo-2-2	NC_007613.1	-	908336	908298	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→

Eco-30-2	NC_010498.1	+	1018107	1018145	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→ Glycos_transf_2 (pfam00535)→ CysE (COG1045)→ RfaG (COG0438)→ colanic acid polymerase WcaD→ Glycos_transf_2 (pfam00535)→ WbbJ (COG0110)→ Gmd (COG1089)→
Eco-29-2	NC_011415.1	-	2393146	2393108	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→
Eco-10-2	NZ_AAJV02000005.1	-	95395	95357	**RNA**→ ←hypo
env-4	contig37229	+	213	251	**RNA**→ ←hypo
Eco-11-2	NC_009801.1	-	2320565	2320527	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→
Sfl-3-2	NC_008258.1	-	2145723	2145685	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→ Glycos_transf_2 (pfam00535)→ CysE (COG1045)→ RfaG (COG0438)→ putative colanic acid biosynthesis protein→ Glycos_transf_2 (pfam00535)→
Sfl-2-2	NC_004337.1	-	2142820	2142782	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→ Mrp (COG0489)GumC (COG3206)→
Sfl-1-2	NC_004741.1	-	2125685	2125647	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→ Mrp (COG0489)GumC (COG3206)→
Eco-5-2	NC_010468.1	+	1755756	1755794	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→ Glycos_transf_2 (pfam00535)→ CysE (COG1045)→ RfaG (COG0438)→ putative colanic acid polymerase→ Glycos_transf_2 (pfam00535)→
Eco-3-2	NZ_AAKB02000001.1	+	4000405	4000443	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→ Glycos_transf_2 (pfam00535)→ CysE (COG1045)→ RfaG (COG0438)→ colanic acid polymerase WcaD→ Glycos_transf_2 (pfam00535)→ WbbJ (COG0110)→ Gmd (COG1089)→
env-5	contig65962	-	267	229	**RNA**→ ←hypo
Eco-31-2	NC_007946.1	-	2275109	2275071	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→ Mrp (COG0489)GumC (COG3206)→
Eco-27-2	NC_002695.1	-	2812835	2812797	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→ Glycos_transf_2 (pfam00535)→ CysE (COG1045)→ RfaG (COG0438)→ putative colanic acid biosynthesis protein→ Glycos_transf_2 (pfam00535)→ WbbJ (COG0110)→ Gmd (COG1089)→
Eco-14-2	NC_002655.2	-	2883049	2883011	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→ Mrp (COG0489)GumC (COG3206)→ Glycos_transf_2 (pfam00535)→ CysE (COG1045)→ RfaG (COG0438)→ putative colanic acid biosynthesis protein→ Glycos_transf_2 (pfam00535)→ WbbJ (COG0110)→ Gmd (COG1089)→
Eco-24-2	NZ_ABHT01000012.1	-	35525	35487	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→ Glycos_transf_2 (pfam00535)→ CysE (COG1045)→ RfaG (COG0438)→ colanic acid polymerase WcaD→ Glycos_transf_2 (pfam00535)→ WbbJ (COG0110)→ Gmd (COG1089)→
Eco-26-2	NZ_ABHU01000016.1	-	69731	69693	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→ Glycos_transf_2 (pfam00535)→ CysE (COG1045)→ RfaG (COG0438)→ colanic acid polymerase WcaD→ Glycos_transf_2 (pfam00535)→ WbbJ (COG0110)→ Gmd (COG1089)→
Eco-25-2	NZ_ABHW01000009.1	+	90525	90563	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→ Glycos_transf_2 (pfam00535)→ CysE (COG1045)→ RfaG (COG0438)→ colanic acid polymerase WcaD→ Glycos_transf_2 (pfam00535)→ WbbJ (COG0110)→ Gmd (COG1089)→
Eco-28-2	NZ_ABKY01000013.1	+	93495	93533	**RNA**→ ←hypo
Eco-8-2	NC_004431.1	-	2424993	2424955	**RNA**→ Poly_export (pfam02563)→ hypo→
Sdy-1-2	NZ_AAMJ02000002.1	+	193675	193713	**RNA**→ Mrp (COG0489)GumC (COG3206)→ Glycos_transf_2 (pfam00535)→ CysE (COG1045)→ COG3385 (COG3385)→ maltose O-acetyltransferase→ RfaG (COG0438)→ colanic acid polymerase WcaD→ Glycos_transf_2 (pfam00535)→ WbbJ (COG0110)→ Gmd (COG1089)→ WcaG (COG0451)→
Eco-32-2	NC_010473.1	-	2226492	2226454	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→ Glycos_transf_2 (pfam00535)→ CysE (COG1045)→ RfaG (COG0438)→ colanic acid polymerase→ Glycos_transf_2 (pfam00535)→ WbbJ (COG0110)→ Gmd (COG1089)→
Eco-34-2	AC_000091.1	-	2139597	2139559	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→ Glycos_transf_2 (pfam00535)→ CysE (COG1045)→ RfaG (COG0438)→ predicted colanic acid polymerase→ Glycos_transf_2 (pfam00535)→ WbbJ (COG0110)→ Gmd (COG1089)→
Eco-33-2	NC_000913.2	-	2135484	2135446	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→ Glycos_transf_2 (pfam00535)→ CysE (COG1045)→ RfaG (COG0438)→ predicted colanic acid polymerase→ Glycos_transf_2 (pfam00535)→ WbbJ (COG0110)→ Gmd (COG1089)→
Eco-1-1	NZ_AAMK02000002.1	-	71951	71913	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→ Glycos_transf_2 (pfam00535)→ CysE (COG1045)→ RfaG (COG0438)→ colanic acid polymerase WcaD→ Glycos_transf_2 (pfam00535)→ WbbJ (COG0110)→ Gmd (COG1089)→

Eco-2-2	NC_008253.1	-	2175301	2175263	RNA→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→ Glycos_transf_2 (pfam00535)→ CysE (COG1045)→ RfaG (COG0438)→ putative colanic acid biosynthesis protein→ Glycos_transf_2 (pfam00535)→ WbbJ (COG0110)→ Gmd (COG1089)→
Eco-12-2	NZ_AAJU02000001.1	-	159686	159648	RNA→ ←hypo
Eco-4-2	NC_008563.1	-	2224614	2224576	RNA→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→
Eco-7-2	NZ_AAJT02000072.1	+	24123	24161	RNA→ ←hypo
env-6	BABF01002607.1	-	437	399	RNA→ hypo→
Sen-13-1	NC_010102.1	+	842962	843000	RNA→ ←hypo
Sbo-1-2	NC_010658.1	+	1072629	1072667	RNA→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→
Eco-9-2	NZ_AAJW02000014.1	-	20958	20920	RNA→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→ Glycos_transf_2 (pfam00535)→
Eco-6-2	NZ_AAJX02000021.1	+	16863	16901	RNA→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→ Glycos_transf_2 (pfam00535)→ CysE (COG1045)→
Eco-3-3	NZ_AAKB02000001.1	+	4026168	4026206	RNA→ NorM (COG0534)→ FrhB (COG1035)→ pyruvyl transferase→ Glycos_transf_2 (pfam00535)→ OCH1 (COG3774)→ putative membrane protein→ Glycos_transf_2 (pfam00535)→ Glf (COG0562)→ COG1216 (COG1216)→
env-7	BAAW01012183.1	-	306	268	RNA→ hypo→
Pat-1-2	NC_004547.2	+	1606683	1606720	RNA→ Rfe (COG0472)→
Yen-1-1	NC_008800.1	-	3358260	3358223	RNA→ GCD1 (COG1208)→ WcaG (COG0451)→
Spr-1-1	NC_009832.1	+	1725327	1725364	RNA→ GalU (COG1210)→ GalU (COG1210)→
Yfr-1-1	NZ_AALE01000009.1	-	42200	42163	RNA→ Rfe (COG0472)→
Ype-13-1	NC_005810.1	+	879981	880018	RNA→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ HemY (COG1232)→
Ype-1-1	NC_010159.1	+	3041298	3041335	RNA→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→
Ype-2-1	NC_008150.1	-	2905292	2905255	RNA→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ hypo→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ HemY (COG1232)→
Ype-7-1	NC_008149.1	+	1130002	1130039	RNA→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ hypo→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ HemY (COG1232)→
Ype-8-1	NC_009381.1	-	3121763	3121726	RNA→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ HemY (COG1232)→
Yps-4-1	NC_010465.1	-	3492037	3492000	RNA→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→
Yps-3-1	NC_010634.1	+	1180681	1180718	RNA→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ HemY (COG1232)→
Ype-6-1	NC_004088.1	+	1202152	1202189	RNA→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→
Ype-4-1	NC_003143.1	-	3473620	3473583	RNA→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→
Ype-15-1	NZ_AAOS02000001.1	-	131115	131078	RNA→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→
Ype-5-1	NZ_AAUB01000101.1	+	5240	5277	RNA→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ putative CDP-D-glucose-4,6-dehydratase→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ HemY (COG1232)→ O-unit flippase-like protein→ Glycos_transf_2 (pfam00535)→
Ype-11-1	NZ_AAYR01000001.1	+	860626	860663	RNA→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→
Ype-16-1	NZ_AAYS01000016.1	+	9718	9755	RNA→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→
Ype-12-1	NZ_AAYT01000001.1	+	525982	526019	RNA→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→
Ype-9-1	NZ_AAYU01000006.1	-	10985	10948	RNA→ GCD1 (COG1208)→
Ype-10-1	NZ_AAYV01000011.1	+	176906	176943	RNA→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→
Ype-14-1	NZ_ABAT01000020.1	+	29861	29898	RNA→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→
Ype-3-1	NZ_ABCD01000001.1	+	176071	176108	RNA→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→
Yps-2-1	NC_006155.1	+	1198616	1198653	RNA→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ HemY (COG1232)→ putative O-unit flippase→ Glycos_transf_2 (pfam00535)→
Yps-1-1	NC_009708.1	-	3437807	3437770	RNA→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ HemY (COG1232)→
Sen-17-1	NZ_CAAY01005146.1	-	1391	1353	RNA→
Sen-27-1	NC_004631.1	+	867247	867285	RNA→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ Epimerase (pfam01370)→ RfbX (COG2244)→

Sen-18-1	NC_003198.1	-	2136681	2136643	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ Epimerase (pfam01370)→ RfbX (COG2244)→
Sen-16-1	NZ_CAAQ01001590.1	+	136	174	**RNA**→
Sen-19-1	NZ_CAAR01000188.1	+	906	944	**RNA**→
Sen-22-1	NZ_CAAU01003370.1	-	3544	3506	**RNA**→
Sen-25-1	NZ_CAAW01000978.1	-	3577	3539	**RNA**→
Sen-26-1	NZ_CAAX01002844.1	+	5441	5479	**RNA**→
Sen-24-1	NZ_CAAZ01000140.1	-	5878	5840	**RNA**→
Sen-20-1	NZ_CAAS01000602.1	+	1454	1492	**RNA**→
Sen-21-1	NZ_CAAT01000069.1	+	43133	43171	**RNA**→
Sen-23-1	NZ_CAAV01001296.1	+	322	360	**RNA**→
Sen-28-1	NZ_ABFH02000001.1	+	2947499	2947537	**RNA**→ O-antigen polymerase→ RfaG (COG0438)→ Glycos_transf_1 (pfam00534)→ RfaG (COG0438)→ ManC (COG0662)ManC (COG0836)→ ManB (COG1109)→
Sen-13-2	NC_010102.1	+	867107	867145	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ RfbX (COG2244)→ Glycos_transf_2 (pfam00535)→
Sen-4-1	NC_011205.1	-	2310345	2310307	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→
Sen-7-1	NC_011083.1	-	2226311	2226273	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ RfbX (COG2244)→ Glycos_transf_2 (pfam00535)→
Sen-15-1	NZ_ABAN01000008.3	-	43446	43408	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ RfbX (COG2244)→ Glycos_transf_2 (pfam00535)→
Sen-8-1	NZ_ABEL01000004.1	-	84390	84352	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ RfbX (COG2244)→ Glycos_transf_2 (pfam00535)→
Sen-2-1	NC_006511.1	+	859951	859989	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ Epimerase (pfam01370)→ RfbX (COG2244)→
Sen-1-1	NC_011147.1	+	859896	859934	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ Epimerase (pfam01370)→ RfbX (COG2244)→
Sen-11-1	NC_011080.1	-	2190132	2190094	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→
Sen-12-1	NZ_ABEW01000006.1	+	349979	350017	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→
Sen-6-1	NZ_ABFG01000001.1	+	143362	143400	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→
Sen-29-1	NC_011094.1	-	2200235	2200197	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ RfbX (COG2244)→ Glycos_transf_2 (pfam00535)→
Sen-34-1	NC_006905.1	-	2191196	2191158	**RNA**→ O-antigen polymerase→ RfaG (COG0438)NOG20060 (NOG20060)→ Glycos_transf_1 (pfam00534)→ RfaG (COG0438)→ ManC (COG0662)ManC (COG0836)→ ManB (COG1109)→
Sen-32-1	NZ_ABFF01000001.1	+	450375	450413	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→
Sen-33-1	NC_011294.1	-	2183328	2183290	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ Epimerase (pfam01370)→ RfbX (COG2244)→
Sen-5-1	NC_011274.1	-	2174336	2174298	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ Epimerase (pfam01370)→ RfbX (COG2244)→
Sen-31-1	NZ_ABEH02000005.1	-	99573	99535	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→
Sen-3-1	NC_011149.1	-	2136558	2136520	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ RfbX (COG2244)→ Glycos_transf_2 (pfam00535)→
Sen-10-1	NZ_ABAK02000001.1	-	1568305	1568267	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→
Sen-9-1	NZ_ABEI01000001.1	-	92205	92167	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→

Sen-29-3	NZ_ABEJ01000005.1	-	54760	54722	**RNA**→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ RfbX (COG2244)→ Glycos_transf_2 (pfam00535)→
Sty-1-1	NC_003197.1	-	2178692	2178654	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ RfbX (COG2244)→ Glycos_transf_2 (pfam00535)→
Sen-30-1	NZ_ABAO01000005.1	-	82478	82440	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ RfbX (COG2244)→ Glycos_transf_2 (pfam00535)→
Sen-14-1	NZ_ABAM02000001.1	-	2637791	2637753	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ fer2 (cd00207)UbiB (COG0543)→ GCD1 (COG1208)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ WcaG (COG0451)→ RfbX (COG2244)→ Glycos_transf_2 (pfam00535)→
Sbo-2-3	NC_007613.1	-	883782	883744	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ putative glycosyl transferase→ putative glycosyl transferase→ Polysacc_synt (pfam01943)→ RfaG (COG0438)→ O-antigen polymerase→ RfaG (COG0438)→
Eco-2-3	NC_008253.1	-	2150637	2150599	**RNA**→ Polysacc_synt (pfam01943)→ antigen polymerase→ Glycos_transf_2 (pfam00535)→ RfaG (COG0438)→ glycosyltransferase→ GalE (COG1087)→ RfaG (COG0438)→ ManC (COG0662)ManC (COG0836)→ ManB (COG1109)→
Eco-8-3	NC_004431.1	-	2400429	2400391	**RNA**→ RfbX (COG2244)→
Eco-12-3	NZ_AAJU02000001.1	-	135022	134984	**RNA**→ RfbX (COG2244)→ hypothetical antigen polymerase→
Sbo-1-3	NC_010658.1	+	1098308	1098346	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→
Eco-10-3	NZ_AAJV02000005.1	-	70839	70801	**RNA**→ RfbB (COG1088)→ RfbA (COG1209)→ WxcM_C (pfam05523)→ RimL (COG1670)→ DegT_DnrJ_EryC1 (pfam01041)→ RfbX (COG2244)→ Glycos_transf_2 (pfam00535)→ Glycos_transf_2 (pfam00535)→ Wzy→ Glycos_transf_1 (pfam00534)→
Eco-7-3	NZ_AAJT02000037.1	+	11951	11989	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ Polysacc_synt (pfam01943)→ Wzy→ COG1216 (COG1216)→ COG1216 (COG1216)→ RfaG (COG0438)→
Eco-4-3	NC_008563.1	-	2200336	2200298	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ Polysacc_synt (pfam01943)→
Sdy-2-2	NC_007606.1	+	2006669	2006707	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ Polysacc_synt (pfam01943)→ O-antigen polymerase→ COG1216 (COG1216)→ COG1216 (COG1216)→
Eco-32-3	NC_010473.1	-	2202207	2202169	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ Polysacc_synt (pfam01943)→ Glf (COG0562)→ O-antigen polymerase→ hypo→ WbbJ (COG0110)→ Glycos_transf_1 (pfam00534)→
Eco-34-3	AC_000091.1	-	2115312	2115274	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ Polysacc_synt (pfam01943)→ Glf (COG0562)→ O-antigen polymerase→ hypo→ WbbJ (COG0110)→ Glycos_transf_1 (pfam00534)→
Eco-33-3	NC_000913.2	-	2111199	2111161	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ Polysacc_synt (pfam01943)→ Glf (COG0562)→ O-antigen polymerase→ conserved protein→ WbbJ (COG0110)→ Glycos_transf_1 (pfam00534)→ → →
Eco-11-3	NC_009801.1	-	2296098	2296060	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ Glycos_transf_2 (pfam00535)→ COG1216 (COG1216)→ hypo→
Eco-29-3	NC_011415.1	-	2368874	2368836	**RNA**→ COG5433 (COG5433)→
Eco-5-3	NC_010468.1	+	1781475	1781513	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ Nitro_FMN_reductase (cd02062)→ putative transmembrane protein→ hypo→ RfaG (COG0438)→ hypo→ COG1216 (COG1216)→
Eco-1-2	NZ_AAMK02000002.1	-	45579	45541	**RNA**→ RfbB (COG1088)→ RfbA (COG1209)→
Eco-31-3	NC_007946.1	-	2250240	2250202	**RNA**→ ←hypo
env-8	RicEn_FSXC77884_b1	-	424	386	**RNA**→ dTDP-glucose 4,6-dehydratase (EC 4.2.1.46)→
Sfl-3-3	NC_008258.1	-	2120667	2120629	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ Polysacc_synt (pfam01943)→
Sfl-2-3	NC_004337.1	-	2118802	2118764	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ Polysacc_synt (pfam01943)→ COG1216 (COG1216)→ Glycos_transf_2 (pfam00535)→ O-antigen polymerase→ COG2246 (COG2246)→ Glycos_transf_2 (pfam00535)→ hypo→
Sfl-1-3	NC_004741.1	-	2101668	2101630	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ dTDP_sugar_isom (pfam00908)→ Polysacc_synt (pfam01943)→ COG1216 (COG1216)→ Glycos_transf_2 (pfam00535)→ O-antigen polymerase→ COG2246 (COG2246)→ Glycos_transf_2 (pfam00535)→ hypo→

Eca-1-1	NZ_ABWM01000050.1	-	11062	11024	**RNA**→ RfbB (COG1088)→ RfbA (COG1209)→ RfbX (COG2244)→ DegT_DnrJ_EryC1 (pfam01041)→ WbbJ (COG0110)→ Glycos_transf_2 (pfam00535)→
Eco-30-3	NC_010498.1	+	1043408	1043446	**RNA**→ Glycos_transf_2 (pfam00535)→ COG5610 (COG5610)→ Glycos_transf_2 (pfam00535)→ hypo→ RfaG (COG0438)→ Polysacc_synt (pfam01943)→ hypo→ IS1 transposase orfA→ Transposase_27 (pfam03400)→
env-9	BABF01000416.1	-	6317	6279	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→ WxcM_C (pfam05523)→ WbbJ (COG0110)→ DegT_DnrJ_EryC1 (pfam01041)→ RfbX (COG2244)→
Eco-9-3	NZ_AAJW02000014.1	-	6070	6032	**RNA**→ Glycos_transf_1 (pfam00534)→ Gmd (COG1089)→ NUDIX (pfam00293)→ ManC (COG0662)ManC (COG0836)→
Eco-6-3	NZ_AAJX02000021.1	+	30437	30475	**RNA**→ Glycos_transf_1 (pfam00534)→ Gmd (COG1089)→ NUDIX (pfam00293)→ ManC (COG0662)ManC (COG0836)→ ManB (COG1109)→ WcaG (COG0451)→ DegT_DnrJ_EryC1 (pfam01041)→ Polysacc_synt (pfam01943)→ O antigen polymerase Wzy→
Esa-1-2	NC_009778.1	+	1124151	1124188	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→
Eca-1-2	NZ_ABWM01000051.1	-	4002	3965	**RNA**→ Poly_export (pfam02563)→ LMWPc (pfam01451)→ Mrp (COG0489)GumC (COG3206)→
env-10	BISONP_C5880	+	997	1034	**RNA**→ Poly_export (pfam02563)→
env-11	BAAV01013333.1	+	262	299	**RNA**→ Poly_export (pfam02563)→
Esp-1-1	NC_009436.1	-	2879815	2879778	**RNA**→ Poly_export (pfam02563)→ LMWPc (cd00115)→ Mrp (COG0489)GumC (COG3206)→
Sen-2-2	NC_006511.1	+	4232657	4232694	**RNA**→ hypo→ putative integral membrane protein→ TolC (COG1538)→ HlyD (pfam00529)→ putative inner membrane protein→
Sen-11-2	NC_011080.1	+	4468905	4468942	**RNA**→ hypo→ putative integral membrane protein→
Sen-1-2	NC_011147.1	+	4228344	4228381	**RNA**→ hypo→ putative integral membrane protein→ TolC (COG1538)→ HlyD (pfam00529)→ putative inner membrane protein→
Sen-33-2	NC_011294.1	+	4334958	4334995	**RNA**→ hypo→ putative integral membrane protein→ TolC (COG1538)→ HlyD (pfam00529)→ large repetitive protein→
Sen-10-2	NZ_ABAK02000001.1	+	3783277	3783314	**RNA**→ ←hypo
Sen-9-2	NZ_ABEI01000006.1	+	77805	77842	**RNA**→ ←hypo
Sen-34-2	NC_006905.1	+	4412354	4412391	**RNA**→ hypo→ putative methyl-accepting chemotaxis protein→ TolC (COG1538)→ HlyD (pfam00529)→ putative inner membrane protein→
Sen-3-2	NC_011149.1	+	4397367	4397404	**RNA**→ ←hypo
Sen-4-2	NC_011205.1	+	4493196	4493233	**RNA**→ hypo→ putative methyl-accepting chemotaxis protein→
Sen-15-2	NZ_ABAN01000019.1	-	3643	3606	**RNA**→ hypo→ putative integral membrane protein→
Sen-7-2	NC_011083.1	+	4473688	4473725	**RNA**→ hypo→ putative integral membrane protein→
Sen-8-2	NZ_ABEL01000009.1	+	78446	78483	**RNA**→ hypo→ putative integral membrane protein→
Sen-6-2	NZ_ABFG01000009.1	-	155732	155695	**RNA**→ hypo→ putative integral membrane protein→
Sen-29-2	NC_011094.1	+	4368291	4368328	**RNA**→ putative inner membrane or exported→ putative integral membrane protein→
Sen-31-2	NZ_ABEH02000001.1	+	66400	66437	**RNA**→ hypo→ putative integral membrane protein→
Sen-29-4	NZ_ABEJ01000008.1	+	40873	40910	**RNA**→ hypo→ putative integral membrane protein→
Sen-14-2	NZ_ABAM02000001.1	+	113856	113893	**RNA**→ hypo (cons)→ putative methyl-accepting chemotaxis protein→
Sty-1-2	NC_003197.1	+	4477737	4477774	**RNA**→ hypo→ putative methyl-accepting chemotaxis protein→ TolC (COG1538)→ HlyD (pfam00529)→ putative inner membrane protein→
Sen-30-2	NZ_ABAO01000007.1	+	77848	77885	**RNA**→ hypo→ putative methyl-accepting chemotaxis protein→
Sen-5-2	NC_011274.1	+	4312866	4312903	**RNA**→ hypo→ putative integral membrane protein→ TolC (COG1538)→ HlyD (pfam00529)→ large repetitive protein→
Sen-28-2	NZ_ABFH02000001.1	-	664795	664758	**RNA**→ hypo→ putative integral membrane protein→
Sen-32-2	NZ_ABFF01000011.1	-	111812	111775	**RNA**→ hypo→ putative integral membrane protein→ TolC (COG1538)→ HlyD (pfam00529)→ putative Ig domain protein→
Sen-13-3	NC_010102.1	+	4403693	4403730	**RNA**→ ←hypo
Sen-12-2	NZ_ABEW01000009.1	+	78780	78817	**RNA**→ hypo (cons)→ putative integral membrane protein→
env-12	JCVI_SCAF_1096628049729	-	117	80	**RNA**→
Spr-1-2	NC_009832.1	-	3180995	3180958	**RNA**→ Rfe (COG0472)→ Poly_export (pfam02563)→ Mrp (COG0489)GumC (COG3206)→ Polysacc_synt (pfam01943)→ hypo→
Sso-1-2	NC_007385.1	+	193310	193347	**RNA**→ WzzB (COG3765)→ WecC (COG0677)→ Epimerase (pfam01370)→
env-13	JCVI_SCAF_1096627390305	+	10333	10370	**RNA**→ ? (GOS12037713)→
Ahy-1-1	NC_008570.1	+	4001368	4001405	**RNA**→ ABC_tran (pfam00005)→
Ahy-1-2	NC_008570.1	-	2238472	2238435	**RNA**→ hypo→

Ahy-1-3	NC_008570.1	-	3246313	3246279	**RNA**→ Rfe (COG0472)→ RimK (COG0189)→ Ald (COG0686)→ Polysacc_synt (pfam01943)→ RfaG (COG0438)→ putative glycosyl transferase→ Glycos_transf_2 (pfam00535)→ hypo→ Epimerase (pfam01370)→ Ugd (COG1004)→
Ahy-1-4	NC_008570.1	-	3268141	3268104	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→
env-14	JCVI_SCAF_1096627187753	-	5472	5435	**RNA**→ RfbB (COG1088)? (GOS4294)→
Asa-1-2	NC_009348.1	+	1482254	1482291	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→
env-15	JCVI_SCAF_1096627199953	+	23504	23541	**RNA**→ PldB (COG2267)? (GOS11241327)→
Ahy-1-5	NC_008570.1	+	3999712	3999749	**RNA**→ PldB (COG2267)→
Asa-1-3	NC_009348.1	+	3828635	3828672	**RNA**→ Glyoxalase (pfam00903)→
env-16	JCVI_SCAF_1096627193069	-	15474	15437	**RNA**→ ? (GOS11649279)→
Ahy-1-6	NC_008570.1	+	4017751	4017788	**RNA**→ Glyoxalase (pfam00903)→ COG2865 (COG2865)→
Asa-1-4	NC_009348.1	+	3822863	3822900	**RNA**→ hypo→
Asa-1-5	NC_009348.1	+	2746012	2746048	**RNA**→ hypo→
Van-1-2	NZ_AAOJ01000002.1	-	773534	773497	**RNA**→ Poly_export (pfam02563)→ Mrp (COG0489)GumC (COG3206)→
Ssp-1-1	NC_008321.1	+	1523017	1523054	**RNA**→ Poly_export (pfam02563)DUF1017 (pfam06251)→
Son-1-1	NC_004347.1	-	3334245	3334208	**RNA**→ Poly_export (pfam02563)DUF1017 (pfam06251)→
Spu-1-1	NZ_AAWY01000010.1	+	75852	75889	**RNA**→ Poly_export (pfam02563)DUF1017 (pfam06251)→
Ssp-2-1	NC_008322.1	+	1598550	1598587	**RNA**→ Poly_export (pfam02563)DUF1017 (pfam06251)→
env-17	JCVI_SCAF_1096627193923	-	129019	128982	**RNA**→ Poly_export (pfam02563)DUF1017 (pfam06251)? (GOS987)→
Sfr-1-1	NC_008345.1	+	1634047	1634084	**RNA**→ Ugd (COG1004)→ Epimerase (pfam01370)→ WecG (COG1922)→
Asa-1-6	NC_009348.1	+	3267272	3267309	**RNA**→ COG1943 (COG1943)→ hypo→
Asa-1-7	NC_009348.1	-	1331858	1331821	**RNA**→ hypo→
Ahy-1-7	NC_008570.1	-	1408724	1408687	**RNA**→ hypo→ hypo→
Asa-1-8	NC_009348.1	-	1330418	1330381	**RNA**→ hypo→
env-18	JCVI_SCAF_1096627187969	-	600	563	**RNA**→
Sfr-1-2	NC_008345.1	-	3427572	3427535	**RNA**→ Poly_export (pfam02563)DUF1017 (pfam06251)→
Sam-1-1	NC_008700.1	-	2730356	2730319	**RNA**→ Poly_export (pfam02563)DUF1017 (pfam06251)→
Ahy-1-8	NC_008570.1	+	3395992	3396029	**RNA**→ COG2357 (COG2357)→
Ahy-1-9	NC_008570.1	-	1556882	1556845	**RNA**→ FruA (COG1299)PtsN (COG1762)PTS_IIB_fruc (pfam02379)→
env-19	JCVI_SCAF_1096628082004	-	476	439	**RNA**→

43.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 43.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one JUMPstart RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

cd00115 (1) Substituted updates: Aug 22, 2001
cd00207 (35) 2Fe-2S iron-sulfur cluster binding domain.
cd02062 (1) Proteins of this family catalyze the reduction of flavin or nitrocompounds using NAD(P)H as electron donor in a obligatory two-electron transfer, utilizing FMN or FAD as cofactor.
COG0110 (18) Acetyltransferase (isoleucine patch superfamily) [General function prediction only]
COG0189 (1) Glutathione synthase/Ribosomal protein S6 modification enzyme (glutaminyl transferase) [Coenzyme metabolism / Translation, ribosomal structure and biogenesis]
COG0438 (27) Glycosyltransferase [Cell envelope biogenesis, outer membrane]
COG0451 (55) Nucleoside-diphosphate-sugar epimerases [Cell envelope biogenesis, outer membrane / Carbohydrate transport and metabolism]
COG0472 (4) UDP-N-acetylmuramyl pentapeptide phosphotransferase/UDP-N-acetylglucosamine-1-phosphate transferase [Cell envelope biogenesis, outer membrane]
COG0489 (35) ATPases involved in chromosome partitioning [Cell division and chromosome partitioning]
COG0534 (1) Na+-driven multidrug efflux pump [Defense mechanisms]
COG0543 (35) 2-polyprenylphenol hydroxylase and related flavodoxin oxidoreductases [Coenzyme metabolism / Energy production and conversion]
COG0562 (4) UDP-galactopyranose mutase [Cell envelope biogenesis, outer membrane]
COG0662 (5) Mannose-6-phosphate isomerase [Carbohydrate transport and metabolism]
COG0677 (1) UDP-N-acetyl-D-mannosaminuronate dehydrogenase [Cell envelope biogenesis, outer membrane]
COG0686 (1) Alanine dehydrogenase [Amino acid transport and metabolism]

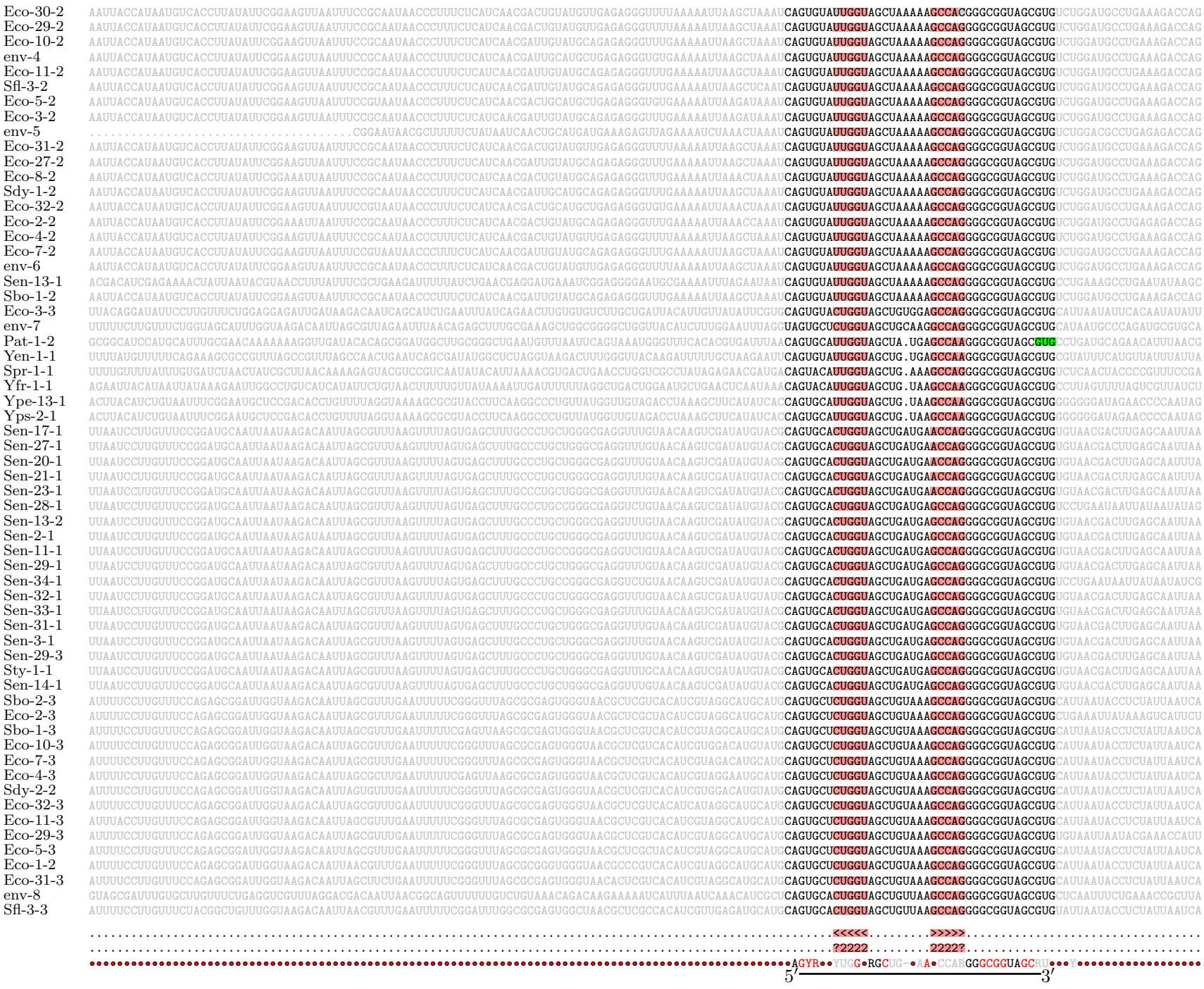
COG0836 (5) Mannose-1-phosphate guanylyltransferase [Cell envelope biogenesis, outer membrane]
COG0859 (40) ADP-heptose:LPS heptosyltransferase [Cell envelope biogenesis, outer membrane]
COG1004 (2) Predicted UDP-glucose 6-dehydrogenase [Cell envelope biogenesis, outer membrane]
COG1035 (1) Coenzyme F420-reducing hydrogenase, beta subunit [Energy production and conversion]
COG1045 (16) Serine acetyltransferase [Amino acid transport and metabolism]
COG1087 (1) UDP-glucose 4-epimerase [Cell envelope biogenesis, outer membrane]
COG1088 (44) dTDP-D-glucose 4,6-dehydratase [Cell envelope biogenesis, outer membrane]
COG1089 (16) GDP-D-mannose dehydratase [Cell envelope biogenesis, outer membrane]
COG1091 (40) dTDP-4-dehydrorhamnose reductase [Cell envelope biogenesis, outer membrane]
COG1109 (4) Phosphomannomutase [Carbohydrate transport and metabolism]
COG1175 (2) ABC-type sugar transport systems, permease components [Carbohydrate transport and metabolism]
COG1208 (38) Nucleoside-diphosphate-sugar pyrophosphorylase involved in lipopolysaccharide biosynthesis/translation initiation factor 2B, gamma/epsilon subunits (eIF-2Bgamma/eIF-2Bepsilon) [Cell envelope biogenesis, outer membrane / Translation, ribosomal structure and biogenesis]
COG1209 (42) dTDP-glucose pyrophosphorylase [Cell envelope biogenesis, outer membrane]
COG1210 (2) UDP-glucose pyrophosphorylase [Cell envelope biogenesis, outer membrane]
COG1216 (9) Predicted glycosyltransferases [General function prediction only]
COG1232 (8) Protoporphyrinogen oxidase [Coenzyme metabolism]
COG1299 (1) Phosphotransferase system, fructose-specific IIC component [Carbohydrate transport and metabolism]
COG1442 (92) Lipopolysaccharide biosynthesis proteins, LPS:glycosyltransferases [Cell envelope biogenesis, outer membrane]
COG1538 (7) Outer membrane protein [Cell envelope biogenesis, outer membrane / Intracellular trafficking and secretion]
COG1670 (1) Acetyltransferases, including N-acetylases of ribosomal proteins [Translation, ribosomal structure and biogenesis]
COG1696 (1) Predicted membrane protein involved in D-alanine export [Cell envelope biogenesis, outer membrane]
COG1762 (1) Phosphotransferase system mannitol/fructose-specific IIA domain (Ntr-type) [Carbohydrate transport and metabolism / Signal transduction mechanisms]
COG1874 (2) Beta-galactosidase [Carbohydrate transport and metabolism]
COG1922 (1) Teichoic acid biosynthesis proteins [Cell envelope biogenesis, outer membrane]
COG1943 (1) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
COG2244 (22) Membrane protein involved in the export of O-antigen and teichoic acid [General function prediction only]
COG2246 (2) Predicted membrane protein [Function unknown]
COG2267 (2) Lysophospholipase [Lipid metabolism]
COG2357 (1) Uncharacterized protein conserved in bacteria [Function unknown]
COG2865 (1) Predicted transcriptional regulator containing an HTH domain and an uncharacterized domain shared with the mammalian protein Schlafen [Transcription]
COG3206 (35) Uncharacterized protein involved in exopolysaccharide biosynthesis [Cell envelope biogenesis, outer membrane]
COG3307 (8) Lipid A core - O-antigen ligase and related enzymes [Cell envelope biogenesis, outer membrane]
COG3385 (1) FOG: Transposase and inactivated derivatives [DNA replication, recombination, and repair]
COG3765 (1) Chain length determinant protein [Cell envelope biogenesis, outer membrane]
COG3774 (1) Mannosyltransferase OCH1 and related enzymes [Cell envelope biogenesis, outer membrane]
COG3833 (2) ABC-type maltose transport systems, permease component [Carbohydrate transport and metabolism]
COG4638 (1) Phenylpropionate dioxygenase and related ring-hydroxylating dioxygenases, large terminal subunit [Inorganic ion transport and metabolism / General function prediction only]
COG5338 (1) Uncharacterized protein conserved in bacteria [Function unknown]
COG5433 (1) Transposase [DNA replication, recombination, and repair]
COG5610 (1) Predicted hydrolase (HAD superfamily) [General function prediction only]
GOS11241327 (1) ?
GOS11649279 (1) ?
GOS12037713 (1) ?
GOS3665 (1) ?
GOS4294 (1) ?
GOS7219 (1) ?
GOS987 (1) ?
NOG06227 (7)
NOG10262 (8)
NOG20060 (1)
pfam00005 (1) ABC transporter.
pfam00293 (2) NUDIX domain.
pfam00529 (7) HlyD family secretion protein.
pfam00534 (69) Glycosyl transferases group 1.
pfam00535 (73) Glycosyl transferase.
pfam00903 (2) Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily.
pfam00908 (36) dTDP-4-dehydrorhamnose 3,5-epimerase.
pfam01041 (31) DegT/DnrJ/EryC1/StrS aminotransferase family.
pfam01075 (4) Glycosyltransferase family 9 (heptosyltransferase).
pfam01370 (9) NAD dependent epimerase/dehydratase family.
pfam01451 (27) Low molecular weight phosphotyrosine protein phosphatase.
pfam01755 (2) Glycosyltransferase family 25 (LPS biosynthesis protein).
pfam01943 (15) Polysaccharide biosynthesis protein.
pfam02379 (1) PTS system, Fructose specific IIB subunit.
pfam02563 (43) Polysaccharide biosynthesis/export protein.
pfam03400 (1) IS1 transposase.
pfam04932 (7) O-Antigen Polymerase.
pfam05523 (3) WxcM-like, C-terminal.
pfam06176 (34) Lipopolysaccharide core biosynthesis protein (WaaY).
pfam06251 (7) Protein of unknown function (DUF1017).
pfam06293 (42) Lipopolysaccharide kinase (Kdo/WaaP) family.
pfam06925 (1) Monogalactosyldiacylglycerol (MGDG) synthase.

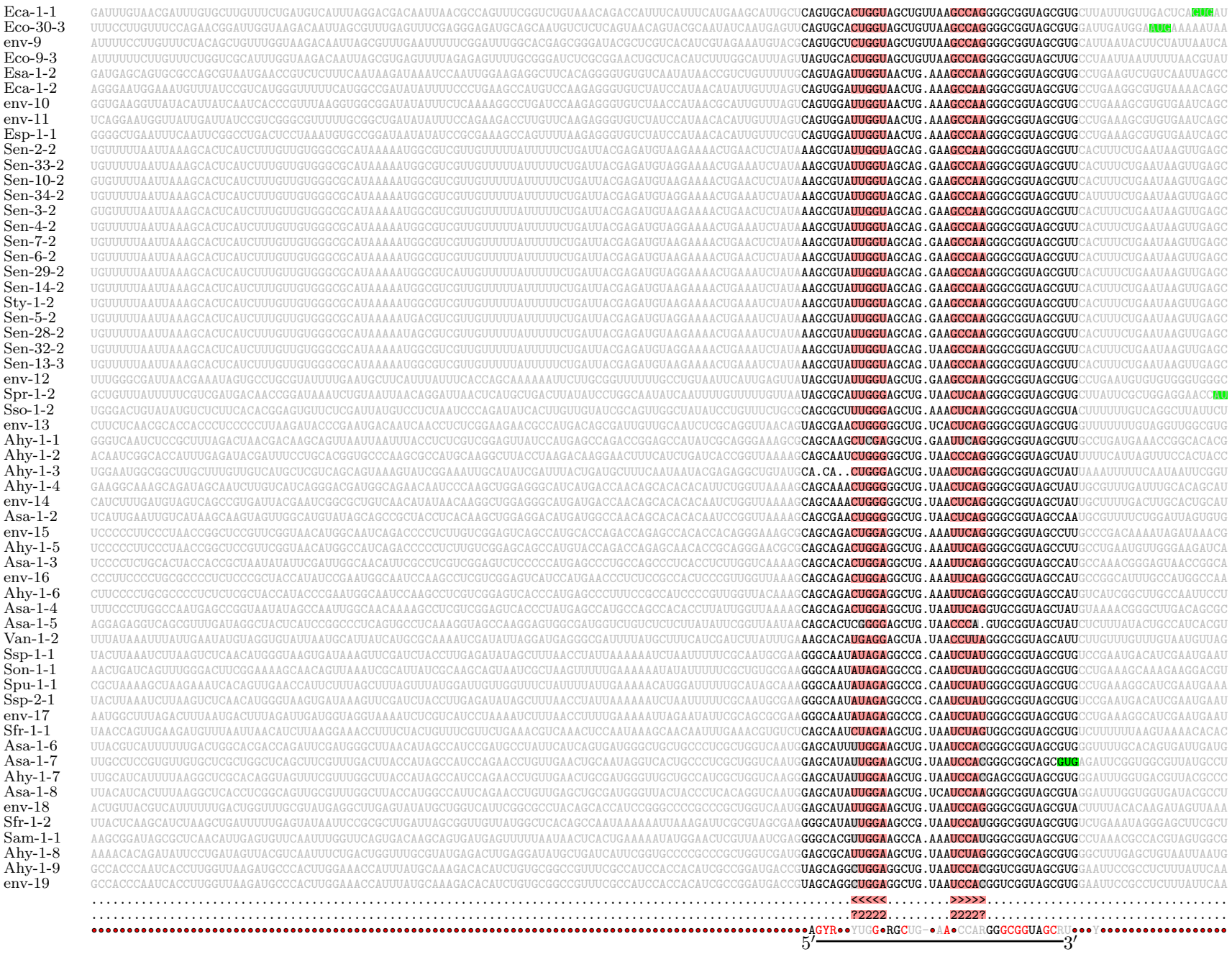
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Bli-1-2,

Eco-1-1, Eco-10-1, Eco-11-1, Eco-12-1, Eco-12-2, Eco-12-3, Eco-13-1, Eco-14-1, Eco-14-2, Eco-15-1, Eco-16-1, Eco-17-1, Eco-18-1, Eco-19-1, Eco-20-1, Eco-21-1, Eco-22-1, Eco-23-1, Eco-24-1, Eco-24-2, Eco-25-1, Eco-25-2, Eco-26-1, Eco-26-2, Eco-27-1, Eco-28-1, Eco-28-2, Eco-29-1, Eco-3-1, Eco-31-1, Eco-33-1, Eco-33-2, Eco-33-3, Eco-34-1, Eco-34-2, Eco-34-3, Eco-4-1, Eco-5-1, Eco-6-1, Eco-6-2, Eco-6-3, Eco-7-1, Eco-8-3, Eco-9-1, Eco-9-2, Sbo-1-1, Sen-1-1, Sen-1-2, Sen-10-1, Sen-11-2, Sen-12-1, Sen-12-2, Sen-15-1, Sen-15-2, Sen-16-1, Sen-18-1, Sen-19-1, Sen-22-1, Sen-24-1, Sen-25-1, Sen-26-1, Sen-29-4, Sen-30-1, Sen-30-2, Sen-31-2, Sen-4-1, Sen-5-1, Sen-6-1, Sen-7-1, Sen-8-1, Sen-8-2, Sen-9-1, Sen-9-2, Sfl-1-1, Sfl-1-2, Sfl-1-3, Sfl-2-1, Sfl-2-2, Sfl-2-3, Ype-1-1, Ype-10-1, Ype-11-1, Ype-12-1, Ype-14-1, Ype-15-1, Ype-16-1, Ype-2-1, Ype-3-1, Ype-4-1, Ype-5-1, Ype-6-1, Ype-7-1, Ype-8-1, Ype-9-1, Yps-1-1, Yps-3-1, Yps-4-1, env-3

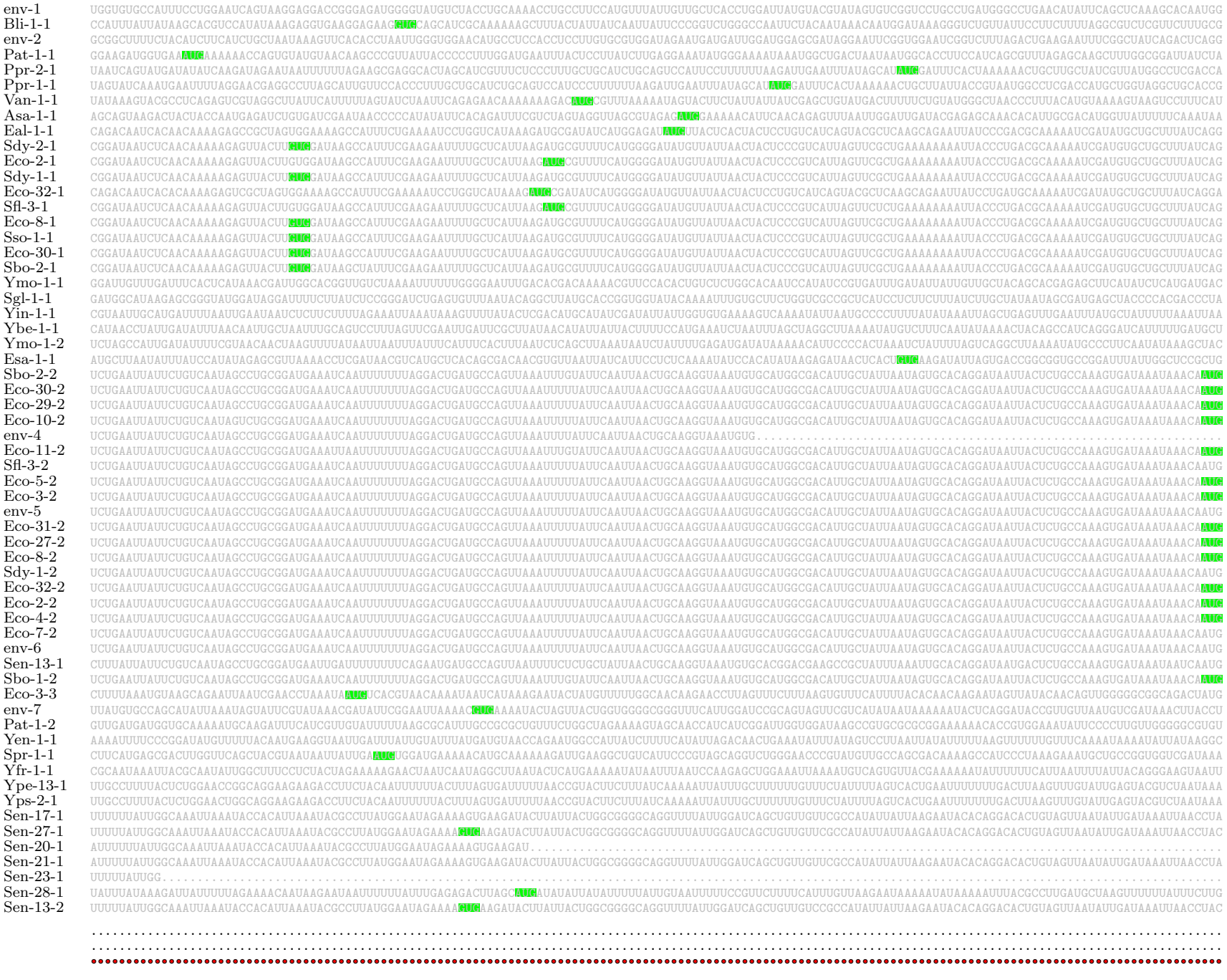
43.4 Multiple-sequence alignment

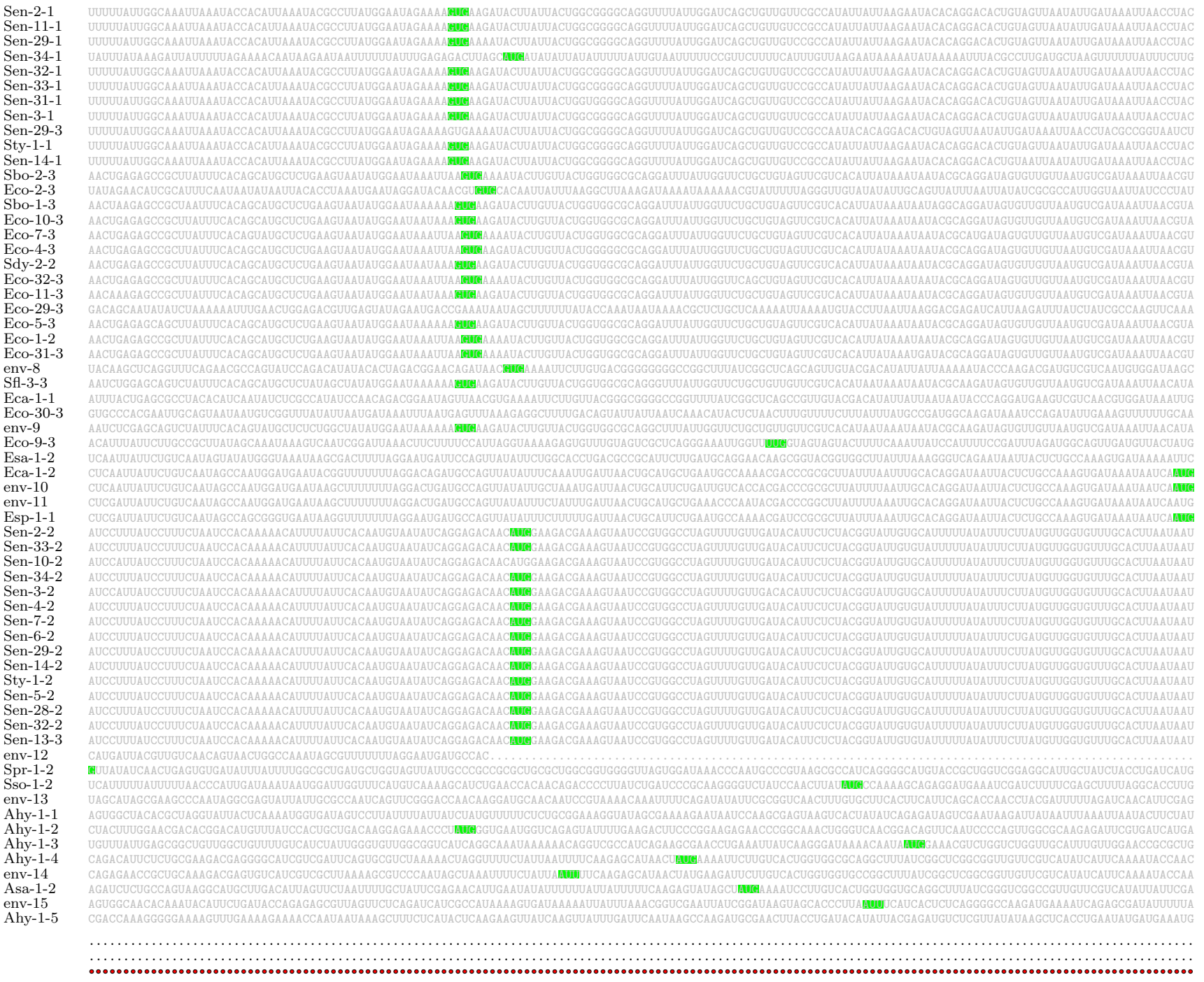
Each JUMPstart RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 43.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The JUMPstart RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.


Eco-30-2
Eco-29-2
Eco-10-2
env-4
Eco-11-2
Sfl-3-2
Eco-5-2
Eco-3-2
env-5
Eco-31-2
Eco-27-2
Eco-8-2
Sdy-1-2
Eco-32-2
Eco-2-2
Eco-4-2
Eco-7-2
env-6
Sen-13-1
Sbo-1-2
Eco-3-3
env-7
Pat-1-2
Yen-1-1
Spr-1-1
Yfr-1-1
Ype-13-1
Yps-2-1
Sen-17-1
Sen-27-1
Sen-20-1
Sen-21-1
Sen-23-1
Sen-28-1
Sen-13-2
Sen-2-1
Sen-11-1
Sen-29-1
Sen-34-1
Sen-32-1
Sen-33-1
Sen-31-1
Sen-3-1
Sen-29-3
Sty-1-1
Sen-14-1
Sbo-2-3
Eco-2-3
Sbo-1-3
Eco-10-3
Eco-7-3
Eco-4-3
Sdy-2-2
Eco-32-3
Eco-11-3
Eco-29-3
Eco-5-3
Eco-1-2
Eco-31-3
env-8
Sfl-3-3
<<<<<
>>>>>
?2222
2222?
5′
3′

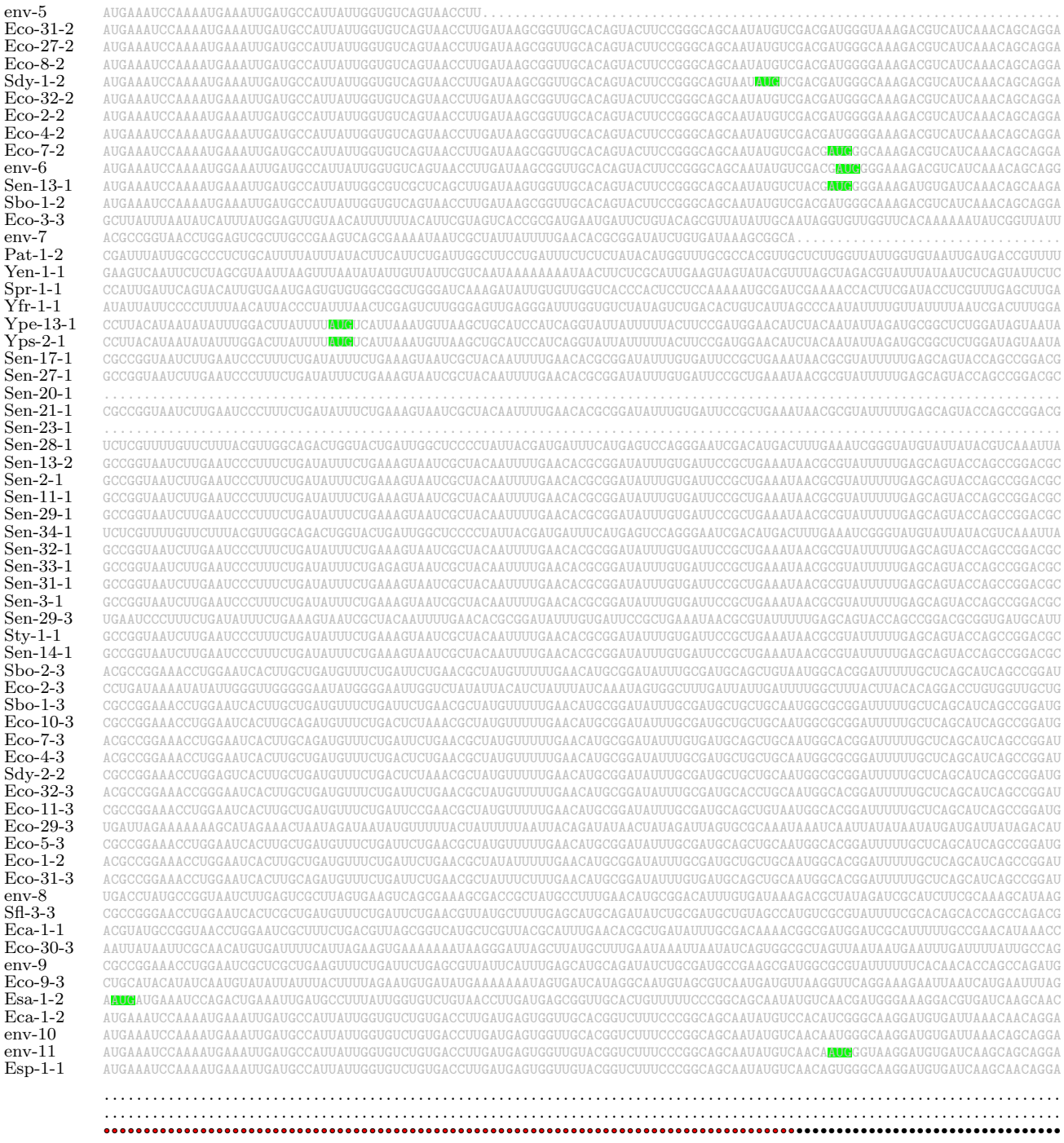

Eca-1-1
Eco-30-3
env-9
Eco-9-3
Esa-1-2
Eca-1-2
env-10
env-11
Esp-1-1
Sen-2-2
Sen-33-2
Sen-10-2
Sen-34-2
Sen-3-2
Sen-4-2
Sen-7-2
Sen-6-2
Sen-29-2
Sen-14-2
Sty-1-2
Sen-5-2
Sen-28-2
Sen-32-2
Sen-13-3
env-12
Spr-1-2
Sso-1-2
env-13
Ahy-1-1
Ahy-1-2
Ahy-1-3
Ahy-1-4
env-14
Asa-1-2
env-15
Ahy-1-5
Asa-1-3
env-16
Ahy-1-6
Asa-1-4
Asa-1-5
Van-1-2
Ssp-1-1
Son-1-1
Spu-1-1
Ssp-2-1
env-17
Sfr-1-1
Asa-1-6
Asa-1-7
Ahy-1-7
Asa-1-8
env-18
Sfr-1-2
Sam-1-1
Ahy-1-8
Ahy-1-9
env-19
5′
3′

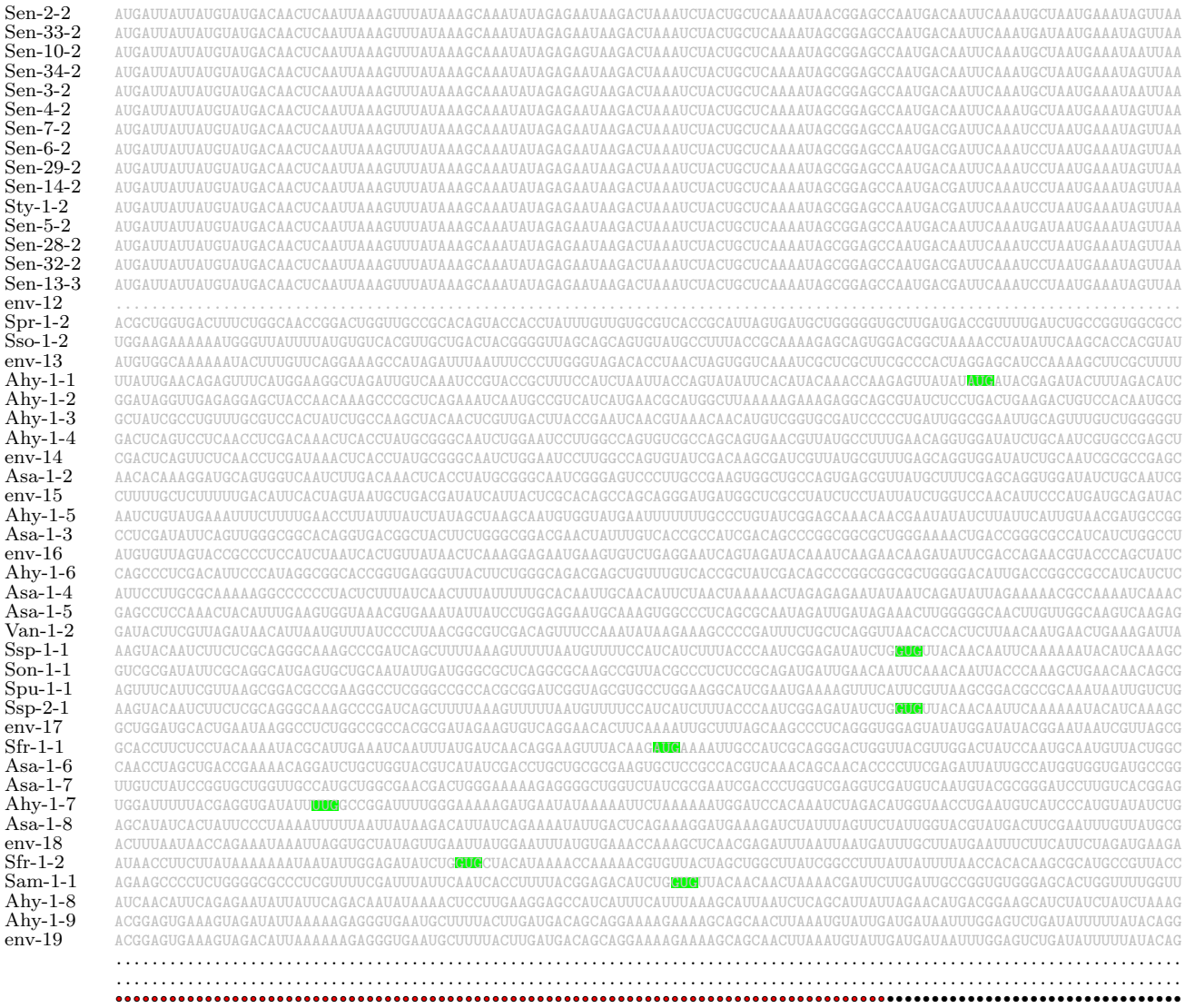
alignment positions 161 · · · 320




Sen-2-1
Sen-11-1
Sen-29-1
Sen-34-1
Sen-32-1
Sen-33-1
Sen-31-1
Sen-3-1
Sen-29-3
Sty-1-1
Sen-14-1
Sbo-2-3
Eco-2-3
Sbo-1-3
Eco-10-3
Eco-7-3
Eco-4-3
Sdy-2-2
Eco-32-3
Eco-11-3
Eco-29-3
Eco-5-3
Eco-1-2
Eco-31-3
env-8
Sfl-3-3
Eca-1-1
Eco-30-3
env-9
Eco-9-3
Esa-1-2
Eca-1-2
env-10
env-11
Esp-1-1
Sen-2-2
Sen-33-2
Sen-10-2
Sen-34-2
Sen-3-2
Sen-4-2
Sen-7-2
Sen-6-2
Sen-29-2
Sen-14-2
Sty-1-2
Sen-5-2
Sen-28-2
Sen-32-2
Sen-13-3
env-12
Spr-1-2
Sso-1-2
env-13
Ahy-1-1
Ahy-1-2
Ahy-1-3
Ahy-1-4
env-14
Asa-1-2
env-15
Ahy-1-5

alignment positions 321 · · · 439


env-5
Eco-31-2
Eco-27-2
Eco-8-2
Sdy-1-2
Eco-32-2
Eco-2-2
Eco-4-2
Eco-7-2
env-6
Sen-13-1
Sbo-1-2
Eco-3-3
env-7
Pat-1-2
Yen-1-1
Spr-1-1
Yfr-1-1
Ype-13-1
Yps-2-1
Sen-17-1
Sen-27-1
Sen-20-1
Sen-21-1
Sen-23-1
Sen-28-1
Sen-13-2
Sen-2-1
Sen-11-1
Sen-29-1
Sen-34-1
Sen-32-1
Sen-33-1
Sen-31-1
Sen-3-1
Sen-29-3
Sty-1-1
Sen-14-1
Sbo-2-3
Eco-2-3
Sbo-1-3
Eco-10-3
Eco-7-3
Eco-4-3
Sdy-2-2
Eco-32-3
Eco-11-3
Eco-29-3
Eco-5-3
Eco-1-2
Eco-31-3
env-8
Sfl-3-3
Eca-1-1
Eco-30-3
env-9
Eco-9-3
Esa-1-2
Eca-1-2
env-10
env-11
Esp-1-1



44 L17 downstream element

44.1 Taxa

The taxonomy of each organism containing a putative L17 downstream element RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual L17 downstream element RNA in Sections 44.2 and 44.4.

abbrev. of hits	taxonomy of species
Lin-1-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria innocua* Clip11262
Lmo-1-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* 10403S
Lmo-2-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* EGD-e
Lmo-3-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* F6900
Lmo-4-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* FSL F2-515
Lmo-5-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* FSL J1-175
Lmo-6-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* FSL J1-194
Lmo-7-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* FSL J1-208
Lmo-8-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* FSL J2-003
Lmo-9-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* FSL J2-064
Lmo-10-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* FSL J2-071
Lmo-11-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* FSL N1-017
Lmo-12-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* FSL N3-165
Lmo-13-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* FSL R2-503
Lmo-14-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* HPB2262
Lmo-15-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* J0161
Lmo-16-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* J2818
Lmo-17-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* LO28
Lmo-18-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes str.* 1/2a F6854
Lmo-19-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes str.* 4b F2365
Lmo-20-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes str.* 4b H7858
Lwe-1-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria welshimeri serovar* 6b str. SLCC5334
Efa-1-1	Bacteria Firmicutes Lactobacillales Enterococcaceae *Enterococcus faecalis* V583
Efa-2-1	Bacteria Firmicutes Lactobacillales Enterococcaceae *Enterococcus faecium* DO
Lac-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus acidophilus* NCFM
Lbr-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus brevis* ATCC 367
Lca-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus casei* ATCC 334
Lde-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus delbrueckii subsp. bulgaricus* ATCC 11842
Lde-2-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus delbrueckii subsp. bulgaricus* ATCC BAA-365
Lga-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus gasseri* ATCC 33323
Ljo-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus johnsonii* NCC 533
Lpl-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus plantarum* WCFS1
Lre-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus reuteri* 100-23
Lre-2-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus reuteri* DSM 20016
Lsa-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus sakei subsp. sakei* 23K
Lsa-2-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus salivarius* UCC118
Ppe-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Pediococcus pentosaceus* ATCC 25745
Lla-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. cremoris* MG1363
Lla-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. cremoris* SK11
Lla-3-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. lactis* Il1403
Sag-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 18RS21
Sag-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 2603V/R
Sag-3-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 515
Sag-4-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* A909
Sag-5-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* CJB111
Sag-6-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* COH1
Sag-7-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* H36B
Sag-8-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* NEM316
Smu-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus mutans* UA159
Spn-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* D39
Spn-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* R6
Spn-3-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP11-BS70
Spn-4-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP14-BS69
Spn-5-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP18-BS74
Spn-6-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP19-BS75
Spn-7-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP23-BS72
Spn-8-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP3-BS71
Spn-9-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP6-BS73

Spn-10-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP9-BS68
Spn-11-1 to Spn-11-2	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* TIGR4
Spy-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* M1 GAS
Spy-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* M49 591
Spy-3-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10270
Spy-4-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10394
Spy-5-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10750
Spy-6-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS2096
Spy-7-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS315
Spy-8-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS5005
Spy-9-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS6180
Spy-10-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS8232
Spy-11-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS9429
Spy-12-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* SSI-1
Spy-13-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes str.* Manfredo
Ssa-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus sanguinis* SK36
Ssu-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus suis* 05ZYH33
Ssu-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus suis* 89/1591
Ssu-3-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus suis* 98HAH33
Sth-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* CNRZ1066
Sth-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* LMD-9
Sth-3-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* LMG 18311
env-1	environmental sample

44.2 Gene contexts

Each L17 downstream element RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one L17 downstream element RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 44.3. The accession number of the sequence containing each L17 downstream element RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 44.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Lsa-1-1	NC_007576.1	-	1723553	1723485	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ ABC_cobalt_transport_domain1 (cd03225)→
Lbr-1-1	NC_008497.1	-	1632895	1632838	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ ABC_cobalt_transport_domain1 (cd03225)→
Lre-2-1	NC_009513.1	-	1515876	1515820	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ ABC_cobalt_transport_domain1 (cd03225)→
Lre-1-1	NZ_AAPZ01000074.1	-	3407	3351	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ ABC_cobalt_transport_domain1 (cd03225)→
Lpl-1-1	NC_004567.1	+	966908	966986	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ ←Transposase_12 (pfam01610)
Lmo-20-1	NZ_AADR01000037.1	-	21840	21767	**RNA**→
Lmo-7-1	NZ_AARL01001264.1	-	347	272	**RNA**→
Lmo-17-1	NZ_AARY01000487.1	-	7439	7364	**RNA**→
Lmo-4-1	NZ_AARI01001268.1	-	308	233	**RNA**→
Lmo-8-1	NZ_AARM01000389.1	-	2131	2056	**RNA**→
Lmo-15-1	NZ_AARW01000011.1	+	79122	79197	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ hypo (cons)→
Lmo-2-1	NC_003210.1	-	2691650	2691575	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ hypo→
Lmo-1-1	NZ_AARZ01000009.1	-	21798	21723	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ hypo (cons)→
Lmo-18-1	NZ_AADQ01000051.1	-	6006	5931	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ hypo (cons)→
Lmo-3-1	NZ_AARU01000033.1	-	21056	20981	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ hypo (cons)→
Lmo-16-1	NZ_AARX01000037.1	-	21052	20977	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ hypo (cons)→
Lmo-12-1	NZ_AARQ01000033.1	-	22377	22302	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ hypo (cons)→

Lwe-1-1	NC_008555.1	-	2597533	2597458	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ hypo→
Lin-1-1	NC_003212.1	-	2782594	2782519	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ hypo→
Lmo-10-1	NZ_AARN02000043.1	-	13390	13315	**RNA**→
Lmo-9-1	NZ_AARO01000300.1	-	7470	7395	**RNA**→
Lmo-19-1	NC_002973.6	-	2640188	2640113	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ hypo→
Lmo-6-1	NZ_AARJ01000040.1	-	20406	20331	**RNA**→
Lmo-5-1	NZ_AARK01000323.1	-	14604	14529	**RNA**→
Lmo-11-1	NZ_AARP02000068.1	-	22443	22368	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ hypo (cons)→
Lmo-14-1	NZ_AATL01000060.1	-	17187	17112	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ hypo (cons)→
Lmo-13-1	NZ_AARR01000044.1	-	22468	22393	**RNA**→
Lla-3-1	NC_002662.1	-	2151650	2151528	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ tRNA-Thr→
Lla-2-1	NC_008527.1	-	2215255	2215137	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ tRNA-Thr→
Lla-1-1	NC_009004.1	-	2310910	2310792	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ tRNA-Thr→
Lde-1-1	NC_008054.1	+	360216	360270	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ ABC_cobalt_transport_domain1 (cd03225)→
Lde-2-1	NC_008529.1	+	354775	354829	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ ABC_cobalt_transport_domain1 (cd03225)→
Lga-1-1	NC_008530.1	+	346685	346736	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ ABC_cobalt_transport_domain1 (cd03225)→
Ljo-1-1	NC_005362.1	+	395992	396043	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ ABC_cobalt_transport_domain1 (cd03225)→
Lac-1-1	NC_006814.2	+	304248	304302	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ ABC_cobalt_transport_domain1 (cd03225)→
Lca-1-1	NC_008526.1	-	2457871	2457801	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ hypo→
Efa-2-1	NZ_AAAK03000044.1	+	11970	12043	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ CbiO (COG1122)→
Lsa-2-1	NC_007929.1	-	1482423	1482351	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ ABC_cobalt_transport_domain1 (cd03225)→
Efa-1-1	NC_004668.1	+	212775	212844	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ hypo→
Ppe-1-1	NC_008525.1	-	1379186	1379131	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ ABC_cobalt_transport_domain1 (cd03225)→
Smu-1-1	NC_004350.1	-	1880017	1879945	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ GshA (COG3572)→
Ssu-2-1	NZ_AAFA02000017.1	+	1760	1834	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ COG3464 (COG3464)→
Spn-5-1	NZ_ABAE01000002.1	+	97263	97335	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ COG3830 (COG3830)→
Spn-3-1	NZ_ABAC01000001.1	+	195756	195827	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ COG3830 (COG3830)→
Spn-11-1	NC_003028.2	+	207852	207923	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ COG3830 (COG3830)→
Spn-11-2	NZ_AAGY02000002.1	-	15136	15065	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ COG3830 (COG3830)→
Spn-8-1	NZ_AAZZ01000001.1	+	97257	97328	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ COG3830 (COG3830)→
Spn-7-1	NZ_ABAG01000001.1	+	136722	136793	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ COG3830 (COG3830)→
Spn-1-1	NC_008533.1	+	210361	210432	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ COG3830 (COG3830)→
Spn-2-1	NC_003098.1	+	210361	210432	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ COG3830 (COG3830)→
Spn-9-1	NZ_ABAA01000001.1	+	181348	181419	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ COG3830 (COG3830)→
Spn-10-1	NZ_ABAB01000005.1	+	84013	84084	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ COG3830 (COG3830)→
Spn-4-1	NZ_ABAD01000001.1	-	109540	109469	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ COG3830 (COG3830)→
Spn-6-1	NZ_ABAF01000002.1	+	180446	180517	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ COG3830 (COG3830)→
Ssa-1-1	NC_009009.1	+	122640	122714	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ 16S small subunit ribosomal RNA→
env-1	BAAZ01018606.1	+	70	141	**RNA**→
Spy-4-1	NC_006086.1	+	119898	119975	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ ←hypo
Spy-13-1	NC_009332.1	+	79773	79850	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ Tra8 (COG2826)→
Spy-9-1	NC_007296.1	+	78525	78602	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ ←hypo
Spy-11-1	NC_008021.1	+	78847	78924	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ ←hypo
Spy-6-1	NC_008023.1	+	78823	78900	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ ←hypo
Spy-10-1	NC_003485.1	+	79130	79207	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ 16S ribosomal RNA→
Spy-12-1	NC_004606.1	+	73239	73316	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ 16S ribosomal RNA→
Spy-7-1	NC_004070.1	+	79240	79317	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ 16S ribosomal RNA→
Spy-8-1	NC_007297.1	+	79037	79114	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ ←hypo
Spy-3-1	NC_008022.1	+	79217	79294	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ ←hypo
Spy-1-1	NC_002737.1	+	78480	78557	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ 16S ribosomal RNA→
Spy-5-1	NC_008024.1	+	80320	80397	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ ←Tra8 (COG2826)
Spy-2-1	NZ_AAFV01000275.1	+	858	935	**RNA**→

Sag-2-1	NC_004116.1	+	90643	90714	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ 16S ribosomal RNA→
Sag-1-1	NZ_AAJO01000021.1	-	2711	2640	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ 16S ribosomal RNA→
Sag-5-1	NZ_AAJQ01000070.1	+	1744	1815	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→
Sag-6-1	NZ_AAJR01000123.1	+	1060	1131	**RNA**→
Sag-8-1	NC_004368.1	+	90859	90930	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ 16S ribosomal RNA→
Sag-3-1	NZ_AAJP01000113.1	+	1707	1778	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→
Sag-4-1	NC_007432.1	+	90353	90424	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ 16S ribosomal RNA→
Sag-7-1	NZ_AAJS01000028.1	+	12182	12253	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ protein of unknown function/lipoprotein, putative→
Ssu-1-1	NC_009442.1	+	87133	87206	RplQ (COG0203)→ RplQ (COG0203)→ **RNA**→ 16S ribosomal RNA→
Ssu-3-1	NC_009443.1	+	87129	87202	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ 16S ribosomal RNA→
Sth-2-1	NC_008532.1	-	1755085	1755008	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ tRNA-Leu→
Sth-3-1	NC_006448.1	-	1696818	1696743	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ ←ISSth4, transposase, IS30 family, truncated
Sth-1-1	NC_006449.1	-	1697442	1697367	RpoA (COG0202)→ RplQ (COG0203)→ **RNA**→ ←ISSth4, transposase, IS30 family, truncated

44.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 44.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one L17 downstream element RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

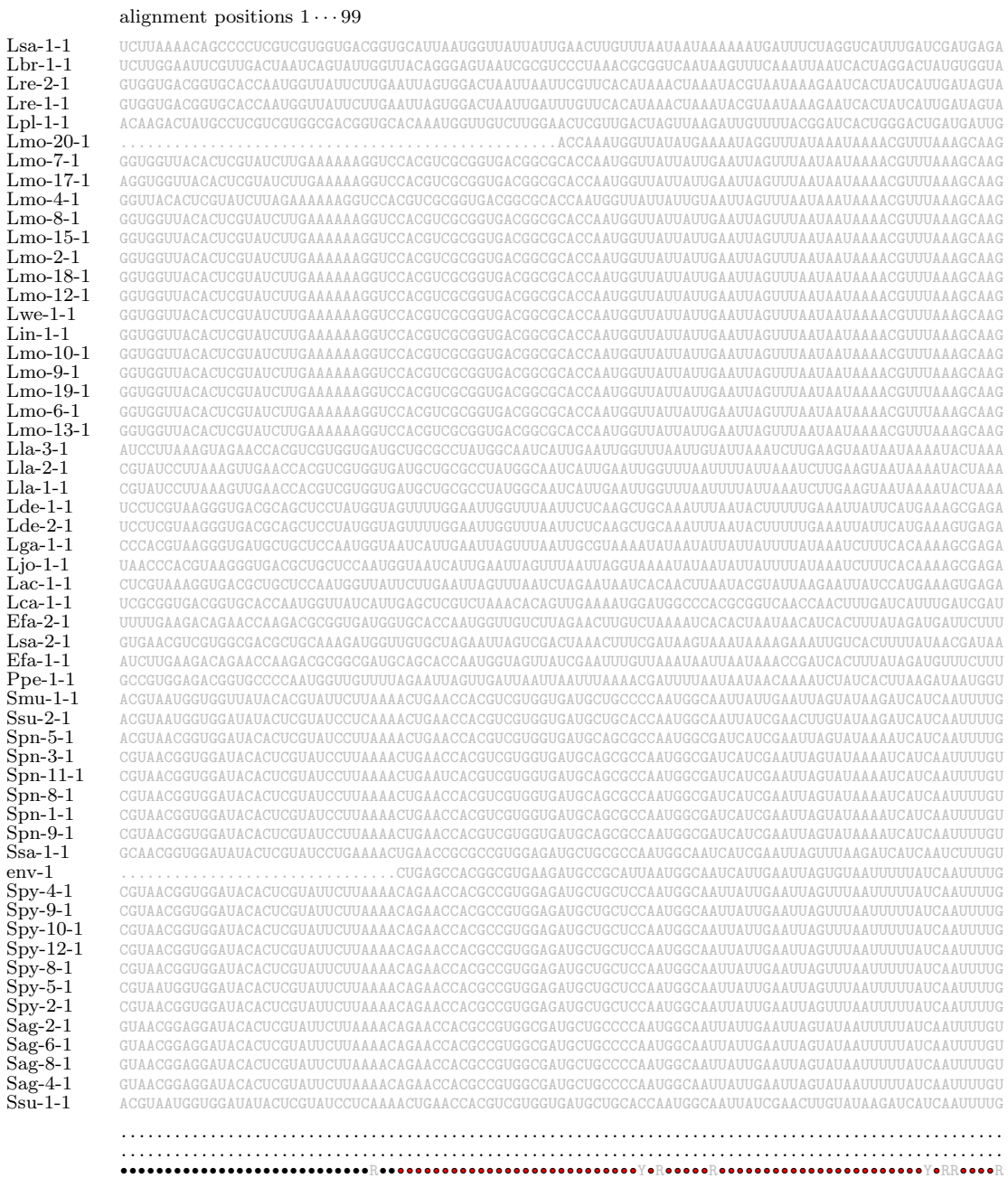
cd03225 (11) Domain I of the ATPase component of a cobalt transport family found in both bacteria and archaea.
COG0202 (68) DNA-directed RNA polymerase, alpha subunit/40 kD subunit [Transcription]
COG0203 (70) Ribosomal protein L17 [Translation, ribosomal structure and biogenesis]
COG1122 (1) ABC-type cobalt transport system, ATPase component [Inorganic ion transport and metabolism]
COG2826 (2) Transposase and inactivated derivatives, IS30 family [DNA replication, recombination, and repair]
COG3464 (1) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
COG3572 (1) Gamma-glutamylcysteine synthetase [Coenzyme metabolism]
COG3830 (12) ACT domain-containing protein [Signal transduction mechanisms]
pfam01610 (1) Transposase.

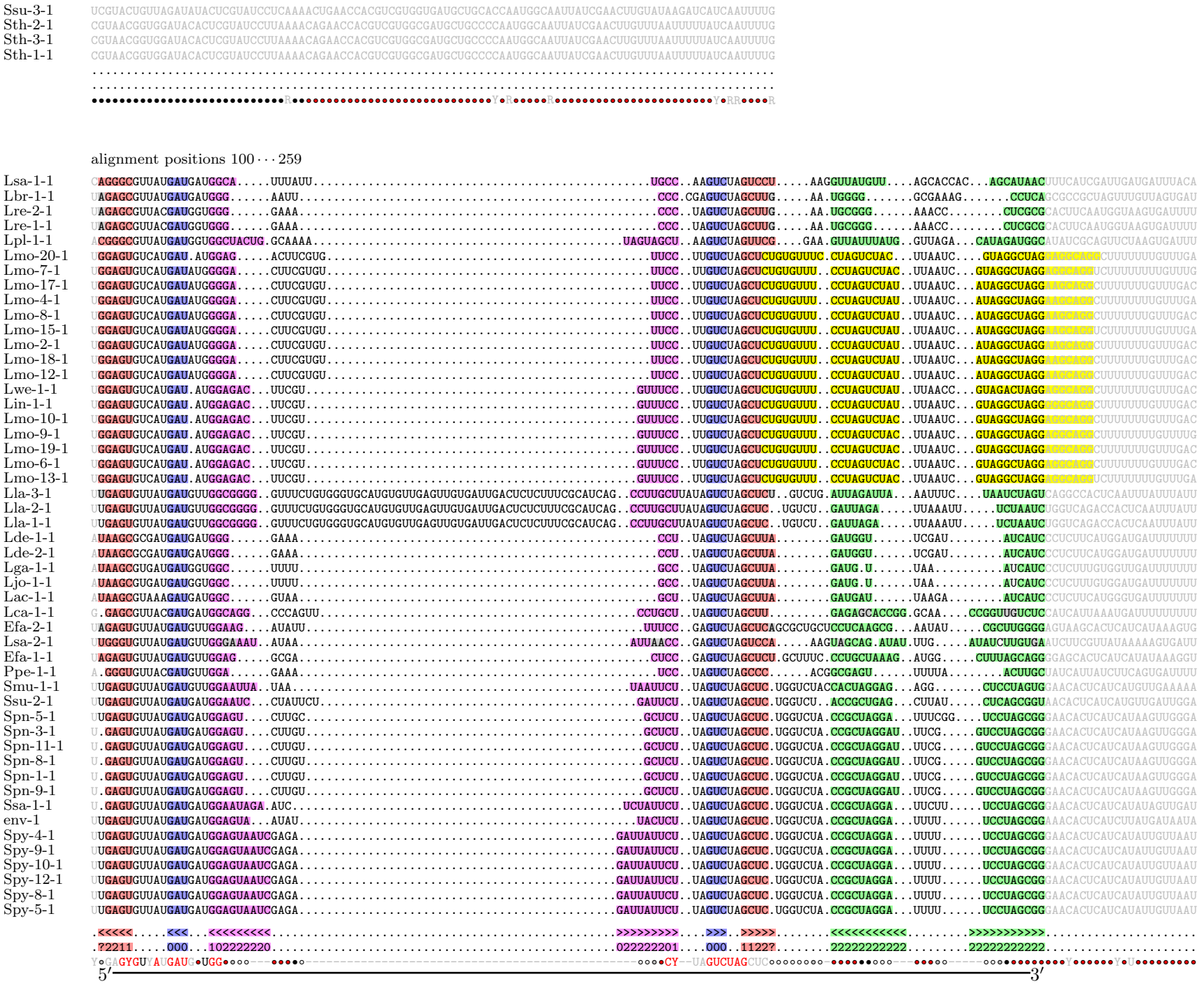
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Lmo-1-1, Lmo-11-1, Lmo-14-1, Lmo-16-1, Lmo-3-1, Lmo-5-1, Sag-1-1, Sag-3-1, Sag-5-1, Sag-7-1, Spn-10-1, Spn-11-2, Spn-2-1, Spn-4-1, Spn-6-1, Spn-7-1, Spy-1-1, Spy-11-1, Spy-13-1, Spy-3-1, Spy-6-1, Spy-7-1

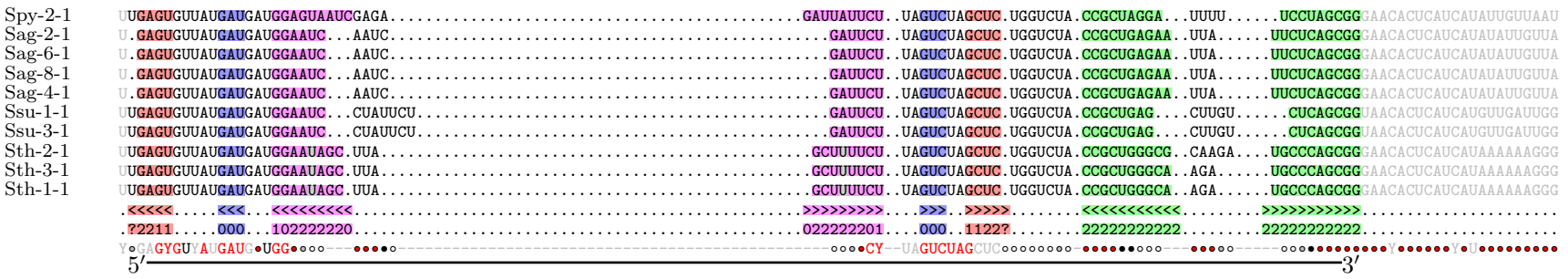
44.4 Multiple-sequence alignment

Each L17 downstream element RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 44.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The L17 downstream element RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red

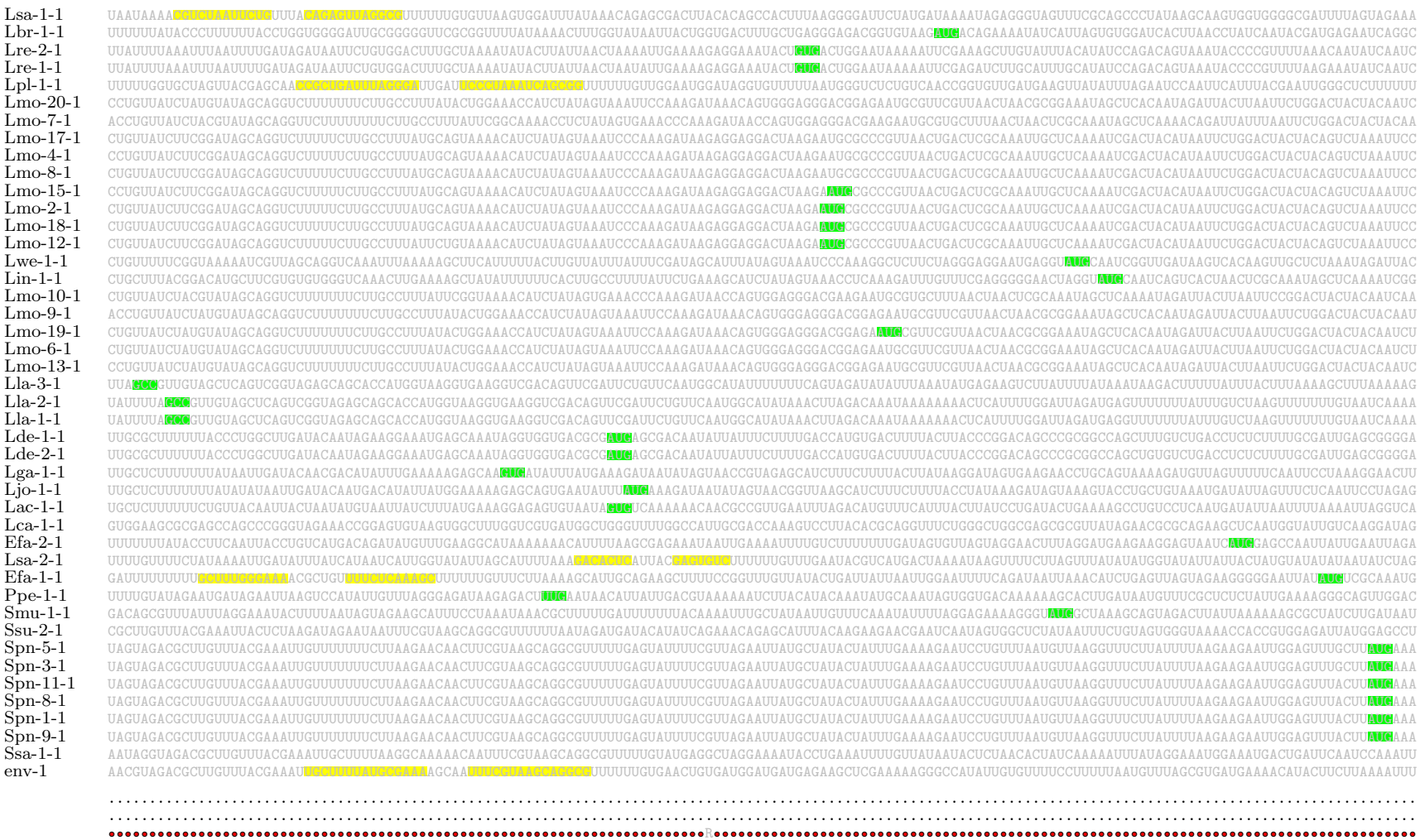
circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

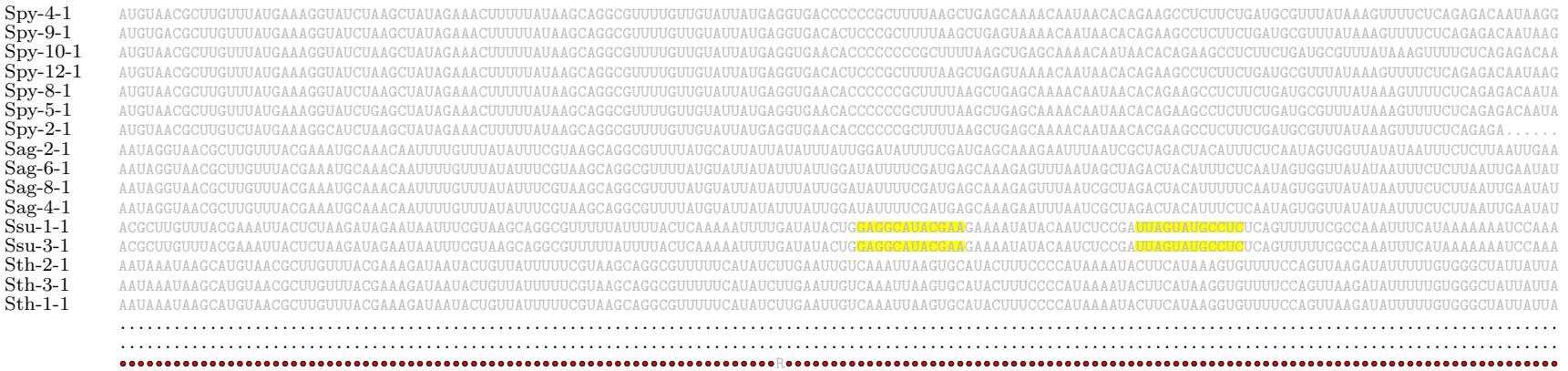



Ssu-3-1
Sth-2-1
Sth-3-1
Sth-1-1
alignment positions 100 · · · 259
Lsa-1-1
Lbr-1-1
Lre-2-1
Lre-1-1
Lpl-1-1
Lmo-20-1
Lmo-7-1
Lmo-17-1
Lmo-4-1
Lmo-8-1
Lmo-15-1
Lmo-2-1
Lmo-18-1
Lmo-12-1
Lwe-1-1
Lin-1-1
Lmo-10-1
Lmo-9-1
Lmo-19-1
Lmo-6-1
Lmo-13-1
Lla-3-1
Lla-2-1
Lla-1-1
Lde-1-1
Lde-2-1
Lga-1-1
Ljo-1-1
Lac-1-1
Lca-1-1
Efa-2-1
Lsa-2-1
Efa-1-1
Ppe-1-1
Smu-1-1
Ssu-2-1
Spn-5-1
Spn-3-1
Spn-11-1
Spn-8-1
Spn-1-1
Spn-9-1
Ssa-1-1
env-1
Spy-4-1
Spy-9-1
Spy-10-1
Spy-12-1
Spy-8-1
Spy-5-1
5′
3′

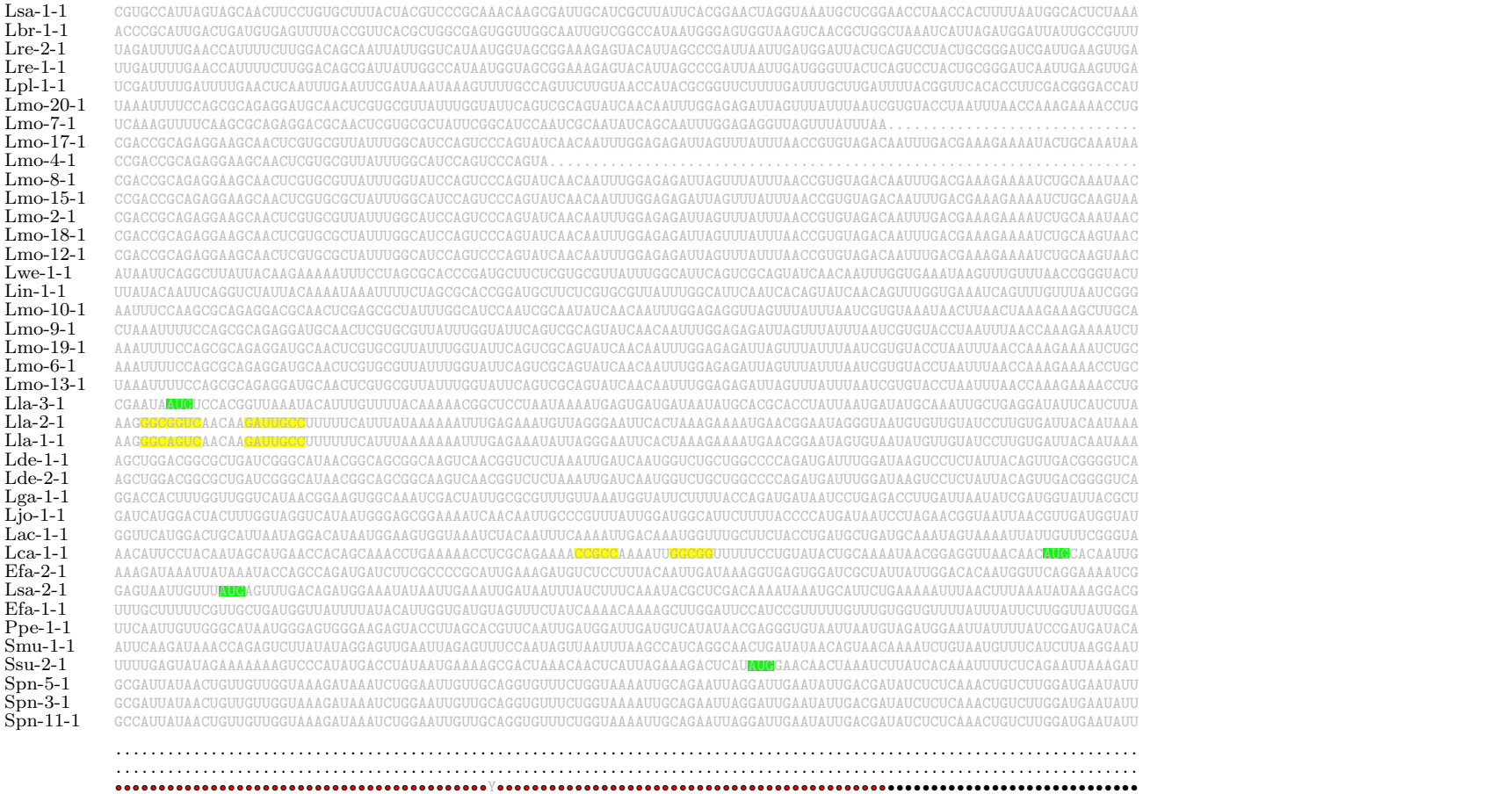


alignment positions 260 · · · 419





alignment positions 420 · · · 537



45 *lactis*-plasmid

45.1 Taxa

The taxonomy of each organism containing a putative *lactis*-plasmid RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *lactis*-plasmid RNA in Sections 45.2 and 45.4.

abbrev. of hits	**taxonomy of species**
Lfe-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus fermentum*
Lme-1-1	Bacteria Firmicutes Lactobacillales *Leuconostoc mesenteroides*
Lla-1-1 to Lla-1-6	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis*
Lla-2-1 to Lla-2-3	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. cremoris*
Lla-3-1 to Lla-3-2	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. cremoris* SK11
Lla-4-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. lactis*

45.2 Gene contexts

Each *lactis*-plasmid RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *lactis*-plasmid RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 45.3. The accession number of the sequence containing each *lactis*-plasmid RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 45.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Lme-1-1	NC_005702.1	-	2304	2198	putative replication initiation protein→ **RNA**→ ←hypo
Lla-3-1	NC_008505.1	-	26010	25905	hypo→ **RNA**→ MobC, mobilisation protein (MobC)→
Lla-2-1	NC_002137.1	-	4166	4061	←Methylase_S (pfam01420) **RNA**→ **RNA**→ hypo→
Lla-2-3	NC_002137.1	-	39289	39184	(shown above)
Lla-1-1	NC_000906.2	+	6041	6146	hypo→ **RNA**→ ←COG3316 (COG3316)
Lla-2-2	NC_004847.1	+	3083	3188	RepC→ **RNA**→ ←COG3316 (COG3316)
Lla-3-2	NC_008504.1	+	3486	3591	Transposase_8 (pfam01527)→ **RNA**→ LytTR (pfam04397)→
Lla-4-1	NC_004966.1	+	6908	7013	Orf8→ **RNA**→ Orf9→
Lla-1-2	NC_009435.1	-	22275	22170	rve (pfam00665)→ **RNA**→ mobilization protein→
Lla-1-3	NC_003101.1	-	6099	5994	←COG5527 (COG5527) **RNA**→ ←Na_citrate (pfam03390)
Lla-1-4	NC_008436.1	+	3235	3340	RepC→ **RNA**→
Lfe-1-1	NC_004947.1	-	1907	1800	hypo→ **RNA**→ ←Rep_3 (pfam01051)
Lla-1-5	NC_004981.2	+	245	352	**RNA**→ transcriptional repressor protein→
Lla-1-6	NC_002799.1	+	2486	2593	Rep_3 (pfam01051)→ **RNA**→ ←Glyoxalase (pfam00903)

45.3 Conserved domains

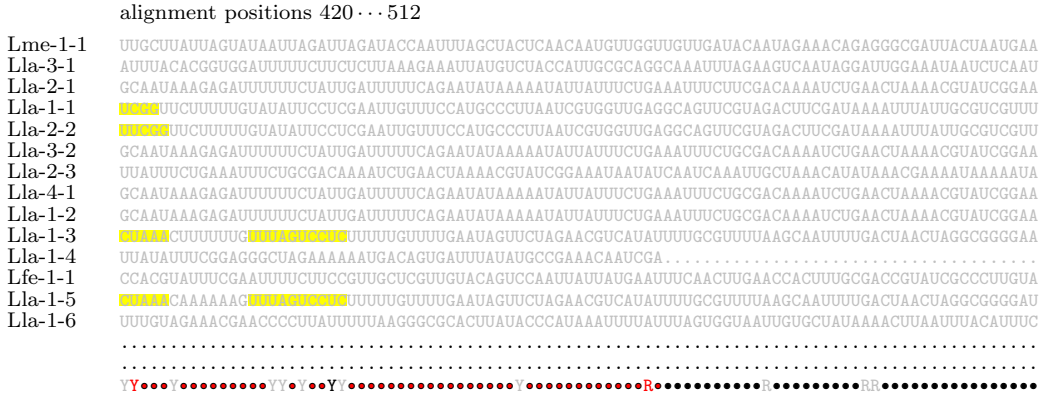
Conserved domains found in protein-coding genes listed in Section 45.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *lactis*-plasmid RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

`d` COG3316 (2) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
`u` COG5527 (1) Protein involved in initiation of plasmid replication [DNA replication, recombination, and repair]
`u` pfam00665 (1) Integrase core domain.
`d` pfam00903 (1) Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily.
`u d` pfam01051 (2) Initiator Replication protein.
`u` pfam01420 (1) Type I restriction modification DNA specificity domain.
`u` pfam01527 (1) Transposase.
`d` pfam03390 (1) Bacterial sodium:citrate symporter.
`D` pfam04397 (1) LytTr DNA-binding domain.

45.4 Multiple-sequence alignment

Each *lactis*-plasmid RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 45.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *lactis*-plasmid RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

alignment positions 1 · · · 99
Lme-1-1
Lla-3-1
Lla-2-1
Lla-1-1
Lla-2-2
Lla-3-2
Lla-2-3
Lla-4-1
Lla-1-2
Lla-1-3
Lla-1-4
Lfe-1-1
Lla-1-5
Lla-1-6
alignment positions 100 · · · 259
Lme-1-1
Lla-3-1
Lla-2-1
Lla-1-1
Lla-2-2
Lla-3-2
Lla-2-3
Lla-4-1
Lla-1-2
Lla-1-3
Lla-1-4
Lfe-1-1
Lla-1-5
Lla-1-6
5′
3′
alignment positions 260 · · · 419
Lme-1-1
Lla-3-1
Lla-2-1
Lla-1-1
Lla-2-2
Lla-3-2
Lla-2-3
Lla-4-1
Lla-1-2
Lla-1-3
Lla-1-4
Lfe-1-1
Lla-1-5
Lla-1-6



46 Lacto-*int*

46.1 Taxa

The taxonomy of each organism containing a putative Lacto-*int* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Lacto-*int* RNA in Sections 46.2 and 46.4.

abbrev. of hits	taxonomy of species
Lla-1-1 to Lla-1-2	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. cremoris* MG1363
Lla-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. cremoris* SK11
Lla-3-1 to Lla-3-4	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. lactis* Il1403
Sag-1-1 to Sag-1-3	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 2603V/R
Sag-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 515
Sag-3-1 to Sag-3-2	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* A909
Sag-4-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* CJB111
Sag-5-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* COH1
Sag-6-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* H36B
Sag-7-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* NEM316
Spn-1-1 to Spn-1-2	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP14-BS69
Spn-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP18-BS74
Spn-3-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP19-BS75
Spn-4-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP23-BS72
Spn-5-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP6-BS73
Spn-6-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP9-BS68
Spy-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* M1 GAS
Spy-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10270
Spy-3-1 to Spy-3-2	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10394
Spy-4-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10750
Spy-5-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS2096
Spy-6-1 to Spy-6-3	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS5005
Spy-7-1 to Spy-7-2	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS6180
Spy-8-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS8232
Spy-9-1 to Spy-9-2	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS9429
Spy-10-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* SSI-1
Spy-11-1 to Spy-11-2	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes str.* Manfredo
Ssu-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus suis* 89/1591
Sth-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* LMD-9
Sph-2-1	Viruses dsDNA viruses, no RNA stage Caudovirales Myoviridae *Streptococcus phage* EJ-1
Lph-1-1	Viruses dsDNA viruses, no RNA stage Caudovirales Siphoviridae *Lactococcus phage* BK5-T

Lph-4-1	Viruses dsDNA viruses, no RNA stage Caudovirales Siphoviridae *Lactococcus phage phi*LC3
Lph-2-1	Viruses dsDNA viruses, no RNA stage Caudovirales Siphoviridae *Lactococcus phage* TP901-1
Lph-3-1	Viruses dsDNA viruses, no RNA stage Caudovirales Siphoviridae *Lactococcus phage* Tuc2009
Lph-5-1	Viruses dsDNA viruses, no RNA stage Caudovirales Siphoviridae *Lactococcus phage ul*36
Sph-3-1	Viruses dsDNA viruses, no RNA stage Caudovirales Siphoviridae *Streptococcus phage* MM1
Sph-4-1	Viruses dsDNA viruses, no RNA stage Caudovirales Siphoviridae *Streptococcus phage* SM1
Sph-5-1	Viruses dsDNA viruses, no RNA stage Caudovirales *Streptococcus phage* SMP
Sph-1-1	Viruses *Streptococcus phage* P9
env-1	environmental sample

46.2 Gene contexts

Each Lacto-*int* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Lacto-*int* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 46.3. The accession number of the sequence containing each Lacto-*int* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 46.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Lph-2-1	NC_002747.1	-	1592	1482	**RNA**→
Lla-3-1	NC_002662.1	-	448475	448367	**RNA**→ INT_phiLC3_C (cd01189)→
Lla-1-1	NC_009004.1	-	584662	584552	**RNA**→ INT_phiLC3_C (cd01189)→
Lla-3-2	NC_002662.1	-	503976	503866	**RNA**→ INT_Tn1545_C (cd01199)→
Lla-3-3	NC_002662.1	+	1455765	1455872	**RNA**→ INT_phiLC3_C (cd01189)→
Lph-4-1	NC_005822.1	-	1323	1215	**RNA**→
Lph-3-1	NC_002703.1	-	1323	1215	**RNA**→
Lph-5-1	NC_004066.1	-	1686	1578	**RNA**→
Sth-1-1	NC_008532.1	-	747056	746949	**RNA**→ Transposase_8 (pfam01527)→
Lla-1-2	NC_009004.1	-	780269	780160	**RNA**→ INT_Tn1545_C (cd01199)→
Lph-1-1	NC_002796.1	-	27143	27032	**RNA**→
Lla-2-1	NC_008527.1	-	1037989	1037878	**RNA**→ INT_phiLC3_C (cd01189)→
Lla-3-4	NC_002662.1	-	1037872	1037763	**RNA**→ INT_phiLC3_C (cd01189)→
Sag-7-1	NC_004368.1	+	1260051	1260156	**RNA**→ INT_phiLC3_C (cd01189)→
Sag-4-1	NZ_AAJQ01000072.1	-	2348	2241	**RNA**→ INT_phiLC3_C (cd01189)→
Sag-3-1	NC_007432.1	-	629913	629806	**RNA**→ ←ISSag6, transposase orfA
Sag-1-1	NC_004116.1	-	601723	601616	**RNA**→ ←hypo
Sag-5-1	NZ_AAJR01000019.1	-	22009	21902	**RNA**→ ←rve (pfam00665)
Sag-1-2	NC_004116.1	+	1865944	1866052	**RNA**→ INT_phiLC3_C (cd01189)→ hypo→
Spy-3-1	NC_006086.1	-	36847	36734	**RNA**→ INT_phiLC3_C (cd01189)→
Spy-7-1	NC_007296.1	+	1267871	1267985	**RNA**→ INT_phiLC3_C (cd01189)→
Spn-3-1	NZ_ABAF01000010.1	-	26501	26397	**RNA**→ INT_Tn1545_C (cd01199)→
Ssu-1-1	NZ_AAFA02000055.1	+	7300	7401	**RNA**→ INT_phiLC3_C (cd01189)→
Sag-6-1	NZ_AAJS01000006.1	-	15270	15173	**RNA**→ INT_Tn1545_C (cd01199)→
Sph-5-1	NC_008721.1	-	11034	10922	**RNA**→
Spy-1-1	NC_002737.1	-	779887	779776	**RNA**→ INT_Tn1545_C (cd01199)→
Spy-9-1	NC_008021.1	-	780623	780514	**RNA**→ INT_Tn1545_C (cd01199)→
Spy-4-1	NC_008024.1	+	1254902	1255012	**RNA**→ INT_Tn1545_C (cd01199)→

Spy-11-1	NC_009332.1	+	1068604	1068715	**RNA**→ INT_Tn1545_C (cd01199)→
Spy-5-1	NC_008023.1	-	533026	532917	**RNA**→ PinR (COG1961)→
Spy-9-2	NC_008021.1	-	531117	531006	**RNA**→ PinR (COG1961)→ DNA integration/recombination/inversion protein→
Spy-11-2	NC_009332.1	+	1308454	1308564	**RNA**→ PinR (COG1961)→
Sag-3-2	NC_007432.1	-	550209	550101	**RNA**→ INT_Tn1545_C (cd01199)→
Spy-2-1	NC_008022.1	-	774133	774024	**RNA**→ INT_Tn1545_C (cd01199)→
Sag-2-1	NZ_AAJP01000049.1	-	1802	1690	**RNA**→ hypo→ INT_Tn1545_C (cd01199)→
Sag-1-3	NC_004116.1	-	560012	559900	**RNA**→ INT_Tn1545_C (cd01199)→
Spn-1-1	NZ_ABAD01000012.1	+	1417	1534	**RNA**→ PinR (COG1961)→
Spn-1-2	NZ_ABAD01000019.1	+	1925	2029	**RNA**→ PinR (COG1961)→
Spy-6-1	NC_007297.1	+	1424779	1424885	**RNA**→ INT_phiLC3_C (cd01189)→
Sph-1-1	NC_009819.1	-	1316	1218	**RNA**→
Sph-2-1	NC_005294.1	-	1349	1243	**RNA**→
Spy-6-2	NC_007297.1	+	1177541	1177649	**RNA**→ INT_Tn1545_C (cd01199)→
env-1	BAAZ01011818.1	-	446	339	**RNA**→ hypo→
Sph-4-1	NC_004996.1	-	1323	1225	**RNA**→
Sph-3-1	NC_003050.2	-	1332	1225	**RNA**→
Spn-4-1	NZ_ABAG01000012.1	+	32680	32787	**RNA**→ INT_Tn1545_C (cd01199)→
Spn-2-1	NZ_ABAE01000018.1	-	1343	1237	**RNA**→ INT_Tn1545_C (cd01199)→
Spn-5-1	NZ_ABAA01000017.1	-	5989	5892	**RNA**→ INT_Tn1545_C (cd01199)→
Spn-6-1	NZ_ABAB01000010.1	-	35088	34991	**RNA**→ INT_Tn1545_C (cd01199)→
Spy-3-2	NC_006086.1	+	1019150	1019257	**RNA**→ INT_phiLC3_C (cd01189)→
Spy-6-3	NC_007297.1	+	1020477	1020584	**RNA**→ INT_phiLC3_C (cd01189)→
Spy-7-2	NC_007296.1	+	1031151	1031258	**RNA**→ INT_phiLC3_C (cd01189)→
Spy-8-1	NC_003485.1	+	1086549	1086656	**RNA**→ INT_phiLC3_C (cd01189)→
Spy-10-1	NC_004606.1	-	879880	879773	**RNA**→ INT_phiLC3_C (cd01189)→

46.3 Conserved domains

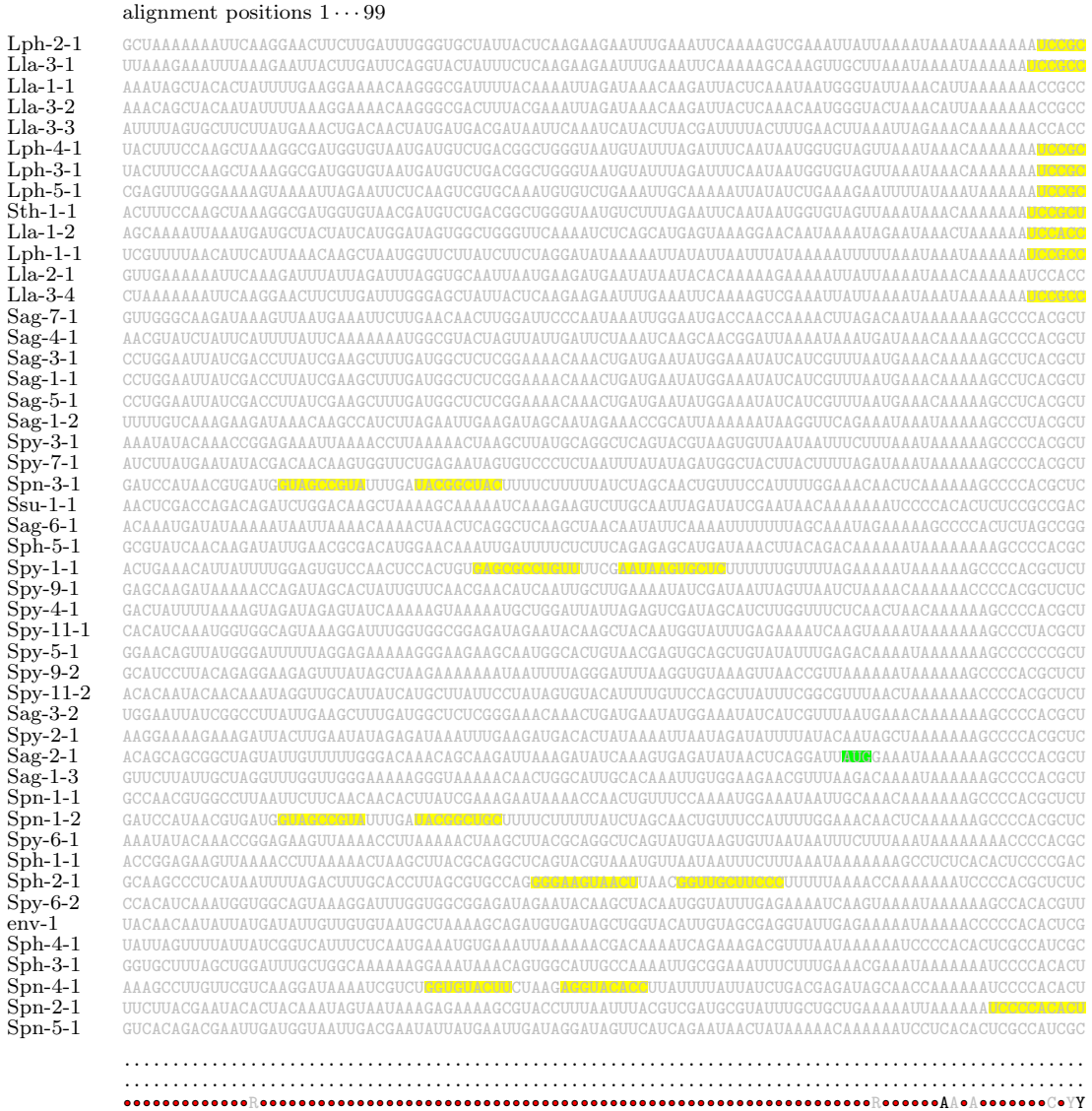
Conserved domains found in protein-coding genes listed in Section 46.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Lacto-*int* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

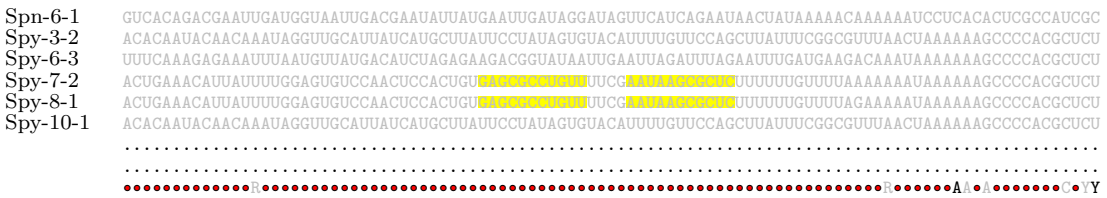
cd01189 (17) phiLC3 phage and phage-related integrases, site-specific recombinases, DNA breaking-rejoining enzymes, C-terminal catalytic domain.
cd01199 (17) Tn1545-related conjugative transposon integrases, site-specific recombinases, DNA breaking-rejoining enzymes, C-terminal catalytic domain.
COG1961 (5) Site-specific recombinases, DNA invertase Pin homologs [DNA replication, recombination, and repair]
pfam00665 (1) Integrase core domain.
pfam01527 (1) Transposase.

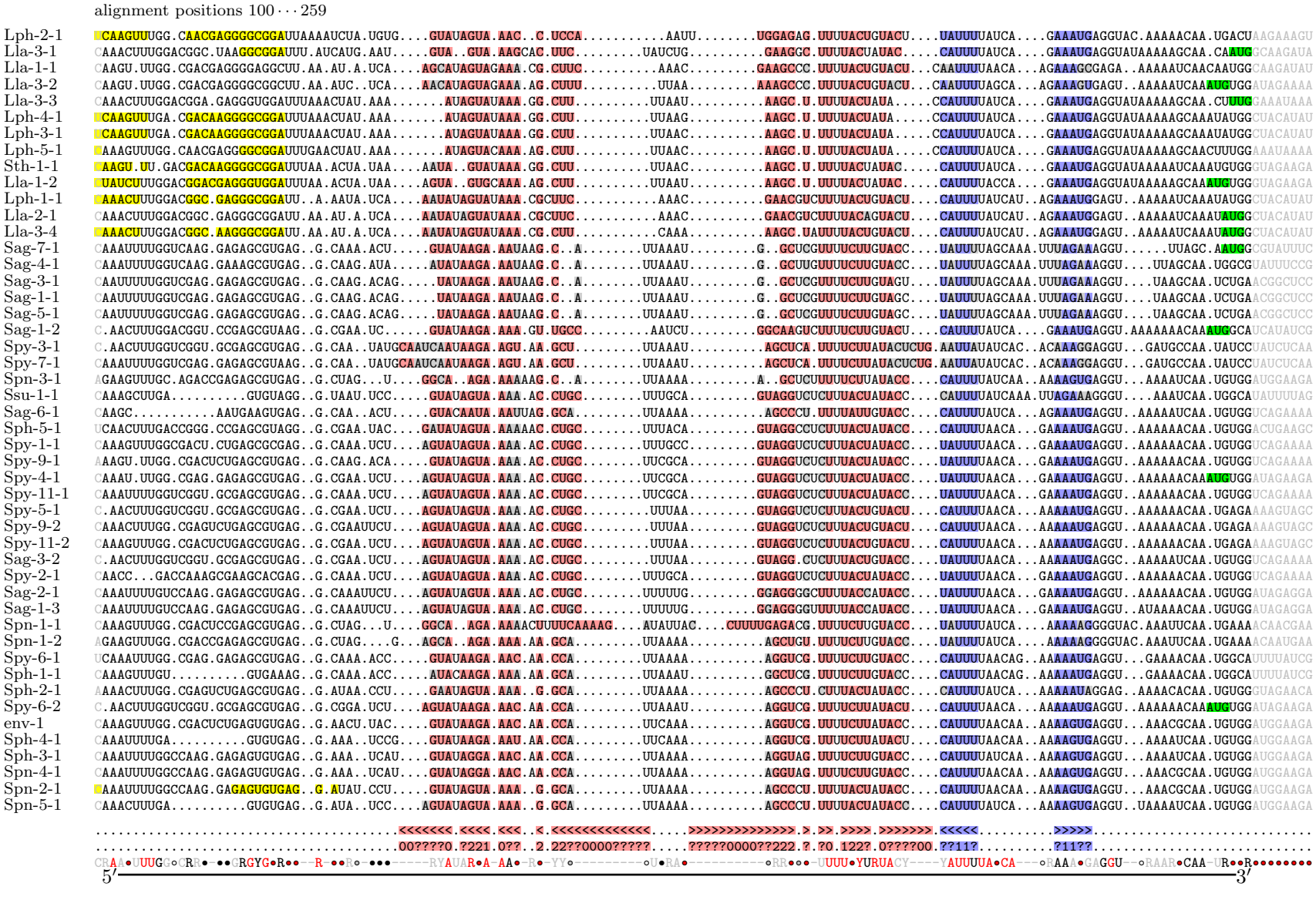
46.4 Multiple-sequence alignment

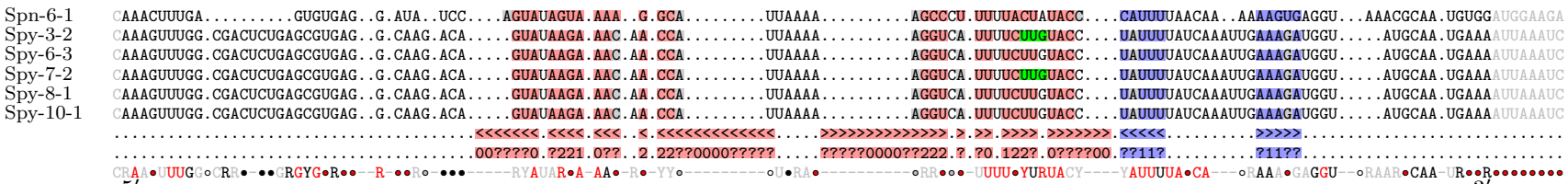
Each Lacto-*int* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 46.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Lacto-*int* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when

they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

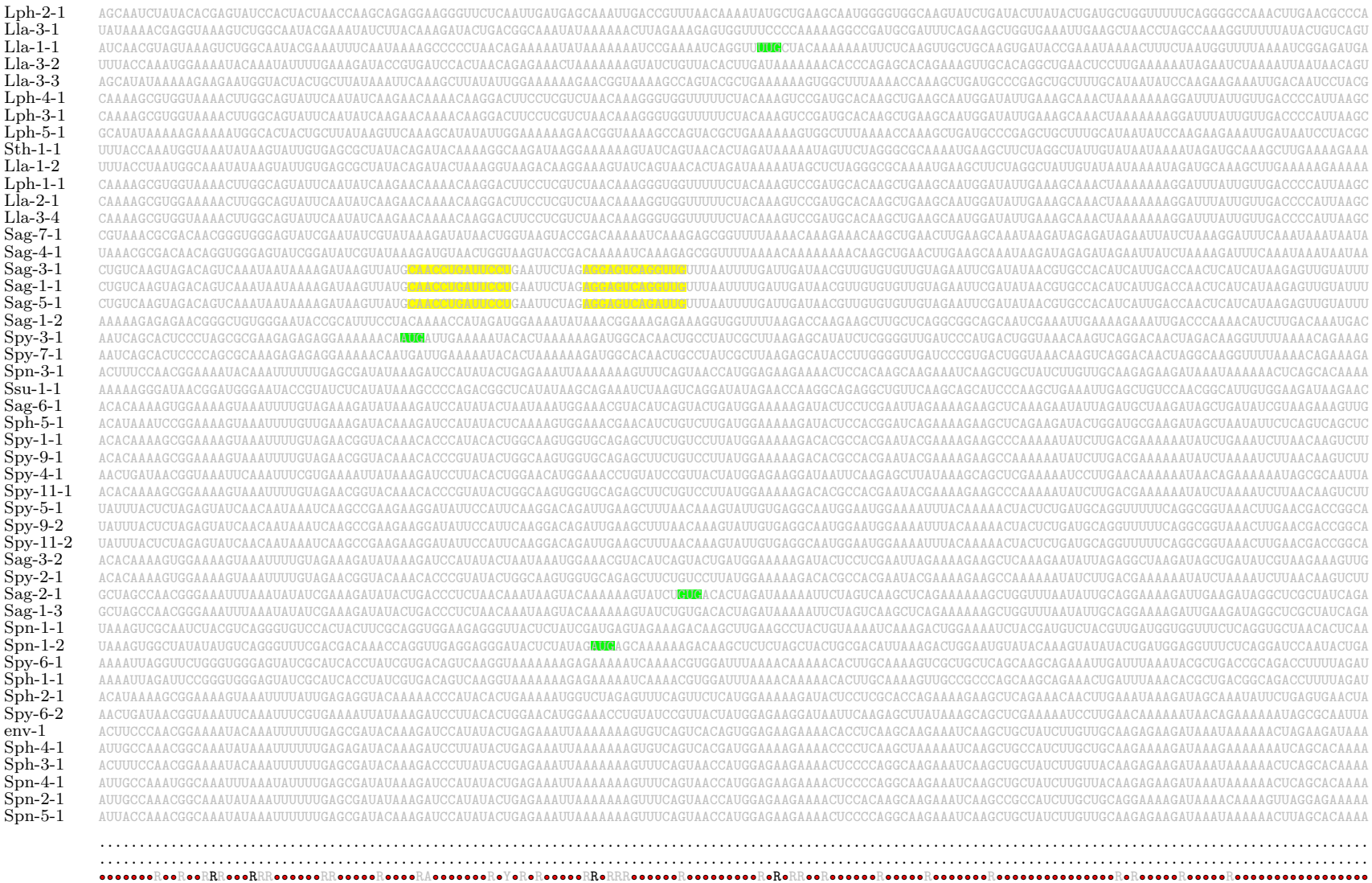


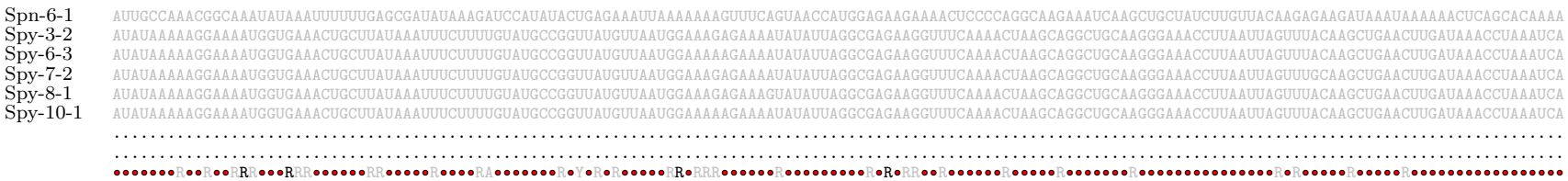

Spn-6-1
Spy-3-2
Spy-6-3
Spy-7-2
Spy-8-1
Spy-10-1


alignment positions 100 · · · 259
Lph-2-1
Lla-3-1
Lla-1-1
Lla-3-2
Lla-3-3
Lph-4-1
Lph-3-1
Lph-5-1
Sth-1-1
Lla-1-2
Lph-1-1
Lla-2-1
Lla-3-4
Sag-7-1
Sag-4-1
Sag-3-1
Sag-1-1
Sag-5-1
Sag-1-2
Spy-3-1
Spy-7-1
Spn-3-1
Ssu-1-1
Sag-6-1
Sph-5-1
Spy-1-1
Spy-9-1
Spy-4-1
Spy-11-1
Spy-5-1
Spy-9-2
Spy-11-2
Sag-3-2
Spy-2-1
Sag-2-1
Sag-1-3
Spn-1-1
Spn-1-2
Spy-6-1
Sph-1-1
Sph-2-1
Spy-6-2
env-1
Sph-4-1
Sph-3-1
Spn-4-1
Spn-2-1
Spn-5-1
5′
3′

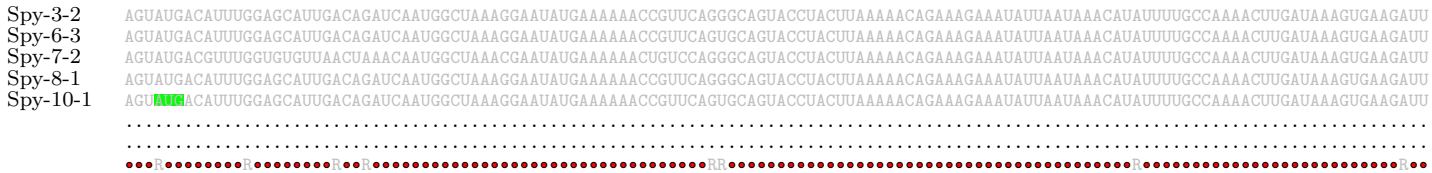


alignment positions 260 · · · 419





alignment positions 420 · · · 551



47 Lacto-*rpoB*

47.1 Taxa

The taxonomy of each organism containing a putative Lacto-*rpoB* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Lacto-*rpoB* RNA in Sections 47.2 and 47.4.

abbrev. of hits	taxonomy of species
Efa-1-1	Bacteria Firmicutes Lactobacillales Enterococcaceae *Enterococcus faecalis* V583
Efa-2-1	Bacteria Firmicutes Lactobacillales Enterococcaceae *Enterococcus faecium* DO
Lac-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus acidophilus* NCFM
Lbr-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus brevis* ATCC 367
Lca-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus casei* ATCC 334
Lde-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus delbrueckii subsp. bulgaricus* ATCC 11842
Lde-2-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus delbrueckii subsp. bulgaricus* ATCC BAA-365
Lga-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus gasseri* ATCC 33323
Ljo-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus johnsonii* NCC 533
Lre-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus reuteri* 100-23
Lre-2-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus reuteri* DSM 20016
Lsa-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus sakei subsp. sakei* 23K
Lsa-2-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus salivarius* UCC118
Ppe-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Pediococcus pentosaceus* ATCC 25745
Ooe-1-1	Bacteria Firmicutes Lactobacillales *Oenococcus oeni* ATCC BAA-1163
Ooe-2-1	Bacteria Firmicutes Lactobacillales *Oenococcus oeni* PSU-1
Lla-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. cremoris* MG1363
Lla-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. cremoris* SK11
Lla-3-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. lactis* Il1403
Sag-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 18RS21
Sag-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 2603V/R
Sag-3-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 515
Sag-4-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* A909
Sag-5-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* CJB111
Sag-6-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* COH1
Sag-7-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* H36B
Sag-8-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* NEM316
Smu-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus mutans* UA159
Spn-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* D39
Spn-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* R6
Spn-3-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP11-BS70
Spn-4-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP14-BS69
Spn-5-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP18-BS74
Spn-6-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP19-BS75
Spn-7-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP23-BS72
Spn-8-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP3-BS71
Spn-9-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP6-BS73
Spn-10-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP9-BS68
Spn-11-1 to Spn-11-2	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* TIGR4
Spy-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* M1 GAS
Spy-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* M49 591

Spy-3-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10270
Spy-4-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10394
Spy-5-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10750
Spy-6-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS2096
Spy-7-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS315
Spy-8-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS5005
Spy-9-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS6180
Spy-10-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS8232
Spy-11-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS9429
Spy-12-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* SSI-1
Spy-13-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes str.* Manfredo
Ssa-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus sanguinis* SK36
Ssu-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus suis* 05ZYH33
Ssu-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus suis* 89/1591
Ssu-3-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus suis* 98HAH33
Sth-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* CNRZ1066
Sth-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* LMD-9
Sth-3-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* LMG 18311

47.2 Gene contexts

Each Lacto-*rpoB* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Lacto-*rpoB* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 47.3. The accession number of the sequence containing each Lacto-*rpoB* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 47.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Lla-2-1	NC_008527.1	-	1876302	1876244	**RNA**→ RNA_pol_B_RPB2 (cd00653)→
Lla-3-1	NC_002662.1	-	1864719	1864661	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→
Lla-1-1	NC_009004.1	-	1970359	1970301	**RNA**→ RNA_pol_B_RPB2 (cd00653)→
Lca-1-1	NC_008526.1	-	2496672	2496615	**RNA**→ RNA_pol_Rpb2_1 (pfam04563)RNA_pol_B_RPB2 (cd00653)→ RpoC (COG0086)→
Ppe-1-1	NC_008525.1	-	1406071	1406013	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→ RpoC (COG0086)→
Ooe-2-1	NC_008528.1	-	1286567	1286534	**RNA**→ RpoB (COG0085)→ RpoC (COG0086)→
Ooe-1-1	NZ_AAUV01000045.1	-	134	101	**RNA**→
Lsa-2-1	NC_007929.1	+	234847	234900	**RNA**→ RNA_pol_B_RPB2 (cd00653)→ RpoC (COG0086)→
Efa-1-1	NC_004668.1	-	3112617	3112562	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→
Efa-2-1	NZ_AAAK03000029.1	-	24725	24667	**RNA**→
Lbr-1-1	NC_008497.1	-	1660516	1660460	**RNA**→ RNA_pol_B_RPB2 (cd00653)→ RpoC (COG0086)→
Lde-2-1	NC_008529.1	+	328930	328957	**RNA**→ RNA_pol_B_RPB2 (cd00653)→ RpoC (COG0086)→
Lde-1-1	NC_008054.1	+	334371	334398	**RNA**→ RNA_pol_B_RPB2 (cd00653)→ RpoC (COG0086)→
Lga-1-1	NC_008530.1	+	321365	321392	**RNA**→ RNA_pol_Rpb2_1 (pfam04563)RNA_pol_B_RPB2 (cd00653)→ RpoC (COG0086)→
Ljo-1-1	NC_005362.1	+	370674	370701	**RNA**→ RNA_pol_Rpb2_1 (pfam04563)RNA_pol_B_RPB2 (cd00653)→ RpoC (COG0086)→
Lac-1-1	NC_006814.2	+	278920	278947	**RNA**→ RNA_pol_B_RPB2 (cd00653)→ RpoC (COG0086)→
Lsa-1-1	NC_007576.1	-	1751463	1751404	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→ RpoC (COG0086)→
Lre-2-1	NC_009513.1	-	1544739	1544681	**RNA**→ RNA_pol_B_RPB2 (cd00653)→ RpoC (COG0086)→ COG3371 (COG3371)→
Lre-1-1	NZ_AAPZ01000014.1	-	9767	9709	**RNA**→ RNA_pol_B_RPB2 (cd00653)→ RpoC (COG0086)→ ABC transporter, permease protein (putative)→
Spy-8-1	NC_007297.1	+	93919	93976	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→ RpoC (COG0086)→
Spy-1-1	NC_002737.1	+	93363	93420	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→ RpoC (COG0086)→

Spy-10-1	NC_003485.1	+	94021	94078	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→ RpoC (COG0086)→
Spy-12-1	NC_004606.1	+	88129	88186	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→
Spy-7-1	NC_004070.1	+	94129	94186	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→ RpoC (COG0086)→
Spy-9-1	NC_007296.1	+	93407	93464	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→ RpoC (COG0086)→
Spy-11-1	NC_008021.1	+	93729	93786	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→ RNA_pol_Rpb2_7 (pfam04560)→ RpoC (COG0086)→
Spy-3-1	NC_008022.1	+	94100	94157	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→ RpoC (COG0086)→
Spy-6-1	NC_008023.1	+	93707	93764	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→ RpoC (COG0086)→
Spy-13-1	NC_009332.1	+	95746	95803	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→ RpoC (COG0086)→
Spy-5-1	NC_008024.1	+	97549	97606	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→ RpoC (COG0086)→
Spy-2-1	NZ_AAFV01000007.1	-	16617	16560	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→
Spy-4-1	NC_006086.1	+	134778	134835	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→ RpoC (COG0086)→
Spn-1-1	NC_008533.1	-	1759407	1759350	**RNA**→ RpoB (COG0085)→ RpoC (COG0086)→
Spn-2-1	NC_003098.1	-	1751875	1751818	**RNA**→ RpoB (COG0085)→ RpoC (COG0086)→
Spn-10-1	NZ_ABAB01000008.1	+	32551	32608	**RNA**→ RpoB (COG0085)→ RpoC (COG0086)→
Spn-7-1	NZ_ABAG01000011.1	-	37093	37036	**RNA**→ RpoB (COG0085)→ RpoC (COG0086)→
Spn-8-1	NZ_AAZZ01000010.1	-	47522	47465	**RNA**→ RpoB (COG0085)→ RpoC (COG0086)→
Spn-9-1	NZ_ABAA01000014.1	-	8798	8741	**RNA**→ RpoB (COG0085)→ RpoC (COG0086)→
Spn-11-1	NC_003028.2	-	1870209	1870152	**RNA**→ RpoB (COG0085)→ RpoC (COG0086)→
Spn-11-2	NZ_AAGY02000001.1	-	26869	26812	**RNA**→ RpoB (COG0085)→ RpoC (COG0086)→
Spn-5-1	NZ_ABAE01000010.1	-	47412	47355	**RNA**→ RpoB (COG0085)→ RpoC (COG0086)→
Spn-6-1	NZ_ABAF01000008.1	+	39164	39221	**RNA**→ RpoB (COG0085)→ RpoC (COG0086)→
Spn-3-1	NZ_ABAC01000008.1	-	47353	47296	**RNA**→ RpoB (COG0085)→ RpoC (COG0086)→
Spn-4-1	NZ_ABAD01000012.1	-	41928	41871	**RNA**→ RpoB (COG0085)→ RpoC (COG0086)→
Ssu-1-1	NC_009442.1	+	110341	110398	**RNA**→ RpoB (COG0085)→
Ssu-3-1	NC_009443.1	+	110247	110304	**RNA**→ RpoB (COG0085)→
Ssu-2-1	NZ_AAFA02000016.1	-	14132	14075	**RNA**→ RpoB (COG0085)→
Ssa-1-1	NC_009009.1	+	174628	174686	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→ RpoC (COG0086)→
Smu-1-1	NC_004350.1	-	1864453	1864396	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→
Sag-2-1	NC_004116.1	+	179028	179085	**RNA**→ RpoB (COG0085)→
Sag-1-1	NZ_AAJO01000041.1	-	7940	7883	**RNA**→ RpoB (COG0085)→ RpoC (COG0086)→
Sag-6-1	NZ_AAJR01000005.1	+	2154	2211	**RNA**→ RpoB (COG0085)→ RpoC (COG0086)→
Sag-8-1	NC_004368.1	+	178933	178990	**RNA**→ RpoB (COG0085)→
Sag-3-1	NZ_AAJP01000013.1	-	30093	30036	**RNA**→ RpoB (COG0085)→ RpoC (COG0086)→
Sag-4-1	NC_007432.1	+	188955	189012	**RNA**→ RpoB (COG0085)→
Sag-5-1	NZ_AAJQ01000006.1	+	7848	7905	**RNA**→ RpoB (COG0085)→ RpoC (COG0086)→
Sag-7-1	NZ_AAJS01000039.1	-	14129	14072	**RNA**→ RpoB (COG0085)→ RpoC (COG0086)→
Sth-3-1	NC_006448.1	-	1667715	1667658	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→
Sth-1-1	NC_006449.1	-	1668340	1668283	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→
Sth-2-1	NC_008532.1	-	1720910	1720853	**RNA**→ RpoB (COG0085)RNA_pol_Rpb2_1 (pfam04563)→

47.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 47.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Lacto-*rpoB* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

cd00653 (12) RNA polymerase beta subunit.
COG0085 (46) DNA-directed RNA polymerase, beta subunit/140 kD subunit [Transcription]
COG0086 (42) DNA-directed RNA polymerase, beta' subunit/160 kD subunit [Transcription]
COG3371 (1) Predicted membrane protein [Function unknown]
pfam04560 (1) RNA polymerase Rpb2, domain 7.
pfam04563 (25) RNA polymerase beta subunit.

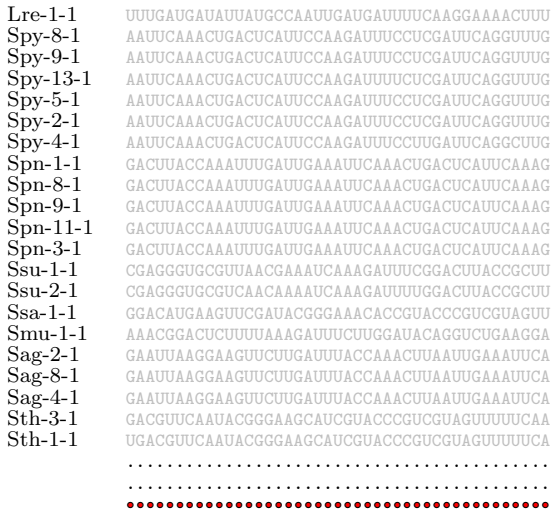
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Lde-1-1, Sag-1-1, Sag-3-1, Sag-5-1, Sag-6-1, Sag-7-1, Spn-10-1, Spn-11-2, Spn-2-1, Spn-4-1, Spn-5-1, Spn-6-1, Spn-7-1, Spy-1-1, Spy-10-1, Spy-11-1, Spy-12-1, Spy-3-1, Spy-6-1, Spy-7-1, Ssu-3-1, Sth-2-1

47.4 Multiple-sequence alignment

Each Lacto-*rpoB* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 47.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Lacto-*rpoB* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

Ssu-2-1
Ssa-1-1
Smu-1-1
Sag-2-1
Sag-8-1
Sag-4-1
Sth-3-1
Sth-1-1
alignment positions 100 · · · 259
Lla-2-1
Lla-3-1
Lla-1-1
Lca-1-1
Ppe-1-1
Ooe-2-1
Ooe-1-1
Lsa-2-1
Efa-1-1
Efa-2-1
Lbr-1-1
Lde-2-1
Lga-1-1
Ljo-1-1
Lac-1-1
Lsa-1-1
Lre-2-1
Lre-1-1
Spy-8-1
Spy-9-1
Spy-13-1
Spy-5-1
Spy-2-1
Spy-4-1
Spn-1-1
Spn-8-1
Spn-9-1
Spn-11-1
Spn-3-1
Ssu-1-1
Ssu-2-1
Ssa-1-1
Smu-1-1
Sag-2-1
Sag-8-1
Sag-4-1
Sth-3-1
Sth-1-1
5′
3′
alignment positions 260 · · · 419

Lla-2-1
Lla-3-1
Lla-1-1
Lca-1-1
Ppe-1-1
Ooe-2-1
Ooe-1-1
Lsa-2-1
Efa-1-1
Efa-2-1
Lbr-1-1
Lde-2-1
Lga-1-1
Ljo-1-1
Lac-1-1
Lsa-1-1
Lre-2-1
Lre-1-1
Spy-8-1
Spy-9-1
Spy-13-1
Spy-5-1
Spy-2-1
Spy-4-1
Spn-1-1
Spn-8-1
Spn-9-1
Spn-11-1
Spn-3-1
Ssu-1-1
Ssu-2-1
Ssa-1-1
Smu-1-1
Sag-2-1
Sag-8-1
Sag-4-1
Sth-3-1
Sth-1-1
alignment positions 420 · · · 462
Lla-2-1
Lla-3-1
Lla-1-1
Lca-1-1
Ppe-1-1
Ooe-2-1
Ooe-1-1
Lsa-2-1
Efa-1-1
Efa-2-1
Lbr-1-1
Lde-2-1
Lga-1-1
Ljo-1-1
Lac-1-1
Lsa-1-1
Lre-2-1



48 Lacto-*usp*

48.1 Taxa

The taxonomy of each organism containing a putative Lacto-*usp* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Lacto-*usp* RNA in Sections 48.2 and 48.4.

abbrev. of hits	taxonomy of species
Efa-1-1	Bacteria Firmicutes Lactobacillales Enterococcaceae *Enterococcus faecalis* V583
Lac-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus acidophilus* NCFM
Lde-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus delbrueckii subsp. bulgaricus* ATCC 11842
Lde-2-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus delbrueckii subsp. bulgaricus* ATCC BAA-365
Lga-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus gasseri* ATCC 33323
Ljo-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus johnsonii* NCC 533

48.2 Gene contexts

Each Lacto-*usp* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Lacto-*usp* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 48.3. The accession number of the sequence containing each Lacto-*usp* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 48.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Efa-1-1	NC_004668.1	+	1051896	1051978	hypo→ **RNA**→ hypo→
Ljo-1-1	NC_005362.1	-	846317	846228	MGS1 (COG2256)→ **RNA**→ hypo→
Lga-1-1	NC_008530.1	+	1232606	1232695	MGS1 (COG2256)→ **RNA**→ hypo→
Lac-1-1	NC_006814.2	-	767168	767078	MGS1 (COG2256)→ **RNA**→ Usp (pfam00582)→
Lde-2-1	NC_008529.1	-	603031	602943	MGS1 (COG2256)→ **RNA**→ hypo→
Lde-1-1	NC_008054.1	-	608889	608801	MGS1 (COG2256)→ **RNA**→ hypo→

48.3 Conserved domains

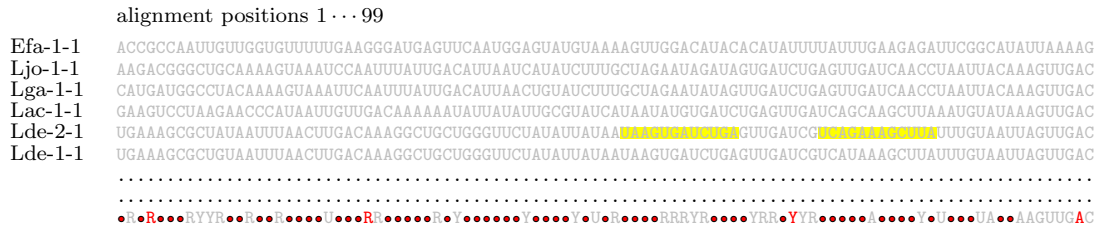
Conserved domains found in protein-coding genes listed in Section 48.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Lacto-*usp* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

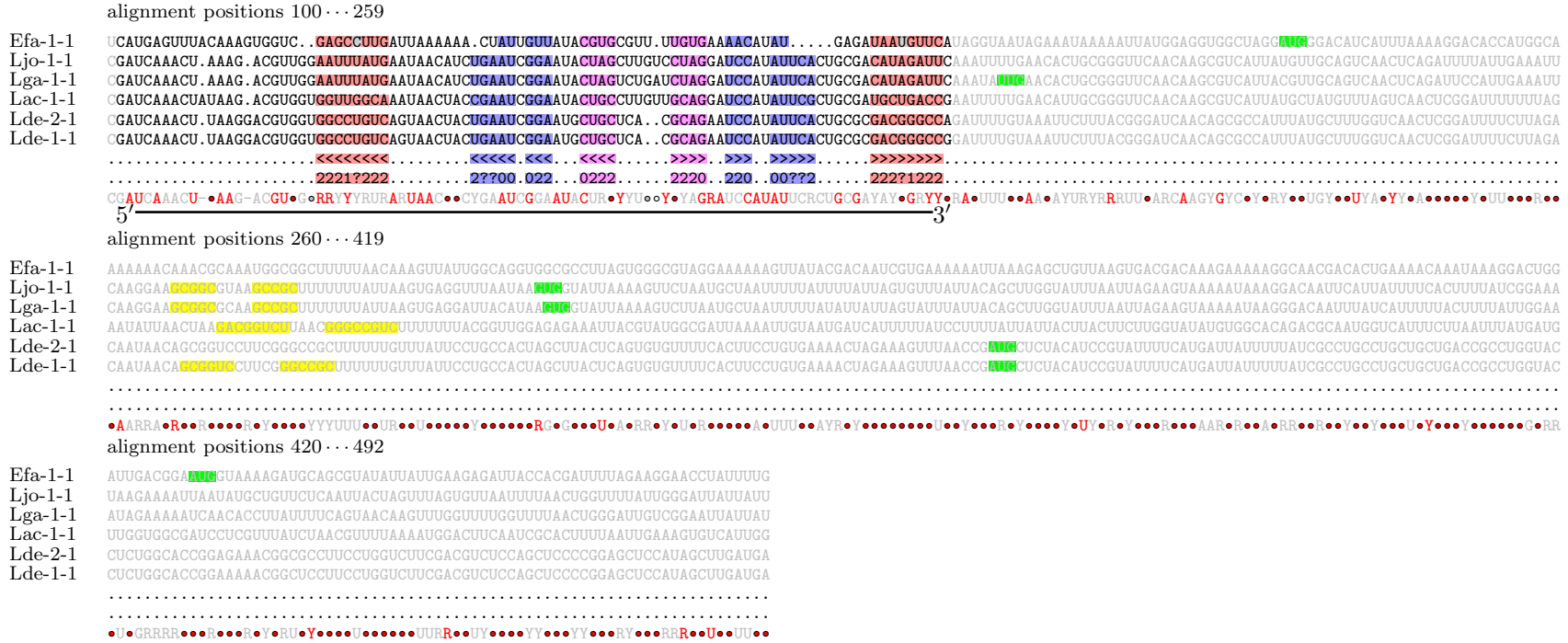
u COG2256 (5) ATPase related to the helicase subunit of the Holliday junction resolvase [DNA replication, recombination, and repair]

D pfam00582 (1) Universal stress protein family.

48.4 Multiple-sequence alignment

Each Lacto-*usp* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 48.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Lacto-*usp* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.





49 Leu/phe leader

49.1 Taxa

The taxonomy of each organism containing a putative Leu/phe leader RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Leu/phe leader RNA in Sections 49.3 and 49.5.

abbrev. of hits	taxonomy of species
Lla-1-1 to Lla-1-3	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. cremoris* MG1363
Lla-2-1 to Lla-2-3	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. cremoris* SK11
Lla-3-1 to Lla-3-3	Bacteria Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. lactis* Il1403

49.2 Notes

Superscript numbers are used to annotate any Leu/phe leader RNA that has special characteristics (described below). These numeric annotations will be used in Sections 49.3 and 49.5. The meaning of the numbers are as follows.

[1] This RNA contains a short ORF with three consecutive phenylalanine codons (UUC or UUU).

[2] This RNA contains a short ORF with three consecutive leucine codons (UUA, UUG, CUA, CUC, CUG or CUU).

49.3 Gene contexts

Each Leu/phe leader RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Leu/phe leader RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 49.4. The accession number of the sequence containing each Leu/phe leader RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 49.1. Superscript numbers refer to annotations listed in Section 49.2

abbrev.	Seq. accession		5′ at	3′ at	genes
[1]Lla-3-1	NC_002662.1	-	1281659	1281495	**RNA**→ AroG (COG0722)→
[1]Lla-2-1	NC_008527.1	-	1276225	1276061	**RNA**→ hypo→
[1]Lla-1-1	NC_009004.1	+	1198427	1198591	**RNA**→ AroG (COG0722)→
[2]Lla-3-2	NC_002662.1	-	820178	820018	**RNA**→ AnsP (COG1113)→
[2]Lla-1-2	NC_009004.1	+	1728588	1728748	**RNA**→ AnsP (COG1113)→
[2]Lla-2-2	NC_008527.1	-	808768	808614	**RNA**→ AnsP (COG1113)→
[2]Lla-2-3	NC_008527.1	-	1233390	1233230	**RNA**→ LeuA (COG0119)→ hypo→ COG3152 (COG3152)→
[2]Lla-1-3	NC_009004.1	-	1258179	1258019	**RNA**→ Aconitase (pfam00330)→ hypo→ LeuD (COG0066)→ ModF (COG1119)→
[2]Lla-3-3	NC_002662.1	+	1239635	1239796	**RNA**→ Iso_dh (pfam00180)→

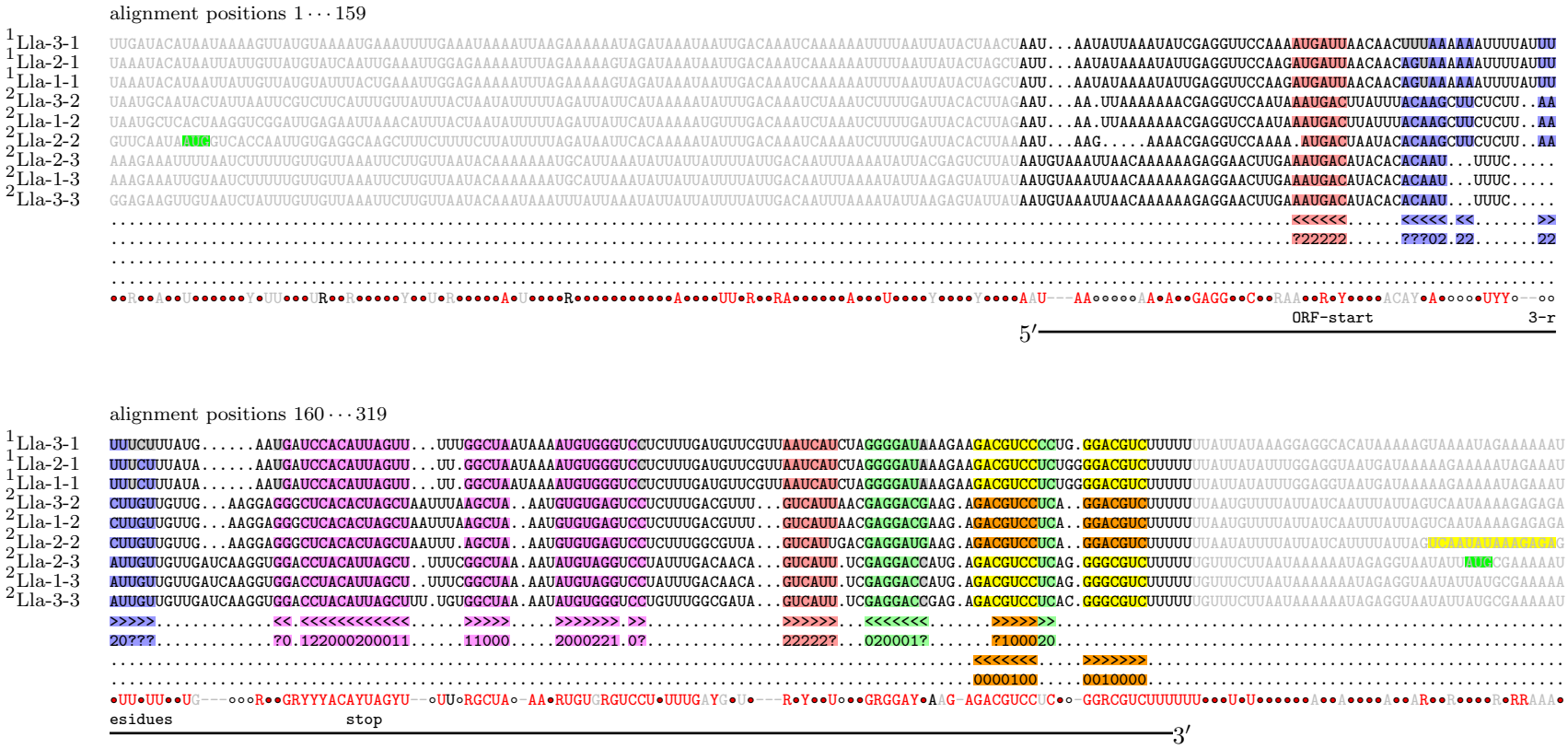
49.4 Conserved domains

Conserved domains found in protein-coding genes listed in Section 49.3 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Leu/phe leader RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

COG0066 (1) 3-isopropylmalate dehydratase small subunit [Amino acid transport and metabolism]
COG0119 (1) Isopropylmalate/homocitrate/citramalate synthases [Amino acid transport and metabolism]
COG0722 (2) 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase [Amino acid transport and metabolism]
COG1113 (3) Gamma-aminobutyrate permease and related permeases [Amino acid transport and metabolism]
COG1119 (1) ABC-type molybdenum transport system, ATPase component/photorepair protein PhrA [Inorganic ion transport and metabolism]
COG3152 (1) Predicted membrane protein [Function unknown]
pfam00180 (1) Isocitrate/isopropylmalate dehydrogenase.
pfam00330 (1) Aconitase family (aconitate hydratase).

49.5 Multiple-sequence alignment

Each Leu/phe leader RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 49.1. Superscript numbers refer to annotations listed in Section 49.2. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Leu/phe leader RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



50 *livK*

50.1 Taxa

The taxonomy of each organism containing a putative *livK* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *livK* RNA in Sections 50.2 and 50.4.

abbrev. of hits	taxonomy of species
Psy-1-1 to Psy-1-2	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. phaseolicola* 1448A
Psy-2-1 to Psy-2-3	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1 to Psy-3-3	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000

50.2 Gene contexts

Each *livK* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *livK* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 50.3. The accession number of the sequence containing each *livK* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 50.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Psy-2-1	NC_007005.1	-	2563011	2562928	**RNA**→ LivK (COG0683)→ LivH (COG0559)→ LivM (COG4177)→ COG4674 (COG4674)→ ABC_TM1139_LivF_branched (cd03224)→ UreE (cd00571)→ UreF (COG0830)→ Urease_gamma (pfam00547)Urease_beta (pfam00699)→ Urease_alpha (cd00375)UreC (COG0804)→ HypB (COG0378)→ UreD (pfam01774)→
Psy-3-1	NC_004578.1	-	2667241	2667158	**RNA**→ LivK (COG0683)→ LivH (COG0559)→ LivM (COG4177)→ COG4674 (COG4674)→ ABC_TM1139_LivF_branched (cd03224)→ UreE (cd00571)→ UreF (COG0830)→ Urease_gamma (pfam00547)Urease_beta (pfam00699)→
Psy-1-1	NC_005773.3	-	3459167	3459084	**RNA**→ LivK (COG0683)→ AmiR (COG3707)→
Psy-2-2	NC_007005.1	+	2563385	2563468	**RNA**→ LivK (COG0683)→ AmiR (COG3707)→
Psy-3-2	NC_004578.1	+	2667662	2667745	**RNA**→ LivK (COG0683)→ AmiR (COG3707)→
Psy-3-3	NC_004578.1	+	2669688	2669769	**RNA**→ Amidase (pfam01425)→
Psy-1-2	NC_005773.3	-	3457138	3457057	**RNA**→ Amidase (pfam01425)→
Psy-2-3	NC_007005.1	+	2565410	2565491	**RNA**→ Amidase (pfam01425)→

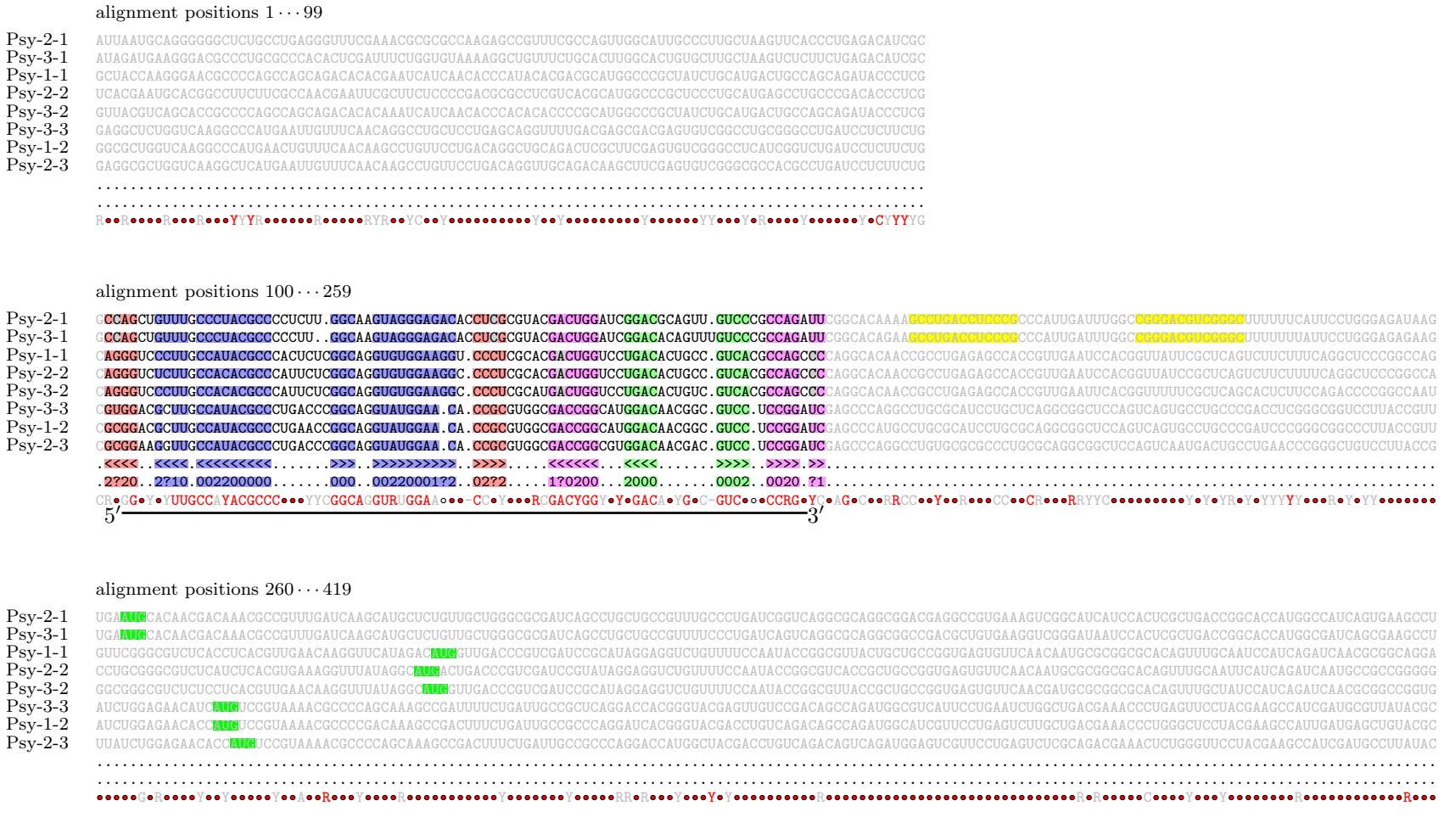
50.3 Conserved domains

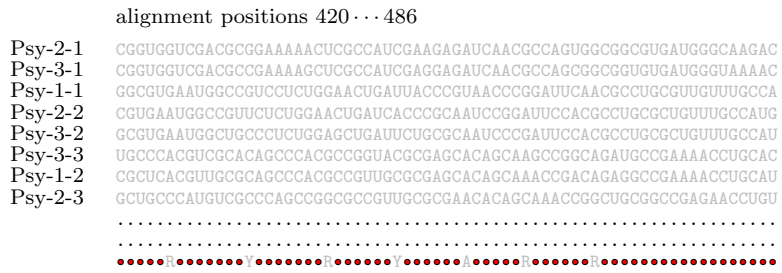
Conserved domains found in protein-coding genes listed in Section 50.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *livK* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

cd00375 (1) Urease alpha-subunit; Urease is a nickel-dependent metalloenzyme that catalyzes the hydrolysis of urea to form ammonia and carbon dioxide.
cd00571 (2) UreE urease accessory protein.
cd03224 (2) LivF (TM1139) is part of the LIV-I bacterial ABC-type two-component transport system that imports neutral, branched-chain amino acids.
COG0378 (1) Ni2+-binding GTPase involved in regulation of expression and maturation of urease and hydrogenase [Posttranslational modification, protein turnover, chaperones / Transcription]
COG0559 (2) Branched-chain amino acid ABC-type transport system, permease components [Amino acid transport and metabolism]
COG0683 (5) ABC-type branched-chain amino acid transport systems, periplasmic component [Amino acid transport and metabolism]
COG0804 (1) Urea amidohydrolase (urease) alpha subunit [Amino acid transport and metabolism]
COG0830 (2) Urease accessory protein UreF [Posttranslational modification, protein turnover, chaperones]
COG3707 (3) Response regulator with putative antiterminator output domain [Signal transduction mechanisms]
COG4177 (2) ABC-type branched-chain amino acid transport system, permease component [Amino acid transport and metabolism]
COG4674 (2) Uncharacterized ABC-type transport system, ATPase component [General function prediction only]
pfam00547 (2) Urease, gamma subunit.
pfam00699 (2) Urease beta subunit.
pfam01425 (3) Amidase.
pfam01774 (1) UreD urease accessory protein.

50.4 Multiple-sequence alignment

Each *livK* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 50.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *livK* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.





51 Lnt

51.1 Taxa

The taxonomy of each organism containing a putative Lnt RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Lnt RNA in Sections 51.2 and 51.4.

abbrev.	of hits	taxonomy of species
Cte-1-1		Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobaculum *Chlorobium tepidum* TLS
Cch-1-1		Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium chlorochromatii* CaD3
Cfe-1-1		Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium ferrooxidans* DSM 13031
Cli-1-1		Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium limicola* DSM 245
Cph-1-1		Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium phaeobacteroides* BS1
Cph-2-1		Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium phaeobacteroides* DSM 266
Cph-3-1		Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium phaeovibrioides* DSM 265 (Prosthecochloris vibrioformis
Plu-1-1		Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Pelodictyon luteolum* DSM 273
Pph-1-1		Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Pelodictyon phaeoclathratiforme* BU-1
Pae-1-1		Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae *Prosthecochloris aestuarii* DSM 271

51.2 Gene contexts

Each Lnt RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Lnt RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 51.3. The accession number of the sequence containing each Lnt RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 51.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Cch-1-1	NC_007514.1	+	1796390	1796458	**RNA**→ Lnt (COG0815)→
Cph-3-1	NC_009337.1	-	1481205	1481140	**RNA**→ Lnt (COG0815)→
Plu-1-1	NC_007512.1	-	1687310	1687245	**RNA**→ Lnt (COG0815)→
Pae-1-1	NZ_AAIJ01000012.1	-	6941	6882	**RNA**→ Lnt (COG0815)→

Cph-1-1	NZ_AAIC01000051.1	-	9191	9131	**RNA**→ Lnt (COG0815)→
Pph-1-1	NZ_AAIK01000033.1	-	13712	13649	**RNA**→ Lnt (COG0815)→
Cli-1-1	NZ_AAHJ01000019.1	+	1286	1345	**RNA**→ Lnt (COG0815)→
Cfe-1-1	NZ_AASE01000007.1	-	53638	53577	**RNA**→ Lnt (COG0815)→
Cte-1-1	NC_002932.3	-	1404024	1403967	**RNA**→ Lnt (COG0815)→
Cph-2-1	NC_008639.1	-	2186687	2186628	**RNA**→ Lnt (COG0815)→

51.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 51.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Lnt RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

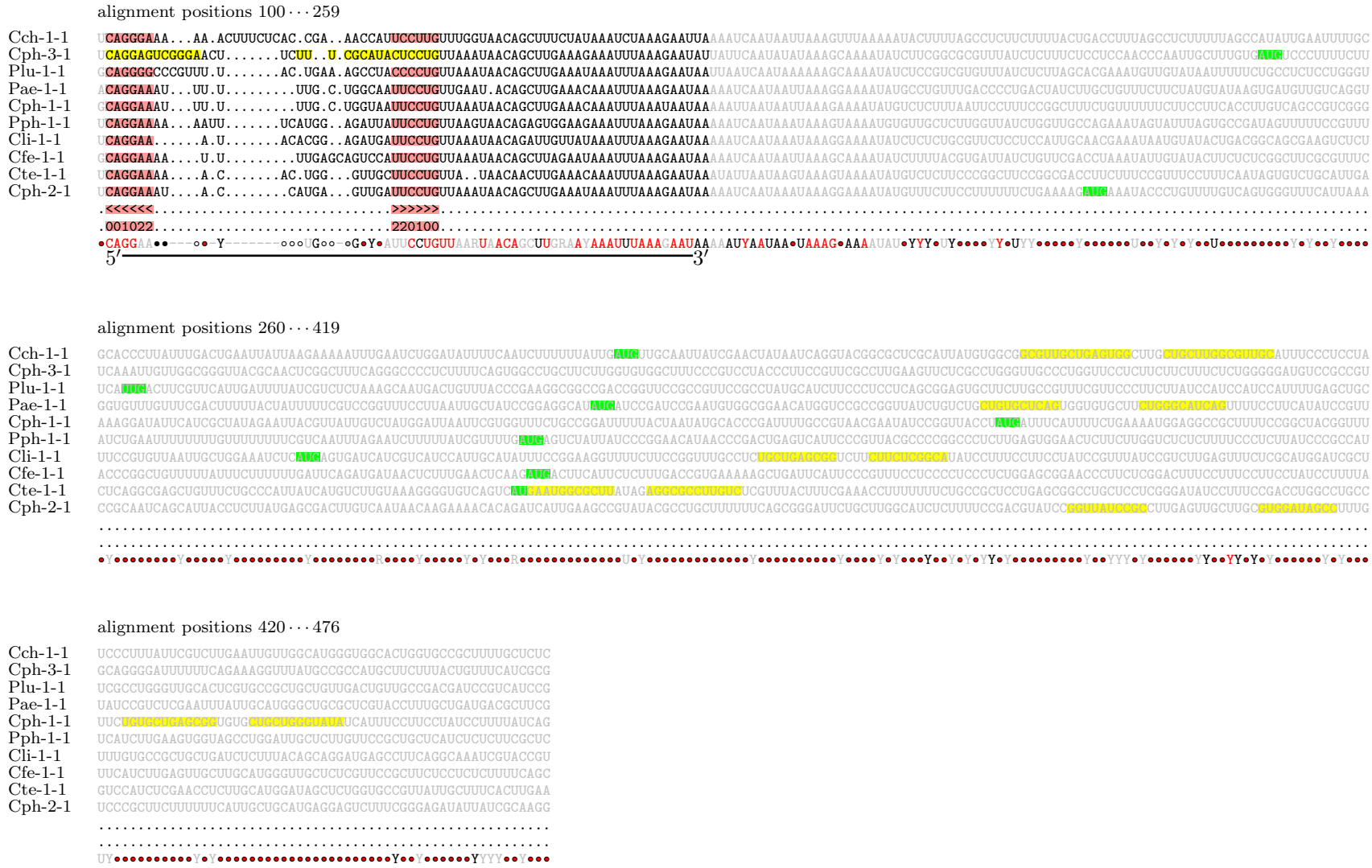
COG0815 (10) Apolipoprotein N-acyltransferase [Cell envelope biogenesis, outer membrane]

51.4 Multiple-sequence alignment

Each Lnt RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 51.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Lnt RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

alignment positions 1 · · · 99

```
Cch-1-1   AACGAUCUCAUAGUCUGAGUGAGCGGUAGUGGUGUCUUUAGUGCCGAAAAGAGCCAUAAUCAGUUCUCCUUUUUCUGGUGGCGUUGAAAUUGGAUAUGU
Cph-3-1   UACGAUUUCGUAGUCGGAGUGUGCGGUGGUUGCGUCUUUAGUGCCGAAAAGAGCCAUAAUGAUUUCUCCUUUUUCUGAUGGCGUUGAAAUUGGAUAUUU
Plu-1-1   AACGAUUUCGUAAUCGGAGUGCGCGGUGGUCGCGUCUUUAGUGCCGAAAAGAGCCAUAAUGAUUUCUCCUUUUUCUGAUGGCGUUGAAAUUGGAUAUUU
Pae-1-1   GAGAAUAAUCUCAUAAUCGGAAUGGGCGGUUGUAGUGUCUUUUGUGCCGAAUAGAGCCAUAAUUGAAUCUCCUUUUUCUGGUGGCGUUGAAAAUUUGUU
Cph-1-1   UCGAGAAUAAUCUCGUAAUCAGCAUGGGCGGUUGUGUUGUCUUUUGUGCCGAAUAGAGCCAUAAGUCAAUCUCCUUUUUGGUGGCGUUUAGAAUUUUCA
Pph-1-1   AACGAUUUCGUAAUCUGAGUGUGCGGUAGUGGUGUCUUUUGUGCCGAAAAGAGCCAUAAUGAGUUCUCCUUUUUCUAAUGGCGUUGAAAUUGGAUAUUU
Cli-1-1   AAUGAUCUCAUAGUCCGAGUGAGCGGUGGUGGUGUCUUUUGUGCCGAAAAGAGCCAUAAUGAUUUCUCCUUUUUCUGGUGGCGUUGAAAUUGGAUAUUU
Cfe-1-1   AACGAUCUCGUAAUCGGAGUGUGCGGUGGUGGUGUCUUUAGUGCCGAAAAGAGCCAUAGUGAUUUUCUCCUUUUUCUAAUGGCGUUGAAAUUGGAUAUUU
Cte-1-1   AGUACGAUUUCAUAAUCGGAGUGAGCGGUCGUCACGUCGUUUGAGCCGAAAAGAGCCAUAAUGAUUCUCCUUUUUCUGGUGGCGUUUUAAAAUGUAUUU
Cph-2-1   AAUGAUUUCAUAAUCGGAAUGGGCGGUGGUGGUGUCUUUUGUGCCGAAAAGAGCCAUAGUGAUUUCUCCUUUUUCUGGUGGCGUUGAAAUUGGAUAUUU
          ....................................................................................................
          ....................................................................................................
          RAYRAU•UYRUARU••RAGYG•G•GGYGGU•GYGUYUUY••UGYYG•RAAGAGY•RYARURA•UUYYYCUUUUUYU•RUGGYG•UG••A••RRRUAUUU
```



52 *manA*

52.1 Taxa

The taxonomy of each organism containing a putative *manA* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *manA* RNA in Sections 52.2 and 52.4.

abbrev. of hits	taxonomy of species
Psp-1-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Photobacterium sp.* SKA34
Van-1-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Vibrio angustum* S14
Pph-1-1	Viruses dsDNA viruses, no RNA stage Caudovirales Myoviridae T4-like viruses *Prochlorococcus phage* P-SSM2
Pph-2-1	Viruses dsDNA viruses, no RNA stage Caudovirales Myoviridae T4-like viruses *Prochlorococcus phage* P-SSM4
env-1 to env-191	environmental samples

52.2 Gene contexts

Each *manA* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *manA* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 52.3. The accession number of the sequence containing each *manA* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 52.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	BABC01000017.1	-	4078	3906	**RNA**→
env-2	JCVI_SCAF_1101668077758	+	594	782	**RNA**→
env-3	JCVI_SCAF_1101668133563	+	1274	1446	**RNA**→ Asn tRNA→
env-4	JCVI_SCAF_1101668685699	+	846	1019	**RNA**→
env-5	JCVI_SCAF_1096626939122	+	284	490	**RNA**→
env-6	JCVI_SCAF_1096626582184	+	627	807	**RNA**→
env-7	JCVI_SCAF_1101667041980	+	196	396	**RNA**→
env-8	JCVI_SCAF_1101667233102	+	752	951	**RNA**→
env-9	JCVI_SCAF_1101668483252	+	508	708	**RNA**→ IbpA (COG0071)? (GOS1219)→
env-10	JCVI_SCAF_1101668008650	-	691	491	**RNA**→ ? (GOS7381)→
env-11	JCVI_SCAF_1096627847578	-	529	333	**RNA**→
Van-1-1	NZ_AAOJ01000001.1	+	468855	469051	**RNA**→ hypo→
Psp-1-1	NZ_AAOU01000012.1	+	1213	1409	**RNA**→ hypo→
env-12	JCVI_SCAF_1101668063070	-	1157	956	**RNA**→ Pseudo tRNA→
env-13	JCVI_SCAF_1101667156636	-	446	261	**RNA**→
env-14	JCVI_SCAF_1096627831874	+	360	543	**RNA**→ ? (GOS11389660)→
env-15	JCVI_SCAF_1096627576479	+	1253	1464	**RNA**→
env-16	JCVI_SCAF_1101668015939	-	539	330	**RNA**→
env-17	JCVI_SCAF_1096627496128	-	1315	1114	**RNA**→ COG1059 (COG1059)? (GOS11384261)→
env-18	JCVI_SCAF_1096628383680	-	543	372	**RNA**→
env-19	JCVI_SCAF_1101668335187	+	1175	1398	**RNA**→ ? (GOS1582)→
env-20	JCVI_SCAF_1096628395834	+	371	593	**RNA**→ ? (GOS11025)→ ? (GOS17718)→
env-21	JCVI_SCAF_1101668007515	-	1391	1178	**RNA**→ Gly tRNA→ ? (GOS18444)→ MannoseP_isomer (pfam01050)? (GOS1274)→ ? (GOS15843)→ ? (GOS17089)→
env-22	JCVI_SCAF_1101668455072	+	466	666	**RNA**→ Thr tRNA→ Gly tRNA→ ? (GOS18444)→
env-23	JCVI_SCAF_1096627787760	+	727	927	**RNA**→ Gly tRNA→ ? (GOS18444)→ Asn tRNA→ Val tRNA→ ? (GOS15843)→ Ser tRNA→ Pro tRNA→ 5S_rRNA→
env-24	JCVI_SCAF_1101667251432	-	356	156	**RNA**→ Thr tRNA→
env-25	JCVI_SCAF_1096627182270	+	995	1197	**RNA**→ Copper-bind (pfam00127)? (GOS9362)→ Transaldolase_FSA (cd00956)? (GOS133)→
env-26	JCVI_SCAF_1096628019100	+	61	263	**RNA**→ PetE (COG3794)? (GOS9362)→ ? (GOS133)→ Transaldolase_FSA (cd00956)? (GOS133)→
env-27	JCVI_SCAF_1101668197246	-	1203	993	**RNA**→ Transaldolase_FSA (cd00956)? (GOS133)→

env-28	JCVI_SCAF_1096627922983	+	1555	1758	**RNA**→
env-29	JCVI_SCAF_1101668723105	-	635	443	**RNA**→ ? (GOS6732)→
env-30	JCVI_SCAF_1096627173391	-	1868	1676	**RNA**→ ? (GOS6732)→
env-31	JCVI_SCAF_1101668704815	+	344	544	**RNA**→ ? (GOS11589240)→
env-32	JCVI_SCAF_1096627081214	+	162	356	**RNA**→
env-33	JCVI_SCAF_1096626975268	+	410	600	**RNA**→ ←? (GOS14155524)
env-34	JCVI_SCAF_1096626936369	+	684	882	**RNA**→ ←? (GOS14155524)
env-35	JCVI_SCAF_1101667114676	+	114	306	**RNA**→ ←? (GOS14155524)
env-36	JCVI_SCAF_1096627908392	-	1160	959	**RNA**→ ? (GOS6732)→
env-37	JCVI_SCAF_1101668469441	+	966	1168	**RNA**→
env-38	JCVI_SCAF_1096627918656	+	518	718	**RNA**→ ? (GOS19566)→ ? (GOS13866)→
env-39	JCVI_SCAF_1096627611218	+	1543	1738	**RNA**→ ? (GOS6732)→
env-40	JCVI_SCAF_1101668210533	+	1366	1555	**RNA**→ ? (GOS11583239)→
env-41	JCVI_SCAF_1101668718596	-	655	470	**RNA**→ Transaldolase_FSA (cd00956)? (GOS133)→
env-42	JCVI_SCAF_1101668713148	-	1022	835	**RNA**→ ? (GOS11357225)→
env-43	JCVI_SCAF_1096627048294	-	1122	923	**RNA**→ IbpA (COG0071)? (GOS1219)→ PurM_N (cd02196)? (GOS4853)→
env-44	JCVI_SCAF_1096627160072	+	462	661	**RNA**→ IbpA (COG0071)? (GOS1219)→ PurM_N (cd02196)? (GOS4853)→
env-45	JCVI_SCAF_1096627748636	+	1307	1497	**RNA**→
env-46	JCVI_SCAF_1096627307099	-	1677	1488	**RNA**→ Transaldolase_FSA (cd00956)? (GOS133)→
env-47	JCVI_SCAF_1101668289945	-	833	644	**RNA**→ PMM (pfam03332)? (GOS7317)→
env-48	JCVI_SCAF_1096626329237	+	610	800	**RNA**→
env-49	JCVI_SCAF_1101668294092	+	46	234	**RNA**→ PurS (COG1828)? (GOS13869)→ GcvT (COG0404)? (GOS1233)→
env-50	JCVI_SCAF_1101668296375	+	242	432	**RNA**→ IbpA (COG0071)? (GOS1219)→ PurM (COG0150)? (GOS4853)→
env-51	JCVI_SCAF_1101667215635	-	598	408	**RNA**→ PMM (pfam03332)? (GOS7317)→
env-52	JCVI_SCAF_1096627045643	-	598	408	**RNA**→ IbpA (COG0071)? (GOS1219)→
env-53	JCVI_SCAF_1096627054734	-	489	301	**RNA**→ ? (GOS133)→
env-54	JCVI_SCAF_1096627559863	+	482	670	**RNA**→ ? (GOS133)→
env-55	JCVI_SCAF_1096626980180	+	1523	1711	**RNA**→ Transaldolase_FSA (cd00956)? (GOS133)→
env-56	JCVI_SCAF_1101668211897	+	1319	1506	**RNA**→ ? (GOS11088388)→
env-57	JCVI_SCAF_1101667356596	+	84	271	**RNA**→ ←? (GOS14195600)
env-58	JCVI_SCAF_1101668222853	-	566	377	**RNA**→ Transaldolase_FSA (cd00956)? (GOS133)→
env-59	JCVI_SCAF_1096626939304	+	49	238	**RNA**→ ? (GOS7556)→ ? (GOS4491)→
env-60	JCVI_SCAF_1096627737247	-	610	421	**RNA**→ ? (GOS7556)→
env-61	JCVI_SCAF_1101668196475	-	349	160	**RNA**→ ? (GOS1274)→
env-62	JCVI_SCAF_1096627033404	+	1144	1334	**RNA**→ ? (GOS15843)→
env-63	JCVI_SCAF_1101668225031	+	377	567	**RNA**→ ? (GOS15843)→
env-64	JCVI_SCAF_1096627736474	-	579	389	**RNA**→ ? (GOS15843)→
env-65	JCVI_SCAF_1101668259949	-	1118	928	**RNA**→ ? (GOS15843)→
env-66	JCVI_SCAF_1096627732527	-	215	25	**RNA**→
env-67	JCVI_SCAF_1096627743775	+	1190	1379	**RNA**→ ? (GOS15843)→
env-68	JCVI_SCAF_1096627712495	+	1339	1529	**RNA**→ MannoseP_isomer (pfam01050)? (GOS1274)→
env-69	JCVI_SCAF_1096627047562	-	440	251	**RNA**→ ? (GOS7556)→ MannoseP_isomer (pfam01050)? (GOS1274)→
env-70	JCVI_SCAF_1096627706162	+	1568	1754	**RNA**→
Pph-2-1	NC_006884.1	+	165110	165295	**RNA**→ hypo→
env-71	JCVI_SCAF_1096627268413	+	736	921	**RNA**→ Transaldolase_FSA (cd00956)? (GOS133)→ ? (GOS15843)→ ? (GOS13042657)→ ? (GOS11950)→
env-72	JCVI_SCAF_1101668751778	+	315	500	**RNA**→
env-73	JCVI_SCAF_1096626957881	+	1314	1499	**RNA**→ ? (GOS15843)→ ? (GOS7556)→
env-74	JCVI_SCAF_1096626978678	-	902	718	**RNA**→ MannoseP_isomer (pfam01050)? (GOS1274)→ ? (GOS15843)→
env-75	JCVI_SCAF_1101667156622	-	245	59	**RNA**→
env-76	JCVI_SCAF_1096627631107	+	989	1175	**RNA**→ ? (GOS17289)→
env-77	JCVI_SCAF_1101667203659	+	460	647	**RNA**→ ? (GOS11430566)→
env-78	JCVI_SCAF_1096627592454	-	418	231	**RNA**→

env-79	JCVI_SCAF_1101668322298	-	317	130	**RNA**→ ? (GOS11430566)→
env-80	JCVI_SCAF_1101668231230	+	187	373	**RNA**→ IbpA (COG0071)? (GOS1219)→ PurM (COG0150)? (GOS4853)→
env-81	JCVI_SCAF_1096627056617	+	96	282	**RNA**→ Thy1 (pfam02511)? (GOS5954)→ ? (GOS13561)→ PhoH (pfam02562)? (GOS1288)→
env-82	JCVI_SCAF_1096627750439	+	1267	1454	**RNA**→ ? (GOS6732)→
env-83	JCVI_SCAF_1096627750132	+	1368	1554	**RNA**→
env-84	JCVI_SCAF_1101668230000	-	623	437	**RNA**→ ? (GOS6732)→
env-85	JCVI_SCAF_1101668308241	+	1006	1192	**RNA**→
env-86	JCVI_SCAF_1096627606460	+	1351	1537	**RNA**→ ? (GOS6732)→
env-87	JCVI_SCAF_1096627518205	-	243	57	**RNA**→
env-88	JCVI_SCAF_1096627546301	-	614	428	**RNA**→ ? (GOS6732)→
env-89	JCVI_SCAF_1096627620789	-	229	43	**RNA**→
env-90	JCVI_SCAF_1101667152283	-	758	572	**RNA**→ ? (GOS6732)→
env-91	JCVI_SCAF_1096626951480	+	541	727	**RNA**→ ? (GOS6732)→
env-92	JCVI_SCAF_1096627631984	-	465	278	**RNA**→ ? (GOS6732)→
env-93	JCVI_SCAF_1096626962396	+	1150	1337	**RNA**→ ? (GOS6732)→
env-94	JCVI_SCAF_1096626974384	-	924	737	**RNA**→ ? (GOS6732)→
env-95	JCVI_SCAF_1096626134187	+	8	194	**RNA**→ ? (GOS6732)→
env-96	JCVI_SCAF_1096626957409	+	2648	2835	**RNA**→ ? (GOS6732)→
env-97	JCVI_SCAF_1096627734913	-	341	154	**RNA**→ ? (GOS133)→
env-98	JCVI_SCAF_1096626967180	+	1201	1388	**RNA**→ ? (GOS6732)→
env-99	JCVI_SCAF_1101668352769	+	839	1026	**RNA**→
env-100	JCVI_SCAF_1096627706424	-	352	165	**RNA**→ ? (GOS133)→
env-101	JCVI_SCAF_1101668231541	+	588	775	**RNA**→ ? (GOS6732)→
env-102	JCVI_SCAF_1096627060928	+	1032	1221	**RNA**→
env-103	JCVI_SCAF_1101668300139	+	471	659	**RNA**→ PMM (pfam03332)? (GOS7317)→
env-104	JCVI_SCAF_1096627323972	+	3658	3846	**RNA**→
env-105	JCVI_SCAF_1101668321206	+	408	596	**RNA**→ IbpA (COG0071)? (GOS1219)→
env-106	JCVI_SCAF_1096627769377	-	1507	1319	**RNA**→ IbpA (COG0071)? (GOS1219)→ PurM (COG0150)? (GOS4853)→
env-107	JCVI_SCAF_1096627697478	+	1381	1569	**RNA**→
env-108	JCVI_SCAF_1096626966619	-	1167	979	**RNA**→ ? (GOS17289)→ Thy1 (pfam02511)? (GOS5954)→
env-109	JCVI_SCAF_1096627780060	+	278	466	**RNA**→ IbpA (COG0071)? (GOS1219)→
env-110	JCVI_SCAF_1096627324210	+	1868	2056	**RNA**→ ←? (GOS14195600)
env-111	JCVI_SCAF_1096627732798	-	245	56	**RNA**→
env-112	JCVI_SCAF_1096626936888	-	692	503	**RNA**→ ? (GOS6732)→
env-113	JCVI_SCAF_1096626955402	-	1932	1743	**RNA**→ IbpA (COG0071)? (GOS1219)→ PurM (COG0150)? (GOS4853)→
env-114	JCVI_SCAF_1096626975016	-	1814	1625	**RNA**→ IbpA (COG0071)? (GOS1219)→ PurM (COG0150)? (GOS4853)→
env-115	JCVI_SCAF_1096627625997	+	1299	1488	**RNA**→ ? (GOS13465662)→ ? (GOS1219)→
env-116	JCVI_SCAF_1096627026722	+	725	914	**RNA**→ IbpA (COG0071)? (GOS1219)→ PurM (COG0150)? (GOS4853)→
env-117	JCVI_SCAF_1096626976928	+	3010	3199	**RNA**→ PMM (pfam03332)? (GOS7317)→ IbpA (COG0071)? (GOS1219)→ PurM (COG0150)? (GOS4853)→
env-118	JCVI_SCAF_1096627057469	+	370	559	**RNA**→ IbpA (COG0071)? (GOS1219)→ PurM (COG0150)? (GOS4853)→
env-119	JCVI_SCAF_1096627054103	+	1130	1318	**RNA**→ MannoseP_isomer (pfam01050)? (GOS1274)→
env-120	JCVI_SCAF_1101668311091	-	924	736	**RNA**→ IbpA (COG0071)? (GOS1219)→
env-121	JCVI_SCAF_1096626984808	-	469	284	**RNA**→ ? (GOS133)→
env-122	JCVI_SCAF_1096627307461	+	953	1137	**RNA**→ ? (GOS11583239)→ ? (GOS6732)→
env-123	JCVI_SCAF_1101667124967	+	594	779	**RNA**→
env-124	JCVI_SCAF_1096627381047	+	2336	2521	**RNA**→ Transaldolase_FSA (cd00956)? (GOS133)→
env-125	JCVI_SCAF_1096627755306	+	466	650	**RNA**→ ? (GOS8297)→
env-126	JCVI_SCAF_1096627050529	-	623	439	**RNA**→ ? (GOS6732)→
env-127	JCVI_SCAF_1096626976647	+	1403	1587	**RNA**→ ? (GOS6732)→
env-128	JCVI_SCAF_1096626969882	+	674	858	**RNA**→ ? (GOS6732)→
env-129	JCVI_SCAF_1096627569486	-	365	181	**RNA**→ ? (GOS133)→
env-130	JCVI_SCAF_1096627733466	+	469	653	**RNA**→ ? (GOS133)→ ? (GOS11025)→

env-131	JCVI_SCAF_1096627590788	-	205	21	**RNA**→
env-132	JCVI_SCAF_1096627750656	-	762	575	**RNA**→ Transaldolase_FSA (cd00956)? (GOS133)→
env-133	JCVI_SCAF_1096627057290	+	731	918	**RNA**→ Transaldolase_FSA (cd00956)? (GOS133)→ ? (GOS11025)→
env-134	JCVI_SCAF_1101668231536	+	745	932	**RNA**→ Transaldolase_FSA (cd00956)? (GOS133)→
env-135	JCVI_SCAF_1096626971351	+	698	885	**RNA**→ Transaldolase_FSA (cd00956)? (GOS133)→
env-136	JCVI_SCAF_1101667105674	-	323	136	**RNA**→
env-137	JCVI_SCAF_1096627387858	-	1139	952	**RNA**→ ? (GOS6732)→
env-138	JCVI_SCAF_1096626957510	+	352	539	**RNA**→ ? (GOS6732)→
env-139	JCVI_SCAF_1101667125164	+	216	403	**RNA**→ Transaldolase_FSA (cd00956)? (GOS133)→
env-140	JCVI_SCAF_1101667071738	-	393	206	**RNA**→ ? (GOS11583239)→
env-141	JCVI_SCAF_1096627571067	-	627	440	**RNA**→ ? (GOS11430566)→ ? (GOS19514)→
env-142	JCVI_SCAF_1101667207126	-	699	512	**RNA**→ ? (GOS19514)→
env-143	JCVI_SCAF_1096627770179	-	1559	1372	**RNA**→ ? (GOS19514)→
env-144	JCVI_SCAF_1096627041192	-	2458	2271	**RNA**→ ? (GOS11430566)→ ? (GOS19514)→
env-145	JCVI_SCAF_1096626929735	-	1080	892	**RNA**→ ? (GOS11583239)→ IbpA (COG0071)? (GOS1219)→
env-146	JCVI_SCAF_1096627700131	+	116	301	**RNA**→ Transaldolase_FSA (cd00956)? (GOS133)→
env-147	JCVI_SCAF_1096626970126	-	3204	3019	**RNA**→ ? (GOS6732)→ ? (GOS11026)→
env-148	JCVI_SCAF_1096627056381	-	1978	1792	**RNA**→ ? (GOS11088388)→
env-149	JCVI_SCAF_1096627572274	+	1109	1306	**RNA**→ Transaldolase_FSA (cd00956)? (GOS133)→
Pph-1-1	NC_006883.1	+	144844	145042	**RNA**→ PurL→
env-150	JCVI_SCAF_1101668462221	-	903	703	**RNA**→ ? (GOS1457)→ ? (GOS1457)→
env-151	JCVI_SCAF_1096627117046	+	338	538	**RNA**→ Asn tRNA→ PurS (COG1828)? (GOS13869)→ IbpA (COG0071)? (GOS1219)→
env-152	JCVI_SCAF_1101668525944	-	893	704	**RNA**→
env-153	JCVI_SCAF_1096627921795	-	853	655	**RNA**→ ? (GOS6732)→
env-154	JCVI_SCAF_1096627734146	-	481	287	**RNA**→ ? (GOS18774)→
env-155	JCVI_SCAF_1101668715696	+	173	368	**RNA**→ Transaldolase_FSA (cd00956)? (GOS133)→ ? (GOS15843)→
env-156	JCVI_SCAF_1096627059528	+	264	456	**RNA**→ Transaldolase_FSA (cd00956)? (GOS133)→
env-157	JCVI_SCAF_1101668699012	+	656	851	**RNA**→ ? (GOS15843)→
env-158	JCVI_SCAF_1101668713317	-	705	511	**RNA**→
env-159	JCVI_SCAF_1101668182917	-	757	563	**RNA**→ ? (GOS1457)→ ? (GOS1457)→
env-160	JCVI_SCAF_1101668180709	-	1537	1343	**RNA**→ ? (GOS1457)→ ? (GOS1457)→
env-161	JCVI_SCAF_1101667227178	+	25	217	**RNA**→ ←? (GOS14071281)
env-162	JCVI_SCAF_1101667226279	-	572	380	**RNA**→ ←? (GOS14071281)
env-163	JCVI_SCAF_1096627318828	+	1140	1334	**RNA**→ ←? (GOS11349735)
env-164	JCVI_SCAF_1096627086450	-	1375	1181	**RNA**→ ? (GOS1457)→ ? (GOS1457)→
env-165	JCVI_SCAF_1096627319509	+	103	295	**RNA**→ ? (GOS11364937)→ ? (GOS1457)→ ? (GOS11034974)→ DNA_pol_B_exo (pfam03104)? (GOS2655)→ POLBc (smart00486)? (GOS2655)→
env-166	JCVI_SCAF_1096626948055	+	631	823	**RNA**→ ? (GOS11364937)→ ? (GOS1457)→ ? (GOS11034974)→ PolB (COG0417)? (GOS2655)→
env-167	JCVI_SCAF_1096626954078	+	1361	1550	**RNA**→ ? (GOS17590)→ ? (GOS13115571)→
env-168	JCVI_SCAF_1096627227614	+	2007	2197	**RNA**→
env-169	JCVI_SCAF_1096627374940	+	1435	1624	**RNA**→ ? (GOS8297)→
env-170	JCVI_SCAF_1101667215213	-	458	267	**RNA**→
env-171	JCVI_SCAF_1096627787443	+	1082	1276	**RNA**→ ? (GOS17289)→
env-172	JCVI_SCAF_1096627038945	-	1710	1521	**RNA**→ ? (GOS11069700)→ ABC_NikE_OppD_transporters (cd03257)? (GOS11069700)→
env-173	JCVI_SCAF_1096628005072	+	1204	1394	**RNA**→
env-174	JCVI_SCAF_1096626973123	+	776	970	**RNA**→ ? (GOS19514)→
env-175	JCVI_SCAF_1096627058092	-	1052	857	**RNA**→ ? (GOS19514)→
env-176	JCVI_SCAF_1101668193024	-	584	398	**RNA**→ ? (GOS6732)→
env-177	JCVI_SCAF_1096627570335	+	13	202	**RNA**→ MannoseP_isomer (pfam01050)? (GOS1274)→ ? (GOS11950)→
env-178	JCVI_SCAF_1096627708510	-	1218	1028	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-179	JCVI_SCAF_1096627310704	+	268	459	**RNA**→ ? (GOS13281280)→ Photo_RC (pfam00124)? (GOS5184)→
env-180	JCVI_SCAF_1101667224646	+	708	899	**RNA**→

env-181	JCVI_SCAF_1101667147392	-	510	318	**RNA**→
env-182	JCVI_SCAF_1096627742844	+	91	290	**RNA**→ ? (GOS18404)→ ? (GOS13716)→
env-183	JCVI_SCAF_1101667347489	+	167	366	**RNA**→ ? (GOS13716)→
env-184	JCVI_SCAF_1101668528260	-	684	484	**RNA**→ MannoseP_isomer (pfam01050)? (GOS1274)→
env-185	JCVI_SCAF_1101668337341	+	198	396	**RNA**→ Asn tRNA→ PurS (COG1828)? (GOS13869)→ Transaldolase_FSA (cd00956)? (GOS133)→ ? (GOS15843)→
env-186	JCVI_SCAF_1096627725200	-	763	570	**RNA**→
env-187	JCVI_SCAF_1096626402892	-	565	365	**RNA**→ Ala tRNA→
env-188	JCVI_SCAF_1101668352580	-	225	27	**RNA**→
env-189	JCVI_SCAF_1096627174089	+	1231	1429	**RNA**→ Ala tRNA→ ? (GOS12117925)→
env-190	JCVI_SCAF_1101668668425	-	1115	917	**RNA**→ Ala tRNA→ Asn tRNA→ ? (GOS11088388)→ ? (GOS2365)→
env-191	JCVI_SCAF_1096627743248	-	1361	1161	**RNA**→ ? (GOS4800)→

52.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 52.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *manA* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

cd00956 (19) Transaldolase-like fructose-6-phosphate aldolases (FSA) found in bacteria and archaea, which are member of the MipB/TalC subfamily of class I aldolases.
cd02196 (2) PurM (Aminoimidazole Ribonucleotide [AIR]
cd03257 (1) The ABC transporter subfamily specific for the transport of dipeptides, oligopeptides (OppD) and nickel (NikE).
COG0071 (17) Molecular chaperone (small heat shock protein) [Posttranslational modification, protein turnover, chaperones]
COG0150 (8) Phosphoribosylaminoimidazole (AIR) synthetase [Nucleotide transport and metabolism]
COG0404 (1) Glycine cleavage system T protein (aminomethyltransferase) [Amino acid transport and metabolism]
COG0417 (1) DNA polymerase elongation subunit (family B) [DNA replication, recombination, and repair]
COG1059 (1) Thermostable 8-oxoguanine DNA glycosylase [DNA replication, recombination, and repair]
COG1828 (3) Phosphoribosylformylglycinamidine (FGAM) synthase, PurS component [Nucleotide transport and metabolism]
COG3794 (1) Plastocyanin [Energy production and conversion]
GOS11025 (3) ?
GOS11026 (1) ?
GOS11034974 (2) ?
GOS11069700 (2) ?
GOS11088388 (3) ?
GOS11349735 (1) ?
GOS11357225 (1) ?
GOS11364937 (2) ?
GOS11384261 (1) ?
GOS11389660 (1) ?
GOS11430566 (4) ?
GOS11583239 (4) ?
GOS11589240 (1) ?
GOS11950 (2) ?
GOS12117925 (1) ?
GOS1219 (18) ?
GOS1233 (1) ?
GOS1274 (8) ?
GOS1288 (1) ?
GOS13042657 (1) ?
GOS13115571 (1) ?
GOS13281280 (1) ?
GOS133 (27) ?
GOS13465662 (1) ?
GOS13561 (1) ?
GOS13716 (2) ?
GOS13866 (1) ?
GOS13869 (3) ?
GOS14071281 (2) ?
GOS14155524 (3) ?
GOS14195600 (2) ?
GOS1457 (10) ?
GOS1582 (1) ?
GOS15843 (13) ?
GOS17089 (1) ?
GOS17289 (3) ?
GOS17590 (1) ?
GOS17718 (1) ?
GOS18404 (1) ?
GOS18444 (3) ?
GOS18774 (1) ?
GOS19514 (6) ?

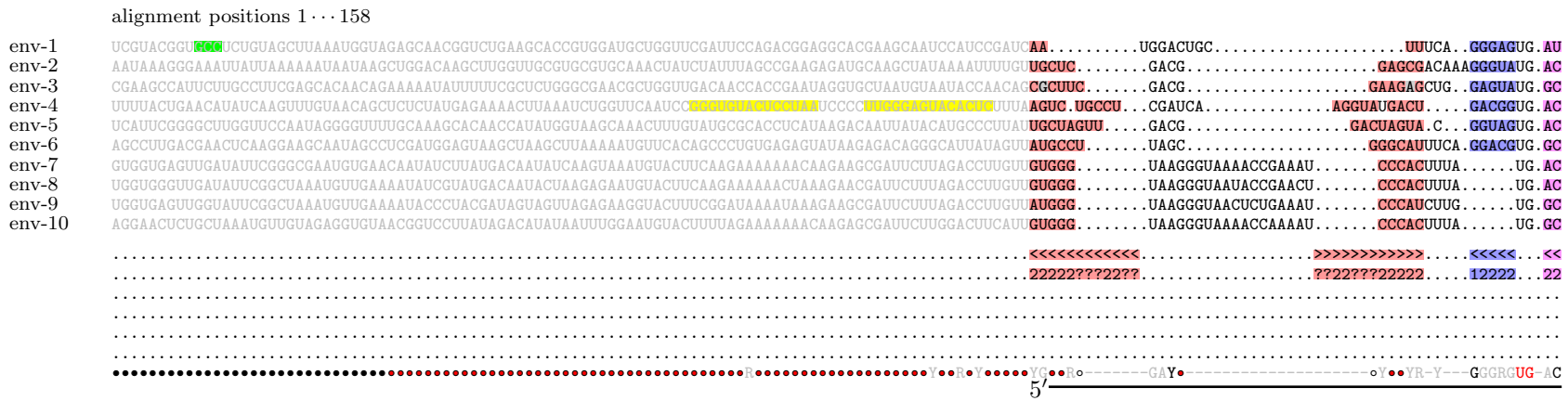
GOS19566 (1) ?
GOS2365 (1) ?
GOS2655 (3) ?
GOS4491 (1) ?
GOS4800 (1) ?
GOS4853 (10) ?
GOS5184 (2) ?
GOS5954 (2) ?
GOS6732 (27) ?
GOS7317 (4) ?
GOS7381 (1) ?
GOS7556 (4) ?

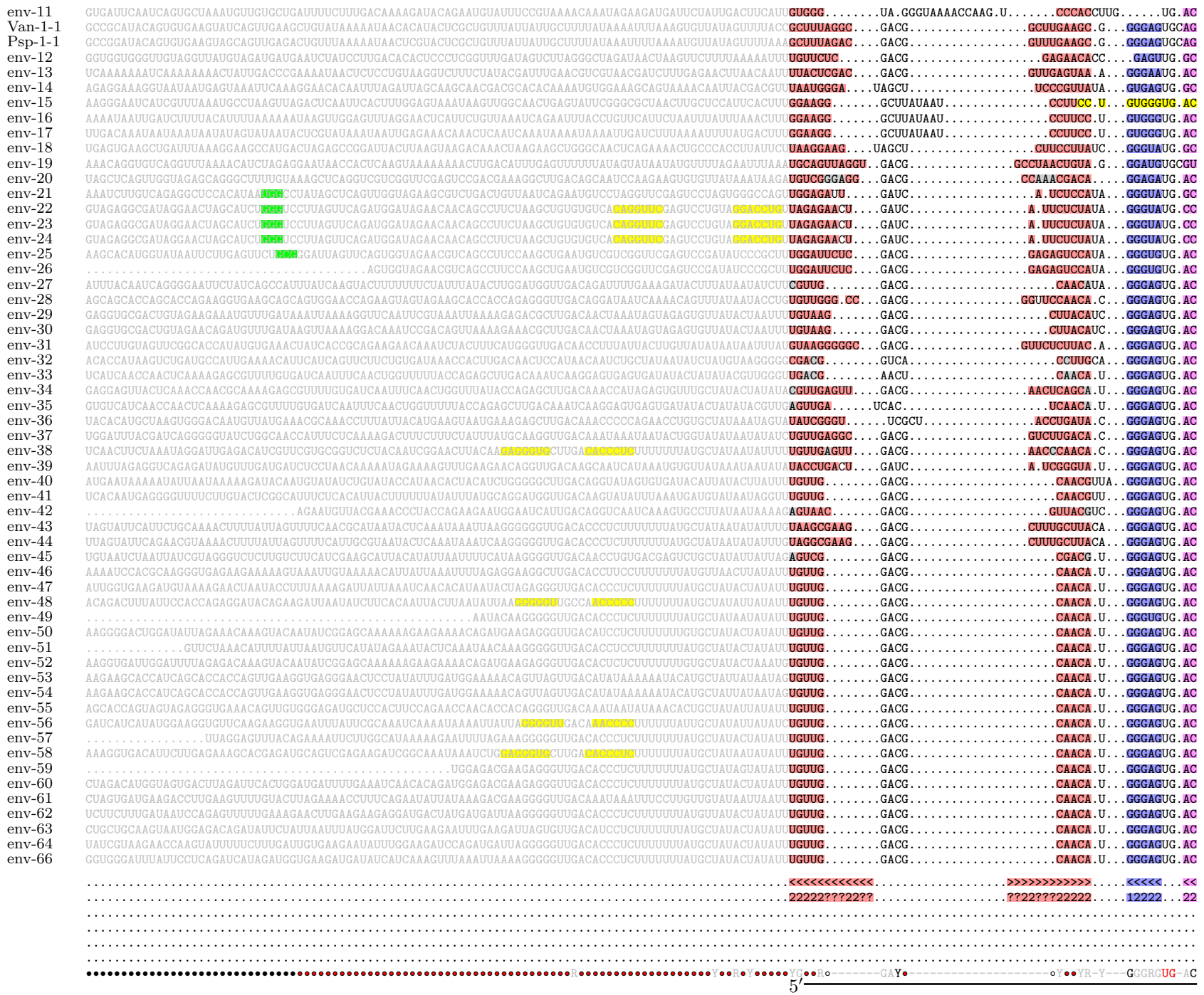
GOS8297 (2) ?
GOS9362 (2) ?
pfam00124 (2) Photosynthetic reaction centre protein.
pfam00127 (1) Copper binding proteins, plastocyanin/azurin family.
pfam01050 (7) Mannose-6-phosphate isomerase.
pfam02511 (2) Thymidylate synthase complementing protein.
pfam02562 (1) PhoH-like protein.
pfam03104 (1) DNA polymerase family B, exonuclease domain.
pfam03332 (4) Eukaryotic phosphomannomutase.
smart00486 (1) DNA polymerase type-B family; DNA polymerase alpha, delta, epsilon and zeta chain (eukaryota), DNA polymerases in archaea, DNA polymerase II in e.

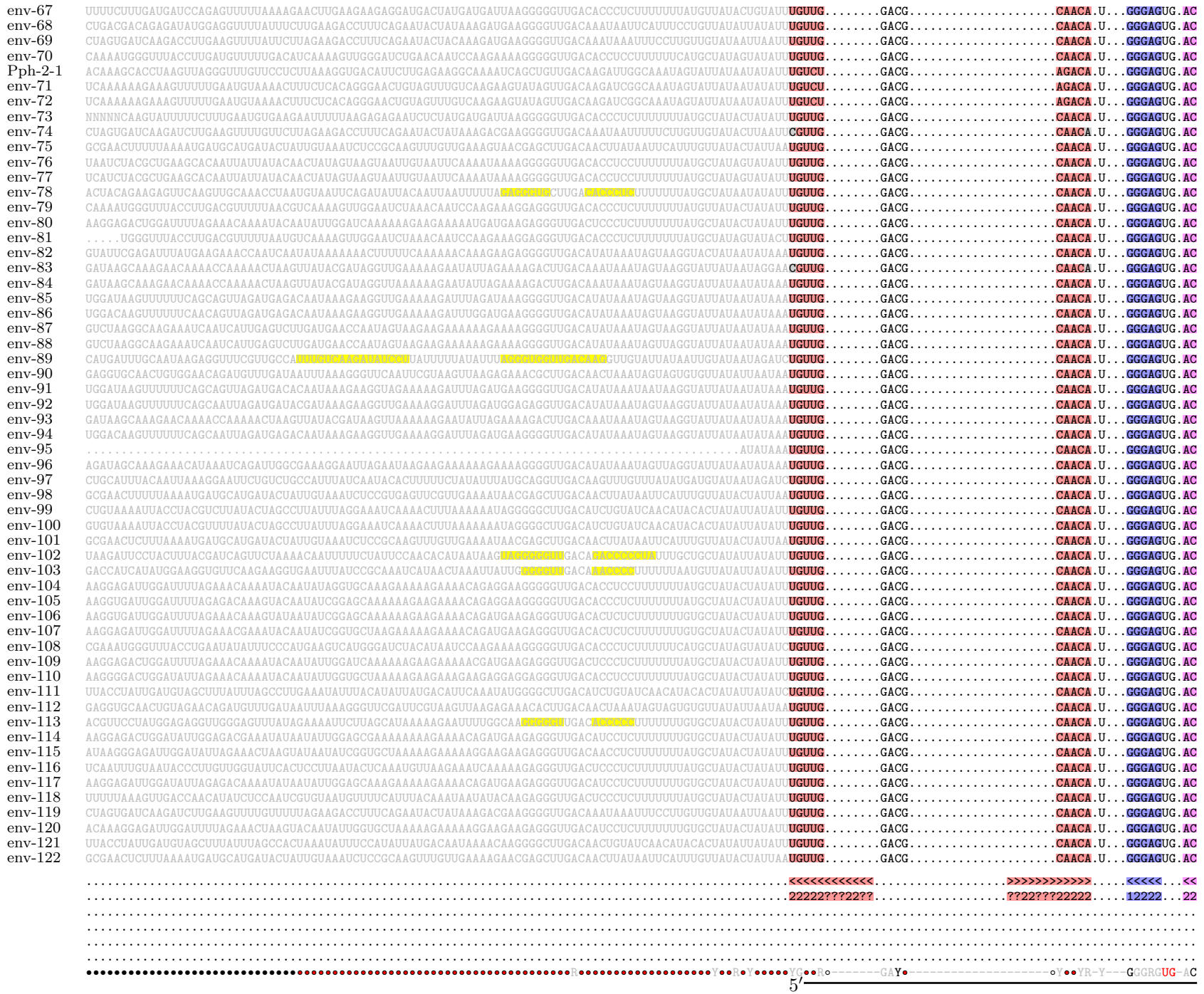
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: env-135, env-65

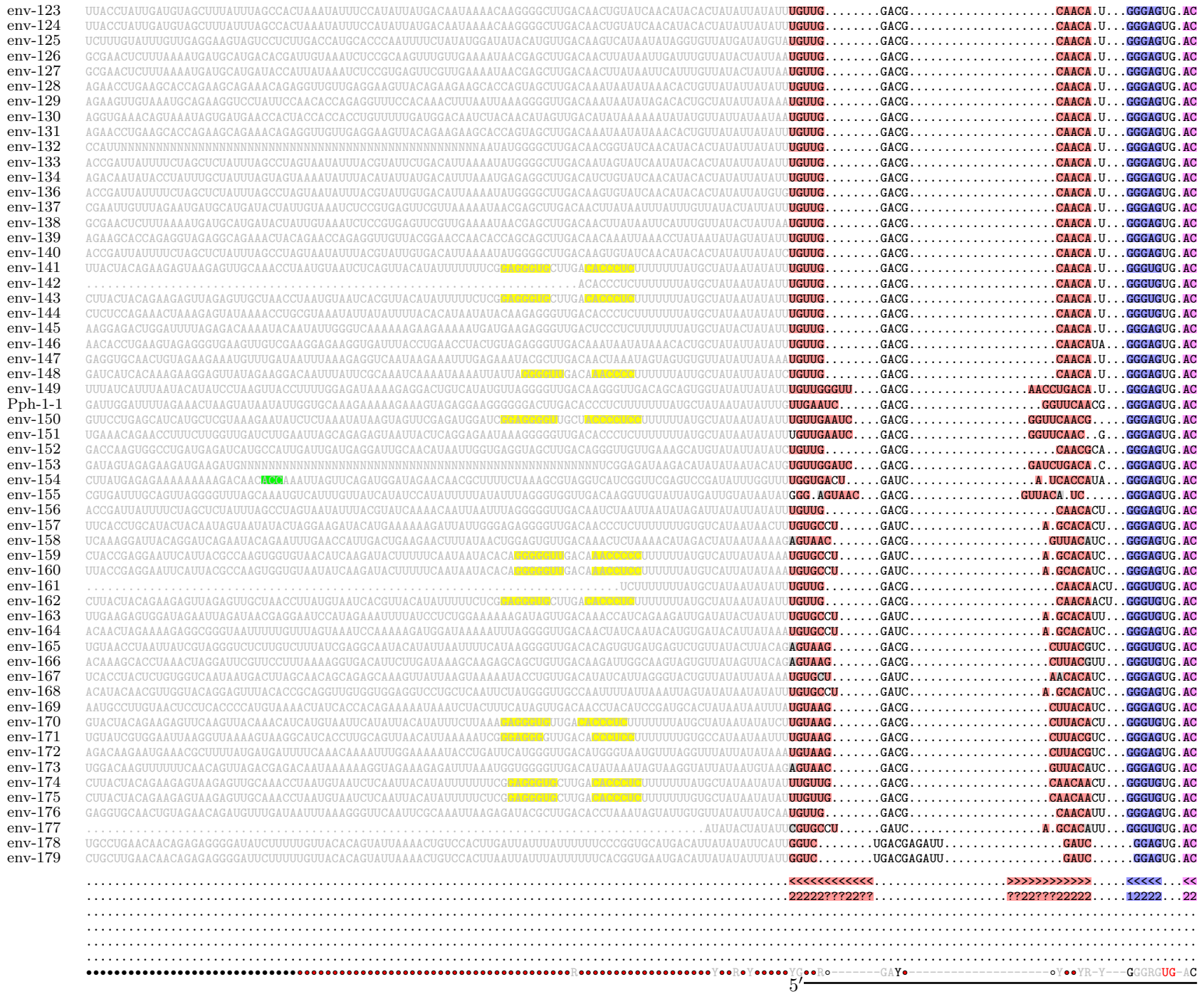
52.4 Multiple-sequence alignment

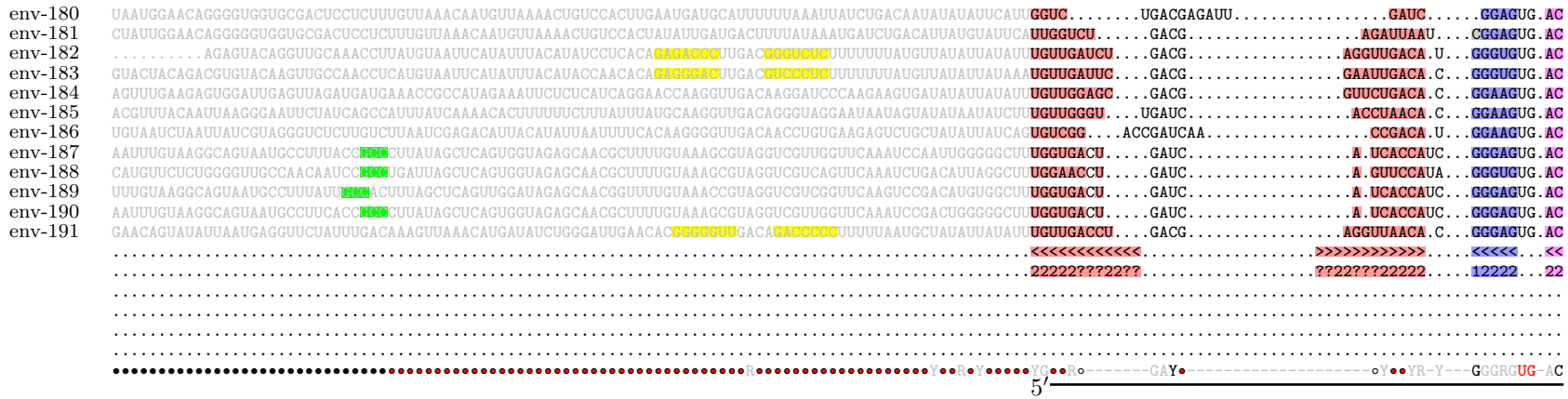
Each *manA* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 52.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *manA* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.




env-11
Van-1-1
Psp-1-1
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-66
5′


env-67
Pph-2-1
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
5′


env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-136
env-137
env-138
env-139
env-140
env-141
env-142
env-143
env-144
env-145
env-146
env-147
env-148
env-149
Pph-1-1
env-150
env-151
env-152
env-153
env-154
env-155
env-156
env-157
env-158
env-159
env-160
env-161
env-162
env-163
env-164
env-165
env-166
env-167
env-168
env-169
env-170
env-171
env-172
env-173
env-174
env-175
env-176
env-177
env-178
env-179
5′

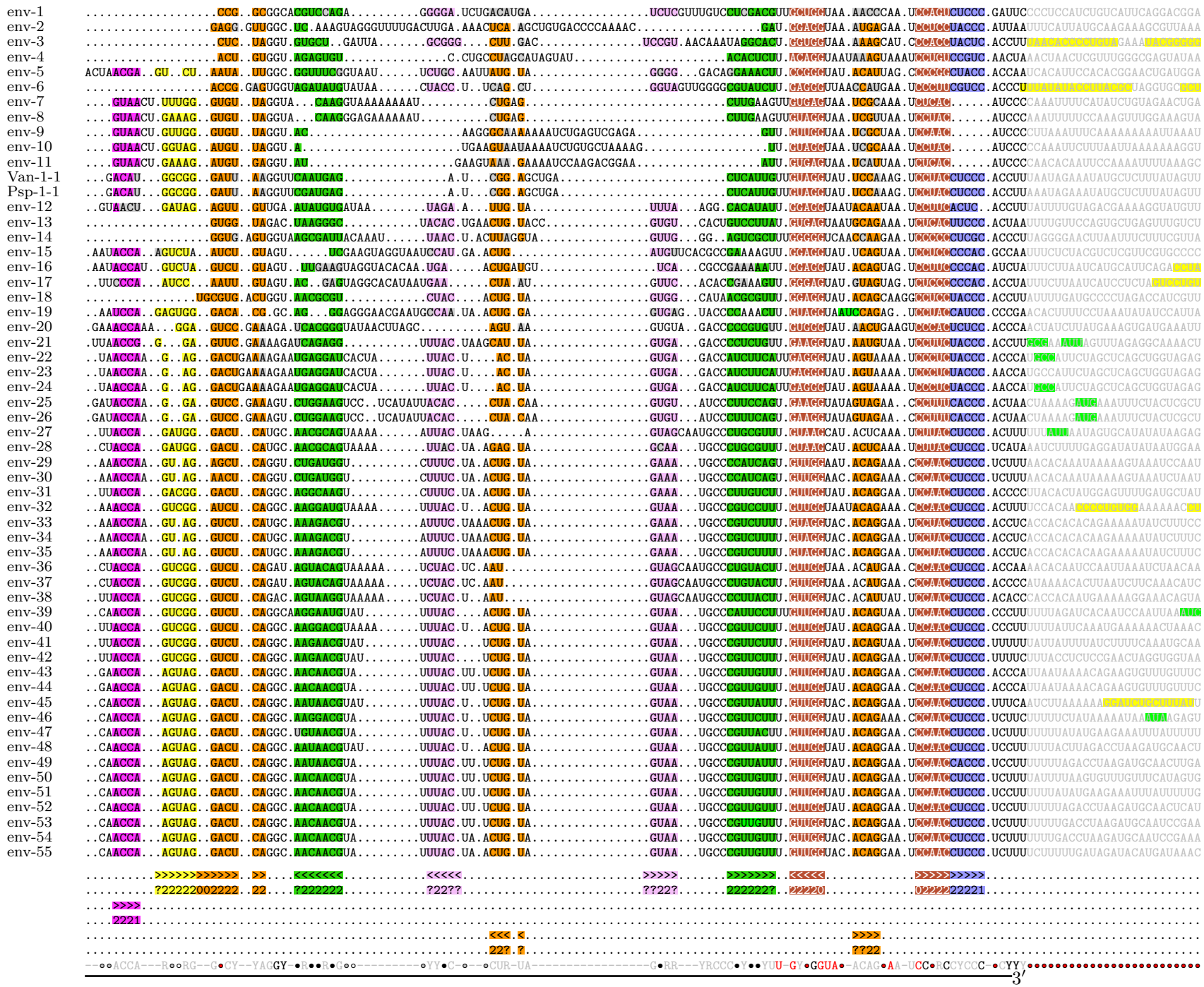


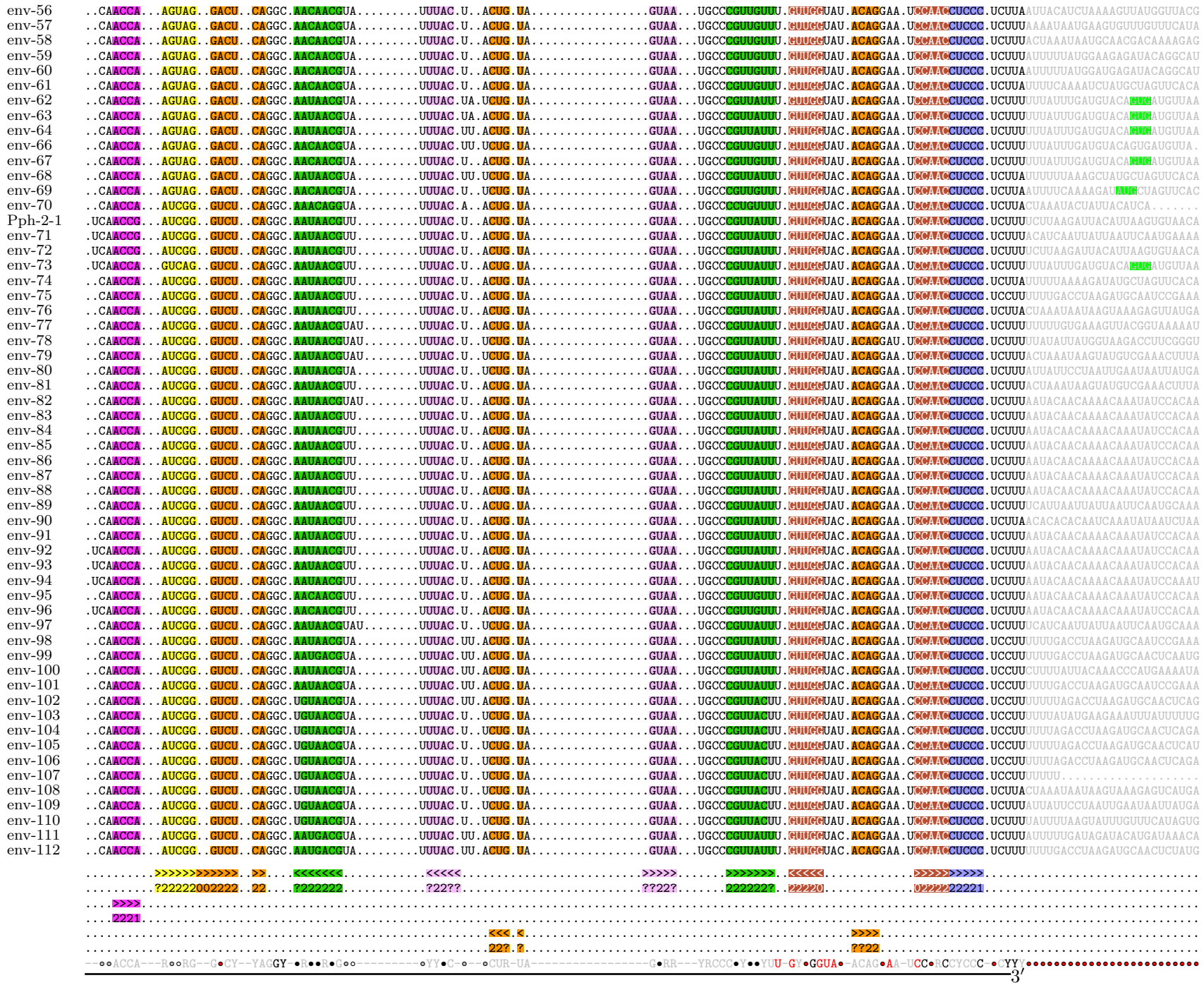
alignment positions 159 · · · 315

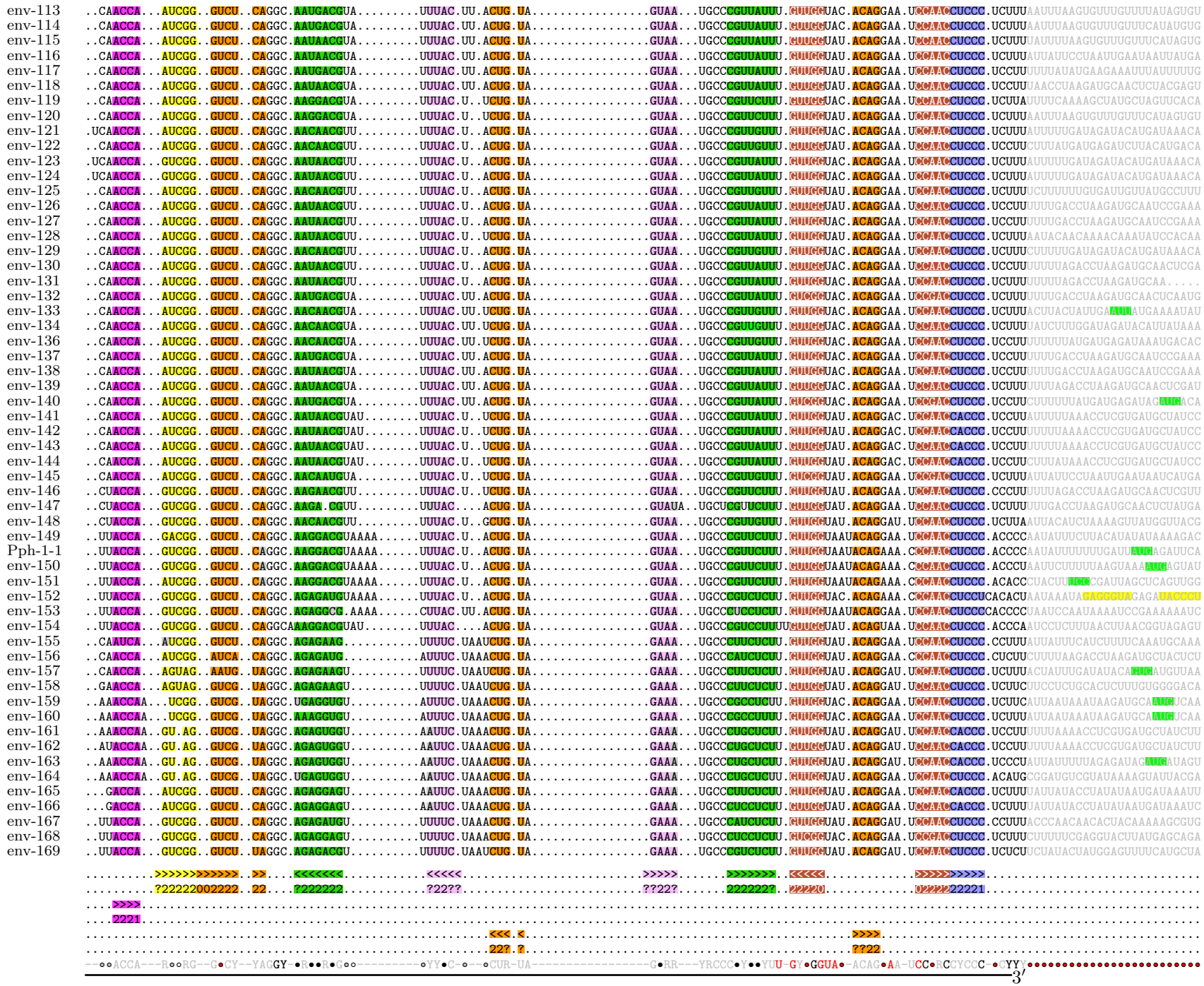
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-66
env-67
env-68
env-69
env-70
Pph-2-1
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88

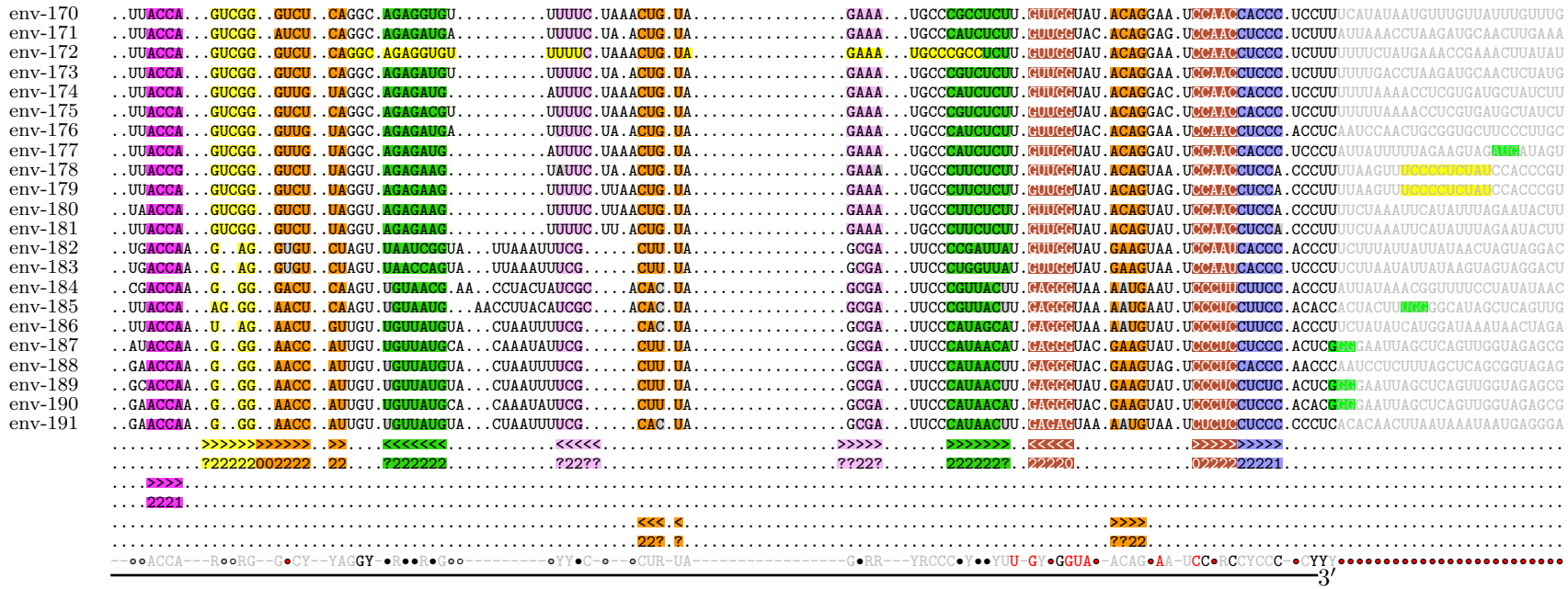
env-89 UGAAUAAA.CUU.ACUGGC..A.AC.U.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCA.....GCGA.....UGCAG.AACCGAU.
env-90 UGAAUAAA.CUU.ACUGGC..A.AC.U.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCA.....GCGA.....UGCAG.AACCGAU.
env-91 UGAAUAAA.CUU.ACUGGC..A.AC.U.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCA.....GCGA.....UGCAG.AACCGAU.
env-92 UGAAUAAA.CUU.ACUGGC..A.AC.U.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCA.....GCGA.....UGCAG.AACCGAU.
env-93 UGAAUAAA.CUU.ACUGGC..A.AC.U.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCA.....GCGA.....UGCAG.AACCGAU.
env-94 UGAAUAAA.CUU.ACUGGC..A.AC.U.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCA.....GCGA.....UGCAG.AACCGAU.
env-95 UGAAUAAA.CUU.ACUGGC..A.AC.U.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCA.....GCGA.....UGCAG.AACCGAU.
env-96 UGAAUAAA.CUU.ACUGGC..A.AC.U.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCA.....GCGA.....UGCAG.AACCGAU.
env-97 UGAAUAAA.CUU.ACUGGC..A.AC.U.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCA.....GCGA.....UGCAG.AACCGAU.
env-98 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCA.....GCGA.....UGUAG.AACCGAU.
env-99 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCA.....GCGA.....UGCAG.AACCGAU.
env-100 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCA.....GCGA.....UGCAG.AACCGAU.
env-101 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCA.....GCGA.....UGCAG.AACCGAU.
env-102 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCUC.....GCAA.....GGGUAG.AACCGAU.
env-103 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCUC.....GCAA.....GGGUAG.AACCGAU.
env-104 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCUC.....GCAA.....GGGUAG.AACCGAU.
env-105 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCUC.....GCAA.....GGGUAG.AACCGAU.
env-106 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCUC.....GCAA.....GGGUAG.AACCGAU.
env-107 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCUC.....GCAA.....GGGUAG.AACCGAU.
env-108 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCUC.....GCAA.....GGGUAG.AACCGAU.
env-109 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCUC.....GCAA.....GGGUAG.AACCGAU.
env-110 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCUC.....GCAA.....GGGUAG.AACCGAU.
env-111 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCUC.....GCAA.....GGGUAG.AACCGAU.
env-112 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCUC.....GCAA.....GGGUAG.AACCGAU.
env-113 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCUC.....GCAA.....GGGUAG.AACCGAU.
env-114 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCUC.....GCAA.....GGGUAG.AACCGAU.
env-115 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCUC.....GCAA.....GGGUAG.AACCGAU.
env-116 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCUC.....GCAA.....GGGUAG.AACCGAU.
env-117 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCUC.....GCAA.....GGGUAG.AACCGAU.
env-118 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCUC.....GCAA.....GGGUAG.AACCGAU.
env-119 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCUC.....GCAA.....GGGUAG.AACCGAU.
env-120 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCUC.....GCAA.....GGGUAG.AACCGAU.
env-121 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUAC.....GAAA.....GUAG.AACCGAU.
env-122 UGAAUAAA.CUU.ACUGGC..A.AC.U.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUAC.....GAAA.....GUAG.AACCGAU.
env-123 UGAAUAAA.CUU.ACUGGC..A.AC.U.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUUC.....GAGA.....GAAG.AACCGAC.
env-124 UGAAUAAA.CUU.ACUGGC..A.AC.U.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUUC.....GAGA.....GAAG.AACCGAC.
env-125 UGAAUAAA.CUU.ACUGGC..A.AC.U.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUCC.....GAAA.....GGAG.AACCGAU.
env-126 UGAAUAAA.CUU.ACUGGC..A.AC.U.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUUC.....GAGA.....GAAG.AACCGAU.
env-127 UGAAUAAA.CUU.ACUGGC..A.AC.U.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUAC.....GAAA.....GUAG.AACCGAU.
env-128 UGAAUAAA.CUU.ACUGGC..A.AC.U.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUUC.....GAAA.....GAAG.AACCGAU.
env-129 UGAAUAAA.CUU.ACUGGC..A.AC.U.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUUC.....GAAA.....GAAG.AACCGAU.
env-130 UGAAUAAA.CUU.ACUGGC..A.AC.U.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUUC.....GAAA.....GAAG.AACCGAU.
env-131 UGAAUAAA.CUU.ACUGGC..A.AC.U.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUUC.....GAAA.....GAAG.AACCGAU.
env-132 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUACC.....GAGA.....GGUAG.AACCGAU.
env-133 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUACC.....GAGA.....GGUAG.AACCGAU.
env-134 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUACC.....GAGA.....GGUAG.AACCGAU.
env-136 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUACC.....GAGA.....GGUAG.AACCGAU.
env-137 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUAUC.....GAGA.....GGUAG.AACCGAU.
env-138 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUAUC.....GAGA.....GGUAG.AACCGAU.
env-139 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUACC.....GAGA.....GGUAG.AACCGAU.
env-140 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUACC.....GCAG.....GGUAG.AACCGAU.
env-141 UGAACAAA.CUU.ACUGGC..A.AC.C.....GCUAGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCA.....GCGA.....UGCAG.AACCGAU.
env-142 UGAACAAA.CUU.ACUGGC..A.AC.C.....GCUAGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCA.....GCGA.....UGCAG.AACCGAU.
env-143 UGAACAAA.CUU.ACUGGC..A.AC.C.....GCUAGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCA.....GCGA.....UGCAG.AACCGAU.
env-144 UGAACAAA.CUU.ACUGGC..A.AC.C.....GCUAGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCA.....GCGA.....UGCAG.AACCGAU.
env-145 UGAAUAAA.CUU.ACUGGC..A.AC.C.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUGCUC.....GCAA.....GGGUAG.AACCGAU.
env-146 UGAAUAAA.CUU.ACUGGC..A.AC.U.....GCUGGUUAAGGU.G.AUG.AGAC..A.....CAG.GUGGUGCUG.CUAC.....GAAA.....GUAG.AACCGAC.

env-147
env-148
env-149
Pph-1-1
env-150
env-151
env-152
env-153
env-154
env-155
env-156
env-157
env-158
env-159
env-160
env-161
env-162
env-163
env-164
env-165
env-166
env-167
env-168
env-169
env-170
env-171
env-172
env-173
env-174
env-175
env-176
env-177
env-178
env-179
env-180
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
env-190
env-191
alignment positions 316 · · · 475
3′

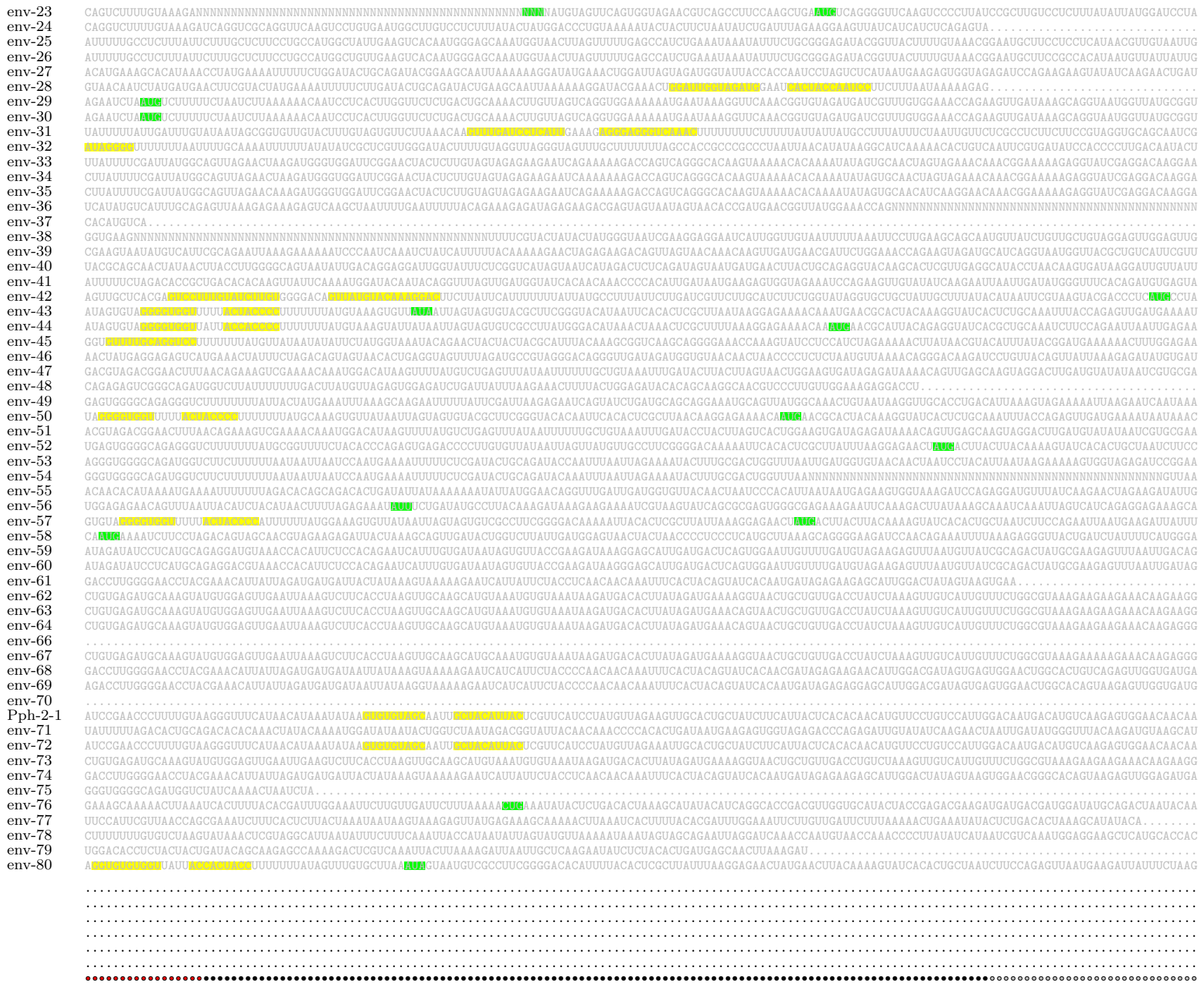

env-1
env-2
env-3
env-4
env-5
env-6
env-7
env-8
env-9
env-10
env-11
Van-1-1
Psp-1-1
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
3′

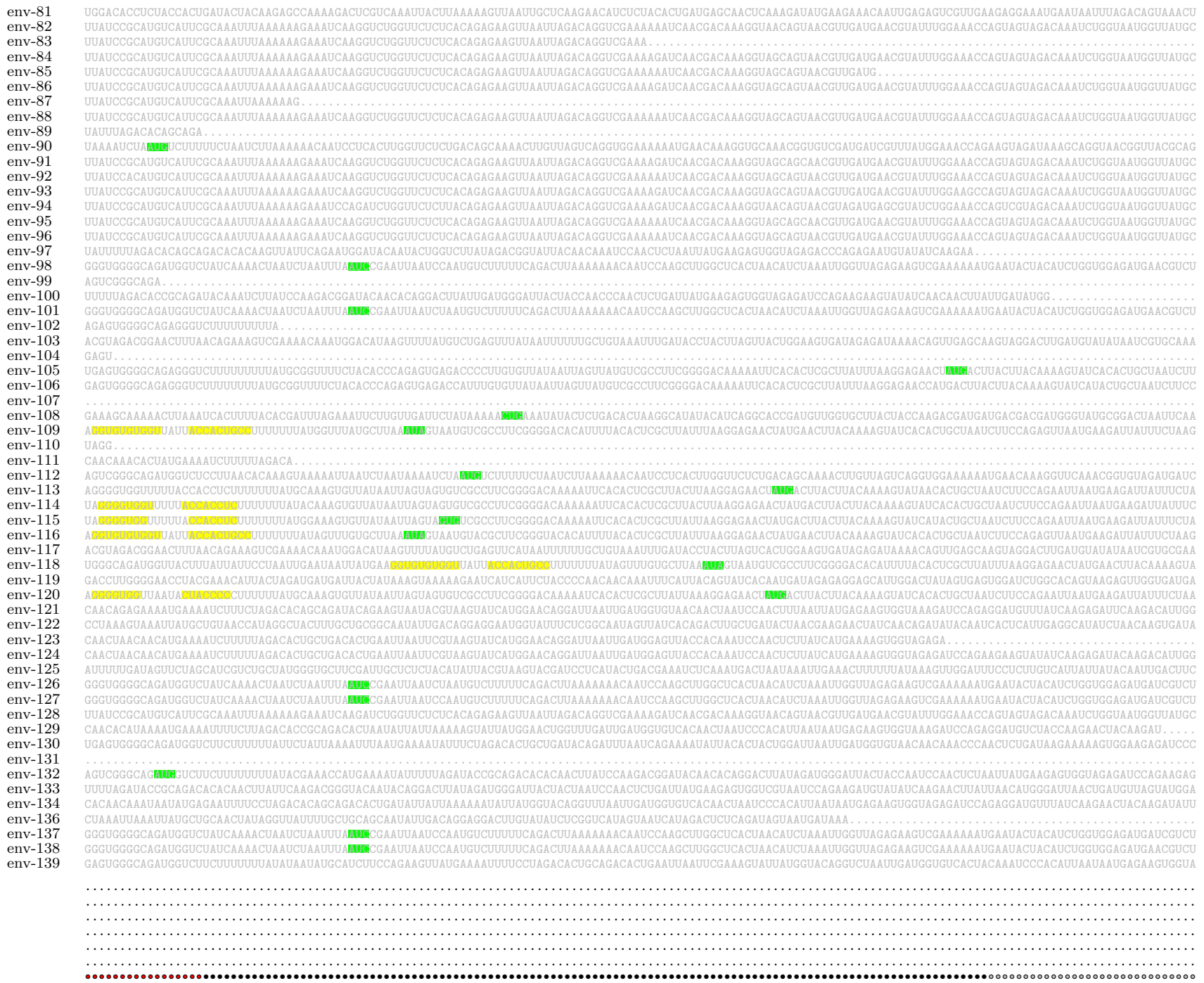

env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-66
env-67
env-68
env-69
env-70
Pph-2-1
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
3′


env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-136
env-137
env-138
env-139
env-140
env-141
env-142
env-143
env-144
env-145
env-146
env-147
env-148
env-149
Pph-1-1
env-150
env-151
env-152
env-153
env-154
env-155
env-156
env-157
env-158
env-159
env-160
env-161
env-162
env-163
env-164
env-165
env-166
env-167
env-168
env-169
3′

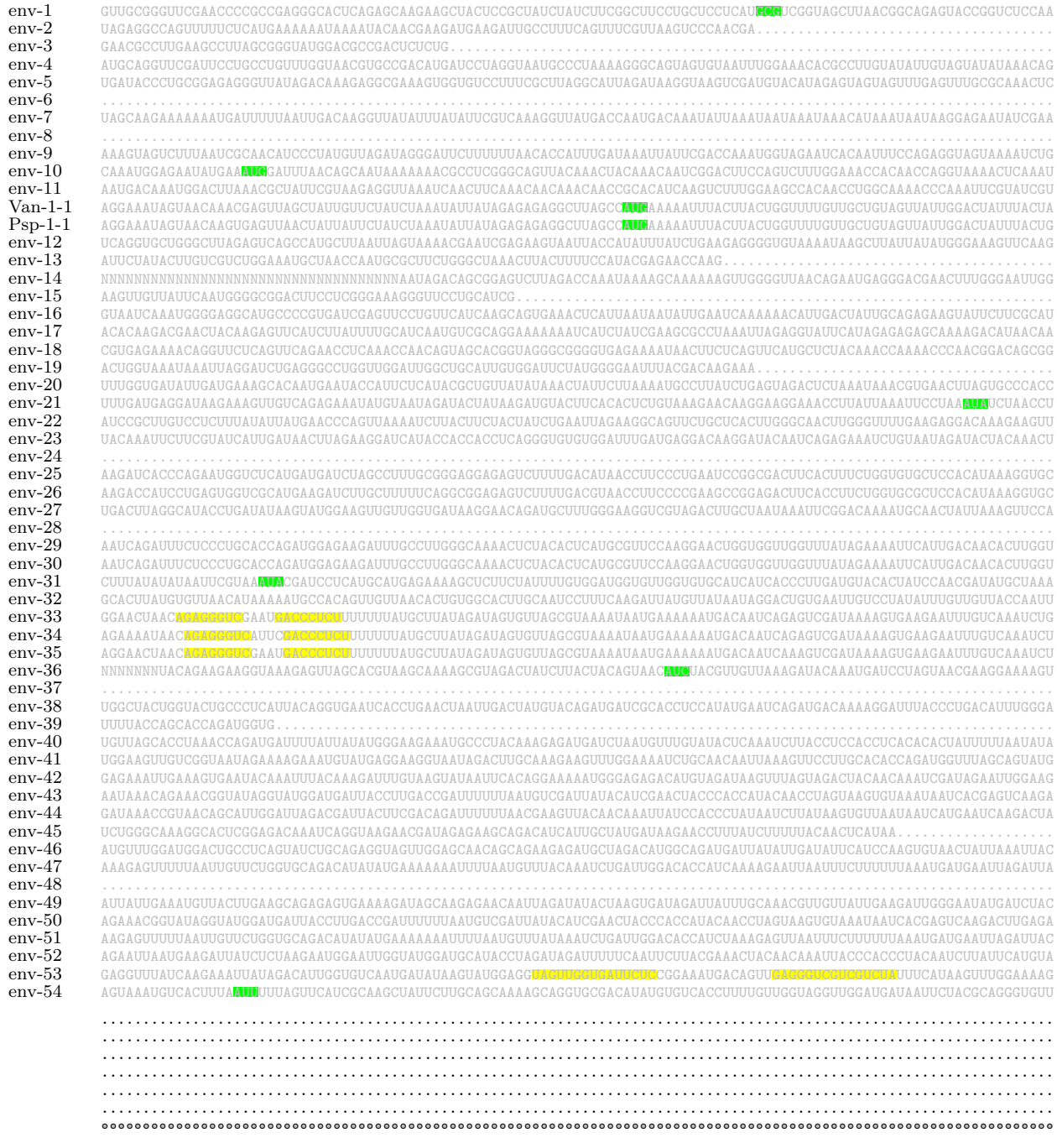


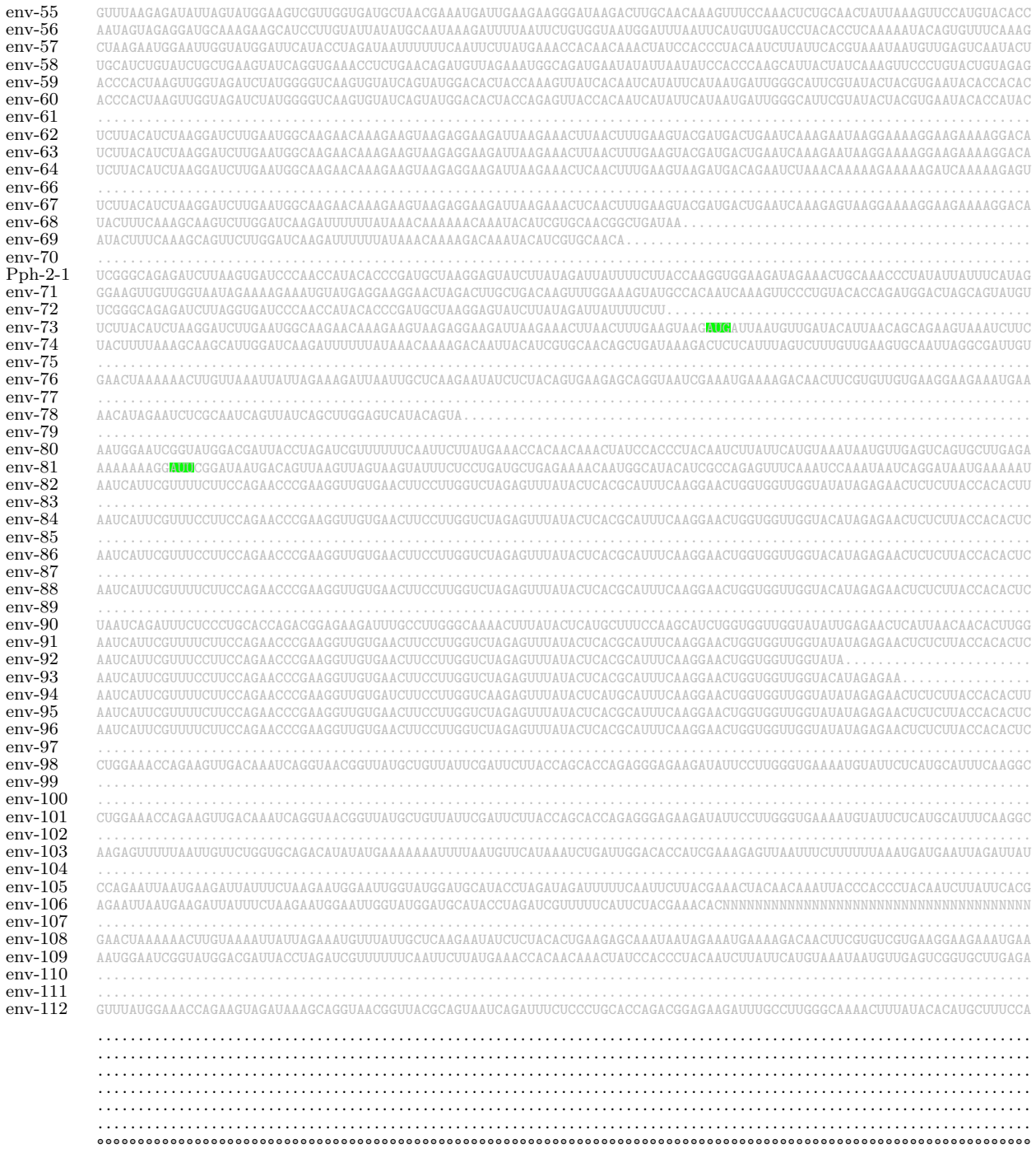
alignment positions 476 · · · 635

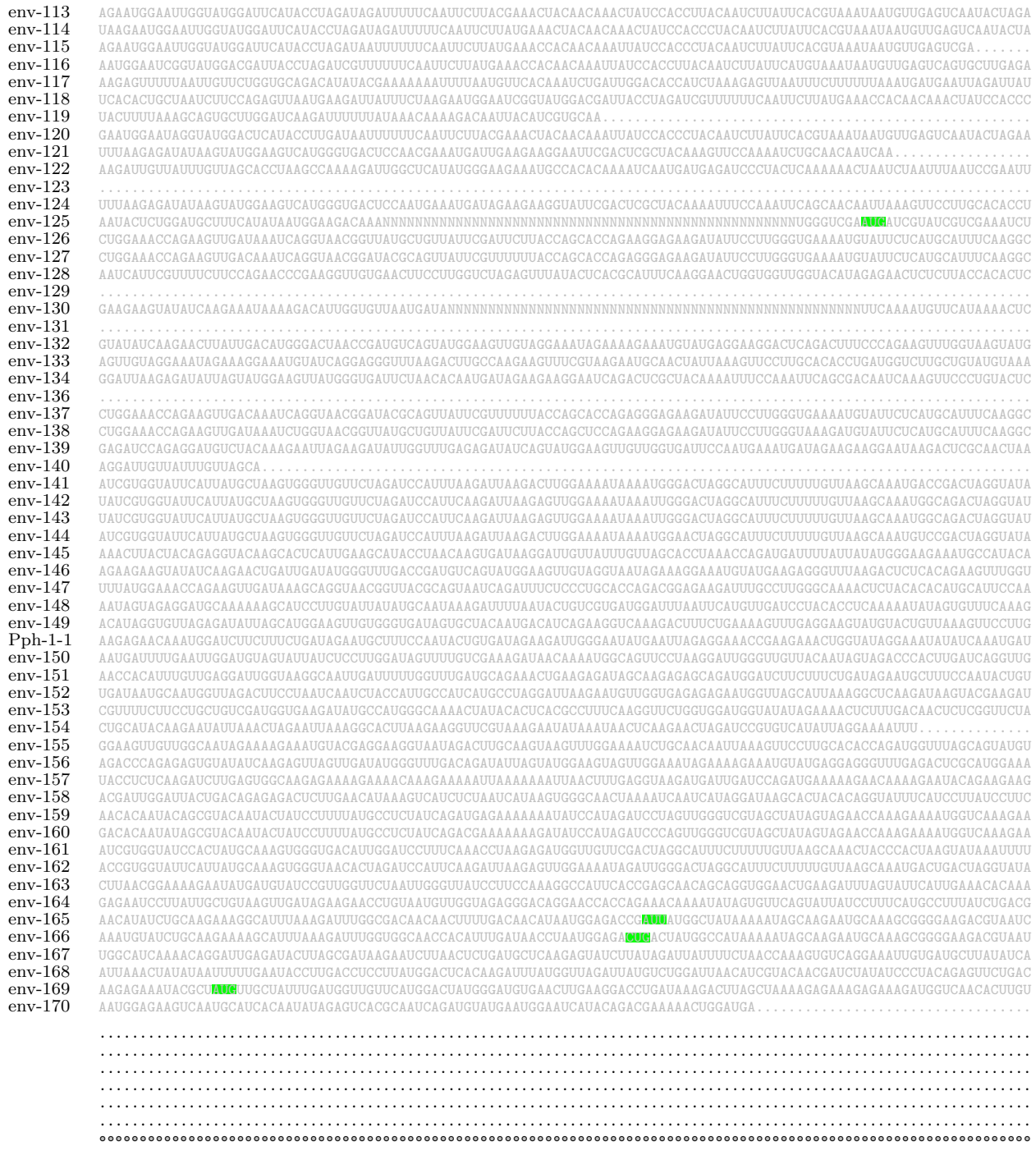

env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-66
env-67
env-68
env-69
env-70
Pph-2-1
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80

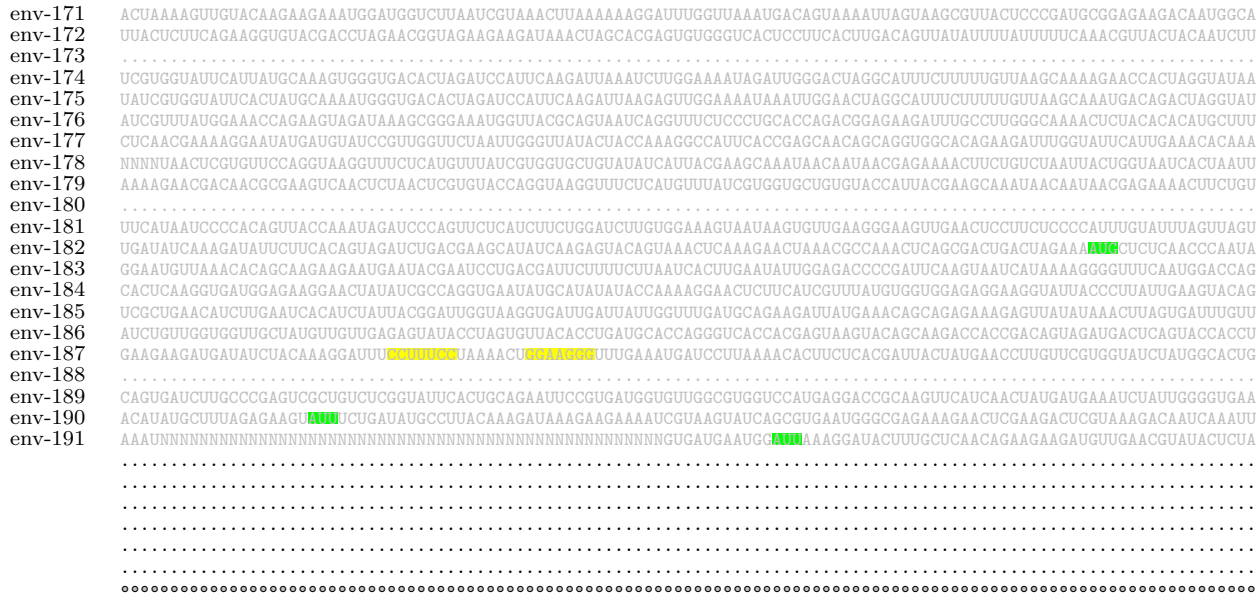

env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-136
env-137
env-138
env-139

env-140
env-141
env-142
env-143
env-144
env-145
env-146
env-147
env-148
env-149
Pph-1-1
env-150
env-151
env-152
env-153
env-154
env-155
env-156
env-157
env-158
env-159
env-160
env-161
env-162
env-163
env-164
env-165
env-166
env-167
env-168
env-169
env-170
env-171
env-172
env-173
env-174
env-175
env-176
env-177
env-178
env-179
env-180
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
env-190
env-191


alignment positions 636 · · · 750
env-1
env-2
env-3
env-4
env-5
env-6
env-7
env-8
env-9
env-10
env-11
Van-1-1
Psp-1-1
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54


env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-66
env-67
env-68
env-69
env-70
Pph-2-1
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112


env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-136
env-137
env-138
env-139
env-140
env-141
env-142
env-143
env-144
env-145
env-146
env-147
env-148
env-149
Pph-1-1
env-150
env-151
env-152
env-153
env-154
env-155
env-156
env-157
env-158
env-159
env-160
env-161
env-162
env-163
env-164
env-165
env-166
env-167
env-168
env-169
env-170



53 *Methylobacterium*-1

53.1 Taxa

The taxonomy of each organism containing a putative *Methylobacterium*-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *Methylobacterium*-1 RNA in Sections 53.2 and 53.4.

abbrev. of hits	taxonomy of species
Msp-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Methylobacteriaceae *Methylobacterium sp.* 4-46
env-1 to env-35	environmental samples

53.2 Gene contexts

Each *Methylobacterium*-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *Methylobacterium*-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 53.3. The accession number of the sequence containing each *Methylobacterium*-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 53.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Msp-1-1	NZ_ABAY01000058.1	-	29133	29042	COG3378 (COG3378)→ **RNA**→ ←TPR (cd00189)CyaA (COG2114)COG5616 (COG5616)
env-1	JCVI_SCAF_1101668649939	-	681	568	? (GOS11095947)→ **RNA**→
env-2	JCVI_SCAF_1096627284418	-	1712	1609	? (GOS11203544)→ **RNA**→ ? (GOS13499732)→
env-3	JCVI_SCAF_1101667319971	+	112	215	**RNA**→ ←? (GOS1843)
env-4	JCVI_SCAF_1101668071277	+	1147	1250	PinR (COG1961)? (GOS4893)→ **RNA**→ ? (GOS13499732)→
env-5	JCVI_SCAF_1096626880745	+	1055	1158	←? (GOS8700) **RNA**→ ? (GOS13499732)→
env-6	JCVI_SCAF_1101667037823	+	230	333	**RNA**→ Transposase_8 (pfam01527)? (GOS11798)→
env-7	JCVI_SCAF_1101668649760	-	527	424	**RNA**→ ? (GOS11693766)→
env-8	JCVI_SCAF_1096626028711	+	371	471	**RNA**→ ? (GOS11853573)→
env-9	JCVI_SCAF_1101668212413	+	1217	1321	? (GOS13055994)→ **RNA**→
env-10	JCVI_SCAF_1101668019257	-	672	568	**RNA**→
env-11	JCVI_SCAF_1101668418363	+	235	337	**RNA**→ ? (GOS13447208)→
env-12	JCVI_SCAF_1096626984948	+	563	664	? (GOS11392543)→ **RNA**→ SSL2 (COG1061)Mrr_cat (pfam04471)? (GOS3636)→
env-13	JCVI_SCAF_1101668036528	-	664	563	? (GOS11050863)→ **RNA**→ ←? (GOS11016987)
env-14	JCVI_SCAF_1101667597257	+	1	99	**RNA**→ ? (GOS13239929)→
env-15	JCVI_SCAF_1096628290018	+	669	770	**RNA**→ ←MT-A70 (pfam05063)? (GOS9797)
env-16	JCVI_SCAF_1096628012523	-	1038	937	←? (GOS12042) **RNA**→
env-17	JCVI_SCAF_1096628085782	-	642	541	**RNA**→ ? (GOS13126317)→
env-18	JCVI_SCAF_1096626921006	-	1561	1458	**RNA**→ ←? (GOS12403)
env-19	ABEF01010507.1	-	763	658	**RNA**→ hypo→
env-20	JCVI_SCAF_1096628164139	+	1143	1246	? (GOS11511305)→ **RNA**→ ? (GOS11511305)→
env-21	JCVI_SCAF_1101667431649	+	507	607	? (GOS11511305)→ **RNA**→ ? (GOS11511305)→
env-22	ABEF01000929.1	-	382	280	**RNA**→
env-23	ABEF01013431.1	+	35	137	**RNA**→ ←IstB (pfam01695)
env-24	ABEF01039662.1	+	104	207	**RNA**→ hypo→
env-25	JCVI_SCAF_1096627424305	+	720	823	←Ser tRNA **RNA**→ ? (GOS13128678)→
env-26	JCVI_SCAF_1096628046848	-	1489	1386	**RNA**→ ? (GOS13126317)→
env-27	JCVI_SCAF_1096628005677	+	62	163	**RNA**→ ? (GOS11674492)→
env-28	JCVI_SCAF_1096627976719	-	1070	969	←? (GOS1694) **RNA**→ Dcm (COG0270)? (GOS606)→
env-29	JCVI_SCAF_1101668090775	-	196	95	←Arg tRNA **RNA**→
env-30	JCVI_SCAF_1096626256100	+	719	820	**RNA**→
env-31	JCVI_SCAF_1101668326084	+	758	860	**RNA**→ ? (GOS13210440)→
env-32	JCVI_SCAF_1101668094001	-	775	673	INT_Rci (cd00796)? (GOS1694)→ **RNA**→ ? (GOS13210440)→
env-33	JCVI_SCAF_1096627997289	-	677	574	←Pseudo tRNA **RNA**→
env-34	JCVI_SCAF_1096628371309	+	176	283	**RNA**→
env-35	JCVI_SCAF_1096628236590	-	1318	1214	←? (GOS7137) **RNA**→

53.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 53.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *Methylobacterium*-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

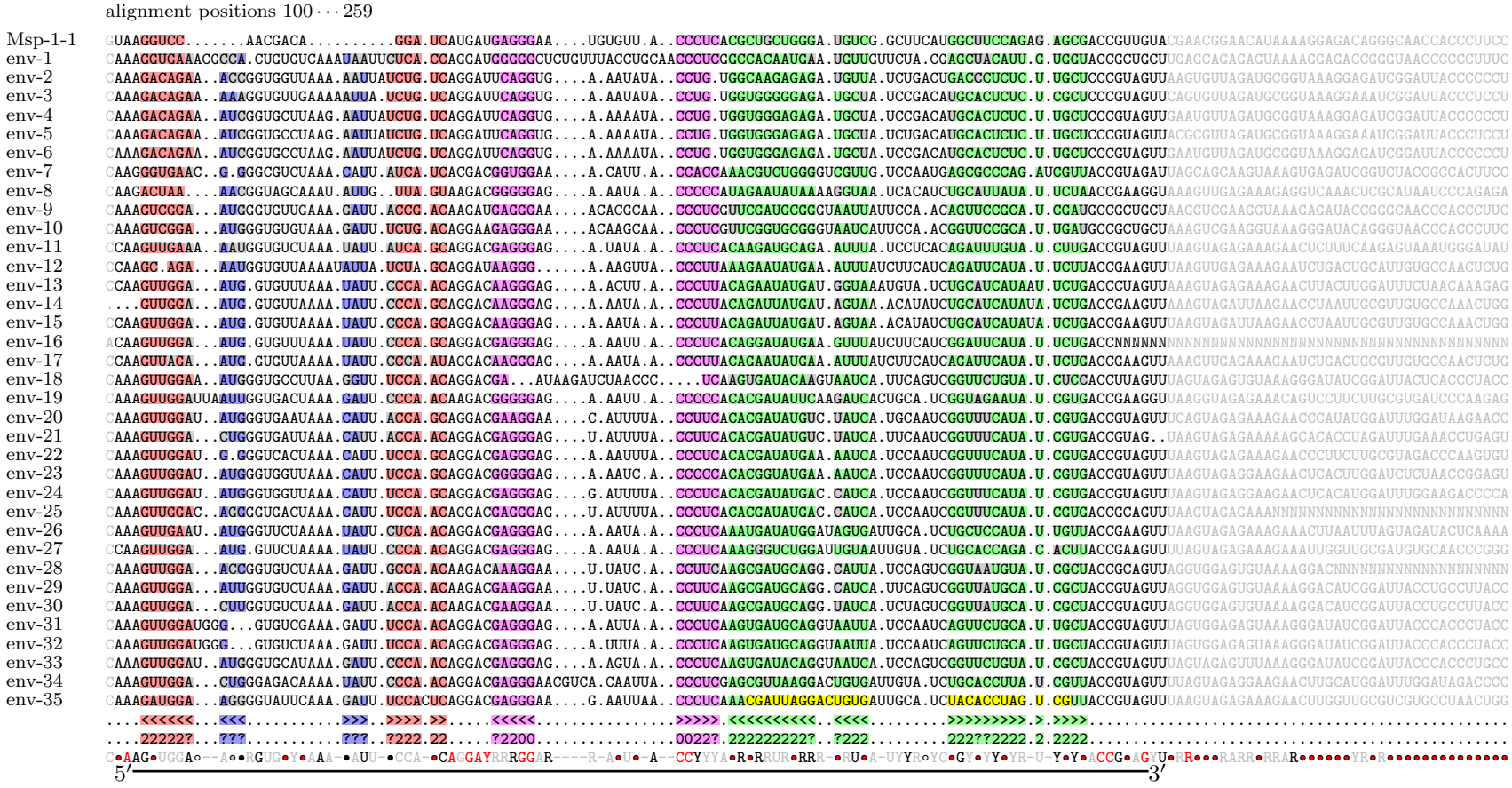
d cd00189 (1) Tetratricopeptide repeat domain; typically contains 34 amino acids [WLF]-X(2)-[LIM]-[GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-X(2)-[ASL]-X(4)-[PKE]
u cd00796 (1) Rci recombinase, C-terminal catalytic domain.
D COG0270 (1) Site-specific DNA methylase [DNA replication, recombination, and repair]
D COG1061 (1) DNA or RNA helicases of superfamily II [Transcription / DNA replication, recombination, and repair]
u COG1961 (1) Site-specific recombinases, DNA invertase Pin homologs [DNA replication, recombination, and repair]
d COG2114 (1) Adenylate cyclase, family 3 (some proteins contain HAMP domain) [Signal transduction mechanisms]

u COG3378 (1) Predicted ATPase [General function prediction only]
d COG5616 (1) Predicted integral membrane protein [Function unknown]
d GOS11016987 (1) ?
u GOS11050863 (1) ?
u GOS11095947 (1) ?
u GOS11203544 (1) ?
u GOS11392543 (1) ?
uD GOS11511305 (4) ?
D GOS11674492 (1) ?
D GOS11693766 (1) ?
D GOS11798 (1) ?
D GOS11853573 (1) ?
u GOS12042 (1) ?
d GOS12403 (1) ?
u GOS13055994 (1) ?
D GOS13126317 (2) ?
D GOS13128678 (1) ?
D GOS13210440 (2) ?
D GOS13239929 (1) ?
D GOS13447208 (1) ?
D GOS13499732 (3) ?
u GOS1694 (2) ?
d GOS1843 (1) ?
D GOS3636 (1) ?
u GOS4893 (1) ?
D GOS606 (1) ?
u GOS7137 (1) ?
u GOS8700 (1) ?
d GOS9797 (1) ?
D pfam01527 (1) Transposase.
d pfam01695 (1) IstB-like ATP binding protein.
D pfam04471 (1) Restriction endonuclease.
d pfam05063 (1) MT-A70.

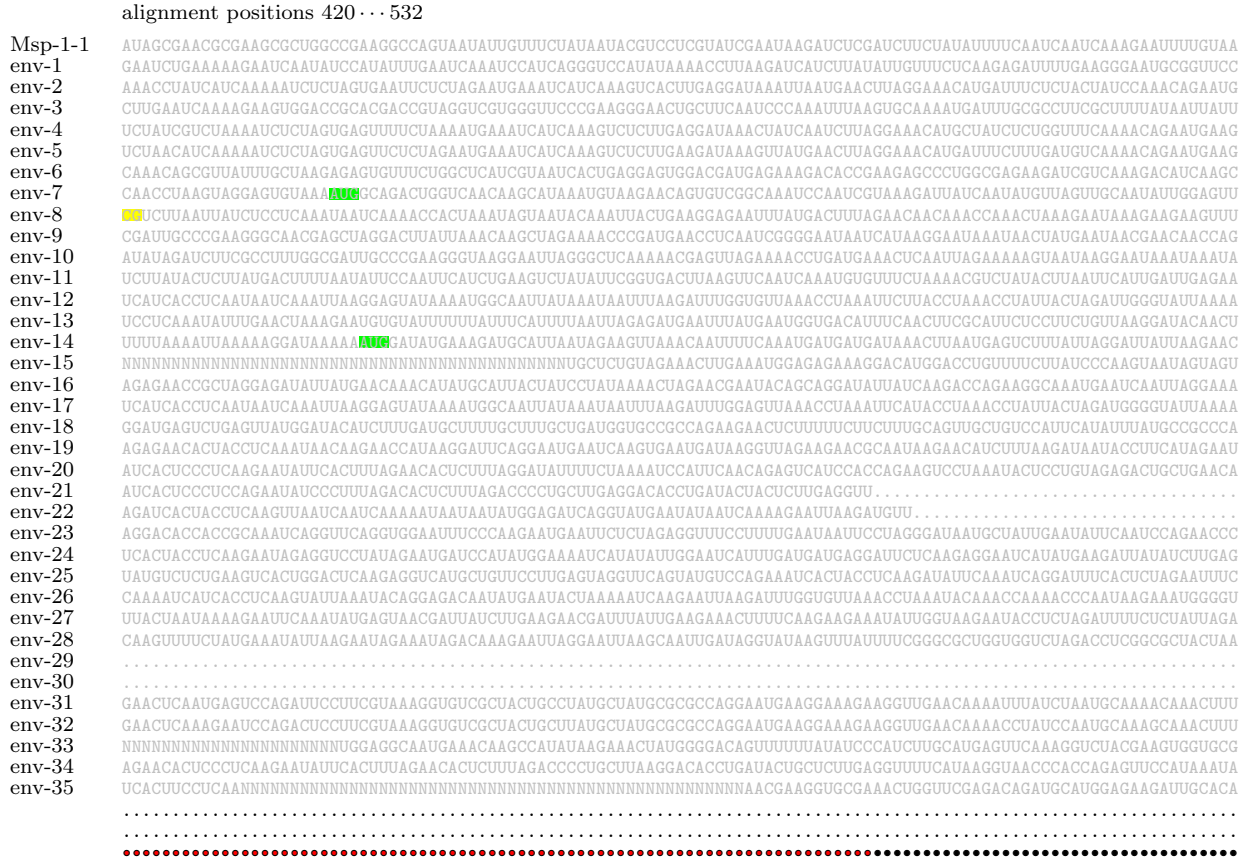
53.4 Multiple-sequence alignment

Each *Methylobacterium*-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 53.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *Methylobacterium*-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

alignment positions 1 · · · 99
Msp-1-1
env-1
env-2
env-3
env-4
env-5
env-6
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35


alignment positions 100 · · · 259
Msp-1-1
env-1
env-2
env-3
env-4
env-5
env-6
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
5′
3′

alignment positions 260 · · · 419



54 Moco-II

54.1 Taxa

The taxonomy of each organism containing a putative Moco-II RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Moco-II RNA in Sections 54.2 and 54.4.

abbrev. of hits	**taxonomy of species**
Tde-1-1	Bacteria Proteobacteria Betaproteobacteria Hydrogenophilales Hydrogenophilaceae *Thiobacillus denitrificans* ATCC 25259
Gur-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Geobacteraceae *Geobacter uraniireducens* Rf4
Pca-1-1 to Pca-1-4	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Pelobacteraceae *Pelobacter carbinolicus* DSM 2380
Ppr-1-1 to Ppr-1-2	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Pelobacteraceae *Pelobacter propionicus* DSM 2379

54.2 Gene contexts

Each Moco-II RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Moco-II RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 54.3. The accession number of the sequence containing each Moco-II RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 54.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Tde-1-1	NC_007404.1	-	2557036	2556993	**RNA**→ Nitrate_red_gam (pfam02665)→ GlpC (COG0247)→ GltD (COG0493)COG1144 (COG1144)→ hypo→ HybA (COG0437)→ NrfD (pfam03916)→ CobB (COG1797)→ IscA (COG0316)→ CysG (COG0007)CysG (COG1648)→
Pca-1-1	NC_007498.2	+	280705	280753	**RNA**→ MoeA (COG0303)→ CoxS (COG2080)Ald_Xan_dh_C (pfam01315)Ald_Xan_dh_C2 (pfam02738)→
Pca-1-2	NC_007498.2	+	189530	189575	**RNA**→ Pcm (COG2518)→ ligand_gated_channel (cd01347)→
Pca-1-3	NC_007498.2	+	186813	186856	**RNA**→ ←hypo
Gur-1-1	NC_009483.1	+	1183100	1183148	**RNA**→ MoeA (COG0303)→
Ppr-1-1	NC_008609.1	-	1659585	1659542	**RNA**→ MoeA (COG0303)→
Ppr-1-2	NC_008609.1	-	1639928	1639883	**RNA**→ DsrC (pfam04358)→ NarI (COG2181)→ GlpC (COG0247)→ hypo→
Pca-1-4	NC_007498.2	+	278681	278728	**RNA**→ MTH1175 (cd00851)→

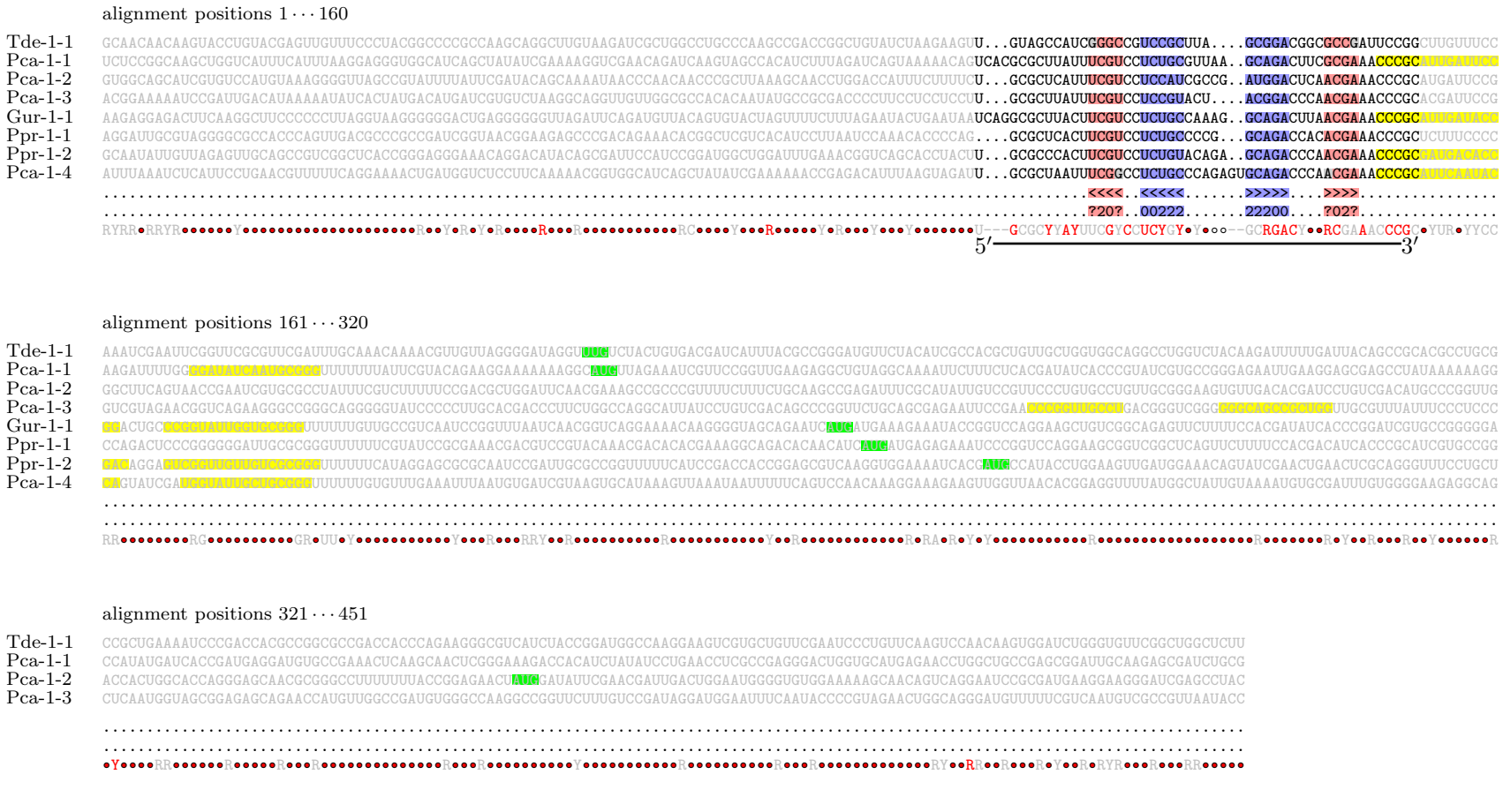
54.3 Conserved domains

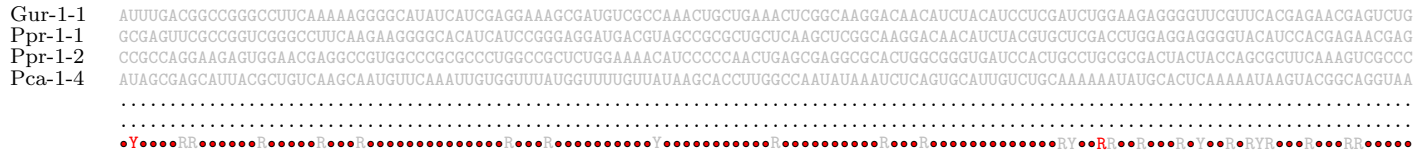
Conserved domains found in protein-coding genes listed in Section 54.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Moco-II RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

cd00851 (1) This uncharacterized conserved protein belongs to a family of iron-molybdenum cluster-binding proteins that includes NifX, NifB, and NifY, all of which are involved in the synthesis of an iron-molybdenum cofactor (FeMo-co) that binds the active site of the dinitrogenase enzyme.
cd01347 (1) TonB dependent/Ligand-Gated channels are created by a monomeric 22 strand (22,24) anti-parallel beta-barrel.
COG0007 (1) Uroporphyrinogen-III methylase [Coenzyme metabolism]
COG0247 (2) Fe-S oxidoreductase [Energy production and conversion]
COG0303 (3) Molybdopterin biosynthesis enzyme [Coenzyme metabolism]
COG0316 (1) Uncharacterized conserved protein [Function unknown]
COG0437 (1) Fe-S-cluster-containing hydrogenase components 1 [Energy production and conversion]
COG0493 (1) NADPH-dependent glutamate synthase beta chain and related oxidoreductases [Amino acid transport and metabolism / General function prediction only]
COG1144 (1) Pyruvate:ferredoxin oxidoreductase and related 2-oxoacid:ferredoxin oxidoreductases, delta subunit [Energy production and conversion]
COG1648 (1) Siroheme synthase (precorrin-2 oxidase/ferrochelatase domain) [Coenzyme metabolism]
COG1797 (1) Cobyrinic acid a,c-diamide synthase [Coenzyme metabolism]
COG2080 (1) Aerobic-type carbon monoxide dehydrogenase, small subunit CoxS/CutS homologs [Energy production and conversion]
COG2181 (1) Nitrate reductase gamma subunit [Energy production and conversion]
COG2518 (1) Protein-L-isoaspartate carboxylmethyltransferase [Posttranslational modification, protein turnover, chaperones]
pfam01315 (1) Aldehyde oxidase and xanthine dehydrogenase, a/b hammerhead domain.
pfam02665 (1) Nitrate reductase gamma subunit.
pfam02738 (1) Aldehyde oxidase and xanthine dehydrogenase, molybdopterin binding domain.
pfam03916 (1) Polysulphide reductase, NrfD.
pfam04358 (1) DsrC like protein.

54.4 Multiple-sequence alignment

Each Moco-II RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 54.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Moco-II RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (◦): 75%, white circle (◦): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.





55 *mraW*

55.1 Taxa

The taxonomy of each organism containing a putative *mraW* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *mraW* RNA in Sections 55.2 and 55.4.

abbrev. of hits	taxonomy of species
Cdi-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Corynebacteriaceae *Corynebacterium diphtheriae* NCTC 13129
Cef-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Corynebacteriaceae *Corynebacterium efficiens* YS-314
Cgl-1-1 to Cgl-1-2	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Corynebacteriaceae *Corynebacterium glutamicum* ATCC 13032
Cgl-2-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Corynebacteriaceae *Corynebacterium glutamicum* R
Cje-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Corynebacteriaceae *Corynebacterium jeikeium* K411
Mgi-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium gilvum* PYR-GCK
Mle-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium leprae* TN
Msm-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium smegmatis str.* MC2 155
Msp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium sp.* JLS
Msp-2-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium sp.* KMS
Msp-3-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium sp.* MCS
Mul-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium ulcerans* Agy99
Mva-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium vanbaalenii* PYR-1
Mav-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumavium complex (MAC) *Mycobacterium avium* 104
Mav-2-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumavium complex (MAC) *Mycobacterium avium subs*
Mbo-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium bovis* AF2122
Mbo-2-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium bovis* BCG st
Mtu-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium tuberculosis* C
Mtu-2-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium tuberculosis* C
Mtu-3-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium tuberculosis* H
Mtu-4-1 to Mtu-4-2	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium tuberculosis* H
Mtu-5-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium tuberculosis* H
Mtu-6-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium tuberculosis s*
Nfa-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Nocardiaceae *Nocardia farcinica* IFM 10152
Rsp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Nocardiaceae *Rhodococcus sp.* RHA1
Kra-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Frankineae Kineosporiaceae *Kineococcus radiotolerans* SRS30216
Jsp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Intrasporangiaceae *Janibacter sp.* HTCC2649
Cmi-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Microbacteriaceae *Clavibacter michiganensis subsp. michiganensis* NCPPB 382
Lxy-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Microbacteriaceae *Leifsonia xyli subsp. xyli str.* CTCB07
Aau-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Micrococcaceae *Arthrobacter aurescens* TC1
Asp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Micrococcaceae *Arthrobacter sp.* FB24
Sar-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micromonosporineae Micromonosporaceae *Salinispora arenicola* CNS205
Str-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micromonosporineae Micromonosporaceae *Salinispora tropica* CNB-440
Pac-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Propionibacterineae Propionibacteriaceae *Propionibacterium acnes* KPA171202
Ser-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Pseudonocardineae Pseudonocardiaceae *Saccharopolyspora erythraea* NRRL 2338
Tfu-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Streptosporangineae Nocardiopsaceae *Thermobifida fusca* YX
mac-1-1	Bacteria Actinobacteria marine actinobacterium PHSC20C1
env-1 to env-21	environmental samples

55.2 Gene contexts

Each *mraW* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *mraW* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 55.3. The accession number of the sequence containing each *mraW* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 55.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	JCVI_SCAF_1096627313926	-	125	13	**RNA**→
Cje-1-1	NC_007164.1	+	884904	885017	**RNA**→ Methyltransf_5 (pfam01795)→ hypo→
env-2	JCVI_SCAF_1096627142908	-	2459	2361	**RNA**→ Methyltransf_5 (pfam01795)? (GOS4290)→ ? (GOS13551726)→ FtsI (COG0768)? (GOS14362002)→
env-3	JCVI_SCAF_1096627149805	+	3981	4079	**RNA**→ COG0220 (COG0220)? (GOS4290)→
env-4	JCVI_SCAF_1096627945542	+	507	605	**RNA**→ ? (GOS4290)→ Methyltransf_5 (pfam01795)? (GOS4290)→
env-5	JCVI_SCAF_1096627351645	+	458	556	**RNA**→ Methyltransf_5 (pfam01795)? (GOS4290)→ ? (GOS13551726)→ PBP_dimer (pfam03717)? (GOS14362477)→
env-6	JCVI_SCAF_1096627352722	+	2839	2937	**RNA**→ Methyltransf_5 (pfam01795)? (GOS4290)→
Asp-1-1	NC_008541.1	+	1740203	1740305	**RNA**→ Methyltransf_5 (pfam01795)→ hypo→ FtsI (COG0768)→ MurE (COG0769)→ MurF (COG0770)→ Rfe (COG0472)→ MurD (COG0771)→ FTSW_RODA_SPOVE (pfam01098)→
Aau-1-1	NC_008711.1	+	1859969	1860072	**RNA**→ Methyltransf_5 (pfam01795)→ hypo→ FtsI (COG0768)→
env-7	JCVI_SCAF_1101668147195	+	882	987	**RNA**→ ? (GOS4290)→
env-8	JCVI_SCAF_1101668137697	-	1371	1266	**RNA**→ ? (GOS4290)→ Methyltransf_5 (pfam01795)? (GOS4290)→ ? (GOS11537123)→
Tfu-1-1	NC_007333.1	+	1283680	1283770	**RNA**→ Methyltransf_5 (pfam01795)→ hypo→ FtsI (COG0768)→
mac-1-1	NZ_AAOB01000011.1	+	3947	4056	**RNA**→ Methyltransf_5 (pfam01795)→ hypo→ FtsI (COG0768)→ MurF (COG0770)→ Rfe (COG0472)→ MurD (COG0771)→ FTSW_RODA_SPOVE (pfam01098)→ MurG (COG0707)→ MurC (COG0773)→
Str-1-1	NC_009380.1	-	3691177	3691072	**RNA**→ Methyltransf_5 (pfam01795)→ hypo→ FtsI (COG0768)→ MurE (COG0769)→ MurF (COG0770)→ Rfe (COG0472)→ FTSW_RODA_SPOVE (pfam01098)→ MurG (COG0707)→
Sar-1-1	NZ_AAWA01000005.1	-	102975	102868	**RNA**→ Methyltransf_5 (pfam01795)→ hypo→ FtsI (COG0768)→ MurE (COG0769)→ MurF (COG0770)→ Rfe (COG0472)→
Cdi-1-1	NC_002935.2	-	1640095	1639987	**RNA**→ Methyltransf_5 (pfam01795)→ hypo→ FtsI (COG0768)→
env-9	BABE01001411.1	-	1117	1012	**RNA**→ ←hypo
Pac-1-1	NC_006085.1	+	823520	823625	**RNA**→ Methyltransf_5 (pfam01795)→ hypo→ FtsI (COG0768)→ MurE (COG0769)→ MurF (COG0770)→ Rfe (COG0472)→ MurD (COG0771)→ FTSW_RODA_SPOVE (pfam01098)→
Cef-1-1	NC_004369.1	-	2185285	2185181	**RNA**→ Methyltransf_5 (pfam01795)→ hypo→ FtsI (COG0768)→ MurE (COG0769)→ MurF (COG0770)→ Rfe (COG0472)→ MurD (COG0771)→ FTSW_RODA_SPOVE (pfam01098)→
Cgl-1-1	NC_003450.3	-	2295350	2295245	**RNA**→ Methyltransf_5 (pfam01795)→ hypo→
Cgl-1-2	NC_006958.1	-	2267021	2266916	**RNA**→ Methyltransf_5 (pfam01795)→
Cgl-2-1	NC_009342.1	-	2259341	2259236	**RNA**→ Methyltransf_5 (pfam01795)→ hypo→
env-10	JCVI_SCAF_1096627776797	-	463	356	**RNA**→
env-11	JCVI_SCAF_1096627355507	+	3936	4043	**RNA**→ Methyltransf_5 (pfam01795)? (GOS4290)→
env-12	JCVI_SCAF_1096627944833	-	288	181	**RNA**→
env-13	JCVI_SCAF_1096627354316	+	1769	1876	**RNA**→ ←? (GOS14444919)
env-14	JCVI_SCAF_1096627132945	-	1705	1587	**RNA**→ Methyltransf_5 (pfam01795)? (GOS4290)→
env-15	JCVI_SCAF_1096627361752	+	84	181	**RNA**→
env-16	JCVI_SCAF_1096627360054	-	1492	1395	**RNA**→ Methyltransf_5 (pfam01795)? (GOS4290)→ ? (GOS14674019)→
env-17	JCVI_SCAF_1096627946713	-	173	76	**RNA**→
env-18	JCVI_SCAF_1096627352118	+	2428	2525	**RNA**→
env-19	JCVI_SCAF_1101668609352	-	1325	1228	**RNA**→
env-20	JCVI_SCAF_1096627936744	+	631	728	**RNA**→ Methyltransf_5 (pfam01795)? (GOS4290)→

env-21	JCVI_SCAF_1096626904734	+	687	788	**RNA**→ Methyltransf_5 (pfam01795)? (GOS4290)→
Ser-1-1	NC_009142.1	-	6587882	6587766	**RNA**→ Methyltransf_5 (pfam01795)→ hypo→ FtsI (COG0768)→ IclR (COG1414)→
Kra-1-1	NC_009664.1	-	3297371	3297257	**RNA**→ Methyltransf_5 (pfam01795)→ hypo→ FtsI (COG0768)→
Rsp-1-1	NC_008268.1	-	1168667	1168567	**RNA**→ Methyltransf_5 (pfam01795)→ possible conserved membrane protein→ FtsI (COG0768)→
Msp-3-1	NC_008146.1	-	3463004	3462909	**RNA**→ Methyltransf_5 (pfam01795)→ FHA domain containing protein→ FtsI (COG0768)→ MurE (COG0769)→ MurF (COG0770)→ Rfe (COG0472)→ MurD (COG0771)→ FTSW_RODA_SPOVE (pfam01098)→
Msp-2-1	NC_008705.1	-	3491286	3491191	**RNA**→ Methyltransf_5 (pfam01795)→ FHA domain containing protein→ FtsI (COG0768)→ MurE (COG0769)→ MurF (COG0770)→ Rfe (COG0472)→ MurD (COG0771)→ FTSW_RODA_SPOVE (pfam01098)→
Msp-1-1	NC_009077.1	-	3436786	3436691	**RNA**→ Methyltransf_5 (pfam01795)→ FHA domain containing protein→ FtsI (COG0768)→ MurE (COG0769)→ MurF (COG0770)→ Rfe (COG0472)→ MurD (COG0771)→ FTSW_RODA_SPOVE (pfam01098)→
Nfa-1-1	NC_006361.1	+	1911134	1911235	**RNA**→ Methyltransf_5 (pfam01795)→
Msm-1-1	NC_008596.1	-	4320284	4320175	**RNA**→ Methyltransf_5 (pfam01795)→ hypo→ FtsI (COG0768)→ MurE (COG0769)→ MurF (COG0770)→ phospho-N-acetylmuramoyl-pentapeptide-transferase→ MurD (COG0771)→ FTSW_RODA_SPOVE (pfam01098)→
Mle-1-1	NC_002677.1	+	1071520	1071629	**RNA**→ Methyltransf_5 (pfam01795)→ putative conserved membrane protein→ FtsI (COG0768)→
Mul-1-1	NC_008611.1	-	3895253	3895150	**RNA**→ Methyltransf_5 (pfam01795)→ conserved proline rich membrane protein→ FtsI (COG0768)→
Mav-2-1	NC_002944.2	-	2107071	2106968	**RNA**→ Methyltransf_5 (pfam01795)→ hypo→ FtsI (COG0768)→
Mav-1-1	NC_008595.1	+	2329935	2330038	**RNA**→ Methyltransf_5 (pfam01795)→ hypo→ FtsI (COG0768)→
Mtu-3-1	NC_009565.1	-	2442984	2442881	**RNA**→ Methyltransf_5 (pfam01795)→ conserved proline rich membrane protein→ FtsI (COG0768)→
Mbo-2-1	NC_008769.1	-	2409311	2409208	**RNA**→ Methyltransf_5 (pfam01795)→ putative proline rich membrane protein→ FtsI (COG0768)→
Mtu-4-1	NC_009525.1	-	2439325	2439222	**RNA**→ Methyltransf_5 (pfam01795)→ putative conserved proline rich membrane protein→ FtsI (COG0768)→
Mtu-2-1	NC_002755.2	-	2428115	2428012	**RNA**→ Methyltransf_5 (pfam01795)→ hypo→ FtsI (COG0768)→
Mtu-5-1	NC_000962.2	-	2429423	2429320	**RNA**→ Methyltransf_5 (pfam01795)→ PROBABLE CONSERVED PROLINE RICH MEMBRANE PROTEIN→ FtsI (COG0768)→
Mbo-1-1	NC_002945.3	-	2409815	2409712	**RNA**→ Methyltransf_5 (pfam01795)→ PROBABLE CONSERVED PROLINE RICH MEMBRANE PROTEIN→ FtsI (COG0768)→
Mtu-6-1	NZ_AASN01000036.1	-	5206	5103	**RNA**→
Mtu-4-2	NZ_AAYK01000127.1	+	1111	1214	**RNA**→
Mtu-1-1	NZ_AAKR01000069.1	-	5321	5218	**RNA**→ COG0275: Predicted S-adenosylmethionine-dependent methyltransferase involved in cell envelope biogenesis→
Mgi-1-1	NC_009338.1	+	3149681	3149784	**RNA**→ Methyltransf_5 (pfam01795)→ hypo→ FtsI (COG0768)→ MurE (COG0769)→ MurF (COG0770)→ phospho-N-acetylmuramoyl-pentapeptide-transferase→ MurD (COG0771)→ FTSW_RODA_SPOVE (pfam01098)→
Mva-1-1	NC_008726.1	-	3758649	3758546	**RNA**→ Methyltransf_5 (pfam01795)→ FHA domain containing protein→ FtsI (COG0768)→
Jsp-1-1	NZ_AAMN01000002.1	+	113267	113369	**RNA**→ Methyltransf_5 (pfam01795)→ putative membrane protein→
Cmi-1-1	NC_009480.1	-	2111756	2111656	**RNA**→ Methyltransf_5 (pfam01795)→ hypo→ FtsI (COG0768)→ MurE (COG0769)→ MurF (COG0770)→ Rfe (COG0472)→ MurD (COG0771)→ FTSW_RODA_SPOVE (pfam01098)→
Lxy-1-1	NC_006087.1	-	1598755	1598652	**RNA**→ Methyltransf_5 (pfam01795)→ hypo→ FtsI (COG0768)→ MurE (COG0769)→ MurF (COG0770)→ Rfe (COG0472)→ MurD (COG0771)→ FTSW_RODA_SPOVE (pfam01098)→

55.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 55.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *mraW* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

COG0220 (1) Predicted S-adenosylmethionine-dependent methyltransferase [General function prediction only]

COG0472 (11) UDP-N-acetylmuramyl pentapeptide phosphotransferase/UDP-N-acetylglucosamine-1-phosphate transferase [Cell envelope biogenesis, outer membrane]

COG0707 (2) UDP-N-acetylglucosamine:LPS N-acetylglucosamine transferase [Cell envelope biogenesis, outer membrane]

COG0768 (31) Cell division protein FtsI/penicillin-binding protein 2 [Cell envelope biogenesis, outer membrane]

COG0769 (12) UDP-N-acetylmuramyl tripeptide synthase [Cell envelope biogenesis, outer membrane]

COG0770 (13) UDP-N-acetylmuramyl pentapeptide synthase [Cell envelope biogenesis, outer membrane]

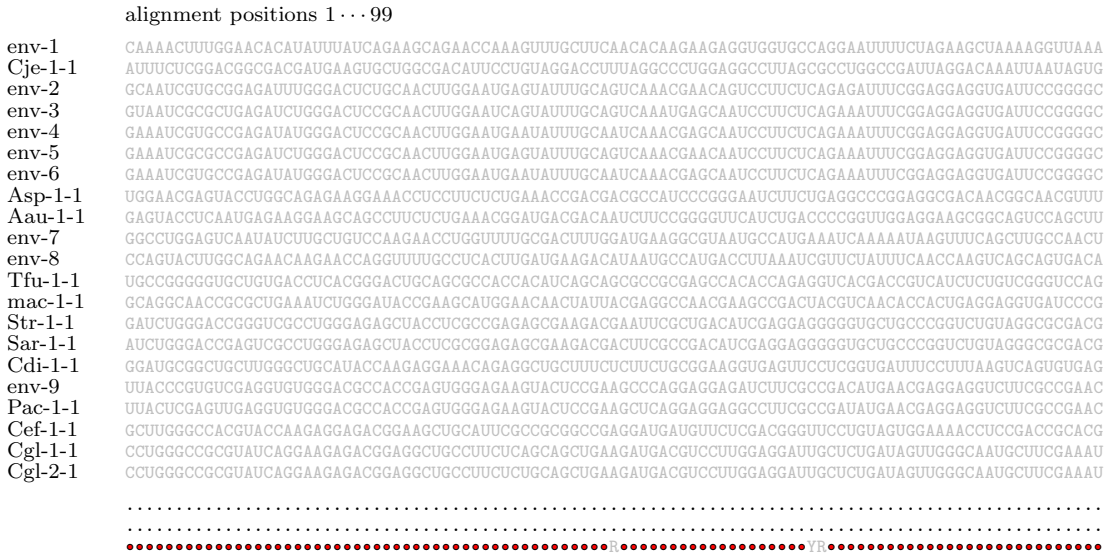
COG0771 (11) UDP-N-acetylmuramoylalanine-D-glutamate ligase [Cell envelope biogenesis, outer membrane]

COG0773 (1) UDP-N-acetylmuramate-alanine ligase [Cell envelope biogenesis, outer membrane]
COG1414 (1) Transcriptional regulator [Transcription]
GOS11537123 (1) ?
GOS13551726 (2) ?
GOS14362002 (1) ?
GOS14362477 (1) ?
GOS14444919 (1) ?
GOS14674019 (1) ?
GOS4290 (14) ?
pfam01098 (12) Cell cycle protein.
pfam01795 (46) MraW methylase family.
pfam03717 (1) Penicillin-binding Protein dimerisation domain.

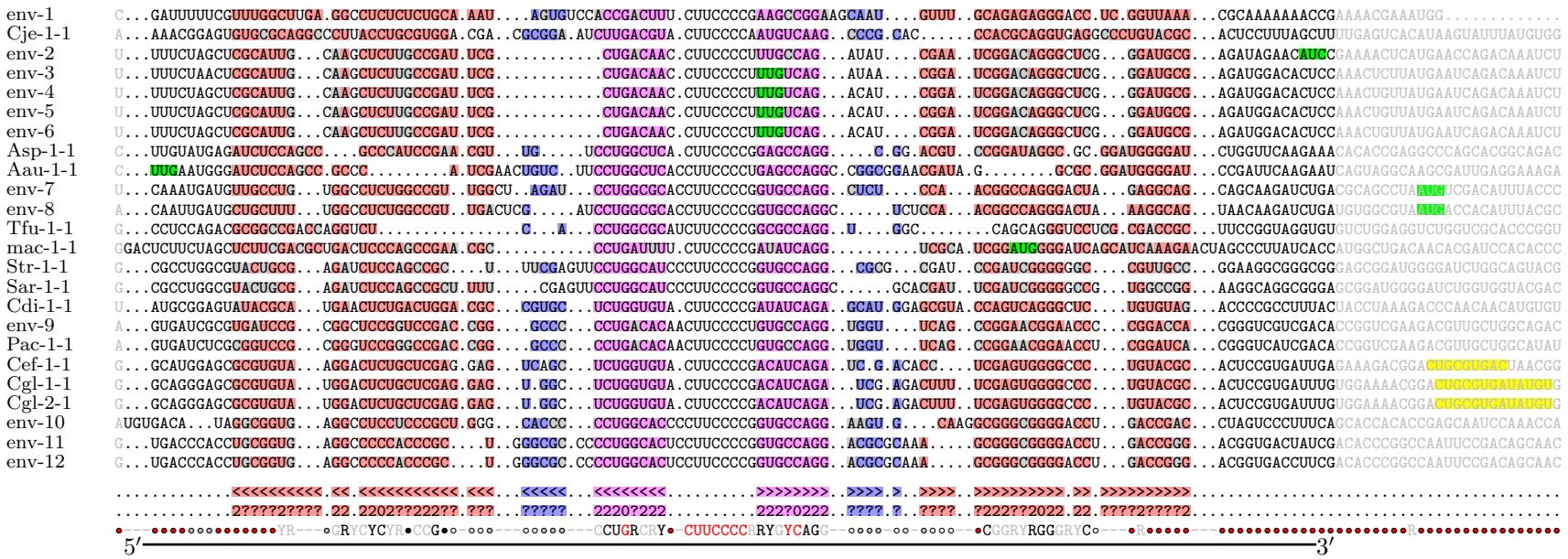
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Cgl-1-2, Mbo-1-1, Mbo-2-1, Msp-2-1, Mtu-2-1, Mtu-4-1, Mtu-4-2, Mtu-5-1, Mtu-6-1

55.4 Multiple-sequence alignment

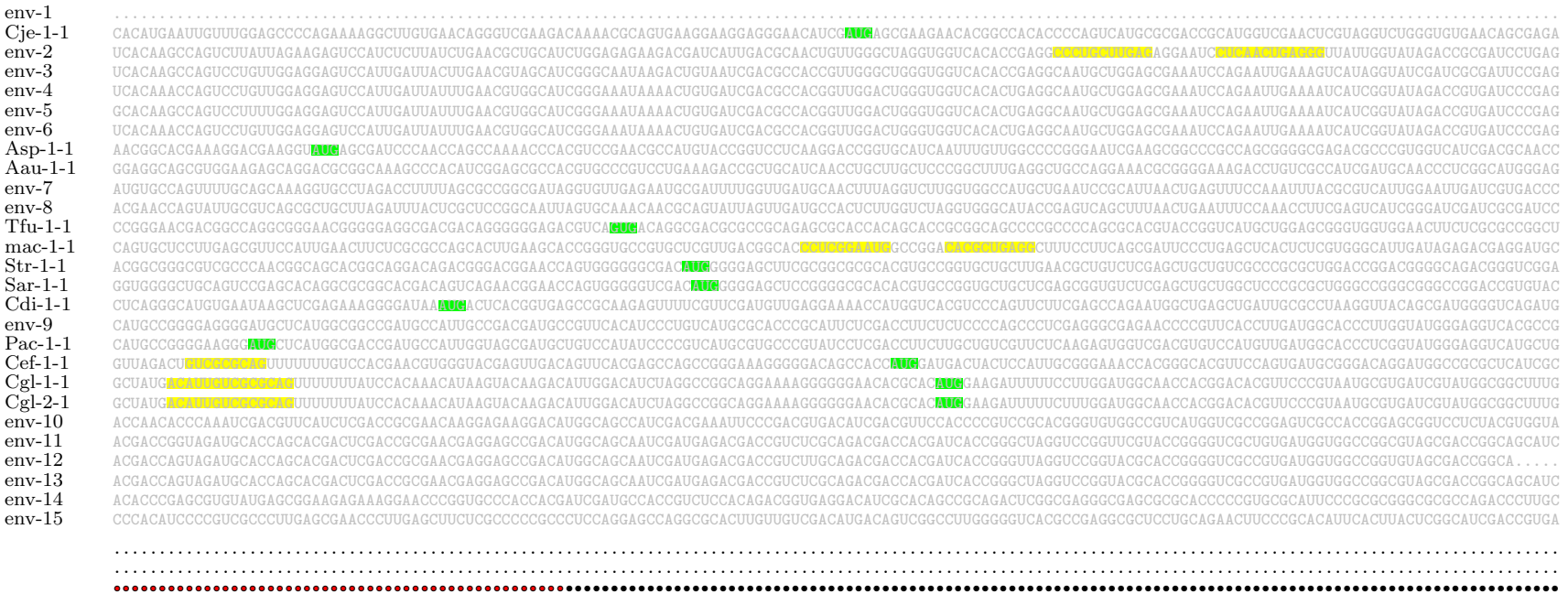
Each *mraW* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 55.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *mraW* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

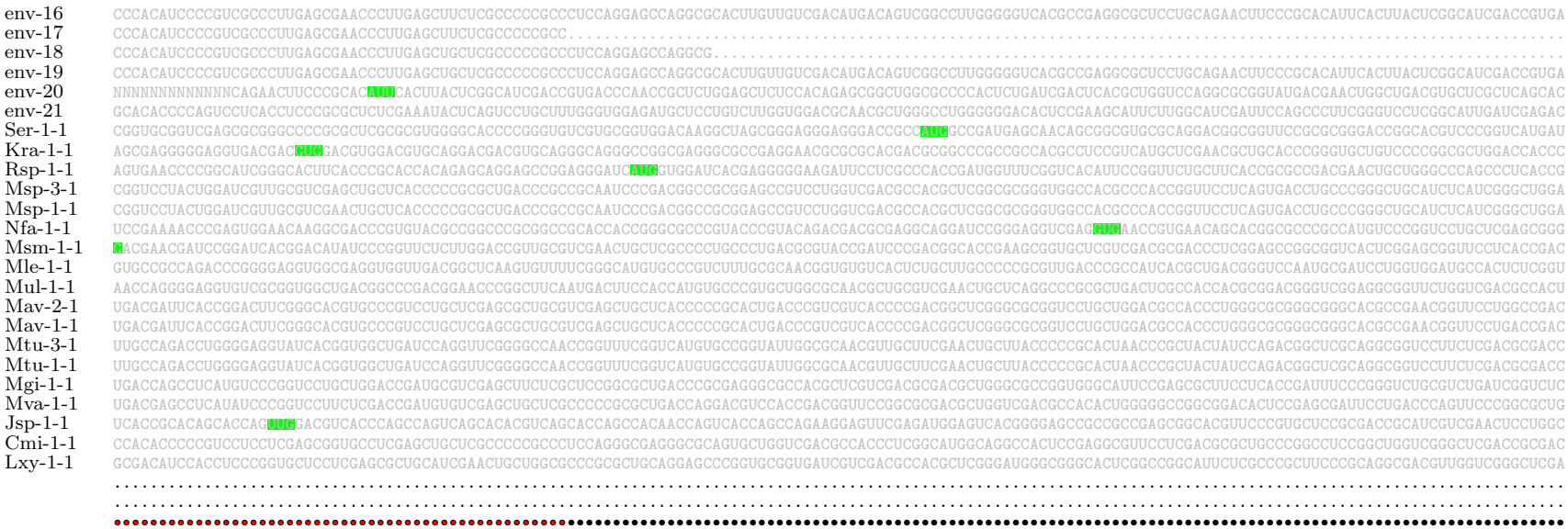

alignment positions 1 · · · 99
env-1 CAAAACUUUGGAACACAUAUUUAUCAGAAGCAGAACCAAAGUUUGCUUCAACACAAGAAGAGGUGGUGCCAGGAAUUUUCUAGAAGCUAAAAGGUUAAA
Cje-1-1 AUUUCUCGGACGGCGACGAUGAAGUGCUGGCGACAUUCCUGUAGGACCUUUAGGCCCUGGAGGCCUUAGCGCCUGGCCGAUUAGGACAAAUUAAUAGUG
env-2 GCAAUCGUGCGGAGAUUUGGGACUCUGCAACUUGGAAUGAGUAUUUGCAGUCAAACGAACAGUCCUUCUCAGAGAUUUCGGAGGAGGUGAUUCCGGGGC
env-3 GUAAUCGCGCUGAGAUCUGGGACUCCGCAACUUGGAAUCAGUAUUUGCAGUCAAAUGAGCAAUCCUUCUCAGAAAUUUCGGAGGAGGUGAUUCCGGGGC
env-4 GAAAUCGUGCCGAGAUAUGGGACUCCGCAACUUGGAAUGAAUAUUUGCAAUCAAACGAGCAAUCCUUCUCAGAAAUUUCGGAGGAGGUGAUUCCGGGGC
env-5 GAAAUCGCGCCGAGAUCUGGGACUCCGCAACUUGGAAUGAGUAUUUGCAGUCAAACGAACAAUCCUUCUCAGAAAUUUCGGAGGAGGUGAUUCCGGGGC
env-6 GAAAUCGUGCCGAGAUAUGGGACUCCGCAACUUGGAAUGAAUAUUUGCAAUCAAACGAGCAAUCCUUCUCAGAAAUUUCGGAGGAGGUGAUUCCGGGGC
Asp-1-1 UGGAACGAGUACCUGGCAGAGAAGGAAACCUCCUUCUCUGAAACCGACGACGCCAUCCCGGGAAUCUUCUGAGGCCCGGAGGCGACAACGGCAACGUUU
Aau-1-1 GAGUACCUCAAUGAGAAGGAAGCAGCCUUCUCUGAAACGGAUGACGACAAUCUUCCGGGGUUCAUCUGACCCCGGUUGGAGGAAGCGGCAGUCCAGCUU
env-7 GGCCUGGAGUCAAUAUCUUGCUGUCCAAGAACCUGGUUUUGCGACUUUGGAUGAAGGCGUAAUGCCAUGAAAUCAAAAAUAAGUUUCAGCUUGCCAACU
env-8 CCAGUACUUGGCAGAACAAGAACCAGGUUUUGCCUCACUUGAUGAAGACAUAAUGCCAUGACCUUAAAUCGUUCUAUUUCAACCAAGUCAGCAGUGACA
Tfu-1-1 UGCCGGGGGUGCUGUGACCUCACGGGACUGCAGCGCCACCACAUCAGCAGGCGCCGCGAGCCACACCAGAGGUCACGACCGUCAUCCUCUGUCGGGUCCAG
mac-1-1 GCAGGCAACCGCGCUGAAAUCUGGGAUACCGAAGCAUGGAACAACUAUUACGAGGCCAACGAAGCCGACUACGUCAACACCACUGAGGAGGUGAUCCCG
Str-1-1 GAUCUGGGACCGGGUCGCCUGGGAGAGCUACCUCGCCGAGAGCGAAGACGAAUUCGCUGACAUCGAGGAGGGGGUGCUGCCCGGUCUGUAGGGCGCGACG
Sar-1-1 AUCUGGGACCGAGUCGCCUGGGAGAGCUACCUCGCGGAGAGCGAAGACGACUUCGCCGACAUCGAGGAGGGGGUGCUGCCCGGUCUGUAGGGCGCGACG
Cdi-1-1 GGAUGCGGCUGCUUGGGCUGCAUACCAAGAGGAAACAGAGGCUGCUUUCUCUUCUGCGGAAGGUGAGUUCCUCGGUGAUUUCCUUUAAGUCAGUGUGAG
env-9 UUACCCGUGUCGAGGUGUGGGACGCCACCGAGUGGGAGAAGUACUCCGAAGCCCAGGAGGAGAUCUUCGCCGACAUGAACGAGGAGGUCUUCGCCGAAC
Pac-1-1 UUACUCGAGUUGAGGUGUGGGACGCCACCGAGUGGGAGAAGUACUCCGAAGCUCAGGAGGAGGCCUUCGCCGAUAUGAACGAGGAGGUCUUCGCCGAAC
Cef-1-1 GCUUGGGCCACGUACCAAGAGGAGACGGAAGCUGCAUUCGCCGCGGCCGAGGAUGAUGUUCUCGACGGGUUCCUGUAGUGGAAAACCUCCGACCGCACG
Cgl-1-1 CCUGGGCCGCGUAUCAGGAAGAGACGGAGGCUGCCUUCUCAGCAGCUGAAGAUGACGUCCUUGGAGGAUUGCUCUGAUAGUUGGGCAAUGCUUCGAAAU
Cgl-2-1 CCUGGGCCGCGUAUCAGGAAGAGACGGAGGCUGCCUUCUCUGCAGCUGAAGAUGACGUCCUUGGAGGAUUGCUCUGAUAGUUGGGCAAUGCUUCGAAAU
. . . . . .
. . . . . .
• • • •

alignment positions 100 · · · 259

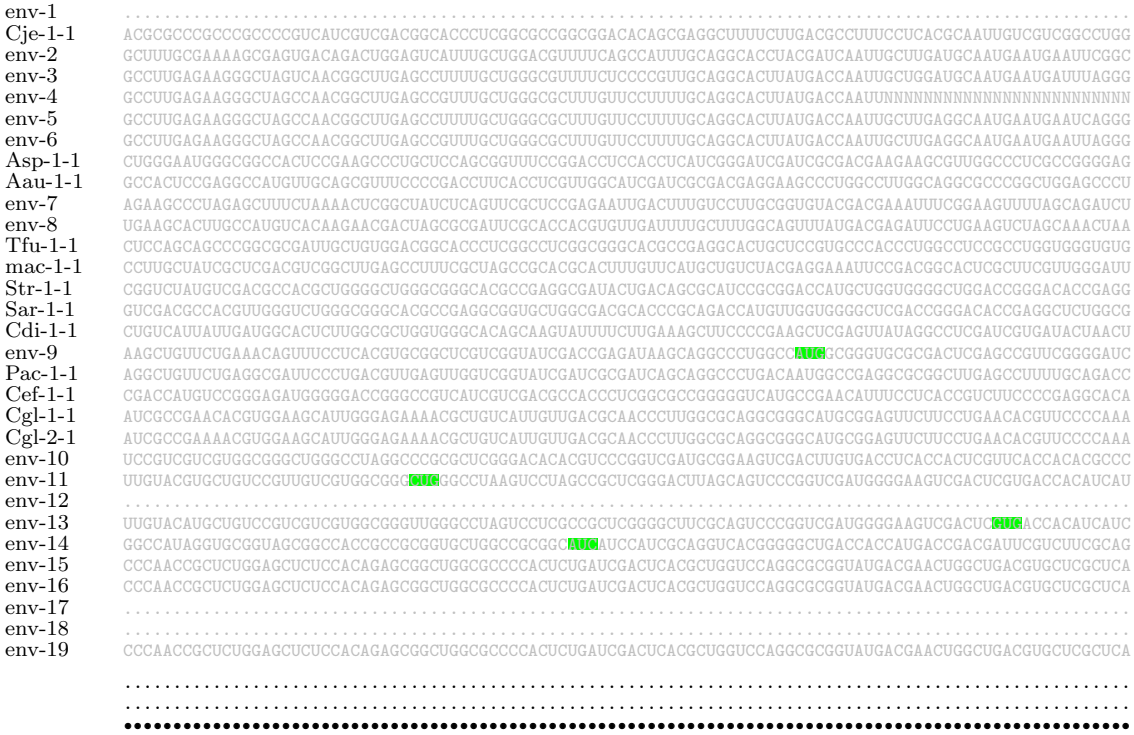


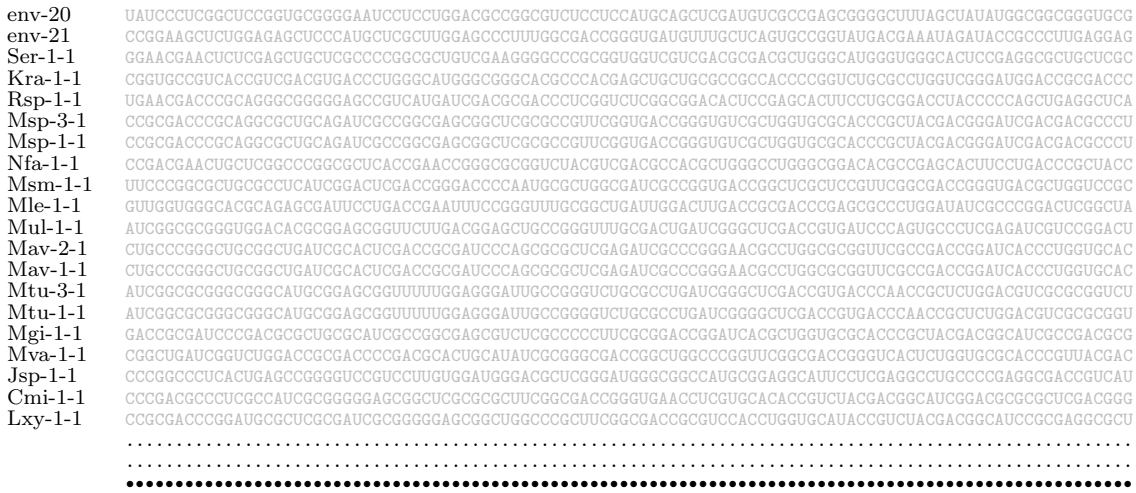
alignment positions 260 · · · 419





alignment positions 420 · · · 521





56 *msiK*

56.1 Taxa

The taxonomy of each organism containing a putative *msiK* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *msiK* RNA in Sections 56.2 and 56.4.

abbrev. of hits	**taxonomy of species**
Aod-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Actinomycineae Actinomycetaceae *Actinomyces odontolyticus* ATCC 17982
Cdi-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Corynebacteriaceae *Corynebacterium diphtheriae* NCTC 13129
Cgl-1-1 to Cgl-1-2	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Corynebacteriaceae *Corynebacterium glutamicum* ATCC 13032
Cgl-2-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Corynebacteriaceae *Corynebacterium glutamicum* R
Nfa-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Nocardiaceae *Nocardia farcinica* IFM 10152
Rsp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Nocardiaceae *Rhodococcus sp.* RHA1
Kra-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Frankineae Kineosporiaceae *Kineococcus radiotolerans* SRS30216
Jsp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Intrasporangiaceae *Janibacter sp.* HTCC2649
Cmi-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Microbacteriaceae *Clavibacter michiganensis subsp. michiganensis* NCPPB 382
Lxy-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Microbacteriaceae *Leifsonia xyli subsp. xyli str.* CTCB07
Aau-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Micrococcaceae *Arthrobacter aurescens* TC1
Asp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Micrococcaceae *Arthrobacter sp.* FB24
Sar-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micromonosporineae Micromonosporaceae *Salinispora arenicola* CNS205
Str-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micromonosporineae Micromonosporaceae *Salinispora tropica* CNB-440
Nsp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Propionibacterineae Nocardioidaceae *Nocardioides sp.* JS614
Pac-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Propionibacterineae Propionibacteriaceae *Propionibacterium acnes* KPA171202
Sav-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Streptomycineae Streptomycetaceae *Streptomyces avermitilis* MA-4680
Sco-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Streptomycineae Streptomycetaceae *Streptomyces coelicolor* A3(2)
Bad-1-1	Bacteria Actinobacteria Actinobacteridae Bifidobacteriales Bifidobacteriaceae *Bifidobacterium adolescentis* ATCC 15703
Bad-2-1	Bacteria Actinobacteria Actinobacteridae Bifidobacteriales Bifidobacteriaceae *Bifidobacterium adolescentis* L2-32
Blo-1-1	Bacteria Actinobacteria Actinobacteridae Bifidobacteriales Bifidobacteriaceae *Bifidobacterium longum* DJO10A
Blo-2-1	Bacteria Actinobacteria Actinobacteridae Bifidobacteriales Bifidobacteriaceae *Bifidobacterium longum* NCC2705
mac-1-1	Bacteria Actinobacteria marine actinobacterium PHSC20C1
env-1 to env-112	environmental samples

56.2 Gene contexts

Each *msiK* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *msiK* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 56.3. The accession number of the sequence containing each *msiK* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 56.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Cdi-1-1	NC_002935.2	-	509334	509270	**RNA**→ MalK (COG3839)→
Cgl-1-1	NC_003450.3	-	2610701	2610636	**RNA**→ MalK (COG3839)→ Kdo (pfam06293)→
Cgl-1-2	NC_006958.1	-	2582375	2582310	**RNA**→ MalK (COG3839)→ Kdo (pfam06293)NOG04040 (NOG04040)→
Cgl-2-1	NC_009342.1	-	2605619	2605554	**RNA**→ MalK (COG3839)→ Kdo (pfam06293)→
env-1	JCVI_SCAF_1096627011185	+	1024	1088	**RNA**→ ←TlyC (COG1253)? (GOS1639)
env-2	JCVI_SCAF_1101668116881	-	928	864	**RNA**→ MalK (COG3839)? (GOS2568)→
Aod-1-1	NZ_AAYI02000004.1	+	407474	407536	**RNA**→ MalK (COG3839)→
env-3	JCVI_SCAF_1101667048432	+	812	869	**RNA**→
Sco-1-1	NC_003888.3	-	4646377	4646310	**RNA**→ MalK (COG3839)→
env-4	JCVI_SCAF_1101667182974	-	200	141	**RNA**→
env-5	JCVI_SCAF_1096627485321	+	838	902	**RNA**→ MalK (COG3839)? (GOS2568)→
env-6	JCVI_SCAF_1101668131180	+	260	323	**RNA**→ MalK (COG3839)? (GOS2568)→
env-7	JCVI_SCAF_1101668101207	+	1006	1063	**RNA**→
env-8	JCVI_SCAF_1101668098689	-	1308	1245	**RNA**→ ABC_MalK_N (cd03301)? (GOS2568)→
env-9	JCVI_SCAF_1096626895685	-	729	665	**RNA**→ MalK (COG3839)? (GOS2568)→
env-10	JCVI_SCAF_1096626903559	-	116	55	**RNA**→
env-11	JCVI_SCAF_1096627644396	+	415	479	**RNA**→ MalK (COG3839)? (GOS2568)→
env-12	JCVI_SCAF_1101668241821	+	1454	1515	**RNA**→
env-13	JCVI_SCAF_1096627640173	+	1704	1765	**RNA**→
env-14	JCVI_SCAF_1096627382797	+	6628	6692	**RNA**→ MalK (COG3839)? (GOS2568)→
env-15	JCVI_SCAF_1096627641758	+	1536	1597	**RNA**→
env-16	JCVI_SCAF_1096627018908	+	315	379	**RNA**→ MalK (COG3839)? (GOS2568)→
env-17	JCVI_SCAF_1096627661581	+	175	239	**RNA**→ MalK (COG3839)? (GOS2568)→ ? (GOS346)→
env-18	JCVI_SCAF_1101668250029	-	506	442	**RNA**→ ←TlyC (COG1253)? (GOS1639)
env-19	JCVI_SCAF_1096627007057	+	5444	5508	**RNA**→ MalK (COG3839)? (GOS2568)→
env-20	JCVI_SCAF_1101667176518	+	65	129	**RNA**→
env-21	JCVI_SCAF_1096627654446	+	396	460	**RNA**→ ABC_MalK_N (cd03301)? (GOS2568)→ MalK (COG3839)? (GOS2568)→
env-22	JCVI_SCAF_1096627017980	-	132	71	**RNA**→
env-23	JCVI_SCAF_1096627640505	-	1179	1115	**RNA**→ ? (GOS2568)→ MalK (COG3839)? (GOS2568)→
env-24	JCVI_SCAF_1101668102051	+	958	1022	**RNA**→ MalK (COG3839)? (GOS2568)→
env-25	JCVI_SCAF_1101668100322	-	264	205	**RNA**→
env-26	JCVI_SCAF_1101668237940	+	1576	1636	**RNA**→
env-27	JCVI_SCAF_1096627296405	+	553	620	**RNA**→ MalK (COG3839)? (GOS2568)→
env-28	JCVI_SCAF_1096627674912	-	178	120	**RNA**→
env-29	JCVI_SCAF_1096627017852	+	1738	1802	**RNA**→ MalK (COG3839)? (GOS2568)→
env-30	JCVI_SCAF_1101668251211	-	741	677	**RNA**→ MalK (COG3839)? (GOS2568)→
env-31	JCVI_SCAF_1096627014293	-	652	588	**RNA**→ MalK (COG3839)? (GOS2568)→
env-32	JCVI_SCAF_1101667169701	+	772	830	**RNA**→

env-33	JCVI_SCAF_1101668138352	-	307	249	**RNA**→
env-34	JCVI_SCAF_1101668122183	+	131	195	**RNA**→ MalK (COG3839)? (GOS2568)→ ? (GOS346)→
env-35	JCVI_SCAF_1101668123036	-	655	591	**RNA**→ MalK (COG3839)? (GOS2568)→
env-36	JCVI_SCAF_1096626060652	+	13	77	**RNA**→
env-37	JCVI_SCAF_1101668098278	+	440	504	**RNA**→ ABC_MalK_N (cd03301)? (GOS2568)→ predicted protein→ ? (GOS2568)→
env-38	JCVI_SCAF_1096627296044	-	1124	1060	**RNA**→ MalK (COG3839)? (GOS2568)→
env-39	JCVI_SCAF_1101668125363	-	501	437	**RNA**→ MalK (COG3839)? (GOS2568)→
env-40	JCVI_SCAF_1101668115963	-	1044	980	**RNA**→ MalK (COG3839)? (GOS2568)→
env-41	JCVI_SCAF_1101668126374	-	289	231	**RNA**→
env-42	JCVI_SCAF_1101668139245	-	236	178	**RNA**→
env-43	JCVI_SCAF_1101668147889	+	662	726	**RNA**→ MalK (COG3839)? (GOS2568)→
Pac-1-1	NC_006085.1	-	439778	439714	**RNA**→ MalK (COG3839)→
env-44	JCVI_SCAF_1096627011507	-	1281	1217	**RNA**→ MalK (COG3839)? (GOS2568)→
env-45	ABEF01020574.1	+	174	236	**RNA**→ MalK (COG3839)→
env-46	BABF01001037.1	+	4132	4196	**RNA**→ MalK (COG3839)→
env-47	BABE01000010.1	-	12297	12233	**RNA**→ MalK (COG3839)→
env-48	BABC01000051.1	-	3300	3236	**RNA**→ MalK (COG3839)→
env-49	BABE01002017.1	-	388	324	**RNA**→ ABC_MalK_N (cd03301)→
env-50	BABE01000234.1	-	10755	10691	**RNA**→ MalK (COG3839)→
Blo-2-1	NC_004307.2	+	1661060	1661124	**RNA**→ MalK (COG3839)→
env-51	BAAY01012752.1	+	1	61	**RNA**→
env-52	BAAZ01002988.1	+	638	702	**RNA**→ MalK (COG3839)→
env-53	BAAY01007686.1	-	255	195	**RNA**→
env-54	BABC01000362.1	-	2644	2580	**RNA**→ MalK (COG3839)→
env-55	hgutS7_s7_164770	+	85	149	**RNA**→
env-56	BAAY01007496.1	-	1139	1075	**RNA**→ MalK (COG3839)→
Bad-2-1	NZ_AAXD02000031.1	-	37139	37075	**RNA**→ ←hypo
env-57	BAAV01027802.1	+	168	232	**RNA**→ MalK (COG3839)→
Bad-1-1	NC_008618.1	-	2026471	2026407	**RNA**→ MalK (COG3839)→
Blo-1-1	NZ_AABM02000008.1	-	53477	53413	**RNA**→ MalK (COG3839)→
env-58	BABD01003636.1	+	708	772	**RNA**→ MalK (COG3839)→
Nfa-1-1	NC_006361.1	+	5869874	5869937	**RNA**→ MalK (COG3839)→
Sav-1-1	NC_003155.3	+	4887082	4887145	**RNA**→ MalK (COG3839)→
Nsp-1-1	NC_008699.1	+	1122407	1122471	**RNA**→ MalK (COG3839)→
env-59	JCVI_SCAF_1101668115049	+	644	708	**RNA**→ MalK (COG3839)? (GOS2568)→
Cmi-1-1	NC_009480.1	+	2799611	2799675	**RNA**→ MalK (COG3839)→
env-60	JCVI_SCAF_1096626856533	+	378	443	**RNA**→ MalK (COG3839)? (GOS2568)→
mac-1-1	NZ_AAOB01000005.1	+	162025	162090	**RNA**→ MalK (COG3839)→ hypo→
Aau-1-1	NC_008711.1	-	988504	988439	**RNA**→ MalK (COG3839)→
Asp-1-1	NC_008541.1	-	812415	812350	**RNA**→ MalK (COG3839)→
env-61	JCVI_SCAF_1101668162609	+	868	933	**RNA**→ MalK (COG3839)? (GOS2568)→
env-62	JCVI_SCAF_1096627015612	+	171	236	**RNA**→ MalK (COG3839)? (GOS2568)→
Lxy-1-1	NC_006087.1	+	1880494	1880558	**RNA**→ MalK (COG3839)→ hypo→
env-63	JCVI_SCAF_1096626908686	+	673	737	**RNA**→ MalK (COG3839)? (GOS2568)→ ? (GOS4505)→
env-64	JCVI_SCAF_1096626902982	+	867	931	**RNA**→ MalK (COG3839)? (GOS2568)→ MalK (COG3839)? (GOS2568)→
env-65	JCVI_SCAF_1101668063560	+	29	93	**RNA**→
env-66	JCVI_SCAF_1096627299376	-	2403	2339	**RNA**→ MalK (COG3839)? (GOS2568)→ ? (GOS4505)→
env-67	JCVI_SCAF_1101667060800	+	402	466	**RNA**→ ABC_MalK_N (cd03301)? (GOS2568)→
env-68	JCVI_SCAF_1101668160299	-	821	757	**RNA**→ MalK (COG3839)? (GOS2568)→
env-69	JCVI_SCAF_1096627941248	-	796	734	**RNA**→ MalK (COG3839)? (GOS2568)→
env-70	JCVI_SCAF_1101668626336	-	972	910	**RNA**→ MalK (COG3839)? (GOS2568)→
env-71	JCVI_SCAF_1101668634135	-	354	292	**RNA**→

env-72	JCVI_SCAF_1096627136573	+	3720	3782	**RNA**→ MalK (COG3839)? (GOS2568)→
env-73	JCVI_SCAF_1101668634120	-	1080	1018	**RNA**→ MalK (COG3839)? (GOS2568)→
env-74	JCVI_SCAF_1096627352346	-	2197	2135	**RNA**→ MalK (COG3839)? (GOS2568)→ ? (GOS4505)→
env-75	JCVI_SCAF_1101668613380	+	782	844	**RNA**→ MalK (COG3839)? (GOS2568)→
env-76	JCVI_SCAF_1101667308769	+	1	54	**RNA**→
env-77	JCVI_SCAF_1101668600310	+	1420	1476	**RNA**→
env-78	JCVI_SCAF_1096627134351	+	1956	2012	**RNA**→
env-79	JCVI_SCAF_1096627136532	-	2203	2141	**RNA**→ ABC_MalK_N (cd03301)? (GOS2568)→ MalK (COG3839)? (GOS2568)→ ? (GOS4505)→
env-80	JCVI_SCAF_1096627931309	-	836	774	**RNA**→ MalK (COG3839)? (GOS2568)→
env-81	JCVI_SCAF_1101668643763	-	551	489	**RNA**→ MalK (COG3839)? (GOS2568)→
env-82	JCVI_SCAF_1101668648367	+	464	526	**RNA**→ MalK (COG3839)? (GOS2568)→ MalK (COG3839)? (GOS2568)→ ? (GOS4505)→
env-83	JCVI_SCAF_1101668624842	+	7	71	**RNA**→
env-84	JCVI_SCAF_1101668629913	+	1011	1074	**RNA**→ ABC_MalK_N (cd03301)→
env-85	JCVI_SCAF_1096627956280	-	1015	950	**RNA**→ MalK (COG3839)? (GOS2568)→
env-86	JCVI_SCAF_1096627366818	-	527	462	**RNA**→ MalK (COG3839)? (GOS2568)→
env-87	JCVI_SCAF_1096627361765	-	706	641	**RNA**→ MalK (COG3839)? (GOS2568)→
env-88	JCVI_SCAF_1096627148878	+	212	277	**RNA**→ MalK (COG3839)? (GOS2568)→ Kdo (pfam06293)? (GOS4505)→
env-89	JCVI_SCAF_1096627954948	+	1586	1645	**RNA**→
env-90	JCVI_SCAF_1101668636189	-	892	827	**RNA**→ MalK (COG3839)? (GOS2568)→
env-91	JCVI_SCAF_1096627143060	+	68	133	**RNA**→
env-92	JCVI_SCAF_1096627358736	+	1402	1467	**RNA**→ MalK (COG3839)? (GOS2568)→
env-93	JCVI_SCAF_1101668641574	-	295	236	**RNA**→
env-94	JCVI_SCAF_1101668652296	+	1161	1226	**RNA**→ ABC_MalK_N (cd03301)? (GOS2568)→
env-95	JCVI_SCAF_1096627353881	+	587	652	**RNA**→ MalK (COG3839)? (GOS2568)→ ? (GOS4505)→
env-96	JCVI_SCAF_1096627947013	+	1126	1185	**RNA**→
env-97	JCVI_SCAF_1101668602492	-	1172	1107	**RNA**→ MalK (COG3839)? (GOS2568)→
env-98	JCVI_SCAF_1101668645534	+	673	738	**RNA**→ MalK (COG3839)? (GOS2568)→
env-99	JCVI_SCAF_1096627352959	-	159	100	**RNA**→
env-100	JCVI_SCAF_1096627951462	-	291	232	**RNA**→
env-101	JCVI_SCAF_1101668606824	-	826	761	**RNA**→ MalK (COG3839)? (GOS2568)→
env-102	JCVI_SCAF_1096627146777	+	1756	1821	**RNA**→ MalK (COG3839)? (GOS2568)→
env-103	JCVI_SCAF_1101668615963	+	943	1008	**RNA**→ MalK (COG3839)? (GOS2568)→
env-104	JCVI_SCAF_1101668608466	+	426	491	**RNA**→ MalK (COG3839)? (GOS2568)→
env-105	JCVI_SCAF_1096627935864	-	892	827	**RNA**→ MalK (COG3839)? (GOS2568)→
env-106	JCVI_SCAF_1101668610543	-	795	730	**RNA**→ MalK (COG3839)? (GOS2568)→
env-107	JCVI_SCAF_1101667309074	-	625	560	**RNA**→
env-108	JCVI_SCAF_1101668623937	+	1430	1488	**RNA**→
env-109	JCVI_SCAF_1101668643446	-	167	109	**RNA**→
env-110	JCVI_SCAF_1101668031962	-	1289	1225	**RNA**→ MalK (COG3839)? (GOS2568)→
Kra-1-1	NC_009664.1	-	947419	947352	**RNA**→ MalK (COG3839)→
Jsp-1-1	NZ_AAMN01000005.1	-	40921	40857	**RNA**→ MalK (COG3839)→
Sar-1-1	NZ_AAWA01000003.1	+	75360	75423	**RNA**→ MalK (COG3839)→
Str-1-1	NC_009380.1	+	4785498	4785561	**RNA**→ MalK (COG3839)→
Rsp-1-1	NC_008268.1	-	4502789	4502725	**RNA**→ hypo→ MalK (COG3839)→
env-111	JCVI_SCAF_1096627384802	-	1408	1344	**RNA**→ MalK (COG3839)? (GOS2568)→
env-112	JCVI_SCAF_1096627135960	-	375	311	**RNA**→ ←UgpB (COG1653)? (GOS3613)

56.3 Conserved domains

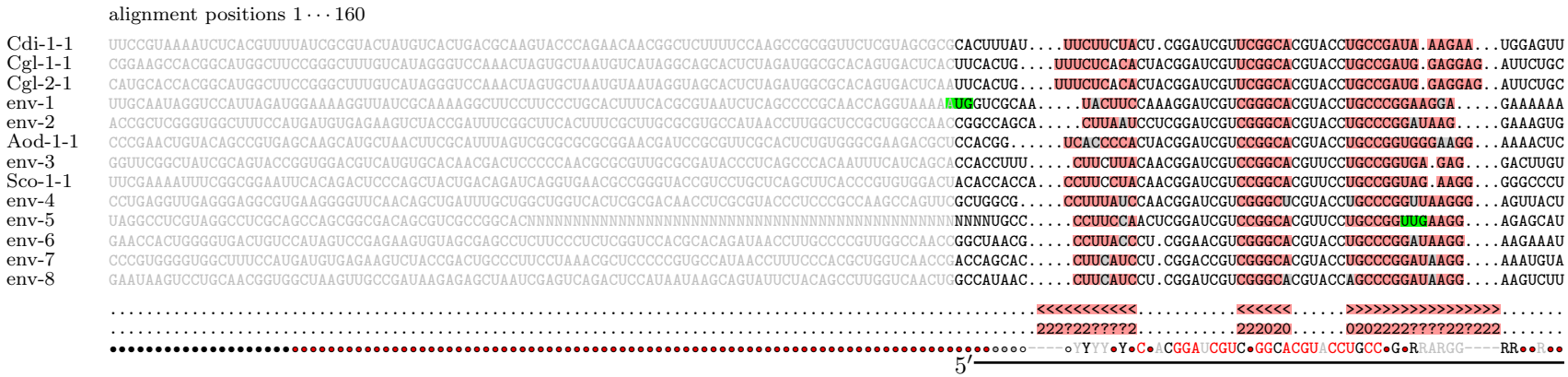
Conserved domains found in protein-coding genes listed in Section 56.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *msiK* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

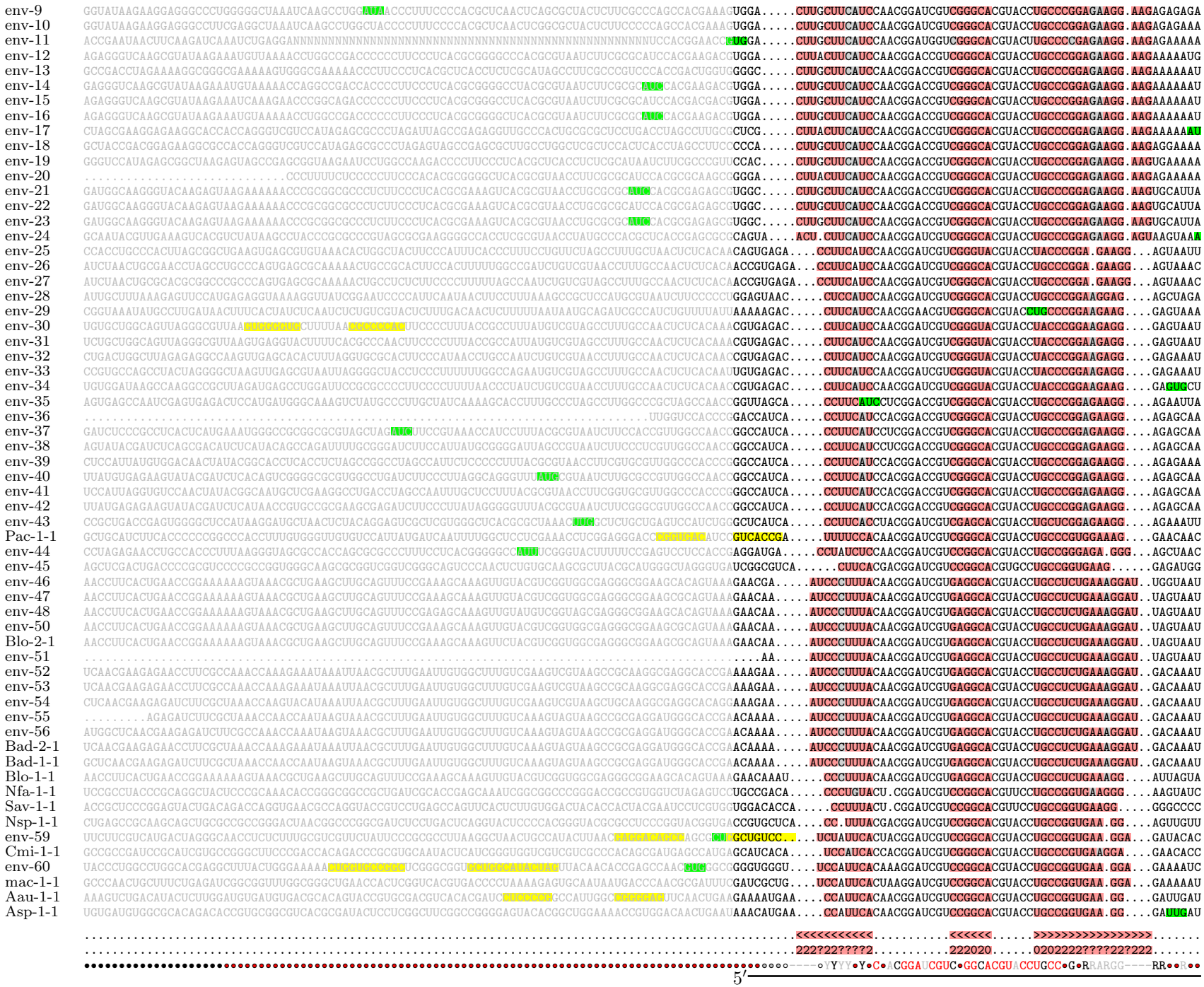
cd03301 (8) The N-terminal ATPase domain of the maltose transporter, MalK.
COG1253 (2) Hemolysins and related proteins containing CBS domains [General function prediction only]
COG1653 (1) ABC-type sugar transport system, periplasmic component [Carbohydrate transport and metabolism]
COG3839 (93) ABC-type sugar transport systems, ATPase components [Carbohydrate transport and metabolism]
GOS1639 (2) ?
GOS2568 (68) ?
GOS346 (2) ?
GOS3613 (1) ?
GOS4505 (7) ?
NOG04040 (1)
pfam06293 (4) Lipopolysaccharide kinase (Kdo/WaaP) family.

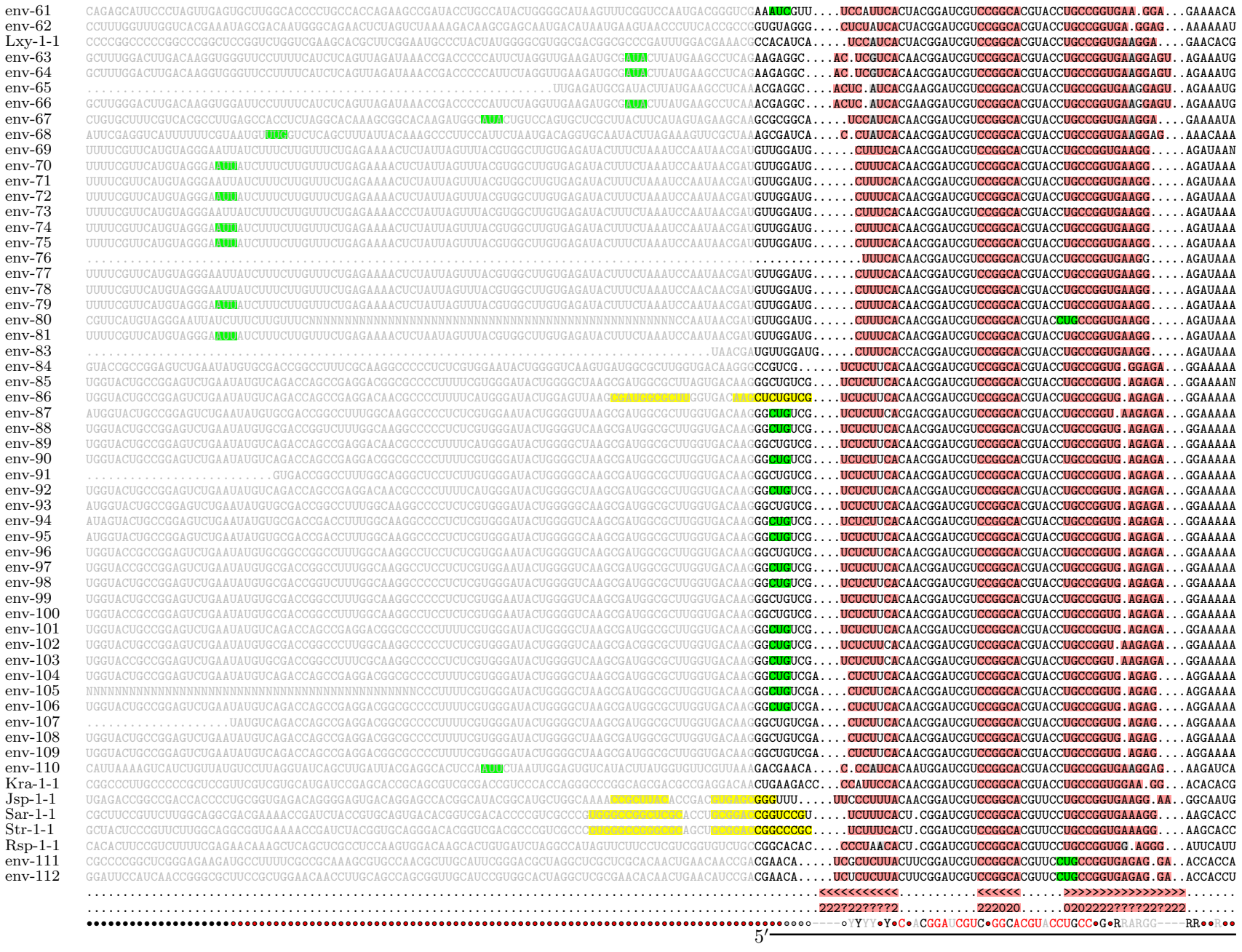
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Cgl-1-2, env-49, env-57, env-58, env-82

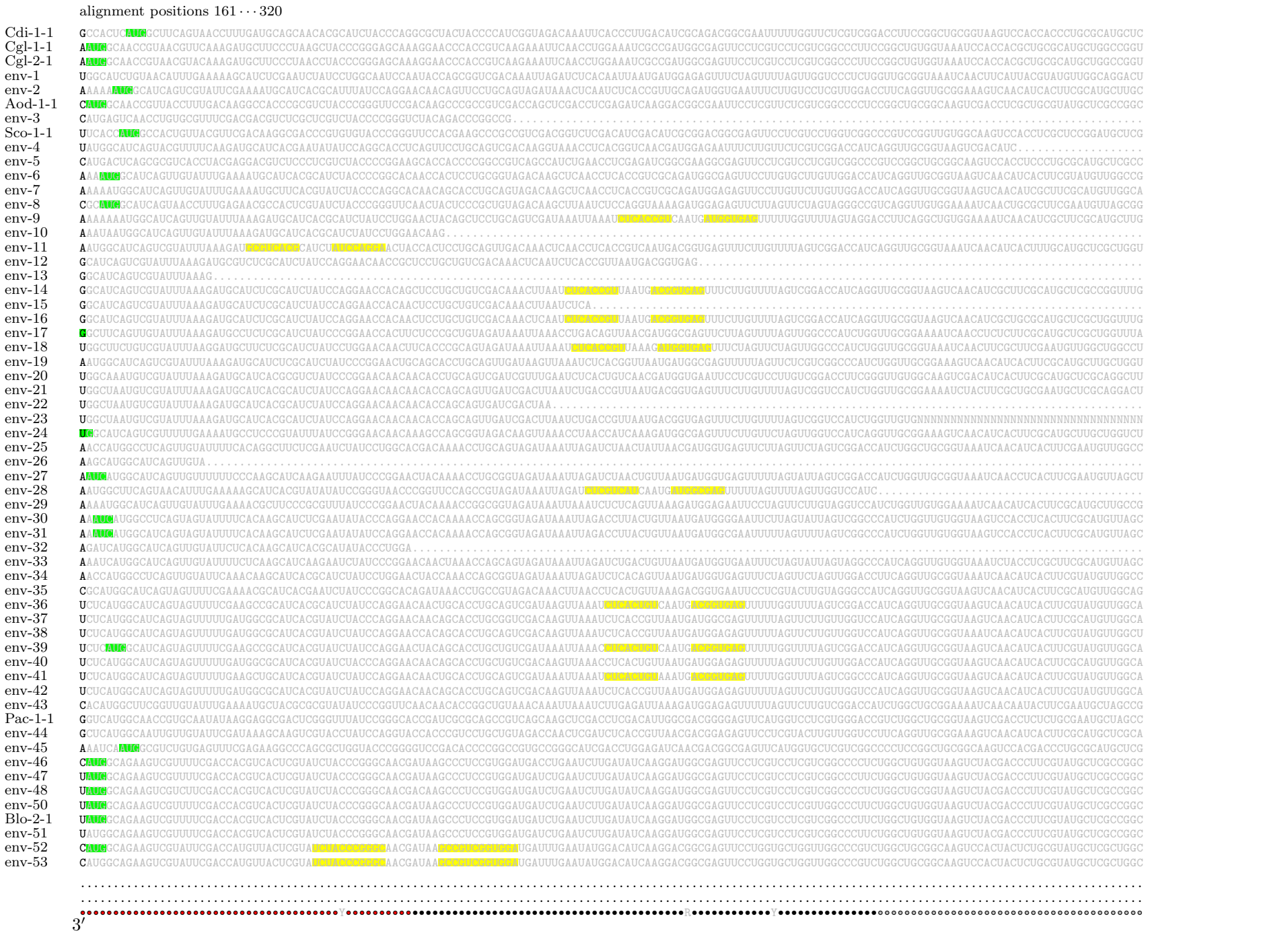
56.4 Multiple-sequence alignment

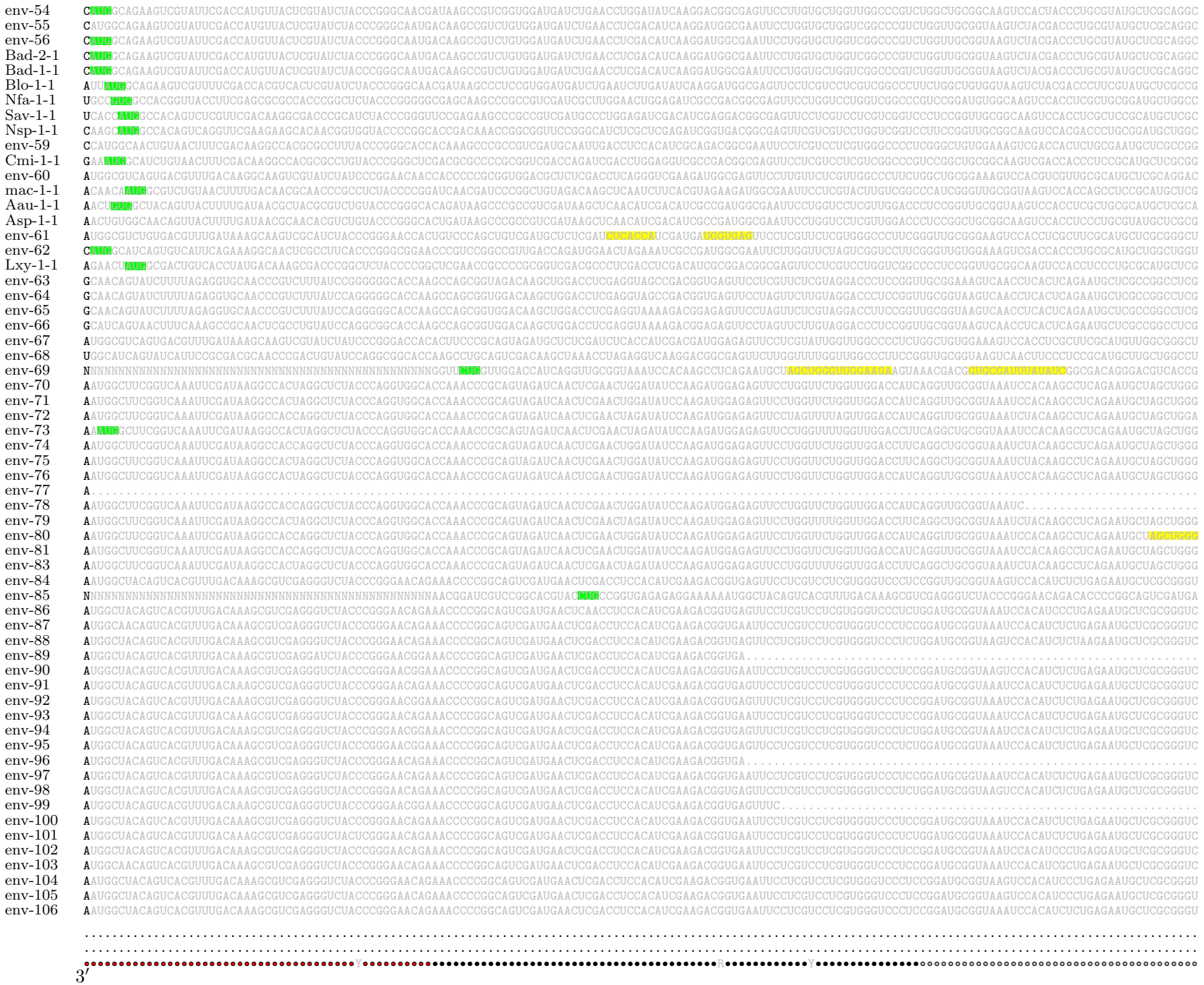
Each *msiK* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 56.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *msiK* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

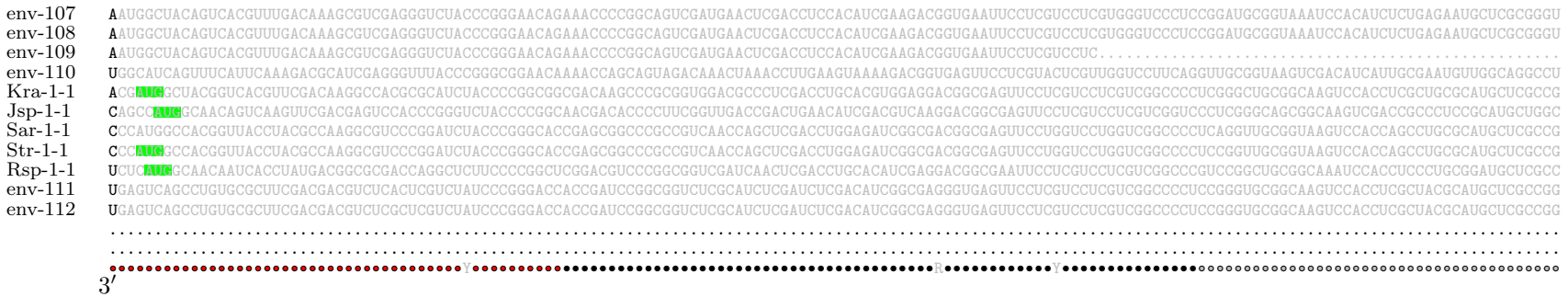



env-9
env-10
env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
Pac-1-1
env-44
env-45
env-46
env-47
env-48
env-50
Blo-2-1
env-51
env-52
env-53
env-54
env-55
env-56
Bad-2-1
Bad-1-1
Blo-1-1
Nfa-1-1
Sav-1-1
Nsp-1-1
env-59
Cmi-1-1
env-60
mac-1-1
Aau-1-1
Asp-1-1
5′

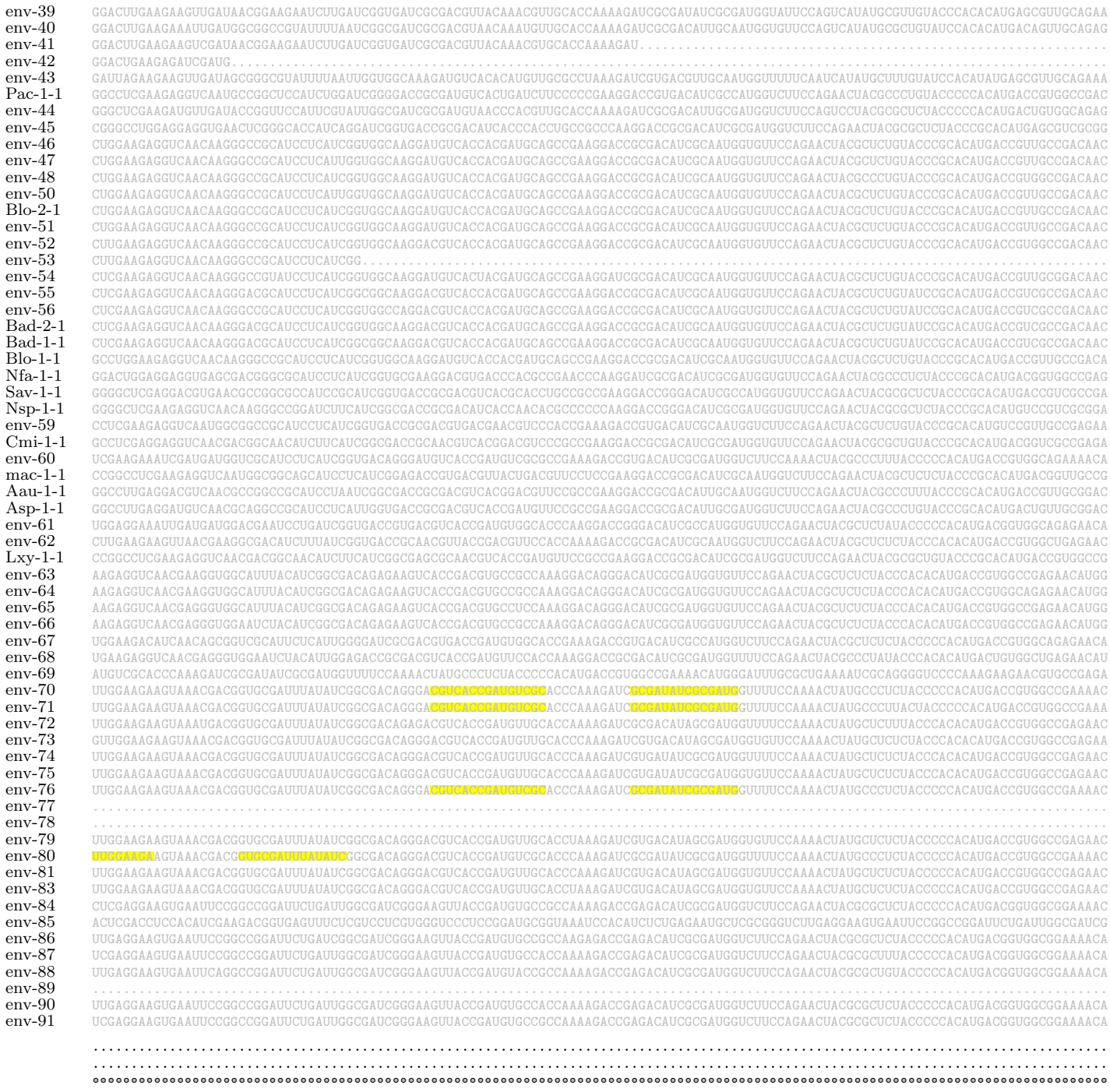

env-61
env-62
Lxy-1-1
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
Kra-1-1
Jsp-1-1
Sar-1-1
Str-1-1
Rsp-1-1
env-111
env-112
5′


alignment positions 161 · · · 320
Cdi-1-1
Cgl-1-1
Cgl-2-1
env-1
env-2
Aod-1-1
env-3
Sco-1-1
env-4
env-5
env-6
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
Pac-1-1
env-44
env-45
env-46
env-47
env-48
env-50
Blo-2-1
env-51
env-52
env-53
3′


env-54
env-55
env-56
Bad-2-1
Bad-1-1
Blo-1-1
Nfa-1-1
Sav-1-1
Nsp-1-1
env-59
Cmi-1-1
env-60
mac-1-1
Aau-1-1
Asp-1-1
env-61
env-62
Lxy-1-1
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
3′



alignment positions 321 · · · 452


env-39
env-40
env-41
env-42
env-43
Pac-1-1
env-44
env-45
env-46
env-47
env-48
env-50
Blo-2-1
env-51
env-52
env-53
env-54
env-55
env-56
Bad-2-1
Bad-1-1
Blo-1-1
Nfa-1-1
Sav-1-1
Nsp-1-1
env-59
Cmi-1-1
env-60
mac-1-1
Aau-1-1
Asp-1-1
env-61
env-62
Lxy-1-1
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91

57 *Nitrosococcus*-1

57.1 Taxa

The taxonomy of each organism containing a putative *Nitrosococcus*-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *Nitrosococcus*-1 RNA in Sections 57.2 and 57.4.

abbrev. of hits	taxonomy of species
Nth-1-1	Bacteria Firmicutes Clostridia Natranaerobiales Natranaerobiaceae *Natranaerobius thermophilus* JW/NM-WN-LF
Noc-1-1 to Noc-1-6	Bacteria Proteobacteria Gammaproteobacteria Chromatiales Chromatiaceae *Nitrosococcus oceani* ATCC 19707
Noc-2-1 to Noc-2-6	Bacteria Proteobacteria Gammaproteobacteria Chromatiales Chromatiaceae *Nitrosococcus oceani* C-27
env-1	environmental sample

57.2 Gene contexts

Each *Nitrosococcus*-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *Nitrosococcus*-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 57.3. The accession number of the sequence containing each *Nitrosococcus*-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 57.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Nth-1-1	NC_010718.1	+	856446	856499	hypo→ **RNA**→ hypo→
Noc-1-1	NC_007484.1	+	1054957	1055010	hypo→ **RNA**→ COG1561 (COG1561)→
Noc-2-1	NZ_ABSG01000004.1	+	83128	83181	**RNA**→
Noc-1-2	NC_007484.1	-	305965	305912	Vgb (COG4257)→ **RNA**→ ←TtrR (COG4566)
Noc-2-2	NZ_ABSG01000032.1	-	143681	143628	**RNA**→
Noc-1-3	NC_007484.1	-	2282238	2282184	hypo→ **RNA**→ Poly_export (pfam02563)→
Noc-2-3	NZ_ABSG01000019.1	-	254130	254076	**RNA**→
env-1	wf3_Contig2836	-	632	566	Transposase_35 (pfam07282)→ **RNA**→ hypo→
Noc-1-4	NC_007484.1	-	1448886	1448832	←CysJ (COG0369) **RNA**→ hypo→
Noc-2-4	NZ_ABSG01000013.1	-	3521	3467	**RNA**→
Noc-1-5	NC_007484.1	+	1561911	1561965	MgtA (COG0474)XkdP (COG1652)→ **RNA**→ hypo→
Noc-2-5	NZ_ABSG01000013.1	+	116544	116598	**RNA**→
Noc-1-6	NC_007484.1	+	658624	658677	TPR (cd00189)→ **RNA**→ ←XseA (COG1570)
Noc-2-6	NZ_ABSG01000039.1	+	5143	5196	**RNA**→

57.3 Conserved domains

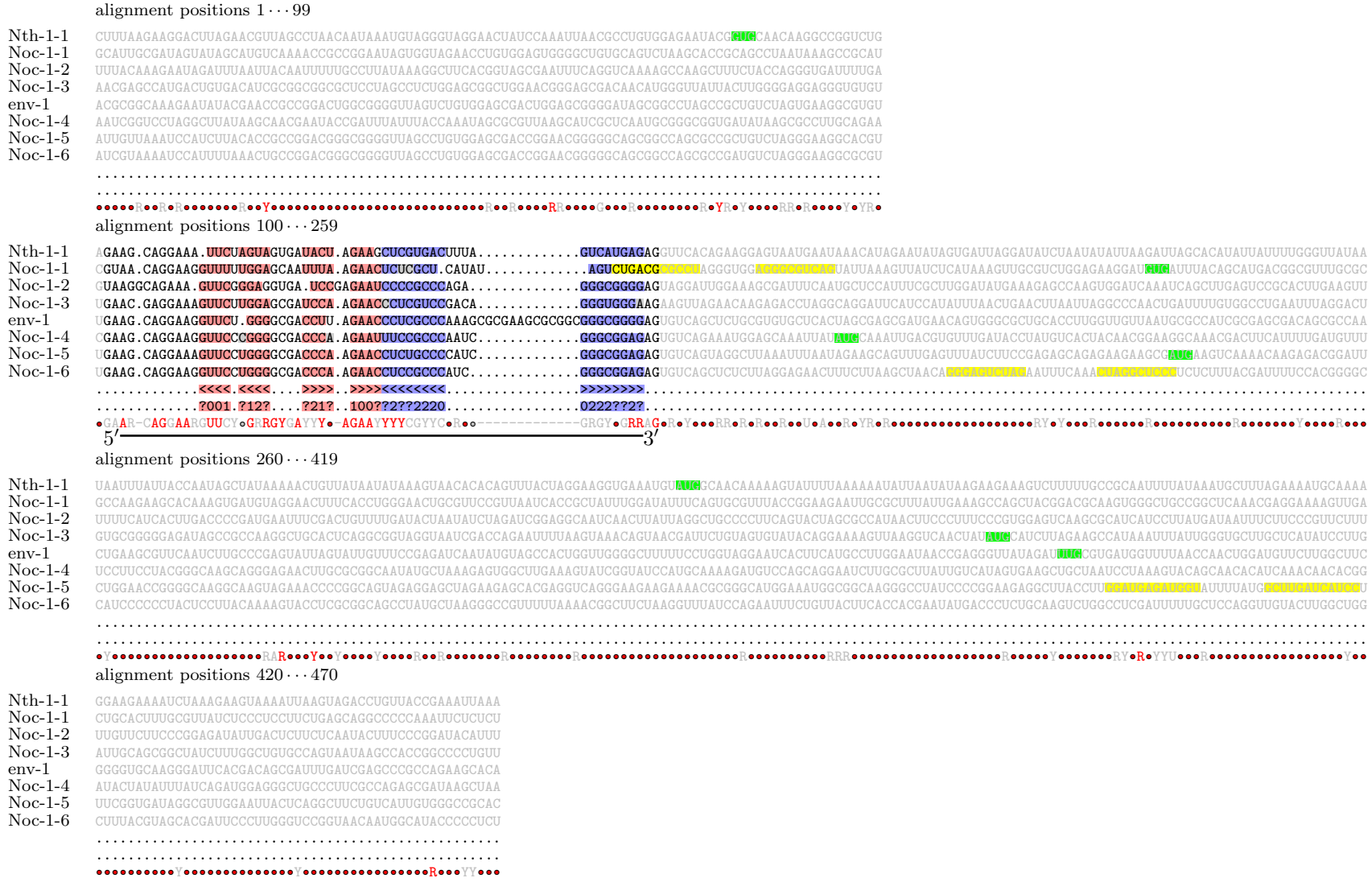
Conserved domains found in protein-coding genes listed in Section 57.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *Nitrosococcus*-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

u cd00189 (1) Tetratricopeptide repeat domain; typically contains 34 amino acids [WLF]-X(2)-[LIM]-[GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-X(2)-[ASL]-X(4)-[PKE]
u COG0369 (1) Sulfite reductase, alpha subunit (flavoprotein) [Inorganic ion transport and metabolism]
u COG0474 (1) Cation transport ATPase [Inorganic ion transport and metabolism]
D COG1561 (1) Uncharacterized stress-induced protein [Function unknown]
d COG1570 (1) Exonuclease VII, large subunit [DNA replication, recombination, and repair]
u COG1652 (1) Uncharacterized protein containing LysM domain [Function unknown]
u COG4257 (1) Streptogramin lyase [Defense mechanisms]
d COG4566 (1) Response regulator [Signal transduction mechanisms]
D pfam02563 (1) Polysaccharide biosynthesis/export protein.
u pfam07282 (1) Putative transposase DNA-binding domain.

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Noc-2-1, Noc-2-2, Noc-2-3, Noc-2-4, Noc-2-5, Noc-2-6

57.4 Multiple-sequence alignment

Each *Nitrosococcus*-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 57.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *Nitrosococcus*-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



58 *nuoG*

58.1 Taxa

The taxonomy of each organism containing a putative *nuoG* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *nuoG* RNA in Sections 58.2 and 58.4.

abbrev. of hits	taxonomy of species
Cko-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Citrobacter koseri* ATCC BAA-895
Esa-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Cronobacter *Enterobacter sakazakii* ATCC BAA-894
Eca-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Enterobacter cancerogenus* ATCC 35316
Esp-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Enterobacter sp.* 638
Eta-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Erwinia tasmaniensis* Et1/99
Eal-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia albertii* TW07627
Eco-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* 101-1
Eco-2-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* 536
Eco-3-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* 53638
Eco-4-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* APEC O1
Eco-5-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* ATCC 8739
Eco-6-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* B171
Eco-7-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* B7A
Eco-8-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* CFT073
Eco-9-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* E110019
Eco-10-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* E22
Eco-11-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* E24377A
Eco-12-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* F11
Eco-13-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* HS
Eco-14-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 EDL933
Eco-15-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4024
Eco-16-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4042
Eco-17-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4045
Eco-18-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4076
Eco-19-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4113
Eco-20-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4115
Eco-21-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4196
Eco-22-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4206
Eco-23-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4401
Eco-24-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4486
Eco-25-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4501
Eco-26-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC508
Eco-27-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC869
Eco-28-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. Sakai
Eco-29-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. TW14588
Eco-30-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* SE11
Eco-31-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* SMS-3-5
Eco-33-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli str.* K12 substr. DH10B
Eco-34-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli str.* K12 substr. MG1655
Eco-35-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli str.* K12 substr. W3110
Eco-32-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* UTI89
Pat-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Pectobacterium atrosepticum* SCRI1043
Pca-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Pectobacterium carotovorum subsp. brasiliensis* PBR1692
Pca-2-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Pectobacterium carotovorum subsp. carotovorum* WPP14
Plu-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Photorhabdus luminescens subsp. laumondii* TTO1
Pmi-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Proteus mirabilis* HI4320
Pal-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Providencia alcalifaciens* DSM 30120
Pru-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Providencia rustigianii* DSM 4541
Pst-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Providencia stuartii* ATCC 25827
Spr-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Serratia proteamaculans* 568
Sbo-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella boydii* CDC 3083-94
Sbo-2-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella boydii* Sb227
Sdy-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella dysenteriae* 1012
Sdy-2-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella dysenteriae* Sd197
Sfl-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella flexneri* 2a str. 2457T
Sfl-2-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella flexneri* 2a str. 301
Sfl-3-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella flexneri* 5 str. 8401
Sso-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella sonnei* Ss046
Sgl-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Sodalis glossinidius str. 'morsitans'*
Ybe-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia bercovieri* ATCC 43970
Yen-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia enterocolitica subsp. enterocolitica* 8081
Yin-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia intermedia* ATCC 29909
Ymo-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia mollaretii* ATCC 43969
Ype-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* Angola

Ype-2-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* Antiqua
Ype-8-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Antiqua str. B42003004
Ype-9-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Antiqua str. E1979001
Ype-10-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Antiqua str. UG05-0454
Ype-11-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Mediaevalis str. K1973002
Ype-12-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Microtus str. 91001
Ype-13-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Orientalis str. F1991016
Ype-14-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Orientalis str. IP275
Ype-15-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Orientalis str. MG05-1020
Ype-3-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* CA88-4125
Ype-4-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* CO92
Ype-5-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* KIM
Ype-6-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* Nepal516
Ype-7-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* Pestoides F
Yps-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pseudotuberculosis* IP 31758
Yps-2-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pseudotuberculosis* IP 32953
Yps-3-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pseudotuberculosis* PB1/+
Yps-4-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pseudotuberculosis* YPIII
env-1 to env-4	environmental samples

58.2 Gene contexts

Each *nuoG* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *nuoG* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 58.3. The accession number of the sequence containing each *nuoG* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 58.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	BABF01000162.1	-	50	8	**RNA**→
Eco-2-1	NC_008253.1	-	2431502	2431460	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-12-1	NZ_AAJU02000001.1	-	415860	415818	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
env-2	BAAV01005753.1	-	455	413	**RNA**→ fer2 (cd00207)→
Eco-32-1	NC_007946.1	-	2532357	2532315	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-4-1	NC_008563.1	-	2505047	2505005	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-8-1	NC_004431.1	-	2682720	2682678	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-3-1	NZ_AAKB02000001.1	+	3649726	3649768	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-1-1	NZ_AAMK02000002.1	-	329675	329633	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Sdy-1-1	NZ_AAMJ02000014.1	-	91911	91869	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-33-1	NC_010473.1	-	2490002	2489960	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-35-1	AC_000091.1	-	2404885	2404843	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→

Eco-5-1	NC_010468.1	+	1500709	1500751	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-34-1	NC_000913.2	-	2398237	2398195	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-13-1	NC_009800.1	-	2441786	2441744	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
env-3	BAAW01001350.1	-	1417	1375	**RNA**→ NuoG (COG1034)→
Eco-28-1	NC_002695.1	-	3126385	3126343	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-20-1	NC_011353.1	-	3167747	3167705	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-14-1	NC_002655.2	-	3196341	3196299	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-22-1	NZ_ABHK02000001.1	-	441965	441923	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-17-1	NZ_ABHL02000001.1	-	310396	310354	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-16-1	NZ_ABHM02000001.1	-	1430121	1430079	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-21-1	NZ_ABHO01000002.1	+	187909	187951	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-19-1	NZ_ABHP01000006.1	-	18431	18389	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-18-1	NZ_ABHQ01000004.1	-	168121	168079	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-23-1	NZ_ABHR01000006.1	-	107830	107788	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-24-1	NZ_ABHS01000006.1	-	168125	168083	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-25-1	NZ_ABHT01000008.1	+	84180	84222	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-27-1	NZ_ABHU01000007.1	-	98697	98655	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-26-1	NZ_ABHW01000006.1	-	80674	80632	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-15-1	NZ_ABJT01000003.1	-	189505	189463	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-29-1	NZ_ABKY01000010.1	+	76081	76123	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-31-1	NC_010498.1	-	2487091	2487049	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-30-1	NC_011415.1	-	2637288	2637246	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eta-1-1	NC_010694.1	+	1345590	1345632	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Cko-1-1	NC_009792.1	+	489806	489849	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Sgl-1-1	NC_007712.1	-	2743668	2743627	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eca-1-1	NZ_ABWM01000054.1	-	18086	18045	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Esp-1-1	NC_009436.1	-	3066128	3066087	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
env-4	BAAW01009584.1	-	396	355	**RNA**→ fer2 (cd00207)→
Sbo-2-1	NC_007613.1	-	2307033	2306991	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Sso-1-1	NC_007384.1	-	2471215	2471173	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→

Eco-11-1	NC_009801.1	-	2570657	2570615	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Sbo-1-1	NC_010658.1	-	2460767	2460725	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Sfl-3-1	NC_008258.1	-	2400881	2400839	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Sfl-2-1	NC_004337.1	-	2408328	2408286	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Sfl-1-1	NC_004741.1	-	2387785	2387743	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-7-1	NZ_AAJT02000015.1	+	21943	21985	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-10-1	NZ_AAJV02000003.1	-	174833	174791	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-9-1	NZ_AAJW02000006.1	-	119928	119886	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eco-6-1	NZ_AAJX02000006.1	-	168192	168150	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Eal-1-1	NZ_ABKX01000003.1	+	317786	317828	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Sdy-2-1	NC_007606.1	-	2279970	2279928	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Pst-1-1	NZ_ABJD02000101.1	-	1598670	1598628	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Pru-1-1	NZ_ABXV01000054.1	+	60557	60597	**RNA**→
Pal-1-1	NZ_ABXW01000062.1	+	94927	94967	**RNA**→
Pmi-1-1	NC_010554.1	-	1883702	1883663	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Yin-1-1	NZ_AALF01000077.1	+	4669	4754	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Esa-1-1	NC_009778.1	+	903670	903711	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Ybe-1-1	NZ_AALC01000035.1	+	13690	13774	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Pat-1-1	NC_004547.2	-	3380818	3380775	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→
Pca-2-1	NZ_ABVY01000005.1	-	55699	55656	**RNA**→
Pca-1-1	NZ_ABVX01000005.1	-	121832	121789	**RNA**→
Plu-1-1	NC_005126.1	-	3598173	3598131	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Ymo-1-1	NZ_AALD01000037.1	+	8930	9016	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Yps-1-1	NC_009708.1	+	1698464	1698506	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Ype-12-1	NC_005810.1	-	2629223	2629181	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Ype-1-1	NC_010159.1	-	1877225	1877183	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Ype-2-1	NC_008150.1	-	2292180	2292138	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Ype-6-1	NC_008149.1	-	2426958	2426916	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Ype-7-1	NC_009381.1	-	2215366	2215324	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Yps-4-1	NC_010465.1	+	1729846	1729888	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Ype-5-1	NC_004088.1	+	1807018	1807060	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→

Ype-4-1	NC_003143.1	-	2865144	2865102	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Ype-14-1	NZ_AAOS02000009.1	-	100594	100552	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Ype-10-1	NZ_AAYR01000001.1	+	1475516	1475558	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Ype-15-1	NZ_AAYS01000011.1	+	21539	21581	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Ype-11-1	NZ_AAYT01000003.1	-	37937	37895	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Ype-8-1	NZ_AAYU01000011.1	-	98561	98519	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Ype-9-1	NZ_AAYV01000001.1	-	512026	511984	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Ype-13-1	NZ_ABAT01000004.1	-	77358	77316	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Ype-3-1	NZ_ABCD01000001.1	+	785303	785345	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Yps-2-1	NC_006155.1	-	3056609	3056567	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Yps-3-1	NC_010634.1	-	2974507	2974465	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Yen-1-1	NC_008800.1	+	1502345	1502387	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→ NuoM (COG1008)→ NuoN (COG1007)→
Spr-1-1	NC_009832.1	-	3658544	3658502	**RNA**→ MopB_CT_NDH-1_NuoG2-N7 (cd02788)NuoG (COG1034)→ NuoH (COG1005)→ NuoI (COG1143)→ NuoJ (COG0839)→ NuoK (COG0713)→ NuoL (COG1009)→

58.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 58.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *nuoG* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

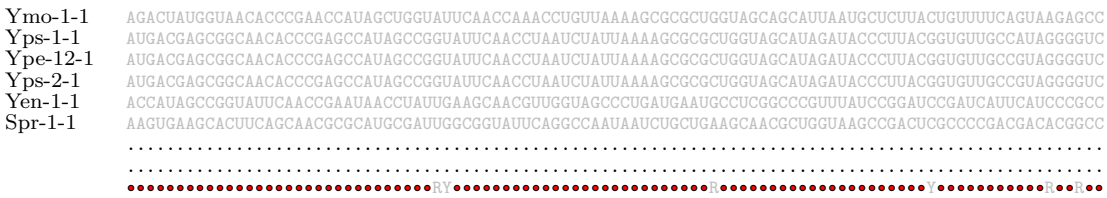
cd00207 (2) 2Fe-2S iron-sulfur cluster binding domain.
cd02788 (78) MopB_CT_NDH-1_NuoG2-N7: C-terminal region of the NuoG-like subunit (of the variant with a [4Fe-4S] cluster, N7) of the NADH-quinone oxidoreductase/NADH dehydrogenase-1 (NDH-1) found in various bacteria. The NDH-1 is the first energy-transducting complex in the respiratory chain and functions as a redox pump that uses the redox energy to translocate H+ ions across the membrane, resulting in a significant contribution to energy production. In Escherichia coli NDH-1, the largest subunit is encoded by the nuoG gene, and is part of the 14 distinct subunits constituting the functional enzyme. The NuoG subunit is made of two domains: the first contains three binding sites for FeS clusters (the fer2 domain), the second domain, is of unknown function or, as postulated, has lost an ancestral formate dehydrogenase activity that became redundant during the evolution of the complex I enzyme. Unique to this group, compared to the other prokaryotic and eukaryotic groups in this domain protein family (NADH-Q-OR-NuoG2), is an N-terminal [4Fe-4S]
COG0713 (78) NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) [Energy production and conversion]
COG0839 (78) NADH:ubiquinone oxidoreductase subunit 6 (chain J) [Energy production and conversion]
COG1005 (78) NADH:ubiquinone oxidoreductase subunit 1 (chain H) [Energy production and conversion]
COG1007 (26) NADH:ubiquinone oxidoreductase subunit 2 (chain N) [Energy production and conversion]
COG1008 (26) NADH:ubiquinone oxidoreductase subunit 4 (chain M) [Energy production and conversion]
COG1009 (78) NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na+/H+ antiporter, MnhA subunit [Energy production and conversion / Inorganic ion transport and metabolism]
COG1034 (79) NADH dehydrogenase/NADH:ubiquinone oxidoreductase 75 kD subunit (chain G) [Energy production and conversion]
COG1143 (78) Formate hydrogenlyase subunit 6/NADH:ubiquinone oxidoreductase 23 kD subunit (chain I) [Energy production and conversion]

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Eco-1-1,

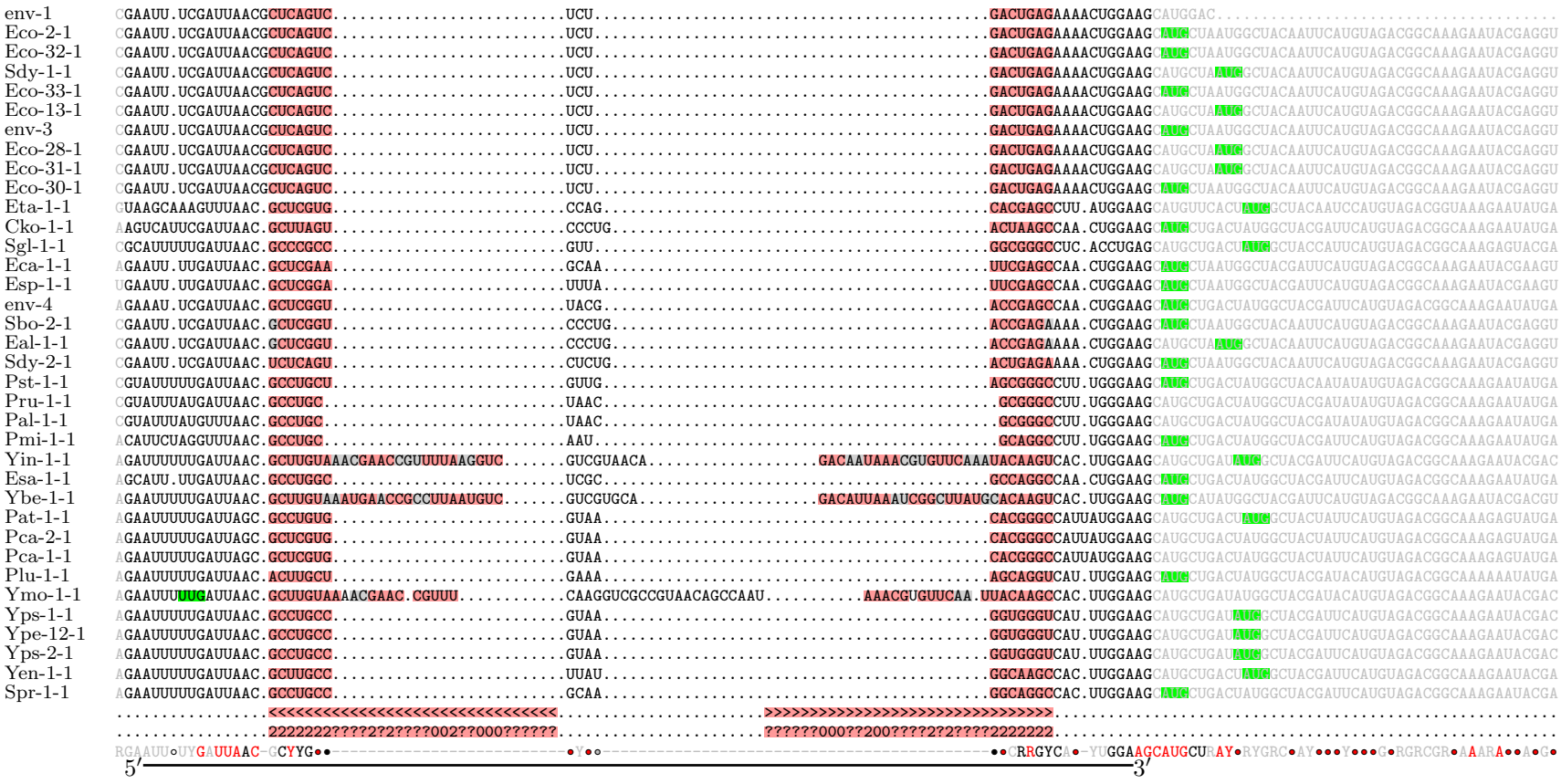
Eco-10-1, Eco-11-1, Eco-12-1, Eco-14-1, Eco-15-1, Eco-16-1, Eco-17-1, Eco-18-1, Eco-19-1, Eco-20-1, Eco-21-1, Eco-22-1, Eco-23-1, Eco-24-1, Eco-25-1, Eco-26-1, Eco-27-1, Eco-29-1, Eco-3-1, Eco-34-1, Eco-35-1, Eco-4-1, Eco-5-1, Eco-6-1, Eco-7-1, Eco-8-1, Eco-9-1, Sbo-1-1, Sfl-1-1, Sfl-2-1, Sfl-3-1, Sso-1-1, Ype-1-1, Ype-10-1, Ype-11-1, Ype-13-1, Ype-14-1, Ype-15-1, Ype-2-1, Ype-3-1, Ype-4-1, Ype-5-1, Ype-6-1, Ype-7-1, Ype-8-1, Ype-9-1, Yps-3-1, Yps-4-1, env-2

58.4 Multiple-sequence alignment

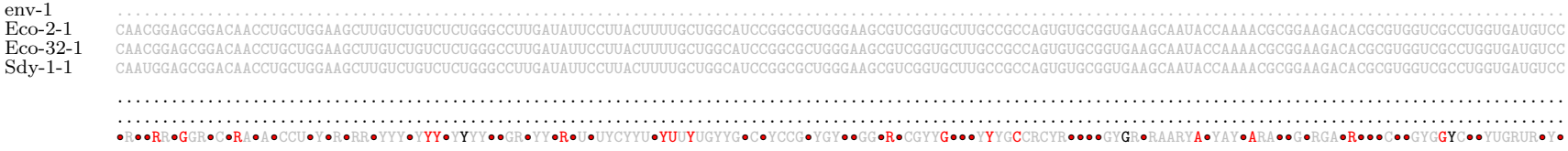
Each *nuoG* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 58.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *nuoG* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



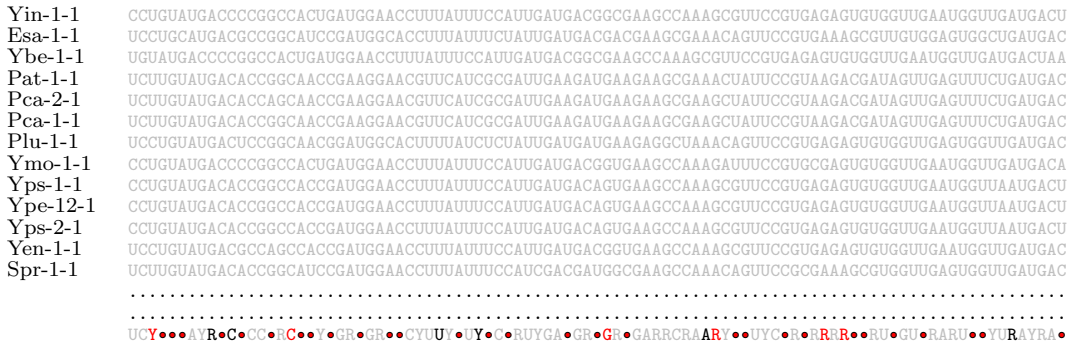
alignment positions 100 · · · 259



alignment positions 260 · · · 419



alignment positions 420 · · · 514



59 Ocean-V

59.1 Taxa

The taxonomy of each organism containing a putative Ocean-V RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Ocean-V RNA in Sections 59.2 and 59.4.

abbrev. of hits	taxonomy of species
env-1 to env-3	environmental samples

59.2 Gene contexts

Each Ocean-V RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Ocean-V RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 59.3. The accession number of the sequence containing each Ocean-V RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 59.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	ABEF01002814.1	+	681	753	hypo→ **RNA**→
env-2	ABEF01002040.1	-	258	187	hypo→ **RNA**→
env-3	ABEF01025170.1	-	751	680	hypo→ **RNA**→ ←COG1357 (COG1357)

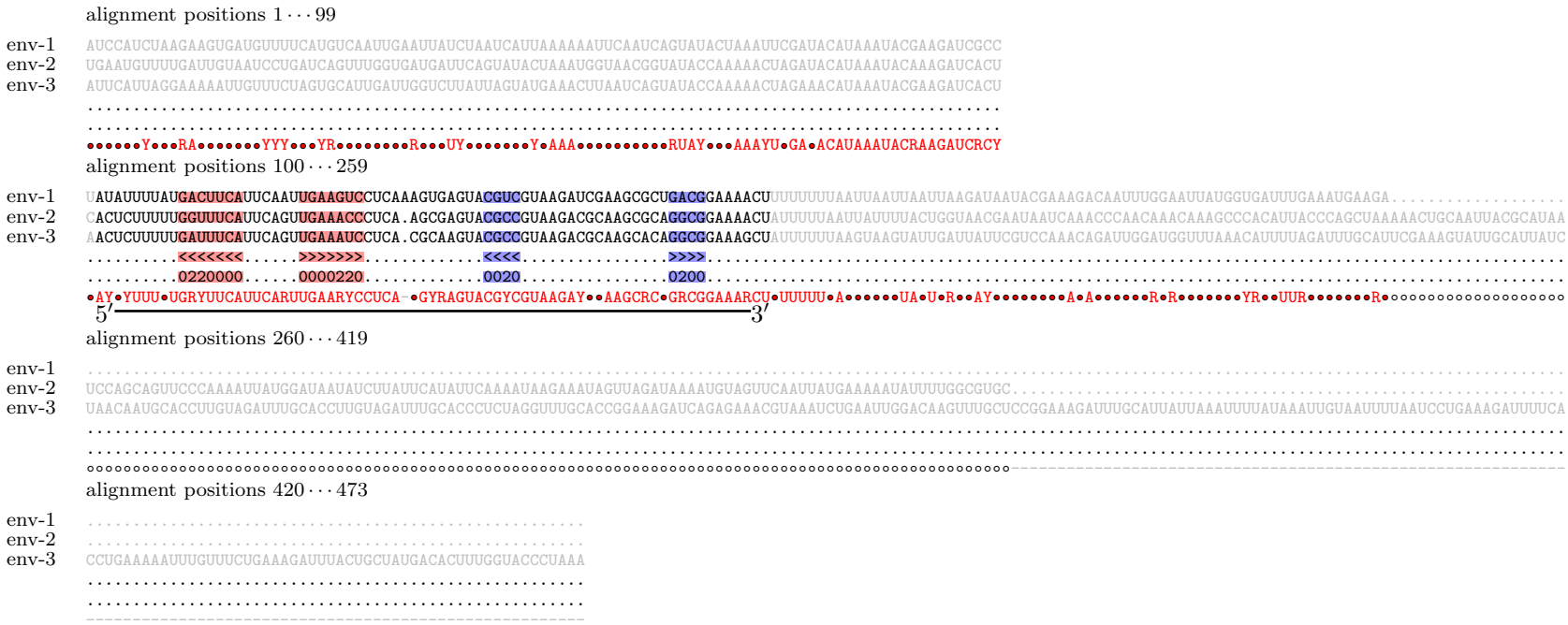
59.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 59.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Ocean-V RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

d COG1357 (1) Uncharacterized low-complexity proteins [Function unknown]

59.4 Multiple-sequence alignment

Each Ocean-V RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 59.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Ocean-V RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



60 Ocean-VI

60.1 Taxa

The taxonomy of each organism containing a putative Ocean-VI RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Ocean-VI RNA in Sections 60.2 and 60.4.

abbrev. of hits	taxonomy of species
env-1 to env-308	environmental samples

60.2 Gene contexts

Each Ocean-VI RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Ocean-VI RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 60.3. The accession number of the sequence containing each Ocean-VI RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 60.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	JCVI_SCAF_1101668022703	-	916	758	? (GOS10359)→ **RNA**→ CypX (COG2124)? (GOS3357)→
env-2	JCVI_SCAF_1096626021559	+	624	782	? (GOS10359)→ **RNA**→
env-3	JCVI_SCAF_1101668094393	+	431	589	? (GOS10359)→ **RNA**→ ←COG1247 (COG1247)? (GOS7913)
env-4	JCVI_SCAF_1096627379785	+	5155	5313	? (GOS10359)→ **RNA**→ ←COG1247 (COG1247)? (GOS7913)
env-5	JCVI_SCAF_1101667242610	-	417	260	? (GOS10359)→ **RNA**→
env-6	JCVI_SCAF_1096626894301	+	805	962	? (GOS10359)→ **RNA**→ ←COG1247 (COG1247)? (GOS7913)
env-7	JCVI_SCAF_1096627379785	+	5836	5994	←COG1247 (COG1247)? (GOS7913) **RNA**→ CypX (COG2124)? (GOS3357)→
env-8	JCVI_SCAF_1096626894301	+	1484	1641	←COG1247 (COG1247)? (GOS7913) **RNA**→ CypX (COG2124)? (GOS3357)→
env-9	JCVI_SCAF_1101667049214	-	566	409	←? (GOS7913) **RNA**→ ? (GOS3357)→
env-10	JCVI_SCAF_1101667256835	-	346	188	? (GOS10359)→ **RNA**→
env-11	JCVI_SCAF_1101668734638	+	106	261	**RNA**→ **RNA**→ Nfo (COG0648)? (GOS623)→
env-299	JCVI_SCAF_1101668734638	+	289	445	(shown above)
env-12	JCVI_SCAF_1096627154416	+	1933	2088	? (GOS10359)→ **RNA**→ **RNA**→ Nfo (COG0648)? (GOS623)→
env-245	JCVI_SCAF_1096627154416	+	2114	2271	(shown above)
env-17	JCVI_SCAF_1101668269783	+	788	943	? (GOS10359)→ **RNA**→ **RNA**→ Nfo (COG0648)? (GOS623)→
env-236	JCVI_SCAF_1101668269783	+	969	1126	(shown above)
env-19	JCVI_SCAF_1096626937785	+	271	428	? (GOS10359)→ **RNA**→ CypX (COG2124)? (GOS3357)→
env-20	JCVI_SCAF_1096627380505	+	3602	3759	? (GOS10359)→ **RNA**→ CypX (COG2124)? (GOS3357)→
env-21	JCVI_SCAF_1096628290566	-	542	385	? (GOS10359)→ **RNA**→ ? (GOS3357)→
env-22	JCVI_SCAF_1096626953993	-	761	604	? (GOS10359)→ **RNA**→ ? (GOS3357)→
env-23	JCVI_SCAF_1096626922112	-	221	64	? (GOS10359)→ **RNA**→
env-24	JCVI_SCAF_1101667103670	-	711	554	**RNA**→ ? (GOS3357)→
env-25	JCVI_SCAF_1096626958530	+	1417	1574	? (GOS10359)→ **RNA**→ ? (GOS3357)→
env-26	JCVI_SCAF_1096626952504	+	984	1141	? (GOS10359)→ **RNA**→

env-27	JCVI_SCAF_1096627051058	-	1899	1744	? (GOS10359)→ RNA→ CypX (COG2124)? (GOS3357)→
env-28	JCVI_SCAF_1101667150070	+	14	171	RNA→ ? (GOS3357)→
env-29	JCVI_SCAF_1101668309932	+	502	659	? (GOS10359)→ RNA→ CypX (COG2124)? (GOS3357)→
env-30	JCVI_SCAF_1101668672019	+	194	351	? (GOS10359)→ RNA→ ? (GOS3357)→
env-31	JCVI_SCAF_1096626984476	-	2163	2006	? (GOS10359)→ RNA→ CypX (COG2124)? (GOS3357)→
env-32	JCVI_SCAF_1096627258928	+	975	1134	? (GOS10359)→ RNA→ CypX (COG2124)? (GOS3357)→
env-33	JCVI_SCAF_1096628200096	-	685	528	? (GOS6677)→ RNA→ ? (GOS3357)→
env-34	JCVI_SCAF_1096628388632	+	1093	1250	←? (GOS670) RNA→ ? (GOS3357)→
env-35	JCVI_SCAF_1101667129577	+	526	683	? (GOS10359)→ RNA→
env-36	JCVI_SCAF_1101667135207	+	442	599	? (GOS10359)→ RNA→ ? (GOS3357)→
env-37	JCVI_SCAF_1101667136619	+	163	320	RNA→ ? (GOS3357)→
env-38	JCVI_SCAF_1101667151626	+	338	495	? (GOS10359)→ RNA→ ? (GOS3357)→
env-39	JCVI_SCAF_1096627856691	+	1067	1224	? (GOS10359)→ RNA→ ? (GOS3357)→
env-40	JCVI_SCAF_1101668177688	-	963	806	? (GOS10359)→ RNA→ CypX (COG2124)? (GOS3357)→
env-41	JCVI_SCAF_1101668271340	+	669	826	? (GOS10359)→ RNA→ ? (GOS3357)→
env-42	JCVI_SCAF_1096628399231	-	995	838	? (GOS10359)→ RNA→ CypX (COG2124)? (GOS3357)→
env-43	JCVI_SCAF_1096626930856	-	831	674	? (GOS10359)→ RNA→ CypX (COG2124)? (GOS3357)→
env-44	JCVI_SCAF_1096627052790	+	1936	2093	? (GOS10359)→ RNA→ ? (GOS3357)→
env-45	JCVI_SCAF_1096627701267	+	826	983	? (GOS10359)→ RNA→ ? (GOS3357)→
env-46	JCVI_SCAF_1096627686808	-	325	167	MetK (COG0192)? (GOS4376)→ RNA→
env-47	JCVI_SCAF_1096627180770	-	1011	853	MetK (COG0192)? (GOS4376)→ RNA→ 53EXOc (cd00008)? (GOS6979)→
env-48	JCVI_SCAF_1096627174199	-	1262	1104	? (GOS10359)→ RNA→ ←COG1247 (COG1247)? (GOS7913)
env-49	JCVI_SCAF_1101668542586	-	934	776	MET2 (COG2021)? (GOS3619)→ RNA→ 53EXOc (smart00475)? (GOS6979)→
env-50	JCVI_SCAF_1101668164062	+	130	288	RNA→ Tdh (COG1063)? (GOS333)→
env-51	JCVI_SCAF_1096627733888	+	688	843	? (GOS10359)→ RNA→
env-52	JCVI_SCAF_1096626957117	+	2686	2844	MET17 (COG2873)? (GOS3285)→ RNA→ 53EXOc (cd00008)? (GOS6979)→
env-53	JCVI_SCAF_1096627163044	-	419	261	MET17 (COG2873)? (GOS3285)→ RNA→ 5_3_exonuc_N (pfam02739)? (GOS6979)→
env-54	JCVI_SCAF_1101668690854	-	768	609	MET17 (COG2873)? (GOS3285)→ RNA→ 53EXOc (cd00008)? (GOS6979)→
env-55	JCVI_SCAF_1101668532346	+	387	545	? (GOS3285)→ RNA→ 53EXOc (smart00475)? (GOS6979)→
env-56	JCVI_SCAF_1101667235180	-	885	727	RNA→ ←? (GOS13057899)
env-57	JCVI_SCAF_1101668334939	-	1656	1498	RNA→ ←? (GOS13057899)
env-58	JCVI_SCAF_1096627998000	-	1412	1254	? (GOS3285)→ RNA→ 53EXOc (smart00475)? (GOS6979)→
env-59	JCVI_SCAF_1096627119051	+	1067	1225	? (GOS10359)→ RNA→ RNA→ ? (GOS11273709)→
env-228	JCVI_SCAF_1096627119051	+	1255	1410	(shown above)
env-61	JCVI_SCAF_1096627130271	-	2383	2225	MET17 (COG2873)? (GOS3285)→ RNA→ 53EXOc (smart00475)? (GOS6979)→
env-62	JCVI_SCAF_1096627323041	-	1619	1461	MET17 (COG2873)? (GOS3285)→ RNA→ 53EXOc (cd00008)? (GOS6979)→
env-63	JCVI_SCAF_1096628011377	+	600	758	? (GOS10359)→ RNA→ Nfo (COG0648)? (GOS623)→
env-64	JCVI_SCAF_1101668328517	-	907	749	? (GOS10359)→ RNA→ RNA→ Nfo (COG0648)? (GOS623)→
env-294	JCVI_SCAF_1101668328517	-	721	566	(shown above)
env-65	JCVI_SCAF_1096627030455	-	2781	2623	RNA→ 53EXOc (cd00008)? (GOS6979)→
env-66	JCVI_SCAF_1096627225102	+	801	959	? (GOS10359)→ RNA→ RNA→ ? (GOS13557689)→
env-253	JCVI_SCAF_1096627225102	+	1013	1170	(shown above)
env-68	JCVI_SCAF_1101668263088	+	1376	1534	? (GOS10359)→ RNA→
env-69	JCVI_SCAF_1096628027439	-	1647	1489	RNA→ ←COG1247 (COG1247)? (GOS7913)
env-70	JCVI_SCAF_1096626665462	+	1	159	RNA→ RNA→ AP2Ec (smart00518)? (GOS623)→
env-272	JCVI_SCAF_1096626665462	+	216	373	(shown above)
env-71	JCVI_SCAF_1101668669207	-	1200	1042	? (GOS11574206)→ RNA→ RNA→ Nfo (COG0648)? (GOS623)→
env-274	JCVI_SCAF_1101668669207	-	985	828	(shown above)
env-72	JCVI_SCAF_1096627026801	+	723	881	? (GOS11574206)→ RNA→
env-73	JCVI_SCAF_1096627078692	+	425	583	? (GOS11574206)→ RNA→ RNA→ Nfo (COG0648)? (GOS623)→
env-275	JCVI_SCAF_1096627078692	+	640	797	(shown above)
env-74	JCVI_SCAF_1096627312287	+	962	1120	? (GOS11574206)→ RNA→ RNA→ ? (GOS13077636)→

env-264	JCVI_SCAF_1096627312287	+	1149	1306	(shown above)
env-75	JCVI_SCAF_1096626914871	+	957	1115	? (GOS11574206)→ **RNA**→ ←COG1247 (COG1247)? (GOS7913)
env-76	JCVI_SCAF_1096627856069	+	1141	1299	? (GOS11574206)→ **RNA**→
env-78	JCVI_SCAF_1096627385161	-	3249	3091	**RNA**→ ←COG1247 (COG1247)? (GOS7913)
env-79	JCVI_SCAF_1096627099539	+	4705	4860	? (GOS11574206)→ **RNA**→
env-80	JCVI_SCAF_1096627177870	-	1289	1131	? (GOS11574206)→ **RNA**→ ←COG1247 (COG1247)? (GOS7913)
env-81	JCVI_SCAF_1096627174996	+	890	1048	? (GOS11574206)→ **RNA**→ ←COG1247 (COG1247)? (GOS7913)
env-82	JCVI_SCAF_1096628023873	+	879	1037	? (GOS10359)→ **RNA**→ ←COG1247 (COG1247)? (GOS7913)
env-83	JCVI_SCAF_1096628033048	-	1566	1408	? (GOS11574206)→ **RNA**→ ←COG1247 (COG1247)? (GOS7913)
env-84	JCVI_SCAF_1096627034149	+	501	659	? (GOS11574206)→ **RNA**→ **RNA**→ ? (GOS11057767)→
env-267	JCVI_SCAF_1096627034149	+	715	872	(shown above)
env-85	JCVI_SCAF_1101668439724	-	222	64	? (GOS11574206)→ **RNA**→
env-87	JCVI_SCAF_1096627514636	+	154	312	**RNA**→ ←COG1247 (COG1247)? (GOS7913)
env-88	JCVI_SCAF_1101667084868	-	467	309	? (GOS11574206)→ **RNA**→ ←COG1247 (COG1247)? (GOS7913)
env-89	JCVI_SCAF_1096628018580	-	365	207	? (GOS11528218)→ **RNA**→ ←COG1247 (COG1247)? (GOS7913)
env-90	JCVI_SCAF_1096628019946	+	1196	1354	? (GOS11574206)→ **RNA**→ ←COG1247 (COG1247)? (GOS7913)
env-91	JCVI_SCAF_1096626418595	-	254	96	? (GOS11574206)→ **RNA**→
env-92	JCVI_SCAF_1101668286173	+	344	502	? (GOS11574206)→ **RNA**→ **RNA**→ ? (GOS13557689)→
env-260	JCVI_SCAF_1101668286173	+	559	716	(shown above)
env-94	JCVI_SCAF_1096627961062	-	1254	1096	? (GOS3285)→ **RNA**→ ←? (GOS13057899)
env-95	JCVI_SCAF_1101667045594	+	569	727	MET2 (COG2021)? (GOS3619)→ **RNA**→
env-96	JCVI_SCAF_1096626889960	-	1466	1308	MET2 (COG2021)? (GOS3619)→ **RNA**→ 53EXOc (smart00475)? (GOS6979)→
env-97	JCVI_SCAF_1096627296996	+	337	495	? (GOS10359)→ **RNA**→ ←COG1247 (COG1247)? (GOS7913)
env-98	JCVI_SCAF_1096627547445	-	1432	1274	**RNA**→ ←COG1247 (COG1247)? (GOS7913)
env-99	JCVI_SCAF_1096626917744	+	742	900	←? (GOS3874) **RNA**→ 53EXOc (smart00475)? (GOS6979)→
env-100	JCVI_SCAF_1101668288696	+	697	855	? (GOS10359)→ **RNA**→ **RNA**→ Nfo (COG0648)? (GOS623)→
env-235	JCVI_SCAF_1101668288696	+	881	1038	(shown above)
env-101	JCVI_SCAF_1101668347478	+	934	1092	←? (GOS3874) **RNA**→
env-102	JCVI_SCAF_1096627175966	-	1684	1526	? (GOS10359)→ **RNA**→ ←COG1247 (COG1247)? (GOS7913)
env-103	JCVI_SCAF_1096627500144	+	360	518	? (GOS3619)→ **RNA**→ 5_3_exonuc_N (pfam02739)? (GOS6979)→
env-105	JCVI_SCAF_1096627317949	+	393	551	? (GOS10359)→ **RNA**→ **RNA**→ Nfo (COG0648)? (GOS623)→
env-292	JCVI_SCAF_1096627317949	+	579	734	(shown above)
env-106	JCVI_SCAF_1096627371151	-	994	836	? (GOS10359)→ **RNA**→ **RNA**→ Nfo (COG0648)? (GOS623)→
env-250	JCVI_SCAF_1096627371151	-	810	653	(shown above)
env-107	JCVI_SCAF_1096627696463	-	1803	1645	**RNA**→ **RNA**→ ? (GOS623)→
env-252	JCVI_SCAF_1096627696463	-	1619	1462	(shown above)
env-108	JCVI_SCAF_1096627304742	-	401	243	? (GOS10359)→ **RNA**→ ←? (GOS13057899)
env-109	JCVI_SCAF_1096627994896	+	211	369	? (GOS10359)→ **RNA**→ **RNA**→ ? (GOS12085149)→
env-249	JCVI_SCAF_1096627994896	+	395	552	(shown above)
env-110	JCVI_SCAF_1096627544027	-	1305	1147	? (GOS10359)→ **RNA**→ **RNA**→ Nfo (COG0648)? (GOS623)→
env-290	JCVI_SCAF_1096627544027	-	1119	964	(shown above)
env-111	JCVI_SCAF_1096627329633	-	287	131	? (GOS10359)→ **RNA**→
env-112	JCVI_SCAF_1096626974071	-	244	86	←? (GOS3874) **RNA**→
env-113	JCVI_SCAF_1096627085809	+	1193	1351	←? (GOS3874) **RNA**→ 53EXOc (smart00475)? (GOS6979)→
env-114	JCVI_SCAF_1096626924876	-	1118	961	MET2 (COG2021)? (GOS3619)→ **RNA**→ 53EXOc (smart00475)? (GOS6979)→
env-115	JCVI_SCAF_1101668330247	+	1049	1207	MET17 (COG2873)? (GOS3285)→ **RNA**→ 53EXOc (cd00008)? (GOS6979)→
env-116	JCVI_SCAF_1096626219711	-	775	617	? (GOS11574206)→ **RNA**→ **RNA**→ ? (GOS623)→
env-293	JCVI_SCAF_1096626219711	-	589	434	(shown above)
env-117	JCVI_SCAF_1096627827704	-	1145	987	? (GOS3285)→ **RNA**→ 5_3_exonuc_N (pfam02739)? (GOS6979)→
env-118	JCVI_SCAF_1096627854582	-	1417	1259	**RNA**→ 5_3_exonuc_N (pfam02739)? (GOS6979)→
env-119	JCVI_SCAF_1101668665373	+	1070	1228	MET17 (COG2873)? (GOS3285)→ **RNA**→ 5_3_exonuc_N (pfam02739)? (GOS6979)→
env-120	JCVI_SCAF_1096627263379	-	1219	1061	MET17 (COG2873)? (GOS3285)→ **RNA**→ 53EXOc (cd00008)? (GOS6979)→

env-121	JCVI_SCAF_1096627306307	-	233	75	MET17 (COG2873)? (GOS3285)→ **RNA**→
env-122	JCVI_SCAF_1096627598829	+	638	796	? (GOS10359)→ **RNA**→ Nfo (COG0648)? (GOS623)→
env-123	JCVI_SCAF_1096627179664	+	653	811	←? (GOS3874) **RNA**→ 53EXOc (smart00475)? (GOS6979)→
env-124	JCVI_SCAF_1096627323562	-	202	44	MET2 (COG2021)? (GOS3619)→ **RNA**→
env-125	JCVI_SCAF_1096626914748	-	425	267	MET2 (COG2021)? (GOS3619)→ **RNA**→ 5_3_exonuc_N (pfam02739)? (GOS6979)→
env-126	JCVI_SCAF_1096628205814	+	506	662	? (GOS10359)→ **RNA**→ **RNA**→ Nfo (COG0648)? (GOS623)→
env-269	JCVI_SCAF_1096628205814	+	712	869	(shown above)
env-127	JCVI_SCAF_1101668355515	-	516	358	? (GOS10359)→ **RNA**→ ? (GOS11452072)→
env-128	JCVI_SCAF_1096627085167	-	1375	1217	? (GOS10359)→ **RNA**→ ? (GOS11452072)→
env-129	JCVI_SCAF_1096627097745	-	2122	1964	? (GOS10359)→ **RNA**→ Tdh (COG1063)? (GOS333)→
env-130	JCVI_SCAF_1101668706160	+	666	824	MET17 (COG2873)? (GOS3285)→ **RNA**→ 53EXOc (cd00008)? (GOS6979)→
env-131	JCVI_SCAF_1096627089652	-	1256	1098	MET2 (COG2021)? (GOS3619)→ **RNA**→ 53EXOc (smart00475)? (GOS6979)→
env-132	JCVI_SCAF_1096627862656	-	488	330	MET2 (COG2021)? (GOS3619)→ **RNA**→ 53EXOc (cd00008)? (GOS6979)→
env-133	JCVI_SCAF_1101668659723	+	979	1137	MET2 (COG2021)? (GOS3619)→ **RNA**→ ←? (GOS13057899)
env-134	JCVI_SCAF_1096628029665	-	805	647	**RNA**→ ←? (GOS13057899)
env-135	JCVI_SCAF_1096627044366	-	1911	1753	? (GOS10359)→ **RNA**→ **RNA**→ Nfo (COG0648)? (GOS623)→
env-270	JCVI_SCAF_1096627044366	-	1697	1540	(shown above)
env-136	JCVI_SCAF_1096627598378	-	1621	1463	? (GOS10359)→ **RNA**→ **RNA**→ Nfo (COG0648)? (GOS623)→
env-257	JCVI_SCAF_1096627598378	-	1406	1249	(shown above)
env-137	JCVI_SCAF_1101667114249	-	607	449	**RNA**→
env-138	JCVI_SCAF_1096626984528	-	1549	1391	**RNA**→ **RNA**→ ? (GOS11057767)→
env-263	JCVI_SCAF_1096626984528	-	1334	1177	(shown above)
env-140	JCVI_SCAF_1096627041458	+	191	349	**RNA**→ **RNA**→ Nfo (COG0648)? (GOS623)→
env-271	JCVI_SCAF_1096627041458	+	404	561	(shown above)
env-141	JCVI_SCAF_1101668282956	-	602	444	MET2 (COG2021)? (GOS3619)→ **RNA**→ 53EXOc (cd00008)? (GOS6979)→
env-142	JCVI_SCAF_1101668275397	-	1163	1005	**RNA**→ **RNA**→ Nfo (COG0648)? (GOS623)→
env-258	JCVI_SCAF_1101668275397	-	948	791	(shown above)
env-143	JCVI_SCAF_1101668691255	+	987	1145	MET2 (COG2021)? (GOS3619)→ **RNA**→ ←? (GOS13057899)
env-144	JCVI_SCAF_1096626147351	-	243	85	MET17 (COG2873)? (GOS3285)→ **RNA**→
env-145	JCVI_SCAF_1096627601418	-	718	560	? (GOS3619)→ **RNA**→ 53EXOc (cd00008)? (GOS6979)→
env-146	JCVI_SCAF_1096627553258	+	414	572	? (GOS3619)→ **RNA**→ ←? (GOS13057899)
env-147	JCVI_SCAF_1096628273620	+	129	287	**RNA**→ 53EXOc (cd00008)? (GOS6979)→
env-148	JCVI_SCAF_1096628032488	-	657	499	MET2 (COG2021)? (GOS3619)→ **RNA**→ ←? (GOS13057899)
env-149	JCVI_SCAF_1096627814143	-	244	86	MET2 (COG2021)? (GOS3619)→ **RNA**→ ←? (GOS13057899)
env-150	JCVI_SCAF_1101668168496	+	251	409	? (GOS3619)→ **RNA**→ 53EXOc (smart00475)? (GOS6979)→
env-151	JCVI_SCAF_1096627089463	+	481	639	? (GOS3619)→ **RNA**→ ←? (GOS13057899)
env-152	JCVI_SCAF_1096627087692	-	597	439	MET2 (COG2021)? (GOS3619)→ **RNA**→ ←? (GOS13057899)
env-153	JCVI_SCAF_1096627067668	-	3788	3630	? (GOS3619)→ **RNA**→ ←? (GOS13057899)
env-154	JCVI_SCAF_1101667075568	+	158	316	**RNA**→ ←? (GOS13057899)
env-155	JCVI_SCAF_1101668765051	-	656	498	**RNA**→ ←? (GOS13057899)
env-156	JCVI_SCAF_1096627071800	-	827	669	MET2 (COG2021)? (GOS3619)→ **RNA**→ 53EXOc (cd00008)? (GOS6979)→
env-157	JCVI_SCAF_1096627552814	-	1556	1398	**RNA**→ ←? (GOS13057899)
env-158	JCVI_SCAF_1096627178922	-	1849	1691	**RNA**→ ←? (GOS13057899)
env-159	JCVI_SCAF_1096627309562	+	1314	1472	MET2 (COG2021)? (GOS3619)→ **RNA**→ ←? (GOS13057899)
env-160	JCVI_SCAF_1101667275442	+	740	898	MET2 (COG2021)? (GOS3619)→ **RNA**→ ←? (GOS13057899)
env-161	JCVI_SCAF_1096628030576	+	44	202	**RNA**→ ←? (GOS13057899)
env-162	JCVI_SCAF_1101668201157	-	183	25	? (GOS10359)→ **RNA**→
env-163	JCVI_SCAF_1101667143321	+	422	579	? (GOS10359)→ **RNA**→
env-164	JCVI_SCAF_1096628037417	-	306	148	MET2 (COG2021)? (GOS3619)→ **RNA**→ ←? (GOS13057899)
env-165	JCVI_SCAF_1101667059383	-	259	103	? (GOS10359)→ **RNA**→ ←? (GOS13057899)
env-166	JCVI_SCAF_1096627321686	+	563	721	? (GOS11574206)→ **RNA**→ **RNA**→ Nfo (COG0648)? (GOS623)→
env-237	JCVI_SCAF_1096627321686	+	747	904	(shown above)

env-167	JCVI_SCAF_1096627171749	-	2621	2463	MET17 (COG2873)? (GOS3285)→ **RNA**→ 53EXOc (smart00475)? (GOS6979)→
env-168	JCVI_SCAF_1096628002503	-	856	698	←GpmB (COG0406)? (GOS670) **RNA**→ **RNA**→ Nfo (COG0648)? (GOS623)→
env-230	JCVI_SCAF_1096628002503	-	652	495	(shown above)
env-169	JCVI_SCAF_1096627389052	-	1016	858	? (GOS10359)→ **RNA**→ CypX (COG2124)? (GOS3357)→
env-170	JCVI_SCAF_1096627207889	+	1673	1831	? (GOS10359)→ **RNA**→
env-171	JCVI_SCAF_1101667417475	+	52	210	**RNA**→ p450 (pfam00067)? (GOS3357)→
env-172	JCVI_SCAF_1101668750253	-	431	273	? (GOS10359)→ **RNA**→ ? (GOS3357)→
env-173	JCVI_SCAF_1096626915383	-	1658	1502	**RNA**→ **RNA**→ Nfo (COG0648)? (GOS623)→
env-301	JCVI_SCAF_1096626915383	-	1474	1319	(shown above)
env-174	JCVI_SCAF_1096626979050	-	1422	1266	? (GOS11528218)→ **RNA**→ **RNA**→ Nfo (COG0648)? (GOS623)→
env-302	JCVI_SCAF_1096626979050	-	1237	1082	(shown above)
env-175	JCVI_SCAF_1096626955441	-	154	1	COG4912 (COG4912)? (GOS11003827)→ **RNA**→
env-176	JCVI_SCAF_1096627330894	-	1209	1053	COG4912 (COG4912)? (GOS11003827)→ **RNA**→ CypX (COG2124)? (GOS3357)→
env-177	JCVI_SCAF_1096627117257	-	2122	1966	**RNA**→ CypX (COG2124)? (GOS3357)→
env-178	JCVI_SCAF_1096627345659	-	777	621	COG4912 (COG4912)? (GOS11003827)→ **RNA**→ CypX (COG2124)? (GOS3357)→
env-179	JCVI_SCAF_1101668489426	+	866	1022	COG4912 (COG4912)? (GOS11003827)→ **RNA**→ ? (GOS3357)→
env-180	JCVI_SCAF_1096627128747	+	740	896	COG4912 (COG4912)? (GOS11003827)→ **RNA**→ CypX (COG2124)? (GOS3357)→
env-181	JCVI_SCAF_1101668473248	+	271	427	COG4912 (COG4912)? (GOS11003827)→ **RNA**→ CypX (COG2124)? (GOS3357)→
env-182	JCVI_SCAF_1096626955532	-	1435	1279	COG4912 (COG4912)? (GOS11003827)→ **RNA**→ CypX (COG2124)? (GOS3357)→
env-183	JCVI_SCAF_1096627114189	-	1985	1829	COG4912 (COG4912)? (GOS11003827)→ **RNA**→ CypX (COG2124)? (GOS3357)→
env-184	JCVI_SCAF_1096627390438	+	21426	21582	COG4912 (COG4912)? (GOS11003827)→ **RNA**→ CypX (COG2124)? (GOS3357)→
env-185	JCVI_SCAF_1096627383440	+	1288	1444	COG4912 (COG4912)? (GOS11003827)→ **RNA**→ CypX (COG2124)? (GOS3357)→
env-186	JCVI_SCAF_1096626917624	-	1452	1296	? (GOS11003827)→ **RNA**→ CypX (COG2124)? (GOS3357)→
env-187	JCVI_SCAF_1101668191992	-	1430	1274	**RNA**→ CypX (COG2124)? (GOS3357)→
env-188	JCVI_SCAF_1096627117257	-	653	496	CypX (COG2124)? (GOS3357)→ **RNA**→ ? (GOS11048107)→
env-189	JCVI_SCAF_1101667285953	+	70	228	**RNA**→ ? (GOS11016378)→
env-190	JCVI_SCAF_1096627128747	+	2210	2368	CypX (COG2124)? (GOS3357)→ **RNA**→ ? (GOS11016378)→
env-191	JCVI_SCAF_1096627125335	-	756	600	CypX (COG2124)? (GOS3357)→ **RNA**→ ? (GOS12009452)→
env-192	JCVI_SCAF_1096626467676	-	622	464	? (GOS3357)→ **RNA**→ ? (GOS13372899)→
env-193	JCVI_SCAF_1101668746994	+	272	430	? (GOS3357)→ **RNA**→ ? (GOS13372899)→
env-194	JCVI_SCAF_1096626928341	+	81	239	**RNA**→ ←? (GOS13372499)
env-195	JCVI_SCAF_1096626995814	-	382	224	CypX (COG2124)? (GOS3357)→ **RNA**→
env-196	JCVI_SCAF_1096627383440	+	2754	2912	CypX (COG2124)? (GOS3357)→ **RNA**→ ←? (GOS13372499)
env-197	JCVI_SCAF_1096627114189	-	517	359	CypX (COG2124)? (GOS3357)→ **RNA**→ ? (GOS12009452)→
env-198	JCVI_SCAF_1096627187175	-	581	425	? (GOS3357)→ **RNA**→ ? (GOS623)→
env-199	JCVI_SCAF_1096626597251	-	262	104	CypX (COG2124)? (GOS3357)→ **RNA**→
env-200	JCVI_SCAF_1096627378495	-	2325	2167	CypX (COG2124)? (GOS3357)→ **RNA**→ ? (GOS11016378)→
env-201	JCVI_SCAF_1101667485944	-	745	587	? (GOS3357)→ **RNA**→ ? (GOS11048107)→
env-202	JCVI_SCAF_1096628203042	-	1377	1219	**RNA**→ ? (GOS11048107)→
env-203	JCVI_SCAF_1101667104153	+	584	742	CypX (COG2124)? (GOS3357)→ **RNA**→
env-204	JCVI_SCAF_1096626968305	+	594	752	CypX (COG2124)? (GOS3357)→ **RNA**→ ? (GOS11016378)→
env-205	JCVI_SCAF_1096626438244	+	323	481	? (GOS3357)→ **RNA**→ ? (GOS13077636)→
env-206	JCVI_SCAF_1096626974722	+	1419	1577	CypX (COG2124)? (GOS3357)→ **RNA**→ ? (GOS11016378)→
env-207	JCVI_SCAF_1096626438521	-	558	400	? (GOS3357)→ **RNA**→ ? (GOS623)→
env-208	JCVI_SCAF_1101667145458	-	534	376	? (GOS3357)→ **RNA**→ ? (GOS11016378)→
env-209	JCVI_SCAF_1096627380505	+	5069	5227	CypX (COG2124)? (GOS3357)→ **RNA**→
env-210	JCVI_SCAF_1096627004813	-	1214	1056	? (GOS3357)→ **RNA**→ ? (GOS11057767)→
env-211	JCVI_SCAF_1096626726379	-	514	356	? (GOS3357)→ **RNA**→ ? (GOS623)→
env-212	JCVI_SCAF_1096627049841	-	985	827	CypX (COG2124)? (GOS3357)→ **RNA**→ AP2Ec (smart00518)? (GOS623)→
env-213	JCVI_SCAF_1096626938145	-	1562	1404	CypX (COG2124)? (GOS3357)→ **RNA**→ ? (GOS11016378)→
env-214	JCVI_SCAF_1096626984476	-	692	534	CypX (COG2124)? (GOS3357)→ **RNA**→ ? (GOS11048107)→
env-215	JCVI_SCAF_1096628386287	+	197	355	? (GOS3357)→ **RNA**→ ? (GOS11048107)→

env-216	JCVI_SCAF_1096626956850	+	963	1121	CypX (COG2124)? (GOS3357)→ **RNA**→ ? (GOS11048107)→
env-217	JCVI_SCAF_1096627731141	-	195	37	CypX (COG2124)? (GOS3357)→ **RNA**→
env-218	JCVI_SCAF_1101668737719	-	1252	1094	? (GOS3357)→ **RNA**→ ? (GOS14670336)→
env-219	JCVI_SCAF_1096627051058	-	430	272	CypX (COG2124)? (GOS3357)→ **RNA**→ ? (GOS11016378)→
env-220	JCVI_SCAF_1096627296996	+	1007	1160	←COG1247 (COG1247)? (GOS7913) **RNA**→ ←? (GOS11295005)
env-221	JCVI_SCAF_1096627073520	-	178	24	? (GOS10359)→ **RNA**→
env-60	JCVI_SCAF_1096627073520	-	367	209	(shown above)
env-222	JCVI_SCAF_1096627098763	+	100	254	**RNA**→ Nfo (COG0648)? (GOS623)→
env-223	JCVI_SCAF_1096626279136	-	824	670	**RNA**→ AP2Ec (smart00518)? (GOS623)→
env-224	JCVI_SCAF_1096627459803	-	624	468	←COG1247 (COG1247)? (GOS7913) **RNA**→ ←? (GOS11295005)
env-225	JCVI_SCAF_1101668710949	-	1644	1487	**RNA**→ ? (GOS11057767)→
env-226	JCVI_SCAF_1096627085167	-	975	818	←? (GOS13303413) **RNA**→ ? (GOS11057767)→
env-227	JCVI_SCAF_1096627117138	-	1897	1740	**RNA**→ ? (GOS11057767)→
env-229	JCVI_SCAF_1096627475504	+	969	1124	BioF (COG0156)? (GOS869)→ **RNA**→ Nfo (COG0648)? (GOS623)→
env-231	JCVI_SCAF_1101667235180	-	692	535	←? (GOS13057899) **RNA**→ ? (GOS11176326)→
env-232	JCVI_SCAF_1101668334939	-	1463	1306	←? (GOS13057899) **RNA**→ ? (GOS11176326)→
env-233	JCVI_SCAF_1096627866583	+	66	223	**RNA**→ ? (GOS11176326)→
env-234	JCVI_SCAF_1096627175966	-	1005	848	←COG1247 (COG1247)? (GOS7913) **RNA**→ ? (GOS11176326)→
env-238	JCVI_SCAF_1101667257400	-	799	642	**RNA**→ Nfo (COG0648)? (GOS623)→
env-239	JCVI_SCAF_1096627163927	-	382	225	? (GOS10359)→ **RNA**→
env-14	JCVI_SCAF_1096627163927	-	563	408	(shown above)
env-240	JCVI_SCAF_1101667252126	+	393	550	? (GOS10359)→ **RNA**→
env-15	JCVI_SCAF_1101667252126	+	212	367	(shown above)
env-241	JCVI_SCAF_1096627590455	+	1432	1589	? (GOS10359)→ **RNA**→
env-13	JCVI_SCAF_1096627590455	+	1251	1406	(shown above)
env-242	JCVI_SCAF_1096627032172	+	125	282	**RNA**→ Nfo (COG0648)? (GOS623)→
env-243	JCVI_SCAF_1096627048162	-	171	14	? (GOS10359)→ **RNA**→
env-16	JCVI_SCAF_1096627048162	-	352	197	(shown above)
env-244	JCVI_SCAF_1096627064429	-	479	322	? (GOS10359)→ **RNA**→
env-18	JCVI_SCAF_1096627064429	-	660	505	(shown above)
env-246	JCVI_SCAF_1096627984317	-	1799	1642	**RNA**→ Nfo (COG0648)? (GOS623)→
env-247	JCVI_SCAF_1096627304742	-	217	60	←? (GOS13057899) **RNA**→
env-248	JCVI_SCAF_1096626980093	-	2725	2568	**RNA**→ Nfo (COG0648)? (GOS623)→
env-251	JCVI_SCAF_1096627694029	-	1623	1466	**RNA**→ Nfo (COG0648)? (GOS623)→
env-254	JCVI_SCAF_1101668168297	+	359	516	? (GOS11302951)→ **RNA**→ **RNA**→ ? (GOS11302951)→
env-77	JCVI_SCAF_1101668168297	+	144	302	(shown above)
env-255	JCVI_SCAF_1101667072392	-	271	114	←COG1247 (COG1247)? (GOS7913) **RNA**→
env-256	JCVI_SCAF_1096626914871	+	1625	1782	←COG1247 (COG1247)? (GOS7913) **RNA**→
env-259	JCVI_SCAF_1096627174996	+	1558	1715	←COG1247 (COG1247)? (GOS7913) **RNA**→ ? (GOS13077636)→
env-261	JCVI_SCAF_1096627502744	-	1681	1527	**RNA**→ Nfo (COG0648)? (GOS623)→
env-262	JCVI_SCAF_1101667114249	-	392	235	**RNA**→
env-265	JCVI_SCAF_1101667294195	+	19	176	**RNA**→
env-266	JCVI_SCAF_1101668338801	+	112	269	**RNA**→ ? (GOS11057767)→
env-268	JCVI_SCAF_1096627320519	-	977	820	? (GOS11574206)→ **RNA**→
env-86	JCVI_SCAF_1096627320519	-	1191	1033	(shown above)
env-273	JCVI_SCAF_1096627687017	+	86	243	**RNA**→ Nfo (COG0648)? (GOS623)→
env-276	JCVI_SCAF_1101667188900	-	832	675	**RNA**→ Nfo (COG0648)? (GOS623)→
env-277	JCVI_SCAF_1101668185452	-	464	307	? (GOS11302951)→ **RNA**→ **RNA**→
env-93	JCVI_SCAF_1101668185452	-	679	521	(shown above)
env-278	JCVI_SCAF_1096627037261	+	1366	1523	? (GOS11016378)→ **RNA**→
env-279	JCVI_SCAF_1101667356111	-	798	641	**RNA**→ ? (GOS13077636)→
env-280	JCVI_SCAF_1101667357902	-	529	372	←COG1247 (COG1247)? (GOS7913) **RNA**→ ? (GOS13077636)→

env-281	JCVI_SCAF_1096627385161	-	2581	2424	←COG1247 (COG1247)? (GOS7913) **RNA**→ ? (GOS13077636)→
env-282	JCVI_SCAF_1096627177870	-	621	464	←COG1247 (COG1247)? (GOS7913) **RNA**→ ? (GOS13077636)→
env-283	JCVI_SCAF_1101668174371	+	706	863	? (GOS11016378)→ **RNA**→ ? (GOS14670336)→
env-284	JCVI_SCAF_1101667354916	-	421	264	? (GOS11016378)→ **RNA**→
env-285	JCVI_SCAF_1101667357269	+	647	804	←COG1247 (COG1247)? (GOS7913) **RNA**→
env-286	JCVI_SCAF_1096626993603	+	660	817	←COG1247 (COG1247)? (GOS7913) **RNA**→ Nfo (COG0648)? (GOS623)→
env-287	JCVI_SCAF_1101668168636	-	350	193	? (GOS10359)→ **RNA**→
env-67	JCVI_SCAF_1101668168636	-	564	406	(shown above)
env-288	JCVI_SCAF_1096628135575	-	459	304	←? (GOS10117) **RNA**→
env-139	JCVI_SCAF_1096628135575	-	665	507	(shown above)
env-289	JCVI_SCAF_1101667084060	+	49	205	**RNA**→ ? (GOS13557689)→
env-291	JCVI_SCAF_1101668457312	+	607	762	? (GOS10359)→ **RNA**→
env-104	JCVI_SCAF_1101668457312	+	421	579	(shown above)
env-295	JCVI_SCAF_1101668725097	+	166	321	**RNA**→ Nfo (COG0648)? (GOS623)→
env-296	JCVI_SCAF_1096628000408	-	875	719	? (GOS10359)→ **RNA**→ Nfo (COG0648)? (GOS623)→
env-297	JCVI_SCAF_1096627851272	+	131	286	**RNA**→ AP2Ec (smart00518)? (GOS623)→
env-298	JCVI_SCAF_1096628019168	-	546	391	←COG1247 (COG1247)? (GOS7913) **RNA**→ ? (GOS623)→
env-300	JCVI_SCAF_1101667214751	+	9	164	**RNA**→ ? (GOS623)→
env-303	JCVI_SCAF_1096626978143	-	2264	2109	**RNA**→ Nfo (COG0648)? (GOS623)→
env-304	JCVI_SCAF_1096627174223	-	1770	1614	←? (GOS7913) **RNA**→ ? (GOS11057767)→
env-305	JCVI_SCAF_1096627174199	-	591	435	←COG1247 (COG1247)? (GOS7913) **RNA**→ ? (GOS623)→
env-306	JCVI_SCAF_1096627290531	+	819	973	MET2 (COG2021)? (GOS3619)→ **RNA**→ Thr tRNA→
env-307	JCVI_SCAF_1096626873027	+	1148	1302	MET2 (COG2021)? (GOS3619)→ **RNA**→ Thr tRNA→
env-308	JCVI_SCAF_1096626892538	-	1100	947	MET2 (COG2021)? (GOS3619)→ **RNA**→ Thr tRNA→

60.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 60.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Ocean-VI RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

D cd00008 (13) 5'-3' exonuclease; T5 type 5'-3' exonuclease domains may co-occur with DNA polymerase I (Pol I) domains, or be part of Pol I containing complexes.
u COG0156 (1) 7-keto-8-aminopelargonate synthetase and related enzymes [Coenzyme metabolism]
u COG0192 (2) S-adenosylmethionine synthetase [Coenzyme metabolism]
u COG0406 (1) Fructose-2,6-bisphosphatase [Carbohydrate transport and metabolism]
D COG0648 (35) Endonuclease IV [DNA replication, recombination, and repair]
D COG1063 (2) Threonine dehydrogenase and related Zn-dependent dehydrogenases [Amino acid transport and metabolism / General function prediction only]
u d COG1247 (33) Sortase and related acyltransferases [Cell envelope biogenesis, outer membrane]
u COG2021 (21) Homoserine acetyltransferase [Amino acid transport and metabolism]
uD COG2124 (42) Cytochrome P450 [Secondary metabolites biosynthesis, transport, and catabolism]
u COG2873 (12) O-acetylhomoserine sulfhydrylase [Amino acid transport and metabolism]
u COG4912 (10) Predicted DNA alkylation repair enzyme [DNA replication, recombination, and repair]

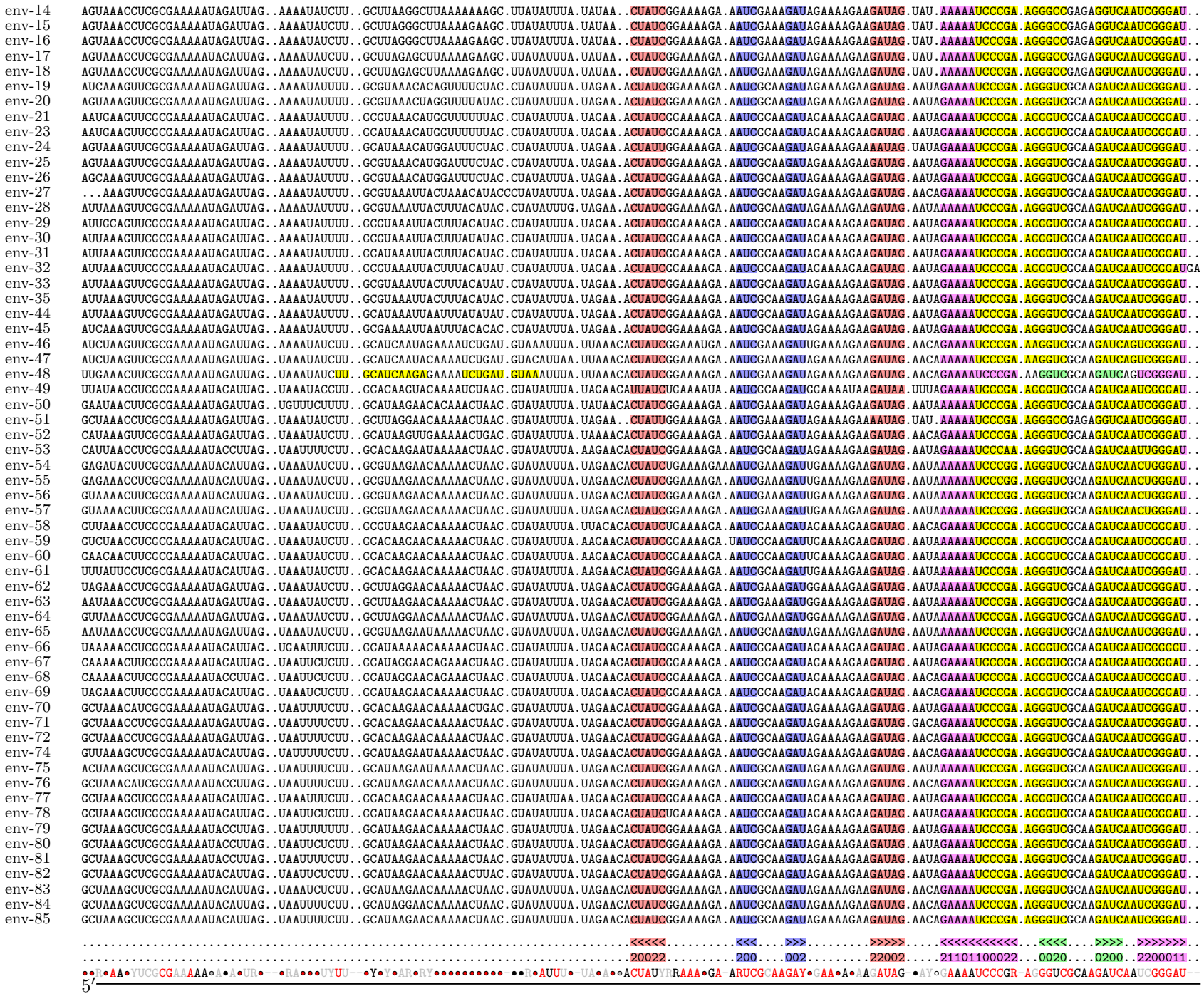
u GOS10117 (1) ?
u GOS10359 (70) ?
u GOS11003827 (11) ?
uD GOS11016378 (11) ?
D GOS11048107 (6) ?
D GOS11057767 (8) ?
D GOS11176326 (4) ?
D GOS11273709 (1) ?
d GOS11295005 (2) ?
uD GOS11302951 (3) ?
D GOS11452072 (2) ?
u GOS11528218 (2) ?
u GOS11574206 (19) ?
D GOS12009452 (2) ?
D GOS12085149 (1) ?
u d GOS13057899 (25) ?
D GOS13077636 (7) ?
u GOS13303413 (1) ?

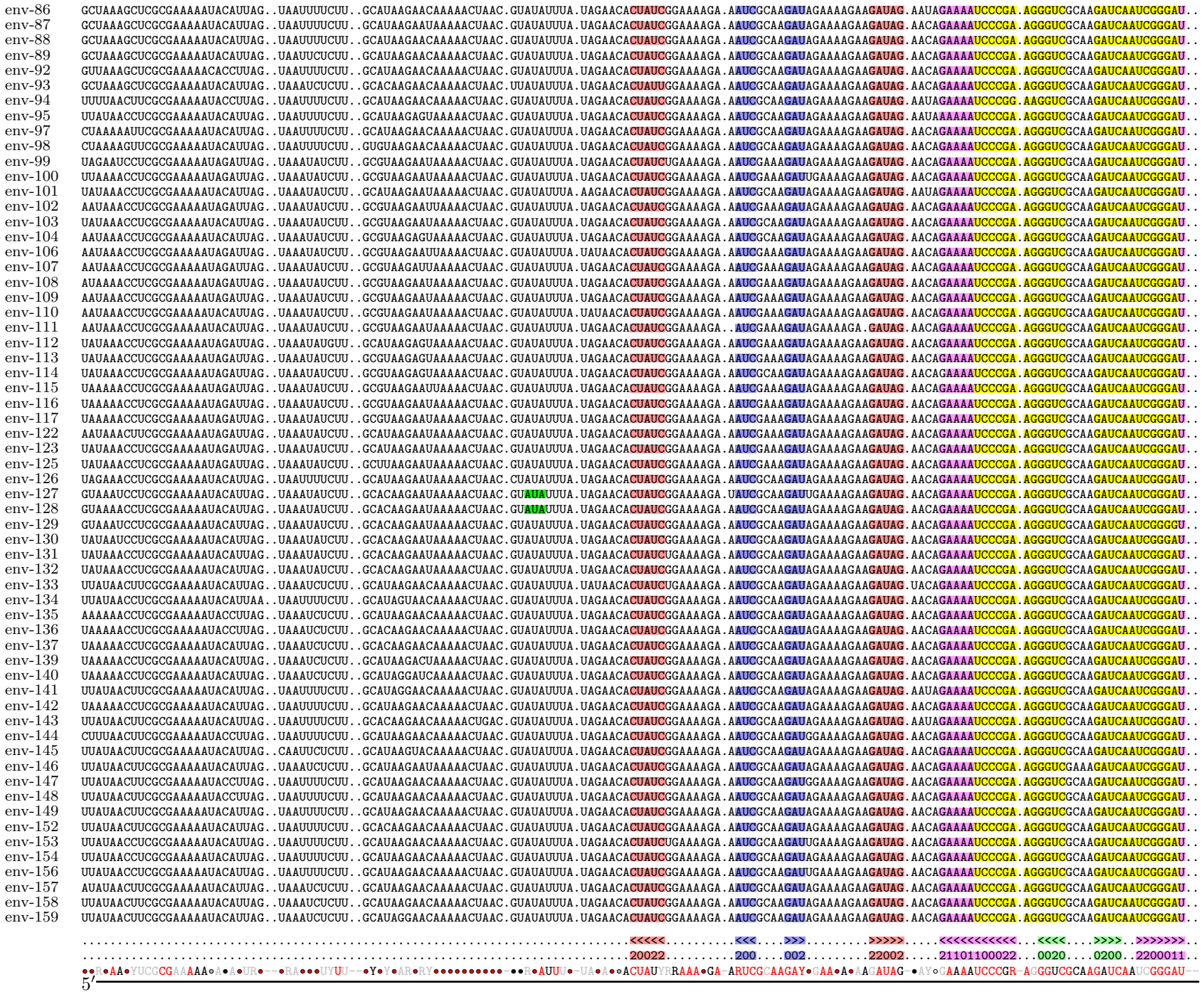
d GOS13372499 (2) ?
D GOS13372899 (2) ?
D GOS13557689 (3) ?
D GOS14670336 (2) ?
u GOS3285 (16) ?
D GOS333 (2) ?
uD GOS3357 (72) ?
u GOS3619 (27) ?
u GOS3874 (5) ?
u GOS4376 (2) ?
D GOS623 (47) ?
u GOS6677 (1) ?
u GOS670 (2) ?
D GOS6979 (31) ?
u d GOS7913 (35) ?
u GOS869 (1) ?
D pfam00067 (1) Cytochrome P450.
D pfam02739 (6) 5'-3' exonuclease, N-terminal resolvase-like domain.
D smart00475 (12) 5'-3' exonuclease;
D smart00518 (4) AP endonuclease family 2; These endonucleases play a role in DNA repair.

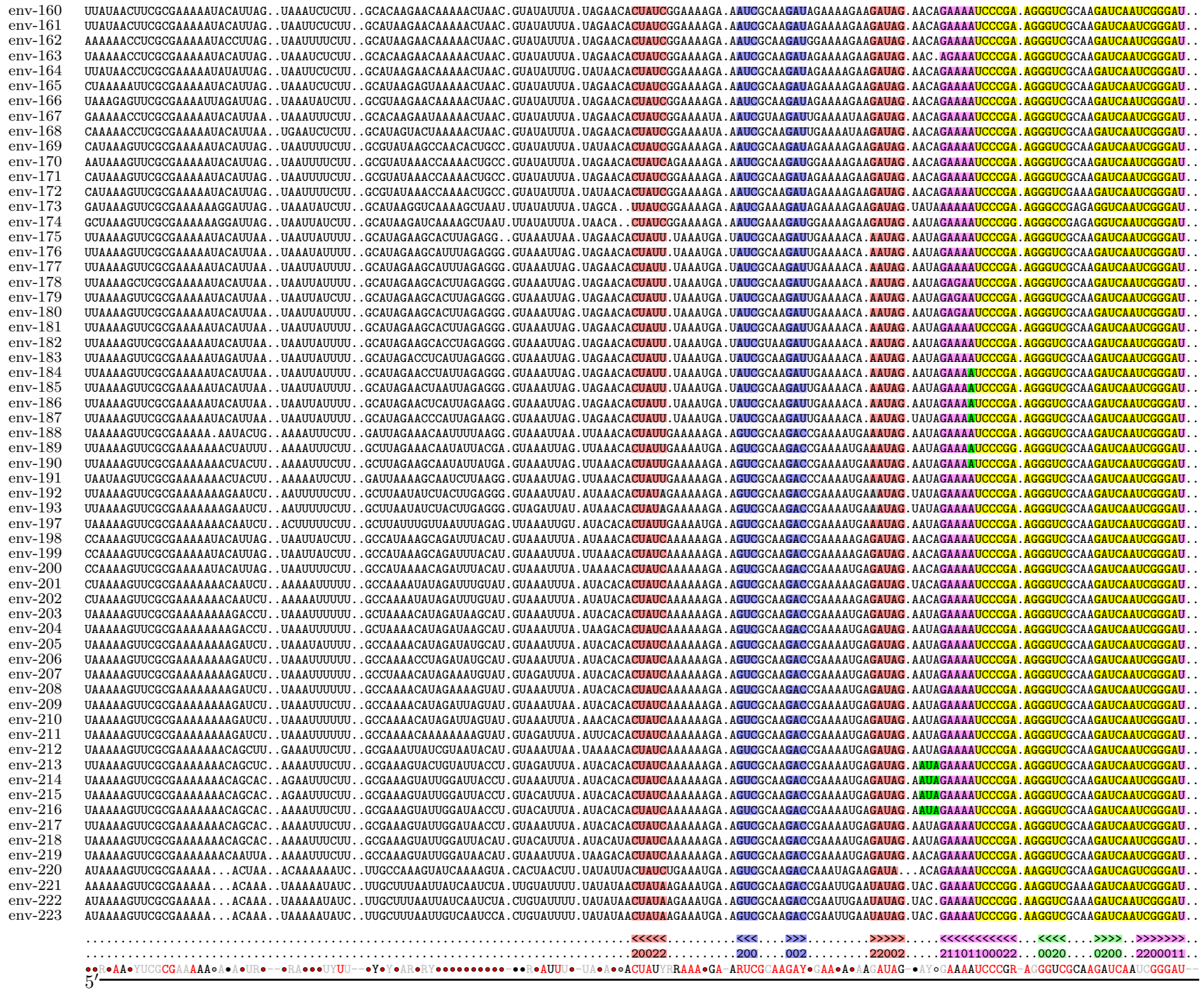
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: env-105, env-118, env-119, env-120, env-121, env-124, env-138, env-150, env-151, env-155, env-194, env-195, env-196, env-22, env-226, env-242, env-243, env-244, env-245, env-252, env-256, env-258, env-263, env-266, env-267, env-268, env-275, env-280, env-281, env-291, env-292, env-293, env-295, env-34, env-36, env-37, env-38, env-39, env-4, env-40, env-41, env-42, env-43, env-73, env-90, env-91, env-96

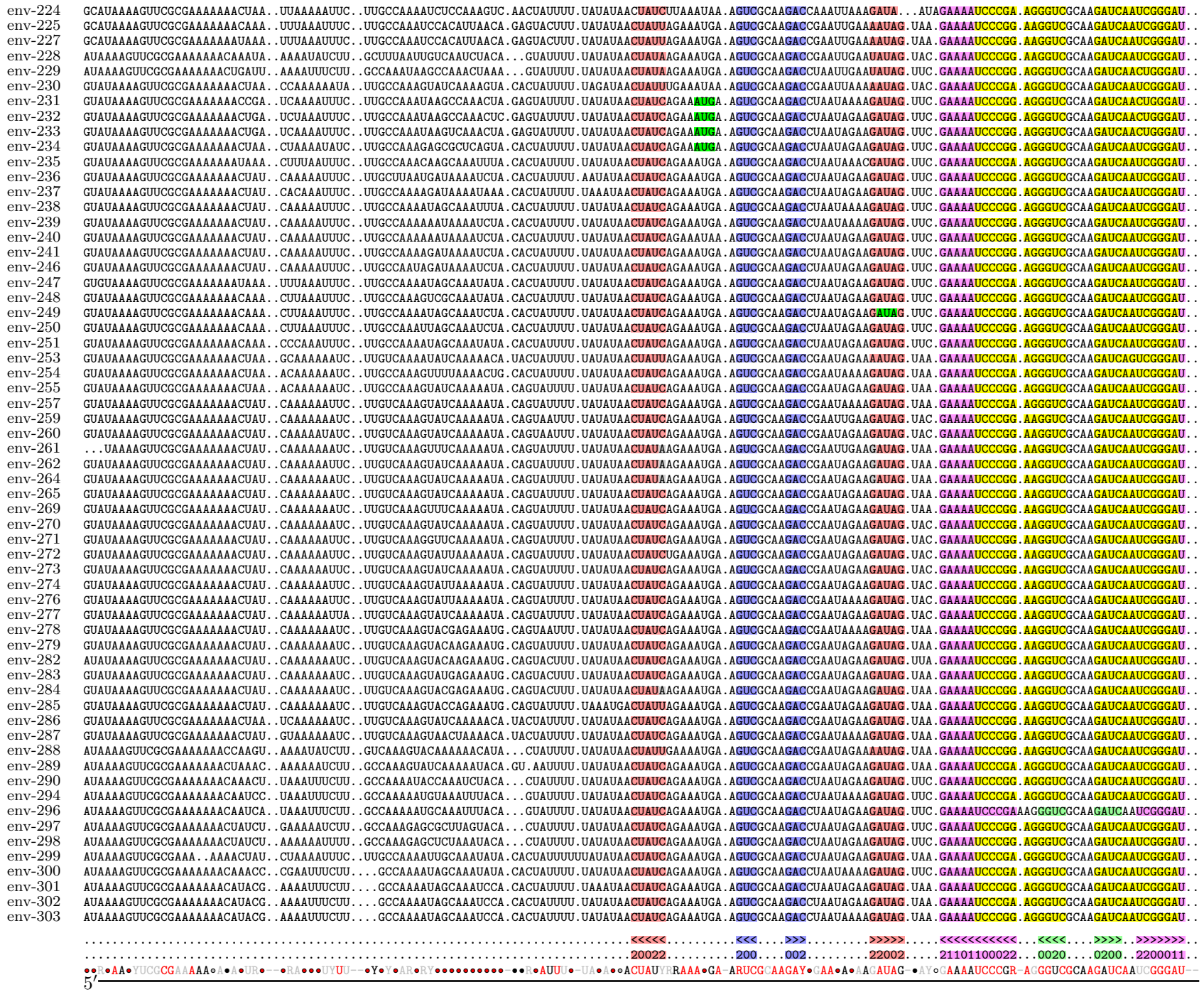
60.4 Multiple-sequence alignment

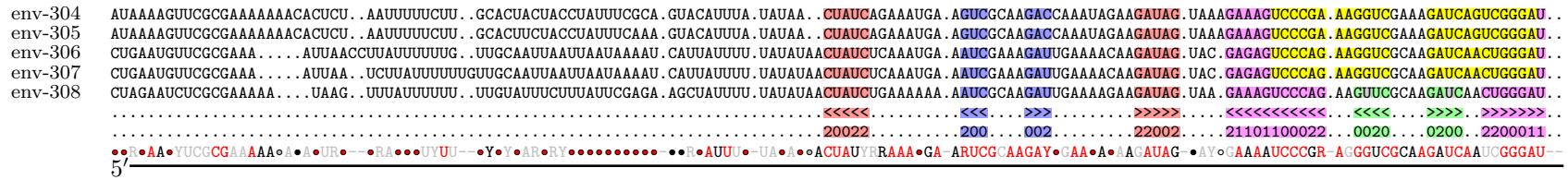
Each Ocean-VI RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 60.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Ocean-VI RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.


env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-35
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
5′

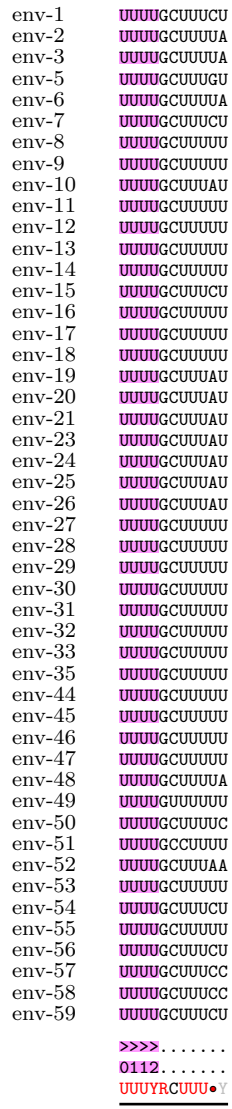

env-86 GCUAAAGCUCGCGAAAAAUACAUUAG..UAAUUUUCUU..GCAUAAGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AAUAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-87 GCUAAAGCUCGCGAAAAAUACAUUAG..UAAUUUUCUU..GCAUAAGAAUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-88 GCUAAAGCUCGCGAAAAAUACAUUAG..UAAUUUUCUU..GCAUAAGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-89 GCUAAAGCUCGCGAAAAAUACAUUAG..UAAUUCUCUU..GCAUAAGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-92 GUUAAAGCUCGCGAAAAACACCUUAG..UAAUUUUCUU..GCAUAAGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-93 GCUAAAGCUCGCGAAAAAUACAUUAG..UAAAUCUCUU..GCACAAGAACAAAAACUAAC.UUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-94 UUUUAACUUCGCGAAAAAUACCUUAG..UAAUUUUCUU..GCAUAAGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AAUAGAAAAUCCCGG.AAGGUCGCAAGAUCAAUCGGGAU..
env-95 UUAUAACCUCGCGAAAAAUACAUUAG..UAAUUUUCUU..GCAUAAGAGUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AAUAAAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-97 CUAAAAAUUCGCGAAAAAUACAUUAG..UAAUUUUCUU..GCAUAAGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-98 CUAAAAGUUCGCGAAAAAUACAUUAG..UAAUUUUCUU..GUGUAAGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-99 UAGAAUCCUCGCGAAAAAUAGAUUAG..UAAAUAUCUU..GCGUAAGAAUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCUGAAAAGA.AAUCGAAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-100 UUAAAACCUCGCGAAAAAUAGAUUAG..UAAAUAUCUU..GCGUAAGAAUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGAAAGAUUGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-101 UAUAAACCUCGCGAAAAAUACAUUAG..UAAAUAUCUU..GCAUAAGAAUAAAAACUAAC.GUAUAUUUA.AAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AAUAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-102 AAUAAACCUCGCGAAAAAUAGAUUAG..UAAAUAUCUU..GCGUAAGAAUUAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGAAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-103 UAUAAACCUCGCGAAAAAUACAUUAG..UAAAUAUCUU..GCGUAAGAAUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-104 AAUAAACCUCGCGAAAAAUACAUUAG..UAAAUAUCUU..GCGUAAGAGUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-106 AAUAAACCUCGCGAAAAAUAGAUUAG..UAAAUAUCUU..GCGUAAGAAUUAAAACUAAC.GUAUAUUUA.UAUAACACUAUCGGAAAAGA.AAUCGAAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-107 AAUAAACCUCGCGAAAAAUAGAUUAG..UAAAUAUCUU..GCGUAAGAUUAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGAAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-108 AUAAAACCUCGCGAAAAAUAGAUUAG..UAAAUAUCUU..GCGUAAGAAUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGAAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-109 AAUAAACCUCGCGAAAAAUAGAUUAG..UAAAUAUCUU..GCGUAAGAAUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGAAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-110 AAUAAACCUCGCGAAAAAUACAUUAG..UAAAUAUCUU..GCGUAAGAAUAAAAACUAAC.GUAUAUUUA.UAUAACACUAUCGGAAAAGA.AAUCGAAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-111 AAUAAACCUCGCGAAAAAUACAUUAG..UAAAUAUCUU..GCGUAAGAAUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA..AUCGAAAGAUAGAAAAGA.GAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-112 UAUAAACCUCGCGAAAAAUAGAUUAG..UAAAUAUGUU..GCAUAAGAGUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGAAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-113 UAUAAACCUCGCGAAAAAUAGAUUAG..UAAAUAUCUU..GCGUAAGAGUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGAAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-114 UAUAAACCUCGCGAAAAAUAGAUUAG..UAAAUAUCUU..GCGUAAGAGUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-115 UAAAAACCUCGCGAAAAAUAGAUUAG..UAAAUAUCUU..GCGUAAGAAUUAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGAAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-116 UAAAAACCUCGCGAAAAAUAGAUUAG..UAAAUAUCUU..GCGUAAGAAUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGAAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-117 UAAAAACCUCGCGAAAAAUAGAUUAG..UAAAUAUCUU..GCGUAAGAAUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGAAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-122 AAUAAACUUCGCGAAAAAUAGAUUAG..UAAAUAUCUU..GCAUAAGAAUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGAAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-123 UAUAAACCUCGCGAAAAAUAGAUUAG..UAAAUAUCUU..GCAUAAGAAUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGAAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-125 UAUAAACCUCGCGAAAAAUAGAUUAG..UAAAUAUCUU..GCUUAAGAAUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGAAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-126 UAGAAACCUCGCGAAAAAUACAUUAG..UAAUUUUCUU..GCAUAAGAAUAAAAACUAAC.CUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-127 GUAAAUCCUCGCGAAAAAUACAUUAG..UAAAUAUCUU..GCACAAGAAUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.UAUCGCAAGAUUGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-128 GUAAAACCUCGCGAAAAAUACAUUAG..UAAAUAUCUU..GCACAAGAAUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-129 GUAAAUCCUCGCGAAAAAUACAUUAG..UAAAUAUCUU..GCACAAGAAUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGGU..
env-130 UAUAAUCCUCGCGAAAAAUACAUUAG..UAAAUAUCUU..GCACAAGAAUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-131 UAUAAACCUCGCGAAAAAUACAUUAG..UAAAUAUCUU..GCACAAGAAUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCUGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-132 UAUAAACCUCGCGAAAAAUACAUUAG..UAAAUAUCUU..GCACAAGAAUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-133 UUAUAACUUCGCGAAAAAUACAUUAG..UAAAUCUCUU..GCAUAAGAACAAAAACUAAC.GUAUAUUUA.UAUAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.UACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-134 UUAUAACCUCGCGAAAAAUACAUUAA..UAAUUUUCUU..GCAUAGUAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-135 AAAAAACCUCGCGAAAAAUACCUUAG..UAAAUCUCUU..GCAUAAGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-136 UAAAAACCUCGCGAAAAAUACCUUAG..UAAAUCUCUU..GCACAAGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-137 UAAAAACCUCGCGAAAAAUACAUUAG..UAAAUCUCUU..GCACAAGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-139 UAAAAACCUCGCGAAAAAUACAUUAG..UAAAUCUCUU..GCAUAAGACUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-140 UAAAAACCUCGCGAAAAAUACAUUAG..UAAAUCUCUU..GCAUAGGAUCAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-141 UUAUAACUUCGCGAAAAAUACAUUAG..UAAUUUUCUU..GCAUAGGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AAUAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-142 UAAAAACCUCGCGAAAAAUACAUUAG..UAAUUUUCUU..GCAUAAGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-143 UUAUAACUUCGCGAAAAAUACAUUAG..UAAUUUUCUU..GCACAAGAACAAAAACUGAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AAUAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-144 CUUUAACUUCGCGAAAAAUACCUUAG..UAAUUUUCUU..GCAUAAGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUGGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-145 UUAUAACUUCGCGAAAAAUACAUUAG..CAAUUCUCUU..GCAUAAGUACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-146 UUAUAACUUCGCGAAAAAUACAUUAG..UAAAUCUCUU..GCAUAAGAAUAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGAAAGAUCAAUCGGGAU..
env-147 UUAUAACUUCGCGAAAAAUACCUUAG..UAAUUUUCUU..GCAUAAGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUGGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-148 UUAUAACUUCGCGAAAAAUACCUUAG..UAAUUUUCUU..GCAUAAGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-149 UUAUAACUUCGCGAAAAAUACAUUAG..UAAUUUUCUU..GCAUAAGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-152 UUAUAACUUCGCGAAAAAUACAUUAG..UAAUUUUCUU..GCACAAGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-153 UUAUAACCUCGCGAAAAAUACAUUAG..UAAUUUUCUU..GCAUAAGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCUGAAAAGA.AAUCGCAAGAUUGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-154 UUAUAACCUCGCGAAAAAUACAUUAG..UAAUUUUCUU..GCAUAAGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-156 UUAUAACCUCGCGAAAAAUACAUUAG..UAAUUUUCUU..GCAUAAGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUUGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-157 AUAUAACUUCGCGAAAAAUACAUUAG..UAAAUCUCUU..GCAUAAGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-158 UUAUAACUUCGCGAAAAAUACAUUAG..UAAAUCUCUU..GCAUAAGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
env-159 UUAUAACUUCGCGAAAAAUACAUUAG..UAAAUCUCUU..GCAUAGGAACAAAAACUAAC.GUAUAUUUA.UAGAACACUAUCGGAAAAGA.AAUCGCAAGAUAGAAAAGAAGAUAG.AACAGAAAAUCCCGA.AGGGUCGCAAGAUCAAUCGGGAU..
<<<<<.....<<<....>>>.....>>>>>.....<<<<<<<<<<<...<<<<....>>>>..>>>>>>>..
20022.....200....002.....22002.....21101100022...0020....0200..2200011..
5′


env-160
env-161
env-162
env-163
env-164
env-165
env-166
env-167
env-168
env-169
env-170
env-171
env-172
env-173
env-174
env-175
env-176
env-177
env-178
env-179
env-180
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
env-190
env-191
env-192
env-193
env-197
env-198
env-199
env-200
env-201
env-202
env-203
env-204
env-205
env-206
env-207
env-208
env-209
env-210
env-211
env-212
env-213
env-214
env-215
env-216
env-217
env-218
env-219
env-220
env-221
env-222
env-223
5′


env-224
env-225
env-227
env-228
env-229
env-230
env-231
env-232
env-233
env-234
env-235
env-236
env-237
env-238
env-239
env-240
env-241
env-246
env-247
env-248
env-249
env-250
env-251
env-253
env-254
env-255
env-257
env-259
env-260
env-261
env-262
env-264
env-265
env-269
env-270
env-271
env-272
env-273
env-274
env-276
env-277
env-278
env-279
env-282
env-283
env-284
env-285
env-286
env-287
env-288
env-289
env-290
env-294
env-296
env-297
env-298
env-299
env-300
env-301
env-302
env-303
5′



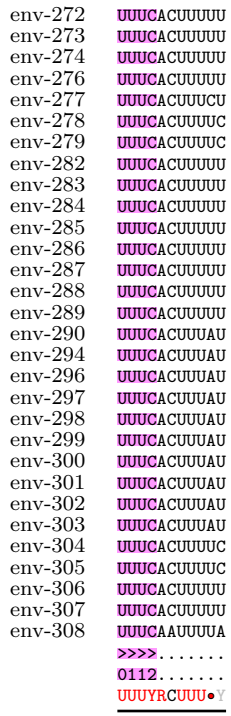
alignment positions 160 · · · 170



```
env-60    UUUUGCUUUUU
env-61    UUUUGCUUUUU
env-62    UUUUGCUUUCC
env-63    UUUUGCUUUAU
env-64    UUUUGCUUUAU
env-65    UUUUGCUUUUC
env-66    UUUUGCUUUUU
env-67    UUUUGCUUUAU
env-68    UUUUGCUUUUU
env-69    UUUUGCUUUUA
env-70    UUUUGCUUUUU
env-71    UUUUGCCUUUU
env-72    UUUUGCUUUUU
env-74    UUUUGCUUUUU
env-75    UUUUGCUUUUA
env-76    UUUUGCUUUUU
env-77    UUUUGCUUUUU
env-78    UUUUGCUUUUA
env-79    UUUUGCUU...
env-80    UUUUGCUUUUA
env-81    UUUUGCUUUUA
env-82    UUUUGCUUUUA
env-83    UUUUGCUUUUA
env-84    UUUUGCUUUUU
env-85    UUUUGCUUUUU
env-86    UUUUGCUUUUU
env-87    UUUUGCUUUUA
env-88    UUUUGCUUUUA
env-89    UUUUGCUUUUA
env-92    UUUUGCUUUCU
env-93    UUUUGCUUUUU
env-94    UUUUGCUUUUU
env-95    UUUUGCUUUUU
env-97    UUUUGCUUUUA
env-98    UUUUGCUUUUA
env-99    UUUUGCUUUCC
env-100   UUUUGCUUUUU
env-101   UUUUGCUUUCC
env-102   UUUUGCUUUUA
env-103   UUUUGCUUUCC
env-104   UUUUGCUUUCC
env-106   UUUUGCUUUCU
env-107   UUUUGCUUUCC
env-108   UUUUGCUUUCU
env-109   UUUUGCUUUCU
env-110   UUUUGCUUUCU
env-111   UUUUGCUUUCU
env-112   UUUUGCUUUCC
env-113   UUUUGCUUUCC
env-114   UUU.GCUUUCC
env-115   UUUUGCUUUCC
env-116   UUUUGCUUUCU
env-117   UUUUGCUUUCC
env-122   UUUUGCUUUCC
env-123   UUUUGCUUUCC
env-125   UUUUGCUUUCC
env-126   UUUUGCUUU..
env-127   UUUUGCUUUUU
env-128   UUUUGCUUUUU
env-129   UUUUGCUUUUU
env-130   UUUUGCUUUUU

          >>>>.......
          0112.......
          UUUYRCUUU•Y
```

```
env-131   UUUUGCUUUUU
env-132   UUUUGCUUUUU
env-133   UUUUGCUUUUU
env-134   UUUUGCUUUUU
env-135   UUUUGCUUUUU
env-136   UUUUGCUUUUU
env-137   UUUUGCUUUUU
env-139   UUUUGCUUUUU
env-140   UUUUGCUUUUU
env-141   UUUUGCUUUUU
env-142   UUUUGCUUUUU
env-143   UUUUGCUUUUU
env-144   UUUUGCUUUUU
env-145   UUUUGCUUUUU
env-146   UUUUGCUUUUU
env-147   UUUUGCUUUUU
env-148   UUUUGCUUUUU
env-149   UUUUGCUUUUU
env-152   UUUUGCUUUUU
env-153   UUUUGCUUUUU
env-154   UUUUGCUUUUU
env-156   UUUUGCUUUUU
env-157   UUUUGCUUUUU
env-158   UUUUGCUUUUU
env-159   UUUUGCUUUUU
env-160   UUUUGCUUUUU
env-161   UUUUGCUUUUU
env-162   UUUUGCUUUUU
env-163   UUUUGCUUUUA
env-164   UUUUGCUUUUU
env-165   UUUUGCUUU..
env-166   UUUUGCUUUCU
env-167   UUUUGCUUUUU
env-168   UUUUGCUUUUU
env-169   UUUUGCUUUUU
env-170   UUUUGCUUUGU
env-171   UUUUGCUUUGU
env-172   UUUUGCUUUGU
env-173   UUUUGCUUUAU
env-174   UUUUGCUUUUA
env-175   UUUUGCUUU..
env-176   UUUUGCUUUUC
env-177   UUUUGCUUUUC
env-178   UUUUGCUUUUC
env-179   UUUUGCUUUUC
env-180   UUUUGCUUUUC
env-181   UUUUGCUUUUC
env-182   UUUUGCUUUUC
env-183   UUUUGCUUUUC
env-184   UUUUGCUUUUC
env-185   UUUUGCUUUUC
env-186   UUUUGCUUUUC
env-187   UUUUGCUUUUC
env-188   UUUUGCUUUUU
env-189   UUUUGUUUUUA
env-190   UUUUGCUUUUA
env-191   UUUUGCUUU..
env-192   UUUUGCUUUUC
env-193   UUUUGCUUUUC
env-197   UUUUGCUUUUU
env-198   UUUUGCUUU..

          >>>>.......
          0112.......
          UUUYRCUUU•Y
```

```
env-199  UUUUGCUUUGU
env-200  UUUUUGCUUUGU
env-201  UUUUUGCUUUAU
env-202  UUUUUGCUUUAU
env-203  UUUUGCUUUAU
env-204  UUUUGCUUUAU
env-205  UUUUGCUUUAC
env-206  UUUUGCUUUAU
env-207  UUUUGCUUUUU
env-208  UUUUGAUUUCU
env-209  UUUUGCUUUCU
env-210  UUUUGCUUUCU
env-211  UUUUGAUUUCU
env-212  UUUUGCUUUAU
env-213  UUUUGCUUUAU
env-214  UUUUGCUUUAU
env-215  UUUUGCUUUAU
env-216  UUUUGCUUUAU
env-217  UUUUGCUUUCU
env-218  UUUUGCUUUUU
env-219  UUUUUGCUUUAU
env-220  UUUUGCUUUUU
env-221  UUUCACUUUAU
env-222  UUUCACUUUAU
env-223  UUUCACUUUAU
env-224  UUUUGCUUUUU
env-225  UUUCACUUUUA
env-227  UUUCACUUUUA
env-228  UUUCACUUUAU
env-229  UUUCACUUUUC
env-230  UUUCACUUUAU
env-231  UUUCACUUUUC
env-232  UUUCACUUUUC
env-233  UUUCACUUUUC
env-234  UUUCACUUUAU
env-235  UUUCACUUUAU
env-236  UUUCACUUUGU
env-237  UUUCACUUUAU
env-238  UUUCACUUUAU
env-239  UUUCACUUUAU
env-240  UUUCACUUUAU
env-241  UUUCACUUUAU
env-246  UUUCACUUUAU
env-247  UUUCACUUUAU
env-248  UUUCACUUUAU
env-249  UUUCACUUUAU
env-250  UUUCACUUUAU
env-251  UUUCACUUUAU
env-253  UUUCACUUUCU
env-254  UUUCACUUUUU
env-255  UUUCACUUUUU
env-257  UUUCACUUUUU
env-259  UUUCACUUUUU
env-260  UUUCACUUUUU
env-261  UUUCACUUUUU
env-262  UUUCACUUUUU
env-264  UUUCACUUUUU
env-265  UUUCACUUUUU
env-269  UUUCACUUUUU
env-270  UUUCACUUUUU
env-271  UUUCACUUUUU

         >>>>.......
         0112.......
         UUUYRCUUU•Y
```



61 *pan*

61.1 Taxa

The taxonomy of each organism containing a putative *pan* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *pan* RNA in Sections 61.2 and 61.4.

abbrev. of hits	taxonomy of species
Det-1-1	Bacteria Chloroflexi Dehalococcoidetes *Dehalococcoides ethenogenes* 195
Dsp-1-1	Bacteria Chloroflexi Dehalococcoidetes *Dehalococcoides sp.* BAV1
Dsp-2-1	Bacteria Chloroflexi Dehalococcoidetes *Dehalococcoides sp.* CBDB1
Hau-1-1	Bacteria Chloroflexi Herpetosiphonales Herpetosiphonaceae *Herpetosiphon aurantiacus* ATCC 23779
Bam-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus amyloliquefaciens* FZB42
Bcl-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus clausii* KSM-K16
Bco-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus coagulans* 36D1
Bha-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus halodurans* C-125
Bli-1-1 to Bli-1-2	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus licheniformis* ATCC 14580 (DSM 13)
Bsp-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus sp.* B14905
Bsp-2-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus sp.* NRRL B-14911
Bsp-3-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus sp.* SG-1
Bsu-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus subtilis subsp. subtilis str.* 168
Ban-2-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* 'Ames Ancestor'
Ban-3-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* A1055

Ban-4-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* A2012
Ban-5-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Ames
Ban-6-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Australia 94
Ban-7-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* CNEVA-9066
Ban-8-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Kruger B
Ban-9-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Sterne
Ban-10-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Vollum
Ban-11-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Western North America USA6153
Ban-1-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis* Tsiankovskii-I
Bce-1-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* 03BB108
Bce-2-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* AH1134
Bce-3-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* AH187
Bce-4-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* AH820
Bce-5-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* ATCC 10987
Bce-6-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* ATCC 14579
Bce-7-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* B4264
Bce-8-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* E33L
Bce-9-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* G9241
Bce-10-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* G9842
Bce-11-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* H3081.97
Bce-12-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* NVH0597-99
Bce-14-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus subsp. cytotoxis* NVH 391-98
Bce-13-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* W
Bth-1-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus thuringiensis serovar israelensis* ATCC 35646
Bth-2-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus thuringiensis serovar konkukian str.* 97-27
Bth-3-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus thuringiensis str.* Al Hakam
Bwe-1-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus weihenstephanensis* KBAB4
Esi-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Exiguobacterium sibiricum* 255-15
Gka-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Geobacillus kaustophilus* HTA426
Gth-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Geobacillus thermodenitrificans* NG80-2
Lin-1-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria innocua* Clip11262
Lmo-1-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* 10403S
Lmo-2-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* EGD-e
Lmo-3-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* F6900
Lmo-4-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* FSL F2-515
Lmo-5-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* FSL J1-175
Lmo-6-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* FSL J1-194
Lmo-7-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* FSL J2-064
Lmo-8-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* FSL J2-071
Lmo-9-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* FSL N1-017
Lmo-10-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* FSL N3-165
Lmo-11-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* FSL R2-503
Lmo-12-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* HPB2262
Lmo-13-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* J0161
Lmo-14-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* J2818
Lmo-15-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes* LO28
Lmo-16-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes str.* 1/2a F6854
Lmo-17-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes str.* 4b F2365
Lmo-18-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria monocytogenes str.* 4b H7858
Lwe-1-1	Bacteria Firmicutes Bacillales Listeriaceae *Listeria welshimeri serovar* 6b str. SLCC5334
Pla-1-1	Bacteria Firmicutes Bacillales Paenibacillaceae *Paenibacillus larvae subsp. larvae* BRL-230010
Cac-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium acetobutylicum* ATCC 824
Cbe-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium beijerinckii* NCIMB 8052
Cbo-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* A str. ATCC 19397
Cbo-2-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* A str. ATCC 3502
Cbo-3-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* A str. Hall
Cbo-4-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* C str. Eklund
Cbo-5-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* F str. Langeland
Cbo-6-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* G
Cbo-7-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* NCTC 2916
Cbo-8-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum str.* Iwanei E
Cbu-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium butyricum* 5521
Cdi-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium difficile* 630
Cdi-2-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium difficile* QCD-32g58
Cdi-3-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium difficile* QCD-66c26

Ckl-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium kluyveri* DSM 555
Cno-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium novyi* NT
Cth-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium thermocellum* ATCC 27405
Eve-1-1	Bacteria Firmicutes Clostridia Clostridiales Eubacteriaceae *Eubacterium ventriosum* ATCC 27560
Dlo-1-1	Bacteria Firmicutes Clostridia Clostridiales Lachnospiraceae *Dorea longicatena* DSM 13814
Dha-1-1 to Dha-1-2	Bacteria Firmicutes Clostridia Clostridiales Peptococcaceae *Desulfitobacterium hafniense* DCB-2
Dha-2-1 to Dha-2-2	Bacteria Firmicutes Clostridia Clostridiales Peptococcaceae *Desulfitobacterium hafniense* Y51
Dre-1-1 to Dre-1-2	Bacteria Firmicutes Clostridia Clostridiales Peptococcaceae *Desulfotomaculum reducens* MI-1
Pth-1-1	Bacteria Firmicutes Clostridia Clostridiales Peptococcaceae *Pelotomaculum thermopropionicum* SI
Tca-1-1	Bacteria Firmicutes Clostridia Clostridiales Veillonellaceae *Thermosinus carboxydivorans* Nor1
Chy-1-1	Bacteria Firmicutes Clostridia Thermoanaerobacterales Thermoanaerobacteraceae *Carboxydothermus hydrogenoformans* Z-2901
Mth-1-1	Bacteria Firmicutes Clostridia Thermoanaerobacterales Thermoanaerobacteraceae Moorella group *Moorella thermoacetica* ATCC 39073
Csa-1-1	Bacteria Firmicutes Clostridia Thermoanaerobacterales Thermoanaerobacterales Family III. Incertae Sedis *Caldicellulosiruptor saccharolyticus* DSM 8903
Tet-1-1	Bacteria Firmicutes Clostridia Thermoanaerobacteriales Thermoanaerobacteriaceae *Thermoanaerobacter ethanolicus* ATCC 33223
Tet-2-1	Bacteria Firmicutes Clostridia Thermoanaerobacteriales Thermoanaerobacteriaceae *Thermoanaerobacter ethanolicus* X514
Efa-1-1	Bacteria Firmicutes Lactobacillales Enterococcaceae *Enterococcus faecalis* V583
Efa-2-1	Bacteria Firmicutes Lactobacillales Enterococcaceae *Enterococcus faecium* DO
Dac-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Desulfuromonadaceae *Desulfuromonas acetoxidans* DSM 684
Gbe-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Geobacteraceae *Geobacter bemidjiensis* Bem
Glo-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Geobacteraceae *Geobacter lovleyi* SZ
Gme-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Geobacteraceae *Geobacter metallireducens* GS-15
Gsu-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Geobacteraceae *Geobacter sulfurreducens* PCA
Gur-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Geobacteraceae *Geobacter uraniireducens* Rf4
Pca-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Pelobacteraceae *Pelobacter carbinolicus* DSM 2380
Ppr-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Pelobacteraceae *Pelobacter propionicus* DSM 2379
Ade-1-1	Bacteria Proteobacteria Deltaproteobacteria Myxococcales Cystobacterineae Myxococcaceae *Anaeromyxobacter dehalogenans* 2CP-C
Asp-1-1	Bacteria Proteobacteria Deltaproteobacteria Myxococcales Cystobacterineae Myxococcaceae *Anaeromyxobacter sp.* Fw109-5
Sac-1-1 to Sac-1-2	Bacteria Proteobacteria Deltaproteobacteria Syntrophobacterales Syntrophaceae *Syntrophus aciditrophicus* SB
Sfu-1-1	Bacteria Proteobacteria Deltaproteobacteria Syntrophobacterales Syntrophobacteraceae *Syntrophobacter fumaroxidans* MPOB
env-1 to env-7	environmental samples

61.2 Gene contexts

Each *pan* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *pan* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 61.3. The accession number of the sequence containing each *pan* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 61.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Dlo-1-1	NZ_AAXB02000027.1	+	5609	5713	**RNA**→ Asp_decarbox (pfam02261)→
Asp-1-1	NC_009675.1	-	4440640	4440607	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ SmpB (COG0691)→
Ade-1-1	NC_007760.1	-	4219175	4219142	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ SmpB (COG0691)→ COG2206 (COG2206)→
Csa-1-1	NC_009437.1	-	2883401	2883291	**RNA**→ COG5495 (COG5495)→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Eve-1-1	NZ_AAVL02000032.1	+	78823	78915	**RNA**→ COG5495 (COG5495)→ PanB (COG0413)→ PanC (cd00560)→
env-1	BABD01006986.1	+	1324	1416	**RNA**→
Cac-1-1	NC_003030.1	+	3049173	3049265	**RNA**→ hypo→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Ckl-1-1	NC_009706.1	+	620335	620416	**RNA**→ COG5495 (COG5495)→ PanB (COG0413)→ PanC (cd00560)→
Cbo-2-1	NC_009495.1	-	507696	507607	**RNA**→ PanB (COG0413)→ PanC (cd00560)→
Cbo-1-1	NC_009697.1	-	472473	472384	**RNA**→ PanB (COG0413)→ PanC (cd00560)→
Cbo-3-1	NC_009698.1	-	492861	492772	**RNA**→ PanB (COG0413)→ PanC (cd00560)→

Cbo-5-1	NC_009699.1	-	519010	518921	**RNA**→ PanB (COG0413)→ PanC (cd00560)→
Cbo-7-1	NZ_ABDO01000005.1	+	231573	231662	**RNA**→
Cbo-8-1	NZ_ABDR01000003.1	-	14998	14909	**RNA**→
Cbo-6-1	NZ_ABDS01000009.1	+	59652	59741	**RNA**→
Cbu-1-1	NZ_ABDT01000064.1	+	8154	8236	**RNA**→
Cbe-1-1	NC_009617.1	-	3038116	3038029	**RNA**→ COG5495 (COG5495)→ PanB (COG0413)→
Cbo-4-1	NZ_ABDQ01000017.1	+	20598	20693	**RNA**→
Cno-1-1	NC_008593.1	+	189304	189393	**RNA**→ COG5495 (COG5495)→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Cdi-1-1	NC_009089.1	-	1753012	1752923	**RNA**→ COG5495 (COG5495)→ PanB (COG0413)→ PanC (cd00560)→
Cdi-2-1	NZ_AAML04000010.1	-	512346	512257	**RNA**→ COG5495 (COG5495)→ hypo→ PanB (COG0413)→ PanC (cd00560)→
Cdi-3-1	NZ_ABFD01000041.1	-	27339	27250	**RNA**→
Efa-2-1	NZ_AAAK03000022.1	+	21023	21104	**RNA**→ COG5495 (COG5495)→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Efa-1-1	NC_004668.1	-	1807268	1807187	**RNA**→ COG5495 (COG5495)→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Gur-1-1	NC_009483.1	+	2436108	2436193	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→ Met_dep_hydrolase_D (cd01312)→ ComEA (COG1555)→ SmpB (COG0691)→ tmRNA→
Gbe-1-1	NZ_AAZA01000002.1	-	132354	132270	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Bcl-1-1	NC_006582.1	-	2188202	2188112	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→ TPR (cd00189)→ DnaQ (COG0847)DinG (COG1199)→ ATZ_TRZ_like (cd01298)→
Bsp-1-1	NZ_AAXV01000082.1	-	1922	1833	**RNA**→ PanB (COG0413)→ PanC (cd00560)→
Esi-1-1	NZ_AADW02000001.1	-	474525	474433	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Glo-1-1	NZ_AAVG01000012.1	-	25672	25578	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ GadB (COG0076)→
Dac-1-1	NZ_AAEW02000005.1	-	99021	98925	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Met_dep_hydrolase_D (cd01312)→ hypo (cons)→ SmpB (COG0691)→
Pla-1-1	NZ_AARF01000121.1	+	6916	7046	**RNA**→
Hau-1-1	NZ_AATI01000127.1	-	6444	6361	**RNA**→ COG5495 (COG5495)→
env-2	BABF01000941.1	+	246	346	**RNA**→ COG5495 (COG5495)→ PanB (COG0413)→
Tet-1-1	NZ_AAKQ01000028.1	+	5416	5505	**RNA**→ COG5495 (COG5495)→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Tet-2-1	NZ_AATV01000038.1	+	1917	2006	**RNA**→ COG5495 (COG5495)→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Cth-1-1	NC_009012.1	-	1079695	1079609	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Ppr-1-1	NC_008609.1	-	3908711	3908623	**RNA**→ PanB (COG0413)→ PanC (cd00560)→
Tca-1-1	NZ_AAWL01000012.1	-	11046	10947	**RNA**→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Gth-1-1	NC_009328.1	-	2247695	2247598	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→ DnaQ (COG0847)DinG (COG1199)→
Gka-1-1	NC_006510.1	-	2224448	2224346	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→ DnaQ (COG0847)DinG (COG1199)→
Dha-2-1	NC_007907.1	+	495034	495118	**RNA**→ PanB (COG0413)→
Dha-1-1	NZ_AAAW04000003.1	-	157429	157345	**RNA**→ PanB (COG0413)→
env-3	BAAV01019155.1	+	11	106	**RNA**→
Bsu-1-1	NC_000964.2	-	2354075	2353984	**RNA**→ PanB (COG0413)→ Pantoate_ligase (pfam02569)→ Asp_decarbox (pfam02261)→
Bam-1-1	NC_009725.1	-	2160220	2160132	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Bli-1-1	NC_006322.1	-	2325492	2325403	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Bli-1-2	NC_006270.2	-	2324634	2324545	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Bth-1-1	NZ_AAJM01000082.1	-	9999	9903	**RNA**→ ←hypo
Bce-10-1	NZ_ABDJ01000005.1	-	406860	406764	**RNA**→
Bth-3-1	NC_008600.1	+	1530923	1531019	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Bce-13-1	NZ_ABCZ01000003.1	-	344383	344287	**RNA**→
Bce-1-1	NZ_ABDM01000018.1	-	72753	72657	**RNA**→
Bce-4-1	NZ_AAUE01000001.1	+	490843	490939	**RNA**→
Bth-2-1	NC_005957.1	+	1495897	1495993	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Bce-5-1	NC_003909.8	+	1625746	1625842	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Ban-9-1	NC_005945.1	+	1471530	1471626	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Ban-2-1	NC_007530.2	+	1471576	1471672	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Ban-5-1	NC_003997.3	+	1471453	1471549	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→

Ban-7-1	NZ_AAEN01000011.1	+	41029	41125	**RNA**→
Ban-3-1	NZ_AAEO01000025.1	+	40728	40824	**RNA**→
Ban-10-1	NZ_AAEP01000035.1	+	43247	43343	**RNA**→
Ban-8-1	NZ_AAEQ01000029.1	+	44759	44855	**RNA**→
Ban-11-1	NZ_AAER01000023.1	+	32064	32160	**RNA**→
Ban-6-1	NZ_AAES01000034.1	+	43003	43099	**RNA**→
Ban-1-1	NZ_ABDN01000005.1	-	313337	313241	**RNA**→
Bce-8-1	NC_006274.1	+	1506507	1506603	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Ban-4-1	NZ_AAAC02000001.1	+	1959897	1959993	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Bce-3-1	NZ_AAUF01000003.1	+	11672	11768	**RNA**→
Bce-11-1	NZ_ABDL01000004.1	+	40698	40794	**RNA**→
Bce-12-1	NZ_ABDK01000005.1	-	174006	173910	**RNA**→
Bce-2-1	NZ_ABDA01000001.1	-	269821	269725	**RNA**→
Bce-7-1	NZ_ABDI01000003.1	-	284591	284495	**RNA**→
Bce-6-1	NC_004722.1	+	1488675	1488771	**RNA**→ ←hypo
Bce-9-1	NZ_AAEK01000008.1	-	121285	121188	**RNA**→ PanB (COG0413)→ PanC (cd00560)→
Bwe-1-1	NZ_AAOY01000005.1	-	112109	112012	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Bce-14-1	NC_009674.1	+	1369573	1369667	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Gme-1-1	NC_007517.1	+	1843118	1843197	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ GadB (COG0076)→ Met_dep_hydrolase_D (cd01312)→ SmpB (COG0691)→
Bsp-3-1	NZ_ABCF01000055.1	-	10245	10154	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
env-4	BAAV01003081.1	-	1285	1187	**RNA**→
env-5	BAAX01003760.1	-	141	43	**RNA**→
Dre-1-1	NC_009253.1	+	1959508	1959602	**RNA**→ COG5495 (COG5495)→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Bsp-2-1	NZ_AAOX01000023.1	-	58833	58746	**RNA**→ hypo→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→ aspartate 1-decarboxylase precursor→ DnaQ (COG0847)DinG (COG1199)→ ComEC (COG2333)→
Bco-1-1	NZ_AAWV01000088.1	-	4126	4038	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Sac-1-1	NC_007759.1	+	1255322	1255406	**RNA**→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Mth-1-1	NC_007644.1	+	135673	135762	**RNA**→ COG5495 (COG5495)→
Gsu-1-1	NC_002939.4	+	1868704	1868781	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ GadB (COG0076)→ Met_dep_hydrolase_D (cd01312)→ SmpB (COG0691)→ sRNA→
Pth-1-1	NC_009454.1	+	221041	221141	**RNA**→ COG5495 (COG5495)→ PanB (COG0413)→
Dre-1-2	NC_009253.1	+	168083	168168	**RNA**→ COG5495 (COG5495)→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Sac-1-2	NC_007759.1	-	1569009	1568972	**RNA**→ COG5495 (COG5495)→ PanB (COG0413)→
Pca-1-1	NC_007498.2	-	2140848	2140761	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→ Met_dep_hydrolase_D (cd01312)→
Det-1-1	NC_002936.3	+	736022	736109	**RNA**→ COG5495 (COG5495)→ Asp_decarbox (pfam02261)→
Dsp-2-1	NC_007356.1	+	635641	635728	**RNA**→ COG5495 (COG5495)→ Asp_decarbox (pfam02261)→ PanB (COG0413)→ PanC (cd00560)→
Dsp-1-1	NC_009455.1	+	714742	714829	**RNA**→ COG5495 (COG5495)→ Asp_decarbox (pfam02261)→ PanB (COG0413)→ PanC (cd00560)→
Sfu-1-1	NC_008554.1	-	441863	441827	**RNA**→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Dha-2-2	NC_007907.1	+	226528	226558	**RNA**→ COG5495 (COG5495)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Dha-1-2	NZ_AAAW04000003.1	-	417198	417168	**RNA**→ COG5495 (COG5495)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
env-6	JCVI_SCAF_1101668085111	-	890	859	**RNA**→ PanB (COG0413)? (GOS5559)→
env-7	ABEF01016840.1	-	293	262	**RNA**→
Chy-1-1	NC_007503.1	-	2132748	2132660	**RNA**→ hypo→
Bha-1-1	NC_002570.2	+	1757772	1757872	**RNA**→ COG3464 (COG3464)→
Lmo-4-1	NZ_AARI01000945.1	-	1535	1450	**RNA**→
Lwe-1-1	NC_008555.1	-	1952862	1952776	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→ DnaQ (COG0847)DinG (COG1199)→
Lmo-15-1	NZ_AARY01000381.1	-	4310	4225	**RNA**→
Lmo-17-1	NC_002973.6	-	1958220	1958134	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→ DnaQ (COG0847)DinG (COG1199)→
Lmo-12-1	NZ_AATL01000027.1	-	107334	107248	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→ DnaQ (COG0847)DinG (COG1199)→ hypo (cons)→ COG0436 (COG0436)→ AsnS (COG0017)→

Lmo-18-1	NZ_AADR01000086.1	+	2839	2925	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→ DnaQ (COG0847)DinG (COG1199)→ DNA polymerase III, epsilon subunit/ATP-dependent helicase DinG, putative→
Lmo-9-1	NZ_AARP02000048.1	-	105793	105707	**RNA**→ PanB (COG0413)→
Lmo-11-1	NZ_AARR01000032.1	-	105907	105821	**RNA**→
Lmo-6-1	NZ_AARJ01000030.1	-	122886	122800	**RNA**→
Lmo-7-1	NZ_AARO01000226.1	-	879	793	**RNA**→
Lmo-8-1	NZ_AARN02000030.1	-	125562	125476	**RNA**→
Lmo-2-1	NC_003210.1	-	1977238	1977152	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→ DnaQ (COG0847)DinG (COG1199)→ hypo→ COG0436 (COG0436)→ AsnS (COG0017)→
Lmo-16-1	NZ_AADQ01000024.1	-	16374	16288	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→ DnaQ (COG0847)DinG (COG1199)→ hypo (cons)→ COG0436 (COG0436)→ AsnS (COG0017)→
Lmo-3-1	NZ_AARU01000023.1	-	40306	40220	**RNA**→ PanB (COG0413)→
Lmo-13-1	NZ_AARW01000001.1	-	105421	105335	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→ DnaQ (COG0847)DinG (COG1199)→ hypo (cons)→ COG0436 (COG0436)→ AsnS (COG0017)→
Lmo-14-1	NZ_AARX01000027.1	-	105755	105669	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→
Lmo-1-1	NZ_AARZ01000003.1	-	105674	105588	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→ DnaQ (COG0847)DinG (COG1199)→ hypo (cons)→ COG0436 (COG0436)→ AsnS (COG0017)→
Lmo-10-1	NZ_AARQ01000027.1	-	107185	107099	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→ DnaQ (COG0847)DinG (COG1199)→ hypo (cons)→ COG0436 (COG0436)→ AsnS (COG0017)→
Lmo-5-1	NZ_AARK01000251.1	-	2227	2142	**RNA**→
Lin-1-1	NC_003212.1	-	2031651	2031565	**RNA**→ PanB (COG0413)→ PanC (cd00560)→ Asp_decarbox (pfam02261)→ DnaQ (COG0847)DinG (COG1199)→

61.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 61.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *pan* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

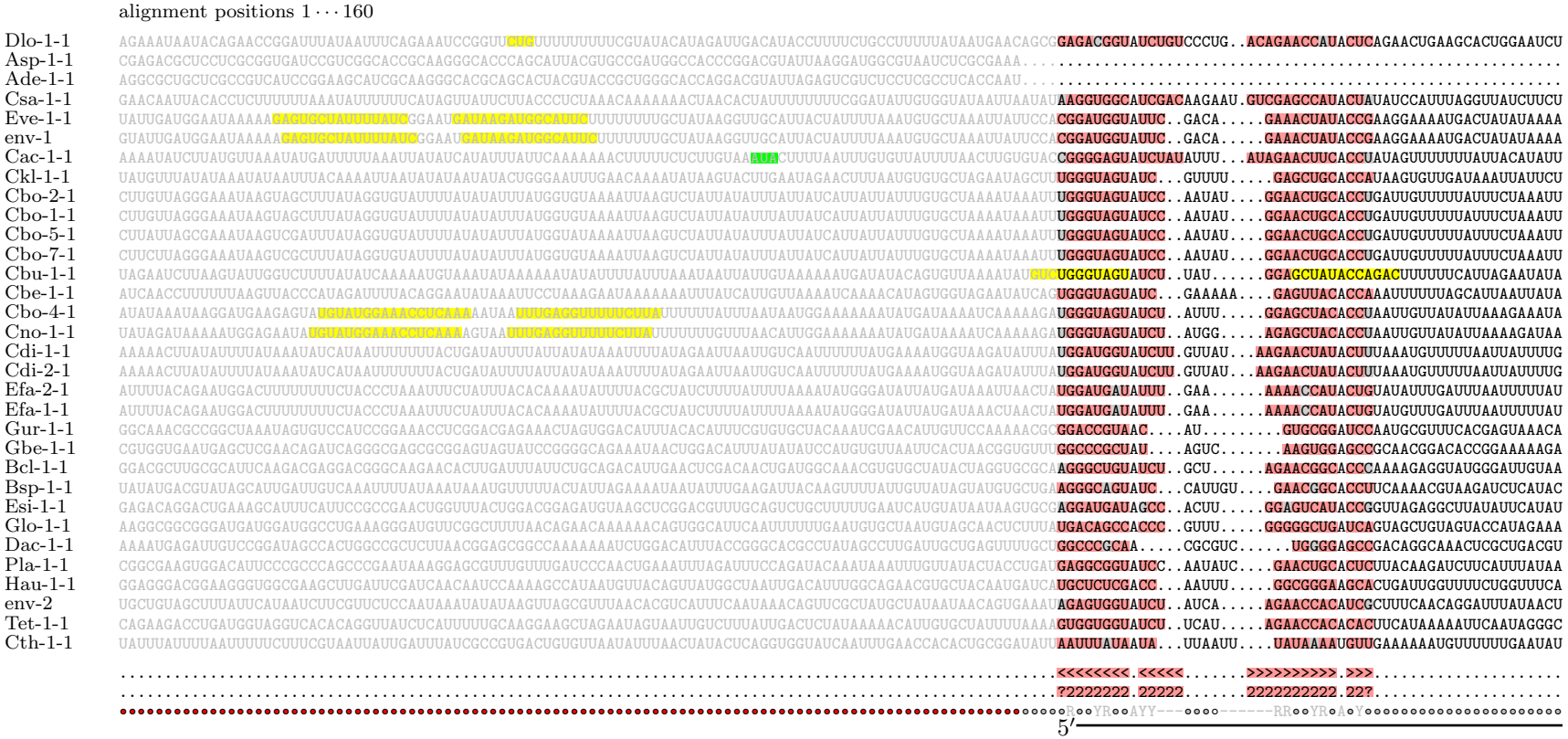
cd00189 (1) Tetratricopeptide repeat domain; typically contains 34 amino acids [WLF]-X(2)-[LIM]-[GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-X(2)-[ASL]-X(4)-[PKE]
cd00560 (68) PanC Pantoate-beta-alanine ligase, also known as pantothenate synthase, catalyzes the formation of pantothenate from pantoate and alanine.
cd01298 (1) TRZ/ATZ family contains enzymes from the atrazine degradation pathway and related hydrolases.
cd01312 (5) Metallo-dependent hydrolases, subgroup D is part of the superfamily of metallo-dependent hydrolases, a large group of proteins that show conservation in their 3-dimensional fold (TIM barrel) and in details of their active site.
COG0017 (6) Aspartyl/asparaginyl-tRNA synthetases [Translation, ribosomal structure and biogenesis]
COG0076 (3) Glutamate decarboxylase and related PLP-dependent proteins [Amino acid transport and metabolism]
COG0413 (73) Ketopantoate hydroxymethyltransferase [Coenzyme metabolism]
COG0436 (6) Aspartate/tyrosine/aromatic aminotransferase [Amino acid transport and metabolism]
COG0691 (6) tmRNA-binding protein [Posttranslational modification, protein turnover, chaperones]
COG0847 (14) DNA polymerase III, epsilon subunit and related 3'-5' exonucleases [DNA replication, recombination, and repair]
COG1199 (14) Rad3-related DNA helicases [Transcription / DNA replication, recombination, and repair]
COG1555 (1) DNA uptake protein and related DNA-binding proteins [DNA replication, recombination, and repair]
COG2206 (1) HD-GYP domain [Signal transduction mechanisms]
COG2333 (1) Predicted hydrolase (metallo-beta-lactamase superfamily) [General function prediction only]
COG3464 (1) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
COG5495 (23) Uncharacterized conserved protein [Function unknown]
GOS5559 (1) ?
pfam02261 (54) Aspartate decarboxylase.
pfam02569 (1) Pantoate-beta-alanine ligase.

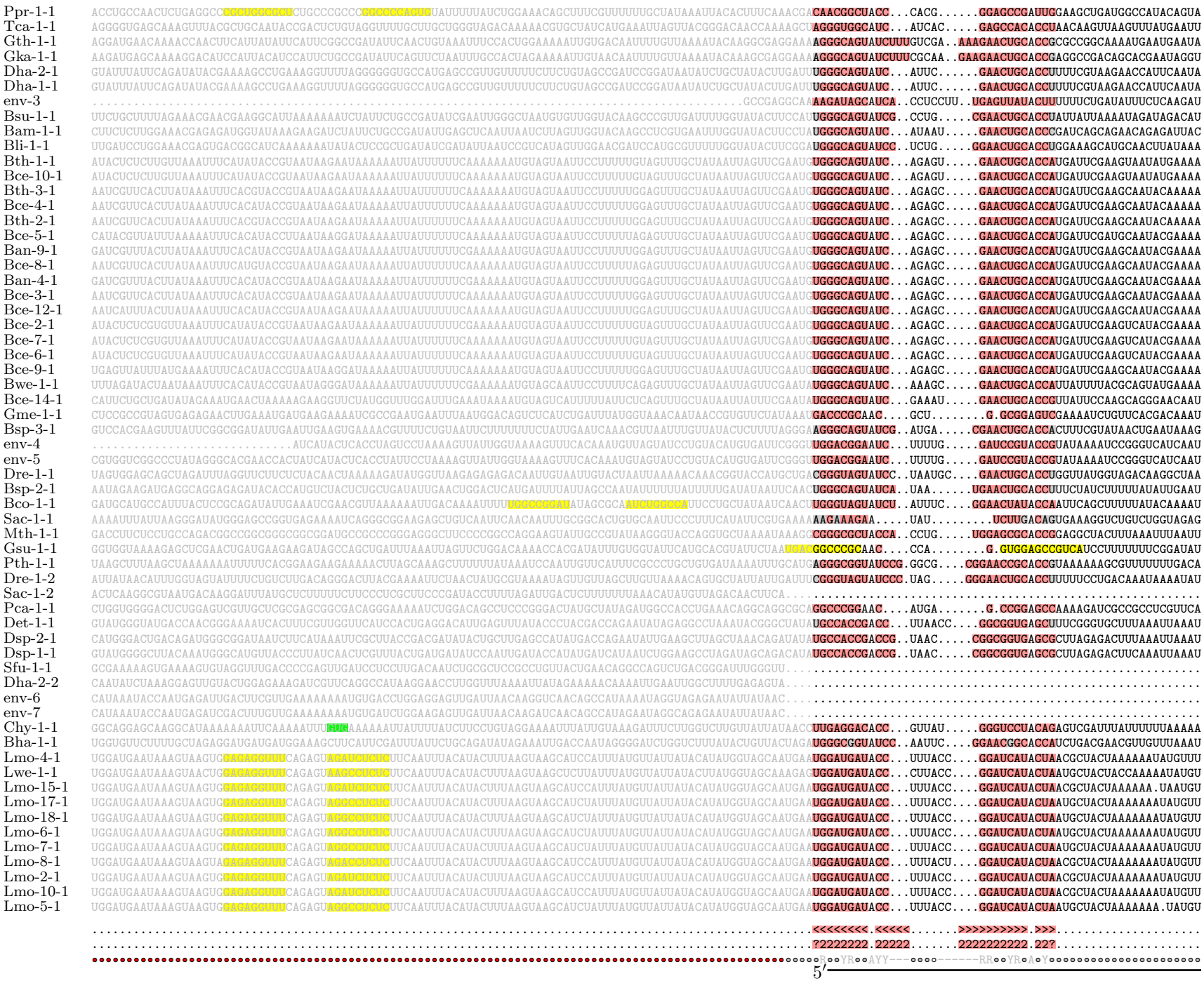
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Ban-1-1, Ban-10-1, Ban-11-1, Ban-2-1, Ban-3-1, Ban-5-1, Ban-6-1, Ban-7-1, Ban-8-1, Bce-1-1, Bce-11-1, Bce-13-1, Bli-1-2, Cbo-3-1, Cbo-6-1, Cbo-8-1, Cdi-3-1, Dha-1-2,

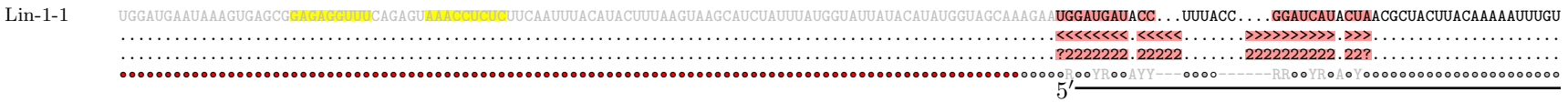
Lmo-1-1, Lmo-11-1, Lmo-12-1, Lmo-13-1, Lmo-14-1, Lmo-16-1, Lmo-3-1, Lmo-9-1, Tet-2-1

61.4 Multiple-sequence alignment

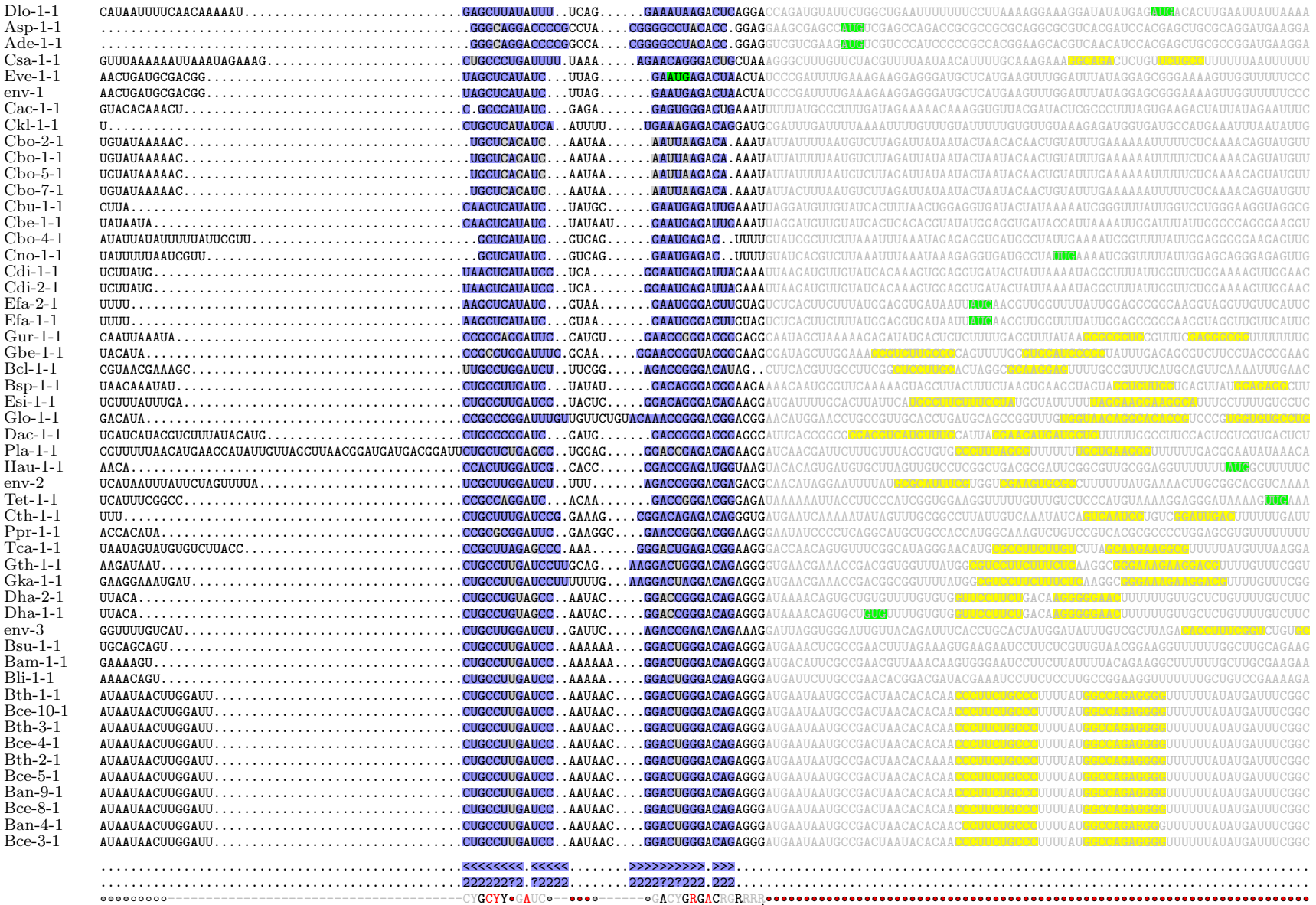
Each *pan* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 61.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *pan* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

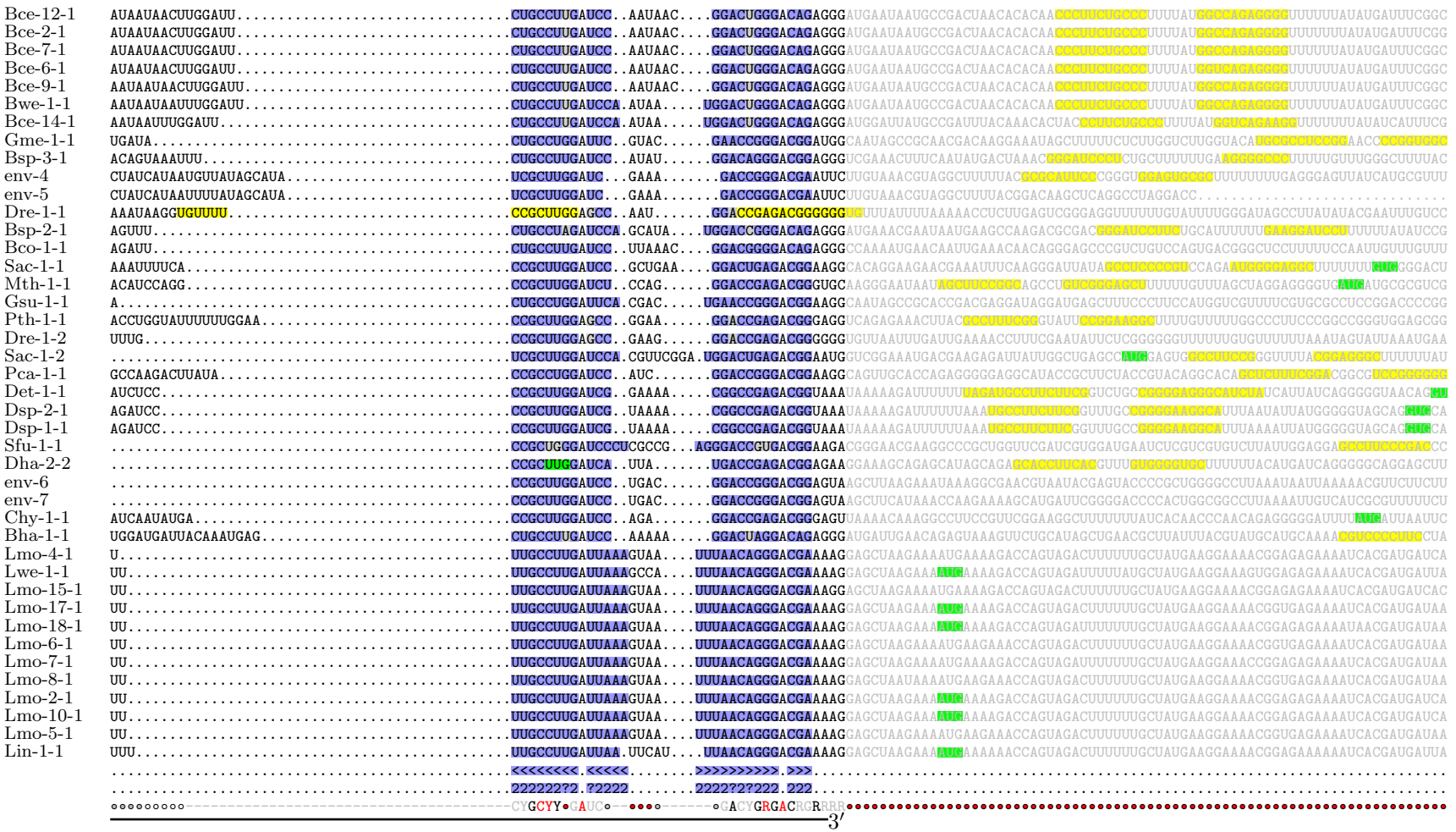



Ppr-1-1
Tca-1-1
Gth-1-1
Gka-1-1
Dha-2-1
Dha-1-1
env-3
Bsu-1-1
Bam-1-1
Bli-1-1
Bth-1-1
Bce-10-1
Bth-3-1
Bce-4-1
Bth-2-1
Bce-5-1
Ban-9-1
Bce-8-1
Ban-4-1
Bce-3-1
Bce-12-1
Bce-2-1
Bce-7-1
Bce-6-1
Bce-9-1
Bwe-1-1
Bce-14-1
Gme-1-1
Bsp-3-1
env-4
env-5
Dre-1-1
Bsp-2-1
Bco-1-1
Sac-1-1
Mth-1-1
Gsu-1-1
Pth-1-1
Dre-1-2
Sac-1-2
Pca-1-1
Det-1-1
Dsp-2-1
Dsp-1-1
Sfu-1-1
Dha-2-2
env-6
env-7
Chy-1-1
Bha-1-1
Lmo-4-1
Lwe-1-1
Lmo-15-1
Lmo-17-1
Lmo-18-1
Lmo-6-1
Lmo-7-1
Lmo-8-1
Lmo-2-1
Lmo-10-1
Lmo-5-1
5′

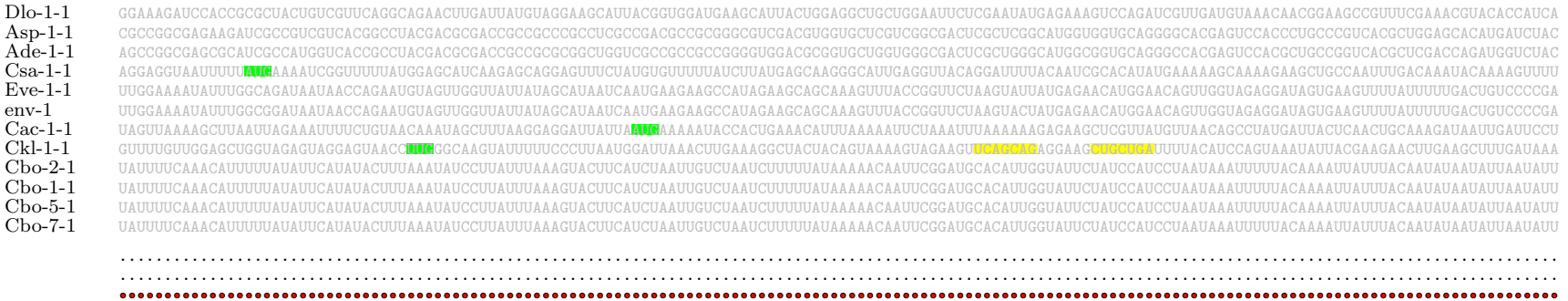


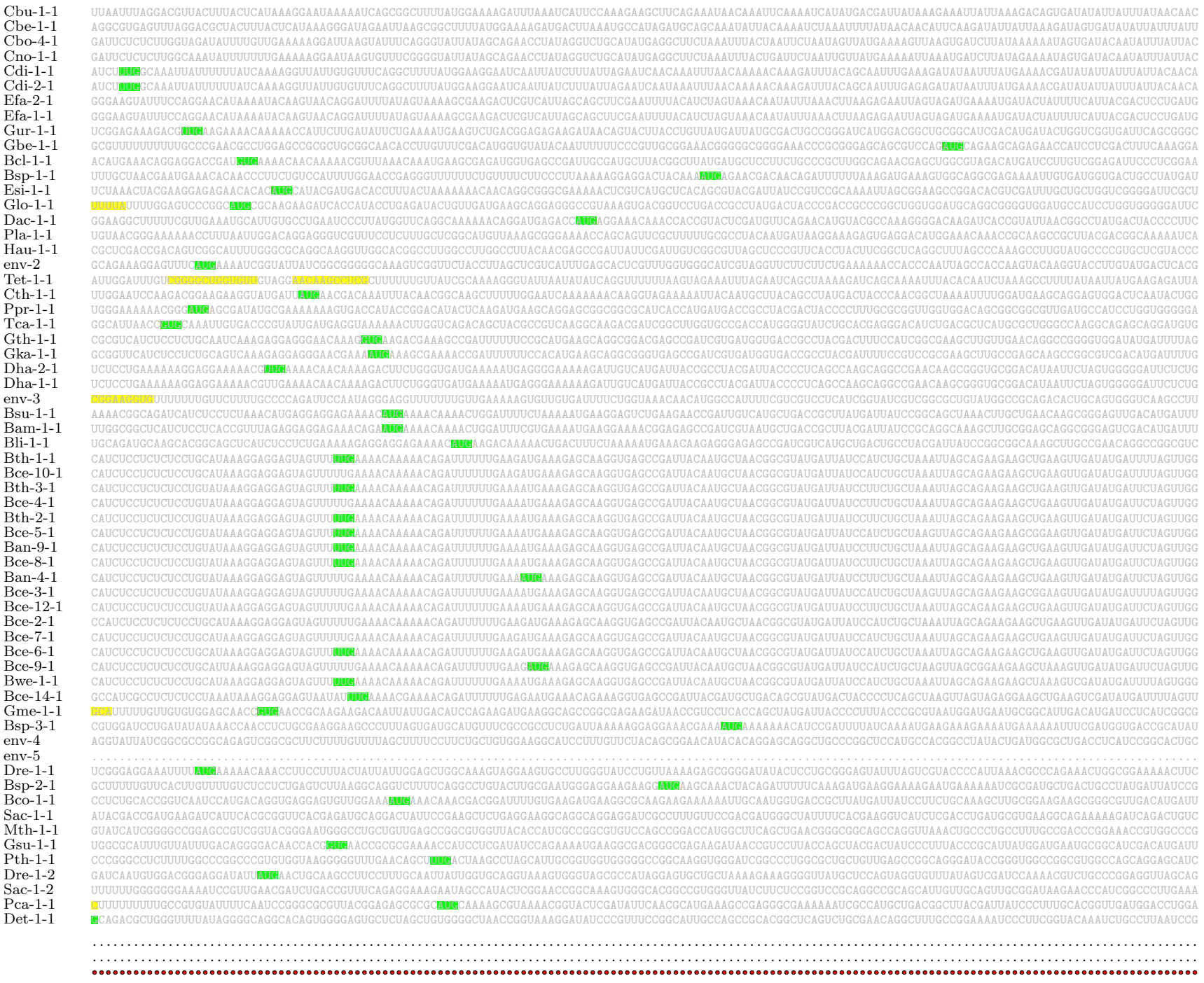
alignment positions 161 · · · 320

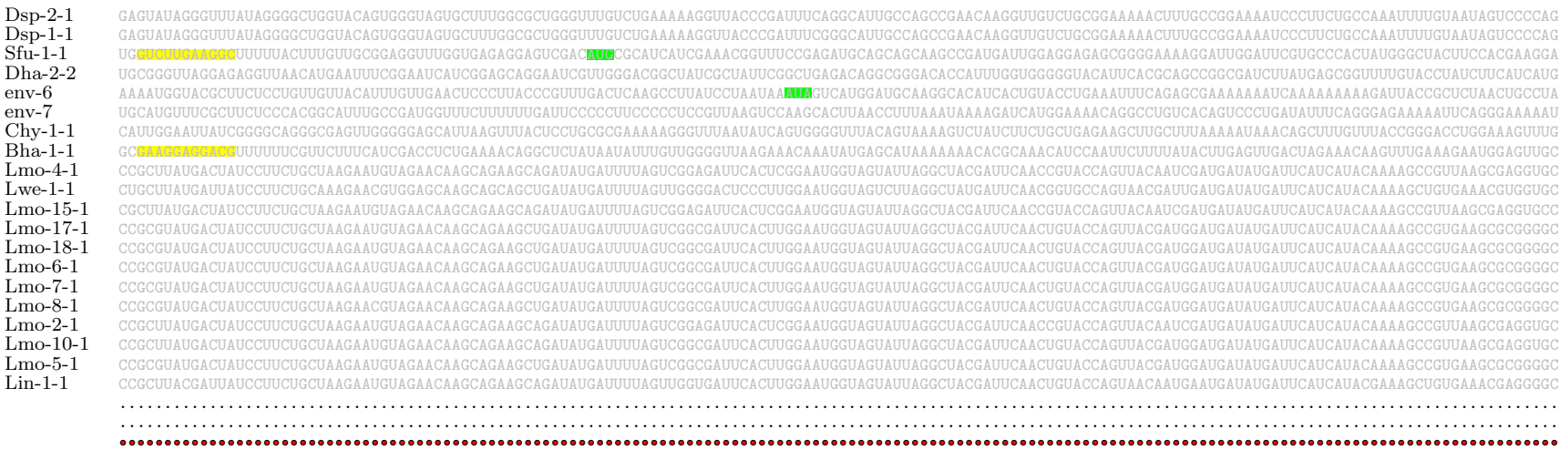




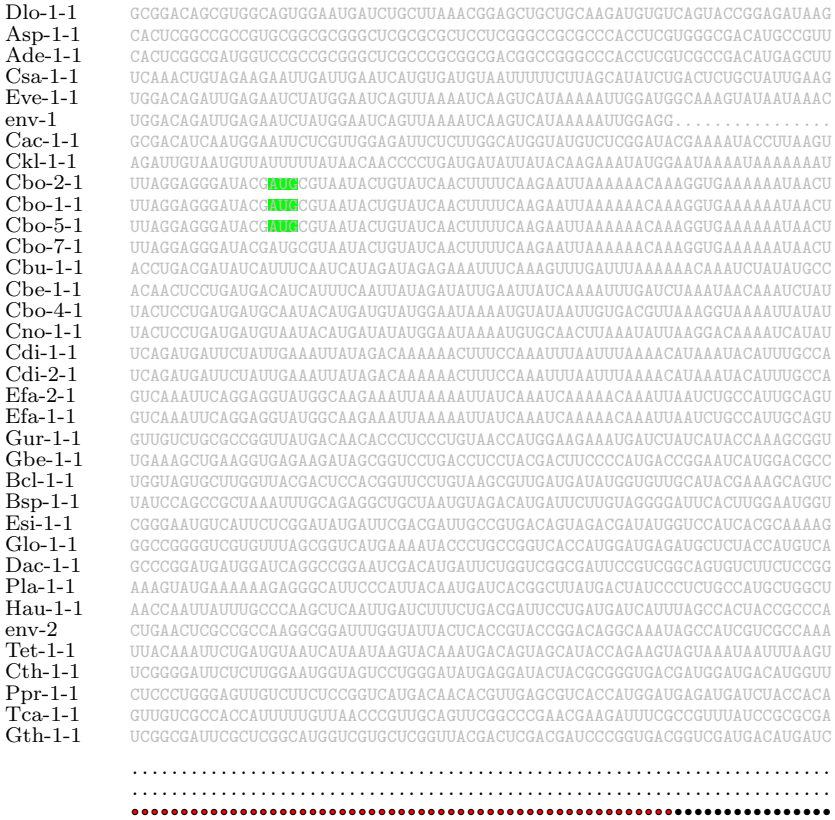
alignment positions 321 · · · 480

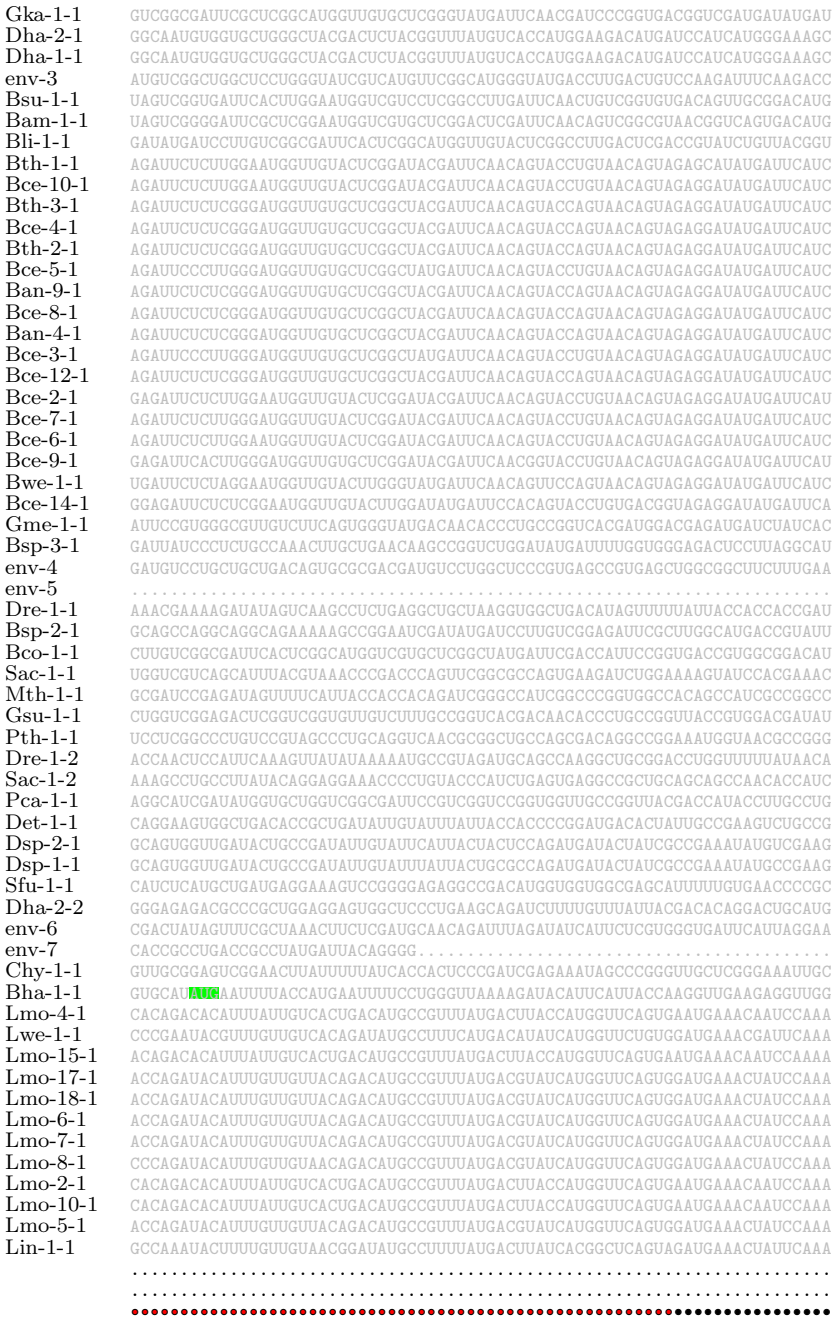



Cbu-1-1
Cbe-1-1
Cbo-4-1
Cno-1-1
Cdi-1-1
Cdi-2-1
Efa-2-1
Efa-1-1
Gur-1-1
Gbe-1-1
Bcl-1-1
Bsp-1-1
Esi-1-1
Glo-1-1
Dac-1-1
Pla-1-1
Hau-1-1
env-2
Tet-1-1
Cth-1-1
Ppr-1-1
Tca-1-1
Gth-1-1
Gka-1-1
Dha-2-1
Dha-1-1
env-3
Bsu-1-1
Bam-1-1
Bli-1-1
Bth-1-1
Bce-10-1
Bth-3-1
Bce-4-1
Bth-2-1
Bce-5-1
Ban-9-1
Bce-8-1
Ban-4-1
Bce-3-1
Bce-12-1
Bce-2-1
Bce-7-1
Bce-6-1
Bce-9-1
Bwe-1-1
Bce-14-1
Gme-1-1
Bsp-3-1
env-4
env-5
Dre-1-1
Bsp-2-1
Bco-1-1
Sac-1-1
Mth-1-1
Gsu-1-1
Pth-1-1
Dre-1-2
Sac-1-2
Pca-1-1
Det-1-1



alignment positions $481 \cdots 551$




Gka-1-1
Dha-2-1
Dha-1-1
env-3
Bsu-1-1
Bam-1-1
Bli-1-1
Bth-1-1
Bce-10-1
Bth-3-1
Bce-4-1
Bth-2-1
Bce-5-1
Ban-9-1
Bce-8-1
Ban-4-1
Bce-3-1
Bce-12-1
Bce-2-1
Bce-7-1
Bce-6-1
Bce-9-1
Bwe-1-1
Bce-14-1
Gme-1-1
Bsp-3-1
env-4
env-5
Dre-1-1
Bsp-2-1
Bco-1-1
Sac-1-1
Mth-1-1
Gsu-1-1
Pth-1-1
Dre-1-2
Sac-1-2
Pca-1-1
Det-1-1
Dsp-2-1
Dsp-1-1
Sfu-1-1
Dha-2-2
env-6
env-7
Chy-1-1
Bha-1-1
Lmo-4-1
Lwe-1-1
Lmo-15-1
Lmo-17-1
Lmo-18-1
Lmo-6-1
Lmo-7-1
Lmo-8-1
Lmo-2-1
Lmo-10-1
Lmo-5-1
Lin-1-1

62 *Pedo*-repair

62.1 Taxa

The taxonomy of each organism containing a putative *Pedo*-repair RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *Pedo*-repair RNA in Sections 62.2 and 62.4.

abbrev. of hits	taxonomy of species
Psp-1-1 to Psp-1-6	Bacteria Bacteroidetes Sphingobacteria Sphingobacteriales Sphingobacteriaceae *Pedobacter sp.* BAL39

62.2 Gene contexts

Each *Pedo*-repair RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *Pedo*-repair RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 62.3. The accession number of the sequence containing each *Pedo*-repair RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 62.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Psp-1-1	NZ_ABCM01000010.1	+	52472	52680	**RNA**→ hypo→ RadC (COG2003)→ hypo→
Psp-1-2	NZ_ABCM01000003.1	+	261261	261464	**RNA**→ hypo→ RadC (COG2003)→ hypo→
Psp-1-3	NZ_ABCM01000036.1	-	10445	10236	**RNA**→ hypo→
Psp-1-4	NZ_ABCM01000009.1	+	161544	161756	**RNA**→ hypo→ hypo→ RadC (pfam04002)→
Psp-1-5	NZ_ABCM01000001.1	+	430655	430865	**RNA**→ ←hypo
Psp-1-6	NZ_ABCM01000014.1	-	11514	11305	**RNA**→ hypo→

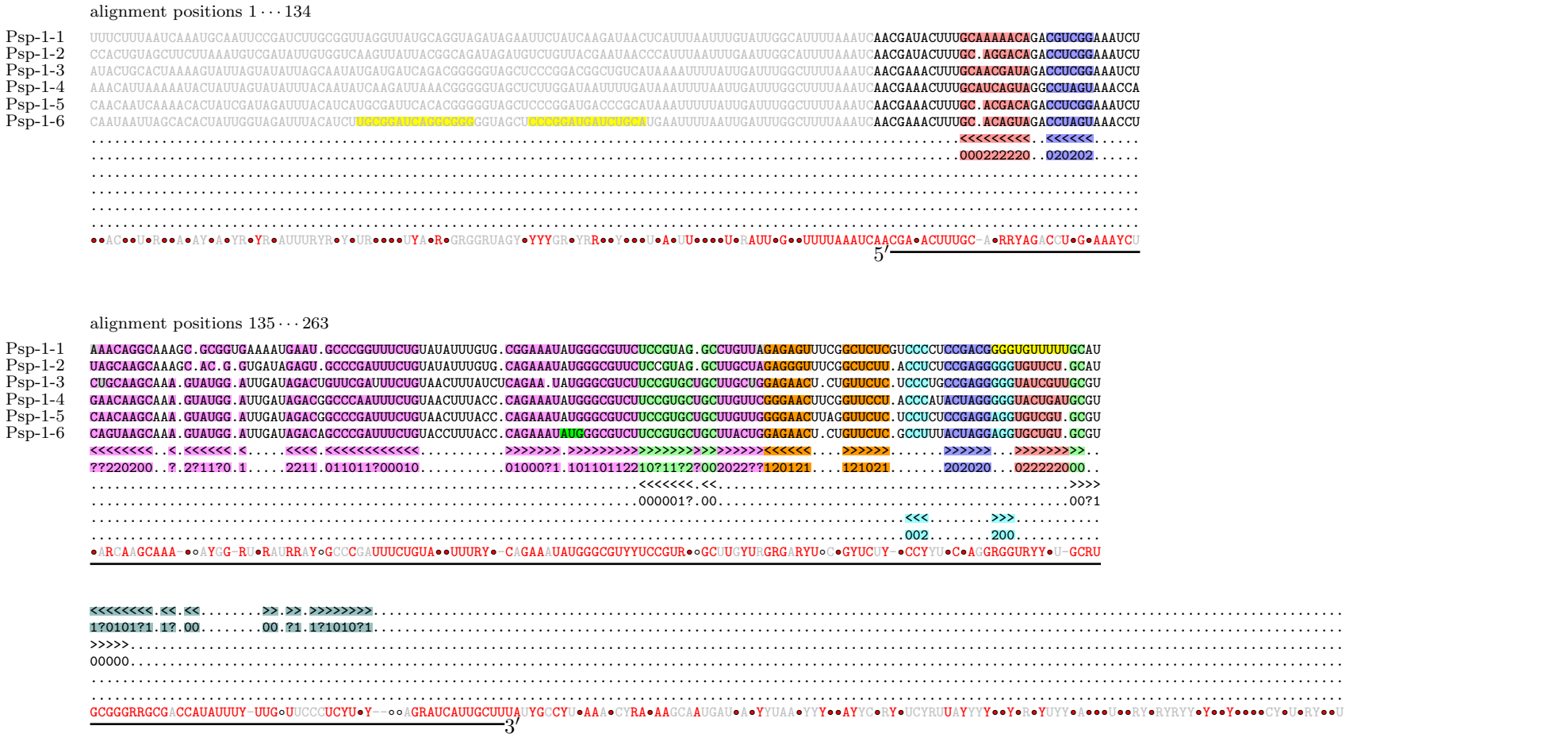
62.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 62.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *Pedo*-repair RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

COG2003 (2) DNA repair proteins [DNA replication, recombination, and repair]
pfam04002 (1) RadC, DNA repair protein.

62.4 Multiple-sequence alignment

Each *Pedo*-repair RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 62.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *Pedo*-repair RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



63 *pfl*

63.1 Taxa

The taxonomy of each organism containing a putative *pfl* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *pfl* RNA in Sections 63.3 and 63.5.

abbrev. of hits	taxonomy of species
Aod-1-1 to Aod-1-2	Bacteria Actinobacteria Actinobacteridae Actinomycetales Actinomycineae Actinomycetaceae *Actinomyces odontolyticus* ATCC 17982
Fal-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Frankineae Frankiaceae *Frankia alni* ACN14a
Fsp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Frankineae Frankiaceae *Frankia sp.* CcI3
Fsp-2-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Frankineae Frankiaceae *Frankia sp.* EAN1pec
Kra-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Frankineae Kineosporiaceae *Kineococcus radiotolerans* SRS30216
Jsp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Intrasporangiaceae *Janibacter sp.* HTCC2649
Cmi-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Microbacteriaceae *Clavibacter michiganensis subsp. michiganensis* NCPPB 382
Lxy-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Microbacteriaceae *Leifsonia xyli subsp. xyli str.* CTCB07
Aau-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Micrococcaceae *Arthrobacter aurescens* TC1
Asp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Micrococcaceae *Arthrobacter sp.* FB24
Sar-1-1 to Sar-1-2	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micromonosporineae Micromonosporaceae *Salinispora arenicola* CNS205
Str-1-1 to Str-1-2	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micromonosporineae Micromonosporaceae *Salinispora tropica* CNB-440
Nsp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Propionibacterineae Nocardioidaceae *Nocardioides sp.* JS614
Sav-1-1 to Sav-1-2	Bacteria Actinobacteria Actinobacteridae Actinomycetales Streptomycineae Streptomycetaceae *Streptomyces avermitilis* MA-4680
Sco-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Streptomycineae Streptomycetaceae *Streptomyces coelicolor* A3(2)
Bca-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides capillosus* ATCC 29799
Hau-1-1	Bacteria Chloroflexi Herpetosiphonales Herpetosiphonaceae *Herpetosiphon aurantiacus* ATCC 23779
Dge-1-1	Bacteria Deinococcus-Thermus Deinococci Deinococcales Deinococcaceae *Deinococcus geothermalis* DSM 11300
Dra-1-1	Bacteria Deinococcus-Thermus Deinococci Deinococcales Deinococcaceae *Deinococcus radiodurans* R1
Bcl-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus clausii* KSM-K16
Bsp-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus sp.* NRRL B-14911
Bsp-2-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus sp.* SG-1
Pla-1-1 to Pla-1-2	Bacteria Firmicutes Bacillales Paenibacillaceae *Paenibacillus larvae subsp. larvae* BRL-230010
Cac-1-1 to Cac-1-2	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium acetobutylicum* ATCC 824
Cbe-1-1 to Cbe-1-2	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium beijerinckii* NCIMB 8052
Cbo-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium bolteae* ATCC BAA-613
Cbo-2-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* A str. ATCC 19397
Cbo-3-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* A str. ATCC 3502
Cbo-4-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* A str. Hall
Cbo-5-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* Bf
Cbo-6-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* F str. Langeland
Cbo-7-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* G
Cbo-8-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* NCTC 2916
Cbo-9-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum str.* Iwanei E
Cbu-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium butyricum* 5521
Cce-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium cellulolyticum* H10
Ckl-1-1 to Ckl-1-2	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium kluyveri* DSM 555
Cph-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium phytofermentans* ISDg
Fpr-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Faecalibacterium prausnitzii* M21/2
Esi-1-1	Bacteria Firmicutes Clostridia Clostridiales Eubacteriaceae *Eubacterium siraeum* DSM 15702
Eve-1-1	Bacteria Firmicutes Clostridia Clostridiales Eubacteriaceae *Eubacterium ventriosum* ATCC 27560
Aca-1-1 to Aca-1-2	Bacteria Firmicutes Clostridia Clostridiales Lachnospiraceae *Anaerostipes caccae* DSM 14662
Dlo-1-1	Bacteria Firmicutes Clostridia Clostridiales Lachnospiraceae *Dorea longicatena* DSM 13814
Dre-1-1	Bacteria Firmicutes Clostridia Clostridiales Peptococcaceae *Desulfotomaculum reducens* MI-1
Rob-1-1 to Rob-1-2	Bacteria Firmicutes Clostridia Clostridiales Ruminococcaceae *Ruminococcus obeum* ATCC 29174
Tca-1-1	Bacteria Firmicutes Clostridia Clostridiales Veillonellaceae *Thermosinus carboxydivorans* Nor1
Hor-1-1	Bacteria Firmicutes Clostridia Halanaerobiales Halanaerobiaceae *Halothermothrix orenii* H 168
Edo-1-1	Bacteria Firmicutes Mollicutes *Eubacterium dolichum* DSM 3991
Ccr-1-1 to Ccr-1-2	Bacteria Proteobacteria Alphaproteobacteria Caulobacterales Caulobacteraceae *Caulobacter crescentus* CB15
Csp-1-1 to Csp-1-2	Bacteria Proteobacteria Alphaproteobacteria Caulobacterales Caulobacteraceae *Caulobacter sp.* K31
Bja-1-1 to Bja-1-2	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Bradyrhizobium japonicum* USDA 110
Bsp-3-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Bradyrhizobium sp.* BTAi1
Bsp-4-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Bradyrhizobium sp.* ORS278
Rpa-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* BisA53
Rpa-2-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* CGA009
Pla-2-1 to Pla-2-2	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Phyllobacteriaceae *Parvibaculum lavamentivorans* DS-1
Bam-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia ambifaria* AMMD
Bam-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia ambifaria* MC40-6
Bce-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* AU 1054
Bce-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* HI2424
Bce-3-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* MC0-3
Bce-4-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* PC184
Bdo-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia dolosa* AUO158
Bmu-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia multivorans* ATCC 17616

Bsp-5-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia sp.* 383
Bvi-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia vietnamiensis* G4
Bma-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* 2002721280
Bma-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* ATCC 10399
Bma-3-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* ATCC 23344
Bma-4-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* FMH
Bma-5-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* GB8 horse 4
Bma-6-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* JHU
Bma-7-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* NCTC 10229
Bma-8-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* NCTC 10247
Bma-9-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* PRL-20
Bma-10-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* SAVP1
Bph-1-1 to Bph-1-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia phymatum* STM815
Bph-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia phytofirmans* PsJN
Bub-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia ubonensis* Bu
Bxe-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia xenovorans* LB400
Bok-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia oklahomensis* C6786
Bok-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia oklahomensis* EO147
Bps-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1106a
Bps-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1106b
Bps-3-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 112
Bps-4-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 14
Bps-5-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1655
Bps-6-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1710a
Bps-7-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1710b
Bps-8-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 305
Bps-9-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 406e
Bps-10-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 668
Bps-11-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 7894
Bps-12-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 9
Bps-13-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 91
Bps-14-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* B7210
Bps-15-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* BCC215
Bps-16-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* DM98
Bps-17-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* K96243
Bps-18-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* NCTC 13177
Bps-19-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* Pasteur 52237
Bps-20-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* S13
Bth-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia thailandensis* Bt4
Bth-2-1 to Bth-2-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia thailandensis* E264
Bth-3-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia thailandensis* MSMB43
Bth-4-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia thailandensis* TXDOH
Reu-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Cupriavidus *Ralstonia eutropha* H16
Reu-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Cupriavidus *Ralstonia eutropha* JMP134
Rme-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Cupriavidus *Ralstonia metallidurans* CH34
Rpi-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Ralstonia pickettii* 12D
Rpi-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Ralstonia pickettii* 12J
Rso-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Ralstonia solanacearum* GMI1000
Rso-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Ralstonia solanacearum* UW551
Aav-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae *Acidovorax avenae subsp. citrulli* AAC00-1
Asp-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae *Acidovorax sp.* JS42
Cte-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae *Comamonas testosteroni* KF-1
Dac-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae *Delftia acidovorans* SPH-1
Pna-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae *Polaromonas naphthalenivorans* CJ2
Psp-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae *Polaromonas sp.* JS666
Rfe-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae *Rhodoferax ferrireducens* T118
Vei-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae *Verminephrobacter eiseniae* EF01-2
Mpe-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales *Methylibium petroleiphilum* PM1
Har-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Oxalobacteraceae *Herminiimonas arsenicoxydans*
Jsp-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Oxalobacteraceae *Janthinobacterium sp.* Marseille
Pca-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Pelobacteraceae *Pelobacter carbinolicus* DSM 2380
env-1 to env-93	environmental samples

63.2 Notes

Superscript numbers are used to annotate any *pfl* RNA that has special characteristics (described below). These numeric annotations will be used in Sections 63.3 and 63.5. The meaning of the numbers are as follows.

[1] This motif instance is presumed to be an ON switch. A predicted transcription terminator overlaps the 3′ end of the conserved motif. If ligand binding stabilizes the conserved structure, then high ligand concentrations destabalize the terminator.

[2] Similar to note 1, this motif instance is probably an ON switch. However, the predicted transcription terminator is 4 nucleotides away from the conserved structure, so the interaction between ligand binding and terminator formation is less clear.

[3] Unclear homology.

63.3 Gene contexts

Each *pfl* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *pfl* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 63.4. The accession number of the sequence containing each *pfl* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 63.1. Superscript numbers refer to annotations listed in Section 63.2

abbrev.	Seq. accession		5′ at	3′ at	genes
Bacteria Actinobacteria					
Kra-1-1	NC_009664.1	-	4174991	4174888	**RNA**→ GlyA (COG0112)→ FolD (COG0190)→ XthA (COG0708)→
Lxy-1-1	NC_006087.1	-	1921306	1921208	**RNA**→ GlyA (COG0112)→ FolD (COG0190)→ Glyoxalase (pfam00903)→
Aau-1-1	NC_008711.1	+	1308908	1308991	**RNA**→ GlyA (COG0112)→
Asp-1-1	NC_008541.1	+	1209742	1209827	**RNA**→ GlyA (COG0112)→ FolD (COG0190)→
Cmi-1-1	NC_009480.1	-	2838330	2838247	**RNA**→ GlyA (COG0112)→ FolD (COG0190)→
Aod-1-1	NZ_AAYI02000004.1	-	189562	189478	**RNA**→ GlyA (COG0112)→ FolD (COG0190)→
Sav-1-1	NC_003155.3	-	688247	688162	**RNA**→ FolD (COG0190)→
Str-1-1	NC_009380.1	-	2367924	2367834	**RNA**→ GlyA (COG0112)→
Sar-1-1	NZ_AAWA01000026.1	-	94720	94631	**RNA**→ GlyA (COG0112)→
Jsp-1-1	NZ_AAMN01000004.1	+	194357	194435	**RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→ hypo→
Sar-1-2	NZ_AAWA01000025.1	+	75111	75190	**RNA**→ FolD (COG0190)→
Nsp-1-1	NC_008699.1	-	3801746	3801670	**RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→ FolD (COG0190)→ putative integral membrane protein→ hypo→ hypo→
Aod-1-2	NZ_AAYI02000004.1	-	2328470	2328386	**RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Fsp-1-1	NC_007777.1	+	708324	708410	**RNA**→ GlyA (COG0112)→
Fal-1-1	NC_008278.1	+	1191460	1191545	**RNA**→ ←hypo
Fsp-2-1	NZ_AAII01000097.1	-	5218	5130	**RNA**→ ←hypo
Sav-1-2	NC_003155.3	-	4242005	4241916	**RNA**→ GlyA (COG0112)→
Sco-1-1	NC_003888.3	+	5269161	5269250	**RNA**→ GlyA (COG0112)→
Str-1-2	NC_009380.1	-	2571803	2571720	**RNA**→ FolD (COG0190)→
Bacteria Bacteroidetes					

[1]Bca-1-1	NZ_AAXG02000001.1	+	53805	53882	**RNA**→ FTHFS (COG2759)→
Bacteria Chloroflexi					
Hau-1-1	NZ_AATI01000112.1	-	11951	11866	**RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→ HisI (COG0140)PRA-CH (pfam01502)→ WecG (COG1922)→ hypo→ COG3603 (COG3603)→ SpoIIAA (COG1366)GAF (smart00065)→ SpoIIAA (COG1366)GAF (smart00065)→
Bacteria Deinococcus-Thermus					
Dra-1-1	NC_001263.1	-	880781	880696	**RNA**→ IMPCH (cd01421)→ FolD (COG0190)→ RfaG (COG0438)→
Dge-1-1	NC_008025.1	+	539910	539993	**RNA**→ IMPCH (cd01421)→ FolD (COG0190)→
Bacteria Firmicutes Bacillales					
Bsp-2-1	NZ_ABCF01000005.1	-	78206	78117	**RNA**→ FolD (COG0190)→
Bsp-1-1	NZ_AAOX01000008.1	+	65010	65093	**RNA**→ FolD (COG0190)→
[2]Pla-1-1	NZ_AARF01000325.1	-	402	322	**RNA**→
[2]Pla-1-2	NZ_AARF01000447.1	-	2648	2568	**RNA**→
Bcl-1-1	NC_006582.1	-	3516508	3516431	**RNA**→ FolD (COG0190)→
Bacteria Firmicutes Clostridia					
[1]Fpr-1-1	NZ_ABED01000023.1	-	84011	83925	**RNA**→
[1]Esi-1-1	NZ_ABCA01000033.1	-	43572	43468	**RNA**→
Dre-1-1	NC_009253.1	-	48258	48175	**RNA**→ PflD (COG1882)→ PflA (COG1180)→
Cbo-1-1	NZ_ABCC01000056.1	+	15768	15857	**RNA**→
[1]Rob-1-1	NZ_AAVO02000015.1	+	35242	35324	**RNA**→ hypo→ FTHFS (COG2759)→
[1]Dlo-1-1	NZ_AAXB02000001.1	+	5878	5961	**RNA**→ hypo→ hypo→ FTHFS (COG2759)→
[1]Tca-1-1	NZ_AAWL01000017.1	-	16393	16312	**RNA**→ PflD (COG1882)→ PflA (COG1180)→
[1]Hor-1-1	NZ_AAOZ01000030.1	-	13151	13063	**RNA**→ PflD (COG1882)→ PflA (COG1180)→
Cph-1-1	NZ_AAQT01000028.1	-	28984	28882	**RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
[1]Aca-1-1	NZ_ABAX01000052.1	-	14224	14145	**RNA**→
[1]Cac-1-1	NC_003030.1	+	2562148	2562228	**RNA**→ AICARFT_IMPCHas (pfam01808)→
Ckl-1-1	NC_009706.1	+	3903782	3903862	**RNA**→ hypo→ HutG (COG3741)→ UbiE (COG2226)→
[1]Aca-1-2	NZ_ABAX01000016.1	-	73917	73833	**RNA**→
Cbe-1-1	NC_009617.1	+	1211931	1212029	**RNA**→ PflD (COG1882)→
[1]Ckl-1-2	NC_009706.1	-	1156219	1156141	**RNA**→ PflD (COG1882)→ PflA (COG1180)→
[1]Cac-1-2	NC_003030.1	+	1127408	1127485	**RNA**→ PflD (COG1882)→ PflA (COG1180)→
[1]Cbo-6-1	NC_009699.1	-	3519944	3519854	**RNA**→ PflD (COG1882)→
[1]Cbo-8-1	NZ_ABDO01000004.1	-	3776	3686	**RNA**→
Cbo-9-1	NZ_ABDR01000012.1	-	89895	89805	**RNA**→
Cbo-7-1	NZ_ABDS01000001.1	+	290591	290681	**RNA**→
[1]Cbo-5-1	NZ_ABDP01000011.1	-	87327	87237	**RNA**→
[1]Cbo-3-1	NC_009495.1	-	3426738	3426648	**RNA**→ PflD (COG1882)→
Cbo-2-1	NC_009697.1	-	3402727	3402637	**RNA**→ PflD (COG1882)→
Cbo-4-1	NC_009698.1	-	3299834	3299744	**RNA**→ PflD (COG1882)→
Cbe-1-2	NC_009617.1	-	4111764	4111685	**RNA**→ AICARFT_IMPCHas (pfam01808)→
Cbu-1-1	NZ_ABDT01000011.1	-	36280	36201	**RNA**→
[1]Cce-1-1	NZ_AAVC01000001.1	-	179151	179072	**RNA**→ PflD (COG1882)→
[1]Eve-1-1	NZ_AAVL02000036.1	-	176021	175941	**RNA**→ hypo→
[1]Rob-1-2	NZ_AAVO02000004.1	+	47771	47851	**RNA**→ COG3363 (COG3363)→ AICARFT_IMPCHas (pfam01808)→ hypo→
Bacteria Firmicutes Mollicutes					
Edo-1-1	NZ_ABAW01000002.1	-	38872	38789	**RNA**→
Bacteria Proteobacteria Alphaproteobacteria					

[1]Pla-2-1	NC_009719.1	-	674390	674303	**RNA**→ GlyA (COG0112)→
Ccr-1-1	NC_002696.2	+	32806	32887	**RNA**→ ←Fiu (COG4774)
Csp-1-1	NZ_AATH01000002.1	-	35515	35436	**RNA**→ RhtB (COG1280)→
Rpa-1-1	NC_008435.1	-	30517	30436	**RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Bja-1-1	NC_004463.1	+	625249	625331	**RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Rpa-2-1	NC_005296.1	-	33689	33607	**RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Bsp-4-1	NC_009445.1	-	352539	352459	**RNA**→ IMPCH (cd01421)AICARFT_IMPCHas (pfam01808)→
Bsp-3-1	NC_009485.1	-	329370	329290	**RNA**→ IMPCH (cd01421)AICARFT_IMPCHas (pfam01808)→
[1]Ccr-1-2	NC_002696.2	+	88947	89027	**RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→ COG0412 (COG0412)→
Csp-1-2	NZ_AATH01000001.1	-	292277	292198	**RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Pla-2-2	NC_009719.1	+	1504618	1504701	**RNA**→ FolD (COG0190)→ Tas (COG0667)→
Bja-1-2	NC_004463.1	-	6500011	6499926	**RNA**→ GlyA (COG0112)→
Bacteria Proteobacteria Betaproteobacteria					
Har-1-1	NC_009138.1	-	2356811	2356707	**RNA**→ GlyA (COG0112)→
Jsp-2-1	NC_009659.1	-	2768093	2767971	**RNA**→ GlyA (COG0112)→
Mpe-1-1	NC_008825.1	+	3122956	3123073	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Pna-1-1	NC_008781.1	+	3062025	3062156	**RNA**→ GlyA (COG0112)→
Cte-1-1	NZ_AAUJ01000003.1	-	402218	402064	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Dac-1-1	NZ_AAVD01000001.1	-	886035	885860	**RNA**→ GlyA (COG0112)→
Psp-1-1	NC_007948.1	+	3034848	3034979	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Asp-2-1	NC_008782.1	-	1721986	1721827	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Aav-1-1	NC_008752.1	+	3298357	3298515	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Vei-1-1	NC_008786.1	+	4684351	4684518	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Rfe-1-1	NC_007908.1	+	2925225	2925356	**RNA**→ GlyA (COG0112)→
Bph-1-1	NZ_AAUG01000006.1	-	223188	223068	**RNA**→ RhtB (COG1280)→
Reu-1-1	NC_008313.1	+	3060261	3060406	**RNA**→ GlyA (COG0112)→
Rme-1-1	NC_007973.1	+	2912877	2912979	**RNA**→ GlyA (COG0112)→
Reu-2-1	NC_007347.1	-	861611	861511	**RNA**→ GlyA (COG0112)→
Rso-2-1	NZ_AAKL01000005.1	+	56385	56514	**RNA**→ GlyA (COG0112)→
Rso-1-1	NC_003295.1	-	778049	777920	**RNA**→ GlyA (COG0112)→
Rpi-1-1	NZ_ABDZ01000039.1	-	2612	2483	**RNA**→ GlyA (COG0112)→
Rpi-2-1	NZ_AAWK01000010.1	+	84446	84574	**RNA**→ GlyA (COG0112)→
Bps-5-1	NZ_AAHR02000001.1	+	112548	112648	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bps-16-1	NZ_ABBI01001009.1	-	7408	7308	**RNA**→
Bps-17-1	NC_006350.1	-	3298400	3298300	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bps-7-1	NC_007434.1	-	3557008	3556908	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bps-10-1	NC_009074.1	-	3132286	3132186	**RNA**→ ←hypo
Bma-10-1	NC_008785.1	+	834517	834617	**RNA**→ ←hypo
Bma-8-1	NC_009080.1	-	1919893	1919793	**RNA**→ ←hypo
Bps-1-1	NC_009076.1	-	3148595	3148495	**RNA**→ ←hypo
Bma-7-1	NC_008836.1	+	2699085	2699185	**RNA**→ ←hypo
Bma-3-1	NC_006348.1	-	2172337	2172237	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bma-2-1	NZ_AAHN02000003.1	+	185000	185100	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bma-5-1	NZ_AAHO01000001.1	-	161030	160930	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bps-6-1	NZ_AAHS02000002.1	-	201942	201842	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bps-20-1	NZ_AAHW02000016.1	+	90806	90906	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bma-4-1	NZ_AAIQ02000001.1	-	143315	143215	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bma-6-1	NZ_AAIR02000001.1	-	154654	154554	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bps-2-1	NZ_AAMB01000001.1	-	303588	303488	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bps-9-1	NZ_AAMM02000001.1	-	179260	179160	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bma-1-1	NZ_AANX02000052.1	+	12272	12372	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→

Bps-8-1	NZ_AAYX01000005.1	-	205375	205275	**RNA**→ ←hypo
Bma-9-1	NZ_AAZP01000001.1	-	11062	10962	**RNA**→
Bps-4-1	NZ_ABBJ01001847.1	+	926	1026	**RNA**→
Bps-13-1	NZ_ABBK01000750.1	-	9165	9065	**RNA**→
Bps-14-1	NZ_ABBN01000630.1	-	10713	10613	**RNA**→
Bps-18-1	NZ_ABBQ01000477.1	-	10831	10731	**RNA**→
Bps-19-1	NZ_AAHV02000001.1	+	89560	89660	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bps-12-1	NZ_ABBL01000768.1	-	9786	9686	**RNA**→
Bps-3-1	NZ_ABBP01000556.1	-	10728	10628	**RNA**→
Bps-15-1	NZ_ABBR01000430.1	-	9813	9713	**RNA**→
Bps-11-1	NZ_ABBO01000706.1	-	10613	10513	**RNA**→
Bok-2-1	NZ_ABBF01000384.1	-	15529	15428	**RNA**→
Bok-1-1	NZ_ABBG01000261.1	-	19106	19005	**RNA**→
Bth-3-1	NZ_ABBM01000487.1	-	14069	13969	**RNA**→
Bth-2-1	NC_007651.1	+	1555456	1555556	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bth-2-2	NZ_AACX01000007.1	+	312508	312608	**RNA**→
Bth-4-1	NZ_ABBD01000528.1	+	32605	32705	**RNA**→
Bth-1-1	NZ_ABBH01000544.1	+	39713	39813	**RNA**→
Bsp-5-1	NC_007510.1	+	810912	811012	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bdo-1-1	NZ_AAKY01000114.1	-	32845	32745	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bvi-1-1	NC_009256.1	+	815126	815226	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bam-1-1	NC_008390.1	+	762627	762727	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bam-2-1	NZ_AAUZ01000015.1	+	20320	20420	**RNA**→ ←hypo (cons)
Bce-4-1	NZ_AAKX01000025.1	-	10448	10348	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bce-3-1	NZ_AAVA01000023.1	+	152021	152121	**RNA**→ ←hypo (cons)
Bce-1-1	NC_008060.1	+	359253	359353	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bce-2-1	NC_008542.1	+	907214	907314	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bub-1-1	NZ_ABBE01000365.1	+	29421	29520	**RNA**→
Bmu-1-1	NZ_AAVB01000011.1	-	74866	74767	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bph-1-2	NZ_AAUG01000007.1	+	135554	135668	**RNA**→ ←COG4221 (COG4221)
Bph-2-1	NZ_AAUH01000005.1	-	372359	372251	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bxe-1-1	NC_007951.1	+	613445	613553	**RNA**→ GlyA (COG0112)→ ATP-cone (pfam03477)→
Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales					
[1]Pca-1-1	NC_007498.2	+	3231175	3231262	**RNA**→ COG0075 (COG0075)→
environmental					
env-1	JCVI_SCAF_1096626882634	+	1074	1209	**RNA**→ SHMT (COG0112)? (COG0112)→
env-2	JCVI_SCAF_1101668120280	-	255	91	**RNA**→
env-3	JCVI_SCAF_1096627597967	-	1537	1424	**RNA**→ ←ORF
env-4	JCVI_SCAF_1101667483037	+	109	271	**RNA**→ ←ORF
env-5	JCVI_SCAF_1101667065799	+	413	611	**RNA**→
[1]env-6	BAAV01000055.1	+	2897	2982	**RNA**→ PurE (COG0041)→ PurM (COG0150)→ FocA (COG2116)→ Formyl_trans_N (pfam00551)→ COG3363 (COG3363)→ AICARFT_IMPCHas (pfam01808)→ PurD (COG0151)→ GATase1_FGAR_AT (cd01740)PurL (COG0046)→
[1]env-7	BAAU01012305.1	-	993	909	**RNA**→ PurE (COG0041)→ PurM_N (cd02196)→
env-8	JCVI_SCAF_1096627470168	-	1159	963	**RNA**→ ←ORF
env-9	soil_XZS116095_g1	-	262	102	**RNA**→
[1]env-10	BABG01015168.1	+	550	636	**RNA**→ hypo→
env-11	BAAY01025872.1	-	84	1	**RNA**→
[1]env-12	BAAU01017791.1	-	392	305	**RNA**→ hypo→
[1]env-13	BAAY01004625.1	+	493	580	**RNA**→ hypo→ hypo→
[1]env-14	BAAU01002505.1	+	255	341	**RNA**→ hypo→ MIS1 (COG2759)→

[1]env-15	BAAX01023463.1	-	777	691	**RNA**→ hypo→
[1]env-16	BAAU01027299.1	+	88	174	**RNA**→
env-17	hgutS7_s7_163498	-	535	449	**RNA**→ COG3363 (COG3363)→
env-18	JCVI_SCAF_1096627480979	+	474	633	**RNA**→ ←ORF
env-19	sludgePhrap_Contig5234	-	252	87	**RNA**→
env-20	JCVI_SCAF_1096628077994	+	1289	1442	**RNA**→ ←ORF
env-21	JCVI_SCAF_1096627425047	+	594	676	**RNA**→ ←ORF
env-22	JCVI_SCAF_1096627352220	-	190	107	**RNA**→
env-23	JCVI_SCAF_1101668620448	-	490	407	**RNA**→ ←ORF
env-24	JCVI_SCAF_1096627136696	+	43	126	**RNA**→
env-25	JCVI_SCAF_1101668652585	+	1152	1235	**RNA**→
env-26	JCVI_SCAF_1096627368357	+	444	527	**RNA**→ ←ORF
env-27	JCVI_SCAF_1096627933793	+	416	499	**RNA**→ ←ORF
env-28	JCVI_SCAF_1096627350145	+	559	642	**RNA**→ ←ORF
env-29	JCVI_SCAF_1096627935183	+	74	157	**RNA**→
env-30	JCVI_SCAF_1096627947405	-	321	238	**RNA**→
env-31	JCVI_SCAF_1096627930887	-	190	107	**RNA**→
env-32	JCVI_SCAF_1101668614554	-	1083	1000	**RNA**→ ←ORF
env-33	JCVI_SCAF_1096626701152	-	572	489	**RNA**→ ←ORF
env-34	JCVI_SCAF_1096627951868	-	503	420	**RNA**→ ←ORF
env-35	JCVI_SCAF_1101668606318	-	1144	1061	**RNA**→ ←ORF
env-36	JCVI_SCAF_1096627150400	+	1673	1756	**RNA**→ ←ORF
env-37	JCVI_SCAF_1096626377935	-	1198	1115	**RNA**→ ←ORF
env-38	JCVI_SCAF_1096627932431	+	1010	1093	**RNA**→ ←ORF
env-39	JCVI_SCAF_1101667313259	+	566	649	**RNA**→
env-40	JCVI_SCAF_1096627368011	+	41	124	**RNA**→
env-41	JCVI_SCAF_1096626706762	-	611	528	**RNA**→ ←ORF
env-42	JCVI_SCAF_1096628402287	-	421	338	**RNA**→ ←ORF
env-43	JCVI_SCAF_1101668616378	-	335	252	**RNA**→
env-44	JCVI_SCAF_1096627354086	+	1032	1115	**RNA**→ ←ORF
env-45	JCVI_SCAF_1096627366227	+	826	909	**RNA**→ ←ORF
env-46	JCVI_SCAF_1096626704610	-	883	800	**RNA**→ ←ORF
env-47	BABF01003966.1	+	1020	1104	**RNA**→ GlyA (COG0112)→
env-48	JCVI_SCAF_1101667568502	-	491	314	**RNA**→ ←ORF
env-49	BABD01019874.1	+	885	962	**RNA**→
[1]env-50	BABD01002377.1	+	1579	1672	**RNA**→
[1]env-51	BABD01022060.1	-	785	692	**RNA**→ hypo→
[1]env-52	BAAY01025330.1	-	152	59	**RNA**→
[1]env-53	BABA01016569.1	+	75	168	**RNA**→
[1]env-54	BAAX01001391.1	+	197	281	**RNA**→ PurE (COG0041)→ SAICAR_synt (cd00476)→ PurF (COG0034)→ PurM_N (cd02196)→
env-55	BAAX01000081.1	-	4230	4146	**RNA**→ NarK (COG2223)→
env-56	JCVI_SCAF_1096627941471	+	1439	1523	**RNA**→
env-57	hgutS8_s8_161320	-	670	587	**RNA**→ COG3363 (COG3363)→
[1]env-58	ABEF01039966.1	+	1068	1153	**RNA**→
[1]env-59	BAAV01020765.1	-	806	725	**RNA**→ hypo→
[1]env-60	BAAY01013765.1	+	896	1010	**RNA**→
[1]env-61	ABEF01017404.1	-	1032	947	**RNA**→ MetH (COG0646)→
env-62	hgutS7_s7_172793	-	471	387	**RNA**→ Pyruvate-formate lyase(EC:2.3.1.54)→
[1]env-63	BABE01005334.1	-	246	164	**RNA**→
env-64	hgutS7_s7_161798	+	259	341	**RNA**→ FocA (COG2116)→
[1]env-65	BAAY01012400.1	-	201	108	**RNA**→
env-66	BABF01008002.1	-	701	620	**RNA**→

[1]env-67	BABA01021702.1	+	659	750	**RNA**→
[1]env-68	BAAY01000038.1	+	8293	8388	**RNA**→ PFL1 (COG1882)→
[1]env-69	BAAV01000327.1	-	435	340	**RNA**→ hypo→
[1]env-70	BABG01000244.1	+	657	739	**RNA**→ ←hypo
env-71	hgutS8_s8_178015	-	981	899	**RNA**→ ←Glycosidases(EC:3.2.1.10)
env-72	hgutS8_s8_168186	+	1291	1373	**RNA**→ PflA (COG1180)→
env-73	BABD01005042.1	+	856	936	**RNA**→ COG3363 (COG3363)→
[3]env-74	BABG01018752.1	-	281	200	**RNA**→
env-75	hgutS8_s8_159115	+	335	416	**RNA**→ hypo→ MIS1 (COG2759)→
env-76	hgutS7_s7_161123	+	401	486	**RNA**→ AICARFT_IMPCHas (pfam01808)→
env-77	hgutS7_s7_160400	+	824	909	**RNA**→
env-78	BAAV01011945.1	-	1060	974	**RNA**→ AICARFT_IMPCHas (pfam01808)→
env-79	BAAX01007901.1	+	363	439	**RNA**→ COG3363 (COG3363)→
env-80	BABG01000490.1	+	211	374	**RNA**→ COG3363 (COG3363)→
env-81	hgutS8_s8_168764	-	915	834	**RNA**→ COG3363 (COG3363)→
env-82	BAAX01006362.1	-	459	383	**RNA**→ hypo→
[1]env-83	BABG01005810.1	+	129	214	**RNA**→ hypo→
[1]env-84	BAAY01015723.1	+	323	408	**RNA**→ ←hypo
[1]env-85	BABA01029967.1	-	180	94	**RNA**→
env-86	BAAW01014873.1	-	349	270	**RNA**→
env-87	BAAW01002219.1	-	737	658	**RNA**→ RhtB (COG1280)→
env-88	sludgePhrap_Contig4910	-	725	609	**RNA**→ SHMT (COG0112)→
env-89	JCVI_SCAF_1101668310123	-	158	61	**RNA**→
env-90	JCVI_SCAF_1096626951197	+	280	377	**RNA**→ ORF→ GlyA (COG0112)? (COG0112)→ ORF→
env-91	JCVI_SCAF_1096627376771	+	323715	323815	**RNA**→ ←ORF
env-92	JCVI_SCAF_1096627190670	+	44626	44734	**RNA**→ ←COG4221 (COG4221)
env-93	wf2_AHAA11011_y1	+	151	240	**RNA**→ GcvT (COG0404)→

63.4 Conserved domains

Conserved domains found in protein-coding genes listed in Section 63.3 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *pfl* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

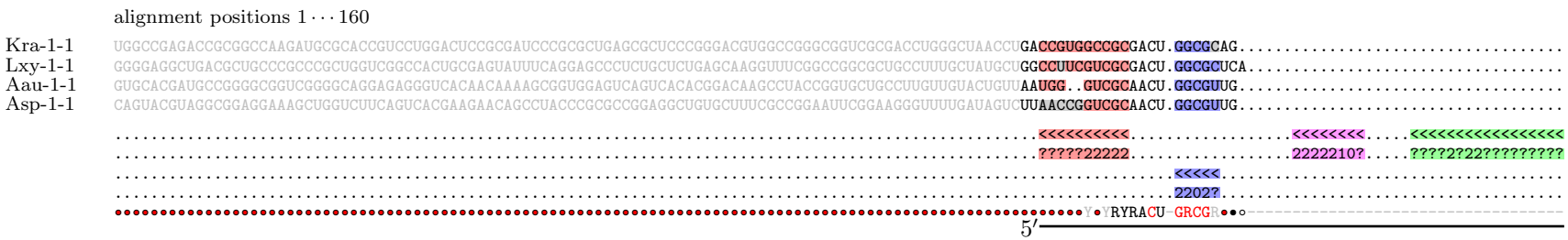
cd00476 (1) SAICAR synthetase catalyzes the seventh step of the de novo biosynthesis of purine nucleotides, the conversion of carboximideaminoimidazole ribonucleotide (CAIR) into succinoaminoimidazolecarboximide ribonucleotide (SAICAR).
cd01421 (4) Inosine monophosphate cyclohydrolase domain.
cd01740 (1) Type 1 glutamine amidotransferase (GATase1)-like domain found in Formylglycinamide ribonucleotide amidotransferase (FGAR-AT).
cd02196 (2) PurM (Aminoimidazole Ribonucleotide [AIR]
COG0034 (1) Glutamine phosphoribosylpyrophosphate amidotransferase [Nucleotide transport and metabolism]
COG0041 (3) Phosphoribosylcarboxyaminoimidazole (NCAIR) mutase [Nucleotide transport and metabolism]
COG0046 (1) Phosphoribosylformylglycinamidine (FGAM) synthase, synthetase domain [Nucleotide transport and metabolism]
COG0075 (1) Serine-pyruvate aminotransferase/archaeal aspartate aminotransferase [Amino acid transport and metabolism]
COG0112 (62) Glycine/serine hydroxymethyltransferase [Amino acid transport and metabolism]
COG0140 (1) Phosphoribosyl-ATP pyrophosphohydrolase [Amino acid transport and metabolism]
COG0150 (1) Phosphoribosylaminoimidazole (AIR) synthetase [Nucleotide transport and metabolism]
COG0151 (1) Phosphoribosylamine-glycine ligase [Nucleotide transport and metabolism]
COG0190 (15) 5,10-methylene-tetrahydrofolate dehydrogenase/Methenyl tetrahydrofolate cyclohydrolase [Coenzyme metabolism]
COG0404 (1) Glycine cleavage system T protein (aminomethyltransferase) [Amino acid transport and metabolism]
COG0412 (1) Dienelactone hydrolase and related enzymes [Secondary metabolites biosynthesis, transport, and catabolism]
COG0438 (1) Glycosyltransferase [Cell envelope biogenesis, outer membrane]

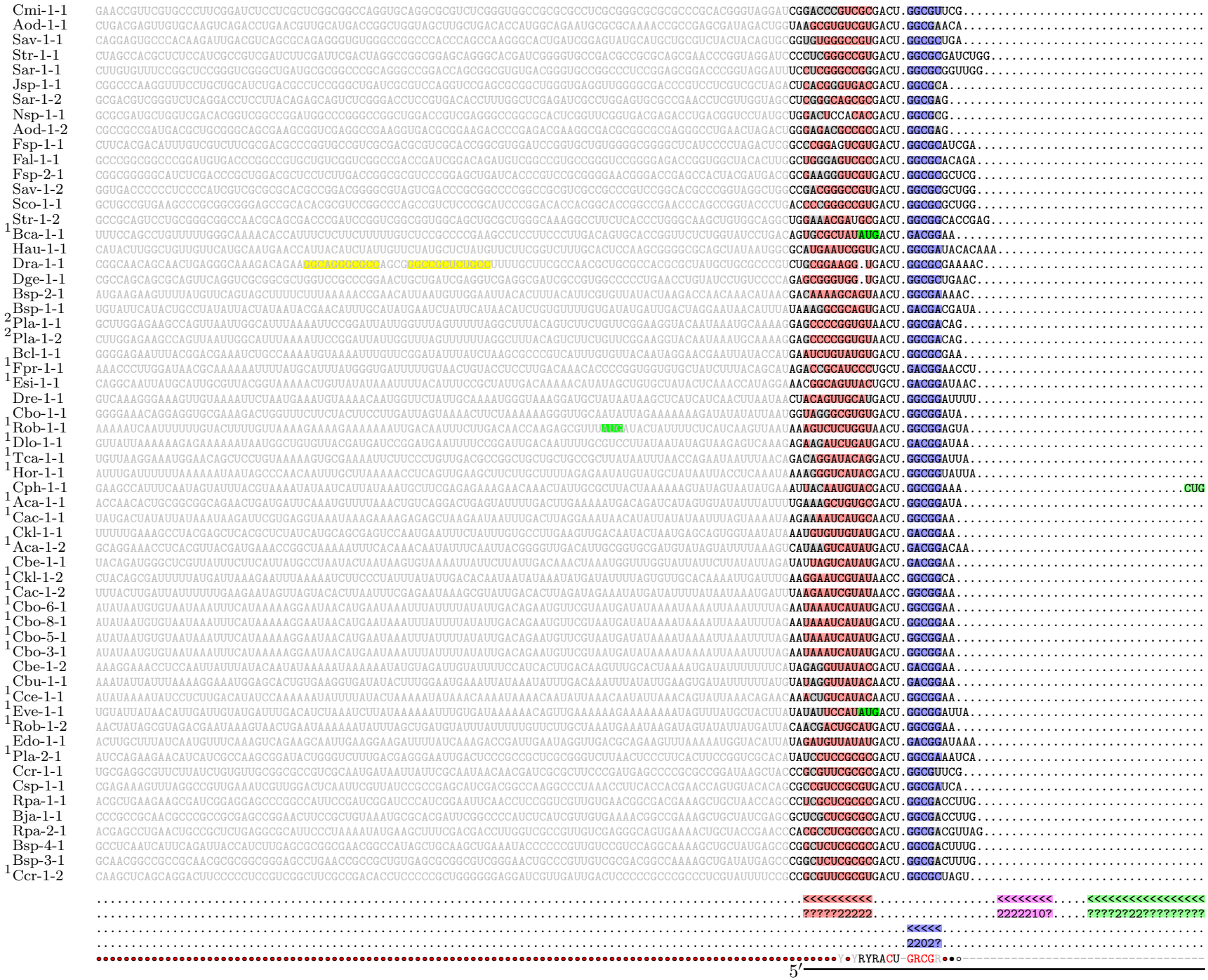
COG0646 (1) Methionine synthase I (cobalamin-dependent), methyltransferase domain [Amino acid transport and metabolism]
COG0667 (1) Predicted oxidoreductases (related to aryl-alcohol dehydrogenases) [Energy production and conversion]
COG0708 (1) Exonuclease III [DNA replication, recombination, and repair]
COG1180 (6) Pyruvate-formate lyase-activating enzyme [Posttranslational modification, protein turnover, chaperones]
COG1280 (3) Putative threonine efflux protein [Amino acid transport and metabolism]
COG1366 (2) Anti-anti-sigma regulatory factor (antagonist of anti-sigma factor) [Signal transduction mechanisms]
COG1882 (12) Pyruvate-formate lyase [Energy production and conversion]
COG1922 (1) Teichoic acid biosynthesis proteins [Cell envelope biogenesis, outer membrane]
COG2116 (2) Formate/nitrite family of transporters [Inorganic ion transport and metabolism]
COG2223 (1) Nitrate/nitrite transporter [Inorganic ion transport and metabolism]
COG2226 (1) Methylase involved in ubiquinone/menaquinone biosynthesis [Coenzyme metabolism]
COG2759 (5) Formyltetrahydrofolate synthetase (FTHFS) catalyzes the ATP-dependent activation of formate ion via its addition to the N10 position of tetrahydrofolate.
COG3363 (8) Archaeal IMP cyclohydrolase [Nucleotide transport and metabolism]
COG3603 (1) Uncharacterized conserved protein [Function unknown]
COG3741 (1) N-formylglutamate amidohydrolase [Amino acid transport and metabolism]
COG4221 (2) Short-chain alcohol dehydrogenase of unknown specificity [General function prediction only]
COG4774 (1) Outer membrane receptor for monomeric catechols [Inorganic ion transport and metabolism]
pfam00551 (1) Formyl transferase.
pfam00903 (1) Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily.
pfam01502 (1) Phosphoribosyl-AMP cyclohydrolase.
pfam01808 (18) AICARFT/IMPCHase bienzyme.
pfam02142 (10) MGS-like domain.
pfam03477 (31) ATP cone domain.
smart00065 (2) Domain present in phytochromes and cGMP-specific phosphodiesterases.

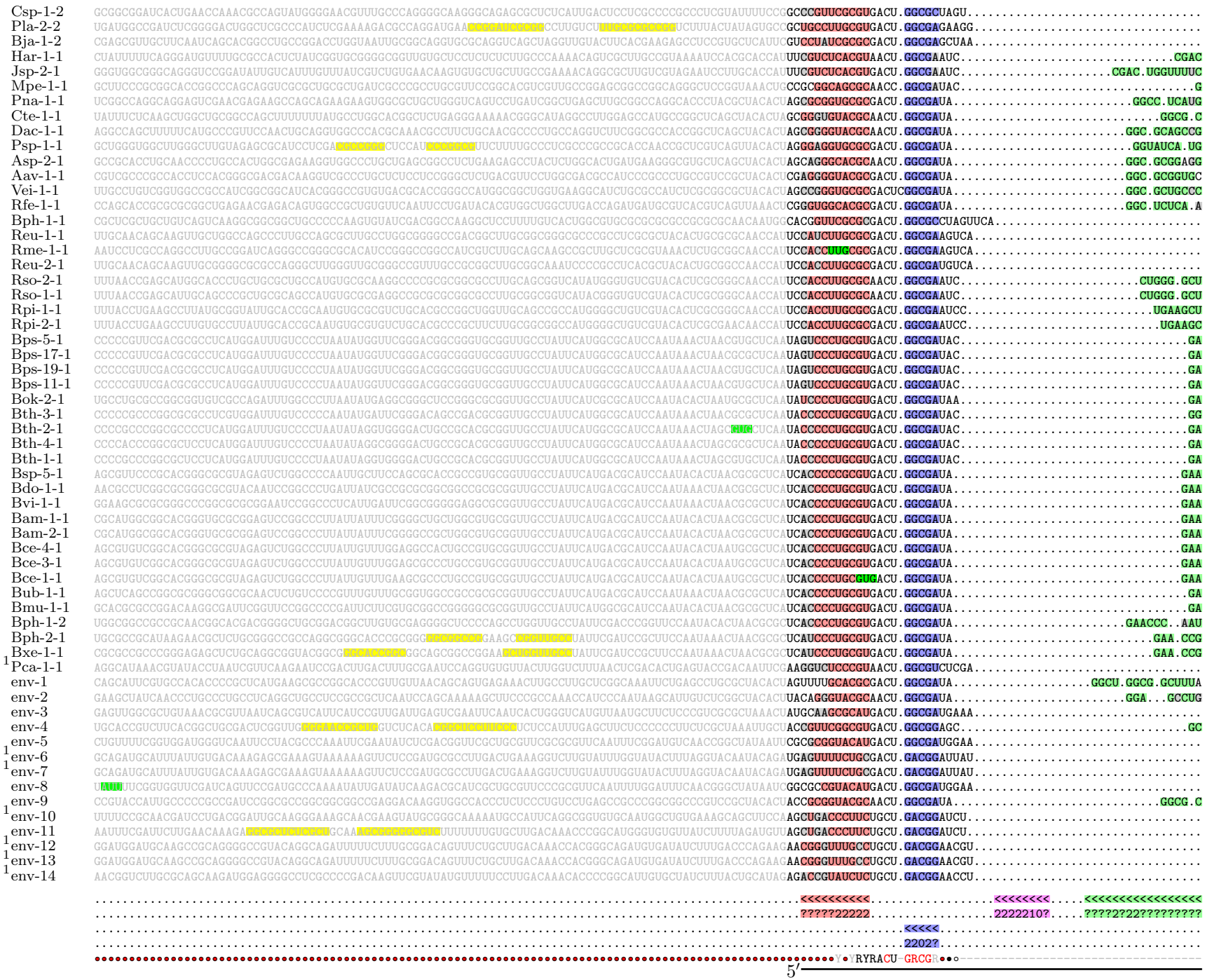
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Bce-2-1, Bma-1-1, Bma-10-1, Bma-2-1, Bma-3-1, Bma-4-1, Bma-5-1, Bma-6-1, Bma-7-1, Bma-8-1, Bma-9-1, Bok-1-1, Bps-1-1, Bps-10-1, Bps-12-1, Bps-13-1, Bps-14-1, Bps-15-1, Bps-16-1, Bps-18-1, Bps-2-1, Bps-20-1, Bps-3-1, Bps-4-1, Bps-6-1, Bps-7-1, Bps-8-1, Bps-9-1, Bth-2-2, Cbo-2-1, Cbo-4-1, Cbo-7-1, Cbo-9-1, env-45

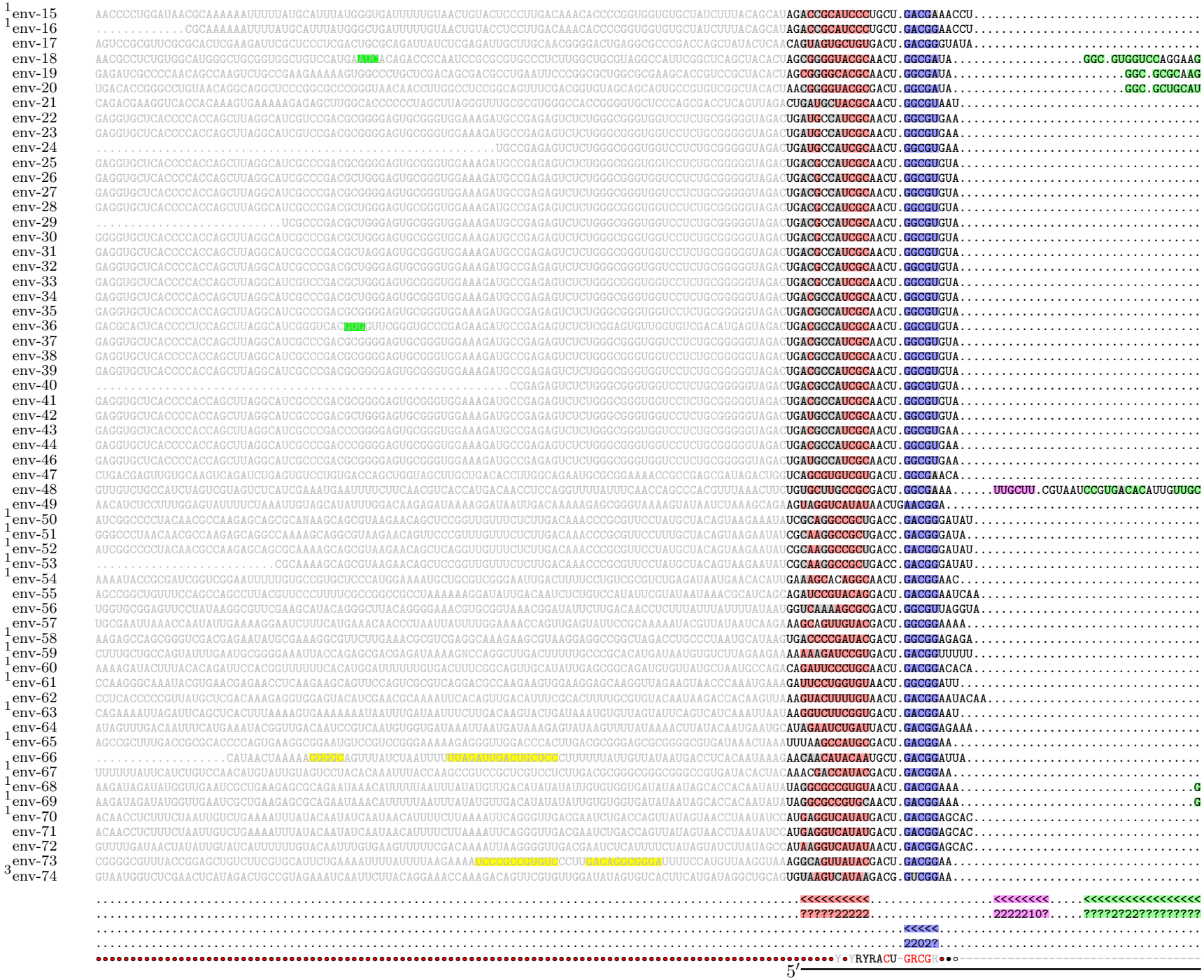
63.5 Multiple-sequence alignment

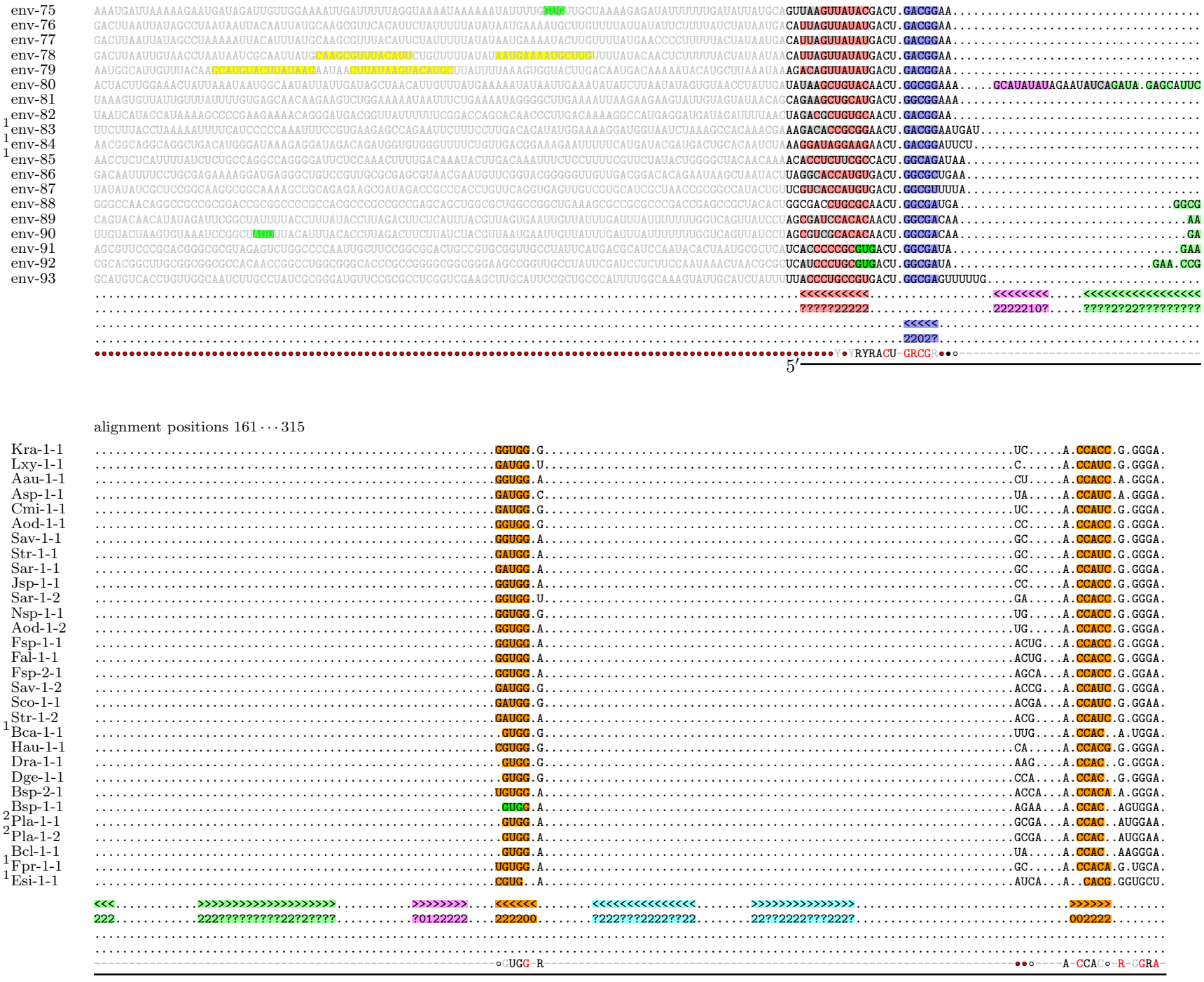
Each *pfl* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 63.1. Superscript numbers refer to annotations listed in Section 63.2. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *pfl* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

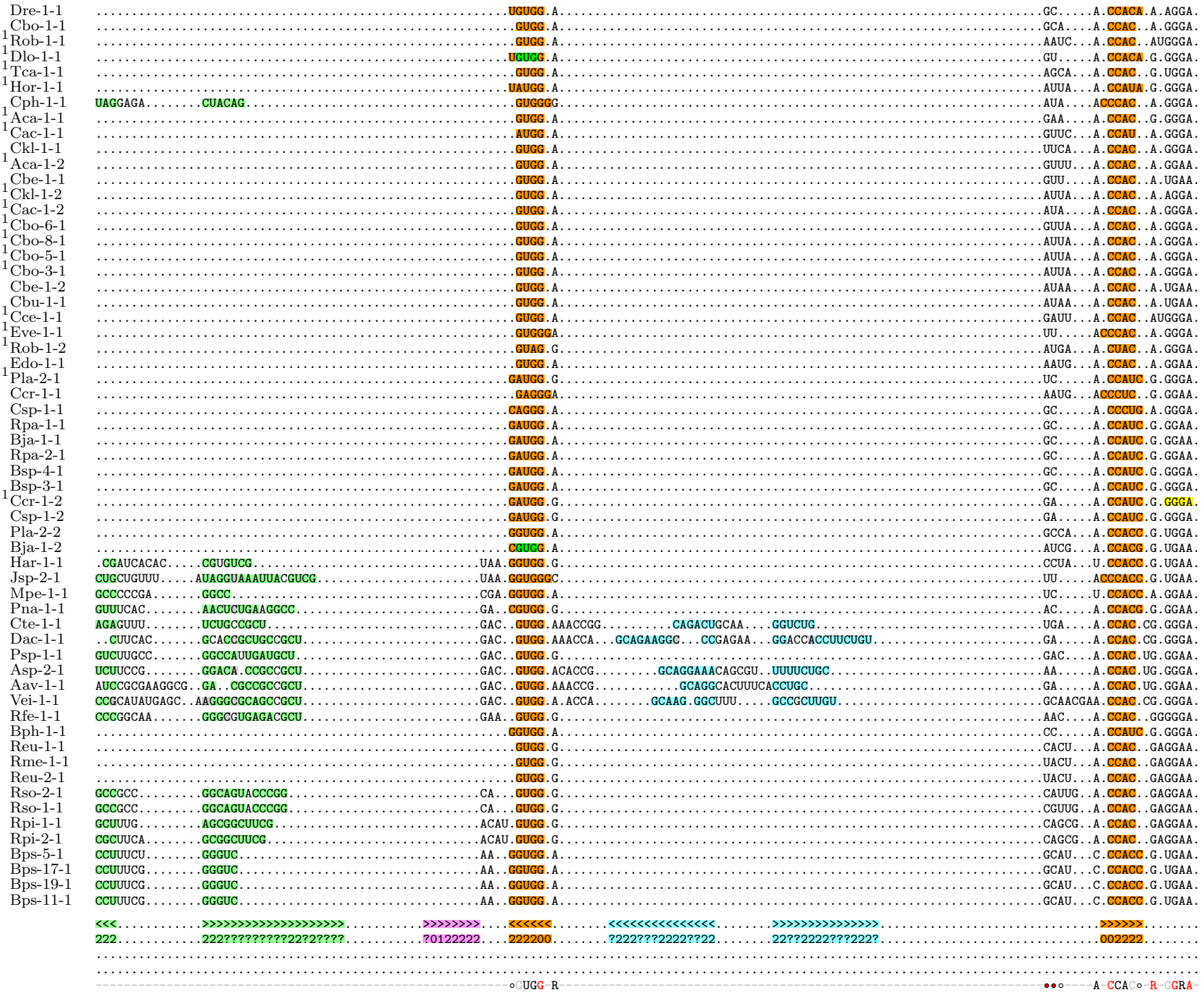


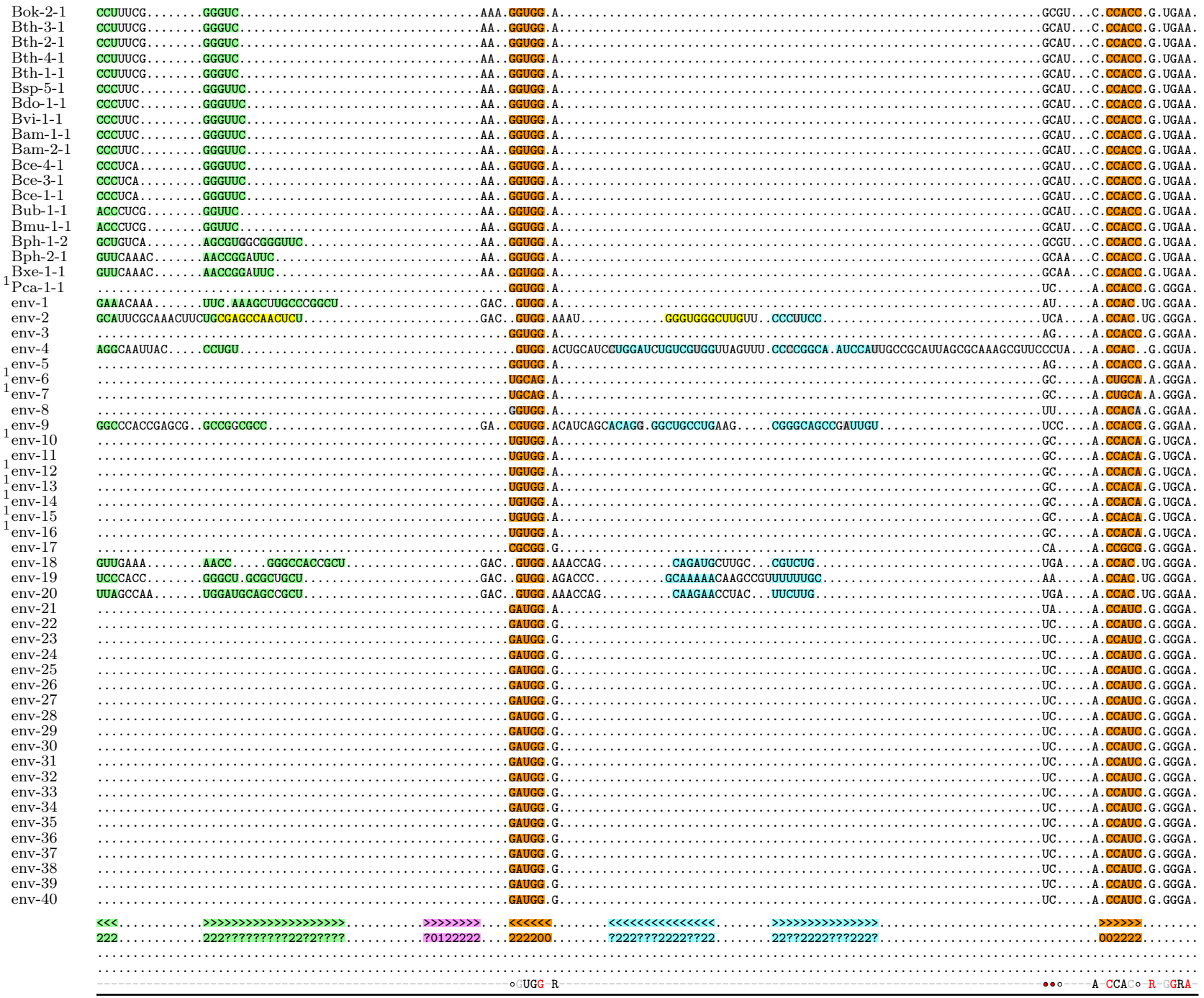

Cmi-1-1
Aod-1-1
Sav-1-1
Str-1-1
Sar-1-1
Jsp-1-1
Sar-1-2
Nsp-1-1
Aod-1-2
Fsp-1-1
Fal-1-1
Fsp-2-1
Sav-1-2
Sco-1-1
Str-1-2
1Bca-1-1
Hau-1-1
Dra-1-1
Dge-1-1
Bsp-2-1
Bsp-1-1
2Pla-1-1
2Pla-1-2
Bcl-1-1
1Fpr-1-1
1Esi-1-1
Dre-1-1
Cbo-1-1
1Rob-1-1
1Dlo-1-1
1Tca-1-1
1Hor-1-1
Cph-1-1
1Aca-1-1
1Cac-1-1
Ckl-1-1
1Aca-1-2
Cbe-1-1
1Ckl-1-2
1Cac-1-2
1Cbo-6-1
1Cbo-8-1
1Cbo-5-1
1Cbo-3-1
Cbe-1-2
Cbu-1-1
1Cce-1-1
1Eve-1-1
1Rob-1-2
Edo-1-1
1Pla-2-1
Ccr-1-1
Csp-1-1
Rpa-1-1
Bja-1-1
Rpa-2-1
Bsp-4-1
Bsp-3-1
1Ccr-1-2
5′

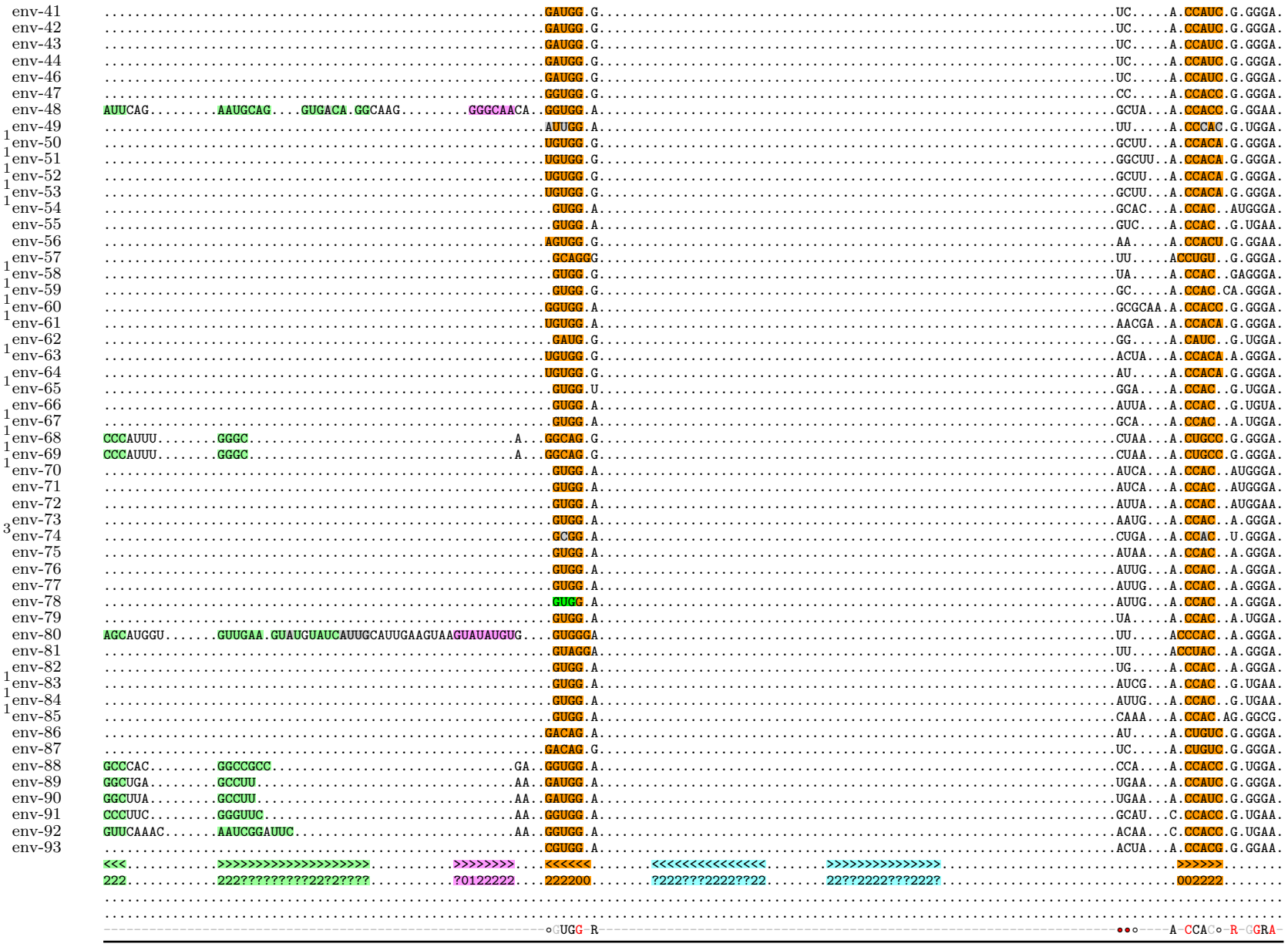

Csp-1-2
Pla-2-2
Bja-1-2
Har-1-1
Jsp-2-1
Mpe-1-1
Pna-1-1
Cte-1-1
Dac-1-1
Psp-1-1
Asp-2-1
Aav-1-1
Vei-1-1
Rfe-1-1
Bph-1-1
Reu-1-1
Rme-1-1
Reu-2-1
Rso-2-1
Rso-1-1
Rpi-1-1
Rpi-2-1
Bps-5-1
Bps-17-1
Bps-19-1
Bps-11-1
Bok-2-1
Bth-3-1
Bth-2-1
Bth-4-1
Bth-1-1
Bsp-5-1
Bdo-1-1
Bvi-1-1
Bam-1-1
Bam-2-1
Bce-4-1
Bce-3-1
Bce-1-1
Bub-1-1
Bmu-1-1
Bph-1-2
Bph-2-1
Bxe-1-1
1Pca-1-1
env-1
env-2
env-3
env-4
env-5
1env-6
1env-7
env-8
env-9
1env-10
env-11
1env-12
1env-13
1env-14
5′

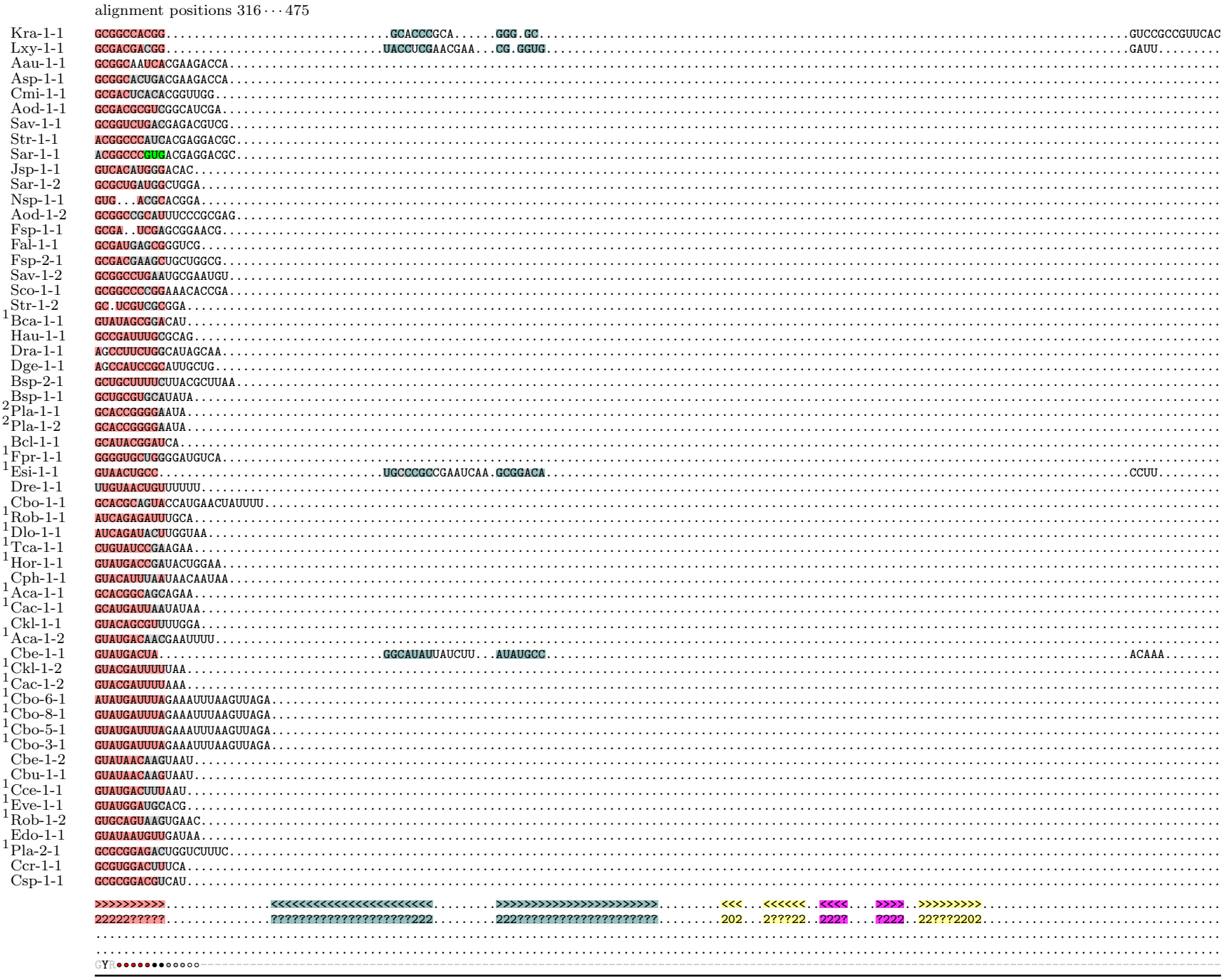

env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
5′

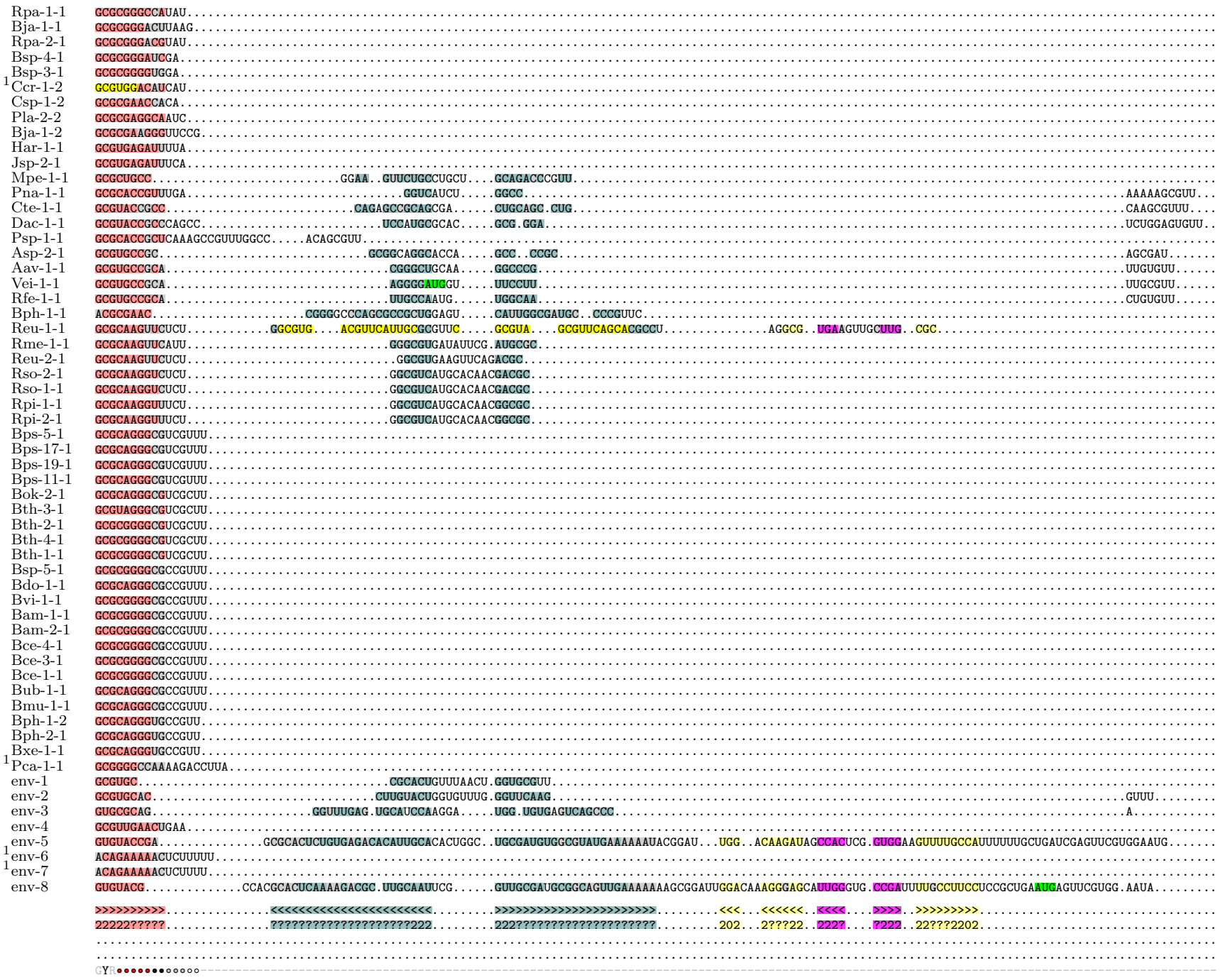

env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
5′
alignment positions 161 · · · 315
Kra-1-1
Lxy-1-1
Aau-1-1
Asp-1-1
Cmi-1-1
Aod-1-1
Sav-1-1
Str-1-1
Sar-1-1
Jsp-1-1
Sar-1-2
Nsp-1-1
Aod-1-2
Fsp-1-1
Fal-1-1
Fsp-2-1
Sav-1-2
Sco-1-1
Str-1-2
Bca-1-1
Hau-1-1
Dra-1-1
Dge-1-1
Bsp-2-1
Bsp-1-1
Pla-1-1
Pla-1-2
Bcl-1-1
Fpr-1-1
Esi-1-1

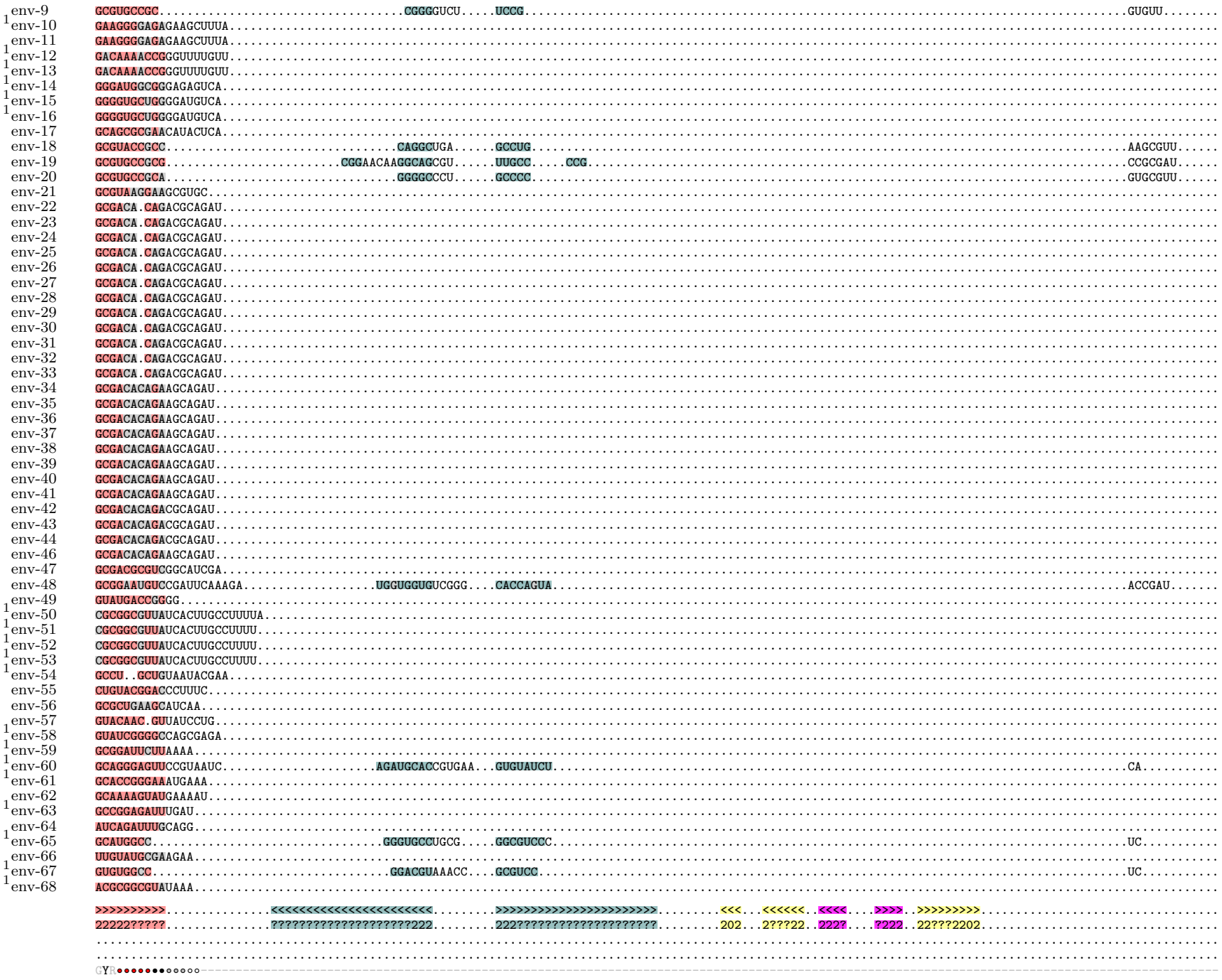

Dre-1-1
Cbo-1-1
1Rob-1-1
1Dlo-1-1
1Tca-1-1
1Hor-1-1
Cph-1-1
1Aca-1-1
1Cac-1-1
Ckl-1-1
1Aca-1-2
Cbe-1-1
1Ckl-1-2
1Cac-1-2
1Cbo-6-1
1Cbo-8-1
1Cbo-5-1
1Cbo-3-1
Cbe-1-2
Cbu-1-1
1Cce-1-1
1Eve-1-1
1Rob-1-2
Edo-1-1
1Pla-2-1
Ccr-1-1
Csp-1-1
Rpa-1-1
Bja-1-1
Rpa-2-1
Bsp-4-1
Bsp-3-1
1Ccr-1-2
Csp-1-2
Pla-2-2
Bja-1-2
Har-1-1
Jsp-2-1
Mpe-1-1
Pna-1-1
Cte-1-1
Dac-1-1
Psp-1-1
Asp-2-1
Aav-1-1
Vei-1-1
Rfe-1-1
Bph-1-1
Reu-1-1
Rme-1-1
Reu-2-1
Rso-2-1
Rso-1-1
Rpi-1-1
Rpi-2-1
Bps-5-1
Bps-17-1
Bps-19-1
Bps-11-1

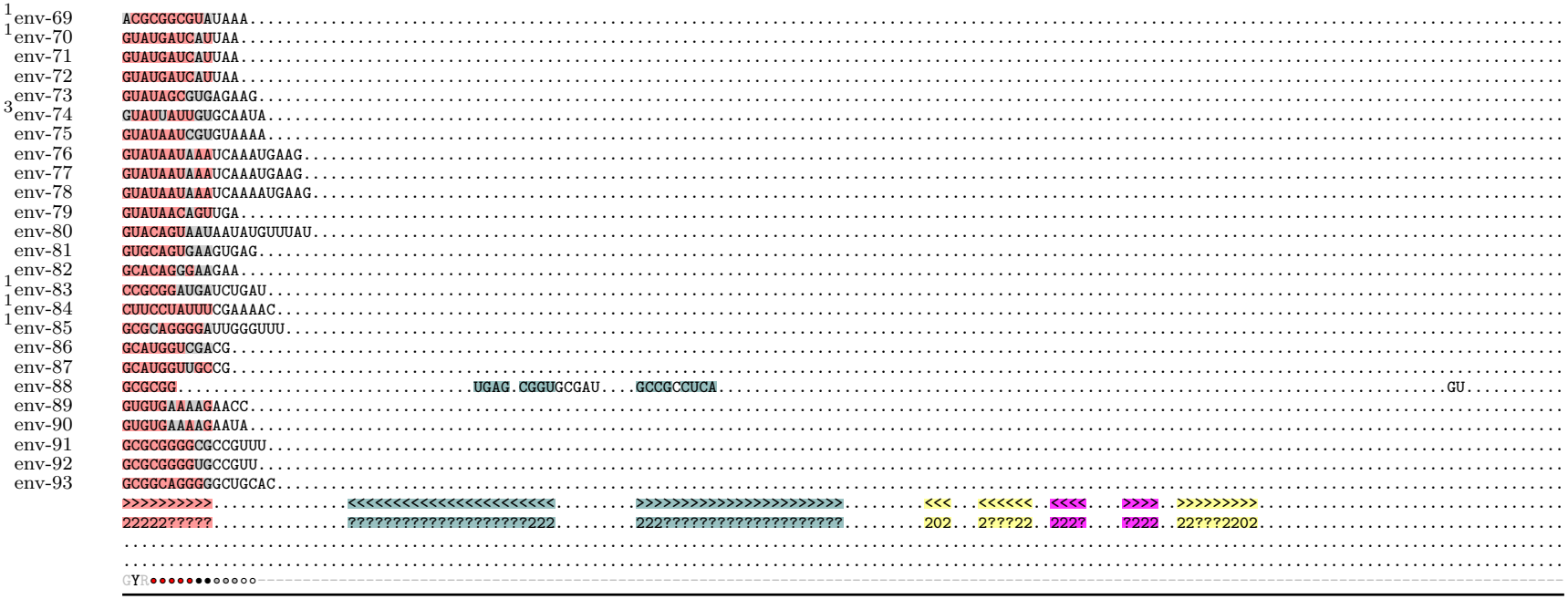

Bok-2-1
Bth-3-1
Bth-2-1
Bth-4-1
Bth-1-1
Bsp-5-1
Bdo-1-1
Bvi-1-1
Bam-1-1
Bam-2-1
Bce-4-1
Bce-3-1
Bce-1-1
Bub-1-1
Bmu-1-1
Bph-1-2
Bph-2-1
Bxe-1-1
Pca-1-1
env-1
env-2
env-3
env-4
env-5
env-6
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40


env-41
env-42
env-43
env-44
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93

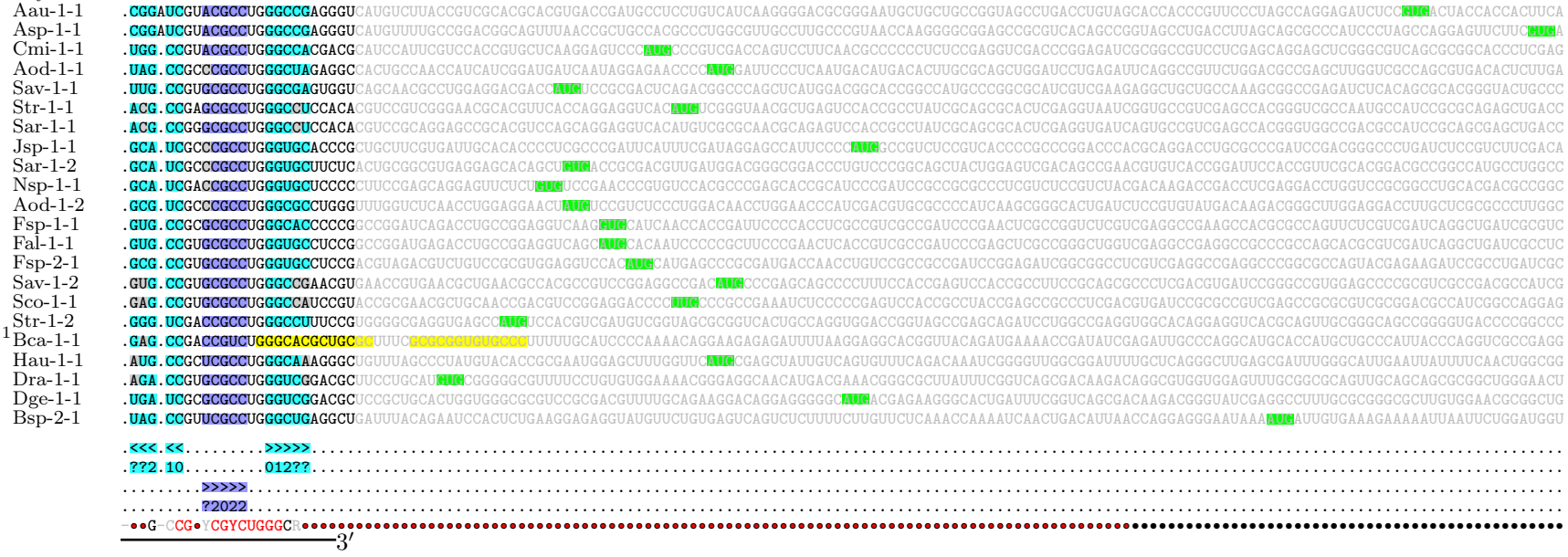

alignment positions 316 · · · 475
Kra-1-1
Lxy-1-1
Aau-1-1
Asp-1-1
Cmi-1-1
Aod-1-1
Sav-1-1
Str-1-1
Sar-1-1
Jsp-1-1
Sar-1-2
Nsp-1-1
Aod-1-2
Fsp-1-1
Fal-1-1
Fsp-2-1
Sav-1-2
Sco-1-1
Str-1-2
1 Bca-1-1
Hau-1-1
Dra-1-1
Dge-1-1
Bsp-2-1
Bsp-1-1
2 Pla-1-1
2 Pla-1-2
Bcl-1-1
1 Fpr-1-1
1 Esi-1-1
Dre-1-1
Cbo-1-1
1 Rob-1-1
1 Dlo-1-1
1 Tca-1-1
1 Hor-1-1
Cph-1-1
1 Aca-1-1
1 Cac-1-1
Ckl-1-1
1 Aca-1-2
Cbe-1-1
1 Ckl-1-2
1 Cac-1-2
1 Cbo-6-1
1 Cbo-8-1
1 Cbo-5-1
1 Cbo-3-1
Cbe-1-2
Cbu-1-1
1 Cce-1-1
1 Eve-1-1
1 Rob-1-2
Edo-1-1
1 Pla-2-1
Ccr-1-1
Csp-1-1

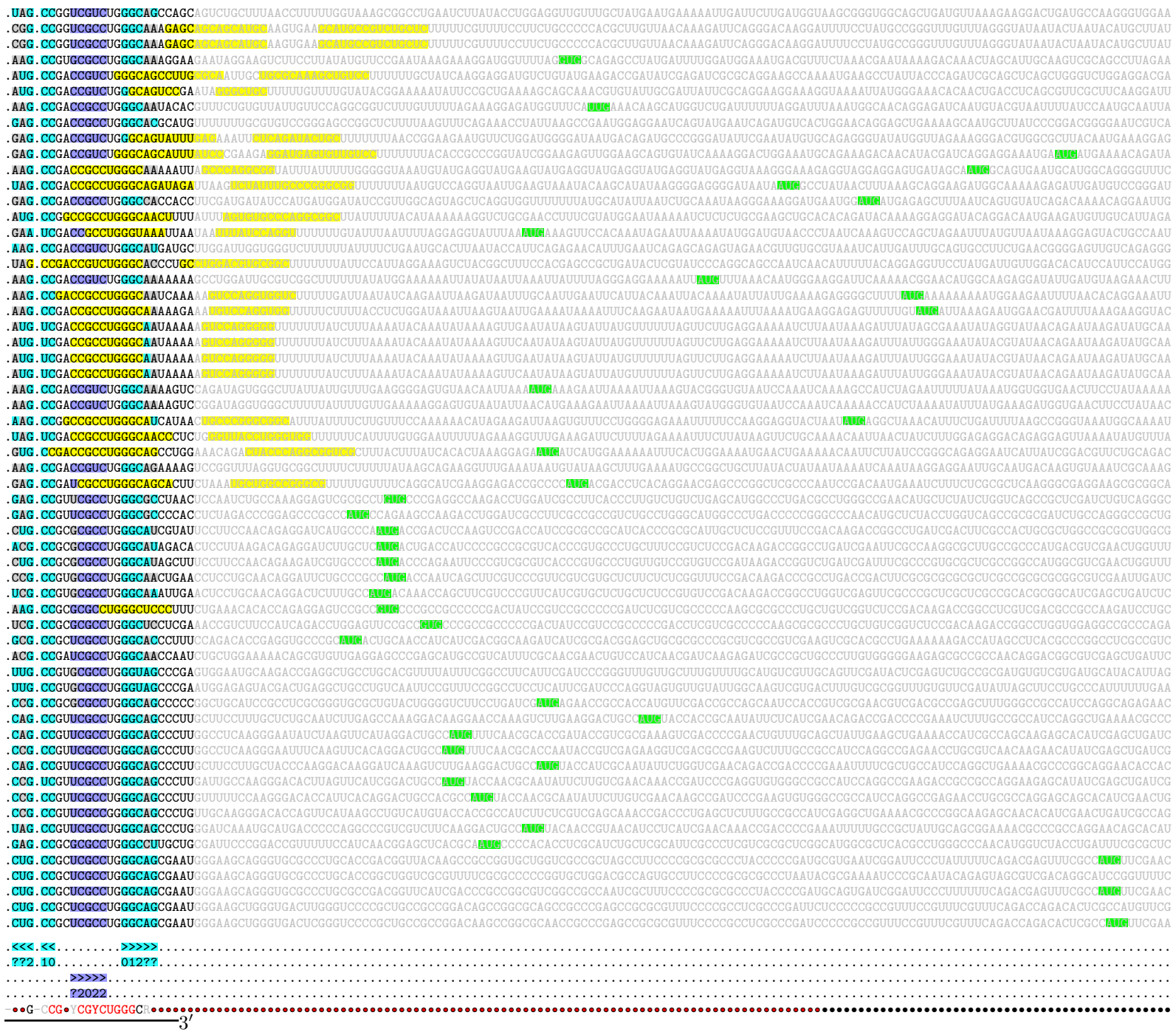

Rpa-1-1
Bja-1-1
Rpa-2-1
Bsp-4-1
Bsp-3-1
1 Ccr-1-2
Csp-1-2
Pla-2-2
Bja-1-2
Har-1-1
Jsp-2-1
Mpe-1-1
Pna-1-1
Cte-1-1
Dac-1-1
Psp-1-1
Asp-2-1
Aav-1-1
Vei-1-1
Rfe-1-1
Bph-1-1
Reu-1-1
Rme-1-1
Reu-2-1
Rso-2-1
Rso-1-1
Rpi-1-1
Rpi-2-1
Bps-5-1
Bps-17-1
Bps-19-1
Bps-11-1
Bok-2-1
Bth-3-1
Bth-2-1
Bth-4-1
Bth-1-1
Bsp-5-1
Bdo-1-1
Bvi-1-1
Bam-1-1
Bam-2-1
Bce-4-1
Bce-3-1
Bce-1-1
Bub-1-1
Bmu-1-1
Bph-1-2
Bph-2-1
Bxe-1-1
1 Pca-1-1
env-1
env-2
env-3
env-4
env-5
1 env-6
1 env-7
env-8

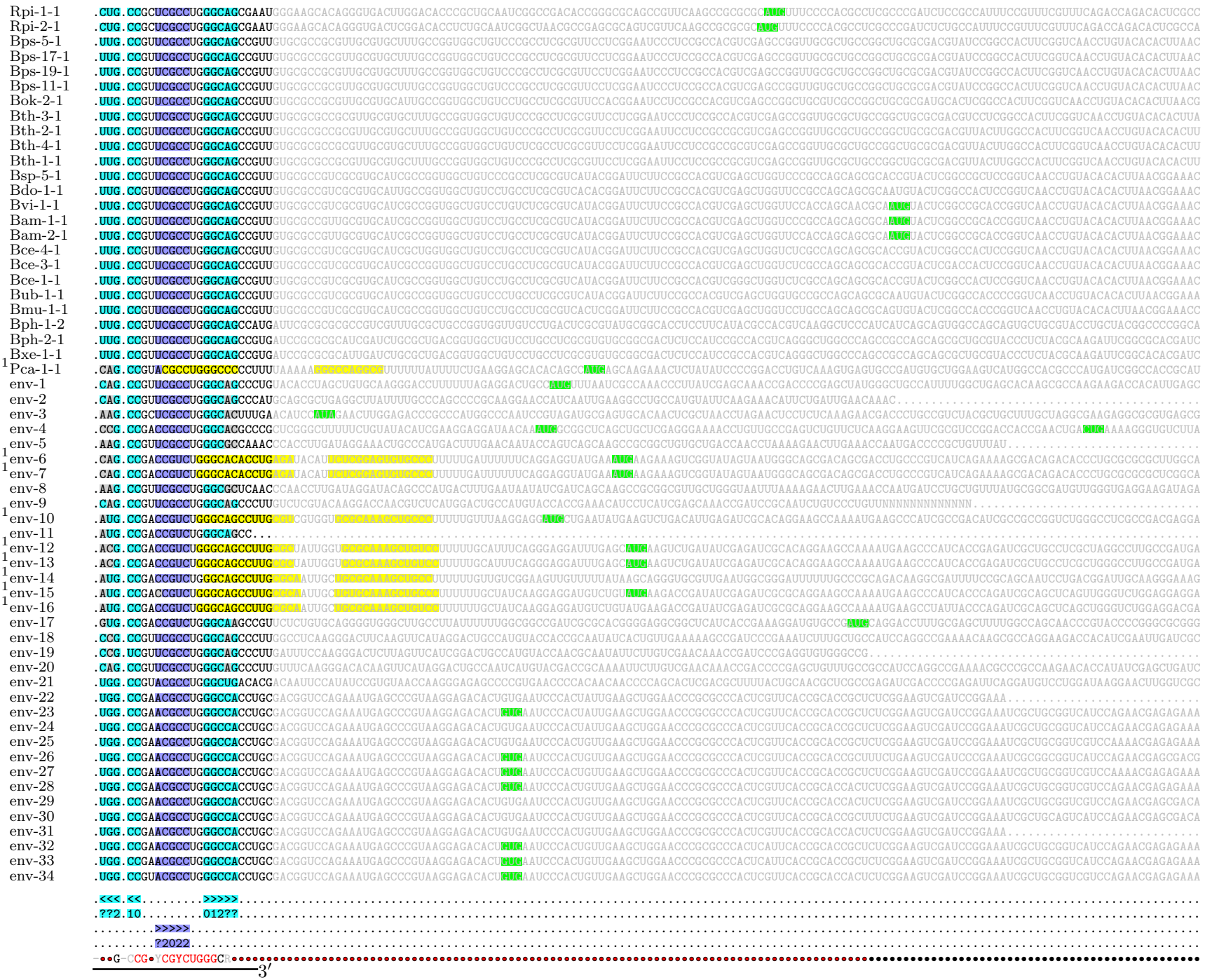

env-9
env-10
env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68

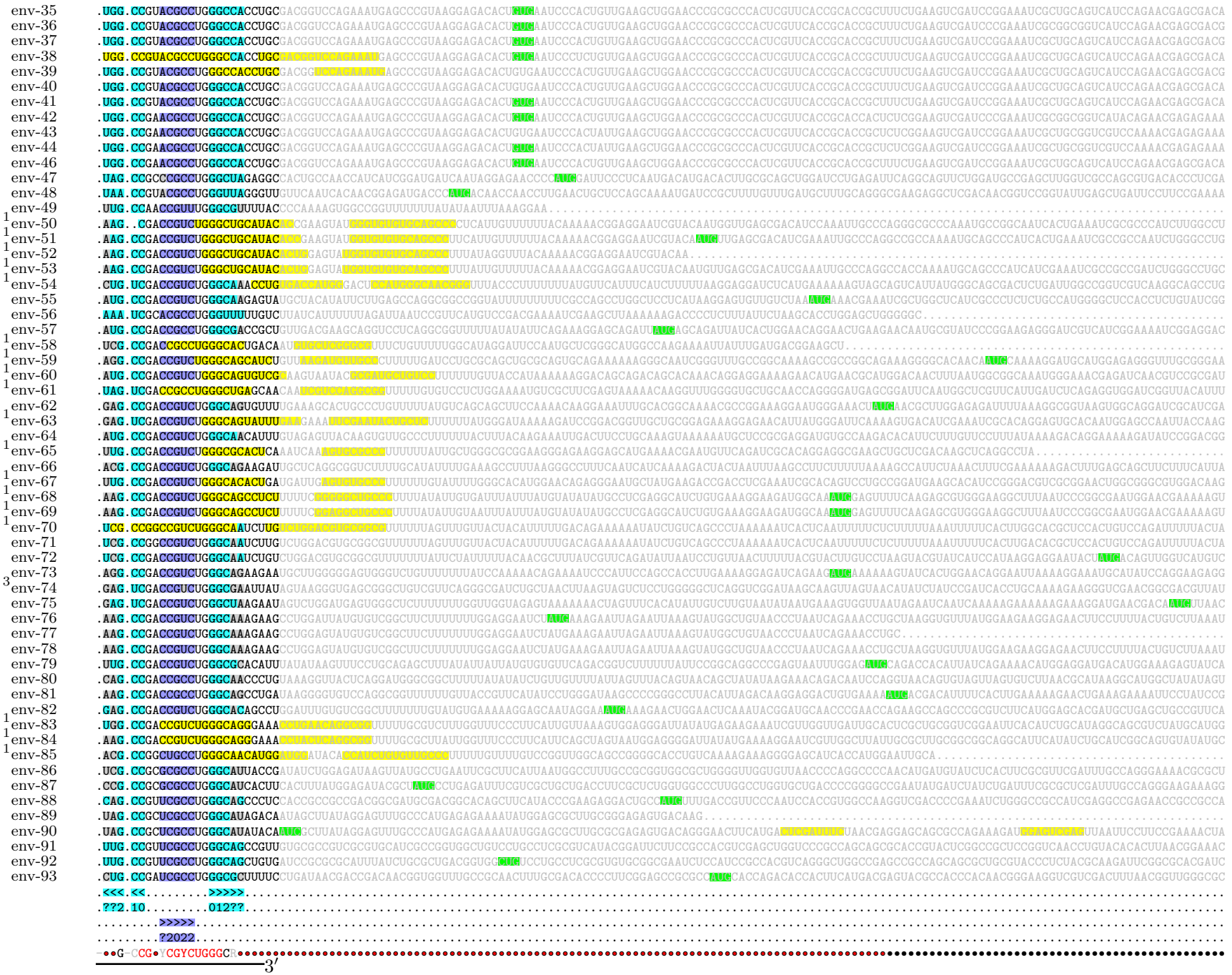


alignment positions 476 · · · 635

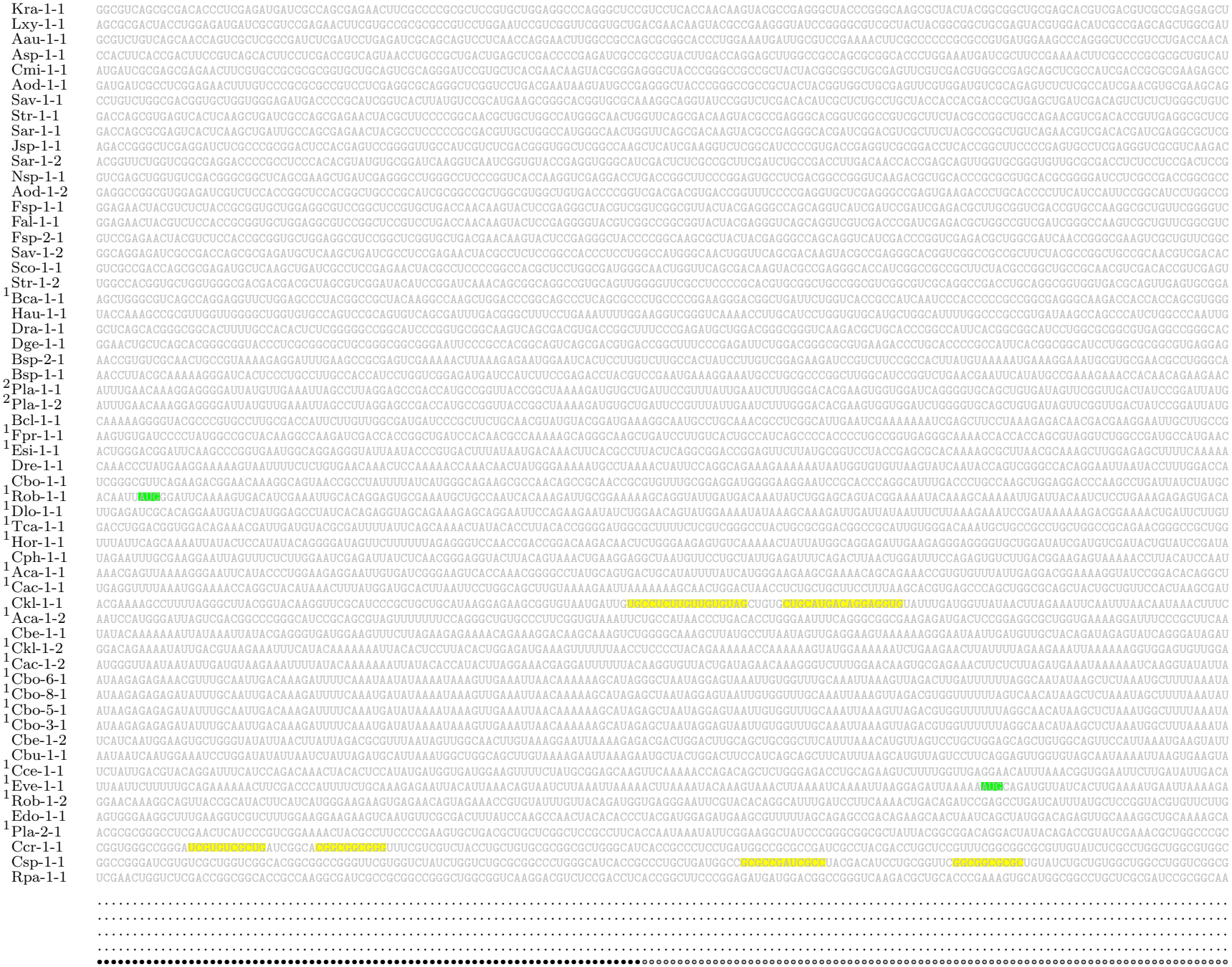


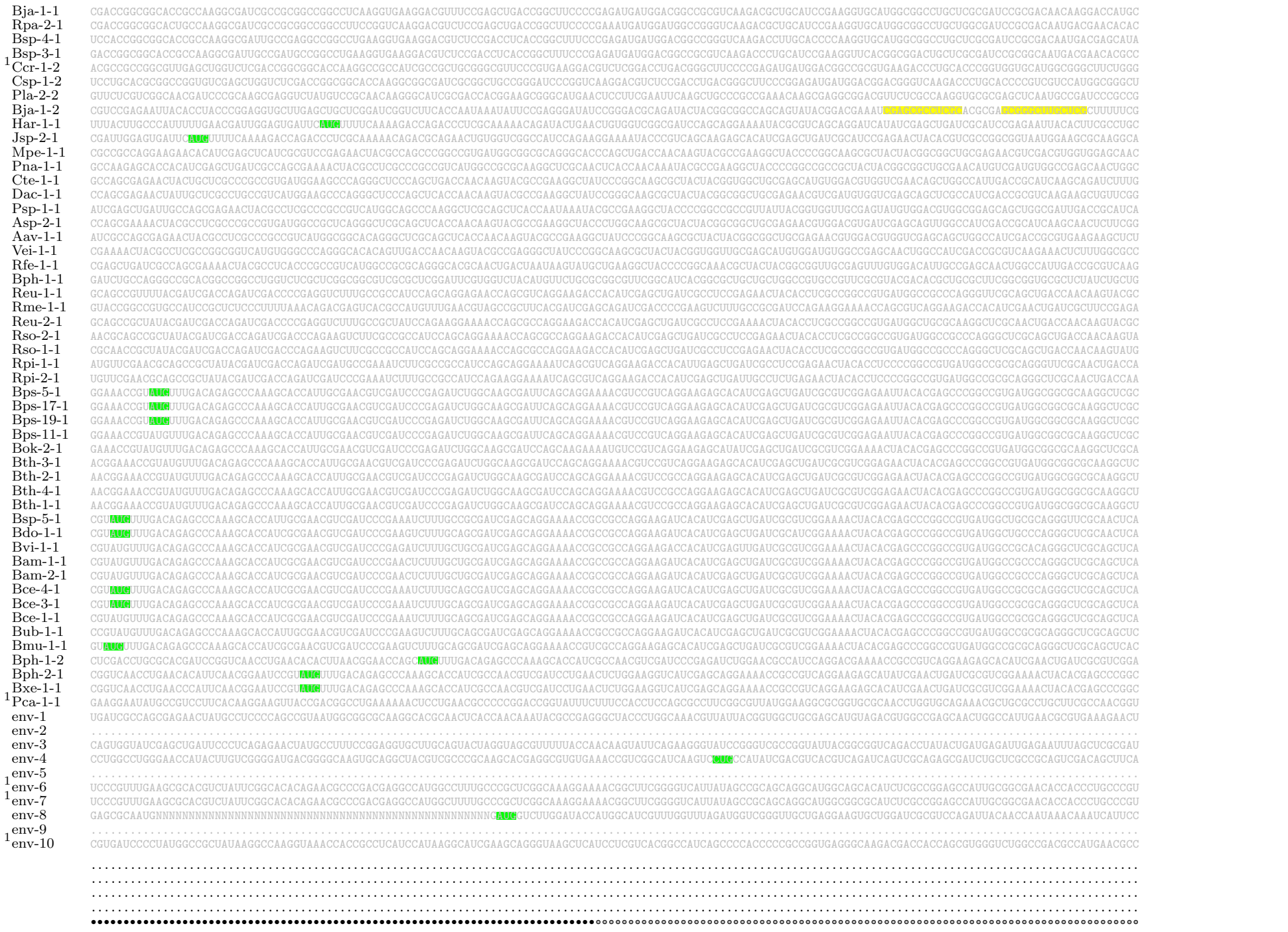

Bsp-1-1
2Pla-1-1
2Pla-1-2
Bcl-1-1
1Fpr-1-1
1Esi-1-1
Dre-1-1
Cbo-1-1
1Rob-1-1
1Dlo-1-1
1Tca-1-1
1Hor-1-1
Cph-1-1
1Aca-1-1
1Cac-1-1
Ckl-1-1
1Aca-1-2
Cbe-1-1
1Ckl-1-2
1Cac-1-2
1Cbo-6-1
1Cbo-8-1
1Cbo-5-1
1Cbo-3-1
Cbe-1-2
Cbu-1-1
1Cce-1-1
1Eve-1-1
1Rob-1-2
Edo-1-1
1Pla-2-1
Ccr-1-1
Csp-1-1
Rpa-1-1
Bja-1-1
Rpa-2-1
Bsp-4-1
Bsp-3-1
1Ccr-1-2
Csp-1-2
Pla-2-2
Bja-1-2
Har-1-1
Jsp-2-1
Mpe-1-1
Pna-1-1
Cte-1-1
Dac-1-1
Psp-1-1
Asp-2-1
Aav-1-1
Vei-1-1
Rfe-1-1
Bph-1-1
Reu-1-1
Rme-1-1
Reu-2-1
Rso-2-1
Rso-1-1
3′


Rpi-1-1
Rpi-2-1
Bps-5-1
Bps-17-1
Bps-19-1
Bps-11-1
Bok-2-1
Bth-3-1
Bth-2-1
Bth-4-1
Bth-1-1
Bsp-5-1
Bdo-1-1
Bvi-1-1
Bam-1-1
Bam-2-1
Bce-4-1
Bce-3-1
Bce-1-1
Bub-1-1
Bmu-1-1
Bph-1-2
Bph-2-1
Bxe-1-1
1 Pca-1-1
env-1
env-2
env-3
env-4
1 env-5
1 env-6
1 env-7
env-8
1 env-9
1 env-10
env-11
1 env-12
1 env-13
1 env-14
1 env-15
1 env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
3′

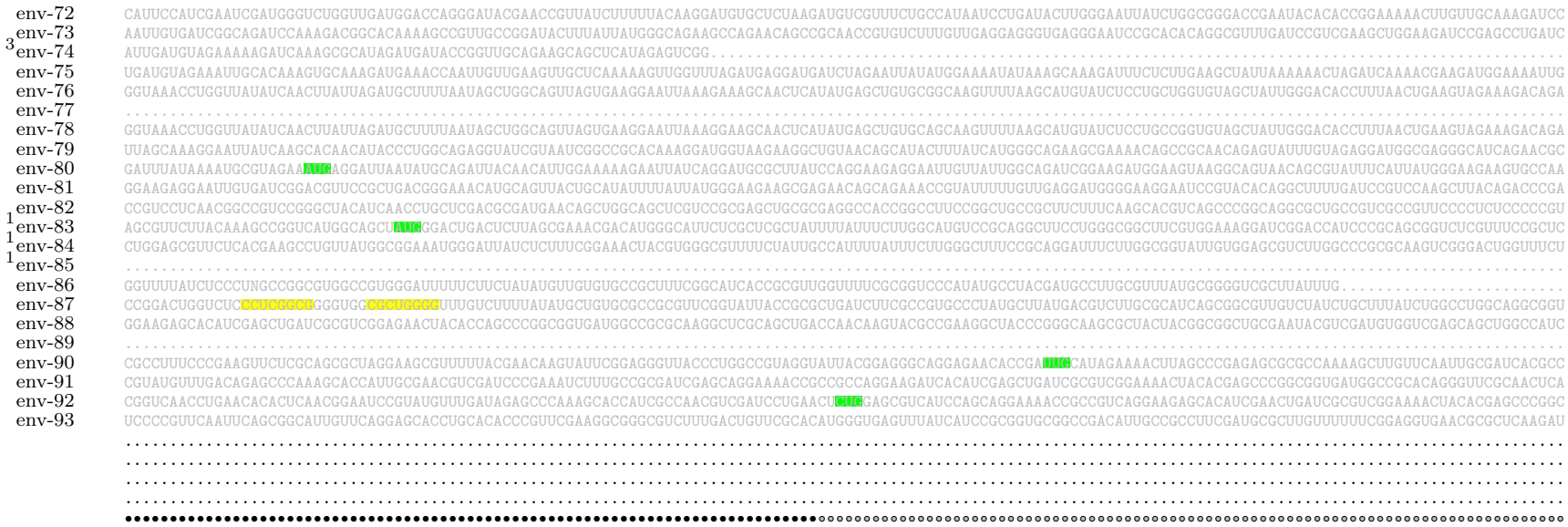

env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
3′

alignment positions 636 · · · 795




Bja-1-1
Rpa-2-1
Bsp-4-1
Bsp-3-1
Ccr-1-2
Csp-1-2
Pla-2-2
Bja-1-2
Har-1-1
Jsp-2-1
Mpe-1-1
Pna-1-1
Cte-1-1
Dac-1-1
Psp-1-1
Asp-2-1
Aav-1-1
Vei-1-1
Rfe-1-1
Bph-1-1
Reu-1-1
Rme-1-1
Reu-2-1
Rso-2-1
Rso-1-1
Rpi-1-1
Rpi-2-1
Bps-5-1
Bps-17-1
Bps-19-1
Bps-11-1
Bok-2-1
Bth-3-1
Bth-2-1
Bth-4-1
Bth-1-1
Bsp-5-1
Bdo-1-1
Bvi-1-1
Bam-1-1
Bam-2-1
Bce-4-1
Bce-3-1
Bce-1-1
Bub-1-1
Bmu-1-1
Bph-1-2
Bph-2-1
Bxe-1-1
Pca-1-1
env-1
env-2
env-3
env-4
env-5
env-6
env-7
env-8
env-9
env-10

env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71



alignment positions 796 · · · 801

```
[1]Esi-1-1     AGAGCG
  Dre-1-1      GGUUUU
  Cbo-1-1      CCCUGU
[1]Rob-1-1     AGAAAG
[1]Dlo-1-1     GACGGC
[1]Tca-1-1     CGGCGU
[1]Hor-1-1     UUAACG
  Cph-1-1      UGUACA
[1]Aca-1-1     UUUGAU
[1]Cac-1-1     ACUCUA
  Ckl-1-1      UAUGUC
[1]Aca-1-2     GCCCUA
  Cbe-1-1      AAUUAU
[1]Ckl-1-2     UGUAGA
[1]Cac-1-2     AAAGCA
[1]Cbo-6-1     UAUUAA
[1]Cbo-8-1     AUUAAA
[1]Cbo-5-1     UAUUAA
[1]Cbo-3-1     UAUUAA
  Cbe-1-2      AAAAAA
  Cbu-1-1      UUAAAA
[1]Cce-1-1     CAGAAA
[1]Eve-1-1     AAAUUC
[1]Rob-1-2     GAAAUA
  Edo-1-1      AGAAAG
[1]Pla-2-1     GACAGG
  Ccr-1-1      AGGCGA
  Csp-1-1      AUCCGG
  Rpa-1-1      UGCCGA
  Bja-1-1      GGAUGC
  Rpa-2-1      CCAGGC
  Bsp-4-1      CCGAGG
  Bsp-3-1      GAGGCG
[1]Ccr-1-2     CGUGCG
  Csp-1-2      GCUGGG
  Pla-2-2      UUCACG
  Bja-1-2      UGCCGA
  Har-1-1      CGUGAU
  Jsp-2-1      CGCAAC
  Mpe-1-1      UGGCCA
  Pna-1-1      CAUCGA
  Cte-1-1      GCGCCG
  Dac-1-1      CGCCGA
  Psp-1-1      AGCAGA
  Asp-2-1      CGCCGA
  Aav-1-1      UCGGCG
  Vei-1-1      GAGGCC
  Rfe-1-1      CAAAUC
  Bph-1-1      UAUCUC
  Reu-1-1      CGAGGG
  Rme-1-1      ACUACA
  Reu-2-1      CGAGGG
  Rso-2-1      CGCCGA
  Rso-1-1      CCGAGG
  Rpi-1-1      ACAAGU
  Rpi-2-1      CAAGUA
  Bps-5-1      AACUGA
  Bps-17-1     AACUGA
  Bps-19-1     AACUGA
  Bps-11-1     AACUGA

               ......
               ......
               ......
               ......
               °°°°°°
```

```
Bok-2-1     GCUCAC
Bth-3-1     GCAGCU
Bth-2-1     CGCAGC
Bth-4-1     CGCAGC
Bth-1-1     CGCAGC
Bsp-5-1     CGAACA
Bdo-1-1     CGAACA
Bvi-1-1     CGAACA
Bam-1-1     CGAACA
Bam-2-1     CGAACA
Bce-4-1     CGAACA
Bce-3-1     CGAACA
Bce-1-1     CGAACA
Bub-1-1     ACGAAC
Bmu-1-1     GAACAA
Bph-1-2     AAACUA
Bph-2-1     UGUUAU
Bxe-1-1     UGUCAU
1Pca-1-1    GCCUUC
env-1       GUUUGG
env-2       ......
env-3       CGUGCA
env-4       ACGAAA
env-5       ......
1env-6      UAUCGG
1env-7      UAUCGG
env-8       UGAUGA
env-9       ......
1env-10     AUAGGC
env-11      ......
1env-12     GCUCUG
1env-13     GCUCUG
1env-14     GAGAAC
1env-15     GCACUG
1env-16     GCACUG
env-17      CCAUGG
env-18      CGGUGG
env-19      ......
env-20      GCGCCG
env-21      CAACUU
env-22      ......
env-23      GCCAUU
env-24      GCCAUU
env-25      ......
env-26      GCCAUU
env-27      GCCAUU
env-28      GCCAUU
env-29      NNNNNN
env-30      ......
env-31      ......
env-32      CCAUUG
env-33      GCCAUU
env-34      GCCAUU
env-35      CCAUUG
env-36      GCCAUU
env-37      GCCAUU
env-38      GCCAUU
env-39      ......
env-40      GCCAUU
env-41      GCCAUU

            ......
            ......
            ......
            ......
            °°°°°°
```

```
      env-42    GCCAUU
      env-43    ......
      env-44    GCCAUU
      env-46    GCCAUU
      env-47    GUCUUC
      env-48    GCCAAC
      env-49    ......
    1 env-50    ......
    1 env-51    CCAUGA
    1 env-52    ......
    1 env-53    CAUGAA
    1 env-54    GCCGGC
      env-55    GGCCUC
      env-56    ......
      env-57    GUAAUA
    1 env-58    ......
    1 env-59    UUACGA
    1 env-60    ......
    1 env-61    CGUGUA
      env-62    GUCAUC
    1 env-63    ......
      env-64    GUAAUG
    1 env-65    ......
      env-66    GCUAUU
    1 env-67    ......
    1 env-68    ACCGUU
    1 env-69    ACCGUU
    1 env-70    UUUUCC
      env-71    UUUUCC
      env-72    GACGAU
      env-73    AUCUAC
    3 env-74    ......
      env-75    AUCUUA
      env-76    UGUAUU
      env-77    ......
      env-78    UGUAUU
      env-79    AGGCAU
      env-80    CAGCCG
      env-81    GCCUGA
      env-82    CCUCGA
    1 env-83    UUUUCU
    1 env-84    GCCUGU
    1 env-85    ......
      env-86    ......
      env-87    GAAGCC
      env-88    GACCGC
      env-89    ......
      env-90    AAUGUC
      env-91    CGAACA
      env-92    CGUGAU
      env-93    UGGCCC
                ......
                ......
                ......
                ......
                °°°°°°
```

64 *pheA*

64.1 Taxa

The taxonomy of each organism containing a putative *pheA* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *pheA* RNA in Sections 64.2 and 64.4.

abbrev. of hits	taxonomy of species
Mgi-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium gilvum* PYR-GCK
Mle-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium leprae* TN
Msm-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium smegmatis str.* MC2 155
Msp-1-1 to Msp-1-2	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium sp.* JLS
Msp-2-1 to Msp-2-2	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium sp.* KMS
Msp-3-1 to Msp-3-2	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium sp.* MCS
Mul-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium ulcerans* Agy99
Mva-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium vanbaalenii* PYR-1
Mav-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumavium complex (MAC) *Mycobacterium avium* 104
Mav-2-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumavium complex (MAC) *Mycobacterium avium subs*
Mbo-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium bovis* AF2122
Mbo-2-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium bovis* BCG st
Mtu-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium tuberculosis* C
Mtu-2-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium tuberculosis* C
Mtu-3-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium tuberculosis* H
Mtu-4-1 to Mtu-4-2	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium tuberculosis* H
Mtu-5-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium tuberculosis* H
Mtu-6-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumtuberculosis complex *Mycobacterium tuberculosis s*
Rsp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Nocardiaceae *Rhodococcus sp.* RHA1
Cae-1-1	Bacteria Actinobacteria Coriobacteridae Coriobacteriales Coriobacterineae Coriobacteriaceae *Collinsella aerofaciens* ATCC 25986
env-1 to env-5	environmental samples

64.2 Gene contexts

Each *pheA* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *pheA* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 64.3. The accession number of the sequence containing each *pheA* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 64.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Msp-3-1	NC_008146.1	+	4606235	4606322	**RNA**→ ←UvrD (COG0210)
Msp-2-1	NC_008705.1	+	4645440	4645527	**RNA**→ ←UvrD (COG0210)
Msp-1-1	NC_009077.1	+	4954634	4954721	**RNA**→ ←UvrD (COG0210)
env-1	hgutS7_s7_166638	+	622	707	**RNA**→ PTS_IIA_fru (cd00211)BglG (COG3711)→
env-2	BAAY01007111.1	+	589	675	**RNA**→
env-3	BAAX01005125.1	+	698	783	**RNA**→ ←HTH_XRE (cd00093)LexA (COG1974)
env-4	BABD01014011.1	+	353	438	**RNA**→
Cae-1-1	NZ_AAVN02000004.1	-	99317	99232	**RNA**→ ←Peptidase_C39A (cd02549)
env-5	hgutS8_s8_179076	-	1381	1296	**RNA**→ ←Peptidase_C39A (cd02549)
Mav-1-1	NC_008595.1	-	1008641	1008568	**RNA**→ PheA (COG1605)→
Mav-2-1	NC_002944.2	-	919307	919234	**RNA**→ PheA (COG1605)→
Mul-1-1	NC_008611.1	+	4895238	4895326	**RNA**→ PheA (COG1605)→
Rsp-1-1	NC_008268.1	-	5989802	5989717	**RNA**→ PheA (COG1605)→

Mle-1-1	NC_002677.1	-	213605	213517	**RNA**→ PheA (COG1605)→
Msm-1-1	NC_008596.1	+	5625243	5625335	**RNA**→ hypo→ PheA (COG1605)→
Msp-1-2	NC_009077.1	+	4954829	4954916	**RNA**→ PheA (COG1605)→
Msp-3-2	NC_008146.1	+	4606430	4606517	**RNA**→ PheA (COG1605)→
Msp-2-2	NC_008705.1	+	4645635	4645722	**RNA**→ PheA (COG1605)→
Mgi-1-1	NC_009338.1	-	1933485	1933397	**RNA**→ PheA (COG1605)→
Mva-1-1	NC_008726.1	+	5219413	5219500	**RNA**→ PheA (COG1605)→
Mtu-3-1	NC_009565.1	-	1061901	1061813	**RNA**→ PheA (COG1605)→
Mbo-2-1	NC_008769.1	-	1088552	1088464	**RNA**→ PheA (COG1605)→
Mtu-4-1	NC_009525.1	-	1059383	1059295	**RNA**→ PheA (COG1605)→ hypo→
Mtu-2-1	NC_002755.2	-	1058088	1058000	**RNA**→ PheA (COG1605)→ PheA (COG1605)→
Mtu-5-1	NC_000962.2	-	1058075	1057987	**RNA**→ PheA (COG1605)→
Mbo-1-1	NC_002945.3	-	1058542	1058454	**RNA**→ PheA (COG1605)→
Mtu-1-1	NZ_AAKR01000023.1	-	51233	51145	**RNA**→ PheA (COG1605)→
Mtu-6-1	NZ_AASN01000010.1	-	89007	88919	**RNA**→
Mtu-4-2	NZ_AAYK01000213.1	-	12222	12134	**RNA**→ PheA (COG1605)→

64.3 Conserved domains

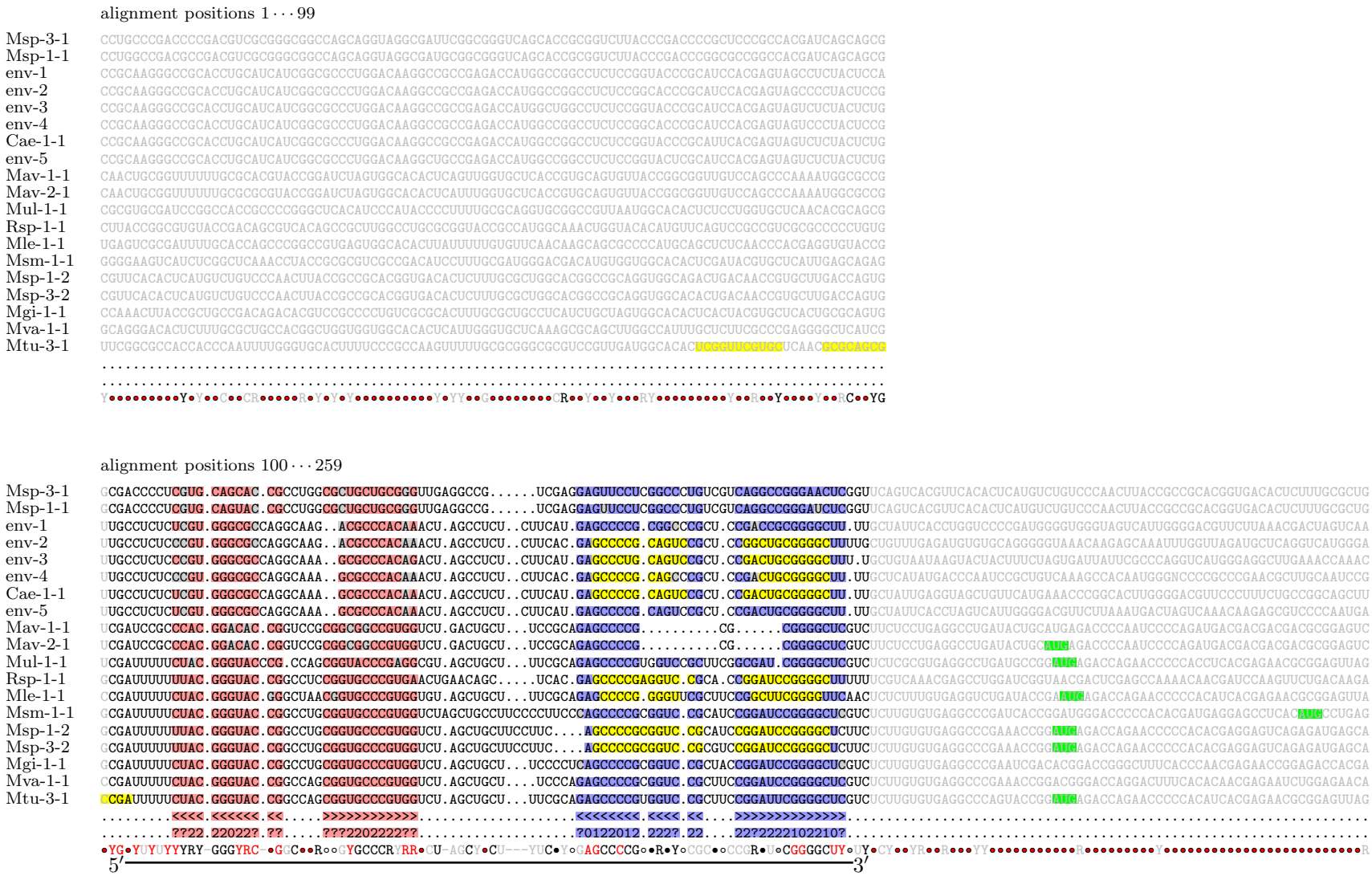
Conserved domains found in protein-coding genes listed in Section 64.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *pheA* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

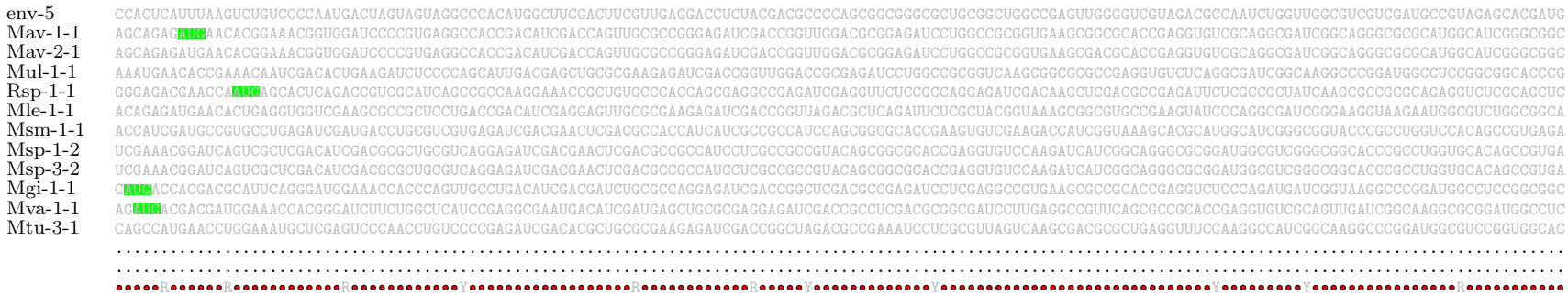
cd00093 (1) Helix-turn-helix XRE-family like proteins.
cd00211 (1) PTS_IIA, PTS system, fructose/mannitol specific IIA subunit.
cd02549 (2) A sub-family of peptidase family C39.
COG0210 (3) Superfamily I DNA and RNA helicases [DNA replication, recombination, and repair]
COG1605 (20) Chorismate mutase [Amino acid transport and metabolism]
COG1974 (1) SOS-response transcriptional repressors (RecA-mediated autopeptidases) [Transcription / Signal transduction mechanisms]
COG3711 (1) Transcriptional antiterminator [Transcription]

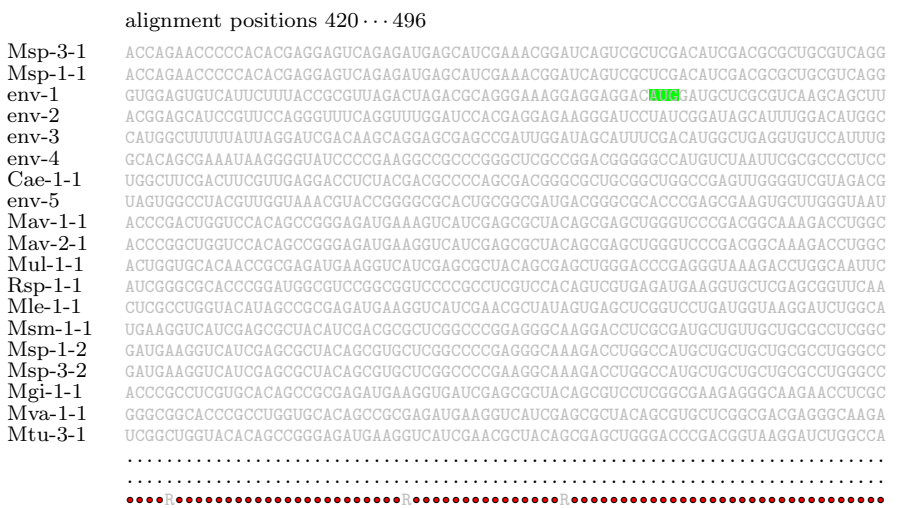
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Mbo-1-1, Mbo-2-1, Msp-2-1, Msp-2-2, Mtu-1-1, Mtu-2-1, Mtu-4-1, Mtu-4-2, Mtu-5-1, Mtu-6-1

64.4 Multiple-sequence alignment

Each *pheA* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 64.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *pheA* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.


alignment positions 1 · · · 99
Msp-3-1
Msp-1-1
env-1
env-2
env-3
env-4
Cae-1-1
env-5
Mav-1-1
Mav-2-1
Mul-1-1
Rsp-1-1
Mle-1-1
Msm-1-1
Msp-1-2
Msp-3-2
Mgi-1-1
Mva-1-1
Mtu-3-1
alignment positions 100 · · · 259
Msp-3-1
Msp-1-1
env-1
env-2
env-3
env-4
Cae-1-1
env-5
Mav-1-1
Mav-2-1
Mul-1-1
Rsp-1-1
Mle-1-1
Msm-1-1
Msp-1-2
Msp-3-2
Mgi-1-1
Mva-1-1
Mtu-3-1
5′
3′
alignment positions 260 · · · 419
Msp-3-1
Msp-1-1
env-1
env-2
env-3
env-4
Cae-1-1





65 PhotoRC-I

65.1 Taxa

The taxonomy of each organism containing a putative PhotoRC-I RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual PhotoRC-I RNA in Sections 65.2 and 65.4.

abbrev. of hits	taxonomy of species
Ssp-1-1 to Ssp-1-2	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* BL107
Ssp-2-1 to Ssp-2-2	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* CC9311
Ssp-3-1 to Ssp-3-2	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* CC9605
Ssp-4-1 to Ssp-4-2	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* CC9902
Ssp-5-1 to Ssp-5-2	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* RS9916
Ssp-6-1 to Ssp-6-2	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* RS9917
Ssp-7-1 to Ssp-7-2	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 7803
Ssp-8-1 to Ssp-8-2	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 7805

Ssp-9-1 to Ssp-9-2 Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 8102
env-1 to env-48 environmental samples

65.2 Gene contexts

Each PhotoRC-I RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one PhotoRC-I RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 65.3. The accession number of the sequence containing each PhotoRC-I RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 65.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Ssp-7-1	NC_009481.1	-	385675	385635	**RNA**→ Photo_RC (pfam00124)→
Ssp-7-2	NC_009481.1	-	1915281	1915241	**RNA**→ Photo_RC (pfam00124)→
Ssp-6-1	NZ_AANP01000006.1	+	165195	165236	**RNA**→ Photo_RC (pfam00124)→
Ssp-6-2	NZ_AANP01000005.1	+	140760	140801	**RNA**→ Photo_RC (pfam00124)→
env-1	JCVI_SCAF_1101668648542	-	1433	1392	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ DUF985 (pfam06172)? (GOS8942)→
env-2	JCVI_SCAF_1096627358224	-	1240	1199	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-3	JCVI_SCAF_1096627357520	+	730	771	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-4	JCVI_SCAF_1101668617945	+	196	237	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-5	JCVI_SCAF_1101668628378	-	1462	1421	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-6	JCVI_SCAF_1101668644252	-	556	515	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-7	JCVI_SCAF_1096627360362	+	2203	2244	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-8	JCVI_SCAF_1096627958055	+	460	501	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-9	JCVI_SCAF_1101668623404	+	812	853	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-10	JCVI_SCAF_1096627351991	-	456	415	**RNA**→ ? (GOS5184)→
env-11	JCVI_SCAF_1101667313216	-	478	437	**RNA**→ ? (GOS5184)→
env-12	JCVI_SCAF_1101668642101	-	421	380	**RNA**→ ? (GOS5184)→
env-13	JCVI_SCAF_1101668637008	-	1220	1179	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-14	JCVI_SCAF_1101667321149	+	22	63	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
Ssp-8-1	NZ_AAOK01000002.1	+	170526	170567	**RNA**→ Photo_RC (pfam00124)→ COG0385 (COG0385)→
Ssp-8-2	NZ_AAOK01000001.1	+	197766	197807	**RNA**→ Photo_RC (pfam00124)→
env-15	JCVI_SCAF_1096627984084	+	1580	1621	**RNA**→ ? (GOS5184)→
env-16	JCVI_SCAF_1096627070640	+	641	682	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-17	JCVI_SCAF_1096627383909	-	1588	1547	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS5184)→
env-18	JCVI_SCAF_1096627078472	-	633	592	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-19	JCVI_SCAF_1101668430728	-	287	246	**RNA**→ ? (GOS5184)→
env-20	JCVI_SCAF_1096628007816	+	535	576	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
Ssp-5-1	NZ_AAUA01000001.1	-	704108	704066	**RNA**→ Photo_RC (pfam00124)→
Ssp-5-2	NZ_AAUA01000004.1	-	159482	159440	**RNA**→ Photo_RC (pfam00124)→
Ssp-2-1	NC_008319.1	-	372998	372956	**RNA**→ Photo_RC (pfam00124)→
Ssp-2-2	NC_008319.1	-	2093117	2093075	**RNA**→ Photo_RC (pfam00124)→
env-21	JCVI_SCAF_1096626847236	+	34	75	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-22	JCVI_SCAF_1096627194864	-	853	811	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-23	JCVI_SCAF_1096627787846	+	983	1025	**RNA**→

env-24	JCVI_SCAF_1101668329101	-	1115	1073	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-25	JCVI_SCAF_1101668676638	+	634	676	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-26	JCVI_SCAF_1096627795270	-	405	363	**RNA**→ ? (GOS5184)→
env-27	JCVI_SCAF_1101667392251	-	632	589	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
Ssp-3-1	NC_007516.1	+	510580	510623	**RNA**→ Photo_RC (pfam00124)NOG06868 (NOG06868)→
Ssp-3-2	NC_007516.1	-	311054	311011	**RNA**→ Photo_RC (pfam00124)NOG06868 (NOG06868)→
env-28	JCVI_SCAF_1101668664162	+	682	725	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-29	JCVI_SCAF_1101667383566	+	127	170	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-30	JCVI_SCAF_1101668668415	+	1393	1436	**RNA**→
env-31	JCVI_SCAF_1096627623626	-	671	628	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-32	JCVI_SCAF_1096627754332	+	357	400	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS12083936)→
env-33	JCVI_SCAF_1096627317655	-	1935	1892	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS12083936)→ DUF985 (pfam06172)? (GOS8942)→
env-34	JCVI_SCAF_1096627370624	+	1310	1353	**RNA**→
env-35	JCVI_SCAF_1101668184704	-	710	667	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-36	JCVI_SCAF_1096627595498	+	1400	1443	**RNA**→ ? (GOS5184)→
env-37	JCVI_SCAF_1096627747941	+	1515	1558	**RNA**→
env-38	JCVI_SCAF_1101668188871	-	1204	1161	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-39	JCVI_SCAF_1101668670661	-	1225	1182	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-40	JCVI_SCAF_1101667330780	-	153	110	**RNA**→
env-41	JCVI_SCAF_1101668214255	+	1134	1177	**RNA**→ ? (GOS5184)→
env-42	JCVI_SCAF_1096627695962	-	109	66	**RNA**→
env-43	JCVI_SCAF_1096627194253	-	692	649	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-44	JCVI_SCAF_1096627198665	-	1324	1281	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS11942174)→
env-45	JCVI_SCAF_1101667428658	+	72	115	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
Ssp-9-1	NC_005070.1	-	1826079	1826036	**RNA**→ Photo_RC (pfam00124)NOG06868 (NOG06868)→
env-46	JCVI_SCAF_1096626093187	+	35	78	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
Ssp-9-2	NC_005070.1	+	2043166	2043209	**RNA**→ Photo_RC (pfam00124)NOG06868 (NOG06868)→ hypo→
env-47	JCVI_SCAF_1096628062325	-	827	784	**RNA**→ ? (GOS5712)→ ? (GOS721)→
env-48	JCVI_SCAF_1096628059963	+	1442	1485	**RNA**→ ? (GOS5184)→
Ssp-4-1	NC_007513.1	-	1741889	1741846	**RNA**→ Photo_RC (pfam00124)NOG06868 (NOG06868)→
Ssp-1-1	NZ_AATZ01000004.1	+	136676	136719	**RNA**→ Photo_RC (pfam00124)→
Ssp-1-2	NZ_AATZ01000002.1	-	657722	657679	**RNA**→ Photo_RC (pfam00124)→
Ssp-4-2	NC_007513.1	+	1945369	1945412	**RNA**→ Photo_RC (pfam00124)NOG06868 (NOG06868)→

65.3 Conserved domains

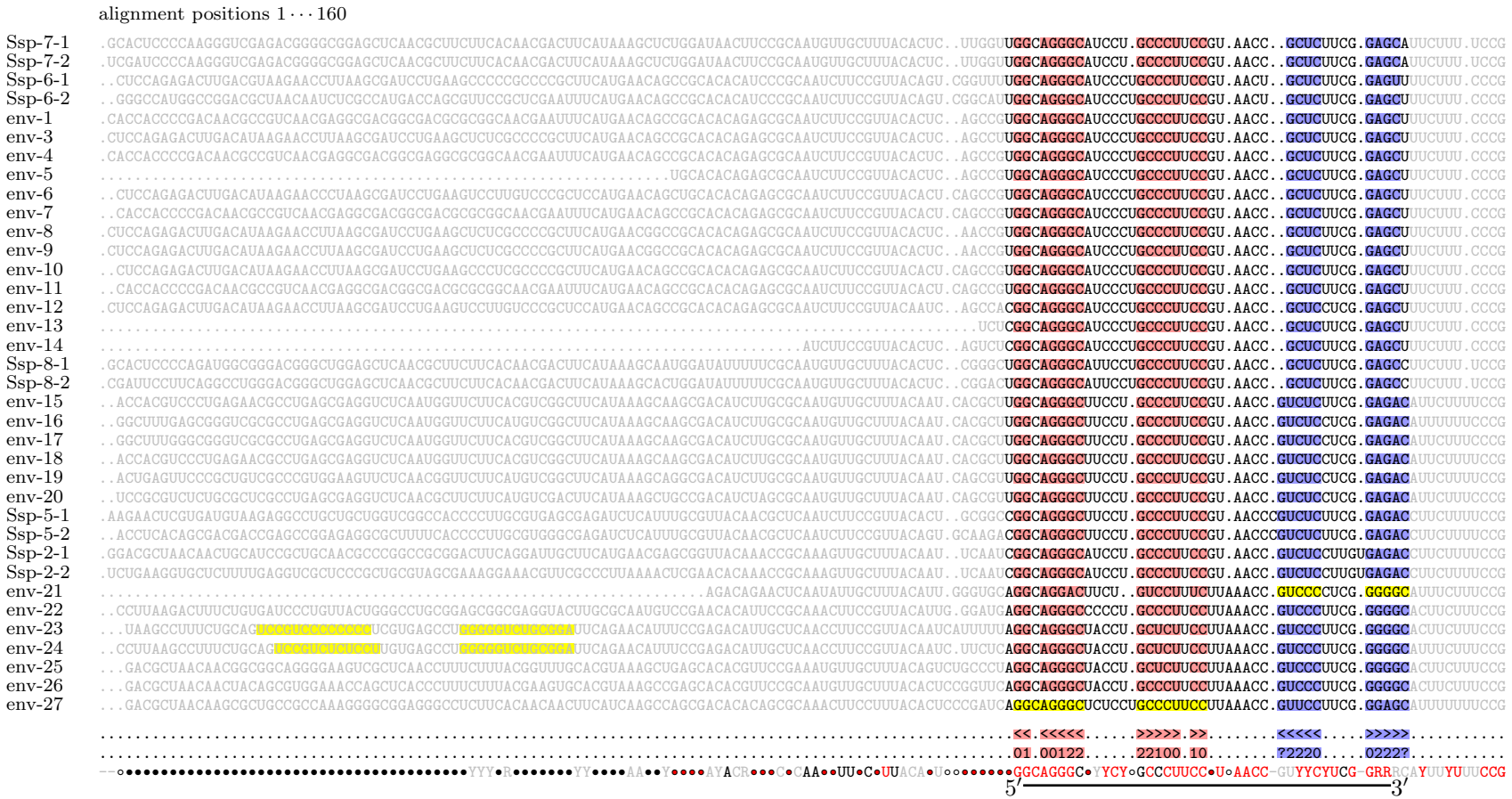
Conserved domains found in protein-coding genes listed in Section 65.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one PhotoRC-I RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

COG0385 (1) Predicted Na+-dependent transporter [General function prediction only]
GOS11942174 (1) ?
GOS12083936 (2) ?
GOS5184 (45) ?
GOS5712 (1) ?
GOS721 (1) ?
GOS8942 (2) ?
NOG06868 (6)
pfam00124 (51) Photosynthetic reaction centre protein.
pfam06172 (2) Protein of unknown function (DUF985).

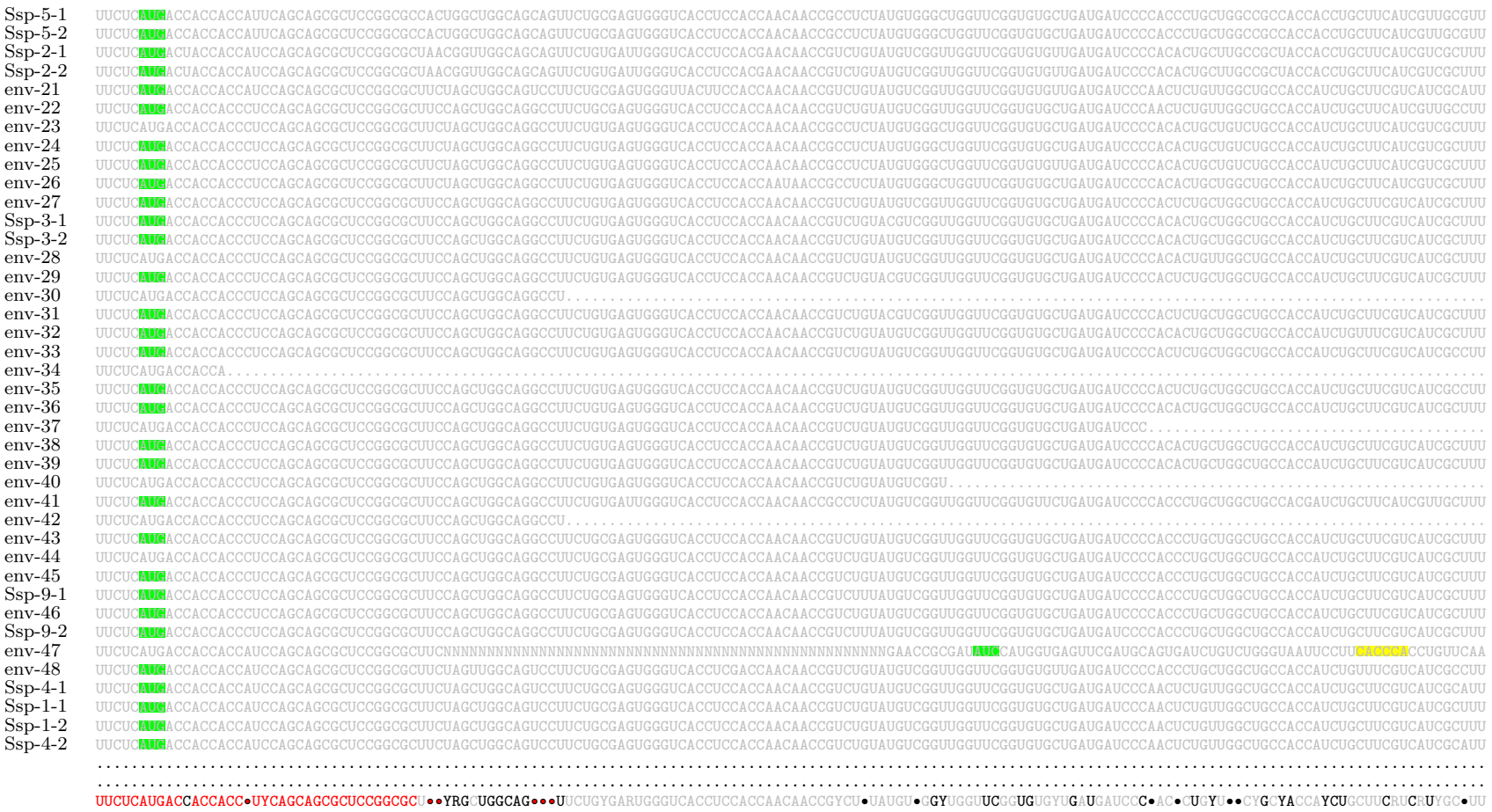
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: env-2

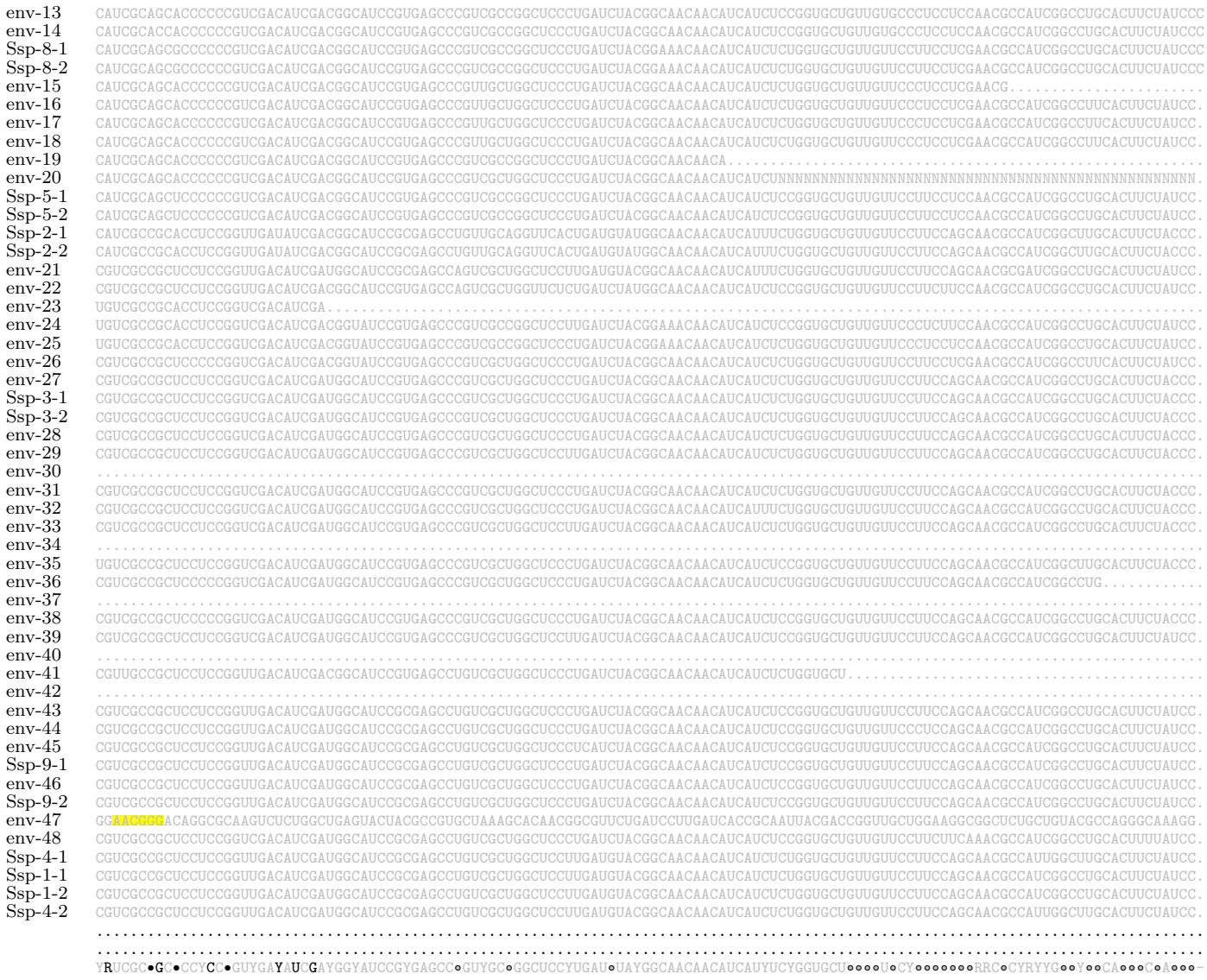
65.4 Multiple-sequence alignment

Each PhotoRC-I RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 65.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The PhotoRC-I RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (◦): 75%, white circle (◦): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



alignment positions 161 · · · 320


Ssp-5-1
Ssp-5-2
Ssp-2-1
Ssp-2-2
env-21
env-22
env-23
env-24
env-25
env-26
env-27
Ssp-3-1
Ssp-3-2
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
Ssp-9-1
env-46
Ssp-9-2
env-47
env-48
Ssp-4-1
Ssp-1-1
Ssp-1-2
Ssp-4-2
alignment positions 321 · · · 450
Ssp-7-1
Ssp-7-2
Ssp-6-1
Ssp-6-2
env-1
env-3
env-4
env-5
env-6
env-7
env-8
env-9
env-10
env-11
env-12



66 PhotoRC-II

66.1 Taxa

The taxonomy of each organism containing a putative PhotoRC-II RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual PhotoRC-II RNA in Sections 66.2 and 66.4.

abbrev. of hits	taxonomy of species
Pph-1-1	Viruses dsDNA viruses, no RNA stage Caudovirales Myoviridae T4-like viruses *Prochlorococcus phage* P-SSM4
env-1 to env-466	environmental samples

66.2 Gene contexts

Each PhotoRC-II RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one PhotoRC-II RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 66.3. The accession number of the sequence containing each PhotoRC-II RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 66.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	JCVI_SCAF_1101668145886	-	835	742	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-2	JCVI_SCAF_1096628394641	+	1565	1661	**RNA**→ ? (GOS5184)→
env-3	JCVI_SCAF_1101668269954	+	472	553	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-4	JCVI_SCAF_1101668735638	+	1372	1454	**RNA**→ ? (GOS14725)→
env-5	JCVI_SCAF_1101668736673	-	331	238	**RNA**→ ? (GOS5184)→
env-6	JCVI_SCAF_1101668461597	+	702	798	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-7	JCVI_SCAF_1101667029380	+	235	321	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-8	JCVI_SCAF_1101668740403	+	1207	1304	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-9	JCVI_SCAF_1096627094069	+	281	358	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-10	JCVI_SCAF_1096627025173	+	3994	4087	**RNA**→ ? (GOS5184)→
env-11	JCVI_SCAF_1096627636189	-	603	510	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-12	JCVI_SCAF_1096627010530	-	1058	952	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-13	JCVI_SCAF_1101667065672	+	555	662	**RNA**→
env-14	JCVI_SCAF_1101668736183	-	920	837	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-15	JCVI_SCAF_1101668361280	-	181	99	**RNA**→
env-16	JCVI_SCAF_1096627470674	-	374	286	**RNA**→ ? (GOS5184)→
env-17	JCVI_SCAF_1096626963273	-	644	558	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-18	JCVI_SCAF_1096626912631	+	441	550	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-19	JCVI_SCAF_1101668170117	+	1078	1186	**RNA**→ predicted protein→
env-20	JCVI_SCAF_1101668738346	+	1224	1332	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-21	JCVI_SCAF_1101668336309	-	263	155	**RNA**→
env-22	JCVI_SCAF_1096627166940	+	1159	1264	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-23	JCVI_SCAF_1096627990528	-	490	385	**RNA**→ ? (GOS14725)→
env-24	JCVI_SCAF_1096627870624	+	856	961	**RNA**→ ? (GOS12122029)→
env-25	JCVI_SCAF_1096627335206	+	1718	1823	**RNA**→
env-26	JCVI_SCAF_1096627166294	-	232	127	**RNA**→
env-27	JCVI_SCAF_1096627651027	+	482	579	**RNA**→ ? (GOS14725)→
env-28	JCVI_SCAF_1096626954874	+	269	349	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ NrdH (cd02976)? (GOS6441)→

env-29	JCVI_SCAF_1101667215307	+	441	521	**RNA**→ ? (GOS5184)→
env-30	JCVI_SCAF_1096626946312	+	658	742	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
Pph-1-1	NC_006884.1	+	170791	170875	**RNA**→
env-31	JCVI_SCAF_1101668662976	+	785	882	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-32	JCVI_SCAF_1096627104624	-	370	273	**RNA**→ ? (GOS5184)→
env-33	JCVI_SCAF_1096627595284	+	688	788	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS5184)→
env-34	JCVI_SCAF_1096626934894	+	1695	1793	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-35	JCVI_SCAF_1101667098365	-	281	191	**RNA**→ ? (GOS5184)→
env-36	JCVI_SCAF_1101668264510	-	1412	1315	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-37	JCVI_SCAF_1101667287216	-	122	30	**RNA**→
env-38	JCVI_SCAF_1096627464384	-	1173	1077	**RNA**→ ? (GOS5184)→
env-39	JCVI_SCAF_1096627177539	+	927	1024	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-40	JCVI_SCAF_1096627731277	-	608	510	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-41	JCVI_SCAF_1096627052295	+	845	944	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-42	JCVI_SCAF_1096627057334	+	801	899	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-43	JCVI_SCAF_1096626926948	+	666	764	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-44	JCVI_SCAF_1096626931783	+	1207	1305	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-45	JCVI_SCAF_1096627761059	+	1397	1495	**RNA**→ ? (GOS5184)→
env-46	JCVI_SCAF_1096627519287	-	1405	1307	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-47	JCVI_SCAF_1096627380576	+	2066	2164	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-48	JCVI_SCAF_1096627770928	+	447	545	**RNA**→ ? (GOS14725)→ Photo_RC (pfam00124)? (GOS5184)→
env-49	JCVI_SCAF_1096626957319	+	245	344	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-50	JCVI_SCAF_1096626996235	+	1380	1477	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-51	JCVI_SCAF_1096627618865	+	449	544	**RNA**→ ? (GOS5184)→
env-52	JCVI_SCAF_1096626969559	-	700	605	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-53	JCVI_SCAF_1096626957050	+	1469	1564	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-54	JCVI_SCAF_1096627612178	-	694	599	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-55	JCVI_SCAF_1096626962154	-	890	795	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-56	JCVI_SCAF_1096626981090	+	1089	1187	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-57	JCVI_SCAF_1096627568392	-	304	206	**RNA**→ ? (GOS5184)→
env-58	JCVI_SCAF_1096626155386	+	204	302	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-59	JCVI_SCAF_1096627574333	-	896	798	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-60	JCVI_SCAF_1096626651488	+	648	746	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-61	JCVI_SCAF_1096627053816	-	394	296	**RNA**→ ? (GOS5184)→ ? (GOS5184)→
env-62	JCVI_SCAF_1101668292268	-	848	750	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-63	JCVI_SCAF_1101667219081	+	141	239	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-64	JCVI_SCAF_1096626971465	-	798	700	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-65	JCVI_SCAF_1096627322220	-	1722	1624	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ FBP_aldolase_I_bact (cd00949)? (GOS1205)→
env-66	JCVI_SCAF_1101668694388	-	412	314	**RNA**→ ? (GOS14725)→
env-67	JCVI_SCAF_1101667122965	+	2	98	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-68	JCVI_SCAF_1096627320100	-	850	753	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-69	JCVI_SCAF_1101667214441	-	853	759	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-70	JCVI_SCAF_1096626985974	+	1073	1169	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-71	JCVI_SCAF_1096627057731	+	816	912	**RNA**→ ? (GOS5184)→
env-72	JCVI_SCAF_1096627058321	+	1529	1625	**RNA**→ ? (GOS5184)→
env-73	JCVI_SCAF_1096628014473	+	195	287	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS5184)→ ? (GOS5184)→
env-74	JCVI_SCAF_1101668018289	+	548	641	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-75	JCVI_SCAF_1096627094464	-	393	297	**RNA**→ ? (GOS5184)→
env-76	JCVI_SCAF_1096628018827	-	485	389	**RNA**→ ? (GOS5184)→
env-77	JCVI_SCAF_1096626917691	+	1417	1513	**RNA**→
env-78	JCVI_SCAF_1101668661405	+	879	977	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-79	JCVI_SCAF_1101668287630	-	475	380	**RNA**→ ? (GOS5184)→

env-80	JCVI_SCAF_1096627175407	-	802	706	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-81	JCVI_SCAF_1096628037299	-	352	256	**RNA**→ ? (GOS14725)→
env-82	JCVI_SCAF_1101667347463	-	339	243	**RNA**→ ? (GOS14725)→
env-83	JCVI_SCAF_1101668668819	+	1427	1523	**RNA**→
env-84	JCVI_SCAF_1096627181509	+	434	539	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-85	JCVI_SCAF_1101668260392	-	804	709	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-86	JCVI_SCAF_1096627380323	+	3152	3247	**RNA**→
env-87	JCVI_SCAF_1096626631643	+	744	839	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-88	JCVI_SCAF_1096628022421	+	873	968	**RNA**→ ? (GOS5184)→
env-89	JCVI_SCAF_1096627843116	+	1350	1445	**RNA**→
env-90	JCVI_SCAF_1096627870751	+	803	881	**RNA**→
env-91	JCVI_SCAF_1096627168565	+	2109	2189	**RNA**→
env-92	JCVI_SCAF_1101667073072	+	717	812	**RNA**→
env-93	JCVI_SCAF_1096628034240	-	131	36	**RNA**→
env-94	JCVI_SCAF_1096627318775	-	480	385	**RNA**→ ? (GOS5184)→
env-95	JCVI_SCAF_1096627099408	-	272	177	**RNA**→
env-96	JCVI_SCAF_1101668743230	+	1400	1493	**RNA**→
env-97	JCVI_SCAF_1101668699051	+	39	135	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-98	JCVI_SCAF_1096627986033	-	416	326	**RNA**→ ? (GOS5184)→
env-99	JCVI_SCAF_1101668169754	-	1392	1298	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-100	JCVI_SCAF_1096627168582	+	2114	2209	**RNA**→ ? (GOS5184)→
env-101	JCVI_SCAF_1096628037492	+	1035	1130	**RNA**→ ? (GOS5184)→
env-102	JCVI_SCAF_1101668256124	+	737	833	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-103	JCVI_SCAF_1096627166188	-	320	223	**RNA**→ ? (GOS14725)→
env-104	JCVI_SCAF_1096627165554	+	498	595	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-105	JCVI_SCAF_1096627179518	-	2082	1990	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS16164)→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-106	JCVI_SCAF_1096627369278	+	831	923	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ fer2 (cd00207)? (GOS9508)→ ? (GOS16164)→ ? (GOS11667)→ Photo_RC (pfam00124)? (GOS5184)→
env-107	JCVI_SCAF_1096628011253	+	1240	1334	**RNA**→ ? (GOS5184)→
env-108	JCVI_SCAF_1096627369480	-	980	885	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-109	JCVI_SCAF_1096627370048	-	815	720	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-110	JCVI_SCAF_1101668163438	-	419	324	**RNA**→ ? (GOS5184)→
env-111	JCVI_SCAF_1096627023265	+	1	91	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS5184)→
env-112	JCVI_SCAF_1101667175632	-	706	612	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-113	JCVI_SCAF_1101667176908	-	606	512	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-114	JCVI_SCAF_1096627017221	+	2282	2376	**RNA**→ ? (GOS14725)→
env-115	JCVI_SCAF_1101668244342	+	741	834	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-116	JCVI_SCAF_1096627007159	-	228	137	**RNA**→
env-117	JCVI_SCAF_1096627656396	+	1549	1640	**RNA**→ ? (GOS5184)→
env-118	JCVI_SCAF_1101668242450	-	169	78	**RNA**→
env-119	JCVI_SCAF_1096626191519	+	551	642	**RNA**→ ? (GOS14725)→
env-120	JCVI_SCAF_1096627012170	-	743	652	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-121	JCVI_SCAF_1101667178191	+	111	202	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-122	JCVI_SCAF_1096626195557	-	761	670	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-123	JCVI_SCAF_1096627010007	-	998	907	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-124	JCVI_SCAF_1101668237307	+	532	624	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-125	JCVI_SCAF_1096627012505	+	1310	1401	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-126	JCVI_SCAF_1096627007497	-	934	841	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-127	JCVI_SCAF_1096627024160	+	682	774	**RNA**→ ? (GOS5184)→ ? (GOS14691955)→ Photo_RC (pfam00124)? (GOS5184)→ Endonuclease_7 (pfam02945)? (GOS11484)→ Photo_RC (pfam00124)? (GOS5184)→
env-128	JCVI_SCAF_1101668253167	+	57	151	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-129	JCVI_SCAF_1096627392877	-	1451	1358	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→

env-130	JCVI_SCAF_1096627182374	-	1016	922	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-131	JCVI_SCAF_1096628026872	+	1130	1226	**RNA**→
env-132	JCVI_SCAF_1101667001806	-	257	162	**RNA**→
env-133	JCVI_SCAF_1096628024106	-	429	334	**RNA**→ ? (GOS5184)→
env-134	JCVI_SCAF_1096627396296	-	1490	1395	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS17023)→
env-135	JCVI_SCAF_1096627177034	+	583	678	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS16164)→ ? (GOS11049691)→ Photo_RC (pfam00124)? (GOS5184)→
env-136	JCVI_SCAF_1096628029483	-	1774	1680	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ predicted protein→ ? (GOS5184)→ ? (GOS11667)→ ? (GOS17023)→
env-137	JCVI_SCAF_1101667083080	-	794	701	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-138	JCVI_SCAF_1096627168342	+	39	133	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-139	JCVI_SCAF_1096627980574	+	1199	1293	**RNA**→ ? (GOS5184)→
env-140	JCVI_SCAF_1096628019210	+	522	615	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→ fer2 (cd00207)? (GOS9508)→
env-141	JCVI_SCAF_1096627182510	-	838	746	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-142	JCVI_SCAF_1096627286923	+	1085	1179	**RNA**→ ? (GOS5184)→
env-143	JCVI_SCAF_1096626848495	-	1075	980	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-144	JCVI_SCAF_1096627285375	+	3902	3997	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS5184)→
env-145	JCVI_SCAF_1101668005604	-	1423	1328	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-146	JCVI_SCAF_1101668008941	-	1042	947	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-147	JCVI_SCAF_1101667033800	+	39	134	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-148	JCVI_SCAF_1101667002217	-	682	587	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-149	JCVI_SCAF_1101667002377	-	628	533	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-150	JCVI_SCAF_1096626848975	-	333	238	**RNA**→ ? (GOS5184)→
env-151	JCVI_SCAF_1096626611574	+	9	104	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-152	JCVI_SCAF_1096626848446	+	742	837	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-153	JCVI_SCAF_1096627291072	+	568	663	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS11667)→ ? (GOS17023)→ Photo_RC (pfam00124)? (GOS5184)→
env-154	JCVI_SCAF_1096626857392	+	1261	1356	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-155	JCVI_SCAF_1096627405323	+	1422	1517	**RNA**→ ? (GOS5184)→
env-156	JCVI_SCAF_1101668046479	+	1153	1248	**RNA**→ ? (GOS5184)→
env-157	JCVI_SCAF_1096626875644	+	1208	1303	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-158	JCVI_SCAF_1101668040437	+	917	1012	**RNA**→ ? (GOS5184)→
env-159	JCVI_SCAF_1101668016070	+	1052	1147	**RNA**→ ? (GOS5184)→
env-160	JCVI_SCAF_1096628023051	-	361	269	**RNA**→ ? (GOS14725)→
env-161	JCVI_SCAF_1096627682273	+	355	449	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-162	JCVI_SCAF_1101668226855	-	1141	1047	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-163	JCVI_SCAF_1096627286105	+	1556	1647	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS16164)→ ? (GOS15599)→ Photo_RC (pfam00124)? (GOS5184)→
env-164	JCVI_SCAF_1101667450443	-	288	193	**RNA**→
env-165	JCVI_SCAF_1101667355987	-	252	155	**RNA**→
env-166	JCVI_SCAF_1096627302966	-	633	537	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-167	JCVI_SCAF_1096626716218	-	155	58	**RNA**→
env-168	JCVI_SCAF_1101668273469	-	1413	1313	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-169	JCVI_SCAF_1096627751112	-	506	409	**RNA**→ predicted protein→
env-170	JCVI_SCAF_1101668737534	-	577	483	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-171	JCVI_SCAF_1096628038160	+	708	804	**RNA**→ ? (GOS5184)→
env-172	JCVI_SCAF_1096627046103	-	1092	996	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-173	JCVI_SCAF_1101667201221	-	409	311	**RNA**→ ? (GOS5184)→
env-174	JCVI_SCAF_1096626936052	-	1404	1308	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-175	JCVI_SCAF_1101667201261	+	307	404	**RNA**→
env-176	JCVI_SCAF_1096627756729	-	666	569	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-177	JCVI_SCAF_1096627604052	-	1631	1538	**RNA**→ ? (GOS5184)→
env-178	JCVI_SCAF_1096627760234	-	558	465	**RNA**→ ? (GOS5184)→

env-179	JCVI_SCAF_1101667225197	+	222	315	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-180	JCVI_SCAF_1096627055940	+	673	766	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-181	JCVI_SCAF_1101667199569	+	463	556	**RNA**→ ? (GOS5184)→ ? (GOS5184)→
env-182	JCVI_SCAF_1096627055486	-	1578	1485	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-183	JCVI_SCAF_1096627722016	-	203	107	**RNA**→
env-184	JCVI_SCAF_1101668503669	+	573	657	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-185	JCVI_SCAF_1101668682639	+	954	1042	**RNA**→ ? (GOS5184)→
env-186	JCVI_SCAF_1101668516185	+	794	881	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-187	JCVI_SCAF_1101668541498	-	466	381	**RNA**→ ? (GOS5184)→
env-188	JCVI_SCAF_1101668355240	-	114	28	**RNA**→
env-189	JCVI_SCAF_1101668459296	-	411	324	**RNA**→ ? (GOS14725)→
env-190	JCVI_SCAF_1101668478872	-	700	616	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-191	JCVI_SCAF_1096627890505	-	907	823	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-192	JCVI_SCAF_1096628396232	+	916	1002	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-193	JCVI_SCAF_1096628234340	-	249	163	**RNA**→
env-194	JCVI_SCAF_1096628232739	+	151	233	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ Transaldolase_FSA (cd00956)? (GOS133)→
env-195	JCVI_SCAF_1101667116905	-	437	328	**RNA**→ ? (GOS5184)→
env-196	JCVI_SCAF_1096626958699	-	2416	2307	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-197	JCVI_SCAF_1101667102314	-	154	45	**RNA**→
env-198	JCVI_SCAF_1101668543023	-	1148	1059	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-199	JCVI_SCAF_1096627160490	+	1148	1237	**RNA**→ ? (GOS12122029)→
env-200	JCVI_SCAF_1101668283203	+	635	724	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-201	JCVI_SCAF_1101668511277	+	997	1085	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-202	JCVI_SCAF_1101668523685	+	821	909	**RNA**→ ? (GOS5184)→
env-203	JCVI_SCAF_1096627113251	-	692	603	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-204	JCVI_SCAF_1096627888393	+	1152	1241	**RNA**→ ? (GOS5184)→
env-205	JCVI_SCAF_1101668685064	+	934	1023	**RNA**→ ? (GOS5184)→
env-206	JCVI_SCAF_1101668536575	+	1533	1622	**RNA**→
env-207	JCVI_SCAF_1096627618648	+	315	415	**RNA**→ ? (GOS5184)→ ? (GOS5184)→ ? (GOS13236823)→
env-208	JCVI_SCAF_1096627769958	-	921	821	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-209	JCVI_SCAF_1101667258422	+	664	762	**RNA**→
env-210	JCVI_SCAF_1096626963113	+	649	747	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS13236823)→
env-211	JCVI_SCAF_1096626111347	+	160	258	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-212	JCVI_SCAF_1096627754377	-	232	134	**RNA**→
env-213	JCVI_SCAF_1101668288475	-	597	499	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-214	JCVI_SCAF_1096626984839	+	1379	1478	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-215	JCVI_SCAF_1096627569467	+	749	845	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-216	JCVI_SCAF_1101668277535	+	966	1064	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-217	JCVI_SCAF_1096627044725	+	129	224	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-218	JCVI_SCAF_1096628005240	+	1454	1551	**RNA**→
env-219	JCVI_SCAF_1096628009532	+	1340	1437	**RNA**→ ? (GOS5184)→
env-220	JCVI_SCAF_1101668723508	-	409	315	**RNA**→ ? (GOS5184)→
env-221	JCVI_SCAF_1096626722011	-	1186	1091	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-222	JCVI_SCAF_1101668730926	+	588	683	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-223	JCVI_SCAF_1101668715312	+	344	437	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-224	JCVI_SCAF_1101667349428	-	665	572	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-225	JCVI_SCAF_1101668736635	+	914	1007	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-226	JCVI_SCAF_1096627999758	+	1090	1183	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-227	JCVI_SCAF_1096628004431	+	1318	1411	**RNA**→ ? (GOS5184)→
env-228	JCVI_SCAF_1096628013145	+	159	252	**RNA**→ ? (GOS5184)→ ? (GOS5184)→
env-229	JCVI_SCAF_1096628009689	+	727	820	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-230	JCVI_SCAF_1101668728652	-	1162	1066	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→

env-231	JCVI_SCAF_1096626122918	-	943	846	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-232	JCVI_SCAF_1101667222016	+	445	538	**RNA**→ ? (GOS5184)→
env-233	JCVI_SCAF_1096627863713	-	1156	1060	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-234	JCVI_SCAF_1096627730548	-	912	817	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-235	JCVI_SCAF_1096627596685	+	1263	1358	**RNA**→ ? (GOS5184)→
env-236	JCVI_SCAF_1096627321943	-	487	393	**RNA**→ ? (GOS5184)→
env-237	JCVI_SCAF_1096627315747	-	1860	1766	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-238	JCVI_SCAF_1096627626344	-	701	607	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-239	JCVI_SCAF_1096627767613	-	199	105	**RNA**→
env-240	JCVI_SCAF_1101667226368	-	317	223	**RNA**→ ? (GOS5184)→
env-241	JCVI_SCAF_1096626176586	+	359	453	**RNA**→ ? (GOS5184)→
env-242	JCVI_SCAF_1096627316135	+	1012	1106	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-243	JCVI_SCAF_1101668182803	+	998	1092	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-244	JCVI_SCAF_1101668180624	-	994	900	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-245	JCVI_SCAF_1096626984236	-	919	825	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-246	JCVI_SCAF_1096627551591	+	443	537	**RNA**→ ? (GOS5184)→
env-247	JCVI_SCAF_1101667099554	+	168	262	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-248	JCVI_SCAF_1101668228317	+	1266	1360	**RNA**→ ? (GOS5184)→
env-249	JCVI_SCAF_1101667126431	+	203	297	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-250	JCVI_SCAF_1096627608448	-	721	627	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-251	JCVI_SCAF_1096626984912	-	1841	1747	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-252	JCVI_SCAF_1101667122650	+	53	147	**RNA**→ ? (GOS5184)→
env-253	JCVI_SCAF_1096627536554	+	473	567	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-254	JCVI_SCAF_1096626984647	-	482	388	**RNA**→ ? (GOS5184)→
env-255	JCVI_SCAF_1096627050473	+	3058	3152	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-256	JCVI_SCAF_1096627542854	+	1324	1419	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-257	JCVI_SCAF_1096627746705	+	1316	1414	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-258	JCVI_SCAF_1101668182082	+	903	1001	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-259	JCVI_SCAF_1101668232856	-	336	238	**RNA**→ ? (GOS5184)→
env-260	JCVI_SCAF_1101668181423	+	549	643	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-261	JCVI_SCAF_1096626958514	-	945	851	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-262	JCVI_SCAF_1096627775462	+	843	937	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-263	JCVI_SCAF_1096627632630	+	256	350	**RNA**→ ? (GOS5184)→
env-264	JCVI_SCAF_1101668227572	+	1163	1258	**RNA**→ ? (GOS5184)→
env-265	JCVI_SCAF_1101667101678	+	699	793	**RNA**→
env-266	JCVI_SCAF_1101668730427	+	238	335	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-267	JCVI_SCAF_1101667353824	+	6	86	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-268	JCVI_SCAF_1101668727855	+	982	1077	**RNA**→
env-269	JCVI_SCAF_1096628002645	+	301	397	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS5184)→ ? (GOS5184)→
env-270	JCVI_SCAF_1101667585598	+	107	202	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-271	JCVI_SCAF_1096627050228	-	1554	1459	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS5184)→
env-272	JCVI_SCAF_1096627026548	-	1850	1755	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-273	JCVI_SCAF_1096627054082	+	1294	1389	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-274	JCVI_SCAF_1101668201319	-	1120	1025	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-275	JCVI_SCAF_1101668320141	-	1405	1310	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-276	JCVI_SCAF_1101667142490	-	761	666	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-277	JCVI_SCAF_1096626930282	+	796	891	**RNA**→ ? (GOS5184)→
env-278	JCVI_SCAF_1096626994968	+	629	724	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-279	JCVI_SCAF_1101667130594	-	277	182	**RNA**→ ? (GOS5184)→
env-280	JCVI_SCAF_1096627710961	-	723	628	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-281	JCVI_SCAF_1101667333165	+	56	151	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-282	JCVI_SCAF_1096627707988	-	1418	1323	**RNA**→ ? (GOS5184)→ ? (GOS5184)→

env-283	JCVI_SCAF_1096627058196	-	518	423	**RNA**→ ? (GOS5184)→
env-284	JCVI_SCAF_1096627838011	-	1284	1189	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-285	JCVI_SCAF_1101667142929	-	719	624	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-286	JCVI_SCAF_1096627381298	-	1392	1297	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-287	JCVI_SCAF_1101667224709	+	255	350	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-288	JCVI_SCAF_1096627710197	+	398	493	**RNA**→ ? (GOS5184)→ ? (GOS5184)→
env-289	JCVI_SCAF_1096627741804	-	1300	1205	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-290	JCVI_SCAF_1101668285781	-	1474	1379	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-291	JCVI_SCAF_1101668299507	-	1380	1285	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-292	JCVI_SCAF_1096626932606	-	315	220	**RNA**→ ? (GOS5184)→
env-293	JCVI_SCAF_1096627767996	+	659	754	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-294	JCVI_SCAF_1101667209554	-	187	93	**RNA**→
env-295	JCVI_SCAF_1101667132249	-	102	7	**RNA**→
env-296	JCVI_SCAF_1101667225521	+	323	418	**RNA**→
env-297	JCVI_SCAF_1101668225210	+	927	1023	**RNA**→ ? (GOS5184)→
env-298	JCVI_SCAF_1101667130225	-	518	422	**RNA**→ ? (GOS5184)→
env-299	JCVI_SCAF_1096627308383	+	356	452	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-300	JCVI_SCAF_1096626972338	+	2141	2237	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-301	JCVI_SCAF_1096627558561	-	889	793	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-302	JCVI_SCAF_1101667141403	+	603	699	**RNA**→
env-303	JCVI_SCAF_1096627566621	+	1417	1513	**RNA**→ ? (GOS5184)→
env-304	JCVI_SCAF_1101667203478	+	224	319	**RNA**→ ? (GOS5184)→
env-305	JCVI_SCAF_1096628273009	+	1582	1678	**RNA**→
env-306	JCVI_SCAF_1096626929650	+	5	100	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-307	JCVI_SCAF_1101668267155	-	993	898	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-308	JCVI_SCAF_1096626951160	+	1099	1199	**RNA**→ ? (GOS5184)→
env-309	JCVI_SCAF_1101667104108	-	726	626	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-310	JCVI_SCAF_1096627560839	-	726	626	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-311	JCVI_SCAF_1096626307653	+	67	160	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-312	JCVI_SCAF_1096627056996	-	910	812	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-313	JCVI_SCAF_1096627065901	+	1153	1249	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-314	JCVI_SCAF_1096627263771	-	1374	1279	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-315	JCVI_SCAF_1101668346464	-	755	654	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-316	JCVI_SCAF_1096626225851	+	310	402	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-317	JCVI_SCAF_1096627595241	+	1	95	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS5184)→
env-318	JCVI_SCAF_1096627551456	-	1405	1313	**RNA**→ ? (GOS5184)→ ? (GOS5184)→ ? (GOS13909)→ ? (GOS7664)→
env-319	JCVI_SCAF_1096627748636	+	199	306	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS5184)→
env-320	JCVI_SCAF_1096627716232	+	475	573	**RNA**→ ? (GOS5184)→ ? (GOS5184)→ ? (GOS1205)→
env-321	JCVI_SCAF_1096627526776	+	666	762	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-322	JCVI_SCAF_1096627041526	-	161	60	**RNA**→ ? (GOS5184)→
env-323	JCVI_SCAF_1096627055819	-	2005	1907	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-324	JCVI_SCAF_1096627785158	-	162	64	**RNA**→
env-325	JCVI_SCAF_1096627557851	-	143	48	**RNA**→
env-326	JCVI_SCAF_1101668271280	+	945	1041	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-327	JCVI_SCAF_1096627760811	+	1519	1616	**RNA**→
env-328	JCVI_SCAF_1096627593034	+	1389	1487	**RNA**→ ? (GOS5184)→
env-329	JCVI_SCAF_1096627560201	-	634	540	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-330	JCVI_SCAF_1096627705536	+	1437	1530	**RNA**→ ? (GOS5184)→
env-331	JCVI_SCAF_1101668208180	+	763	857	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-332	JCVI_SCAF_1096627543309	+	152	245	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-333	JCVI_SCAF_1096627714386	-	582	489	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-334	JCVI_SCAF_1096627758463	+	364	458	**RNA**→ ? (GOS5184)→ ? (GOS5184)→ ? (GOS13909)→

env-335	JCVI_SCAF_1096627382162	+	2321	2415	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-336	JCVI_SCAF_1096626980834	+	1659	1754	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-337	JCVI_SCAF_1101668291157	+	1623	1721	**RNA**→
env-338	JCVI_SCAF_1096627389395	-	3808	3717	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS17023)→
env-339	JCVI_SCAF_1096627602047	-	428	331	**RNA**→ ? (GOS5184)→
env-340	JCVI_SCAF_1101668202738	+	764	861	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-341	JCVI_SCAF_1096627737735	-	721	627	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-342	JCVI_SCAF_1096627782477	+	910	1003	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-343	JCVI_SCAF_1101667107204	-	447	354	**RNA**→ ? (GOS5184)→
env-344	JCVI_SCAF_1096626954123	+	1060	1153	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-345	JCVI_SCAF_1096626960853	+	102	196	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ FBP_aldolase_I_a (cd00948)? (GOS1205)→
env-346	JCVI_SCAF_1096627724170	-	1371	1278	**RNA**→ ? (GOS5184)→ ? (GOS5184)→ ? (GOS1205)→
env-347	JCVI_SCAF_1096626930686	+	586	680	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-348	JCVI_SCAF_1096627597054	-	979	886	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-349	JCVI_SCAF_1096626981244	+	1396	1490	**RNA**→ ? (GOS5184)→
env-350	JCVI_SCAF_1096627055616	+	148	239	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS5184)→
env-351	JCVI_SCAF_1101668271599	+	984	1076	**RNA**→ ←? (GOS13482251)
env-352	JCVI_SCAF_1101668315155	-	489	397	**RNA**→ ? (GOS5184)→
env-353	JCVI_SCAF_1101667138541	-	527	435	**RNA**→ ←? (GOS13482251)
env-354	JCVI_SCAF_1101667211228	-	705	613	**RNA**→ ←? (GOS13482251)
env-355	JCVI_SCAF_1101668303290	-	1272	1180	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-356	JCVI_SCAF_1101667218481	+	694	786	**RNA**→ ? (GOS5184)→
env-357	JCVI_SCAF_1096627769584	-	1203	1111	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS5184)→
env-358	JCVI_SCAF_1101667135048	+	663	755	**RNA**→
env-359	JCVI_SCAF_1096626942091	+	483	575	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-360	JCVI_SCAF_1096627047791	+	1949	2041	**RNA**→
env-361	JCVI_SCAF_1096626967995	+	662	754	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-362	JCVI_SCAF_1096626997884	-	2089	1997	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-363	JCVI_SCAF_1096626978634	-	924	832	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-364	JCVI_SCAF_1096627572272	-	1341	1249	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-365	JCVI_SCAF_1101668284553	-	647	555	**RNA**→ ←? (GOS13482251)
env-366	JCVI_SCAF_1101668296868	-	383	291	**RNA**→ ←? (GOS13482251)
env-367	JCVI_SCAF_1096627058848	+	691	783	**RNA**→ ←? (GOS13482251)
env-368	JCVI_SCAF_1096627382288	-	2964	2872	**RNA**→ ←? (GOS13482251)
env-369	JCVI_SCAF_1101667222949	+	365	457	**RNA**→ ←? (GOS13482251)
env-370	JCVI_SCAF_1096627764980	-	636	541	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-371	JCVI_SCAF_1096627771283	+	1177	1272	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-372	JCVI_SCAF_1101668298883	+	1169	1264	**RNA**→ ? (GOS5184)→
env-373	JCVI_SCAF_1096627171064	+	1434	1529	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-374	JCVI_SCAF_1096627999423	-	1171	1078	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS5184)→
env-375	JCVI_SCAF_1096627972420	+	1176	1269	**RNA**→ ? (GOS5184)→
env-376	JCVI_SCAF_1101668733431	+	612	706	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ predicted protein→
env-377	JCVI_SCAF_1096627171368	-	975	882	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-378	JCVI_SCAF_1101667350638	-	573	478	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-379	JCVI_SCAF_1096627170071	+	1551	1647	**RNA**→ ? (GOS5184)→
env-380	JCVI_SCAF_1096627765614	+	809	905	**RNA**→ ? (GOS5184)→
env-381	JCVI_SCAF_1096627162087	+	1157	1253	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-382	JCVI_SCAF_1096627705999	-	953	860	**RNA**→ ? (GOS5184)→ ? (GOS5184)→
env-383	JCVI_SCAF_1101667353653	+	491	584	**RNA**→ ? (GOS5184)→ ? (GOS5184)→
env-384	JCVI_SCAF_1096628002965	-	260	166	**RNA**→ ? (GOS5184)→
env-385	JCVI_SCAF_1096628011544	+	42	136	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-386	JCVI_SCAF_1101668704329	+	24	118	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→

env-387	JCVI_SCAF_1096627088370	-	919	825	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-388	JCVI_SCAF_1096626958669	+	500	594	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-389	JCVI_SCAF_1096627731752	-	491	398	**RNA**→ ? (GOS5184)→
env-390	JCVI_SCAF_1096626938383	+	1232	1325	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-391	JCVI_SCAF_1101668210846	-	1065	972	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-392	JCVI_SCAF_1101668693828	+	836	931	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-393	JCVI_SCAF_1101668708608	+	357	452	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-394	JCVI_SCAF_1096627169396	+	311	406	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-395	JCVI_SCAF_1101667211999	-	571	479	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-396	JCVI_SCAF_1101667531418	+	177	271	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-397	JCVI_SCAF_1096627568803	-	759	665	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-398	JCVI_SCAF_1096627738885	-	613	519	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-399	JCVI_SCAF_1096626963926	-	1077	983	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-400	JCVI_SCAF_1096626965120	-	1969	1875	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-401	JCVI_SCAF_1096627715831	+	910	1004	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-402	JCVI_SCAF_1096627055876	+	2612	2705	**RNA**→ ? (GOS5184)→
env-403	JCVI_SCAF_1096627058540	+	2318	2414	**RNA**→ ? (GOS5184)→
env-404	JCVI_SCAF_1101668667692	+	1220	1316	**RNA**→ ? (GOS5184)→
env-405	JCVI_SCAF_1101667205985	-	727	632	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-406	JCVI_SCAF_1096626249123	+	143	234	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-407	JCVI_SCAF_1101667106088	+	629	722	**RNA**→ ? (GOS5184)→
env-408	JCVI_SCAF_1101668187804	+	1183	1275	**RNA**→ ? (GOS5184)→
env-409	JCVI_SCAF_1096627725473	-	1300	1208	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS5184)→
env-410	JCVI_SCAF_1096627051824	+	561	653	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-411	JCVI_SCAF_1096627719197	+	592	684	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-412	JCVI_SCAF_1096627775356	-	314	222	**RNA**→ ? (GOS5184)→
env-413	JCVI_SCAF_1101668297370	+	1089	1181	**RNA**→ ? (GOS5184)→
env-414	JCVI_SCAF_1096626248335	-	134	40	**RNA**→
env-415	JCVI_SCAF_1096627044404	-	754	660	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-416	JCVI_SCAF_1096627057497	+	723	817	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS5184)→
env-417	JCVI_SCAF_1096627710339	+	213	306	**RNA**→ ? (GOS5184)→
env-418	JCVI_SCAF_1096627722063	-	269	176	**RNA**→ ? (GOS5184)→
env-419	JCVI_SCAF_1101668232687	-	1007	914	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-420	JCVI_SCAF_1096627043198	+	395	488	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS11484)→
env-421	JCVI_SCAF_1096627744946	+	1514	1607	**RNA**→ ? (GOS5184)→
env-422	JCVI_SCAF_1101667214905	+	456	549	**RNA**→ ? (GOS5184)→ ? (GOS5184)→
env-423	JCVI_SCAF_1096627317941	-	941	848	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-424	JCVI_SCAF_1096627963947	-	479	386	**RNA**→ ? (GOS5184)→
env-425	JCVI_SCAF_1101668311345	+	1166	1259	**RNA**→ ? (GOS5184)→
env-426	JCVI_SCAF_1101668290907	+	1398	1491	**RNA**→ ? (GOS5184)→
env-427	JCVI_SCAF_1096627055378	-	763	670	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-428	JCVI_SCAF_1101668223745	-	1534	1440	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-429	JCVI_SCAF_1096627304898	+	908	1002	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-430	JCVI_SCAF_1096626973470	+	1074	1168	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-431	JCVI_SCAF_1096627046483	+	2313	2406	**RNA**→
env-432	JCVI_SCAF_1101668294065	+	1380	1473	**RNA**→
env-433	JCVI_SCAF_1096627055630	-	2148	2068	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-434	JCVI_SCAF_1096627047491	+	791	884	**RNA**→ ? (GOS5184)→
env-435	JCVI_SCAF_1101668295461	-	332	239	**RNA**→ ? (GOS5184)→
env-436	JCVI_SCAF_1096627730061	-	999	904	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS5184)→
env-437	JCVI_SCAF_1096627596052	+	911	1005	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-438	JCVI_SCAF_1101667218504	-	686	591	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→

env-439	JCVI_SCAF_1096627745463	+	1169	1264	**RNA**→ ? (GOS5184)→
env-440	JCVI_SCAF_1096627773919	-	492	396	**RNA**→ ? (GOS5184)→
env-441	JCVI_SCAF_1096626952722	-	1604	1510	**RNA**→ predicted protein→ Photo_RC (pfam00124)? (GOS5184)→ HNHc (cd00085)? (GOS11970)→
env-442	JCVI_SCAF_1096626991215	+	66	160	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-443	JCVI_SCAF_1096626949868	-	1213	1116	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-444	JCVI_SCAF_1096627043739	-	724	627	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-445	JCVI_SCAF_1096627172387	+	1228	1323	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS5184)→
env-446	JCVI_SCAF_1096627710828	+	215	308	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-447	JCVI_SCAF_1101667097099	-	507	414	**RNA**→ ? (GOS5184)→
env-448	JCVI_SCAF_1096627245141	+	1468	1561	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-449	JCVI_SCAF_1101667124924	+	427	520	**RNA**→ ? (GOS5184)→
env-450	JCVI_SCAF_1096627718236	-	1105	1011	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-451	JCVI_SCAF_1096627305863	+	431	525	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-452	JCVI_SCAF_1101667543230	+	269	363	**RNA**→ ? (GOS5184)→
env-453	JCVI_SCAF_1096627003584	-	1277	1183	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-454	JCVI_SCAF_1096626250697	-	826	733	**RNA**→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-455	JCVI_SCAF_1096627720223	+	49	142	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS5184)→ Photo_RC (pfam00124)? (GOS5184)→
env-456	JCVI_SCAF_1096628022640	+	975	1072	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-457	JCVI_SCAF_1101668693837	+	1018	1114	**RNA**→ ? (GOS5184)→
env-458	JCVI_SCAF_1096627054839	+	102	201	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS1205)→ FBP_aldolase_I_a (cd00948)? (GOS1205)→
env-459	JCVI_SCAF_1096627675232	+	535	631	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→ ? (GOS5184)→
env-460	JCVI_SCAF_1096627025173	+	2592	2685	**RNA**→ ? (GOS11938944)→
env-461	JCVI_SCAF_1096627103424	-	728	633	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-462	JCVI_SCAF_1096627980844	+	1153	1248	**RNA**→ ? (GOS14725)→
env-463	JCVI_SCAF_1101668686991	+	802	901	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-464	JCVI_SCAF_1096627074648	+	863	955	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-465	JCVI_SCAF_1101668147882	+	617	725	**RNA**→ Photo_RC (pfam00124)? (GOS5184)→
env-466	JCVI_SCAF_1096627018250	-	651	540	**RNA**→ ? (GOS14725)→

66.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 66.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one PhotoRC-II RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

cd00085 (1) HNH nucleases; HNH endonuclease signature which is found in viral, prokaryotic, and eukaryotic proteins.
cd00207 (2) 2Fe-2S iron-sulfur cluster binding domain.
cd00948 (2) Fructose-1,6-bisphosphate aldolase.
cd00949 (1) Fructose-1.
cd00956 (1) Transaldolase-like fructose-6-phosphate aldolases (FSA) found in bacteria and archaea, which are member of the MipB/TalC subfamily of class I aldolases.
cd02976 (1) NrdH-redoxin (NrdH) family; NrdH is a small monomeric protein with a conserved redox active CXXC motif within a TRX fold, characterized by a glutaredoxin (GRX)-like sequence and TRX-like activity profile.
GOS11049691 (1) ?
GOS11484 (2) ?
GOS11667 (3) ?
GOS11938944 (1) ?
GOS11970 (1) ?
GOS1205 (6) ?
GOS12122029 (2) ?
GOS13236823 (2) ?
GOS133 (1) ?
GOS13482251 (8) ?
GOS13909 (2) ?
GOS14691955 (1) ?
GOS14725 (14) ?
GOS15599 (1) ?
GOS16164 (4) ?
GOS17023 (4) ?
GOS5184 (474) ?
GOS6441 (1) ?

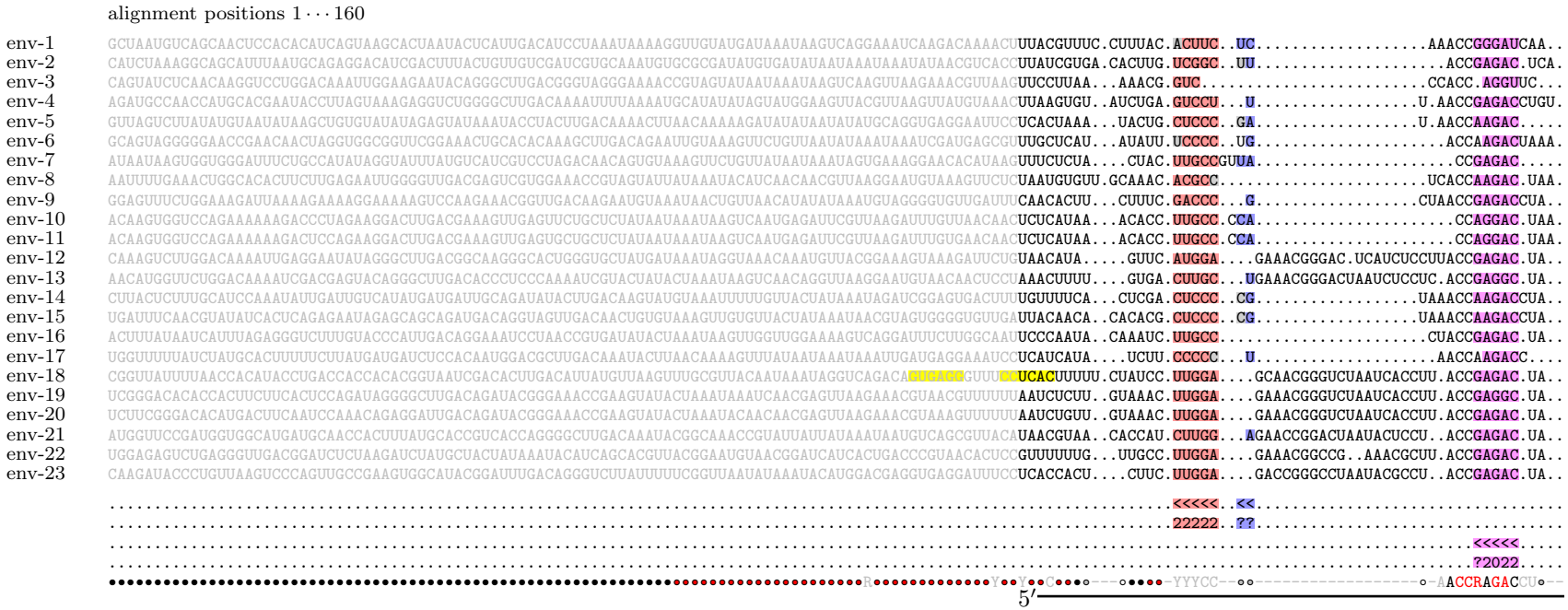
GOS7664 (1) ?
GOS9508 (2) ?

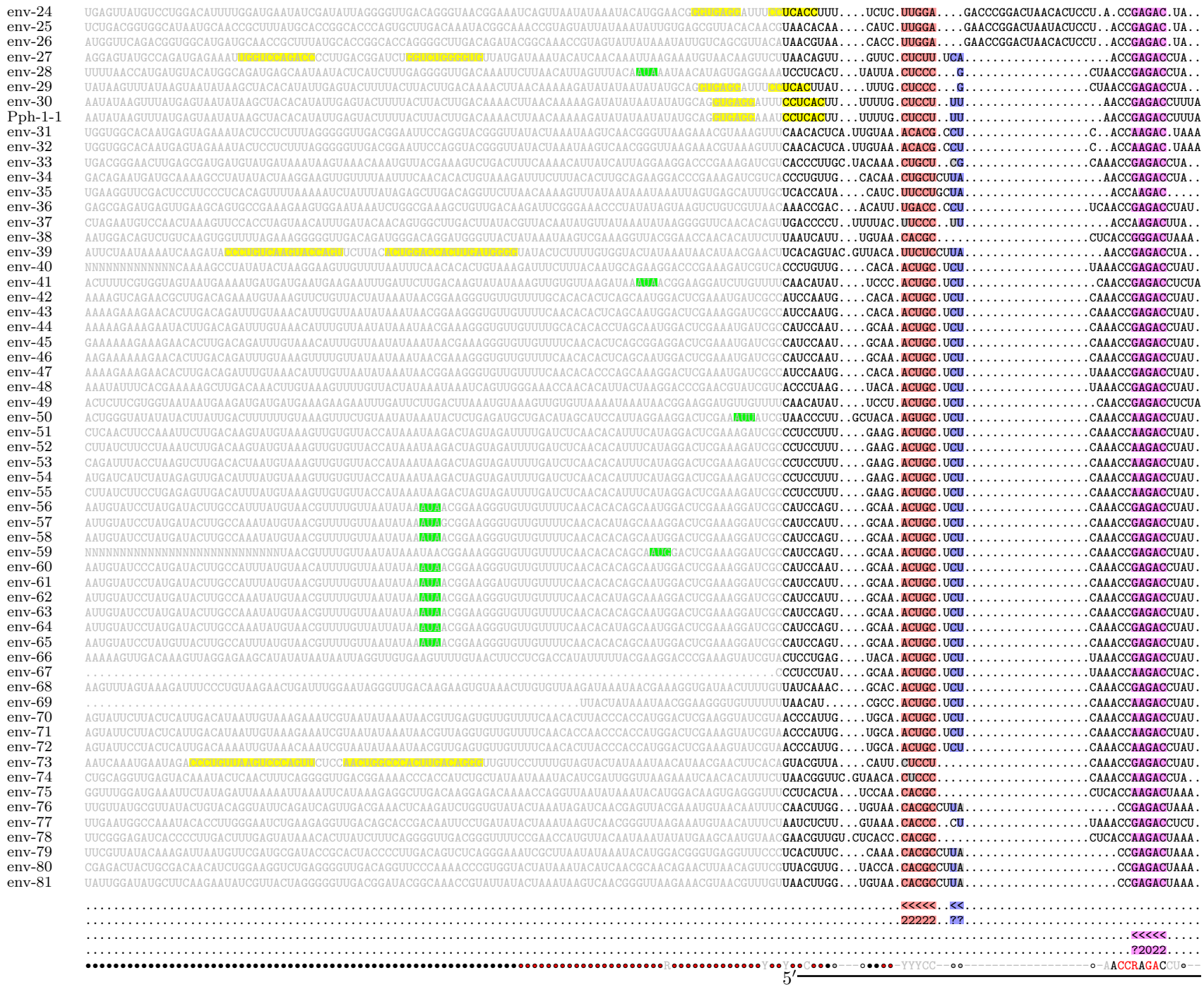
pfam00124 (297) Photosynthetic reaction centre protein.
pfam02945 (1) Recombination endonuclease VII.

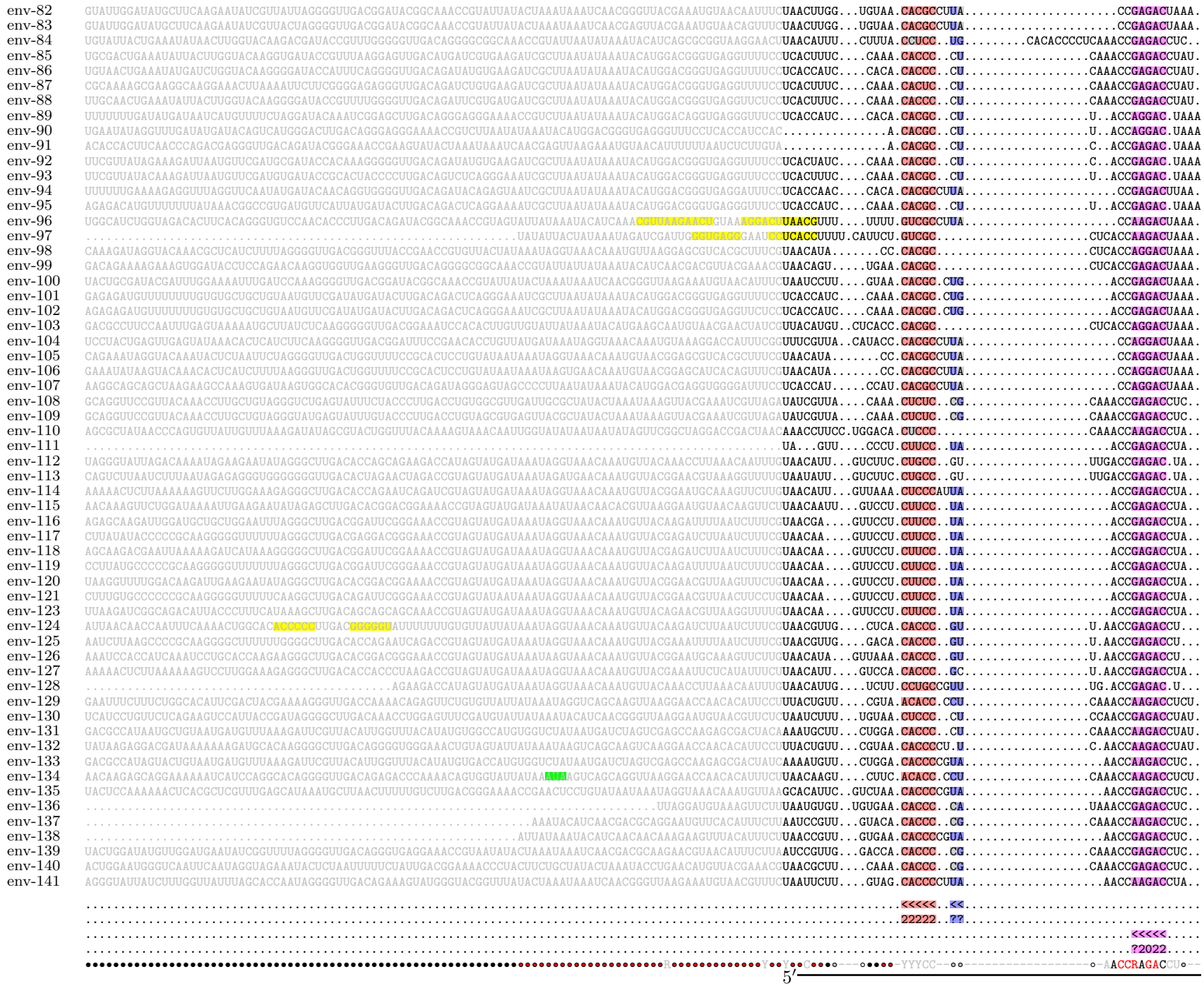
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: env-122, env-149, env-283, env-286, env-299, env-393, env-448

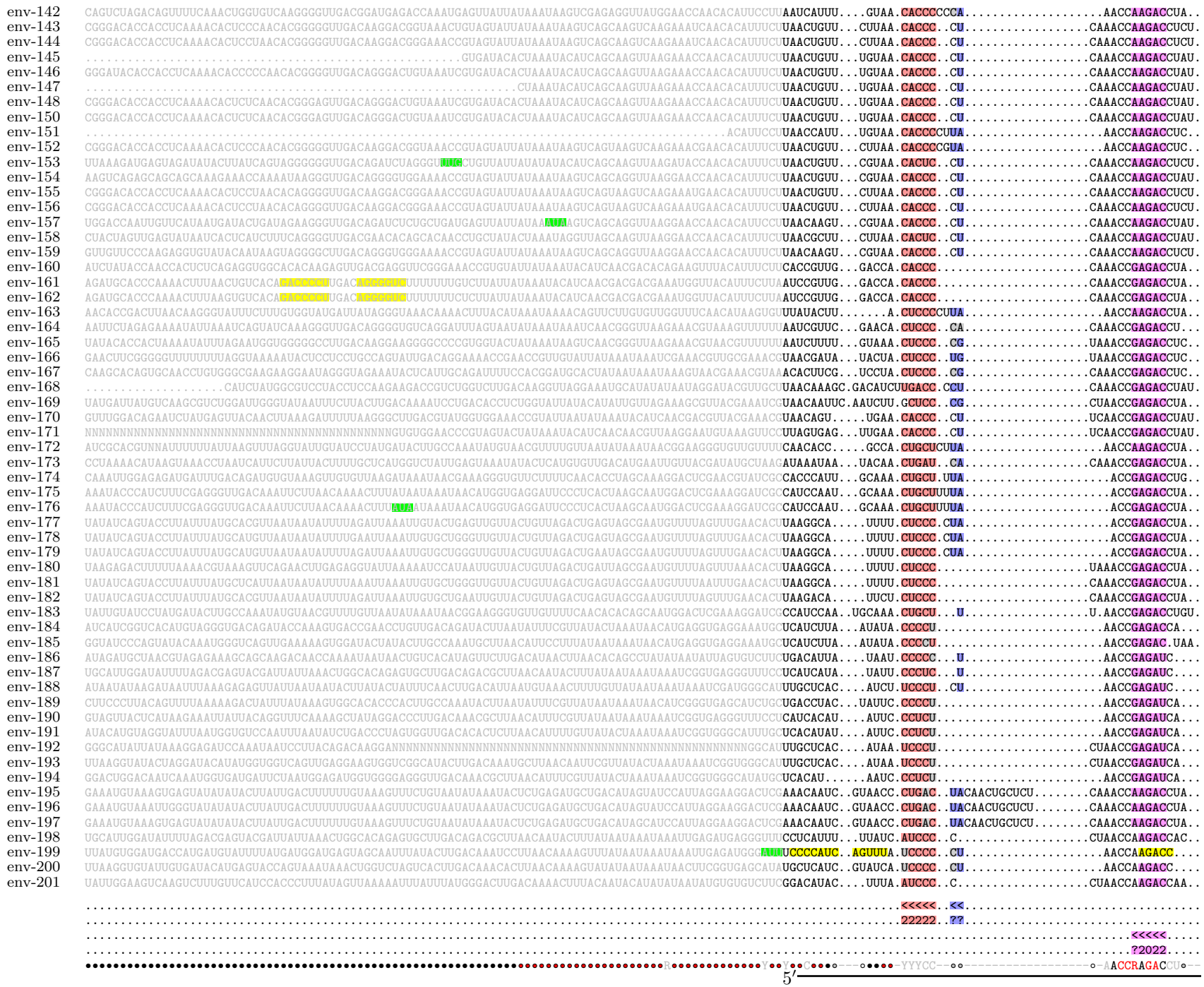
66.4 Multiple-sequence alignment

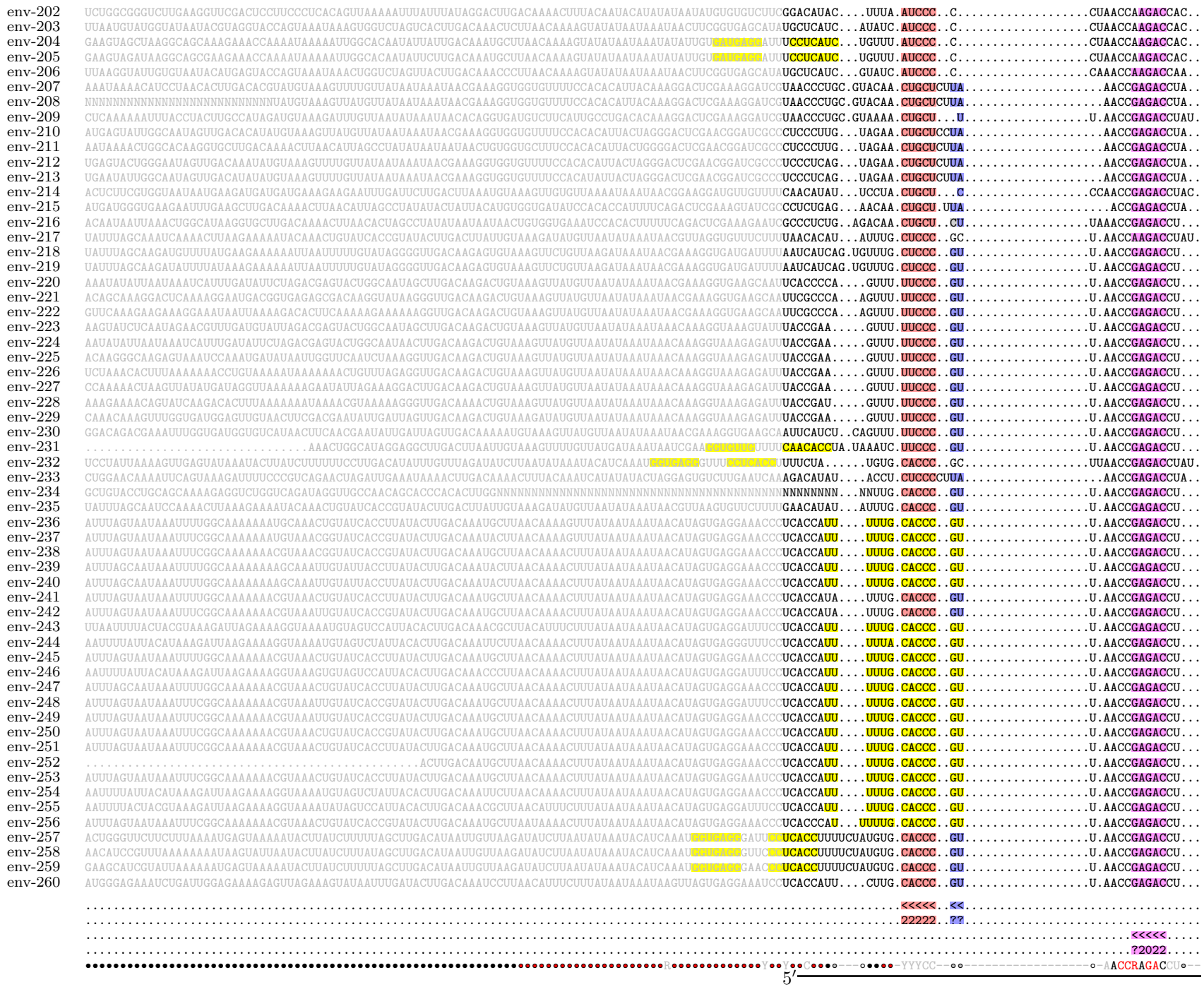
Each PhotoRC-II RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 66.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The PhotoRC-II RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

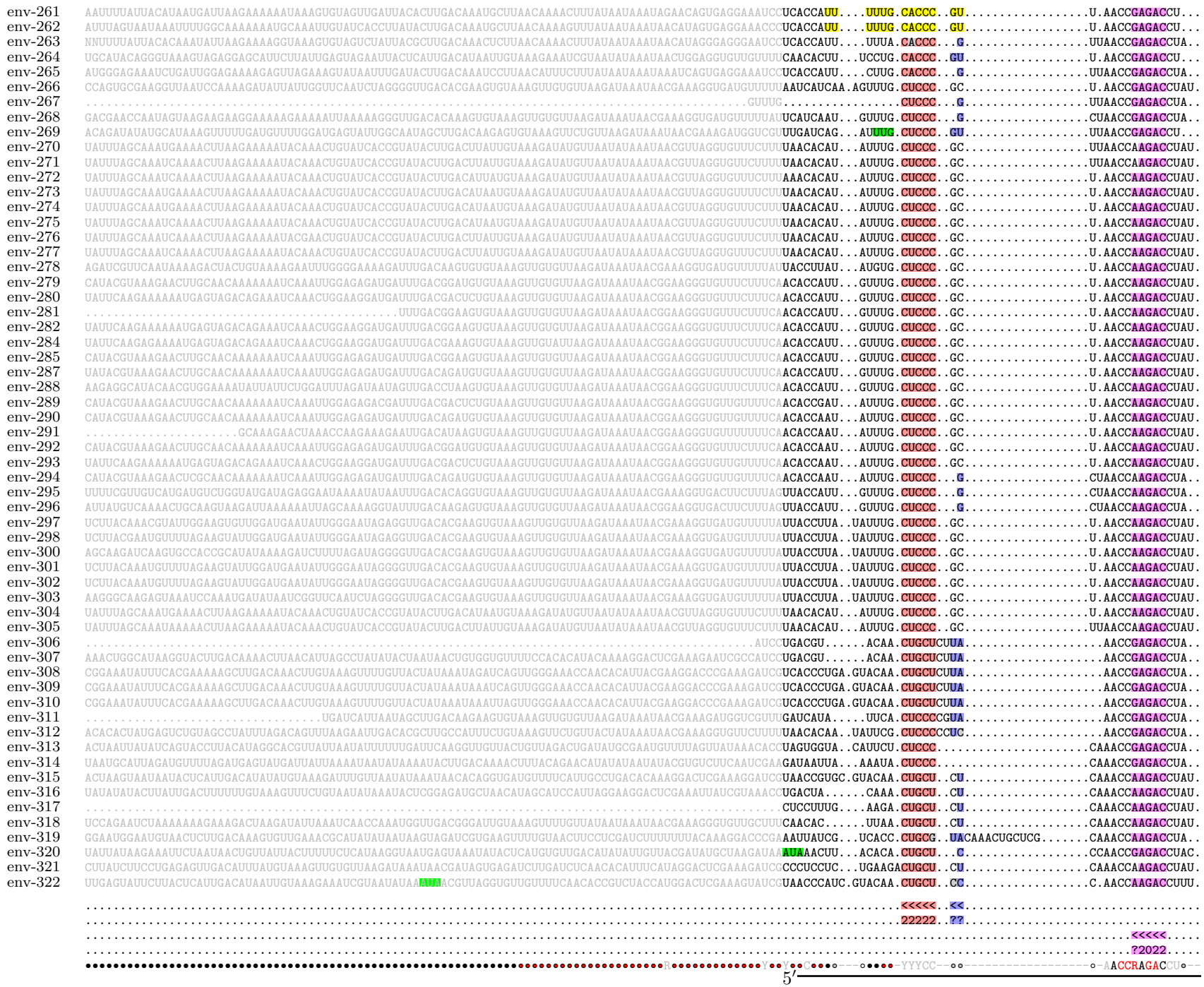


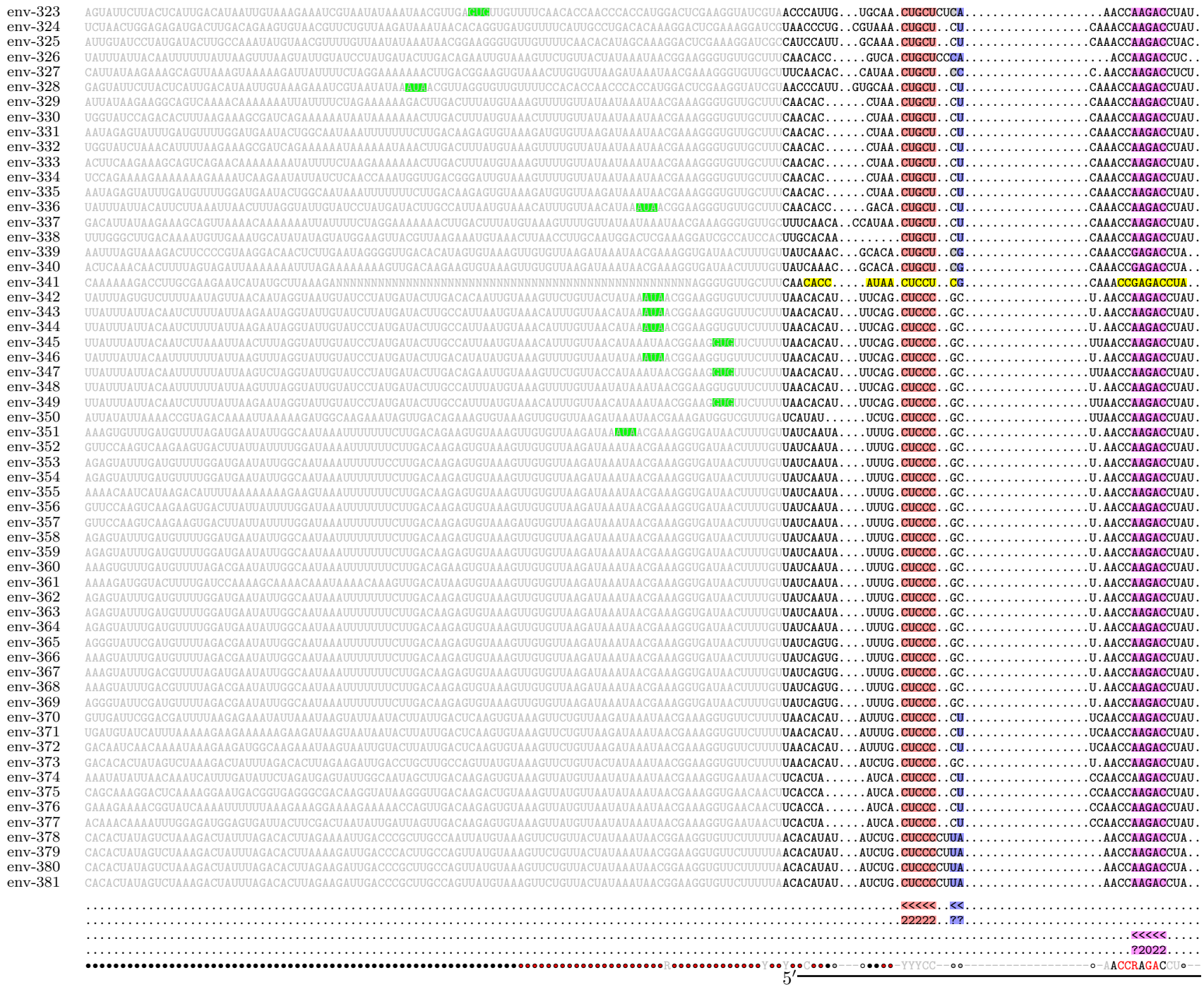

env-24
env-25
env-26
env-27
env-28
env-29
env-30
Pph-1-1
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
5′

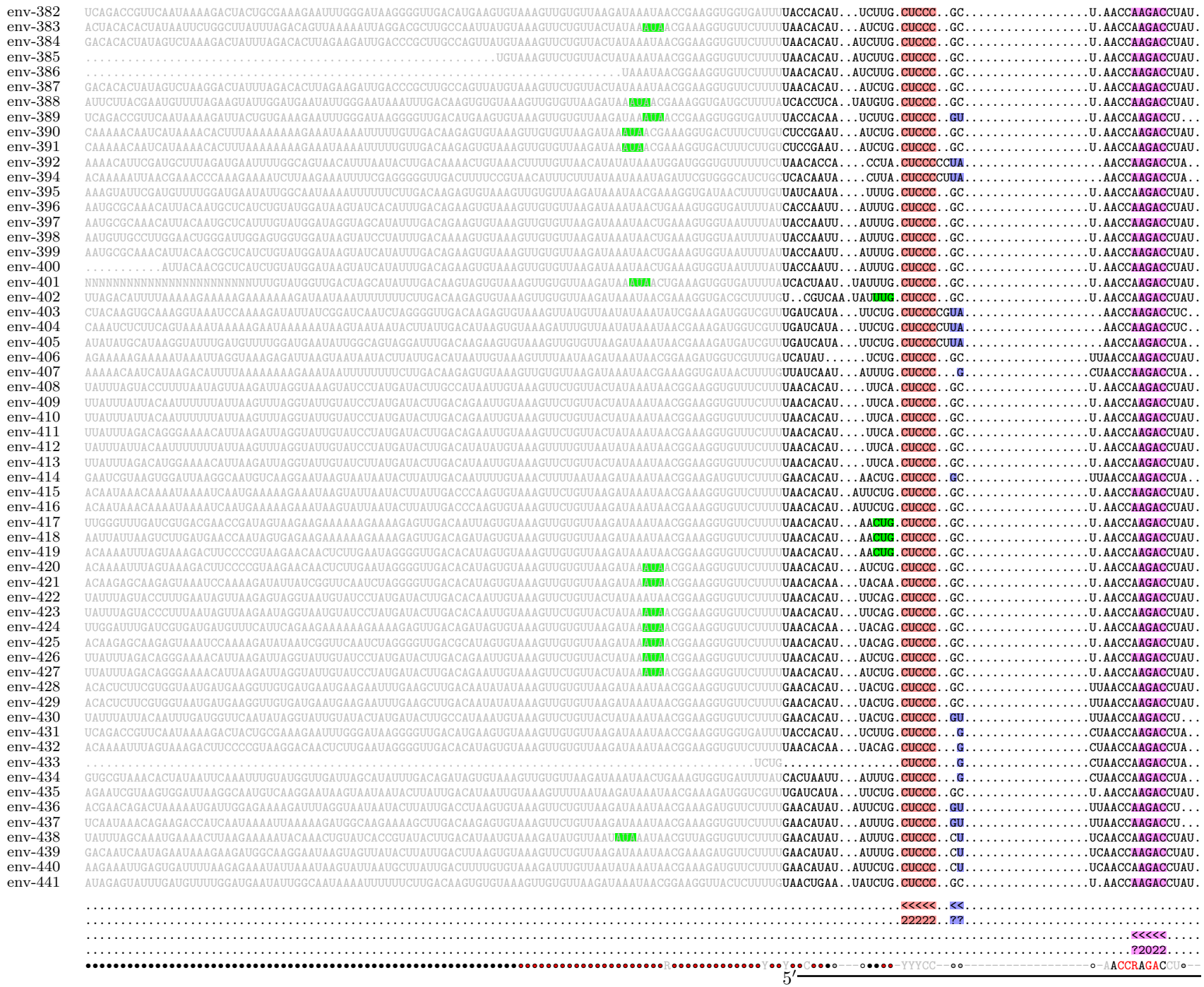

env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-135
env-136
env-137
env-138
env-139
env-140
env-141
5′

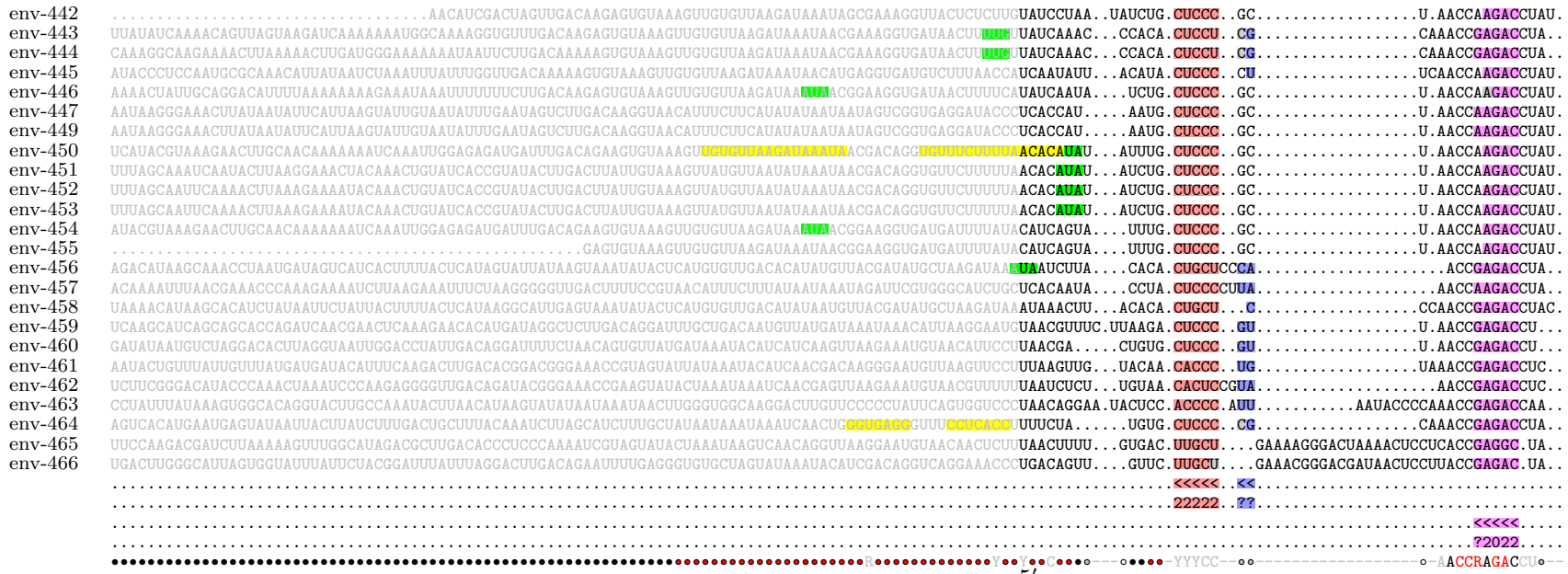

env-142
env-143
env-144
env-145
env-146
env-147
env-148
env-150
env-151
env-152
env-153
env-154
env-155
env-156
env-157
env-158
env-159
env-160
env-161
env-162
env-163
env-164
env-165
env-166
env-167
env-168
env-169
env-170
env-171
env-172
env-173
env-174
env-175
env-176
env-177
env-178
env-179
env-180
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
env-190
env-191
env-192
env-193
env-194
env-195
env-196
env-197
env-198
env-199
env-200
env-201
5′


env-202
env-203
env-204
env-205
env-206
env-207
env-208
env-209
env-210
env-211
env-212
env-213
env-214
env-215
env-216
env-217
env-218
env-219
env-220
env-221
env-222
env-223
env-224
env-225
env-226
env-227
env-228
env-229
env-230
env-231
env-232
env-233
env-234
env-235
env-236
env-237
env-238
env-239
env-240
env-241
env-242
env-243
env-244
env-245
env-246
env-247
env-248
env-249
env-250
env-251
env-252
env-253
env-254
env-255
env-256
env-257
env-258
env-259
env-260
5′

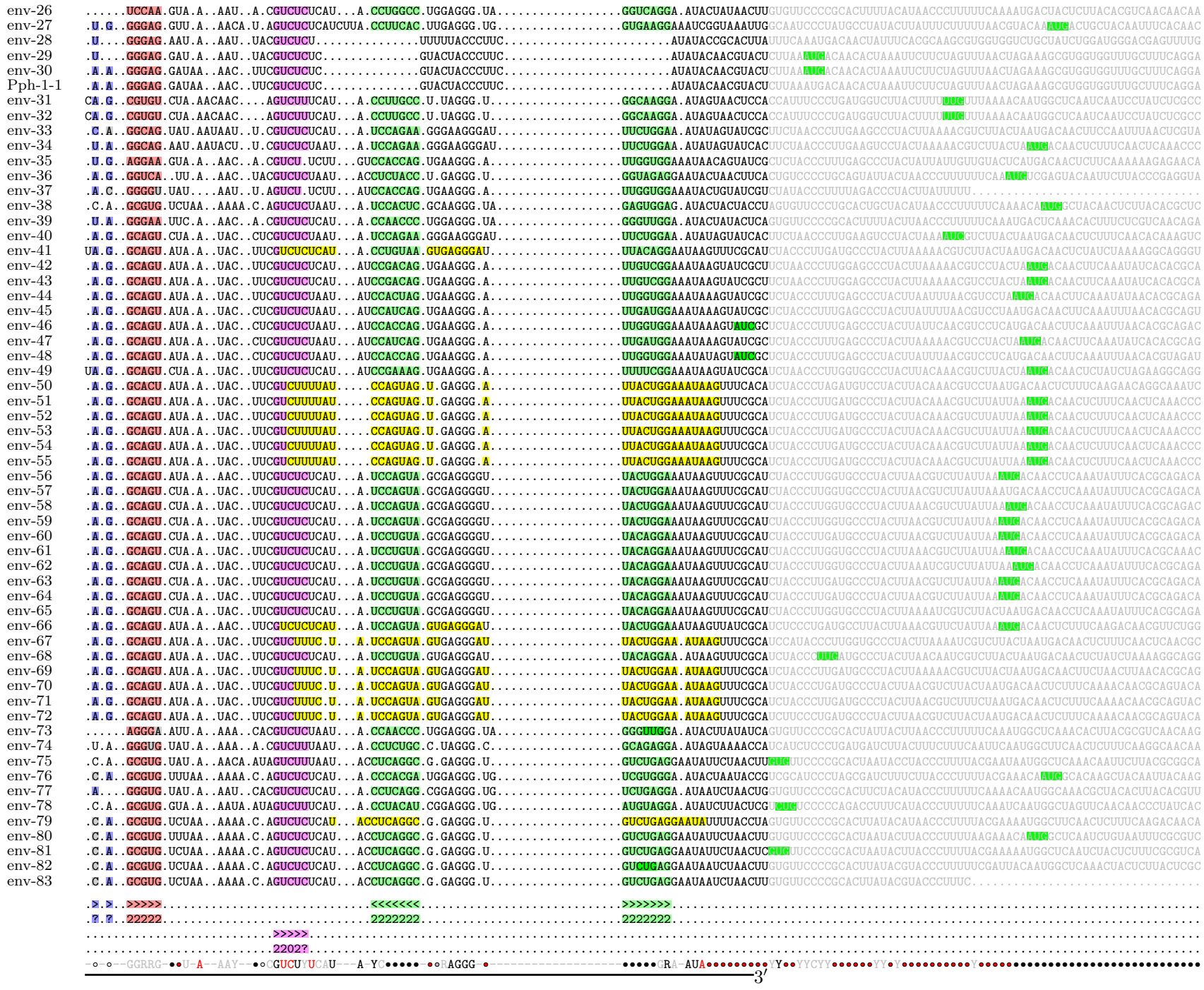

env-261
env-262
env-263
env-264
env-265
env-266
env-267
env-268
env-269
env-270
env-271
env-272
env-273
env-274
env-275
env-276
env-277
env-278
env-279
env-280
env-281
env-282
env-284
env-285
env-287
env-288
env-289
env-290
env-291
env-292
env-293
env-294
env-295
env-296
env-297
env-298
env-300
env-301
env-302
env-303
env-304
env-305
env-306
env-307
env-308
env-309
env-310
env-311
env-312
env-313
env-314
env-315
env-316
env-317
env-318
env-319
env-320
env-321
env-322
5′

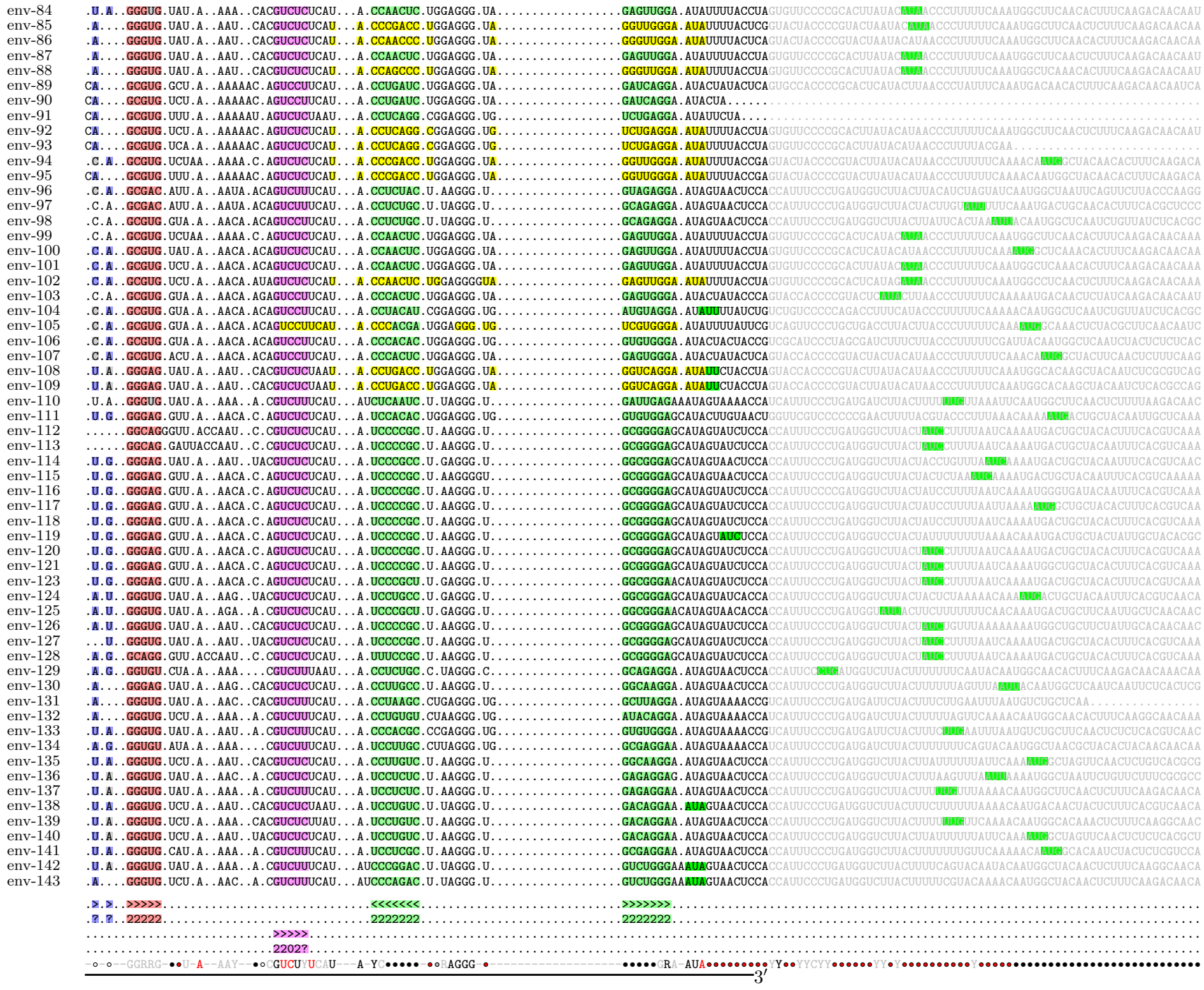

env-323
env-324
env-325
env-326
env-327
env-328
env-329
env-330
env-331
env-332
env-333
env-334
env-335
env-336
env-337
env-338
env-339
env-340
env-341
env-342
env-343
env-344
env-345
env-346
env-347
env-348
env-349
env-350
env-351
env-352
env-353
env-354
env-355
env-356
env-357
env-358
env-359
env-360
env-361
env-362
env-363
env-364
env-365
env-366
env-367
env-368
env-369
env-370
env-371
env-372
env-373
env-374
env-375
env-376
env-377
env-378
env-379
env-380
env-381
5′

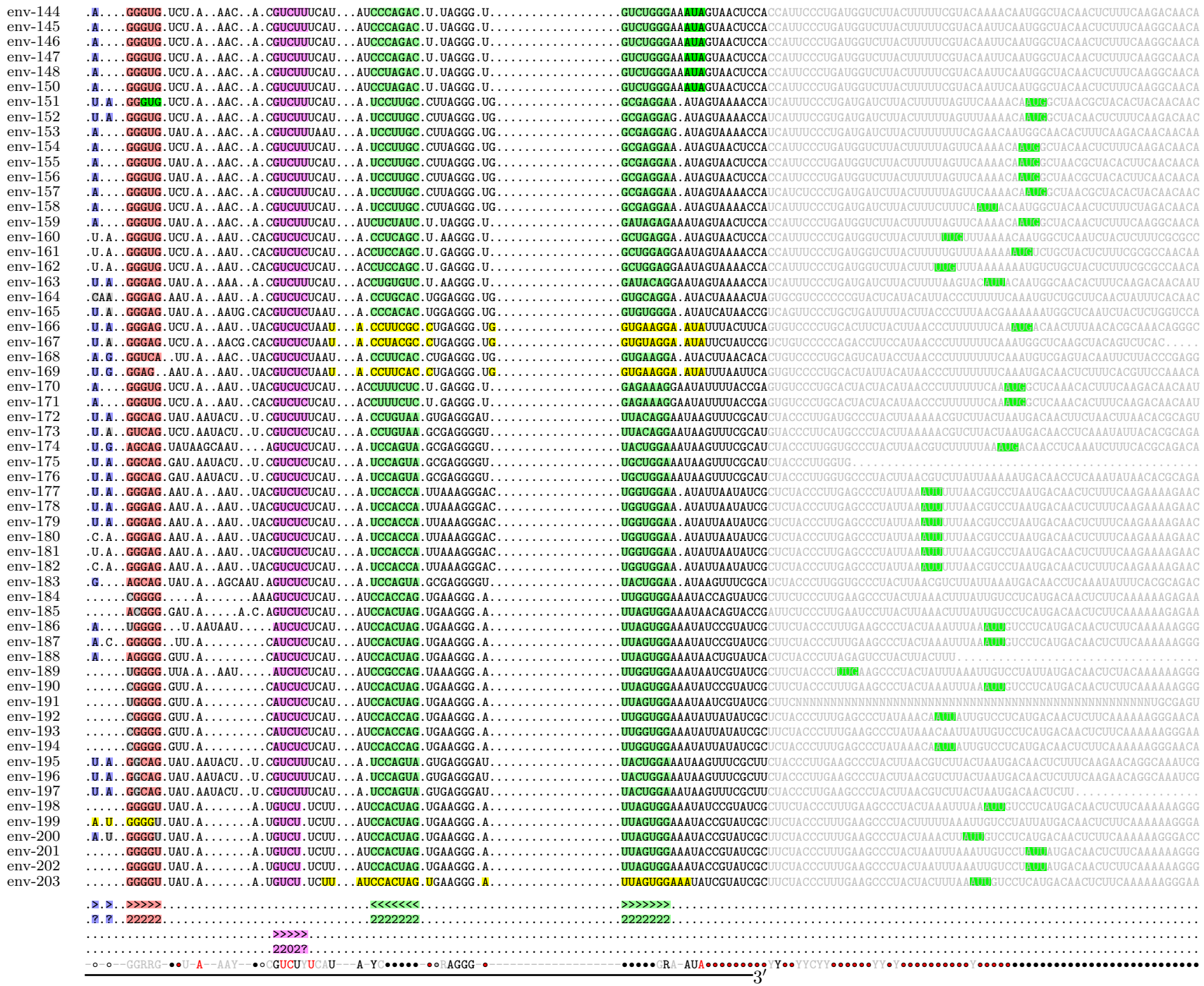

env-382
env-383
env-384
env-385
env-386
env-387
env-388
env-389
env-390
env-391
env-392
env-394
env-395
env-396
env-397
env-398
env-399
env-400
env-401
env-402
env-403
env-404
env-405
env-406
env-407
env-408
env-409
env-410
env-411
env-412
env-413
env-414
env-415
env-416
env-417
env-418
env-419
env-420
env-421
env-422
env-423
env-424
env-425
env-426
env-427
env-428
env-429
env-430
env-431
env-432
env-433
env-434
env-435
env-436
env-437
env-438
env-439
env-440
env-441
5′

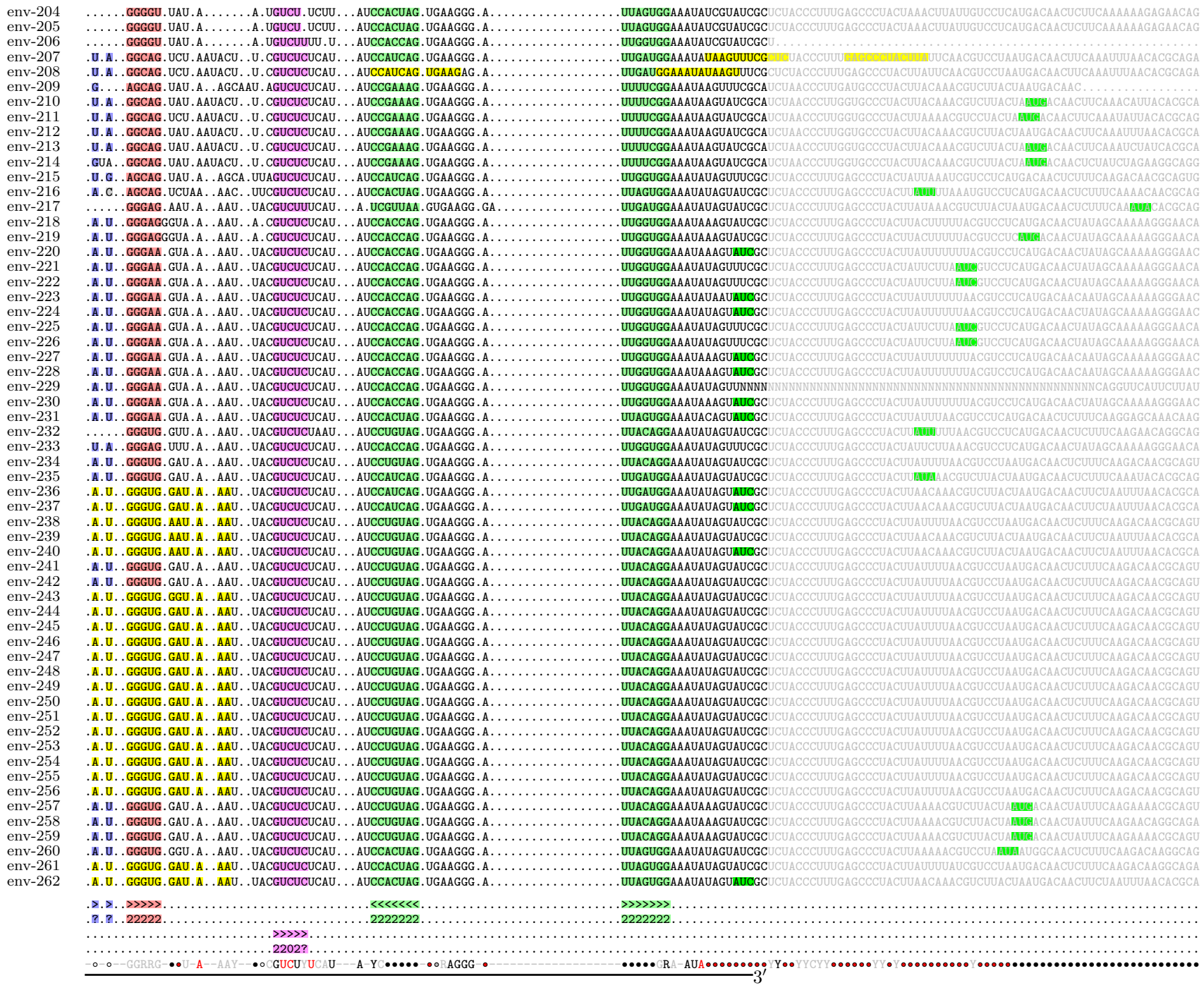


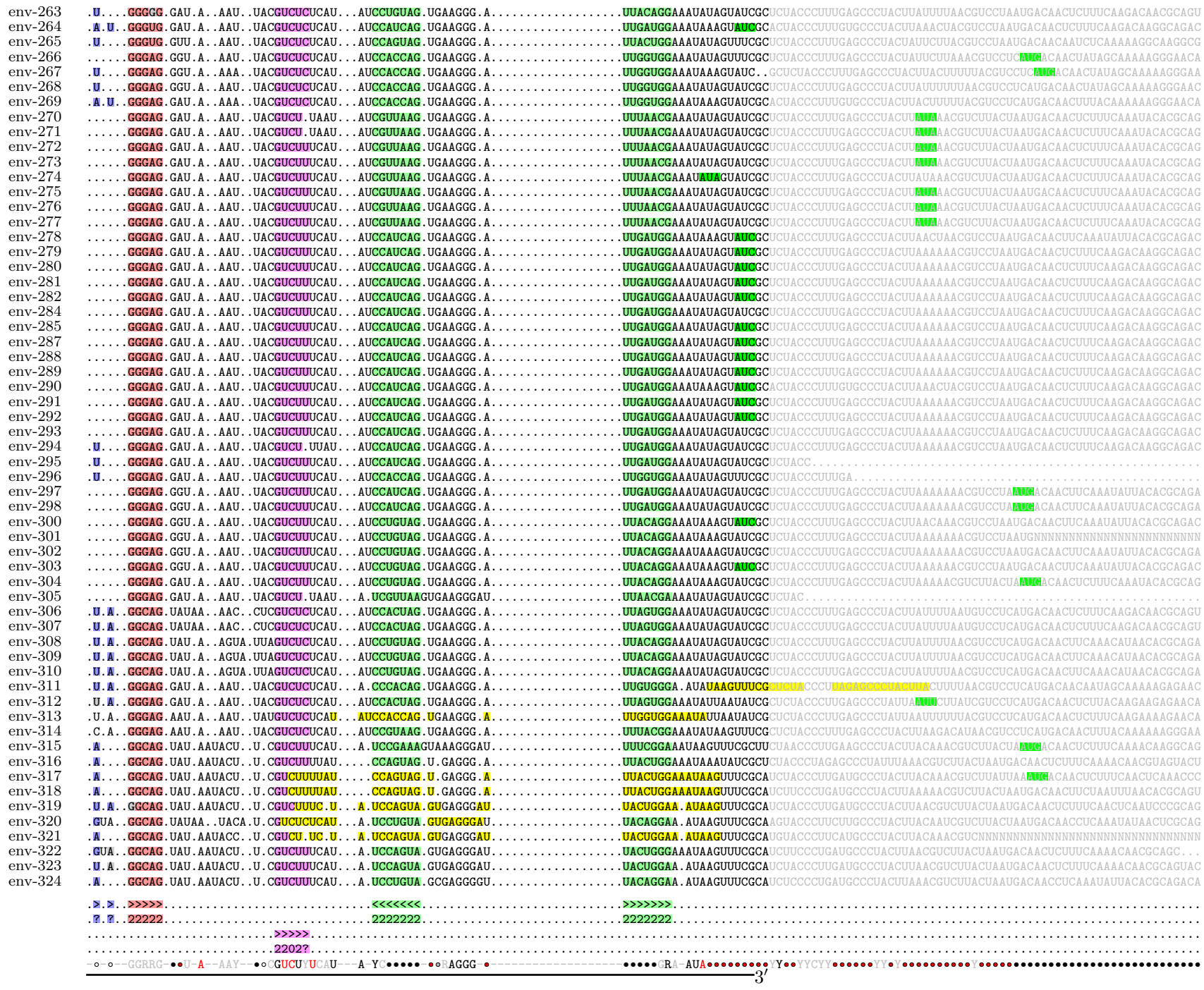
alignment positions 161 · · · 320

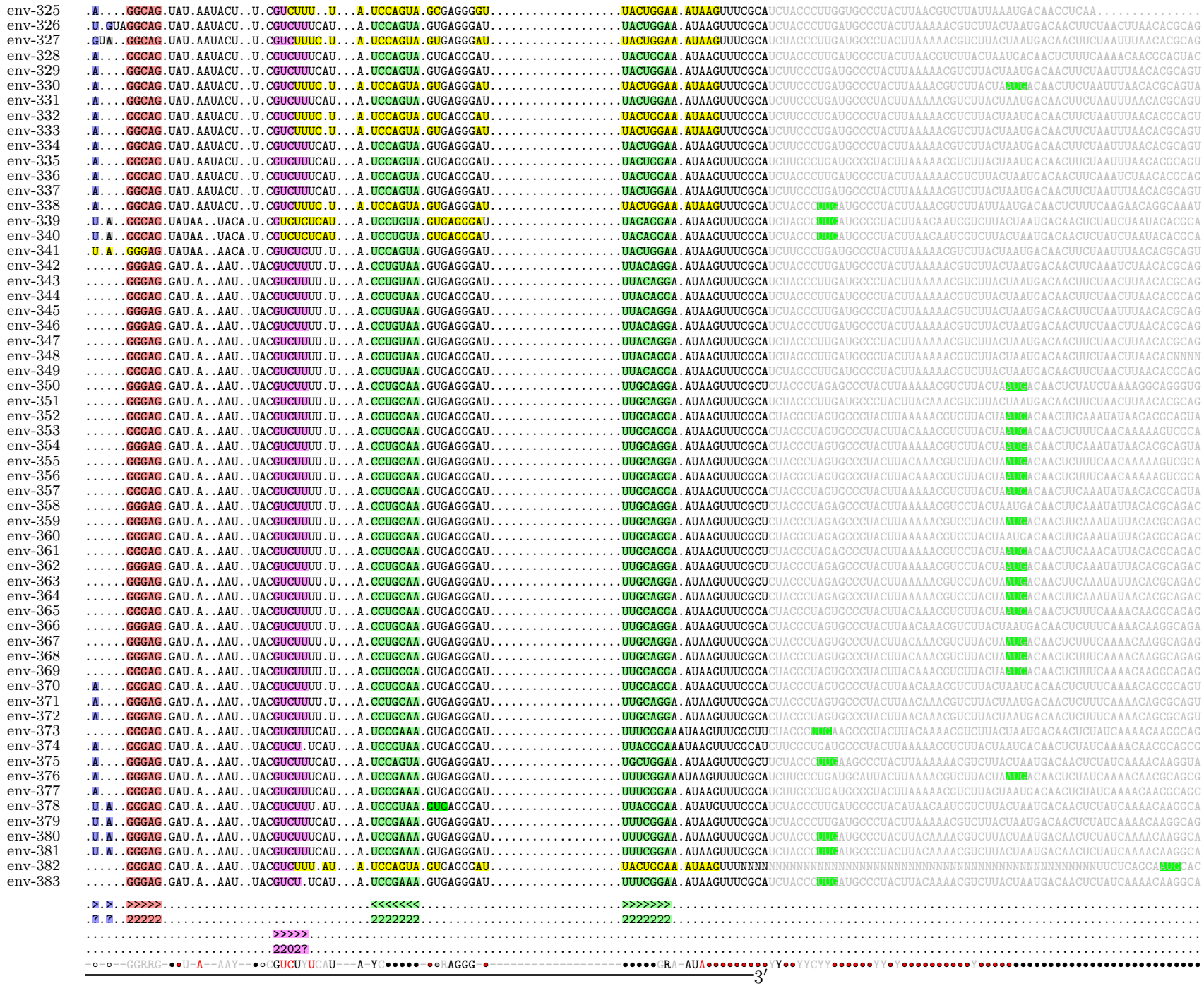

env-26
env-27
env-28
env-29
env-30
Pph-1-1
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
3′

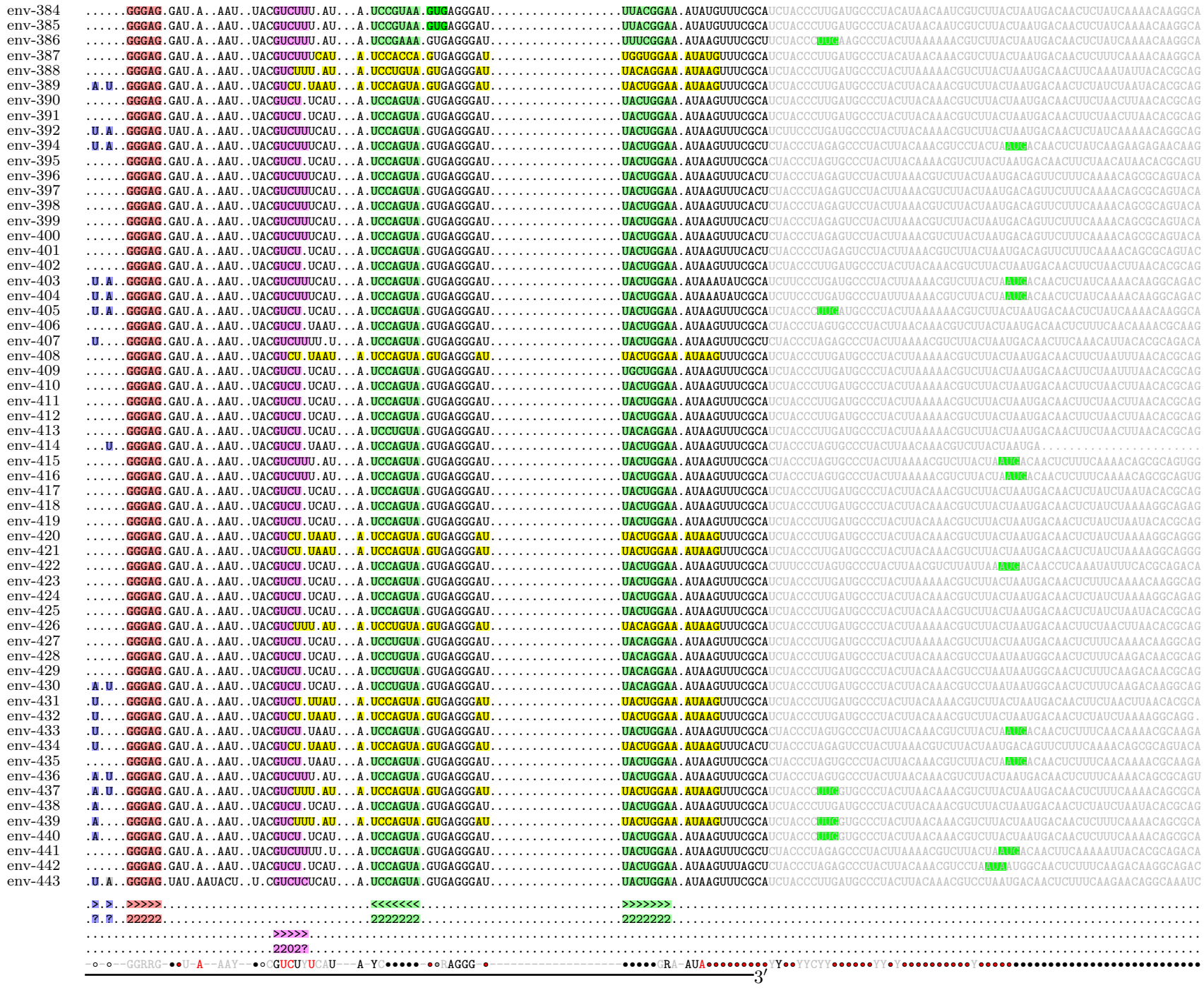

env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-135
env-136
env-137
env-138
env-139
env-140
env-141
env-142
env-143
3′

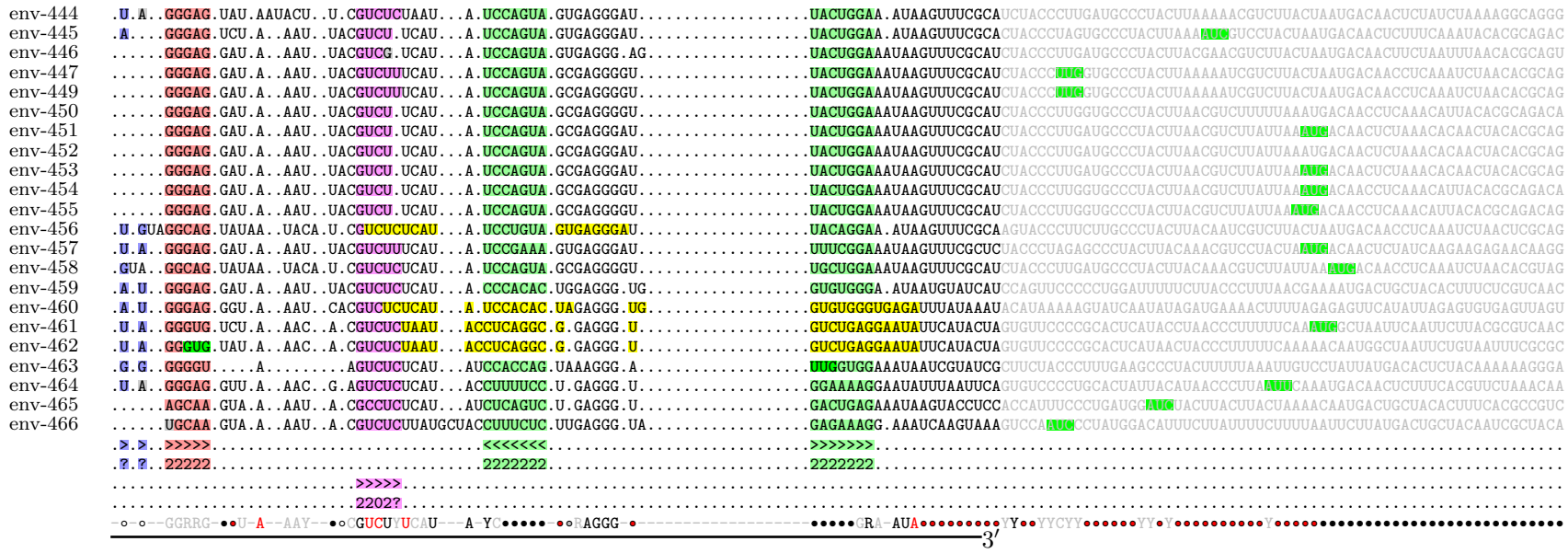

env-144
env-145
env-146
env-147
env-148
env-150
env-151
env-152
env-153
env-154
env-155
env-156
env-157
env-158
env-159
env-160
env-161
env-162
env-163
env-164
env-165
env-166
env-167
env-168
env-169
env-170
env-171
env-172
env-173
env-174
env-175
env-176
env-177
env-178
env-179
env-180
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
env-190
env-191
env-192
env-193
env-194
env-195
env-196
env-197
env-198
env-199
env-200
env-201
env-202
env-203
3′


env-204
env-205
env-206
env-207
env-208
env-209
env-210
env-211
env-212
env-213
env-214
env-215
env-216
env-217
env-218
env-219
env-220
env-221
env-222
env-223
env-224
env-225
env-226
env-227
env-228
env-229
env-230
env-231
env-232
env-233
env-234
env-235
env-236
env-237
env-238
env-239
env-240
env-241
env-242
env-243
env-244
env-245
env-246
env-247
env-248
env-249
env-250
env-251
env-252
env-253
env-254
env-255
env-256
env-257
env-258
env-259
env-260
env-261
env-262
3′


env-263
env-264
env-265
env-266
env-267
env-268
env-269
env-270
env-271
env-272
env-273
env-274
env-275
env-276
env-277
env-278
env-279
env-280
env-281
env-282
env-284
env-285
env-287
env-288
env-289
env-290
env-291
env-292
env-293
env-294
env-295
env-296
env-297
env-298
env-300
env-301
env-302
env-303
env-304
env-305
env-306
env-307
env-308
env-309
env-310
env-311
env-312
env-313
env-314
env-315
env-316
env-317
env-318
env-319
env-320
env-321
env-322
env-323
env-324
3′

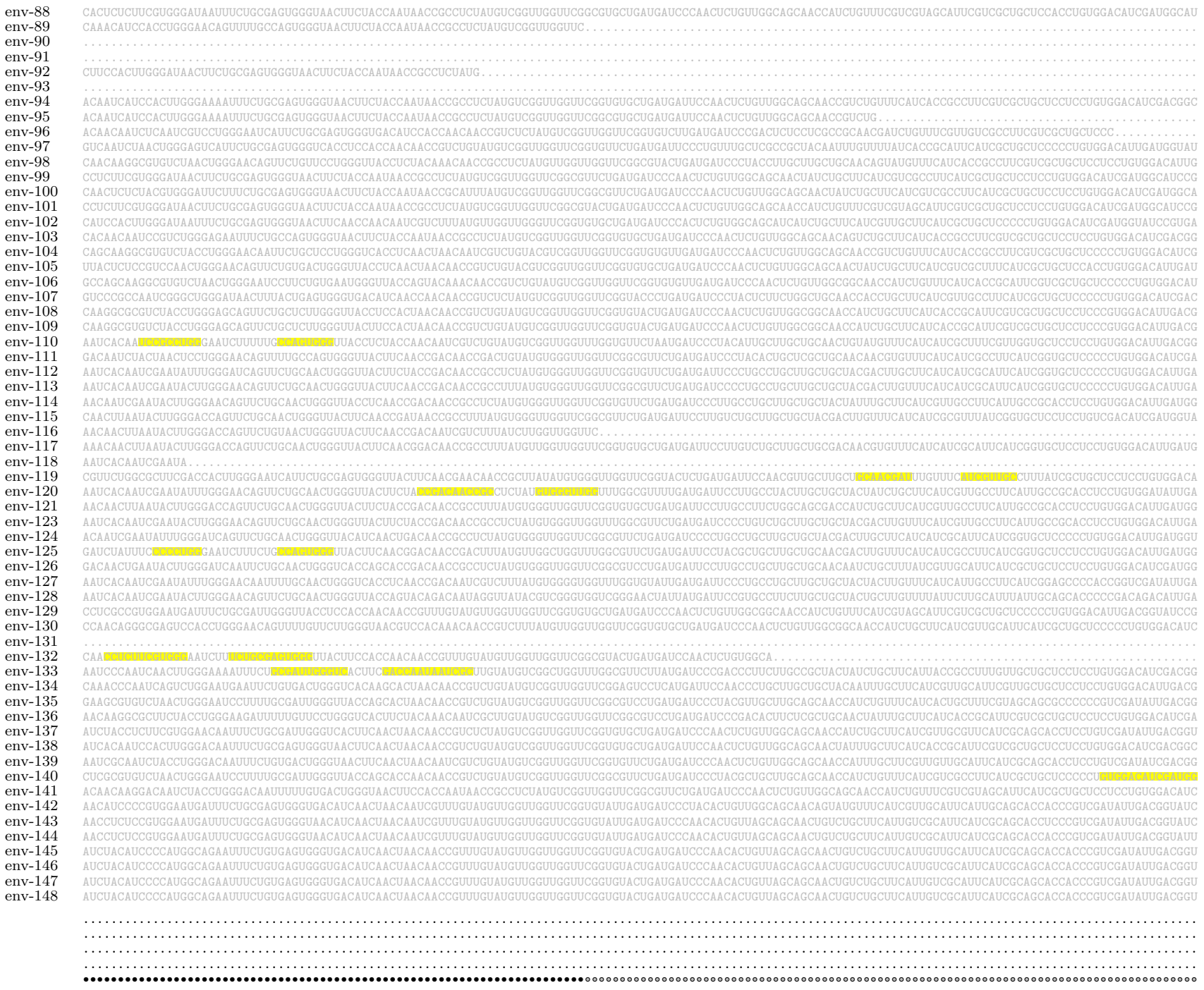

env-325
env-326
env-327
env-328
env-329
env-330
env-331
env-332
env-333
env-334
env-335
env-336
env-337
env-338
env-339
env-340
env-341
env-342
env-343
env-344
env-345
env-346
env-347
env-348
env-349
env-350
env-351
env-352
env-353
env-354
env-355
env-356
env-357
env-358
env-359
env-360
env-361
env-362
env-363
env-364
env-365
env-366
env-367
env-368
env-369
env-370
env-371
env-372
env-373
env-374
env-375
env-376
env-377
env-378
env-379
env-380
env-381
env-382
env-383
3′

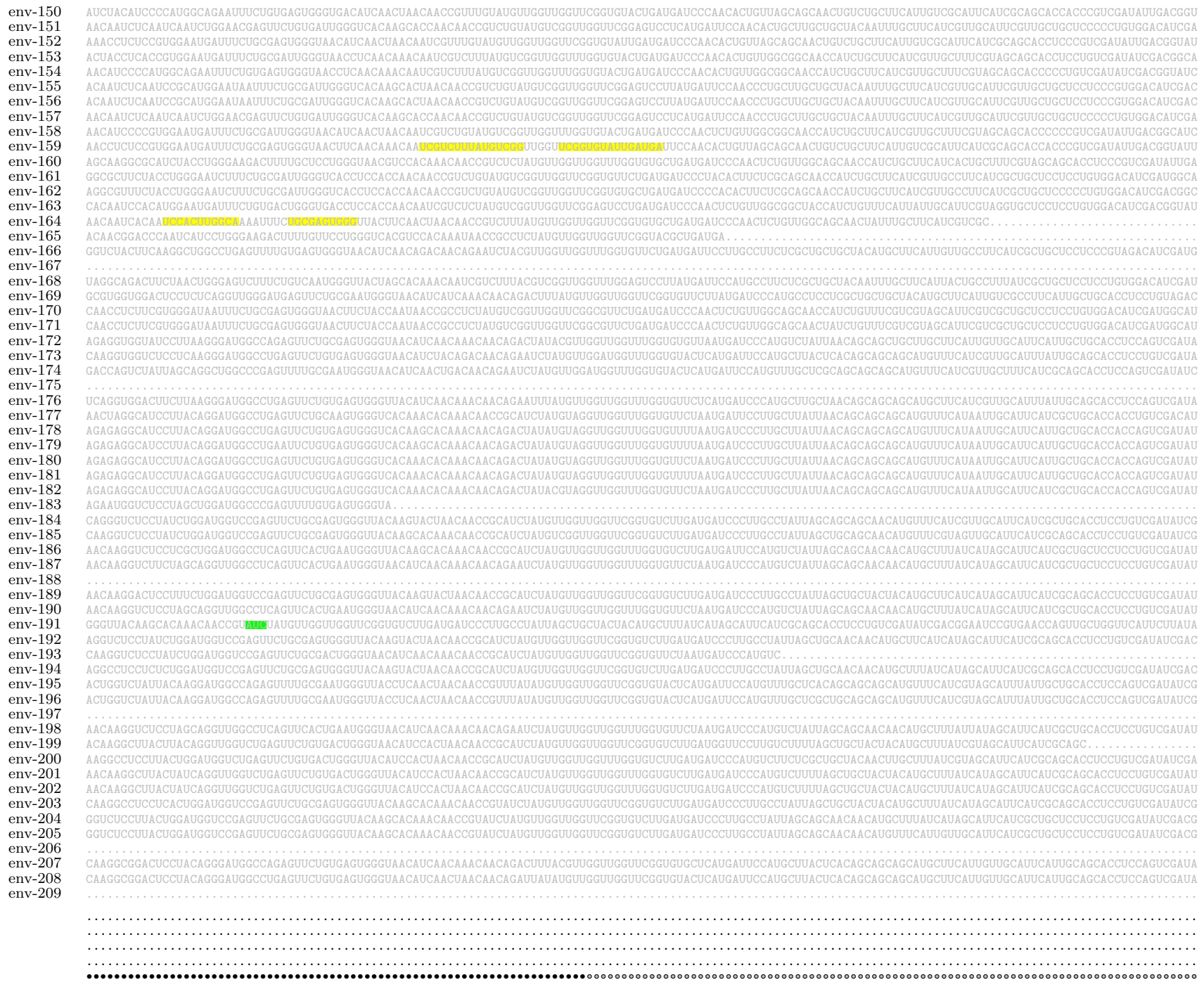

env-384
env-385
env-386
env-387
env-388
env-389
env-390
env-391
env-392
env-394
env-395
env-396
env-397
env-398
env-399
env-400
env-401
env-402
env-403
env-404
env-405
env-406
env-407
env-408
env-409
env-410
env-411
env-412
env-413
env-414
env-415
env-416
env-417
env-418
env-419
env-420
env-421
env-422
env-423
env-424
env-425
env-426
env-427
env-428
env-429
env-430
env-431
env-432
env-433
env-434
env-435
env-436
env-437
env-438
env-439
env-440
env-441
env-442
env-443
3′



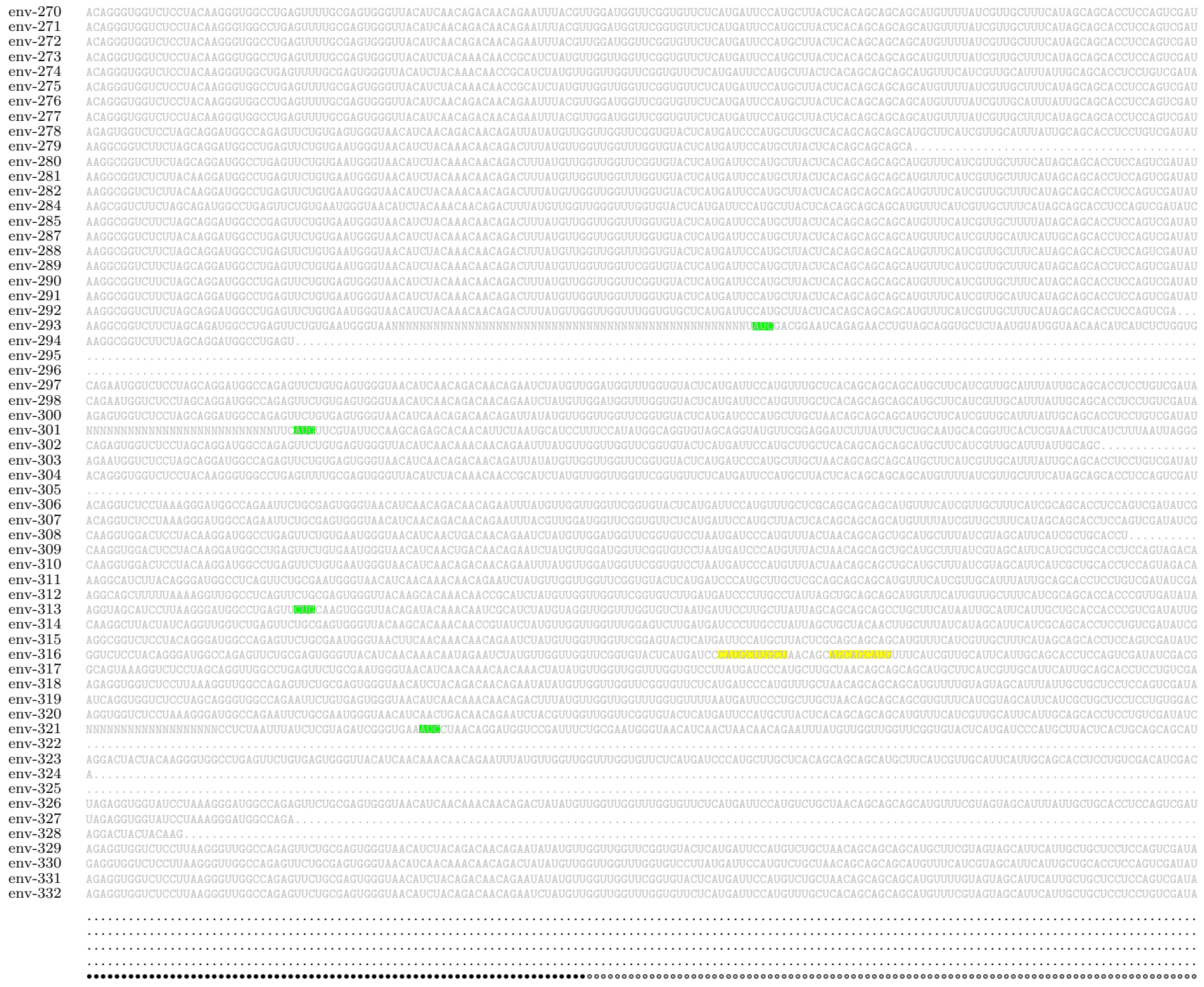
alignment positions 321 · · · 480

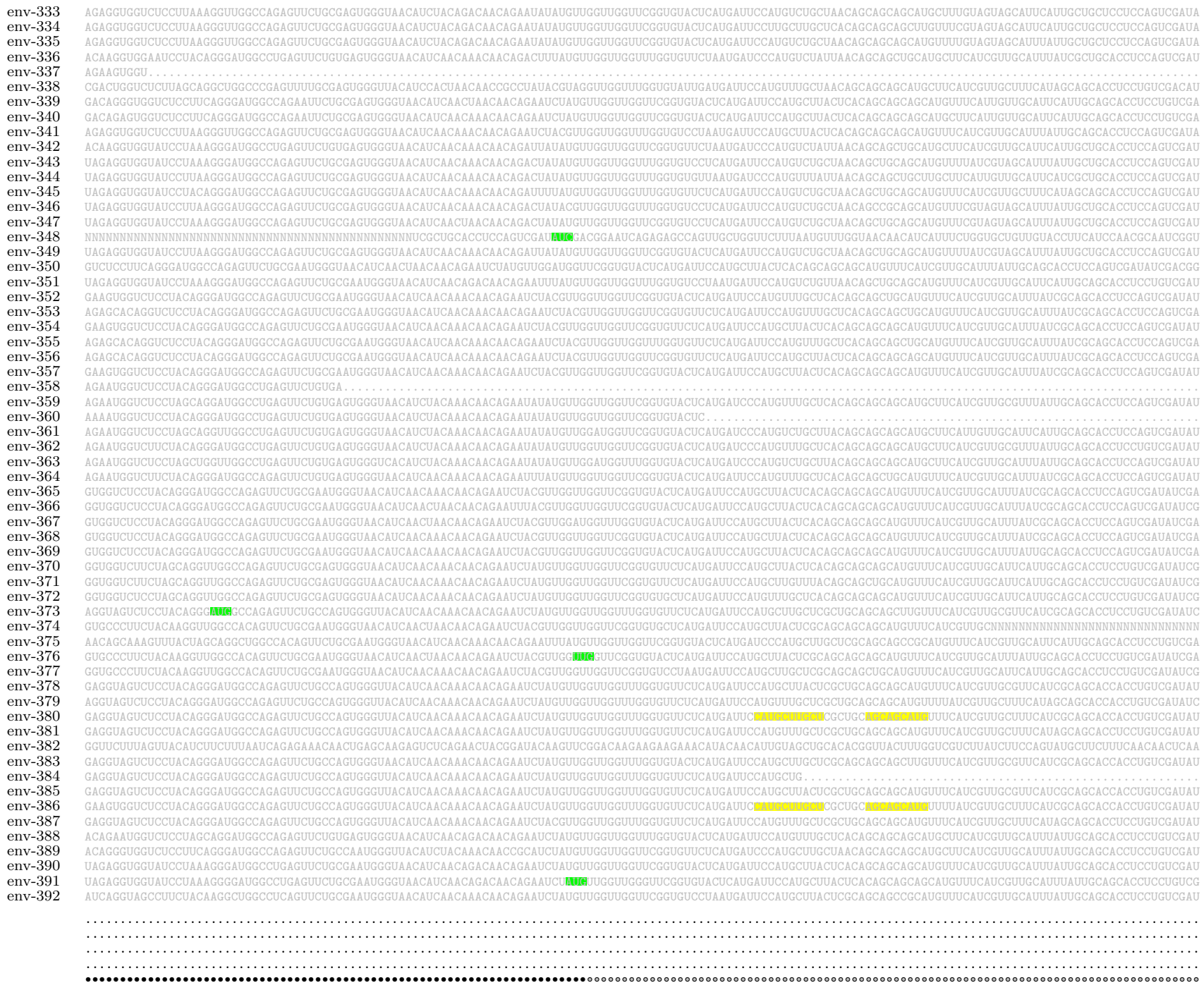
env-29
env-30
Pph-1-1
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87


env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-135
env-136
env-137
env-138
env-139
env-140
env-141
env-142
env-143
env-144
env-145
env-146
env-147
env-148

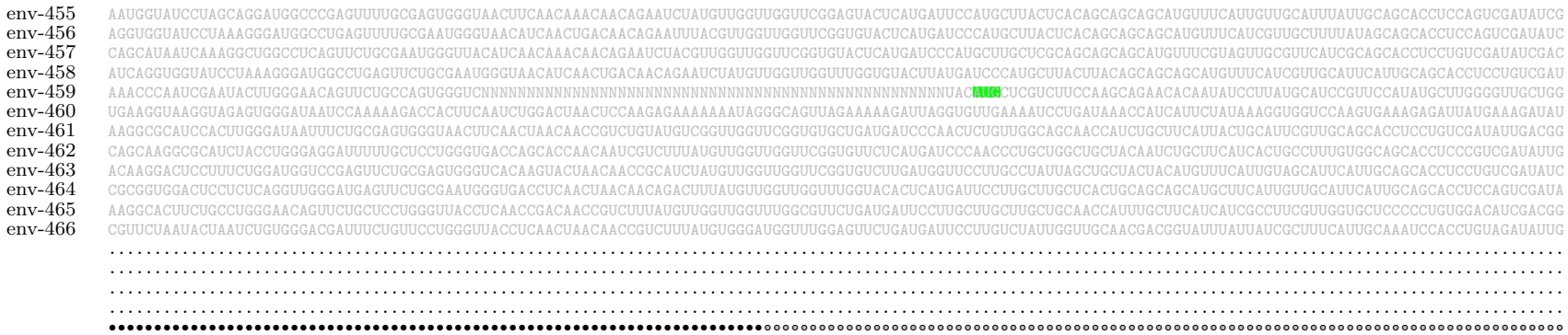

env-150
env-151
env-152
env-153
env-154
env-155
env-156
env-157
env-158
env-159
env-160
env-161
env-162
env-163
env-164
env-165
env-166
env-167
env-168
env-169
env-170
env-171
env-172
env-173
env-174
env-175
env-176
env-177
env-178
env-179
env-180
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
env-190
env-191
env-192
env-193
env-194
env-195
env-196
env-197
env-198
env-199
env-200
env-201
env-202
env-203
env-204
env-205
env-206
env-207
env-208
env-209

env-210
env-211
env-212
env-213
env-214
env-215
env-216
env-217
env-218
env-219
env-220
env-221
env-222
env-223
env-224
env-225
env-226
env-227
env-228
env-229
env-230
env-231
env-232
env-233
env-234
env-235
env-236
env-237
env-238
env-239
env-240
env-241
env-242
env-243
env-244
env-245
env-246
env-247
env-248
env-249
env-250
env-251
env-252
env-253
env-254
env-255
env-256
env-257
env-258
env-259
env-260
env-261
env-262
env-263
env-264
env-265
env-266
env-267
env-268
env-269


env-270
env-271
env-272
env-273
env-274
env-275
env-276
env-277
env-278
env-279
env-280
env-281
env-282
env-284
env-285
env-287
env-288
env-289
env-290
env-291
env-292
env-293
env-294
env-295
env-296
env-297
env-298
env-300
env-301
env-302
env-303
env-304
env-305
env-306
env-307
env-308
env-309
env-310
env-311
env-312
env-313
env-314
env-315
env-316
env-317
env-318
env-319
env-320
env-321
env-322
env-323
env-324
env-325
env-326
env-327
env-328
env-329
env-330
env-331
env-332

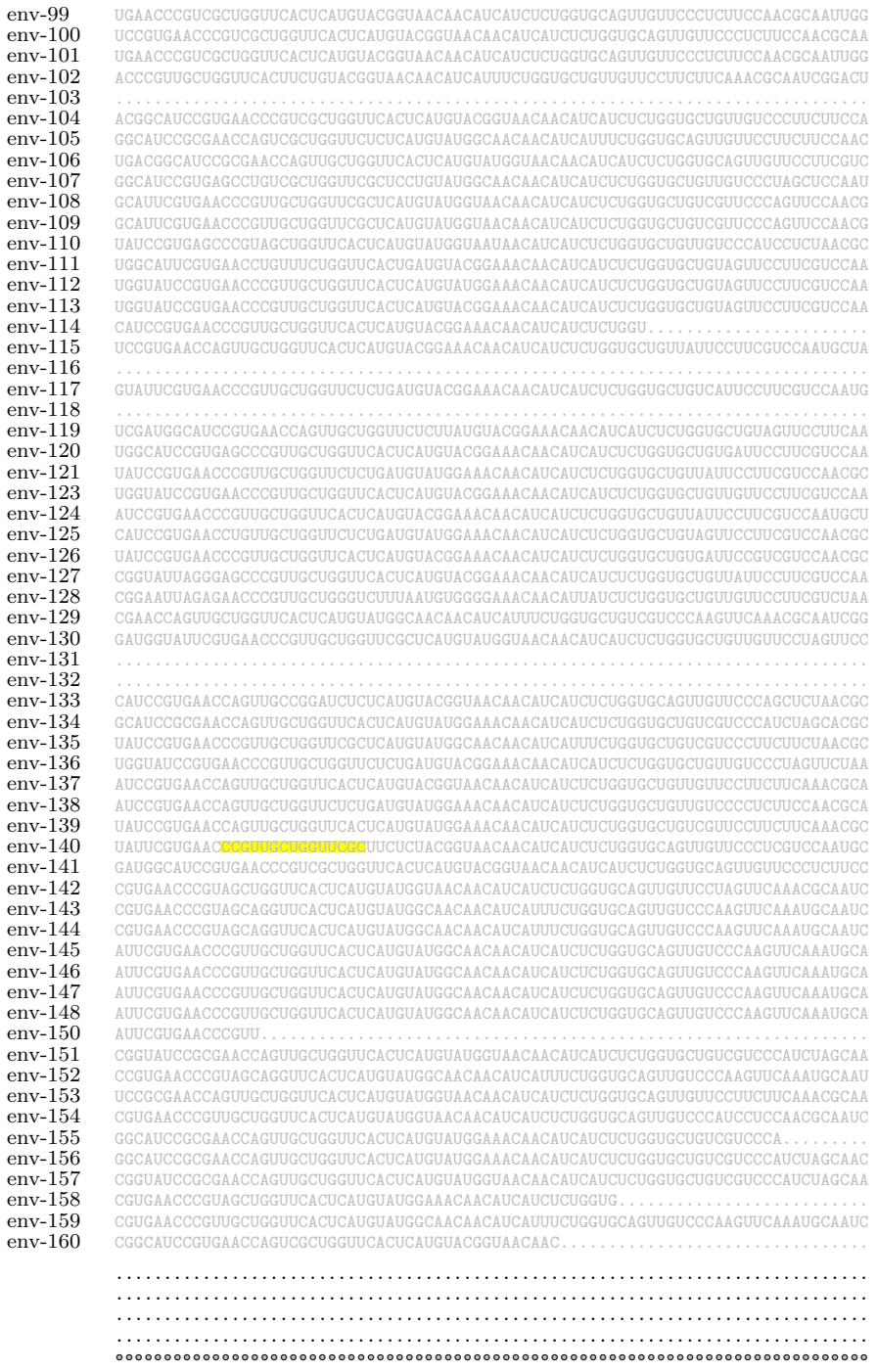

env-333
env-334
env-335
env-336
env-337
env-338
env-339
env-340
env-341
env-342
env-343
env-344
env-345
env-346
env-347
env-348
env-349
env-350
env-351
env-352
env-353
env-354
env-355
env-356
env-357
env-358
env-359
env-360
env-361
env-362
env-363
env-364
env-365
env-366
env-367
env-368
env-369
env-370
env-371
env-372
env-373
env-374
env-375
env-376
env-377
env-378
env-379
env-380
env-381
env-382
env-383
env-384
env-385
env-386
env-387
env-388
env-389
env-390
env-391
env-392

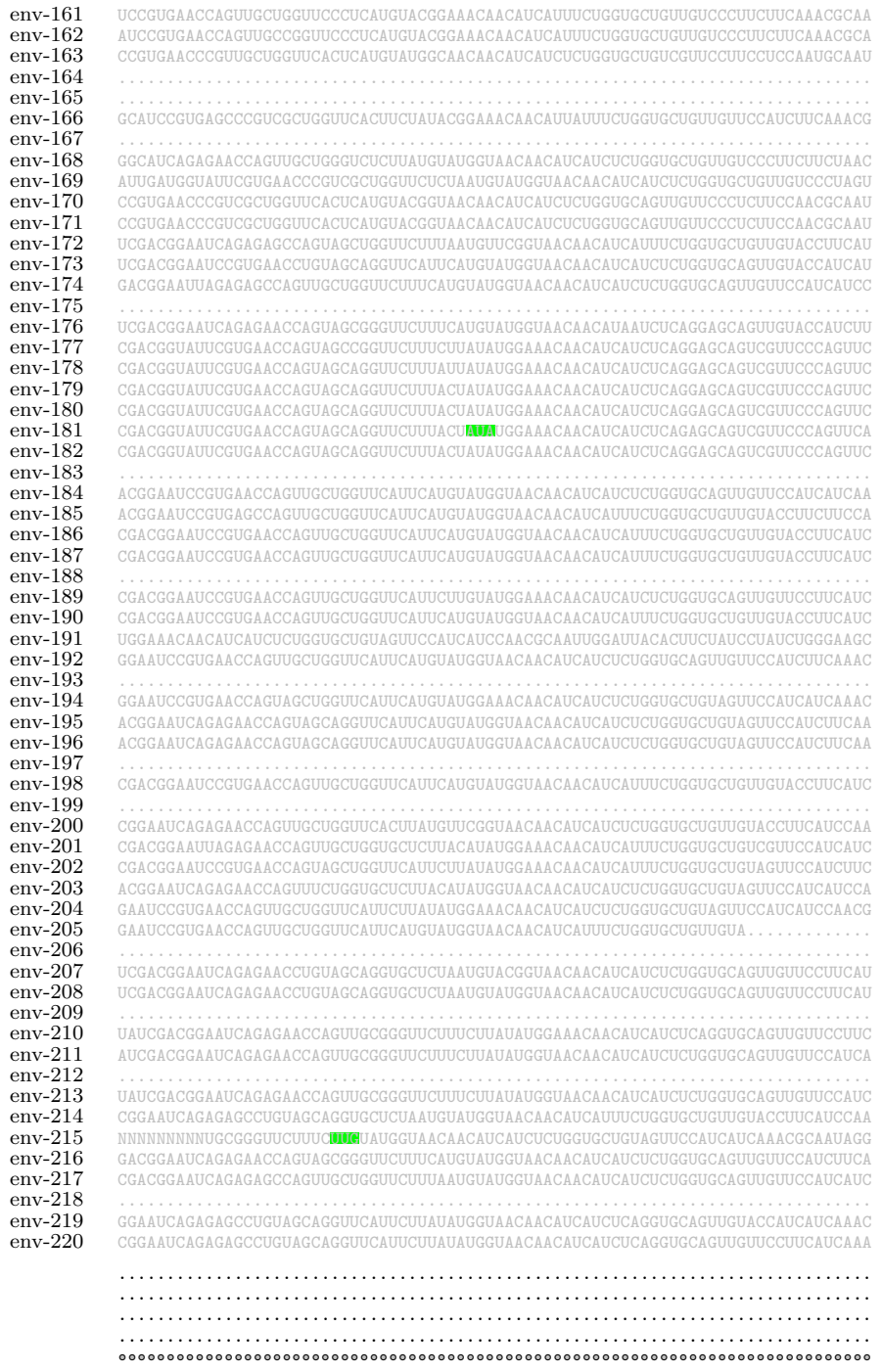
env-394
env-395
env-396
env-397
env-398
env-399
env-400
env-401
env-402
env-403
env-404
env-405
env-406
env-407
env-408
env-409
env-410
env-411
env-412
env-413
env-414
env-415
env-416
env-417
env-418
env-419
env-420
env-421
env-422
env-423
env-424
env-425
env-426
env-427
env-428
env-429
env-430
env-431
env-432
env-433
env-434
env-435
env-436
env-437
env-438
env-439
env-440
env-441
env-442
env-443
env-444
env-445
env-446
env-447
env-448
env-449
env-450
env-451
env-452
env-453
env-454



alignment positions 481 · · · 558

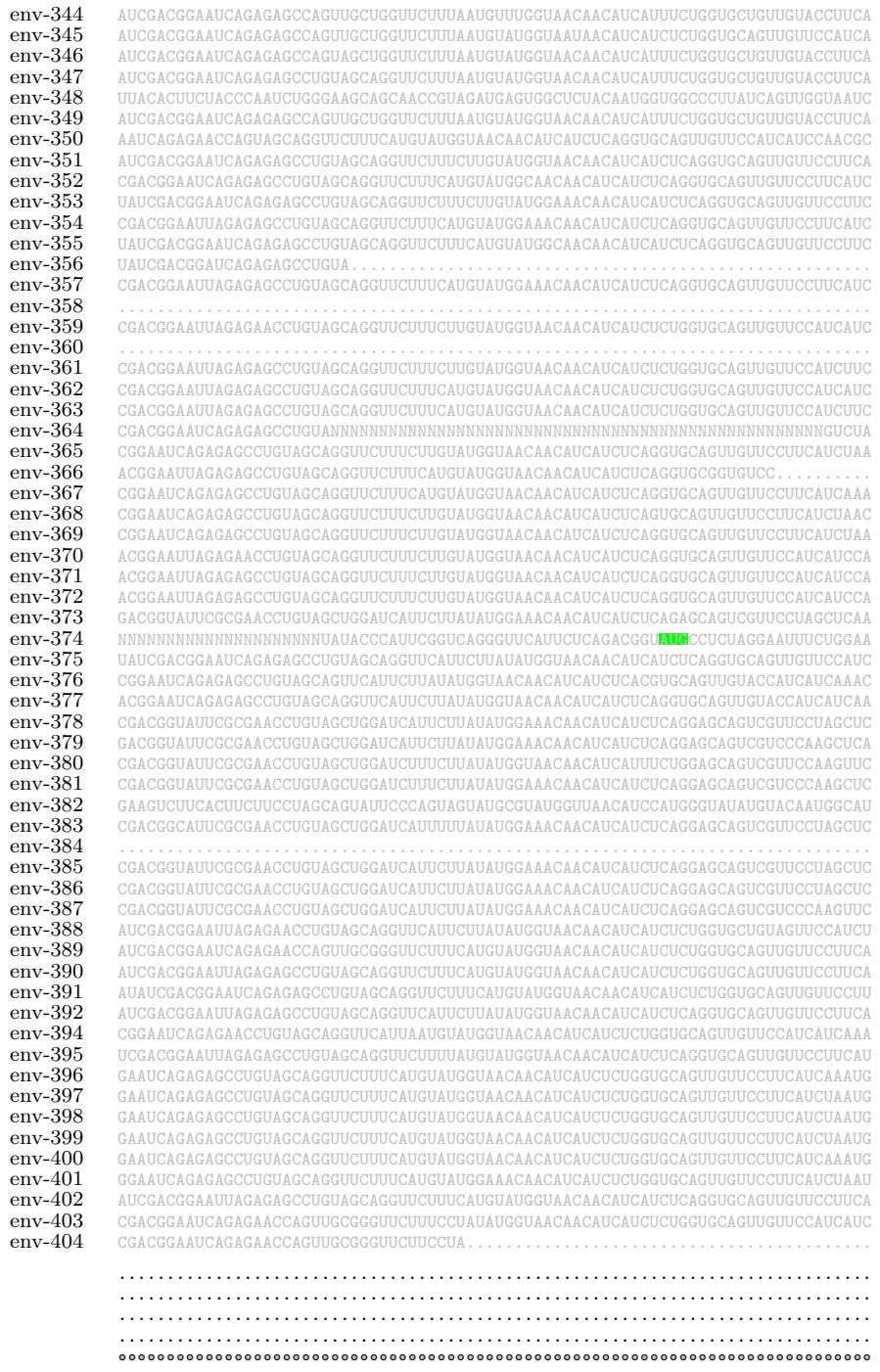
```
env-39  UGAACCCGUUGCUGGUUCACUCAUGUACGGAAACAACAUCAUCUCUGGUGCUGUCGUCCCCUCUAGCAAUGCCAUCGG
env-40  CGAUAUCGACGGUAUCCGCGAACCAGUAGCAGGAUCAUUCUUAUAUGGAAACAACAUCAUCUCAGGAGCAGUCGUGCC
env-41  GAAUCAGAGAGCCUGUAGCAGGAUCUCUUAUGUUUGGUAACAACAUCAUUUCUGGUGCUGUUGUACCUUCAUCCAACG
env-42  UAUCGACGGAAUUAGAGAACCAGUAGCGGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCAGGUGCAGUUGUACCAUC
env-43  UAUCGACGGAAUCAGAGAACCAGUAGCGGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCUGGUGCGGUUGUUCCUUC
env-44  UCGACGGAAUCAGAGAACCAGUGGCGGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCUU
env-45  UCGACGGAAUCAGAGAACCAGUGGCGGGUUCUUUAUUGUUUGGUAACAACAUCAUCUCAGGUGCAGUUGUACCAUCAU
env-46  CGACGGAAUCAGAGAACCAGUGGCGGGUUCUUUAAUGUUUGGUAACAACAUCAUCUCAGGUGCAGUUGUACCAUCAUC
env-47  AUCGACGGAAUCAGAGAACCAGUAGCGGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUACCAUCU
env-48  UGAUGGAAUCAGAGAACCUGUAGCAGGUUCAUUCAUGUAUGGUAACAACAUCAUCNNNNNNNNNNNNNNNNNNNNNN
env-49  CGGAAUCAGAGAGCCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUUUCUGGUGCUGUUGUACCUUCAUCCAA
env-50  GACGGAAUCAGAGAACCAGUAGCAGGUUCAUUCAUGUAUGGUAACAACAUCAUCUCUGGUGCUGUAGUUCCAUCUUCA
env-51  UAUCGACGGAAUCAGAGAACCCGUGGCAGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCAGUGCAGUUGUUCCAUCU
env-52  UAUCGACGGAAUCAGAGAACCCGUGGCAGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCAGGUGCAGUUGUUCCAUC
env-53  UAUCGACGGAAUCAGAGAACCCGUGGCAGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCAGGUGCAGUUGUUCCUUC
env-54  UAUCGACGGAAUCAGAGAACCCGUGGCAGGUUCUUUCAUGUAUGGUAACAACAUCAUUUCUGGUGCAGUUGUUCCAUC
env-55  UAUCGACGGAAUCAGAGAACCCGUGGCAGGUUCUUUCAUGUAUGGUAACAACAUCAUUUCUGGUGCAGUUGUUCCAUC
env-56  GACGGAAUCAGAGAGCCAGUAGCUGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCAGGUGCUGUAGUUCCAUCAUCC
env-57  ........................................................................................
env-58  CGACGGAAUCAGAGAGCCAGUAGCUGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCAGGUGCUGUAGUUCCAUCAUC
env-59  GACGGAAUCAGAGAGCCAGUAGCUGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCAGGUGCUGUAGUUCCAUCAUCC
env-60  GACGGAAUCAGAGAGCCAGUAGCUGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCAGGUGCUGUCGUUCCAUCAUCC
env-61  CGACGGAAUCAGAGAGCCAGUAGCUGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCAGGUGCUGUAGUUCCA.....
env-62  UCGACGGAAUCAGAGAGCCAGUUGCUGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCAGGUGCUGUAGUUCCAUCAU
env-63  GACGGAAUCAGAGAGCCAGUAGCUGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCAGGUGCUGUAGUUCCAUCAUCC
env-64  GACGGAAUCAGAGAGCCAGUAGCUGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCAGGUGCUGUAGUUCCAUCAUCC
env-65  UCGACGGAAUCAGAGAGCCAGUAGCUGGUUCUUUCUUGUAUGGUAACAACAUCAUCUCAGGUGCUGUAGUUCCAUCAU
env-66  GGAAUCAGAGAACCAGUAGCGGGUUCUUUCUUAUAUGGUAAUAACAUCAUCUCUGGUGCUAUUGUUCCAUCAUCAAAC
env-67  AUAUCGACGGAAUCAGAGAACCAGUAGCGGGUUCUUUCUUAUAUGGCAACAAUAUCAUCUCUGGUGCAGUUGUUCCAU
env-68  CGGAAUCAGAGAACCAGUAGCAGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCAGGUGCAGUUGUUCCAUCAUCCAA
env-69  AUCGACGGAAUCAGAGAGCCAGUAGCUGGUUCUUUAAUGUAUGGUAACAACAUCAUUUCUGGUGCUGUUGUACCUUCA
env-70  GAAUCAGAGAACCAGUUGCGGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCUUCAAACG
env-71  GGAAUCAGAGAACCAGUUGCGGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCUUCAAAC
env-72  GAAUCAGAGAACCAGUUGCGGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCUUCAAACG
env-73  CCGUGAACCCGUUGCUGGUUCACUCAUGUACGGAAACAACAUCAUCUCUGGUGCUGUCGUCCCCUCUAGCAAUGCCAU
env-74  UCCGUGAACCCGUAGCUGGUUCACUCAUGUACGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCUUCUUCAAACGCAA
env-75  GGCAUCCGUGAACCAGUCGCUGGUUCUCUGAUGUAUGGUAACAACAUCAUUUCUGGUGCAGUUGUACCUUCUUC....
env-76  GAUGGCAUCCGUGAACCCGUCGCUGGUUCACUCAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCCUCUUCC
env-77  ........................................................................................
env-78  CGAUGGCAUCCGUGAACCCGUCGCAGGUUCACUCAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCUUCUUC
env-79  CGUGAACCCGUUGCAGGUUCACUCAUGUACGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCUUCUUCCAACGCAAUU
env-80  UGGCAUCCGCGAACCUGUUGCAGGUUCUCUCAUGUAUGGCAACAACAUCAUCUCUGGUGCAGUUGUCCCCAGUUCCAA
env-81  GGUAUCCGUGAACCCGUUGCUGGUUCGCUCAUC.....................................................
env-82  ACGGCAUCCGUGAACCCGUC.....................................................................
env-83  ........................................................................................
env-84  CCGUGAACCCGUCGCUGGUUCACUCAUGUACGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCCUCUUCCAACGCAAU
env-85  UCCGUGAACCCGUCGCUGGUUCACUCAUGUACGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCCUCUUCCAACGCAA
env-86  ........................................................................................
env-87  CCGUGAACCCGUCGCUGGUUCACUCAUGUACGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCCUCUUCCAACGCAAU
env-88  CCGCGAACCUGUUGCAGGUUCUCUCAUGUACGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCCUCUUCCAACGCAAU
env-89  ........................................................................................
env-90  ........................................................................................
env-91  ........................................................................................
env-92  ........................................................................................
env-93  ........................................................................................
env-94  AUCCGUGAACCAGUCGCUGGUUCUCUGAUGUACGGUAACAACAUCAUCUCUGGUGCUGUUGUUCCCUCUUCCAACGCA
env-95  ........................................................................................
env-96  ........................................................................................
env-97  UCGUGAACCUGUCGCAGGUUCUCUAAUGUACGGUAACAACAUCAUCUCUGGUGCUGUCGUUCCGUCUUCUAACGCAAU
env-98  ACGGUAUUCGUGAACCAGUUGCUGGUUCUCUCAUGUAUGGCAACAACAUCAUCUCUGGUGCUGUCGUCCCUAGUUCCA

        ........................................................................................
        ........................................................................................
        ........................................................................................
        ........................................................................................
        °°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°
```

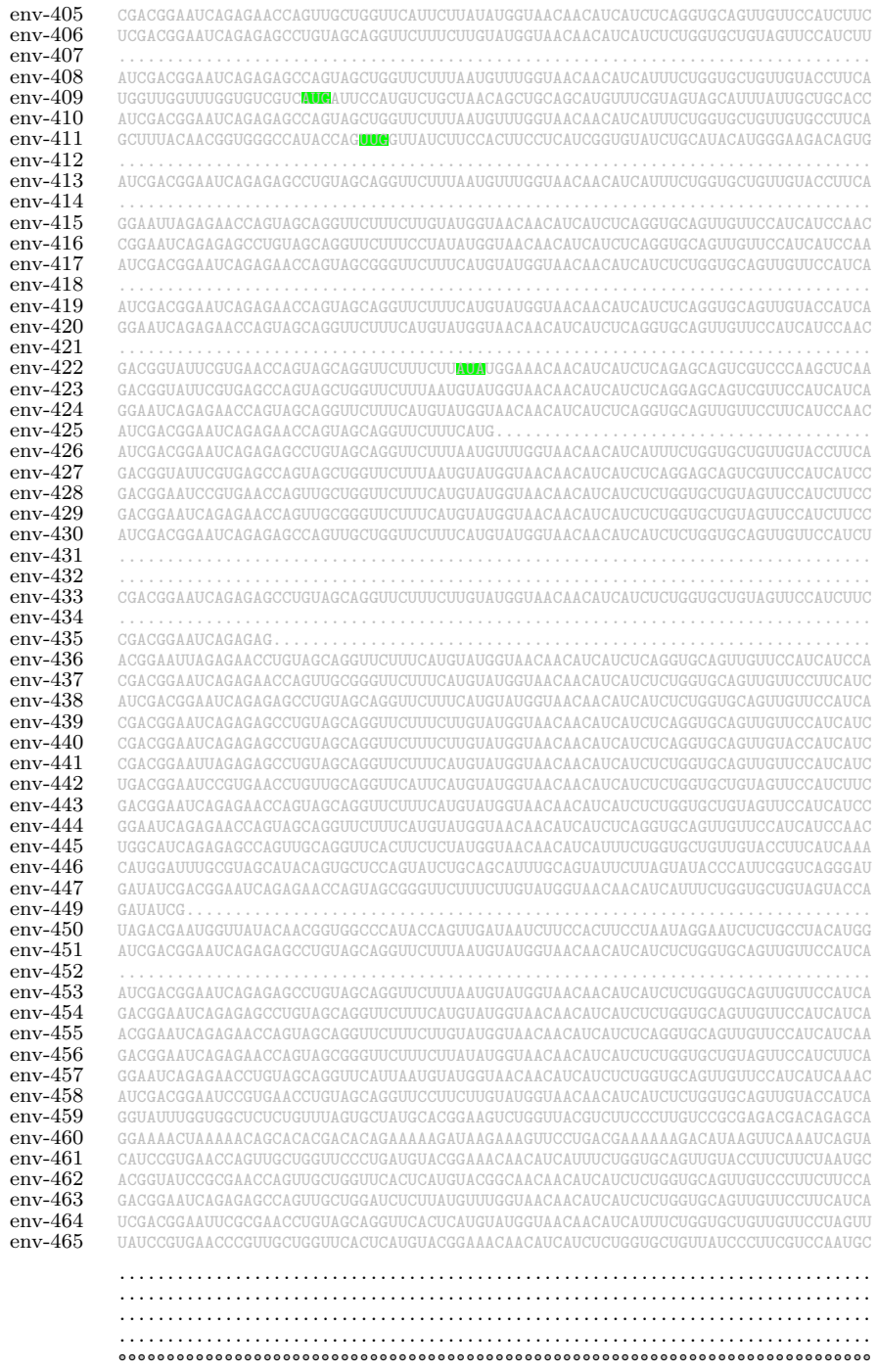

env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-135
env-136
env-137
env-138
env-139
env-140
env-141
env-142
env-143
env-144
env-145
env-146
env-147
env-148
env-150
env-151
env-152
env-153
env-154
env-155
env-156
env-157
env-158
env-159
env-160


env-161
env-162
env-163
env-164
env-165
env-166
env-167
env-168
env-169
env-170
env-171
env-172
env-173
env-174
env-175
env-176
env-177
env-178
env-179
env-180
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
env-190
env-191
env-192
env-193
env-194
env-195
env-196
env-197
env-198
env-199
env-200
env-201
env-202
env-203
env-204
env-205
env-206
env-207
env-208
env-209
env-210
env-211
env-212
env-213
env-214
env-215
env-216
env-217
env-218
env-219
env-220

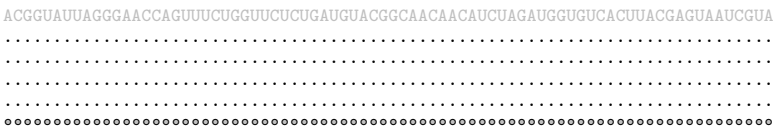
```
env-221  GGAAUUAGAGAGCCUGUAGCAGGUUCAUUCUUAUAUGGUAACAACAUCAUCUCAGGUGCAGUUGUACCAUCCUCAAAC
env-222  GGAAUCAGAGAACCUGUAGCAGGUUCAUUCUUAUAUGGAAACAACAUCAUCUCAGGUGCAGUUGUUCCAUCCUCAAAC
env-223  CGGAAUCAGAGAGCCUGUAGCAGGUUCAUUCUUAUAUGGUAACAACAUCAUCUCAAGUGCAGUUGUUCCUUCAUCAAA
env-224  CGGAAUCAGAGAGCCUGUAGCAGGUUCAUUCUUAUAUGGUAACAACAUCAUCUCAGGUGCAGUUGUACCAUCCUCAAA
env-225  GGAAUCAGAGAACCAGUUGCUGGUUCAUUCUUAUAUGGUAACAACAUCAUCUCAGGUGCAGUUGUACCAUCCUCAAAC
env-226  GGAAUCAGAGAACCUGUAGCAGGUUCAUUCUUAUAUGGUAACAACAUCAUCUCAGGUGCAGUUGUACCAUCCUCAAAC
env-227  CGGAAUCAGAGAGCCUGUAGCAGGUUCAUUCUUAUAUGGUAACAACAUCAUCUCAGGUGCAGUUGUUCCAUCCUCAAA
env-228  CGGAAUCAGAGAACCUGUAGCAGGUUCAUUCUUAUAUGGUAACAACAUCAUCUCAGGUGCAGUUGUUCCUUCAUCAAA
env-229  CAUGGGCAGACAGUGGGAACUAUCAUACAGAUUAGGUAUGAGACCUUGGAUUUGCGUUGCAUAUUCCGCACCAGUAUC
env-230  CGGAAUUAGAGAGCCUGUAGCAGGUUCAUUCUUAUAUGGUAACAACAUCAUCUCAGGUGCAGUUGUUCCUUCAUCAAA
env-231  CGGAAUUAGAGAACCAGUUGCGGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCUGGUGCUGUAGUUCCAUCAUCAAA
env-232  AUCGACGGAAUUAGAGAACCAGUUGCGGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCUUCA
env-233  AUAUCGACGGAAUCAGAGAGCCUGUAGCAGGUUCAUUCUUAUAUGGUAACAACAUCAUCUCUGGUGCGGUUGUUCCAU
env-234  ACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCGGUUGUUCCUUCUUCAA
env-235  AUCGACGGAAUCAGAGAGCCAGUUGCUGGUUCUUUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCA
env-236  UAUCGACGGAAUCAGAGAGCCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCUUC
env-237  UAUCGACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCUUC
env-238  ACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGAAACAACAUCAUCUCAGGUGCAGUUGUUCCUUCAUCUA
env-239  ........................................................................................
env-240  ........................................................................................
env-241  ACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCGGUUGUUCCU.......
env-242  ACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGCAACAACAUCAUCUCUGGUGCAGUUGUUCCUUCUUCAA
env-243  ACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCGGUUGUUCCUUCUUCAA
env-244  ACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCGGUUGUUCCUUCUUCAA
env-245  ACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCUUCUUCAA
env-246  ACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCGGUUGUUCCUUCUUCAA
env-247  ACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCUUCUUCAA
env-248  ........................................................................................
env-249  ACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCGGUUGUUCCUUCUUCAA
env-250  ACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCGGUUGUUCCUUCUUCAA
env-251  ACGGAAUCAGAGAGCCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCAGGUGCGGUUGUUCCUUCUUCAA
env-252  ACGGAAUCAGAGAACCU.....................................................................
env-253  GCAACAGUAGACGAGUGGUUAUACAACGGUGGCCCUUAUCAGUUAGUAAUCUUCCACUUCCUAAUCGGAAUCUCUGCC
env-254  ACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCGGUUGUUCCUUCUUCAA
env-255  ACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCGGUUGUUCCUUCUUCAA
env-256  ACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCGGUUGUUCCUUCUUCAA
env-257  ACGGCAUCAGAGAGCCUGUAGCAGGUGCUCUAAUGUACGGAAACAACAUCAUCUCAGGUGCAGUUGUUCCUUCAUCAA
env-258  UUGACGGCAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGAAACAAUAUCAUCUCAGGUGCAGUUGUUCCUUCAU
env-259  ACGGCAUCAGAGAGC...........................................................................
env-260  AUCGACGGAAUCCGUGAACCAGUUGCUGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCA
env-261  UCGACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGAAACAACAUCAUCUCAGGUGCAGUUGUUCCUUCAU
env-262  UAUCGACGGAAUCAGAGAGCCUGUAGCAGGUGCUCUAAUGUAUGGAAACAACAUCAUCUCAGGUGCAGUUGUUCCUUC
env-263  ........................................................................................
env-264  ........................................................................................
env-265  ........................................................................................
env-266  GGAAUCAGAGAGCCUGUAGCAGGUUCAUUCUUAUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUACCAUCAUCCAAC
env-267  ACGGAAUCAGAGAGCCUGUAGCAGGUUCAUUCUUAUAUGGUAACAACAUCAUCUCAGGUGCAGUUGUACCAUCAUCAA
env-268  ........................................................................................
env-269  GACGGAAUCAGAGAACCAGUUGCGGGUUCUUUCUUAUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCUUCA
env-270  AUCGACGGAAUCAGAGAGCCAGUUGCUGGUUCUUUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCA
env-271  AUCGACGGAAUCAGAGAGCCAGUUGCUGGUUCUUUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCA
env-272  AUCGACGGAAUCAGAGAGCCAGUUGCUGGUUCUUUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCA
env-273  AUCGACGGAAUCAGAGAGCCUGUAGCAGGUUCUUUCAUGUAUGGUAACAAUAUCAUCUCUGGUGCAAUUGUUCCAUCA
env-274  UCGACGGAAUCAGAGAGCCAGUUGCUGGUUCUUUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCAU
env-275  AUCGACGGAAUCAGAGAGCCUGUAGCAGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCUGGUGCAAUUGUUCCAUCA
env-276  AUCGACGGAAUCAGAGAGCCAGUUGCUGGUUCUUUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCA
env-277  AUCGACGGAAUCAGAGAGCCAGUUGCUGGUUCUUUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCA
env-278  CGACGGAAUUAGAGAACCUGUAGCAGGUUCAUUCUUAUAUGGUAACAACAUCAUCUCUGGUGCUGUAGUUCCAUCUUC
env-279  ........................................................................................
env-280  CGACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCUUC

         ........................................................................................
         ........................................................................................
         ........................................................................................
         ........................................................................................
         °°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°°
```

```
env-281  CGACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCUUC
env-282  CGACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCUUC
env-284  GACGAAUCAGAGACCUGUAGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNACACUU
env-285  CGACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCUUC
env-287  CGACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCUUC
env-288  CGACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCUUC
env-289  CGACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCUUC
env-290  CGACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCUUC
env-291  CGACGGAAUCAGAGAACCUGUAGCAGGUGCUCUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCUUC
env-292  ................................................................................
env-293  CAGUUGUUCCCAGUUCAAACGCAAUCGGUCUACAUUUCUACCCAAUUUGGGAAGCAGCAACUGUAGACGAGUGGUUAU
env-294  ................................................................................
env-295  ................................................................................
env-296  ................................................................................
env-297  UCGACGGAAUUAGAGAACCUGUAGCAGGUUCAUUCUUAUAUGGUAACAACAUCAUCUCUGGUGCUGUAGUUCCAUCUU
env-298  UCGACGGAAUUAGAGAACCUGUAGCAGGUUCAUUCUUAUAUGGUAACAACAUCAUCUCUGGUGCUGUAGUUCCAUCUU
env-300  CGACGGAAUUAGAGAACCUGUAGCAGGUUCAUUCUUAUAUGGUAACAACAUCAUCUCUGGUGCUGUAGUUCCAUCUUC
env-301  AGACAACAGAAAAUGAAUCUCAAAAUUAUGGGUAUAAGUUUGGACAAGAAGAAGAAACAUACAACAUUGUCGCUGCUC
env-302  ................................................................................
env-303  CGACGGAAUUAGAGAACCUGUAGCAGGUUCAUUCUUAUAUGGUAACAACAUCAUCUCUGGUGCUGUAGUUCCAUCUUC
env-304  AUCGACGGAAUCAGAGAGCCAGUUGCUGGUUCUUUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCA
env-305  ................................................................................
env-306  ACGGAAUCAGAGAACCAGUAGCGGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCUUCAUCAA
env-307  ACGGAAUCAGAGAGCCAGUUGCUGGUUCUUUAAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCAUCAA
env-308  ................................................................................
env-309  UUGAUGGAAUCAGAGAACCUGUAGCAGGUUCAUUCAUGUAUGGUAACAACAUCAUCUCAGGUGCAGUUGUACCAUCUU
env-310  UUGAUGGAAUCAGAGAACCUGUAGCAGGUUCAUUCAUGUAUGGUAACAACAUCAUCUCAGGUGCAGUUGUACCAUCUU
env-311  CGGAAUCAGAGAACCUGUAGCAGGUUCAUUCUUAUAUGGAAACAACAUCAUCUCAGGUGCAGUUGUUCCAUCCUCAAA
env-312  GACGGAAUCCGUGAACCAGUAGCUGGAUCUUUCUUAUAUGGUAACAACAUCAUUUCAGGAGCAGUCGUUCCAAGUUCC
env-313  ACGGUAUUCGUGAACCAGUAGCUGGUUCUUUCUUAUAUGGAAACAACAUCAUCUCAGGAGCAGUCGUUCCUAGCUCAA
env-314  ACGGAAUCCGUGAACCAGUUGCUGGUUCAUUUAUGUAUGGAAACAACAUCAUCUCUGGUGCUGUAGUUCCAUCUUCAA
env-315  GACGGAAUCAGAGAACCAGUAGCUGGUUCAUUCAUGUAUGGUAACAACAUCAUUUCUGGUGCUGUAGUUCCAUCUUCC
env-316  GAAUCAGAGAACCAGUUGCUGGUUCAUUCAUGUAUGGUAACAACAUCAUCUCUGGUGCUGUAGUUCCAUCUUCAAACG
env-317  UAUCGACGGAAUCAGAGAACCCGUGGCAGGUUCUUUCAUGUAUGGUAACAACAUCAUUUCUGGUGCAGUUGUUCCAUC
env-318  UCGACGGAAUCAGAGAGCCUGUAGCAGGUUCAUUUAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCAU
env-319  AUCGAUGGAAUCAGAGAACCAGUUGCGGGUUCUUUUAUGUAUGGUAACAACAUCAUCUCAGGUGCUGUAGUUCCAUCU
env-320  GACGGAAUCAGAGAACCAGUAGCGGGUUCUUUCUUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUACCAUCUUCA
env-321  GUUUCAUCGUUGCAUUCAUUGCAGCACCUCCUGUCGAUAUCGACGGAAUCAGAGAACCCGUGGCAGGUUCUUUCAUGU
env-322  ................................................................................
env-323  GGAAUCAGAGAACCAGUAGCGGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCUUCAUCAAAC
env-324  ................................................................................
env-325  ................................................................................
env-326  AUCGACGGAAUCAGAGAGCCAGUAGCUGGUUCUUUAAUGUAUGGUAACAACAUCAUUUCUGGUGCUGUUGUACCUUCA
env-327  ................................................................................
env-328  ................................................................................
env-329  UCGACGGAAUCAGAGAGCCUGUAGCAGGUUCAUUUAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCAU
env-330  CGACGGAAUCAGAGAGCCUGUAGCAGGUUCUUUAAUGUAUGGUAACAACAUCAUCUCUGGUGCUGUAGUUCCAUCUUC
env-331  UCGACGGAAUCAGAGAGCCUGUAGCAGGUUCAUUUAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCAU
env-332  UCGACGGAAUCAGAGAGCCUGUAGCAGGUUCAUUUAUGUAUGGUAACAACAUCAUCUCUGGUGCUGUAGUUCCAUCAU
env-333  UCGACGGAAUCAGAGAGCCUGUAGCAGGUUCAUUUAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCAU
env-334  UCGACGGAAUCAGAGAGCCUGUAGCAGGUUCAUUUAUGUAUGGUAACAACAUCAUCUCUGGUGCAGUUGUUCCAUCAU
env-335  UCGACGGAAUCAGAGAGCCUGUAGCAGGUUCAUUUAUGUAUGGUAACAACAUCAUCUCAGGUGCAGUUGUUCCAUCAU
env-336  AUCGACGGAAUCAGAGAGCCAGUUGCUGGUUCUUUAAUGUAUGGAAACAACAUCAUUUCUGGUGCAGUUGUUCCAUCA
env-337  ................................................................................
env-338  CGACGGAAUCAGAGAGCCUGUAGCAGGUUCUUUCUUGUAUGGUAACAACAUCAUUUCUGGUGCUGUUGUCCCAUCAUC
env-339  UAUCGACGGAAUCAGAGAACCAGUAGCAGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCAGGUGCAGUUGUUCCAUC
env-340  UAUCGACGGAAUCAGAGAACCAGUAGCAGGUUCUUUCAUGUAUGGUAACAACAUCAUCUCAGGUGCAGUUGUACCAUC
env-341  UCGACGGAAUCAGAGAGCCUGUAGCAGGUUCUUUCAUGUAUGGAAACAACAUCAUCUCAGGUGCAGUUGUUCCUUCAU
env-342  AUCGACGGUAUUCGUGAGCCAGUAGCUGGUUCUUUAAUGUAUGGUAACAACAUCAUCUCAGGAGCAGUCGUUCCAUCA
env-343  AUCGACGGAAUCAGAGAGCCUGUAGCAGGUUCUUUAAUGUAUGGUAACAACAUCAUUUCUGGUGCUGUUGUACCUUCA

         ................................................................................
         ................................................................................
         ................................................................................
         ................................................................................
         oooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooo
```


env-344
env-345
env-346
env-347
env-348
env-349
env-350
env-351
env-352
env-353
env-354
env-355
env-356
env-357
env-358
env-359
env-360
env-361
env-362
env-363
env-364
env-365
env-366
env-367
env-368
env-369
env-370
env-371
env-372
env-373
env-374
env-375
env-376
env-377
env-378
env-379
env-380
env-381
env-382
env-383
env-384
env-385
env-386
env-387
env-388
env-389
env-390
env-391
env-392
env-394
env-395
env-396
env-397
env-398
env-399
env-400
env-401
env-402
env-403
env-404


env-405
env-406
env-407
env-408
env-409
env-410
env-411
env-412
env-413
env-414
env-415
env-416
env-417
env-418
env-419
env-420
env-421
env-422
env-423
env-424
env-425
env-426
env-427
env-428
env-429
env-430
env-431
env-432
env-433
env-434
env-435
env-436
env-437
env-438
env-439
env-440
env-441
env-442
env-443
env-444
env-445
env-446
env-447
env-449
env-450
env-451
env-452
env-453
env-454
env-455
env-456
env-457
env-458
env-459
env-460
env-461
env-462
env-463
env-464
env-465

env-466 ACGGUAUUAGGGAACCAGUUUCUGGUUCUCUGAUGUACGGCAACAACAUCUAGAUGGUGUCACUUACGAGUAAUCGUA



67 *Polynucleobacter*-1

67.1 Taxa

The taxonomy of each organism containing a putative *Polynucleobacter*-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *Polynucleobacter*-1 RNA in Sections 67.2 and 67.4.

abbrev. of hits	taxonomy of species
Psp-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Polynucleobacter sp.* QLW-P1DMWA-1
env-1 to env-15	environmental samples

67.2 Gene contexts

Each *Polynucleobacter*-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *Polynucleobacter*-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 67.3. The accession number of the sequence containing each *Polynucleobacter*-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 67.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	JCVI_SCAF_1096626897839	-	1248	1077	←? (GOS10837) **RNA**→
Psp-1-1	NC_009379.1	+	2130772	2130906	←hypo **RNA**→ hypo→
env-2	JCVI_SCAF_1096627456148	-	488	350	←? (GOS10837) **RNA**→
env-3	JCVI_SCAF_1101668104807	+	141	279	**RNA**→
env-4	JCVI_SCAF_1096627448889	+	610	913	? (GOS11034)→ **RNA**→ ←? (GOS11352119)
env-5	JCVI_SCAF_1096627016889	+	1328	1469	? (GOS11780615)→ **RNA**→ ←? (GOS11352119)
env-6	JCVI_SCAF_1101668617863	+	1108	1398	? (GOS11034)→ **RNA**→
env-7	JCVI_SCAF_1101668644630	-	563	273	? (GOS11034)→ **RNA**→
env-8	JCVI_SCAF_1096627450512	+	351	648	**RNA**→
env-9	JCVI_SCAF_1101668251682	-	997	720	? (GOS11034)→ **RNA**→
env-10	JCVI_SCAF_1096627012967	-	2433	2156	? (GOS11034)→ **RNA**→ ←? (GOS11023188)
env-11	JCVI_SCAF_1096627005925	-	385	112	? (GOS11034)→ **RNA**→
env-12	JCVI_SCAF_1096627670229	-	475	196	FabG (COG1028)? (GOS13118986)→ **RNA**→
env-13	JCVI_SCAF_1101668234032	-	1153	872	? (GOS11034)→ **RNA**→ ←? (GOS11023188)
env-14	JCVI_SCAF_1096627671665	+	558	837	? (GOS11034)→ **RNA**→ ←? (GOS11023188)
env-15	JCVI_SCAF_1101668145292	-	911	630	? (GOS11034)→ **RNA**→ ←? (GOS1582)

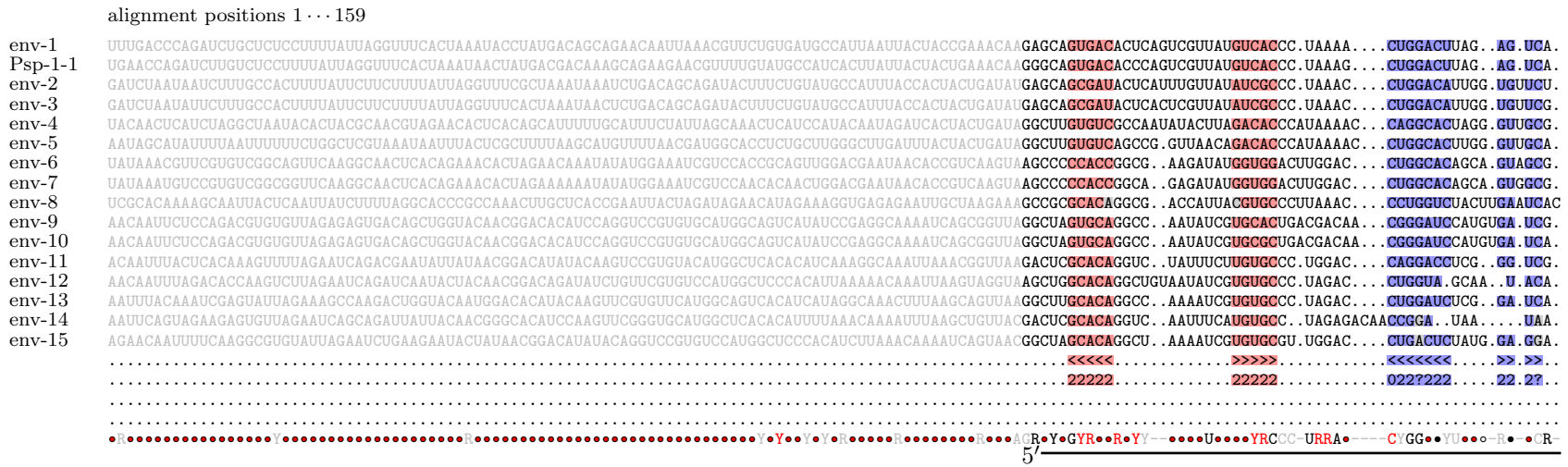
67.3 Conserved domains

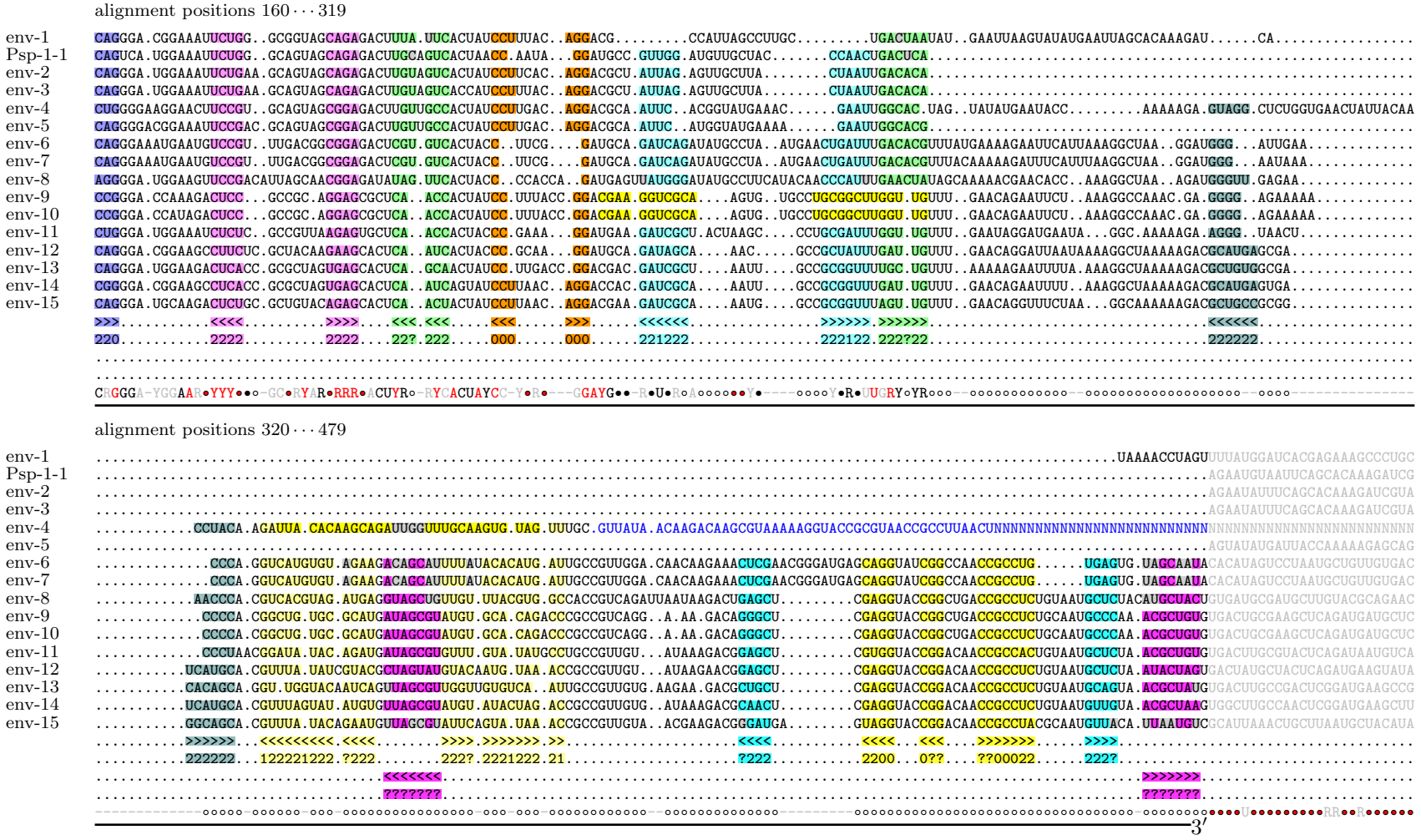
Conserved domains found in protein-coding genes listed in Section 67.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *Polynucleobacter*-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

u COG1028 (1) Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases) [Secondary metabolites biosynthesis, transport, and catabolism / General function prediction only]
u GOS10837 (2) ?
d GOS11023188 (3) ?
u GOS11034 (9) ?
d GOS11352119 (2) ?
u GOS11780615 (1) ?
u GOS13118986 (1) ?
d GOS1582 (1) ?

67.4 Multiple-sequence alignment

Each *Polynucleobacter*-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 67.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *Polynucleobacter*-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below this base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.




alignment positions 160 · · · 319
env-1
Psp-1-1
env-2
env-3
env-4
env-5
env-6
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-14
env-15
alignment positions 320 · · · 479
env-1
Psp-1-1
env-2
env-3
env-4
env-5
env-6
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-14
env-15
3′

68 *potC*

68.1 Taxa

The taxonomy of each organism containing a putative *potC* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *potC* RNA in Sections 68.2 and 68.4.

abbrev. of hits	**taxonomy of species**
env-1 to env-136	environmental samples

68.2 Gene contexts

Each *potC* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *potC* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 68.3. The accession number of the sequence containing each *potC* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 68.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	JCVI_SCAF_1101667512334	-	407	341	**RNA**→ ? (GOS1374)→
env-2	JCVI_SCAF_1096627239868	+	703	769	**RNA**→ ? (GOS1374)→
env-3	JCVI_SCAF_1101668037411	+	756	820	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-4	JCVI_SCAF_1101667035129	-	546	482	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-5	JCVI_SCAF_1101667013937	+	285	349	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-6	JCVI_SCAF_1101668026636	-	1197	1133	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-7	JCVI_SCAF_1101668041109	-	993	929	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-8	JCVI_SCAF_1101668698884	+	365	430	**RNA**→ CysU (COG0555)? (GOS1374)→
env-9	JCVI_SCAF_1101667002397	-	418	349	**RNA**→
env-10	JCVI_SCAF_1101668530459	+	723	792	**RNA**→ CysU (COG0555)? (GOS1374)→
env-11	JCVI_SCAF_1096627913997	-	1529	1458	**RNA**→ CysU (COG0555)? (GOS1374)→ ThiP (COG1178)? (GOS1374)→
env-12	JCVI_SCAF_1101667385554	+	419	491	**RNA**→ PotC (COG1177)? (GOS1374)→
env-13	JCVI_SCAF_1096626860626	-	684	621	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-14	JCVI_SCAF_1096626663908	-	922	854	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ? (GOS1374)→
env-15	JCVI_SCAF_1101668477367	-	507	438	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-16	JCVI_SCAF_1096627910193	-	1593	1524	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ThiP (COG1178)? (GOS1374)→
env-17	JCVI_SCAF_1096627347580	-	385	316	**RNA**→ ? (GOS1374)→
env-18	JCVI_SCAF_1096627328274	+	1024	1091	**RNA**→
env-19	JCVI_SCAF_1101668435226	-	689	620	**RNA**→ ? (GOS1374)→
env-20	JCVI_SCAF_1101668420392	-	179	111	**RNA**→
env-21	JCVI_SCAF_1096627337314	+	2906	2975	**RNA**→
env-22	JCVI_SCAF_1096627963680	+	202	271	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ? (GOS1374)→
env-23	JCVI_SCAF_1101668361177	-	369	300	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-24	JCVI_SCAF_1101667343497	-	505	437	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-25	JCVI_SCAF_1096627325638	-	1587	1518	**RNA**→ AHP1 (COG0678)? (GOS7685)→ CysU (COG0555)? (GOS1374)→
env-26	JCVI_SCAF_1096628256610	-	199	130	**RNA**→
env-27	JCVI_SCAF_1101668329984	-	591	522	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-28	JCVI_SCAF_1096627072684	-	1905	1838	**RNA**→ ThiP (COG1178)? (GOS1374)→
env-29	JCVI_SCAF_1096627095442	-	303	236	**RNA**→ ? (GOS7685)→
env-30	JCVI_SCAF_1096627788803	+	647	715	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-31	JCVI_SCAF_1101667278325	-	68	1	**RNA**→
env-32	JCVI_SCAF_1096627841184	+	1512	1580	**RNA**→
env-33	JCVI_SCAF_1096627099136	-	1404	1336	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ThiP (COG1178)? (GOS1374)→
env-34	JCVI_SCAF_1101668687970	-	933	865	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ? (GOS1374)→
env-35	JCVI_SCAF_1096627131098	-	2171	2103	**RNA**→ AHP1 (COG0678)? (GOS7685)→ CysU (COG0555)? (GOS1374)→ PotC (COG1177)? (GOS1374)→
env-36	JCVI_SCAF_1096627161299	+	646	714	**RNA**→ AHP1 (COG0678)? (GOS7685)→ CysU (COG0555)? (GOS1374)→
env-37	JCVI_SCAF_1101667288740	-	801	733	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ? (GOS1374)→
env-38	JCVI_SCAF_1096627100744	-	1734	1666	**RNA**→ AHP1 (COG0678)? (GOS7685)→ CysU (COG0555)? (GOS1374)→

env-39	JCVI_SCAF_1101668403979	+	763	831	**RNA**→
env-40	JCVI_SCAF_1096627072641	-	1598	1530	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ThiP (COG1178)? (GOS1374)→
env-41	JCVI_SCAF_1096627841344	-	1169	1101	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-42	JCVI_SCAF_1101668734904	-	1527	1459	**RNA**→ ThiP (COG1178)? (GOS1374)→
env-43	JCVI_SCAF_1096626946579	+	656	724	**RNA**→ AHP1 (COG0678)? (GOS7685)→ CysU (COG0555)? (GOS1374)→
env-44	JCVI_SCAF_1101668429312	-	815	747	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ? (GOS1374)→
env-45	JCVI_SCAF_1096627855974	-	482	414	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-46	JCVI_SCAF_1101668574907	+	43	112	**RNA**→ AHP1 (COG0678)? (GOS7685)→ CysU (COG0555)? (GOS1374)→
env-47	JCVI_SCAF_1096627779946	+	408	477	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-48	JCVI_SCAF_1096627331178	-	1298	1230	**RNA**→ ThiP (COG1178)? (GOS1374)→
env-49	JCVI_SCAF_1096627174886	+	1616	1684	**RNA**→
env-50	JCVI_SCAF_1101668270954	+	130	204	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ThiP (COG1178)? (GOS1374)→
env-51	JCVI_SCAF_1101668590969	+	1144	1213	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-52	JCVI_SCAF_1101668455117	-	76	7	**RNA**→
env-53	JCVI_SCAF_1096627072716	-	925	856	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ? (GOS1374)→
env-54	JCVI_SCAF_1096627318085	+	325	394	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ThiP (COG1178)? (GOS1374)→
env-55	JCVI_SCAF_1101667344869	+	864	933	**RNA**→
env-56	JCVI_SCAF_1096628036144	-	1639	1570	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-57	JCVI_SCAF_1096627105346	+	1566	1635	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-58	JCVI_SCAF_1101668282680	-	597	528	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-59	JCVI_SCAF_1096628399615	+	1005	1074	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-60	JCVI_SCAF_1096627992131	+	225	294	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-61	JCVI_SCAF_1101668430573	-	675	606	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-62	JCVI_SCAF_1096627701009	+	509	578	**RNA**→ ? (GOS7685)→ ? (GOS1374)→
env-63	JCVI_SCAF_1101667273902	+	669	738	**RNA**→
env-64	JCVI_SCAF_1101668259161	+	761	830	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ? (GOS1374)→
env-65	JCVI_SCAF_1101667192012	+	52	121	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-66	JCVI_SCAF_1101668718540	+	945	1014	**RNA**→ ? (GOS1374)→
env-67	JCVI_SCAF_1101668480152	+	234	303	**RNA**→ ThiP (COG1178)? (GOS1374)→
env-68	JCVI_SCAF_1101668353387	-	822	753	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ? (GOS1374)→
env-69	JCVI_SCAF_1096628155684	-	413	344	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-70	JCVI_SCAF_1096627325938	-	1890	1821	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ThiP (COG1178)? (GOS1374)→
env-71	JCVI_SCAF_1096627180704	+	2373	2442	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-72	JCVI_SCAF_1096627248519	+	1545	1614	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ? (GOS1374)→
env-73	JCVI_SCAF_1101668226220	+	1227	1296	**RNA**→ ? (GOS7685)→
env-74	JCVI_SCAF_1101668376505	-	933	864	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ? (GOS1374)→
env-75	JCVI_SCAF_1096628171252	+	390	460	**RNA**→ AHP1 (COG0678)? (GOS7685)→ PotA (COG3842)? (GOS2568)→
env-76	JCVI_SCAF_1096627566413	-	689	617	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-77	JCVI_SCAF_1096627782036	-	1600	1531	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-78	JCVI_SCAF_1101668365470	+	1123	1192	**RNA**→
env-79	JCVI_SCAF_1101667100719	+	625	694	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-80	JCVI_SCAF_1101668425195	+	266	335	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ? (GOS1374)→
env-81	JCVI_SCAF_1096627253543	+	1381	1449	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-82	JCVI_SCAF_1101668498086	+	548	620	**RNA**→ AHP1 (COG0678)? (GOS7685)→ PotC (COG1177)? (GOS1374)→
env-83	JCVI_SCAF_1096627856650	-	677	608	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-84	JCVI_SCAF_1101668738454	-	1358	1286	**RNA**→ AHP1 (COG0678)? (GOS7685)→ CysU (COG0555)? (GOS1374)→
env-85	JCVI_SCAF_1096627567281	-	1362	1293	**RNA**→ AHP1 (COG0678)? (GOS7685)→ CysU (COG0555)? (GOS1374)→
env-86	JCVI_SCAF_1101668219540	-	454	385	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-87	JCVI_SCAF_1101668290024	+	1406	1475	**RNA**→
env-88	JCVI_SCAF_1096627834566	-	73	4	**RNA**→
env-89	JCVI_SCAF_1101668371053	+	1048	1117	**RNA**→
env-90	JCVI_SCAF_1096627039410	+	1452	1521	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ? (GOS1374)→

env-91	JCVI_SCAF_1096627790331	-	1449	1380	**RNA**→ AHP1 (COG0678)? (GOS7685)→ CysU (COG0555)? (GOS1374)→
env-92	JCVI_SCAF_1096627996692	-	363	294	**RNA**→ ? (GOS1374)→
env-93	JCVI_SCAF_1096626992844	+	351	420	**RNA**→ AHP1 (COG0678)? (GOS7685)→ CysU (COG0555)? (GOS1374)→ 16S_rRNA→
env-94	JCVI_SCAF_1101668507005	-	805	736	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-95	JCVI_SCAF_1101668513320	-	397	328	**RNA**→ ? (GOS1374)→
env-96	JCVI_SCAF_1096627169596	-	597	528	**RNA**→ PotC (COG1177)? (GOS1374)→
env-97	JCVI_SCAF_1096627998679	-	946	877	**RNA**→ CysU (COG0555)? (GOS1374)→
env-98	JCVI_SCAF_1101668584466	+	600	671	**RNA**→ CysU (COG0555)? (GOS1374)→
env-99	JCVI_SCAF_1101668690257	+	93	162	**RNA**→ ThiP (COG1178)? (GOS1374)→ ? (GOS1374)→
env-100	JCVI_SCAF_1101668533625	+	971	1040	**RNA**→ PotC (COG1177)? (GOS1374)→
env-101	JCVI_SCAF_1096627330646	+	79	148	**RNA**→ ThiP (COG1178)? (GOS1374)→
env-102	JCVI_SCAF_1101668664340	-	191	122	**RNA**→
env-103	JCVI_SCAF_1101668529877	-	1099	1030	**RNA**→ ThiP (COG1178)? (GOS1374)→
env-104	JCVI_SCAF_1101668495371	-	135	66	**RNA**→
env-105	JCVI_SCAF_1101668658199	+	390	460	**RNA**→ AHP1 (COG0678)? (GOS7685)→ UgpE (COG0395)? (GOS1374)→
env-106	JCVI_SCAF_1096626321794	-	206	137	**RNA**→
env-107	JCVI_SCAF_1101667337990	-	574	505	**RNA**→ PotC (COG1177)? (GOS1374)→
env-108	JCVI_SCAF_1096627982342	-	934	860	**RNA**→ ThiP (COG1178)? (GOS1374)→
env-109	JCVI_SCAF_1096627484529	+	1359	1428	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-110	JCVI_SCAF_1101668686095	+	591	659	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ? (GOS1374)→
env-111	JCVI_SCAF_1101668465438	+	475	544	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-112	JCVI_SCAF_1096627348752	+	537	605	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ThiP (COG1178)? (GOS1374)→
env-113	JCVI_SCAF_1101668323967	-	1207	1139	**RNA**→ AHP1 (COG0678)? (GOS7685)→ AHP1 (COG0678)? (GOS7685)→ ? (GOS1374)→
env-114	JCVI_SCAF_1096627084081	-	1586	1518	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ? (GOS7685)→ CysU (COG0555)? (GOS1374)→
env-115	JCVI_SCAF_1101667303053	-	490	422	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-116	JCVI_SCAF_1101668356818	+	956	1024	**RNA**→
env-117	JCVI_SCAF_1096627101375	+	1029	1097	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-118	JCVI_SCAF_1096627912514	+	1447	1515	**RNA**→
env-119	JCVI_SCAF_1101668538228	-	1294	1226	**RNA**→ AHP1 (COG0678)? (GOS7685)→ CysU (COG0555)? (GOS1374)→
env-120	JCVI_SCAF_1096627131838	+	401	469	**RNA**→ AHP1 (COG0678)? (GOS7685)→ ThiP (COG1178)? (GOS1374)→ PotA (COG3842)? (GOS2568)→
env-121	JCVI_SCAF_1101668710754	+	1642	1710	**RNA**→
env-122	JCVI_SCAF_1096627902609	+	39	106	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-123	JCVI_SCAF_1096627153683	+	85	153	**RNA**→ ThiP (COG1178)? (GOS1374)→
env-124	JCVI_SCAF_1096627372012	-	634	566	**RNA**→ PotC (COG1177)? (GOS1374)→
env-125	JCVI_SCAF_1096627331539	-	2658	2590	**RNA**→ ThiP (COG1178)? (GOS1374)→ PotA (COG3842)? (GOS2568)→
env-126	JCVI_SCAF_1096627107251	-	233	165	**RNA**→
env-127	JCVI_SCAF_1096627328443	+	1904	1972	**RNA**→ ? (GOS1374)→
env-128	JCVI_SCAF_1101668425231	-	951	883	**RNA**→ ThiP (COG1178)? (GOS1374)→
env-129	JCVI_SCAF_1096627163365	+	3046	3114	**RNA**→ ThiP (COG1178)? (GOS1374)→
env-130	JCVI_SCAF_1101668562326	-	436	368	**RNA**→ ? (GOS1374)→
env-131	JCVI_SCAF_1101667231885	-	196	128	**RNA**→
env-132	JCVI_SCAF_1101668667072	+	1201	1273	**RNA**→
env-133	JCVI_SCAF_1096627838700	+	734	806	**RNA**→ PotC (COG1177)? (GOS1374)→
env-134	JCVI_SCAF_1096627166035	-	565	493	**RNA**→ AHP1 (COG0678)? (GOS7685)→
env-135	JCVI_SCAF_1101668473132	-	1217	1146	**RNA**→ CysU (COG0555)? (GOS1374)→
env-136	JCVI_SCAF_1101668708390	+	445	514	**RNA**→ ThiP (COG1178)? (GOS1374)→

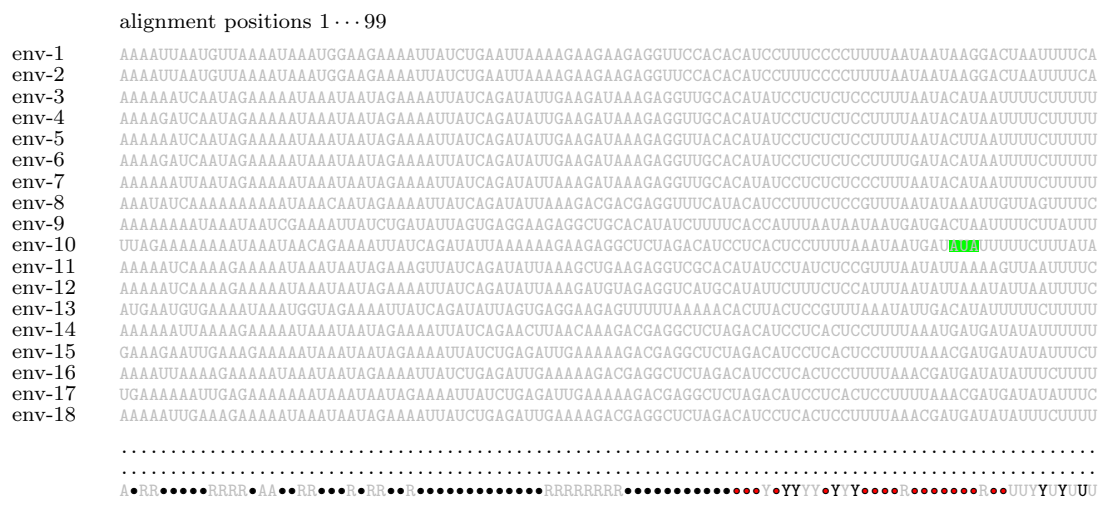
68.3 Conserved domains

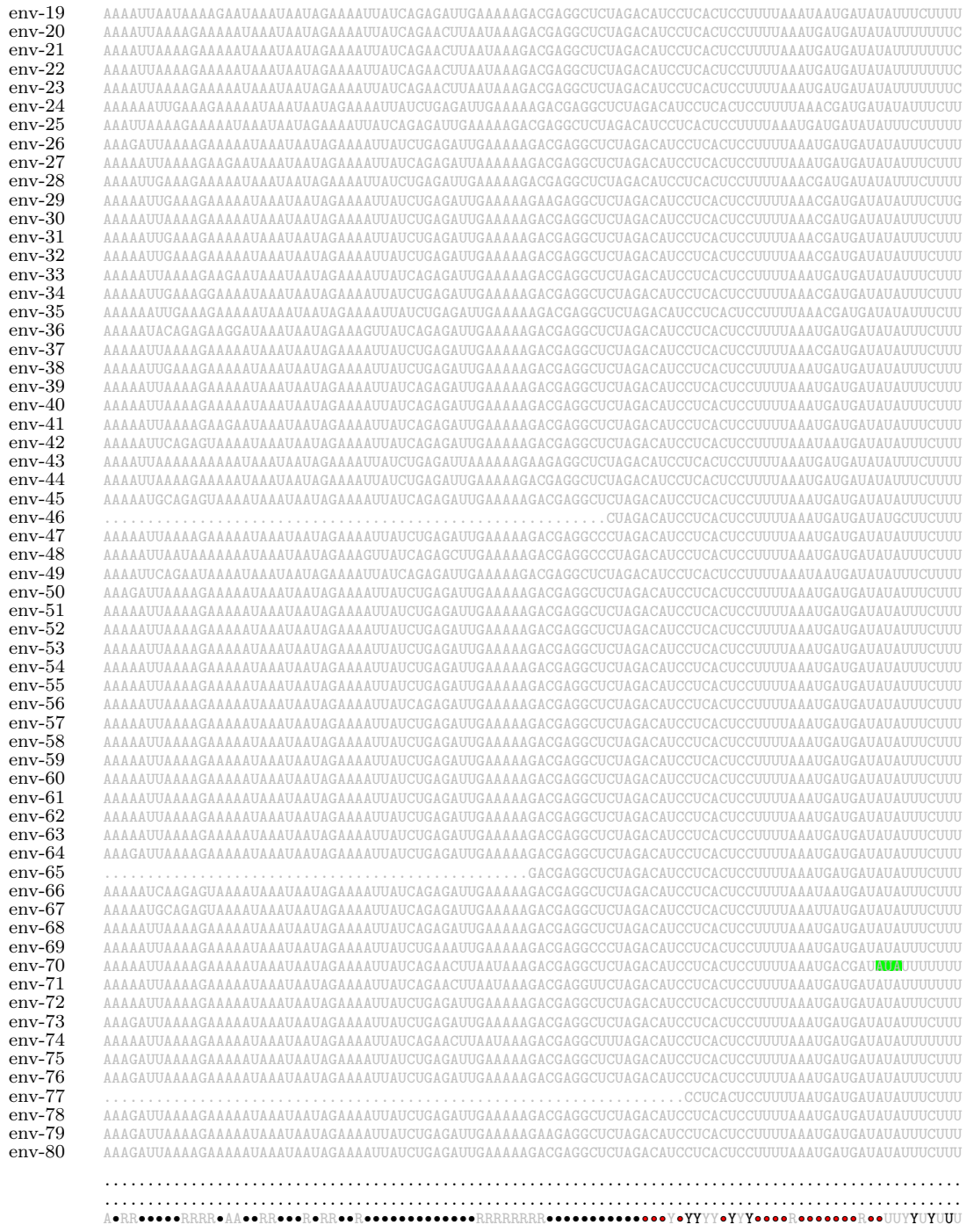
Conserved domains found in protein-coding genes listed in Section 68.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *potC* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

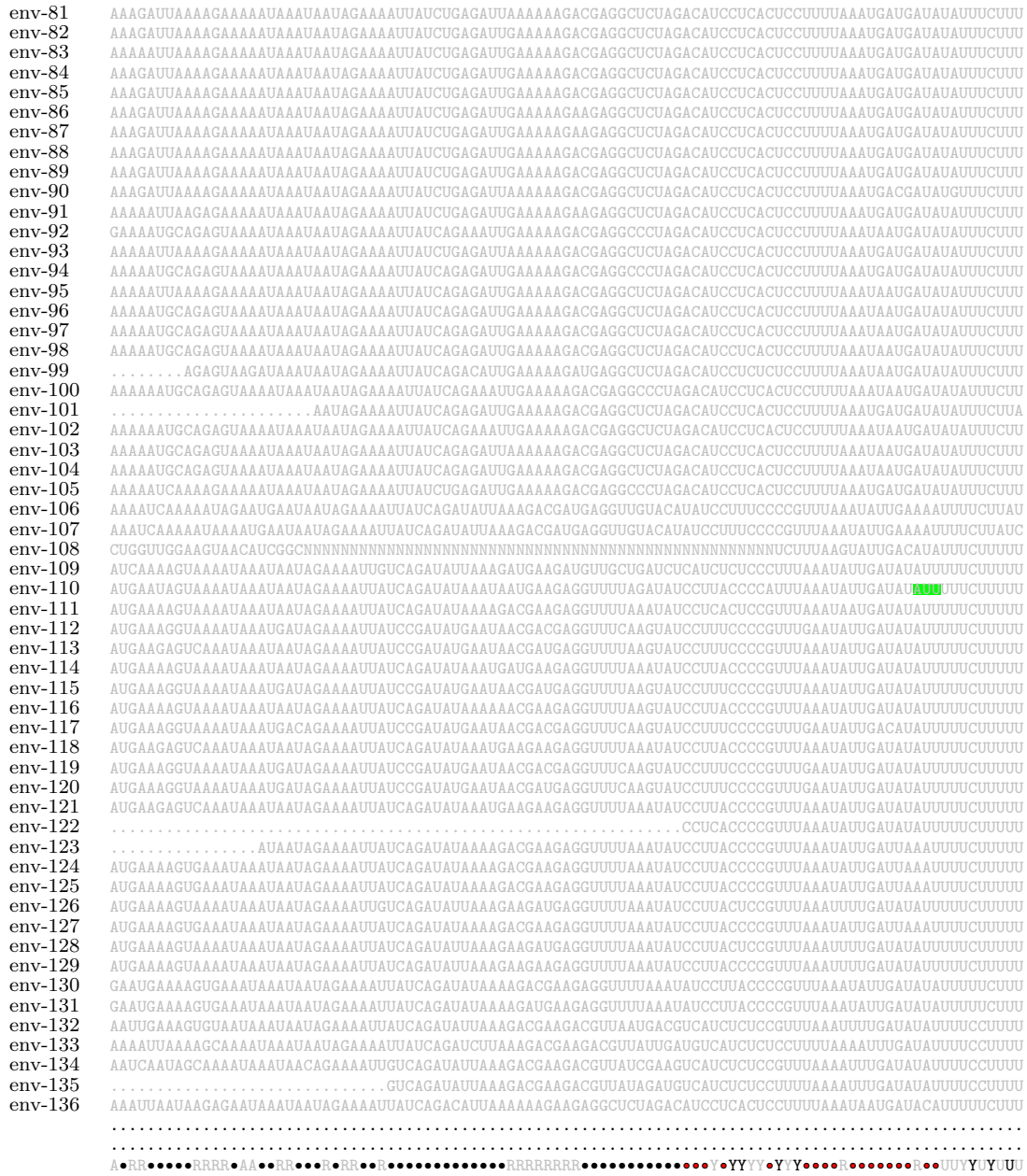
COG0395 (1) ABC-type sugar transport system, permease component [Carbohydrate transport and metabolism]
COG0555 (18) ABC-type sulfate transport system, permease component [Posttranslational modification, protein turnover, chaperones]
COG0678 (75) Peroxiredoxin [Posttranslational modification, protein turnover, chaperones]
COG1177 (8) ABC-type spermidine/putrescine transport system, permease component II [Amino acid transport and metabolism]
COG1178 (22) ABC-type Fe3+ transport system, permease component [Inorganic ion transport and metabolism]
COG3842 (3) ABC-type spermidine/putrescine transport systems, ATPase components [Amino acid transport and metabolism]
GOS1374 (74) ?
GOS2568 (3) ?
GOS7685 (79) ?

68.4 Multiple-sequence alignment

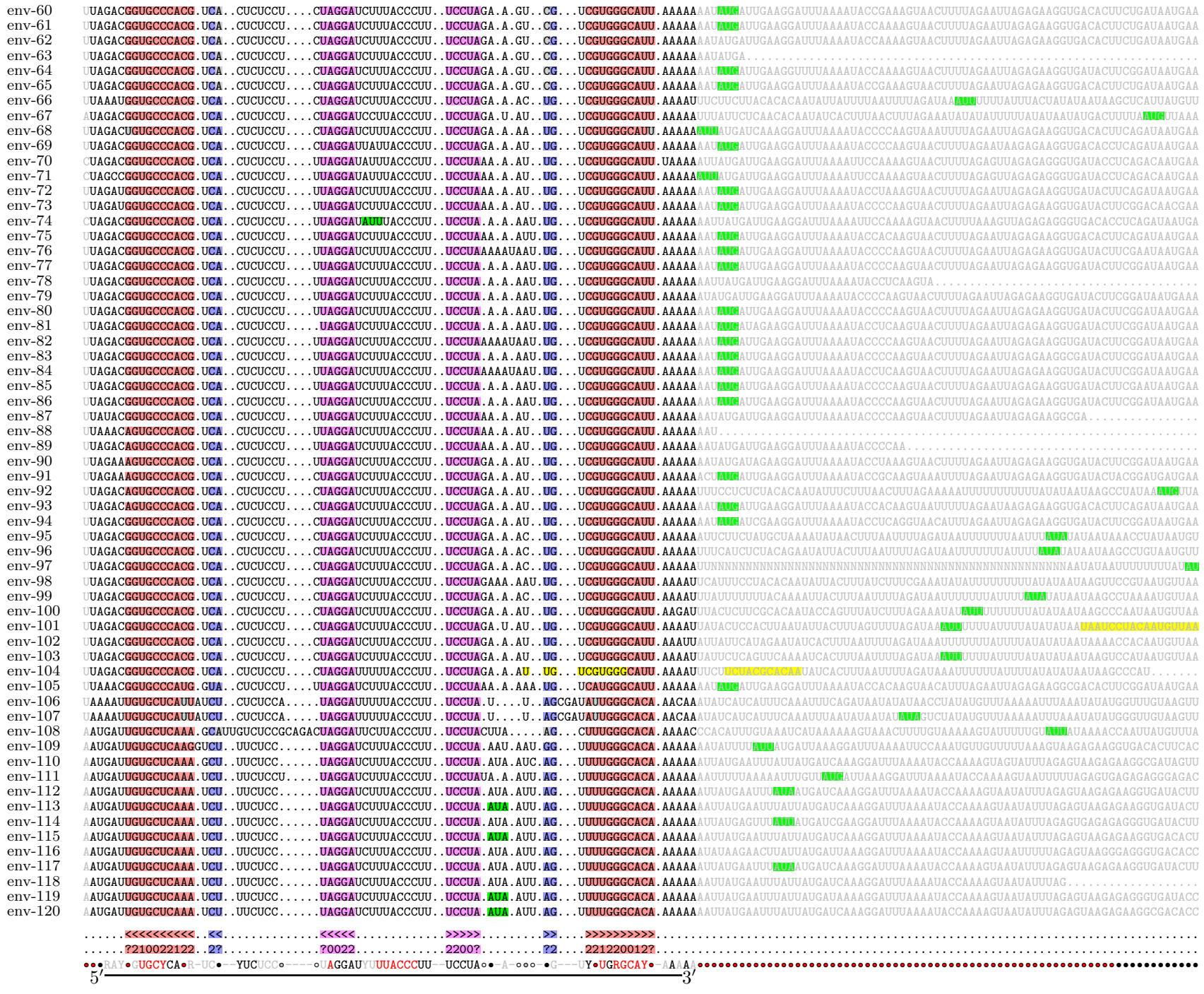
Each *potC* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 68.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *potC* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

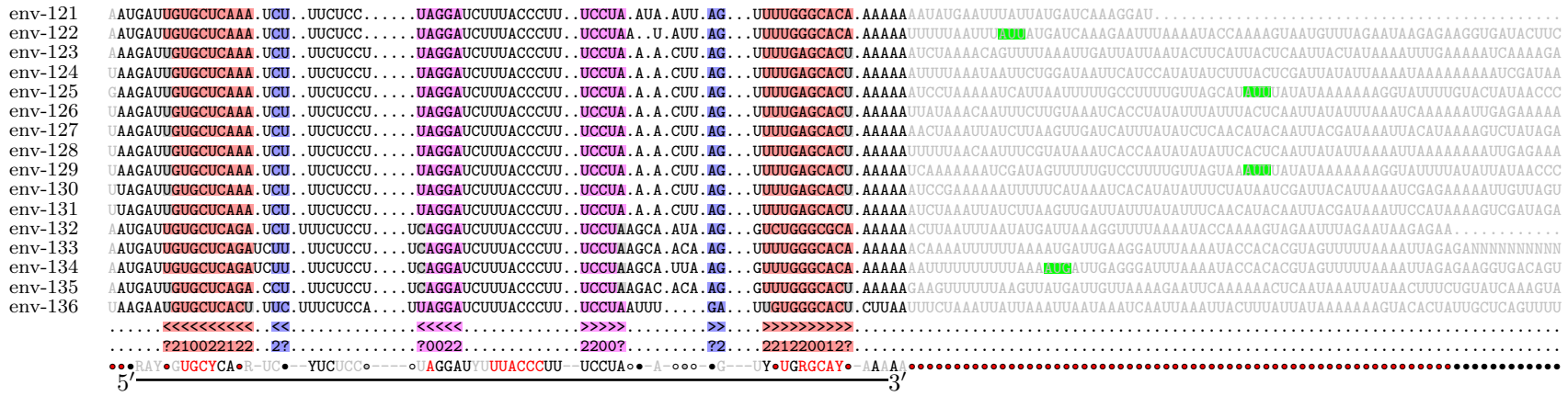



env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80


env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-135
env-136

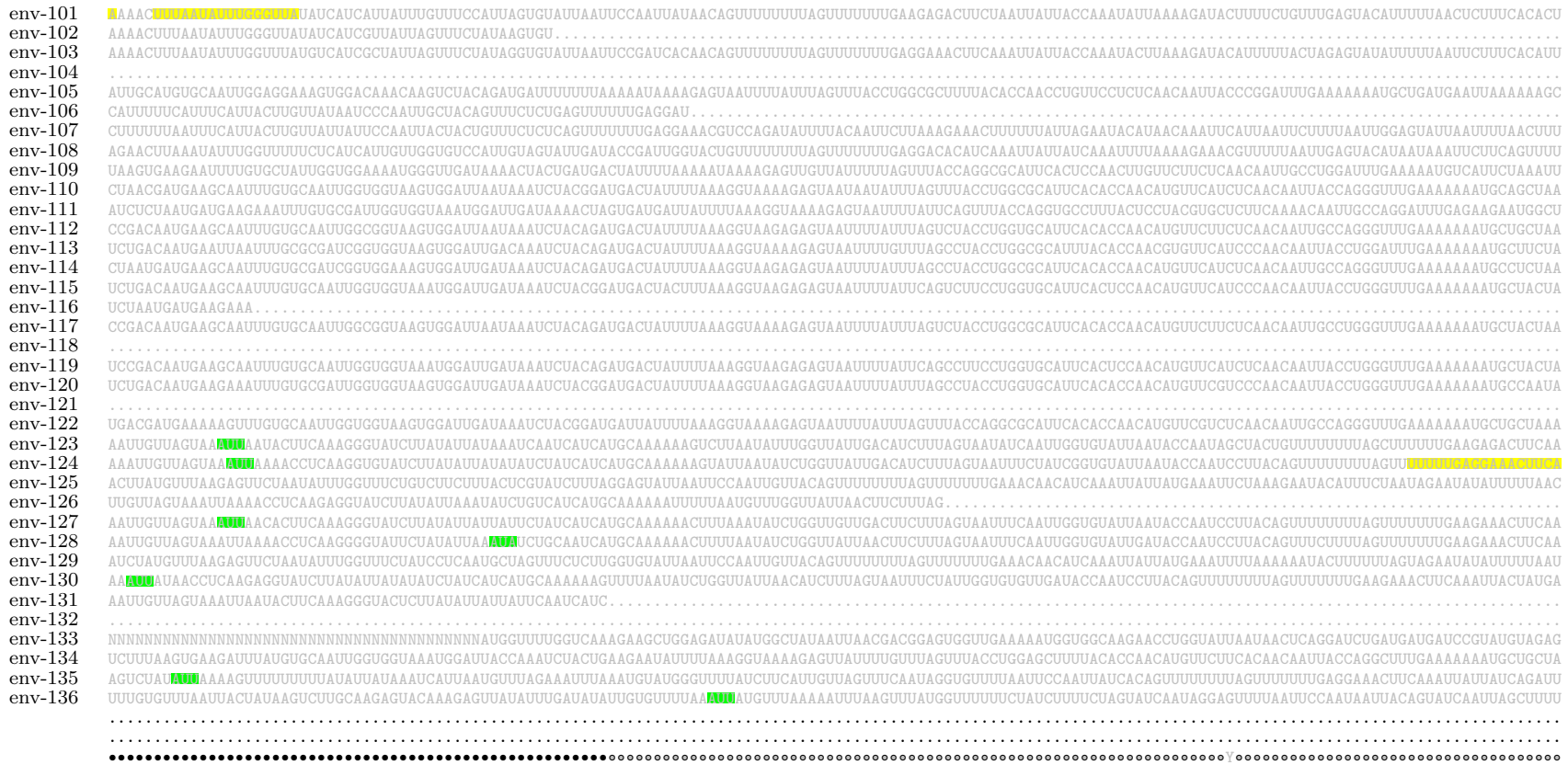
alignment positions 100 · · · 259


env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
5′
3′

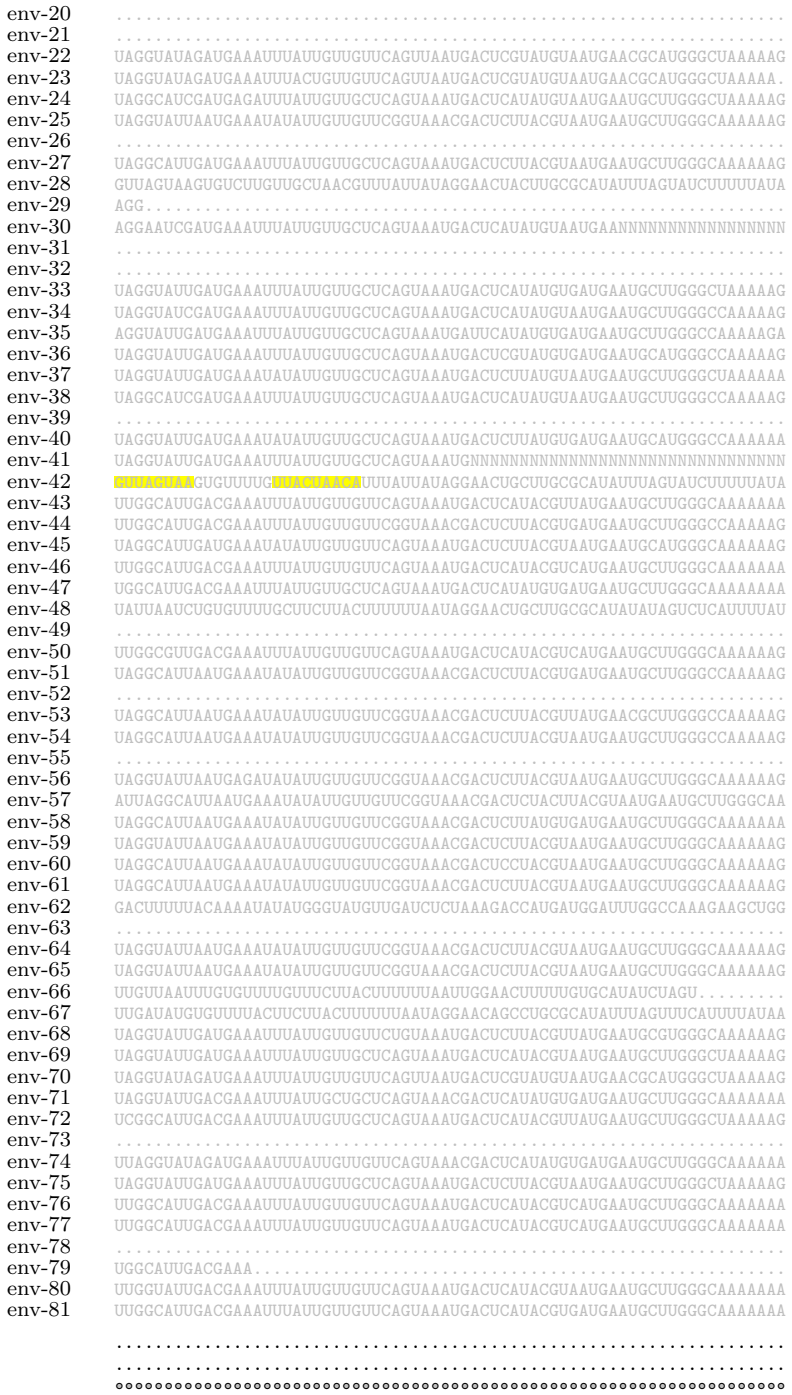


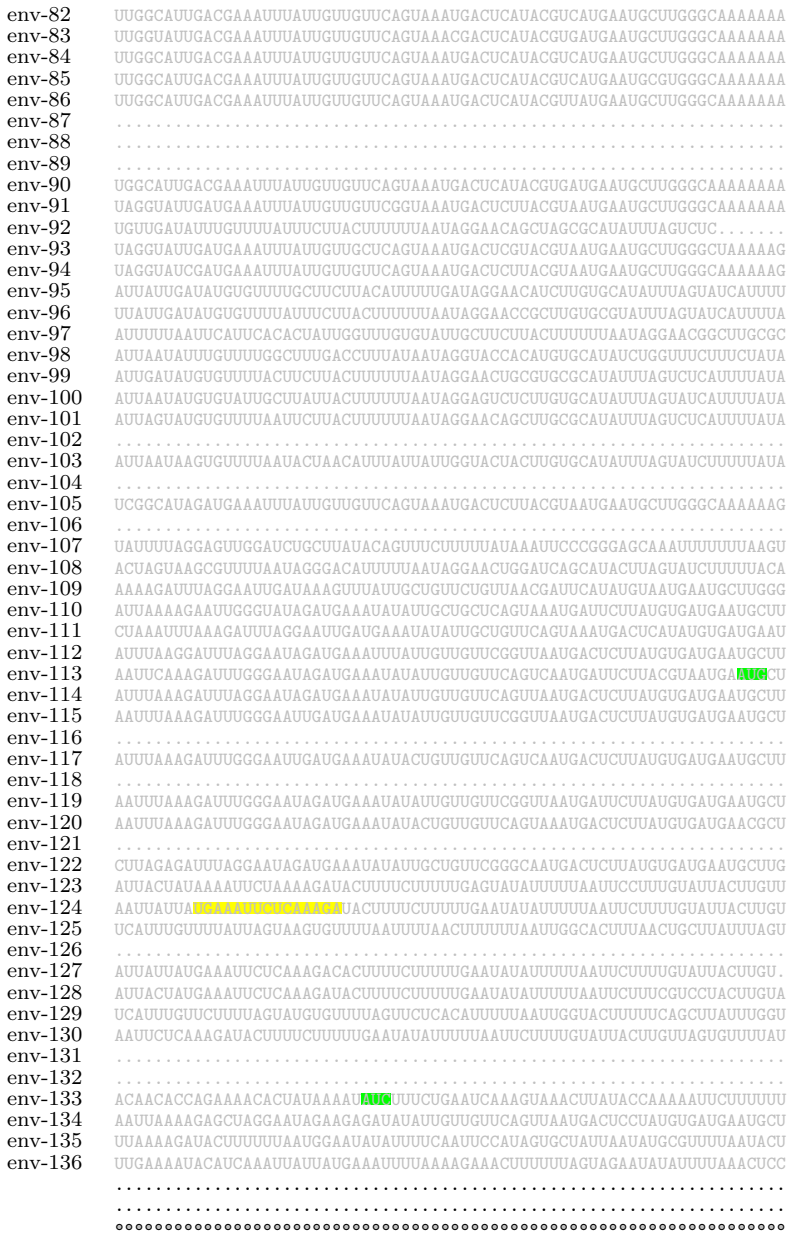
alignment positions 260 · · · 419

env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100



alignment positions 420 · · · 487


env-20
env-21
env-22 UAGGUAUAGAUGAAAUUUAUUGUUGUUCAGUUAAUGACUCGUAUGUAAUGAACGCAUGGGCUAAAAAG
env-23 UAGGUAUAGAUGAAAUUUACUGUUGUUCAGUUAAUGACUCGUAUGUAAUGAACGCAUGGGCUAAAAA.
env-24 UAGGCAUCGAUGAGAUUUAUUGUUGCUCAGUAAAUGACUCAUAUGUAAUGAAUGCUUGGGCUAAAAAG
env-25 UAGGUAUUAAUGAAAUAUAUUGUUGUUCGGUAAACGACUCUUACGUAAUGAAUGCUUGGGCAAAAAAG
env-26
env-27 UAGGCAUUGAUGAAAUUUAUUGUUGCUCAGUAAAUGACUCUUACGUAAUGAAUGCUUGGGCAAAAAAG
env-28 GUUAGUAAGUGUCUUGUUGCUAACGUUUAUUAUAGGAACUACUUGCGCAUAUUUAGUAUCUUUUUAUA
env-29 AGG.....
env-30 AGGAAUCGAUGAAAUUUAUUGUUGCUCAGUAAAUGACUCAUAUGUAAUGAANNNNNNNNNNNNNNNN
env-31
env-32
env-33 UAGGUAUUGAUGAAAUUUAUUGUUGCUCAGUAAAUGACUCAUAUGUGAUGAAUGCUUGGGCUAAAAAG
env-34 UAGGUAUCGAUGAAAUUUAUUGUUGCUCAGUAAAUGACUCAUAUGUAAUGAAUGCUUGGGCCAAAAAG
env-35 AGGUAUUGAUGAAAUUUAUUGUUGCUCAGUAAAUGAUUCAUAUGUGAUGAAUGCUUGGGCCAAAAAGA
env-36 UAGGUAUUGAUGAAAUUUAUUGUUGCUCAGUAAAUGACUCGUAUGUGAUGAAUGCAUGGGCCAAAAAG
env-37 UAGGUAUUGAUGAAAUAUAUUGUUGCUCAGUAAAUGACUCUUAUGUAAUGAAUGCUUGGGCUAAAAAA
env-38 UAGGCAUCGAUGAAAUUUAUUGUUGCUCAGUAAAUGACUCAUAUGUAAUGAAUGCUUGGGCCAAAAAG
env-39
env-40 UAGGUAUUGAUGAAAUAUAUUGUUGCUCAGUAAAUGACUCUUAUGUGAUGAAUGCAUGGGCCAAAAAA
env-41 UAGGUAUUGAUGAAAUUUAUUGUUGCUCAGUAAAUGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
env-42 CUUAGUAAAGUGUUUUGGUACUAACAUUUAUUAUAGGAACUGCUUGCGCAUAUUUAGUAUCUUUUUAUA
env-43 UUGGCAUUGACGAAAUUUAUUGUUGUUCAGUAAAUGACUCAUACGUUAUGAAUGCUUGGGCAAAAAAA
env-44 UUGGCAUUGACGAAAUUUAUUGUUGUUCGGUAAACGACUCUUACGUGAUGAAUGCUUGGGCCAAAAAG
env-45 UAGGCAUUGAUGAAAUAUAUUGUUGUUCAGUAAAUGACUCUUACGUAAUGAAUGCAUGGGCAAAAAAG
env-46 UUGGCAUUGACGAAAUUUAUUGUUGUUCAGUAAAUGACUCAUACGUCAUGAAUGCUUGGGCAAAAAAA
env-47 UGGCAUUGACGAAAUUUAUUGUUGCUCAGUAAAUGACUCAUAUGUGAUGAAUGCUUGGGCAAAAAAAA
env-48 UAUUAAUCUGUGUUUUGCUUCUUACUUUUUUAAUAGGAACUGCUUGCGCAUAUAUAGUCUCAUUUUAU
env-49
env-50 UUGGCGUUGACGAAAUUUAUUGUUGUUCAGUAAAUGACUCAUACGUCAUGAAUGCUUGGGCAAAAAAG
env-51 UAGGCAUUAAUGAAAUAUAUUGUUGUUCGGUAAACGACUCUUACGUGAUGAAUGCUUGGGCCAAAAAG
env-52
env-53 UAGGCAUUAAUGAAAUAUAUUGUUGUUCGGUAAACGACUCUUACGUUAUGAACGCUUGGGCCAAAAAG
env-54 UAGGCAUUAAUGAAAUAUAUUGUUGUUCGGUAAACGACUCUUACGUAAUGAAUGCUUGGGCCAAAAAG
env-55
env-56 UAGGUAUUAAUGAGAUAUAUUGUUGUUCGGUAAACGACUCUUACGUAAUGAAUGCUUGGGCAAAAAAG
env-57 AUUAGGCAUUAAUGAAAUAUAUUGUUGUUCGGUAAACGACUCUACUUACGUAAUGAAUGCUUGGGCAA
env-58 UAGGCAUUAAUGAAAUAUAUUGUUGUUCGGUAAACGACUCUUAUGUGAUGAAUGCUUGGGCAAAAAAA
env-59 UAGGUAUUAAUGAAAUAUAUUGUUGUUCGGUAAACGACUCUUACGUAAUGAAUGCUUGGGCAAAAAAG
env-60 UAGGCAUUAAUGAAAUAUAUUGUUGUUCGGUAAACGACUCCUACGUAAUGAAUGCUUGGGCAAAAAAG
env-61 UAGGCAUUAAUGAAAUAUAUUGUUGUUCGGUAAACGACUCUUACGUAAUGAAUGCUUGGGCAAAAAAG
env-62 GACUUUUUACAAAAUAUAUGGGUAUGUUGAUCUCUAAAGACCAUGAUGGAUUUGGCCAAAGAAGCUGG
env-63
env-64 UAGGUAUUAAUGAAAUAUAUUGUUGUUCGGUAAACGACUCUUACGUAAUGAAUGCUUGGGCAAAAAAG
env-65 UAGGUAUUAAUGAAAUAUAUUGUUGUUCGGUAAACGACUCUUACGUAAUGAAUGCUUGGGCAAAAAAG
env-66 UUGUUAAUUUGUGUUUUGUUUCUUACUUUUUUAAUUGGAACUUUUUGUGCAUAUCUAGU.....
env-67 UUGAUAUGUGUUUUACUUCUUACUUUUUUAAUAGGAACAGCCUGCGCAUAUUUAGUUUCAUUUUAUAA
env-68 UAGGUAUUGAUGAAAUUUAUUGUUGUUCUGUAAAUGACUCUUACGUUAUGAAUGCGUGGGCAAAAAAG
env-69 UAGGUAUUGAUGAAAUUUAUUGUUGCUCAGUAAAUGACUCAUACGUAAUGAAUGCUUGGGCUAAAAAG
env-70 UAGGUAUAGAUGAAAUUUAUUGUUGUUCAGUUAAUGACUCGUAUGUAAUGAACGCAUGGGCUAAAAAG
env-71 UAGGUAUUGACGAAAUUUAUUGCUGCUCAGUAAACGACUCAUAUGUGAUGAAUGCUUGGGCAAAAAAA
env-72 UCGGCAUUGACGAAAUUUAUUGUUGCUCAGUAAAUGACUCAUACGUUAUGAAUGCUUGGGCUAAAAAG
env-73
env-74 UUAGGUAUAGAUGAAAUUUAUUGUUGUUCAGUAAACGACUCAUAUGUGAUGAAUGCUUGGGCAAAAAA
env-75 UAGGUAUUGAUGAAAUUUAUUGUUGCUCAGUAAAUGACUCUUACGUAAUGAAUGCUUGGGCUAAAAAG
env-76 UUGGCAUUGACGAAAUUUAUUGUUGUUCAGUAAAUGACUCAUACGUCAUGAAUGCUUGGGCAAAAAAA
env-77 UUGGCAUUGACGAAAUUUAUUGUUGUUCAGUAAAUGACUCAUACGUCAUGAAUGCUUGGGCAAAAAAA
env-78
env-79 UGGCAUUGACGAAA.....
env-80 UUGGUAUUGACGAAAUUUAUUGUUGUUCAGUAAAUGACUCAUACGUAAUGAAUGCUUGGGCAAAAAAA
env-81 UUGGCAUUGACGAAAUUUAUUGUUGUUCAGUAAAUGACUCAUACGUGAUGAAUGCUUGGGCAAAAAAA


env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-135
env-136

69 *psaA*

69.1 Taxa

The taxonomy of each organism containing a putative *psaA* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *psaA* RNA in Sections 69.2 and 69.4.

abbrev. of hits	taxonomy of species
Ssp-1-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* BL107
Ssp-2-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* CC9311
Ssp-3-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* CC9605
Ssp-4-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* CC9902
Ssp-5-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* RS9916
Ssp-6-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* RS9917
Ssp-7-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 5701
Ssp-8-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 7803
Ssp-9-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 7805
Ssp-10-1	Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 8102
Pma-1-1	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9303
Pma-2-1	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9313
env-1 to env-19	environmental samples

69.2 Gene contexts

Each *psaA* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *psaA* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 69.3. The accession number of the sequence containing each *psaA* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 69.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Ssp-7-1	NZ_AANO01000001.1	+	318481	318691	**RNA**→ PsaA_PsaB (pfam00223)→ PsaA_PsaB (pfam00223)→
Pma-2-1	NC_005071.1	-	1864565	1864369	**RNA**→ PsaA_PsaB (pfam00223)NOG04762 (NOG04762)→ PsaA_PsaB (pfam00223)NOG04763 (NOG04763)→
Pma-1-1	NC_008820.1	-	2066948	2066752	**RNA**→ PsaA_PsaB (pfam00223)→ PsaA_PsaB (pfam00223)→
env-1	JCVI_SCAF_1101668672762	+	787	979	**RNA**→ ? (GOS2689)→
Ssp-2-1	NC_008319.1	+	402382	402566	**RNA**→ PsaA_PsaB (pfam00223)→ PsaA_PsaB (pfam00223)→
env-2	JCVI_SCAF_1101667380532	-	620	429	**RNA**→ ? (GOS2689)→
env-3	JCVI_SCAF_1096627759113	+	274	465	**RNA**→ ? (GOS2689)→ ? (GOS2689)→
Ssp-1-1	NZ_AATZ01000002.1	+	703831	704023	**RNA**→ PsaA_PsaB (pfam00223)→ PsaA_PsaB (pfam00223)→
Ssp-4-1	NC_007513.1	-	1900065	1899872	**RNA**→ PsaA_PsaB (pfam00223)NOG04762 (NOG04762)→ PsaA_PsaB (pfam00223)NOG04763 (NOG04763)→
env-4	JCVI_SCAF_1101668660072	+	182	374	**RNA**→
env-5	JCVI_SCAF_1101668664310	-	244	53	**RNA**→
env-6	JCVI_SCAF_1101668663170	-	413	222	**RNA**→
env-7	JCVI_SCAF_1096627369725	-	1376	1184	**RNA**→ PsaA_PsaB (pfam00223)? (GOS2689)→
env-8	JCVI_SCAF_1096627967060	+	66	257	**RNA**→

env-9	JCVI_SCAF_1096627322457	-	1015	823	**RNA**→ ? (GOS2689)→
Ssp-3-1	NC_007516.1	+	340021	340213	**RNA**→ PsaA_PsaB (pfam00223)NOG04762 (NOG04762)→ PsaA_PsaB (pfam00223)NOG04763 (NOG04763)→ hypo→
env-10	JCVI_SCAF_1096626105447	+	165	356	**RNA**→
env-11	JCVI_SCAF_1096627967065	+	629	820	**RNA**→
env-12	JCVI_SCAF_1096628050846	+	55	247	**RNA**→
Ssp-10-1	NC_005070.1	-	2010986	2010794	**RNA**→ PsaA_PsaB (pfam00223)NOG04762 (NOG04762)→ PsaA_PsaB (pfam00223)NOG04763 (NOG04763)→
env-13	JCVI_SCAF_1096626106119	-	660	468	**RNA**→ ? (GOS2689)→
env-14	JCVI_SCAF_1096627278248	+	941	1133	**RNA**→ PsaA_PsaB (pfam00223)? (GOS2689)→
env-15	JCVI_SCAF_1096627080268	-	6072	5880	**RNA**→ PsaA_PsaB (pfam00223)? (GOS2689)→ PsaA_PsaB (pfam00223)? (GOS2689)→
Ssp-9-1	NZ_AAOK01000002.1	-	133205	133012	**RNA**→ PsaA_PsaB (pfam00223)→ PsaA_PsaB (pfam00223)→
Ssp-8-1	NC_009481.1	+	414736	414928	**RNA**→ hypo→ PsaA_PsaB (pfam00223)→ PsaA_PsaB (pfam00223)→
Ssp-5-1	NZ_AAUA01000004.1	+	189443	189638	**RNA**→ PsaA_PsaB (pfam00223)→ PsaA_PsaB (pfam00223)→
Ssp-6-1	NZ_AANP01000005.1	-	107675	107479	**RNA**→ PsaA_PsaB (pfam00223)→ PsaA_PsaB (pfam00223)→
env-16	JCVI_SCAF_1101668607881	-	428	232	**RNA**→
env-17	JCVI_SCAF_1096627368091	+	694	890	**RNA**→ PsaA_PsaB (pfam00223)? (GOS2689)→ PsaA_PsaB (pfam00223)? (GOS2689)→ ? (GOS7128)→
env-18	JCVI_SCAF_1096626596788	-	576	384	**RNA**→
env-19	JCVI_SCAF_1096626471076	+	140	333	**RNA**→

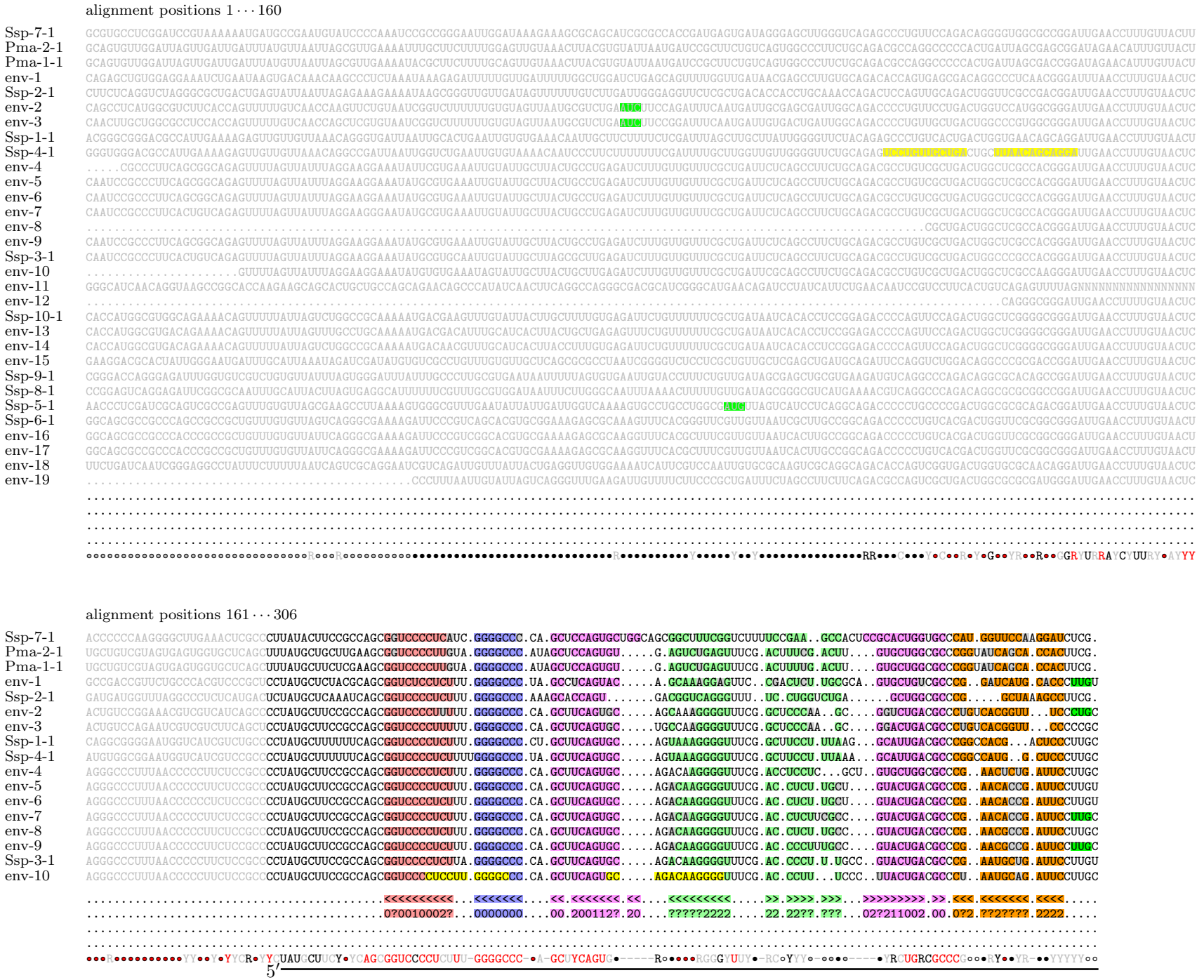
69.3 Conserved domains

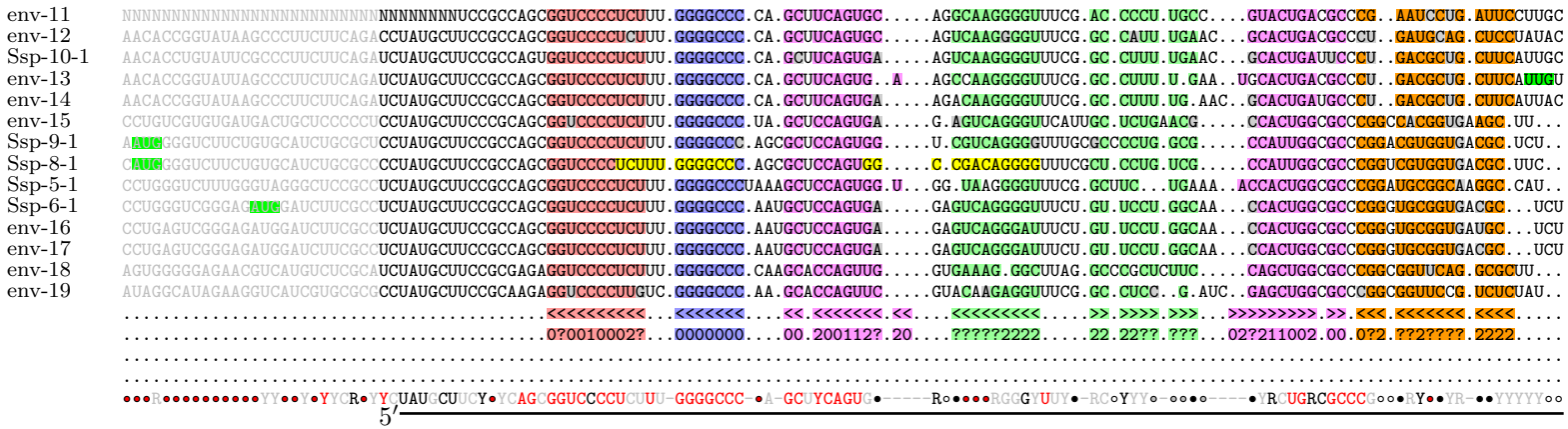
Conserved domains found in protein-coding genes listed in Section 69.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *psaA* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

GOS2689 (12) ?
GOS7128 (1) ?
NOG04762 (4)
NOG04763 (4)
pfam00223 (30) Photosystem I psaA/psaB protein.

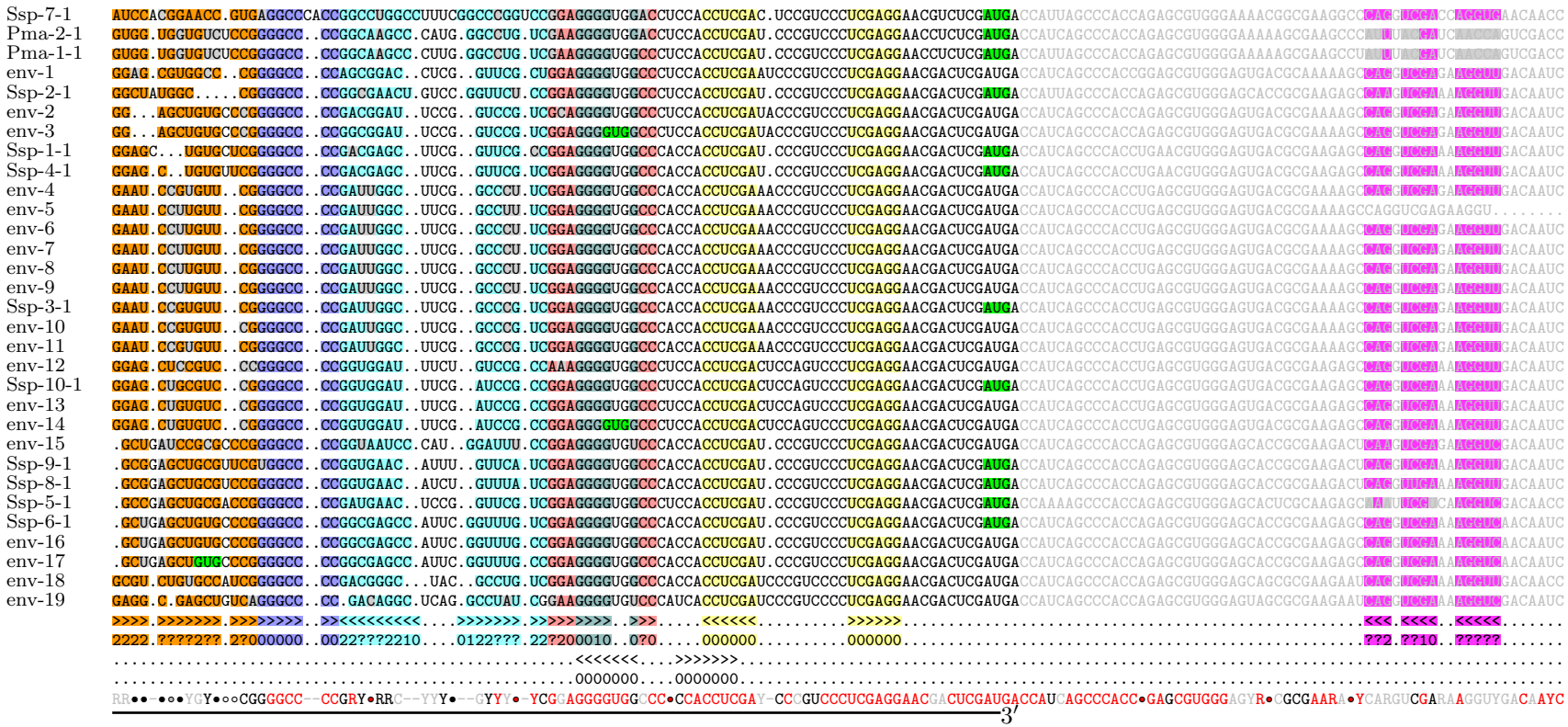
69.4 Multiple-sequence alignment

Each *psaA* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 69.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *psaA* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

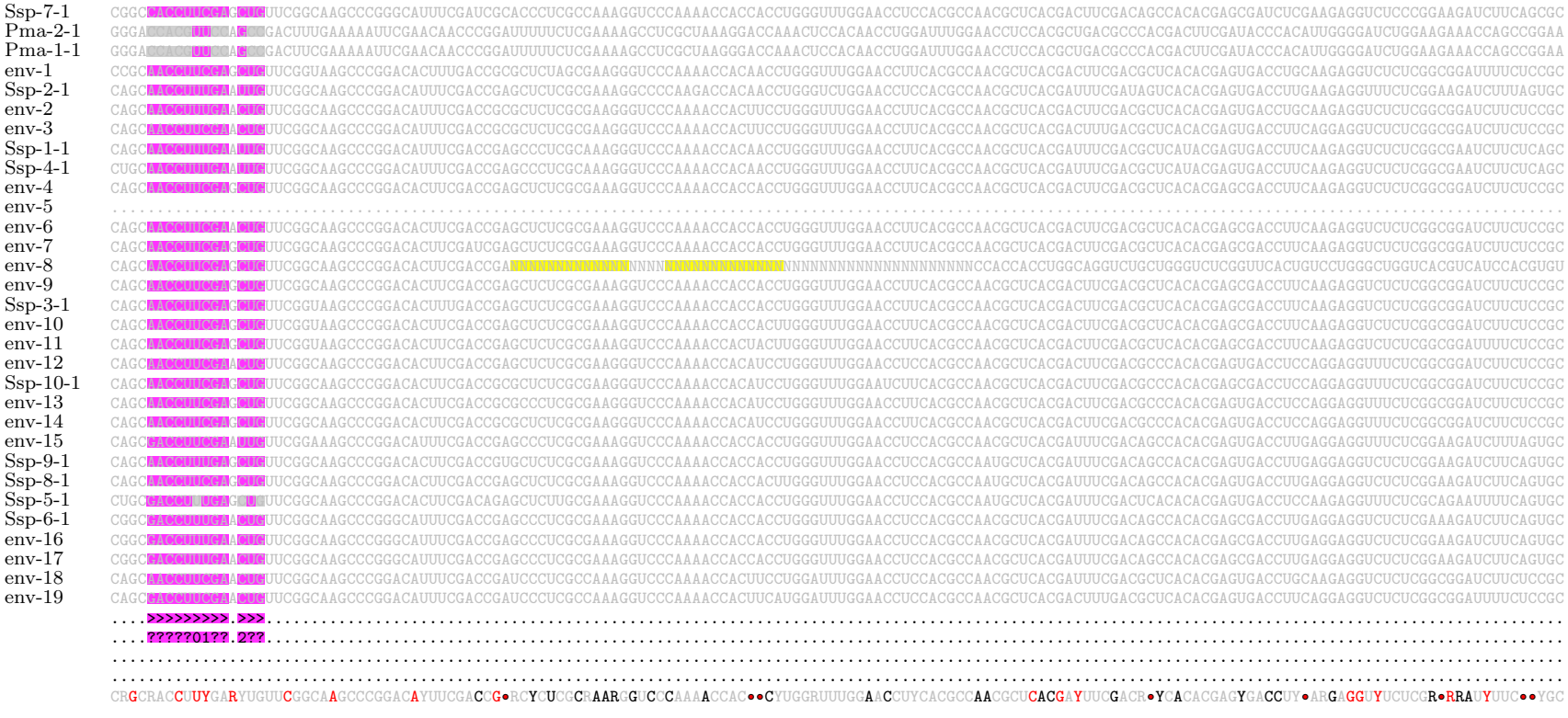

alignment positions 1 · · · 160
Ssp-7-1
Pma-2-1
Pma-1-1
env-1
Ssp-2-1
env-2
env-3
Ssp-1-1
Ssp-4-1
env-4
env-5
env-6
env-7
env-8
env-9
Ssp-3-1
env-10
env-11
env-12
Ssp-10-1
env-13
env-14
env-15
Ssp-9-1
Ssp-8-1
Ssp-5-1
Ssp-6-1
env-16
env-17
env-18
env-19
alignment positions 161 · · · 306
Ssp-7-1
Pma-2-1
Pma-1-1
env-1
Ssp-2-1
env-2
env-3
Ssp-1-1
Ssp-4-1
env-4
env-5
env-6
env-7
env-8
env-9
Ssp-3-1
env-10
5′



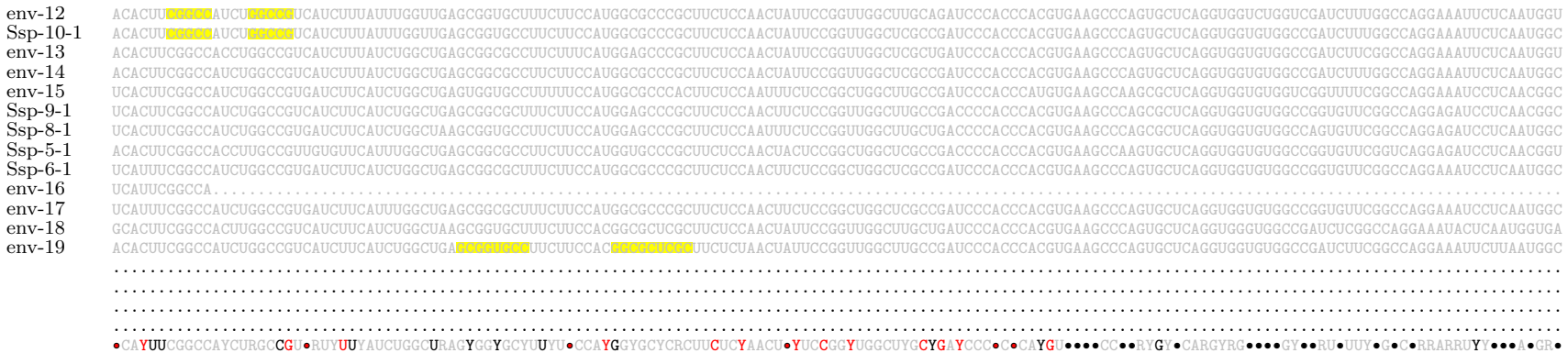
alignment positions 307 · · · 466



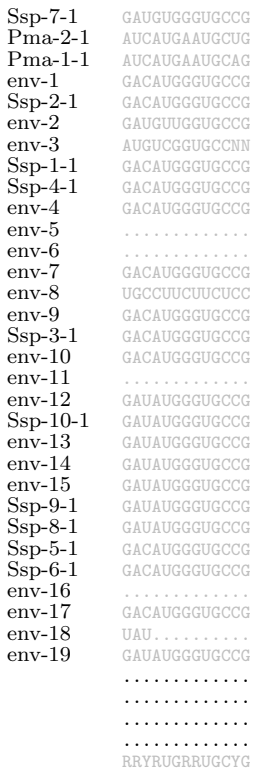
alignment positions 467 · · · 626



alignment positions 627 · · · 786



alignment positions 787 · · · 799



70 *psbNH*

70.1 Taxa

The taxonomy of each organism containing a putative *psbNH* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *psbNH* RNA in Sections 70.2 and 70.4.

abbrev.	of hits	taxonomy of species
Ssp-1-1		Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* BL107
Ssp-2-1		Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* CC9311
Ssp-3-1		Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* CC9605
Ssp-4-1		Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* CC9902
Ssp-5-1		Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* RS9916
Ssp-6-1		Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* RS9917
Ssp-7-1		Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 7803
Ssp-8-1		Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 7805
Ssp-9-1		Bacteria Cyanobacteria Chroococcales *Synechococcus sp.* WH 8102
Pma-1-1		Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9303
Pma-2-1		Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9313
env-1 to env-28		environmental samples

70.2 Gene contexts

Each *psbNH* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *psbNH* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 70.3. The accession number of the sequence containing each *psbNH* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 70.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Pma-2-1	NC_005071.1	+	1943364	1943422	←PsbH (pfam00737) **RNA**→ PsbN (pfam02468)NOG21533 (NOG21533)→
Pma-1-1	NC_008820.1	+	2161522	2161580	←PsbH (pfam00737) **RNA**→ PsbN (pfam02468)→
Ssp-8-1	NZ_AAOK01000002.1	+	214718	214775	←PsbH (pfam00737) **RNA**→ PsbN (pfam02468)→
Ssp-7-1	NC_009481.1	-	340673	340616	←PsbH (pfam00737) **RNA**→ PsbN (pfam02468)→
env-1	JCVI_SCAF_1096627066797	+	210	268	←PsbH (pfam00737) **RNA**→ ←? (GOS2285)
env-2	JCVI_SCAF_1096627081534	+	1261	1319	←PsbH (pfam00737) **RNA**→ ←? (GOS2285)
env-3	JCVI_SCAF_1101668667375	-	1426	1368	**RNA**→ ←? (GOS2285)
env-4	JCVI_SCAF_1096627073860	-	1256	1198	**RNA**→ ←? (GOS2285)
env-5	JCVI_SCAF_1096628013422	+	482	540	←PsbH (pfam00737) **RNA**→ ←? (GOS2285)
env-6	JCVI_SCAF_1096627950878	+	1284	1341	←PsbH (pfam00737)? (GOS13170996) **RNA**→ PsbN (pfam02468)→
Ssp-6-1	NZ_AANP01000005.1	+	188511	188568	←PsbH (pfam00737) **RNA**→ PsbN (pfam02468)→
env-7	JCVI_SCAF_1096626388708	-	76	19	←PsbH (pfam00737)? (GOS13170996) **RNA**→
env-8	JCVI_SCAF_1096627356520	-	1029	972	←PsbH (pfam00737)? (GOS13170996) **RNA**→ PsbN (pfam02468)→
env-9	JCVI_SCAF_1096627359014	+	2012	2069	←PsbH (pfam00737)? (GOS13170996) **RNA**→ PsbN (pfam02468)→
env-10	JCVI_SCAF_1096627360792	-	4804	4747	←PsbH (pfam00737)? (GOS13170996) **RNA**→ PsbN (pfam02468)→

env-11	JCVI_SCAF_1096627949840	+	689	746	←PsbH (pfam00737)? (GOS13170996) **RNA**→ ←? (GOS13442019)
Ssp-5-1	NZ_AAUA01000004.1	-	95600	95543	←PsbH (pfam00737) **RNA**→ photosystem II reaction center protein PsbH→
Ssp-2-1	NC_008319.1	-	322025	321967	←PsbH (pfam00737) **RNA**→ PsbN (pfam02468)→
Ssp-1-1	NZ_AATZ01000002.1	-	616384	616327	←PsbH (pfam00737) **RNA**→ ←photosystem II reaction center protein PsbH
Ssp-4-1	NC_007513.1	+	1985986	1986043	←PsbH (pfam00737)NOG16382 (NOG16382) **RNA**→ PsbN (pfam02468)NOG21533 (NOG21533)→
env-12	JCVI_SCAF_1096627155346	-	1394	1337	←PsbH (pfam00737)? (GOS17710) **RNA**→ PsbN (pfam02468)→
Ssp-9-1	NC_005070.1	-	271369	271312	←PsbH (pfam00737)NOG16382 (NOG16382) **RNA**→ PsbN (pfam02468)NOG21533 (NOG21533)→
env-13	JCVI_SCAF_1101667407209	-	763	706	**RNA**→ PsbN (pfam02468)→
env-14	JCVI_SCAF_1096626959936	+	1236	1293	←PsbH (pfam00737) **RNA**→ PsbN (pfam02468)→
env-15	JCVI_SCAF_1101667591879	+	275	332	←PsbH (pfam00737) **RNA**→ ←? (GOS2285)
env-16	JCVI_SCAF_1096627750156	+	186	243	←PsbH (pfam00737)? (GOS13170996) **RNA**→ ←? (GOS2285)
env-17	JCVI_SCAF_1096627371787	+	1564	1621	←PsbH (pfam00737)? (GOS13170996) **RNA**→ ←? (GOS2285)
Ssp-3-1	NC_007516.1	-	272727	272670	←PsbH (pfam00737)NOG16382 (NOG16382) **RNA**→ PsbN (pfam02468)→
env-18	JCVI_SCAF_1096627751564	+	380	437	←PsbH (pfam00737)? (GOS13170996) **RNA**→ ←? (GOS2285)
env-19	JCVI_SCAF_1096627370986	-	2468	2411	**RNA**→ ←Glyco_hydro_38 (pfam01074)? (GOS2285)
env-20	JCVI_SCAF_1101668657376	-	642	585	←PsbH (pfam00737)? (GOS13170996) **RNA**→ ←? (GOS2285)
env-21	JCVI_SCAF_1096627612675	+	466	523	←PsbH (pfam00737)? (GOS13170996) **RNA**→ ←? (GOS2285)
env-22	JCVI_SCAF_1096627369297	+	743	800	←PsbH (pfam00737)? (GOS13170996) **RNA**→ ←? (GOS2285)
env-23	JCVI_SCAF_1096627772150	-	261	204	←PsbH (pfam00737)? (GOS13170996) **RNA**→
env-24	JCVI_SCAF_1096627507000	-	1446	1389	←PsbH (pfam00737)? (GOS13170996) **RNA**→ ←? (GOS2285)
env-25	JCVI_SCAF_1096627371491	+	1394	1451	←PsbH (pfam00737)? (GOS13170996) **RNA**→
env-26	JCVI_SCAF_1101668231100	-	759	702	←PsbH (pfam00737)? (GOS13170996) **RNA**→ ←? (GOS2285)
env-27	JCVI_SCAF_1096628059629	-	1169	1112	←PsbH (pfam00737)? (GOS17710) **RNA**→ ? (GOS12019938)→
env-28	JCVI_SCAF_1096626726490	+	266	323	←PsbH (pfam00737)? (GOS17710) **RNA**→ PsbN (pfam02468)→

70.3 Conserved domains

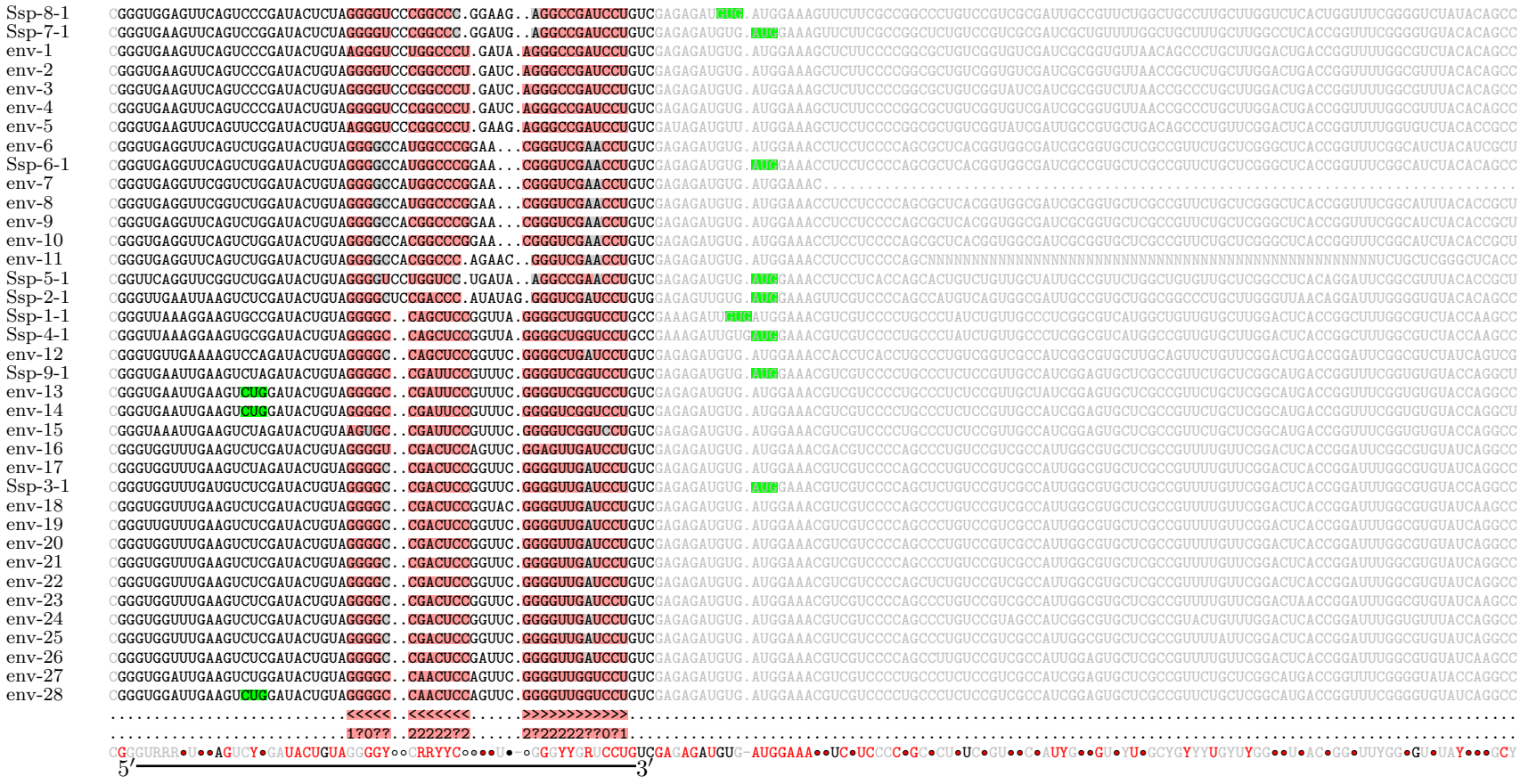
Conserved domains found in protein-coding genes listed in Section 70.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *psbNH* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

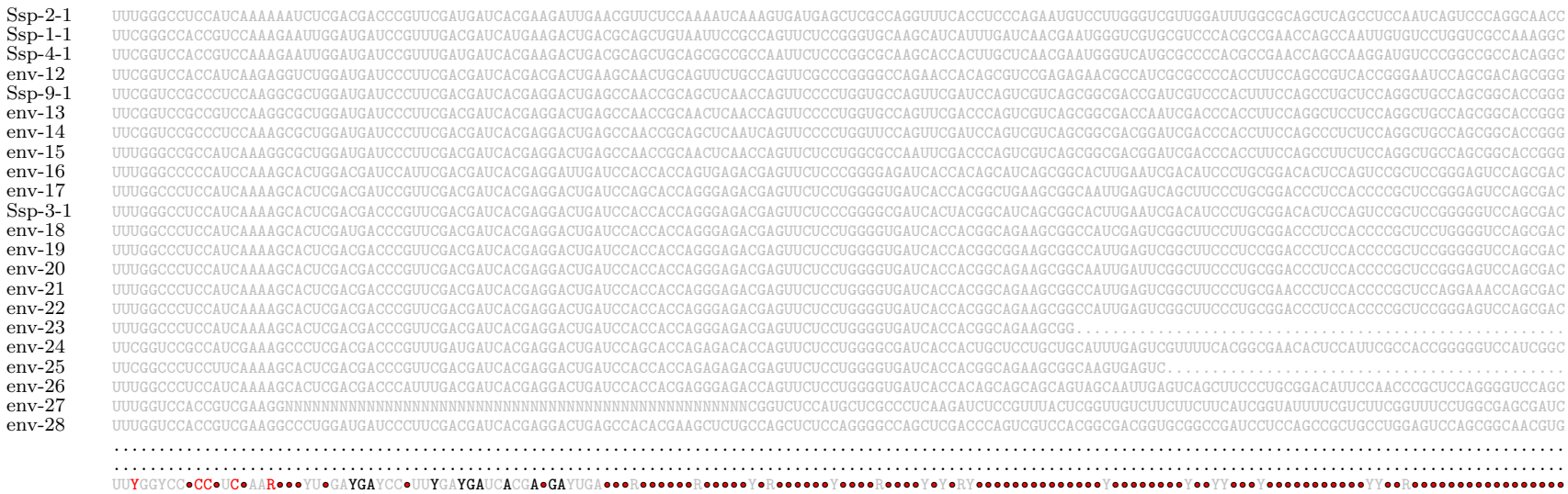
GOS12019938 (1) ?
GOS13170996 (16) ?
GOS13442019 (1) ?
GOS17710 (3) ?
GOS2285 (15) ?
NOG16382 (3)
NOG21533 (3)
pfam00737 (35) Photosystem II 10 kDa phosphoprotein.
pfam01074 (1) Glycosyl hydrolases family 38.
pfam02468 (17) Photosystem II reaction centre N protein (psbN).

70.4 Multiple-sequence alignment

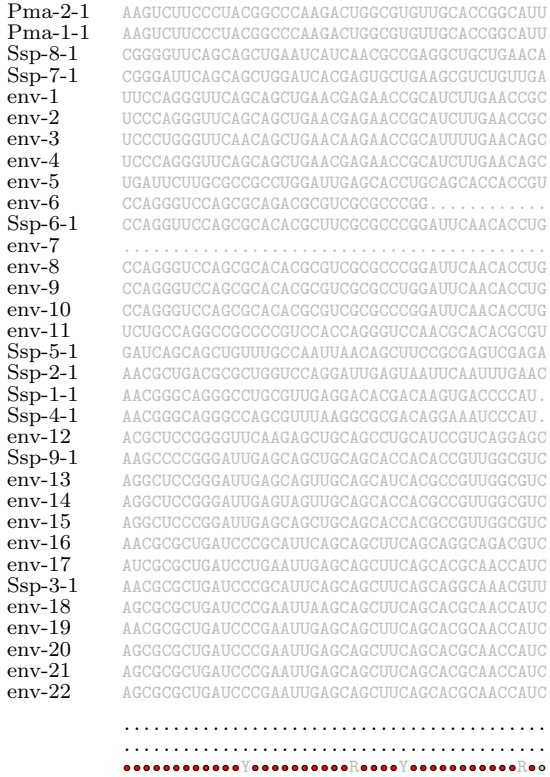
Each *psbNH* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 70.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *psbNH* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching <

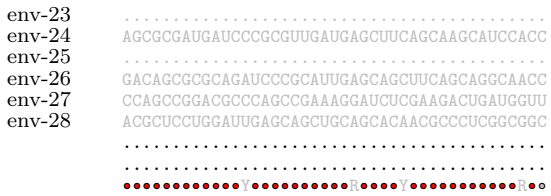
and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below this base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.


Ssp-8-1
Ssp-7-1
env-1
env-2
env-3
env-4
env-5
env-6
Ssp-6-1
env-7
env-8
env-9
env-10
env-11
Ssp-5-1
Ssp-2-1
Ssp-1-1
Ssp-4-1
env-12
Ssp-9-1
env-13
env-14
env-15
env-16
env-17
Ssp-3-1
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
5′
3′
alignment positions 260 · · · 419
Pma-2-1
Pma-1-1
Ssp-8-1
Ssp-7-1
env-1
env-2
env-3
env-4
env-5
env-6
Ssp-6-1
env-7
env-8
env-9
env-10
env-11
Ssp-5-1



alignment positions 420 · · · 462





71 *Pseudomon*-1

71.1 Taxa

The taxonomy of each organism containing a putative *Pseudomon*-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *Pseudomon*-1 RNA in Sections 71.2 and 71.4.

abbrev. of hits	taxonomy of species
Avi-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Azotobacter vinelandii*
Pae-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* 2192
Pae-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* C3719
Pae-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PA7
Pae-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PACS2
Pae-5-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PAO1
Pae-6-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* UCBPP-PA14
Pen-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas entomophila* L48
Pfl-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf-5
Pfl-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf0-1
Pme-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas mendocina ymp*
Ppu-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* F1
Ppu-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* GB-1
Ppu-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* KT2440
Ppu-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* W619
Pst-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas stutzeri* A1501
Psy-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. phaseolicola* 1448A
Psy-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000
env-1 to env-4	environmental samples

71.2 Gene contexts

Each *Pseudomon*-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *Pseudomon*-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 71.3. The accession number of the sequence containing each *Pseudomon*-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 71.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	wf3_AHAI8448_y1	+	45	164	**RNA**→
Avi-1-1	NZ_AAAU03000001.1	+	218429	218539	53EXOc (cd00008)PolA (COG0749)→ **RNA**→ ←TetR_N (pfam00440)
Psy-3-1	NC_004578.1	-	373116	372999	53EXOc (cd00008)PolA (COG0749)→ **RNA**→ ←COG0218 (COG0218)
Psy-2-1	NC_007005.1	-	287565	287450	53EXOc (cd00008)PolA (COG0749)→ **RNA**→ ←COG0218 (COG0218)
Psy-1-1	NC_005773.3	-	299668	299553	53EXOc (cd00008)PolA (COG0749)→ **RNA**→ ←COG0218 (COG0218)
Pst-1-1	NC_009434.1	+	135860	135973	53EXOc (cd00008)PolA (COG0749)→ **RNA**→ ←hypo
Pae-2-1	NZ_AAKV01000085.1	+	45982	46097	COG0749: DNA polymerase I - 3'-5' exonuclease and polymerase domains→ **RNA**→ ←COG0218 (COG0218)
Pae-1-1	NZ_AAKW01000050.1	-	2192	2078	PolA (COG0749)→ **RNA**→ ←COG0218 (COG0218)
Pae-4-1	NZ_AAQW01000001.1	-	6411526	6411411	53EXOc (cd00008)PolA (COG0749)→ **RNA**→ ←COG0218 (COG0218)
Pae-3-1	NC_009656.1	-	6507817	6507702	53EXOc (cd00008)PolA (COG0749)→ **RNA**→ ←hypo
Pae-5-1	NC_002516.2	-	6183637	6183522	53EXOc (cd00008)PolA (COG0749)→ **RNA**→ ←COG0218 (COG0218)
Pae-6-1	NC_008463.1	-	6456548	6456433	53EXOc (cd00008)PolA (COG0749)→ **RNA**→ ←COG0218 (COG0218)
Pfl-2-1	NC_007492.1	-	60881	60764	53EXOc (cd00008)PolA (COG0749)→ **RNA**→ ←COG0218 (COG0218)
env-2	JCVI_SCAF_1096628074974	-	693	577	PolA (COG0749)? (GOS2377)→ **RNA**→ ←COG0218 (COG0218)? (GOS798)
env-3	JCVI_SCAF_1096628070897	-	776	660	53EXOc (cd00008)? (GOS2377)→ **RNA**→ ←COG0218 (COG0218)? (GOS798)
env-4	JCVI_SCAF_1101668316662	-	776	660	DNA_pol_A (pfam00476)? (GOS2377)→ **RNA**→ ←COG0218 (COG0218)? (GOS798)
Pme-1-1	NC_009439.1	+	155482	155598	53EXOc (cd00008)PolA (COG0749)→ **RNA**→ ←COG0218 (COG0218)
Pfl-1-1	NC_004129.6	+	81738	81854	53EXOc (cd00008)PolA (COG0749)→ **RNA**→ ←COG0218 (COG0218)
Pen-1-1	NC_008027.1	+	74508	74623	53EXOc (cd00008)PolA (COG0749)→ **RNA**→ ←COG0218 (COG0218)
Ppu-4-1	NZ_AAVY01000030.1	-	41020	40905	53EXOc (cd00008)PolA (COG0749)→ **RNA**→ ←COG0218 (COG0218)
Ppu-1-1	NC_009512.1	+	144034	144149	53EXOc (cd00008)PolA (COG0749)→ **RNA**→ ←COG0218 (COG0218)
Ppu-3-1	NC_002947.3	+	130441	130556	53EXOc (cd00008)PolA (COG0749)→ **RNA**→ ←COG0218 (COG0218)
Ppu-2-1	NZ_AAXR01000002.1	-	40490	40375	53EXOc (cd00008)PolA (COG0749)→ **RNA**→ ←COG0218 (COG0218)

71.3 Conserved domains

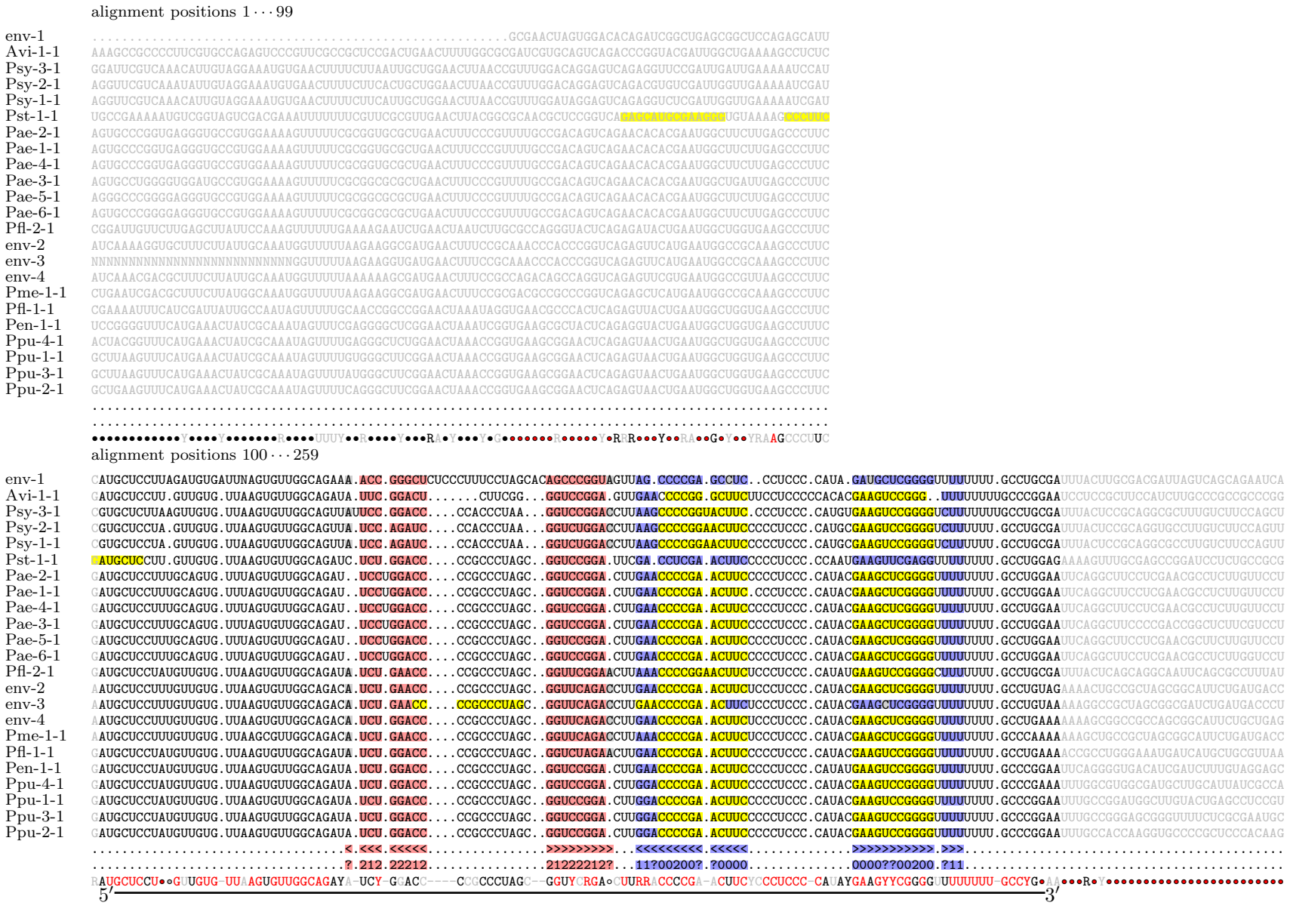
Conserved domains found in protein-coding genes listed in Section 71.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *Pseudomon*-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

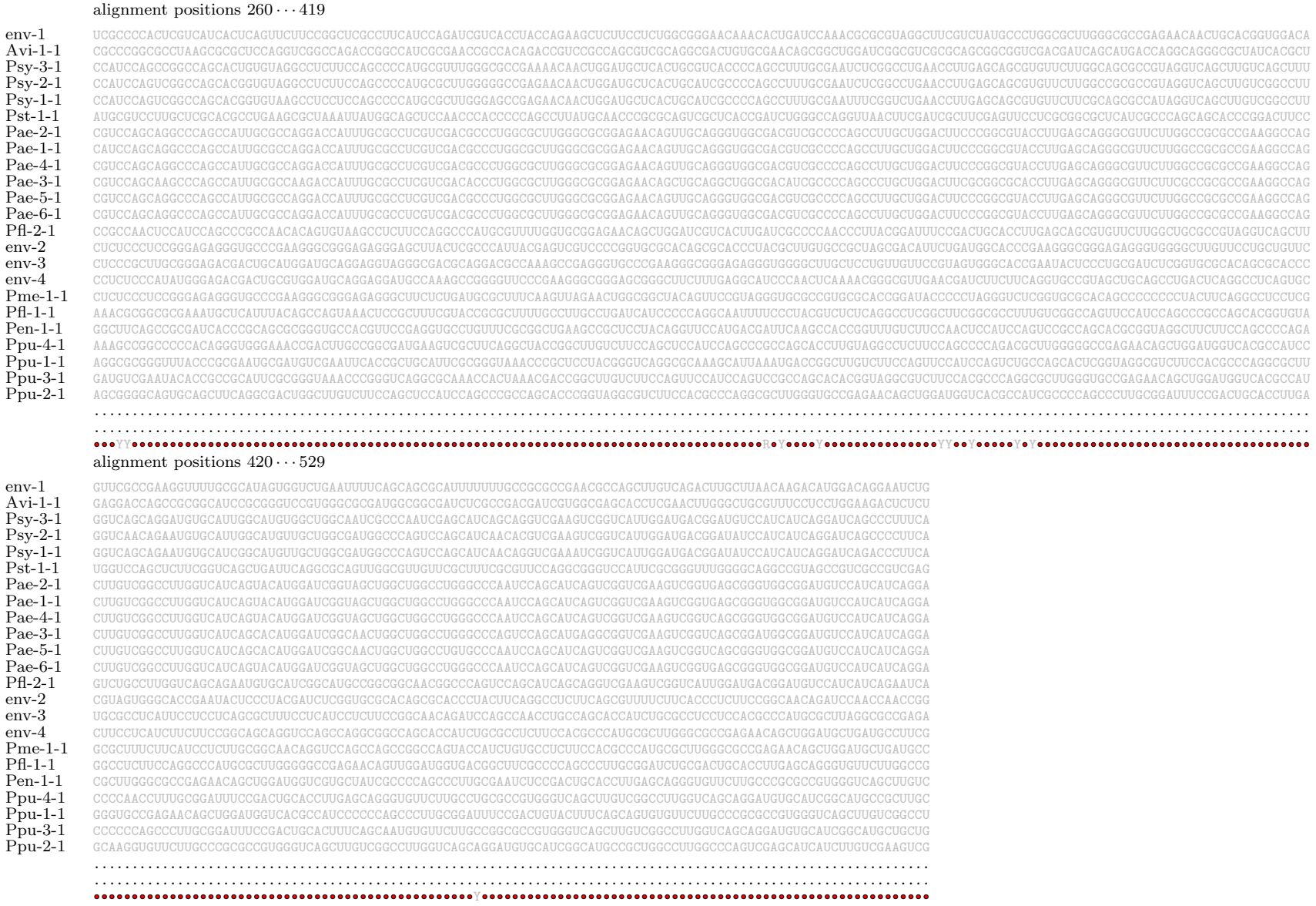
u cd00008 (18) 5'-3' exonuclease; T5 type 5'-3' exonuclease domains may co-occur with DNA polymerase I (Pol I) domains, or be part of Pol I containing complexes.
d COG0218 (19) Predicted GTPase [General function prediction only]
u COG0749 (19) DNA polymerase I - 3'-5' exonuclease and polymerase domains [DNA replication, recombination, and repair]
u GOS2377 (3) ?
d GOS798 (3) ?
d pfam00440 (1) Bacterial regulatory proteins, tetR family.
u pfam00476 (1) DNA polymerase family A.

71.4 Multiple-sequence alignment

Each *Pseudomon*-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 71.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *Pseudomon*-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs

exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.





72 *Pseudomon*-2

72.1 Taxa

The taxonomy of each organism containing a putative *Pseudomon*-2 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *Pseudomon*-2 RNA in Sections 72.2 and 72.4.

abbrev. of hits	taxonomy of species
Pae-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* 2192
Pae-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* C3719
Pae-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PA7
Pae-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PACS2
Pae-5-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PAO1
Pae-6-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* UCBPP-PA14
Pen-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas entomophila* L48
Pfl-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf-5
Pfl-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf0-1
Ppu-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* F1
Ppu-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* GB-1
Ppu-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* KT2440
Ppu-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* W619
Psy-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. phaseolicola* 1448A
Psy-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000

72.2 Gene contexts

Each *Pseudomon*-2 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *Pseudomon*-2 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 72.3. The accession number of the sequence containing each *Pseudomon*-2 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 72.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Pae-5-1	NC_002516.2	-	883590	883497	←COG4270 (COG4270) **RNA**→ hypo→
Pae-2-1	NZ_AAKV01000120.1	-	2685	2591	←COG4270 (COG4270) **RNA**→ ←ARA1 (COG0656)
Pae-3-1	NC_009656.1	+	4838512	4838605	←COG4270 (COG4270) **RNA**→ hypo→
Pae-6-1	NC_008463.1	+	4773797	4773890	←COG4270 (COG4270) **RNA**→ hypo→
Pae-1-1	NZ_AAKW01000080.1	-	46559	46466	←COG4270 (COG4270) **RNA**→ ←ARA1 (COG0656)
Pae-4-1	NZ_AAQW01000001.1	-	1453231	1453138	←COG4270 (COG4270) **RNA**→ ←ARA1 (COG0656)
Pfl-2-1	NC_007492.1	-	1846573	1846471	←hypo **RNA**→ hypo→
Pfl-1-1	NC_004129.6	+	5126929	5127029	←hypo **RNA**→ hypo→
Psy-1-1	NC_005773.3	+	4236121	4236222	hypo→ **RNA**→ hypo→
Psy-3-1	NC_004578.1	-	1892115	1892014	hypo→ **RNA**→ hypo→
Psy-2-1	NC_007005.1	+	4354612	4354713	hypo→ **RNA**→ hypo→
Pen-1-1	NC_008027.1	-	1458308	1458226	←hypo **RNA**→ hypo→
Ppu-4-1	NZ_AAVY01000037.1	+	27678	27778	←hypo (cons) **RNA**→ hypo (cons)→
Ppu-2-1	NZ_AAXR01000007.1	+	309	409	**RNA**→ hypo (cons)→
Ppu-3-1	NC_002947.3	-	1884027	1883927	←hypo **RNA**→ hypo→
Ppu-1-1	NC_009512.1	+	4546771	4546872	←Sel1 domain protein repeat-containing protein **RNA**→ hypo→

72.3 Conserved domains

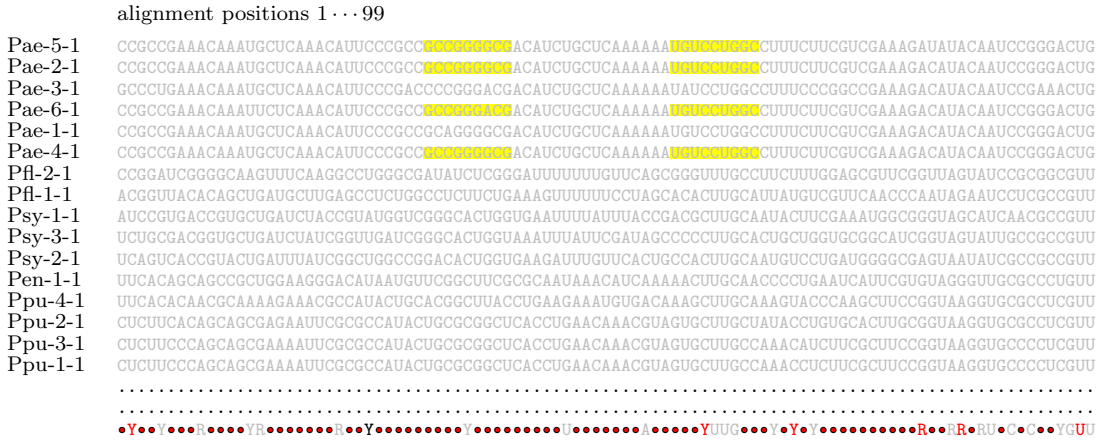
Conserved domains found in protein-coding genes listed in Section 72.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *Pseudomon*-2 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

d COG0656 (3) Aldo/keto reductases, related to diketogulonate reductase [General function prediction only]

u COG4270 (6) Predicted membrane protein [Function unknown]

72.4 Multiple-sequence alignment

Each *Pseudomon*-2 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 72.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *Pseudomon*-2 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



alignment positions 100 · · · 259
Pae-5-1
Pae-2-1
Pae-3-1
Pae-6-1
Pae-1-1
Pae-4-1
Pfl-2-1
Pfl-1-1
Psy-1-1
Psy-3-1
Psy-2-1
Pen-1-1
Ppu-4-1
Ppu-2-1
Ppu-3-1
Ppu-1-1
5′
3′
alignment positions 260 · · · 419
Pae-5-1
Pae-2-1
Pae-3-1
Pae-6-1
Pae-1-1
Pae-4-1
Pfl-2-1
Pfl-1-1
Psy-1-1
Psy-3-1
Psy-2-1
Pen-1-1
Ppu-4-1
Ppu-2-1
Ppu-3-1
Ppu-1-1
alignment positions 420 · · · 520
Pae-5-1
Pae-2-1
Pae-3-1
Pae-6-1
Pae-1-1
Pae-4-1
Pfl-2-1
Pfl-1-1
Psy-1-1
Psy-3-1
Psy-2-1
Pen-1-1
Ppu-4-1
Ppu-2-1
Ppu-3-1
Ppu-1-1

73 *Pseudomon*-GGDEF

73.1 Taxa

The taxonomy of each organism containing a putative *Pseudomon*-GGDEF RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *Pseudomon*-GGDEF RNA in Sections 73.2 and 73.4.

abbrev. of hits	taxonomy of species
Psy-1-1 to Psy-1-4	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. phaseolicola* 1448A
Psy-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1 to Psy-3-2	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000

73.2 Gene contexts

Each *Pseudomon*-GGDEF RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *Pseudomon*-GGDEF RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 73.3. The accession number of the sequence containing each *Pseudomon*-GGDEF RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 73.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Psy-1-1	NC_005773.3	-	4038706	4038584	**RNA**→ COG5001 (COG5001)→
Psy-3-1	NC_004633.1	-	36935	36813	**RNA**→ COG5001 (COG5001)→
Psy-1-2	NC_007274.1	+	85953	86075	**RNA**→ COG5001 (COG5001)→ COG3293 (COG3293)→ Resolvase (pfam00239)HTH_7 (pfam02796)→ YacA→ ParE (COG3668)→
Psy-1-3	NC_007274.1	-	24986	24864	**RNA**→ COG5001 (COG5001)→
Psy-1-4	NC_005773.3	+	2854245	2854367	**RNA**→ COG5001 (COG5001)→ GntR (COG1802)→
Psy-3-2	NC_004578.1	-	3272676	3272554	**RNA**→ COG5001 (COG5001)→ GntR (COG1802)→
Psy-2-1	NC_007005.1	-	3294081	3293955	**RNA**→ COG5001 (COG5001)→ GntR (COG1802)→

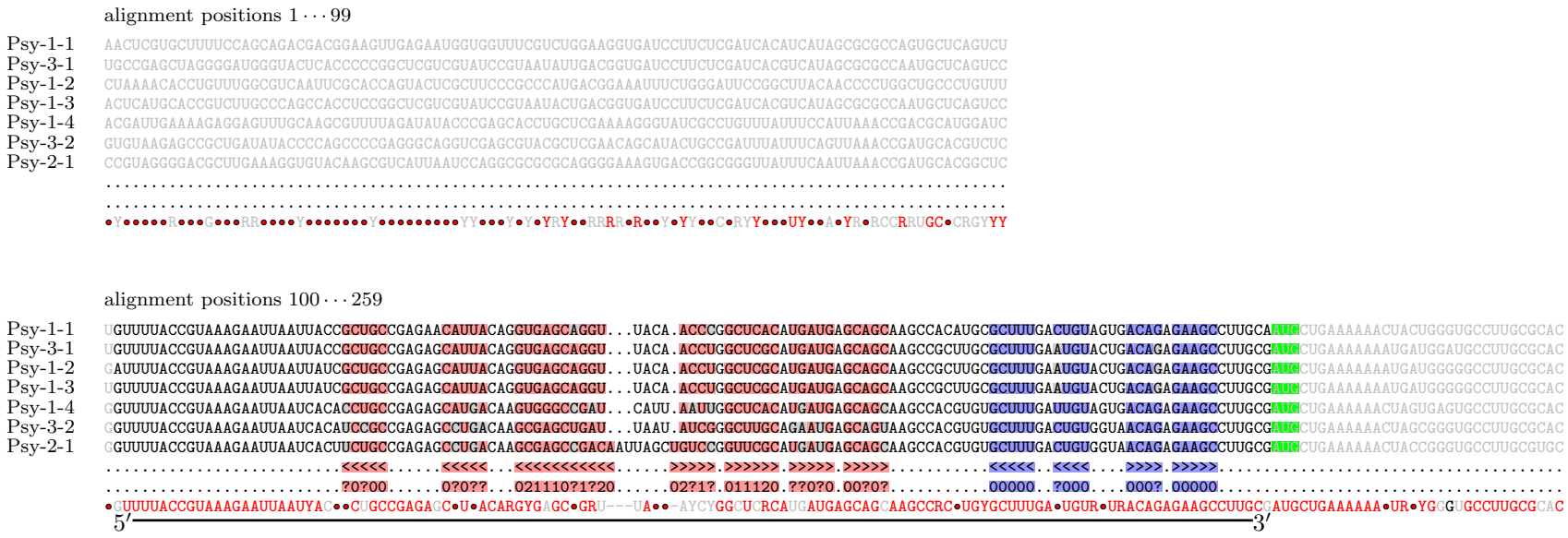
73.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 73.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *Pseudomon*-GGDEF RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

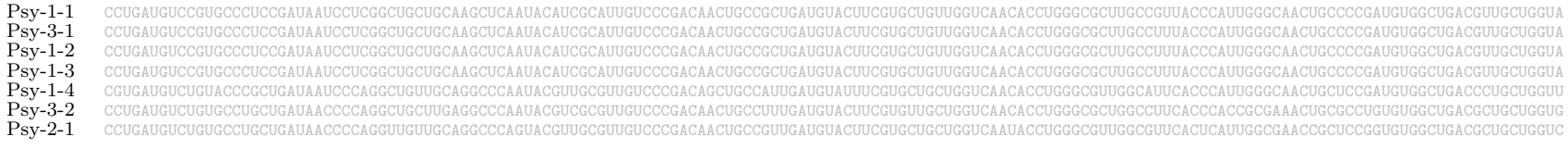
COG1802 (3) Transcriptional regulators [Transcription]
COG3293 (1) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
COG3668 (1) Plasmid stabilization system protein [General function prediction only]
COG5001 (7) Predicted signal transduction protein containing a membrane domain, an EAL and a GGDEF domain [Signal transduction mechanisms]
pfam00239 (1) Resolvase, N terminal domain.
pfam02796 (1) Helix-turn-helix domain of resolvase.

73.4 Multiple-sequence alignment

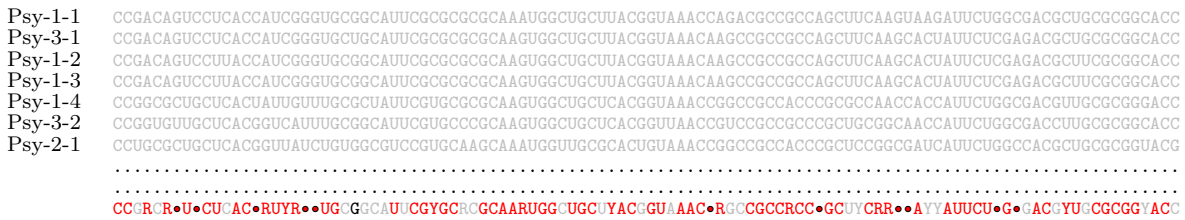
Each *Pseudomon*-GGDEF RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 73.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *Pseudomon*-GGDEF RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



alignment positions 260 · · · 419



alignment positions 420 · · · 527



74 *Pseudomon-groES*

74.1 Taxa

The taxonomy of each organism containing a putative *Pseudomon-groES* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *Pseudomon-groES* RNA in Sections 74.2 and 74.4.

abbrev.	# of hits	taxonomy of species
Avi-1-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Azotobacter vinelandii*
Pae-1-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* 2192
Pae-2-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* C3719
Pae-3-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PA7
Pae-4-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PACS2
Pae-5-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PAO1
Pae-6-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* UCBPP-PA14
Pen-1-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas entomophila* L48
Pfl-1-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf-5
Pfl-2-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf0-1
Pme-1-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas mendocina ymp*
Ppu-1-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* F1
Ppu-2-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* GB-1
Ppu-3-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* KT2440
Ppu-4-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* W619
Pst-1-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas stutzeri* A1501
Psy-1-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. phaseolicola* 1448A
Psy-2-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1		Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000
env-1		environmental sample

74.2 Gene contexts

Each *Pseudomon-groES* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *Pseudomon-groES* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 74.3. The accession number of the sequence containing each *Pseudomon-groES* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 74.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Pme-1-1	NC_009439.1	+	4070023	4070138	**RNA**→ Cpn10 (pfam00166)→ GroEL (cd03344)→
Pae-3-1	NC_009656.1	-	5112607	5112494	**RNA**→ Cpn10 (pfam00166)→ GroEL (cd03344)→
Pae-6-1	NC_008463.1	-	5081170	5081056	**RNA**→ Cpn10 (pfam00166)→ GroEL (cd03344)→
Pae-5-1	NC_002516.2	-	4917616	4917502	**RNA**→ Cpn10 (pfam00166)→ GroEL (cd03344)→
Pae-2-1	NZ_AAKV01000114.1	+	21675	21789	**RNA**→ Cpn10 (pfam00166)→ GroEL (cd03344)→ GroEL (cd03344)→
Pae-1-1	NZ_AAKW01000074.1	+	184532	184646	**RNA**→ Cpn10 (pfam00166)→ GroEL (cd03344)→
Pae-4-1	NZ_AAQW01000001.1	+	1172268	1172382	**RNA**→ Cpn10 (pfam00166)→ GroEL (cd03344)→
Pen-1-1	NC_008027.1	-	4773909	4773793	**RNA**→ Cpn10 (pfam00166)→ GroEL (cd03344)→
Ppu-4-1	NZ_AAVY01000005.1	-	58884	58767	**RNA**→ Cpn10 (pfam00166)→ GroEL (cd03344)→
Ppu-2-1	NZ_AAXR01000015.1	+	67156	67274	**RNA**→ Cpn10 (pfam00166)→ GroEL (cd03344)→
Ppu-3-1	NC_002947.3	+	1549106	1549223	**RNA**→ Cpn10 (pfam00166)→ GroEL (cd03344)→
Ppu-1-1	NC_009512.1	-	4882595	4882478	**RNA**→ Cpn10 (pfam00166)→ GroEL (cd03344)→
Avi-1-1	NZ_AAAU03000004.1	-	179284	179172	**RNA**→ Cpn10 (pfam00166)→ GroEL (cd03344)→
Pst-1-1	NC_009434.1	-	3389964	3389849	**RNA**→ Cpn10 (pfam00166)→ GroEL (cd03344)→
env-1	JCVI_SCAF_1096628077435	+	1237	1351	**RNA**→ Cpn10 (pfam00166)? (GOS12636)→
Psy-3-1	NC_004578.1	-	4940472	4940362	**RNA**→ Cpn10 (pfam00166)→ GroEL (cd03344)→
Psy-1-1	NC_005773.3	-	4661390	4661280	**RNA**→ Cpn10 (pfam00166)→ GroEL (cd03344)→
Psy-2-1	NC_007005.1	-	4857654	4857544	**RNA**→ Cpn10 (pfam00166)→ GroEL (cd03344)→
Pfl-2-1	NC_007492.1	-	5078998	5078885	**RNA**→ Cpn10 (pfam00166)→ GroEL (cd03344)→
Pfl-1-1	NC_004129.6	-	5584544	5584432	**RNA**→ Cpn10 (pfam00166)→ GroEL (cd03344)→

74.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 74.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *Pseudomon-groES* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

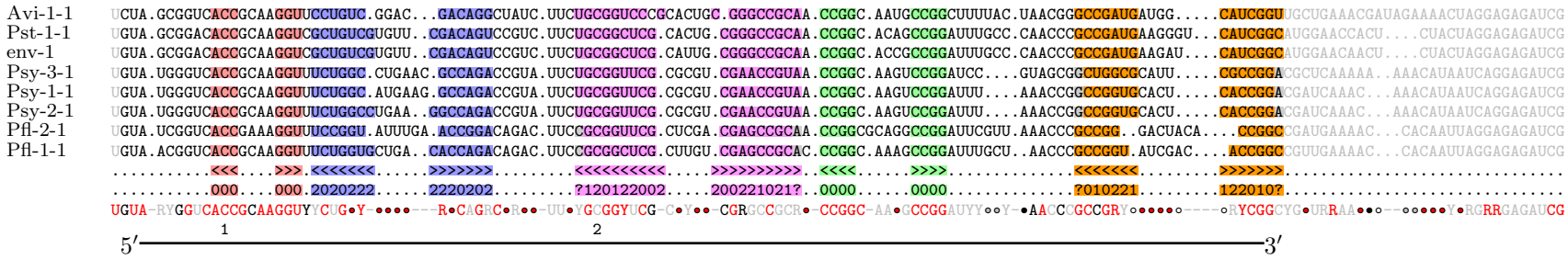
cd03344 (20) GroEL_like type I chaperonin.
GOS12636 (1) ?

pfam00166 (20) Chaperonin 10 Kd subunit.

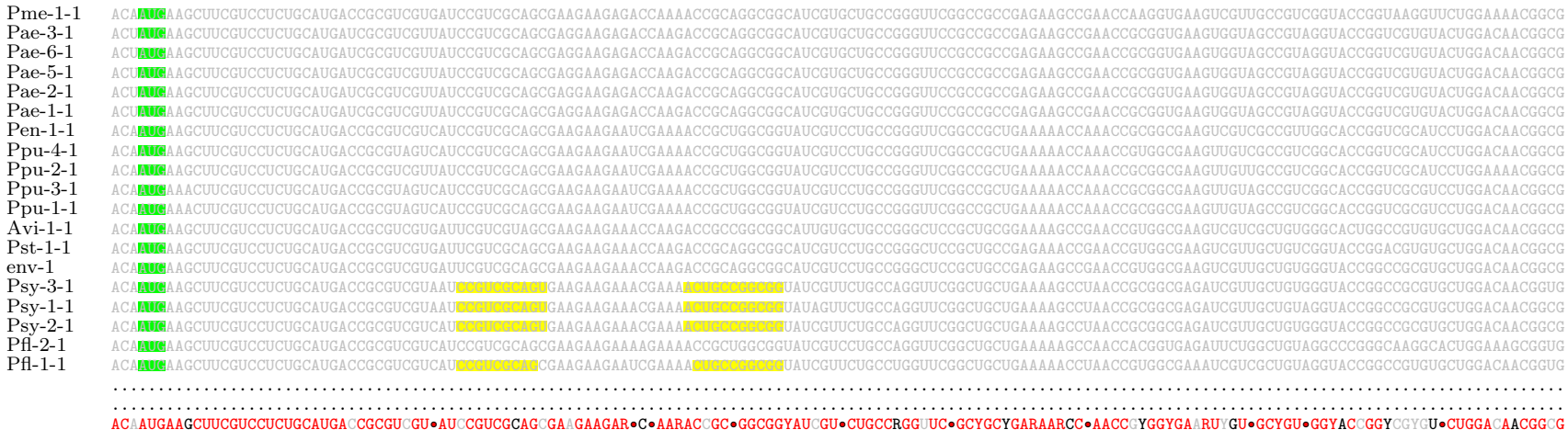
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Pae-4-1

74.4 Multiple-sequence alignment

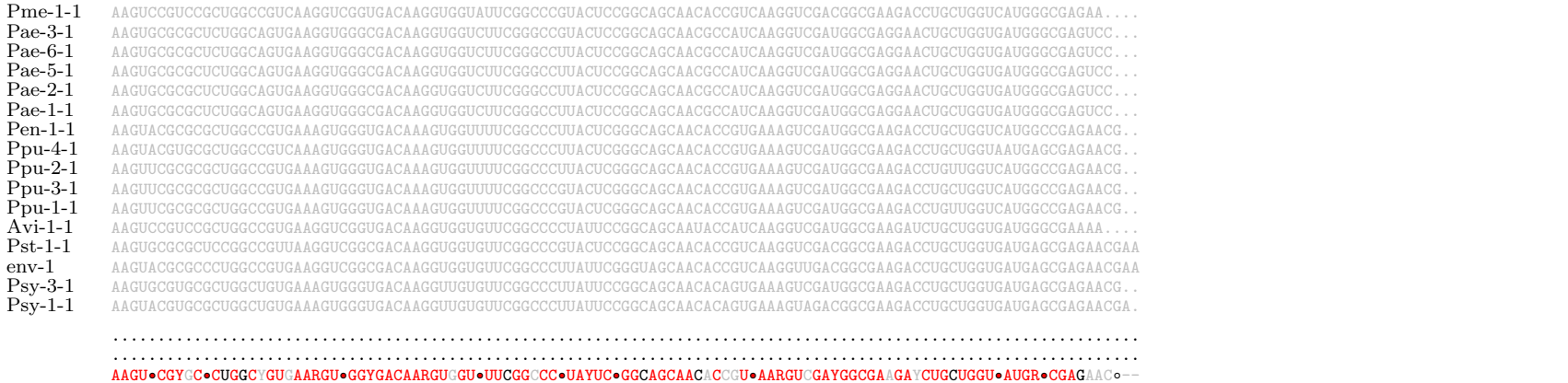
Each *Pseudomon-groES* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 74.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *Pseudomon-groES* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

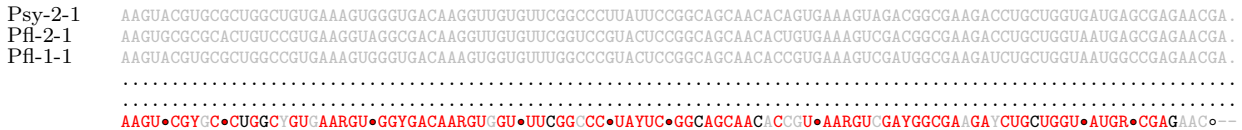


alignment positions 260 · · · 419



alignment positions 420 · · · 532





75 *Pseudomon*-Rho

75.1 Taxa

The taxonomy of each organism containing a putative *Pseudomon*-Rho RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *Pseudomon*-Rho RNA in Sections 75.2 and 75.4.

abbrev. of hits	taxonomy of species
Avi-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Azotobacter vinelandii*
Pae-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* 2192
Pae-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* C3719
Pae-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PA7
Pae-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PACS2
Pae-5-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PAO1
Pae-6-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* UCBPP-PA14
Pen-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas entomophila* L48
Pfl-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf-5
Pfl-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf0-1
Pme-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas mendocina ymp*
Ppu-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* F1
Ppu-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* GB-1
Ppu-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* KT2440
Ppu-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* W619
Pst-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas stutzeri* A1501
Psy-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. phaseolicola* 1448A
Psy-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000
env-1 to env-3	environmental samples

75.2 Gene contexts

Each *Pseudomon*-Rho RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *Pseudomon*-Rho RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 75.3. The accession number of the sequence containing each *Pseudomon*-Rho RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 75.1.

abbrev.	Seq. accession	5′ at	3′ at	genes

Pst-1-1	NC_009434.1	+	602149	602251	**RNA**→ Rho (COG1158)→
Pae-3-1	NC_009656.1	-	6182387	6182288	**RNA**→ Rho (COG1158)→ COG1835 (COG1835)→ UbiD (pfam01977)→ fer2 (cd00207)UbiB (COG0543)→
Pae-6-1	NC_008463.1	-	6172494	6172394	**RNA**→ Rho (COG1158)→ COG1835 (COG1835)→ UbiD (pfam01977)→ fer2 (cd00207)UbiB (COG0543)→
Pae-5-1	NC_002516.2	-	5900281	5900181	**RNA**→ Rho (COG1158)→ COG1835 (COG1835)→ UbiD (pfam01977)→ fer2 (cd00207)UbiB (COG0543)→
Pae-2-1	NZ_AAKV01000091.1	+	82803	82903	**RNA**→ Rho (COG1158)→
Pae-1-1	NZ_AAKW01000062.1	+	141534	141634	**RNA**→ Rho (COG1158)→
Pae-4-1	NZ_AAQW01000001.1	-	6128167	6128067	**RNA**→ Rho (COG1158)→
Avi-1-1	NZ_AAAU03000002.1	-	309629	309521	**RNA**→ Rho (COG1158)→ UbiD (pfam01977)→ fer2 (cd00207)UbiB (COG0543)→
env-1	JCVI_SCAF_1096628042271	-	1643	1543	**RNA**→ Rho (COG1158)? (GOS2904)→ UbiB (COG0543)? (GOS230)→
env-2	JCVI_SCAF_1096628075208	+	1034	1134	**RNA**→ Rho (COG1158)? (GOS2904)→
Pme-1-1	NC_009439.1	+	316056	316157	**RNA**→ Rho (COG1158)→
Psy-3-1	NC_004578.1	-	5963068	5962969	**RNA**→ Rho (COG1158)→
Psy-1-1	NC_005773.3	+	330473	330572	**RNA**→ Rho (COG1158)→
Psy-2-1	NC_007005.1	+	320164	320263	**RNA**→ Rho (COG1158)→
Pfl-2-1	NC_007492.1	-	6124613	6124513	**RNA**→ Rho (COG1158)→
Pfl-1-1	NC_004129.6	-	6800498	6800398	**RNA**→ Rho (COG1158)→
env-3	JCVI_SCAF_1096626956697	-	1873	1774	**RNA**→ ←? (GOS6636)
Ppu-4-1	NC_010501.1	-	5563696	5563597	**RNA**→ Rho (COG1158)→
Pen-1-1	NC_008027.1	-	5652749	5652649	**RNA**→ Rho (COG1158)→
Ppu-2-1	NC_010322.1	-	5894889	5894790	**RNA**→ Rho (COG1158)→
Ppu-3-1	NC_002947.3	-	5948619	5948520	**RNA**→ Rho (COG1158)→ UbiD (pfam01977)→ fer2 (cd00207)UbiB (COG0543)→
Ppu-1-1	NC_009512.1	-	5757664	5757565	**RNA**→ Rho (COG1158)→

75.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 75.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *Pseudomon*-Rho RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

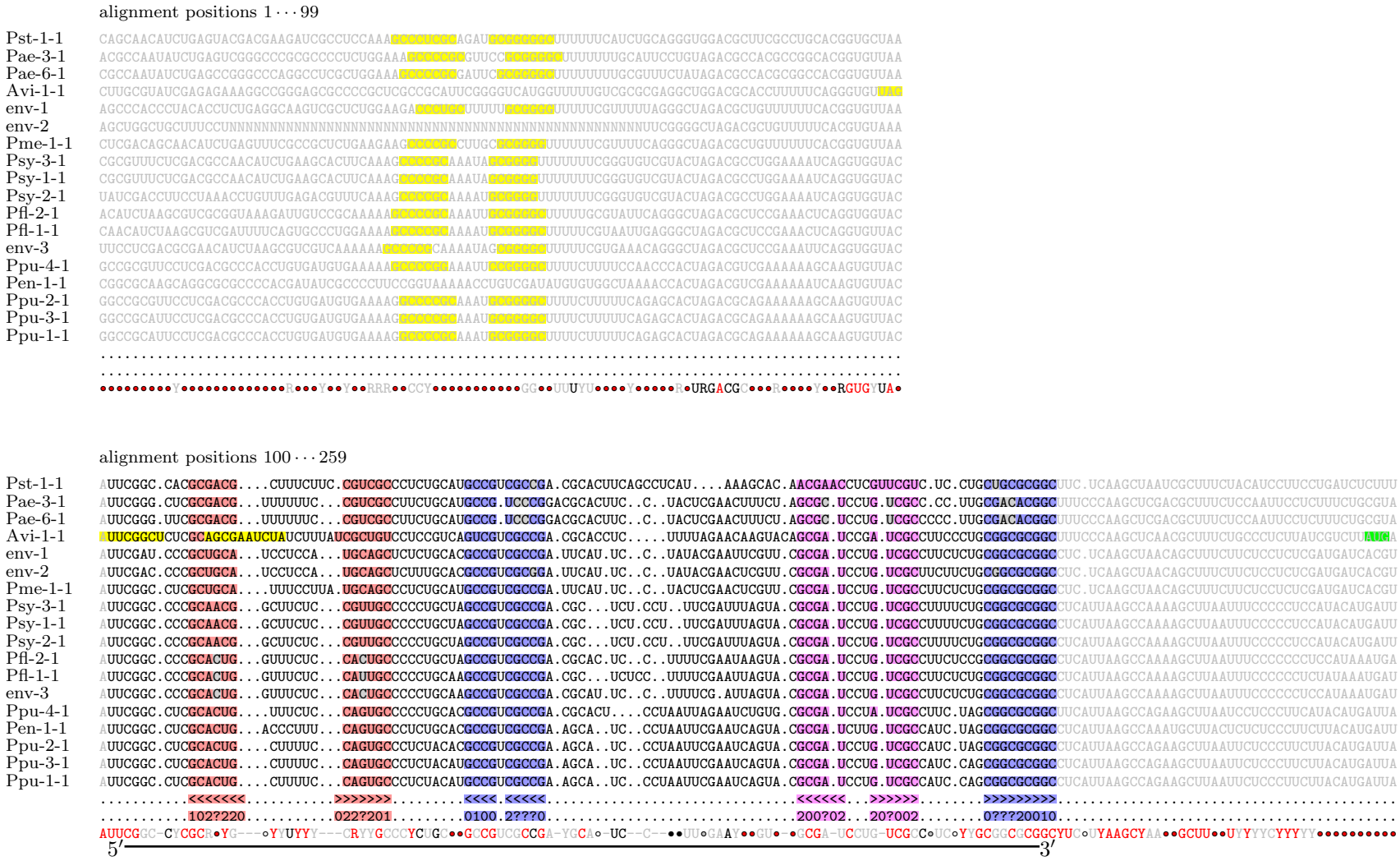
cd00207 (5) 2Fe-2S iron-sulfur cluster binding domain.
COG0543 (6) 2-polyprenylphenol hydroxylase and related flavodoxin oxidoreductases [Coenzyme metabolism / Energy production and conversion]
COG1158 (21) Transcription termination factor [Transcription]
COG1835 (3) Predicted acyltransferases [Lipid metabolism]
GOS230 (1) ?
GOS2904 (2) ?
GOS6636 (1) ?
pfam01977 (5) 3-octaprenyl-4-hydroxybenzoate carboxy-lyase.

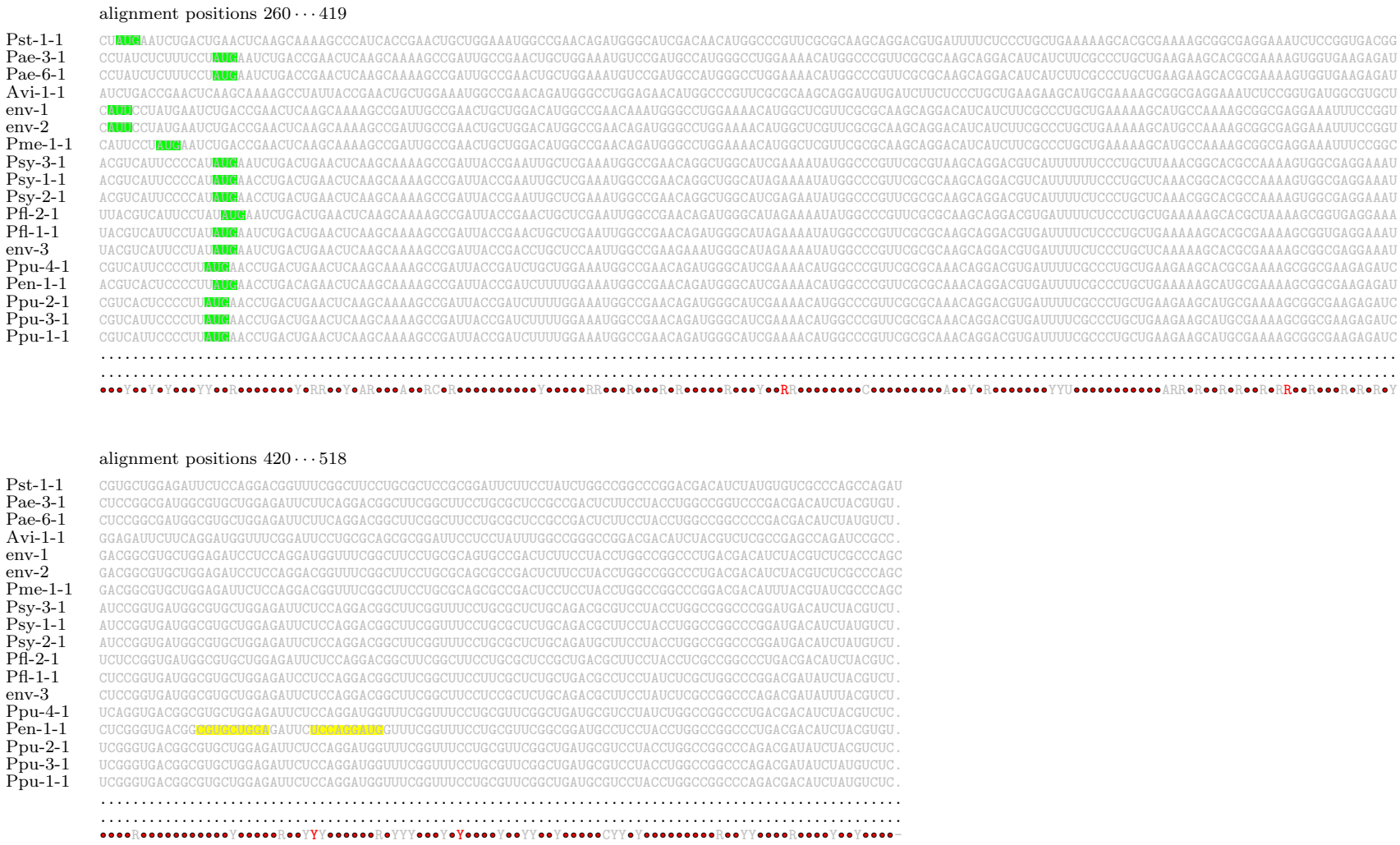
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Pae-1-1, Pae-2-1, Pae-4-1, Pae-5-1

75.4 Multiple-sequence alignment

Each *Pseudomon*-Rho RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 75.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *Pseudomon*-Rho RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets,

where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.





76 *Pyrobac*-1

76.1 Taxa

The taxonomy of each organism containing a putative *Pyrobac*-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *Pyrobac*-1 RNA in Sections 76.2 and 76.4.

abbrev. of hits	taxonomy of species
Pae-1-1 to Pae-1-2	Archaea Crenarchaeota Thermoprotei Thermoproteales Thermoproteaceae *Pyrobaculum aerophilum str.* IM2
Par-1-1 to Par-1-2	Archaea Crenarchaeota Thermoprotei Thermoproteales Thermoproteaceae *Pyrobaculum arsenaticum* DSM 13514
Pca-1-1 to Pca-1-2	Archaea Crenarchaeota Thermoprotei Thermoproteales Thermoproteaceae *Pyrobaculum calidifontis* JCM 11548
Pis-1-1 to Pis-1-2	Archaea Crenarchaeota Thermoprotei Thermoproteales Thermoproteaceae *Pyrobaculum islandicum* DSM 4184

76.2 Gene contexts

Each *Pyrobac*-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *Pyrobac*-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 76.3. The accession number of the sequence containing each *Pyrobac*-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 76.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Pis-1-1	NC_008701.1	+	1105103	1105165	←UPF0027 (pfam01139) **RNA**→ Asparaginase (cd00411)→
Pca-1-1	NC_009073.1	+	138640	138702	←HTH_ARSR (smart00418) **RNA**→ hypo→
Pae-1-1	NC_003364.1	+	587219	587280	P. aerophilum family 3 protein→ **RNA**→ conserved within P. aerophilum→
Par-1-1	NC_009376.1	-	650735	650674	←tRNA-Val **RNA**→ hypo→
Pae-1-2	NC_003364.1	+	1080081	1080140	Ribosomal_L15e (pfam00827)→ **RNA**→ ←hypo
Pca-1-2	NC_009073.1	-	482784	482726	←COG1325 (COG1325) **RNA**→ ←hypo
Par-1-2	NC_009376.1	-	634088	634027	Ribosomal_L15e (pfam00827)→ **RNA**→ ←hypo
Pis-1-2	NC_008701.1	+	1584379	1584438	Ribosomal_L15e (pfam00827)→ **RNA**→ ←hypo

76.3 Conserved domains

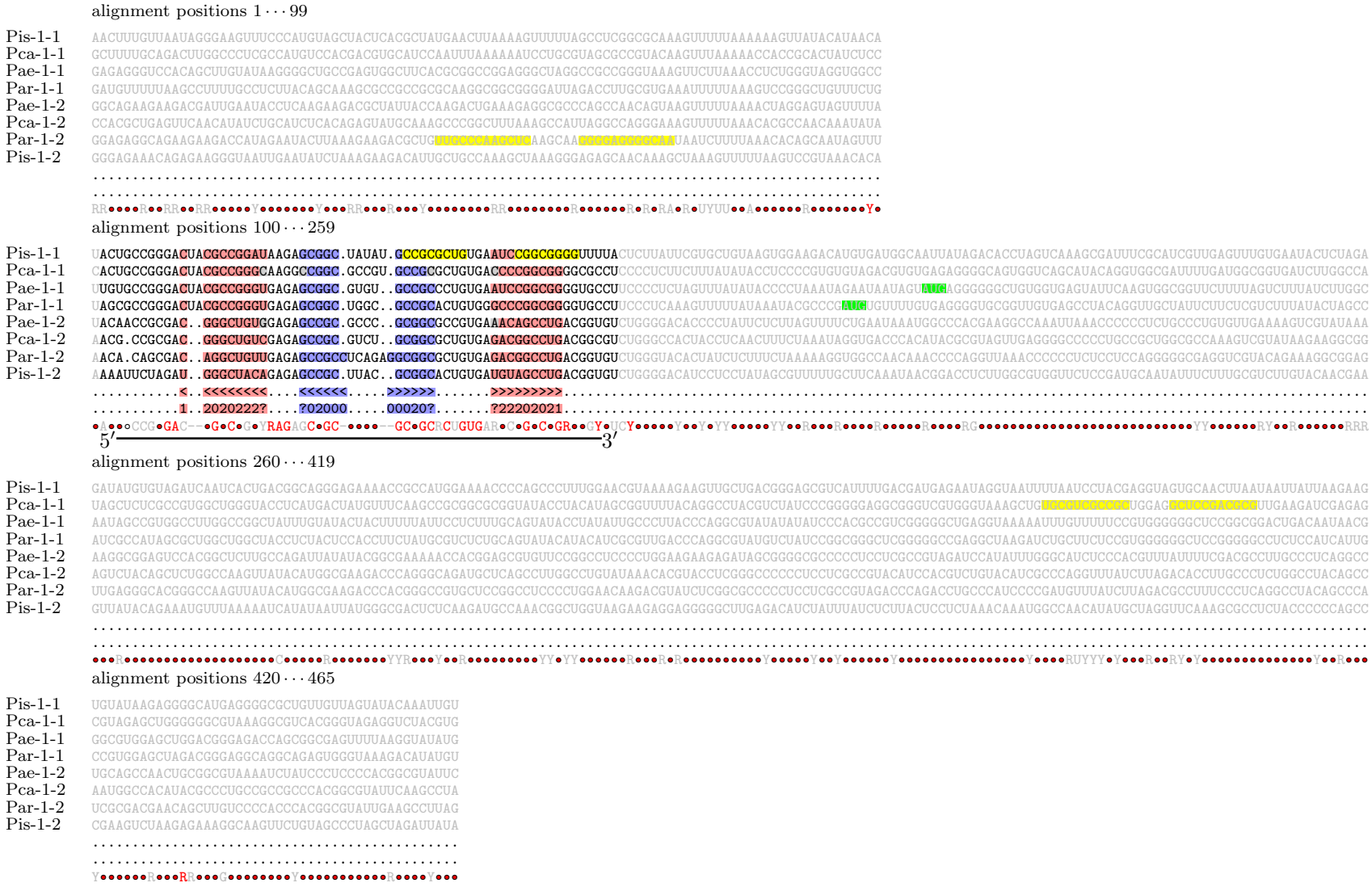
Conserved domains found in protein-coding genes listed in Section 76.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *Pyrobac*-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

D cd00411 (1) Asparaginase (amidohydrolase): Asparaginases are tetrameric enzymes that catalyze the hydrolysis of asparagine to aspartic acid and ammonia.
u COG1325 (1) Predicted exosome subunit [Translation, ribosomal structure and biogenesis]
u pfam00827 (3) Ribosomal L15.
u pfam01139 (1) Uncharacterized protein family UPF0027.
u smart00418 (1) helix_turn_helix, Arsenical Resistance Operon Repressor;

76.4 Multiple-sequence alignment

Each *Pyrobac*-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 76.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *Pyrobac*-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons

are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



77 *Pyrobac*-HINT

77.1 Taxa

The taxonomy of each organism containing a putative *Pyrobac*-HINT RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *Pyrobac*-HINT RNA in Sections 77.2 and 77.4.

abbrev. of hits	taxonomy of species
Pae-1-1	Archaea Crenarchaeota Thermoprotei Thermoproteales Thermoproteaceae *Pyrobaculum aerophilum str.* IM2
Par-1-1	Archaea Crenarchaeota Thermoprotei Thermoproteales Thermoproteaceae *Pyrobaculum arsenaticum* DSM 13514
Pca-1-1	Archaea Crenarchaeota Thermoprotei Thermoproteales Thermoproteaceae *Pyrobaculum calidifontis* JCM 11548
Pis-1-1	Archaea Crenarchaeota Thermoprotei Thermoproteales Thermoproteaceae *Pyrobaculum islandicum* DSM 4184

77.2 Gene contexts

Each *Pyrobac*-HINT RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *Pyrobac*-HINT RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 77.3. The accession number of the sequence containing each *Pyrobac*-HINT RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 77.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Pca-1-1	NC_009073.1	+	1673183	1673248	**RNA**→ HINT_subgroup (cd01277)→
Pis-1-1	NC_008701.1	+	493855	493922	**RNA**→ HINT_subgroup (cd01277)→ HTH_LUXR (smart00421)→
Par-1-1	NC_009376.1	-	1502914	1502850	**RNA**→ HINT_subgroup (cd01277)→ HTH_LUXR (smart00421)→
Pae-1-1	NC_003364.1	+	1883639	1883702	**RNA**→ HINT_subgroup (cd01277)→ HTH_LUXR (smart00421)→

77.3 Conserved domains

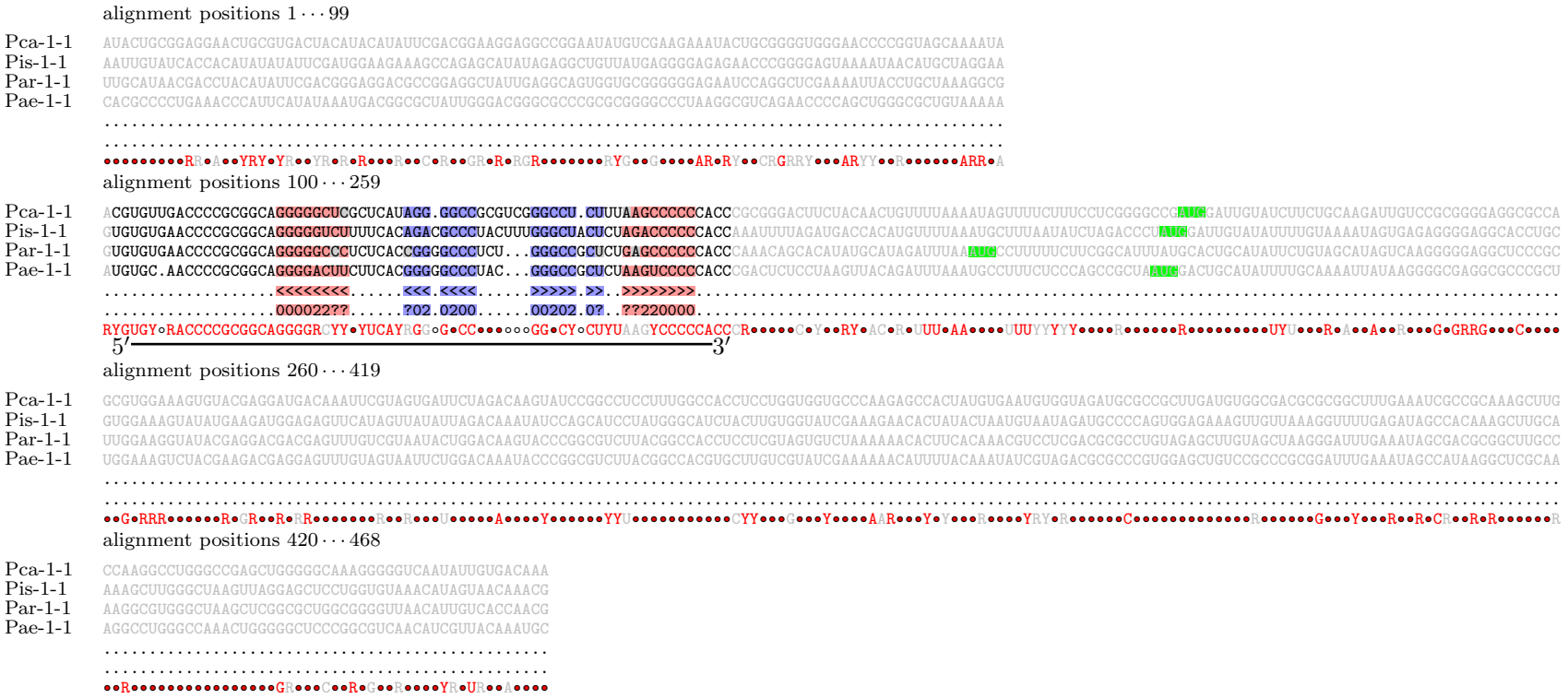
Conserved domains found in protein-coding genes listed in Section 77.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *Pyrobac*-HINT RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

D cd01277 (4) HINT (histidine triad nucleotide-binding protein) subgroup: Members of this CD belong to the superfamily of histidine triad hydrolases that act on alpha-phosphate of ribonucleotides.

D smart00421 (3) helix_turn_helix, Lux Regulon; lux regulon (activates the bioluminescence operon

77.4 Multiple-sequence alignment

Each *Pyrobac*-HINT RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 77.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *Pyrobac*-HINT RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



78 *radC*

78.1 Taxa

The taxonomy of each organism containing a putative *radC* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *radC* RNA in Sections 78.2 and 78.4.

abbrev. of hits	taxonomy of species
Pla-1-1 to Pla-1-4	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Phyllobacteriaceae *Parvibaculum lavamentivorans* DS-1
Bpe-1-1 to Bpe-1-16	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Alcaligenaceae *Bordetella petrii* DSM 12804
Bxe-1-1 to Bxe-1-5	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Burkholderia xenovorans* LB400
Bps-1-1 to Bps-1-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 406e
Bps-2-1 to Bps-2-3	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* B7210
Bps-3-1 to Bps-3-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* S13
Rme-1-1 to Rme-1-4	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Cupriavidus *Ralstonia metallidurans* CH34
Rpi-1-1 to Rpi-1-5	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Ralstonia pickettii* 12D
Aav-1-1 to Aav-1-4	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae *Acidovorax avenae subsp. citrulli* AAC00-1
Asp-1-1 to Asp-1-12	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae *Acidovorax sp.* JS42
Cte-1-1 to Cte-1-4	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae *Comamonas testosteroni* KF-1
Dac-1-1 to Dac-1-4	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae *Delftia acidovorans* SPH-1
Mpe-1-1 to Mpe-1-3	Bacteria Proteobacteria Betaproteobacteria Burkholderiales *Methylibium petroleiphilum* PM1
Har-1-1 to Har-1-3	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Oxalobacteraceae *Herminiimonas arsenicoxydans*
Neu-1-1 to Neu-1-3	Bacteria Proteobacteria Betaproteobacteria Nitrosomonadales Nitrosomonadaceae *Nitrosomonas eutropha* C91
Aar-1-1 to Aar-1-4	Bacteria Proteobacteria Betaproteobacteria Rhodocyclales Rhodocyclaceae *Aromatoleum aromaticum* EbN1
Tsp-1-1 to Tsp-1-4	Bacteria Proteobacteria Betaproteobacteria Rhodocyclales Rhodocyclaceae *Thauera sp.* MZ1T
Tsp-2-1 to Tsp-2-4	Bacteria Proteobacteria Gammaproteobacteria Chromatiales Ectothiorhodospiraceae *Thioalkalivibrio sp.* HL-EbGR7
Sen-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Virchow *Salmonella enterica subsp. en*
Pae-1-1 to Pae-1-8	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* 2192
Pae-2-1 to Pae-2-4	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* C3719
Pae-3-1 to Pae-3-3	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PA7
Pae-4-1 to Pae-4-7	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PACS2
Pae-5-1 to Pae-5-3	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* UCBPP-PA14
Pfl-1-1 to Pfl-1-3	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf-5
Pfl-2-1 to Pfl-2-3	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf0-1
Ppu-1-1 to Ppu-1-3	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* F1
Ppu-2-1 to Ppu-2-3	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* W619
Psy-1-1 to Psy-1-2	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Xax-1-1 to Xax-1-4	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas axonopodis pv. citri str.* 306
Xca-1-1 to Xca-1-4	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas campestris pv. campestris str.* 8004
Xca-2-1 to Xca-2-2	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas campestris pv. campestris str.* B100
Xca-3-1 to Xca-3-7	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas campestris pv. vesicatoria str.* 85-10
Xfa-1-1 to Xfa-1-4	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xylella fastidiosa* 9a5c
env-1 to env-20	environmental samples

78.2 Gene contexts

Each *radC* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *radC* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 78.3. The accession number of the sequence containing each *radC* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 78.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Bps-2-1	NZ_ABBN01001396.1	-	99650	99488	**RNA**→ hypo→ hypo→ hypo→
Xax-1-1	NC_003919.1	+	2605044	2605188	**RNA**→ hypo→ hypo→

env-1	JCVI_SCAF_1101668523961	-	181	27	**RNA**→
Rpi-1-1	NZ_ABDZ01000015.1	-	55684	55530	**RNA**→ hypo (cons)→ COG4123 (COG4123)→
Tsp-2-1	NZ_ABYM01000018.1	-	15519	15373	**RNA**→ hypo (cons)→ conserved hypothetical plasmid protein→
Neu-1-1	NC_008344.1	-	108863	108709	**RNA**→ hypo→
Pae-1-1	NZ_AAKW01000063.1	-	23023	22892	**RNA**→ ←hypo
Pae-4-1	NZ_AAQW01000001.1	+	6030570	6030704	**RNA**→ RfbB (COG1088)→ RfbD (COG1091)→ RfbA (COG1209)→
Ppu-2-1	NC_010501.1	+	5450160	5450296	**RNA**→ phage integrase→
Ppu-1-1	NC_009512.1	+	5156412	5156550	**RNA**→ hypo→ hypo→ Phage_integrase (pfam00589)→ hypo→
Ppu-1-2	NC_009512.1	+	3339390	3339527	**RNA**→ hypo→
Pfl-2-1	NC_007492.1	+	3444997	3445139	**RNA**→ NOG13362 (NOG13362)NOG18463 (NOG18463)→ hypo→
Ppu-2-2	NC_010501.1	+	3676296	3676402	**RNA**→ hypo→
Pfl-2-2	NC_007492.1	+	3469195	3469339	**RNA**→ hypo→
env-2	JCVI_SCAF_1096627186203	-	1265	1125	**RNA**→ COG4227 (COG4227)? (GOS11123422)→
env-3	JCVI_SCAF_1096627929620	-	725	593	**RNA**→
Pfl-2-3	NC_007492.1	+	873551	873694	**RNA**→ hypo→ hypo→ Phage_integrase (pfam00589)NOG18361 (NOG18361)→
Ppu-1-3	NC_009512.1	+	3312138	3312275	**RNA**→ hypo→ hypo→ RadC (pfam04002)→ hypo→
Ppu-2-3	NC_010501.1	+	3637927	3638064	**RNA**→ hypo→ hypo→ RadC (pfam04002)→ hypo→
Xca-1-1	NC_007086.1	-	2470043	2469891	**RNA**→ hypo→
Pla-1-1	NC_009719.1	+	3635395	3635546	**RNA**→ hypo→
Cte-1-1	NZ_AAUJ01000001.1	+	1027307	1027458	**RNA**→ hypo (cons)→
Asp-1-1	NC_008782.1	-	1372297	1372146	**RNA**→ hypo→
Rme-1-1	NC_007973.1	+	2530420	2530571	**RNA**→ hypo→
Pae-1-2	NZ_AAKW01000022.1	+	180373	180524	**RNA**→ COG1100: GTPase SAR1 and related small G proteins→
Bpe-1-1	NC_010170.1	+	1116746	1116892	**RNA**→ hypo→
env-4	sludgeOz_Contig2460	+	592	738	**RNA**→ (,,,)→
Bpe-1-2	NC_010170.1	-	4560875	4560729	**RNA**→ hypo→
Dac-1-1	NC_010002.1	+	4547442	4547588	**RNA**→ hypo→
Xfa-1-1	NC_002488.3	-	1681828	1681680	**RNA**→ hypo→
Asp-1-2	NC_008782.1	-	1499393	1499248	**RNA**→ hypo→ hypo→ hypo→
Xca-3-1	NC_007508.1	-	2724046	2723901	**RNA**→ hypo→ hypo→ hypo→ hypo→ hypo→ hypo→
Bxe-1-1	NC_007951.1	+	3576431	3576582	**RNA**→ hypo→
Bpe-1-3	NC_010170.1	+	1515330	1515481	**RNA**→ hypo→
env-5	JCVI_SCAF_1096628044192	-	1759	1608	**RNA**→ ? (GOS11886727)→ ? (GOS13495353)→ ? (GOS13066296)→
Rpi-1-2	NZ_ABDZ01000015.1	-	60751	60591	**RNA**→ hypo (cons)→ hypo (cons)→
Asp-1-3	NC_008782.1	-	2364053	2363900	**RNA**→ hypo→ hypo→
Bpe-1-4	NC_010170.1	+	1373516	1373669	**RNA**→ hypo→
Tsp-1-1	NZ_ABKT01000008.1	+	47409	47564	**RNA**→ hypo (cons)→ hypo (cons)→
Xax-1-2	NC_003919.1	+	2668338	2668489	**RNA**→ RadC (pfam04002)→
Pae-3-1	NC_009656.1	+	4617908	4618045	**RNA**→ hypo→ hypo→ hypo→ hypo→ hypo→ hypo→ O-methyl transferase family protein→
Pae-5-1	NC_008463.1	+	5292563	5292700	**RNA**→ hypo→ hypo→ hypo→ hypo→ hypo→ hypo→ COG4123 (COG4123)→
Pae-2-1	NZ_AAKV01000112.1	-	73439	73302	**RNA**→ hypo→ hypo→ hypo→
Pae-1-3	NZ_AAKW01000073.1	-	69153	69016	**RNA**→ hypo→ hypo→
Pae-4-2	NZ_AAQW01000001.1	-	958497	958360	**RNA**→ hypo→ hypo→
Pfl-1-1	NC_004129.6	+	5414183	5414317	**RNA**→ hypo→
Pae-1-4	NZ_AAKW01000073.1	-	72757	72610	**RNA**→ hypo→
Pae-4-3	NZ_AAQW01000001.1	-	962101	961954	**RNA**→ hypo→
Pae-2-2	NZ_AAKV01000112.1	-	77053	76905	**RNA**→ hypo→
Pae-3-2	NC_009656.1	+	4614401	4614549	**RNA**→ hypo→
Pae-5-2	NC_008463.1	+	5288768	5288916	**RNA**→ ←hypo
Pfl-1-2	NC_004129.6	+	5411071	5411217	**RNA**→ hypo→
Pfl-1-3	NC_004129.6	+	5412448	5412599	**RNA**→ hypo→
Pae-2-3	NZ_AAKV01000112.1	-	59821	59674	**RNA**→ hypo→ hypo→ COG0741: Soluble lytic murein transglycosylase and related regulatory proteins (some contain LysM/invasin domains)→

Pae-2-4	NZ_AAKV01000112.1	-	75975	75825	**RNA**→ hypo→
Pae-3-3	NC_009656.1	+	4615660	4615809	**RNA**→ hypo→
Pae-1-5	NZ_AAKW01000073.1	-	71679	71530	**RNA**→ hypo→
Pae-4-4	NZ_AAQW01000001.1	-	961023	960874	**RNA**→ hypo→
Pae-5-3	NC_008463.1	+	5290021	5290173	**RNA**→ hypo→
Aav-1-1	NC_008752.1	-	531393	531251	**RNA**→ RadC (COG2003)→
Psy-1-1	NC_007005.1	-	1657819	1657666	**RNA**→ ←hypo
Psy-1-2	NC_007005.1	-	1658497	1658347	**RNA**→ hypo→
Bps-3-1	NZ_AAHW02000025.1	+	23298	23445	**RNA**→ hypo→
Bps-1-1	NZ_AAMM02000010.1	-	124116	123970	**RNA**→ hypo→
Bps-2-2	NZ_ABBN01001396.1	-	102254	102109	**RNA**→ hypo→ hypo→
env-6	JCVI_SCAF_1096626436570	-	521	354	**RNA**→ ? (GOS14103005)→
Xca-3-2	NC_007508.1	-	2727905	2727768	**RNA**→ hypo→
Bxe-1-2	NC_007951.1	+	3572546	3572683	**RNA**→ hypo→
Bpe-1-5	NC_010170.1	+	1511445	1511582	**RNA**→ hypo→
Xfa-1-2	NC_002488.3	-	1689983	1689846	**RNA**→ hypo→
Sen-1-1	NZ_ABFH02000001.1	-	1099005	1098868	**RNA**→ ←FmrO (pfam07091)
env-7	JCVI_SCAF_1096628063987	-	1311	1177	**RNA**→ ? (GOS11302078)→ ? (GOS13495353)→
Neu-1-2	NC_008344.1	-	78726	78571	**RNA**→ RadC (pfam04002)→
Asp-1-4	NC_008782.1	-	1502364	1502228	**RNA**→ hypo→
Asp-1-5	NC_008782.1	-	1375308	1375172	**RNA**→ hypo→
Bpe-1-6	NC_010170.1	+	1113738	1113874	**RNA**→ hypo→ hypo→ hypo→ hypo→ hypo→
Cte-1-2	NZ_AAUJ01000001.1	+	1024305	1024441	**RNA**→ hypo (cons)→ hypo (cons)→
Xfa-1-3	NC_002488.3	-	1689085	1688948	**RNA**→ hypo→ hypo→
Xca-3-3	NC_007508.1	-	2727007	2726872	**RNA**→ hypo→ hypo→
Bxe-1-3	NC_007951.1	+	3573444	3573579	**RNA**→ hypo→ hypo→
Bpe-1-7	NC_010170.1	+	1512343	1512478	**RNA**→ hypo→ hypo→
env-8	JCVI_SCAF_1101667372641	-	455	320	**RNA**→ ? (GOS11195583)→
Bpe-1-8	NC_010170.1	-	4563132	4562996	**RNA**→ hypo→ hypo→
Dac-1-2	NC_010002.1	+	4545185	4545321	**RNA**→ hypo→ hypo→
Xca-1-2	NC_007086.1	-	2474704	2474565	**RNA**→ hypo→
Xca-3-4	NC_007508.1	-	2618952	2618813	**RNA**→ hypo→ hypo→
Har-1-1	NC_009138.1	+	1993906	1994045	**RNA**→ hypo→ hypo→
Pae-4-5	NZ_AAQW01000001.1	+	3510184	3510323	**RNA**→ hypo→ hypo→
env-9	JCVI_SCAF_1096628039386	-	392	253	**RNA**→ ? (GOS14058337)→
Rme-1-2	NC_007973.1	+	2527412	2527552	**RNA**→ hypo→ hypo→
Pla-1-2	NC_009719.1	+	3633128	3633268	**RNA**→ hypo→ hypo→
Pae-1-6	NZ_AAKW01000022.1	+	177365	177505	**RNA**→ hypo→ hypo→
Har-1-2	NC_009138.1	+	2030494	2030641	**RNA**→ RadC (pfam04002)→
Pae-4-6	NZ_AAQW01000001.1	+	3546772	3546919	**RNA**→ RadC (pfam04002)→
Bpe-1-9	NC_010170.1	-	4524050	4523902	**RNA**→ RadC (pfam04002)→
Dac-1-3	NC_010002.1	+	4572348	4572496	**RNA**→ RadC (pfam04002)→
Xca-1-3	NC_007086.1	-	2444672	2444524	**RNA**→ RadC (pfam04002)→
Xca-2-1	NC_010688.1	+	2848835	2848983	**RNA**→ hypo→ RadC (pfam04002)→
Pla-1-3	NC_009719.1	+	3663513	3663661	**RNA**→ RadC (pfam04002)→
Asp-1-6	NC_008782.1	-	1332744	1332597	**RNA**→ RadC (pfam04002)→
Asp-1-7	NC_008782.1	-	1463970	1463823	**RNA**→ RadC (pfam04002)→
Rme-1-3	NC_007973.1	+	2567864	2568011	**RNA**→ RadC (pfam04002)→
Pae-1-7	NZ_AAKW01000022.1	+	217815	217962	**RNA**→ RadC (pfam04002)→
env-10	sludgeOz_Contig9137	+	396	544	**RNA**→ RadC (pfam04002)→
Cte-1-3	NZ_AAUJ01000001.1	+	1065457	1065605	**RNA**→ RadC (pfam04002)→
Bpe-1-10	NC_010170.1	+	1148088	1148234	**RNA**→ RadC (pfam04002)→

Bxe-1-4	NC_007951.1	+	3604961	3605108	**RNA**→ RadC (pfam04002)→
Bpe-1-11	NC_010170.1	+	1547113	1547260	**RNA**→ RadC (COG2003)→ INT_pAE1 (cd01188)Phage_integr_N (pfam02899)→ INT_SG4 (cd01187)→ INT_SG5 (cd01190)→
Xca-3-5	NC_007508.1	-	2694905	2694758	**RNA**→ RadC (pfam04002)→
env-11	JCVI_SCAF_1096627691118	-	772	624	**RNA**→ RadC (pfam04002)? (GOS9318)→
Rpi-1-3	NZ_ABDZ01000015.1	-	36427	36279	**RNA**→ RadC (pfam04002)→
Aar-1-1	NC_006513.1	-	1434850	1434698	**RNA**→ RadC (pfam04002)→
Aar-1-2	NC_006513.1	-	1513825	1513675	**RNA**→ RadC (pfam04002)→
Asp-1-8	NC_008782.1	-	2312741	2312591	**RNA**→ RadC (pfam04002)→
env-12	sludgePhrap_Contig6275	+	675	825	**RNA**→
Mpe-1-1	NC_008825.1	+	2560890	2561040	**RNA**→ RadC (pfam04002)→
Tsp-1-2	NZ_ABKT01000008.1	+	90856	91006	**RNA**→ RadC (pfam04002)→
env-13	JCVI_SCAF_1096628079251	-	637	487	**RNA**→ RadC (pfam04002)? (GOS9318)→
Bps-2-3	NZ_ABBN01001396.1	-	97405	97260	**RNA**→ hypo→ hypo→ hypo→ conserved hypothetical plasmid protein→
Tsp-2-2	NZ_ABYM01000006.1	-	223191	223045	**RNA**→ RadC (pfam04002)→
Tsp-2-3	NZ_ABYM01000018.1	-	18883	18735	**RNA**→ hypo (cons)→
env-14	ABEF01007626.1	+	149	295	**RNA**→ ←hypo
Bpe-1-12	NC_010170.1	+	1375792	1375941	**RNA**→ hypo→
env-15	sludgePhrap_Contig13626	+	1466	1615	**RNA**→
Rpi-1-4	NZ_ABDZ01000015.1	-	58265	58115	**RNA**→ hypo (cons)→
env-16	sludgeOz_Contig6458	+	798	944	**RNA**→
Mpe-1-2	NC_008825.1	+	2530285	2530433	**RNA**→ hypo→ hypo→ tRNA-Ile→ tRNA-Ala→
Tsp-1-3	NZ_ABKT01000008.1	+	48927	49075	**RNA**→ hypo (cons)→ hypo (cons)→ hypo (cons)→ hypo (cons)→ COG4123 (COG4123)→
env-17	sludgeOz_Contig8211	+	192	340	**RNA**→ (,,,)→ (,,,)→
Asp-1-9	NC_008782.1	-	2362535	2362387	**RNA**→ hypo→ hypo→ hypo→ hypo→ COG4123 (COG4123)→
Aar-1-3	NC_006513.1	-	1470276	1470128	**RNA**→ hypo→
Bps-1-2	NZ_AAMM02000010.1	-	119452	119305	**RNA**→ hypo→ hypo→
Bps-3-2	NZ_AAHW02000025.1	+	26964	27111	**RNA**→ hypo→ hypo→
Aav-1-2	NC_008752.1	-	593292	593146	**RNA**→ hypo→ hypo→
Xca-1-4	NC_007086.1	-	2466860	2466713	**RNA**→ hypo→ hypo→ plasmid-related protein→
Xca-3-6	NC_007508.1	-	2602146	2601999	**RNA**→ hypo→ hypo→ hypo→
Xca-2-2	NC_010688.1	+	2821286	2821433	**RNA**→ hypo→ hypo→ hypo→
Bxe-1-5	NC_007951.1	+	3581310	3581455	**RNA**→ ←hypo
Bpe-1-13	NC_010170.1	+	1519640	1519785	**RNA**→ hypo→ hypo→ hypo→ hypo→ hypo→ hypo→ COG1041 (COG1041)→
Pla-1-4	NC_009719.1	+	3637020	3637165	**RNA**→ hypo→ hypo→ hypo→
Cte-1-4	NZ_AAUJ01000001.1	+	1029477	1029622	**RNA**→ hypo (cons)→
Bpe-1-14	NC_010170.1	+	1118341	1118486	**RNA**→ hypo→
Xfa-1-4	NC_002488.3	-	1678879	1678735	**RNA**→ hypo→ hypo→ hypo→
Asp-1-10	NC_008782.1	-	1497620	1497476	**RNA**→ hypo→ hypo→
Xca-3-7	NC_007508.1	-	2722273	2722129	**RNA**→ hypo→ hypo→ hypo→
Asp-1-11	NC_008782.1	-	1366182	1366037	**RNA**→ hypo→ hypo→ hypo→
Rme-1-4	NC_007973.1	+	2536533	2536678	**RNA**→ hypo→
Pae-1-8	NZ_AAKW01000022.1	+	186488	186633	**RNA**→ hypo→
env-18	JCVI_SCAF_1096628076737	+	1187	1332	**RNA**→ ? (GOS11989196)→
Har-1-3	NC_009138.1	+	1996181	1996326	**RNA**→ hypo→
Pae-4-7	NZ_AAQW01000001.1	+	3512459	3512604	**RNA**→ hypo→
Bpe-1-15	NC_010170.1	-	4559739	4559594	**RNA**→ hypo→
Dac-1-4	NC_010002.1	+	4548578	4548723	**RNA**→ hypo→
env-19	JCVI_SCAF_1096626463033	+	358	504	**RNA**→ ? (GOS11989196)→ ? (GOS8497)→
Aav-1-3	NC_008752.1	-	595771	595624	**RNA**→ hypo→ COG3415 (COG3415)rve (pfam00665)→
Xax-1-3	NC_003919.1	+	2619655	2619805	**RNA**→ hypo→ hypo→ plasmid-related protein→
Neu-1-3	NC_008344.1	-	107737	107590	**RNA**→ hypo→

Bpe-1-16	NC_010170.1	+	1370935	1371077	**RNA**→ hypo→
Aav-1-4	NC_008752.1	-	598085	597946	**RNA**→ hypo→ hypo→
Xax-1-4	NC_003919.1	+	2603319	2603458	**RNA**→ hypo→ hypo→
Aar-1-4	NC_006513.1	-	1477955	1477810	**RNA**→ hypo→
Mpe-1-3	NC_008825.1	+	2516445	2516590	**RNA**→ hypo→ hypo→
Tsp-1-4	NZ_ABKT01000008.1	+	38351	38496	**RNA**→ hypo (cons)→ hypo (cons)→
Asp-1-12	NC_008782.1	-	2365997	2365852	**RNA**→ hypo→ hypo→
Tsp-2-4	NZ_ABYM01000018.1	-	23385	23239	**RNA**→ hypo→ hypo (cons)→ hypo (cons)→
Rpi-1-5	NZ_ABDZ01000015.1	-	63434	63295	**RNA**→ hypo→ hypo (cons)→
env-20	JCVI_SCAF_1096628083237	+	389	531	**RNA**→ RadC (pfam04002)? (GOS9318)→ RadC (pfam04002)? (GOS9318)→ ? (GOS13072590)→

78.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 78.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *radC* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

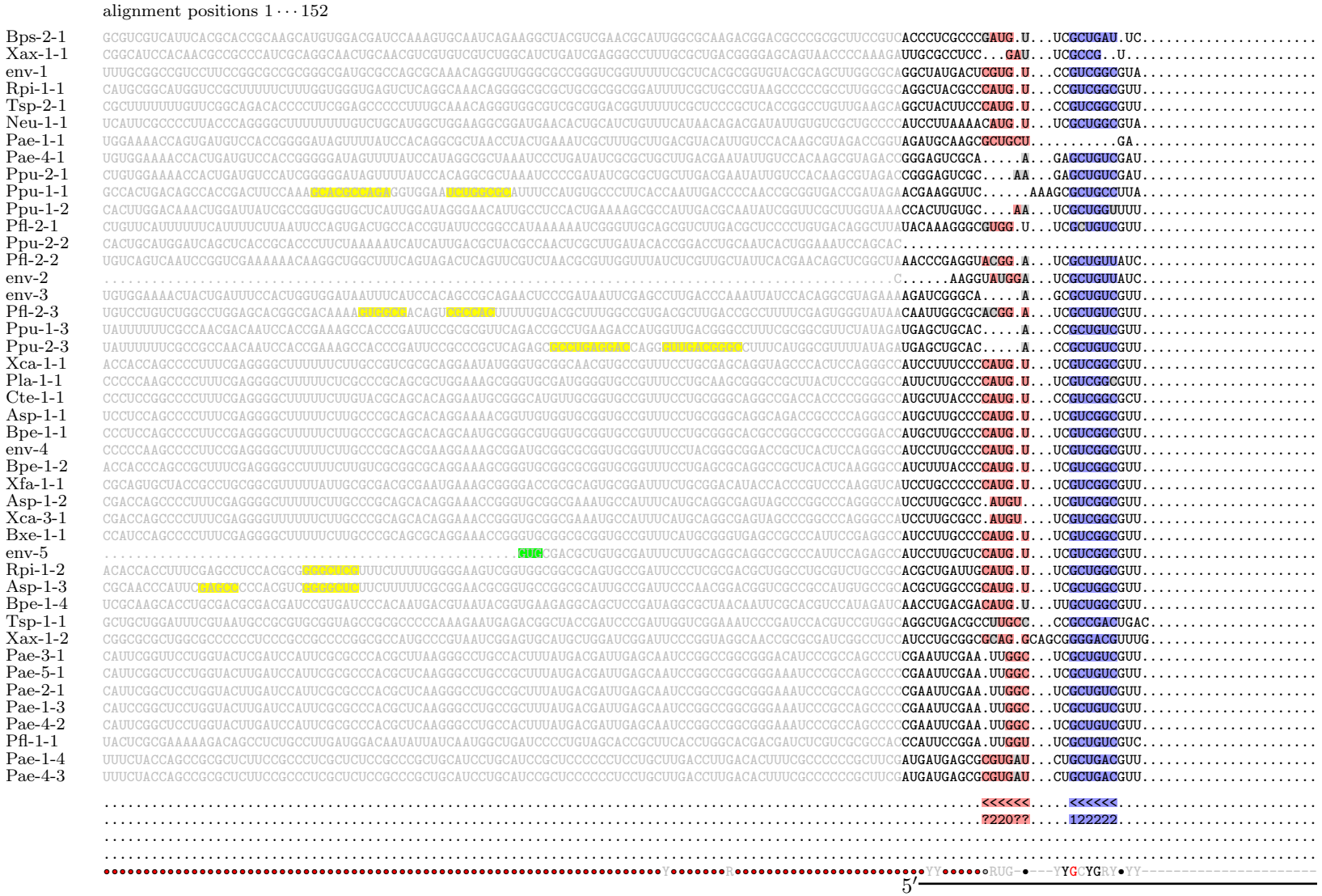
cd01187 (1) INT_SG4, DNA breaking-rejoining enzymes, integrase/recombinases subgroup 4, N- and C-terminal domains.
cd01188 (1) pAE1 and related integrases, DNA breaking-rejoining enzymes, integrase/recombinases, C-terminal domain.
cd01190 (1) INT_SG5, DNA breaking-rejoining enzymes, integrase/recombinases subgroup 5, N- and C-terminal domains.
COG1041 (1) Predicted DNA modification methylase [DNA replication, recombination, and repair]
COG1088 (1) dTDP-D-glucose 4,6-dehydratase [Cell envelope biogenesis, outer membrane]
COG1091 (1) dTDP-4-dehydrorhamnose reductase [Cell envelope biogenesis, outer membrane]
COG1209 (1) dTDP-glucose pyrophosphorylase [Cell envelope biogenesis, outer membrane]
COG2003 (2) DNA repair proteins [DNA replication, recombination, and repair]
COG3415 (1) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
COG4123 (4) Predicted O-methyltransferase [General function prediction only]
COG4227 (1) Antirestriction protein [DNA replication, recombination, and repair]
GOS11123422 (1) ?
GOS11195583 (1) ?
GOS11302078 (1) ?
GOS11886727 (1) ?
GOS11989196 (2) ?
GOS13066296 (1) ?
GOS13072590 (1) ?
GOS13495353 (2) ?
GOS14058337 (1) ?
GOS14103005 (1) ?
GOS8497 (1) ?
GOS9318 (4) ?
NOG13362 (1)
NOG18361 (1)
NOG18463 (1)
pfam00589 (2) Phage integrase family.
pfam00665 (1) Integrase core domain.
pfam02899 (1) Phage integrase, N-terminal SAM-like domain.
pfam04002 (31) RadC, DNA repair protein.
pfam07091 (1) Ribosomal RNA methyltransferase (FmrO).

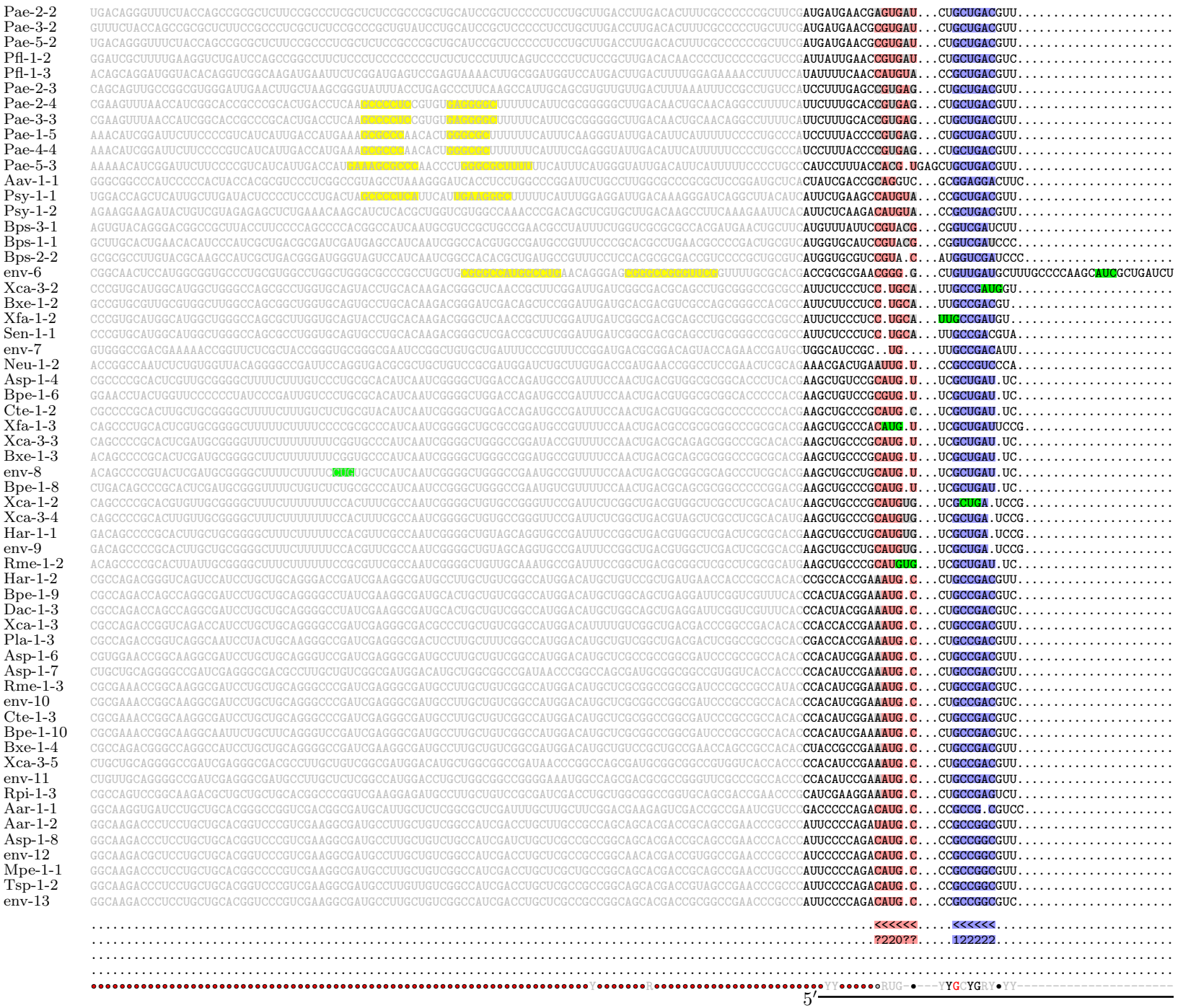
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Asp-1-5, Bpe-1-11, Bpe-1-13, Bpe-1-3, Bpe-1-5, Bpe-1-7, Dac-1-1, Dac-1-2, Dac-1-4, Pae-1-2, Pae-1-6, Pae-1-7, Pae-1-8, Pae-4-5, Pae-4-6, Pae-4-7, Pla-1-2, Rme-1-1, Xca-2-1

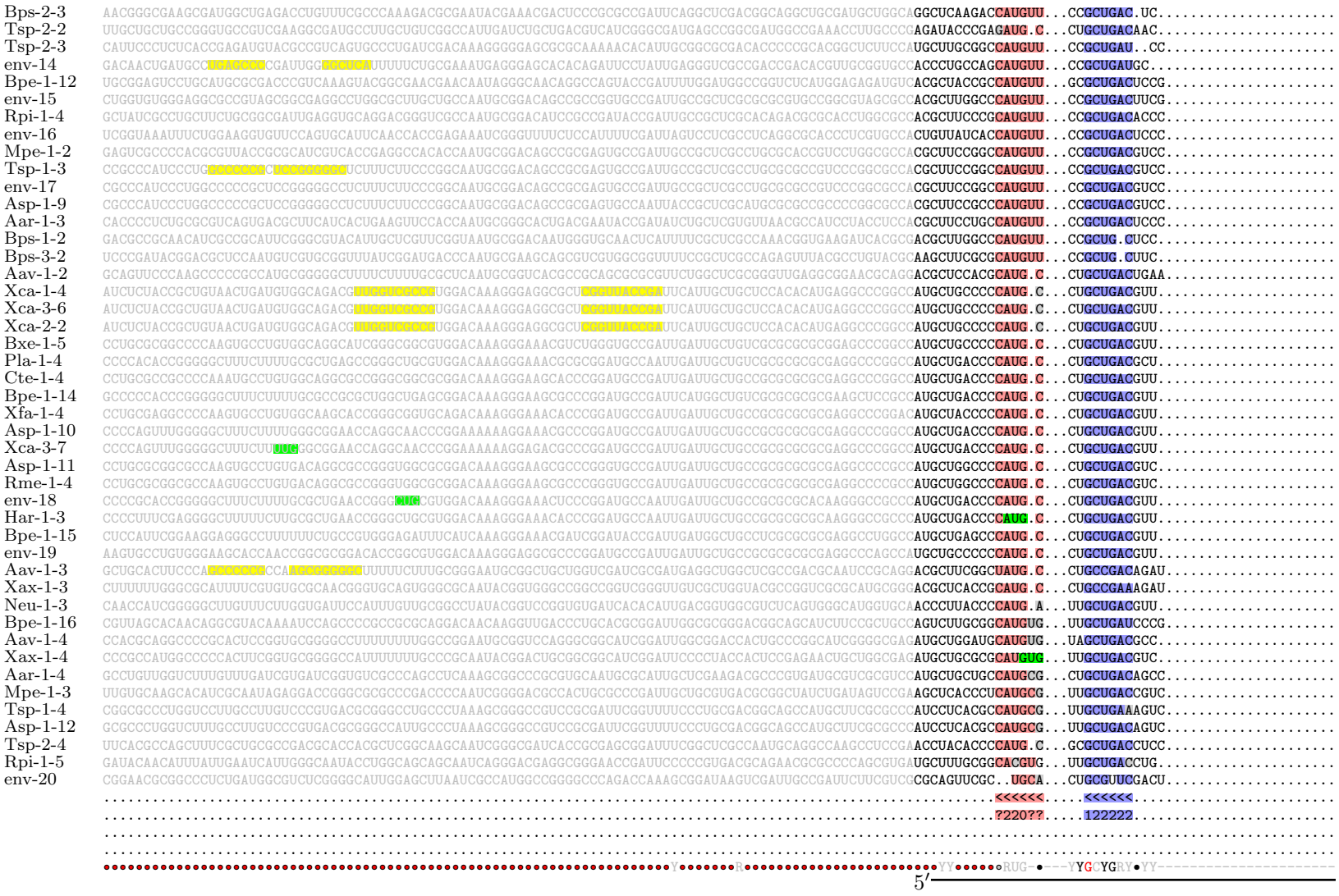
78.4 Multiple-sequence alignment

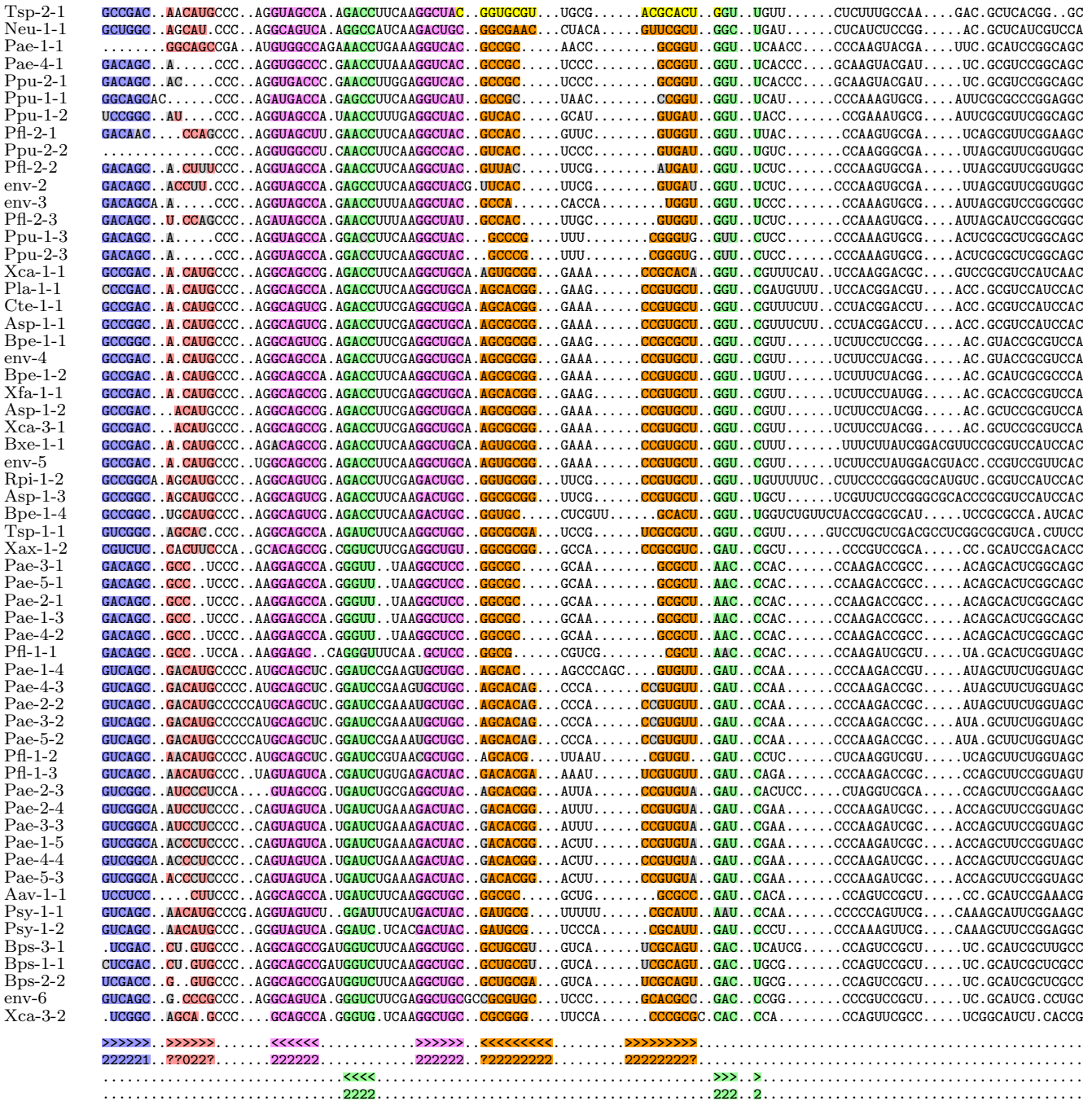
Each *radC* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 78.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *radC* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently mis-

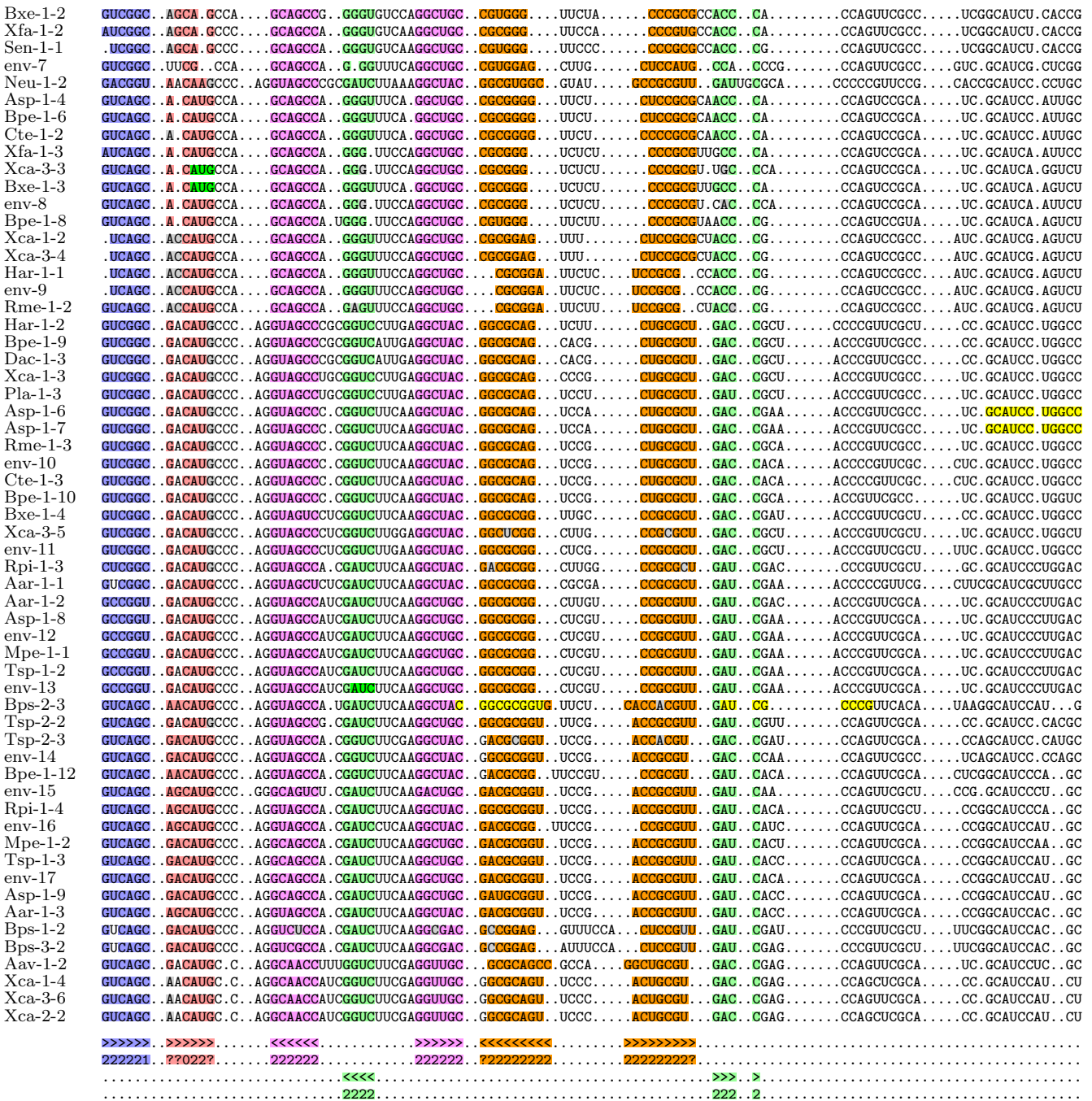
annotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



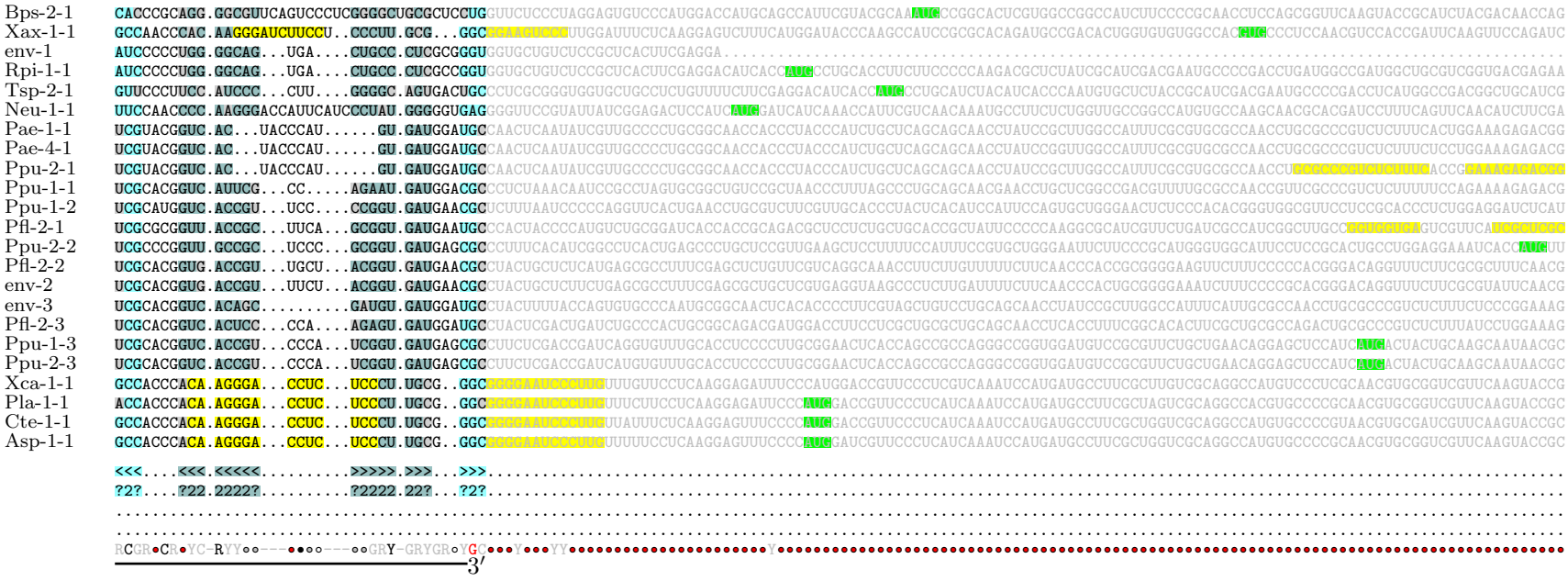

Pae-2-2
Pae-3-2
Pae-5-2
Pfl-1-2
Pfl-1-3
Pae-2-3
Pae-2-4
Pae-3-3
Pae-1-5
Pae-4-4
Pae-5-3
Aav-1-1
Psy-1-1
Psy-1-2
Bps-3-1
Bps-1-1
Bps-2-2
env-6
Xca-3-2
Bxe-1-2
Xfa-1-2
Sen-1-1
env-7
Neu-1-2
Asp-1-4
Bpe-1-6
Cte-1-2
Xfa-1-3
Xca-3-3
Bxe-1-3
env-8
Bpe-1-8
Xca-1-2
Xca-3-4
Har-1-1
env-9
Rme-1-2
Har-1-2
Bpe-1-9
Dac-1-3
Xca-1-3
Pla-1-3
Asp-1-6
Asp-1-7
Rme-1-3
env-10
Cte-1-3
Bpe-1-10
Bxe-1-4
Xca-3-5
env-11
Rpi-1-3
Aar-1-1
Aar-1-2
Asp-1-8
env-12
Mpe-1-1
Tsp-1-2
env-13
5′

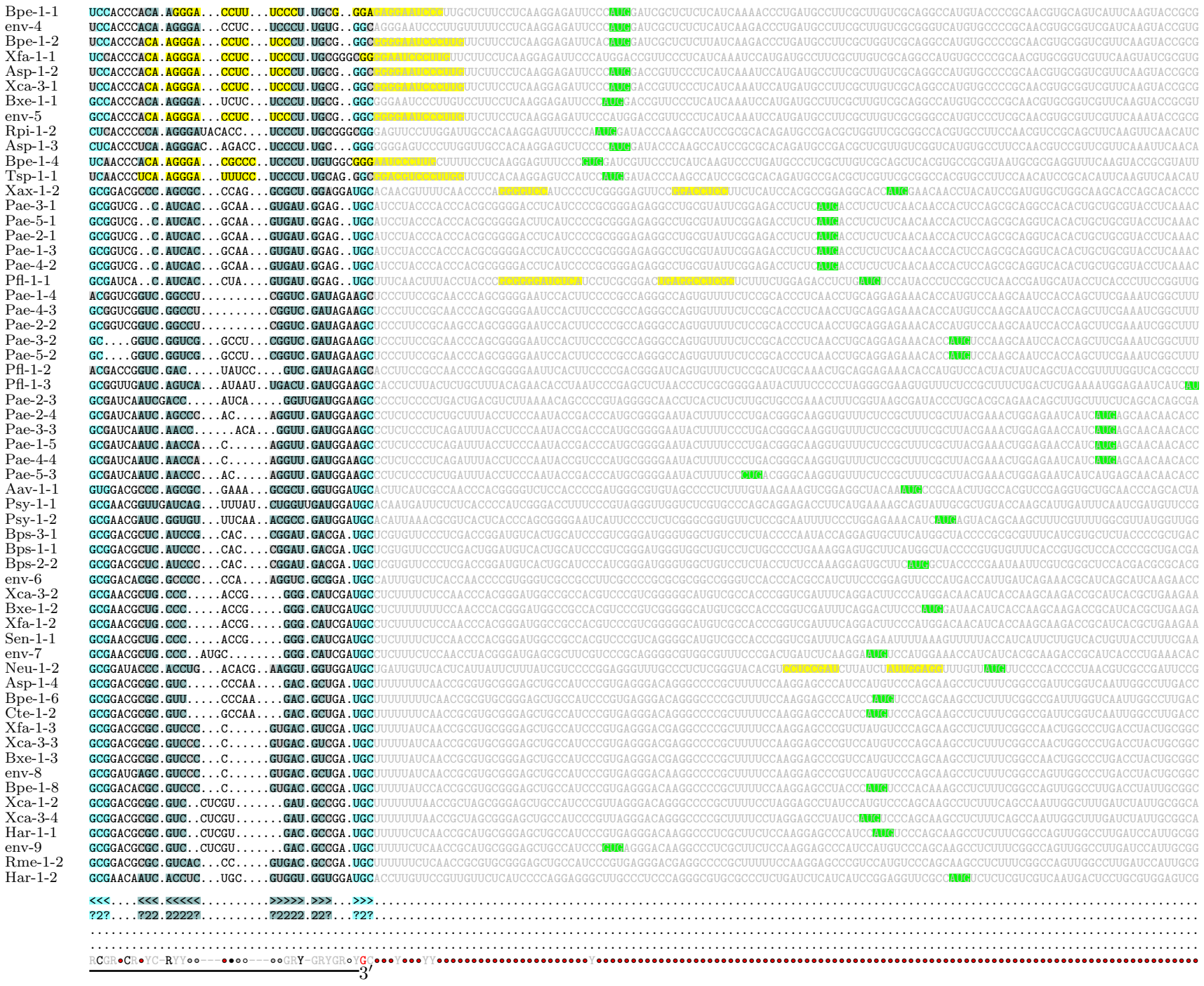

Bps-2-3
Tsp-2-2
Tsp-2-3
env-14
Bpe-1-12
env-15
Rpi-1-4
env-16
Mpe-1-2
Tsp-1-3
env-17
Asp-1-9
Aar-1-3
Bps-1-2
Bps-3-2
Aav-1-2
Xca-1-4
Xca-3-6
Xca-2-2
Bxe-1-5
Pla-1-4
Cte-1-4
Bpe-1-14
Xfa-1-4
Asp-1-10
Xca-3-7
Asp-1-11
Rme-1-4
env-18
Har-1-3
Bpe-1-15
env-19
Aav-1-3
Xax-1-3
Neu-1-3
Bpe-1-16
Aav-1-4
Xax-1-4
Aar-1-4
Mpe-1-3
Tsp-1-4
Asp-1-12
Tsp-2-4
Rpi-1-5
env-20
5′
alignment positions 153 · · · 274
Bps-2-1
Xax-1-1
env-1
Rpi-1-1

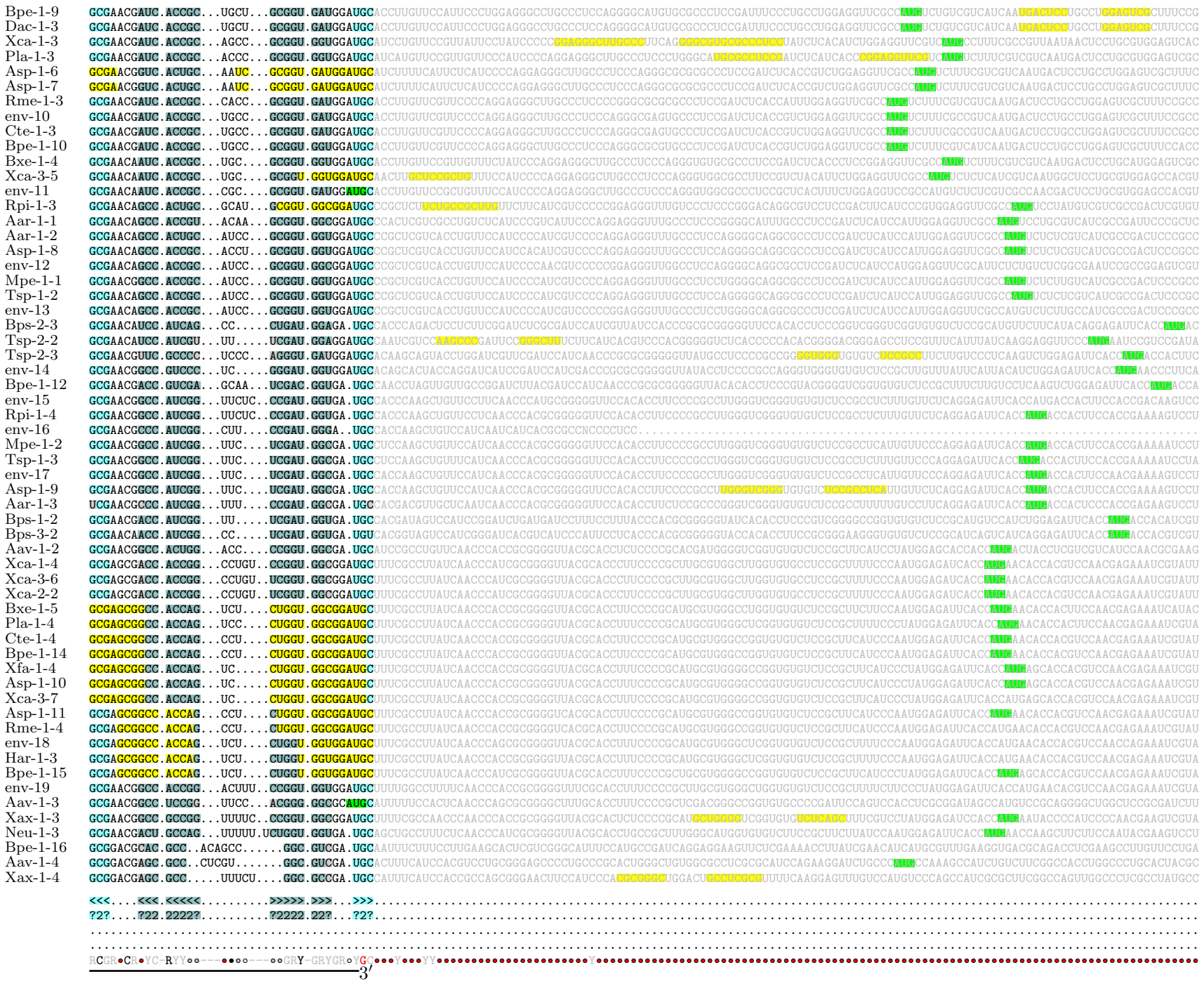

Tsp-2-1
Neu-1-1
Pae-1-1
Pae-4-1
Ppu-2-1
Ppu-1-1
Ppu-1-2
Pfl-2-1
Ppu-2-2
Pfl-2-2
env-2
env-3
Pfl-2-3
Ppu-1-3
Ppu-2-3
Xca-1-1
Pla-1-1
Cte-1-1
Asp-1-1
Bpe-1-1
env-4
Bpe-1-2
Xfa-1-1
Asp-1-2
Xca-3-1
Bxe-1-1
env-5
Rpi-1-2
Asp-1-3
Bpe-1-4
Tsp-1-1
Xax-1-2
Pae-3-1
Pae-5-1
Pae-2-1
Pae-1-3
Pae-4-2
Pfl-1-1
Pae-1-4
Pae-4-3
Pae-2-2
Pae-3-2
Pae-5-2
Pfl-1-2
Pfl-1-3
Pae-2-3
Pae-2-4
Pae-3-3
Pae-1-5
Pae-4-4
Pae-5-3
Aav-1-1
Psy-1-1
Psy-1-2
Bps-3-1
Bps-1-1
Bps-2-2
env-6
Xca-3-2

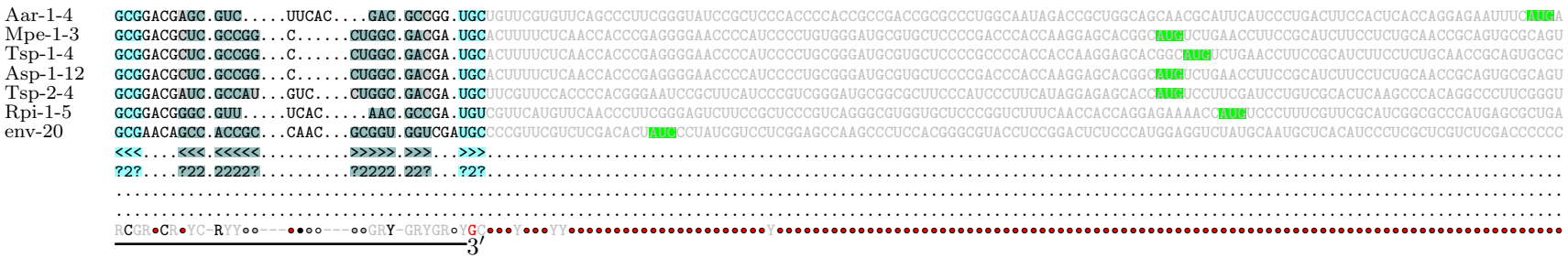

Bxe-1-2
Xfa-1-2
Sen-1-1
env-7
Neu-1-2
Asp-1-4
Bpe-1-6
Cte-1-2
Xfa-1-3
Xca-3-3
Bxe-1-3
env-8
Bpe-1-8
Xca-1-2
Xca-3-4
Har-1-1
env-9
Rme-1-2
Har-1-2
Bpe-1-9
Dac-1-3
Xca-1-3
Pla-1-3
Asp-1-6
Asp-1-7
Rme-1-3
env-10
Cte-1-3
Bpe-1-10
Bxe-1-4
Xca-3-5
env-11
Rpi-1-3
Aar-1-1
Aar-1-2
Asp-1-8
env-12
Mpe-1-1
Tsp-1-2
env-13
Bps-2-3
Tsp-2-2
Tsp-2-3
env-14
Bpe-1-12
env-15
Rpi-1-4
env-16
Mpe-1-2
Tsp-1-3
env-17
Asp-1-9
Aar-1-3
Bps-1-2
Bps-3-2
Aav-1-2
Xca-1-4
Xca-3-6
Xca-2-2

alignment positions 275 · · · 434

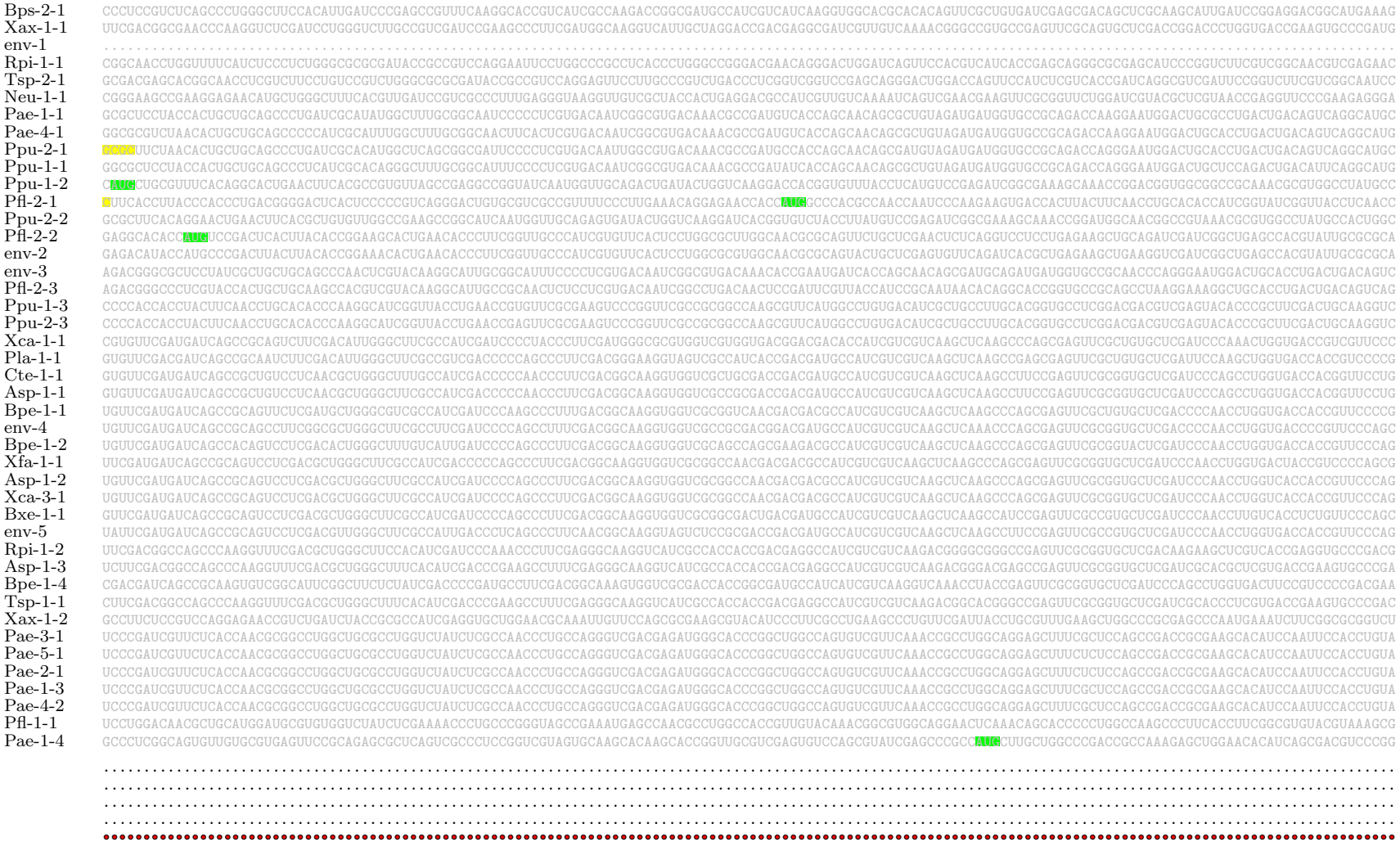


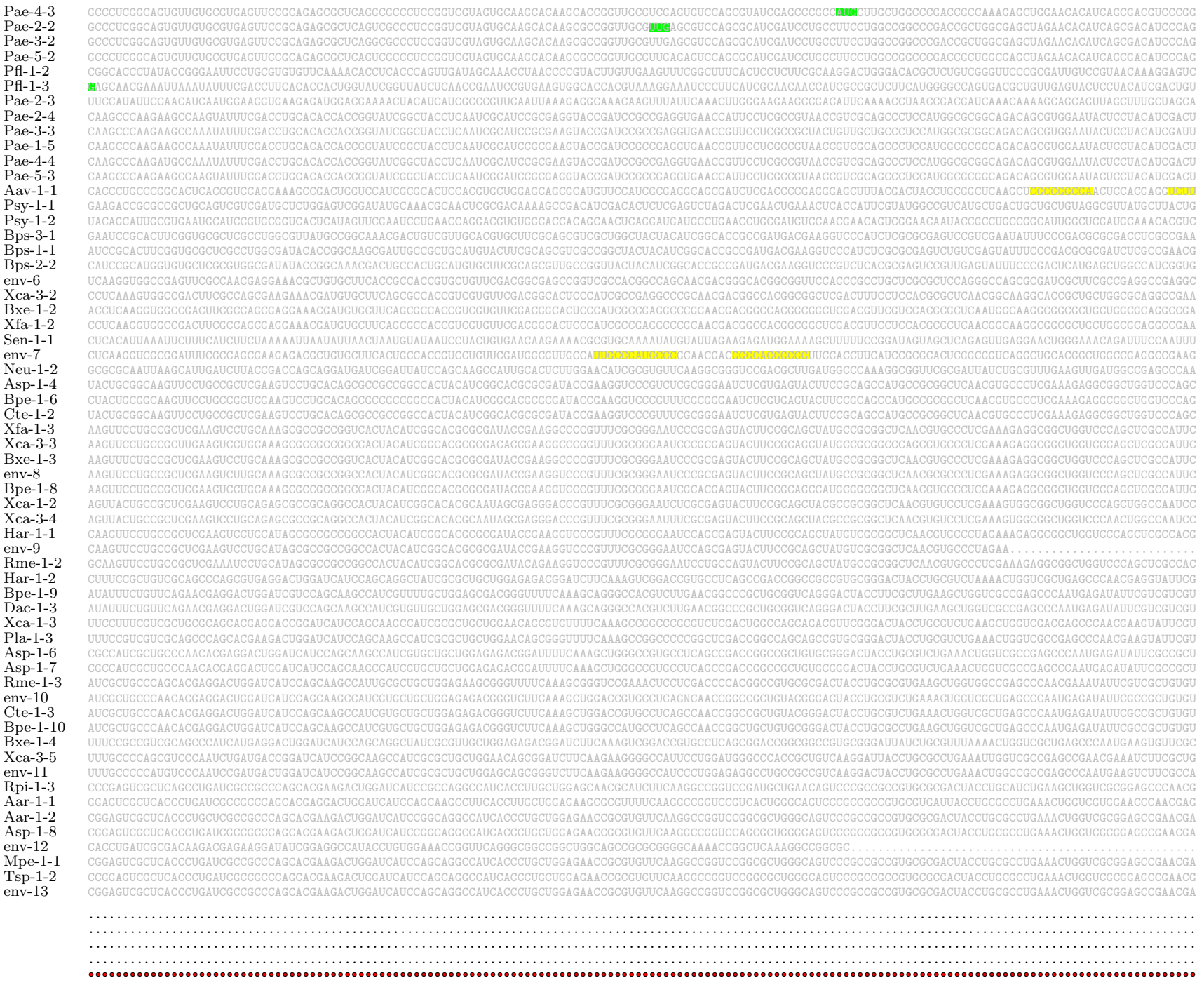

Bpe-1-1
env-4
Bpe-1-2
Xfa-1-1
Asp-1-2
Xca-3-1
Bxe-1-1
env-5
Rpi-1-2
Asp-1-3
Bpe-1-4
Tsp-1-1
Xax-1-2
Pae-3-1
Pae-5-1
Pae-2-1
Pae-1-3
Pae-4-2
Pfl-1-1
Pae-1-4
Pae-4-3
Pae-2-2
Pae-3-2
Pae-5-2
Pfl-1-2
Pfl-1-3
Pae-2-3
Pae-2-4
Pae-3-3
Pae-1-5
Pae-4-4
Pae-5-3
Aav-1-1
Psy-1-1
Psy-1-2
Bps-3-1
Bps-1-1
Bps-2-2
env-6
Xca-3-2
Bxe-1-2
Xfa-1-2
Sen-1-1
env-7
Neu-1-2
Asp-1-4
Bpe-1-6
Cte-1-2
Xfa-1-3
Xca-3-3
Bxe-1-3
env-8
Bpe-1-8
Xca-1-2
Xca-3-4
Har-1-1
env-9
Rme-1-2
Har-1-2
3′

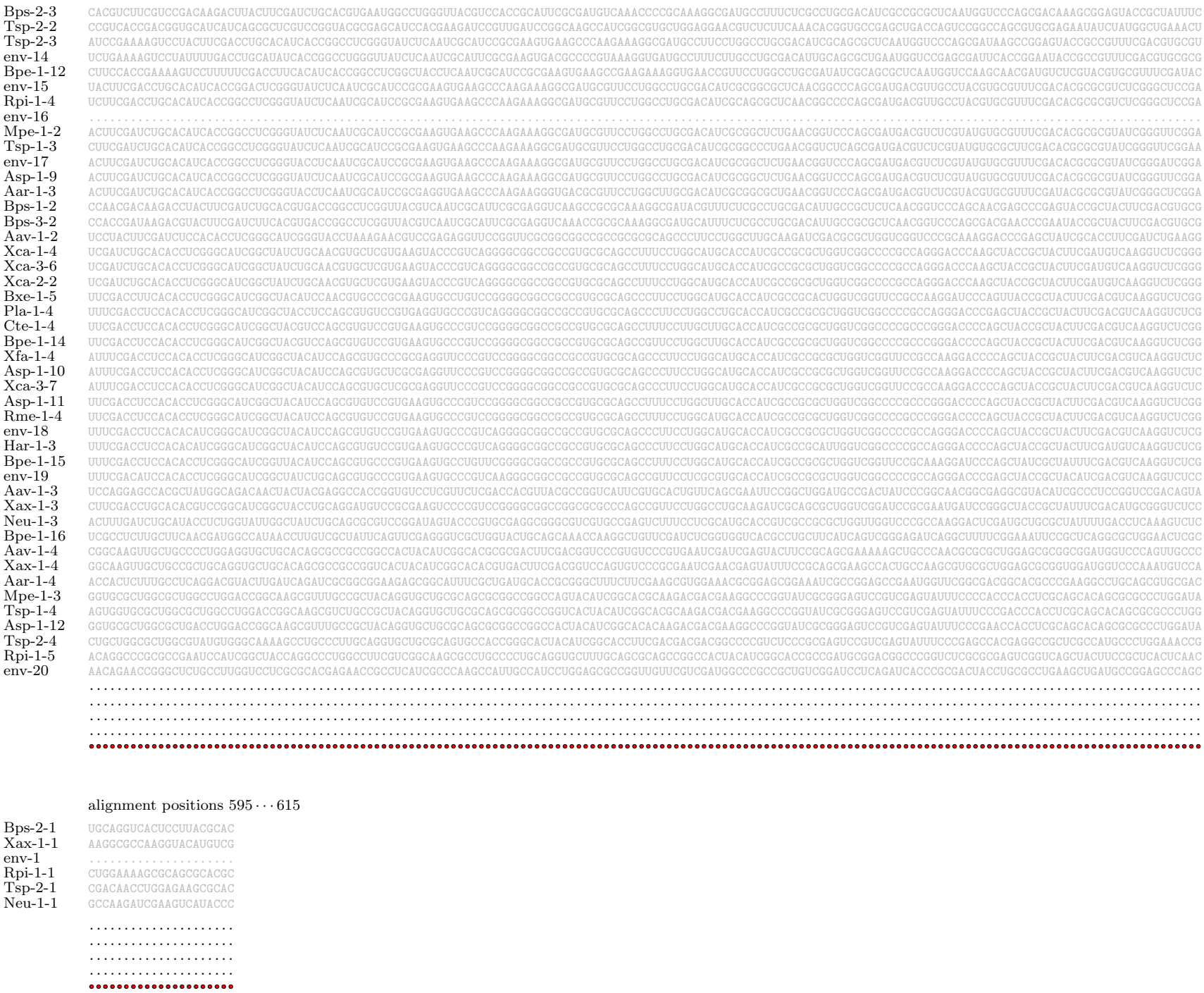

Bpe-1-9
Dac-1-3
Xca-1-3
Pla-1-3
Asp-1-6
Asp-1-7
Rme-1-3
env-10
Cte-1-3
Bpe-1-10
Bxe-1-4
Xca-3-5
env-11
Rpi-1-3
Aar-1-1
Aar-1-2
Asp-1-8
env-12
Mpe-1-1
Tsp-1-2
env-13
Bps-2-3
Tsp-2-2
Tsp-2-3
env-14
Bpe-1-12
env-15
Rpi-1-4
env-16
Mpe-1-2
Tsp-1-3
env-17
Asp-1-9
Aar-1-3
Bps-1-2
Bps-3-2
Aav-1-2
Xca-1-4
Xca-3-6
Xca-2-2
Bxe-1-5
Pla-1-4
Cte-1-4
Bpe-1-14
Xfa-1-4
Asp-1-10
Xca-3-7
Asp-1-11
Rme-1-4
env-18
Har-1-3
Bpe-1-15
env-19
Aav-1-3
Xax-1-3
Neu-1-3
Bpe-1-16
Aav-1-4
Xax-1-4
3′

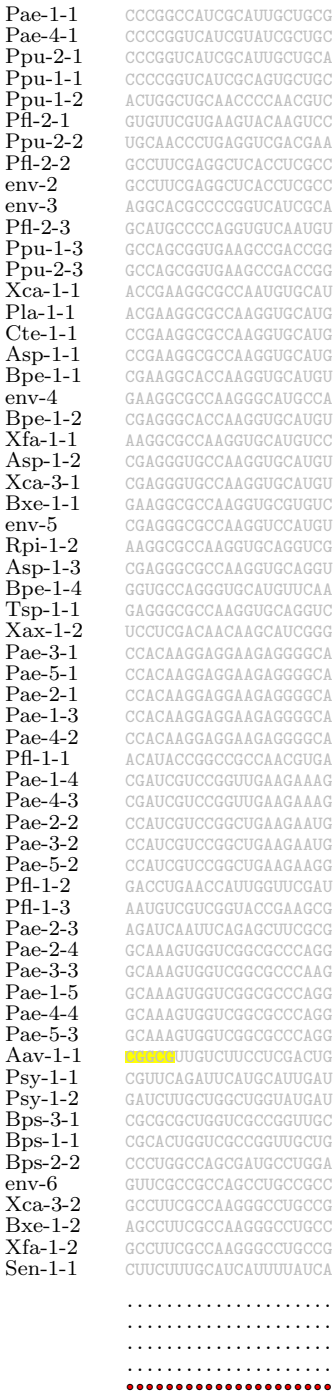


alignment positions 435 · · · 594




Pae-4-3
Pae-2-2
Pae-3-2
Pae-5-2
Pfl-1-2
Pfl-1-3
Pae-2-3
Pae-2-4
Pae-3-3
Pae-1-5
Pae-4-4
Pae-5-3
Aav-1-1
Psy-1-1
Psy-1-2
Bps-3-1
Bps-1-1
Bps-2-2
env-6
Xca-3-2
Bxe-1-2
Xfa-1-2
Sen-1-1
env-7
Neu-1-2
Asp-1-4
Bpe-1-6
Cte-1-2
Xfa-1-3
Xca-3-3
Bxe-1-3
env-8
Bpe-1-8
Xca-1-2
Xca-3-4
Har-1-1
env-9
Rme-1-2
Har-1-2
Bpe-1-9
Dac-1-3
Xca-1-3
Pla-1-3
Asp-1-6
Asp-1-7
Rme-1-3
env-10
Cte-1-3
Bpe-1-10
Bxe-1-4
Xca-3-5
env-11
Rpi-1-3
Aar-1-1
Aar-1-2
Asp-1-8
env-12
Mpe-1-1
Tsp-1-2
env-13


Bps-2-3
Tsp-2-2
Tsp-2-3
env-14
Bpe-1-12
env-15
Rpi-1-4
env-16
Mpe-1-2
Tsp-1-3
env-17
Asp-1-9
Aar-1-3
Bps-1-2
Bps-3-2
Aav-1-2
Xca-1-4
Xca-3-6
Xca-2-2
Bxe-1-5
Pla-1-4
Cte-1-4
Bpe-1-14
Xfa-1-4
Asp-1-10
Xca-3-7
Asp-1-11
Rme-1-4
env-18
Har-1-3
Bpe-1-15
env-19
Aav-1-3
Xax-1-3
Neu-1-3
Bpe-1-16
Aav-1-4
Xax-1-4
Aar-1-4
Mpe-1-3
Tsp-1-4
Asp-1-12
Tsp-2-4
Rpi-1-5
env-20
alignment positions 595 · · · 615
Bps-2-1
Xax-1-1
env-1
Rpi-1-1
Tsp-2-1
Neu-1-1


Pae-1-1 CCCGGCCAUCGCAUUGCUGCG
Pae-4-1 CCCCGGUCAUCGUAUCGCUGC
Ppu-2-1 CCCGGUCAUCGCAUUGCUGCA
Ppu-1-1 CCCCGGUCAUCGCAGUGCUGC
Ppu-1-2 ACUGGCUGCAACCCCAACGUC
Pfl-2-1 GUGUUCGUGAAGUACAAGUCC
Ppu-2-2 UGCAACCCUGAGGUCGACGAA
Pfl-2-2 GCCUUCGAGGCUCACCUCGCC
env-2 GCCUUCGAGGCUCACCUCGCC
env-3 AGGCACGCCCCGGUCAUCGCA
Pfl-2-3 GCAUGCCCCAGGUGUCAAUGU
Ppu-1-3 GCCAGCGGUGAAGCCGACCGG
Ppu-2-3 GCCAGCGGUGAAGCCGACCGG
Xca-1-1 ACCGAAGGCGCCAAUGUGCAU
Pla-1-1 ACGAAGGCGCCAAGGUGCAUG
Cte-1-1 CCGAAGGCGCCAAGGUGCAUG
Asp-1-1 CCGAAGGCGCCAAGGUGCAUG
Bpe-1-1 CGAAGGCACCAAGGUGCAUGU
env-4 GAAGGCGCCAAGGGCAUGCCA
Bpe-1-2 CGAGGGCACCAAGGUGCAUGU
Xfa-1-1 AAGGCGCCAAGGUGCAUGUCC
Asp-1-2 CGAGGGUGCCAAGGUGCAUGU
Xca-3-1 CGAGGGUGCCAAGGUGCAUGU
Bxe-1-1 GAAGGCGCCAAGGUGCGUGUC
env-5 CGAGGGCGCCAAGGUCCAUGU
Rpi-1-2 AAGGCGCCAAGGUGCAGGUCG
Asp-1-3 CGAGGGCGCCAAGGUGCAGGU
Bpe-1-4 GGUGCCAGGGUGCAUGUUCAA
Tsp-1-1 GAGGGCGCCAAGGUGCAGGUC
Xax-1-2 UCCUCGACAACAAGCAUCGGG
Pae-3-1 CCACAAGGAGGAAGAGGGGCA
Pae-5-1 CCACAAGGAGGAAGAGGGGCA
Pae-2-1 CCACAAGGAGGAAGAGGGGCA
Pae-1-3 CCACAAGGAGGAAGAGGGGCA
Pae-4-2 CCACAAGGAGGAAGAGGGGCA
Pfl-1-1 ACAUACCGGCCGCCAACGUGA
Pae-1-4 CGAUCGUCCGGUUGAAGAAAG
Pae-4-3 CGAUCGUCCGGUUGAAGAAAG
Pae-2-2 CCAUCGUCCGGCUGAAGAAUG
Pae-3-2 CCAUCGUCCGGCUGAAGAAUG
Pae-5-2 CCAUCGUCCGGCUGAAGAAGG
Pfl-1-2 GACCUGAACCAUUGGUUCGAU
Pfl-1-3 AAUGUCGUCGGUACCGAAGCG
Pae-2-3 AGAUCAAUUCAGAGCUUCGCG
Pae-2-4 GCAAAGUGGUCGGCGCCCAGG
Pae-3-3 GCAAAGUGGUCGGCGCCCAAG
Pae-1-5 GCAAAGUGGUCGGCGCCCAGG
Pae-4-4 GCAAAGUGGUCGGCGCCCAGG
Pae-5-3 GCAAAGUGGUCGGCGCCCAGG
Aav-1-1 UGGGGUUGUCUUCCUCGACUG
Psy-1-1 CGUUCAGAUUCAUGCAUUGAU
Psy-1-2 GAUCUUGCUGGCUGGUAUGAU
Bps-3-1 CGCGCGCUGGUCGCCGGUUGC
Bps-1-1 CGCACUGGUCGCCGGUUGCUG
Bps-2-2 CCCUGGCCAGCGAUGCCUGGA
env-6 GUUCGCCGCCAGCCUGCCGCC
Xca-3-2 GCCUUCGCCAAGGGCCUGCCG
Bxe-1-2 AGCCUUCGCCAAGGGCCUGCC
Xfa-1-2 GCCUUCGCCAAGGGCCUGCCG
Sen-1-1 CUUCUUUGCAUCAUUUUAUCA

```
env-7       CCUUCGCCAGGAGCCUGCCCC
Neu-1-2     CGAGGUCUUUGUCGCGGUAUU
Asp-1-4     UCGCCAUUCCCUGAUCCAACU
Bpe-1-6     CUCGCCAUUCCCUGAUCCAAC
Cte-1-2     UCGCCGUUCCCUGAUCCAACU
Xfa-1-3     CCUGAUCCAACUGGAGGAAUC
Xca-3-3     CCUGAUCCAACUGGAGGAAUC
Bxe-1-3     CCUGAUCCAACUGGAGGAAUC
env-8       CCUGAUCCAACUGGAGAAAUC
Bpe-1-8     CCUGAUCCAACUGGAGGAAUC
Xca-1-2     CUGAUCCAACUAGAGGAGCCA
Xca-3-4     CUGAUCCAACUAGAGGAGCCA
Har-1-1     CCCUGAUCCAACUGGAGGAAU
env-9       ......................
Rme-1-2     ACCCUGAUCCAACUGGAGGAA
Har-1-2     CAAUCGUUUUCCUGGACAGCA
Bpe-1-9     GUUCAUGAACAGCAUGCACGA
Dac-1-3     GUUCAUGAACAGCAUGCACGA
Xca-1-3     CGUUGUGUUCAUGGACAAUCU
Pla-1-3     CGUUGUGUUCAUGGACAGCCU
Asp-1-6     GUGUUCCUGGACAGCAUGCAC
Asp-1-7     GUGUUCCUGGACAGCAUGCAC
Rme-1-3     UCCUGGACAGCAUGCACCAGG
env-10      UCCUGGACAGCCUGCACCAGG
Cte-1-3     UCCUGGACAGCCUGCACCAGG
Bpe-1-10    UCCUGGACAGCAUGCACCAGG
Bxe-1-4     AGCCGUGUUCCUGGACAGCAU
Xca-3-5     UCGUUUUCCUCAACAACAAGC
env-11      UCGUGUUCCUCAACAGCAUGC
Rpi-1-3     AGGUCUUCUCCGUCGUGUUCC
Aar-1-1     AUCUUCGCCGUCGUGUUCCUC
Aar-1-2     AAUCUUCGCCGUCGUGUUUCU
Asp-1-8     AAUCUUCGCCGUCGUGUUUCU
env-12      ......................
Mpe-1-1     AAUCUUCGCCGUCGUGUUUCU
Tsp-1-2     AAAUCUUCGCCGUCGUGUUUC
env-13      AAUCUUCGCCGUCGUGUUUCU
Bps-2-3     GACGUCCGGGUAUCCGGCGCC
Tsp-2-2     CGCCAGCGAAGCACUGGAGGU
Tsp-2-3     GUAUCCGGCAGCGAAGCGCAG
env-14      UUUCCGGCAGCGAGGCCCAGC
Bpe-1-12    ACGUGUUUCGGGGCACCGAUGC
env-15      AGCGCAACACCUGGUGCGCCG
Rpi-1-4     AGCGCAGCACCUGGUGCGCCG
env-16      ......................
Mpe-1-2     AGCGCAGCACCUGGUGCGCCG
Tsp-1-3     GCGCAGCACCUGGUGCGCCGC
env-17      AGCGCAGCACCUGGUGCGCCG
Asp-1-9     AGCGCAGCACCUGGUGCGCCG
Aar-1-3     AGCGCAGCACCUGGUGCGUCG
Bps-1-2     CGUCUCGGGCGCCGAUGCGCA
Bps-3-2     CGUCUCGGGUGCCGAUGCCCA
Aav-1-2     UCUCCGGCGCCAAGGCCAAGG
Xca-1-4     UGCCGAGGCCAAGAAGCUCGU
Xca-3-6     UGCCGAGGCCAAGAAGCUCGU
Xca-2-2     UGCCGAGGCCAAGAAGCUCGU
Bxe-1-5     GUGCCGAGGCCAAGAAACUGG
Pla-1-4     GGUGCCGAGGCCAAGAAGCUC
Cte-1-4     GUGCCGAGGCCAAGAAGCUCG
Bpe-1-14    GUGCCGAGGCCAAGAAGCUCG

            ......................
            ......................
            ......................
            ......................
            ●●●●●●●●●●●●●●●●●●●●●
```

```
Xfa-1-4     GGGUGCCGAGGCCAAGAAACU
Asp-1-10    GGGUGCCGAGGCCAAGAAACU
Xca-3-7     GGGUGCCGAGGCCAAGAAACU
Asp-1-11    GAGCCGAGGCCAAGAAGCUCG
Rme-1-4     GUGCCGAGGCCAAGAAGCUCG
env-18      GGUGCCGAGGCCAAGAAGCUC
Har-1-3     GGUGCCGAGGCCAAGAAGCUC
Bpe-1-15    GGUGCCGAGGCCAAGAAGCUG
env-19      GGUGCCGAGGCCAAGAAGCUG
Aav-1-3     CACCCAGCUGGGACGAUCUGC
Xax-1-3     GGUGCCGAGGCCAAGAAGCUG
Neu-1-3     UGGCGCUGAUGCCAAGACACU
Bpe-1-16    AACUGAUCUGCAUAGGGCGCU
Aav-1-4     AACCCCUGAUCCCGCGAAGGA
Xax-1-4     AUCCCUGAUCCGCAGAAGGAG
Aar-1-4     CAGCAAGUGGGAUCUGCGCCC
Mpe-1-3     CCCAUGCCUGGACGCAGCGCG
Tsp-1-4     AUACCCAUGCCUGGACACAGC
Asp-1-12    CCCAUGCCUGGACACAGCGCG
Tsp-2-4     GCGCAUGGACGCAGCGCCAGU
Rpi-1-5     GCCGCCGACCACGCCCUGAAA
env-20      GAAGUCUUCGUCGCCUUAUUU
            .....................
            .....................
            .....................
            .....................
            ●●●●●●●●●●●●●●●●●●●●●
```

79 Rhizobiales-1

79.1 Taxa

The taxonomy of each organism containing a putative Rhizobiales-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Rhizobiales-1 RNA in Sections 79.2 and 79.4.

abbrev. of hits	taxonomy of species
Ccr-1-1	Bacteria Proteobacteria Alphaproteobacteria Caulobacterales Caulobacteraceae *Caulobacter crescentus* CB15
Csp-1-1 to Csp-1-8	Bacteria Proteobacteria Alphaproteobacteria Caulobacterales Caulobacteraceae *Caulobacter sp.* K31
Bja-1-1 to Bja-1-27	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Bradyrhizobium japonicum* USDA 110
Bsp-1-1 to Bsp-1-16	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Bradyrhizobium sp.* BTAi1
Bsp-2-1 to Bsp-2-2	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Bradyrhizobium sp.* ORS278
Nha-1-1 to Nha-1-92	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Nitrobacter hamburgensis* X14
Nsp-1-1 to Nsp-1-57	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Nitrobacter sp.* Nb-311A
Nwi-1-1 to Nwi-1-74	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Nitrobacter winogradskyi* Nb-255
Rpa-1-1 to Rpa-1-2	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* BisA53
Rpa-2-1 to Rpa-2-7	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* BisB18
Rpa-3-1 to Rpa-3-41	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* BisB5
Rpa-4-1 to Rpa-4-6	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* HaA2
Bab-1-1 to Bab-1-39	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Brucellaceae *Brucella abortus bv.* 1 str. 9-941
Bme-1-1 to Bme-1-38	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Brucellaceae *Brucella melitensis* 16M
Bme-2-1 to Bme-2-39	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Brucellaceae *Brucella melitensis biovar* Abortus 2308
Bov-1-1 to Bov-1-41	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Brucellaceae *Brucella ovis* ATCC 25840
Bsu-1-1 to Bsu-1-43	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Brucellaceae *Brucella suis* 1330
Oan-1-1 to Oan-1-71	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Brucellaceae *Ochrobactrum anthropi* ATCC 49188
Mch-1-1 to Mch-1-16	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Methylobacteriaceae *Methylobacterium chloromethanicum* CM4
Mex-1-1 to Mex-1-20	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Methylobacteriaceae *Methylobacterium extorquens* PA1
Mlo-1-1 to Mlo-1-25	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Phyllobacteriaceae *Mesorhizobium loti* MAFF303099
Msp-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Phyllobacteriaceae *Mesorhizobium sp.* BNC1

Atu-1-1 to Atu-1-2	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Rhizobiaceae Rhizobium/Agrobacterium group *Agrobacterium tumefaciens str.* C58
Ret-1-1 to Ret-1-10	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Rhizobiaceae Rhizobium/Agrobacterium group *Rhizobium etli* CFN 42
Rle-1-1 to Rle-1-48	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Rhizobiaceae Rhizobium/Agrobacterium group *Rhizobium leguminosarum bv. viciae* 3841
Sme-1-1 to Sme-1-4	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Rhizobiaceae Sinorhizobium/Ensifer group *Sinorhizobium medicae* WSM419
Sme-2-1 to Sme-2-4	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Rhizobiaceae Sinorhizobium/Ensifer group *Sinorhizobium meliloti* 1021
Mma-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodospirillales Rhodospirillaceae *Magnetospirillum magnetotacticum* MS-1
Nar-1-1	Bacteria Proteobacteria Alphaproteobacteria Sphingomonadales Sphingomonadaceae *Novosphingobium aromaticivorans* DSM 12444
env-1 to env-59	environmental samples

79.2 Gene contexts

Each Rhizobiales-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Rhizobiales-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 79.3. The accession number of the sequence containing each Rhizobiales-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 79.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Sme-2-1	NC_003037.1	-	766175	766114	FixA (COG2086)→ **RNA**→ FixB (COG2025)→
Sme-2-2	NC_003047.1	+	2851442	2851503	FixA (COG2086)→ **RNA**→ FixB (COG2025)→
Ret-1-1	NC_007761.1	-	2610579	2610520	←ProX (COG2113) **RNA**→ CobQ (COG1492)→
Atu-1-1	NC_003062.1	+	84150	84211	←RsmC (COG2813) **RNA**→ ←Pnp (COG1185)
Atu-1-2	NC_003304.1	+	84351	84412	←RsmC (COG2813) **RNA**→ ←Pnp (COG1185)
Nar-1-1	NC_007794.1	+	2330843	2330904	←SSB (pfam00436) **RNA**→ ←hypo
env-1	soil_XZS29820_x1	-	211	151	BioF (COG0156)→ **RNA**→
env-2	soil_XZS100099_b1	+	726	786	hypo→ **RNA**→
env-3	soil_XZS31810_y1	-	903	843	**RNA**→ ←hypo
Mlo-1-1	NC_002678.2	+	6282032	6282093	PulA (COG1523)→ **RNA**→ Aamy (smart00642)→
Bsu-1-1	NC_004310.3	+	414890	414951	hypo→ **RNA**→ ←CaiD (COG1024)
Bme-1-1	NC_003317.1	-	1568007	1567946	hypo→ **RNA**→ ←CaiD (COG1024)
Bov-1-1	NC_009505.1	+	436880	436941	hypo→ **RNA**→ ←CaiD (COG1024)
Nwi-1-1	NC_007406.1	+	1186034	1186096	←probable transcriptional regulator **RNA**→ ThiD (COG0351)→
Nwi-1-2	NC_007406.1	+	1345493	1345555	Lpd (COG1249)→ **RNA**→ DAHP_synth_2 (pfam01474)→
Nwi-1-3	NC_007406.1	-	185179	185117	←COG2863 (COG2863) **RNA**→ ←AICARFT_IMPCHas (pfam01808)MGS (pfam02142)
Nwi-1-4	NC_007406.1	+	1460070	1460132	PRA-CH (pfam01502)→ **RNA**→ hypo→
Nwi-1-5	NC_007406.1	+	2007964	2008026	←SecG (pfam03840) **RNA**→ ←TpiA (COG0149)
Nwi-1-6	NC_007406.1	-	2722478	2722416	PrsA (COG0462)→ **RNA**→ ←hypo
Nwi-1-7	NC_007406.1	-	3225557	3225495	←COG2133 (COG2133) **RNA**→ ←GST_C_1 (cd03179)GST_N_4 (cd03056)
Nwi-1-8	NC_007406.1	-	112662	112600	Soj (COG1192)→ **RNA**→ Spo0J (COG1475)→
Rle-1-1	NC_008380.1	+	3097635	3097696	←DUF1501 (pfam07394) **RNA**→ ←COG5267 (COG5267)
env-4	ABEF01031871.1	-	759	698	hypo→ **RNA**→ ←COG1051 (COG1051)
env-5	soil_Contig27976	-	515	455	←hypo **RNA**→ Predicted integral membrane protein→
Nha-1-1	NC_007964.1	+	1409568	1409629	←BaeS (COG0642)PAS (smart00091) **RNA**→ ←BaeS (COG0642)PAS (smart00091)
Nwi-1-9	NC_007406.1	-	1914706	1914645	H_PPase (pfam03030)→ **RNA**→ ←hypo
Nwi-1-10	NC_007406.1	-	1842019	1841958	←hypo **RNA**→ adh_short (pfam00106)→
Oan-1-1	NC_009668.1	+	633686	633747	MgtE (COG2239)→ **RNA**→ TypA (COG1217)→
env-6	soil_XZS57818_b1	+	562	623	←DUF140 (pfam02405) **RNA**→ ←Ala_racemase_C (pfam00842)

env-7	soil_Contig9153	-	95	34	←77 Permeases of the major facilitator superfamily **RNA**→
env-8	soil_XZS60340_g1	+	442	504	hypo→ **RNA**→ 1-acyl-sn-glycerol-3-phosphate acyltransferase→
env-9	soil_XZS30534_y1	+	444	507	←hypo **RNA**→ hypo→
env-10	soil_Contig38734	-	303	242	NuoN (COG1007)→ **RNA**→ Biotin-(acetyl-CoA carboxylase) ligase→
env-11	soil_XZS41526_y1	+	280	342	hypo→ **RNA**→ hypo→
Oan-1-2	NC_009667.1	-	940830	940769	AccA (COG0825)→ **RNA**→ DUF949 (pfam06104)→
Bsp-2-1	NC_009445.1	+	5476794	5476856	LivH (COG0559)→ **RNA**→ LivM (COG4177)→
env-12	soil_Contig33468	+	665	725	HATPase_c (smart00387)→ **RNA**→
env-13	soil_Contig33468	+	614	664	(shown above)
Oan-1-3	NC_009667.1	-	2380203	2380143	←TyrS (COG0162) **RNA**→ ←EmrE (COG2076)
Oan-1-4	NC_009668.1	-	1857592	1857532	LivK (COG0683)→ **RNA**→ LivH (COG0559)→
Oan-1-5	NC_009668.1	-	198512	198452	←Fba (COG0191) **RNA**→ ←Fbp (COG0158)
Bov-1-2	NC_009505.1	+	833874	833933	RpsD (COG0522)→ **RNA**→ MesJ (COG0037)→
Bsu-1-2	NC_004310.3	+	809118	809177	RpsD (COG0522)→ **RNA**→ MesJ (COG0037)→
Bme-1-2	NC_003317.1	-	1176521	1176462	RpsD (COG0522)→ **RNA**→ MesJ (COG0037)→
Bab-1-1	NC_006932.1	+	830819	830878	RpsD (COG0522)→ **RNA**→ MesJ (COG0037)→
Bme-2-1	NC_007618.1	+	827097	827156	RpsD (COG0522)→ **RNA**→ MesJ (COG0037)→
Oan-1-6	NC_009667.1	-	1300888	1300829	←SerA (COG0111) **RNA**→ ←SerC (COG1932)
Bov-1-3	NC_009504.1	+	269534	269593	←TauB (COG1116) **RNA**→ transposase OrfA→
Bsu-1-3	NC_004311.2	+	270563	270622	←hypo **RNA**→ ←Peptidase_U32 (pfam01136)
Bab-1-2	NC_006933.1	-	901055	900996	hypo→ **RNA**→ ←Peptidase_U32 (pfam01136)
Bme-2-2	NC_007624.1	-	895812	895753	←hypo **RNA**→ ←Peptidase_U32 (pfam01136)
Bme-1-3	NC_003318.1	-	999969	999910	←TauB (COG1116) **RNA**→ ←Peptidase_U32 (pfam01136)
Bov-1-4	NC_009505.1	+	385250	385309	XdhC (COG1975)→ **RNA**→ ←LysR (COG0583)
Bsu-1-4	NC_004310.3	+	362861	362920	XdhC (COG1975)→ **RNA**→ ←LysR (COG0583)
Bme-1-4	NC_003317.1	-	1619454	1619395	XdhC (COG1975)→ **RNA**→ ←LysR (COG0583)
Bab-1-3	NC_006932.1	+	384629	384688	XdhC (COG1975)→ **RNA**→ ←LysR (COG0583)
Bme-2-3	NC_007618.1	+	380997	381056	XdhC (COG1975)→ **RNA**→ COG3748 (COG3748)→
Oan-1-7	NC_009667.1	-	1447858	1447799	COG2319 (COG2319)→ **RNA**→ ←hypo
Oan-1-8	NC_009667.1	-	344968	344909	←COG5473 (COG5473) **RNA**→ ←tRNA-Thr
Bab-1-4	NC_006932.1	+	1959139	1959198	hypo→ **RNA**→ ←HTH_CueR (cd01108)
Bme-2-4	NC_007618.1	+	1956258	1956317	hypo→ **RNA**→ ←HTH_CueR (cd01108)
Bsu-1-5	NC_004310.3	+	1942414	1942473	hypo→ **RNA**→ ←HTH_CueR (cd01108)
Bme-1-5	NC_003317.1	-	51419	51360	hypo→ **RNA**→ ←HTH_CueR (cd01108)
Bov-1-5	NC_009505.1	+	1946990	1947049	hypo→ **RNA**→ ←HTH_CueR (cd01108)
Oan-1-9	NC_009667.1	-	1768002	1767943	CarA (COG0505)→ **RNA**→ ←AraJ (COG2814)
Bov-1-6	NC_009504.1	+	813447	813506	hypo→ **RNA**→ ←COG1407 (COG1407)
Oan-1-10	NC_009667.1	-	137650	137591	COG0384 (COG0384)→ **RNA**→ ←PncB_like (cd01401)
Bov-1-7	NC_009504.1	-	167239	167180	ARA1 (COG0656)→ **RNA**→ ←GlcD (COG0277)
Bme-1-6	NC_003318.1	+	1102795	1102854	ARA1 (COG0656)→ **RNA**→ ←GlpC (COG0247)
Bsu-1-6	NC_004311.2	-	167807	167748	ARA1 (COG0656)→ **RNA**→ ←GlpC (COG0247)
Bab-1-5	NC_006933.1	-	170465	170406	ARA1 (COG0656)→ **RNA**→ ←GlpC (COG0247)
Bme-2-5	NC_007624.1	-	170513	170454	ARA1 (COG0656)→ **RNA**→ ←GlpC (COG0247)
Bme-1-7	NC_003318.1	+	498226	498285	ACR_tran (pfam00873)→ **RNA**→ ←UxuA (COG1312)
Bab-1-6	NC_006933.1	+	416930	416989	←CrcB (COG0239) **RNA**→ ←UxuA (COG1312)
Bme-2-6	NC_007624.1	+	416943	417002	←CrcB (COG0239) **RNA**→ ←UxuA (COG1312)
Bov-1-8	NC_009504.1	-	801891	801832	ACR_tran (pfam00873)→ **RNA**→ ←UxuA (COG1312)
Bsu-1-7	NC_004311.2	-	800246	800187	AcrA (COG0845)→ **RNA**→ ←UxuA (COG1312)
Oan-1-11	NC_009668.1	-	1671300	1671241	COG3153 (COG3153)→ **RNA**→ ←GumC (COG3206)
Bab-1-7	NC_006932.1	+	1675686	1675745	RpmE (COG0254)→ **RNA**→ ←DUF28 (pfam01709)
Bme-2-7	NC_007618.1	+	1672843	1672902	RpmE (COG0254)→ **RNA**→ ←DUF28 (pfam01709)
Bov-1-9	NC_009505.1	+	1665538	1665597	RpmE (COG0254)→ **RNA**→ ←DUF28 (pfam01709)

Bsu-1-8	NC_004310.3	+	1658762	1658821	RpmE (COG0254)→ **RNA**→ ←DUF28 (pfam01709)
Bme-1-8	NC_003317.1	-	330490	330431	RpmE (COG0254)→ **RNA**→ ←DUF28 (pfam01709)
Bab-1-8	NC_006933.1	-	750327	750268	←HTH_MARR (smart00347) **RNA**→ COG3324 (COG3324)→
Bme-2-8	NC_007624.1	-	750315	750256	←HTH_MARR (smart00347) **RNA**→ COG3324 (COG3324)→
Bov-1-10	NC_009504.1	+	437375	437434	←HTH_MARR (smart00347) **RNA**→ ←COG3293 (COG3293)
Bsu-1-9	NC_004311.2	+	464707	464766	←HTH_MARR (smart00347) **RNA**→ COG3324 (COG3324)→
Bme-1-9	NC_003318.1	-	824324	824265	←HTH_MARR (smart00347) **RNA**→ COG3324 (COG3324)→
Bme-1-10	NC_003318.1	-	182689	182630	Fur (COG0735)→ **RNA**→ ←hypo
Bab-1-9	NC_006932.1	+	1487365	1487424	CSD (pfam00313)→ **RNA**→ hypo→
Bme-2-9	NC_007618.1	+	1484519	1484578	CSD (pfam00313)→ **RNA**→ Glutelin:Proline-rich region→
Bsu-1-10	NC_004310.3	+	1469304	1469363	CSD (pfam00313)→ **RNA**→ ←hypo
Bme-1-11	NC_003317.1	-	518833	518774	CSD (pfam00313)→ **RNA**→ Hypothetical Membrane Spanning Protein→
Bab-1-10	NC_006933.1	+	1071810	1071869	Fur (COG0735)→ **RNA**→ hypo→
Bme-2-10	NC_007624.1	+	1066566	1066625	Fur (COG0735)→ **RNA**→ hypo→
Bov-1-11	NC_009504.1	+	1074741	1074800	Fur (COG0735)→ **RNA**→ hypo→
Bsu-1-11	NC_004311.2	+	1116698	1116757	Fur (COG0735)→ **RNA**→ hypo→
Bov-1-12	NC_009505.1	+	1476797	1476856	CSD (pfam00313)→ **RNA**→ ←hypo
Bov-1-13	NC_009505.1	+	72769	72828	MltE (COG0741)→ **RNA**→ COG1272 (COG1272)→
Bsu-1-12	NC_004310.3	+	71949	72008	MltE (COG0741)→ **RNA**→ COG1272 (COG1272)→
Bab-1-11	NC_006932.1	+	71617	71676	MltE (COG0741)→ **RNA**→ COG1272 (COG1272)→
Bme-2-11	NC_007618.1	+	67978	68037	MltE (COG0741)→ **RNA**→ COG1272 (COG1272)→
Bme-1-12	NC_003317.1	-	1931433	1931374	SLT (pfam01464)→ **RNA**→ COG1272 (COG1272)→
Oan-1-12	NC_009667.1	+	84451	84510	SLT (pfam01464)→ **RNA**→ COG1272 (COG1272)→
Bme-1-13	NC_003317.1	-	1424118	1424059	DUF853 (pfam05872)→ **RNA**→ ←ErfK (COG1376)
Bab-1-12	NC_006932.1	+	580443	580502	DUF853 (pfam05872)→ **RNA**→ ←ErfK (COG1376)
Bme-2-12	NC_007618.1	+	576740	576799	DUF853 (pfam05872)→ **RNA**→ ←ErfK (COG1376)
Bsu-1-13	NC_004310.3	+	559019	559078	DUF853 (pfam05872)→ **RNA**→ ←ErfK (COG1376)
Bov-1-14	NC_009505.1	+	580452	580511	DUF853 (pfam05872)→ **RNA**→ ←COG1011 (COG1011)
Bab-1-13	NC_006933.1	+	876598	876657	←hypo **RNA**→ ←MviM (COG0673)
Bme-2-13	NC_007624.1	+	876578	876637	←hypo **RNA**→ ←MviM (COG0673)
Bov-1-15	NC_009504.1	-	294386	294327	←hypo **RNA**→ ←MviM (COG0673)
Bsu-1-14	NC_004311.2	-	295012	294953	←hypo **RNA**→ ←MviM (COG0673)
Bme-1-14	NC_003318.1	+	975526	975585	←hypo **RNA**→ ←MviM (COG0673)
Bab-1-14	NC_006933.1	-	405424	405365	hypo→ **RNA**→ ←COG1407 (COG1407)
Bme-2-14	NC_007624.1	-	405494	405435	hypo→ **RNA**→ ←COG1407 (COG1407)
Bsu-1-15	NC_004311.2	+	811693	811752	hypo→ **RNA**→ ←COG1407 (COG1407)
Bme-1-15	NC_003318.1	-	486674	486615	Hypothetical Membrane Associated Protein→ **RNA**→ ←COG1407 (COG1407)
Bab-1-15	NC_006933.1	+	974901	974960	hypothetical membrane protein→ **RNA**→ ←TruA (COG0101)
Bme-2-15	NC_007624.1	+	969658	969717	hypo→ **RNA**→ ←TruA (COG0101)
Bov-1-16	NC_009504.1	+	1022147	1022206	hypo→ **RNA**→ ←TruA (COG0101)
Bsu-1-16	NC_004311.2	+	1019508	1019567	hypo→ **RNA**→ ←Fmt (COG0223)
Bme-1-16	NC_003318.1	-	279495	279436	Hypothetical Membrane Spanning Protein→ **RNA**→ ←TruA (COG0101)
Bov-1-17	NC_009504.1	-	495556	495497	COG5455 (COG5455)→ **RNA**→ ←COG3293 (COG3293)
Bsu-1-17	NC_004311.2	-	521206	521147	COG5455 (COG5455)→ **RNA**→ hypo→
Bme-1-17	NC_003318.1	+	767817	767876	COG5455 (COG5455)→ **RNA**→ OppA (COG4166)→
Bab-1-16	NC_006933.1	+	693038	693097	COG5455 (COG5455)→ **RNA**→ hypo→
Bme-2-16	NC_007624.1	+	693026	693085	COG5455 (COG5455)→ **RNA**→ hypo→
Bab-1-17	NC_006932.1	+	724765	724824	hypo→ **RNA**→ ←LldP (COG1620)
Bme-2-17	NC_007618.1	+	721060	721119	←hypo **RNA**→ ←LldP (COG1620)
Bsu-1-18	NC_004310.3	+	703060	703119	hypo→ **RNA**→ ←LldP (COG1620)
Bov-1-18	NC_009505.1	+	724516	724575	hypo→ **RNA**→ ←LldP (COG1620)
Bab-1-18	NC_006932.1	+	625260	625319	BaeS (COG0642)→ **RNA**→ GlnE (COG1391)→

Bme-2-18	NC_007618.1	+	621557	621616	BaeS (COG0642)→ **RNA**→ GlnE (COG1391)→
Bov-1-19	NC_009505.1	+	625222	625281	BaeS (COG0642)→ **RNA**→ GlnE (COG1391)→
Bsu-1-19	NC_004310.3	+	603763	603822	BaeS (COG0642)→ **RNA**→ GlnE (COG1391)→
Oan-1-13	NC_009667.1	-	2356155	2356096	←GST_C_2 (cd03180)GST_N_2 (cd03047) **RNA**→ ←ArtM (COG4160)
Bov-1-20	NC_009505.1	-	151380	151321	←COG4567 (COG4567) **RNA**→ ←COG3293 (COG3293)
Bsu-1-20	NC_004310.3	-	151086	151027	←hypo **RNA**→ ←COG3293 (COG3293)
Bme-1-18	NC_003317.1	+	1851995	1852054	←hypo **RNA**→ ←COG3293 (COG3293)
Bab-1-19	NC_006932.1	-	152439	152380	←hypo **RNA**→ ←COG3293 (COG3293)
Bme-2-19	NC_007618.1	-	148806	148747	←hypo **RNA**→ ←COG3293 (COG3293)
Bab-1-20	NC_006932.1	+	1867220	1867279	COG3735 (COG3735)→ **RNA**→ ←Lpd (COG1249)
Bme-2-20	NC_007618.1	+	1864348	1864407	COG3735 (COG3735)→ **RNA**→ ←Lpd (COG1249)
Bov-1-21	NC_009505.1	+	1853992	1854051	←hypo **RNA**→ ←Lpd (COG1249)
Bsu-1-21	NC_004310.3	+	1850242	1850301	COG3735 (COG3735)→ **RNA**→ ←Lpd (COG1249)
Bme-1-19	NC_003317.1	-	143557	143498	COG3735 (COG3735)→ **RNA**→ ←Lpd (COG1249)
Oan-1-14	NC_009667.1	+	2486548	2486607	COG3577 (COG3577)→ **RNA**→ ←COG0042 (COG0042)
Bov-1-22	NC_009504.1	-	134811	134752	UhpC (COG2271)→ **RNA**→ ←HTH_ARAC (smart00342)
Bab-1-21	NC_006932.1	+	1942624	1942683	←hypo **RNA**→ LolA (COG2834)→
Bme-2-21	NC_007618.1	+	1939746	1939805	←hypo **RNA**→ LolA (COG2834)→
Bsu-1-22	NC_004310.3	+	1925889	1925948	←hypo **RNA**→ LolA (COG2834)→
Bme-1-20	NC_003317.1	-	67933	67874	MIP (cd00333)→ **RNA**→ LolA (COG2834)→
Bab-1-22	NC_006933.1	-	134661	134602	UhpC (COG2271)→ **RNA**→ ←Transposase_8 (pfam01527)
Bme-2-22	NC_007624.1	-	134724	134665	←hypo **RNA**→ ←hypo
Bsu-1-23	NC_004311.2	-	135238	135179	←hypo **RNA**→ ←hypo
Bov-1-23	NC_009505.1	+	1929637	1929696	MIP (cd00333)→ **RNA**→ LolA (COG2834)→
Oan-1-15	NC_009668.1	+	1371195	1371254	G6PD_N (pfam00479)G6PD_C (pfam02781)→ **RNA**→ NagB (COG0363)→
Bab-1-23	NC_006932.1	-	1699282	1699223	Pyruvate_Kinase (cd00288)→ **RNA**→ ←TPR (cd00189)
Bme-2-23	NC_007618.1	-	1696439	1696380	Pyruvate_Kinase (cd00288)→ **RNA**→ ←TPR (cd00189)
Bov-1-24	NC_009505.1	-	1688963	1688904	Pyruvate_Kinase (cd00288)→ **RNA**→ ←TPR (cd00189)
Bme-1-21	NC_003317.1	+	306892	306951	Pyruvate_Kinase (cd00288)→ **RNA**→ ←TPR (cd00189)
Bab-1-24	NC_006932.1	-	33415	33356	Cons_hypoth698 (pfam03601)→ **RNA**→ ←RpsA (COG0539)
Bme-2-24	NC_007618.1	-	29816	29757	Cons_hypoth698 (pfam03601)→ **RNA**→ ←RpsA (COG0539)
Bab-1-25	NC_006932.1	+	1537112	1537171	←MMT1 (COG0053) **RNA**→ ←TrpE (COG0147)PabA (COG0512)
Bme-2-25	NC_007618.1	+	1534266	1534325	hypo→ **RNA**→ hypo→
Bab-1-26	NC_006933.1	+	380984	381043	ArgD (COG4992)→ **RNA**→ ←hypo
Bov-1-25	NC_009504.1	-	837878	837819	ArgD (COG4992)→ **RNA**→ ←hypo
Bsu-1-24	NC_004311.2	-	836135	836076	ArgD (COG4992)→ **RNA**→ ←hypo
Bme-1-22	NC_003318.1	+	462228	462287	ArgD (COG4992)→ **RNA**→ ←Hypothetical Membrane Spanning Protein
Bsu-1-25	NC_004311.2	+	214919	214978	HTH_CueR (cd01108)→ **RNA**→ ←GATase1_Hsp31_like (cd03141)
Bme-1-23	NC_003318.1	-	1055641	1055582	HTH_CueR (cd01108)→ **RNA**→ ←GATase1_Hsp31_like (cd03141)
Bme-1-24	NC_003317.1	-	469081	469022	←MMT1 (COG0053) **RNA**→ ←TrpE (COG0147)PabA (COG0512)
Bme-1-25	NC_003317.1	+	1969807	1969866	Cons_hypoth698 (pfam03601)→ **RNA**→ ←RpsA (COG0539)
Bsu-1-26	NC_004310.3	-	1682361	1682302	Pyruvate_Kinase (cd00288)→ **RNA**→ ←TPR (cd00189)
Bov-1-26	NC_009505.1	+	1526484	1526543	hypo→ **RNA**→ ←TrpE (COG0147)PabA (COG0512)
Bsu-1-27	NC_004310.3	+	860545	860604	Pcm (COG2518)→ **RNA**→ NlpD (COG0739)→
Bme-1-26	NC_003317.1	-	1125082	1125023	Pcm (COG2518)→ **RNA**→ NlpD (COG0739)→
Bsu-1-28	NC_004310.3	-	1109588	1109529	Eno (COG0148)→ **RNA**→ TehB (pfam03848)→
Bsu-1-29	NC_004310.3	-	33554	33495	Cons_hypoth698 (pfam03601)→ **RNA**→ ←RpsA (COG0539)
Bsu-1-30	NC_004310.3	+	1945471	1945530	HMA (cd00371)ZntA (COG2217)→ **RNA**→ hypo→
Bov-1-27	NC_009505.1	+	885762	885821	Pcm (COG2518)→ **RNA**→ ←COG3293 (COG3293)
Bov-1-28	NC_009505.1	-	33552	33493	COG3293 (COG3293)→ **RNA**→ ←RpsA (COG0539)
Bme-2-26	NC_007624.1	+	381053	381112	ArgD (COG4992)→ **RNA**→ ←hypo
Bab-1-27	NC_006932.1	+	1962166	1962225	HMA (cd00371)ZntA (COG2217)→ **RNA**→ hypo→

Bme-2-27	NC_007618.1	+	1959285	1959344	HMA (cd00371)ZntA (COG2217)→ **RNA**→ hypo→
Bov-1-29	NC_009505.1	-	1112331	1112272	Eno (COG0148)→ **RNA**→ ←transposase OrfA
Bsu-1-31	NC_004310.3	+	1519072	1519131	hypo→ **RNA**→ hypo→
Oan-1-16	NC_009667.1	-	1310021	1309961	RluA (COG0564)→ **RNA**→ RpoD (COG0568)→
env-14	soil_Contig10577	+	108	169	**RNA**→ COG2740 (COG2740)→
env-15	soil_Contig26760	+	337	398	hypo→ **RNA**→ **RNA**→ ←DadA (COG0665)
env-36	soil_Contig26760	+	219	280	(shown above)
Nha-1-2	NC_007964.1	+	3594976	3595037	Ferritin_like_AB (cd01045)CCC1_like_1 (cd02437)→ **RNA**→ Icc (COG1409)→
Mlo-1-2	NC_002678.2	+	4014611	4014672	TrpF (COG0135)→ **RNA**→ TrpB (COG0133)→
env-16	soil_XZS38199_g1	-	247	186	←hypo **RNA**→ ←hypo
env-17	soil_Contig10442	+	881	942	COG2840 (COG2840)→ **RNA**→
Bja-1-1	NC_004463.1	-	5927841	5927780	Ribosomal_L30 (cd01658)→ **RNA**→ Ribosomal_L15 (pfam01305)→
Oan-1-17	NC_009667.1	-	1154710	1154649	←hypo **RNA**→ ←PriA (COG1198)
Rle-1-2	NC_008380.1	+	1355847	1355908	AtoS (COG2202)COG5001 (COG5001)→ **RNA**→ ←UgpQ (COG0584)
Mlo-1-3	NC_002678.2	+	4360359	4360419	HTH_MARR (smart00347)→ **RNA**→ COG4231 (COG4231)→
Bab-1-28	NC_006932.1	-	1427839	1427778	UbiE (COG2226)HTH_ARSR (smart00418)→ **RNA**→ MTHFR (cd00537)→
Bme-2-28	NC_007618.1	-	1424989	1424928	UbiE (COG2226)HTH_ARSR (smart00418)→ **RNA**→ MTHFR (cd00537)→
Bov-1-30	NC_009505.1	-	1417320	1417259	UbiE (COG2226)HTH_ARSR (smart00418)→ **RNA**→ MTHFR (cd00537)→
Bsu-1-32	NC_004310.3	-	1409785	1409724	UbiE (COG2226)HTH_ARSR (smart00418)→ **RNA**→ MTHFR (cd00537)→
Bme-1-27	NC_003317.1	+	578361	578422	UbiE (COG2226)HTH_ARSR (smart00418)→ **RNA**→ MTHFR (cd00537)→
Oan-1-18	NC_009667.1	+	1437802	1437862	hypo→ **RNA**→ ←COG0218 (COG0218)
Bov-1-31	NC_009505.1	+	215180	215241	Aldedh (pfam00171)→ **RNA**→ AdhP (COG1064)→
Bsu-1-33	NC_004310.3	+	215127	215188	Aldedh (pfam00171)→ **RNA**→ AdhP (COG1064)→
Bab-1-29	NC_006932.1	+	216489	216550	AcoR (COG3284)→ **RNA**→ AdhP (COG1064)→
Bme-2-29	NC_007618.1	+	212856	212917	AcoR (COG3284)→ **RNA**→ AdhP (COG1064)→
Bme-1-28	NC_003317.1	-	1787951	1787890	Aldedh (pfam00171)→ **RNA**→ AdhP (COG1064)→
Rle-1-3	NC_008380.1	+	620156	620217	←putative transmembrane protein **RNA**→ COG4976 (COG4976)→
Rle-1-4	NC_008380.1	+	1811296	1811357	DnaE (COG0587)→ **RNA**→ ←COG3409 (COG3409)ErfK_YbiS_YhnG (pfam03734)
Rle-1-5	NC_008380.1	+	2167323	2167384	←COG2607 (COG2607) **RNA**→ YajC (COG1862)→
Ret-1-2	NC_007766.1	-	212815	212754	←BioA (COG0161) **RNA**→ ←Aldedh (pfam00171)
Rle-1-6	NC_008380.1	+	1408637	1408698	RhtB (COG1280)→ **RNA**→ Cytochrom_C_2 (pfam01322)→
Ret-1-3	NC_007761.1	+	1145025	1145086	hypo→ **RNA**→ hypo→
Ret-1-4	NC_007761.1	-	597435	597374	UvrD (COG0210)→ **RNA**→ ←Lactamase_B (pfam00753)
Rle-1-7	NC_008380.1	+	1767935	1767996	←GGDEF (cd01949) **RNA**→ ClpP (COG0740)→
Sme-1-1	NC_009636.1	+	3540819	3540880	COG1079 (COG1079)→ **RNA**→ Cdd (COG0295)→
Rle-1-8	NC_008380.1	+	660405	660466	BON (pfam04972)→ **RNA**→ hypo→
Oan-1-19	NC_009667.1	-	865615	865554	←PrfB (COG1186) **RNA**→ ←PEPCK_ATP (cd00484)
Mlo-1-4	NC_002678.2	+	4009251	4009311	←STE14 (COG2020) **RNA**→ ←InfC (COG0290)
Oan-1-20	NC_009667.1	-	743980	743919	CaiD (COG1024)→ **RNA**→ hypo→
Oan-1-21	NC_009667.1	+	1417200	1417261	←Gst (COG0625) **RNA**→ BacA (pfam02673)→
Oan-1-22	NC_009667.1	-	1432252	1432191	DapD (COG2171)→ **RNA**→ ←hypo
Oan-1-23	NC_009668.1	-	1892375	1892314	Spo0J (COG1475)→ **RNA**→ replication protein C→
Bab-1-30	NC_006932.1	-	448334	448273	←hypo **RNA**→ hypo→
Bme-2-30	NC_007618.1	-	444638	444577	←hypo **RNA**→ hypo→
Bov-1-32	NC_009505.1	-	448468	448407	←hypo **RNA**→ ←hypo
Bsu-1-34	NC_004310.3	-	426468	426407	←hypo **RNA**→ hypo→
Bme-1-29	NC_003317.1	+	1556419	1556480	←hypo **RNA**→ ←CSD (pfam00313)
Oan-1-24	NC_009667.1	-	1035419	1035362	ArgH (COG0165)→ **RNA**→ COG5567 (COG5567)→
Bab-1-31	NC_006932.1	-	570315	570254	COG3226 (COG3226)→ **RNA**→ Aldedh (pfam00171)→
Bme-2-31	NC_007618.1	-	566610	566549	BetA (COG2303)→ **RNA**→ Aldedh (pfam00171)→
Bov-1-33	NC_009505.1	-	570285	570224	BetA (COG2303)→ **RNA**→ Aldedh (pfam00171)→
Bsu-1-35	NC_004310.3	-	548851	548790	BetA (COG2303)→ **RNA**→ Aldedh (pfam00171)→

Bme-1-30	NC_003317.1	+	1434274	1434335	GMC_oxred_C (pfam05199)→ **RNA**→ Aldedh (pfam00171)→
Oan-1-25	NC_009667.1	-	2828307	2828246	←conserved hypothetical membrane spanning protein **RNA**→ ←PtrB (COG1770)
Oan-1-26	NC_009667.1	+	1307075	1307135	←pyrophosphatase (cd00412) **RNA**→ ←COG4977 (COG4977)
Oan-1-27	NC_009668.1	+	1842520	1842581	COG3791 (COG3791)→ **RNA**→ ←HTH_ARAC (smart00342)
Bab-1-32	NC_006933.1	+	1107448	1107505	MglA (COG1129)→ **RNA**→ XylH (COG4214)→
Bme-2-32	NC_007624.1	+	1102192	1102249	MglA (COG1129)→ **RNA**→ XylH (COG4214)→
Bov-1-34	NC_009504.1	+	1109758	1109815	←COG3268 (COG3268) **RNA**→ XylH (COG4214)→
Bsu-1-36	NC_004311.2	+	1152340	1152397	MglA (COG1129)→ **RNA**→ XylH (COG4214)→
Bme-1-31	NC_003318.1	-	147066	147009	MglA (COG1129)→ **RNA**→ XylH (COG4214)→
Oan-1-28	NC_009667.1	+	464612	464669	←AraJ (COG2814) **RNA**→ LysA (COG0019)→
Oan-1-29	NC_009668.1	-	1840111	1840054	G_glu_transpept (pfam01019)→ **RNA**→ ←PepD (COG2195)
Oan-1-30	NC_009667.1	+	2825932	2825993	←transcriptional regulator, MucR family **RNA**→ PtrB (COG1770)→
Oan-1-31	NC_009667.1	+	1042491	1042552	COG2096 (COG2096)→ **RNA**→ ←NrdG (COG0602)
Oan-1-32	NC_009668.1	+	1700107	1700168	COG5338 (COG5338)→ **RNA**→ ←GntP_permease (pfam02447)
Oan-1-33	NC_009668.1	+	1608002	1608063	←CbiK (COG5266) **RNA**→ ←ZnuA (COG4531)
Oan-1-34	NC_009667.1	+	286391	286452	←MdoB (COG1368) **RNA**→ ←MdlB (COG1132)
Oan-1-35	NC_009668.1	-	151737	151676	←tRNA-Sec **RNA**→ ←SelB (COG3276)
Oan-1-36	NC_009667.1	-	2011276	2011215	COG3637 (COG3637)→ **RNA**→ COG2355 (COG2355)→
Oan-1-37	NC_009667.1	+	425646	425707	←MetC (COG0626) **RNA**→ DCD (pfam06559)→
Oan-1-38	NC_009668.1	+	134797	134858	DUF849 (pfam05853)→ **RNA**→ Acetyltransf_1 (pfam00583)→
Oan-1-39	NC_009667.1	+	2107005	2107066	hypo→ **RNA**→ SDH_alpha (pfam03313)SDH_beta (pfam03315)→
Oan-1-40	NC_009667.1	+	582327	582388	HTH_ARSR (smart00418)→ **RNA**→ ATM1 (COG5265)→
Oan-1-41	NC_009668.1	+	662096	662157	COG0412 (COG0412)→ **RNA**→ ←DUF808 (pfam05661)
Oan-1-42	NC_009668.1	-	251428	251367	←NrdH (cd02976) **RNA**→ ←DltE (COG0300)
Bab-1-33	NC_006932.1	+	174035	174096	hypo→ **RNA**→ COG1934 (COG1934)→
Bme-2-33	NC_007618.1	+	170402	170463	hypo→ **RNA**→ COG1934 (COG1934)→
Bsu-1-37	NC_004310.3	+	172660	172721	hypo→ **RNA**→ COG1934 (COG1934)→
Bme-1-32	NC_003317.1	-	1830401	1830340	hypo→ **RNA**→ COG1934 (COG1934)→
Bov-1-35	NC_009505.1	+	172976	173037	hypo→ **RNA**→ COG1934 (COG1934)→
Oan-1-43	NC_009667.1	-	1410984	1410923	MGS (cd01422)→ **RNA**→ Glk (COG0837)→
Oan-1-44	NC_009667.1	-	2525127	2525066	ABC_MJ0796_Lo1CDE_FtsE (cd03255)→ **RNA**→ DnaE (COG0587)→
Bab-1-34	NC_006932.1	-	987679	987618	GlnK (COG0347)→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)→
Bme-2-34	NC_007618.1	-	984799	984738	GlnK (COG0347)→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)→
Bsu-1-38	NC_004310.3	-	968679	968618	GlnK (COG0347)→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)→
Bme-1-33	NC_003317.1	+	1019673	1019734	GlnK (COG0347)→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)→
Bov-1-36	NC_009505.1	-	977764	977703	GlnK (COG0347)→ **RNA**→ Gln-synt_C (pfam00120)Gln-synt_N (pfam03951)→
Oan-1-45	NC_009668.1	+	644942	645003	COG1739 (COG1739)→ **RNA**→ ←GGDEF (cd01949)
Oan-1-46	NC_009667.1	-	2699672	2699611	WcaG (COG0451)→ **RNA**→ ←hypo
Oan-1-47	NC_009667.1	-	2769728	2769667	COG3743 (COG3743)→ **RNA**→ ←MltE (COG0741)
Bab-1-35	NC_006932.1	+	452330	452391	Chorismate_synt (pfam01264)→ **RNA**→ Hist_deacetyl (pfam00850)→
Bme-2-35	NC_007618.1	+	448634	448695	RibA (COG0807)DHBP_synthase (pfam00926)→ **RNA**→ Hist_deacetyl (pfam00850)→
Bov-1-37	NC_009505.1	+	452464	452525	RibA (COG0807)DHBP_synthase (pfam00926)→ **RNA**→ Hist_deacetyl (pfam00850)→
Bme-1-34	NC_003317.1	-	1552427	1552366	RibA (COG0807)DHBP_synthase (pfam00926)→ **RNA**→ Hist_deacetyl (pfam00850)→
Bsu-1-39	NC_004310.3	+	430465	430526	RibA (COG0807)DHBP_synthase (pfam00926)→ **RNA**→ Hist_deacetyl (pfam00850)→
Oan-1-48	NC_009667.1	+	85301	85362	COG1272 (COG1272)→ **RNA**→ ←DnaE (COG0587)
Oan-1-49	NC_009667.1	+	513332	513393	COG3826 (COG3826)→ **RNA**→ **RNA**→ ArsC_Yffb (cd03035)→
Oan-1-68	NC_009667.1	+	513477	513538	(shown above)
Oan-1-50	NC_009668.1	+	614057	614118	COG3118 (COG3118)→ **RNA**→ COG2802 (COG2802)→
Oan-1-51	NC_009668.1	+	747730	747791	AraH (COG1172)→ **RNA**→ RbsB (COG1879)→
Oan-1-52	NC_009668.1	-	1864804	1864747	DctQ (COG1593)→ **RNA**→ DctP (COG1638)→
Oan-1-53	NC_009667.1	-	772096	772035	Ribosomal_S15p_S13e (cd00353)→ **RNA**→ Pnp (COG1185)→
Oan-1-54	NC_009668.1	-	168530	168469	Aes (COG0657)→ **RNA**→ AcrA (COG0845)→

Oan-1-55	NC_009668.1	-	1379266	1379206	AppF (COG4608)→ **RNA**→ ←ProC (COG0345)
Oan-1-56	NC_009668.1	+	1233552	1233613	←COG2128 (COG2128) **RNA**→ ←PRX_Typ2cys (cd03015)
env-18	JCVI_SCAF_1096627928814	+	1235	1296	**RNA**→
env-19	JCVI_SCAF_1101668623320	-	927	866	EF_TS (pfam00889)? (GOS1222)→ **RNA**→
Oan-1-57	NC_009667.1	+	2323512	2323573	←HflX (COG2262) **RNA**→ ←Hfq (COG1923)
Oan-1-58	NC_009668.1	-	792371	792310	←PS_pyruv_trans (pfam04230) **RNA**→ ←PS_pyruv_trans (pfam04230)
Mlo-1-5	NC_002678.2	+	6362866	6362927	LipA (COG4520)→ **RNA**→ COG4235 (COG4235)→
Rle-1-9	NC_008380.1	-	622066	622005	COG2852 (COG2852)→ **RNA**→ ←COG1284 (COG1284)
Rle-1-10	NC_008380.1	-	3548225	3548164	←LasI (COG3916) **RNA**→ ←CsgD (COG2771)Autoind_bind (pfam03472)
Rle-1-11	NC_008380.1	-	2977707	2977646	AcpP (COG0236)→ **RNA**→ FabA (COG0764)→
Rle-1-12	NC_008384.1	-	419133	419072	GCD (cd01151)→ **RNA**→ DUF1209 (pfam06731)→
Rle-1-13	NC_008380.1	+	1778883	1778944	COG1917 (COG1917)→ **RNA**→ ←hypo
Rle-1-14	NC_008378.1	+	258216	258277	←COG4119 (COG4119) **RNA**→ PRX_like2 (cd02970)→
Rle-1-15	NC_008380.1	+	19721	19782	TrpA (COG0159)→ **RNA**→ AccD (COG0777)→
Rle-1-16	NC_008380.1	+	920367	920428	hypo→ **RNA**→ ←COG5403 (COG5403)
Rle-1-17	NC_008378.1	-	49072	49011	←HisM (COG0765) **RNA**→ ←SBP_bac_3 (pfam00497)
Rle-1-18	NC_008380.1	-	4591966	4591905	TlpA_like_family (cd02966)→ **RNA**→ ←FadB (COG1250)
Rle-1-19	NC_008384.1	-	35665	35604	DHDPS-like (cd00408)→ **RNA**→ ←ATZ_TRZ_like (cd01298)
Rle-1-20	NC_008380.1	+	1715294	1715355	←REC (cd00156) **RNA**→ REC (cd00156)BaeS (COG0642)→
Rle-1-21	NC_008380.1	-	624939	624878	←Lhr (COG1201) **RNA**→ ←COG1284 (COG1284)
Rle-1-22	NC_008378.1	+	361671	361732	G_glu_transpept (pfam01019)→ **RNA**→ DppB (COG0601)→
Rle-1-23	NC_008380.1	+	4728649	4728710	hypo→ **RNA**→ ←TdcF (COG0251)
Rle-1-24	NC_008380.1	+	2189256	2189317	←YqgFc (smart00732) **RNA**→ COG2220 (COG2220)→
Rle-1-25	NC_008380.1	-	745577	745516	←FliF (COG1766) **RNA**→ ←hypo
Ret-1-5	NC_007762.1	+	44547	44608	virulence VirF1 protein→ **RNA**→ ←hypo
Rle-1-26	NC_008380.1	+	1597637	1597698	DUF159 (pfam02586)→ **RNA**→ ←PssA (COG1183)
Rle-1-27	NC_008380.1	-	499722	499661	←TPR (cd00189)COG5616 (COG5616) **RNA**→ ←TPR (cd00189)CyaA (COG2114)COG5616 (COG5616)
Rle-1-28	NC_008380.1	-	1472476	1472415	MviM (COG0673)→ **RNA**→ ←hypo
Rle-1-29	NC_008380.1	+	4913151	4913212	putative transmembrane protein→ **RNA**→ ←adh_short (pfam00106)
Rle-1-30	NC_008380.1	+	1744940	1745001	←COG5653 (COG5653) **RNA**→ CtaA (COG1612)→
Rle-1-31	NC_008380.1	+	2601664	2601725	PAS (cd00130)→ **RNA**→ putative transmembrane protein→
Rle-1-32	NC_008380.1	+	481729	481790	PrpB (COG2513)→ **RNA**→ ←GOX_like_FMN_family (cd02809)
Rle-1-33	NC_008380.1	+	1463391	1463452	←putative N-terminal transmembrane anchored protein **RNA**→ ←hypo
Rle-1-34	NC_008380.1	+	2322487	2322548	hypo→ **RNA**→ ←GlnQ (COG1126)
Rle-1-35	NC_008380.1	-	111712	111651	IolE (COG1082)→ **RNA**→ MviM (COG0673)→
Rle-1-36	NC_008380.1	+	3798542	3798603	COG3437 (COG3437)→ **RNA**→ ←ILVD_EDD (pfam00920)
Ret-1-6	NC_007761.1	+	923346	923407	LemA (pfam04011)→ **RNA**→ EPSP_synthase (cd01556)→
Rle-1-37	NC_008380.1	+	1410371	1410432	CccA (COG2010)→ **RNA**→ ←GST_C_1 (cd03179)GST_N_4 (cd03056)
Oan-1-59	NC_009667.1	-	1769382	1769321	←COG1610 (COG1610) **RNA**→ CarA (COG0505)→
Rle-1-38	NC_008380.1	-	948943	948882	←GST_C_Ure2p_like (cd03178)GST_N_Ure2p_like (cd03048) **RNA**→ COG2259 (COG2259)→
Ret-1-7	NC_007761.1	+	4023864	4023925	LysA (COG0019)→ **RNA**→ hypo→
Oan-1-60	NC_009667.1	+	589136	589197	←PurF (COG0034) **RNA**→ ←Colicin_V (pfam02674)
Mlo-1-6	NC_002678.2	+	3838344	3838405	Amidase (pfam01425)→ **RNA**→ ArgB (COG0548)→
Mlo-1-7	NC_002678.2	+	6416832	6416893	←MltE (COG0741) **RNA**→ DHDPS (cd00950)→
env-20	JCVI_SCAF_1096627928814	+	94	155	**RNA**→
env-21	soil_XZS61489_y1	-	560	499	hypo→ **RNA**→ ←COG3293 (COG3293)
env-23	soil_XZS33276_y1	-	647	586	←Uncharacterized protein conserved in bacteria **RNA**→ Methyltransf_6 (pfam03737)→
Bja-1-2	NC_004463.1	-	4516243	4516182	←adh_short (pfam00106) **RNA**→ ←FabD (COG0331)
Rpa-4-1	NC_007778.1	-	3106482	3106421	hypo→ **RNA**→ ←Ion_trans (pfam00520)
Mch-1-1	NZ_ABEX01000006.1	-	302172	302111	COG0388 (COG0388)→ **RNA**→ DUF1178 (pfam06676)→
Mex-1-1	NZ_ABEC01000009.1	-	270048	269987	COG0388 (COG0388)→ **RNA**→ DUF1178 (pfam06676)→
Mch-1-2	NZ_ABEX01000005.1	-	39935	39873	←FlhB (COG1377) **RNA**→ ←FliR (COG1684)

Mex-1-2	NZ_ABEC01000005.1	+	403396	403458	hypo→ **RNA**→ hypo→
Mex-1-3	NZ_ABEC01000001.1	-	174903	174841	PhnP (COG1235)→ **RNA**→ ←Trunc_globin (cd00454)
Mch-1-3	NZ_ABEX01000002.1	+	70145	70207	PhnP (COG1235)→ **RNA**→ ←Trunc_globin (cd00454)
Mex-1-4	NZ_ABEC01000001.1	+	159757	159819	BaeS (COG0642)→ **RNA**→ REC (cd00156)→
Mex-1-5	NZ_ABEC01000007.1	+	235592	235653	NlpD (COG0739)→ **RNA**→ ←PrfB (COG1186)
Mch-1-4	NZ_ABEX01000003.1	+	183833	183894	NlpD (COG0739)→ **RNA**→ ←PrfB (COG1186)
Mex-1-6	NZ_ABEC01000004.1	-	180536	180474	COG2363 (COG2363)→ **RNA**→ ←AtoE (COG2031)
Mex-1-7	NZ_ABEC01000004.1	-	188653	188591	←Acd (COG2515) **RNA**→ ←MdlB (COG1132)
Mch-1-5	NZ_ABEX01000013.1	-	195751	195690	←COG4338 (COG4338) **RNA**→ ←Fmt (COG0223)
Mch-1-6	NZ_ABEX01000008.1	-	259866	259805	major facilitator superfamily MFS_1→ **RNA**→ DUF140 (pfam02405)→
Mma-1-1	NZ_AAAP01003055.1	+	1628	1690	←NrdG (COG0602) **RNA**→ COG1686: D-alanyl-D-alanine carboxypeptidase→
Mch-1-7	NZ_ABEX01000009.1	-	184407	184346	PAS (cd00130)COG3920 (COG3920)→ **RNA**→ **RNA**→ ←hypo
Mch-1-11	NZ_ABEX01000009.1	-	184345	184285	(shown above)
Mex-1-8	NZ_ABEC01000005.1	-	378367	378305	←GST_C_GTT1_like (cd03189)GST_N_GTT1_like (cd03046) **RNA**→ **RNA**→ ModC (COG4148)→
Mex-1-11	NZ_ABEC01000005.1	-	378430	378369	(shown above)
Mch-1-8	NZ_ABEX01000025.1	-	26678	26617	←GST_C_GTT1_like (cd03189)GST_N_GTT1_like (cd03046) **RNA**→ ModC (COG4148)→
Mex-1-9	NZ_ABEC01000004.1	-	297350	297288	SufI (COG2132)→ **RNA**→ ←COG1032 (COG1032)
Mex-1-10	NZ_ABEC01000009.1	+	331676	331737	←SCO (cd02968) **RNA**→ ←CysG (COG0007)CysG (COG1648)
Mch-1-9	NZ_ABEX01000014.1	+	25079	25140	←SCO (cd02968) **RNA**→ ←CysG (COG0007)CysG (COG1648)
env-24	soil_XZS65113_y1	-	549	488	←hypo **RNA**→ ←AlkB (COG3145)
Rpa-3-1	NC_007958.1	-	3601658	3601597	←hypo **RNA**→ SLT (pfam01464)NOG07949 (NOG07949)→
env-25	soil_XZS14719_g1	+	571	633	hypo→ **RNA**→
Nsp-1-1	NZ_AAMY01000003.1	-	136543	136482	←PRX5_like (cd03013) **RNA**→ ←CaiC (COG0318)
Rpa-1-1	NC_008435.1	-	758058	757997	gltB_C (cd00982)GltB (COG0069)GltS (cd00713)→ **RNA**→ GltD (COG0493)→
Oan-1-61	NC_009667.1	-	1763705	1763644	COG1835 (COG1835)→ **RNA**→ MGS_CPS_II (cd01424)CarB (COG0458)CPSase_L_D3 (pfam02787)→
Nwi-1-11	NC_007406.1	+	1504032	1504093	RpsC (COG0092)→ **RNA**→ Ribosomal_L16 (pfam00252)→
Nsp-1-2	NZ_AAMY01000007.1	-	60580	60519	RpsC (COG0092)→ **RNA**→ Ribosomal_L16 (pfam00252)→
Bja-1-3	NC_004463.1	-	5441459	5441398	NuoK (COG0713)→ **RNA**→ NuoL (COG1009)→
Bja-1-4	NC_004463.1	-	753743	753682	PheT (COG0072)tRNA_bind_bactPheRS (cd02796)→ **RNA**→ hypo→
Mlo-1-8	NC_002678.2	+	3420275	3420336	←Lpd (COG1249) **RNA**→ ←adh_short (pfam00106)
Rle-1-39	NC_008380.1	-	3174159	3174099	←hypo **RNA**→ AraJ (COG2814)→
Rle-1-40	NC_008380.1	-	121608	121548	←CotS (COG0510) **RNA**→ ←MalK (COG3839)
Rle-1-41	NC_008380.1	-	3242275	3242215	COG3791 (COG3791)→ **RNA**→ DUF704 (pfam05146)→
Mlo-1-9	NC_002678.2	+	3584606	3584667	←CysK (COG0031)COG0517 (COG0517) **RNA**→ ←YjgH (cd02198)
Mlo-1-10	NC_002678.2	+	1404644	1404705	←GntK (COG3265) **RNA**→ ←AraH (COG1172)
Ret-1-8	NC_007761.1	-	1057560	1057499	←hypo **RNA**→ hypo→
Mlo-1-11	NC_002678.2	-	1422849	1422788	←CoxL (COG1529) **RNA**→ ←CoxS (COG2080)
Mlo-1-12	NC_002678.2	-	6906940	6906879	←RibC (COG0307) **RNA**→ ←RibD (COG0117)RibD (COG1985)
Rle-1-42	NC_008380.1	-	1171415	1171354	←putative transmembrane protein **RNA**→ hypo→
Bja-1-5	NC_004463.1	+	5190730	5190791	COG3212 (COG3212)NOG32137 (NOG32137)→ **RNA**→ Asd (COG0136)→
env-26	soil_Contig24937	+	278	341	ABC-type amino acid transport system, permease component→ **RNA**→ GlnQ (COG1126)→
Bsu-1-40	NC_004310.3	+	1766385	1766446	LysR_substrate (pfam03466)→ **RNA**→ ←DUF1275 (pfam06912)
Bab-1-36	NC_006932.1	+	1783278	1783339	LysR_substrate (pfam03466)→ **RNA**→ ←DUF1275 (pfam06912)
Bme-2-36	NC_007618.1	+	1780425	1780486	LysR_substrate (pfam03466)→ **RNA**→ ←DUF1275 (pfam06912)
Bov-1-38	NC_009505.1	+	1774505	1774566	LysR_substrate (pfam03466)→ **RNA**→ ←DUF1275 (pfam06912)
Bme-1-35	NC_003317.1	-	223280	223219	LysR_substrate (pfam03466)→ **RNA**→ ←DUF1275 (pfam06912)
Mlo-1-13	NC_002678.2	+	2081632	2081693	COG3921 (COG3921)→ **RNA**→ ←hypo
Nha-1-3	NC_007964.1	-	585766	585705	OsmC (COG1764)→ **RNA**→ AtpD (COG0055)→
Rpa-2-1	NC_007925.1	-	839231	839170	gltB_C (cd00982)GltB (COG0069)Glu_synth_NTN (pfam04897)→ **RNA**→ GltD (COG0493)→
env-27	soil_XZS50792_g1	-	809	748	**RNA**→ ←hypo
env-28	soil_XZS57871_g1	+	91	152	**RNA**→ ←PyrF (COG0284)
Bja-1-6	NC_004463.1	+	696568	696630	←protease_HslV (cd01913) **RNA**→ IGPD (pfam00475)→

Rpa-3-2	NC_007958.1	-	4444159	4444098	YdeM (cd03454)→ **RNA**→ YdeM (cd03454)→
Nsp-1-3	NZ_AAMY01000001.1	+	105868	105929	hypo→ **RNA**→ ←Chorismate_synt (pfam01264)
env-29	soil_XZS22442_b2	+	761	822	hypo→ **RNA**→
env-22	soil_XZS22442_b2	+	637	698	(shown above)
Bov-1-39	NC_009505.1	-	850064	850003	HtrB (COG1560)→ **RNA**→ Rpe (COG0036)→
Bsu-1-41	NC_004310.3	-	824865	824804	HtrB (COG1560)→ **RNA**→ Rpe (COG0036)→
Bme-1-36	NC_003317.1	+	1160774	1160835	HtrB (COG1560)→ **RNA**→ Rpe (COG0036)→
Bab-1-37	NC_006932.1	-	846566	846505	HtrB (COG1560)→ **RNA**→ Rpe (COG0036)→
Bme-2-37	NC_007618.1	-	842844	842783	HtrB (COG1560)→ **RNA**→ Rpe (COG0036)→
env-30	soil_Contig37826	-	373	312	**RNA**→ DUF1099 (pfam06499)→
Mlo-1-14	NC_002678.2	+	1281545	1281606	←Lig (COG0272) **RNA**→ ←RecN (COG0497)
Bsp-1-1	NC_009485.1	+	3669586	3669647	COG0436 (COG0436)→ **RNA**→ PotD (COG0687)→
Mlo-1-15	NC_002678.2	+	3348856	3348917	DUF861 (pfam05899)→ **RNA**→ DUF861 (pfam05899)→
Rpa-3-3	NC_007958.1	+	1988245	1988306	EFh (cd00051)NOG18955 (NOG18955)→ **RNA**→ ←DmpA (COG3191)
Mlo-1-16	NC_002678.2	+	574515	574576	←CoaD (COG0669) **RNA**→ hypo→
Bja-1-7	NC_004463.1	-	4610762	4610701	hypo→ **RNA**→ ←COG0790 (COG0790)
env-31	soil_Contig30687	+	696	757	←FadB (COG1250) **RNA**→
Bja-1-8	NC_004463.1	-	8090778	8090717	COG2105 (COG2105)NOG18972 (NOG18972)→ **RNA**→ ←hypo
Bja-1-9	NC_004463.1	+	228265	228326	hypo→ **RNA**→ ←COG2128 (COG2128)
Bja-1-10	NC_004463.1	+	2369934	2369995	REC (cd00156)CheA (COG0643)Hpt (pfam01627)→ **RNA**→ CheW (cd00732)→
Bja-1-11	NC_004463.1	-	423142	423081	REC (cd00156)CheA (COG0643)Hpt (pfam01627)→ **RNA**→ CheW (cd00732)→
Bja-1-12	NC_004463.1	-	5004801	5004740	HolB (COG0470)→ **RNA**→ MetG (COG0143)→
Bja-1-13	NC_004463.1	+	6433376	6433437	←DadA (COG0665) **RNA**→ SixA (COG2062)→
Bja-1-14	NC_004463.1	-	7538986	7538925	DLH (pfam01738)→ **RNA**→ ←Beta-lactamase (pfam00144)
Bsp-1-2	NC_009485.1	-	7356079	7356018	TadG (COG4961)→ **RNA**→ ←CSD (pfam00313)
Rpa-4-2	NC_007778.1	+	3237917	3237978	UbiH (COG0654)→ **RNA**→ ←COG0679 (COG0679)
Rpa-2-2	NC_007925.1	+	4166371	4166432	←hypo **RNA**→ ←glycosyl transferase, family 39
Nha-1-4	NC_007964.1	-	1668382	1668321	BaeS (COG0642)→ **RNA**→ COG3577 (COG3577)→
Bov-1-40	NC_009505.1	+	1492864	1492925	←COG0433 (COG0433) **RNA**→ RplY (COG1825)→
Bsu-1-42	NC_004310.3	+	1485435	1485496	←COG0433 (COG0433) **RNA**→ RplY (COG1825)→
Bab-1-38	NC_006932.1	+	1503498	1503559	←COG0433 (COG0433) **RNA**→ RplY (COG1825)→
Bme-2-38	NC_007618.1	+	1500652	1500713	←COG0433 (COG0433) **RNA**→ RplY (COG1825)→
Bme-1-37	NC_003317.1	-	502706	502645	←COG0433 (COG0433) **RNA**→ RplY (COG1825)→
Rpa-3-4	NC_007958.1	-	1061777	1061716	←COG1434 (COG1434) **RNA**→ ←FtsX (COG2177)
Rpa-3-5	NC_007958.1	+	1936550	1936611	Crp (COG0664)→ **RNA**→ ←hypo
Mlo-1-17	NC_002678.2	+	5895178	5895239	COG2423 (COG2423)→ **RNA**→ PepP (COG0006)→
Oan-1-62	NC_009667.1	-	1662273	1662212	ProW (COG4176)→ **RNA**→ ProV (COG4175)→
Bja-1-15	NC_004463.1	+	1486769	1486830	←TrkA (COG2072) **RNA**→ ←hypo
Rpa-3-6	NC_007958.1	-	2091093	2091032	CaiC (COG0318)→ **RNA**→ PAS (cd00130)Tar (COG0840)→
Rpa-4-3	NC_007778.1	-	2025157	2025096	PheA (COG0077)→ **RNA**→ ←COG2141 (COG2141)
Nha-1-5	NC_007964.1	-	3992533	3992472	COG0820 (COG0820)→ **RNA**→ hypo→
Oan-1-63	NC_009667.1	-	1111125	1111064	LysR_substrate (pfam03466)→ **RNA**→ ←DUF1275 (pfam06912)
Bja-1-16	NC_004463.1	-	4190698	4190637	←NAD_synthase (cd00553)COG0388 (COG0388) **RNA**→ ←hypo
Bja-1-17	NC_004463.1	-	6227812	6227751	←MhpC (COG0596) **RNA**→ ←pCLME (cd01597)
Bja-1-18	NC_004463.1	-	5798916	5798855	ARA1 (COG0656)→ **RNA**→ hypo→
Bja-1-19	NC_004463.1	+	705771	705832	←REC (cd00156)BaeS (COG0642)PAS (smart00091) **RNA**→ ←hypo
Bja-1-20	NC_004463.1	-	7397623	7397562	←FMO-like (pfam00743) **RNA**→ ←COG1917 (COG1917)NOG23552 (NOG23552)
Bja-1-21	NC_004463.1	-	5720459	5720398	hypo→ **RNA**→ ←PanC (cd00560)
Bja-1-22	NC_004463.1	-	1046794	1046733	Gly_His_Pro_Ser_Thr_tRS_core (cd00670)→ **RNA**→ hypo→
Bja-1-23	NC_004463.1	-	4209651	4209590	←Methyltransf_6 (pfam03737)NOG11613 (NOG11613) **RNA**→ ←PssA (COG1183)
Bja-1-24	NC_004463.1	+	837493	837554	←Pnp (COG1185) **RNA**→ ←Ribosomal_S15p_S13e (cd00353)
Bsp-2-2	NC_009445.1	+	7265079	7265140	hypo→ **RNA**→ ←AcrB (COG0841)

Bsp-1-3	NC_009485.1	-	475693	475632	←putative transposase **RNA**→ ←HupE_UreJ (pfam04955)
Bja-1-25	NC_004463.1	-	6681913	6681852	COG1042 (COG1042)Acetyltransf_1 (pfam00583)→ **RNA**→ ←hypo
Bja-1-26	NC_004463.1	+	7329902	7329963	←hypo **RNA**→ SLT (pfam01464)→
Rpa-3-7	NC_007958.1	+	3408581	3408642	←DnaJ (pfam00226) **RNA**→ ←hypo
env-32	soil_XZS20120_g2	-	320	259	G6PD_C (pfam02781)→ **RNA**→ ←hypo
Nwi-1-12	NC_007406.1	-	2648313	2648252	←hypo **RNA**→ ←QPRTase (cd01572)
Rpa-3-8	NC_007958.1	+	957456	957517	←GreA (COG0782) **RNA**→ ←YjgH (cd02198)
Mlo-1-18	NC_002678.2	+	2150099	2150160	BaeS (COG0642)→ **RNA**→ RfaG (COG0438)→
Nwi-1-13	NC_007406.1	-	1236260	1236198	←tetratricopeptide TPR_4 protein **RNA**→ ←COG3334 (COG3334)NOG08792 (NOG08792)
Nha-1-6	NC_007964.1	-	4070674	4070613	DapD (COG2171)→ **RNA**→ ArgE (COG0624)→
Nha-1-7	NC_007964.1	-	2885300	2885239	transcriptional regulator, MucR family→ **RNA**→ **RNA**→ ←COG2166 (COG2166)
Nha-1-16	NC_007964.1	-	2885416	2885355	(shown above)
Mlo-1-19	NC_002678.2	+	974799	974860	Lrp (COG1522)→ **RNA**→ ←MauG (COG1858)
Oan-1-64	NC_009667.1	-	1710158	1710097	←COG0433 (COG0433) **RNA**→ RplY (COG1825)→
Bsp-1-4	NC_009485.1	-	4025091	4025030	COG1910 (COG1910)ModE (COG2005)→ **RNA**→ ←TerC (COG0861)
env-33	soil_Contig36811	+	533	594	adh_short (pfam00106)→ **RNA**→ CaiC (COG0318)→
env-34	soil_Contig22746	-	1563	1502	**RNA**→ hypo→
Mlo-1-20	NC_002678.2	+	2387353	2387414	hypo→ **RNA**→ hypo→
Nha-1-8	NC_007964.1	+	2687762	2687823	Transposase_11 (pfam01609)→ **RNA**→ ←Uncharacterised conserved protein UCP032025
Nwi-1-14	NC_007406.1	+	3268245	3268306	DUF697 (pfam05128)→ **RNA**→ ←ArnT (COG1807)
Rpa-3-9	NC_007958.1	-	663527	663468	←tRNA-Ser **RNA**→ ←NagC (COG1940)
Nha-1-9	NC_007964.1	+	1463938	1463998	WcaG (COG0451)→ **RNA**→ RfaE (COG2870)→
Nwi-1-15	NC_007406.1	-	3039517	3039457	←OmpA (COG2885) **RNA**→ ←FrdB (COG0479)
Nsp-1-4	NZ_AAMY01000002.1	+	126936	126997	DUF1052 (pfam06319)→ **RNA**→ hypo→
Rpa-4-4	NC_007778.1	-	335740	335679	←tRNA-Gly **RNA**→ COG3118 (COG3118)→
Nsp-1-5	NZ_AAMY01000024.1	-	21836	21775	AcrA (COG0845)→ **RNA**→ ←COG4249 (COG4249)
Rpa-3-10	NC_007958.1	-	2197407	2197346	←hypo **RNA**→ Ntn_CGH (cd01902)→
Rpa-3-11	NC_007958.1	+	4358727	4358788	COG2860 (COG2860)→ **RNA**→ ←AroE (COG0169)
Oan-1-65	NC_009667.1	-	2298950	2298889	PhnD (COG3221)→ **RNA**→ COG3639 (COG3639)→
Rpa-3-12	NC_007958.1	+	1792753	1792814	MetK (COG0192)→ **RNA**→ AdoHcyase (pfam05221)→
Nwi-1-16	NC_007406.1	+	520956	521017	hypo→ **RNA**→ ←PyrD (COG0167)
Nwi-1-17	NC_007406.1	-	54953	54892	←GCD1 (COG1208) **RNA**→ ←hypo
Nha-1-10	NC_007964.1	-	2356391	2356330	Rieske_cytochrome_bc1 (cd03470)→ **RNA**→ ←Rnd (COG0349)
Nha-1-11	NC_007964.1	+	1579938	1579999	FliQ (COG1987)→ **RNA**→ FliR (COG1684)→
Bsp-1-5	NC_009485.1	+	5440377	5440438	CsgD (COG2771)→ **RNA**→ ←hypo
env-35	soil_XZS72327_b1	+	699	760	Zn-dependent proteases→ **RNA**→
Rpa-3-13	NC_007958.1	-	4336343	4336282	←hypo **RNA**→ DUF1330 (pfam07045)NOG31707 (NOG31707)→
Nha-1-12	NC_007964.1	+	1487075	1487136	Ttg2C (COG1463)→ **RNA**→ ←Herpes_gp2 (pfam05955)
Nsp-1-6	NZ_AAMY01000001.1	+	22406	22468	←molybdate transport permease protein **RNA**→ ModC (COG4149)→
env-37	soil_XZS97685_g1	-	518	457	hypo→ **RNA**→ SCO (cd02968)→
Nwi-1-18	NC_007406.1	+	3184150	3184211	COG3335 (COG3335)→ **RNA**→ Bacterial_IMPase_like_1 (cd01641)→
Nsp-1-7	NZ_AAMY01000013.1	+	90994	91056	ligand_gated_channel (cd01347)→ **RNA**→ MbtH (pfam03621)→
Nwi-1-19	NC_007406.1	-	131538	131477	DUF1052 (pfam06319)→ **RNA**→ ←COG4567 (COG4567)
Bsp-1-6	NC_009485.1	-	7307545	7307484	COG4095 (COG4095)→ **RNA**→ ←ErfK (COG1376)
Nwi-1-20	NC_007406.1	+	1079530	1079591	←hypo **RNA**→ ←Pyruvate_Kinase (cd00288)
Nwi-1-21	NC_007406.1	+	588983	589044	←FlhA (COG1298) **RNA**→ tRNA-Met→
Nsp-1-8	NZ_AAMY01000014.1	-	36577	36516	PRC (pfam05239)→ **RNA**→ ←hypo
Nwi-1-22	NC_007406.1	+	2248953	2249013	COG3335 (COG3335)→ **RNA**→ Crp (COG0664)→
env-38	soil_XZS18544_b2	+	67	128	**RNA**→ ←TPP_ComE (cd03372)
Oan-1-66	NC_009667.1	-	1409908	1409847	Glk (COG0837)→ **RNA**→ MdlB (COG1132)→
Oan-1-67	NC_009667.1	-	2830634	2830573	←hypo **RNA**→ ←CaiC (COG0318)
Oan-1-69	NC_009668.1	+	1661220	1661281	MviM (COG0673)→ **RNA**→ ←MviM (COG0673)

Nwi-1-23	NC_007406.1	+	177499	177560	LysS (COG1384)→ **RNA**→ ErfK_YbiS_YhnG (pfam03734)→
Nwi-1-24	NC_007406.1	+	175764	175825	SBP_bac_3 (pfam00497)→ **RNA**→ LysS (COG1384)→
Nha-1-13	NC_007964.1	+	1877445	1877506	←EmrE (COG2076) **RNA**→ PrfC (COG4108)→
Nha-1-14	NC_007964.1	-	2266124	2266063	Qor (COG0604)→ **RNA**→ hypo→
Nwi-1-25	NC_007406.1	-	2200896	2200835	←Cfa (COG2230) **RNA**→ hypo→
Nwi-1-26	NC_007406.1	+	973329	973390	RpoE (COG1595)→ **RNA**→ FecR (COG3712)→
Nsp-1-9	NZ_AAMY01000028.1	-	6956	6895	Gln_GluRS_non_core (cd00807)tRNA-synt_1c_C (pfam03950)→ **RNA**→ Competence (pfam03772)→
Rpa-3-14	NC_007958.1	+	643892	643953	←DUF143 (pfam02410) **RNA**→ ←NadD (COG1057)
Nsp-1-10	NZ_AAMY01000010.1	+	60695	60756	COG2329 (COG2329)→ **RNA**→ ←RpmG (COG0267)
Nsp-1-11	NZ_AAMY01000014.1	-	11894	11835	RpoD (COG0568)Sigma70_r1_2 (pfam00140)Sigma70_r1_1 (pfam03979)→ **RNA**→ Transposase_8 (pfam01527)→
Nsp-1-12	NZ_AAMY01000034.1	-	26988	26927	COG2192 (COG2192)→ **RNA**→ ETC_C1_NDUFA4 (pfam04800)→
env-39	soil_XZS87262_y1	-	173	112	hypo→ **RNA**→
Nha-1-15	NC_007964.1	+	491389	491450	PckA (COG1866)→ **RNA**→ PenP (COG2367)→
Rpa-3-15	NC_007958.1	-	4614181	4614120	OmpA (COG2885)→ **RNA**→ COG1729 (COG1729)→
env-40	soil_XZS76604_x1	+	261	322	←hypo **RNA**→ PaaI (COG2050)→
Rpa-3-16	NC_007958.1	-	4632013	4631952	Ndh (COG1252)→ **RNA**→ COG4645 (COG4645)→
Nwi-1-27	NC_007406.1	+	1336232	1336293	MRP-like (cd02037)→ **RNA**→ FcbT1 (COG4663)→
Rpa-3-17	NC_007958.1	+	934551	934612	DsbA_HCCA_Iso (cd03022)→ **RNA**→ hypo→
Rpa-3-18	NC_007958.1	-	3823955	3823894	←CypX (COG2124) **RNA**→ ←hypo
Nha-1-17	NC_007964.1	+	1450276	1450337	LpxC (pfam03331)→ **RNA**→ ComL (COG4105)→
Rpa-3-19	NC_007958.1	-	2522521	2522460	←GppA (COG0248) **RNA**→ CoxL (COG1529)→
Nha-1-18	NC_007964.1	-	3321684	3321623	hypo→ **RNA**→ ←hypo
Nsp-1-13	NZ_AAMY01000010.1	+	43655	43715	←hypo **RNA**→ ←hypo
Nsp-1-14	NZ_AAMY01000011.1	+	436	496	←Nitrous oxidase accessory protein **RNA**→ ←AraJ (COG2814)
Nwi-1-28	NC_007406.1	-	434926	434866	←SecA_protein (pfam01043)SEC-C (pfam02810) **RNA**→ Rotamase (pfam00639)→
Rpa-2-3	NC_007925.1	+	2129897	2129958	COG2822 (COG2822)→ **RNA**→ Cobalamin riboswitch→
Nha-1-19	NC_007964.1	-	1256968	1256907	GOX_like_FMN_family (cd02809)→ **RNA**→ ←FeoB (pfam02421)
Nwi-1-29	NC_007406.1	-	166244	166183	←CrcB (COG0239) **RNA**→ ←DCD (pfam06559)
Nsp-1-15	NZ_AAMY01000010.1	+	63940	64001	hypo→ **RNA**→ ValRS_core (cd00817)→
Nsp-1-16	NZ_AAMY01000002.1	+	135024	135085	←MrcB (COG0744) **RNA**→ ←GlnD (COG2844)
Nwi-1-30	NC_007406.1	+	143465	143526	←CoaA (COG1072) **RNA**→ ←HisI (COG0140)
Nwi-1-31	NC_007406.1	+	2206723	2206784	Sms (COG1066)→ **RNA**→ Colicin_V (pfam02674)→
Nha-1-20	NC_007964.1	+	1567296	1567357	←FlgA (pfam03240)NOG07957 (NOG07957) **RNA**→ ←FlgG (COG4786)
Nha-1-21	NC_007964.1	+	2895903	2895964	FolK (COG0801)→ **RNA**→ MM_CoA_mutase (pfam01642)→
Nwi-1-32	NC_007406.1	-	2544048	2543987	←COG2329 (COG2329) **RNA**→ HutB (cd01149)→
Nha-1-22	NC_007964.1	+	1442993	1443054	MurB (COG0812)→ **RNA**→ Dala_Dala_ligas (pfam01820)→
Nwi-1-33	NC_007406.1	+	2380828	2380889	←Peptidase_M29 (pfam02073) **RNA**→ ←hypo
Nwi-1-34	NC_007406.1	+	3336665	3336726	←ArgB (COG0548) **RNA**→ ←Protein of unknown function DUF423
Bsp-1-7	NC_009485.1	+	2908203	2908265	DsbD (pfam02683)→ **RNA**→ ←COG3920 (COG3920)
Bsp-1-8	NC_009485.1	-	3144925	3144864	hypo→ **RNA**→ ←DUF1109 (pfam06532)
Bsp-1-9	NC_009485.1	+	1123596	1123657	COG1835 (COG1835)→ **RNA**→ Mfd (COG1197)→
Bsp-1-10	NC_009485.1	+	174004	174065	Spo0J (COG1475)→ **RNA**→ ←HolA (COG1466)
Bsp-1-11	NC_009485.1	-	3175906	3175845	putative membrane protein of unknown function→ **RNA**→ ←PutA (COG1012)
Bsp-1-12	NC_009485.1	+	1880787	1880848	hypo→ **RNA**→ ←GOX_like_FMN_family (cd02809)
Bsp-1-13	NC_009485.1	+	1816679	1816740	←ProC (COG0345) **RNA**→ ←COG5465 (COG5465)
Bsp-1-14	NC_009485.1	+	3186351	3186412	putative exported protein of unknown function→ **RNA**→ ←ACH1 (COG0427)
Bsp-1-15	NC_009485.1	-	4352542	4352481	Qor (COG0604)→ **RNA**→ ←hypo
Bsp-1-16	NC_009485.1	+	6463308	6463369	LivK (COG0683)→ **RNA**→ OmpR (COG0745)→
Nwi-1-35	NC_007406.1	-	1220601	1220540	hypo→ **RNA**→ ←hypo
Nha-1-23	NC_007964.1	-	3556411	3556349	MutY (COG1194)→ **RNA**→ ←GST_C_Ure2p_like (cd03178)GST_N_Ure2p_like (cd03048)
Nsp-1-17	NZ_AAMY01000014.1	+	41333	41394	ThiS (cd00565)→ **RNA**→ ThiG (COG2022)→
Nwi-1-36	NC_007406.1	+	2689024	2689085	ThiS (cd00565)NOG15857 (NOG15857)→ **RNA**→ ThiG (COG2022)→

Mlo-1-21	NC_002678.2	+	2148552	2148613	OmpR (COG0745)→ **RNA**→ BaeS (COG0642)→
Nha-1-24	NC_007964.1	+	1862150	1862211	←HPT (cd00088)NOG09935 (NOG09935) **RNA**→ ←REC (cd00156)
Nsp-1-18	NZ_AAMY01000001.1	-	284311	284249	COG2932 (COG2932)→ **RNA**→ ←Acs (COG0365)
Nsp-1-19	NZ_AAMY01000010.1	+	83452	83513	HflX (COG2262)→ **RNA**→ PutA (COG1012)→
Rpa-3-20	NC_007958.1	-	775180	775119	←protein of unknown function DUF1486 **RNA**→ ←DltE (COG0300)
Nsp-1-20	NZ_AAMY01000010.1	+	31195	31256	hypo→ **RNA**→ ←Polysacc_deac_1 (pfam01522)
Rpa-2-4	NC_007925.1	-	5171648	5171588	hypo→ **RNA**→ ←COG4191 (COG4191)CHASE3 (pfam05227)
Rpa-3-21	NC_007958.1	+	1018799	1018860	DCD (pfam06559)→ **RNA**→ ←Twin-arginine translocation pathway signal
Rpa-3-22	NC_007958.1	-	951254	951193	←COG0523 (COG0523) **RNA**→ ←PsaA (cd01137)
Rpa-3-23	NC_007958.1	+	3621746	3621807	COG3920 (COG3920)→ **RNA**→ ←hypo
Rpa-3-24	NC_007958.1	+	225176	225237	FolC (COG0285)→ **RNA**→ ←Thioredoxin (pfam00085)
Rpa-3-25	NC_007958.1	-	374758	374697	putative Omp2b porin→ **RNA**→ ACH1 (COG0427)→
Rpa-3-26	NC_007958.1	+	2144251	2144312	hypo→ **RNA**→ ←TolC (COG1538)
Rpa-3-27	NC_007958.1	-	4634449	4634388	RuvB (COG2255)→ **RNA**→ Plasmid_stabil (pfam05016)→
Nha-1-25	NC_007964.1	-	2725326	2725265	←Poly_export (pfam02563) **RNA**→ GumC (COG3206)→
Nha-1-26	NC_007964.1	-	3217704	3217643	←hypo **RNA**→ ←EmrA (COG1566)
Nha-1-27	NC_007964.1	+	189341	189402	protein of unknown function DUF1150→ **RNA**→ ILVD_EDD (pfam00920)→
Nha-1-28	NC_007964.1	-	192468	192407	PtsN (COG1762)→ **RNA**→ hypo→
Nwi-1-37	NC_007406.1	+	2950118	2950180	←hypo **RNA**→ ←COG5001 (COG5001)
Nha-1-29	NC_007964.1	-	2552577	2552516	←COG2339 (COG2339) **RNA**→ ←hypo
Nsp-1-21	NZ_AAMY01000003.1	+	40297	40358	←hypo **RNA**→ ←RnhB (COG0164)
Nha-1-30	NC_007964.1	-	691015	690954	hypo→ **RNA**→ ←EAL (cd01948)DUF1 (smart00267)
Nsp-1-22	NZ_AAMY01000001.1	-	110406	110345	DnaJ (COG0484)DnaJ (pfam00226)→ **RNA**→ ←hypo
Nsp-1-23	NZ_AAMY01000024.1	-	30457	30396	AsmA (COG2982)→ **RNA**→ ←COG3576 (COG3576)
Nwi-1-38	NC_007406.1	+	1455909	1455970	hypo→ **RNA**→ ←DUF37 (pfam01809)
Nha-1-31	NC_007964.1	-	3038401	3038339	←Ugd (COG1004) **RNA**→ hypo→
Nsp-1-24	NZ_AAMY01000006.1	-	124549	124488	hypo→ **RNA**→ ←hypo
Nsp-1-25	NZ_AAMY01000018.1	-	56257	56196	ClpA (COG0542)→ **RNA**→ hypo→
Rpa-2-5	NC_007925.1	+	5321038	5321099	HigB (COG3549)→ **RNA**→ VapI (COG3093)→
Rpa-1-2	NC_008435.1	-	5377089	5377028	hypo→ **RNA**→ hypo→
Rpa-2-6	NC_007925.1	+	1529801	1529862	hypo→ **RNA**→ hypo→
Rpa-2-7	NC_007925.1	-	862139	862078	protein of unknown function DUF81→ **RNA**→ ←RhtB (COG1280)
Nha-1-32	NC_007964.1	-	254891	254834	←BaeS (COG0642) **RNA**→ DepA (COG4553)→
Nha-1-33	NC_007964.1	-	138281	138222	HslU (COG1220)→ **RNA**→ ←COG2840 (COG2840)
Nsp-1-26	NZ_AAMY01000009.1	-	46615	46554	hypo→ **RNA**→ MMT1 (COG0053)→
Nwi-1-39	NC_007406.1	+	3227629	3227690	←hypo **RNA**→ ←hypo
Nsp-1-27	NZ_AAMY01000001.1	-	287171	287110	CDC9 (COG1793)→ **RNA**→ hypo→
Nsp-1-28	NZ_AAMY01000002.1	+	152664	152725	LysS (COG1384)→ **RNA**→ ErfK_YbiS_YhnG (pfam03734)→
Nsp-1-29	NZ_AAMY01000003.1	-	152371	152310	←COG1051 (COG1051) **RNA**→ ←hypo
Nwi-1-40	NC_007406.1	+	1971474	1971535	←CoaD (COG0669) **RNA**→ hypo→
Nwi-1-41	NC_007406.1	-	221165	221108	HemN (COG0635)→ **RNA**→ ←LivK (COG0683)
Nsp-1-30	NZ_AAMY01000002.1	-	229763	229702	DnaQ (COG0847)→ **RNA**→ ←SecB (COG1952)
Nha-1-34	NC_007964.1	+	614924	614985	hypo→ **RNA**→ ←PyrD (COG0167)
Nha-1-35	NC_007964.1	-	2030886	2030825	←PrfB (COG1186) **RNA**→ ←MrcA (COG5009)
Nha-1-36	NC_007964.1	+	2511451	2511512	COG1573 (COG1573)→ **RNA**→ ←hypo
Nha-1-37	NC_007964.1	+	4015482	4015543	LivK (COG0683)→ **RNA**→ LivH (COG0559)→
Nha-1-38	NC_007964.1	-	619626	619564	DUF589 (pfam04543)→ **RNA**→ Acs (COG0365)→
Nha-1-39	NC_007964.1	-	437751	437691	←TadE-like **RNA**→ ←hypo
Nha-1-40	NC_007964.1	+	3068182	3068243	GalE (COG1087)→ **RNA**→ ManC (COG0662)ManC (COG0836)→
Nha-1-41	NC_007964.1	-	1683868	1683807	←DAHP_synth_2 (pfam01474) **RNA**→ ←Lpd (COG1249)
Nha-1-42	NC_007964.1	-	3609764	3609703	COG5001 (COG5001)→ **RNA**→ SGA1 (COG3387)→
Nha-1-43	NC_007964.1	-	1687618	1687557	hypo→ **RNA**→ ←NAD_synthase (cd00553)COG0388 (COG0388)

Nsp-1-31	NZ_AAMY01000003.1	-	158402	158341	hypo→ **RNA**→ ←COG5350 (COG5350)
Nha-1-44	NC_007964.1	-	4382843	4382782	←COG1051 (COG1051) **RNA**→ COG3415 (COG3415)→
Nha-1-45	NC_007964.1	+	237745	237806	←CrcB (COG0239) **RNA**→ ←DCD (pfam06559)
Nwi-1-42	NC_007406.1	-	2600690	2600629	←Ugd (COG1004) **RNA**→ hypo→
Nwi-1-43	NC_007406.1	+	1446183	1446244	hypo→ **RNA**→ **RNA**→ COG3415 (COG3415)→
Nwi-1-70	NC_007406.1	+	1446286	1446346	(shown above)
Nsp-1-32	NZ_AAMY01000023.1	+	34850	34911	←SpoU (COG0566) **RNA**→ ←hypo
Nsp-1-33	NZ_AAMY01000007.1	-	9750	9689	hypo→ **RNA**→ ←MltE (COG0741)
Nsp-1-34	NZ_AAMY01000004.1	-	168261	168201	←PDZ_serine_protease (cd00987)DegQ (COG0265) **RNA**→ ←hypo
Nsp-1-35	NZ_AAMY01000004.1	+	144448	144510	COG4312 (COG4312)→ **RNA**→ ←APP (cd01085)
Nsp-1-36	NZ_AAMY01000008.1	-	83799	83739	hypo→ **RNA**→ hypo→
Nsp-1-37	NZ_AAMY01000003.1	+	254237	254298	←Asd (COG0136) **RNA**→ ←hypo
Nsp-1-38	NZ_AAMY01000038.1	-	20019	19958	←hypo **RNA**→ ←hypo
Nsp-1-39	NZ_AAMY01000007.1	+	35322	35383	hypo→ **RNA**→ ←COG4799 (COG4799)
Nsp-1-40	NZ_AAMY01000003.1	-	95228	95167	COG5001 (COG5001)→ **RNA**→ SGA1 (COG3387)→
Nsp-1-41	NZ_AAMY01000020.1	-	16833	16772	←tRNA-Lys **RNA**→ ←COG1215 (COG1215)COG5309 (COG5309)
Nsp-1-42	NZ_AAMY01000001.1	+	307273	307334	Lpd (COG1249)→ **RNA**→ tyrosine recombinase→
Nwi-1-44	NC_007406.1	-	420497	420436	HemK (COG2890)→ **RNA**→ hypo→
Nsp-1-43	NZ_AAMY01000008.1	-	40151	40090	PldB (COG2267)→ **RNA**→ ←Bacterial_IMPase_like_1 (cd01641)
Nsp-1-44	NZ_AAMY01000019.1	-	21765	21704	Sms (COG1066)→ **RNA**→ Colicin_V (pfam02674)→
Nsp-1-45	NZ_AAMY01000002.1	-	132802	132741	MrcB (COG0744)→ **RNA**→ ←COG1451 (COG1451)
Nsp-1-46	NZ_AAMY01000002.1	+	207865	207926	HemN (COG0635)→ **RNA**→ ←LivK (COG0683)
Nsp-1-47	NZ_AAMY01000016.1	-	24092	24031	AtoS (COG2202)COG4191 (COG4191)HAMP (pfam00672)→ **RNA**→ ←hypo
Nwi-1-45	NC_007406.1	-	705107	705046	LasI (COG3916)→ **RNA**→ hypo→
env-41	soil_XZS13595_g1	-	269	208	EAL (cd01948)→ **RNA**→
Rpa-4-5	NC_007778.1	-	3377298	3377237	ThrA (COG0460)→ **RNA**→ FBPase_glpX (cd01516)→
Rpa-3-28	NC_007958.1	-	393532	393471	ChrR (COG3806)→ **RNA**→ ←PerM (COG0628)
Rpa-3-29	NC_007958.1	+	4876078	4876139	←COG3212 (COG3212) **RNA**→ ←hypo
Rpa-3-30	NC_007958.1	+	3943773	3943834	HAMP (pfam00672)MA (smart00283)→ **RNA**→ ←adh_short (pfam00106)
Rpa-3-31	NC_007958.1	-	3218698	3218637	hypo→ **RNA**→ ←PepD (COG2195)
Rpa-3-32	NC_007958.1	+	378091	378152	MarR (pfam01047)NOG19002 (NOG19002)→ **RNA**→ ←hypo
Rpa-3-33	NC_007958.1	-	3125707	3125646	COG0790 (COG0790)→ **RNA**→ ←XthA (COG0708)
Nsp-1-48	NZ_AAMY01000014.1	+	30906	30967	LysR (COG0583)→ **RNA**→ ←Soj (COG1192)
Nsp-1-49	NZ_AAMY01000017.1	-	40243	40182	hypo→ **RNA**→ COG3268 (COG3268)→
Rpa-3-34	NC_007958.1	-	742333	742272	DHDPS (cd00950)→ **RNA**→ ←hypo
Nwi-1-46	NC_007406.1	-	631429	631368	←hypo **RNA**→ ←hypo
Mlo-1-22	NC_002678.2	+	3685297	3685359	TrxB (COG0492)→ **RNA**→ TrxB (COG0492)→
env-42	soil_Contig34393	+	62	124	**RNA**→ ←PurR (COG1609)
Csp-1-1	NZ_AATH01000009.1	-	106195	106134	hypo (cons)→ **RNA**→ ←NirD (COG2146)
Csp-1-2	NZ_AATH01000004.1	-	137699	137638	←hypo (cons) **RNA**→ ←hypo
Csp-1-3	NZ_AATH01000004.1	-	319472	319411	COG3920 (COG3920)→ **RNA**→ Pyridox_oxidase (pfam01243)→
Ccr-1-1	NC_002696.2	-	705859	705798	COG2081 (COG2081)→ **RNA**→ ←HTH_ARAC (smart00342)
Oan-1-70	NC_009667.1	+	1008604	1008665	←COG3897 (COG3897) **RNA**→ ←COG2947 (COG2947)
env-43	wf1_AGZO15032_g2	+	153	213	**RNA**→ ←COG3536 (COG3536)
env-44	soil_Contig36324	-	1052	991	←Spy (COG3914) **RNA**→ ←Spy (COG3914)
env-45	soil_XZS50760_g1	+	382	442	Fe-S-cluster-containing hydrogenase components 1→ **RNA**→ Cytochrome b subunit of formate dehydrogenase→
env-46	soil_XZS59981_b1	-	700	641	hypo→ **RNA**→ DUF1194 (pfam06707)→
Sme-2-3	NC_003047.1	+	3086653	3086711	TPR (cd00189)→ **RNA**→ CONSERVED HYPOTHETICAL SIGNAL PEPTIDE PROTEIN→
Oan-1-71	NC_009667.1	-	871155	871095	hypo→ **RNA**→ hypo→
env-47	JCVI_SCAF_1101668558192	+	363	423	? (GOS1286)→ **RNA**→ ←LeuB (COG0473)? (GOS1809)
env-48	soil_XZS39199_g1	+	240	300	ABC-type sugar transport systems, ATPase components→ **RNA**→ Adenylate cyclase, family 3 (some proteins contain HAMP d
Rle-1-43	NC_008380.1	+	514685	514746	COG4782 (COG4782)→ **RNA**→ DadA (COG0665)→

Rle-1-44	NC_008380.1	+	503730	503790	putative transmembrane protein→ **RNA**→ hypo→
Sme-1-2	NC_009636.1	-	593521	593461	←PpsA (COG0574) **RNA**→ ←Smc (COG1196)SMC_N (pfam02463)
Sme-2-4	NC_003078.1	-	507620	507560	cysteine_hydrolases (cd00431)→ **RNA**→ ←AraD (COG0235)
env-49	sludgePhrap_Contig1238	+	590	648	STE14 (COG2020)→ **RNA**→ ←(,,,)
env-50	soil_XZS96631_b1	+	90	150	**RNA**→ PTSIIA_gutA (pfam03829)→
Mlo-1-23	NC_002678.2	+	2844564	2844624	←COG3576 (COG3576) **RNA**→ ←DUF1348 (pfam07080)
Mlo-1-24	NC_002678.2	+	4580854	4580914	MscL (pfam01741)→ **RNA**→ COG0436 (COG0436)→
Mlo-1-25	NC_002678.2	+	3857112	3857172	←PseudoU_synth_EcTruA (cd02570) **RNA**→ ←COG3153 (COG3153)
Sme-1-3	NC_009636.1	-	1682338	1682277	COG0790 (COG0790)SEL1 (smart00671)→ **RNA**→ ←MobB (cd03116)
Rle-1-45	NC_008380.1	-	4532957	4532897	←ACR_tran (pfam00873) **RNA**→ ←AcrA (COG0845)
Ret-1-9	NC_007761.1	+	3823222	3823282	←PntA (COG3288) **RNA**→ ←hypo
Rle-1-46	NC_008380.1	+	276951	277011	GumC (COG3206)→ **RNA**→ hypo→
Rle-1-47	NC_008380.1	+	3735475	3735535	←COG4649 (COG4649) **RNA**→ ←NnrU (pfam07298)
Rle-1-48	NC_008380.1	+	4581433	4581493	CAP_ED (cd00038)RosB (COG4651)→ **RNA**→ ←hypo
Ret-1-10	NC_007761.1	+	930867	930927	COG4564 (COG4564)MA (smart00283)→ **RNA**→ ←DUF523 (pfam04463)
Msp-1-1	NC_008254.1	+	1016961	1017021	hypo→ **RNA**→ ←COG4770 (COG4770)
Sme-1-4	NC_009636.1	-	404908	404848	←GloB (COG0491) **RNA**→ AraJ (COG2814)→
Bab-1-39	NC_006933.1	+	402280	402340	Lhr (COG1201)→ **RNA**→ Lhr (COG1201)→
Bme-2-39	NC_007624.1	+	402350	402410	Lhr (COG1201)→ **RNA**→ Lhr (COG1201)→
Bov-1-41	NC_009504.1	-	816582	816522	hypo→ **RNA**→ COG1407 (COG1407)→
Bsu-1-43	NC_004311.2	-	814837	814777	Lhr (COG1201)→ **RNA**→ Lhr (COG1201)→
Bme-1-38	NC_003318.1	+	483530	483590	Lhr (COG1201)→ **RNA**→ Lhr (COG1201)→
env-51	wf2_AHAA9600_y1	-	273	212	**RNA**→ ←hypo
Nha-1-46	NC_007964.1	-	3172960	3172900	←GreA (COG0782) **RNA**→ ←MGS_CPS_II (cd01424)CarB (COG0458)CPSase_L_D3 (pfam02787)
Mch-1-10	NZ_ABEX01000002.1	-	242550	242490	←XthA (COG0708) **RNA**→ ←BLUF (pfam04940)
Mch-1-12	NZ_ABEX01000006.1	-	3969	3909	←hypo **RNA**→ ←hypo
Mex-1-12	NZ_ABEC01000006.1	+	198293	198353	HrcA (COG1420)→ **RNA**→ DUF6 (pfam00892)→
Mch-1-13	NZ_ABEX01000023.1	-	33592	33532	←Nitroreductase_2 (cd02138) **RNA**→ CaiD (COG1024)→
Mex-1-13	NZ_ABEC01000002.1	-	126653	126593	←COG2917 (COG2917) **RNA**→ ←FtsY (COG0552)
Mex-1-14	NZ_ABEC01000003.1	-	91817	91757	COG3404 (COG3404)→ **RNA**→ ←hypo
Mex-1-15	NZ_ABEC01000008.1	+	157323	157383	FabB (COG0304)→ **RNA**→ KAS_I_II (cd00834)→
Mex-1-16	NZ_ABEC01000008.1	+	58043	58103	COG0714 (COG0714)→ **RNA**→ COG3609 (COG3609)→
Mch-1-14	NZ_ABEX01000009.1	-	20577	20517	COG0790 (COG0790)SEL1 (smart00671)→ **RNA**→ hypo→
Mex-1-17	NZ_ABEC01000009.1	+	276290	276350	MltB (COG2951)PG_binding_1 (pfam01471)→ **RNA**→ SGNH_hydrolase_peri2 (cd01829)→
Mch-1-15	NZ_ABEX01000006.1	+	308518	308578	MltB (COG2951)→ **RNA**→
Mch-1-16	NZ_ABEX01000002.1	-	226561	226501	NnrU (pfam07298)→ **RNA**→ COG4649 (COG4649)→
Mex-1-18	NZ_ABEC01000002.1	+	770130	770190	COG2070 (COG2070)→ **RNA**→ ←RpsD (COG0522)
Mex-1-19	NZ_ABEC01000010.1	-	72120	72060	hypo (cons)→ **RNA**→ RimK (COG0189)→
Mex-1-20	NZ_ABEC01000001.1	+	223012	223072	←ERG9 (COG1562) **RNA**→ ←COG3737 (COG3737)
env-52	soil_XZS12286_b1	+	214	274	**RNA**→ AccD (COG0777)→
Csp-1-4	NZ_AATH01000001.1	+	1279608	1279668	MntH (COG1914)→ **RNA**→ ←USP_Like (cd00293)
Csp-1-5	NZ_AATH01000001.1	-	1259144	1259084	REC (cd00156)BaeS (COG0642)AtoS (COG2202)→ **RNA**→ ←ArcC (COG0549)
Csp-1-6	NZ_AATH01000008.1	+	76268	76328	←LpqC (COG3509) **RNA**→ ←adh_short (pfam00106)
Csp-1-7	NZ_AATH01000004.1	+	328205	328265	COG0780 (COG0780)→ **RNA**→ ←TonB_dep_Rec (pfam00593)
Csp-1-8	NZ_AATH01000001.1	-	1538529	1538469	SCP-x_thiolase (cd00829)→ **RNA**→ ←LysR (COG0583)
env-53	JCVI_SCAF_1096628067764	-	278	218	AdhP (COG1064)? (GOS333)→ **RNA**→
env-54	JCVI_SCAF_1096628062076	-	1798	1738	**RNA**→ ←UhpC (COG2271)? (GOS274)
Nwi-1-47	NC_007406.1	-	1723711	1723651	←GST_C_Ure2p_like (cd03178)GST_N_Ure2p_like (cd03048) **RNA**→ NagC (COG1940)→
Rpa-3-35	NC_007958.1	+	191956	192016	hypo→ **RNA**→ ←InfC (COG0290)
Bja-1-27	NC_004463.1	-	4655603	4655543	hypo→ **RNA**→ ←adh_short (pfam00106)
Rpa-4-6	NC_007778.1	-	94765	94705	alkPPc (cd00016)→ **RNA**→ ←SNase (pfam00565)
Rpa-3-36	NC_007958.1	-	324977	324917	hypo→ **RNA**→ HELICc (cd00079)DEAD (pfam00270)→

Rpa-3-37	NC_007958.1	-	1970154	1970094	Acyl_transf_3 (pfam01757)NOG09695 (NOG09695)→ **RNA**→ ←PurL (COG0046)
Rpa-3-38	NC_007958.1	+	3724732	3724792	protein of unknown function DUF1321→ **RNA**→ Fumarase_classII (cd01362)→
Rpa-3-39	NC_007958.1	-	3596750	3596690	hypo→ **RNA**→ ←COG1002 (COG1002)
Rpa-3-40	NC_007958.1	+	4027592	4027652	hypo→ **RNA**→ ←COG3637 (COG3637)
Nwi-1-48	NC_007406.1	-	829977	829917	DegQ (COG0265)→ **RNA**→ ←COG1959 (COG1959)
Nwi-1-49	NC_007406.1	-	2713985	2713925	hypo→ **RNA**→ RimL (COG1670)→
Nwi-1-50	NC_007406.1	-	3221881	3221821	Protein of unknown function DUF992→ **RNA**→ FusA (COG0480)→
Nha-1-47	NC_007964.1	+	3225937	3225997	URO-D (cd00717)→ **RNA**→ ←SirA_YedF_YeeD (cd00291)
Nha-1-48	NC_007964.1	-	2557888	2557828	←peptidase M23B **RNA**→ ←EFP (pfam01132)
Nha-1-49	NC_007964.1	+	3179066	3179126	←hypo **RNA**→ ←hypo
Nha-1-50	NC_007964.1	-	3827269	3827209	hypo→ **RNA**→ hypo→
Nha-1-51	NC_007964.1	-	1181506	1181446	←RgpF (COG3754) **RNA**→ ←COG0790 (COG0790)SEL1 (smart00671)
Nha-1-52	NC_007964.1	+	3167754	3167814	UbiE (COG2226)NOG15446 (NOG15446)→ **RNA**→ tRNA-OTHER→
Nha-1-53	NC_007964.1	+	1142903	1142963	FixC (COG0644)→ **RNA**→ FixC (COG0644)→
Nha-1-54	NC_007964.1	-	2022919	2022859	CafA (COG1530)→ **RNA**→ ←COG3666 (COG3666)Transposase_11 (pfam01609)
Nha-1-55	NC_007964.1	+	134345	134405	DnaQ (COG0847)→ **RNA**→ ←SecB (COG1952)
Nwi-1-51	NC_007406.1	+	1363987	1364047	Kup (COG3158)→ **RNA**→ RHOD_YgaP (cd01527)→
Nha-1-56	NC_007964.1	+	3422561	3422621	hypo→ **RNA**→ ←COG0012 (COG0012)
Nha-1-57	NC_007964.1	-	252013	251953	COG1451 (COG1451)→ **RNA**→ ←MrcB (COG0744)
Nha-1-58	NC_007964.1	+	537100	537160	SecA_protein (pfam01043)SEC-C (pfam02810)→ **RNA**→ ←DUF949 (pfam06104)
Nha-1-59	NC_007964.1	+	832225	832285	KefB (COG0475)TrkA_N (pfam02254)→ **RNA**→ RibA (COG0807)DHBP_synthase (pfam00926)→
Nha-1-60	NC_007964.1	+	2997557	2997617	OpuBB (COG1174)OpuBC (COG1732)→ **RNA**→ ←IlvC (COG0059)
Nha-1-61	NC_007964.1	-	284071	284011	Invasion associated locus B→ **RNA**→ TldD (COG0312)→
Nha-1-62	NC_007964.1	-	15433	15373	←COG3666 (COG3666) **RNA**→ ←TrpS (COG0180)
Nha-1-63	NC_007964.1	+	1400509	1400569	AtoS (COG2202)COG5001 (COG5001)→ **RNA**→ PotB (COG1176)→
Nha-1-64	NC_007964.1	+	2886507	2886567	COG2166 (COG2166)→ **RNA**→ ←transcriptional regulator, MucR family
Nha-1-65	NC_007964.1	-	3205699	3205639	ErfK_YbiS_YhnG (pfam03734)→ **RNA**→ ←hypo
Nha-1-66	NC_007964.1	-	47593	47533	←Fur (COG0735) **RNA**→ FabA (cd01287)→
Nha-1-67	NC_007964.1	+	583383	583443	NUDIX (pfam00293)→ **RNA**→ ←ATP-synt_DE_N (pfam02823)
Nha-1-68	NC_007964.1	+	1931787	1931847	←MoeA (COG0303) **RNA**→ hypo→
Nha-1-69	NC_007964.1	+	3426735	3426795	COG5465 (COG5465)→ **RNA**→ ProC (COG0345)→
Nha-1-70	NC_007964.1	-	49616	49556	←LdhA (COG1052) **RNA**→ COG3807 (COG3807)→
Nha-1-71	NC_007964.1	+	484074	484134	←LepA (COG0481) **RNA**→ ←FruB (COG1925)
Nha-1-72	NC_007964.1	+	756656	756716	CysI (COG0155)→ **RNA**→ MetC (COG0626)→
Nha-1-73	NC_007964.1	+	3435033	3435093	EbsC (COG2606)→ **RNA**→ COG3666 (COG3666)Transposase_11 (pfam01609)→
Nha-1-74	NC_007964.1	+	3620975	3621035	←COG4235 (COG4235)NOG27282 (NOG27282) **RNA**→ ←COG1959 (COG1959)
Nha-1-75	NC_007964.1	-	3818453	3818393	←COG0523 (COG0523) **RNA**→ ←hypo
Nwi-1-52	NC_007406.1	-	809666	809606	←FecR (COG3712) **RNA**→ ←hypo
Nwi-1-53	NC_007406.1	+	3288346	3288406	AsmA (COG2982)→ **RNA**→ ←COG3576 (COG3576)
Nha-1-76	NC_007964.1	-	2412454	2412394	TPR (cd00189)NOG06853 (NOG06853)→ **RNA**→ Guanylate_cyc (pfam00211)CHASE2 (pfam05226)→
Nha-1-77	NC_007964.1	-	2617526	2617466	ADC_lyase (pfam02618)→ **RNA**→ COG1561 (COG1561)→
Nha-1-78	NC_007964.1	+	273642	273702	←protein of unknown function DUF81 **RNA**→ ←Transposase_11 (pfam01609)NOG35082 (NOG35082)
Nsp-1-50	NZ_AAMY01000009.1	+	24047	24107	hypo→ **RNA**→ COG3179 (COG3179)→
Nsp-1-51	NZ_AAMY01000003.1	+	57334	57394	COG3293 (COG3293)→ **RNA**→ hypo→
Nsp-1-52	NZ_AAMY01000017.1	-	66317	66257	ArsC_ArsC (cd03034)→ **RNA**→ ←COG3335 (COG3335)
Nsp-1-53	NZ_AAMY01000001.1	+	403034	403094	DUF949 (pfam06104)→ **RNA**→ ←SecA_protein (pfam01043)SEC-C (pfam02810)
Nsp-1-54	NZ_AAMY01000010.1	-	53893	53833	excinuclease ABC subunit A→ **RNA**→ RNR_1_like (cd02888)→
Nwi-1-54	NC_007406.1	-	417006	416946	hypo→ **RNA**→ hypo→
Nsp-1-55	NZ_AAMY01000002.1	+	290323	290383	TrpA (COG0159)→ **RNA**→ AccD (COG0777)→
Nsp-1-56	NZ_AAMY01000004.1	+	160861	160921	←COG4321 (COG4321) **RNA**→ ←COG4321 (COG4321)
Nha-1-79	NC_007964.1	+	1774111	1774171	RplQ (COG0203)→ **RNA**→ PDZ_serine_protease (cd00987)DegQ (COG0265)→
Nha-1-80	NC_007964.1	-	2847082	2847022	hypo→ **RNA**→ DsbA_HCCA_Iso (cd03022)→

Nha-1-81	NC_007964.1	-	4121012	4120952	←hypo **RNA**→ ←COG1238 (COG1238)
Rpa-3-41	NC_007958.1	+	768481	768541	hypo→ **RNA**→ ←DUF894 (pfam05977)
Nha-1-82	NC_007964.1	-	7041	6981	hypo→ **RNA**→ ←GyrB (COG0187)DNA_gyraseB_C (pfam00986)
Nha-1-83	NC_007964.1	-	92907	92847	PAS (cd00130)AtoS (COG2202)→ **RNA**→ COG3673 (COG3673)→
Nha-1-84	NC_007964.1	+	2573667	2573727	←DedA (COG0586) **RNA**→ hypo→
Nha-1-85	NC_007964.1	+	2374397	2374457	hypo→ **RNA**→ ←hypo
Nha-1-86	NC_007964.1	+	4388479	4388539	COG5350 (COG5350)→ **RNA**→ ←COG2071 (COG2071)
Nha-1-87	NC_007964.1	-	2540028	2539968	hypo→ **RNA**→ ←Smf (COG0758)
Nwi-1-55	NC_007406.1	+	1636783	1636843	UbiH (COG0654)→ **RNA**→ COG3785 (COG3785)→
Nwi-1-56	NC_007406.1	-	696881	696821	CaiC (COG0318)→ **RNA**→ ←COG1741 (COG1741)
Nwi-1-57	NC_007406.1	-	2048564	2048504	COG4549 (COG4549)→ **RNA**→ ProS (COG0442)→
Nwi-1-58	NC_007406.1	-	3135899	3135839	←hypo **RNA**→ ←hypo
Nwi-1-59	NC_007406.1	-	2654434	2654374	rve (pfam00665)Transposase_8 (pfam01527)→ **RNA**→ ←tRNA-Glu
Nwi-1-60	NC_007406.1	-	3156400	3156340	←rve (pfam00665) **RNA**→ ←rve (pfam00665)
Nwi-1-61	NC_007406.1	+	2877578	2877638	←SGA1 (COG3387) **RNA**→ ←COG5001 (COG5001)
Nwi-1-62	NC_007406.1	+	1884315	1884375	←DMRL_synthase (pfam00885) **RNA**→ ←RibC (COG0307)
Nwi-1-63	NC_007406.1	+	544919	544979	←Amino_oxidase (pfam01593) **RNA**→ ←Amino_oxidase (pfam01593)
Nwi-1-64	NC_007406.1	+	2073194	2073254	←hypo **RNA**→ ←tRNA-Asp
Nwi-1-65	NC_007406.1	+	1711366	1711426	COG3637 (COG3637)→ **RNA**→ ABC_UvrA_I (cd03270)ABC_UvrA_II (cd03271)→
Nwi-1-66	NC_007406.1	-	692763	692703	←GloB (COG0491) **RNA**→ hypo→
Nwi-1-67	NC_007406.1	-	2354990	2354930	CaiD (COG1024)→ **RNA**→ MmsB (COG2084)→
Nwi-1-68	NC_007406.1	-	701523	701463	adh_short (pfam00106)→ **RNA**→ ←hypo
Nwi-1-69	NC_007406.1	+	1269897	1269957	←COG3415 (COG3415) **RNA**→ hypo→
Nwi-1-71	NC_007406.1	-	751115	751055	COG2863 (COG2863)→ **RNA**→ ←ABC_TM1139_LivF_branched (cd03224)
Nwi-1-72	NC_007406.1	-	1969870	1969810	PpiB (COG0652)→ **RNA**→ Queuosine_synth (pfam02547)→
env-55	soil_XZS45806_b2	+	234	293	**RNA**→ ←hypo
env-56	soil_Contig32804	+	553	614	←hypo **RNA**→
env-57	soil_XZS17656_b2	-	251	190	Predicted hydrolase of the metallo-beta-lactamase superfamily→ **RNA**→ hypo→
Nha-1-88	NC_007964.1	-	3939360	3939299	←Ribosomal_L19 (pfam01245) **RNA**→ ←tRNA_m1G_MT (pfam01746)
Nha-1-89	NC_007964.1	-	1821162	1821101	←tRNA-synt_2c (pfam01411) **RNA**→ →
Nwi-1-73	NC_007406.1	+	3286251	3286312	←Crp (COG0664) **RNA**→ AsmA (COG2982)→
Nha-1-90	NC_007964.1	+	1884022	1884083	COG1160 (COG1160)→ **RNA**→ ←adh_short (pfam00106)
Nsp-1-57	NZ_AAMY01000002.1	-	347795	347734	MviN (COG0728)→ **RNA**→ TrpS (COG0180)→
Nha-1-91	NC_007964.1	-	1203350	1203289	FocA (COG2116)→ **RNA**→ ←COG1695 (COG1695)
Nwi-1-74	NC_007406.1	-	2697551	2697490	COG3293 (COG3293)→ **RNA**→ ←collagen triple helix repeat protein
Nha-1-92	NC_007964.1	+	3172846	3172907	MGS_CPS_II (cd01424)CarB (COG0458)CPSase_L_D3 (pfam02787)→ **RNA**→ GreA (COG0782)→
env-58	soil_XZS39472_b1	-	1390	1330	**RNA**→
env-59	soil_XZS99445_g1	+	1020	1080	**RNA**→

79.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 79.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Rhizobiales-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

u cd00016 (1) Alkaline phosphatase homologues; alkaline phosphatases are non-specific phosphomonoesterases that catalyze the hydrolysis reaction via a phosphoseryl intermediate to produce inorganic phosphate and the corresponding alcohol, optimally at high pH.

u cd00038 (1) effector domain of the CAP family of transcription factors; members include CAP (or cAMP receptor protein (CRP)), which binds cAMP, FNR (fumarate and nitrate reduction), which uses an iron-sulfur cluster to sense oxygen) and CooA, a heme containing CO sensor.

u cd00051 (1) EF-hand, calcium binding motif; A diverse superfamily of calcium sensors and calcium signal modulators; most examples in this alignment model have 2 active canonical EF

hands.
D cd00079 (1) Helicase superfamily c-terminal domain; associated with DEXDc-, DEAD-, and DEAH-box proteins, yeast initiation factor 4A, Ski2p, and Hepatitis C virus NS3 helicases; this domain is found in a wide variety of helicases and helicase related proteins; may not be an autonomously folding unit, but an integral part of the helicase; 4 helicase superfamilies at present according to the organization of their signature motifs; all helicases share the ability to unwind nucleic acid duplexes with a distinct directional polarity; they utilize the free energy from nucleoside triphosphate hydrolysis to fuel their translocation along DNA, unwinding the duplex in the process
u cd00088 (1) Histidine Phosphotransfer domain, involved in signalling through a two part component systems in which an autophosphorylating histidine protein kinase serves as a phosphoryl donor to a response regulator protein; the response regulator protein is modulated by phosphorylation and dephosphorylation of a conserved aspartic acid residue; two-component proteins are abundant in most eubacteria; In E.
uD cd00130 (4) PAS domain; PAS motifs appear in archaea, eubacteria and eukarya.
uDd cd00156 (8) Signal receiver domain; originally thought to be unique to bacteria (CheY, OmpR, NtrC, and PhoB), now recently identified in eukaroytes ETR1 Arabidopsis thaliana; this domain receives the signal from the sensor partner in a two-component systems; contains a phosphoacceptor site that is phosphorylated by histidine kinase homologs; usually found N-terminal to a DNA binding effector domain; forms homodimers
u d cd00189 (9) Tetratricopeptide repeat domain; typically contains 34 amino acids [WLF]-X(2)-[LIM]-[GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-X(2)-[ASL]-X(4)-[PKE]
u d cd00288 (6) Pyruvate kinase (PK): Large allosteric enzyme that regulates glycolysis through binding of the substrate, phosphoenolpyruvate, and one or more allosteric effectors.
d cd00291 (1) SirA, YedF, and YeeD.
d cd00293 (1) Usp: Universal stress protein family.
u cd00333 (2) Major intrinsic protein (MIP) superfamily.
u d cd00353 (2) Ribosomal protein S15 (prokaryotic)_S13 (eukaryotic) binds the central domain of 16S rRNA and is required for assembly of the small ribosomal subunit and for intersubunit association, thus representing a key element in the assembly of the whole ribosome.
u cd00371 (3) Heavy-metal-associated domain (HMA) is a conserved domain of approximately 30 amino acid residues found in a number of proteins that transport or detoxify heavy metals, for example, the CPx-type heavy metal ATPases and copper chaperones.
u cd00408 (1) Dihydrodipicolinate synthase family.
u cd00412 (1) Inorganic pyrophosphatase.
u cd00431 (1) Cysteine hydrolases; This family contains amidohydrolases, like CSHase (N-carbamoylsarcosine amidohydrolase), involved in creatine metabolism and nicotinamidase, converting nicotinamide to nicotinic acid and ammonia in the pyridine nucleotide cycle.
d cd00454 (2) Truncated hemoglobins (trHbs) are a family of oxygen-binding heme proteins found in cyanobacteria, eubacteria, unicellular eukaryotes, and plants.
d cd00484 (1) Phosphoenolpyruvate carboxykinase (PEPCK), a critical gluconeogenic enzyme, catalyzes the first committed step in the diversion of tricarboxylic acid cycle intermediates toward gluconeogenesis.
D cd00537 (5) Methylenetetrahydrofolate reductase (MTHFR).
u d cd00553 (2) NAD+ synthase is a homodimer, which catalyzes the final step in de novo nicotinamide adenine dinucleotide (NAD+) biosynthesis, an amide transfer from either ammonia or glutamine to nicotinic acid adenine dinucleotide (NaAD).
d cd00560 (1) PanC Pantoate-beta-alanine ligase, also known as pantothenate synthase, catalyzes the formation of pantothenate from pantoate and alanine.
u cd00565 (2) ThiS (ThiaminS) is a sulfur carrier protein involved in thiamin biosynthesis in bacteria.
u cd00670 (1) Gly_His_Pro_Ser_Thr_tRNA synthetase class II core domain.
u cd00713 (1) Glutamine amidotransferases class-II (Gn-AT), glutamate synthase (GltS)-type.
u cd00717 (1) Uroporphyrinogen decarboxylase (URO-D) is a dimeric cytosolic enzyme that decarboxylates the four acetate side chains of uroporphyrinogen III (uro-III) to create coproporphyrinogen III, without requiring any prosthetic groups or cofactors.
D cd00732 (2) CheW, a small regulator protein, unique to the chemotaxis signalling in prokaryotes and archea.
u cd00807 (1) Glutaminyl-tRNA synthetase (GlnRS) and non-descriminating Glutamyl-tRNA synthetase (GluRS) cataytic core domain.
D cd00817 (1) This is the catalytic core domain of valine amino-acyl tRNA synthetases (ValRS)
.
u cd00829 (1) Thiolase domain associated with sterol carrier protein (SCP)-x isoform and related proteins; SCP-2 has multiple roles in intracellular lipid circulation and metabolism.
D cd00834 (1) Beta-ketoacyl-acyl carrier protein (ACP) synthase (KAS), type I and II.
uD cd00950 (2) Dihydrodipicolinate synthase (DHDPS) is a key enzyme in lysine biosynthesis.
u cd00982 (2) gltb_C.
uD cd00987 (2) PDZ domain of tryspin-like serine proteases, such as DegP/HtrA, which are oligomeric proteins involved in heat-shock response, chaperone function, and apoptosis.
u cd01045 (1) Ferritin-like domain found in Archaea and Bacteria (Ferritin_like_AB) is a member of a broad superfamily of ferritin-like diiron-carboxylate proteins.
d cd01085 (1) X-Prolyl Aminopeptidase 2.
u d cd01108 (7) Helix-turn-helix transcription regulator CueR.
d cd01137 (1) Metal binding protein PsaA.
D cd01149 (1) Hemin binding protein HutB.
u cd01151 (1) Glutaryl-CoA dehydrogenase (GCD).
D cd01287 (1) FabA, beta-hydroxydecanoyl-acyl carrier protein (ACP)-dehydratase: Bacterial protein of the type II, fatty acid synthase system that binds ACP and catalyzes both dehydration and isomerization reactions, apparently in the same active site.
d cd01298 (1) TRZ/ATZ family contains enzymes from the atrazine degradation pathway and related hydrolases.
u cd01347 (1) TonB dependent/Ligand-Gated channels are created by a monomeric 22 strand (22,24) anti-parallel beta-barrel.
D cd01362 (1) Class II fumarases.
d cd01401 (1) Nicotinate phosphoribosyltransferase (NAPRTase), related to PncB.
u cd01422 (1) Methylglyoxal synthase catalyzes the enolization of dihydroxyacetone phosphate (DHAP) to produce methylglyoxal.
uDd cd01424 (3) Methylglyoxal synthase-like domain from type II glutamine-dependent carbamoyl phosphate synthetase (CSP).
D cd01516 (1) Bacterial fructose-1,6-bisphosphatase, glpX-encoded.
D cd01527 (1) Member of the Rhodanese Homology Domain superfamily.
D cd01556 (1) EPSP synthase domain.
d cd01572 (1) Quinolinate phosphoribosyl transferase (QAPRTase or QPRTase), also called nicotinate-nucleotide pyrophosphorylase, is involved in the de novo synthesis of NAD in both prokaryotes and eukaryotes.
d cd01597 (1) pCLME: prokaryotic 3-carboxy-cis,cis-muconate cycloisomerase (CMLE)_like

This group contains proteins similar to pCLME, a member of the Lyase class I family.
Dd cd01641 (2) Predominantly bacterial family of Mg++ dependend phosphatases, related to inositol monophosphatases.
u cd01658 (1) Ribosomal protein L30 (known as L7 in eukaryotes) is one of the smallest ribosomal proteins with a molecular mass of about 7kDa.
D cd01829 (1) SGNH_peri2; putative periplasmic member of the SGNH-family of hydrolases, a diverse family of lipases and esterases.
D cd01902 (1) Choloylglycine hydrolase (CGH) is a bile salt-modifying enzyme that hydrolyzes non-peptide carbon-nitrogen bonds in choloylglycine and choloyltaurine, both of which are present in bile.
u cd01913 (1) Protease HslV and the ATPase/chaperone HslU are part of an ATP-dependent proteolytic system that is the prokaryotic homolog of the proteasome.
u d cd01948 (2) EAL domain.
u d cd01949 (2) Diguanylate-cyclase (DGC) or GGDEF domain: Originally named after a conserved residue pattern, and initially described as domain of unknown function 1 (DUF1).
u cd02037 (1) MRP (Multiple Resistance and pH adaptation) is a homologue of the Fer4_NifH superfamily.
u cd02138 (1) Nitroreductase-like family 2.
d cd02198 (2) YjgH belong to a large family of proteins present in bacteria, archaea, and eukaryotes with no definitive function.
u cd02437 (1) CCC1_like_1: This is a protein family closely related to CCC1, a family of proteins involved in iron and manganese transport.
u cd02570 (1) PseudoU_synth_EcTruA: Pseudouridine synthase, Escherichia coli TruA like.
u cd02796 (1) tRNA-binding-domain-containing prokaryotic phenylalanly tRNA synthetase (PheRS) beta chain.
u d cd02809 (3) Glycolate oxidase-like (GOX-like) FMN-binding domain.
D cd02888 (1) RNR, class I_like family.
u cd02966 (1) TlpA-like family; composed of TlpA, ResA, DsbE and similar proteins.
uD cd02968 (3) SCO (an acronym for Synthesis of Cytochrome c Oxidase) family; composed of proteins similar to Sco1, a membrane-anchored protein possessing a soluble domain with a TRX fold.
D cd02970 (1) Peroxiredoxin (PRX)-like 2 family; hypothetical proteins that show sequence similarity to PRXs.
u cd02976 (1) NrdH-redoxin (NrdH) family; NrdH is a small monomeric protein with a conserved redox active CXXC motif within a TRX fold, characterized by a glutaredoxin (GRX)-like sequence and TRX-like activity profile.
u cd03013 (1) Peroxiredoxin (PRX) family, PRX5-like subfamily; members are similar to the human protein, PRX5, a homodimeric TRX peroxidase, widely expressed in tissues and found cellularly in mitochondria, peroxisomes and the cytosol.
d cd03015 (1) Peroxiredoxin (PRX) family, Typical 2-Cys PRX subfamily; PRXs are thiol-specific antioxidant (TSA) proteins, which confer a protective role in cells through its peroxidase activity by reducing hydrogen peroxide, peroxynitrite, and organic hydroperoxides.
uD cd03022 (2) DsbA family, 2-hydroxychromene-2-carboxylate (HCCA) isomerase subfamily; HCCA isomerase is a glutathione (GSH) dependent enzyme involved in the naphthalene catabolic pathway.
u cd03034 (1) Arsenate Reductase (ArsC) family, ArsC subfamily; arsenic reductases similar to that encoded by arsC on the R733 plasmid of Escherichia coli. [As(V)] to arsenite [As(III)]
D cd03035 (1) Arsenate Reductase (ArsC) family, Yffb subfamily; Yffb is an uncharacterized bacterial protein encoded by the yffb gene, related to the thioredoxin-fold arsenic reductases, ArsC. [As(V)] to arsenite [As(III)]
u cd03046 (2) GST_N family, Saccharomyces cerevisiae GTT1-like subfamily; composed of predominantly uncharacterized proteins with similarity to the S.
u cd03047 (1) GST_N family, unknown subfamily 2; composed of uncharacterized bacterial proteins with similarity to GSTs.
u d cd03048 (3) GST_N family, Ure2p-like subfamily; composed of the Saccharomyces cerevisiae Ure2p and related GSTs. [Ure3], an inactive, self-propagating and infectious amyloid. Ure2p displays a GST fold containing an N-terminal TRX-fold domain and a C-terminal alpha helical domain, with an active site located in a cleft between the two domains. The N-terminal TRX-fold domain is sufficient to induce the [Ure3]
d cd03056 (2) GST_N family, unknown subfamily 4; composed of uncharacterized bacterial proteins with similarity to GSTs.
d cd03116 (1) Molybdenum is an essential trace element in the form of molybdenum cofactor (Moco) which is associated with the metabolism of nitrogen, carbon and sulfur by redox active enzymes.
d cd03141 (2) Type 1 glutamine amidotransferase (GATase1)-like domain found in proteins similar to Escherichia coli Hsp31 protein (EcHsp31).
u d cd03178 (3) GST_C family, Ure2p-like subfamily; composed of the Saccharomyces cerevisiae Ure2p and related GSTs. [Ure3], an inactive, self-propagating and infectious amyloid. Ure2p displays a GST fold containing an N-terminal thioredoxin-fold domain and a C-terminal alpha helical domain. The N-terminal thioredoxin-fold domain is sufficient to induce the [Ure3]
d cd03179 (2) GST_C family, unknown subfamily 1; composed of uncharacterized bacterial proteins, with similarity to GSTs.
u cd03180 (1) GST_C family, unknown subfamily 2; composed of uncharacterized bacterial proteins, with similarity to GSTs.
u cd03189 (2) GST_C family, Saccharomyces cerevisiae GTT1-like subfamily; composed of predominantly uncharacterized proteins with similarity to the S.
d cd03224 (1) LivF (TM1139) is part of the LIV-I bacterial ABC-type two-component transport system that imports neutral, branched-chain amino acids.
u cd03255 (1) This family is comprised of MJ0796 ATP-binding cassette, macrolide-specific ABC-type efflux carrier (MacAB), and proteins involved in cell division (FtsE) and release of liporoteins from the cytoplasmic membrane (LolCDE).
D cd03270 (1) The excision repair protein UvrA domain I; Nucleotide excision repair in eubacteria is a process that repairs DNA damages by the removal of a 12-13-mer oligonucleotide containing the lesion.
D cd03271 (1) The excision repair protein UvrA domain II; Nucleotide excision repair in eubacteria is a process that repairs DNA damages by the removal of a 12-13-mer oligonucleotide containing the lesion.
d cd03372 (1) Thiamine pyrophosphate (TPP) family, ComE subfamily, TPP-binding module; composed of proteins similar to Methanococcus jannaschii sulfopyruvate decarboxylase beta subunit (ComE).
uD cd03454 (2) YdeM is a Bacillus subtilis protein that belongs to a family of prokaryotic proteins of unkown function.
u cd03470 (1) Iron-sulfur protein (ISP) component of the bc(1) complex family, Rieske domain; The Rieske domain is a [2Fe-2S]
D COG0006 (1) Xaa-Pro aminopeptidase [Amino acid transport and metabolism]

d COG0007 (2) Uroporphyrinogen-III methylase [Coenzyme metabolism]
d COG0012 (1) Predicted GTPase, probable translation factor [Translation, ribosomal structure and biogenesis]
uD COG0019 (2) Diaminopimelate decarboxylase [Amino acid transport and metabolism]
u COG0031 (1) Cysteine synthase [Amino acid transport and metabolism]
u COG0034 (1) Glutamine phosphoribosylpyrophosphate amidotransferase [Nucleotide transport and metabolism]
D COG0036 (5) Pentose-5-phosphate-3-epimerase [Carbohydrate transport and metabolism]
D COG0037 (5) Predicted ATPase of the PP-loop superfamily implicated in cell cycle control [Cell division and chromosome partitioning]
d COG0042 (1) tRNA-dihydrouridine synthase [Translation, ribosomal structure and biogenesis]
d COG0046 (1) Phosphoribosylformylglycinamidine (FGAM) synthase, synthetase domain [Nucleotide transport and metabolism]
uD COG0053 (3) Predicted Co/Zn/Cd cation transporters [Inorganic ion transport and metabolism]
D COG0055 (1) F0F1-type ATP synthase, beta subunit [Energy production and conversion]
d COG0059 (1) Ketol-acid reductoisomerase [Amino acid transport and metabolism / Coenzyme metabolism]
u COG0069 (2) Glutamate synthase domain 2 [Amino acid transport and metabolism]
u COG0072 (1) Phenylalanyl-tRNA synthetase beta subunit [Translation, ribosomal structure and biogenesis]
u COG0077 (1) Prephenate dehydratase [Amino acid transport and metabolism]
u COG0092 (2) Ribosomal protein S3 [Translation, ribosomal structure and biogenesis]
d COG0101 (4) Pseudouridylate synthase [Translation, ribosomal structure and biogenesis]
u COG0111 (1) Phosphoglycerate dehydrogenase and related dehydrogenases [Amino acid transport and metabolism]
d COG0117 (1) Pyrimidine deaminase [Coenzyme metabolism]
D COG0133 (1) Tryptophan synthase beta chain [Amino acid transport and metabolism]
u COG0135 (1) Phosphoribosylanthranilate isomerase [Amino acid transport and metabolism]
uD COG0136 (2) Aspartate-semialdehyde dehydrogenase [Amino acid transport and metabolism]
d COG0140 (1) Phosphoribosyl-ATP pyrophosphohydrolase [Amino acid transport and metabolism]
D COG0143 (1) Methionyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
d COG0147 (3) Anthranilate/para-aminobenzoate synthases component I [Amino acid transport and metabolism / Coenzyme metabolism]
u COG0148 (2) Enolase [Carbohydrate transport and metabolism]
d COG0149 (1) Triosephosphate isomerase [Carbohydrate transport and metabolism]
u COG0155 (1) Sulfite reductase, beta subunit (hemoprotein) [Inorganic ion transport and metabolism]
u COG0156 (1) 7-keto-8-aminopelargonate synthetase and related enzymes [Coenzyme metabolism]
d COG0158 (1) Fructose-1,6-bisphosphatase [Carbohydrate transport and metabolism]
u COG0159 (2) Tryptophan synthase alpha chain [Amino acid transport and metabolism]
u COG0161 (1) Adenosylmethionine-8-amino-7-oxononanoate aminotransferase [Coenzyme metabolism]
u COG0162 (1) Tyrosyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
d COG0164 (1) Ribonuclease HII [DNA replication, recombination, and repair]
u COG0165 (1) Argininosuccinate lyase [Amino acid transport and metabolism]
d COG0167 (2) Dihydroorotate dehydrogenase [Nucleotide transport and metabolism]
d COG0169 (1) Shikimate 5-dehydrogenase [Amino acid transport and metabolism]
Dd COG0180 (2) Tryptophanyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
d COG0187 (1) Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), B subunit [DNA replication, recombination, and repair]
D COG0189 (1) Glutathione synthase/Ribosomal protein S6 modification enzyme (glutaminyl transferase) [Coenzyme metabolism / Translation, ribosomal structure and biogenesis]
u COG0191 (1) Fructose/tagatose bisphosphate aldolase [Carbohydrate transport and metabolism]
u COG0192 (1) S-adenosylmethionine synthetase [Coenzyme metabolism]
u COG0203 (1) Ribosomal protein L17 [Translation, ribosomal structure and biogenesis]
u COG0210 (1) Superfamily I DNA and RNA helicases [DNA replication, recombination, and repair]
d COG0218 (1) Predicted GTPase [General function prediction only]
d COG0223 (2) Methionyl-tRNA formyltransferase [Translation, ribosomal structure and biogenesis]
d COG0235 (1) Ribulose-5-phosphate 4-epimerase and related epimerases and aldolases [Carbohydrate transport and metabolism]
u COG0236 (1) Acyl carrier protein [Lipid metabolism / Secondary metabolites biosynthesis, transport, and catabolism]
u COG0239 (4) Integral membrane protein possibly involved in chromosome condensation [Cell division and chromosome partitioning]
d COG0247 (4) Fe-S oxidoreductase [Energy production and conversion]
u COG0248 (1) Exopolyphosphatase [Nucleotide transport and metabolism / Inorganic ion transport and metabolism]
d COG0251 (1) Putative translation initiation inhibitor, yjgF family [Translation, ribosomal structure and biogenesis]
u COG0254 (5) Ribosomal protein L31 [Translation, ribosomal structure and biogenesis]
uD COG0265 (3) Trypsin-like serine proteases, typically periplasmic, contain C-terminal PDZ domain [Posttranslational modification, protein turnover, chaperones]
d COG0267 (1) Ribosomal protein L33 [Translation, ribosomal structure and biogenesis]
u COG0272 (1) NAD-dependent DNA ligase (contains BRCT domain type II) [DNA replication, recombination, and repair]
d COG0277 (1) FAD/FMN-containing dehydrogenases [Energy production and conversion]
d COG0284 (1) Orotidine-5'-phosphate decarboxylase [Nucleotide transport and metabolism]
u COG0285 (1) Folylpolyglutamate synthase [Coenzyme metabolism]
d COG0290 (2) Translation initiation factor 3 (IF-3) [Translation, ribosomal structure and biogenesis]
D COG0295 (1) Cytidine deaminase [Nucleotide transport and metabolism]
d COG0300 (2) Short-chain dehydrogenases of various substrate specificities [General function prediction only]
u COG0303 (1) Molybdopterin biosynthesis enzyme [Coenzyme metabolism]
u COG0304 (1) 3-oxoacyl-(acyl-carrier-protein) synthase [Lipid metabolism / Secondary metabolites biosynthesis, transport, and catabolism]
u d COG0307 (2) Riboflavin synthase alpha chain [Coenzyme metabolism]
D COG0312 (1) Predicted Zn-dependent proteases and their inactivated homologs [General function prediction only]
uDd COG0318 (5) Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II [Lipid metabolism

/ Secondary metabolites biosynthesis, transport, and catabolism]
d COG0331 (1) (acyl-carrier-protein) S-malonyltransferase [Lipid metabolism]
uDd COG0345 (3) Pyrroline-5-carboxylate reductase [Amino acid transport and metabolism]
u COG0347 (5) Nitrogen regulatory protein PII [Amino acid transport and metabolism]
d COG0349 (1) Ribonuclease D [Translation, ribosomal structure and biogenesis]
D COG0351 (1) Hydroxymethylpyrimidine/phosphomethylpyrimidine kinase [Coenzyme metabolism]
D COG0363 (1) 6-phosphogluconolactonase/Glucosamine-6-phosphate isomerase/deaminase [Carbohydrate transport and metabolism]
Dd COG0365 (2) Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases [Lipid metabolism]
u COG0384 (1) Predicted epimerase, PhzC/PhzF homolog [General function prediction only]
u d COG0388 (4) Predicted amidohydrolase [General function prediction only]
u COG0412 (1) Dienelactone hydrolase and related enzymes [Secondary metabolites biosynthesis, transport, and catabolism]
Dd COG0427 (2) Acetyl-CoA hydrolase [Energy production and conversion]
u COG0433 (6) Predicted ATPase [General function prediction only]
uD COG0436 (2) Aspartate/tyrosine/aromatic aminotransferase [Amino acid transport and metabolism]
D COG0438 (1) Glycosyltransferase [Cell envelope biogenesis, outer membrane]
D COG0442 (1) Prolyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
u COG0451 (2) Nucleoside-diphosphate-sugar epimerases [Cell envelope biogenesis, outer membrane / Carbohydrate transport and metabolism]
uDd COG0458 (3) Carbamoylphosphate synthase large subunit (split gene in MJ) [Amino acid transport and metabolism / Nucleotide transport and metabolism]
u COG0460 (1) Homoserine dehydrogenase [Amino acid transport and metabolism]
u COG0462 (1) Phosphoribosylpyrophosphate synthetase [Nucleotide transport and metabolism / Amino acid transport and metabolism]
u COG0470 (1) ATPase involved in DNA replication [DNA replication, recombination, and repair]
d COG0473 (1) Isocitrate/isopropylmalate dehydrogenase [Amino acid transport and metabolism]
u COG0475 (1) Kef-type K+ transport systems, membrane components [Inorganic ion transport and metabolism]
d COG0479 (1) Succinate dehydrogenase/fumarate reductase, Fe-S protein subunit [Energy production and conversion]
D COG0480 (1) Translation elongation factors (GTPases) [Translation, ribosomal structure and biogenesis]
u COG0481 (1) Membrane GTPase LepA [Cell envelope biogenesis, outer membrane]
u COG0484 (1) DnaJ-class molecular chaperone with C-terminal Zn finger domain [Posttranslational modification, protein turnover, chaperones]
u COG0491 (2) Zn-dependent hydrolases, including glyoxylases [General function prediction only]
uD COG0492 (2) Thioredoxin reductase [Posttranslational modification, protein turnover, chaperones]
D COG0493 (2) NADPH-dependent glutamate synthase beta chain and related oxidoreductases [Amino acid transport and metabolism / General function prediction only]
d COG0497 (1) ATPase involved in DNA repair [DNA replication, recombination, and repair]
uD COG0505 (2) Carbamoylphosphate synthase small subunit [Amino acid transport and metabolism / Nucleotide transport and metabolism]
u COG0510 (1) Predicted choline kinase involved in LPS biosynthesis [Cell envelope biogenesis, outer membrane]
d COG0512 (3) Anthranilate/para-aminobenzoate synthases component II [Amino acid transport and metabolism / Coenzyme metabolism]
u COG0517 (1) FOG: CBS domain [General function prediction only]
u d COG0522 (6) Ribosomal protein S4 and related proteins [Translation, ribosomal structure and biogenesis]
u COG0523 (2) Putative GTPases (G3E family) [General function prediction only]
d COG0539 (5) Ribosomal protein S1 [Translation, ribosomal structure and biogenesis]
u COG0542 (1) ATPases with chaperone activity, ATP-binding subunit [Posttranslational modification, protein turnover, chaperones]
uD COG0548 (2) Acetylglutamate kinase [Amino acid transport and metabolism]
d COG0549 (1) Carbamate kinase [Amino acid transport and metabolism]
d COG0552 (1) Signal recognition particle GTPase [Intracellular trafficking and secretion]
uD COG0559 (3) Branched-chain amino acid ABC-type transport system, permease components [Amino acid transport and metabolism]
u COG0564 (1) Pseudouridylate synthases, 23S RNA-specific [Translation, ribosomal structure and biogenesis]
u COG0566 (1) rRNA methylases [Translation, ribosomal structure and biogenesis]
uD COG0568 (2) DNA-directed RNA polymerase, sigma subunit (sigma70/sigma32) [Transcription]
u COG0574 (1) Phosphoenolpyruvate synthase/pyruvate phosphate dikinase [Carbohydrate transport and metabolism]
u d COG0583 (6) Transcriptional regulator [Transcription]
d COG0584 (1) Glycerophosphoryl diester phosphodiesterase [Energy production and conversion]
u COG0586 (1) Uncharacterized membrane-associated protein [Function unknown]
uDd COG0587 (3) DNA polymerase III, alpha subunit [DNA replication, recombination, and repair]
u COG0596 (1) Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily) [General function prediction only]
D COG0601 (1) ABC-type dipeptide/oligopeptide/nickel transport systems, permease components [Amino acid transport and metabolism / Inorganic ion transport and metabolism]
u d COG0602 (2) Organic radical activating enzymes [Posttranslational modification, protein turnover, chaperones]
u COG0604 (2) NADPH:quinone reductase and related Zn-dependent oxidoreductases [Energy production and conversion / General function prediction only]
D COG0624 (1) Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase and related deacylases [Amino acid transport and metabolism]
u COG0625 (1) Glutathione S-transferase [Posttranslational modification, protein turnover, chaperones]
uD COG0626 (2) Cystathionine beta-lyases/cystathionine gamma-synthases [Amino acid transport and metabolism]
d COG0628 (1) Predicted permease [General function prediction only]
u COG0635 (2) Coproporphyrinogen III oxidase and related Fe-S oxidoreductases [Coenzyme metabolism]
uDd COG0642 (14) Signal transduction histidine kinase [Signal transduction mechanisms]
u COG0643 (2) Chemotaxis protein histidine kinase and related kinases [Cell motility and

secretion / Signal transduction mechanisms]
uD COG0644 (2) Dehydrogenases (flavoproteins) [Energy production and conversion]
u COG0652 (1) Peptidyl-prolyl cis-trans isomerase (rotamase) - cyclophilin family [Posttranslational modification, protein turnover, chaperones]
u COG0654 (2) 2-polyprenyl-6-methoxyphenol hydroxylase and related FAD-dependent oxidoreductases [Coenzyme metabolism / Energy production and conversion]
u COG0656 (6) Aldo/keto reductases, related to diketogulonate reductase [General function prediction only]
u COG0657 (1) Esterase/lipase [Lipid metabolism]
D COG0662 (1) Mannose-6-phosphate isomerase [Carbohydrate transport and metabolism]
uD COG0664 (3) cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases [Signal transduction mechanisms]
uDd COG0665 (3) Glycine/D-amino acid oxidases (deaminating) [Amino acid transport and metabolism]
u COG0669 (2) Phosphopantetheine adenylyltransferase [Coenzyme metabolism]
uDd COG0673 (9) Predicted dehydrogenases and related proteins [General function prediction only]
d COG0679 (1) Predicted permeases [General function prediction only]
u d COG0683 (5) ABC-type branched-chain amino acid transport systems, periplasmic component [Amino acid transport and metabolism]
D COG0687 (1) Spermidine/putrescine-binding periplasmic protein [Amino acid transport and metabolism]
u d COG0708 (2) Exonuclease III [DNA replication, recombination, and repair]
u COG0713 (1) NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) [Energy production and conversion]
u COG0714 (1) MoxR-like ATPases [General function prediction only]
u COG0728 (1) Uncharacterized membrane protein, putative virulence factor [General function prediction only]
u COG0735 (6) Fe2+/Zn2+ uptake regulation proteins [Inorganic ion transport and metabolism]
uD COG0739 (4) Membrane proteins related to metalloendopeptidases [Cell envelope biogenesis, outer membrane]
D COG0740 (1) Protease subunit of ATP-dependent Clp proteases [Posttranslational modification, protein turnover, chaperones / Intracellular trafficking and secretion]
u d COG0741 (7) Soluble lytic murein transglycosylase and related regulatory proteins (some contain LysM/invasin domains) [Cell envelope biogenesis, outer membrane]
u d COG0744 (3) Membrane carboxypeptidase (penicillin-binding protein) [Cell envelope biogenesis, outer membrane]
uD COG0745 (2) Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain [Signal transduction mechanisms / Transcription]
d COG0758 (1) Predicted Rossmann fold nucleotide-binding protein involved in DNA uptake [DNA replication, recombination, and repair / Intracellular trafficking and secretion]
D COG0764 (1) 3-hydroxymyristoyl/3-hydroxydecanoyl-(acyl carrier protein) dehydratases [Lipid metabolism]
u COG0765 (1) ABC-type amino acid transport system, permease component [Amino acid transport and metabolism]
D COG0777 (3) Acetyl-CoA carboxylase beta subunit [Lipid metabolism]
u COG0780 (1) Enzyme related to GTP cyclohydrolase I [General function prediction only]
uD COG0782 (3) Transcription elongation factor [Transcription]
u d COG0790 (5) FOG: TPR repeat, SEL1 subfamily [General function prediction only]
u COG0801 (1) 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase [Coenzyme metabolism]
uD COG0807 (5) GTP cyclohydrolase II [Coenzyme metabolism]
u COG0812 (1) UDP-N-acetylmuramate dehydrogenase [Cell envelope biogenesis, outer membrane]
u COG0820 (1) Predicted Fe-S-cluster redox enzyme [General function prediction only]
u COG0825 (1) Acetyl-CoA carboxylase alpha subunit [Lipid metabolism]
D COG0836 (1) Mannose-1-phosphate guanylyltransferase [Cell envelope biogenesis, outer membrane]
uD COG0837 (2) Glucokinase [Carbohydrate transport and metabolism]
D COG0840 (1) Methyl-accepting chemotaxis protein [Cell motility and secretion / Signal transduction mechanisms]
d COG0841 (1) Cation/multidrug efflux pump [Defense mechanisms]
uDd COG0845 (4) Membrane-fusion protein [Cell envelope biogenesis, outer membrane]
u COG0847 (2) DNA polymerase III, epsilon subunit and related 3'-5' exonucleases [DNA replication, recombination, and repair]
d COG0861 (1) Membrane protein TerC, possibly involved in tellurium resistance [Inorganic ion transport and metabolism]
d COG1002 (1) Type II restriction enzyme, methylase subunits [Defense mechanisms]
u COG1004 (2) Predicted UDP-glucose 6-dehydrogenase [Cell envelope biogenesis, outer membrane]
u COG1007 (1) NADH:ubiquinone oxidoreductase subunit 2 (chain N) [Energy production and conversion]
D COG1009 (1) NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na+/H+ antiporter, MnhA subunit [Energy production and conversion / Inorganic ion transport and metabolism]
d COG1011 (1) Predicted hydrolase (HAD superfamily) [General function prediction only]
Dd COG1012 (2) NAD-dependent aldehyde dehydrogenases [Energy production and conversion]
uDd COG1024 (6) Enoyl-CoA hydratase/carnithine racemase [Lipid metabolism]
d COG1032 (1) Fe-S oxidoreductase [Energy production and conversion]
u COG1042 (1) Acyl-CoA synthetase (NDP forming) [Energy production and conversion]
u d COG1051 (3) ADP-ribose pyrophosphatase [Nucleotide transport and metabolism]
u COG1052 (1) Lactate dehydrogenase and related dehydrogenases [Energy production and conversion / Coenzyme metabolism / General function prediction only]
d COG1057 (1) Nicotinic acid mononucleotide adenylyltransferase [Coenzyme metabolism]
uD COG1064 (6) Zn-dependent alcohol dehydrogenases [General function prediction only]
u COG1066 (2) Predicted ATP-dependent serine protease [Posttranslational modification, protein turnover, chaperones]
u COG1072 (1) Panthothenate kinase [Coenzyme metabolism]
u COG1079 (1) Uncharacterized ABC-type transport system, permease component [General function prediction only]
u COG1082 (1) Sugar phosphate isomerases/epimerases [Carbohydrate transport and metabolism]
u COG1087 (1) UDP-glucose 4-epimerase [Cell envelope biogenesis, outer membrane]
u COG1116 (2) ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component [Inorganic ion transport and metabolism]
Dd COG1126 (2) ABC-type polar amino acid transport system, ATPase component [Amino acid transport and metabolism]
u COG1129 (4) ABC-type sugar transport system, ATPase component [Carbohydrate transport and metabolism]

Dd COG1132 (3) ABC-type multidrug transport system, ATPase and permease components [Defense mechanisms]
u COG1160 (1) Predicted GTPases [General function prediction only]
u d COG1172 (2) Ribose/xylose/arabinose/galactoside ABC-type transport systems, permease components [Carbohydrate transport and metabolism]
u COG1174 (1) ABC-type proline/glycine betaine transport systems, permease component [Amino acid transport and metabolism]
D COG1176 (1) ABC-type spermidine/putrescine transport system, permease component I [Amino acid transport and metabolism]
d COG1183 (2) Phosphatidylserine synthase [Lipid metabolism]
uDd COG1185 (4) Polyribonucleotide nucleotidyltransferase (polynucleotide phosphorylase) [Translation, ribosomal structure and biogenesis]
u d COG1186 (4) Protein chain release factor B [Translation, ribosomal structure and biogenesis]
u d COG1192 (2) ATPases involved in chromosome partitioning [Cell division and chromosome partitioning]
u COG1194 (1) A/G-specific DNA glycosylase [DNA replication, recombination, and repair]
d COG1196 (1) Chromosome segregation ATPases [Cell division and chromosome partitioning]
D COG1197 (1) Transcription-repair coupling factor (superfamily II helicase) [DNA replication, recombination, and repair / Transcription]
d COG1198 (1) Primosomal protein N' (replication factor Y) - superfamily II helicase [DNA replication, recombination, and repair]
uD COG1201 (9) Lhr-like helicases [General function prediction only]
u COG1208 (1) Nucleoside-diphosphate-sugar pyrophosphorylase involved in lipopolysaccharide biosynthesis/translation initiation factor 2B, gamma/epsilon subunits (eIF-2Bgamma/eIF-2Bepsilon) [Cell envelope biogenesis, outer membrane / Translation, ribosomal structure and biogenesis]
d COG1215 (1) Glycosyltransferases, probably involved in cell wall biogenesis [Cell envelope biogenesis, outer membrane]
D COG1217 (1) Predicted membrane GTPase involved in stress response [Signal transduction mechanisms]
u COG1220 (1) ATP-dependent protease HslVU (ClpYQ), ATPase subunit [Posttranslational modification, protein turnover, chaperones]
u COG1235 (2) Metal-dependent hydrolases of the beta-lactamase superfamily I [General function prediction only]
d COG1238 (1) Predicted membrane protein [Function unknown]
u d COG1249 (9) Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component, and related enzymes [Energy production and conversion]
u d COG1250 (2) 3-hydroxyacyl-CoA dehydrogenase [Lipid metabolism]
u COG1252 (1) NADH dehydrogenase, FAD-containing subunit [Energy production and conversion]
uD COG1272 (7) Predicted membrane protein, hemolysin III homolog [General function prediction only]
u d COG1280 (2) Putative threonine efflux protein [Amino acid transport and metabolism]
d COG1284 (2) Uncharacterized conserved protein [Function unknown]
u COG1298 (1) Flagellar biosynthesis pathway, component FlhA [Cell motility and secretion / Intracellular trafficking and secretion]
d COG1312 (5) D-mannonate dehydratase [Carbohydrate transport and metabolism]
u COG1368 (1) Phosphoglycerol transferase and related proteins, alkaline phosphatase superfamily [Cell envelope biogenesis, outer membrane]
d COG1376 (5) Uncharacterized protein conserved in bacteria [Function unknown]
u COG1377 (1) Flagellar biosynthesis pathway, component FlhB [Cell motility and secretion / Intracellular trafficking and secretion]
uD COG1384 (3) Lysyl-tRNA synthetase (class I) [Translation, ribosomal structure and biogenesis]
D COG1391 (4) Glutamine synthetase adenylyltransferase [Posttranslational modification, protein turnover, chaperones / Signal transduction mechanisms]
Dd COG1407 (6) Predicted ICC-like phosphoesterases [General function prediction only]
D COG1409 (1) Predicted phosphohydrolases [General function prediction only]
u COG1420 (1) Transcriptional regulator of heat shock gene [Transcription]
u COG1434 (1) Uncharacterized conserved protein [Function unknown]
u d COG1451 (2) Predicted metal-dependent hydrolase [General function prediction only]
u COG1463 (1) ABC-type transport system involved in resistance to organic solvents, periplasmic component [Secondary metabolites biosynthesis, transport, and catabolism]
d COG1466 (1) DNA polymerase III, delta subunit [DNA replication, recombination, and repair]
uD COG1475 (3) Predicted transcriptional regulators [Transcription]
D COG1492 (1) Cobyric acid synthase [Coenzyme metabolism]
u COG1522 (1) Transcriptional regulators [Transcription]
u COG1523 (1) Type II secretory pathway, pullulanase PulA and related glycosidases [Carbohydrate transport and metabolism]
uD COG1529 (2) Aerobic-type carbon monoxide dehydrogenase, large subunit CoxL/CutL homologs [Energy production and conversion]
u COG1530 (1) Ribonucleases G and E [Translation, ribosomal structure and biogenesis]
d COG1538 (1) Outer membrane protein [Cell envelope biogenesis, outer membrane / Intracellular trafficking and secretion]
u COG1560 (5) Lauroyl/myristoyl acyltransferase [Cell envelope biogenesis, outer membrane]
D COG1561 (1) Uncharacterized stress-induced protein [Function unknown]
u COG1562 (1) Phytoene/squalene synthetase [Lipid metabolism]
d COG1566 (1) Multidrug resistance efflux pump [Defense mechanisms]
u COG1573 (1) Uracil-DNA glycosylase [DNA replication, recombination, and repair]
u COG1593 (1) TRAP-type C4-dicarboxylate transport system, large permease component [Carbohydrate transport and metabolism]
u COG1595 (1) DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog [Transcription]
d COG1609 (1) Transcriptional regulators [Transcription]
u COG1610 (1) Uncharacterized conserved protein [Function unknown]
D COG1612 (1) Uncharacterized protein required for cytochrome oxidase assembly [Posttranslational modification, protein turnover, chaperones]
d COG1620 (4) L-lactate permease [Energy production and conversion]
D COG1638 (1) TRAP-type C4-dicarboxylate transport system, periplasmic component [Carbohydrate transport and metabolism]
d COG1648 (2) Siroheme synthase (precorrin-2 oxidase/ferrochelatase domain) [Coenzyme metabolism]
D COG1670 (1) Acetyltransferases, including N-acetylases of ribosomal proteins [Translation, ribosomal structure and biogenesis]
Dd COG1684 (2) Flagellar biosynthesis pathway, component FliR [Cell motility and secretion /

Intracellular trafficking and secretion]
d COG1695 (1) Predicted transcriptional regulators [Transcription]
D COG1729 (1) Uncharacterized protein conserved in bacteria [Function unknown]
u COG1732 (1) Periplasmic glycine betaine/choline-binding (lipo)protein of an ABC-type transport system (osmoprotectant binding protein) [Cell envelope biogenesis, outer membrane]
u COG1739 (1) Uncharacterized conserved protein [Function unknown]
d COG1741 (1) Pirin-related protein [General function prediction only]
u COG1762 (1) Phosphotransferase system mannitol/fructose-specific IIA domain (Ntr-type) [Carbohydrate transport and metabolism / Signal transduction mechanisms]
u COG1764 (1) Predicted redox protein, regulator of disulfide bond formation [Posttranslational modification, protein turnover, chaperones]
u COG1766 (1) Flagellar biosynthesis/type III secretory pathway lipoprotein [Cell motility and secretion / Intracellular trafficking and secretion]
Dd COG1770 (2) Protease II [Amino acid transport and metabolism]
u COG1793 (1) ATP-dependent DNA ligase [DNA replication, recombination, and repair]
d COG1807 (1) 4-amino-4-deoxy-L-arabinose transferase and related glycosyltransferases of PMT family [Cell envelope biogenesis, outer membrane]
D COG1825 (6) Ribosomal protein L25 (general stress protein Ctc) [Translation, ribosomal structure and biogenesis]
u COG1835 (2) Predicted acyltransferases [Lipid metabolism]
d COG1858 (1) Cytochrome c peroxidase [Inorganic ion transport and metabolism]
D COG1862 (1) Preprotein translocase subunit YajC [Intracellular trafficking and secretion]
u COG1866 (1) Phosphoenolpyruvate carboxykinase (ATP) [Energy production and conversion]
D COG1879 (1) ABC-type sugar transport system, periplasmic component [Carbohydrate transport and metabolism]
u COG1910 (1) Periplasmic molybdate-binding protein/domain [Inorganic ion transport and metabolism]
u COG1914 (1) Mn2+ and Fe2+ transporters of the NRAMP family [Inorganic ion transport and metabolism]
u d COG1917 (2) Uncharacterized conserved protein, contains double-stranded beta-helix domain [Function unknown]
d COG1923 (1) Uncharacterized host factor I protein [General function prediction only]
d COG1925 (1) Phosphotransferase system, HPr-related proteins [Carbohydrate transport and metabolism]
d COG1932 (1) Phosphoserine aminotransferase [Coenzyme metabolism / Amino acid transport and metabolism]
D COG1934 (5) Uncharacterized protein conserved in bacteria [Function unknown]
Dd COG1940 (2) Transcriptional regulator/sugar kinase [Transcription / Carbohydrate transport and metabolism]
d COG1952 (2) Preprotein translocase subunit SecB [Intracellular trafficking and secretion]
d COG1959 (2) Predicted transcriptional regulator [Transcription]
u COG1975 (5) Xanthine and CO dehydrogenases maturation factor, XdhC/CoxF family [Posttranslational modification, protein turnover, chaperones]
d COG1985 (1) Pyrimidine reductase, riboflavin biosynthesis [Coenzyme metabolism]
u COG1987 (1) Flagellar biosynthesis pathway, component FliQ [Cell motility and secretion / Intracellular trafficking and secretion]
u COG2005 (1) N-terminal domain of molybdenum-binding protein [General function prediction only]
u COG2010 (1) Cytochrome c, mono- and diheme variants [Energy production and conversion]
u COG2020 (2) Putative protein-S-isoprenylcysteine methyltransferase [Posttranslational modification, protein turnover, chaperones]
D COG2022 (2) Uncharacterized enzyme of thiazole biosynthesis [Nucleotide transport and metabolism]
D COG2025 (2) Electron transfer flavoprotein, alpha subunit [Energy production and conversion]
d COG2031 (1) Short chain fatty acids transporter [Lipid metabolism]
D COG2050 (1) Uncharacterized protein, possibly involved in aromatic compounds catabolism [Secondary metabolites biosynthesis, transport, and catabolism]
D COG2062 (1) Phosphohistidine phosphatase SixA [Signal transduction mechanisms]
u COG2070 (1) Dioxygenases related to 2-nitropropane dioxygenase [General function prediction only]
d COG2071 (1) Predicted glutamine amidotransferases [General function prediction only]
u COG2072 (1) Predicted flavoprotein involved in K+ transport [Inorganic ion transport and metabolism]
u d COG2076 (2) Membrane transporters of cations and cationic drugs [Inorganic ion transport and metabolism]
d COG2080 (1) Aerobic-type carbon monoxide dehydrogenase, small subunit CoxS/CutS homologs [Energy production and conversion]
u COG2081 (1) Predicted flavoproteins [General function prediction only]
D COG2084 (1) 3-hydroxyisobutyrate dehydrogenase and related beta-hydroxyacid dehydrogenases [Lipid metabolism]
u COG2086 (2) Electron transfer flavoprotein, beta subunit [Energy production and conversion]
u COG2096 (1) Uncharacterized conserved protein [Function unknown]
u COG2105 (1) Uncharacterized conserved protein [Function unknown]
u COG2113 (1) ABC-type proline/glycine betaine transport systems, periplasmic components [Amino acid transport and metabolism]
d COG2114 (1) Adenylate cyclase, family 3 (some proteins contain HAMP domain) [Signal transduction mechanisms]
u COG2116 (1) Formate/nitrite family of transporters [Inorganic ion transport and metabolism]
u COG2124 (1) Cytochrome P450 [Secondary metabolites biosynthesis, transport, and catabolism]
u d COG2128 (2) Uncharacterized conserved protein [Function unknown]
u COG2132 (1) Putative multicopper oxidases [Secondary metabolites biosynthesis, transport, and catabolism]
u COG2133 (1) Glucose/sorbosone dehydrogenases [Carbohydrate transport and metabolism]
d COG2141 (1) Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases [Energy production and conversion]
d COG2146 (1) Ferredoxin subunits of nitrite reductase and ring-hydroxylating dioxygenases [Inorganic ion transport and metabolism / General function prediction only]
u d COG2166 (2) SufE protein probably involved in Fe-S center assembly [General function prediction only]
u COG2171 (2) Tetrahydrodipicolinate N-succinyltransferase [Amino acid transport and metabolism]
d COG2177 (1) Cell division protein [Cell division and chromosome partitioning]
u COG2192 (1) Predicted carbamoyl transferase, NodU family [Posttranslational modification, protein turnover, chaperones]
d COG2195 (2) Di- and tripeptidases [Amino acid transport and metabolism]
u COG2202 (5) FOG: PAS/PAC domain [Signal transduction mechanisms]

u COG2217 (3) Cation transport ATPase [Inorganic ion transport and metabolism]
D COG2220 (1) Predicted Zn-dependent hydrolases of the beta-lactamase fold [General function prediction only]
u COG2226 (6) Methylase involved in ubiquinone/menaquinone biosynthesis [Coenzyme metabolism]
u COG2230 (1) Cyclopropane fatty acid synthase and related methyltransferases [Cell envelope biogenesis, outer membrane]
u COG2239 (1) Mg/Co/Ni transporter MgtE (contains CBS domain) [Inorganic ion transport and metabolism]
u COG2255 (1) Holliday junction resolvasome, helicase subunit [DNA replication, recombination, and repair]
D COG2259 (1) Predicted membrane protein [Function unknown]
u COG2262 (2) GTPases [General function prediction only]
u COG2267 (1) Lysophospholipase [Lipid metabolism]
u d COG2271 (3) Sugar phosphate permease [Carbohydrate transport and metabolism]
u COG2303 (3) Choline dehydrogenase and related flavoproteins [Amino acid transport and metabolism]
u COG2319 (1) FOG: WD40 repeat [General function prediction only]
u COG2329 (2) Uncharacterized enzyme involved in biosynthesis of extracellular polysaccharides [General function prediction only]
u COG2339 (1) Predicted membrane protein [Function unknown]
D COG2355 (1) Zn-dependent dipeptidase, microsomal dipeptidase homolog [Amino acid transport and metabolism]
u COG2363 (1) Uncharacterized small membrane protein [Function unknown]
D COG2367 (1) Beta-lactamase class A [Defense mechanisms]
u COG2423 (1) Predicted ornithine cyclodeaminase, mu-crystallin homolog [Amino acid transport and metabolism]
u COG2513 (1) PEP phosphonomutase and related enzymes [Carbohydrate transport and metabolism]
u COG2515 (1) 1-aminocyclopropane-1-carboxylate deaminase [Amino acid transport and metabolism]
u COG2518 (3) Protein-L-isoaspartate carboxylmethyltransferase [Posttranslational modification, protein turnover, chaperones]
u COG2606 (1) Uncharacterized conserved protein [Function unknown]
u COG2607 (1) Predicted ATPase (AAA+ superfamily) [General function prediction only]
D COG2740 (1) Predicted nucleic-acid-binding protein implicated in transcription termination [Transcription]
u d COG2771 (2) DNA-binding HTH domain-containing proteins [Transcription]
D COG2802 (1) Uncharacterized protein, similar to the N-terminal domain of Lon protease [General function prediction only]
u COG2813 (2) 16S RNA G1207 methylase RsmC [Translation, ribosomal structure and biogenesis]
uDd COG2814 (5) Arabinose efflux permease [Carbohydrate transport and metabolism]
u COG2822 (1) Predicted periplasmic lipoprotein involved in iron transport [Inorganic ion transport and metabolism]
D COG2834 (5) Outer membrane lipoprotein-sorting protein [Cell envelope biogenesis, outer membrane]
u d COG2840 (2) Uncharacterized protein conserved in bacteria [Function unknown]
d COG2844 (1) UTP:GlnB (protein PII) uridylyltransferase [Posttranslational modification, protein turnover, chaperones]
u COG2852 (1) Uncharacterized protein conserved in bacteria [Function unknown]
u COG2860 (1) Predicted membrane protein [Function unknown]
u COG2863 (2) Cytochrome c553 [Energy production and conversion]
D COG2870 (1) ADP-heptose synthase, bifunctional sugar kinase/adenylyltransferase [Cell envelope biogenesis, outer membrane]
u COG2885 (2) Outer membrane protein and related peptidoglycan-associated (lipo)proteins [Cell envelope biogenesis, outer membrane]
u COG2890 (1) Methylase of polypeptide chain release factors [Translation, ribosomal structure and biogenesis]
u COG2917 (1) Intracellular septation protein A [Cell division and chromosome partitioning]
u COG2932 (1) Predicted transcriptional regulator [Transcription]
d COG2947 (1) Uncharacterized conserved protein [Function unknown]
u COG2951 (2) Membrane-bound lytic murein transglycosylase B [Cell envelope biogenesis, outer membrane]
uD COG2982 (3) Uncharacterized protein involved in outer membrane biogenesis [Cell envelope biogenesis, outer membrane]
D COG3093 (1) Plasmid maintenance system antidote protein [General function prediction only]
uD COG3118 (2) Thioredoxin domain-containing protein [Posttranslational modification, protein turnover, chaperones]
d COG3145 (1) Alkylated DNA repair protein [DNA replication, recombination, and repair]
u d COG3153 (2) Predicted acetyltransferase [General function prediction only]
u COG3158 (1) K+ transporter [Inorganic ion transport and metabolism]
D COG3179 (1) Predicted chitinase [General function prediction only]
d COG3191 (1) L-aminopeptidase/D-esterase [Amino acid transport and metabolism / Secondary metabolites biosynthesis, transport, and catabolism]
uDd COG3206 (3) Uncharacterized protein involved in exopolysaccharide biosynthesis [Cell envelope biogenesis, outer membrane]
u COG3212 (2) Predicted membrane protein [Function unknown]
u COG3221 (1) ABC-type phosphate/phosphonate transport system, periplasmic component [Inorganic ion transport and metabolism]
u COG3226 (1) Uncharacterized protein conserved in bacteria [Function unknown]
u COG3265 (1) Gluconate kinase [Carbohydrate transport and metabolism]
uD COG3268 (2) Uncharacterized conserved protein [Function unknown]
d COG3276 (1) Selenocysteine-specific translation elongation factor [Translation, ribosomal structure and biogenesis]
u COG3284 (2) Transcriptional activator of acetoin/glycerol metabolism [Secondary metabolites biosynthesis, transport, and catabolism / Transcription]
u COG3288 (1) NAD/NADP transhydrogenase alpha subunit [Energy production and conversion]
u d COG3293 (12) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
D COG3324 (4) Predicted enzyme related to lactoylglutathione lyase [General function prediction only]
d COG3334 (1) Uncharacterized conserved protein [Function unknown]
u d COG3335 (3) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
uD COG3387 (3) Glucoamylase and related glycosyl hydrolases [Carbohydrate transport and

metabolism]
u COG3404 (1) Methenyl tetrahydrofolate cyclohydrolase [Amino acid transport and metabolism]
d COG3409 (1) Putative peptidoglycan-binding domain-containing protein [Cell envelope biogenesis, outer membrane]
uD COG3415 (3) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
u COG3437 (1) Response regulator containing a CheY-like receiver domain and an HD-GYP domain [Transcription / Signal transduction mechanisms]
u COG3509 (1) Poly(3-hydroxybutyrate) depolymerase [Secondary metabolites biosynthesis, transport, and catabolism]
d COG3536 (1) Uncharacterized protein conserved in bacteria [Function unknown]
u COG3549 (1) Plasmid maintenance system killer protein [General function prediction only]
u d COG3576 (3) Predicted flavin-nucleotide-binding protein structurally related to pyridoxine 5'-phosphate oxidase [General function prediction only]
uD COG3577 (2) Predicted aspartyl protease [General function prediction only]
D COG3609 (1) Predicted transcriptional regulators containing the CopG/Arc/MetJ DNA-binding domain [Transcription]
u d COG3637 (3) Opacity protein and related surface antigens [Cell envelope biogenesis, outer membrane]
D COG3639 (1) ABC-type phosphate/phosphonate transport system, permease component [Inorganic ion transport and metabolism]
uDd COG3666 (3) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
D COG3673 (1) Uncharacterized conserved protein [Function unknown]
uD COG3712 (2) Fe2+-dicitrate sensor, membrane component [Inorganic ion transport and metabolism / Signal transduction mechanisms]
u COG3735 (4) Uncharacterized protein conserved in bacteria [Function unknown]
d COG3737 (1) Uncharacterized conserved protein [Function unknown]
u COG3743 (1) Uncharacterized conserved protein [Function unknown]
D COG3748 (1) Predicted membrane protein [Function unknown]
u COG3754 (1) Lipopolysaccharide biosynthesis protein [Cell envelope biogenesis, outer membrane]
D COG3785 (1) Uncharacterized conserved protein [Function unknown]
u COG3791 (2) Uncharacterized conserved protein [Function unknown]
u COG3806 (1) Transcriptional activator [Transcription]
D COG3807 (1) Uncharacterized protein conserved in bacteria [Function unknown]
u COG3826 (1) Uncharacterized protein conserved in bacteria [Function unknown]
d COG3839 (1) ABC-type sugar transport systems, ATPase components [Carbohydrate transport and metabolism]
u COG3897 (1) Predicted methyltransferase [General function prediction only]
u d COG3914 (2) Predicted O-linked N-acetylglucosamine transferase, SPINDLY family [Post-translational modification, protein turnover, chaperones]
u COG3916 (2) N-acyl-L-homoserine lactone synthetase [Signal transduction mechanisms / Secondary metabolites biosynthesis, transport, and catabolism]
u d COG3920 (4) Signal transduction histidine kinase [Signal transduction mechanisms]
u COG3921 (1) Uncharacterized protein conserved in bacteria [Function unknown]
u COG4095 (1) Uncharacterized conserved protein [Function unknown]
D COG4105 (1) DNA uptake lipoprotein [General function prediction only]
D COG4108 (1) Peptide chain release factor RF-3 [Translation, ribosomal structure and biogenesis]
u COG4119 (1) Predicted NTP pyrophosphohydrolase [DNA replication, recombination, and repair / General function prediction only]
D COG4148 (2) ABC-type molybdate transport system, ATPase component [Inorganic ion transport and metabolism]
D COG4149 (1) ABC-type molybdate transport system, permease component [Inorganic ion transport and metabolism]
d COG4160 (1) ABC-type arginine/histidine transport system, permease component [Amino acid transport and metabolism]
D COG4166 (1) ABC-type oligopeptide transport system, periplasmic component [Amino acid transport and metabolism]
D COG4175 (1) ABC-type proline/glycine betaine transport system, ATPase component [Amino acid transport and metabolism]
u COG4176 (1) ABC-type proline/glycine betaine transport system, permease component [Amino acid transport and metabolism]
D COG4177 (1) ABC-type branched-chain amino acid transport system, permease component [Amino acid transport and metabolism]
u d COG4191 (2) Signal transduction histidine kinase regulating C4-dicarboxylate transport system [Signal transduction mechanisms]
D COG4214 (5) ABC-type xylose transport system, permease component [Carbohydrate transport and metabolism]
D COG4231 (1) Indolepyruvate ferredoxin oxidoreductase, alpha and beta subunits [Energy production and conversion]
uD COG4235 (2) Cytochrome c biogenesis factor [Posttranslational modification, protein turnover, chaperones]
d COG4249 (1) Uncharacterized protein containing caspase domain [General function prediction only]
u COG4312 (1) Uncharacterized protein conserved in bacteria [Function unknown]
u d COG4321 (2) Uncharacterized protein related to arylsulfate sulfotransferase involved in siderophore biosynthesis [General function prediction only]
u COG4338 (1) Uncharacterized protein conserved in bacteria [Function unknown]
u COG4520 (1) Surface antigen [Cell envelope biogenesis, outer membrane]
d COG4531 (1) ABC-type Zn2+ transport system, periplasmic component/surface adhesin [Inorganic ion transport and metabolism]
u COG4549 (1) Uncharacterized protein conserved in bacteria [Function unknown]
D COG4553 (1) Poly-beta-hydroxyalkanoate depolymerase [Lipid metabolism]
u COG4564 (1) Signal transduction histidine kinase [Signal transduction mechanisms]
u d COG4567 (2) Response regulator consisting of a CheY-like receiver domain and a Fis-type HTH domain [Signal transduction mechanisms / Transcription]
u COG4608 (1) ABC-type oligopeptide transport system, ATPase component [Amino acid transport and metabolism]
D COG4645 (1) Uncharacterized protein conserved in bacteria [Function unknown]
uD COG4649 (2) Uncharacterized protein conserved in bacteria [Function unknown]
u COG4651 (1) Kef-type K+ transport system, predicted NAD-binding component [Inorganic ion transport and metabolism]
D COG4663 (1) TRAP-type mannitol/chloroaromatic compound transport system, periplasmic component [Secondary metabolites biosynthesis, transport, and catabolism]

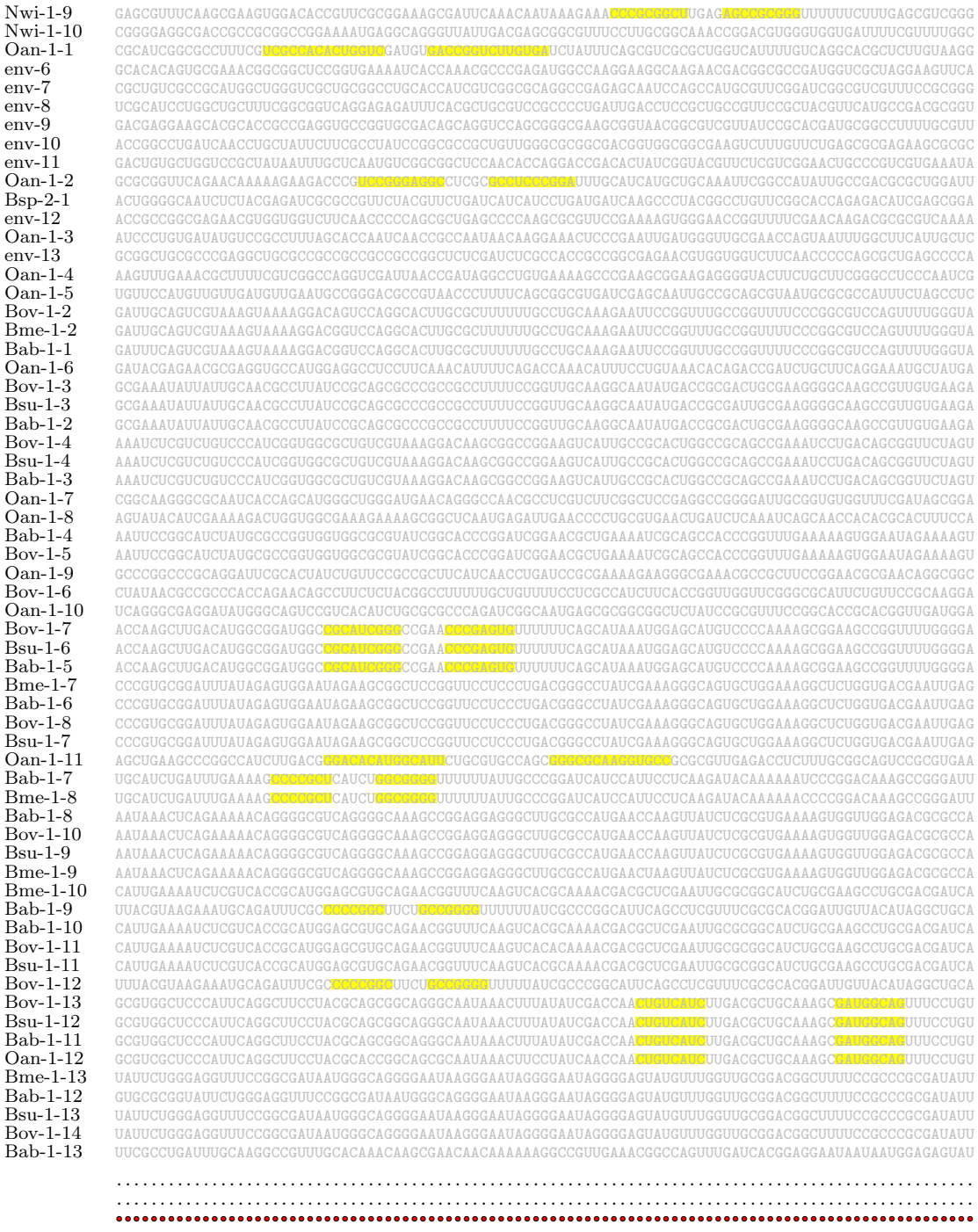
d COG4770 (1) Acetyl/propionyl-CoA carboxylase, alpha subunit [Lipid metabolism]
u COG4782 (1) Uncharacterized protein conserved in bacteria [Function unknown]
d COG4786 (1) Flagellar basal body rod protein [Cell motility and secretion]
d COG4799 (1) Acetyl-CoA carboxylase, carboxyltransferase component (subunits alpha and beta) [Lipid metabolism]
u COG4961 (1) Flp pilus assembly protein TadG [Intracellular trafficking and secretion]
D COG4976 (1) Predicted methyltransferase (contains TPR repeat) [General function prediction only]
d COG4977 (1) Transcriptional regulator containing an amidase domain and an AraC-type DNA-binding HTH domain [Transcription]
u COG4992 (5) Ornithine/acetylornithine aminotransferase [Amino acid transport and metabolism]
u d COG5001 (6) Predicted signal transduction protein containing a membrane domain, an EAL and a GGDEF domain [Signal transduction mechanisms]
d COG5009 (1) Membrane carboxypeptidase/penicillin-binding protein [Cell envelope biogenesis, outer membrane]
D COG5265 (1) ABC-type transport system involved in Fe-S cluster assembly, permease and ATPase components [Posttranslational modification, protein turnover, chaperones]
u COG5266 (1) ABC-type Co2+ transport system, periplasmic component [Inorganic ion transport and metabolism]
d COG5267 (1) Uncharacterized protein conserved in bacteria [Function unknown]
d COG5309 (1) Exo-beta-1,3-glucanase [Carbohydrate transport and metabolism]
u COG5338 (1) Uncharacterized protein conserved in bacteria [Function unknown]
u d COG5350 (2) Predicted protein tyrosine phosphatase [General function prediction only]
d COG5403 (1) Uncharacterized conserved protein [Function unknown]
u COG5455 (5) Predicted integral membrane protein [Function unknown]
u d COG5465 (2) Uncharacterized conserved protein [Function unknown]
u COG5473 (1) Predicted integral membrane protein [Function unknown]
D COG5567 (1) Predicted small periplasmic lipoprotein [Cell motility and secretion]
u d COG5616 (2) Predicted integral membrane protein [Function unknown]
u COG5653 (1) Protein involved in cellulose biosynthesis (CelD) [Cell envelope biogenesis, outer membrane]
u GOS1222 (1) ?
u GOS1286 (1) ?
d GOS1809 (1) ?
d GOS274 (1) ?
u GOS333 (1) ?
u NOG06853 (1)
D NOG07949 (1)
u NOG07957 (1)
d NOG08792 (1)
u NOG09695 (1)
u NOG09935 (1)
u NOG11613 (1)
u NOG15446 (1)
u NOG15857 (1)
u NOG18955 (1)
u NOG18972 (1)
u NOG19002 (1)
d NOG23552 (1)
u NOG27282 (1)
D NOG31707 (1)
u NOG32137 (1)
d NOG35082 (1)
d pfam00085 (1) Thioredoxin.
uDd pfam00106 (10) short chain dehydrogenase.
D pfam00120 (5) Glutamine synthetase, catalytic domain.
u pfam00140 (1) Sigma-70 factor, region 1.
d pfam00144 (1) Beta-lactamase.
uDd pfam00171 (9) Aldehyde dehydrogenase family.
D pfam00211 (1) Adenylate and Guanylate cyclase catalytic domain.
u pfam00226 (2) DnaJ domain.
D pfam00252 (2) Ribosomal protein L16.
D pfam00270 (1) DEAD/DEAH box helicase.
u pfam00293 (1) NUDIX domain.
u d pfam00313 (7) 'Cold-shock' DNA-binding domain.
u pfam00436 (1) Single-strand binding protein family.
D pfam00475 (1) Imidazoleglycerol-phosphate dehydratase.
u pfam00479 (1) Glucose-6-phosphate dehydrogenase, NAD binding domain.
u d pfam00497 (2) Bacterial extracellular solute-binding proteins, family 3.
d pfam00520 (1) Ion transport protein.
d pfam00565 (1) Staphylococcal nuclease homologue. [SNase numbering in parentheses]
uD pfam00583 (2) Acetyltransferase (GNAT) family.
d pfam00593 (1) TonB dependent receptor.
D pfam00639 (1) PPIC-type PPIASE domain.
u d pfam00665 (3) Integrase core domain.
u pfam00672 (2) HAMP domain.
u pfam00743 (1) Flavin-binding monooxygenase-like.
d pfam00753 (1) Metallo-beta-lactamase superfamily.
d pfam00842 (1) Alanine racemase, C-terminal domain.
D pfam00850 (5) Histone deacetylase family.
u pfam00873 (3) AcrB/AcrD/AcrF family.
u pfam00885 (1) 6,7-dimethyl-8-ribityllumazine synthase.
u pfam00889 (1) Elongation factor TS.
D pfam00892 (1) Integral membrane protein DUF6.
Dd pfam00920 (2) Dehydratase family.
uD pfam00926 (5) 3,4-dihydroxy-2-butanone 4-phosphate synthase.
d pfam00986 (1) DNA gyrase B subunit, carboxyl terminus.
u pfam01019 (2) Gamma-glutamyltranspeptidase.
u d pfam01043 (3) SecA protein, amino terminal region.
u pfam01047 (1) MarR family.
d pfam01132 (1) Elongation factor P (EF-P).
d pfam01136 (4) Peptidase family U32.
D pfam01243 (1) Pyridoxamine 5'-phosphate oxidase.
u pfam01245 (1) Ribosomal protein L19.
u d pfam01264 (2) Chorismate synthase.
D pfam01305 (1) Ribosomal protein L15 amino terminal region.
D pfam01322 (1) Cytochrome C'.
u pfam01411 (1) tRNA synthetases class II (A).
u pfam01425 (1) Amidase.
uD pfam01464 (4) Transglycosylase SLT domain.
u pfam01471 (1) Putative peptidoglycan binding domain.

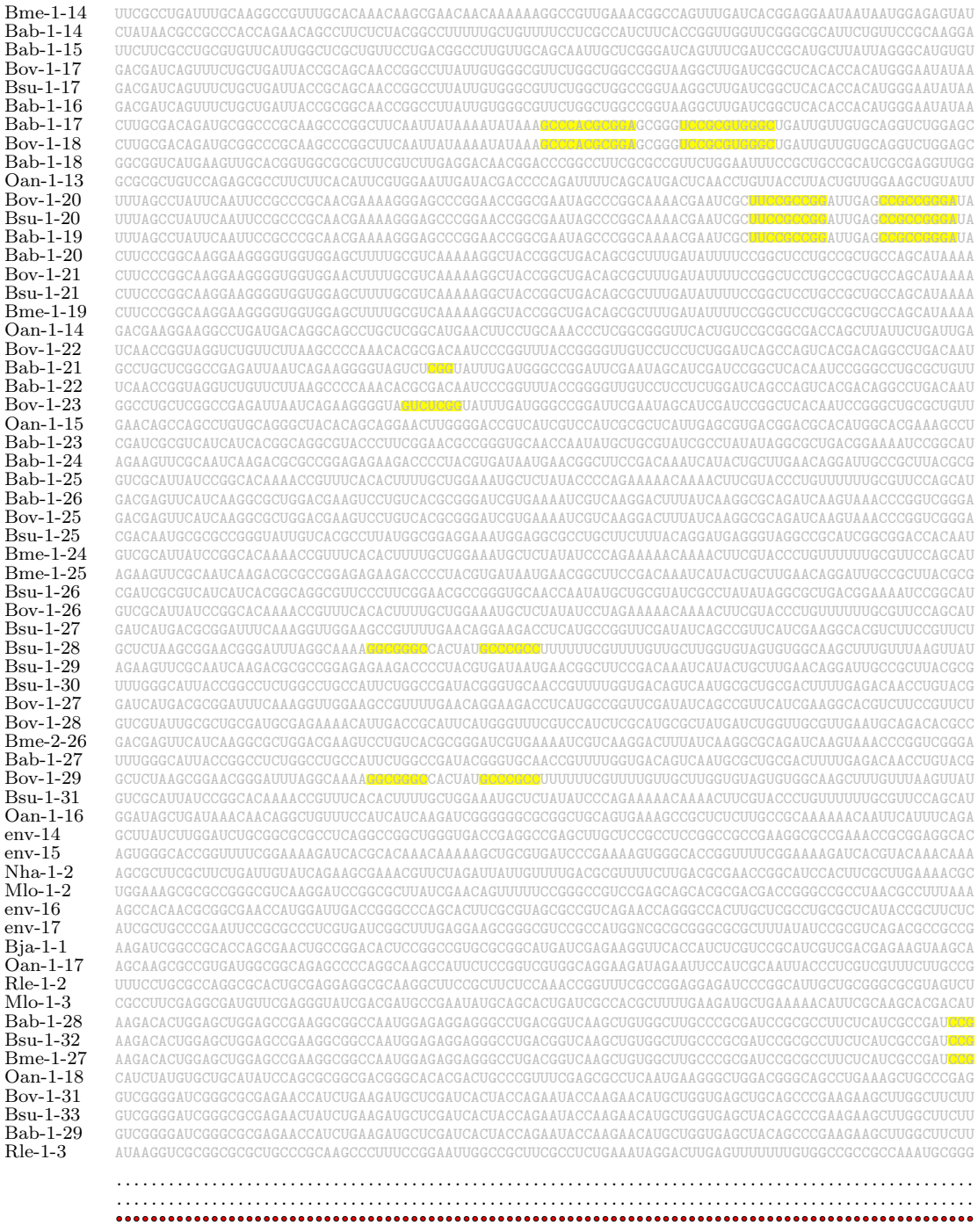
uD pfam01474 (2) Class-II DAHP synthetase family.
u pfam01502 (1) Phosphoribosyl-AMP cyclohydrolase.
d pfam01522 (1) Polysaccharide deacetylase.
uDd pfam01527 (3) Transposase.
u d pfam01593 (2) Flavin containing amine oxidoreductase.
uDd pfam01609 (4) Transposase DDE domain.
u pfam01627 (2) Hpt domain.
D pfam01642 (1) Methylmalonyl-CoA mutase.
d pfam01709 (5) Domain of unknown function DUF28.
u pfam01738 (1) Dienelactone hydrolase family.
u pfam01741 (1) Large-conductance mechanosensitive channel, MscL.
d pfam01746 (1) tRNA (Guanine-1)-methyltransferase.
u pfam01757 (1) Acyltransferase family.
d pfam01808 (1) AICARFT/IMPCHase bienzyme.
d pfam01809 (1) Domain of unknown function DUF37.
D pfam01820 (1) D-ala D-ala ligase.
u pfam02073 (1) Thermophilic metalloprotease (M29).
d pfam02142 (1) MGS-like domain.
u pfam02254 (1) TrkA-N domain.
uD pfam02405 (2) Domain of unknown function DUF140.
u pfam02410 (1) Domain of unknown function DUF143.
d pfam02421 (1) Ferrous iron transport protein B.
d pfam02447 (1) GntP family permease.
d pfam02463 (1) RecF/RecN/SMC N terminal domain.
D pfam02547 (1) Queuosine biosynthesis protein.
u pfam02563 (1) Polysaccharide biosynthesis/export protein.
u pfam02586 (1) Uncharacterised ACR, COG2135.
u pfam02618 (1) Aminodeoxychorismate lyase.
D pfam02673 (1) Bacitracin resistance protein BacA.
Dd pfam02674 (3) Colicin V production protein.
u pfam02683 (1) Cytochrome C biogenesis protein transmembrane region.
u pfam02781 (2) Glucose-6-phosphate dehydrogenase, C-terminal domain.
uDd pfam02787 (3) Carbamoyl-phosphate synthetase large chain, oligomerisation domain.
u d pfam02810 (3) SEC-C motif. [HC]
d pfam02823 (1) ATP synthase, Delta/Epsilon chain, beta-sandwich domain.
u pfam03030 (1) Inorganic H+ pyrophosphatase.
u pfam03240 (1) FlgA family.
D pfam03313 (1) Serine dehydratase alpha chain.
D pfam03315 (1) Serine dehydratase beta chain.
u pfam03331 (1) UDP-3-O-acyl N-acetylglycosamine deacetylase.
u pfam03466 (6) LysR substrate binding domain.
d pfam03472 (1) Autoinducer binding domain.
u pfam03601 (4) Conserved hypothetical protein 698.
D pfam03621 (1) MbtH-like protein.
uDd pfam03734 (4) ErfK/YbiS/YcfS/YnhG.
uD pfam03737 (2) Demethylmenaquinone methyltransferase.
D pfam03772 (1) Competence protein.
D pfam03829 (1) PTS system glucitol/sorbitol-specific IIA component.
u pfam03840 (1) Preprotein translocase SecG subunit.
D pfam03848 (1) Tellurite resistance protein TehB.
u pfam03950 (1) tRNA synthetases class I (E and Q), anti-codon binding domain.
D pfam03951 (5) Glutamine synthetase, beta-Grasp domain.
u pfam03979 (1) Sigma-70 factor, region 1.
u pfam04011 (1) LemA family.
u d pfam04230 (2) Polysaccharide pyruvyl transferase.
d pfam04463 (1) Protein of unknown function (DUF523).
u pfam04543 (1) Family of unknown function (DUF589).
D pfam04800 (1) ETC complex I subunit conserved region.
u pfam04897 (1) Glutamate synthase amidotransferase domain.
d pfam04940 (1) Sensors of blue-light using FAD.
d pfam04955 (1) HupE / UreJ protein.
u pfam04972 (1) Putative phospholipid-binding domain.
D pfam05016 (1) Plasmid stabilisation system protein.
u pfam05128 (1) Family of unknown function (DUF697).
D pfam05146 (1) Aha1 domain.
u pfam05199 (1) GMC oxidoreductase.
D pfam05221 (1) S-adenosyl-L-homocysteine hydrolase.
D pfam05226 (1) CHASE2 domain.
d pfam05227 (1) CHASE3 domain.
u pfam05239 (1) PRC-barrel domain.
d pfam05661 (1) Protein of unknown function (DUF808).
u pfam05853 (1) Prokaryotic protein of unknown function (DUF849).
u pfam05872 (5) Bacterial protein of unknown function (DUF853).
uD pfam05899 (2) Protein of unknown function (DUF861).
d pfam05955 (1) Equine herpesvirus glycoprotein gp2.
d pfam05977 (1) Bacterial protein of unknown function (DUF894).
uDd pfam06104 (3) Bacterial protein of unknown function (DUF949).
u pfam06319 (2) Protein of unknown function (DUF1052).
D pfam06499 (1) Protein of unknown function (DUF1099).
d pfam06532 (1) Protein of unknown function (DUF1109).
uDd pfam06559 (4) 2'-deoxycytidine 5'-triphosphate deaminase (DCD).
D pfam06676 (2) Protein of unknown function (DUF1178).
D pfam06707 (1) Protein of unknown function (DUF1194).
D pfam06731 (1) Protein of unknown function (DUF1209).
d pfam06912 (6) Protein of unknown function (DUF1275).
D pfam07045 (1) Protein of unknown function (DUF1330).
d pfam07080 (1) Protein of unknown function (DUF1348).
u d pfam07298 (2) NnrU protein.
u pfam07394 (1) Protein of unknown function (DUF1501).
u d smart00091 (3) PAS domain; PAS motifs appear in archaea, eubacteria and eukarya.
d smart00267 (1) Domain of Unknown Function with GGDEF motif; Domain apparently occurring exclusively in eubacteria.
u smart00283 (2) Methyl-accepting chemotaxis-like domains (chemotaxis sensory transducer).
d smart00342 (3) helix_turn_helix, arabinose operon control protein;
u smart00347 (6) helix_turn_helix multiple antibiotic resistance protein;
u smart00387 (1) Histidine kinase-like ATPases; Histidine kinase-, DNA gyrase B-, phytochrome-like ATPases.
u smart00418 (6) helix_turn_helix, Arsenical Resistance Operon Repressor;
D smart00642 (1) Alpha-amylase domain;
u d smart00671 (3) Sel1-like repeats.
u smart00732 (1) Likely ribonuclease with RNase H fold.

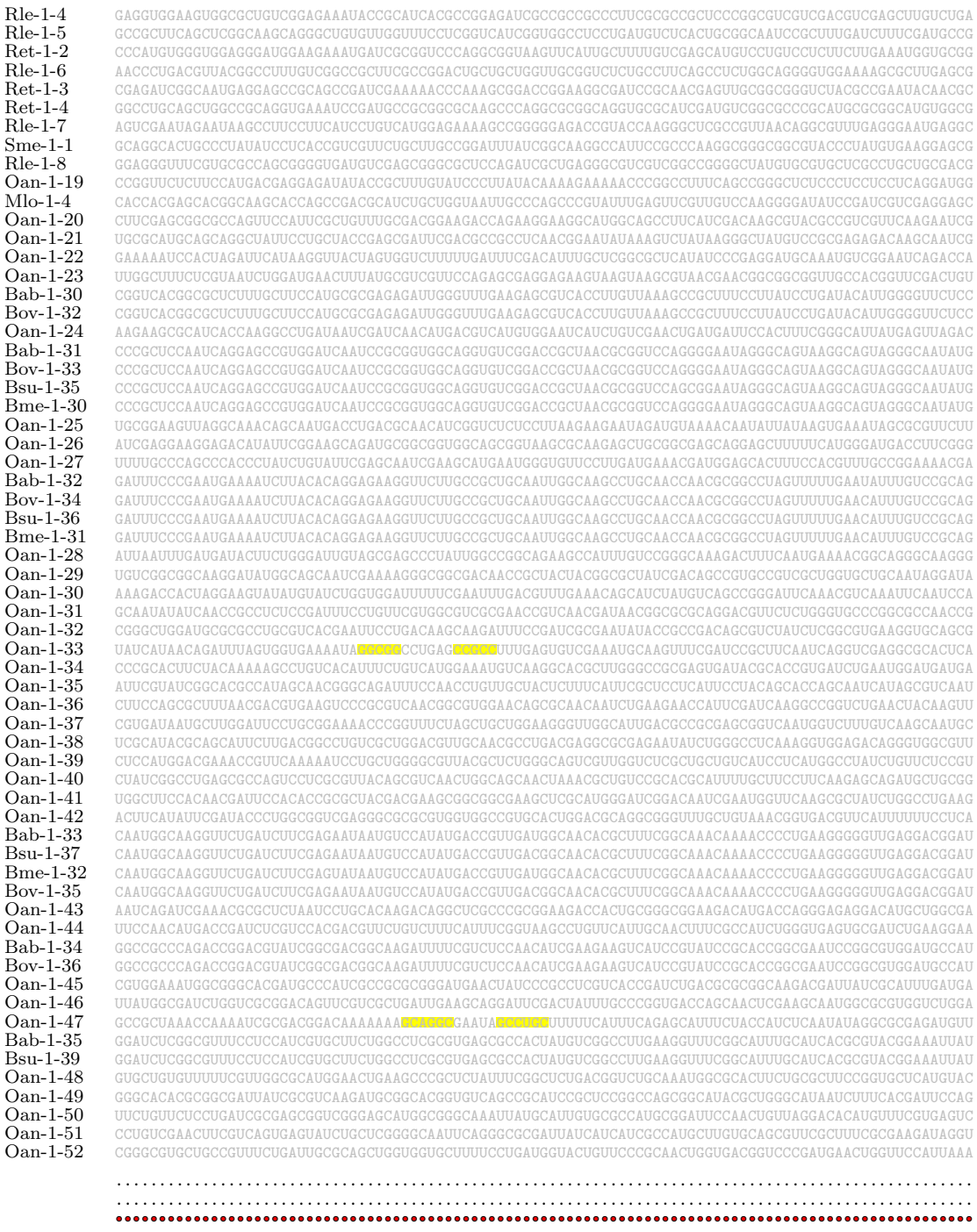
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Atu-1-2, Bme-1-11, Bme-1-12, Bme-1-15, Bme-1-16, Bme-1-17, Bme-1-18, Bme-1-20, Bme-1-21, Bme-1-22, Bme-1-23, Bme-1-26, Bme-1-28, Bme-1-29, Bme-1-3, Bme-1-33, Bme-1-34, Bme-1-35, Bme-1-37, Bme-1-4, Bme-1-5, Bme-1-6, Bme-2-1, Bme-2-10, Bme-2-11, Bme-2-12, Bme-2-13, Bme-2-14, Bme-2-15, Bme-2-16, Bme-2-17, Bme-2-18, Bme-2-19, Bme-2-2, Bme-2-20, Bme-2-21, Bme-2-22, Bme-2-23, Bme-2-24, Bme-2-25, Bme-2-27, Bme-2-28, Bme-2-29, Bme-2-3, Bme-2-30, Bme-2-31, Bme-2-32, Bme-2-33, Bme-2-34, Bme-2-35, Bme-2-36, Bme-2-37, Bme-2-38, Bme-2-39, Bme-2-4, Bme-2-5, Bme-2-6, Bme-2-7, Bme-2-8, Bme-2-9, Bov-1-15, Bov-1-16, Bov-1-19, Bov-1-24, Bov-1-30, Bov-1-37, Bov-1-38, Bov-1-9, Bsu-1-10, Bsu-1-14, Bsu-1-15, Bsu-1-16, Bsu-1-18, Bsu-1-19, Bsu-1-2, Bsu-1-22, Bsu-1-23, Bsu-1-24, Bsu-1-34, Bsu-1-38, Bsu-1-5, Bsu-1-8

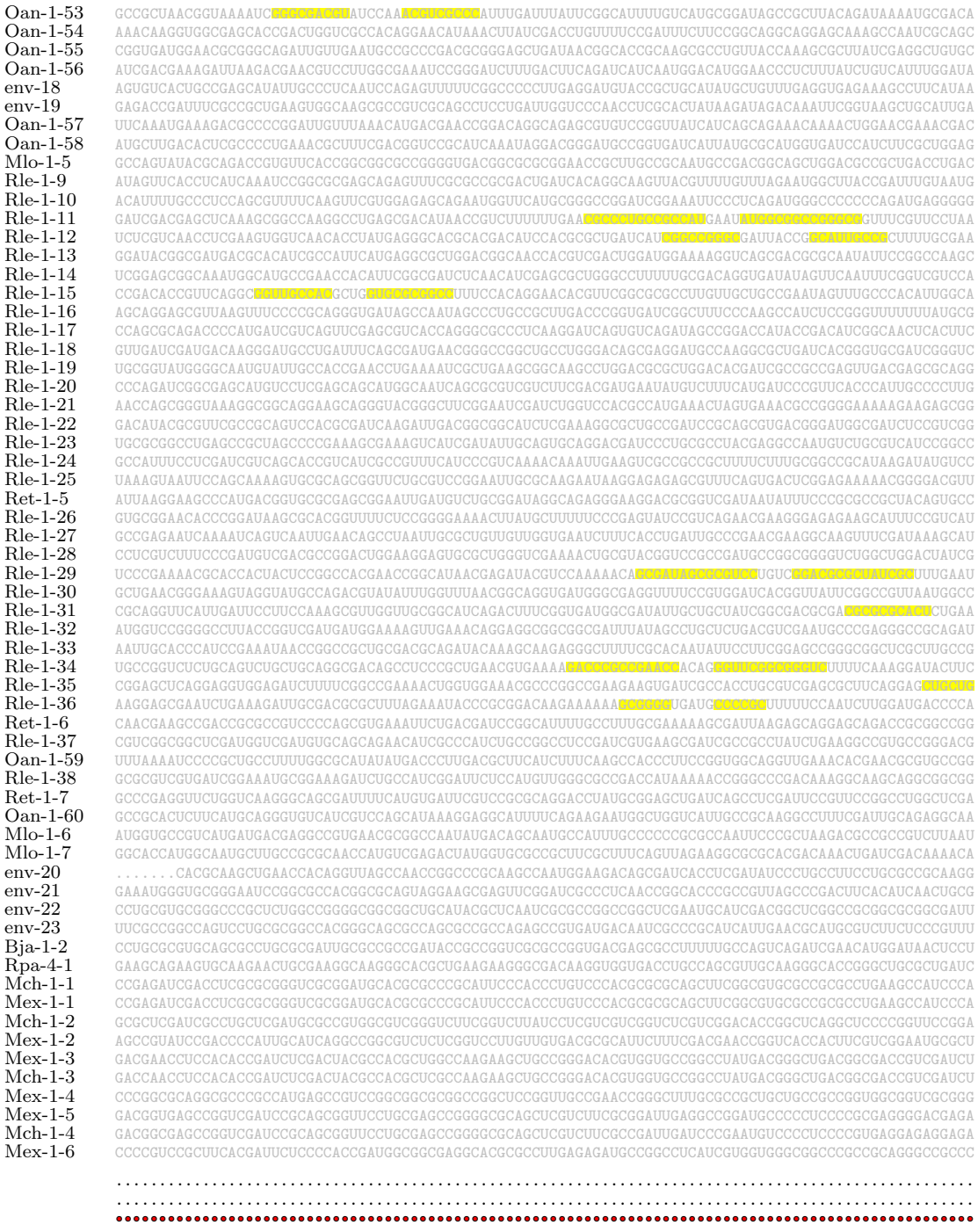
79.4 Multiple-sequence alignment

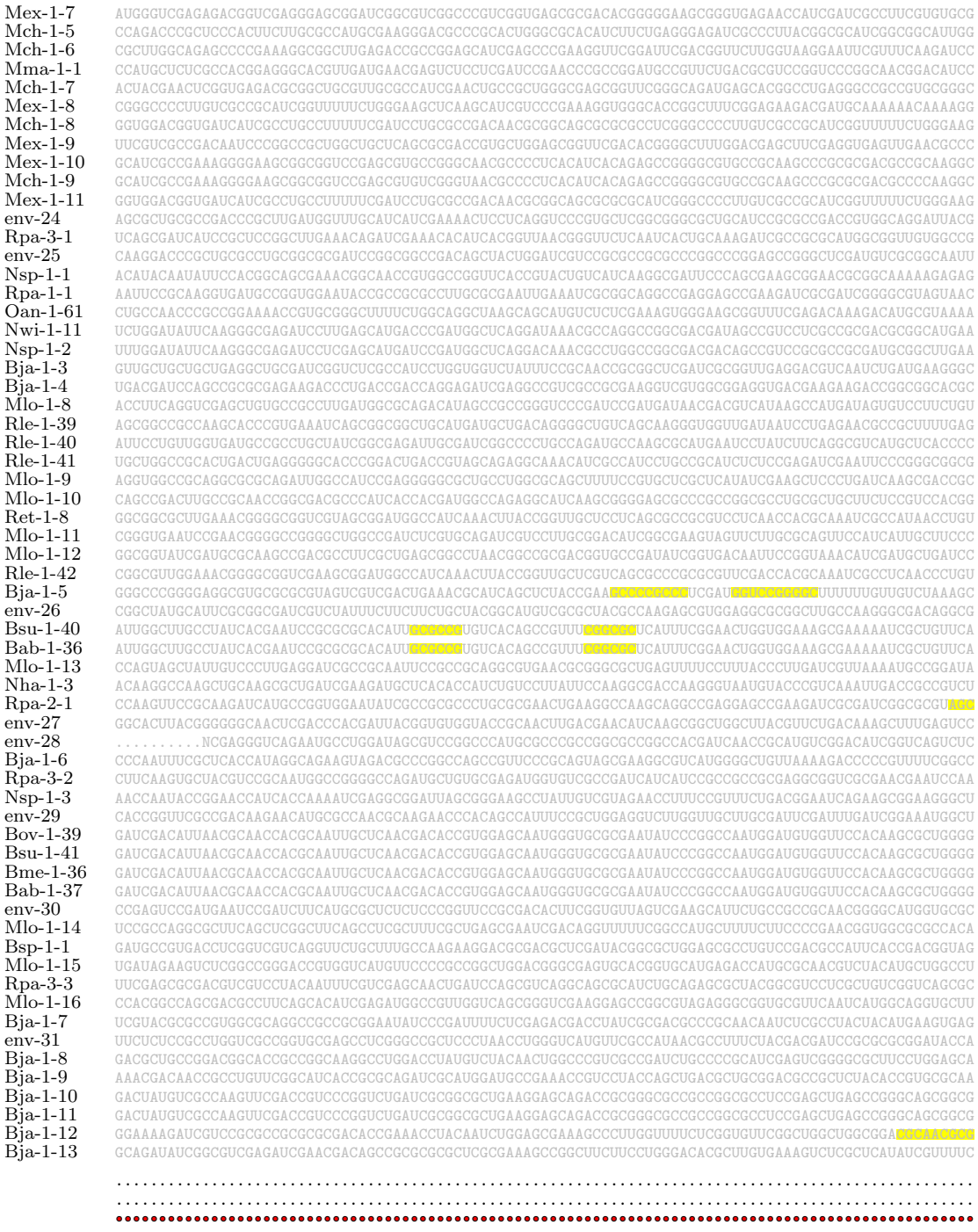
Each Rhizobiales-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 79.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Rhizobiales-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

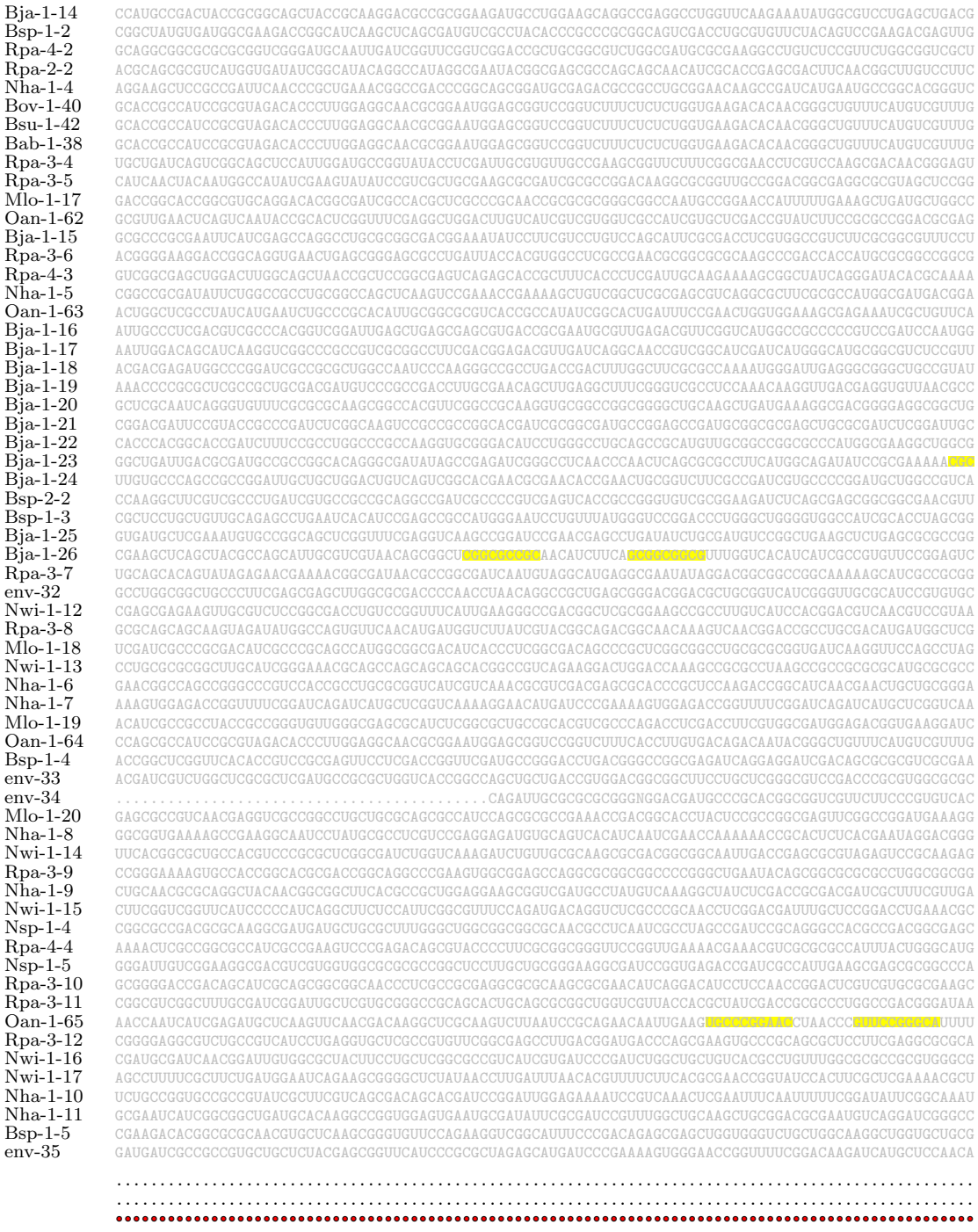

Nwi-1-9
Nwi-1-10
Oan-1-1
env-6
env-7
env-8
env-9
env-10
env-11
Oan-1-2
Bsp-2-1
env-12
Oan-1-3
env-13
Oan-1-4
Oan-1-5
Bov-1-2
Bme-1-2
Bab-1-1
Oan-1-6
Bov-1-3
Bsu-1-3
Bab-1-2
Bov-1-4
Bsu-1-4
Bab-1-3
Oan-1-7
Oan-1-8
Bab-1-4
Bov-1-5
Oan-1-9
Bov-1-6
Oan-1-10
Bov-1-7
Bsu-1-6
Bab-1-5
Bme-1-7
Bab-1-6
Bov-1-8
Bsu-1-7
Oan-1-11
Bab-1-7
Bme-1-8
Bab-1-8
Bov-1-10
Bsu-1-9
Bme-1-9
Bme-1-10
Bab-1-9
Bab-1-10
Bov-1-11
Bsu-1-11
Bov-1-12
Bov-1-13
Bsu-1-12
Bab-1-11
Oan-1-12
Bme-1-13
Bab-1-12
Bsu-1-13
Bov-1-14
Bab-1-13

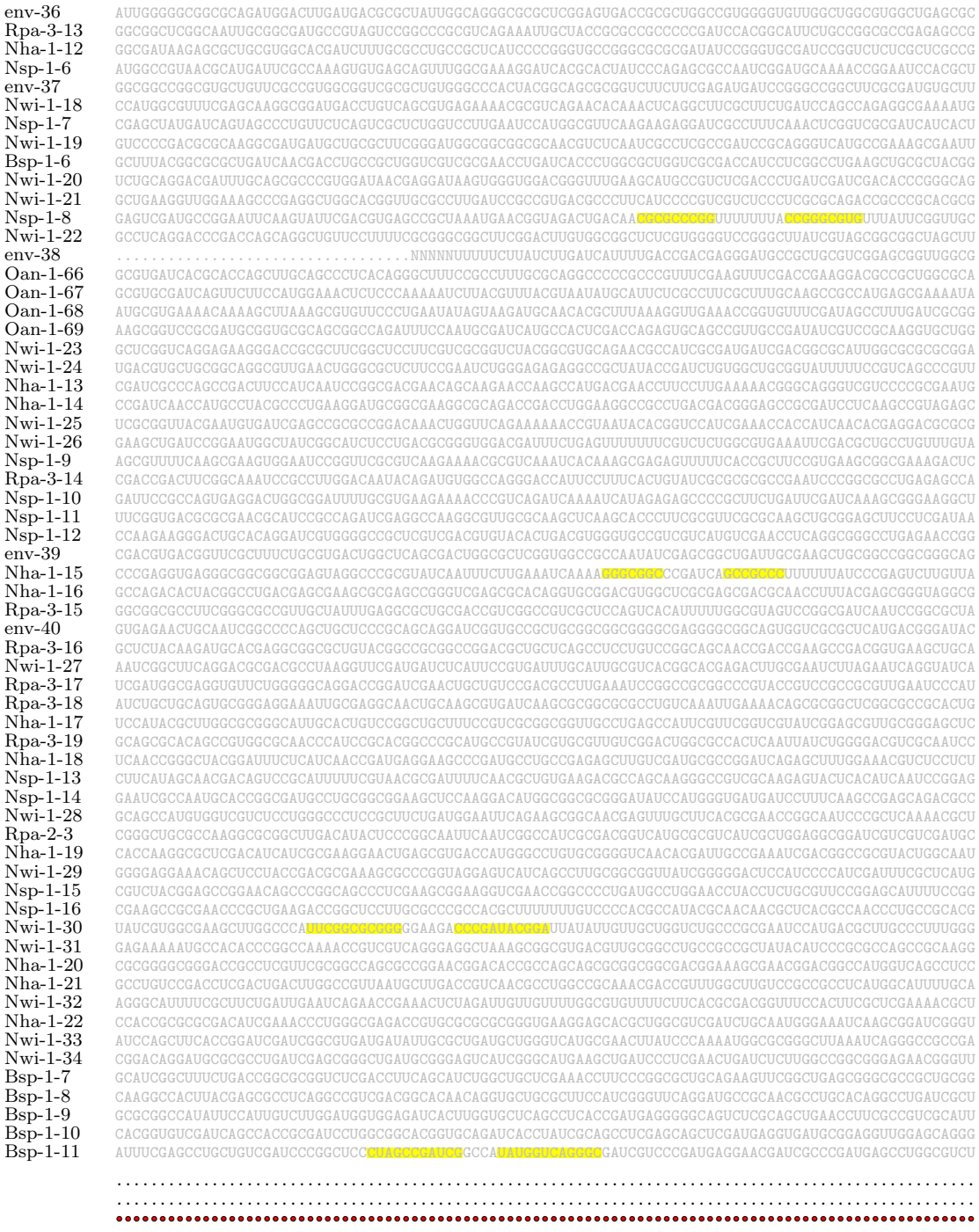

Bme-1-14
Bab-1-14
Bab-1-15
Bov-1-17
Bsu-1-17
Bab-1-16
Bab-1-17
Bov-1-18
Bab-1-18
Oan-1-13
Bov-1-20
Bsu-1-20
Bab-1-19
Bab-1-20
Bov-1-21
Bsu-1-21
Bme-1-19
Oan-1-14
Bov-1-22
Bab-1-21
Bab-1-22
Bov-1-23
Oan-1-15
Bab-1-23
Bab-1-24
Bab-1-25
Bab-1-26
Bov-1-25
Bsu-1-25
Bme-1-24
Bme-1-25
Bsu-1-26
Bov-1-26
Bsu-1-27
Bsu-1-28
Bsu-1-29
Bsu-1-30
Bov-1-27
Bov-1-28
Bme-2-26
Bab-1-27
Bov-1-29
Bsu-1-31
Oan-1-16
env-14
env-15
Nha-1-2
Mlo-1-2
env-16
env-17
Bja-1-1
Oan-1-17
Rle-1-2
Mlo-1-3
Bab-1-28
Bsu-1-32
Bme-1-27
Oan-1-18
Bov-1-31
Bsu-1-33
Bab-1-29
Rle-1-3

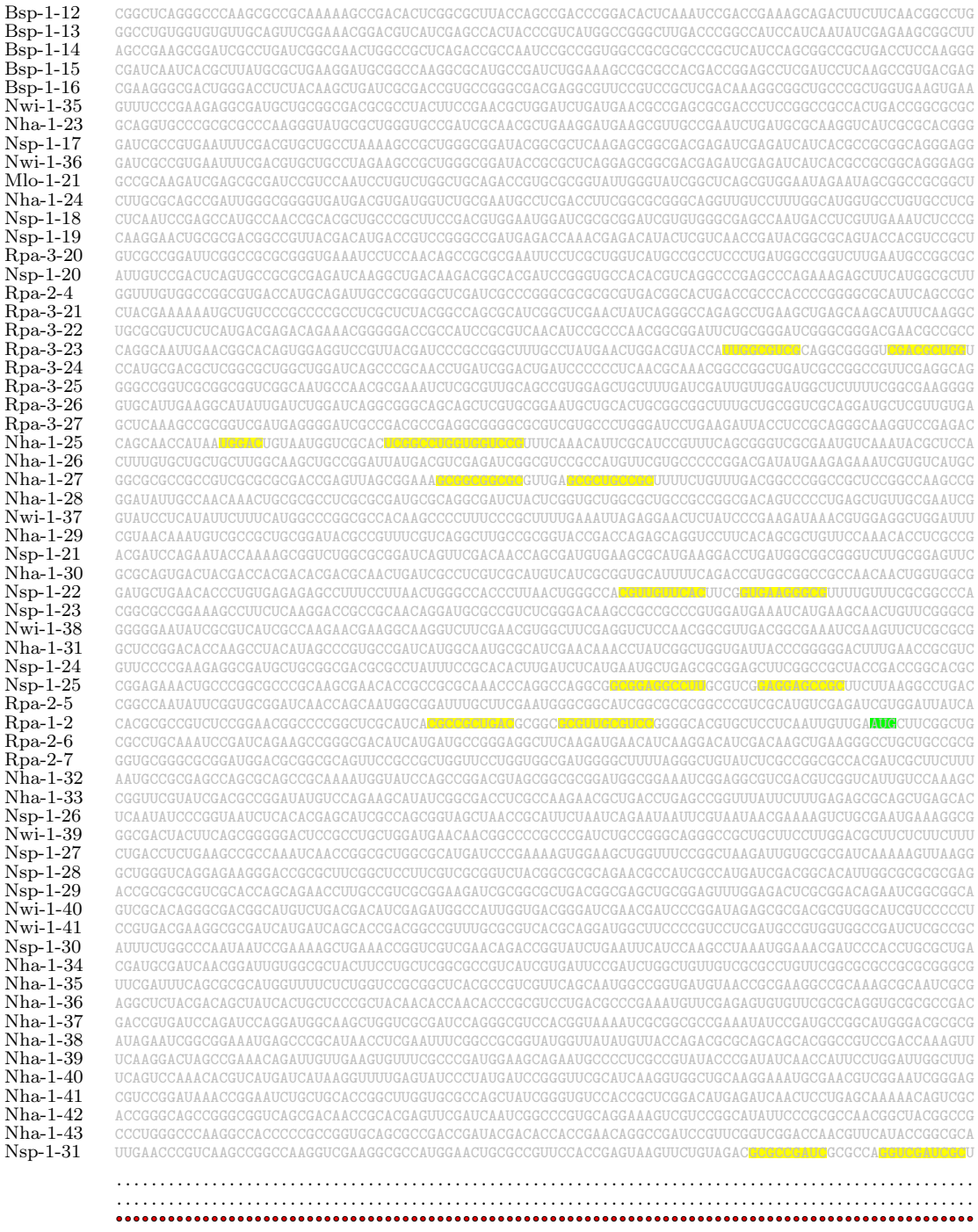

Rle-1-4
Rle-1-5
Ret-1-2
Rle-1-6
Ret-1-3
Ret-1-4
Rle-1-7
Sme-1-1
Rle-1-8
Oan-1-19
Mlo-1-4
Oan-1-20
Oan-1-21
Oan-1-22
Oan-1-23
Bab-1-30
Bov-1-32
Oan-1-24
Bab-1-31
Bov-1-33
Bsu-1-35
Bme-1-30
Oan-1-25
Oan-1-26
Oan-1-27
Bab-1-32
Bov-1-34
Bsu-1-36
Bme-1-31
Oan-1-28
Oan-1-29
Oan-1-30
Oan-1-31
Oan-1-32
Oan-1-33
Oan-1-34
Oan-1-35
Oan-1-36
Oan-1-37
Oan-1-38
Oan-1-39
Oan-1-40
Oan-1-41
Oan-1-42
Bab-1-33
Bsu-1-37
Bme-1-32
Bov-1-35
Oan-1-43
Oan-1-44
Bab-1-34
Bov-1-36
Oan-1-45
Oan-1-46
Oan-1-47
Bab-1-35
Bsu-1-39
Oan-1-48
Oan-1-49
Oan-1-50
Oan-1-51
Oan-1-52

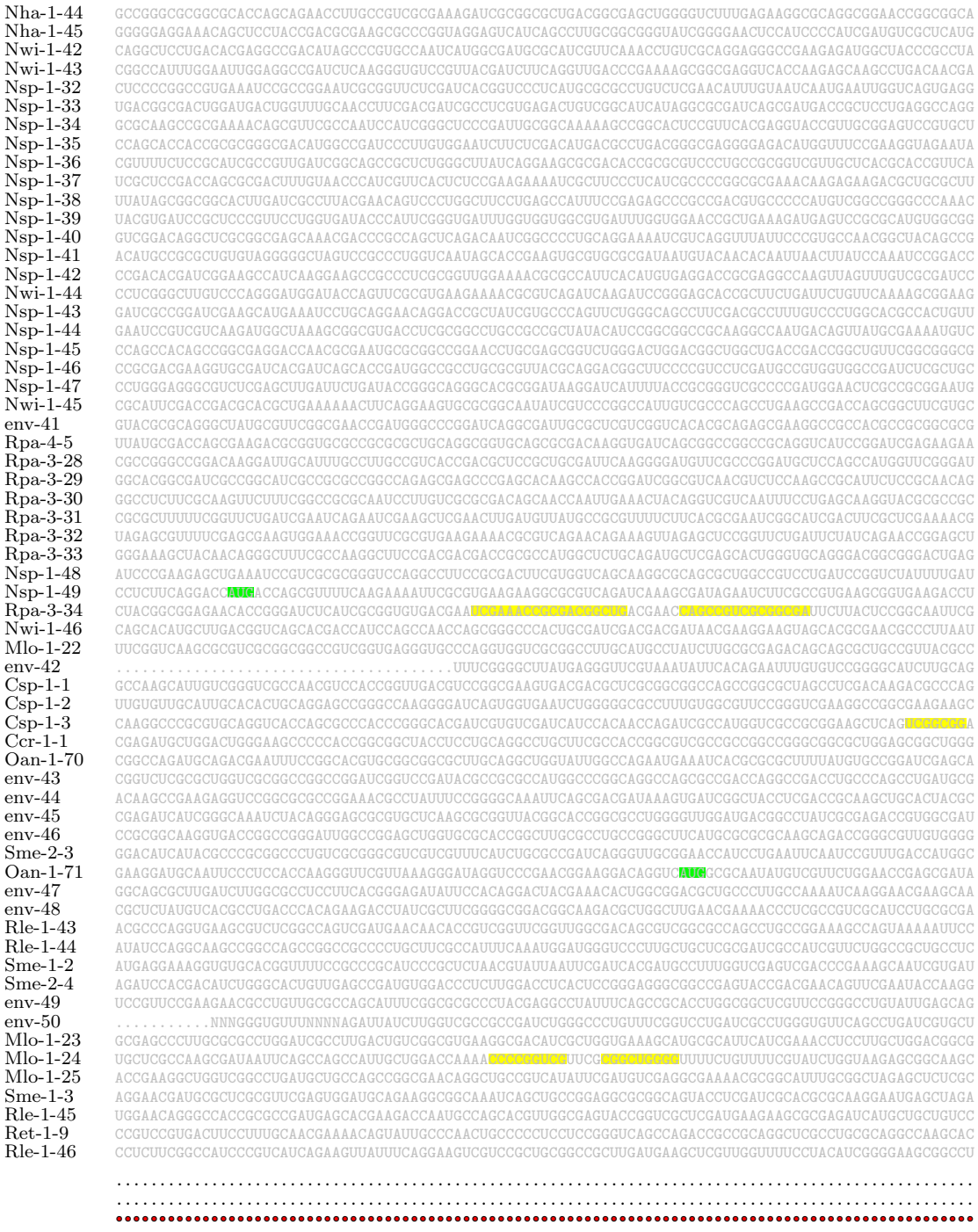

Oan-1-53
Oan-1-54
Oan-1-55
Oan-1-56
env-18
env-19
Oan-1-57
Oan-1-58
Mlo-1-5
Rle-1-9
Rle-1-10
Rle-1-11
Rle-1-12
Rle-1-13
Rle-1-14
Rle-1-15
Rle-1-16
Rle-1-17
Rle-1-18
Rle-1-19
Rle-1-20
Rle-1-21
Rle-1-22
Rle-1-23
Rle-1-24
Rle-1-25
Ret-1-5
Rle-1-26
Rle-1-27
Rle-1-28
Rle-1-29
Rle-1-30
Rle-1-31
Rle-1-32
Rle-1-33
Rle-1-34
Rle-1-35
Rle-1-36
Ret-1-6
Rle-1-37
Oan-1-59
Rle-1-38
Ret-1-7
Oan-1-60
Mlo-1-6
Mlo-1-7
env-20
env-21
env-22
env-23
Bja-1-2
Rpa-4-1
Mch-1-1
Mex-1-1
Mch-1-2
Mex-1-2
Mex-1-3
Mch-1-3
Mex-1-4
Mex-1-5
Mch-1-4
Mex-1-6

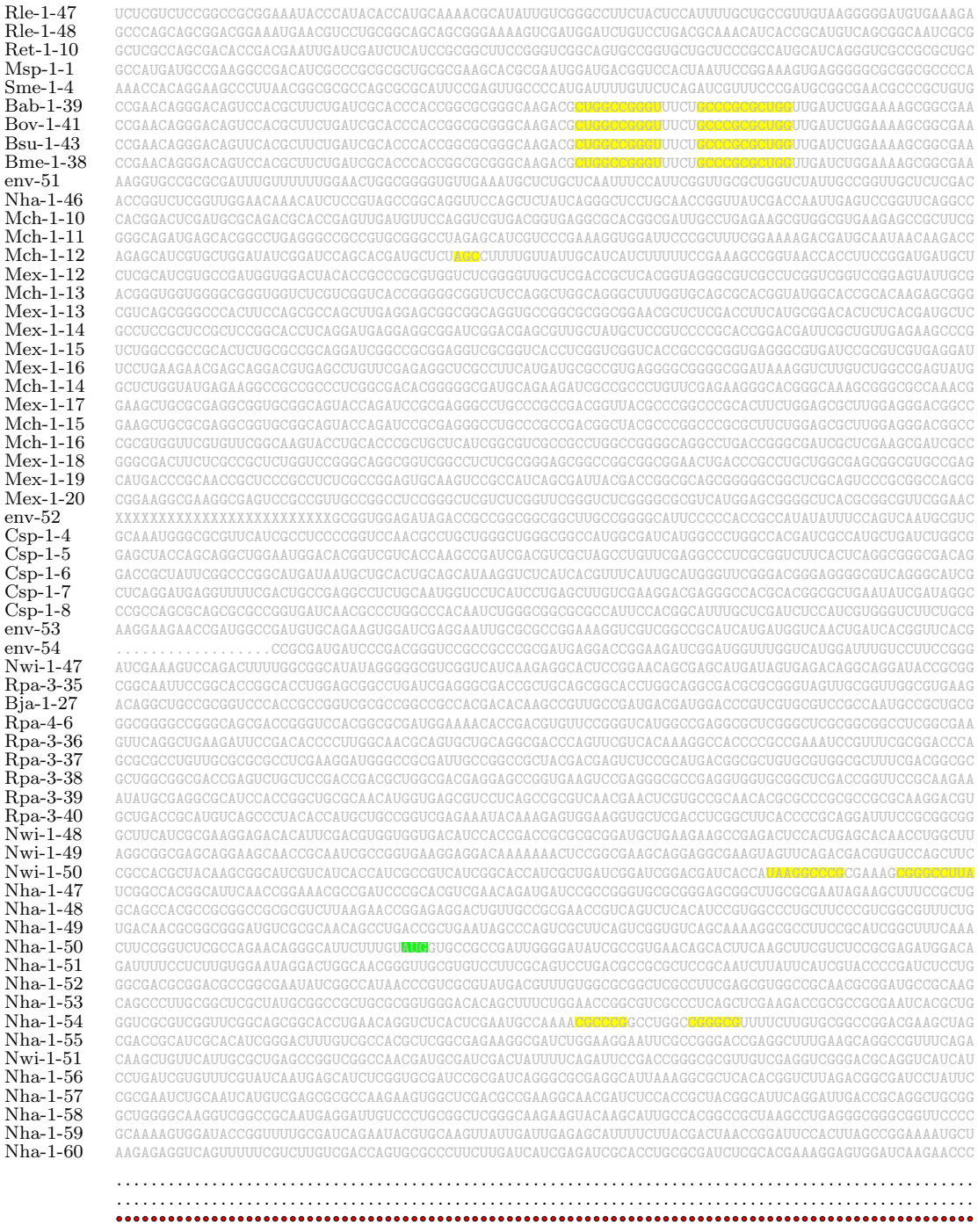

Mex-1-7
Mch-1-5
Mch-1-6
Mma-1-1
Mch-1-7
Mex-1-8
Mch-1-8
Mex-1-9
Mex-1-10
Mch-1-9
Mex-1-11
env-24
Rpa-3-1
env-25
Nsp-1-1
Rpa-1-1
Oan-1-61
Nwi-1-11
Nsp-1-2
Bja-1-3
Bja-1-4
Mlo-1-8
Rle-1-39
Rle-1-40
Rle-1-41
Mlo-1-9
Mlo-1-10
Ret-1-8
Mlo-1-11
Mlo-1-12
Rle-1-42
Bja-1-5
env-26
Bsu-1-40
Bab-1-36
Mlo-1-13
Nha-1-3
Rpa-2-1
env-27
env-28
Bja-1-6
Rpa-3-2
Nsp-1-3
env-29
Bov-1-39
Bsu-1-41
Bme-1-36
Bab-1-37
env-30
Mlo-1-14
Bsp-1-1
Mlo-1-15
Rpa-3-3
Mlo-1-16
Bja-1-7
env-31
Bja-1-8
Bja-1-9
Bja-1-10
Bja-1-11
Bja-1-12
Bja-1-13

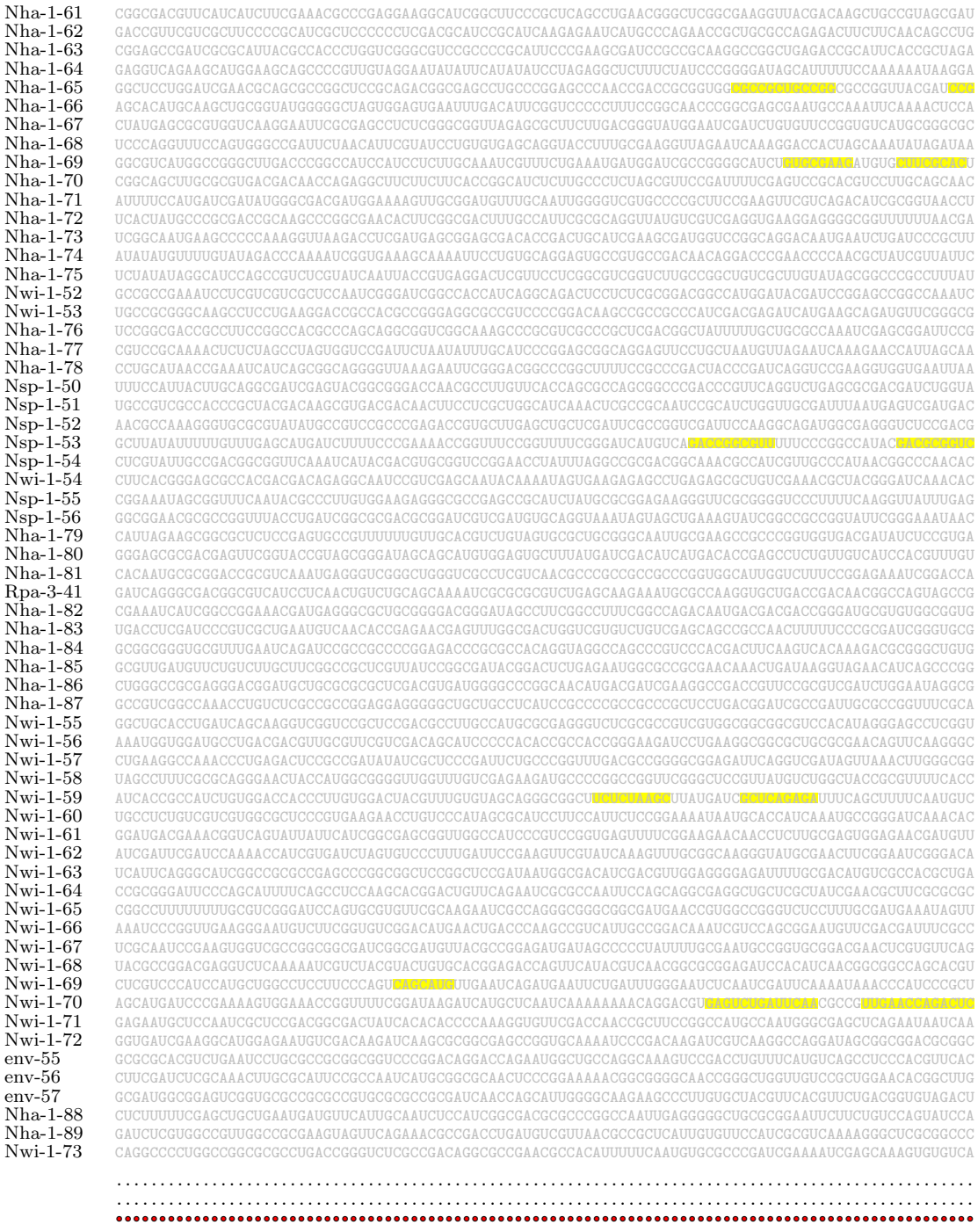

Bja-1-14
Bsp-1-2
Rpa-4-2
Rpa-2-2
Nha-1-4
Bov-1-40
Bsu-1-42
Bab-1-38
Rpa-3-4
Rpa-3-5
Mlo-1-17
Oan-1-62
Bja-1-15
Rpa-3-6
Rpa-4-3
Nha-1-5
Oan-1-63
Bja-1-16
Bja-1-17
Bja-1-18
Bja-1-19
Bja-1-20
Bja-1-21
Bja-1-22
Bja-1-23
Bja-1-24
Bsp-2-2
Bsp-1-3
Bja-1-25
Bja-1-26
Rpa-3-7
env-32
Nwi-1-12
Rpa-3-8
Mlo-1-18
Nwi-1-13
Nha-1-6
Nha-1-7
Mlo-1-19
Oan-1-64
Bsp-1-4
env-33
env-34
Mlo-1-20
Nha-1-8
Nwi-1-14
Rpa-3-9
Nha-1-9
Nwi-1-15
Nsp-1-4
Rpa-4-4
Nsp-1-5
Rpa-3-10
Rpa-3-11
Oan-1-65
Rpa-3-12
Nwi-1-16
Nwi-1-17
Nha-1-10
Nha-1-11
Bsp-1-5
env-35

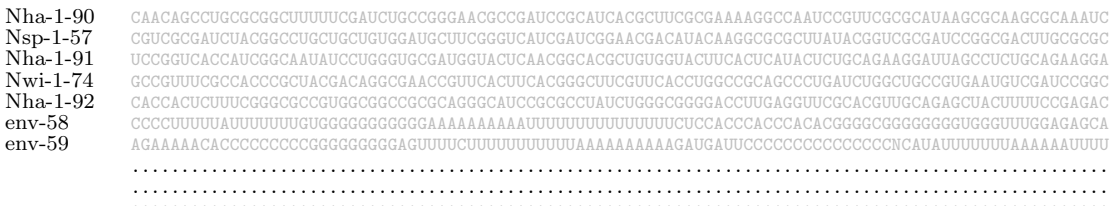

env-36
Rpa-3-13
Nha-1-12
Nsp-1-6
env-37
Nwi-1-18
Nsp-1-7
Nwi-1-19
Bsp-1-6
Nwi-1-20
Nwi-1-21
Nsp-1-8
Nwi-1-22
env-38
Oan-1-66
Oan-1-67
Oan-1-68
Oan-1-69
Nwi-1-23
Nwi-1-24
Nha-1-13
Nha-1-14
Nwi-1-25
Nwi-1-26
Nsp-1-9
Rpa-3-14
Nsp-1-10
Nsp-1-11
Nsp-1-12
env-39
Nha-1-15
Nha-1-16
Rpa-3-15
env-40
Rpa-3-16
Nwi-1-27
Rpa-3-17
Rpa-3-18
Nha-1-17
Rpa-3-19
Nha-1-18
Nsp-1-13
Nsp-1-14
Nwi-1-28
Rpa-2-3
Nha-1-19
Nwi-1-29
Nsp-1-15
Nsp-1-16
Nwi-1-30
Nwi-1-31
Nha-1-20
Nha-1-21
Nwi-1-32
Nha-1-22
Nwi-1-33
Nwi-1-34
Bsp-1-7
Bsp-1-8
Bsp-1-9
Bsp-1-10
Bsp-1-11


Bsp-1-12
Bsp-1-13
Bsp-1-14
Bsp-1-15
Bsp-1-16
Nwi-1-35
Nha-1-23
Nsp-1-17
Nwi-1-36
Mlo-1-21
Nha-1-24
Nsp-1-18
Nsp-1-19
Rpa-3-20
Nsp-1-20
Rpa-2-4
Rpa-3-21
Rpa-3-22
Rpa-3-23
Rpa-3-24
Rpa-3-25
Rpa-3-26
Rpa-3-27
Nha-1-25
Nha-1-26
Nha-1-27
Nha-1-28
Nwi-1-37
Nha-1-29
Nsp-1-21
Nha-1-30
Nsp-1-22
Nsp-1-23
Nwi-1-38
Nha-1-31
Nsp-1-24
Nsp-1-25
Rpa-2-5
Rpa-1-2
Rpa-2-6
Rpa-2-7
Nha-1-32
Nha-1-33
Nsp-1-26
Nwi-1-39
Nsp-1-27
Nsp-1-28
Nsp-1-29
Nwi-1-40
Nwi-1-41
Nsp-1-30
Nha-1-34
Nha-1-35
Nha-1-36
Nha-1-37
Nha-1-38
Nha-1-39
Nha-1-40
Nha-1-41
Nha-1-42
Nha-1-43
Nsp-1-31

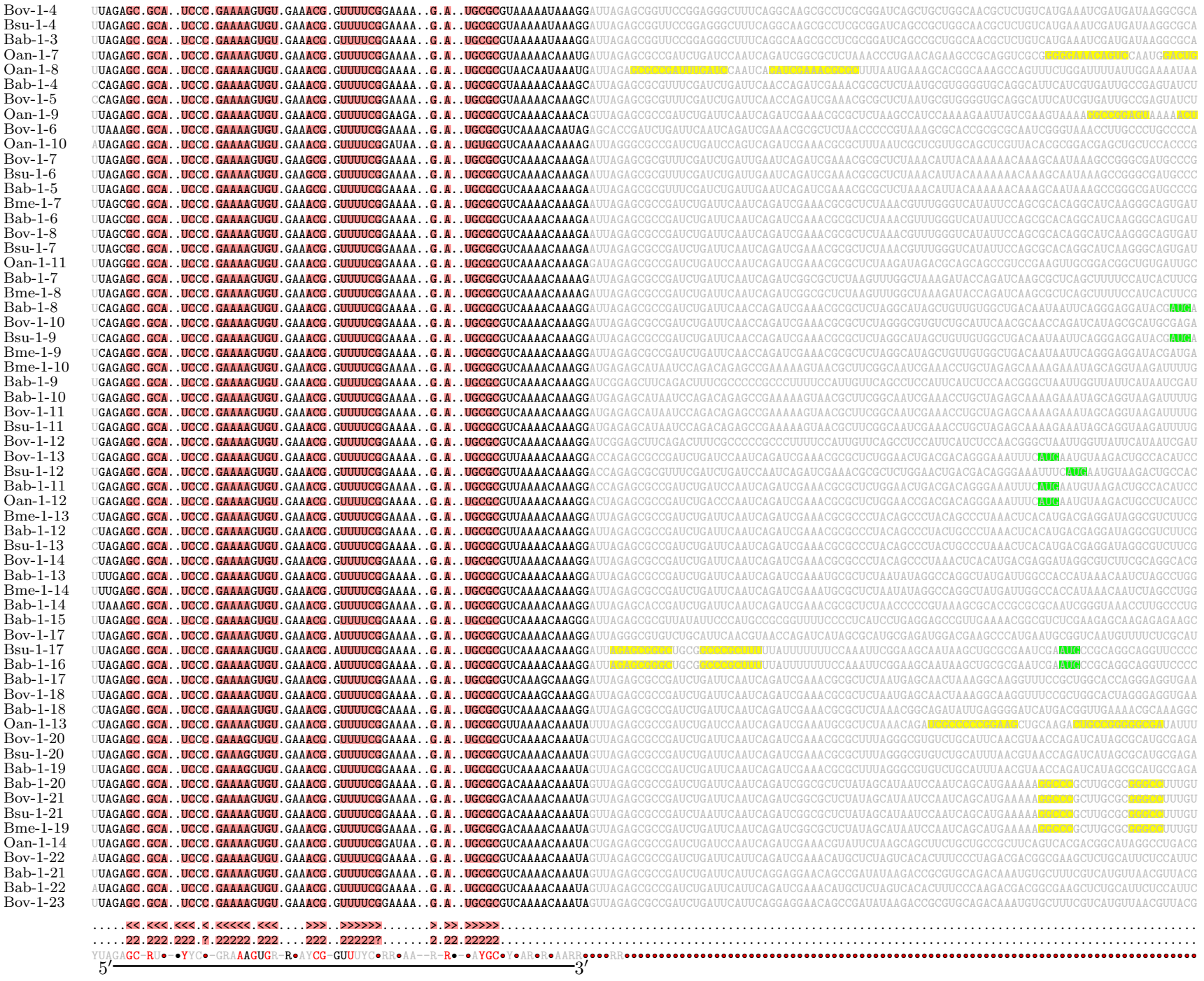

Nha-1-44
Nha-1-45
Nwi-1-42
Nwi-1-43
Nsp-1-32
Nsp-1-33
Nsp-1-34
Nsp-1-35
Nsp-1-36
Nsp-1-37
Nsp-1-38
Nsp-1-39
Nsp-1-40
Nsp-1-41
Nsp-1-42
Nwi-1-44
Nsp-1-43
Nsp-1-44
Nsp-1-45
Nsp-1-46
Nsp-1-47
Nwi-1-45
env-41
Rpa-4-5
Rpa-3-28
Rpa-3-29
Rpa-3-30
Rpa-3-31
Rpa-3-32
Rpa-3-33
Nsp-1-48
Nsp-1-49
Rpa-3-34
Nwi-1-46
Mlo-1-22
env-42
Csp-1-1
Csp-1-2
Csp-1-3
Ccr-1-1
Oan-1-70
env-43
env-44
env-45
env-46
Sme-2-3
Oan-1-71
env-47
env-48
Rle-1-43
Rle-1-44
Sme-1-2
Sme-2-4
env-49
env-50
Mlo-1-23
Mlo-1-24
Mlo-1-25
Sme-1-3
Rle-1-45
Ret-1-9
Rle-1-46

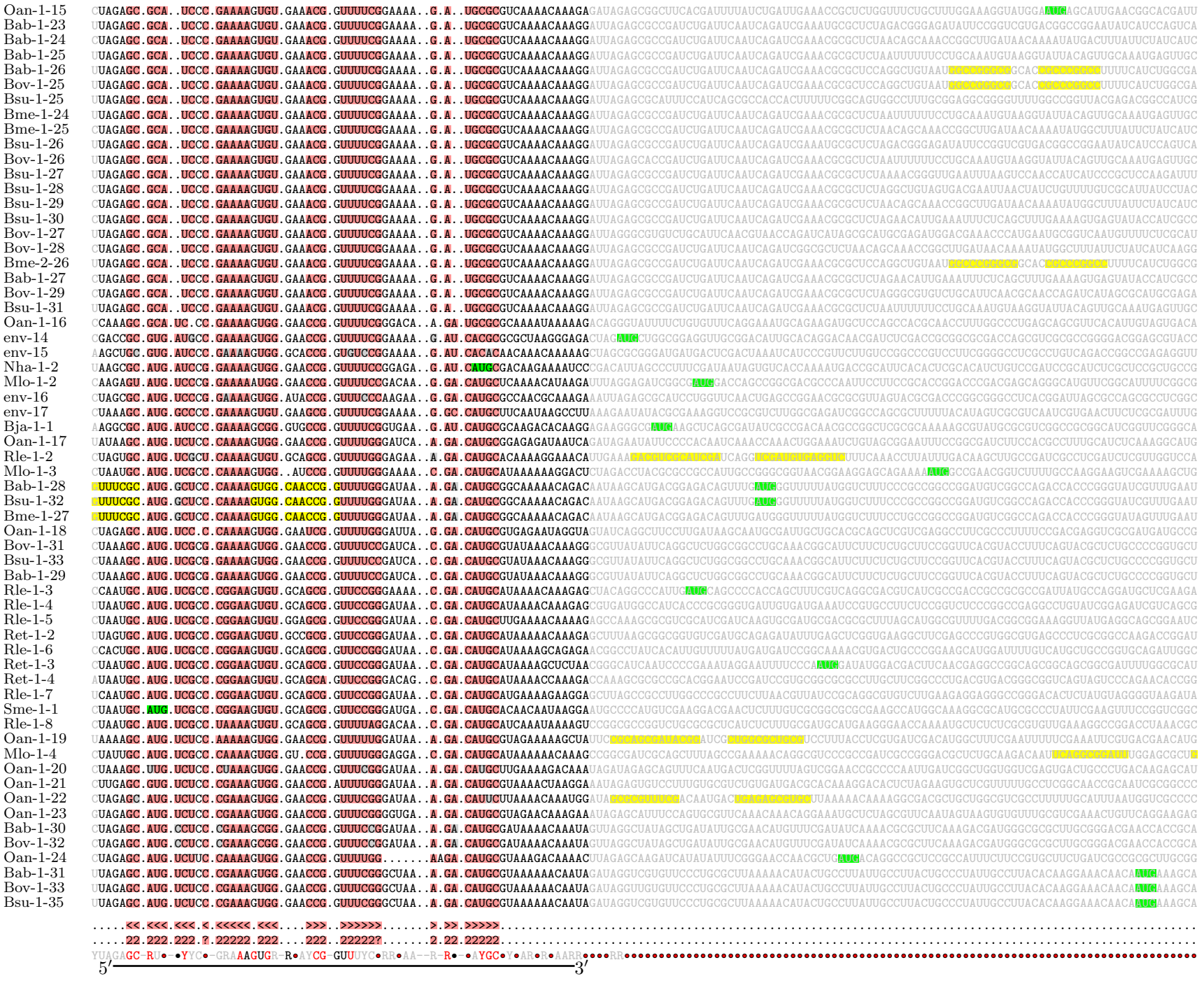

Rle-1-47
Rle-1-48
Ret-1-10
Msp-1-1
Sme-1-4
Bab-1-39
Bov-1-41
Bsu-1-43
Bme-1-38
env-51
Nha-1-46
Mch-1-10
Mch-1-11
Mch-1-12
Mex-1-12
Mch-1-13
Mex-1-13
Mex-1-14
Mex-1-15
Mex-1-16
Mch-1-14
Mex-1-17
Mch-1-15
Mch-1-16
Mex-1-18
Mex-1-19
Mex-1-20
env-52
Csp-1-4
Csp-1-5
Csp-1-6
Csp-1-7
Csp-1-8
env-53
env-54
Nwi-1-47
Rpa-3-35
Bja-1-27
Rpa-4-6
Rpa-3-36
Rpa-3-37
Rpa-3-38
Rpa-3-39
Rpa-3-40
Nwi-1-48
Nwi-1-49
Nwi-1-50
Nha-1-47
Nha-1-48
Nha-1-49
Nha-1-50
Nha-1-51
Nha-1-52
Nha-1-53
Nha-1-54
Nha-1-55
Nwi-1-51
Nha-1-56
Nha-1-57
Nha-1-58
Nha-1-59
Nha-1-60

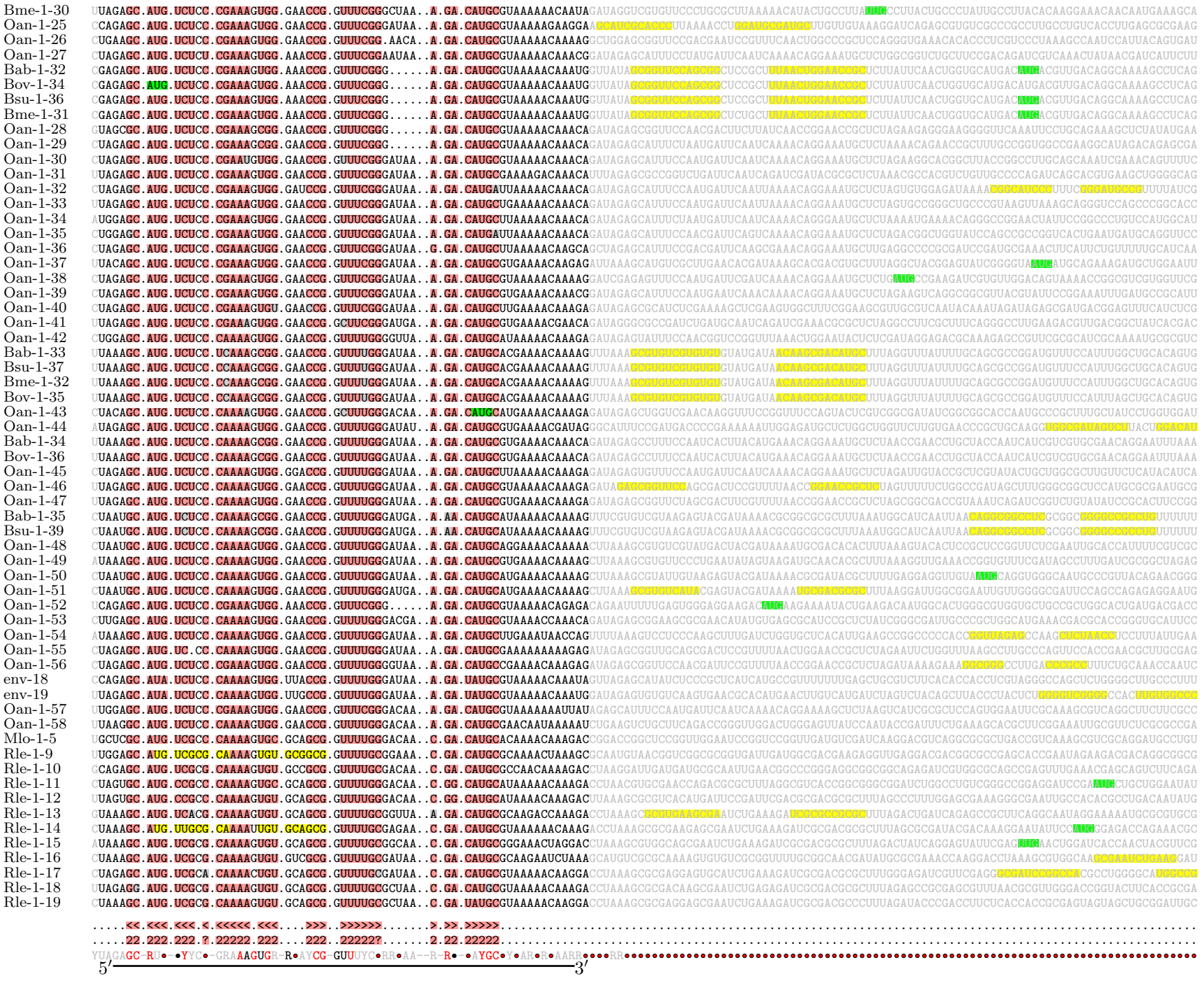

Nha-1-61
Nha-1-62
Nha-1-63
Nha-1-64
Nha-1-65
Nha-1-66
Nha-1-67
Nha-1-68
Nha-1-69
Nha-1-70
Nha-1-71
Nha-1-72
Nha-1-73
Nha-1-74
Nha-1-75
Nwi-1-52
Nwi-1-53
Nha-1-76
Nha-1-77
Nha-1-78
Nsp-1-50
Nsp-1-51
Nsp-1-52
Nsp-1-53
Nsp-1-54
Nwi-1-54
Nsp-1-55
Nsp-1-56
Nha-1-79
Nha-1-80
Nha-1-81
Rpa-3-41
Nha-1-82
Nha-1-83
Nha-1-84
Nha-1-85
Nha-1-86
Nha-1-87
Nwi-1-55
Nwi-1-56
Nwi-1-57
Nwi-1-58
Nwi-1-59
Nwi-1-60
Nwi-1-61
Nwi-1-62
Nwi-1-63
Nwi-1-64
Nwi-1-65
Nwi-1-66
Nwi-1-67
Nwi-1-68
Nwi-1-69
Nwi-1-70
Nwi-1-71
Nwi-1-72
env-55
env-56
env-57
Nha-1-88
Nha-1-89
Nwi-1-73

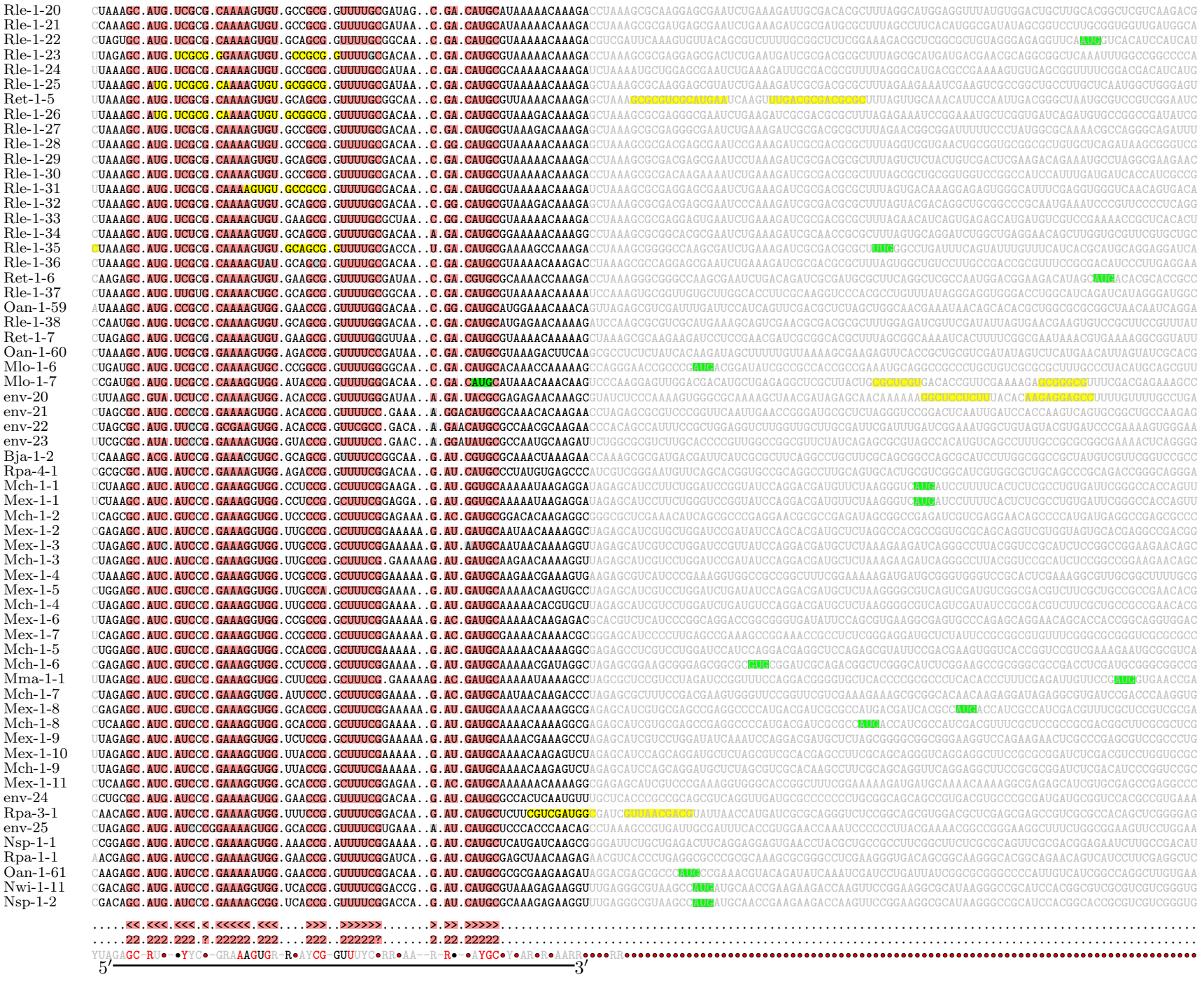


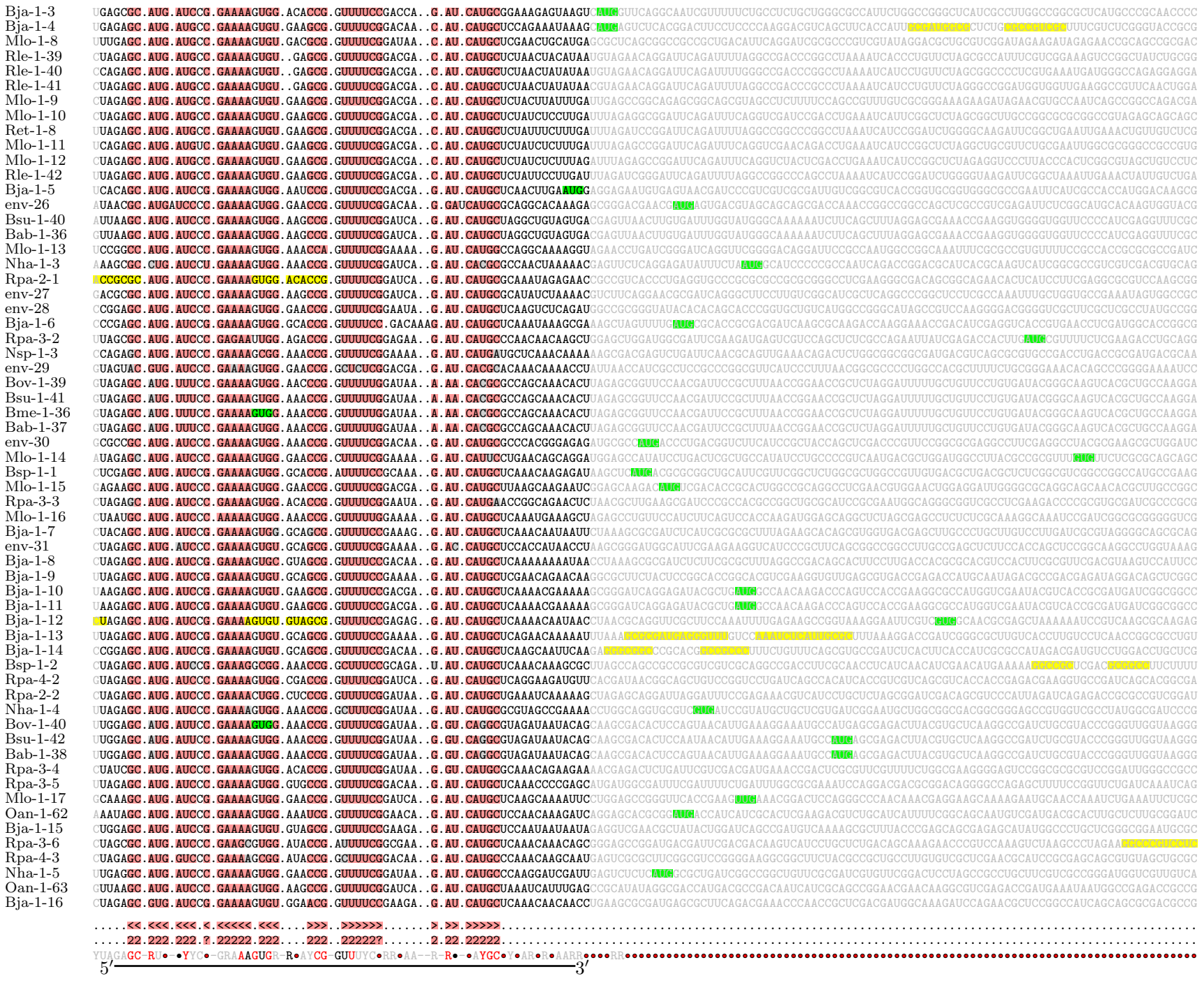
alignment positions 100 · · · 259


Bov-1-4
Bsu-1-4
Bab-1-3
Oan-1-7
Oan-1-8
Bab-1-4
Bov-1-5
Oan-1-9
Bov-1-6
Oan-1-10
Bov-1-7
Bsu-1-6
Bab-1-5
Bme-1-7
Bab-1-6
Bov-1-8
Bsu-1-7
Oan-1-11
Bab-1-7
Bme-1-8
Bab-1-8
Bov-1-10
Bsu-1-9
Bme-1-9
Bme-1-10
Bab-1-9
Bab-1-10
Bov-1-11
Bsu-1-11
Bov-1-12
Bov-1-13
Bsu-1-12
Bab-1-11
Oan-1-12
Bme-1-13
Bab-1-12
Bsu-1-13
Bov-1-14
Bab-1-13
Bme-1-14
Bab-1-14
Bab-1-15
Bov-1-17
Bsu-1-17
Bab-1-16
Bab-1-17
Bov-1-18
Bab-1-18
Oan-1-13
Bov-1-20
Bsu-1-20
Bab-1-19
Bab-1-20
Bov-1-21
Bsu-1-21
Bme-1-19
Oan-1-14
Bov-1-22
Bab-1-21
Bab-1-22
Bov-1-23
5′
3′


Oan-1-15
Bab-1-23
Bab-1-24
Bab-1-25
Bab-1-26
Bov-1-25
Bsu-1-25
Bme-1-24
Bme-1-25
Bsu-1-26
Bov-1-26
Bsu-1-27
Bsu-1-28
Bsu-1-29
Bsu-1-30
Bov-1-27
Bov-1-28
Bme-2-26
Bab-1-27
Bov-1-29
Bsu-1-31
Oan-1-16
env-14
env-15
Nha-1-2
Mlo-1-2
env-16
env-17
Bja-1-1
Oan-1-17
Rle-1-2
Mlo-1-3
Bab-1-28
Bsu-1-32
Bme-1-27
Oan-1-18
Bov-1-31
Bsu-1-33
Bab-1-29
Rle-1-3
Rle-1-4
Rle-1-5
Ret-1-2
Rle-1-6
Ret-1-3
Ret-1-4
Rle-1-7
Sme-1-1
Rle-1-8
Oan-1-19
Mlo-1-4
Oan-1-20
Oan-1-21
Oan-1-22
Oan-1-23
Bab-1-30
Bov-1-32
Oan-1-24
Bab-1-31
Bov-1-33
Bsu-1-35
5′
3′

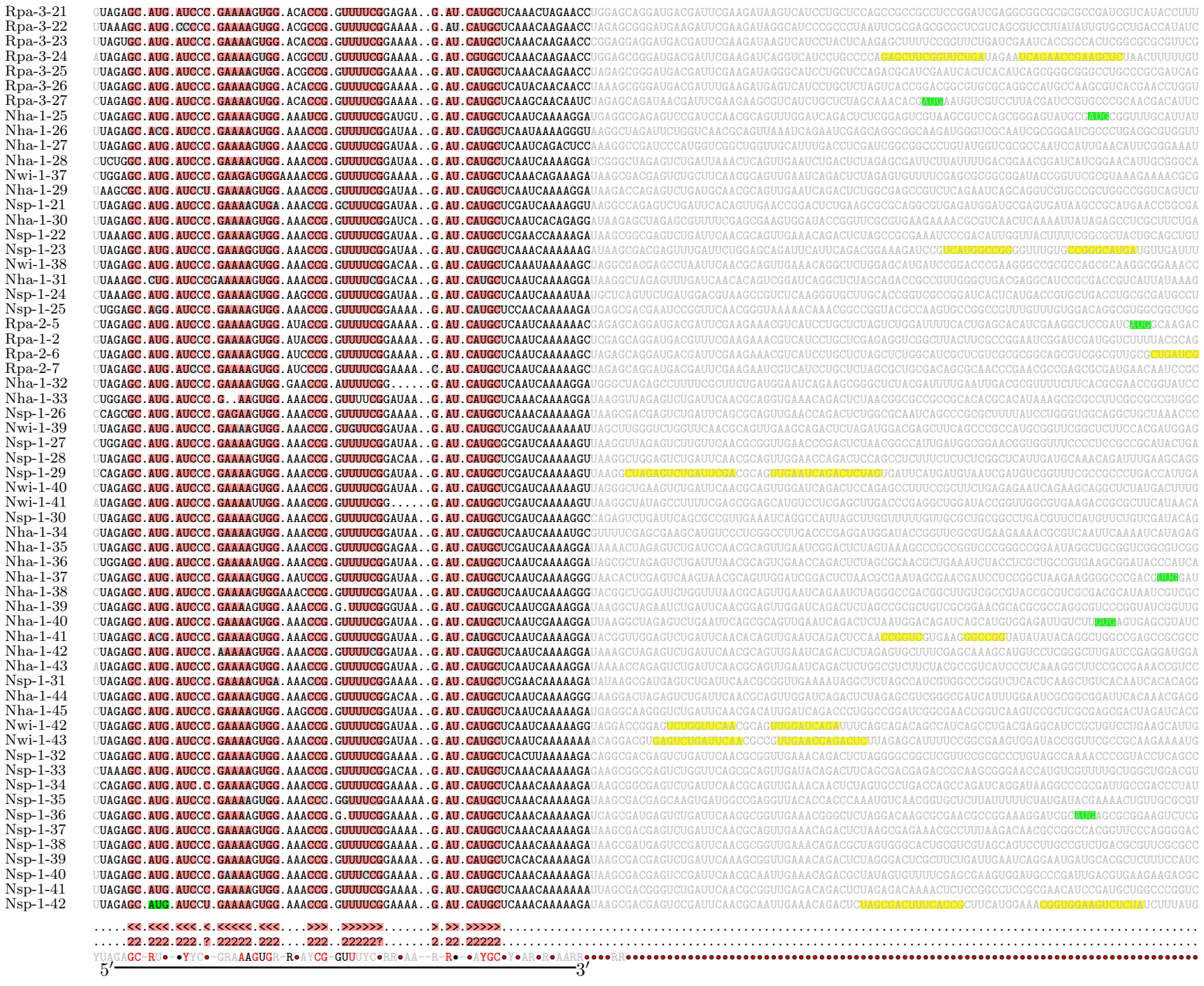

Bme-1-30
Oan-1-25
Oan-1-26
Oan-1-27
Bab-1-32
Bov-1-34
Bsu-1-36
Bme-1-31
Oan-1-28
Oan-1-29
Oan-1-30
Oan-1-31
Oan-1-32
Oan-1-33
Oan-1-34
Oan-1-35
Oan-1-36
Oan-1-37
Oan-1-38
Oan-1-39
Oan-1-40
Oan-1-41
Oan-1-42
Bab-1-33
Bsu-1-37
Bme-1-32
Bov-1-35
Oan-1-43
Oan-1-44
Bab-1-34
Bov-1-36
Oan-1-45
Oan-1-46
Oan-1-47
Bab-1-35
Bsu-1-39
Oan-1-48
Oan-1-49
Oan-1-50
Oan-1-51
Oan-1-52
Oan-1-53
Oan-1-54
Oan-1-55
Oan-1-56
env-18
env-19
Oan-1-57
Oan-1-58
Mlo-1-5
Rle-1-9
Rle-1-10
Rle-1-11
Rle-1-12
Rle-1-13
Rle-1-14
Rle-1-15
Rle-1-16
Rle-1-17
Rle-1-18
Rle-1-19
5′
3′

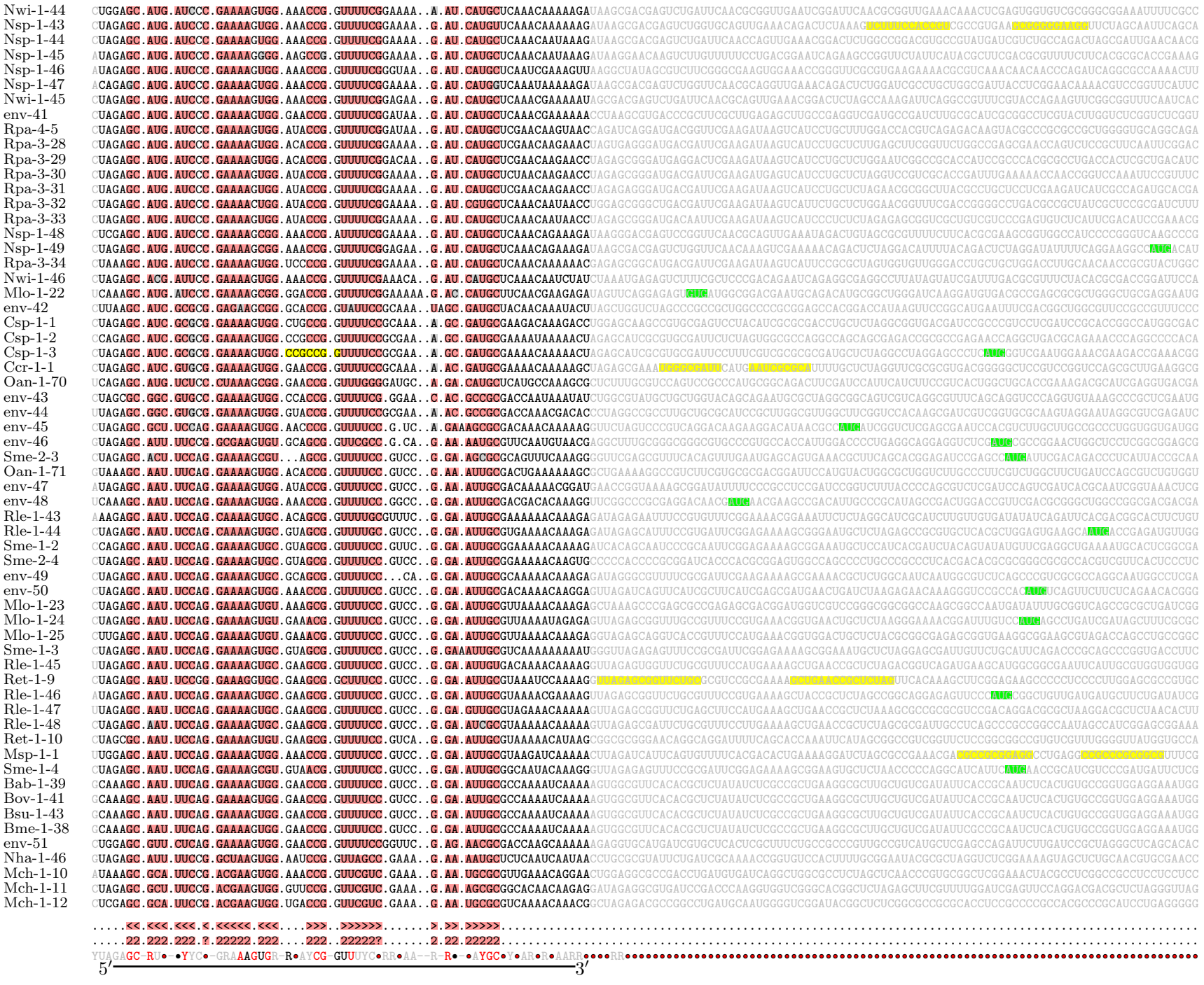

Rle-1-20
Rle-1-21
Rle-1-22
Rle-1-23
Rle-1-24
Rle-1-25
Ret-1-5
Rle-1-26
Rle-1-27
Rle-1-28
Rle-1-29
Rle-1-30
Rle-1-31
Rle-1-32
Rle-1-33
Rle-1-34
Rle-1-35
Rle-1-36
Ret-1-6
Rle-1-37
Oan-1-59
Rle-1-38
Ret-1-7
Oan-1-60
Mlo-1-6
Mlo-1-7
env-20
env-21
env-22
env-23
Bja-1-2
Rpa-4-1
Mch-1-1
Mex-1-1
Mch-1-2
Mex-1-2
Mex-1-3
Mch-1-3
Mex-1-4
Mex-1-5
Mch-1-4
Mex-1-6
Mex-1-7
Mch-1-5
Mch-1-6
Mma-1-1
Mch-1-7
Mex-1-8
Mch-1-8
Mex-1-9
Mex-1-10
Mch-1-9
Mex-1-11
env-24
Rpa-3-1
env-25
Nsp-1-1
Rpa-1-1
Oan-1-61
Nwi-1-11
Nsp-1-2
5′
3′


Bja-1-3
Bja-1-4
Mlo-1-8
Rle-1-39
Rle-1-40
Rle-1-41
Mlo-1-9
Mlo-1-10
Ret-1-8
Mlo-1-11
Mlo-1-12
Rle-1-42
Bja-1-5
env-26
Bsu-1-40
Bab-1-36
Mlo-1-13
Nha-1-3
Rpa-2-1
env-27
env-28
Bja-1-6
Rpa-3-2
Nsp-1-3
env-29
Bov-1-39
Bsu-1-41
Bme-1-36
Bab-1-37
env-30
Mlo-1-14
Bsp-1-1
Mlo-1-15
Rpa-3-3
Mlo-1-16
Bja-1-7
env-31
Bja-1-8
Bja-1-9
Bja-1-10
Bja-1-11
Bja-1-12
Bja-1-13
Bja-1-14
Bsp-1-2
Rpa-4-2
Rpa-2-2
Nha-1-4
Bov-1-40
Bsu-1-42
Bab-1-38
Rpa-3-4
Rpa-3-5
Mlo-1-17
Oan-1-62
Bja-1-15
Rpa-3-6
Rpa-4-3
Nha-1-5
Oan-1-63
Bja-1-16
5′
3′

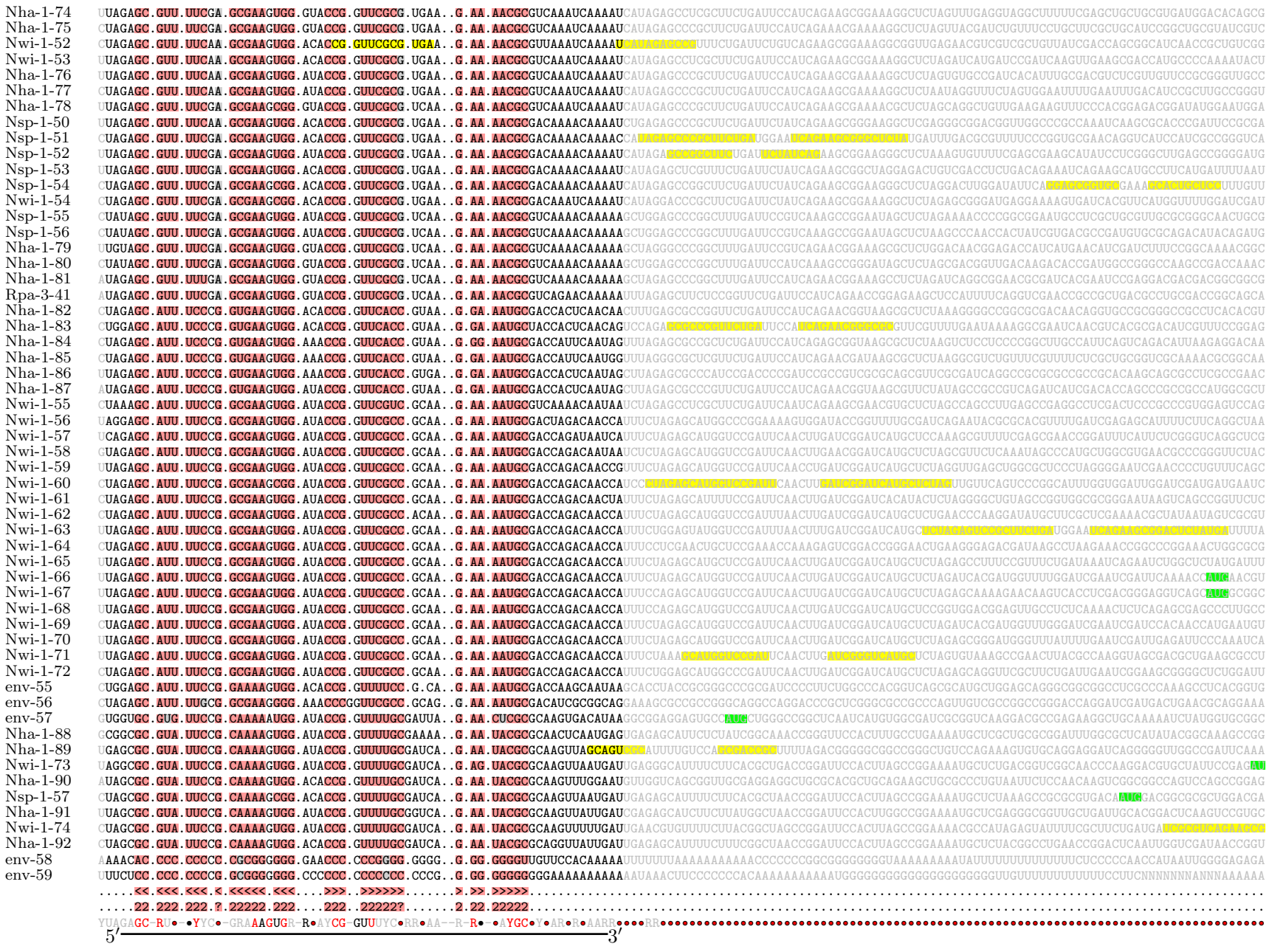
Bja-1-17
Bja-1-18
Bja-1-19
Bja-1-20
Bja-1-21
Bja-1-22
Bja-1-23
Bja-1-24
Bsp-2-2
Bsp-1-3
Bja-1-25
Bja-1-26
Rpa-3-7
env-32
Nwi-1-12
Rpa-3-8
Mlo-1-18
Nwi-1-13
Nha-1-6
Nha-1-7
Mlo-1-19
Oan-1-64
Bsp-1-4
env-33
env-34
Mlo-1-20
Nha-1-8
Nwi-1-14
Rpa-3-9
Nha-1-9
Nwi-1-15
Nsp-1-4
Rpa-4-4
Nsp-1-5
Rpa-3-10
Rpa-3-11
Oan-1-65
Rpa-3-12
Nwi-1-16
Nwi-1-17
Nha-1-10
Nha-1-11
Bsp-1-5
env-35
env-36
Rpa-3-13
Nha-1-12
Nsp-1-6
env-37
Nwi-1-18
Nsp-1-7
Nwi-1-19
Bsp-1-6
Nwi-1-20
Nwi-1-21
Nsp-1-8
Nwi-1-22
env-38
Oan-1-66
Oan-1-67
Oan-1-68
5′
3′

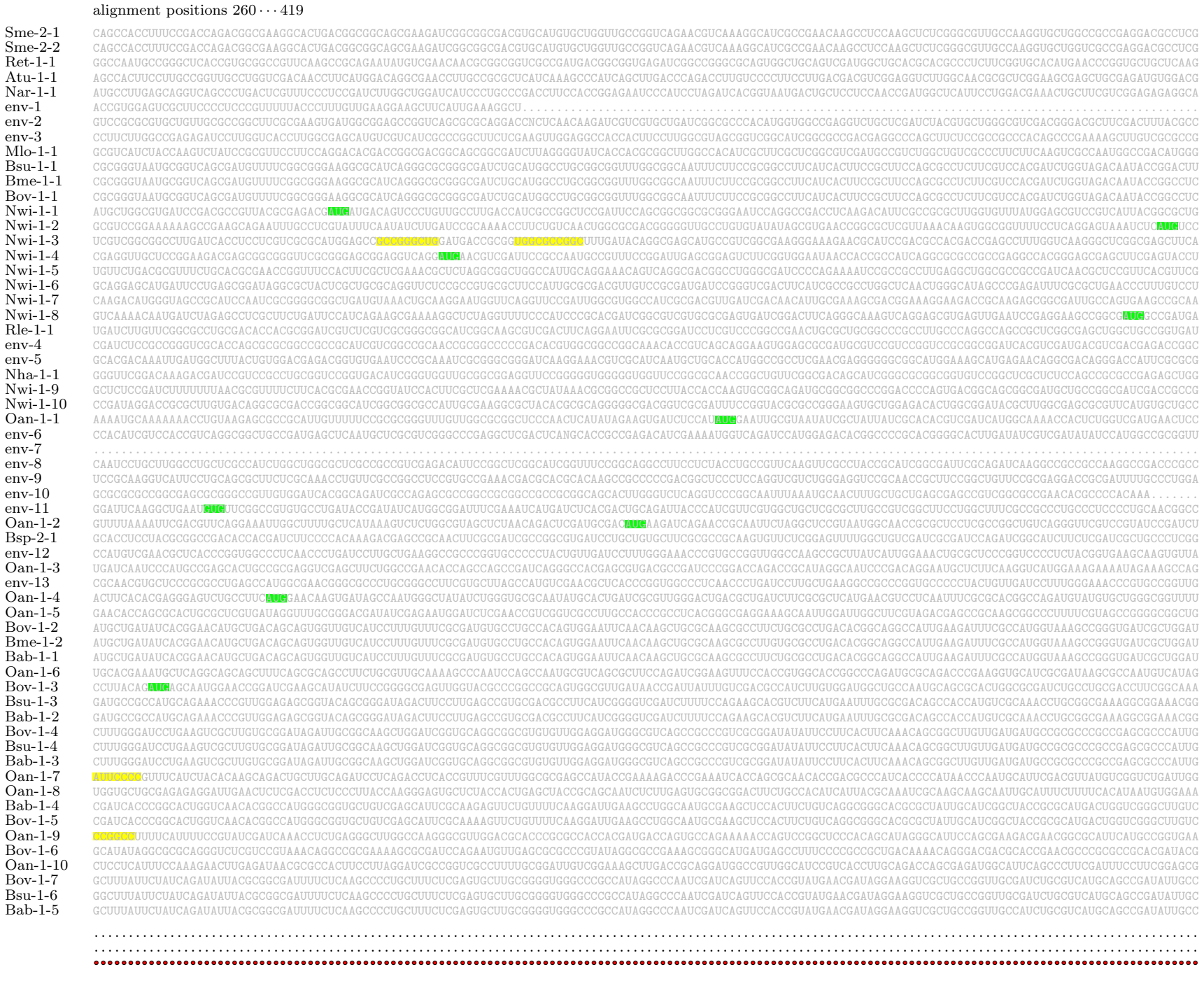
Oan-1-69
Nwi-1-23
Nwi-1-24
Nha-1-13
Nha-1-14
Nwi-1-25
Nwi-1-26
Nsp-1-9
Rpa-3-14
Nsp-1-10
Nsp-1-11
Nsp-1-12
env-39
Nha-1-15
Nha-1-16
Rpa-3-15
env-40
Rpa-3-16
Nwi-1-27
Rpa-3-17
Rpa-3-18
Nha-1-17
Rpa-3-19
Nha-1-18
Nsp-1-13
Nsp-1-14
Nwi-1-28
Rpa-2-3
Nha-1-19
Nwi-1-29
Nsp-1-15
Nsp-1-16
Nwi-1-30
Nwi-1-31
Nha-1-20
Nha-1-21
Nwi-1-32
Nha-1-22
Nwi-1-33
Nwi-1-34
Bsp-1-7
Bsp-1-8
Bsp-1-9
Bsp-1-10
Bsp-1-11
Bsp-1-12
Bsp-1-13
Bsp-1-14
Bsp-1-15
Bsp-1-16
Nwi-1-35
Nha-1-23
Nsp-1-17
Nwi-1-36
Mlo-1-21
Nha-1-24
Nsp-1-18
Nsp-1-19
Rpa-3-20
Nsp-1-20
Rpa-2-4
5′
3′

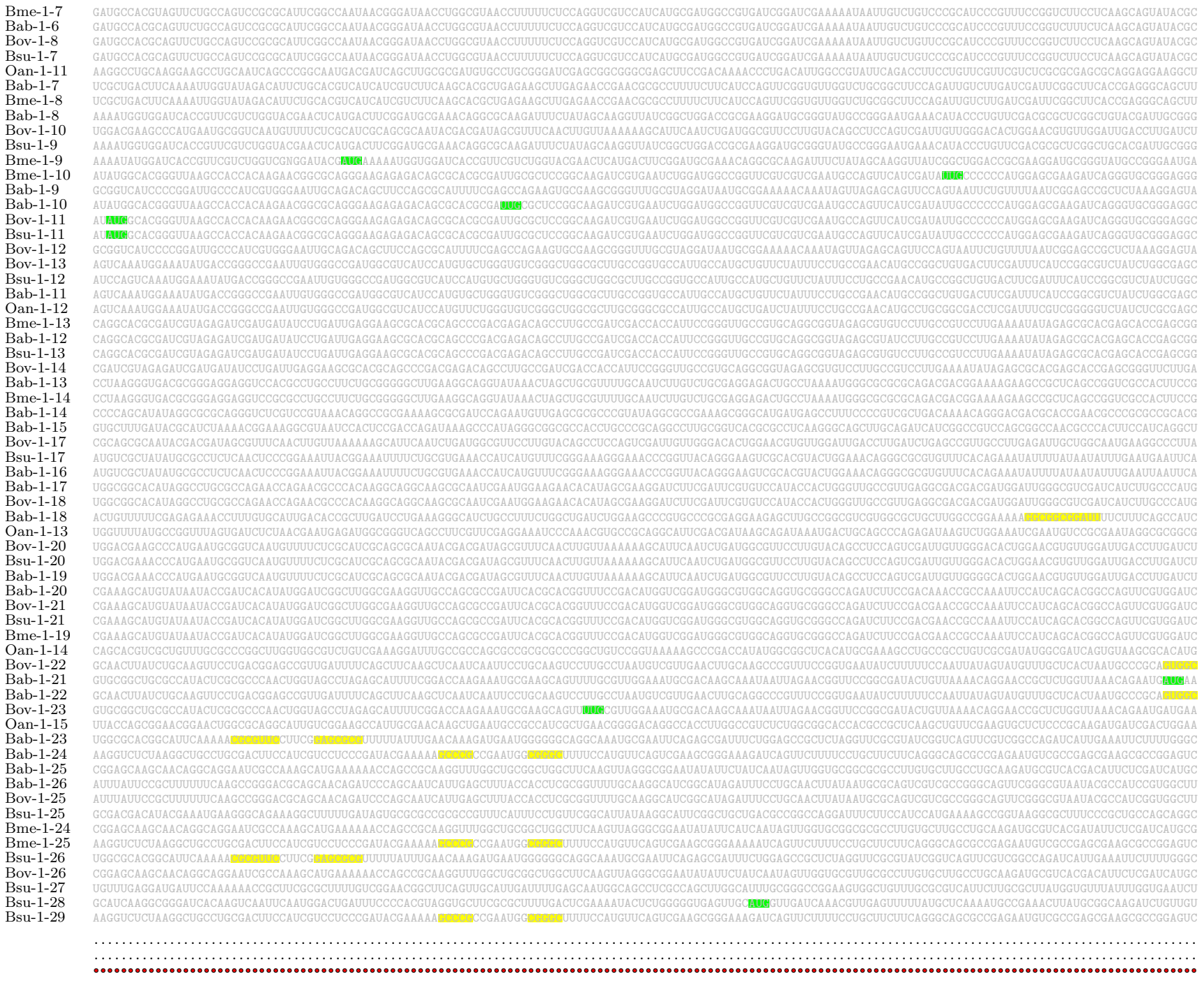

Rpa-3-21
Rpa-3-22
Rpa-3-23
Rpa-3-24
Rpa-3-25
Rpa-3-26
Rpa-3-27
Nha-1-25
Nha-1-26
Nha-1-27
Nha-1-28
Nwi-1-37
Nha-1-29
Nsp-1-21
Nha-1-30
Nsp-1-22
Nsp-1-23
Nwi-1-38
Nha-1-31
Nsp-1-24
Nsp-1-25
Rpa-2-5
Rpa-1-2
Rpa-2-6
Rpa-2-7
Nha-1-32
Nha-1-33
Nsp-1-26
Nwi-1-39
Nsp-1-27
Nsp-1-28
Nsp-1-29
Nwi-1-40
Nwi-1-41
Nsp-1-30
Nha-1-34
Nha-1-35
Nha-1-36
Nha-1-37
Nha-1-38
Nha-1-39
Nha-1-40
Nha-1-41
Nha-1-42
Nha-1-43
Nsp-1-31
Nha-1-44
Nha-1-45
Nwi-1-42
Nwi-1-43
Nsp-1-32
Nsp-1-33
Nsp-1-34
Nsp-1-35
Nsp-1-36
Nsp-1-37
Nsp-1-38
Nsp-1-39
Nsp-1-40
Nsp-1-41
Nsp-1-42
5′
3′

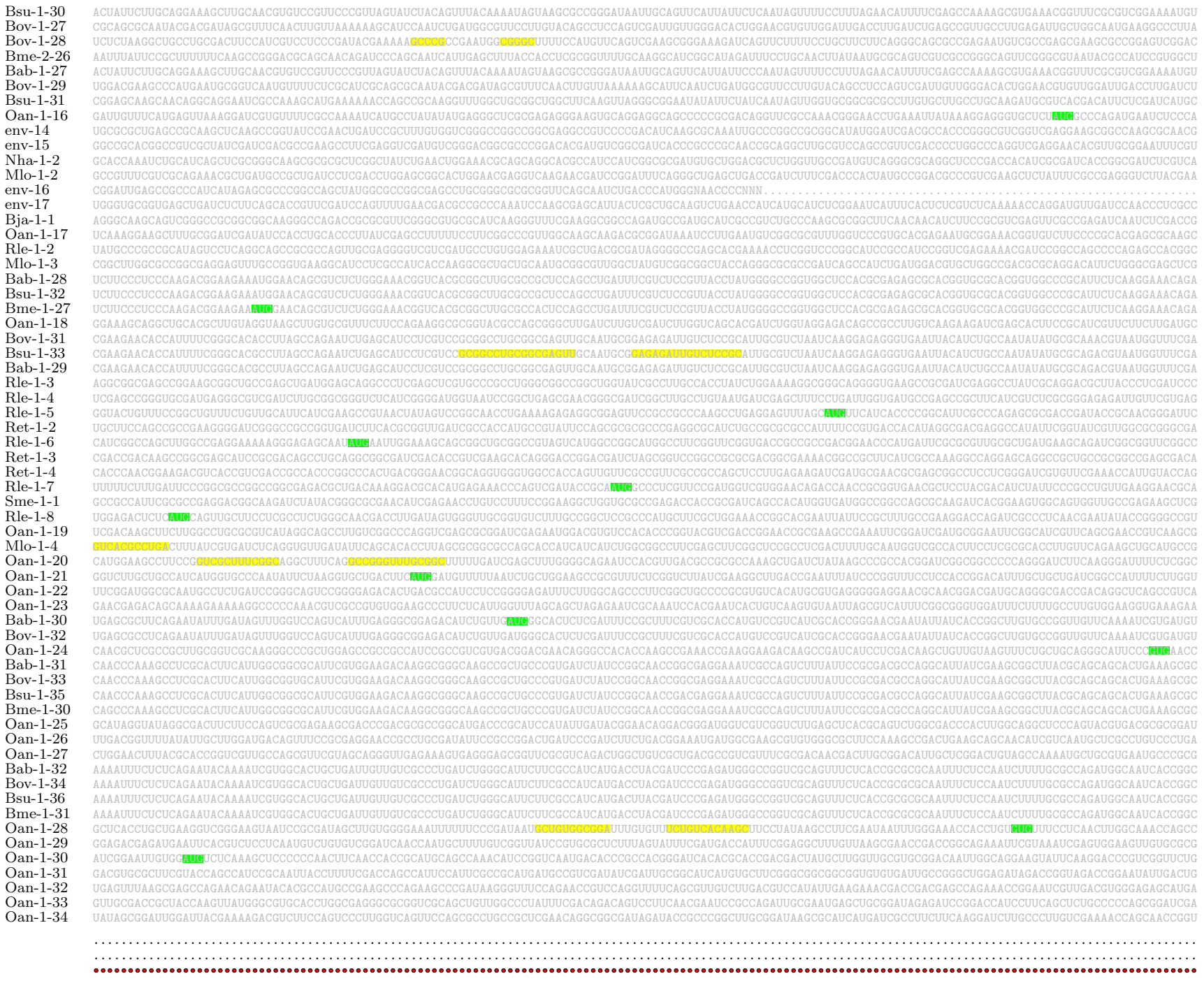

Nwi-1-44
Nsp-1-43
Nsp-1-44
Nsp-1-45
Nsp-1-46
Nsp-1-47
Nwi-1-45
env-41
Rpa-4-5
Rpa-3-28
Rpa-3-29
Rpa-3-30
Rpa-3-31
Rpa-3-32
Rpa-3-33
Nsp-1-48
Nsp-1-49
Rpa-3-34
Nwi-1-46
Mlo-1-22
env-42
Csp-1-1
Csp-1-2
Csp-1-3
Ccr-1-1
Oan-1-70
env-43
env-44
env-45
env-46
Sme-2-3
Oan-1-71
env-47
env-48
Rle-1-43
Rle-1-44
Sme-1-2
Sme-2-4
env-49
env-50
Mlo-1-23
Mlo-1-24
Mlo-1-25
Sme-1-3
Rle-1-45
Ret-1-9
Rle-1-46
Rle-1-47
Rle-1-48
Ret-1-10
Msp-1-1
Sme-1-4
Bab-1-39
Bov-1-41
Bsu-1-43
Bme-1-38
env-51
Nha-1-46
Mch-1-10
Mch-1-11
Mch-1-12
5′
3′

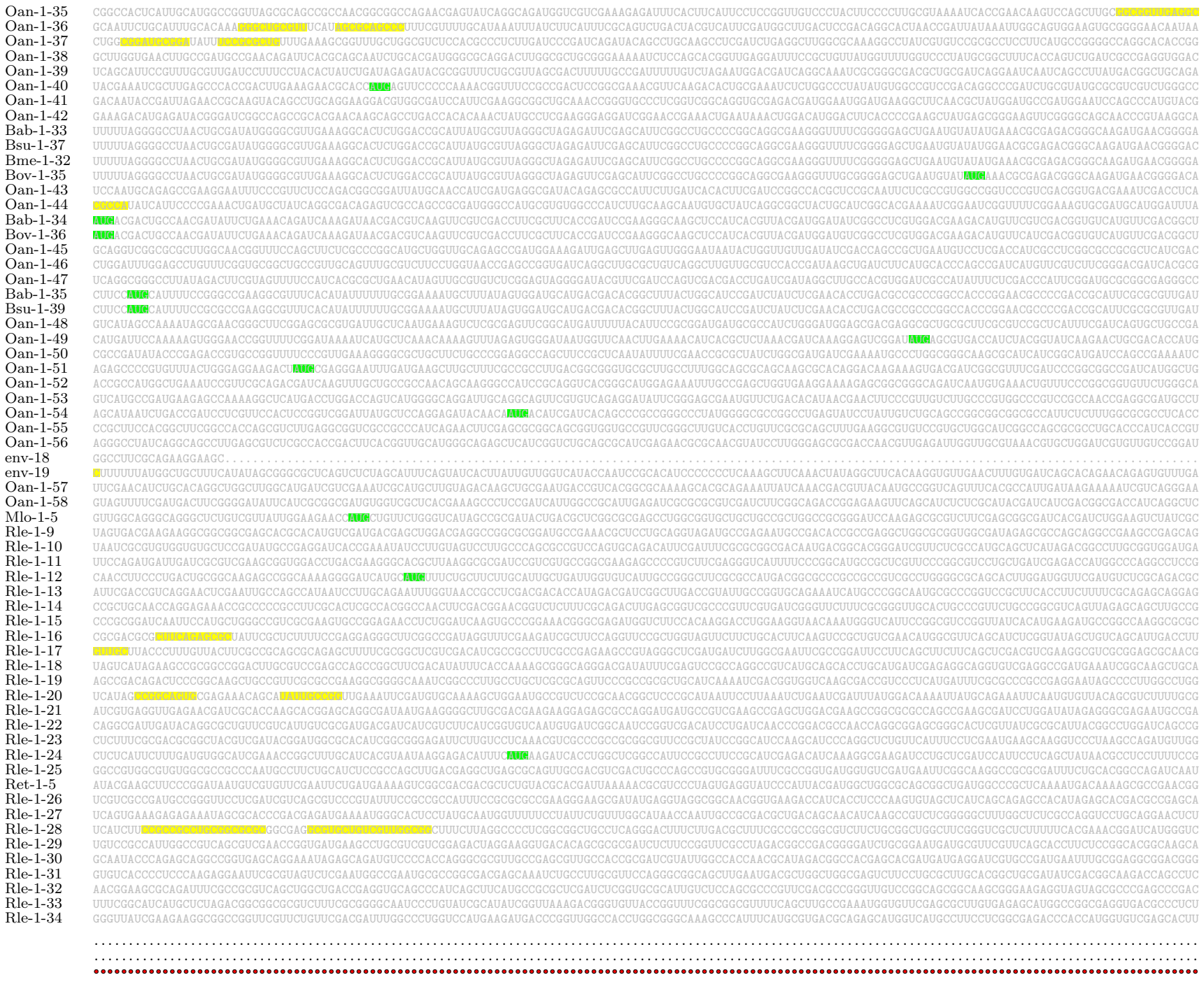
Mex-1-12
Mch-1-13
Mex-1-13
Mex-1-14
Mex-1-15
Mex-1-16
Mch-1-14
Mex-1-17
Mch-1-15
Mch-1-16
Mex-1-18
Mex-1-19
Mex-1-20
env-52
Csp-1-4
Csp-1-5
Csp-1-6
Csp-1-7
Csp-1-8
env-53
env-54
Nwi-1-47
Rpa-3-35
Bja-1-27
Rpa-4-6
Rpa-3-36
Rpa-3-37
Rpa-3-38
Rpa-3-39
Rpa-3-40
Nwi-1-48
Nwi-1-49
Nwi-1-50
Nha-1-47
Nha-1-48
Nha-1-49
Nha-1-50
Nha-1-51
Nha-1-52
Nha-1-53
Nha-1-54
Nha-1-55
Nwi-1-51
Nha-1-56
Nha-1-57
Nha-1-58
Nha-1-59
Nha-1-60
Nha-1-61
Nha-1-62
Nha-1-63
Nha-1-64
Nha-1-65
Nha-1-66
Nha-1-67
Nha-1-68
Nha-1-69
Nha-1-70
Nha-1-71
Nha-1-72
Nha-1-73
5′
3′

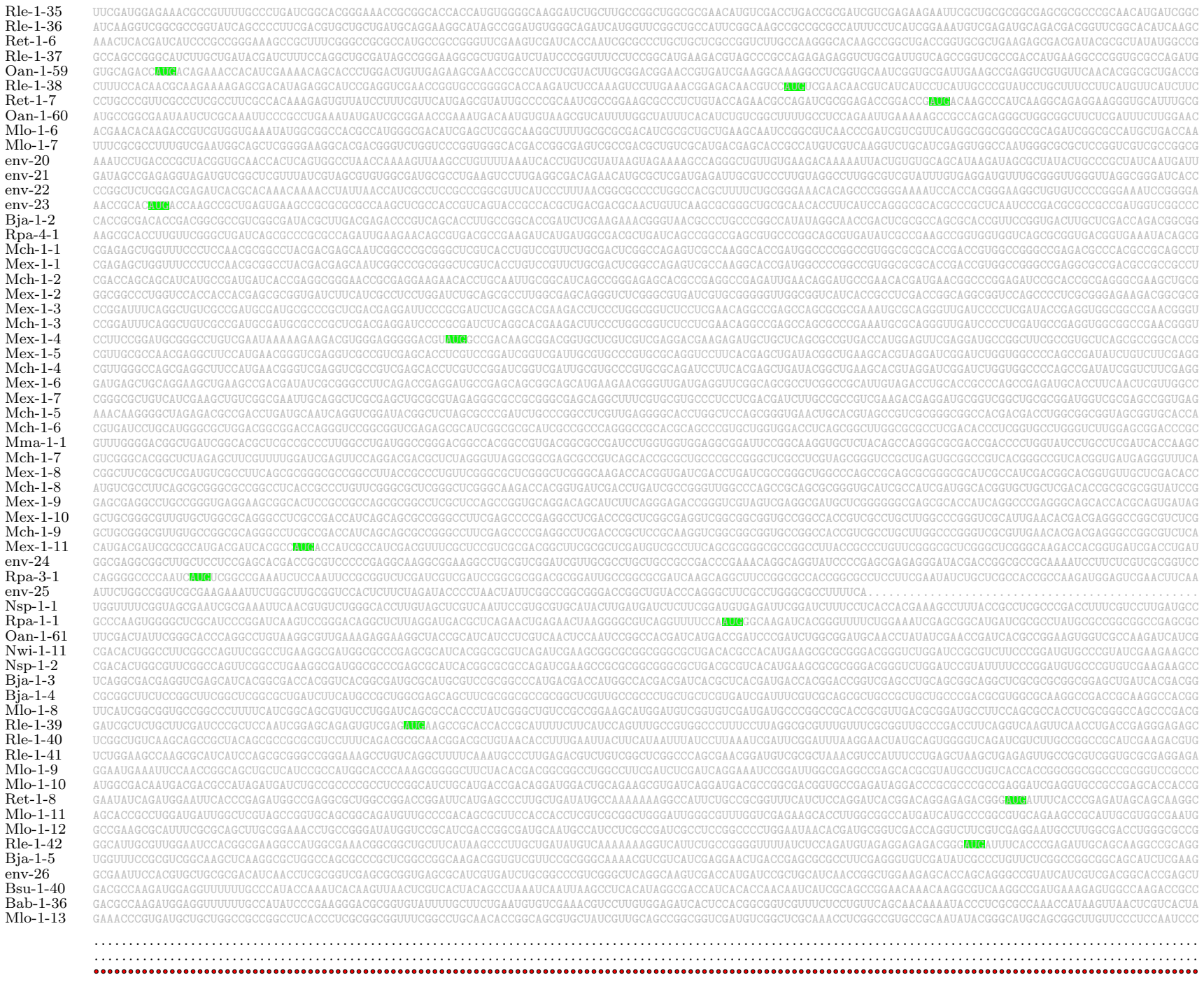

Nha-1-74
Nha-1-75
Nwi-1-52
Nwi-1-53
Nha-1-76
Nha-1-77
Nha-1-78
Nsp-1-50
Nsp-1-51
Nsp-1-52
Nsp-1-53
Nsp-1-54
Nwi-1-54
Nsp-1-55
Nsp-1-56
Nha-1-79
Nha-1-80
Nha-1-81
Rpa-3-41
Nha-1-82
Nha-1-83
Nha-1-84
Nha-1-85
Nha-1-86
Nha-1-87
Nwi-1-55
Nwi-1-56
Nwi-1-57
Nwi-1-58
Nwi-1-59
Nwi-1-60
Nwi-1-61
Nwi-1-62
Nwi-1-63
Nwi-1-64
Nwi-1-65
Nwi-1-66
Nwi-1-67
Nwi-1-68
Nwi-1-69
Nwi-1-70
Nwi-1-71
Nwi-1-72
env-55
env-56
env-57
Nha-1-88
Nha-1-89
Nwi-1-73
Nha-1-90
Nsp-1-57
Nha-1-91
Nwi-1-74
Nha-1-92
env-58
env-59
5′
3′

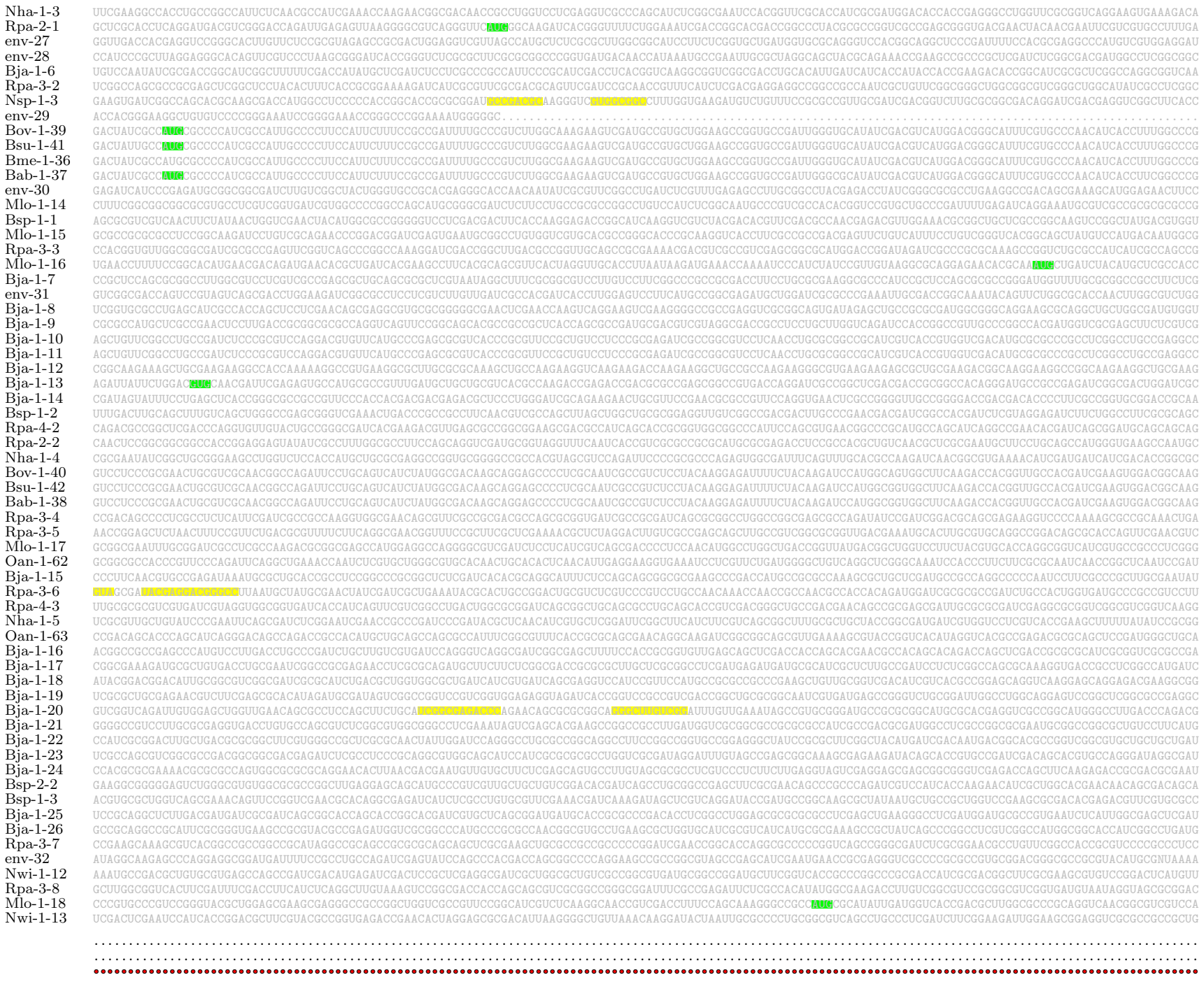

alignment positions 260 · · · 419
Sme-2-1
Sme-2-2
Ret-1-1
Atu-1-1
Nar-1-1
env-1
env-2
env-3
Mlo-1-1
Bsu-1-1
Bme-1-1
Bov-1-1
Nwi-1-1
Nwi-1-2
Nwi-1-3
Nwi-1-4
Nwi-1-5
Nwi-1-6
Nwi-1-7
Nwi-1-8
Rle-1-1
env-4
env-5
Nha-1-1
Nwi-1-9
Nwi-1-10
Oan-1-1
env-6
env-7
env-8
env-9
env-10
env-11
Oan-1-2
Bsp-2-1
env-12
Oan-1-3
env-13
Oan-1-4
Oan-1-5
Bov-1-2
Bme-1-2
Bab-1-1
Oan-1-6
Bov-1-3
Bsu-1-3
Bab-1-2
Bov-1-4
Bsu-1-4
Bab-1-3
Oan-1-7
Oan-1-8
Bab-1-4
Bov-1-5
Oan-1-9
Bov-1-6
Oan-1-10
Bov-1-7
Bsu-1-6
Bab-1-5

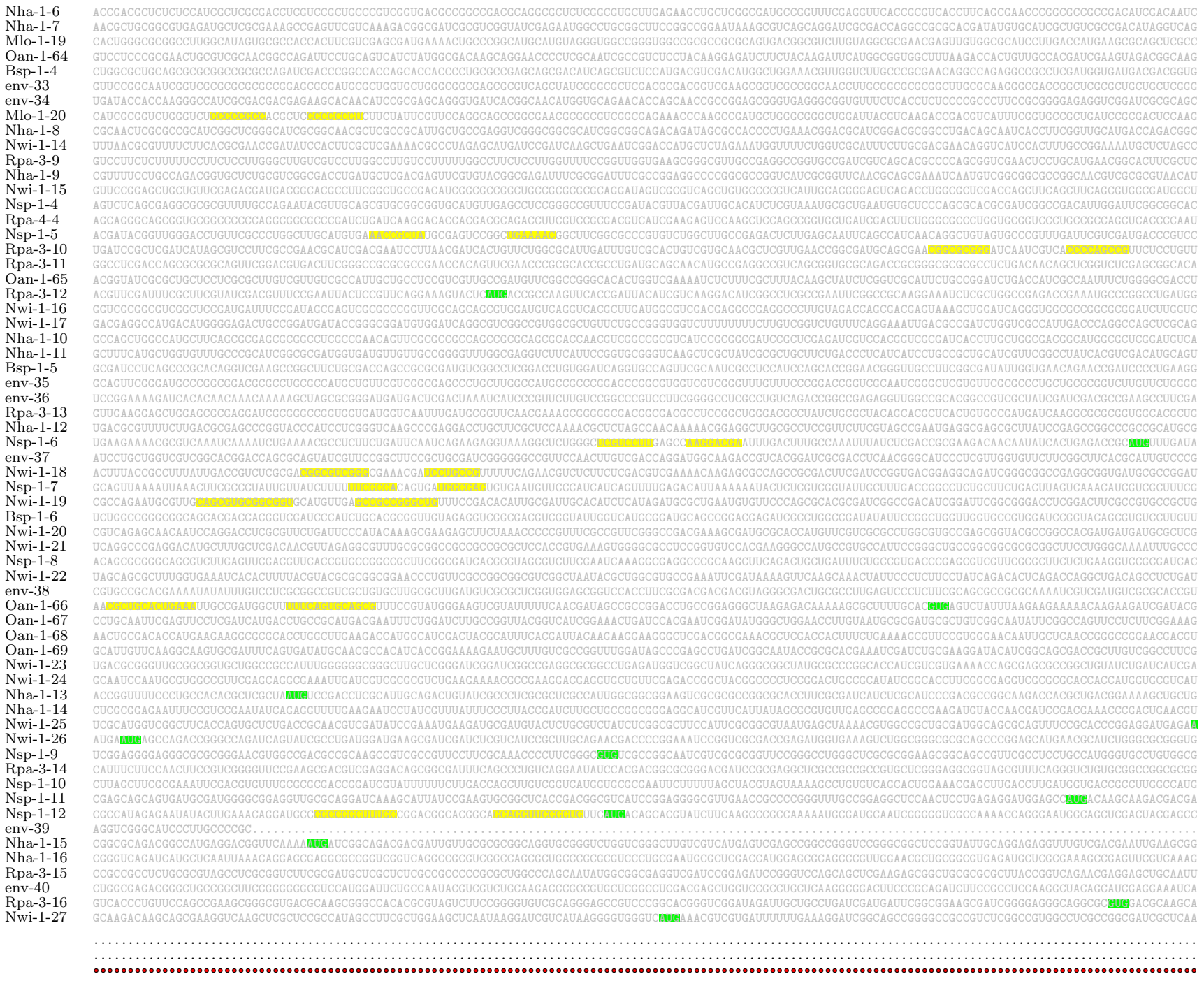

Bme-1-7
Bab-1-6
Bov-1-8
Bsu-1-7
Oan-1-11
Bab-1-7
Bme-1-8
Bab-1-8
Bov-1-10
Bsu-1-9
Bme-1-9
Bme-1-10
Bab-1-9
Bab-1-10
Bov-1-11
Bsu-1-11
Bov-1-12
Bov-1-13
Bsu-1-12
Bab-1-11
Oan-1-12
Bme-1-13
Bab-1-12
Bsu-1-13
Bov-1-14
Bab-1-13
Bme-1-14
Bab-1-14
Bab-1-15
Bov-1-17
Bsu-1-17
Bab-1-16
Bab-1-17
Bov-1-18
Bab-1-18
Oan-1-13
Bov-1-20
Bsu-1-20
Bab-1-19
Bab-1-20
Bov-1-21
Bsu-1-21
Bme-1-19
Oan-1-14
Bov-1-22
Bab-1-21
Bab-1-22
Bov-1-23
Oan-1-15
Bab-1-23
Bab-1-24
Bab-1-25
Bab-1-26
Bov-1-25
Bsu-1-25
Bme-1-24
Bme-1-25
Bsu-1-26
Bov-1-26
Bsu-1-27
Bsu-1-28
Bsu-1-29

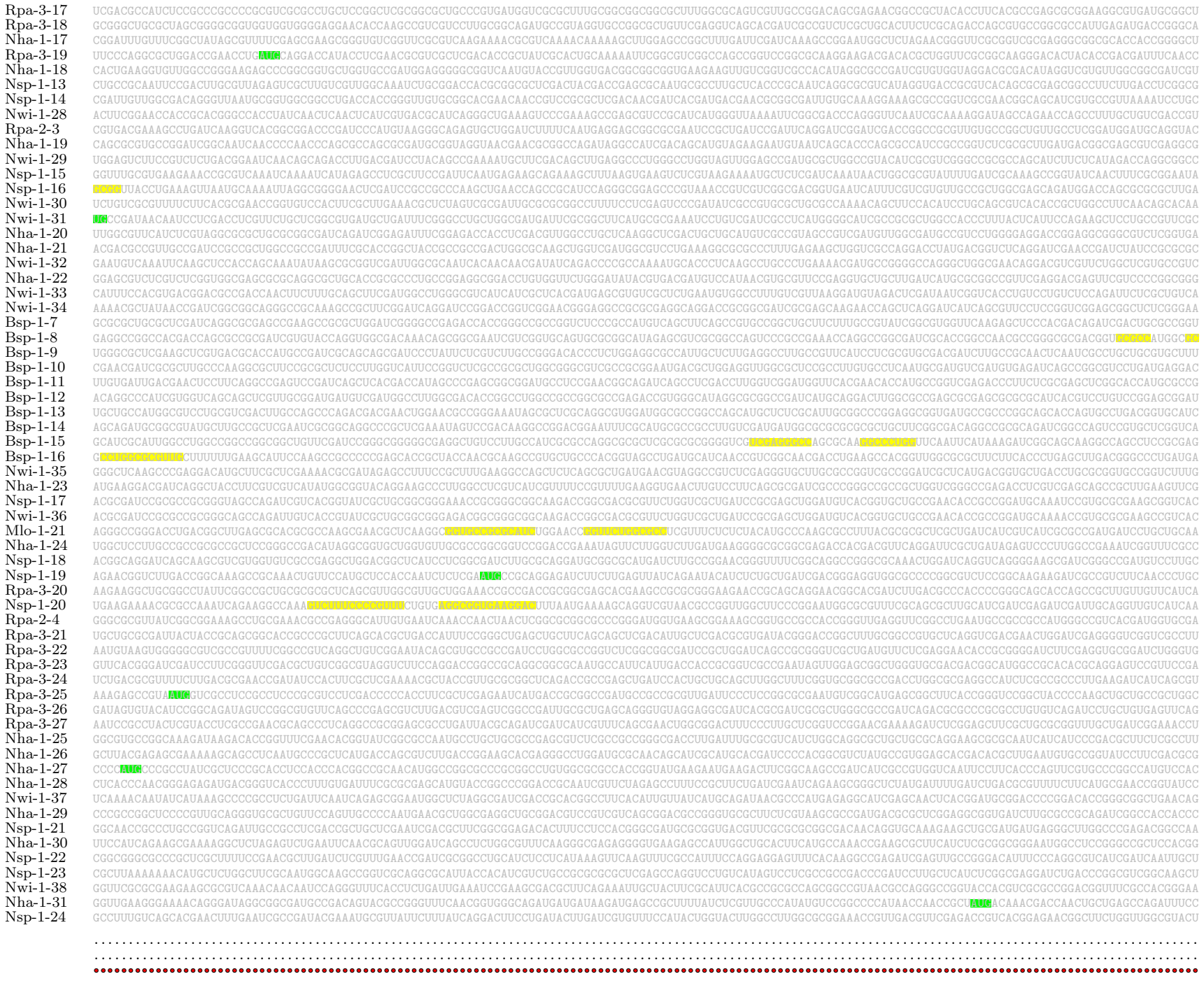

Bsu-1-30
Bov-1-27
Bov-1-28
Bme-2-26
Bab-1-27
Bov-1-29
Bsu-1-31
Oan-1-16
env-14
env-15
Nha-1-2
Mlo-1-2
env-16
env-17
Bja-1-1
Oan-1-17
Rle-1-2
Mlo-1-3
Bab-1-28
Bsu-1-32
Bme-1-27
Oan-1-18
Bov-1-31
Bsu-1-33
Bab-1-29
Rle-1-3
Rle-1-4
Rle-1-5
Ret-1-2
Rle-1-6
Ret-1-3
Ret-1-4
Rle-1-7
Sme-1-1
Rle-1-8
Oan-1-19
Mlo-1-4
Oan-1-20
Oan-1-21
Oan-1-22
Oan-1-23
Bab-1-30
Bov-1-32
Oan-1-24
Bab-1-31
Bov-1-33
Bsu-1-35
Bme-1-30
Oan-1-25
Oan-1-26
Oan-1-27
Bab-1-32
Bov-1-34
Bsu-1-36
Bme-1-31
Oan-1-28
Oan-1-29
Oan-1-30
Oan-1-31
Oan-1-32
Oan-1-33
Oan-1-34

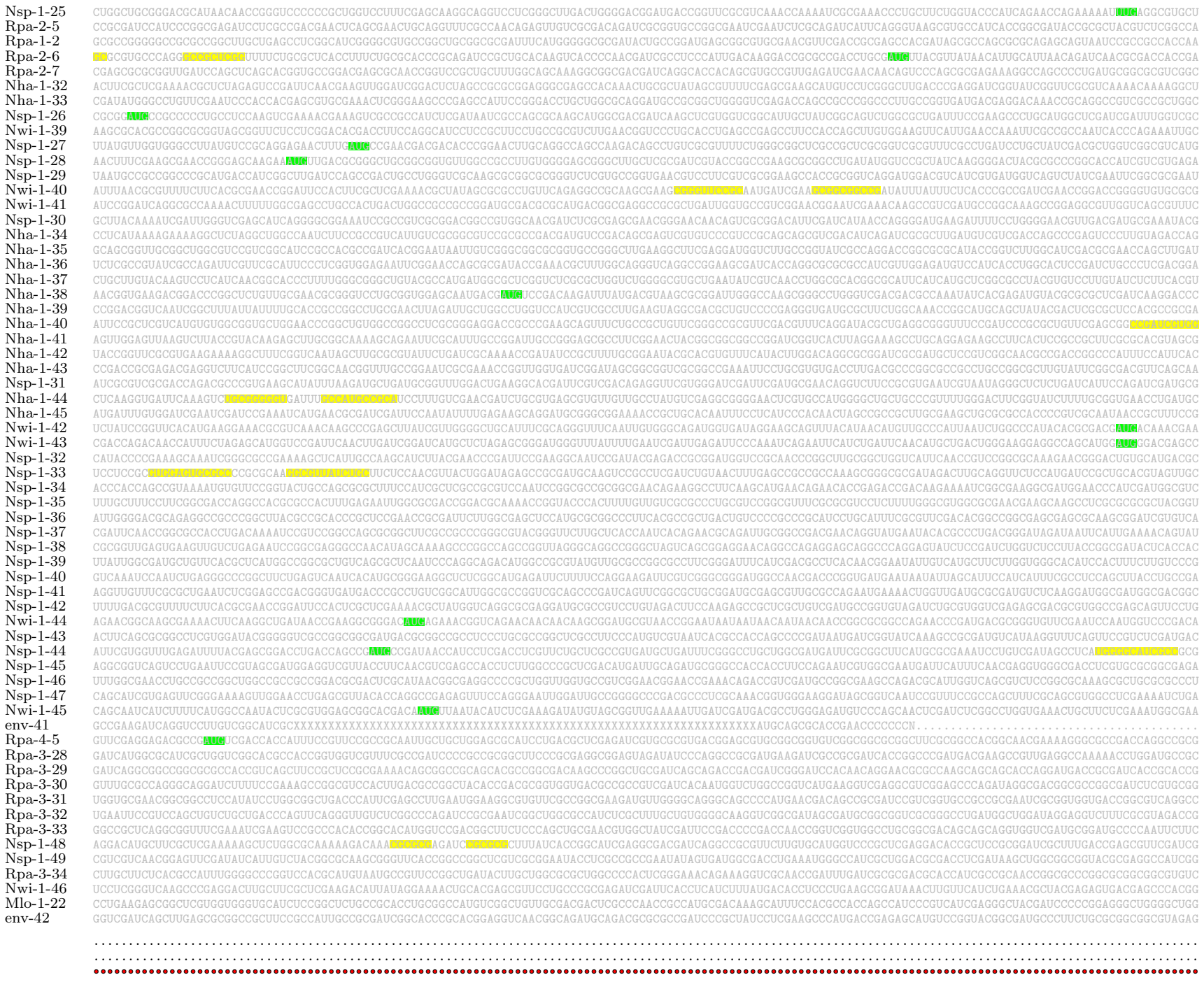

Oan-1-35
Oan-1-36
Oan-1-37
Oan-1-38
Oan-1-39
Oan-1-40
Oan-1-41
Oan-1-42
Bab-1-33
Bsu-1-37
Bme-1-32
Bov-1-35
Oan-1-43
Oan-1-44
Bab-1-34
Bov-1-36
Oan-1-45
Oan-1-46
Oan-1-47
Bab-1-35
Bsu-1-39
Oan-1-48
Oan-1-49
Oan-1-50
Oan-1-51
Oan-1-52
Oan-1-53
Oan-1-54
Oan-1-55
Oan-1-56
env-18
env-19
Oan-1-57
Oan-1-58
Mlo-1-5
Rle-1-9
Rle-1-10
Rle-1-11
Rle-1-12
Rle-1-13
Rle-1-14
Rle-1-15
Rle-1-16
Rle-1-17
Rle-1-18
Rle-1-19
Rle-1-20
Rle-1-21
Rle-1-22
Rle-1-23
Rle-1-24
Rle-1-25
Ret-1-5
Rle-1-26
Rle-1-27
Rle-1-28
Rle-1-29
Rle-1-30
Rle-1-31
Rle-1-32
Rle-1-33
Rle-1-34


Rle-1-35
Rle-1-36
Ret-1-6
Rle-1-37
Oan-1-59
Rle-1-38
Ret-1-7
Oan-1-60
Mlo-1-6
Mlo-1-7
env-20
env-21
env-22
env-23
Bja-1-2
Rpa-4-1
Mch-1-1
Mex-1-1
Mch-1-2
Mex-1-2
Mex-1-3
Mch-1-3
Mex-1-4
Mex-1-5
Mch-1-4
Mex-1-6
Mex-1-7
Mch-1-5
Mch-1-6
Mma-1-1
Mch-1-7
Mex-1-8
Mch-1-8
Mex-1-9
Mex-1-10
Mch-1-9
Mex-1-11
env-24
Rpa-3-1
env-25
Nsp-1-1
Rpa-1-1
Oan-1-61
Nwi-1-11
Nsp-1-2
Bja-1-3
Bja-1-4
Mlo-1-8
Rle-1-39
Rle-1-40
Rle-1-41
Mlo-1-9
Mlo-1-10
Ret-1-8
Mlo-1-11
Mlo-1-12
Rle-1-42
Bja-1-5
env-26
Bsu-1-40
Bab-1-36
Mlo-1-13

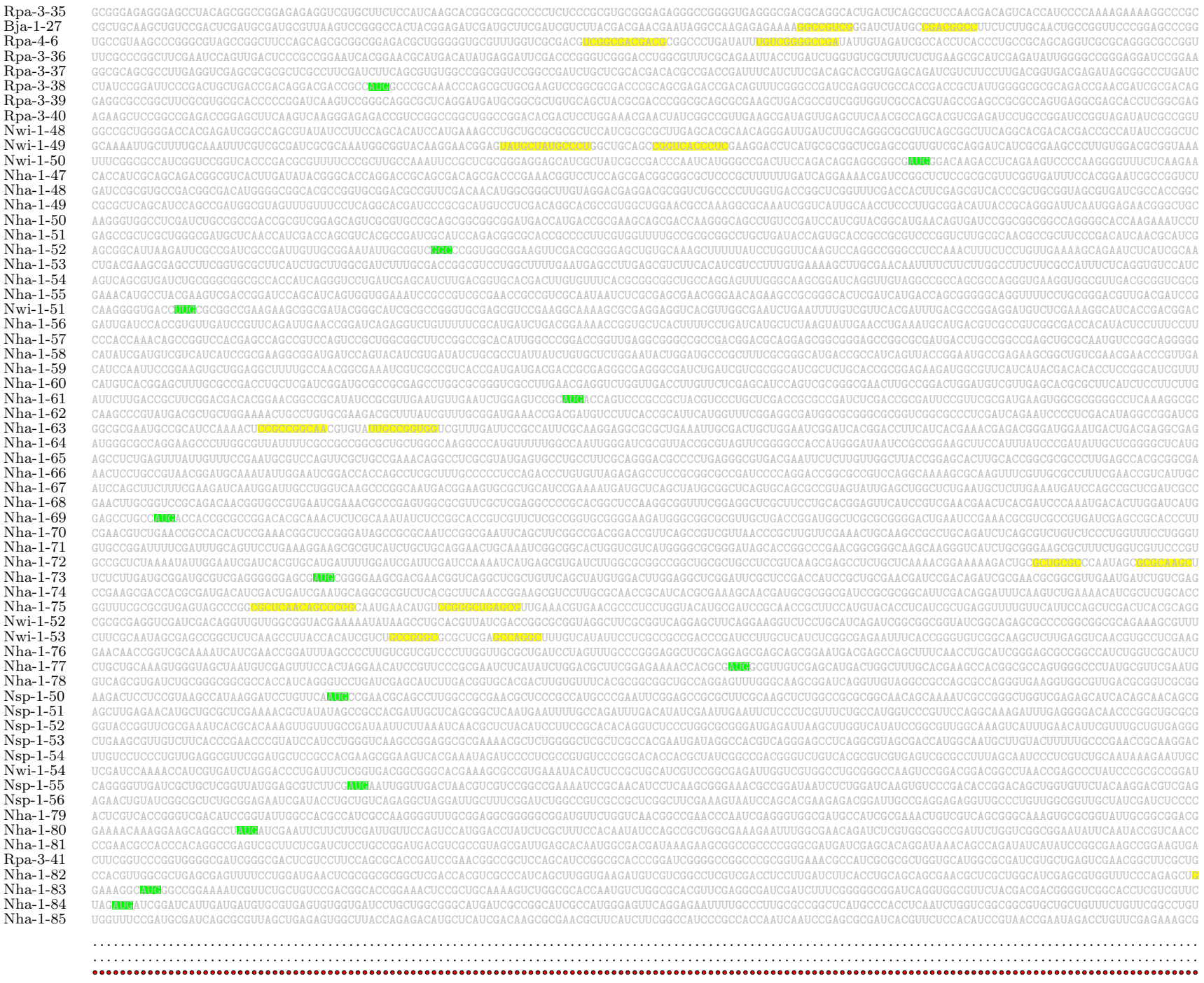

Nha-1-3
Rpa-2-1
env-27
env-28
Bja-1-6
Rpa-3-2
Nsp-1-3
env-29
Bov-1-39
Bsu-1-41
Bme-1-36
Bab-1-37
env-30
Mlo-1-14
Bsp-1-1
Mlo-1-15
Rpa-3-3
Mlo-1-16
Bja-1-7
env-31
Bja-1-8
Bja-1-9
Bja-1-10
Bja-1-11
Bja-1-12
Bja-1-13
Bja-1-14
Bsp-1-2
Rpa-4-2
Rpa-2-2
Nha-1-4
Bov-1-40
Bsu-1-42
Bab-1-38
Rpa-3-4
Rpa-3-5
Mlo-1-17
Oan-1-62
Bja-1-15
Rpa-3-6
Rpa-4-3
Nha-1-5
Oan-1-63
Bja-1-16
Bja-1-17
Bja-1-18
Bja-1-19
Bja-1-20
Bja-1-21
Bja-1-22
Bja-1-23
Bja-1-24
Bsp-2-2
Bsp-1-3
Bja-1-25
Bja-1-26
Rpa-3-7
env-32
Nwi-1-12
Rpa-3-8
Mlo-1-18
Nwi-1-13

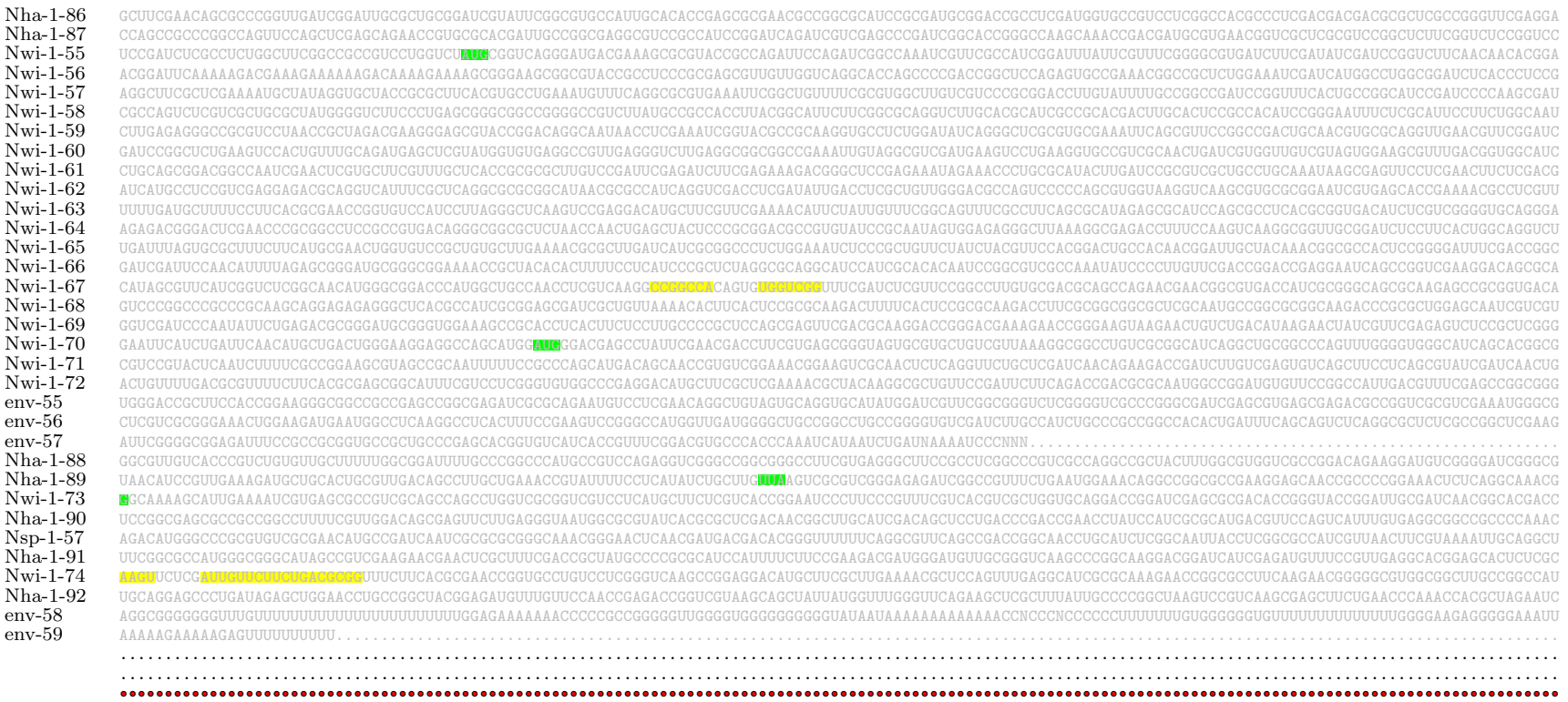

Nha-1-6
Nha-1-7
Mlo-1-19
Oan-1-64
Bsp-1-4
env-33
env-34
Mlo-1-20
Nha-1-8
Nwi-1-14
Rpa-3-9
Nha-1-9
Nwi-1-15
Nsp-1-4
Rpa-4-4
Nsp-1-5
Rpa-3-10
Rpa-3-11
Oan-1-65
Rpa-3-12
Nwi-1-16
Nwi-1-17
Nha-1-10
Nha-1-11
Bsp-1-5
env-35
env-36
Rpa-3-13
Nha-1-12
Nsp-1-6
env-37
Nwi-1-18
Nsp-1-7
Nwi-1-19
Bsp-1-6
Nwi-1-20
Nwi-1-21
Nsp-1-8
Nwi-1-22
env-38
Oan-1-66
Oan-1-67
Oan-1-68
Oan-1-69
Nwi-1-23
Nwi-1-24
Nha-1-13
Nha-1-14
Nwi-1-25
Nwi-1-26
Nsp-1-9
Rpa-3-14
Nsp-1-10
Nsp-1-11
Nsp-1-12
env-39
Nha-1-15
Nha-1-16
Rpa-3-15
env-40
Rpa-3-16
Nwi-1-27


Rpa-3-17
Rpa-3-18
Nha-1-17
Rpa-3-19
Nha-1-18
Nsp-1-13
Nsp-1-14
Nwi-1-28
Rpa-2-3
Nha-1-19
Nwi-1-29
Nsp-1-15
Nsp-1-16
Nwi-1-30
Nwi-1-31
Nha-1-20
Nha-1-21
Nwi-1-32
Nha-1-22
Nwi-1-33
Nwi-1-34
Bsp-1-7
Bsp-1-8
Bsp-1-9
Bsp-1-10
Bsp-1-11
Bsp-1-12
Bsp-1-13
Bsp-1-14
Bsp-1-15
Bsp-1-16
Nwi-1-35
Nha-1-23
Nsp-1-17
Nwi-1-36
Mlo-1-21
Nha-1-24
Nsp-1-18
Nsp-1-19
Rpa-3-20
Nsp-1-20
Rpa-2-4
Rpa-3-21
Rpa-3-22
Rpa-3-23
Rpa-3-24
Rpa-3-25
Rpa-3-26
Rpa-3-27
Nha-1-25
Nha-1-26
Nha-1-27
Nha-1-28
Nwi-1-37
Nha-1-29
Nsp-1-21
Nha-1-30
Nsp-1-22
Nsp-1-23
Nwi-1-38
Nha-1-31
Nsp-1-24

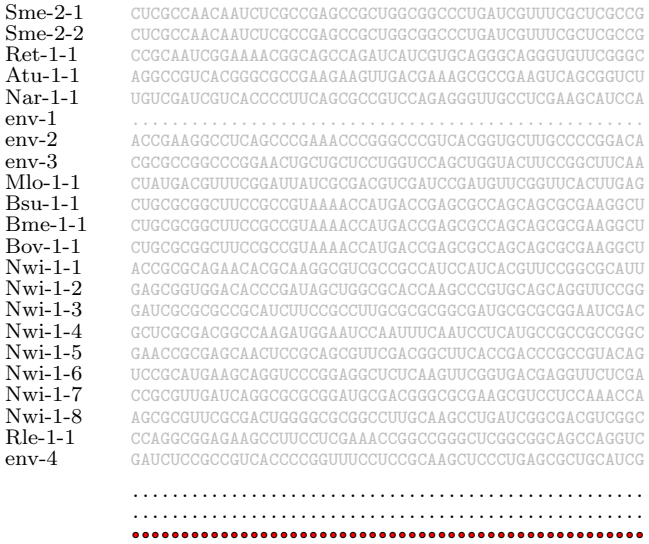

Nsp-1-25
Rpa-2-5
Rpa-1-2
Rpa-2-6
Rpa-2-7
Nha-1-32
Nha-1-33
Nsp-1-26
Nwi-1-39
Nsp-1-27
Nsp-1-28
Nsp-1-29
Nwi-1-40
Nwi-1-41
Nsp-1-30
Nha-1-34
Nha-1-35
Nha-1-36
Nha-1-37
Nha-1-38
Nha-1-39
Nha-1-40
Nha-1-41
Nha-1-42
Nha-1-43
Nsp-1-31
Nha-1-44
Nha-1-45
Nwi-1-42
Nwi-1-43
Nsp-1-32
Nsp-1-33
Nsp-1-34
Nsp-1-35
Nsp-1-36
Nsp-1-37
Nsp-1-38
Nsp-1-39
Nsp-1-40
Nsp-1-41
Nsp-1-42
Nwi-1-44
Nsp-1-43
Nsp-1-44
Nsp-1-45
Nsp-1-46
Nsp-1-47
Nwi-1-45
env-41
Rpa-4-5
Rpa-3-28
Rpa-3-29
Rpa-3-30
Rpa-3-31
Rpa-3-32
Rpa-3-33
Nsp-1-48
Nsp-1-49
Rpa-3-34
Nwi-1-46
Mlo-1-22
env-42

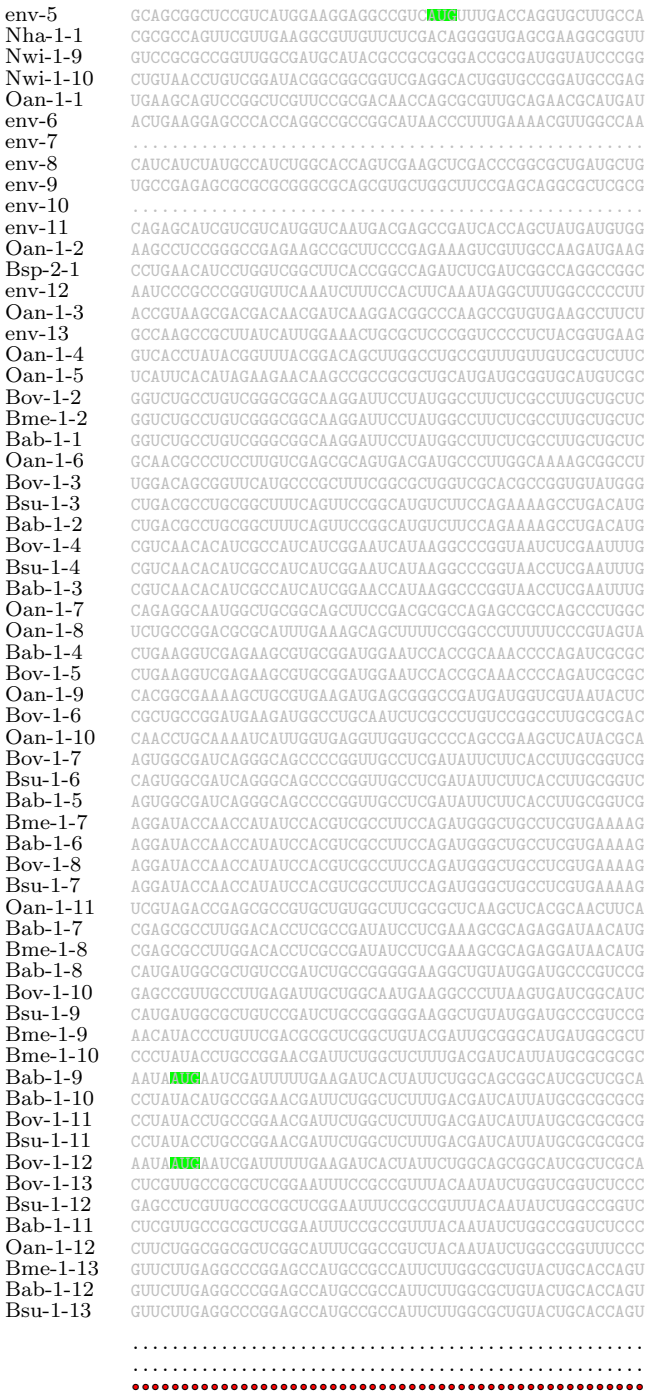
Csp-1-1
Csp-1-2
Csp-1-3
Ccr-1-1
Oan-1-70
env-43
env-44
env-45
env-46
Sme-2-3
Oan-1-71
env-47
env-48
Rle-1-43
Rle-1-44
Sme-1-2
Sme-2-4
env-49
env-50
Mlo-1-23
Mlo-1-24
Mlo-1-25
Sme-1-3
Rle-1-45
Ret-1-9
Rle-1-46
Rle-1-47
Rle-1-48
Ret-1-10
Msp-1-1
Sme-1-4
Bab-1-39
Bov-1-41
Bsu-1-43
Bme-1-38
env-51
Nha-1-46
Mch-1-10
Mch-1-11
Mch-1-12
Mex-1-12
Mch-1-13
Mex-1-13
Mex-1-14
Mex-1-15
Mex-1-16
Mch-1-14
Mex-1-17
Mch-1-15
Mch-1-16
Mex-1-18
Mex-1-19
Mex-1-20
env-52
Csp-1-4
Csp-1-5
Csp-1-6
Csp-1-7
Csp-1-8
env-53
env-54
Nwi-1-47


Rpa-3-35
Bja-1-27
Rpa-4-6
Rpa-3-36
Rpa-3-37
Rpa-3-38
Rpa-3-39
Rpa-3-40
Nwi-1-48
Nwi-1-49
Nwi-1-50
Nha-1-47
Nha-1-48
Nha-1-49
Nha-1-50
Nha-1-51
Nha-1-52
Nha-1-53
Nha-1-54
Nha-1-55
Nwi-1-51
Nha-1-56
Nha-1-57
Nha-1-58
Nha-1-59
Nha-1-60
Nha-1-61
Nha-1-62
Nha-1-63
Nha-1-64
Nha-1-65
Nha-1-66
Nha-1-67
Nha-1-68
Nha-1-69
Nha-1-70
Nha-1-71
Nha-1-72
Nha-1-73
Nha-1-74
Nha-1-75
Nwi-1-52
Nwi-1-53
Nha-1-76
Nha-1-77
Nha-1-78
Nsp-1-50
Nsp-1-51
Nsp-1-52
Nsp-1-53
Nsp-1-54
Nwi-1-54
Nsp-1-55
Nsp-1-56
Nha-1-79
Nha-1-80
Nha-1-81
Rpa-3-41
Nha-1-82
Nha-1-83
Nha-1-84
Nha-1-85

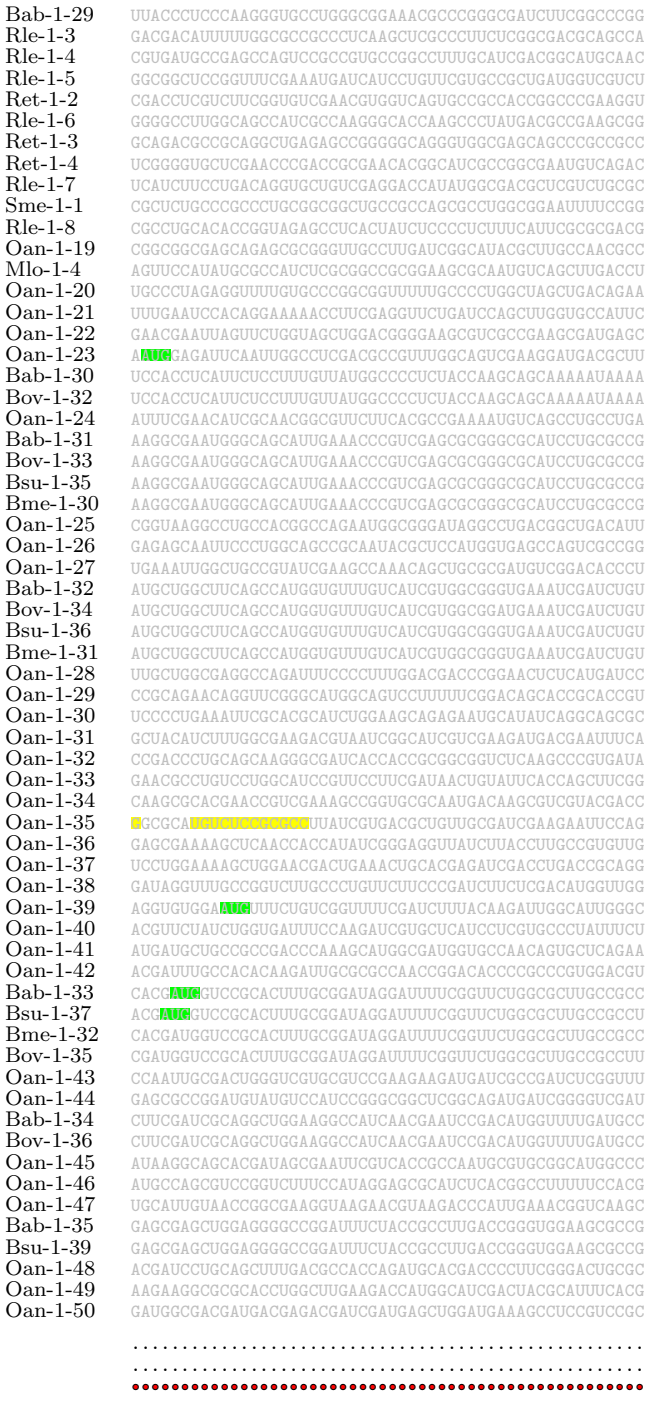


alignment positions $420 \cdots 471$

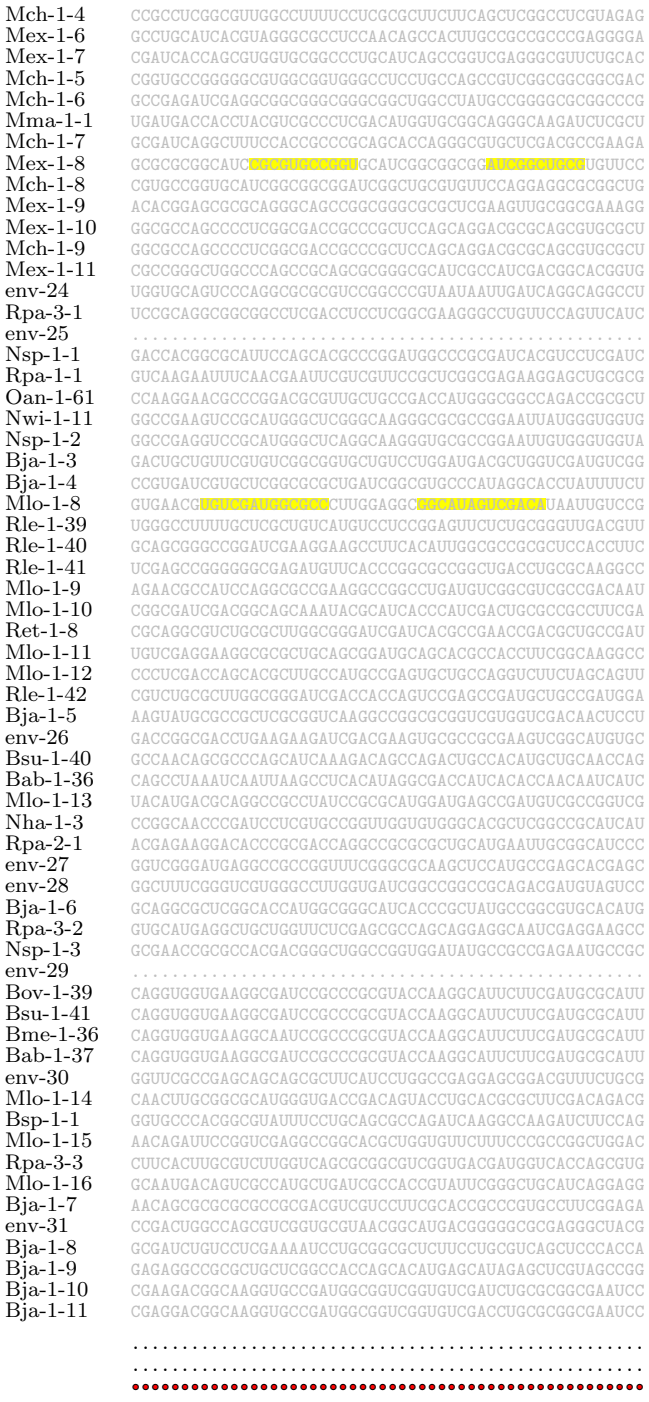


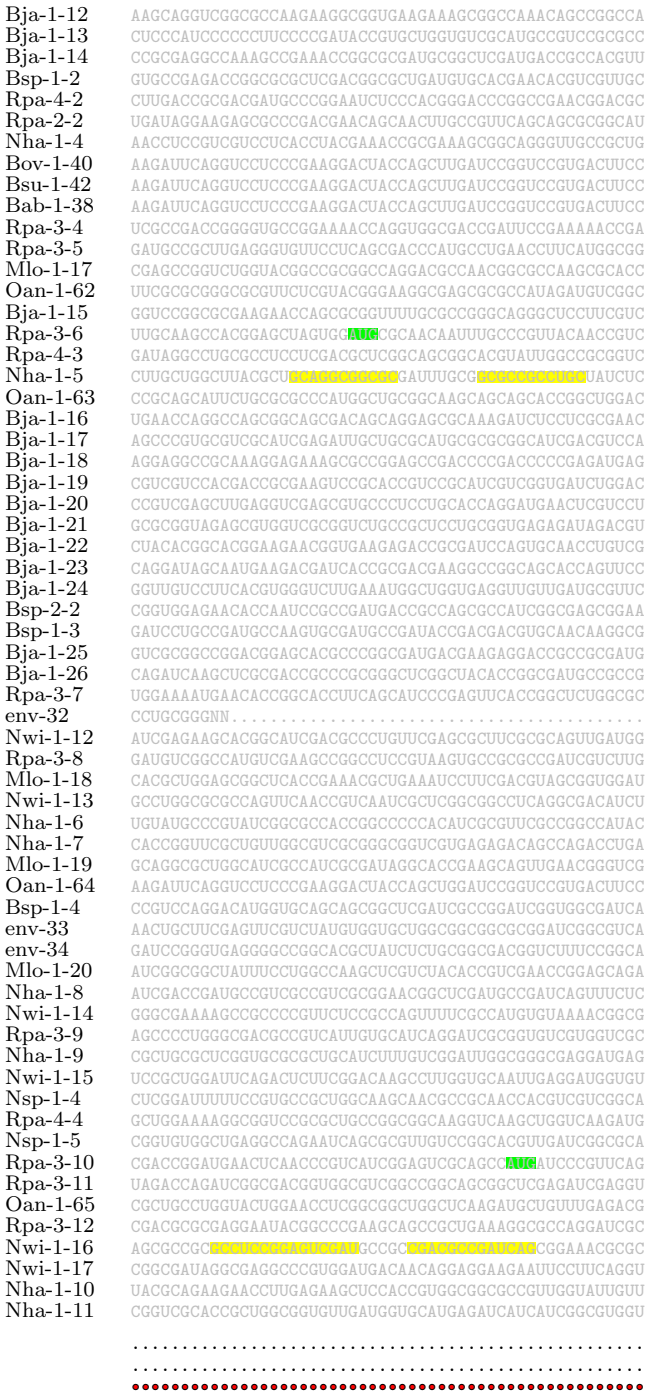

env-5
Nha-1-1
Nwi-1-9
Nwi-1-10
Oan-1-1
env-6
env-7
env-8
env-9
env-10
env-11
Oan-1-2
Bsp-2-1
env-12
Oan-1-3
env-13
Oan-1-4
Oan-1-5
Bov-1-2
Bme-1-2
Bab-1-1
Oan-1-6
Bov-1-3
Bsu-1-3
Bab-1-2
Bov-1-4
Bsu-1-4
Bab-1-3
Oan-1-7
Oan-1-8
Bab-1-4
Bov-1-5
Oan-1-9
Bov-1-6
Oan-1-10
Bov-1-7
Bsu-1-6
Bab-1-5
Bme-1-7
Bab-1-6
Bov-1-8
Bsu-1-7
Oan-1-11
Bab-1-7
Bme-1-8
Bab-1-8
Bov-1-10
Bsu-1-9
Bme-1-9
Bme-1-10
Bab-1-9
Bab-1-10
Bov-1-11
Bsu-1-11
Bov-1-12
Bov-1-13
Bsu-1-12
Bab-1-11
Oan-1-12
Bme-1-13
Bab-1-12
Bsu-1-13

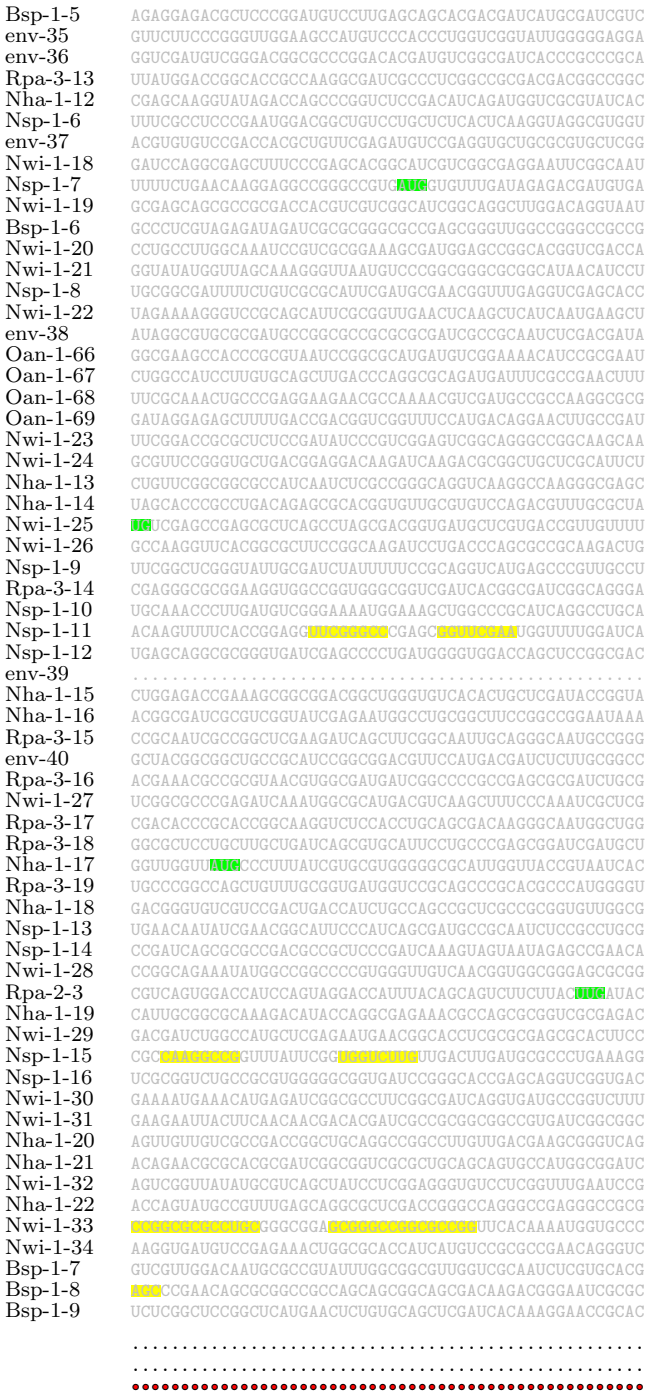
Bov-1-14
Bab-1-13
Bme-1-14
Bab-1-14
Bab-1-15
Bov-1-17
Bsu-1-17
Bab-1-16
Bab-1-17
Bov-1-18
Bab-1-18
Oan-1-13
Bov-1-20
Bsu-1-20
Bab-1-19
Bab-1-20
Bov-1-21
Bsu-1-21
Bme-1-19
Oan-1-14
Bov-1-22
Bab-1-21
Bab-1-22
Bov-1-23
Oan-1-15
Bab-1-23
Bab-1-24
Bab-1-25
Bab-1-26
Bov-1-25
Bsu-1-25
Bme-1-24
Bme-1-25
Bsu-1-26
Bov-1-26
Bsu-1-27
Bsu-1-28
Bsu-1-29
Bsu-1-30
Bov-1-27
Bov-1-28
Bme-2-26
Bab-1-27
Bov-1-29
Bsu-1-31
Oan-1-16
env-14
env-15
Nha-1-2
Mlo-1-2
env-16
env-17
Bja-1-1
Oan-1-17
Rle-1-2
Mlo-1-3
Bab-1-28
Bsu-1-32
Bme-1-27
Oan-1-18
Bov-1-31
Bsu-1-33


Bab-1-29 UUACCCUCCCAAGGGUGCCUGGGCGGAAACGCCCGGGCGAUCUUCGGCCCGG
Rle-1-3 GACGACAUUUUUGGCGCCGCCCCUCAAGCUCGCCCUUCUCGGCGACGCAGCCA
Rle-1-4 CGUGAUGCCGAGCCAGUCCGCCGUGCCGGCCUUUGCAUCGACGGCAUGCAAC
Rle-1-5 GGCGGCUCCGGUUUCGAAAUGAUCAUCCUGUUCGUGCCGCUGAUGGUCGUCU
Ret-1-2 CGACCUCGUCUUCGGUGUCGAACGUGGUCAGUGCCGCCACCGGCCCGAAGGU
Rle-1-6 GGGGCCUUGGCAGCCAUCGCCAAGGGCACCAAGCCCUAUGACGCCGAAGCGG
Ret-1-3 GCAGACGCCGCAGGCUGAGAGCCGGGGGCAGGGUGGCGAGCAGCCCGCCGCC
Ret-1-4 UCGGGGUGCUCGAACCCGACCGCGAACACGGCAUCGCCCGGCGAAUGUCAGAC
Rle-1-7 UCAUCUUCCUGACAGGUGCUGUCGAGGACCAUAUGGCGACGCUCGUCUGCGC
Sme-1-1 CGCUCUGCCCGCCCUGCGGCGGCUGCCGCCAGCGCCUGGCGGAAUUUUCCGG
Rle-1-8 CGCCUGCACACCGGUAGAGCCUCACUAUCUCCCCUCUUUCAUUCGCGCGACG
Oan-1-19 CGGCGGCGAGCAGAGCGCGGGUUGCCUUGAUCGGCAUACGCUUGCCAACGCC
Mlo-1-4 AGUUCCAUAUGGCGCCAUCUCGCGGCCGCGGAAGCGCAAUGUCAGCUUGACCU
Oan-1-20 UGCCCUAGAGGUUUUGUGCCCGGCGGUUUUUGCCCCUGGCUAGCUGACAGAA
Oan-1-21 UUUGAAUCCACAGGAAAAACCUUCGAGGUUCUGAUCCAGCUUGGUGCCAUUC
Oan-1-22 GAACGAAUUAGUUCUGGUAGCUGGACGGGGAAGCGUCGCCGAAGCGAUGAGC
Oan-1-23 AAUGGAGAUUCAAUUGGCCUCGACGCCGUUUGGCAGUCGAAGGAUGACGCUU
Bab-1-30 UCCACCUCAUUCUCCUUUGUUAUGGCCCCUCUACCAAGCAGCAAAAAUAAAA
Bov-1-32 UCCACCUCAUUCUCCUUUGUUAUGGCCCCUCUACCAAGCAGCAAAAAUAAAA
Oan-1-24 AUUUCGAACAUCGCAACGGCGUUCUUCACGCCGAAAAUGUCAGCCUGCCUGA
Bab-1-31 AAGGCGAAUGGGCAGCAUUGAAACCCGUCGAGCGCGGGCGCAUCCUGCGCCG
Bov-1-33 AAGGCGAAUGGGCAGCAUUGAAACCCGUCGAGCGCGGGCGCAUCCUGCGCCG
Bsu-1-35 AAGGCGAAUGGGCAGCAUUGAAACCCGUCGAGCGCGGGCGCAUCCUGCGCCG
Bme-1-30 AAGGCGAAUGGGCAGCAUUGAAACCCGUCGAGCGCGGGCGCAUCCUGCGCCG
Oan-1-25 CGGUAAGGCCUGCCACGGCCAGAAUGGCGGGAUAGGCCUGACGGCUGACAUU
Oan-1-26 GAGAGCAAUUCCCUGGCAGCCGCAAUACGCUCCAUGGUGAGCCAGUCGCCGG
Oan-1-27 UGAAAUUGGCUGCCGUAUCGAAGCCAAACAGCUGCGCGAUGUCGGACACCCU
Bab-1-32 AUGCUGGCUUCAGCCAUGGUGUUUGUCAUCGUGGCGGGUGAAAUCGAUCUGU
Bov-1-34 AUGCUGGCUUCAGCCAUGGUGUUUGUCAUCGUGGCGGAUGAAAUCGAUCUGU
Bsu-1-36 AUGCUGGCUUCAGCCAUGGUGUUUGUCAUCGUGGCGGGUGAAAUCGAUCUGU
Bme-1-31 AUGCUGGCUUCAGCCAUGGUGUUUGUCAUCGUGGCGGGUGAAAUCGAUCUGU
Oan-1-28 UUGCUGGCGAGGCCAGAUUUCCCCUUUGGACGACCCGGAACUCUCAUGAUCC
Oan-1-29 CCGCAGAACAGGUUCGGGCAUGGCAGUCCUUUUUCGGACAGCACCGCACCGU
Oan-1-30 UCCCCUGAAAUUCGCACGCAUCUGGAAGCAGAGAAUGCAUAUCAGGCAGCGC
Oan-1-31 GCUACAUCUUUGGCGAAGACGUAAUCGGCAUCGUCGAAGAUGACGAAUUUCA
Oan-1-32 CCGACCCUGCAGCAAGGGCGAUCACCACCGCGGCGGUCUCAAGCCCGUGAUA
Oan-1-33 GAACGCCUGUCCUGGCAUCCGUUCCUUCGAUAACUGUAUUCACCAGCUUCGG
Oan-1-34 CAAGCGCACGAACCGUCGAAAGCCGGUGCGCAAUGACAAGCGUCGUACGACC
Oan-1-35 GGCGCAUGUCUCCGGCGCUUAUCGUGACGCUGUUGCGAUCGAAGAAUUCCAG
Oan-1-36 GAGCGAAAAGCUCAACCACCAUAUCGGGAGGUUAUCUUACCUUGCCGUGUUG
Oan-1-37 UCCUGGAAAAGCUGGAACGACUGAAACUGCACGAGAUCGACCUGACCGCAGG
Oan-1-38 GAUAGGUUUGCCGGUCUUGCCCUGUUCUUCCCGAUCUUCUCGACAUGGUUGG
Oan-1-39 AGGUGUGGAAUGCUUUCUGUCGGUUUUCGAUCUUUACAAGAUUGGCAUUGGGC
Oan-1-40 ACGUUCUAUCUGGUGAUUUCCAAGAUCGUGCUCAUCCUCGUGCCCUAUUUCU
Oan-1-41 AUGAUGCUGCCGCCGACCCAAAGCAUGGCGAUGGUGCCAACAGUGCUCAGAA
Oan-1-42 ACGAUUUGCCACACAAGAUUGCGCGCCAACCGGACACCCCGCCCGGUGGACGU
Bab-1-33 CACGAUGGUCCGCACUUUGCGGAUAGGAUUUUCGGUUCUGGCGCUUGCCGCC
Bsu-1-37 ACGAUGGUCCGCACUUUGCGGAUAGGAUUUUCGGUUCUGGCGCUUGCCGCCU
Bme-1-32 CACGAUGGUCCGCACUUUGCGGAUAGGAUUUUCGGUUCUGGCGCUUGCCGCC
Bov-1-35 CGAUGGUCCGCACUUUGCGGAUAGGAUUUUCGGUUCUGGCGCUUGCCGCCUU
Oan-1-43 CCAAUUGCGACUGGGUCGUGCGUCCGAAGAAGAUGAUCGCCGAUCUCGGUUU
Oan-1-44 GAGCGCCGGAUGUAUGUCCAUCCGGGCGGCUCGGCAGAUGAUCGGGGUCGAU
Bab-1-34 CUUCGAUCGCAGGCUGGAAGGCCAUCAACGAAUCCGACAUGGUUUUGAUGCC
Bov-1-36 CUUCGAUCGCAGGCUGGAAGGCCAUCAACGAAUCCGACAUGGUUUUGAUGCC
Oan-1-45 AUAAGGCAGCACGAUAGCGAAUUCGUCACCGCCAAUGCGUGCGGCAUGGCCC
Oan-1-46 AUGCCAGCGUCCGGUCUUUCCAUAGGAGCGCAUCUCACGGCCUUUUUCCACG
Oan-1-47 UGCAUUGUAACCGGCGAAGGUAAGAACGUAAGACCCAUUGAAACGGUCAAGC
Bab-1-35 GAGCGAGCUGGAGGGGCCGGAUUUCUACCGCCUUGACCGGGUGGAAGCGCCG
Bsu-1-39 GAGCGAGCUGGAGGGGCCGGAUUUCUACCGCCUUGACCGGGUGGAAGCGCCG
Oan-1-48 ACGAUCCUGCAGCUUUGACGCCACCAGAUGCACGACCCCUUCGGGACUGCGC
Oan-1-49 AAGAAGGCGCGCACCUGGCUUGAAGACCAUGGCAUCGACUACGCAUUUCACG
Oan-1-50 GAUGGCGACGAUGACGAGACGAUCGAUGAGCUGGAUGAAAGCCUCCGUCCGC

Oan-1-51 GACGGCAGGCGUCGUCUAUCAUGCCGAACGCGUGGCGAAGCUUCUGAUGGCU
Oan-1-52 GCGAUCCGCAGACACUUUCCGGCUUGCAGGGCGGCGUCGUGGAAAUGACGGU
Oan-1-53 UCAGGCGCUUUUGCGCGGCAUCACCACAUGAGGACAAGAUAUGUUCAAUACC
Oan-1-54 UCAAGGCCGAUGCUGCAACCGAUAUUUCCGCUCCACCGCCCGCCGUGCCGGU
Oan-1-55 CUGCAUGGCCAGCAAAGCUGCCGUUUCCUGCGGCAGACCGGCGGUCACACCG
Oan-1-56 CGACAACGAAGGUGUAGCGGUACGCAACACCAUCUGGCGAACGAACGCCGAG
env-18 ....
env-19 CUUGGCGAAUAUUUCUUACAACUAAGCCCAGCCAAGUCAAUAUUACCGAAAA
Oan-1-57 AUUUUCUGCUUACGUACAGAAUUCAGAAAAAGGUCUUGUAGAUUUUGCGAUC
Oan-1-58 GGCACACCAGCCUGCAUUGAAAUAACAUUGCCAUGGAAGCGGCGACCGAAGG
Mlo-1-5 GACCAGCUGUCCGAGCUUGACCGCGAUGCCGGGCGCGGCUUGAUCCAGCCGG
Rle-1-9 CAACCCGCCGAGAAGUGCCGCCCAGGCCGGAUGAAUGCUCGAAAUGGACAGG
Rle-1-10 GAUGGGCGAUCUCGAUCCAUUCGUUGCGGCAGCGGCGCACGAGUUCGGCCCA
Rle-1-11 GCAUGCUGGUGCUUGCCGUCACCAAUUUUGCCGCCAUGCCCUUCCUGAUGAC
Rle-1-12 CGCUUGCCAUCAUGGGUUACCAAUGGACGCCGUGGAUCUUCACGCUUCUCGC
Rle-1-13 UUCUUGCCAGGUGAGCUCGAAGGAAUGAUCGGUAUCGAUAUAGGCCCAGUCG
Rle-1-14 GCGUACAGCCGACCUUGCGCGCCGCCGGGGUGGAGACCGUGGCCGUUAUCAA
Rle-1-15 CUGGCCGAUCUUUUCGACAAUGGCGAGUUUGAAUCGCUGCCGCAGCCGAAGG
Rle-1-16 CGUCUCUUCCUUCGGCGCCUCCGGCGUCUUGGUCUUGGUUGGGGCCGCCGGC
Rle-1-17 UCCUGCUUGCGGAAAGCCGCACCAUCGCAAUAGACCGGCGCAUUGACGACGG
Rle-1-18 GCGGUCCCAUCGGAUGGUUGGCGCCGAGCUUCAUUGCCGUAUCGAUGGCGUC
Rle-1-19 GCUGAUGUCAUGCGGGUCAACAGCACCUCGGCCUCAUCGAUAUCGUCAAACA
Rle-1-20 GCUCUCGAAAAGACGCGCGGCGCUGUAAGAAGGGGUGUCGAAUCGAUUUGCU
Rle-1-21 UGCCGCCGAGGCUGGCGCGGUGGCGGAAGAGCGCCAGCAUGCCGGCGCCGAU
Rle-1-22 CUGUGGAAACAGUAUCUGCUGUUUCUCCAAGGGCUUCUGCACGGGGACUUCG
Rle-1-23 CGAGAAUGAUCUCUGACUUGUCGGUACCUUCGCGCAGCAGCAUGUCAUCGAU
Rle-1-24 GCCAGGAUAUCAAGGAUGUUUCUGCAGGCAUCACUCACAUCCUGUUGACGCA
Rle-1-25 GCCCUGCAUGACCUUCAUGGCAUCCGGCGAAUUCAGGAUUUCGGCGCCGCCG
Ret-1-5 GCAGCUGUUUUAACGAUUGAGUGUGGCUGGCAACGGACGACCGUACCCGCUA
Rle-1-26 UCAUCGGCAGCACGAGAUCGCGCCGCAAGCGAUUGCCUUCCGAUUUGCCCGA
Rle-1-27 UCCAUCAAGGGCUCGUCGCCGAACCGGCGAUAAGCCCUUUCCGCUUCCAGCG
Rle-1-28 AUUUCCUGCUUCAACAGUAGGAGCGCUCCGCACAUAUAUCGGAGCGCUCUGG
Rle-1-29 CCAUUUCCGUUCCGAUGUAACCGGGGCAAAUAGCGUUGACGGUAAUGUUCUU
Rle-1-30 CAUCGGUUAUCUCUCGGAAUGCAGGGAUCGAACCGGCGCGACGGCAUGCCGC
Rle-1-31 UCGGAGCGCGGUCACAAGGCGGUCUCGCGGGCGCUCGUCAAUCUGCUUAUCA
Rle-1-32 GGCCUUCGCCCCCAGCGACAGCGCUUUAAGAACAUGCGUUCCCCGCUGCACG
Rle-1-33 UCGAACAGCCGGAUCAACAGCAUAGCGGCGACGUUGAAGACACGUUCGUUAC
Rle-1-34 CCUUGAUCAUUUCGGGAUCGAGCGCCGAGGUCGGCUCGUCGAACAGCAUGAU
Rle-1-35 AGCAAUGCGCCGCUCUACGAAAACAUCGAUUCGAUGCUUGCUGAGCAGAAGC
Rle-1-36 UCGAUCAUGUCGCCGUUUUUCACGACCGCCAGCGGCCCGCCGACAGCUGCUU
Ret-1-6 AUGCGCUACGCGCCAUGGGCGUUGCGAUCGACGAGCCCGACGACACCACCUU
Rle-1-37 UUCGUCAACAAUGGCAUAGGCAGCUUUGGCUCUCAGUCUCAAAAACUCGACG
Oan-1-59 GUUACGAAGAAAUCCUGACCGACCCGUCUUAUGCCGGUCAGAUCGUAACCUU
Rle-1-38 GCCGGCUUCGGCAAGGUGACCGAUCCGGCCUCGACCGCGAGGCAUGAUCGCCG
Ret-1-7 CUCCGUCCCUCGCUUGCCCGGCUGGUAACGACGAAACGCCUGCUAGCGCGCA
Oan-1-60 AGCUUAUUGCUGUGCAGGUGUAUCGUCCGGAACAGGCGCUUCUUCCUGCUCG
Mlo-1-6 GAUGGGCAUCGAAUCCAAGUUCGAGGGUGGGCUGCGCGUCACCGACCAGAAG
Mlo-1-7 CCGGCUCCAACAACACCGCGGAAGCGGUCGGGCUUGUGCAAUAUGCCGAAAA
env-20 CGGUUAAGCGUAUCUCCCAAAAGUGGAUCCCGGUUUUGGGACGAAAAUACGU
env-21 GCCUUUACUCCGCGUUCCGUCAGCCAAUUCCGCAGGCUGUUGGCGUCGUAGG
env-22 AACCGGGCCCGGAAAAUGGGGGC.....................
env-23 GGCCUACACGCUGCGCUAUAUCCCGGCACGCGAGGACAUCGACACGCUCGAU
Bja-1-2 CGGAUCUCGUCGGGAUCGGUGAUGGCGGUUGCCAGCACGUUCGACACCAGCG
Rpa-4-1 CAUCGGCGUAGUGGGCGAUCUGCGGAUUGUGCGAGCGCUGGGUCUCGUAGAC
Mch-1-1 CCGGUUCCGGCGGCGUCAGCCGGGGUGCCGAUGCCGAUGCUCGCCGCCGAGC
Mex-1-1 CCGGUUCCGGCGGCGUCAGCCGGGGUGCCGAUGCCGAUGCUCGCCGCCGAGC
Mch-1-2 ACUGAAGGCAUGGAGCGCGAGCUUGGCCGCAUCCGCCAUCGCCGGCACGCCG
Mex-1-2 GCGCACGUCGCCGUCGGUGAGGAUGCCGGCGAGCGCGCCCUUUUCAUCGACC
Mex-1-3 UGCCCUUGUAGCGGCCGGUCUUGAGCAGCACCGAGGACCAGAAAUCGUGGAU
Mch-1-3 UGCCCUUGUAGCGGCCGGUCUUGAGCAGCACCGAGGACCAGAAAUCGUGGAU
Mex-1-4 CGGAGGAGGCCUACGGCCUGCUCCAGGGGGCCGGGCGCGUCGAUCUCCUGUU
Mex-1-5 CCGCCUCGGCGUUGGCCUUUUCCUCGCGCUUCUUCAGCUCGGCCUCGUAGAG


Mch-1-4
Mex-1-6
Mex-1-7
Mch-1-5
Mch-1-6
Mma-1-1
Mch-1-7
Mex-1-8
Mch-1-8
Mex-1-9
Mex-1-10
Mch-1-9
Mex-1-11
env-24
Rpa-3-1
env-25
Nsp-1-1
Rpa-1-1
Oan-1-61
Nwi-1-11
Nsp-1-2
Bja-1-3
Bja-1-4
Mlo-1-8
Rle-1-39
Rle-1-40
Rle-1-41
Mlo-1-9
Mlo-1-10
Ret-1-8
Mlo-1-11
Mlo-1-12
Rle-1-42
Bja-1-5
env-26
Bsu-1-40
Bab-1-36
Mlo-1-13
Nha-1-3
Rpa-2-1
env-27
env-28
Bja-1-6
Rpa-3-2
Nsp-1-3
env-29
Bov-1-39
Bsu-1-41
Bme-1-36
Bab-1-37
env-30
Mlo-1-14
Bsp-1-1
Mlo-1-15
Rpa-3-3
Mlo-1-16
Bja-1-7
env-31
Bja-1-8
Bja-1-9
Bja-1-10
Bja-1-11


Bja-1-12
Bja-1-13
Bja-1-14
Bsp-1-2
Rpa-4-2
Rpa-2-2
Nha-1-4
Bov-1-40
Bsu-1-42
Bab-1-38
Rpa-3-4
Rpa-3-5
Mlo-1-17
Oan-1-62
Bja-1-15
Rpa-3-6
Rpa-4-3
Nha-1-5
Oan-1-63
Bja-1-16
Bja-1-17
Bja-1-18
Bja-1-19
Bja-1-20
Bja-1-21
Bja-1-22
Bja-1-23
Bja-1-24
Bsp-2-2
Bsp-1-3
Bja-1-25
Bja-1-26
Rpa-3-7
env-32
Nwi-1-12
Rpa-3-8
Mlo-1-18
Nwi-1-13
Nha-1-6
Nha-1-7
Mlo-1-19
Oan-1-64
Bsp-1-4
env-33
env-34
Mlo-1-20
Nha-1-8
Nwi-1-14
Rpa-3-9
Nha-1-9
Nwi-1-15
Nsp-1-4
Rpa-4-4
Nsp-1-5
Rpa-3-10
Rpa-3-11
Oan-1-65
Rpa-3-12
Nwi-1-16
Nwi-1-17
Nha-1-10
Nha-1-11

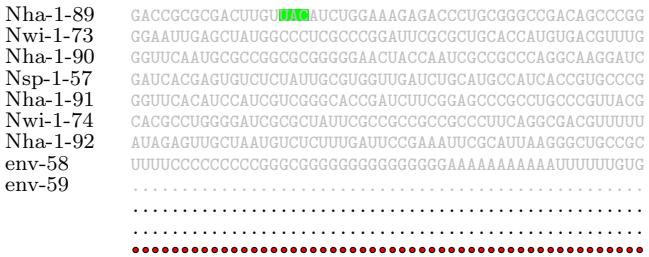

Bsp-1-5
env-35
env-36
Rpa-3-13
Nha-1-12
Nsp-1-6
env-37
Nwi-1-18
Nsp-1-7
Nwi-1-19
Bsp-1-6
Nwi-1-20
Nwi-1-21
Nsp-1-8
Nwi-1-22
env-38
Oan-1-66
Oan-1-67
Oan-1-68
Oan-1-69
Nwi-1-23
Nwi-1-24
Nha-1-13
Nha-1-14
Nwi-1-25
Nwi-1-26
Nsp-1-9
Rpa-3-14
Nsp-1-10
Nsp-1-11
Nsp-1-12
env-39
Nha-1-15
Nha-1-16
Rpa-3-15
env-40
Rpa-3-16
Nwi-1-27
Rpa-3-17
Rpa-3-18
Nha-1-17
Rpa-3-19
Nha-1-18
Nsp-1-13
Nsp-1-14
Nwi-1-28
Rpa-2-3
Nha-1-19
Nwi-1-29
Nsp-1-15
Nsp-1-16
Nwi-1-30
Nwi-1-31
Nha-1-20
Nha-1-21
Nwi-1-32
Nha-1-22
Nwi-1-33
Nwi-1-34
Bsp-1-7
Bsp-1-8
Bsp-1-9

Bsp-1-10 CGAUCAGCCCUUCCGCCGCGCAUUCACCGCGAUCGACAACAACGUCCGCUGC
Bsp-1-11 GGUUGGCGGUGAAGACCGAGCCGCCGAGACCGAACGGUGAAUCAUUGGCGAU
Bsp-1-12 GCCGCCAUCGAACAUGAUCUCGAUCUUGUCGCCGACCGCGUCCGCGAUGCCC
Bsp-1-13 ACCAUGCCGGUGUGGGUCCAGAUGUCGAAAUGCCCGAGCCACAGUUGCUCGU
Bsp-1-14 CGAGGAUCUGCACGUCGUGCUCGGUGUGGUCGACAUGGGUCACCAUCGGCAC
Bsp-1-15 GGCUGACCGCAGCGCAUCCUGGAUCAAGCCGCGCAAGGCGAUAUGAGGUGCC
Bsp-1-16 AAGCAGCAACAGGUAGGAGAGACGUUUCACUGGCUUGCCUGACAUUGCCGCC
Nwi-1-35 UCAGCACGAAUUUUGAAUCGCCGACCCGAAACCCGUUAUUCUUGAUCAGGAC
Nha-1-23 GAUAUUUGGCAACGGCACUCUCUCCGAUCACGCGGGGCGCCAUGCGCGCCCA
Nsp-1-17 CACGGCGAAGCUCGCGCGCGAGCUGUUCGGUACGUCCUGGAUCAAGCUCGAA
Nwi-1-36 UACGGCGAAGCUCGCGCGCGAACUGUUCGGCACGUCCUGGAUCAAGCUCGAA
Mlo-1-21 UCGGUGGUCGCUUUCGUCUUCAUGGAGCGCCAUUGGGCAACGGUGACGCAGC
Nha-1-24 AGAAAUUCGGUGACCCCGGCAUCGCGCGCAACGGUAACGCGCCGCUUCUCCG
Nsp-1-18 GGACCCACGCGACCAGUUCCUUGCGCAGUGUCUCGCUCGGCUCCACGCCCUG
Nsp-1-19 UACCGAAGAGCAGAUCGGAACGGUCGCUCAUGCCGAUCGUGCCGAUCUCGAC
Rpa-3-20 GCCCCUGAUAGCCCUGGCGGGCGACUUCGGCCGGCGACACGUUGAGGAUCGC
Nsp-1-20 UAGAAUGACGCGUCCCAUCCGCUGGCUGGAUUUGCUGCGGUAGAGGAAGAGC
Rpa-2-4 CAUAACGAGAGUCCGACGCCCAUGCCGGUGGACUUGGUGGUCAUGAACGGCA
Rpa-3-21 GGGAUCGUCGCUUUGCGGCCUCGUCGGGUCCAUGCUCGGGACCUGCAGCGUG
Rpa-3-22 AUGAUGCGGGCGACGUCGCGGGCGCUGACCUCGGACUCGGUCGAGACACCCU
Rpa-3-23 GAGGAUUCGAGCGCCGCGGAUCGGUCAAACCGUCACCUGGCCGGCCGAGCGA
Rpa-3-24 CGGGAUCGACAUCACGCCGUAUUUCGACGCGGUGACCGGGCUCUCGUCAACA
Rpa-3-25 GCUGGCCGAGCGCAUCAAUGCACAUCACGCCCGCGGCCAGAAGUUCAAGGUC
Rpa-3-26 CACGUCGACGGUGGUCCGCUGACCGCCCUGGGCCUCGCGGCUGACGCCCUUC
Rpa-3-27 CAUCGCCUUCUACUAUGCCGAGGCCGACCAUAUCGUCGUCGUUCGAGUGCUG
Nha-1-25 CGCCCUUUCGCUUCUCGUCGUCAAUCUGAUCACGCCCAGAUACAAGUCCGAA
Nha-1-26 AUCUGGUCGUGGUUGCUGGUAUAGGCGUUGACGACGCAACUGCUUCCGGGUG
Nha-1-27 CUCAAGGAUCUCGGCCAGCUCGUGGCGCGCGAGAUCGAGAAGGCCGGCGGUG
Nha-1-28 ACUUCGCUUGAAAACGCUCUCAAUCUGUCCGGAGCCAUGAUCCCGUCCGCGA
Nwi-1-37 GUAGCCUUGCGCUUCCGUCACGGUGCCGUCGAGGCUGACGAGUUCAAGCUGC
Nha-1-29 GUUGCUUCUCGGCGCGCGCUUGCGGCUCCGCCGUCUCGCCGAAACGGAAACG
Nsp-1-21 GCGCCUCGAGGUGCUCGGGGACCGCGUAGCCCUUGUGUGUCUCGAAGCCGUA
Nha-1-30 CCGUUCCGGCGAGGCAAGACCGAUCAGCCGCCCCUGGAUGUAGUCGCAGCCC
Nsp-1-22 CAACCGCAACGUAUCUUUCGAGACGUAUUAACAUUGUCGUGCCGCUCCAACG
Nsp-1-23 CUUGGGGUCGACGCGCCUGGAAGGAUCCCAGAGGUGCGAGCGGAUGAUGGCG
Nwi-1-38 CAUUAUCGAUGCCGGACGUUCCCCACGGACGGCAGCGCAGGAUCCGCGCCAG
Nha-1-31 GACUAUUGCCGUCGGGCCGCAAUGGCGGAAUCGACCUUUGGCCGGCGUGCCG
Nsp-1-24 UGCCGCUCCACGUCCCCUCGAAAACAUCAGGAUCGACGUUAACGUAUUUUCC
Nsp-1-25 GGCCGAUCGUUUGGAGGUUGCGUGGCGAAUACUUUUCUACACGGCACAGUGC
Rpa-2-5 UUGGUUUCAGACCAGCGCGCAAUUCUGGCUCAAUCUGCAAAGCGCCUACGAU
Rpa-1-2 UCCGCCCGCGACCAGCUUAAAGUACCUCUCGCCGCAUGUACCAGCCUCAGAU
Rpa-2-6 GUACUUGCCGCUGUGGACUGGUUUCGCUCCAACAUCCGAACGGGCGUGCGGC
Rpa-2-7 AUCGACGAAUUGCGGCAGGAAGGCCAGGAAGAACAGCGCCACUUUCGGAUUG
Nha-1-32 GGAGUCCGGCUUUGAUUCCAUCCGAGCCGGACUCCAGGUCAACGCGAAAGCC
Nha-1-33 GCGCAGCAGAAAGUUGUGCAGCGCCCGAUGGGCGCGGGCCUGCGUCAUGCCG
Nsp-1-26 CACCAUCGUCACGUGGAUUGUGGUGCGGAUUUCCGAUCGUCCCGCCGAUGCC
Nwi-1-39 UACGGGUCAUGAGAACCUCCUCGCGUGUUCAACCGCUUUACUCAGCGGCCUG
Nsp-1-27 CUCGCGGCGAUCGGCUUCGACCCCUGGAAUAUCGUGCAAAGCAUUCGCCUGC
Nsp-1-28 ACCAACGAGCGCCGGCUGUAUCUGAUCAUCGAUUCGGGUCGCGCUCUGCGAU
Nsp-1-29 GUCGGGGGGCGCGUUUUCAUAAGCCUCGAUCGCCUGCUCGCGCGCCUUGAAA
Nwi-1-40 UCUUAUUCAUGUAUCGCCGAGACAACUGAAAUCAUUGCGGAUUUCAGGUUUG
Nwi-1-41 CGCCGCGAAGCGCUGCGAUCCCUGAGUACGGGCGAGAGCCGCCACCAGCGCG
Nsp-1-30 CCAGCAUAGGCGAGCUCGAAACUGAACAUCAGCUUGCCGGUGUGUUCAGCCU
Nha-1-34 CGACGAAUAAAGCUGGAUCAGGGUAGCACCGGCGCGAAUCUUGGUCAGUGCG
Nha-1-35 GCCGGCGGGGAUACGGAACGGAGUGGCCGGCUUGUCGGCAAGCGCGACCUUC
Nha-1-36 GUCGAGCACGAGCCACGCCCGUUGAUACCGGCGGCGCAUGCGCCGAAUGGCG
Nha-1-37 ACCUGCAUAUCGACCCGUUUUUGGGCCUGCCGAUCACCGCGGCGGUGAUGUU
Nha-1-38 GAACGGGUUCUGGUCCGAGGCCGCCAAGCGGGUCGACUGGAUCACGCCCUUC
Nha-1-39 UAGACCGUCACCACCUGAUCGCUCGGCCCUUCGACCUGAAUGACGCGAUCCA
Nha-1-40 CACCAACUCGGCCGUACGGGUUCAUGGUGUUUGAACAGCUCGCGGAGAUCGG
Nha-1-41 CGCGGUCGGCGUCCGCAUCGUCACCAGUUCUUCGGCCGCCGUGGGGUGCAAU
Nha-1-42 ACGAAAUUGUUUCCCAAAUAAAAUUGACUUCGCGGGAUGGCCAACAGCCCGC

Nha-1-43 GCGAUCGAUUCGAGCGACCAUGUCGCGGUCGAAUCCCUCGGCGACGAUCGAC
Nsp-1-31 GAACGGUCGGCCUUCAAUCGUCAUGUCGCCGGGGCCCAUCUCGUCGACCGCC
Nha-1-44 GGCGCUGGCGGGAGACGGGUCGGGUCGAUCCGCGUCCGCGUGGCGGGUUCCG
Nha-1-45 CCUCCGGUUCGCUACACCCGGAAAUGCUUCGACAGCUUGAGGCCCUGGGCCU
Nwi-1-42 GAACUGCUGAGCCAGAUUUCCGACUAUUGCCGCCGGGUCGGCAUGGCCGAAU
Nwi-1-43 UAUUCGAACGACCUUCGUGAGCGGGUAGUGCGUGCUGUCGUUAAGGCGGCC
Nsp-1-32 ACGGGACGGACGCGAUGGCUCUCCUAGCAAACCCUCCGCGCGCCCGAACCUU
Nsp-1-33 GGGUUUCCGAGAACGGAAUCAGUUCGACCCAGUCCACCGCAUCGACUCCCGG
Nsp-1-34 UCCCAACCCUCGCGAUGUCGAUUUCACGGCCUCGGUGCGGAGGCCGCCGACU
Nsp-1-35 AGCGGCGCGGUCGCGGCGUCGAGCCAGACUUUCGUGGCCUCGUCGAGAUGCG
Nsp-1-36 ACCUGGCGAAAUUCUGCUCGGCGCGCGGGCUGCCCGUCCCGACCUUCUCUGA
Nsp-1-37 CUUAAACUCAUAUCGCCGCGUGGCGACCGCUUCAGACCGCAGCUUCUCAAGG
Nsp-1-38 ACUUAACAUGACGCCAGAACGUUUGAUUUGCGCUGGAUCAAUCUCGGUCAAU
Nsp-1-39 CAGCAUCGCCAGCGCGCGUGCGAUGCGCCUGCGCGUCGAGGAUGGCAUGAUG
Nsp-1-40 GUUUUGCAAAUUGCCCAGCAACAUUGAAGACUACGGAAUUAUCGGCGACGGC
Nsp-1-41 CGACAGGAACGGCAGGCUCUGGAGCACGAGCACGCCCGCGAAGAUGUAGAUC
Nsp-1-42 GAUCGCGCGCAAGUCGCCGCCGCGGCUCAAGAGGUGCGUGGCGAAUGAAUGG
Nwi-1-44 UCAAGAUCCGCGGCACUGCCUCUCACGUCGCCGAAAAAUACCUCCAACUCGC
Nsp-1-43 GAGAUCGAGAUGGCCAGCGGCCAGCAUGCAAUAGGAAUAGCAGUCGCCUCCA
Nsp-1-44 CUGGGGAGGGUGUUAUUCGUUCCAAAAGCUCCUGCCGUUCGCCAAGACCUAUA
Nsp-1-45 UAAUCGAGCACGAACGGCGGCGCCAGGAUGAGCCGCCAUGAGAACGACAGCG
Nsp-1-46 GGGUGCGGGCGAGAGCCGCCACCAGCGCGACGGCGUCAUAGGCAAGGGUAGC
Nsp-1-47 CUCGCCUUUCCGACAAGCGCCGUGCCGUUCAUCGUGCCGGUCGUAUCCGUGA
Nwi-1-45 ACUUGAGCUUUAUGAGGAGAUGCAGGCCGUCAUCGGUUCAAACGAGAUCAAG
env-41 ....
Rpa-4-5 GUCGACGCGAUGCGCCGCGAACUCAACAAGAUGCCGAUCGAAGGCACCGUGG
Rpa-3-28 CGAGCGUCGAGAUCAGCACGACGUAGUCCGGCAUCCGCGUGUCCUUGCCGAC
Rpa-3-29 AUCGCCAGCGCCAUCAGGAACAAGCCGAAUGGCCCGACUUUGCCGACGAAGA
Rpa-3-30 UUCGCCGAUCCGGCGCAGCGGCGUGCCGGCGGUGCGGCGCUUCAACACCGCC
Rpa-3-31 GGCGCGCCGAAUCGCCUCGCGGGCGUUCUCGACCAGCUCGGGAUGGCGGUCG
Rpa-3-32 GAUCGAUGCCGCUGCCCGGCAGCAAUGGGAUCAGCGUCGGGCAACGGAUCGC
Rpa-3-33 UGCCAGGCGCCGGCCUGAUAGUCCCAGAAAGUGUACUGACCCGGGGCGUCGG
Nsp-1-48 CCGCCCACAGCCGCCCGAUCUCGGAGGCAGCGCGGCCCUGGGAAUCGUAUUC
Nsp-1-49 CGCGCCCCGCGAUCUUCCGCUGAUCGCCGCGGACUCCGACAAUGCCGCAUCG
Rpa-3-34 AGCGGCCGGCGCCUUAGGCUGAGGCUGAGGCUGAGGCUGGCGAGGCGCAGCG
Nwi-1-46 CCCGGAGAUGCCAGCCAGGACAUGCCAGAGUGCGUGGCGCUCAAUCGGCCCU
Mlo-1-22 UGUUAUGUCGAUGAGGUCUUUCUCGACCUCGGCGACCGCACGACGCCGCAGA
env-42 GAUACCGGCGGCCAUCUCGUCGUUGCUGGCCAGGAUCGCGGUGGGGCAGGGG
Csp-1-1 CGAACGCCGCCUCGCCCCGGCGCACCACGAAGCCGUGGAAGCGCUUGUCGCC
Csp-1-2 UGUGGAGGCGACGAACUCAGCUUGGGUCCCGAUGAUCACCGGGAUGCACGGC
Csp-1-3 GGCCCGCGACCCGCGCAUUUCCGUCGCCAUCGGCCAUCACGAGCCUGUCGAG
Ccr-1-1 GGCGGCUGAAGCCCAGGGCGCGGGCGACAGCCUCGGCGCUCACCUCGCCGGU
Oan-1-70 CAGCCCUUCGUUGGAAUGAUAGAAGAAGCCCUUGUCGCCCAGUUUCAUGGCG
env-43 GCGCCCGACGAUCCGUCAUCGAACUCGGCCUUCAGUUCGCGGGCCGCGCGCC
env-44 AAGUUUCUCCAGGCUCGCCACGAGCAGCGGCGUGUUGAACAGCACGCACGUG
env-45 GCCUGCCGGCCGCGCCUGGCGCUACUUUCACGAGGUGAUCCUGCACUGGACC
env-46 AGCUGCAGCGCGAGGGCUAUAUCGAGGCGAUCACCUCGCCGGAGUUUCUCAA
Sme-2-3 UGCGGGCAAGAUCACCAAGAUCGGCCUGGAUAUCGGCUCCACCGGCAAAACC
Oan-1-71 UGUCGAAAAAGCGGUCAUUGAUGCCUCCGAUCUGGAAGAAGCAGAGAUCCGC
env-47 CUGUCGGAUUGGCGACAUUCUUUCCCGCGACAUCCGGCGCCGAUCCAUGCUG
env-48 CCAACCGCCGGGAGCGACAACGAGGAAAGCUGGCGCCGCAGCCCCUUCUACU
Rle-1-43 CGGCAAUAUGGCAAGCAUCGCGCUCGACUUUGCCGCCGGUUCCGGCCCCUCG
Rle-1-44 AAAACACCCUUCGCCAGAGACGCUCGUCGCGCGUGGGCCGUUCCGGCAGAUC
Sme-1-2 GGGCCAUUGUAAUCGGAUUGUGGGUGAUGAUGACGAAACGGGUCUCAGUCGA
Sme-2-4 GUUUUGCCGAGCGAUAUCUGGCCGUGCUGCCCGAGCAGGCAUGCAAGGCGGU
env-49 GAAAUGGCCGGGCGCGUCCACUGGGGGAACCCUCCAGUAUGCGCUUCUUGGG
env-50 CGGCGGUGUUUCCGCCACGAUCACCGCGGUCGGCCCGAGCGCCUGGUCCAAG
Mlo-1-23 AUCAACCGGUAAUCGAGCUCGCGGCUCCAUUUGCGCGUCAGGAAAGCCUGGA
Mlo-1-24 UACACCUGGCAGAGGGGCAUCCCCGACCUCAGGCAGGCGCUGGCACGCUACU
Mlo-1-25 GCCUCGCGGAUCAGCGCGCCGCCUAUGCCGCUGCCAUGGAACGAAGGUUCCA
Sme-1-3 CCGGCUCGCGCUUGUAGCCCUCUAUCAGCACGAGGUCGCAGGGUGCGAGCCG
Rle-1-45 GCCGGUCUUGACCUGCUUCUGCUCGACGACGUCGUCCUUACCGAGAACAAGC
. . . . . .
. . . . . .

Ret-1-9 CGCCGCCGAAGAAACCGGCGGUCAUCGCCAGCAGGAGAACGACGAGAAGCAU
Rle-1-46 AAACCGUGCAGAUCCGCCGUCGUUAUGACGAAAUCUUCCGUUCCGGCGGCAU
Rle-1-47 GGCCGAGACGAACGGCCGGAGCGUGAUCUCGCCCGCCCGCUGCCGUCGUUUC
Rle-1-48 GCACGUCCUCUUUGCUUGGUCCGUGCCCCUCUUCGGACCAGUCAUCCAUUUG
Ret-1-10 CGCGUAGCGGCAGCCGGUUUGCCGUGCGAGCGCCACGGCAUUCUCCGCCGCC
Msp-1-1 GCUUGGCCCGGCGCUGCCCGCACAAGCUUGACAAGCCCCGGCAUCGGCGCGC
Sme-1-4 GGAUCGGUUCGGCGCCAAACGCAUCUUCCGCGCGGCCAUCCUCGUCUUCGUC
Bab-1-39 UGUGGUGCUGGAUGAACUGCACUCGCUGGUGAUUUCCAAGCGCGGGCAUUUG
Bov-1-41 UGUGGUGCUGGAUGAACUGCACUCGCUGGUGAUUUCCAAGCGCGGGCAUUUG
Bsu-1-43 UGUGGUGCUGGAUGAACUGCACUCGCUGGUGAUUUCCAAGCGCGGGCAUUUG
Bme-1-38 UGUGGUGCUGGAUGAACUGCACUCGCUGGUGAUUUCCAAGCGCGGGCAUUUG
env-51 ....
Nha-1-46 CCUCGCCGUUGGUGAUGGCGUCCACGAUAUGCGGCCGGCCCUCCAGAACCUU
Mch-1-10 CGAAGGCGAGCAGCGAGACCUCGCCGUGGCGCUCGUCCUGCUGGAUGCGCUC
Mch-1-11 AGCUCCGGCUGUCGGCCGUCAGCCCGAGAUGCAGGCCGGCGAGGAUGCCGGC
Mch-1-12 AACACGUUUCCCCCCGCGGCCCUGAGAUCCGGGCCCGGAGGACGGCUGGCCG
Mex-1-12 CUGCCGCCUCUGACGCUCGUCGUCCUGCGUGUCGGAUUGGCUGCGCUGAUCC
Mch-1-13 CCGACUUUGCGGGGGACGACACGGUGCGGGCCGUGGUGCUGGCGGCGGCGGG
Mex-1-13 CUGCGAGAACAACUCGACCUGGCUCAACGCGUUCUGCCCGACGGUGGCGUCG
Mex-1-14 UCAGCCCGGAUCAAGCGGUCUACGCGGCGACCGAGCUCGGCGGGUCGUUCUG
Mex-1-15 CACCUCGCUCGGCCGCGGCGCCACCGAGAAUUGGGACGCGCUGACGACCGGA
Mex-1-16 CGGCUUCAGCGGGCCUGGGACGCCGGCAAGGCCUCCGGCCGGCACGGCGUGC
Mch-1-14 GCUGCCGCCGGGCCUGGGCACCGCACCUUGAAGCGAGCCGGGCGGUGAUGCG
Mex-1-17 CGCUUUCCUGCGAUCCUCGGCCUCCUCCUCCUCGCGGCCCUCUGGCUCACGG
Mch-1-15 CGCUUUCCUGCGAUCCUCGGCCUUCUCCUCCUCGCGGCCUUCUGGCUCACGG
Mch-1-16 CUCGAUACGACGACGAGCGAGGCGAGCGCGCUCCGCGAGGAGCGCCCGGAUU
Mex-1-18 GAGGCUUCCUUCACCUCGAUCACGUCGCCGGCCUUGACCUGGUAGCUCGGGA
Mex-1-19 CGUUGACCGUCGCGGUUCAGAUGGACCCGAUCCAGUCGAUCCGCAUCGCGGG
Mex-1-20 CCGAGGCUCGUCCCGUCGAUGCCGGCGGGCUCGUUCACCUCCCAGGCGCGCA
env-52 UGCCCGAGGUGACCGCCGAUCCGCUGAAAUUCCGCGACGAACGCAAAUAUGC
Csp-1-4 GUCUUGGCGUAGUAGGGCUCGAGACGACGGCGCAAGGCCUGAUCGUGCGGUC
Csp-1-5 CCUGAGAGGCAUGGCGUUAGGACCGCCCCAGUCGCGGUAGACCGCGUCCACG
Csp-1-6 AGGAGGUCGGCCAUCAGCGUCUGCAUCGGCUGUGGGCUGACCGUGUAGCGGA
Csp-1-7 UCGAACAGGUUCUGGGCCUGUAGCUGCACUCGCGUGCCACGCAGCCACGGAU
Csp-1-8 CGACGCACAAACCGUAGCAGACCCCAUCCAUGAGCUUUGAUGAAAAGGACGC
env-53 ....
env-54 AGCUGACGGGCGUCACCAUGCCCCAGAACAGUCCGCUGAGGCCGAUCAUCAG
Nwi-1-47 GCGUCCCGCGCAAGUUCGAAUCCGGACCCGGUCUCACGGCGAAAGAGAUGGU
Rpa-3-35 CAGAUCAUCUGACGGGCCACACGCGUACUCAAAUGUAAAACGGGCUUGCUGC
Bja-1-27 UUCUGUUGCAAUUCCAUGCCUCAUGAUCACGCUGACAUAUUUAAUAAAAUCU
Rpa-4-6 CGCGGCUUCGGCAAGCUUGAGUUCCUGCGCGCACUGCGCCUUCAGCUCCGGC
Rpa-3-36 UCAUCGAGUCUGCGAUUCGGCCGCGUUUCGUUCGGGACUCGUUCCACACCUU
Rpa-3-37 CUCGCCGAACCACCAGGCAUGCGGCACGAUGCCGUCCGGCGGCGCCUCGAGG
Rpa-3-38 AACUUCCGGAUCGGCGCGGAGCGGAUGCCGAUCGGCCUCGUGCACGCGCUCG
Rpa-3-39 UUUCUUCUCGGUCGUCUUCGUGCCCCUGAACUGCCCGGGGAGCCAUUUGGCG
Rpa-3-40 CGGCGCGAAGGCGACGUCGCGGCUGCCGAGGAAAAUGUGAUUGUACUCGGCG
Nwi-1-48 GGUCAUGCGAAUGACUUCCGCCAGCCCGAUCGCCUCGACGGGCCUGGCGAGA
Nwi-1-49 AGACGGCGAAAUCUGGAAGAUCUGGUACACCUUCGUUCCGGCGCCGGAAAAA
Nwi-1-50 AGGGGUUCUCAAGGCACUCAGGCGUCGGGUCCCCGGUGCAUUGCGCUGGUGG
Nha-1-47 CCCGGACCAGAUUCGGAAUAUCGAUCACCGAGAGCGGAUCGGUGCAAUACAC
Nha-1-48 ACCAUCUCGUGCAGGGACAGUUUCACCACCAUGUUUUCCUGCAGGUAGGGCG
Nha-1-49 GCGGGUUCCACUGGGCGGCAAAUCUUCGAAAUCGCCGGGGUCUGCUCCGCCU
Nha-1-50 UUAUCCUCGACCCGCUGCUCGAUAGCCACUGAGGCGCGACGCUCAUGCGCCA
Nha-1-51 CCGAGCGCAACCUGCGCUUCCGCCAUAUCCUGCUCAGCCGCUUUCUCGAACC
Nha-1-52 UGCCGCUGAAUAUGCGAUGAAGGACCUCGUCGGAUUUUCGCCUCAGUGCAGC
Nha-1-53 AUUUGUUGAGGAUAGUGCGUCUUACGAAGAUGGUCACGUGAUGUAAAGCUGG
Nha-1-54 GCCUCGCAGCUUGAUCUUGGCAAGUCCGGCGGAGGCCUUGAUCCAGCCGAAC
Nha-1-55 GAACAGGCCCGCAUAGGCCAGUUCGAAGCGAAACAUCAGCUUGUCGUUAUGU
Nwi-1-51 GCUAUCUGCUGGUUGACGUCCGCGAGCCGAACGAAAUUGCGGCGGAGGCCUU
Nha-1-56 CCAGCCGCAGCUUGCCGGCAUCGCGCGCACCGGCCUCGCCGCCGAGCGCCGC
Nha-1-57 CCGCCCCGCUGGGUGCCGGGGAUCGGUUCGGGCUGCAUGCCCUCAUGCGCGG
Nha-1-58 CGCCACCGGCGCUGCACCCGGAUUGGGGGUCGGCGCAUCCGUGGCGGCCGCG
. . . . . .
. . . . . .

Nha-1-59
Nha-1-60
Nha-1-61
Nha-1-62
Nha-1-63
Nha-1-64
Nha-1-65
Nha-1-66
Nha-1-67
Nha-1-68
Nha-1-69
Nha-1-70
Nha-1-71
Nha-1-72
Nha-1-73
Nha-1-74
Nha-1-75
Nwi-1-52
Nwi-1-53
Nha-1-76
Nha-1-77
Nha-1-78
Nsp-1-50
Nsp-1-51
Nsp-1-52
Nsp-1-53
Nsp-1-54
Nwi-1-54
Nsp-1-55
Nsp-1-56
Nha-1-79
Nha-1-80
Nha-1-81
Rpa-3-41
Nha-1-82
Nha-1-83
Nha-1-84
Nha-1-85
Nha-1-86
Nha-1-87
Nwi-1-55
Nwi-1-56
Nwi-1-57
Nwi-1-58
Nwi-1-59
Nwi-1-60
Nwi-1-61
Nwi-1-62
Nwi-1-63
Nwi-1-64
Nwi-1-65
Nwi-1-66
Nwi-1-67
Nwi-1-68
Nwi-1-69
Nwi-1-70
Nwi-1-71
Nwi-1-72
env-55
env-56
env-57
Nha-1-88



80 Rhizobiales-2

80.1 Taxa

The taxonomy of each organism containing a putative Rhizobiales-2 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Rhizobiales-2 RNA in Sections 80.2 and 80.4.

abbrev. of hits	taxonomy of species
Bja-1-1 to Bja-1-2	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Bradyrhizobium japonicum* USDA 110
Bsp-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Bradyrhizobium sp.* BTAi1
Bsp-2-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Bradyrhizobium sp.* ORS278
Nha-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Nitrobacter hamburgensis* X14
Nsp-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Nitrobacter sp.* Nb-311A
Nwi-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Nitrobacter winogradskyi* Nb-255
Rpa-1-1 to Rpa-1-2	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* BisA53
Rpa-2-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* BisB18
Rpa-3-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* BisB5
Rpa-4-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* CGA009
Rpa-5-1 to Rpa-5-2	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* HaA2
Rpa-6-1 to Rpa-6-2	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* TIE-1

80.2 Gene contexts

Each Rhizobiales-2 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Rhizobiales-2 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 80.3. The accession number of the sequence containing each Rhizobiales-2 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 80.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Bja-1-1	NC_004463.1	+	86174	86270	hypo→ **RNA**→ DUF1194 (pfam06707)NOG12842 (NOG12842)→
Rpa-4-1	NC_005296.1	+	2048180	2048283	hypo→ **RNA**→ hypo→
Rpa-6-1	NC_011004.1	+	2195886	2195989	hypo→ **RNA**→ hypo→
Rpa-6-2	NZ_ABJR01000001.1	+	180800	180903	hypo (cons)→ **RNA**→ hypo (cons)→
Bsp-1-1	NC_009485.1	+	3149215	3149317	hypo→ **RNA**→ ←hypo
Bja-1-2	NC_004463.1	+	3362455	3362556	Peptidase_S8 (pfam00082)→ **RNA**→ hypo→
Bsp-2-1	NC_009445.1	+	2873967	2874068	←hypo **RNA**→ hypo→
Rpa-1-1	NC_008435.1	+	2056469	2056569	hypo→ **RNA**→ hypo→
Nwi-1-1	NC_007406.1	-	2442749	2442653	Peptidase_S8 (pfam00082)→ **RNA**→ COG3237 (COG3237)→
Nsp-1-1	NZ_AAMY01000022.1	-	71516	71421	←hypo **RNA**→ COG3237 (COG3237)→
Rpa-2-1	NC_007925.1	+	1899601	1899703	hypo→ **RNA**→ hypo→
Rpa-3-1	NC_007958.1	+	2153720	2153823	hypo→ **RNA**→ hypo→
Rpa-5-1	NC_007778.1	-	4059694	4059591	Peptidase_S8 (pfam00082)→ **RNA**→ hypo→
Nha-1-1	NC_007964.1	-	2918364	2918268	Peptidase_S8 (pfam00082)→ **RNA**→ hypo→
Rpa-5-2	NC_007778.1	-	641631	641545	hypo→ **RNA**→ COG3637 (COG3637)OprB (pfam04966)→
Rpa-1-2	NC_008435.1	+	126470	126561	hypo→ **RNA**→ ←GltS_FMN (cd02808)

80.3 Conserved domains

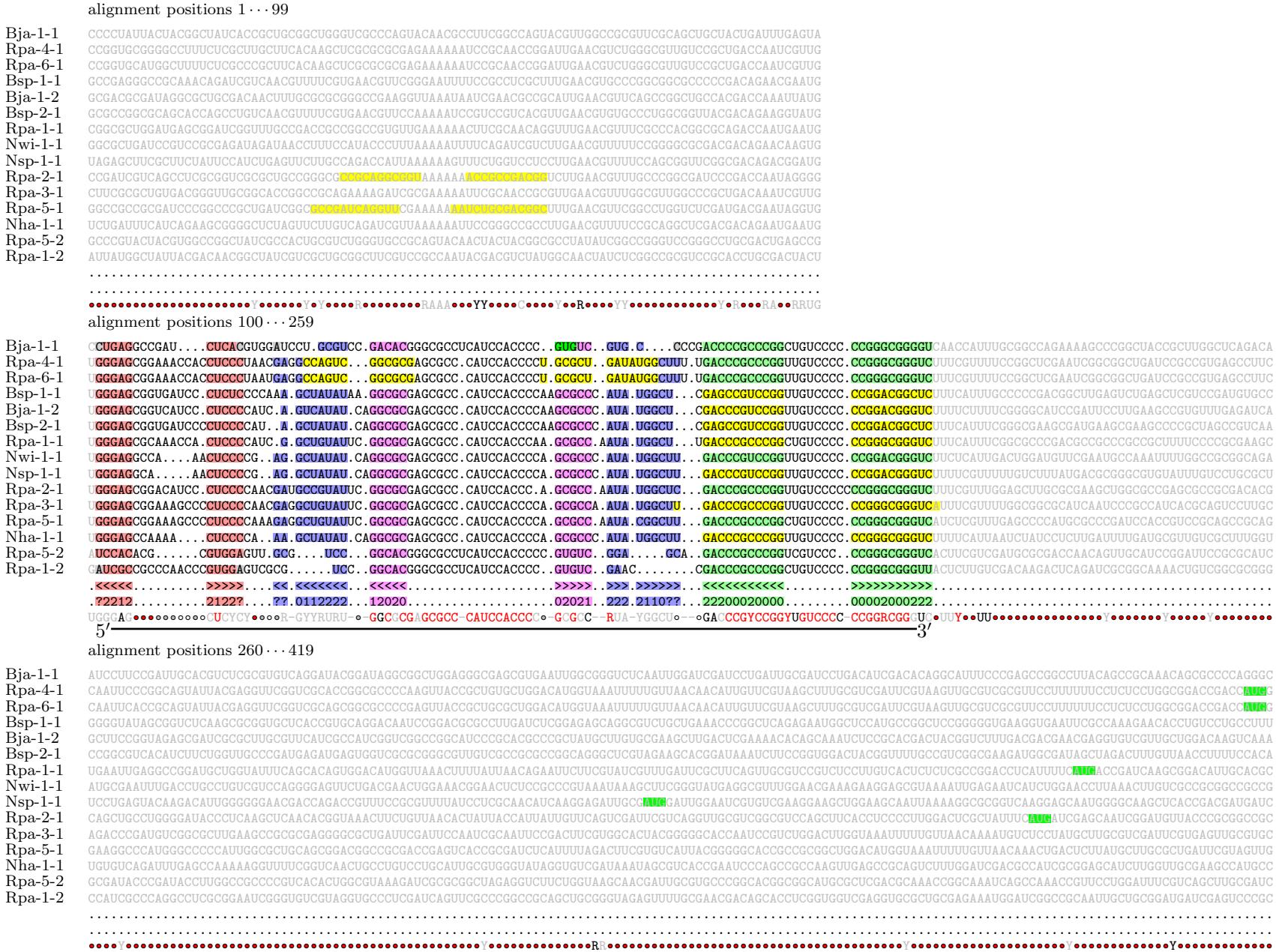
Conserved domains found in protein-coding genes listed in Section 80.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Rhizobiales-2 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

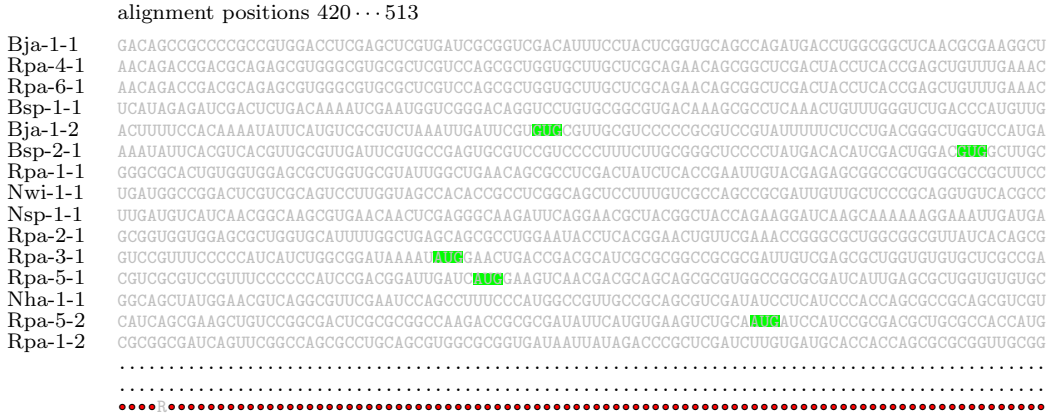
d cd02808 (1) Glutamate synthase (GltS) FMN-binding domain.
D COG3237 (2) Uncharacterized protein conserved in bacteria [Function unknown]
D COG3637 (1) Opacity protein and related surface antigens [Cell envelope biogenesis, outer membrane]
D NOG12842 (1)
u pfam00082 (4) Subtilase family.
D pfam04966 (1) Carbohydrate-selective porin, OprB family.
D pfam06707 (1) Protein of unknown function (DUF1194).

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Rpa-6-2

80.4 Multiple-sequence alignment

Each Rhizobiales-2 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 80.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Rhizobiales-2 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.


alignment positions 1 · · · 99
Bja-1-1
Rpa-4-1
Rpa-6-1
Bsp-1-1
Bja-1-2
Bsp-2-1
Rpa-1-1
Nwi-1-1
Nsp-1-1
Rpa-2-1
Rpa-3-1
Rpa-5-1
Nha-1-1
Rpa-5-2
Rpa-1-2
alignment positions 100 · · · 259
Bja-1-1
Rpa-4-1
Rpa-6-1
Bsp-1-1
Bja-1-2
Bsp-2-1
Rpa-1-1
Nwi-1-1
Nsp-1-1
Rpa-2-1
Rpa-3-1
Rpa-5-1
Nha-1-1
Rpa-5-2
Rpa-1-2
5′
3′
alignment positions 260 · · · 419
Bja-1-1
Rpa-4-1
Rpa-6-1
Bsp-1-1
Bja-1-2
Bsp-2-1
Rpa-1-1
Nwi-1-1
Nsp-1-1
Rpa-2-1
Rpa-3-1
Rpa-5-1
Nha-1-1
Rpa-5-2
Rpa-1-2



81 Rhodopirellula-1

81.1 Taxa

The taxonomy of each organism containing a putative Rhodopirellula-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Rhodopirellula-1 RNA in Sections 81.2 and 81.4.

abbrev. of hits	taxonomy of species
Pae-1-1 to Pae-1-2	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae *Prosthecochloris aestuarii* DSM 271
Bma-1-1 to Bma-1-12	Bacteria Planctomycetes Planctomycetacia Planctomycetales Planctomycetaceae *Blastopirellula marina* DSM 3645
Pma-1-1 to Pma-1-2	Bacteria Planctomycetes Planctomycetacia Planctomycetales Planctomycetaceae *Planctomyces maris* DSM 8797
Rba-1-1 to Rba-1-36	Bacteria Planctomycetes Planctomycetacia Planctomycetales Planctomycetaceae *Rhodopirellula baltica* SH 1
Lsp-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Limnobacter sp.* MED105
Psp-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae *Polaromonas sp.* JS666
Tde-1-1	Bacteria Proteobacteria Betaproteobacteria Hydrogenophilales Hydrogenophilaceae *Thiobacillus denitrificans* ATCC 25259
dpr-1-1	Bacteria Proteobacteria Deltaproteobacteria delta proteobacterium MLMS-1
CDe-1-1 to CDe-1-2	Bacteria Proteobacteria Deltaproteobacteria Desulfobacterales Desulfobacteraceae Desulfococcus *Candidatus* Desulfococcus oleovorans Hxd3
Pca-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Pelobacteraceae *Pelobacter carbinolicus* DSM 2380
Maq-1-1	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Alteromonadaceae *Marinobacter aquaeolei* VT8
Msp-1-1	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Alteromonadaceae *Marinobacter sp.* ELB17
Sde-1-1	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Alteromonadaceae *Saccharophagus degradans* 2-40
Spe-1-1	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella pealeana* ATCC 700345
Noc-1-1	Bacteria Proteobacteria Gammaproteobacteria Chromatiales Chromatiaceae *Nitrosococcus oceani* ATCC 19707
Epe-1-1 to Epe-1-6	Bacteria Proteobacteria Gammaproteobacteria sulfur-oxidizingsymbionts *Endoriftia persephone* 'Hot96_1+Hot96_2'
Mfe-1-1	Bacteria Proteobacteria *Mariprofundus ferrooxydans* PV-1
env-1 to env-82	environmental samples

81.2 Gene contexts

Each Rhodopirellula-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Rhodopirellula-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 81.3. The accession number of the sequence containing each Rhodopirellula-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 81.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	JCVI_SCAF_1101668587780	+	239	350	**RNA**→ ANK (cd00204)→
env-2	JCVI_SCAF_1096627145484	+	436	522	**RNA**→ ? (GOS12064495)→ ? (GOS13313260)→
env-3	JCVI_SCAF_1101668605391	+	903	989	**RNA**→
env-4	soil_Contig17737	+	763	844	**RNA**→
env-5	sludgeOz_Contig686	+	230	335	**RNA**→ (,,,)→
env-6	JCVI_SCAF_1101668636129	-	948	827	**RNA**→ ? (GOS5566)→
Pma-1-1	NZ_ABCE01000004.1	-	216649	216566	**RNA**→ hypo→ putative IS4 family transposase→ hypo→ putative serine/threonine-protein kinase→
env-7	gws2_d1_0273_1	+	5677	5757	**RNA**→ ←hypo
env-8	JCVI_SCAF_1101668552826	-	475	397	**RNA**→
env-9	soil_XZS53285_y1	+	313	391	**RNA**→ hypo→
env-10	wf1_AGZO10865_b2	-	595	481	**RNA**→ hypo→
Bma-1-1	NZ_AANZ01000008.1	+	145850	145933	**RNA**→ hypo→
env-11	wf2_AHAA3541_b1	-	601	480	**RNA**→
env-12	JCVI_SCAF_1101668653671	+	729	828	**RNA**→
env-13	soil_XZS56238_g1	-	340	257	**RNA**→
Rba-1-1	NC_005027.1	+	2059640	2059720	**RNA**→ hypothetical protein-signal peptide and transmembrane prediction→ hypothetical protein-transmembrane prediction→ hypo→
env-14	sludgeOz_Contig11279	+	7675	7732	**RNA**→ ←(,,,)
Rba-1-2	NC_005027.1	+	653164	653246	**RNA**→ FKBP_C (pfam00254)→ hypo→
Bma-1-2	NZ_AANZ01000002.1	+	315381	315462	**RNA**→ hypo→
Bma-1-3	NZ_AANZ01000004.1	-	6670	6592	**RNA**→ ←COG1733 (COG1733)
Rba-1-3	NC_005027.1	+	658090	658172	**RNA**→ hypo→
env-15	soil_Contig27	+	396	477	**RNA**→
env-16	soil_Contig36107	-	1105	1024	**RNA**→ hypo→ hypo→
Noc-1-1	NC_007484.1	+	1207069	1207152	**RNA**→ hypo→ hypo→ COG2929 (COG2929)→
Rba-1-4	NC_005027.1	-	6703040	6702957	**RNA**→ hypo→ hypothetical protein-transmembrane prediction→ hypothetical protein-signal peptide and transmembrane prediction→ hypothetical protein-signal peptide and transmembrane prediction→
Rba-1-5	NC_005027.1	-	4140138	4140052	**RNA**→ hypothetical protein-putative membrane bound protein→ hypo→ hypothetical protein-transmembrane region and signal peptide prediction→
Rba-1-6	NC_005027.1	-	4139398	4139314	**RNA**→ hypo→ hypothetical protein-transmembrane region and signal peptide prediction→
env-17	soil_XZS92484_b1	-	257	174	**RNA**→
env-18	soil_Contig38576	+	968	1052	**RNA**→ ←hypo
env-19	soil_Contig6927	-	1374	1290	**RNA**→ Lactoylglutathione lyase and related lyases→
env-20	soil_Contig934	+	429	513	**RNA**→ COG2929 (COG2929)→ hypo→
env-21	soil_Contig6927	-	944	860	**RNA**→ Lactoylglutathione lyase and related lyases→
env-22	soil_Contig937	-	252	168	**RNA**→
env-23	soil_Contig6927	-	359	275	**RNA**→
Bma-1-4	NZ_AANZ01000004.1	-	269423	269342	**RNA**→ hypo→ hypo→
Rba-1-7	NC_005027.1	-	1616609	1616528	**RNA**→ hypothetical protein-signal peptide and transmembrane prediction→ hypo→ hypo→ hypothetical protein-signal peptide and transmembrane prediction→ hypothetical protein-signal peptide and transmembrane prediction→

Spe-1-1	NZ_AAVJ01000053.1	-	7018	6914	**RNA**→ ←DUF388 (pfam04076)
env-24	JCVI_SCAF_1096627149823	+	211	293	**RNA**→ ? (GOS11558247)→
Rba-1-8	NC_005027.1	-	6703745	6703662	**RNA**→ hypo→
Rba-1-9	NC_005027.1	-	20628	20546	**RNA**→ hypothetical protein-transmembrane prediction→
Rba-1-10	NC_005027.1	-	1620038	1619956	**RNA**→ ←hypo
Bma-1-5	NZ_AANZ01000008.1	+	146662	146742	**RNA**→ hypo→
env-25	JCVI_SCAF_1096627470609	+	174	256	**RNA**→ Ribosomal_S23p (pfam05635)? (GOS12064012)→
Bma-1-6	NZ_AANZ01000002.1	+	311758	311842	**RNA**→ hypo→ hypo→
env-26	JCVI_SCAF_1096627143648	-	674	590	**RNA**→ ANK (cd00204)? (GOS6)→
Bma-1-7	NZ_AANZ01000002.1	+	314601	314683	**RNA**→ hypo→ hypo→
Rba-1-11	NC_005027.1	+	2058946	2059030	**RNA**→ ←hypo
Pma-1-2	NZ_ABCE01000013.1	+	18825	18908	**RNA**→ hypo→ hypo→
Rba-1-12	NC_005027.1	+	611384	611465	**RNA**→ FimT (COG4970)→
Rba-1-13	NC_005027.1	-	1618073	1617992	**RNA**→ hypo→ GH16_laminarinase (cd02180)CBM_4_9 (pfam02018)→
Rba-1-14	NC_005027.1	-	1620666	1620584	**RNA**→ hypo→
Rba-1-15	NC_005027.1	-	6702477	6702396	**RNA**→ hypothetical protein-transmembrane prediction→ hypothetical protein-signal peptide and transmembrane prediction→ hypothetical protein-signal peptide and transmembrane prediction→
Bma-1-8	NZ_AANZ01000008.1	+	147634	147716	**RNA**→ ←hypo
Rba-1-16	NC_005027.1	-	1621103	1621021	**RNA**→ ←hypo
Rba-1-17	NC_005027.1	+	609527	609608	**RNA**→ ←hypothetical protein-transmembrane prediction
env-27	JCVI_SCAF_1096627351219	+	24	106	**RNA**→
Rba-1-18	NC_005027.1	+	656133	656215	**RNA**→ hypo→ hypo→
Pae-1-1	NZ_AAIJ01000014.1	+	44799	44881	**RNA**→ REC (cd00156)→
Rba-1-19	NC_005027.1	+	2064508	2064589	**RNA**→ ←INT_IntI (cd01193)
Rba-1-20	NC_005027.1	+	2058496	2058576	**RNA**→ hypo→
Bma-1-9	NZ_AANZ01000002.1	+	312520	312602	**RNA**→ hypo→
Rba-1-21	NC_005027.1	-	1618509	1618428	**RNA**→ ←hypo
Rba-1-22	NC_005027.1	-	1221800	1221719	**RNA**→ hypothetical protein-signal peptide and transmembrane prediction→
Rba-1-23	NC_005027.1	-	1619591	1619510	**RNA**→ ←hypo
Rba-1-24	NC_005027.1	-	1615924	1615842	**RNA**→ hypo→ hypo→ hypothetical protein-signal peptide and transmembrane prediction→ hypothetical protein-signal peptide and transmembrane prediction→
Rba-1-25	NC_005027.1	-	1621708	1621626	**RNA**→ ←hypothetical protein-transmembrane prediction
Rba-1-26	NC_005027.1	+	2060203	2060287	**RNA**→ hypothetical protein-signal peptide and transmembrane prediction→ hypothetical protein-transmembrane prediction→ hypo→
Bma-1-10	NZ_AANZ01000002.1	+	310729	310810	**RNA**→ hypo→ hypo→
Rba-1-27	NC_005027.1	-	1624939	1624857	**RNA**→ hypothetical protein-transmembrane prediction→
Rba-1-28	NC_005027.1	-	1625566	1625485	**RNA**→ hypothetical protein-signal peptide and transmembrane prediction→ hypothetical protein-transmembrane prediction→
Rba-1-29	NC_005027.1	+	610259	610340	**RNA**→ Acetyltransf_1 (pfam00583)NOG21735 (NOG21735)→ hypothetical protein-signal peptide prediction→ FimT (COG4970)→
Bma-1-11	NZ_AANZ01000002.1	+	313391	313472	**RNA**→ hypo→
Rba-1-30	NC_005027.1	+	2061199	2061279	**RNA**→ hypo→ hypo→
Bma-1-12	NZ_AANZ01000002.1	+	314027	314107	**RNA**→ ←hypo
Rba-1-31	NC_005027.1	+	610845	610928	**RNA**→ hypothetical protein-signal peptide prediction→ FimT (COG4970)→
Rba-1-32	NC_005027.1	-	1219215	1219133	**RNA**→ hypo→ hypo→
Rba-1-33	NC_005027.1	+	5397382	5397465	**RNA**→ hypo→
Rba-1-34	NC_005027.1	-	1619046	1618964	**RNA**→ hypo→ hypothetical protein-signal peptide prediction→
Rba-1-35	NC_005027.1	-	16585	16503	**RNA**→ RimK (COG0189)→ hypo→
env-28	JCVI_SCAF_1096627470609	+	1339	1421	**RNA**→
env-29	JCVI_SCAF_1101668308184	-	269	217	**RNA**→
env-30	gws2_g3_0002_13	+	1351345	1351409	**RNA**→ protein conserved in bacteria→ Transposase_8 (pfam01527)→ hypo→
Maq-1-1	NC_008740.1	-	4271794	4271725	**RNA**→ Glyoxalase (pfam00903)→ hypo→ hypo→
Pca-1-1	NC_007498.2	+	1618852	1618924	**RNA**→ hypo→ hypo→

Rba-1-36	NC_005027.1	-	20041	19960	**RNA**→ hypo→ Tra8 (COG2826)→
Epe-1-1	NZ_AASF01000436.1	-	263	198	**RNA**→
CDe-1-1	NZ_AAWN01000004.1	-	90381	90303	**RNA**→ ←COG2378 (COG2378)
CDe-1-2	NZ_AAWN01000064.1	-	10449	10383	**RNA**→ rve (pfam00665)→ hypo→
env-31	soil_XZS99204_b1	-	490	420	**RNA**→ hypo→
env-32	soil_Contig40575	-	1489	1390	**RNA**→ hypo→
Sde-1-1	NC_007912.1	+	2334095	2334161	**RNA**→ COG3576 (COG3576)→
env-33	gws2_d1_0290_1	-	999	879	**RNA**→ PA0570→
Psp-1-1	NC_007948.1	+	1988488	1988558	**RNA**→ COG3666 (COG3666)Transposase_11 (pfam01609)→ PLDc (cd00138)→
env-34	sludgePhrap_Contig16220	-	3651	3579	**RNA**→ Glyoxalase (pfam00903)→
Epe-1-2	NZ_AASF01002149.1	+	2146	2229	**RNA**→
Epe-1-3	NZ_AASF01001610.1	-	453	369	**RNA**→
Msp-1-1	NZ_AAXY01000001.1	+	44260	44338	**RNA**→ ←hypo
Epe-1-4	NZ_AASF01001702.1	+	96	175	**RNA**→
Epe-1-5	NZ_AASF01000597.1	-	154	82	**RNA**→
Epe-1-6	NZ_AASF01000398.1	-	418	341	**RNA**→
env-35	wf1_AGZO15577_g2	+	371	452	**RNA**→ hypo→
env-36	gws2_d1_0267_32	+	13113	13208	**RNA**→ membrane protein→
env-37	soil_Contig40575	-	835	735	**RNA**→ hypo→ hypo→
env-38	soil_XZS21013_g3	-	667	586	**RNA**→ ←hypo
env-39	JCVI_SCAF_1096627388974	-	241731	241667	**RNA**→ COG1741 (COG1741)? (GOS5791)→
Pae-1-2	NZ_AAIJ01000025.1	+	10188	10252	**RNA**→ ←hypo
dpr-1-1	NZ_AAQF01000008.1	+	16783	16834	**RNA**→ COG3472 (COG3472)DUF262 (pfam03235)→
Lsp-1-1	NZ_ABCT01000001.1	+	166411	166462	**RNA**→ hypo→
Mfe-1-1	NZ_AATS01000002.1	-	130478	130414	**RNA**→ hypo→ hypo→
env-40	JCVI_SCAF_1096627924300	+	948	1038	**RNA**→ OsmC (COG1764)? (GOS10947)→
env-41	JCVI_SCAF_1101668162688	-	740	661	**RNA**→
env-42	JCVI_SCAF_1101668607686	-	316	234	**RNA**→
env-43	JCVI_SCAF_1096627137814	-	4926	4846	**RNA**→
env-44	JCVI_SCAF_1096627366631	+	686	766	**RNA**→
env-45	JCVI_SCAF_1096627145484	+	5898	5977	**RNA**→
env-46	JCVI_SCAF_1096627362722	+	2940	3020	**RNA**→ Doc (COG3654)? (GOS12539)→
env-47	JCVI_SCAF_1096627146406	+	3593	3673	**RNA**→ ? (GOS11475520)→
env-48	JCVI_SCAF_1101668135096	-	893	813	**RNA**→
env-49	JCVI_SCAF_1101668135096	-	286	206	**RNA**→
env-50	JCVI_SCAF_1101668607686	-	1537	1457	**RNA**→ Doc (COG3654)? (GOS12539)→
env-51	JCVI_SCAF_1096627145484	+	4765	4845	**RNA**→ ? (GOS13075438)→ ? (GOS13130160)→
env-52	JCVI_SCAF_1096626373840	-	740	660	**RNA**→ Doc (COG3654)? (GOS12539)→
env-53	JCVI_SCAF_1096627362722	+	3766	3846	**RNA**→ ←DUF1080 (pfam06439)? (GOS7388)
env-54	JCVI_SCAF_1096627145484	+	1277	1357	**RNA**→ ? (GOS12064495)→ ? (GOS13313260)→
env-55	JCVI_SCAF_1101667177312	+	541	621	**RNA**→
env-56	JCVI_SCAF_1101667176249	+	419	499	**RNA**→
env-57	JCVI_SCAF_1096627366921	-	2282	2202	**RNA**→ ←? (GOS14648470)
env-58	JCVI_SCAF_1096627151317	+	2406	2487	**RNA**→ predicted protein→
env-59	JCVI_SCAF_1096627006799	+	346	426	**RNA**→ ? (GOS11217110)→ ? (GOS11057793)→
env-60	JCVI_SCAF_1096627147689	-	1263	1182	**RNA**→ ←COG4634 (COG4634)? (GOS13152162)
env-61	JCVI_SCAF_1096627147689	-	1928	1847	**RNA**→ ←COG4634 (COG4634)? (GOS13152162)
env-62	JCVI_SCAF_1096627931031	-	393	313	**RNA**→ ? (GOS13552515)→
env-63	JCVI_SCAF_1101668236600	-	528	448	**RNA**→
env-64	JCVI_SCAF_1096627145484	+	3334	3413	**RNA**→ ? (GOS13552515)→ ? (GOS12158022)→
env-65	JCVI_SCAF_1101668619749	-	845	766	**RNA**→ ? (GOS6844)→
env-66	JCVI_SCAF_1096627143814	+	2867	2946	**RNA**→ ? (GOS11217110)→ ? (GOS11057793)→

env-67	JCVI_SCAF_1101668607686	-	1139	1060	**RNA**→ Doc (COG3654)? (GOS12539)→
env-68	JCVI_SCAF_1096627644915	+	101	180	**RNA**→
env-69	JCVI_SCAF_1096627467316	+	1125	1204	**RNA**→ ? (GOS11113800)→ ? (GOS11246048)→
env-70	JCVI_SCAF_1096627644915	+	954	1033	**RNA**→
env-71	JCVI_SCAF_1101668236600	-	1237	1158	**RNA**→ ? (GOS11606920)→ ? (GOS13517746)→
env-72	JCVI_SCAF_1101667320291	+	593	672	**RNA**→
env-73	JCVI_SCAF_1096627650354	-	234	155	**RNA**→
env-74	JCVI_SCAF_1096627006799	+	1016	1097	**RNA**→ ? (GOS11217110)→ ? (GOS11057793)→
env-75	JCVI_SCAF_1096627138362	+	2545	2626	**RNA**→
env-76	ABEF01012645.1	+	329	410	**RNA**→ hypo→
Tde-1-1	NC_007404.1	-	1266332	1266274	**RNA**→ ←hypo
env-77	sludgePhrap_Contig5344	-	495	437	**RNA**→ ←Transposase and inactivated derivatives(,,,)
env-78	JCVI_SCAF_1096627149823	+	1429	1511	**RNA**→ ? (GOS13075438)→
env-79	gws2_d1_0050_10	-	5715	5598	**RNA**→ hypo→ Transposase_11 (pfam01609)→
env-80	gws2_d1_0074_21	+	23046	23164	**RNA**→ COG1569 (COG1569)→ hypo (cons)→
env-81	gws2_d1_0050_10	-	3184	3066	**RNA**→ ←hypo
env-82	gws2_d1_0532_1	-	1846	1728	**RNA**→ hypo (cons)→

81.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 81.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Rhodopirellula-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

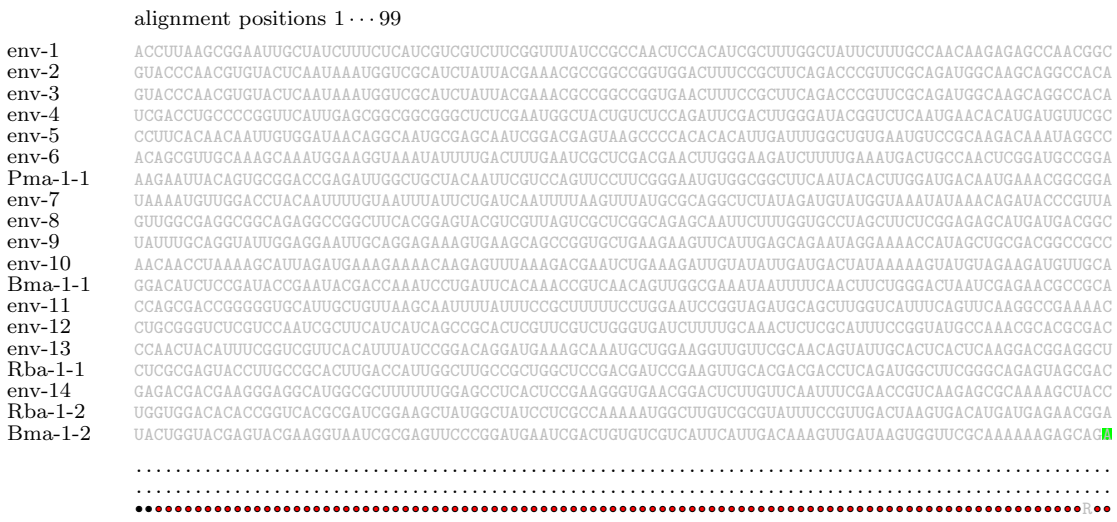
cd00138 (1) Phospholipase D.
cd00156 (1) Signal receiver domain; originally thought to be unique to bacteria (CheY, OmpR, NtrC, and PhoB), now recently identified in eukaryotes ETR1 Arabidopsis thaliana; this domain receives the signal from the sensor partner in a two-component systems; contains a phosphoacceptor site that is phosphorylated by histidine kinase homologs; usually found N-terminal to a DNA binding effector domain; forms homodimers
cd00204 (2) ankyrin repeats; ankyrin repeats mediate protein-protein interactions in very diverse families of proteins.
cd01193 (1) IntI (E2) integrases, site-specific tyrosine recombinases, DNA breaking-rejoining enzymes, N- and C-terminal domains.
cd02180 (1) Laminarinase, also known as glucan endo-1,3-beta-D-glucosidase, is a glycosyl hydrolase family 16 member that hydrolyzes 1,3-beta-D-glucosidic linkages in 1,3-beta-D-glucans such as laminarins, curdlans, paramylons, and pachymans, with very limited action on mixed-link (1,3-1,4-)-beta-D-glucans.
COG0189 (1) Glutathione synthase/Ribosomal protein S6 modification enzyme (glutaminyl transferase) [Coenzyme metabolism / Translation, ribosomal structure and biogenesis]
COG1569 (1) Predicted nucleic acid-binding protein, contains PIN domain [General function prediction only]
COG1733 (1) Predicted transcriptional regulators [Transcription]
COG1741 (1) Pirin-related protein [General function prediction only]
COG1764 (1) Predicted redox protein, regulator of disulfide bond formation [Posttranslational modification, protein turnover, chaperones]
COG2378 (1) Predicted transcriptional regulator [Transcription]
COG2826 (1) Transposase and inactivated derivatives, IS30 family [DNA replication, recombination, and repair]
COG2929 (2) Uncharacterized protein conserved in bacteria [Function unknown]
COG3472 (1) Uncharacterized conserved protein [Function unknown]
COG3576 (1) Predicted flavin-nucleotide-binding protein structurally related to pyridoxine 5'-phosphate oxidase [General function prediction only]
COG3654 (4) Prophage maintenance system killer protein [General function prediction only]
COG3666 (1) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
COG4634 (2) Uncharacterized protein conserved in bacteria [Function unknown]
COG4970 (3) Tfp pilus assembly protein FimT [Cell motility and secretion / Intracellular trafficking and secretion]
GOS10947 (1) ?
GOS11057793 (3) ?
GOS11113800 (1) ?
GOS11217110 (3) ?
GOS11246048 (1) ?
GOS11475520 (1) ?
GOS11558247 (1) ?
GOS11606920 (1) ?
GOS12064012 (1) ?
GOS12064495 (2) ?
GOS12158022 (1) ?
GOS12539 (4) ?
GOS13075438 (2) ?

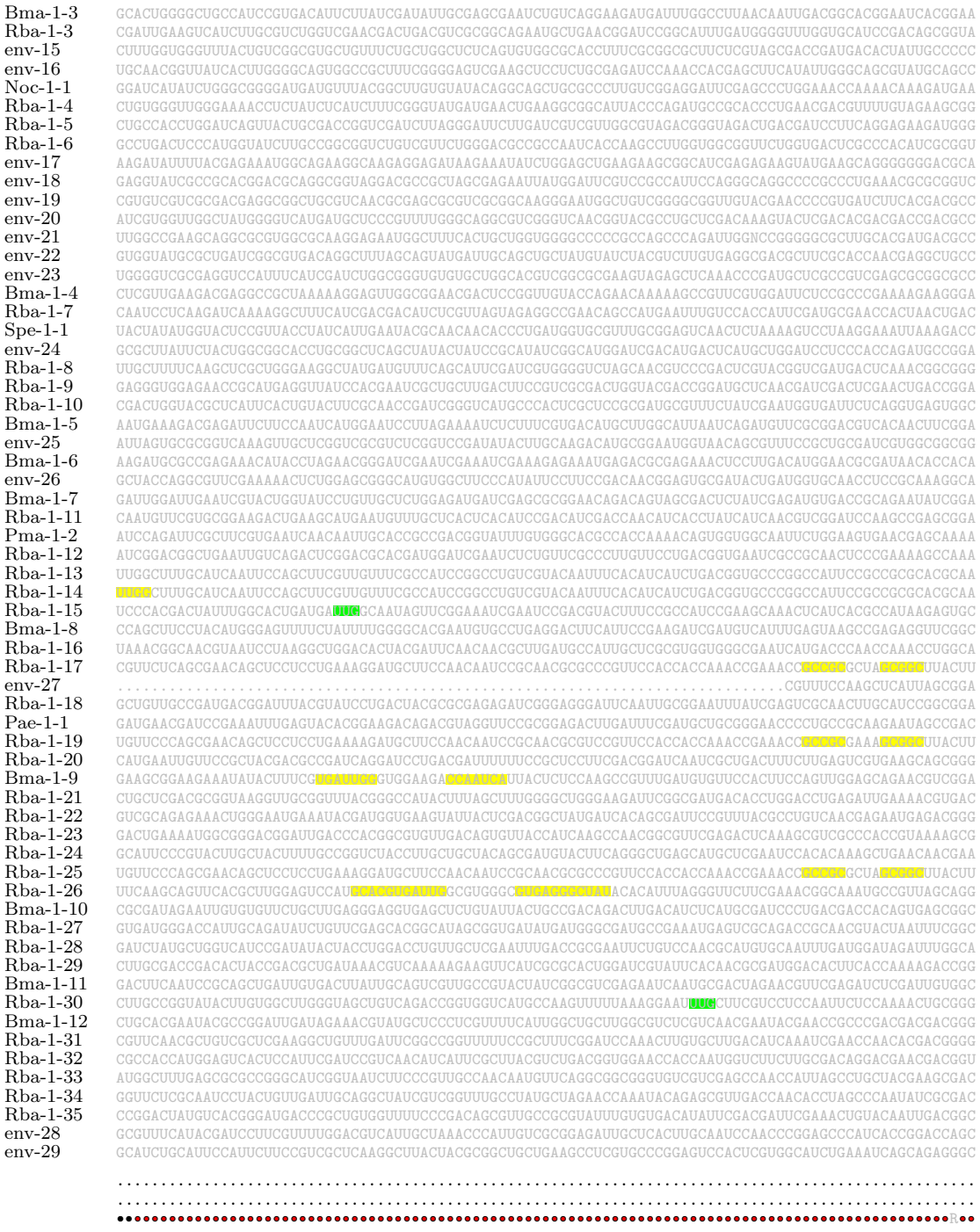
GOS13130160 (1) ?
GOS13152162 (2) ?
GOS13313260 (2) ?
GOS13517746 (1) ?
GOS13552515 (2) ?
GOS14648470 (1) ?
GOS5566 (1) ?
GOS5791 (1) ?
GOS6 (1) ?
GOS6844 (1) ?
GOS7388 (1) ?
NOG21735 (1)

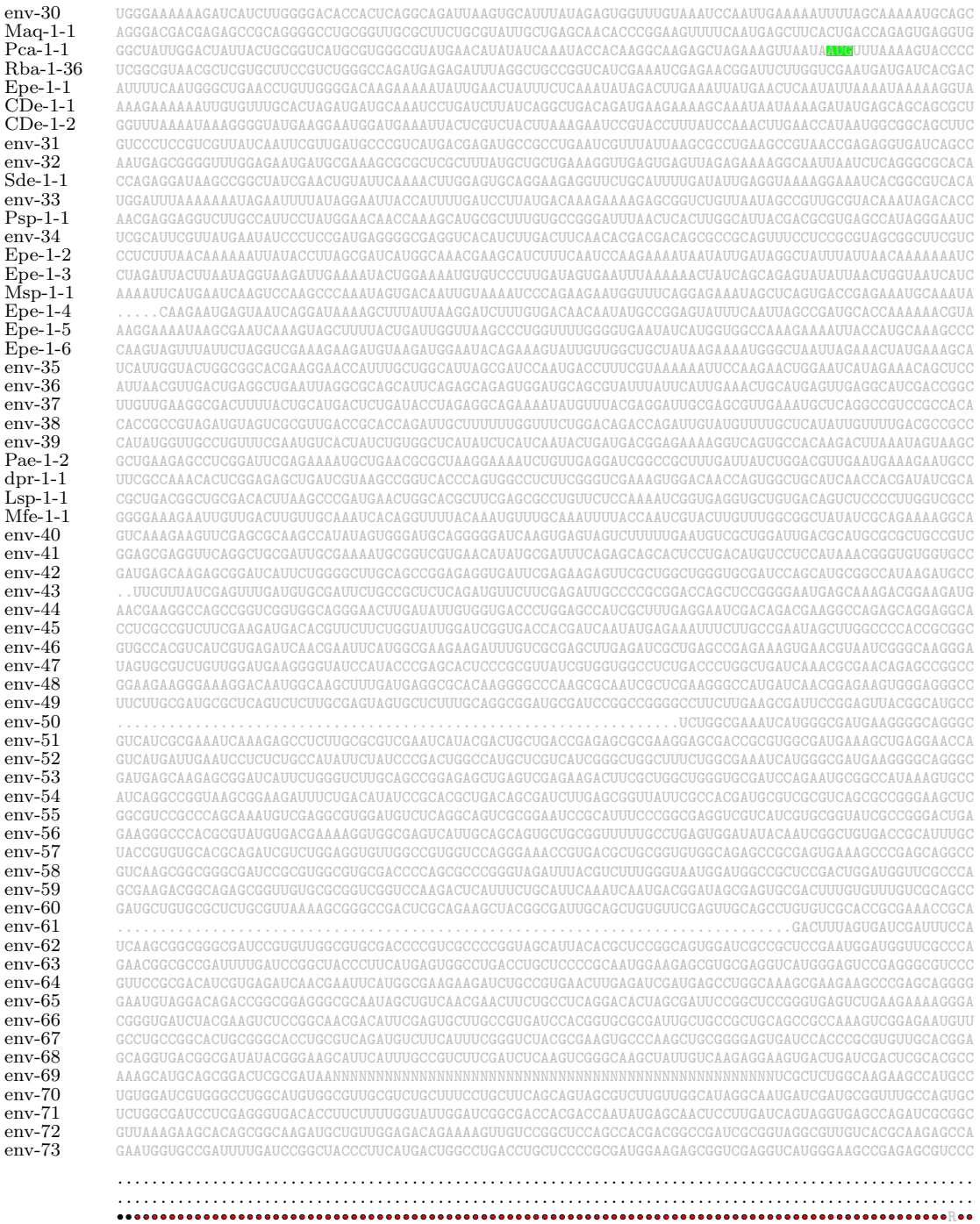
pfam00254 (1) FKBP-type peptidyl-prolyl cis-trans isomerase.
pfam00583 (1) Acetyltransferase (GNAT) family.
pfam00665 (1) Integrase core domain.
pfam00903 (2) Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily.
pfam01527 (1) Transposase.
pfam01609 (2) Transposase DDE domain.
pfam02018 (1) Carbohydrate binding domain.
pfam03235 (1) Protein of unknown function DUF262.
pfam04076 (1) Domain unknown function (DUF388).
pfam05635 (1) S23 ribosomal protein.
pfam06439 (1) Domain of Unknown Function (DUF1080).

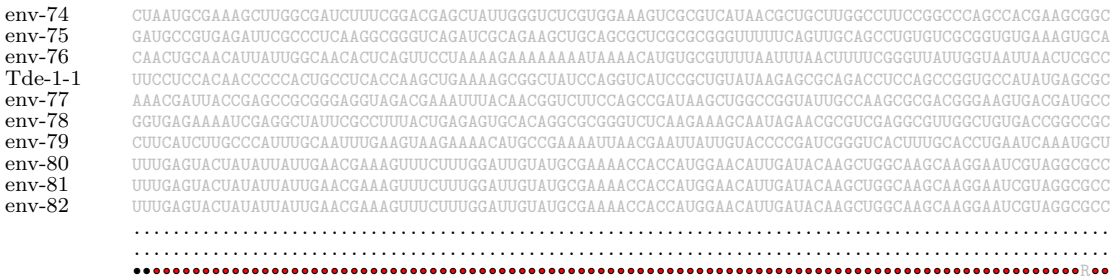
81.4 Multiple-sequence alignment

Each Rhodopirellula-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 81.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Rhodopirellula-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

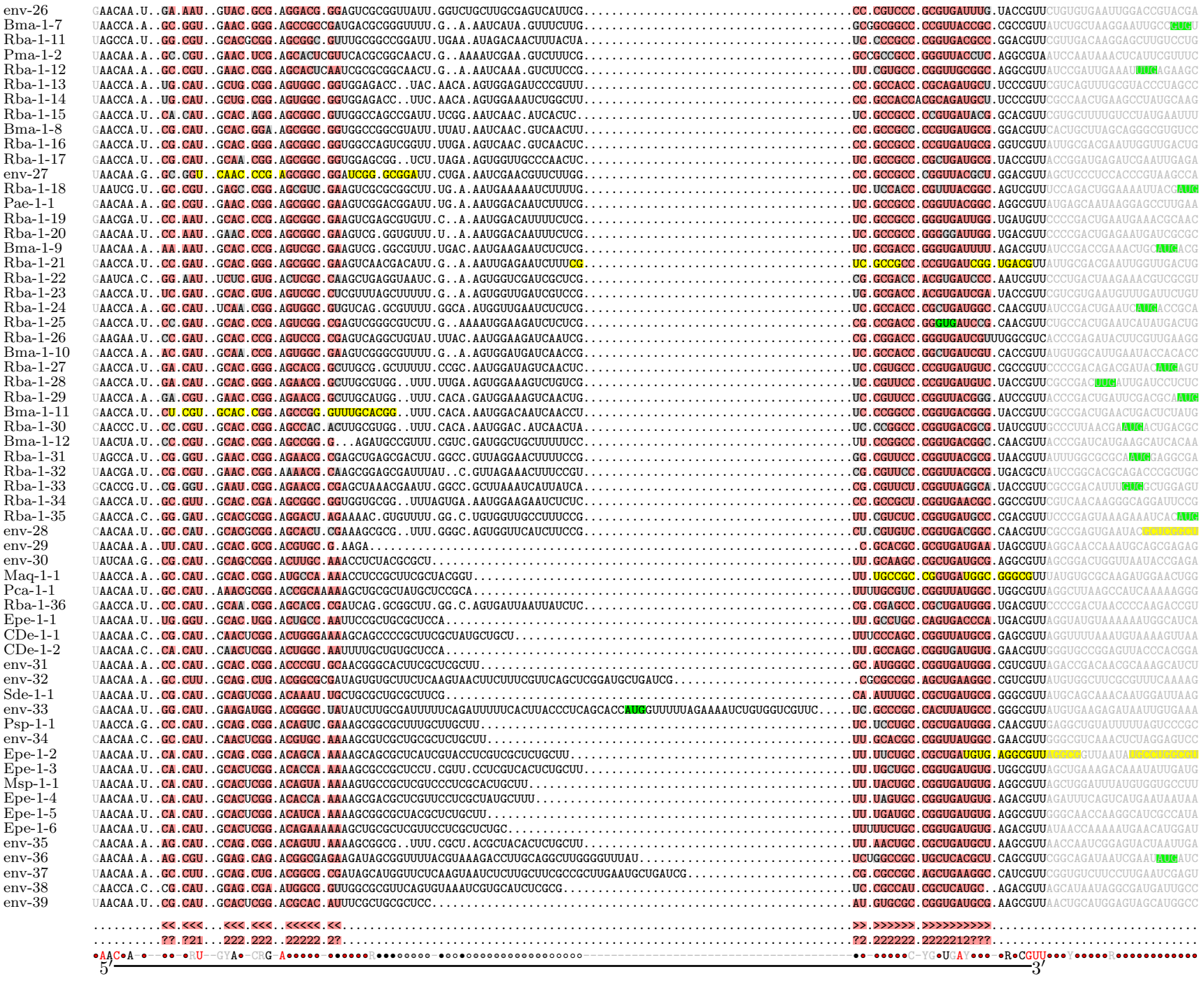



Bma-1-3
Rba-1-3
env-15
env-16
Noc-1-1
Rba-1-4
Rba-1-5
Rba-1-6
env-17
env-18
env-19
env-20
env-21
env-22
env-23
Bma-1-4
Rba-1-7
Spe-1-1
env-24
Rba-1-8
Rba-1-9
Rba-1-10
Bma-1-5
env-25
Bma-1-6
env-26
Bma-1-7
Rba-1-11
Pma-1-2
Rba-1-12
Rba-1-13
Rba-1-14
Rba-1-15
Bma-1-8
Rba-1-16
Rba-1-17
env-27
Rba-1-18
Pae-1-1
Rba-1-19
Rba-1-20
Bma-1-9
Rba-1-21
Rba-1-22
Rba-1-23
Rba-1-24
Rba-1-25
Rba-1-26
Bma-1-10
Rba-1-27
Rba-1-28
Rba-1-29
Bma-1-11
Rba-1-30
Bma-1-12
Rba-1-31
Rba-1-32
Rba-1-33
Rba-1-34
Rba-1-35
env-28
env-29

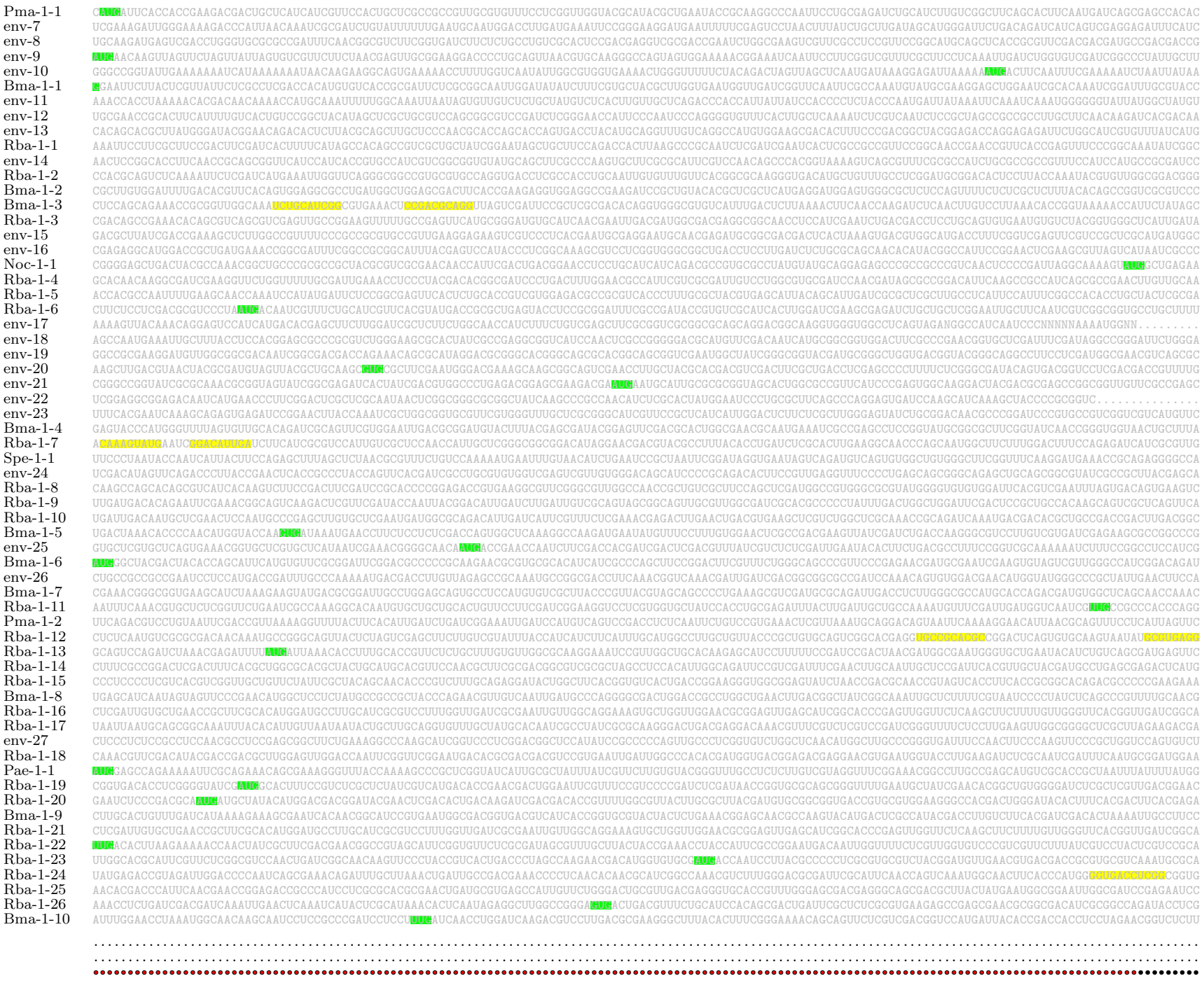

env-30
Maq-1-1
Pca-1-1
Rba-1-36
Epe-1-1
CDe-1-1
CDe-1-2
env-31
env-32
Sde-1-1
env-33
Psp-1-1
env-34
Epe-1-2
Epe-1-3
Msp-1-1
Epe-1-4
Epe-1-5
Epe-1-6
env-35
env-36
env-37
env-38
env-39
Pae-1-2
dpr-1-1
Lsp-1-1
Mfe-1-1
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73



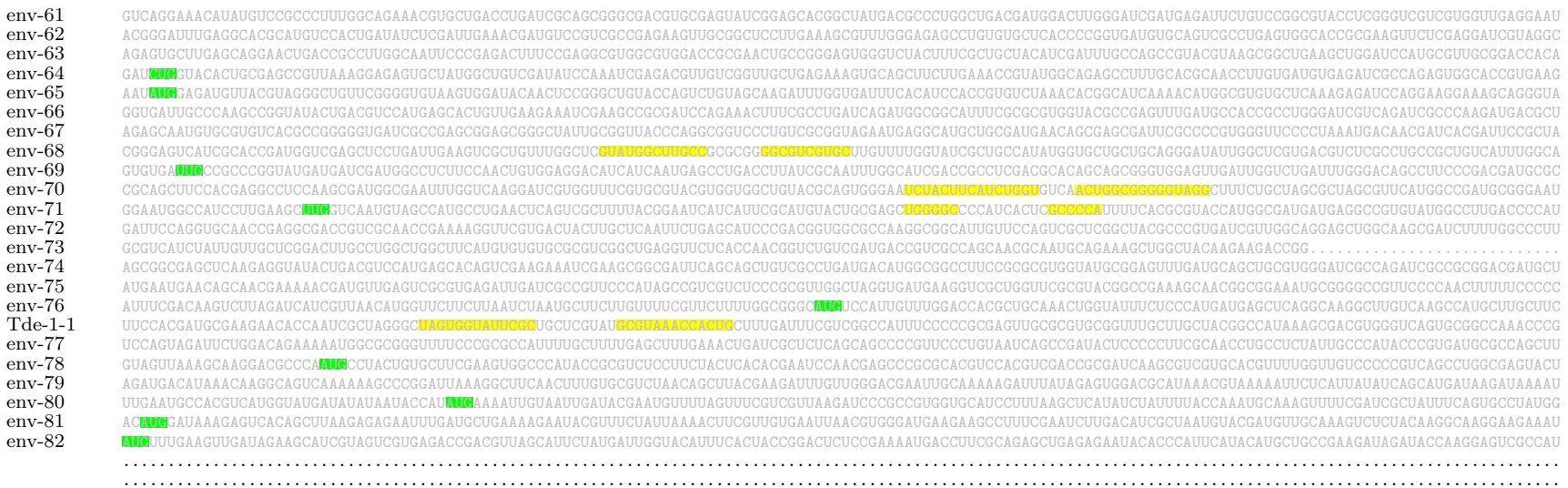
alignment positions 100 · · · 259


env-26
Bma-1-7
Rba-1-11
Pma-1-2
Rba-1-12
Rba-1-13
Rba-1-14
Rba-1-15
Bma-1-8
Rba-1-16
Rba-1-17
env-27
Rba-1-18
Pae-1-1
Rba-1-19
Rba-1-20
Bma-1-9
Rba-1-21
Rba-1-22
Rba-1-23
Rba-1-24
Rba-1-25
Rba-1-26
Bma-1-10
Rba-1-27
Rba-1-28
Rba-1-29
Bma-1-11
Rba-1-30
Bma-1-12
Rba-1-31
Rba-1-32
Rba-1-33
Rba-1-34
Rba-1-35
env-28
env-29
env-30
Maq-1-1
Pca-1-1
Rba-1-36
Epe-1-1
CDe-1-1
CDe-1-2
env-31
env-32
Sde-1-1
env-33
Psp-1-1
env-34
Epe-1-2
Epe-1-3
Msp-1-1
Epe-1-4
Epe-1-5
Epe-1-6
env-35
env-36
env-37
env-38
env-39
5′
3′

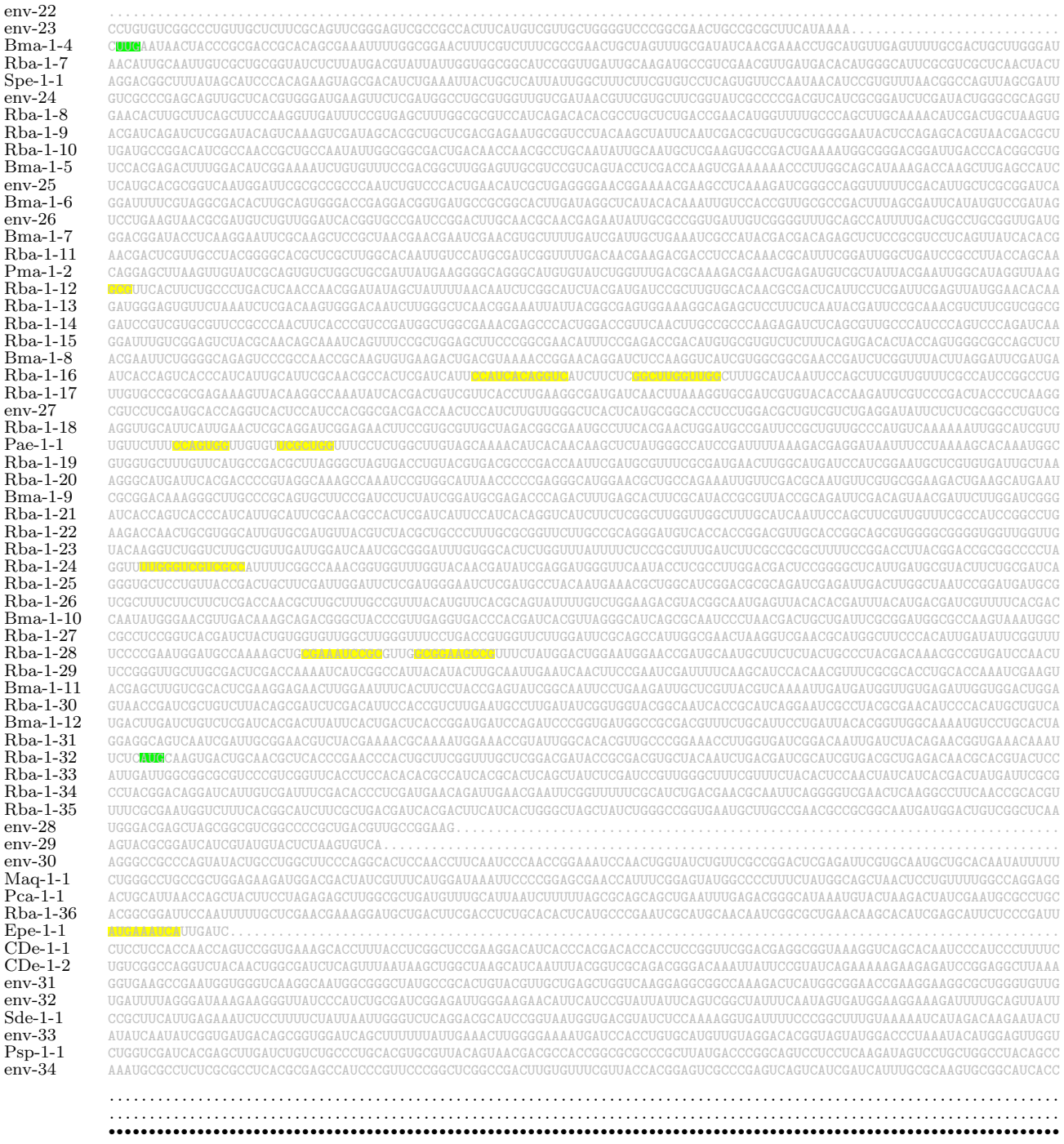
Pae-1-2
dpr-1-1
Lsp-1-1
Mfe-1-1
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
Tde-1-1
env-77
env-78
env-79
env-80
env-81
env-82
5′
3′
alignment positions 260 · · · 419
env-1
env-2
env-3
env-4
env-5
env-6


Pma-1-1
env-7
env-8
env-9
env-10
Bma-1-1
env-11
env-12
env-13
Rba-1-1
env-14
Rba-1-2
Bma-1-2
Bma-1-3
Rba-1-3
env-15
env-16
Noc-1-1
Rba-1-4
Rba-1-5
Rba-1-6
env-17
env-18
env-19
env-20
env-21
env-22
env-23
Bma-1-4
Rba-1-7
Spe-1-1
env-24
Rba-1-8
Rba-1-9
Rba-1-10
Bma-1-5
env-25
Bma-1-6
env-26
Bma-1-7
Rba-1-11
Pma-1-2
Rba-1-12
Rba-1-13
Rba-1-14
Rba-1-15
Bma-1-8
Rba-1-16
Rba-1-17
env-27
Rba-1-18
Pae-1-1
Rba-1-19
Rba-1-20
Bma-1-9
Rba-1-21
Rba-1-22
Rba-1-23
Rba-1-24
Rba-1-25
Rba-1-26
Bma-1-10

Rba-1-27
Rba-1-28
Rba-1-29
Bma-1-11
Rba-1-30
Bma-1-12
Rba-1-31
Rba-1-32
Rba-1-33
Rba-1-34
Rba-1-35
env-28
env-29
env-30
Maq-1-1
Pca-1-1
Rba-1-36
Epe-1-1
CDe-1-1
CDe-1-2
env-31
env-32
Sde-1-1
env-33
Psp-1-1
env-34
Epe-1-2
Epe-1-3
Msp-1-1
Epe-1-4
Epe-1-5
Epe-1-6
env-35
env-36
env-37
env-38
env-39
Pae-1-2
dpr-1-1
Lsp-1-1
Mfe-1-1
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60



alignment positions 420 · · · 537


env-22
env-23
Bma-1-4
Rba-1-7
Spe-1-1
env-24
Rba-1-8
Rba-1-9
Rba-1-10
Bma-1-5
env-25
Bma-1-6
env-26
Bma-1-7
Rba-1-11
Pma-1-2
Rba-1-12
Rba-1-13
Rba-1-14
Rba-1-15
Bma-1-8
Rba-1-16
Rba-1-17
env-27
Rba-1-18
Pae-1-1
Rba-1-19
Rba-1-20
Bma-1-9
Rba-1-21
Rba-1-22
Rba-1-23
Rba-1-24
Rba-1-25
Rba-1-26
Bma-1-10
Rba-1-27
Rba-1-28
Rba-1-29
Bma-1-11
Rba-1-30
Bma-1-12
Rba-1-31
Rba-1-32
Rba-1-33
Rba-1-34
Rba-1-35
env-28
env-29
env-30
Maq-1-1
Pca-1-1
Rba-1-36
Epe-1-1
CDe-1-1
CDe-1-2
env-31
env-32
Sde-1-1
env-33
Psp-1-1
env-34

Epe-1-2
Epe-1-3
Msp-1-1
Epe-1-4
Epe-1-5
Epe-1-6
env-35
env-36
env-37
env-38
env-39
Pae-1-2
dpr-1-1
Lsp-1-1
Mfe-1-1
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
Tde-1-1
env-77
env-78
env-79
env-80
env-81
env-82

82 *rmf*

82.1 Taxa

The taxonomy of each organism containing a putative *rmf* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *rmf* RNA in Sections 82.2 and 82.4.

abbrev. of hits	taxonomy of species
Avi-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Azotobacter vinelandii*
Pae-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* 2192
Pae-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* C3719
Pae-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PA7
Pae-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PACS2
Pae-5-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PAO1
Pae-6-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* UCBPP-PA14
Pen-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas entomophila* L48
Pfl-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf-5
Pfl-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf0-1
Pme-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas mendocina ymp*
Ppu-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* F1
Ppu-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* GB-1
Ppu-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* KT2440
Ppu-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* W619
Pst-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas stutzeri* A1501
Psy-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. phaseolicola* 1448A
Psy-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000
env-1 to env-3	environmental samples

82.2 Gene contexts

Each *rmf* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *rmf* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 82.3. The accession number of the sequence containing each *rmf* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 82.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Pae-3-1	NC_009656.1	-	2140830	2140703	**RNA**→ RMF (pfam04957)→
Pae-6-1	NC_008463.1	-	2155099	2154971	**RNA**→ RMF (pfam04957)→
Pae-5-1	NC_002516.2	+	3414255	3414382	**RNA**→ RMF (pfam04957)→
Pae-2-1	NZ_AAKV01000040.1	+	88056	88183	**RNA**→ ←PyrD (COG0167)
Pae-1-1	NZ_AAKW01000025.1	+	81584	81711	**RNA**→ ←PyrD (COG0167)
Pae-4-1	NZ_AAQW01000001.1	+	4117524	4117651	**RNA**→ ←PyrD (COG0167)
Avi-1-1	NZ_AAAU03000015.1	+	28165	28290	**RNA**→ RMF (pfam04957)→
Pst-1-1	NC_009434.1	+	2608193	2608318	**RNA**→ RMF (pfam04957)→

Psy-2-1	NC_007005.1	-	2459729	2459599	**RNA**→ RMF (pfam04957)→
Psy-3-1	NC_004578.1	-	2557709	2557579	**RNA**→ RMF (pfam04957)→
Psy-1-1	NC_005773.3	-	2438030	2437900	**RNA**→ RMF (pfam04957)→
Pme-1-1	NC_009439.1	+	3334674	3334807	**RNA**→ RMF (pfam04957)→
env-1	JCVI_SCAF_1096627046503	+	1032	1166	**RNA**→ RMF (pfam04957)? (GOS11125526)→
Ppu-1-1	NC_009512.1	+	4073813	4073942	**RNA**→ RMF (pfam04957)→
Ppu-3-1	NC_002947.3	-	2388767	2388638	**RNA**→ ←PyrD (COG0167)
Ppu-4-1	NZ_AAVY01000028.1	-	16492	16363	**RNA**→ RMF (pfam04957)→
Ppu-2-1	NZ_AAXR01000003.1	+	127284	127413	**RNA**→ RMF (pfam04957)→
Pen-1-1	NC_008027.1	-	1832021	1831892	**RNA**→ RMF (pfam04957)→
env-2	JCVI_SCAF_1096626875023	-	744	610	**RNA**→ RMF (pfam04957)? (GOS11125526)→
env-3	ABEF01018510.1	+	527	662	**RNA**→ hypo→
Pfl-1-1	NC_004129.6	-	2087729	2087597	**RNA**→ ←PyrD (COG0167)
Pfl-2-1	NC_007492.1	-	1989142	1989010	**RNA**→ RMF (pfam04957)→

82.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 82.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *rmf* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

COG0167 (5) Dihydroorotate dehydrogenase [Nucleotide transport and metabolism]
GOS11125526 (2) ?

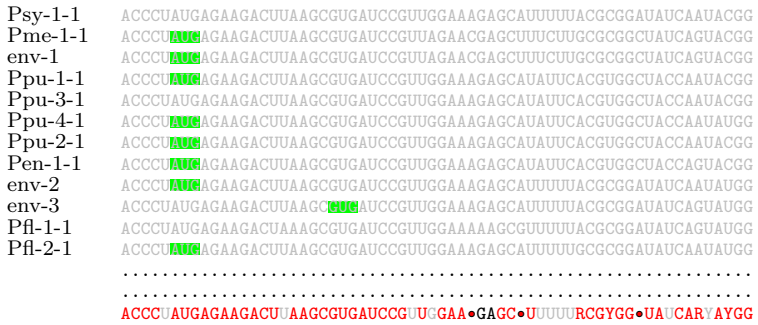
pfam04957 (16) Ribosome modulation factor.

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Pae-1-1, Pae-2-1, Pae-4-1

82.4 Multiple-sequence alignment

Each *rmf* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 82.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *rmf* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

alignment positions 1 · · · 98
Pae-3-1
Pae-6-1
Pae-5-1
Avi-1-1
Pst-1-1
Psy-2-1
Psy-3-1
Psy-1-1
Pme-1-1
env-1
Ppu-1-1
Ppu-3-1
Ppu-4-1
Ppu-2-1
Pen-1-1
env-2
env-3
Pfl-1-1
Pfl-2-1
alignment positions 99 · · · 258
Pae-3-1
Pae-6-1
Pae-5-1
Avi-1-1
Pst-1-1
Psy-2-1
Psy-3-1
Psy-1-1
Pme-1-1
env-1
Ppu-1-1
Ppu-3-1
Ppu-4-1
Ppu-2-1
Pen-1-1
env-2
env-3
Pfl-1-1
Pfl-2-1
5′
3′
alignment positions 259 · · · 322
Pae-3-1
Pae-6-1
Pae-5-1
Avi-1-1
Pst-1-1
Psy-2-1
Psy-3-1



83 *rne*-II

83.1 Taxa

The taxonomy of each organism containing a putative *rne*-II RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *rne*-II RNA in Sections 83.2 and 83.4.

abbrev. of hits	taxonomy of species
Avi-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Azotobacter vinelandii*
Pae-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* C3719
Pae-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PA7
Pae-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PACS2
Pae-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PAO1
Pae-5-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* UCBPP-PA14
Pen-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas entomophila* L48
Pfl-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf-5
Pfl-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf0-1
Pme-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas mendocina ymp*
Ppu-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* F1
Ppu-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* GB-1
Ppu-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* KT2440
Ppu-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* W619
Pst-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas stutzeri* A1501
Psy-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. phaseolicola* 1448A
Psy-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000
env-1 to env-3	environmental samples

83.2 Gene contexts

Each *rne*-II RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *rne*-II RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 83.3. The accession number of the sequence containing each *rne*-II RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present

on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 83.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Avi-1-1	NZ_AAAU03000001.1	-	1332545	1332276	**RNA**→ CafA (COG1530)→
Pst-1-1	NC_009434.1	+	2848416	2848682	**RNA**→ CafA (COG1530)→
Pae-5-1	NC_008463.1	-	2236725	2236458	**RNA**→ CafA (COG1530)→
Pae-4-1	NC_002516.2	+	3332603	3332870	**RNA**→ CafA (COG1530)→
Pae-3-1	NZ_AAQW01000001.1	+	4035886	4036153	**RNA**→ CafA (COG1530)→
Pae-1-1	NZ_AAKV01000040.1	+	6408	6675	**RNA**→ CafA (COG1530)→
Pae-2-1	NC_009656.1	-	2242576	2242309	**RNA**→ CafA (COG1530)→
Pen-1-1	NC_008027.1	-	1679634	1679264	**RNA**→ CafA (COG1530)→
Ppu-4-1	NZ_AAVY01000015.1	+	97770	98155	**RNA**→ CafA (COG1530)→
Ppu-2-1	NZ_AAXR01000003.1	+	274860	275231	**RNA**→ CafA (COG1530)→
Ppu-3-1	NC_002947.3	-	2150898	2150526	**RNA**→ CafA (COG1530)→
Ppu-1-1	NC_009512.1	+	4270107	4270479	**RNA**→ CafA (COG1530)→
env-1	JCVI_SCAF_1096628050550	-	447	174	**RNA**→
Pme-1-1	NC_009439.1	-	1785155	1784883	**RNA**→ CafA (COG1530)→
Psy-3-1	NC_004578.1	+	4346971	4347259	**RNA**→ CafA (COG1530)→
Psy-2-1	NC_007005.1	-	1845898	1845610	**RNA**→ CafA (COG1530)→
Psy-1-1	NC_005773.3	-	1897287	1896999	**RNA**→ CafA (COG1530)→
env-2	JCVI_SCAF_1101668580737	-	755	470	**RNA**→ S1 (smart00316)? (GOS1511)→
Pfl-1-1	NC_004129.6	-	1992152	1991866	**RNA**→ CafA (COG1530)→
env-3	JCVI_SCAF_1101668162775	+	1387	1654	**RNA**→
Pfl-2-1	NC_007492.1	+	4701518	4701785	**RNA**→ CafA (COG1530)→

83.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 83.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *rne*-II RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

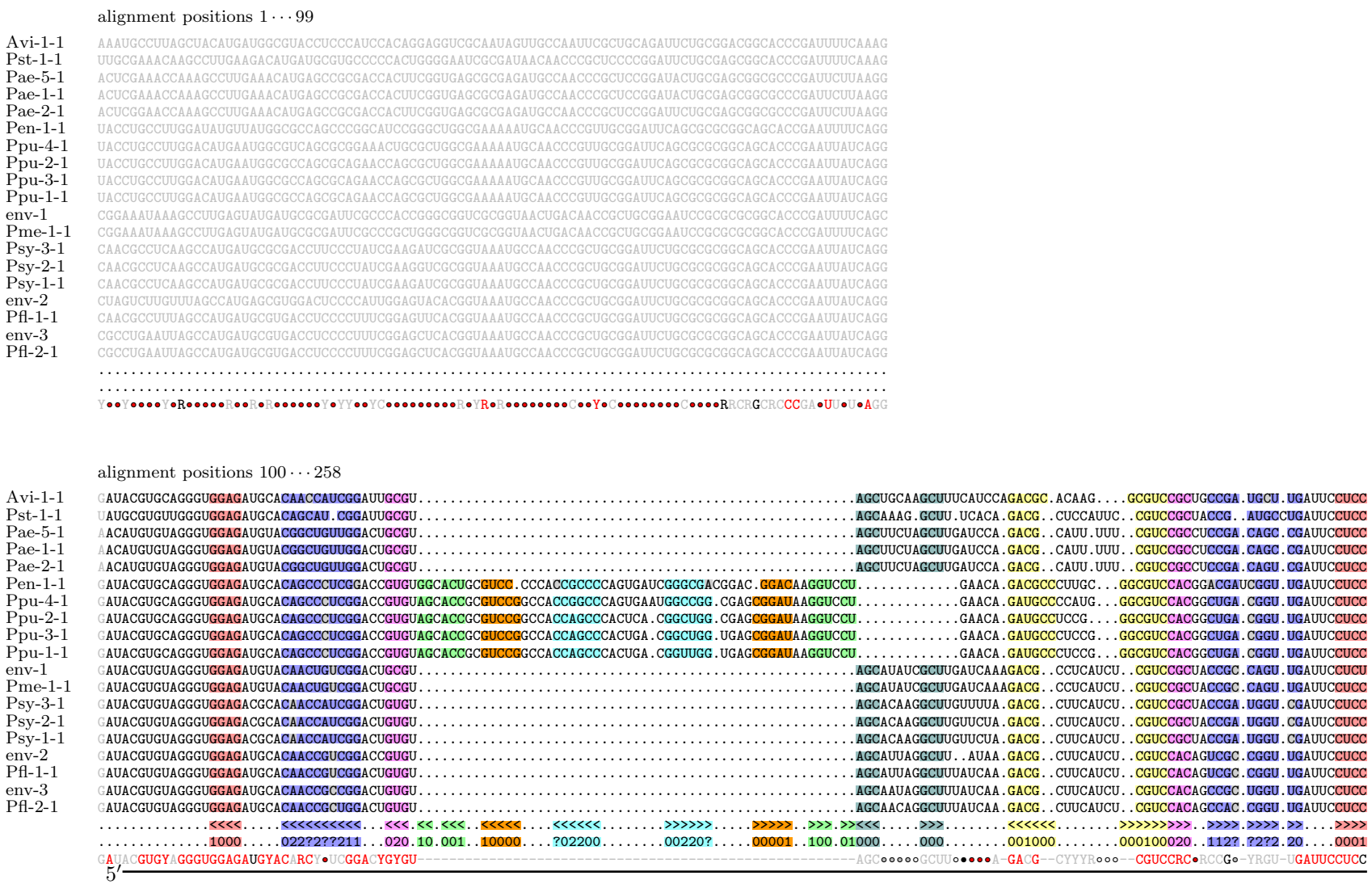
COG1530 (18) Ribonucleases G and E [Translation, ribosomal structure and biogenesis]
GOS1511 (1) ?
smart00316 (1) Ribosomal protein S1-like RNA-binding domain;

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Pae-3-1, Pae-4-1

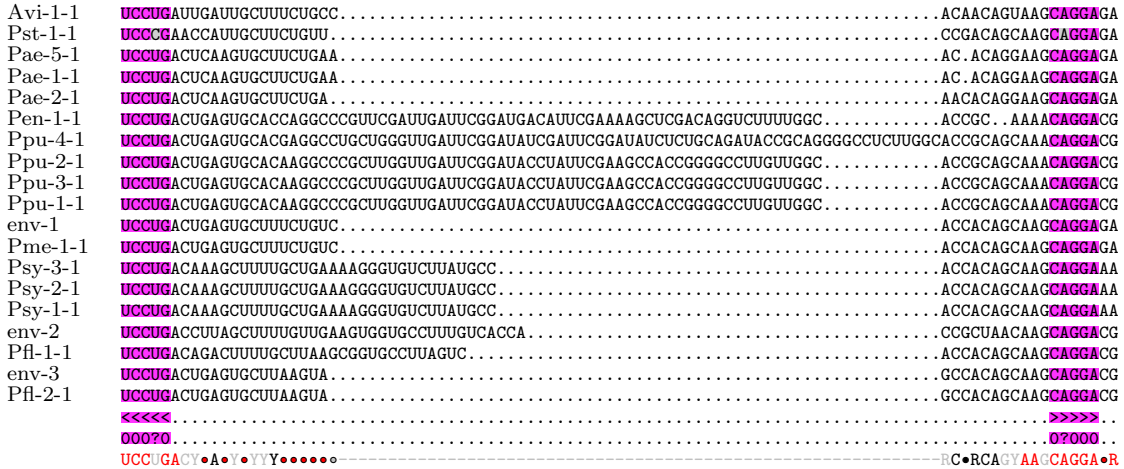
83.4 Multiple-sequence alignment

Each *rne*-II RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 83.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *rne*-II RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-

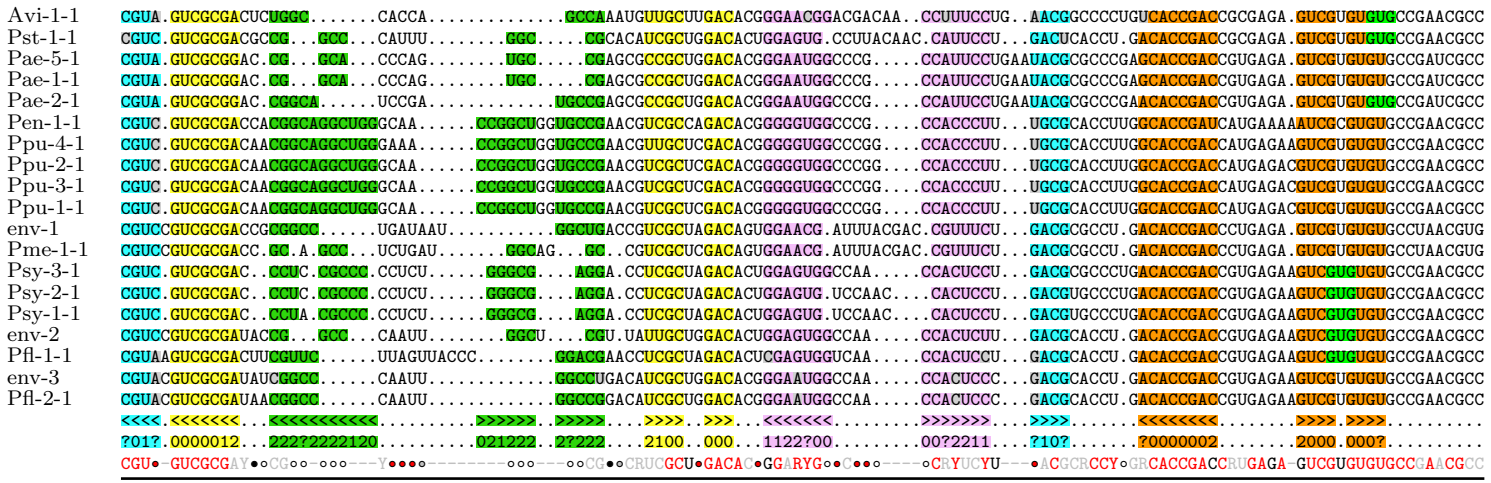
Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



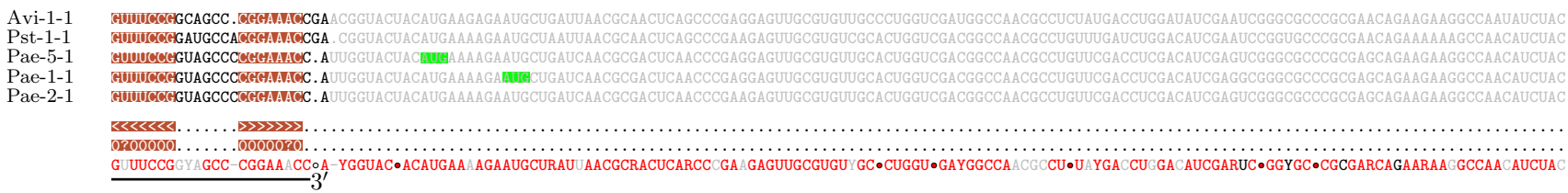
alignment positions 259 · · · 359

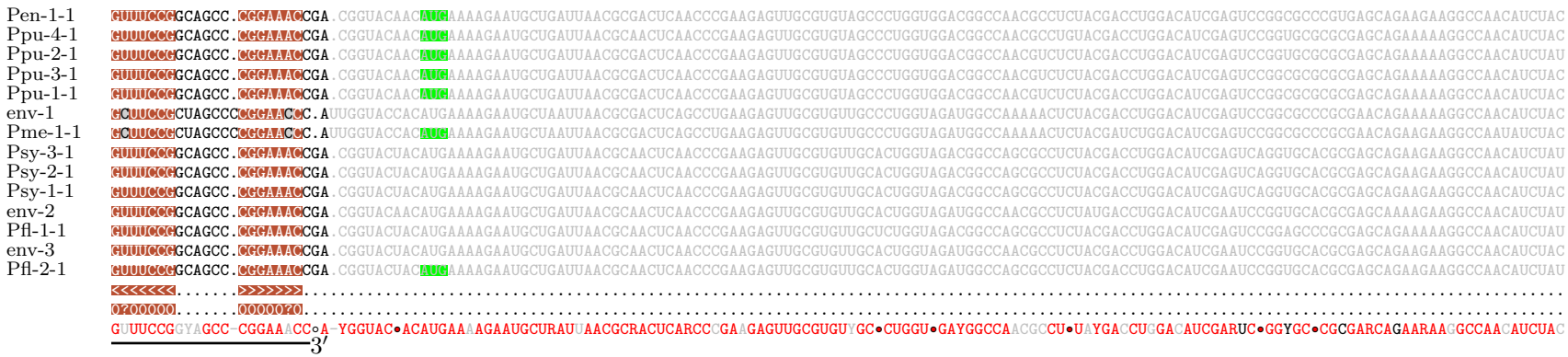


alignment positions 360 · · · 497

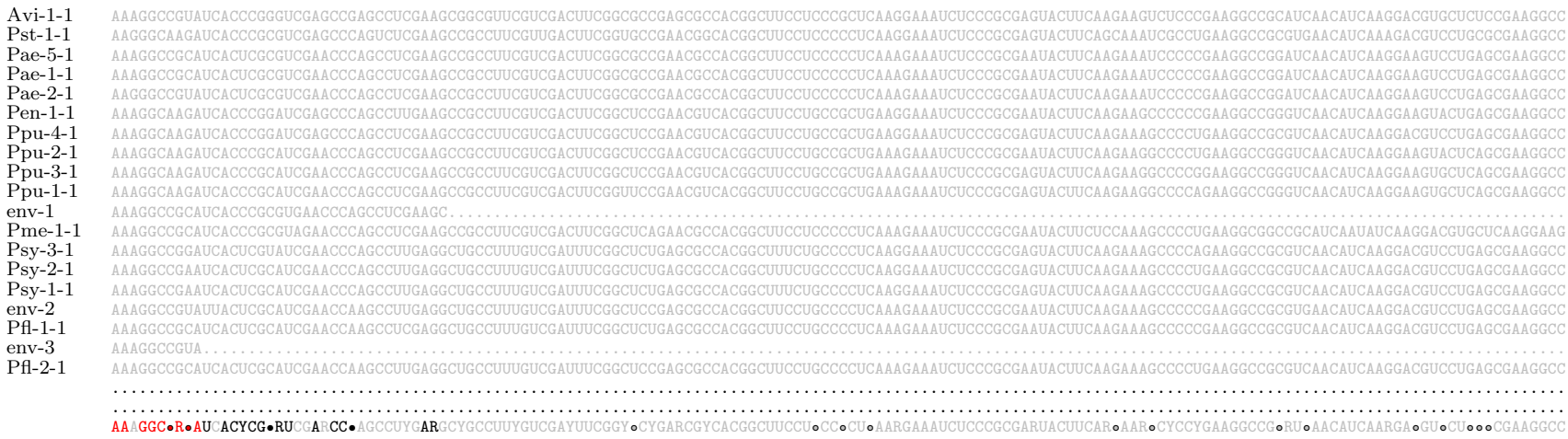


alignment positions 498 · · · 657

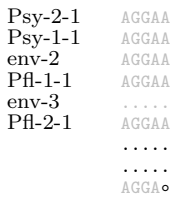




alignment positions 658 · · · 817



alignment positions 818 · · · 822



84 SAM-Chlorobi

84.1 Taxa

The taxonomy of each organism containing a putative SAM-Chlorobi RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual SAM-Chlorobi RNA in Sections 84.2 and 84.4.

abbrev. of hits	taxonomy of species
Cte-1-1	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobaculum *Chlorobium tepidum* TLS
Cch-1-1	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium chlorochromatii* CaD3
Cfe-1-1	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium ferrooxidans* DSM 13031
Cli-1-1	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium limicola* DSM 245
Cph-1-1	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium phaeobacteroides* BS1
Cph-2-1	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium phaeobacteroides* DSM 266
Cph-3-1	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Chlorobium phaeovibrioides* DSM 265 (Prosthecochloris vibrioformis
Plu-1-1	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Pelodictyon luteolum* DSM 273
Pph-1-1	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae Chlorobium/Pelodictyon group *Pelodictyon phaeoclathratiforme* BU-1
Pae-1-1	Bacteria Chlorobi Chlorobia Chlorobiales Chlorobiaceae *Prosthecochloris aestuarii* DSM 271

84.2 Gene contexts

Each SAM-Chlorobi RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one SAM-Chlorobi RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 84.3. The accession number of the sequence containing each SAM-Chlorobi RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 84.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Cch-1-1	NC_007514.1	+	2027693	2027794	**RNA**→ MetK (COG0192)→
Pae-1-1	NZ_AAIJ01000005.1	-	72374	72279	**RNA**→ MetK (COG0192)→ AdoHcyase (cd00401)→
Cph-1-1	NZ_AAIC01000035.1	+	8811	8905	**RNA**→ MetK (COG0192)→ AdoHcyase (cd00401)→
Cph-2-1	NC_008639.1	-	1051404	1051314	**RNA**→ MetK (COG0192)→
Cfe-1-1	NZ_AASE01000037.1	-	8265	8171	**RNA**→ MetK (COG0192)→ AdoHcyase (pfam05221)→
Pph-1-1	NZ_AAIK01000059.1	+	6046	6141	**RNA**→ MetK (COG0192)→ AdoHcyase (pfam05221)→
Cph-3-1	NC_009337.1	+	1264526	1264622	**RNA**→ MetK (COG0192)→

Cte-1-1	NC_002932.3	-	705942	705847	**RNA**→ hypo→ MetK (COG0192)→ AdoHcyase (cd00401)→
Cli-1-1	NZ_AAHJ01000034.1	+	5797	5892	**RNA**→ ←hypo
Plu-1-1	NC_007512.1	-	776880	776779	**RNA**→ MetK (COG0192)→

84.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 84.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one SAM-Chlorobi RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

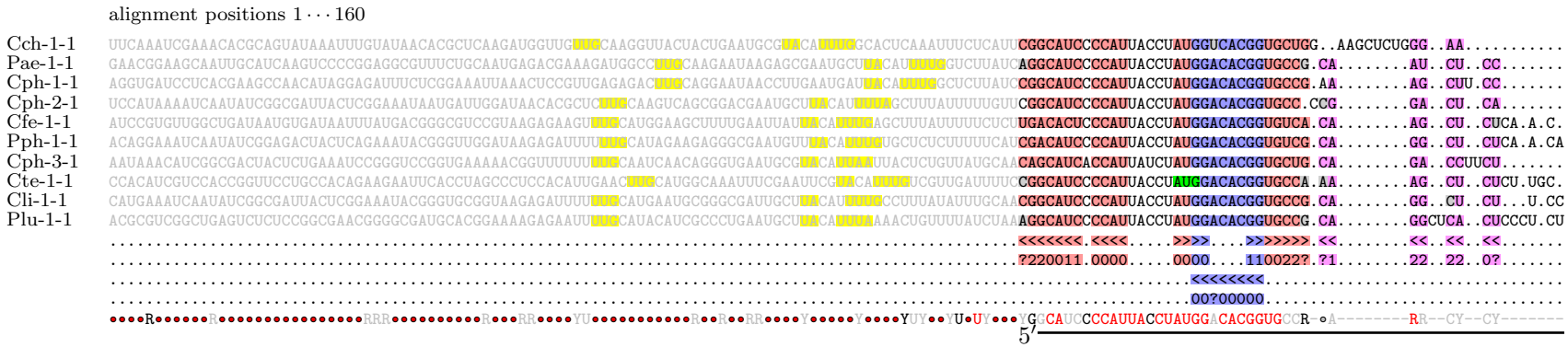
cd00401 (3) S-adenosyl-L-homocysteine hydrolase (AdoHycase) catalyzes the hydrolysis of S-adenosyl-L-homocysteine (AdoHyc) to form adenosine (Ado) and homocysteine (Hcy).

COG0192 (9) S-adenosylmethionine synthetase [Coenzyme metabolism]
pfam05221 (2) S-adenosyl-L-homocysteine hydrolase.

84.4 Multiple-sequence alignment

Each SAM-Chlorobi RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 84.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The SAM-Chlorobi RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (◦): 75%, white circle (◦): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

(Features specific to this RNA motif) Predicted Bacteroidetes-like promoter sequences upstream of SAM-Chlorobi RNAs are highlighted in yellow. They follow the consensus TTG ... TAnnnTTTG (although T residues are written as 'U' in this RNA-centric alignment).



85 SAM-I-IV-variant

85.1 Taxa

The taxonomy of each organism containing a putative SAM-I-IV-variant RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual SAM-I-IV-variant RNA in Sections 85.3 and 85.5.

abbrev. of hits	**taxonomy of species**
Aba-1-1	Bacteria Acidobacteria Acidobacteriales Acidobacteriaceae *Acidobacteria bacterium* Ellin345
Fba-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales *Flavobacteria bacterium* BAL38
Cat-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Croceibacter atlanticus* HTCC2559
Psp-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Polaribacter sp.* MED152
Pto-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Psychroflexus torquis* ATCC 700755
Rbi-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Robiginitalea biformata* HTCC2501
Lar-1-1	Bacteria Lentisphaerae Lentisphaerales Lentisphaeraceae *Lentisphaera araneosa* HTCC2155
Bab-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Brucellaceae *Brucella abortus bv.* 1 str. 9-941
Bme-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Brucellaceae *Brucella melitensis* 16M
Bme-2-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Brucellaceae *Brucella melitensis biovar* Abortus 2308
Bov-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Brucellaceae *Brucella ovis* ATCC 25840
Bsu-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Brucellaceae *Brucella suis* 1330
Oan-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Brucellaceae *Ochrobactrum anthropi* ATCC 49188
Lag-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Labrenzia aggregata* IAM 12614
Rsp-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Roseobacter sp.* AzwK-3b
Acr-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodospirillales Acetobacteraceae *Acidiphilium cryptum* JF-5
Nar-1-1	Bacteria Proteobacteria Alphaproteobacteria Sphingomonadales Sphingomonadaceae *Novosphingobium aromaticivorans* DSM 12444
Swi-1-1	Bacteria Proteobacteria Alphaproteobacteria Sphingomonadales Sphingomonadaceae *Sphingomonas wittichii* RW1
Bce-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* HI2424
Bce-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* MC0-3
Bce-3-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* PC184
Bsp-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia sp.* 383
Pna-1-1 to Pna-1-3	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae *Polaromonas naphthalenivorans* CJ2
Ngo-1-1	Bacteria Proteobacteria Betaproteobacteria Neisseriales Neisseriaceae *Neisseria gonorrhoeae* FA 1090
Nme-1-1	Bacteria Proteobacteria Betaproteobacteria Neisseriales Neisseriaceae *Neisseria meningitidis* FAM18
Nme-2-1	Bacteria Proteobacteria Betaproteobacteria Neisseriales Neisseriaceae *Neisseria meningitidis* MC58
Nme-3-1	Bacteria Proteobacteria Betaproteobacteria Neisseriales Neisseriaceae *Neisseria meningitidis* Z2491
Dac-1-1	Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Desulfuromonadaceae *Desulfuromonas acetoxidans* DSM 684
Mxa-1-1	Bacteria Proteobacteria Deltaproteobacteria Myxococcales Cystobacterineae Myxococcaceae *Myxococcus xanthus* DK 1622
Cli-1-1 to Cli-1-2	Bacteria Proteobacteria Gammaproteobacteria *Congregibacter litoralis* KT71
Hch-1-1	Bacteria Proteobacteria Gammaproteobacteria Oceanospirillales Hahellaceae *Hahella chejuensis* KCTC 2396
Nca-1-1	Bacteria Proteobacteria Gammaproteobacteria Oceanospirillales *Neptuniibacter caesariensis*
env-1 to env-438	environmental samples

85.2 Notes

Superscript numbers are used to annotate any SAM-I-IV-variant RNA that has special characteristics (described below). These numeric annotations will be used in Sections 85.3 and 85.5. The meaning of the numbers are as follows.

[1] This sequence might have an internal loop, although it was not found by a homology search of SAM-I or SAM-IV riboswitches (both of which contain internal loops in P2).

85.3 Gene contexts

Each SAM-I-IV-variant RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one SAM-I-IV-variant RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 85.4. The accession number of the sequence containing each SAM-I-IV-variant RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 85.1. Superscript numbers refer to annotations listed in Section 85.2

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	ABEF01044326.1	+	1056	1176	**RNA**→ hypo→
env-2	JCVI_SCAF_1096627944084	+	868	1007	**RNA**→ ? (GOS11176883)→
env-3	JCVI_SCAF_1096627022509	-	9550	9417	**RNA**→ ←? (GOS11954949)
env-4	JCVI_SCAF_1096627517752	+	199	326	**RNA**→ ←? (GOS11954949)
env-5	JCVI_SCAF_1096627382769	+	6070	6201	**RNA**→ MetK (COG0192)? (GOS4376)→
Aba-1-1	NC_008009.1	-	5414708	5414571	**RNA**→ MetK (COG0192)→
env-6	JCVI_SCAF_1096627472915	+	170	325	**RNA**→
env-7	JCVI_SCAF_1096628350211	-	1267	1114	**RNA**→ predicted protein→ thiolase (cd00751)? (GOS3111)→
Lar-1-1	NZ_ABCK01000024.1	+	45217	45334	**RNA**→ FGase (pfam05013)→ hypo→
env-8	ABEF01007799.1	+	1141	1268	**RNA**→
env-9	soil_XZS33110_y1	+	600	756	**RNA**→ hypo→
Dac-1-1	NZ_AAEW02000015.1	+	42526	42673	**RNA**→ MET17 (COG2873)→ PSRA_1 (cd02558)→
env-10	JCVI_SCAF_1096627669635	+	1192	1296	**RNA**→
env-11	JCVI_SCAF_1096627669171	-	440	305	**RNA**→
env-12	JCVI_SCAF_1096627022492	-	1342	1208	**RNA**→ ? (GOS11785628)→
env-13	JCVI_SCAF_1096627017618	+	887	1021	**RNA**→ SEL1 (smart00671)? (GOS697)→ ? (GOS11041706)→
[1]env-14	JCVI_SCAF_1096627453466	-	317	190	**RNA**→
env-15	JCVI_SCAF_1096627674920	-	305	190	**RNA**→
env-16	JCVI_SCAF_1101667310288	+	198	320	**RNA**→ COG0714 (COG0714)? (GOS14278197)→
Cat-1-1	NZ_AAMP01000002.1	-	208403	208291	**RNA**→ ←hypo
env-17	wf3_Contig7109	-	1398	1286	**RNA**→ NhaC (COG1757)→
env-18	JCVI_SCAF_1096627112481	-	1114	962	**RNA**→
env-19	JCVI_SCAF_1101667169972	-	313	197	**RNA**→
env-20	JCVI_SCAF_1096627015080	+	742	870	**RNA**→ MET2 (COG2021)? (GOS3619)→
env-21	JCVI_SCAF_1096627016816	+	301	428	**RNA**→ MET2 (COG2021)? (GOS3619)→
env-22	JCVI_SCAF_1096627415864	+	109	261	**RNA**→
env-23	JCVI_SCAF_1101667194498	-	309	189	**RNA**→
env-24	soil_XZS67444_y1	-	249	123	**RNA**→
Mxa-1-1	NC_008095.1	-	1123541	1123405	**RNA**→ ←hypo
env-25	JCVI_SCAF_1096627148500	-	717	610	**RNA**→ ? (GOS11697307)→
env-26	JCVI_SCAF_1096627960261	-	1417	1310	**RNA**→ ←? (GOS11701787)
env-27	JCVI_SCAF_1096627133175	+	279	386	**RNA**→ ←? (GOS11701787)
env-28	JCVI_SCAF_1101667507259	+	98	224	**RNA**→
env-29	JCVI_SCAF_1096627090946	+	987	1105	**RNA**→ FGase (pfam05013)? (GOS7466)→
env-30	JCVI_SCAF_1101668135845	+	1231	1342	**RNA**→
env-31	JCVI_SCAF_1096627156733	+	234	345	**RNA**→ ? (GOS13297217)→
env-32	ABEF01017161.1	-	942	808	**RNA**→ hypo→
env-33	soil_Contig37056	-	748	609	**RNA**→ Homoserine acetyltransferase(EC:2.3.1.31)→
env-34	JCVI_SCAF_1096627458374	-	911	767	**RNA**→ MET2 (COG2021)? (GOS3619)→
Hch-1-1	NC_007645.1	-	783086	782963	**RNA**→ hypo→
env-35	JCVI_SCAF_1096626883249	+	723	820	**RNA**→

env-36	JCVI_SCAF_1096624996060	-	852	740	**RNA**→
env-37	JCVI_SCAF_1101668216426	+	925	1032	**RNA**→ ? (GOS18368)→
env-38	JCVI_SCAF_1096627738665	-	872	765	**RNA**→ ←FAD_binding_7 (pfam03441)? (GOS1285)
env-39	JCVI_SCAF_1096627969816	+	120	227	**RNA**→ ? (GOS11151719)→
env-40	JCVI_SCAF_1096626941806	-	1332	1226	**RNA**→ ? (GOS11151719)→
Fba-1-1	NZ_AAXX01000006.1	-	136094	135985	**RNA**→ hypo→ hypo→ hypo→
env-41	JCVI_SCAF_1096627706058	+	152	306	**RNA**→ ? (GOS18368)→
env-42	JCVI_SCAF_1096628183775	+	310	465	**RNA**→ ? (GOS18368)→
env-43	JCVI_SCAF_1096627526279	+	340	444	**RNA**→ ? (GOS18368)→
env-44	JCVI_SCAF_1096627570996	+	134	238	**RNA**→ ? (GOS18368)→
env-45	JCVI_SCAF_1096627732474	-	688	583	**RNA**→ ? (GOS18368)→
env-46	JCVI_SCAF_1096627304127	+	547	653	**RNA**→ ? (GOS18368)→
env-47	JCVI_SCAF_1096627254203	+	3046	3152	**RNA**→ ? (GOS18368)→
env-48	JCVI_SCAF_1096627505884	+	353	458	**RNA**→ ? (GOS18368)→
env-49	JCVI_SCAF_1101668173975	-	990	884	**RNA**→ ? (GOS18368)→
env-50	JCVI_SCAF_1101668181848	+	754	860	**RNA**→ ? (GOS18368)→
env-51	JCVI_SCAF_1101668374940	+	257	363	**RNA**→ ? (GOS18368)→ ? (GOS13245863)→
env-52	JCVI_SCAF_1096628011786	+	472	594	**RNA**→ ? (GOS18368)→ ? (GOS14602209)→ HTH_MlrA (cd01104)? (GOS606)→
env-53	JCVI_SCAF_1096627565360	-	619	514	**RNA**→ ? (GOS18368)→
env-54	JCVI_SCAF_1096627154973	+	171	277	**RNA**→ ? (GOS18368)→
Rbi-1-1	NZ_AAOI01000001.1	+	638994	639099	**RNA**→ hypo→ hypo→
Psp-1-1	NZ_AANA01000001.1	+	1651504	1651612	**RNA**→ hypo→ hypo→ hypo→
env-55	JCVI_SCAF_1101668246057	+	681	795	**RNA**→
env-56	JCVI_SCAF_1096627662020	-	1463	1354	**RNA**→ COG2340 (COG2340)? (GOS14414208)→
env-57	soil_Contig16206	-	348	230	**RNA**→
env-58	JCVI_SCAF_1096627273291	-	956	821	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-59	JCVI_SCAF_1101668527817	+	280	406	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-60	BAAX01001997.1	+	1769	1874	**RNA**→
env-61	BAAX01003673.1	-	1052	934	**RNA**→
env-62	JCVI_SCAF_1096627685829	+	932	1035	**RNA**→ MetK (COG0192)? (GOS4376)→
env-63	JCVI_SCAF_1096626931677	+	1258	1361	**RNA**→ MetK (COG0192)? (GOS4376)→
env-64	JCVI_SCAF_1101668372096	+	571	674	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-65	JCVI_SCAF_1096626928819	+	695	798	**RNA**→ MetK (COG0192)? (GOS4376)→ COG0780 (COG0780)? (GOS7114)→
env-66	JCVI_SCAF_1096626944644	+	425	528	**RNA**→ MetK (COG0192)? (GOS4376)→ COG0780 (COG0780)? (GOS7114)→ ? (GOS9454)→
env-67	JCVI_SCAF_1101668320074	-	650	547	**RNA**→ MetK (COG0192)? (GOS4376)→
env-68	JCVI_SCAF_1101668370130	-	750	647	**RNA**→ MetK (COG0192)? (GOS4376)→
env-69	JCVI_SCAF_1096626894466	+	855	967	**RNA**→ ? (GOS11074455)→ ? (GOS13205716)→ ? (GOS13085086)→ ? (GOS13055046)→
env-70	JCVI_SCAF_1096627863184	-	503	390	**RNA**→ ? (GOS12011754)→
env-71	JCVI_SCAF_1096626913066	-	866	753	**RNA**→ ? (GOS11074455)→
env-72	JCVI_SCAF_1101668741091	+	780	892	**RNA**→ ? (GOS13476120)→
env-73	JCVI_SCAF_1101667347443	+	322	435	**RNA**→ ? (GOS12011754)→
env-74	JCVI_SCAF_1096626949907	-	2258	2145	**RNA**→ ? (GOS11879195)→ ? (GOS13085086)→ ? (GOS13055046)→ ? (GOS13327453)→
env-75	JCVI_SCAF_1096628136528	+	1361	1461	**RNA**→
env-76	JCVI_SCAF_1101668314905	+	1049	1149	**RNA**→
env-77	JCVI_SCAF_1096627489354	-	1374	1261	**RNA**→ ? (GOS11074455)→ ? (GOS11074455)→ ? (GOS13085086)→
env-78	JCVI_SCAF_1101668257290	-	987	874	**RNA**→ ? (GOS11074455)→
env-79	JCVI_SCAF_1101668280796	+	207	320	**RNA**→ ? (GOS11074455)→
env-80	JCVI_SCAF_1101667266397	+	406	518	**RNA**→ ? (GOS8636)→
env-81	JCVI_SCAF_1101668167402	-	671	559	**RNA**→ SEST_like (cd01823)? (GOS8636)→
env-82	JCVI_SCAF_1096627716901	+	259	371	**RNA**→ ? (GOS8636)→ ? (GOS8636)→ vWFA (cd00198)? (GOS8133)→
env-83	JCVI_SCAF_1101668235177	+	25	153	**RNA**→
Cli-1-1	NZ_AAOA01000009.1	+	175160	175281	**RNA**→ hypo→ hypo→ Transposase_8 (pfam01527)→ rve (pfam00665)→ hypo→

Cli-1-2	NZ_AAOA01000016.1	+	47143	47265	**RNA**→ hypo→
env-84	JCVI_SCAF_1101667135423	+	92	225	**RNA**→
Nme-3-1	NC_003116.1	+	650530	650653	**RNA**→ MetK (COG0192)→
Nme-1-1	NC_008767.1	+	441618	441741	**RNA**→ MetK (COG0192)→
Nme-2-1	NC_003112.2	-	1888986	1888863	**RNA**→ ←hypo
Ngo-1-1	NC_002946.2	+	118224	118347	**RNA**→ MetK (COG0192)→
env-85	ABEF01031733.1	+	190	316	**RNA**→ hypo→
env-86	JCVI_SCAF_1101668119308	-	783	645	**RNA**→ MET2 (COG2021)? (GOS3619)→
env-87	ABEF01012528.1	-	889	758	**RNA**→ MET2 (COG2021)→
env-88	JCVI_SCAF_1096628168180	+	1380	1510	**RNA**→ ? (GOS3619)→
env-89	JCVI_SCAF_1096626920933	+	2446	2568	**RNA**→ ←? (GOS11262717)
env-90	JCVI_SCAF_1096627539577	+	1198	1320	**RNA**→ ←? (GOS11262717)
env-91	JCVI_SCAF_1101668513251	+	919	1039	**RNA**→ ←? (GOS11262717)
env-92	JCVI_SCAF_1096626948304	+	245	382	**RNA**→ ? (GOS12164296)→
env-93	ABEF01005861.1	+	299	440	**RNA**→ hypo→ hypo→
env-94	JCVI_SCAF_1096628157798	-	1040	915	**RNA**→
env-95	JCVI_SCAF_1096627527469	+	721	844	**RNA**→
env-96	JCVI_SCAF_1096628274490	+	341	469	**RNA**→ Acetyltransf_1 (pfam00583)? (GOS11675326)→
env-97	JCVI_SCAF_1101668716121	+	692	820	**RNA**→ ? (GOS11238854)→
env-98	JCVI_SCAF_1096627738853	-	864	737	**RNA**→ SRA (smart00466)? (GOS2778)→
env-99	JCVI_SCAF_1101668257019	-	862	734	**RNA**→ ? (GOS11238854)→
env-100	JCVI_SCAF_1096627610786	+	997	1125	**RNA**→ ? (GOS11238854)→
env-101	JCVI_SCAF_1101668211147	-	1244	1116	**RNA**→ ? (GOS11238854)→
env-102	JCVI_SCAF_1096627987638	+	224	342	**RNA**→ ? (GOS11175113)→
env-103	JCVI_SCAF_1101667468172	+	255	383	**RNA**→
env-104	JCVI_SCAF_1096628228675	-	1290	1162	**RNA**→ ? (GOS11094545)→
env-105	JCVI_SCAF_1101668191496	+	1243	1358	**RNA**→
env-106	JCVI_SCAF_1096628059263	-	1134	1007	**RNA**→
env-107	JCVI_SCAF_1096628136383	-	1169	1032	**RNA**→ ←Ribosomal_S15p_S13e (cd00353)? (GOS13206)
env-108	JCVI_SCAF_1096627345166	-	1643	1506	**RNA**→ ? (GOS11175113)→ COG4319 (COG4319)? (GOS11052592)→ DltE (COG0300)? (GOS1303)→
env-109	JCVI_SCAF_1101668203373	-	1219	1083	**RNA**→ ? (GOS11175113)→
env-110	JCVI_SCAF_1101668716771	-	125	1	**RNA**→
env-111	JCVI_SCAF_1096626942020	-	1122	986	**RNA**→ ? (GOS13093079)→
env-112	JCVI_SCAF_1096628151960	+	68	205	**RNA**→
env-113	JCVI_SCAF_1101667124163	+	476	613	**RNA**→ ? (GOS13093079)→
env-114	JCVI_SCAF_1101668174748	+	199	336	**RNA**→ ? (GOS11175113)→
env-115	JCVI_SCAF_1096626558299	+	447	584	**RNA**→ ? (GOS11513631)→
env-116	JCVI_SCAF_1101667522721	+	503	630	**RNA**→
env-117	JCVI_SCAF_1096627539543	+	463	592	**RNA**→ ? (GOS11689481)→ ? (GOS10829)→
env-118	JCVI_SCAF_1101668217711	+	1196	1313	**RNA**→
env-119	JCVI_SCAF_1096626974317	-	1210	1073	**RNA**→ ←? (GOS12077705)
env-120	JCVI_SCAF_1101668291408	+	321	447	**RNA**→ ? (GOS11689481)→
env-121	JCVI_SCAF_1096628162704	-	1743	1606	**RNA**→
env-122	ABEF01000072.1	+	359	486	**RNA**→ hypo→
env-123	JCVI_SCAF_1096627627363	+	281	408	**RNA**→ ? (GOS13542770)→ ? (GOS13171742)→
env-124	JCVI_SCAF_1096626932565	+	512	639	**RNA**→ ←Dcm (COG0270)? (GOS606)
env-125	JCVI_SCAF_1096627453093	+	715	811	**RNA**→
env-126	JCVI_SCAF_1101668021261	+	401	527	**RNA**→ OmpW (COG3047)? (GOS8568)→ OmpW (COG3047)? (GOS8568)→
env-127	JCVI_SCAF_1096627288222	+	1603	1729	**RNA**→ OmpW (COG3047)? (GOS8568)→
env-128	JCVI_SCAF_1101668396607	-	1022	890	**RNA**→
env-129	JCVI_SCAF_1096627887852	-	201	81	**RNA**→
env-130	JCVI_SCAF_1096627892392	-	390	273	**RNA**→

env-131	JCVI_SCAF_1096627634917	-	1356	1211	**RNA**→
env-132	JCVI_SCAF_1096627447827	+	93	211	**RNA**→
env-133	JCVI_SCAF_1101668309352	-	877	756	**RNA**→ COG3378 (COG3378)? (GOS11844258)→
env-134	ABEF01046272.1	+	143	263	**RNA**→ COG0714 (COG0714)→
env-135	JCVI_SCAF_1096627663773	+	43	155	**RNA**→
env-136	JCVI_SCAF_1101668122355	+	1135	1237	**RNA**→
env-137	JCVI_SCAF_1096628170041	-	715	606	**RNA**→ ? (GOS11663299)→
env-138	JCVI_SCAF_1101667504948	+	645	755	**RNA**→
env-139	JCVI_SCAF_1096628269986	+	1347	1458	**RNA**→
env-140	JCVI_SCAF_1096627265999	+	1266	1356	**RNA**→
env-141	JCVI_SCAF_1101667417344	-	839	734	**RNA**→
env-142	JCVI_SCAF_1096627239866	-	227	137	**RNA**→
env-143	JCVI_SCAF_1096627264432	+	7044	7149	**RNA**→ OmpW (pfam03922)? (GOS8568)→ MET2 (COG2021)? (GOS3619)→ MetW (pfam07021)? (GOS721)→
env-144	JCVI_SCAF_1101668544598	-	614	508	**RNA**→ OmpW (COG3047)? (GOS8568)→
env-145	JCVI_SCAF_1101667421803	-	536	431	**RNA**→ OmpW (pfam03922)? (GOS8568)→
env-146	JCVI_SCAF_1101668574702	-	867	747	**RNA**→ ←Phe tRNA
Nca-1-1	NZ_AAOW01000005.1	+	152632	152749	**RNA**→ hypo→
env-147	JCVI_SCAF_1096627559129	-	1384	1266	**RNA**→ SNF2_N (pfam00176)? (GOS1367)→
Bce-3-1	NZ_AAKX01000119.1	-	136189	136072	**RNA**→ ←GlmS (COG0449)
Bce-2-1	NZ_AAVA01000017.1	-	155074	154957	**RNA**→ hypo→
Bce-1-1	NC_008543.1	+	271286	271403	**RNA**→ hypo→
Bsp-1-1	NC_007511.1	-	3090724	3090607	**RNA**→ ←hypo
env-148	JCVI_SCAF_1096627388964	+	422118	422235	**RNA**→ COG0702 (COG0702)? (GOS589)→ ? (GOS11533962)→ DltE (COG0300)? (GOS49)→ HTH_MARR (smart00347)? (GOS1235)→
env-149	JCVI_SCAF_1101668208178	-	1711	1605	**RNA**→ ←? (GOS8069)
Rsp-1-1	NZ_ABCR01000016.1	+	10470	10588	**RNA**→ hypo→
Oan-1-1	NC_009668.1	+	851733	851856	**RNA**→ MetA (COG1897)→
Bab-1-1	NC_006933.1	+	745625	745747	**RNA**→ MetA (COG1897)→
Bme-2-1	NC_007624.1	+	745613	745735	**RNA**→ MetA (COG1897)→
Bov-1-1	NC_009504.1	+	441133	441255	**RNA**→ MetA (COG1897)→
Bsu-1-1	NC_004311.2	+	467621	467743	**RNA**→ MetA (COG1897)→
Bme-1-1	NC_003318.1	+	819611	819733	**RNA**→ MetA (COG1897)→
env-150	JCVI_SCAF_1096627600807	+	1516	1630	**RNA**→ ? (GOS11284089)→
env-151	JCVI_SCAF_1096627499403	-	1190	1078	**RNA**→
env-152	JCVI_SCAF_1101668701353	-	606	494	**RNA**→ ? (GOS11284089)→
env-153	ABEF01001233.1	+	402	519	**RNA**→
env-154	ABEF01017755.1	+	247	364	**RNA**→ COG0714 (COG0714)→
env-155	JCVI_SCAF_1101667126092	+	342	449	**RNA**→
env-156	JCVI_SCAF_1101668177802	+	640	748	**RNA**→ TonB (pfam03544)→
env-157	JCVI_SCAF_1096627742735	+	1126	1254	**RNA**→ ? (GOS13361308)→
env-158	JCVI_SCAF_1096627840920	-	1606	1494	**RNA**→ ? (GOS11593333)→
env-159	JCVI_SCAF_1101668380184	-	528	416	**RNA**→ ? (GOS11593333)→
env-160	JCVI_SCAF_1096627528317	+	1621	1718	**RNA**→
Pna-1-1	NC_008781.1	-	4203375	4203257	**RNA**→ hypo→
env-161	JCVI_SCAF_1101667480781	-	619	510	**RNA**→
env-162	JCVI_SCAF_1096627185926	-	1189	1081	**RNA**→ ? (GOS14032410)→
env-163	JCVI_SCAF_1096628187598	-	1389	1281	**RNA**→ ←? (GOS13274897)
env-164	JCVI_SCAF_1096628087356	+	1423	1531	**RNA**→ ? (GOS4776)→
env-165	JCVI_SCAF_1101668329945	-	207	113	**RNA**→
env-166	JCVI_SCAF_1096627090207	+	1780	1874	**RNA**→
env-167	JCVI_SCAF_1096627233128	-	1365	1257	**RNA**→ MHT1 (COG2040)? (GOS4776)→
env-168	JCVI_SCAF_1096627323785	+	2582	2690	**RNA**→ MHT1 (COG2040)? (GOS4776)→

env-169	JCVI_SCAF_1096628272096	+	46	154	**RNA**→
env-170	JCVI_SCAF_1096627209927	+	1515	1623	**RNA**→ MHT1 (COG2040)? (GOS4776)→
env-171	JCVI_SCAF_1101667398021	+	572	668	**RNA**→
env-172	JCVI_SCAF_1096627803147	+	1667	1764	**RNA**→
env-173	JCVI_SCAF_1096627060512	+	1020	1130	**RNA**→
env-174	JCVI_SCAF_1096627578586	+	350	456	**RNA**→ FabG (COG1028)? (GOS49)→
env-175	JCVI_SCAF_1096627779147	+	499	607	**RNA**→ ? (GOS3744)→
env-176	JCVI_SCAF_1096628113336	-	1005	896	**RNA**→ ? (GOS14663863)→
env-177	JCVI_SCAF_1096628383540	-	167	72	**RNA**→
env-178	JCVI_SCAF_1101668225452	+	1085	1194	**RNA**→ ? (GOS14663863)→
env-179	JCVI_SCAF_1101668464039	-	97	2	**RNA**→
env-180	JCVI_SCAF_1101667517533	+	387	497	**RNA**→ ? (GOS12092552)→
env-181	JCVI_SCAF_1101667496723	+	160	270	**RNA**→ ←Cyt_C5_DNA_methylase (cd00315)? (GOS606)
env-182	JCVI_SCAF_1101667572226	-	482	372	**RNA**→ ? (GOS12092552)→
env-183	JCVI_SCAF_1096627260064	-	744	634	**RNA**→ ? (GOS12092552)→ ? (GOS606)→
env-184	JCVI_SCAF_1101667520909	+	484	594	**RNA**→ ? (GOS12092552)→
env-185	JCVI_SCAF_1101668206055	-	1264	1155	**RNA**→ ? (GOS11631472)→
env-186	JCVI_SCAF_1096627167517	-	953	843	**RNA**→ ? (GOS11858204)→
env-187	JCVI_SCAF_1096627700322	-	106	11	**RNA**→
env-188	JCVI_SCAF_1096627002423	+	861	970	**RNA**→
env-189	JCVI_SCAF_1101668177386	+	1421	1515	**RNA**→
env-190	JCVI_SCAF_1101668428450	-	135	40	**RNA**→
env-191	JCVI_SCAF_1096627381257	+	6570	6680	**RNA**→
env-192	JCVI_SCAF_1096626557110	+	778	875	**RNA**→
env-193	JCVI_SCAF_1096627773291	-	139	44	**RNA**→
env-194	JCVI_SCAF_1096628252679	-	555	446	**RNA**→ ←DUF37 (pfam01809)? (GOS9549)
env-195	JCVI_SCAF_1096627703570	+	368	478	**RNA**→ ←? (GOS899)
env-196	JCVI_SCAF_1101668084382	+	313	423	**RNA**→ ? (GOS13467305)→ ? (GOS11259886)→ ? (GOS11259886)→
env-197	JCVI_SCAF_1101668209998	-	1053	944	**RNA**→ ? (GOS11631472)→
env-198	JCVI_SCAF_1101668437921	-	819	709	**RNA**→ ? (GOS13467305)→ ? (GOS11259886)→ ? (GOS11575626)→
env-199	JCVI_SCAF_1101668103677	-	566	456	**RNA**→ ? (GOS13467305)→ ? (GOS11259886)→
env-200	JCVI_SCAF_1096627330083	-	2164	2055	**RNA**→ ? (GOS13247137)→
env-201	JCVI_SCAF_1096626469133	+	396	505	**RNA**→ ? (GOS11631472)→
env-202	JCVI_SCAF_1101667233265	+	237	346	**RNA**→ ? (GOS13247137)→ ? (GOS13226794)→
env-203	JCVI_SCAF_1101668706240	+	1101	1196	**RNA**→
env-204	JCVI_SCAF_1101668708185	+	994	1103	**RNA**→ ? (GOS13247137)→
env-205	JCVI_SCAF_1096627209634	-	1502	1393	**RNA**→ ? (GOS13247137)→
env-206	JCVI_SCAF_1096627414922	+	724	834	**RNA**→ ? (GOS11575626)→ DUF55 (pfam01878)? (GOS11158481)→
env-207	JCVI_SCAF_1096627110561	+	535	644	**RNA**→ ←? (GOS11343493)
env-208	JCVI_SCAF_1096627036518	+	941	1050	**RNA**→
env-209	JCVI_SCAF_1101668720396	+	570	679	**RNA**→ ←? (GOS1733)
env-210	JCVI_SCAF_1101668678585	-	221	126	**RNA**→
env-211	JCVI_SCAF_1101668463146	-	998	889	**RNA**→ ? (GOS11631472)→ Pseudo tRNA→
env-212	JCVI_SCAF_1096626411095	+	154	263	**RNA**→ ? (GOS11631472)→
env-213	JCVI_SCAF_1096627269300	-	1023	914	**RNA**→ ? (GOS11631472)→
env-214	JCVI_SCAF_1096627208345	+	456	565	**RNA**→ ? (GOS11631472)→ Ser tRNA→ ? (GOS14253017)→
env-215	JCVI_SCAF_1096626942975	-	760	652	**RNA**→
env-216	JCVI_SCAF_1096627196160	-	3049	2941	**RNA**→ ←? (GOS11802)
env-217	JCVI_SCAF_1101668313011	+	626	734	**RNA**→ ←? (GOS11343493)
env-218	JCVI_SCAF_1096627614817	-	1578	1470	**RNA**→ ←? (GOS11343493)
env-219	JCVI_SCAF_1096627771297	+	476	585	**RNA**→ ? (GOS13247137)→
env-220	JCVI_SCAF_1096627779582	+	868	977	**RNA**→ ? (GOS11631472)→

env-221	JCVI_SCAF_1101668300707	-	1222	1113	**RNA**→ ? (GOS13247137)→ RpoE (COG1595)? (GOS5504)→ ? (GOS11390543)→
env-222	JCVI_SCAF_1096627726096	+	1426	1521	**RNA**→
env-223	JCVI_SCAF_1101668215902	-	1374	1265	**RNA**→ ? (GOS13361308)→
env-224	JCVI_SCAF_1101668741346	+	1174	1282	**RNA**→
env-225	JCVI_SCAF_1096627927794	+	454	564	**RNA**→ ? (GOS14660636)→
env-226	JCVI_SCAF_1101667048185	-	583	473	**RNA**→ ? (GOS14660636)→
env-227	JCVI_SCAF_1096627916643	-	558	450	**RNA**→
env-228	JCVI_SCAF_1096627635954	+	1091	1211	**RNA**→
env-229	ABEF01023758.1	+	461	570	**RNA**→ hypo→
Lag-1-1	NZ_AAUW01000010.1	-	69484	69347	**RNA**→ hypo→
env-230	JCVI_SCAF_1101668648929	-	259	157	**RNA**→
env-231	JCVI_SCAF_1096627360979	-	3243	3125	**RNA**→ ? (GOS10845)→
env-232	ABEF01026777.1	+	1438	1540	**RNA**→
env-233	ABEF01032110.1	-	492	379	**RNA**→
env-234	ABEF01016903.1	+	179	298	**RNA**→ COG0714 (COG0714)→
env-235	JCVI_SCAF_1101667040313	+	263	381	**RNA**→ ? (GOS12105745)→
env-236	JCVI_SCAF_1101668061828	+	1093	1197	**RNA**→
Acr-1-1	NC_009470.1	-	3559	3436	**RNA**→ Acetyltransf_1 (pfam00583)→
Swi-1-1	NC_009511.1	-	636052	635931	**RNA**→ MetA (COG1897)→
env-237	JCVI_SCAF_1096627469829	-	999	880	**RNA**→ ←? (GOS11044062)
env-238	JCVI_SCAF_1096627680375	+	521	637	**RNA**→ ? (GOS13297217)→ ? (GOS11765158)→
env-239	JCVI_SCAF_1096627120551	-	1762	1647	**RNA**→ ? (GOS13297217)→ ? (GOS11923052)→ ? (GOS11765158)→
env-240	JCVI_SCAF_1101668394919	-	603	487	**RNA**→ ? (GOS13297217)→
env-241	JCVI_SCAF_1096628351919	-	323	221	**RNA**→
env-242	JCVI_SCAF_1096628001838	-	418	302	**RNA**→
env-243	JCVI_SCAF_1096628364210	+	1204	1320	**RNA**→ ? (GOS13297217)→
env-244	JCVI_SCAF_1096627826832	-	271	176	**RNA**→
Pna-1-2	NC_008760.1	-	31707	31584	**RNA**→ Transposase_25 (pfam03050)→
env-245	JCVI_SCAF_1096627496650	+	356	463	**RNA**→ ? (GOS11475618)→
env-246	JCVI_SCAF_1096626908750	+	1218	1327	**RNA**→ ? (GOS12031351)→
env-247	JCVI_SCAF_1096627379551	-	2636	2525	**RNA**→
env-248	JCVI_SCAF_1096626988086	+	1428	1520	**RNA**→
env-249	JCVI_SCAF_1101668506008	+	321	427	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-250	JCVI_SCAF_1101667077280	-	401	295	**RNA**→ ? (GOS3285)→
env-251	JCVI_SCAF_1101668426160	-	135	43	**RNA**→
env-252	JCVI_SCAF_1096627171199	+	716	822	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-253	JCVI_SCAF_1101668715390	-	555	449	**RNA**→ ? (GOS3285)→
Pto-1-1	NZ_AAPR01000181.1	-	352	260	**RNA**→
env-254	JCVI_SCAF_1096627323041	-	3034	2928	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-255	JCVI_SCAF_1096627263379	-	2634	2528	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-256	JCVI_SCAF_1101668521348	+	222	328	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-257	JCVI_SCAF_1096627318695	-	636	530	**RNA**→ Cys_Met_Meta_PP (pfam01053)? (GOS3285)→
env-258	JCVI_SCAF_1096626957117	+	1282	1388	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-259	JCVI_SCAF_1096626874379	+	1636	1742	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-260	JCVI_SCAF_1096626875614	-	475	369	**RNA**→ ? (GOS3285)→
env-261	JCVI_SCAF_1096627260474	+	3543	3649	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-262	JCVI_SCAF_1101668588966	-	453	347	**RNA**→ ? (GOS3285)→
env-263	JCVI_SCAF_1101668093721	-	297	205	**RNA**→
env-264	JCVI_SCAF_1096626892538	-	3548	3442	**RNA**→ MET17 (COG2873)? (GOS3285)→ MET2 (COG2021)? (GOS3619)→
env-265	JCVI_SCAF_1096627533744	-	1567	1461	**RNA**→ MET17 (COG2873)? (GOS3285)→ MET17 (COG2873)? (GOS3285)→ ? (GOS3619)→
env-266	JCVI_SCAF_1096627243045	+	4986	5092	**RNA**→ ? (GOS3285)→
env-267	JCVI_SCAF_1096626987395	-	933	827	**RNA**→ MET17 (COG2873)? (GOS3285)→

env-268	JCVI_SCAF_1096626463060	+	479	585	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-269	JCVI_SCAF_1096628293622	-	1219	1113	**RNA**→ ? (GOS3285)→ MET2 (COG2021)? (GOS3619)→
env-270	JCVI_SCAF_1101667516259	-	316	224	**RNA**→
env-271	JCVI_SCAF_1096626930172	+	1639	1731	**RNA**→ ? (GOS3285)→
env-272	JCVI_SCAF_1101667286361	+	492	584	**RNA**→
env-273	JCVI_SCAF_1096627117530	-	449	343	**RNA**→ ? (GOS3285)→
env-274	JCVI_SCAF_1096627129211	-	479	373	**RNA**→ ? (GOS3285)→
env-275	JCVI_SCAF_1096627344587	-	1635	1529	**RNA**→ MET17 (COG2873)? (GOS3285)→ MET17 (COG2873)? (GOS3285)→ ? (GOS3619)→
env-276	JCVI_SCAF_1096627346237	+	661	767	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-277	JCVI_SCAF_1101668504176	+	714	820	**RNA**→ ? (GOS3285)→
env-278	JCVI_SCAF_1096627344546	+	1516	1622	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-279	JCVI_SCAF_1096627121962	+	1346	1447	**RNA**→ ? (GOS3285)→
env-280	JCVI_SCAF_1101668093170	+	1291	1397	**RNA**→ ? (GOS3285)→
env-281	JCVI_SCAF_1096627320893	-	566	460	**RNA**→ ? (GOS3285)→
env-282	JCVI_SCAF_1096627704766	-	443	337	**RNA**→ ? (GOS3285)→
env-283	JCVI_SCAF_1096626924876	-	3577	3471	**RNA**→
env-284	JCVI_SCAF_1101668165124	-	1559	1453	**RNA**→ MET17 (COG2873)? (GOS3285)→ ? (GOS3619)→
env-285	JCVI_SCAF_1096627870892	-	770	664	**RNA**→ ? (GOS3285)→
env-286	JCVI_SCAF_1101668256392	+	987	1093	**RNA**→ ? (GOS3285)→
env-287	JCVI_SCAF_1096628238831	-	437	331	**RNA**→ ? (GOS3285)→
env-288	JCVI_SCAF_1101668291307	+	181	287	**RNA**→ MET17 (COG2873)? (GOS3285)→ MET17 (COG2873)? (GOS3285)→ ? (GOS3619)→
env-289	JCVI_SCAF_1096627091613	+	2025	2117	**RNA**→
env-290	JCVI_SCAF_1096627190972	-	1121	1015	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-291	JCVI_SCAF_1096628081542	+	282	388	**RNA**→ ? (GOS3285)→ ? (GOS3285)→ ? (GOS3619)→
env-292	JCVI_SCAF_1096627622563	+	61	167	**RNA**→
env-293	JCVI_SCAF_1096627319097	-	511	405	**RNA**→ ? (GOS3285)→
env-294	JCVI_SCAF_1096627093771	+	634	740	**RNA**→ ? (GOS3285)→ MET17 (COG2873)? (GOS3285)→
env-295	JCVI_SCAF_1096627343486	+	1918	2024	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-296	JCVI_SCAF_1096627130271	-	3798	3692	**RNA**→
env-297	JCVI_SCAF_1096627327375	+	2371	2477	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-298	JCVI_SCAF_1101668668770	-	882	776	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-299	JCVI_SCAF_1096626942706	-	973	867	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-300	JCVI_SCAF_1096627703617	+	56	162	**RNA**→
env-301	JCVI_SCAF_1096626476896	-	1047	941	**RNA**→
env-302	JCVI_SCAF_1101668211932	-	902	796	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-303	JCVI_SCAF_1096626957446	+	340	446	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-304	JCVI_SCAF_1096626125585	-	239	147	**RNA**→
env-305	JCVI_SCAF_1101668214721	+	1041	1147	**RNA**→ ? (GOS3285)→
env-306	JCVI_SCAF_1096626137780	-	508	402	**RNA**→ ? (GOS3285)→
env-307	JCVI_SCAF_1101668191391	-	1454	1348	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-308	JCVI_SCAF_1096626959948	-	502	396	**RNA**→ ? (GOS3285)→
env-309	JCVI_SCAF_1096627245548	+	1377	1483	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-310	JCVI_SCAF_1096627311787	-	90	1	**RNA**→
env-311	JCVI_SCAF_1101667276848	-	434	328	**RNA**→ ? (GOS3285)→
env-312	JCVI_SCAF_1096627036390	+	11	117	**RNA**→
env-313	JCVI_SCAF_1096627382404	-	5481	5375	**RNA**→ MET17 (COG2873)? (GOS3285)→ MET2 (COG2021)? (GOS3619)→
env-314	JCVI_SCAF_1096627381557	+	101	207	**RNA**→
env-315	JCVI_SCAF_1096627591498	-	194	102	**RNA**→
env-316	JCVI_SCAF_1096627193851	-	559	453	**RNA**→ ? (GOS3285)→
env-317	JCVI_SCAF_1096626135635	-	385	283	**RNA**→
env-318	JCVI_SCAF_1096628022488	-	1457	1351	**RNA**→ ? (GOS3285)→ MET17 (COG2873)? (GOS3285)→
env-319	JCVI_SCAF_1096627053142	+	117	223	**RNA**→ MET17 (COG2873)? (GOS3285)→ ? (GOS3619)→

env-320	JCVI_SCAF_1101667216735	+	289	381	**RNA**→
env-321	JCVI_SCAF_1101667128491	-	632	526	**RNA**→
env-322	JCVI_SCAF_1096627306307	-	1643	1537	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-323	JCVI_SCAF_1096626562225	+	279	385	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-324	JCVI_SCAF_1096628016253	+	1155	1261	**RNA**→ ? (GOS3285)→
env-325	JCVI_SCAF_1096627181677	-	1693	1587	**RNA**→
env-326	JCVI_SCAF_1096627180162	-	1065	959	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-327	JCVI_SCAF_1101667138126	-	680	574	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-328	JCVI_SCAF_1101667096925	+	326	432	**RNA**→ ? (GOS3285)→
env-329	JCVI_SCAF_1096627323401	+	1481	1587	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-330	JCVI_SCAF_1096626664166	-	297	205	**RNA**→
env-331	JCVI_SCAF_1101668318849	+	688	794	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-332	JCVI_SCAF_1096626891366	+	1120	1226	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-333	JCVI_SCAF_1101668281990	+	1515	1607	**RNA**→
env-334	JCVI_SCAF_1101667475927	-	606	500	**RNA**→ Cys_Met_Meta_PP (pfam01053)? (GOS3285)→
env-335	JCVI_SCAF_1101668265031	+	173	279	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-336	JCVI_SCAF_1096627689605	-	887	781	**RNA**→ ? (GOS3285)→
env-337	JCVI_SCAF_1096627334308	+	3030	3136	**RNA**→ ? (GOS3285)→
env-338	JCVI_SCAF_1101667256753	-	732	626	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-339	JCVI_SCAF_1101667332039	+	248	354	**RNA**→ Cys_Met_Meta_PP (pfam01053)? (GOS3285)→
env-340	JCVI_SCAF_1101667123461	+	377	483	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-341	JCVI_SCAF_1096627323692	-	1138	1032	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-342	JCVI_SCAF_1096627066632	-	666	560	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-343	JCVI_SCAF_1096627316088	+	863	969	**RNA**→ MET17 (COG2873)? (GOS3285)→ ? (GOS3619)→
env-344	JCVI_SCAF_1096627320428	+	1757	1863	**RNA**→ ? (GOS3285)→
env-345	JCVI_SCAF_1096628261629	-	590	484	**RNA**→ ? (GOS3285)→
env-346	JCVI_SCAF_1096627029411	+	838	930	**RNA**→
env-347	JCVI_SCAF_1101668262705	+	428	534	**RNA**→ ? (GOS3285)→
env-348	JCVI_SCAF_1101668275645	-	539	433	**RNA**→ ? (GOS3285)→
env-349	JCVI_SCAF_1096627033813	+	264	370	**RNA**→ MET17 (COG2873)? (GOS3285)→ MET2 (COG2021)? (GOS3619)→
env-350	JCVI_SCAF_1096627268641	+	7434	7540	**RNA**→ ? (GOS3285)→
env-351	JCVI_SCAF_1096626941368	-	583	477	**RNA**→ ? (GOS3285)→
env-352	JCVI_SCAF_1101668538318	+	1126	1231	**RNA**→ MTHFR (cd00537)? (GOS6193)→
env-353	JCVI_SCAF_1096626993230	-	305	217	**RNA**→
env-354	JCVI_SCAF_1096627260474	+	2307	2411	**RNA**→ MetK (COG0192)? (GOS4376)→
env-355	JCVI_SCAF_1101668537071	-	424	320	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-356	JCVI_SCAF_1096628061480	+	448	552	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→ Hpt (COG0634)? (GOS1256)→
env-357	JCVI_SCAF_1101668477138	-	612	508	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-358	JCVI_SCAF_1096627243045	+	3750	3854	**RNA**→ MetK (COG0192)? (GOS4376)→
env-359	JCVI_SCAF_1096627344546	+	278	381	**RNA**→ MetK (COG0192)? (GOS4376)→
env-360	JCVI_SCAF_1096627117530	-	1687	1584	**RNA**→ MetK (COG0192)? (GOS4376)→
env-361	JCVI_SCAF_1096627129211	-	1718	1614	**RNA**→ MetK (COG0192)? (GOS4376)→
env-362	JCVI_SCAF_1096626930172	+	403	507	**RNA**→ MetK (COG0192)? (GOS4376)→
env-363	JCVI_SCAF_1096627240702	-	538	434	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-364	JCVI_SCAF_1101668484636	-	918	814	**RNA**→ MetK (COG0192)? (GOS4376)→
env-365	JCVI_SCAF_1096626978410	-	828	725	**RNA**→ MetK (COG0192)? (GOS4376)→
env-366	JCVI_SCAF_1096627376428	+	648	747	**RNA**→ MetK (COG0192)? (GOS4376)→
env-367	JCVI_SCAF_1096627091613	+	791	893	**RNA**→ MetK (COG0192)? (GOS4376)→
env-368	JCVI_SCAF_1096627320428	+	519	621	**RNA**→ MetK (COG0192)? (GOS4376)→
env-369	JCVI_SCAF_1096627422617	+	718	820	**RNA**→ MetK (COG0192)? (GOS4376)→
env-370	JCVI_SCAF_1096627305633	-	717	615	**RNA**→ MetK (COG0192)? (GOS4376)→
env-371	JCVI_SCAF_1101667362380	+	106	208	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→

env-372	JCVI_SCAF_1096627302018	-	833	731	**RNA**→ MetK (COG0192)? (GOS4376)→
env-373	JCVI_SCAF_1096627682899	-	326	240	**RNA**→
env-374	JCVI_SCAF_1096626988086	+	192	294	**RNA**→ MetK (COG0192)? (GOS4376)→
env-375	JCVI_SCAF_1101668168132	-	549	447	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-376	JCVI_SCAF_1096627319575	+	1967	2069	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-377	JCVI_SCAF_1101667280867	-	143	57	**RNA**→
env-378	JCVI_SCAF_1096627066632	-	1900	1798	**RNA**→ MetK (COG0192)? (GOS4376)→
env-379	JCVI_SCAF_1096627025785	+	1250	1352	**RNA**→ MetK (COG0192)? (GOS4376)→
env-380	JCVI_SCAF_1096628261629	-	1099	997	**RNA**→ S-AdoMet_synt_C (pfam02773)? (GOS4376)→
env-381	JCVI_SCAF_1096626989768	-	839	737	**RNA**→ MetK (COG0192)? (GOS4376)→
env-382	JCVI_SCAF_1101667240237	-	280	194	**RNA**→
env-383	JCVI_SCAF_1096627049315	+	1640	1742	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-384	JCVI_SCAF_1096628016253	+	55	157	**RNA**→ MetK (COG0192)? (GOS4376)→ S-AdoMet_synt_C (pfam02773)? (GOS4376)→
env-385	JCVI_SCAF_1096628095332	-	616	514	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-386	JCVI_SCAF_1096627323401	+	245	347	**RNA**→ MetK (COG0192)? (GOS4376)→
env-387	JCVI_SCAF_1101667264418	-	564	462	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-388	JCVI_SCAF_1096626716519	+	477	579	**RNA**→ MetK (COG0192)? (GOS4376)→
env-389	JCVI_SCAF_1096627374561	+	1354	1456	**RNA**→ MetK (COG0192)? (GOS4376)→
env-390	JCVI_SCAF_1096627170658	+	853	957	**RNA**→ MetK (COG0192)? (GOS4376)→
env-391	JCVI_SCAF_1096627382071	-	791	687	**RNA**→ MetK (COG0192)? (GOS4376)→
env-392	JCVI_SCAF_1096627083383	-	177	88	**RNA**→
env-393	JCVI_SCAF_1096627318695	-	1868	1764	**RNA**→ MetK (COG0192)? (GOS4376)→
env-394	JCVI_SCAF_1096627121962	+	114	217	**RNA**→ MetK (COG0192)? (GOS4376)→
env-395	JCVI_SCAF_1101668093170	+	50	153	**RNA**→ MetK (COG0192)? (GOS4376)→
env-396	JCVI_SCAF_1096627343486	+	670	772	**RNA**→ MetK (COG0192)? (GOS4376)→ S-AdoMet_synt_C (pfam02773)? (GOS4376)→
env-397	JCVI_SCAF_1101668508058	-	1037	935	**RNA**→ MetK (COG0192)? (GOS4376)→
env-398	JCVI_SCAF_1096627327375	+	1130	1232	**RNA**→ MetK (COG0192)? (GOS4376)→
env-399	JCVI_SCAF_1096627099999	+	1669	1771	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-400	JCVI_SCAF_1096626957117	+	41	144	**RNA**→ MetK (COG0192)? (GOS4376)→
env-401	JCVI_SCAF_1096627272433	-	958	855	**RNA**→ MetK (COG0192)? (GOS4376)→
env-402	JCVI_SCAF_1096627320893	-	1806	1703	**RNA**→ MetK (COG0192)? (GOS4376)→
env-403	JCVI_SCAF_1096628238831	-	1269	1166	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-404	JCVI_SCAF_1101668344519	-	911	808	**RNA**→ MetK (COG0192)? (GOS4376)→
env-405	JCVI_SCAF_1101668545521	+	418	520	**RNA**→ MetK (COG0192)? (GOS4376)→
env-406	JCVI_SCAF_1101668281990	+	280	382	**RNA**→ MetK (COG0192)? (GOS4376)→
env-407	JCVI_SCAF_1101668278776	+	1443	1530	**RNA**→
env-408	JCVI_SCAF_1096627961917	+	1209	1296	**RNA**→
env-409	JCVI_SCAF_1101668261317	-	898	796	**RNA**→ MetK (COG0192)? (GOS4376)→
env-410	JCVI_SCAF_1096627334308	+	1795	1897	**RNA**→ MetK (COG0192)? (GOS4376)→
env-411	JCVI_SCAF_1101668263083	+	308	410	**RNA**→ MetK (COG0192)? (GOS4376)→ MetK (COG0192)? (GOS4376)→
env-412	JCVI_SCAF_1096627294171	-	343	252	**RNA**→
env-413	JCVI_SCAF_1096626874379	+	395	500	**RNA**→ MetK (COG0192)? (GOS4376)→
env-414	JCVI_SCAF_1096626875614	-	1716	1611	**RNA**→ MetK (COG0192)? (GOS4376)→
env-415	JCVI_SCAF_1096626892538	-	4787	4683	**RNA**→ MetK (COG0192)? (GOS4376)→
env-416	JCVI_SCAF_1096627343101	+	2517	2620	**RNA**→ MetK (COG0192)? (GOS4376)→
env-417	JCVI_SCAF_1101668093721	-	1537	1434	**RNA**→ MetK (COG0192)? (GOS4376)→
Nar-1-1	NC_007794.1	+	722861	723008	**RNA**→ MetA (COG1897)→
env-418	JCVI_SCAF_1101668087361	+	509	617	**RNA**→ OmpW (COG3047)? (GOS8568)→
env-419	JCVI_SCAF_1096627384377	+	9293	9403	**RNA**→ OmpW (COG3047)? (GOS8568)→ ? (GOS3619)→
env-420	JCVI_SCAF_1101668301974	-	1058	948	**RNA**→ OmpW (COG3047)? (GOS8568)→
env-421	JCVI_SCAF_1101667471706	-	203	105	**RNA**→
env-422	JCVI_SCAF_1101668680057	-	1205	1095	**RNA**→ ? (GOS12118661)→

env-423	JCVI_SCAF_1101667148821	-	512	401	**RNA**→
Pna-1-3	NC_008760.1	+	62361	62482	**RNA**→ hypo→
env-424	JCVI_SCAF_1101668080534	-	156	55	**RNA**→
env-425	JCVI_SCAF_1096627299933	-	1557	1430	**RNA**→ MET2 (COG2021)? (GOS3619)→
env-426	JCVI_SCAF_1096627300646	+	364	482	**RNA**→ MET2 (COG2021)? (GOS3619)→
env-427	JCVI_SCAF_1096627651683	+	756	881	**RNA**→ MET2 (COG2021)? (GOS3619)→
env-428	JCVI_SCAF_1096627676845	-	807	690	**RNA**→ MET2 (COG2021)? (GOS3619)→
env-429	JCVI_SCAF_1101667178866	-	120	15	**RNA**→
env-430	JCVI_SCAF_1096627009086	+	1628	1733	**RNA**→
env-431	JCVI_SCAF_1096626192109	-	481	364	**RNA**→ ? (GOS3619)→
env-432	JCVI_SCAF_1101667164259	-	132	27	**RNA**→
env-433	JCVI_SCAF_1101668234415	-	146	10	**RNA**→
env-434	JCVI_SCAF_1096627008193	-	510	385	**RNA**→
env-435	JCVI_SCAF_1101668248391	-	452	329	**RNA**→
env-436	JCVI_SCAF_1096627647150	-	1279	1155	**RNA**→ ? (GOS12166317)→
env-437	JCVI_SCAF_1096627024754	+	425	548	**RNA**→
env-438	JCVI_SCAF_1096626189621	+	41	164	**RNA**→

85.4 Conserved domains

Conserved domains found in protein-coding genes listed in Section 85.3 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one SAM-I-IV-variant RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

cd00198 (1) Von Willebrand factor type A (vWA) domain was originally found in the blood coagulation protein von Willebrand factor (vWF).
cd00315 (1) Cytosine-C5 specific DNA methylases; Methyl transfer reactions play an important role in many aspects of biology.
cd00353 (1) Ribosomal protein S15 (prokaryotic)_S13 (eukaryotic) binds the central domain of 16S rRNA and is required for assembly of the small ribosomal subunit and for intersubunit association, thus representing a key element in the assembly of the whole ribosome.
cd00537 (1) Methylenetetrahydrofolate reductase (MTHFR).
cd00751 (1) Thiolase are ubiquitous enzymes that catalyze the reversible thiolytic cleavage of 3-ketoacyl-CoA into acyl-CoA and acetyl-CoA, a 2-step reaction involving a covalent intermediate formed with a catalytic cysteine.
cd01104 (1) Helix-turn-helix transcription regulator MlrA (merR-like regulator A).
cd01823 (1) SEST_like.
cd02558 (1) PSRA_1: Pseudouridine synthase, a subgroup of the RluA family.
COG0192 (56) S-adenosylmethionine synthetase [Coenzyme metabolism]
COG0270 (1) Site-specific DNA methylase [DNA replication, recombination, and repair]
COG0300 (2) Short-chain dehydrogenases of various substrate specificities [General function prediction only]
COG0449 (1) Glucosamine 6-phosphate synthetase, contains amidotransferase and phosphosugar isomerase domains [Cell envelope biogenesis, outer membrane]
COG0634 (1) Hypoxanthine-guanine phosphoribosyltransferase [Nucleotide transport and metabolism]
COG0702 (1) Predicted nucleoside-diphosphate-sugar epimerases [Cell envelope biogenesis, outer membrane / Carbohydrate transport and metabolism]
COG0714 (4) MoxR-like ATPases [General function prediction only]
COG0780 (2) Enzyme related to GTP cyclohydrolase I [General function prediction only]
COG1028 (1) Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases) [Secondary metabolites biosynthesis, transport, and catabolism / General function prediction only]
COG1595 (1) DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog [Transcription]
COG1757 (1) Na+/H+ antiporter [Energy production and conversion]
COG1897 (8) Homoserine trans-succinylase [Amino acid transport and metabolism]
COG2021 (14) Homoserine acetyltransferase [Amino acid transport and metabolism]
COG2040 (3) Homocysteine/selenocysteine methylase (S-methylmethionine-dependent) [Amino acid transport and metabolism]
COG2340 (1) Uncharacterized protein with SCP/PR1 domains [Function unknown]
COG2873 (50) O-acetylhomoserine sulfhydrylase [Amino acid transport and metabolism]
COG3047 (7) Outer membrane protein W [Cell envelope biogenesis, outer membrane]
COG3378 (1) Predicted ATPase [General function prediction only]
COG4319 (1) Ketosteroid isomerase homolog [Function unknown]
GOS10829 (1) ?
GOS10845 (1) ?
GOS11041706 (1) ?
GOS11044062 (1) ?
GOS11052592 (1) ?
GOS11074455 (6) ?
GOS11094545 (1) ?

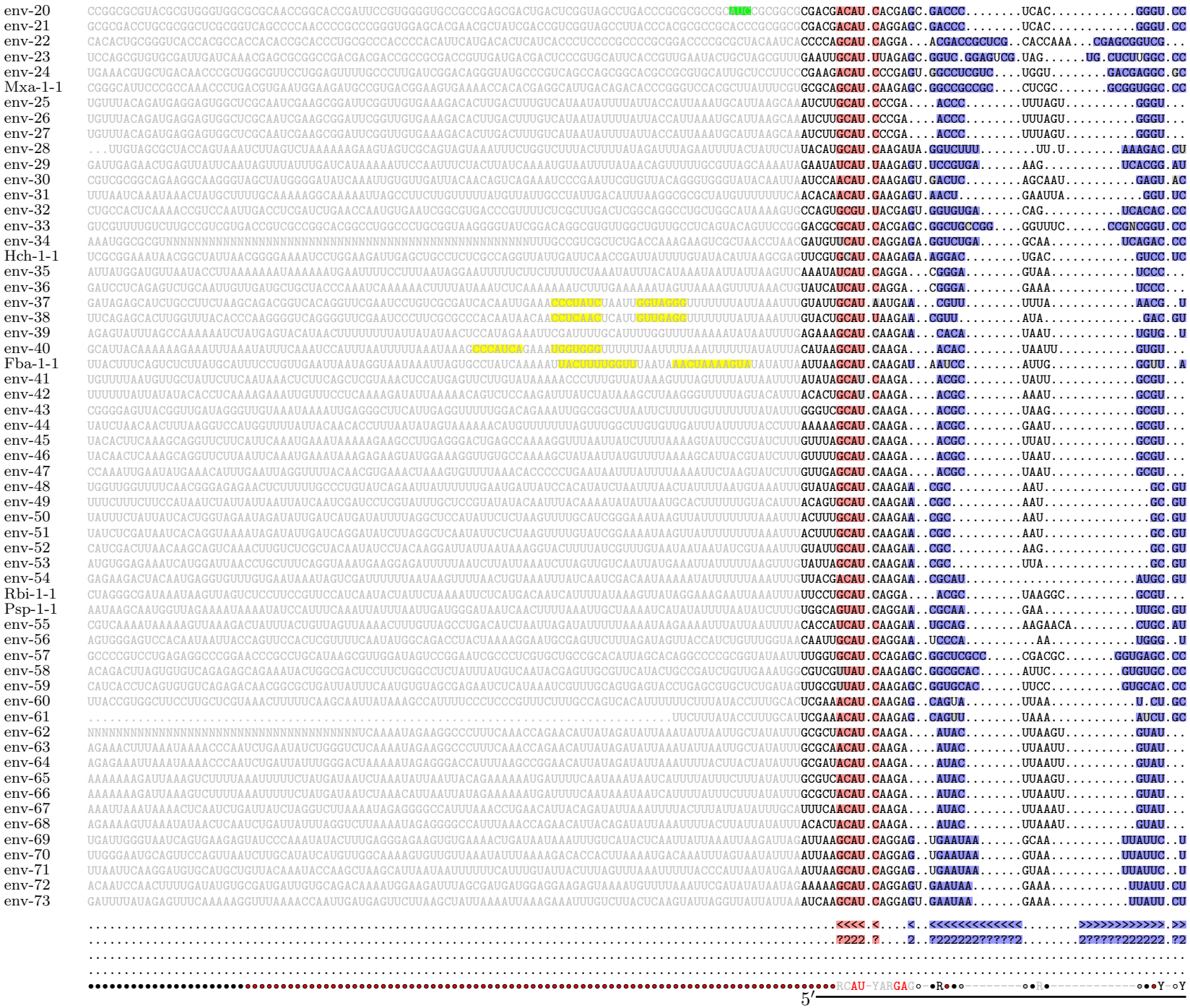
GOS11151719 (2) ?
GOS11158481 (1) ?
GOS11175113 (4) ?
GOS11176883 (1) ?
GOS11238854 (4) ?
GOS11259886 (4) ?
GOS11262717 (3) ?
GOS11284089 (2) ?
GOS11343493 (3) ?
GOS11390543 (1) ?
GOS11475618 (1) ?
GOS11513631 (1) ?
GOS11533962 (1) ?
GOS11575626 (2) ?
GOS11593333 (2) ?
GOS11631472 (8) ?
GOS11663299 (1) ?
GOS11675326 (1) ?
GOS11689481 (2) ?
GOS11697307 (1) ?
GOS11701787 (2) ?
GOS11765158 (2) ?
GOS11785628 (1) ?
GOS11802 (1) ?
GOS11844258 (1) ?
GOS11858204 (1) ?
GOS11879195 (1) ?
GOS11923052 (1) ?
GOS11954949 (2) ?
GOS12011754 (2) ?
GOS12031351 (1) ?
GOS12077705 (1) ?
GOS12092552 (4) ?
GOS12105745 (1) ?
GOS12118661 (1) ?
GOS12164296 (1) ?
GOS12166317 (1) ?
GOS1235 (1) ?
GOS1256 (1) ?
GOS1285 (1) ?
GOS1303 (1) ?
GOS13055046 (2) ?
GOS13085086 (3) ?
GOS13093079 (2) ?
GOS13171742 (1) ?
GOS13205716 (1) ?
GOS13206 (1) ?
GOS13226794 (1) ?
GOS13245863 (1) ?
GOS13247137 (6) ?
GOS13274897 (1) ?
GOS13297217 (5) ?
GOS13327453 (1) ?
GOS13361308 (2) ?
GOS13467305 (3) ?
GOS13476120 (1) ?
GOS13542770 (1) ?
GOS1367 (1) ?
GOS14032410 (1) ?
GOS14253017 (1) ?
GOS14278197 (1) ?
GOS14414208 (1) ?
GOS14602209 (1) ?
GOS14660636 (2) ?
GOS14663863 (2) ?
GOS1733 (1) ?
GOS18368 (15) ?
GOS2778 (1) ?
GOS3111 (1) ?
GOS3285 (89) ?
GOS3619 (23) ?
GOS3744 (1) ?
GOS4376 (68) ?
GOS4776 (4) ?
GOS49 (2) ?
GOS5504 (1) ?
GOS589 (1) ?
GOS606 (4) ?
GOS6193 (1) ?
GOS697 (1) ?
GOS7114 (2) ?
GOS721 (1) ?
GOS7466 (1) ?
GOS8069 (1) ?
GOS8133 (1) ?
GOS8568 (9) ?
GOS8636 (4) ?
GOS899 (1) ?
GOS9454 (1) ?
GOS9549 (1) ?
pfam00176 (1) SNF2 family N-terminal domain.
pfam00438 (13) S-adenosylmethionine synthetase, N-terminal domain.
pfam00583 (2) Acetyltransferase (GNAT) family.
pfam00665 (1) Integrase core domain.
pfam01053 (3) Cys/Met metabolism PLP-dependent enzyme.
pfam01527 (1) Transposase.
pfam01809 (1) Domain of unknown function DUF37.
pfam01878 (1) Protein of unknown function DUF55.
pfam02773 (3) S-adenosylmethionine synthetase, C-terminal domain.
pfam03050 (1) Transposase IS66 family.
pfam03441 (1) FAD binding domain of DNA photolyase.
pfam03544 (1) Gram-negative bacterial tonB protein.
pfam03922 (2) OmpW family.
pfam05013 (2) N-formylglutamate amidohydrolase.

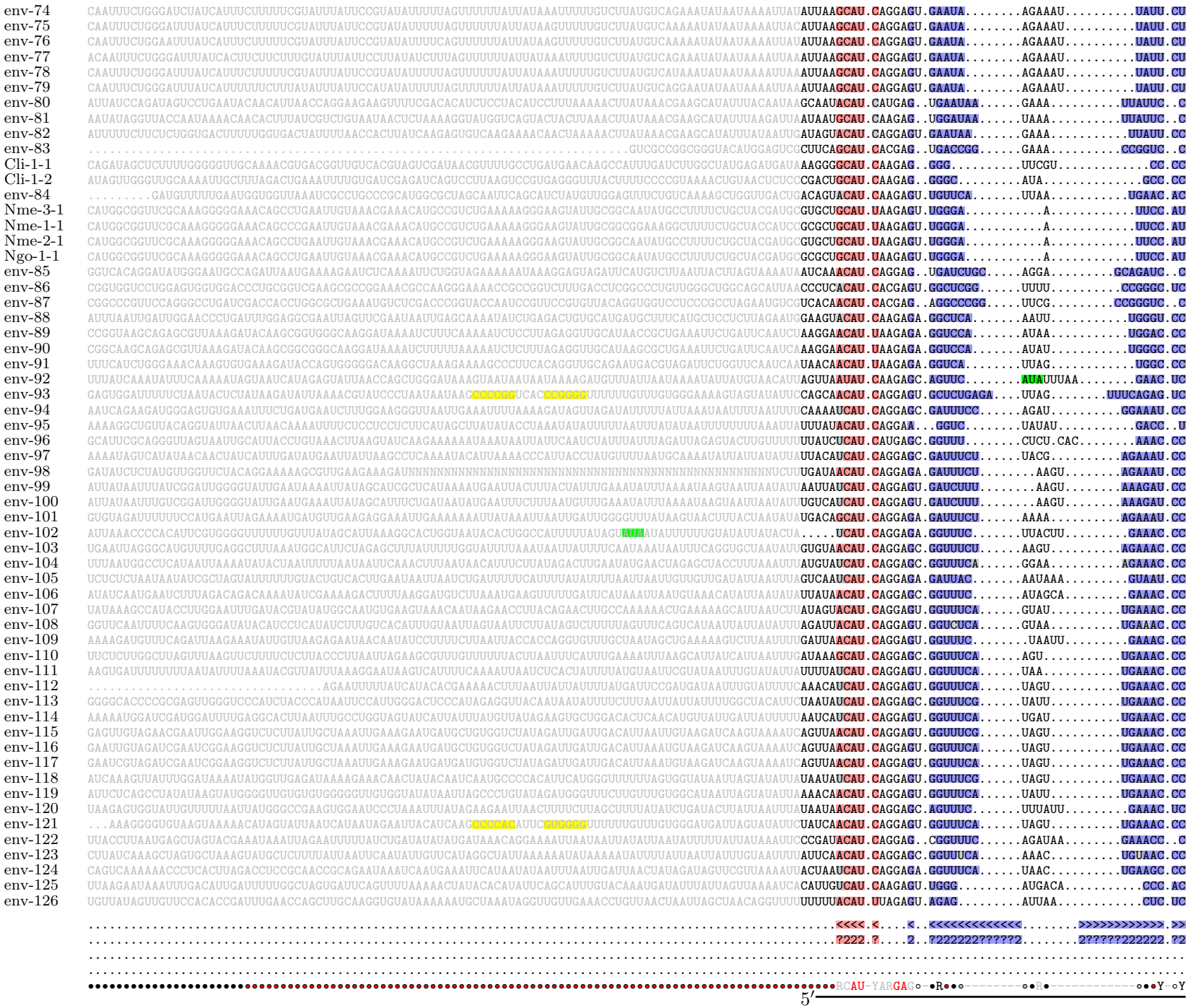
pfam07021 (1) Methionine biosynthesis protein MetW.
smart00347 (1) helix_turn_helix multiple antibiotic resistance protein;
smart00466 (1) SET and RING finger associated domain.
smart00671 (1) Sel1-like repeats.

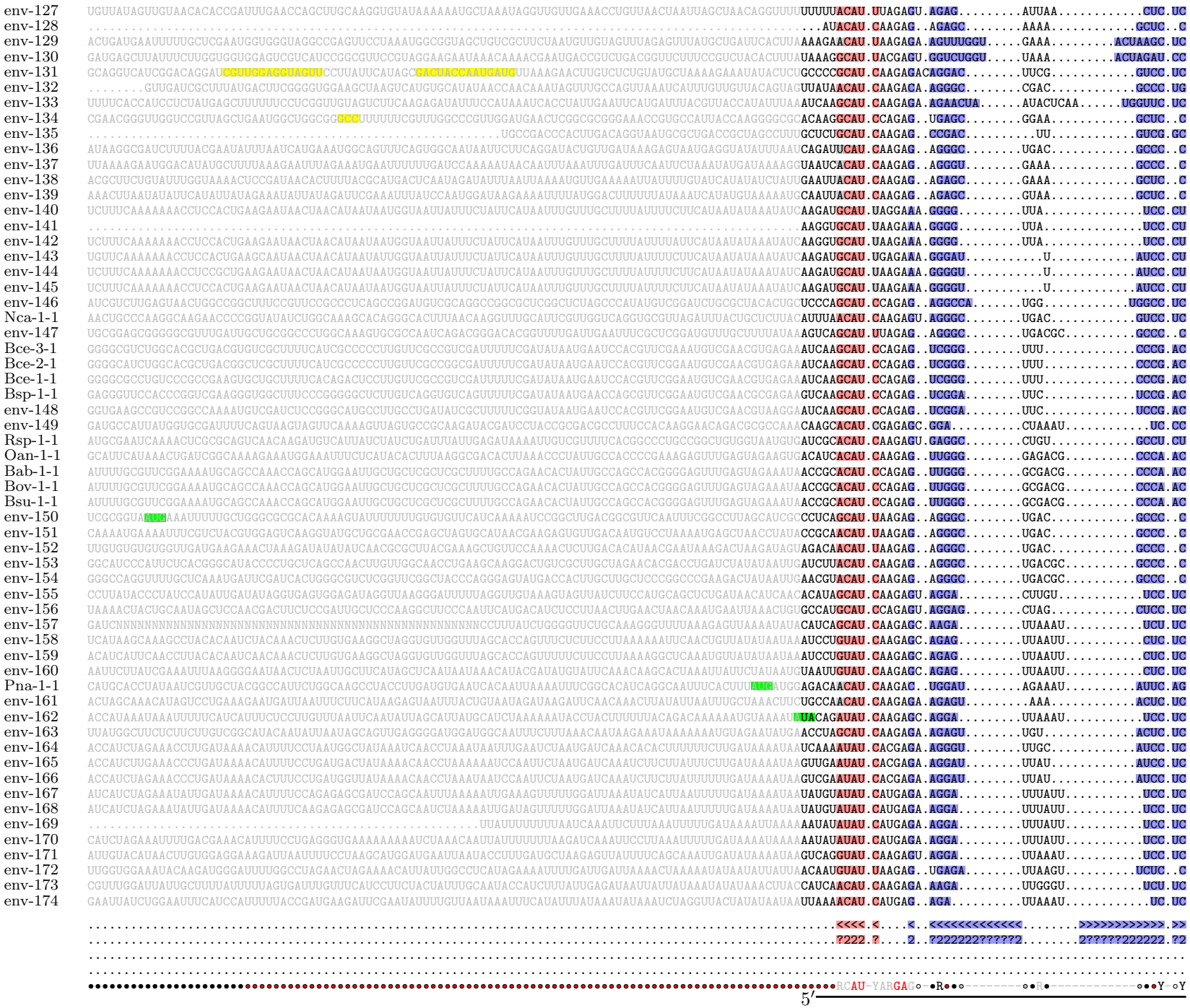
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Bme-1-1, Bme-2-1, env-344

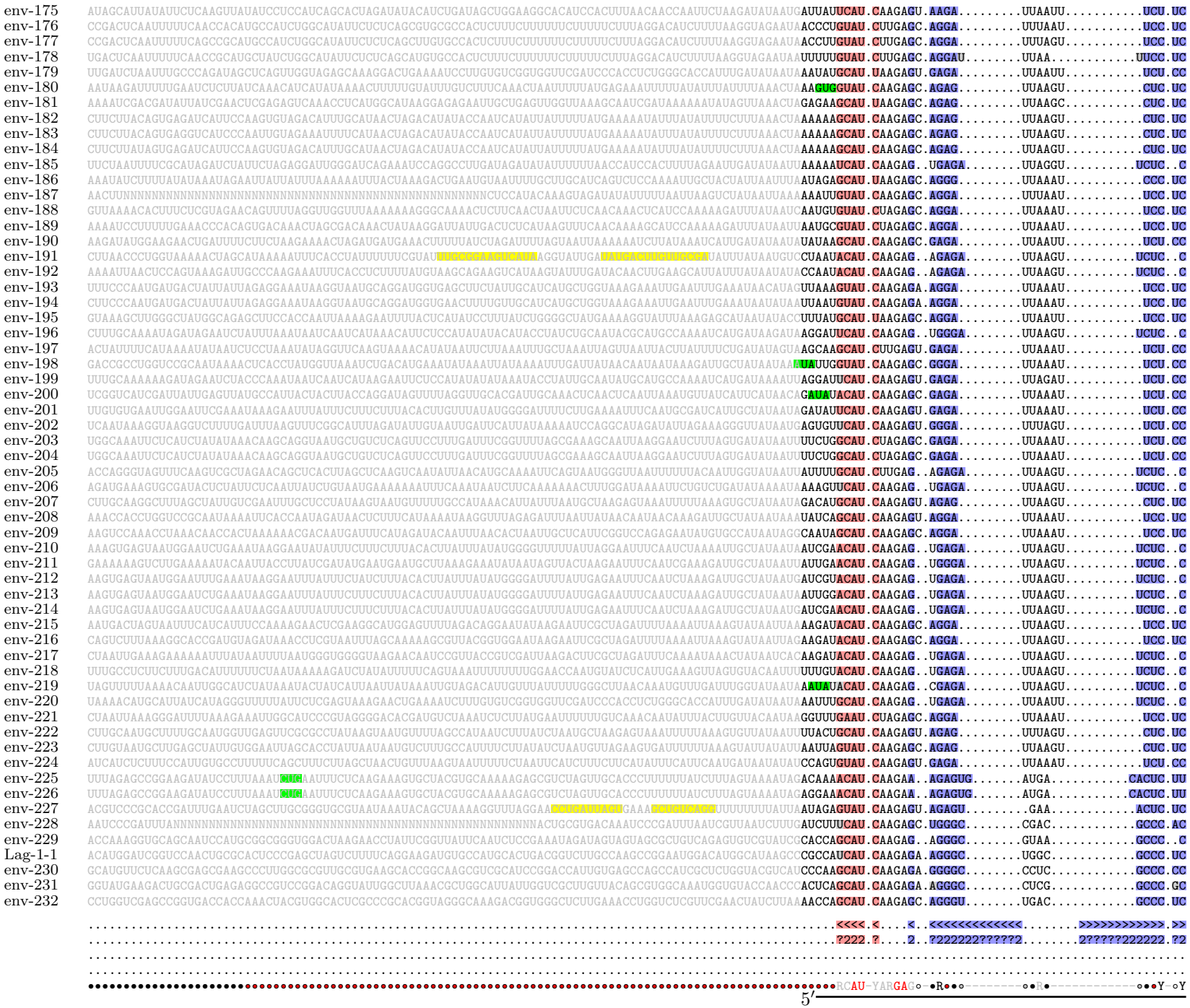
85.5 Multiple-sequence alignment

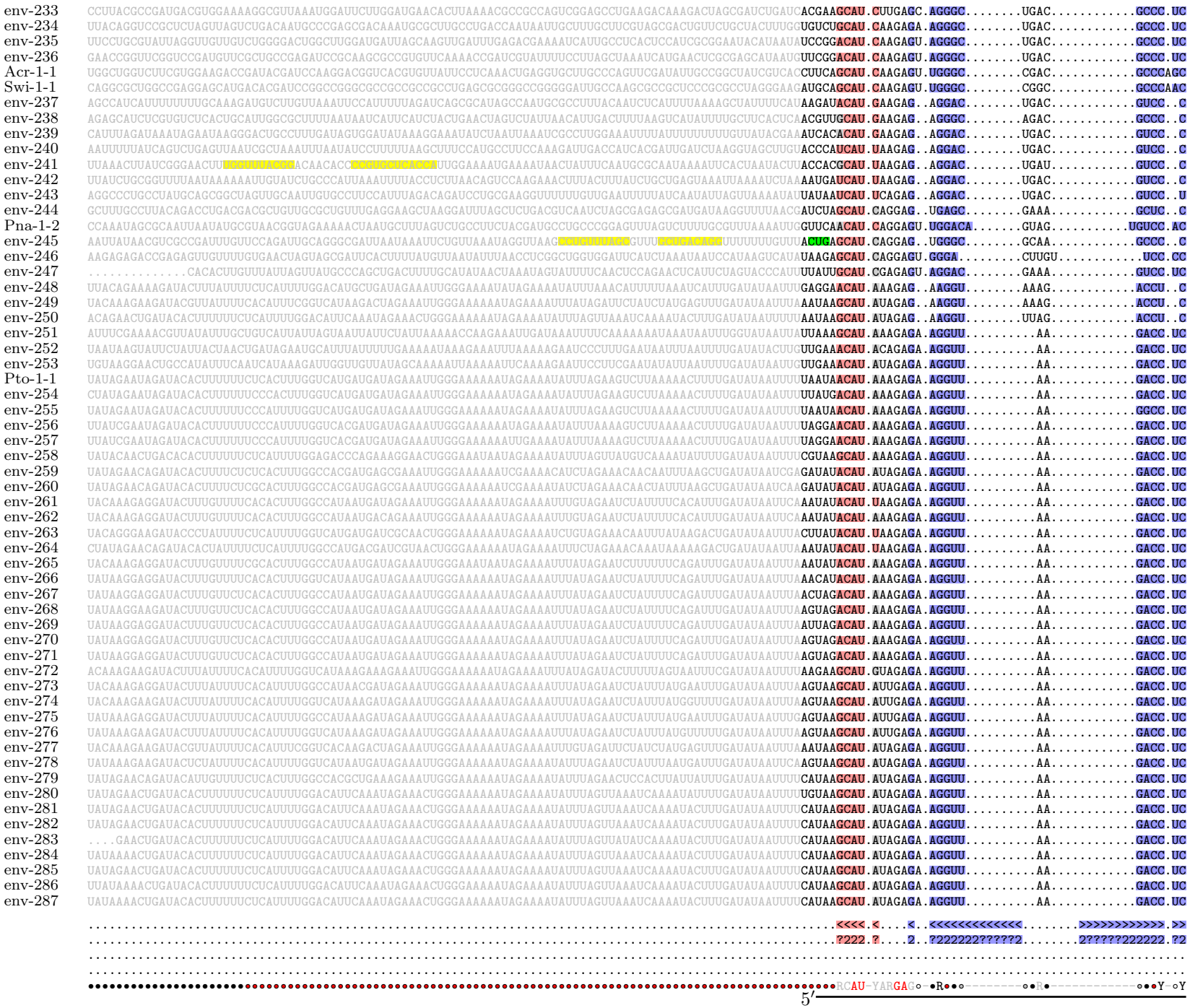
Each SAM-I-IV-variant RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 85.1. Superscript numbers refer to annotations listed in Section 85.2. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The SAM-I-IV-variant RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

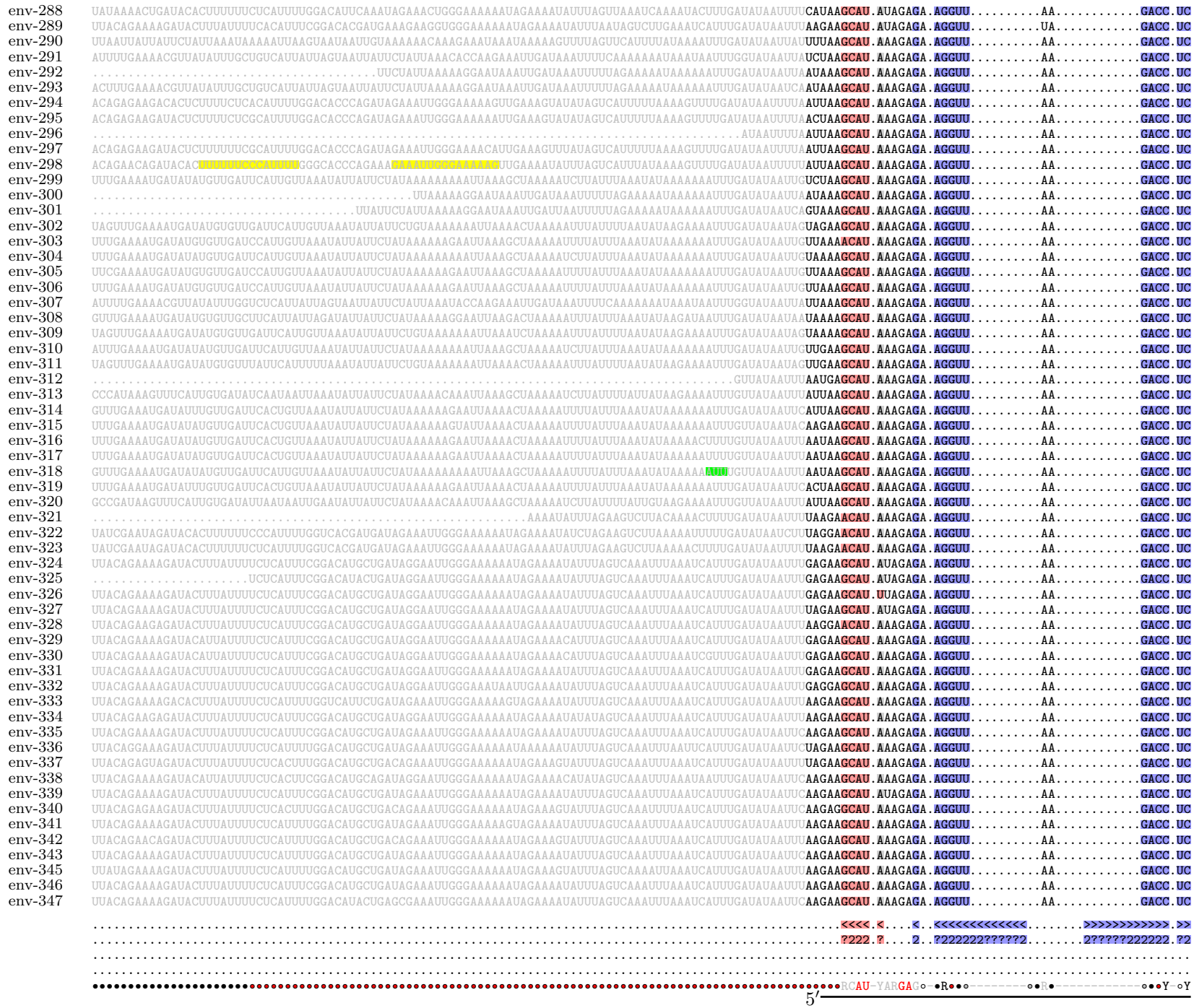

env-20
env-21
env-22
env-23
env-24
Mxa-1-1
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
Hch-1-1
env-35
env-36
env-37
env-38
env-39
env-40
Fba-1-1
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
Rbi-1-1
Psp-1-1
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
5′

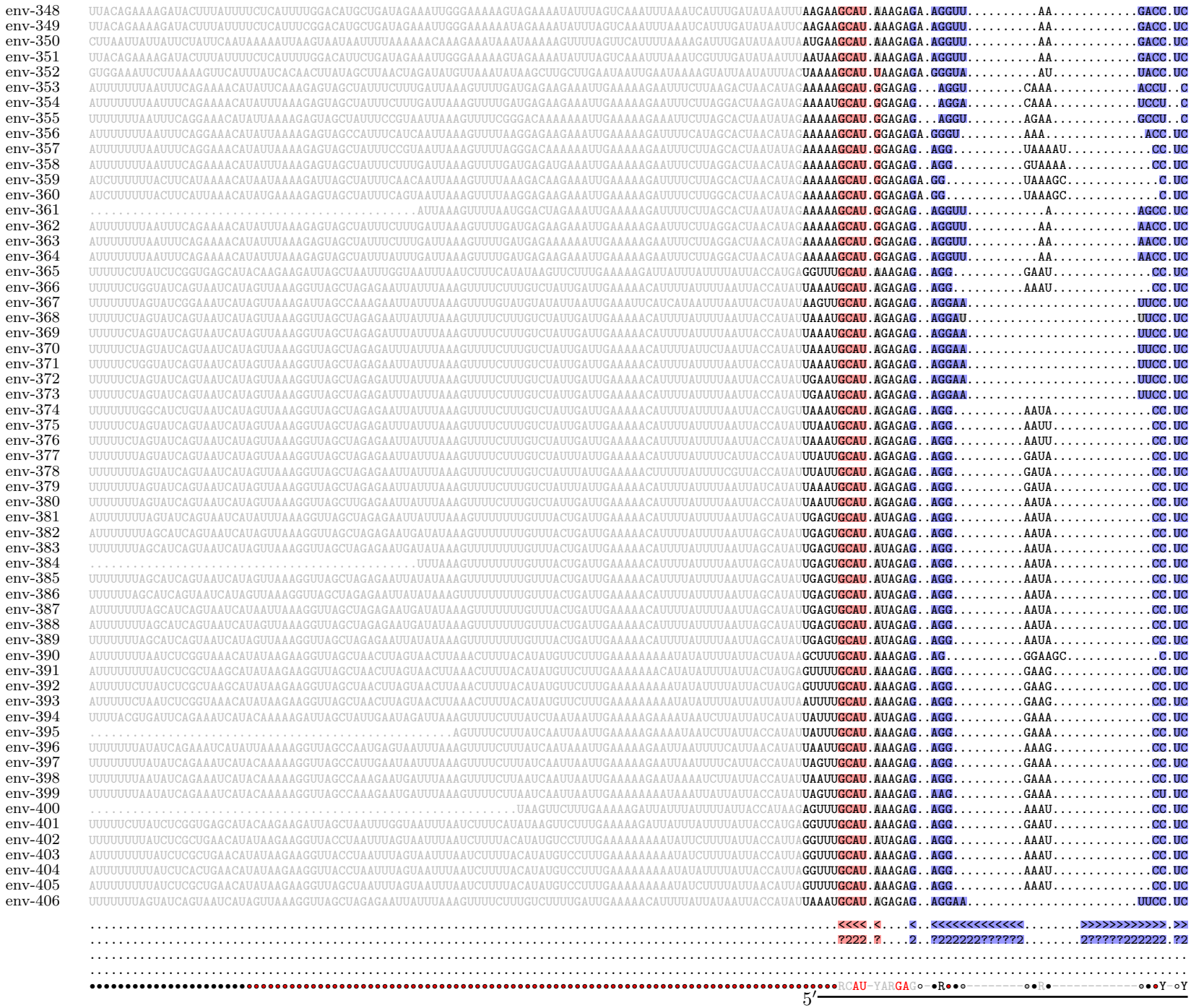

env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
Cli-1-1
Cli-1-2
env-84
Nme-3-1
Nme-1-1
Nme-2-1
Ngo-1-1
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
env-123
env-124
env-125
env-126
5′


env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-135
env-136
env-137
env-138
env-139
env-140
env-141
env-142
env-143
env-144
env-145
env-146
Nca-1-1
env-147
Bce-3-1
Bce-2-1
Bce-1-1
Bsp-1-1
env-148
env-149
Rsp-1-1
Oan-1-1
Bab-1-1
Bov-1-1
Bsu-1-1
env-150
env-151
env-152
env-153
env-154
env-155
env-156
env-157
env-158
env-159
env-160
Pna-1-1
env-161
env-162
env-163
env-164
env-165
env-166
env-167
env-168
env-169
env-170
env-171
env-172
env-173
env-174
5′

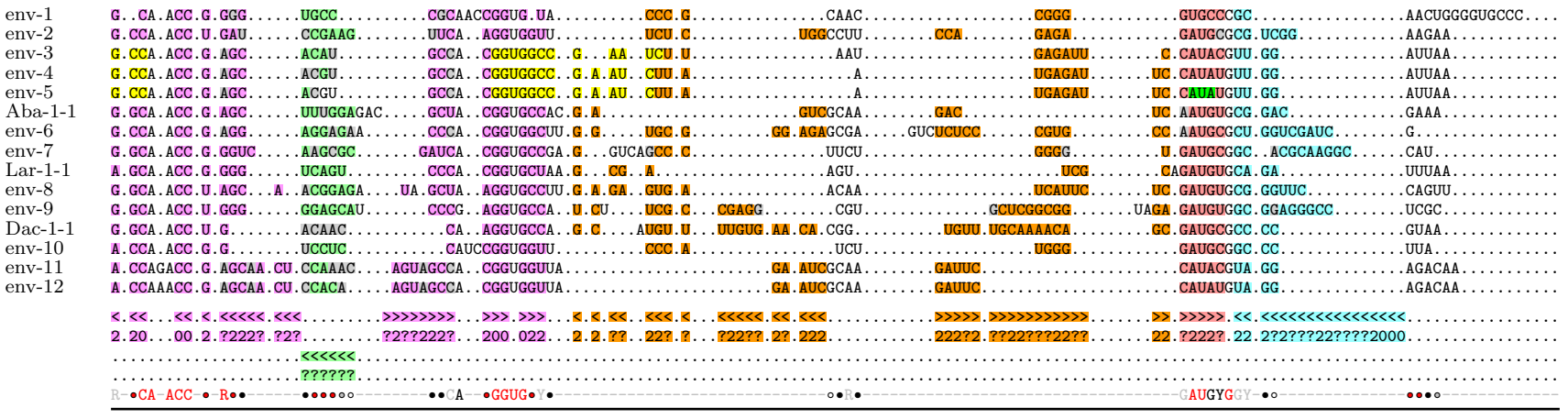

env-175
env-176
env-177
env-178
env-179
env-180
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
env-190
env-191
env-192
env-193
env-194
env-195
env-196
env-197
env-198
env-199
env-200
env-201
env-202
env-203
env-204
env-205
env-206
env-207
env-208
env-209
env-210
env-211
env-212
env-213
env-214
env-215
env-216
env-217
env-218
env-219
env-220
env-221
env-222
env-223
env-224
env-225
env-226
env-227
env-228
env-229
Lag-1-1
env-230
env-231
env-232
5′


env-233
env-234
env-235
env-236
Acr-1-1
Swi-1-1
env-237
env-238
env-239
env-240
env-241
env-242
env-243
env-244
Pna-1-2
env-245
env-246
env-247
env-248
env-249
env-250
env-251
env-252
env-253
Pto-1-1
env-254
env-255
env-256
env-257
env-258
env-259
env-260
env-261
env-262
env-263
env-264
env-265
env-266
env-267
env-268
env-269
env-270
env-271
env-272
env-273
env-274
env-275
env-276
env-277
env-278
env-279
env-280
env-281
env-282
env-283
env-284
env-285
env-286
env-287
5′

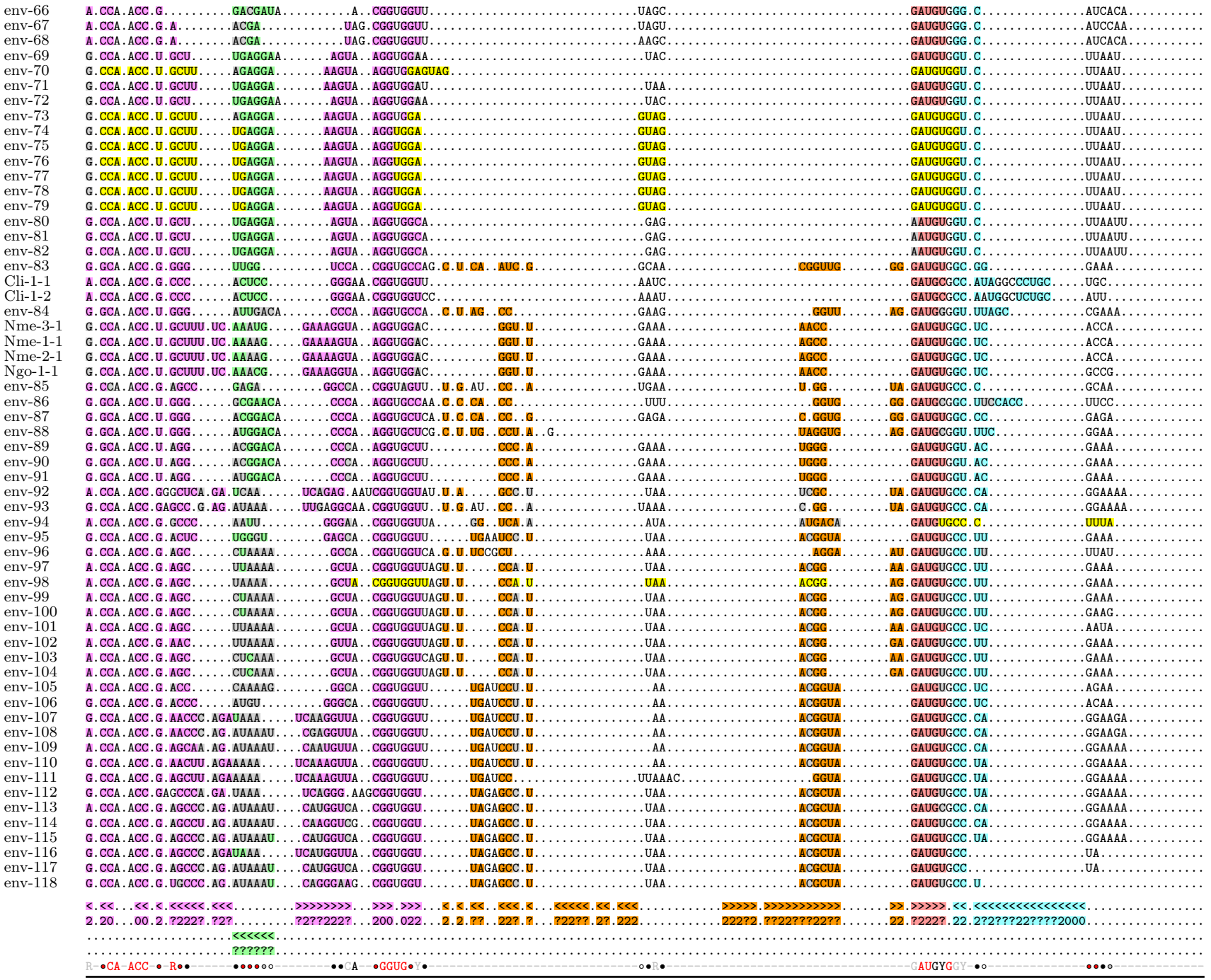

env-288
env-289
env-290
env-291
env-292
env-293
env-294
env-295
env-296
env-297
env-298
env-299
env-300
env-301
env-302
env-303
env-304
env-305
env-306
env-307
env-308
env-309
env-310
env-311
env-312
env-313
env-314
env-315
env-316
env-317
env-318
env-319
env-320
env-321
env-322
env-323
env-324
env-325
env-326
env-327
env-328
env-329
env-330
env-331
env-332
env-333
env-334
env-335
env-336
env-337
env-338
env-339
env-340
env-341
env-342
env-343
env-345
env-346
env-347
5′

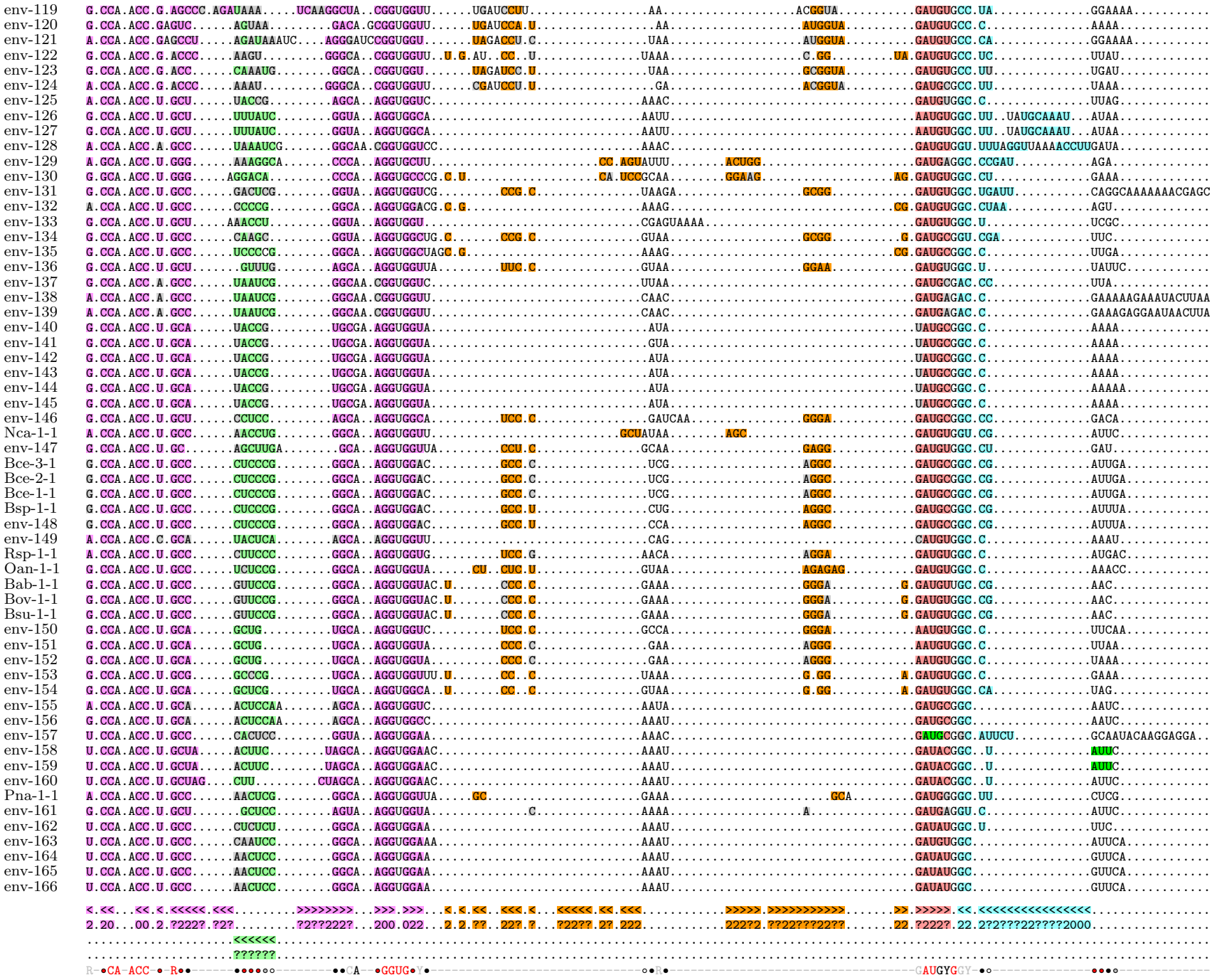

env-348
env-349
env-350
env-351
env-352
env-353
env-354
env-355
env-356
env-357
env-358
env-359
env-360
env-361
env-362
env-363
env-364
env-365
env-366
env-367
env-368
env-369
env-370
env-371
env-372
env-373
env-374
env-375
env-376
env-377
env-378
env-379
env-380
env-381
env-382
env-383
env-384
env-385
env-386
env-387
env-388
env-389
env-390
env-391
env-392
env-393
env-394
env-395
env-396
env-397
env-398
env-399
env-400
env-401
env-402
env-403
env-404
env-405
env-406
5′

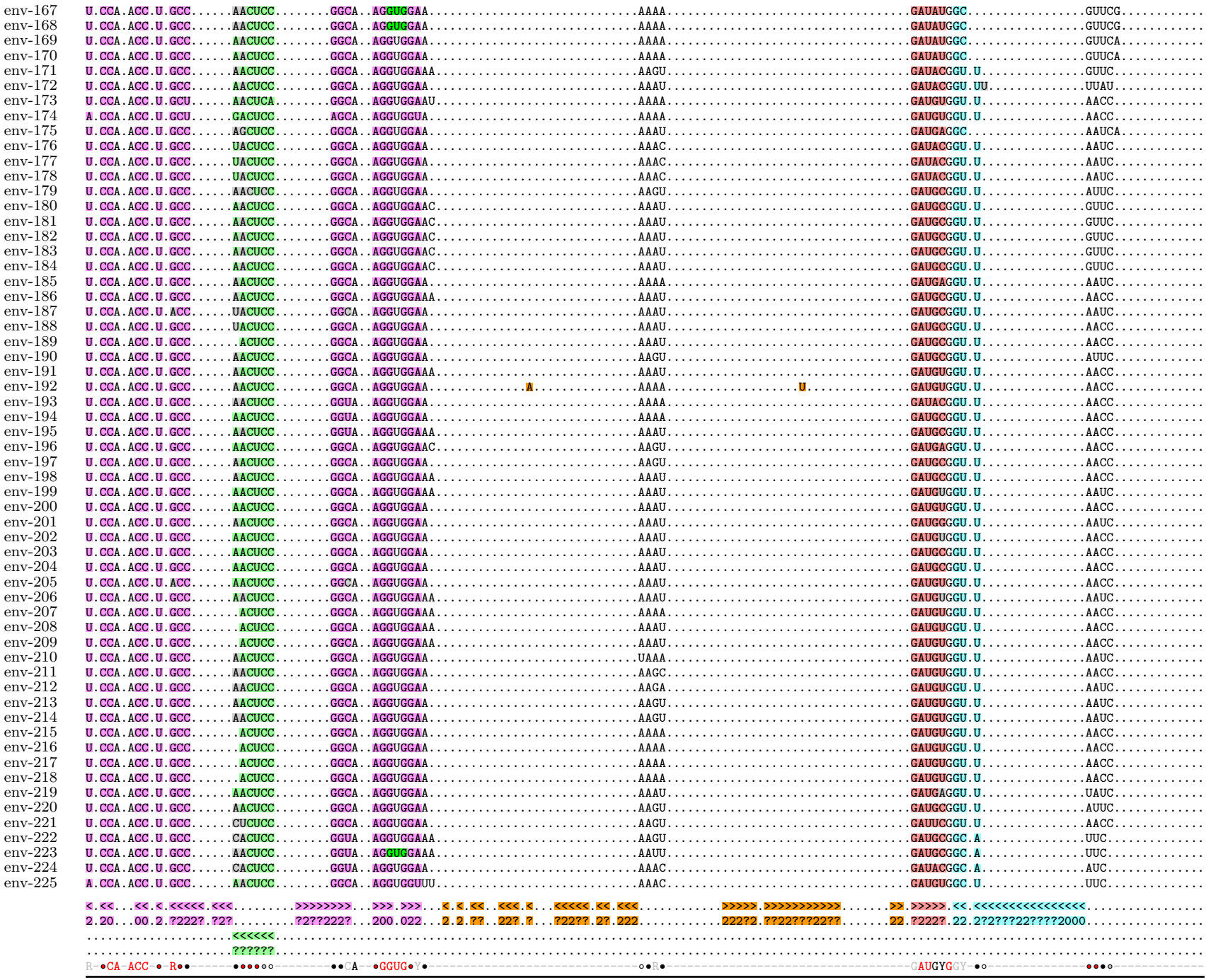
alignment positions 155 · · · 314

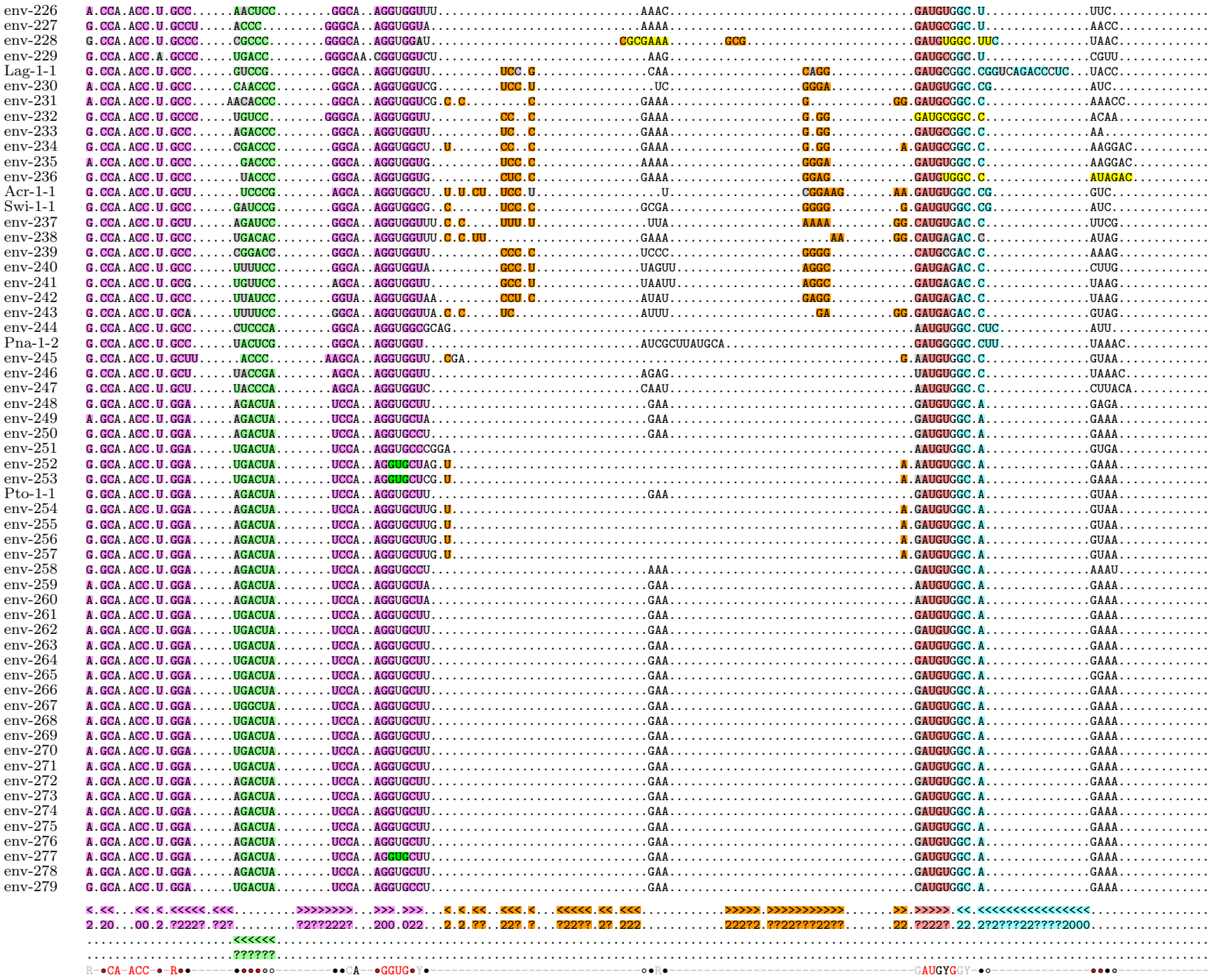


1
env-13
env-14
env-15
env-16
Cat-1-1
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
Mxa-1-1
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
Hch-1-1
env-35
env-36
env-37
env-38
env-39
env-40
Fba-1-1
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
Rbi-1-1
Psp-1-1
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65


env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
Cli-1-1
Cli-1-2
env-84
Nme-3-1
Nme-1-1
Nme-2-1
Ngo-1-1
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118

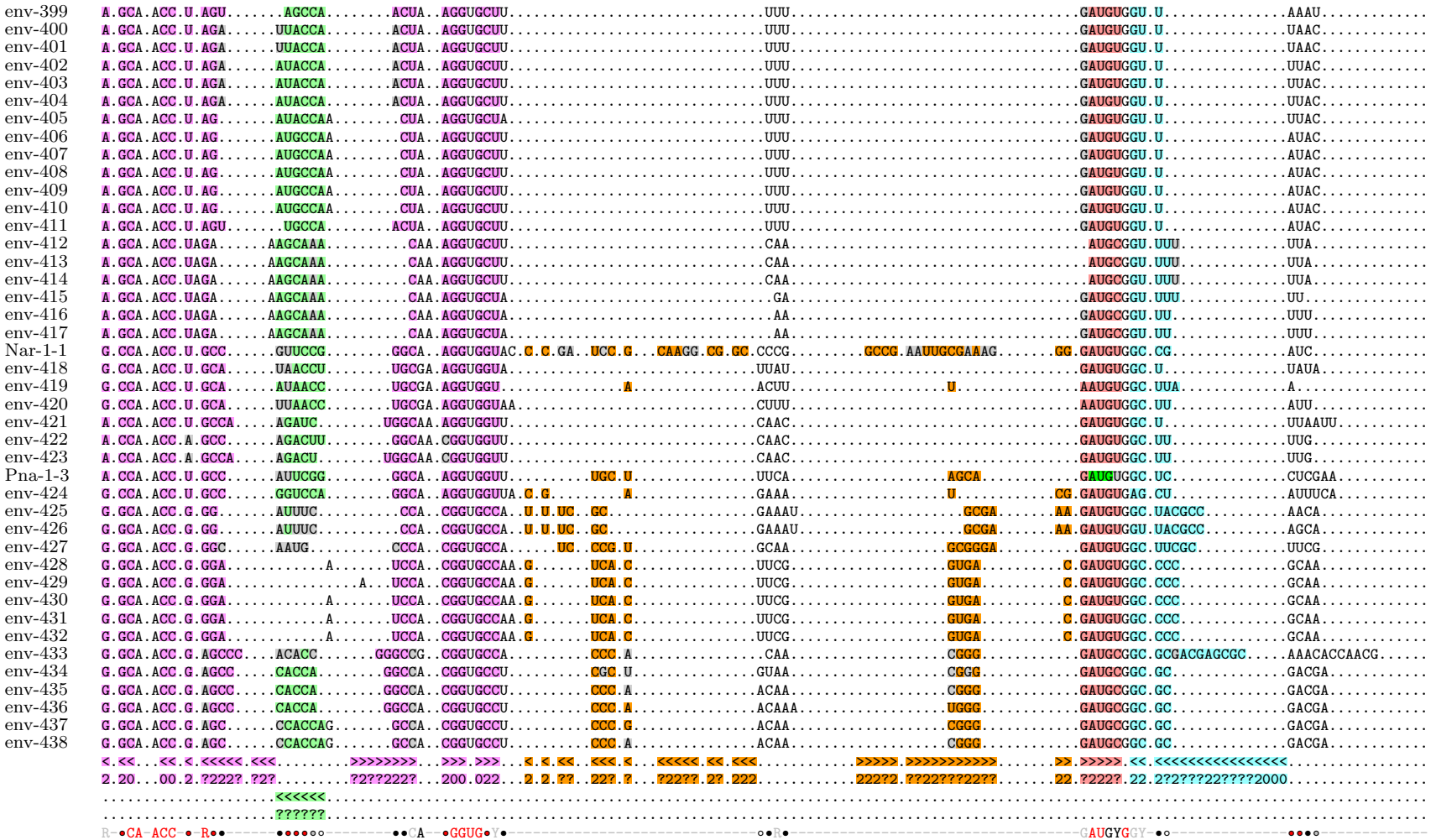

env-119
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-135
env-136
env-137
env-138
env-139
env-140
env-141
env-142
env-143
env-144
env-145
env-146
Nca-1-1
env-147
Bce-3-1
Bce-2-1
Bce-1-1
Bsp-1-1
env-148
env-149
Rsp-1-1
Oan-1-1
Bab-1-1
Bov-1-1
Bsu-1-1
env-150
env-151
env-152
env-153
env-154
env-155
env-156
env-157
env-158
env-159
env-160
Pna-1-1
env-161
env-162
env-163
env-164
env-165
env-166

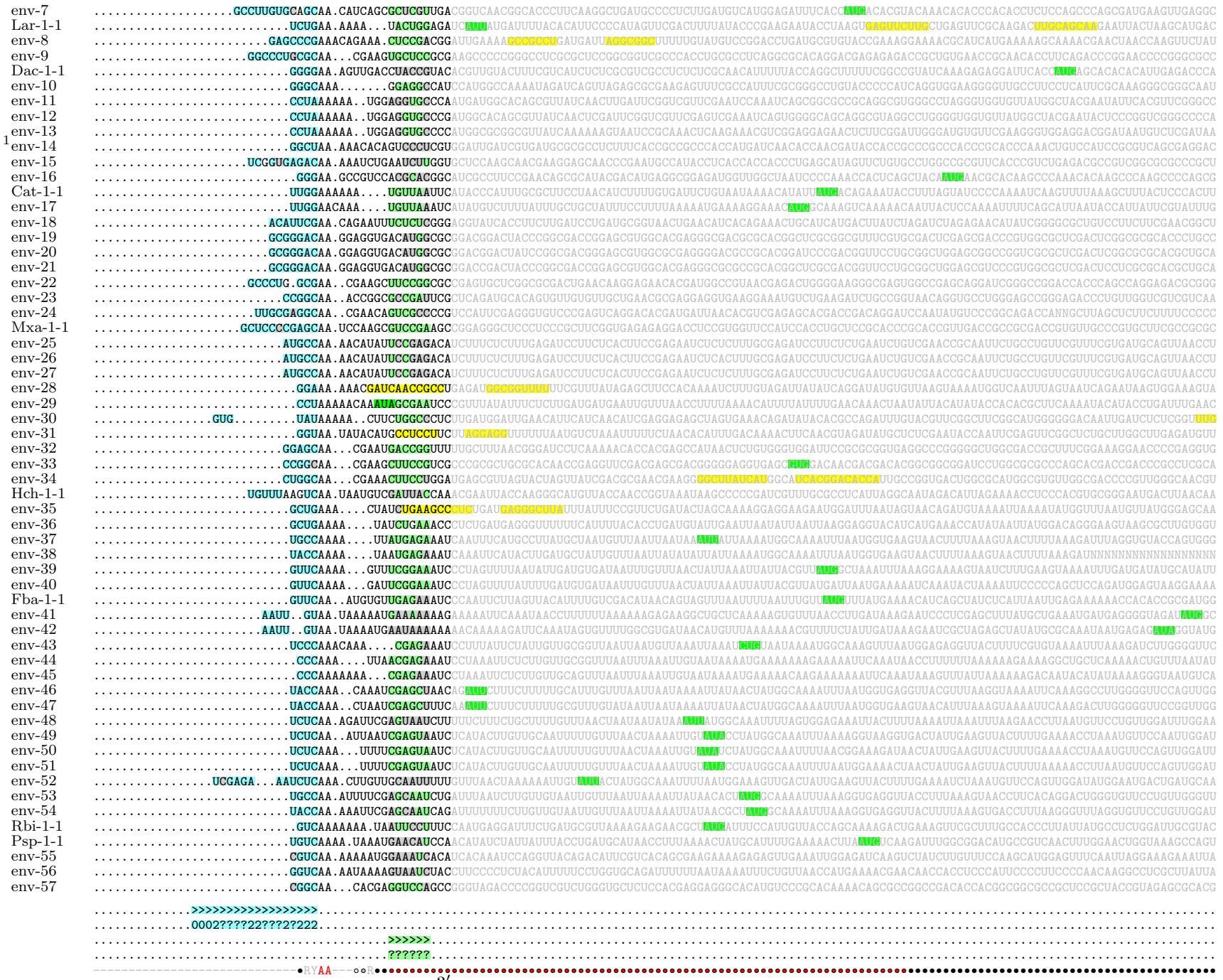

env-167
env-168
env-169
env-170
env-171
env-172
env-173
env-174
env-175
env-176
env-177
env-178
env-179
env-180
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
env-190
env-191
env-192
env-193
env-194
env-195
env-196
env-197
env-198
env-199
env-200
env-201
env-202
env-203
env-204
env-205
env-206
env-207
env-208
env-209
env-210
env-211
env-212
env-213
env-214
env-215
env-216
env-217
env-218
env-219
env-220
env-221
env-222
env-223
env-224
env-225

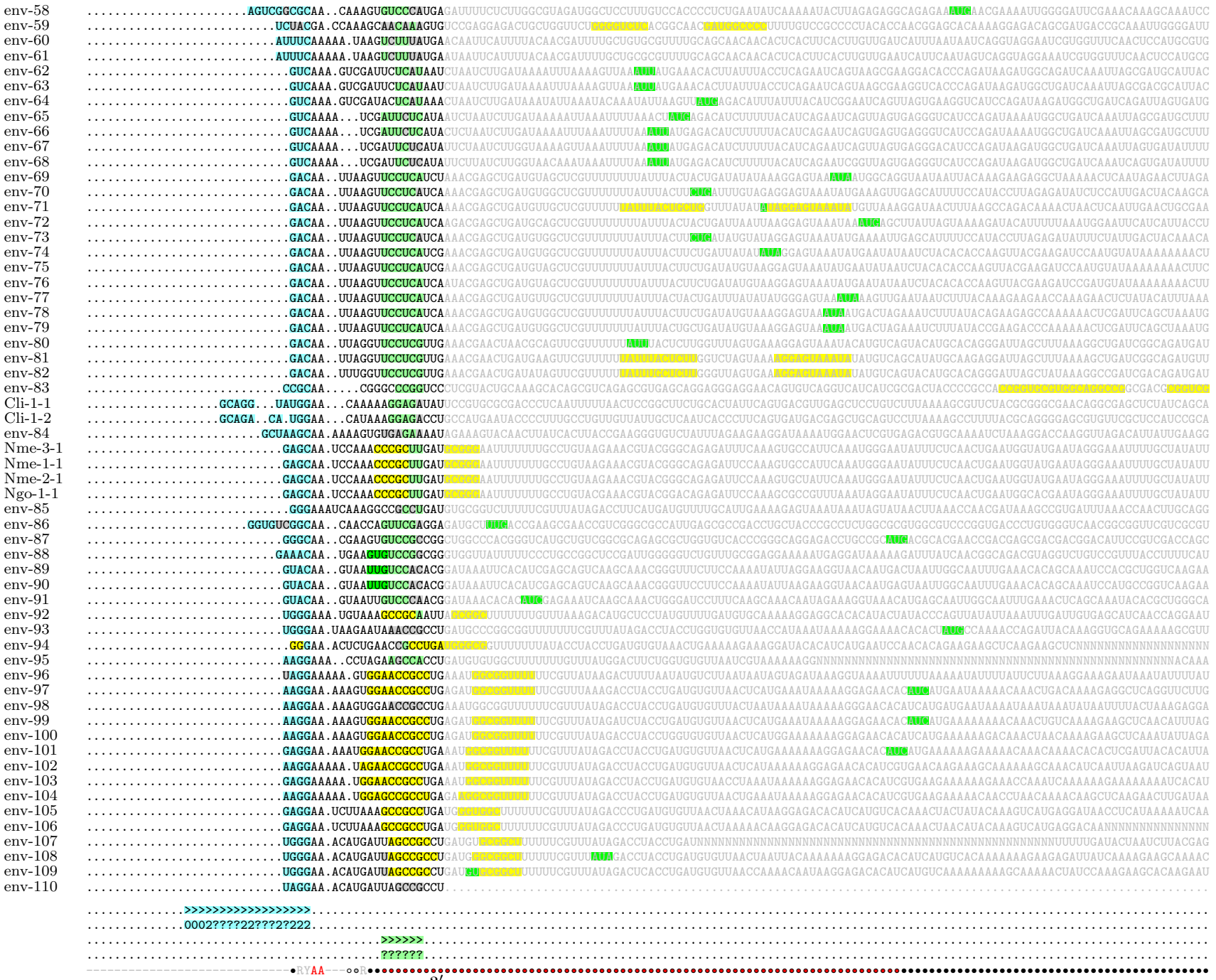

env-226
env-227
env-228
env-229
Lag-1-1
env-230
env-231
env-232
env-233
env-234
env-235
env-236
Acr-1-1
Swi-1-1
env-237
env-238
env-239
env-240
env-241
env-242
env-243
env-244
Pna-1-2
env-245
env-246
env-247
env-248
env-249
env-250
env-251
env-252
env-253
Pto-1-1
env-254
env-255
env-256
env-257
env-258
env-259
env-260
env-261
env-262
env-263
env-264
env-265
env-266
env-267
env-268
env-269
env-270
env-271
env-272
env-273
env-274
env-275
env-276
env-277
env-278
env-279

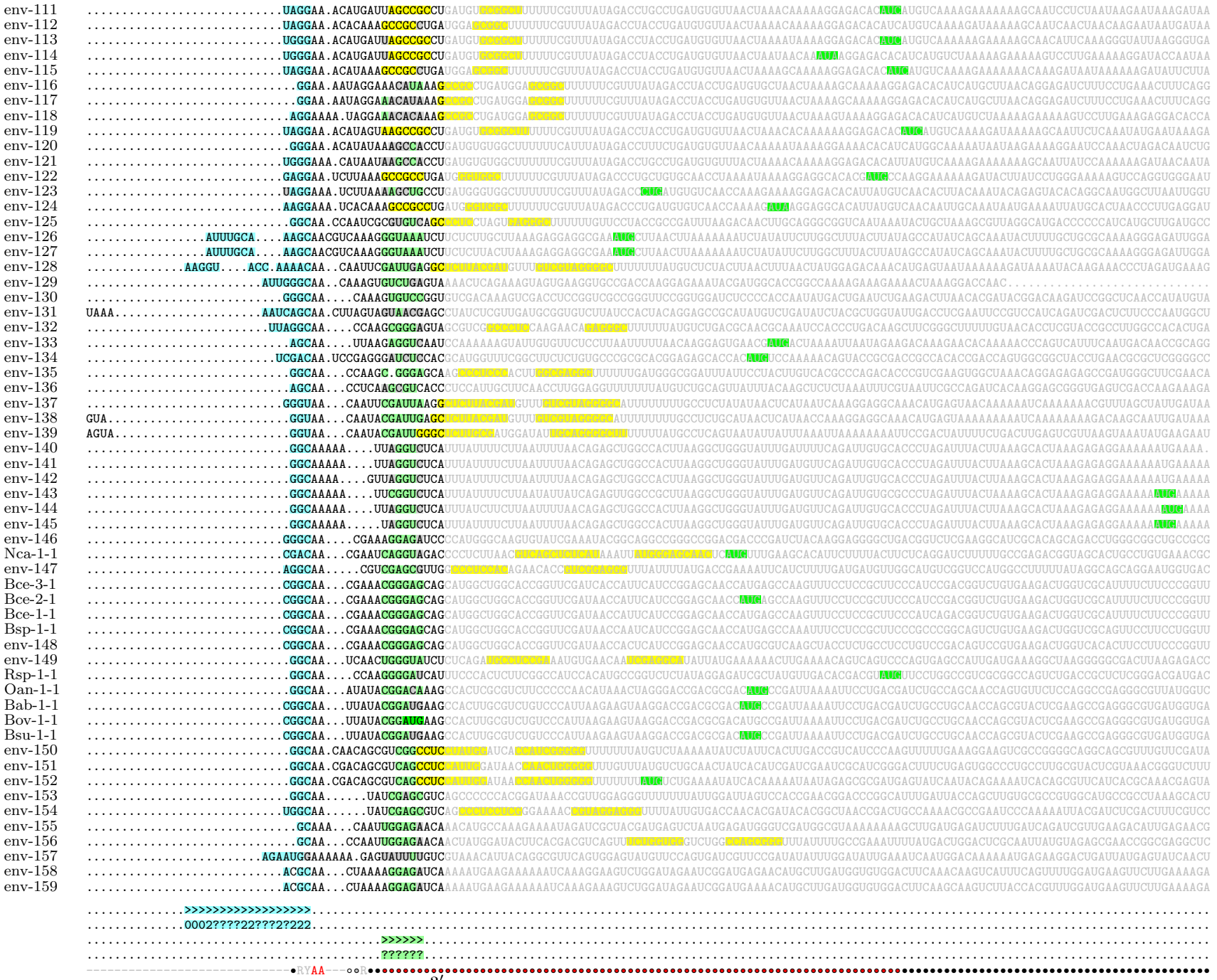
env-280
env-281
env-282
env-283
env-284
env-285
env-286
env-287
env-288
env-289
env-290
env-291
env-292
env-293
env-294
env-295
env-296
env-297
env-298
env-299
env-300
env-301
env-302
env-303
env-304
env-305
env-306
env-307
env-308
env-309
env-310
env-311
env-312
env-313
env-314
env-315
env-316
env-317
env-318
env-319
env-320
env-321
env-322
env-323
env-324
env-325
env-326
env-327
env-328
env-329
env-330
env-331
env-332
env-333
env-334
env-335
env-336
env-337
env-338

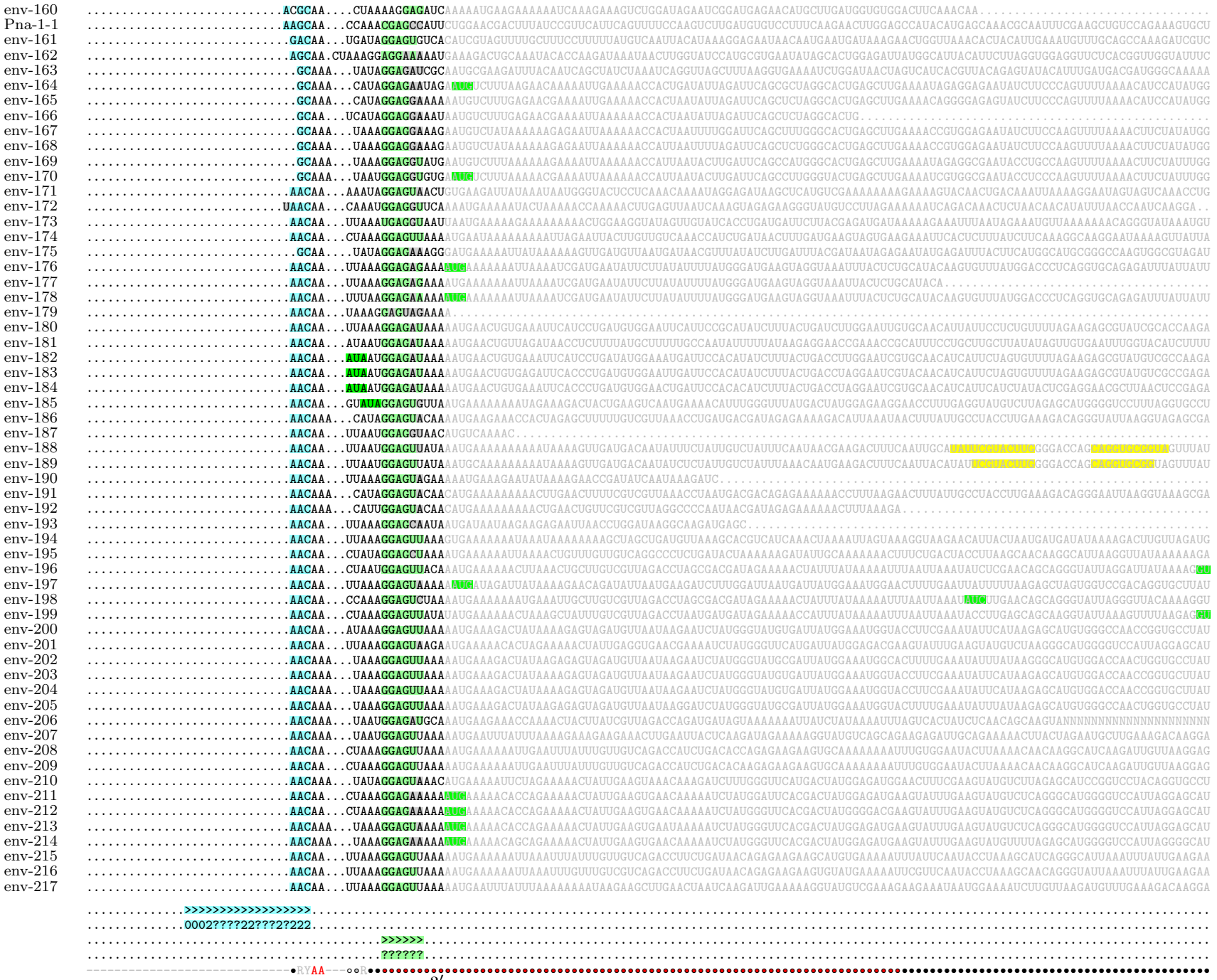
env-339
env-340
env-341
env-342
env-343
env-345
env-346
env-347
env-348
env-349
env-350
env-351
env-352
env-353
env-354
env-355
env-356
env-357
env-358
env-359
env-360
env-361
env-362
env-363
env-364
env-365
env-366
env-367
env-368
env-369
env-370
env-371
env-372
env-373
env-374
env-375
env-376
env-377
env-378
env-379
env-380
env-381
env-382
env-383
env-384
env-385
env-386
env-387
env-388
env-389
env-390
env-391
env-392
env-393
env-394
env-395
env-396
env-397
env-398

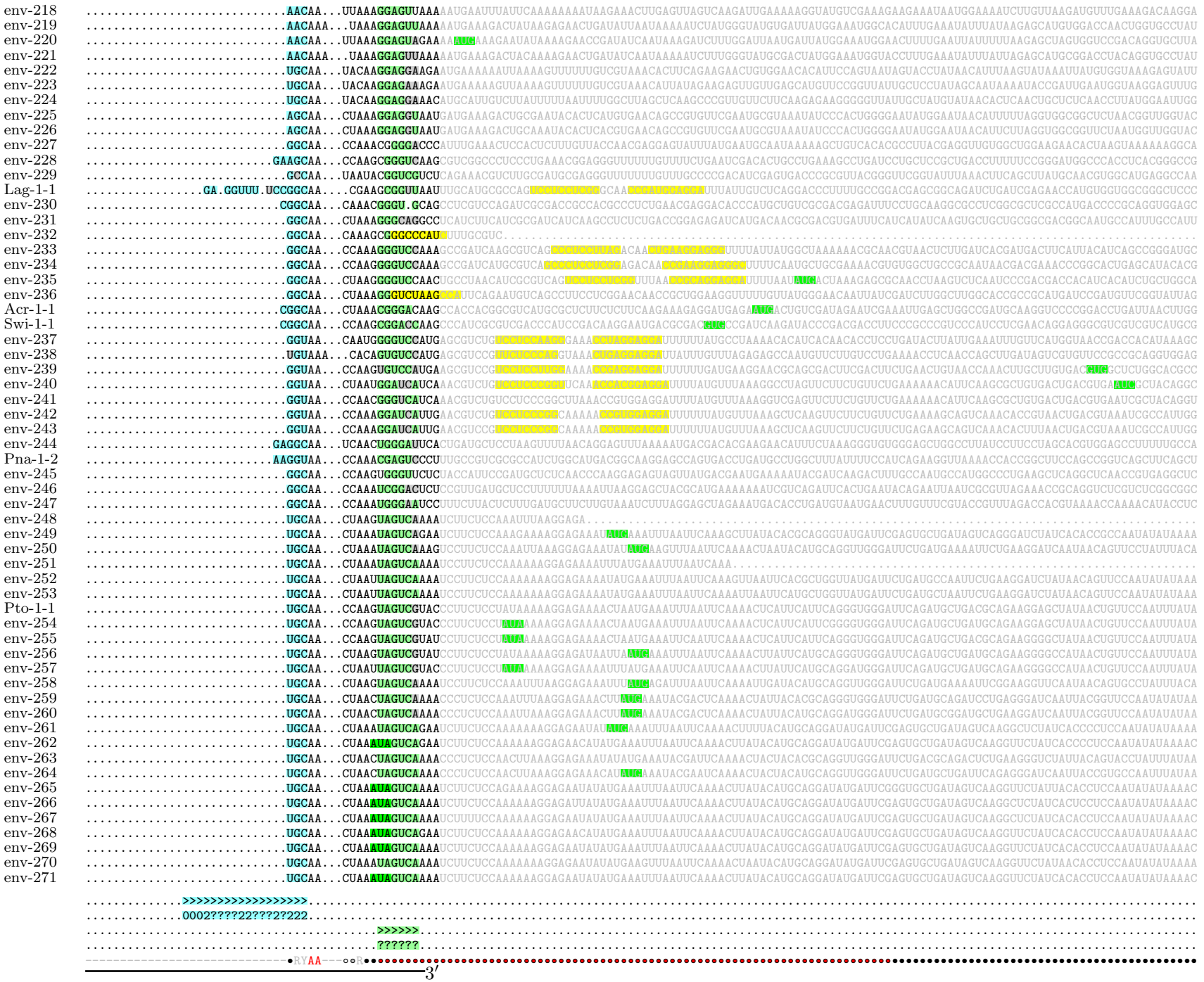

env-399
env-400
env-401
env-402
env-403
env-404
env-405
env-406
env-407
env-408
env-409
env-410
env-411
env-412
env-413
env-414
env-415
env-416
env-417
Nar-1-1
env-418
env-419
env-420
env-421
env-422
env-423
Pna-1-3
env-424
env-425
env-426
env-427
env-428
env-429
env-430
env-431
env-432
env-433
env-434
env-435
env-436
env-437
env-438
alignment positions 315 · · · 474
env-1
env-2
env-3
env-4
env-5
Aba-1-1
env-6
3′

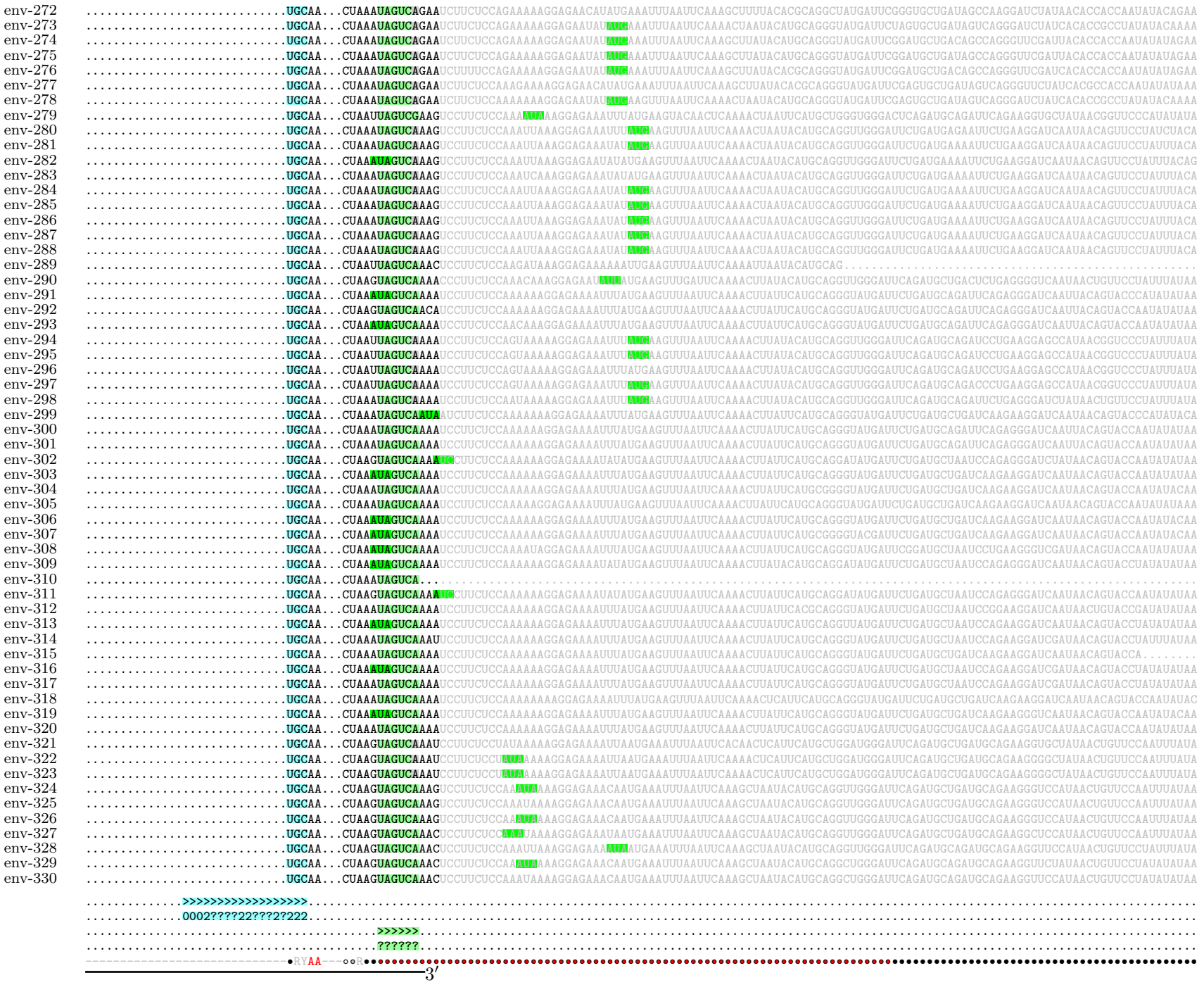

env-7
Lar-1-1
env-8
env-9
Dac-1-1
env-10
env-11
env-12
env-13
env-14
env-15
env-16
Cat-1-1
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
Mxa-1-1
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
Hch-1-1
env-35
env-36
env-37
env-38
env-39
env-40
Fba-1-1
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
Rbi-1-1
Psp-1-1
env-55
env-56
env-57
3′

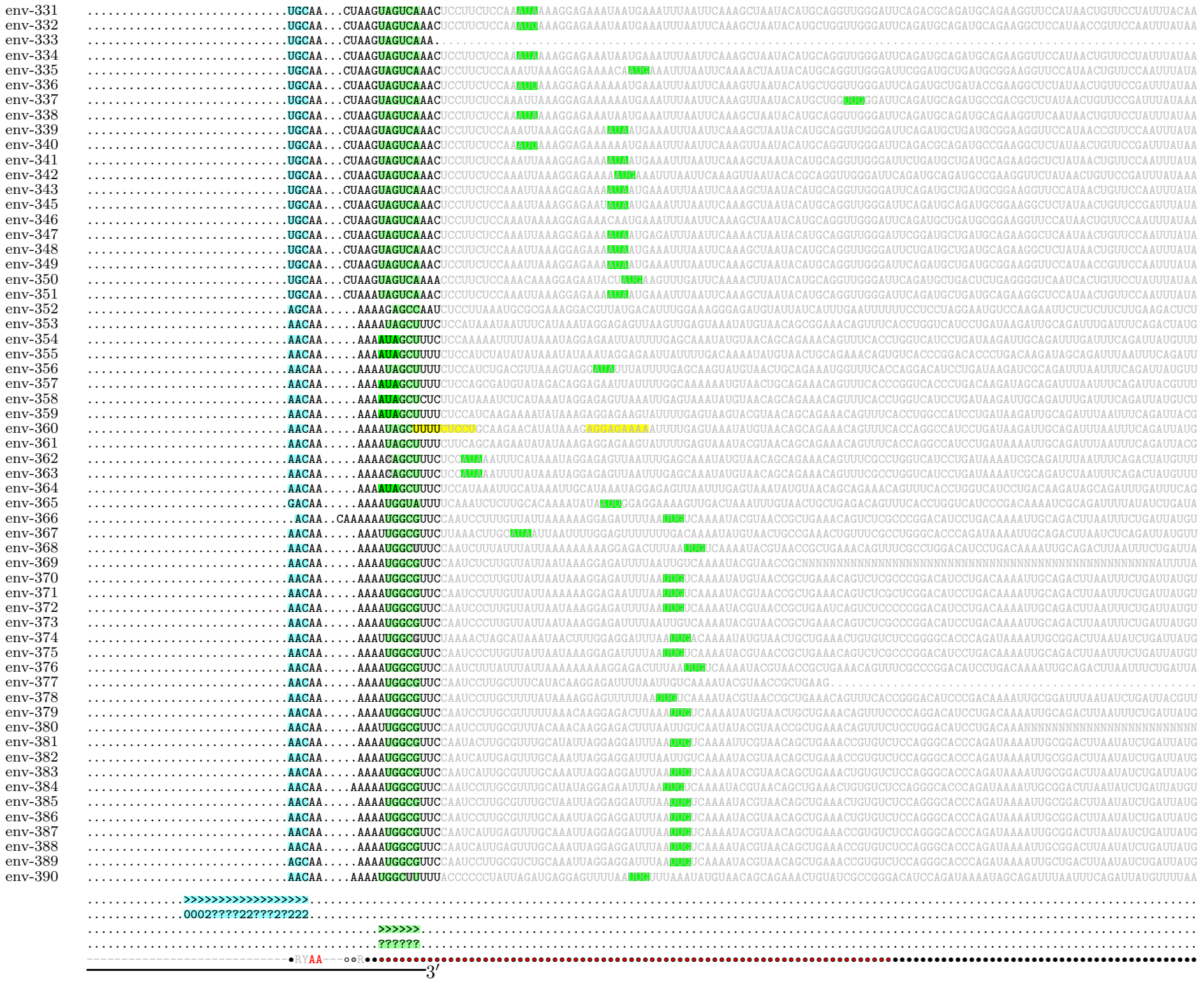

env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
Cli-1-1
Cli-1-2
env-84
Nme-3-1
Nme-1-1
Nme-2-1
Ngo-1-1
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
3′


env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-135
env-136
env-137
env-138
env-139
env-140
env-141
env-142
env-143
env-144
env-145
env-146
Nca-1-1
env-147
Bce-3-1
Bce-2-1
Bce-1-1
Bsp-1-1
env-148
env-149
Rsp-1-1
Oan-1-1
Bab-1-1
Bov-1-1
Bsu-1-1
env-150
env-151
env-152
env-153
env-154
env-155
env-156
env-157
env-158
env-159
3′


env-160
Pna-1-1
env-161
env-162
env-163
env-164
env-165
env-166
env-167
env-168
env-169
env-170
env-171
env-172
env-173
env-174
env-175
env-176
env-177
env-178
env-179
env-180
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
env-190
env-191
env-192
env-193
env-194
env-195
env-196
env-197
env-198
env-199
env-200
env-201
env-202
env-203
env-204
env-205
env-206
env-207
env-208
env-209
env-210
env-211
env-212
env-213
env-214
env-215
env-216
env-217
3′


env-218
env-219
env-220
env-221
env-222
env-223
env-224
env-225
env-226
env-227
env-228
env-229
Lag-1-1
env-230
env-231
env-232
env-233
env-234
env-235
env-236
Acr-1-1
Swi-1-1
env-237
env-238
env-239
env-240
env-241
env-242
env-243
env-244
Pna-1-2
env-245
env-246
env-247
env-248
env-249
env-250
env-251
env-252
env-253
Pto-1-1
env-254
env-255
env-256
env-257
env-258
env-259
env-260
env-261
env-262
env-263
env-264
env-265
env-266
env-267
env-268
env-269
env-270
env-271
3′


env-272
env-273
env-274
env-275
env-276
env-277
env-278
env-279
env-280
env-281
env-282
env-283
env-284
env-285
env-286
env-287
env-288
env-289
env-290
env-291
env-292
env-293
env-294
env-295
env-296
env-297
env-298
env-299
env-300
env-301
env-302
env-303
env-304
env-305
env-306
env-307
env-308
env-309
env-310
env-311
env-312
env-313
env-314
env-315
env-316
env-317
env-318
env-319
env-320
env-321
env-322
env-323
env-324
env-325
env-326
env-327
env-328
env-329
env-330
3′

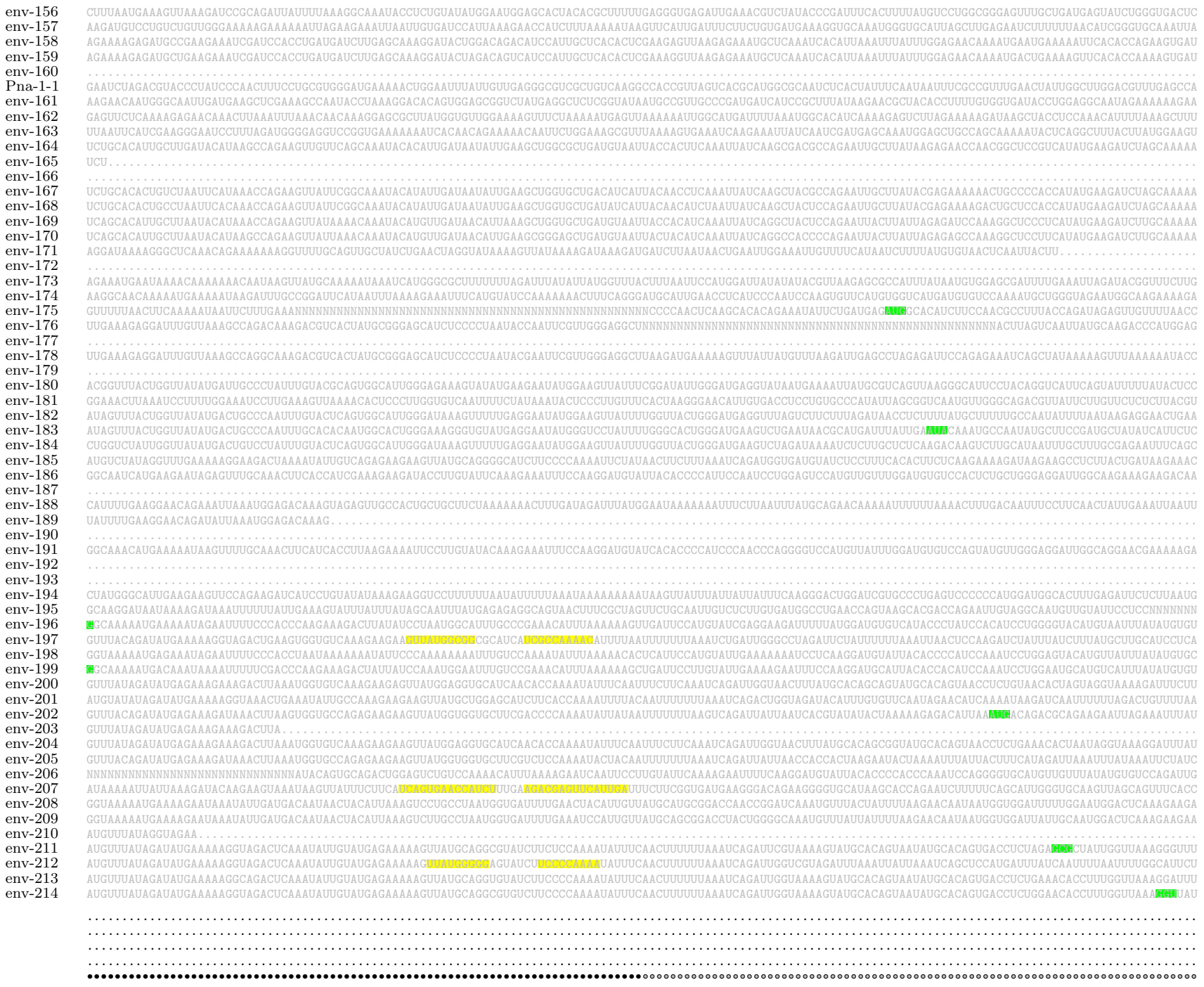

env-331
env-332
env-333
env-334
env-335
env-336
env-337
env-338
env-339
env-340
env-341
env-342
env-343
env-345
env-346
env-347
env-348
env-349
env-350
env-351
env-352
env-353
env-354
env-355
env-356
env-357
env-358
env-359
env-360
env-361
env-362
env-363
env-364
env-365
env-366
env-367
env-368
env-369
env-370
env-371
env-372
env-373
env-374
env-375
env-376
env-377
env-378
env-379
env-380
env-381
env-382
env-383
env-384
env-385
env-386
env-387
env-388
env-389
env-390
3′

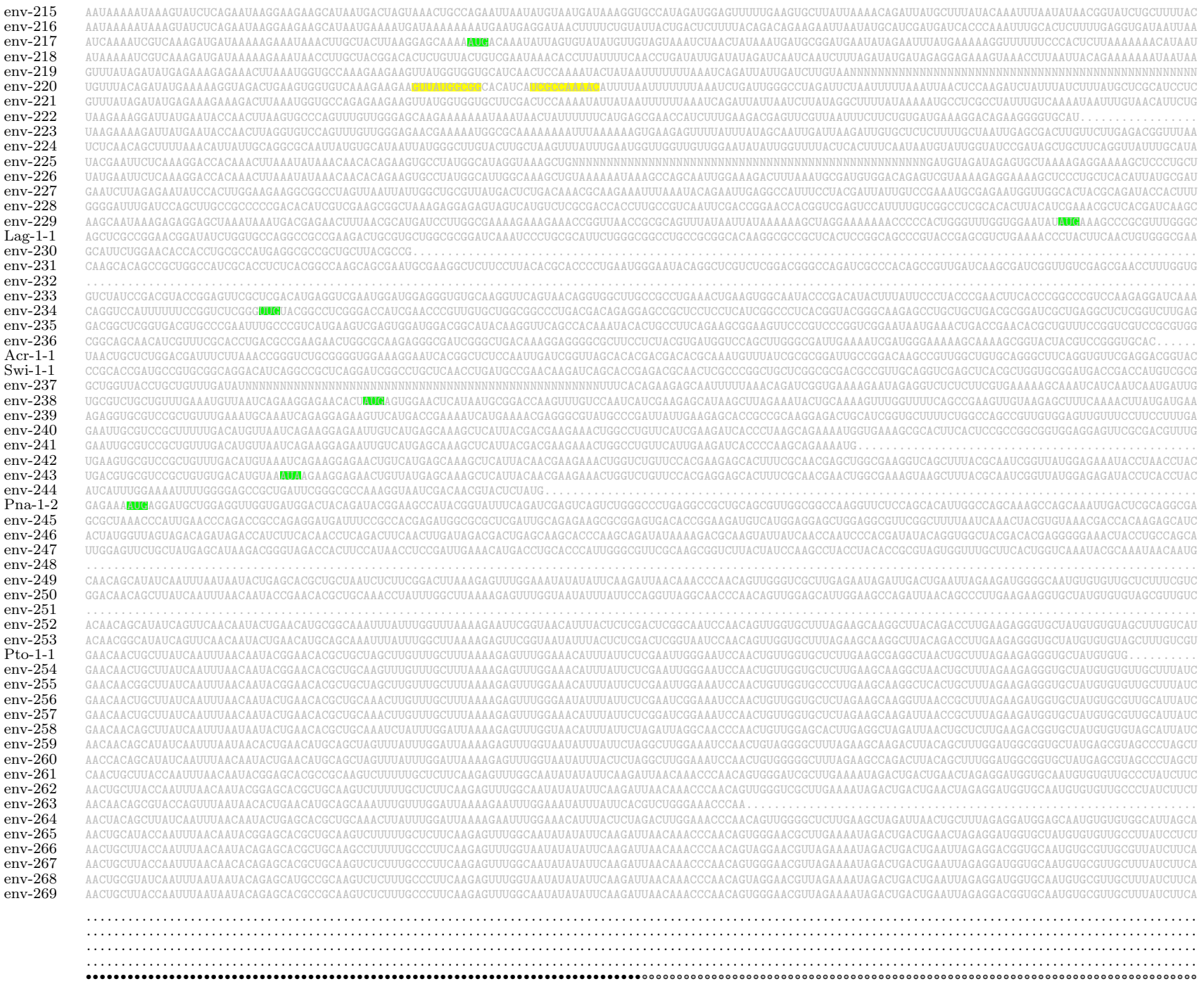
env-391
env-392
env-393
env-394
env-395
env-396
env-397
env-398
env-399
env-400
env-401
env-402
env-403
env-404
env-405
env-406
env-407
env-408
env-409
env-410
env-411
env-412
env-413
env-414
env-415
env-416
env-417
Nar-1-1
env-418
env-419
env-420
env-421
env-422
env-423
Pna-1-3
env-424
env-425
env-426
env-427
env-428
env-429
env-430
env-431
env-432
env-433
env-434
env-435
env-436
env-437
env-438
3′
alignment positions 475 · · · 634

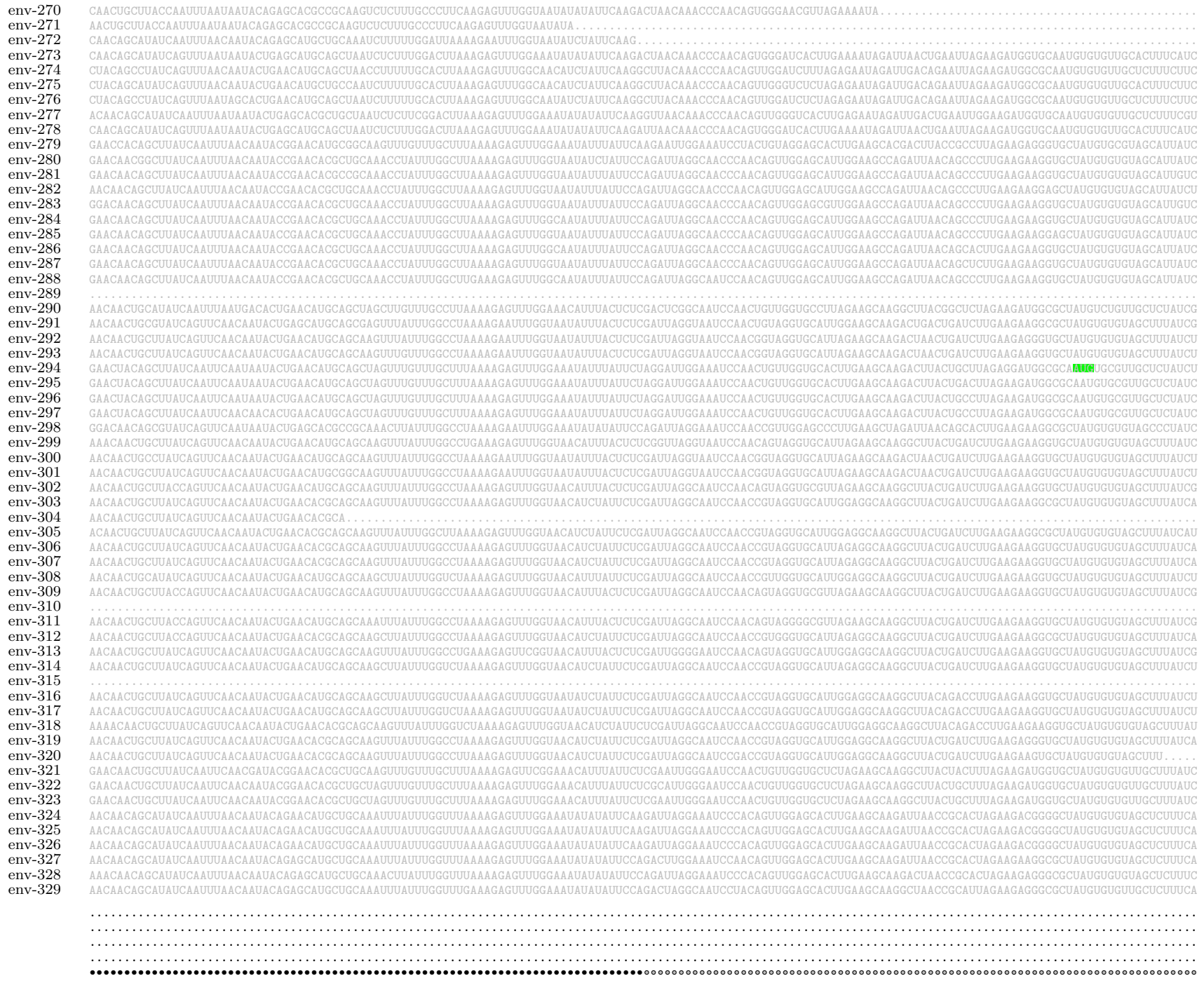
env-1
env-2
env-3
env-4
env-5
Aba-1-1
env-6
env-7
Lar-1-1
env-8
env-9
Dac-1-1
env-10
env-11
env-12
env-13
env-14
env-15
env-16
Cat-1-1
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
Mxa-1-1
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
Hch-1-1
env-35
env-36
env-37
env-38
env-39
env-40
Fba-1-1
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53

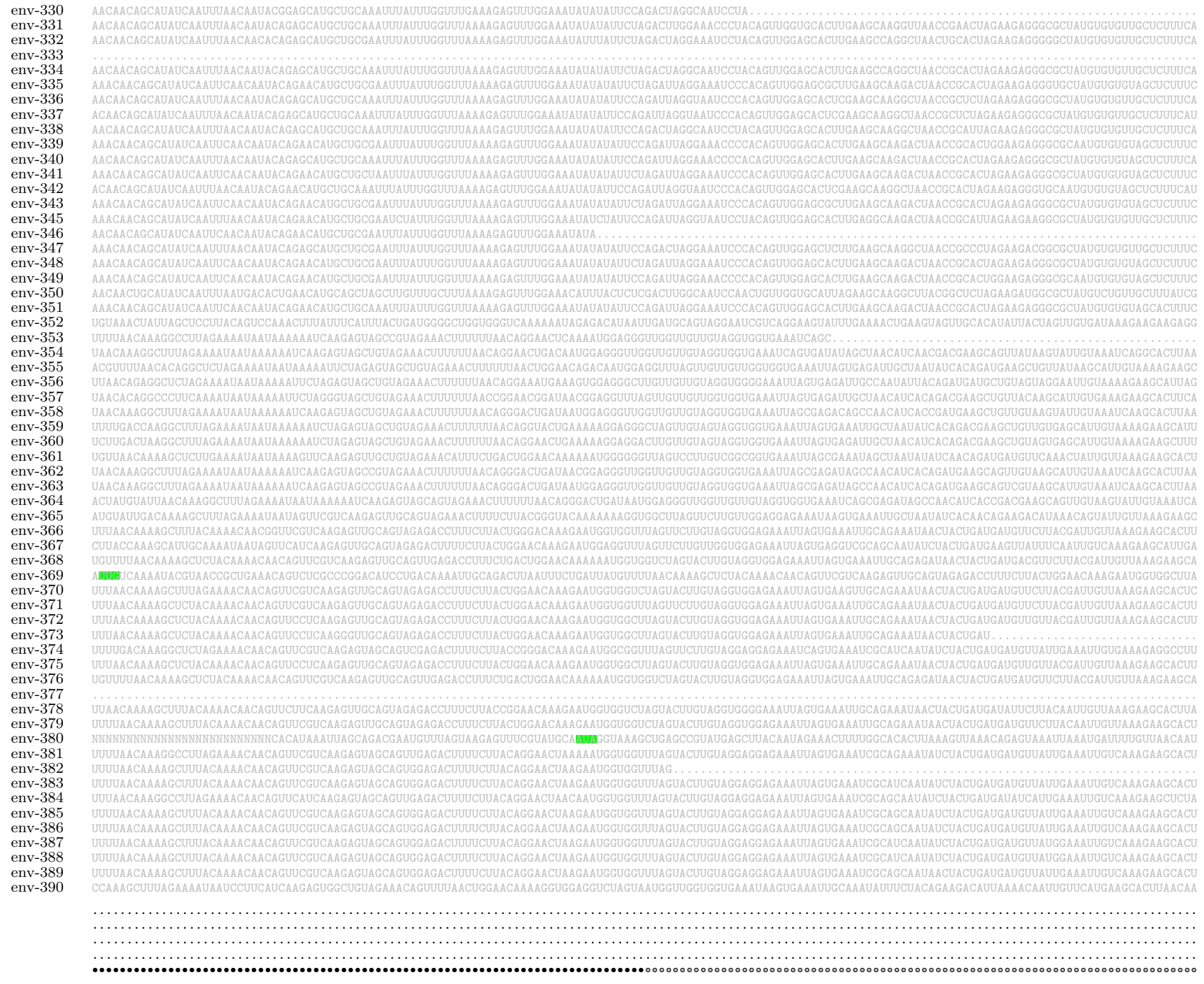
env-54
Rbi-1-1
Psp-1-1
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
Cli-1-1
Cli-1-2
env-84
Nme-3-1
Nme-1-1
Nme-2-1
Ngo-1-1
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105

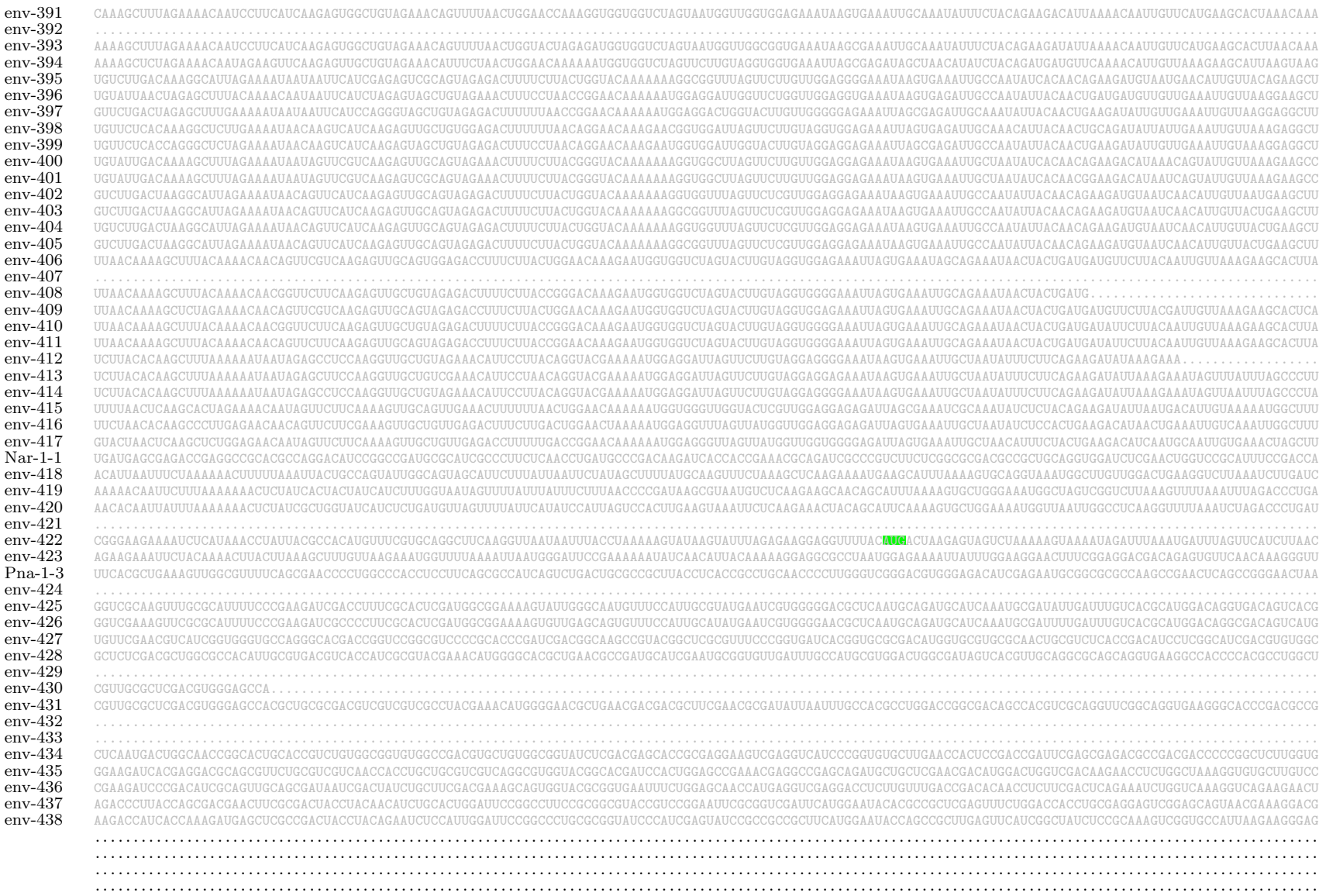
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-135
env-136
env-137
env-138
env-139
env-140
env-141
env-142
env-143
env-144
env-145
env-146
Nca-1-1
env-147
Bce-3-1
Bce-2-1
Bce-1-1
Bsp-1-1
env-148
env-149
Rsp-1-1
Oan-1-1
Bab-1-1
Bov-1-1
Bsu-1-1
env-150
env-151
env-152
env-153
env-154
env-155


env-156
env-157
env-158
env-159
env-160
Pna-1-1
env-161
env-162
env-163
env-164
env-165
env-166
env-167
env-168
env-169
env-170
env-171
env-172
env-173
env-174
env-175
env-176
env-177
env-178
env-179
env-180
env-181
env-182
env-183
env-184
env-185
env-186
env-187
env-188
env-189
env-190
env-191
env-192
env-193
env-194
env-195
env-196
env-197
env-198
env-199
env-200
env-201
env-202
env-203
env-204
env-205
env-206
env-207
env-208
env-209
env-210
env-211
env-212
env-213
env-214


env-215
env-216
env-217
env-218
env-219
env-220
env-221
env-222
env-223
env-224
env-225
env-226
env-227
env-228
env-229
Lag-1-1
env-230
env-231
env-232
env-233
env-234
env-235
env-236
Acr-1-1
Swi-1-1
env-237
env-238
env-239
env-240
env-241
env-242
env-243
env-244
Pna-1-2
env-245
env-246
env-247
env-248
env-249
env-250
env-251
env-252
env-253
Pto-1-1
env-254
env-255
env-256
env-257
env-258
env-259
env-260
env-261
env-262
env-263
env-264
env-265
env-266
env-267
env-268
env-269


env-270
env-271
env-272
env-273
env-274
env-275
env-276
env-277
env-278
env-279
env-280
env-281
env-282
env-283
env-284
env-285
env-286
env-287
env-288
env-289
env-290
env-291
env-292
env-293
env-294
env-295
env-296
env-297
env-298
env-299
env-300
env-301
env-302
env-303
env-304
env-305
env-306
env-307
env-308
env-309
env-310
env-311
env-312
env-313
env-314
env-315
env-316
env-317
env-318
env-319
env-320
env-321
env-322
env-323
env-324
env-325
env-326
env-327
env-328
env-329


env-330
env-331
env-332
env-333
env-334
env-335
env-336
env-337
env-338
env-339
env-340
env-341
env-342
env-343
env-345
env-346
env-347
env-348
env-349
env-350
env-351
env-352
env-353
env-354
env-355
env-356
env-357
env-358
env-359
env-360
env-361
env-362
env-363
env-364
env-365
env-366
env-367
env-368
env-369
env-370
env-371
env-372
env-373
env-374
env-375
env-376
env-377
env-378
env-379
env-380
env-381
env-382
env-383
env-384
env-385
env-386
env-387
env-388
env-389
env-390


env-391
env-392
env-393
env-394
env-395
env-396
env-397
env-398
env-399
env-400
env-401
env-402
env-403
env-404
env-405
env-406
env-407
env-408
env-409
env-410
env-411
env-412
env-413
env-414
env-415
env-416
env-417
Nar-1-1
env-418
env-419
env-420
env-421
env-422
env-423
Pna-1-3
env-424
env-425
env-426
env-427
env-428
env-429
env-430
env-431
env-432
env-433
env-434
env-435
env-436
env-437
env-438
alignment positions 635 · · · 647
env-1

```
env-2       UCUCCGCGACACC
env-3       GCGGCAUCAUGGU
env-4       UGCGGGAUUAUGG
env-5       CUGCGGCAUCUUG
Aba-1-1     CAAGGGCGACAAU
env-6       AGUCGGCACGCAA
env-7       NNNNNNNNNNNNN
Lar-1-1     UCACUGCGUCGCC
env-8       .............
env-9       CUCCAGCCCGGUG
Dac-1-1     ACUGAUGUGCUGG
env-10      A............
env-11      UCUGGAGGAAUCC
env-12      UGAGGAAUCCGAG
env-13      GAACCUGUACGCC
env-14      .............
env-15      .............
env-16      CCCUCCGUCUCAC
Cat-1-1     AUGACAAAAUCCU
env-17      AUUAACCAUGUGC
env-18      ACGUCGGGACGAA
env-19      .............
env-20      CGUGCGAUCGACA
env-21      CGCGCGAUCGACA
env-22      CGCACUGGUGACG
env-23      .............
env-24      .............
Mxa-1-1     GGGCCCGCCGGCC
env-25      CCGCAGGCGCAGU
env-26      GCGCAACCCACGC
env-27      CUGCAGCAGGCAA
env-28      UGGAAUAAAUAUC
env-29      UGAUGGCGAAAUU
env-30      .............
env-31      UAAUUAAAUGCUC
env-32      UGCCAAGGUGCCA
env-33      CCGUCGAGCGCCG
env-34      GUCGUGCCGAAGA
Hch-1-1     GAAUCAGCCCGAG
env-35      .............
env-36      UUAUGGUCGAUUA
env-37      AUUACUAGCUGCU
env-38      UAUGCUUGCAUCA
env-39      AUUAAAAAAAAUA
env-40      UGCAUAUACCUCU
Fba-1-1     CCCUAUGAAAUUU
env-41      UUUGAGGUUGAAA
env-42      GUUUCGAAGUUGA
env-43      CCUUAUUUAGGNN
env-44      AUAUGUAAGAUCA
env-45      AUUUAUGUAAGAU
env-46      UCAGAAAAAGGGG
env-47      AAUCGCAAUAAAG
env-48      GUUACAUCAUUAU
env-49      AUUGAGUGUAGGA
env-50      GUUUUAGAUUCGU
env-51      UAAAUAUGGGUUU
env-52      NNNNNNNNNNNNN
env-53      AUUUCAUAGAAAA
env-54      AAUUUCAAAAGAA

            .............
            .............
            .............
            .............
            ooooooooooooo
```

1

```
Rbi-1-1     GGAUUAUGUCCUG
Psp-1-1     GUGAUCUAGAGCA
env-55      UGUAUUUUAUGAC
env-56      UCAAUUCAAAGUG
env-57      .............
env-58      GCUAGAAGCACGU
env-59      AACCCAACCCAGA
env-60      .............
env-61      GUUGAUGAUGCCG
env-62      UGAAUAUCAGUUU
env-63      CGAAUAUCAGUUU
env-64      UAAGAGUGAAUAU
env-65      AGUCAAUACAAAU
env-66      AGUCAAUACCAAU
env-67      AGCGAAUAUCAGU
env-68      AGUGAAUACCAGU
env-69      CUUUACAACCGUG
env-70      AAACUUCAAUGGA
env-71      UCAGCCUUGGUUU
env-72      UCAGUAAACUCAG
env-73      AAACUUCAAUGGA
env-74      CAUUGUAUCAACA
env-75      .............
env-76      .............
env-77      GUUCUCCAACCAU
env-78      ACUGCAACCAUGG
env-79      AUUGCAACCAUGG
env-80      GGUGAGCUAUCAG
env-81      GGUGAGUUAUCAG
env-82      GGUGAGCUAUCAG
env-83      CGCCGCCGGACCC
Cli-1-1     UUAACAGAACAGG
Cli-1-2     GUGUGUUCAGAGC
env-84      GGUGGUAGCCUCG
Nme-3-1     CGAUUCUGGCGCA
Nme-1-1     CCAUCUUGGCGCA
Nme-2-1     CCAUCUUGGCGCA
Ngo-1-1     CCAUCUUGGCGCA
env-85      UAGAUCCCCGCAG
env-86      CGGUCAUCCCACA
env-87      GUGGGGGGAGCUC
env-88      AGCUCUCUCCAAA
env-89      UCUCAGCUCUCGA
env-90      UCUCAGCUCUCGA
env-91      AGAGUAUCUGCUU
env-92      CCAAGCCUAUGUG
env-93      UAAGUCCAGAUUG
env-94      UAAAAGGUAUCGU
env-95      UUUGUAAAAGGAA
env-96      NNNNNNNNNNNNN
env-97      UAUUCUAGACCAC
env-98      UUCGUCUUAGUAA
env-99      AAUCUUAGAUCAU
env-100     AUCCUAGAUCAUA
env-101     AAAUAUUAACGAC
env-102     CUCCAAGGAGCAA
env-103     AAUGGAGAGGAGG
env-104     AUUAACUAUUUCA
env-105     .............
env-106     GACUAUGAUGUUG

            .............
            .............
            .............
            .............
            ooooooooooooo
```

```
env-107     GGCAUAUUUUACU
env-108     GAAGGAAGCAAAA
env-109     UUGGAUGGUGACA
env-110     .............
env-111     GCAAAGUUUCUUG
env-112     GCAAAGUUCCUUG
env-113     AAGUUCCUUGAGU
env-114     UUUGAUUAGCCUU
env-115     GGAGUAAUGCCAA
env-116     GAAAUUACUUGGU
env-117     GAAAUUACUUGGU
env-118     .............
env-119     GAAGCAAAGUUCC
env-120     UGAAAAAGGAAAG
env-121     UUUUGAGGGUGAU
env-122     GAGAGAUUGCCCC
env-123     CUACCAUGAACUC
env-124     GGACCAUAAGAUC
env-125     .............
env-126     UGCUGGUGGAGUA
env-127     UGCUGGUGGAGUA
env-128     AGGAUUAUCGAUA
env-129     .............
env-130     GCC..........
env-131     UCAUGGGCGGUGA
env-132     CAGUUACAGGCUC
env-133     UAGCACUCGAGAC
env-134     CCUUCGACGACCU
env-135     UACUACUCGUUCG
env-136     .............
env-137     UAUGCGAGAGAGA
env-138     .............
env-139     .............
env-140     .............
env-141     GUUUCAAUUGGUG
env-142     .............
env-143     GUUUCAAUCGGUG
env-144     UGUUUCAAUUGGU
env-145     GUUUCAAUUGGUG
env-146     AGAAGAAGGACGA
Nca-1-1     AUCCUGAUUGCCU
env-147     CGCAUGUUUUGGG
Bce-3-1     GGAAGGGCAAGCG
Bce-2-1     GGAAGGGCAAGCG
Bce-1-1     GGAAGGGCAGGCG
Bsp-1-1     GGAAGGGCGAGCA
env-148     GGAAGGGCAAGCG
env-149     CGAUUGACAUCGU
Rsp-1-1     UUCGGUGAGGGCA
Oan-1-1     UGGCGCGCCAUAC
Bab-1-1     CGCGCGCCACACA
Bov-1-1     CGCGCGCCACACA
Bsu-1-1     CGCGCGCCACACA
env-150     UGAUUAUUUUGAA
env-151     ACUUUUAUACAGG
env-152     AGUUACGUGAGUA
env-153     AACCCCAGAGAUG
env-154     GCAUUACGCCAAA
env-155     UGAAAAACCGCCU
env-156     GGAAGGAUGGGUA

            .............
            .............
            .............
            .............
            ooooooooooooo
```

```
env-157    GCUGUUUCACCUG
env-158    UAAGAUUAACGAU
env-159    CAAGAUUAACGAU
env-160    .............
Pna-1-1    CCACUGAUAGCUC
env-161    AGCCCAAGAAUGA
env-162    ACAAAUGGAGAUA
env-163    AUCCAUGCGAAGA
env-164    AGUCACUUGAACU
env-165    .............
env-166    .............
env-167    AAUCACUUGAACU
env-168    AAUCACUUGAACU
env-169    AAUCACUUGAAUU
env-170    AAUCACUUGAAUU
env-171    .............
env-172    .............
env-173    GGUCUCAUUGGAU
env-174    AAAGAGUUUCUCC
env-175    UGGAACGAUAAUC
env-176    UAUCCAAAGCACG
env-177    .............
env-178    UUAAUAAUCAGGG
env-179    .............
env-180    CCUUAUUUUCUUU
env-181    ACUUCCUUCUCCA
env-182    ACCGCAUUGCCUG
env-183    AUAAGGUUUAAAA
env-184    AAAGUCAACAGCG
env-185    UUUUCAUUAAGAU
env-186    CGAUCUGGUUUUA
env-187    .............
env-188    AAGUUUUUUAGUU
env-189    .............
env-190    .............
env-191    AAAUCUUCAUAUA
env-192    .............
env-193    .............
env-194    GCAAGCCAAAGAC
env-195    NNNNNNNNNNNNN
env-196    CCCGAUUGCUGGA
env-197    AUUCCCAUUAGCU
env-198    CCAGAUUGUUGGA
env-199    CCAGAUUGUUGGG
env-200    AACUAGACACAAG
env-201    UGGCAAAUUAGAU
env-202    AAAAAAUACAAAG
env-203    .............
env-204    UAACUAGACACAA
env-205    CAUUUGUUUAUCU
env-206    UUGGGAAGAUCGU
env-207    UGAAGUUUUGCAU
env-208    AGUUAUGGUGGUG
env-209    GUUAUGGCGGUAU
env-210    .............
env-211    CUAGAUUAGACAU
env-212    CCAUUCCGAAUGA
env-213    CUUAAUUAGACAC
env-214    AUAAAUUAGACAC
env-215    AAUGAUUAGGUGU

           .............
           .............
           .............
           .............
           ooooooooooooo
```

```
env-216     AACUGAAUAUGCU
env-217     GGAACAUUUGUUA
env-218     UGGCAAAAGAAAU
env-219     NUUUUUUUUAUCC
env-220     AAUUCCCAUUAGC
env-221     CGCCCUGCUACGU
env-222     .............
env-223     ACCAAUAUUUUCG
env-224     AAAAAAUGGUCUC
env-225     CACCUUAUGCGAU
env-226     UGGUAAAGUGGAG
env-227     UUCCGAUAAGAAA
env-228     CGAUCGAGGCUCA
env-229     UGGGUAUGAGGAC
Lag-1-1     CGGCUGCACCUGC
env-230     .............
env-231     GCGGCACCUACGA
env-232     .............
env-233     GGCAAUGAUUGAC
env-234     UUCCGCCACUACA
env-235     AGCAACUGCAAUG
env-236     .............
Acr-1-1     AGGGGAGGUGAAG
Swi-1-1     CGCCAUACCCCCG
env-237     AUGGCAAUGUCGU
env-238     AGGGCGCCUCUGG
env-239     GGAGAGAGGAGGG
env-240     GUGAAAGAACAAA
env-241     .............
env-242     GAAGAACGCGACA
env-243     GAAGAACGUGACA
env-244     .............
Pna-1-2     AGACCAUCGAGUU
env-245     AUUUGAGUAGAAU
env-246     UCAACCACCUAGC
env-247     AAUAAGCUGAUAG
env-248     .............
env-249     UGGAACAGCUUCA
env-250     UUCAGGAACCUCU
env-251     .............
env-252     CUGGUACGUCUGC
env-253     CUGGUACGUCUGC
Pto-1-1     .............
env-254     UUCUGGUACAUCU
env-255     UUCUGGUACAUCU
env-256     UUCUGGUACAUCU
env-257     UUCUGGUACAUCU
env-258     UUCAGGAACCUCU
env-259     UCAGGAACAUCUG
env-260     UCAGGAACAUCUG
env-261     UGGAACAUCUUCA
env-262     GGAACAUCUUCAG
env-263     .............
env-264     UCUGGGACAUCUG
env-265     GGAACAUCUUCAG
env-266     GGUACAUCAUCGG
env-267     GGUACAUCAUCGG
env-268     GGUACAUCAUCGG
env-269     GGUACAUCAUCGG
env-270     .............

            .............
            .............
            .............
            .............
            °°°°°°°°°°°°°
```

```
env-271  ............
env-272  ............
env-273  UGGAACAGCUUCU
env-274  UGGUACUGCAGCU
env-275  UGGUACUGCAGCU
env-276  CGGUACUGCAGCU
env-277  CUGGAACAGCUUC
env-278  UGGGACAGCUUCC
env-279  UUCUGGCACUUCU
env-280  UUCAGGAACCUCU
env-281  UUCAGGAACCUCU
env-282  UCAGGAACCUCUG
env-283  UUCAGGAACCUCU
env-284  UUCAGGAACCUCU
env-285  UUCAGGAACCUCU
env-286  UUCAGGAACCUCU
env-287  UUCAGGAACCUCU
env-288  UUCAGGAACCUCU
env-289  ............
env-290  UCGGGAACAUCUG
env-291  UCUGGUACGUCUG
env-292  UCUGGUACGUCUG
env-293  UCUGGUACGUCUG
env-294  UCAGGCACAUCAG
env-295  UUCAGGCACAUCU
env-296  UUCAGGCACAUCU
env-297  UUCAGGUACAUCU
env-298  UUCAGGAACUUCU
env-299  GUCUGGUACAUCU
env-300  UCUGGUACGUCUG
env-301  UCUGGUACGUCUG
env-302  UCUGGUACAUCUG
env-303  UCUGGUACAUCUG
env-304  ............
env-305  CUGGUACAUCUGC
env-306  UCUGGUACAUCUG
env-307  UCUGGUACAUCUG
env-308  UCUGGUACAUCUG
env-309  UCUGGUACAUCUG
env-310  ............
env-311  UCUGGUACAUCUG
env-312  UCUGGUACAUCUG
env-313  UCUGGUACAUCUG
env-314  UCUGGUACAUCUG
env-315  ............
env-316  UCUGGUACGUCUG
env-317  UCUGGUACGUCUG
env-318  CUUCUGGUACGUC
env-319  UCUGGUACAUCUG
env-320  ............
env-321  UUCUGGUACAUCU
env-322  UUCUGGUACAUCU
env-323  UUCUGGUACAUCU
env-324  UCAGGAACAUCUG
env-325  UCAGGAACAUCUG
env-326  UCAGGAACAUCUG
env-327  UCAGGUACAUCUG
env-328  AUCAGGAACAUCU
env-329  UCAGGAACAUCUG
env-330  ............

         ............
         ............
         ............
         ............
         ooooooooooooo
```

```
env-331    UCAGGAACGUCUG
env-332    UCAGGAACAUCUG
env-333    .............
env-334    UCAGGAACAUCUG
env-335    AUCAGGAACAUCU
env-336    UCACGAACAUCUG
env-337    CUGGAACAUCUGC
env-338    UCAGGAACAUCAG
env-339    AUCAGGAACAUCU
env-340    UCAGGAACAUCUG
env-341    AUCAGGAACAUCU
env-342    CAGGAACAUCUGC
env-343    AUCAGGAACAUCU
env-345    AUCAGGAACAUCU
env-346    .............
env-347    AUCAGAACCUCUG
env-348    AUCAGGAACAUCU
env-349    AUCAGGAACAUCU
env-350    UCAGGUACAUCUG
env-351    AUCAGGAACAUCU
env-352    AUGUAGAUAAUAU
env-353    .............
env-354    AAAAACAAUAAAA
env-355    ACUUAAAAAAACA
env-356    AGAAAACAAUAAA
env-357    AAAAACAAUAAAA
env-358    AAAAACAAUAAAA
env-359    AGAGAAAACAAUA
env-360    AGAGAAAACAAUA
env-361    AAGUAAAACAAUU
env-362    AAAAACAAUAAAA
env-363    AAAAACAAUAAAA
env-364    AGCACUUAAAAAA
env-365    CUUAAAUAAAACA
env-366    AACAAAACUAUCA
env-367    AAAUAACUAUAAA
env-368    CUUAACAAAACCA
env-369    GUACUUGUAGGUG
env-370    AACAAAACUAUAA
env-371    AACAAAACUAUCA
env-372    AACAAAACUAUCA
env-373    .............
env-374    AAAAAUAACUAUC
env-375    AACAAAACUAUCA
env-376    CUUAACAAAACUA
env-377    .............
env-378    ACAAAACUAUCAA
env-379    UAACAAAACUAUA
env-380    UUUGAUAUGAAGC
env-381    AAAAAUAACUAUC
env-382    .............
env-383    AAAAAUAACUAUC
env-384    AAAAUAACUAUCA
env-385    AAAAAUAACUAUC
env-386    AAAAAUAACUAUC
env-387    AAAAAUAACUAUC
env-388    AAAAAUAACUAUC
env-389    AAAAAUAACUAUU
env-390    AACUAUAAAAACU
env-391    ACUAUAAAGACUA

           .............
           .............
           .............
           .............
           °°°°°°°°°°°°
```

```
env-392    ..............
env-393    ACUAUAAAAACUA
env-394    ACAAUUAAAACAA
env-395    CUAAAUAAAACAA
env-396    UUGAACAAAACUA
env-397    UAAAUAAGACUAU
env-398    CUUAAUAAAACUA
env-399    CUUAAUAAAACUA
env-400    UUAAAUAAAACAA
env-401    UUAAAUAAAACAA
env-402    UAAAUAAAACAAU
env-403    UAAAUAAAACAAU
env-404    UUAAAUAAAACAA
env-405    UAAAUAAAACAAU
env-406    ACAAAACUAUAAA
env-407    ..............
env-408    ..............
env-409    ACAAAACUAUAAA
env-410    ACAAAACUAUCAA
env-411    ACAAAACUAUCAA
env-412    ..............
env-413    AGUAAAACUAUUA
env-414    AGUAAAACUAUUA
env-415    AGAUAAAACAAUC
env-416    GAGUAAAACAAUC
env-417    UGAGUAAAACAAU
Nar-1-1    CGUCAGCAAGAAC
env-418    CAGAUGUUGACUC
env-419    UGAAAAAUCAGUA
env-420    GAAAAAUCAGUUA
env-421    ..............
env-422    UCUUAUUUUAUAA
env-423    CGUAAUAUCGUCA
Pna-1-3    AAGGCAGGAGCCU
env-424    ..............
env-425    UCACUGGUGCCGC
env-426    UCACUGGUGCCGC
env-427    ACAGCGUCGUCGG
env-428    GGUGGGAAGGAAU
env-429    ..............
env-430    ..............
env-431    GGUUGGUGGGAAG
env-432    ..............
env-433    ..............
env-434    CGCACCGACACCA
env-435    GAUGCUGUGCACG
env-436    GACGCUCGACGCN
env-437    CCACUUCAGUGGC
env-438    AGCACUUCACUGG
           .............
           .............
           .............
           .............
           °°°°°°°°°°°°°
```

86 SAM-II long loops

86.1 Taxa

The taxonomy of each organism containing a putative SAM-II long loops RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual SAM-II long loops RNA in Sections 86.2 and 86.4.

abbrev. of hits	taxonomy of species
Bca-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides caccae* ATCC 43185
Bdo-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides dorei* DSM 17855
Bfr-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides fragilis* NCTC 9343
Bfr-2-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides fragilis* YCH46
Bov-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides ovatus* ATCC 8483
Bth-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides thetaiotaomicron* VPI-5482
Bun-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides uniformis* ATCC 8492
Bvu-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae *Bacteroides vulgatus* ATCC 8482
Pdi-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Porphyromonadaceae *Parabacteroides distasonis* ATCC 8503
Pjo-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Porphyromonadaceae *Parabacteroides johnsonii* DSM 18315
Pme-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Porphyromonadaceae *Parabacteroides merdae* ATCC 43184
Pgi-1-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Porphyromonadaceae *Porphyromonas gingivalis* ATCC 33277
Pgi-2-1	Bacteria Bacteroidetes Bacteroidia Bacteroidales Porphyromonadaceae *Porphyromonas gingivalis* W83
ueu-1-1	Bacteria Bacteroidetes environmental samples unidentified eubacterium SCB49
Fba-1-1	Bacteria Bacteroidetes *Flavobacteria bacterium* BBFL7
Fba-2-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales *Flavobacteria bacterium* BAL38
Cat-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Croceibacter atlanticus* HTCC2559
Ddo-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Dokdonia donghaensis* MED134
Fjo-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Flavobacterium johnsoniae* UW101
Fps-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Flavobacterium psychrophilum* JIP02/86
Gfo-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Gramella forsetii* KT0803
Lbl-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Leeuwenhoekiella blandensis* MED217
Pir-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Polaribacter irgensii* 23-P
Psp-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Polaribacter sp.* MED152
Pto-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Psychroflexus torquis* ATCC 700755
Rbi-1-1	Bacteria Bacteroidetes Flavobacteria Flavobacteriales Flavobacteriaceae *Robiginitalea biformata* HTCC2501
Fba-3-1	Bacteria Bacteroidetes Flavobacteria *Flavobacteriales bacterium* HTCC2170
Psp-2-1	Bacteria Bacteroidetes Sphingobacteria Sphingobacteriales Sphingobacteriaceae *Pedobacter sp.* BAL39
Pma-1-1	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* AS9601
Pma-2-1	Bacteria Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9301
env-1 to env-132	environmental samples

86.2 Gene contexts

Each SAM-II long loops RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one SAM-II long loops RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 86.3. The accession number of the sequence containing each SAM-II long loops RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 86.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	JCVI_SCAF_1096627609663	-	1072	1013	**RNA**→ MetC (COG0626)? (GOS3285)→
env-2	JCVI_SCAF_1096626175348	-	843	784	**RNA**→ MetC (COG0626)? (GOS3285)→
env-3	JCVI_SCAF_1101668208614	-	470	411	**RNA**→ ? (GOS3285)→
env-4	JCVI_SCAF_1096627323424	-	1278	1219	**RNA**→ MetC (COG0626)? (GOS3285)→

env-5	JCVI_SCAF_1096627004500	-	581	522	**RNA**→ MetC (COG0626)? (GOS3285)→
env-6	JCVI_SCAF_1096628013975	-	274	215	**RNA**→ ? (GOS3285)→
env-7	JCVI_SCAF_1101667225678	+	427	486	**RNA**→ ? (GOS3285)→
Pma-2-1	NC_009091.1	-	375563	375504	**RNA**→ MetC (COG0626)→ MetC (COG0626)→ CysK (COG0031)→
env-8	JCVI_SCAF_1096627321893	-	1008	949	**RNA**→ MetC (COG0626)? (GOS3285)→
env-9	JCVI_SCAF_1096628398720	-	1316	1257	**RNA**→ MetC (COG0626)? (GOS3285)→ ? (GOS3285)→
env-10	JCVI_SCAF_1096627720042	-	587	528	**RNA**→ MetC (COG0626)? (GOS3285)→
env-11	JCVI_SCAF_1096628352492	-	1249	1190	**RNA**→ ? (GOS3285)→ ? (GOS3285)→
env-12	JCVI_SCAF_1096628229633	+	1734	1793	**RNA**→
Pma-1-1	NC_008816.1	-	401135	401076	**RNA**→ MetC (COG0626)→ MetC (COG0626)→ CysK (COG0031)→
env-13	JCVI_SCAF_1096627713106	-	683	624	**RNA**→ MetC (COG0626)? (GOS3285)→
env-14	JCVI_SCAF_1096627709846	+	725	784	**RNA**→ MetC (COG0626)? (GOS3285)→
env-15	JCVI_SCAF_1101667106167	-	237	178	**RNA**→ ? (GOS3285)→
env-16	JCVI_SCAF_1101667534115	-	613	554	**RNA**→ MetC (COG0626)? (GOS3285)→
env-17	JCVI_SCAF_1096626983612	+	1440	1499	**RNA**→ MetC (COG0626)? (GOS3285)→
env-18	JCVI_SCAF_1101667128252	-	233	174	**RNA**→ ? (GOS3285)→
env-19	JCVI_SCAF_1096627714531	-	1321	1262	**RNA**→ ? (GOS3285)→ ? (GOS3285)→ ? (GOS3285)→
env-20	JCVI_SCAF_1096628356103	-	1081	1022	**RNA**→ MetC (COG0626)? (GOS3285)→
env-21	JCVI_SCAF_1096627305652	-	2394	2335	**RNA**→ MetC (COG0626)? (GOS3285)→ MetC (COG0626)? (GOS3285)→
env-22	JCVI_SCAF_1096626921896	+	1713	1772	**RNA**→ ? (GOS3285)→
env-23	JCVI_SCAF_1096627396212	-	1087	1028	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-24	JCVI_SCAF_1096627292244	+	1095	1156	**RNA**→ MET17 (COG2873)? (GOS3285)→ ? (GOS10901)→
env-25	JCVI_SCAF_1101668527332	-	910	847	**RNA**→ MET17 (COG2873)? (GOS3285)→
Psp-2-1	NZ_ABCM01000030.1	+	26135	26215	**RNA**→ MetK (COG0192)→
Gfo-1-1	NC_008571.1	-	366446	366372	**RNA**→ MetK (COG0192)→
env-26	JCVI_SCAF_1096626866251	-	671	611	**RNA**→ MetK (COG0192)? (GOS4376)→
env-27	JCVI_SCAF_1096626854522	-	3178	3120	**RNA**→ MetK (COG0192)? (GOS4376)→
Fba-1-1	NZ_AAPD01000001.1	-	935844	935772	**RNA**→ MetK (COG0192)→
env-28	JCVI_SCAF_1096627402240	-	283	225	**RNA**→ ? (GOS4376)→
env-29	JCVI_SCAF_1096626915432	-	1598	1540	**RNA**→ MetK (COG0192)? (GOS4376)→
ueu-1-1	NZ_ABCO01000001.1	-	556577	556445	**RNA**→ MetK (COG0192)→
Pgi-1-1	NC_010729.1	+	2055065	2055124	**RNA**→ MetK (COG0192)→
Pgi-2-1	NC_002950.2	+	1994578	1994637	**RNA**→ MetK (COG0192)→
Pdi-1-1	NC_009615.1	-	2817240	2817181	**RNA**→ MetK (COG0192)→
Pjo-1-1	NZ_ABYH01000277.1	-	7168	7109	**RNA**→
Pme-1-1	NZ_AAXE02000111.1	+	39496	39555	**RNA**→ ←hypo
env-30	JCVI_SCAF_1096627401769	-	297	240	**RNA**→ ? (GOS4376)→
env-31	JCVI_SCAF_1096627396805	+	383	441	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→ MetK (COG0192)? (GOS4376)→
env-32	JCVI_SCAF_1101668574804	-	1015	958	**RNA**→ MetK (COG0192)? (GOS4376)→
Ddo-1-1	NZ_AAMZ01000003.1	-	48269	48210	**RNA**→ MetK (COG0192)→
env-33	JCVI_SCAF_1096627474725	+	1073	1131	**RNA**→ MetK (COG0192)? (GOS4376)→
env-34	JCVI_SCAF_1101668602500	-	997	939	**RNA**→ MetK (COG0192)? (GOS4376)→
env-35	JCVI_SCAF_1096627366491	+	2684	2742	**RNA**→ MetK (COG0192)? (GOS4376)→
env-36	JCVI_SCAF_1096627277433	+	380	438	**RNA**→ MetK (COG0192)? (GOS4376)→
Rbi-1-1	NZ_AAOI01000005.1	-	209548	209469	**RNA**→ MetK (COG0192)→
Lbl-1-1	NZ_AANC01000003.1	-	237275	237201	**RNA**→ MetK (COG0192)→
Fba-3-1	NZ_AAOC01000005.1	-	243743	243673	**RNA**→ MetK (COG0192)→
env-37	JCVI_SCAF_1096626346811	-	840	782	**RNA**→ MetK (COG0192)? (GOS4376)→
Cat-1-1	NZ_AAMP01000001.1	-	1344288	1344189	**RNA**→ MetK (COG0192)→
Pto-1-1	NZ_AAPR01000008.1	+	157663	157721	**RNA**→ MetK (COG0192)→
env-38	JCVI_SCAF_1096626870652	+	2307	2366	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-39	JCVI_SCAF_1096627277718	+	7951	8010	**RNA**→ MetK (COG0192)? (GOS4376)→ MET17 (COG2873)? (GOS3285)→ MET2 (COG2021)? (GOS3619)→ COG1765 (COG1765)? (GOS10947)→ ? (GOS3285)→ ? (GOS6193)→

env-40	JCVI_SCAF_1096628033852	+	1488	1547	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-41	JCVI_SCAF_1096627099057	-	1566	1507	**RNA**→ MetK (COG0192)? (GOS4376)→ ? (GOS3285)→
env-42	JCVI_SCAF_1096628269516	+	873	932	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-43	JCVI_SCAF_1096628178924	-	1342	1283	**RNA**→ ? (GOS3285)→
env-44	JCVI_SCAF_1096627066950	+	506	565	**RNA**→ MET17 (COG2873)? (GOS3285)→ MET2 (COG2021)? (GOS3619)→
env-45	JCVI_SCAF_1096627402583	-	1217	1158	**RNA**→ ? (GOS3285)→ MET17 (COG2873)? (GOS3285)→
env-46	JCVI_SCAF_1101668441589	+	717	776	**RNA**→ ? (GOS3285)→
env-47	JCVI_SCAF_1096627411931	-	1684	1625	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-48	JCVI_SCAF_1096626857991	-	795	736	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-49	JCVI_SCAF_1101668376995	-	754	695	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-50	JCVI_SCAF_1101668450894	-	630	571	**RNA**→ Cys_Met_Meta_PP (pfam01053)? (GOS3285)→
env-51	JCVI_SCAF_1101668416052	+	684	743	**RNA**→ ? (GOS3285)→
env-52	JCVI_SCAF_1096626330588	-	92	33	**RNA**→
env-53	JCVI_SCAF_1101668552424	+	891	950	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-54	JCVI_SCAF_1101668299256	+	399	458	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-55	JCVI_SCAF_1096627308126	+	588	647	**RNA**→ MET17 (COG2873)? (GOS3285)→ ? (GOS3619)→
env-56	JCVI_SCAF_1101667432870	-	730	671	**RNA**→ MET17 (COG2873)? (GOS3285)→
env-57	JCVI_SCAF_1101668049517	-	491	431	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-58	JCVI_SCAF_1101668043036	-	470	410	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-59	JCVI_SCAF_1096627438223	-	1727	1669	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→ S-AdoMet_synt_M (pfam02772)? (GOS4376)→
env-60	JCVI_SCAF_1101668370776	+	759	817	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-61	JCVI_SCAF_1096627980867	+	883	941	**RNA**→ MetK (COG0192)? (GOS4376)→
env-62	JCVI_SCAF_1096626009591	+	142	202	**RNA**→ MetK (COG0192)? (GOS4376)→
env-63	JCVI_SCAF_1096626616824	-	1414	1354	**RNA**→ MetK (COG0192)? (GOS4376)→
env-64	JCVI_SCAF_1096626856299	-	593	534	**RNA**→ MetK (COG0192)? (GOS4376)→
env-65	JCVI_SCAF_1096628389487	+	303	365	**RNA**→ MetK (COG0192)? (GOS4376)→
env-66	JCVI_SCAF_1101668210493	-	1259	1199	**RNA**→ MetK (COG0192)? (GOS4376)→
env-67	JCVI_SCAF_1101668750758	-	647	587	**RNA**→ MetK (COG0192)? (GOS4376)→
env-68	JCVI_SCAF_1096627860676	+	1223	1283	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-69	JCVI_SCAF_1101667335610	-	329	269	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-70	JCVI_SCAF_1101668586181	-	696	636	**RNA**→ MetK (COG0192)? (GOS4376)→
env-71	JCVI_SCAF_1101667257829	-	878	818	**RNA**→ MetK (COG0192)? (GOS4376)→
env-72	JCVI_SCAF_1096626894984	-	281	221	**RNA**→
env-73	JCVI_SCAF_1096627098520	-	992	932	**RNA**→ MetK (COG0192)? (GOS4376)→
env-74	JCVI_SCAF_1096626891259	+	683	743	**RNA**→ ←? (GOS11251012)
env-75	JCVI_SCAF_1096627317691	-	2946	2886	**RNA**→ MetK (COG0192)? (GOS4376)→ MET17 (COG2873)? (GOS3285)→ ? (GOS3619)→
env-76	JCVI_SCAF_1096627001039	-	993	933	**RNA**→ MetK (COG0192)? (GOS4376)→
env-77	JCVI_SCAF_1101667432231	+	159	218	**RNA**→ MetK (COG0192)? (GOS4376)→
env-78	JCVI_SCAF_1096626910058	+	602	660	**RNA**→ MetK (COG0192)? (GOS4376)→ MetK (COG0192)? (GOS4376)→ ? (GOS3285)→
env-79	JCVI_SCAF_1101668505994	-	1374	1315	**RNA**→ MetK (COG0192)? (GOS4376)→
env-80	JCVI_SCAF_1101668435854	-	749	690	**RNA**→ MetK (COG0192)? (GOS4376)→
env-81	JCVI_SCAF_1096627219953	+	38	97	**RNA**→ MetK (COG0192)? (GOS4376)→
env-82	JCVI_SCAF_1096628389438	+	499	558	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-83	JCVI_SCAF_1096627186799	-	304	245	**RNA**→
env-84	JCVI_SCAF_1096627155384	-	1646	1587	**RNA**→ MetK (COG0192)? (GOS4376)→ ? (GOS3285)→
env-85	JCVI_SCAF_1096627988331	+	276	335	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-86	JCVI_SCAF_1096626991476	+	877	936	**RNA**→ MetK (COG0192)? (GOS4376)→ MET17 (COG2873)? (GOS3285)→
env-87	JCVI_SCAF_1101667341953	-	688	629	**RNA**→ MetK (COG0192)? (GOS4376)→
env-88	JCVI_SCAF_1101668423433	-	79	20	**RNA**→
env-89	JCVI_SCAF_1096626894819	-	274	215	**RNA**→
env-90	JCVI_SCAF_1101667116281	+	363	422	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-91	JCVI_SCAF_1096627787351	-	852	793	**RNA**→ S-AdoMet_synt_C (pfam02773)? (GOS4376)→ ? (GOS3285)→

env-92	JCVI_SCAF_1101668762251	-	309	250	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-93	JCVI_SCAF_1101668742341	+	1181	1240	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-94	JCVI_SCAF_1096627611673	-	1578	1519	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→ MetK (COG0192)? (GOS4376)→ ? (GOS3285)→
env-95	JCVI_SCAF_1101668177421	+	230	289	**RNA**→ MetK (COG0192)? (GOS4376)→
env-96	JCVI_SCAF_1096626958110	+	839	898	**RNA**→ MetK (COG0192)? (GOS4376)→
env-97	JCVI_SCAF_1101668185252	+	249	307	**RNA**→ MetK (COG0192)? (GOS4376)→
env-98	JCVI_SCAF_1096627103077	+	1161	1220	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-99	JCVI_SCAF_1096627505900	+	424	483	**RNA**→ MetK (COG0192)? (GOS4376)→
env-100	JCVI_SCAF_1101668759094	+	60	119	**RNA**→ MetK (COG0192)? (GOS4376)→
env-101	JCVI_SCAF_1096627973299	+	603	662	**RNA**→ MetK (COG0192)? (GOS4376)→
env-102	JCVI_SCAF_1096628086426	-	1274	1215	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→ Chorismate_mut (pfam01817)? (GOS69)→ COG0436 (COG0436)? (GOS113)→
env-103	JCVI_SCAF_1096627265055	+	143	202	**RNA**→ MetK (COG0192)? (GOS4376)→ S-AdoMet_synt_C (pfam02773)? (GOS4376)→
env-104	JCVI_SCAF_1096627546860	+	329	388	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-105	JCVI_SCAF_1101668520745	+	744	803	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-106	JCVI_SCAF_1096627902800	+	118	177	**RNA**→ MetK (COG0192)? (GOS4376)→ MetK (COG0192)? (GOS4376)→
env-107	JCVI_SCAF_1096626871642	-	1329	1270	**RNA**→ MetK (COG0192)? (GOS4376)→
Pir-1-1	NZ_AAOG01000001.1	+	143195	143292	**RNA**→ MetK (COG0192)→
env-108	JCVI_SCAF_1101668180177	-	421	361	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-109	JCVI_SCAF_1096627807455	-	878	819	**RNA**→ MetK (COG0192)? (GOS4376)→
env-110	JCVI_SCAF_1096627512370	-	1190	1131	**RNA**→ MetK (COG0192)? (GOS4376)→
Fjo-1-1	NC_009441.1	-	1712338	1712232	**RNA**→ MetK (COG0192)→
env-111	JCVI_SCAF_1096627426589	+	376	466	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→ MetK (COG0192)? (GOS4376)→
Fps-1-1	NC_009613.1	-	321748	321688	**RNA**→ MetK (COG0192)→
env-112	JCVI_SCAF_1096627455278	+	1053	1112	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
Fba-2-1	NZ_AAXX01000001.1	+	754085	754145	**RNA**→ MetK (COG0192)→
Psp-1-1	NZ_AANA01000001.1	+	1408910	1408969	**RNA**→ MetK (COG0192)→
env-113	JCVI_SCAF_1096624997329	+	37	96	**RNA**→ MetK (COG0192)? (GOS4376)→
env-114	JCVI_SCAF_1096626857523	+	2002	2058	**RNA**→ MetK (COG0192)? (GOS4376)→
env-115	JCVI_SCAF_1096628217660	+	958	1015	**RNA**→ MetK (COG0192)? (GOS4376)→
Bvu-1-1	NC_009614.1	+	1599296	1599355	**RNA**→ MetK (COG0192)→
env-116	BAAU01000101.1	+	970	1029	**RNA**→ MetK (COG0192)→
env-117	BAAV01002129.1	+	648	707	**RNA**→ MetK (COG0192)→
env-118	BAAY01000229.1	-	4860	4801	**RNA**→ MetK (COG0192)→
env-119	BABB01000128.1	-	2054	1995	**RNA**→ MetK (COG0192)→
Bdo-1-1	NZ_ABWZ01000039.1	+	10487	10546	**RNA**→
env-120	BABA01032719.1	-	684	625	**RNA**→ MetK (COG0192)→
env-121	BABD01010212.1	-	361	302	**RNA**→ S-AdoMet_synt_N (pfam00438)→
env-122	BABD01004898.1	-	734	675	**RNA**→ S-AdoMet_synt_N (pfam00438)S-AdoMet_synt_M (pfam02772)→
env-123	BAAU01003527.1	-	70	11	**RNA**→
env-124	BAAV01005021.1	+	856	915	**RNA**→ S-AdoMet_synt_N (pfam00438)→
env-125	BABD01000682.1	-	2075	2016	**RNA**→ MetK (COG0192)→ hypo→
env-126	BAAX01034998.1	-	534	475	**RNA**→ S-AdoMet_synt_N (pfam00438)→
env-127	BABA01000840.1	+	862	921	**RNA**→ MetK (COG0192)→
env-128	BAAY01003005.1	+	1818	1877	**RNA**→
Bun-1-1	NZ_AAYH02000034.1	+	58632	58691	**RNA**→ MetK (COG0192)→
Bfr-1-1	NC_003228.3	+	69593	69652	**RNA**→ MetK (COG0192)→
Bfr-2-1	NC_006347.1	+	64060	64119	**RNA**→ MetK (COG0192)→
env-129	BABA01005785.1	+	593	652	**RNA**→ MetK (COG0192)→
env-130	BABG01001786.1	+	1819	1878	**RNA**→ MetK (COG0192)→
Bca-1-1	NZ_AAVM02000002.1	+	266706	266765	**RNA**→ MetK (COG0192)→
Bth-1-1	NC_004663.1	+	4108976	4109035	**RNA**→ MetK (COG0192)→

Bov-1-1	NZ_AAXF02000049.1	+	418563	418622	**RNA**→ MetK (COG0192)→
env-131	JCVI_SCAF_1096627783900	-	312	255	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-132	JCVI_SCAF_1096627836680	-	1384	1327	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→ S-AdoMet_synt_C (pfam02773)? (GOS4376)→

86.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 86.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one SAM-II long loops RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

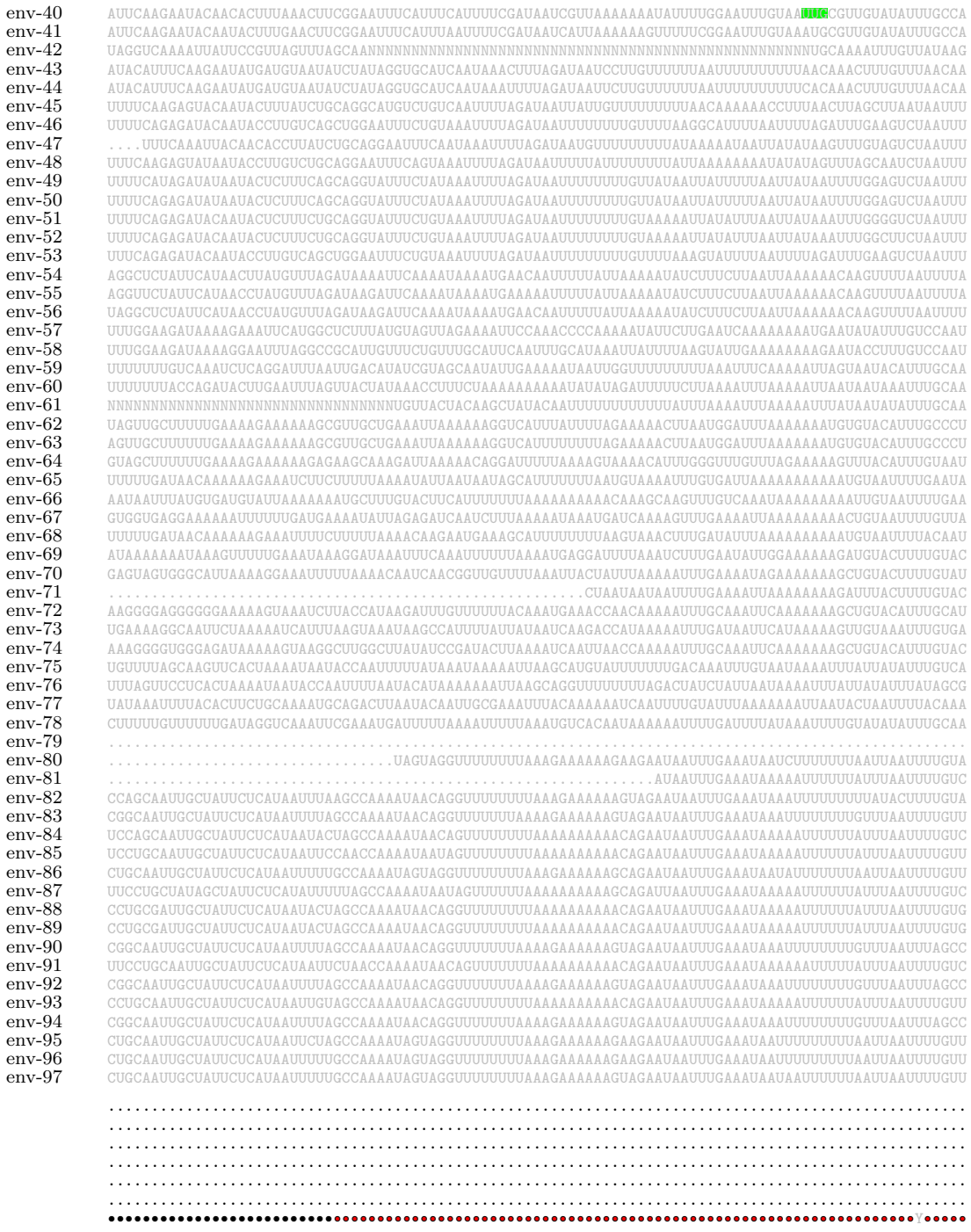
COG0031 (2) Cysteine synthase [Amino acid transport and metabolism]
COG0192 (84) S-adenosylmethionine synthetase [Coenzyme metabolism]
COG0436 (1) Aspartate/tyrosine/aromatic aminotransferase [Amino acid transport and metabolism]
COG0626 (18) Cystathionine beta-lyases/cystathionine gamma-synthases [Amino acid transport and metabolism]
COG1765 (1) Predicted redox protein, regulator of disulfide bond formation [Posttranslational modification, protein turnover, chaperones]
COG2021 (2) Homoserine acetyltransferase [Amino acid transport and metabolism]
COG2873 (16) O-acetylhomoserine sulfhydrylase [Amino acid transport and metabolism]
GOS10901 (1) ?
GOS10947 (1) ?
GOS11251012 (1) ?
GOS113 (1) ?
GOS3285 (53) ?
GOS3619 (4) ?
GOS4376 (80) ?
GOS6193 (1) ?
GOS69 (1) ?
pfam00438 (28) S-adenosylmethionine synthetase, N-terminal domain.
pfam01053 (1) Cys/Met metabolism PLP-dependent enzyme.
pfam01817 (1) Chorismate mutase.
pfam02772 (2) S-adenosylmethionine synthetase, central domain.
pfam02773 (3) S-adenosylmethionine synthetase, C-terminal domain.

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: env-116, env-117, env-119, env-129

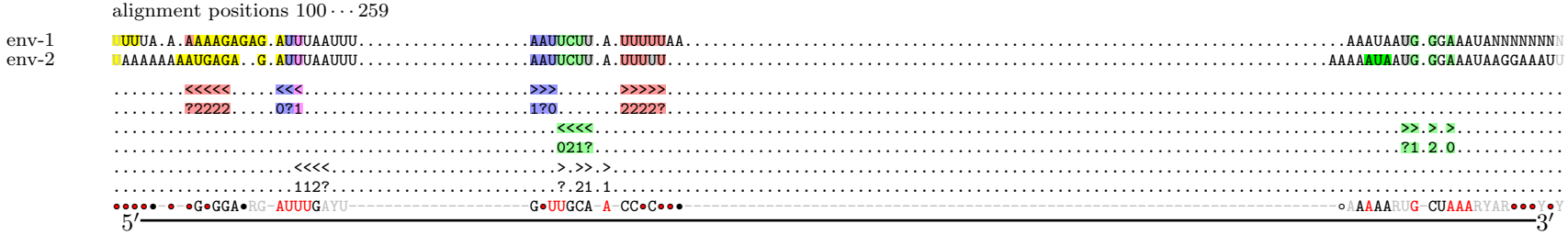
86.4 Multiple-sequence alignment

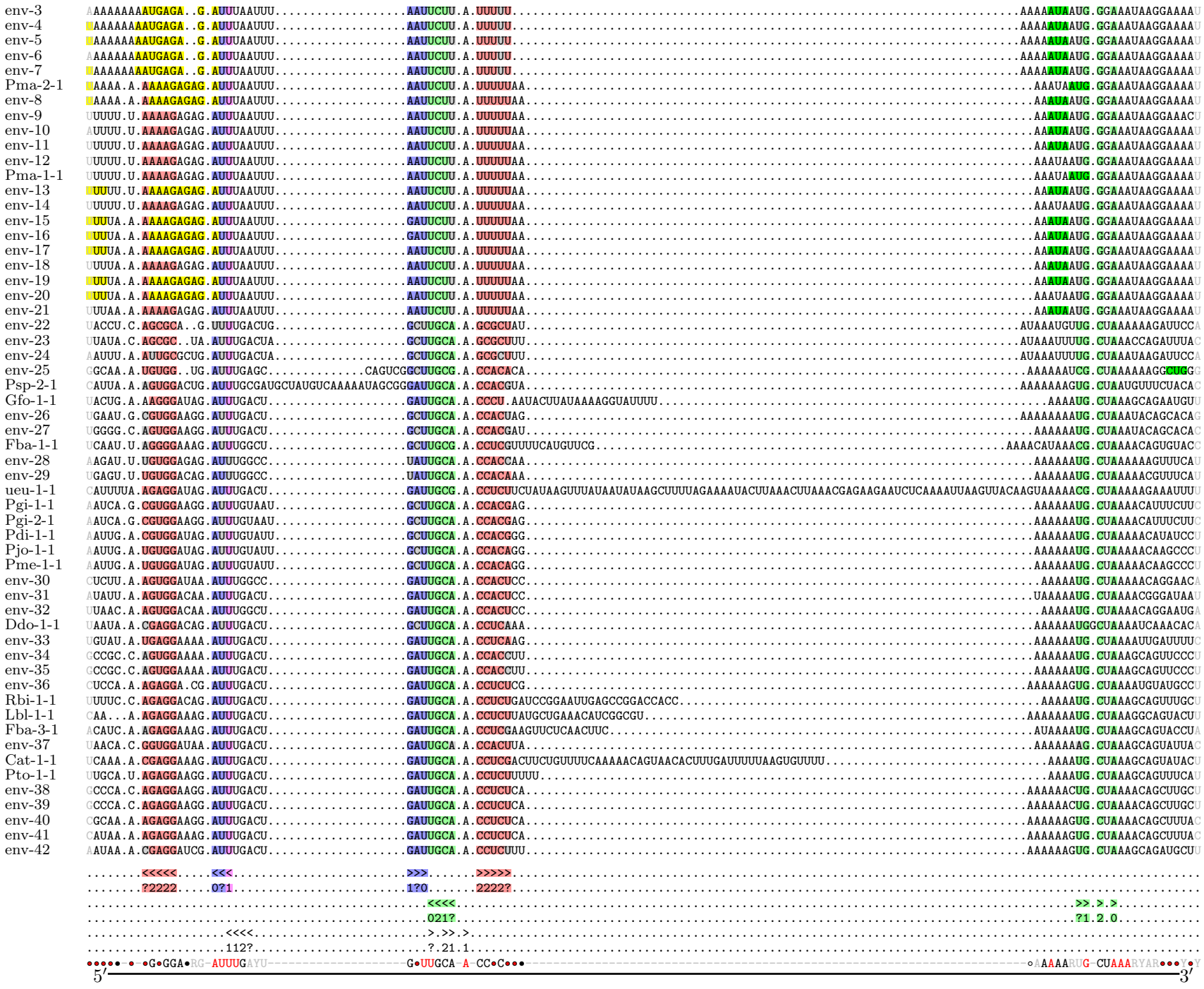
Each SAM-II long loops RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 86.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The SAM-II long loops RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

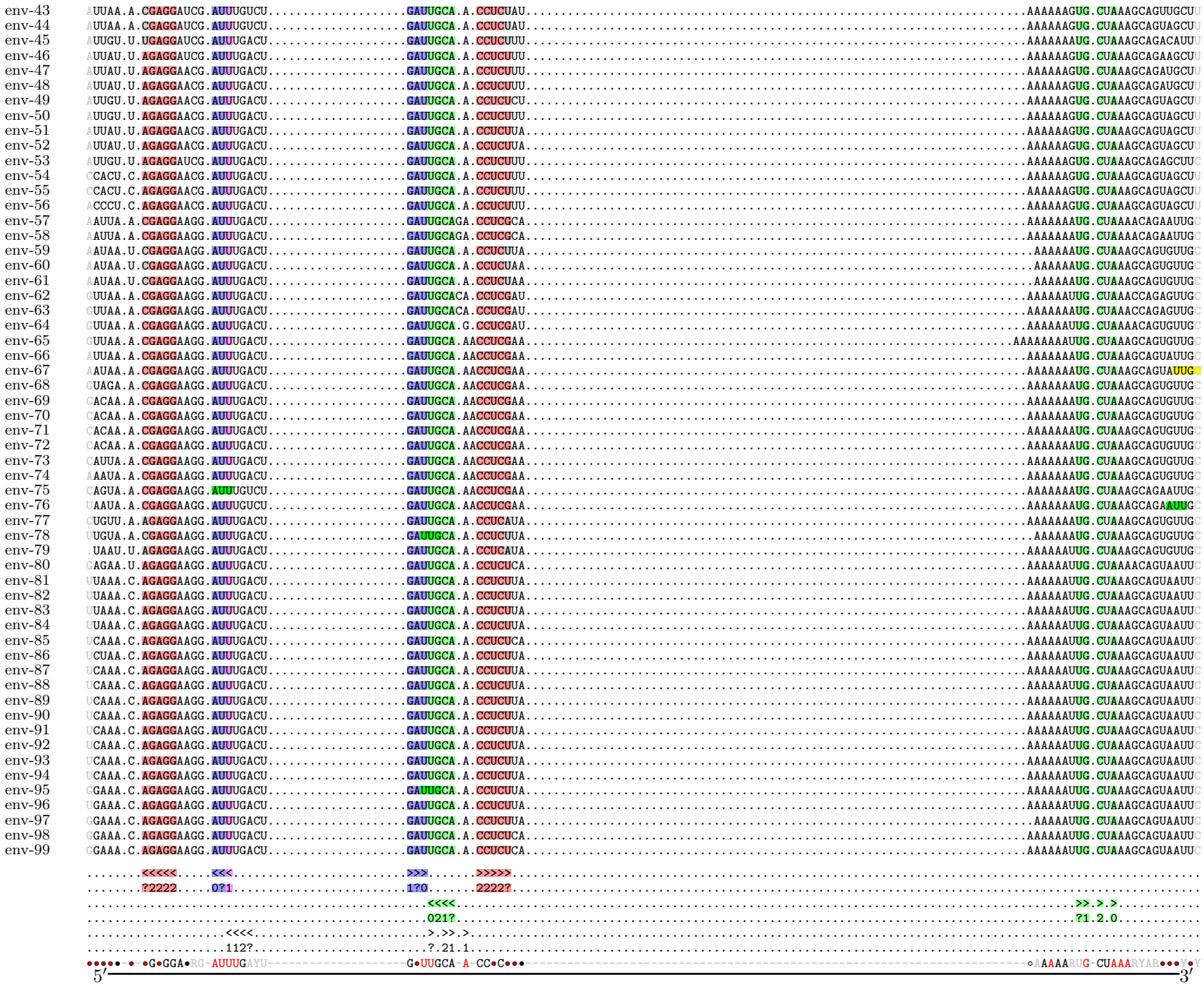
alignment positions 1 · · · 99
env-1
env-2
env-3
env-4
env-5
env-6
env-7
Pma-2-1
env-8
env-9
env-10
env-11
env-12
Pma-1-1
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
Psp-2-1
Gfo-1-1
env-26
env-27
Fba-1-1
env-28
env-29
ueu-1-1
Pgi-1-1
Pgi-2-1
Pdi-1-1
Pjo-1-1
Pme-1-1
env-30
env-31
env-32
Ddo-1-1
env-33
env-34
env-35
env-36
Rbi-1-1
Lbl-1-1
Fba-3-1
env-37
Cat-1-1
Pto-1-1
env-38
env-39

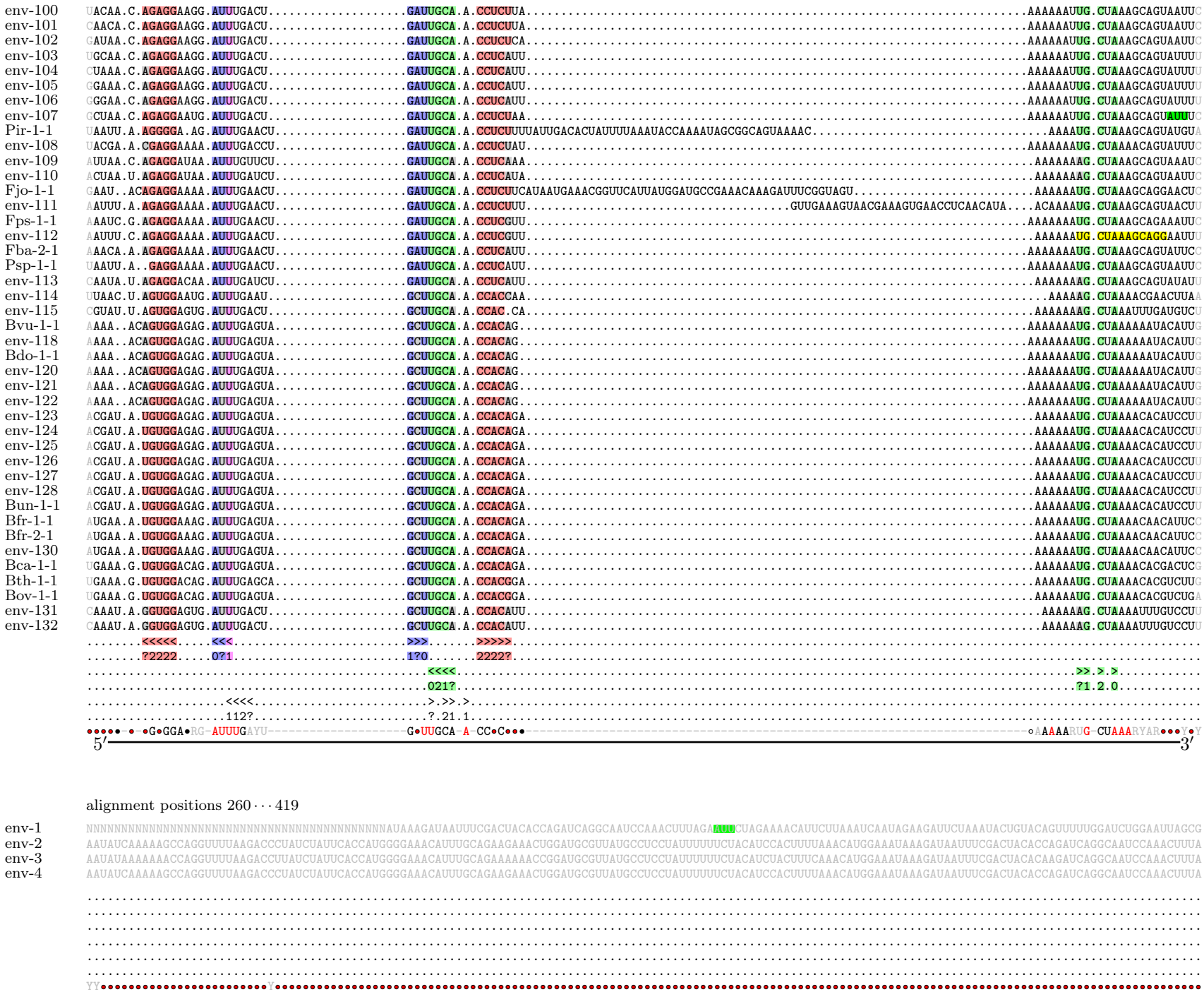

env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97

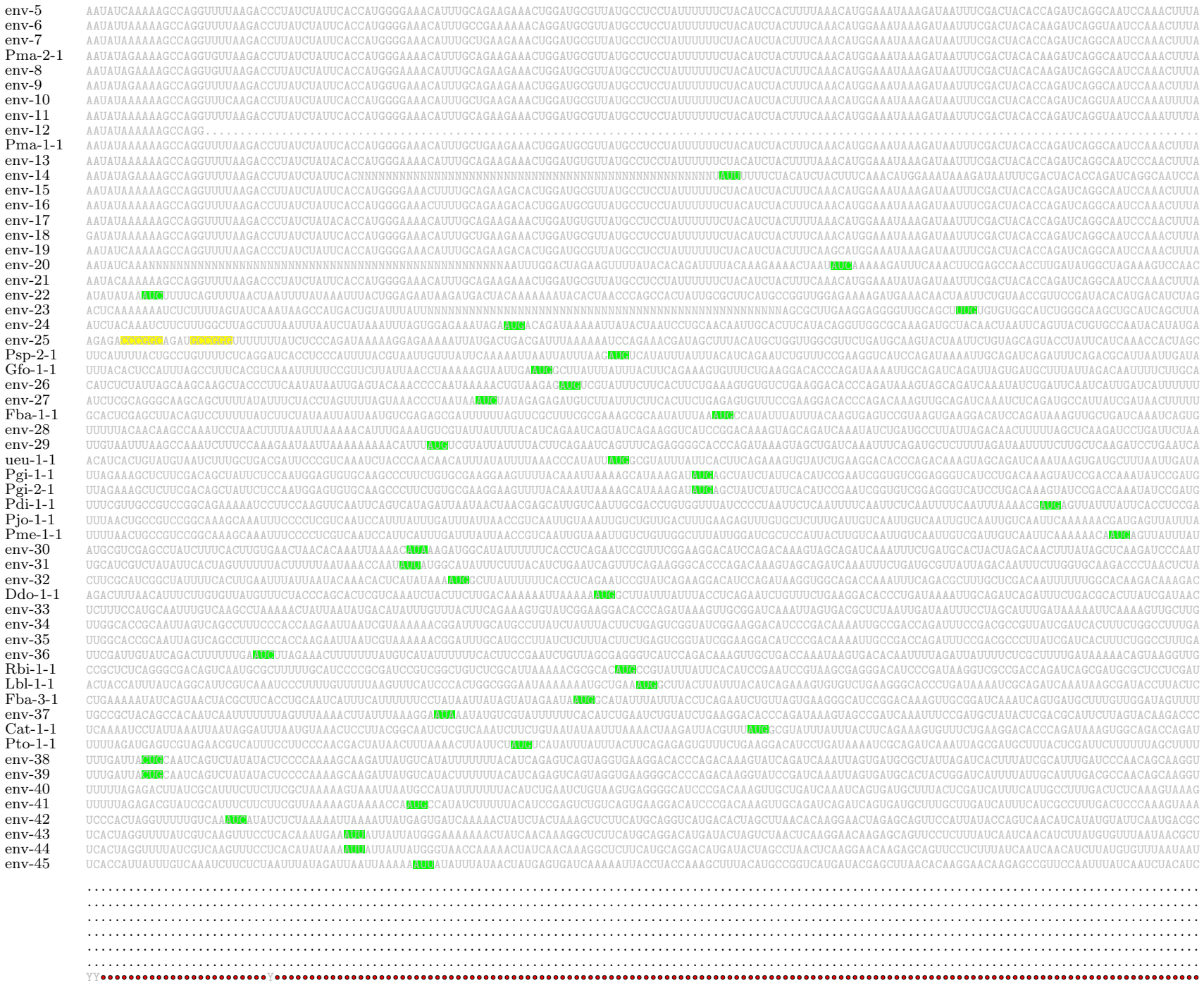
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
Pir-1-1
env-108
env-109
env-110
Fjo-1-1
env-111
Fps-1-1
env-112
Fba-2-1
Psp-1-1
env-113
env-114
env-115
Bvu-1-1
env-118
Bdo-1-1
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
Bun-1-1
Bfr-1-1
Bfr-2-1
env-130
Bca-1-1
Bth-1-1
Bov-1-1
env-131
env-132

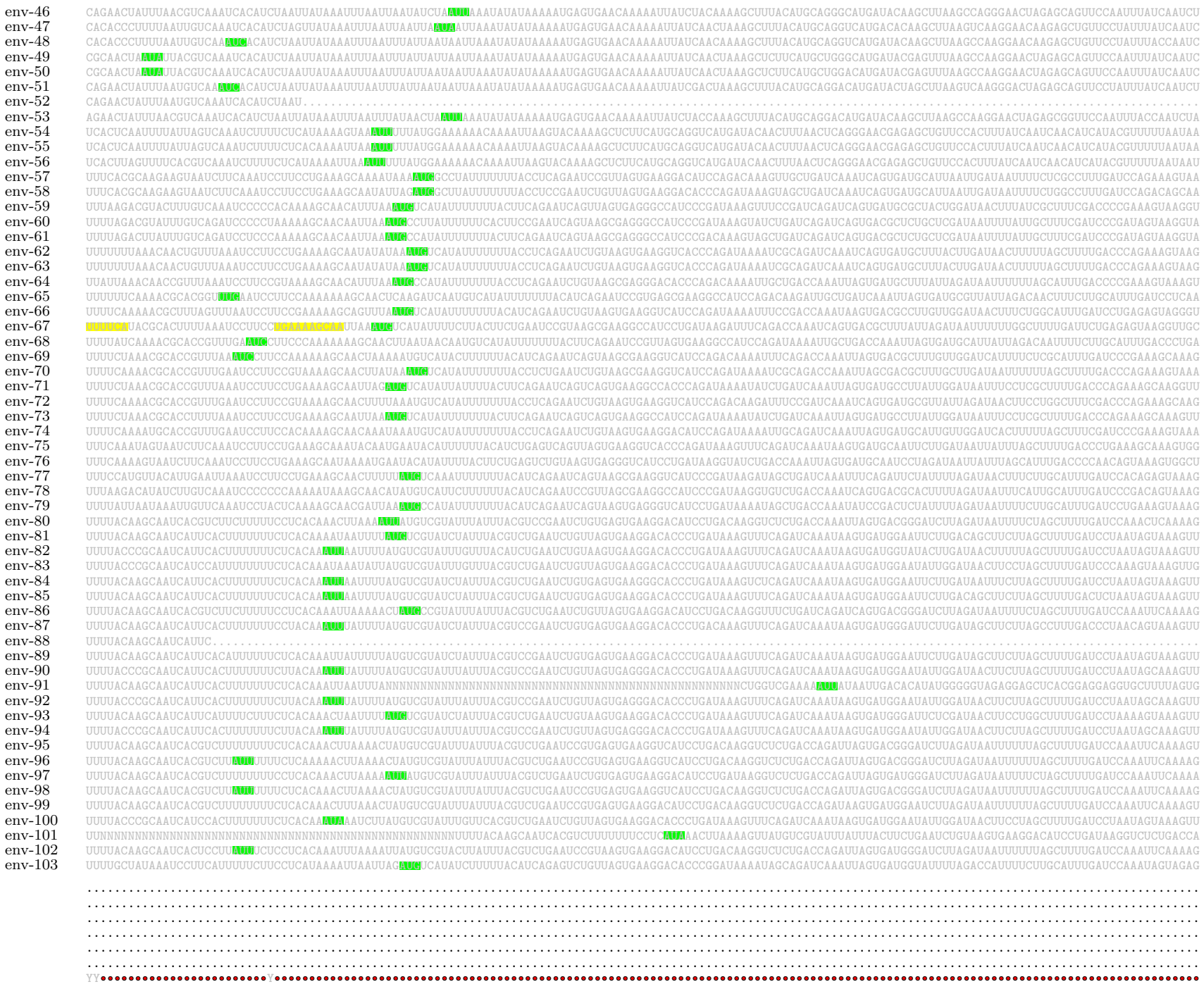

alignment positions 100 · · · 259
env-1
env-2
5′
3′


env-3
Pma-2-1
env-8
env-9
env-10
env-11
env-12
Pma-1-1
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
Psp-2-1
Gfo-1-1
env-26
env-27
Fba-1-1
env-28
env-29
ueu-1-1
Pgi-1-1
Pgi-2-1
Pdi-1-1
Pjo-1-1
Pme-1-1
env-30
env-31
env-32
Ddo-1-1
env-33
env-34
env-35
env-36
Rbi-1-1
Lbl-1-1
Fba-3-1
env-37
Cat-1-1
Pto-1-1
env-38
env-39
env-40
env-41
env-42
5′
3′


env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
5′
3′


env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
Pir-1-1
env-108
env-109
env-110
Fjo-1-1
env-111
Fps-1-1
env-112
Fba-2-1
Psp-1-1
env-113
env-114
env-115
Bvu-1-1
env-118
Bdo-1-1
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
Bun-1-1
Bfr-1-1
Bfr-2-1
env-130
Bca-1-1
Bth-1-1
Bov-1-1
env-131
env-132
5′
3′
alignment positions 260 · · · 419
env-1
env-2
env-3
env-4


env-5
Pma-2-1
env-8
env-9
env-10
env-11
env-12
Pma-1-1
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
Psp-2-1
Gfo-1-1
env-26
env-27
Fba-1-1
env-28
env-29
ueu-1-1
Pgi-1-1
Pgi-2-1
Pdi-1-1
Pjo-1-1
Pme-1-1
env-30
env-31
env-32
Ddo-1-1
env-33
env-34
env-35
env-36
Rbi-1-1
Lbl-1-1
Fba-3-1
env-37
Cat-1-1
Pto-1-1
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45


env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103

env-104
env-105
env-106
env-107
Pir-1-1
env-108
env-109
env-110
Fjo-1-1
env-111
Fps-1-1
env-112
Fba-2-1
Psp-1-1
env-113
env-114
env-115
Bvu-1-1
env-118
Bdo-1-1
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
Bun-1-1
Bfr-1-1
Bfr-2-1
env-130
Bca-1-1
Bth-1-1
Bov-1-1
env-131
env-132
alignment positions 420 · · · 558
env-1
env-2
env-3
env-4
env-5
env-6
env-7
Pma-2-1
env-8

env-9
env-10
env-11
env-12
Pma-1-1
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
Psp-2-1
Gfo-1-1
env-26
env-27
Fba-1-1
env-28
env-29
ueu-1-1
Pgi-1-1
Pgi-2-1
Pdi-1-1
Pjo-1-1
Pme-1-1
env-30
env-31
env-32
Ddo-1-1
env-33
env-34
env-35
env-36
Rbi-1-1
Lbl-1-1
Fba-3-1
env-37
Cat-1-1
Pto-1-1
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50

env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
Pir-1-1

87 SAM/SAH riboswitch

87.1 Taxa

The taxonomy of each organism containing a putative SAM/SAH riboswitch RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual SAM/SAH riboswitch RNA in Sections 87.2 and 87.4.

abbrev. of hits	taxonomy of species
Dsh-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Dinoroseobacter shibae* DFL 12
Jsp-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Jannaschia sp.* CCS1
Lve-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Loktanella vestfoldensis* SKA53
Oin-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Oceanibulbus indolifex* HEL-45
Oba-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Oceanicola batsensis* HTCC2597
Ogr-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Oceanicola granulosus* HTCC2516
Oan-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Octadecabacter antarcticus* 238
Oan-2-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Octadecabacter antarcticus* 307

Pde-1-1 to Pde-1-2	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Paracoccus denitrificans* PD1222
Pga-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Phaeobacter gallaeciensis* 2.10
Pga-2-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Phaeobacter gallaeciensis* BS107
Rsp-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Rhodobacter sphaeroides* 2.4.1
Rsp-2-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Rhodobacter sphaeroides* ATCC 17025
Rsp-3-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Rhodobacter sphaeroides* ATCC 17029
Rde-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Roseobacter denitrificans* OCh 114
Rli-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Roseobacter litoralis* Och 149
Rsp-4-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Roseobacter sp.* AzwK-3b
Rsp-5-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Roseobacter sp.* CCS2
Rsp-6-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Roseobacter sp.* GAI101
Rsp-7-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Roseobacter sp.* MED193
Rsp-8-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Roseobacter sp.* SK209-2-6
Rnu-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Roseovarius nubinhibens* ISM
Rsp-9-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Roseovarius sp.* 217
Rsp-10-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Roseovarius sp.* HTCC2601
Rsp-11-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Roseovarius sp.* TM1035
Rsp-12-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Ruegeria sp.* R11
Sst-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Sagittula stellata* E-37
Spo-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Silicibacter pomeroyi* DSS-3
Ssp-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Silicibacter sp.* TM1040
Ssp-2-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Sulfitobacter sp.* EE-36
Ssp-3-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Sulfitobacter sp.* NAS-14.1
Rba-1-1	Bacteria Proteobacteria Alphaproteobacteria *Rhodobacterales bacterium* HTCC2083
Rba-2-1	Bacteria Proteobacteria Alphaproteobacteria *Rhodobacterales bacterium* HTCC2150
Rba-3-1	Bacteria Proteobacteria Alphaproteobacteria *Rhodobacterales bacterium* HTCC2255
Rba-4-1	Bacteria Proteobacteria Alphaproteobacteria *Rhodobacterales bacterium* HTCC2654
Rba-5-1	Bacteria Proteobacteria Alphaproteobacteria *Rhodobacterales bacterium* Y4I
env-1 to env-27	environmental samples

87.2 Gene contexts

Each SAM/SAH riboswitch RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one SAM/SAH riboswitch RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 87.3. The accession number of the sequence containing each SAM/SAH riboswitch RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 87.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Rba-3-1	NZ_AATR01000003.1	+	91067	91114	**RNA**→ MetK (COG0192)→ COG0220 (COG0220)→
env-1	JCVI_SCAF_1101667010544	-	622	575	**RNA**→ MetK (COG0192)? (GOS4376)→
Oan-1-1	NZ_ABSK01000009.1	-	52599	52551	**RNA**→
Dsh-1-1	NC_009952.1	-	1046963	1046915	**RNA**→ MetK (COG0192)→
env-2	JCVI_SCAF_1096627477496	-	1357	1311	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→ MetK (COG0192)? (GOS4376)→ ? (GOS209)→
env-3	JCVI_SCAF_1096627420315	+	1577	1623	**RNA**→
env-4	JCVI_SCAF_1096626908805	-	1378	1332	**RNA**→ MetK (COG0192)? (GOS4376)→
env-5	JCVI_SCAF_1096627427430	-	1264	1217	**RNA**→ MetK (COG0192)? (GOS4376)→ S-AdoMet_synt_C (pfam02773)? (GOS4376)→
env-6	JCVI_SCAF_1096626925902	+	989	1034	**RNA**→
env-7	JCVI_SCAF_1101668680689	+	90	135	**RNA**→ MetK (COG0192)? (GOS4376)→
env-8	JCVI_SCAF_1096626987208	-	1022	977	**RNA**→ MetK (COG0192)? (GOS4376)→
env-9	JCVI_SCAF_1096627600923	+	1335	1380	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→

env-10	JCVI_SCAF_1096627832706	+	1214	1261	**RNA**→
env-11	JCVI_SCAF_1096627086091	+	1080	1127	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-12	JCVI_SCAF_1096627847607	+	224	271	**RNA**→ MetK (COG0192)? (GOS4376)→ S-AdoMet_synt_C (pfam02773)? (GOS4376)→ COG0220 (COG0220)? (GOS209)→
env-13	JCVI_SCAF_1096627093504	+	246	293	**RNA**→ MetK (COG0192)? (GOS4376)→ COG0220 (COG0220)? (GOS209)→
env-14	JCVI_SCAF_1101668567300	-	393	346	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-15	JCVI_SCAF_1096627095309	-	785	738	**RNA**→ MetK (COG0192)? (GOS4376)→
env-16	JCVI_SCAF_1096627841121	-	508	461	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-17	JCVI_SCAF_1096627064213	-	3303	3256	**RNA**→ MetK (COG0192)? (GOS4376)→ COG0220 (COG0220)? (GOS209)→ EPSP_synthase (cd01556)? (GOS3778)→
env-18	wf1_Contig785	+	253	301	**RNA**→ MetK (COG0192)→
Oan-2-1	NZ_ABSH01000031.1	-	84911	84863	**RNA**→
Oba-1-1	NZ_AAMO01000011.1	+	96876	96925	**RNA**→ MetK (COG0192)→
Rba-2-1	NZ_AAXZ01000012.1	+	36217	36265	**RNA**→ MetK (COG0192)→
Lve-1-1	NZ_AAMS01000003.1	+	12007	12055	**RNA**→ MetK (COG0192)→ COG0220 (COG0220)→ EPSP_synthase (cd01556)→ Cmk (COG0283)→
Jsp-1-1	NC_007802.1	+	678143	678190	**RNA**→ MetK (COG0192)→ COG0220 (COG0220)→
Rsp-7-1	NZ_AANB01000001.1	+	292634	292681	**RNA**→ MetK (COG0192)→
env-19	JCVI_SCAF_1096627354361	-	1534	1483	**RNA**→ MetK (COG0192)? (GOS4376)→ ? (GOS6241)→
env-20	JCVI_SCAF_1096627970754	-	251	203	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-21	JCVI_SCAF_1096627362274	-	2663	2615	**RNA**→ MetK (COG0192)? (GOS4376)→
Rsp-8-1	NZ_AAYC01000001.1	+	142050	142098	**RNA**→ MetK (COG0192)→
env-22	wf2_AHAA9457_y1	-	319	270	**RNA**→ S-AdoMet_synt_N (pfam00438)→
env-23	JCVI_SCAF_1101668636529	+	189	238	**RNA**→ MetK (COG0192)? (GOS4376)→
env-24	JCVI_SCAF_1096627147823	+	1017	1066	**RNA**→ MetK (COG0192)? (GOS4376)→
Rsp-9-1	NZ_AAMV01000006.1	+	34838	34885	**RNA**→ MetK (COG0192)→ COG0220 (COG0220)→
Rsp-11-1	NZ_ABCL01000003.1	-	82568	82521	**RNA**→ MetK (COG0192)→ COG0220 (COG0220)→
Rsp-2-1	NC_009429.1	+	776419	776468	**RNA**→ MetK (COG0192)→
Rsp-3-1	NC_009050.1	-	343169	343120	**RNA**→ MetK (COG0192)→
Rsp-1-1	NC_007494.1	-	688624	688575	**RNA**→ MetK (COG0192)→
Rba-4-1	NZ_AAMT01000035.1	-	1434	1386	**RNA**→ MetK (COG0192)→
Pde-1-1	NC_008686.1	+	1874599	1874647	**RNA**→ MetK (COG0192)→
Pde-1-2	NZ_AAIT01000003.1	+	251041	251089	**RNA**→ MetK (COG0192)→
Ssp-1-1	NC_008044.1	+	211359	211407	**RNA**→ MetK (COG0192)→
env-25	JCVI_SCAF_1096627351014	-	2673	2625	**RNA**→ MetK (COG0192)? (GOS4376)→
Sst-1-1	NZ_AAYA01000008.1	-	121966	121918	**RNA**→ MetK (COG0192)→ COG0220 (COG0220)→
Rsp-10-1	NZ_AATQ01000051.1	-	20133	20085	**RNA**→ MetK (COG0192)→
Rsp-4-1	NZ_ABCR01000003.1	-	393478	393430	**RNA**→ MetK (COG0192)→
Rnu-1-1	NZ_AALY01000001.1	-	139779	139732	**RNA**→ MetK (COG0192)→
Spo-1-1	NC_006569.1	+	9372	9422	**RNA**→ MetK (COG0192)→ Adk (COG0563)→
Pga-1-1	NZ_ABIE01000005.1	-	70619	70571	**RNA**→ MetK (COG0192)→
Pga-2-1	NZ_ABIF01000007.1	+	192066	192114	**RNA**→ MetK (COG0192)→
Rsp-12-1	NZ_ABXM01000011.1	-	241752	241702	**RNA**→
Rba-5-1	NZ_ABXF01000008.1	+	119484	119533	**RNA**→
Ogr-1-1	NZ_AAOT01000020.1	-	3283	3235	**RNA**→ MetK (COG0192)→ S-adenosylmethionine synthetase→ hypo→
env-26	JCVI_SCAF_1096627934609	-	246	198	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
env-27	JCVI_SCAF_1096627355058	+	2154	2202	**RNA**→ S-AdoMet_synt_N (pfam00438)? (GOS4376)→
Rba-1-1	NZ_ABXE01000008.1	-	209807	209759	**RNA**→
Ssp-3-1	NZ_AALZ01000001.1	-	70323	70276	**RNA**→ MetK (COG0192)→
Ssp-2-1	NZ_AALV01000001.1	-	126167	126120	**RNA**→ MetK (COG0192)→
Rsp-5-1	NZ_AAYB01000002.1	+	503668	503715	**RNA**→ MetK (COG0192)→ COG0220 (COG0220)→
Rsp-6-1	NZ_ABXS01000051.1	+	24818	24865	**RNA**→
Oin-1-1	NZ_ABID01000011.1	-	17084	17036	**RNA**→ MetK (COG0192)→
Rde-1-1	NC_008209.1	-	3781791	3781742	**RNA**→ MetK (COG0192)→

Rli-1-1 NZ_ABIG01000009.1 + 69706 69755 **RNA**→ MetK (COG0192)→

87.3 Conserved domains

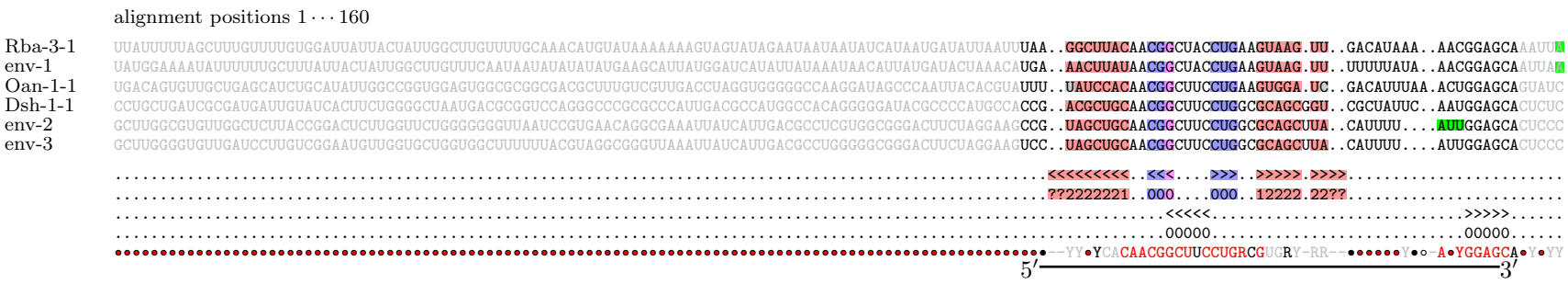
Conserved domains found in protein-coding genes listed in Section 87.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one SAM/SAH riboswitch RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

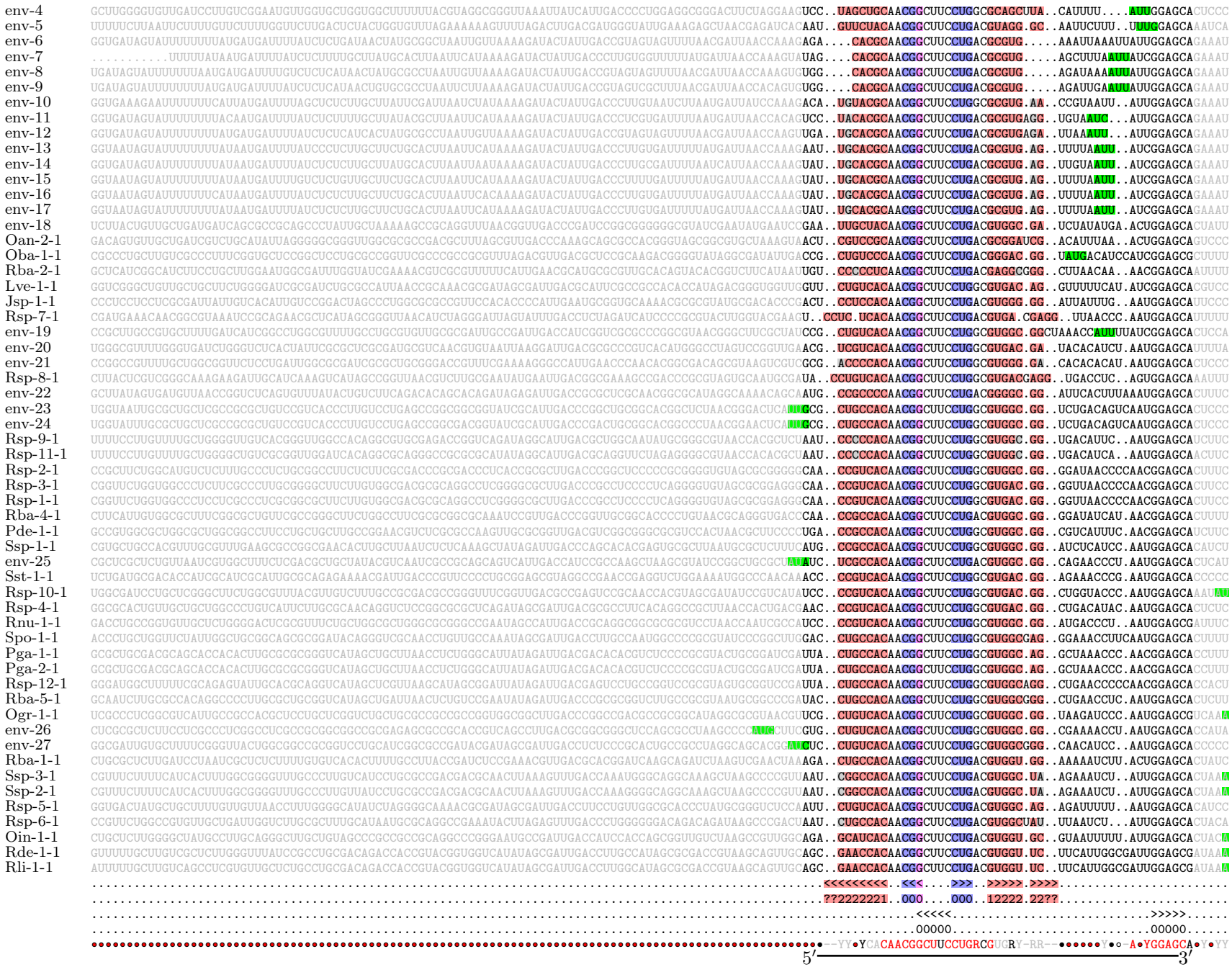
cd01556 (2) EPSP synthase domain.
COG0192 (47) S-adenosylmethionine synthetase [Coenzyme metabolism]
COG0220 (10) Predicted S-adenosylmethionine-dependent methyltransferase [General function prediction only]
COG0283 (1) Cytidylate kinase [Nucleotide transport and metabolism]
COG0563 (1) Adenylate kinase and related kinases [Nucleotide transport and metabolism]
GOS209 (4) ?
GOS3778 (1) ?
GOS4376 (25) ?
GOS6241 (1) ?
pfam00438 (9) S-adenosylmethionine synthetase, N-terminal domain.
pfam02773 (2) S-adenosylmethionine synthetase, C-terminal domain.

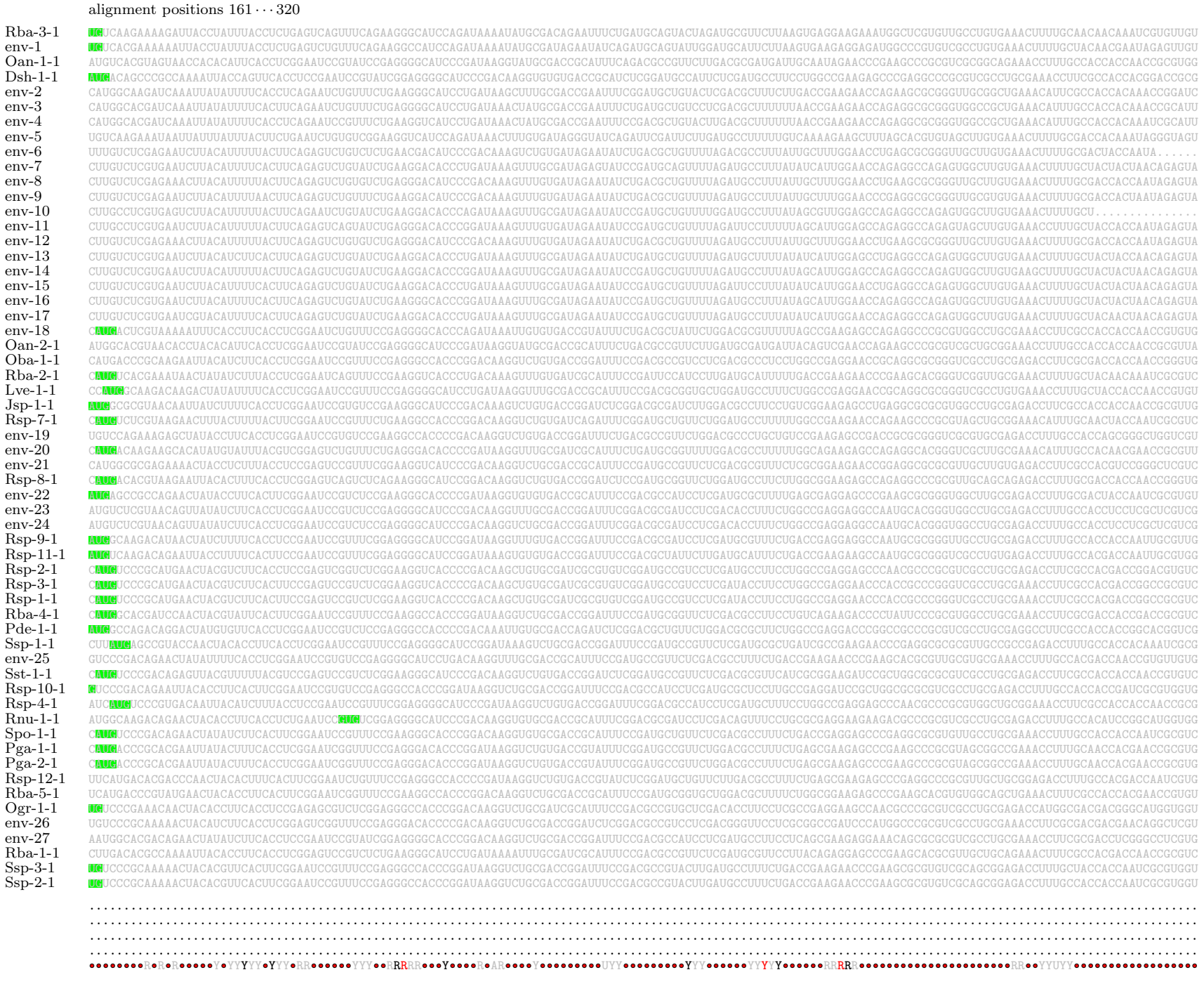
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Pde-1-2

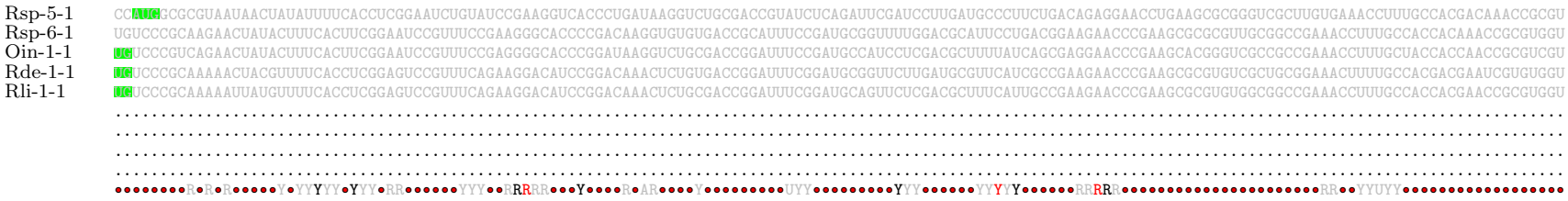
87.4 Multiple-sequence alignment

Each SAM/SAH riboswitch RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 87.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The SAM/SAH riboswitch RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

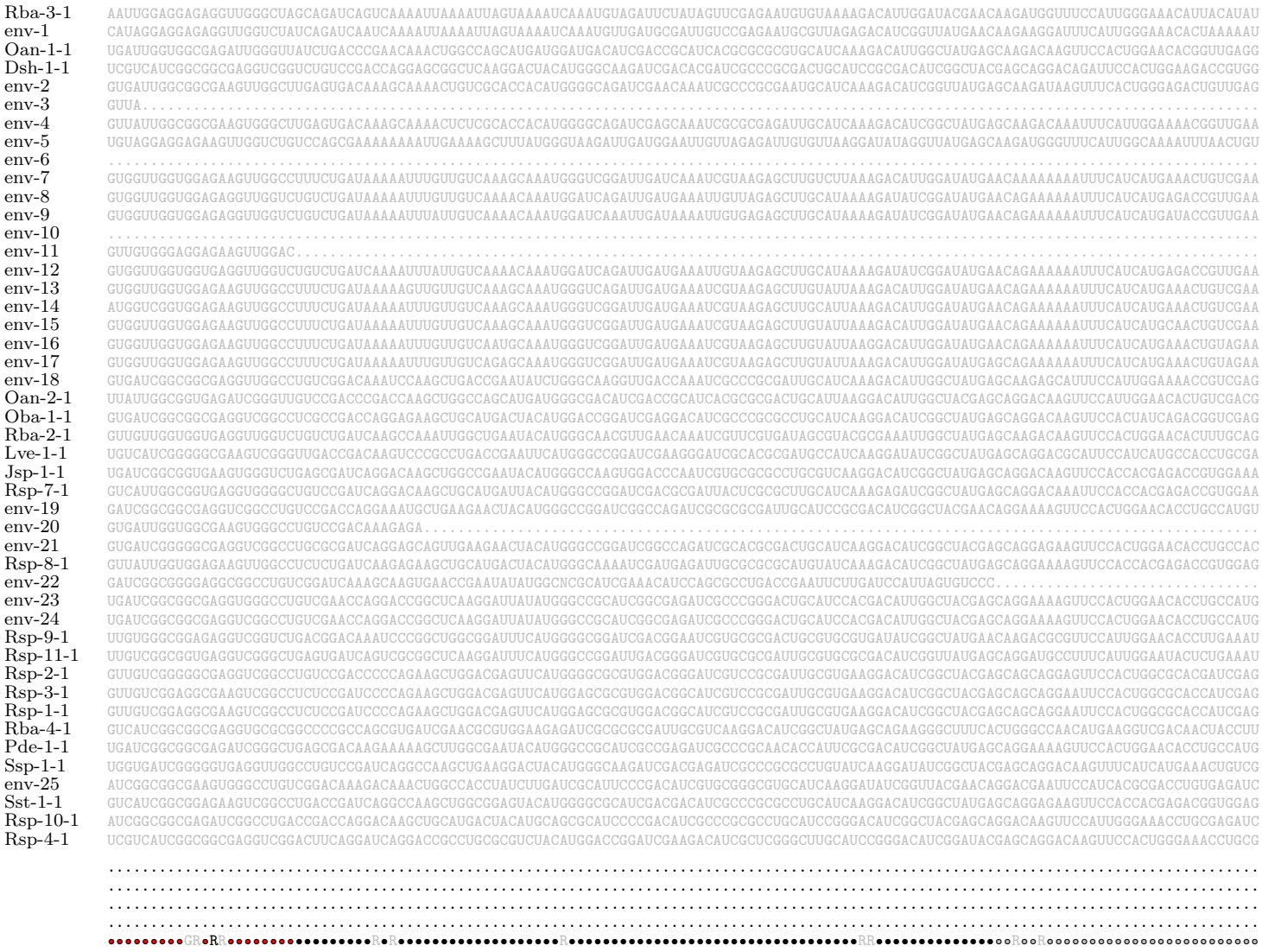



env-4
env-5
env-6
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
Oan-2-1
Oba-1-1
Rba-2-1
Lve-1-1
Jsp-1-1
Rsp-7-1
env-19
env-20
env-21
Rsp-8-1
env-22
env-23
env-24
Rsp-9-1
Rsp-11-1
Rsp-2-1
Rsp-3-1
Rsp-1-1
Rba-4-1
Pde-1-1
Ssp-1-1
env-25
Sst-1-1
Rsp-10-1
Rsp-4-1
Rnu-1-1
Spo-1-1
Pga-1-1
Pga-2-1
Rsp-12-1
Rba-5-1
Ogr-1-1
env-26
env-27
Rba-1-1
Ssp-3-1
Ssp-2-1
Rsp-5-1
Rsp-6-1
Oin-1-1
Rde-1-1
Rli-1-1
5′
3′


alignment positions 161 · · · 320
Rba-3-1
env-1
Oan-1-1
Dsh-1-1
env-2
env-3
env-4
env-5
env-6
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
Oan-2-1
Oba-1-1
Rba-2-1
Lve-1-1
Jsp-1-1
Rsp-7-1
env-19
env-20
env-21
Rsp-8-1
env-22
env-23
env-24
Rsp-9-1
Rsp-11-1
Rsp-2-1
Rsp-3-1
Rsp-1-1
Rba-4-1
Pde-1-1
Ssp-1-1
env-25
Sst-1-1
Rsp-10-1
Rsp-4-1
Rnu-1-1
Spo-1-1
Pga-1-1
Pga-2-1
Rsp-12-1
Rba-5-1
Ogr-1-1
env-26
env-27
Rba-1-1
Ssp-3-1
Ssp-2-1



alignment positions 321 · · · 455



88 *sanguinis*-hairpin

88.1 Taxa

The taxonomy of each organism containing a putative *sanguinis*-hairpin RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *sanguinis*-hairpin RNA in Sections 88.2 and 88.4.

abbrev. of hits	**taxonomy of species**
Ssa-1-1 to Ssa-1-9	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus sanguinis* SK36
Sth-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* CNRZ1066
Sth-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* LMD-9
Sth-3-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* LMG 18311

88.2 Gene contexts

Each *sanguinis*-hairpin RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *sanguinis*-hairpin RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 88.3. The accession number of the sequence containing each *sanguinis*-hairpin RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 88.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Ssa-1-1	NC_009009.1	-	1327053	1327003	COG1837 (COG1837)→ **RNA**→ hypo→ TrkG (COG0168)→ TrkA (COG0569)→
Ssa-1-2	NC_009009.1	-	2097852	2097798	Conserved uncharacterized protein→ **RNA**→ **RNA**→ **RNA**→ SgaT_UlaA (pfam04215)→ SgaB (COG3414)→
Ssa-1-3	NC_009009.1	-	2097936	2097886	(shown above)
Ssa-1-4	NC_009009.1	-	2097765	2097715	(shown above)
Ssa-1-5	NC_009009.1	+	72979	73030	PutA (COG1012)EutG (COG1454)→ **RNA**→ Hypothetical protein, possibly membrane-associated→
Ssa-1-6	NC_009009.1	-	2183081	2183029	component of signal recognition particle (SRP)→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ Pgi (COG0166)→
Ssa-1-7	NC_009009.1	-	2182905	2182853	(shown above)
Ssa-1-8	NC_009009.1	-	2182987	2182936	(shown above)
Ssa-1-9	NC_009009.1	-	2183163	2183112	(shown above)
Sth-3-1	NC_006448.1	+	454927	454979	hypo→ **RNA**→ COG1853 (COG1853)→ COG1853 (COG1853)→ RimL (COG1670)→
Sth-1-1	NC_006449.1	+	450152	450204	hypo→ **RNA**→ COG1853 (COG1853)→ RimL (COG1670)→
Sth-2-1	NC_008532.1	+	459182	459234	COG1853 (COG1853)→ **RNA**→ COG1853 (COG1853)→ RimL (COG1670)→

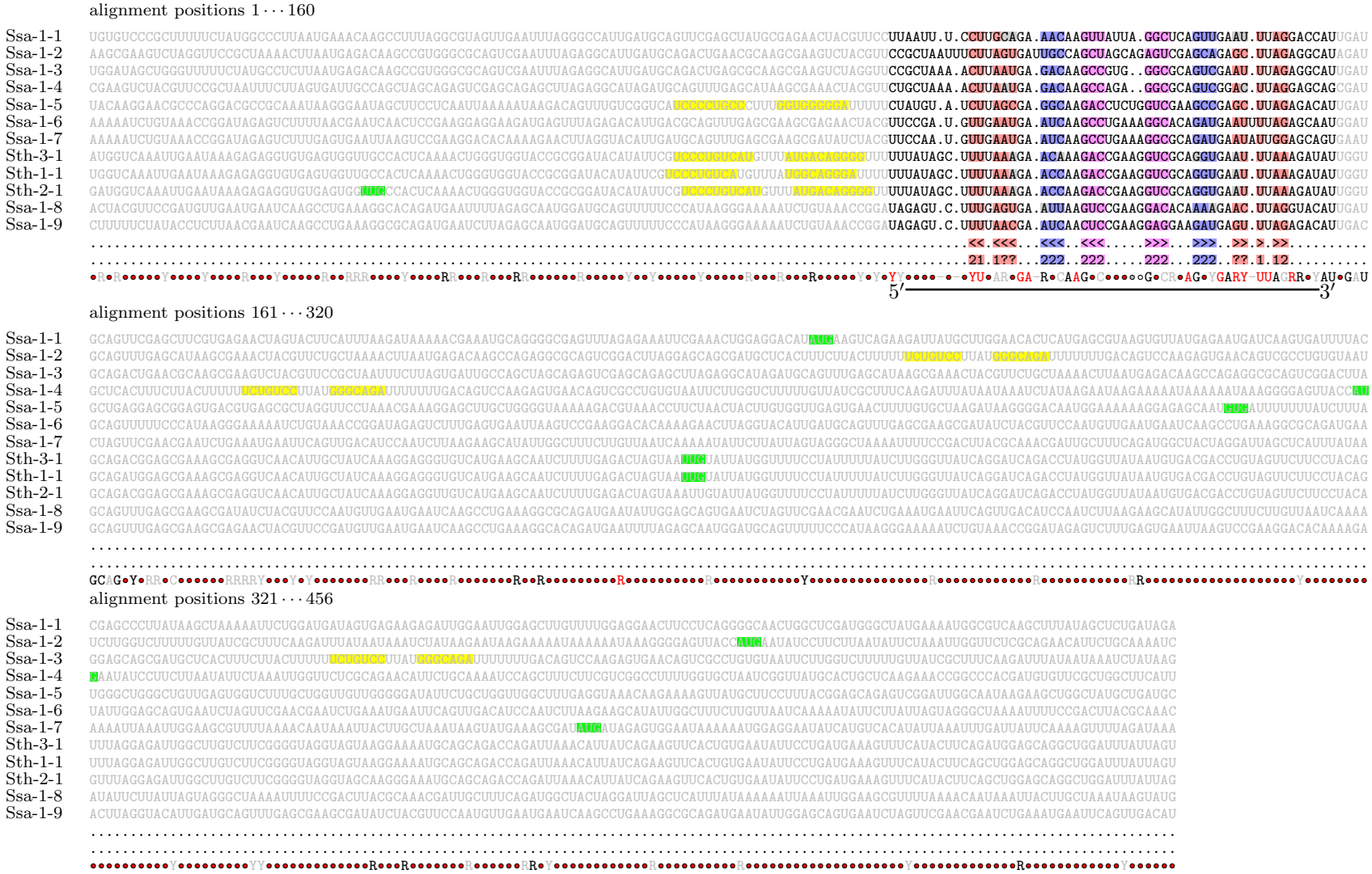
88.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 88.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *sanguinis*-hairpin RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

D COG0166 (1) Glucose-6-phosphate isomerase [Carbohydrate transport and metabolism]
D COG0168 (1) Trk-type K+ transport systems, membrane components [Inorganic ion transport and metabolism]
D COG0569 (1) K+ transport systems, NAD-binding component [Inorganic ion transport and metabolism]
u COG1012 (1) NAD-dependent aldehyde dehydrogenases [Energy production and conversion]
u COG1454 (1) Alcohol dehydrogenase, class IV [Energy production and conversion]
D COG1670 (3) Acetyltransferases, including N-acetylases of ribosomal proteins [Translation, ribosomal structure and biogenesis]
u COG1837 (1) Predicted RNA-binding protein (contains KH domain) [General function prediction only]
uD COG1853 (5) Conserved protein/domain typically associated with flavoprotein oxygenases, DIM6/NTAB family [General function prediction only]
D COG3414 (1) Phosphotransferase system, galactitol-specific IIB component [Carbohydrate transport and metabolism]
D pfam04215 (1) Putative sugar-specific permease, SgaT/UlaA.

88.4 Multiple-sequence alignment

Each *sanguinis*-hairpin RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 88.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *sanguinis*-hairpin RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



89 *sbcD*

89.1 Taxa

The taxonomy of each organism containing a putative *sbcD* RNA is listed, with abbreviations identifying each hit (e.g., “Eco-1-1” and “Eco-1-2” might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *sbcD* RNA in Sections 89.2 and 89.4.

abbrev. of hits	taxonomy of species
Bam-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia ambifaria* MC40-6
Bam-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia ambifaria* MEX-5
Bvi-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia vietnamiensis* G4
Reu-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Cupriavidus *Ralstonia eutropha* H16
Rme-1-1 to Rme-1-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Cupriavidus *Ralstonia metallidurans* CH34
Rpi-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Ralstonia pickettii* 12D
Rpi-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae *Ralstonia pickettii* 12J

89.2 Gene contexts

Each *sbcD* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *sbcD* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 89.3. The accession number of the sequence containing each *sbcD* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 89.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Bam-1-1	NC_010553.1	-	294183	294046	←hypo **RNA**→ hypo→
Bam-2-1	NZ_ABLK01000025.1	-	8990	8853	←hypo (cons) **RNA**→ hypo (cons)→
Bvi-1-1	NC_009227.1	+	92362	92499	←hypo **RNA**→ hypo→
Rpi-2-1	NC_010682.1	+	1712737	1712884	←hypo **RNA**→ hypo→
Rpi-1-1	NZ_ABDZ01000013.1	+	70607	70754	←hypo (cons) **RNA**→ hypo→
Rme-1-1	NC_006525.1	+	102218	102372	←hypo **RNA**→ hypo→
Rme-1-2	NC_007972.1	-	42518	42364	←hypo **RNA**→ hypo→
Reu-1-1	NC_005241.1	-	312212	312058	←hypo **RNA**→ ←hypo

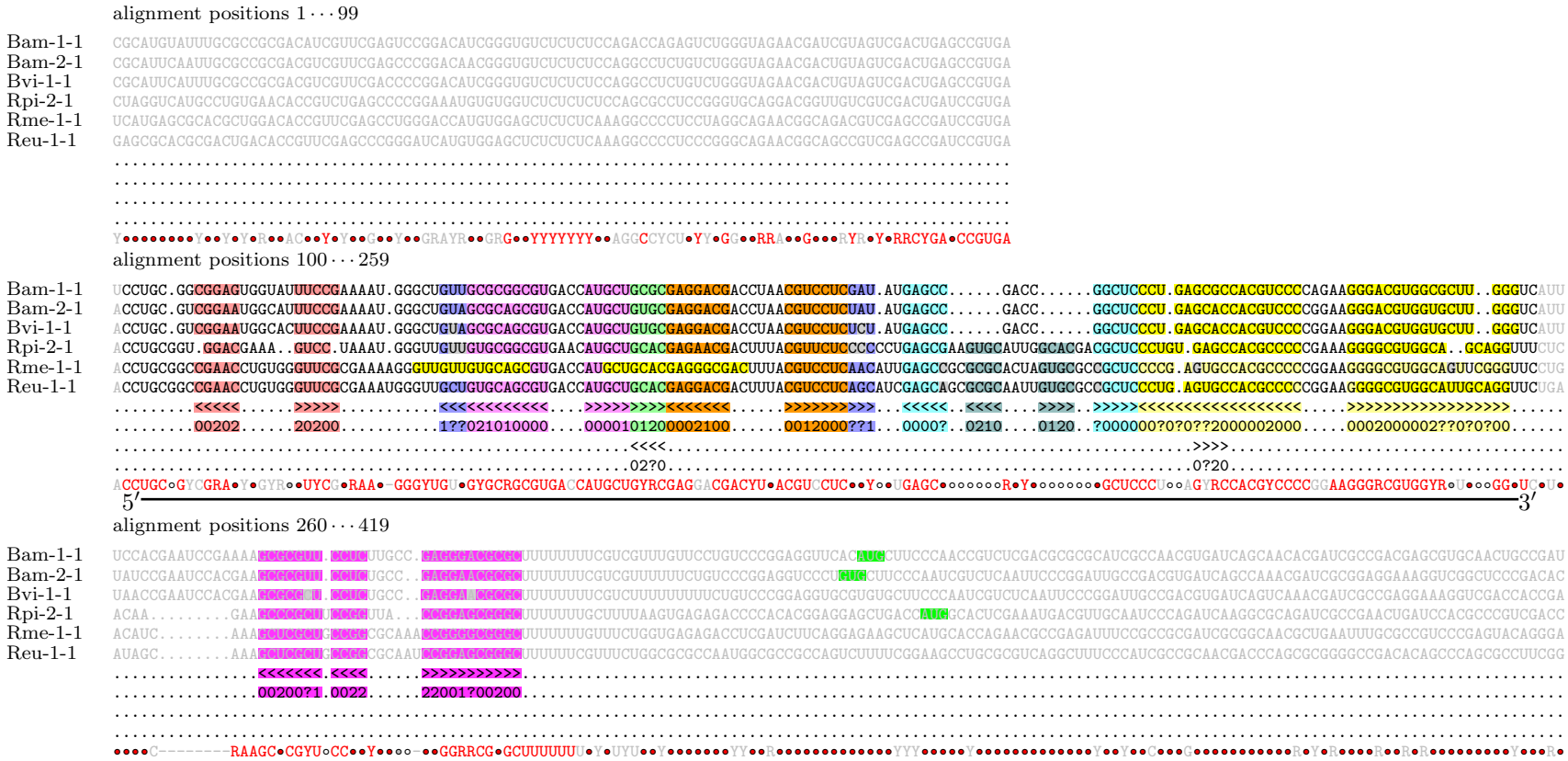
89.3 Conserved domains

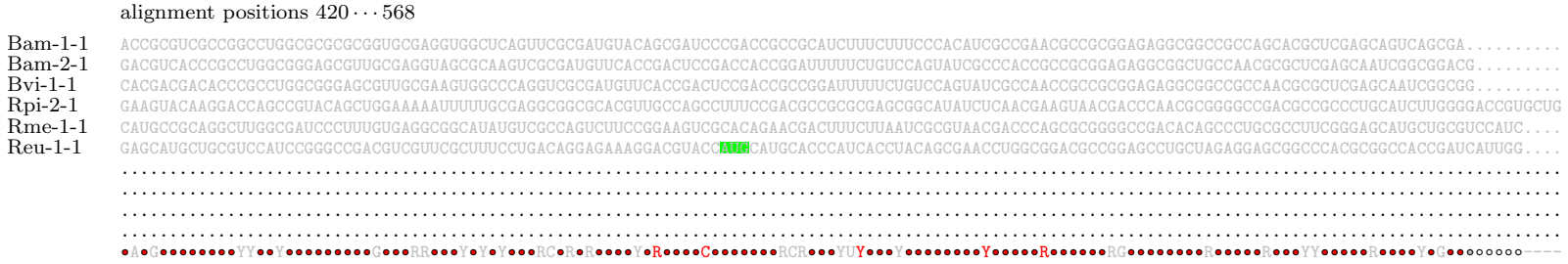
Conserved domains found in protein-coding genes listed in Section 89.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *sbcD* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Rme-1-2, Rpi-1-1

89.4 Multiple-sequence alignment

Each *sbcD* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 89.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *sbcD* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.





90 ScRE

90.1 Taxa

The taxonomy of each organism containing a putative ScRE RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual ScRE RNA in Sections 90.2 and 90.4.

abbrev. of hits	taxonomy of species
Spn-1-1 to Spn-1-5	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* D39
Spn-2-1 to Spn-2-5	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* R6
Spn-3-1 to Spn-3-5	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP11-BS70
Spn-4-1 to Spn-4-5	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP14-BS69
Spn-5-1 to Spn-5-5	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP18-BS74
Spn-6-1 to Spn-6-5	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP19-BS75
Spn-7-1 to Spn-7-5	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP23-BS72
Spn-8-1 to Spn-8-5	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP3-BS71
Spn-9-1 to Spn-9-5	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP6-BS73
Spn-10-1 to Spn-10-5	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* SP9-BS68
Spn-11-1 to Spn-11-7	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* TIGR4
Sth-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* CNRZ1066
Sth-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* LMG 18311

90.2 Gene contexts

Each ScRE RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one ScRE RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 90.3. The accession number of the sequence containing each ScRE RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 90.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Sth-2-1	NC_006448.1	-	931217	931134	**RNA**→ NUDIX (pfam00293)→ UDP-N-acetylglucosamine enolpyruvyl transferase, putative, truncated→
Sth-1-1	NC_006449.1	-	937217	937134	**RNA**→ NUDIX (pfam00293)→ UdpNAET (cd01555)→ hypo→ Tra8 (COG2826)→

Spn-5-1	NZ_ABAE01000001.1	+	178781	178861	**RNA**→ LacZ (COG3250)COG3583 (COG3583)→
Spn-1-1	NC_008533.1	+	579577	579656	**RNA**→ LacZ (COG3250)→
Spn-2-1	NC_003098.1	+	572046	572125	**RNA**→ LacZ (COG3250)→
Spn-6-1	NZ_ABAF01000001.1	-	323318	323239	**RNA**→ LacZ (COG3250)COG3583 (COG3583)→
Spn-9-1	NZ_ABAA01000005.1	+	77694	77773	**RNA**→ LacZ (COG3250)→ COG3583 (COG3583)→
Spn-11-1	NC_003028.2	+	615365	615444	**RNA**→ LacZ (COG3250)COG3583 (COG3583)→
Spn-8-1	NZ_AAZZ01000008.1	+	1875	1954	**RNA**→ LacZ (COG3250)COG3583 (COG3583)→
Spn-10-1	NZ_ABAB01000002.1	-	38629	38550	**RNA**→ LacZ (COG3250)COG3583 (COG3583)→
Spn-3-1	NZ_ABAC01000002.1	-	266993	266914	**RNA**→ LacZ (COG3250)COG3583 (COG3583)→
Spn-4-1	NZ_ABAD01000003.1	+	115602	115681	**RNA**→ LacZ (COG3250)→ COG3583 (COG3583)→
Spn-7-1	NZ_ABAG01000005.1	+	9898	9977	**RNA**→ LacZ (COG3250)COG3583 (COG3583)→
Spn-11-2	NZ_AAGY02000050.1	+	85	164	**RNA**→
Spn-9-2	NZ_ABAA01000018.1	+	9217	9297	**RNA**→ Pep_deformylase (pfam01327)→ SpoU (COG0566)→
Spn-11-3	NC_003028.2	+	1372820	1372900	**RNA**→ Pep_deformylase (pfam01327)→ SpoU (COG0566)→
Spn-11-4	NZ_AAGY02000082.1	+	8273	8353	**RNA**→ Pep_deformylase (pfam01327)→ SpoU (COG0566)→
Spn-6-2	NZ_ABAF01000004.1	+	90557	90637	**RNA**→ Pep_deformylase (pfam01327)→ SpoU (COG0566)→
Spn-1-2	NC_008533.1	+	1308216	1308296	**RNA**→ Pep_deformylase (pfam01327)→ SpoU (COG0566)→
Spn-2-2	NC_003098.1	+	1300685	1300765	**RNA**→ Pep_deformylase (pfam01327)→ SpoU (COG0566)→
Spn-8-2	NZ_AAZZ01000005.1	+	23855	23935	**RNA**→ Pep_deformylase (pfam01327)→ SpoU (COG0566)→
Spn-10-2	NZ_ABAB01000004.1	-	96178	96098	**RNA**→ Pep_deformylase (pfam01327)→ SpoU (COG0566)→
Spn-3-2	NZ_ABAC01000003.1	-	123521	123441	**RNA**→ Pep_deformylase (pfam01327)→ SpoU (COG0566)→
Spn-4-2	NZ_ABAD01000006.1	+	92832	92912	**RNA**→ Pep_deformylase (pfam01327)→ SpoU (COG0566)→
Spn-7-2	NZ_ABAG01000002.1	+	190789	190869	**RNA**→ Pep_deformylase (pfam01327)→ TrmH family RNA methyltransferase group 3→ SpoU (COG0566)→
Spn-5-2	NZ_ABAE01000009.1	-	17089	17009	**RNA**→ Pep_deformylase (pfam01327)→ SpoU (COG0566)→
Spn-9-3	NZ_ABAA01000013.1	-	6287	6205	**RNA**→ MdlB (COG1132)→
Spn-6-3	NZ_ABAF01000005.1	-	69982	69900	**RNA**→ MdlB (COG1132)→
Spn-7-3	NZ_ABAG01000016.1	-	6265	6183	**RNA**→ MdlB (COG1132)→
Spn-5-3	NZ_ABAE01000006.1	+	128580	128662	**RNA**→ MdlB (COG1132)→
Spn-1-3	NC_008533.1	-	1870856	1870774	**RNA**→ MdlB (COG1132)→ COG2071 (COG2071)→ hypo→ Pgi (COG0166)→
Spn-2-3	NC_003098.1	-	1863324	1863242	**RNA**→ MdlB (COG1132)→ COG2071 (COG2071)→ hypo→ Pgi (COG0166)→
Spn-8-3	NZ_AAZZ01000003.1	-	11732	11650	**RNA**→ MdlB (COG1132)→
Spn-4-3	NZ_ABAD01000007.1	-	6271	6189	**RNA**→ MdlB (COG1132)→
Spn-3-3	NZ_ABAC01000005.1	+	185721	185803	**RNA**→ hypo→ MdlB (COG1132)→
Spn-11-5	NC_003028.2	-	1982404	1982322	**RNA**→ MdlB (COG1132)→ COG2071 (COG2071)→ hypo→ Pgi (COG0166)→
Spn-10-3	NZ_ABAB01000001.1	+	146115	146197	**RNA**→ MdlB (COG1132)→
Spn-3-4	NZ_ABAC01000001.1	+	388076	388157	**RNA**→ BacA (pfam02673)→ undecaprenyl pyrophosphate phosphatase→
Spn-11-6	NC_003028.2	+	429828	429909	**RNA**→ BacA (pfam02673)→
Spn-10-4	NZ_ABAB01000058.1	-	1780	1699	**RNA**→ BacA (pfam02673)→ DNA polymerase IV→
Spn-4-4	NZ_ABAD01000013.1	+	758	839	**RNA**→ BacA (pfam02673)→ undecaprenyl pyrophosphate phosphatase→
Spn-8-4	NZ_AAZZ01000001.1	+	290382	290463	**RNA**→ BacA (pfam02673)→ undecaprenyl pyrophosphate phosphatase→
Spn-5-4	NZ_ABAE01000001.1	+	163	244	**RNA**→ BacA (pfam02673)→ hypo→
Spn-6-4	NZ_ABAF01000006.1	+	108389	108470	**RNA**→ BacA (pfam02673)→
Spn-9-4	NZ_ABAA01000001.1	+	396277	396358	**RNA**→ BacA (pfam02673)→ undecaprenyl pyrophosphate phosphatase→
Spn-1-4	NC_008533.1	+	417116	417197	**RNA**→ BacA (pfam02673)→ hypo→
Spn-2-4	NC_003098.1	+	409587	409668	**RNA**→ BacA (pfam02673)→
Spn-1-5	NC_008533.1	+	828526	828607	**RNA**→ SpeE (COG0421)→ Saccharop_dh (pfam03435)→ LysA (COG0019)→ COG2957 (COG2957)→ COG0388 (COG0388)→
Spn-2-5	NC_003098.1	+	820996	821077	**RNA**→ SpeE (COG0421)→ Saccharop_dh (pfam03435)→ LysA (COG0019)→ COG2957 (COG2957)→ COG0388 (COG0388)→
Spn-9-5	NZ_ABAA01000002.1	+	80158	80239	**RNA**→ SpeE (COG0421)→ Saccharop_dh (pfam03435)→ Saccharop_dh (pfam03435)→ LysA (COG0019)→ COG2957 (COG2957)→ C
Spn-7-4	NZ_ABAG01000015.1	+	56131	56212	**RNA**→
Spn-8-5	NZ_AAZZ01000004.1	+	74304	74385	**RNA**→ SpeE (COG0421)→ Saccharop_dh (pfam03435)→ LysA (COG0019)→ COG2957 (COG2957)→ COG0388 (COG0388)→
Spn-10-5	NZ_ABAB01000006.1	-	75003	74922	**RNA**→ SpeE (COG0421)→ Saccharop_dh (pfam03435)→ LysA (COG0019)→ COG2957 (COG2957)→ COG0388 (COG0388)→
Spn-3-5	NZ_ABAC01000002.1	-	19592	19511	**RNA**→ SpeE (COG0421)→ Saccharop_dh (pfam03435)→ LysA (COG0019)→ COG2957 (COG2957)→ COG0388 (COG0388)→

Spn-11-7	NC_003028.2	+	872203	872284	**RNA**→ SpeE (COG0421)→ Saccharop_dh (pfam03435)→ LysA (COG0019)→ COG2957 (COG2957)→ COG0388 (COG0388)→
Spn-6-5	NZ_ABAF01000001.1	-	73257	73176	**RNA**→ SpeE (COG0421)→ Saccharop_dh (pfam03435)→ LysA (COG0019)→ COG2957 (COG2957)→ COG0388 (COG0388)→
Spn-4-5	NZ_ABAD01000004.1	+	60482	60563	**RNA**→ SpeE (COG0421)→ Saccharop_dh (pfam03435)→ LysA (COG0019)→ COG2957 (COG2957)→ COG0388 (COG0388)→
Spn-5-5	NZ_ABAE01000003.1	-	213546	213465	**RNA**→ SpeE (COG0421)→ Saccharop_dh (pfam03435)→ LysA (COG0019)→ COG2957 (COG2957)→ COG0388 (COG0388)→
Spn-7-5	NZ_ABAG01000003.1	+	26920	27001	**RNA**→ SpeE (COG0421)→ Saccharop_dh (pfam03435)→ LysA (COG0019)→ COG2957 (COG2957)→ COG0388 (COG0388)→

90.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 90.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one ScRE RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

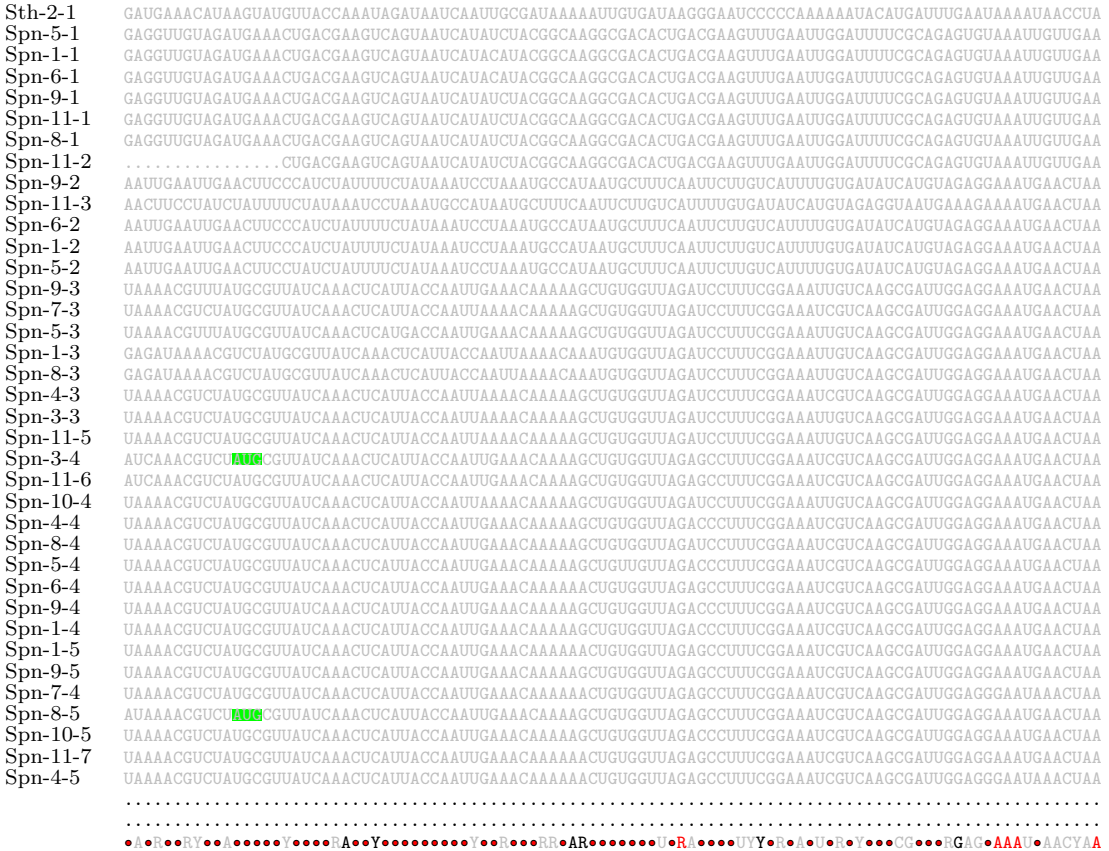
D cd01555 (1) UDP-N-acetylglucosamine enolpyruvyl transferase catalyzes enolpyruvyl transfer as part of the first step in the biosynthesis of peptidoglycan, a component of the bacterial cell wall.
D COG0019 (11) Diaminopimelate decarboxylase [Amino acid transport and metabolism]
D COG0166 (3) Glucose-6-phosphate isomerase [Carbohydrate transport and metabolism]
D COG0388 (11) Predicted amidohydrolase [General function prediction only]
D COG0421 (11) Spermidine synthase [Amino acid transport and metabolism]
D COG0566 (12) rRNA methylases [Translation, ribosomal structure and biogenesis]
D COG1132 (11) ABC-type multidrug transport system, ATPase and permease components [Defense mechanisms]
D COG2071 (3) Predicted glutamine amidotransferases [General function prediction only]
D COG2826 (1) Transposase and inactivated derivatives, IS30 family [DNA replication, recombination, and repair]
D COG2957 (11) Peptidylarginine deiminase and related enzymes [Amino acid transport and metabolism]
D COG3250 (11) Beta-galactosidase/beta-glucuronidase [Carbohydrate transport and metabolism]
D COG3583 (9) Uncharacterized protein conserved in bacteria [Function unknown]
D pfam00293 (2) NUDIX domain.
D pfam01327 (12) Polypeptide deformylase.
D pfam02673 (10) Bacitracin resistance protein BacA.
D pfam03435 (12) Saccharopine dehydrogenase.

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Spn-10-1, Spn-10-2, Spn-10-3, Spn-11-4, Spn-2-1, Spn-2-2, Spn-2-3, Spn-2-4, Spn-2-5, Spn-3-1, Spn-3-2, Spn-3-5, Spn-4-1, Spn-4-2, Spn-5-5, Spn-6-3, Spn-6-5, Spn-7-1, Spn-7-2, Spn-7-5, Spn-8-2, Sth-1-1

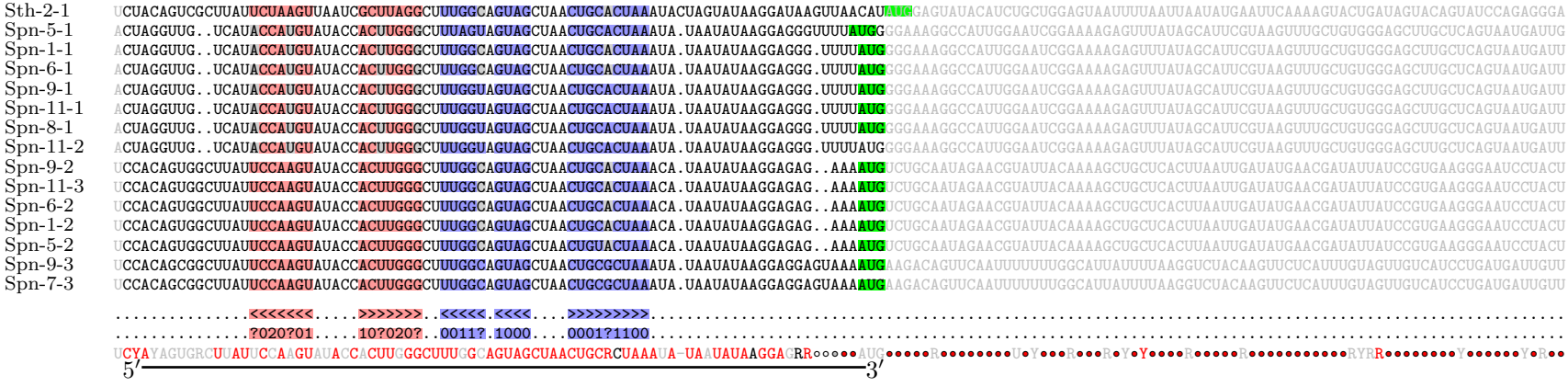
90.4 Multiple-sequence alignment

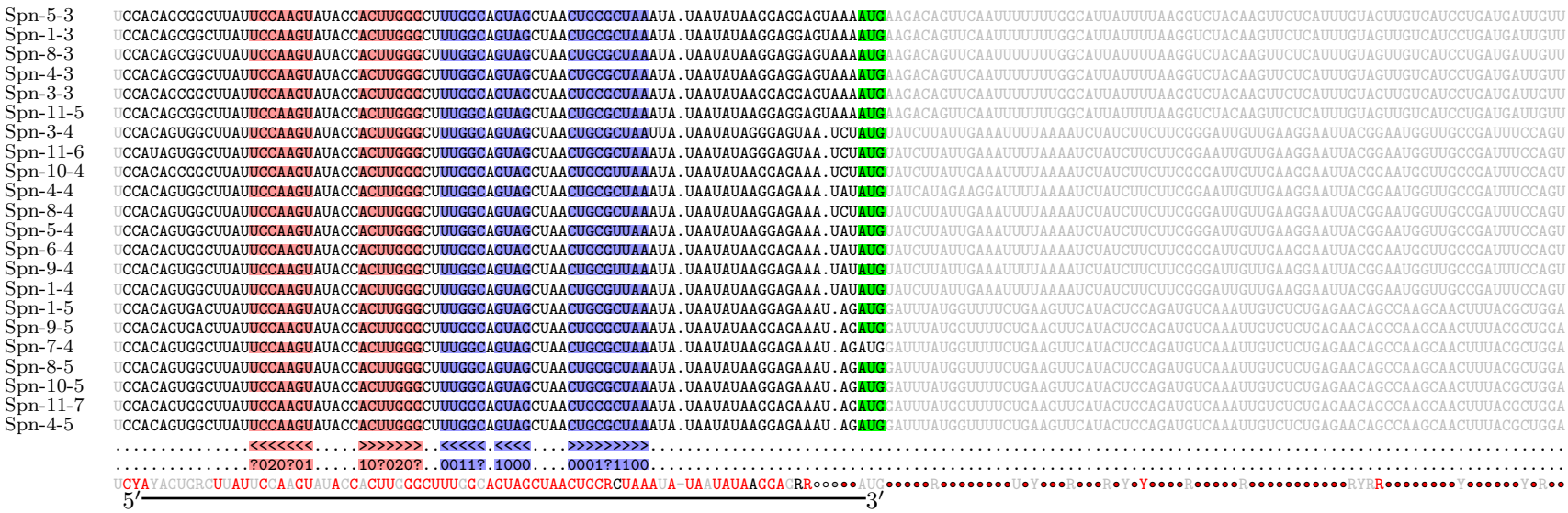
Each ScRE RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 90.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The ScRE RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

alignment positions 1 · · · 99

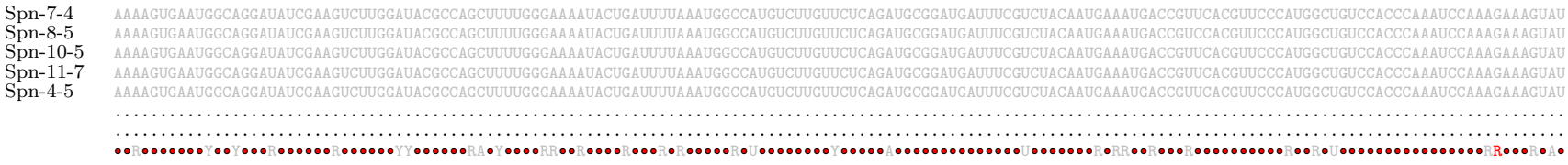


alignment positions 100 · · · 259

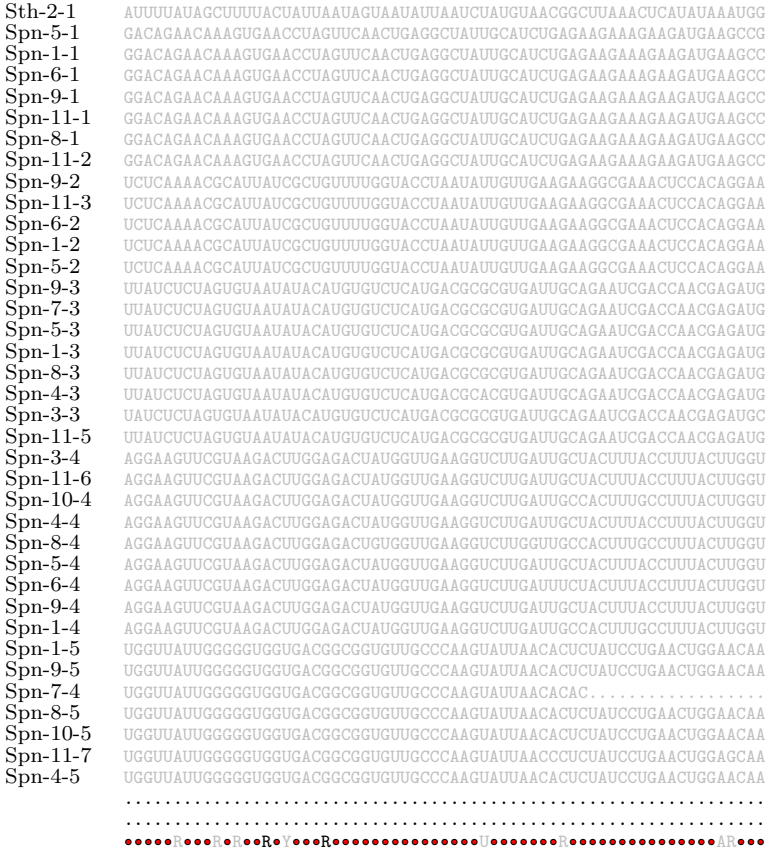




alignment positions 260 · · · 419



alignment positions 420 · · · 484



91 Soil-1

91.1 Taxa

The taxonomy of each organism containing a putative Soil-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Soil-1 RNA in Sections 91.2 and 91.4.

abbrev. of hits	taxonomy of species
env-1 to env-15	environmental samples

91.2 Gene contexts

Each Soil-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Soil-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 91.3. The accession number of the sequence containing each Soil-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 91.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	soil_Contig15146	+	505	554	hypo→ **RNA**→
env-2	soil_XZS40657_b1	+	909	959	**RNA**→
env-3	soil_XZS39717_b1	+	355	405	hypo→ **RNA**→ ←hypo
env-4	soil_Contig20171	-	979	929	PanC (cd00560)→ **RNA**→ Dfp (COG0452)→
env-5	soil_Contig34509	-	654	604	hypo→ **RNA**→ ←hypo
env-6	soil_Contig991	+	792	844	**RNA**→
env-7	soil_XZS80995_x1	+	324	374	**RNA**→ hypo→
env-8	soil_Contig12227	-	115	65	**RNA**→
env-9	soil_XZS48150_g1	-	440	390	←hypo **RNA**→ ←DNA polymerase III, alpha subunit
env-10	soil_XZS2613_g1	+	429	479	**RNA**→ ←hypo
env-11	soil_Contig43	+	223	273	**RNA**→ hypo→
env-12	soil_Contig15146	-	606	557	**RNA**→ ←hypo
env-13	soil_XZS28760_g1	+	643	695	ThrRS_anticodon (cd00860)→ **RNA**→
env-14	soil_XZS88675_x1	+	114	164	**RNA**→ Beta-lactamase class C and other penicillin binding proteins→
env-15	soil_XZS72798_g1	+	499	549	hypo→ **RNA**→ ←hypo

91.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 91.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Soil-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

u cd00560 (1) PanC Pantoate-beta-alanine ligase, also known as pantothenate synthase, catalyzes the formation of pantothenate from pantoate and alanine.
u cd00860 (1) ThrRS Threonyl-anticodon binding domain.

D COG0452 (1) Phosphopantothenoylcysteine synthetase/decarboxylase [Coenzyme metabolism]

91.4 Multiple-sequence alignment

Each Soil-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 91.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Soil-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

92 *Solibacter*-1

92.1 Taxa

The taxonomy of each organism containing a putative *Solibacter*-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *Solibacter*-1 RNA in Sections 92.2 and 92.4.

abbrev. of hits	**taxonomy of species**
Sus-1-1 to Sus-1-34	Bacteria Acidobacteria Solibacteres Solibacterales Solibacteraceae *Solibacter usitatus* Ellin6076

92.2 Gene contexts

Each *Solibacter*-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *Solibacter*-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 92.3. The accession number of the sequence containing each *Solibacter*-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 92.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Sus-1-1	NC_008536.1	+	2330824	2330941	←COG1146 (COG1146) **RNA**→ ←hypo
Sus-1-2	NC_008536.1	-	2151496	2151379	←PqqL (COG0612) **RNA**→ TPR (cd00189)→
Sus-1-3	NC_008536.1	-	3021428	3021311	high-affinity nickel-transporter→ **RNA**→ ←PotD (COG0687)
Sus-1-4	NC_008536.1	-	5777073	5776956	TPR (cd00189)→ **RNA**→ ←hypo
Sus-1-5	NC_008536.1	-	851583	851466	←Aldedh (pfam00171) **RNA**→ ←MviM (COG0673)
Sus-1-6	NC_008536.1	-	4666235	4666118	EFP (pfam01132)→ **RNA**→ ←SelA (pfam03841)
Sus-1-7	NC_008536.1	+	7704774	7704891	ligand_gated_channel (cd01347)→ **RNA**→ ←IolE (COG1082)
Sus-1-8	NC_008536.1	+	9473468	9473585	←ValS (COG0525) **RNA**→ COG1695 (COG1695)→
Sus-1-9	NC_008536.1	+	8809416	8809534	←MviM (COG0673) **RNA**→ TadD (COG5010)→
Sus-1-10	NC_008536.1	-	2362547	2362430	←COG2887 (COG2887)AddB (COG3857) **RNA**→ ←hypo
Sus-1-11	NC_008536.1	-	8429170	8429053	←MgtE (COG2239) **RNA**→ cytochrome C family protein→
Sus-1-12	NC_008536.1	-	1098516	1098399	←AslA (COG3119) **RNA**→ ←TPR (cd00189)
Sus-1-13	NC_008536.1	+	5423626	5423743	←COG3503 (COG3503) **RNA**→ CaiB (COG1804)→
Sus-1-14	NC_008536.1	+	7292050	7292167	YjgP_YjgQ (pfam03739)→ **RNA**→ Beta-lactamase (pfam00144)→
Sus-1-15	NC_008536.1	-	4303293	4303176	vWFA (cd00198)→ **RNA**→ ←Met_dep_hydrolase_A (cd01299)
Sus-1-16	NC_008536.1	+	9133756	9133873	GalE (COG1087)→ **RNA**→ ←Dcp (COG0339)
Sus-1-17	NC_008536.1	+	3435144	3435261	←COG4221 (COG4221) **RNA**→ hypo→
Sus-1-18	NC_008536.1	+	2241345	2241462	←conserved hypothetical protein, transmembrane prediction **RNA**→ ←Glycos_transf_2 (pfam00535)
Sus-1-19	NC_008536.1	-	9067233	9067116	hypo→ **RNA**→ ←ArgE (COG0624)
Sus-1-20	NC_008536.1	-	4096063	4095946	←Gcd (COG4993) **RNA**→ COG3537 (COG3537)→
Sus-1-21	NC_008536.1	+	652598	652715	DltE (COG0300)→ **RNA**→ ←hypo
Sus-1-22	NC_008536.1	-	8250698	8250581	YjgH (cd02198)→ **RNA**→ ←hypo
Sus-1-23	NC_008536.1	-	4379763	4379646	Nei (COG0266)→ **RNA**→ ←COG0412 (COG0412)
Sus-1-24	NC_008536.1	+	1943480	1943597	COG2936 (COG2936)→ **RNA**→ ←hypo
Sus-1-25	NC_008536.1	+	8987142	8987259	Amidase (pfam01425)→ **RNA**→ ←Amidohydro_2 (pfam04909)
Sus-1-26	NC_008536.1	-	5176402	5176285	←Gcd (COG4993) **RNA**→ COG5316 (COG5316)→
Sus-1-27	NC_008536.1	-	1803145	1803028	←GloB (COG0491) **RNA**→ LolA (COG2834)→
Sus-1-28	NC_008536.1	+	5502532	5502649	RfaG (COG0438)COG1216 (COG1216)→ **RNA**→ ←DUF1211 (pfam06736)
Sus-1-29	NC_008536.1	-	9400998	9400881	←IlvA (COG1171) **RNA**→ hypo→
Sus-1-30	NC_008536.1	-	2561850	2561733	DUF72 (pfam01904)→ **RNA**→ ←hypo
Sus-1-31	NC_008536.1	-	7065224	7065109	hypo→ **RNA**→ ←hypo
Sus-1-32	NC_008536.1	-	6073691	6073574	COG3467 (COG3467)→ **RNA**→ hypo→
Sus-1-33	NC_008536.1	-	5102769	5102652	TauA (COG0715)→ **RNA**→ ←FtsX (pfam02687)
Sus-1-34	NC_008536.1	-	2933282	2933165	Peptidase_M20 (pfam01546)→ **RNA**→ ←REC (cd00156)AtoS (COG2202)COG4191 (COG4191)PAS (smart00091)

92.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 92.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *Solibacter*-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

d cd00156 (1) Signal receiver domain; originally thought to be unique to bacteria (CheY, OmpR, NtrC, and PhoB), now recently identified in eukaroytes ETR1 Arabidopsis thaliana; this domain receives the signal from the sensor partner in a two-component systems; contains a phosphoacceptor site that is phosphorylated by histidine kinase homologs; usually found N-terminal to a DNA binding effector domain; forms homodimers

uDd cd00189 (3) Tetratricopeptide repeat domain; typically contains 34 amino acids [WLF]-X(2)-[LIM]-[GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-X(2)-[ASL]-X(4)-[PKE]

u cd00198 (1) Von Willebrand factor type A (vWA) domain was originally found in the blood coagulation protein von Willebrand factor (vWF).

d cd01299 (1) Metallo-dependent hydrolases, subgroup A is part of the superfamily of metallo-dependent hydrolases, a large group of proteins that show conservation in their 3-dimensional fold (TIM barrel) and in details of their active site.

u cd01347 (1) TonB dependent/Ligand-Gated channels are created by a monomeric 22 strand (22,24) anti-parallel beta-barrel.

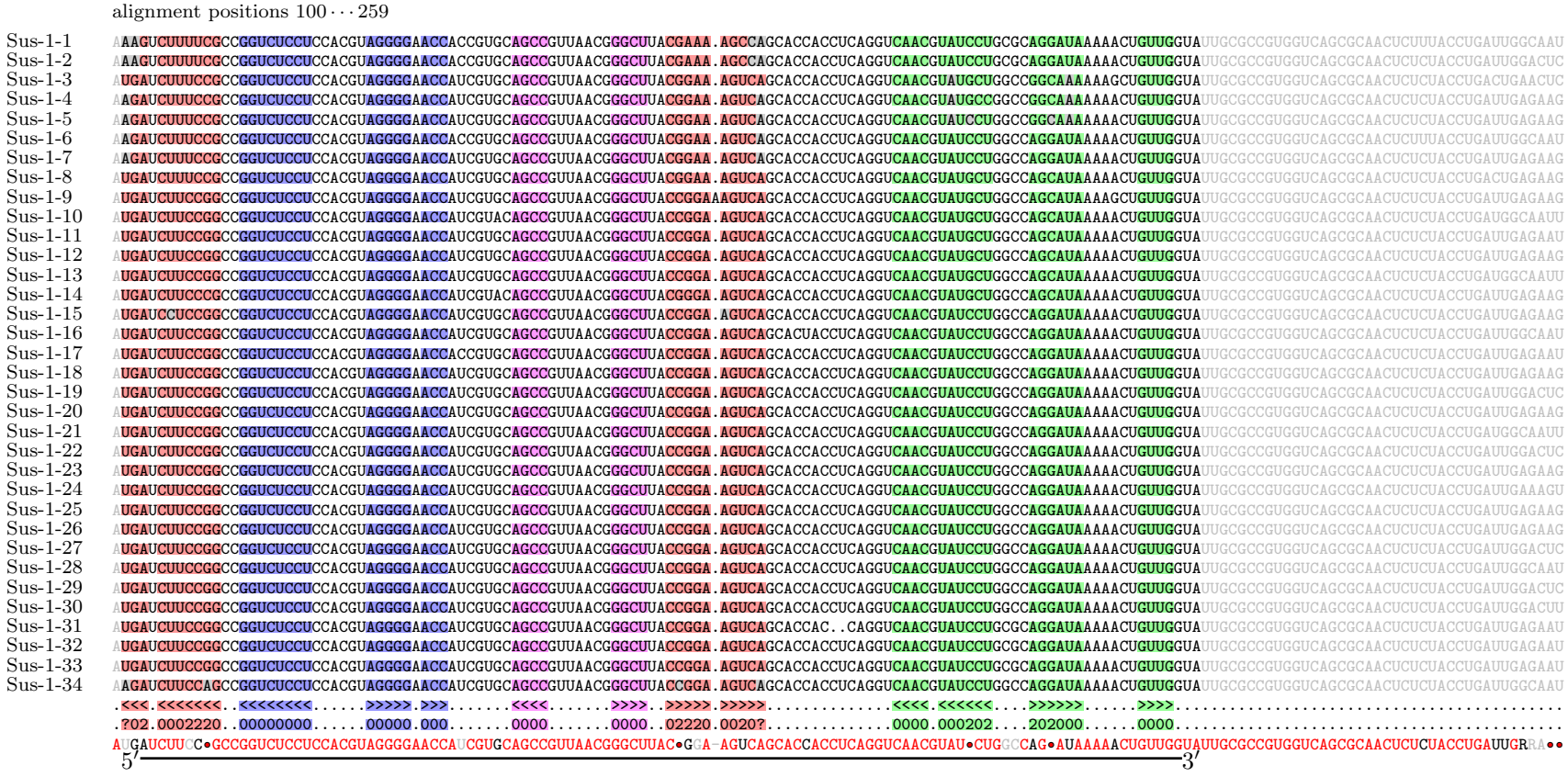
u cd02198 (1) YjgH belong to a large family of proteins present in bacteria, archaea, and eukaryotes with no definitive function.
u COG0266 (1) Formamidopyrimidine-DNA glycosylase [DNA replication, recombination, and repair]
u COG0300 (1) Short-chain dehydrogenases of various substrate specificities [General function prediction only]
d COG0339 (1) Zn-dependent oligopeptidases [Amino acid transport and metabolism]
d COG0412 (1) Dienelactone hydrolase and related enzymes [Secondary metabolites biosynthesis, transport, and catabolism]
u COG0438 (1) Glycosyltransferase [Cell envelope biogenesis, outer membrane]
u COG0491 (1) Zn-dependent hydrolases, including glyoxylases [General function prediction only]
u COG0525 (1) Valyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
u COG0612 (1) Predicted Zn-dependent peptidases [General function prediction only]
d COG0624 (1) Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase and related deacylases [Amino acid transport and metabolism]
u d COG0673 (2) Predicted dehydrogenases and related proteins [General function prediction only]
d COG0687 (1) Spermidine/putrescine-binding periplasmic protein [Amino acid transport and metabolism]
u COG0715 (1) ABC-type nitrate/sulfonate/bicarbonate transport systems, periplasmic components [Inorganic ion transport and metabolism]
d COG1082 (1) Sugar phosphate isomerases/epimerases [Carbohydrate transport and metabolism]
u COG1087 (1) UDP-glucose 4-epimerase [Cell envelope biogenesis, outer membrane]
u COG1146 (1) Ferredoxin [Energy production and conversion]
u COG1171 (1) Threonine dehydratase [Amino acid transport and metabolism]
u COG1216 (1) Predicted glycosyltransferases [General function prediction only]
D COG1695 (1) Predicted transcriptional regulators [Transcription]
D COG1804 (1) Predicted acyl-CoA transferases/carnitine dehydratase [Energy production and conversion]
d COG2202 (1) FOG: PAS/PAC domain [Signal transduction mechanisms]
u COG2239 (1) Mg/Co/Ni transporter MgtE (contains CBS domain) [Inorganic ion transport and metabolism]
D COG2834 (1) Outer membrane lipoprotein-sorting protein [Cell envelope biogenesis, outer membrane]
u COG2887 (1) RecB family exonuclease [DNA replication, recombination, and repair]
u COG2936 (1) Predicted acyl esterases [General function prediction only]
u COG3119 (1) Arylsulfatase A and related enzymes [Inorganic ion transport and metabolism]
u COG3467 (1) Predicted flavin-nucleotide-binding protein [General function prediction only]
u COG3503 (1) Predicted membrane protein [Function unknown]
D COG3537 (1) Putative alpha-1,2-mannosidase [Carbohydrate transport and metabolism]
u COG3857 (1) ATP-dependent nuclease, subunit B [DNA replication, recombination, and repair]
d COG4191 (1) Signal transduction histidine kinase regulating C4-dicarboxylate transport system [Signal transduction mechanisms]
u COG4221 (1) Short-chain alcohol dehydrogenase of unknown specificity [General function prediction only]
u COG4993 (2) Glucose dehydrogenase [Carbohydrate transport and metabolism]
D COG5010 (1) Flp pilus assembly protein TadD, contains TPR repeats [Intracellular trafficking and secretion]
D COG5316 (1) Uncharacterized conserved protein [Function unknown]
D pfam00144 (1) Beta-lactamase.
u pfam00171 (1) Aldehyde dehydrogenase family.
d pfam00535 (1) Glycosyl transferase.
u pfam01132 (1) Elongation factor P (EF-P).
u pfam01425 (1) Amidase.
u pfam01546 (1) Peptidase family M20/M25/M40.
u pfam01904 (1) Protein of unknown function DUF72.
d pfam02687 (1) Predicted permease.
u pfam03739 (1) Predicted permease YjgP/YjgQ family.
d pfam03841 (1) L-seryl-tRNA selenium transferase.
d pfam04909 (1) Amidohydrolase.
d pfam06736 (1) Protein of unknown function (DUF1211).
d smart00091 (1) PAS domain; PAS motifs appear in archaea, eubacteria and eukarya.

92.4 Multiple-sequence alignment

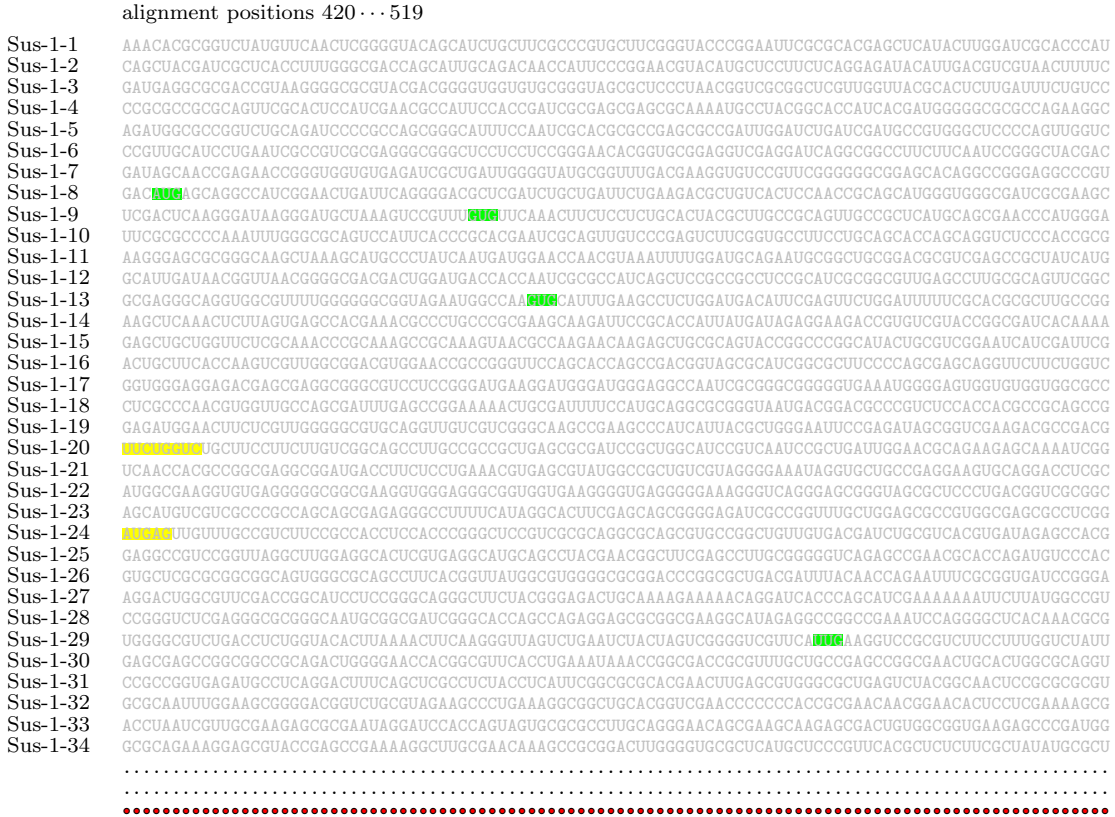
Each *Solibacter*-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 92.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *Solibacter*-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

alignment positions 1 · · · 99

```
Sus-1-1   UCGAUCGUCCCCUCGUUUCGCUUGCAAGUCUCUGAAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUAUUAUCGGUUCGAACCG
Sus-1-2   UCGAUCGUCCCCUCGUUUCGCUUGCAAGUCUCUGAAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUAUUAUCGGUUCGAACCG
Sus-1-3   UCGAUCGUCCCCCUUGUUUCGCGUAGCAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUAUUAUCGGUUCGAACCG
Sus-1-4   UCGAUCGUCCCCCUUGUUUCGCGUAGCAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUGUUAUCGGUUCGAACCG
Sus-1-5   UCGAUCGUCCCCCUUGUUUCGCGUAACAGUCUCUGUAUCCAGGAGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUGUUAUCGGUUCGAACCG
Sus-1-6   UCGAUCGUCCCCCUUGUUUCGCGUAGCAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUAUUAUCGGUUCGAACCG
Sus-1-7   UCGAUCGUCCCCCUUGUUUCGCGUAGCAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUAUUAUCGGUUCGAACCG
Sus-1-8   UCGAUCGUCCCCCUUGUUUCGCGUAGCAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUAUUGUCGGUUCGAACCG
Sus-1-9   UCGAUCGUCCCCCUUGUUUCGCGUAGCAGUCUCUGAAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUGUUAUCGGUUCGAACCG
Sus-1-10  UCGAUCGUCCCCCUUGUUUCGCGUAGCAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUGUUAUCGGUUCGAACCG
Sus-1-11  UCGAUCGUCCCCCUUGUUUCGCGUAGCAGUCUCUGAAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUAUUAUCGGUUCGAACCG
Sus-1-12  UCGAUCGUCCCCCUUGUUUCGCGUAGCAGCCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUGUUAUCGGUUCGAACCG
Sus-1-13  UCGAUCGUCCCCCUUGUUUCGCGUAGCAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUAUUAUCGGUUCGAACCG
Sus-1-14  UCGAUCGUCCCCCUUGUUUCGCGUAACAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUGUUAUCGGUUCGAACCG
Sus-1-15  UCGAUCGUCCCCCUUGUUUCGCGUAGCAGCCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUGUUAUCGGUUCGAACCG
Sus-1-16  UCGAUCGUCCCCCUUGUUUCGCGUAGCAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUGUUAUCGGUUCGAACCG
Sus-1-17  UCGAUCGUCCCCCUUGUUUCGCGUAGCAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUGUUAUCGGUUCGAACCG
Sus-1-18  UCGAUCGUCCCCCUUGUUUCGCGUAGCAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUGUUAUCGGUUCGAACCG
Sus-1-19  UCGAUCGUCCCCCUUGUUUCGCGUGUCAGCCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUAUUAUCGGUUCGAACCG
Sus-1-20  UCGAUCGUCCCCCUUGUUUCGCGUAGCAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUAUUAUCGGUUCGAACCG
Sus-1-21  UCGAUCGUCCCCCUUGUUUCGCGUAGCAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUAUUAUCGGUUCGAACCG
Sus-1-22  UCGAUCGUCCCCCUUGUUUCGCGUAGCAGCCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUGUUAUCGGUUCGAACCG
Sus-1-23  UCGAUCGUCCCCCUUGUUUCGCGUAGCAGUCUCUGAAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUCUUAUCGGUUCGAACCG
Sus-1-24  CGAUCGUCCCCCUUGUUUCGCGUAAGCAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUAUUAUCGGUUCGAACCG
Sus-1-25  UCGAUCGUCCCCCUUGUUUCGCGUAGCAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUGUUAUCGGUUCGAACCG
Sus-1-26  UCGAUCGUCCCCCUUGUUUCGCGUAGCAGCCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUAUUAUCGGUUCGAACCG
Sus-1-27  UCGAUCGUCCCCCUUGUUUCGCGUAGCAGCCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACAGUCCCCCGGGUGUUAUCGGUUCGAACCG
Sus-1-28  UCGAUCGUCCCCCUUGUUUCGCGUAGCAAUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUGUUAUCGGUUCGAACCG
Sus-1-29  UCGAUCGUCCCCCUUGUUUCGCGUAGCAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUGUUAUCGGUUCGAACCG
Sus-1-30  UCGAUCGUCCCCUUGUUUCGCGUAGCAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUAUUAUCGGUUCGAACCG
Sus-1-31  UCGAUCGUCCCCCUUGUUUCGCGUAACAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUAUUAUCGGUUCGAACCG
Sus-1-32  UCGAUCGUCCCCCUUGUUUCGCGUAGCAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUAUUAUCGGUUCGAACCG
Sus-1-33  UCGAUCGUCCCCCUUGUUUCGCGUAGCAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUGUUAUCGGUUCGAACCG
Sus-1-34  UCGAUCGUCCCCUUGUUUCGCGUAGCAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGUAUUAUCGGUUCGAACCG
          ....................................................................................................
          ....................................................................................................
          UCGAUCGUCCCCUUGUUUCGCGUAGCAGUCUCUGUAUCCAGGGGUGCGGCAAUCAAACGCUAUCCGCGACGGUCCCCCGGGURUUAUCGGUUCGAACCG
```


alignment positions 100 · · · 259
Sus-1-1
Sus-1-2
Sus-1-3
Sus-1-4
Sus-1-5
Sus-1-6
Sus-1-7
Sus-1-8
Sus-1-9
Sus-1-10
Sus-1-11
Sus-1-12
Sus-1-13
Sus-1-14
Sus-1-15
Sus-1-16
Sus-1-17
Sus-1-18
Sus-1-19
Sus-1-20
Sus-1-21
Sus-1-22
Sus-1-23
Sus-1-24
Sus-1-25
Sus-1-26
Sus-1-27
Sus-1-28
Sus-1-29
Sus-1-30
Sus-1-31
Sus-1-32
Sus-1-33
Sus-1-34
5′
3′

alignment positions 260 · · · 419



93 STAXI

93.1 Taxa

The taxonomy of each organism containing a putative STAXI RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual STAXI RNA in Sections 93.2 and 93.4.

abbrev. of hits	taxonomy of species
Eco-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* E110019
Eco-2-1 to Eco-2-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 EDL933
Eco-3-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4024
Eco-4-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4042
Eco-5-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4045
Eco-6-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4076
Eco-7-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4113
Eco-8-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4115
Eco-9-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4196
Eco-10-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4206
Eco-11-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4401

Eco-12-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4486
Eco-13-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC4501
Eco-14-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC508
Eco-15-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. EC869
Eco-16-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. Sakai
Eco-17-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. TW14588
Pat-1-1 to Pat-1-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Pectobacterium atrosepticum* SCRI1043
Plu-1-1 to Plu-1-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Photorhabdus luminescens subsp. laumondii* TTO1
Pmi-1-1 to Pmi-1-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Proteus mirabilis* HI4320
Sen-1-1 to Sen-1-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Paratyphi B *Salmonella enterica subsp*
Sen-2-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*
Sen-3-1 to Sen-3-4	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*
Sen-4-1 to Sen-4-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*
Sen-5-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*
Sen-6-1 to Sen-6-4	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*
Sen-7-1 to Sen-7-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*
Sen-8-1 to Sen-8-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*
Sen-9-1 to Sen-9-4	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*
Sen-10-1 to Sen-10-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*
Sen-11-1 to Sen-11-4	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Typhi *Salmonella enterica subsp. ente*
Sen-12-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Virchow *Salmonella enterica subsp. en*
Spr-1-1 to Spr-1-8	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Serratia proteamaculans* 568
Yen-1-1 to Yen-1-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia enterocolitica subsp. enterocolitica* 8081
Yin-1-1 to Yin-1-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia intermedia* ATCC 29909
Yps-1-1 to Yps-1-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pseudotuberculosis* IP 31758
env-1 to env-2	environmental samples

93.2 Gene contexts

Each STAXI RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one STAXI RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 93.3. The accession number of the sequence containing each STAXI RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 93.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Sen-12-1	NZ_ABFH02000001.1	+	1901911	1902034	←hypo **RNA**→ ←hypo
Sen-5-1	NZ_CAAS01001077.1	+	1629	1752	**RNA**→
Sen-2-1	NZ_CAAY01001733.1	+	729	852	**RNA**→
Sen-11-1	NC_004631.1	+	4520498	4520621	←hypo **RNA**→ ←hypo
Sen-3-1	NC_003198.1	+	4537689	4537812	←hypo **RNA**→ ←hypo
Sen-4-1	NZ_CAAR01000001.1	+	10605	10728	**RNA**→
Sen-6-1	NZ_CAAT01000371.1	-	2032	1909	**RNA**→
Sen-7-1	NZ_CAAU01001295.1	-	1948	1825	**RNA**→
Sen-8-1	NZ_CAAV01000293.1	+	1911	2034	**RNA**→
Sen-9-1	NZ_CAAW01000272.1	+	5160	5283	**RNA**→
Sen-10-1	NZ_CAAX01000510.1	+	7169	7292	**RNA**→
Yin-1-1	NZ_AALF01000047.1	+	31503	31609	hypo→ **RNA**→
Yin-1-2	NZ_AALF01000086.1	+	136	243	**RNA**→ hypo→
env-1	BAAW01011162.1	-	362	187	**RNA**→ hypo→

Pat-1-1	NC_004547.2	+	602159	602313	←hypo **RNA**→ hypo→
Pat-1-2	NC_004547.2	+	648593	648750	hypo→ **RNA**→ hypo→
Pat-1-3	NC_004547.2	+	650861	651015	←putative plasmid-related protein **RNA**→ hypo→
Spr-1-1	NC_009832.1	-	4503612	4503435	hypo→ **RNA**→ hypo→
Spr-1-2	NC_009832.1	-	4501549	4501373	hypo→ **RNA**→ hypo→
Spr-1-3	NC_009832.1	-	4560724	4560545	TopA (COG0550)→ **RNA**→ hypo→
Pmi-1-1	NC_010554.1	-	2805661	2805504	hypo→ **RNA**→ ←hypo
Pmi-1-2	NC_010554.1	-	2803227	2803068	hypo→ **RNA**→ hypo→
Pmi-1-3	NC_010554.1	-	2873824	2873664	TopA (COG0550)zf-C4_Topoisom (pfam01396)→ **RNA**→ hypo→
Spr-1-4	NC_009832.1	+	440732	440887	hypo→ **RNA**→ ←hypo
Spr-1-5	NC_009832.1	+	366915	367070	TopA (COG0550)→ **RNA**→ hypo→
Spr-1-6	NC_009832.1	-	4581610	4581456	←HTH_XRE (cd00093) **RNA**→ hypo→
Spr-1-7	NC_009832.1	-	4584247	4584093	hypo→ **RNA**→ hypo→
Spr-1-8	NC_009832.1	+	437613	437768	hypo→ **RNA**→ INT_XerDC (cd00798)→
env-2	BABG01007993.1	+	655	806	**RNA**→
Eco-16-1	NC_002695.1	+	1399050	1399201	←hypo **RNA**→ hypo→
Eco-8-1	NC_011353.1	+	1342636	1342787	←hypo **RNA**→ hypo→
Eco-2-1	NC_002655.2	+	1087216	1087367	←hypo **RNA**→ ←hypo
Eco-2-2	NC_002655.2	+	1482823	1482974	←hypo **RNA**→ ←hypo
Eco-1-1	NZ_AAJW02000034.1	-	31791	31640	←hypo (cons) **RNA**→ hypo→
Eco-10-1	NZ_ABHK02000001.1	+	4190518	4190669	←hypo (cons) **RNA**→ hypo→
Eco-5-1	NZ_ABHL02000002.1	+	1034870	1035021	←hypo (cons) **RNA**→ hypo→
Eco-4-1	NZ_ABHM02000002.1	+	475079	475230	←hypo (cons) **RNA**→ hypo→
Eco-9-1	NZ_ABHO01000005.1	+	108854	109005	←hypo (cons) **RNA**→ hypo→
Eco-6-1	NZ_ABHQ01000007.1	+	191155	191306	←hypo (cons) **RNA**→ hypo→
Eco-11-1	NZ_ABHR01000003.1	+	166182	166333	←hypo (cons) **RNA**→ hypo→
Eco-12-1	NZ_ABHS01000005.1	+	181302	181453	←hypo (cons) **RNA**→ hypo→
Eco-13-1	NZ_ABHT01000005.1	-	74646	74495	←hypo (cons) **RNA**→ hypo→
Eco-15-1	NZ_ABHU01000017.1	+	68213	68364	←hypo (cons) **RNA**→ hypo→
Eco-14-1	NZ_ABHW01000010.1	+	74936	75087	←hypo (cons) **RNA**→ hypo→
Eco-3-1	NZ_ABJT01000002.1	+	189401	189552	←hypo **RNA**→ ←COG3436 (COG3436)
Eco-17-1	NZ_ABKY01000008.1	-	177749	177598	←hypo **RNA**→ ←COG3436 (COG3436)
Eco-7-1	NZ_ABHP01000002.1	-	105833	105682	←hypo (cons) **RNA**→ hypo→
Sen-4-2	NZ_CAAR01000141.1	-	205	49	**RNA**→
Sen-10-2	NZ_CAAX01000002.1	+	18122	18278	**RNA**→
Sen-11-2	NC_004631.1	+	4448069	4448225	INT_XerDC (cd00798)XerD (COG4974)→ **RNA**→ hypo→
Sen-3-2	NC_003198.1	+	4464476	4464632	INT_XerDC (cd00798)XerD (COG4974)→ **RNA**→ hypo→
Sen-6-2	NZ_CAAT01000318.1	-	9701	9545	**RNA**→
Sen-9-2	NZ_CAAW01001007.1	+	6567	6723	**RNA**→
Sen-11-3	NC_004631.1	+	4405724	4405878	TopA (COG0550)zf-C4_Topoisom (pfam01396)→ **RNA**→ hypo→
Sen-3-3	NC_003198.1	+	4421074	4421228	TopA (COG0550)zf-C4_Topoisom (pfam01396)→ **RNA**→ hypo→
Sen-6-3	NZ_CAAT01000366.1	+	4733	4887	**RNA**→
Sen-7-2	NZ_CAAU01000131.1	-	1514	1360	**RNA**→
Sen-9-3	NZ_CAAW01000133.1	+	1718	1872	**RNA**→
Plu-1-1	NC_005126.1	+	1220999	1221153	TopA (COG0550)zf-C4_Topoisom (pfam01396)→ **RNA**→ hypo→
Plu-1-2	NC_005126.1	+	1338968	1339125	←FhaC (COG2831) **RNA**→ hypo→
Plu-1-3	NC_005126.1	+	1340745	1340904	hypo→ **RNA**→ hypo→
Sen-8-2	NZ_CAAV01003008.1	-	2147	1986	**RNA**→
Sen-9-4	NZ_CAAW01001007.1	+	4425	4586	**RNA**→
Sen-11-4	NC_004631.1	+	4445927	4446088	←DUF1062 (pfam06353) **RNA**→ INT_XerDC (cd00798)XerD (COG4974)→
Sen-3-4	NC_003198.1	+	4462334	4462495	←DUF1062 (pfam06353) **RNA**→ INT_XerDC (cd00798)XerD (COG4974)→
Sen-6-4	NZ_CAAT01000318.1	-	11843	11682	**RNA**→

Sen-7-3	NZ_CAAU01003246.1	+	549	710	**RNA**→
Sen-10-3	NZ_CAAX01000002.1	+	15980	16141	**RNA**→
Sen-1-1	NC_010102.1	+	4503610	4503764	←hypo **RNA**→ hypo→
Sen-1-2	NC_010102.1	+	4566669	4566826	hypo→ **RNA**→ hypo→
Yps-1-1	NC_009708.1	-	4162859	4162704	←COG5435 (COG5435) **RNA**→ COG4227 (COG4227)→
Yps-1-2	NC_009708.1	-	4214684	4214528	hypo→ **RNA**→ hypo→
Yen-1-1	NC_008800.1	-	3821315	3821159	hypo→ **RNA**→ hypo→
Yen-1-2	NC_008800.1	-	3772495	3772339	putative phage integrase→ **RNA**→ COG4227 (COG4227)→
Sen-1-3	NC_010102.1	+	4564505	4564663	hypo→ **RNA**→ hypo→
Yps-1-3	NC_009708.1	-	4172667	4172518	←hypo **RNA**→ INT_XerDC (cd00798)→
Yen-1-3	NC_008800.1	-	3774300	3774141	←hypo **RNA**→ putative phage integrase→

93.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 93.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one STAXI RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

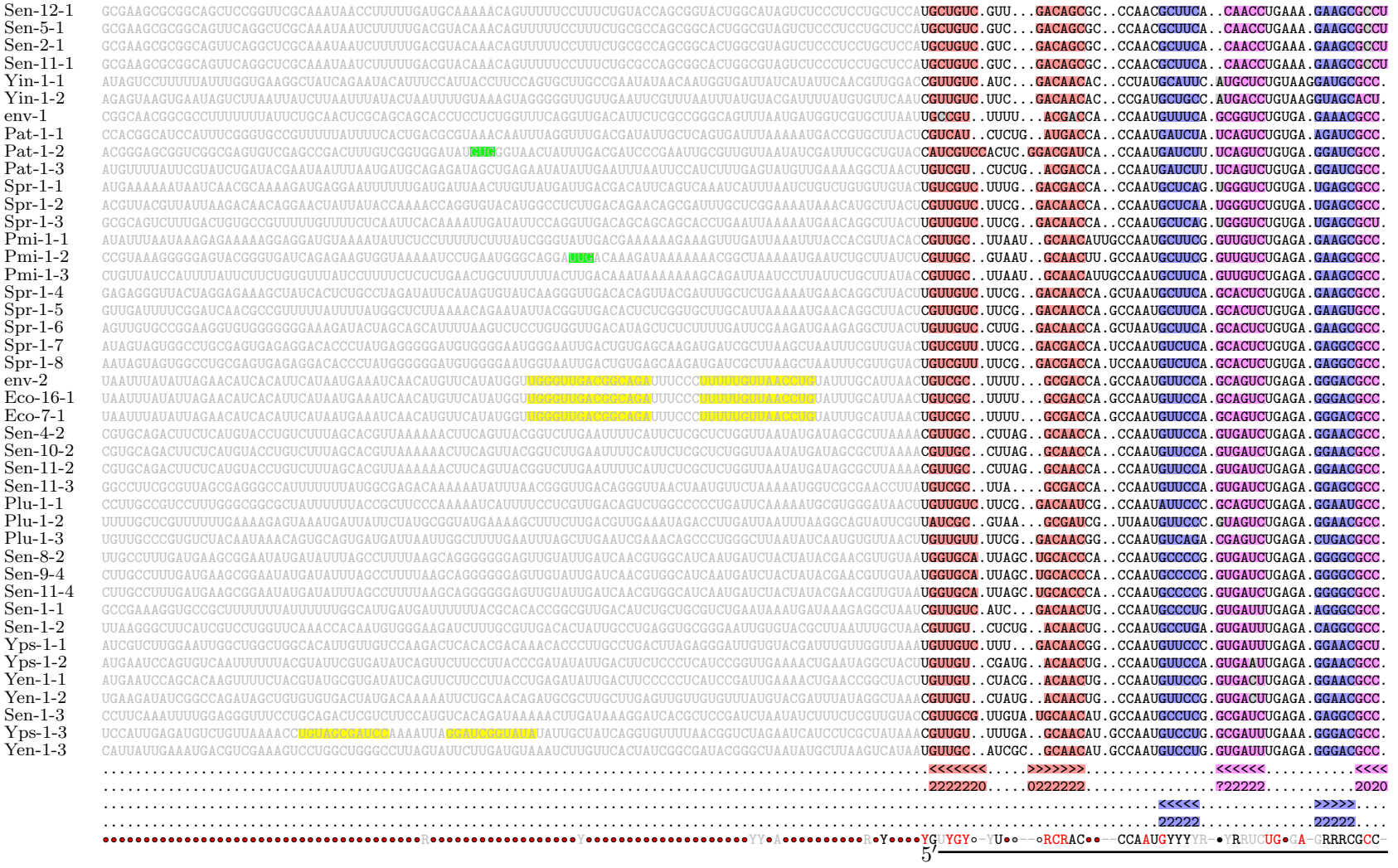
u cd00093 (1) Helix-turn-helix XRE-family like proteins.
uD cd00798 (6) XerD and XerC integrases, DNA breaking-rejoining enzymes, N- and C-terminal domains.
u COG0550 (6) Topoisomerase IA [DNA replication, recombination, and repair]
u COG2831 (1) Hemolysin activation/secretion protein [Intracellular trafficking and secretion]
d COG3436 (2) Transposase and inactivated derivatives [DNA replication, recombination, and repair]
D COG4227 (2) Antirestriction protein [DNA replication, recombination, and repair]
uD COG4974 (4) Site-specific recombinase XerD [DNA replication, recombination, and repair]
u COG5435 (1) Uncharacterized conserved protein [Function unknown]
u pfam01396 (4) Topoisomerase DNA binding C4 zinc finger.
u pfam06353 (2) Protein of unknown function (DUF1062).

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Eco-1-1, Eco-10-1, Eco-11-1, Eco-12-1, Eco-13-1, Eco-14-1, Eco-15-1, Eco-17-1, Eco-2-1, Eco-2-2, Eco-3-1, Eco-4-1, Eco-5-1, Eco-6-1, Eco-8-1, Eco-9-1, Sen-10-1, Sen-10-3, Sen-3-1, Sen-3-2, Sen-3-3, Sen-3-4, Sen-4-1, Sen-6-1, Sen-6-2, Sen-6-3, Sen-6-4, Sen-7-1, Sen-7-2, Sen-7-3, Sen-8-1, Sen-9-1, Sen-9-2, Sen-9-3

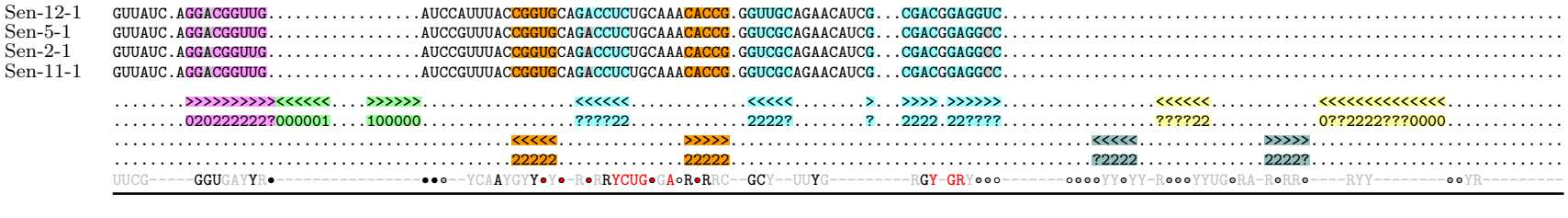
93.4 Multiple-sequence alignment

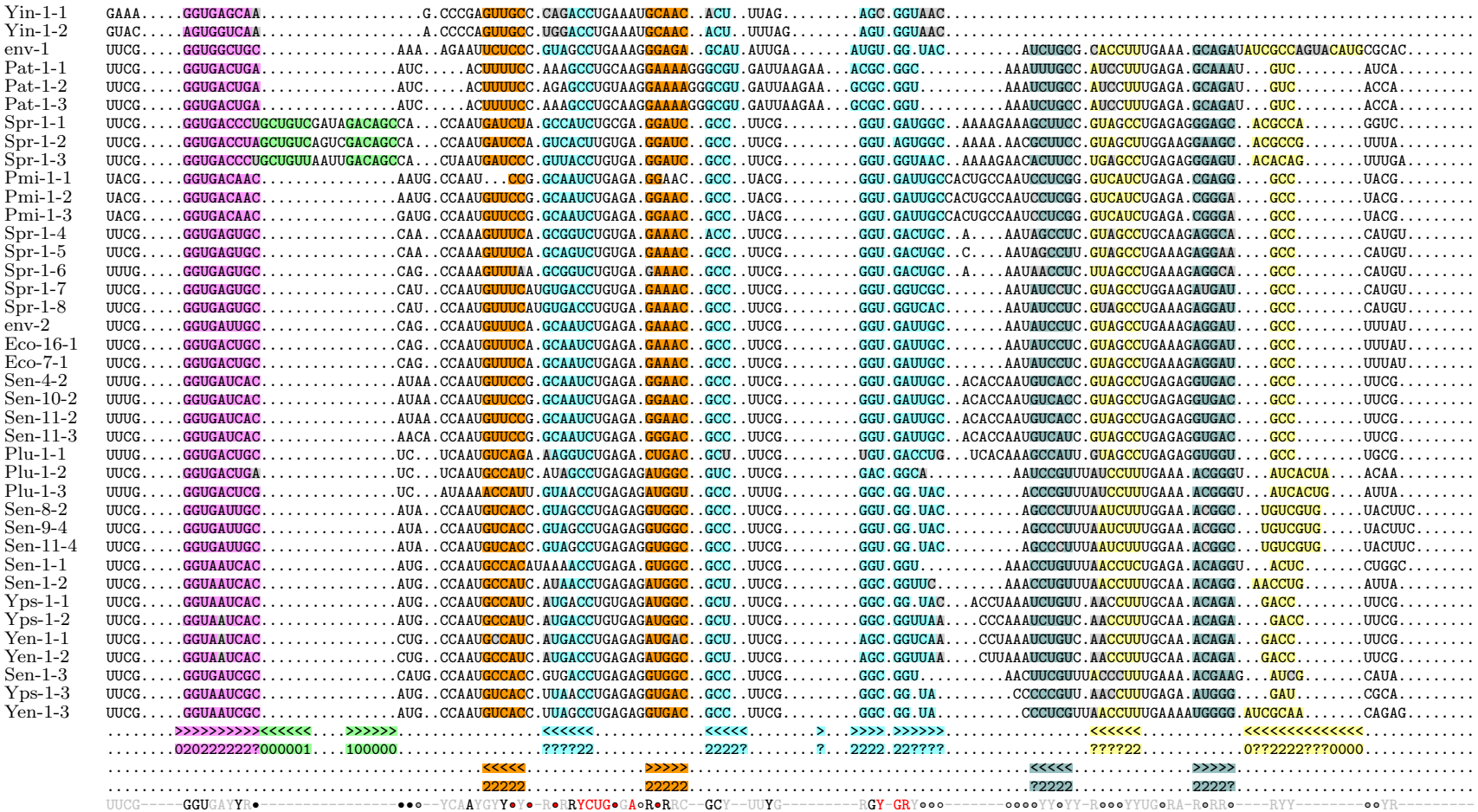
Each STAXI RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 93.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The STAXI RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

alignment positions 1 · · · 157

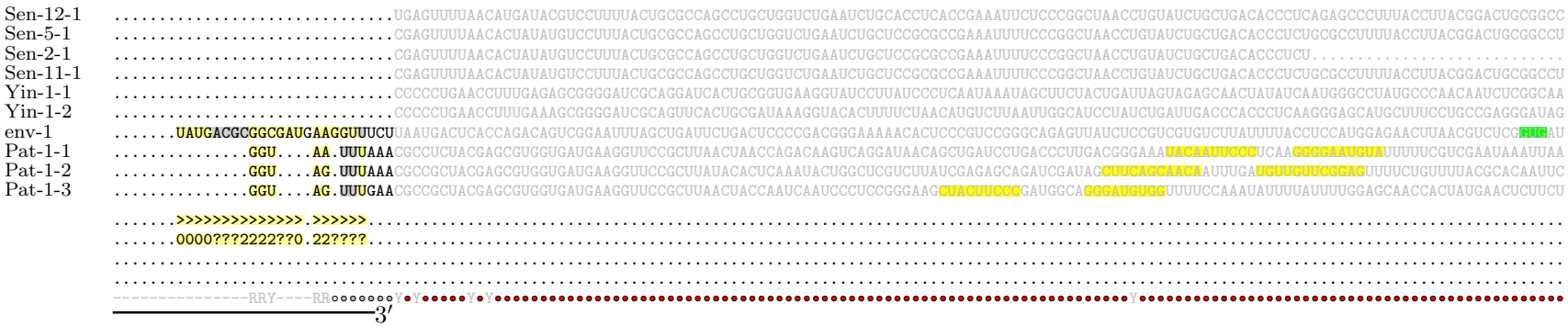


alignment positions 158 · · · 317

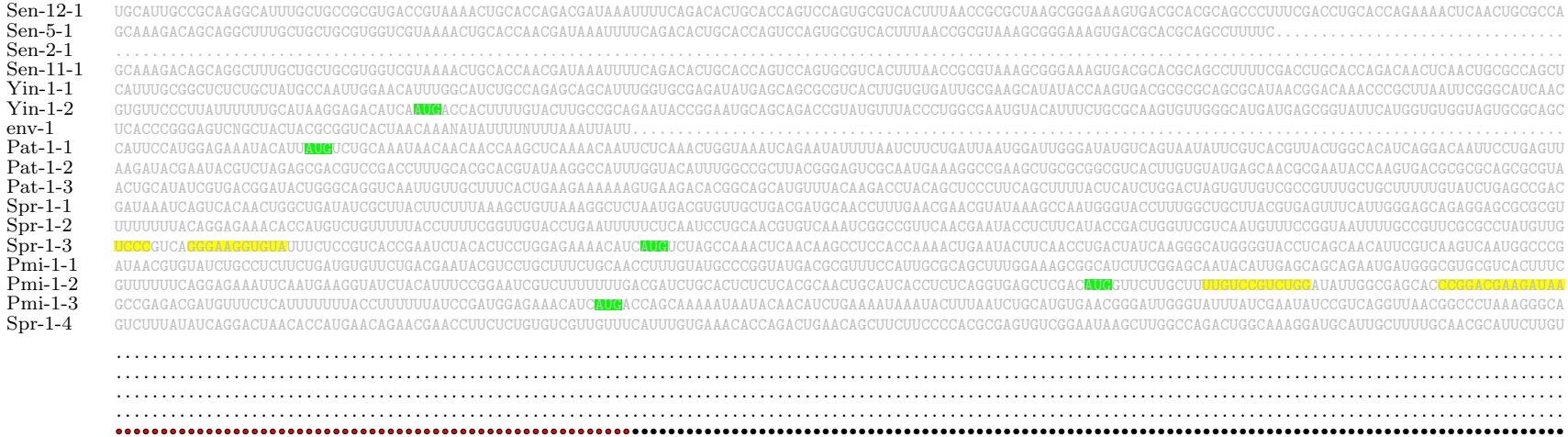


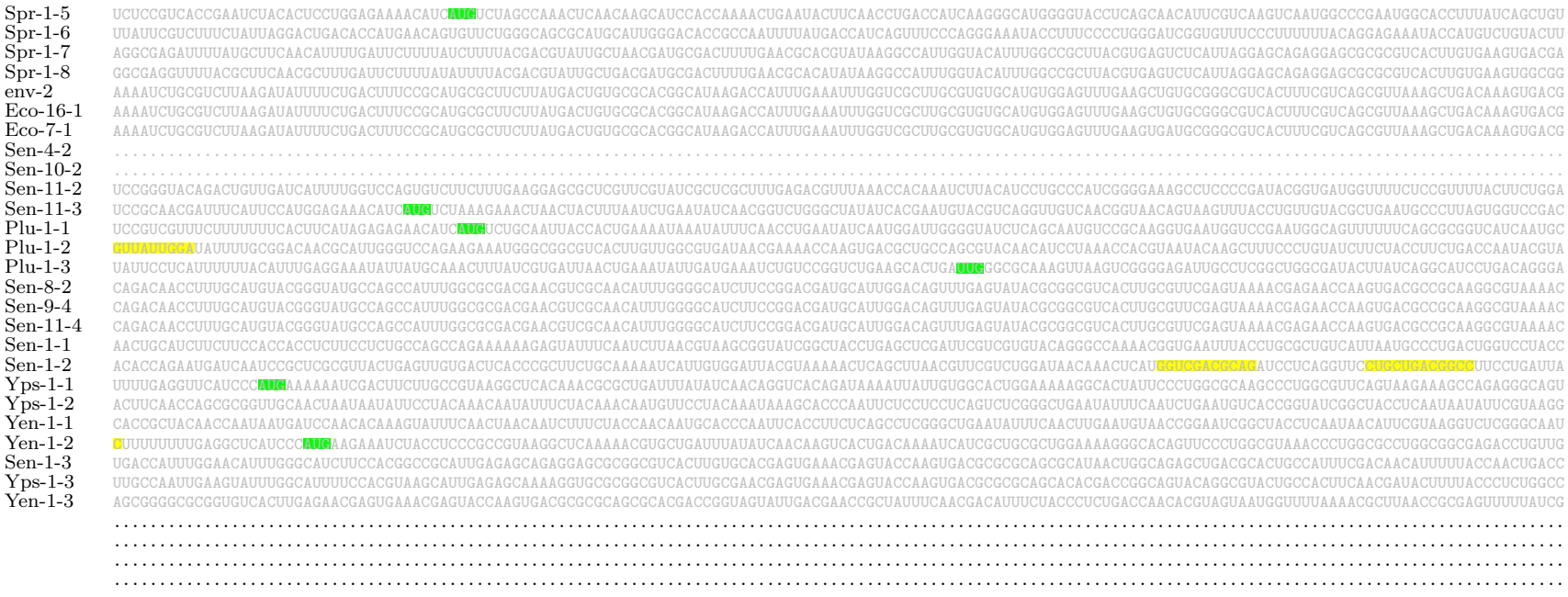


alignment positions 318 · · · 477

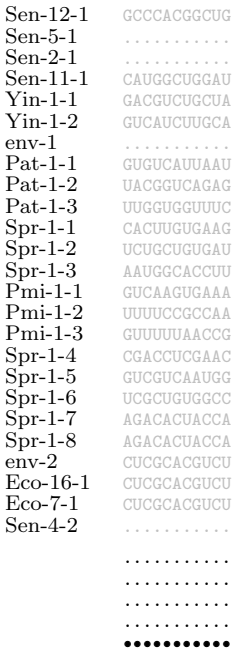


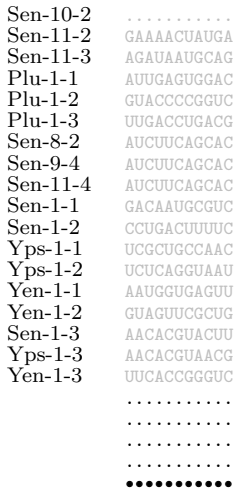
alignment positions 478 · · · 637





alignment positions 638 · · · 648





94 *sucA*-II

94.1 Taxa

The taxonomy of each organism containing a putative *sucA*-II RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *sucA*-II RNA in Sections 94.2 and 94.4.

abbrev. of hits	**taxonomy of species**
Avi-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Azotobacter vinelandii*
Pae-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* 2192
Pae-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* C3719
Pae-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PA7
Pae-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PACS2
Pae-5-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PAO1
Pae-6-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* UCBPP-PA14
Pen-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas entomophila* L48
Pfl-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf-5
Pfl-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf0-1
Pme-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas mendocina ymp*
Ppu-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* F1
Ppu-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* GB-1
Ppu-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* KT2440
Ppu-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* W619
Pst-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas stutzeri* A1501
Psy-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. phaseolicola* 1448A
Psy-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000
env-1 to env-3	environmental samples

94.2 Gene contexts

Each *sucA*-II RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *sucA*-II RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 94.3. The accession number of the sequence containing each *sucA*-II RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 94.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	JCVI_SCAF_1096628054928	+	1330	1614	**RNA**→
Avi-1-1	NZ_AAAU03000006.1	+	163403	163616	**RNA**→ SucA (COG0567)→ AceF (COG0508)→ Lpd (COG1249)→
Pae-1-1	NZ_AAKW01000006.1	+	22649	22862	**RNA**→ SucA (COG0567)→ AceF (COG0508)→ Lpd (COG1249)→
Pae-3-1	NC_009656.1	-	3824583	3824370	**RNA**→ SucA (COG0567)→ AceF (COG0508)→ Lpd (COG1249)→
Pae-5-1	NC_002516.2	+	1724027	1724240	**RNA**→ SucA (COG0567)→ AceF (COG0508)→ Lpd (COG1249)→
Pae-2-1	NZ_AAKV01000011.1	+	22622	22835	**RNA**→ SucA (COG0567)→ AceF (COG0508)→ Lpd (COG1249)→
Pae-4-1	NZ_AAQW01000001.1	+	2284313	2284526	**RNA**→ SucA (COG0567)→ AceF (COG0508)→ Lpd (COG1249)→
Pae-6-1	NC_008463.1	-	3919482	3919269	**RNA**→ SucA (COG0567)→ AceF (COG0508)→ Lpd (COG1249)→
Pen-1-1	NC_008027.1	-	3928150	3927934	**RNA**→ SucA (COG0567)→ AceF (COG0508)→ Lpd (COG1249)→
Ppu-4-1	NZ_AAVY01000007.1	+	142075	142292	**RNA**→ SucA (COG0567)→ AceF (COG0508)→ Lpd (COG1249)→
Ppu-1-1	NC_009512.1	+	1868297	1868514	**RNA**→ SucA (COG0567)→ AceF (COG0508)→ Lpd (COG1249)→
Ppu-3-1	NC_002947.3	-	4735854	4735637	**RNA**→ SucA (COG0567)→ AceF (COG0508)→ Lpd (COG1249)→
Ppu-2-1	NZ_AAXR01000004.1	+	295124	295341	**RNA**→ SucA (COG0567)→ AceF (COG0508)→ Lpd (COG1249)→
Psy-1-1	NC_005773.3	+	2325229	2325446	**RNA**→ SucA (COG0567)→ AceF (COG0508)→
Psy-3-1	NC_004578.1	+	2419253	2419470	**RNA**→ SucA (COG0567)→ AceF (COG0508)→
Psy-2-1	NC_007005.1	+	2329440	2329657	**RNA**→ SucA (COG0567)→ AceF (COG0508)→
Pfl-1-1	NC_004129.6	+	1921629	1921848	**RNA**→ SucA (COG0567)→ AceF (COG0508)→
Pfl-2-1	NC_007492.1	+	1802825	1803041	**RNA**→ SucA (COG0567)→ AceF (COG0508)→ Lpd (COG1249)→
env-2	JCVI_SCAF_1096627720741	-	447	228	**RNA**→
env-3	JCVI_SCAF_1096626432556	-	595	376	**RNA**→ ? (GOS2186)→
Pme-1-1	NC_009439.1	-	2769676	2769456	**RNA**→ SucA (COG0567)→ AceF (COG0508)→ Lpd (COG1249)→
Pst-1-1	NC_009434.1	+	2022862	2023085	**RNA**→ SucA (COG0567)→ AceF (COG0508)→ Lpd (COG1249)→

94.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 94.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *sucA*-II RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

COG0508 (19) Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component, and related enzymes [Energy production and conversion]

COG0567 (19) 2-oxoglutarate dehydrogenase complex, dehydrogenase (E1) component, and related enzymes [Energy production and conversion]

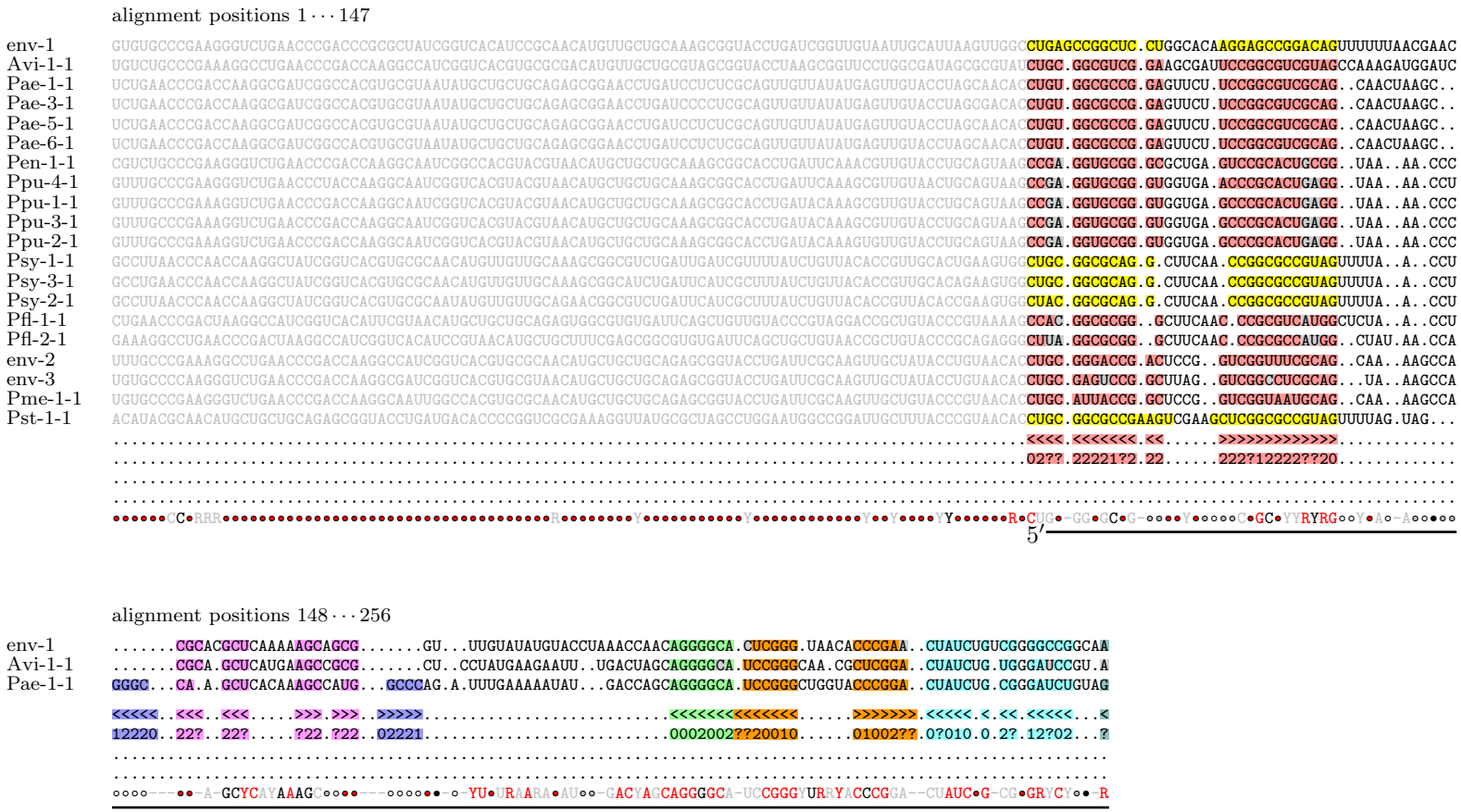
COG1249 (15) Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component, and related enzymes [Energy production and conversion]

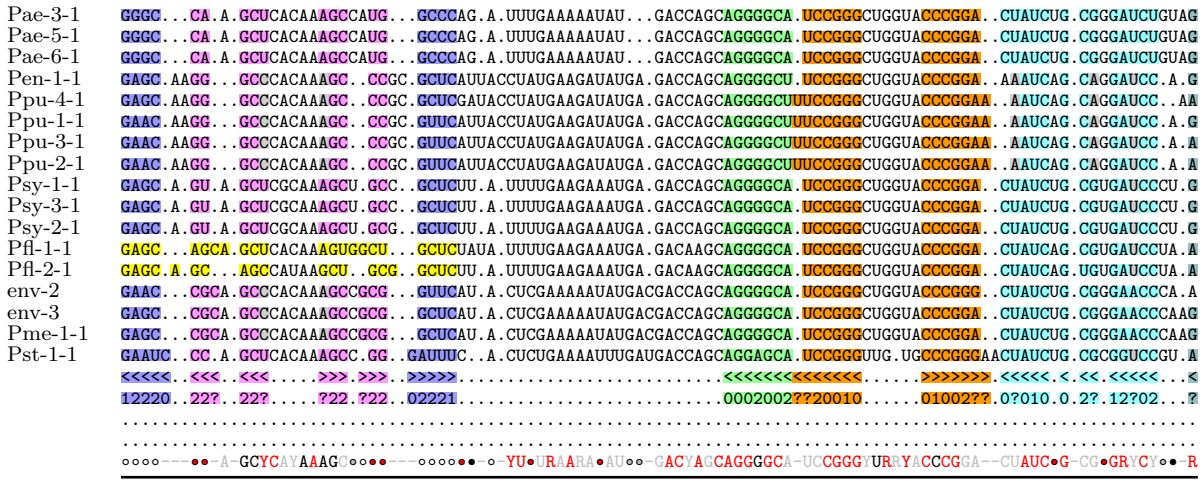
GOS2186 (1) ?

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Pae-2-1, Pae-4-1

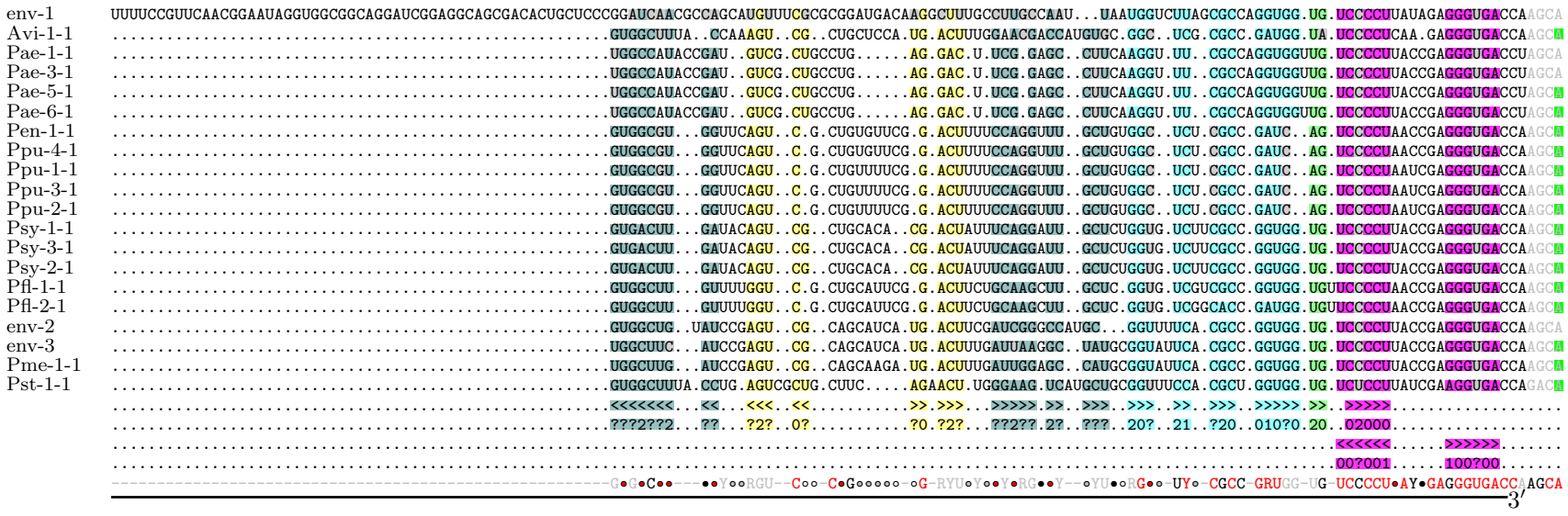
94.4 Multiple-sequence alignment

Each *sucA*-II RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 94.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *sucA*-II RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

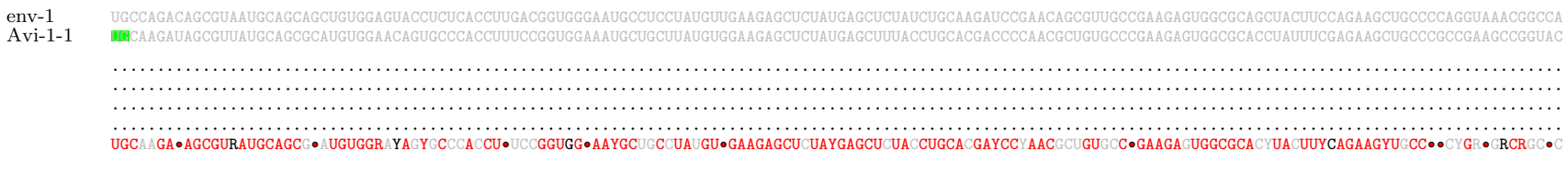




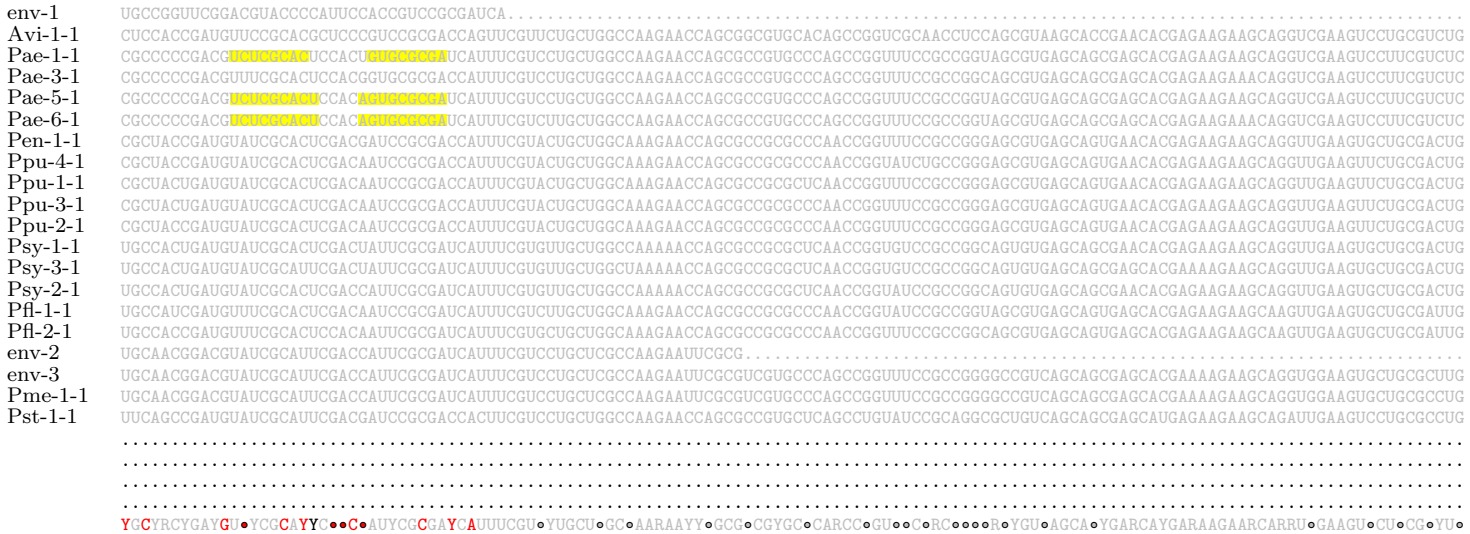
alignment positions 257 · · · 416



alignment positions 417 · · · 576



alignment positions 577 · · · 712



95 *sucC*

95.1 Taxa

The taxonomy of each organism containing a putative *sucC* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *sucC* RNA in Sections 95.3 and 95.5.

abbrev. of hits	taxonomy of species
Msp-1-1	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Alteromonadaceae *Marinobacter sp.* ELB17
Pae-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* 2192
Pae-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* C3719
Pae-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PA7
Pae-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PACS2
Pae-5-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PAO1
Pae-6-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* UCBPP-PA14
Pen-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas entomophila* L48
Pfl-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf-5
Pfl-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf0-1
Pme-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas mendocina ymp*
Ppu-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* F1
Ppu-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* GB-1
Ppu-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* KT2440
Ppu-4-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* W619
Psy-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. phaseolicola* 1448A
Psy-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000
env-1 to env-7	environmental samples

95.2 Notes

Superscript numbers are used to annotate any *sucC* RNA that has special characteristics (described below). These numeric annotations will be used in Sections 95.3 and 95.5. The meaning of the numbers are as follows.

[1] Homology is less clear because the hit was accepted with an E-value of ≈1, and because the species is not in the order Pseudomonadales.

95.3 Gene contexts

Each *sucC* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *sucC* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 95.4. The accession number of the sequence containing each *sucC* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 95.1. Superscript numbers refer to annotations listed in Section 95.2

abbrev.	Seq. accession		5′ at	3′ at	genes
[1]Msp-1-1	NZ_AAXY01000001.1	+	309539	309608	**RNA**→ PP_kinase (pfam02503)→ COG5339 (COG5339)→
env-1	JCVI_SCAF_1101667378642	+	278	346	**RNA**→ SucC (COG0045)? (GOS4231)→
env-2	JCVI_SCAF_1101668315038	-	978	910	**RNA**→ SucC (COG0045)? (GOS4231)→
env-3	JCVI_SCAF_1096628081269	-	276	208	**RNA**→
Pme-1-1	NC_009439.1	-	2763649	2763581	**RNA**→ SucC (COG0045)→ SucD (COG0074)→

env-4	JCVI_SCAF_1096627694513	-	702	632	**RNA**→ SucC (COG0045)? (GOS4231)→
Pae-6-1	NC_008463.1	-	3913400	3913330	**RNA**→ SucC (COG0045)→ SucD (COG0074)→
Pae-3-1	NC_009656.1	-	3818499	3818429	**RNA**→ SucC (COG0045)→ SucD (COG0074)→ BrnQ (COG1114)→
Pae-1-1	NZ_AAKW01000006.1	+	28731	28801	**RNA**→ SucC (COG0045)→ SucD (COG0074)→
Pae-5-1	NC_002516.2	+	1730109	1730179	**RNA**→ SucC (COG0045)→ SucD (COG0074)→
Pae-4-1	NZ_AAQW01000001.1	+	2290395	2290465	**RNA**→ SucC (COG0045)→ SucD (COG0074)→
Pae-2-1	NZ_AAKV01000011.1	+	28701	28771	**RNA**→ SucC (COG0045)→ SucD (COG0074)→
Psy-3-1	NC_004578.1	+	2425315	2425383	**RNA**→ SucC (COG0045)→ SucD (COG0074)→
Psy-1-1	NC_005773.3	+	2331289	2331357	**RNA**→ SucC (COG0045)→ SucD (COG0074)→ BrnQ (COG1114)→
Psy-2-1	NC_007005.1	+	2335517	2335585	**RNA**→ SucC (COG0045)→ SucD (COG0074)→
Pen-1-1	NC_008027.1	-	3922206	3922138	**RNA**→ SucC (COG0045)→ SucD (COG0074)→
Ppu-4-1	NZ_AAVY01000007.1	+	148008	148076	**RNA**→ SucC (COG0045)→ SucD (COG0074)→
Ppu-2-1	NZ_AAXR01000004.1	+	301072	301140	**RNA**→ SucC (COG0045)→ SucD (COG0074)→
Ppu-3-1	NC_002947.3	-	4729904	4729836	**RNA**→ SucC (COG0045)→ SucD (COG0074)→
Ppu-1-1	NC_009512.1	+	1874247	1874315	**RNA**→ SucC (COG0045)→ SucD (COG0074)→
Pfl-2-1	NC_007492.1	+	1808806	1808876	**RNA**→ SucC (COG0045)→ SucD (COG0074)→
Pfl-1-1	NC_004129.6	+	1927644	1927715	**RNA**→ SucC (COG0045)→ SucD (COG0074)→
env-5	JCVI_SCAF_1101668621703	-	978	908	**RNA**→ SucC (COG0045)? (GOS4231)→
env-6	JCVI_SCAF_1096626888236	+	334	404	**RNA**→ SucC (COG0045)? (GOS4231)→
env-7	JCVI_SCAF_1096627624180	+	1296	1366	**RNA**→ SucC (COG0045)? (GOS4231)→

95.4 Conserved domains

Conserved domains found in protein-coding genes listed in Section 95.3 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *sucC* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

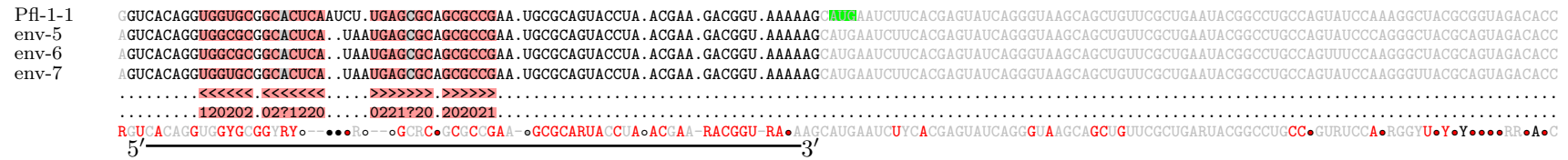
COG0045 (23) Succinyl-CoA synthetase, beta subunit [Energy production and conversion]
COG0074 (17) Succinyl-CoA synthetase, alpha subunit [Energy production and conversion]
COG1114 (2) Branched-chain amino acid permeases [Amino acid transport and metabolism]
COG5339 (1) Uncharacterized protein conserved in bacteria [Function unknown]
GOS4231 (6) ?
pfam02503 (1) Polyphosphate kinase.

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Pae-4-1

95.5 Multiple-sequence alignment

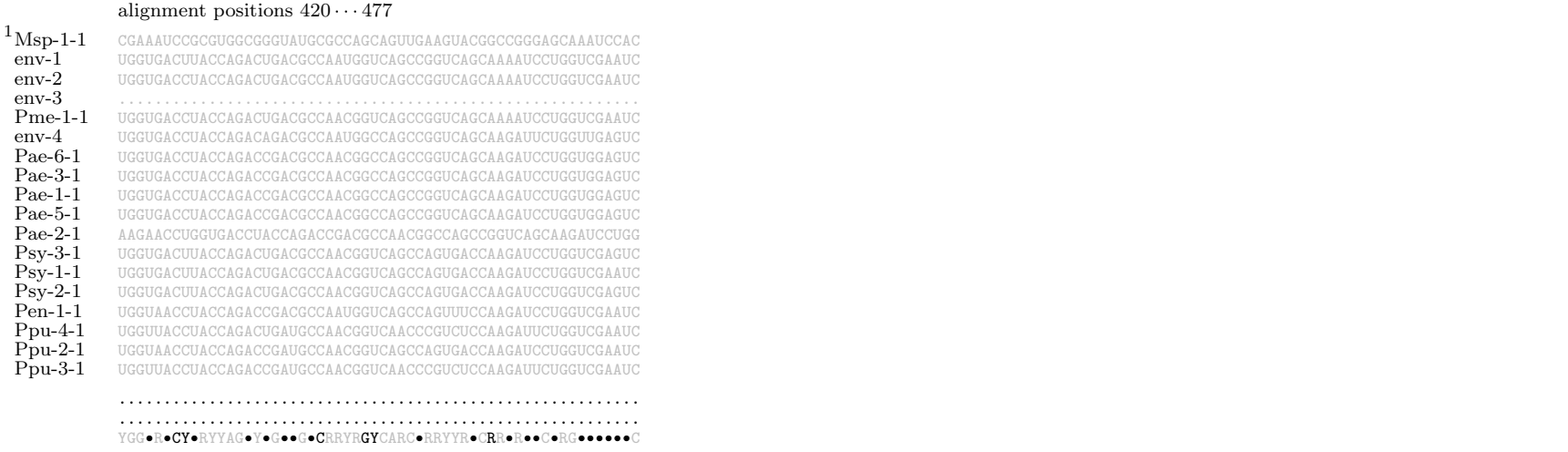
Each *sucC* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 95.1. Superscript numbers refer to annotations listed in Section 95.2. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *sucC* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" =

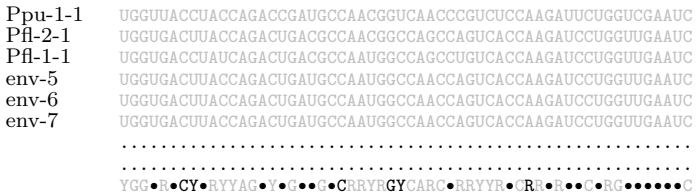
"A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



alignment positions 260 · · · 419

alignment positions 420 · · · 477





96 Termite-*flg*

96.1 Taxa

The taxonomy of each organism containing a putative Termite-*flg* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Termite-*flg* RNA in Sections 96.2 and 96.4.

abbrev. of hits	taxonomy of species
env-1 to env-13	environmental samples

96.2 Gene contexts

Each Termite-*flg* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Termite-*flg* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 96.3. The accession number of the sequence containing each Termite-*flg* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 96.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	ABDH01006538.1	+	207	311	**RNA**→ FlgL (COG1344)→
env-2	ABDH01005516.1	+	187	291	**RNA**→ Flagellin_N (pfam00669)→
env-3	ABDH01031000.1	-	179	73	**RNA**→
env-4	ABDH01011841.1	+	172	278	**RNA**→ Flagellin_N (pfam00669)→
env-5	ABDH01019639.1	-	580	474	**RNA**→ FlgL (COG1344)→
env-6	ABDH01040472.1	+	816	922	**RNA**→ Flagellin_N (pfam00669)Flagellin_C (pfam00700)→
env-7	ABDH01025479.1	+	472	571	**RNA**→ FlgL (COG1344)→
env-8	ABDH01004069.1	+	90	191	**RNA**→ FlgL (COG1344)→
env-9	ABDH01009620.1	+	89	189	**RNA**→ FlgL (COG1344)→
env-10	ABDH01005032.1	-	522	422	**RNA**→ Flagellin_N (pfam00669)→
env-11	ABDH01040089.1	-	2908	2808	**RNA**→ FlgL (COG1344)→
env-12	ABDH01022490.1	+	329	430	**RNA**→ Flagellin_N (pfam00669)→
env-13	ABDH01011825.1	+	35	136	**RNA**→ hypo→ Flagellin_N (pfam00669)→

96.3 Conserved domains

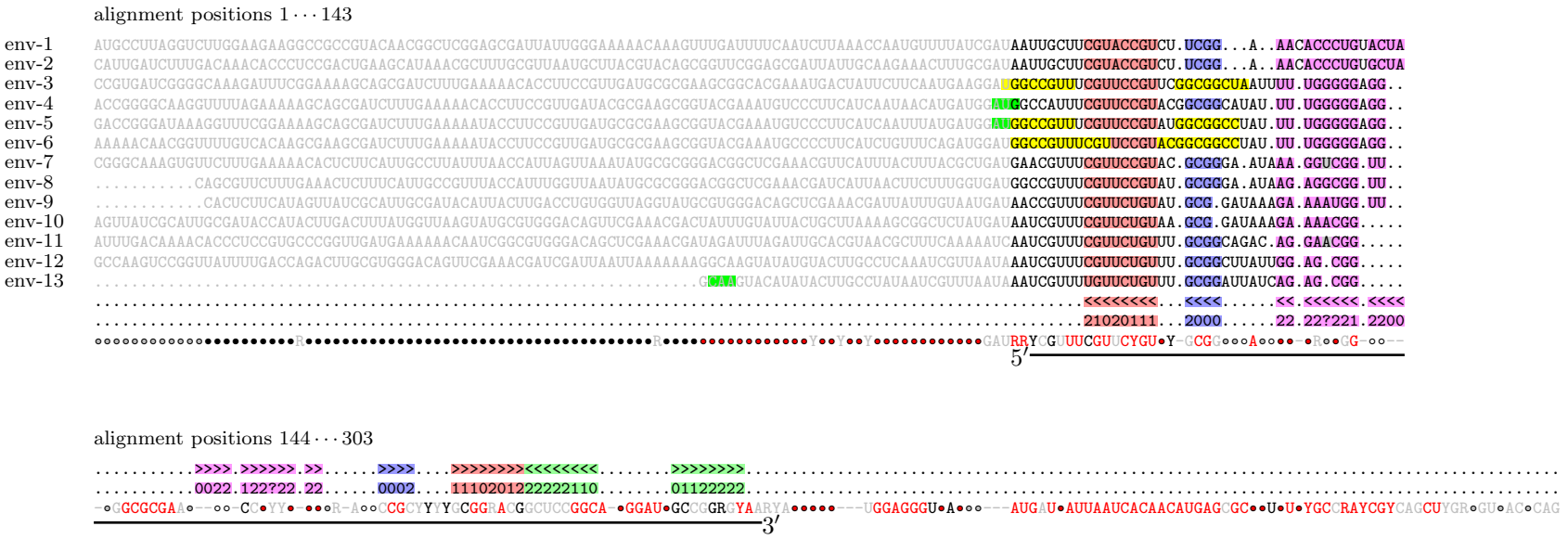
Conserved domains found in protein-coding genes listed in Section 96.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Termite-*flg* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

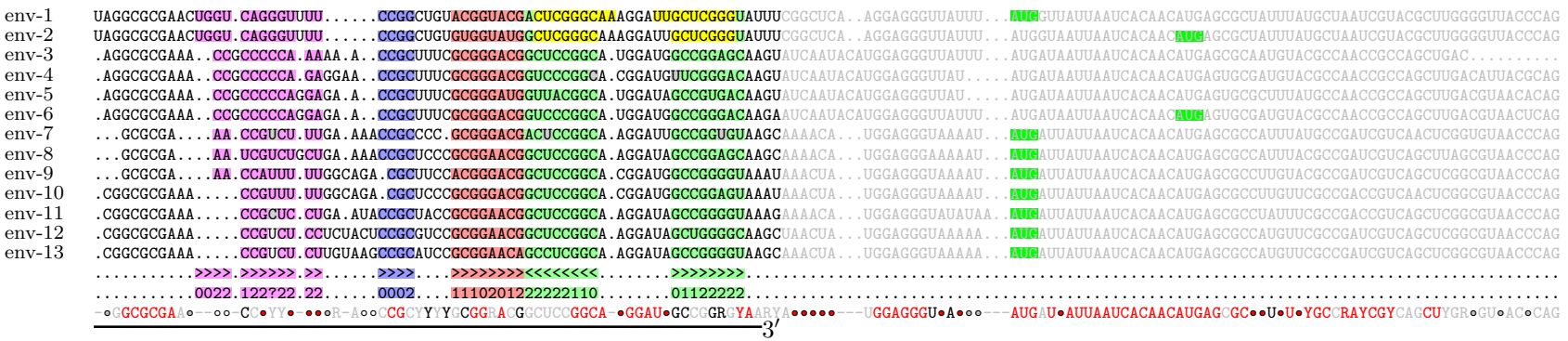
COG1344 (6) Flagellin and related hook-associated proteins [Cell motility and secretion]
pfam00669 (6) Bacterial flagellin N-terminus.

pfam00700 (1) Bacterial flagellin C-terminus.

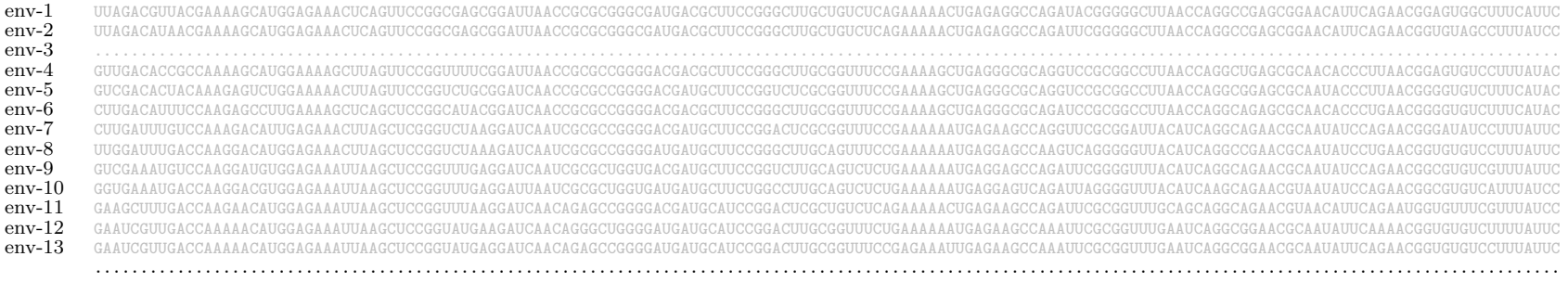
96.4 Multiple-sequence alignment

Each Termite-*flg* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 96.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Termite-*flg* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.





alignment positions 304 · · · 463



alignment positions 464 · · · 524

97 Termite-*leu*

97.1 Taxa

The taxonomy of each organism containing a putative Termite-*leu* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Termite-*leu* RNA in Sections 97.3 and 97.5.

abbrev. of hits	taxonomy of species
env-1 to env-20	environmental samples

97.2 Notes

Superscript numbers are used to annotate any Termite-*leu* RNA that has special characteristics (described below). These numeric annotations will be used in Sections 97.3 and 97.5. The meaning of the numbers are as follows.

[1] accepted with Evalue 1 in local miniscan, but upstream of leuA

[2] accepted with Evalue 1 in local miniscan, but upstream of leuB

97.3 Gene contexts

Each Termite-*leu* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Termite-*leu* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 97.4. The accession number of the sequence containing each Termite-*leu* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 97.1. Superscript numbers refer to annotations listed in Section 97.2

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	ABDH01034349.1	-	657	580	**RNA**→ ←hypo
[2]env-2	ABDH01039360.1	-	692	617	**RNA**→ LeuB (COG0473)→
env-3	ABDH01040491.1	+	1714	1789	**RNA**→ ligand_gated_channel (cd01347)→
env-4	ABDH01041047.1	+	7614	7689	**RNA**→ hypo→
env-5	ABDH01040822.1	-	1467	1392	**RNA**→ COG3263 (COG3263)→
env-6	ABDH01024454.1	+	562	639	**RNA**→ ←DsbDgamma (cd02953)
env-7	ABDH01040894.1	-	1192	1117	**RNA**→ ←SUA5 (COG0009)
env-8	ABDH01004321.1	+	335	412	**RNA**→ hypo→
env-9	ABDH01040812.1	+	1997	2072	**RNA**→ ←ThrA (COG0460)ACT (pfam01842)
env-10	ABDH01040182.1	-	3096	3021	**RNA**→ hypo→
env-11	ABDH01040994.1	-	213	138	**RNA**→
env-12	ABDH01016605.1	+	80	155	**RNA**→ ←hypo
env-13	ABDH01015524.1	-	202	125	**RNA**→
env-14	ABDH01041052.1	+	3073	3148	**RNA**→ ←Gln_GluRS_non_core (cd00807)tRNA-synt_1c_C (pfam03950)

[1]env-15	ABDH01044059.1	+	83	186	**RNA**→ LeuA (COG0119)→
env-16	ABDH01037712.1	+	11	126	**RNA**→ LeuA (COG0119)→
env-17	ABDH01021984.1	-	236	127	**RNA**→
env-18	ABDH01040673.1	-	4423	4312	**RNA**→ IlvC (COG0059)→ hypo→ LeuA (COG0119)→ Iso_dh (pfam00180)→
env-19	ABDH01022123.1	+	160	273	**RNA**→ IlvC (COG0059)→
env-20	ABDH01039021.1	+	112	223	**RNA**→ IlvC (COG0059)→ HMGL-like (pfam00682)→

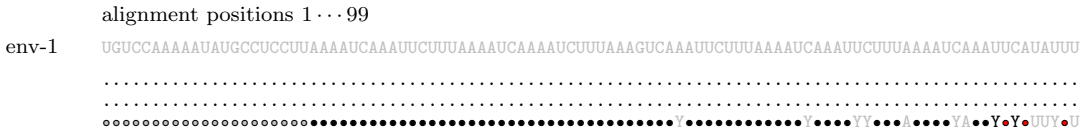
97.4 Conserved domains

Conserved domains found in protein-coding genes listed in Section 97.3 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Termite-*leu* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

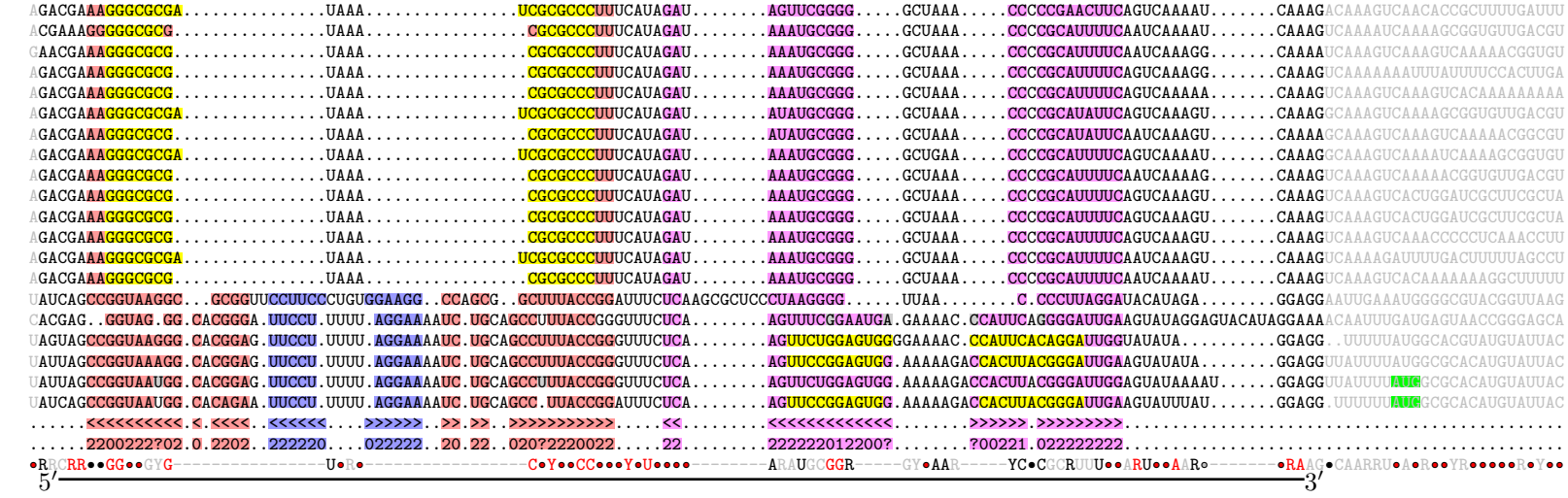
cd00807 (1) Glutaminyl-tRNA synthetase (GlnRS) and non-descriminating Glutamyl-tRNA synthetase (GluRS) cataytic core domain.
cd01347 (1) TonB dependent/Ligand-Gated channels are created by a monomeric 22 strand (22,24) anti-parallel beta-barrel.
cd02953 (1) DsbD gamma family; DsbD gamma is the C-terminal periplasmic domain of the bacterial protein DsbD.
COG0009 (1) Putative translation factor (SUA5) [Translation, ribosomal structure and biogenesis]
COG0059 (3) Ketol-acid reductoisomerase [Amino acid transport and metabolism / Coenzyme metabolism]
COG0119 (3) Isopropylmalate/homocitrate/citramalate synthases [Amino acid transport and metabolism]
COG0460 (1) Homoserine dehydrogenase [Amino acid transport and metabolism]
COG0473 (1) Isocitrate/isopropylmalate dehydrogenase [Amino acid transport and metabolism]
COG3263 (1) NhaP-type Na+/H+ and K+/H+ antiporters with a unique C-terminal domain [Inorganic ion transport and metabolism]
pfam00180 (1) Isocitrate/isopropylmalate dehydrogenase.
pfam00682 (1) HMGL-like.
pfam01842 (1) ACT domain.
pfam03950 (1) tRNA synthetases class I (E and Q), anti-codon binding domain.

97.5 Multiple-sequence alignment

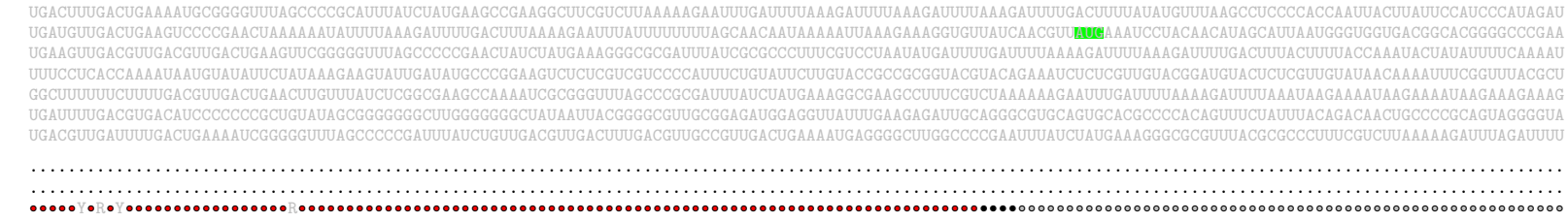
Each Termite-*leu* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 97.1. Superscript numbers refer to annotations listed in Section 97.2. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Termite-*leu* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



alignment positions 100 · · · 259



alignment positions 260 · · · 419



98 *traJ*-II

98.1 Taxa

The taxonomy of each organism containing a putative *traJ*-II RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *traJ*-II RNA in Sections 98.2 and 98.4.

abbrev. of hits	taxonomy of species
uba-1-1 to uba-1-4	Bacteria environmental samples uncultured bacterium
Efa-1-1 to Efa-1-11	Bacteria Firmicutes Lactobacillales Enterococcaceae *Enterococcus faecium* DO
Rpa-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* BisB18
Rpa-2-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* CGA009
Bsu-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Brucellaceae *Brucella suis* 1330
Rru-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodospirillales Rhodospirillaceae *Rhodospirillum rubrum* ATCC 11170

Ssp-1-1	Bacteria Proteobacteria Alphaproteobacteria Sphingomonadales Sphingomonadaceae *Sphingomonas sp.* A1
Ade-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Alcaligenaceae *Achromobacter denitrificans*
Axy-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Alcaligenaceae *Achromobacter xylosoxidans*
Bpe-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Alcaligenaceae *Bordetella pertussis*
Bam-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia ambifaria* AMMD
Bps-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 9
Bps-2-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 91
Bps-3-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* K96243
Reu-1-1 to Reu-1-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Cupriavidus *Ralstonia eutropha* JMP134
Asp-1-1 to Asp-1-2	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae *Acidovorax sp.* JS42
Dac-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae *Delftia acidovorans*
Neu-1-1 to Neu-1-3	Bacteria Proteobacteria Betaproteobacteria Nitrosomonadales Nitrosomonadaceae *Nitrosomonas eutropha* C91
Aar-1-1	Bacteria Proteobacteria Betaproteobacteria Rhodocyclales Rhodocyclaceae *Aromatoleum aromaticum* EbN1
Maq-1-1	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Alteromonadaceae *Marinobacter aquaeolei* VT8
Sba-1-1	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella baltica* OS195
Eae-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Enterobacter aerogenes*
Eco-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli*
Eco-2-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* B7A
Eco-3-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* HS
Sfl-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella flexneri* 2a str. 2457T
Yen-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia enterocolitica subsp. enterocolitica* 8081
Yps-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pseudotuberculosis* IP 31758
Lpn-1-1 to Lpn-1-2	Bacteria Proteobacteria Gammaproteobacteria Legionellales Legionellaceae *Legionella pneumophila str.* Corby
Lpn-2-1	Bacteria Proteobacteria Gammaproteobacteria Legionellales Legionellaceae *Legionella pneumophila str.* Paris
Hso-1-1	Bacteria Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae Histophilus *Haemophilus somnus* 129PT
Aba-1-1 to Aba-1-2	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Moraxellaceae *Acinetobacter baumannii* ATCC 17978
Pae-1-1 to Pae-1-2	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa*
Pae-2-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* UCBPP-PA14
Pfu-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fulva*
Vsp-1-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Vibrio sp.* Ex25
Xfa-1-1	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xylella fastidiosa* Ann-1
Xfa-2-1	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xylella fastidiosa* Dixon
Psp-1-1	Bacteria Proteobacteria *Pseudomonas sp.* ADP
env-1 to env-12	environmental samples

98.2 Gene contexts

Each *traJ*-II RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *traJ*-II RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 98.3. The accession number of the sequence containing each *traJ*-II RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 98.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Rpa-1-1	NC_007925.1	-	3949352	3949235	**RNA**→ hypo→
Rpa-2-1	NC_005296.1	-	4675439	4675322	**RNA**→ hypo→
Hso-1-1	NC_006298.1	-	2725	2610	**RNA**→ Rep_3 (pfam01051)→ hypo→
Neu-1-1	NC_008341.1	+	49656	49768	**RNA**→ TraJ protein→ Relaxase (pfam03432)→ TRAG (pfam02534)→ TraF (COG4959)→ SLT (pfam01464)→ COG4227 (COG4227)COG4643 (COG4643)→
Aba-1-1	NC_009083.1	-	12579	12464	**RNA**→ regulatory protein LysR→
Rru-1-1	NC_007641.1	+	12307	12420	**RNA**→ ←Resolvase (pfam00239)HTH_7 (pfam02796)
Bsu-1-1	NC_004311.2	+	351801	351905	**RNA**→ traJ protein→ Relaxase (pfam03432)NOG06529 (NOG06529)→

Eco-2-1	NZ_AAJT01000036.1	+	37600	37713	**RNA**→ hypo→
Sba-1-1	NZ_AATK01000113.1	-	4369	4249	**RNA**→ putative plasmid conjugal transfer protein→ Relaxase (pfam03432)→ relaxase/mobilization nuclease domain→
env-1	JCVI_SCAF_1096627192326	-	6254	6135	**RNA**→ ? (GOS13060501)→ Relaxase (pfam03432)? (GOS13078013)→
Pfu-1-1	NC_004951.1	+	1746	1860	**RNA**→
Eco-3-1	NZ_AAJY01000001.1	-	2661881	2661766	**RNA**→ COG5314 (COG5314)→
Vsp-1-1	NZ_AAKK02000007.1	-	48960	48845	**RNA**→ COG5314 (COG5314)→ hypo→ TrbL (COG3846)→
Lpn-2-1	NC_006368.1	-	68995	68919	**RNA**→ HTH_XRE (cd00093)NOG18680 (NOG18680)→ HipA (COG3550)→
Maq-1-1	NC_008740.1	-	1924450	1924330	**RNA**→ StbC (COG4691)→
Aar-1-1	NC_006824.1	-	58535	58414	**RNA**→ TraJ protein of DNA transfer system→
Lpn-1-1	NC_009494.1	-	204445	204334	**RNA**→ TraJ protein→ Relaxase (pfam03432)→
Lpn-1-2	NC_009494.1	+	3056748	3056860	**RNA**→ traJ protein→ Relaxase (pfam03432)→
env-2	sludgePhrap_Contig16293	+	3941	4057	**RNA**→ ←Integrase(,,,)
env-3	sludgePhrap_Contig13274	+	2060	2134	**RNA**→
env-4	sludgePhrap_Contig15290	+	1561	1637	**RNA**→ (,,,)→
env-5	sludgeOz_Contig9677	+	478	595	**RNA**→ (,,,)→ Relaxase (pfam03432)→
env-6	sludgePhrap_Contig16329	+	3197	3314	**RNA**→ RelB (COG3077)→ COG3041 (COG3041)→
env-7	sludgePhrap_Contig8904	+	115	232	**RNA**→ (,,,)→ Relaxase (pfam03432)→
env-8	sludgePhrap_Contig16340	+	16727	16844	**RNA**→ (,,,)→ Relaxase (pfam03432)→ TRAG (pfam02534)→ TraF (COG4959)→ (,,,)→ COG4227 (COG4227)COG4643 (COG4643)→
env-9	sludgePhrap_Contig3329	-	469	352	**RNA**→ (,,,)→
env-10	sludgePhrap_Contig14398	+	262	379	**RNA**→ (,,,)→ Relaxase (pfam03432)→
env-11	sludgePhrap_Contig8607	+	326	443	**RNA**→ (,,,)→ Relaxase (pfam03432)→
Efa-1-1	NZ_AAAK03000145.1	-	327	207	**RNA**→
Efa-1-2	NZ_AAAK03000149.1	+	2174	2294	**RNA**→
Efa-1-3	NZ_AAAK03000042.1	-	404	284	**RNA**→
Efa-1-4	NZ_AAAK03000148.1	+	2246	2366	**RNA**→
Efa-1-5	NZ_AAAK03000031.1	+	24224	24344	**RNA**→
Efa-1-6	NZ_AAAK03000093.1	-	403	283	**RNA**→
Efa-1-7	NZ_AAAK03000024.1	-	350	230	**RNA**→
Efa-1-8	NZ_AAAK03000117.1	+	5053	5173	**RNA**→ ←FtsK_SpoIIIE (pfam01580)
Pae-1-1	NC_008357.1	-	80290	80170	**RNA**→ TraJ oriT recognising protein→ TraX protein→
Eco-1-1	NC_005923.1	-	3801	3681	**RNA**→ CAT (pfam00302)→
uba-1-1	NC_006352.1	-	55497	55377	**RNA**→ oriT-binding protein→
Efa-1-9	NZ_AAAK03000025.1	+	25250	25370	**RNA**→ hypo→
Efa-1-10	NZ_AAAK03000061.1	-	643	523	**RNA**→
Efa-1-11	NZ_AAAK03000128.1	+	4010	4130	**RNA**→ hypo→
Xfa-2-1	NZ_AAAL02000016.1	-	12113	11995	**RNA**→ Relaxase (pfam03432)→ TRAG (pfam02534)→ TraF (COG4959)→ TopA (COG0550)zf-C4_Topoisom (pfam01396)→ conjugal transfer protein→ COG4227 (COG4227)COG4643 (COG4643)→
Xfa-1-1	NZ_AAAM03000037.1	-	13794	13676	**RNA**→ Relaxase (pfam03432)→ TRAG (pfam02534)→ TraF (COG4959)→ TopA (COG0550)zf-C4_Topoisom (pfam01396)→ conjugal transfer protein→ COG4227 (COG4227)COG4643 (COG4643)→
Bps-3-1	NC_006350.1	-	3882620	3882503	**RNA**→ putative plasmid conjugal transfer protein→ Relaxase (pfam03432)NOG06529 (NOG06529)→
Bps-2-1	NZ_ABBK01000836.1	-	18196	18079	**RNA**→
Ade-1-1	NC_005793.1	-	74417	74299	**RNA**→ TraJ→ TraX→
Yps-1-1	NC_009704.1	-	50565	50446	**RNA**→ hypo→ conjugal transfer protein TraJ→ Relaxase (pfam03432)→ TRAG (pfam02534)→ TraF (COG4959)→ TopA (COG0550)zf-C4_Topoisom (pfam01396)→ putative conjugal transfer protein TraD→ COG4227 (COG4227)COG4643 (COG4643)→
Yen-1-1	NC_008800.1	-	1319525	1319415	**RNA**→ COG5314 (COG5314)→ hypo→ TrbL (COG3846)→
Asp-1-1	NC_008766.1	+	55378	55496	**RNA**→ hypo→ TraJ protein of DNA transfer system→ Relaxase (pfam03432)→ putative entry/exclusion protein TrbK→ TRAG (pfam02534)→ TraF (COG4959)→ TopA (COG0550)zf-C4_Topoisom (pfam01396)→ conjugal transfer protein→
uba-1-2	NC_003430.1	-	72113	71995	**RNA**→ TraJ protein→ hypo→

Ssp-1-1	NC_007353.2	-	16811	16692	**RNA**→ TraJ→
Axy-1-1	NC_006830.1	-	68726	68608	**RNA**→ conjugal transfer protein TraJ→
Reu-1-1	NC_007337.1	-	85852	85733	**RNA**→ TraJ protein of DNA transfer system→ Relaxase (pfam03432)NOG06529 (NOG06529)→ TRAG (pfam02534)→ TraF (COG4959)→ TopA (COG0550)zf-C4_Topoisom (pfam01396)→ conjugal transfer protein→ COG4227 (COG4227)COG4643 (COG4643)→
Reu-1-2	NC_005912.1	-	85852	85733	**RNA**→ oriT binding protein→
Eae-1-1	NC_001735.4	-	51588	51469	**RNA**→ TraJ→ ORFX→
Psp-1-1	NC_004956.1	+	11995	12114	**RNA**→ TraJ protein→ Relaxase (pfam03432)→ TraH protein→
Bam-1-1	NC_008385.1	+	14717	14836	**RNA**→ TraJ protein of DNA transfer system→ Relaxase (pfam03432)→ TRAG (pfam02534)→ TraF (COG4959)→ TopA (COG0550)zf-C4_Topoisom (pfam01396)→ conjugal transfer protein→ COG4227 (COG4227)COG4643 (COG4643)→
uba-1-3	NC_004840.1	-	31396	31277	**RNA**→ oriT binding protein→
uba-1-4	NC_007680.1	-	52509	52390	**RNA**→ TraJ protein→ TraX protein→
Dac-1-1	NC_005088.1	-	56634	56515	**RNA**→ TraJ→
env-12	sludgeOz_Contig11623	+	12970	13089	**RNA**→ (,,,)→ Relaxase (pfam03432)→ TRAG (pfam02534)→ TraF (COG4959)→ TopA (COG0550)→ Topoisom_bac (pfam01131)zf-C4_Topoisom (pfam01396)→ (,,,)→ COG4227 (COG4227)COG4643 (COG4643)→
Neu-1-2	NC_008344.1	+	2382211	2382330	**RNA**→ StbC (COG4691)→
Neu-1-3	NC_008344.1	-	683049	682930	**RNA**→ StbC (COG4691)→
Aba-1-2	NC_009085.1	+	788981	789100	**RNA**→ COG5314 (COG5314)→
Sfl-1-1	NC_004741.1	-	2609584	2609465	**RNA**→ StbC (COG4691)→
Bpe-1-1	NC_008459.1	-	39467	39348	**RNA**→ TraJ→
Bps-1-1	NZ_ABBL01001689.1	+	9564	9680	**RNA**→
Pae-1-2	NC_009739.1	-	22315	22194	**RNA**→ TraJ protein→ Relaxase (pfam03432)→
Asp-1-2	NC_008782.1	-	2704928	2704812	**RNA**→ OriT-binding protein, TraJ→
Pae-2-1	NC_008463.1	+	1312525	1312644	**RNA**→ oriT-binding protein, TraJ→

98.3 Conserved domains

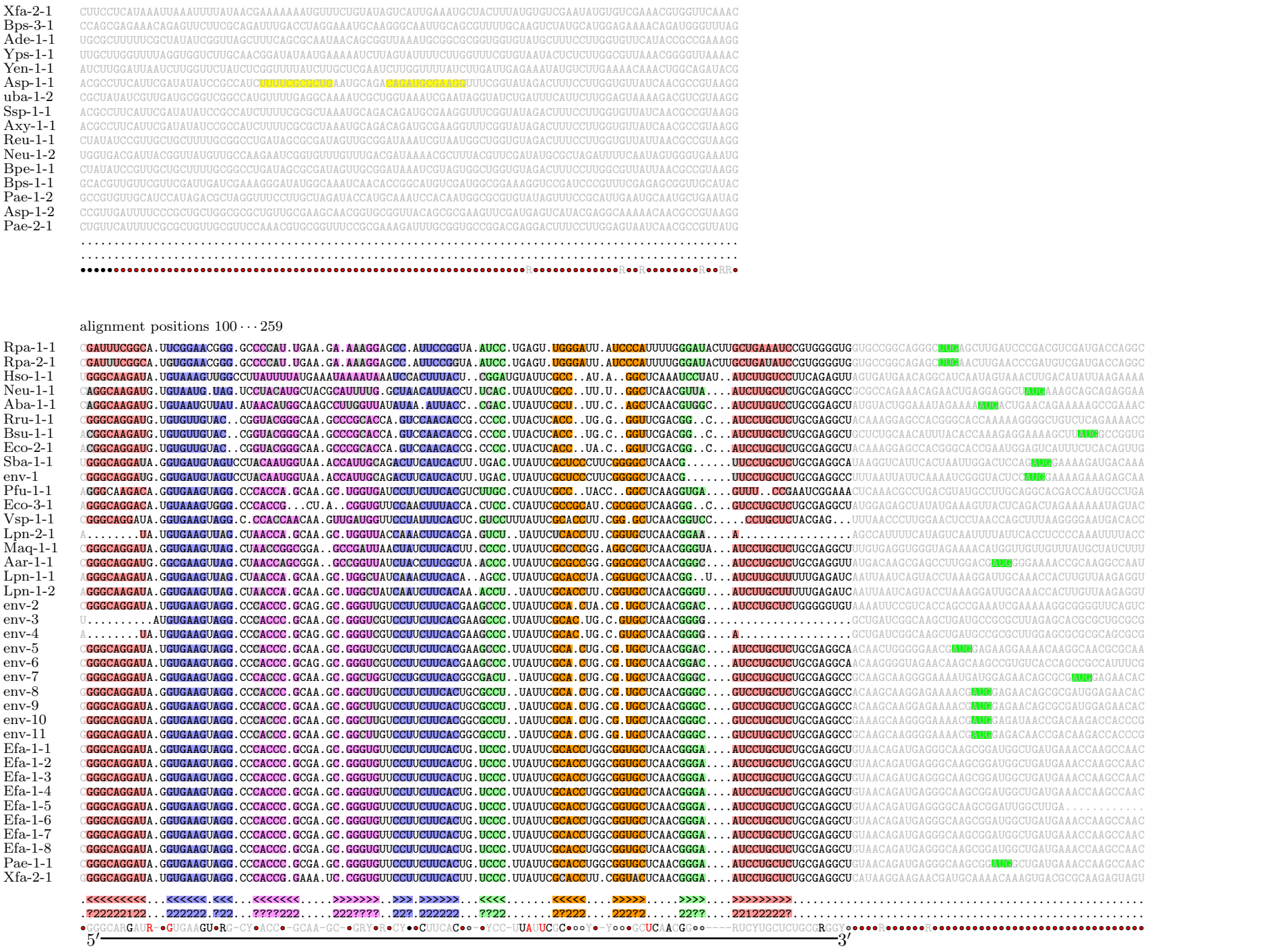
Conserved domains found in protein-coding genes listed in Section 98.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *traJ*-II RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

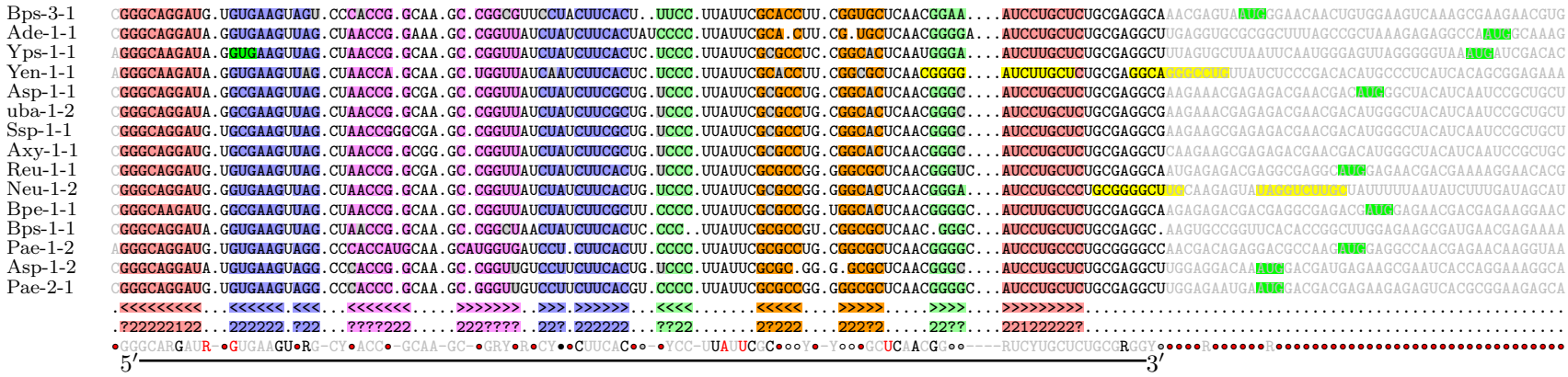
cd00093 (1) Helix-turn-helix XRE-family like proteins.
COG0550 (7) Topoisomerase IA [DNA replication, recombination, and repair]
COG3041 (1) Uncharacterized protein conserved in bacteria [Function unknown]
COG3077 (1) DNA-damage-inducible protein J [DNA replication, recombination, and repair]
COG3550 (1) Uncharacterized protein related to capsule biosynthesis enzymes [General function prediction only]
COG3846 (2) Type IV secretory pathway, TrbL components [Intracellular trafficking and secretion]
COG4227 (8) Antirestriction protein [DNA replication, recombination, and repair]
COG4643 (8) Uncharacterized protein conserved in bacteria [Function unknown]
COG4691 (4) Plasmid stability protein [General function prediction only]
COG4959 (9) Type IV secretory pathway, protease TraF [Posttranslational modification, protein turnover, chaperones / Intracellular trafficking and secretion]
COG5314 (4) Conjugal transfer/entry exclusion protein [Intracellular trafficking and secretion]
GOS13060501 (1) ?
GOS13078013 (1) ?
NOG06529 (3)
NOG18680 (1)
pfam00239 (1) Resolvase, N terminal domain.
pfam00302 (1) Chloramphenicol acetyltransferase.
pfam01051 (1) Initiator Replication protein.
pfam01131 (1) DNA topoisomerase.
pfam01396 (7) Topoisomerase DNA binding C4 zinc finger.
pfam01464 (1) Transglycosylase SLT domain.
pfam01580 (1) FtsK/SpoIIIE family.
pfam02534 (9) TraG/TraD family.
pfam02796 (1) Helix-turn-helix domain of resolvase.
pfam03432 (21) Relaxase/Mobilisation nuclease domain.

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Aba-1-2, Bam-1-1, Bps-2-1, Dac-1-1, Eae-1-1, Eco-1-1, Efa-1-10, Efa-1-11, Efa-1-9, Neu-1-3, Psp-1-1, Reu-1-2, Sfl-1-1, Xfa-1-1, env-12, uba-1-1, uba-1-3, uba-1-4

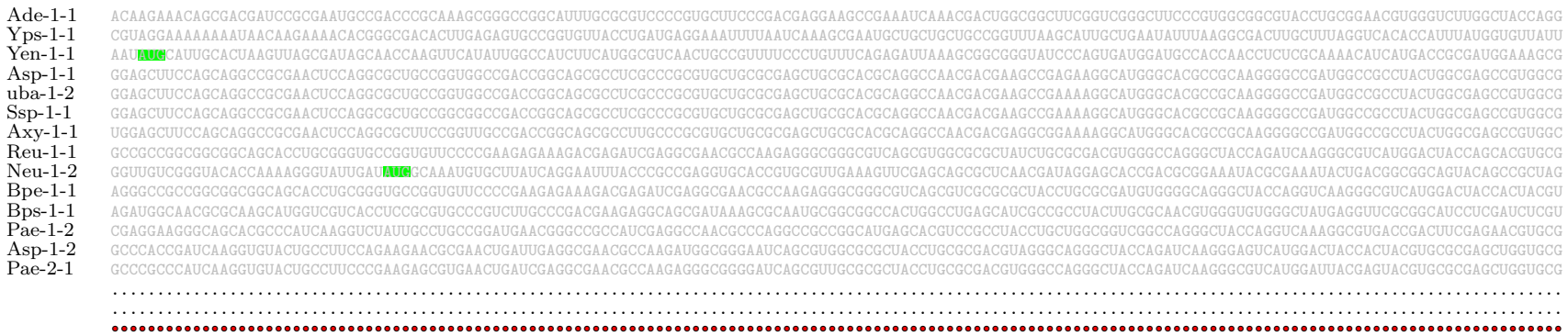
98.4 Multiple-sequence alignment

Each *traJ*-II RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 98.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *traJ*-II RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

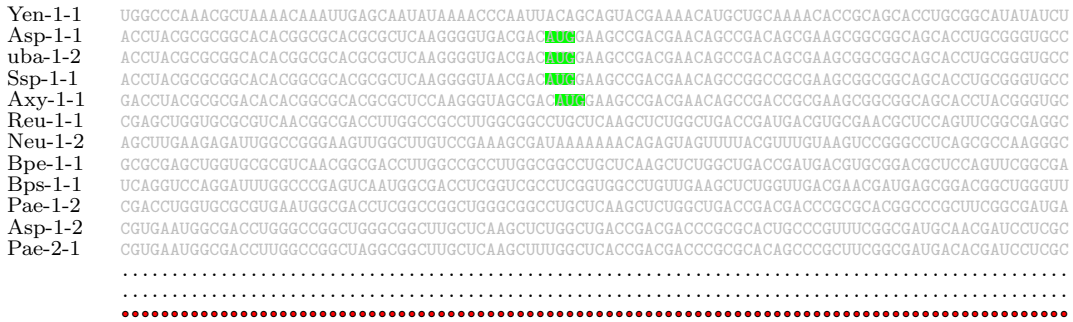

alignment positions 100 · · · 259
5′
3′



alignment positions 260 · · · 419



alignment positions $420 \cdots 515$



99 Transposase-resistance

99.1 Taxa

The taxonomy of each organism containing a putative Transposase-resistance RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Transposase-resistance RNA in Sections 99.2 and 99.4.

abbrev. of hits	taxonomy of species
Cgl-1-1 to Cgl-1-2	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Corynebacteriaceae *Corynebacterium glutamicum*
Csp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Corynebacteriaceae *Corynebacterium sp.* L2-79-05
Cau-1-1	Bacteria Chloroflexi Chloroflexales Chloroflexaceae *Chloroflexus aurantiacus* J-10-fl
uba-1-1 to uba-1-2	Bacteria environmental samples uncultured bacterium
Ogr-1-1	Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae *Oceanicola granulosus* HTCC2516
Mpe-1-1	Bacteria Proteobacteria Betaproteobacteria Burkholderiales *Methylibium petroleiphilum* PM1
Tde-1-1	Bacteria Proteobacteria Betaproteobacteria Hydrogenophilales Hydrogenophilaceae *Thiobacillus denitrificans* ATCC 25259
Mfl-1-1	Bacteria Proteobacteria Betaproteobacteria Methylophilales Methylophilaceae *Methylobacillus flagellatus* KT
Neu-1-1 to Neu-1-2	Bacteria Proteobacteria Betaproteobacteria Nitrosomonadales Nitrosomonadaceae *Nitrosomonas eutropha* C91
Apu-1-1	Bacteria Proteobacteria Gammaproteobacteria Aeromonadales Aeromonadaceae *Aeromonas punctata (*Aeromonas caviae)
Asa-1-1	Bacteria Proteobacteria Gammaproteobacteria Aeromonadales Aeromonadaceae *Aeromonas salmonicida subsp. salmonicida* A449
Cfr-1-1 to Cfr-1-3	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Citrobacter freundii*
Eae-1-1 to Eae-1-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Enterobacter aerogenes*
Eco-1-1 to Eco-1-13	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli*
Eco-2-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* B171
Eco-3-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* B7A
Eco-4-1 to Eco-4-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* SECEC SMS-3-5
Kpn-1-1 to Kpn-1-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Klebsiella pneumoniae subsp. pneumoniae* MGH 78578
Sen-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Newport *Salmonella enterica subsp. en*
Sdy-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella dysenteriae* 1012
Sso-1-1	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Shigella sonnei* Ss046
Ype-1-1 to Ype-1-2	Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Orientalis str. IP275
Pmu-1-1	Bacteria Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae *Pasteurella multocida*
Asp-1-1	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Moraxellaceae *Acinetobacter sp.* SUN
Pae-1-1 to Pae-1-2	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa*
Vch-1-1	Bacteria Proteobacteria Gammaproteobacteria Vibrionales Vibrionaceae *Vibrio cholerae* B33
Xax-1-1 to Xax-1-2	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas axonopodis pv. citri str.* 306
Xca-1-1 to Xca-1-13	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas campestris pv. campestris str.* 8004
Xca-2-1 to Xca-2-13	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas campestris pv. campestris str.* ATCC 33913
Xor-1-1	Bacteria Proteobacteria Gammaproteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas oryzae pv. oryzicola* BLS256
Sen-2-1 to Sen-2-6	CVM19633Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Salmonella enterica subsp. enterica serovar* Schwarzengrund str
Sen-3-1 to Sen-3-6	SC-B67Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Salmonella Salmonella enterica subsp. entericaserovar Choleraesuis *Salmonella enteric*
Sen-2-1 to Sen-2-6	SL480Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae *Salmonella enterica subsp. enterica serovar* Schwarzengrund str

99.2 Gene contexts

Each Transposase-resistance RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Transposase-resistance RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 99.3. The accession number of the sequence containing each Transposase-resistance RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 99.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	JCVI_SCAF_1096627926921	-	332	191	**RNA**→
Ogr-1-1	NZ_AAOT01000085.1	-	885	748	hypo→ **RNA**→ ←Glf (COG0562)
Eae-1-1	NC_001735.4	+	33417	33474	DhfrIIc→ **RNA**→ orfD→
Tde-1-1	NC_007404.1	-	1266337	1266270	COG4333 (COG4333)→ **RNA**→ ←hypo
env-2	sludgeOz_Contig10390	-	728	661	(,,,)→ **RNA**→ (,,,)→
env-3	sludgePhrap_Contig16243	-	1593	1526	(,,,)→ **RNA**→ (,,,)→
env-4	sludgeOz_Contig11624	+	16482	16552	Dcd (COG0717)→ **RNA**→ (,,,)→
env-5	sludgePhrap_Contig5236	+	819	889	(,,,)→ **RNA**→ (,,,)→
Sen-3-1	NC_006856.1	+	25069	25159	DHFR (cd00209)→ **RNA**→ ←hypo
Cfr-1-1	NC_004464.2	+	66491	66581	DHFR (cd00209)→ **RNA**→ hypo→
Eco-1-1	NC_009131.1	+	39372	39462	**RNA**→
Eco-1-2	NC_009132.1	-	24993	24903	DHFR (cd00209)→ **RNA**→ **RNA**→ NTP_transf_2 (pfam01909)→
Eco-1-4	NC_009132.1	-	24643	24583	(shown above)
Sen-2-4	NZ_ABEJ01000032.1	-	29026	28936	**RNA**→
Sen-3-2	NC_006856.1	+	25419	25479	←hypo **RNA**→ NTP_transf_2 (pfam01909)→
Cfr-1-2	NC_004464.2	+	66841	66901	hypo→ **RNA**→ NTP_transf_2 (pfam01909)→
Eco-1-3	NC_009131.1	+	39722	39782	**RNA**→
Sen-2-5	NZ_ABEJ01000032.1	-	28676	28616	**RNA**→
Eae-1-2	NC_001735.4	+	33734	33794	orfD→ **RNA**→ EmrE (COG2076)→
env-6	sludgePhrap_Contig13421	-	1261	1193	(,,,)→ **RNA**→ COG3344 (COG3344)→
env-7	sludgeOz_Contig11452	-	4911	4848	COG1479 (COG1479)→ **RNA**→ C4dic_mal_tran (pfam03595)→
env-8	sludgeOz_Contig5990	-	490	421	←(,,,) **RNA**→ ←(,,,)
env-9	sludgeOz_Contig10284	-	3037	2925	**RNA**→ NTP_transf_2 (pfam01909)→
env-10	sludgeOz_Contig10886	-	793	681	(,,,)→ **RNA**→ ←(,,,)
env-11	sludgeOz_Contig9019	-	1005	937	←(,,,) **RNA**→ ←(,,,)
env-12	sludgeOz_Contig9306	+	443	511	←(,,,) **RNA**→ ←(,,,)
env-13	sludgeOz_Contig10657	+	3443	3511	←(,,,) **RNA**→ ←(,,,)
env-14	sludgePhrap_Contig14205	-	346	287	Glyoxalase (pfam00903)→ **RNA**→ ←(,,,)
env-15	sludgePhrap_Contig13455	+	142	201	**RNA**→ (,,,)→
Mpe-1-1	NC_008825.1	+	410415	410529	hypo→ **RNA**→ hypo→
env-16	sludgeOz_Contig8564	-	565	501	(,,,)→ **RNA**→ ←(,,,)
env-17	sludgeOz_Contig1514	-	314	236	(,,,)→ **RNA**→ (,,,)→
env-18	sludgePhrap_Contig16286	+	21577	21655	(,,,)→ **RNA**→ Glyoxalase (pfam00903)→
env-19	sludgeOz_Contig11569	+	11337	11415	(,,,)→ **RNA**→ (,,,)→

env-20	sludgeOz_Contig11505	+	12411	12460	(,,,)→ **RNA**→ TIR (smart00255)→
env-21	sludgePhrap_Contig15228	-	886	827	DinB (COG2318)→ **RNA**→ (,,,)→
env-22	sludgePhrap_Contig14698	+	1090	1198	COG3791 (COG3791)→ **RNA**→ (,,,)→
Eco-1-5	NC_004998.1	-	82813	82713	Acetyltransf_1 (pfam00583)→ **RNA**→ Acetyltransf_1 (pfam00583)→
env-23	sludgeOz_Contig8967	-	427	369	Abi (pfam02517)→ **RNA**→
env-24	sludgePhrap_Contig7109	-	1050	990	**RNA**→ (,,,)→
env-25	sludgeOz_Contig11452	-	3660	3602	C4dic_mal_tran (pfam03595)→ **RNA**→ (,,,)→
env-26	sludgeOz_Contig9696	-	213	155	←INT_IntI (cd01193) **RNA**→
env-27	sludgeOz_Contig8967	-	1402	1344	←(,,,) **RNA**→ Abi (pfam02517)→
env-29	soil_XZS11214_b1	+	408	516	**RNA**→
env-30	sludgePhrap_Contig8783	-	1502	1444	**RNA**→ ←(,,,)
env-31	sludgeOz_Contig10536	-	1679	1589	(,,,)→ **RNA**→ **RNA**→ (,,,)→
env-39	sludgeOz_Contig10536	-	2028	1936	(shown above)
env-32	sludgePhrap_Contig1119	+	29	141	**RNA**→ (,,,)→
env-34	sludgePhrap_Contig291	-	431	319	(,,,)→ **RNA**→
env-35	sludgePhrap_Contig14656	+	618	730	(,,,)→ **RNA**→ (,,,)→
env-36	sludgePhrap_Contig15602	+	2356	2468	(,,,)→ **RNA**→ (,,,)→
Eco-4-1	NZ_ABAQ01000017.1	+	4029	4089	**RNA**→
Sen-3-3	NC_006856.1	+	26275	26335	NTP_transf_2 (pfam01909)→ **RNA**→ Multi_Drug_Res (pfam00893)→
Cfr-1-3	NC_004464.2	+	67697	67757	NTP_transf_2 (pfam01909)→ **RNA**→ Multi_Drug_Res (pfam00893)→
Eco-1-6	NC_009131.1	+	40578	40638	**RNA**→
Eco-1-7	NC_009132.1	-	23787	23727	NTP_transf_2 (pfam01909)→ **RNA**→ Multi_Drug_Res (pfam00893)→
Eco-1-8	NC_007682.3	+	12884	12944	NTP_transf_2 (pfam01909)→ **RNA**→ Multi_Drug_Res (pfam00893)→
Apu-1-1	NC_006143.1	-	37578	37518	NTP_transf_2 (pfam01909)→ **RNA**→ Multi_Drug_Res (pfam00893)→
Cgl-1-1	NC_004945.1	+	15772	15832	NTP_transf_2 (pfam01909)→ **RNA**→ Multi_Drug_Res (pfam00893)→
Ype-1-1	NC_009141.1	-	153872	153812	NTP_transf_2 (pfam01909)→ **RNA**→ Multi_Drug_Res (pfam00893)→
Eco-3-1	NZ_AAJT01000012.1	+	78938	78998	NTP_transf_2 (pfam01909)→ **RNA**→ DHPS (cd00739)→
Ype-1-2	NZ_AAOS01000105.1	+	6141	6201	NTP_transf_2 (pfam01909)→ **RNA**→ DHPS (cd00739)→
Sen-2-6	NZ_ABEJ01000032.1	-	27820	27760	**RNA**→
env-37	sludgeOz_Contig10284	-	2129	2069	NTP_transf_2 (pfam01909)→ **RNA**→ Multi_Drug_Res (pfam00893)→
Vch-1-1	NZ_AAWE01000022.1	-	14332	14272	NTP_transf_2 (pfam01909)→ **RNA**→ DHPS (cd00739)→
Pae-1-1	NC_007100.1	-	5299	5192	←Soj (COG1192) **RNA**→ NTP_transf_2 (pfam01909)→
Xca-1-1	NC_007086.1	-	409369	409309	hypo→ **RNA**→ ←xanthomonadin biosynthesis related protein
Xca-2-1	NC_003902.1	-	408775	408715	hypo→ **RNA**→ ←xanthomonadin biosynthesis related protein
Xca-1-2	NC_007086.1	-	406598	406486	hypo→ **RNA**→ **RNA**→ hypo→
Xca-1-12	NC_007086.1	-	406034	405974	(shown above)
Xca-2-2	NC_003902.1	-	406004	405892	hypo→ **RNA**→ **RNA**→ hypo→
Xca-2-12	NC_003902.1	-	405449	405389	(shown above)
env-38	sludgePhrap_Contig15228	-	1904	1844	(,,,)→ **RNA**→ DinB (COG2318)→
env-40	sludgePhrap_Contig13301	-	1488	1428	(,,,)→ **RNA**→ ←(,,,)
env-41	sludgePhrap_Contig89	+	675	735	(,,,)→ **RNA**→ **RNA**→ (,,,)→
env-82	sludgePhrap_Contig89	+	374	432	(shown above)
env-42	sludgePhrap_Contig2626	+	688	748	(,,,)→ **RNA**→
env-81	sludgePhrap_Contig2626	+	385	445	(shown above)
env-43	sludgePhrap_Contig15830	-	371	311	(,,,)→ **RNA**→
Xor-1-1	NZ_AAQN01000001.1	-	685152	685092	**RNA**→
env-44	sludgeOz_Contig10536	-	566	506	COG4271 (COG4271)→ **RNA**→ (,,,)→
uba-1-1	NC_006352.1	+	36625	36685	NTP_transf_2 (pfam01909)→ **RNA**→ ←Resolvase (pfam00239)
Pmu-1-1	NC_004772.1	+	1341	1401	NTP_transf_2 (pfam01909)→ **RNA**→ PenP (COG2367)→
Sso-1-1	NC_007384.1	-	4110457	4110397	NTP_transf_2 (pfam01909)→ **RNA**→ ←transposition regulatory protein
Pae-1-2	NC_008357.1	-	64005	63945	NTP_transf_2 (pfam01909)→ **RNA**→ hypo→
Eco-1-9	NC_002525.1	+	14615	14675	NTP_transf_2 (pfam01909)→ **RNA**→ hypo→

uba-1-2	NC_006385.1	-	17103	17043	NTP_transf_2 (pfam01909)→ **RNA**→ putative acyl-carrier protein→
Eco-1-10	NC_007635.1	+	8218	8278	←Aph (COG3231) **RNA**→ Fimbrial_CS1 (pfam04449)→
Asa-1-1	NC_009349.1	-	91503	91443	NTP_transf_2 (pfam01909)→ **RNA**→ Multi_Drug_Res (pfam00893)→
Sen-3-4	NC_006856.1	-	39722	39662	NTP_transf_2 (pfam01909)→ **RNA**→ EmrE (COG2076)→
Sen-2-1	NZ_ABAL01000047.1	-	11290	11230	**RNA**→
Eco-1-11	NC_004998.1	-	80976	80916	NTP_transf_2 (pfam01909)→ **RNA**→ Multi_Drug_Res (pfam00893)→
Eco-1-12	NC_009133.1	-	13535	13475	NTP_transf_2 (pfam01909)→ **RNA**→ Multi_Drug_Res (pfam00893)→
Eco-2-1	NZ_AAJX01000089.1	-	2614	2554	NTP_transf_2 (pfam01909)→ **RNA**→ DHPS (cd00739)→
Sen-1-1	NC_009140.1	+	126896	126956	NTP_transf_2 (pfam01909)→ **RNA**→ Antibiotic_NAT (pfam02522)→
Neu-1-1	NC_008344.1	+	1437538	1437598	hypo→ **RNA**→ conserved hypothetical protein; putative conserved domain→
env-45	soil_XZS70175_g1	+	531	591	hypo→ **RNA**→ hypo→
env-46	sludgePhrap_Contig14656	+	235	295	(,,,)→ **RNA**→ (,,,)→
env-47	sludgePhrap_Contig10490	-	1302	1242	(,,,)→ **RNA**→ **RNA**→ (,,,)→
env-103	sludgePhrap_Contig10490	-	940	880	(shown above)
Eco-1-13	NC_006671.1	+	90832	90892	WbbJ (COG0110)→ **RNA**→ NTP_transf_2 (pfam01909)→
env-48	sludgeOz_Contig112	-	826	765	**RNA**→ (,,,)→
env-49	soil_XZS11214_g1	-	459	399	**RNA**→ hypo→
Sen-3-5	NC_006856.1	-	42983	42923	SerB (COG0560)→ **RNA**→ NTP_transf_2 (pfam01909)→
Sen-2-2	NZ_ABAL01000047.1	-	14551	14491	**RNA**→
Eco-4-2	NZ_ABAQ01000017.1	+	3173	3233	**RNA**→
env-50	soil_XZS15696_g1	+	336	444	**RNA**→
env-51	sludgePhrap_Contig9073	+	157	215	**RNA**→ **RNA**→ (,,,)→
env-28	sludgePhrap_Contig9073	+	525	640	(shown above)
env-52	sludgePhrap_Contig10923	-	1299	1239	**RNA**→ **RNA**→ (,,,)→
env-93	sludgePhrap_Contig10923	-	940	880	(shown above)
Xca-1-3	NC_007086.1	-	411457	411397	←xanthomonadin biosynthesis related protein 1 **RNA**→ hypo→
Xca-2-3	NC_003902.1	-	410863	410803	←xanthomonadin biosynthesis related protein 1 **RNA**→ hypo→
Xca-1-4	NC_007086.1	-	409028	408968	←xanthomonadin biosynthesis related protein **RNA**→ ←xanthomonadin biosynthesis related protein 1
Xca-2-4	NC_003902.1	-	408434	408374	←xanthomonadin biosynthesis related protein **RNA**→ ←xanthomonadin biosynthesis related protein 1
Xca-1-5	NC_007086.1	-	405341	405281	←xanthomonadin biosynthesis related protein 1 **RNA**→ hypo→
Xca-2-5	NC_003902.1	-	404756	404696	←xanthomonadin biosynthesis related protein 1 **RNA**→ ←xanthomonadin biosynthesis related protein 1
Xca-1-6	NC_007086.1	-	413002	412942	←xanthomonadin biosynthesis related protein **RNA**→ **RNA**→ ←xanthomonadin biosynthesis related protein 1
Xca-1-9	NC_007086.1	-	411797	411737	(shown above)
Xca-2-6	NC_003902.1	-	412408	412348	←xanthomonadin biosynthesis related protein **RNA**→ **RNA**→ ←xanthomonadin biosynthesis related protein 1
Xca-2-9	NC_003902.1	-	411203	411143	(shown above)
Xca-1-7	NC_007086.1	-	413342	413282	hypo→ **RNA**→ ←xanthomonadin biosynthesis related protein
Xca-2-7	NC_003902.1	-	412748	412688	hypo→ **RNA**→ ←xanthomonadin biosynthesis related protein
Xca-1-8	NC_007086.1	-	405681	405621	hypo→ **RNA**→ ←xanthomonadin biosynthesis related protein 1
Xca-2-8	NC_003902.1	-	405096	405036	hypo→ **RNA**→ ←xanthomonadin biosynthesis related protein 1
env-53	sludgePhrap_Contig12098	-	1297	1237	(,,,)→ **RNA**→ (,,,)→
env-54	sludgePhrap_Contig15715	+	1227	1287	(,,,)→ **RNA**→ ←(,,,)
env-55	sludgePhrap_Contig15602	+	1973	2033	(,,,)→ **RNA**→ **RNA**→ **RNA**→ (,,,)→
env-59	sludgePhrap_Contig15602	+	1610	1670	(shown above)
env-79	sludgePhrap_Contig15602	+	1270	1330	(shown above)
env-56	sludgePhrap_Contig15739	-	2456	2395	(,,,)→ **RNA**→ **RNA**→ (,,,)→
env-60	sludgePhrap_Contig15739	-	2090	2030	(shown above)
env-57	sludgePhrap_Contig15602	+	248	306	(,,,)→ **RNA**→ ←(,,,)
env-58	sludgePhrap_Contig1119	+	530	590	(,,,)→ **RNA**→ (,,,)→
env-61	sludgePhrap_Contig14698	+	598	658	(,,,)→ **RNA**→ COG3791 (COG3791)→
env-62	sludgePhrap_Contig9738	-	1318	1258	(,,,)→ **RNA**→ (,,,)→
env-63	sludgePhrap_Contig3120	+	51	111	**RNA**→ (,,,)→
env-64	sludgePhrap_Contig14698	+	1513	1570	(,,,)→ **RNA**→

env-65	sludgePhrap_Contig15830	-	1814	1756	(,,,)→ **RNA**→ (,,,)→
env-66	sludgePhrap_Contig15677	+	2486	2544	(,,,)→ **RNA**→ **RNA**→ (,,,)→
env-88	sludgePhrap_Contig15677	+	2849	2909	(shown above)
env-67	sludgePhrap_Contig15602	+	2824	2882	(,,,)→ **RNA**→ (,,,)→
env-68	sludgePhrap_Contig15677	+	1559	1617	(,,,)→ **RNA**→ (,,,)→
env-70	sludgePhrap_Contig15677	+	115	173	**RNA**→ **RNA**→ (,,,)→
env-84	sludgePhrap_Contig15677	+	478	538	(shown above)
env-71	sludgePhrap_Contig10490	-	549	491	(,,,)→ **RNA**→ **RNA**→ (,,,)→
env-87	sludgePhrap_Contig10490	-	192	132	(shown above)
env-72	sludgePhrap_Contig15677	+	3657	3717	(,,,)→ **RNA**→
env-33	sludgePhrap_Contig15677	+	3243	3355	(shown above)
env-73	sludgePhrap_Contig15739	-	4007	3947	(,,,)→ **RNA**→ (,,,)→
env-74	sludgePhrap_Contig15677	+	1147	1207	(,,,)→ **RNA**→ (,,,)→
env-75	sludgePhrap_Contig15739	-	2857	2797	(,,,)→ **RNA**→ (,,,)→
env-76	sludgePhrap_Contig3140	-	597	537	**RNA**→ (,,,)→
env-77	sludgePhrap_Contig14656	+	2323	2383	(,,,)→ **RNA**→
env-78	sludgePhrap_Contig15602	+	756	816	←(,,,) **RNA**→ (,,,)→
env-80	sludgePhrap_Contig3120	+	428	488	(,,,)→ **RNA**→
env-83	sludgePhrap_Contig15830	-	3768	3708	**RNA**→ (,,,)→
env-85	sludgePhrap_Contig14656	+	1991	2051	(,,,)→ **RNA**→ (,,,)→
env-86	sludgePhrap_Contig15739	-	3189	3129	(,,,)→ **RNA**→ **RNA**→ (,,,)→
env-112	sludgePhrap_Contig15739	-	3551	3493	(shown above)
env-89	sludgePhrap_Contig12098	-	429	369	(,,,)→ **RNA**→ **RNA**→ (,,,)→
env-101	sludgePhrap_Contig12098	-	794	734	(shown above)
env-90	sludgePhrap_Contig9738	-	442	382	(,,,)→ **RNA**→
env-69	sludgePhrap_Contig9738	-	805	747	(shown above)
env-91	sludgePhrap_Contig15739	-	1625	1567	(,,,)→ **RNA**→ (,,,)→
env-92	sludgePhrap_Contig14771	-	188	128	(,,,)→ **RNA**→
env-94	sludgePhrap_Contig15715	+	2517	2577	(,,,)→ **RNA**→ (,,,)→
env-95	sludgePhrap_Contig15739	-	874	814	(,,,)→ **RNA**→ (,,,)→
env-96	sludgePhrap_Contig15107	-	777	717	(,,,)→ **RNA**→ (,,,)→
env-97	sludgePhrap_Contig15739	-	1278	1218	(,,,)→ **RNA**→ (,,,)→
env-98	sludgePhrap_Contig3140	-	235	175	(,,,)→ **RNA**→
env-99	sludgePhrap_Contig15715	+	2128	2188	(,,,)→ **RNA**→ (,,,)→
env-100	sludgePhrap_Contig15107	-	316	256	(,,,)→ **RNA**→
env-102	sludgePhrap_Contig15715	+	869	929	(,,,)→ **RNA**→ (,,,)→
env-104	sludgePhrap_Contig15715	+	1636	1696	←(,,,) **RNA**→ (,,,)→
env-105	sludgePhrap_Contig15830	-	1375	1315	(,,,)→ **RNA**→ (,,,)→
env-106	sludgeOz_Contig3418	-	1637	1576	**RNA**→ COG2856 (COG2856)→
env-107	soil_Contig12933	-	836	776	**RNA**→ hypo→
env-108	soil_Contig20109	+	1142	1202	hypo→ **RNA**→ COG5515 (COG5515)→
Asp-1-1	NC_000923.1	+	1175	1235	hypo→ **RNA**→ ←COG1357 (COG1357)
Kpn-1-1	NC_009651.1	-	37673	37613	aminoglycoside adenylyltransferase→ **RNA**→ DGPF (pfam04946)→
Sen-3-6	NC_006856.1	-	39211	39151	EmrE (COG2076)→ **RNA**→ ←Transposase_mut (pfam00872)
Sen-2-3	NZ_ABAL01000047.1	-	10779	10719	**RNA**→
Sdy-1-1	NZ_AAMJ01000001.1	+	14770	14830	Acetyltransf_1 (pfam00583)→ **RNA**→ hypo→
Cgl-1-2	NC_003227.1	+	27610	27670	NTP_transf_2 (pfam01909)→ **RNA**→ ←Multi_Drug_Res (pfam00893)
Csp-1-1	NC_009128.1	-	3095	3035	NTP_transf_2 (pfam01909)→ **RNA**→ Multi_Drug_Res (pfam00893)→
env-109	sludgePhrap_Contig14656	+	1183	1243	(,,,)→ **RNA**→ (,,,)→
env-110	sludgeOz_Contig10536	-	1093	1033	(,,,)→ **RNA**→ COG4271 (COG4271)→
env-111	sludgeOz_Contig1153	+	950	1009	←(,,,) **RNA**→
Xca-1-10	NC_007086.1	-	407734	407674	←xanthomonadin biosynthesis related protein 1 **RNA**→ hypo→

Xca-2-10	NC_003902.1	-	407140	407080	←xanthomonadin biosynthesis related protein 1 **RNA**→ hypo→
Xca-1-11	NC_007086.1	-	402732	402672	hypo→ **RNA**→ Acetyltransf_1 (pfam00583)→
Xca-2-11	NC_003902.1	-	402147	402087	hypo→ **RNA**→ Acetyltransf_1 (pfam00583)→
Kpn-1-2	NC_009651.1	-	36967	36907	DGPF (pfam04946)→ **RNA**→ AraJ (COG2814)→
Xax-1-1	NC_003919.1	-	2937644	2937584	hypo→ **RNA**→ **RNA**→ transposase→
Xax-1-2	NC_003919.1	-	2937253	2937193	(shown above)
env-113	sludgeOz_Contig112	-	297	237	(,,,)→ **RNA**→
Xca-1-13	NC_007086.1	-	409885	409825	hypo→ **RNA**→ hypo→
Xca-2-13	NC_003902.1	-	409291	409231	hypo→ **RNA**→ hypo→
env-114	soil_Contig9569	-	1152	1092	**RNA**→ hypo→
env-115	soil_Contig3259	-	480	420	**RNA**→ hypo→
env-116	soil_Contig39158	+	1053	1113	hypo→ **RNA**→
env-117	soil_Contig39158	+	678	738	hypo→ **RNA**→ hypo→
env-118	soil_Contig9569	-	136	76	hypo→ **RNA**→
env-119	soil_Contig9569	-	647	587	(shown above)
env-120	soil_Contig12942	+	673	733	hypo→ **RNA**→ hypo→
env-121	soil_Contig12942	+	291	351	**RNA**→ hypo→
env-122	soil_Contig12933	-	434	374	hypo→ **RNA**→
env-123	soil_Contig26669	-	803	743	**RNA**→ hypo→
env-124	sludgePhrap_Contig14698	+	57	115	**RNA**→ (,,,)→
env-125	sludgePhrap_Contig15715	+	436	494	(,,,)→ **RNA**→ (,,,)→
env-126	sludgePhrap_Contig10923	-	376	318	(,,,)→ **RNA**→ (,,,)→
env-127	sludgePhrap_Contig15830	-	3314	3253	(,,,)→ **RNA**→ (,,,)→
Neu-1-2	NC_008344.1	-	1103939	1103879	Qor (COG0604)→ **RNA**→ transposase IS3/IS911 family protein→
Mfl-1-1	NC_007947.1	-	416298	416238	CzcD (COG1230)→ **RNA**→ COG4886 (COG4886)→
Cau-1-1	NZ_AAAH02000041.1	+	38722	38782	COG1359 (COG1359)→ **RNA**→ hypo→
env-128	sludgePhrap_Contig15830	-	2292	2232	(,,,)→ **RNA**→ (,,,)→
env-129	sludgePhrap_Contig16237	-	5795	5699	(,,,)→ **RNA**→ (,,,)→

99.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 99.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Transposase-resistance RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

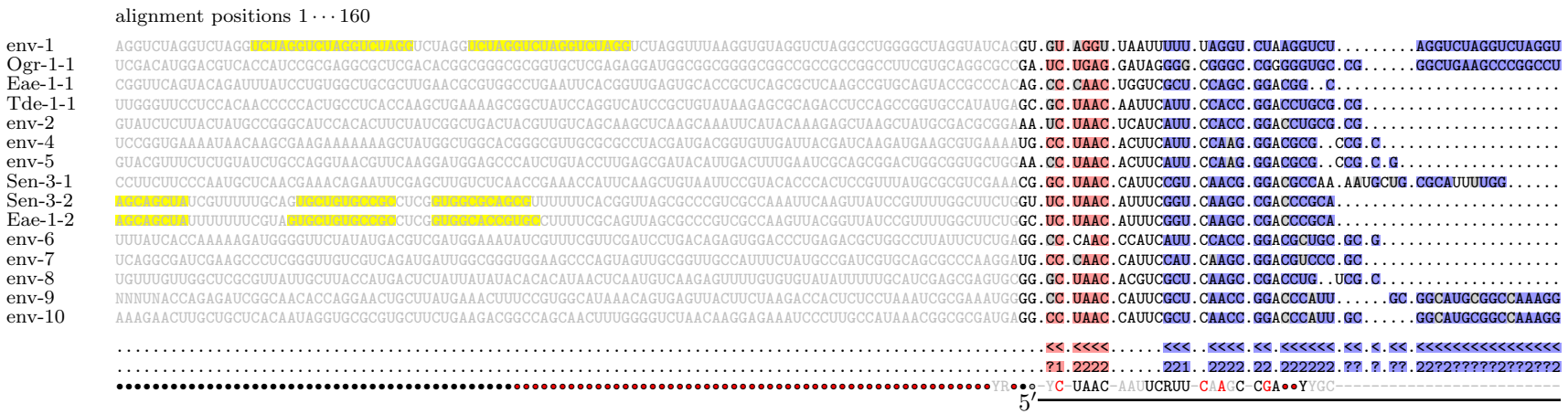
u cd00209 (3) Dihydrofolate reductase (DHFR).
D cd00739 (4) DHPS subgroup of Pterin binding enzymes.
u cd01193 (1) IntI (E2) integrases, site-specific tyrosine recombinases, DNA breaking-rejoining enzymes, N- and C-terminal domains.
u COG0110 (1) Acetyltransferase (isoleucine patch superfamily) [General function prediction only]
u COG0560 (1) Phosphoserine phosphatase [Amino acid transport and metabolism]
d COG0562 (1) UDP-galactopyranose mutase [Cell envelope biogenesis, outer membrane]
u COG0604 (1) NADPH:quinone reductase and related Zn-dependent oxidoreductases [Energy production and conversion / General function prediction only]
u COG0717 (1) Deoxycytidine deaminase [Nucleotide transport and metabolism]
u COG1192 (1) ATPases involved in chromosome partitioning [Cell division and chromosome partitioning]
u COG1230 (1) Co/Zn/Cd efflux system component [Inorganic ion transport and metabolism]
d COG1357 (1) Uncharacterized low-complexity proteins [Function unknown]
u COG1359 (1) Uncharacterized conserved protein [Function unknown]
u COG1479 (1) Uncharacterized conserved protein [Function unknown]
uD COG2076 (3) Membrane transporters of cations and cationic drugs [Inorganic ion transport and metabolism]
uD COG2318 (2) Uncharacterized protein conserved in bacteria [Function unknown]
D COG2367 (1) Beta-lactamase class A [Defense mechanisms]
D COG2814 (1) Arabinose efflux permease [Carbohydrate transport and metabolism]
D COG2856 (1) Predicted Zn peptidase [Amino acid transport and metabolism]
u COG3231 (1) Aminoglycoside phosphotransferase [Translation, ribosomal structure and biogenesis]
D COG3344 (1) Retron-type reverse transcriptase [DNA replication, recombination, and repair]
uD COG3791 (2) Uncharacterized conserved protein [Function unknown]
uD COG4271 (2) Predicted nucleotide-binding protein containing TIR -like domain [Transcrip-

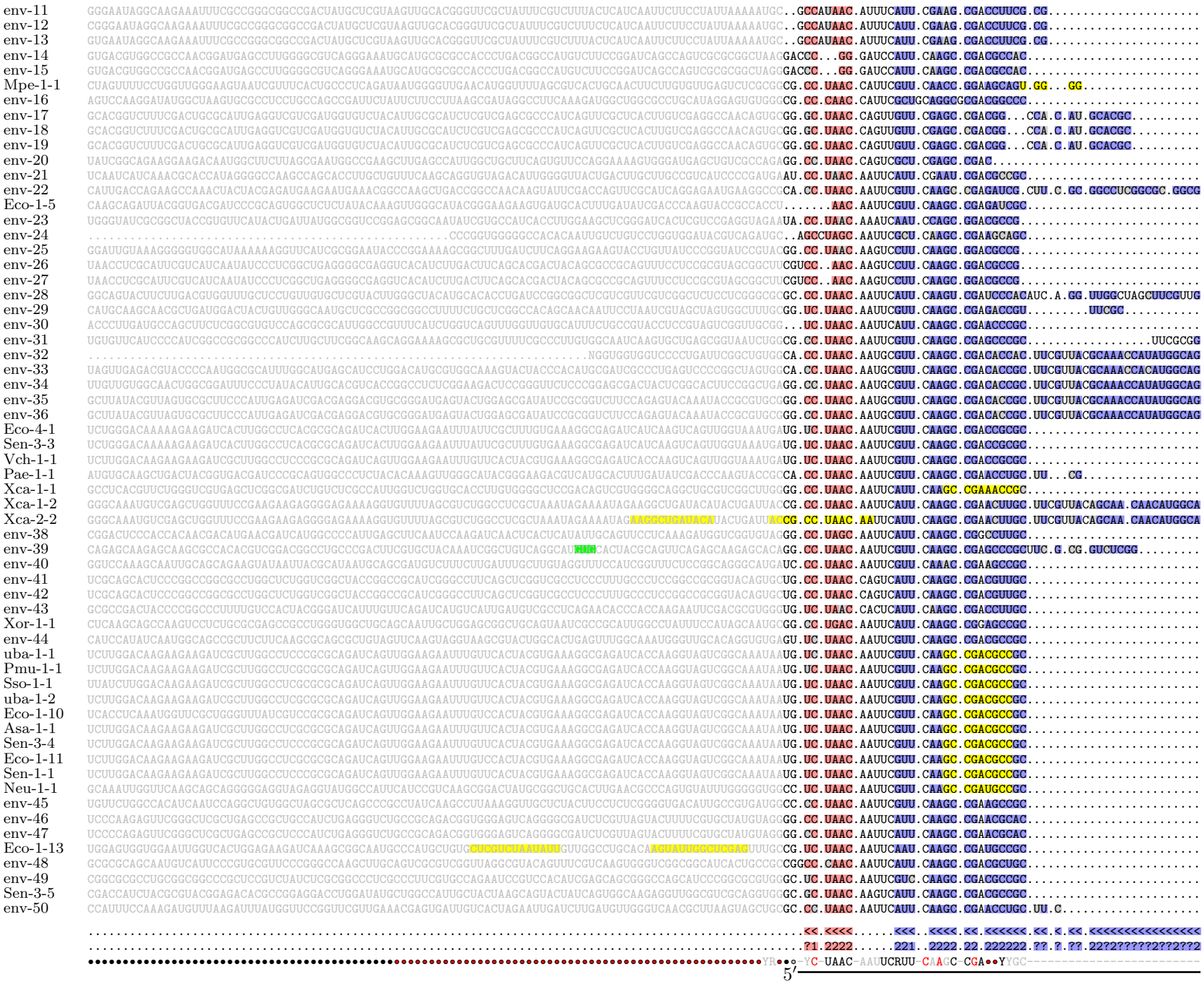
tion]
u COG4333 (1) Uncharacterized protein conserved in bacteria [Function unknown]
D COG4886 (1) Leucine-rich repeat (LRR) protein [Function unknown]
D COG5515 (1) Uncharacterized conserved small protein [Function unknown]
d pfam00239 (1) Resolvase, N terminal domain.
uD pfam00583 (5) Acetyltransferase (GNAT) family.
d pfam00872 (1) Transposase, Mutator family.
Dd pfam00893 (13) Small Multidrug Resistance protein.
uD pfam00903 (2) Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily.
uD pfam01909 (32) Nucleotidyltransferase domain.
uD pfam02517 (2) CAAX amino terminal protease family.
D pfam02522 (1) Aminoglycoside 3-N-acetyltransferase.
uD pfam03595 (2) C4-dicarboxylate transporter/malic acid transport protein.
D pfam04449 (1) CS1 type fimbrial major subunit.
uD pfam04946 (2) DGPF domain.
D smart00255 (1) Toll - interleukin 1 - resistance;

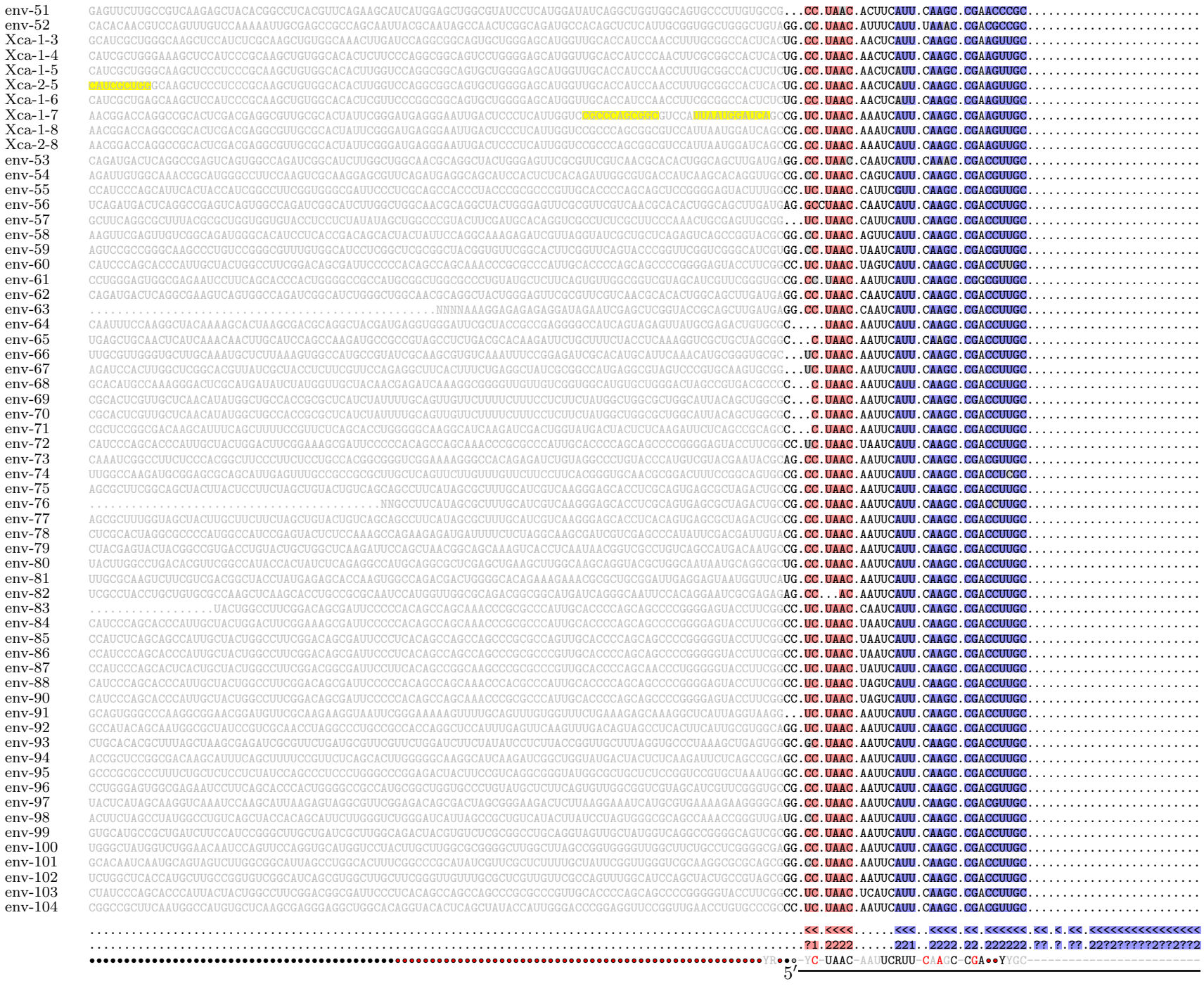
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Apu-1-1, Cfr-1-1, Cfr-1-2, Cfr-1-3, Cgl-1-1, Eco-1-1, Eco-1-12, Eco-1-2, Eco-1-3, Eco-1-4, Eco-1-6, Eco-1-7, Eco-1-8, Eco-1-9, Eco-2-1, Eco-3-1, Eco-4-2, Pae-1-2, Sen-2-1, Sen-2-2, Sen-2-3, Sen-2-4, Sen-2-5, Sen-2-6, Xca-2-1, Xca-2-10, Xca-2-11, Xca-2-12, Xca-2-13, Xca-2-3, Xca-2-4, Xca-2-6, Xca-2-7, Xca-2-9, Ype-1-1, Ype-1-2, env-3, env-37, env-89

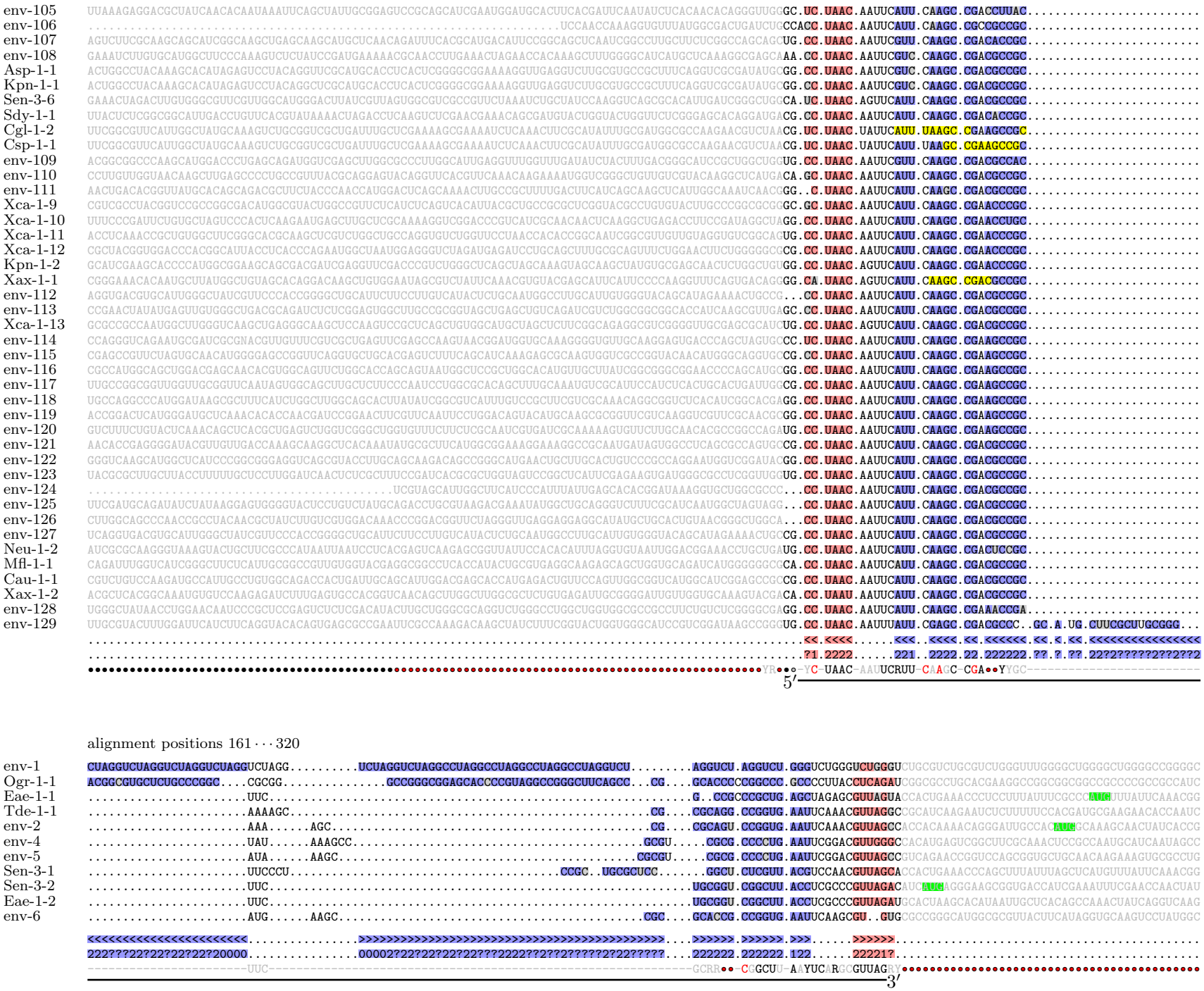
99.4 Multiple-sequence alignment

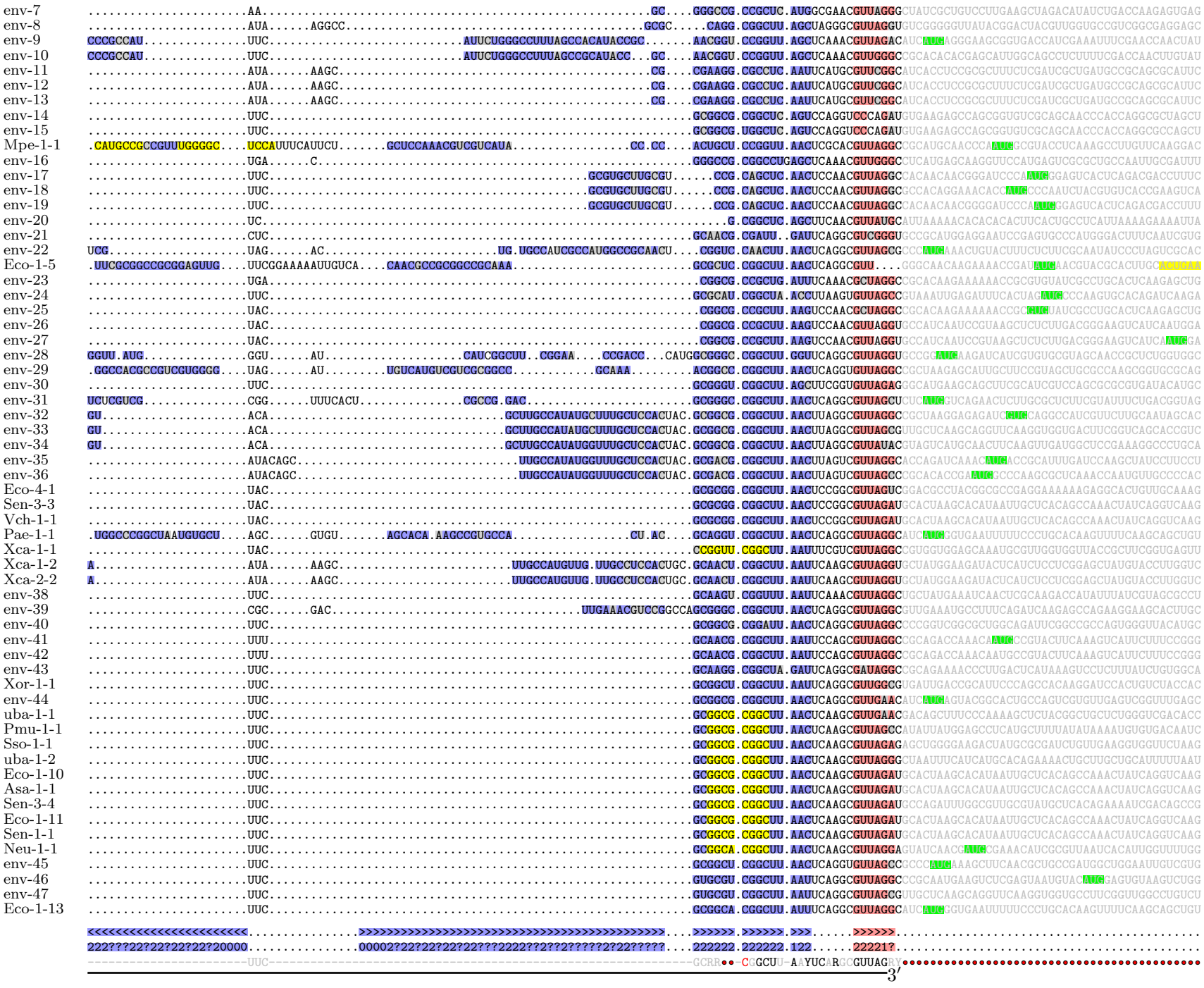
Each Transposase-resistance RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 99.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Transposase-resistance RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

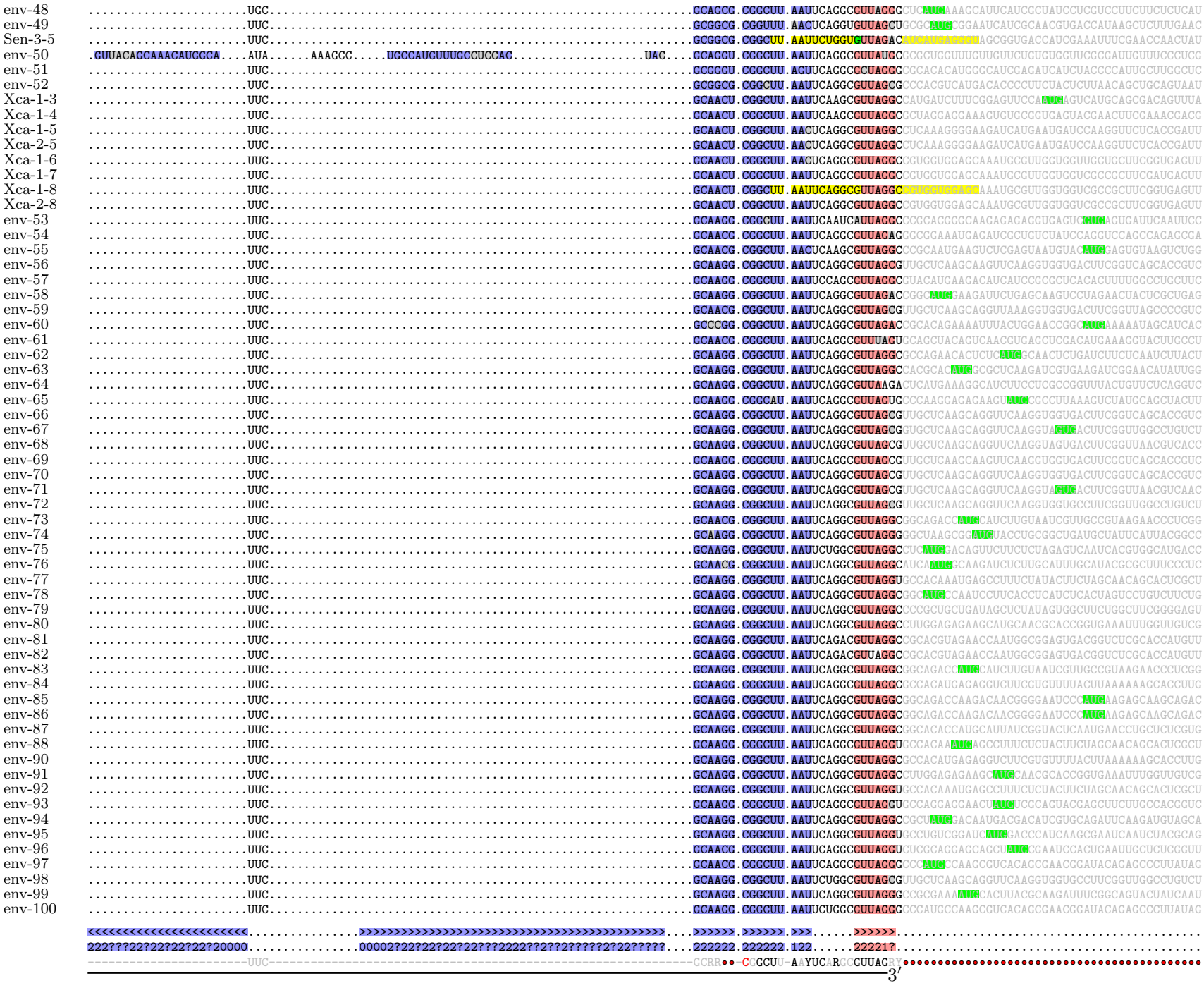


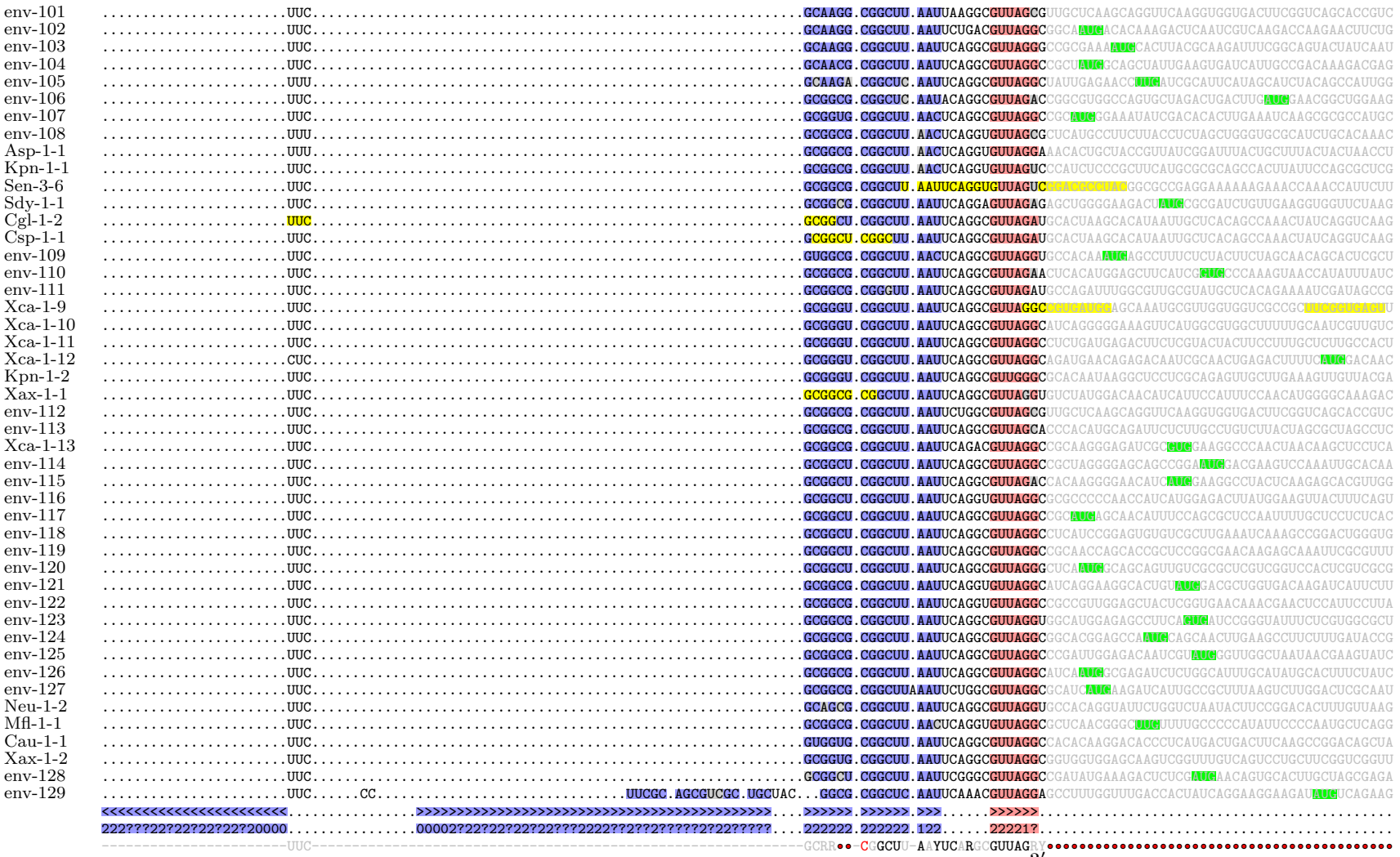

env-11
env-12
env-13
env-14
env-15
Mpe-1-1
env-16
env-17
env-18
env-19
env-20
env-21
env-22
Eco-1-5
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
Eco-4-1
Sen-3-3
Vch-1-1
Pae-1-1
Xca-1-1
Xca-1-2
Xca-2-2
env-38
env-39
env-40
env-41
env-42
env-43
Xor-1-1
env-44
uba-1-1
Pmu-1-1
Sso-1-1
uba-1-2
Eco-1-10
Asa-1-1
Sen-3-4
Eco-1-11
Sen-1-1
Neu-1-1
env-45
env-46
env-47
Eco-1-13
env-48
env-49
Sen-3-5
env-50
5′

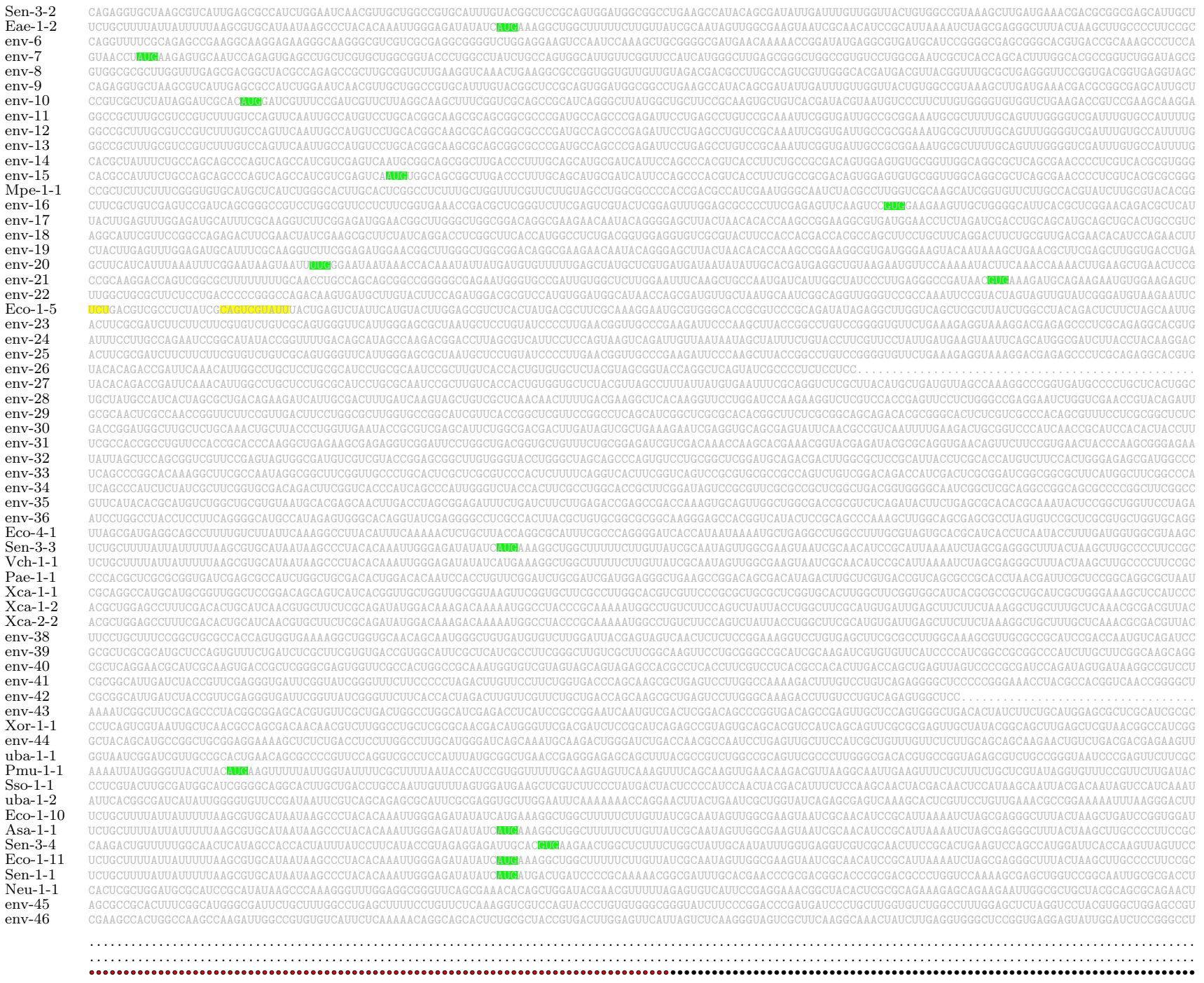

env-51
env-52
Xca-1-3
Xca-1-4
Xca-1-5
Xca-2-5
Xca-1-6
Xca-1-7
Xca-1-8
Xca-2-8
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
5′


env-105
env-106
env-107
env-108
Asp-1-1
Kpn-1-1
Sen-3-6
Sdy-1-1
Cgl-1-2
Csp-1-1
env-109
env-110
env-111
Xca-1-9
Xca-1-10
Xca-1-11
Xca-1-12
Kpn-1-2
Xax-1-1
env-112
env-113
Xca-1-13
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
Neu-1-2
Mfl-1-1
Cau-1-1
Xax-1-2
env-128
env-129
5′
alignment positions 161 · · · 320
env-1
Ogr-1-1
Eae-1-1
Tde-1-1
env-2
env-4
env-5
Sen-3-1
Sen-3-2
Eae-1-2
env-6
3′

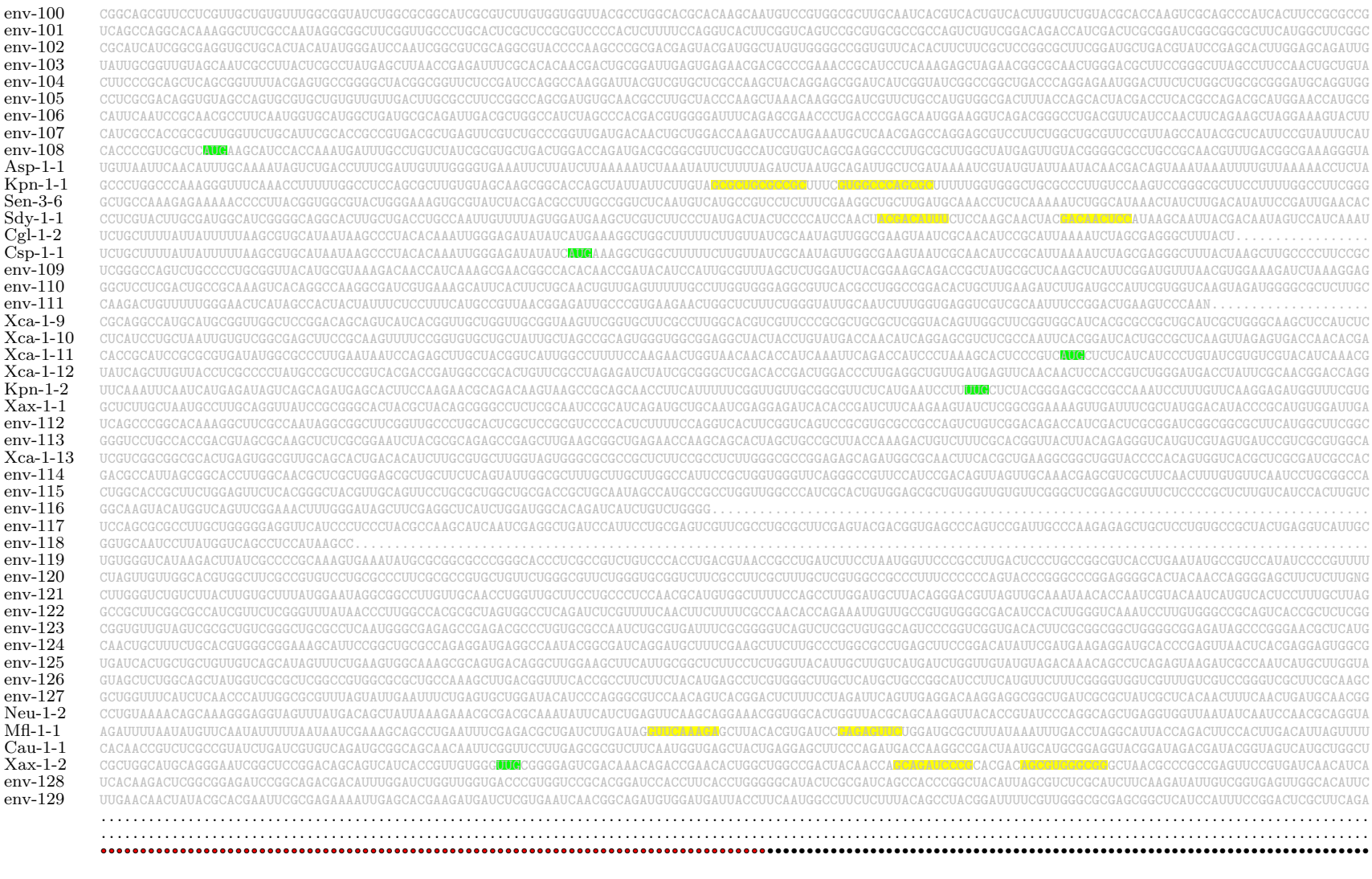

env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-14
env-15
Mpe-1-1
env-16
env-17
env-18
env-19
env-20
env-21
env-22
Eco-1-5
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
Eco-4-1
Sen-3-3
Vch-1-1
Pae-1-1
Xca-1-1
Xca-1-2
Xca-2-2
env-38
env-39
env-40
env-41
env-42
env-43
Xor-1-1
env-44
uba-1-1
Pmu-1-1
Sso-1-1
uba-1-2
Eco-1-10
Asa-1-1
Sen-3-4
Eco-1-11
Sen-1-1
Neu-1-1
env-45
env-46
env-47
Eco-1-13
3′

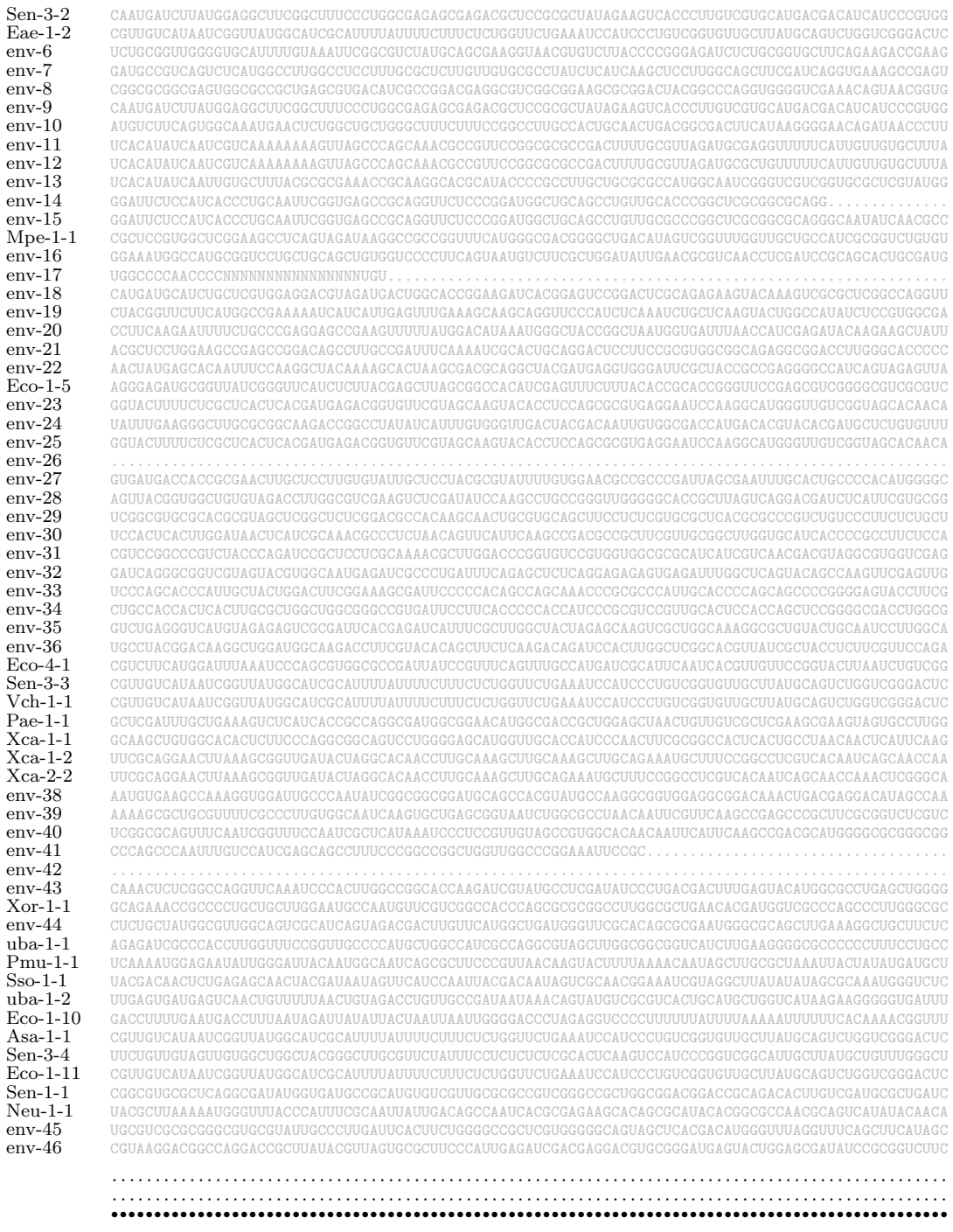

env-48
env-49
Sen-3-5
env-50
env-51
env-52
Xca-1-3
Xca-1-4
Xca-1-5
Xca-2-5
Xca-1-6
Xca-1-7
Xca-1-8
Xca-2-8
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
3′


env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
Asp-1-1
Kpn-1-1
Sen-3-6
Sdy-1-1
Cgl-1-2
Csp-1-1
env-109
env-110
env-111
Xca-1-9
Xca-1-10
Xca-1-11
Xca-1-12
Kpn-1-2
Xax-1-1
env-112
env-113
Xca-1-13
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
Neu-1-2
Mfl-1-1
Cau-1-1
Xax-1-2
env-128
env-129
3′
alignment positions 321 · · · 480
env-1
Ogr-1-1
Eae-1-1
Tde-1-1
env-2
env-4
env-5
Sen-3-1


Sen-3-2
Eae-1-2
env-6
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-14
env-15
Mpe-1-1
env-16
env-17
env-18
env-19
env-20
env-21
env-22
Eco-1-5
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
Eco-4-1
Sen-3-3
Vch-1-1
Pae-1-1
Xca-1-1
Xca-1-2
Xca-2-2
env-38
env-39
env-40
env-41
env-42
env-43
Xor-1-1
env-44
uba-1-1
Pmu-1-1
Sso-1-1
uba-1-2
Eco-1-10
Asa-1-1
Sen-3-4
Eco-1-11
Sen-1-1
Neu-1-1
env-45
env-46

env-47
Eco-1-13
env-48
env-49
Sen-3-5
env-50
env-51
env-52
Xca-1-3
Xca-1-4
Xca-1-5
Xca-2-5
Xca-1-6
Xca-1-7
Xca-1-8
Xca-2-8
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99


env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
Asp-1-1
Kpn-1-1
Sen-3-6
Sdy-1-1
Cgl-1-2
Csp-1-1
env-109
env-110
env-111
Xca-1-9
Xca-1-10
Xca-1-11
Xca-1-12
Kpn-1-2
Xax-1-1
env-112
env-113
Xca-1-13
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
Neu-1-2
Mfl-1-1
Cau-1-1
Xax-1-2
env-128
env-129
alignment positions 481 · · · 577
env-1
Ogr-1-1
Eae-1-1
Tde-1-1
env-2
env-4
env-5
Sen-3-1


Sen-3-2
Eae-1-2
env-6
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-14
env-15
Mpe-1-1
env-16
env-17
env-18
env-19
env-20
env-21
env-22
Eco-1-5
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
Eco-4-1
Sen-3-3
Vch-1-1
Pae-1-1
Xca-1-1
Xca-1-2
Xca-2-2
env-38
env-39
env-40
env-41
env-42
env-43
Xor-1-1
env-44
uba-1-1
Pmu-1-1
Sso-1-1
uba-1-2
Eco-1-10
Asa-1-1
Sen-3-4
Eco-1-11
Sen-1-1
Neu-1-1
env-45
env-46

```
env-47      AUCCCAGCACCCAUUACUACUGGCCUUCGGACGGCGAUUCCCUCACAGCCAGCCAGCCCGCGCCCGUUGCACCCCAGCAGCCCCGGGGGUACCUUCG
Eco-1-13    GCUCGAUUUGCUGAAAGUCUCAUCACCGCCAGGCGAUGGCGGAACAUGGCGACCGCUGGAGCUAACUGUUGUCGCUCGAAGCGAAGUAGUGCCUUGG
env-48      CGCACUCGUUGCAGCAGAGGUAGUAGCUGCCGCGGCUGGCCAUCCCAGCCCCGACUUGCCAGACGAUCUGAACGCCUGGAUUGCCCGCCAAUCAGCG
env-49      CGGUACGAGCAGCCCGGCAAAGGCACCGCGUACGCACGUUUACUGCAUCCAAUCUUCGCAUGGCAUUAUUUGCAGGCGUCUGGGCCGCUGCUUGGUU
Sen-3-5     CAAUGACCUUAUGGAGGCUUCGGCUUUCCCUGGCGAGAGCGAGACGCUCCGCGCUAUAGAAGUCACCCUUGUCGUGCAUGACGACAUCAUCCCGUGG
env-50      GGCCCCGCCCCCGAUUGGCCUGCAAGGCUGGUUGGGCUUCGCUUGGAUCAGCUCAAACAGCUGCAAAAAACUCGUUAAGGCCAAACCCCCUUCCAA
env-51      CCUUCUGGCAGUACUUCUUGACGUGGUUUGCUCCUGUUGUGCUCGUACUUGGGCUACAUGCACACCUGAUCCGGCGGCUCGUCGUUCGUCGGCUCUC
env-52      CACGCUUUAGCUAAGCGAGAUCGCGUUCUGAUGCGUUCGUUCUGGAUCUUCUAUAUCCUCUUACCGGUUGCUUUAGGUGCCCUAAAGCUGAGUGGGC
Xca-1-3     CAAGCGAUAAUCAACAAGCAGGUUCUUAAUAACUGCGACUUGCUUAUAGGAGUUUUCUGGACGCGAAUCGGCACUCCAACCGGCGACUACGCUAGCG
Xca-1-4     GCAUCCCGAUCUGGAUAGUUGUCUGGCAGUACGUACAGCCAUACUUGCUGUAUGGGCCAGACUCCUAUGCUCACUUUCGGAAAGUGGGAAGCAUUGG
Xca-1-5     AGCUCAUCCUAGCGAUGGGAUCCGUUUUCAUCAUUAACUUAAUUUUUUAUUUCUUAUCAGGGUCGCUUCCGCCGAUCGCAUUGAUAGCAGACAAGAU
Xca-2-5     AGCUCAUCCUAGCGAUGGGAUCCGUUUUCAUCAUUAACUUAAUUUUUUAUUUCUUAUCAGGGUCGCUUCCGCCGAUCGCAUUGAUAGCAGACAAGAU
Xca-1-6     GCAAGCUGUGGCACACUCGGUCAAGGCGGCAGUGCUGGGGAGCAGGGUUGCACCAUCCAACCUUCGCGGCCACUCACUGCCUAACAACUCAUUCAAG
Xca-1-7     CAAGCUGUGGCACACUCGUUCCCGGCGGCAGUGCUGGGGAGCAUGGUUGCACCAUCCAACCUUCGCGGCCACUCUCUGCCUAACAACUCAUUCAAGC
Xca-1-8     CAAGCUGUGGCACACUUGGUCCAGGCGGCAGUGCUGGGGAGCAUGGUUGCACCAUCCAACCUUUGCGGCCACUCUCUGCCUAACAACUCAUUCAAGC
Xca-2-8     CAAGCUGUGGCACACUUGGUCCAGGCGGCAGUGCUGGGGAGCAUGGUUGCACCAUCCAACCUUUGCGGCCACUCACUGCCUAACAACUCAUUCAAGC
env-53      UUGCUUGCGUUUACUUACGUGGCAUCGGCAUGCGCGCUGUUUUUUCCAACGUACUGCACCUCUUGGUUUGGUGCUCCUUGCCCCGGUUGCAGUCAUCA
env-54      UGCAGAGCACAAUCUUUCAGGCCCCAGAUUUUCGCGACCAUCUAGCGGCCGCUUCAAUGGCCAUGUGCUCAAGCGAGGGAGGCUGGCACAGGUACAC
env-55      CGUAAGGACGGCCAGGACCGCUUAUACGUUAGUGCGCUUCCCAUUGAGAUCGACGAGGACGUGCGGGAUGAGUACUGGAGCGAUAUCCGCGGUCUUC
env-56      CCAUCCCAGCACCCAUUGCUACUGGCCUUCGGACAGCGAUUCCCCCACAGCCAGCAAACCCGCGCCCAUUGCACCCCAGCAGCCCCGGGGAGUACCU
env-57      CCCAGAUCAAAUAUCAAAUUGUUGGUAGAGCUGAUCCAAGCGAGUGUAGUGAUUCGCCUUGCACUGCCUUGGCACUCCGGCGUGCAGAGCUUUUUCA
env-58      GCACGCUCCAGAUCAAGGGGCGCUGCCUCCCCACGGUUGCACGCCUGGGCAGCGGCUUGGCUCAAUCCAGCGGACAAAGAGUGGGACGCCGUCUUAU
env-59      CAUCCCAGCAUUCACUACCAUCGGCCUUCGGUGGGCGAUUCCCUCGCAGCCACCCUACCCGCGCCCGUUGCACCCCAGCAGCUCCGGGGAGUACUUU
env-60      CUUUCAGCGUGCGACCGCUACUACUUGUCGCUGCUCGAUGCACAAAAAUUGGAUCUUAAGCAACGAGCAGGCGCUUCUAUCGAGUCUCAACAAAGAG
env-61      UGGGCGGAACGCGGGUUCUGUGACAAGUGCGGGACGCAUCUGUUUUAUCACCUACUCCCUGGGAACCAAUACAUCCUUCCGGCUGGCAUCUUUCAGG
env-62      GUUUUUGGCUGCGCUCAAGCGAUCAUUGCCACAUCUCUAGUCGGGCGGGUUGUUGCUCGCAUGAUCGCUCGCUCGAUCACCUUGUCGGCUAAGACUC
env-63      AGCCUGAUGAACCUACUUGGCCUGACACGUUCGGCCAUACCACUAUCUCAGAGGCCAUGCAGGCGCUCGAGCUGAAGCUUGGCAAGCAGGUACGCUG
env-64      ....................................................................................................
env-65      GCCUUCGAUACGCGCAAAUGAGACAAACGGAUCCGUGCACAUCGAGGUGUUCUACACAGGGACUCAGACUGGGGUCUUUAAAGAGGACGCUAUCAAC
env-66      CCAUCCCAGCACCCAUUGCUACUGGCCUUCGGACAGCGAUUCCCCCACAGCCAGCAAACCCACGCCCAUUGCACCCCAGCAGCCCCGGGGAGUACCU
env-67      AUCUCAGCAGCCAUUGCUACUGGCCUUCGGACAGCGAUUCCCUCACAGCCAGCCCGCGCCCAGUUGCACCCCAGCAGCCCCGGGGGUACCUUCG
env-68      CCAGCACUCACUGCUACUGGCCUUCGGACGGCGAUUCCUUCACAGCCGGCAAGCCCGCGCCCGUUGCACCCCAGCAACCCUGGGGGUACCUUCGGCC
env-69      CCAUCCCAGCACCCAUUGCUACUGGCCUUCGGACAGCGAUUCCCCCACAGCCAGCAAACCCGCGCCCAUUGCACCCCAGCAGCCCCGGGGAGUACCU
env-70      CCAUCCCAGCACCCAUUGCUACUGGACUUCGGAAAGCGAUUCCCCCACAGCCAGCAAACCCGCGCCCAUUGCACCCCAGCAGCCCCGGGGAGUACCU
env-71      CCAGCACUCACUGCUACUGGCCUUCGGACGGCGAUUCCUUCACAGCCGGCAAGCCCGCGCCCGUUGCACCCCAGCAACCCUGGGGGUACCUUCGGCC
env-72      AUCCCAGCACUCACUUCUACUGGCCUUCGGAGGGCGAUUCCUUCACAGCCAGCCAGCCCGCGCCCGUUGCACCCCAGCAGCCCCGGGGGUACCUUCG
env-73      GGCGGCAAGUACAAAGUGCUGCUAGCUGCCUAUGCAAUUUCUUGUAUGGCGCUUGGCUACACCCUCGUUCGUGACAUUGCUCGCCACCUGUCAGGUG
env-74      GUAUGUGCAAGGCCGCCUUAGAUGCCAGUAUCGGCACUACUGGCAGCGGCACAUGCCAAAGGGACUCGCAUGAUAUCUAUGGUUGCUACAACGAGAU
env-75      CCAGCUCGAGAUCUCUGGAUUCACUUAUACUCAUAGCUCAGAUGACUCAGGCCGAGUCAGUGGCCAGAUCGGCAUCUUGGCUGGCAACGCAGGCUAC
env-76      UUCUAGCCUAUGGCCUGUCAGCUACCACAGCAUUCUUGGGUCUGGGAUCAUUAGCCGCUGUCAUACUUAUCCUAGUGGGCGCAGCCAAACCGGGUUG
env-77      ....................................................................................................
env-78      UGUAUCGUUUAGUGGGGGCGUUGUUGCGGGUCUAUUCGCAAACAGUCCUGACGCUUGCACGCCAAAGCAGUUGUGCCCCGGGGAUAGUCUAGACAAG
env-79      GGCGUUUGUGGCAUCCUCGGCUCGCGGCUACGGUGUUCGGCACUUCGGUUCAGUACCCGGUUCGGUCGGGCAUCGUGGCCUAACUAAUCAUUCAAGC
env-80      ....................................................................................................
env-81      UCGGUCGCCUCCCUUUGCCCUCCGGCCGCGGUACAGUGCUGCCUAACCAGUCAUUCAAGCCGACGUUGCUUUGCAACGCGGCUUAAUUCCAGCGUUA
env-82      UCGGUCGCCUCCCUUUGCCCUCCGGCCGCGGUACAGUGCUGCCUAACCAGUCAUUCAAGCCGACGUUGCUUUGCAACGCGGCUUAAUUCCAGCGUUA
env-83      GGCGGCAAGUACAAAGUGCUGCUAGCUGCCUAUGCAAUUUCUUGUAUGGCGCUUGGCUACACCCUCGUUCGUGACAUUGCUCGCCACCUGUCAGGUG
env-84      GGCUGUUGUUCAGGCGUACUACCCAUCUGCAUCGACUGUAGAUGAUGCAAUCCCGGAUCGUGCCCGGGCCUAUCUCCAGCAGGCUUAUGAGUCGCUU
env-85      GUACUGUCAGCAGCCUUCAUAGCGCUUUGCAUCGUCAAGGGAGCACCUCACAGUGAGCGCUAGACUGCCGCCUAACAAUUCAUUCAAGCCGACCUUG
env-86      GUACUGUCAGCAGCCUUCAUAGCGCUUUGCAUCGUCAAGGGAGCACCUCGCAGUGAGCGCUAGACUGCCGCCUAACAAUUCAUUCAAGCCGACCUUG
env-87      ....................................................................................................
env-88      AGAAGCACGAGCCGGAGACUGGUUACGUCUUAGUUGAGACGUACCCCAAUGGCGCAUUUGGCAUGAGCAUCCUGGACAUGCGUGGCAAAGUACUACC
env-90      GGCUGUUGUUCAGGCGUACUACCCAUCUGCAUCGACUGUAGAUGAUGCAAUCCCGGAUCGUGCCCGGGCCUAUCUCCAGCAGGCUUAUGAGUCGCUU
env-91      CAAAUCCAAGCAUUAAGAGUAGGCGUUCGGAGACAGCGACUAGCGGGAAGACUCUUAAGGAAAUCAUGCGUGAAAAGAAGGGGCAGGCCUAACAAUU
env-92      ....................................................................................................
env-93      CACUAGAUCUCUACCGGGAUGGCAAAGGCUUCCGCUUCAAGUUGACUAACGUCGGAGAGGCCGAGGCUAGAGACGCAGAAAUAAAGUUCGUCCUUCA
env-94      GUGUUCCCGUUUUCAUUCAUUGGCACCGUCACUAUCAGAGCAUUCUUCGGCAACUAGACUCUCUUAAUCGCCGCGUUUCGAGCGGAGAGACUGUCUA
env-95      UCUUAACCUAGACAUUAAUGCUGAACUAUACGUCCCGGAAAAUUUCUCGCCAGCUAACUUCUUUGACCGGCUAAAUACUGUGUGGUGGAUUACGUCG
env-96      GUACGUCCUAUGGCUUCGAUCUCGCAAUGCACUGACAUGGCUCAACGUGUGCCUCGGCUCGGCGGCUGUAGGGGCACUGGCAAUGGCUGUUCUCAGC
env-97      CGCUCUUUUGGCUGUUGCUGUUGUAUGCACGUACUUCCUAGCCCGCGCCCUUUCUGCUCUCCUCUAUCCAGCCCUCCCUGGGCCCGGAGACUACUUC
env-98      ....................................................................................................
env-99      GGCACUUGCAGCUAUGUGUACGAUUACCGCUCCGGCGACAAGCAUUUCAGCUUUUCCGUUCUCAGCACUUGGGGGCAAGGCAUCAAGAUCGGCUGGU

            ....................................................................................................
            ....................................................................................................
            ****************************************************************************************************
```

100 TwoAYGGAY

100.1 Taxa

The taxonomy of each organism containing a putative TwoAYGGAY RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual TwoAYGGAY RNA in Sections 100.2 and 100.4.

abbrev. of hits taxonomy of species

Rin-1-1 to Rin-1-13	Bacteria Firmicutes Clostridia Clostridiales Lachnospiraceae *Roseburia intestinalis* L1-82
Pth-1-1 to Pth-1-4	Bacteria Firmicutes Clostridia Clostridiales Peptococcaceae *Pelotomaculum thermopropionicum* SI
Aba-1-1	Bacteria Proteobacteria Gammaproteobacteria *Alteromonadales bacterium* TW-7
Pha-1-1	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Pseudoalteromonadaceae *Pseudoalteromonas haloplanktis* TAC125
Ptu-1-1	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Pseudoalteromonadaceae *Pseudoalteromonas tunicata* D2
Pin-1-1 to Pin-1-7	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Psychromonadaceae *Psychromonas ingrahamii* 37
Sam-1-1	Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae *Shewanella amazonensis* SB2B
Pfl-1-1 to Pfl-1-3	Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf-5
env-1 to env-185	environmental samples

100.2 Gene contexts

Each TwoAYGGAY RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one TwoAYGGAY RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 100.3. The accession number of the sequence containing each TwoAYGGAY RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 100.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
Pin-1-1	NC_008709.1	-	4227970	4227804	**RNA**→ ←RplI (COG0359)
Pin-1-2	NC_008709.1	+	3672869	3673034	**RNA**→ ←hypo
Pin-1-3	NC_008709.1	+	1228464	1228624	**RNA**→ ←PleD (COG3706)
Pin-1-4	NC_008709.1	+	1237521	1237683	**RNA**→ ←PseudoU_synth_RluE (cd02566)
Pin-1-5	NC_008709.1	-	3670339	3670175	**RNA**→ Ribosomal_S2 (pfam00318)→
Pin-1-6	NC_008709.1	+	4230002	4230171	**RNA**→ ←COG0618 (COG0618)
Pin-1-7	NC_008709.1	-	1227322	1227152	**RNA**→ ←ModF (COG1119)ABC_ATPase (cd00267)
env-1	BAAZ01005038.1	+	287	455	**RNA**→ ←SleB (COG3773)
env-2	BAAX01027538.1	-	775	609	**RNA**→ ←STAS (pfam01740)
env-3	BAAW01000251.1	-	9337	9272	**RNA**→ ←Nei (COG0266)
Sam-1-1	NC_008700.1	+	4244752	4244914	**RNA**→ ←FdnI (COG2864)
env-4	BAAU01025327.1	+	674	835	**RNA**→ ←hypo
env-5	BAAU01012842.1	-	247	79	**RNA**→
env-6	BAAY01006061.1	+	789	955	**RNA**→
env-7	BAAU01018933.1	-	896	731	**RNA**→ ←DUF6 (pfam00892)
env-8	BAAU01010825.1	-	211	45	**RNA**→
env-9	BAAZ01011936.1	-	179	13	**RNA**→
env-10	BABA01001199.1	-	531	366	**RNA**→ hypo→
env-11	BABD01029348.1	-	766	601	**RNA**→ PstS (COG0226)→
env-12	BABA01008647.1	-	531	365	**RNA**→
env-13	BAAU01005978.1	-	405	243	**RNA**→ ←hypo
env-14	BAAU01001958.1	-	671	504	**RNA**→ ←hypo
env-15	BAAU01003657.1	-	644	477	**RNA**→ ←hypo
env-16	BAAU01005466.1	-	843	675	**RNA**→ ←MGS_CPS_II (cd01424)
env-17	BAAU01021915.1	-	521	357	**RNA**→
env-18	BAAU01002859.1	+	30	193	**RNA**→ ←FldA (COG0716)
env-19	BAAU01005978.1	+	732	895	**RNA**→
env-20	BAAU01027214.1	+	624	783	**RNA**→ ←hypo

env-21	BABG01023967.1	-	528	365	**RNA**→ ←dUTPase (pfam00692)
env-22	BABG01000271.1	-	2224	2061	**RNA**→ ←Idi (COG1443)
env-23	BABG01000436.1	-	2508	2345	**RNA**→ ABC_MJ0796_Lo1CDE_FtsE (cd03255)→ FtsX (pfam02687)→
env-24	BABG01005079.1	+	1103	1266	**RNA**→ ←PGM_PMM_IV (pfam00408)
env-25	BABG01000221.1	-	4934	4771	**RNA**→ ←FliS (COG1516)
env-26	BABG01000508.1	-	3021	2858	**RNA**→ COG1408 (COG1408)→
env-27	BABG01030864.1	-	800	634	**RNA**→
env-28	BAAU01025988.1	-	467	305	**RNA**→
env-29	BABG01000900.1	-	2303	2139	**RNA**→ FeoA (pfam04023)→
env-30	BABG01034670.1	-	387	221	**RNA**→
env-31	BABG01000393.1	-	2168	2002	**RNA**→ ←hypo
env-32	BABG01000236.1	-	4066	3900	**RNA**→ ←EAL (cd01948)
env-33	BAAU01005512.1	-	763	597	**RNA**→ ←hypo
env-34	BABG01000690.1	+	1558	1724	**RNA**→ ←Peptidase_M37 (pfam01551)
env-35	BABG01000393.1	+	2416	2582	**RNA**→ hypo→
env-36	BABG01014358.1	+	541	707	**RNA**→ hypo→
env-37	BABG01016197.1	+	197	363	**RNA**→ hypo→
env-38	BABG01000271.1	+	2472	2638	**RNA**→ ←hypo
env-39	BABG01001478.1	+	2351	2518	**RNA**→ hypo→
env-40	BAAU01027214.1	-	514	349	**RNA**→ hypo→
env-41	BABG01006543.1	+	58	224	**RNA**→ hypo→
env-42	BAAX01012202.1	+	62	228	**RNA**→ ←hypo
env-43	BABG01000704.1	-	141	1	**RNA**→
env-44	BABG01000436.1	+	2753	2925	**RNA**→ hypo→
env-45	BABG01030578.1	+	419	590	**RNA**→ hypo→
env-46	BABG01000752.1	+	502	671	**RNA**→ AslB (COG0641)→
env-47	BABG01000982.1	-	2748	2608	**RNA**→ ←MetG (COG0143)tRNA_bind_EcMetRS_like (cd02800)
env-48	BABG01026739.1	-	558	388	**RNA**→ ←hypo
env-49	BABG01000222.1	+	98	267	**RNA**→ hypo→
env-50	BABG01002686.1	-	791	622	**RNA**→ ←GckA (COG2379)
env-51	BABG01009970.1	+	135	304	**RNA**→ ←hypo
env-52	BABG01000900.1	+	2431	2600	**RNA**→
env-53	BABG01002620.1	-	1696	1527	**RNA**→ ←ThiC (pfam01964)
env-54	BABG01000236.1	+	4311	4480	**RNA**→ ←hypo
env-55	BABG01001970.1	-	2862	2722	**RNA**→ ←hypo
env-56	BABG01005381.1	-	998	829	**RNA**→ hypo→
env-57	BABG01000272.1	+	136	305	**RNA**→ ←FtsX (pfam02687)
env-58	BABG01000704.1	+	846	1015	**RNA**→ ←ThrS (COG0441)
env-59	BABG01000732.1	-	2801	2632	**RNA**→ ←HA (pfam03457)
env-60	BAAU01001873.1	-	176	7	**RNA**→
env-61	BABG01005079.1	-	856	687	**RNA**→ hypo→
env-62	BAAU01024793.1	+	91	230	**RNA**→ ←Polysacc_deac_1 (pfam01522)
env-63	BAAU01025327.1	-	250	81	**RNA**→
env-64	BABG01034670.1	+	631	803	**RNA**→
env-65	BAAU01001873.1	+	419	587	**RNA**→ ←Methionine_synt (pfam01717)
env-66	BABG01000690.1	-	1307	1140	**RNA**→ ←PpiB (COG0652)
env-67	BABG01000098.1	-	5643	5475	**RNA**→ ←AsnB (COG0367)
env-68	BABG01015786.1	+	315	483	**RNA**→
env-69	BABG01000704.1	-	651	483	**RNA**→
env-70	BABG01002686.1	+	1035	1203	**RNA**→ ←hypo
env-71	BABG01001970.1	+	3125	3293	**RNA**→
env-72	BABG01002029.1	+	55	223	**RNA**→ ←GatB (COG0064)

env-73	wf3_Contig7384	+	647	721	**RNA**→ hypo→ hypo→
Ptu-1-1	NZ_AAOH01000004.1	+	570569	570658	**RNA**→ hypo→
env-74	JCVI_SCAF_1096628099914	-	663	497	**RNA**→ ←? (GOS13244646)
Aba-1-1	NZ_AAVS01000041.1	-	17899	17733	**RNA**→ ←COG2932 (COG2932)HNHc (smart00507)
env-75	JCVI_SCAF_1096628101805	+	1140	1306	**RNA**→
Pha-1-1	NC_007481.1	+	2620603	2620680	**RNA**→ ←GcvP (COG0403)GcvP (COG1003)
env-76	wf3_AHAI1282_g1	+	357	445	**RNA**→ hypo→
env-77	BABB01000035.1	+	5490	5654	**RNA**→ ←CitB (COG2197)
env-78	hgutS8_s8_168446	-	1809	1645	**RNA**→ ←CitB (COG2197)
env-79	BAAY01004664.1	+	626	790	**RNA**→
env-80	BABB01000035.1	+	5824	5988	**RNA**→ ←CitB (COG2197)
env-81	BABB01000035.1	+	6159	6323	**RNA**→ ←CitB (COG2197)
env-82	BABG01033689.1	+	201	365	**RNA**→
env-83	BABG01033688.1	-	603	439	**RNA**→ ←CitB (COG2197)
env-84	BABB01000035.1	+	6534	6698	**RNA**→ ←CitB (COG2197)
env-85	BAAY01004664.1	+	1001	1165	**RNA**→
env-86	BABD01000901.1	-	293	129	**RNA**→
env-87	hgutS8_s8_168446	-	1366	1202	**RNA**→ ←CitB (COG2197)
env-88	hgutS7_s7_165842_1	+	346	511	**RNA**→ Valyl-tRNA synthetase(EC:6.1.1.9)→
env-89	BAAU01010141.1	-	959	794	**RNA**→ tRNA-synt_1 (pfam00133)→
env-90	BABB01000443.1	-	5578	5413	**RNA**→ ValS (COG0525)→
env-91	BABD01002028.1	-	2013	1848	**RNA**→ class_I_aaRS_core (cd00802)→ ValS (COG0525)→
env-92	BAAV01016980.1	+	240	405	**RNA**→
env-93	BAAY01025688.1	+	25	193	**RNA**→ ←LrgB (pfam04172)
env-94	BAAU01014849.1	+	675	838	**RNA**→
env-95	BABB01001311.1	-	954	792	**RNA**→ HTH_ARSR (smart00418)→ hypo→
env-96	BABG01001480.1	+	1057	1219	**RNA**→ HTH_ARSR (smart00418)→
env-97	BAAU01014848.1	-	872	709	**RNA**→ HTH_ARSR (smart00418)→ hypo→
env-98	BABG01027172.1	+	434	599	**RNA**→
env-99	BAAY01029070.1	-	262	101	**RNA**→ ←hypo
env-100	BAAU01017280.1	+	548	713	**RNA**→ ←hypo
env-101	BABB01000615.1	+	1519	1684	**RNA**→ ←GalT (COG4468)
env-102	BABB01001694.1	-	1795	1629	**RNA**→ ←GalE (COG1087)
env-103	BABG01003660.1	-	430	264	**RNA**→ ←hypo
env-104	BABD01004484.1	-	414	248	**RNA**→
env-105	BABB01000759.1	+	2789	2955	**RNA**→ ←ArgS (COG0018)
env-106	BABB01012412.1	-	295	129	**RNA**→
env-107	hgutS7_s7_164358_2	-	794	627	**RNA**→ Alpha-amylase_N (pfam02903)→
env-108	BABB01000218.1	+	10258	10425	**RNA**→ Alpha-amylase (pfam00128)Alpha-amylase_N (pfam02903)→ hypo→
env-109	BAAV01011582.1	-	1247	1090	**RNA**→ Alpha-amylase_N (pfam02903)Aamy (smart00642)→
env-110	BAAY01014418.1	-	579	411	**RNA**→ hypo→
env-111	BABB01012429.1	-	422	254	**RNA**→ hypo→
env-112	BABD01008410.1	+	4	171	**RNA**→
env-113	BABB01000982.1	+	9	177	**RNA**→ ←PulF (COG1459)
env-114	BAAV01009409.1	-	543	375	**RNA**→ ←hypo
env-115	BAAV01021692.1	+	208	376	**RNA**→ ←hypo
env-116	BAAY01014083.1	+	441	609	**RNA**→ ←hypo
env-117	BABD01026585.1	-	392	224	**RNA**→
env-118	BABG01025746.1	+	314	481	**RNA**→ hypo→
env-119	BABB01000066.1	-	9671	9504	**RNA**→ HisS (COG0124)→ AspS (COG0173)→ hypo→
env-120	BABG01011318.1	-	737	576	**RNA**→ ABC_membrane (pfam00664)→
env-121	BABB01010186.1	-	350	180	**RNA**→

env-122	BAAV01001293.1	-	234	67	**RNA**→ ←hypo
env-123	BABD01036649.1	-	946	792	**RNA**→ hypo→
env-124	BABB01000222.1	+	7557	7727	**RNA**→ CapC (COG4464)→ COG3944 (COG3944)→
env-125	BABB01000066.1	+	11493	11661	**RNA**→ ←PTS_IIB_glc (cd00212)PtsG (COG1263)
env-126	BABB01000043.1	+	4826	4994	**RNA**→ ←MdlB (COG1132)
env-127	BABD01000800.1	-	2312	2144	**RNA**→ ←MdlB (COG1132)
env-128	BAAY01002367.1	-	1949	1782	**RNA**→ ←hypo
env-129	BABD01001878.1	+	2188	2332	**RNA**→
env-130	BAAY01018278.1	+	527	698	**RNA**→ ←hypo
env-131	BABB01000120.1	+	4140	4308	**RNA**→ ←COG1853 (COG1853)
env-132	BABG01008261.1	+	622	790	**RNA**→ ←hypo
env-133	BABD01011793.1	-	129	1	**RNA**→
env-134	BAAV01008811.1	+	615	783	**RNA**→ ←COG1853 (COG1853)
env-135	BABD01016812.1	-	730	558	**RNA**→ hypo→
env-136	BABB01000628.1	+	3990	4158	**RNA**→ Tyr tRNA→
env-137	BABG01023578.1	+	628	797	**RNA**→
env-138	BAAU01014784.1	+	416	585	**RNA**→
env-139	BABB01000738.1	+	4375	4544	**RNA**→ hypo→
env-140	BABD01036650.1	+	491	660	**RNA**→
env-141	BAAV01030986.1	+	399	567	**RNA**→
env-142	BABG01019423.1	-	456	362	**RNA**→ ←hypo
env-143	BABB01000639.1	+	4945	5113	**RNA**→
env-144	BABB01001805.1	+	359	528	**RNA**→ ←DapB_C (pfam05173)
env-145	BAAV01024931.1	+	129	298	**RNA**→
env-146	BABB01018585.1	+	578	742	**RNA**→ ←hypo
env-147	BABB01000031.1	+	589	760	**RNA**→ ←Sulfatase (pfam00884)
env-148	BABG01001700.1	-	1701	1529	**RNA**→ ←Sulfatase (pfam00884)
env-149	BAAY01006880.1	+	32	203	**RNA**→ ←hypo
env-150	BAAU01016443.1	+	70	136	**RNA**→ ←hypo
env-151	BAAV01012556.1	-	776	710	**RNA**→ ←hypo
env-152	BABG01018343.1	-	594	502	**RNA**→
env-153	BABB01008784.1	+	236	405	**RNA**→ hypo→
env-154	BAAU01006263.1	+	1601	1770	**RNA**→
env-155	BAAV01012488.1	-	890	721	**RNA**→ ←Flavoprotein (pfam02441)
env-156	BABB01011528.1	+	151	320	**RNA**→
env-157	BABB01000321.1	+	5277	5446	**RNA**→ ARO8 (COG1167)→
env-158	BABD01000536.1	+	5186	5325	**RNA**→
env-159	BABB01001613.1	+	799	965	**RNA**→ ←COG3581 (COG3581)
env-160	BABG01000872.1	+	465	630	**RNA**→ ←COG3581 (COG3581)
env-161	BAAY01007451.1	+	25	190	**RNA**→ ←hypo
env-162	BABG01025117.1	+	31	202	**RNA**→ ←COG3581 (COG3581)
env-163	BABB01000925.1	+	486	653	**RNA**→ ←COG1924 (COG1924)COG3580 (COG3580)COG3581 (COG3581)
env-164	ABEF01021936.1	-	751	686	**RNA**→ ←Aldolase (pfam01081)
Pfl-1-1	NC_004129.6	-	3142534	3142468	**RNA**→ hypo→
Pfl-1-2	NC_004129.6	-	2830169	2830103	**RNA**→ ←PMSR (pfam01625)
Pfl-1-3	NC_004129.6	-	2989975	2989909	**RNA**→ ←LysP (COG0833)
env-165	BAAV01011978.1	+	813	978	**RNA**→
env-166	BAAX01016096.1	+	604	769	**RNA**→ ←hypo
env-167	mgutLn2_U_FF_aab65f06_b1	+	140	306	**RNA**→ Rhs family protein→
env-168	BABG01022683.1	-	528	364	**RNA**→ ←hypo
env-169	BABG01010186.1	-	991	843	**RNA**→ ←PhnK (COG1101)
env-170	BAAX01032386.1	+	191	337	**RNA**→ ←HMA (cd00371)

env-171	BABD01036144.1	+	678	842	**RNA**→
env-172	BABA01026201.1	+	110	272	**RNA**→ ←ClpX (COG1219)
env-173	BABG01018589.1	-	701	539	**RNA**→ ←HTH_3 (pfam01381)
env-174	BABA01015165.1	+	783	945	**RNA**→
env-175	BAAX01033270.1	-	298	136	**RNA**→
env-176	BABG01004015.1	+	650	812	**RNA**→ ←GGDEF (cd01949)
env-177	BABG01005149.1	-	1125	964	**RNA**→ ←INT_XerDC (cd00798)
env-178	BABG01028760.1	-	763	600	**RNA**→ ←PgsA (COG0558)
env-179	BABG01012947.1	-	690	527	**RNA**→ ←hypo
env-180	BABG01005904.1	-	1457	1294	**RNA**→ ←hypo
env-181	BAAX01016096.1	-	154	8	**RNA**→
Pth-1-1	NC_009454.1	-	2855527	2855462	**RNA**→ COG1284 (COG1284)→ TPP_TK (cd02012)→ COG3958 (COG3958)→ COG4942 (COG4942)→
Pth-1-2	NC_009454.1	+	96011	96076	**RNA**→ IlvA (COG1171)→
Pth-1-3	NC_009454.1	-	2842660	2842505	**RNA**→ UvrB (COG0556)→
Pth-1-4	NC_009454.1	-	444630	444476	**RNA**→ ←LivF (COG0410)
Rin-1-1	NZ_ABYJ01000412.1	+	150	319	**RNA**→
Rin-1-2	NZ_ABYJ01000381.1	+	208	375	**RNA**→
Rin-1-3	NZ_ABYJ01000220.1	-	8608	8441	**RNA**→
Rin-1-4	NZ_ABYJ01000445.1	+	547	714	**RNA**→
Rin-1-5	NZ_ABYJ01000556.1	+	707	875	**RNA**→
Rin-1-6	NZ_ABYJ01000216.1	+	20250	20418	**RNA**→
Rin-1-7	NZ_ABYJ01000220.1	+	9045	9213	**RNA**→
Rin-1-8	NZ_ABYJ01000023.1	-	6715	6547	**RNA**→
Rin-1-9	NZ_ABYJ01000380.1	-	13234	13065	**RNA**→
env-182	BAAX01017353.1	+	47	216	**RNA**→ Gp_dh_N (pfam00044)→
Rin-1-10	NZ_ABYJ01000556.1	-	214	45	**RNA**→
env-183	hgutS7_s7_177709_1	+	640	806	**RNA**→ hypo→
env-184	hgutS8_s8_177709_1	+	70	235	**RNA**→ hypo→
Rin-1-11	NZ_ABYJ01000044.1	+	847	1014	**RNA**→
Rin-1-12	NZ_ABYJ01000404.1	+	11801	11968	**RNA**→
Rin-1-13	NZ_ABYJ01000512.1	-	14600	14434	**RNA**→
env-185	BAAZ01022944.1	-	390	224	**RNA**→ ←hypo

100.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 100.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one TwoAYGGAY RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

cd00212 (1) PTS_IIB, PTS system, glucose/sucrose specific IIB subunit.
cd00267 (1) ABC (ATP-binding cassette) transporter nucleotide-binding domain; ABC transporters are a large family of proteins involved in the transport of a wide variety of different compounds, like sugars, ions, peptides and more complex organic molecules.
cd00371 (1) Heavy-metal-associated domain (HMA) is a conserved domain of approximately 30 amino acid residues found in a number of proteins that transport or detoxify heavy metals, for example, the CPx-type heavy metal ATPases and copper chaperones.
cd00798 (1) XerD and XerC integrases, DNA breaking-rejoining enzymes, N- and C-terminal domains.
cd00802 (1) Class I amino acyl-tRNA synthetase (aaRS) catalytic core domain.
cd01424 (1) Methylglyoxal synthase-like domain from type II glutamine-dependent carbamoyl phosphate synthetase (CSP).
cd01948 (1) EAL domain.
cd01949 (1) Diguanylate-cyclase (DGC) or GGDEF domain: Originally named after a conserved residue pattern, and initially described as domain of unknown function 1 (DUF1).
cd02012 (1) Thiamine pyrophosphate (TPP) family, Transketolase (TK) subfamily, TPP-binding module; TK catalyzes the transfer of a two-carbon unit from ketose phosphates to aldose phosphates.
cd02566 (1) PseudoU_synth_RluE: Pseudouridine synthase, Escherichia coli RluE.

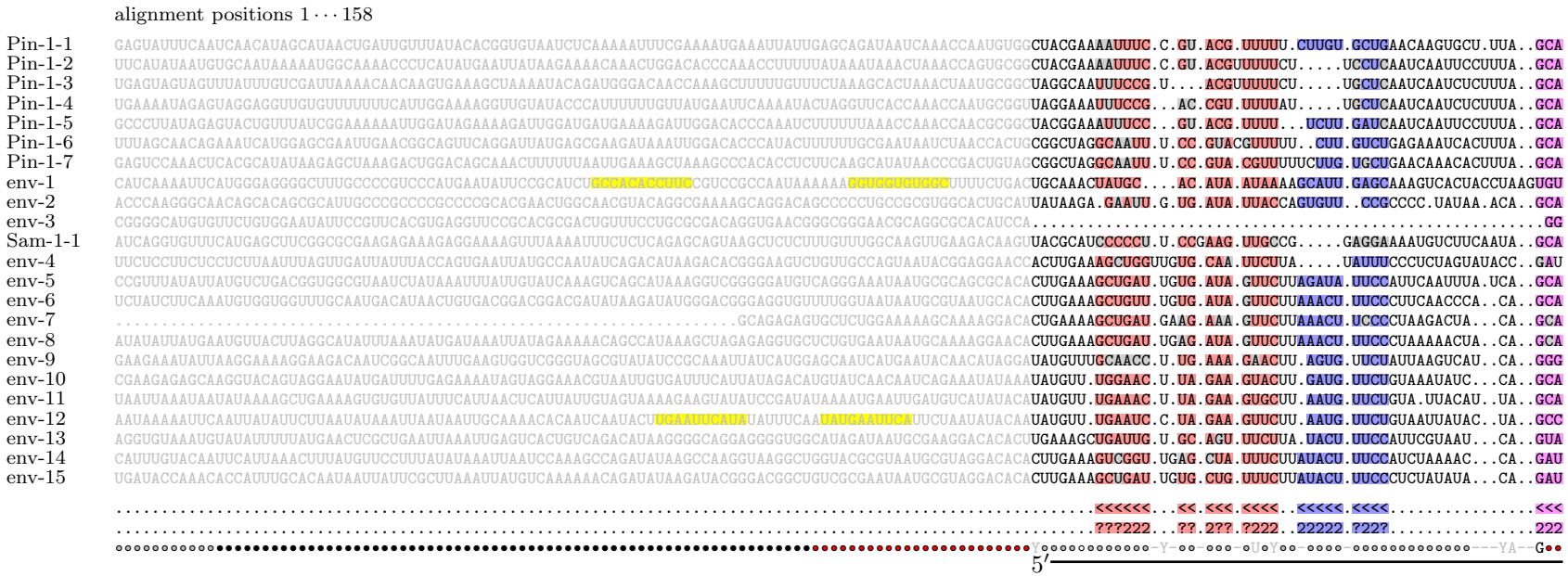
cd02800 (1) tRNA-binding-domain-containing Escherichia coli methionyl-tRNA synthetase (EcMetRS)-like proteins.
cd03255 (1) This family is comprised of MJ0796 ATP-binding cassette, macrolide-specific ABC-type efflux carrier (MacAB), and proteins involved in cell division (FtsE) and release of liporoteins from the cytoplasmic membrane (LolCDE).
COG0018 (1) Arginyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
COG0064 (1) Asp-tRNAAsn/Glu-tRNAGln amidotransferase B subunit (PET112 homolog) [Translation, ribosomal structure and biogenesis]
COG0124 (1) Histidyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
COG0143 (1) Methionyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
COG0173 (1) Aspartyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
COG0226 (1) ABC-type phosphate transport system, periplasmic component [Inorganic ion transport and metabolism]
COG0266 (1) Formamidopyrimidine-DNA glycosylase [DNA replication, recombination, and repair]
COG0359 (1) Ribosomal protein L9 [Translation, ribosomal structure and biogenesis]
COG0367 (1) Asparagine synthase (glutamine-hydrolyzing) [Amino acid transport and metabolism]
COG0403 (1) Glycine cleavage system protein P (pyridoxal-binding), N-terminal domain [Amino acid transport and metabolism]
COG0410 (1) ABC-type branched-chain amino acid transport systems, ATPase component [Amino acid transport and metabolism]
COG0441 (1) Threonyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
COG0525 (2) Valyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
COG0556 (1) Helicase subunit of the DNA excision repair complex [DNA replication, recombination, and repair]
COG0558 (1) Phosphatidylglycerophosphate synthase [Lipid metabolism]
COG0618 (1) Exopolyphosphatase-related proteins [General function prediction only]
COG0641 (1) Arylsulfatase regulator (Fe-S oxidoreductase) [General function prediction only]
COG0652 (1) Peptidyl-prolyl cis-trans isomerase (rotamase) - cyclophilin family [Posttranslational modification, protein turnover, chaperones]
COG0716 (1) Flavodoxins [Energy production and conversion]
COG0833 (1) Amino acid transporters [Amino acid transport and metabolism]
COG1003 (1) Glycine cleavage system protein P (pyridoxal-binding), C-terminal domain [Amino acid transport and metabolism]
COG1087 (1) UDP-glucose 4-epimerase [Cell envelope biogenesis, outer membrane]
COG1101 (1) ABC-type uncharacterized transport system, ATPase component [General function prediction only]
COG1119 (1) ABC-type molybdenum transport system, ATPase component/photorepair protein PhrA [Inorganic ion transport and metabolism]
COG1132 (2) ABC-type multidrug transport system, ATPase and permease components [Defense mechanisms]
COG1167 (1) Transcriptional regulators containing a DNA-binding HTH domain and an aminotransferase domain (MocR family) and their eukaryotic orthologs [Transcription / Amino acid transport and metabolism]
COG1171 (1) Threonine dehydratase [Amino acid transport and metabolism]
COG1219 (1) ATP-dependent protease Clp, ATPase subunit [Posttranslational modification, protein turnover, chaperones]
COG1263 (1) Phosphotransferase system IIC components, glucose/maltose/N-acetylglucosamine-specific [Carbohydrate transport and metabolism]
COG1284 (1) Uncharacterized conserved protein [Function unknown]
COG1408 (1) Predicted phosphohydrolases [General function prediction only]
COG1443 (1) Isopentenyldiphosphate isomerase [Lipid metabolism]
COG1459 (1) Type II secretory pathway, component PulF [Cell motility and secretion / Intracellular trafficking and secretion]
COG1516 (1) Flagellin-specific chaperone FliS [Cell motility and secretion / Intracellular trafficking and secretion / Posttranslational modification, protein turnover, chaperones]
COG1853 (2) Conserved protein/domain typically associated with flavoprotein oxygenases, DIM6/NTAB family [General function prediction only]
COG1924 (1) Activator of 2-hydroxyglutaryl-CoA dehydratase (HSP70-class ATPase domain) [Lipid metabolism]
COG2197 (7) Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain [Signal transduction mechanisms / Transcription]
COG2379 (1) Putative glycerate kinase [Carbohydrate transport and metabolism]
COG2864 (1) Cytochrome b subunit of formate dehydrogenase [Energy production and conversion]
COG2932 (1) Predicted transcriptional regulator [Transcription]
COG3580 (1) Uncharacterized protein conserved in bacteria [Function unknown]
COG3581 (4) Uncharacterized protein conserved in bacteria [Function unknown]
COG3706 (1) Response regulator containing a CheY-like receiver domain and a GGDEF domain [Signal transduction mechanisms]
COG3773 (1) Cell wall hydrolyses involved in spore germination [Cell envelope biogenesis, outer membrane]
COG3944 (1) Capsular polysaccharide biosynthesis protein [Cell envelope biogenesis, outer membrane]
COG3958 (1) Transketolase, C-terminal subunit [Carbohydrate transport and metabolism]
COG4464 (1) Capsular polysaccharide biosynthesis protein [Carbohydrate transport and metabolism / Cell envelope biogenesis, outer membrane]
COG4468 (1) Galactose-1-phosphate uridyltransferase [Carbohydrate transport and metabolism]
COG4942 (1) Membrane-bound metallopeptidase [Cell division and chromosome partitioning]
GOS13244646 (1) ?
pfam00044 (1) Glyceraldehyde 3-phosphate dehydrogenase, NAD binding domain.
pfam00128 (1) Alpha amylase, catalytic domain.
pfam00133 (1) tRNA synthetases class I (I, L, M and V).
pfam00318 (1) Ribosomal protein S2.
pfam00408 (1) Phosphoglucomutase/phosphomannomutase, C-terminal domain.
pfam00664 (1) ABC transporter transmembrane region.
pfam00692 (1) dUTPase.
pfam00884 (2) Sulfatase.
pfam00892 (1) Integral membrane protein DUF6.
pfam01081 (1) KDPG and KHG aldolase.
pfam01381 (1) Helix-turn-helix.
pfam01522 (1) Polysaccharide deacetylase.
pfam01551 (1) Peptidase family M23/M37.
pfam01625 (1) Peptide methionine sulfoxide reductase.
pfam01717 (1) Methionine synthase, vitamin-B12 independent.
pfam01740 (1) STAS domain.
pfam01964 (1) ThiC family.
pfam02441 (1) Flavoprotein.

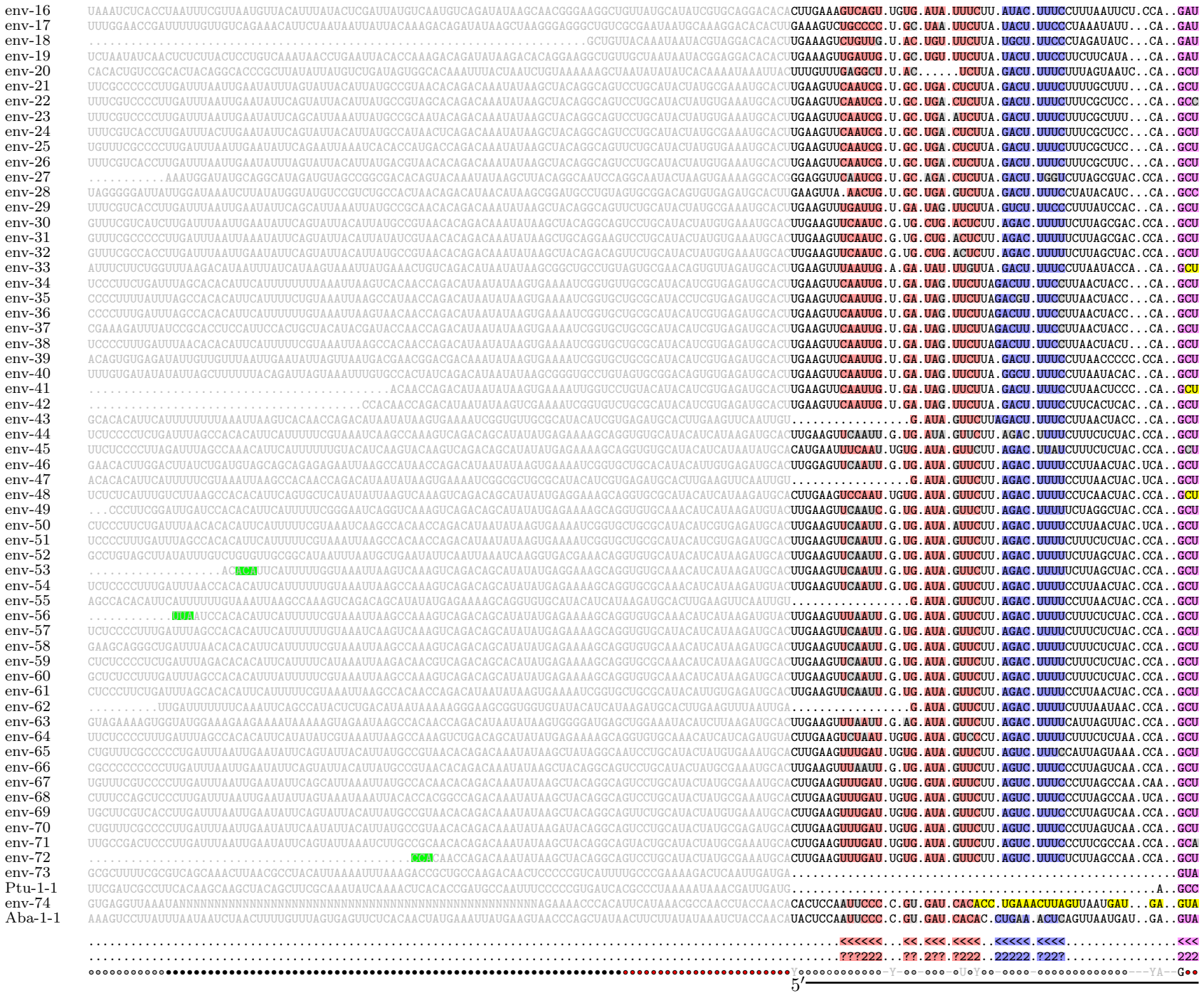
pfam02687 (2) Predicted permease.
pfam02903 (3) Alpha amylase, N-terminal ig-like domain.
pfam03457 (1) Helicase associated domain.
pfam04023 (1) FeoA domain.
pfam04172 (1) LrgB-like family.
pfam05173 (1) Dihydrodipicolinate reductase, C-terminus.
smart00418 (3) helix_turn_helix, Arsenical Resistance Operon Repressor;
smart00507 (1) HNH nucleases;
smart00642 (1) Alpha-amylase domain;

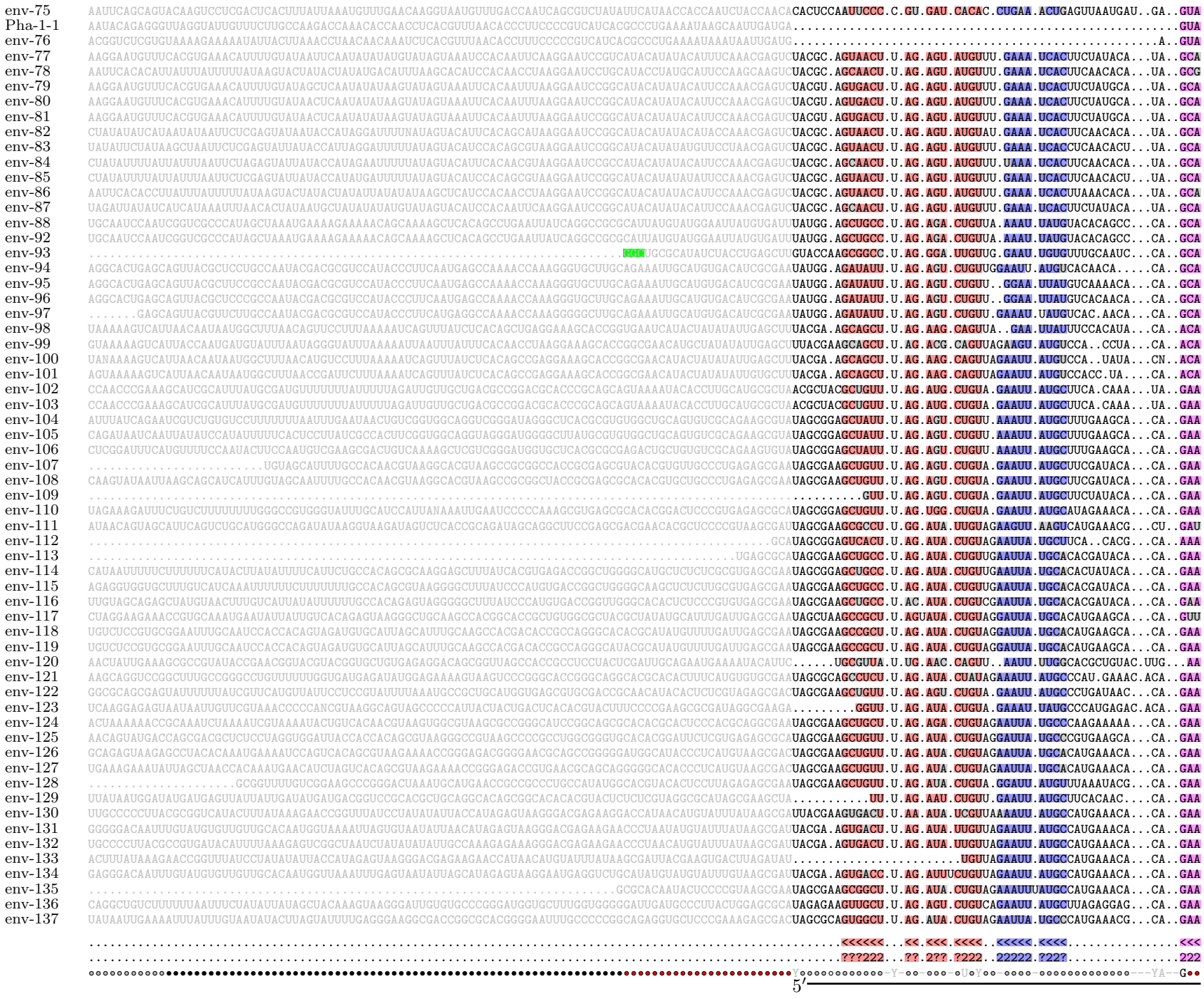
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: env-89, env-90, env-91

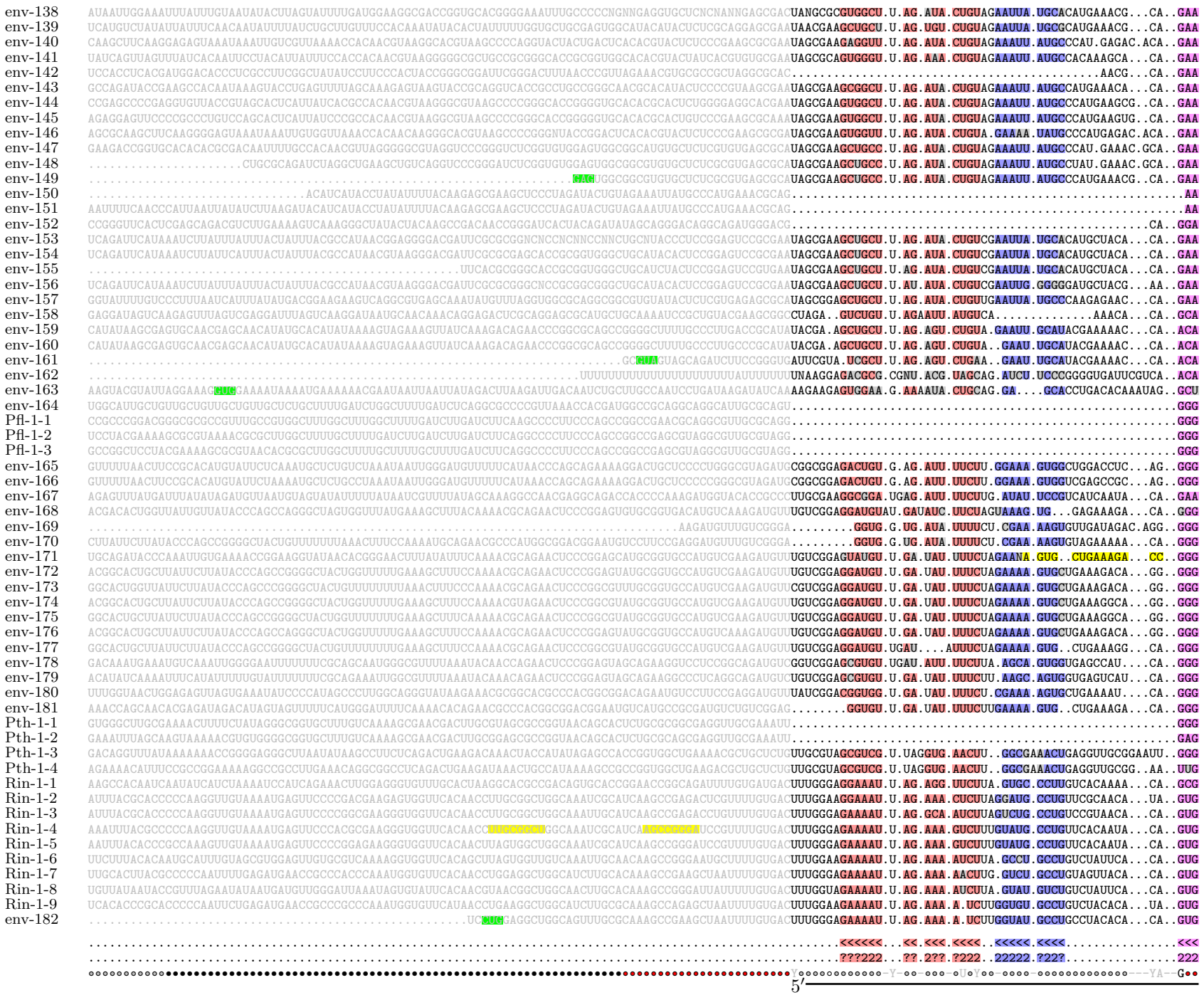
100.4 Multiple-sequence alignment

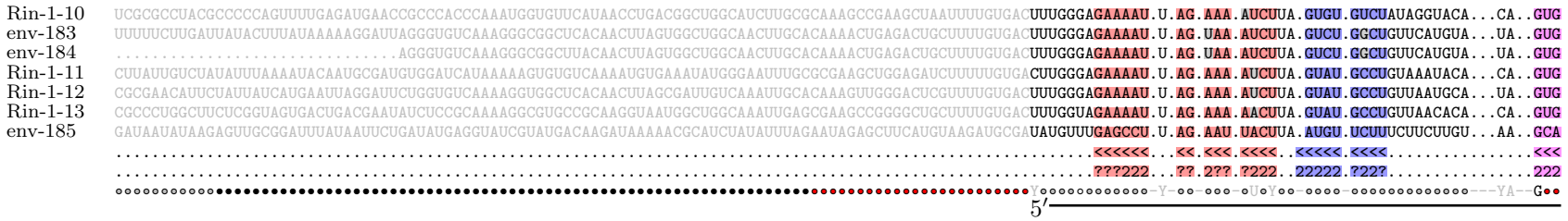
Each TwoAYGGAY RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 100.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The TwoAYGGAY RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



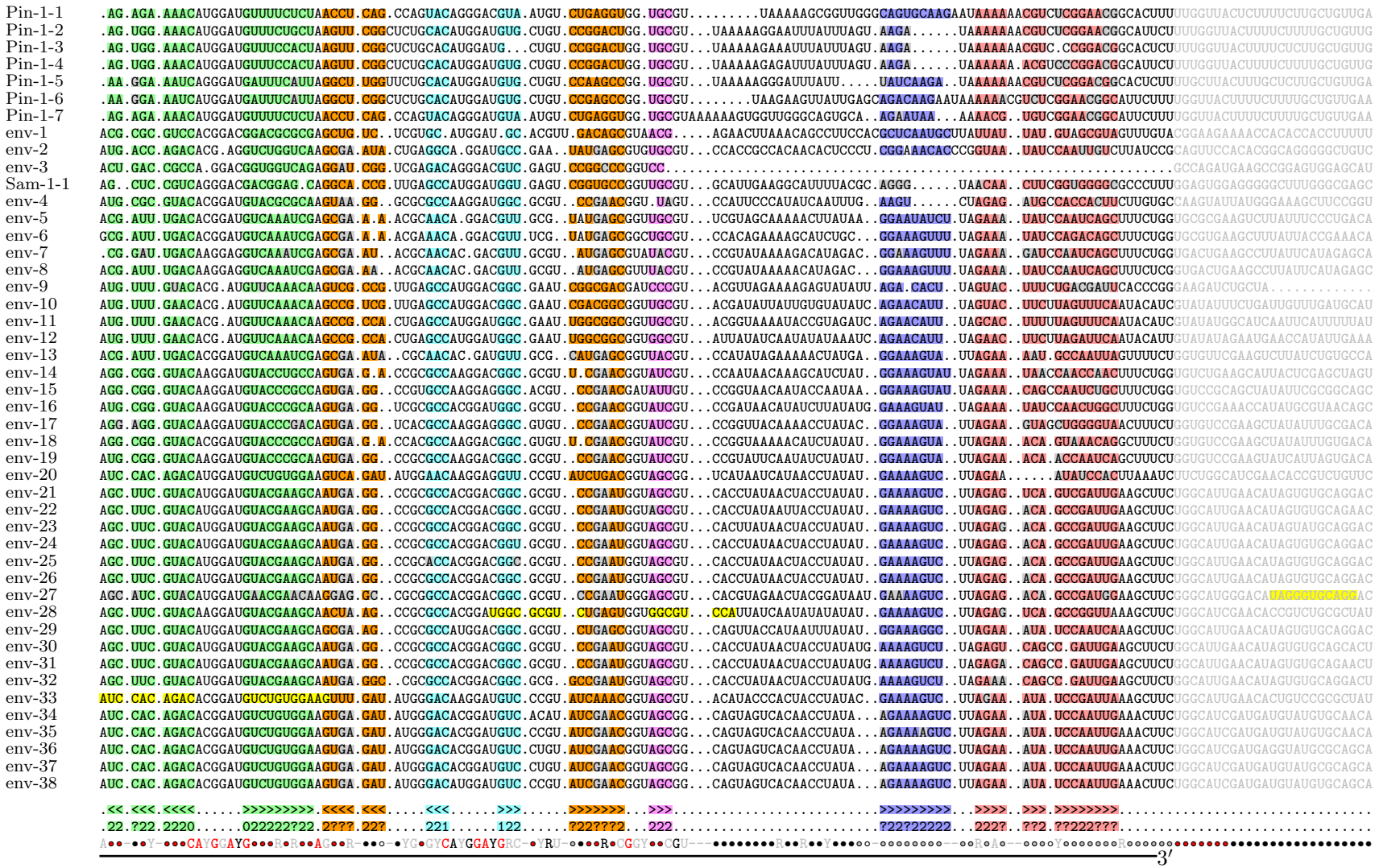

env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
Ptu-1-1
env-74
Aba-1-1
5′

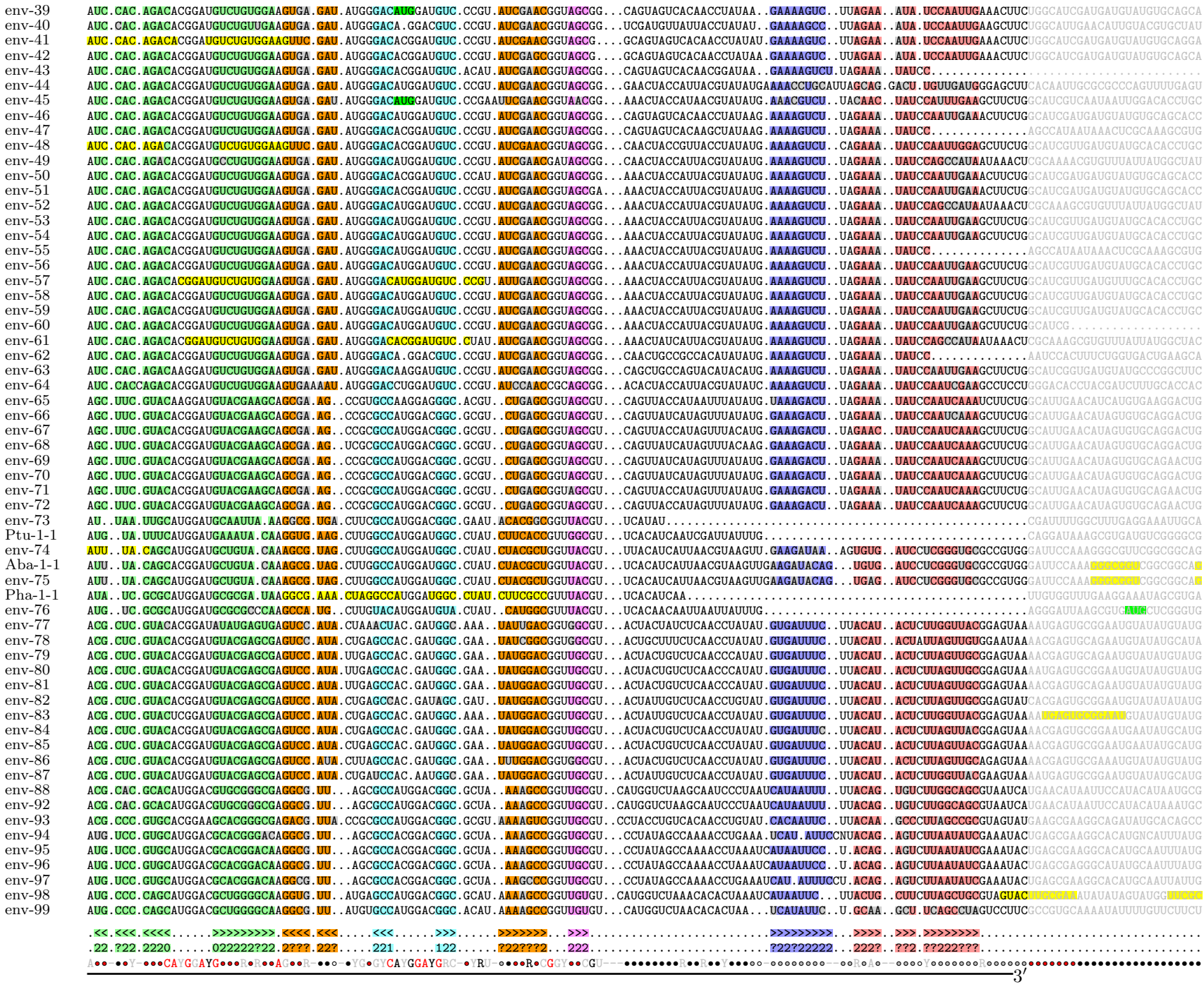

env-75
Pha-1-1
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-135
env-136
env-137
5′

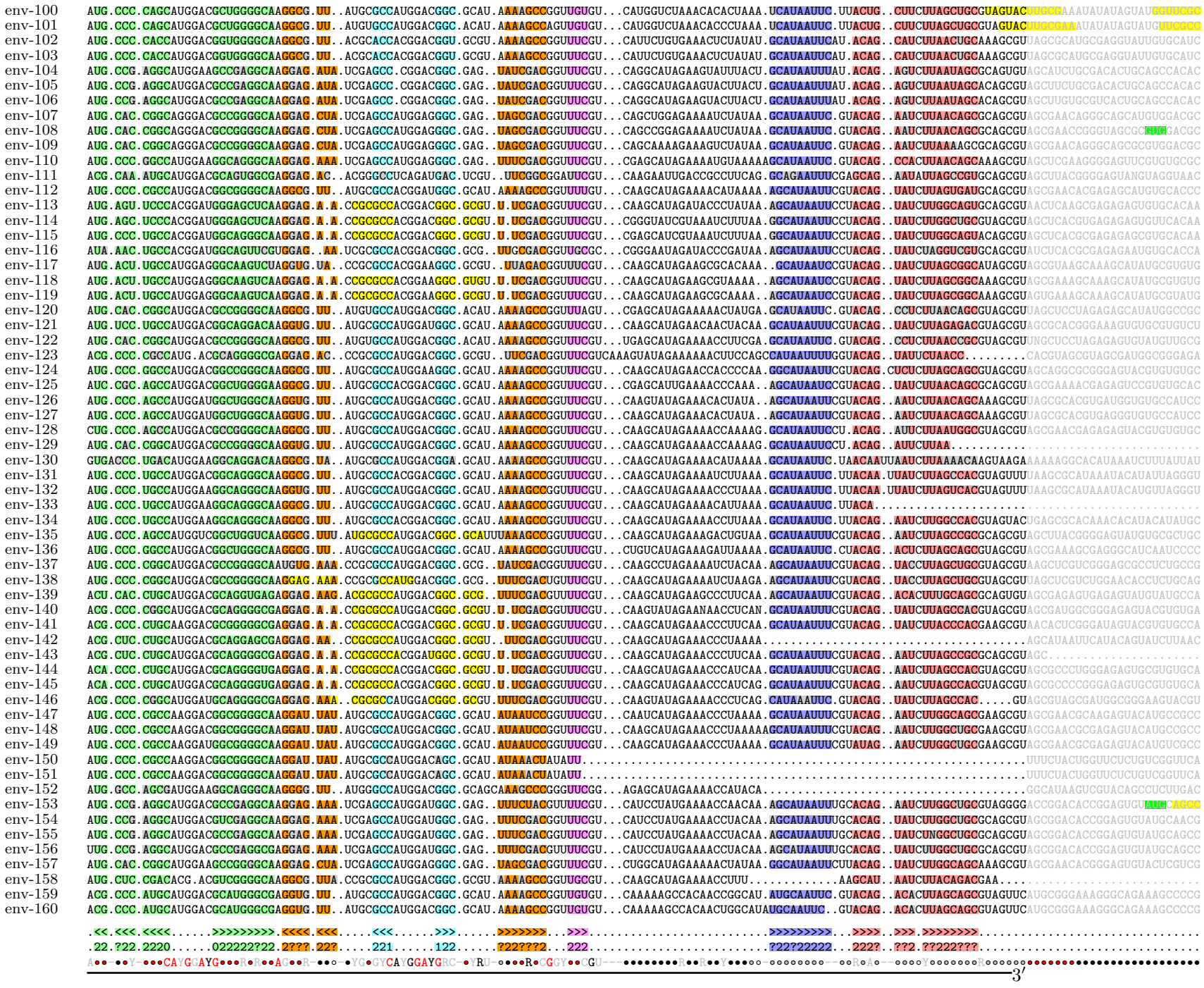

env-138
env-139
env-140
env-141
env-142
env-143
env-144
env-145
env-146
env-147
env-148
env-149
env-150
env-151
env-152
env-153
env-154
env-155
env-156
env-157
env-158
env-159
env-160
env-161
env-162
env-163
env-164
Pfl-1-1
Pfl-1-2
Pfl-1-3
env-165
env-166
env-167
env-168
env-169
env-170
env-171
env-172
env-173
env-174
env-175
env-176
env-177
env-178
env-179
env-180
env-181
Pth-1-1
Pth-1-2
Pth-1-3
Pth-1-4
Rin-1-1
Rin-1-2
Rin-1-3
Rin-1-4
Rin-1-5
Rin-1-6
Rin-1-7
Rin-1-8
Rin-1-9
env-182
5′

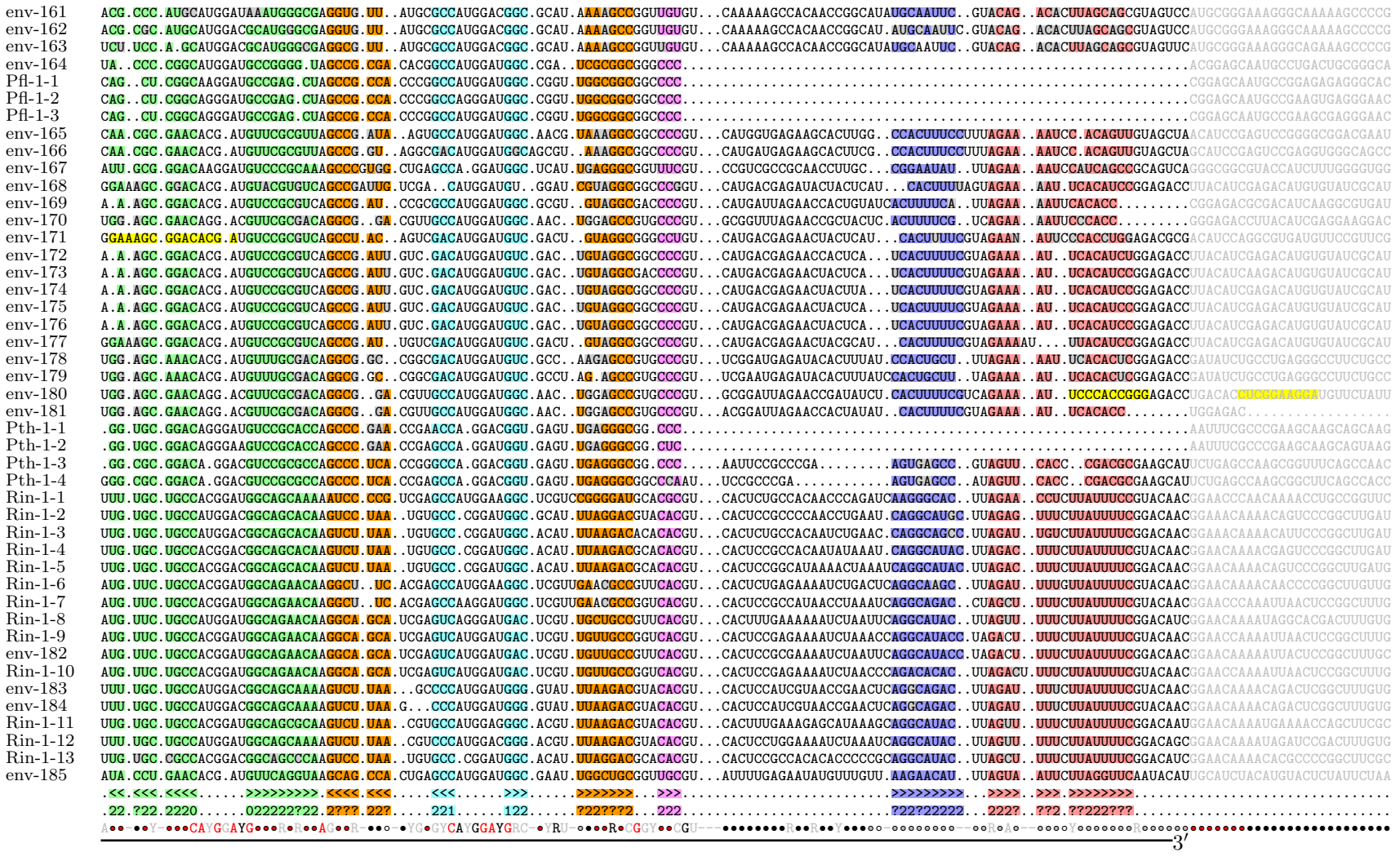


alignment positions 159 · · · 318

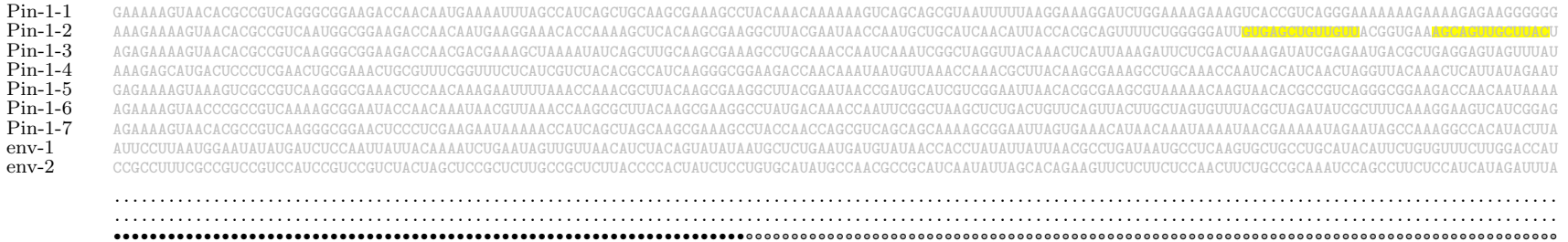



env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
Ptu-1-1
env-74
Aba-1-1
env-75
Pha-1-1
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
3′

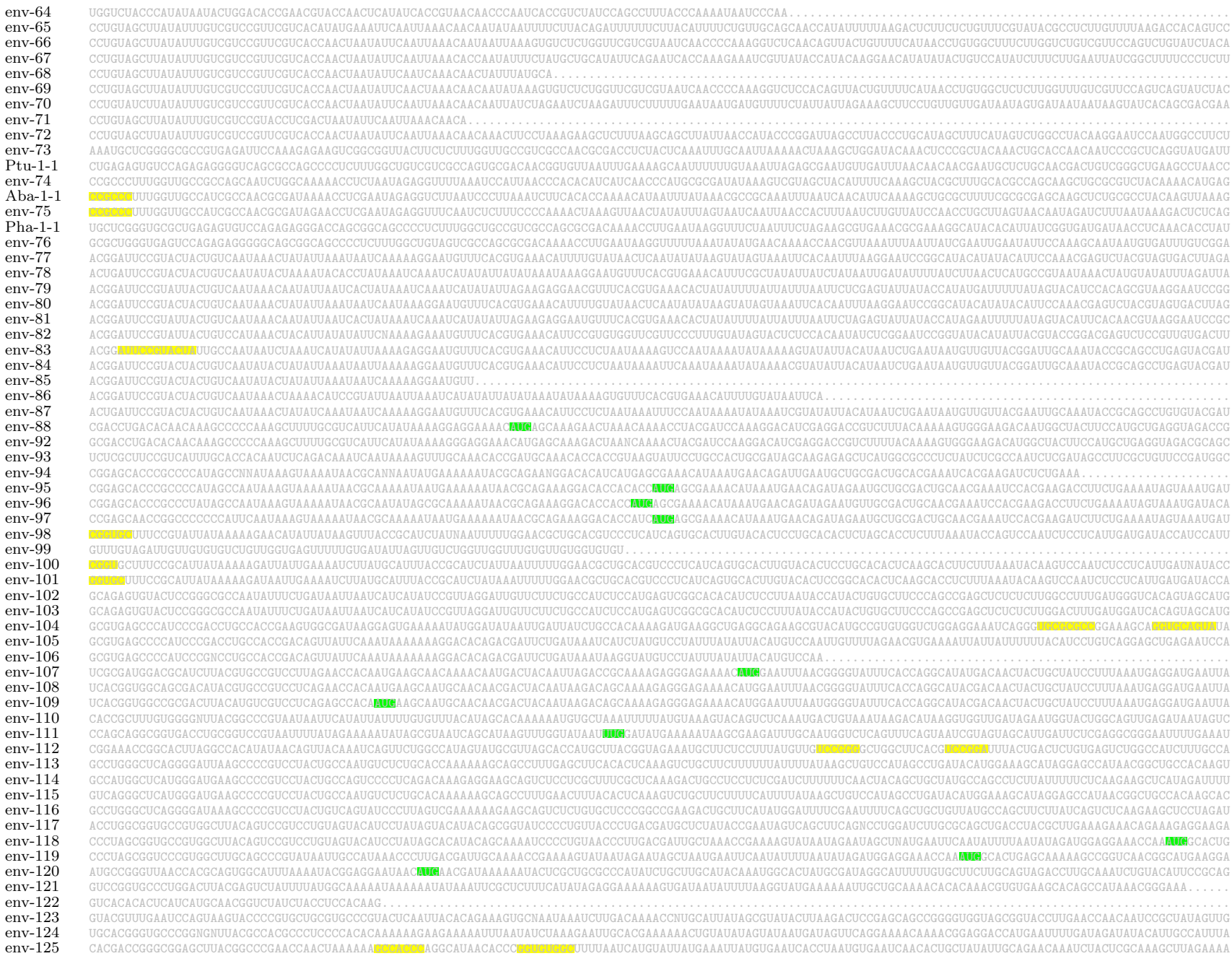

env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-135
env-136
env-137
env-138
env-139
env-140
env-141
env-142
env-143
env-144
env-145
env-146
env-147
env-148
env-149
env-150
env-151
env-152
env-153
env-154
env-155
env-156
env-157
env-158
env-159
env-160
3′



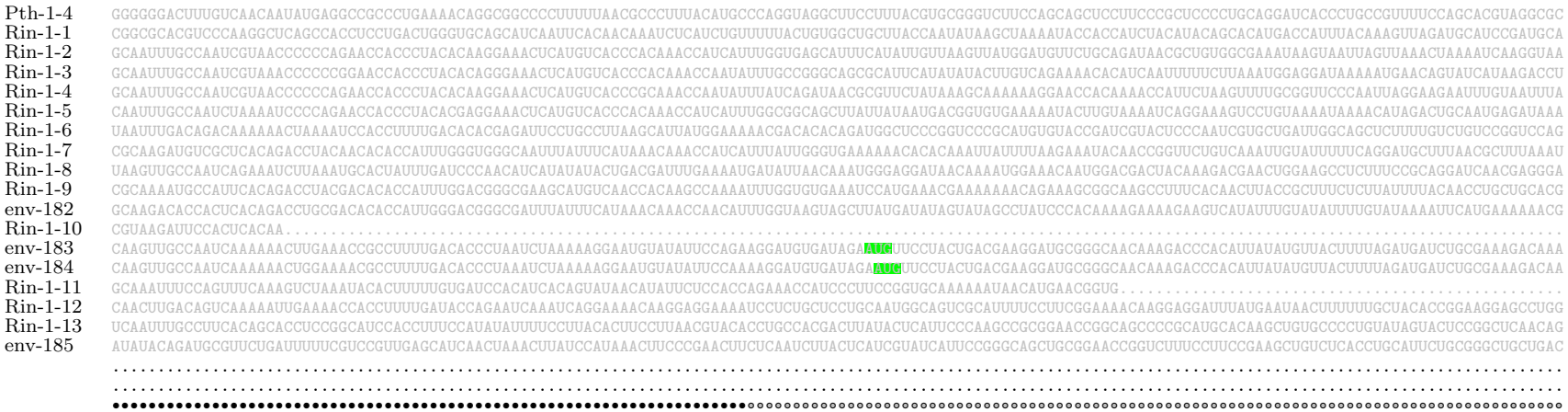
alignment positions 319 · · · 478



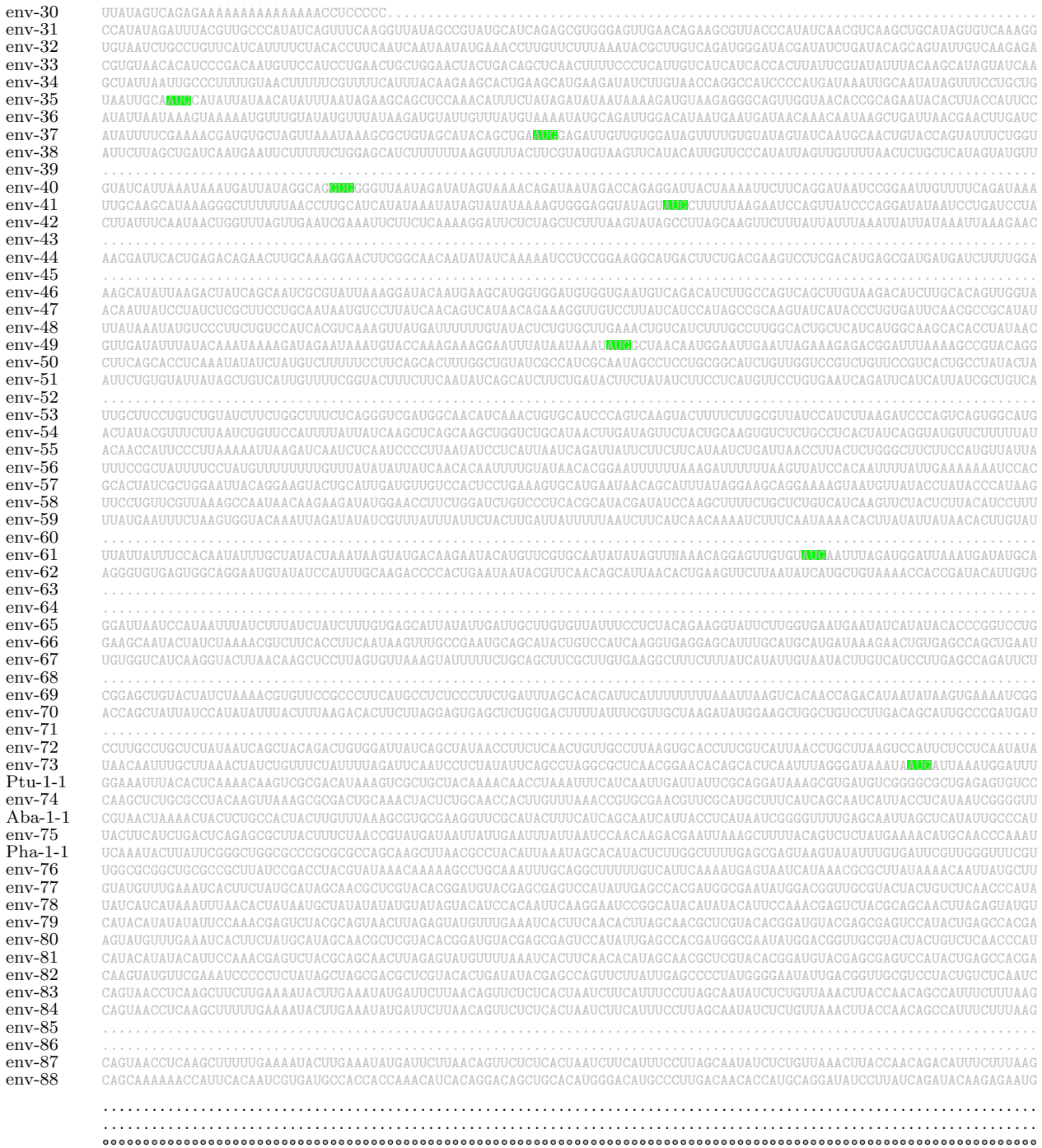
env-3
Sam-1-1
env-4
env-5
env-6
env-7
env-8
env-9
env-10
env-11
env-12
env-13
env-14
env-15
env-16
env-17
env-18
env-19
env-20
env-21
env-22
env-23
env-24
env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63

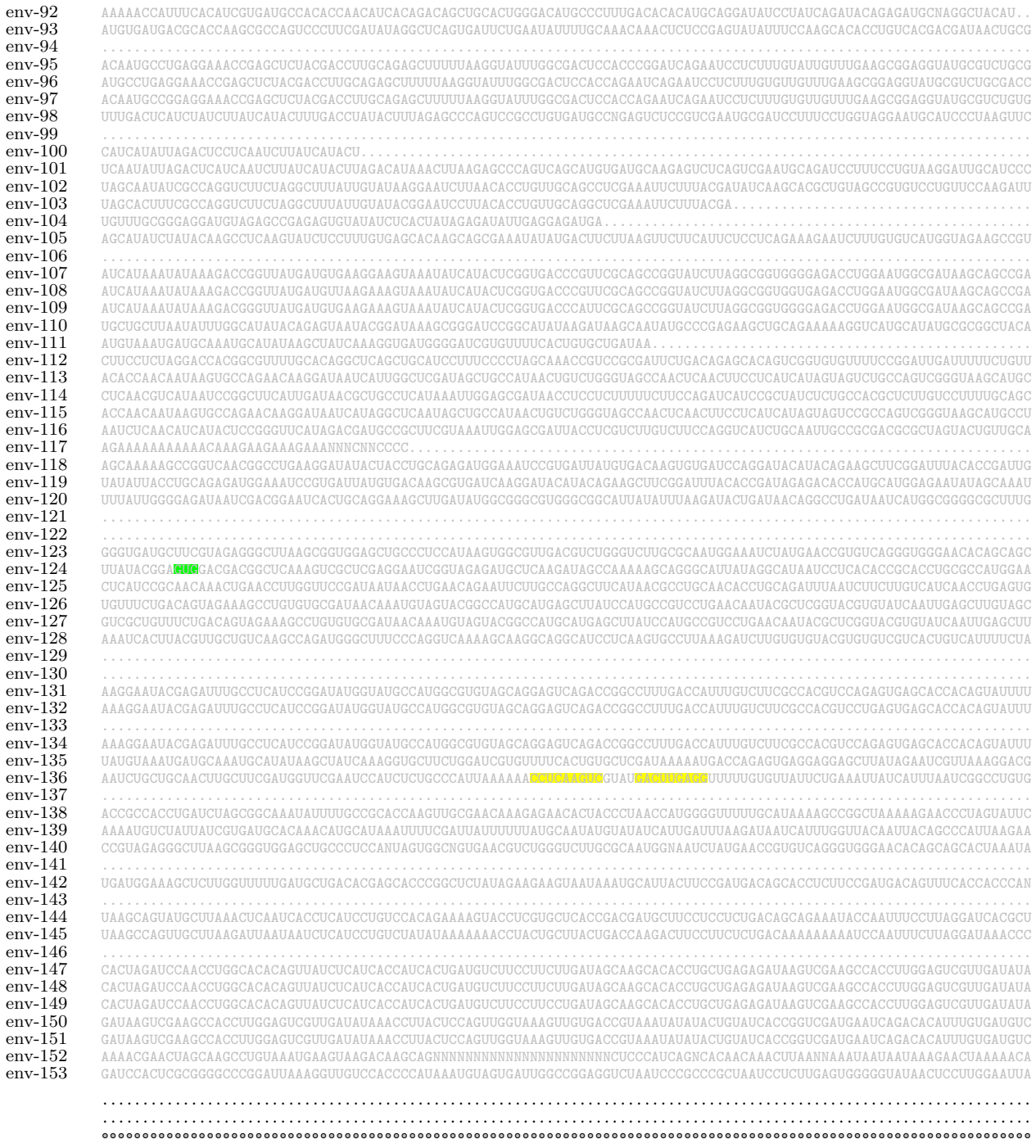

env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
Ptu-1-1
env-74
Aba-1-1
env-75
Pha-1-1
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
env-123
env-124
env-125

env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-135
env-136
env-137
env-138
env-139
env-140
env-141
env-142
env-143
env-144
env-145
env-146
env-147
env-148
env-149
env-150
env-151
env-152
env-153
env-154
env-155
env-156
env-157
env-158
env-159
env-160
env-161
env-162
env-163
env-164
Pfl-1-1
Pfl-1-2
Pfl-1-3
env-165
env-166
env-167
env-168
env-169
env-170
env-171
env-172
env-173
env-174
env-175
env-176
env-177
env-178
env-179
env-180
env-181
Pth-1-1
Pth-1-2
Pth-1-3



alignment positions $479 \cdots 593$


env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
Ptu-1-1
env-74
Aba-1-1
env-75
Pha-1-1
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88


env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108
env-109
env-110
env-111
env-112
env-113
env-114
env-115
env-116
env-117
env-118
env-119
env-120
env-121
env-122
env-123
env-124
env-125
env-126
env-127
env-128
env-129
env-130
env-131
env-132
env-133
env-134
env-135
env-136
env-137
env-138
env-139
env-140
env-141
env-142
env-143
env-144
env-145
env-146
env-147
env-148
env-149
env-150
env-151
env-152
env-153

101 *wcaG*

101.1 Taxa

The taxonomy of each organism containing a putative *wcaG* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *wcaG* RNA in Sections 101.2 and 101.4.

abbrev. of hits	taxonomy of species
Sph-1-1	Viruses dsDNA viruses, no RNA stage Caudovirales Myoviridae *Synechococcus phage syn*9
Sph-2-1	Viruses dsDNA viruses, no RNA stage Caudovirales Myoviridae T4-like viruses *Synechococcus phage* S-PM2
env-1 to env-118	environmental samples

101.2 Gene contexts

Each *wcaG* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *wcaG* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 101.3. The accession number of the sequence containing each *wcaG* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 101.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	JCVI_SCAF_1096627988848	+	1360	1459	**RNA**→ ? (GOS1696)→
env-2	JCVI_SCAF_1101668710796	-	226	128	**RNA**→
env-3	JCVI_SCAF_1101667281504	-	651	557	**RNA**→ Gly tRNA→ ? (GOS11184583)→
env-4	JCVI_SCAF_1096627349584	-	641	548	**RNA**→ ? (GOS7556)→
env-5	JCVI_SCAF_1096628390190	+	595	676	**RNA**→ Gmd (COG1089)? (GOS5158)→ ? (GOS721)→
env-6	JCVI_SCAF_1096627284163	+	1003	1097	**RNA**→ WcaG (COG0451)? (GOS4294)→
env-7	JCVI_SCAF_1101667028124	-	199	100	**RNA**→ Gly tRNA→
env-8	JCVI_SCAF_1096626871951	+	1789	1887	**RNA**→ P4Hc (smart00702)? (GOS8792)→ ? (GOS11184583)→
env-9	JCVI_SCAF_1101668738526	-	1535	1436	**RNA**→ ? (GOS13195016)→
env-10	JCVI_SCAF_1101668672382	+	116	206	**RNA**→ Gmd (COG1089)? (GOS5158)→
env-11	JCVI_SCAF_1101668240460	+	137	234	**RNA**→ WcaG (COG0451)? (GOS4294)→
env-12	JCVI_SCAF_1096628017379	+	1409	1506	**RNA**→
env-13	JCVI_SCAF_1096627180125	+	593	690	**RNA**→ Gmd (COG1089)? (GOS5158)→
env-14	JCVI_SCAF_1096627326740	-	1278	1183	**RNA**→ WcaG (COG0451)? (GOS4294)→
env-15	JCVI_SCAF_1096627287958	-	734	637	**RNA**→ RfbB (COG1088)? (GOS4294)→
env-16	JCVI_SCAF_1096627515878	+	1605	1702	**RNA**→
env-17	JCVI_SCAF_1101668743236	-	561	491	**RNA**→ ? (GOS4294)→
env-18	JCVI_SCAF_1101668740399	+	1193	1295	**RNA**→ ? (GOS13195016)→
env-19	JCVI_SCAF_1096627314381	-	3304	3207	**RNA**→ Glyco_transf_11 (pfam01531)? (GOS6720)→ ? (GOS1625)→ ? (GOS1625)→ ? (GOS8260)→ COG0637 (COG0637)LicC (COG4750)? (GOS4678)→
env-20	JCVI_SCAF_1096627301567	+	1523	1620	**RNA**→
env-21	JCVI_SCAF_1101668606612	-	492	390	**RNA**→ Gly tRNA→ SLH (pfam00395)? (GOS13552031)→
env-22	JCVI_SCAF_1096627146507	-	1699	1597	**RNA**→ Gly tRNA→ SLH (pfam00395)? (GOS13552031)→ ? (GOS13866)→ ? (GOS14287987)→
env-23	JCVI_SCAF_1096627362151	+	856	958	**RNA**→ Gly tRNA→ SLH (pfam00395)? (GOS13552031)→
env-24	JCVI_SCAF_1096626846327	+	1082	1177	**RNA**→ ? (GOS4294)→
env-25	JCVI_SCAF_1096627987509	+	991	1089	**RNA**→ ? (GOS4491)→
env-26	JCVI_SCAF_1096627174567	-	666	568	**RNA**→ Gly tRNA→ ? (GOS14287987)→
env-27	JCVI_SCAF_1096626415574	+	417	514	**RNA**→ Leu tRNA→ ? (GOS8589)→

env-28	JCVI_SCAF_1096627018987	-	2128	2032	**RNA**→ Gly tRNA→ ? (GOS13176163)→
env-29	JCVI_SCAF_1096627315204	-	926	830	**RNA**→ Gly tRNA→ ? (GOS4491)→
env-30	JCVI_SCAF_1096627012476	-	1252	1156	**RNA**→ Met tRNA→ Leu tRNA→
env-31	JCVI_SCAF_1096627022633	-	1426	1330	**RNA**→ Gly tRNA→ ? (GOS13226979)→
env-32	JCVI_SCAF_1096627315058	+	450	555	**RNA**→ ? (GOS4294)→ WcaG (COG0451)? (GOS4294)→
env-33	JCVI_SCAF_1096627638638	+	262	367	**RNA**→ ←? (GOS11441011)
env-34	JCVI_SCAF_1101667281504	-	551	454	**RNA**→ Gly tRNA→ ? (GOS11184583)→
env-35	JCVI_SCAF_1096627178336	-	1719	1622	**RNA**→ Gly tRNA→ ? (GOS11184583)→
env-36	JCVI_SCAF_1096626664626	-	487	392	**RNA**→ ? (GOS7556)→
env-37	JCVI_SCAF_1096627986644	-	191	93	**RNA**→
env-38	JCVI_SCAF_1096627155304	-	451	353	**RNA**→ Gly tRNA→
env-39	JCVI_SCAF_1101667007739	+	475	570	**RNA**→ ? (GOS4294)→
env-40	JCVI_SCAF_1096627153909	+	3337	3436	**RNA**→ Gly tRNA→
env-41	JCVI_SCAF_1096627656112	+	423	518	**RNA**→ ? (GOS4294)→ ? (GOS4294)→ dTDP_sugar_isom (pfam00908)? (GOS8594)→
env-42	JCVI_SCAF_1101667166959	+	715	811	**RNA**→
env-43	JCVI_SCAF_1096627505316	-	157	60	**RNA**→
env-44	JCVI_SCAF_1101668355177	-	1166	1069	**RNA**→ WcaG (COG0451)? (GOS4294)→
env-45	JCVI_SCAF_1096627750892	+	62	161	**RNA**→ WcaG (COG0451)? (GOS4294)→ ? (GOS4294)→ ? (GOS11038676)→
env-46	JCVI_SCAF_1096627057077	-	1049	954	**RNA**→ ? (GOS7556)→ ? (GOS5158)→
env-47	JCVI_SCAF_1101667346321	+	517	612	**RNA**→ ? (GOS5158)→
env-48	JCVI_SCAF_1096628017369	+	492	587	**RNA**→ ? (GOS5158)→ Gmd (COG1089)? (GOS5158)→ ? (GOS6866)→
env-49	JCVI_SCAF_1096627178865	+	240	335	**RNA**→ WcaG (COG0451)? (GOS4294)→ Glyco_transf_11 (pfam01531)? (GOS6720)→
env-50	JCVI_SCAF_1096626921072	+	391	486	**RNA**→ ? (GOS7797)→
env-51	JCVI_SCAF_1096627175424	-	2449	2354	**RNA**→ Gmd (COG1089)? (GOS5158)→ WcaG (COG0451)? (GOS4294)→ ? (GOS6720)→
env-52	JCVI_SCAF_1096627164048	+	52	147	**RNA**→ Gmd (COG1089)? (GOS5158)→ WcaG (COG0451)? (GOS4294)→ Glyco_transf_11 (pfam01531)? (GOS6720)→
env-53	JCVI_SCAF_1096627181460	+	1685	1780	**RNA**→ Gmd (COG1089)? (GOS5158)→
env-54	JCVI_SCAF_1096627326740	-	1176	1081	**RNA**→ WcaG (COG0451)? (GOS4294)→
env-55	JCVI_SCAF_1096627905439	+	372	471	**RNA**→ ? (GOS4294)→
env-56	JCVI_SCAF_1096627163819	-	1766	1666	**RNA**→ Gmd (COG1089)? (GOS5158)→ WcaG (COG0451)? (GOS4294)→
env-57	JCVI_SCAF_1096628017574	+	132	232	**RNA**→ Gmd (COG1089)? (GOS5158)→ ? (GOS5158)→ WcaG (COG0451)? (GOS4294)→
env-58	JCVI_SCAF_1101668742161	+	336	436	**RNA**→ Gmd (COG1089)? (GOS5158)→ ? (GOS4294)→
env-59	JCVI_SCAF_1101668276391	+	185	282	**RNA**→ WcaG (COG0451)? (GOS4294)→
env-60	JCVI_SCAF_1096626914328	+	563	659	**RNA**→ WcaG (COG0451)? (GOS4294)→
env-61	JCVI_SCAF_1096627854511	-	286	191	**RNA**→
env-62	JCVI_SCAF_1096627509257	+	441	536	**RNA**→ ? (GOS4294)→ WcaG (COG0451)? (GOS4294)→ ? (GOS6844)→
env-63	JCVI_SCAF_1101668171059	+	876	974	**RNA**→ RfbB (COG1088)? (GOS4294)→
env-64	JCVI_SCAF_1101668336457	+	239	334	**RNA**→ WcaG (COG0451)? (GOS4294)→
env-65	JCVI_SCAF_1096628033863	+	1622	1721	**RNA**→
env-66	JCVI_SCAF_1096627078485	+	519	614	**RNA**→ WcaG (COG0451)? (GOS4294)→ ? (GOS6844)→ ? (GOS6957)→
env-67	JCVI_SCAF_1096628036419	-	933	835	**RNA**→ WcaG (COG0451)? (GOS4294)→
env-68	JCVI_SCAF_1101668323164	-	1399	1302	**RNA**→ RfbB (COG1088)? (GOS4294)→ ? (GOS4294)→
env-69	JCVI_SCAF_1101668738374	+	219	314	**RNA**→ Gmd (COG1089)? (GOS5158)→ ? (GOS4294)→
env-70	JCVI_SCAF_1096627793364	+	1399	1496	**RNA**→
env-71	JCVI_SCAF_1096627078401	-	2256	2159	**RNA**→ ? (GOS5158)→ ? (GOS5158)→ Gmd (COG1089)? (GOS5158)→ WcaG (COG0451)? (GOS4294)→
env-72	JCVI_SCAF_1101668365448	-	628	529	**RNA**→ Epimerase (pfam01370)? (GOS4294)→
env-73	JCVI_SCAF_1101668346227	+	509	608	**RNA**→ ? (GOS4294)→ ? (GOS4294)→
env-74	JCVI_SCAF_1096627177447	+	1019	1117	**RNA**→ WcaG (COG0451)? (GOS4294)→
env-75	JCVI_SCAF_1096627316054	+	475	573	**RNA**→ WcaG (COG0451)? (GOS4294)→ ? (GOS13350831)→
env-76	JCVI_SCAF_1096628027324	-	1399	1301	**RNA**→ ? (GOS4294)→ ? (GOS4294)→ ? (GOS13350831)→
env-77	JCVI_SCAF_1096627178851	+	547	645	**RNA**→ WcaG (COG0451)? (GOS4294)→
env-78	JCVI_SCAF_1101668387418	+	846	943	**RNA**→

env-79	JCVI_SCAF_1096627765138	+	354	453	**RNA**→ ECT (cd02173)? (GOS8446)→ RfbB (COG1088)? (GOS4294)→
env-80	JCVI_SCAF_1096627372755	-	218	121	**RNA**→
env-81	JCVI_SCAF_1096626851108	-	1315	1216	**RNA**→ WcaG (COG0451)? (GOS4294)→
env-82	JCVI_SCAF_1101668331227	+	945	1044	**RNA**→ WcaG (COG0451)? (GOS4294)→
env-83	JCVI_SCAF_1096628386119	-	1531	1432	**RNA**→ ? (GOS5158)→ WcaG (COG0451)? (GOS4294)→ ? (GOS453)→
env-84	JCVI_SCAF_1096627064305	-	1598	1501	**RNA**→ WcaG (COG0451)? (GOS4294)→ ? (GOS6720)→
env-85	JCVI_SCAF_1096627180401	-	670	571	**RNA**→ RfbB (COG1088)? (GOS4294)→
env-86	JCVI_SCAF_1101667348771	-	383	284	**RNA**→ ? (GOS7797)→
env-87	JCVI_SCAF_1096627180449	+	516	608	**RNA**→ ? (GOS18444)→ Thr tRNA→ Ala tRNA→
env-88	JCVI_SCAF_1096627007615	-	1316	1219	**RNA**→ ? (GOS7556)→ Glyco_transf_11 (pfam01531)? (GOS6720)→
env-89	JCVI_SCAF_1096627009603	+	709	807	**RNA**→ cytidylyltransferase (cd02170)? (GOS8446)→ RfbB (COG1088)? (GOS4294)→ dTDP_sugar_isom (pfam00908)? (GOS8594)→ Epimerase (pfam01370)? (GOS4294)→ ? (GOS13311056)→ EGL-9 (COG3751)? (GOS7796)→
env-90	JCVI_SCAF_1101668252080	-	383	288	**RNA**→ ←? (GOS11737729)
env-91	JCVI_SCAF_1101668244897	-	946	849	**RNA**→ ? (GOS8217)→
env-92	JCVI_SCAF_1096627673078	-	390	290	**RNA**→
env-93	JCVI_SCAF_1096627648430	+	713	811	**RNA**→ ? (GOS11956858)→ MannoseP_isomer (pfam01050)? (GOS1274)→ ? (GOS705)→
env-94	JCVI_SCAF_1096627008095	+	1869	1966	**RNA**→ Glyco_transf_11 (pfam01531)? (GOS6720)→
env-95	JCVI_SCAF_1096627636250	+	530	627	**RNA**→ ? (GOS5158)→ Gmd (COG1089)? (GOS5158)→ ManC (COG0662)? (GOS9050)→
env-96	JCVI_SCAF_1101668249861	+	1000	1097	**RNA**→ ? (GOS6720)→
env-97	JCVI_SCAF_1101667171027	+	152	249	**RNA**→
env-98	JCVI_SCAF_1096627023165	-	560	463	**RNA**→ ? (GOS6720)→
env-99	JCVI_SCAF_1096627660620	+	441	538	**RNA**→ ? (GOS7796)→
env-100	JCVI_SCAF_1096627007582	-	731	634	**RNA**→ ? (GOS7797)→
env-101	JCVI_SCAF_1096627660981	+	1029	1126	**RNA**→ WcaG (COG0451)? (GOS4294)→
env-102	JCVI_SCAF_1101668243849	+	646	743	**RNA**→ ←? (GOS11737729)
env-103	JCVI_SCAF_1096627017934	+	31	128	**RNA**→ ? (GOS4294)→ WcaG (COG0451)? (GOS4294)→ ? (GOS12026525)→
env-104	JCVI_SCAF_1096627379300	-	2476	2379	**RNA**→ ? (GOS7796)→ ? (GOS13311056)→ ? (GOS13524081)→ ? (GOS6720)→
env-105	JCVI_SCAF_1101668236646	+	699	796	**RNA**→ ? (GOS7796)→ ? (GOS13311056)→
env-106	JCVI_SCAF_1101668254996	-	875	779	**RNA**→ ECT (cd02173)? (GOS8446)→ ? (GOS4294)→
env-107	JCVI_SCAF_1101668154241	-	1037	937	**RNA**→ WcaG (COG0451)? (GOS4294)→
env-108	JCVI_SCAF_1096627678759	+	335	432	**RNA**→ ←? (GOS11737729)
env-109	JCVI_SCAF_1101667168679	+	548	643	**RNA**→
env-110	JCVI_SCAF_1096627015860	-	308	207	**RNA**→
env-111	JCVI_SCAF_1096627663931	+	265	363	**RNA**→ ? (GOS7797)→ ? (GOS7797)→
env-112	JCVI_SCAF_1096627639048	-	1412	1312	**RNA**→ ? (GOS5158)→ Gmd (COG1089)? (GOS5158)→ ? (GOS4294)→
env-113	JCVI_SCAF_1096627674418	-	925	823	**RNA**→ WcaG (COG0451)? (GOS4294)→
Sph-1-1	NC_008296.2	+	163184	163278	**RNA**→ gp208→ high light inducible protein→ high light inducible protein→ putative thioredoxin→ gp212→ gp213→ gp214→ gp215→
env-114	JCVI_SCAF_1096627012844	+	1767	1863	**RNA**→ RfbB (COG1088)? (GOS4294)→
Sph-2-1	NC_006820.1	+	122806	122905	**RNA**→ tRNA-Gly→ putative S-layer protein→ hypo→ putative high-light inducible protein→ hypo→ putative high-light inducible protein→
env-115	JCVI_SCAF_1096627158073	+	175	274	**RNA**→ Gly tRNA→ ? (GOS7556)→ ? (GOS11670541)→
env-116	JCVI_SCAF_1096628029727	-	454	355	**RNA**→ Gly tRNA→
env-117	JCVI_SCAF_1101668239056	+	989	1088	**RNA**→ ? (GOS4491)→
env-118	JCVI_SCAF_1096626914328	+	216	310	**RNA**→ ? (GOS18334)→

101.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 101.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *wcaG* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

cd02170 (1) The cytidylyltransferase family includes cholinephosphate cytidylyltransferase (CCT), glycerol-3-phosphate cytidylyltransferase, RafE and phosphoethanolamine cytidylyltransferase (ECT).
cd02173 (2) CTP:phosphoethanolamine cytidylyltransferase (ECT) catalyzes the conversion of phosphoethanolamine to CDP-ethanolamine as part of the CDP#ethanolamine biosynthesis pathway.
COG0451 (30) Nucleoside-diphosphate-sugar epimerases [Cell envelope biogenesis, outer membrane / Carbohydrate transport and metabolism]
COG0637 (1) Predicted phosphatase/phosphohexomutase [General function prediction only]
COG0662 (1) Mannose-6-phosphate isomerase [Carbohydrate transport and metabolism]
COG1088 (7) dTDP-D-glucose 4,6-dehydratase [Cell envelope biogenesis, outer membrane]
COG1089 (14) GDP-D-mannose dehydratase [Cell envelope biogenesis, outer membrane]
COG3751 (1) Predicted proline hydroxylase [Posttranslational modification, protein turnover, chaperones]
COG4750 (1) CTP:phosphocholine cytidylyltransferase involved in choline phosphorylation for cell surface LPS epitopes [Cell envelope biogenesis, outer membrane]
GOS11038676 (1) ?
GOS11184583 (4) ?
GOS11441011 (1) ?
GOS11670541 (1) ?
GOS11737729 (3) ?
GOS11956858 (1) ?
GOS12026525 (1) ?
GOS1274 (1) ?
GOS13176163 (1) ?
GOS13195016 (2) ?
GOS13226979 (1) ?
GOS13311056 (3) ?
GOS13350831 (2) ?
GOS13524081 (1) ?
GOS13552031 (3) ?
GOS13866 (1) ?
GOS14287987 (2) ?
GOS1625 (2) ?
GOS1696 (1) ?
GOS18334 (1) ?
GOS18444 (1) ?
GOS4294 (58) ?
GOS4491 (3) ?
GOS453 (1) ?
GOS4678 (1) ?
GOS5158 (23) ?
GOS6720 (10) ?
GOS6844 (2) ?
GOS6866 (1) ?
GOS6957 (1) ?
GOS705 (1) ?
GOS721 (1) ?
GOS7556 (5) ?
GOS7796 (4) ?
GOS7797 (5) ?
GOS8217 (1) ?
GOS8260 (1) ?
GOS8446 (3) ?
GOS8589 (1) ?
GOS8594 (2) ?
GOS8792 (1) ?
GOS9050 (1) ?
pfam00395 (3) S-layer homology domain.
pfam00908 (2) dTDP-4-dehydrorhamnose 3,5-epimerase.
pfam01050 (1) Mannose-6-phosphate isomerase.
pfam01370 (2) NAD dependent epimerase/dehydratase family.
pfam01531 (5) Glycosyl transferase family 11.
smart00702 (1) Prolyl 4-hydroxylase alpha subunit homologues.

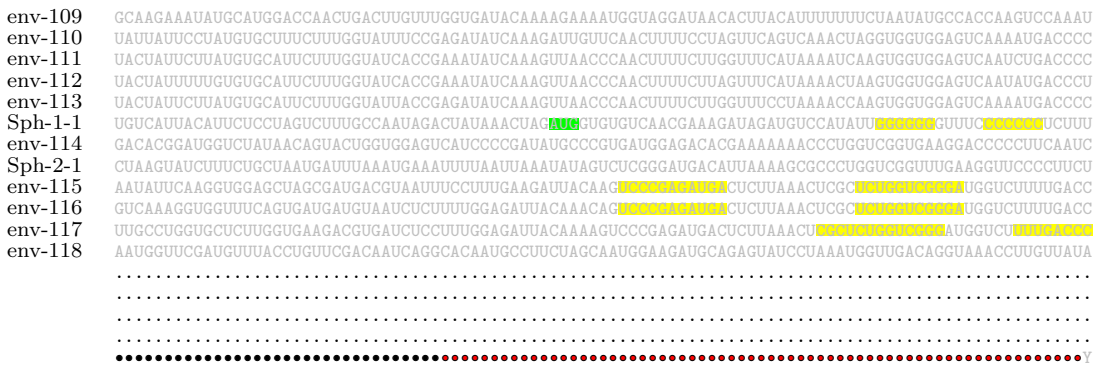
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: env-22, env-23

101.4 Multiple-sequence alignment

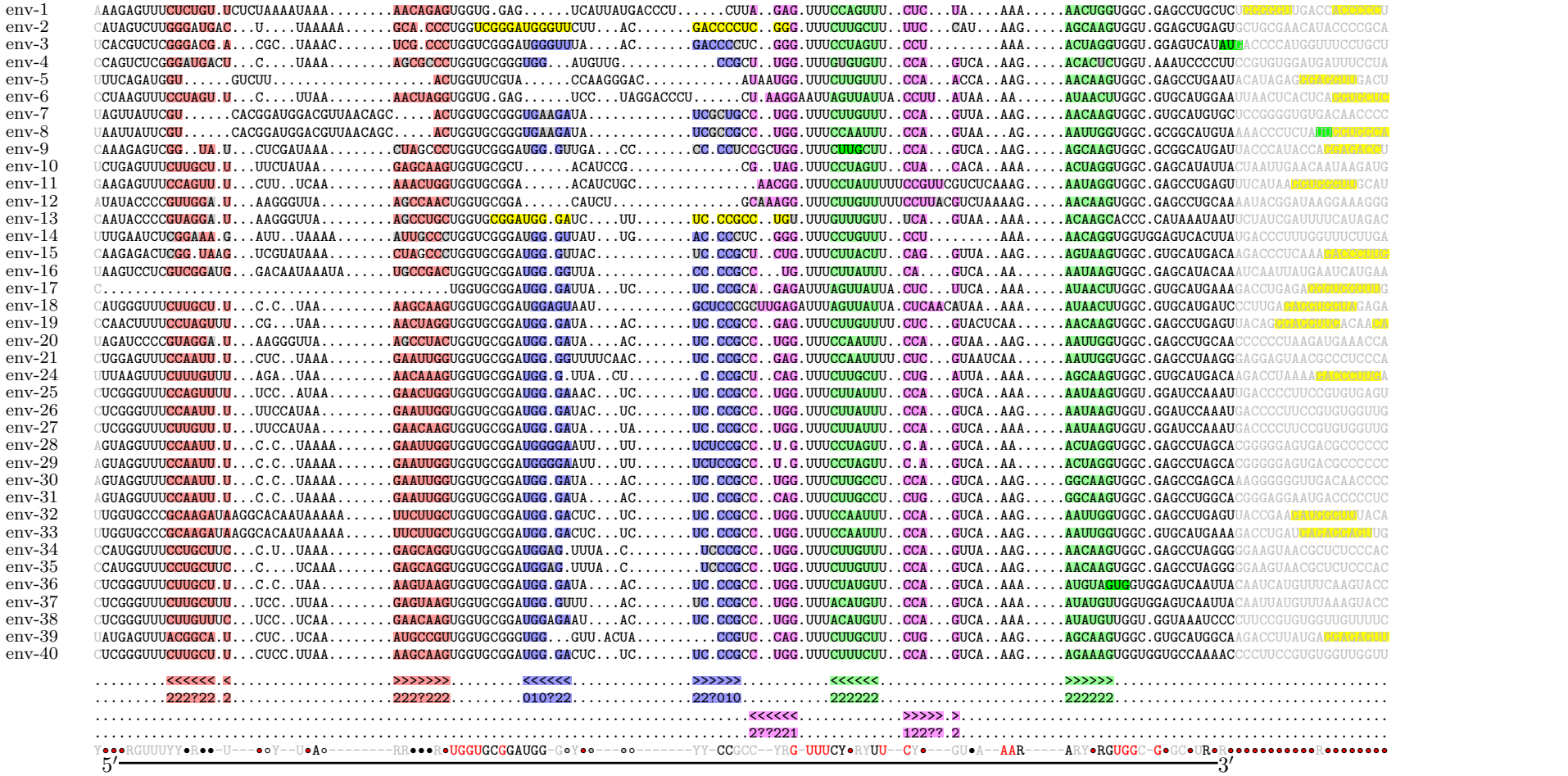
Each *wcaG* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 101.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *wcaG* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when

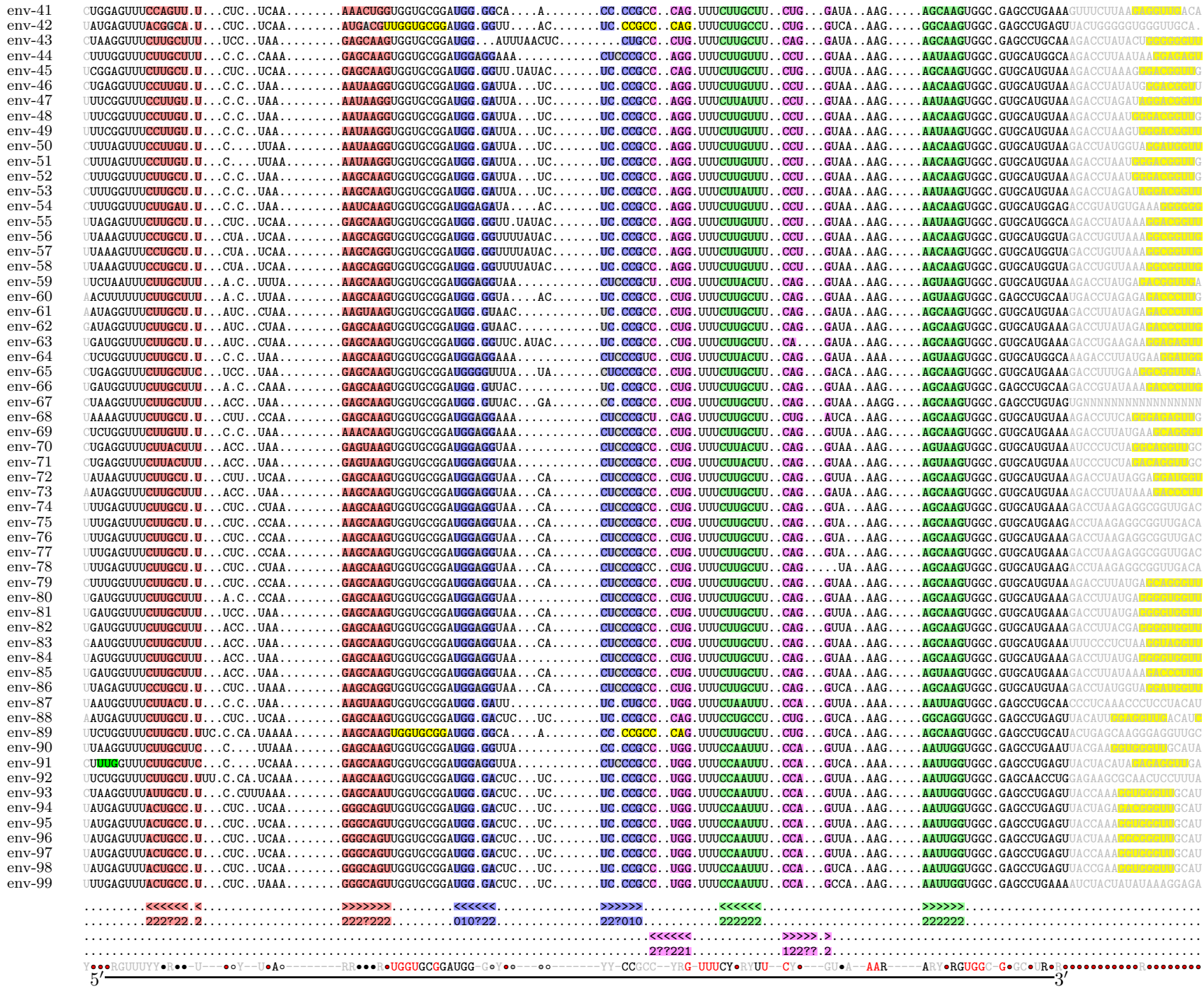
they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (◎): 75%, white circle (◎): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

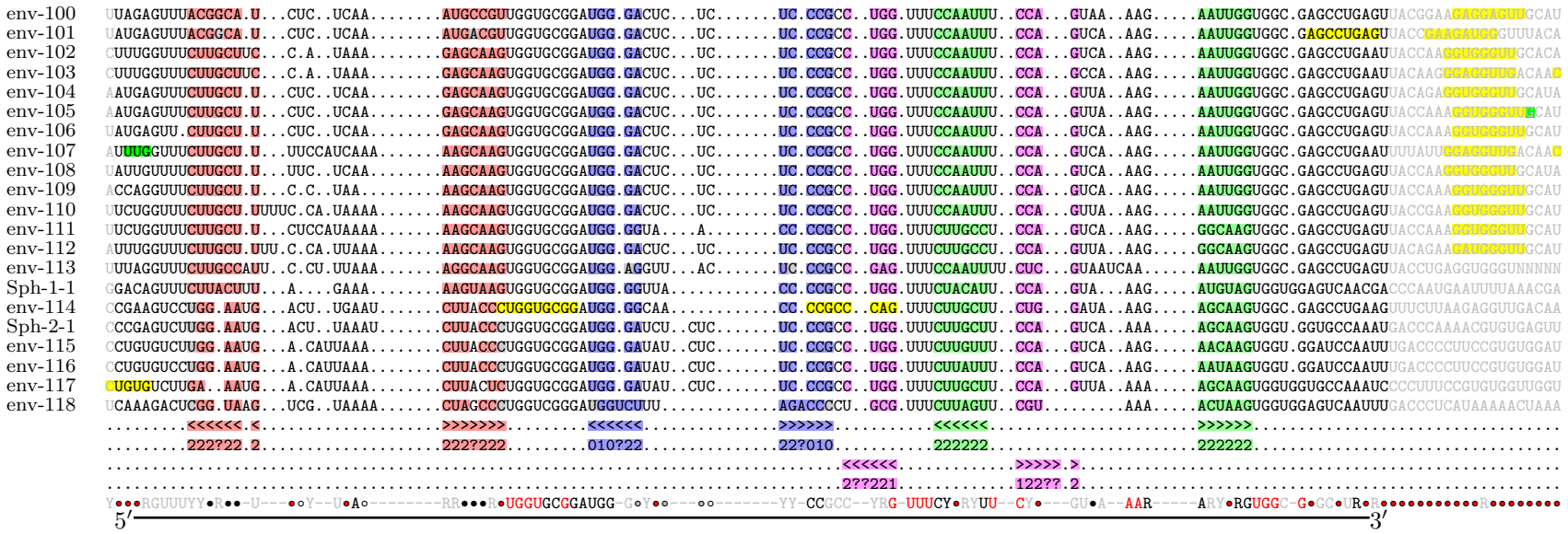
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
env-100
env-101
env-102
env-103
env-104
env-105
env-106
env-107
env-108



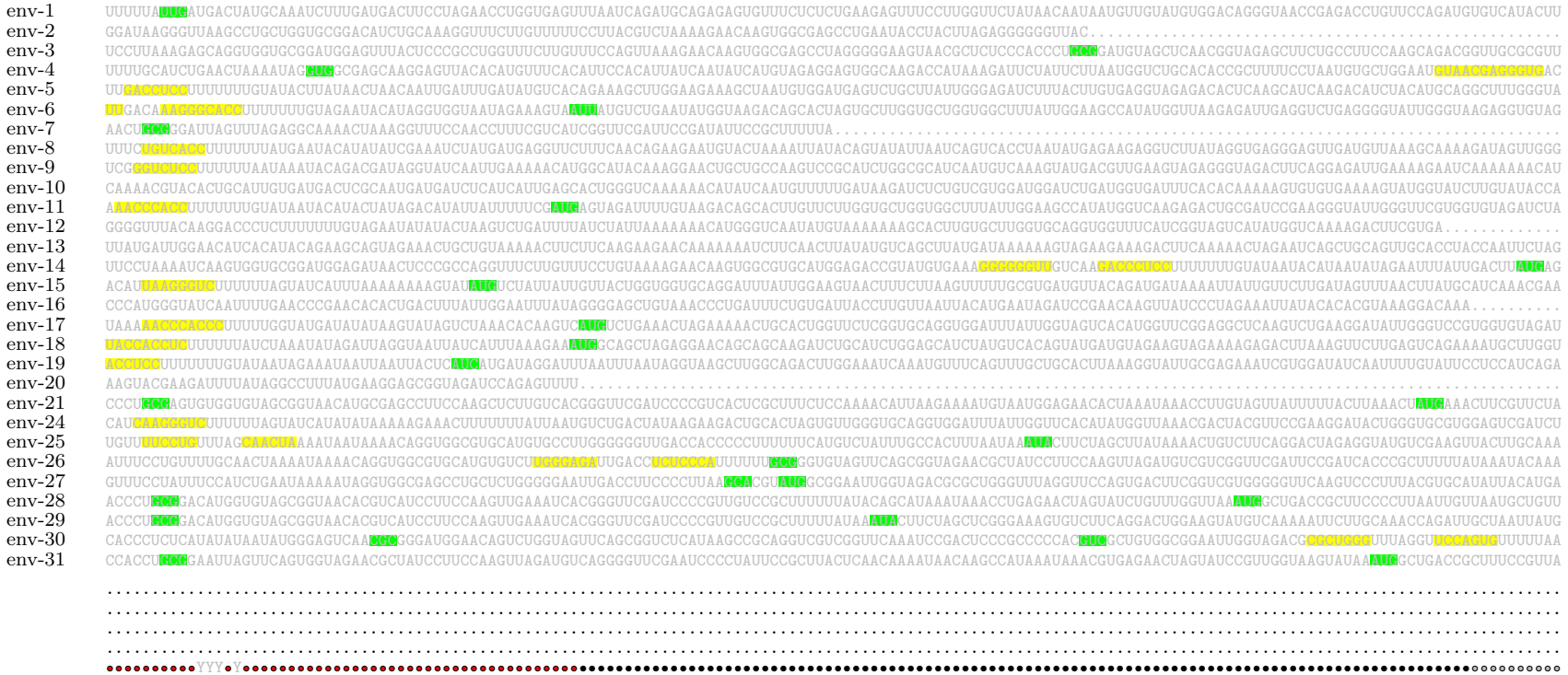
alignment positions 100 · · · 259

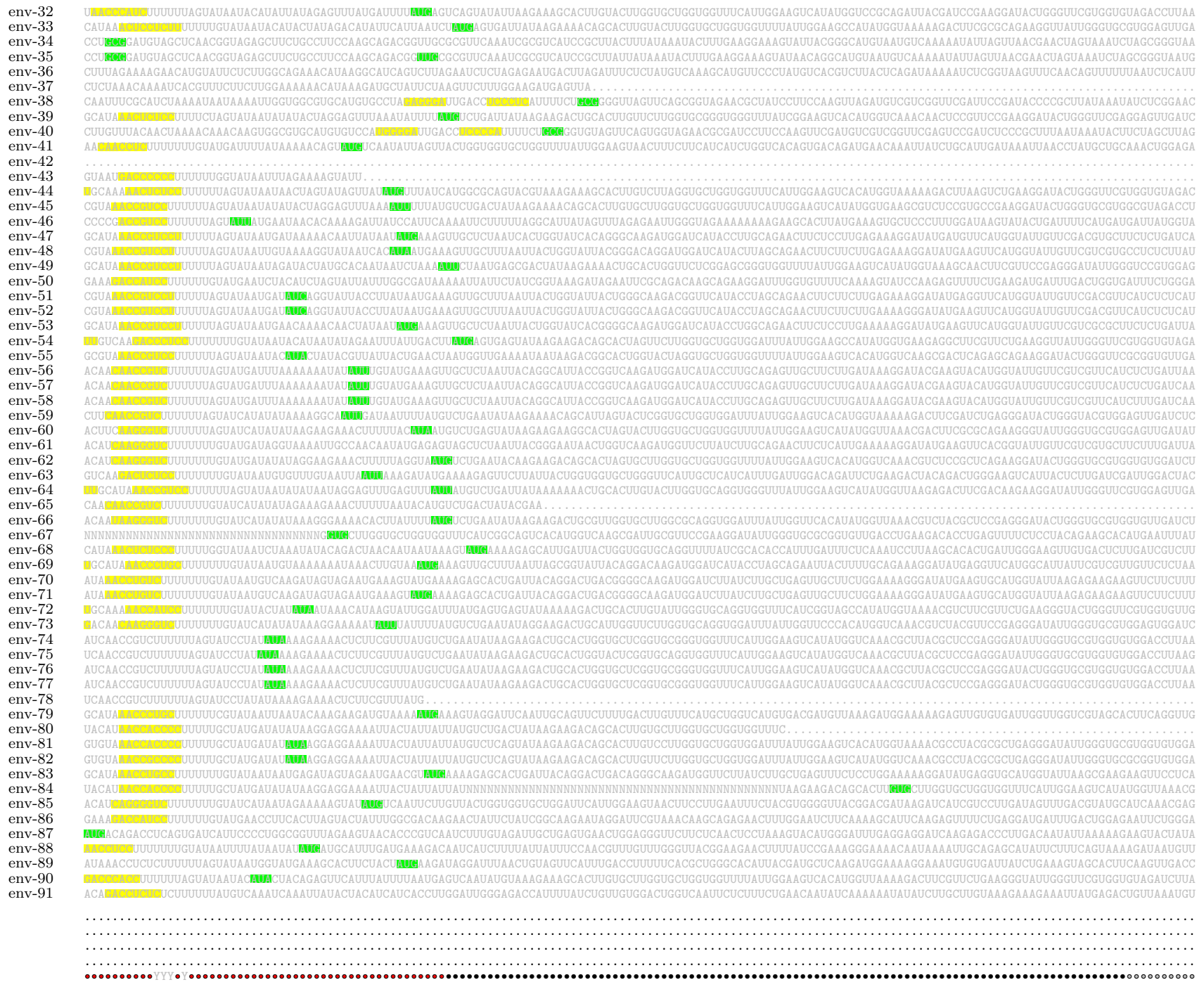


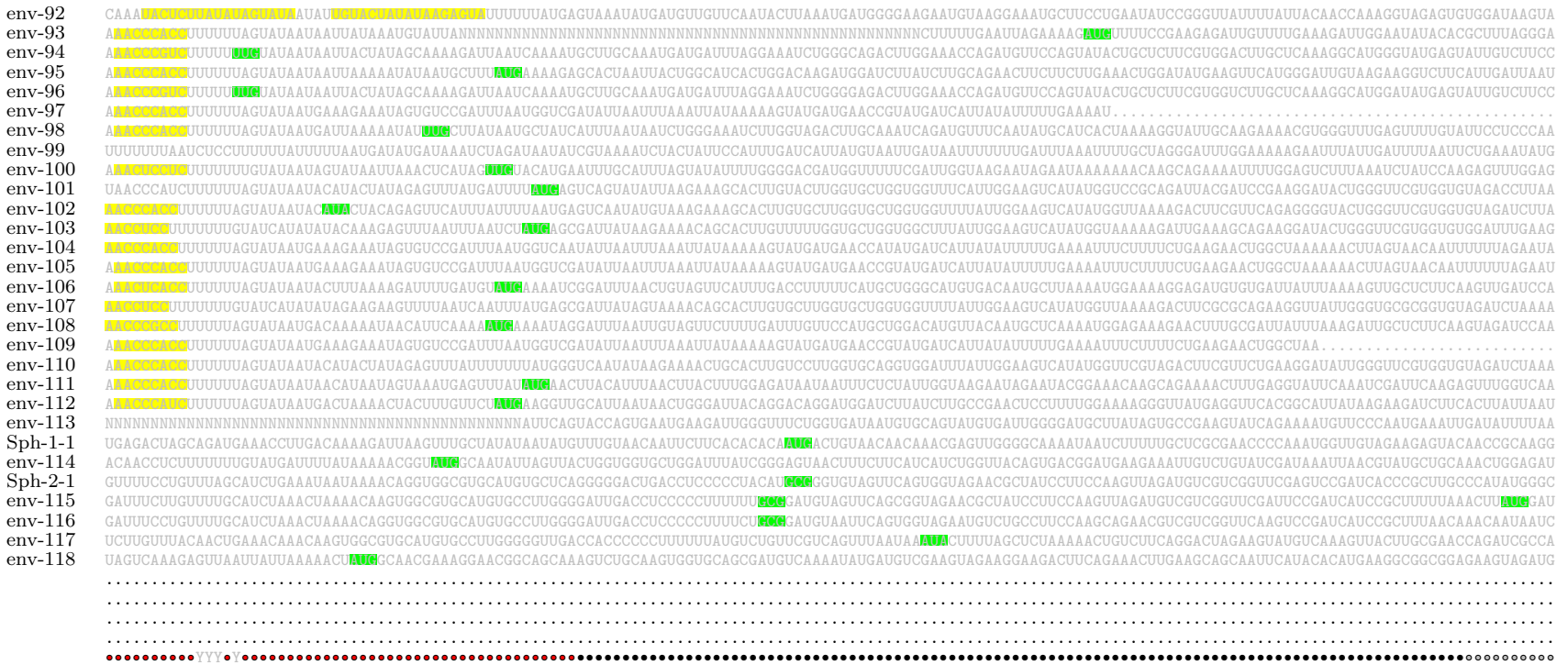

env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91
env-92
env-93
env-94
env-95
env-96
env-97
env-98
env-99
5′
3′



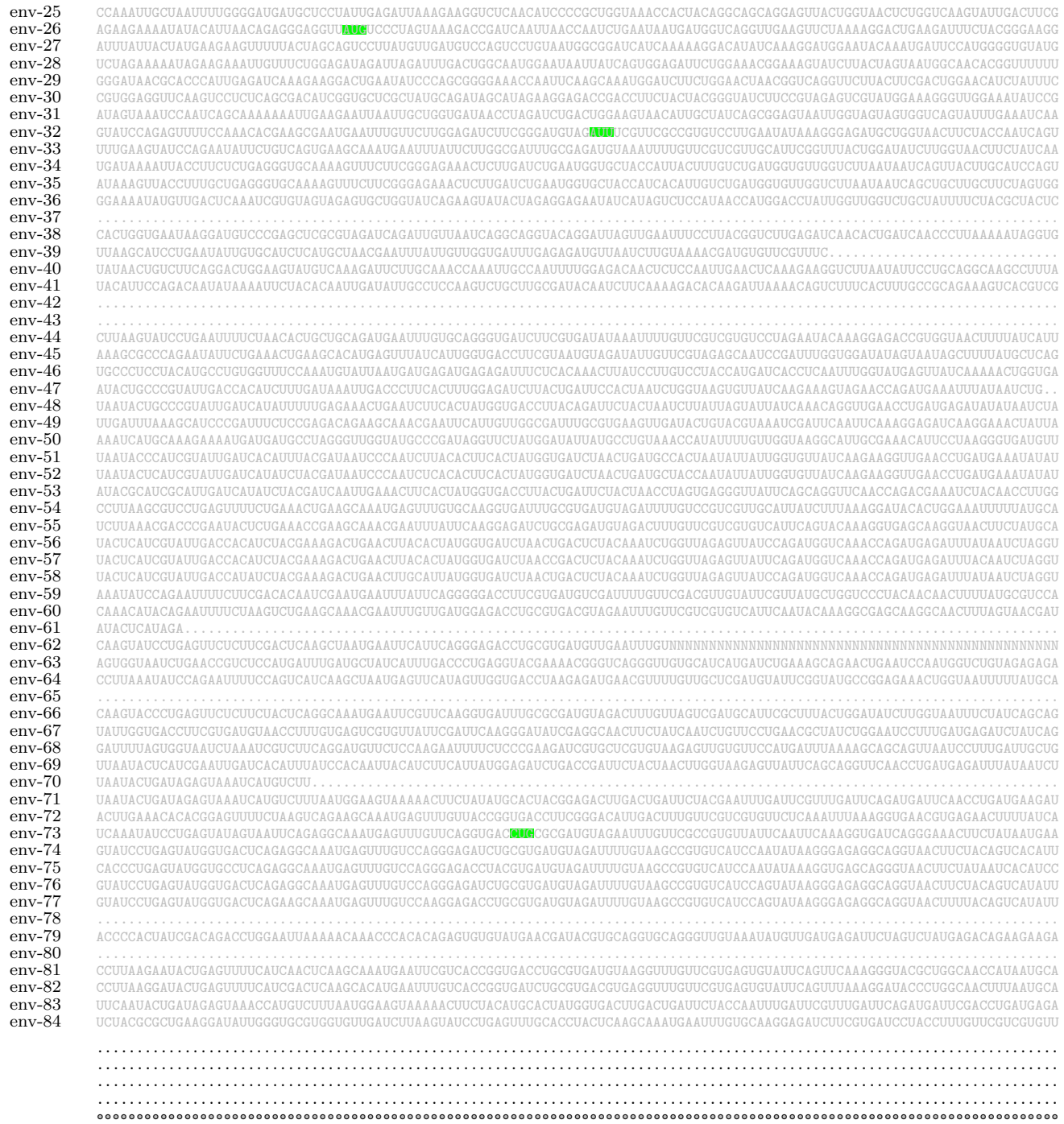
alignment positions 260 · · · 419




env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84
env-85
env-86
env-87
env-88
env-89
env-90
env-91



alignment positions 420 · · · 540


env-25
env-26
env-27
env-28
env-29
env-30
env-31
env-32
env-33
env-34
env-35
env-36
env-37
env-38
env-39
env-40
env-41
env-42
env-43
env-44
env-45
env-46
env-47
env-48
env-49
env-50
env-51
env-52
env-53
env-54
env-55
env-56
env-57
env-58
env-59
env-60
env-61
env-62
env-63
env-64
env-65
env-66
env-67
env-68
env-69
env-70
env-71
env-72
env-73
env-74
env-75
env-76
env-77
env-78
env-79
env-80
env-81
env-82
env-83
env-84

102 Whalefall-1

102.1 Taxa

The taxonomy of each organism containing a putative Whalefall-1 RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual Whalefall-1 RNA in Sections 102.2 and 102.4.

abbrev. of hits	**taxonomy of species**
env-1 to env-16	environmental samples

102.2 Gene contexts

Each Whalefall-1 RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one Whalefall-1 RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 102.3. The accession number of the sequence containing each Whalefall-1 RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 102.1.

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	wf1_Contig538	+	699	752	**RNA**→
env-2	wf2_Contig5173	+	1823	1878	COG3021 (COG3021)→ **RNA**→ COG4922 (COG4922)→
env-3	wf2_Contig3569	+	1214	1270	hypo→ **RNA**→
env-4	wf2_Contig4944	+	714	769	**RNA**→
env-5	wf2_Contig42	+	200	255	**RNA**→ FolA (COG0262)→
env-6	wf2_Contig168	+	688	743	**RNA**→
env-7	wf2_Contig264	-	127	72	←hypo **RNA**→
env-8	wf2_Contig5380	-	778	723	Glyco_hydro_20 (pfam00728)→ **RNA**→ ←Na+/H+ antiporter
env-9	wf2_Contig264	+	161	216	**RNA**→ hypo→
env-10	wf2_Contig5380	+	812	867	Na+/H+ antiporter→ **RNA**→ ←Glyco_hydro_20 (pfam00728)
env-11	wf2_Contig5173	-	1772	1717	←COG4922 (COG4922) **RNA**→ ←COG3021 (COG3021)
env-12	wf2_Contig42	-	142	87	←FolA (COG0262) **RNA**→
env-13	wf2_Contig1327	-	105	50	**RNA**→
env-14	wf2_Contig1741	-	360	306	**RNA**→
env-15	wf2_Contig4944	-	663	608	**RNA**→
env-16	wf2_Contig3569	-	1164	1109	**RNA**→ ←hypo

102.3 Conserved domains

Conserved domains found in protein-coding genes listed in Section 102.2 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one Whalefall-1 RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

`uD` COG0262 (2) Dihydrofolate reductase [Coenzyme metabolism]
`u d` COG3021 (2) Uncharacterized protein conserved in bacteria [Function unknown]
`uD` COG4922 (2) Uncharacterized protein conserved in bacteria [Function unknown]

`u d` pfam00728 (2) Glycosyl hydrolase family 20, catalytic domain.

102.4 Multiple-sequence alignment

Each Whalefall-1 RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 102.1. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The Whalefall-1 RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs

exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

103 *yjdF*

103.1 Taxa

The taxonomy of each organism containing a putative *yjdF* RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *yjdF* RNA in Sections 103.3 and 103.5.

abbrev. of hits	taxonomy of species
cdi-1-1	Bacteria candidate division TM7 single-cell isolate TM7a
Bam-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus amyloliquefaciens* FZB42
Bsu-1-1	Bacteria Firmicutes Bacillales Bacillaceae *Bacillus subtilis subsp. subtilis str.* 168
Ban-1-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* 'Ames Ancestor'
Ban-2-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* A1055
Ban-3-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* A2012
Ban-4-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Ames
Ban-5-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Australia 94

Ban-6-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* CNEVA-9066
Ban-7-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Kruger B
Ban-8-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Sterne
Ban-9-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Vollum
Ban-10-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Western North America USA6153
Bce-1-1 to Bce-1-2	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* 03BB108
Bce-2-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* AH1134
Bce-3-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* AH820
Bce-4-1 to Bce-4-2	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* ATCC 10987
Bce-5-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* ATCC 14579
Bce-6-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* B4264
Bce-7-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* E33L
Bce-8-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* G9241
Bce-9-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* G9842
Bce-10-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* NVH0597-99
Bce-12-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus subsp. cytotoxis* NVH 391-98
Bce-11-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* W
Bth-1-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus thuringiensis serovar israelensis* ATCC 35646
Bth-2-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus thuringiensis serovar konkukian str.* 97-27
Bth-3-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus thuringiensis str.* Al Hakam
Bwe-1-1	Bacteria Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus weihenstephanensis* KBAB4
Pla-1-1	Bacteria Firmicutes Bacillales Paenibacillaceae *Paenibacillus larvae subsp. larvae* BRL-230010
Sau-1-1	Bacteria Firmicutes Bacillales *Staphylococcus aureus* RF122
Sau-2-1	Bacteria Firmicutes Bacillales *Staphylococcus aureus subsp. aureus* MSSA476
Sau-3-1	Bacteria Firmicutes Bacillales *Staphylococcus aureus subsp. aureus* MW2
Sep-1-1	Bacteria Firmicutes Bacillales *Staphylococcus epidermidis* ATCC 12228
Sep-2-1	Bacteria Firmicutes Bacillales *Staphylococcus epidermidis* RP62A
Sha-1-1	Bacteria Firmicutes Bacillales *Staphylococcus haemolyticus* JCSC1435
Cac-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium acetobutylicum* ATCC 824
Cbe-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium beijerinckii* NCIMB 8052
Cbo-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium bolteae* ATCC BAA-613
Cbo-2-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* A str. ATCC 19397
Cbo-3-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* A str. ATCC 3502
Cbo-4-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* A str. Hall
Cbo-5-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* Bf
Cbo-6-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* C str. Eklund
Cbo-7-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* F str. Langeland
Cbo-8-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* G
Cbo-9-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum* NCTC 2916
Cbo-10-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium botulinum str.* Iwanei E
Cbu-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium butyricum* 5521
Cdi-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium difficile* 630
Cdi-2-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium difficile* QCD-32g58
Cdi-3-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium difficile* QCD-66c26
Ckl-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium kluyveri* DSM 555
Csp-1-1 to Csp-1-2	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium sp.* L2-50
Cte-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium tetani* E88
Fpr-1-1	Bacteria Firmicutes Clostridia Clostridiales Clostridiaceae *Faecalibacterium prausnitzii* M21/2
Aca-1-1	Bacteria Firmicutes Clostridia Clostridiales Lachnospiraceae *Anaerostipes caccae* DSM 14662
Dha-1-1	Bacteria Firmicutes Clostridia Clostridiales Peptococcaceae *Desulfitobacterium hafniense* DCB-2
Dha-2-1	Bacteria Firmicutes Clostridia Clostridiales Peptococcaceae *Desulfitobacterium hafniense* Y51
Rob-1-1	Bacteria Firmicutes Clostridia Clostridiales Ruminococcaceae *Ruminococcus obeum* ATCC 29174
Lbr-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus brevis* ATCC 367
Lca-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus casei* ATCC 334
Lga-1-1 to Lga-1-2	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus gasseri* ATCC 33323
Ljo-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus johnsonii* NCC 533
Lpl-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus plantarum* WCFS1
Lre-1-1	Bacteria Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus reuteri* 100-23
Sag-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 18RS21
Sag-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 2603V/R
Sag-3-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 515
Sag-4-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* A909
Sag-5-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* CJB111
Sag-6-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* COH1
Sag-7-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* H36B

Sag-8-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* NEM316
Smu-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus mutans* UA159
Ssa-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus sanguinis* SK36
Sth-1-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* CNRZ1066
Sth-2-1	Bacteria Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* LMG 18311
env-1 to env-28	environmental samples

103.2 Notes

Superscript numbers are used to annotate any *yjdF* RNA that has special characteristics (described below). These numeric annotations will be used in Sections 103.3 and 103.5. The meaning of the numbers are as follows.

[1] This sequence does not contain a likely Shine-Dalgarno sequence (or nearby annotated start codon) in the expected location that can be sequestered by the alternate stem.

[2] This sequence has a downstream transcription terminator, where the 5′ part of the stem of the terminator may be sequestered by the alternate stem. Thus, the sequence might be transcriptionally regulated.

[3] This sequence has an abberant sequence for P4, and all sequence features 3′ to that.

103.3 Gene contexts

Each *yjdF* RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *yjdF* RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 103.4. The accession number of the sequence containing each *yjdF* RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 103.1. Superscript numbers refer to annotations listed in Section 103.2

abbrev.	Seq. accession		5′ at	3′ at	genes
env-1	BABD01035521.1	-	272	169	**RNA**→
env-2	hgutS7_s7_171693	+	768	874	**RNA**→ YjdF (YjdF)→
env-3	hgutS7_s7_161053	+	2210	2316	**RNA**→ YjdF (YjdF)→
env-4	mgutLn3_U_BK_aaa32c03_b1	-	686	570	**RNA**→ YjdF (YjdF)→
[3]Sth-2-1	NC_006448.1	+	228890	229028	**RNA**→ NAPRTase_A (cd01570)→ NadE (COG0171)→ DUF805 (pfam05656)→
Sth-1-1	NC_006449.1	+	228896	229034	**RNA**→ NAPRTase_A (cd01570)→ NadE (COG0171)→ DUF805 (pfam05656)→
Smu-1-1	NC_004350.1	-	367605	367494	**RNA**→ NOG18391 (NOG18391) (YjdF)→
Sag-4-1	NC_007432.1	+	99777	99884	**RNA**→ NOG18391 (NOG18391) (YjdF)→
Sag-7-1	NZ_AAJS01000038.1	+	4969	5076	**RNA**→ YjdF (YjdF)→
Sag-2-1	NC_004116.1	+	100059	100168	**RNA**→ YjdF (YjdF)→
Sag-8-1	NC_004368.1	+	100087	100196	**RNA**→ YjdF (YjdF)→
Sag-1-1	NZ_AAJO01000035.1	-	11328	11219	**RNA**→ YjdF (YjdF)→
Sag-3-1	NZ_AAJP01000070.1	+	4621	4730	**RNA**→ YjdF (YjdF)→
Sag-5-1	NZ_AAJQ01000035.1	+	5391	5500	**RNA**→ YjdF (YjdF)→
Sag-6-1	NZ_AAJR01000071.1	-	2075	1968	**RNA**→ YjdF (YjdF)→
Cbo-5-1	NZ_ABDP01000029.1	+	23660	23767	**RNA**→

Cbo-7-1	NC_009699.1	-	2191605	2191498	**RNA**→ YjdF (YjdF)→ RpiR (COG1737)→
Cbo-3-1	NC_009495.1	-	2133357	2133250	**RNA**→ YjdF (YjdF)→
Cbo-2-1	NC_009697.1	-	2061998	2061891	**RNA**→ YjdF (YjdF)→
Cbo-4-1	NC_009698.1	-	2062217	2062110	**RNA**→ RpiR (COG1737)→
Cbo-9-1	NZ_ABDO01000027.1	+	26264	26371	**RNA**→
Cbo-10-1	NZ_ABDR01000006.1	+	198390	198497	**RNA**→
Cbo-8-1	NZ_ABDS01000027.1	-	9321	9214	**RNA**→
env-5	BAAY01022519.1	+	731	858	**RNA**→
env-6	BAAX01018354.1	-	777	637	**RNA**→
env-7	BAAU01013946.1	-	312	180	**RNA**→
env-8	BAAY01000296.1	-	1743	1603	**RNA**→ YjdF (YjdF)→
Fpr-1-1	NZ_ABED01000035.1	-	71701	71561	**RNA**→
env-9	BABA01004437.1	-	1434	1294	**RNA**→ YjdF (YjdF)→
env-10	BAAX01006676.1	+	492	632	**RNA**→ YjdF (YjdF)→
[1]Dha-2-1	NC_007907.1	+	2980037	2980152	**RNA**→ hypo→
Dha-1-1	NZ_AAAW04000001.1	+	1137932	1138046	**RNA**→ ←NuoG (COG1034)
Lre-1-1	NZ_AAPZ01000005.1	-	19879	19763	**RNA**→ YjdF (YjdF)→
Ljo-1-1	NC_005362.1	+	275419	275532	**RNA**→ NOG18391 (NOG18391) (YjdF)→
Lga-1-1	NC_008530.1	+	273421	273532	**RNA**→ YjdF (YjdF)→ hypo→
Lga-1-2	NC_008530.1	+	275566	275681	**RNA**→ YjdF (YjdF)→
Lca-1-1	NC_008526.1	-	2847352	2847247	**RNA**→ YjdF (YjdF)→
[1,2]Pla-1-1	NZ_AARF01000365.1	+	613	727	**RNA**→
env-11	hgutS8_s8_159859	-	979	851	**RNA**→ YjdF (YjdF)→
Bce-4-1	NC_003909.8	+	1016995	1017114	**RNA**→ YjdF (YjdF)→
Bce-1-1	NZ_ABDM01000004.1	-	15683	15564	**RNA**→
Bsu-1-1	NC_000964.2	+	1275069	1275177	**RNA**→ NOG18391 (NOG18391) (YjdF)→
Bam-1-1	NC_009725.1	-	2746219	2746110	**RNA**→ YjdF (YjdF)→
Bce-12-1	NC_009674.1	-	2923703	2923597	**RNA**→ YjdF (YjdF)→
Bwe-1-1	NZ_AAOY01000052.1	-	29052	28944	**RNA**→ YjdF (YjdF)→
Bce-8-1	NZ_AAEK01000054.1	+	27482	27591	**RNA**→ YjdF (YjdF)→
Bce-4-2	NC_003909.8	-	3941275	3941165	**RNA**→ YjdF (YjdF)→
Bce-10-1	NZ_ABDK01000052.1	+	1053	1162	**RNA**→
Bce-7-1	NC_006274.1	-	4030705	4030596	**RNA**→ YjdF (YjdF)→
Bth-1-1	NZ_AAJM01000387.1	-	527	419	**RNA**→ YjdF (YjdF)→
Bce-9-1	NZ_ABDJ01000002.1	-	298383	298275	**RNA**→
Bce-2-1	NZ_ABDA01000005.1	+	109676	109784	**RNA**→
Bce-5-1	NC_004722.1	-	4116917	4116809	**RNA**→ YjdF (YjdF)→
Bce-6-1	NZ_ABDI01000010.1	+	127567	127675	**RNA**→
Ban-3-1	NZ_AAAC02000001.1	-	4442224	4442115	**RNA**→ DUF1094 (pfam06491)→
Ban-2-1	NZ_AAEO01000019.1	-	334800	334691	**RNA**→
Ban-8-1	NC_005945.1	-	3993678	3993569	**RNA**→ YjdF (YjdF)→
Ban-1-1	NC_007530.2	-	3993306	3993197	**RNA**→ YjdF (YjdF)→
Ban-4-1	NC_003997.3	-	3993179	3993070	**RNA**→ NOG14194 (NOG14194) (YjdF)→
Ban-9-1	NZ_AAEP01000031.1	+	112300	112409	**RNA**→
Ban-10-1	NZ_AAER01000035.1	-	506024	505915	**RNA**→
Ban-5-1	NZ_AAES01000022.1	+	660079	660188	**RNA**→
Ban-6-1	NZ_AAEN01000013.1	-	247688	247579	**RNA**→
Ban-7-1	NZ_AAEQ01000038.1	-	269069	268960	**RNA**→
Bth-3-1	NC_008600.1	-	3979564	3979454	**RNA**→ YjdF (YjdF)→
Bce-1-2	NZ_ABDM01000016.1	+	131030	131140	**RNA**→
Bce-3-1	NZ_AAUE01000003.1	-	291718	291608	**RNA**→
Bce-11-1	NZ_ABCZ01000008.1	-	123755	123645	**RNA**→

Bth-2-1	NC_005957.1	-	3979781	3979671	**RNA**→ NOG14194 (NOG14194) (YjdF)→
Lbr-1-1	NC_008497.1	-	2190740	2190629	**RNA**→ YjdF (YjdF)→
Cbu-1-1	NZ_ABDT01000028.1	-	47084	46979	**RNA**→
Ckl-1-1	NC_009706.1	-	3572617	3572507	**RNA**→ YjdF (YjdF)→
Cbe-1-1	NC_009617.1	-	1839650	1839541	**RNA**→ YjdF (YjdF)→
Cbo-6-1	NZ_ABDQ01000001.1	+	474183	474289	**RNA**→
Cte-1-1	NC_004557.1	+	625938	626043	**RNA**→ NOG14194 (NOG14194) (YjdF)→
Aca-1-1	NZ_ABAX01000014.1	-	163914	163803	**RNA**→
env-12	BABE01003771.1	-	405	298	**RNA**→ YjdF (YjdF)→
env-13	BABG01026286.1	+	697	819	**RNA**→
Csp-1-1	NZ_AAYW01000011.1	-	104977	104847	**RNA**→
env-14	BAAV01003796.1	-	1332	1200	**RNA**→
env-15	BABG01006156.1	-	1375	1241	**RNA**→ YjdF (YjdF)→ hypo→ hypo→
env-16	hgutS8_s8_176847	-	1594	1461	**RNA**→ YjdF (YjdF)→
env-17	hgutS8_s8_169626	+	2932	3056	**RNA**→
env-18	BABD01001501.1	+	1231	1364	**RNA**→ YjdF (YjdF)→ RimI (COG0456)→
Csp-1-2	NZ_AAYW01000013.1	+	30067	30200	**RNA**→
env-19	hgutS8_s8_161563	-	649	516	**RNA**→ YjdF (YjdF)→
env-20	BAAX01025816.1	+	7	140	**RNA**→
env-21	BAAV01035733.1	-	423	292	**RNA**→
Rob-1-1	NZ_AAVO02000014.1	+	31411	31540	**RNA**→ YjdF (YjdF)→
env-22	hgutS8_s8_174773	+	321	450	**RNA**→ ←hypo
Sha-1-1	NC_007168.1	+	192943	193053	**RNA**→ NOG18391 (NOG18391) (YjdF)→
Sep-2-1	NC_002976.3	+	2348458	2348569	**RNA**→ ←RimL (COG1670)
Sep-1-1	NC_004461.1	-	268742	268631	**RNA**→ NOG14194 (NOG14194) (YjdF)→
Sau-1-1	NC_007622.1	+	408566	408677	**RNA**→ NOG18391 (NOG18391) (YjdF)→ DUF523 (pfam04463)→
Sau-2-1	NC_002953.3	+	423980	424091	**RNA**→ YjdF (YjdF)→ DUF523 (pfam04463)→
Sau-3-1	NC_003923.1	+	425335	425446	**RNA**→ YjdF (YjdF)→ DUF523 (pfam04463)→
Lpl-1-1	NC_004567.1	-	1767848	1767740	**RNA**→ NOG14194 (NOG14194) (YjdF)→
Cac-1-1	NC_003030.1	+	1718166	1718284	**RNA**→ NOG14194 (NOG14194) (YjdF)→
Ssa-1-1	NC_009009.1	-	2058556	2058444	**RNA**→ YjdF (YjdF)→ hypo→ hypo→ TRX_family (cd02947)→
Cdi-1-1	NC_009089.1	+	2040488	2040594	**RNA**→ YjdF (YjdF)→
Cdi-2-1	NZ_AAML04000011.1	+	225559	225665	**RNA**→ YjdF (YjdF)→
Cdi-3-1	NZ_ABFD01000045.1	+	70809	70915	**RNA**→
cdi-1-1	NZ_ABBV01000209.1	-	1887	1784	**RNA**→
env-23	BAAZ01016242.1	-	137	25	**RNA**→
env-24	BABF01001261.1	+	3242	3358	**RNA**→ YjdF (YjdF)→
env-25	BABA01034220.1	+	495	608	**RNA**→
Cbo-1-1	NZ_ABCC01000057.1	-	34328	34171	**RNA**→
env-26	BAAV01030689.1	-	295	177	**RNA**→
env-27	BAAV01036128.1	+	32	151	**RNA**→
env-28	BABD01028836.1	+	160	279	**RNA**→

103.4 Conserved domains

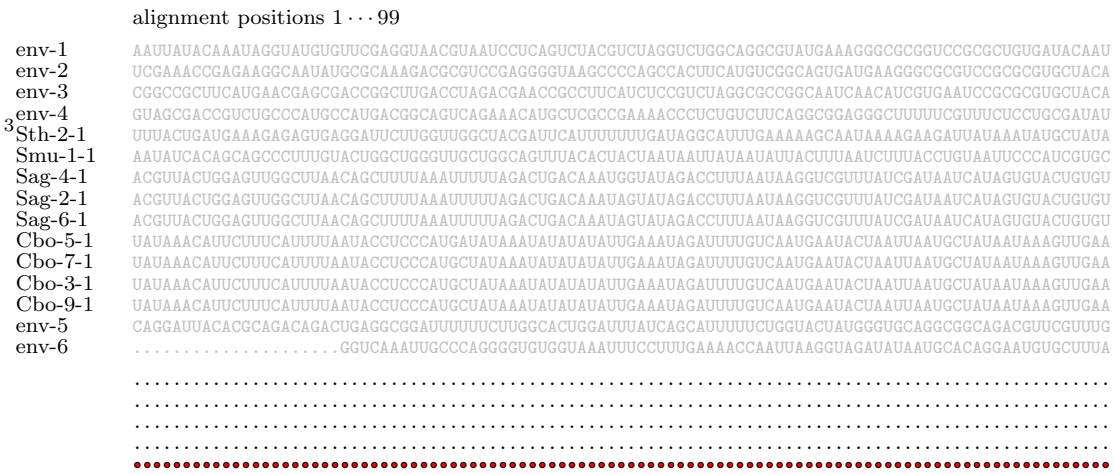
Conserved domains found in protein-coding genes listed in Section 103.3 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *yjdF* RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

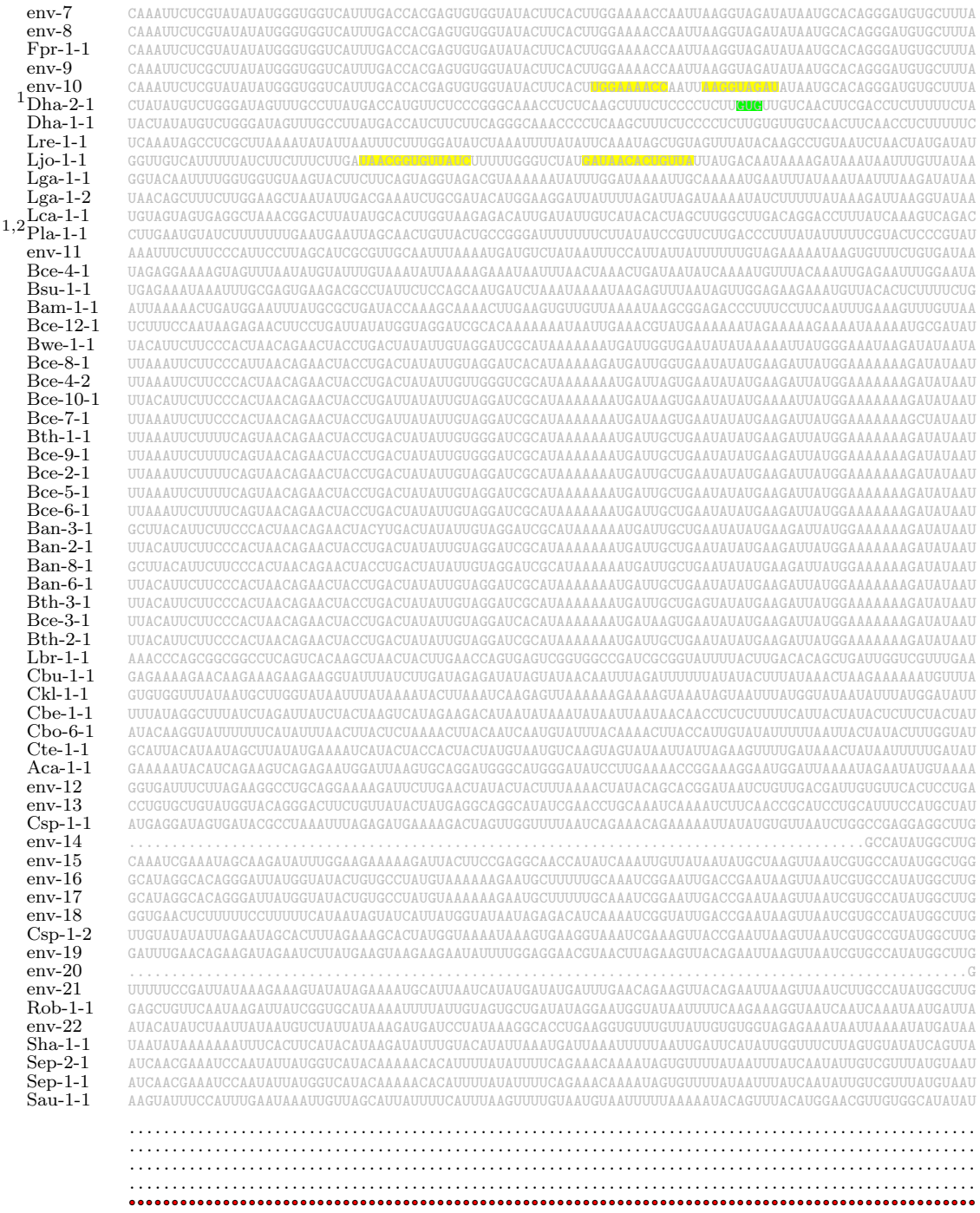
cd01570 (2) Nicotinate phosphoribosyltransferase (NAPRTase), subgroup A.
cd02947 (1) TRX family; composed of two groups: Group I, which includes proteins that exclusively encode a TRX domain; and Group II, which are composed of fusion proteins of TRX and additional domains.
COG0171 (2) NAD synthase [Coenzyme metabolism]
COG0456 (1) Acetyltransferases [General function prediction only]
COG1034 (1) NADH dehydrogenase/NADH:ubiquinone oxidoreductase 75 kD subunit (chain G) [Energy production and conversion]
COG1670 (1) Acetyltransferases, including N-acetylases of ribosomal proteins [Translation, ribosomal structure and biogenesis]
COG1737 (2) Transcriptional regulators [Transcription]
NOG14194 (6)
NOG18391 (6)
pfam04463 (3) Protein of unknown function (DUF523).
pfam05656 (2) Protein of unknown function (DUF805).
pfam06491 (1) Protein of unknown function (DUF1094).
YjdF (60) YjdF protein

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Ban-1-1, Ban-10-1, Ban-4-1, Ban-5-1, Ban-7-1, Ban-9-1, Bce-1-1, Bce-1-2, Bce-11-1, Cbo-10-1, Cbo-2-1, Cbo-4-1, Cbo-8-1, Cdi-3-1, Sag-1-1, Sag-3-1, Sag-5-1, Sag-7-1, Sag-8-1, Sau-3-1, Sth-1-1

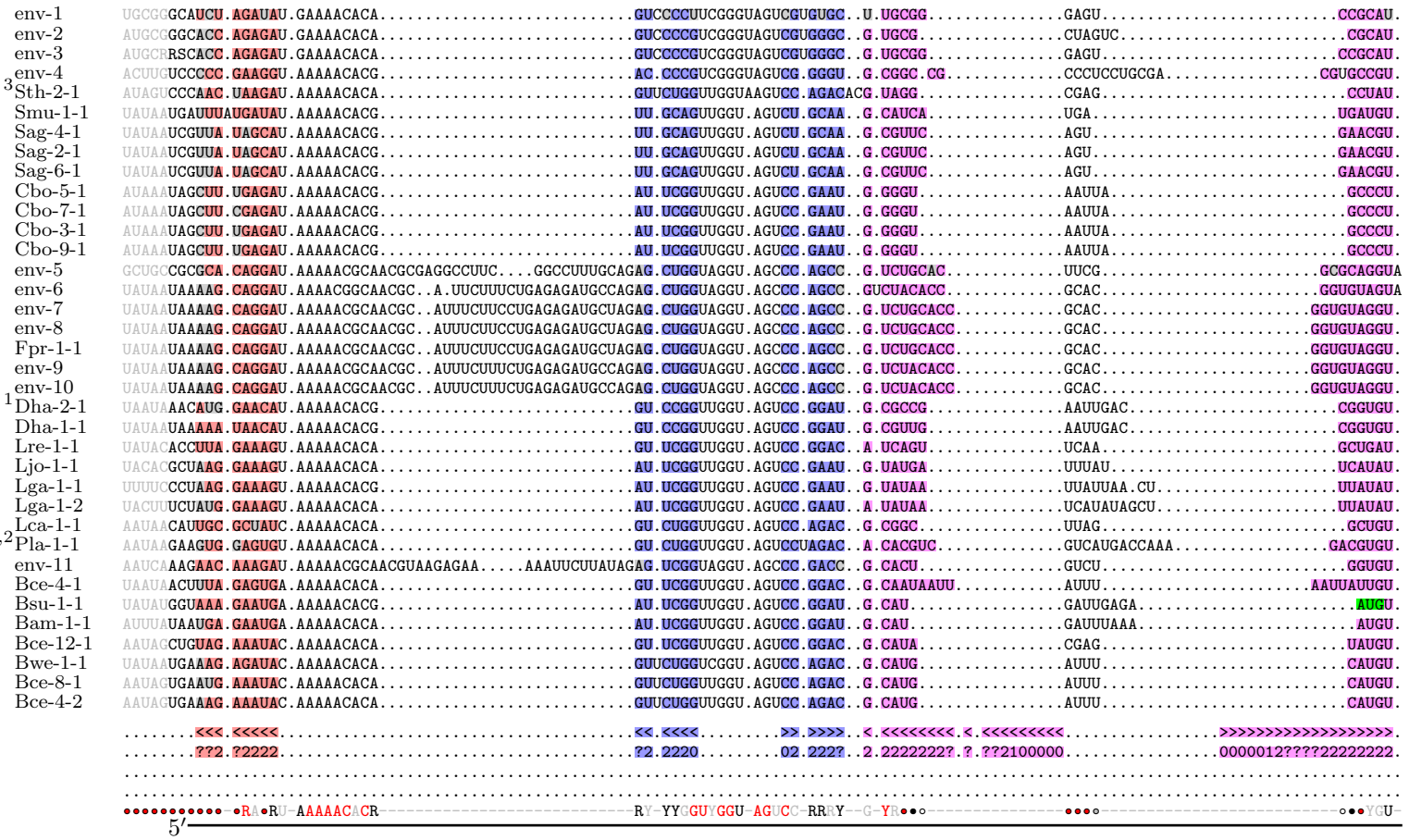
103.5 Multiple-sequence alignment

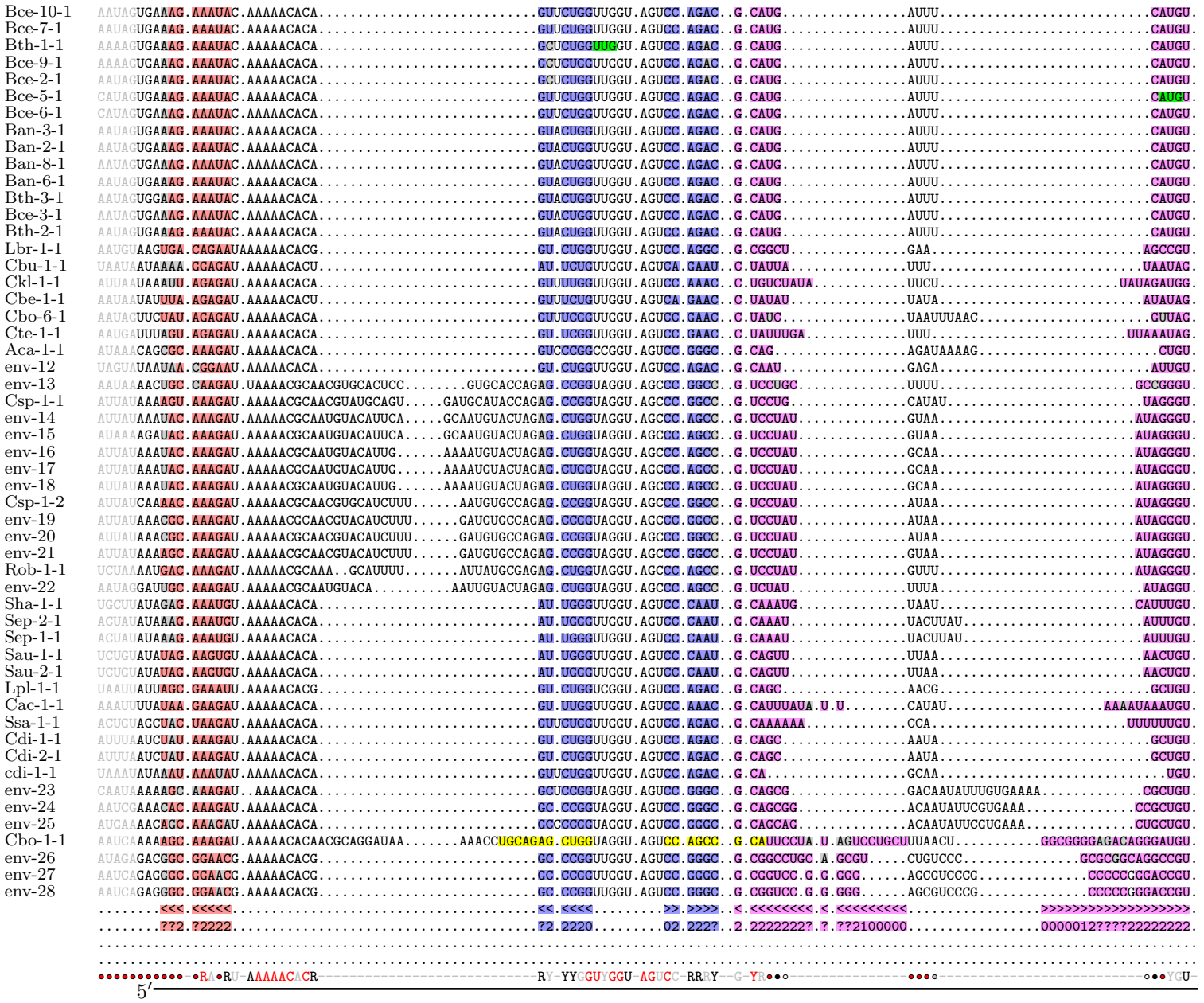
Each *yjdF* RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 103.1. Superscript numbers refer to annotations listed in Section 103.2. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *yjdF* RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.



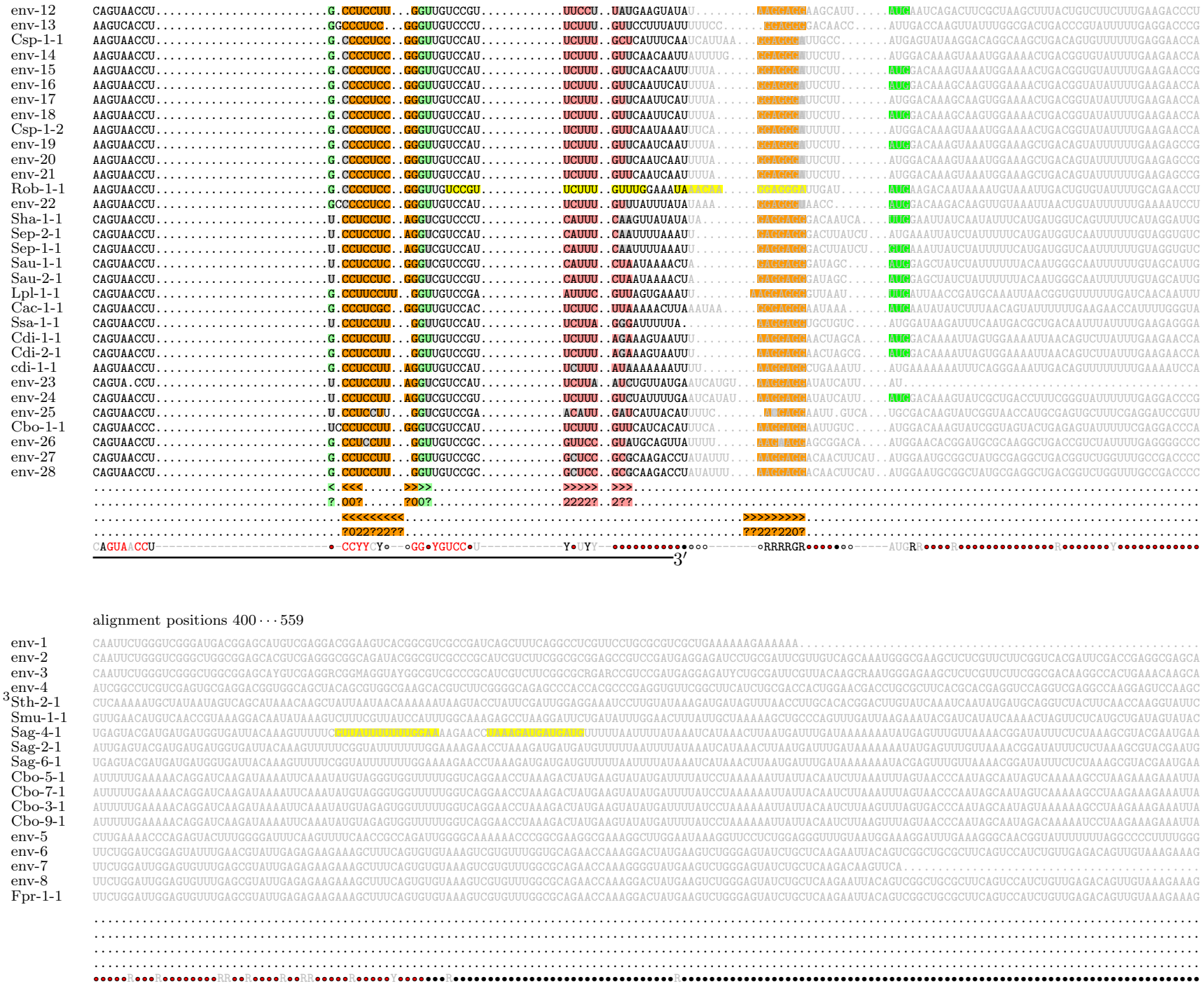

env-7
env-8
Fpr-1-1
env-9
env-10
1 Dha-2-1
Dha-1-1
Lre-1-1
Ljo-1-1
Lga-1-1
Lga-1-2
Lca-1-1
1,2 Pla-1-1
env-11
Bce-4-1
Bsu-1-1
Bam-1-1
Bce-12-1
Bwe-1-1
Bce-8-1
Bce-4-2
Bce-10-1
Bce-7-1
Bth-1-1
Bce-9-1
Bce-2-1
Bce-5-1
Bce-6-1
Ban-3-1
Ban-2-1
Ban-8-1
Ban-6-1
Bth-3-1
Bce-3-1
Bth-2-1
Lbr-1-1
Cbu-1-1
Ckl-1-1
Cbe-1-1
Cbo-6-1
Cte-1-1
Aca-1-1
env-12
env-13
Csp-1-1
env-14
env-15
env-16
env-17
env-18
Csp-1-2
env-19
env-20
env-21
Rob-1-1
env-22
Sha-1-1
Sep-2-1
Sep-1-1
Sau-1-1

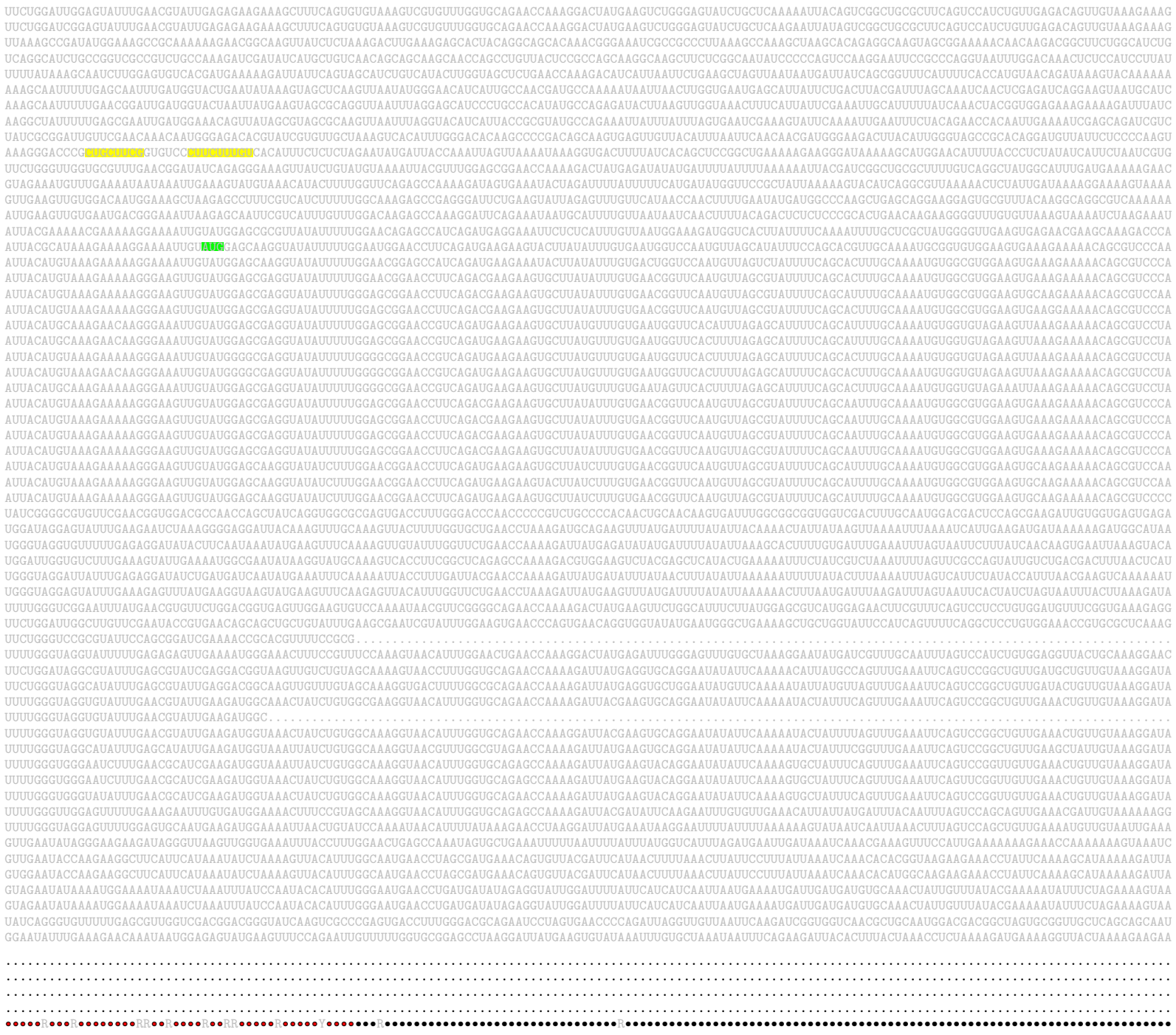
alignment positions 100 · · · 239

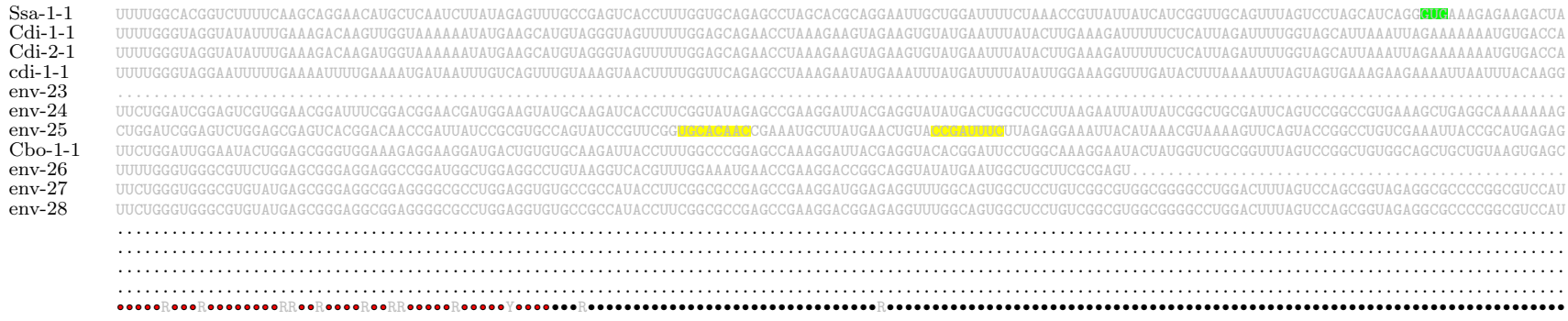



Bce-10-1
Bce-7-1
Bth-1-1
Bce-9-1
Bce-2-1
Bce-5-1
Bce-6-1
Ban-3-1
Ban-2-1
Ban-8-1
Ban-6-1
Bth-3-1
Bce-3-1
Bth-2-1
Lbr-1-1
Cbu-1-1
Ckl-1-1
Cbe-1-1
Cbo-6-1
Cte-1-1
Aca-1-1
env-12
env-13
Csp-1-1
env-14
env-15
env-16
env-17
env-18
Csp-1-2
env-19
env-20
env-21
Rob-1-1
env-22
Sha-1-1
Sep-2-1
Sep-1-1
Sau-1-1
Sau-2-1
Lpl-1-1
Cac-1-1
Ssa-1-1
Cdi-1-1
Cdi-2-1
cdi-1-1
env-23
env-24
env-25
Cbo-1-1
env-26
env-27
env-28
5′

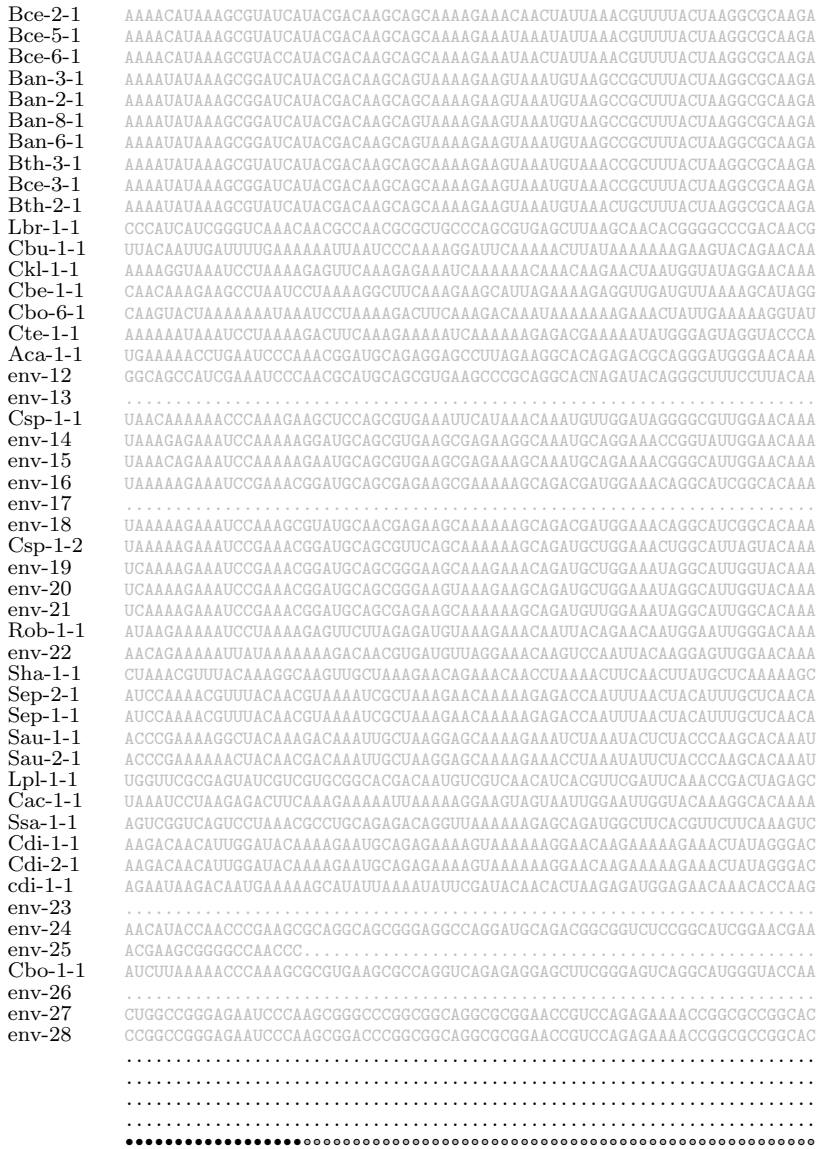
alignment positions 240 · · · 399
env-1
env-2
env-3
env-4
3Sth-2-1
Smu-1-1
Sag-4-1
Sag-2-1
Sag-6-1
Cbo-5-1
Cbo-7-1
Cbo-3-1
Cbo-9-1
env-5
env-6
env-7
env-8
Fpr-1-1
env-9
1env-10
Dha-2-1
Dha-1-1
Lre-1-1
Ljo-1-1
Lga-1-1
Lga-1-2
Lca-1-1
1,2Pla-1-1
env-11
Bce-4-1
Bsu-1-1
Bam-1-1
Bce-12-1
Bwe-1-1
Bce-8-1
Bce-4-2
Bce-10-1
Bce-7-1
Bth-1-1
Bce-9-1
Bce-2-1
Bce-5-1
Bce-6-1
Ban-3-1
Ban-2-1
Ban-8-1
Ban-6-1
Bth-3-1
Bce-3-1
Bth-2-1
Lbr-1-1
Cbu-1-1
Ckl-1-1
Cbe-1-1
Cbo-6-1
Cte-1-1
Aca-1-1
CAGUAACCU
3′


env-12
env-13
Csp-1-1
env-14
env-15
env-16
env-17
env-18
Csp-1-2
env-19
env-20
env-21
Rob-1-1
env-22
Sha-1-1
Sep-2-1
Sep-1-1
Sau-1-1
Sau-2-1
Lpl-1-1
Cac-1-1
Ssa-1-1
Cdi-1-1
Cdi-2-1
cdi-1-1
env-23
env-24
env-25
Cbo-1-1
env-26
env-27
env-28
3′
alignment positions 400 · · · 559
env-1
env-2
env-3
env-4
3Sth-2-1
Smu-1-1
Sag-4-1
Sag-2-1
Sag-6-1
Cbo-5-1
Cbo-7-1
Cbo-3-1
Cbo-9-1
env-5
env-6
env-7
env-8
Fpr-1-1


env-9
env-10
1 Dha-2-1
Dha-1-1
Lre-1-1
Ljo-1-1
Lga-1-1
Lga-1-2
Lca-1-1
1,2 Pla-1-1
env-11
Bce-4-1
Bsu-1-1
Bam-1-1
Bce-12-1
Bwe-1-1
Bce-8-1
Bce-4-2
Bce-10-1
Bce-7-1
Bth-1-1
Bce-9-1
Bce-2-1
Bce-5-1
Bce-6-1
Ban-3-1
Ban-2-1
Ban-8-1
Ban-6-1
Bth-3-1
Bce-3-1
Bth-2-1
Lbr-1-1
Cbu-1-1
Ckl-1-1
Cbe-1-1
Cbo-6-1
Cte-1-1
Aca-1-1
env-12
env-13
Csp-1-1
env-14
env-15
env-16
env-17
env-18
Csp-1-2
env-19
env-20
env-21
Rob-1-1
env-22
Sha-1-1
Sep-2-1
Sep-1-1
Sau-1-1
Sau-2-1
Lpl-1-1
Cac-1-1



alignment positions $560 \cdots 629$



104 *ykkC*-III

104.1 Taxa

The taxonomy of each organism containing a putative *ykkC*-III RNA is listed, with abbreviations identifying each hit (e.g., "Eco-1-1" and "Eco-1-2" might hypothetically represent two distinct RNAs in *E. coli*). The abbreviations will be used to identify each individual *ykkC*-III RNA in Sections 104.3 and 104.5.

abbrev. of hits	**taxonomy of species**
Cps-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Corynebacteriaceae *Corynebacterium pseudogenitalium* ATCC 33035
Mgi-1-1 to Mgi-1-2	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium gilvum* PYR-GCK
Mma-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium marinum* M
Msm-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium smegmatis str.* MC2 155
Msp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium sp.* JLS
Msp-2-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium sp.* KMS
Msp-3-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium sp.* MCS
Mul-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium ulcerans* Agy99
Mva-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium vanbaalenii* PYR-1
Mav-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumavium complex (MAC) *Mycobacterium avium* 104
Mav-2-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumavium complex (MAC) *Mycobacterium avium subs*
Min-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae Mycobacterium Mycobacteriumavium complex (MAC) *Mycobacterium intracellula*
Nfa-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Nocardiaceae *Nocardia farcinica* IFM 10152
Rsp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Nocardiaceae *Rhodococcus sp.* RHA1
Kra-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Frankineae Kineosporiaceae *Kineococcus radiotolerans* SRS30216
Jsp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Intrasporangiaceae *Janibacter sp.* HTCC2649
Cmi-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Microbacteriaceae *Clavibacter michiganensis subsp. michiganensis* NCPPB 382
Cmi-2-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Microbacteriaceae *Clavibacter michiganensis subsp. sepedonicus*
Aau-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Micrococcaceae *Arthrobacter aurescens* TC1
Ach-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Micrococcaceae *Arthrobacter chlorophenolicus* A6
Asp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Micrococcaceae *Arthrobacter sp.* FB24
Rsa-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Micrococcaceae *Renibacterium salmoninarum* ATCC 33209
Scl-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Streptomycineae Streptomycetaceae *Streptomyces clavuligerus* ATCC 27064
Sco-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Streptomycineae Streptomycetaceae *Streptomyces coelicolor* A3(2)
Spr-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Streptomycineae Streptomycetaceae *Streptomyces pristinaespiralis* ATCC 25486
Ssp-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Streptomycineae Streptomycetaceae *Streptomyces sp.* Mg1
Ssp-2-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Streptomycineae Streptomycetaceae *Streptomyces sp.* SPB74
Tfu-1-1	Bacteria Actinobacteria Actinobacteridae Actinomycetales Streptosporangineae Nocardiopsaceae *Thermobifida fusca* YX
mac-1-1	Bacteria Actinobacteria marine actinobacterium PHSC20C1
Dge-1-1	Bacteria Deinococcus-Thermus Deinococci Deinococcales Deinococcaceae *Deinococcus geothermalis* DSM 11300
Bma-1-1	Bacteria Planctomycetes Planctomycetacia Planctomycetales Planctomycetaceae *Blastopirellula marina* DSM 3645
Bba-1-1	Bacteria Proteobacteria Deltaproteobacteria Bdellovibrionales Bdellovibrionaceae *Bdellovibrio bacteriovorus* HD100
Ade-1-1	Bacteria Proteobacteria Deltaproteobacteria Myxococcales Cystobacterineae Myxococcaceae *Anaeromyxobacter dehalogenans* 2CP-C
Asp-2-1	Bacteria Proteobacteria Deltaproteobacteria Myxococcales Cystobacterineae Myxococcaceae *Anaeromyxobacter sp.* K
Mxa-1-1 to Mxa-1-2	Bacteria Proteobacteria Deltaproteobacteria Myxococcales Cystobacterineae Myxococcaceae *Myxococcus xanthus* DK 1622
Sce-1-1	Bacteria Proteobacteria Deltaproteobacteria Myxococcales Sorangiineae Polyangiaceae *Sorangium cellulosum* 'So ce 56'
Lpn-1-1	Bacteria Proteobacteria Gammaproteobacteria Legionellales Legionellaceae *Legionella pneumophila str.* Corby
Lpn-2-1	Bacteria Proteobacteria Gammaproteobacteria Legionellales Legionellaceae *Legionella pneumophila str.* Lens
Lpn-3-1	Bacteria Proteobacteria Gammaproteobacteria Legionellales Legionellaceae *Legionella pneumophila str.* Paris
Lpn-4-1	Bacteria Proteobacteria Gammaproteobacteria Legionellales Legionellaceae *Legionella pneumophila subsp. pneumophila str.* Philadelphia 1
env-1 to env-2	environmental samples

104.2 Notes

Superscript numbers are used to annotate any *ykkC*-III RNA that has special characteristics (described below). These numeric annotations will be used in Sections 104.3 and 104.5. The meaning of the numbers are as follows.

[1] A translated BLAST search indicates that an *emrE* gene is immediately downstream of this RNA.

104.3 Gene contexts

Each *ykkC*-III RNA (indicated by "RNA→") is listed with its surrounding genes. Environmental sequences and some RefSeq entries lack gene annotations, and no genes are listed for such sequences. The direction of each gene is indicated with an arrow (→), and each conserved domain in the gene is colored. Conserved domains associated with more than one *ykkC*-III RNA are assigned a color; other domains are gray. Information about these conserved domains is given in Section 104.4. The accession number of the sequence containing each *ykkC*-III RNA is given. Accessions beginning with "NC_" or "NZ_" are RefSeq, while other accession refer to environmental samples. Nucleotide coordinates are given for the 5′ and 3′ boundaries of each. If the 5′ coordinate is greater than the 3′ coordinate, the RNA is present on the reverse-complement strand of the containing genomic DNA sequence. Each hit is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 104.1. Superscript numbers refer to annotations listed in Section 104.2

abbrev.	Seq. accession		5′ at	3′ at	genes
Mgi-1-1	NC_009338.1	-	5615876	5615812	**RNA**→ EmrE (COG2076)→ EmrE (COG2076)→ OSBS (cd03320)→ COG4977 (COG4977)→ GATase1_PfpI_2 (cd03139)→
Lpn-4-1	NC_002942.5	-	593290	593225	**RNA**→ EmrE (COG2076)→ PlsC (COG0204)→
Lpn-2-1	NC_006369.1	-	644770	644705	**RNA**→ EmrE (COG2076)→ PlsC (COG0204)→
Lpn-3-1	NC_006368.1	-	661276	661211	**RNA**→ EmrE (COG2076)→ PlsC (COG0204)→
Lpn-1-1	NC_009494.1	+	3009308	3009373	**RNA**→ EmrE (COG2076)→ PlsC (COG0204)→
Mxa-1-1	NC_008095.1	+	446679	446738	**RNA**→ EmrE (COG2076)→
Bma-1-1	NZ_AANZ01000001.1	+	452592	452653	**RNA**→ EmrE (COG2076)→ hypo→ hypo→
Bba-1-1	NC_005363.1	-	2412225	2412158	**RNA**→ EmrE (COG2076)→
Aau-1-1	NC_008711.1	+	3894383	3894446	**RNA**→ EmrE (COG2076)→ EmrE (COG2076)→
Asp-1-1	NC_008541.1	+	4240999	4241061	**RNA**→ EmrE (COG2076)→ EmrE (COG2076)→
Ach-1-1	NZ_ABKU01000008.1	-	73600	73537	**RNA**→ small multidrug resistance protein→ EmrE (COG2076)→ CoaA (COG1072)→
[1]env-1	JCVI_SCAF_1096627927426	-	1115	1048	**RNA**→ DUF839 (pfam05787)? (GOS4424)→
Cmi-2-1	NC_010407.1	-	2863085	2863022	**RNA**→ EmrE (COG2076)→ HTH_XRE (cd00093)→ HipA (COG3550)→
Cmi-1-1	NC_009480.1	-	3063871	3063808	**RNA**→ EmrE (COG2076)→
Cps-1-1	NZ_ABYQ01000020.1	-	60074	60010	**RNA**→
mac-1-1	NZ_AAOB01000006.1	-	32110	32049	**RNA**→ EmrE (COG2076)→
Msp-3-1	NC_008146.1	-	1314056	1313992	**RNA**→ EmrE (COG2076)→
Msp-2-1	NC_008705.1	-	1318365	1318301	**RNA**→ EmrE (COG2076)→
Msp-1-1	NC_009077.1	-	1319367	1319303	**RNA**→ EmrE (COG2076)→
Min-1-1	NZ_ABIN01000081.1	+	32257	32321	**RNA**→
Mav-2-1	NC_002944.2	+	3831493	3831556	**RNA**→ EmrE (COG2076)→
Mav-1-1	NC_008595.1	+	4438905	4438968	**RNA**→ EmrE (COG2076)→
Msm-1-1	NC_008596.1	-	1755118	1755054	**RNA**→ EmrE (COG2076)→
Rsp-1-1	NC_008268.1	-	5005863	5005803	**RNA**→ EmrE (COG2076)→ hypo→ hypo→ hypo→
env-2	BABE01000739.1	+	1806	1870	**RNA**→
Kra-1-1	NC_009664.1	-	2110497	2110434	**RNA**→ EmrE (COG2076)→
Nfa-1-1	NC_006361.1	+	133448	133513	**RNA**→ EmrE (COG2076)→
Mma-1-1	NC_010612.1	+	3478073	3478137	**RNA**→ EmrE (COG2076)→
Mul-1-1	NC_008611.1	-	3207260	3207196	**RNA**→ EmrE (COG2076)→
Jsp-1-1	NZ_AAMN01000002.1	+	564690	564754	**RNA**→ EmrE (COG2076)→
Dge-1-1	NC_008025.1	+	2285879	2285949	**RNA**→ EmrE (COG2076)→
Scl-1-1	NZ_ABJH01000172.1	+	10128	10203	**RNA**→ EmrE (COG2076)→
Spr-1-1	NZ_ABJI01000368.1	+	23663	23724	**RNA**→ EmrE (COG2076)→
Ssp-1-1	NZ_ABJF01000294.1	-	7123	7042	**RNA**→ EmrE (COG2076)→
Asp-2-1	NC_011145.1	+	4934528	4934594	**RNA**→ EmrE (COG2076)→
Ade-1-1	NC_007760.1	+	4887717	4887781	**RNA**→ EmrE (COG2076)→ AraJ (COG2814)→ hypo→
Mgi-1-2	NC_009338.1	+	85042	85109	**RNA**→ SpeB (COG0010)→ COG0388 (COG0388)→ hypo→
Mva-1-1	NC_008726.1	-	851388	851320	**RNA**→ SpeB (COG0010)→ COG0388 (COG0388)→ hypo→
Ssp-2-1	NZ_ABJG01000230.1	+	7855	7920	**RNA**→ EmrE (COG2076)→
Sco-1-1	NC_003888.3	-	3153808	3153739	**RNA**→ EmrE (COG2076)→

Mxa-1-2	NC_008095.1	+	4374426	4374486	**RNA**→ EmrE (COG2076)→
Tfu-1-1	NC_007333.1	-	1465653	1465594	**RNA**→ EmrE (COG2076)→
Sce-1-1	NC_010162.1	-	8188359	8188289	**RNA**→ EmrE (COG2076)→ rDP_like (cd01301)→ UbiE (COG2226)→ FixC (COG0644)→ S_TKc (smart00220)→
Rsa-1-1	NC_010168.1	-	2096309	2096249	**RNA**→ EmrE (COG2076)→

104.4 Conserved domains

Conserved domains found in protein-coding genes listed in Section 104.3 are listed, with the first sentence in their description from the Conserved Domain Database. Conserved domains associated with more than one *ykkC*-III RNA are assigned a color, while others are shown in gray. The number in parentheses is the number of occurrences of the given gene.

cd00093 (1) Helix-turn-helix XRE-family like proteins.
cd01301 (1) renal dipeptidase (rDP), best studied in mammals and also called membrane or microsomal dipeptidase, is a membrane-bound glycoprotein hydrolyzing dipeptides and is involved in hydrolytic metabolism of penem and carbapenem beta-lactam antibiotics.
cd03139 (1) Type 1 glutamine amidotransferase (GATase1)-like domain found in a subgroup of proteins similar to PfpI from Pyrococcus furiosus.
cd03320 (1) o-Succinylbenzoate synthase (OSBS) catalyzes the conversion of 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate (SHCHC) to 4-(2'-carboxyphenyl)-4-oxobutyrate (o-succinylbenzoate or OSB), a reaction in the menaquinone biosynthetic pathway.
COG0010 (2) Arginase/agmatinase/formimionoglutamate hydrolase, arginase family [Amino acid transport and metabolism]
COG0204 (4) 1-acyl-sn-glycerol-3-phosphate acyltransferase [Lipid metabolism]
COG0388 (2) Predicted amidohydrolase [General function prediction only]
COG0644 (1) Dehydrogenases (flavoproteins) [Energy production and conversion]
COG1072 (1) Panthothenate kinase [Coenzyme metabolism]
COG2076 (41) Membrane transporters of cations and cationic drugs [Inorganic ion transport and metabolism]
COG2226 (1) Methylase involved in ubiquinone/menaquinone biosynthesis [Coenzyme metabolism]
COG2814 (1) Arabinose efflux permease [Carbohydrate transport and metabolism]
COG3550 (1) Uncharacterized protein related to capsule biosynthesis enzymes [General function prediction only]
COG4977 (1) Transcriptional regulator containing an amidase domain and an AraC-type DNA-binding HTH domain [Transcription]
GOS4424 (1) ?
pfam05787 (1) Bacterial protein of unknown function (DUF839).
smart00220 (1) Serine/Threonine protein kinases, catalytic domain; Phosphotransferases.

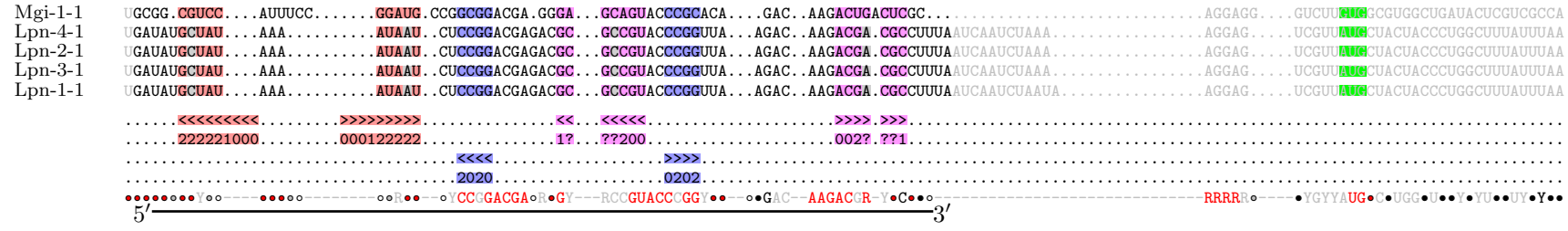
Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Msp-2-1

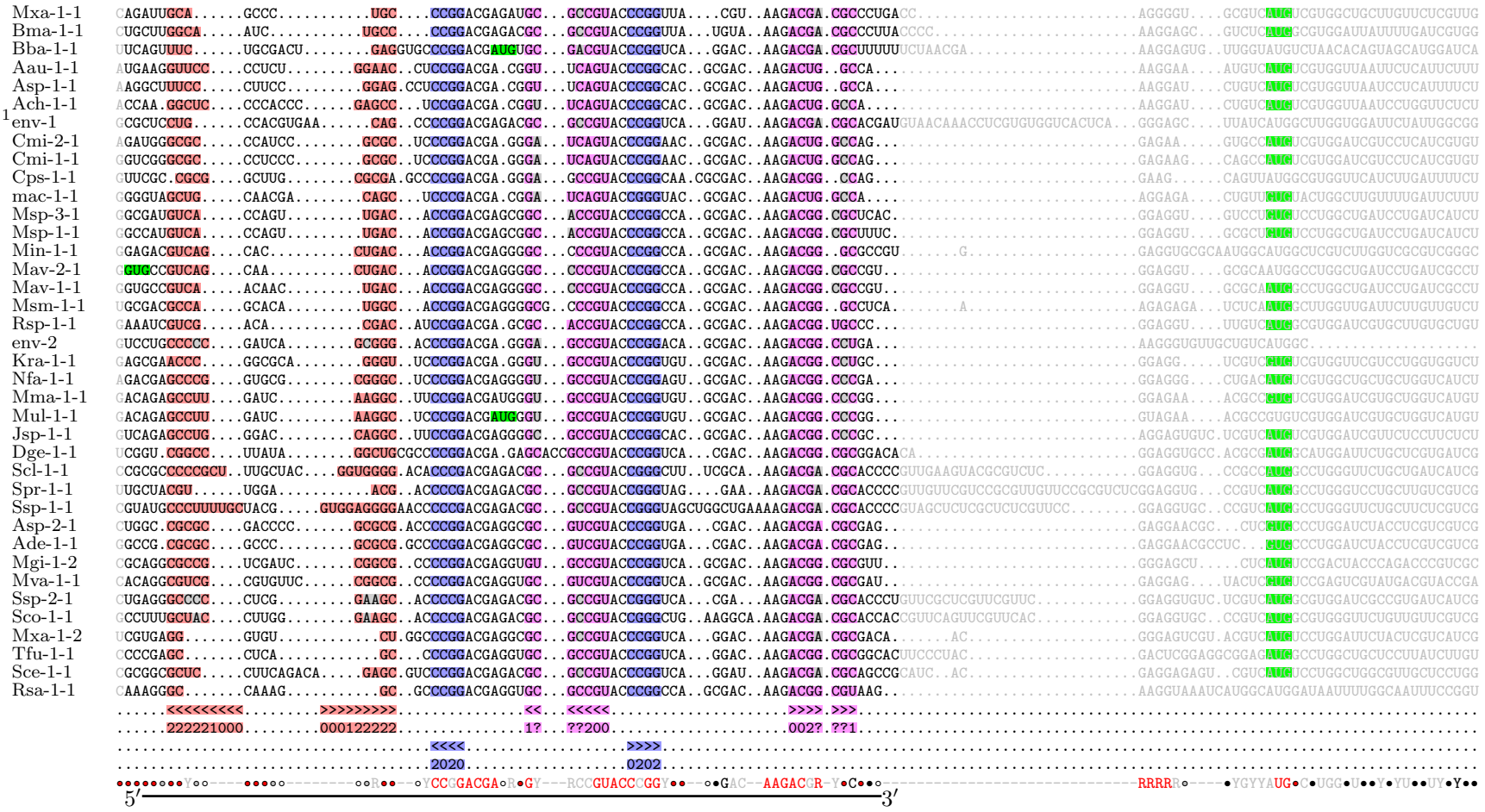
104.5 Multiple-sequence alignment

Each *ykkC*-III RNA is denoted by an abbreviation (like "Eco-1-1") that refers to a taxonomy given in Section 104.1. Superscript numbers refer to annotations listed in Section 104.2. The alignment includes sequences containing the RNA motif, as well as flanking sequence. The *ykkC*-III RNA itself is denoted by the line underneath marked 5′ and 3′ on either end. Nucleotides in the flanking sequences (i.e., not part of the motif) are written in gray letters. Stems of predicted rho-independent transcription terminators are shaded yellow. (Note: terminator predictions have not been analyzed manually, and many are likely to be false positives.) Start codons are shaded green. (Note: start codons are frequently misannotated, especially in environmental samples.) Nucleotides proposed to basepair as part of the consensus structure are shaded in color when they comprise Watson-Crick or G-U pairs. Otherwise they are shaded gray. Conserved stems are also indicated at the bottom of the alignment by angle brackets, where matching < and > denote base-paired columns. Below these angle brackets, the symbol "2" denotes base pairs exhibiting covariation, "1" denotes base pairs exhibiting compatible mutations, "0" denotes base pairs that are not observed to mutate and "?" denotes base pairs that have a significant frequency of non-canonical nucleotides for Watson-Crick or G-U pairs (> 5%). Below these base pair annotation is the consensus sequence: "R" = "A" or "G", "Y" = "C" or "U", red nucleotides: nucleotide identity conserved more than 97% of the time, black nucleotides: 90%, gray nucleotides: 75%, red circle (•): nucleotide is present 97% of the time, black circle (•): 90%, gray circle (•): 75%, white circle (•): 50%. All percentages of sequences just described (e.g. 97% conserved) assume that sequences have been weighted by the GSC algorithm implemented by the Infernal software.

alignment positions 1 · · · 99

alignment positions 100 · · · 259




Mxa-1-1
Bma-1-1
Bba-1-1
Aau-1-1
Asp-1-1
Ach-1-1
1env-1
Cmi-2-1
Cmi-1-1
Cps-1-1
mac-1-1
Msp-3-1
Msp-1-1
Min-1-1
Mav-2-1
Mav-1-1
Msm-1-1
Rsp-1-1
env-2
Kra-1-1
Nfa-1-1
Mma-1-1
Mul-1-1
Jsp-1-1
Dge-1-1
Scl-1-1
Spr-1-1
Ssp-1-1
Asp-2-1
Ade-1-1
Mgi-1-2
Mva-1-1
Ssp-2-1
Sco-1-1
Mxa-1-2
Tfu-1-1
Sce-1-1
Rsa-1-1
5′
3′
alignment positions 260 · · · 419
Mgi-1-1
Lpn-4-1
Lpn-2-1
Lpn-3-1
Lpn-1-1
Mxa-1-1
Bma-1-1
Bba-1-1
Aau-1-1
Asp-1-1
Ach-1-1
1env-1

Cmi-2-1
Cmi-1-1
Cps-1-1
mac-1-1
Msp-3-1
Msp-1-1
Min-1-1
Mav-2-1
Mav-1-1
Msm-1-1
Rsp-1-1
env-2
Kra-1-1
Nfa-1-1
Mma-1-1
Mul-1-1
Jsp-1-1
Dge-1-1
Scl-1-1
Spr-1-1
Ssp-1-1
Asp-2-1
Ade-1-1
Mgi-1-2
Mva-1-1
Ssp-2-1
Sco-1-1
Mxa-1-2
Tfu-1-1
Sce-1-1
Rsa-1-1
alignment positions 420 · · · 528
Mgi-1-1
Lpn-4-1
Lpn-2-1
Lpn-3-1
Lpn-1-1
Mxa-1-1
Bma-1-1
Bba-1-1
Aau-1-1
Asp-1-1
Ach-1-1
1 env-1
Cmi-2-1
Cmi-1-1
Cps-1-1
mac-1-1
Msp-3-1
Msp-1-1
Min-1-1
Mav-2-1

References

References are listed in the manuscript.