

Supplementary figures for:

Analysis of metagenome-assembled viral genomes from the human gut reveals diverse putative CrAss-like phages with unique genomic features

Natalya Yutin¹, Sean Benler¹, Sergei A. Shmakov¹, Yuri I. Wolf¹, Igor Tolstoy¹, Mike Rayko², Dmitry Antipov², Pavel A. Pevzner³, Eugene V. Koonin^{1*}

¹, National Center for Biotechnology Information, National Library of Medicine, Bethesda, Maryland 20894, USA;

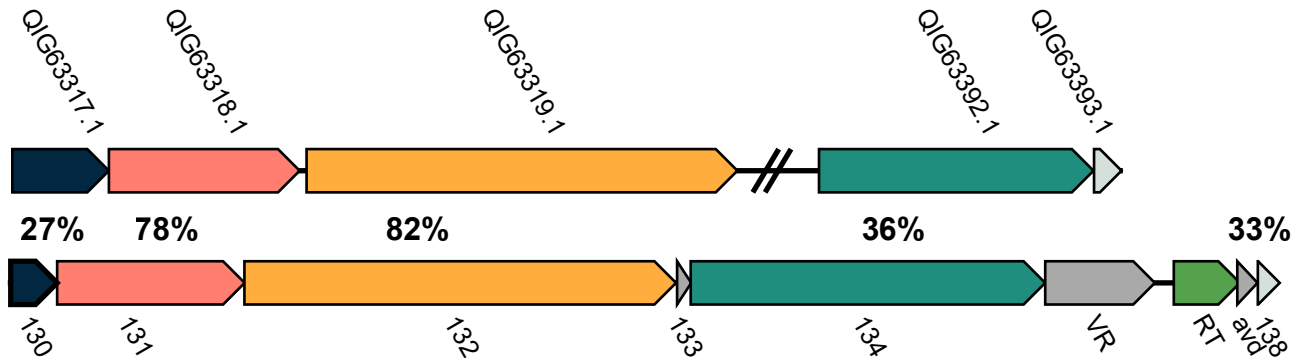
², Center for Algorithmic Biotechnology, Institute for Translational Biomedicine, St. Petersburg State University, 199004 St. Petersburg, Russia ;

³, Department of Computer Science and Engineering, University of California-San Diego, La Jolla, CA 92093, USA.

*Correspondence : koonin@ncbi.nlm.nih.gov

Supplementary Figure 1

Bacteroides phage DAC16 (MT074137.1)



Zeta family crAss-like phage (OLOE01000004.1)

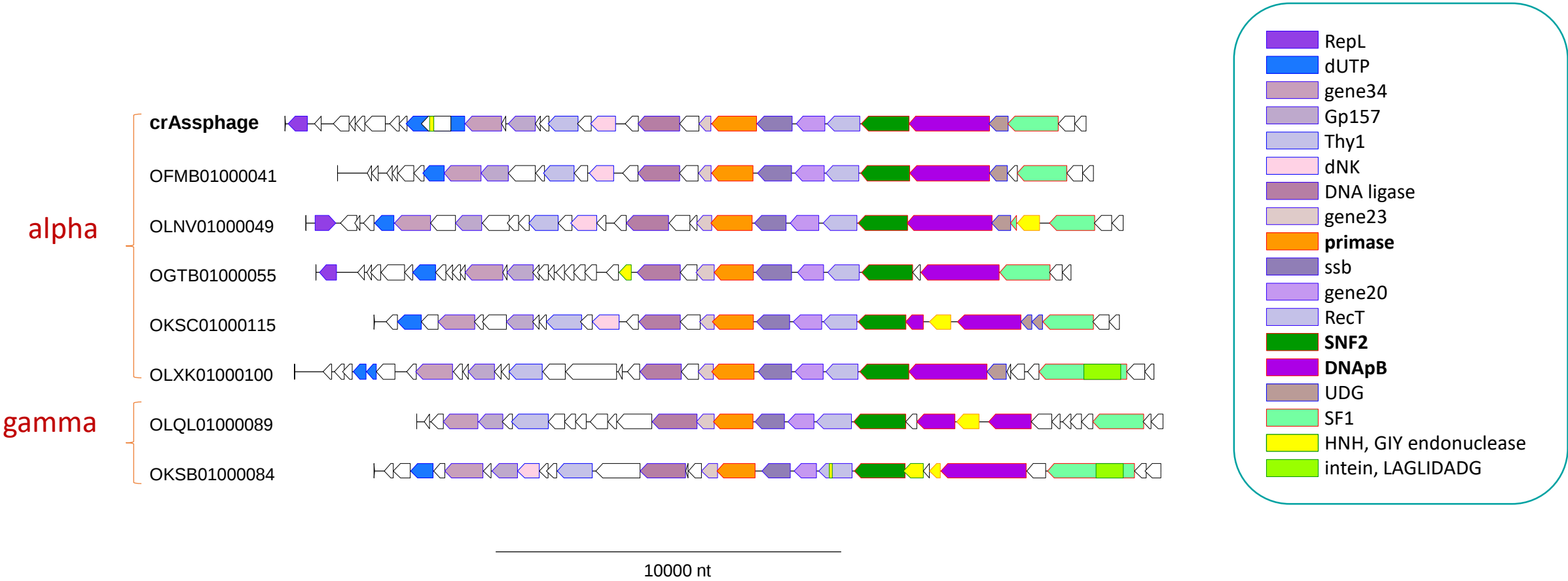
1000 nt

A syntenic block of homologous genes between a Zeta family crAss-like phage and *Bacteroides* phage DAC16. Homologous ORFs are color-coded and the amino acid sequence similarity is indicated in the middle.

Supplementary Figure 2

A

alpha/gamma group



beta group

B

IAS_virus_KJ003983

OLRZ01000056

OJNZ01000055

OLSD01000074

OCLL01000003

OFRY01000050

crAss001_MH675552

PPYF01195288

OBAM01000118

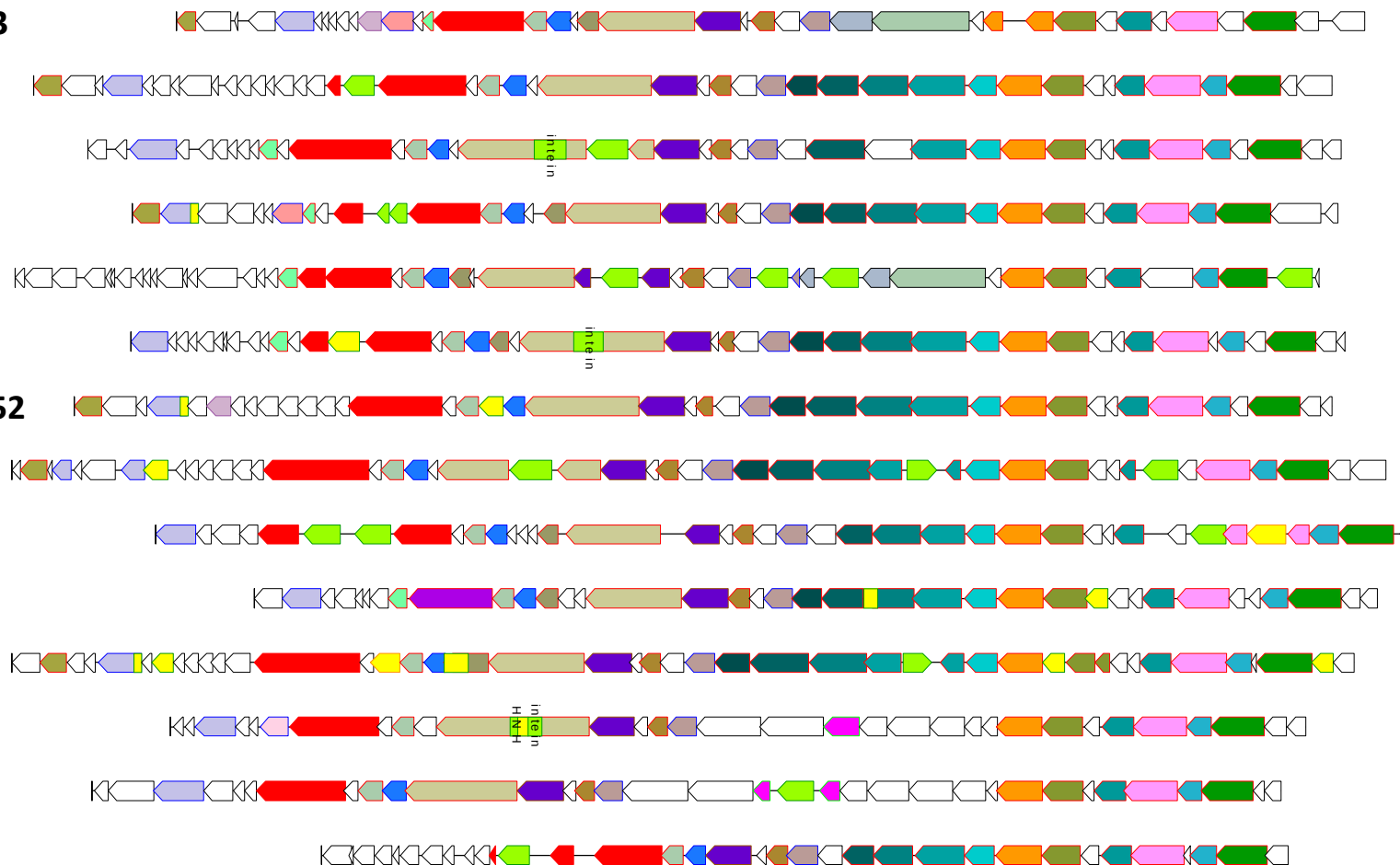
OHUS01000015

OJOM01000034

OGZO01000002

OJQL01000090

OKXB01000124

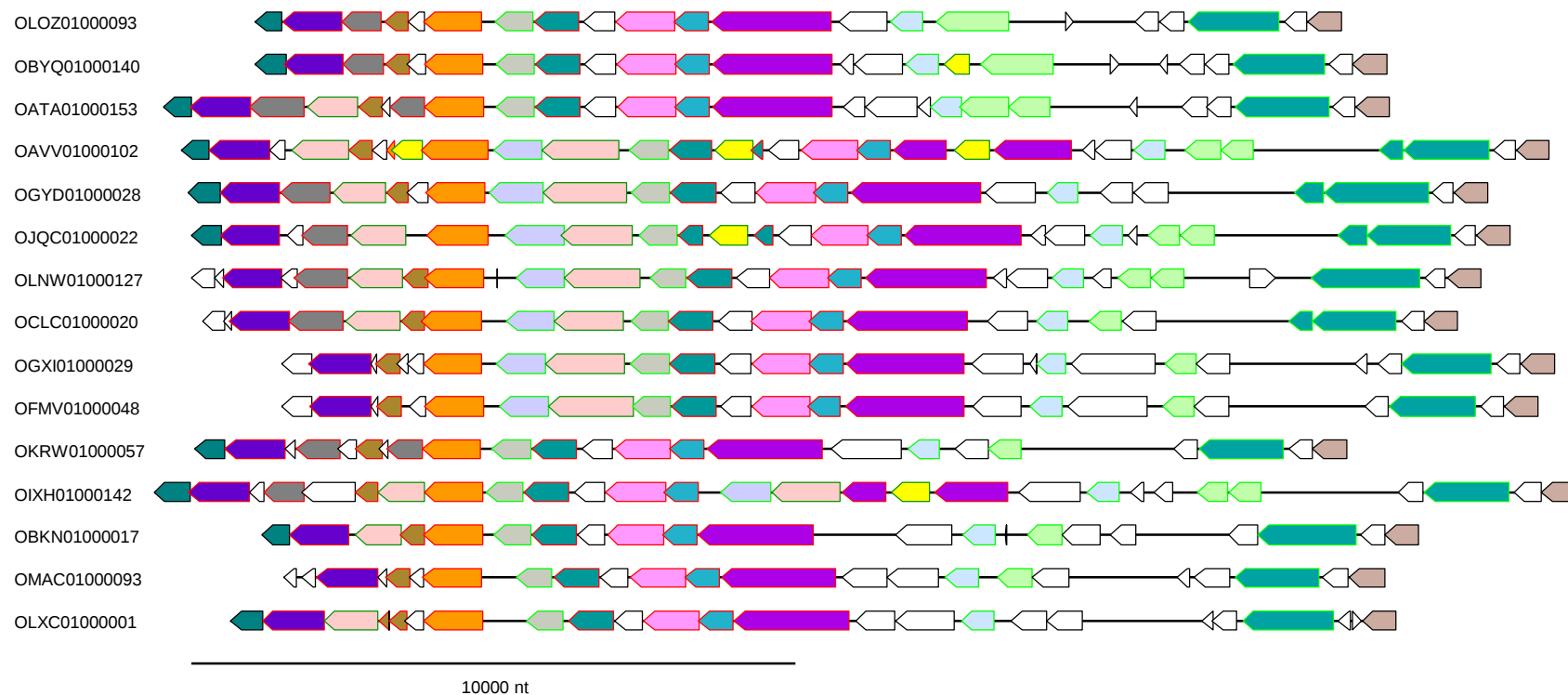


- gene51_IAS
- PoiA**
- aspartate protease (gene66_IAS)
- dUTPase
- NrdG, ribonucleotide reductase activating protein
- RNR, ribonucleotide reductase
- PDDEXK nuclease**
- phage endonuclease 48b
- UDG
- uncharacterized_ AXQ62718
- ThiF family
- MPN, metalloprotease
- uncharacterized_ AXQ62721
- uncharacterized_ AXQ62722
- primase**
- 45b
- ATP_43b**
- DnaB**
- Rep_Org**
- SNF2**

10000 nt

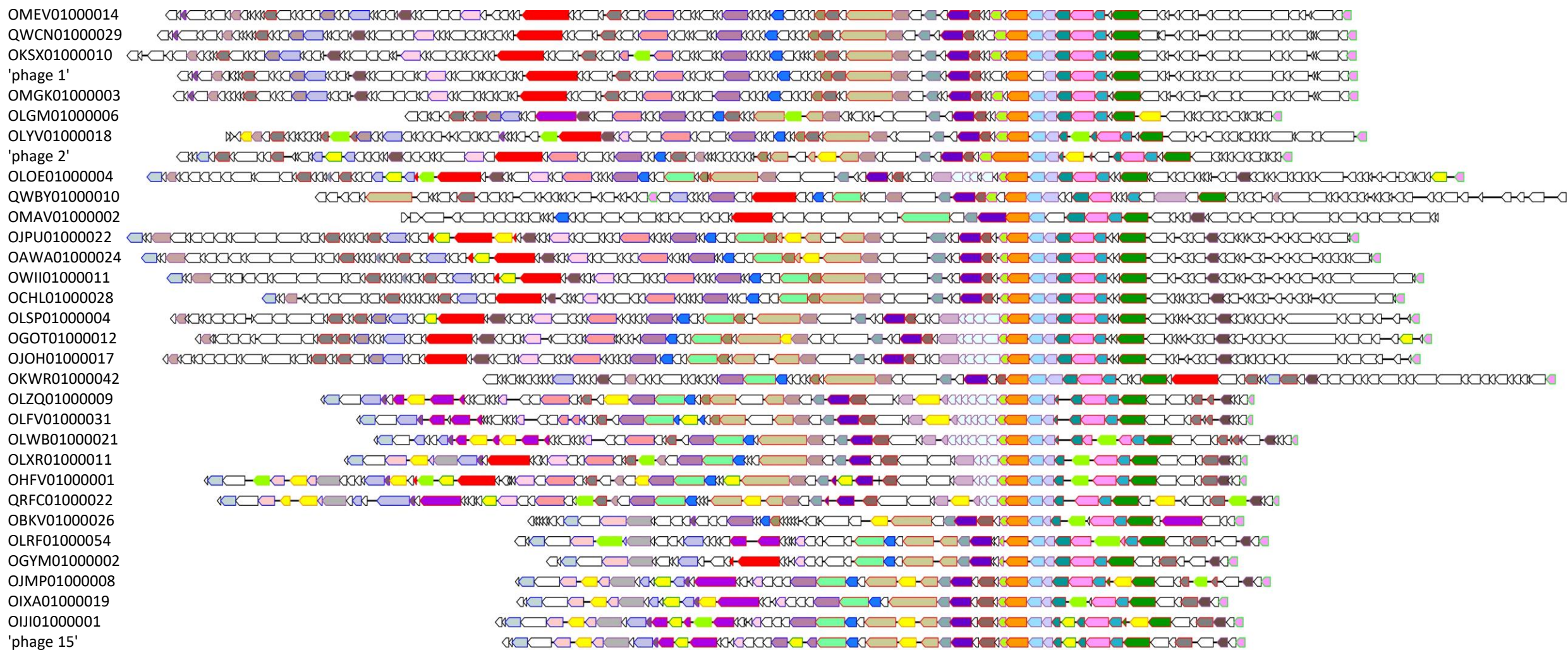
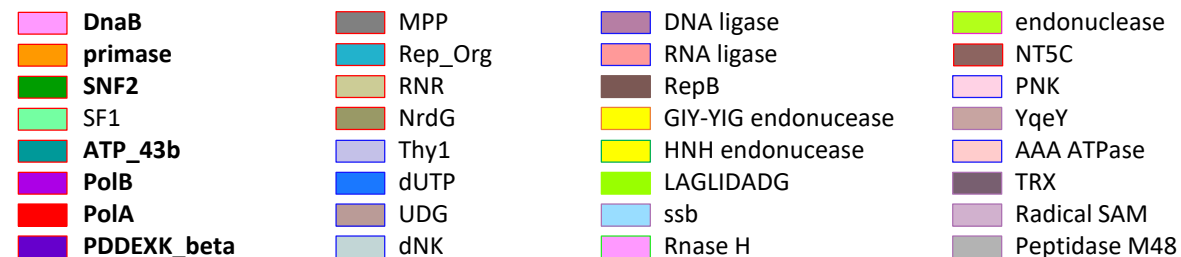
C

delta group



D

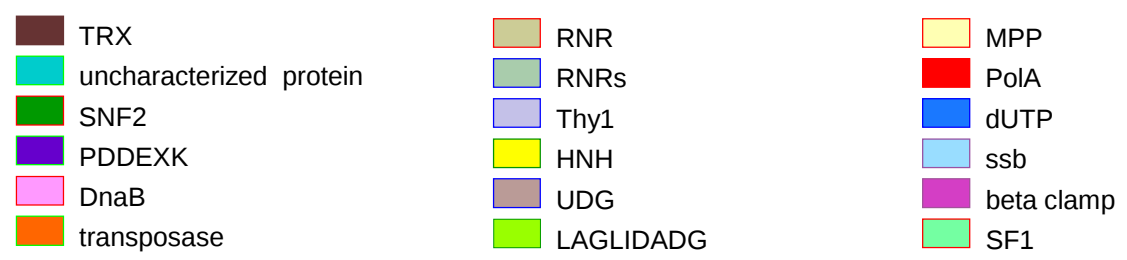
zeta group



10000 nt

E

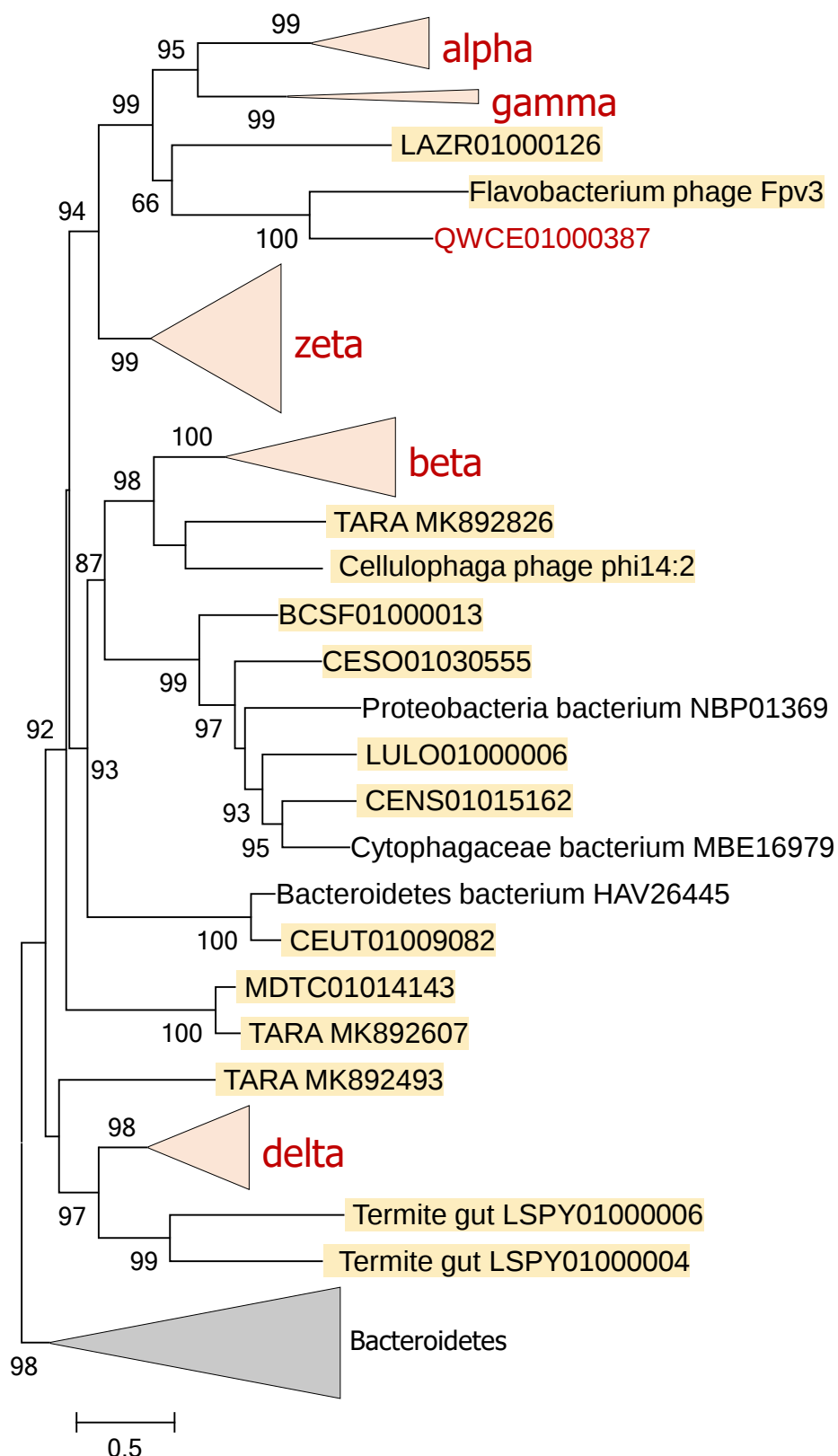
epsilon group



10000 nt

Group-specific genome maps of the replication gene block of the crAss-like phages. A, Alpha and Gamma groups; B, Beta group; C, Delta group; D, Zeta group; E, Epsilon group

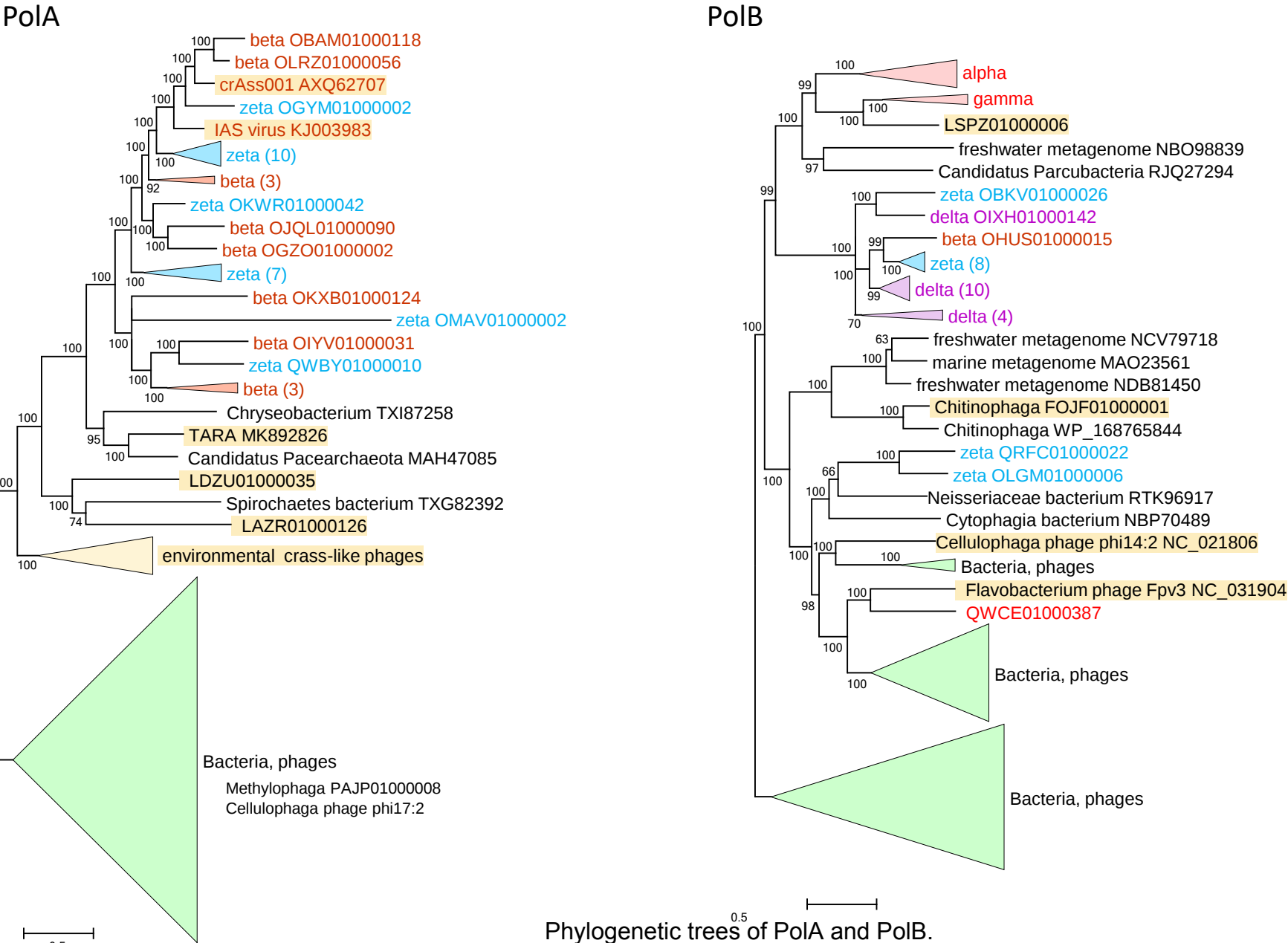
Supplementary Figure 3



Phylogenetic tree of the DnaG family primase.

Previously analyzed crAss-like sequences are highlighted in yellow.

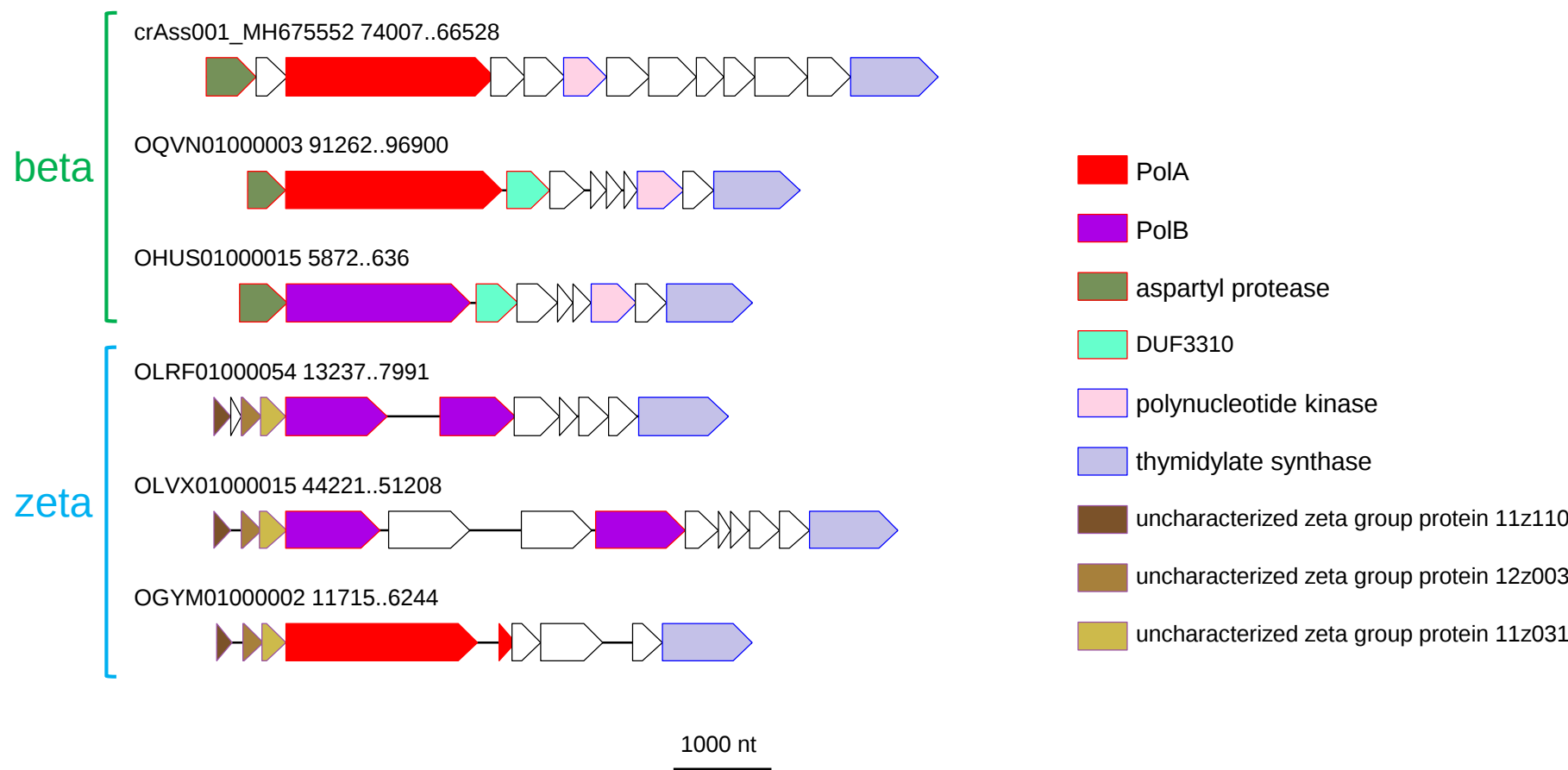
Supplementary Figure 4



Supplementary Figure 5

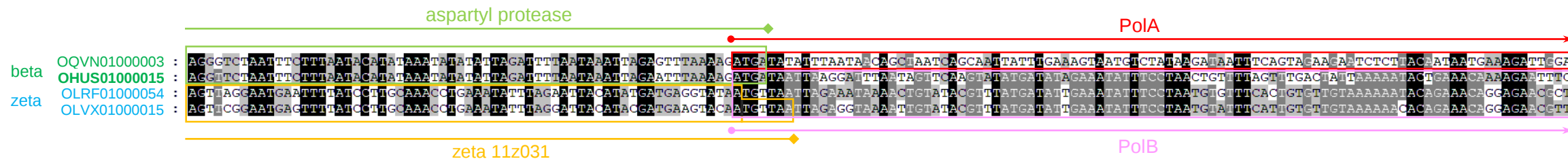
Two examples of in situ DNAP replacement in crAss-like phages.
A, Genomic organization around the DNAP genes
B, Nucleotide sequence alignments of the 5'-terminal regions of
DNAP genes and the upstream regions

A

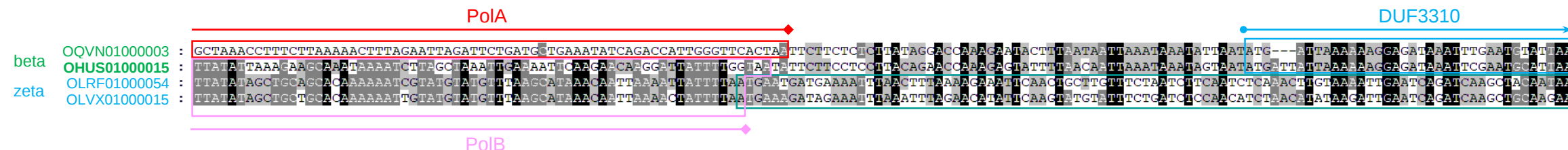


B

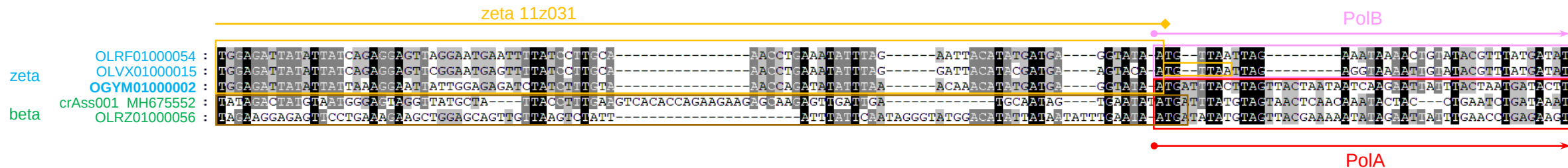
acquired PolB in OHUS01000015 (start)



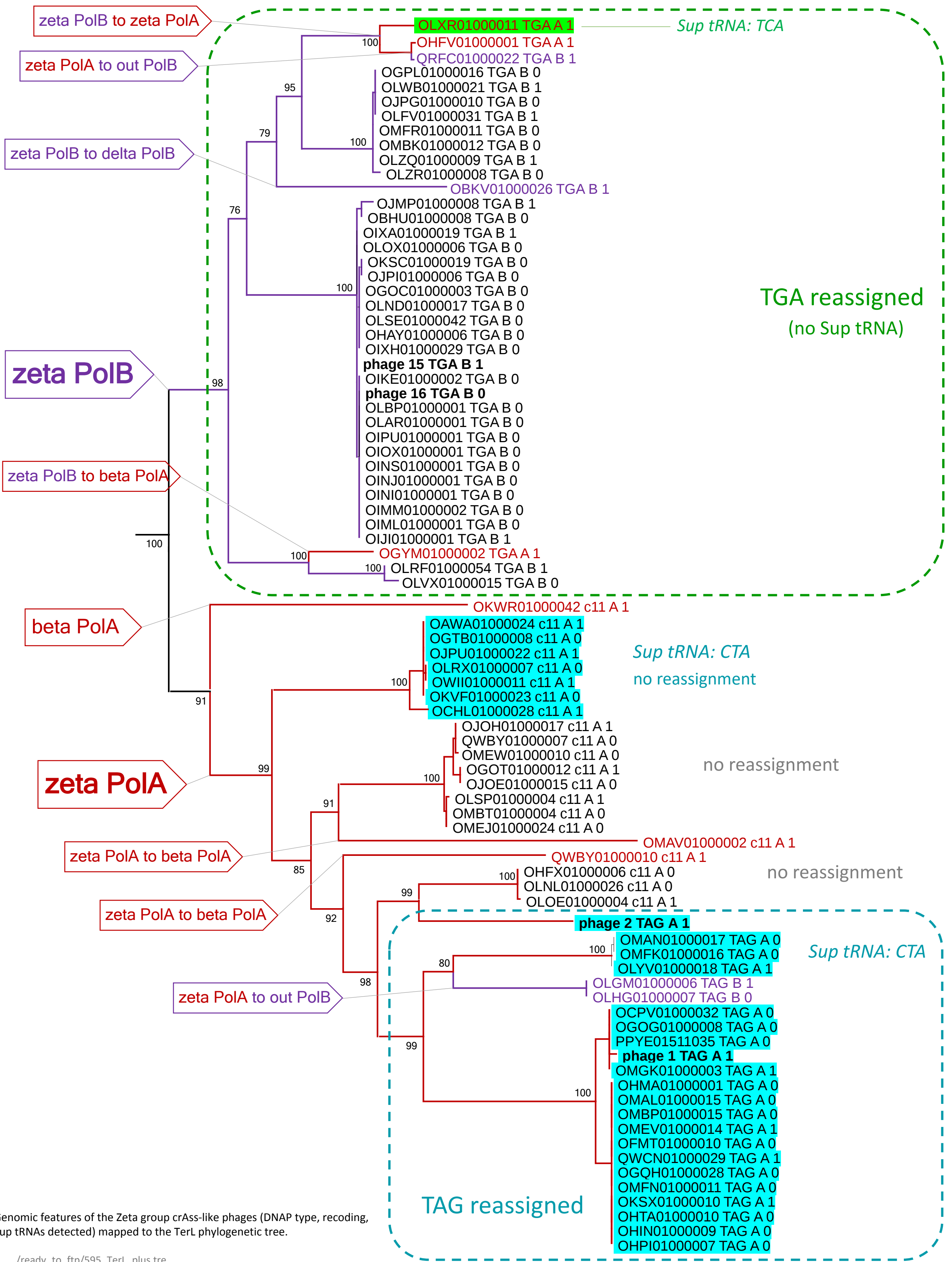
acquired PolB in OHUS01000015 (end)



acquired PolA in OGYM01000002 (start)

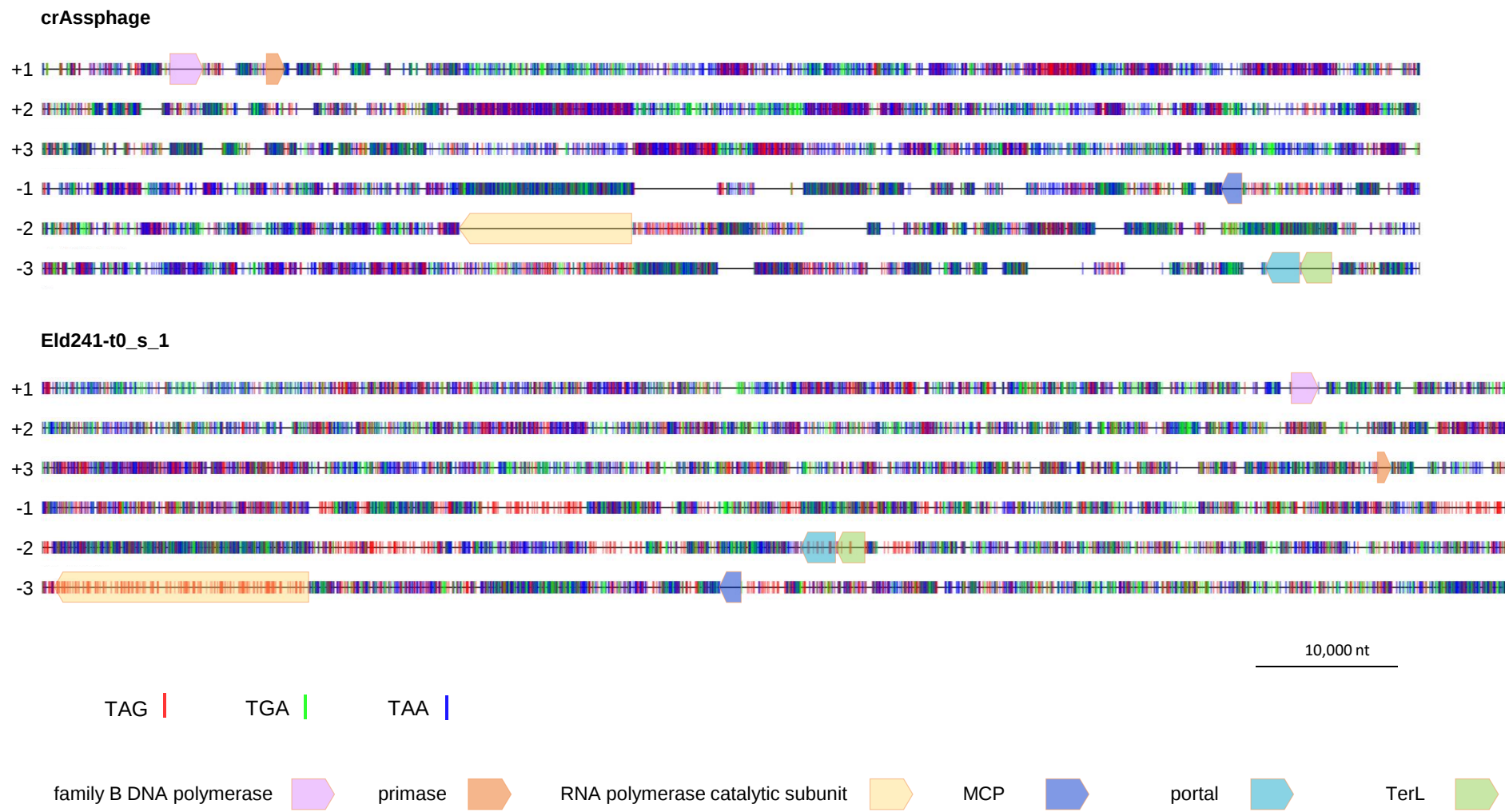


Supplementary Figure 6



Genomic features of the Zeta group crAss-like phages (DNAP type, recoding, Sup tRNAs detected) mapped to the TerL phylogenetic tree.

Supplementary Figure 7



TAG, TGA, and TAA codons in the genomes of crAssphage and contig Eld241-t0_s_1 of Ref. 13

Supplementary Figure 8

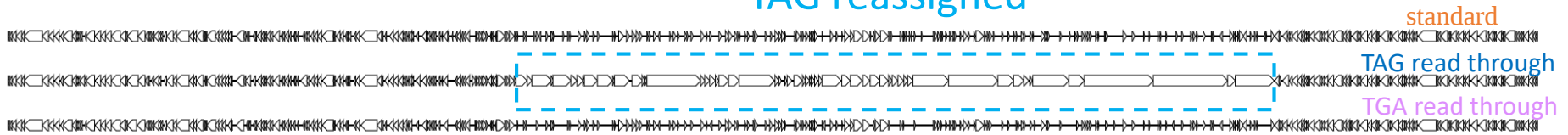
A

no reassignment



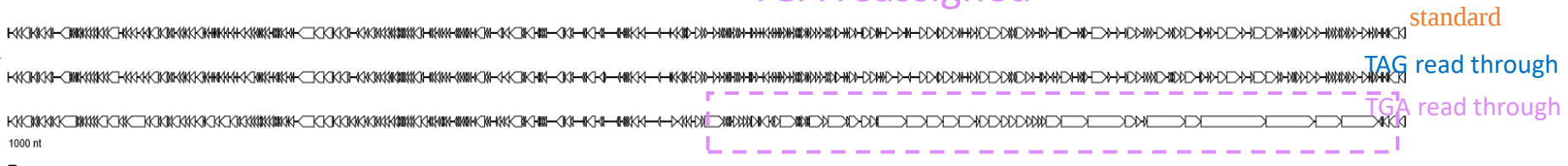
B

TAG reassigned



C

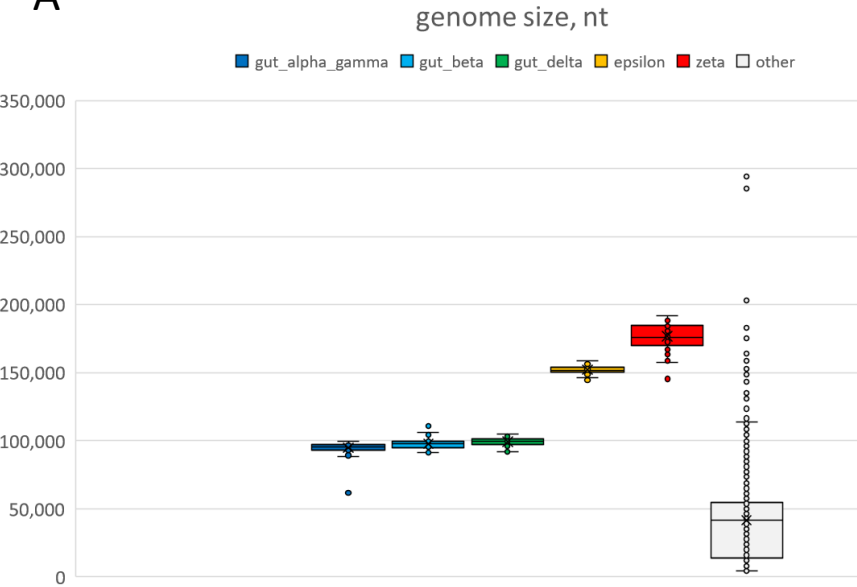
TGA reassigned



Three zeta group phage genomes, OLOE01000004 (A), OKSX01000010 (B), and OIXA01000019 (C), translated with standard bacterial code (top row), standard bacterial code with TAG read-trough (middle row), and standard bacterial code with TGA read-trough (bottom row). Accounting for read-trough TAG and TGA results in longer ORFs and higher coding density in parts of the OKSX01000010 and OIXA01000019 genomes.

Supplementary Figure 9

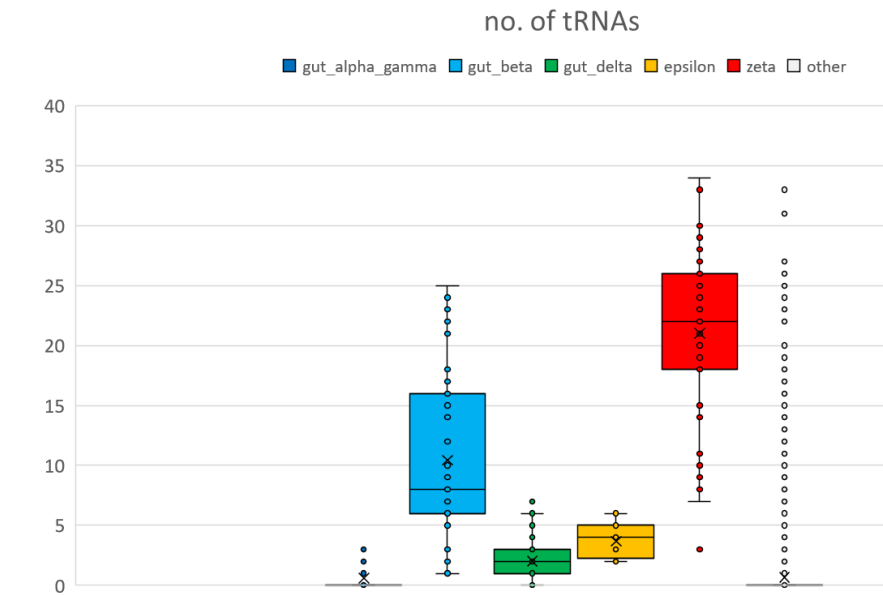
A



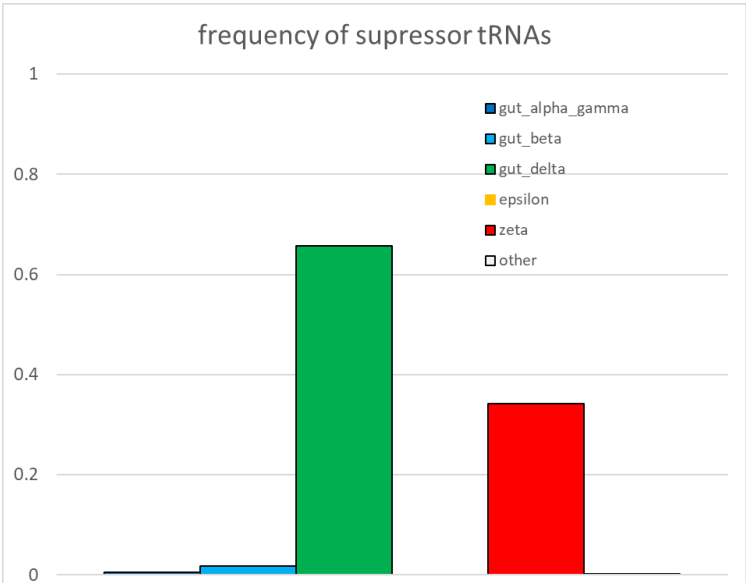
B

	# of contigs	# of tRNA	# of Sup tRNA	anticodon	reassigned codons
alpha gamma	190	115	1	TTA	TAA
beta	57	595	1	CTA	TAG
delta	233	472	153	CTA	TAG
epsilon	36	132	0		
zeta	79	1662	27	26 CTA, 1 TCA	26 TAG, 1 TCA
Flavob_phages	1	10	0		
other cMAGs	3343	2203	9	1 TCA, 2 CTA, 6 TTA	1 TCA, 2 CTA, 6 TTA

C



D



Supplementary figure 9

Genome size, number of tRNAs and suppressor tRNAs in crass-like cMAGs compared to other gut phage cMAGs (4-300 kb).

A. Distribution of genome sizes of alpha_gamma (n=190), beta (n=57), delta (n=233), epsilon (n=36), zeta (n=79) crass-like phages and other gut cMAGs (n=3344). Box boundaries correspond to the interquartile range (median indicated); whiskers indicate the lowest and the highest values within the range of 1.5 times the interquartile distance from the median; all points are shown.

B. Genome features of gut cMAGs.

C. Distribution of the number of tRNAs in the genomes of alpha_gamma (n=190), beta (n=57), delta (n=233), epsilon (n=36), zeta (n=79) crass-like phages and other gut cMAGs (n=3344). Box boundaries correspond to the interquartile range (median indicated); whiskers indicate the lowest and the highest values within the range of 1.5 times the interquartile distance from the median; all points are shown.

D. Frequency of suppressor tRNA occurrence in gut cMAGs.

A

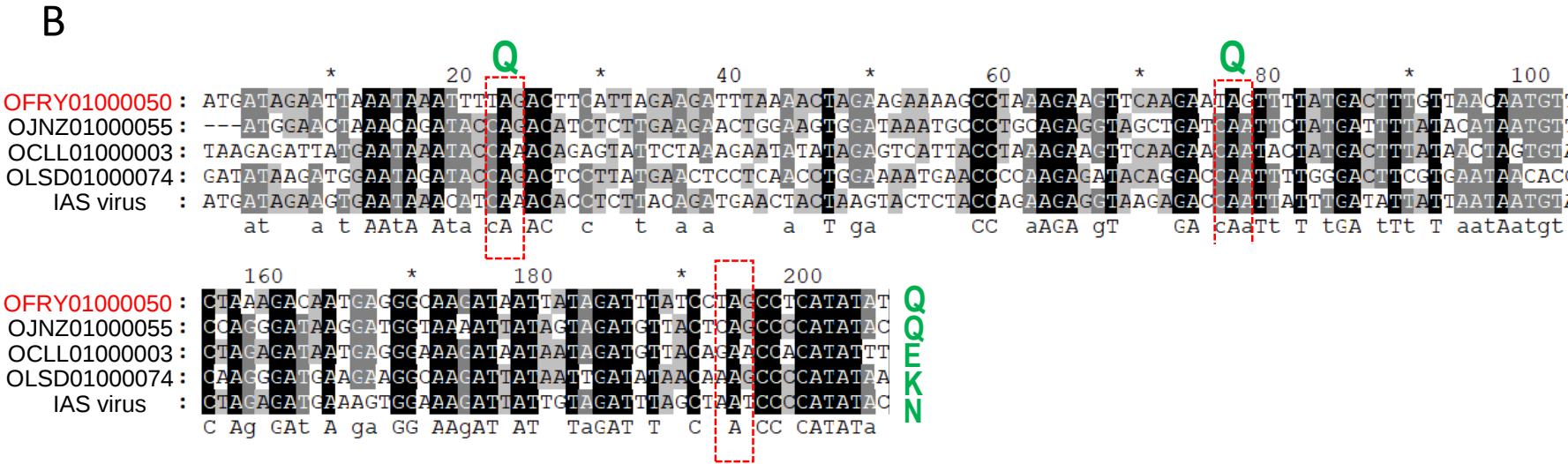
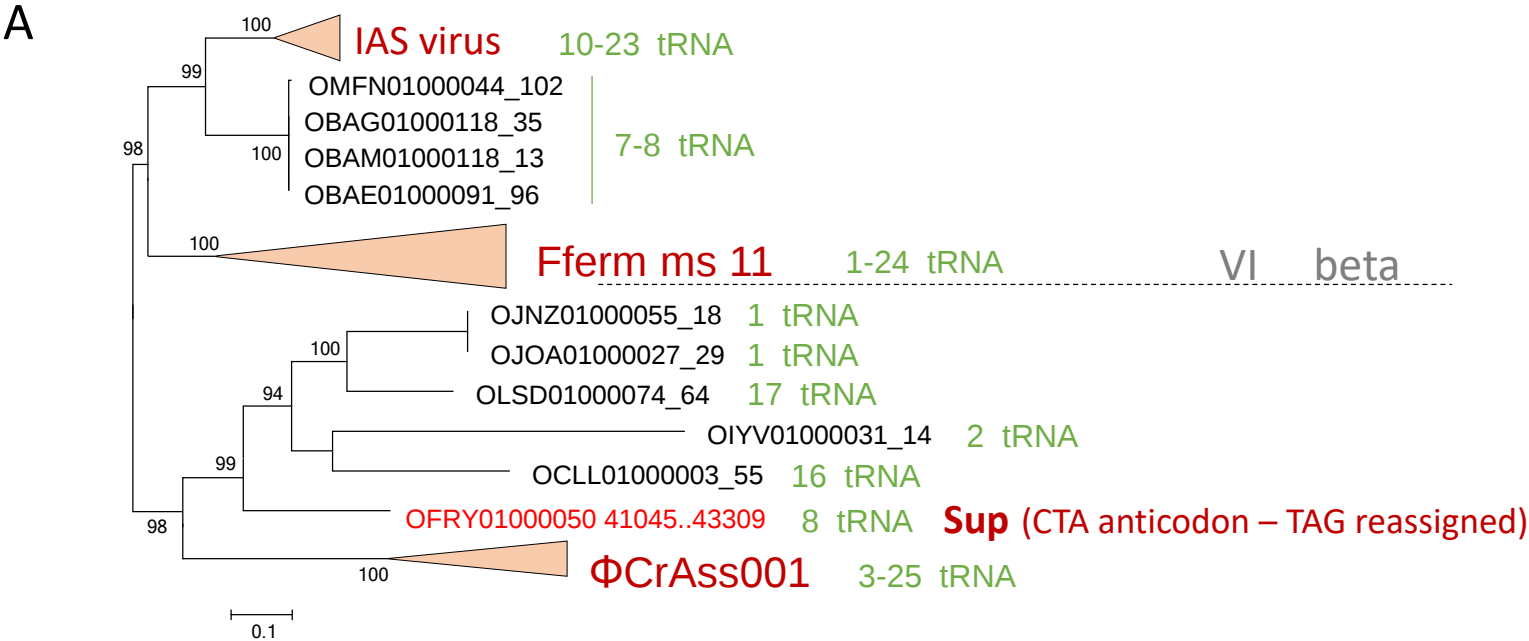
A, TerL phylogenetic tree of alpha/gamma group phages.

B

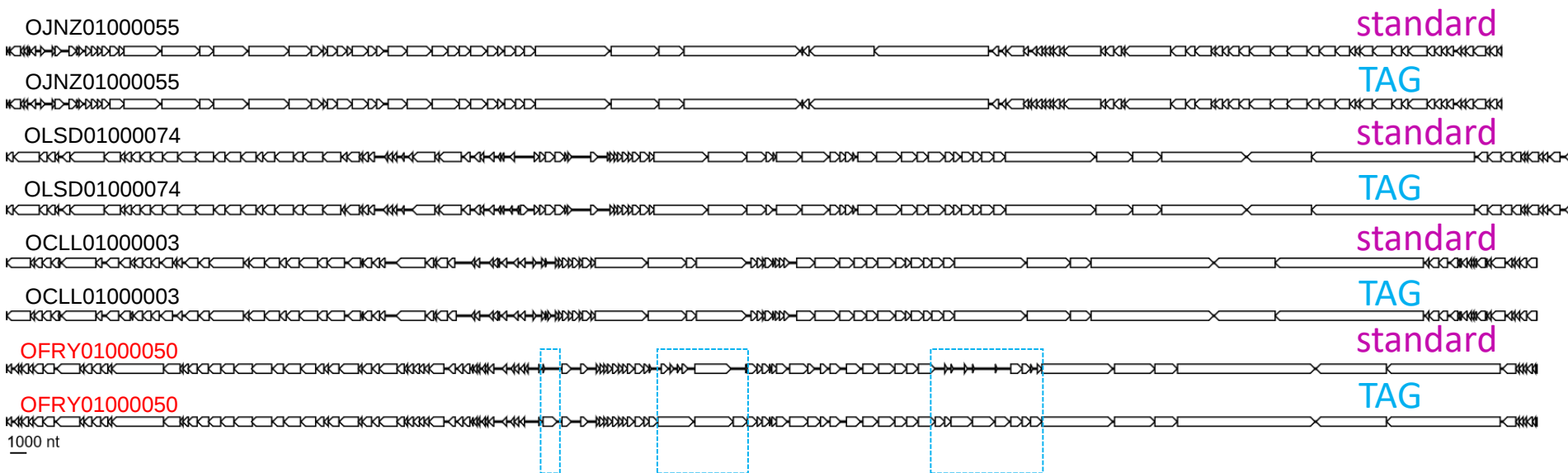
OGRL01000078 TGTTCATGGTGAATGGTAACACGTCAGATTGTTCTGAAATTCCAAGTTCGAACTTGGTAGAATAA
 OGRL01000073 TGTTCATGGTGAATGGTAACACGTCAGATTGTTCTGAAATTCCAAGTTCGAACTTGGTAGAATAA
 OJNL01000044 TGTTCATGGTGAATGGTAACACGTCAGATTGTTCTGAAATTCCAAGTTCGAACTTGGTAGAATAA
 OGQI01000045 TGTTCATGGTGAATGGTAACACGTCAGATTGTTCTGAAATTCCAAGTTCGAACTTGGTAGAATAA
 OLPH01000160 TGTTCATGGTGAATGGTAACACGTCAGATTGTTCTGAAATTCCAAGTTCGAACTTGGTAGAATAA
OKSO01000115 TGTTCATGGTGAATGGTAACACGTCAGATTGTTCTGAAATTCCAAGTTCGAACTTGGTAGAATAA



Supplementary Figure 11



C



Suppressor tRNA in the Beta group of crAss-like phages.

A, TerL phylogenetic tree of Beta group phages.

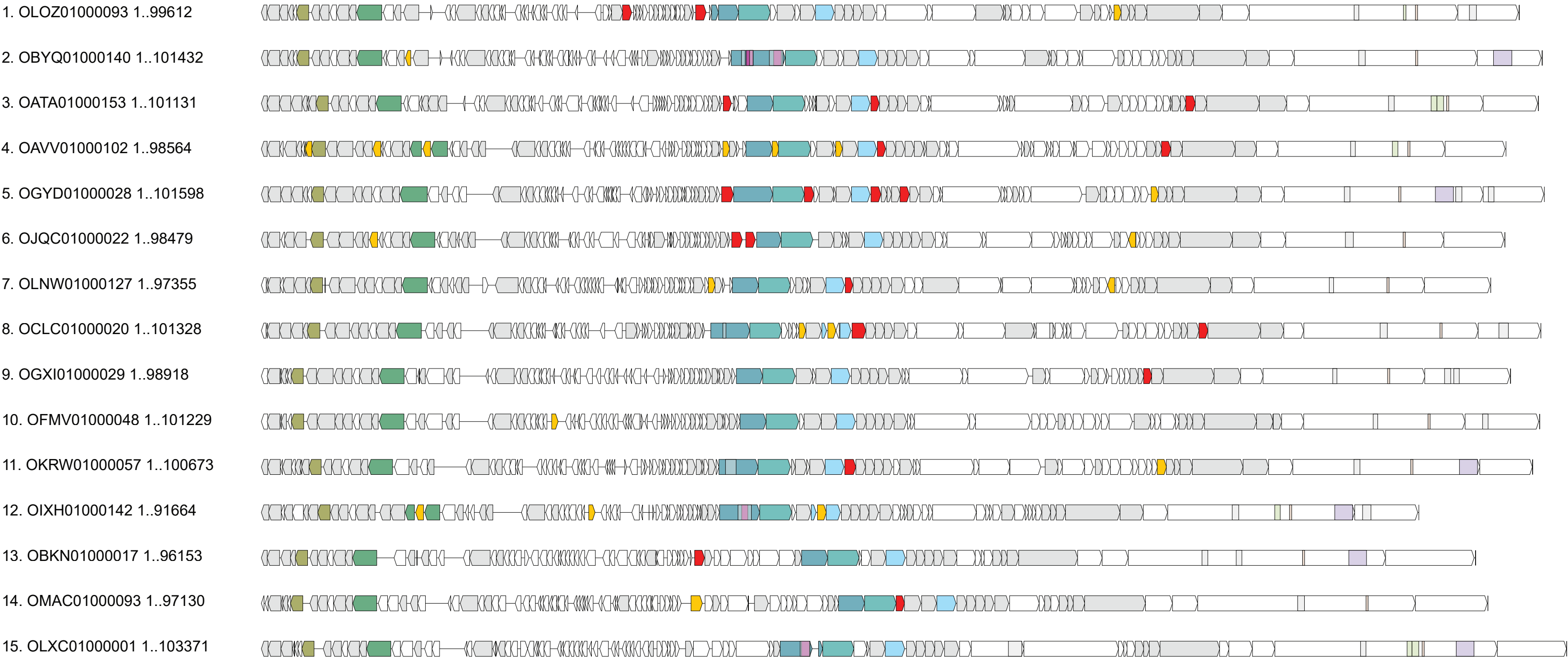
B, A fragment of TerL gene alignment; the first three TAGs reassigned to Q in OFRY01000050 genome are shown; the whole TerL gene of OFRY01000050 has 15 reassigned TAG codons.

C, Emergence of alternative coding in OFRY01000050 genome. Four closely related genomes were translated with standard bacterial code and standard bacterial code with TAG read-trough. Regions of alternative coding in OFRY01000050 are marked with dashed boxes.

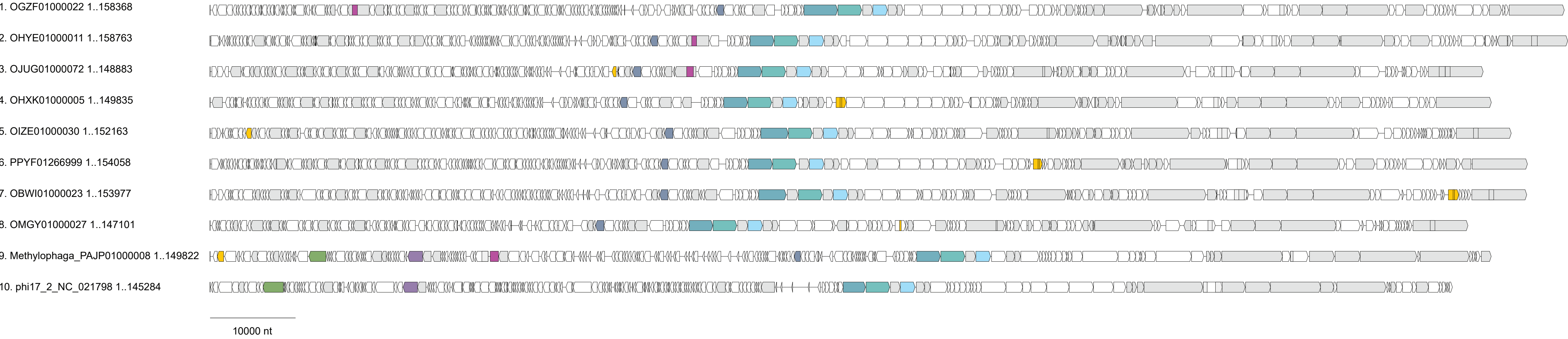
Supplementary Figure 12

Introns and inteins in crass-like phages. Whole genome maps for representative alpha/beta, delta, zeta, and epsilon phages are shown; whole-genome annotations of these genomes are here: [ftp://ftp.ncbi.nih.gov/pub/yutinn/crassfamily_2020/](http://ftp.ncbi.nih.gov/pub/yutinn/crassfamily_2020/)





10000 nt





Supplementary Figure 13

The transcription gene block in the Zeta group crAss-like phages. Conserved domains were identified with translating blast searches against a custom set of profiles and superimposed on genome map constructed with Prodigal.

