



HAL
open science

Reconstructing Dynamic Evolutionary Events in Diatom Nuclear and Organelle Genomes

Richard Dorrell, Fuhai Liu, Chris Bowler

► **To cite this version:**

Richard Dorrell, Fuhai Liu, Chris Bowler. Reconstructing Dynamic Evolutionary Events in Diatom Nuclear and Organelle Genomes. Angela Falciatore; Thomas Mock. The Molecular Life of Diatoms, Springer International Publishing, pp.147-177, 2022, 978-3-030-92498-0. 10.1007/978-3-030-92499-7_6. hal-03859168

HAL Id: hal-03859168

<https://cnrs.hal.science/hal-03859168>

Submitted on 17 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Chapter 6: Reconstructing dynamic evolutionary events in diatom nuclear and organelle genomes

Richard G. Dorrell^{1*}, Fuhai Liu^{1,2,3}, Chris Bowler¹

¹Institut de Biologie de l'Ecole Normale Supérieure (IBENS), PSL Research Université, Paris 75005

²Centre de Recherches Interdisciplinaires, Paris 75014

³Tsinghua University Berkeley Campus, Shenzhen China

Correspondence: dorrell@bio.ens.psl.eu

Abstract

The diatoms evolved within the stramenopiles, an ecologically important and diverse assemblage of eukaryotes that includes both photosynthetic macrophytes and microalgae, as well as non-photosynthetic heterotrophs and parasites. The evolutionary history of the stramenopiles, which stretches back to the Palaeozoic, has been marked by the acquisition of chloroplasts in a recent common ancestor of their photosynthetic members, the ochrophytes; and progressive gains of genes in the nuclear genome by horizontal and endosymbiotic gene transfer. Here, we place diatoms in their actual evolutionary context within the stramenopiles; identify gene transfers that have shaped the coding content of the diatom nucleus; and profile sources of differences in chloroplast and mitochondrial genome content between different stramenopiles including diatoms. We underline the importance of considering diatoms as evolutionary mosaics, supported by genes of bacterial, red, green and other eukaryotic algal origin as illustrated by multiple phylogenomic studies realised over the last two decades; and the relatively limited changes to organelle genome content in diatoms compared to other stramenopile lineages. However, we identify a previously undocumented transfer of a novel open reading frame of the chloroplasts of green algae into the ochrophytes, underlining the importance of changes in organelle and nuclear gene content, in defining the current biology of diatoms.

Keywords: horizontal gene transfer; endosymbiosis; shopping bag model; cryptomonads; haptophytes; dinoflagellates

Table of Contents

1. The evolutionary context of diatoms within the stramenopiles	4
Diversity of photosynthetic and non-photosynthetic stramenopile groups	4
Fig. 1. Evolutionary positions of diatoms within the eukaryotes and the stramenopiles	4
Phylogenetic arrangement and endosymbiotic histories of the stramenopiles	5
2. Phylogenomic insights into diatom evolution	6
Fig. 2. Schematic diagram of the diatom chloroplast	7
Table 1. Sizes and sources of gene transfers into and out of diatom nuclear genomes	8-9
Methods and problems when exploring origins of diatom nuclear genes	9
Prokaryotic signals	11
Red algal-related signals	12
Green algal-derived signals	13
Have diatoms donated genes to other organisms?	14
3. Evolution of the diatom organelle genomes	14
Diatom mitochondria	14
Fig. 3. Organisation of the diatom mitogenome	15
Diatom chloroplasts	16
Fig. 4. Plastid genomes of diatoms and their closest relatives	16
Fig. 5. Genes with sporadic distributions across diatom chloroplast genomes	17
Diatom leucoplasts	18
Why are certain genes lost from diatom chloroplast genomes ?	18
Fig. 6. Dynamic of evolutionary rates in diatom chloroplast genomes	19
Loss versus transfer of diatom organelle genes	20
Gains and transfers of novel diatom ORFs	20
Fig. 7. Evolution of diatom chloroplast <i>ycf66</i>	20
4. Concluding remarks	21
5. References	21-29

1. The evolutionary context of diatoms within the stramenopiles

Diversity of photosynthetic and non-photosynthetic stramenopile groups

Diatoms belong to the stramenopile phylum, which sits within the “SAR clade” (Stramenopiles, Alveolates and Rhizaria) supergroup of the tree of life (2, 6, 7) (Fig. 1). Besides stramenopiles, the SAR clade contains the alveolates (including dinoflagellates, which are the principal photosynthetic component of corals, and apicomplexans such as the malaria parasite *Plasmodium*), and rhizarians (including the important fossil markers,

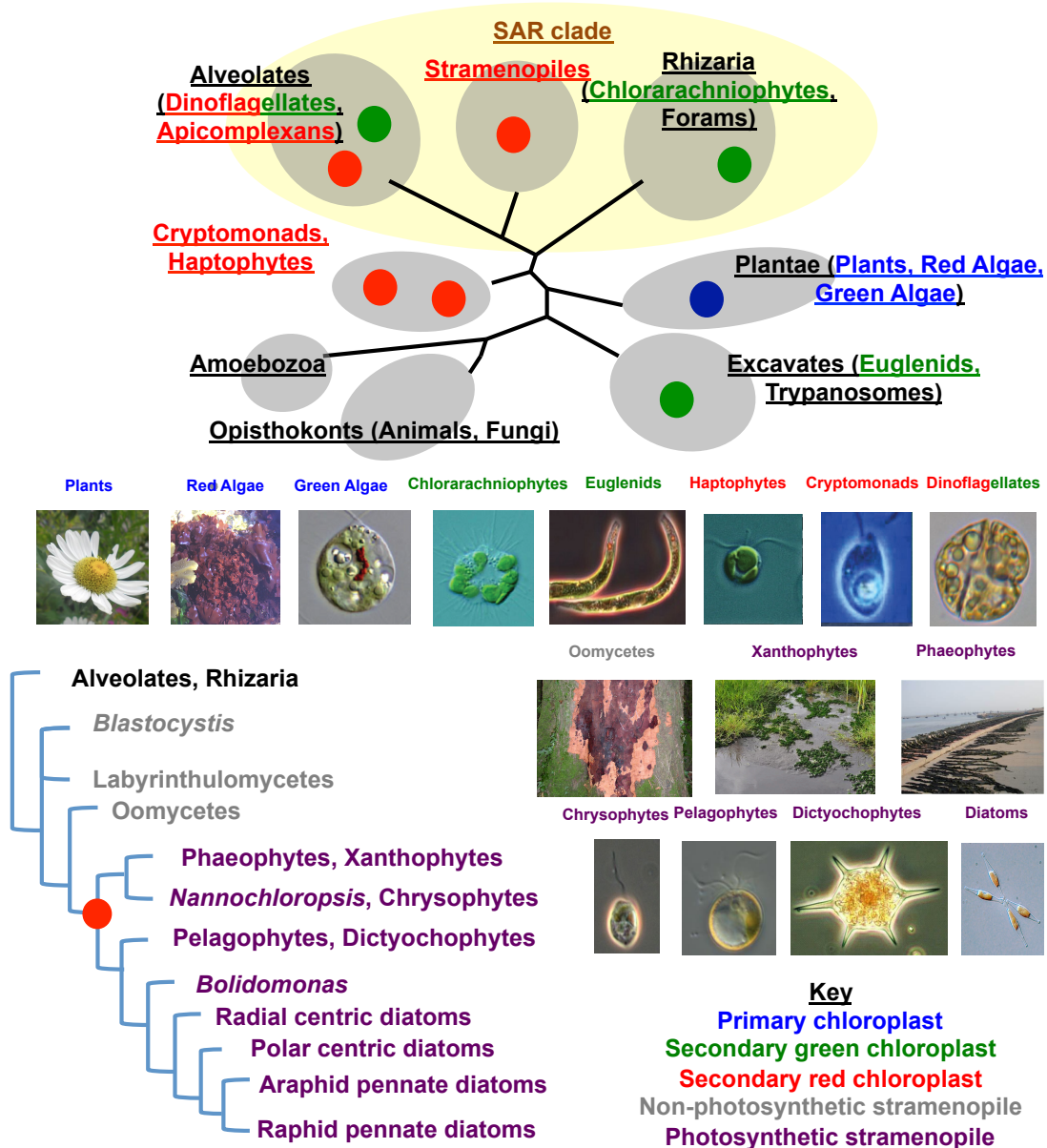


Fig. 1. Evolutionary positions of diatoms within the eukaryotes and the stramenopiles. Top : a schematic tree of eukaryotic diversity, adapted from (5) ; bottom, a closeup of stramenopile lineages, adapted from (2), showing the global and local evolutionary context of diatoms. Cell images are reprinted from Encyclopedia of Life (www.eol.org), Prof. Connie Lovejoy (Université Laval), and Dr. Zhanru Shao (Institute of Oceanology, Chinese Academy of Sciences), with permissions.

forams). The group is distantly related both to plants and to animals, sharing a last common ancestor with each no more recently than one billion years ago (16). The stramenopiles themselves are evolutionarily ancient, radiating at least six hundred million years before the present (17), placing them on a common level of antiquity to plants and animals (5).

The stramenopiles are an ecologically diverse, and environmentally important group of eukaryotes. They are characterised, in particular, by their flagellar organisation, which typically includes one short flagellum, and one long flagellum covered in tubular « hair »-like structures; with the uneven length of their flagella giving rise to an alternate name of « heterokonts » (7, 18). Beyond this, stramenopiles contain few conserved features, although their monophyly is well supported by molecular data (19, 20).

Alongside the diatoms, the stramenopiles contain numerous other photosynthetic taxa, referred to collectively as “ochrophytes” (Fig. 1). These include, but are not limited to: giant macrophytic kelps (found within the phaeophytes), which can grow to nearly one hundred meters in length (21), and unicellular and colonial algae (e.g., within the chrysophytes, or “golden algae”, dictyochophytes, or “silicoflagellates”, and pelagophytes), which can have cell diameters of only a couple of micrometers. Several other stramenopile groups (particularly phaeophytes, dictyochophytes and chrysophytes) are important contributors to marine and freshwater photosynthesis (5) Some of the ochrophyte groups are silicifying, like diatoms (e.g., silicoflagellates within the dictyochophytes; synurophytes within the chrysophytes), while others are not (2, 6, 22). Some of these photosynthetic groups are also known to engage in mixotrophic strategies, either through the phagocytotic consumption of bacteria (e.g., in chrysophytes and dictyochophytes (23, 24)), or through the osmotrophic uptake of extracellular organic nutrients (e.g., in diatoms and pelagophytes) (25), while in others (e.g., kelps) neither strategy has been reported (2).

Moreover, there is a substantial non-photosynthetic component to stramenopile diversity (Fig. 1). These include some lineages that resolve within otherwise photosynthetic groups, e.g., non-photosynthetic diatoms (26, 27), dictyochophytes (28, 29), and chrysophytes (23, 30), and accordingly retain leucoplasts (non-photosynthetic plastids). Other non-photosynthetic stramenopile groups contain no photosynthetic members and no trace of plastids. These “aplastidic stramenopiles” include important pathogens: e.g., oomycetes, parasites of algae and plants, such as the causative agents of potato blight and sudden oak death (31, 32); labyrinthulomycetes, which are important pathogens of marine algae and invertebrates (33); and *Blastocystis*, a human gut commensal which may have harmful effects in immunodeficient hosts (34). In addition, the aplastidic stramenopiles include important marine saprotrophs (e.g., hypochytriomycetes, related to oomycetes (35)) and predator groups (e.g., bicosoecids, related to *Blastocystis* (36)).

Phylogenetic arrangement and endosymbiotic histories of the stramenopiles

The phylogenetic relationships within the stramenopiles have been revealed progressively by morphology, electron microscopy and biochemistry (7, 18), alongside single-gene, and subsequently multigene phylogenies (4, 6, 37) (Fig. 1). The exclusively non-photosynthetic groups form the most basally divergent stramenopile nodes, with oomycetes forming the closest major sister-group to the photosynthetic ochrophytes (6, 37) (Fig. 1). Within the ochrophytes, three major groups are apparent (Fig. 1) : the “chrysisista”, which include phaeophytes; chrysophytes, and the oil-producing alga *Nannochloropsis* (6, 38); the « hypogyristea », containing pelagophytes and dictyochophytes (6, 39); and the “khakista” or

diatoms, and their immediate sister group, a single-genus group alternatively labelled *Bolidomonas* or *Triparma* (6, 11). The hypogyristera and khakista are clearly sister-groups, whereas the monophyly of chrysista is uncertain, with some phylogenomic studies resolving the constituent lineages as a single group (37); and others indicating a basally divergent group of pinguiophytes, synchronophytes, synurophytes and chrysophytes, which precede the divergence of phaeophytes, raphidophytes and eustigmatophytes from the remaining ochrophytes (40, 41).

Chloroplasts originate via the endosymbiotic uptake of free-living photosynthetic bacteria (in the case of primary endosymbiosis), or of single-celled eukaryotic algae containing chloroplasts of red or green origin (in the case of secondary endosymbiosis; Fig. 1). The engulfed cells are converted into stable, intracellular organelles; and this process has been observed to occur in multiple algal groups across the tree of life (Fig. 1) (24). The chloroplasts found in ochrophytes are typically characterised by the presence of four surrounding membranes, the outermost of which is contiguous to the endoplasmic reticulum (42), and possess chlorophyll c as a light harvesting pigment (43), although exceptions to these paradigms are known (2, 44) (Fig. 2). Both of these features are characteristic of chloroplasts acquired through the secondary endosymbioses of red algae (i.e., those found in cryptomonads, haptophytes, dinoflagellates, and photosynthetic apicomplexans; Fig. 1), although are not known in red algae themselves (2). In contrast, ochrophytes do not retain phycobiliprotein subunits, which are found in red algae and in cryptomonad chloroplasts (45, 46). The ochrophyte chloroplast additionally is defined by a « girdle lamella », a ring-like structure consisting of three appressed thylakoids that cover the exterior of the chloroplast stroma (47), and an additional synapomorphy (the accessory pigments diatoxanthin and diadinoxanthin) unifies the Khakista (diatoms and *Bolidomonas*) and hypogyristera (pelagophytes and dictyochophytes) (2, 48, 49). (Figs. 1, 2).

Cosnsistent with the classical paradigm based on biochemical and ultrastructural data, multigene phylogenies of chloroplast genomes robustly resolve ochrophytes as a monophyletic group, with the chloroplast originating within red algae, and closely related to other chloroplast groups acquired through the secondary endosymbiosis of red algae (39, 50, 51). Thus, the most parsimonious scenario for the origin of the ochrophyte chloroplast is through a single endosymbiosis event in a recent common ancestor (39), prior to the divergence of diatoms (Fig. 1).

2. Phylogenomic insights into diatom evolution

Prior to the genomics era, the deep evolutionary history of the eukaryotic tree was resolvable only through ultrastructural similarities, and single-gene trees realised with individual markers (e.g., 18S and chloroplast 16S ribosomal DNA (18, 52)). Based on the conserved ultrastructural features of their chloroplasts, and single-gene phylogenetic data, it was posited that the major groups of algae with chloroplasts of secondary red origin, i.e., stramenopiles, cryptomonads, haptophytes, and dinoflagellates, descended from a common ancestor which acquired its chloroplast through a single secondary endosymbiosis involving a red alga (18, 52). These lineages were therefore proposed to form a monophyletic group, termed the “chromalveolates”; and non-photosynthetic members of these lineages, such as oomycetes, were posited to have therefore once possessed chloroplasts, and subsequently to have lost them (18).

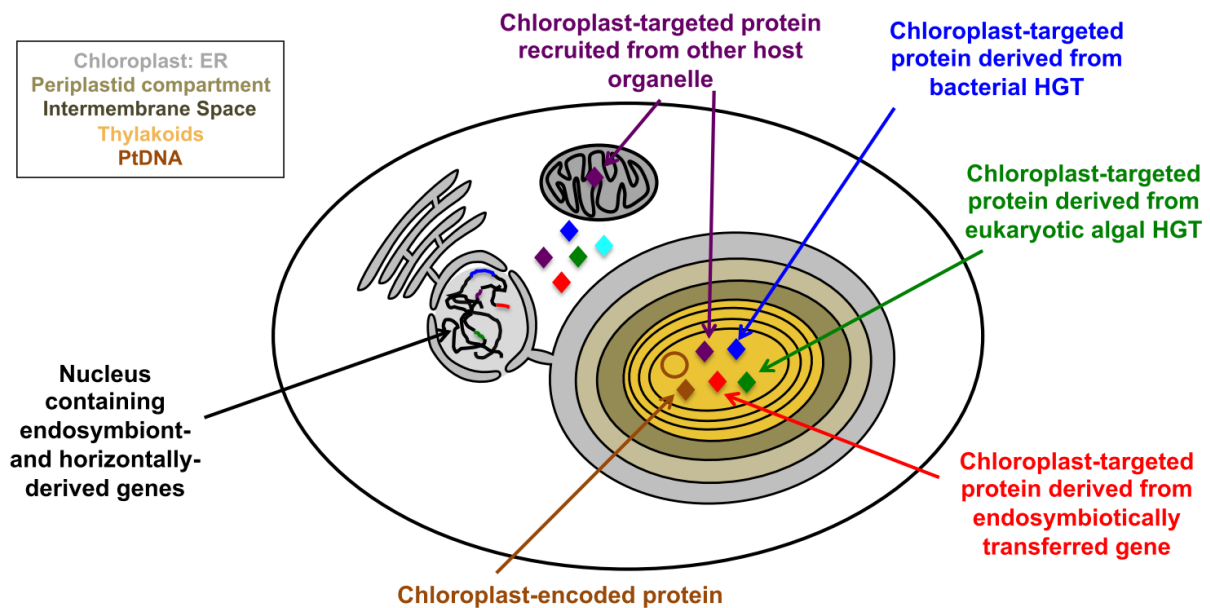


Fig. 2. Schematic diagram of the diatom chloroplast, adapted from (1, 2) ; demonstrating the ultrastructure of the chloroplast, and the major evolutionary classes of chloroplast- and nucleus-encoded chloroplast proteins.

Whole genome sequencing in diatoms commenced with the centric species *Thalassiosira pseudonana* and the pennate *Phaeodactylum tricoratum* (54, 55). This has also opened new windows of insight into the phylogenetic composition of the diatom nucleus (54-56), which has undergone substantial elaborations over its evolutionary history. The advent of high-throughput genomic sequencing, and therefore multigene phylogenomics revolutionised our understanding of diatom evolution, for example through the incorporation of rhizaria (which do not contain red algal chloroplasts) with stramenopiles and alveolates, as part of the “SAR clade” (19, 24). Further studies have positioned the nuclear lineages of cryptomonads as being more closely related to that of plants and red algae than that of diatoms, directly questioning the occurrence of a single, secondary endosymbiosis event (41, 57). Moreover, genomic investigation of major non-photosynthetic stramenopile groups, such as oomycetes (31) and hypochytriomycetes (35), has revealed an absence of genes of red algal origin that are shared with photosynthetic lineages (51, 58), placing the acquisition of the ochrophyte chloroplast after their divergence from non-photosynthetic relatives (Fig. 1).

The acquisition of a chloroplast by the common ancestor of the ochrophytes dramatically altered the composition of their genomes, adding a new organelle, with its own complex biochemical and physiological needs, to the host genome. Some of these functions are encoded in the chloroplast genome itself, and others in the genome of the stramenopile nucleus (Fig. 2). These nuclear genes, encoding chloroplast-targeted proteins, might have originated within the nucleus itself, and adapted to support the biology of the chloroplast (59-61), or might have been derived from either the nucleus or the chloroplast of the red algal ancestor of the ochrophyte chloroplast, and relocated to the host via endosymbiotic gene transfer (2, 39) (Fig. 2). At least some of the proteins in the chloroplast proteome may additionally have bifunctional roles in other organelles, e.g. through dual-targeting to the mitochondria (30, 39, 62) or to the plasma membrane (63) (Fig. 2).

Alongside this, both the nuclear and mitochondrial genomes of ochrophytes and aplastidic stramenopiles have undergone their own, independent gene transfer events, e.g., the relocation

of mitochondria-encoded functions to the nucleus, and the horizontal acquisition of non-chloroplast related genes of non-stramenopile origin by the stramenopile nuclear genome (55, 64, 65). Here, we summarise current knowledge (as of 2021) regarding different evolutionary sources of novelty in diatom nuclear genomes that have arisen through the course of stramenopile evolution, focusing on the genome of the model diatom *P. tricornutum*, which remains the best-studied system for phylogenomic reconstructions of gene transfer in the diatom lineage (55, 65) (Table 1). Subsequently, we explore the dynamic evolution of diatom mitochondria and chloroplasts, and identify features underpinning the progressive loss, and occasional gain, of novel coding functions in the organelle genomes of individual diatom species.

Source	Method	Bacterial genes	Red genes	Green genes	Other genes
Armbrust et al. (2004) (54)	First annotation of <i>T. pseudonana</i> (Tp) genome ; BLAST comparison to plant, red algae and animal genomes	n.a.	182 genes with exclusive homology to <i>Cyanidioschyzon merolae</i> (red algal reference)	865 genes with exclusive homology to <i>Arabidopsis thaliana</i>	806 genes with exclusive homology to <i>Mus musculus</i>
Bowler et al. (2008) (55)	First annotation of <i>P. tricornutum</i> (Pt) genome ; automated phylogenies	784, of which 587 unique to Pt; 267 shared with Tp ; and 73 conserved across stramenopiles	171, of which 160 only present in ochrophytes	n.a.	n.a.
Moustafa et al. (2009) (56)	Automated phylogenies of Pt and Tp genomes; <i>C. merolae</i> only complete red model	n.a.	418 in Pt ; 460 in Tp	1,757 in Pt ; 1,862 in Tp; 647 and 716 respectively of prasinophyte origin	248 Pt and 211 Tp genes of ambiguous green/red origin
Deschamps and Moreira (2012) (66)	Manual reannotation of data from (56)	n.a.	n.a.	89 « green » trees in (56) consistent with green origins ; 91 with red origins; remainder unresolved	n.a.
Dorrell et al. (2017) (39)	Manual phylogenies of 770 plastid proteins conserved across stramenopiles with densely sampled (genome and transcriptome) reference dataset	26/ 502 (5%) tractable proteins	255/502 (50%) tractable proteins ; of which only 4 shared with oomycetes	105/ 502 (20%) tractable proteins ; majority (76%) suggest position in chlorophytes to exclusion of streptophytes	243 inferred transfers from ochrophytes to haptophytes ; most likely from pelagophyte/dictyochophyte ancestor
Rastogi et al. (2018) (65)	BLAST best hit annotation of version 3 Pt genome with	584, of which 133 tractable to diatom common ancestor, 226 to	459, of which 50 tractable to diatom common ancestor, 353 to	1,981, of which 181 tractable to diatom	882 genes of inferred amorphean (animal, fungi,

	same reference dataset as (39) ; considering to last common SAR ancestor	ochrophyte ancestor, and 148 to stramenopiles	ochrophyte ancestor, and 28 to stramenopiles	common ancestor, 805 to ochrophyte ancestor, and 691 to stramenopiles	amoebozoan, or excavate) origin, most (480) tractable to stramenopile ancestor
Vancaester et al. (2020) (67)	Automated phylogenies of 20 genomes from diatoms and diatom relatives ; considering to last common SAR ancestor	E.g. 509 in Pt ; 983 in Tp ; 299 tractable to diatom common ancestor ; 327 to earlier in SAR evolution	n.a.	n.a.	n.a.
Fan et al. (2020) (68)	Automated phylogenies of 23 genomes of secondary chloroplast-containing algae	64 in Pt and 130 in Tp; 47 Pt and Tp genes shared with other species	n.a.	n.a.	n.a.
Dorrell et al. (2021) (40)	Manual phylogenies same dataset as (65) ; considering to last common ochrophyte ancestor, transfers with other secondary plastids, and direction (in/ out of diatoms)	285, of which 132 tractable to diatoms and 153 to earlier in ochrophyte evolution	193 with red algae, 231 with haptophytes, 156 with dinoflagellates, 127 with cryptomonads ; majority (631/ 707, 89%) resolved to early timepoints in ochrophyte evolution	260, of which 219 (84%) resolved to early timepoints in ochrophyte evolution	817 transfers from ochrophytes into haptophytes, 439 into dinoflagellates ; inferred to occur from pelagophyte/ dictyochophyte donor
Dorrell et al. (2021) (40)	Automated phylogenies all diatom genomes and transcriptomes (39) ; considering to last common ochrophyte ancestor	E.g. 106 in Pt and 134 in Tp; 429 transfers in polar centric/ pennate diatom ancestor; 102 in diatom common ancestor	n.a.	n.a.	n.a.

Table 1. Sizes and sources of gene transfers into and out of diatom nuclear genomes with both prokaryotic and eukaryotic partners, assembled from previous phylogenomic studies (39, 40, 54-56, 65-68)

Methods and problems when exploring origins of diatom nuclear genes

Dynamic transitions in the evolutionary history of the diatom nuclear, mitochondrial, and chloroplast genomes can be identified, and quantified, by comparative genomic and phylogenomic approaches (55, 56, 65) (Table 1). Understanding these transitions is important as it may reveal adaptations of diatoms to their current environments, and allow us to pinpoint genomic novelties that may explain their success over red or green algae, or even other stramenopile groups, in the contemporary ocean (5). However, the exact size, timing, reason, and even the veracity, of individual gene transfer events can be controversial (66, 69), depending particularly on the methodology used.

Firstly, different results may be obtained dependent on the reference library used, which may be constrained both by taxonomic undersampling and oversampling of key lineages. For example, early estimates of the green algal signal in diatom genomes (54, 55, 70) were likely to be overinflated by the paucity of red algal gene models at the time of analysis, being largely dependent on the highly reduced genome of the extremophilic alga *Cyanidioschyzon merolae* as the only publicly available red algal gene reference (71, 72).

In contrast, the indiscriminate use of transcriptome data and poorly curated reference genomes can lead to the misidentification of contaminants as horizontally transferred genes (39, 67, 73), or the preferential identification of genes with unresolvable or ambiguous origins as being related to oversampled taxa (66). These strategies may be avoided by judicious taxonomic sampling when constructing reference datasets (39, 61); alongside the use of new long-read sequencing technologies (e.g. ; PacBio (74)), assembly approaches (e.g.; Dovetail (75)), gene annotation pipelines (incorporative of transcriptome or proteomic data (65, 76)), and *in silico* cleaning of reference genome and transcriptome libraries (e.g., through considering nucleotide composition, (40, 77, 78)) to improve curation of gene models, and remove probable contaminant sequences.

Different results can also be obtained because of the methodology used, e.g., BLAST-based analyses such as ranking of top hits (51, 54, 65, 79) via whole-genome phylogenomic analysis (40, 61, 66-68). BLAST top hit analysis is less computationally intensive than phylogenetic techniques, and provides several technical advantages, e.g., the ability to infer possible evolutionary histories for genes that are too short or divergently evolved (e.g., chimeric gene fusions) to be resolved through classical phylogenetic approaches (39, 40, 79), but its resolution is necessarily poorer, for example not being able to determine the directions of gene transfers that can be inferred from the topology (e.g., monophyly versus paraphyly of stramenopile lineages) within a phylogenetic output (39, 61).

The threshold stringencies applied when identifying evolutionary origins may also bias the evolutionary relationships reconstructed. Too permissive criteria can admit poorly resolved or artifactual relationships, whereas too stringent parameters can conversely eliminate usable phylogenetic signals. For example, using phylogenomic approaches and automated tree sorting, Moustafa et al identified 418 red and 1,757 green genes in the *P. tricornutum* genome (56) (Table 1). Following manual reinspection of the green gene dataset identified by Moustafa et al (56), Deschamps and Moreira found that 91 of the putative green gene trees had topologies consistent with a vertical transfer of genes from the green lineage into the diatoms, whereas 89 of the putative green genes had histories more consistent with a transfer from the red lineage into the diatoms (66), although they did not perform a comparable analysis of the red genes identified by Moustafa et al. (Table 1).

Finally, even within well-sampled datasets, the exact questions that are posed by a study of horizontal gene transfer will determine the results obtained. Early studies on diatoms necessarily focussed on gene transfers from animals, fungi, bacteria, red algae and green algae, as these were the reference genomic models available at the time (54-56). However, diatoms may also have exchanged genes with other algae with secondary chloroplasts (e.g., cryptomonads, haptophytes and dinoflagellates); either as part of chloroplast endosymbiosis events (39, 40, 51) or through independent horizontal gene transfers (1, 63); and some of the “bacterial”, “red” or “green” genes identified in early studies may have passed through one or more of these algal groups as intermediates prior to arriving in diatom genomes (40). Genes

acquired horizontally into diatom genomes may have variously been received in early ancestors of the SAR clade, stramenopiles or ochrophytes, in the diatom ancestor, or even specifically within the diatom species considered; and recent studies have profited from the expanded number of diatom and stramenopile genomes and transcriptomes available to infer the probable timing of individual gene transfers (40, 67, 68). These more densely sampled datasets may even be able to identify occasions in which diatoms or ancestors have acted as donors, rather than recipients, in gene transfers with other lineages (39, 40).

In Table 1, we profile the key results of different studies of horizontal gene transfers involving the diatom nucleus, focussing on the *T. pseudonana* and *P. tricornutum* genomes (39, 40, 54-56, 65-68). The most recent of these analyses, involving a densely sampled reference dataset of all currently available eukaryotic and prokaryotic genomes and transcriptomes, and manual sorting of trees by partner lineage, timing and direction of horizontal transfer (40); identified 1,347/ 12,177 (11.0%) genes in the *P. tricornutum* genome to have been horizontally acquired since the origin of the ochrophytes, and a further 1,771 examples, distributed over 1,184 (9.7%) genes, of gene transfers from the ochrophytes into the other branches of the tree of life (40). Below, we discuss the likely origin points and functions of different horizontally transferred genes, including genes acquired from bacteria, red algae and their endosymbiotic descendants, green algae, as well as gene transfers from diatoms into other eukaryotic lineages.

Prokaryotic signals

Phylogenomic annotations of the *P. tricornutum* have identified large numbers of genes of prokaryotic origin, inferred to constitute 7.4% (784 genes) of the original genome annotation (55); and reduced to a still substantial 285 genes (2.3%) (40) in the most recent estimates from the third genome annotation with more densely sampled taxonomic references (Table 1) (40). Most of these genes pertain to bacteria, with limited or zero estimated archaeal and viral contributions into diatom genomes (40, 65). These genes have been acquired progressively through diatom evolution, with both studies focussed exclusively on *Phaeodactylum* (40, 65); and indeed on wider diatom pan-genomes (40, 67, 68), detecting large numbers of bacterially derived genes into early ancestors of the SAR clade, stramenopiles or ochrophytes; into the diatom common ancestor; and even into individual diatom groups or species (Table 1).

Two recent studies from Vancaester et al. (2020) (67) and from our group (40), which have reconciled automatically resolved bacterial transfers from densely sampled datasets with multigene reference tree topologies for SAR clade members, have attributed large numbers of bacterial gene transfer to early ancestors of the diatoms, following their divergence from other stramenopiles (40, 67) (Table 1). This may reflect a greater flux of bacterial DNA into diatoms than related algae, which may contribute to their comparative ecological success (40). However, it is possible that this asymmetry reflects the greater number of diatom reference sequence libraries in which gene transfers may be identified, and a parallel study by Fan et al. (2020) using a smaller number of diatom genomes, but balanced taxonomic sampling of diatom and non-diatom species, did not appear to yield greater number of bacterial genes in diatoms than other lineages (68) (Table 1). An alternative explanation is a high rate of transfer, but also secondary loss of bacterial-derived genes into diatoms, which may enable specific diatom species to explore new evolutionary niches and fluctuating environments (40).

Regardless of the frequencies of their arrival, bacterial genes contribute a wide range of different functions to diatoms. For example, within the dataset of 770 genes encoding chloroplast-targeted proteins that are shared across (and presumably ancestral to) all ochrophytes, we identified 49 of probable prokaryotic origin, including many implicated in the expression of the chloroplast genome (39). More recently, considering both *in silico* targeting predictions and GFP labelling, we have shown that bacterial genes acquired early during ochrophyte evolution are enriched in chloroplast-targeted proteins, consistent with origins tied to the ochrophyte chloroplast; whereas more recently acquired proteins predominantly function in the diatom secretome (40). Many of these more recently acquired bacterial genes have functions pertaining to environmental stress, e.g., ice-binding proteins in polar-adapted diatoms (80, 81); and proteins involved in the synthesis of Vitamin B₁₂, which may be a limiting micronutrient in many of the Southern Ocean habitats where diatoms are abundant (67, 82).

Red algal-related signals

Early estimates posited approximately 400 genes (4% of the coding content) in the *P. tricornutum* and *T. pseudonana* genomes were of red algal origin (Table 1) (55, 56, 65). These genes principally encode proteins of chloroplast function, and are shared across ochrophytes, e.g., constituting one half (255 of the 502 genes) of the chloroplast-targeted proteins shared across all ochrophytes identified by Dorrell et al. 2017 for which a tractable phylogenetic signal could be found (Table 1) (39). In contrast, only four of these genes were found to be shared with non-photosynthetic stramenopile groups, e.g., in oomycetes and labyrinthulomycetes (39) (Table 1). This, alongside the previously discussed lack of red algal signal in the genomes of non-photosynthetic stramenopiles (31, 35, 51, 58) is consistent with a late acquisition of the ochrophyte chloroplast, after their divergence from oomycetes (2).

Despite the massive expansion in the availability of red algal genome sequences and transcriptomes in the last decade (39, 72, 83), the number of red algal-derived genes identified in stramenopile genomes has remained relatively constant. For example, whereas Moustafa et al. (2009) identified 418 genes of putative red origin in the *P. tricornutum* genome using the extremophilic red alga *C. merolae* as a principle reference (56), only 459 red genes were identified by Rastogi et al. (2018) using a much more exhaustive dataset including five complete red algal genomes, and thirteen different red algal transcriptomes (65) (Table 1). One possible reason for this quite limited red contribution is the relatively high level of reduction observed in red algal genomes, which lack many of the accessory functions identified in diatoms and in other eukaryotic groups (65, 84).

An alternative scenario is that the red algal signal in stramenopile lineages does not itself directly come from red algae, but from a secondary, red-chloroplast containing lineage that was acquired by the stramenopiles through a tertiary or more complex endosymbiosis event. Using an innovative methodology based on the similarity of BLAST hit signals, Stiller et al. (2014) concluded a probable cryptomonad origin for the ochrophyte chloroplast (51). In our most recent analyses, considering algae with secondary chloroplasts as possible donors into diatom genomes, we could only manually identify 193 *P. tricornutum* genes that resolved phylogenetically as a direct sister to red algae, but found a further 514 that resolved with cryptomonads, haptophytes or dinoflagellates, with haptophytes contributing the greatest number (Table 1) (40). In this case, whatever functions that might have been inherited by the stramenopile ancestor from red algae would have been winnowed by not only one, but potentially two or more successive endosymbiotic events. Ultimately, identifying why

stramenopiles contain the red algal genes that they do will depend on tracing the permeation of the red signal across all major groups of algae with secondary, red-derived chloroplasts.

Green algal-derived signals

As discussed above, diatoms contain a sizeable number of genes of green algal origin, although the number identified varies based on the methodology employed (56, 61, 66) (Table 1). In our most recent estimates, we identified 260 (2%) genes in the *P. tricornutum* genome showing relationships consistent with a direct gene transfer from the green algae into the ochrophytes, most of which resolve to deep nodes in ochrophyte evolution (Table 1) (40), and 105 out of 502 (20%) of conserved ochrophyte chloroplast-targeted proteins for which a tractable origin was identified for green algae (39) (Table 1). The exact numbers and identities of the green genes in diatom genomes is likely to undergo further revision with deeper sequencing of red algae (e.g., xanthophyll cycle genes (85)), and the reassignment of some putatively “green” genes as having cryptomonad, haptophyte or dinoflagellate origins; although even the most stringent pipelines and sampling thresholds fail to completely eliminate the green signal from diatom genomes (40, 66).

Beyond the presence and size of the green signal in diatoms, other questions remain, namely: where do these genes come from, and how were they acquired? Phylogenomic analyses typically pinpoint green genes as arising within chlorophytes (i.e., single-celled green algae) rather than streptophytes (plants and their closest relatives), which helps argue against their origin as being artifactual because of phylogenetic misannotation (39, 56), although it remains to be determined where within the chlorophytes this signal preferentially falls (Table 1). The green genes that are present in diatom genomes preferentially encode chloroplast-targeted proteins (39), which might be consistent with a chloroplast endosymbiotic origin. This could be reconcilable either through a cryptic green endosymbiosis within the evolutionary history of the stramenopile host, or through the acquisition of a tertiary or higher chloroplast containing mixed signals of red and green origin (2, 5). Answering precisely how green genes were transferred into diatoms will depend both on deeper sequencing of the green algal tree (particularly in the context of early-diverging members, e.g. *Prasinoderma* (86)); and ideally the verification of the exact evolutionary origin of the ochrophyte chloroplasts (from red algae, cryptomonads, or another group entirely) (51).

Perhaps most interesting is to consider what functional advantages green genes might contribute to diatom genomes. Many of the green genes encoding chloroplast-targeted proteins have functions in biosynthetic pathways (e.g., chloroplast and carotenoid synthesis (2, 87, 88)). Other diatom green genes have functions in both chloroplast- and non-chloroplast related environmental stress responses, e.g., including *Lhcx* light-harvesting complex genes implicated in photoadaptation to aberrant light conditions (48, 89), and the iron-stress related protein *ISIP2a*, which mediates non-reductive iron uptake across the plasma membrane (90, 91). Understanding the functional significance of the green footprint in diatom genomes will depend on large scale characterisation of their encoded properties, e.g., using environmental sequence datasets, and targeted mutagenesis in model taxa (5).

Have diatoms donated genes to other organisms ?

Alongside the mosaic origin of diatom genomes, it is possible that diatoms, or ochrophytes in general, have donated genes to other organisms. For example, we previously identified 243 conserved chloroplast-targeted proteins that have been shared between the ochrophytes and

the haptophytes (39); which we have subsequently expanded to 817 genes (6.7%) of the *Phaeodactylum* genome supporting horizontal gene transfer from ochrophytes to haptophytes (40) (Table 1). Phylogenetic analysis of these genes indicated a mass transfer event from the hypogyristea (pelagophytes and dictyochophytes) into a common haptophyte ancestor (2, 39, 40). These genes predominantly have chloroplast-targeted functions, and may be footprints of an ancient endosymbiotic transfer of the ochrophyte chloroplast into the haptophytes (39, 51). We have subsequently inferred the same enrichment in pelagophyte/ dictyochophyte affinities for genes transferred from the ochrophytes to the dinoflagellates, which may relate to a direct or indirect (e.g., via haptophyte) gene transfer (40) (Table 1). Further phylogenetic analysis, of the chloroplast genomes of the photosynthetic alveolates *Chromera* and *Vitrella* has revealed a possible relationship with chrysophytes in the ochrophytes (12, 40, 92). These serial transfer events may help resolve the still uncertain origins, dynamics and functions of the red and green algal signals associated with marine algal genomes.

Despite the gene transfers associated with other ochrophytes, there is relatively little evidence that diatoms themselves have transferred large numbers of genes into other algal groups- with one exception. The « dinotoms », a closely related group of dinoflagellate algae within the order Peridiniales, possess whole cell endosymbionts of diatom origin (93, 94). These endosymbionts, which retain a complete chloroplast, mitochondria, and nucleus, appear to have been acquired, lost and replaced on multiple occasions, from both pennate and centric diatom lineages (95-97). Phylogenomic analysis of dinotom nuclear transcriptomes reveals very little evidence for the loss of functions from the symbionts, or transfer of genes to the host, consistent with their relatively transient evolutionary associations (77, 98). It remains to be determined what genetic integration events, if any, are required for the stable domestication of the dinotom endosymbionts.

Despite their relatively limited evolutionary interactions with other algal groups, diatoms may exchange genes with each other, which may have adaptive functions to certain environmental stresses. For example, the iron stress-associated protein ISIP1, which facilitates the non-reductive uptake of extracellular siderophores by endocytosis, shows a complex phylogeny, consistent with either multiple transfer events between diatoms, or independent paralogy and gene loss events (63). Similarly, phylogenies of ice-binding proteins, which confer cold stress tolerance in polar native algae, reveal probable horizontal transfer events between centric and pennate species (e.g., between the centric diatom *Chaetoceros* and the pennate *Navicula* sp.; or between the centric *Attheya* and the pennate *Amphora* sp. (80, 81)). Resolving these more recent gene transfer events is more challenging, as they are more likely to be biased by limitations in taxonomic sampling, and limited resolution of short protein sequences (99). However, explorations of adaptive and within-genus diatom gene transfers is becoming increasingly possible due to dense transcriptomic sampling of specific genera (e.g., *Thalassiosira*, *Chaetoceros* ; Fig. 1 (6)) or genomes (*Thalassiosira*, *Pseudo-nitzschia* ; (100)).

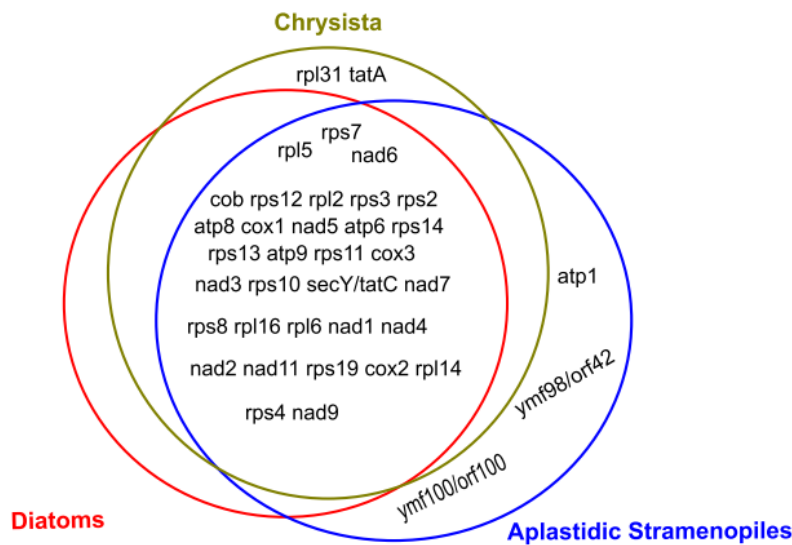
3. Evolution of the diatom organelle genomes

Diatom mitochondria

Diatom mitochondrial genomes display a largely well-conserved organisation, forming a single, circular chromosome with a single repeat region (8, 101). These genomes typically contain 33 core protein-coding genes, alongside genes for tRNAs and rRNAs (Fig. 3A) (101-103). This coding content is less than in some microbial eukaryotes (e.g., 47 protein-coding genes in the free-living stramenopile relative *Ancoracystis*), but somewhat greater than the

three protein-coding genes retained in dinoflagellate mitochondria (104, 105). The coding content of diatom mitochondria is similar to that of both ochrophytes and non-photosynthetic stramenopiles (Fig. 4A), except for the probable ancestral loss of *atp1* in diatoms, which is mitochondria-encoded in both photosynthetic eustigmatophytes (106) and non-photosynthetic oomycetes and bicoseocids (36). A further seven genes have more sporadic distributions across diatom mitochondria, including *tatA*, which is only mitochondria-encoded in a few select genera (13, 101, 107) (Fig. 3B).

Alongside this relatively conserved genomic content, more unusual organisations have evolved in individual mitochondrial genomes. These include structural changes, e.g., a possible linear organisation of the mitochondrial genome in the secondarily non-photosynthetic diatom *Nitzschia* (8), and expansions (103) or losses (108) of the repeat regions in different diatom species. These also include changes likely to impact on the



expression pathways of individual genes, e.g., the dynamic transfer and independent inheritance of introns in diatom mitochondrial *cox1* and *rnL* genes (101, 102, 107), independent origins of translationally fused gene pairs in the pennate species *Halamphora* and *Phaeodactylum* (101, 103), and the possible use of UGA-stop codons to encode tryptophan in the centric diatoms *Thalassiosira* and *Skeletonema* (109).



Fig. 3. Organisation of the diatom mitogenome. Top: Venn Diagram of mitochondrial genome-encoded proteins assignable to the common ancestors of diatoms, other ochrophytes (chrysisista), and aplastidic stramenopiles (oomycetes, labyrinthulomycetes, biocoseocids). **Bottom :** heatmap of the occurrence of seven genes with sporadic distribution in published diatom mitochondrial genomes. Red cells indicate presence and blank cells indicate absence.

Most dramatically, a change in the post-translational mitochondrial biology of raphid pennate diatoms (including *Fragilariopsis*, *Phaeodactylum*, and the diatom endosymbionts of dinotom algae) has been noted, in which the *nad11* gene is divided into two separately located and independently transcribed ORFs, corresponding to the

iron-sulfur binding and the molybdopterin-binding domains (103, 108, 110). This configuration is not known in araphid pennate or centric diatom mitochondria, and its functional consequences remain unknown (110).

Diatom chloroplast genomes

The chloroplast genomes associated with diatoms are typically arranged as a single, circular chromosome, with between 134 and 180 protein-coding genes, alongside ribosomal and transfer RNAs (Fig. 4) (2, 3, 111, 112). Phylogenomic analysis, and the presence of discrete molecular synapomorphies (e.g., the presence of a form ID type rubisco) robustly place this chloroplast genome within red algae, and closely related to the chloroplasts of cryptomonads and haptophytes (50, 113, 114). However, the coding content of the diatom chloroplast genome is somewhat less than the ca. 250 genes associated with red algal chloroplasts (50, 84), and indeed is less than the chloroplast gene contents of other ochrophyte groups (2), e.g., raphidophytes (*Heterosigma*) (13), phaeophytes (*Ectocarpus*) (14) and chrysophytes (*Ochromonas*) (12), pointing to both endosymbiotic- and post-endosymbiotic reductions in diatom chloroplast genome content (Fig. 3). This notwithstanding, the diatom chloroplast genome has undergone less dramatic reductions than some other ochrophyte groups (e.g., pelagophytes, dictyochophytes), in which even the loss of the chloroplast inverted repeat is known (9, 10).

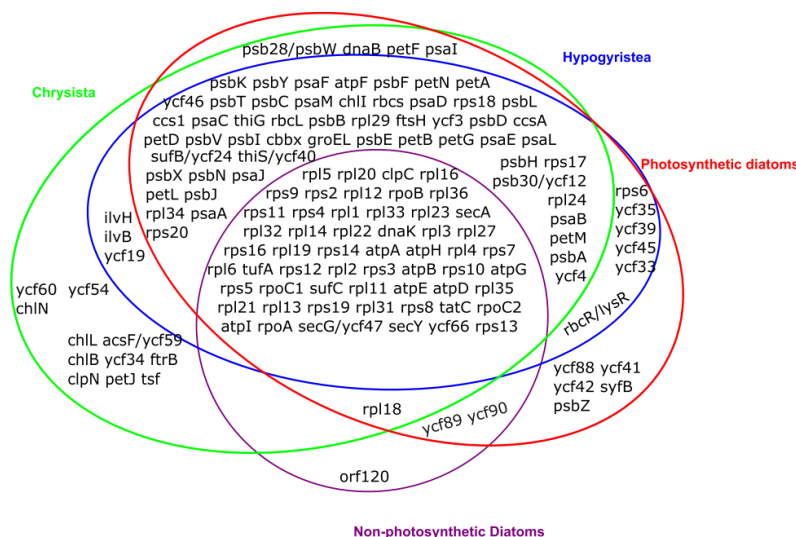
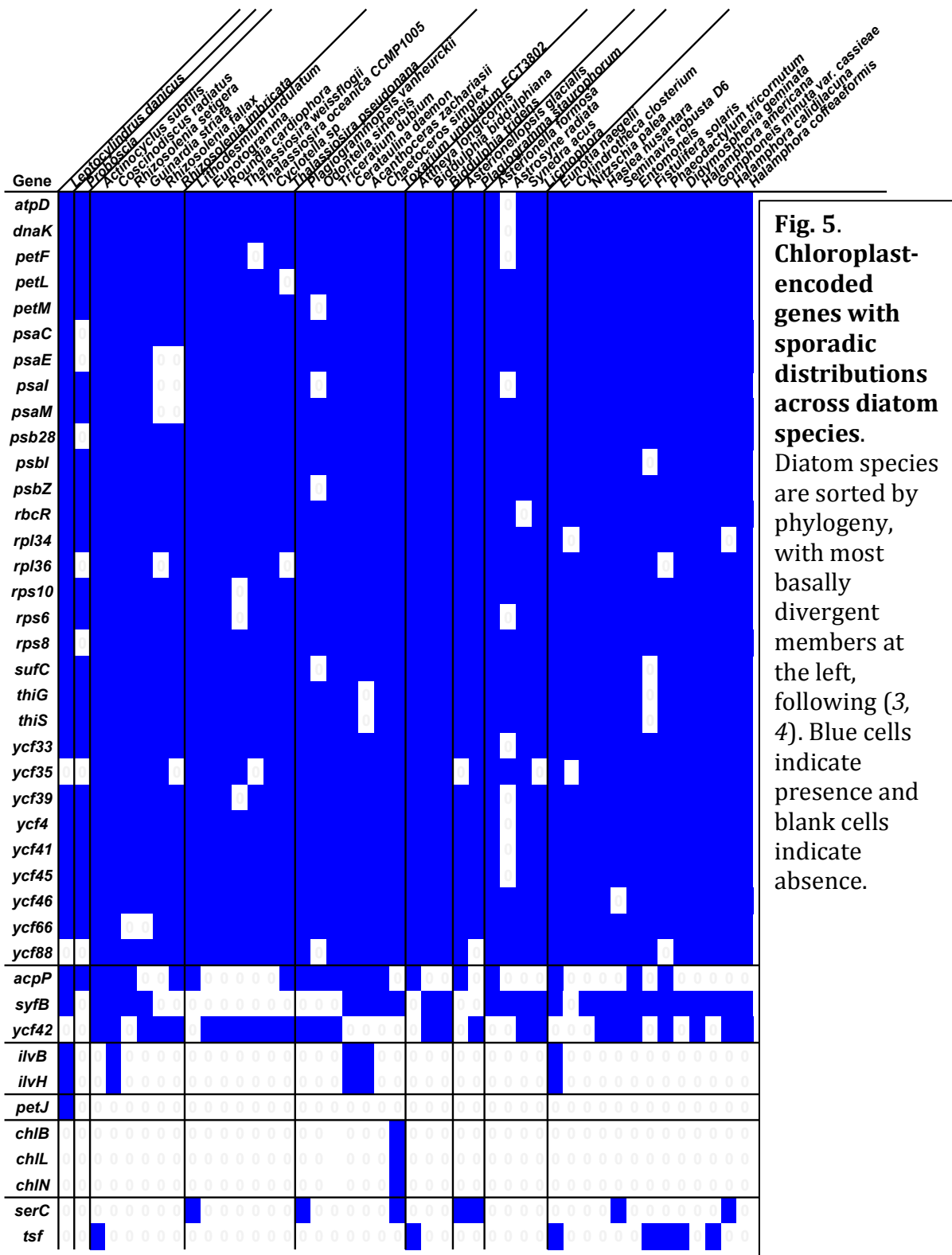


Fig. 4. Plastid genomes of diatoms and their relatives. Venn Diagram of chloroplast – encoded genes commonly found in photosynthetic diatoms (3), non-photosynthetic diatoms e.g. *Nitzschia* (8), ochrophytes assigned to the hypogyrystea (pelagophytes, dictyochophytes, *Triparma*) (9-11), and chrysiesta (chrysophytes, *Nannochloropsis*, phaeophytes) (12-15). Genes identified in fewer than 25% of published genomes within a particular lineage are not shown.

Alongside these general trends, different diatom species have retained and lost different patterns of genes (Fig. 5). For example, the basally divergent genus *Leptocylindrus* (4) retains a chloroplast *petJ* gene encoding cytochrome c_8 , which is also retained in other ochrophytes (e.g., *Ectocarpus*) (2, 14), suggesting it was present in the diatom common ancestor and lost from other species. Similarly, the *ilvB* and *ilvH* genes, encoding the small and large subunits of acetolactate synthase, have been retained in the chloroplast genomes of the diatom genera *Leptocylindrus*, *Coscinodiscus*, *Cerataulina*, *Acanthoceros* and *Eunotia* (3, 115). Conversely, the light-independent protochlorophyllide oxidoreductase complex, encoded by the *chlB*, *chlL* and *chlN* genes, is uniquely chloroplast-encoded in the diatom *Toxarium* (116). These diatoms are distantly positioned to one another and (except for *Leptocylindrus*) to the base of the diatom tree (4), but single-gene phylogenies reveal likely vertical origins of each gene (116), suggesting that they have been independently lost in a wide range of other diatoms.



Other diatoms have lost genes typically retained in the chloroplast genomes of other species (Figs. 5, 6). Substantial losses of chloroplast-encoded genes have been noted in the centric species *Proboscia* and the pennate species *Astrosyne radiata*, which form extremely long branches in diatom chloroplast gene trees (3, 117). *Astrosyne* was isolated from a shallow-water coral reef habitat in Guam (118), whereas *Proboscia* was isolated from the Red Sea, and it remains to be determined if this reductive evolution is correlated to high-light or temperature adaptations in either species (3).

Diatom leucoplast genomes

An even more dramatic degree of reduction is known in some diatoms within the genus *Nitzschia*, which have secondarily lost the capacity for photosynthesis, and may have originated through one (119) or independent evolutionary events (26). These species retain vestigial, non-photosynthetic plastids known as « leucoplasts », which perform essential biosynthetic functions and still retain genomes, but lack plastid-encoded genes associated with photosynthetic functions, e.g., genes for photosystem I, II, cytochrome b_6/f , the Calvin Cycle, and chlorophyll biosynthesis (Fig. 4) (8, 120). Notably, however, the *Nitzschia* leucoplast genome does retain genes encoding subunits of plastid ATP synthase (Fig. 4), which has been proposed to function in this organelle in the catabolic consumption of ATP, allowing the maintenance of a thylakoid proton gradient that would permit function of the thylakoid Tat protein import complex (120, 121). Comparative analysis of different non-photosynthetic *Nitzschia* species reveals largely conserved leucoplast genome content, consistent either with convergent trajectories following independent losses of photosynthesis, or limited divergence in genome content following a single loss of photosynthesis (8).

The *Nitzschia* leucoplast genome is notably less reduced in coding content compared to that of other secondarily non-photosynthetic plastids of similar evolutionary origin, e.g., those identified in non-photosynthetic chrysophytes or in apicomplexans (30, 122), which have lost a wider range of leucoplast-encoded functions, including ATP synthase. It has been proposed that the osmotrophic feeding strategies of non-photosynthetic *Nitzschia* spp., may limit their ability to supplement plastidial functions with heterotrophically acquired metabolites, explaining the greater relative functional autonomy of their leucoplasts compared to phagotrophic (chrysophyte) or parasitic (apicomplexan) lineages (30, 120). Ultimately, further functional characterisation is required to understand the metabolic contributions of the *Nitzschia* leucoplast in the context of other, secondarily non-photosynthetic algal species.; which will be aided by the recent completion of a nuclear transcriptome (120) and draft genome (27).

Why are certain genes lost from diatom chloroplasts?

It remains to be determined what physiological processes underpin the reductive evolution of diatom organelle genomes. This is the case both for genes that have been lost from diatom chloroplasts compared to other algal groups, and genes that have been lost from individual diatom lineages. In other groups of eukaryotes, e.g., plants and dinoflagellates, the loss of genes from the chloroplast genome is correlated to mutation rate and a loss in constraining selective pressure (123, 124), and it is notable that *Astrosyne* and *Proboscia* are two of the fastest-evolving diatom species, considering chloroplast-encoded substitution rates (117).

As a test of this principle, we have compared the degree of conservation between proteins encoded with universal, sporadic and occasional presence in diatom chloroplast genomes (Fig. 6). As controls, we have considered the conservation observed between orthologues from non-diatom stramenopiles, and between diatom and non-diatom species (Fig. 6). We note that the proteins with sporadic distributions in diatom species are much more divergent than would be expected compared to both proteins with conserved distributions, and orthologues from non-diatom stramenopiles (Fig. 6). Thus, the genes that are most frequently lost from diatom chloroplast genomes are likely to be those already under relaxed selection in the diatom chloroplast.

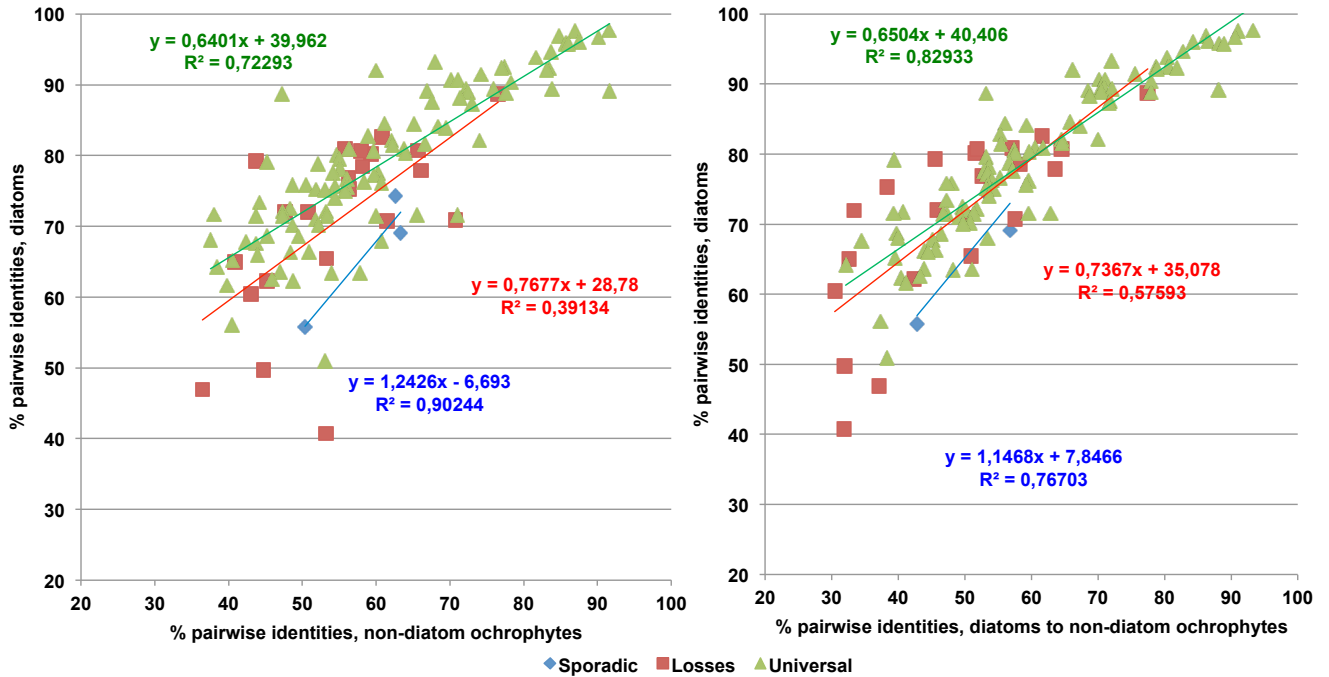


Fig. 6. Dynamic of evolutionary rates in diatom chloroplast genomes. Scatterplots showing the mean pairwise identities observed by BLAST searches between complete (>90% coverage) protein coding sequences with universal presence (green), identified to be infrequently lost (red), or to have sporadic distributions (*serC*, *tsf*, *ilvB*, *ilvH*) in diatom chloroplast genomes, plotted against similar pairwise identity scores for pairs of non-diatom stramenopile sequences (left), and pairs of diatom and non-diatom sequences (right). The regression gradients for the chloroplast-encoded proteins with incomplete presences in diatom chloroplasts are much steeper than those with universal presence, i.e., a small decrease in non-diatom chloroplast sequence identity correlates to a much larger decrease in diatom chloroplast sequence identity. This indicates that the chloroplast-encoded proteins that are variably present in diatom chloroplasts are more rapidly diverging, even within their chloroplast milieu, than genes that are never lost.

Loss versus transfer of diatom organelle genes

Another related question is under what circumstances genes are lost completely from diatom organelles, versus being relocated to the nucleus. Typically, genes are lost from chloroplast and mitochondrial genomes if they no longer perform necessary functions in these organisms.

In contrast, genes may be relocated to the nucleus to permit differential regulation, to protect from elevated chloroplast mutation rates, or to change the stoichiometry of their expression relative to chloroplast-encoded copies (124-126). In the context of diatoms, the open ocean species *Thalassiosira oceanica*, which is adapted to tolerate chronic iron limitation (127), has been proposed to have relocated the gene *petF*, encoding the iron-sulfur protein ferredoxin, to the nucleus, to allow its regulation in response to environmental iron availability (128). In another case, the *psb28* gene of the centric diatom *T. pseudonana* has been shown to be present as both chloroplast- and nucleus-encoded copies, suggesting it is midway through a functional chloroplast to nuclear transfer event (129). *Psb28* has a non-essential role in photosynthetic complex assembly, but its absence from the cyanobacteria *Synechococcus* and

Synechocystis leads to retarded growth under high light conditions (130). It remains to be determined whether the nucleus-encoded copies of these proteins permit photo-adaptation in *T. pseudonana*.

Gains and transfers of novel diatom organelle ORFs

Finally, alongside the loss and reduction of organelle-encoded genes, diatom chloroplasts and mitochondria possess novel ORFs not identified in other lineages. For example, some pennate diatom chloroplasts are known to contain *serC* and *tsf* genes, encoding serine and tyrosine recombinases (111). These genes are frequently located on plasmids, and may represent the traces of mobile genetic elements integrated into diatom chloroplasts (131, 132).

Other ORFs are either conserved across all diatoms, or indeed all stramenopiles, hence might form ancestral components of stramenopile organelle genomes. Among these is *ycf66*, a gene found in plant and green algal chloroplast genomes, and conserved in cyanobacteria (133), although it is frequently lost in some lineages (134, 135), and is not known in the chloroplasts of red algae. A phylogeny of *ycf66* positions the ochrophytes as a monophyletic group, within the green lineage, as a sister-group to the streptophyte lineage (i.e., land plants and their close relatives), except for the early-branching species *Klebsormidium* (Fig. 7).

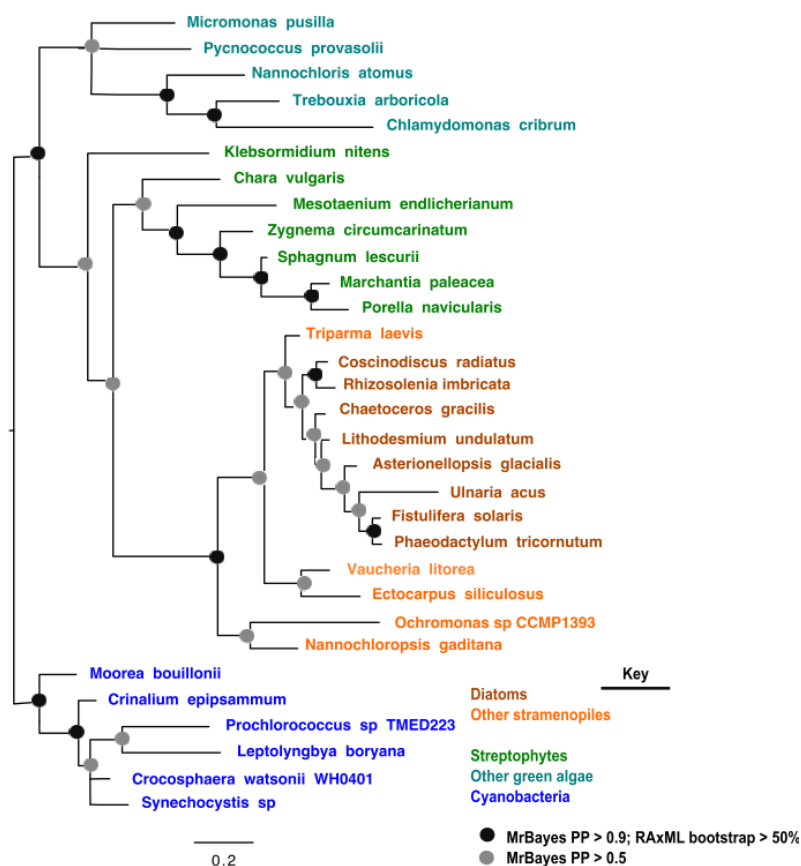


Fig. 7. Evolution of diatom chloroplast *ycf66*. Consensus MrBayes and RAxML tree, realised under three substitution matrices (GTR, Jones/JTT, WAG), for a 31 taxa x 93 aa alignment of *ycf66* from cyanobacteria and chloroplast genomes. The MrBayes topology supports placement of the ochrophyte sequences within the green algae, as a sister-group to all streptophyte species except *Klebsormidium sp.*

The most parsimonious explanation for this topology is a transfer of the *ycf66* gene from the green chloroplast lineage into a common ancestor of the ochrophyte chloroplast, concomitant with its endosymbiotic uptake, which would represent the first known example of a horizontal gene transfer event between two distantly related chloroplast genomes. Understanding the functional consequences of this transfer will depend on characterisation of the physiological role performed by *ycf66*, which remains unknown.

4. Concluding remarks

In this chapter, we have explored the phylogenetic context of the diatoms within their broader constituent lineage, the ochrophytes, within the stramenopiles (Fig. 1). Diatoms are evolutionary mosaics, containing complex chloroplasts with complicated evolutionary histories (Fig. 2); and nuclear genomes that have been enriched via by the horizontal and endosymbiotic acquisition of genes with chloroplast and non-chloroplast functions (Table 1). We show that the diatom nucleus is an evolutionary mosaic, supported by genes of prokaryotic, red, green, and other eukaryotic algal origin (Table 1). While the exact numbers, and even the individual evolutionary events that have given rise to these genes remain uncertain, their presence is indelible, and may have contributed unique biological functions to diatom cells.

Alongside this, we have compared reductive trajectories in mitochondrial (Fig. 3) and chloroplast genome content (Figs. 4, 5) across the stramenopile tree of life, focusing on sources of evolutionary difference within the diatom lineage. Diatom organelles are marked by reductive evolution, although greater extremes exist in both non-diatom and diatom branches of the stramenopile tree (e.g., non-photosynthetic members of the genus *Nitzschia*) than those observed in the diatom lineage as a whole. We show that the genes are frequently lost from diatom chloroplast genomes tend to exhibit lower global similarity than well-conserved genes (Fig. 6). It remains to be determined whether this is driven by chloroplast mutation rate, which appears to be somewhat lower in diatoms than that of the mitochondria, despite having a more conserved coding content (136) Figs. 4, 6); or via relaxed selection pressure, which has been shown to substantially vary across different chloroplast genes in other algal lineages (e.g., dinoflagellates (137)).

Finally, a possible gain of a chloroplast-encoded function has been identified, the uncharacterised open reading frame *ycf66*, which appears to have been acquired by ochrophyte chloroplasts via horizontal gene transfer of an equivalent open reading frame from green algal chloroplasts (Fig. 7). The role of this transfer event awaits functional and environmental characterisation. We stress the importance of considering function alongside the number, tempo, and mode of acquisition of novel diatom genes, when exploring the extraordinary environmental success of this lineage. In this context, emergent transformable systems (138-140), and environmental sequence datasets (141, 142), may cast new functional insights into the consequences of dynamic evolutionary events in diatom nuclei and organellular genomes.

Acknowledgments

RGD acknowledges a CNRS Momentum Fellowship (awarded 2019-2021).

5. References

1. T. Nonoyama *et al.*, Metabolic Innovations Underpinning the Origin and Diversification of the Diatom Chloroplast. *Biomolecules* **9**, 322 (2019).
2. R. G. Dorrell, C. Bowler, Secondary plastids of stramenopiles. In *Adv. Bot. Res.: Secondary Endosymbiosis*, Y. Hirakawa, Ed. (Elsevier, 2017) **84**, 57-103.
3. M. Yu *et al.*, Evolution of the plastid genomes in diatoms. In *Adv. Bot. Res. Plastid Genome Evolution*, S. Chaw, R. Jansen, Eds. (Elsevier, 2018) **85**, 129-155.

4. M. B. Parks, N. J. Wickett, A. J. Alverson, Signal, uncertainty, and conflict in phylogenomic data for a diverse lineage of microbial eukaryotes (diatoms, Bacillariophyta). *Mol Biol Evol* **35**, 80-93 (2018).
5. R. G. Dorrell, A. G. Smith, Do red and green make brown?: Perspectives on plastid acquisitions within chromalveolates. *Eukaryot Cell* **10**, 856-868 (2011).
6. M. Ichinomiya, A. Lopes dos Santos, P. Gourvil, S. Yoshikawa, M. Kamiya, K. Ohki, S. Audic, C. de Vargas, M.H. Noël, D. Vaultot, A. Kuwata. Diversity and oceanic distribution of the Parmales (Bolidophyceae), a picoplanktonic group closely related to diatoms. *ISME J* **10**, 2419-2436 (2016).
7. S. M. Adl *et al.*, The revised classification of eukaryotes. *J Euk Microbiol* **59**, 429-493 (2012).
8. R. Kamikawa, T. Azuma, K. I. Ishii, Y. Matsuno, H. Miyashita, Diversity of organellar genomes in non-photosynthetic diatoms. *Protist* **169**, 351-361 (2018).
9. K. Y. Han *et al.*, Dictyochophyceae plastid genomes reveal unusual variability in their organization. *J Phycol* **55**, 1166-1180 (2019).
10. H. C. Ong *et al.*, Analyses of the complete chloroplast genome sequences of two members of the Pelagophyceae: *Aureococcus anophagefferens* CCMP1984 and *Aureooumbra lagunensis* CCMP1507. *J Phycol* **46**, 602-615 (2010).
11. N. Tajima *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).
12. T. Ševčíková *et al.*, Updating algal evolutionary relationships through plastid genome sequencing: did alveolate plastids emerge through endosymbiosis of an ochrophyte? *Sci Rep* **5**, 10134 (2015).
13. R. A. Cattolico *et al.*, Chloroplast genome sequencing analysis of *Heterosigma akashiwo* CCMP452 (West Atlantic) and NIES293 (West Pacific) strains. *Bmc Genomics* **9**, 211 (2008).
14. G. Le Corguillé *et al.*, Plastid genomes of two brown algae, *Ectocarpus siliculosus* and *Fucus vesiculosus*: further insights on the evolution of red-algal derived plastids. *BMC Evol Biol* **9**, 253 (2009).
15. T. Ševčíková *et al.*, Plastid genomes and proteins illuminate the evolution of eustigmatophyte algae and their bacterial endosymbionts. *Genom Biol Evol* **11**, 362-379 (2019).
16. L. W. Parfrey, D. J. Lahr, A. H. Knoll, L. A. Katz, Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc Natl Acad Sci USA* **108**, 13624-13629 (2011).
17. J. W. Brown, U. Sorhannus, A molecular genetic timescale for the diversification of autotrophic stramenopiles (Ochrophyta): substantive underestimation of putative fossil ages. *Plos One* **6**, 12759 (2010).
18. T. Cavalier-Smith, A revised six-kingdom system of life. *Biol Rev* **73**, 203-266 (1998).
19. F. Burki *et al.*, Phylogenomics reshuffles the eukaryotic supergroups. *PLoS One* **2**, 000790 (2007).
20. M. Elias, N. J. Patron, P. J. Keeling, The RAB family GTPase Rab1A from *Plasmodium falciparum* defines a unique paralog shared by chromalveolates and Rhizaria. *Journal of Eukaryotic Microbiology* **56**, 348-356 (2009).
21. D. C. Fork, S. K. Herbert, S. Malkin, Light energy distribution in the brown alga *Macrocystis pyrifera* (giant kelp). *Plant Physiol* **95**, 731-739 (1991).
22. K. R. Hendry *et al.*, Competition between silicifiers and non-silicifiers in the past and present ocean and its evolutionary impacts. *Front Mar Sci* **5**, 22 (2018).

23. N. Graupner *et al.*, Evolution of heterotrophy in chrysophytes as reflected by comparative transcriptomics. *FEMS Microbiol Ecol* **94**, fiy039 (2018).
24. G. Walker, R. G. Dorrell, A. Schlacht, J. B. Dacks, Eukaryotic systematics: a user's guide for cell biologists and parasitologists. *Parasitol* **138**, 1638-1663 (2011).
25. V. Villanova *et al.*, Investigating mixotrophic metabolism in the model diatom *Phaeodactylum tricornutum*. *Philos Trans R Soc B* **372**, 20160404 (2017).
26. R. Kamikawa *et al.*, Multiple losses of photosynthesis in *Nitzschia* (Bacillariophyceae). *Phycol Res* **63**, 19-28 (2015).
27. A. Pendergrass, W. R. Roberts, E. C. Ruck, J. A. Lewis, A. J. Alverson, The genome of the nonphotosynthetic diatom, *Nitzschia sp.*: Insights into the metabolic shift to heterotrophy and the rarity of loss of photosynthesis in diatoms. *bioRxiv* **05.28.115543**, (2020).
28. H. Sekiguchi, M. Moriya, T. Nakayama, I. Inouye, Vestigial chloroplasts in heterotrophic stramenopiles *Pteridomonas danica* and *Ciliophrys infusionum* (Dictyochophyceae). *Protist* **153**, 157-167 (2002).
29. M. Kayama *et al.*, Highly reduced plastid genomes of the non-photosynthetic dictyochophyceans *Pteridomonas* spp. (Ochrophyta, SAR) are retained for tRNA-Glu-based organellar heme biosynthesis. *Front Plant Sci* **11**, 602455 (2020).
30. R. G. Dorrell *et al.*, Principles of plastid reductive evolution illuminated by nonphotosynthetic chrysophytes. *Proc Natl Acad Sci USA* **116**, 6914-6923 (2019).
31. J. W. Stiller, J. L. Huang, Q. Ding, J. Tian, C. Goodwillie, Are algal genes in nonphotosynthetic protists evidence of historical plastid endosymbioses? *BMC Genom* **10**, 484 (2009).
32. C. A. Levesque *et al.*, Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genome Biology* **11**, 73 (2010).
33. C. K. M. Tsui *et al.*, Labyrinthulomycetes phylogeny and its implications for the evolutionary loss of chloroplasts and gain of ectoplasmic gliding. *Mol Phylogenet Evol* **50**, 129-140 (2009).
34. L. Eme, E. Gentekaki, B. Curtis, J. M. Archibald, A. J. Roger, Lateral gene transfer in the adaptation of the anaerobic parasite *Blastocystis* to the gut. *Curr Biol* **27**, 807-820 (2017).
35. G. Leonard *et al.*, Comparative genomic analysis of the 'pseudofungus' *Hyphochytrium catenoides*. *Open Biol* **8**, 170184 (2018).
36. D. Jirsová, Z. Füssy, J. Richtová, A. Gruber, M. Oborník, Morphology, ultrastructure, and mitochondrial genome of the marine non-photosynthetic bicosoecid *Cafileria marina* gen. et sp. nov. *Microorganisms* **7**, 240 (2019).
37. R. Derelle, P. López-García, H. Timpano, D. Moreira, A phylogenomic framework to study the diversity and evolution of stramenopiles (heterokonts). *Mol Biol Evol* **33**, 2890-2898 (2016).
38. L. Rodolfi *et al.*, Microalgae for oil: strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor. *Biotechnol Bioengineering* **102**, 100-112 (2009).
39. R. G. Dorrell *et al.*, Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. *Elife* **6**, 23717 (2017).
40. R. G. Dorrell *et al.*, Phylogenomic fingerprinting of tempo and functions of horizontal gene transfer within ochrophytes. *Proc Natl Acad Sci USA* **118**, 2009974118 (2021).
41. F. Burki *et al.*, Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc Biol Sci* **283**, 20152802 (2016).

42. K. Ishida, T. Cavalier-Smith, B. R. Green, Endomembrane structure and the chloroplast protein targeting pathway in *Heterosigma akashiwo* (Raphidophyceae, Chromista). *J Phycol* **36**, 1135-1144 (2000).
43. K. V. Kowallik, B. Stoebe, I. Schaffran, P. Kroth-Pancic, U. Freier, The chloroplast genome of a chlorophyll a+c-containing alga, *Odontella sinensis*. *Plant Mol Biol Reporter* **13**, 336-342 (1995).
44. R. Wetherbee *et al.*, The golden paradox - a new heterokont lineage with chloroplasts surrounded by two membranes. *J Phycol* **55**, 257-278 (2019).
45. D. Bhattacharya *et al.*, Genome of the red alga *Porphyridium purpureum*. *Nat Commun* **4**, 1941 (2013).
46. S. Sturm *et al.*, A novel type of light-harvesting antenna protein of red algal origin in algae with secondary plastids. *BMC Evol Biol* **13**, 159 (2013).
47. R. A. Andersen, G. W. Saunders, M. P. Paskind, J. P. Sexton, Ultrastructure and 18S ribosomal RNA gene sequence for *Pelagomonas calceolata* gen. et sp. nov. and the description of a new algal class, the pelagophyceae *classis nov.*. *J Phycol* **29**, 701-715 (1993).
48. J. M. Buck *et al.*, Lhcx proteins provide photoprotection via thermal dissipation of absorbed light in the diatom *Phaeodactylum tricornutum*. *Nat Commun* **10**, 4167 (2019).
49. P. Kuczynska, M. Jemiola-Rzeminska, K. Strzalka, Photosynthetic pigments in diatoms. *Mar Drugs* **13**, 5847-5881 (2015).
50. S. A. M.-F. Muñoz-Gómez, F.G. Durnin, K. Colp, M. Grisdale, C.J. Archibald, J.M. Slamovits, C.H., The new red algal subphylum proteorhodophytina comprises the largest and most divergent plastid genomes known. *Curr Biol* **27**, 1677-1684 (2017).
51. J. W. Stiller *et al.*, The evolution of photosynthesis in chromist algae through serial endosymbioses. *Nat Commun* **5**, 5764 (2014).
52. T. Cavalier-Smith, Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J Euk Microbiol* **46**, 347-366 (1999).
53. J. W. Stiller, B. D. Hall, Long-branch attraction and the rDNA model of early eukaryotic evolution. *Mol Biol Evol* **16**, 1270-1279 (1999).
54. E. V. Armbrust *et al.*, The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**, 79-86 (2004).
55. C. Bowler *et al.*, The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239-244 (2008).
56. A. Moustafa *et al.*, Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* **324**, 1724-1726 (2009).
57. G. Lax *et al.*, Hemimastigophora is a novel supra-kingdom-level lineage of eukaryotes. *Nature* **564**, 410-414 (2018).
58. Q. Wang, H. Sun, J. Huang, Re-analyses of "algal" genes suggest a complex evolutionary history of oomycetes. *Front Plant Sci* **8**, 1540 (2017).
59. A. M. G. Novák Vanclová *et al.*, Metabolic quirks and the colourful history of the *Euglena gracilis* secondary plastid. *New Phytol* **225**, 1578-1592 (2020).
60. A. W. D. Larkum, P. J. Lockhart, C. J. Howe, Shopping for plastids. *Trends Plant Sci* **12**, 189-195 (2007).
61. A. A. Morozov, Y. P. Galachyants, Diatom genes originating from red and green algae: Implications for the secondary endosymbiosis models. *Mar Genomics* **45**, 72-78 (2019).

62. G. H. Gile, D. Moog, C. H. Slamovits, U. G. Maier, J. M. Archibald, Dual organellar targeting of aminoacyl-tRNA synthetases in diatoms and cryptophytes. *Genome Biol Evol* **7**, 1728-1742 (2015).
63. E. Kazamia *et al.*, Endocytosis-mediated siderophore uptake as a strategy for Fe acquisition in diatoms. *Sci Adv* **4**, 4536 (2018).
64. P. J. Keeling, J. D. Palmer, Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* **9**, 605-618 (2008).
65. A. Rastogi *et al.*, Integrative analysis of large scale transcriptome data draws a comprehensive landscape of *Phaeodactylum tricornutum* genome and evolutionary origin of diatoms. *Sci Rep* **8**, 4834 (2018).
66. P. Deschamps, D. Moreira, Re-evaluating the green contribution to diatom genomes. *Genome Biol Evol* **4**, 683-688 (2012).
67. E. Vancaester, T. Depuydt, C. M. Osuna-Cruz, K. Vandepoele, Comprehensive and functional analysis of horizontal gene transfer events in diatoms. *Mol Biol Evol* **37**, 3243-3257 (2020).
68. X. Fan *et al.*, Phytoplankton pangenome reveals extensive prokaryotic horizontal gene transfer of diverse functions. *Sci Adv* **6**, 0011 (2020).
69. T. Dagan, W. Martin, Seeing green and red in diatom genomes. *Science* **324**, 1651-1652 (2009).
70. R. I. Ponce-Toledo, P. López-García, D. Moreira. Horizontal and endosymbiotic gene transfer in early plastid evolution. *New Phytol* **224**, 618–624 (2019).
71. M. Matsuzaki *et al.*, Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* **428**, 653-657 (2004).
72. J. Collén *et al.*, Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *Proc Natl Acad Sci USA* **110**, 5247-5252 (2013).
73. F. Burki *et al.*, Re-evaluating the green versus red signal in eukaryotes with secondary plastid of red algal origin. *Genom Biol Evol* **4**, 626-635 (2012).
74. A. Rhoads, K. F. Au, PacBio Sequencing and Its Applications. *Genom Proteom Bioinform* **13**, 278-289 (2015).
75. K. M. Moll *et al.*, Strategies for optimizing BioNano and Dovetail explored through a second reference quality assembly for the legume model, *Medicago trunculata*. *BMC Genom* **18**, 578 (2017).
76. M. Yang, X. Lin, X. Liu, J. Zhang, F. Ge, Genome annotation of a model diatom *Phaeodactylum tricornutum* using an integrated proteogenomic pipeline. *Mol Plant* **11**, 1292-1307 (2018).
77. E. Hehenberger, F. Burki, M. Kolisko, P. J. Keeling, Functional relationship between a dinoflagellate host and its diatom endosymbiont. *Mol Biol Evol* **33**, 2376-2390 (2016).
78. S. Sato *et al.*, Genome-enabled phylogenetic and functional reconstruction of an araphid pennate diatom *Plagiosiriata* sp. CCMP470, previously assigned as a radial centric diatom, and its bacterial commensal. *Sci Rep* **10**, 9449 (2020).
79. R. Méheust, E. Zelzion, D. Bhattacharya, P. Lopez, E. Bapteste, Protein networks identify novel symbiogenetic genes resulting from plastid endosymbiosis. *Proc Natl Acad Sci USA* **113**, 3579-3584 (2016).
80. U. Sorhannus, Evolution of antifreeze protein genes in the diatom genus *fragilariopsis*: evidence for horizontal gene transfer, gene duplication and episodic diversifying selection. *Evol Bioinform Online* **7**, 279-289 (2011).
81. J. A. Raymond, H. J. Kim, Possible role of horizontal gene transfer in the colonization of sea ice by algae. *PLoS One* **7**, e35968 (2012).

82. T. J. Browning *et al.*, Nutrient co-limitation at the boundary of an oceanic gyre. *Nature* **551**, 242-246 (2017).
83. S. H. Brawley *et al.*, Insights into the red algae and eukaryotic evolution from the genome of *Porphyra umbilicalis* (Bangiophyceae, Rhodophyta). *Proc Natl Acad Sci USA* **114**, 6361-6370 (2017).
84. H. Qiu, J. M. Lee, H. S. Yoon, D. Bhattacharya, Hypothesis: Gene-rich plastid genomes in red algae may be an outcome of nuclear genome reduction. *J Phycol* **53**, 715-719 (2017).
85. O. Dautermann, M. Lohr, A functional zeaxanthin epoxidase from red algae shedding light on the evolution of light-harvesting carotenoids and the xanthophyll cycle in photosynthetic eukaryotes. *Plant J* **92**, 879-891 (2017).
86. L. Li *et al.*, The genome of *Prasinoderma coloniale* unveils the existence of a third phylum within green plants. *Nat Ecol Evol* **4**, 1220-1231 (2020).
87. S. Coesel, M. Obornik, J. Varela, A. Falciatore, C. Bowler, Evolutionary origins and functions of the carotenoid biosynthetic pathway in marine diatoms. *PLoS One* **3**, 2896 (2008).
88. R. Frommolt *et al.*, Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis. *Mol Biol Evol* **25**, 2653-2667 (2008).
89. C. Büchel, Evolution and function of light harvesting proteins. *J Plant Physiol* **172**, 62-75 (2015).
90. J. B. McQuaid *et al.*, Carbonate-sensitive phytoferritin controls high-affinity iron uptake in diatoms. *Nature* **555**, 534-537 (2018).
91. A. E. Allen *et al.*, Whole-cell response of the pennate diatom *Phaeodactylum tricornerutum* to iron starvation. *Proc Natl Acad Sci USA* **105**, 10438-10443 (2008).
92. J. I. Kim *et al.*, Evolutionary dynamics of cryptophyte plastid genomes. *Genome Biol Evol* **9**, 1859-1872 (2017).
93. B. Imanian, J. F. Pombert, P. J. Keeling, The complete plastid genomes of the two 'dinotoms' *Durinskia baltica* and *Kryptoperidinium foliaceum*. *Plos One* **5**, 107111 (2010).
94. N. Yamada *et al.*, Discovery of a kleptoplastic 'dinotom' dinoflagellate and the unique nuclear dynamics of converting kleptoplastids to permanent plastids. *Sci Rep* **9**, 10474 (2019).
95. J. Kretschmann, A. Žerdoner Čalasan, M. Gottschling, Molecular phylogenetics of dinophytes harboring diatoms as endosymbionts (Kryptoperidiniaceae, Peridinales), with evolutionary interpretations and a focus on the identity of *Durinskia oculata* from Prague. *Mol Phylogenet Evol* **118**, 392-402 (2018).
96. N. Yamada, S. D. Sym, T. Horiguchi, Identification of highly divergent diatom-derived chloroplasts in dinoflagellates, including a description of *Durinskia kwazulunatalensis* sp. nov. (Peridinales, Dinophyceae). *Mol Biol Evol* **34**, 1335-1351 (2017).
97. N. Yamada, H. Sakai, R. Onuma, P. G. Kroth, T. Horiguchi, Five non-motile dinotom dinoflagellates of the genus *Dinothrix*. *Front Plant Sci* **11**, 591050 (2020).
98. F. Burki *et al.*, Endosymbiotic gene transfer in tertiary plastid-containing dinoflagellates. *Eukaryot Cell* **13**, 00299-13 (2014).
99. N. G. Swenson, Phylogenetic resolution and quantifying the phylogenetic diversity and dispersion of communities. *PLoS One* **4**, e4390 (2009).
100. S. Basu *et al.*, Finding a partner in the ocean: molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. *New Phytol* **215**, 140-156 (2017).
101. R. M. Crowell, J. A. Nienow, A. B. Cahoon, The complete chloroplast and mitochondrial genomes of the diatom *Nitzschia palea* (Bacillariophyceae) demonstrate

- high sequence similarity to the endosymbiont organelles of the dinotom *Durinskia baltica*. *J Phycol* **55**, 352-364 (2019).
102. N. V. Ravin *et al.*, Complete sequence of the mitochondrial genome of a diatom alga *Synedra acus* and comparative analysis of diatom mitochondrial genomes. *Curr Genet* **56**, 215-223 (2010).
 103. M. P. Oudot-Le Secq, B. R. Green, Complex repeat structures and novel features in the mitochondrial genomes of the diatoms *Phaeodactylum tricorutum* and *Thalassiosira pseudonana*. *Gene* **476**, 20-26 (2011).
 104. J. Janoušek *et al.*, A new lineage of eukaryotes illuminates early mitochondrial genome reduction. *Curr Biol* **27**, 3717-3724.e3715 (2017).
 105. E. A. Nash *et al.*, Organization of the mitochondrial genome in the dinoflagellate *Amphidinium carterae*. *Mol Biol Evol* **24**, 1528-1536 (2007).
 106. T. Ševčíková *et al.*, A comparative analysis of mitochondrial genomes in eustigmatophyte algae. *Genom Biol Evol* **8**, 705-722 (2016).
 107. W. X. Guillory *et al.*, Recurrent loss, horizontal transfer, and the obscure origins of mitochondrial introns in diatoms (Bacillariophyta). *Genom Biol Evol* **10**, 1504-1515 (2018).
 108. S. M. An, J. H. Noh, D. H. Choi, J. H. Lee, E. C. Yang, Repeat region absent in mitochondrial genome of tube-dwelling diatom *Berkeleya fennica* (Naviculales, Bacillariophyceae). *Mitochondrial DNA* **27**, 2137-2138 (2016).
 109. M. Ehara, Y. Inagaki, K. I. Watanabe, T. Ohama, Phylogenetic analysis of diatom *coxI* genes and implications of a fluctuating GC content on mitochondrial genetic code evolution. *Curr Genet* **37**, 29-33 (2000).
 110. B. Imanian, J. F. Pombert, R. G. Dorrell, F. Burki, P. J. Keeling, Tertiary endosymbiosis in two dinotoms has generated little change in the mitochondrial genomes of their dinoflagellate hosts and diatom endosymbionts. *PLoS One* **7**, e43763 (2012).
 111. S. E. Hamsher *et al.*, Extensive chloroplast genome rearrangement amongst three closely related *Halamphora* spp. (Bacillariophyceae), and evidence for rapid evolution as compared to land plants. *PLoS One* **14**, e0217824 (2019).
 112. F. Prasetya *et al.*, *Haslea nusantara* (Bacillariophyceae), a new blue diatom from the Java Sea, Indonesia: morphology, biometry and molecular characterization *Plant Ecol Evol* **152**, 15 (2019).
 113. J. Janoušek, A. Horák, M. Oborník, J. Lukes, P. J. Keeling, A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proc Natl Acad Sci USA* **107**, 10949-10954 (2010).
 114. F. R. Tabita, S. Satagopan, T. E. Hanson, N. E. Kreel, S. S. Scott, Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *J Exp Bot* **59**, 1515-1524 (2008).
 115. J. S. M. Sabir *et al.*, Conserved gene order and expanded inverted repeats characterize plastid genomes of Thalassiosirales. *Plos One* **9**, 107854 (2014).
 116. E. C. Ruck, S. R. Linard, T. Nakov, E. C. Theriot, A. J. Alverson, Hoarding and horizontal transfer led to an expanded gene and intron repertoire in the plastid genome of the diatom, *Toxarium undulatum* (Bacillariophyta). *Curr Genet* **63**, 499-507 (2017).
 117. Y. Ren *et al.*, Nucleotide substitution rates of diatom plastid encoded protein genes are correlated with genome architecture. *Sci Rep* **10**, 14358 (2020).
 118. M. P. Ashworth, E. C. Ruck, C. S. Lobban, D. K. Romanovicz, E. C. Theriot, A revision of the genus *Cyclophora* and description of *Astrosyne* gen. nov

- (Bacillariophyta), two genera with the pyrenoids contained within pseudosepta. *Phycologia* **51**, 684-699 (2012).
119. A. Onyshchenko, E. C. Ruck, T. Nakov, A. J. Alverson, A single loss of photosynthesis in the diatom order Bacillariales (Bacillariophyta). *Am J Bot* **106**, 560-572 (2019).
 120. R. Kamikawa *et al.*, A non-photosynthetic diatom reveals early steps of reductive evolution in plastids. *Mol Biol Evol* **34**, 2355-2366 (2017).
 121. R. Kamikawa *et al.*, Proposal of a twin arginine translocator system-mediated constraint against loss of ATP synthase genes from nonphotosynthetic plastid genomes. *Mol Biol Evol* **32**, 2598-2604 (2015).
 122. L. Hadariová, M. Vesteg, V. Hampl, J. Krajšovič, Reductive evolution of chloroplasts in non-photosynthetic plants, algae and protists. *Curr Genet* **64**, 365-387 (2018).
 123. R. G. Dorrell *et al.*, Progressive and biased divergent evolution underpins the origin and diversification of peridinin dinoflagellate plastids. *Mol Biol Evol* **34**, 361-379 (2017).
 124. A. M. Magee *et al.*, Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genom Res* **20**, 1700-1710 (2010).
 125. R. G. Dorrell, C. J. Howe, What makes a chloroplast? Reconstructing the establishment of photosynthetic symbioses. *J Cell Sci* **125**, 1865-1875 (2012).
 126. Z. B. Noordally *et al.*, Circadian control of chloroplast transcription by a nuclear-encoded timing signal. *Science* **339**, 1316-1319 (2013).
 127. X. Gao, C. Bowler, E. Kazamia, , Iron metabolism strategies in diatoms. *J Exp Bot* **72**, 2165-2180 (2021).
 128. M. Lommer *et al.*, Recent transfer of an iron-regulated gene from the plastid to the nuclear genome in an oceanic diatom adapted to chronic iron limitation. *BMC Genom* **11**, 718 (2010).
 129. K. Jiroutova, L. Koreny, C. Bowler, M. Obornik, A gene in the process of endosymbiotic transfer. *Plos One* **5**, 013234 (2010).
 130. M. Bečková *et al.*, Association of Psb28 and Psb27 proteins with PSII-PSI supercomplexes upon exposure of *Synechocystis* sp. PCC 6803 to high light. *Mol Plant* **10**, 62-72 (2017).
 131. M. Hildebrand *et al.*, Plasmids in diatom species. *J Bacteriol* **173**, 5924-5927 (1991).
 132. J. D. Jacobs *et al.*, Characterization of two circular plasmids from the marine diatom *Cylindrotheca fusiformis*: plasmids hybridize to chloroplast and nuclear DNA. *Mol Gen Genet* **233**, 302-310 (1992).
 133. B. Stoebe, W. Martin, K. Kowallik, Distribution and nomenclature of protein-coding genes in 12 sequenced chloroplast genomes. *Plant Mol Biol Rep* **16**, 13 (1998).
 134. F. Leliaert *et al.*, Chloroplast phylogenomic analyses reveal the deepest-branching lineage of the chlorophyta, Palmophyllophyceae class. nov. *Sci Rep* **6**, 25367 (2016).
 135. L. Gao, Y. Zhou, Z. W. Wang, Y. J. Su, T. Wang, Evolution of the *rpoB-psbZ* region in fern plastid genomes: notable structural rearrangements and highly variable intergenic spacers. *BMC Plant Biol* **11**, 64 (2011).
 136. M. Krasovec, S. Sanchez-Brosseau, G. Piganeau, First estimation of the spontaneous mutation rate in diatoms. *Genom Biol Evol* **11**, 1829-1837 (2019).
 137. C. M. Klinger *et al.*, Plastid transcript editing across dinoflagellate lineages shows lineage-specific application but conserved trends. *Genom Biol Evol* **10**, 1019-1038 (2018).
 138. T. Mock *et al.*, Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature* **541**, 536-540 (2017).

139. J. K. Brunson *et al.*, Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom. *Science* **361**, 1356-1358 (2018).
140. A. K. Sharma, M. Nymark, T. Sparstad, A. M. Bones, P. Winge, Transgene-free genome editing in marine algae by bacterial conjugation - comparison with biolistic CRISPR/Cas9 transformation. *Sci Rep* **8**, 14401 (2018).
141. L. Caputi *et al.*, Community-level responses to iron availability in open ocean planktonic ecosystems. *Global Biogeochemical Cycles* **10**, 1029 (2019).
142. Q. Carradec *et al.*, A global ocean atlas of eukaryotic genes. *Nat Commun* **9**, 373 (2018).