Rampant nuclear–mitochondrial–plastid phylogenomic discordance in globally distributed calcifying microalgae

Tzu-Tong Kao (iD)**, Tzu-Haw Wang** (iD) **and Chuan Ku** (iD)

Institute of Plant and Microbial Biology, Academia Sinica, Taipei 11529, Taiwan

Summary

- Incongruent phylogenies have been widely observed between nuclear and plastid or mitochondrial genomes in terrestrial plants and animals. However, few studies have examined these patterns in microalgae or the discordance between the two organelles.
- Here we investigated the nuclear–mitochondrial–plastid phylogenomic incongruence in *Emiliania–Gephyrocapsa*, a group of cosmopolitan calcifying phytoplankton with enormous populations and recent speciations. We assembled mitochondrial and plastid genomes of 27 strains from across global oceans and temperature regimes, and analyzed the phylogenomic histories of the three compartments using concatenation and coalescence methods.
- Six major clades with varying morphology and distribution are well recognized in the nuclear phylogeny, but such relationships are absent in the mitochondrial and plastid phylogenies, which also differ substantially from each other. The rampant phylogenomic discordance is due to a combination of organellar capture (introgression), organellar genome recombination, and incomplete lineage sorting of ancient polymorphic organellar genomes. Hybridization can lead to replacements of whole organellar genomes without introgression of nuclear genes and the two organelles are not inherited as a single cytoplasmic unit.
- This study illustrates the convoluted evolution and inheritance of organellar genomes in isogamous haplodiplontic microalgae and provides a window into the phylogenomic complexity of marine unicellular eukaryotes.

Introduction

Eukaryotes differ from prokaryotes by their nucleus and organelles of endosymbiotic origins (Zimorski *et al.*, 2014). Most genes in the bacterial ancestors of mitochondria and plastids have been transferred to the nuclear genome (Ku *et al.*, 2015). However, endosymbiotic organelles generally retain a remnant genome encoding fewer than *c.* 200 genes (Smith & Keeling, 2015), whose localization could be essential for the redox regulation of their expression and bioenergetic functions (Allen, 2015). The simple organization, relatively conserved gene contents, and generally clonal and uniparental inheritance (Birky, 2008) of mitochondrial genomes (mitogenomes) and plastid genomes (plastomes) make them widely used for phylogenetic analyses of different eukaryotes at various taxonomic levels (Galtier *et al.*, 2009; Dong *et al.*, 2012). Organellar markers are also analyzed together with nuclear markers to provide better resolution in phylogenies (Phillips *et al.*, 2006; Liu *et al.*, 2010; Lockwood *et al.*, 2013).

However, incongruent phylogenies have been observed between nuclear and organellar genes or genomes. Such cytonuclear discordance, either mito-nuclear or plastid-nuclear, can be caused by a variety of processes, including sex-biased inheritance, introgression leading to organellar capture (whole-genome replacement) (Soltis & Kuzoff, 1995; Folk *et al.*, 2017; Lee-Yaw *et al.*, 2019; Zhang *et al.*, 2019; Sarver *et al.*, 2021), and incomplete lineage sorting (ILS) (Meleshko *et al.*, 2021; Rose *et al.*, 2021). Cytonuclear phylogenetic discordance can result from inter-compartmental differences, such as higher level of introgression in mitochondrial than nuclear genome in some cases (Bonnet *et al.*, 2017; Sarver *et al.*, 2021). Modes of organelle inheritance, which might not be strictly clonal and uniparental, are also important factors. Seed plants have different mechanisms of organellar exclusion, with maternal organelle inheritance in most angiosperms but paternal plastid inheritance in gymnosperms (Greiner *et al.*, 2015). In contrast to sperm-egg mating (oogamy) in plants and animals, in isogamous eukaryotes, two morphologically indistinguishable gametic cells with similar abundance of organelles and organellar DNA fuse into a zygote. One example is budding yeasts, which have biparental inheritance of mitochondria and transient heteroplasmy (i.e. different organellar haplotypes in one cell) in the zygote (Solieri, 2010), where mitochondria fuse and different haplotypes might recombine (Fritsch *et al.*, 2014). In another isogamous eukaryote, the green microalga *Chlamydomonas reinhardtii*, plastid DNA is inherited from the mt^+ and mitochondrial DNA from the mt^- mating type through selective elimination of organellar DNA from the opposite type (Miyamura, 2010). For most eukaryotes,

which are unicellular and isogamous (Lehtonen *et al.*, 2016), the modes of organelle inheritance are yet to be characterized.

Phylogenomic analysis is a powerful method for unraveling the inheritance history of different genomes, but our current understanding of cytonuclear discordance at the whole-genome level is largely based on terrestrial animal and plant systems. In addition, most studies focused on only the nuclear and mitochondrial or only the nuclear and plastid genome. Despite the importance and widespread distribution of plastids across the eukaryotic tree of life—plants, diverse algae, and other plastid-bearing protists (Keeling, 2013; Zimorski *et al.*, 2014), differences in the phylogenomic history between the two organelles, plastids and mitochondria, remain poorly understood. Multigene phylogenetic analyses of all three compartments have been limited to a few studies on plants, which revealed varying degrees of incongruence among the three genomes (Folk *et al.*, 2017; Vargas *et al.*, 2017; Meleshko *et al.*, 2021). In multicellular coralline red algae, mitogenomes and plastomes support different groupings of taxonomic orders (Lee *et al.*, 2018). For other algae, including the most abundant microalgae, very little is known about their inter-compartmental phylogenomic discordance.

Microalgae are unicellular photosynthetic eukaryotes that have been shaping global biogeochemical cycling and marine food webs for at least 1.5 billion years (Falkowski *et al.*, 2004; Worden *et al.*, 2015). Investigating the origins and evolution of dominant microalgal species is key to our understanding of marine microbial communities and their succession through geological time. One unique group of microalgae is the calcifying phytoplankton, coccolithophores (Haptophyta), which convert carbon into organic photosynthates as well as the inorganic calcium carbonate ($CaCO_3$) plates called coccoliths (Taylor *et al.*, 2017). Coccolithophores account for 10% of annual global primary production (Rousseaux & Gregg, 2013). The most common coccolithophore, *Emiliania huxleyi*, is ubiquitous across global oceans and forms massive annual blooms in temperate and subpolar regions (Balch, 2018). Based on fossil and oxygen isotope records, *E. huxleyi* originated < 0.3 Ma (Raffi *et al.*, 2006) and became the most dominant coccolithophore morphospecies around 0.085 Ma (Thierstein *et al.*, 1977), making it an excellent model for studying recent global expansion of marine microbes. Nuclear phylogenomic analyses suggested that *E. huxleyi* evolved from within *Gephyrocapsa* (Noëlaerhabdaceae, Isochrysidales), which has been abundant throughout the Quaternary (Bendif *et al.*, 2019). Despite its extremely large population size, the nuclear genetic diversity of *E. huxleyi* is unexpectedly limited, possibly as a result of strong linked selection rather than unusually low mutation rates (Filatov, 2019; Krasovec *et al.*, 2020). Coccolithophores are apparently isogamous (Houdan *et al.*, 2004), but their mating process is still obscure. Similar to *C. reinhardtii*, *E. huxleyi* has a single cup-shaped plastid, and its mitochondria have variable shapes characteristic of dynamic fusion and fission (Uwizeye *et al.*, 2021).

The complete mitogenome and plastome sequences are available from *E. huxleyi* strains CCMP373 and CCMP1516 (Sánchez-Puerta *et al.*, 2004, 2005; Smith & Keeling, 2012). The mitogenome is *c.* 29 kb, with a 353-bp difference between the two strains, and encodes 21 proteins, which are more than in animals and fungi, but fewer than in most plants (Zardoya, 2020). The plastome is *c.* 105.3 kb in both strains, with a much smaller inverted repeat region than that of plants (Zhu *et al.*, 2016) and more proteins encoded than in plants, which is consistent with its origin from gene-rich plastomes of red algae (Sánchez-Puerta *et al.*, 2005; Green, 2011). Phylogenies of nuclear ribosomal DNA (rDNA), plastid *tufA*, and mitochondrial *cox1* and *cox3* genes have apparently incongruent topologies for some strains in the *Emiliania–Gephyrocapsa* clade (Bendif *et al.*, 2016). However, these trees had largely polytomous nodes, likely because one or two genes generally do not bear sufficient phylogenetic information to resolve the relationships between closely related organisms. Besides, genes within the same genome can have conflicting phylogenies due to distinct evolutionary histories. To further understand the extent and potential mechanisms of inter-compartmental discordance, whole-genome assemblies and sequences are required.

In this study, we explore whole genomes of *Emiliania–Gephyrocapsa* to assess nuclear–mitochondrial–plastid phylogenomic incongruence in these isogamous marine unicellular eukaryotes with astronomically large populations, recent speciation, and rapid global expansion. We assembled complete or near-complete mitogenomes and plastomes, inferred their highly resolved phylogenies, characterized intergenomic and intragenomic incongruences, and finally elucidated the underlying evolutionary processes. Our results reveal widespread phylogenomic discordance across the three compartments, which resulted from a combination of introgression, recombination, and ILS.

Materials and Methods

Strains and genomic data

Genomic data of 29 *Emiliania–Gephyrocapsa* strains (23 *E. huxleyi*, one *G. ericsonii*, one *G. muellerae*, two *G. oceanica*, and two *G. parvula*) from across global oceans and temperature regimes were retrieved from the National Center for Biotechnology Information (NCBI) or newly sequenced in this study (Supporting Information Table S1). The data include one nuclear (CCMP1516; Read *et al.*, 2013) and two mitochondrial and plastid (CCMP373 and CCMP1516; Sánchez-Puerta *et al.*, 2004, 2005; Smith & Keeling, 2012) genome assemblies, as well as 27 Illumina sequence read archives (Read *et al.*, 2013; von Dassow *et al.*, 2015; Bendif *et al.*, 2019; this study). For the newly sequenced strains, a single-cell-derived culture was prepared by isolating individual cells using a MoFlo XDP cell sorter (Beckman Coulter Inc., Brea, CA, USA), followed by incubation in modified K/2 medium (Keller *et al.*, 1987; Probert, 2019) with an antibiotic cocktail of cefotaxime (50 ppm), carbenicillin (50 ppm), kanamycin (20 ppm), and amoxicillin (20 ppm) to reduce bacterial contamination. Approximately 1×10^8 cells were harvested by centrifugation (15 min at 3000 ***g***) at the end of log growing phase. The cells were embedded in 1.5% low-melting-point agarose (50101; Lonza Group AG, Basel, Switzerland) and digested with proteinase K following a published

method (Agarkova *et al.*, 2006). From the gel plugs, high-molecular-weight DNA was extracted according to Bionano Prep Cell Culture DNA Isolation protocol (30026, revision F). Linked-read libraries were prepared using Chromium™ Genome Library & Gel Bead Kit v.2 (10× Genomics Inc., Pleasanton, CA, USA) and sequenced on Illumina NovaSeq 6000 (Illumina Inc., San Diego, CA, USA).

Organellar genome assembly and annotation

Adapters and low-quality ends (Phred score < 20) of sequencing reads were trimmed using BBDuk in BBTools (http://sourceforge.net/projects/bbmap). Both mitogenome and plastome sequences were assembled either using NOVOPlasty v.4.1 (Dierckxsens *et al.*, 2016; for all strains except B39, B11, EH2, M217, and M219) with the corresponding CCMP1516 genomes as seed or *de novo* using SPADES v.3.15 (Bankevich *et al.*, 2012; for low-coverage single-end sequencing data of B39, B11, EH2, M217, and M219) with the default parameters. The assemblies were polished through mapping reads with Bwa-Mem v.0.7.17 (Li, 2013), inspecting with IGV v.2.10.2 (Thorvaldsdottir *et al.*, 2013), and editing in AliView v.1.26 (Larsson, 2014). We were unable to circularize most mitogenomes due to a highly repetitive region between *trnI* and *rps12*. In addition, the highly fragmented mitogenome assemblies of B39 and M219 missed at least two genes. The highly repetitive region and the two strains were thus excluded from our subsequent analyses of mitogenomes. Organellar protein-coding genes were predicted using BLAT (Kent, 2002) and annotated and visualized using GeSeq (Tillich *et al.*, 2017) (Dataset S1).

Identification of putative single-copy core nuclear genes

Whole-genome *de novo* assemblies were constructed using SPADES (plus scaffolding using ARKS in ARCS v.1.1.1 (Coombe *et al.*, 2018) for the strains with 10× linked-read data) and annotated using AUGUSTUS v.3.3 (Stanke *et al.*, 2006, 2008). The gene model was trained with GenomeThreader (Gremme *et al.*, 2005) in Braker2 v.2.1.5 (Hoff *et al.*, 2016, 2019; Brůna *et al.*, 2021) using the nuclear coding sequences of CCMP1516 (Read *et al.*, 2013) and the assembly of the haploid strain RCC1217 (this study). The single-copy genes shared by all strains were identified by OrthoFinder v.2.5.2 (Emms & Kelly, 2019).

Phylogenomic analyses

Five datasets were prepared (Dataset S2), including three coding sequence datasets: (1) 132 putative single-copy nuclear genes shared by all 29 strains except CCMP373, which lacks nuclear data; (2) 20 mitogenome shared genes (excluding *dam*, which was incompletely assembled for some strains); (3) 112 plastome shared genes (excluding open reading frames (ORFs) not widely present across strains); and two whole-organellar nucleotide datasets: (4) whole-mitogenome (excluding the highly repetitive region); (5) whole-plastome. The coding sequences were aligned using Muscle v.3.8.1 in AliView (Edgar, 2004) and the

organellar whole-genome nucleotide sequences were aligned using MAFFT v.7.310 (Katoh & Standley, 2013). All alignments were inspected and edited manually to exclude poorly aligned regions flanking indels.

Phylogenies of nuclear, mitochondrial, and plastid genomes were inferred separately using a concatenation maximum likelihood (ML) approach (IQ-Tree 2 v.2.1.4-beta; Minh *et al.*, 2020b) and two coalescent methods, including one based on gene trees inferred by IQ-Tree 2 (Astral-III v.5.7.7; Zhang *et al.*, 2018) and the other based on single-nucleotide polymorphisms (SVDQuartets in PAUP* 4.0a build 168; Chifman & Kubatko, 2014, 2015). The best partition schemes and codon evolutionary models were selected using PartitionFinder (Lanfear *et al.*, 2012). Branch supports were calculated using 1000 replicates of ultrafast bootstrap that resamples both genes and sites (Hoang *et al.*, 2018) and SH-like aLRT (Guindon *et al.*, 2010). We also tried to infer the trees using amino acid or nucleotide models, to partition the datasets using both genes and coding positions, or to reduce the effects of positive selection by using only third codon positions (Dataset S3), but no conflict in strongly supported branches (UFBoot > 95%) was observed.

Analyses of phylogenomic discordance

To visualize inter-genomic discordance, ML trees were fitted to chronograms using chronos function in R package APE (Paradis & Schliep, 2019) and tanglegrams were plotted between different compartments using R package DENDEXTEND (Galili, 2015). Gene and site concordance factors were calculated using the '--gcf' and '--scf' functions in IQ-Tree (Minh *et al.*, 2020a). Phylogenomic networks were built using SplitsTree4 (Huson & Bryant, 2006), with super-networks inferred from individual IQ-Tree gene trees and neighbor-nets from ML distances under GTR + I + G model. Presence of hybridization among major clades was tested with an 'ABBA-BABA'-like method (HyDe; Blischak *et al.*, 2018). Pairwise substitution rates within each coding datasets and whole-genome datasets were calculated using codeml (F3X4 model) and baseml (GTR model), respectively, in PamL 4 (Yang, 2007). Recombination in organellar genomes was analyzed using RDP5 v.5.5 with seven different methods: RDP, GENECONV, MAXCHI, CHIMAERA, BOOTSCAN, SISCAN, and 3SEQ (Martin *et al.*, 2021).

Results

Genome assemblies

Nuclear, mitochondrial, and plastid genomes of 29 *Emiliania–Gephyrocapsa* strains (Fig. 1a) were analyzed in this study, including new assemblies from 27 strains (Table S1; Figs S1, S2; Dataset S1). To ensure the assembly correctness, we inspected the mapping of reads onto each assembly. We found a few ambiguous nucleotides (usually R and Y), especially in the highly repetitive regions of mitogenomes and the inverted repeats of plastomes, but there was no sign of heteroplasmy or DNA contaminations. We also double-checked the sequence identifiers

(IDs) and confirmed that there was no mis-labeling of any strain and genomes. In support of the accuracy of genome assemblies, strains originally derived from the same laboratory source, including RCC1216/RCC1217 (von Dassow *et al.*, 2009), CCMP1516/AWI1516/M217 (Read *et al.*, 2013; Zhang *et al.*, 2016), or M219/NZEH (Read *et al.*, 2013), are closely grouped together in phylogenies (see later), although they have been maintained at and transferred between culture collection centers and laboratories for 20 yr or longer.

The mitogenome and plastome median size are 28.8 kb (25.3–31.1 kb except 45.4 kb in RCC1296) and 105.3 kb (105.3–109.2 kb), respectively. The median GC contents of whole genomes, mitogenomes, and plastomes are 63.4% (53.4–65.9%), 26.6% (26.0–31.3%), and 36.8% (36.7–37.5%), respectively (Table S1). The whole-genome GC contents of *Emiliania* (61.3–65.9%) are generally higher than those of *Gephyrocapsa* (*G. muellerae*: 53.4%; *G. oceanica*: 55.3 and 60.6%; *G. parvula*: 60.7 and 61.1%; *G. ericsonii*: 64.34%), but the mitogenome GC contents of *Emiliania* (26.0–28.6%) are mostly lower than those of *Gephyrocapsa* (28.5–31.3%).

The mean pairwise synonymous substitution rates (dS) from high to low are nuclear genome (0.0334), mitogenome (0.0289), and plastome (0.0235) (Table 1). The average dN/dS values of all three genomes are far below 1 (nuclear: 0.100 (0.000–1.212); mitochondrial: 0.055 (0.000–0.606); plastid: 0.055 (0.000–0.478); Table S2), indicating the coding sequences are generally under purifying selection (Yang & Bielawski, 2000). The dN/dS value of one nuclear gene, EOD22827 (EMIHUDRAFT 447689; a cold-shock protein), is > 1, which is suggestive of strong diversifying selection.

Nuclear core-gene phylogeny

To avoid confounding effects of gene paralogy, we only included putatively single-copy genes shared by the nuclear genomes. The ML phylogeny inferred from the concatenated alignment of the coding sequences (Fig. 1b) is largely congruent with the trees inferred using coalescent methods (Dataset S3), as well as the ones inferred using other derived datasets (e.g. only third codon positions), partitions, and evolutionary models. No conflict was observed between the trees at any nodes with strong support (SH-aLRT support $\geq$ 80% and UFBoot support $\geq$ 95% (IQ-TREE 2), bootstrap support $\geq$ 70% (ASTRAL-III and SVDQUARTETS), or posterior possibility $\geq$ 0.95 (ASTRAL-III)).

The ML phylogeny divides the *Emiliania–Gephyrocapsa* strains into six major clades: GOC, GMU, GEP, EHB, EHA2, and EHA1 (Fig. 1b). Overall the nuclear genome phylogeny reflects the evolution of observable traits – morphological, physiological, or ecological. Large, medium, and small *Gephyrocapsa* cells correspond to clades GOC, GMU, and GEP, respectively. *Gephyrocapsa oceanica* (GOC) is inferred as the earliest divergent lineage within *Emiliania–Gephyrocapsa* based on midpoint rooting, which is consistent with previous studies (Bendif *et al.*, 2019; Filatov *et al.*, 2021). The other three *Gephyrocapsa* species are in two clades, GMU (*G. muellerae*) and GEP (*G. ericsonii* and *G. parvula*). Together they form a clade (GMEP) sister to the
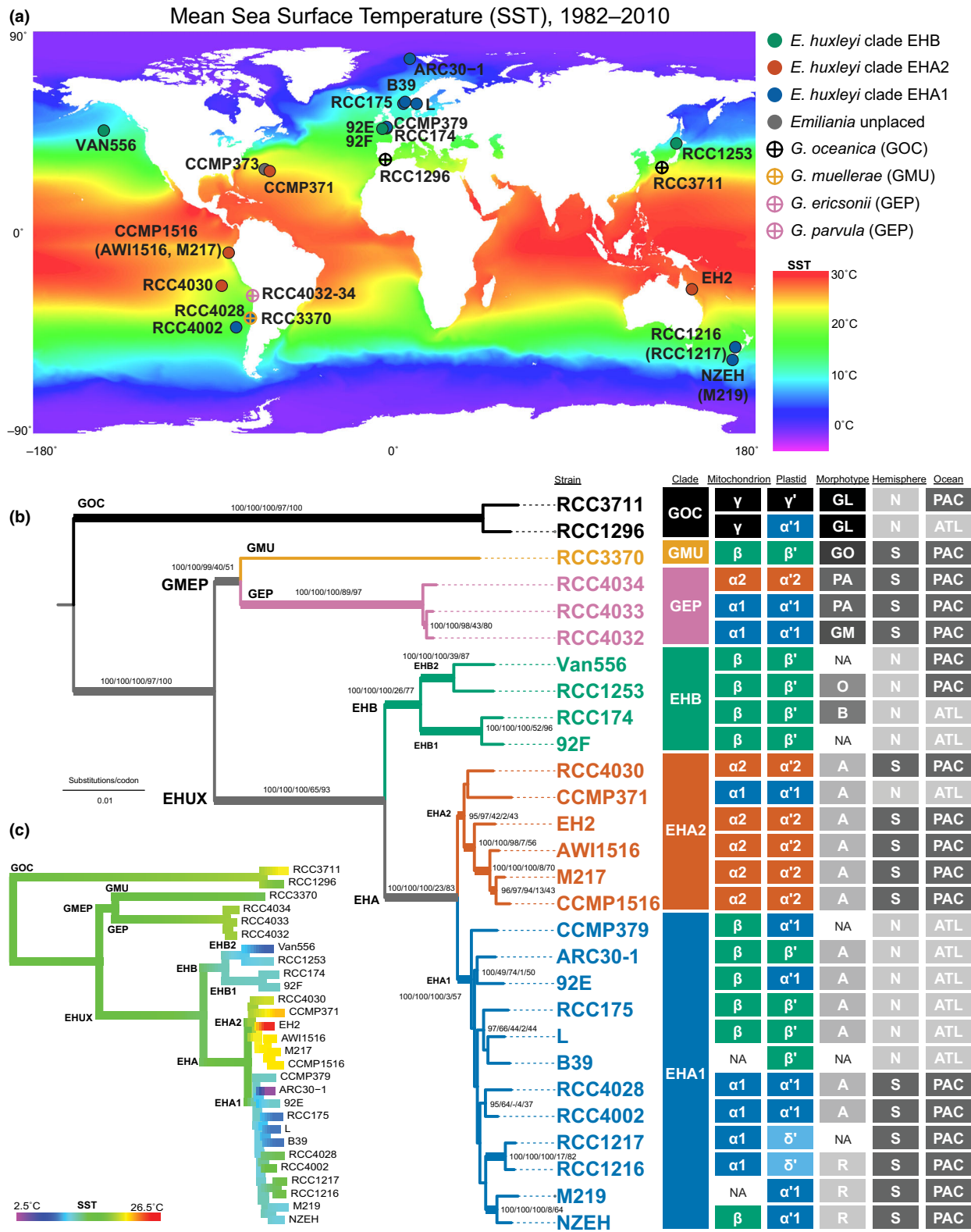
strongly supported monophyletic species *E. huxleyi* (EHUX). Within EHUX, two well-supported clades with different coccolith morphology are recognized – EHB with lightly calcified coccoliths (morphotypes B and O) and EHA with heavily calcified coccoliths (morphotypes A and R). Each of them consists of two subgroups: the Atlantic subgroup EHB1, Pacific EHB2, temperate EHA1, and low-latitude EHA2. According to the mean sea surface temperatures (SSTs) between 1982 and 2010 (Fig. 1a) and the reconstructed ancestral temperatures (Fig. 1c), *Gephyrocapsa* clades are within the range of SST 18–22°C. EHB (SST 9.0–12.7°C) and EHA1 (SST 2.6–16.7°C) are distributed in temperate and even artic zones, whereas EHA2 is found in much warmer tropical regions (SST 20.9–26.4°C) similar to the ancestral state of EHUX as inferred from fossil records (Filatov *et al.*, 2021).

Organellar genome diversity and phylogenies

In addition to the published sequences of CCMP373 and CCMP1516, we successfully assembled complete or near-complete organellar genome sequences, including a total of 27 plastome and 25 mitogenome (excluding M219 and B39 which had insufficient mitogenome reads) sequences (Table S1). We inferred the phylogenies of mitogenomes and plastomes from the concatenated alignments of their coding (Fig. 2) and whole genome sequences (Dataset S2) using the IQ-TREE ML method with different models and partitions as well as two coalescent methods, which resulted in largely congruent topologies for each organelle. Several well-supported clades can be recognized in the mitogenome and plastome phylogenies, which are additionally supported by some phylogenetically informative indels (4–3698 bp; Fig. 2). Following the naming system used for mitochondrial *cox1-cox3* and plastid *tufA* phylogenies (Bendif *et al.*, 2016), we named the mitogenome clades α (subdivided into α1 and α2), β, and γ, and the plastome clades α′ (subdivided into α′1 and α′2), β′, γ′, and δ′.

The γ mitogenome of *G. oceanica* is the most divergent mitogenome lineage (Fig. 2a) with two distinct 0.8–2.6 kb indels absent in the other *Emiliania–Gephyrocapsa* mitogenomes. One of them contains ORF584, also reported in *Chrysochromulina* (Nishimura *et al.*, 2014), and the other contains partially duplicated *nad3* and *cox2* genes. The mitogenome size and GC content are notably different between the two *G. oceanica* strains (RCC3711: 28.2 kb, 28.9%; RCC1296: 45.4 kb, 31.3%). The much larger size in RCC1296 is due to the presence of several unique regions, including a 2419-bp insertion upstream of exon a of *cox1*, a 2595-bp insertion in *cob*, a 758-bp insertion between *nad2* and *nad3* that includes partially duplicated *atp6* and ORF104, and a 8271-bp insertion in the second intron of *cox1* that contains ORF627 and ORF636, which are also found in the noncalcifying haptophytes *Chrysochromulina* (Nishimura *et al.*, 2014) and *Diacronema* (Hulatt *et al.*, 2020), respectively (Fig. S1; Notes S1).

The β mitogenomes mainly occur in temperate strains (SST 9.0–15.0°C), including all members of the nuclear clade EHB, more than half of EHA1 strains, and GMU. The α2

(a) Mean Sea Surface Temperature (SST), 1982–2010

- ● *E. huxleyi* clade EHB
- ● *E. huxleyi* clade EHA2
- ● *E. huxleyi* clade EHA1
- ● *Emiliania* unplaced
- ⊕ *G. oceanica* (GOC)
- ⊕ *G. muellerae* (GMU)
- ⊕ *G. ericsonii* (GEP)
- ⊕ *G. parvula* (GEP)

mitogenomes are from strains distributed in the tropics of South Pacific (SST 19.6–26.4°C), including five out of six EHA2 strains and one GEP strain. The α1 mitogenomes are geographically widely distributed (SST 2.6–23.1°C) and found in four EHA1, one EHA2, and two GEP strains (Fig. 2a).

Different from the nuclear and mitogenome phylogenies, GOC plastomes are in two distinct lineages, γ′ and α′. In the plastome phylogeny with midpoint rooting (Fig. 2b), γ′ is less divergent from α′ than the β′ and δ′ plastomes, suggesting the latter two evolved much faster and/or diverged much earlier than

Fig. 1 Geographical distribution and nuclear phylogeny of *Emiliania–Gephyrocapsa* strains included in this study. (a) The geographical origins of the 29 stains on a world map showing the mean sea surface temperature (SST) from 1982 to 2010. The SST source data from National Centers for Environmental Information, National Oceanic and Atmospheric Administration, USA, were analyzed and visualized in R (R Core Team, 2021). Strains in different nuclear clades are represented by different colors and symbols. (b) The maximum likelihood (ML) phylogeny of the nuclear genome was inferred from a concatenated alignment of 132 putative single-copy genes with partitions and codon models selected by PARTITIONFINDER and presented with midpoint rooting. Branches with both SH-aLRT support ≥ 80% and UFBoot support ≥ 95% are thickened. For branches with SH-aLRT support ≥ 80% or UFBoot support ≥ 95%, five numbers are shown: SH-aLRT support, UFBoot support, SVDQUARTETS bootstrap support, gene concordance factor, and site concordance factor (all in percentage). '-': the topology is not supported by the SVDQUARTETS coalescent phylogeny. Six major clades are recognized and colored. GOC (*Gephyrocapsa oceanica*), GMU (*Gephyrocapsa muellerae*), GEP (*Gephyrocapsa ericsonii* and *Gephyrocapsa parvula*), EHB, EHA2, and EHA1 (*Emiliania huxleyi*). The major clades in the mitogenome and plastome phylogenies (Fig. 2), morphotypes, and geographic origins (northern (N)/southern (S) hemisphere and Pacific (PAC)/Atlantic (ATL) Oceans) are also indicated for each strain. NA: not available. (c) Ancestral state reconstruction of SST based on the fastAnc function in the R package PHYTOOLS (Revell, 2012).

the other types of plastomes. The β′ plastomes are found in GMU and temperate northern hemisphere strains of EHB and EHA1. The δ′ lineage is unique to RCC1216 and its derived haploid strain RCC1217. These two highly divergent plastome lineages have slightly larger genome sizes (β′: 107.0–109.2 kb; δ′: 109.0 kb) than the other *E. huxleyi* (all 105.3 kb) and *Gephyrocapsa* (105.3–107.1 kb) plastomes (Table S1). The α′2 is mainly found in low-latitude EHUX (EHA2) plus one GEP strain. The α′1 clade is the most heterogeneous in terms of its corresponding nuclear clades, which include GOC, GEP, EHA2, and EHA1 (half of the strains).

Phylogenomic networks and organellar recombination

To examine the congruence of phylogenetic signals from different genes and single nucleotide polymorphisms (SNPs) within the nuclear and organellar genomes, gene and site concordance factors (Minh *et al.*, 2020a) were calculated for the nodes in the ML phylogenies with strong SH-aLRT or UFBoot supports (Figs 1b, 2). Overall, most major clades in the ML phylogenies (i.e. nuclear GOC, GEP, EHB; mitochondrial α, β, γ; and plastid α′, β′, δ′) are supported by at least a quarter of genes and three-quarters of decisive sites. In some rare cases, alternative quartets were strongly supported. One notable case in the nuclear phylogeny is that 40% of the genes support GMU and GEP as sister groups and 39% support either GEP or GMU as sister to EHUX. The discordance among gene trees was also revealed by the nuclear super-network inferred from gene trees (Fig. 3a). Nevertheless, in addition to ML trees, the sister relationship between GMU and GEP is strongly supported by trees inferred using two

different coalescent methods (SVDQUARTETS: bootstrap support 99%; ASTRAL: bootstrap support 92%, posterior possibility 0.99) and the neighbor-net inferred from ML distance (Fig. 3b). The nuclear phylogenomic grouping of GMU and GEP is therefore well supported, despite the relatively short branch uniting these two lineages (Fig. 1b).

Another relationship with notable conflicts is the monophyly of mitochondrial clade β. Although 52% of the genes support the clade of L, ARC30-1, and RCC175 as sister to the other β mitogenomes (Fig. 2a), 14% of the genes support it is more closely related to clade α. This discordance is evident in the mitogenome neighbor-net (Fig. 3c). Among the two coalescent methods, ASTRAL strongly supported the monophyly of β (bootstrap support 100%, posterior possibility 1.00), but the SVDQUARTETS bootstrap support is only 61%. To resolve these conflicting signals, we further tested recombination in mitochondrial genomes using various methods implemented in RDP5 v.5.5 (Martin *et al.*, 2021) and found clear evidence for a recombined origin of L, ARC30-1, and RCC175 mitogenomes. One recombination breakpoint is located around the spacer between *atp6* and *ORF104* and the other in the highly repetitive region between *trnI* and *rps12*. The sequence of the larger fragment (*c.* 18 kb) is highly similar to the other β mitogenomes, whereas the smaller fragment (*c.* 9 kb) is more closely related to clade α (Fig. S3). Recombination with an unknown source was also detected in the mitogenome of RCC1296, at the sequences flanking its unique 758 bp insert between *nad2* and *nad3*. Some putative recombination signals were found in almost all strains at the highly repetitive region between *trnI* and *rps12*, which could not be well aligned and was excluded from the alignment dataset for

Table 1 List of alignment datasets and sequence substitution rates.

	CDS	Alignment length (bp)	GC (%)	Missing data and gap (%)	Informative codons	Informative sites	Substitutions per site	dS	dN/dS
Nuclear CDS	132	92 025	67.14	12.86	2895	3277	0.0101	0.0334	0.1000
Mitochondrial CDS	20	15 519	30.51	0.61	394	406	0.0074	0.0289	0.0551
Whole mitogenome	—	28 090	28.53	4.68	—	981	0.0110	—	—
Plastid CDS	112	79 062	37.37	0.25	1329	1345	0.0060	0.0235	0.0552
Whole plastome	—	108 119	36.82	2.53	—	1898	0.0062	—	—

CDS, coding sequence.
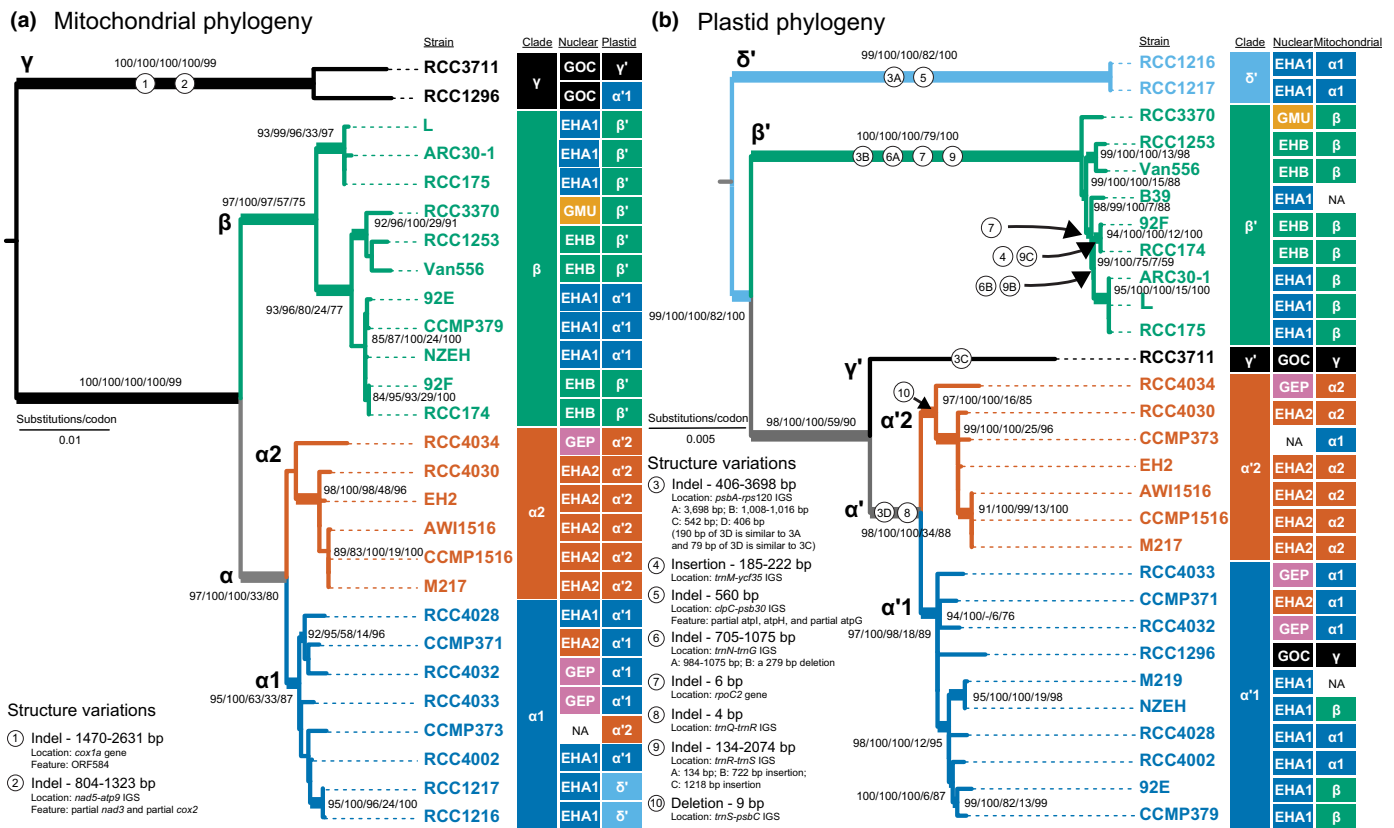
Fig. 2 Phylogenies of organellar genomes of *Emiliania–Gephyrocapsa*. (a) The mitochondrial maximum likelihood (ML) tree inferred from a concatenated alignment of 20 coding genes. (b) The plastid ML tree inferred from a concatenated alignment of 112 coding genes. Both ML trees were midpoint-rooted. Branches with both SH-aLRT test support ≥ 80% and UFBoot support ≥ 95% are thickened. For branches with SH-aLRT support ≥ 80% or UFBoot support ≥ 95%, five numbers are shown: SH-aLRT support, UFBoot support, SVDQUARTETS bootstrap support, gene concordance factor, and site concordance factor (all in percentage). '-': the topology is not supported by the SVDQUARTETS coalescent phylogeny. Four major mitogenome and five plastome clades are recognized. The clades in the nuclear phylogeny (Fig. 1b) and the other organellar phylogeny are also indicated. Nonhomopolymer indels > 3 bp (up to 3698 bp) that are phylogenetically informative (shared by ≥ 2 strains) are marked on the trees with symbols ①–⑩. NA, not available.

inferring phylogenies. We also inferred the plastome neighbor-net based on the ML distances (Fig. 3d). In contrast to mitogenomes, no strong signal of recombination was detected among the plastomes, and the clades recognized in the ML phylogeny (Fig. 2b) are also well separated from each other in the neighbor-net (Fig. 3d).

In addition, the presence of hybridizations (including introgressive hybridizations) was tested using HYDE v.0.4.3 (Blischak *et al.*, 2018). When *G. oceanica* was set as the outgroup, no admixture was detected among the clades GMU, GEP, EHA, and EHB based on the nuclear dataset (Bonferroni corrected *P*-values all > 0.1; Table S3). When all three genomes were combined in one analysis based on the nuclear topology, GEP was inferred as a mixture of EHUX (EHA or EHB) and GMU, with a higher probability that it is sister to EHUX ($\hat{\gamma} > 0.5$; EHA–GMU $\hat{\gamma} = 0.836$, EHB–GMU $\hat{\gamma} = 0.761$), and EHA is a mixture of EHB and GMEP (GMU or GEP), with a higher probability that it is sister to EHB (EHB–GMU $\hat{\gamma} = 0.919$) or GEP (EHB–GEP $\hat{\gamma} = 0.400$). However, the $\hat{\gamma}$ value is biased due to dataset sizes nonproportional to the genome sizes, and the degree of hybridization of different strains within a clade is not uniform.

Intergenomic comparisons of tree topologies

We compared the nuclear, mitochondrial, and plastid ML trees in pairwise tanglegrams, which reveal widespread discordance at different phylogenetic depths among the three genomes (Fig. 4). In both nuclear and mitochondrial phylogenies, the two GOC strains RCC3711 and RCC1296 form a distinct clade sister to all the other taxa (Fig. 4a), whereas in the plastid phylogeny, RCC3711 remains distinct (γ'), but the other strain RCC1296 is firmly nested within clade α' (Fig. 4b,c). The phylogenetic placements of the other four *Gephyrocapsa* strains show even starker nuclear-organellar incongruence. They form a well-supported clade (GMEP) in the nuclear tree and are subdivided into the sister clades GMU and GEP, where *G. ericsonii* (RCC4032) is closer to one *G. parvula* strain (RCC4033) than to the other (RCC4034). In contrast to their monophyly in the nuclear phylogeny, they are scattered across each of the organellar phylogenies, with RCC4032 and RCC4033 both located (but not as sister groups) within α1 and α'1, RCC4034 as the most basal lineage in α2 and α'2, and RCC3370 placed in β and β'.

All members of the lightly calcified temperate clade EHB are nested within β and β', which contain other temperate strains
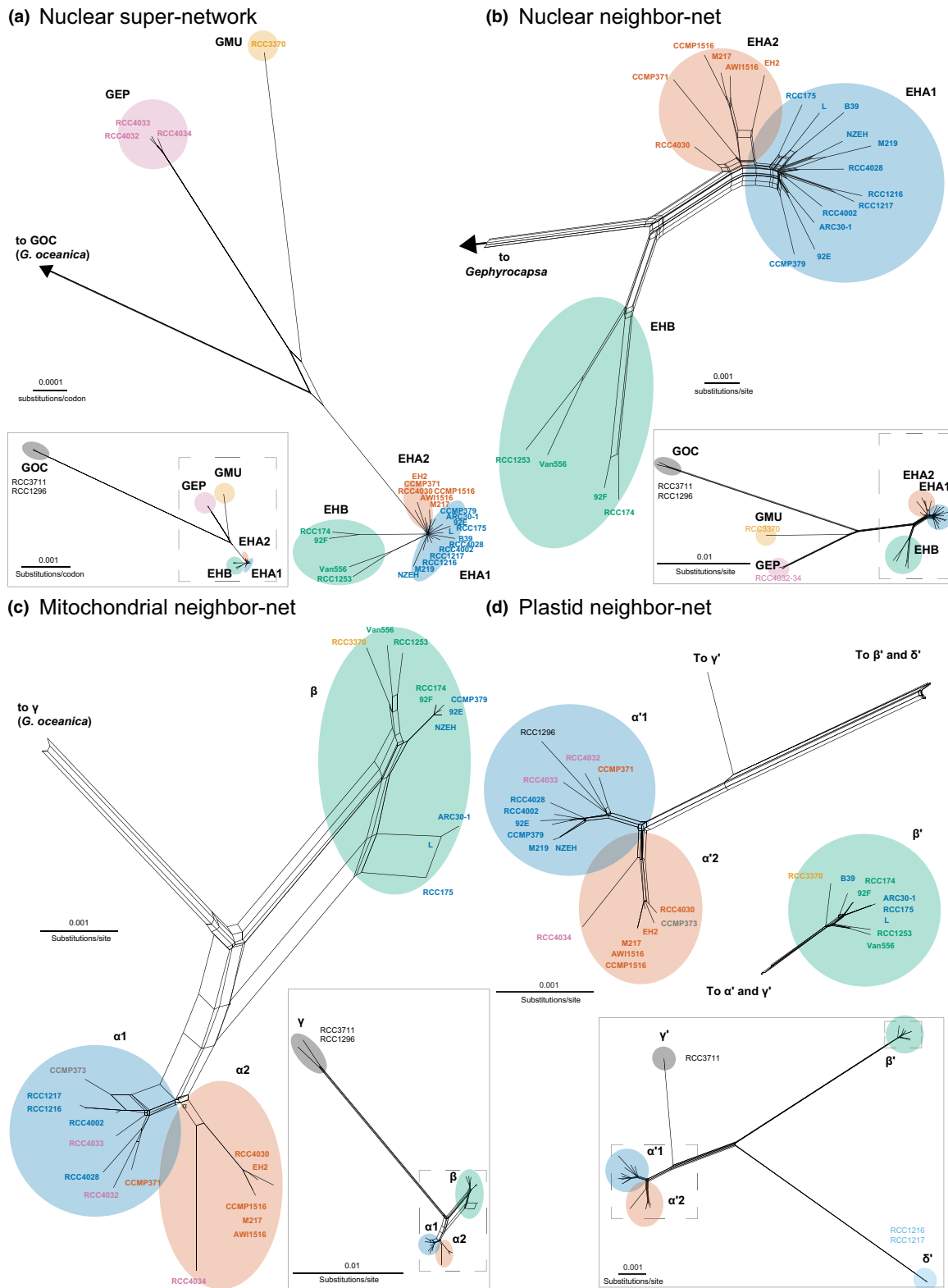
Fig. 3 Phylogenetic networks of nuclear and organellar genomes of *Emiliania–Gephyrocapsa* inferred in SPLITSTREE4. (a) Nuclear super-network inferred from maximum likelihood (ML) trees of 132 putative single-copy nuclear genes. (b–d) Neighbor-nets inferred from ML distances of the nuclear (b), mitochondrial (c), and plastid (d) genomes. The clades recognized in the respective ML nuclear, mitochondrial, and plastid trees (Figs 1b, 2a,b) are highlighted with color patches. Strain names are colored based on the major nuclear clades. Long branches were truncated to show more details of the phylogenetic networks. Full networks are provided in small boxes, with the enlarged areas indicated with dashed boxes.
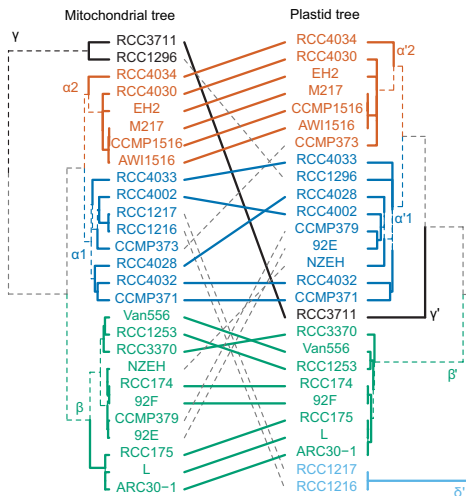
(a) Tanglegram between nuclear and mitochondrial tree
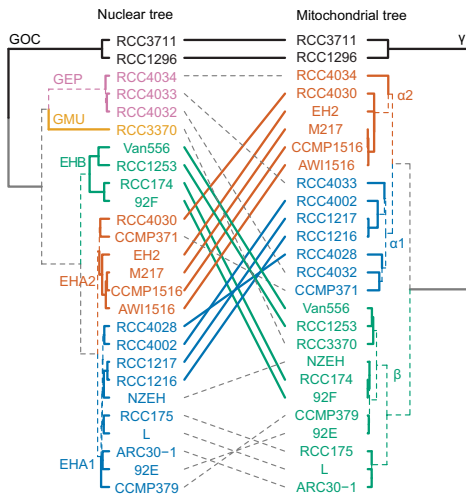
(b) Tanglegram between nuclear and plastid tree
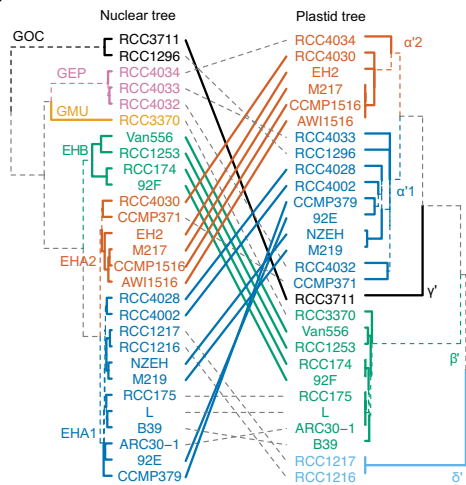
(c) Tanglegram between mitochondrial and plastid tree

Fig. 4 Tanglegrams visualizing the discordance between the maximum likelihood (ML) phylogenies of nuclear, mitochondrial, and plastid genomes of *Emiliania–Gephyrocapsa*. (a) Nuclear and mitochondrial trees. (b) Nuclear and plastid trees. (c) Mitochondrial and plastid trees. Only strains represented in both trees were retained in each tanglegram. The major clades are colored as in Figs 1(b) and 2(a,b). Tree branches are dotted when the subtended clades are absent in the other tree. Connecting lines of taxa are colored they are found in matching clades between the two trees; otherwise they are gray and dotted. Matching clades (nuclear/mitochondrial/plastid): GOC/γ/γ', EHB/β/β', EHA1/α1/ α'1, and EHA2/α2/α'2.

EHA2 member from North Atlantic, is most closely related to *G. ericsonii* and nested in clades α1 and α'1.

The members of EHA1 possess either an α1 or β mitogenome and an α'1, β', or δ' plastome. There is clear phylogenomic discordance, both cyto-nuclear and inter-organellar (Fig. 4c). For example, NZEH, CCMP379, and 92E have β mitogenomes but α'1 plastomes. Another peculiar case is CCMP373, with the mitogenome nested within α1 and plastome within α'2. A particularly notable lineage is formed by RCC1216 and its haploid strain RCC1217 derived from laboratory culture about 22 yr ago (von Dassow *et al.*, 2009). They are deeply nested within EHA1 and α1, but form their own distinct clade δ' in the plastome phylogeny (Fig. 2b), which is highly divergent from all the other plastomes and likely represents an early-diverging plastome lineage.

By comparing the distribution of individual strains across the major clades in each tree, nuclear–mitochondrial–plastid discordance can be identified at the level of major clades when one strain is placed in unmatched clades. With the nuclear clades as the reference, the matching nuclear/mitochondrial/plastid clades are GOC/γ/γ', EHB/β/β', EHA1/α1/α'1, and EHA2/α2/α'2 (coded with the same colors in Figs 1b, 2), with GMU, GEP and δ' having no matching clades in the other trees. Based on occurrence of a strain in matching or unmatched clades, it can be classified into one of the four types (Fig. 5): (1) concordance between major clades (occurring in matching clades across the three genomes); (2) cytonuclear discordance (nuclear clade unmatched with both organellar clades, which are matching); (3) inter-organellar discordance with the plastome clade unmatched with the other two; and (4) inter-organellar discordance with the mitogenome clade unmatched with the other two.

Among the 26 strains found in all three trees, 12 occur in matching clades (Fig. 5). They include M217, AWI1516, and CCMP1516, which share the same laboratory origin. It should also be pointed out that this type of concordance only takes into account clade matching but not relationships between and within these clades, which can differ greatly among the three trees. Such topological differences include, for example, the sister group relationship between α' and γ' in the plastome tree but EHA/α and EHB/β in the others, as well as the different positions of the subclade formed by L, ARC30-1, and RCC175 within β and β' (a basal position in β but a nested position in β'). Given that these conflicting topologies were not considered, the distribution of 14 out of 26 strains in unmatched clades (Fig. 5) merely depicts part

including GMU and certain EHA1 members. All South Pacific members of EHA2 have roughly the same phylogenetic relationships among them across all three genomes. In both organellar trees, one *G. parvula* strain, RCC4034, consistently forms the sister lineage to these EHA2 strains. However, CCMP371, the only
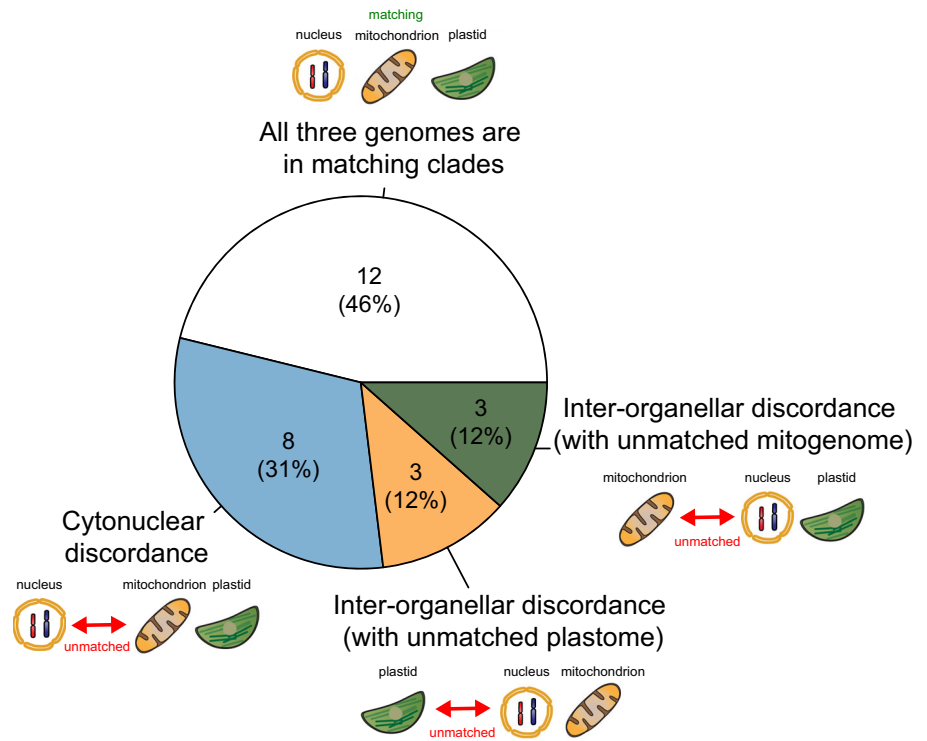
Fig. 5 Summary of phylogenomic concordance and discordance patterns of 26 *Emiliania–Gephyrocapsa* strains based on distribution across major nuclear, mitochondrial, and plastid clades. The major clades of nuclear genomes (GOC, GMEP, EHB, EHA1, and EHA2), mitogenomes (γ, β, α1, and α2), and plastomes (δ′, γ′, β′, α′1 and α′2) are defined in Figs 1 and 2, with the 'matching clades' of different genomes (GOC/γ/γ′, EHB/β/β′, EHA1/α1/α′1, and EHA2/α2/α′2) coded with the same color.

of the phylogenomic discordance among the three compartments.

Discussion

Nuclear–mitochondrial–plastid discordance

The nuclear genome phylogeny (Fig. 1) slightly differs from those of two earlier studies (Read *et al.*, 2013; Bendif *et al.*, 2019), but is consistent with recently published reference-based analysis of whole genomes (Filatov *et al.*, 2021). Both this (Fig. 3a; Table S1; Dataset S3) and previous studies (Bendif *et al.*, 2019; Filatov *et al.*, 2021) found little or only modest discordance between nuclear genes and weak signals of nuclear hybridization or introgression. Similarly, there is little intragenomic discordance within most organellar genomes, suggesting the phylogenetic trees of concatenated alignments can generally illustrate the whole-genome evolutionary history of organellar haplotypes.

In terms of intergenomic discordance across major clades (Fig. 5), cytonuclear discordance is found in all four GMEP and four EHUX strains (Notes S2). Cytonuclear discordance is the best studied type of intergenomic discordance, partially because most studies focused on plastid-less eukaryotes or only the nuclear and one organellar genome. Inter-organellar comparisons can also be challenging in plants due to frequent mitogenome rearrangements (Rose *et al.*, 2021). Here we illustrate the extent of cytoplasmic incongruence at the whole-genome level in algae, identifying strains with unmatched plastome (RCC1296, RCC1216, and RCC1217) or mitogenome (CCMP379, 92E,

and NZEH) clades. The extent of inter-organellar discordance in these microalgae has rarely been observed in other photosynthetic eukaryotes with phylogenomic analyses of all three genomes. For example, the angiosperm genus *Heuchera* is well known for its frequent hybridization and cytonuclear discordance (Soltis & Kuzoff, 1995), but it has largely congruent mitogenome and plastome phylogenies (Folk *et al.*, 2017). Extensive cytonuclear discordance and overall inter-organellar congruence are also seen in *Sphagnum* mosses (Meleshko *et al.*, 2021).

Evolutionary processes underlying intergenomic discordance

Rampant nuclear–mitochondrial–plastid discordance in *Emiliania–Gephyrocapsa* was likely caused by a combination of evolutionary processes, including recombination, introgression, and ILS. Organellar genomes were traditionally assumed to be nonrecombining, but intermolecular recombination has been documented for mitogenomes of animals, fungi, and plants (Barr *et al.*, 2005; White *et al.*, 2008; Fritsch *et al.*, 2014) and plastomes of plants (Wolfe & Randle, 2004; Sullivan *et al.*, 2017). Here we provide strong evidence for inter-haplotype mitogenome recombination in L, ARC30-1, and RCC175, which have a major β fragment and minor α fragment (Figs 3c, S3). This clear sign of past mitochondrial heteroplasmy suggests polymorphic haplotypes can co-exist in one cell. The absence of heteroplasmy in all 29 sampled strains indicates that heteroplasmy is uncommon or transient or that heteroplasmy in the original field-isolated cells has been lost due to organellar sorting at each cell division and long-term laboratory maintenance. Transient
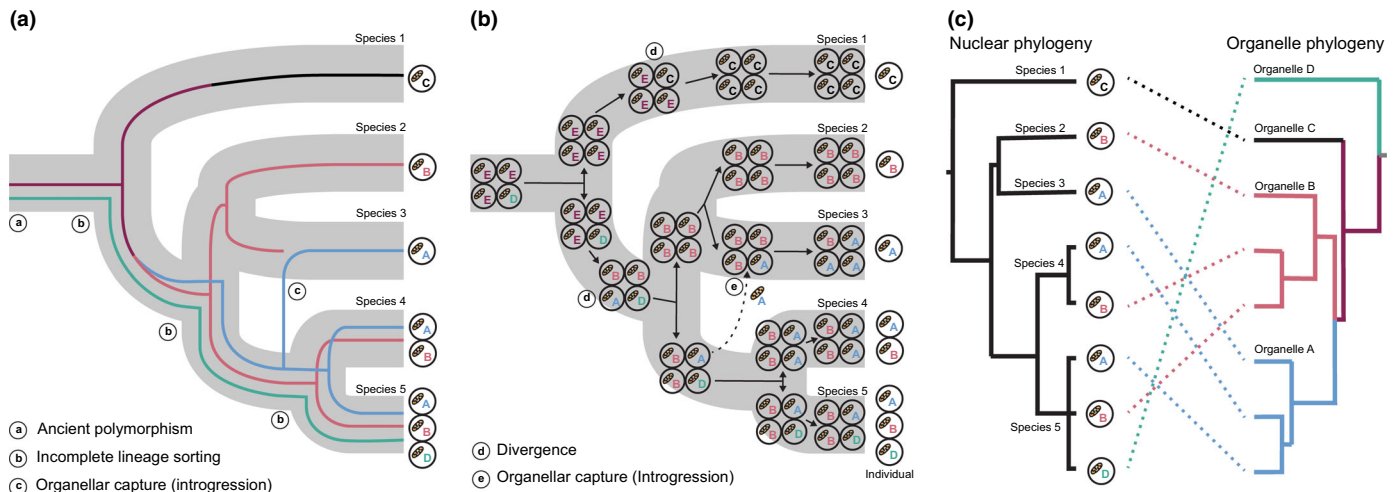
Fig. 6 Schematic illustration of intergenomic discordance as a result of introgression and incomplete lineage sorting (ILS) of organellar genomes. For simplicity, the symbol of mitochondrion is used to represent any of the two organelles. (a) Species and nuclear genome diversification history (gray shading) can differ from the organellar history (lines) due to ILS of ancient polymorphisms and/or introgression (organellar capture). (b) One possible scenario of organellar population genomics corresponding to (a). Individual genomes are sampled from present populations of the five hypothetical species. (c) The species/nuclear and organellar phylogenies inferred from the individuals sampled in (b). The diversification history of the type C organellar genomes is concordant with species phylogeny. The diversification history of the type A includes an introgression event between species 3 and the common ancestor of species 4 and 5 (a). Although different processes were involved (with or without introgression), the type A and B subtrees have similar topologies. The type D represents the relic of ancient polymorphic organellar genomes that has only been retained in species 5 after ILS.

mitogenome heteroplasmy is found in budding yeasts after syngamy, where mitochondrial fusion makes inter-haplotype recombination possible (Fritsch *et al.*, 2014). The variably shaped mitochondria of *Emiliania–Gephyrocapsa* probably have dynamic fusion and fission (Uwizeye *et al.*, 2021), whereas the plastid is generally a single compartment. This might explain the observation of only mitogenome but not plastome recombination in this study.

The vast majority of discordant patterns can be attributed to introgression and ILS (Fig. 6). It was previously suggested that introgressive hybridization might have caused incongruent topologies of individual gene trees in these microalgae (Bendif *et al.*, 2015, 2016). Here the recombination-based evidence of past heteroplasmy further indicates that hybridization can combine cytoplasmic components (organelles) from the two gametes. In addition to inter-haplotype recombination, another possible outcome is random loss of the haplotype from one gamete. In that case, if the nuclear genome is inherited only from the other gamete due to some unknown mechanism, the nuclear and this organellar genome would be unmatched (Fig. 6c), resulting in apparent replacement of the organellar genome by one captured from another lineage (e in Fig. 6b). Some possible examples of strains with organellar captures include the four GMEP strains with matching mitogenome and plastome clades. The original cytoplasm of L, ARC30-1, and RCC175 was probably α1/α′1, matching their EHA1 nuclear genomes, but it might have been fused with β/β′ organelles, completely replacing α′1 plastome and generating an α1-β recombined mitogenome. The ancestor of RCC1296 might have been fused with α1/α′1, but only the α′1 plastome was captured whereas the α1 mitogenome was lost. In contrast, 92E, CCMP379, and NZEH captured only the β

mitogenome but not the β′ plastome. Many organellar introgressions involve EHB and EHA1 clades, each being equally represented in Atlantic and Pacific Oceans, and the introgressions led to cells with EHA1 nuclear genomes capturing β/β′ organellar genomes, but not vice versa. This unidirectional introgression might be partially due to more EHA1 strains sampled than EHB, but it is possible that EHB nuclei only co-exist with β/β′ organelles or EHA1 nuclei are more receptive to divergent organellar genomes.

A combination of low nuclear intragenomic and strong mito-nuclear or plastid-nuclear discordance has also been observed in other eukaryotes (Folk *et al.*, 2017; Sarver *et al.*, 2021). The substantial effect of introgressive hybridization on mitochondrial but not nuclear genomes in some animals has been explained by mechanisms such as sex-linked biases, negative selection against nuclear introgression, and adaptive introgression of mitogenomes (Toews & Brelsford, 2012; Bonnet *et al.*, 2017). However, coccolithophores are haplodiplontic (Frada *et al.*, 2019), alternating between diploid and haploid cells that can reproduce asexually through mitosis, and they are apparently isogamous (Houdan *et al.*, 2004). Without egg-sperm or any sexual dimorphism, coccolithophores seem much less likely than animals or plants to have sex-linked biases. In addition, the average dN/dS of coding genes is below 0.1 for both organellar genomes, suggesting the organellar genes are generally under purifying selection. Unlike the nuclear phylogeny, there is no apparent associations between organellar phylogenies and geographical distribution or physical environments. These observations suggest that some other mechanisms are responsible for such disparity in the amount of introgression between nuclear and organellar genomes in coccolithophores.

Introgression (including organellar capture) can explain much of the observed nuclear–mitochondrial–plastid discordance, but we cannot exclude potential contributions from ILS (Fig. 6). The presence of δ′ plastomes in RCC1216/RCC1217 clearly demonstrates divergent organellar haplotypes can be found within *E. huxleyi*, a species with extremely large population size (Filatov, 2019; Filatov *et al.*, 2021). While δ′ in RCC1216 could be captured from an unsampled donor lineage, an alternative explanation is the existence of ancient polymorphic haplotypes in *E. huxleyi* populations where the rare δ′ was inherited by RCC1216. The cytonuclear discordance in GEP can be explained by three recent events of organellar capture in GEP strains, where RCC4034 captured α2/α′2 organelles and RCC4032 and RCC4033, sisters in the nuclear phylogeny, each captured different α1/α′1 haplotypes. However, an alternative scenario is that the ancestral GEP population had sampled different α/α′ haplotypes, already diverged into α1 and α2 mitogenomes and α′1 and α′2 plastomes, which were later sorted into the three GEP strains. With the separation of GEP and EHUX α/α′ haplotypes since the ancient introgression, this scenario might better explain the more basal positions of GEP strains in α1, α2, α′1, and α′2. In addition, ILS can explain why RCC4032 and RCC4033 do not cluster together in the organellar phylogenies despite their well-supported sister-group relationship in the nuclear tree. One aforementioned unknown in coccolithophore hybridization is how it impacts organellar genomes while barely affecting nuclear genomes. The patterns of ILS, however, can be independent for each genome and cause strong cytonuclear or inter-organellar discordance despite low nuclear intragenomic discordance. The random segregation of organellar genomes during cell divisions also has a potential role in evolution by creating segregational drift, as shown for multicopy plasmids (Ilhan *et al.*, 2019).

Implications for coccolithophore biology

The phylogenomic analyses of the three genomes provide important insights into coccolithophore biology and life cycles. Mitogenome inter-haplotype recombination points to biparental inheritance and fusion of mitochondria following plasmogamy in coccolithophores. Isogamy, fusion between equally sized gametes, and biparental inheritance likely contributed to the rampant discordance in coccolithophores. High mito-nuclear discordance has also been reported from isogamous yeasts (De Chiara *et al.*, 2020) with known mitochondrial biparental inheritance, fusion, and recombination (Fritsch *et al.*, 2014). Biparental inheritance of cytoplasmic components coupled with possible recombination or differential loss of haplotypes would mean that the cytoplasm is not inherited as a single unit and that introgression can greatly entangle organellar genome phylogenies. This is in contrast with eukaryotes having canonical uniparental inheritance with only occasional leakage from the other parent. Nonetheless, it should be noted that isogamy is not necessarily associated with biparental organellar inheritance. *Chlamydomonas reinhardtii* has uniparental inheritance of both organelles (Miyamura, 2010), whereas isogamous multicellular brown algae (e.g. *Ectocarpus*

siliculosus) have biparentally inherited plastids but maternally inherited mitochondria (Motomura *et al.*, 2010). Further research should be undertaken to investigate if coccolithophore plastids are biparentally inherited like their mitochondria.

In addition to organelle transmission, the mode of nuclear genome inheritance can be an important factor in intergenomic discordance. Given the low nuclear intragenomic discordance despite introgressive hybridization, we hypothesize that coccolithophore gamete mating leads to plasmogamy, but not necessarily karyogamy and retention of biparental nuclear genomes. It is similar to androgenesis (Hedtke & Hillis, 2011; Schwander & Oldroyd, 2016) and gynogenesis (sperm-cell-dependent parthenogenesis) (Schlupp, 2005; Jacquier *et al.*, 2020) in some animals and plants. Because of the lack of heteromorphic gametes, we term this kind of false syngamy, which results in cytoplasmic but not nuclear hybridization, 'unigenesis'. The significance of unigenesis is unclear, but it is possible that gamete fusion between strains with large nuclear genome size differences (Read *et al.*, 2013) might fail to combine the two haploid genomes, eventually eliminating one of them. In contrast to normal syngamy, unigenesis would enable organellar but not nuclear introgression between distant lineages.

Besides enormous population size, isogamy, and unigenesis, the unicellular haplodiplontic life cycle also sets coccolithophores apart from animals and plants. Animals are typically diplontic, whereas most plants alternate between the diploid sporophyte and haploid gametophyte with one dependent on the other. In contrast, both haploid and diploid stages of coccolithophores are free-living unicellular organisms that can reproduce asexually (Frada *et al.*, 2019). It was shown that haplodiploid animals (haploid males and diploid females) exhibit reduced nuclear introgression relative to mitochondrial introgression, leading to mito-nuclear incongruence (Lohse & Ross, 2015; Patten *et al.*, 2015). The haplodiplontic life cycle of coccolithophores might also contribute to intergenomic discordance. One possibility is that haploid coccolithophores are fully independent, which would allow for more efficient selection against any deleterious genes or gene combinations during the haploid phase (Otto & Gerstein, 2008) and thus reduce the chances of nuclear introgression.

With their unique calcifying physiology and global ecological importance, coccolithophores have been widely used for studies on cell biology (e.g. Taylor *et al.*, 2017) and viral infection of algae (e.g. Ku *et al.*, 2020). High-throughput sequencing has greatly advanced our knowledge on the genomics of these microalgae, leading to discoveries about their pangenome variation (Read *et al.*, 2013), speciation mode (Bendif *et al.*, 2019; Filatov *et al.*, 2021), and evolution of extremely large populations (Filatov, 2019). Here we further establish coccolithophores as an underexplored unicellular model for studying the relationships between nuclear, mitochondrial, and plastid phylogenomic histories. The extensive phylogenetic discordance among these compartmentalized genomes hints at intriguing modes of reproduction and life cycling in coccolithophores, of which future research would also shed light on the biology of other microalgae and unicellular eukaryotes.

Author contributions

T-TK and CK planned and designed the research. T-TK and T-HW performed laboratory work, collected sequences and carried out the analyses with the help of CK. T-TK and CK wrote the manuscript with support from T-HW.

ORCID

Tzu-Tong Kao https://orcid.org/0000-0003-1372-7324
Chuan Ku https://orcid.org/0000-0001-6414-4423
Tzu-Haw Wang https://orcid.org/0000-0001-7358-8966

Data availability

The datasets supporting the conclusions of this article are included within the article and its Supporting Information. The organellar genome assemblies have been deposited in the NCBI under the accession numbers listed in Table S1.

References

Agarkova IV, Dunigan DD, Van Etten JL. 2006. Virion-associated restriction endonucleases of chloroviruses. *Journal of Virology* 80: 8114–8123.

Allen JF. 2015. Why chloroplasts and mitochondria retain their own genomes and genetic systems: colocation for redox regulation of gene expression. *Proceedings of the National Academy of Sciences, USA* 112: 10231–10238.

Balch WM. 2018. The ecology, biogeochemistry, and optical properties of coccolithophores. *Annual Review of Marine Science* 10: 71–98.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD et al. 2012. SPADES: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19: 455–477.

Barr CM, Neiman M, Taylor DR. 2005. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytologist* 168: 39–50.

Bendif EM, Nevado B, Wong ELY, Hagino K, Probert I, Young JR, Rickaby REM, Filatov DA. 2019. Repeated species radiations in the recent evolution of the key marine phytoplankton lineage *Gephyrocapsa*. *Nature Communications* 10: 4234.

Bendif EM, Probert I, Díaz-Rosas F, Thomas D, van den Engh G, Young JR, von Dassow P. 2016. Recent reticulate evolution in the ecologically dominant lineage of coccolithophores. *Frontiers in Microbiology* 7: 784.

Bendif EM, Probert I, Young JR, von Dassow P. 2015. Morphological and phylogenetic characterization of new *Gephyrocapsa* isolates suggests introgressive hybridization in the *Emiliania/Gephyrocapsa* complex (Haptophyta). *Protist* 166: 323–336.

Birky CW. 2008. Uniparental inheritance of organelle genes. *Current Biology* 18: R692–R695.

Blischak PD, Chifman J, Wolfe AD, Kubatko LS. 2018. HyDe: a python package for genome-scale hybridization detection. *Systematic Biology* 67: 821–829.

Bonnet T, Leblois R, Rousset F, Crochet P-A. 2017. A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. *Evolution* 71: 2140–2158.

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics* 3: lqaa108.

Chifman J, Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30: 3317–3324.

Chifman J, Kubatko L. 2015. Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. *Journal of Theoretical Biology* 374: 35–47.

Coombe L, Zhang J, Vandervalk BP, Chu J, Jackman SD, Birol I, Warren RL. 2018. ARKS: chromosome-scale scaffolding of human genome drafts with linked read kmers. *BMC Bioinformatics* 19: 234.

von Dassow P, John U, Ogata H, Probert I, Bendif EM, Kegel JU, Audic S, Wincker P, Da Silva C, Claverie J-M et al. 2015. Life-cycle modification in open oceans accounts for genome variability in a cosmopolitan phytoplankton. *The ISME Journal* 9: 1365–1377.

von Dassow P, Ogata H, Probert I, Wincker P, Da Silva C, Audic S, Claverie J-M, de Vargas C. 2009. Transcriptome analysis of functional differentiation between haploid and diploid cells of *Emiliania huxleyi*, a globally significant photosynthetic calcifying cell. *Genome Biology* 10: R114.

De Chiara M, Friedrich A, Barré B, Breitenbach M, Schacherer J, Liti G. 2020. Discordant evolution of mitochondrial and nuclear yeast genomes at population level. *BMC Biology* 18: 49.

Dierckxsens N, Mardulyn P, Smits G. 2016. NOVOPLASTY: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Research* 45: e18.

Dong W, Liu J, Yu J, Wang L, Zhou S. 2012. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE* 7: e35071.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.

Emms DM, Kelly S. 2019. ORTHOFINDER: phylogenetic orthology inference for comparative genomics. *Genome Biology* 20: 238.

Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield O, Taylor FJR. 2004. The evolution of modern eukaryotic phytoplankton. *Science* 305: 354–360.

Filatov DA. 2019. Extreme Lewontin's paradox in ubiquitous marine phytoplankton species. *Molecular Biology and Evolution* 36: 4–14.

Filatov DA, Bendif EM, Archontikis OA, Hagino K, Rickaby REM. 2021. The mode of speciation during a recent radiation in open-ocean phytoplankton. *Current Biology* 31: 5439–5449.

Folk RA, Mandel JR, Freudenstein JV. 2017. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. *Systematic Biology* 66: 320–337.

Frada MJ, Bendif EM, Keuter S, Probert I. 2019. The private life of coccolithophores. *Perspectives in Phycology* 6: 11–30.

Fritsch ES, Chabbert CD, Klaus B, Steinmetz LM. 2014. A genome-wide map of mitochondrial DNA recombination in yeast. *Genetics* 198: 755–771.

Galili T. 2015. DENDEXTEND: an R package for visualizing, adjusting, and comparing trees of hierarchical clustering. *Bioinformatics* 31: 3718–3720.

Galtier N, Nabholz B, Glémin S, Hurst GDD. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology* 18: 4541–4550.

Green BR. 2011. Chloroplast genomes of photosynthetic eukaryotes. *The Plant Journal* 66: 34–44.

Greiner S, Sobanski J, Bock R. 2015. Why are most organelle genomes transmitted maternally? *BioEssays* 37: 80–94.

Gremme G, Brendel V, Sparks ME, Kurtz S. 2005. Engineering a software tool for gene structure prediction in higher organisms. *Information and Software Technology* 47: 965–978.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* 59: 307–321.

Hedtke SM, Hillis DM. 2011. The potential role of androgenesis in cytoplasmic–nuclear phylogenetic discordance. *Systematic Biology* 60: 87–96.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* 35: 518–522.

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and Augustus. *Bioinformatics* 32: 767–769.

Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-genome annotation with BRAKER. *Methods in Molecular Biology* 1962: 65–95.

Houdan A, Billard C, Marie D, Not F, Sáez AG, Young JR, Probert I. 2004. Holococcolithophore-heterococcolithophore (Haptophyta) life cycles: flow cytometric analysis of relative ploidy levels. *Systematics and Biodiversity* 1: 453–465.

Hulatt CJ, Wijffels RH, Viswanath K, Posewitz MC. 2020. The complete mitogenome and plastome of the haptophyte *Pavlova lutheri* NIVA-4/92. *Mitochondrial DNA Part B, Resources* 5: 2748–2749.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23: 254–267.

Ilhan J, Kupczok A, Woehle C, Wein T, Hülter NF, Rosenstiel P, Landan G, Mizrahi I, Dagan T. 2019. Segregational drift and the interplay between plasmid copy number and evolvability. *Molecular Biology and Evolution* 36: 472–486.

Jacquier NMA, Gilles LM, Pyott DE, Martinant J-P, Rogowsky PM, Widiez T. 2020. Puzzling out plant reproduction by haploid induction for innovations in plant breeding. *Nature Plants* 6: 610–619.

Katoh K, Standley DM. 2013. Mafft multiple sequence alignment software v.7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.

Keeling PJ. 2013. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annual Review of Plant Biology* 64: 583–607.

Keller MD, Selvin RC, Claus W, Guillard RRL. 1987. Media for the culture of oceanic ultraphytoplankton. *Journal of Phycology* 23: 633–638.

Kent WJ. 2002. Blat: the Blast-like alignment tool. *Genome Research* 12: 656–664.

Krasovec M, Rickaby REM, Filatov DA. 2020. Evolution of mutation rate in astronomically large phytoplankton populations. *Genome Biology and Evolution* 12: 1051–1059.

Ku C, Nelson-Sathi S, Roettger M, Sousa FL, Lockhart PJ, Bryant D, Hazkani-Covo E, McInerney JO, Landan G, Martin WF. 2015. Endosymbiotic origin and differential loss of eukaryotic genes. *Nature* 524: 427–432.

Ku C, Sheyn U, Sebé-Pedrós A, Ben-Dor S, Schatz D, Tanay A, Rosenwasser S, Vardi A. 2020. A single-cell view on alga-virus interactions reveals sequential transcriptional programs and infection states. *Science Advances* 6: eaba4137.

Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* 29: 1695–1701.

Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30: 3276–3278.

Lee JM, Song HJ, Park SI, Lee YM, Jeong SY, Cho TO, Kim JH, Choi H-G, Choi CG, Nelson WA *et al.* 2018. Mitochondrial and plastid genomes from coralline red algae provide insights into the incongruent evolutionary histories of organelles. *Genome Biology and Evolution* 10: 2961–2972.

Lee-Yaw JA, Grassa CJ, Joly S, Andrew RL, Rieseberg LH. 2019. An evaluation of alternative explanations for widespread cytonuclear discordance in annual sunflowers (*Helianthus*). *New Phytologist* 221: 515–526.

Lehtonen J, Kokko H, Parker GA. 2016. What do isogamous organisms teach us about sex and the two sexes? *Philosophical Transactions of the Royal Society B: Biological Sciences* 371: 20150532.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with Bwa-Mem. *arXiv preprint arXiv:1303.3997*.

Liu H, Aris-Brosou S, Probert I, de Vargas C. 2010. A time line of the environmental genetics of the haptophytes. *Molecular Biology and Evolution* 27: 161–176.

Lockwood JD, Aleksić JM, Zou J, Wang J, Liu J, Renner SS. 2013. A new phylogeny for the genus *Picea* from plastid, mitochondrial, and nuclear sequences. *Molecular Phylogenetics and Evolution* 69: 717–727.

Lohse K, Ross L. 2015. What haplodiploids can teach us about hybridization and speciation. *Molecular Ecology* 24: 5075–5077.

Martin DP, Varsani A, Roumagnac P, Botha G, Maslamoney S, Schwab T, Kelz Z, Kumar V, Murrell B. 2021. Rdp5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evolution* 7: veaa087.

Meleshko O, Martin MD, Korneliussen TS, Schröck C, Lamkowski P, Schmutz J, Healey A, Piatkowski BT, Shaw AJ, Weston DJ *et al.* 2021. Extensive genome-wide phylogenetic discordance is due to incomplete lineage sorting and not ongoing introgression in a rapidly radiated bryophyte genus. *Molecular Biology and Evolution* 38: 2750–2766.

Minh BQ, Hahn MW, Lanfear R. 2020a. New methods to calculate concordance factors for phylogenomic datasets. *Molecular Biology and Evolution* 37: 2727–2733.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020b. IQ-Tree 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37: 1530–1534.

Miyamura S. 2010. Cytoplasmic inheritance in green algae: patterns, mechanisms and relation to sex type. *Journal of Plant Research* 123: 171–184.

Motomura T, Nagasato C, Kimura K. 2010. Cytoplasmic inheritance of organelles in brown algae. *Journal of Plant Research* 123: 185–192.

Nishimura Y, Kamikawa R, Hashimoto T, Inagaki Y. 2014. An intronic open reading frame was released from one of group II introns in the mitochondrial genome of the haptophyte *Chrysochromulina* sp. NIES-1333. *Mobile Genetic Elements* 4: e29384.

Otto SP, Gerstein AC. 2008. The evolution of haploidy and diploidy. *Current Biology* 18: R1121–R1124.

Paradis E, Schliep K. 2019. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526–528.

Patten MM, Carioscia SA, Linnen CR. 2015. Biased introgression of mitochondrial and nuclear genes: a comparison of diploid and haplodiploid systems. *Molecular Ecology* 24: 5200–5210.

Phillips MJ, McLenachan PA, Down C, Gibb GC, Penny D. 2006. Combined mitochondrial and nuclear DNA sequences resolve the interrelations of the major Australasian marsupial radiations. *Systematic Biology* 55: 122–137.

Probert I. 2019. K/2 Ian/K-ET V.2. doi: 10.17504/protocols.io.7bxhipn.

R Core Team. 2021. *R: a language and environment for statistical computing.* [WWW document] URL https://www.r-project.org/ [accessed 6 October 2021].

Raffi I, Backman J, Fornaciari E, Pälike H, Rio D, Lourens L, Hilgen F. 2006. A review of calcareous nannofossil astrobiochronology encompassing the past 25 million years☆. *Quaternary Science Reviews* 25: 3113–3137.

Read BA, Kegel J, Klute MJ, Kuo A, Lefebvre SC, Maumus F, Mayer C, Miller J, Monier A, Salamov A *et al.* 2013. Pan genome of the phytoplankton *Emiliania* underpins its global distribution. *Nature* 499: 209–213.

Revell LJ. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3: 217–223.

Rose JP, Toledo CAP, Lemmon EM, Lemmon AR, Sytsma KJ. 2021. Out of sight, out of mind: widespread nuclear and plastid-nuclear discordance in the flowering plant genus *Polemonium* (Polemoniaceae) suggests widespread historical gene flow despite limited nuclear signal. *Systematic Biology* 70: 162–180.

Rousseaux C, Gregg W. 2013. Interannual variation in phytoplankton primary production at a global scale. *Remote Sensing* 6: 1–19.

Sánchez-Puerta MV, Bachvaroff TR, Delwiche CF. 2004. The complete mitochondrial genome sequence of the haptophyte *Emiliania huxleyi* and its relation to heterokonts. *DNA Research* 11: 1–10.

Sánchez-Puerta MV, Bachvaroff TR, Delwiche CF. 2005. The complete plastid genome sequence of the haptophyte *Emiliania huxleyi*: a comparison to other plastid genomes. *DNA Research* 12: 151–156.

Sarver BAJ, Herrera ND, Sneddon D, Hunter SS, Settles ML, Kronenberg Z, Demboski JR, Good JM, Sullivan J. 2021. Diversification, introgression, and rampant cytonuclear discordance in rocky mountains chipmunks (Sciuridae: Tamias). *Systematic Biology* 70: 908–921.

Schlupp I. 2005. The evolutionary ecology of gynogenesis. *Annual Review of Ecology, Evolution, and Systematics* 36: 399–417.

Schwander T, Oldroyd BP. 2016. Androgenesis: where males hijack eggs to clone themselves. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371: 20150534.

Smith DR, Keeling PJ. 2012. Twenty-fold difference in evolutionary rates between the mitochondrial and plastid genomes of species with secondary red plastids. *Journal of Eukaryotic Microbiology* 59: 181–184.

Smith DR, Keeling PJ. 2015. Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. *Proceedings of the National Academy of Sciences, USA* 112: 10177–10184.

Solieri L. 2010. Mitochondrial inheritance in budding yeasts: towards an integrated understanding. *Trends in Microbiology* 18: 521–530.

Soltis DE, Kuzoff RK. 1995. Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). *Evolution* 49: 727–742.

Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24: 637–644.

Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 7: 62.

Sullivan AR, Schiffthaler B, Thompson SL, Street NR, Wang X-R. 2017. Interspecific plastome recombination reflects ancient reticulate evolution in *Picea* (Pinaceae). *Molecular Biology and Evolution* 34: 1689–1701.

Taylor AR, Brownlee C, Wheeler G. 2017. Coccolithophore cell biology: chalking up progress. *Annual Review of Marine Science* 9: 283–310.

Thierstein HR, Geitzenauer KR, Molfino B, Shackleton NJ. 1977. Global synchroneity of late Quaternary coccolith datum levels validation by oxygen isotopes. *Geology* 5: 400.

Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* 14: 178–192.

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* 45: W6–W11.

Toews DPL, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology* 21: 3907–3930.

Uwizeye C, Decelle J, Jouneau P-H, Flori S, Gallet B, Keck J-B, Bo DD, Moriscot C, Seydoux C, Chevalier F et al. 2021. Morphological bases of phytoplankton energy management and physiological responses unveiled by 3D subcellular imaging. *Nature Communications* 12: 1049.

Vargas OM, Ortiz EM, Simpson BB. 2017. Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). *New Phytologist* 214: 1736–1750.

White DJ, Wolff JN, Pierson M, Gemmell NJ. 2008. Revealing the hidden complexities of mtDNA inheritance. *Molecular Ecology* 17: 4925–4942.

Wolfe AD, Randle CP. 2004. Recombination, heteroplasmy, haplotype polymorphism, and paralogy in plastid genes: implications for plant molecular systematics. *Systematic Botany* 29: 1011–1020.

Worden AZ, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, Keeling PJ. 2015. Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science* 347: 1257594.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24: 1586–1591.

Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution* 15: 496–503.

Zardoya R. 2020. Recent advances in understanding mitochondrial genome diversity. *F1000Research* 9: 270.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19: 153.

Zhang D, Tang L, Cheng Y, Hao Y, Xiong Y, Song G, Qu Y, Rheindt FE, Alström P, Jia C et al. 2019. "Ghost introgression" as a cause of deep mitochondrial divergence in a bird species complex. *Molecular Biology and Evolution* 36: 2375–2386.

Zhang X, Gamarra J, Castro S, Carrasco E, Hernandez A, Mock T, Hadaegh AR, Read BA. 2016. Characterization of the small RNA transcriptome of the marine coccolithophorid, *Emiliania huxleyi*. *PLoS ONE* 11: e0154279.

Zhu A, Guo W, Gupta S, Fan W, Mower JP. 2016. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytologist* 209: 1747–1756.

Zimorski V, Ku C, Martin WF, Gould SB. 2014. Endosymbiotic theory for organelle origins. *Current Opinion in Microbiology* 22: 38–48.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Dataset S1 Maps of newly assembled organellar genomes (25 mitochondrial and 27 plastid).

Dataset S2 Alignments of nuclear (CDS), mitochondrial (whole genome and CDS), and plastid (whole genome and CDS) sequences.

Dataset S3 Phylogenomic trees and networks inferred using different datasets and methods.

Fig. S1 Mitogenome maps of CCMP373 and RCC1296.

Fig. S2 Plastome maps of CCMP373 and RCC1216.

Fig. S3 The recombination origin of the mitogenomes in L, ARC30-1, and RCC175.

Notes S1 Organellar genome organization.

Notes S2 Cytonuclear discordance at the level of major clades.

Table S1 List of coccolithophore strains included in this study.

Table S2 The dN/dS ratios of the nuclear, mitochondrial, and plastid coding genes analyzed.

Table S3 HyDe hybridization tests for four clades delineated in the nuclear phylogeny (GEP, GMU, EHA, and EHB) of *Emiliania–Gephyrocapsa* with GOC set as the outgroup.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.