New deep-branching environmental plastid genomes on the algal tree of life

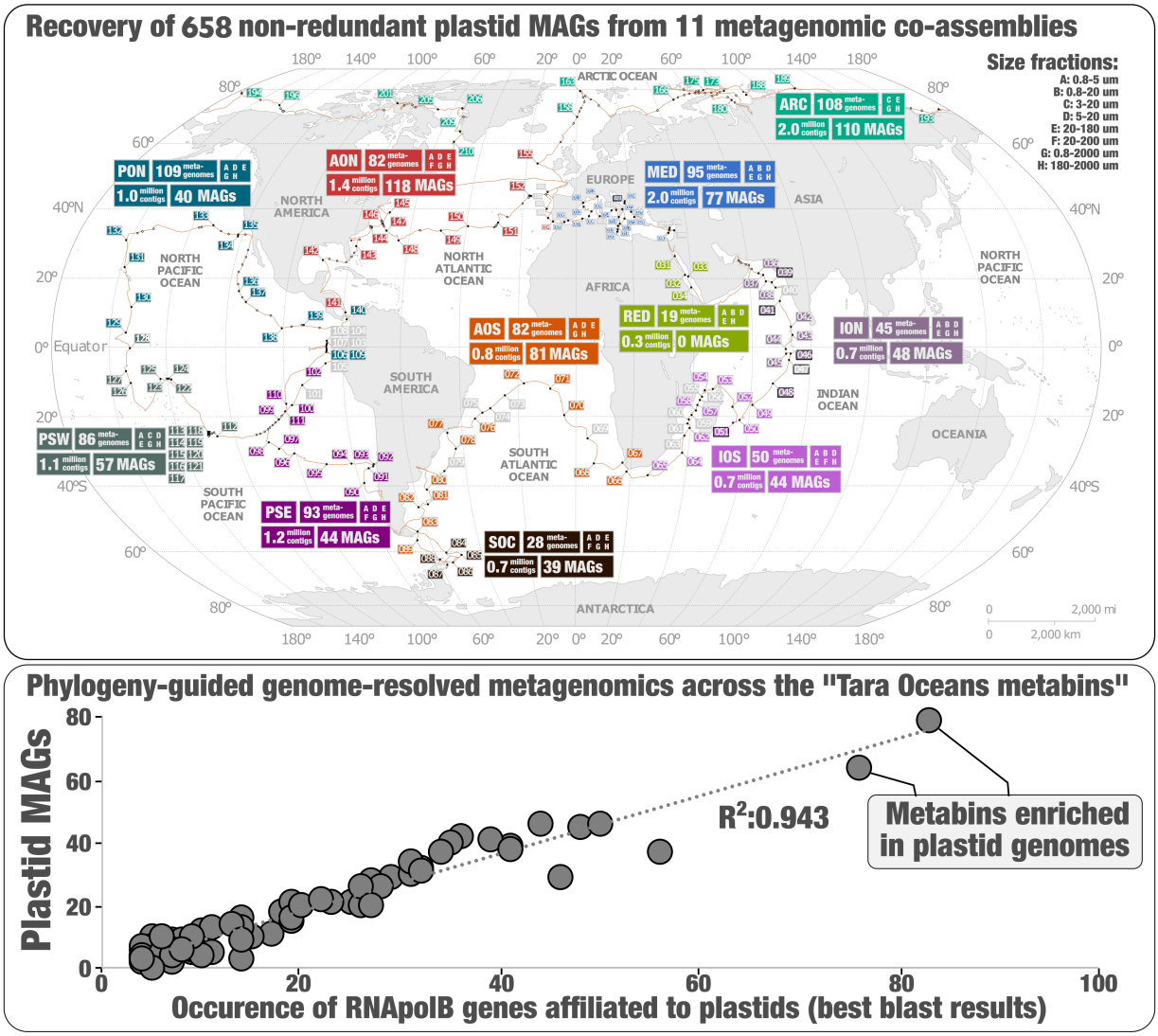
Supplementary Information

Mahwash Jamy[1], Thomas Huber[2], Thibault Antoine[3], Hans-Joachim Ruscheweyh[4], Lucas Paoli[5], Eric Pelletier[3], Tom O. Delmont[3]*, Fabien Burki[2]*

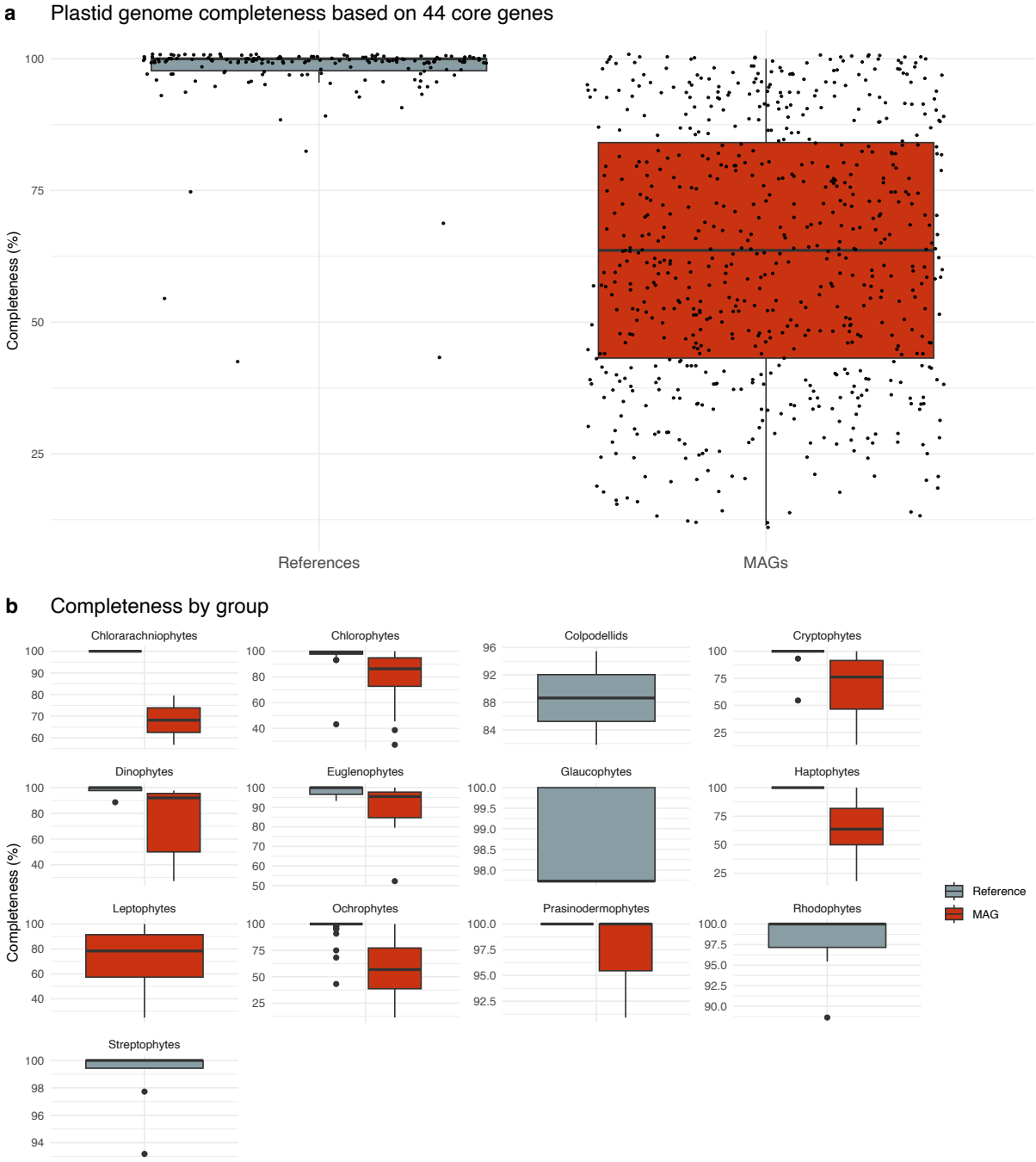
[1] Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala, Sweden
[2] Department of Organismal Biology (Systematic Biology), Uppsala University, Uppsala, Sweden
[3] Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France
[4] Department of Biology, Institute of Microbiology and Swiss Institute of Bioinformatics, ETH Zürich, 8093, Zürich, Switzerland
[5] Global Health Institute, School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

*co-corresponding authors

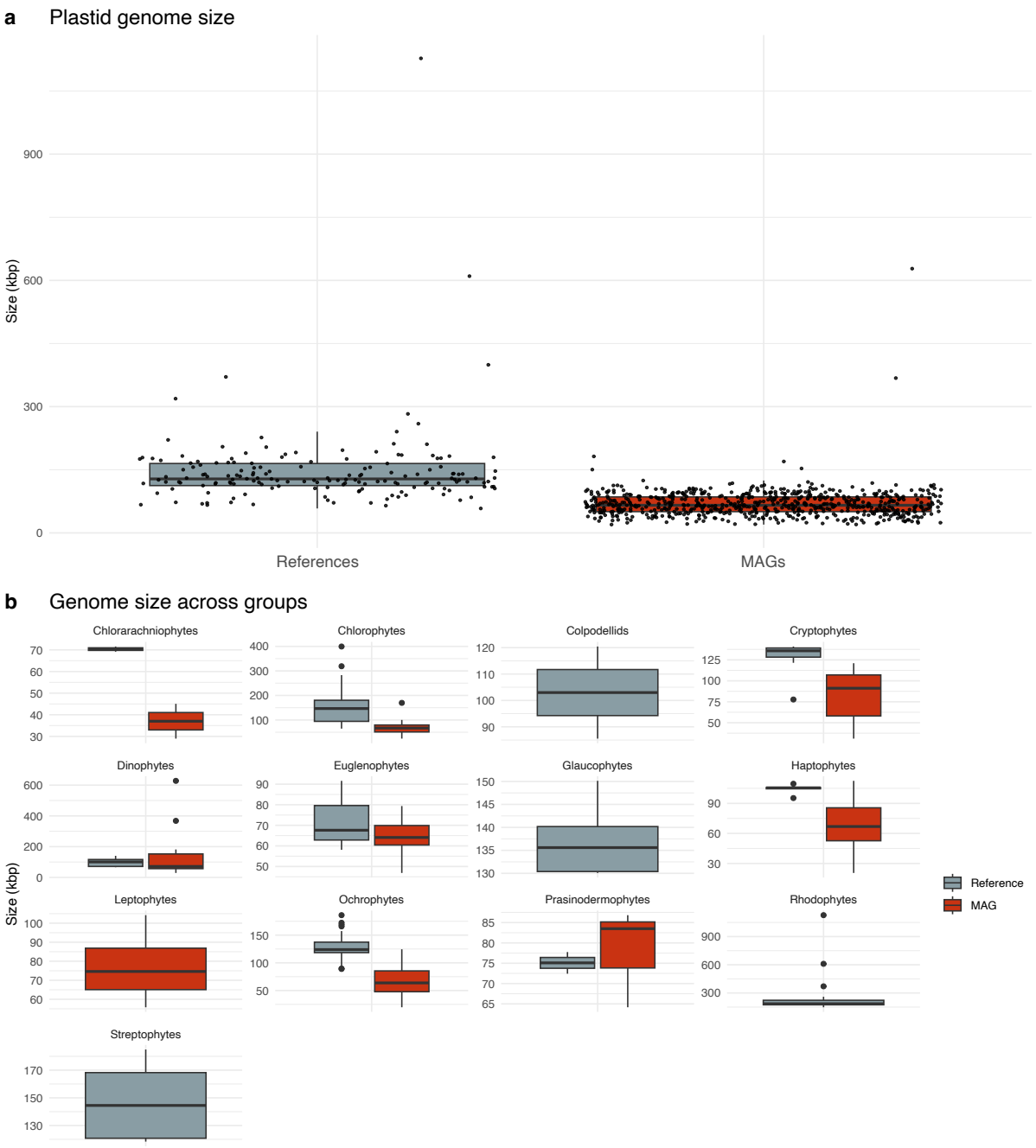
This pdf file contains



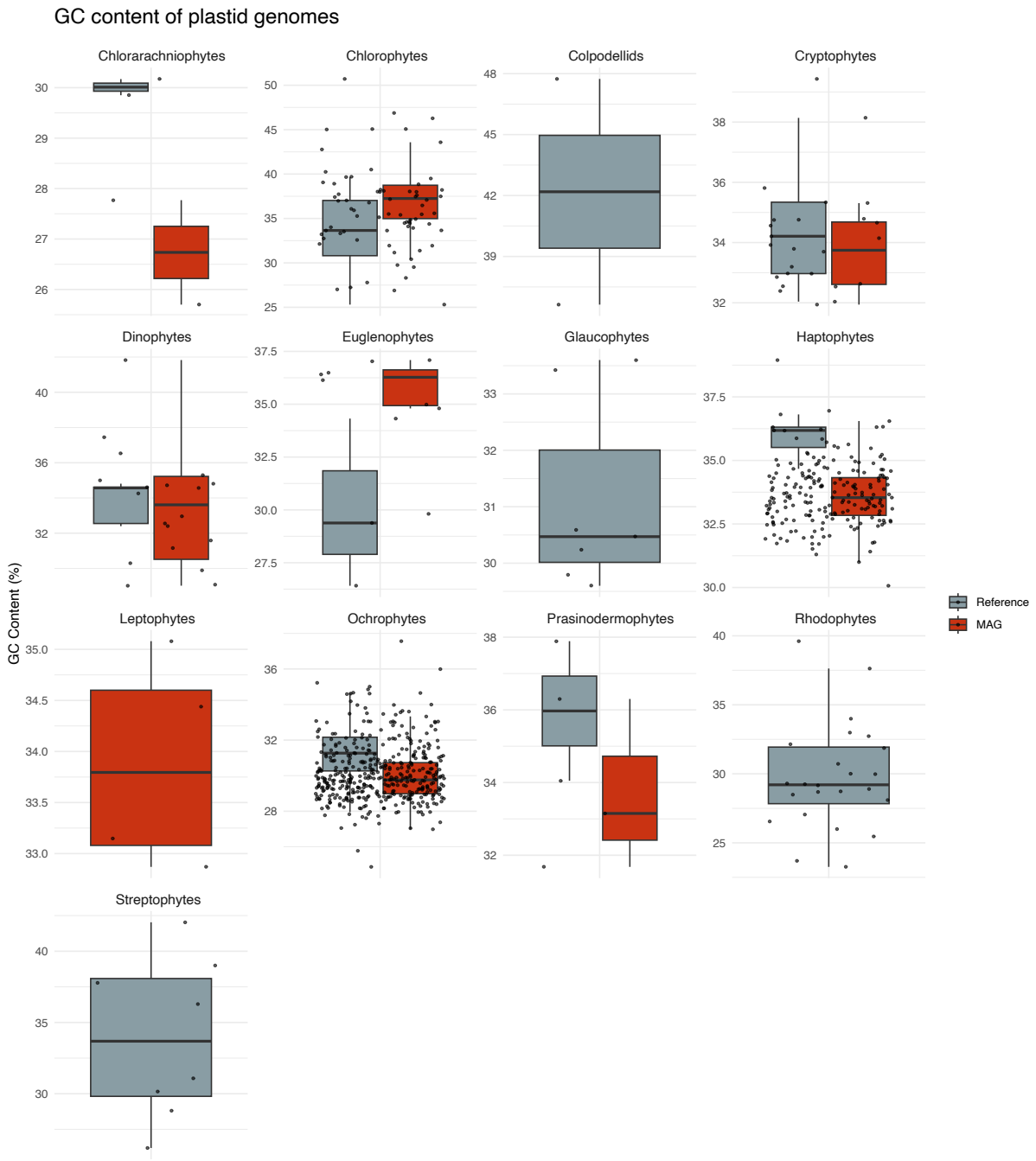
Supplementary Fig. 1. Top panel summarizes how many non-redundant ptMAGs were characterized from each of the 11 large *Tara* Oceans co-assemblies. Bottom panel shows the strong correlation between (1) the number of plastid-affiliated RNApolB genes occurring in each *Tara* Oceans metabin (each metabin contains a subset of contigs from one co-assembly that were automatically binned using differential coverage and sequence composition), (2) and the number of manually characterized ptMAGs across the same metabins.



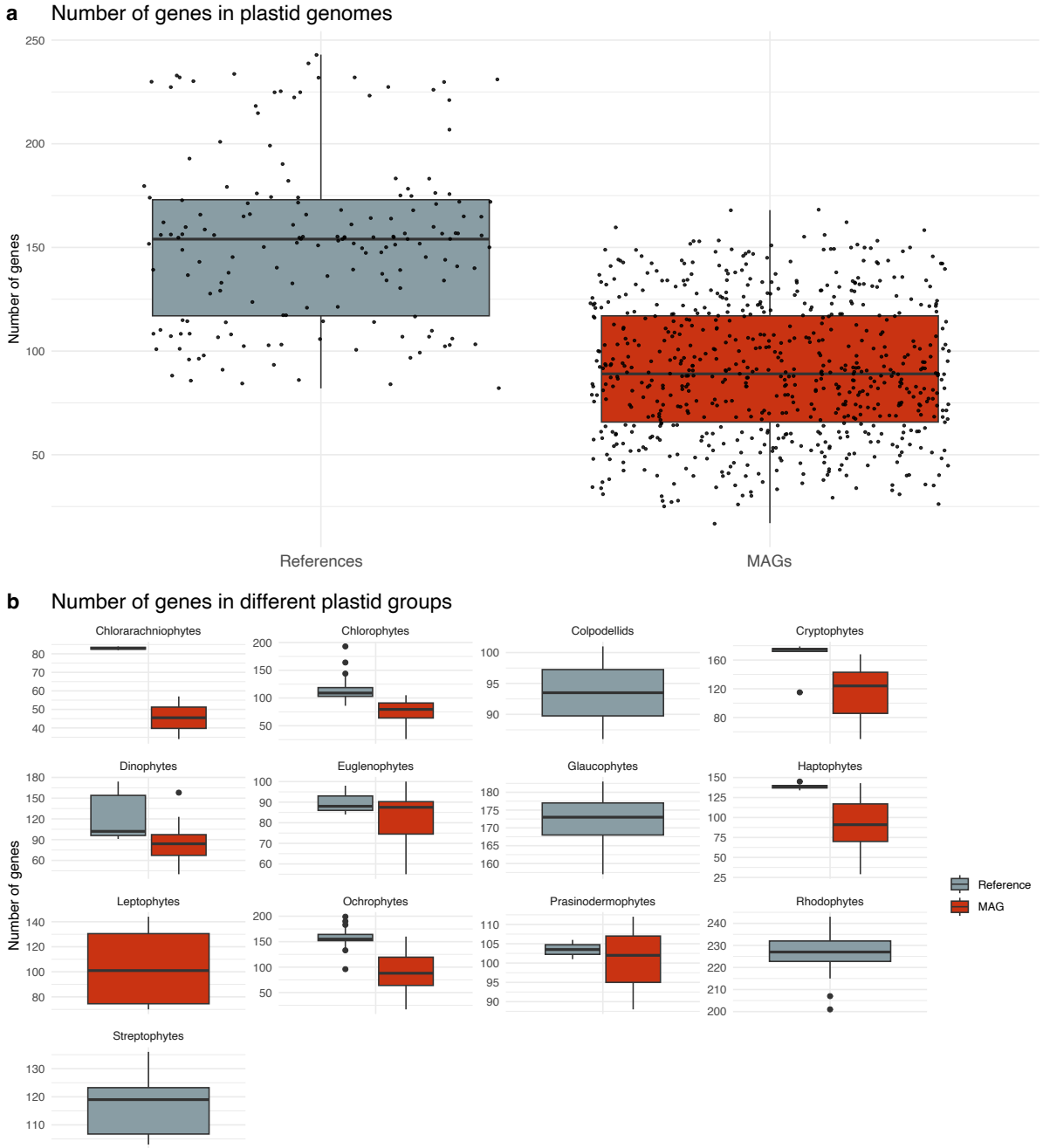
Supplementary Fig. 2. Estimated completeness of the reference plastid genomes (n=165, not including cyanobacteria and *Paulinella*) and ptMAGs (n=660). Panel (a) displays an overall comparison between references and ptMAGs with dots representing individual genomes, while panel (b) presents a comparison between groups with dots representing outlier genomes. Completeness was estimated based on the presence of 44 core genes (listed in Supplementary Table 6). Source data are provided in Supplementary Data 1.



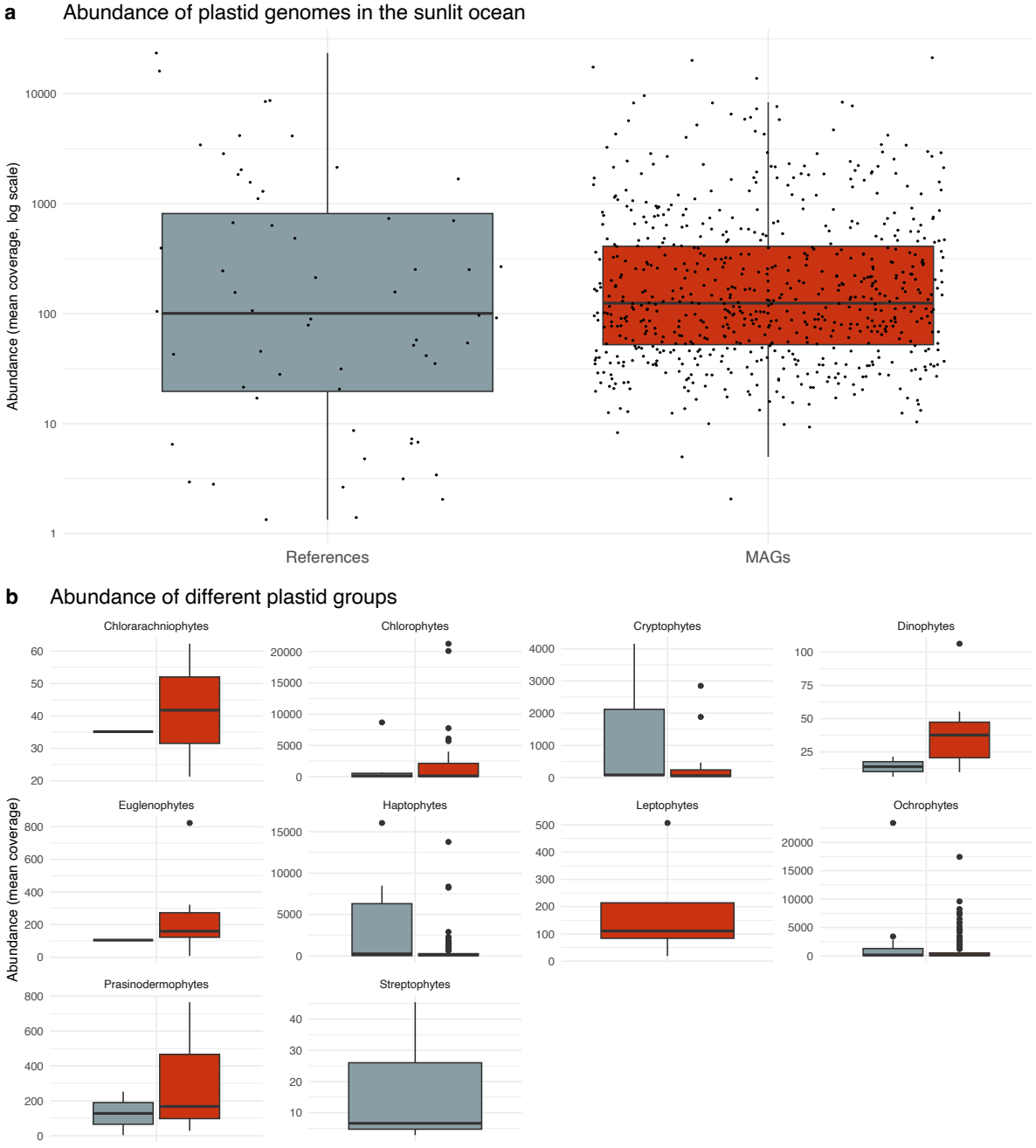
Supplementary Fig. 3. Size of the reference plastid genomes (n=165, not including cyanobacteria and *Paulinella*) and ptMAGs (n=660). Panel (a) shows that ptMAGs are smaller than reference genomes in general, with dots representing individual genomes, while panel (b) presents a comparison between groups with dots representing outlier genomes. Source data are provided in Supplementary Data 1.



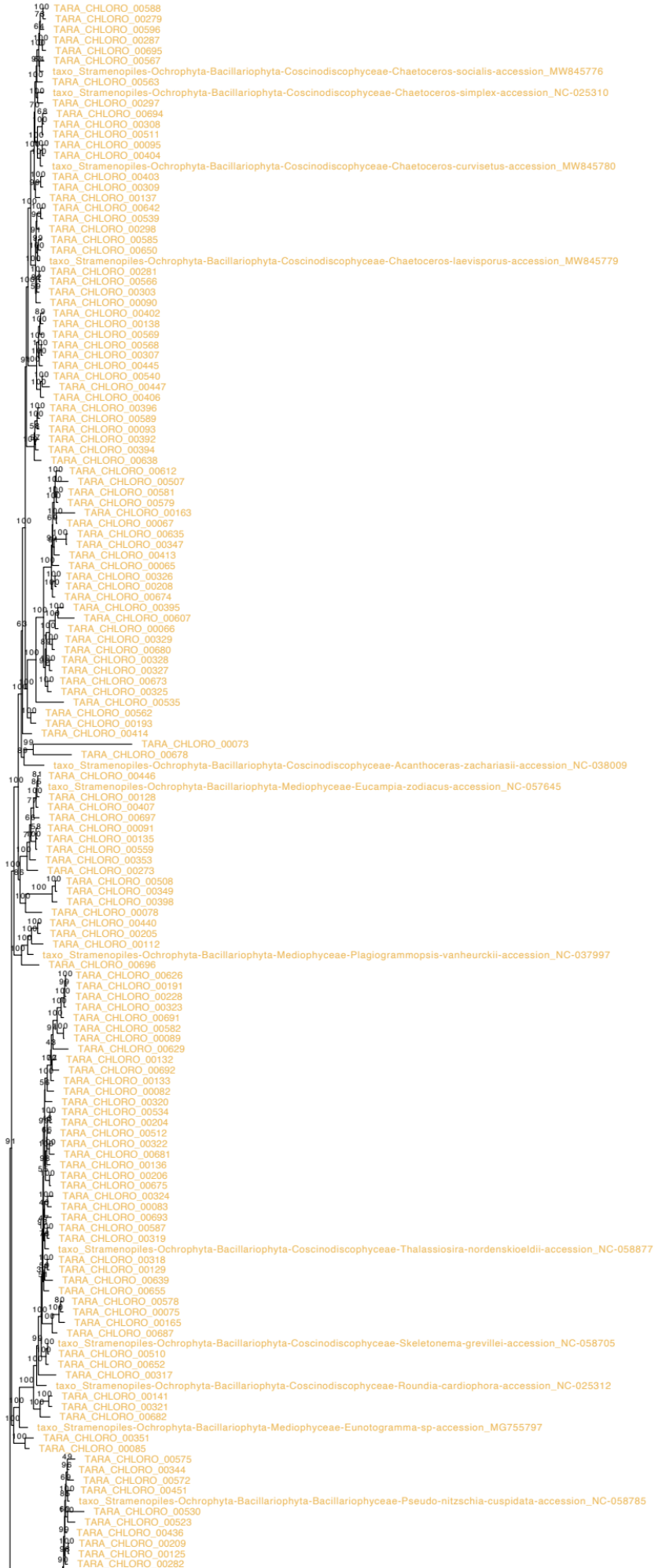
Supplementary Fig. 4. GC content of the plastid genomes across algal groups including references (n=165, not including cyanobacteria and *Paulinella*) and ptMAGs (n=660). Source data are provided in Supplementary Data 1.

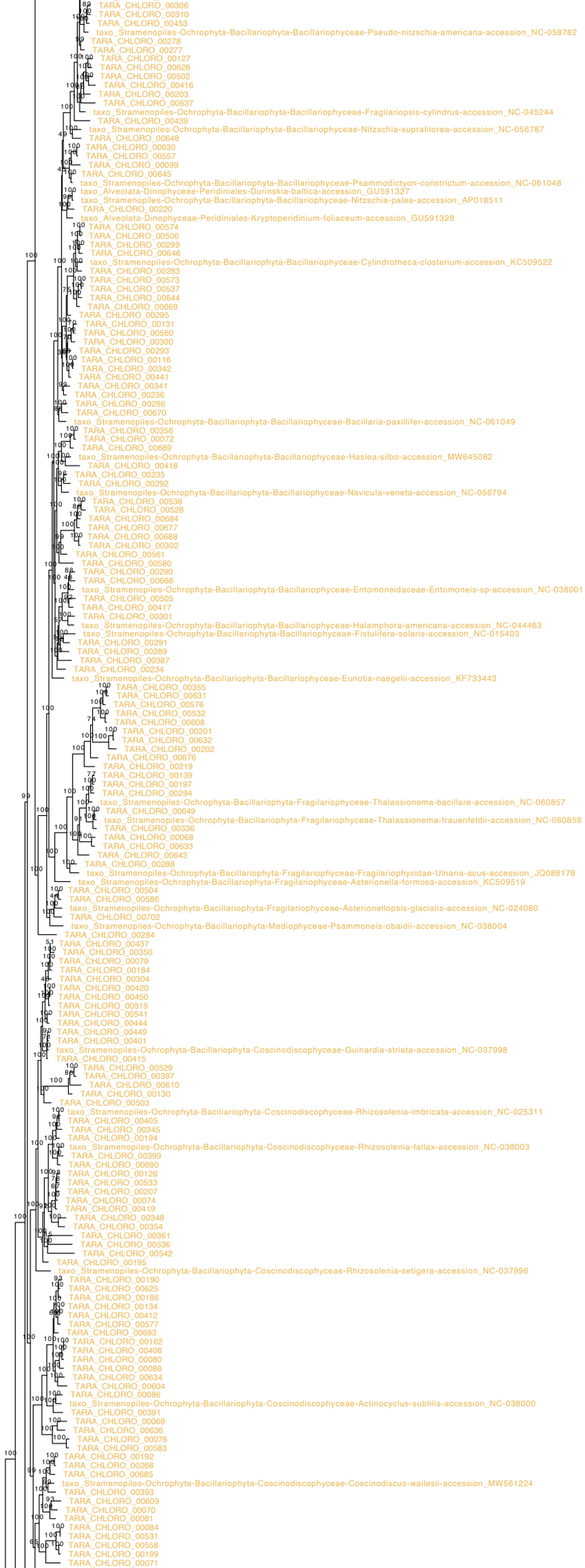


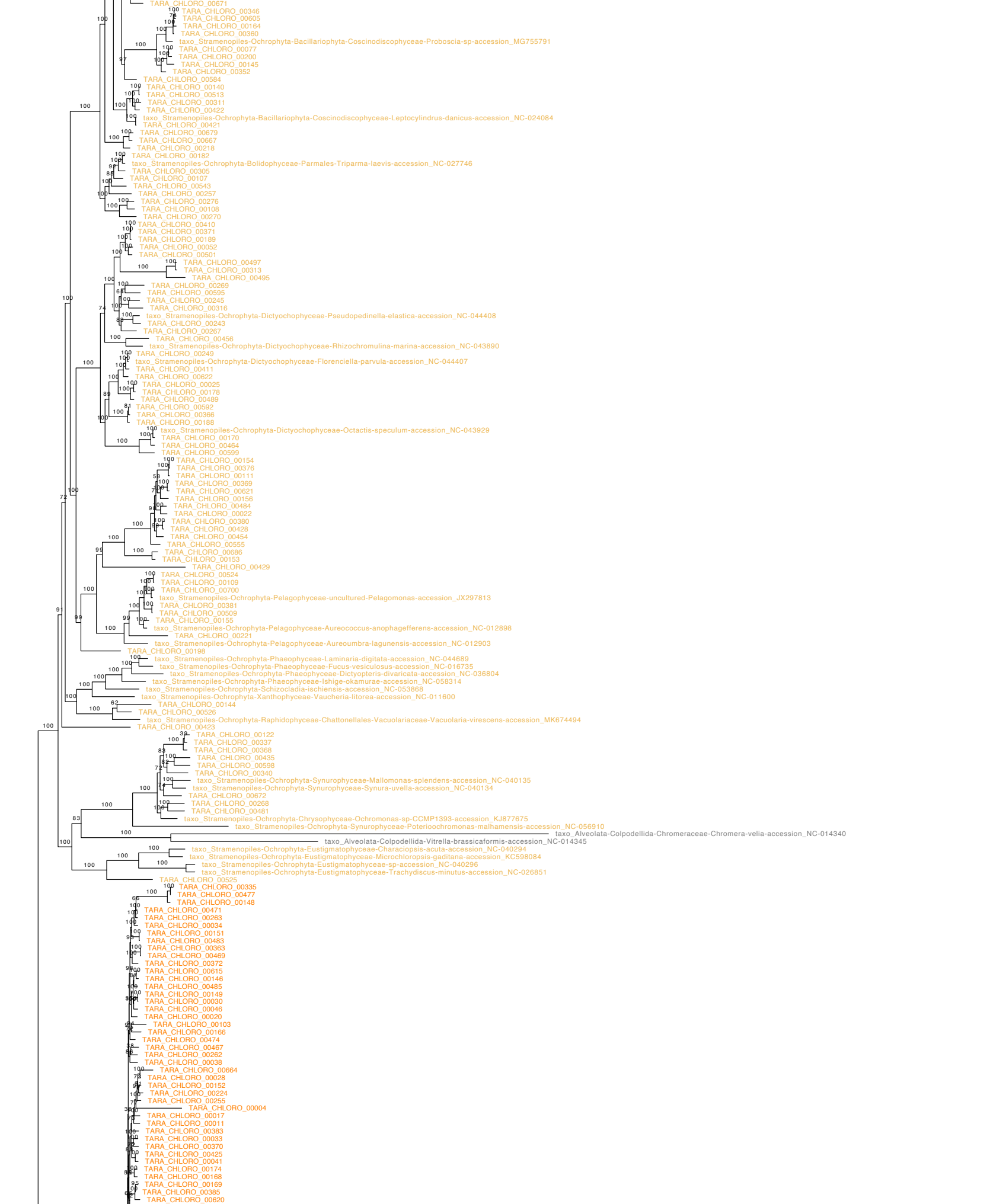
Supplementary Fig. 5. Number of genes encoded in the reference plastid genomes (n=165, not including cyanobacteria and *Paulinella*) and ptMAGs (n=660) included in the Maximum Likelihood phylogeny shown in Figure 1. Panel (a) shows that ptMAGs generally encode fewer genes due to lower completeness than reference genomes, with dots representing individual genomes, while panel (b) presents a comparison between groups with dots representing outlier genomes. Source data are provided in Supplementary Data 1.

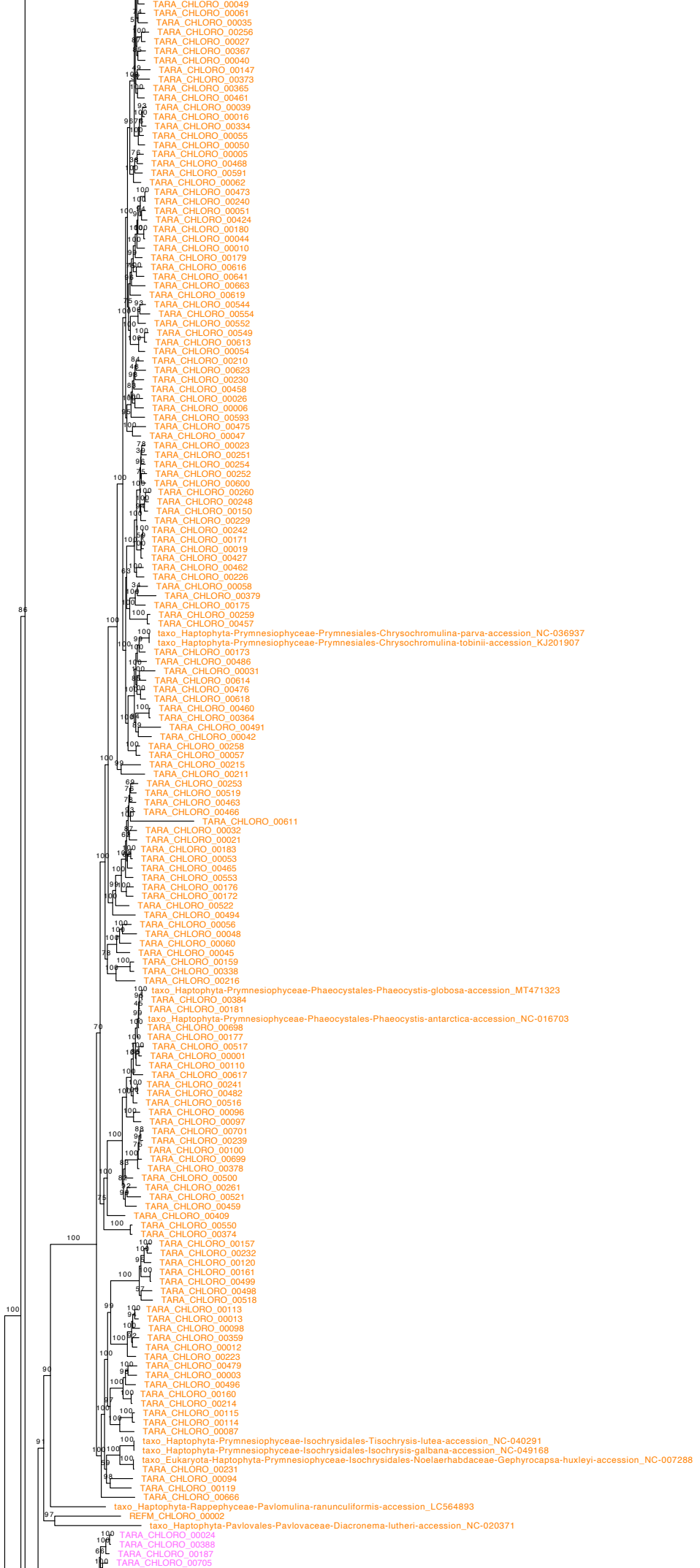


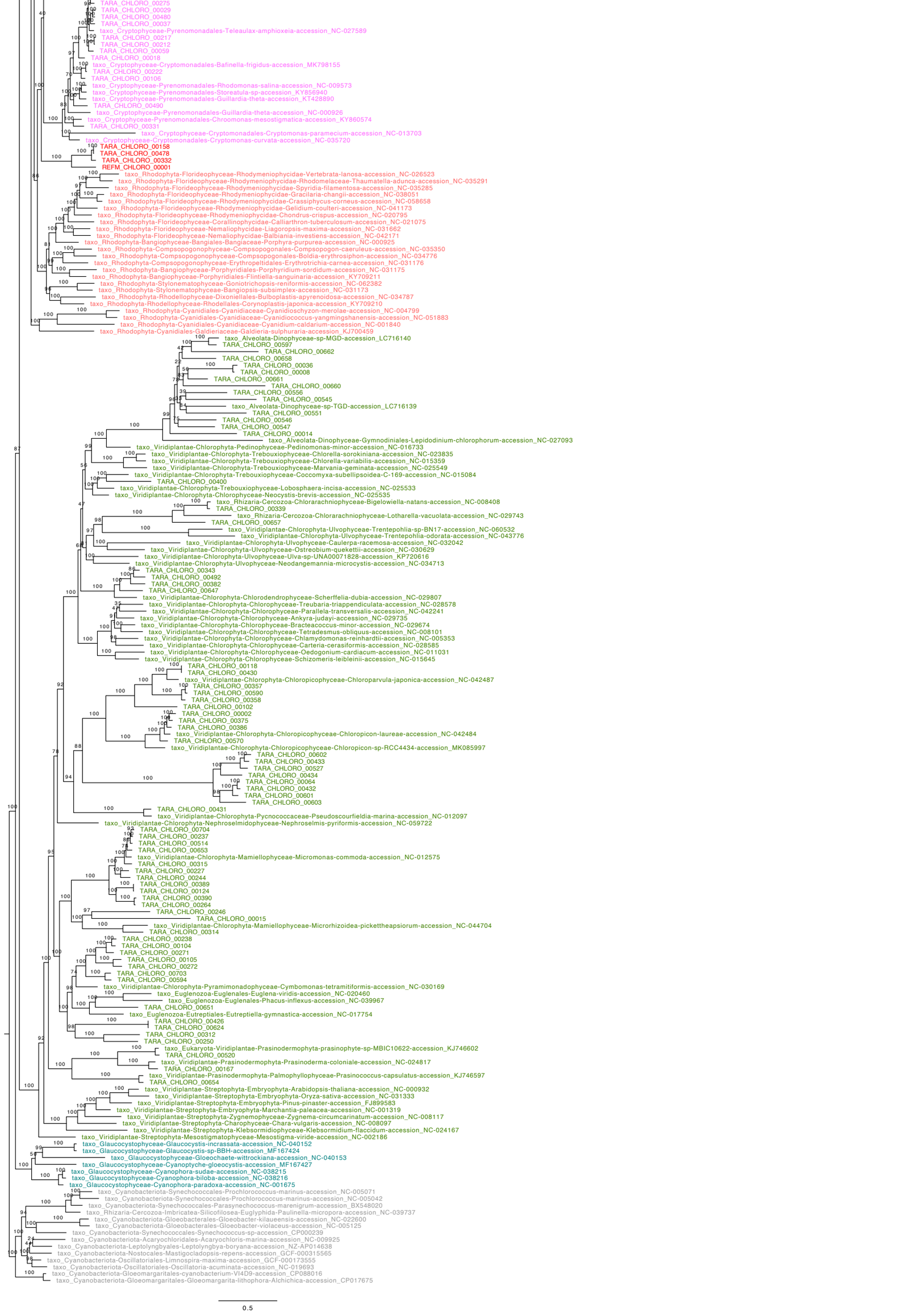
Supplementary Fig. 6. Abundance of plastid genomes in the sunlit ocean as estimated by mean vertical coverage obtained by mapping 937 *Tara* Oceans metagenomes to each genome. To account for non-specific read recruitment, a genome was only considered to be detected if at least 25% of its length was covered by reads. Panel (a) and (b) are based on 60 reference genomes (not including cyanobacteria, *Paulinella*, and non-marine taxa) and 660 ptMAGs. ptMAGs represent nearly 85% of the total signal, in part due to the low number of selected reference genomes, and in many cases, references still represent the most abundant taxa (e.g., ochrophytes, haptophytes, and cryptophytes). Source data are provided in Supplementary Data 1.


TARA_CHLORO_00588
TARA_CHLORO_00279
TARA_CHLORO_00596
TARA_CHLORO_00287
TARA_CHLORO_00695
TARA_CHLORO_00567
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Coscinodiscophyceae-Chaetoceros-socialis-accession_MW845776
TARA_CHLORO_00563
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Coscinodiscophyceae-Chaetoceros-simplex-accession_NC-025310
TARA_CHLORO_00297
TARA_CHLORO_00694
TARA_CHLORO_00308
TARA_CHLORO_00511
TARA_CHLORO_00095
TARA_CHLORO_00404
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Coscinodiscophyceae-Chaetoceros-curvisetus-accession_MW845780
TARA_CHLORO_00403
TARA_CHLORO_00309
TARA_CHLORO_00137
TARA_CHLORO_00642
TARA_CHLORO_00539
TARA_CHLORO_00298
TARA_CHLORO_00585
TARA_CHLORO_00650
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Coscinodiscophyceae-Chaetoceros-laevisporus-accession_MW845779
TARA_CHLORO_00281
TARA_CHLORO_00566
TARA_CHLORO_00303
TARA_CHLORO_00090
TARA_CHLORO_00402
TARA_CHLORO_00138
TARA_CHLORO_00569
TARA_CHLORO_00568
TARA_CHLORO_00307
TARA_CHLORO_00445
TARA_CHLORO_00540
TARA_CHLORO_00447
TARA_CHLORO_00406
TARA_CHLORO_00396
TARA_CHLORO_00589
TARA_CHLORO_00093
TARA_CHLORO_00392
TARA_CHLORO_00394
TARA_CHLORO_00638
TARA_CHLORO_00612
TARA_CHLORO_00507
TARA_CHLORO_00581
TARA_CHLORO_00579
TARA_CHLORO_00163
TARA_CHLORO_00067
TARA_CHLORO_00635
TARA_CHLORO_00347
TARA_CHLORO_00413
TARA_CHLORO_00065
TARA_CHLORO_00326
TARA_CHLORO_00208
TARA_CHLORO_00674
TARA_CHLORO_00395
TARA_CHLORO_00607
TARA_CHLORO_00066
TARA_CHLORO_00329
TARA_CHLORO_00680
TARA_CHLORO_00328
TARA_CHLORO_00327
TARA_CHLORO_00673
TARA_CHLORO_00325
TARA_CHLORO_00535
TARA_CHLORO_00562
TARA_CHLORO_00193
TARA_CHLORO_00414
TARA_CHLORO_00073
TARA_CHLORO_00678
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Coscinodiscophyceae-Acanthoceras-zachariasii-accession_NC-038009
TARA_CHLORO_00446
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Mediophyceae-Eucampia-zodiacus-accession_NC-057645
TARA_CHLORO_00128
TARA_CHLORO_00407
TARA_CHLORO_00697
TARA_CHLORO_00091
TARA_CHLORO_00135
TARA_CHLORO_00559
TARA_CHLORO_00353
TARA_CHLORO_00273
TARA_CHLORO_00508
TARA_CHLORO_00349
TARA_CHLORO_00398
TARA_CHLORO_00078
TARA_CHLORO_00440
TARA_CHLORO_00205
TARA_CHLORO_00112
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Mediophyceae-Plagiogrammopsis-vanheurckii-accession_NC-037997
TARA_CHLORO_00696
TARA_CHLORO_00626
TARA_CHLORO_00191
TARA_CHLORO_00228
TARA_CHLORO_00323
TARA_CHLORO_00691
TARA_CHLORO_00582
TARA_CHLORO_00089
TARA_CHLORO_00629
TARA_CHLORO_00132
TARA_CHLORO_00692
TARA_CHLORO_00133
TARA_CHLORO_00082
TARA_CHLORO_00320
TARA_CHLORO_00534
TARA_CHLORO_00204
TARA_CHLORO_00512
TARA_CHLORO_00322
TARA_CHLORO_00681
TARA_CHLORO_00136
TARA_CHLORO_00206
TARA_CHLORO_00675
TARA_CHLORO_00324
TARA_CHLORO_00083
TARA_CHLORO_00693
TARA_CHLORO_00587
TARA_CHLORO_00319
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Coscinodiscophyceae-Thalassiosira-nordenskioeldii-accession_NC-058877
TARA_CHLORO_00318
TARA_CHLORO_00129
TARA_CHLORO_00639
TARA_CHLORO_00655
TARA_CHLORO_00578
TARA_CHLORO_00075
TARA_CHLORO_00165
TARA_CHLORO_00687
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Coscinodiscophyceae-Skeletonema-grevillei-accession_NC-058705
TARA_CHLORO_00510
TARA_CHLORO_00652
TARA_CHLORO_00317
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Coscinodiscophyceae-Roundia-cardiophora-accession_NC-025312
TARA_CHLORO_00141
TARA_CHLORO_00321
TARA_CHLORO_00682
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Mediophyceae-Eunotogramma-sp-accession_MG755797
TARA_CHLORO_00351
TARA_CHLORO_00085
TARA_CHLORO_00575
TARA_CHLORO_00344
TARA_CHLORO_00572
TARA_CHLORO_00451
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Bacillariophyceae-Pseudo-nitzschia-cuspidata-accession_NC-058785
TARA_CHLORO_00530
TARA_CHLORO_00523
TARA_CHLORO_00436
TARA_CHLORO_00209
TARA_CHLORO_00125
TARA_CHLORO_00282

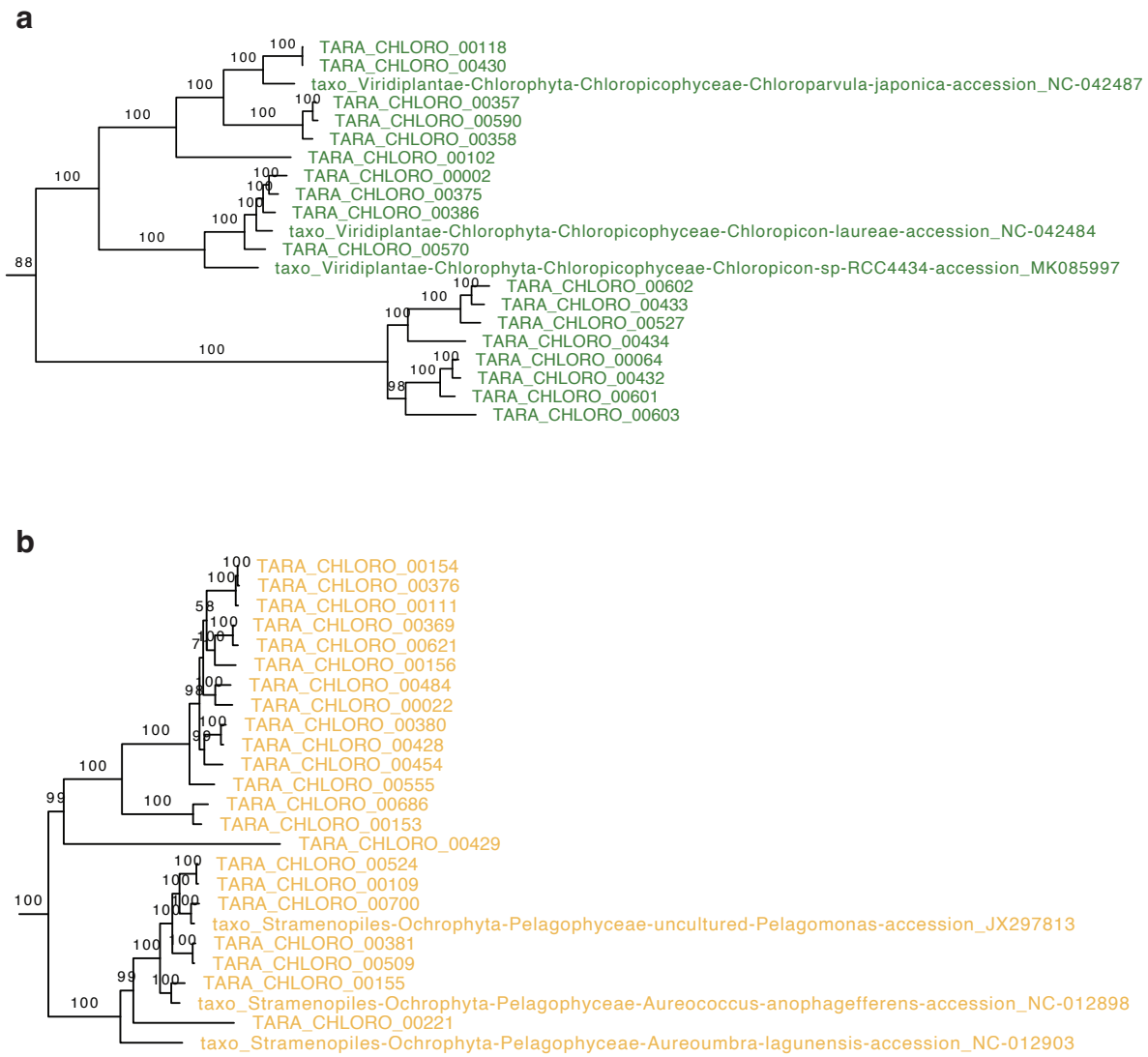

TARA_CHLORO_00306
TARA_CHLORO_00310
TARA_CHLORO_00453
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Bacillariophyceae-Pseudo-nitzschia-americana-accession_NC-058782
TARA_CHLORO_00278
TARA_CHLORO_00277
TARA_CHLORO_00127
TARA_CHLORO_00628
TARA_CHLORO_00502
TARA_CHLORO_00416
TARA_CHLORO_00203
TARA_CHLORO_00637
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Bacillariophyceae-Fragilariopsis-cylindrus-accession_NC-045244
TARA_CHLORO_00439
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Bacillariophyceae-Nitzschia-supralitorea-accession_NC-056787
TARA_CHLORO_00648
TARA_CHLORO_00630
TARA_CHLORO_00557
TARA_CHLORO_00099
TARA_CHLORO_00645
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Bacillariophyceae-Psammodictyon-constrictum-accession_NC-061048
taxo_Alveolata-Dinophyceae-Peridiniales-Durinskia-baltica-accession_GU591327
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Bacillariophyceae-Nitzschia-palea-accession_AP018511
TARA_CHLORO_00220
taxo_Alveolata-Dinophyceae-Peridiniales-Kryptoperidinium-foliaceum-accession_GU591328
TARA_CHLORO_00574
TARA_CHLORO_00506
TARA_CHLORO_00299
TARA_CHLORO_00646
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Bacillariophyceae-Cylindrotheca-closterium-accession_KC509522
TARA_CHLORO_00283
TARA_CHLORO_00573
TARA_CHLORO_00537
TARA_CHLORO_00644
TARA_CHLORO_00669
TARA_CHLORO_00295
TARA_CHLORO_00131
TARA_CHLORO_00560
TARA_CHLORO_00300
TARA_CHLORO_00293
TARA_CHLORO_00116
TARA_CHLORO_00342
TARA_CHLORO_00441
TARA_CHLORO_00341
TARA_CHLORO_00236
TARA_CHLORO_00286
TARA_CHLORO_00670
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Bacillariophyceae-Bacillaria-paxillifer-accession_NC-061049
TARA_CHLORO_00356
TARA_CHLORO_00072
TARA_CHLORO_00689
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Bacillariophyceae-Haslea-silbo-accession_MW645082
TARA_CHLORO_00418
TARA_CHLORO_00235
TARA_CHLORO_00292
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Bacillariophyceae-Navicula-veneta-accession_NC-056794
TARA_CHLORO_00538
TARA_CHLORO_00528
TARA_CHLORO_00684
TARA_CHLORO_00677
TARA_CHLORO_00688
TARA_CHLORO_00302
TARA_CHLORO_00561
TARA_CHLORO_00580
TARA_CHLORO_00290
TARA_CHLORO_00668
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Bacillariophyceae-Entomoneidaceae-Entomoneis-sp-accession_NC-038001
TARA_CHLORO_00505
TARA_CHLORO_00417
TARA_CHLORO_00301
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Bacillariophyceae-Halamphora-americana-accession_NC-044463
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Bacillariophyceae-Fistulifera-solaris-accession_NC-015403
TARA_CHLORO_00291
TARA_CHLORO_00289
TARA_CHLORO_00387
TARA_CHLORO_00234
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Bacillariophyceae-Eunotia-naegelii-accession_KF733443
TARA_CHLORO_00355
TARA_CHLORO_00631
TARA_CHLORO_00576
TARA_CHLORO_00532
TARA_CHLORO_00608
TARA_CHLORO_00201
TARA_CHLORO_00632
TARA_CHLORO_00202
TARA_CHLORO_00676
TARA_CHLORO_00219
TARA_CHLORO_00139
TARA_CHLORO_00197
TARA_CHLORO_00294
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Fragilariophyceae-Thalassionema-bacillare-accession_NC-060857
TARA_CHLORO_00649
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Fragilariophyceae-Thalassionema-frauenfeldii-accession_NC-060858
TARA_CHLORO_00336
TARA_CHLORO_00068
TARA_CHLORO_00633
TARA_CHLORO_00643
TARA_CHLORO_00288
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Fragilariophyceae-Fragilariophycidae-Ulnaria-acus-accession_JQ088178
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Fragilariophyceae-Asterionella-formosa-accession_KC509519
TARA_CHLORO_00504
TARA_CHLORO_00586
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Fragilariophyceae-Asterionellopsis-glacialis-accession_NC-024080
TARA_CHLORO_00702
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Mediophyceae-Psammoneis-obaidii-accession_NC-038004
TARA_CHLORO_00284
TARA_CHLORO_00437
TARA_CHLORO_00350
TARA_CHLORO_00079
TARA_CHLORO_00184
TARA_CHLORO_00304
TARA_CHLORO_00420
TARA_CHLORO_00450
TARA_CHLORO_00515
TARA_CHLORO_00541
TARA_CHLORO_00444
TARA_CHLORO_00449
TARA_CHLORO_00401
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Coscinodiscophyceae-Guinardia-striata-accession_NC-037998
TARA_CHLORO_00415
TARA_CHLORO_00529
TARA_CHLORO_00397
TARA_CHLORO_00610
TARA_CHLORO_00130
TARA_CHLORO_00503
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Coscinodiscophyceae-Rhizosolenia-imbricata-accession_NC-025311
TARA_CHLORO_00405
TARA_CHLORO_00345
TARA_CHLORO_00194
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Coscinodiscophyceae-Rhizosolenia-fallax-accession_NC-038003
TARA_CHLORO_00399
TARA_CHLORO_00690
TARA_CHLORO_00126
TARA_CHLORO_00533
TARA_CHLORO_00207
TARA_CHLORO_00074
TARA_CHLORO_00419
TARA_CHLORO_00348
TARA_CHLORO_00354
TARA_CHLORO_00361
TARA_CHLORO_00536
TARA_CHLORO_00542
TARA_CHLORO_00195
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Coscinodiscophyceae-Rhizosolenia-setigera-accession_NC-037996
TARA_CHLORO_00190
TARA_CHLORO_00625
TARA_CHLORO_00186
TARA_CHLORO_00134
TARA_CHLORO_00412
TARA_CHLORO_00577
TARA_CHLORO_00683
TARA_CHLORO_00162
TARA_CHLORO_00408
TARA_CHLORO_00080
TARA_CHLORO_00088
TARA_CHLORO_00634
TARA_CHLORO_00604
TARA_CHLORO_00086
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Coscinodiscophyceae-Actinocyclus-subtilis-accession_NC-038000
TARA_CHLORO_00391
TARA_CHLORO_00069
TARA_CHLORO_00636
TARA_CHLORO_00076
TARA_CHLORO_00583
TARA_CHLORO_00192
TARA_CHLORO_00266
TARA_CHLORO_00685
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Coscinodiscophyceae-Coscinodiscus-wailesii-accession_MW561224
TARA_CHLORO_00393
TARA_CHLORO_00609
TARA_CHLORO_00070
TARA_CHLORO_00081
TARA_CHLORO_00084
TARA_CHLORO_00531
TARA_CHLORO_00558
TARA_CHLORO_00199
TARA_CHLORO_00071


TARA_CHLORO_00671
TARA_CHLORO_00346
TARA_CHLORO_00605
TARA_CHLORO_00164
TARA_CHLORO_00360
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Coscinodiscophyceae-Proboscia-sp-accession_MG755791
TARA_CHLORO_00077
TARA_CHLORO_00200
TARA_CHLORO_00145
TARA_CHLORO_00352
TARA_CHLORO_00584
TARA_CHLORO_00140
TARA_CHLORO_00513
TARA_CHLORO_00311
TARA_CHLORO_00422
taxo_Stramenopiles-Ochrophyta-Bacillariophyta-Coscinodiscophyceae-Leptocylindrus-danicus-accession_NC-024084
TARA_CHLORO_00421
TARA_CHLORO_00679
TARA_CHLORO_00667
TARA_CHLORO_00218
TARA_CHLORO_00182
taxo_Stramenopiles-Ochrophyta-Bolidophyceae-Parmales-Triparma-laevis-accession_NC-027746
TARA_CHLORO_00305
TARA_CHLORO_00107
TARA_CHLORO_00543
TARA_CHLORO_00257
TARA_CHLORO_00276
TARA_CHLORO_00108
TARA_CHLORO_00270
TARA_CHLORO_00410
TARA_CHLORO_00371
TARA_CHLORO_00189
TARA_CHLORO_00052
TARA_CHLORO_00501
TARA_CHLORO_00497
TARA_CHLORO_00313
TARA_CHLORO_00495
TARA_CHLORO_00269
TARA_CHLORO_00595
TARA_CHLORO_00245
TARA_CHLORO_00316
taxo_Stramenopiles-Ochrophyta-Dictyochophyceae-Pseudopedinella-elastica-accession_NC-044408
TARA_CHLORO_00243
TARA_CHLORO_00267
TARA_CHLORO_00456
taxo_Stramenopiles-Ochrophyta-Dictyochophyceae-Rhizochromulina-marina-accession_NC-043890
TARA_CHLORO_00249
taxo_Stramenopiles-Ochrophyta-Dictyochophyceae-Florenciella-parvula-accession_NC-044407
TARA_CHLORO_00411
TARA_CHLORO_00622
TARA_CHLORO_00025
TARA_CHLORO_00178
TARA_CHLORO_00489
TARA_CHLORO_00592
TARA_CHLORO_00366
TARA_CHLORO_00188
taxo_Stramenopiles-Ochrophyta-Dictyochophyceae-Octactis-speculum-accession_NC-043929
TARA_CHLORO_00170
TARA_CHLORO_00464
TARA_CHLORO_00599
TARA_CHLORO_00154
TARA_CHLORO_00376
TARA_CHLORO_00111
TARA_CHLORO_00369
TARA_CHLORO_00621
TARA_CHLORO_00156
TARA_CHLORO_00484
TARA_CHLORO_00022
TARA_CHLORO_00380
TARA_CHLORO_00428
TARA_CHLORO_00454
TARA_CHLORO_00555
TARA_CHLORO_00686
TARA_CHLORO_00153
TARA_CHLORO_00429
TARA_CHLORO_00524
TARA_CHLORO_00109
TARA_CHLORO_00700
taxo_Stramenopiles-Ochrophyta-Pelagophyceae-uncultured-Pelagomonas-accession_JX297813
TARA_CHLORO_00381
TARA_CHLORO_00509
TARA_CHLORO_00155
taxo_Stramenopiles-Ochrophyta-Pelagophyceae-Aureococcus-anophagefferens-accession_NC-012898
TARA_CHLORO_00221
taxo_Stramenopiles-Ochrophyta-Pelagophyceae-Aureoumbra-lagunensis-accession_NC-012903
TARA_CHLORO_00198
taxo_Stramenopiles-Ochrophyta-Phaeophyceae-Laminaria-digitata-accession_NC-044689
taxo_Stramenopiles-Ochrophyta-Phaeophyceae-Fucus-vesiculosus-accession_NC-016735
taxo_Stramenopiles-Ochrophyta-Phaeophyceae-Dictyopteris-divaricata-accession_NC-036804
taxo_Stramenopiles-Ochrophyta-Phaeophyceae-Ishige-okamurae-accession_NC-058314
taxo_Stramenopiles-Ochrophyta-Schizocladia-ischiensis-accession_NC-053868
taxo_Stramenopiles-Ochrophyta-Xanthophyceae-Vaucheria-litorea-accession_NC-011600
TARA_CHLORO_00144
TARA_CHLORO_00526
taxo_Stramenopiles-Ochrophyta-Raphidophyceae-Chattonellales-Vacuolariaceae-Vacuolaria-virescens-accession_MK674494
TARA_CHLORO_00423
TARA_CHLORO_00122
TARA_CHLORO_00337
TARA_CHLORO_00368
TARA_CHLORO_00435
TARA_CHLORO_00598
TARA_CHLORO_00340
taxo_Stramenopiles-Ochrophyta-Synurophyceae-Mallomonas-splendens-accession_NC-040135
taxo_Stramenopiles-Ochrophyta-Synurophyceae-Synura-uvella-accession_NC-040134
TARA_CHLORO_00672
TARA_CHLORO_00268
TARA_CHLORO_00481
taxo_Stramenopiles-Ochrophyta-Chrysophyceae-Ochromonas-sp-CCMP1393-accession_KJ877675
taxo_Stramenopiles-Ochrophyta-Synurophyceae-Poterioochromonas-malhamensis-accession_NC-056910
taxo_Alveolata-Colpodellida-Chromeraceae-Chromera-velia-accession_NC-014340
taxo_Alveolata-Colpodellida-Vitrella-brassicaformis-accession_NC-014345
taxo_Stramenopiles-Ochrophyta-Eustigmatophyceae-Characiopsis-acuta-accession_NC-040294
taxo_Stramenopiles-Ochrophyta-Eustigmatophyceae-Microchloropsis-gaditana-accession_KC598084
taxo_Stramenopiles-Ochrophyta-Eustigmatophyceae-sp-accession_NC-040296
taxo_Stramenopiles-Ochrophyta-Eustigmatophyceae-Trachydiscus-minutus-accession_NC-026851
TARA_CHLORO_00525
TARA_CHLORO_00335
TARA_CHLORO_00477
TARA_CHLORO_00148
TARA_CHLORO_00471
TARA_CHLORO_00263
TARA_CHLORO_00034
TARA_CHLORO_00151
TARA_CHLORO_00483
TARA_CHLORO_00363
TARA_CHLORO_00469
TARA_CHLORO_00372
TARA_CHLORO_00615
TARA_CHLORO_00146
TARA_CHLORO_00485
TARA_CHLORO_00149
TARA_CHLORO_00030
TARA_CHLORO_00046
TARA_CHLORO_00020
TARA_CHLORO_00103
TARA_CHLORO_00166
TARA_CHLORO_00474
TARA_CHLORO_00467
TARA_CHLORO_00262
TARA_CHLORO_00038
TARA_CHLORO_00664
TARA_CHLORO_00028
TARA_CHLORO_00152
TARA_CHLORO_00224
TARA_CHLORO_00255
TARA_CHLORO_00004
TARA_CHLORO_00017
TARA_CHLORO_00011
TARA_CHLORO_00383
TARA_CHLORO_00033
TARA_CHLORO_00370
TARA_CHLORO_00425
TARA_CHLORO_00041
TARA_CHLORO_00174
TARA_CHLORO_00168
TARA_CHLORO_00169
TARA_CHLORO_00385
TARA_CHLORO_00620

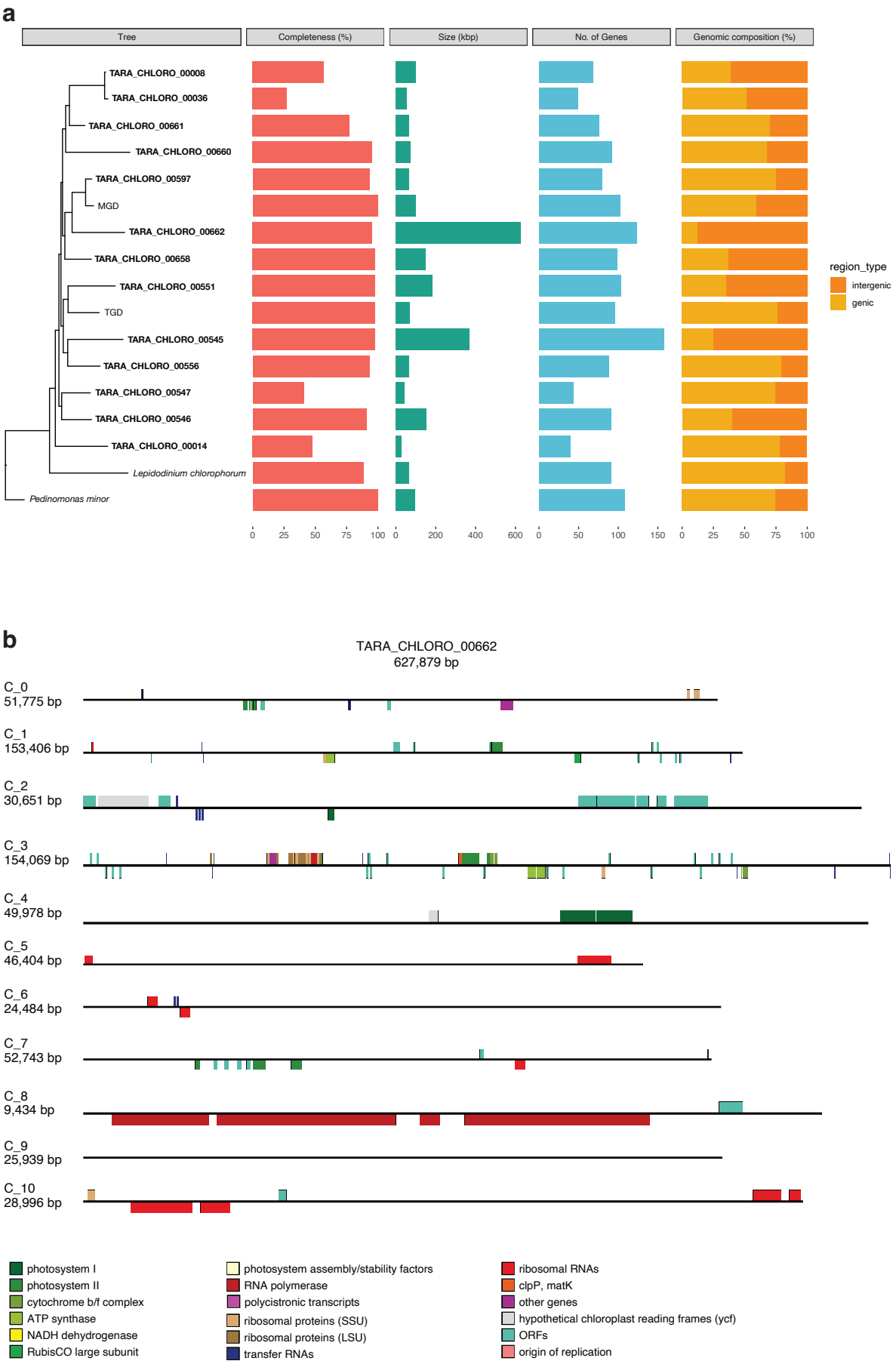

TARA_CHLORO_00049
TARA_CHLORO_00061
TARA_CHLORO_00035
TARA_CHLORO_00256
TARA_CHLORO_00027
TARA_CHLORO_00367
TARA_CHLORO_00040
TARA_CHLORO_00147
TARA_CHLORO_00373
TARA_CHLORO_00365
TARA_CHLORO_00461
TARA_CHLORO_00039
TARA_CHLORO_00016
TARA_CHLORO_00334
TARA_CHLORO_00055
TARA_CHLORO_00050
TARA_CHLORO_00005
TARA_CHLORO_00468
TARA_CHLORO_00591
TARA_CHLORO_00062
TARA_CHLORO_00473
TARA_CHLORO_00240
TARA_CHLORO_00051
TARA_CHLORO_00424
TARA_CHLORO_00180
TARA_CHLORO_00044
TARA_CHLORO_00010
TARA_CHLORO_00179
TARA_CHLORO_00616
TARA_CHLORO_00641
TARA_CHLORO_00663
TARA_CHLORO_00619
TARA_CHLORO_00544
TARA_CHLORO_00554
TARA_CHLORO_00552
TARA_CHLORO_00549
TARA_CHLORO_00613
TARA_CHLORO_00054
TARA_CHLORO_00210
TARA_CHLORO_00623
TARA_CHLORO_00230
TARA_CHLORO_00458
TARA_CHLORO_00026
TARA_CHLORO_00006
TARA_CHLORO_00593
TARA_CHLORO_00475
TARA_CHLORO_00047
TARA_CHLORO_00023
TARA_CHLORO_00251
TARA_CHLORO_00254
TARA_CHLORO_00252
TARA_CHLORO_00600
TARA_CHLORO_00260
TARA_CHLORO_00248
TARA_CHLORO_00150
TARA_CHLORO_00229
TARA_CHLORO_00242
TARA_CHLORO_00171
TARA_CHLORO_00019
TARA_CHLORO_00427
TARA_CHLORO_00462
TARA_CHLORO_00226
TARA_CHLORO_00058
TARA_CHLORO_00379
TARA_CHLORO_00175
TARA_CHLORO_00259
TARA_CHLORO_00457
taxo_Haptophyta-Prymnesiophyceae-Prymnesiales-Chrysochromulina-parva-accession_NC-036937
taxo_Haptophyta-Prymnesiophyceae-Prymnesiales-Chrysochromulina-tobinii-accession_KJ201907
TARA_CHLORO_00173
TARA_CHLORO_00486
TARA_CHLORO_00031
TARA_CHLORO_00614
TARA_CHLORO_00476
TARA_CHLORO_00618
TARA_CHLORO_00460
TARA_CHLORO_00364
TARA_CHLORO_00491
TARA_CHLORO_00042
TARA_CHLORO_00258
TARA_CHLORO_00057
TARA_CHLORO_00215
TARA_CHLORO_00211
TARA_CHLORO_00253
TARA_CHLORO_00519
TARA_CHLORO_00463
TARA_CHLORO_00466
TARA_CHLORO_00611
TARA_CHLORO_00032
TARA_CHLORO_00021
TARA_CHLORO_00183
TARA_CHLORO_00053
TARA_CHLORO_00465
TARA_CHLORO_00553
TARA_CHLORO_00176
TARA_CHLORO_00172
TARA_CHLORO_00522
TARA_CHLORO_00494
TARA_CHLORO_00056
TARA_CHLORO_00048
TARA_CHLORO_00060
TARA_CHLORO_00045
TARA_CHLORO_00159
TARA_CHLORO_00338
TARA_CHLORO_00216
taxo_Haptophyta-Prymnesiophyceae-Phaeocystales-Phaeocystis-globosa-accession_MT471323
TARA_CHLORO_00384
TARA_CHLORO_00181
taxo_Haptophyta-Prymnesiophyceae-Phaeocystales-Phaeocystis-antarctica-accession_NC-016703
TARA_CHLORO_00698
TARA_CHLORO_00177
TARA_CHLORO_00517
TARA_CHLORO_00001
TARA_CHLORO_00110
TARA_CHLORO_00617
TARA_CHLORO_00241
TARA_CHLORO_00482
TARA_CHLORO_00516
TARA_CHLORO_00096
TARA_CHLORO_00097
TARA_CHLORO_00701
TARA_CHLORO_00239
TARA_CHLORO_00100
TARA_CHLORO_00699
TARA_CHLORO_00378
TARA_CHLORO_00500
TARA_CHLORO_00261
TARA_CHLORO_00521
TARA_CHLORO_00459
TARA_CHLORO_00409
TARA_CHLORO_00550
TARA_CHLORO_00374
TARA_CHLORO_00157
TARA_CHLORO_00232
TARA_CHLORO_00120
TARA_CHLORO_00161
TARA_CHLORO_00499
TARA_CHLORO_00498
TARA_CHLORO_00518
TARA_CHLORO_00113
TARA_CHLORO_00013
TARA_CHLORO_00098
TARA_CHLORO_00359
TARA_CHLORO_00012
TARA_CHLORO_00223
TARA_CHLORO_00479
TARA_CHLORO_00003
TARA_CHLORO_00496
TARA_CHLORO_00160
TARA_CHLORO_00214
TARA_CHLORO_00115
TARA_CHLORO_00114
TARA_CHLORO_00087
taxo_Haptophyta-Prymnesiophyceae-Isochrysidales-Tisochrysis-lutea-accession_NC-040291
taxo_Haptophyta-Prymnesiophyceae-Isochrysidales-Isochrysis-galbana-accession_NC-049168
taxo_Eukaryota-Haptophyta-Prymnesiophyceae-Isochrysidales-Noelaerhabdaceae-Gephyrocapsa-huxleyi-accession_NC-007288
TARA_CHLORO_00231
TARA_CHLORO_00094
TARA_CHLORO_00119
TARA_CHLORO_00666
taxo_Haptophyta-Rappephyceae-Pavlomulina-ranunculiformis-accession_LC564893
REFM_CHLORO_00002
taxo_Haptophyta-Pavlovales-Pavlovaceae-Diacronema-lutheri-accession_NC-020371
TARA_CHLORO_00024
TARA_CHLORO_00388
TARA_CHLORO_00187
TARA_CHLORO_00705


TARA_CHLORO_00275
TARA_CHLORO_00029
TARA_CHLORO_00480
TARA_CHLORO_00037
taxo_Cryptophyceae-Pyrenomonadales-Teleaulax-amphioxeia-accession_NC-027589
TARA_CHLORO_00217
TARA_CHLORO_00212
TARA_CHLORO_00059
TARA_CHLORO_00018
taxo_Cryptophyceae-Cryptomonadales-Bafinella-frigidus-accession_MK798155
TARA_CHLORO_00222
TARA_CHLORO_00106
taxo_Cryptophyceae-Pyrenomonadales-Rhodomonas-salina-accession_NC-009573
taxo_Cryptophyceae-Pyrenomonadales-Storeatula-sp-accession_KY856940
taxo_Cryptophyceae-Pyrenomonadales-Guillardia-theta-accession_KT428890
TARA_CHLORO_00490
taxo_Cryptophyceae-Pyrenomonadales-Guillardia-theta-accession_NC-000926
taxo_Cryptophyceae-Pyrenomonadales-Chroomonas-mesostigmatica-accession_KY860574
TARA_CHLORO_00331
taxo_Cryptophyceae-Cryptomonadales-Cryptomonas-paramecium-accession_NC-013703
taxo_Cryptophyceae-Cryptomonadales-Cryptomonas-curvata-accession_NC-035720
TARA_CHLORO_00158
TARA_CHLORO_00478
TARA_CHLORO_00332
REFM_CHLORO_00001
taxo_Rhodophyta-Florideophyceae-Rhodymeniophycidae-Vertebrata-lanosa-accession_NC-026523
taxo_Rhodophyta-Florideophyceae-Rhodymeniophycidae-Rhodomelaceae-Thaumatella-adunca-accession_NC-035291
taxo_Rhodophyta-Florideophyceae-Rhodymeniophycidae-Spyridia-filamentosa-accession_NC-035285
taxo_Rhodophyta-Florideophyceae-Rhodymeniophycidae-Gracilaria-changii-accession_NC-038051
taxo_Rhodophyta-Florideophyceae-Rhodymeniophycidae-Crassiphycus-corneus-accession_NC-058658
taxo_Rhodophyta-Florideophyceae-Rhodymeniophycidae-Gelidium-coulteri-accession_NC-041173
taxo_Rhodophyta-Florideophyceae-Rhodymeniophycidae-Chondrus-crispus-accession_NC-020795
taxo_Rhodophyta-Florideophyceae-Corallinophycidae-Calliarthron-tuberculosum-accession_NC-021075
taxo_Rhodophyta-Florideophyceae-Nemaliophycidae-Liagoropsis-maxima-accession_NC-031662
taxo_Rhodophyta-Florideophyceae-Nemaliophycidae-Balbiania-investiens-accession_NC-042171
taxo_Rhodophyta-Bangiophyceae-Bangiales-Bangiaceae-Porphyra-purpurea-accession_NC-000925
taxo_Rhodophyta-Compsopogonophyceae-Compsopogonales-Compsopogon-caeruleus-accession_NC-035350
taxo_Rhodophyta-Compsopogonophyceae-Compsopogonales-Boldia-erythrosiphon-accession_NC-034776
taxo_Rhodophyta-Compsopogonophyceae-Erythropeltidales-Erythrotrichia-carnea-accession_NC-031176
taxo_Rhodophyta-Bangiophyceae-Porphyridiales-Porphyridium-sordidum-accession_NC-031175
taxo_Rhodophyta-Bangiophyceae-Porphyridiales-Flintiella-sanguinaria-accession_KY709211
taxo_Rhodophyta-Stylonematophyceae-Goniotrichopsis-reniformis-accession_NC-062382
taxo_Rhodophyta-Stylonematophyceae-Bangiopsis-subsimplex-accession_NC-031173
taxo_Rhodophyta-Rhodellophyceae-Dixoniellales-Bulboplastis-apyrenoidosa-accession_NC-034787
taxo_Rhodophyta-Rhodellophyceae-Rhodellales-Corynoplastis-japonica-accession_KY709210
taxo_Rhodophyta-Cyanidiales-Cyanidiaceae-Cyanidioschyzon-merolae-accession_NC-004799
taxo_Rhodophyta-Cyanidiales-Cyanidiaceae-Cyanidiococcus-yangmingshanensis-accession_NC-051883
taxo_Rhodophyta-Cyanidiales-Cyanidiaceae-Cyanidium-caldarium-accession_NC-001840
taxo_Rhodophyta-Cyanidiales-Galdieriaceae-Galdieria-sulphuraria-accession_KJ700459
taxo_Alveolata-Dinophyceae-sp-MGD-accession_LC716140
TARA_CHLORO_00597
TARA_CHLORO_00662
TARA_CHLORO_00658
TARA_CHLORO_00036
TARA_CHLORO_00008
TARA_CHLORO_00661
TARA_CHLORO_00660
TARA_CHLORO_00556
TARA_CHLORO_00545
taxo_Alveolata-Dinophyceae-sp-TGD-accession_LC716139
TARA_CHLORO_00551
TARA_CHLORO_00546
TARA_CHLORO_00547
TARA_CHLORO_00014
taxo_Alveolata-Dinophyceae-Gymnodiniales-Lepidodinium-chlorophorum-accession_NC-027093
taxo_Viridiplantae-Chlorophyta-Pedinophyceae-Pedinomonas-minor-accession_NC-016733
taxo_Viridiplantae-Chlorophyta-Trebouxiophyceae-Chlorella-sorokiniana-accession_NC-023835
taxo_Viridiplantae-Chlorophyta-Trebouxiophyceae-Chlorella-variabilis-accession_NC-015359
taxo_Viridiplantae-Chlorophyta-Trebouxiophyceae-Marvania-geminata-accession_NC-025549
taxo_Viridiplantae-Chlorophyta-Trebouxiophyceae-Coccomyxa-subellipsoidea-C-169-accession_NC-015084
TARA_CHLORO_00400
taxo_Viridiplantae-Chlorophyta-Trebouxiophyceae-Lobosphaera-incisa-accession_NC-025533
taxo_Viridiplantae-Chlorophyta-Chlorophyceae-Neocystis-brevis-accession_NC-025535
taxo_Rhizaria-Cercozoa-Chlorarachniophyceae-Bigelowiella-natans-accession_NC-008408
TARA_CHLORO_00339
taxo_Rhizaria-Cercozoa-Chlorarachniophyceae-Lotharella-vacuolata-accession_NC-029743
TARA_CHLORO_00657
taxo_Viridiplantae-Chlorophyta-Ulvophyceae-Trentepohlia-sp-BN17-accession_NC-060532
taxo_Viridiplantae-Chlorophyta-Ulvophyceae-Trentepohlia-odorata-accession_NC-043776
taxo_Viridiplantae-Chlorophyta-Ulvophyceae-Caulerpa-racemosa-accession_NC-032042
taxo_Viridiplantae-Chlorophyta-Ulvophyceae-Ostreobium-quekettii-accession_NC-030629
taxo_Viridiplantae-Chlorophyta-Ulvophyceae-Ulva-sp-UNA00071828-accession_KP720616
taxo_Viridiplantae-Chlorophyta-Ulvophyceae-Neodangemannia-microcystis-accession_NC-034713
TARA_CHLORO_00343
TARA_CHLORO_00492
TARA_CHLORO_00382
TARA_CHLORO_00647
taxo_Viridiplantae-Chlorophyta-Chlorodendrophyceae-Scherffelia-dubia-accession_NC-029807
taxo_Viridiplantae-Chlorophyta-Chlorophyceae-Treubaria-triappendiculata-accession_NC-028578
taxo_Viridiplantae-Chlorophyta-Chlorophyceae-Parallela-transversalis-accession_NC-042241
taxo_Viridiplantae-Chlorophyta-Chlorophyceae-Ankyra-judayi-accession_NC-029735
taxo_Viridiplantae-Chlorophyta-Chlorophyceae-Bracteacoccus-minor-accession_NC-029674
taxo_Viridiplantae-Chlorophyta-Chlorophyceae-Tetradesmus-obliquus-accession_NC-008101
taxo_Viridiplantae-Chlorophyta-Chlorophyceae-Chlamydomonas-reinhardtii-accession_NC-005353
taxo_Viridiplantae-Chlorophyta-Chlorophyceae-Carteria-cerasiformis-accession_NC-028585
taxo_Viridiplantae-Chlorophyta-Chlorophyceae-Oedogonium-cardiacum-accession_NC-011031
taxo_Viridiplantae-Chlorophyta-Chlorophyceae-Schizomeris-leibleinii-accession_NC-015645
TARA_CHLORO_00118
TARA_CHLORO_00430
taxo_Viridiplantae-Chlorophyta-Chloropicophyceae-Chloroparvula-japonica-accession_NC-042487
TARA_CHLORO_00357
TARA_CHLORO_00590
TARA_CHLORO_00358
TARA_CHLORO_00102
TARA_CHLORO_00002
TARA_CHLORO_00375
TARA_CHLORO_00386
taxo_Viridiplantae-Chlorophyta-Chloropicophyceae-Chloropicon-laureae-accession_NC-042484
TARA_CHLORO_00570
taxo_Viridiplantae-Chlorophyta-Chloropicophyceae-Chloropicon-sp-RCC4434-accession_MK085997
TARA_CHLORO_00602
TARA_CHLORO_00433
TARA_CHLORO_00527
TARA_CHLORO_00434
TARA_CHLORO_00064
TARA_CHLORO_00432
TARA_CHLORO_00601
TARA_CHLORO_00603
TARA_CHLORO_00431
taxo_Viridiplantae-Chlorophyta-Pycnococcaceae-Pseudoscourfieldia-marina-accession_NC-012097
taxo_Viridiplantae-Chlorophyta-Nephroselmidophyceae-Nephroselmis-pyriformis-accession_NC-059722
TARA_CHLORO_00704
TARA_CHLORO_00237
TARA_CHLORO_00514
TARA_CHLORO_00653
taxo_Viridiplantae-Chlorophyta-Mamiellophyceae-Micromonas-commoda-accession_NC-012575
TARA_CHLORO_00315
TARA_CHLORO_00227
TARA_CHLORO_00244
TARA_CHLORO_00389
TARA_CHLORO_00124
TARA_CHLORO_00390
TARA_CHLORO_00264
TARA_CHLORO_00246
TARA_CHLORO_00015
taxo_Viridiplantae-Chlorophyta-Mamiellophyceae-Microrhizoidea-picketheapsiorum-accession_NC-044704
TARA_CHLORO_00314
TARA_CHLORO_00238
TARA_CHLORO_00104
TARA_CHLORO_00271
TARA_CHLORO_00105
TARA_CHLORO_00272
TARA_CHLORO_00703
TARA_CHLORO_00594
taxo_Viridiplantae-Chlorophyta-Pyramimonadophyceae-Cymbomonas-tetramitiformis-accession_NC-030169
taxo_Euglenozoa-Euglenales-Euglena-viridis-accession_NC-020460
taxo_Euglenozoa-Euglenales-Phacus-inflexus-accession_NC-039967
TARA_CHLORO_00651
taxo_Euglenozoa-Eutreptiales-Eutreptiella-gymnastica-accession_NC-017754
TARA_CHLORO_00426
TARA_CHLORO_00624
TARA_CHLORO_00312
TARA_CHLORO_00250
taxo_Eukaryota-Viridiplantae-Prasinodermophyta-prasinophyte-sp-MBIC10622-accession_KJ746602
TARA_CHLORO_00520
taxo_Viridiplantae-Prasinodermophyta-Prasinoderma-coloniale-accession_NC-024817
TARA_CHLORO_00167
taxo_Viridiplantae-Prasinodermophyta-Palmophyllophyceae-Prasinococcus-capsulatus-accession_KJ746597
TARA_CHLORO_00654
taxo_Viridiplantae-Streptophyta-Embryophyta-Arabidopsis-thaliana-accession_NC-000932
taxo_Viridiplantae-Streptophyta-Embryophyta-Oryza-sativa-accession_NC-031333
taxo_Viridiplantae-Streptophyta-Embryophyta-Pinus-pinaster-accession_FJ899583
taxo_Viridiplantae-Streptophyta-Embryophyta-Marchantia-paleacea-accession_NC-001319
taxo_Viridiplantae-Streptophyta-Zygnemophyceae-Zygnema-circumcarinatum-accession_NC-008117
taxo_Viridiplantae-Streptophyta-Charophyceae-Chara-vulgaris-accession_NC-008097
taxo_Viridiplantae-Streptophyta-Klebsormidiophyceae-Klebsormidium-flaccidum-accession_NC-024167
taxo_Viridiplantae-Streptophyta-Mesostigmatophyceae-Mesostigma-viride-accession_NC-002186
taxo_Glaucocystophyceae-Glaucocystis-incrassata-accession_NC-040152
taxo_Glaucocystophyceae-Glaucocystis-sp-BBH-accession_MF167424
taxo_Glaucocystophyceae-Gloeochaete-wittrockiana-accession_NC-040153
taxo_Glaucocystophyceae-Cyanoptyche-gloeocystis-accession_MF167427
taxo_Glaucocystophyceae-Cyanophora-sudae-accession_NC-038215
taxo_Glaucocystophyceae-Cyanophora-biloba-accession_NC-038216
taxo_Glaucocystophyceae-Cyanophora-paradoxa-accession_NC-001675
taxo_Cyanobacteriota-Synechococcales-Prochlorococcus-marinus-accession_NC-005071
taxo_Cyanobacteriota-Synechococcales-Prochlorococcus-marinus-accession_NC-005042
taxo_Cyanobacteriota-Synechococcales-Parasynechococcus-marenigrum-accession_BX548020
taxo_Rhizaria-Cercozoa-Imbricatea-Silicofilosea-Euglyphida-Paulinella-micropora-accession_NC-039737
taxo_Cyanobacteriota-Gloeobacterales-Gloeobacter-kilaueensis-accession_NC-022600
taxo_Cyanobacteriota-Gloeobacterales-Gloeobacter-violaceus-accession_NC-005125
taxo_Cyanobacteriota-Synechococcales-Synechococcus-sp-accession_CP000239
taxo_Cyanobacteriota-Acaryochloridales-Acaryochloris-marina-accession_NC-009925
taxo_Cyanobacteriota-Leptolyngbyales-Leptolyngbya-boryana-accession_NZ-AP014638
taxo_Cyanobacteriota-Nostocales-Mastigocladopsis-repens-accession_GCF-000315565
taxo_Cyanobacteriota-Oscillatoriales-Limnospira-maxima-accession_GCF-000173555
taxo_Cyanobacteriota-Oscillatoriales-Oscillatoria-acuminata-accession_NC-019693
taxo_Cyanobacteriota-Gloeomargaritales-cyanobacterium-VI4D9-accession_CP088016
taxo_Cyanobacteriota-Gloeomargaritales-Gloeomargarita-lithophora-Alchichica-accession_CP017675
0.5

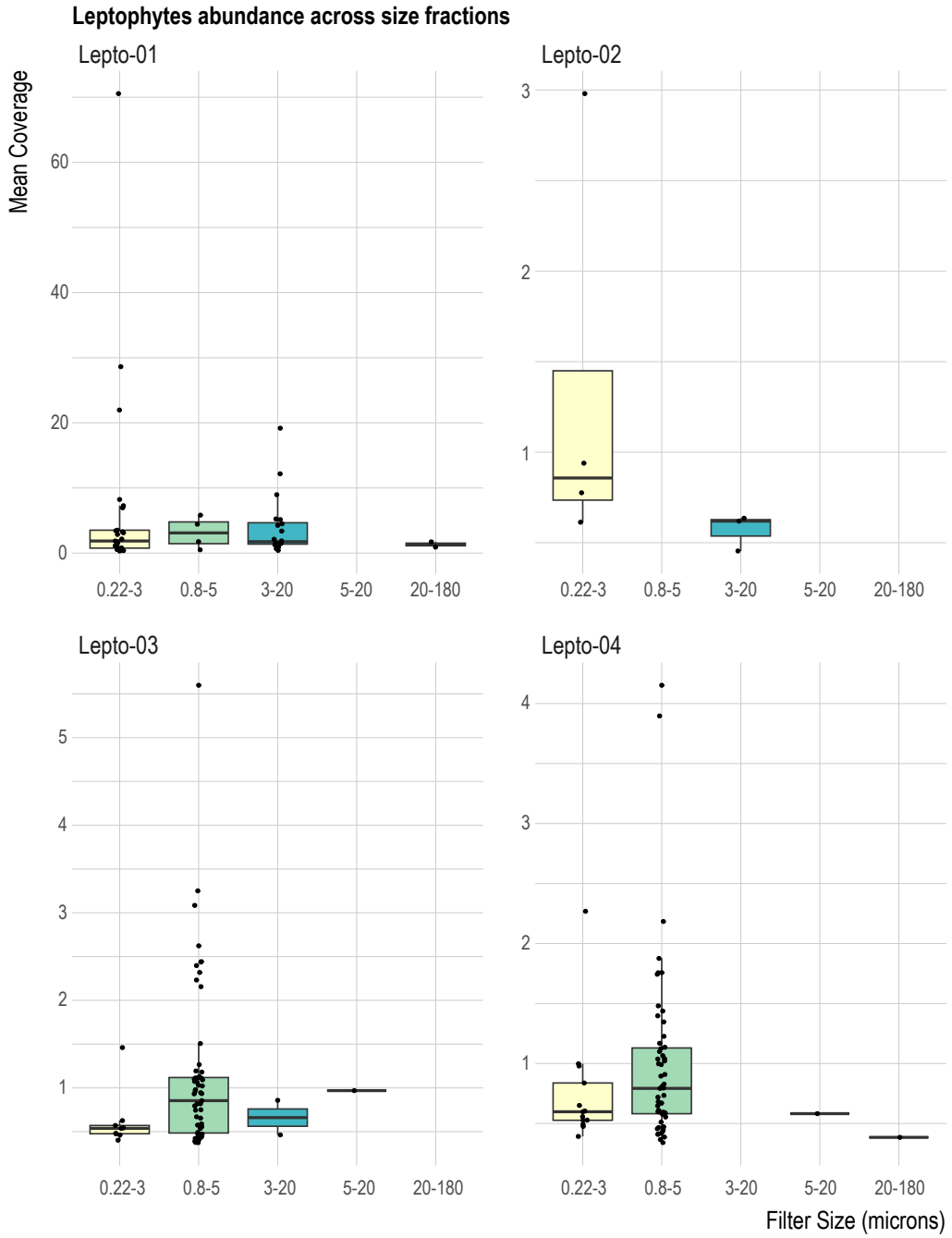
Supplementary Fig. 7. (On previous page). Maximum-likelihood phylogeny of 179 selected reference genomes and 660 ptMAGs based on 93 plastid-encoded genes. The tree was run under the LG+F+I+G4 model, and branch support values were calculated with 1000 ultrafast bootstrap replicates.



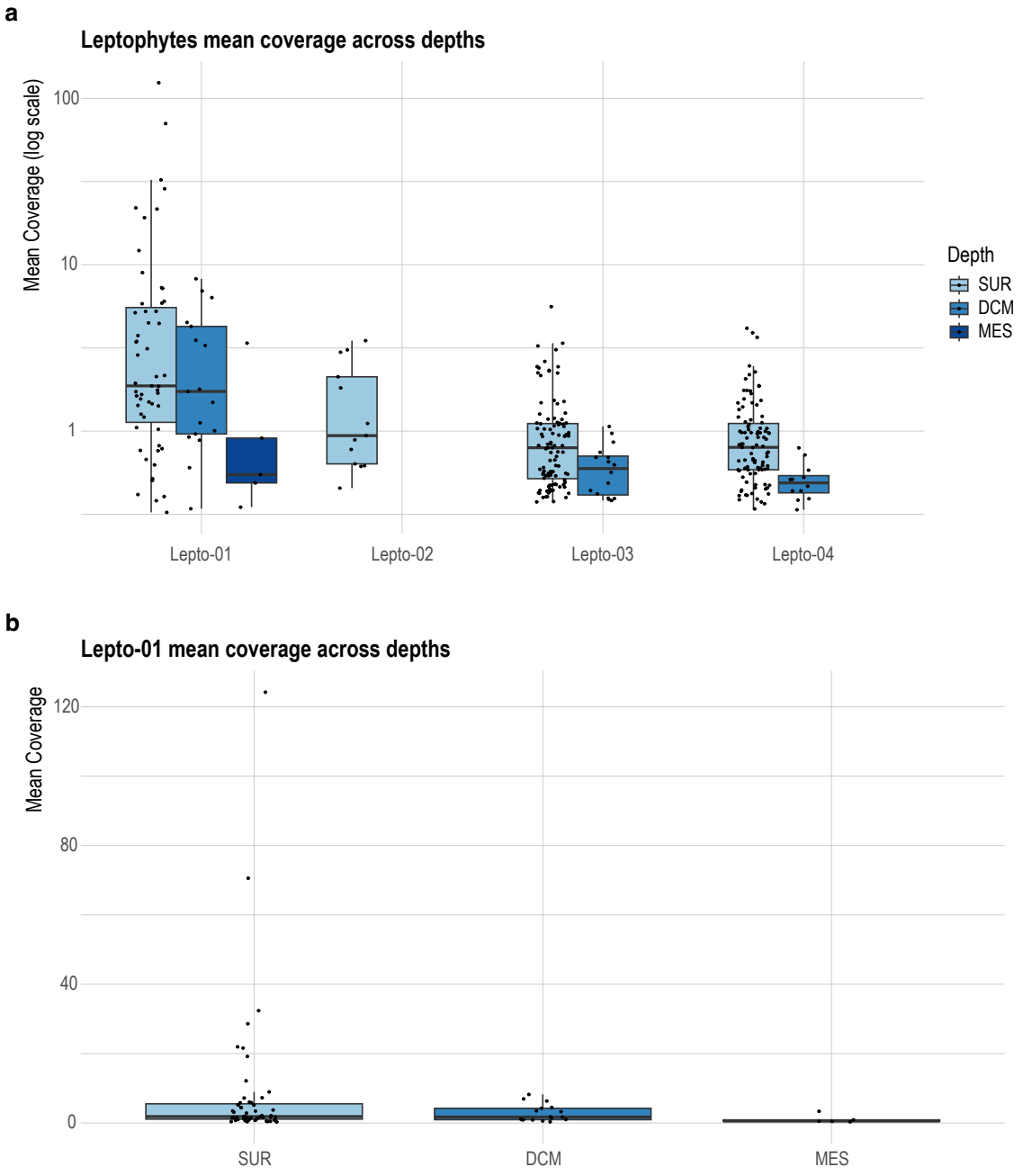
Supplementary Fig. 8. Two examples of deep-branching novel diversity in algal groups detected by ptMAGs. Panel (a) shows a clade of eight ptMAGs that is related to Chloropicophyceae. Panel (b) shows a novel clade of plastid genomes that is related to, but distinct from known pelagophyte references.


a
Tree
Completeness (%)
Size (kbp)
No. of Genes
Genomic composition (%)
TARA_CHLORO_00008
TARA_CHLORO_00036
TARA_CHLORO_00661
TARA_CHLORO_00660
TARA_CHLORO_00597
MGD
TARA_CHLORO_00662
TARA_CHLORO_00658
TARA_CHLORO_00551
TGD
TARA_CHLORO_00545
TARA_CHLORO_00556
TARA_CHLORO_00547
TARA_CHLORO_00546
TARA_CHLORO_00014
Lepidodinium chlorophorum
Pedinomonas minor
0 25 50 75 100
0 200 400 600
0 50 100 150
0 25 50 75 100
region_type
intergenic
genic
b
TARA_CHLORO_00662
627,879 bp
C_0
51,775 bp
C_1
153,406 bp
C_2
30,651 bp
C_3
154,069 bp
C_4
49,978 bp
C_5
46,404 bp
C_6
24,484 bp
C_7
52,743 bp
C_8
9,434 bp
C_9
25,939 bp
C_10
28,996 bp
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
photosystem assembly/stability factors
RNA polymerase
polycistronic transcripts
ribosomal proteins (SSU)
ribosomal proteins (LSU)
transfer RNAs
ribosomal RNAs
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
ORFs
origin of replication

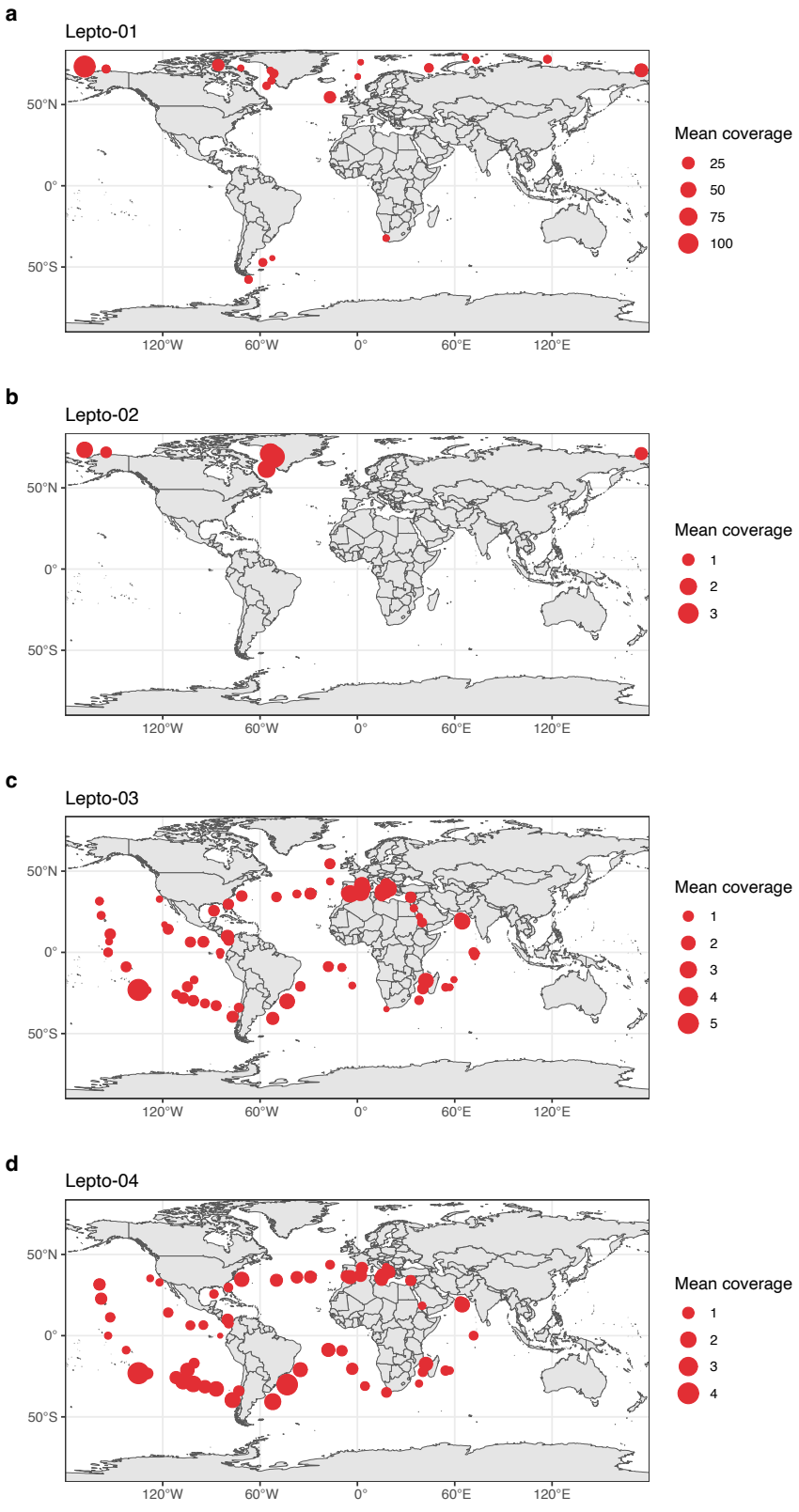
Supplementary Fig. 9. (on previous page) **(a)** A clade of 13 ptMAGs related to green algal endosymbionts of dinoflagellates, showing large variations in size, with the largest genome (TARA_CHLORO_00662) being around 600 kbp. This variation in genome size is accompanied by an inflation in intergenic regions in the plastid genome. Panel **(b)** visualises the plastome organisation (including large intergenic regions) in the largest ptMAG. Contigs are not drawn to scale and size is indicated.



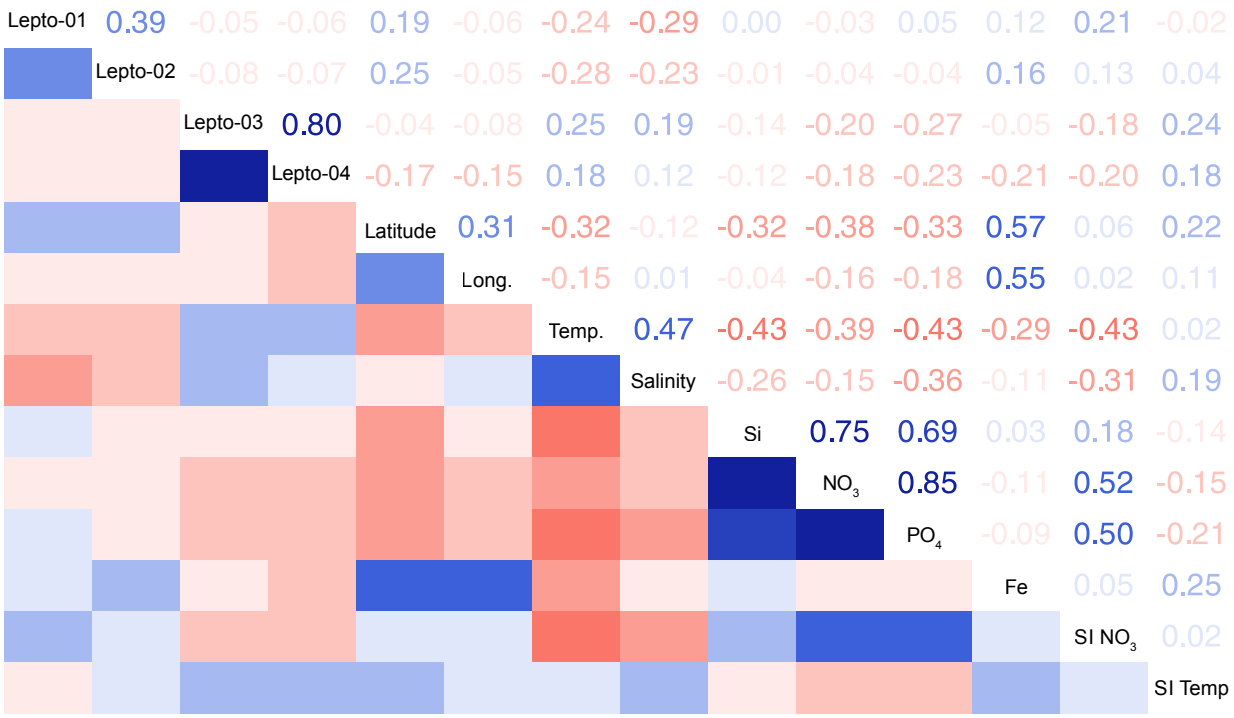
Supplementary Fig. 10. Signal of leptophyte plastid genomes in *Tara* Ocean samples from different size fractions. Leptophytes are detected most frequently in the 0.22-3 μm, 0.8-5 μm, and the 3-20 μm size fractions, and are barely detected in the 5-20 μm and 20-180 μm size fractions, indicating that leptophytes are small algae that are less than 5 μm in size. Source data and code provided in linked GitHub repository[1].



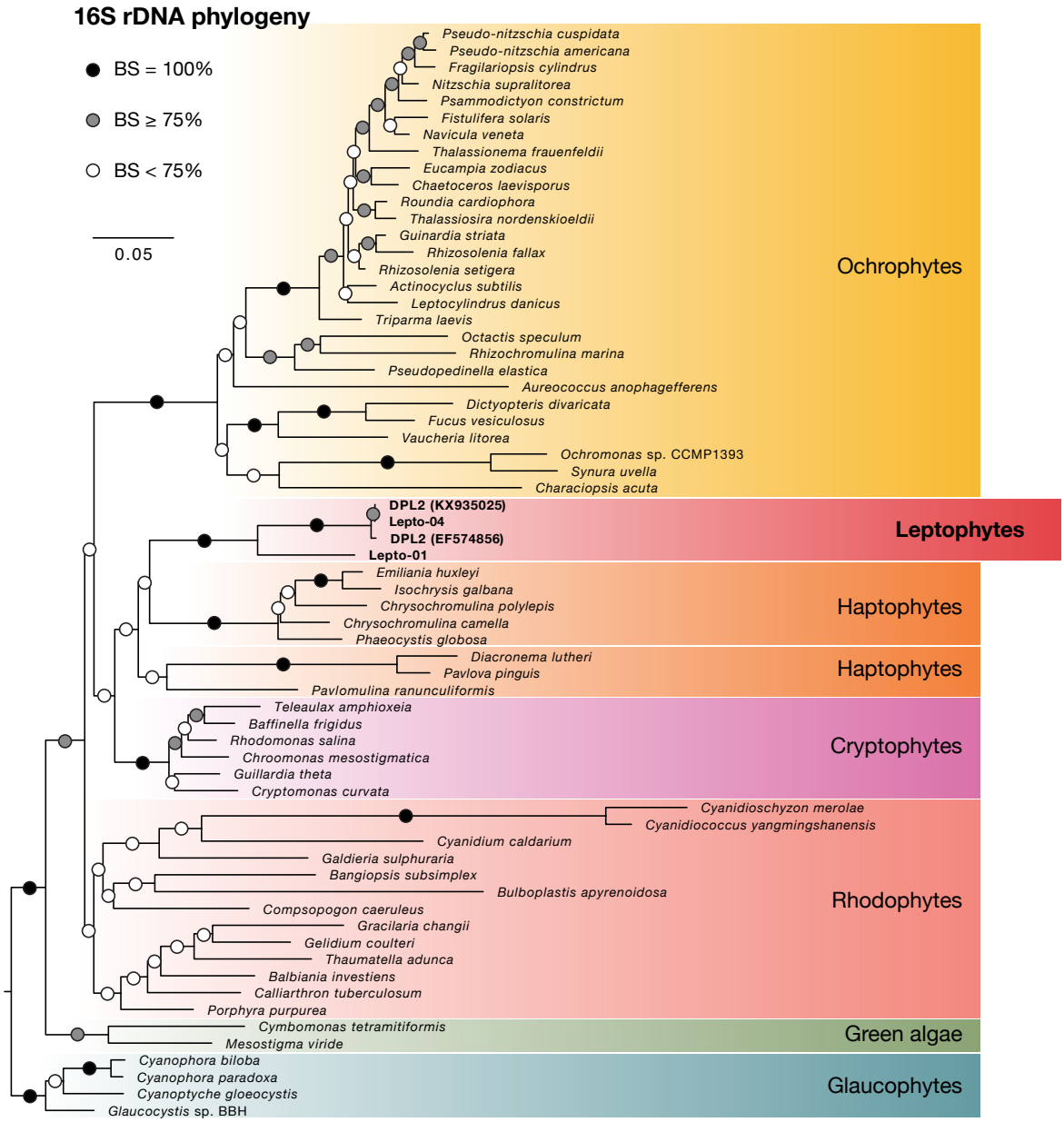
Supplementary Fig. 11. Signal of leptophyte plastid genomes in *Tara* Ocean samples from different depths. Panel **(a)** depicts all four leptophyte ptMAGs while panel **(b)** zooms in on the most abundant ptMAG (Lepto-01). Data points represent mean coverage values from individual samples. Leptophytes are predominantly detected in the surface layer, detected at a lesser abundance in the DCM layer, and are barely detected in the mesopelagic layer. Source data and code provided in linked GitHub repository[1].



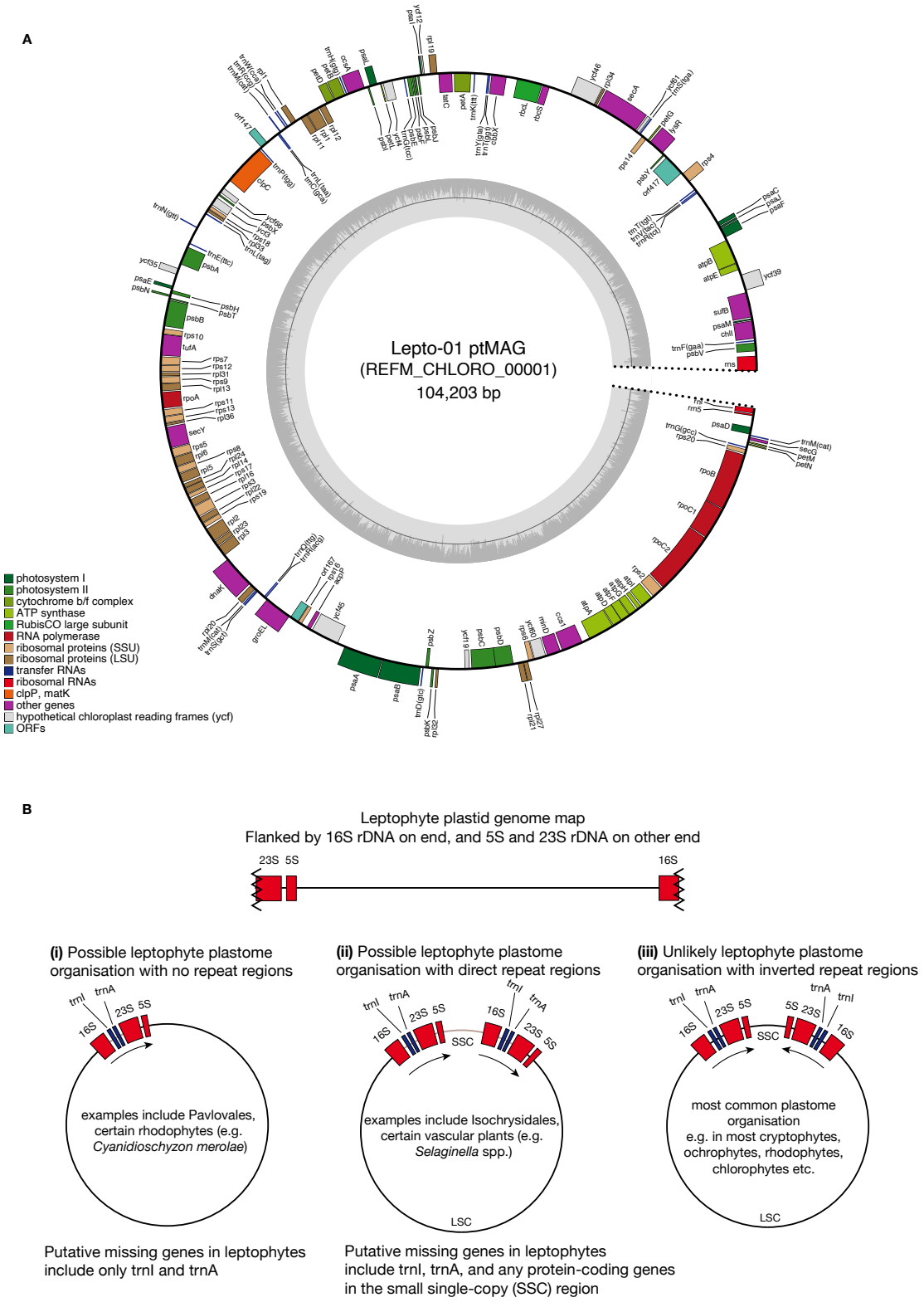
Supplementary Fig. 12. Distribution of leptophyte plastid genomes as determined by mapping of 937 *Tara* Ocean metagenomes against each genome. Lepto-01 and Lepto-02 display an Arctic distribution (panels **a** and **b** respectively), while Lepto-03 and Lepto-04 ptMAGs have a more tropical and subtropical distribution (panels **c** and **d** respectively). The overall low mean coverage indicates that leptophytes are rare in general, with the exception of several stations where Lepto-01 is abundant. Source data and code provided in linked GitHub repository[1].



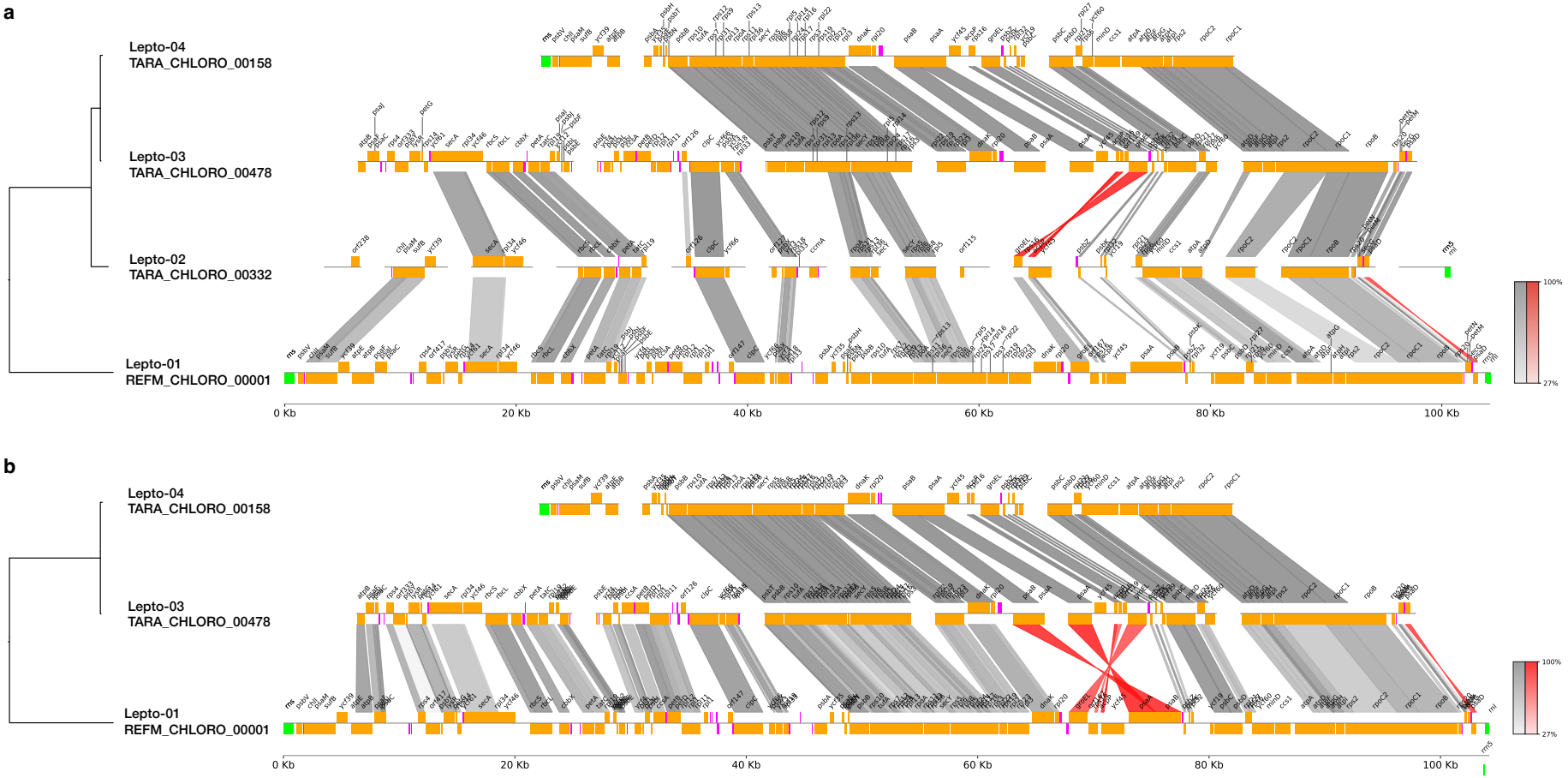
Supplementary Fig. 13. Pearson correlations between the abundance of leptophyte plastids and environmental parameters. SI NO_3 and SI Temp represent the seasonality indices of nitrate and sea surface temperature respectively. These were defined as the range of the nitrate and temperature in one grid cell divided by the total range of that variable across all *Tara* sampling stations.



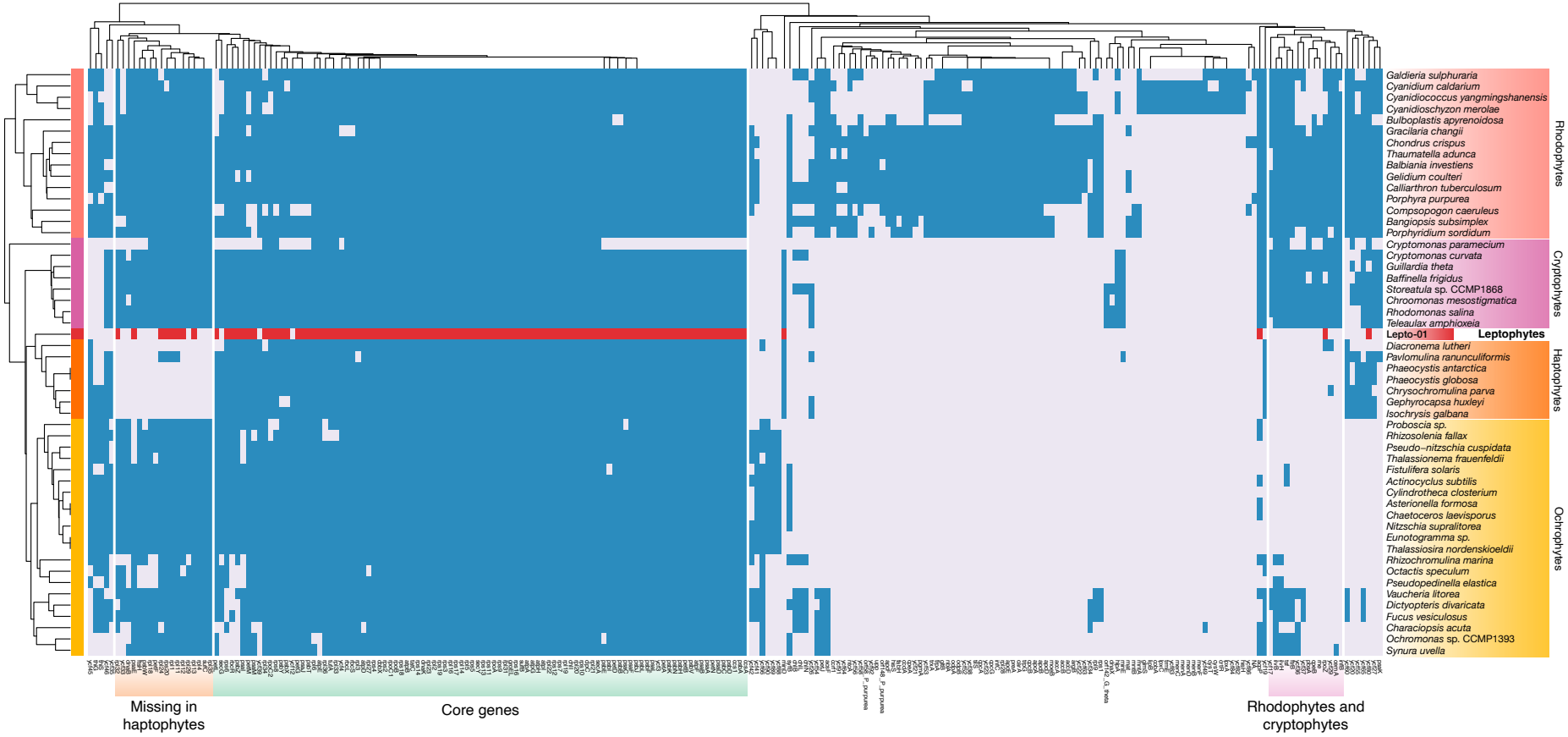
Supplementary Fig. 14. Maximum Likelihood phylogeny inferred with an alignment of the plastid 16S rDNA gene (65 taxa, 1499 sites). The phylogeny was inferred with raxml-ng using the GTR+G model and branch support assessed with 100 non-parametric bootstrap replicates.



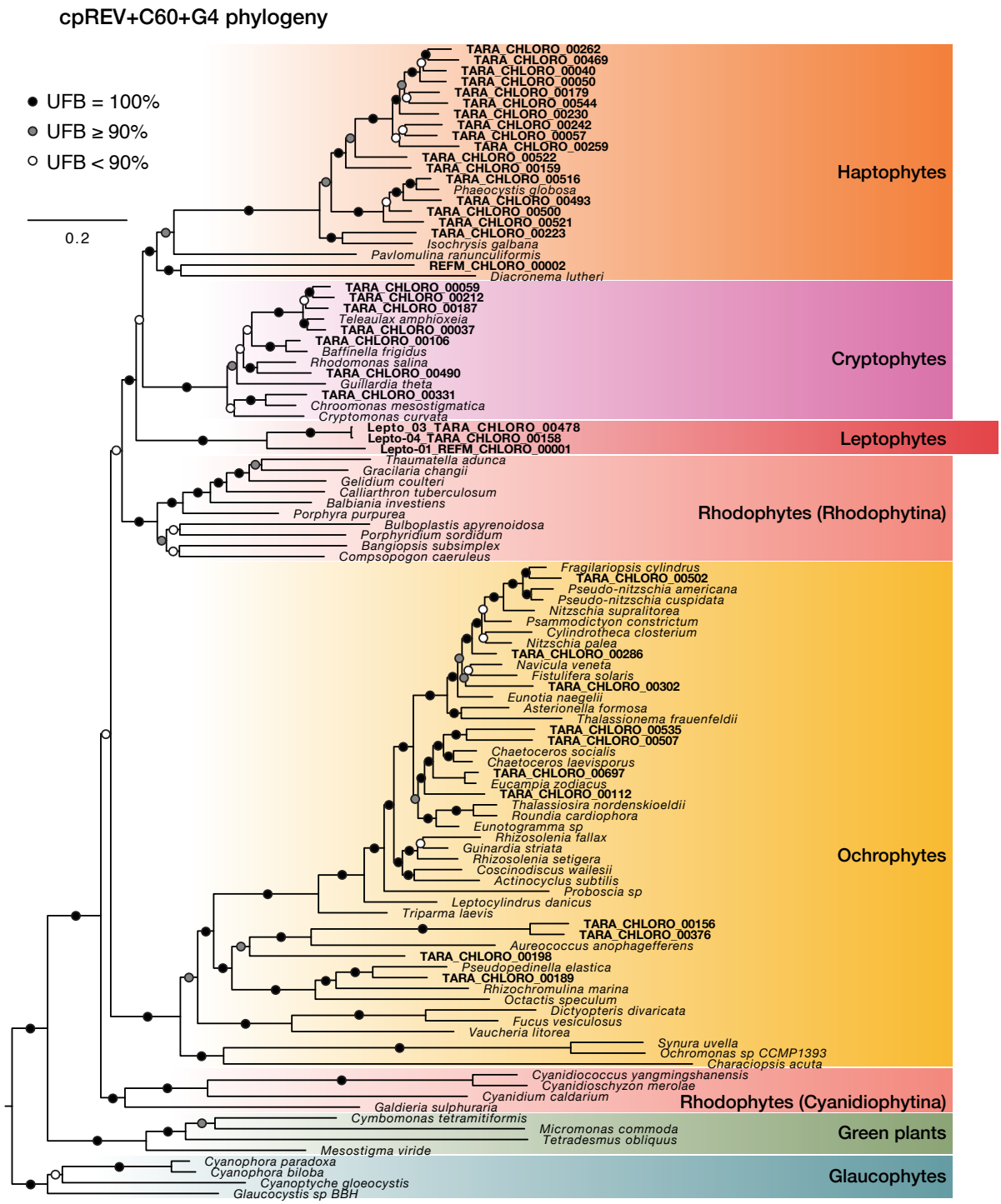
Supplementary Fig. 15. **(A)** Annotated map of the Lepto-01 plastid genome. Genes drawn inside the outer circle are transcribed clockwise, while those drawn outside the outer circle are transcribed counterclockwise. Genes are colour coded based on functional groups. The dark shaded area in the inner circle indicates GC content. The genome map was generated by OGDRAW[2]. **(B)** Cartoons of possible plastome organisation of leptophytes based on the available leptophyte plastid genome map.



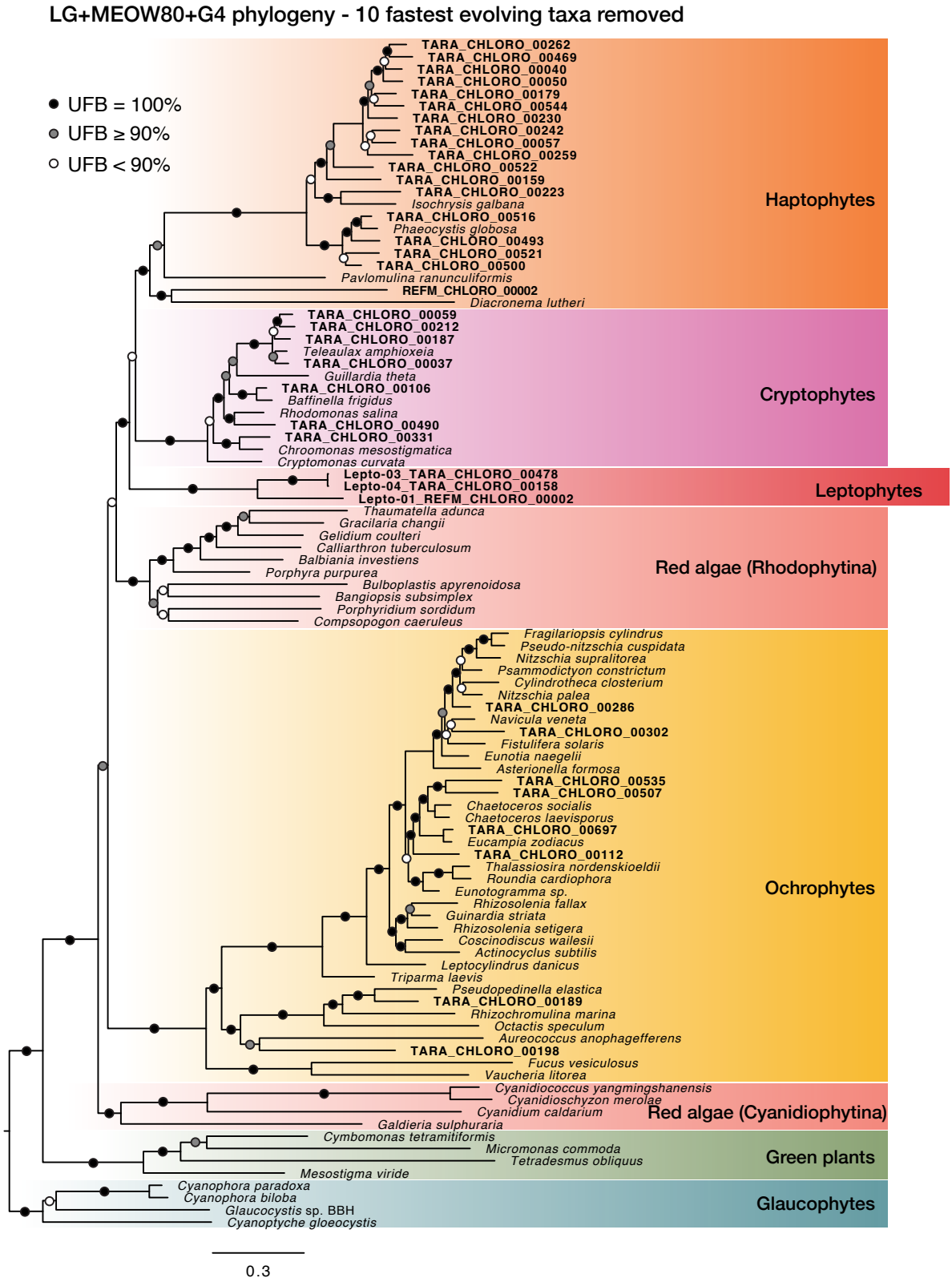
Supplementary Fig. 16. Synteny comparisons of the leptophyte ptMAGs. Panel (**a**) displays all leptophyte ptMAGs, and panel (**b**) displays the three most complete ptMAGs. Genomes were visualized using PyGenomeViz and sequence similarity was calculated using MMSEQs for reciprocal best-hit CDS search[3]. A cartoon phylogeny on the left depicts the relationships between these plastid genomes. Protein coding genes are coloured orange and labelled, while tRNA sequences are coloured magenta and not labelled for clarity. Overall, the leptophyte plastid genomes are very similar in gene content and synteny with the exception of two differences: (1) The *psaD* gene is inverted in Lepto-01 compared to the rest; and (2) A nearly 10 kbp region spanning from the *psaB* gene to the *groEL* gene is inverted in the Lepto-03 and Lepto-04 lineage.



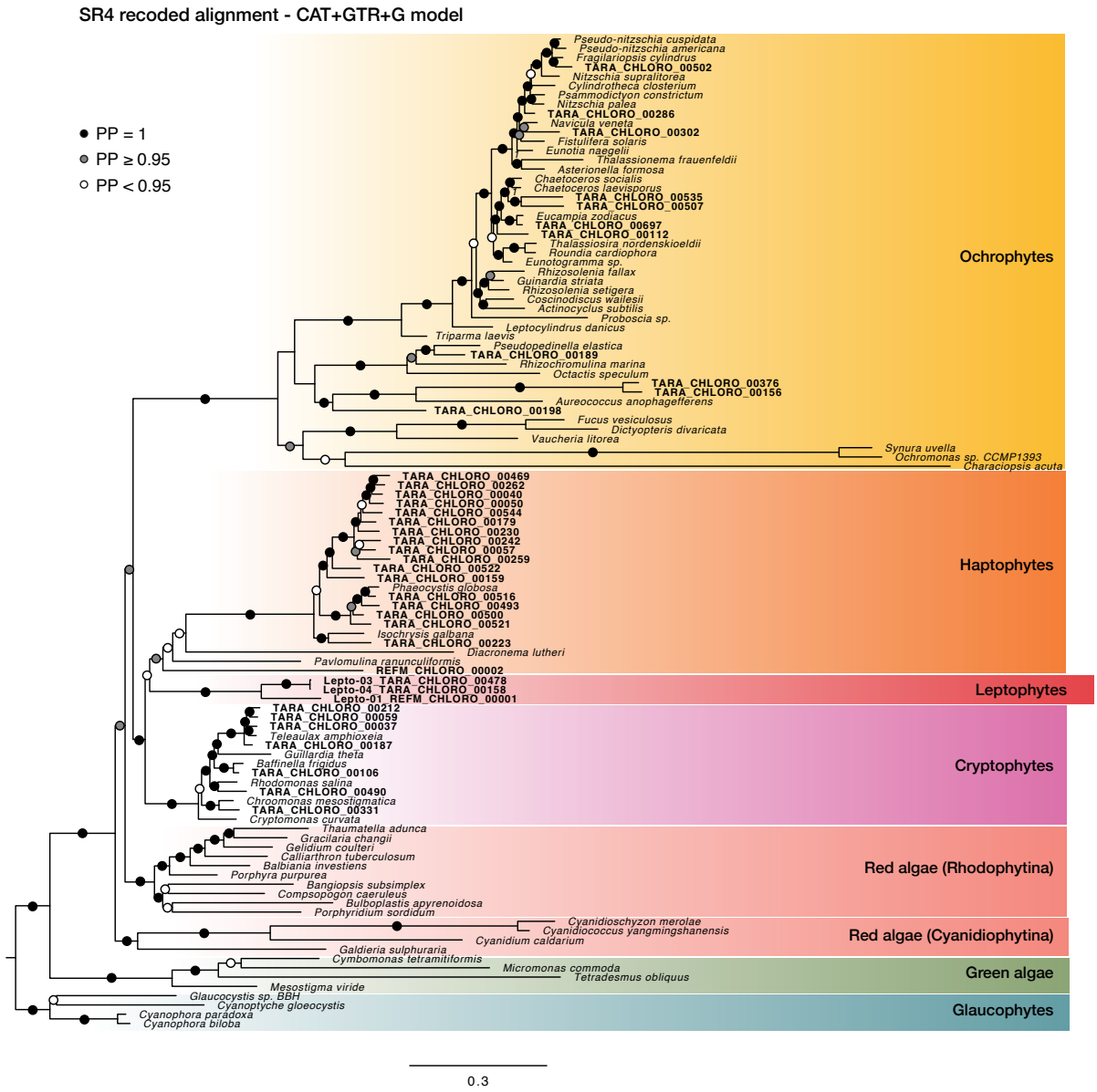
Supplementary Fig. 17. Binary heatmap and dendrograms of genes and taxa generated with UPGMA clustering, based on gene presence/absence of 237 plastid protein-coding genes. Blue/red and grey boxes represent presence and absence respectively. Source data provided in Supplementary Data 5.



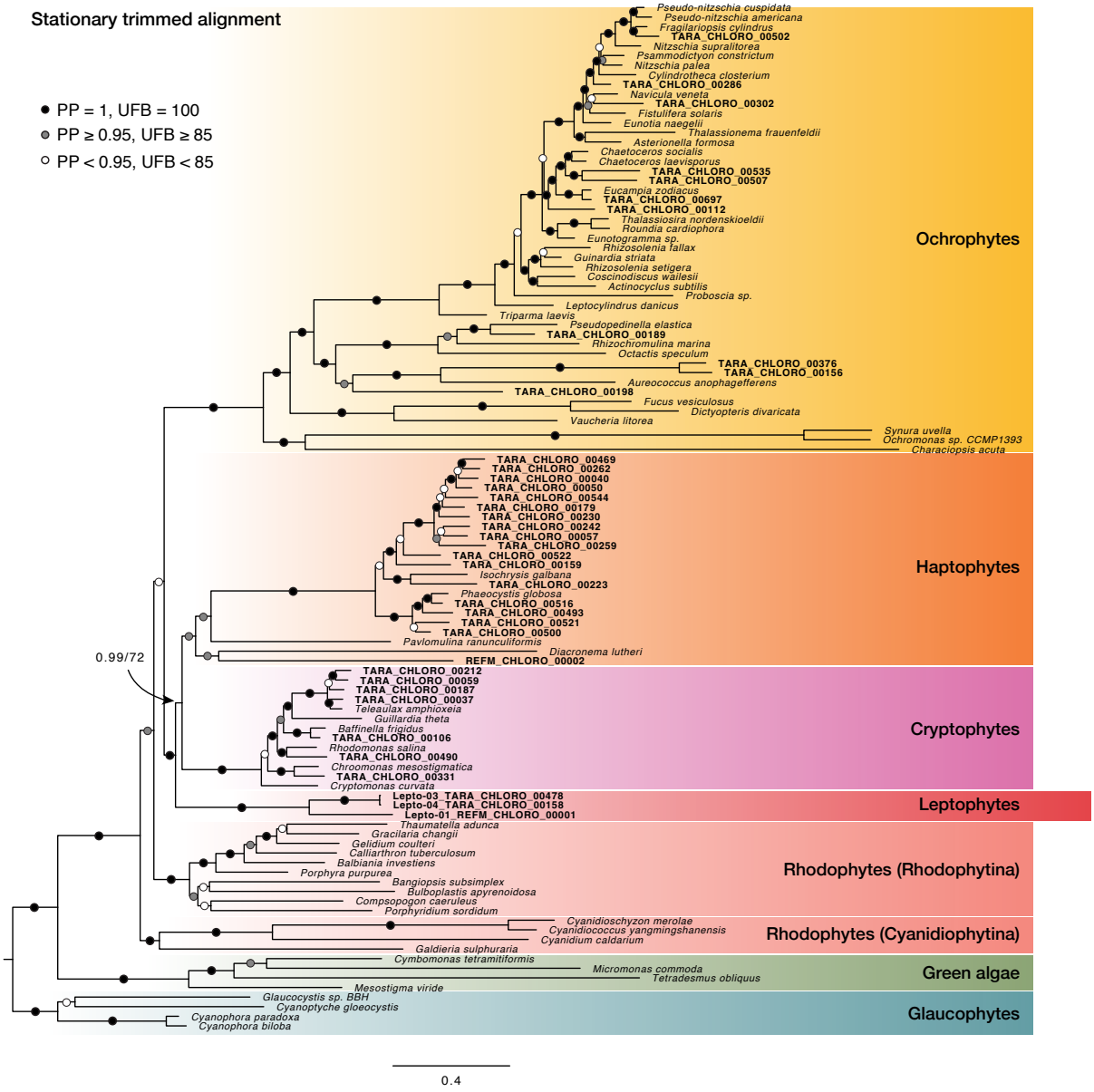
Supplementary Fig. 18. Maximum Likelihood phylogeny inferred with an alignment of 93 plastid-encoded genes (107 taxa, 20,292 sites). The phylogeny was inferred under the cpREV+C60+G4 model along with 1,000 ultrafast bootstraps.



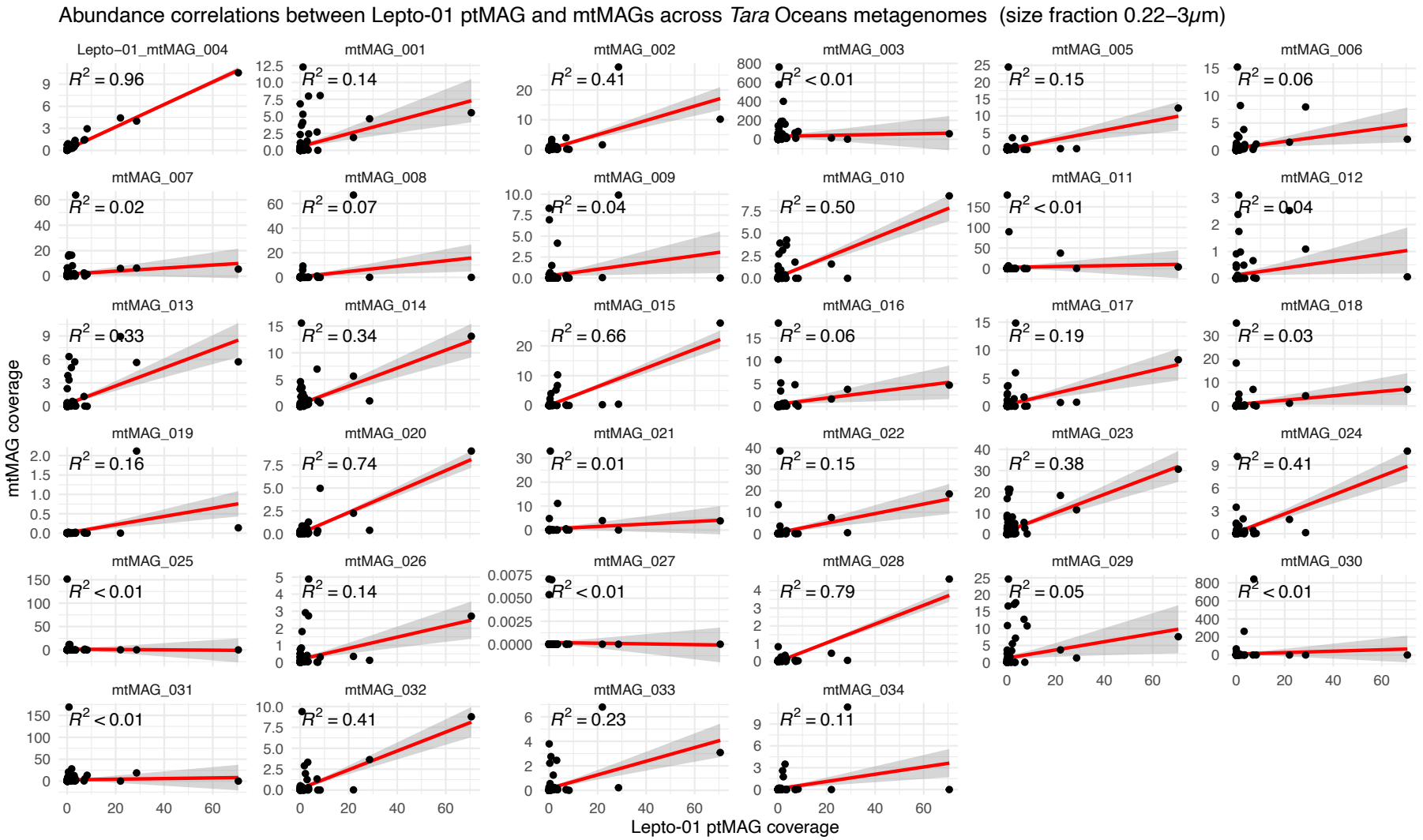
Supplementary Fig. 19. Maximum Likelihood phylogeny inferred with an alignment excluding the 10 fastest evolving taxa (longest root to tip distances). The input alignments included 93 plastid-encoded genes, 97 taxa, and 20,292 sites. The phylogeny was inferred under the LG+MEOW80+G4 model along with 1,000 ultrafast bootstraps.



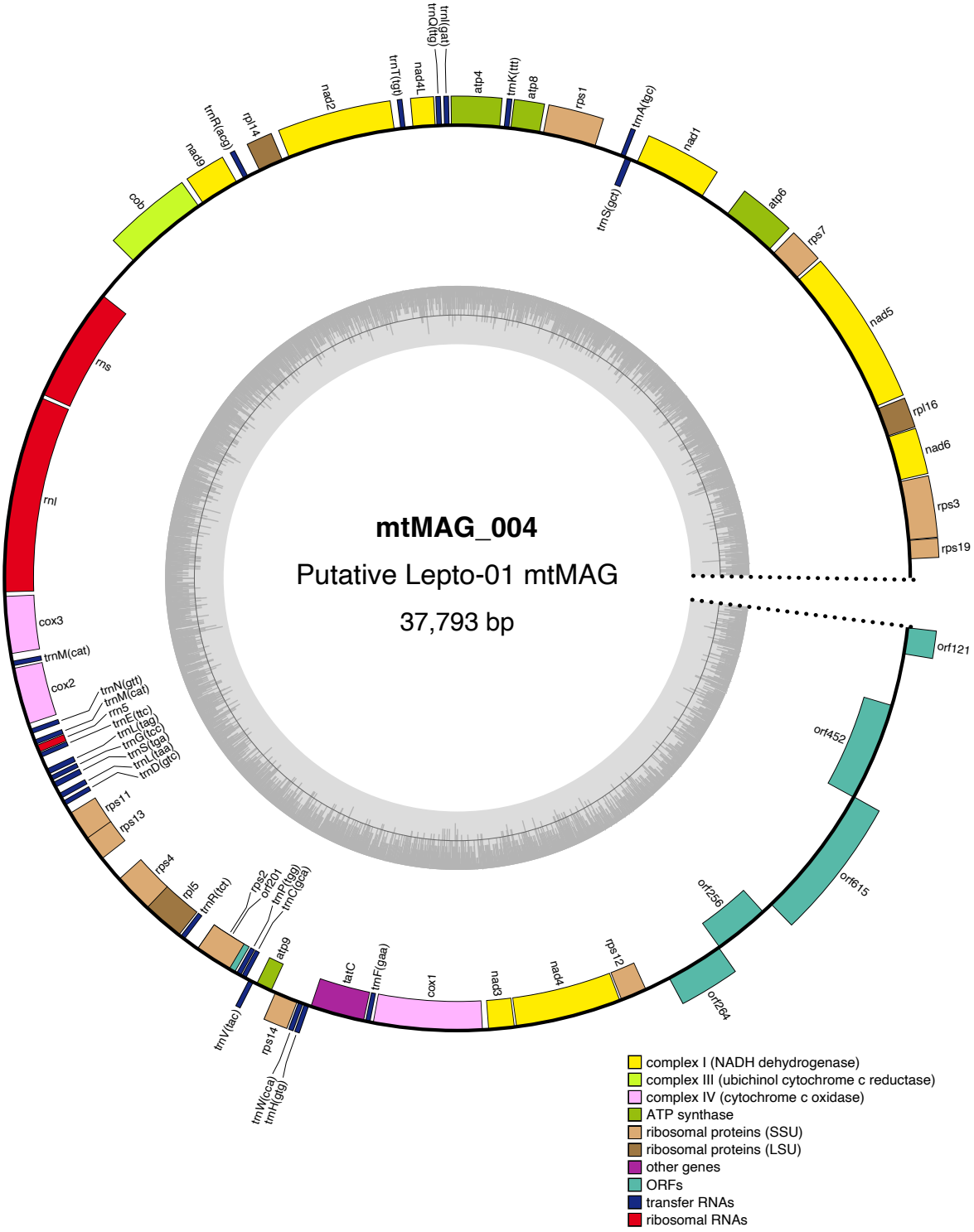
Supplementary Fig. 20. Bayesian inference of the SR4 recoded alignment with the CAT+GTR+G model (107 taxa, 20,292 sites).



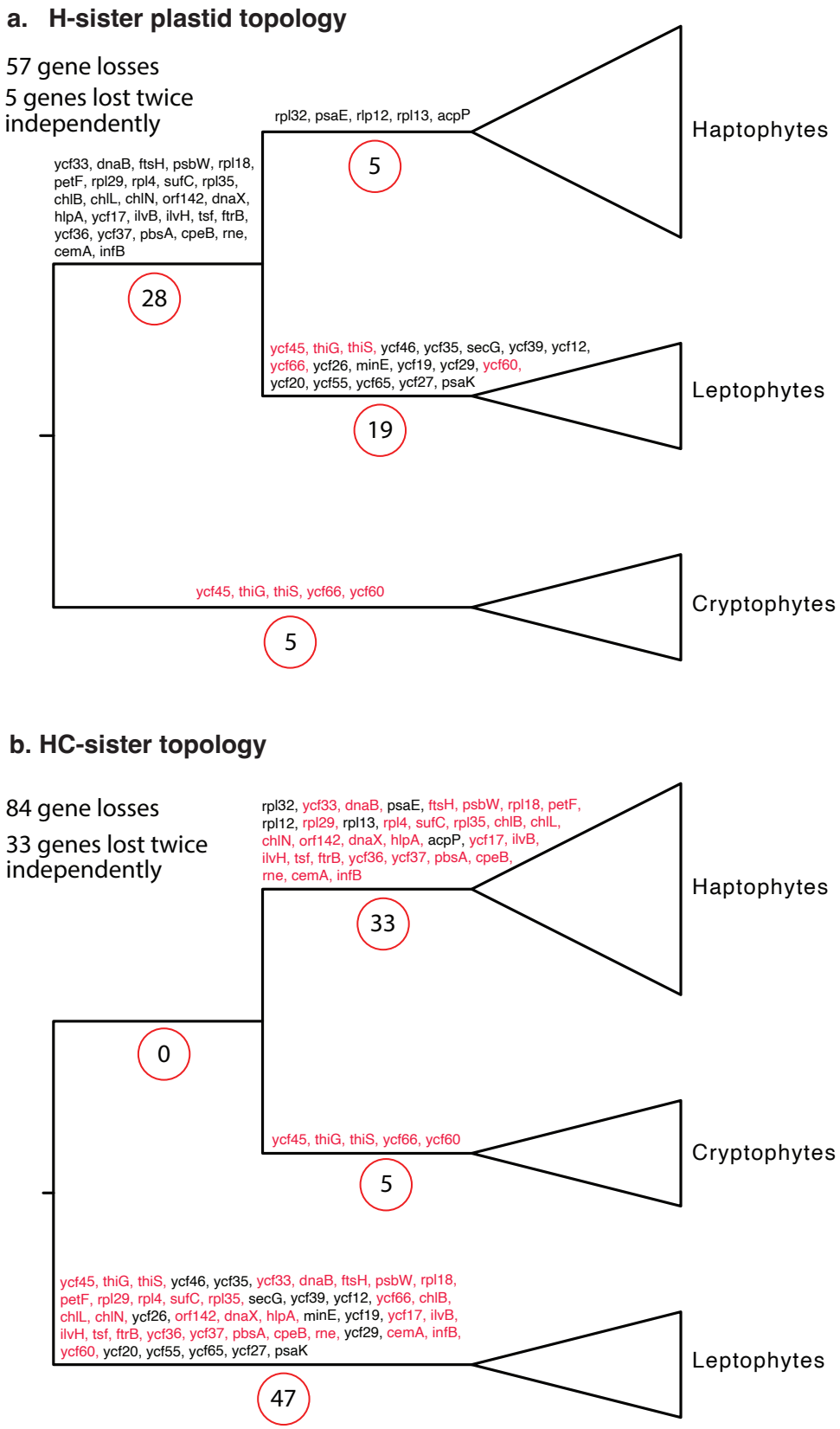
Supplementary Fig. 21. Bayesian inference of the stationary trimmed alignment to remove compositionally heterogenous sites (107 taxa, 15,754 sites). Branch support values are listed in the following order: posterior probabilities under the CAT+GTR+G model, and ultrafast bootstrap support under the LG+MEOW80+G model.



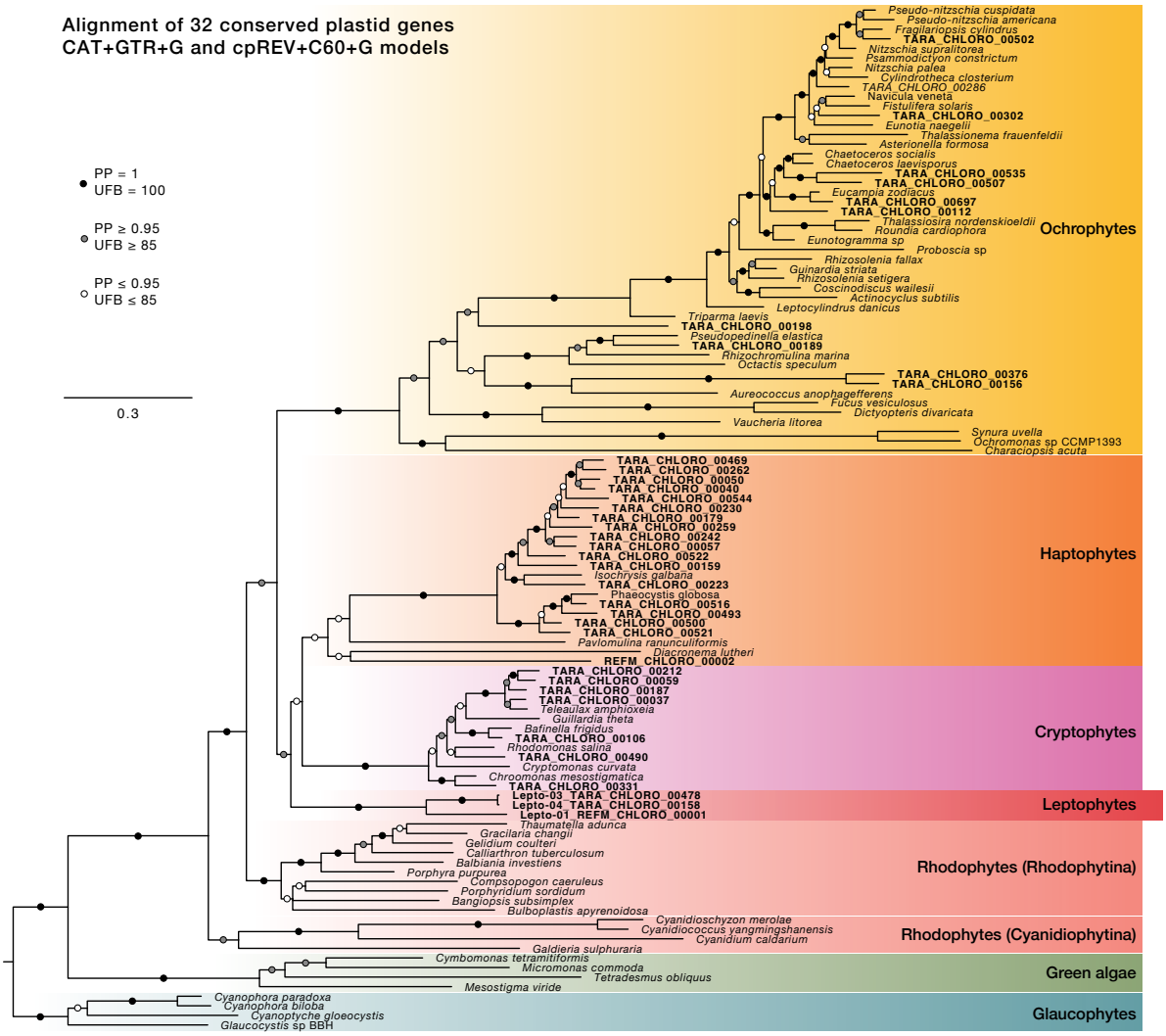
Supplementary Fig. 22. Abundance correlations between the Lepto-1 ptMAG and all 34 mtMAGs in Figure 4 across *Tara* Ocean metagenomes. Source data and code are provided on the associated GitHub repository[1].



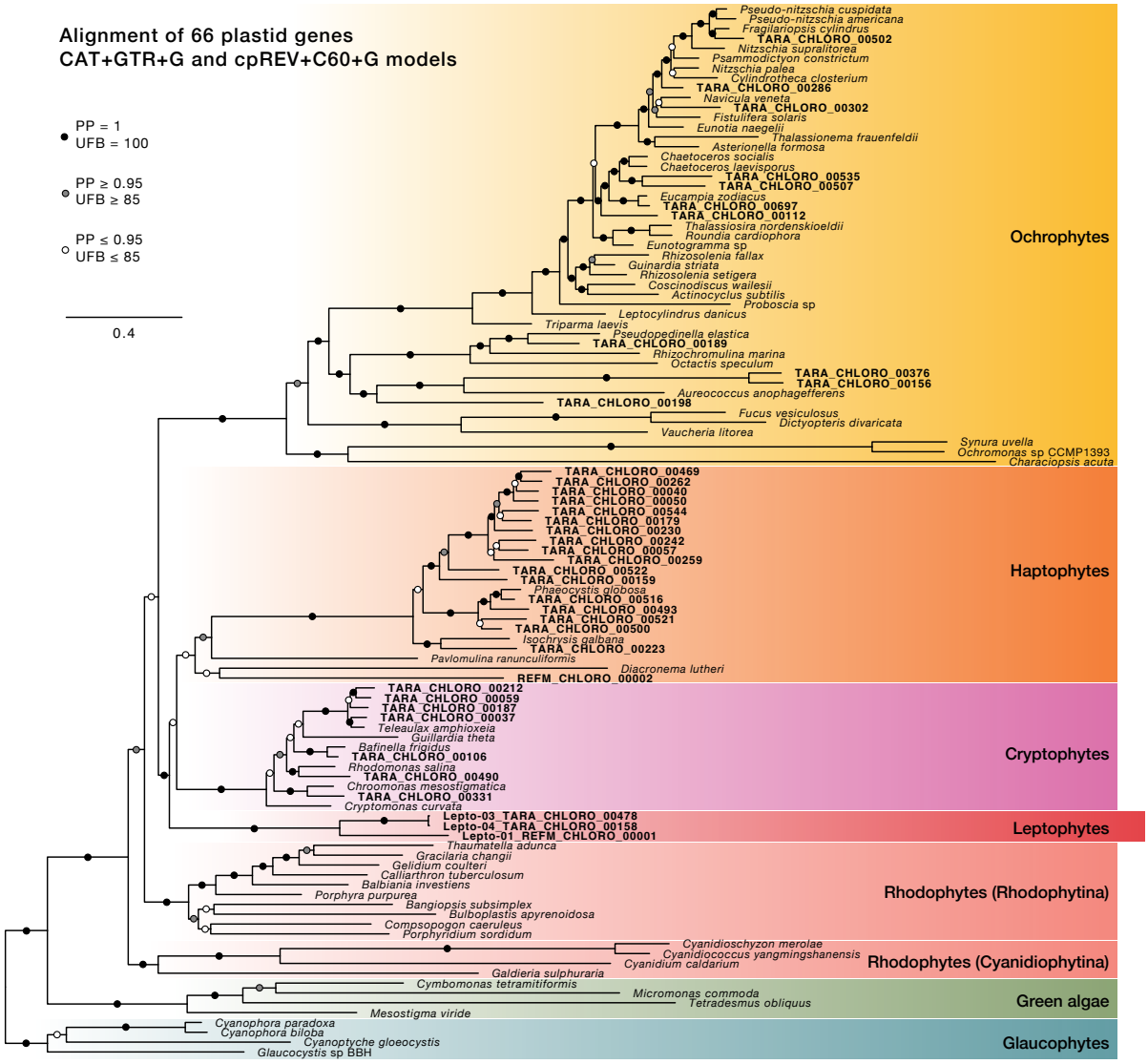
Supplementary Fig. 23. Annotated map of the putative Lepto-01 mitochondrial genome. Genes drawn inside the outer circle are transcribed clockwise, while those drawn outside the outer circle are transcribed counterclockwise. Genes are colour coded based on functional groups. The dark shaded area in the inner circle indicates GC content. The genome map was generated by OGDRAW[2].



Supplementary Fig. 24. Maximum parsimony scenarios for gene losses under the two possible plastid topologies: (**a**) H-sister and (**b**) HC-sister. Genes lost multiple times independently are coloured red. Gene loss numbers (displayed in red circles at every branch) were calculated based on the taxon sampling in Fig. 2 and Supplementary Fig. 16.



Supplementary Fig. 25. Bayesian inference of a preliminary dataset with 107 taxa, and 32 genes (7,921 sites). Branch support values indicate posterior probability obtained with the CAT+GTR+G model and ultrafast bootstrap values obtained with the cpREV+C60+G model.



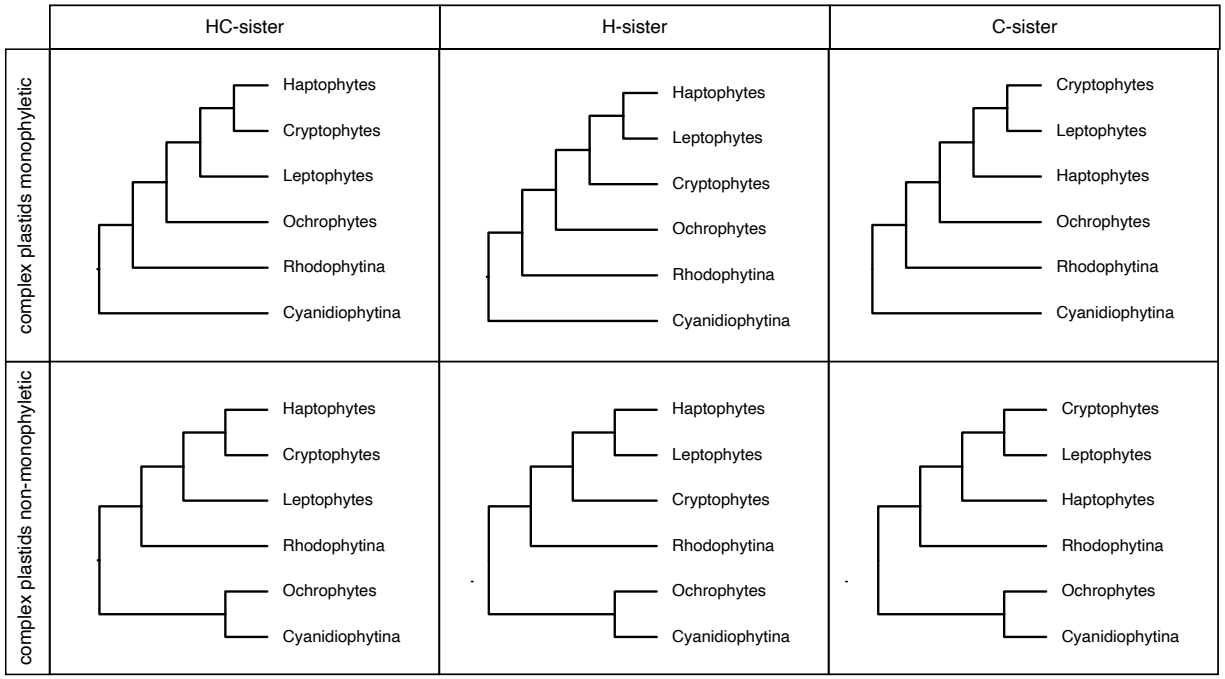
Supplementary Fig. 26. Bayesian inference of a preliminary dataset with 107 taxa, and 66 genes (16,438 sites). Branch support values indicate posterior probability obtained with the CAT+GTR+G model and ultrafast bootstrap values obtained with the cpREV+C60+G model.

Supplementary Table 1. Top 15 most abundant plastid genomes in the two Arctic stations where Lepto-01 (in bold text) is most abundant (0.8-2000 μm size fraction). Mean coverage was calculated by mapping the two listed metagenomes to each genome, and percentage abundance reflects each genome's signal as a proportion of the total plastid signal in each sample.

Region	Arctic				Arctic			
Size fraction	0.8-2000 µm				0.8-2000 µm			
Depth	SUR				SUR			
Station	194				193			
Filter ID	194SUR1GGZZ11				193SUR1GGZZ11			
	Plastid genome	**Clade**	**Mean coverage**	**% abundance**	**Plastid genome**	**Clade**	**Mean coverage**	**% abundance**
	TARA_CHLORO_00329	Ochrophyta (Coscinodiscophyceae)	417.3643034	16.05	NC_027589	Cryptophyta (Teleaulax amphioxeia)	514.6636486	25.84
	TARA_CHLORO_00704	Chlorophyta (Mammielophyceae)	254.054665	9.77	TARA_CHLORO_00704	Chlorophyta (Mammielophyceae)	306.51312	15.39
	TARA_CHLORO_00272	Euglenozoa	186.9111833	7.19	TARA_CHLORO_00237	Chlorophyta (Mammielophyceae)	153.864608	7.72
	TARA_CHLORO_00680	Ochrophyta (Coscinodiscophyceae)	154.8617982	5.95	NC_012575	Chlorophyta (Micromonas commoda)	83.05775298	4.17
	TARA_CHLORO_00275	Cryptophyta (Cryptophyceae)	153.357879	5.90	TARA_CHLORO_00297	Ochrophyta (Coscinodiscophyceae)	65.48039179	3.29
	Lepto-01	**Leptophyte**	124.151819	4.77	TARA_CHLORO_00473	Haptophyta (Prymnesiophyceae)	62.23567659	3.12
	TARA_CHLORO_00327	Ochrophyta (Coscinodiscophyceae)	88.50332919	3.40	TARA_CHLORO_00686	Ochrophyta	53.42210228	2.68
	TARA_CHLORO_00297	Ochrophyta (Coscinodiscophyceae)	84.94463191	3.27	TARA_CHLORO_00326	Ochrophyta (Coscinodiscophyceae)	44.74423352	2.25
	TARA_CHLORO_00473	Haptophyta (Prymnesiophyceae)	82.07171127	3.16	TARA_CHLORO_00327	Ochrophyta (Coscinodiscophyceae)	43.57903498	2.19
	TARA_CHLORO_00240	Haptophyta (Prymnesiophyceae)	75.97484423	2.92	TARA_CHLORO_00240	Haptophyta (Prymnesiophyceae)	42.20396394	2.12
	TARA_CHLORO_00328	Ochrophyta (Coscinodiscophyceae)	74.68082057	2.87	TARA_CHLORO_00299	Ochrophyta (Bacillariophyceae)	33.79422825	1.70
	TARA_CHLORO_00326	Ochrophyta (Coscinodiscophyceae)	70.60133521	2.71	TARA_CHLORO_00282	Ochrophyta (Bacillariophyceae)	33.1461306	1.66
	TARA_CHLORO_00263	Haptophyta (Prymnesiophyceae)	52.65510577	2.02	**Lepto-01**	**Leptophyte**	32.3916202	1.63
	NC_012575	Chlorophyta (Micromonas commoda)	41.60626851	1.60	TARA_CHLORO_00328	Ochrophyta (Coscinodiscophyceae)	29.42184233	1.48
	TARA_CHLORO_00239	Haptophyta (Prymnesiophyceae)	41.37302726	1.59	TARA_CHLORO_00267	Ochrophyta (Dictyochophyceae)	26.61431827	1.34

Supplementary Table 2. (On following page). Companion table for Figure 3c in the main text comparing six alternate topologies. Panel **a** shows the different topologies tested, focusing on three possible positions for leptophytes (HC-sister, H-sister, C-sister) under scenarios where complex plastids are monophyletic (as recovered in [4]) or non-monophyletic (as recovered in [5]). The table extends Figure 3c which displayed results only for the topologies where complex plastids are monophyletic for the sake of clarity. Panel **b** presents maximum likelihood estimates under different models for alternate positions of leptophytes within the CHL clade. The highest scoring topology is highlighted in bold. Blue bars show the difference in log-likelihood scores between each topology and the best-scoring one. A single asterisk (*) indicates topologies rejected by the Bonferroni-corrected chi-squared test, while two asterisks (**) indicate topologies also rejected by an Approximately Unbiased (AU) test. P-values are shown in Supplementary Data 2.

a



b

ML analyses using fixed topology trees					
Dataset	Untreated				Stationary trimmed
Model	GTR-CAT-PMSF	LG-MEOW80	LG-MEOW80 + GHOST (8 cat)	LG-MEOW80 + GF-MIX †	LG-MEOW80
Topology					
HC-sister, complex plastids monophyletic	**-931909.236**	**-1056271.087**	**-1045458.696**	-1055620.569 *	**-592787.09**
H-sister, complex plastids monophyletic	-931920.074 *	-1056274.766	-1045464.078 *	**-1055586.193**	-592793.164 *
C-sister, complex plastids monophyletic	-931924.826 **	-1056273.024	-1045463.152 *	-1055602.069 *	-592795.016 *
HC-sister, complex plastids non-monophyletic	-931915.342	-1056272.398	-1045466.547 *	-1055607.982 *	-592794.338 *
H-sister, complex plastids non-monophyletic	-931929.606 **	-1056281.817 *	-1045478.6 *	-1055617.997 *	-592801.836 *
C-sister, complex plastids non-monophyletic	-931935.101 **	-1056291.179 *	-1045482.379 *	-1055651.186 *	-592806.39 *

* Rejected by Bonferroni corrected χ2 test

** Rejected by Bonferroni corrected χ2 test and AU test

† AU test not possible under GF-MIX model

Supplementary Table 3. The 44 plastid encoded genes used to estimate completeness and redundancy (results shown in Supplementary Figure 2 and Supplementary Data 1).

Gene	Protein name
atpA	ATP synthase CF1 alpha chain
atpB	ATP synthase CF1 beta chain
atpE	ATP synthase epsilon chain
atpF	ATP synthase subunit b
atpH	ATP synthase CF0 C chain subunit III
petB	cytochrome b6
petG	cytochrome b6-f complex subunit V
psaA	photosystem I P700 apoprotein A1
psaB	photosystem I P700 apoprotein A2
psaC	photosystem I iron-sulfur center
psaJ	photosystem II protein J
psbA	photosystem II reaction center protein D1
psbB	photosystem II chlorophyll A core antenna apoprotein
psbC	photosystem II chlorophyll A core antenna
psbD	photosystem II reaction center protein D2
psbE	cytochrome b559 alpha chain
psbF	cytochrome b559 beta chain
psbH	photosystem II protein H
psbJ	photosystem II protein J
psbK	photosystem II protein K
psbL	photosystem II protein L
psbN	photosystem II protein N
psbT	photosystem II protein T
rbcL	ribulose bisphosphate carboxylase large chain
rpl14	50S ribosomal protein L14
rpl16	50S ribosomal protein L16
rpl2	50S ribosomal protein L2
rpl20	50S ribosomal protein L20
rpl36	50S ribosomal protein L36
rpoA	DNA-directed RNA polymerase alpha chain
rpoB	DNA-directed RNA polymerase beta chain
rpoC1	DNA-directed RNA polymerase beta' chain
rpoC2	DNA-directed RNA polymerase beta' chain
rps11	30S ribosomal protein S11
rps12	30S ribosomal protein S12
rps14	30S ribosomal protein S14
rps18	30S ribosomal protein S18
rps19	30S ribosomal protein S19
rps2	30S ribosomal protein S2
rps3	30S ribosomal protein S3
rps4	30S ribosomal protein S4
rps7	30S ribosomal protein S7
rps8	30S ribosomal protein S8
ycf4	photosystem I assembly protein ycf4

Supplementary Table 4. The 25 canonical mitochondrial genes used to check for mitochondrial contamination in the plastid MAGs.

Gene	Protein name
nad1	NADH dehydrogenase subunit 1
nad2	NADH dehydrogenase subunit 2
nad3	NADH dehydrogenase subunit 3
nad4	NADH dehydrogenase subunit 4
nad4L	NADH dehydrogenase subunit 4L
nad5	NADH dehydrogenase subunit 5
nad6	NADH dehydrogenase subunit 6
nad7	NADH dehydrogenase subunit 7
nad8	NADH dehydrogenase subunit 8
nad9	NADH dehydrogenase subunit 9
nad10	NADH dehydrogenase subunit 10
nad11	NADH dehydrogenase subunit 11
sdh2	Succinate dehydrogenase [ubiquinone] iron-sulfur subunit
sdh3	Succinate dehydrogenase cytochrome b560 subunit
sdh4	Succinate dehydrogenase subunit 4
cob	Cytochrome b
cox1	Cytochrome c oxidase subunit 1
cox2	Cytochrome c oxidase subunit 2
cox3	Cytochrome c oxidase subunit 3
atp1	ATP synthase subunit alpha
atp3	ATP synthase subunit gamma
atp4	ATP synthase subunit 4
atp6	ATP synthase F(0) complex subunit a
atp8	ATP synthase F(0) complex subunit 8
atp9	ATP synthase subunit 9

Supplementary Table 5. Convergence of Phylobayes GTR+CAT+G4 analyses for the 107 taxon, 93 gene dataset with a fixed input tree for the CAT-PMSF approach[6]. Two Markov chains were run, and burn-in for tracecomp was set to 1,800.

Parameter	Effective sample size	Relative difference
loglik	35	1.6552
length	113	0.0624932
alpha	330	0.525332
Nmode	146	0.227368
statent	118	3.53681
statalpha	110	0.210322
rrent	104	0.979222
rrmean	3525	0.0110885

Supplementary Table 6. The 28 genes used for mitochondrial phylogenetic analyses (Figure 4).

Gene	Protein name
atp1	ATP synthase subunit alpha
atp4	ATP synthase subunit 4
atp6	ATP synthase F(0) complex subunit a
atp8	ATP synthase F(0) complex subunit 8
cob	Cytochrome b
cox1	Cytochrome c oxidase subunit 1
cox2	Cytochrome c oxidase subunit 2
cox3	Cytochrome c oxidase subunit 3
nad1	NADH dehydrogenase subunit I
nad2	NADH dehydrogenase subunit 2
nad3	NADH dehydrogenase subunit 3
nad4	NADH dehydrogenase subunit 4
nad4L	NADH dehydrogenase subunit 4L
nad5	NADH dehydrogenase subunit 5
nad6	NADH dehydrogenase subunit 6
nad7	NADH dehydrogenase subunit 7
nad9	NADH dehydrogenase subunit 9
nad10	NADH dehydrogenase subunit 10
nad11	NADH dehydrogenase subunit 11
rpl2	50S ribosomal protein L2
rpl6	50S ribosomal protein L6
rpl16	50S ribosomal protein L16
rps2	30S ribosomal protein S2
rps4	30S ribosomal protein S4
rps8	30S ribosomal protein S8
rps14	30S ribosomal protein S14
rps19	30S ribosomal protein S19
tatC	Sec-independent protein translocase component TatC

Supplementary Note 1.
Evaluating the position of leptophytes in the plastid phylogeny

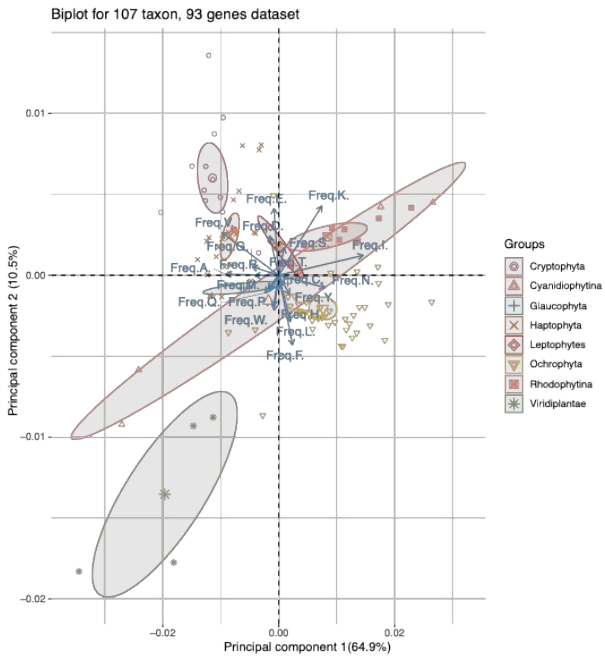
Our phylogenomic analyses based on 107 taxa, 93 plastid-encoded genes and a range of sophisticated phylogenetic models confirmed that leptophytes are related to haptophytes and cryptophytes. However, the position of leptophytes within the CHL group (Cryptophytes-Haptophytes-Leptophytes) could not be confidently resolved and two topologies remained credible: the HC-sister toplogy (leptophytes sister to both hatophytes and cryptophytes), and the H-sister topology (leptophytes sister to haptophytes).

The uncertainty of the position of leptophytes could be caused by lack of phylogenetic signal (as indicated by the small branch lengths separating the groups within the CHL clade), and could also be caused by potential biases and phylogenetic artefacts, such as potential compositional heterogeneity across taxa[7], and heterotachy[8].

The LG+MEOW80+GHOST model with 8 linked site classes supported the HC-sister topology, which was also recovered by the CAT-GTR+G model with strong support (Fig. 3 in main text). However, the GFmix[7] model which accounts for compositional heterogeneity across sites and branches strongly favoured the H-sister topology (maximum likelihood values for the H-sister and HC-sister topology differed by 34 points).

Compositional heterogeneity

Compositional heterogeneity across taxa may be an artefact of our dataset, particularly due to the inclusion of extremophile red algae (Cyanidiophytina). Additionally, recent studies have suggested that ribosomal protein genes, which are highly represented in our dataset, tend to exhibit greater compositional bias due to strong co-evolution among these proteins[9,10]. A principal component analysis (PCA) revealed the amino acid preferences across different groups (Supplementary Fig. 27).



Supplementary Figure 27. PCA analysis showing compositional biases among clades based on the alignment used for phylogenomic analyses.

We accounted for the compositional bias in three ways: (1) evaluating the likelihood of alternative topologies under the GFmix model, (2) removing compositionally heterogenous sites, and (3) recoding the dataset using the SR4 recoding scheme[11].

1. *GFmix model*

The GFmix model adjusts the vector of amino acid frequencies in a branch-specific manner to account for shifts in the relative frequencies of amino acids in different branches of the tree. To identify which amino acids are enriched or depleted in specific lineages, we analyzed our concatenated alignment using a custom R script (courtesy of Charley McCarthey, Dalhousie University; available on the GitHub repository[1]. This script divided the taxa into two groups and identified amino acid groups enriched or depleted in each lineage. The two taxa groups compared were *Cyanidiococcus yangmingshanensis* (accession NC_051883) and *Cyanidioschyzon merolae* (accession NC_004799) versus all other taxa.

The groups of amino acids enriched or depleted in these lineages are as follows:
G-class: I, N, F, K, T, D, S, E
F-class: R, C, V, W, H, M, A, Q

The GFmix program utilized the enriched and depleted amino acid classes to calculate the log-likelihoods for the six alternative topologies listed in Supplementary Table 2a. Among these, the H-sister topology with complex plastids forming a monophyletic grouping, was strongly favoured by the LG+MEOW80+GFmix model (Supplementary Table 2a). All other topologies were statistically rejected based on the Bonferroni-corrected chi-squared test.

2. *Removing compositionally heterogenous sites*

We generated a compositionally homogenized dataset by removing heterogenous sites from the concatenated alignment using Stuart's test of marginal homogeneity as implemented in BMGE[12]. This dataset was then analysed with the conventional site-heterogenous model LG+MEOW80+G, recovering the HC-sister topology (Fig. 3c, Supplementary Table 2). All alternative topologies were rejected by the Bonferroni-corrected chi-squared test, but not by the AU topology test. The HC-sister topology was also strongly supported when analysing the dataset under the CAT+GTR+G model (PP = 0.99).

3. *SR4 recoding*

We recoded the concatenated alignment using the SR4 recoding scheme and inferred a phylogeny under the CAT+GTR+G model in PhyloBayes. This phylogeny also supported the H-sister topology, consistent with the results of the GFmix model, albeit with low support (PP = 0.75). The limited support may stem from the low phylogenetic signal in the dataset, further eroded by amino acid recoding. Additionally, we caution that this result should be interpreted cautiously as recoding datasets has been shown to generate incorrect trees due to the associated of phylogenetic signal[13,14].

Altogether, the GFmix model, which explicitly models both site- and branch-heterogeneity, supports the H-sister topology. However, compositionally homogenized datasets analysed with conventional site-heterogenous models recapitulate the results of the GFmix model in only one out of two cases. While H-sister topology is more likely as it invokes a simpler scenario of plastid transfer and gene content evolution (Fig. 5, Supplementary Fig. 24), we cannot confidently rule out the HC-sister topology based on our current analyses. More data

in the form of cell morphology and nuclear DNA from leptophytes will be required to decide between these scenarios.

Supplementary references

1. Jamy, M. Code for 'Identification of a deep-branching lineage of algae using environmental plastid genomes'. Zenodo https://doi.org/10.5281/zenodo.17635604 (2025).
2. Greiner, S., Lehwark, P. & Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **47**, W59–W64 (2019).
3. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
4. Pietluch, F., Mackiewicz, P., Ludwig, K. & Gagat, P. A New Model and Dating for the Evolution of Complex Plastids of Red Alga Origin. *Genome Biol. Evol.* **16**, evae192 (2024).
5. Kim, J. I. *et al.* Evolutionary Dynamics of Cryptophyte Plastid Genomes. *Genome Biol. Evol.* **9**, 1859–1872 (2017).
6. Szánthó, L. L., Lartillot, N., Szöllősi, G. J. & Schrempf, D. Compositionally Constrained Sites Drive Long-Branch Attraction. *Syst. Biol.* **72**, 767–780 (2023).
7. Muñoz-Gómez, S. A. *et al.* Site-and-branch-heterogeneous analyses of an expanded dataset favour mitochondria as sister to known Alphaproteobacteria. *Nat. Ecol. Evol.* **6**, 253–262 (2022).
8. Crotty, S. M. *et al.* GHOST: Recovering Historical Signal from Heterotachously Evolved Sequence Alignments. *Syst. Biol.* **69**, 249–264 (2020).

9. Petitjean, C., Deschamps, P., López-García, P. & Moreira, D. Rooting the Domain Archaea by Phylogenomic Analysis Supports the Foundation of the New Kingdom Proteoarchaeota. *Genome Biol. Evol.* **7**, 191–204 (2014).
10. Ramulu, H. G. *et al.* Ribosomal proteins: Toward a next generation standard for prokaryotic systematics? *Mol. Phylogenet. Evol.* **75**, 103–117 (2014).
11. Susko, E. & Roger, A. J. On Reduced Amino Acid Alphabets for Phylogenetic Inference. *Mol. Biol. Evol.* **24**, 2139–2150 (2007).
12. Criscuolo, A. & Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 210 (2010).
13. Hernandez, A. M. & Ryan, J. F. Six-State Amino Acid Recoding is not an Effective Strategy to Offset Compositional Heterogeneity and Saturation in Phylogenetic Analyses. *Syst. Biol.* **70**, 1200–1212 (2021).
14. Foster, P. G. *et al.* Recoding Amino Acids to a Reduced Alphabet may Increase or Decrease Phylogenetic Accuracy. *Syst. Biol.* **72**, 723–737 (2023).