

SUPPLEMENTARY INFORMATION

Global metagenomics reveals plastid diversity and unexplored algal lineages

Shrestha et al.

Supplementary Note 1: Expanding the cyanobacterial taxonomy to explore plastid evolution

Our initial phylogenetic reconstruction included 1,973 genomes/MAGs associated with Cyanobacteriota, using four Margulisbacteria species as an outgroup (Supplementary Fig. 1). This dataset was dereplicated to reduce representative cyanobacterial dataset without collapsing and mixing taxonomic orders. Through dereplication, the original dataset was reduced from 1,973 to 259 genomes/MAGs, comprising 16, 96, and 147 from the three clades Sericytochromatia, Vampirovibrionia, and Cyanobacteria, respectively. The phylogenetic relationships between the original and the dereplicated dataset were nearly identical, except the order FACHB-261 branched off prior to Gloeomargaritales in the dereplicated tree (Fig. 1a, Supplementary Fig. 1). Our phylogenetic analysis strongly supports monophyly of three major clades within Cyanobacteriota, placing Sericytochromatia at the base and the non-photosynthetic clade Vampirovibrionia as a sister to Cyanobacteria, consistent with published studies^{1,2}.

In terms of Cyanobacteriota taxon sampling, our study lacks species from the newly described GTDB order JACMPNO1 within Sericytochromatia but includes all other known orders. Within Vampirovibrionia, members formed two strongly supported monophyletic clades (UFBoot 100%), including Vampirovibrionales and Gastranaerophilales as a sister to clade containing orders Obscuribacteriales, LMEP-6097, Caenarcaniphilales, and 2-02-FULL-35-15, consistent with other studies¹. Gastranaerophilales represented the largest order within Vampirovibrionia with a total of 423 MAGs. Cyanobacteria was strongly supported as a monophyletic clade (UFBoot 100%). Gloeobacterales represented the most basal and early diverging order within Cyanobacteria, followed by the diversification of orders Thermostichales, Pseudanabaenales, and FACHB-261 with strong support (UFBoot >95%), consistent with previous studies^{1,3}. Subsequently, the order Gloeomargaritales diversified from remaining cyanobacterial orders (moderate support, UFBoot 75%), followed by the order Thermosynechococcales (full support, UFBoot 100%). Our topology is consistent with a recent cyanobacterial phylogeny incorporating deep-branching lineages³, although alternative relationships of Gloeomargaritales and Thermosynechococcales with other cyanobacterial orders have also been documented¹. The remaining 14 photosynthetic cyanobacterial orders were grouped into two well supported clades: one including CAIUCS01, PCC7407, Elainellales, Phormidesmiales, CACIAM-69d, PCC-9006, Synechococcales, and PCC-9307, and the other containing orders SKYG9, GCF-001939115, FACHB-46, Neosynechococcales, Leptolyngbyales, and Cyanobacterales (Fig. 1a). Cyanobacterales represented the largest order, containing 558 genomes/MAGs in our dataset. Our Cyanobacteria phylogeny included comprehensive taxon sampling from 21 orders, and their relationships resembled the topology presented by Tan et al. (2024), except for the placement of

the orders Elainellales and Phormidesmiales. Our topology strongly supports Elainellales as a sister to Phormidesmiales and together placed as a sister to PCC-6307 and Synechococcales, which is incongruent with the topology presented by Tan et al. (2024), where Elainellales was sister to Phormidesmiales and Cyanobacteriales.

We then utilized the dereplicated Cyanobacteriota representatives along with RefSeq NCBI plastomes to trace the origin of plastids, conducting several phylogenetic analyses while adjusting the phylogenetic markers (UNI56 and PLASTID54) and the ptMAGs dataset generated in our study. We noted that with an increasing presence of UNI56 marker (above 20%), more plastomes and ptMAGs representing Streptophyta and Chlorophyta were filtered out from the final alignment. This is mainly due to the large number of HMM profiles for ribosomal proteins, which constitute the majority of UNI56 marker. It is known that plastids in the red lineage have retained more plastid-encoded ribosomal proteins compared to those in the green lineage^{4,5}. To mitigate this, we used PLASTID54 marker that included additional plastid-encoded photosynthesis-related proteins to retain plastomes and MAGs associated with the green lineage. Similarly, we observed that some ptMAGs, particularly those belonging to Streptophyta, shared high (>90%) average nucleotide identity (ANI) with RefSeq plastome as well mitochondrial genomes. For example, a ptMAG IMG3300027552_BIN20 of a single contig of ~43 kb, had 100% ANI and alignment fraction against *Arabidopsis thaliana* plastome (NC_000932), but also shared 92% ANI with a 91% alignment fraction against *Pulsatilla cernua* (from the family Ranunculaceae) mitochondrial genome (NC_068018). Plastid MAGs with higher ANI and alignment fraction against the plastome compared to the mitochondrial genome were included in our analysis. The presence of plastid sequences in mitochondrial genome due to intracellular and horizontal gene transfer are well documented in land plants⁶⁻⁸, and likely explains the higher ANI shared with mitochondrial genomes. Furthermore, our initial phylogenetic analysis included Alveolata RefSeq plastomes, but due to their extremely long branch lengths, we excluded them and the associated ptMAGs from subsequent analyses (Supplementary Fig. S2).

In our phylogenetic analyses, ptMAGs and RefSeq plastomes together formed a strongly supported monophyletic group, positioned as a sister to cyanobacterial order Gloeomargaritales. This finding aligns with the observation that Archaeplastida plastids originated from deeply branching cyanobacterial lineage⁹⁻¹¹. Whereas *Paulinella* plastids were placed within the recently diverged cyanobacterial order PCC-6307, following the diversification of *Synechococcus* sp. RSCF101. Our dereplicated dataset was much smaller, substantially reducing the taxa representing the order PCC-6307 from 498 to 28 taxa. Therefore, we inferred *Paulinella* plastid relationship by retaining all the original PCC-6307 taxa. Nevertheless, the phylogenetic position of *Paulinella* plastid within PCC-6307 remained unchanged, even with the expanded dataset (Supplementary Fig. 3). The *Synechococcus* sp. RSCF101 is distinct from *Synechococcus* sp. WH5701, which is considered the closest extant relative of *Paulinella* plastid¹². These two *Synechococcus* species belonged to separate phylogenetic clades within the

PCC-6307 (Fig. 1c, Supplementary Fig. 3). Moreover, *Synechococcus* species WH5701 and RSCF101 exhibit noticeable differences in genomic features, including variation in genome sizes (3.04 Mb vs. 2.98 Mb, respectively) and gene counts (3403 vs. 3021 genes, respectively).

Supplementary Note 2: Plastid MAGs and their relationships within Ochrophyta

In their extensive phylogenomic analyses of stramenopiles, Derelle et al. (2016)¹³ initially suggested two major divisions within Ochrophyta: Chrysista, which includes classes Xanthophyceae, Phaeophyceae, Rhodophyceae, Synchromophyceae, Synurophyceae, and Chrysophyceae, and Diatomista, which comprises the classes Pelagophyceae, Dictyochophyceae, Bolidophyceae, and the phylum Bacillariophyta (diatoms). Since then, several phylogenomic studies have aimed to delineate major ochrophyte lineages by sampling unrepresented taxa and improving phylogenetic resolution. However, the current understanding of Ochrophyta relationships remains inconsistent due to incomparable taxon sampling, different markers (nuclear and organellar), and phylogenetic methods used^{14–18}. While the relationships among the major classes within Diatomista are well resolved, incongruent relationships among the classes within Chrysista are prevalent.

In our study, most novel ptMAGs recovered belonged to Ochrophyta (n=155, Supplementary Data 4). We reconstructed phylogenies by combining the novel ptMAGs and reference ochrophyte plastomes to assess the placement of novel ptMAGs within ochrophytes. Our phylogenetic analyses strongly support the monophyly of Ochrophyta but with inconsistent placement of some classes belonging to Chrysista (Fig. 2, Supplementary Figs. 4, 11–12). Our phylogenetic reconstruction, which included only Ochrophyta plastomes and ptMAGs consistently recovered Chrysista as a paraphyletic group (Supplementary Fig. 12). To avoid the influence of novel ptMAGs, we first reconstructed the phylogeny using only 213 reference ochrophyte plastomes (Supplementary Fig. 11). The reference ochrophytes phylogeny recovered two strongly supported clades: one with Chrysophyceae and Eustigmatophyceae and another clade with remaining all other ochrophytes. Olisthodiscophyceae and Pinguicophyceae, each represented by a single species, formed a moderately supported clade (UFBoot 90%), and diverged early as sister to PXR and Diatomista. The PXR clade was strongly supported as a monophyletic group (UFBoot 100%) and was recovered as a sister to Diatomista with relatively low support (UFBoot 78%). Within Diatomista, Pelagophyceae and Dictyochophyceae formed a moderately supported clade as a sister to Bolidophyceae and Bacillariophyta (UFBoot 93%). When we added novel ptMAGs associated with Ochrophyta on the reference plastomes dataset, an identical topology was recovered with much higher bootstrap support (Supplementary Fig. 12). This topology lacks the dichotomy between Chrysista and Diatomista and is incongruent with the current phylogenomic studies^{14,17,18}, likely due to the phylogenetic markers. We selected plastid phylogenetic markers commonly present in red and green lineages with the broad aim to understand plastid evolution across diverse photosynthetic eukaryotes. To resolve evolutionary relationships with Ochrophyta, lineage-specific phylogenetically informative markers, robust

assessment of phylogenetic methods, and comprehensive taxon sampling are required^{14,18}. Despite the incongruency in relationship between Ochrophyta classes, the placement of novel ptMAGs within the classes was always consistent. Therefore, we used the inferred topology to describe the placement of Ochrophyta-related novel ptMAGs (Supplementary Fig. 12).

Among 155 Ochrophyta-related novel ptMAGs, most (n=122) belonged to Diatomista, with the remaining 33 placed within Chrysista (Supplementary Fig. 12). Within Chrysista, the largest number of ptMAGs (n=26) were within Chrysophyceae, with size ranging from ~19 to ~132 kb, and 79% median completeness. The remaining ptMAGs within Chrysista were confined to Raphidophyceae (n=3), Xanthophyceae (n=2), and a single novel ptMAG of ~110 kb (IMGM3300020214_BIN249) within Eustigmatophyceae, placed as a sister to *Nannochloropsis* sp. Within Diatomista, novel ptMAGs were distributed across all major clades. Pelagophyceae contained ten novel ptMAGs, size ranging from ~47 to ~138 kb with 83% median completeness. Similarly, 24 novel ptMAGs, ranging in size from ~21 to ~110 kb with 76% median completeness were recovered within Dictyochrophyceae. We also recovered the plastid of the centropasthelida species *Meringosphaera mediterranea* (a Haptista) between *Rhizochromulina marina* and *Pseudopedinella elastica* plastids. Plastids in *Meringosphaera* are known to be derived through kleptoplasty, a process where plastids of engulfed prey are temporarily retained. Phylogenetic analyses using 16S rDNA and plastid markers have placed them within Dictyochrophyceae¹⁹, consistent with our phylogenetic placement.

We consistently recovered a clade containing only ptMAGs prior to the diversification of Bolidophyceae and Bacillariophyta (Supplementary Fig. 12). The clade included three novel ptMAGs, ranging in size from ~46 to ~49 kb, which lacked reference plastomes, and were strongly supported (UFBoot 100%) as sister to Bolidophyceae and Bacillariophyta. The class Bolidophyceae is represented by a single complete plastome of *Triparma laevis*²⁰. We identified a clade that included four additional ptMAGs, ranging from ~104 to ~106 kb, along with *T. laevis*. These metagenomes have 95% median completeness, and their sizes are comparable to the ~112 kb *T. laevis* plastome (excluding a copy of inverted repeats). Within Bolidophyceae, *T. laevis* plastome, together with four novel ptMAGs formed two distinct clades within Bolidophyceae. One contained the *T. laevis* plastome along with two ptMAGs, IMGM3300031602_BIN118, and IMGM3300035224_BIN177, and the other clade with two metagenomes IMGM3300025666_Ga0209601 and IMGM3300027791_BIN75. Interestingly, two copies of *psbA* gene, which is suggested to be a unique feature of the *T. laevis* plastome, were also found in two ptMAGs IMGM3300031602_BIN118, and IMGM3300035224_BIN177. In contrast, only a single copy of *psbA* gene was present in the sister clade containing only ptMAGs. In IMGM3300035224_BIN177, a copy of *psbA* was much smaller, about 265 bp compared to the full-length gene of 1083 bp and was located at the end of a contig, likely representing a partial *psbA* copy due to incomplete assembly. The two copies of *psbA* in *T. laevis* were attributed to gene duplication after the diversification of Bolidophyceae²⁰. However, the

absence of the duplicated copy in the sister clade within Bolidophyceae suggests the duplication event was confined to a specific clade within Bolidophyceae. Alternatively, the gene duplication may have occurred early during the speciation of Bolidophyceae, with the sister clade lacking the duplicate *psbA* due to incomplete assembly.

Bacillariophyta was strongly supported as a monophyletic clade and included the largest number of novel ptMAGs (n=81), with assembly sizes ranging from ~20 to ~116 kb, and between 28% and 99% completeness (96% median). The relationship among Bacillariophyta families recovered in our phylogenetic reconstructions were consistent with the current comprehensive phylogenomic study using concatenated plastid genes of reference diatom species²¹. The extensive collection of novel ptMAGs associated with Bacillariophyta generated in this study will serve as a valuable resource for future research, significantly enhancing phylogenetic resolution and facilitating comparative analyses within this group.

Lastly, we identified a single novel ptMAG (IMGM3300021091_BIN1003) that is strongly supported (UFBoot 100%) as a sister to plastome of dinoflagellate species *Kryptoperidinium foliaceum* (Supplementary Fig. 12). The ptMAG is represented by two contigs with a total length of ~108 kb and retains nearly all protein coding genes present in *K. foliaceum* plastome except for *ccsA* (Cytochrome C biogenesis), *tyrC* (tyrosine recombinase), and *serC1/C2* (serine recombinases), likely due to incomplete assembly. Plastids of *K. foliaceum* and the closely related dinoflagellate species *Durinskia baltica* are thought to be derived from a diatom through tertiary endosymbiosis and are collectively referred to as dinotoms²². Early phylogenetic studies using single gene markers and rDNA sequences suggested dinotom plastids were closely related to plastids of *Nitzschia* species^{23,24}. Due to the lack of complete plastomes from *Nitzschia* species at the time, dinotom plastomes were compared with those of the free-living diatom species *Phaeodactylum tricornutum*²². With the access to multiple *Nitzschia* plastomes, *D. baltica* plastome was found to be highly similar (92% pairwise identity) to the *Nitzschia palea* plastome, sharing identical gene content and synteny²⁵. However, monophyletic dinotoms were found to retain plastids of multiple *Nitzschia* or other diatom species through independent acquisition and replacement of the endosymbiont²⁶. This complicates the understanding of tertiary endosymbiosis in dinotoms, making it challenging to pinpoint the source of endosymbiont plastids.

In our phylogenetic reconstruction, plastids from two dinotoms (*K. foliaceum* and *D. baltica*) formed a strongly supported (UFBoot 100%) monophyletic group with *Nitzschia palea*, *Tryblionella apiculata*, and the ptMAG IMGM3300021091_BIN1003 (Supplementary Fig. 12). Within the group, *T. apiculata* branched off early at the base, with *K. foliaceum* and IMGM3300021091_BIN1003 grouping as a sister to *D. baltica* and *N. palea*. It is likely that *K. foliaceum* may have retained the plastid of the diatom *T. apiculata*, whereas in *D. baltica*, their plastids were replaced with the endosymbiont *N. palea*. The plastome size (excluding a copy of

inverted repeats) of *K. foliaceum* (134,409 bp) was comparable to that of diatom *T. apiculata* (129,773 bp) and shares all protein coding genes except a copy of putative serine recombinase (*serCI*). Although *T. apiculata* plastome lacks annotated tyrosine recombinase (*tyrC*) found in *K. foliaceum*, it contains a putative integrase/recombinase gene annotated as *xerC*²⁷. The XerC is 101 amino acids long, substantially shorter (due to premature stop codon) compared to TyrC of 312 aa, and shares 52% pairwise aa identity, with both containing an integrase catalytic domain (IPR002104). The pairwise nucleotide identity between these two genes are much higher (63%) with the upstream region of *xerC* gene sharing 59% nucleotide identity with the 5' region of the *tyrC* gene. The nearly identical gene content between the dinotom *K. foliaceum* and the diatom *Tryblionella apiculata* suggests that *K. foliaceum* may have retained the plastid of the diatom *Tryblionella apiculata*, whereas in *D. baltica*, their plastids were replaced with the endosymbiont *N. palea*.

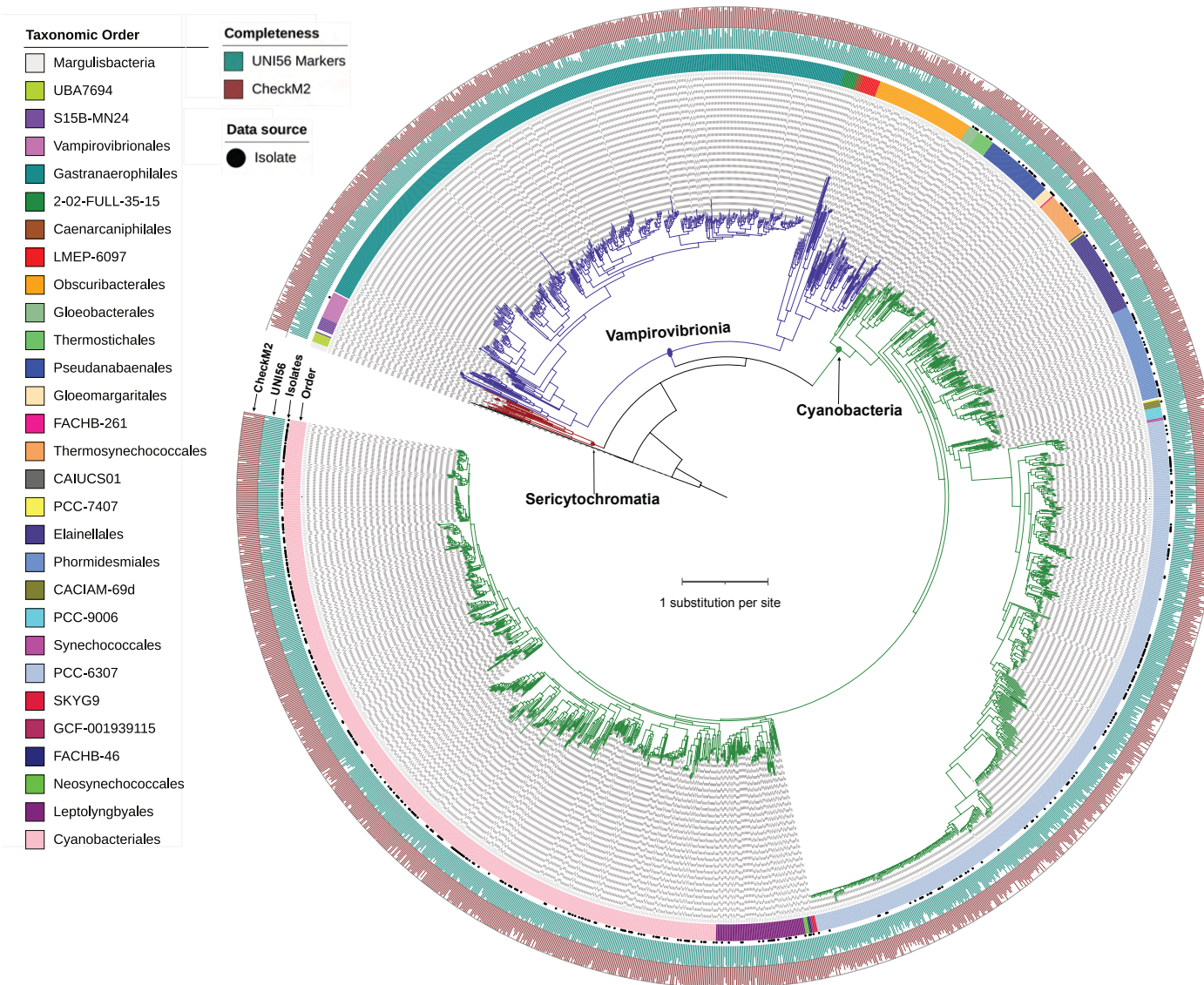
Our phylogenetic reconstruction Ochrophyta, incorporating novel ptMAGs and RefSeq plastomes, supports a monophyletic origin of Ochrophyta plastids but reveals inconsistencies in placements for some Chrysista classes. Subsequent studies with lineage-specific markers and robust phylogenetic methods will likely solve these relationships, and the novel ptMAGs will be resourceful. Despite these inconsistencies, the novel ptMAGs consistently fit within their respective classes, with most belonging to Diatomista. We identified many novel clades represented solely by ptMAGs but lacking reference plastomes including Chrysophyceae, Pelagophyceae, Dictyophyceae, Bolidophyceae and the phylum Bacillariophyta. Among them, an early branching clade sister to Bolidophyceae and Bacillariophyta, likely represents a lineage without reference plastomes with essential implication on our understanding of plastid evolution in diatoms. Furthermore, novel ptMAGs closely related to the dinotoms like *K. foliaceum* and *D. baltica*, show close relationships with specific diatom species and provide resources to investigate evolution of complex plastids in dinoflagellates.

Supplementary Note 3: Potentially misclassified plastomes in the NCBI database

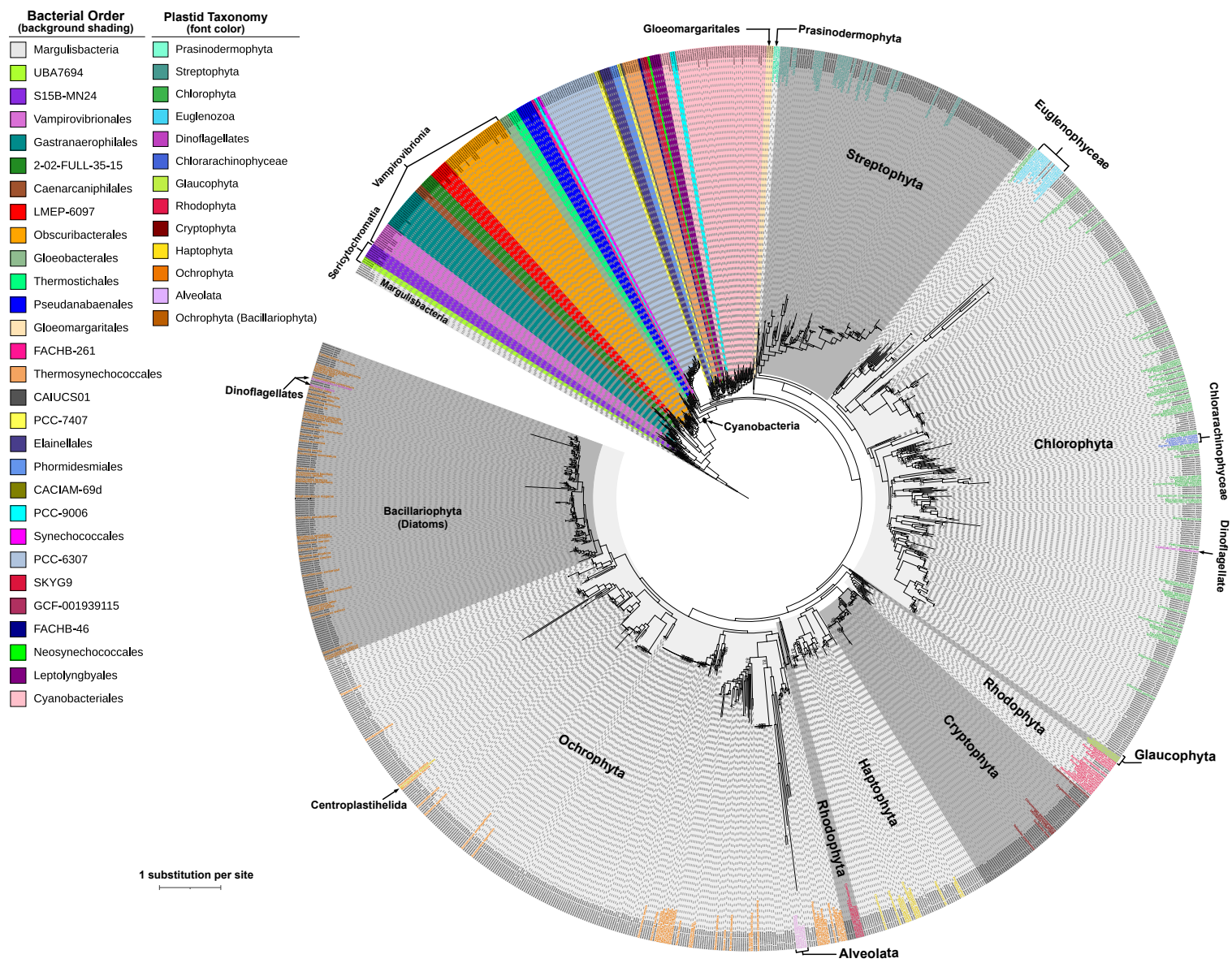
In our study, we identified a few cases of potential mislabeling among plastomes in the NCBI database. Although, we note that these examples are not expected to be an exhaustive list of such database errors. The *Interfilum terricola* plastome (NC_025542) is classified in NCBI as belonging to a streptophyte alga, but our phylogenomic analysis placed this sequence as an early diverging taxon within Trebouxiophyceae, consistent with the position recovered by Lemieux et al. (2014)²⁸ for *Geminella terricola* (KM462881) within chlorophytes. The sequences for these two accessions share 100% nucleotide identity, and the mislabeling likely arose due to confusion caused by renaming of *Geminella terricola* strain SAG 2100 to *Interfilum terricola* by Mikhailiyuk, et al. 2008²⁹. Additional examples of likely mis-labeled plastomes include: (i) plastome NC_058752, attributed to *Grateloupia livida* (Rhodophyta, Florideophyceae), which shares 99.9% identity with plastome NC_036804 of *Dictyopteris divaricate* (Ochrophyta, Phaeophyceae), (ii) proteins of plastome NC_058751 of *Chorda filum* (Ochrophyta, Phaeophyceae), which matched best with proteins from seed plants, specifically Zosteraceae

species (marine monocots), and our phylogenetic analysis placed this plastome within Streptophyta (results not shown), (iii) the *Symbiochloris* sp. SG-2018 plastome (NC_065389), listed as Trebouxiophyceae in NCBI, which was placed within Ulvophyceae in our preliminary analysis. This sequence is published as belonging to *Symbiochlorum hainanensis* (Ulvophyceae), with the NCBI accession ON645925 mis-labeled as *Symbiochloris*. Other than the *Interfilum terricola* plastome (NC_025542) that we corrected to *Geminella terricola* in figures and tables, these mis-labeled plastomes were not found to be sister to an identified ptMAG and were therefore excluded from the phylogeny presented in our study. These findings underscore the necessity for careful verification of taxonomic assignments when using publicly available plastome sequences.

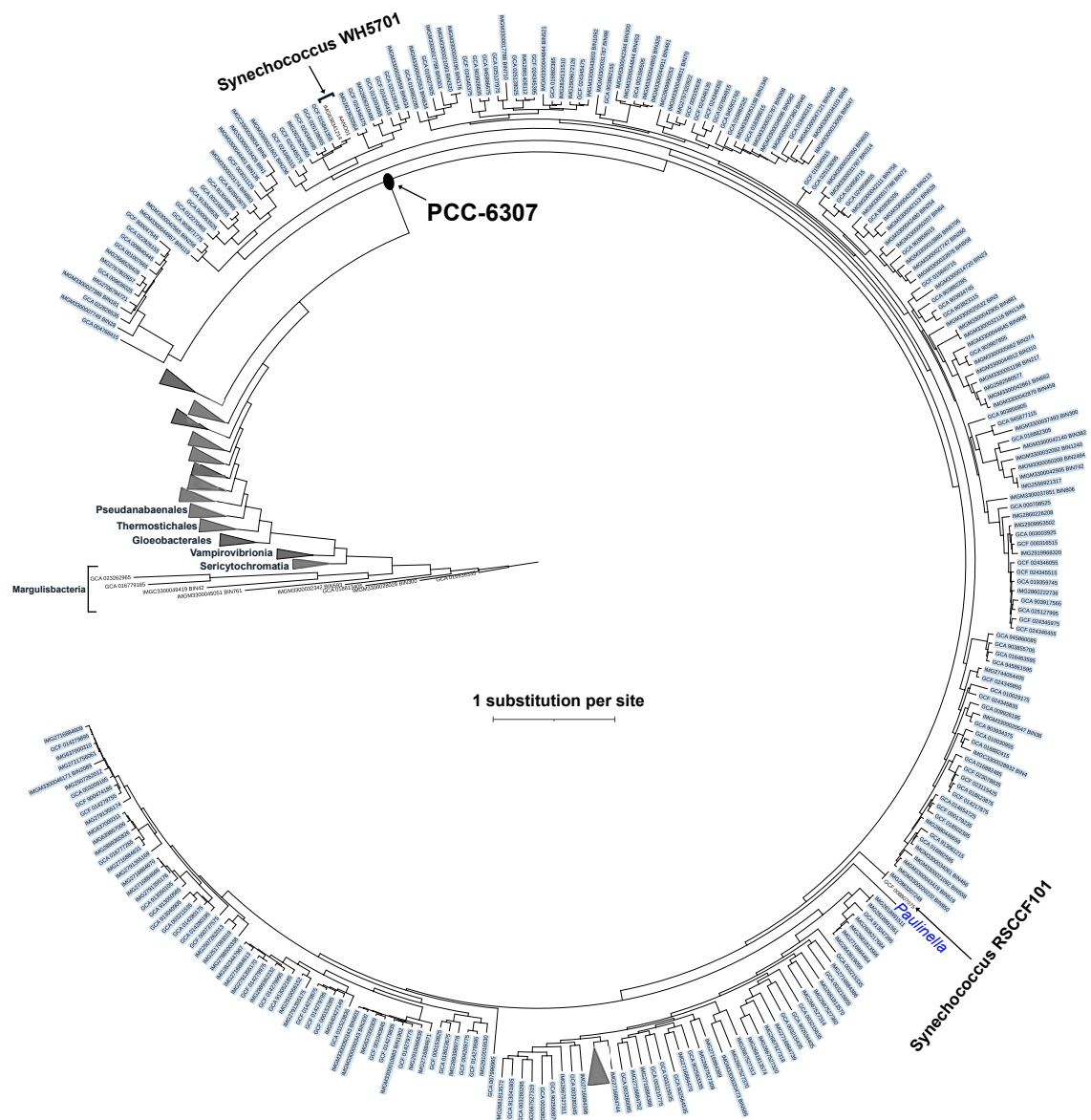
SUPPLEMENTARY INFORMATION



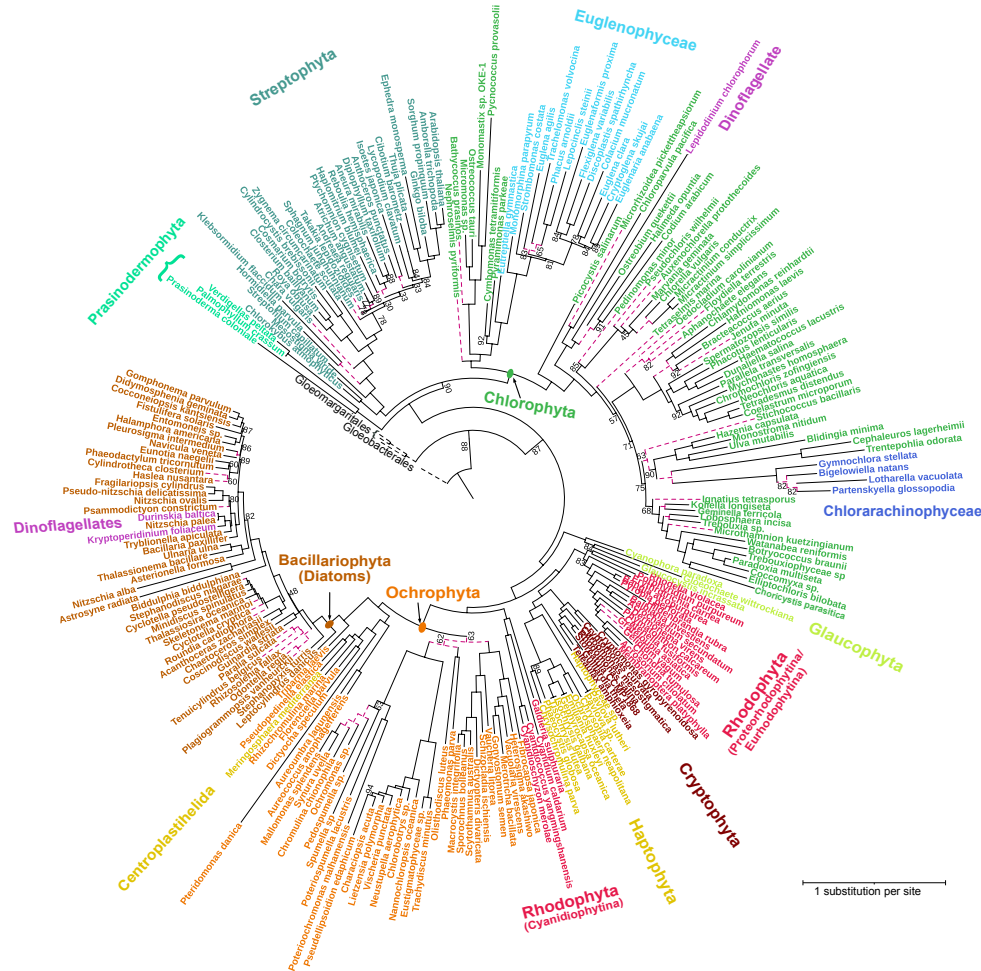
Supplementary Figure 1. The phylogenetic relationships of 1,973 Cyanobacteriota genomes/MAGs were estimated using IQ-TREE, with Margulisbacteria as outgroup to root the tree. Each major clade is separated by color-coded branches: Sericytochromatia (dark red), Vampirovibrionia (dark blue), and Cyanobacteria (green). MAGs identified in this study (labeled as NeLLi2023 in Supplementary Data 2) are highlighted with thicker branches. Bootstrap support values are not shown. The diagram includes four rings: ring 1 indicates the GTDB taxonomic order for each cyanobacterial lineage; ring 2 uses filled circles to indicate isolates; ring 3 provides a bar plot showing UNI56 markers count; ring 4 provides a bar plot displaying CheckM2 completeness.



Supplementary Figure 2. The phylogenetic relationships between cyanobacterial species and plastids from photosynthetic eukaryotes. Cyanobacterial species are represented by dereplicated 259 genomes/MAGs, while plastids are represented by 241 reference plastomes from major photosynthetic eukaryotic groups, and 647 ptMAG contigs containing 16S rRNA genes with at least 10% UNI56 marker. The tree was generated using a concatenated alignment of PLASTID54 marker under the LG+F+I+G4 substitution model with IQ-TREE. Cyanobacterial orders are highlighted with background color, with corresponding color codes for the clade provided in the color key (Bacterial Order). Gloeomargaritales, which is positioned as a sister to all plastids (reference plastomes and ptMAGs), is highlighted with an arrow and labeled. Labels for reference plastomes from diverse photosynthetic eukaryotes are color-coded and indicated in the color key (Plastid Taxonomy). Plastid metagenomes are indicated by labels with black text. Plastids from the major photosynthetic groups are separated with shades of gray and labeled by corresponding species name. The Alveolata clade includes *Chromera* species and four plastomes from Apicomplexa species but excludes dinoflagellates. Dinoflagellate members are separated from Alveolata to indicate their distinct phylogenetic positions.



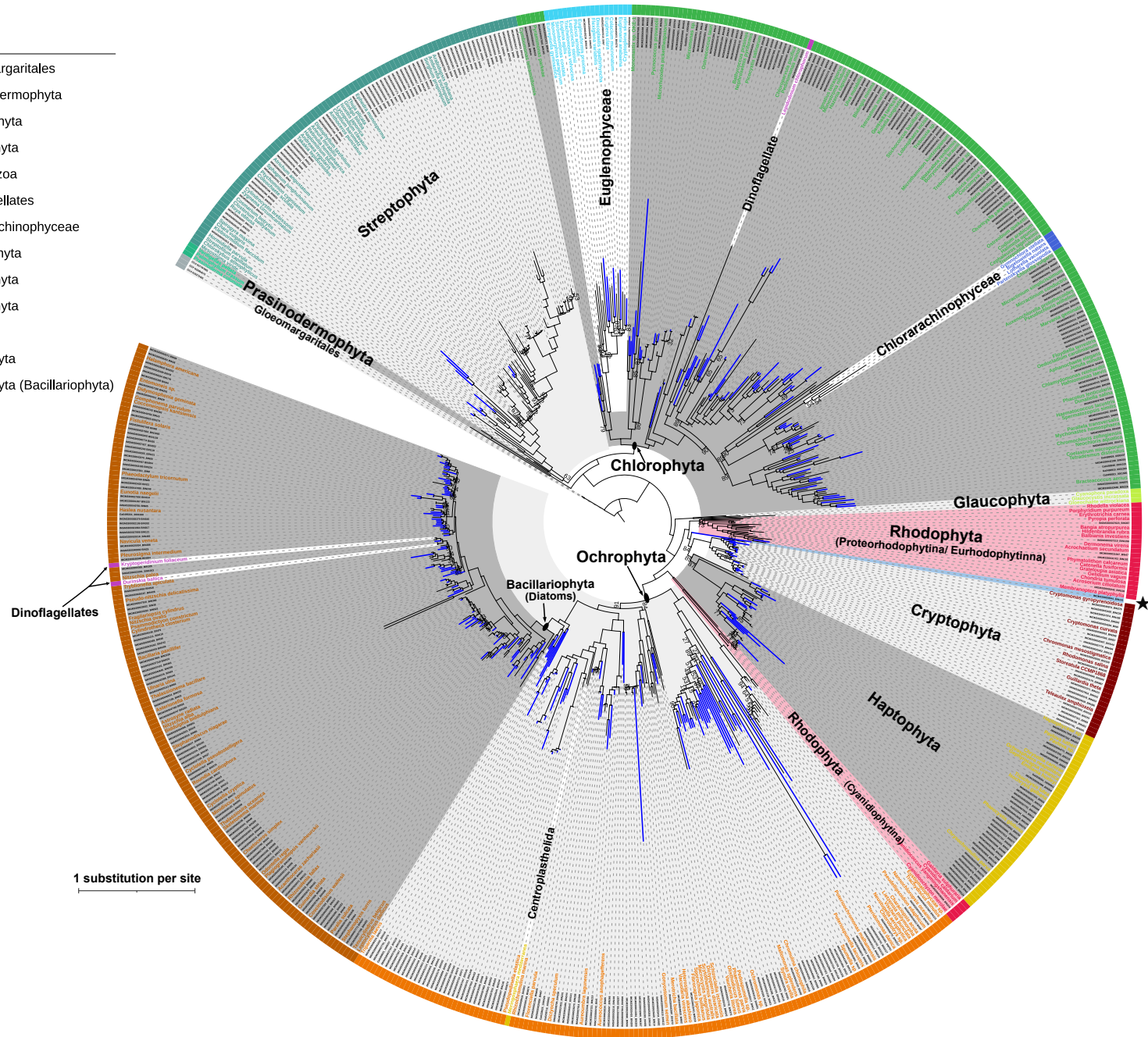
Supplementary Figure 3. Phylogenetic relationship of the *Paulinella chromatophore* plastid compared to cyanobacterial members of the taxonomic order PCC-6307. The tree highlights the relationships of *Synechococcus* sp. WH5701 and *Synechococcus* sp. RSCCF101 in comparison to the *P. chromatophore* plastid. *Synechococcus* RSCCF101 branches off before the diversification of *P. chromatophore* plastid and is phylogenetically closer to *P. chromatophore*. Only members within the taxonomic order PCC-6307 are shown, while other orders within Cyanobacteriota are collapsed. For a complete overview of the relationship among all orders within Cyanobacteriota, please refer to Fig. 1 and Supplementary Fig. 1.



Supplementary Figure 4. Phylogenetic relationships among different photosynthetic eukaryote plastids estimated with concatenated PLASTID54 marker using IQ-TREE. The dataset comprises 238 RefSeq plastomes, three Gloeomargaritales species, and one Gloeobacteriales species, which was used as an outgroup to root the tree. Ultrafast bootstrap values are displayed only for branches with support below 95%. Leaf labels are colored according to the color key. Plastomes whose phylogenetic placement changes upon the addition of ptMAGs (i.e., compared to Fig. 2) are indicated with dotted magenta branches.

Taxonomy

- Gloeomargaritales
- Prasinodermophyta
- Streptophyta
- Chlorophyta
- Euglenozoa
- Dinoflagellates
- Chlorarachniophyceae
- Glaucophyta
- Rhodophyta
- Cryptophyta
- Haptista
- Ochrophyta
- Ochrophyta (Bacillariophyta)

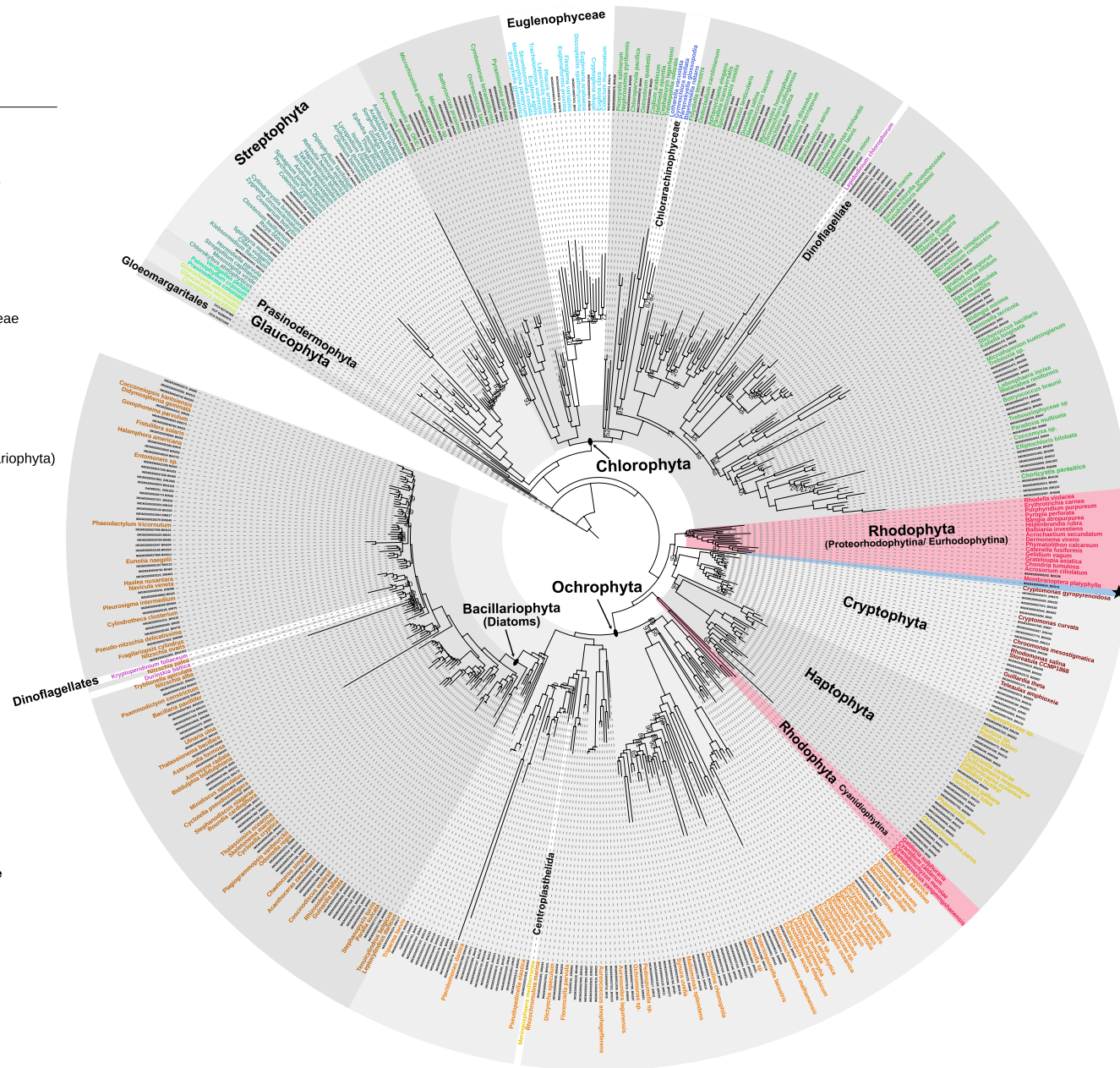


Supplementary Figure 5. Plastid relationship among different photosynthetic eukaryotes. The maximum-likelihood tree was inferred using a concatenated alignment of PLASTID54 marker under the LG+F+I+G4 substitution model, following the nsgtree pipeline and IQ-TREE. Labels for the 238 reference plastids are color-coded as described in the legend, and these clades are shaded in gray for distinction. Secondary and complex plastids within Chlorophyta (Euglenophyceae, Chlorarachinophyceae, and *Lepidodinium* spp.), and in dinoflagellates and Centroplasthelida are excluded from the shades. Rhodophyta clades are shaded in red to highlight two independent origins of secondary red-algal plastids. A single plastid metagenome that branched off before the diversification of Haptophyta and Cryptophyta is highlighted in a blue shade and indicated by a black star. Novel ptMAGs are highlighted with bright blue branches to distinguish them from redundant ptMAGs derived in this study. Ultrafast bootstrap values below 95% are shown under the branches. Reference plastids from the phylum Bacillariophyta are shown in dark orange to distinguish from the rest of other ochrophytes (orange).

Taxonomy

- Gloeomargaritales
- Glaucoophyta
- Prasinodermophyta
- Streptophyta
- Chlorophyta
- Euglenophyceae
- Dinoflagellate
- Chlorarachniophyceae
- Rhodophyta
- Haptista
- Cryptophyta
- Ochrophyta
- Ochrophyta (Bacillariophyta)

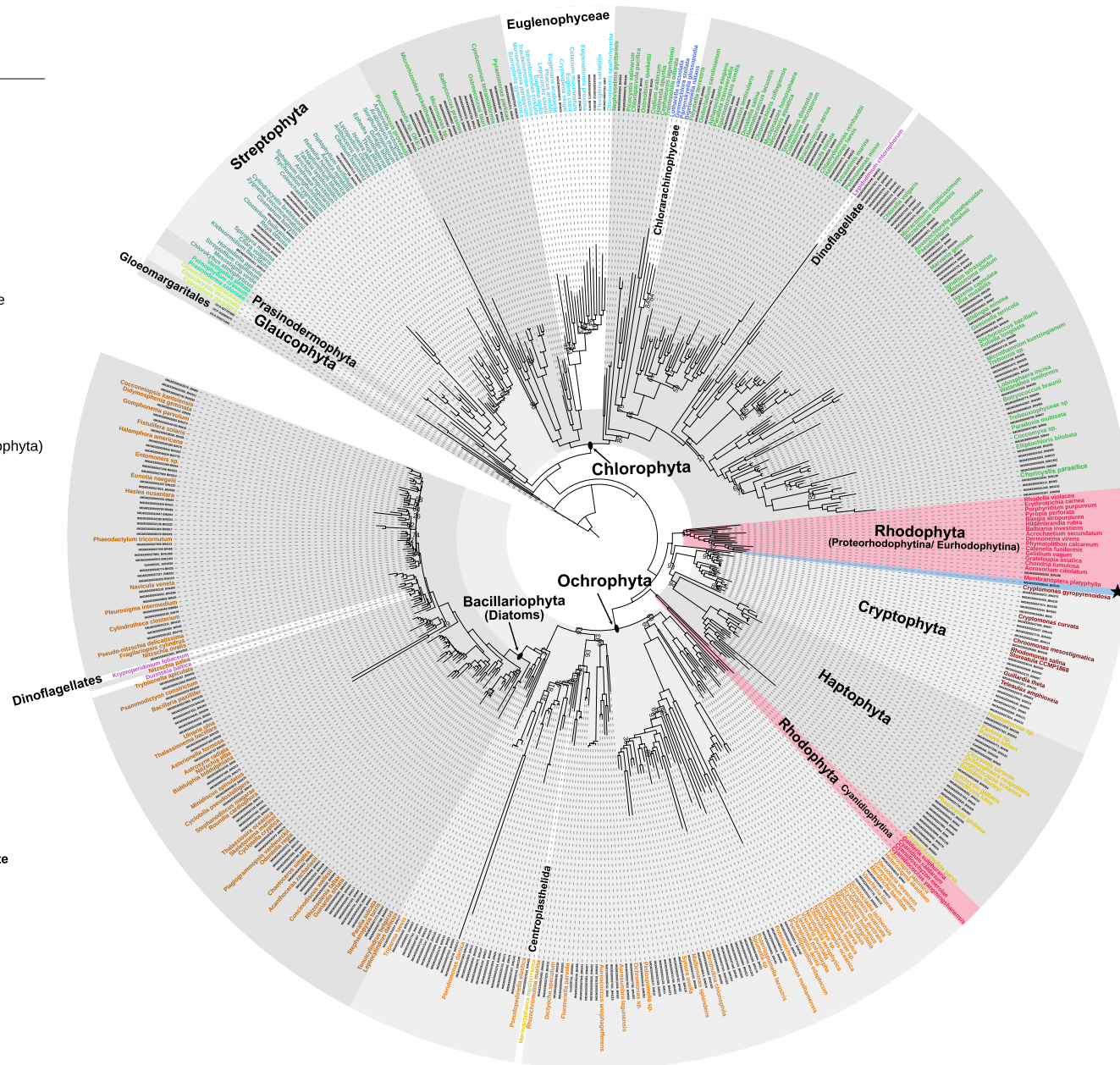
1 substitution per site



Supplementary Figure 6. Maximum-likelihood tree inferred using IQ-TREE under the WAG+C60+PMSF+G4 model, illustrating the evolutionary relationships among photosynthetic eukaryotes based on a 537-taxa dataset. The color key indicates the assigned taxonomy for each clade. A single ptMAG, which diverges before the diversification of Haptophyta and Cryptophyta, is highlighted with a blue background and marked with a black star, indicating its potential significance in understanding early eukaryotic evolution. Ultrafast bootstrap values are displayed for branches with support below 95%.

Taxonomy

- Gloeomargaritales
- Glaucomphyta
- Prasinodermophyta
- Streptophyta
- Chlorophyta
- Euglenophyceae
- Dinoflagellate
- Chlorarachinophyceae
- Rhodophyta
- Cryptophyta
- Haptista
- Ochrophyta
- Ochrophyta (Bacillariophyta)

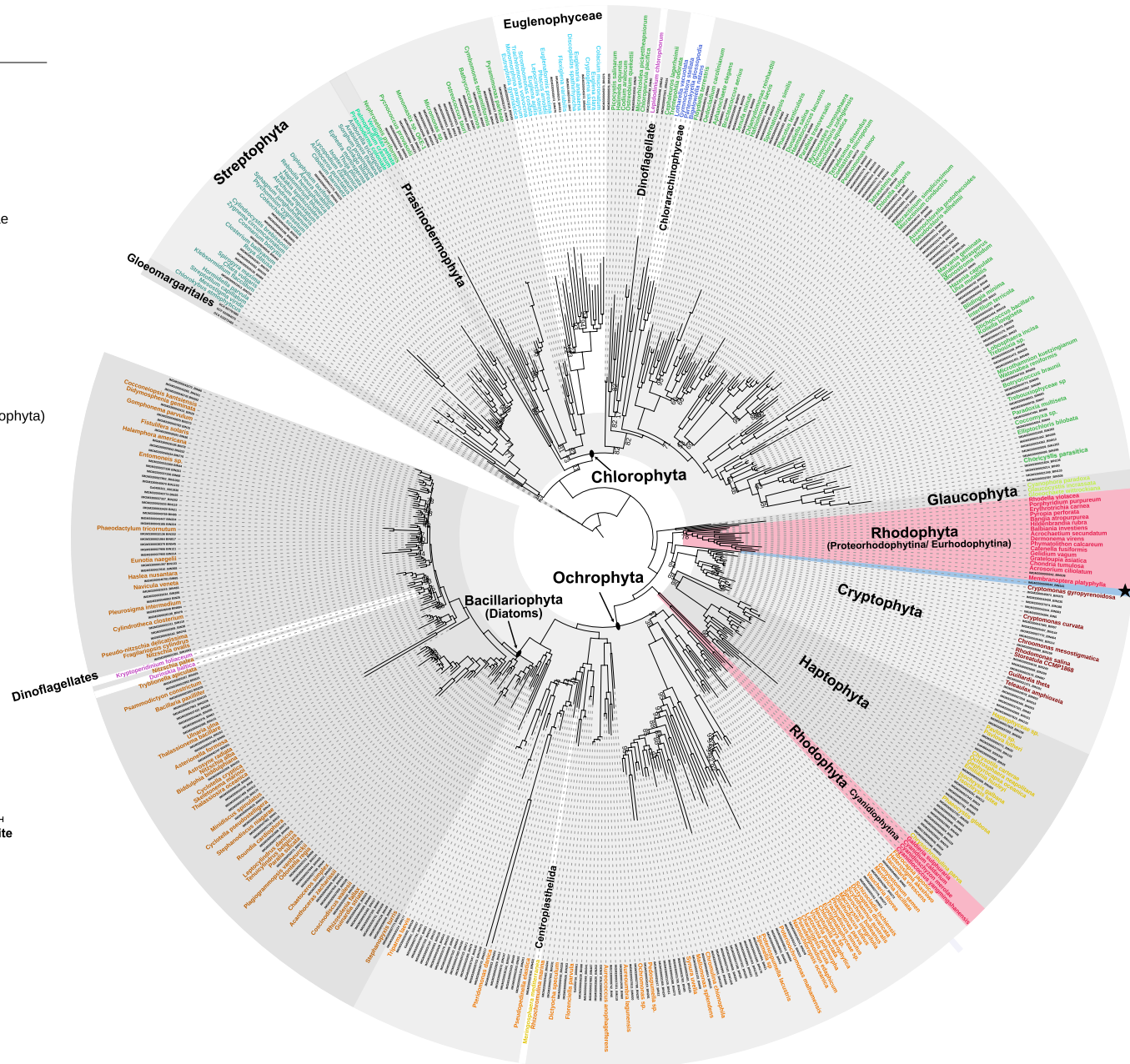


Supplementary Figure 7. Maximum-likelihood tree inferred using IQ-TREE under the LG+C60+PMSF+G4 model, illustrating the evolutionary relationships among photosynthetic eukaryotes based on a 537-taxa dataset. The color key indicates the assigned taxonomy for each clade. A single ptMAG, which diverges before the diversification of Haptophyta and Cryptophyta, is highlighted with a blue background and marked with a black star, indicating its potential significance in understanding early eukaryotic evolution. Ultrafast bootstrap values are displayed for branches with support below 95%.

Taxonomy

- Gloeomargaritales
- Streptophyta
- Prasinodermophyta
- Chlorophyta
- Euglenophyceae
- Chlorarachniophyceae
- Dinoflagellate
- Glaucophyta
- Rhodophyta
- Cryptophyta
- Haptista
- Ochrophyta
- Ochrophyta (Bacillariophyta)

1 substitution per site

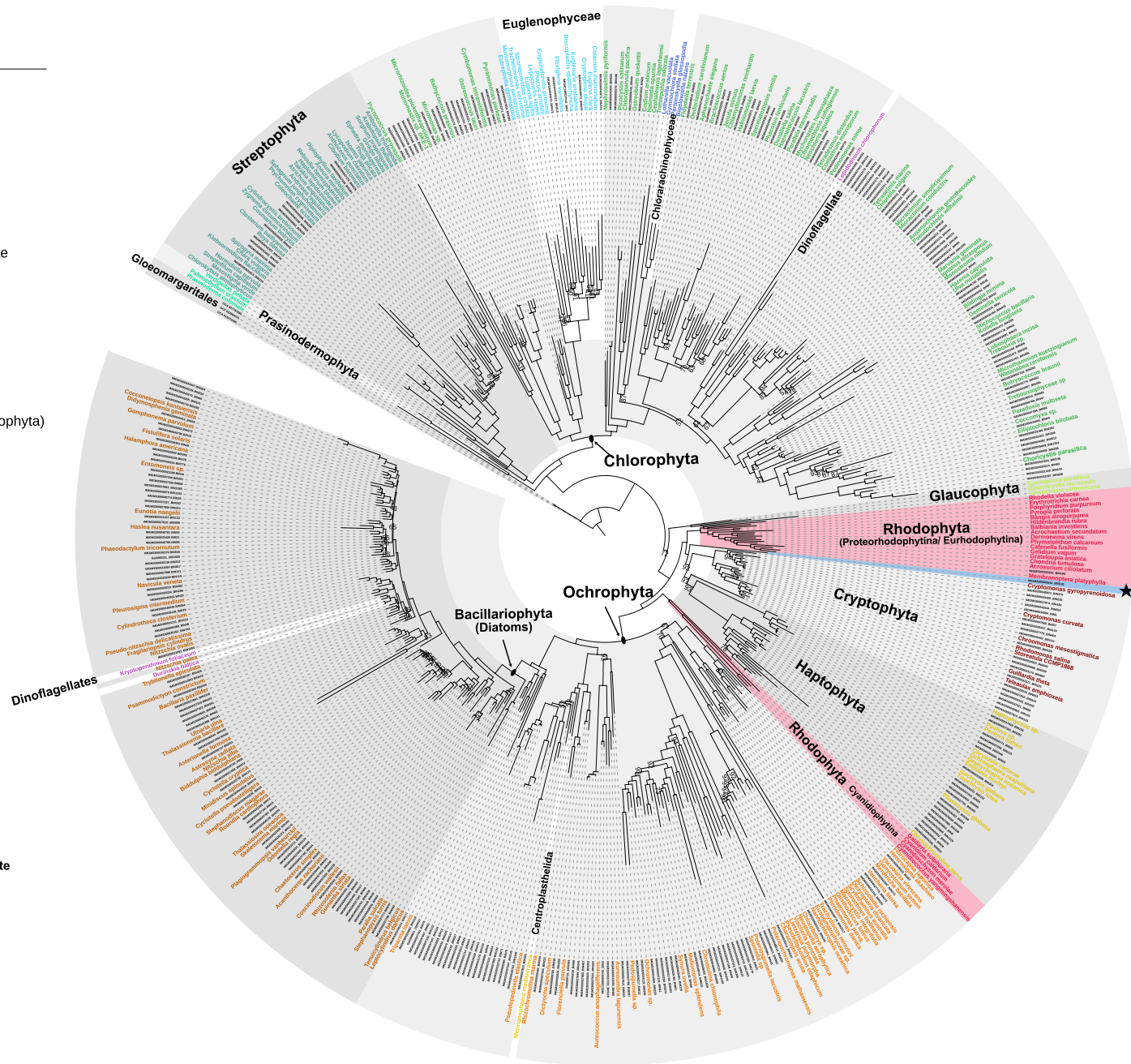


Supplementary Figure 8. Maximum-likelihood tree inferred using IQ-TREE under the LG+PMSF+F+R6 model, illustrating the evolutionary relationships among photosynthetic eukaryotes based on a 537-taxa dataset. The color key indicates the assigned taxonomy for each clade. A single ptMAG, which diverges before the diversification of Haptophyta and Cryptophyta, is highlighted with a blue background and marked with a black star, indicating its potential significance in understanding early eukaryotic evolution. Ultrafast bootstrap values are displayed for branches with support below 95%.

Taxonomy

- Gloeomargaritales
- Prasinodermophyta
- Streptophyta
- Chlorophyta
- Euglenophyceae
- Dinoflagellate
- Chlorarachinophyceae
- Glaucoophyta
- Rhodophyta
- Cryptophyta
- Haptista
- Ochrophyta
- Ochrophyta (Bacillariophyta)

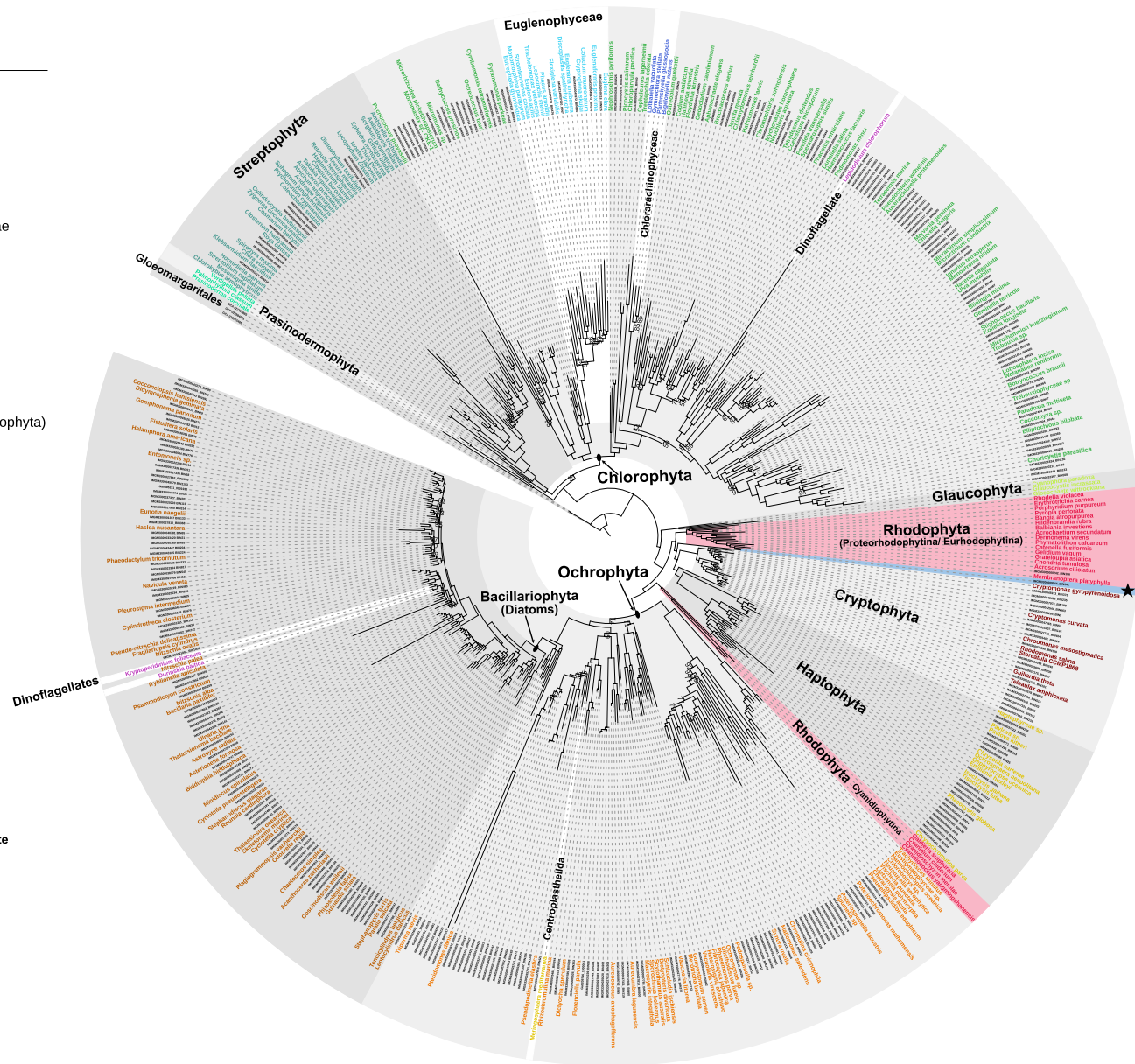
1 substitution per site



Supplementary Figure 9. Maximum-likelihood tree inferred using IQ-TREE under the LG+C60+F+G model, illustrating the evolutionary relationships among photosynthetic eukaryotes based on a 537-taxa dataset. The color key indicates the assigned taxonomy for each clade. A single ptMAG, which diverges before the diversification of Haptophyta and Cryptophyta, is highlighted with a blue background and marked with a black star, indicating its potential significance in understanding early eukaryotic evolution. Ultrafast bootstrap values are displayed for branches with support below 95%.

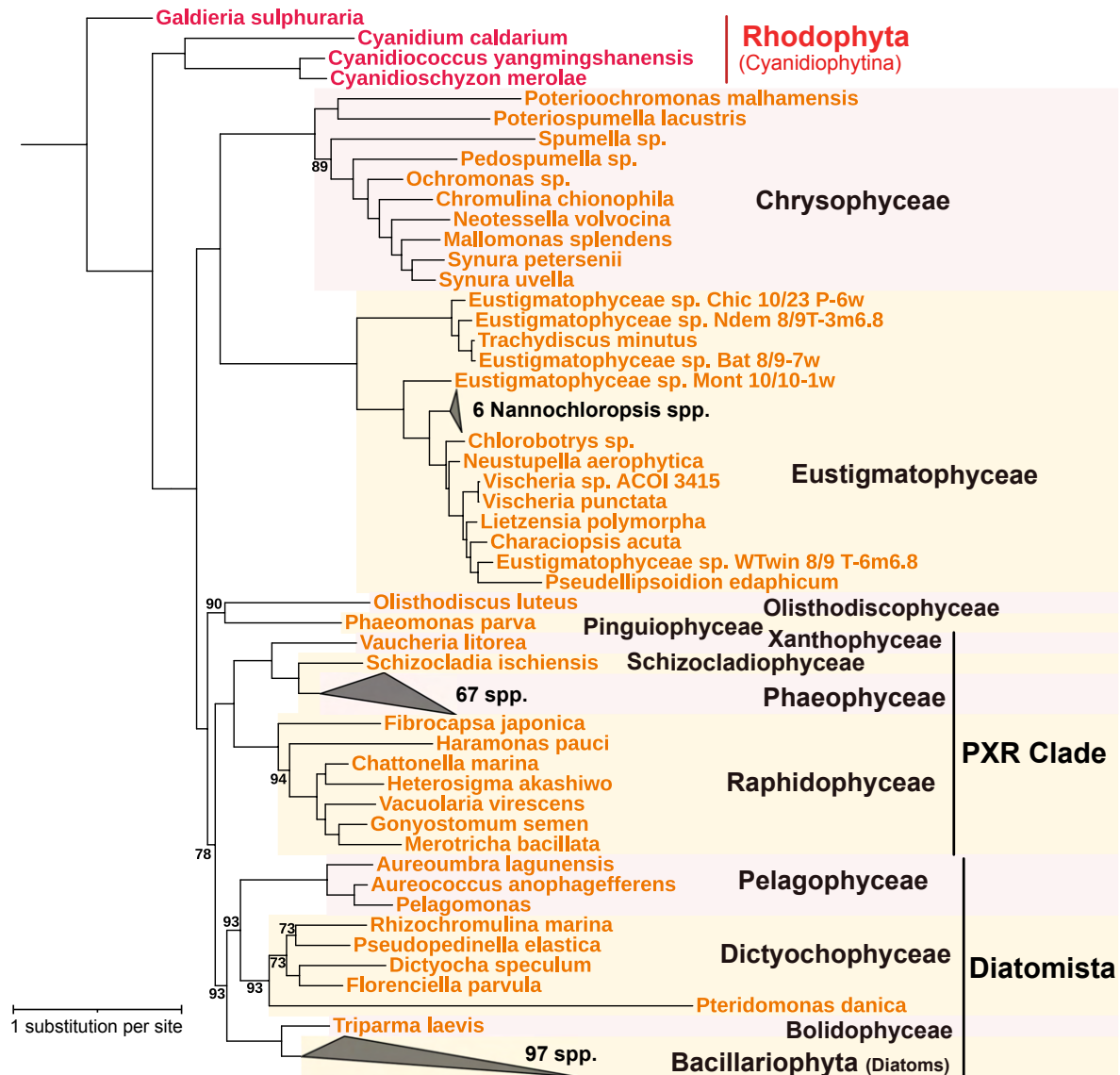
- Taxonomy**
- Gloeomargaritales
 - Prasinodermophyta
 - Streptophyta
 - Chlorophyta
 - Euglenophyceae
 - Chlorarachniophyceae
 - Dinoflagellate
 - Glaucophyta
 - Rhodophyta
 - Cryptophyta
 - Haptista
 - Ochrophyta
 - Ochrophyta (Bacillariophyta)

1 substitution per site

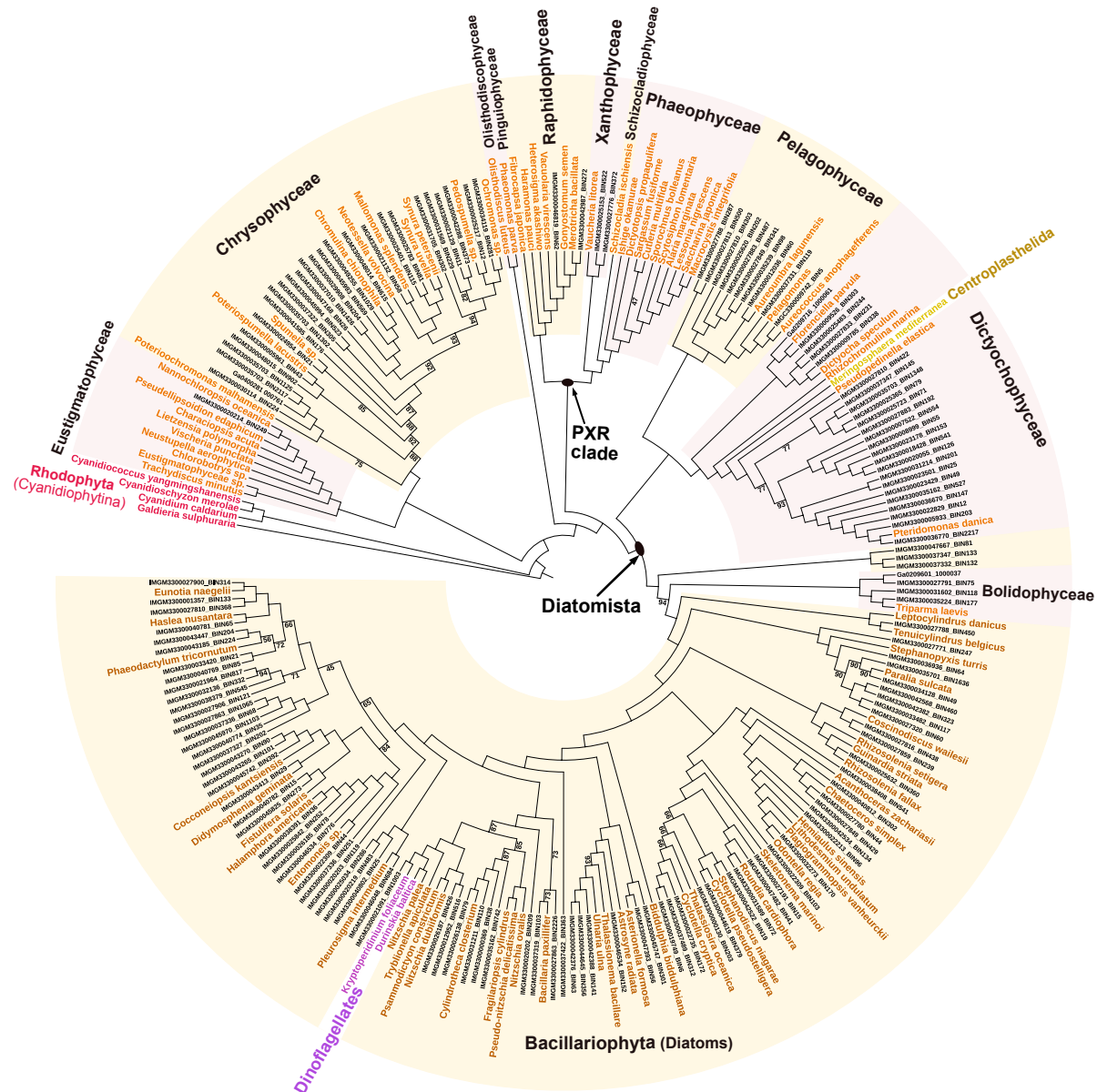


Supplementary Figure 10. Maximum-likelihood tree inferred using IQ-TREE under the LG+C60+F+R6 model, illustrating the evolutionary relationships among photosynthetic eukaryotes based on a 537-taxa dataset. The color key indicates the assigned taxonomy for each clade. A single ptMAG, which diverges before the diversification of Haptophyta and Cryptophyta, is highlighted with a blue background and marked with a black star, indicating its potential significance in understanding early eukaryotic evolution. Ultrafast bootstrap values are displayed for branches with support below 95%.

SUPPLEMENTARY INFORMATION



Supplementary Figure 11. Phylogenetic relationships among reference Ochrophyta plastids. A maximum-likelihood tree was constructed with a concatenated alignment of 54 plastid proteins, applying the LG+F+I+G4 substitution model. The analysis included a total of 213 Ochrophyta reference plastids, with four Rhodophyta plastids serving as outgroup. Only ultrafast bootstrap support values less than 95% are displayed, either above or below the corresponding branches. Clades containing a large number of taxa were collapsed and labeled accordingly. Major groups within Ochrophyta were distinguished with alternating pink or yellow shading and labeled accordingly.



Supplementary Figure 12. A cladogram depicting the phylogenetic relationships of secondary red-algal plastids in Ochrophyta and other photosynthetic eukaryotes. The relationships were estimated using a minimum of 20% PLASTID54 marker using IQ-TREE and the LG+F+I+G4 substitution model. Red algal species (subphylum Cyanidiophytina) was used to root the tree. Only the branches with ultrafast bootstrap support values less than 95% are shown. Complete plastomes available in the NCBI RefSeq/GenBank are labeled with species name, while the ptMAGs are shown in black font. Major classes within Ochrophyta are distinguished with alternating pink or yellow shading and labeled. Plastids from Centropastheliida and dinoflagellates are excluded from the shaded area. Reference ochrophyte species from the phylum Bacillariophyta are highlighted in darker orange compared to species from other ochrophytes.

REFERENCES

1. Soo, R. M., Hemp, J. & Hugenholtz, P. Evolution of photosynthesis and aerobic respiration in the cyanobacteria. *Free Radical Biology and Medicine* **140**, 200–205 (2019).
2. Matheus Carnevali, P. B. *et al.* Hydrogen-based metabolism as an ancestral trait in lineages sibling to the Cyanobacteria. *Nat Commun* **10**, 463 (2019).
3. Tan, S. Exploring the Origins and Evolution of Oxygenic and Anoxygenic Photosynthesis in Deeply Branched Cyanobacteriota.
4. Wicke, S., Schneeweiss, G. M., dePamphilis, C. W., Müller, K. F. & Quandt, D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol* **76**, 273–297 (2011).
5. Lee, J. *et al.* Parallel evolution of highly conserved plastid genome architecture in red seaweeds and seed plants. *BMC Biol* **14**, 75 (2016).
6. Wang, D. *et al.* Transfer of Chloroplast Genomic DNA to Mitochondrial Genome Occurred At Least 300 MYA. *Molecular Biology and Evolution* **24**, 2040–2048 (2007).
7. Sloan, D. B. & Wu, Z. History of Plastid DNA Insertions Reveals Weak Deletion and AT Mutation Biases in Angiosperm Mitochondrial Genomes. *Genome Biology and Evolution* **6**, 3210–3221 (2014).
8. Gandini, C. L. & Sanchez-Puerta, M. V. Foreign Plastid Sequences in Plant Mitochondria are Frequently Acquired Via Mitochondrion-to-Mitochondrion Horizontal Transfer. *Sci Rep* **7**, 43402 (2017).
9. Ponce-Toledo, R. I. *et al.* An Early-Branching Freshwater Cyanobacterium at the Origin of Plastids. *Current Biology* **27**, 386–391 (2017).
10. Moore, K. R. *et al.* An Expanded Ribosomal Phylogeny of Cyanobacteria Supports a Deep Placement of Plastids. *Front. Microbiol.* **10**, (2019).
11. Sánchez-Baracaldo, P., Raven, J. A., Pisani, D. & Knoll, A. H. Early photosynthetic eukaryotes inhabited low-salinity habitats. *Proc. Natl. Acad. Sci. U.S.A.* **114**, (2017).
12. Nowack, E. C. M., Melkonian, M. & Glöckner, G. Chromatophore Genome Sequence of *Paulinella* Sheds Light on Acquisition of Photosynthesis by Eukaryotes. *Current Biology* **18**, 410–418 (2008).
13. Derelle, R., López-García, P., Timpano, H. & Moreira, D. A Phylogenomic Framework to Study the Diversity and Evolution of Stramenopiles (=Heterokonts). *Mol Biol Evol* **33**, 2890–2898 (2016).
14. Di Franco, A., Baurain, D., Glöckner, G., Melkonian, M. & Philippe, H. Lower Statistical Support with Larger Data Sets: Insights from the Ochrophyta Radiation. *Molecular Biology and Evolution* **39**, msab300 (2022).
15. Barcytè, D. *et al.* Olisthodiscus represents a new class of Ochrophyta. *Journal of Phycology* **57**, 1094–1118 (2021).
16. Azuma, T. *et al.* An Enigmatic Stramenopile Sheds Light on Early Evolution in Ochrophyta Plastid Organellogenesis. *Molecular Biology and Evolution* **39**, msac065 (2022).

17. Terpis, K. X. *et al.* Multiple plastid losses within photosynthetic stramenopiles revealed by comprehensive phylogenomics. *Current Biology* **35**, 483-499.e8 (2025).
18. Cho, A., Lax, G. & Keeling, P. J. Phylogenomic analyses of ochrophytes (stramenopiles) with an emphasis on neglected lineages. *Molecular Phylogenetics and Evolution* **198**, 108120 (2024).
19. Sørensen, M. E. S. *et al.* A novel kleptoplastidic symbiosis revealed in the marine centrophelid Meringosphaera with evidence of genetic integration. *Current Biology* **33**, 3571-3584.e6 (2023).
20. Tajima, N. *et al.* Sequencing and analysis of the complete organellar genomes of Parmales, a closely related group to Bacillariophyta (diatoms). *Curr Genet* **62**, 887–896 (2016).
21. Jeong, Y. & Lee, J. Comparative analysis of organelle genomes provides conflicting evidence between morphological similarity and phylogenetic relationship in diatoms. *Front. Mar. Sci.* **10**, (2024).
22. Imanian, B., Pombert, J.-F. & Keeling, P. J. The Complete Plastid Genomes of the Two ‘Dinotoms’ *Durinskia baltica* and *Kryptoperidinium foliaceum*. *PLoS ONE* **5**, e10711 (2010).
23. Chesnick, J. M., Kooistra, W. H. C. F., Wellbrock, U. & Medlin, L. K. Ribosomal RNA Analysis Indicates a Benthic Pennate Diatom Ancestry for the Endosymbionts of the Dinoflagellates *Peridinium foliaceum* and *Peridinium balticum* (Pyrrhophyta). *Journal of Eukaryotic Microbiology* **44**, 314–320 (1997).
24. Tamura, M., Shimada, S. & Horiguchi, T. *Galeidinium Rugatum* Gen. Et Sp. Nov. (dinophyceae), a New Coccoid Dinoflagellate with a Diatom Endosymbiont. *Journal of Phycology* **41**, 658–671 (2005).
25. Crowell, R. M., Nienow, J. A. & Cahoon, A. B. The complete chloroplast and mitochondrial genomes of the diatom *Nitzschia palea* (Bacillariophyceae) demonstrate high sequence similarity to the endosymbiont organelles of the dinotom *Durinskia baltica*. *Journal of Phycology* **55**, 352–364 (2019).
26. Yamada, N., Sym, S. D. & Horiguchi, T. Identification of Highly Divergent Diatom-Derived Chloroplasts in Dinoflagellates, Including a Description of *Durinskia kwazulunatalensis* sp. nov. (Peridiniales, Dinophyceae). *Molecular Biology and Evolution* **34**, 1335–1351 (2017).
27. Gastineau, R. *et al.* Morphological and molecular identification reveals that waters from an isolated oasis in Tamanrasset (extreme South of Algerian Sahara) are colonized by opportunistic and pollution-tolerant diatom species. *Ecological Indicators* **121**, 107104 (2021).
28. Lemieux, C., Otis, C. & Turmel, M. Chloroplast phylogenomic analysis resolves deep-level relationships within the green algal class Trebouxiophyceae. *BMC Evol Biol* **14**, 211 (2014).
29. Mikhailyuk, T. I. *et al.* New Streptophyte Green Algae from Terrestrial Habitats and an Assessment of the Genus *Interfilum* (klebsormidiophyceae, Streptophyta). *Journal of Phycology* **44**, 1586–1603 (2008).