

Global metagenomics reveals plastid diversity and unexplored algal lineages

Received: 26 March 2025

Accepted: 14 January 2026

Published online: 30 January 2026

 Check for updates

Bikash Shrestha¹✉, Miguel F. Romero¹, Juan C. Villada¹, Plastid MAG Consortium*, Crysten E. Blaby-Haas^{1,2}✉ & Frederik Schulz¹✉

Photosynthetic organelles in eukaryotes originated through primary endosymbiosis with a cyanobacterium, an event that profoundly shaped the evolutionary landscape of the eukaryotic tree of life. Primary plastids in Archaeplastida, especially in cultivable plants and algae, contribute most to known plastid diversity. Secondary and higher-order endosymbiosis, involving eukaryotic hosts and algal endosymbionts, further spread photosynthesis among protists within the CASH lineages (Cryptophyta, Alveolata, Stramenopila, and Haptophyta). Despite various hypotheses explaining secondary plastid evolution and distribution, empirical support remains limited. Here, we employ cultivation-independent global metagenomics to expand plastid diversity and investigate plastid origins. We capture 1,027 plastid sequences, including 300 novel sequences belonging to previously unsequenced plastids and representing yet-to-be described microeukaryotes. This includes a new lineage that offers insights into plastid evolution in haptophytes and cryptophytes. Our results confirm that Archaeplastida plastids originate from an early branching cyanobacterial lineage closely related to Gloeomargaritales and identify the closest extant relative of *Paulinella* plastids. Additionally, our findings suggest two independent origins of secondary red-algal plastids, contributing to plastid diversity in CASH lineages and challenging the prevailing model of single secondary plastid origin. Our study highlights the importance of metagenomic data in uncovering biological diversity and advancing understanding of plastid relationships across photosynthetic eukaryotes.

Oxygenic photosynthesis is a crucial metabolic process that has profoundly shaped life on Earth. While oxygenic photosynthesis (referred herein as photosynthesis) first evolved in cyanobacteria, autotrophy has spread between kingdoms, from bacteria to eukaryotes, and across Eukarya through endosymbiosis, leading to the emergence of taxonomically distinct photosynthetic eukaryotes. The resulting plants and algae play essential roles in terrestrial and aquatic ecosystems, with plastids serving as their primary organelles responsible for generating molecular oxygen and converting CO₂ to sugars.

Plastids originated through primary endosymbiosis, where a heterotrophic eukaryote engulfed a now extinct cyanobacterium that eventually evolved into a photosynthetic organelle. To date, while only two confirmed cases of primary endosymbiosis-derived plastids have been documented, other intriguing examples of cyanobacteria evolving from endosymbionts into organelles in marine eukaryotes have been described recently. Notable examples include nitroplasts in *Braarudosphaera bigelowii*¹ and diazoplasts in *Epithemia clementina*², both enabling endosymbiotic nitrogen fixation and highlighting

¹DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA. ²The Molecular Foundry, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA. *A full list of author affiliations appears at the end of the paper.

✉ e-mail: bshrestha@lbl.gov; cblaby@lbl.gov; fschulz@lbl.gov

ongoing cyanobacterial integration in eukaryotes. The first event, occurring approximately 1.5 billion years ago, involved a eukaryote and β -cyanobacterium endosymbiont, leading to the formation of plastids in Archaeplastida. This group comprises green algae and land plants (Viridiplantae), glaucophytes (Glaucophyta), and red algae (Rhodophyta), as well as non-photosynthetic members such as Rhodophelia and Picozoa, which are closely related to Rhodophyta^{3–9}. Despite some uncertainty about the cyanobacterial ancestor of these plastids, recent research indicates that the ancestor is related to early diverging lineages, such as Gloeomargaritales^{10,11}. The monophyly of Archaeplastida plastids is generally well-supported by conserved gene content, synteny, presence of inverted repeats, and high sequence similarity of 16S rRNA genes^{12,13}. However, studies challenging this monophyly are not uncommon and phylogenetic relationships among the Archaeplastida group remain debated^{12–14}. The second primary endosymbiosis took place much more recently, between 90–140 million years ago involving *Paulinella* (Rhizaria), an amoeba that acquired an α -cyanobacterium related to the *Synechococcus/Prochlorococcus* clade^{15–17}. Unlike Archaeplastida plastids, the *Paulinella* plastid retains more of its cyanobacterial genome (-1 Mb vs. <200 kb) and has a distinct membrane structure¹⁸.

Beyond primary endosymbiosis, the genesis of new algal lineages has occurred through secondary or serial endosymbiosis, where non-photosynthetic eukaryotes have engulfed an alga, forming secondary/complex plastids^{13,14}. Algae such as chlorarachniophytes, euglenophytes, and some dinoflagellates acquired secondary plastids through the endosymbiosis of distinct green algae, and the evolutionary origins of these independent events are relatively well understood¹⁹. However, the evolutionary origins and spread of red-alga-derived plastids among algal groups, including Cryptophyta, Haptophyta, Ochrophyta (photosynthetic Stramenopila), and Alveolata (comprising chrompodellids, apicomplexans, and dinoflagellates), collectively known as CASH lineages, remain contentious. Several hypotheses have been proposed to account for the presence of red-alga-derived plastids in distinct lineages of protists. The Chromalveolate hypothesis, first proposed by Cavalier-Smith²⁰, suggests that the CASH lineages share a common ancestor bearing secondary plastids, which would require multiple plastid losses to explain the existence of non-photosynthetic protist lineages. However, there is little empirical evidence to support such widespread plastid loss^{21,22}. Several alternative hypotheses have been proposed to explain the distribution of plastids in CASH lineages. For example, the Rhodoplex hypothesis posits that secondary/complex red-algal plastids originated within one CASH lineage (likely Cryptophyta) and subsequently spread via tertiary or higher-order endosymbiosis (serial endosymbiosis^{22–26}). Despite these competing hypotheses, empirical evidence supporting them remains limited²⁷.

Our understanding of plastid evolution has primarily relied on complete plastid genomes (plastomes) obtained from isolates or culture collections. As a result, currently available plastid sequences likely represent only a fraction of existing diversity, as many environments remain underexplored. The use of metagenome-assembled genomes (MAGs) has significantly expanded our understanding of microbial diversity, since sampled organisms do not need to be isolated or cultured. Research leveraging MAGs has been instrumental in discovering new lineages across the major domains of life and elucidating their evolutionary history and relationships^{28–30}. The analysis of plastid MAGs (ptMAGs) has the potential to uncover previously unidentified plastid lineages, suggest the existence of new algal lineages, and offer deeper insight into our current understanding of plastid evolution.

Here, we aim to investigate plastid diversity and evolution using metagenomic approaches. We generate a dataset of ptMAGs and expand the Cyanobacteriota MAGs dataset using publicly available metagenomic data in the Integrated Microbial Genomes and Microbiomes (IMG/M) database^{31,32}. By integrating these MAGs with

reference plastomes and Cyanobacteriota genomes, we reconstruct plastid evolutionary relationships. Our analyses reveal the closest cyanobacterial relatives of *Paulinella* plastids, confirm that Archaeplastida plastids originate from an early branching cyanobacterial lineage, and provide evidence for two independent events involving engulfment of a red alga. Additionally, we report multiple plastid lineages that likely represent previously undiscovered algal lineages and provide complete or nearly complete genomic sequences for plastids previously identified solely through rRNA sampling, including a novel plastid lineage that offers new insights into plastid evolution within haptophytes and cryptophytes.

Results and discussion

Expanding the cyanobacterial taxonomy for tracing plastid origins

To investigate the origin of plastids, we first expanded the phylogenetic framework of Cyanobacteriota taxonomy. With the extended MAG binning, a total of 1973 genomes/MAGs belonging to Cyanobacteria (referred to as Cyanobacteriia in the GTDB), Vampirovibrionia, and Sericytochromatia were analyzed. This dataset included 438 new metagenomic bins along with 1,539 genomes/metagenomes available in two public databases: the GTDB and the IMG/M databases. Among the 438 new metagenomic bins, GTDB taxonomic classification assigned eight, 84, and 346 bins into Sericytochromatia, Vampirovibrionia, and Cyanobacteria, respectively (Supplementary Data 1). Two groups, Sericytochromatia and Vampirovibrionia, were exclusively represented by metagenomic data, while the Cyanobacteria included metagenomic bins as well as 496 genomes assembled from isolates (Supplementary Data 1). The inclusion of 438 new metagenomic bins within Cyanobacteriota increased the overall phylogenetic diversity by 28%. Our phylogenetic analyses strongly support the monophyly of the photosynthetic clade Cyanobacteria and the non-photosynthetic clades, Sericytochromatia and Vampirovibrionia, with Sericytochromatia positioned at the base of the clade and sister to Vampirovibrionia and Cyanobacteria (Fig. 1a and Supplementary Fig. 1). This topology is consistent with published studies^{33,34}. For detailed description of the Cyanobacteriota phylogeny, see Supplementary Note 1.

To determine the phylogenetic placement of plastids within the Cyanobacteriota taxonomy, we conducted a series of phylogenetic reconstructions by varying the datasets and the phylogenetic markers used for the supermatrix alignments (Supplementary Data 2; detailed description of the iterated analyses can be found in methods). In all phylogenetic reconstructions, regardless of increasing taxon sampling, swapping the phylogenetic markers, or modifying cut-off parameters for tree reconstructions, we observed that plastomes and ptMAGs grouped together in one of two strongly supported monophyletic clades (UFBoot 100%) within the Cyanobacterial tree (Fig. 1c and Supplementary Fig. 2). The first clade represents the primary plastids of *Paulinella* and confirms relatedness to an ancestor of the α -cyanobacterial clade containing *Synechococcus* species^{15,16}. However, while *Synechococcus* sp. WH 5701 was previously assumed to be the closest extant relative¹⁶, our results place the *Paulinella* plastid after the diversification of *Synechococcus* sp. RSCCF101 (Fig. 1c, Supplementary Fig. 3). The *Synechococcus* sp. RSCCF101 and WH 5701 share ~79% nucleotide identity but belong to distinct phylogenetic subclades within the larger *Synechococcales* lineage (referred to as “order PCC-6307” in GTDB taxonomy; Fig. 1d, Supplementary Fig. 3). The second clade contains all plastomes and ptMAGs other than *Paulinella* and is sister to the cyanobacterial order Gloeomargaritales (Fig. 1c and Supplementary Fig. 2). This finding supports the current understanding that primary plastids in Archaeplastida originate from a deeply branching cyanobacterium closely related to *Gloeomargarita* lineage^{10,11,35} and does not support any alternative scenarios^{36,37}.

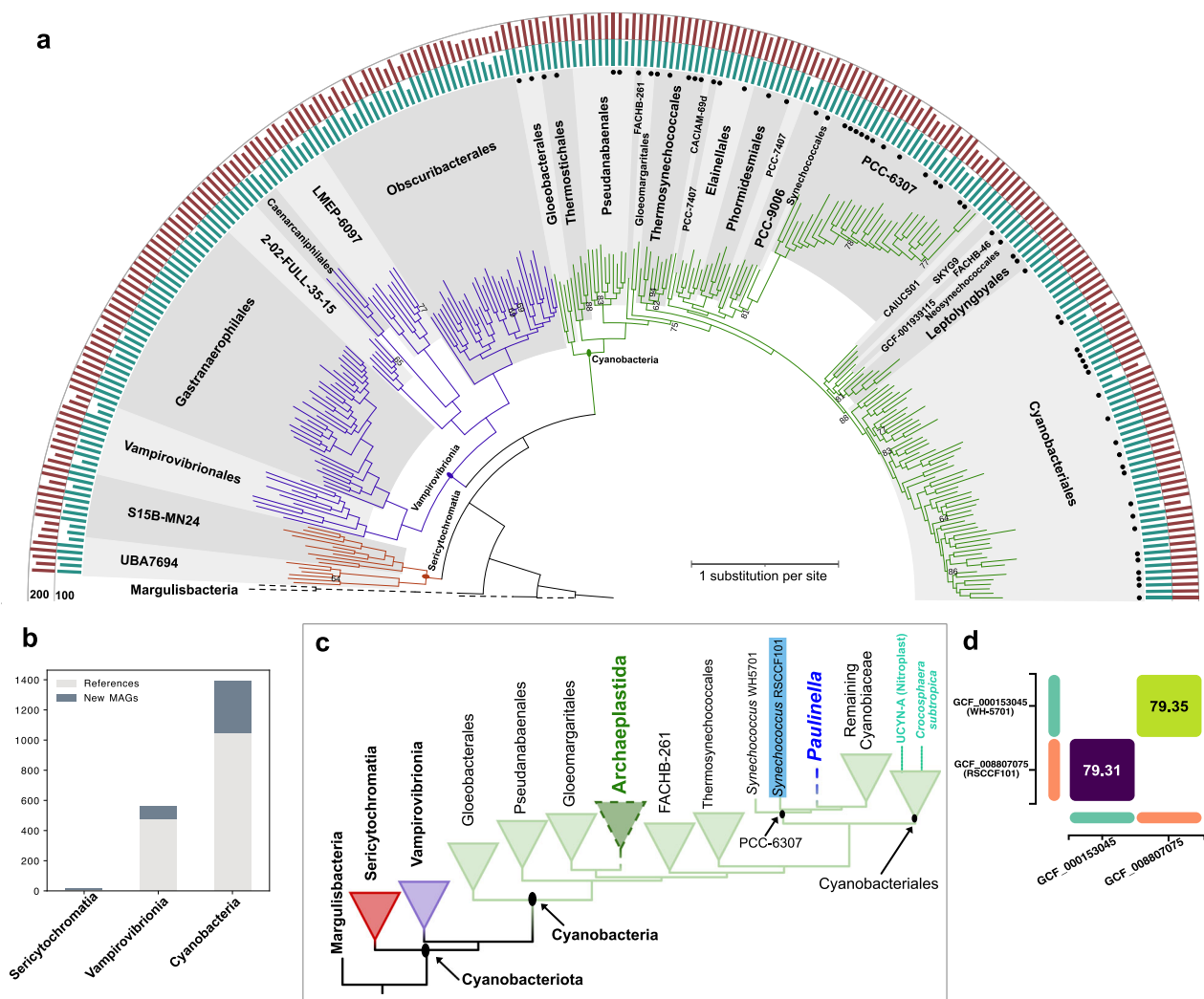


Fig. 1 | Phylogenetic placement of primary plastids within the cyanobacterial lineage. **a** A maximum-likelihood tree of the dereplicated Cyanobacteriota dataset, constructed using concatenated alignment of UNIS6 marker with IQ-TREE. Branches are color-coded: deep red for Sericytochromatia, violet for Vampiromivirionia, and green for Cyanobacteria. The tree is rooted using Margulisbacteria. Taxonomic orders from GTDB are labeled and distinguished with different shades of gray. Only ultrafast bootstrap values < 95% are shown, either above or below the branches. Isolates are marked with filled black circles at the end of the leaves, distinguishing them from MAGs. The outside bar plots represent the count of UNIS6 marker proteins (dark cyan) and CheckM2 completeness (dark red). **b** A stacked histogram showing the count of publicly available and newly generated MAGs for Cyanobacteriota in this study. **c** A simplified illustration showing the phylogenetic

positions of primary plastids (Archaeplastida and *Paulinella*) within the cyanobacterial phylogeny. Selected cyanobacterial taxonomic orders are shown as collapsed triangles to ease visualization of plastid origins, which are highlighted by dotted clade and branch lines. The *Synechococcus* sp. RSCCF101, which is phylogenetically closer to the *Paulinella* plastid, is highlighted with a blue background. Within Cyanobacteriales, UCYN-A represents *Candidatus Atelocyanobacterium thalassa* (nitroplast), an N_2 -fixing endosymbiont in the haptophyte (*Braarudosphaera bigelowii*), while *Crocosphaera subtropica* is identified as the closest free-living relative of another N_2 -fixing endosymbiont, diazoplast, found in the diatom *Epithemia clementina*. **d** Whole-genome ANI comparison between *Synechococcus* sp. WH-5701 and *Synechococcus* sp. RSCCF101.

An extended taxonomic framework of plastids tracing novel ptMAG lineages

Next, we investigated the relationships among plastids in photosynthetic eukaryotes and the impact of our novel ptMAGs on established plastid phylogenetic relationships. The resulting trees revealed a consistent topology of major photosynthetic eukaryotes within green and red lineages, such as Streptophyta, Chlorophyta, Rhodophyta, Cryptophyta, Haptophyta, and Ochrophyta. Within the green lineage, Prasinodermophyta emerged as an early diverging clade and sister to Chlorophyta and Streptophyta with varying UFBoot support values (93–100%), substantiating their classification as a new phylum within Viridiplantae^{38,39}. Streptophytes and chlorophytes were strongly supported as monophyletic groups, and lineages with secondary/

complex plastids, such as Euglenophyceae, chlorarachniophytes, and the dinoflagellates species *Lepidodinium*, which acquired plastids from unrelated green algae, were included within chlorophytes. Similarly, Ochrophyta also included kleptoplasty-derived complex plastids in two unrelated lineages: a marine Centroplasthelida *Meringosphaera mediterranea* (a Haptista), and two dinoflagellate species, *Kryptoperidinium foliaceum* and *Durinskia baltica*. The two strongly supported clades of plastids from green lineages (Streptophyta and Chlorophyta) were recovered as a sister to plastids in the red lineage (Rhodophyta, Cryptophyta, Haptophyta, Ochrophyta) and Glaucophyta (UFBoot 100%; Fig. 2, Supplementary Figs. 4–10). Under the site-homogeneous model, Glaucophyta plastids were strongly supported as a sister group to the red-algal plastids (UFBoot >95%; Fig. 2; Supplementary

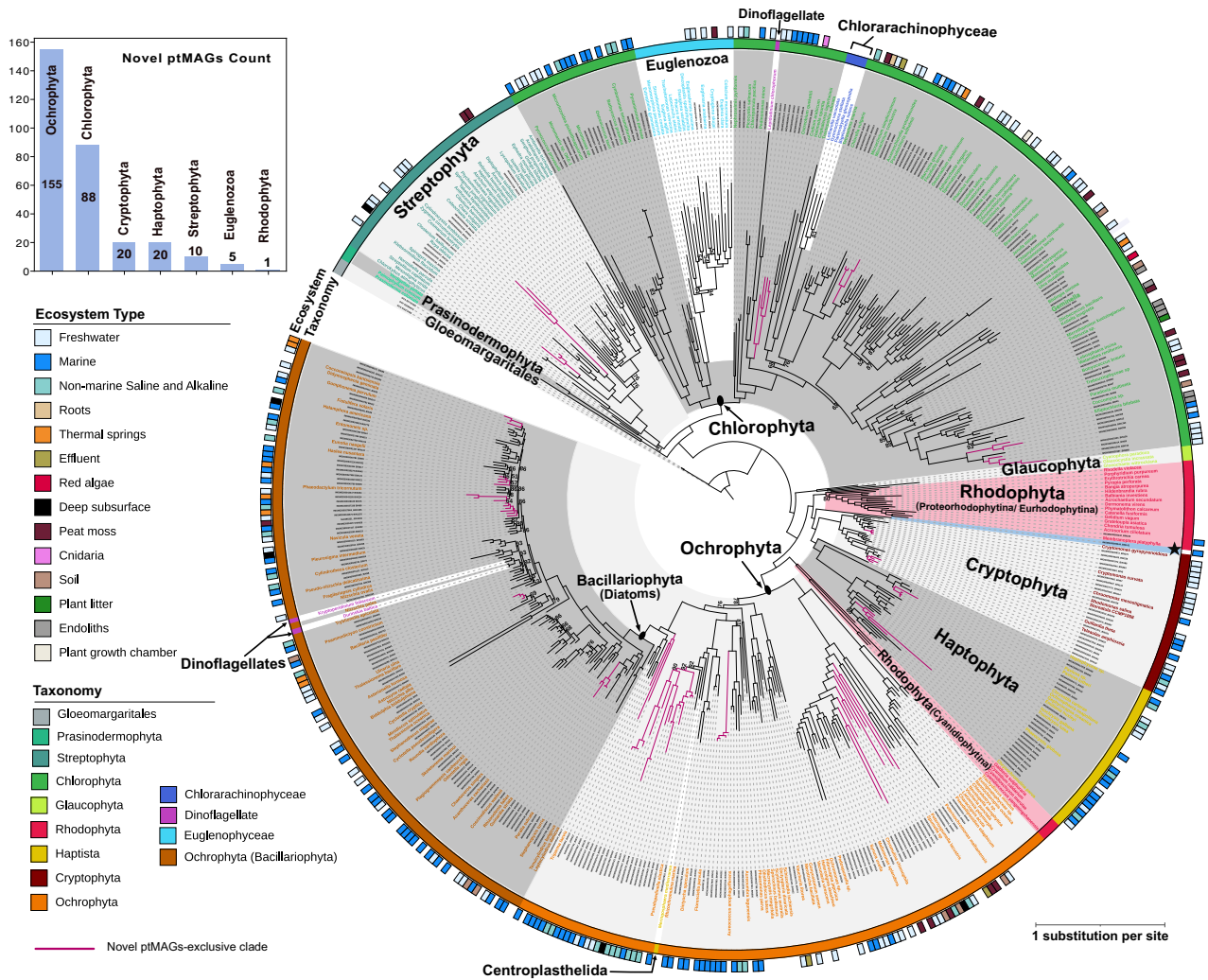


Fig. 2 | Plastid relationships among major photosynthetic eukaryotes. The maximum-likelihood tree was constructed using a concatenated alignment of PLASTID54 marker under the LG + F + I + G4 substitution model, following the nsgtree pipeline and IQ-TREE analysis. A similar topology was recovered under site-heterogeneous models LG + C60 + F + G/R6 (see Supplementary Figs. 6–10). In the tree, the NCBI reference plastome names are colored according to the “Taxonomy” color key, ptMAGs are labeled in black font, and three Gloeomargaritales species were used as the outgroup. Two long-branch taxa (NC_056103, *Pteridomonas danica*, and a ptMAG) were excluded here but are shown in Supplementary Figs. 6–10. The bar plot insert illustrates the number of novel ptMAGs identified per major photosynthetic group. The inner ring indicates the assigned taxonomy for

each clade, as described in the color key. Different background shades of gray are used to distinguish separate clades; clades representing secondary/complex green-algal plastids in Euglenophyceae, Chlorarachinophyceae, and the dinoflagellate *Lepidodinium chlorophorum*, and complex red-algal plastids in dinoflagellates and Centroplasthelida are excluded from the shaded areas. The outer ring indicates the ecosystem of origin for the ptMAGs. A single ptMAG, which diverges before the diversification of Haptophyta and Cryptophyta, is highlighted with blue background shading and marked with a black star. Ultrafast bootstrap values are displayed only for branches with support below 95%. Some examples of clades represented exclusively by two or more ptMAGs are highlighted with magenta branches.

Figs. 4 and 5). In contrast, site-heterogeneous models yielded two alternative topologies: some placed Glaucochyta as the earliest-branching lineage, sister to all other plastids (Supplementary Figs. 6 and 7), while others recovered it as sister to red-algal plastids (Supplementary Figs. 8–10). This model-dependent pattern indicates that the position of Glaucochyta remains unresolved, consistent with the ambiguity present in previous phylogenomic studies⁴⁰. Noteworthy, glaucophytes share more plastid-encoded genes with rhodophytes than with plastids from green lineage^{40,41}.

Our comprehensive phylogenomic analysis also expands the known plastid diversity across most photosynthetic lineages and identifies groups that may correspond to previously described taxa

lacking plastid sequences. We examined the phylogenetic distribution of 300 novel ptMAGs and compared them to selected RefSeq plastomes from various photosynthetic groups. Since these novel ptMAGs were obtained using a dereplicated ptMAGs dataset (Supplementary Data 3), the true diversity of the novel ptMAGs is likely under-represented. Among the novel ptMAGs, the largest number belonged to Ochrophyta ($n = 155$), followed by Chlorophyta ($n = 88$), Haptophyta ($n = 20$), Cryptophyta ($n = 20$), Streptophyta ($n = 10$), Euglenophyceae ($n = 5$) and Rhodophyta ($n = 1$) (Supplementary Data 4 and 5). Interestingly, we identified a novel ptMAG as a sister to Cryptophyta and Haptophyta (discussed in more detail below). We did not recover any ptMAGs representing Glaucochyta and Prasinodermaphyta, and both

phyla were solely represented by RefSeq plastomes. We recovered a single new ptMAG from red algae, suggesting that the available metagenomics data is limited with respect to the plastids of rare or low-abundance algae. Additional deeper sequencing of diverse environmental samples and the identification of new ptMAGs could help to resolve ancient relationships, such as the endosymbiotic events involving red algal endosymbiont(s).

The ptMAGs identified in our study highlight unreported plastid diversity, particularly from clades lacking reference plastomes. Within the green lineages, assembly sizes for Streptophyta ptMAGs ($n = 10$) ranged from -61 to -131 kb, with completeness ranging from 39% to 96% (90% median; Supplementary Data 4 and 5). Among these, seven ptMAGs from freshwater and deep subsurface habitats were associated with Zygnematophyceae algae, indicating they represent new plastid sequences belonging to this class (Fig. 2). In Chlorophyta, we recovered a greater number of ptMAGs, with assembly sizes ranging from -20 to -209 kb and completeness from 33% to 97% (84% median). Particularly, five new ptMAGs (90% median completeness) from marine ecosystems formed a clade sister to Pedinophyceae, while four new ptMAGs (92% median completeness) from marine and freshwater habitats grouped in a distinct clade within the family Chlorellaceae (class Trebouxiophyceae). As observed previously^{42,43}, the plastomes from Trebouxiophyceae were not recovered as monophyletic (Fig. 2 and Supplementary Fig. 4). A single novel rhodophyte ptMAG (-171 kb, 98% completeness) was recovered as a sister taxon to the *Membranoptera platyphylla* plastome, suggesting it belongs to the family Delesseriaceae, subphylum Eurhodophytina. In both Cryptophyta and Haptophyta, we identified 20 novel ptMAGs in each group. Cryptophyta ptMAGs ranged from -96 to -150 kb in size with completeness between 70% and 96%, while Haptophyta ptMAGs varied from -10 to -138 kb with completeness from 9% to 94%. Within Cryptophyta, nine freshwater ptMAGs (89% median completeness) grouped with RefSeq *Cryptomonas* plastomes, while a clade with seven ptMAGs (95% median completeness) from diverse habitats was placed with Geminigeraceae plastomes. In Haptophyta, ptMAGs from marine, freshwater, and non-marine saline and alkaline habitats formed a clade with *Pavlova* plastomes ($n = 6$; 77% median completeness), and with Chrysochromulnaceae plastomes ($n = 9$; 85% median completeness). Given the limited number of complete plastomes for both Cryptophyta and Haptophyta, these novel ptMAGs expand the plastid diversity within these clades.

Considering that the large number of novel ptMAGs identified are from Ochrophyta, we performed an extensive phylogenomic analysis to determine their distribution within the phylum and provide a detailed description of our findings in the Supplementary Note 2. In brief, we identified a total of 155 novel ptMAGs within Ochrophyta, with sizes ranging from -19 to -141 kb and completeness from 14% to 100% (92% median; Supplementary Data 4, 5; Supplementary Figs. 11 and 12). The majority ($n = 81$; 96% median completeness) were classified as Bacillariophyta and originated from diverse habitats including freshwater, marine, deep subsurface, non-marine saline and alkaline, thermal springs, and soil, showcasing significant plastid diversity associated to diatoms. We also identified four novel Bolidophyceae ptMAGs (>100 kb, 85% median completeness), a class sister to Bacillariophyta, which is represented by only a single reference plastome in the public database⁴⁴. Notably, we found a monophyletic clade consisting of three ptMAGs from marine habitats and non-marine saline and alkaline habitats, sister to Bacillariophyta and Bolidophyceae. Despite their smaller size (<50 kb, 48% median completeness), these ptMAGs could be key to advancing our understanding of plastid evolution in diatoms. Additional ptMAGs were identified in other ochrophyte classes, including Chrysophyceae ($n = 27$), Dictyochophyceae ($n = 24$), and Pelagophyceae ($n = 10$), many from diverse habitats such as freshwater, marine, deep subsurface, soil, peat moss, non-marine saline and alkaline, and effluent, forming clades that lack reference plastomes.

The newly recovered ptMAGs from this study, especially those within Chlorophyta and Ochrophyta, hold considerable promise to clarify plastid relationships and evolution within these lineages. To achieve this, it will be necessary to use lineage-specific markers by carefully selecting informative genes to improve phylogenetic resolution. Additionally, employing appropriate evolutionary models and phylogenetic techniques will be crucial for addressing difficult lineages.

Independent origins of secondary red-algal plastids

Interestingly, in all our phylogenetic reconstructions, the Rhodophyta is split (Fig. 2 and Supplementary Figs. 4–10). The mesophilic red algae were monophyletic and strongly supported as a sister to Cryptophyta and Haptophyta. Whereas the early diverging red-algal subphylum Cyanidiophytina was strongly supported as a sister to Ochrophyta (Fig. 2). This finding is consistent with the initial observation by Kim et al.⁴⁵, and a more recent study by Pietluch et al.⁴⁶, which presented a nearly identical topology, suggesting two independent origins of secondary red-algal plastids and proposing a “multiple and serial endosymbiosis” model. This model extends the serial endosymbiosis hypothesis, which posits an initial secondary endosymbiosis of red-algal plastids by cryptophytes, followed by acquisition of cryptophyte plastids by ochrophytes and haptophytes following tertiary or higher-order endosymbiosis^{23–25}. Our study, along with the findings of Pietluch et al.⁴⁶, challenges the notion of a single origin for secondary red-algal plastids and the vertical inheritance of the plastids in all lineages with red-alga-derived secondary/complex plastids, presenting evidence for two distinct engulfment events involving a red alga.

Two separate events are potentially at odds with phenotypic features shared by algae in the CASH lineages, specifically the symbiont-specific ERAD-like machinery (SELMA) complex and occurrence of chlorophyll *c*^{20,47}, which are commonly cited as evidence for a single plastid origin. However, the evolutionary history of these features may be more nuanced and complex. Since the SELMA complex, which facilitates protein import across the periplastidal membrane, is derived from preexisting symbiont-derived ER-associated protein degradation (ERAD) machinery^{48,49}, the frequency of independent derivation and reestablishment of this system remains uncertain. Under our model, two independent adaptations (in cryptophytes and ochrophytes) and two establishments (in haptophytes and alveolates) would be required, whereas under a single origin model, a single adaptation followed by multiple reestablishments are required, making both scenarios comparably parsimonious as suggested by Pietluch et al.⁴⁶. Phylogenetic analyses reveal that SELMA components show mosaic evolutionary origins, with only some SELMA proteins tracing back to red algae and displaying monophyly across CASH lineages⁵⁰. Reconciling these patterns with the single origin and serial endosymbiosis model requires invoking lineage-specific duplications involving ancestral non-endosymbiont genes and replacement of SELMA components rather than simple vertical inheritance⁵⁰. Similarly, while chlorophyll *c* synthase (CHLC) appears monophyletic across CASH lineage, phylogenetic analyses reveal several inconsistencies including polyphyly of ochrophytes, cryptophytes nested within ochrophytes rather than positioned as the basal lineage, and clustering of some ochrophytes and haptophytes members as a clade, potentially due to poor phylogenetic resolution or gene transfer between lineages^{51,52}. Notably, some ochrophytes entirely lack CHLC homologs despite producing chlorophyll *c*, demonstrating that alternative chlorophyll *c* biosynthetic pathways have evolved independently at least once. While this absence has been attributed to gene loss and pathway remodeling⁵², the independent evolution of chlorophyll *c* synthesis indicates multiple independent innovations and complicates interpretation of CHLC monophyly as definitive evidence for a single endosymbiotic origin.

Overall, these observations suggest that while the SELMA complex and chlorophyll *c* have been considered essential features supporting a single origin of red-algal secondary plastids, the molecular phylogenies of the genes underlying these features present a more nuanced picture. These complexities do not refute a single origin hypothesis, as gene losses, replacements, and lateral transfers could account for the observed patterns. However, they equally do not preclude scenarios involving multiple independent endosymbiotic events, particularly considering the distinct red-algal plastid ancestors that we observe. Distinguishing between these competing hypotheses will require expanding sampling across underrepresented algal lineages, including early diverging groups, and characterization of the genes underlying shared phenotypic traits.

New lineage sister to Cryptophyta and Haptophyta

We also discovered a novel ptMAG from a marine ecosystem (Arctic Ocean near Svalbard) that exhibits an intriguing phylogenetic relationship with Cryptophyta and Haptophyta. This ptMAG (IMG3300009544_BIN141) is phylogenetically affiliated with the clade comprising the plastids of Cryptophyta and Haptophyta, and together these three lineages form a monophyletic clade that is sister to mesophilic red algae. This ptMAG could potentially provide new insight into the acquisition of plastids from red algae in Cryptophyta before the spread to Haptophyta. The ptMAG consists of a single 104,344 bp contig likely representing the large single copy region of the plastome with partial 23S rRNA and 16S rRNA gene sequences at either end and retains 108 protein-coding genes with 26 tRNAs genes (Supplementary Data 6). The identified genes include those encoding large and small ribosomal subunits (38 genes), photosystem I/II reaction center (25 genes), ATP synthase (8 genes), cytochrome *b₆f* complex (6 genes), plastid-encoded RNA polymerase (4 genes), carbon fixation (3 genes), cytochrome *c* biogenesis (2 genes), and protein translocase (2 genes). The ptMAG lacks several plastid genes found in haptophytes and/or cryptophytes, such as *ftsH*, *minE*, *dnabX*, and *chlN/LB*⁵³, but we cannot rule out that the absences are due to the partial nature of the assembly. However, the ptMAG contains two genes, *minD* (encoding septum site-determining protein) and eubacterial *c*-type *rpl36* (encoding the large ribosomal subunit protein L36, with the best BLAST hit to RPL36 protein from bacterial order Pirellulales), which are absent in the plastomes of mesophilic red algae but shared by plastomes of cryptophytes and haptophytes. The presence of eubacterial *c*-type *rpl36* in cryptophytes and haptophytes plastomes is attributed to horizontal gene transfer in an ancestor of cryptophytes before engulfment of a cryptophyte by the haptophyte ancestor⁵⁴. This ptMAG could come from an algal lineage that diverged prior to acquisition of the plastid by haptophytes, or this ptMAG is from a red alga that unlike other red algae encodes a Pirellulales-like *rpl36*.

Since our study includes only a single sequence of this novel ptMAG, we searched public databases for other sequences sharing high similarity. Initially, we used the partial 16S rRNA gene sequence from the ptMAG to search for similar sequences in the NCBI database. The nucleotide BLAST search revealed that the closest match is a 16S rRNA gene sequence (KX935025; ~92% nucleotide pairwise identity) from an uncultured marine eukaryote in the DPL2 clade identified by Choi et al.⁵⁵. DPL2 (Deep-branching Plastid Lineage 2) is a globally distributed deep-branching clade positioned as a sister to haptophytes. We also used the protein sequences for plastid genes from the ptMAG IMG3300009544_BIN141 to search the MGnify microbiome database⁵⁶ (<https://www.ebi.ac.uk/metagenomics>) with phmmer. We identified several assemblies with significant *e*-values (1×10^{-208} or lower), mainly corresponding to Tara Oceans assemblies and a few from a marine microbial diversity study from Saanich Inlet. Several contigs were identified with sizes ranging from -1 to -104 kb and sharing >99% nucleotide identity with the novel ptMAG. However,

none were longer than our assembly. The longest contigs from those Tara Oceans assemblies (TARA_B110000971 and TARA_B110000977, from different Arctic Ocean locations) shared identical gene content with our novel ptMAG IMG3300009544_BIN141, including 100% amino acid identity with eubacterial *c*-type RPL36. The 16S rRNA gene sequence from all these sequences exhibit the highest sequence similarity (>90%) with DLPL2 16S rRNA sequence. The highly similar genomic features between the Tara Oceans contigs and our novel ptMAGs suggest that they are likely all from closely related, undescribed marine algae.

Our results were further corroborated by findings described in a recent preprint by Jamy et al.⁵⁷, which identified a new algal group “Leptophytes” based on plastid sequences, placing it deep within the algal tree of life. However, Jamy et al. focused exclusively on the identification of plastid sequences from the oceanic pelagic zone. Our novel ptMAG, IMG3300009544_BIN141, shares 99.9% identity with the best representative plastome for leptophytes (Lepto-01)⁵⁷. The phylogenetic placement of the novel ptMAG as a sister lineage to both cryptophytes and haptophytes was supported by Jamy et al. However, an alternative topology was also presented, suggesting leptophyte plastids branch closer to haptophyte plastids than cryptophytes, leading to uncertainty regarding their exact phylogenetic placement. In addition, the authors were unable to identify candidate nuclear genome sequences belonging to these algae. Nonetheless, the novel ptMAGs recovered in both studies may help to bridge the evolutionary gap between red algae and Cryptophyta/Haptophyta.

Novel ptMAGs linked to secondary/complex green-algal plastids

Complex plastids derived from green algae have been previously identified in three distinct lineages: Chlorarachniophyceae, Euglenophyceae, and the dinoflagellate species *Lepidodinium chlorophorum*. Notably, in *L. chlorophorum*, the original plastid is believed to have been replaced by a chlorophyte-derived plastid^{19,58–60}. Our phylogenetic analyses support the independent origins of green-alga-derived secondary plastids in these three lineages (Fig. 2, and Supplementary Figs. 4, 5). No novel ptMAGs were found within Chlorarachniophyceae; this clade was solely represented by the four RefSeq plastomes of *Lotharella vacuolata*, *Gymnochora stellata*, *Bigelowiella natans*, and *Partenskyella glossopodia*. Chlorarachniophyceae formed a strongly supported monophyletic group, sister to the order Trentepohliales (moderate to high support values; UFBoot 77% to 98%). Chlorarachniophyceae together with Trentepohliales was strongly supported as sister to order Bryopsidales, in agreement with a previous study that concluded siphonous green algae are likely ancestors of Chlorarachniophyceae plastids¹⁹.

Our study consistently recovered secondary plastids in Euglenophyceae as a sister clade to the prasinophyte algae order Pyramimonadales with strong support. In some cases, euglenophyte plastids were placed closer to *Cymbomonas tetramitiformis* than to other Pyramimonadales, albeit with moderate support (UFBoot 85%; Fig. 2 and Supplementary Figs. 4 and 5). Although the specific Pyramimonadales alga closest to euglenophytes has not been determined, the sister relationship between Pyramimonadales and euglenophytes has been consistently recovered in previous studies. Our findings further support that secondary plastids in euglenophytes share a common ancestor with Pyramimonadales algae^{40,59,61}. We identified several novel ptMAGs both within Pyramimonadales and Euglenophyceae. Specifically, novel ptMAGs (*n* = 5) with assembly sizes ranging from -44 to -104 kb and completeness from 52% to 85% (75% median) were discovered within Euglenophyceae, and ptMAGs (*n* = 4) sizes ranging from -70 to -90 kb and completeness from 60% to 93% were placed within Pyramimonadales.

Among the lineages with chlorophyte-derived secondary plastids, we uncovered several novel ptMAGs closely related to *L. chlorophorum* (Fig. 2). To gain insights into the origin of plastids in this dinoflagellate,

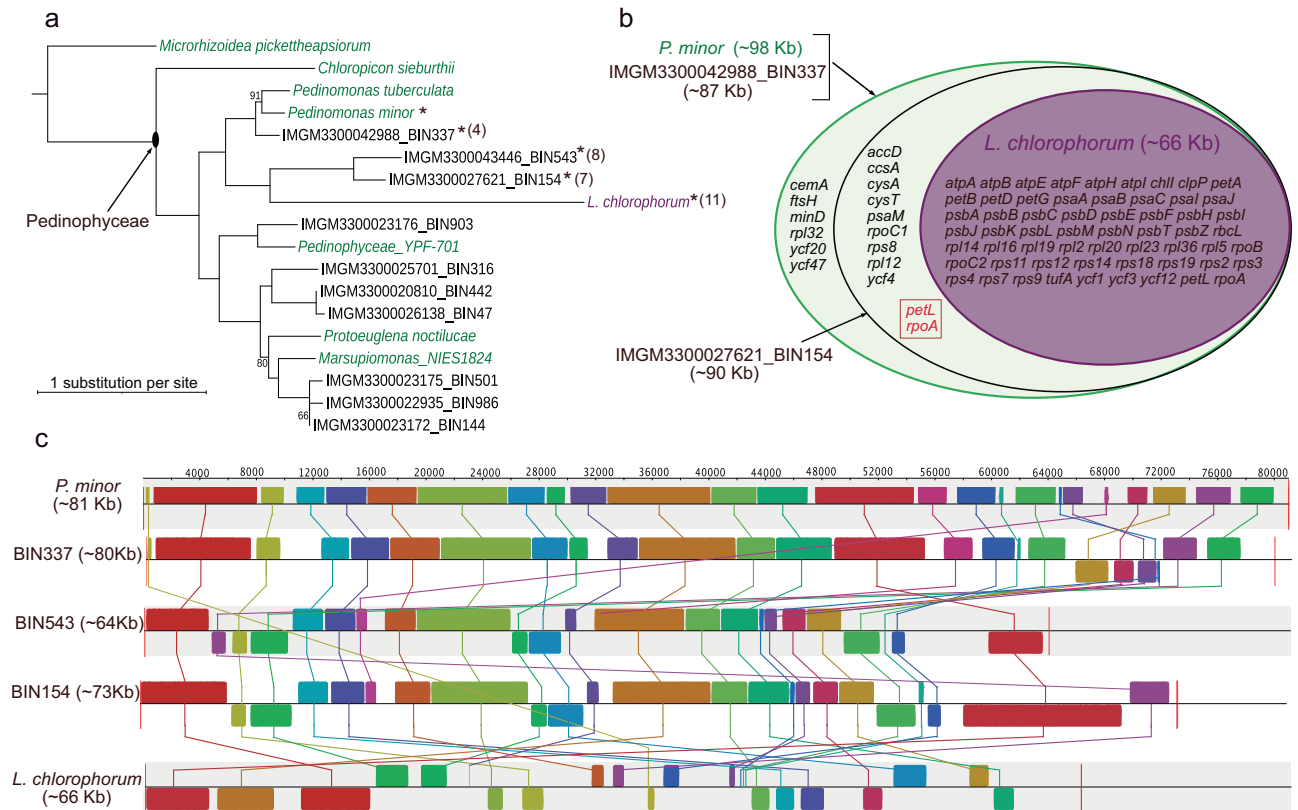


Fig. 3 | Closest relatives of green-alga-derived plastids in the dinoflagellate *Lepidodinium chlorophorum*. **a** Phylogenetic position of the *L. chlorophorum* plastid compared to *Pedinomonas* spp. plastids and novel ptMAGs. Additional Pedinophyceae plastomes available in GenBank are included. *Microrhizoida* and *Chloropicon* spp. were used as an outgroup. Ultrafast bootstrap values < 95% are shown. The reference species are color-coded, green for chlorophytes and purple for *L. chlorophorum*. Asterisks denote plastomes/ptMAGs used for whole genome alignment analysis. The number in the parentheses indicates the number of inverted syntenic blocks compared against *P. minor* (Fig. 3c). **b** A Venn diagram showing gene contents among *P. minor*, ptMAGs (IMG3300042988_BIN337 and

IMG3300027621_BIN154), and *L. chlorophorum*, along with their respective plastome sizes. Two genes, *petL* and *rpoA*, missing only in IMG3300027621_BIN154, are highlighted with a red dotted box. **c** Whole plastome alignment of ptMAGs and *L. chlorophorum* relative to *P. minor* generated using progressiveMauve alignment. Each locally collinear block representing synteny is color-coded; the blocks below the horizontal center indicate inversions. For simplicity, the ptMAG labels have been abbreviated to associated bins. A copy of inverted repeats and single copy regions were removed from the plastome of *Pedinomonas* for the alignment.

we conducted comparative analyses (Fig. 3). The plastid origin in the *L. chlorophorum* has been attributed to an endosymbiotic event involving pedinophyte algae closely related to *Pedinomonas minor*^{49,62}. We identified two novel ptMAGs more closely related to *L. chlorophorum* than *Pedinomonas* species. To determine the exact position of these novel ptMAGs, we included additional Pedinophyceae plastomes available in the GenBank and reconstructed a phylogeny (Fig. 3a). Two novel ptMAGs, IMG3300043446_BIN543, and IMG3300027621_BIN154, with assembly sizes of ~64 kb and ~90 kb, respectively, were strongly supported as sister to *L. chlorophorum*. The ptMAG IMG3300027621_BIN154 was represented by two contigs (~81 kb and ~9 kb), with the larger contig containing ribosomal rRNA operons at both ends, and the shorter contig containing genes, *cysT*, *ccsA*, and *psaC*, which are typically located within the small single copy of *Pedinomonas* species. The ptMAG IMG3300043446_BIN543 consists of a single contig of ~64 kb and shares 65% ANI with the ptMAG IMG3300027621_BIN154. The larger ptMAG IMG3300027621_BIN154 likely represents a near-complete assembly and was selected for the gene content analysis. It retained almost all protein-coding genes present in *L. chlorophorum* except for *petL* and *rpoA*, while also including nine additional genes (Fig. 3b). *L. chlorophorum*, along with the ptMAGs IMG3300043446_BIN543 and IMG3300027621_BIN154, was strongly supported as a sister to a clade containing two *Pedinomonas* plastomes and a novel ptMAG (IMG3300042988_BIN337) of size 87,241 bp and 86% completeness.

The gene content of ptMAG IMG3300042988_BIN337 was identical to *P. minor*, but four syntenic blocks were inverted at the location 64–72 kb (Fig. 3b, c). Considering the overall gene content, synteny, and the phylogenetic position, the ptMAG IMG3300042988_BIN337 likely represents the plastome of a new *Pedinomonas* species. The *Pedinomonas* clade (*Pedinomonas* spp. and IMG3300042988_BIN337) retained all the protein-coding genes present in *L. chlorophorum* and ptMAG IMG3300027621_BIN154, along with six additional genes (Fig. 3b). Gene order in *L. chlorophorum* and its sister ptMAGs (IMG3300043446_BIN543 and IMG3300027621_BIN154) showed substantial rearrangements compared to *P. minor* (Fig. 3c). The ptMAGs IMG3300027621_BIN154 and IMG3300043446_BIN543 had similar gene order with inversions of syntenic blocks at three locations (4–12 kb, 24–32 kb, and 48–68 kb). In contrast, the *L. chlorophorum* plastome was highly reduced and rearranged, and gene order is not conserved. Additionally, the branch leading to *L. chlorophorum* was substantially longer compared to other plastomes and ptMAGs within Pedinophyceae. Considering shared gene content, synteny, and branch lengths, the ptMAGs IMG3300027621_BIN154 and IMG3300043446_BIN543 could potentially represent a novel pedinophyte algae sister to *Pedinomonas*. If so, these two ptMAGs are more closely related to the *L. chlorophorum* plastid than the plastid from *P. minor*. It is also possible that these ptMAGs belong to novel dinoflagellate species that are sister to *L. chlorophorum* and have retained ancestral plastome features subsequently lost in *L. chlorophorum*.

The endosymbiotic origin of primary plastids derived from cyanobacteria and their subsequent distribution among eukaryotes have led to remarkable diversity in photosynthetic eukaryotes. In this study, we explored the evolutionary relationships of primary plastids with cyanobacterial lineages and secondary/complex plastids among major algal groups by analyzing plastid sequences derived from metagenomic assemblies. By expanding the cyanobacterial taxonomic framework to include both photosynthetic and non-photosynthetic sister lineages, we pinpointed the origin of plastids. Our findings confirm that Archaeplastida plastids are closely related to the cyanobacterial order Gloeomargaritales, supporting the hypothesis that primary plastids originated from a deeply branching cyanobacterium, while *Paulinella* plastids arose independently. However, our results challenge existing models of plastid evolution in eukaryotes with secondary/complex plastids, presenting evidence for two independent origins of secondary red-algal plastids. Recent studies align with this finding, emphasizing the need for further research and reassessment of secondary endosymbiosis involving extinct red algae. Additionally, we uncovered numerous novel ptMAGs, particularly within Chlorophyta and Ochrophyta, revealing previously unexplored plastid diversity. Notably, we discovered a novel ptMAG that may represent an evolutionary link between red algae and Cryptophyta/Haptophyta, potentially tracing the common ancestor of secondary red-algal plastids. The novel ptMAGs offers valuable insights into plastid evolution across photosynthetic eukaryotes.

Methods

Phylogenetic markers

We utilized two sets of phylogenetic markers, referred to as UNI56 and PLASTID54, for all our phylogenetic analyses (Supplementary Data 2). The UNI56 marker included 56 universal single-copy marker proteins, comprising 30 large and small ribosomal subunits, three DNA-directed RNA polymerase subunits, 10 tRNA synthetase, and other functional proteins^{33,63}. The PLASTID54 consisted of 34 Hidden Markov Model (HMM) profiles derived from the UNI56 marker, including large and small ribosomal subunits, DNA-directed RNA polymerase subunits, and the preprotein translocase SecY. Additionally, they included 20 HMM profiles for plastid photosynthetic proteins associated with photosystems, ATP synthase, and the cytochrome complex as available in Pfam-A HMMs (last accessed August 2024) and as detailed in Supplementary Data 2. The phylogenomic relationships among Cyanobacteriota species were inferred using the UNI56 marker, while the plastid phylogeny among photosynthetic eukaryotes was estimated using the PLASTID54 marker. The initial placement of plastids within the cyanobacterial lineage was examined using the UNI56 marker and then re-evaluated with the PLASTID54 marker.

Cyanobacteriota genomes and MAGs

We performed metagenomic binning on 31,152 metagenomes from the IMG/M^{31,32}, focusing on non-redundant samples publicly available up to April 2023^{64–72}. Only contigs with lengths of ≥ 5 kb were processed. The binning was conducted on each sample using MetaBAT 2⁷³ with parameters: `--minContig 5000`, `--minClSize 20,000`, and `--cvExt`, and the gene calling on each bin was performed using Prodigal (v2.6.3)⁷⁴.

Reference genomes and bins from publicly accessible repositories were downloaded from the IMG isolates repository ($n = 115,969$) and the bacterial and archaeal representative datasets of GTDB ($n = 85,205$)⁷⁵. All genomes and bins underwent CheckM2 v1.0.2⁷⁶ analysis using UNI56 marker for universal marker detection required for phylogenomic reconstruction, CheckM2 for quality assessment, and SeqKit v2.5.1⁷⁷ for assembly characterization. Genomes and bins with less than 50% UNI56 marker proteins, CheckM2 completeness below 50%, or CheckM2 contamination above 5% were excluded from the genome catalog.

Species-level clusters were inferred with skani (v0.1.4⁷⁸, parameters: `ANI \geq 95%`, `align_fraction_ref \geq 50%`, and `align_fraction_query \geq 50%`) and cluster representatives were selected based on the highest quality metrics (completeness and contamination), with ties resolved by selecting assemblies with the highest number of predicted genes and the highest N50 value. Finally, genomes classified as Cyanobacteriota according to the GTDB taxonomic classification were selected for the downstream analyses.

Cyanobacteriota phylogeny and diversity

The phylogenetic relationship of Cyanobacteriota was inferred with an estimated species tree using the nsgtree pipeline v0.5.1 (<https://github.com/NeLLi-team/nsgtree>) and IQ-TREE v2.0⁷⁹. In brief, the nsgtree pipeline identified orthologs associated with UNI56 marker using hmmsearch v3.1/b2 (www.hmmer.org). For each genome and marker, the ortholog with the highest bitscore was extracted. Extracted orthologs were then aligned using MAFFT v7⁸⁰ and trimmed using trimAl v1.4⁸¹ with a gap threshold of 10% (`-gt 0.1`). A supermatrix alignment was generated by concatenating 56 individual ortholog alignments. Only genomes retaining at least 50% of the UNI56 marker proteins were included in the final supermatrix alignment. A maximum-likelihood (ML) species tree was estimated with the supermatrix alignment using IQ-TREE with the LG + F + I + G4 substitution model, and branch support was evaluated with 1000 ultrafast bootstrap pseudoreplications⁸². The percentage increase in phylogenetic diversity was calculated by comparing the sum of branch lengths of the Cyanobacteriota species trees with and without new bacterial MAGs generated in our study (labeled as NeLLi2023 in Supplementary Data 1).

To generate representative Cyanobacteriota genomes for examining the origin of plastids, we further dereplicated the Cyanobacteriota dataset. Pairwise evolutionary distances (branch lengths) were inferred between all species-level Cyanobacteriota culture representatives utilizing the ML species tree generated with IQ-TREE using PhyloDM v3.2.0⁸³ and then clustered with MCL v22-282 at different distances cutoff values, ranging from 0.2 to 1.0 and an inflation value at 1.5⁸⁴. For each cutoff value, we calculated basic cluster statistics, such as, the number of clusters, number of singletons, average cluster size, and the number of genomes in the largest cluster, and sorted the cluster based on the UNI56 marker proteins count for the genomes/MAGs generated by the nsgtree pipeline. We inspected the clusters at different cutoffs to ensure the genomes belonging to different GTDB taxonomic orders were not collapsed or mixed within clusters, allowing us to select an appropriate cutoff value.

Screening of metagenomes for plastid sequences

We analyzed 26,257 public metagenomes from the IMG/M database to identify contigs containing plastid 16S rRNA gene sequences using cmsearch (Infernal v1.1)⁸⁵ with the bacterial SSU rRNA covariance model RF00177 from the Rfam database. To accommodate longer sequences, we utilized the `--anytrunc` option in cmsearch and parsed the output to identify contigs with sequential and multiple non-overlapping alignments to the covariance model, which were then assembled into single sequences. The recovered SSU rRNA sequences were annotated using the SILVA⁸⁶ and PR2⁸⁷ databases via blastn v2.13.0⁸⁸, retaining those with a plastid sequence as the best hit and an alignment length of at least 500 bp, resulting in 11,298 candidate plastid SSU rRNA genes. For further analyses, we selected 4497 contigs with lengths of at least 5 kb.

Identification of plastid metagenome assembled genomes (ptMAGs)

To identify high-quality ptMAGs among MAGs that contained a contig with a plastid SSU rRNA gene, we performed gene calling and annotation using Prodigal v2.6.3 with the `“-p meta”` option, which employs

precalculated training files, and keeping other parameters at default settings. We selected ptMAGs that contained at least 10% of UNIS6 marker proteins, in accordance with the nsgtree pipeline. To remove non-plastid sequences likely resulting from mis-assembly and mis-binning, we performed a DIAMOND v2.1.8⁸⁹ BLASTp for all predicted proteins in the ptMAGs against the NCBI non-redundant (nr) database (last accessed August 2024), using an e-value cut-off of 1×10^{-6} and minimum of 30% of subject and query coverage. We then assigned taxonomic affiliations for the best hits and categorized the hits based on amino acid percent identity into three groups: low (<70%), medium ($\geq 70\%$ and <90%), and high ($\geq 90\%$). To streamline the taxonomic affiliation of the proteins, we retained taxonomic ranks at the Domain and Phylum levels, along with the assigned percent identity category. We then assigned taxonomic affiliations to the contigs based on the simplified taxonomic ranks for the majority of the proteins in each contig. Contigs assigned to bacteria with the medium and high percent identities that were not associated with the Cyanobacteriota were discarded. Only contigs assigned to Eukaryota, Cyanobacteriota, and Bacteria from any other phylum but with low percent identities were selected for the further analysis.

Similarly, to remove potentially mis-identified mitochondrial metagenomic contigs, we compared the ptMAGs against 18,222 complete RefSeq mitochondrial genomes available in the NCBI (<https://ftp.ncbi.nlm.nih.gov/refseq/release/mitochondrion/>; last access October, 2024). For each plastid contig, we generated BLAST hits (blastn, e-value of 1×10^{-6}) and estimated pairwise identity using skani³⁴ with the “skani dist --slow” option, against the mitochondrial RefSeq database. Based on the percent identity and alignment length (95% nucleotide identity and >95% alignment length), we identified and filtered out any contigs that are likely mitochondrial in origin.

Selecting reference plastomes for phylogenetic analyses

From the total of 15,248 RefSeq plastomes available at NCBI (<https://ftp.ncbi.nlm.nih.gov/refseq/release/plastid/>; last access October 2024), we initially subsampled reference plastomes at the order level to investigate the phylogenetic placement of the ptMAGs among photosynthetic eukaryotes. As the NCBI RefSeq plastomes database is heavily overrepresented by species from the phylum Streptophyta, specifically the class Magnoliopsida (~14,000), we manually selected a few species to represent Magnoliopsida, rather than choosing a species to represent each taxonomic order during the analyses. We then iterated phylogenetic reconstruction per phylum by selecting all ptMAGs belonging to a particular phylum, along with all RefSeq plastomes associated with the corresponding phylum, to select the RefSeq plastomes that represent the closest sisters to the ptMAGs. During this process, we identified that the NCBI plastome for *Interfilum terricola* (NC_025542) is misclassified, with phylogenetic evidence supporting its reassignment to *Geminella terricola* (Chlorophyta). Accordingly, this plastome was renamed to *Geminella terricola* in our figures, and further details on this case, as well as other potential plastome misclassifications, are provided in the Supplementary Note 3.

Dereplicating and identifying novel ptMAGs

To eliminate redundant ptMAGs, we employed a similar approach as for the Cyanobacteriota, but with some modifications. First, we generated an initial ML tree for the ptMAGs, setting a minimum at 10% of PLASTID54 marker proteins for inclusion in the final supermatrix alignment. Using the ML tree, we calculated pairwise evolutionary distances between cluster representatives with PhyloDM and at a 0.99 cutoff value with the inflation value of 1.5 using MCL clustering. Second, we generated cluster statistics and sorted the cluster based on the count of PLASTID54 marker proteins. To represent each cluster, we selected plastid ptMAGs based on the criteria of highest PLASTID54 marker proteins count, fewest contigs per metagenomic bin, overall size of the contigs, and highest number of proteins.

To identify novel ptMAGs, we downloaded 15,248 complete plastomes available at NCBI RefSeq along with all plastid sequences ≥ 3000 bp available in the NCBI GenBank (last access October 2024), creating a dataset of 28,983 publicly available plastid sequences, including complete and partial plastomes. We calculated average nucleotide identity (ANI) for the ptMAGs against the plastid dataset using skani with the “skani dist --slow” option. Plastid MAGs were classified as redundant based on two criteria: (i) ANI $\geq 95\%$, align_fraction_ref $\geq 50\%$, and align_fraction_query $\geq 50\%$ (ptMAGs), and (ii) ANI $\geq 95\%$ and align_fraction_query $\geq 95\%$ with any values for the align_fraction_ref. The latter criterion specifically aimed to filter out smaller ptMAGs with very high ANI that only cover partial regions of a complete reference plastome.

Estimating plastid phylogeny

We inferred plastid phylogeny by calculating a species tree using the nsgtree pipeline (<https://github.com/NeLLi-team/nsgtree>) and IQ-TREE, with adjustments to the dataset and markers. To determine the origin of plastids, we combined a dereplicated Cyanobacteriota dataset with ptMAGs and RefSeq plastomes and inferred phylogenetic relationships using PLASTID54. Our initial ML tree estimation included a dataset containing 259 dereplicated Cyanobacteriota taxa, 241 RefSeq NCBI plastomes, and 647 metagenomic contigs containing plastid 16S rRNA genes. Although we initially had 1027 ptMAGs (Supplementary Data 3) containing 16S rRNA genes, applying the criterion of at least 10% of PLASTID54 marker proteins at the contig-level reduced the number to 647 contigs. Structural annotation was performed on the contigs encoding 16S rRNA using Prodigal to generate protein sequences, which were subsequently used for the ML estimation. We then iterated a series of phylogenetic reconstructions by generating several supermatrix alignments, replacing the proteins from ptMAGs ($n = 1,027$ bins) instead of limiting to contigs with 16S rRNA genes, and adjusting the parameter of minimum numbers of PLASTID54 marker proteins in the final alignment (genomes retaining at least 10%, 20%, 30%, and 40%). The ML species trees were then estimated for each supermatrix alignment using IQ-TREE with the LG + F + I + G4 substitution model, and branch support was evaluated with 1,000 ultrafast bootstrap pseudoreplications. The final dataset used to examine the distribution of novel ptMAGs comprised 537 taxa in total, including 238 reference plastomes, 296 ptMAGs, and three Gloeomargaritales species used as the outgroup to the tree (Supplementary Data 4 and 5).

To further investigate the phylogenetic relationship between photosynthetic eukaryotes, we conducted additional analyses using complex site-heterogenous models implemented in IQ-TREE. We reduced our original 537-taxon dataset to a smaller but representative 157-taxon dataset for best-fit model selection using IQ-TREE. We evaluated site-homogeneous models (LG and WAG) in combination with empirical amino acid frequencies (+F), gamma-distributed rate heterogeneity (+G4), as well as site-heterogenous mixture models (C20, C40 and C60) and their Posterior Mean Site Frequency (PMSF) approximations. Based on the Bayesian Information Criterion (BIC), we selected the best-fitting model as WAG + C60 + PMSF + G4. We inferred maximum likelihood phylogenies using this model for both the 157-taxon dataset and the original 537-taxon dataset, with branch support assessed using 1000 ultrafast bootstrap pseudoreplicates. Additionally, we inferred maximum likelihood trees for the 535-taxon dataset using several other site-heterogenous models, including LG + C60 + PMSF + G4, LG + PMSF + F + R6, LG + C60 + F + G, and LG + C60 + F + R6, each with 1000 ultrafast bootstrap pseudoreplicates.

Annotation and completeness assessment of ptMAGs

Protein-coding genes, ribosomal RNAs, and transfer RNAs in ptMAGs were annotated using GeSeq v2.03⁹⁰, and tRNAscan-SE v2.07⁹¹, followed by manual curation using Geneious Prime v2025.0.2 (www.geneious.com).

geneious.com). Whole-genome alignments were carried out using progressiveMauve v1.1.3⁹² within the Geneious platform. To assess plastid completeness, we conducted OrthoFinder v2.5.5⁹³ analysis using a dataset comprising dereplicated ptMAGs and NCBI reference plastomes that were also used for phylogenetic inference among photosynthetic eukaryotes. OrthoFinder results were parsed to identify core proteins for each phylum (e.g., Streptophyta, Chlorophyta, Rhodophyta, etc.), defined as orthologs present in 90% of the NCBI reference plastomes. For each ptMAG, we calculated the proportion of these core proteins detected for the corresponding phylum, thereby providing a lineage-specific measure of plastid completeness (Supplementary Data 7–13).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The plastid MAGs generated in our study are publicly available through the European Nucleotide Archive under project accession PRJEB106049 and individual accessions are listed in Supplementary Data 3. Additional data generated in our study are available as supplementary dataset via FigShare [<https://doi.org/10.6084/m9.figshare.31146589>]⁹⁴. The supplementary dataset includes concatenated protein alignments used for phylogenetic inferences, phylogenetic trees, annotations of plastid MAGs in GenBank format, and plastid MAGs in FASTA format.

Code availability

The workflow (NSGTree v0.5.1) used to infer phylogenies in this manuscript is available on GitHub [<https://github.com/NeLLi-team/nsgtree>].

References

- Coale, T. H. et al. Nitrogen-fixing organelle in a marine alga. *Science* **384**, 217–222 (2024).
- Moulin, S. L. Y. et al. The endosymbiont of *Epithemia clementina* is specialized for nitrogen fixation within a photosynthetic eukaryote. *ISME Commun.* **4**, ycae055 (2024).
- Whatley, J. M., John, P. & Whatley, F. R. From extracellular to intracellular: the establishment of mitochondria and chloroplasts. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **204**, 165–187 (1979).
- Whatley, J. M. & Whatley, F. R. Chloroplast evolution. *N. Phytol.* **87**, 233–247 (1981).
- Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G. & Bhattacharya, D. A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* **21**, 809–818 (2004).
- Adl, S. M. et al. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J. Eukaryot. Microbiol.* **52**, 399–451 (2005).
- Parfrey, L. W., Lahr, D. J. G., Knoll, A. H. & Katz, L. A. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc. Natl. Acad. Sci. USA* **108**, 13624–13629 (2011).
- Gawryluk, R. M. R. et al. Non-photosynthetic predators are sister to red algae. *Nature* **572**, 240–243 (2019).
- Schön, M. E. et al. Single cell genomics reveals plastid-lacking Picozoa are close relatives of red algae. *Nat. Commun.* **12**, 6651 (2021).
- Ponce-Toledo, R. I. et al. An Early-branching freshwater cyanobacterium at the origin of plastids. *Curr. Biol.* **27**, 386–391 (2017).
- Moore, K. R. et al. An expanded ribosomal phylogeny of cyanobacteria supports a deep placement of plastids. *Front. Microbiol.* **10**, 1612 (2019).
- Rodríguez-Ezpeleta, N. et al. Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr. Biol.* **15**, 1325–1330 (2005).
- Keeling, P. J. The endosymbiotic origin, diversification and fate of plastids. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 729–748 (2010).
- Archibald, J. M. The puzzle of plastid evolution. *Curr. Biol.* **19**, R81–R88 (2009).
- Marin, B., Nowack, E. C. & Melkonian, M. A plastid in the making: evidence for a second primary endosymbiosis. *Protist* **156**, 425–432 (2005).
- Nowack, E. C. M., Melkonian, M. & Glöckner, G. Chromatophore genome sequence of paulinella sheds light on acquisition of photosynthesis by eukaryotes. *Curr. Biol.* **18**, 410–418 (2008).
- Delaye, L., Valadez-Cano, C. & Pérez-Zamorano, B. How really ancient is paulinella chromatophora? *PLoS Curr.* **8**, ecurrent-s.tol.e68a099364bb1a1e129a17b4e06b0c6b (2016).
- Macorano, L. & Nowack, E. C. M. Paulinella chromatophora. *Curr. Biol.* **31**, R1024–R1026 (2021).
- Jackson, C., Knoll, A. H., Chan, C. X. & Verbruggen, H. Plastid phylogenomics with broad taxon sampling further elucidates the distinct evolutionary origins and timing of secondary green plastids. *Sci. Rep.* **8**, 1523 (2018).
- Cavalier-Smith, T. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J. Eukaryot. Microbiol.* **46**, 347–366 (1999).
- Burki, F. et al. Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc. R. Soc. B* **283**, 20152802 (2016).
- Strassert, J. F. H., Irisarri, I., Williams, T. A. & Burki, F. A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. *Nat. Commun.* **12**, 1879 (2021).
- Petersen, J. et al. Chromera velia, endosymbioses and the rhodoplex hypothesis—plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH Lineages). *Genome Biol. Evol.* **6**, 666–684 (2014).
- Stiller, J. W. et al. The evolution of photosynthesis in chromist algae through serial endosymbioses. *Nat. Commun.* **5**, 5764 (2014).
- Bodyl, A., Stiller, J. W. & Mackiewicz, P. Chromalveolate plastids: direct descent or multiple endosymbioses? *Trends Ecol. Evol.* **24**, 119–121 (2009).
- Sanchez-Puerta, M. V. & Delwiche, C. F. A hypothesis for plastid evolution in chromalveolates. *J. Phycol.* **44**, 1097–1107 (2008).
- Sibbald, S. J. & Archibald, J. M. Genomic insights into plastid evolution. *Genome Biol. Evol.* **12**, 978–990 (2020).
- Martijn, J., Vosseberg, J., Guy, L., Offre, P. & Ettema, T. J. G. Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature* **557**, 101–105 (2018).
- Liu, Y. et al. Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature* **593**, 553–557 (2021).
- Delmont, T. O. et al. Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean. *Cell Genom.* **2**, 100123 (2022).
- Chen, I.-M. A. et al. IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res.* **45**, D507–D516 (2017).
- Chen, I.-M. A. et al. The IMG/M data management and analysis system v.7: content updates and new features. *Nucleic Acids Res.* **51**, D723–D732 (2023).
- Matheus Carnevali, P. B. et al. Hydrogen-based metabolism as an ancestral trait in lineages sibling to the Cyanobacteria. *Nat. Commun.* **10**, 463 (2019).

34. Soo, R. M., Hemp, J. & Hugenholtz, P. Evolution of photosynthesis and aerobic respiration in the cyanobacteria. *Free Radic. Biol. Med.* **140**, 200–205 (2019).
35. Sánchez-Baracaldo, P., Raven, J. A., Pisani, D. & Knoll, A. H. Early photosynthetic eukaryotes inhabited low-salinity habitats. *Proc. Natl. Acad. Sci. USA.* **114**, E7737–E7745 (2017).
36. Deusch, O. et al. Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. *Mol. Biol. Evol.* **25**, 748–761 (2008).
37. Ochoa De Alda, J. A. G., Esteban, R., Diago, M. L. & Houmard, J. The plastid ancestor originated among one of the major cyanobacterial lineages. *Nat. Commun.* **5**, 4937 (2014).
38. Li, L. et al. The genome of *Prasinoderma coloniale* unveils the existence of a third phylum within green plants. *Nat. Ecol. Evol.* **4**, 1220–1231 (2020).
39. Yang, Z. et al. Phylotranscriptomics unveil a Paleoproterozoic–Mesoproterozoic origin and deep relationships of the Viridiplantae. *Nat. Commun.* **14**, 5542 (2023).
40. Figueroa-Martinez, F., Jackson, C. & Reyes-Prieto, A. Plastid genomes from diverse glaucophyte genera reveal a largely conserved gene content and limited architectural diversity. *Genome Biol. Evol.* **11**, 174–188 (2018).
41. Prieto, A. R., Russell, S., Martinez, F. F. & Jackson, C. Comparative Plastid Genomics of Glaucophytes. in (eds Chaw, S.-M. & Jansen, R. K.) *Advances in Botanical Research*, Vol. 85, Ch 4, 95–127 (Academic Press, 2018).
42. Fučíková, K. et al. New phylogenetic hypotheses for the core Chlorophyta based on chloroplast sequence data. *Front. Ecol. Evol.* **2**, 63 (2014).
43. 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).
44. 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).
45. Kim, J. I. et al. The plastid genome of the Cryptomonad *Teleaulax amphioxeia*. *PLoS ONE* **10**, e0129284 (2015).
46. 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).
47. Cavalier-Smith, T. Symbiogenesis: mechanisms, evolutionary consequences, and systematic implications. *Annu. Rev. Ecol. Evol. Syst.* **44**, 145–172 (2013).
48. Sommer, M. S. et al. Der1-mediated preprotein import into the periplastid compartment of chromalveolates? *Mol. Biol. Evol.* **24**, 918–928 (2007).
49. Hempel, F., Bullmann, L., Lau, J., Zauner, S. & Maier, U. G. ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms. *Mol. Biol. Evol.* **26**, 1781–1790 (2009).
50. Ponce-Toledo, R. I., Moreira, D., López-García, P. & Deschamps, P. Molecular phylogeny of the SELMA translocation machinery recounts the evolution of complex photosynthetic eukaryotes. *Mol. Biol. Evol.* **42**, msaf167 (2025).
51. Jiang, Y. et al. A chlorophyll *c* synthase widely co-opted by phytoplankton. *Science* **382**, 92–98 (2023).
52. Jinkerson, R. E. et al. Biosynthesis of chlorophyll *c* in a dinoflagellate and heterologous production *in planta*. *Curr. Biol.* **34**, 594–605.e4 (2024).
53. Kim, J. I. et al. Evolutionary dynamics of cryptophyte plastid genomes. *Genome Biol. Evol.* **9**, 1859–1872 (2017).
54. Rice, D. W. & Palmer, J. D. An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. *BMC Biol.* **4**, 31 (2006).
55. Choi, C. J. et al. Newly discovered deep-branching marine plastid lineages are numerically rare but globally distributed. *Curr. Biol.* **27**, R15–R16 (2017).
56. Richardson, L. et al. MGnify: the microbiome sequence data analysis resource in 2023. *Nucleic Acids Res.* **51**, D753–D759 (2023).
57. Jamy, M. et al. Identification of a deep-branching lineage of algae using environmental plastid genomes. *Nat. Commun.* **17**, 662 (2026).
58. Watanabe, M. M., Suda, S., Inouya, I., Sawaguchi, T. & Chihara, M. *Lepidodinium viride* gen. et sp. nov. (gymnodiniales, dinophyta), a green dinoflagellate with a chlorophyll *a*- and *b*-containing endosymbiont. *J. Phycol.* **26**, 741–751 (1990).
59. Bennett, M. S., Wiegert, K. E. & Triemer, R. E. Characterization of *Euglenaformis* gen. nov. and the chloroplast genome of *Euglenaformis* [*Euglena*] *proxima* (Euglenophyta). *Phycologia* **53**, 66–73 (2014).
60. Suzuki, S., Hirakawa, Y., Kofuji, R., Sugita, M. & Ishida, K. Plastid genome sequences of *Gymnochlorella stellata*, *Lotharella vacuolata*, and *Partenskyella glossopodia* reveal remarkable structural conservation among chlorarachniophyte species. *J. Plant Res.* **129**, 581–590 (2016).
61. Turmel, M., Gagnon, M.-C., O’Kelly, C. J., Otis, C. & Lemieux, C. The chloroplast genomes of the green algae *Pyramimonas*, *Monomastix*, and *Pycnococcus* shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of Euglenids. *Mol. Biol. Evol.* **26**, 631–648 (2009).
62. Kamikawa, R. et al. Plastid genome-based phylogeny pinpointed the origin of the green-colored plastid in the dinoflagellate *Lepidodinium chlorophorum*. *Genome Biol. Evol.* **7**, 1133–1140 (2015).
63. Eloë-Fadrosch, E. A. et al. Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs. *Nat. Commun.* **7**, 10476 (2016).
64. Mackelprang, R., Vaishampayan, P. & Fisher, K. Adaptation to environmental extremes structures functional traits in biological soil crust and hypolithic microbial communities. *mSystems* **7**, e01419–e01421 (2022).
65. Tauer, B., Reichert, E. T. & Ward, L. M. Microbial mat metagenomes from Waikite Valley, Aotearoa New Zealand. Preprint at <https://doi.org/10.48550/arXiv.2412.01649> (2024).
66. Podowski, J. C., Paver, S. F., Newton, R. J. & Coleman, M. L. Genome streamlining, Proteorhodopsin, and organic nitrogen metabolism in freshwater nitrifiers. *mBio* **13**, e02379–21 (2022).
67. Fernandes-Martins, M. C., Colman, D. R. & Boyd, E. S. Relationships between fluid mixing, biodiversity, and chemosynthetic primary productivity in Yellowstone hot springs. *Environ. Microbiol.* **25**, 1022–1040 (2023).
68. Alteio, L. V. et al. Complementary metagenomic approaches improve reconstruction of microbial diversity in a forest soil. *mSystems* **5**, <https://doi.org/10.1128/msystems.00768-19> (2020).
69. Rodríguez-R, L. M., Tsementzi, D., Luo, C. & Konstantinidis, K. T. Iterative subtractive binning of freshwater chronoserics metagenomes identifies over 400 novel species and their ecologic preferences. *Environ. Microbiol.* **22**, 3394–3412 (2020).
70. Avila-Magaña, V. et al. Elucidating gene expression adaptation of phylogenetically divergent coral holobionts under heat stress. *Nat. Commun.* **12**, 5731 (2021).
71. Pérez Castro, S. et al. Diversity at single nucleotide to pangenome scales among sulfur cycling bacteria in salt marshes. *Appl Environ. Microbiol.* **89**, e00988-23 (2023).
72. Acinas, S. G. et al. Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities. *Commun. Biol.* **4**, 604 (2021).
73. Kang, D. D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).

74. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**, 119 (2010).
75. Parks, D. H. et al. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* **50**, D785–D794 (2022).
76. Chklovski, A., Parks, D. H., Woodcroft, B. J. & Tyson, G. W. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat. Methods* **20**, 1203–1212 (2023).
77. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **11**, e0163962 (2016).
78. Shaw, J. & Yu, Y. W. Fast and robust metagenomic sequence comparison through sparse chaining with skani. *Nat. Methods* **20**, 1661–1665 (2023).
79. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
80. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
81. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
82. Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
83. Mussig, A. J. PhyloDM. Preprint at Zenodo <https://doi.org/10.5281/zenodo.6910552> (2022).
84. Van Dongen, S. Graph clustering via a discrete uncoupling process. *SIAM J. Matrix Anal. Appl.* **30**, 121–141 (2008).
85. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
86. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2012).
87. Guillou, L. et al. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* **41**, D597–D604 (2012).
88. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
89. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
90. Tillich, M. et al. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **45**, W6–W11 (2017).
91. Chan, P. P., Lin, B. Y., Mak, A. J. & Lowe, T. M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**, 9077–9096 (2021).
92. Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **5**, e11147 (2010).
93. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
94. Shrestha, B. et al. Global Metagenomics Reveals Plastid Diversity and Unexplored Algal Lineages. figshare <https://doi.org/10.6084/m9.figshare.31146589> (2026).
- Facility, is supported by the Office of Science of the U.S. Department of Energy, operated under Contract No. DE-AC02-05CH11231. The work on chlorophyte algal plastids by B.S. and C.E.B-H. were supported by the Department of Energy Office of Science, Biological and Environmental Research program under award no. DE-SC0023027. The work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Author contributions

Conceptualization and methodology: F.S., C.E.B-H., and B.S. Investigation: C.E.B-H. and B.S. Formal analyses: M.F.R. and J.C.V. generated MAGs associated with Cyanobacteriota and plastids. B.S. performed phylogenetic analyses. B.S. drafted the initial manuscript. C.E.B-H. and F.S. revised the manuscript. All authors read and approved the final manuscript.

Competing interests

F.S. serves as CEO of SampleX. This work was not funded by, nor does it benefit SampleX. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-026-68871-w>.

Correspondence and requests for materials should be addressed to Bikash Shrestha, Crysten E. Blaby-Haas or Frederik Schulz.

Peer review information *Nature Communications* thanks Pavel Skaloud and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026

Acknowledgements

The work conducted by the U.S. Department of Energy Joint Genome Institute (<https://ror.org/04xm1d337>), a DOE Office of Science User

Plastid MAG Consortium

Katherine McMahon³, Ramunas Stepanauskas⁴, Alison Buchan⁵, Thomas Mock⁶, Kirsten Fisher⁷, Joan Slonczewski⁸, Luce Ward⁹, Sean Crowe¹⁰ & William Inskeep¹¹

³Department of Civil and Environmental Engineering and Bacteriology, University of Wisconsin-Madison, Madison, WI, USA. ⁴Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, USA. ⁵Department of Microbiology, University of Tennessee-Knoxville, Knoxville, TN, USA. ⁶School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK. ⁷Department of Biological Sciences, California State University, Los Angeles, CA, USA. ⁸Kenyon College, Gambier, OH, USA. ⁹Department of Geosciences, Smith College, Northampton, MA, USA. ¹⁰Department of Microbiology and Immunology, and Department of Earth, Ocean, and Atmospheric Sciences, The University of British Columbia, Vancouver, BC, Canada. ¹¹Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT, USA.