



HAL
open science

Functional impacts of horizontal genome evolution across eukaryotic algae

Richard G Dorrell

► **To cite this version:**

Richard G Dorrell. Functional impacts of horizontal genome evolution across eukaryotic algae. Protistology. Université Paris-Saclay, 2024. tel-04760826

HAL Id: tel-04760826

<https://hal.science/tel-04760826v1>

Submitted on 30 Oct 2024

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.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Functional impacts of horizontal genome evolution across eukaryotic algae

Habilitation à diriger des recherches
de l'Université Paris-Saclay

présentée et soutenue à Gif-sur-Yvette, le 7 Octobre, par

Richard DORRELL

Composition du jury

Hervé PHILIPPE Directeur de Recherche (CNRS), SETE de Moulis/ Université de Toulouse	Rapporteur
Claire GACHON Professeure (MNHN), MCAM / Musée Nationale d'Histoire Naturelle	Rapporteuse
Benjamin FIELD Directeur de Recherche (CNRS), BIAM / Université Aix-Marseille	Rapporteur
Catherine JACKSON Directrice de Recherche (CNRS), IJM / Université Paris-Cité	Examinatrice
Purificación LOPEZ-GARCIA Directrice de Recherche (CNRS), ESE / Université Paris-Saclay	Examinatrice
Olivier VALLON Directeur de Recherche (CNRS), LBCPLM / Institut de Biologie Physico-Chimique	Examineur

Titre : Impacts fonctionnels de l'évolution horizontale du génome chez les algues eucaryotes

Mots clés : Dinoflagellés, Diatomées, Haptophytes, Chloroplastes, Endosymbiose, Transferts Horizontaux de Gènes

Résumé : Les micro-organismes eucaryotes, et en particulier les algues, jouent un rôle essentiel dans l'écosystème mondiale, en tant que producteurs primaires, dans le cycle des nutriments et à la base des réseaux alimentaires aquatiques. Les algues ont une histoire évolutive complexe et polyphylétique, due à l'acquisition endosymbiotique répétée de chloroplastes et des transferts horizontaux de gènes avec des partenaires algaux et bactériens dans les génomes nucléaires. Les algues les plus abondantes dans l'océan contemporain, telles que les dinoflagellés, les haptophytes et les diatomées, possèdent des chloroplastes originaires des endosymbioses secondaire ou supérieure d'algues rouges. Ces chloroplastes sont soutenus par des protéomes complexes codés par le noyau, qui peuvent être considérés comme des mosaïques définies à travers de l'histoire évolutive de l'hôte.

Mon équipe étudie comment l'évolution horizontale des algues informe leur biologie fonctionnelle et leur écologie actuelle. Cela implique trois projets complémentaires : la reconstruction computationnelle des protéomes des chloroplastes d'algues encodés dans le noyau, avec une emphase sur les dinoflagellés ; la caractérisation expérimentale et environnementale de nouvelles protéines soutenant le chloroplaste des diatomées ; et l'évaluation phylogénétique des transferts horizontaux de gènes plus larges qui sous-tendent l'évolution des algues, avec un accent particulier sur les haptophytes et les espèces autochtones de l'océan Arctique. Enfin, je considère comment ces histoires et approches combinées peuvent s'entrelacer pour éclairer notre compréhension de la sensibilité et de la résilience des algues au changement climatique anthropique.

Title : Functional impacts of horizontal genome evolution across eukaryotic algae

Keywords : Dinoflagellates, Diatoms, Haptophytes, Chloroplasts, Endosymbiosis, Horizontal Gene Transfer

Abstract : Eukaryotic micro-organisms, and particularly algae, play essential roles in the Earth ecosystem, as primary producers, in nutrient cycling, and at the basis of aquatic food webs. Algae have a complex and polyphyletic evolutionary history, which has involved the repeated endosymbiotic acquisition of chloroplasts, and horizontal gene transfers with both algal and bacterial partners into the nuclear genomes. The most abundant algae in the contemporary ocean, such as dinoflagellates, haptophytes and diatoms, possess chloroplasts acquired through the secondary or higher endosymbioses of red algae. These chloroplasts further are supported by complex, nucleus-encoded proteomes, which can be considered to be mosaics, defined by the entire evolutionary history of the host.

My group studies how the horizontal evolution of algae with secondary red chloroplasts informs their current functional biology and ecology. This involves three complementary projects: computational reconstruction of the nucleus-encoded proteomes of algal chloroplasts, with a particular focus on dinoflagellates; experimental and environmental characterisation of novel proteins supporting the diatom chloroplast; and phylogenetic evaluation of the broader horizontal gene transfers underpinning algal evolution, with a particular focus on haptophytes and distantly-related species native to the Arctic Ocean. Finally, I consider how these combined stories and approaches may intertwine to inform our understanding of algal sensitivity and resilience to anthropogenic climate change.

(Short) acknowledgments

In a randomly determined order:

My group members (Tomomi, Shun, Anna, Louis, Mathias, Victoria)

The jury for their patience

Many, many students and interns

Some colleagues at the LCQB and the IBENS

Engineers and support staff at both laboratories, for keeping the show running

Collaborators along the way

My husband and family

The algae

My garden, for sanity reasons

Also, for help bringing this together: Emmanuelle Baudry, Martine Boccara, Marianne Delarue, Silvia de Monte, Uta Paszkowski. *Big thanks.*

Table of Contents

Chapter 1: Introduction and context	4
1.1. <i>The importance of protists to the Earth ecosystem.....</i>	4
1.3. <i>Algae from functional and phylogenetic perspectives.....</i>	6
1.3. <i>The endosymbiotic origin and horizontal evolution of eukaryotic algae</i>	10
1.4. <i>Outline of work performed in my group.....</i>	13
Chapter 2: Computational reconstruction of the evolutionary trajectories of dinoflagellate chloroplast proteomes (and beyond)	16
2.1. <i>Dinoflagellates: an open-ended experiment in chloroplast evolution</i>	16
2.2. <i>Tracing the divergent evolution of dinoflagellate chloroplast genes through a transcriptomic approach</i>	19
2.3. <i>Corrective- and diversifying- functions of chloroplast RNA editing in Kareniaceae</i>	23
2.4. <i>A holistic appraisal of the Kareniacean chloroplast proteome reveals an intricate mosaic- and repeated endosymbioses</i>	26
2.5. <i>Independent and progressive reductions of the chloroplast proteomes of chrysophyte algae</i>	32
2.6. <i>Concluding remarks and pedagogical insights.....</i>	39
Chapter 3: Integrative environmental and experimental characterisation of the 'dark matter' of diatom chloroplast proteomes	43
3.1. <i>Diatoms: a model primary producer in a wider constellation of algal diversity</i>	43
3.2. <i>Evaluating- with limitations- the red and green origins of the diatom chloroplast</i>	47
3.3. <i>Moving from evolutionary to functional annotation of the diatom chloroplast</i>	50
3.4. <i>Convergent phenotyping of diatom chloroplast glycolysis from environmental and experimental data.....</i>	57
3.5. <i>Ongoing characterisations of novel molybdenum transporters and ATP synthases associated with the pan-algal chloroplast.....</i>	62
3.6. <i>Concluding remarks and pedagogical insights</i>	67
Chapter 4: Phylogenetic and biogeographical understanding of horizontal evolution across algal pan-genomes	71
4.1. <i>Quantifying the evolutionary impacts of inter-algal HGT.....</i>	71
4.2. <i>Identification of hidden chloroplast links between stramenopiles, haptophytes and dinoflagellates.....</i>	75
4.3. <i>Frequency and localisation biases contextualise bacterial horizontal gene transfer into diatom genomes.....</i>	79
4.4. <i>Convergence and "Arctic-specific" genes in an Arctic pan-algal genome</i>	84
4.5. <i>Taxonomic and biogeographical trends in Vitamin B₁₂ metabolism loss across the photosynthetic tree of life</i>	91
4.6. <i>Concluding remarks and pedagogical insights</i>	93
Chapter 5: Perspectives and future directions	98
References	102

Chapter 1: Introduction and context

1.1. The importance of protists to the Earth ecosystem

The vast majority of earth's biodiversity is microbial. Single-celled organisms (bacteria, archaea, viruses, and protists), while only first visualised by humans in 1677 pre-date us by billions of years (Lane, 2015). Protists, or eukaryotic micro-organisms, are found across all habitats, ranging from the deep-sea abyssal plain to the stratosphere, and from hypersaline geothermal springs to the polar ice-caps (Epstein & López-García, 2007). Many remain anonymous, with some groups only starting to enter laboratory study (Del Campo et al., 2023; Walker et al., 2011).

My interests in eukaryotic microorganisms really began from the perspective of Earth systems science, reading seminal work from James Lovelock and Lynn Margulis (Lovelock, 2000; Margulis & Dolan, 2002) while I was still in secondary school. So I would like to begin by reflecting on the essential functions that protists perform across all Earth ecosystems and in the context of human society, where they are directly implicated in over half of the UN Sustainable Development goals (**Fig. 1.1**) (Nhamo et al., 2021).

Photosynthetic protists, or "algae", which are unified by the presence of chloroplasts, are responsible for nearly half of all planetary primary production (Dorrell & Smith, 2011; Friedlingstein et al., 2022). Algae further support aquatic food webs, forming the trophic basis of worldwide fish stocks (Asch et al., 2019), and acting as essential symbionts of coral reefs (Janouškovec et al., 2012). Multicellular species that have arisen within the algae (i.e., seaweeds) have direct importance to human society as subsistence crops (Msuya & Hurtado, 2017). Unicellular algae may serve as chassis for industrial production (biofuels, fertilisers, high-value products), and even for green carbon capture (Pires, 2017). These applications are appealing given the lower land use requirements (particularly productive arable land) for algal bioreactors than for biofuel and fodder crops (Shareefdeen et al., 2023). Other, harmful bloom-forming algae have severe impacts on marine fauna and thus aquatic ecology, which may be intensified by climate change (Wells et al., 2015).



Fig. 1.1. Varied importance of microbial eukaryotes to the Earth ecosystem. The inner wheel of this figure shows 8 UN Sustainable Development Goals (SDGs), adapted freely from (Nhamo et al., 2021), for which protist biology and diversity are relevant. The outer wheel shows drawings of representative protists, adapted freely from (Keeling & Eglit, 2023), overlaid over SDGs to which they are particularly relevant. The visual organisation of the drawings are independent of taxonomic affiliation, shown in **Fig. 1.3**.

Both photosynthetic and heterotrophic protists play important roles in soil and aquatic ecosystems in decomposing organic carbon and nitrogen, and support global agriculture (Hirt et al., 2023). Some may act as biomarkers of ecosystem health, facilitating nutrient cycling and excluding pathogens (Perez-Lamarque et al., 2023; Xiong et al., 2020). Other non-photosynthetic protists may be important parasites of humans, including the causative agents of malaria, (*Plasmodium falciparum*) and African sleeping sickness (*Trypanosoma brucei*) (Bannerman et al., 2018; Walker et al., 2011), or pathogens of crops and wild plants, e.g. oomycetes (Gachon et al., 2010; Lévesque et al., 2010). Finally, even protists with limited industrial applications may act as evolutionary models, or provide genetic information of new

metabolic pathways that can change our view of the functional cell biology and eco-physiological diversity of life (C. L. Jackson et al., 2023; Richardson et al., 2015).

Group	Type species	Brief description
Cyanobacteria	<i>Prochlorococcus</i> , <i>Synechococcus</i>	Single-celled and colonial bacteria ("blue-green algae") capable of oxygenic photosynthesis, and predominant in nutrient-poor habitats (e.g. the tropical open ocean)
Archaeplastids	-	The grouping of red algae, green algae including plants, and glaucophytes, all of which possess chloroplasts of primary (cyanobacterial) origin
Red algae/ Rhodophytes	<i>Porphyra</i> , <i>Cyanidioschyzon</i>	Regroups both cultivated multicellular seaweeds (e.g., nori) and extremophilic single-celled algae growing in acidic hot springs. Probable ancestor of the diatom, haptophyte, etc plastid.
Green algae/ Chlorophytes	<i>Chlamydomonas</i> , <i>Ostreococcus</i>	The closest unicellular relatives of plants. Include "core Chlorophyceae", found in terrestrial and freshwater habitats, and prasinophytes, a paraphyletic group abundant in the ocean.
Streptophytes	<i>Chara</i> , <i>Arabidopsis</i>	The group of multicellular charophytes and plants. Only found in freshwater and on land with no marine species.
Cryptomonads	<i>Guillardia</i>	Algae with secondary red chloroplasts, potentially related to Archaeplastids, possessing a "nucleomorph" (vestigial nucleus from the plastid endosymbiont)
Haptophytes	<i>Emiliana</i>	Single-celled and often calcareous marine algae (whose exoskeleta form the major component of limestone and chalk). Abundant in marine habitats. Grouped in Fig. 1.2 with cryptomonads (as "CCTH") but more likely related to SAR.
SAR	-	The monophyletic and strongly-supported grouping of Stramenopiles, Alveolates and Rhizaria (the latter including foram and phytomixid protists and chlorarachniophyte algae)
Stramenopiles	-	The grouping of ochrophytes (diatoms, pelagophytes, phaeophytes, chrysophytes) and non-photosynthetic protists including oomycetes and labyrinthulomycetes, within SAR.
Diatoms	<i>Phaeodactylum</i>	Sessile, single-celled and chain-forming algae highly abundant in marine habitats. Include pennate and centric groups, differentiated by cell morphology.
Pelagophytes	<i>Aureococcus</i>	Abundant marine flagellated algae with relatively few (<40) cultured strains. Immediate sister to the dictyochophytes (silicoflagellates) and thereafter the diatoms.
Chrysophytes	<i>Ochromonas</i>	"Golden algae", freshwater and marine ochrophytes. May be basally divergent to all other ochrophytes or closely related to phaeophytes (giant kelps).
Oomycetes	<i>Phytophthora</i>	"Slime molds", typically pathogenic protists including important crop pathogens (downy mildews). Related to diatoms, etc in the stramenopiles, but lacking an obvious plastid.
Alveolates	-	The grouping of dinoflagellates, apicomplexans (including chromerids), and one non-photosynthetic group (ciliates). Sister to stramenopiles within SAR.
Dinoflagellates	<i>Amphidinium</i> , <i>Symbiodinium</i>	Photosynthetic, heterotrophic, parasitic and toxic (red tide-) forming algae. Notably include the photosymbionts of coral reefs.
Apicomplexa	<i>Plasmodium</i> , <i>Chromera</i>	Includes photosynthetic members ("chromerids") associated with coral, and parasites, e.g. the malaria pathogen <i>Plasmodium</i> . The latter are descended from algae, and possess a non-photosynthetic plastid ("apicoplast") that may form a target for drug treatment.
Euglenids	<i>Euglena</i>	Algae with secondary green chloroplasts, distantly related from all other groups.

Table 1.1. Glossary definitions of some of the algal and protists groups discussed in this memoir.

1.3. Algae from functional and phylogenetic perspectives

My previous work, particularly my PhD (Dorrell & Howe, 2012a, 2015) and early post-doctoral studies (Dorrell et al., 2022; Dorrell & Bowler, 2017), has focussed on one specific component of protist biology: the phylogenomic origins of eukaryotic algae and their chloroplasts (**Table 1.1**). These organisms can be taxonomically distant from one another, in terms of their nuclear and mitochondrial genomes (Dorrell & Howe, 2012b; Dorrell & Smith, 2011), but nonetheless are unified by their ability to photosynthesise and perform essential primary metabolism in their chloroplasts (Ait-Mohamed et al., 2020; Nonoyama et al., 2019), acquired through multiple endosymbiosis events.

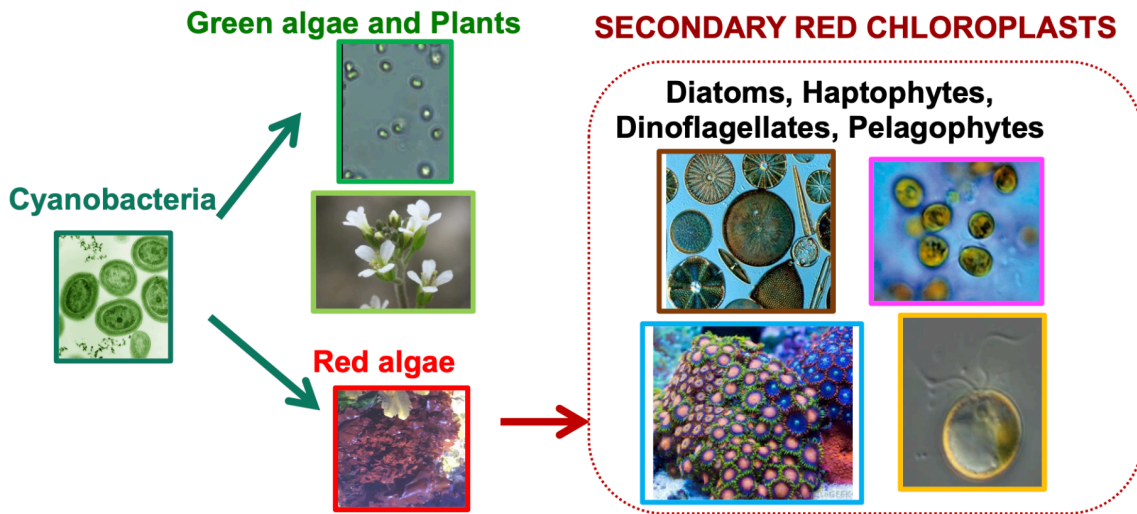


Fig. 1.2. Endosymbiotic evolution of algae and their chloroplasts. A: Schematic diagram of the successive spread of photosynthesis across algal groups by primary and secondary endosymbiosis events. For concision minor groups (glaucophytes, secondary green chloroplasts) are excluded, and all secondary red chloroplasts are shown deriving directly from red algae as opposed to higher (e.g., tertiary, (Stiller et al., 2014)) endosymbiotic origin.

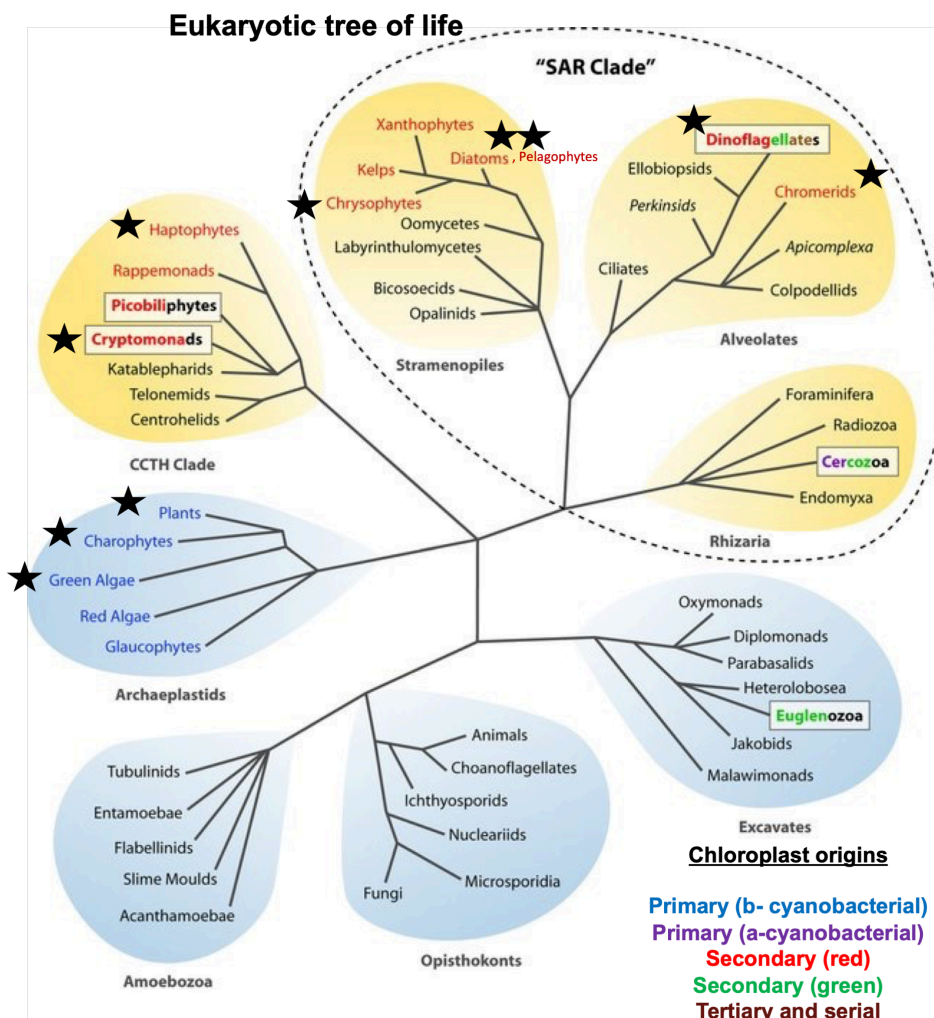


Fig. 1.3. Taxonomic diversity of eukaryotic algae. This eukaryotic tree of life, adapted from (Dorrell & Smith, 2011), shows six "supergroups" of which one (SAR) is split into its three lineages (Stramenopiles, Alveolates, and Rhizaria). Supergroups in which most photosynthetic taxa are found are shaded in yellow. Algae and plants are shown in coloured text, corresponding to the chloroplast evolutionary origin. Lineages studied by my group are asterisked.

In Figs. **1.2-1.4** I present three different views of the diversity of algal life which have shaped the research in my group. **Fig. 1.2** describes two major endosymbiotic events that have given rise to the majority of eukaryotic algal groups. These relate to a primary endosymbiosis of a cyanobacterium, most likely a sister-group to the early-diverging lineage *Gloeomargarita* (**Fig. 1.2**) (Ponce-Toledo et al., 2017) in a common ancestor of the archaeplastids, incorporating plants, primary red and green algae. Beyond this, one or more secondary endosymbiosis of eukaryotic red algae, themselves possessing a chloroplast of primary endosymbiotic origin, have given rise to most of the eukaryotic algal groups, such as diatoms, haptophytes, pelagophytes and dinoflagellates, found in the contemporary ocean. The ancestor of the secondary red chloroplast is deeply positioned within the red algae, without evident living descendants (Muñoz-Gómez et al., 2017; Novak et al., 2024). Other secondary endosymbioses, e.g. of green algae, have also occurred.

Fig. 1.3 shows a rough tree of eukaryotic nuclear and mitochondrial diversity in which the eukaryotes (and predominantly protists) resolve into a number of supra-phylum level assemblies termed “supergroups” (Dorrell & Smith, 2011; Walker et al., 2011). While this tree is out of date with regards to the monophyly of some of these lineages (e.g., “CCTH” and Excavate clades, (Eglit et al., 2024; Strassert et al., 2021)), it shows that eukaryotic algae primarily fall into three supergroups: the archaeplastids; and diverse members of the “SAR” (expanded in **Fig. 2A** into its constituents: Stramenopiles, Alveolates and Rhizaria) and “CCTH clades”, in which the secondary red chloroplast is widespread.

The evolutionary implications of this algal polyphyly are twofold. First, different algal groups can be quite distantly related to plants, to animals, and to one another. Molecular clock estimates put the radiation of the archaeplastid and SAR clades as over one billion years in the past (Dorrell & Smith, 2011; Strassert et al., 2021), and the nucleocytoplasmic biology of these species may be very different from one another. One example are the euglenids, which possess secondary green chloroplasts and stand apart from all other algae within the Excavates (Ebenezer et al., 2019; Vanclová et al., 2017). Euglenids possess a number of features absent or rarely found in other algae, e.g. the absence of otherwise conserved mitochondrial outer membrane transport complexes, and the use of polycistronic transcription and trans-splicing of nuclear transcripts (Cavalier-Smith, 2010). Other groups that are more closely positioned to other algae may also have highly divergent cell biology,

e.g. dinoflagellates within the SAR clade (see chapter 2), which also perform polycistronic nuclear transcription and trans-splicing (Lukes et al., 2009).

The second implication is that photosynthesis has been acquired multiple times across the eukaryotes. Most algae have described non-photosynthetic close relatives, e.g., oomycetes within stramenopiles as a sister to diatoms (Leonard et al., 2018; Lévesque et al., 2010). While there has been some debate how many of these non-photosynthetic sister groups once possessed chloroplasts, (Azuma et al., 2022; Terpis et al., 2024), available genomic data suggest that most never did (see chapter 3) (Stiller et al., 2009; Q. Wang et al., 2017). Thus, most algal groups (stramenopiles, haptophytes, dinoflagellates, cryptomonads) derive from separate endosymbiosis events to one another. These independent chloroplast acquisitions may have been accompanied by different cell biology innovations in each algal group.

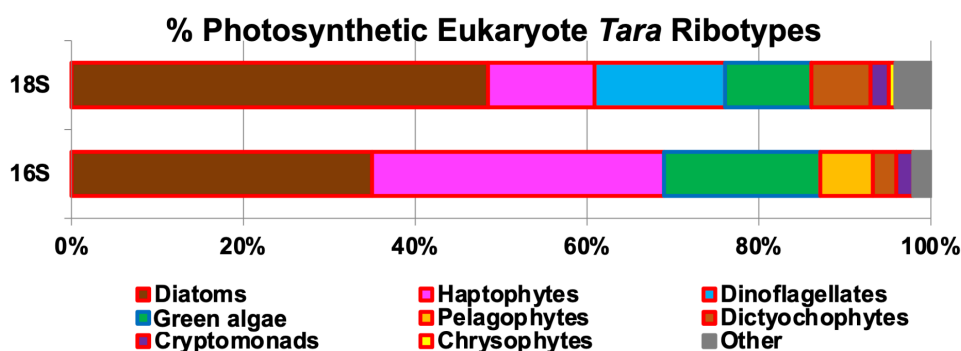


Fig. 1.4. Relative algal abundances in the oceans. These bar charts compare 18S V9 and 16S V4V5 rDNA ribotypes across all *Tara* Oceans stations in surface samples not filtered for size fraction, adapted from (Penot et al., 2022).

The colour of each bar border relates to the chloroplast type shown in Fig. 1.3.

Finally, Fig. 1.4 presents an environmental view of algal diversity, considering mapped 18S V9 and 16S V4V5 chloroplast rDNA sequences from the *Tara* Oceans expedition (Penot et al., 2022; Pierella Karlusich et al., 2023). These data exclude cyanobacteria (i.e., prokaryotic algae), which from marker genes shared with eukaryotes (e.g., *psbO* metaT data, 0.22-3 μ m surface size fractions) probably represent about 25% of total *Tara* algal abundance (Pierella Karlusich et al., 2023). While discrepancies are visible between each marker (e.g., absence of dinoflagellates from 16S, and probable underestimation of haptophytes in 18S), due to sequence divergence (Dorrell, Nisbet, et al., 2019; H. Liu et al., 2009), certain trends are evident in both graphs. Firstly, diatoms are the most abundant algal group, although in 16S data are only fractionally more abundant than haptophytes. Diatoms are differentiated from other eukaryotic algae via larger cell sizes (including colonial and chain-forming species), sessility, and predominant occurrence in nutrient-rich temperate and sub-polar oceans

(Bowler et al., 2010), contrasting to the more global distributions of other orders (H. Liu et al., 2009). Secondly, with the exception of one group (green algae, in the archaeplastids), the most abundant eukaryotic algae (e.g. diatoms, dinoflagellates, haptophytes, pelagophytes, dictyochophytes) possess secondary red chloroplasts. This raises the question of what allows algae with secondary red chloroplasts, or a subset thereof, to be so abundant in the contemporary ocean (Dorrell & Smith, 2011; Nonoyama et al., 2019).

1.3. The endosymbiotic origin and horizontal evolution of eukaryotic algae

Much of the work within my group relates to the cell biology and genetics of eukaryotic algae tree of life, particularly relating to their chloroplasts by endosymbiosis. I present a schematic cell diagram outlining some of these processes in **Fig. 1.5**.

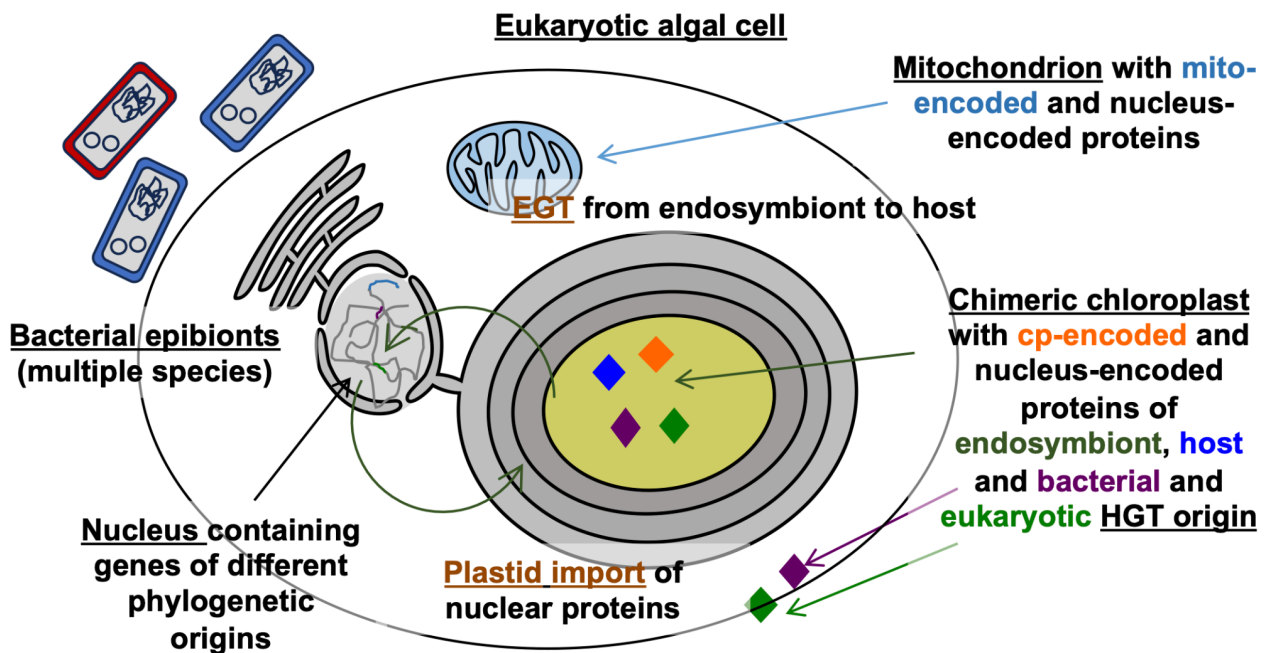


Fig. 1.5. Schematic diagram of chloroplast cell biology. This figure shows a eukaryotic cell, adapted from (Nonoyama et al., 2019), with a nucleus, mitochondrion and secondary chloroplast surrounded by four membranes (the outermost of which is the ER). The nuclear genome of this lineage contains both its endogenous (ancestrally present) genes alongside others acquired by horizontal transfers from endosymbiotic organelles and other sources. The mitochondrion and chloroplast are supported by proteins encoded in their own genome, but also by proteins encoded in the nucleus and imported post-translation.

The endosymbiont hypothesis, as proposed by Mereschkowsky and championed by Lynn Margulis, is that mitochondria and chloroplasts are derived from free-living micro-organisms that were internalised and converted into a stable organelle (Sagan, 1967). This ancestry is evident from the biology of these organelles, e.g. the presence of their own genomes, which

are expressed using bacterial-type (i.e., chloramphenicol- sensitive) ribosomes (Ellis, 1977; Nimmo et al., 2019). The genomes of chloroplasts and mitochondria are much smaller (< 250 genes in the case of chloroplasts) than those of free-living bacteria, suggesting extensive genome reduction (Green, 2011b). This genome reduction may be counterbalanced by infrequent gene gains (e.g. by horizontal transfer), genome fragmentation, and expansions of non-coding DNA (Dorrell et al., 2022; Richardson et al., 2014; van Beveren et al., 2022). Finally, all mitochondria and the chloroplasts of plants, red and green algae in the archaeplastids are surrounded by a double membrane, which is proposed to be derived from the inner and outer envelope of the bacterial cell (Hammond et al., 2022; Huang et al., 2023). This hypothesis is supported by the presence of either a peptidoglycan layer (in glaucophytes) between these two membranes which likely corresponds to the ancestral bacterial cell wall, or a remnant synthesis pathway (in some green algae, and moss) repurposed for plastid division (Matsumoto et al., 2012; Takano & Takechi, 2010).

Experimental proteomics of different plants and algae suggest that between 1000 and 2000 proteins are required for the biogenesis and function of mitochondria and chloroplasts (Barylyuk et al., 2020; Hammond et al., 2022; Huang et al., 2023; Novák Vanclová et al., 2020). The majority of these proteins are encoded in the host nuclear genome, and imported post-translationally across the organelle membranes (Dorrell & Howe, 2012b). For mitochondria and for the double-membrane bound chloroplasts of archaeplastids this is accomplished by a hydrophilic N-terminal targeting sequence, recognised respectively by TIM/ TOM and TIC/ TOC complexes located in the mitochondrial or chloroplast inner or outer membranes. Proteins imported through these complexes are unfolded, cotransported across both membranes, and refolded in the organelle interior following cleavage of the targeting sequence (Jarvis, 2008; Neupert, 1997). A small subset of proteins, e.g. aminoacyl-tRNA synthetases, may be targeted to both organelles using ambiguous targeting sequences that may be recognised by both the TOM and TOC machinery (Duchêne et al., 2005).

The situation is even more complicated for secondary red and green chloroplasts (**Fig. 1.5**). Reflecting their eukaryotic origin, secondary chloroplasts may contain additional associated organelles: a relict nucleus, or “nucleomorph” (Gould et al., 2006; Sarai et al., 2020); and even mitochondria from the endosymbiont (in one group, the dinotoms) (Imanian et al., 2012; Yamada et al., 2017). Secondary chloroplasts are surrounded by additional membranes beyond the two of the archaeplastid chloroplast: this includes a third membrane, presumably

derived from the plasma membrane of the endosymbiont ancestor and encasing a novel organelle derived from the endosymbiont cytoplasm termed the “periplastid compartment” or “PPC” (Flori et al., 2016; Gould et al., 2006). Finally, in all secondary chloroplast lineages except two (peridinin containing dinoflagellates, with chloroplasts of red origin; and euglenids with chloroplasts of green origin) the chloroplast is contained within the endoplasmic reticulum, which forms a fourth, outermost membrane (termed the “chloroplast endoplasmic reticulum” or “cER” (Huang et al., 2023).

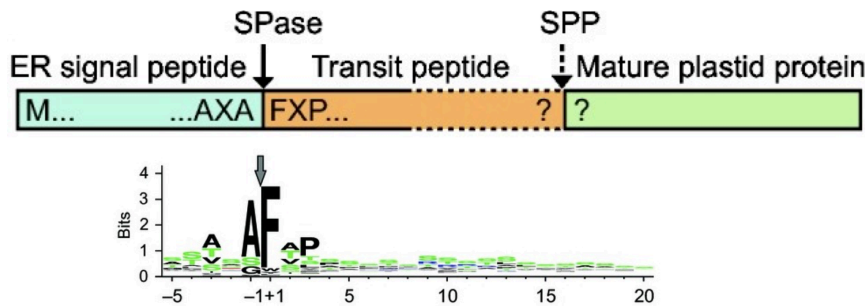


Fig. 1.6. A diatom consensus chloroplast targeting sequence. The signal and transit peptide diagram (top), and bitscore map of the “AXAFAP” cleavage site (bottom) are adapted from (Gruber et al., 2015; Huesgen et al., 2013).

Proteins imported into secondary chloroplasts thus pass through the endoplasmic reticulum first, and are differentiated from proteins imported into primary chloroplasts via the presence of a *hydrophobic* signal peptide at the beginning of their N-terminal targeting sequences (**Fig. 1.6**). This is followed by a specific motif (“AXAFAP”, for example, in diatoms), positioned around the signal peptide cleavage site that allows recognition by a protein complex (SELMA) derived from the ER ubiquitinyl-associated protein degradation machinery that allows import across the third plastid membrane (Gruber et al., 2007; Stork et al., 2012). Finally, proteins inside the PPC may either reside and function in this compartment, as in per some carbon metabolism and sugar transport proteins in diatoms (Moog et al., 2020; Yu et al., 2022), or be imported into the chloroplast via a conventional transit peptide and TIC/ TOC complex. Dual-targeting to secondary chloroplasts and mitochondria is known, dependent either on targeting sequences that are sufficiently ambiguous to be recognised by both ER and mitochondrial protein import machineries, or alternative initiation codons that allow the specific production of each presequence during protein translation (Gile et al., 2015; Hammond et al., 2022).

Finally, the phylogenetic origin of proteins targeted to primary and secondary chloroplasts, or indeed to other organelles in algal cells is heterogeneous, reflecting the global history of the algal host. Many of the proteins imported into chloroplasts are derived from the endosymbiotic ancestor of the chloroplast, encoded by genes that were relocated to the

nuclear genome by endosymbiotic gene transfer (Dorrell & Howe, 2012b). These include most of the proteins that interact with proteins encoded within the chloroplast itself, e.g. extrinsic subunits of photosystems and the chloroplast ribosome (Ishida & Green, 2002; Jarvis, 2008; Pierella Karlusich et al., 2023). Others may derive from the host, e.g. aminoacyl-tRNA synthetases of mitochondrial origin recruited to support the chloroplast (Duchêne et al., 2005; Gile et al., 2015), and RNA processing pathways (Dorrell & Howe, 2012a), or from genes horizontally acquired from bacteria, viruses, or other eukaryotes (Morse et al., 1995; Petersen et al., 2006; K. Tanaka et al., 1996). This may reflect broader histories of horizontal gene transfer into algal and protist genomes, which may remodel the biology of organelles beyond the chloroplast, and confer physiological adaptations to the environment (Cho et al., 2023; Eme et al., 2017; Foflonker et al., 2018; Rius et al., 2023).

The mixed composition of the chloroplast proteome- and more broadly the algal nuclear genome- has been variously termed, a “shopping bag” (Ebenezer et al., 2019; Larkum et al., 2007), “chimeric” (Bodył, 2018; Richtová et al., 2021), and the particularly poetic “red carpet” (Ponce-Toledo et al., 2019). My preferred term is “mosaic”: a complex network of and interacting pieces of different origins; which when viewed from a distance reveals patterns and complexities that are more intricate than its individual components (Dorrell & Howe, 2012b; Green, 2011a; Stairs et al., 2020).

1.4. Outline of work performed in my group

In this *mémoire*, I discuss different projects to understand the mosaic composition of algal genomes from evolutionary, functional, and increasingly environmental perspectives. Much of this work relates specifically to the nucleus-encoded proteomes of algae (particularly dinoflagellates, diatoms, and haptophytes) with chloroplasts of secondary red origin, although over time my initial scope has expanded to consider algae with chloroplasts of different phylogenetic origins, and phenomena of horizontal genome evolution beyond the chloroplast.

Chapter 2 outlines attempts to understand the formation and evolutionary dynamics of the algal chloroplast mosaic, with particular focus on those of dinoflagellates, an algal lineage of special evolutionary significance (Dorrell & Howe, 2015). Beginning with computational reconstructions of dinoflagellate plastid genome evolution (Dorrell, Klinger, et al., 2017; Klinger et al., 2018), we have more recently attempted complete reconstructions of dinoflagellate chloroplast proteomes, particularly of one group, Kareniaceae, which serve as a

model system for understanding chloroplast evolution (Novák Vanclová et al., 2024). These data emphasise the importance of micro-evolutionary variation, down to the species level, in understanding chloroplast functions, and provide evidence for multiple endosymbiotic acquisitions of the Kareniacean chloroplast. In a final project, on a second group of algae, the chrysophytes, we explore how changes to chloroplast proteomes accompany lifestyle transitions (e.g., the loss of photosynthesis), and how the phylogenetic diversity of chloroplasts can be interrogated to identify proteins important to primary metabolism (Dorrell, Azuma, et al., 2019).

Chapter 3 extends on the functional understanding of chloroplast proteomes, focussing on novel chloroplast proteins associated with diatoms, and close relatives within the stramenopiles (Dorrell & Bowler, 2017). Beginning with a phylogenomic interrogation of the timing and composition (in particular the green footprint) of the stramenopile chloroplast (Dorrell, Gile, et al., 2017), we explore what metabolic innovations and transcriptional dynamics characterise the diatom chloroplast, as particularly abundant components of current marine communities (Ait-Mohamed et al., 2020; Nonoyama et al., 2019). Finally, this chapter presents ongoing efforts to characterise novel diatom chloroplast proteins, leveraging meta-genomic data from *Tara* Oceans and CRISPR-mediated mutagenesis in the model diatom species *Phaeodactylum* to explore protein function at both laboratory and environmental levels (Dorrell, Zhang, et al., 2024; S. Liu et al., 2022). These particularly relate to understanding the functions of novel chloroplast carbon metabolism pathways, transporter proteins, and an ATP synthase-related protein found across marine algae but not plants.

Chapter 4 considers broader phenomena of horizontal gene transfer across algal genomes, including genes encoding proteins that function beyond the chloroplast (Dorrell et al., 2022). This begins with the phylogenetic observation that haptophytes, a globally distributed algal order of substantial environmental importance, has acquired much of its nucleus-encoded chloroplast proteome from another algal group, the pelagophytes/ dictyochophytes, in a previously undocumented tertiary or higher endosymbiosis (Dorrell, Gile, et al., 2017; Penot et al., 2022). We then explore the global dynamics and cellular functions of horizontal gene transfer across algal evolution, and in particular the contribution of continuous horizontal gene transfer from bacteria to the predicted secretomes of diatoms (Dorrell et al., 2021). We further consider biogeographical factors that underpin inter-algal horizontal gene transfer, for example detecting within-oceanic gene transfers in the pan-genome of algae indigenous to

the Arctic Ocean (Dorrell et al., 2023). Finally, moving beyond horizontal gene transfer, we explore how taxonomy and biogeography interact to shape algal evolution, in a seminal project evaluating the differential loss and retention of vitamin B₁₂ dependence across algae and land plants (Dorrell, Nef, et al., 2024).

Where possible, I have tried to integrate reflection on the pedagogical components of these projects, and what I have learned in supervising student members of my group. I would like to formally acknowledge here the group members who have contributed to the projects presented: both alumni (Tomomi Nonoyama, Anna Novák Vanclová, Shun Liu) and current personnel (Mathias Penot-Raquin, Louis Graf, Victoria Powell) for their role driving the science forward, and helping me become a more effective, sensitive, and supportive group leader.

Chapter 2: Computational reconstruction of the evolutionary trajectories of dinoflagellate chloroplast proteomes (and beyond)

2.1. Dinoflagellates: an open-ended experiment in chloroplast evolution

Dinoflagellates, to me, have always been very interesting algae. Dinoflagellates have important contributions to global aquatic ecosystems, as the most species-rich algal lineage in *Tara* Oceans data (de Vargas et al., 2015), or as essential symbionts (*Symbiodinium*, *sensu lato*) of coral reefs (Reddy et al., 2023). Dinoflagellate ecosystem contributions extend beyond photosynthesis, e.g. as important predators (*Noctiluca*) (Cooney et al., 2022), pathogens (Sydiniales, Perkinsids) (Gornik et al., 2015) and agents of harmful algal blooms (*Karenia*, *Karlodinium*) (X. Li et al., 2019; Place et al., 2012) (**Fig. 2.1A**). From a taxonomic perspective, dinoflagellates are positioned within the alveolates, as a sister-group to the apicomplexans, a lineage that includes coral symbionts (*Chromera*, *Vitrella*) and pathogens of global relevance (e.g. *Plasmodium*, the causative agent of malaria) (Hicks et al., 2019; Oborník et al., 2012). The global ecology of dinoflagellates and their closest relatives has been effectively synthesised in other reviews (Janouškovec et al., 2012; Le Bescot et al., 2016).

Dinoflagellate genomes, relative to other algae, are highly idiosyncratic. Dinoflagellate nuclear genomes are retained in a permanently condensed state and do not use histones as a core packaging element (Marinov et al., 2021; Nand et al., 2021) (**Fig. 2.1B**). Their mitochondrial genomes are reduced to three essential protein-coding genes (*cob*, *coxI*, *coxII*), which can be encoded on multiple overlapping and partial elements (Nash et al., 2007). The dinoflagellate chloroplast genome is classically viewed to be reduced to just twelve protein-coding genes, rRNAs and tRNAs (Dorrell, Nisbet, et al., 2019; Green, 2011b; Robbens et al., 2007), although a thirteenth dinoflagellate chloroplast gene, coding for *psbT*, has recently been proposed (He et al., 2024) (**Fig. 2.1C**). These genes are coded on unigenic, plasmid-like elements termed “minicircles” (Barbrook & Howe, 2000; Z. Zhang et al., 1999) (**Fig. 2.1D**).

The dinoflagellate nucleus-encoded and chloroplast-targeted proteome is also unusual, using a form II RuBisCo lacking small subunits and acquired from a proteobacterial donor, distinct from the plastid-encoded (in terms of the large subunit) form I RuBisCos shared between plants, most algae and cyanobacteria (Morse et al., 1995; Tabita et al., 2007). Dinoflagellate chloroplast transcripts undergo some essentially unique processing events (**Fig. 2.1E**). These include extensive post-transcriptional substitutional base editing, involving all twelve possible

interconversions and extending to up to a quarter of the residues in some dinoflagellate chloroplast transcripts (Dorrell & Howe, 2015; Zauner et al., 2004); and the addition of a 3' poly(U) tail, which has only been documented in dinoflagellates, the closely related photosynthetic Apicomplexa *Vitrella* and *Chromera*, and the green alga *Pithophora*, where it presumably arose independently (Dorrell, 2020; Meade et al., 2020; Y. Wang & Morse, 2006).

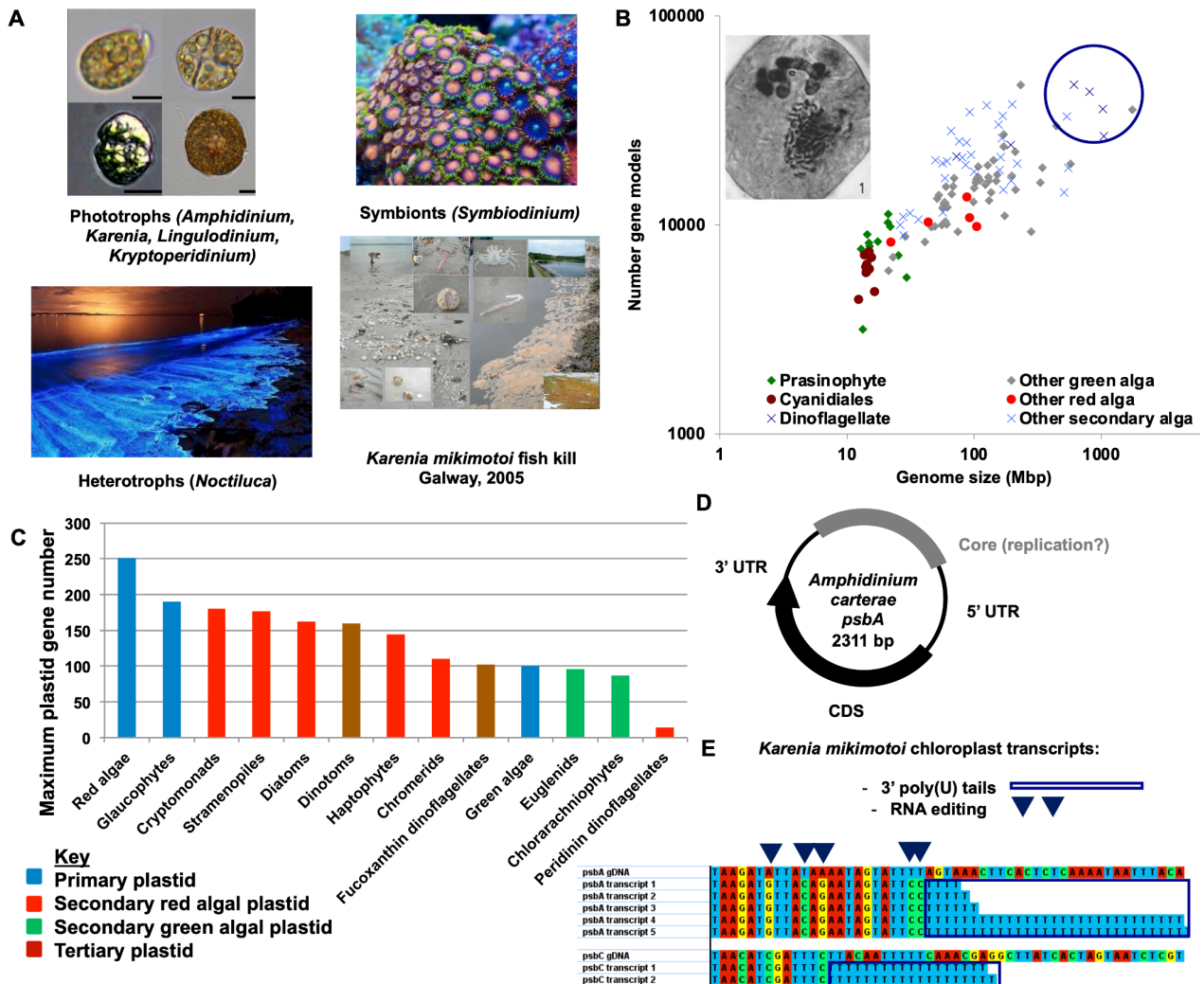


Fig. 2.1. Dinoflagellate ecological diversity and evolutionary idiosyncrasy. **A:** Varied ecosystem services of dinoflagellates, from free-living phototrophs to symbionts, predators, and causative agents of Harmful Algal Blooms. Images taken from (Dorrell & Howe, 2015; Silke et al., 2005). **B:** Graph adapted of algal nuclear genome sizes (horizontal axis) and gene model numbers (vertical axis), using published genome sequences from JGI PhycoCosm (Grigoriev et al., 2021). Both axes are on \log_{10} scales, with dinoflagellates that have far larger genomes than other groups highlighted by a blue circle. *Inset*, TEM image of a *Kryptoperidinium foliaceum* cell (Dodge, 1971), containing a permanently condensed dinoflagellate host nucleus, and conventional diatom endosymbiont nucleus. **C:** data replotted from (Green, 2011b), showing the maximum number of protein-coding genes associated with the chloroplast genomes of different algae, showing peridinin-containing dinoflagellates as a highly reduced outlier. **D:** an exemplar chloroplast minicircle from the model dinoflagellate *Amphidinium carterae*, encoding the D1 subunit of photosystem II (*psbA*) as a single unigenic element. A conserved “core” element shared between all minicircles in a

species, and proposed to be involved in minicircle replication and/ or transcription, is shown in grey. Diagram adapted from (Dorrell, Nisbet, et al., 2019) E: 3' terminal alignment of *Karenia mikimotoi* chloroplast *psbA* and *psbC* transcripts against the corresponding genomic sequences, showing the presence of poly(U) tails (blue cases) and substitutional base editing (inverted triangles). Figure taken from (Dorrell & Howe, 2012a)

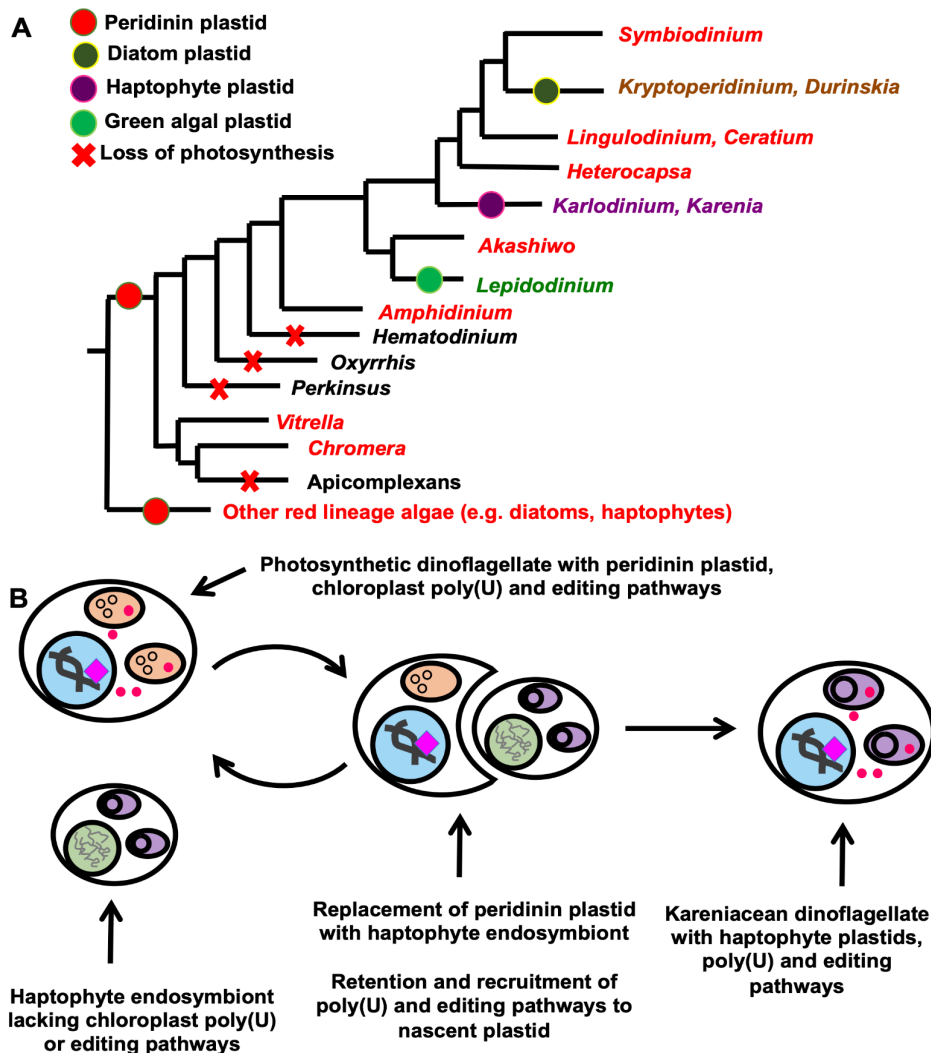


Fig. 2.2. Serial acquisitions of the dinoflagellate chloroplast. A: schematic dinoflagellate tree, showing the relationships between the key nuclear lineages and those of their sister-groups (i.e., *Vitrella*, *Chromera*, and parasitic apicomplexans). Branches are shaded by inferred evolutionary origin (or absence) of a plastid. **B:** cell diagram suggesting a possible model for the origins of the Kareniacean chloroplast, and its associated RNA metabolism, through serial endosymbiosis. Both figures are adapted from (Dorrell & Howe, 2015).

I was most interested in the endosymbiotic promiscuity of dinoflagellates. In addition to possessing a chloroplast of secondary (or higher, see Chapter 4) red endosymbiotic origin, present in most photosynthetic groups and containing the accessory pigment peridinin, some dinoflagellate groups have replaced their chloroplasts (Dorrell & Howe, 2015; Haxo et al., 1976; Novák Vanclová et al., 2024) (**Fig. 2.2A**). These include chloroplasts derived from haptophytes (in the Kareniaceae, i.e. *Karenia*, *Karlodinium*, *Takayama*) (Ishida & Green, 2002), green algae in the distantly-related species *Lingulodinium*, “MGD” and “TGD” (Sarai et al., 2020), and with diatoms, in the “dinotoms” (*Kryptoperidinium*, *Durinskia*) (Imanian et al., 2012), named as they possess in effect a complete diatom endosymbiont cell (including its own

nucleus) within a dinoflagellate host. Phylogenomic data imply many of these endosymbioses have occurred multiple times in different dinoflagellate groups and/ or repeatedly in the same lineage (Novák Vanclová et al., 2024; Sarai et al., 2020; Yamada et al., 2017).

Within my PhD, I explored in particular the evolution of the dinoflagellate chloroplast RNA metabolism from a purely observational level in the ancestral, peridinin-containing chloroplast (Dorrell, Nisbet, et al., 2019), its closest sister-group, the Apicomplexa (Dorrell et al., 2014), and in particular in the serially acquired, haptophyte-derived chloroplasts of the Kareniacean dinoflagellates *Karenia mikimotoi* and *Karlodinium veneficum* (Dorrell et al., 2016; Dorrell & Howe, 2012a; Richardson et al., 2014). This, strikingly, included the observation of both RNA editing and 3' polyU tail addition of Kareniacean chloroplast transcripts, with the most parsimonious explanation for their presence being that they derive from the ancestral peridinin-dinoflagellate chloroplast, and were recruited to the incoming Kareniacean plastid during serial endosymbiosis (**Fig. 2.2B**). This suggests that the serially-acquired dinoflagellate chloroplasts use proteins that originated within the host, and more specifically their plastid predecessors (Dorrell & Howe, 2015). Dinoflagellate chloroplasts, supported by proteins derived from ancient and current endosymbionts, are thus a model system for exploring the mosaic evolution of the chloroplast proteome.

My PhD studies brought me into contact with a succession of highly motivated undergraduate students, one of whom (Elisabeth Richardson) (Richardson et al., 2014) remains a close collaborator in dinoflagellate projects (Dorrell, Klinger, et al., 2017; Klinger et al., 2018) and beyond (Dorrell et al., 2023; Richardson et al., 2015). Both reflecting their special evolutionary importance, and the researchers I interacted with during my doctoral studies, I chose to continue working on dinoflagellates after my arrival in France in a predominantly bioinformatic capacity, firstly working on the evolution of their chloroplast transcriptomes (Dorrell, Klinger, et al., 2017; Klinger et al., 2018), and more recently through a more complete appraisal of the Kareniacean chloroplast proteome, with Anna Novák Vanclová, a post-doctoral researcher recruited to my first CNRS grant (Ait-Mohamed et al., 2020; Dorrell et al., 2023; Novák Vanclová et al., 2024).

2.2. Tracing the divergent evolution of dinoflagellate chloroplast genes through a transcriptomic approach

The first of these projects related to the peridinin-containing dinoflagellate chloroplast, which I had first studied at the beginning of my PhD (Barbrook et al., 2012; Dorrell, Klinger, et al., 2017; Dorrell, Nisbet, et al., 2019). Genes located in peridinin chloroplast genomes are highly divergent, forming extremely long branches in phylogenies and rendering their precise evolutionary origin unclear (Janouskovec et al., 2010; Yoon et al., 2002). At the beginning of my postdoctoral research, only two complete peridinin chloroplast genomes (*Amphidinium carterae*, and the coral endosymbiont *Symbiodinium minutum*) (Dorrell, Nisbet, et al., 2019; Mungpakdee et al., 2014) had been sequenced, although there were minicircle collections sequenced for a handful of other species (e.g., *Heterocapsa triquetra*, *Lingulodinium polyedra*) (M. J. Nelson et al., 2007; Y. Wang & Morse, 2006).

My initial interest was purely to expand on annotated peridinin dinoflagellate chloroplast sequences, via a transcriptomic approach. At the time beginning my post-doctoral research, the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP), a global transcriptome survey of over 600 protists with limited corresponding genomic coverage, had recently been released (Keeling et al., 2014). Anecdotally, I have found that MMETSP libraries have greater depth and transcript completeness than more recent and larger-scale algal transcript sequencing (e.g., OneKp) (One Thousand Plant Transcriptomes Initiative, 2019) and even draft genome collections (D. R. Nelson et al., 2021). MMETSP incorporates transcriptomes for 45 photosynthetic dinoflagellates (**Table 2.1**), important as the size and complexity of their genomes hinders full-scale sequencing (Grigoriev et al., 2021; Janouškovec et al., 2017).

Peridinin dinoflagellates		Known minicircle sequences		Kareniacean dinoflagellates		Known plastid transcripts	
<i>Amphidinium carterae</i>	CCAP1102/6	12					
<i>Symbiodinium</i> sp. Mf.1.05b		12		<i>Karenia brevis</i>	CCMP2281	36	
<i>Heterocapsa triquetra</i>		10		<i>Karenia mikimotoi</i>	RCC1513	65	
<i>Lingulodinium polyedrum</i>	CCMP1738	9		<i>Karlodinium veneficum</i>	UIO083	73	
<i>Symbiodinium</i> sp. C3		9					
<i>Amphidinium carterae</i>	CS21	6					
<i>Ceratium homidum</i>		6					
<i>Akashiwo sanguinea</i>	CCMP1321	5					
<i>Heterocapsa niei</i>	CCMP447	5					
Peridinin dinoflagellate MMETSP libraries		Number of species		Kareniacean dinoflagellate MMETSP libraries		Number of transcripts annotated	
12 genes annotated		10		<i>Karenia brevis</i>	CCMP2229	58	
11 genes annotated		10		<i>Karenia brevis</i>	SP1	72	
10 genes annotated		8		<i>Karenia brevis</i>	SP3	48	
9 genes annotated		4		<i>Karenia brevis</i>	Wilson	71	
				<i>Karlodinium veneficum</i>	CCMP2283	46	

Table 2.1. Overview of peridinin-containing (left) and Kareniacean (right) dinoflagellate chloroplast coding sequences either previously sequenced (top) or assembled de novo from MMETSP transcriptomes (bottom).

This table synthesises data from (Dorrell et al., 2016; Dorrell, Klinger, et al., 2017; Gabrielsen et al., 2011; Mungpakdee et al., 2014; Novák Vanclová et al., 2024). Peridinin dinoflagellate MMETSP libraries are shown by the *total* number of species for which different numbers of chloroplast genes have been annotated.

While RNA sequencing pipelines frequently eliminate organelle-encoded mRNAs, e.g. through poly(A) selection (Small et al., 2013; D. R. Smith & Sanitá Lima, 2017), they may be more successful in recovering peridinin dinoflagellate chloroplast transcripts. This relates to the 3' poly(U) tail added to dinoflagellate chloroplast transcripts, which was first detected amongst poly(A)-enriched mRNA (Dorrell, 2020; Y. Wang & Morse, 2006). Polyuridylylated transcripts have been proposed to dimerise with nuclear polyadenylylated transcripts *in vitro* (Janouskovec et al., 2010; Y. Wang & Morse, 2006), allowing their recovery from poly(A)-enriched fractions, although this hypothesis to my knowledge remains unproven. Nonetheless, MMETSP appears to be a rich source of dinoflagellate chloroplast transcripts, and I could annotate plastid transcriptomes for 32 peridinin-containing dinoflagellate species (**Table 2.1**) (Dorrell, Klinger, et al., 2017). For a subsequent project (see section 2.4), we extended this approach to recover plastid-encoded transcripts from Kareniacean MMETSP libraries, which likewise contain 3' poly(U) tails (**Table 2.1**) (Novák Vanclová et al., 2024).

Next, I considered the evolution of the peridinin plastid genome. Dinoflagellate mRNA sequences, having undergone editing events necessary for the correct expression of plastid proteins, may be more of the peridinin plastid proteome than the underlying genomic DNA sequences (Dang & Green, 2009; Richardson et al., 2014). For this work, I assembled a consortium of largely post-graduate collaborators, some colleagues from my PhD (Elisabeth Richardson, Erin Butterfield) (Dorrell et al., 2013; Richardson et al., 2014) and others new (Robert Newby- MTSU; Christen Klinger- University of Alberta) (Dorrell, Klinger, et al., 2017). I am proud that the resultant publication, and its successor (Klinger et al., 2018), were essentially designed and performed entirely by early-career researchers, working collaboratively, in some sense as an “evening lab”, alongside our day-to-day projects.

The results from this survey suggest specific and sustained divergent sequence evolution in the peridinin dinoflagellate chloroplast. Using a concatenated chloroplast gene tree from Christen Klinger, we demonstrated that dinoflagellates are differentiated from other algae by elevated chloroplast mutation rates, and divergent selection (i.e., elevated K_a/K_s ; **Fig 2.3A**). This divergence was evident even compared to close relatives of the dinoflagellates (e.g., the photosynthetic apicomplexan *Vitrella*) (Janouskovec et al., 2010; Oborník et al., 2012), but was equally true between dinoflagellate species, with extremely high pairwise K_a/K_s ratios observed for example between members of the core dinoflagellate orders Gonyaulacales and Suessiales (Pochon et al., 2006), despite relatively low numbers of total pairwise substitutions.

The implications are that the peridinin chloroplast is not only very fast-evolving relative to other plastid groups, but there may be very high internal variation in peridinin chloroplast coding content, consistent with diversifying selection. This observation was further supported by tabulating alternative initiation codons (Robert Newby), indels (Christen Klinger) and loss of otherwise universally conserved atpA residues (Erin Butterfield), all of which suggested that the majority of this divergent evolution occurred within peridinin dinoflagellates rather than in its common ancestor, and much of it is species-specific (**Fig. 2.3B**).

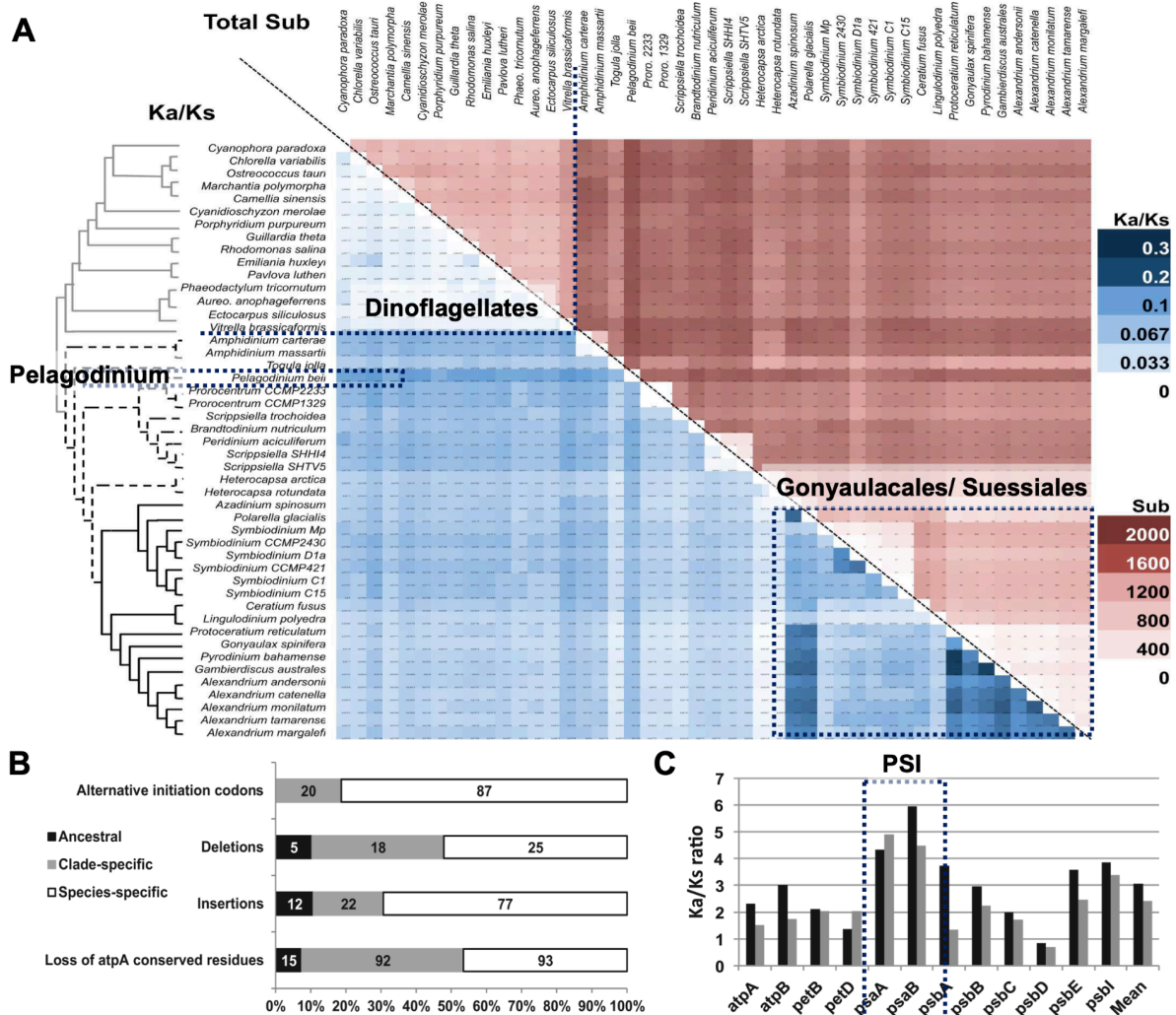


Fig 2.3. Divergent trajectories of peridinin chloroplast genomes. A: heatmap of pairwise total substitutions (horizontal axis) and Ka/ Ks ratios (vertical axis) visualised over a 48 taxa x 4,478 nt gap-free concatenated alignment of six chloroplast genes (*psaA*, *psaB*, *psbA*, *psbC*, *psbD*, *psbE*) shared between dinoflagellates and related algae, reconciled to a consensus Bayesian topology inferred from a 3,410 aa concatenated alignment of protein translations. Three taxonomic groups with elevated Ka/ Ks (dinoflagellates compared to non-dinoflagellates, and *Pelagodinium beii* and Gonyaulacales/ Suessiales compared to other dinoflagellates) are shown by hatched blue lines. **B:** occurrences of non-ATG translation initiation codons, indels, and loss of otherwise conserved atpA residues across all peridinin chloroplast transcripts, reconciled to the plastid tree topology via a last-common

ancestor approach. Most unusual dinoflagellate chloroplast features are species-specific, as opposed to ancestral. **C:** Ka/ Ks ratios calculated using PAML for different plastid genes in the branch leading to the dinoflagellate common ancestor (black bars), or for individual dinoflagellate groups (grey bars), divided by the Ka/Ks ratios calculated for all non-dinoflagellate branches within the tree. Elevated relative Ka/ Ks ratios observed for the genes *psaA* and *psaB*, encoding the A1 and A2 subunits of photosystem I, are highlighted with a blue hatched box.

Next, we considered what dinoflagellate species and chloroplast genes show the most divergent gene evolution. Beyond the elevated Ka/ Ks ratios established for the Suessiales including the coral symbiont *Symbiodinium*, we observed rapid divergence in the Prorocentrales *Brandtodinium nutricula* and *Pelagodinium beii*, both of which are endosymbionts of protists (Graeff & Leblond, 2023; Probert et al., 2014; Siano et al., 2010). Subsequent studies have highlighted that many dinoflagellates and dinoflagellate relatives have symbiotic lifestyles (Kwong et al., 2019; Lajeunesse et al., 2018; Mathur et al., 2018), and I wonder whether the elevated Ka/ Ks ratios observed in the dinoflagellate common ancestor might suggest it is likewise an ancestrally symbiotic lineage. Alongside this, with the aid of Elisabeth Richardson, we noted that the elevated mutation rates and Ka/ Ks ratios associated with peridinin chloroplast genes are concentrated on regions coding for the stromal faces of the core photosystem I (PSI) subunits *psaA* and *psaB* (**Fig 2.3**). Other groups have also detected elevated mutation rates in dinoflagellate PSI (Pochon et al., 2006), alongside unusual structural organisations such as supercomplexes with large numbers of light-harvesting pigments and extensive photoprotection capacities (Aihara et al., 2016; Kato et al., 2020). Tentatively, I would suggest that the elevated Ka/ Ks in dinoflagellate *psaA* and *psaB* may relate to changes to PSI structure specific to this lineage. There is also a non-photosynthetic dinoflagellate group (Warnowiids) (Cooney et al., 2023) that has been recently described to retain and express genes for PSI, but not photosystem II, as part of light-sensing eyespots, and it will be interesting to determine how the PSI complex has evolved in dinoflagellates where it is divested of photosynthetic responsibility.

2.3. Corrective- and diversifying- functions of chloroplast RNA editing in

Karenia

Next, my “evening lab” performed a reappraisal of the serially acquired chloroplast genomes of Karenia (Klinger et al., 2018). As described above, Karenia, typified by the harmful bloom-forming genera *Karenia* and *Karlodinium*, have replaced the ancestral peridinin chloroplast with one from haptophytes, but utilise chloroplast RNA editing and poly(U) tail

addition probably derived from the preceding peridinin plastid (**Fig. 2.2**). During my PhD, alongside documenting these pathways, I had performed preliminary evaluations of the functional roles of Kareniacean RNA editing with Elisabeth Richardson. We, alongside others, showed that RNA editing predominantly has a *corrective* role, removing genomic mutations such as in-frame termination codons that would otherwise prove deleterious (C. J. Jackson et al., 2013; Richardson et al., 2014). We also, however, found that some editing events could have *diversifying* impacts on translated plastid sequences, for example the edited removal of a stop codon creating a translatable 3' *atpA* mRNA extension with no apparent homology to other algal sequences (Richardson et al., 2014).

Following from these initial observations, we sought to explore the global evolutionary impacts of RNA editing in peridinin-containing dinoflagellates and Kareniaceae: does it correct for divergent mutations in dinoflagellate plastid genomes, or does it have a role in driving the divergent evolution of their plastid proteomes? This was explored by determining the complete chloroplast editotype (i.e., combined genome and transcriptome) of the peridinin dinoflagellate *Pyrocystis lunula*, with Robert Newby and colleagues at MTSU. Next, we compared the impacts of plastid RNA editing on transcript sequence homology across peridinin-containing (*Symbiodinium minutum*) and Kareniacean species (*Karenia mikimotoi*, *Karlodinium veneficum*) (Dorrell et al., 2016; Mungpakdee et al., 2014; Richardson et al., 2014). This project was conceived (and largely performed) by Elisabeth Richardson and Christen Klinger, with my role being principally that of a senior author and scientific adviser. A very pleasant surprise was the participation of Lucas Paoli, a Normalien who had previously performed his L3 internship with me (see section 2.5) during an M1 placement with Elisabeth Richardson at the University of Alberta (Dorrell, Azuma, et al., 2019; Klinger et al., 2018). This was my first experience of my students working together, creating their own research ecosystem and defining new ideas beyond my direct supervision.

The resulting work, predominantly performed by Elisabeth, Christen, Robert and Lucas, demonstrated that RNA editing in both peridinin and Kareniacean chloroplasts occurs in non-uniform clusters centred over highly divergent sequence. It has a predominantly corrective role, acting to increase the sequence similarity between dinoflagellate chloroplast mRNAs and those of other algal groups compared to that observed at the level of DNA (Klinger et al., 2018). I nonetheless note one observation from this study, as it has served as a significant source of inspiration for my subsequent research. **Fig 2. 4** shows an aggregate plot

of the impacts of RNA editing on chloroplast sequence homology of seven genes (*atpA*, *atpB*, *psbA*, *psbB*, *psbC*, *psbD*, *psbE*) assessed over two Kareniacean and two peridinin-containing dinoflagellates. This is determined by the aggregate difference in the sequence identity of mRNA (edited) versus gDNA (unedited) sequence for each species, relative to five references (*Amphidinium carterae*, *Chrysochromulina tobin*, *Emiliana huxleyi*, *Phaeodactylum tricornutum*, *Vitrella brassicaformis*) that lack chloroplast RNA editing. Values further away from zero indicate more consistent increases or decreases in homology across different reference organisms. I have marked the zero line between net corrective and net diversifying functions.

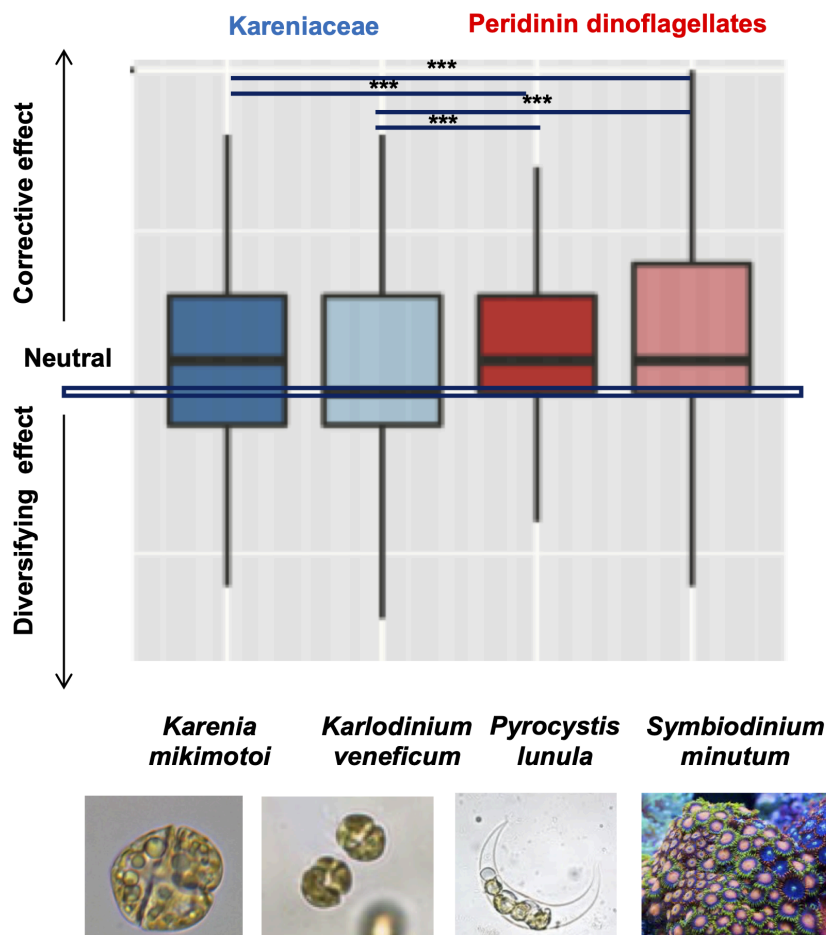


Fig. 2.4. Corrective versus diversifying functions of dinoflagellate chloroplast RNA editing. Boxplots of the average consequences of RNA editing on chloroplast sequence homology over seven chloroplast genes from four dinoflagellates, compared to five reference organisms. The dark blue case indicates the zero line between editing events inferred to increase mRNA sequence homology to reference organisms (i.e., with corrective effects), versus events that decrease mRNA sequence homology (i.e. with diversifying effects). *** indicates a significant difference in average editing effect between species, one-way Tukey's test, $P < 0.001$.

While the majority of editing events in all four species have inferred corrective roles, increasing aggregate sequence identity to the reference organisms, some do not, and may contribute to a diversification of plastid transcript sequence compared to the gDNA template. Editing events that have projected diversifying impacts occur significantly more frequently in the Kareniaceae (48/ 141- 34% considered events in *Karenia mikimotoi*, 87/ 188- 46% in *Karlodinium veneficum*) than in peridinin-containing species (145/ 503 - 29% in *Pyrocystis lunula*, 82/ 299- 27% in *Symbiodinium minutum*; Tukey's test, $P < 0.001$ for all pairwise species

comparisons). Thus, we might propose that RNA editing functions slightly differently in Kareniacean chloroplasts to in their peridinin-containing predecessors. While in both species, it appears to mainly correct DNA mutations that may have deleterious impacts on protein function, there is at least some evidence that in Kareniaceae it is used to create novel protein residues not encoded in genomic DNA at a purely mRNA level. This may serve to diverge these protein sequences from those of their free-living haptophyte ancestors.

This is a very tenuous interpretation of a complex dataset, requiring assumptions about the relationship between mutations in genomic DNA and the evolution of RNA editing sites, and the structure and interactions of chloroplast-encoded proteins. Nonetheless, it prompted me to reflect on what Kareniacean chloroplast proteomes look like globally. Are they more similar to those of their free-living haptophyte ancestors, those of their peridinin-containing dinoflagellate hosts, or something else entirely?

2.4. A holistic appraisal of the Kareniacean chloroplast proteome reveals an intricate mosaic- and repeated endosymbioses

The final project from my group on Kareniaceae, like their chloroplasts, passed through several hands and transformed into something beyond its ancestral form. The project in its initial form, for which I had obtained a Wenner-Gren fellowship to join Fabien Burki's group at Uppsala University (Burki et al., 2014; Strasser et al., 2021), was to build a *pan*-transcriptome for Kareniacean dinoflagellates to understand the holistic evolution of their plastids. While I declined this fellowship, I chose to keep the Kareniacean evolution story alive with a supervised Masters student, Fuhai Liu, in 2019 (Dorrell et al., 2022; Novák Vanclová et al., 2024). Fuhai developed preliminary versions of several of the informatic tools described below, particularly a Kareniacean chloroplast targeting predictor using MMETSP data.

The Kareniacean project however was transformed by the arrival of a post-doctoral researcher, Anna Novák Vanclová, in my group in January 2020. I first encountered Anna as an unusually thorough and creative PhD student in Vladimír Hampl's group at Charles University, working on the freshwater secondary green alga *Euglena gracilis* (Novák Vanclová et al., 2020; Vanclová et al., 2017). The biology of euglenids in many mirrors that of dinoflagellates, with highly expanded and fragmented nuclear genomes (Ebenezer et al., 2019), three membrane-bound chloroplasts structurally similar to the peridinin plastid (Vanclová et al., 2017), trophic flexibility and propensity to loss of photosynthesis (Záhonová et al., 2018), and

an ongoing intractability to experimental manipulation (Faktorová et al., 2020). Anna's PhD focussed on the mosaic evolution of the euglenid chloroplast proteome, which integrates proteins from the green algal symbiont, euglenid (excavate) host and other sources entirely (e.g., red algae and their evolutionary derivatives) (Ebenezer et al., 2019; Novák Vanclová et al., 2020; Ponce-Toledo et al., 2019). Anna initially joined my group to perform experimental proteomics of the model peridinin dinoflagellate *Amphidinium carterae*, in particular incorporating next-generation techniques (e.g., LOPIT) to resolve its plastid proteome (Barylyuk et al., 2020; Mulvey et al., 2017). In the context of the unfolding COVID pandemic, this initial project was simply not technically possible in our laboratory. Instead, Anna reoriented as a capable and thoughtful bioinformatician, contributing to projects in my group understanding the diatom chloroplast proteome and the Arctic algal pan-genome (Ait-Mohamed et al., 2020; Dorrell et al., 2023) discussed in Chapters 3 and 4 respectively.

The finalis project was to assemble transcriptomes of four new Kareniaceae: *Karenia papilionacea*, a toxic and relatively distant relative of the previously studied *Karenia brevis* and *mikimotoi* (Fowler et al., 2015); *Karlodinium armiger*, a relative of *Karlodinium veneficum* with a strong measured preference for mixotrophy over phototrophy (Berge et al., 2008); *Karlodinium* sp. RCC3446/ UIO097, a different strain of the previously sequenced *Karlodinium veneficum* UIO083/ CCMP2283 (Gabrielsen et al., 2011; Keeling et al., 2014); and *Takayama helix*. This latter species is particularly interesting as being to my knowledge the first sequenced member of the genus *Takayama*, a large and dinovorous (i.e., predator of other dinoflagellates) sister-group to the much smaller and bacterivorous *Karlodinium* (De salas et al., 2003; Jeong et al., 2016). The working dataset was completed with a *Karenia mikimotoi* transcriptome completed during my doctoral studies, MMETSP *Karenia brevis* and *Karlodinium veneficum* libraries, and a kleptoplastic species, the "Ross Sea Dinoflagellate" (RSD) (Dorrell et al., 2016; Hehenberger et al., 2019; Keeling et al., 2014).

The second stage of the project was to build a detailed *in silico* portrait of the pan-Kareniacean nucleus-encoded plastid proteome. This involved developing a targeting predictor able to sensitively and specifically identify Kareniacean chloroplast-targeting sequences, which are differentiated from those associated with other secondary red chloroplasts algae via greater length and more complex hydrophobicity (Patron & Waller, 2007; Yokoyama et al., 2011). For this, Fuhai tabulated proteins either experimentally localised or expected to function uniquely in single organelles (the chloroplasts, mitochondria, nucleus, and endomembrane systems),

and Anna constructed targeting motifs that differentiated the chloroplast-targeted proteins from those of other cellular compartments. We were particularly grateful to avail of preliminary data from collaborators, particularly Erin Butterfield for the dinoflagellate mitochondrion (Butterfield et al., 2013, 2016) and Bridget Bannerman for the dinoflagellate nuclear proteome (Bannerman et al., 2018). The retrieved Kareniacean chloroplast targeting motif is distinct from the “ASAFAP” motif associated with diatom protein import (**Fig. 2.5**), containing a post signal-peptide cleavage diarginine motif which had previously been noted by other studies (Gruber et al., 2007, 2015; Patron & Waller, 2007). The signal peptides of Kareniacean plastid-targeted proteins also contains a dicysteine motif associated with plastid- and ER-targeted proteins in the apicomplexan *Toxoplasma* (Barylyuk et al., 2020). A probable structural similarity between the host (dinoflagellate/ apicomplexan) and Kareniacean chloroplast protein import machinery has also been noted in previous studies (Hehenberger et al., 2019; Yokoyama et al., 2011).

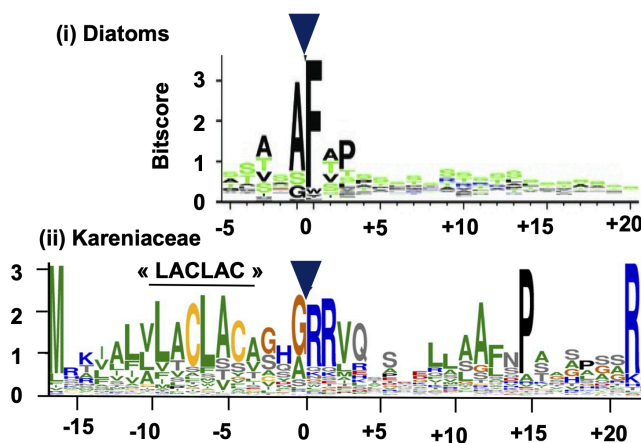


Fig. 2.5. Comparison of chloroplast-targeting presequences established for diatoms and Kareniaceae. The diatom sequence (top) is taken from (Gruber et al., 2015). Sequence consensus bitscore maps are centred around the signal peptide cleavage site (dark blue triangle), with residue numbers given relative to the cleavage site. A consensus “LACLAC” motif that precedes the signal peptide cleavage site in Kareniaceae is underlined.

The final component of this study, designed and conducted by Anna, was to sort the Kareniacean plastid proteome, taken as the assembled transcripts inferred to possess chloroplast-targeting sequences, into different evolutionary categories. Anna considered which proteins derive from the (haptophyte) endosymbiont, the (dinoflagellate) host including the preceding peridinin plastid, horizontal gene transfer from sources such as green algae and bacteria, or have no definable origin (**Fig. 2.6**). Previous studies (Burki et al., 2014; Hehenberger et al., 2019; Matsuo & Inagaki, 2018) had performed a preliminary annotation of parts of the Kareniacean chloroplast proteome, particularly relating to specific core metabolic pathways (e.g., haem and chlorophyll biosynthesis). Anna’s approach was differentiated by its ambition and sheer sense of scale, as reflecting the expanded nature of their nuclear genomes, a total Kareniacean chloroplast proteome may incorporate thousands

of potential non-redundant proteins (Grigoriev et al., 2021). Anna designed a scripted pipeline for building single-gene trees, annotating and sorting the > 30, 000 proteins predicted as supporting a Kareniacean plastid in her dataset by phylogenetic origin (Klinger et al., 2018; Novák Vanclová et al., 2024). This approach also allowed a very specific subtlety to the phylogenetic annotation. By considering each query protein as independent to one another, Anna was able to detect species-specific variations in the evolutionary origin of functionally homologous proteins, as opposed to viewing all Kareniacean dinoflagellates as having entirely the same plastids, with homologous proteins of common evolutionary origin.

These combined approaches (**Fig 2.6**) revealed a highly complex mosaic plastid proteome. The Kareniacean plastid is predominantly of host (dinoflagellate origin), with the haptophyte endosymbiont contributing < 15% of the plastid proteins of all *Karenia*, *Karlodinium*, and *Takayama* species. At the same time, the Kareniacean proteome shows a clear functional structure, with certain metabolic processes (e.g., photosynthesis and translation) deriving predominantly from the endosymbiont, while others (most organic metabolism pathways) derive from the host. This picture is rendered more complex by protein-specific exceptions, including sporadic horizontal acquisitions of genes encoding chloroplast-targeted proteins from green algae, ochrophytes and bacteria (Novák Vanclová et al., 2024). What is perhaps most interesting, from an evolutionary standpoint, is the variation in proteome composition between different Kareniacean species. These included multiple examples of plastid-targeted proteins (e.g., phosphoglycerate kinase, **Fig. 2.6**) where *Karlodinium* and *Takayama* use proteins of different origins to those found in *Karenia*, suggesting that the phylogenetic composition of Kareniacean plastid proteomes vary even over micro-evolutionary scales.

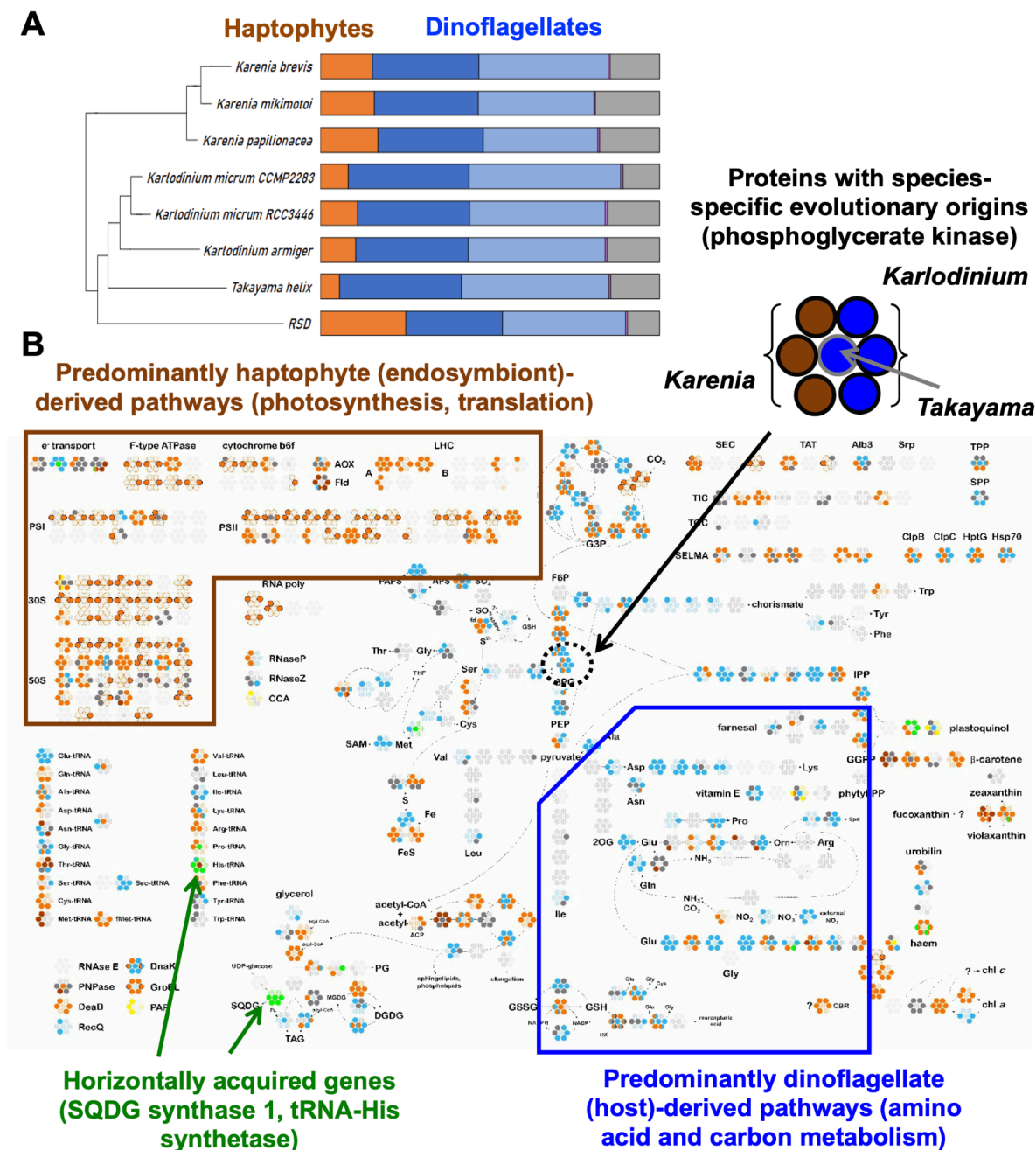


Fig. 2.6. Compositional diversity of the Kareniacean plastid proteome. A: phylogram of eight Kareniacean species, showing evolutionary relationships obtained by a concatenated PhyloFisher topology of 72, 162 aa from 241 well-conserved nuclear genes, and bar charts showing the evolutionary origin identified from single-gene trees of 29, 329 non-redundant plastid-targeted proteins. Proteins for which no non-Kareniacean homologues were identified are excluded. The majority of the signal in each species are for plastid-targeted proteins of dinoflagellate (host) origin. **B:** pathway diagram of core Kareniacean plastid metabolic processes, adapted from (Ait-Mohamed et al., 2020). Individual clusters of seven circles indicate one discrete plastid protein. Each circle within each cluster relates to the homologues identified from a specific Kareniacean species, and each circle is shaded by inferred phylogenetic origin. Callouts show exemplar proteins of predominantly haptophyte origin (brown), dinoflagellate origin (blue), horizontally acquired from green algae (green), and one specific protein (phosphoglycerate kinase) where *Karlodinium* and *Takayama* uniquely possess a dinoflagellate-derived plastid-targeted isoform, whereas *Karenia* possesses a second isoform derived from the haptophyte endosymbiont.

The more surprising finding, however, came from the plastid genomes of these species. Previously, Fuhai and I had assembled partial plastid transcriptomes for the MMETSP *Karenia brevis* and *Karlodinium veneficum* libraries (**Table 2.1**), which supported previous evidence for a phylogenetic origin of their Kareniacean chloroplast deep in the haptophytes, as a sister-group to the class Prymnesiophyceae (Choi et al., 2017; Kawachi et al., 2021; Takishita et al., 1999). The initial reviews of Anna's paper asked the question- given the diverse content of the *pan*-Kareniacean plastid proteome, how divergent are the plastid genomes of this lineage? Anna built a concatenated plastid multigene tree (**Fig 2.7**), using new transcript sequences amplified by me, revealing a surprising story of plastid replacements. Most Kareniacean plastids, including all members of the genus *Karenia* and *Karlodinium veneficum* group together, per previous topologies, as a sister of all Prymnesiophyceae, although with uncertain placement relative to the early-branching Rappephyceae and DPL lineages (Choi et al., 2017). *Takayama helix* and *Karlodinium armiger* however, group in very distant locations in the haptophyte tree to other Kareniaceae and indeed to one another, respectively with members of the orders Phaeocystales and Prymnesiales (**Fig. 2.7**).

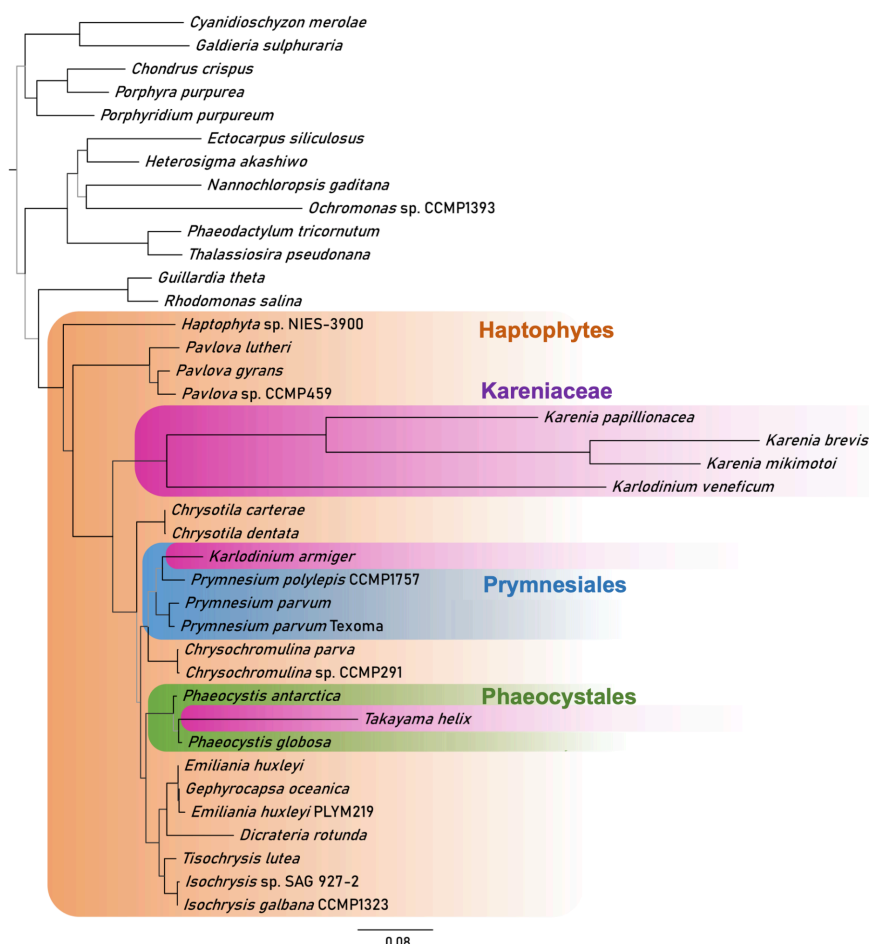


Fig. 2.7. Concatenated multigene topology of Kareniacean plastid genomes. The tree shown is reconstructed by IQ-TREE based on a 39 taxa x 2404 amino acid alignment of five plastid-encoded proteins (PsaA, PsbA, PsbC, PsbD, RbcL), including six Kareniaceae species. Bootstrap support is expressed by branch colour (black for $\geq 90\%$, dark grey for $\geq 75\%$, light grey for $< 75\%$). Kareniaceae are shaded in pink; Prymnesiales in blue; Phaeocystales in green; and all other haptophytes are shaded in orange.

These data, combined with the species-specific variations in plastid proteome content, suggest repeated acquisitions of haptophytes by the Kareniaceae. Anna further reasoned that- given the relatively distant placement of *Karenia* from *Karlodinium veneficum* - their plastids are likely to have also been independently acquired, potentially by a dinoflagellate-to-dinoflagellate transfer (Novák Vanclová et al., 2024). This story raises many questions, particularly the correlation between trophism and plastid establishment, as Kareniaceae are typically considered to be bacterivorous (Jeong et al., 2016, 2021), *Karenia mikimotoi* has been observed to feed on the haptophyte *Isochrysis* (Q.-C. Zhang et al., 2013). Anna also considered the spatial distribution of Kareniaceae in global ecosystems data from *Tara* Oceans with Charlotte Nef at the IBENS, and found little evidence that Kareniaceae and haptophytes co-occur (Lima-Mendez et al., 2015; Nef et al., 2022). I am therefore uncertain why Kareniaceae frequently acquire haptophyte endosymbionts, if they rarely or never encounter them as prey. From a cellular perspective, the frequent acquisition of haptophyte chloroplasts by the Kareniaceae, with different nucleus-encoded proteomes assembled in each case, may provide us with an appealing system for exploring the diverse trajectories that eukaryotes undergo as they gain the capacity for photosynthesis (Hehenberger et al., 2019).

2.5. Independent and progressive reductions of the chloroplast proteomes of chrysophyte algae

While the majority of my work on chloroplast proteomes has focused on endosymbiotic acquisition, particularly within the dinoflagellates, I have periodically been able to explore the consequences of chloroplast reduction, and the loss of photosynthesis. Secondary losses of photosynthesis are known from across the tree of life, yielding multiple non-photosynthetic plant and algal groups (Cooney et al., 2023; Hadariová et al., 2018; Molina et al., 2014; D. R. Smith & Lee, 2014). The most famous of these is the malaria parasite *Plasmodium*, in the dinoflagellate sister-group Apicomplexa (Dorrell et al., 2014; Hicks et al., 2019; Oborník et al., 2012). These “non-photosynthetic algae” often retain plastid-derived organelles (although exceptions are known, c.f. (Gornik et al., 2015)), with associated and essential metabolic functions that do not include photosynthesis (Lim & McFadden, 2010). Studying what proteins are retained or lost from secondarily non-photosynthetic plastids may give us a phylogenetic window into photosynthetic function, by allowing us to identify novel proteins whose presence is associated with photosynthesis. In my PhD, I studied the evolution of apicomplexan plastid RNA processing, showing the loss of pathways (e.g., 3' poly(U) tail addition) associated with its

photosynthetic relatives *Chromera* and *Vitrella* (Dorrell et al., 2014; Janouskovec et al., 2010), which I proposed to be involved in the specific maturation of transcripts of plastid-encoded photosystem subunits.

I have since occasionally revisited the question of photosynthesis loss, in particular in one project focussing on chrysophyte algae (Dorrell, Azuma, et al., 2019). Chrysophytes, or “golden algae”, have acquired their chloroplasts independently to the dinoflagellates, positioned instead within the stramenopiles as closer (but still comparatively distant) to diatoms (Andersen, 2004; Dorrell & Bowler, 2017). Chrysophytes, unusually for algae, are found with almost equal propensity in freshwater (e.g., *Synura*) and marine habitats, with high inferred rates of migration projected between the two from meta-genomics (Jamy et al., 2022; Singer et al., 2021). The environmental ubiquity of chrysophytes extends even further, to polar and continental glacier ecosystems, and marine deep-water and sediments (Gérikas Ribeiro et al., 2020; Gooday et al., 2020; Seeleuthner et al., 2018).

Chrysophytes may allow us to explore at a taxonomically confined level some of the broader principles in chloroplast evolution. Like dinoflagellates, they have rapidly-evolving plastid genomes, with some studies placing the origin of the apicomplexan plastid within or near them (Ševčíková et al., 2015). They are also, like dinoflagellates, trophically flexible, with some genera (e.g., *Synura*, and the paraphyletic *Ochromonas*) possessing yellow chloroplasts capable of photosynthesis, and others (*Paraphysomonas*, and the paraphyletic *Spumella*) not, and reliant exclusively on bacterivory (Grossmann et al., 2016). These boundaries are blurred by the preference of many photosynthetic species (e.g., most *Ochromonas*) for mixotrophy, and the presence of pigments and even thylakoids in some non-photosynthetic species (e.g., *Cornospumella fuschlensis*) (Boenigk et al., 2023; Lie et al., 2018).

The loss of photosynthetic capacity has occurred multiple times across chrysophytes. For example, in one single chrysophyte 18S rDNA tree, constructed with the aid of Louis Graf who has since joined my group as a postdoc, we were able to identify at least eighteen independently evolved non-photosynthetic clades (**Fig. 2.8**) (Andersen et al., 2017; Dorrell, Azuma, et al., 2019). At the time, this was greater than the number of independent photosynthesis losses (14) documented in plants (Hadariová et al., 2018; Krause, 2011). Chrysophytes are thus a particularly good model for understanding the different ways in which photosynthetic species have become dependent on heterotrophy, as the same process has occurred with multiple independent biological replicates across the same order.

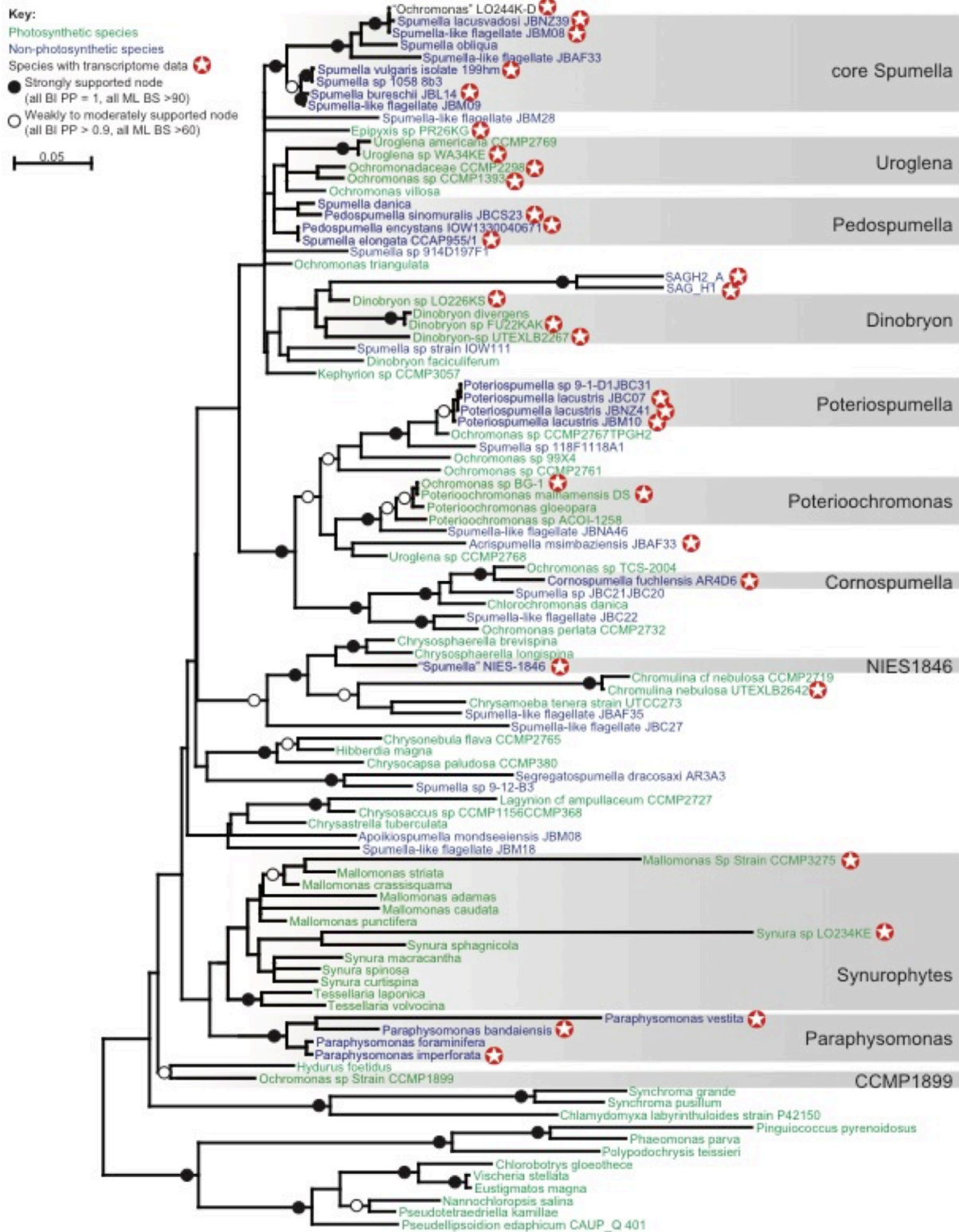


Fig. 2.8. Multiple losses of photosynthesis across chrysophytes. Consensus RAxML topology realised for a 96 species x 1,734 nt 18S rDNA alignment of chrysophytes, trimmed to 50%, 80% and 90% occupancy. Branches are coloured by inferred trophic status. Species for which transcriptome or genome data are available are labelled with red stars. This tree was produced with taxonomic aid from Louis Graf (Sungkyunkwan University).

The chrysophyte project in my group developed iteratively, through exchanges with multiple collaborators and students. My initial focus was on the presence- or not- of plastid genomes in

non-photosynthetic chrysophytes. Previous studies had documented the presence of membrane-bound organelles that structurally resembled chloroplasts (i.e., leucoplasts) in non-photosynthetic chrysophytes, and one study of an unrelated non-photosynthetic stramenopile (the dictyochophyte *Pteridomonas danica*) had identified a probable *rbcL* gene, suggesting the presence of plastidial DNA (Mylnikov et al., 2008; Sekiguchi et al., 2002). At the same time, there was emerging evidence that some secondarily non-photosynthetic plants (e.g., the carrion flower *Rafflesia lagasceae*) and algae (the green alga *Polytomella parva*, and the dinoflagellate *Haematodinium perezii*), had completely lost a plastidial genome or (in the case of *Haematodinium*) had lost all vestiges of a plastid-derived organelle (Gornik et al., 2015; Molina et al., 2014; D. R. Smith & Lee, 2014).

I first used published sequence data to determine the extent that non-photosynthetic chrysophytes retained material from their ancestral chloroplasts. For this, I was able to avail of recently generated MMETSP transcriptomes for three strains of the non-photosynthetic marine chrysophyte *Paraphysomonas*, and the distantly positioned non-photosynthetic freshwater species *Spumella elongata* (Keeling et al., 2014), alongside data in construction for the pan-stramenopile plastid proteomes (see Chapter 3). The initial exploration of the *Paraphysomonas* and *Spumella* plastomes was performed with two Normalien L3 interns, Lucas Paoli and Guillemette Audren de Kerdrel, who joined me in summer 2015 and 2016 to explore non-photosynthetic plastid chrysophytes respectively at a computational and experimental level (Dorrell, Azuma, et al., 2019; Dorrell et al., 2021; Klinger et al., 2018)

The results from these two internships revealed different reductive histories in the *Paraphysomonas* and *Spumella* plastids (**Fig 2.9**). For *Spumella*, Guillemette could amplify both plastidial 16S and 23S rDNA, consistent with the retention of a plastid genome. Using the same primers with *Paraphysomonas*, she could only amplify bacterial sequence, despite being able to confirm the presence of both mitochondrial and nuclear DNA from the same samples (**Fig 2.9A**). We also confirmed these data using deep-sequencing approaches, where we were unable to assemble contigs coherent with the presence of a *Paraphysomonas* plastome from total genomic DNA. These data are consistent with the probable loss of plastidial DNA from *Paraphysomonas*, or at the very least the absence of plastid sequence homologous to barcoding consensus regions.

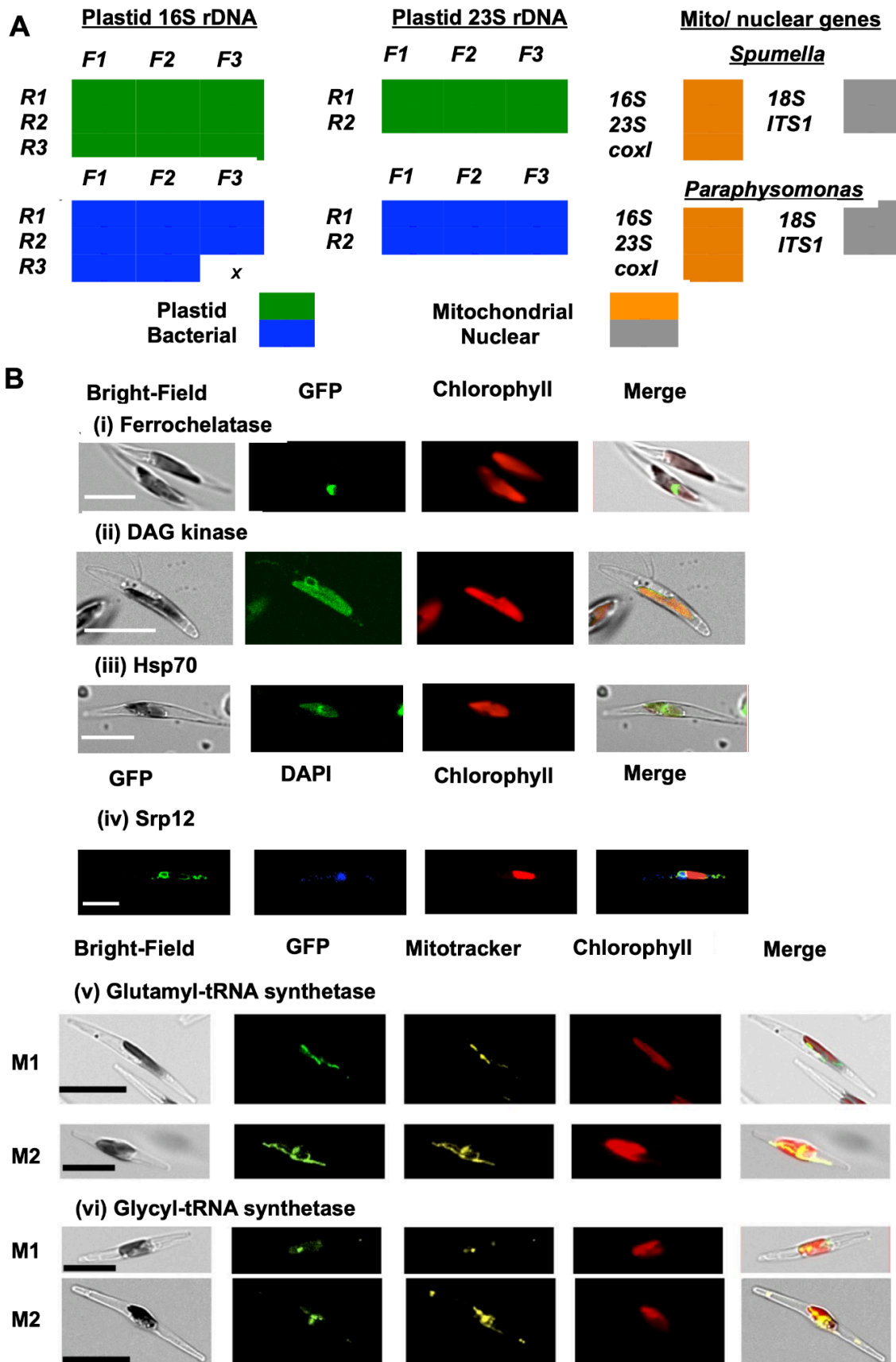


Fig. 2.9. Inference of plastid DNA in non-photosynthetic chrysophytes. **A:** results of PCRs performed with consensus plastid 16S and 23s rDNA primers (three forward primers, labelled F1-3, x two-three reverse primers, labelled R1-R3) for genomic DNA from (top) *Spumella elongata* CCAP 955/1 and (bottom) *Paraphysomonas bandaiensis* RCC383 and, alongside equivalent PCRs performed using consensus eukaryotic mitochondrial (16S, 23S

rDNA, *cox1*) and nuclear (18S, ITS1) barcoding primers (Gachon et al., 2013; Marin et al., 2003). All three organelle genomes were amplifiable for *Spumella*, while PCRs with plastid-specific primers for *Paraphysomonas* amplify bacterial sequence only. **B**: Confocal microscope images for *Phaeodactylum tricornutum* cell lines expressing GFP-tagged sequences of *Paraphysomonas* plastid-derived proteins introduced via biolistic transformation per (Falciatore et al., 1999) lines are stained for mitochondria (MitoTracker Orange) or nuclei (DAPI) following (A. Tanaka et al., 2015). Notably, genes encoding proteins associated with the plastid genome in other stramenopiles (e.g., aminoacyl-tRNA synthetases, **v** and **vi**) are likely to be uniquely targeted to the *Paraphysomonas* mitochondria, regardless of the methionine (**M1**, **M2**) from which the protein is expressed, whereas other proteins appear to successfully localise to the periplastid compartment (Hsp70, ferrochelatase, **i** and **iii**) or chloroplast: endoplasmic reticulum (DAG kinase, Srp12, **ii** and **iv**).

Considering the transcriptomes of these species, Lucas could identify key proteins relating to plastid biogenesis and metabolism in both species, e.g. haem and lipid metabolism (ferrochelatase, DAG kinase) and protein import across the chloroplast-endoplasmic reticulum (Srp12) and periplastid membrane (Hsp70) in *Paraphysomonas* (Gould et al., 2006; Huang et al., 2023; Richtová et al., 2021). We further validated the localisations of some of these proteins around the *Paraphysomonas* plastid via the expression of C-terminal GFP fusions in the transformable diatom *Phaeodactylum tricornutum* (**Fig 2.9Bi-iv**). In contrast, Lucas was unable to identify *Paraphysomonas* homologues of most plastid proteins implicated in genome expression or replication in his transcriptome searches. Together, these data suggest a probable loss of the *Paraphysomonas* plastid DNA.

Next, we attempted to develop a more formal proof for the secondary loss of the *Paraphysomonas* plastid genome. For this, we collaborated with Gillian Gile at Arizona State University, who had recently published the first study in diatoms of aminoacyl-tRNA synthetases, a class of proteins required for charging amino acids onto tRNAs, with a different synthetase required for each amino acid expressed from each genome (Gatti & Tzagoloff, 1991; Gile et al., 2015). Many plant aminoacyl-tRNA synthetases are dual-targeted to the mitochondria and chloroplasts, and Gillian had shown that diatoms typically possess two copies of each aminoacyl tRNA synthetase, one cytoplasmic involved in expression of the nuclear genome, and the other dual-targeted to the mitochondria and chloroplast, involved in the expression of both (Duchêne et al., 2005; Gile et al., 2015). Looking into our *Paraphysomonas* data, Lucas and I were typically able to identify two copies of each synthetase, one corresponding to the nuclear and one corresponding to the dual organelle-targeted copy. However, the organelle-targeted sequences lacked hydrophobic signal peptides necessary for translocation into the plastid (i.e., through the chloroplast:

endoplasmic reticulum (Gruber et al., 2015), and only possessed a hydrophilic N-terminal targeting sequence inferred to allow import into the *Paraphysomonas* mitochondrion (Hammond et al., 2022; Neupert, 1997). Expressing C-terminal GFP-tagged copies of these proteins in *Phaeodactylum*, Guillemette and I noted that the organelle-targeted isoforms indeed uniquely localised to the mitochondrion, with no apparent dual-targeting (**Fig. 2.9Bv, vi**). The mass relocation of a protein family ancestrally involved in the expression of the *Paraphysomonas* plastid genome to another organelle (the mitochondria) could be considered a formal confirmation of the loss of plastid DNA from this species, and I am pleased to see this technique has subsequently adopted by other groups to identify further losses of plastid DNA from other secondarily non-photosynthetic algae (e.g., *Actinophrys sol* (Azuma et al., 2022)).

The final component of the chrysophyte plastid evolution story was a collaboration in 2018 with Ryoma Kamikawa at Kyoto University, who has similarly a longstanding interest in secondarily non-photosynthetic algae (Azuma et al., 2022; Kamikawa, Tanifuji, et al., 2015; Moog et al., 2020). Ryoma's group had recently sequenced the plastid genome of the chrysophyte "*Spumella*" sp. NIES-1486, which is distantly related to the *Spumella* sp. CCAP 955/1 studied in my group, suggesting independent retention of plastid DNA in multiple non-photosynthetic chrysophyte lineages (**Fig. 2.8**). Ryoma, Gillian and I performed a more exhaustive comparison of the inferred plastid-targeted proteomes of *Spumella*, *Paraphysomonas* and other secondarily non-photosynthetic chrysophytes, at this point integrating further transcriptome data from non-photosynthetic freshwater species sequenced in the group of Jens Boenigk (Beisser et al., 2017; Graupner et al., 2018).

These data revealed recurrent trends, but also variation, in the plastid proteomes of individual chrysophytes (**Fig. 2.10**). *Paraphysomonas* represents one extreme, as the only lineage in which proteins implicated in plastid DNA replication or expression are likely to have lost (or relocated to other organelles), consistent with the absence of plastid DNA. On the other extreme, *Cornospumella fuschlensis* (Grossmann et al., 2016) retains much of its photosynthesis machinery, with the notable absence of a detectable oxygen-evolving complex of photosystem II, and appears to encode the majority of its ancestral carbon, lipid and cofactor metabolism pathways (**Fig. 2.10**). I find it nonetheless intriguing that effectively all of the plastidial amino acid biosynthesis pathways (e.g., glutamine/ glutamate synthase, branched chain amino-acid, lysine and shikimate pathway biosynthesis), many of which require plastidial NAD(P)H production, appear to have been lost across the diversity of

non-photosynthetic chrysophytes including *Cornospumella*, and wonder if this is likely to precede the formal loss of photosynthetic capacity in these species (Fig. 2.10).

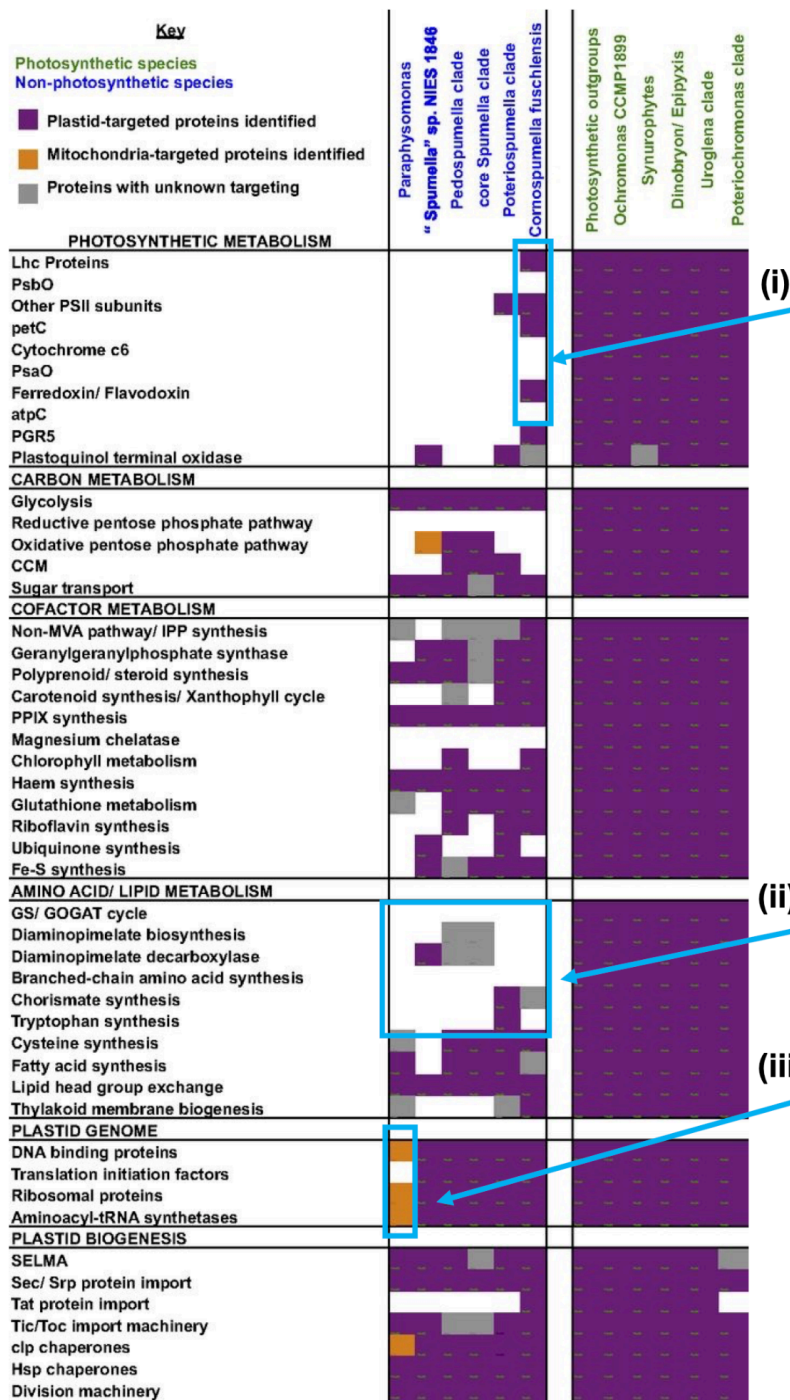


Fig. 2.10. Contrasting degrees of plastid proteome reduction across non-photosynthetic chrysophytes.

Schematic diagram of plastid proteome content of different non-photosynthetic (dark blue) or photosynthetic (green) chrysophyte species, inferred by *in silico* targeting prediction (ASAFind, and modified HECTAR) and single-gene phylogenies (Dorrell, Azuma, et al., 2019; Gruber et al., 2015; Gschloessl et al., 2008). Proteins are divided by biological function, and shaded by predicted localisation (plastid, mitochondria, or elsewhere). Light blue boxes indicate variable tendencies in plastid proteome content across non-photosynthetic species: (i) greater retention of photosystem and LHC proteins in *Cornospumella* than other non-photosynthetic species; (ii) almost universal loss of NAD(P)H- dependent amino acid biosynthesis pathways in non-photosynthetic species; and (iii) unique loss (or mitochondrial relocation) of plastid DNA replication and gene expression in *Paraphysomonas*.

2.6. Concluding remarks and pedagogical insights

Exploring the proteomic diversity of algal chloroplasts, particularly at a computational level, has informed my subsequent attempts to explore the functions of algal chloroplast proteins, and to consider deeper evolutionary phenomena that may explain the diversification of algal

life. Conceptually, the plastid diversity and broader trophic flexibility of dinoflagellates and chrysophytes (**Figs. 2.1, 2.8**), give particularly interesting grounds to explore the biological consequences of natural evolutionary variation. The data from my group and from other researchers have allowed us to project multiple and independent plastid acquisitions in the case of Kareniaceae (Hehenberger et al., 2019), and or losses of photosynthesis in chrysophytes (Graupner et al., 2018) across relatively confined taxonomic spaces (**Figs. 2.7, 2.9**). I note that the repeated acquisition of the same plastid group has also been documented in closely-related dinoflagellates with green algal and diatom endosymbionts (Sarai et al., 2020; Yamada et al., 2017) (**Fig. 2.2**). Multiple closely-related losses of photosynthesis have been inferred in diatoms, dinoflagellates and chlorophytes (Cooney et al., 2022; Kamikawa, Yubuki, et al., 2015; Pánek et al., 2022), and thus neither Kareniaceae nor chrysophytes are unique in this evolutionary plasticity.

Beyond purely exploring what algal life can and cannot do in an evolutionary capacity, I feel that these taxonomic approaches can provide us with important information into protein function. For example, considering dinoflagellate chloroplast gene evolution on a purely observational level, we identified gene families and domains of chloroplast-encoded proteins (e.g., stromal faces of peridinin dinoflagellate photosystem I) that are particularly divergent (**Fig. 2.3**). These domains may be of interest for future biochemical exploration, for example using chloroplast transformation in the model species *Amphidinium* (Nimmo et al., 2019) to probe protein-protein interactions or functional domains. Subsequent work on the consequences of chloroplast transcript editing for sequence homology helped pinpoint similarities and differences between how peridinin and Kareniacean plastids operate at evolutionary levels, with editing functioning to correct divergent evolution in underlying genomic sequences, but with potentially having secondary roles in creating sequence novelty in Kareniacean plastids compared to haptophytes (**Fig. 2.4**). It will be interesting to identify what effectors of RNA editing (e.g. PPR-type or MORF-type proteins) (Barkan & Small, 2014; Hicks et al., 2019; Takenaka et al., 2012), are shared between peridinin dinoflagellates and Kareniaceae, which we may be able to do e.g. by screening Kareniacean plastid proteomes for RNA-binding proteins of inferred host origin (**Fig. 2.6**), prior to exploring their different functions. Finally, alongside identifying differences in core plastid metabolic pathways (**Fig. 2.10**), comparative analysis of photosynthetic and obligately heterotrophic chrysophytes may allow us to identify novel proteins whose absence from heterotrophic groups imply cryptic

functions in photosynthesis. While I have not explicitly yet exploited this approach in my research, two of the proteins under functional exploration in my group (Trans3, and xATPA, see Chapter 3) are projected to have been lost from non-photosynthetic chrysophyte species, which may suggest functions linked to photosynthetic performance.

On broader cellular grounds, these data point to the roles of other organelles within the host cell in shaping chloroplast evolution. The apparent conservation of dicysteine residues in the signal peptide regions of Kareniacean and apicomplexan plastid-targeting motifs (**Fig. 2.5**) (Barylyuk et al., 2020) might reflect common features of the ER protein import machinery derived from the common ancestor of the dinoflagellates and apicomplexans (Janouskovec et al., 2010). Anna's data, more globally, project a strong contribution of the dinoflagellate host to the *pan*-Kareniacean chloroplast proteome, particularly in pathways (e.g. carbon metabolism) also associated with non-photosynthetic plastids (**Fig. 2.6**). Anna has proposed in her paper that this is due to the presence, or evolutionary contribution, of a non-photosynthetic host-derived plastid to Kareniacean chloroplast evolution, a view shared by some of our collaborators (Hehenberger et al., 2019; Novák Vanclová et al., 2024). Finally, the relocalisation of ancestrally dual-targeted aminoacyl-tRNA synthetases to exclusively support the *Paraphysomonas* mitochondrion provides an example of how plastids can shape the biology of other organelles (**Fig. 2.9**). We, and others, have shown that many of the ancestrally dual-targeted synthetases associated with chrysophytes, diatoms and other plastid-containing symbionts are of red algal (i.e., plastid endosymbiont) origin (Dorrell, Azuma, et al., 2019; Gile et al., 2015). Thus, a plastid genome that no longer exists in the *Paraphysomonas* cell has imprinted its associated tRNA synthetases on the expression pathways of another organelle, the mitochondrion. The functional significance of plastid-mitochondria interactions in algae remains an area of strong interest to me (Hammond et al., 2022) and is further explored in the plastid glycolysis work described in Chapter 3.

At an ecological level, I wonder whether the environmental niches dinoflagellates and chrysophytes inhabit influence their evolutionary variability. As described in Anna's paper, Kareniaceae have different distributions to haptophytes in *Tara* Oceans data (Novák Vanclová et al., 2024), and it remains to be determined if the repeated acquisitions of haptophyte plastids relates to ancestral (if not current) niche co-occurrence, trophic preference for haptophytes in the host cell, or a specific predilection for converting consumed haptophytes into stable endosymbionts. More information into the trophic activity of Kareniaceae on

haptophytes (Jeong et al., 2021), or the stability of haptophyte chloroplasts in isolation or inside their host (e.g., in kleptoplastic species) (Cruz & Cartaxana, 2022; Decelle et al., 2012) may help provide insights into why they are so frequently domesticated as endosymbionts. Similarly, the majority of documented non-photosynthetic chrysophytes, including all of the species for which we possess transcriptome data except *Paraphysomonas*, are from freshwater habitats. We might expect more acute fluctuations in nutrients in these ecosystems (e.g., nitrate scarcity in bog ecosystems, or eutrophication) that might select for increased reliance on heterotrophic as opposed to photosynthetic acquisition of amino acids (Gomez Isaza et al., 2020; Levy et al., 2019). Greater dispersal limitation in freshwater than marine ecosystems subject to global circulation, might further allow heterotrophic chrysophytes to outcompete phototrophic relatives (Cornwallis et al., 2023; Richter et al., 2022). That said, non-photosynthetic marine chrysophytes beyond *Paraphysomonas* have been documented, both using single-cell approaches and cultivation, and the full diversity, taxonomy and functional biology of marine chrysophyte heterotrophs awaits deeper exploration (Gérikas Ribeiro et al., 2020; Seeleuthner et al., 2018).

Finally, at a pedagogical level, these explorations of chloroplast proteomes have been an instructive lesson in establishing and maintaining collaborations, whether with project students (Lucas Paoli, Guillemette Audren de Kerdrel, Fuhai Liu), laboratory alumni and former colleagues (Elisabeth Richardson, Erin Butterfield), or other researchers (Christen Klinger, Robert Newby, Gillian Gile, Ryoma Kamikawa). I am particularly struck by how much of this research has happened as “evening projects” with other motivated young scientists, and, within reason, I think that these side-collaborations can be a deep source of motivation and inspiration. I am particularly indebted to Anna Novák Vanclová, for developing the Kareniacean plastome evolution beyond my initial vision, and working under difficult logistic circumstances (i.e., COVID) to produce a nuanced vision of how plastid biology can vary over even micro-evolutionary scales. I hope that I will be able to capture at least some of this sensitivity to the biological variation across algal strains within my ongoing research.

Chapter 3: Integrative environmental and experimental characterisation of the 'dark matter' of diatom chloroplast proteomes

3.1. Diatoms: a model primary producer in a wider constellation of algal diversity

Following my PhD on dinoflagellates in the group of Chris Howe at Cambridge University, I joined the group of Chris Bowler at the IBENS for my post-doctoral research. At the time of joining (in late 2014), the Bowler group was principally known as a diatomist group (Bowler et al., 2010), predating the seminal *Tara* Oceans publications in January 2015 (Bork et al., 2015). Chris' group had led the development of the pennate diatom *Phaeodactylum tricornutum* as a model system, developing molecular techniques in transformation, fluorescence microscopy and evaluation of gene expression (Falciatore et al., 1999; Siau et al., 2007), coordinating the sequencing and annotation of its genome (Bowler et al., 2008; Moustafa et al., 2009), and uncovering unusual Fe stress metabolism and nitrogen assimilation and recycling pathways using molecular biology approaches (Allen et al., 2008, 2011).

Although my work has been greatly shaped by diatom biology, and *Phaeodactylum* as a model experimental system (Ait-Mohamed et al., 2020; Dorrell, Azuma, et al., 2019; Dorrell et al., 2022; Dorrell, Gile, et al., 2017), I have to admit that I do not find diatoms intrinsically the most interesting algal lineage. So I will be relatively succinct in introducing their biology and ecology. It is true that diatoms play an important role in marine primary production, with some sources claiming as much as 40% of total photosynthetic biomass (Benoiston et al., 2017), although I rather feel this contribution is overestimated (e.g., due to amplification biases from commonly used 18S rDNA barcoding primers) (H. Liu et al., 2009; Pierella Karlusich et al., 2023). Diatoms predominate in nutrient-rich and well-mixed but iron-limited oceans at high latitudes, and are thus projected to be susceptible to the global consequences of oceanic heating and climate breakdown (Chaffron et al., 2021; Malviya et al., 2016). Diatoms have unusual biological features that distinguish them from other eukaryotic algae (including their closest sister-group, the bolidophytes), e.g. the predominance of sessile (as opposed to flagellated) forms, the encapsulation (in most species) of the diatom cell in a silicon-rich frustule, and the predominant use of photo-autotrophy or osmotrophy instead of phagotrophic mixotrophy as a life-strategy (Ban et al., 2023; Bowler et al., 2010).

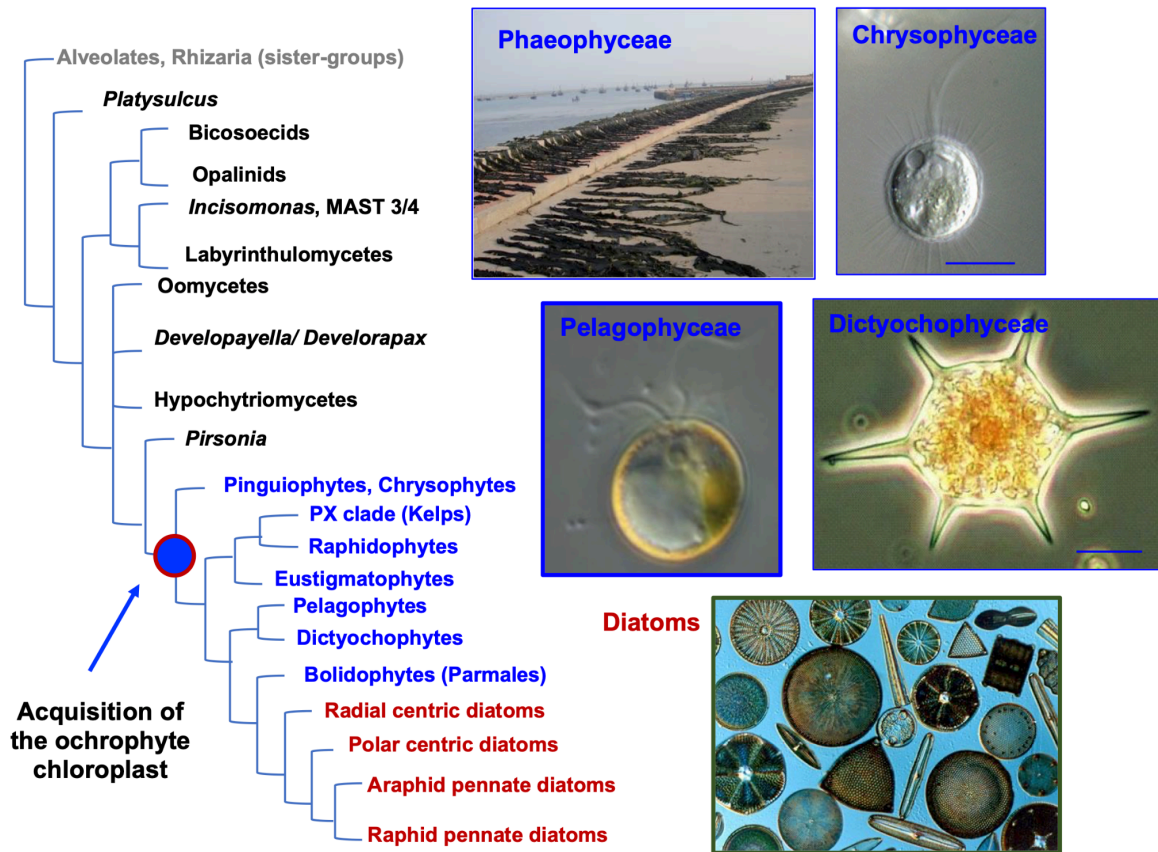


Fig. 3.1. Taxonomic and trophic diversity of stramenopiles. This figure, adapted from (Dorrell & Bowler, 2017), shows a schematic stramenopile tree, including diatoms (brown), other Ochrophyta (blue), and non-photosynthetic species (black). Images of specific groups of particular environmental interest are shown on the right-hand side.

My initial interest in diatoms reflected rather the evolutionary context in which they sit. In the same way as plants might be considered as an abstraction of a much more diverse phylum (i.e., Viridiplantae, or green algae), diatoms form just one tip node in a much more diverse phylum, the Stramenopila (Andersen, 2004; Dorrell & Bowler, 2017). This lineage regroups diverse photosynthetic members within the class Ochrophyta, including not only diatoms, but also chrysophytes, other single-celled algae predominant in freshwater (Synurales, Eustigmatophyceae) or marine habits (Pelagophyceae, Dictyochophyceae), and giant multicellular seaweeds (i.e., brown kelps, within the Phaeophyceae) (Cock et al., 2010; Dorrell & Bowler, 2017) (**Fig. 3.1**). This ecological and morphological diversity, which all possess the same chloroplast lineage (i.e., given monophyly or paraphyly in plastid genome trees (Janouskovec et al., 2010; Ševčíková et al., 2015)), is an appealing system to explore the post-endosymbiotic diversification of algal life. Beyond the ochrophytes, the stramenopiles further include multiple non-photosynthetic orders, including important plant and algal

pathogens (Oomycota) and aquatic and terrestrial decomposers and predators (Labyrinthulomycota, Bicosoecida) (Gachon et al., 2010; Jirsová et al., 2019; Rius et al., 2023). At the time of joining the Bowler group, it was still debated that these non-photosynthetic stramenopiles once possessed chloroplasts and lost them (i.e., per the “chromalveolate hypothesis” (Cavalier-Smith, 1999; Green, 2011a). That said, available genetic data from the first oomycete genomes already indicated that this was not the case, and these non-photosynthetic stramenopiles probably diverged from the ochrophytes prior to the acquisition of a plastid (Lévesque et al., 2010; Stiller et al., 2009) (Fig. 3.1).

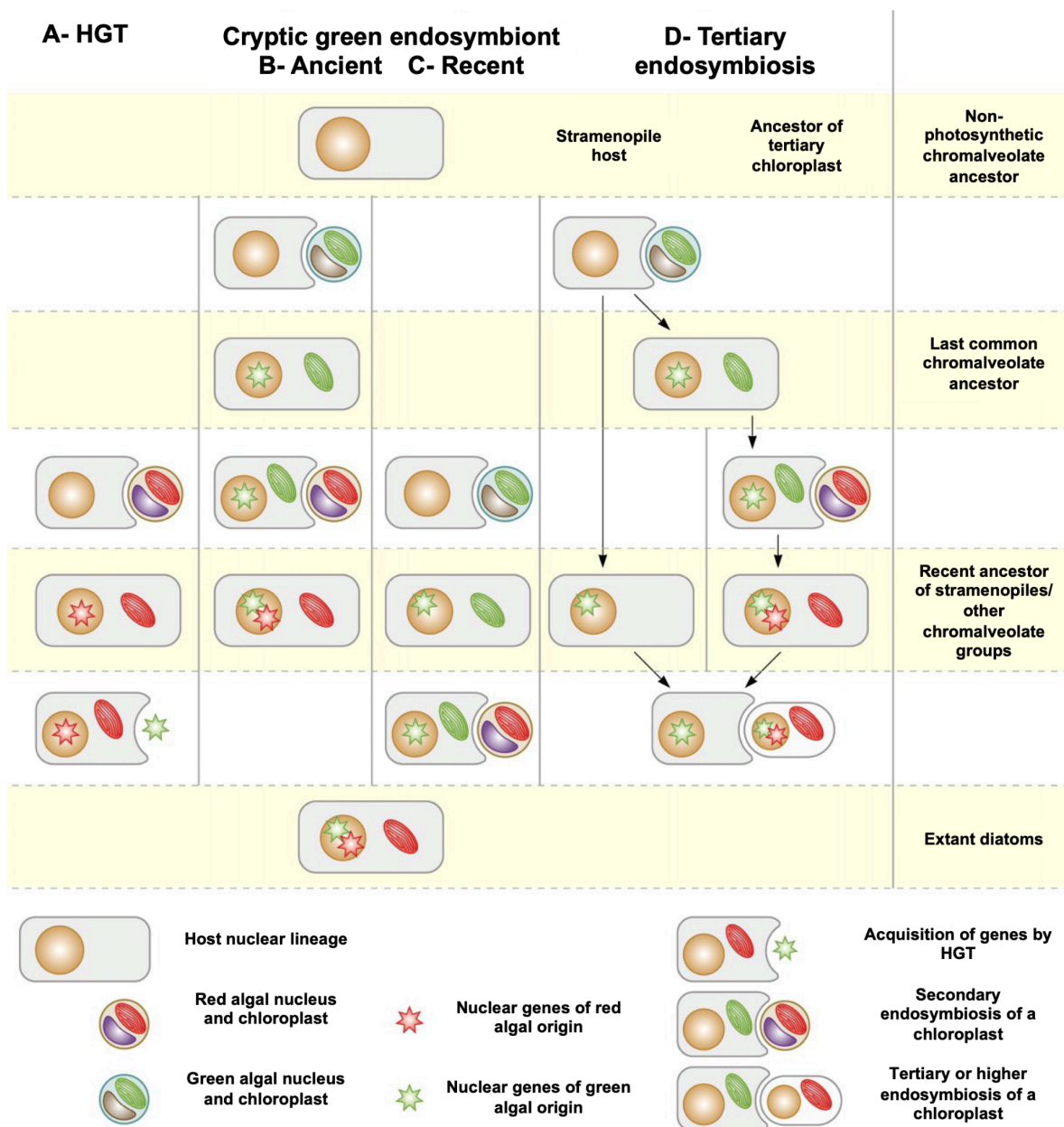


Fig. 3.2. Different hypotheses for the origin of green genes in diatom genomes. This figure has been modified from (Dorrell & Smith, 2011). Four different scenarios for the acquisition of green versus red genes in diatom genomes are shown. **A:** a recent horizontal acquisition of green genes post-endosymbiosis of the red chloroplast. **B:** a cryptic green algal endosymbiont in the last common chromalveolate ancestor that was then replaced by the

secondary endosymbiosis of a red alga. **C**: the same cryptic endosymbiosis, in this case occurring in the specific ancestor of stramenopiles to the exclusion of other chromalveolate groups. **D**: acquisition by the stramenopiles of another chromalveolate lineage, i.e. by tertiary or higher endosymbiosis, which itself possessed a secondary red chloroplast and underlying set of genes of green origin, giving rise to a chimeric, “red-green” endosymbiont.

More broadly, I was interested in the origins and composition of the stramenopile- and by extent diatom- plastid. It had long been agreed that the diatom chloroplast was of secondary (or higher) red algal origin, e.g. reflecting the retention of chlorophyll *c*, and close placement of its plastid genome to that of cryptomonads and haptophytes, within the red algae in plastid phylogenomic trees (Cavalier-Smith, 1999; Khan et al., 2007; Le Corguillé et al., 2009).

However, initial studies of diatom genomes had uncovered clues of a more complex evolutionary story. This began with data published at the time I was an undergraduate, suggesting that diatom genomes not only contained an extensive gene repertoire of green algal origin, which was proposed to be the remnants of a cryptic green endosymbiont acquired and lost deep in diatom evolutionary history (Becker et al., 2008; Moustafa et al., 2009). This story appealed to me as a potential example of a chloroplast with a mosaic evolution, although was also somewhat of a magnet for controversy given the limited taxonomic reference datasets used to confirm its identity and scale (Burki, Flegontov, et al., 2012; Woehle et al., 2011). I therefore wrote my undergraduate honours thesis on the likelihood, extent, and consequences of a green signal in diatom genomes, in effect concluding that at least part of this signal was real, although the exact mechanism by which it arrived in the diatom nuclear genome was uncertain (**Fig. 3.2**) (Dorrell & Smith, 2011).

I initially joined the Bowler group to re-evaluate the scope and timing of the green signal in the *Phaeodactylum* genome, in particular using more balanced taxonomic sampling (e.g., more red algal genomes) than was available at the time the initial analyses were performed. I was particularly interested in identifying which of these green genes were chloroplast-targeted and shared across all photosynthetic stramenopiles, i.e. may have been associated with the ancestral origins of its proteome. I am grateful to a few people who helped shape this idea: Jan Janouškovec for showing me the green gene story while I was an intern at the University of British Columbia (Burki, Flegontov, et al., 2012; Imanian et al., 2012; Janouskovec et al., 2010), Alison Smith for carefully directing my undergraduate thesis at Cambridge (Dorrell, Nef, et al., 2024; Dorrell & Smith, 2011), and Claire Gachon who supervised me for a final internship at SAMS before my PhD, who (at the 2013 British Phycological Society meeting) encouraged me to formally define how I would critically enumerate the green signal in diatom genomes.

I would finally like to note that, during the early years of my postdoc, the Bowler group was a dynamic, friendly and exciting environment, and I was able to interact productively with my colleagues working on diatom biology. I was fortunate to establish collaborations with many of them: Leïla Tirichine and Achal Rastogi exploring genome structure and expression (Ait-Mohamed et al., 2020; Rastogi et al., 2018), Elena Kazamia and Zhanru Shao considering iron and carbon metabolism (Kazamia et al., 2018; Shao et al., 2023), Fabio Rocha Jiménez Vieira and Juan José Pierella Karlusich using large-scale sequence data to explore diatom biology at phylogenetic and environmental levels (Carpenter et al., 2019; Pierella Karlusich et al., 2023), and Martine Boccara working on imaging and cellular dynamics (Bey et al., 2022). I feel that this time was particularly fecund as my colleagues and I were able to strike a balance between working collaboratively on common questions and systems (e.g., the same model organism), while maintaining our own individual interests and projects.

3.2. Evaluating- with limitations- the red and green origins of the diatom chloroplast

My first post-doctoral project was an attempt to evaluate the mosaic evolutionary signals underpinning the origins of the diatom chloroplast. Here, I searched multispecies datasets for genes encoding nucleus-encoded and chloroplast-targeted proteins shared across diatoms and a plurality of other ochrophyte genomes and transcriptomes, and could thus be attributed to their common ancestor; validated the chloroplast localisations of the proteins encoded by some of these genes using GFP-linked constructs in *Phaeodactylum*; and manually inspected protein phylogenetic origins using balanced taxonomic sampling, particularly with regard to the evaluation of red versus green signals (Dorrell, Gile, et al., 2017). This work in particular drew heavily on MMETSP (Keeling et al., 2014), and included some preliminary analysis of dual-targeted diatom aminoacyl-tRNA synthetases using both computational and microscopic techniques with Gillian Gile (Dorrell, Gile, et al., 2017; Gile et al., 2015). Our data validated the mosaic origins of the ochrophyte chloroplast proteome, with tractable red, green, bacterial and host signals extending to its common ancestor (**Fig. 3.3**).

This work has been important for establishing subsequent projects in my group, e.g. providing a ready dataset of 770 putatively ancestral and over 9,000 species-specific ochrophyte chloroplast protein homologues, and I have subsequently been able to reannotate this dataset in supervised student projects on specific features of chrysophyte and diatom plastid

proteome evolution (Dorrell, Azuma, et al., 2019; Nonoyama et al., 2019). In retrospect, I feel that it lacks the taxonomic subtlety of subsequent work from members of my group (c.f., the pan-Karenicean chloroplast proteome discussed in chapter 2, which explores similarly chimeric signals but extending down to the species level (Novák Vanclová et al., 2024). It was, nonetheless, an instructive lesson in what can and cannot be done interpreting a phylogenetic data on a large scale, which I will illustrate using contrasting examples from the green and host-derived plastid proteins in my dataset.

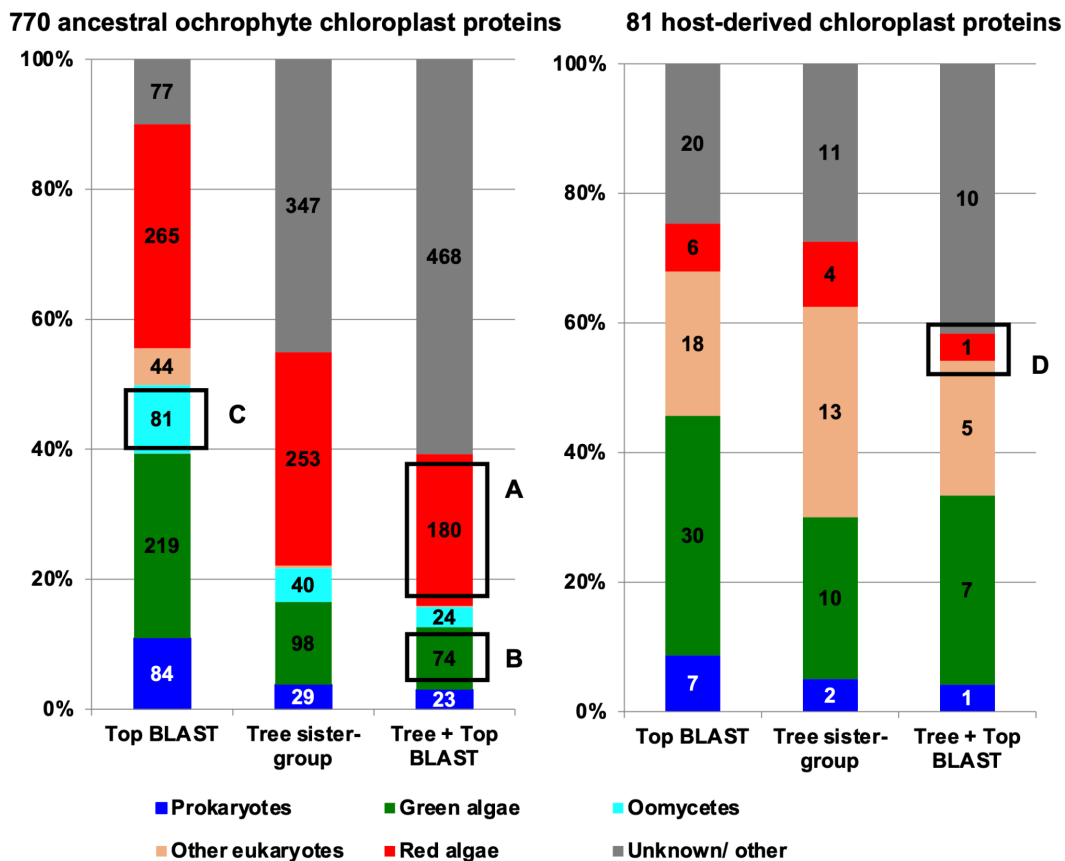


Fig. 3.3. Composition of the ancestral ochrophyte chloroplast proteome. Bar plots of (left) 770 protein orthogroups that appear to be chloroplast-targeted across a plurality of ochrophyte groups, and are likely to be derived from their common ancestor, and (right) those of host origin. Proteins are grouped by inferred phylogenetic affinities from BLAST best hit searches (minimum: three successive best-hits from different members of the same taxonomic group), single-gene phylogeny (single sister-group relationship with a monophyletic ochrophyte clade in RAxML consensus trees generated with the PROTGAMMAJTT substitution matrix), and combined filtration. Four labelled boxes are discussed in the text: **A** and **B**, numbers of proteins of assignable red and green origin, which occur in an approximately five to two ratio; **C** and **D**, number of proteins that can be assigned to the stramenopile host lineage, of which only one has a clear red origin.

Considering proteins of green origin, it is true that my data show a green algal signal in the ochrophyte plastid proteome (**Fig. 3.3**). This signal is present when we consider proteins that

have homologues from across the ochrophytes, i.e. both in diatoms, pelagophytes/ dictyochophytes, and chrysophytes/ phaeophytes; when we correct for taxonomic sampling biases by only considering proteins with homologues from a wide range of red algal groups; or perform rigorous analysis of internal BLAST affinities, tree sister-group positions, and uniquely shared residues. In its most strict evaluation, the green signal is composed of only 74 proteins, compared to > 1,800 in early analyses (Moustafa et al., 2009), but these are using stringent taxonomic criteria that also radically diminish the number of proteins with definable red algal origins in the dataset to 180 (**Fig. 3.3 A, B**). This suggests a conservative ratio of five red to two green genes supporting the ochrophyte plastid proteome (Dorrell, Gile, et al., 2017; Dorrell & Bowler, 2017). The green signal appears to be shared with Chlorophyta and not Streptophyta within the green lineage, arguing against vertical inheritance from a last common ancestor of the green algae and diatoms, and differential loss in other groups (e.g., red algae). Finally, a broader evaluation of all genes conserved across ochrophytes suggested that green genes more frequently encode chloroplast-targeted proteins than expected through random chance, suggesting that these genes do relate specifically in some way to the chloroplast (Dorrell et al., 2021, 2022).

I am hesitant to infer a cryptic green chloroplast deep in ochrophyte evolution. Subsequent studies, using novel genetic resources and more careful searches have continued to erode the number of genes identified as being of green origin in diatom genomes (Dautermann & Lohr, 2017). That said, I do not believe the number of inferred green genes will ever reach zero. From my own searches, again using large datasets incorporative of multiple red algal genomes and transcriptomes, I have found it difficult to expand far beyond 400-500 genes of red origin in the *Phaeodactylum* genome, which is far fewer than the number of proteins inferred computationally experimentally to support its chloroplast proteome (Gruber et al., 2015; Huang et al., 2023; Rastogi et al., 2018).

Nonetheless, the green ochrophyte genes could have been acquired from a “red-green” endosymbiont containing genes of both red and green algal origin (see scenario D.2 in **Fig 3.2**). This would give an ancestral plastid proteome containing both red and green genes, and an apparent bias towards green genes coding for chloroplast proteins in ochrophyte genomes, without there ever having been a direct endosymbiotic interaction between ochrophytes and green algae. I am particularly intrigued by the hypothesis from John Stiller that the ochrophyte chloroplast derives from a tertiary endosymbiosis of a cryptomonad, which is closely

positioned to the green algae (as a probable sister to the archaeplastids) (Stiller et al., 2014; Strassert et al., 2021). Perhaps we are seeing the footprints of a cryptomonad tertiary endosymbiont, which contributed its own endogenous “green” genes and a secondary “red” chloroplast. In the absence of identifying the specific donor of the ochrophyte chloroplast, there is a limit to how deeply we can interpret the phylogenetic composition of its proteome.

I have somewhat greater confidence in the capacity of our data to evaluate the presence of a chloroplast in a common ancestor of all stramenopiles, including ochrophytes and their non-photosynthetic sister-groups (oomycetes, labyrinthulomycetes). Within my data, I could detect approximately 80 of chloroplast-targeted proteins shared across ochrophytes that have direct homologues in non-photosynthetic stramenopiles (**Fig. 3.3C**). I reasoned that by considering the deeper origin of these proteins, I could infer whether they derive from a plastid endosymbiont shared between ochrophytes and oomycetes and subsequently lost from the latter, or if they relate to non-plastid proteins recruited to a plastid acquired after the divergence of ochrophytes from oomycetes. While a plurality of proteins in the total ochrophyte chloroplast proteome are of red algal origin, only one of those shared between the ochrophytes and oomycetes is (**Fig. 3.3A, D**). This argues quite strongly against a shared red chloroplast across stramenopiles subsequently lost from non-photosynthetic groups, i.e., suggests that oomycetes never possessed a chloroplast. Other studies, using similar but complementary approaches (e.g., the number of oomycete genes of putative red origin shared with diatoms) derive essentially no support for a cryptic red plastid in oomycete evolution (Leonard et al., 2018; Q. Wang et al., 2017). More broadly, I feel that this is an example of how phylogenetic data- if we can ask the correct question and design appropriate logical tests- can move beyond simple enumeration of different evolutionary signals, which I have tried to expand on in subsequent projects on HGT (see Chapter 4).

3.3. Moving from evolutionary to functional annotation of the diatom chloroplast

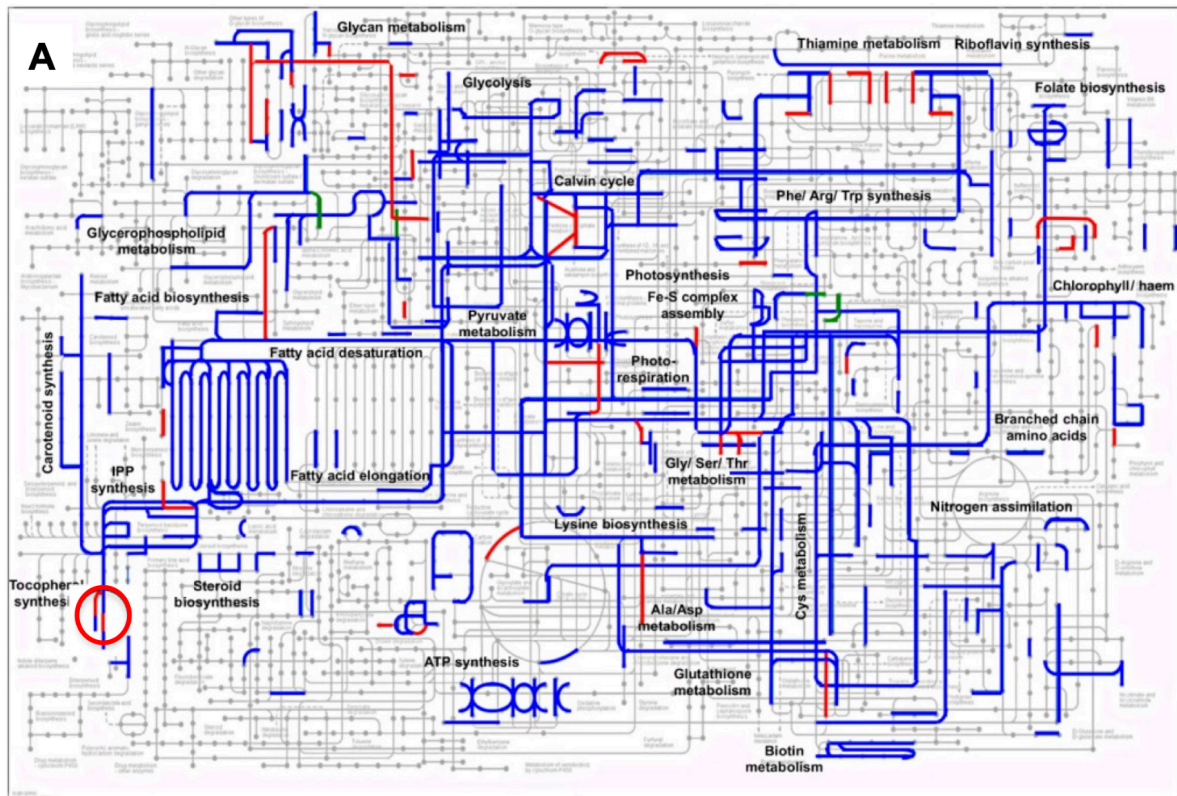
While my initial work on the ochrophyte plastid proteome was largely self-directed, the data generated have helped me more carefully design subsequent projects involving student participation. Several of these (Ait-Mohamed et al., 2020; S. Liu et al., 2022; Nonoyama et al., 2019) have tried to meld evolutionary and gene expression data to ascribe functions to the novel proteins associated with diatom chloroplasts. I particularly acknowledge Ellie McCallum

(Dorrell et al., 2021; Dorrell, Gile, et al., 2017), who performed a Masters internship with me in 2016 and developed some new approaches, looking at correlations of expression dynamics of diatom genes across published microarray data (Ashworth et al., 2016). Ellie has since completed a PhD at Concordia University and began a postdoc at UBC in synthetic microbiology (McCallum & Potvin-Trottier, 2021; McCallum & Tropini, 2023), and I am grateful to have participated in the beginning of her career.

The first of these projects was led by Tomomi Nonoyama, a visiting PhD student from Tsuyoshi Tanaka's group at Tokyo University of Agriculture and Technology who joined me for a one year placement between 2017 and 2018, generating CRISPR mutants for the glycolysis and transporter projects discussed in sections 3.4 and 3.5 (Dorrell, Zhang, et al., 2024; Nonoyama et al., 2019). Alongside this experimental work, Tomomi and I conducted a parallel bioinformatic project to identify protein innovations that separate the chloroplasts of diatoms from other ochrophytes. This project was initially sparked by a collaboration with Elena Kazamia in the Bowler Group, who was working on a diatom iron-stress induced protein (ISIP1) which she showed was involved in the non-reductive uptake of siderophore-bound iron across the plasma membrane, prior to delivery to the chloroplast (Kazamia et al., 2018). Elena and I inferred that this protein was likely to be a unique innovation of diatoms, except for a distantly related homologue that may have been shared between pelagophytes, haptophytes and Kareniaceae by chloroplast endosymbioses (see Chapter 4), and we were interested here, with Tomomi, to identify other proteins unique to the diatom chloroplast.

Broadly, by reconciling all of the > 9,000 putative plastid-targeted protein families identified in my previous work to a species tree developed by Guillemette Audren de Kerdel (see Chapter 4), Tomomi and I inferred substantial elaborations (i.e., protein and metabolic gains) in diatom chloroplasts subsequent to the radiation of the ochrophyte common ancestor (**Fig. 3.4A**). Many of these were associated either with the common ancestor of the diatoms, pelagophytes and dictyochophytes, or the common ancestor of pennate and polar centric diatoms, presaging later work on gene transfers and metabolic innovations specific to these timepoints in diatom evolution (Dorrell et al., 2021; Dorrell, Zhang, et al., 2024). Among the protein gains that were assigned specifically in the diatom chloroplast were a complete ornithine cycle, which has been inferred to facilitate the recycling of excess chloroplast amino acids in collaboration with the diatom mitochondria (Broddrick et al., 2019; S. R. Smith et al., 2019); the iron storage protein ferritin, which may enable diatoms to store transiently

available iron, and to occlude the growth of competitors (Lampe et al., 2018; Marchetti et al., 2009); and a complete biosynthesis pathway for the antioxidant tocopherol, which relates to the diatom-specific recruitment of tocopherol cyclase to the chloroplast (Fig. 3.4B). This latter protein has recently been identified from experimental proteomics of plastid-enriched *Phaeodactylum* fractions (Huang et al., 2023). While a phylogenetic observation is just the beginning of a long process of functional characterisation, I nonetheless hope that Tomomi's data may be useful for future researchers searching for plastid proteins to study in diatoms.



Shared to pelagophytes & dictyochophytes / gained by pennate and polar centric diatoms

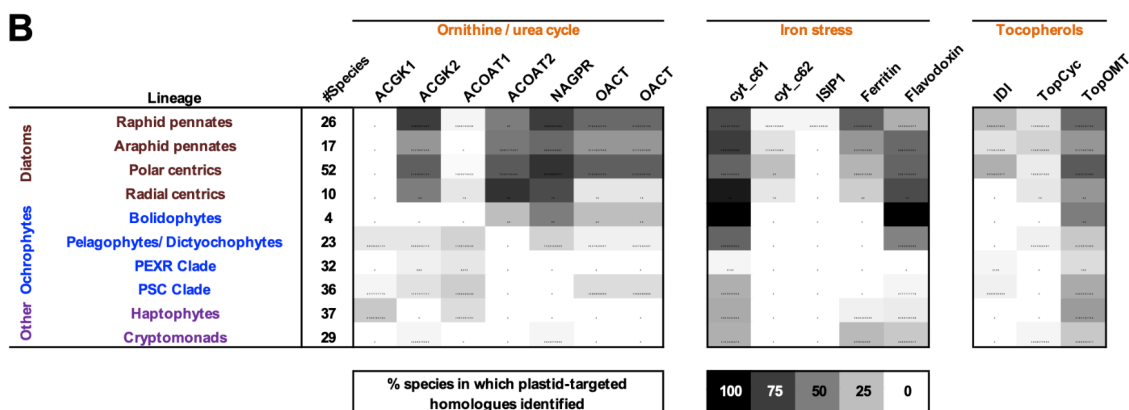


Fig. 3.4. Progressive gains in diatom plastid proteome content. **A:** KEGG map (Kanehisa & Sato, 2020) of plastid metabolic functions assigned to the diatom/ pelagophyte/ dictyochophyte common ancestor (blue) or specific to the pennate and polar centric diatom ancestor (red), inferred by *in silico* prediction and reciprocal BLAST best-hit. The elaborations specific to diatom chloroplasts include a complete tocopherol synthesis pathway (circled)

generated by the diatom-specific plastidial recruitment of tocopherol cyclase. **B**: heatmap of the frequency with which plastid-targeted copies of proteins involved in ornithine, iron stress and tocopherol metabolism were assigned to 266 ochrophyte, cryptomonad and haptophyte genomes and transcriptomes. Abbreviations: ACGK- acetylglutamate kinase; ACOAT- acetylornithine aminotransferase; NAGPR- N-acetylglutamylphosphate reductase; OACT- ornithine acetyltransferase; cytc6- cytochrome c6; ISIP1- iron stress induced protein 1; IDI- isopentenyl diphosphate isomerase; TopCyc- tocopherol cyclase; TopOMT- tocopherol o-methyltransferase.

The second functional project was chiefly done outside my group, with initial work from Leila Tirichine and Ouardia Ait-Mohamed at the IBENS. Leila and Ouardia had generated a WGCNA (weighted gene correlation network association) network for the entire *Phaeodactylum tricornutum* genome, using the version 3 genome annotation and large-scale RNAseq under different nitrate, phosphorus and iron availability conditions (Cruz de Carvalho et al., 2016; McCarthy et al., 2017; Rastogi et al., 2018; S. R. Smith et al., 2016). The objective of this study was not to find genes upregulated under nitrogen, phosphorus or iron stress, but rather to identify ones expressed with the same patterns in different datasets, whose co-expression might suggest linked physiological or cellular functions (Ait-Mohamed et al., 2020). My role in this project was limited to interpreting the WGCNA outputs, comparing it to similar co-expression networks previously generated from *Phaeodactylum* microarray data (i.e., DiatomPortal) previously explored by Ellie McCallum in my group (Ashworth et al., 2016; Dorrell, Gile, et al., 2017). My group further annotated core chloroplast and mitochondrial metabolic pathways in the PhaeoNet dataset, as an early post-doctoral project for Anna Novák Vanclová (Ait-Mohamed et al., 2020).

By comparing Ouardia's WGCNA annotations to *in silico* localisation and functional annotations, Anna was able to assign coexpressed gene modules to specific organelle metabolic activities. For example, most light-harvesting and photosystem complexes belong to the lightcyan1 and lightsteelblue1 WGCNA modules, while the TCA cycle and mitochondrial respiratory complexes are respectively predominantly found in the paleturquoise and orangered4 modules. The genes assigned to these modules may form future targets for functional analysis linked to photosynthetic activity and respiration. I was particularly interested by the blue module, which spans much of core chloroplast metabolism (e.g., amino acid and lipid biosynthesis, and the Calvin cycle), mitochondrial gene expression and biogenesis proteins (e.g., many amino-acyl tRNA synthetases) dual-targeted to both organelles (**Fig. 3.5**). Three of the genes under current functional investigation in my group (cpEnolase, cpPGM1A, and Trans3) belong to this module. As a final proof of concept, we used these data

to identify chloroplast-targeted proteins containing domains of unknown function (DUFs) that are coregulated with the chloroplast photosynthetic machinery, and explore the evolutionary and regulatory diversity of chloroplast sigma factors encoded in the *Phaeodactylum* genome, the latter in collaboration with Nathalie Joli from the Bowler group (Ait-Mohamed et al., 2020).

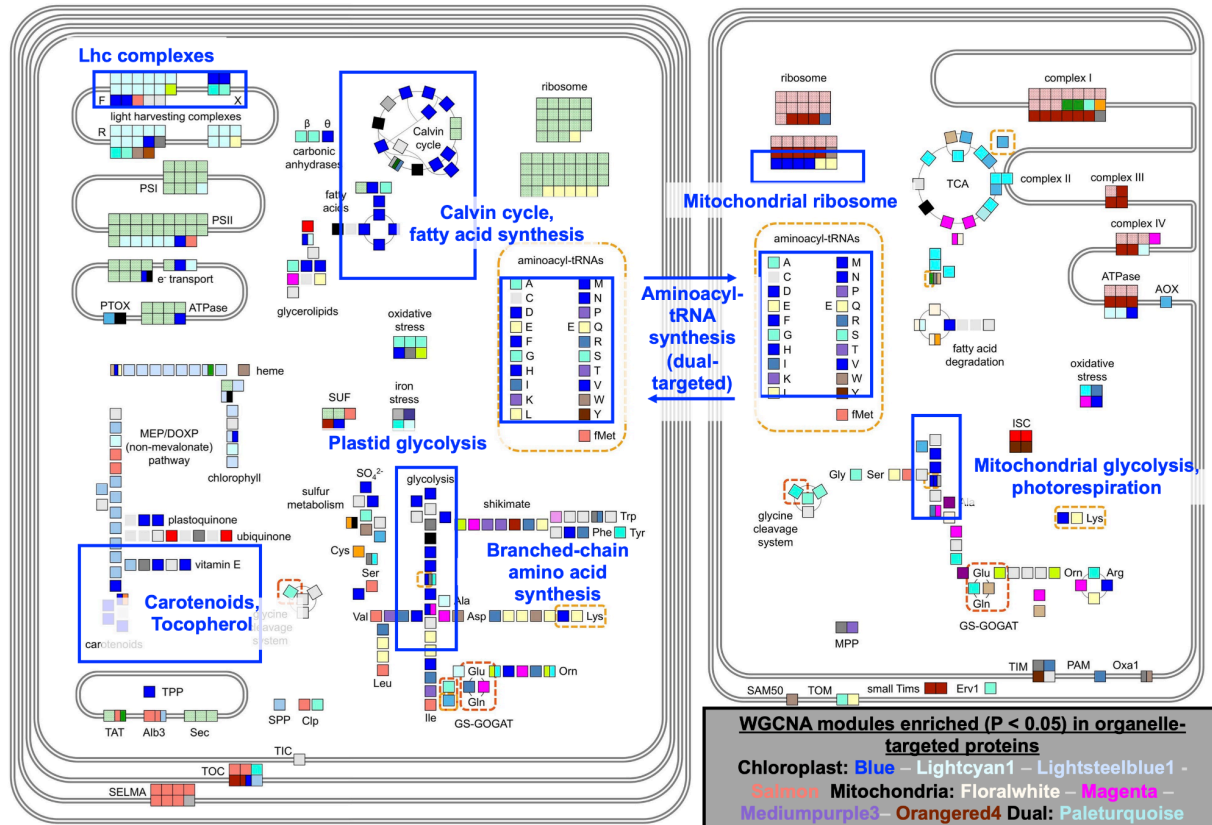


Fig. 3.5. WGCNA map of core *Phaeodactylum* organelle metabolism pathways. This figure shows a schematic diagram of a *Phaeodactylum* chloroplast (left) and mitochondrion (right), with its associated proteins shown as boxes coloured by WGCNA module identity. Organelle-encoded proteins (which are not considered in the WGCNA as the genes encoding them are not detected in most RNAseq data) are shown in pale green. Specific pathways that show a preponderance of blue module proteins are shown with labelled boxes.

In a final project, led by Shun Liu, a motivated student whose PhD I co-supervised with Chris Bowler from 2019-2023, we made forays into incorporating environmental data from *Tara* Oceans in our physiological understanding of chloroplast proteins. Shun joined my group to study an unusual chloroplast transporter (Trans3), discussed in section 3.5, considering experimental localisation, CRISPR mutant phenotypes, and inferred biochemical activity, but began her contract shortly prior to COVID and thus reoriented as a bioinformatician for the first year of her studies. Here, showing resilience and a desire to learn under very difficult circumstances, Shun performed a comprehensive analysis of *Phaeodactylum* chloroplast transporters, to inferring function using phylogenetic and gene expression analysis

approaches (S. Liu et al., 2022). Shun used coregulation approaches previously developed by Ellie McCallum and Tomomi Nonoyama in my group, to identify chloroplast-targeted transporters that are coexpressed with genes encoded on the mitochondrial genome, and might have functions coordinating diatom chloroplast and mitochondrial biology (Brodrick et al. 2019; Prihoda et al. 2012). Finally, building single-gene trees incorporating cultured species and meta-genomic data, Shun identified environmental sequences of diatom chloroplast transporters. Shun quantified the expression of these genes in relation to indices of photosynthetic activity (e.g., chlorophyll concentrations) and primary production (dissolved and particulate carbon) in *Tara* Oceans data (Carradec et al., 2018; Guidi et al., 2016).

While Shun began by considering these variables in isolation, with encouragement she started to compare different data to one another, revealing interesting phenomena of which diatom chloroplast transporters are most intricately connected to oceanic primary production. Shun for example noted that two of the transporters showing the strongest positive (Phatr3_J50540) and negative (Phatr3_J42868) correlations to primary production are uniquely found in diatoms and not known in other algae, and both show strong expression correlations to genes from the mitochondrial genome. Shun hypothesised that chloroplast: mitochondria crosstalk pathways unique to diatoms may be particularly important for directing their photosynthetic activity in the wild (S. Liu et al., 2022). Elsewhere across Shun's data, it is to me visually apparent that many of the transporters of green algal and inferred host origin show less positive correlations to primary production than those originating from red algae and their endosymbiotic derivatives (i.e., cryptomonads, haptophytes, dinoflagellates). That said, the second (Phatr3_J3046) and third (Phatr3_J38786) most positively correlated transporters come from green algae and the stramenopile host respectively (**Fig. 3.6**). Finally, Shun used these data to select some new chloroplast transporters that she explored at the bench within her PhD, e.g. one (Phatr3_J43171) that unusually shows a positive expression correlation to temperature in *Tara* data (S. Liu et al., 2022).

Correlation between environmental parameter and metaT relative expression

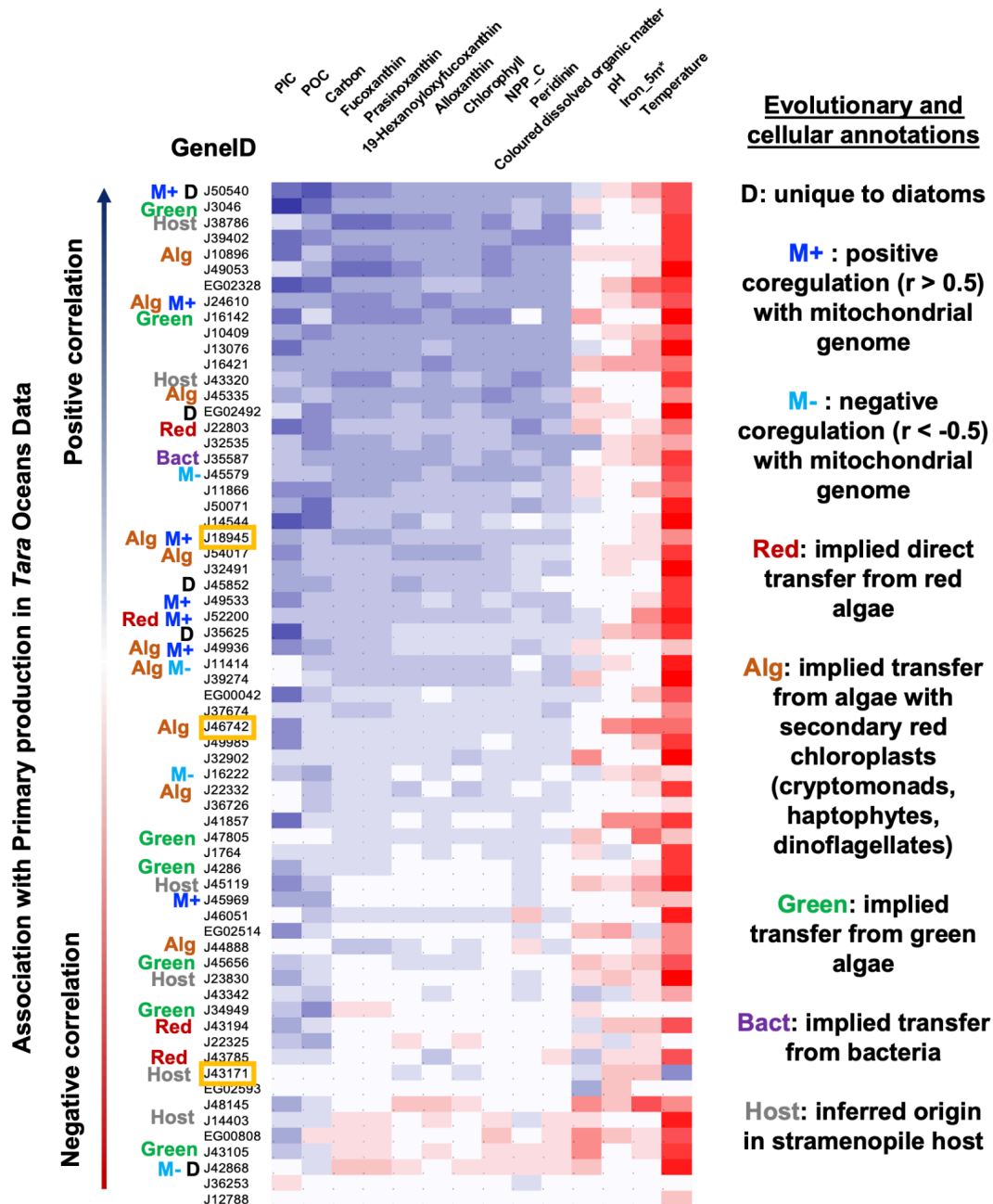


Fig. 3.6. Diatom chloroplast transporter diversity assessed by combined transcriptomic, phylogenetic and environmental approaches. This heatmap, adapted from (S. Liu et al., 2022), shows *Tara* Oceans homologues of 65 genes encoding chloroplast transporters from the diatom *Phaeodactylum*. Environmental homologues were identified by single-gene phylogeny to only retain those that formed monophyletic clades with sequences encoding chloroplast-targeted proteins from cultured diatoms. Homologues are ordered by descending correlation to indices related to primary production (the total number of unique depth and size fraction combinations showing significant correlations with $P < 0.05$) with transporters showing the strongest positive correlation at the top and those showing the strongest negative correlations at the bottom. Transporters are annotated with inferred evolutionary origin (uniqueness to diatoms, and origin via HGT, as shown by single-gene phylogenies in (Dorrell et al., 2021), and inferred coregulation to the *Phaeodactylum* mitochondrial genome, assessed using microarray data from (Ashworth et al., 2016). Transporters explored experimentally in Shun's PhD are shown with orange boxes. Abbreviations: PIC- particulate inorganic carbon; POC- particulate organic carbon; NPP_C - net primary production.

In total, these three “semi-functional” bioinformatic projects were a useful training for me in how to use published computational data to select candidate proteins for wet analysis, and indeed how to provide intellectual space for junior members of my group (particularly Ellie McCallum, Tomomi Nonoyama, and Shun Liu) to identify their own stories and questions for analysis. I will now outline some ongoing projects in my group that have used these data to inform the functional phenotyping of novel proteins associated with the diatom chloroplast.

3.4. Convergent phenotyping of diatom chloroplast glycolysis from environmental and experimental data

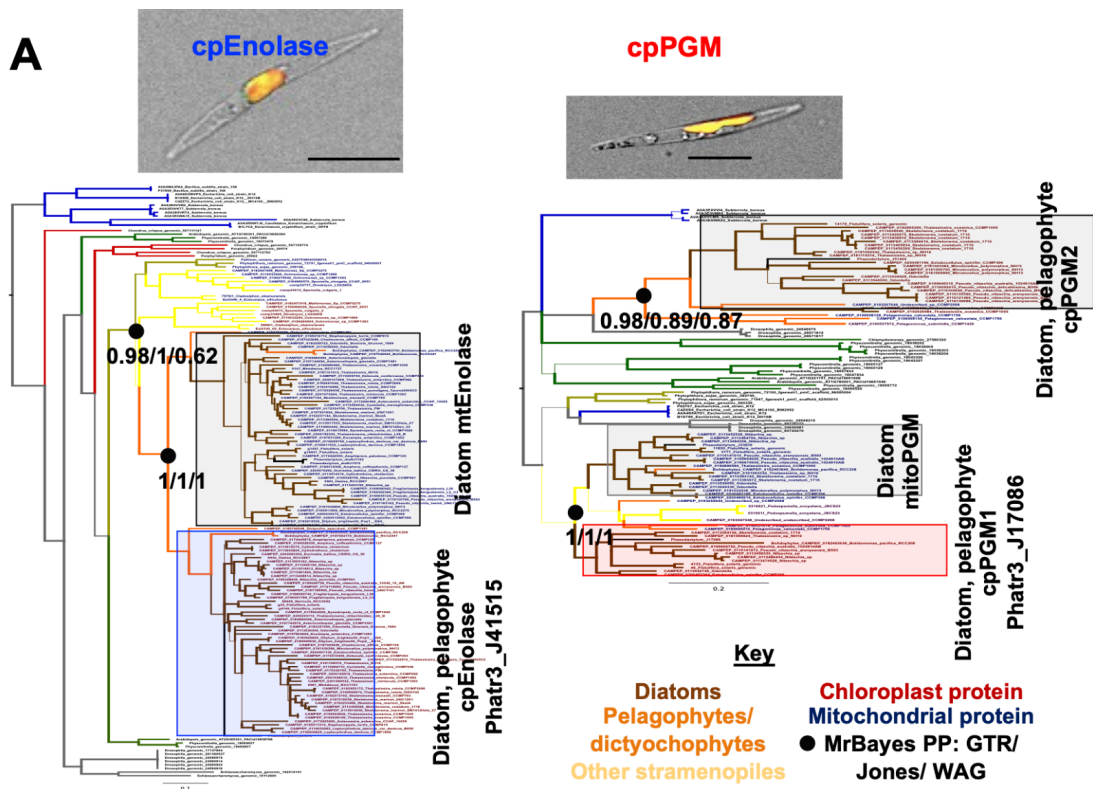
Most recently, my group has explored proteomic novelties of the diatom chloroplast experimentally, combining phylogenomic, transcriptomic and environmental approaches with protein localisation, CRISPR knockout and phenotyping in *Phaeodactylum*. These projects include those of my first two PhD students (Shun Liu, Mathias Penot-Raquin) although I begin with one project led by me, with participation from Tomomi Nonoyama during her exchange PhD year and Mathias during his Masters internship in my group (Dorrell, Zhang, et al., 2024).

This project relates to a complete glycolysis-gluconeogenesis pathway from glucose-6-phosphate to pyruvate annotated in diatom chloroplast proteomes (Kroth et al., 2008; S. R. Smith et al., 2012). Plastidial glycolysis, while seemingly contradictory (i.e.- catabolising carbon in the same compartment that it is fixed by the Calvin Cycle) can occur via the recruitment of two proteins to the plastid, enolase and phospho- glycerate mutase (henceforth cpEnolase and cpPGM) that can act as a bridge between endogenous enzymes of the Calvin Cycle (in the reverse direction to carbon fixation) and the pyruvate hub (implicated in shikimate and plastid fatty acid biosynthesis) (Bromke, 2013; Huang et al., 2023). The roles of plastid glycolysis nonetheless remain elusive, in plants for example being chiefly restricted to sink (root and seed) tissue and *Arabidopsis* mutant lines for both proteins having very limited phenotypes (Andriotis et al., 2010; Prabhakar et al., 2010).

My curiosity to explore diatom chloroplast glycolysis at the bench was provoked by our preceding bioinformatic studies. Both cpEnolase and cpPGM enzymes are recovered within the ancestral ochrophyte proteome and were inferred based on phylogeny to be derived from the host (Dorrell, Gile, et al., 2017; Dorrell & Bowler, 2017), but subsequent analyses from Tomomi Nonoyama projected a slightly more recent origin of the diatom isoform, most likely

in the diatom, pelagophyte and dictyochophyte common ancestor (Nonoyama et al., 2019). Considering Phaeonet data, both enzymes were recovered in the “blue” module containing much of the Calvin cycle and mitochondrial sugar metabolism, suggesting at a transcriptional level that this pathway is coexpressed with diatom carbon fixation (Ait-Mohamed et al., 2020). Nonetheless, considering data from Shun Liu and other researchers, the substrate (glyceraldehyde-3-phosphate) and product (phosphoenolpyruvate, or “PEP”) of diatom cpEnolase and cpPGM may be exchanged with the cytoplasm via chloroplast triose phosphate: phosphate transporters (S. Liu et al., 2022; Moog et al., 2020), and the significance of being able to directly interconvert both metabolites in the chloroplast was thus unclear.

Firstly, I considered diatom plastid glycolysis at a bioinformatic level, constructing localisation-aware trees of cpEnolase and cpPGM from taxonomic libraries explored in Tomomi Nonoyama’s work, and then evaluating *Tara* Oceans expression trends using phylogenetically-aware techniques developed with Shun Liu for her work on chloroplast transporters (S. Liu et al., 2022; Nonoyama et al., 2019) (**Fig. 3.7**). These data broadly suggest a recurrent origin of cpEnolase and cpPGM across the algal tree of life, but a specific duplication and recruitment of the host derived, respiratory enzymes to the chloroplast of the diatom and pelagophyte common ancestor (**Fig. 3.7A**). Thus, chloroplast glycolysis sits amongst one of the metabolic innovations that differentiates the chloroplasts of diatoms and pelagophytes from those of other photosynthetic stramenopiles (and, as we show elsewhere in the same paper, may have been transferred into haptophytes via a cryptic endosymbiosis, discussed more formally in Chapter 4). Considering *Tara* Oceans expression data, the environmental homologues of both chloroplast glycolysis proteins are predominantly expressed in high latitude stations, considering the relative abundance of meta-transcriptome (metaT) sequences that phylogenetically reconcile with genes encoding diatom chloroplast-targeted proteins (**Fig. 3.7B**). Although the predominant occurrence of these meta-genes at high latitudes is not surprising as diatoms are most abundant in these stations, a range of different normalisation techniques (e.g., comparing chloroplast glycolysis metaT to meta-genome/ metaG data, and to all diatom metaT sequences, and performing absolute and parametric correlation tests), applied with the aid of Juan Pierella Karlusich from the Bowler group, suggested that the relative expression level of both genes correlates positively to latitude and environmental daylength, and negatively to temperature (Dorrell, Zhang, et al., 2024).



B Tara Oceans metaT total abundances

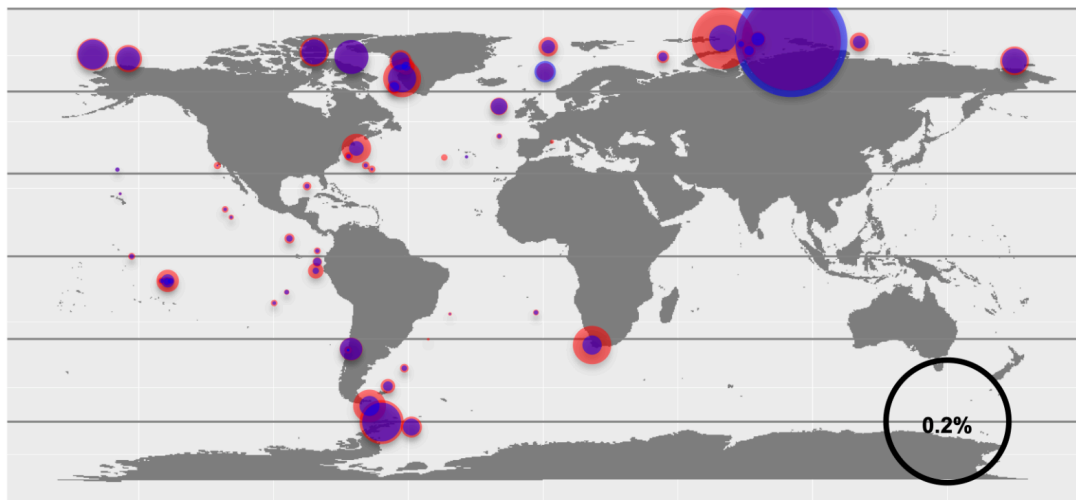


Fig. 3.7. Computational insight into diatom chloroplast glycolysis evolution and ecology. A: GFP localisations (top) and localisation-aware phylogenies (bottom) of diatom cpEnolase (Phatr3_J41515) and cpPGM1 (Phatr3_J17086). Scale bars in GFP images are 10 μ m. GFP fluorescence (green) coincides with chlorophyll (red), confirming plastidial localisation. Both trees suggest duplication and chloroplast retargeting of respiratory (mitochondria-targeted) glycolytic isoforms in a diatom/ pelagophyte common ancestor. **B:** *Tara* Oceans relative meta-transcriptome abundances (expressed as a % of total mapped metaT reads) of meta-genes phylogenetically reconciled to diatom cpEnolase and cpPGM1. Both genes show greatest relative metatranscriptome relative abundances in high latitude (Arctic, Southern Ocean and South Atlantic Ocean) stations.

Next, I phenotyped *Phaeodactylum* CRISPR knockout mutants for two genes encoding cpEnolase (Phatr3_J41515) and one isoform of cpPGM (Phatr3_J17086) (**Fig. 3.8**). These analyses, were performed collaboratively with supervised students and with multiple researchers from outside my group (Dorrell, Zhang, et al., 2024). The CRISPR mutants were predominantly generated and purified by Tomomi Nonoyama, and complementation and growth analyses under a range of conditions were performed by multiple undergraduate and Masters interns, including Mathias Penot-Raquin (**Fig. 3.8A**). Alongside growth rate, we integrated photophysiology (with Benjamin Bailleul and Dany Croteau, IBPC), gene expression (Yue Liang, Institute of Oceanography, Qingdao), lipid profiles (Eric Maréchal and Juliette Jouhet, CEA Grenoble), metabolite profiling and enzyme kinetics (Alisdair Fernie and Youjun Zhang, MPI Potsdam) to build a holistic model of the metabolic impacts of diatom chloroplast glycolysis (**Fig. 3.8B**). These analyses were performed under varying daylength and temperature based on environmental correlations of glycolysis meta-gene expression and latitude. The three tested conditions: 19°C and light: dark cycling (“19C LD”); 19°C and continuous light (“19C CL”); and 8°C and continuous light (“8C CL”).

These data broadly suggest that plastidial glycolysis has different roles dependent on environmental day-length and temperature (**Fig. 3.8**). Briefly, under the “default” 19C LD conditions, the mutant lines show a very limited phenotype, with perhaps a 10% reduction in growth rate compared to control lines. At a metabolic and transcriptomic level, this primarily seems to relate to a downregulation in the expression of chloroplast shikimate biosynthesis and accumulation of chloroplast TAGs, and increase in the expression of mitochondrial glycolysis and chloroplast galactolipid synthesis, consistent with an inhibition of flux in a predominantly glycolytic direction (Bromke, 2013; Huang et al., 2023).

Under 19C CL conditions, the growth defect is augmented to 30%, suggesting increasing importance of plastid glycolysis under continuous light. We observe more dramatic mutant phenotypes, including a general upregulation of most chloroplast photosynthesis and primary metabolism genes, under-accumulation of cytoplasmic sugars, and overaccumulation of small amino acids (e.g., valine, ornithine) implicated in chloroplast: mitochondrial metabolite shuttling (Broddrick et al., 2019). We propose that the enhanced growth defect reflects an inability to effectively divert plastidial photosynthate into PEP pools in mutant lines, necessitating the supply of PEP via mitochondrial glycolysis and respiration, and reducing the degree of excess triose phosphate available for storage sugar synthesis in mutant lines.

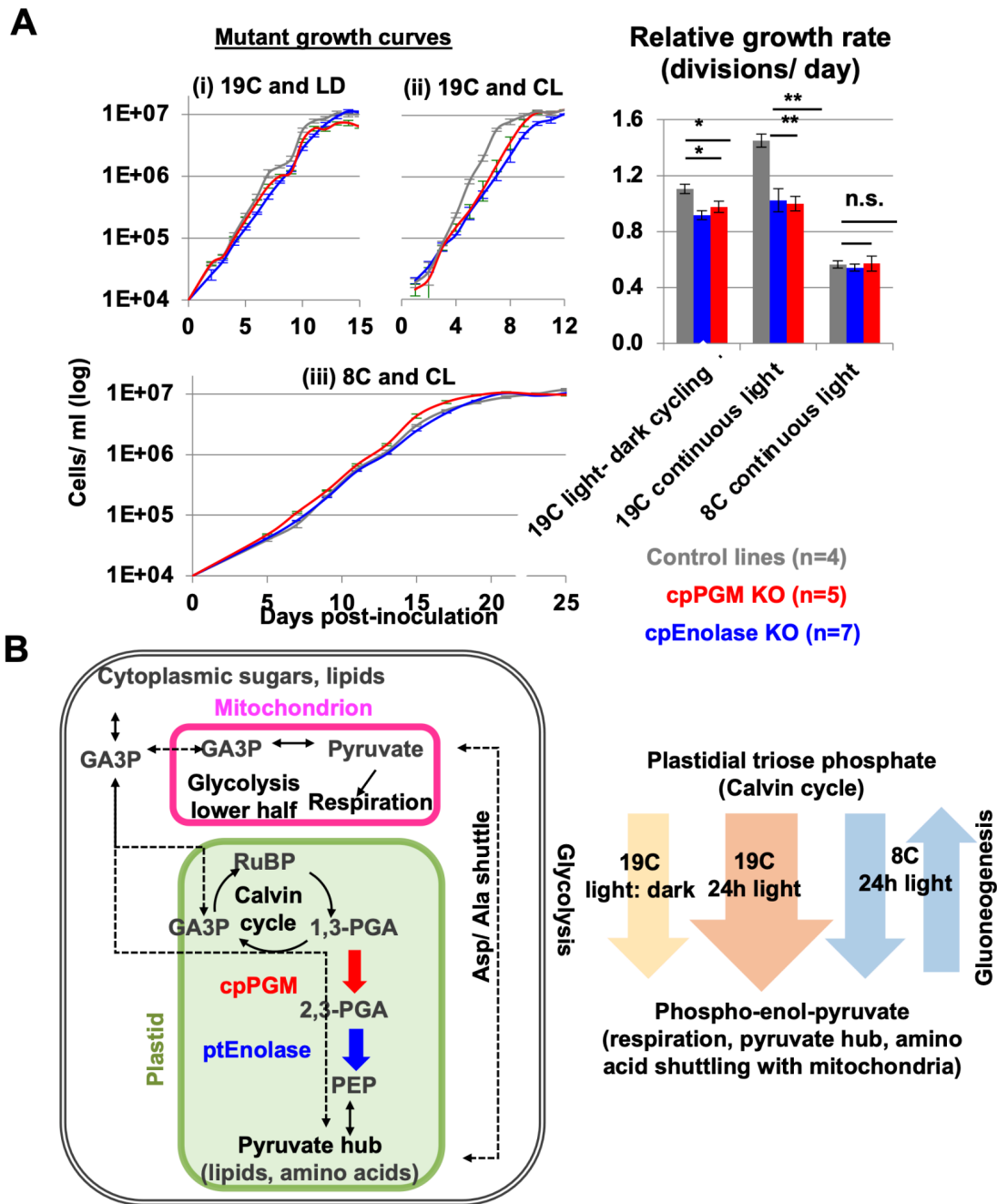


Fig. 3.8. Phenotypes of Phaeodactylum chloroplast glycolysis knockout lines. **A:** growth curves realised for homozygous cpEnolase (Phatr3_J41515) and cpPGM1 (Phatr3_J17086) CRISPR knockouts and empty vector (Cas9 and zeocin-resistant) controls. At least five genetically non-identical mutants were generated for two CRISPR regions of each gene, and growth curves were repeated three to eight times for each culture and condition. Log-phase growth rates were measured using flow cytometry in ESAW medium under $50 \mu\text{E m}^{-2} \text{s}^{-1}$ cool white illumination, under 12h: 12h light: dark cycling or continuous illumination, at 19°C and 8°C. Growth curves were performed with a starting density $10,000 \text{ cells ml}^{-1}$, with growth rates measured between $50,000$ and $4,000,000 \text{ cells ml}^{-1}$. Complementation of mutant lines with blasticidin-resistance linked copies of each gene restored the control growth phenotypes suggesting an absence of off-target CRISPR activity. **B:** an integrative metabolic model for the roles of diatom chloroplast glycolysis mutants under each growth condition, integrating photo-physiological, RNAseq, qPCR, metabolomic, and lipid data, and measured reaction kinetics of each enzyme expressed in *E. coli*.

Finally, at 8C CL conditions, we see a complete suppression of the growth difference between knockout and control lines. This is accompanied by the overaccumulation of lipids (TAGs) and metabolites (aspartate) synthesised from the pyruvate hub side (PEP) of plastid glycolysis (Yu et al., 2022), and we project that the chloroplast glycolysis has a secondary, gluconeogenic role at low temperatures, impeded in the knockout lines. Finally, in response to reviewer comments, we have confirmed the reversibility of both diatom cpEnolase and cpPGM1 (Sutherland et al., 1949), suggesting that they can function in both the glycolytic and the gluconeogenic direction dependent on substrate availability and metabolic context.

This project was time- and labour-consuming, and were I to perform it again it would have been designed differently, with fewer mutants, faster evaluation of knockout phenotypes, and stable isotope labelling of metabolic fluxes in lieu of total cellular metabolite measurements (Godrijan et al., 2020). It nonetheless gave insights into how to work sensitively with a consortium of researchers from different backgrounds, and integrate multiple viewpoints into the manuscript without occluding the science. At a pedagogical level, the growth analyses were my first exposure in supervising BTS students (Amandine Baylet, Yonna Lauruol, Quentin Caris), and I particularly enjoyed the challenge of articulating my science concisely and meaningfully in a different language to my own, for the benefit of young students with a technical vision and ambitions that are different- but complementary- to mine.

3.5. Ongoing characterisations of novel molybdenum transporters and ATP synthases associated with the pan-algal chloroplast

Finally, I will briefly outline two projects continuing in my group, using combined phylogenetic, environmental, physiological, and biochemical approaches to characterise novel proteins targeted to the diatom chloroplast. Both proteins are shared with other marine algae (particularly those with secondary red chloroplasts) but are absent from plants and from the primary red algae. The initial work on these projects has been performed in a largely *Phaeodactylum* background, although I am keen to integrate work on other model systems (e.g., *Chlamydomonas*). Each project has been led by PhD students in my group (Shun Liu, and Mathias Penot-Raquin) and I have been privileged to see them develop the scope of their work beyond what I initially envisioned.

Shun's PhD project, which was successfully defended in summer 2023, relates to a novel chloroplast major facilitator supergroup family transporter (provisionally named "Trans3", Phatr3_J18945). The major facilitator supergroup transporter (PFAM07690) superfamily is highly diverse, including sugar, lipid, amino acid, secondary metabolite, siderophore and some metal ion transporters (Yan, 2013). Prior to Shun's arrival, I had begun to explore diatom chloroplast transporters with Tomomi Nonoyama, selecting Trans3 for deeper investigation due to its presence in the photosynthetic stramenopile common ancestor and *Phaeonet* blue module, and transcriptional co-regulation with the mitochondrial genome (**Figs 3.3-3.6**). Tomomi generated a GFP-fusion construct for Trans3, where it seemed to sit between the *Phaeodactylum* chloroplast and mitochondria, and my initial hypothesis for Shun's PhD was that it was likely to be a novel sugar or amino acid transporter involved in chloroplast: mitochondria metabolite exchange (Broddrick et al., 2019; Prihoda et al., 2012).

Shun's PhD overturned most of these ideas (**Fig. 3.9**). Using self-assembling GFP, Shun showed that Trans3 did not specifically localise at the chloroplast: mitochondria contact points, but rather localises to the diatom periplastid membrane (which, in confocal images, frequently sits between but the plastid and mitochondria), with the N-terminus positioned in the cER lumen and the C-terminus in the periplastid compartment (**Fig. 3.9A**) (Gould et al. 2006). Generating heterozygous Trans3 CRISPR knockdown mutants, Shun saw limited evidence for dysregulation in sugar or amino acid metabolic pathways, but rather in photosynthesis, with her knockdown lines showing limited ability to downregulate photosystem I reaction core expression in response to Fe restriction (**Fig. 3.9B**) (Bey et al., 2022). These latter analyses were performed with cooperation of Giovanni Finazzi and Mattia Storti at the CEA Grenoble (S. Liu et al., 2022).

In parallel, Giovanni, Mattia and Shun made a surprising observation: that Trans3 resembles, at least superficially, the *Chlamydomonas* plasma membrane molybdate transporter MOT2 (Tejada-Jiménez et al., 2011) (**Fig. 3.9C**). In the final year of her PhD, Shun adapted a molybdate uptake assay for *Phaeodactylum*, with her knockdown lines showing greater uptake of molybdate from media than controls (**Fig. 3.9D**). These data position Trans3 as a chloroplast molybdate transporter, which to our knowledge has not been documented elsewhere in the tree of life (Gasber et al., 2011; Tejada-Jiménez et al., 2011).

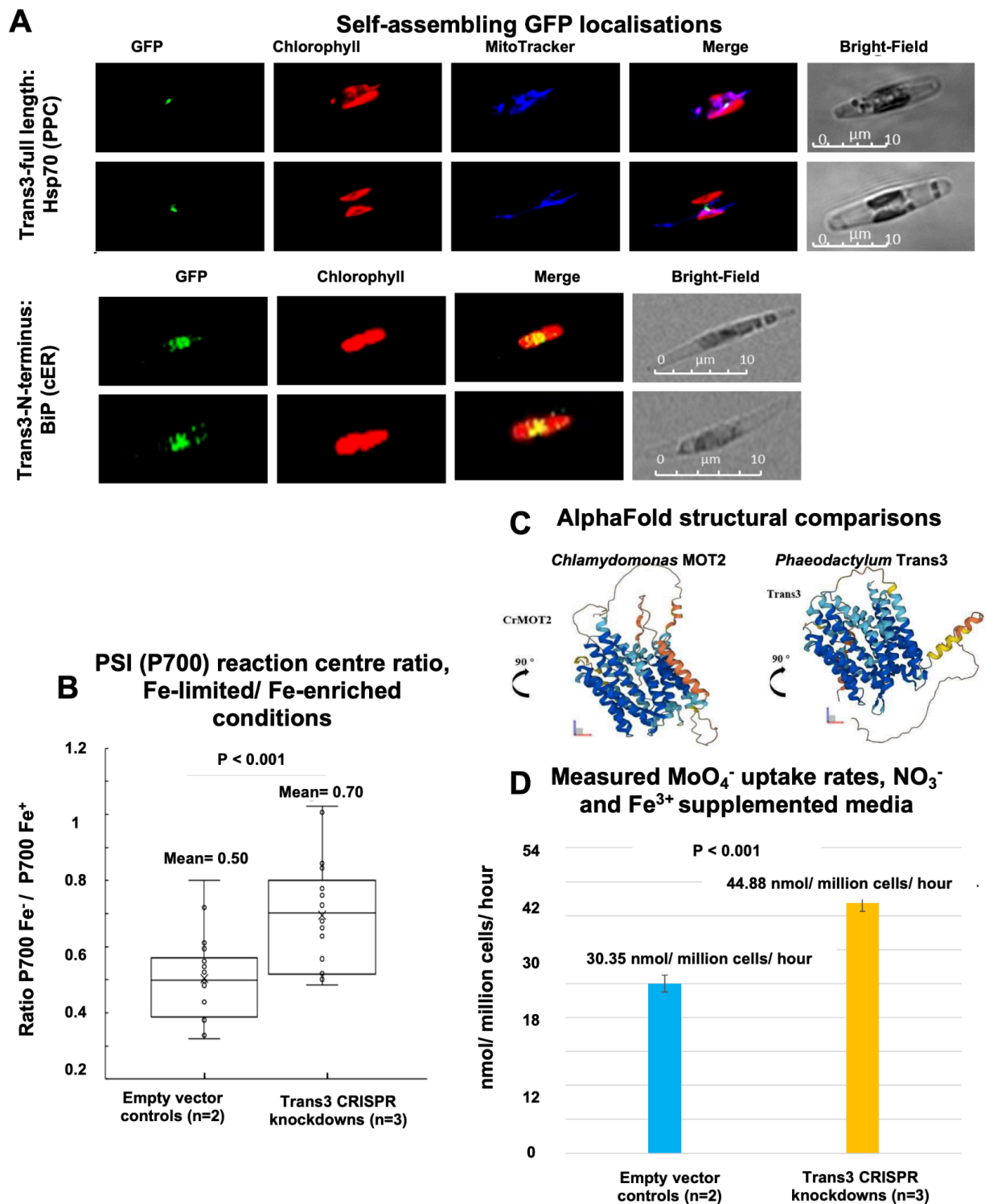


Fig. 3.9. Experimental insights into the function of Trans3 in the diatom chloroplast. **A:** self-assembling GFP constructs of the Trans3 full length protein and N-terminal regions, generated following the methodology of (Moog et al., 2020) and showing coincident localisations with the PPC marker Hsp70 and the cER marker BiP respectively. **B:** measured ratio of PSI P700 reaction centres of cell cultures (10^6 cells ml^{-1}) exposed cultured either under Fe-replete conditions or following three days of culturing in Fe-free media per (Bey et al., 2022), sensitised with 20 μM DCMU, and measured using a JTS-10 spectrophotometer following the methodology of (Bailey et al., 2008). **C:** AlphaFold structural comparisons (Jumper et al., 2021) of *Chlamydomonas reinhardtii* MOT2 (geneID: CHLRE_13g580150v5) (Tejada-Jiménez et al., 2011) and *Phaeodactylum* Trans3 at 90° rotation, showing effective

similarity between both proteins. **D**: measured molybdate uptake rates of Trans3 knockdown versus control cell cultures (10^6 cells ml^{-1}) incubated for 24 hours with high concentrations of nitrate (880 μM), iron (8.69 μM) and molybdate (918 nM) following two days of molybdate starvation. Molybdate concentrations were determined from cell supernatants by colorimetry of green-yellow molybdenum-dithiol complexes generated by incubation with strong sulfuric acid, extracted into isoamyl acetate, and enumerated with a reaction mixture of sodium hydroxide, dimercapto-4-methylbenzene and thioglycolic acid, following the methodology of (Cardenas & Mortenson, 1974). Molybdate uptake was determined deductively (i.e., by measuring the difference in supernatant molybdate concentrations at the end versus beginning of the incubation) and divided by the final cell concentration measured by flow cytometry, measured following (Dorrell, Zhang, et al., 2024). All panels in this figure are reprinted from the PhD thesis of Shun Liu (Université Paris-Saclay, 2023).

I was sceptical of the proposed molybdate transport activity of Trans3, particularly given the absence of known molybdenum-containing proteins in diatom or indeed other chloroplasts (Schwarz & Mendel, 2006). Shun was courageous to persevere with these final analyses from her PhD. I am now committed to validating experimentally if Trans3 transports molybdate, which I am testing with a research engineer in my group, Victoria Powell, via *Chlamydomonas* and yeast transfection. However, I was impressed to see Shun gain autonomy and direct her own research, particularly in maintaining her own collaborations with the Finazzi group at the CEA Grenoble, and in supervising two undergraduate students (Pauline Clémente, Morgane Roquais) who helped her with some of the cloning and molecular biology in her PhD.

In a final project, led by Mathias Penot-Raquin, we are exploring the biology of a novel homologue to the chloroplast ATP synthase a subunit (henceforth, "xATPA"). xATPA was first identified by Anna Novák Vanclová in the plastid proteomes of both the secondary green alga *Euglena gracilis* in her PhD, as a nucleus-encoded and chloroplast targeted protein, contrasting with the conventional ATPa encoded in all studied photosynthetic chloroplast genomes (Green, 2011b; Novák Vanclová et al., 2020). Anna identified a *Phaeodactylum* xATPA homologue (Phatr3_J31465) as part of the Phaeonet project (Ait-Mohamed et al., 2020). Mathias first joined my group for a Masters internship, performing the initial CRISPR mutagenesis and bioinformatic analysis of the *Phaeodactylum* xATPA gene alongside growth characterisation of plastid glycolysis mutants detailed above, and is now in the second year of a PhD with me on the same subject (Dorrell, Zhang, et al., 2024).

Mathias' data suggest that xATPA is widely spread across the algal tree of life, and may have spread with the chloroplast, e.g. with the *Euglena* isoform derived from green algae, and the Kareniacean isoform from haptophytes (**Fig. 3.10A**). Nonetheless, it is not detected in plants,

red algae, most plastid-lacking eukaryotes or bacteria, and forms a separate family to both chloroplast (photosynthetic) and mitochondria (respiratory) ATPA. Mathias has performed environmental and transcriptomic profiling of xATPA, suggesting that like plastid glycolysis the diatom xATPA meta-transcriptome homologues are predominantly expressed at high latitudes. Mathias has also generated CRISPR knockout lines, which show a limited phenotype barring a potential loss of growth under combined low temperature, continuous illumination and low salinity, which may reflect the environmental conditions in which xATPA is highly expressed (**Fig. 3. 10B**). These insights are currently being supported by transcriptomic profiling of xATPA knockout lines under different light, temperature and salinity regimes.

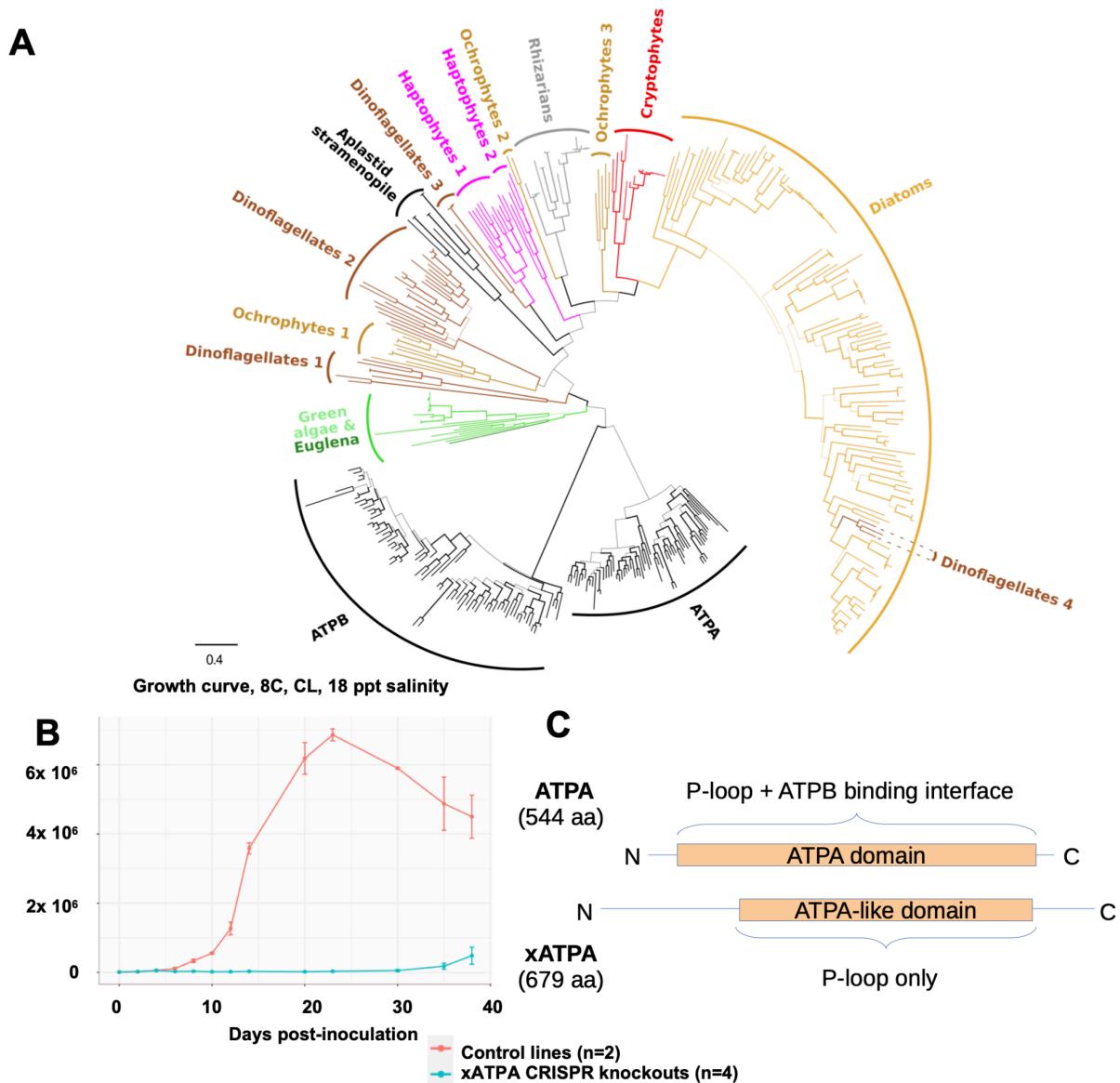


Fig. 3.10. Distribution, phenotype and structure of xATPA. **A:** Consensus MrBayes topology realised with three substitution matrices (GTR, Jones, WAG) of a 197 taxa x 490 aa alignment of ATPA and ATPB sequences (PFAM: 00006) from across the tree of life, rooted on the division between ATPA and ATPB. xATPA forms an outgroup to all conventional (chloroplast and mitochondrial) ATPA sequences although with greater sequence similarity than that

of ATPB, and is distributed across algal groups and protists (labyrinthulomycetes) but absent from other lineages. **B:** growth curve of homozygous xATPA CRISPR knockouts (four lines, targeting three different CRISPR regions) compared to empty vector controls under low temperature, continuous light, and low salinity, suggesting effective loss of growth in xATPA mutants. **C:** Schematic alignment of *Phaeodactylum* chloroplast-encoded ATPA and xATPA sequences, demonstrating the presence of a complete P-loop but absence of ATPB-binding domain in xATPA.

For the second half of his PhD, Mathias is keen to explore the biochemistry of xATPA in more detail, for example via expression in *E. coli* using similar techniques to those from the plastidial glycolysis project (Dorrell, Zhang, et al., 2024). Mathias has noted that xATPA lacks a predicted ATPB-binding domain and is unlikely to function in ATP synthesis or hydrolysis, but we both wonder whether it can bind ATP with an affinity relevant to the chloroplast (**Fig. 3.10C**). A secondary question may be the localisation of homologues of xATPA from different species, in particular given it appears to be present in one plastid-lacking stramenopile group (labyrinthulomycetes) that per our and other data have never possessed a red chloroplast (Rius et al., 2023; Stiller et al., 2009; Q. Wang et al., 2017) (**Fig. 3.10A**). It will be interesting to see what functions xATPA has in these species, and indeed if in photosynthetic organisms it may have moonlighting functions outside the plastid. A final question, given the probable basal position of the green algae to other xATPA sequences (**Fig. 3.10A**), may be to reassess the significance of the green signal in diatom evolution from a functional level. For this, Mathias has already developed some scripts for the high-throughput enumeration of environmental correlations of genes in *Tara* Oceans data (used, for example, to assess global distributions of haptophyte algal orders (Penot et al., 2022)). I feel these approaches could transform our understanding of not only how many green genes have been transferred to diatom genomes, but to what extent it has impacted on their functional eco-physiology.

3.6. Concluding remarks and pedagogical insights

My research on the diatom chloroplast has evolved over time, particularly in terms of the technical framing and the questions posed. I was initially interested purely in its evolutionary history, e.g. the enumeration and timing of red and green signals (**Figs. 3.2, 3.3**), but this has given me a basis to explore function, e.g. identifying protein innovations that differentiate diatoms from other stramenopiles (**Figs 3.1, 3.4, 3.6, 3.7**). On more technical grounds, my initial evolutionary work has provided a basis for quantifying environmental expression of chloroplast proteins with greater precision (e.g., via phylogenetic reconciliation of meta-transcriptome homologues (**Figs. 3.6, 3.7**)). Both this, and the expression networks I have

been able to exploit from *Phaeodactylum* transcriptome data have been very useful in defining experimental approaches for and understanding protein function (**Figs. 3.5, 3.8**). I note that other groups are in parallel exploiting environmental data to pinpoint conditions (e.g., high temperatures) for *Phaeodactylum* mutant phenotyping in-lab (Gao et al., 2024).

Nonetheless, my interest in diatom chloroplasts remains primarily evolutionary. Exploring what biological functions different mosaic signals (e.g., green genes) may have conferred in diatom evolution has interested me since my undergraduate studies (Dorrell & Smith, 2011), and I feel that the most interesting outstanding questions from these projects are effectively taxonomic in nature. I am very intrigued by the evolution of a potential molybdate transporter (i.e., Trans3/ MOT2) in diatom chloroplasts (**Fig. 3.9**), as this may challenge quite a lot of what we know about how diatom chloroplasts function (e.g., as molybdate storage compartments) and indeed how molybdenum is partitioned inside living cells. Likewise, the most exciting feature for me of xATPA is its probable association with chloroplast endosymbiosis (i.e. being transferred from green algae to *Euglena*, and from haptophytes to Kareniaceae) but also its presence in a non-photosynthetic group (labyrinthulomycetes): i.e. that it is a protein co-inherited with, but not intrinsically linked, to the chloroplast (**Fig. 3.10**).

There are clearly some loose ends that I aim to resolve in the near future. The priority is completing the Trans3 work, for which Shun has left me with some good outstanding questions. Firstly, the evolutionary history of Trans3, given its homology to *Chlamydomonas* MOT2 (**Fig. 3.9**), remains to be verified. We have now searched for and localised *in silico* homologues of Trans3 across > 1,500 algal and plant genomes and transcriptomes incorporative of the OneKp transcriptome project (Carpenter et al., 2019). These data imply that Trans3 only has chloroplast-targeted homologues in secondary red chloroplast-containing algae (ochrophytes including diatoms, cryptomonads, haptophytes, dinoflagellates), has non- chloroplast-targeted homologues in green algae, and is absent from red algae. We might infer Trans3 to be a gene transferred from green algae into the endosymbiotic ancestor of the secondary red chloroplast, and recruited to the chloroplast in this lineage (see scenario D, **Fig. 3.2**), but this will formally need phylogenetic validation.

A second question is the localisations of other proteins involved in molybdate trafficking in diatoms. In her thesis, Shun annotated probable molybdate transporters, alongside molybdate-dependent enzymes present in *Phaeodactylum*, several we wish to now localise (**Table 3.1**). These include a second probable chloroplast-targeted MOT2 transporter

(Phatr3_J16421), and two probable plasma-membrane associated molybdate transporters (Phatr3_J38899, MOT1; Phatr3_J42677, MOT2_3) which likely mediate molybdate import into the diatom cell (Gasber et al., 2011; Tejada-Jiménez et al., 2011). To determine if molybdate functions indirectly in the chloroplast, we are investigating two proteins involved in molybdate complex (MoCo) synthesis and insertion (Phatr3_J20318, Cnx5_1; Phatr3_49807, MOSC_3) that may possess chloroplast-targeting sequences, and one final molybdate-dependent enzyme (sulphite oxidase, Phatr3_J44482) which is not predicted to be chloroplast-targeted *in silico*, but has been detected at relatively high abundance in mass spectrometry of *Phaeodactylum* chloroplast-enriched cell fractions (Huang et al., 2023).

GeneID	Name	Function	Localisation
Phatr3_J18945	MOT2_1/Trans3	Molybdate transport	Chloroplast
Phatr3_J16421	MOT2_2		Chloroplast
Phatr3_J42677	MOT2_3		Plasma Membrane
Phatr3_J38899	MOT1		Plasma Membrane
Phatr3_J15625	Cnx2	cPMP synthesis	Mitochondria
Phatr3_J10532	Cnx3		Cytoplasmic
Phatr3_J45404	ATM3	cPMP transport	Mitochondria
Phatr3_J9137	Cnx6	MPT synthesis	Cytoplasmic
Phatr3_EG00664	Cnx7		Cytoplasmic
Phatr3_J20318	Cnx5_1		Dual
Phatr3_J34373	Cnx5_2		Cytoplasmic
Phatr3_J13395	Cnx1G	MPT activation	Cytoplasmic
Phatr3_J44961	Cnx1(E+G)	Mo insertion	Endomembrane
Phatr3_J46641	ABA3	MoCo sulphuration	Cytoplasmic
Phatr3_J54983	Nitrate reductase	MoCo-dependent enzyme	Cytoplasmic
Phatr3_J15968	Xanthine dehydrogenase		Cytoplasmic
Phatr3_J44482	Sulphite oxidase		Cytoplasmic
Phatr3_EG01365	MOSC_1	MOSC-containing proteins	Endomembrane
Phatr3_J48037	MOSC_2		Endomembrane
Phatr3_J49807	MOSC_3		Chloroplast
Phatr3_J40382	MOSC_4		Cytoplasmic

Table 3.1 Putative molybdenum- associated proteins in the *Phaeodactylum* genome. This table is reprinted from the PhD thesis of Shun Liu (Université Paris-Saclay, 2023). Consensus *in silico* localisation predictions are provided following (Ait-Mohamed et al., 2020). Genes under investigation in my group are shown in red text.

The most urgent question is to confirm the biochemical activity of Trans3, for which we are using two approaches: expressing it in yeast, which classically lacks molybdate transport activities and can therefore be used to directly assay for molybdate uptake (Tejada-Jiménez et al., 2011), and complementation of an existing *Chlamydomonas reinhardtii* MOT2 (CLiP) mutant with the *Phaeodactylum* Trans3 sequence (X. Li et al., 2016). In general I am keen to explore the use of *Chlamydomonas* mutants more in my group. As one attempt to define a function for xATPA, Mathias has already identified phenotypes of *Chlamydomonas* xATPA mutants from

high-throughput screens (sensitivity to ultraviolet light) (Fauser et al., 2022), although was unable to detect a convincing phenotype in *Phaeodactylum* mutant lines using previously tested UV sensitivity assays (De Riso et al., 2009). Nonetheless I feel that forward mutagenesis and phenotyping, alongside other techniques (e.g., large-scale localisation) possible in *Chlamydomonas*, will be a very helpful complement to reverse genetic approaches in *Phaeodactylum* to identify protein function (L. Wang et al., 2023).

On a broader scale, I think that my group should begin exploring experimental systems beyond diatoms. This not only reflects an underlying interest in evolution but also the possibility to better understand protein function: we cannot fully understand how plastidial glycolysis, which has evolved independently in diatoms from plants, functions in diatoms based on our knowledge of Plant Physiol; and the same principle should hold true for our still relatively diatom-centric understanding of biochemistry in the secondary red chloroplast tree of life. For this I am encouraged by emerging transformable models in other marine algal groups (Faktorová et al., 2020), and as a preliminary test of this Mathias is planning to localise the labyrinthulomycete homologue of xATPA directly in the transformable species *Aplanochytrium* (Rius et al., 2023). Similarly, while we have exploited Tara Oceans data to date to predict the environmental significance of diatom genes (Dorrell, Zhang, et al., 2024; S. Liu et al., 2022), we could begin to explore how the expression of specific proteins varies amongst different taxonomic groups of phytoplankton, to uncover conserved and lineage-specific biological roles, e.g. using meta-genome assembled genomes (Delmont et al., 2022). Mathias and I have made some preliminary investigations, showing that xATPA expression shows positive correlations to latitude in multiple algal groups (e.g., cryptomonads, haptophytes), whereas plastid glycolysis gene expression is only correlated to latitude in diatoms.

On pedagogical grounds, I have found my first supervisions of PhD students to be deeply rewarding, with the challenge to balance providing support without constraining my students' ideas. For this, I take a lot of inspiration from the dynamic and friendly research environment that I experienced as an early career postdoc in the Bowler Group, where many of my most rewarding collaborations were organically developed with my colleagues. I hope that I can develop and guide a research ecosystem of similar poise and positivity in my own group.

Chapter 4: Phylogenetic and biogeographical understanding of horizontal evolution across algal pan-genomes

4.1. Quantifying the evolutionary impacts of inter-algal HGT

The endosymbiont hypothesis appeals to me not just as an explanation for the origins of mitochondria and chloroplasts, but also as a prime mover of genetic material (Sagan, 1967). Endosymbiosis involves integrating an entire symbiont genome into that of the host, with retention of at least the genes essential to the viability of the new chloroplast (Dorrell & Howe, 2012b). We might therefore view the endosymbiotic acquisition of chloroplasts as a specific form of horizontal gene transfer. If the cyanobacterial ancestor of the chloroplast had approximately 3,000 genes (e.g., 3,034 in the genome of *Gloeomargarita*) (Moreira et al., 2017; Ponce-Toledo et al., 2017) then the much smaller nature of chloroplast genomes (i.e., < 250 genes) (Green, 2011b) opens up the space for extensive gene transfer into the host nucleus. Phylogenetic analysis of primary chloroplast-containing organisms estimate large numbers of endosymbiotic gene transfers (e.g. 900 genes of probable cyanobacterial origin in the *Chlamydomonas* nuclear genome, or up to 2,000 genes of probable cyanobacterial origin in *Arabidopsis*), which is likely to be the largest single source of horizontally acquired DNA in the archaeplastids (W. Martin et al., 2002; Moustafa & Bhattacharya, 2008).

At the same time, there are probable limits to what is transferred from endosymbionts into their host genomes. One question is whether endosymbiont-derived genes only serve to support chloroplast and mitochondria, or can encode proteins that function in other cell organelles (Dorrell & Howe, 2012b). The majority (50-90%) of archaeplastid genes of cyanobacterial origin have inferred plastid functions (Ku et al., 2015; Reyes-Prieto et al., 2006), suggesting that endosymbiotic gene transfer predominantly serves to support chloroplast biogenesis. This concept may even be exploitable for identifying cryptic endosymbioses, e.g. green genes in diatom genomes predominantly encode chloroplast-targeted proteins, suggesting a chloroplast origin (Dorrell, Gile, et al., 2017; Moustafa et al., 2009). At the same time, within my own data, I have identified endosymbiotically derived genes with non-chloroplast localisations, e.g., aminoacyl-tRNA synthetases of red algal origin that exclusively support the *Paraphysomonas* mitochondrion (Dorrell, Azuma, et al., 2019; Gile et al., 2015). I have also identified genes encoding chloroplast-targeted proteins that probably had non-chloroplast origins in the endosymbiont, e.g. Trans3, a chloroplast-targeted

homologue of the *Chlamydomonas* plasma membrane molybdate transporter MOT2 (S. Liu et al., 2022; Tejada-Jiménez et al., 2011).

A second question is what factors limit the number of genes relocated from the endosymbiont to the host. We might expect secondary endosymbionts to contribute a huge number of genes to their host nucleus. The nuclear genomes of secondary endosymbionts are larger than the nucleoid genomes of bacteria (e.g., 5-10,000 genes in the red ancestor of the diatom chloroplast (Brawley et al., 2017), and the genes in these genomes are already under nuclear promoters and polyA elements necessary for expression in the host cell (Dorrell & Howe, 2012b; Nowack et al., 2011; Stegemann & Bock, 2006). Moreover, the symbiont nucleus is eliminated early in endosymbiosis, except in nucleomorph-harboring groups (Gould et al., 2006; Sarai et al., 2020). This lability may create conditions favourable to the lytic release and nuclear relocation of symbiont-derived DNA (Dorrell & Howe, 2012b). However, in my best estimates of the red signal in the *Phaeodactylum* genome, using over 20 red algal genomes and transcriptomes as a phylogenetic reference, I could not extend beyond 400-450 genes of probable red origin (Dorrell et al., 2022; Rastogi et al., 2018). Why there is so little red signal in diatoms is as interesting to me as whether the green signal (see Chapter 3) in these organisms really exists (Dorrell & Smith, 2011; Woehle et al., 2011).

Previous studies have identified phylogenetic signals linking the plastid genomes of distantly related secondary red chloroplasts in cryptomonads, haptophytes, ochrophytes and alveolates (Rice & Palmer, 2006; Ševčíková et al., 2015). Similar chloroplast-targeted proteins have been found in specific subsets of these lineages (Stiller et al., 2014) (**Fig. 4.1**). These signals do not correspond to the underlying nuclear phylogenies of these organisms (Strassert et al., 2021), suggesting successive tertiary and quaternary endosymbioses (**Fig. 4.1**). A progressive spread of the red chloroplast across the algal tree may provide an explanation for the limited red signal observed in some of these groups. For example, if the diatom chloroplast is a tertiary endosymbiont of cryptomonad origin (**Fig. 4.1**, points A, C), then the “red” genes in diatoms might be the result of two successive rounds of endosymbiotic winnowing (first in the secondary chloroplast-containing cryptomonads; then in the ochrophyte nucleus following tertiary endosymbiosis) (Stiller et al., 2014; Strassert et al., 2021). We could, perhaps, instead evaluate the size of the “cryptomonad signal” as a measure of endosymbiotic gene transfer in the diatom genome.

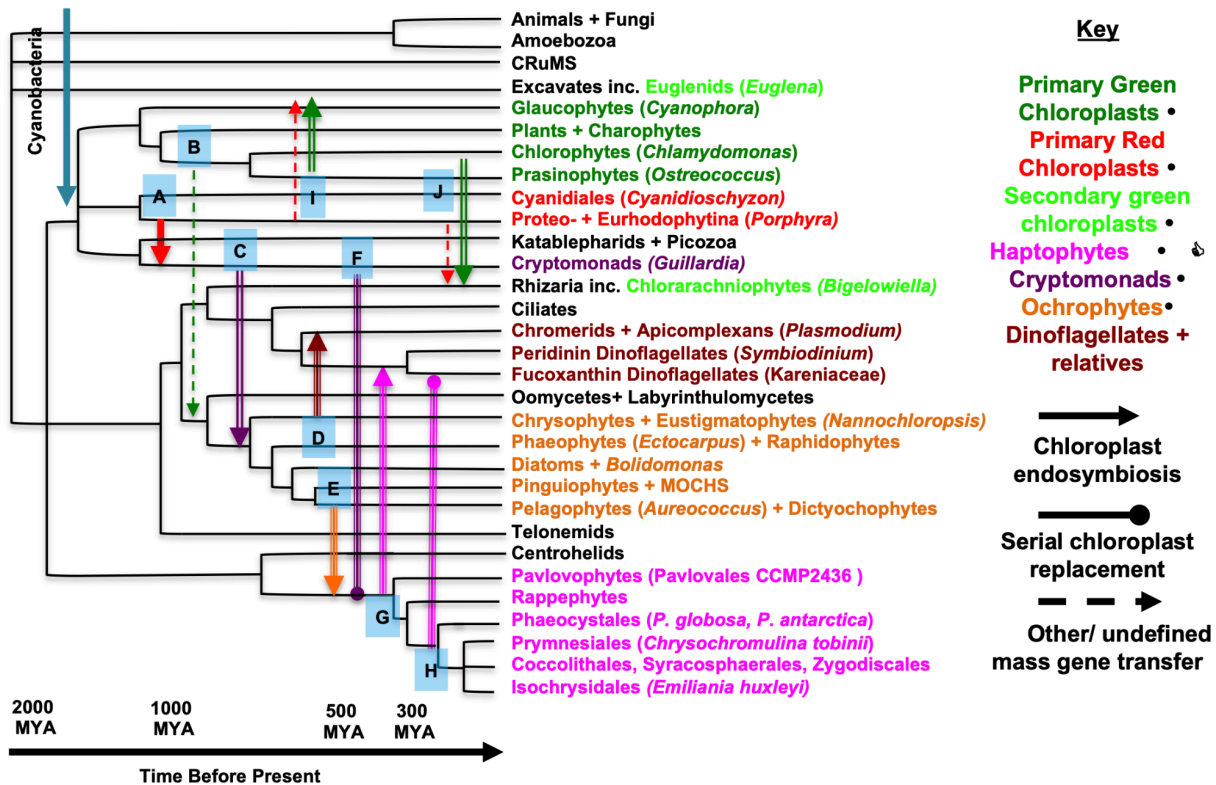


Fig. 4.1 Progressive origins of “secondary endosymbionts” across the eukaryotes. This figure, adapted from (Penot et al., 2022; Pierella Karlusich et al., 2022), shows a time-calibrated tree of eukaryotic nuclear evolution (timeline adapted from (Strasser et al., 2021)) with chloroplast-containing groups shown in coloured text. Type species with sequenced genomes for specific lineages are provided in italics. Ten potential instances of secondary or higher endosymbioses are highlighted with labelled arrows, with colours corresponding to the probable endosymbiont and number of lines the level of endosymbiosis (i.e., single line for primary, double line for secondary, etc). **A:** secondary endosymbiosis of a red alga into cryptomonads per (Cavalier-Smith, 1999; Green, 2011a). **B:** possible endosymbiont-derived green signal in diatoms and other ochrophytes per (Dorrell, Gile, et al., 2017; Moustafa et al., 2009). **C:** tertiary endosymbiosis of a cryptomonad into ochrophytes per (Stiller et al., 2014). **D:** quaternary endosymbiosis of a chrysophyte into chromerids per (Ševčíková et al., 2015). **E:** possible quaternary contribution of a pelagophyte/ dictyochophyte to the haptophyte plastid proteome per (Dorrell, Gile, et al., 2017). **F:** tertiary endosymbiosis of a cryptomonad contributing the current haptophyte plastid genome per (Rice & Palmer, 2006). **G:** possible quinary endosymbiosis of a haptophyte into the peridinin dinoflagellates per (Yoon et al., 2002). **H:** multiple haptophyte (prymnesiophyte) serial endosymbioses in Kareniaceae per (Kawachi et al., 2021; Novák Vanclová et al., 2024). **I, J:** secondary green endosymbioses of green algae, with possible contributions (“red carpet”) genes from red algae or their endosymbiotic derivatives, into euglenids and chlorarachniophytes, per (Novák Vanclová et al., 2020; Ponce-Toledo et al., 2019).

An alternative hypothesis is that historical or cryptic endosymbioses have saturated the host nuclear genomes of some algae containing “secondary chloroplasts”. These cryptic endosymbiotic gene transfers may limit the genetic contribution from their current chloroplasts. For example, the red signal in diatom genomes may have been limited by competition by genes derived from a preceding green endosymbiont (Fig. 4.1, point B). This

phenomenon is visible in Kareniaceae (see chapter 2; **Fig. 4.1**, point H) where we see a limited contribution of the haptophyte endosymbiont to the nucleus-encoded proteome, despite clear reductive evolution of its plastid genome (Klinger et al., 2018; Novák Vanclová et al., 2024). Similarly, algae with secondary green endosymbionts (cryptomonads and euglenids, **Fig. 4.1**, points I and J) may contain genes derived from red algae, in what has been termed the “red carpet” hypothesis (Novák Vanclová et al., 2020; Ponce-Toledo et al., 2019), which may constrain the size of the green signal in this lineage. All of these hypotheses are nonetheless limited by the absence (except in dinoflagellates) of clearly diagnosed serial endosymbiosis events in eukaryotic algae.

An additional question is *how* non-endosymbiotic horizontal gene transfers influence algal genome evolution. Without wishing to enter the sometimes fraught debate about how *much* horizontal gene transfers occur in eukaryotes (Leger et al., 2018; W. F. Martin, 2017), I note that some of the most compelling examples involve species from unusual habitats, *e.g.* hypersaline and thermophilic algae (Cho et al., 2023; Foflonker et al., 2018), and anaerobic protozoa living in sediments and as gut commensals (Eme et al., 2017; Stairs et al., 2020). Alongside considering how much horizontal gene transfer has occurred, and when during algal evolution, I think it is interesting to consider the functional contributions of horizontally acquired algal genes with regard to their environment. Understanding the biogeographical trends that underpin horizontal gene transfer in algae may be important to understanding both its frequency (two species have to live together to exchange genetic material) and selective consequences (adaptive functions conferred relative to the environment).

Finally, most algal species do not live in isolation. Most algae are probably mixotrophic, considering genome-scale predictive models and observations from natural communities (Bock et al., 2021; Unrein et al., 2014). The wild biology of algae is contextualised by interactions with commensals, epibionts, parasites and predators (Dittami et al., 2021; Gachon et al., 2010; Sato et al., 2020). Algal trophism and symbiosis are linked to one another, with different species known to acquire growth-limiting nutrients (*e.g.*, vitamin B₁₂) via scavenging, mutualism, and phagotrophy (Croft et al., 2005; Nef et al., 2019). These biotic interactions may provide a source for horizontal gene transfers but also complicate our enumeration of them, *e.g.* from biological contamination in genomes (Carpenter et al., 2019; Sato et al., 2020).

In this chapter, I outline projects from my group seeking to explore the impact of horizontal interactions on the genome evolution of eukaryotic algae, starting at an endosymbiotic level

and ending with more environmentally-aware questions of recent changes in the gene content of individual species. While my analyses are not explicitly aware of what biotic interactions underpin these exchanges, I hope they may act as starting points for future projects to explore the eco-physiological implications of horizontal gene transfer. One feature that particularly appeals to me is the significance of culture collection sites, which may act as a starting point for ongoing work to understand algal pan-genomes at a biogeographical level.

4.2. Identification of hidden chloroplast links between stramenopiles, haptophytes and dinoflagellates

The first project began with the construction of the ancestral ochrophyte plastid proteome (Dorrell, Gile, et al., 2017). While annotating trees for the 770 chloroplast protein groups in this study, I observed that many showed a paraphyly of ochrophytes relative to other “secondary” chloroplast-containing algae, e.g. cryptomonads, haptophytes, and dinoflagellates (**Fig. 4.2**). These signals were not viewed as disrupting the ancestrality of these proteins, which I had assessed in terms of ochrophyte monophyly relative to archaeplastids (red and green algae), bacteria, and the stramenopile host. I nonetheless attempted a preliminary annotation of the positions of other secondary chloroplast-containing algae to the ochrophytes, considering direction (transfers into a monophyletic ochrophyte clade, and transfers out of ochrophytes into other algae yielding a paraphyletic ochrophyte clade) and closest sister-group.

The strongest signal from the ochrophytes connected them to a distantly-related algal group, the haptophytes. Haptophytes are single-celled, planktonic and mixotrophic algae that are globally distributed and abundant in marine habitats (Penot et al., 2022). Haptophytes are defined by the “haptonema”, a filamentous organelle that sits between two flagella of equal length (Eikrem et al., 2016). Some haptophytes, e.g. *Emiliania huxleyi* are calcifiers, producing blooms that can be visible from space (Burki, Hirakawa, et al., 2012; Vincent et al., 2021). Haptophytes are distantly related to other algae, forming a sister-group to the “SAR clade” of stramenopiles, alveolates and rhizaria, and acquired their chloroplasts independently to ochrophytes (Stiller et al., 2014; Strassert et al., 2021). Haptophytes were already of interest to us both considering ecology, as they are the second most abundant algal group in *Tara* data after diatoms; and also evolution, as the progenitors of the Kareniacean plastid (*see* chapter 2; **Fig. 4.1**) (Novák Vanclová et al., 2024; Pierella Karlusich et al., 2023).

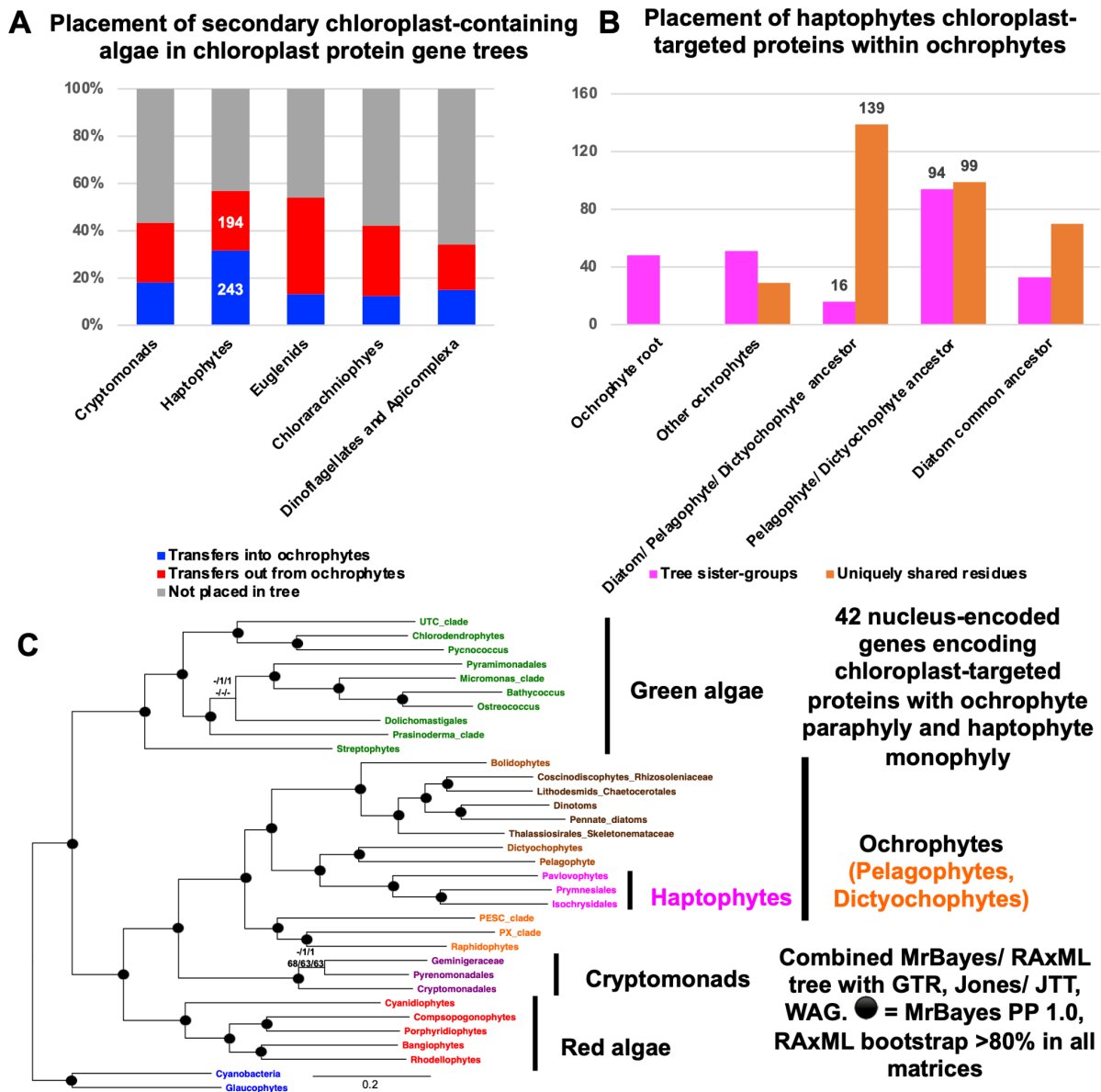


Fig. 4.2. Evidence for a partial pelagophyte/ dictyochophyte origin to the haptophyte plastid proteome. This figure summarises data originally presented in (Dorrell, Gile, et al., 2017). **A:** phylogenetic positions of five algal groups with secondary chloroplasts relative to the ochrophytes in consensus RAXML bootstrap trees (PROTGAMMAJTT matrix) of 770 nucleus-encoded and chloroplast-targeted proteins inferred to be shared across all ochrophytes, in which they are monophyletic to the exclusion of red algae, green algae, bacteria, and non-photosynthetic stramenopiles. **B:** specific phylogenetic positions recovered for haptophytes within the 243 trees in which they position inside the ochrophytes. **C:** concatenated topology of 42 of these genes in which all ochrophyte and haptophyte orders are represented respectively as paraphyletic and monophyletic clades.

Amongst our data, haptophytes were the only algal group placed more often inside (243 gene trees) than outside (194 gene trees; **Fig. 4.2A**). A finer analysis of the sister-group positions in these trees, and the residues uniquely shared between haptophyte and ochrophyte orthologues strongly suggested a placement related to the pelagophytes and dictyochophytes, or potentially the common ancestor of these and diatoms, but excluded a

more ancient or alternative origins (e.g. from phaeophytes, chrysophytes, or within diatoms; **Fig. 4.2B**). Finally, a concatenation of haptophyte genes of ochrophyte origin positioned them as a robustly-supported sister-group to the pelagophyte / dictyochophyte common ancestor (**Fig. 4.2C**, MrBayes PP = 1.0, RAxML bootstrap > 80% with three substitution matrices).

These data imply that the ochrophyte nucleus-encoded and chloroplast-targeted proteome, and more specifically that of a pelagophyte/ dictyochophyte ancestor, has been shared with haptophytes. We also showed that haptophyte proteins of ochrophyte origin frequently have chloroplast-targeting signals, as might be expected for genes acquired via a chloroplast endosymbiosis. Haptophyte, pelagophyte/ dictyochophyte and diatom chloroplasts further share biochemical features (e.g., use of the accessory pigments diatoxanthin/ diadinoxanthin (Andersen, 2004)- and proteins (e.g., lower half plastidial glycolysis, and enzymes involved in chlorophyll c and fucoxanthin/ diatoxanthin biosynthesis) not found elsewhere amongst secondary red chloroplasts (Bai et al., 2022; Dorrell, Zhang, et al., 2024; Jiang et al., 2023; Nonoyama et al., 2019). More recently, we extended the pelagophyte/ dictyochophyte/ haptophyte signal to the peridinin dinoflagellates, considering a more global study of horizontal gene transfer into and out of the *Phaeodactylum* genome, discussed below (**Fig. 4.3**) (Dorrell et al., 2021) and I suspect that the peridinin plastid may also be of haptophyte origin.

The implications of these signals are a probable origin of at least the haptophyte chloroplast-targeted proteome, and potentially that of dinoflagellates as well, via a tertiary or higher endosymbiotic acquisition of a pelagophyte/ dictyochophyte common ancestor (**Fig. 4.1**, point E). This transfer is plausible in terms of timing, with molecular clock analyses suggesting that the pelagophyte/ dictyochophyte ancestor probably existed at the time that haptophytes acquired their plastids (Strassert et al., 2021). I find this association particularly exciting as the groups it unifies, diatoms and pelagophytes/ dictyochophytes (as closely-related members of the ochrophytes), and haptophytes and dinoflagellates (as endosymbiotic derivatives) are the most abundant eukaryotic algae in the ocean (Penot et al., 2022; Pierella Karlusich et al., 2023), alongside chlorophytes. I wonder if proteins specifically shared between these groups account in part for this ecological prominence.

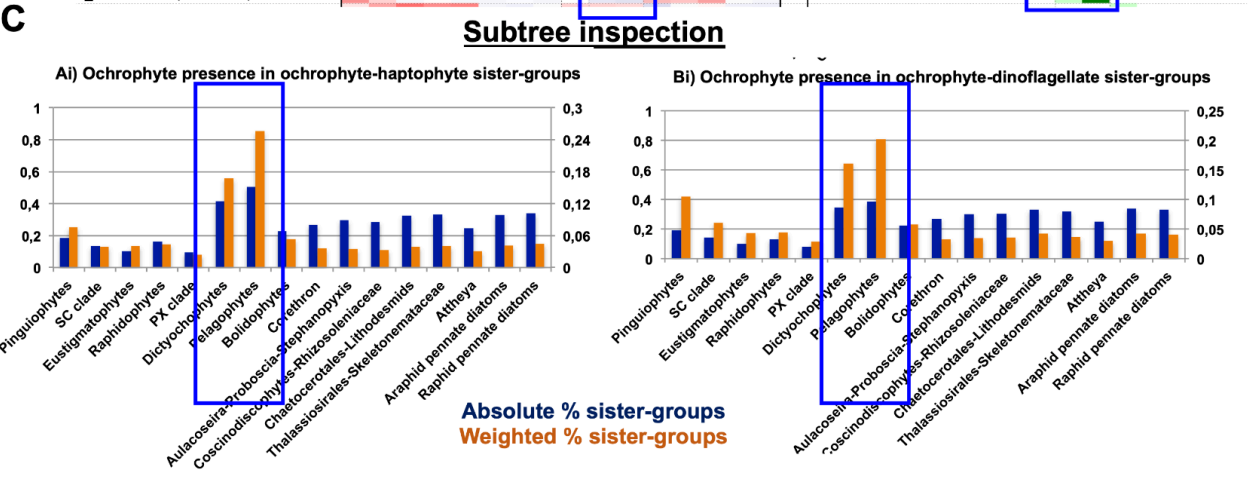
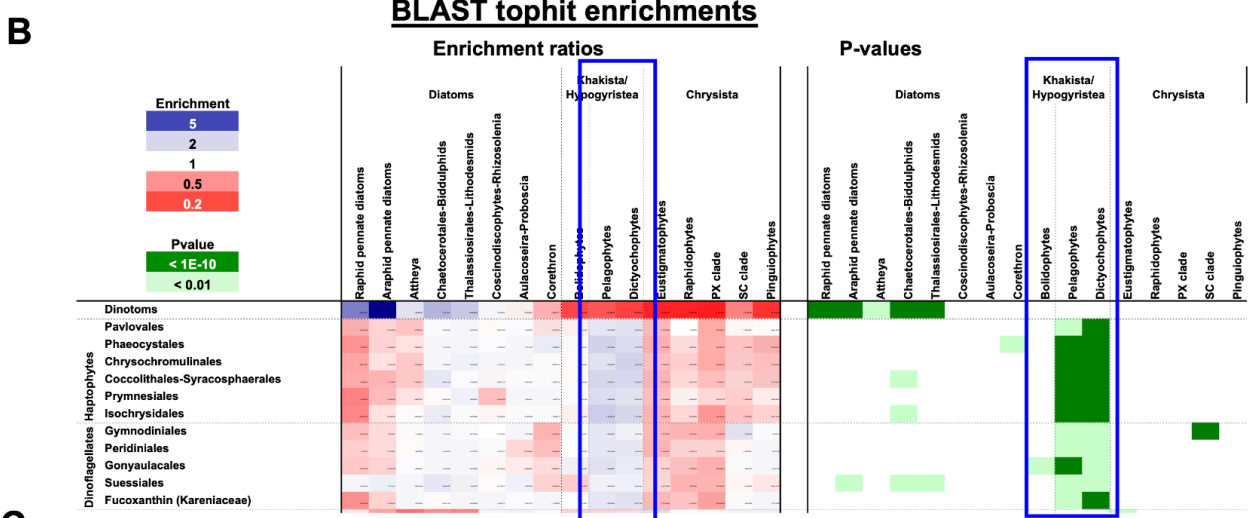
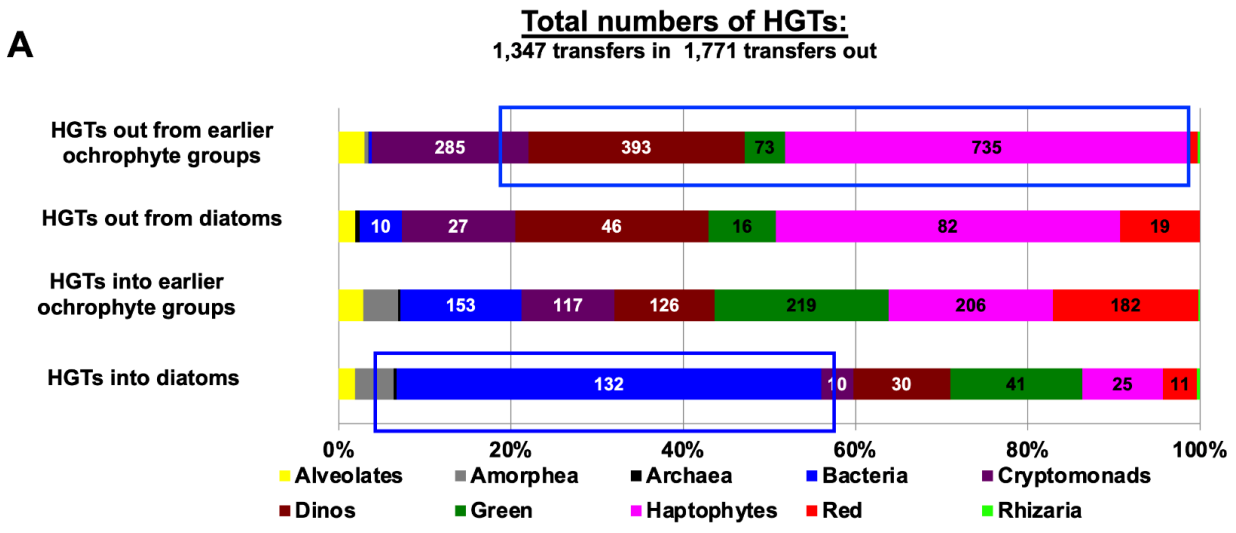


Fig. 4.3. Evidence for a shared genetic signal between pelagophytes/ dictyochophytes, haptophytes and peridinin dinoflagellates. **A:** Summary of > 3,000 horizontal gene transfers into and out of the *Phaeodactylum* genome, manually annotated from RAXML bootstrap consensus trees of single-gene trees constructed from reciprocal BLAST best-hit homologues. Both haptophytes (735 gene trees) and dinoflagellates (393 gene trees) frequently position inside the ochrophytes before the origin of diatoms (highlighted in blue box). Notably also one-half of horizontal gene transfers that have occurred into the diatom common ancestor or later are of bacterial origin (highlighted in blue box). **B:** Best hits of dinoflagellate and haptophyte sequences from each alignment used to construct the single-gene trees in Fig (A) , when searched by internal BLASTp against ochrophyte homologues

from the same alignment. Sequences from haptophytes and all dinoflagellate orders (including peridinin-containing and Kareniacean species) more frequently retrieve pelagophyte and dictyochophyte best hits than expected through random chance (highlighted in blue box, two-tailed χ^2 test, $P < 0.05$), except for dinotoms (dinoflagellates with diatom symbionts) that show a corresponding strong enrichment in diatom best hits (Imanian et al., 2012; Yamada et al., 2017). **C**: positions of haptophytes (left) and dinoflagellates (right) in *Phaeodactylum* gene trees where they branch inside the ochrophytes. The *absolute* frequency (blue bars, plotted on the left-hand y-axis of each graph) is the proportion of trees where each ochrophyte class is found in the haptophyte or dinoflagellate sister-group, and the *weighted* frequency (orange bars, plotted on the right-hand y-axis) is the same statistic but divided by the number of ochrophyte classes found in the sister group (e.g., a sister-group only composed of one ochrophyte class would be scored as 1, a sister-group composed of two classes as 0.5 for each class, and so forth). In both plots haptophyte and dinoflagellate sister-groups are heavily biased towards pelagophytes and dictyochophytes (highlighted in blue box).

A second point of interest is the discrepancy between the signal from the nucleus-encoded and chloroplast-targeted proteins, and the chloroplast genomes of haptophytes. It is clear that haptophyte chloroplast genomes are not of pelagophyte/ dictyochophyte origin, resolving outside of the ochrophytes in chloroplast genome phylogenies (Dorrell, Gile, et al., 2017; Khan et al., 2007; Le Corguillé et al., 2009). Pelagophyte and dictyochophyte chloroplast genomes also contain fewer genes and have lost features (e.g., the inverted repeat) retained in haptophyte chloroplasts, so I find it improbable that they could have acted as a donor of the haptophyte chloroplast genome (Han et al., 2019; Ong et al., 2010). Instead, the haptophyte chloroplast genome resolves as a sister-group of that of cryptomonads, sharing features (e.g., a bacterially-derived *rpl36* gene) not found in other lineages (Khan et al. 2007; Rice and Palmer 2006). These data suggest a serial origin of the haptophyte chloroplast with a pelagophyte/ dictyocophyte derived proteome and cryptomonad-derived genome, extending our view of serial endosymbiosis beyond tractable cases in the dinoflagellates (**Fig. 4.1E, F**).

4.3. Frequency and localisation biases contextualise bacterial horizontal gene transfer into diatom genomes

The second project is a more systemic evaluation of horizontal gene transfers involving diatoms. This began with work on the version 3 *Phaeodactylum tricornutum* genome (Rastogi et al., 2018), where we evaluated the total red, green and bacterial signals via a crude BLAST best-hit approach. As a follow-up, I performed a more comprehensive overview of gene transfers into and out of *Phaeodactylum* extending to the last common ochrophyte ancestor (Dorrell et al., 2021). The workflow used was to retrieve probable homologues of every *Phaeodactylum* gene from a densely sampled tree of life, including uniprot, JGI genome,

MMETSP and OneKp transcriptomes, separated into 151 taxonomic groups; then manually annotate single-gene trees to explore the global dynamics, timing and direction of possible horizontal gene transfers. I was particularly interested to see to what extent diatoms and their relatives act as donors, as well as recipients, of horizontal gene transfers. Extending from previous enumerations of red, green and bacterial content in diatom genomes (Fan et al., 2020; Moustafa et al., 2009; Vancaester et al., 2020), I was also more broadly interested to determine to what extent algae with secondary chloroplasts participated in these transfers.

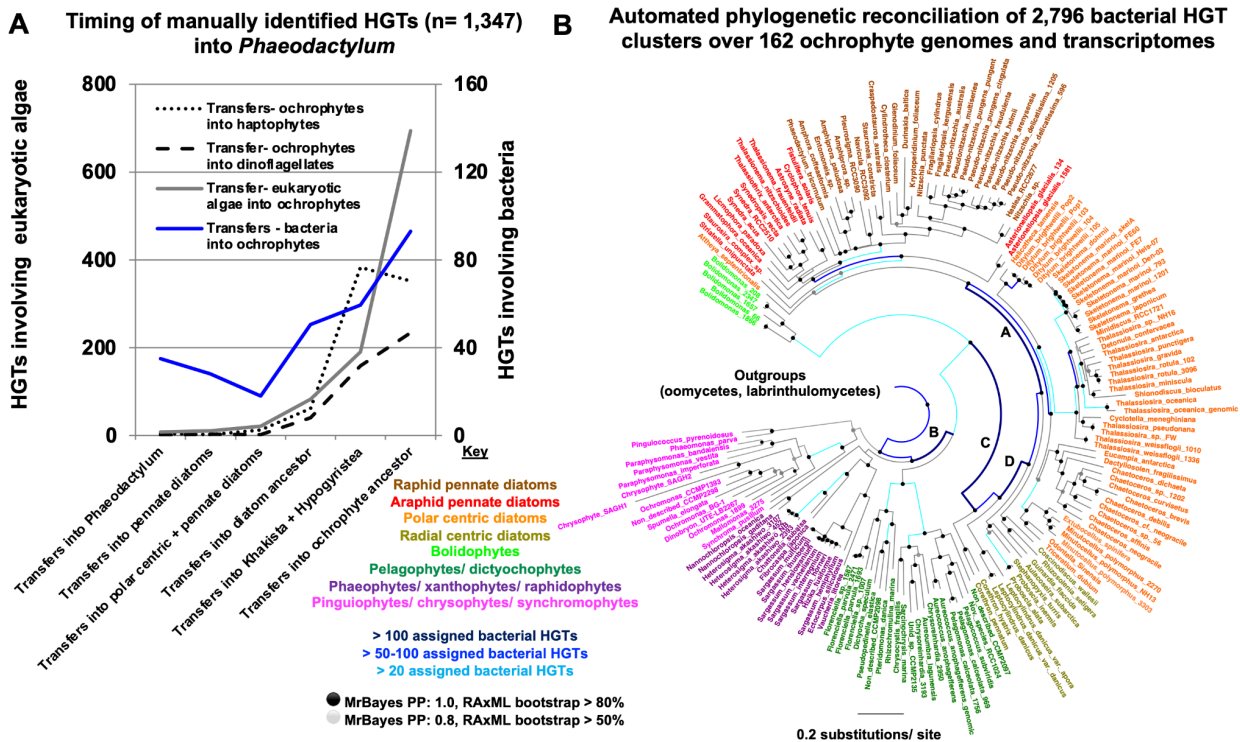


Fig. 4.4. Evidence for largely continuous transfers into diatom genomes through manual and automated phylogenetic reconciliation. **A:** timing of the 1,347 manually identified HGTs into the *Phaeodactylum* genome shown in Fig. 4.3A and separated based on apparent timing, from those specific to *Phaeodactylum*, up to the common ancestor of all ochrophytes. Four major groups of transfers are shown: HGTs from bacteria and from all eukaryotic algae (red and green algae, cryptomonads, dinoflagellates and haptophytes) into ochrophytes; and HGTs from ochrophytes into dinoflagellates and haptophytes only. Fewer HGTs occur involving bacteria than other groups so these are plotted on the right-hand y-axis. **B:** distributions of 2,796 bacterial HGTs identified by automated phylogenetic approaches (using IQTree and ModelFinder) to be shared between at least two of 162 ochrophyte genomes and transcriptomes. HGTs are reconciled by a most recent-common ancestor approach onto the consensus tree of a 63 gene/ 26, 399 aa concatenated multigene alignment (generated using MrBayes and RAxML, with substitution matrices GTR, Jones/ JTT and WAG). Nodes are shaded by the amount of bacterial HGTs identified. Four nodes with > 100 inferred HGTs are labelled: A- polar/ pennate diatom common ancestor, 429 HGTs; B- Khakhista/ Hypogyristea common ancestor- 152 HGTs; C- diatom common ancestor- 106 HGTs; D- common ancestor of Coscinodiscophyceae, Rhizosoleniaceae, polar centric and pennate diatoms- 102 HGTs.

The data obtained suggested a diverse range of gene transfers involving eukaryotic algae in diatom evolution (**Fig. 4.3A**). Most of the phylogenetically confirmed transfers (850/ 967) from algae into *Phaeodactylum* occurred in an ochrophyte common ancestor (**Figs. 4.3A, 4.4A**). These genes had a plurality of different sister-groups, with effectively equal representation (> 100 genes each) of cryptomonads, dinoflagellates, haptophytes, primary green and red algae (**Fig. 4.3A**), which neither confirms nor refutes the hypothesis that the ochrophyte plastid is of cryptomonad origin (Dorrell et al., 2021; Stiller et al., 2014). Most of the gene transfers from ochrophytes into other algae involved haptophytes and/ or dinoflagellates, and were enriched around the pelagophyte/ dictyochophyte common ancestor (**Fig. 4.4A**). Gene transfers specific to diatoms (117 transfers in from eukaryotic algae, 190 transfers out) were much fewer than those assigned to earlier time points in ochrophyte evolution, suggesting genetic isolation of diatoms from other algae in their recent history (**Figs. 4.3A, 4.4A**).

Alongside these predominantly algal HGTs, the *Phaeodactylum* HGT dataset contained a secondary signal of bacterial origin, with a distinct dynamic. Somewhat smaller in size than algal HGTs (285 gene transfers in, 12 gene transfers out), it is consistent with independent estimates of bacterial gene content (50-500 genes) in the *Phaeodactylum* genome (Fan et al., 2020; Vancaester et al., 2020). While the algal signal predominantly relates to the ochrophyte ancestor with a steep diminution within the diatoms, bacterial HGTs are detected with almost equal frequency in diatoms (132 transfers) to earlier timepoints (153 transfers), and are the single largest recent source of horizontally acquired genes in *Phaeodactylum* (**Figs 4.3A, 4.4A**).

At this point, I was able to establish a collaboration with Guillaume Blanc at Université Aix-Marseille, who has a longstanding interest in the phylogenetic evaluation of horizontal gene transfer (Blanc et al., 2007; Maumus & Blanc, 2016). Guillaume's group performed a census of bacterial HGTs across 161 further ochrophyte genomes and transcriptomes to *Phaeodactylum*. We mapped these transfers onto an ochrophyte multigene tree prepared by my former L3 student Guillemette Audren de Kerdrel (**Fig. 4.4B**) (Dorrell, Azuma, et al., 2019; Dorrell et al., 2021). This complementary approach demonstrated a continuous wave of horizontal gene transfers from bacteria into ochrophytes, with greater numbers of events associated with diatom origins (e.g., the diatom common ancestor, or the common ancestor of polar centric and pennate diatoms) than elsewhere in the tree (**Fig. 4.4B**). A similar dynamic (i.e., continuous bacterial gene transfer in ochrophytes, with greater transfers into diatoms)

has been recovered by an independent study using genome data only (Vancaester et al., 2020).

I am nonetheless agnostic about the idea that diatoms have received greater amounts of bacterial HGT than other algae. There is no apparent difference in the numbers of bacterial genes in diatoms versus other algal genomes (Fan et al., 2020), and the number of bacterial HGTs assigned to the polar / pennate diatom common ancestor alone in our data (462) is greater than the number (289) identified in the entire *Phaeodactylum* genome (**Fig. 4.3A, 4.4B**). If there is a greater frequency of bacterial HGT arrival in diatoms, it should logically be counterbalanced by a greater frequency of HGT loss to explain these distributions. More broadly, our reference libraries include many more diatoms (107) than other ochrophytes (55), and we may have more frequently detected bacterial HGTs in diatoms due to sampling bias.

To explore this possibility, I collaborated with Benoît Perez-Lamarque and H el ene Morlon at the IBENS, who employed a range of phylogenetic reconciliation techniques to dissect the HGT signal from the automated phylogenies (Dorrell et al., 2021; Perez-Lamarque & Morlon, 2019). Beno t confirmed that the elevated bacterial HGT signal in diatoms cannot be explained by inter-diatom HGT or gene duplications, considering the topologies of individual gene trees; it is not related to library completeness, as it is present even following correction for BUSCO coverage; it is detectable in subsets of libraries with equal sampling of diatoms and non-diatom ochrophytes; and it is greater than the frequencies of HGT possible into diatoms by chance, assessed by randomisation of leaf nodes. I concede that our data do suggest elevated frequencies of gene transfers from bacteria into diatoms. I nonetheless would be happy to see our data independently verified, with a greater number of non-diatom ochrophytes from the very first step of the HGT identification.

Beyond this, our data revealed interesting asymmetries in the encoded functions of algal versus bacterial-derived genes. Comparing both manually annotated *Phaeodactylum* HGTs and the automatically identified bacterial HGTs across ochrophytes, and our previous annotations of the *Phaeodactylum* genome (e.g., Phaeonet, see Chapter 3), genes acquired from eukaryotic algae were inferred to frequently encode chloroplast-targeted proteins, relate to photosynthesis, possess activatory histone marks (e.g., H₃K₄ methylation, and H₃K₉ acetylation), and are highly expressed in published *Phaeodactylum* transcriptome data (Ait-Mohamed et al., 2020; Dorrell et al., 2021; Rastogi et al., 2018). Combined with the predominant reconciliation of these to the ochrophyte common ancestor, these imply that

much of this signal relates to genes acquired from the algal ancestor of the ochrophyte chloroplast, encoding essential proteins in plastid metabolism and biogenesis (Fig. 4.4A).

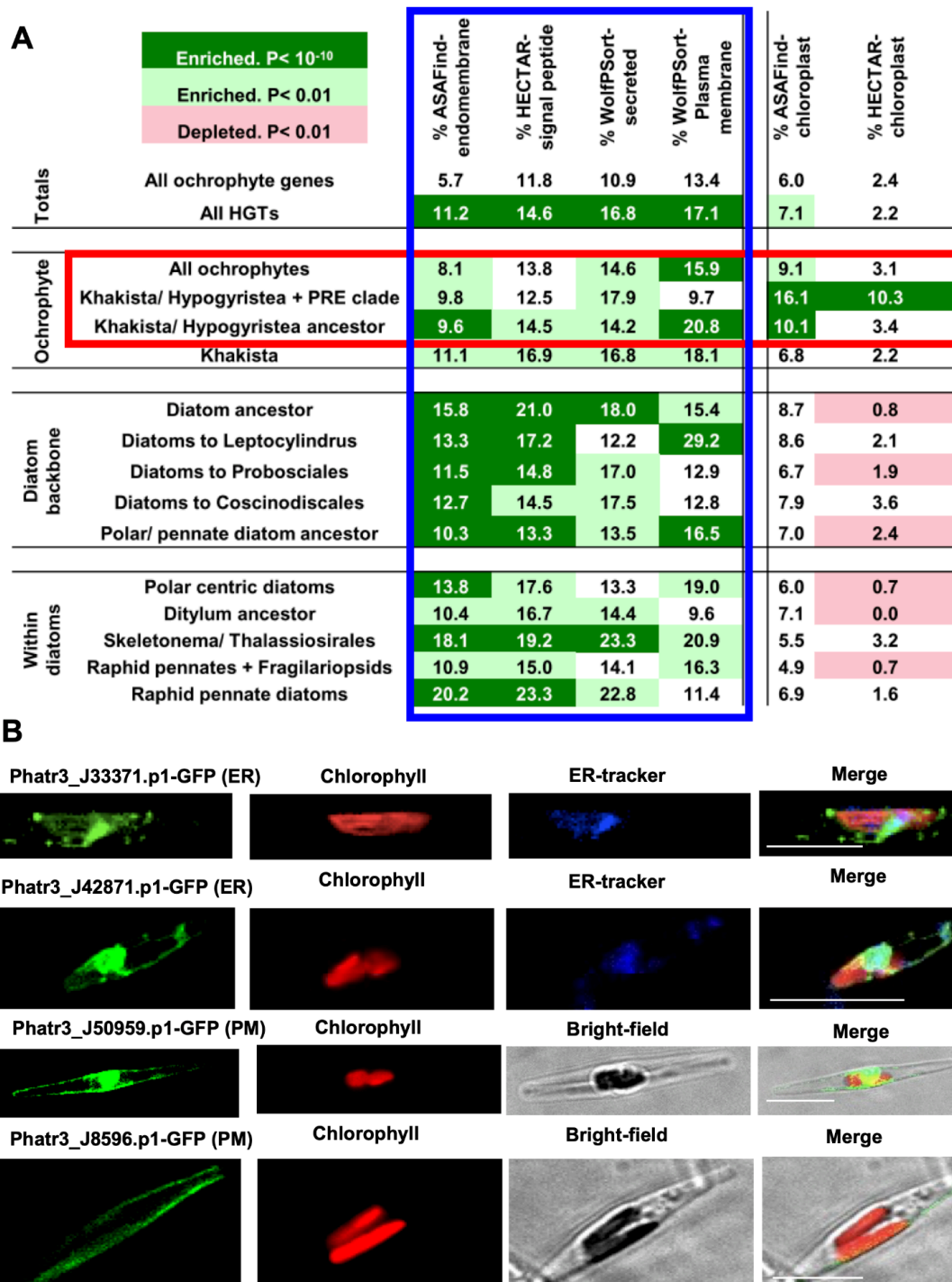


Fig. 4.5. Targeting bias in bacterial HGTs into diatom genomes. A: frequencies with which 2,796 bacterial HGTs identified by automated phylogenetic reconciliation across 162 ochrophyte genomes and transcriptomes possess secretory or chloroplast localisations, inferred by *in silico* prediction. Cells are shaded by P-value of enrichment (two-tailed χ^2 -test) above expected frequencies. Only HGTs reconciled to timepoints extending from the ochrophyte common ancestor to Khakista/ Hypogyristera common ancestor are enriched in chloroplast-targeting sequences (red box), whereas all HGTs irrespective of timepoint are enriched in inferred secretory localisations (blue box). **B:** experimental localisations of full-length: C-terminal GFP fusions for four *Phaeodactylum* genes of

inferred bacterial origin and either endomembrane or plasma membrane localisations. Endomembrane-targeted GFP-fusion proteins are colabelled with ER-tracker blue, following (Shao et al., 2023).

Phaeodactylum bacterial HGTs are lowly expressed and are not enriched in activatory histone marks. Across our data, only those that are associated with the ochrophyte common ancestor are enriched in chloroplast-targeting sequences, while those of more recent origin are not (**Fig. 4.5A**). This implies that the very first bacterial HGTs acquired by ochrophytes were associated with a chloroplast endosymbiosis, while subsequent transfers were not.

Instead, we noted a sustained bias in ochrophyte genes of bacterial origin to encoding proteins with signal peptides, but not chloroplast-targeting or type II signal anchor sequences (**Fig. 4.5A**). This would be consistent with predominant localisations in the endomembrane system, plasma membrane, or secretome. We localised some of these bacterially-derived proteins to the *Phaeodactylum* ER and plasma membrane using GFP, for which I was aided by Kyle Frischkorn within the Bowler group (Dorrell et al., 2021) (**Fig. 4.5B**). Enrichments in secretory targeting motifs have been inferred in HGT studies from other protists (e.g., the gut commensal *Blastocystis*, and in a broader pan-eukaryotic study of viral HGTs (Eme et al., 2017; Irwin et al., 2022), and I wonder if this bias occurs more broadly in eukaryotic evolution.

4.4. Convergence and “Arctic-specific” genes in an Arctic pan-algal genome

Following this work, I was interested to see how HGT in algal species is influenced by biogeography. This coincided with a project between Connie Lovejoy at the Université Laval and the JGI, to sequence the first algal genomes from the Arctic Ocean (Dorrell et al., 2023).

The Arctic is surrounded almost entirely by continents, which isolate it from global ocean circulation (Richter et al., 2022). A heavy influence of continental freshwater produces much lower salinities (15 to 20 ‰) than the 35 ‰ found in most other oceans (Freyria et al., 2022). Arctic water masses are constitutively cold (typically -2 to 4 °C), and are subject to extreme light cycles, from continuous dark in the polar winter to continuous illumination in summer (Joli et al., 2023; Leu et al., 2015). The chemistry of the Arctic changes over annual cycles, with relatively well-mixed and nutrient-rich early spring waters becoming more oligotrophic through the summer by the deposition of nutrient-poor freshwater and meltwater in the surface photic layer. The polar night finally curtails photosynthetic activity in the autumn (Ardyna et al., 2014; Leu et al., 2015). All of these parameters are being impacted radically by human activity, with the Arctic heating at up to twice the other rate of other ocean water

masses, accompanied by declining sea ice cover, increased precipitation, decreased salinity and more oligotrophic summer waters (Ardyna & Arrigo, 2020; W. K. W. Li et al., 2009).

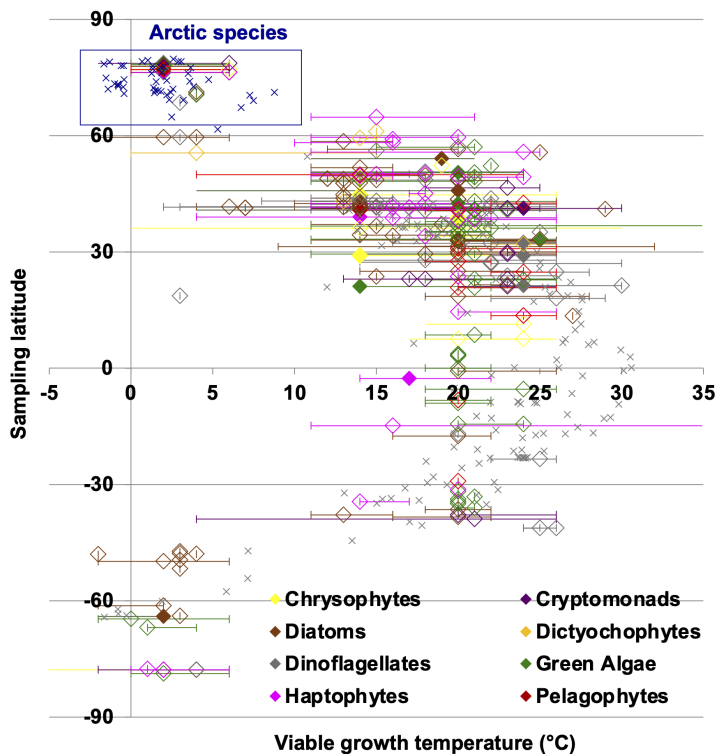


Fig. 4.6. Different thermal niches of Arctic versus non-Arctic algae.

This plot shows the collection latitude (vertical) and growth temperatures (horizontal) of 317 algal species with sequenced genomes (solid points) or transcriptomes (open points) (Grigoriev et al., 2021; Keeling et al., 2014). The growth temperature for each species is centred around the sequenced culture, with viable growth ranges from culture collection records shown as error bars. Equivalent latitudes and measured temperatures are shown as crosses for sampled *Tara* Oceans stations, (Ibarbalz et al., 2023). Arctic stations and species are highlighted with a dark blue box.

The isolated, extreme and fragile environment of the Arctic supports a distinctive biome with defined regional and seasonal variation. The majority of Arctic algae are restricted to this environment, or are uniquely found in the Arctic and Antarctic Southern Ocean (Ibarbalz et al., 2023). While they come from many different lineages that have colonised the Arctic independently (Gérikas Ribeiro et al., 2020), they are characterised by growth temperature optima well below those of species from lower latitudes (Fig. 4.6). Diatoms predominate in the most nutrient-rich spring waters, giving way to smaller flagellated algae in the more oligotrophic summer conditions, and finally dinoflagellates with greater capacity for mixotrophy in the autumn (Freyria et al., 2021; Leu et al., 2015; W. K. W. Li et al., 2009). This biome is changing rapidly, with fragility of many Arctic endemic algae to the predicted impacts of climate change (Chaffron et al., 2021) and more frequent incursions of species from the North Atlantic and Pacific (Anderson et al., 2021).

This project was phrased around annotation of four new algal genomes from the Arctic, isolated by the Lovejoy lab, and sequenced by the JGI: a cryptomonad (*Baffinella* sp. CCMP2293), a haptophyte (*Pavlova* sp. CCMP2436), a chrysophyte (*Ochromonas* sp. CCMP2298) and a pelagophyte (nov. gen. sp. CCMP2097) (Grigoriev et al., 2021; Hamilton et

al., 2008). My initial responsibilities were to quantify HGTs into these genomes, adapting pipelines I had developed for *Phaeodactylum* (Dorrell et al., 2021; Rastogi et al., 2018). I also generated multigene and single-gene phylogenies for each species; confirmed Arctic endemism using phylogenetically reconciled 18SV4, V9 and 16S V4V5 rDNA *Tara* Oceans ribotypes; and identified specific PFAMs and gene families associated with Arctic species (Dorrell et al., 2023; Mistry et al., 2021). These analyses used a custom dataset of 317 algal genomes and transcriptomes from eight different lineages (the four above, plus diatoms, dictyochophytes, dinoflagellates and green algae) with sequenced Arctic relatives (Fig. 4.6). I am grateful for aid from Federico Ibarbalz and Juan José Pierella Karlusich from the Bowler Group, Zoltán Füssy from Charles University, my former student Elisabeth Richardson and Nikola Zarevski, an M2 student who performed some of the phylogenetic analyses in 2020 (Dorrell et al., 2023).

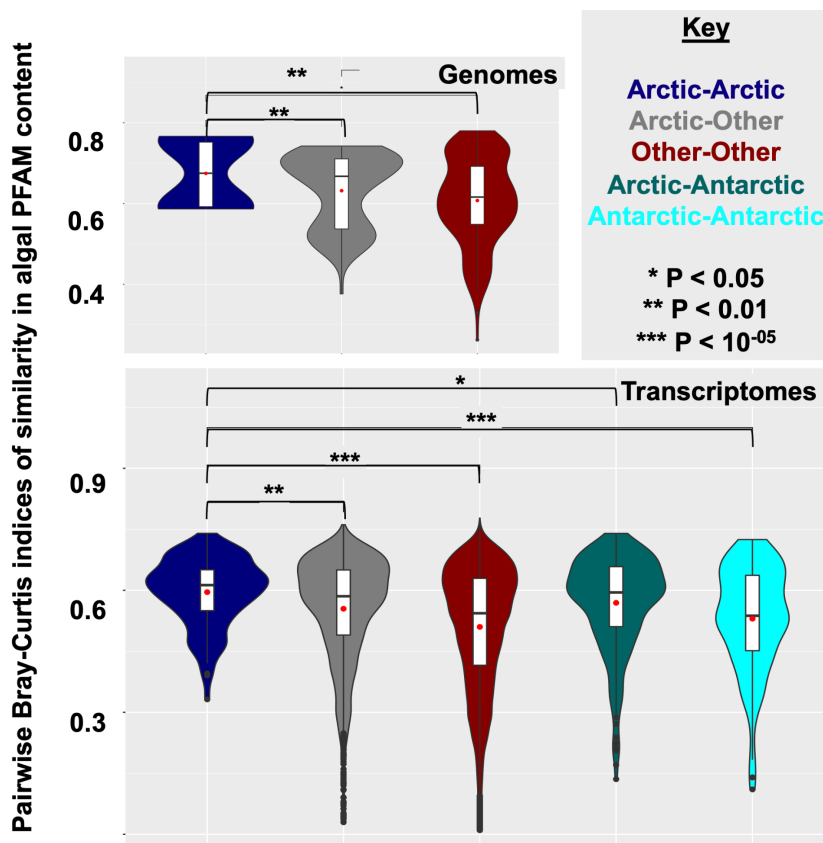


Fig. 4.7. Convergence in the PFAM content of distantly-related Arctic species.

This figure shows Violin Plots of anti Bray-Curtis indices (larger values indicating greater similarity) between the PFAM annotations of pairs of 21 algal genomes (top) and 296 MMETSP transcriptomes (bottom) from groups (shown in Fig. 3.6) with sequenced Arctic members. Genomes and transcriptomes were treated separately due to substantial differences in the PFAMs identified in each kind of library. Pairwise comparisons between species from the same taxonomic group are excluded to minimise confounding effects from

comparing species pairs with uneven phylogenetic distances. Freshwater and non-photosynthetic species were excluded to avoid introducing artifacts by comparing them to marine and photosynthetic species. Species are divided into those collected from the Arctic Ocean, the Antarctic Southern Ocean, and all other oceans grouped as "Other". No "Arctic-Antarctic" or "Antarctic-Antarctic" analyses are shown for genomes as only one Antarctic species (the diatom *Fragilariopsis cylindrus*) was sequenced at the time of analysis (Grigoriev et al., 2021; Joli et al., 2023).

This study broadly indicates convergent and horizontal evolution of the pan-Arctic pan-algal genome. From PFAM content, Arctic species are more similar to one another than they are to taxonomically equivalent non-Arctic relatives. We quantified this using a modified version of the Bray-Curtis similarity index, treating PFAMs in algal genomes and transcriptomes in the same way as species distributions between separate geographical sites (**Fig. 4.7**). This analysis was suggested by Federico Ibarbalz, performed by me, and plotted by Elisabeth Richardson. Surprisingly, we observed convergence between Arctic species pairs even compared to species from the Antarctic Southern Ocean, which may reflect Arctic-specific as opposed to polar-specific convergences (**Fig. 4.7**). We could also identify Arctic-specific convergence in PFAM content of Arctic species using phylogenetically-aware PCA, although with clearer signals between Arctic haptophytes, chrysophytes and pelagophytes than diatoms.

This convergence is likely to be predominantly due to differential expansions, contractions, and losses of gene families. Using phylogenetically-constrained CAFE analysis, we could identify specific PFAM domains that occurred at greater or lower frequency in Arctic species than their non-Arctic relatives (Freitas & Nery, 2020). At the same time, we could find evidence for horizontal gene transfer between some of our Arctic species. The single most enriched PFAM in Arctic genomes and transcriptomes (P-value of both presence and copy number < 10^{-30} , two-tailed χ^2 -test) was the PF11999/ DUF3493 domain associated with ice-binding proteins. Ice-binding proteins are a diverse family with multiple independently-evolved subgroups (Bar Dolev et al., 2016). They are unified by their ability to form complexes with water, many via threonine-rich domains ("TXT motifs"), which depress its freezing temperature ("thermal hysteresis") and can prevent the nucleation or expansion of ice crystals (Vance et al., 2019). PF11999/ DUF3494-type (or "type I") IBPs are unified by a conserved global β -solenoid structure (Bar Dolev et al., 2016; Vance et al., 2019). Beyond this, PF11999/ DUF3494 IBPs are variable in sequence, some containing recognisable TXT motifs and others not; localisation, with some known to be secreted and others probably cytoplasmic (Jung et al., 2016); and activity, with measured thermal hysteresis varying from 0 and 4°C in different purified proteins (Vance et al., 2019).

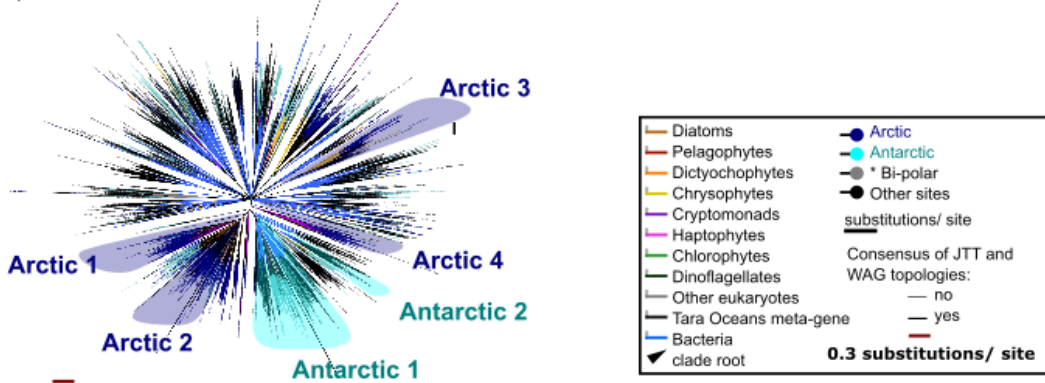
Concerning biological activity, intracellular ice-binding proteins most likely act as compatible solutes, lowering the crystallisation temperature of the cytosol and reducing the risk of cell lysis through osmotic upshock in freezing conditions (Bar Dolev et al., 2016). Secreted ice-binding proteins may have more variable roles, e.g. keeping solute channels open to allow

nutrient exchange across the plasma membrane, or permitting cell adhesion to floating sea-ice (Oude Vrielink et al., 2016). The Lovejoy group has shown that some Arctic algal ice-binding protein genes are transcriptionally regulated in response to hyposaline conditions corresponding to ice-melt, and hypersaline conditions as may occur in the brine pockets of floating sea-ice, suggesting physiological roles in salinity tolerance (Freyria et al., 2022).

Previous studies had suggested probable repeated horizontal transfers of PF11999 genes between distantly-related microbes collected from the Antarctic, including bacteria, fungi, diatoms, and green algae (Arai et al., 2019; Jung et al., 2016; Raymond & Kim, 2012; Stephens et al., 2020), and I wished to determine if similar horizontal gene transfers explained their particular enrichment in Arctic algae. Here, with the aid of Nikola Zarevski, Juan José Pierella Karlusich and Zoltán Füßy, I constructed a 4,862 branch tree of PF11999 domain diversity, incorporating cultured algal and *Tara* Oceans sequences. This tree was annotated based on taxonomy and biogeographical origin (collection site or most abundant *Tara* Oceans province).

The obtained topology revealed a complex story of Arctic- and Antarctic-specific gene families, each biased towards one polar biome only, and containing taxonomically distantly related members (**Fig. 4.8**). For example, “Arctic clade 1” incorporates cryptomonads (*Baffinella* sp. CCMP2293), haptophytes (Pavlovales sp. CCMP2436), pelagophytes (nov. gen. CCMP2097), diatoms (*Chaetoceros neogracile* RCC1993), and three dinoflagellates (*Heterocapsa arctica*, *Polarella glacialis* and the freshwater *Scrippsiella hangoei* *Peridinium aciculiferum* species complex) from distantly-related orders (Gymnodiniales, Suessiales and Peridinales respectively) (Annenkova et al., 2015; Dorrell, Klinger, et al., 2017). The remaining “Arctic clades” incorporate additional diatoms (*Amphora*, *Attheya*, *Entomoneis*), dictyochophytes and prasinophytes (*Pyramimonas*) (clade 2); chrysophytes (*Ochromonas*, *Pedospumella*) and dictyochophytes, (clade 3); and haptophytes and separate prasinophytes (*Micromonas*) to those found in clade 2. Finally, two “Antarctic clades” are distantly positioned to the “Arctic clades”, incorporating diatoms, haptophytes and bacteria (clade 1) and cryptomonads and prasinophytes (clade 2). In many cases, biogeographical origin is a better predictor of the IBPs possessed by individual species, with the closely related prasinophytes *Micromonas polaris* and *Mantoniella antarctica* respectively placed in Arctic clade 4 and Antarctic clade 2; and the Antarctic *Chaetoceros dictyota* and *C. neogracile* sp. Cn-AFP resolving in Antarctic clade 1, despite the closely related *C. neogracilis* sp. RCC1993 falling in Arctic clade 1 (**Fig. 4.8B**) (Gwak et al., 2010; McKie-Krisberg et al., 2015; Simon et al., 2017).

A 4,862-branch *Tara* Oceans-enriched tree



B specific Arctic- and Antarctic-associated clades

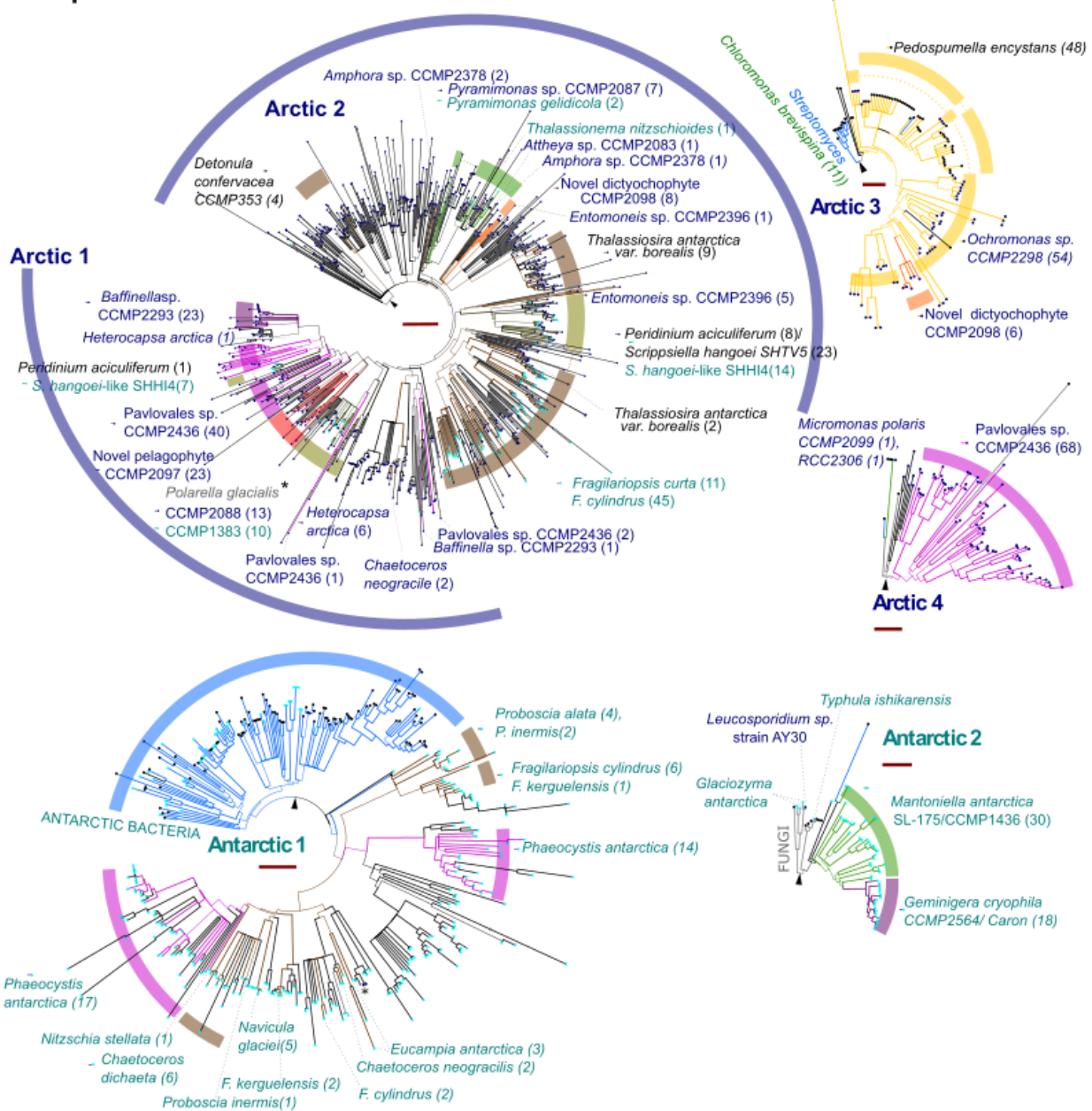


Fig. 4.8. Global taxonomic and biogeographical diversity of the PF11999 domain. Consensus RAXML best-scoring topology (realised with JTT and WAG substitution matrices) of a 4,862 taxa x 193 aa alignment of PF11999-containing / type I ice-binding protein sequences identified from Uniref, JGI genomes, MMETSP

transcriptomes and *Tara* Oceans meta-transcriptome data (Carradec et al., 2018; Grigoriev et al., 2021; Keeling et al., 2014). Branches are coloured by taxonomy using a most ancient common ancestor approach, and leaves by biogeography: “Arctic” sequences are defined as being collected at > 60°N (cultured species) or with > 70% total meta-T relative abundance across all depths and size fractions in Arctic Ocean stations (*Tara* Oceans meta-genes); “Antarctic” sequences those from > 60°S or with > 70% meta-T relative abundance from Southern Ocean stations; and “Bipolar” sequences (*Tara* Oceans meta-genes only) showing at least 70% of total meta-T relative abundance in Arctic and Southern Ocean stations, with a minimum of 30% in each. Panel **A** shows the global topology, with four clades enriched in Arctic algae and two enriched in Antarctic algae labelled with callouts. Panel **B** shows these clades in detail. The numbers of ice-binding proteins associated with each strain are shown with circular brackets.

The most parsimonious explanation for this phylogenetic distribution is that the Arctic clade IBPs represent “Arctic-specific” isoforms of this protein, spread between distantly-related species by horizontal gene transfer, presumably within Arctic water masses. Furthering this story, Nikola was able to identify during his Masters internship a further thirty gene families with signatures of within-Arctic horizontal gene transfer, although these were not included in the final paper (Dorrell et al., 2023).

Beyond HGT, the IBP topology also demonstrates how genome coding content can reveal biogeographical and in particular migratory histories. Within Arctic clade 1, we identified a dinoflagellate (*Polarella glacialis*) which occurs both in the Arctic (strain CCMP2088) and Antarctic (CCMP1383), but whose ancestral habitat was uncertain (Ibarbalz et al., 2023; Stephens et al., 2020) (**Fig. 4.8B**). The placement of this species in a predominantly Arctic IBP clade implies it acquired these genes in the Arctic, and migrated south to the Antarctic, rather than being ancestrally present in the Southern Ocean and migrating north. In a second example, Arctic clade 3 includes both an Arctic marine chrysophyte (*Ochromonas* sp. CCMP2298) and dictyochophyte (CCMP2098), and also terrestrial/ freshwater algae, e.g. the distantly related chrysophyte *Pedospumella encystans* JBM/S11 collected in Austria, and the chlorophyte *Chloromonas brevispina* UTEX:SN096, collected in New England (**Fig 4.8B**) (Dorrell, Azuma, et al., 2019; Grossmann et al., 2016; Raymond, 2014). One possible explanation would be an ancient migration of freshwater algae, possessing ice-binding proteins, from continental masses into the Arctic Ocean, where they may have then donated their IBP genes to marine species (Jamy et al., 2022). I find it particularly interesting that the Arctic is somewhat less saline than other oceans, and IBPs are mediators of hyposalinity tolerance in some of its species (Freyria et al., 2021). I wonder whether this ecosystem has formed a hotspot for interactions between freshwater, euryhaline, and ice-adapted marine organisms.

4.5. Taxonomic and biogeographical trends in Vitamin B₁₂ metabolism loss across the photosynthetic tree of life

In a final project, moving from algae towards plants, I have been interested to consider how phylogeny and environment shape patterns of gene loss in photosynthetic eukaryotes. This project relates to Vitamin B₁₂, an essential cofactor for several enzymes (B₁₂-dependent methionine synthase or METH, methylmalonyl-coA mutase in the mitochondrial propionate shunt, and form II ribonucleotide reductase) (Froese et al., 2019). Eukaryotes including algae cannot synthesise B₁₂, acquiring it from bacteria by symbiosis or phagotrophy, and its scarcity may constrain marine primary production, e.g. in the Southern and South Atlantic oceans (Bertrand et al., 2007; Browning et al., 2017). A subset of eukaryotes- including plants- have eliminated their dependence on B₁₂, using instead a B₁₂-independent isoform of methionine synthase, METE. The absence of B₁₂ from plants may make it a key limiting nutrient in the human diet, with severe health consequences (e.g. pernicious anaemia) (Chan et al., 2016). An estimated 70-90% of peoples in low-income and subsistence economies suffer from chronic B₁₂ deficiency (Gebremedhin, 2021; Stabler & Allen, 2004).

The dependence or not of a species on B₁₂ can be effectively traced by the presence of METH (use of B₁₂) and METE (ability to survive without B₁₂) in its genome. These genes have partially overlapping distributions, with organisms that uniquely possess METH (e.g., animals) dependent on B₁₂, those that only possess METE (e.g., plants) not using it, and those possessing both able to use it facultatively. Previous studies of algae have suggested that most are either obligately dependent on B₁₂ (i.e., possess METH only) or facultative users (possessing METH and METE), with a small number, like plants, only having METE (Brawley et al., 2017; Croft et al., 2005; Helliwell et al., 2011). In this project I was interested to evaluate the occurrence of B₁₂-associated metabolism across algae, considering distributions of METH, METE, and other proteins dependent on B₁₂ (MCM, RNR-^{II}) or involved in B₁₂ uptake and conversion (MTRR; CBLA, B, C, D, F, J; CBA1). I was interested to determine at what point B₁₂ utilisation had been lost in the plants, considering gene distributions in 1, 667 plant and algal genomes and transcriptomes, e.g. from the One Thousand Plant Transcriptomes (OneKp) project (Carpenter et al., 2019; One Thousand Plant Transcriptomes Initiative, 2019). I benefited from advice, particularly on B₁₂ metabolism, from Alison Smith at Cambridge University, who has worked extensively on this question in eukaryotic algae (Croft et al., 2005; Dorrell & Smith, 2011; Helliwell et al., 2011). I also acknowledge Setsen Altan-Ochir, an

undergraduate student who assisted in some of the phylogenies within the study (Dorrell, Nef, et al., 2024).

The principal conclusion from this study is that B₁₂-dependent methionine synthase (METH), alongside the proteins required for its uptake into the cytoplasm (CBLB, C, D, J, CBA1) and conversion into the methylcobalamin form required for methionine synthesis (MTRR) do occur in some plants, specifically in the hornworts (Anthoceroophyta) due to probable vertical inheritance of an ancestral streptophyte isoform (**Fig. 4.9**). We further detected CBLB and CBLJ, required for the synthesis of adenosylcobalamin, in liverworts (Marchantiophyta) (Dorrell, Nef, et al., 2024), although in the absence of known B₁₂-dependent proteins in this lineage it is difficult to establish whether these are functional. Dependent on the taxonomic position of hornworts (i.e., as a sister-group to the Bryophytes, or to all plants) this suggests multiple losses of B₁₂-utilisation in vascular plants and in mosses, and more broadly blurs a little further the functional distinctions between alga and plant (Buschmann & Holzinger, 2020; One Thousand Plant Transcriptomes Initiative, 2019; Wickett et al., 2014).

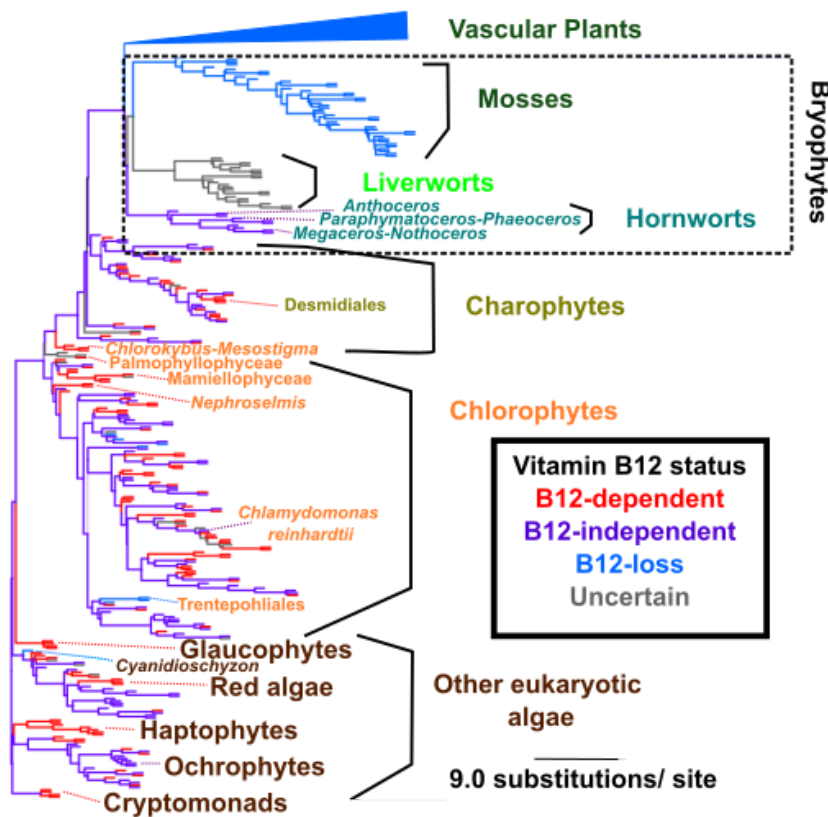


Fig. 4.9. Progressive loss of B₁₂-dependence in land plants.

This figure reproduces the ASTRAL multigene topology of the OneKp project (One Thousand Plant Transcriptomes Initiative, 2019). Taxa are labelled by phylogenetic origin, as are specific leaf nodes with either complete B₁₂-dependence or loss of B₁₂-metabolism. Branches are coloured based on B₁₂ status: either dependence (METH only); B₁₂-independence (METH and METE); or B₁₂-loss (METE, and no METH/ MTRR). Reflecting their loss of METH but retention of CBLB/J, liverworts are labelled as “B₁₂-uncertain”.

From a physiological perspective, perhaps the more interesting story is in what algal species B₁₂-metabolism has been lost. For fifteen strains beyond plants, we detected METE, but not

METH and its associated protein MTRR, suggesting that these species do not use B₁₂ for methionine synthesis (**Table 4.1**). Most of these species come from collection sites in which B₁₂ may not be readily bioavailable (**Table 4.1**). These include the Antarctic (e.g., the cryptomonad *Geminigera*, which is the only member of its order to possess METE, and to lack METH), reflecting evidence for B₁₂- limitation in this ocean region (Bertrand et al., 2007); in acidic hot spring environments (e.g. *Cyanidioschyzon*, *Galdieria*) (Brawley et al., 2017; Cho et al., 2023), with independent losses of B₁₂-dependent metabolism identified in extremophilic hot spring bacteria (Daebeler et al., 2018); and in freshwater ecosystems, which may suggest a parallel loss of B₁₂- use during both algal and plant terrestrialisation.

Species	Taxonomy		Collection site			Notes
	Lineage	Strain	Context	Latitude	Longitude	
Land Plants						
Mosses, Vascular plants, liverworts?						
<i>Blastophysa cf. rhizopus</i>	Chlorophytes	M3368	Marine epiphyte	28.4	14.4	Based on PFAM only
<i>Botryococcus braunii</i>	Chlorophytes	UTEX 2441	Freshwater	-13.5	-72.1	Based on PFAM only
<i>Brandtodinium nutriculum</i>	Dinoflagellates	RCC3387	Foraminiferan symbiont	43.7	7.3	
<i>Cephaleuros virescens</i>	Chlorophytes	SAG 28.93	Leaf parasite	-29.9	31	Based on PFAM only; Trentepohliales
<i>Coccomyxa pringsheimii</i>	Chlorophytes	SAG 216.7	Lichen symbiont	N/A	N/A	
<i>Cyanidioschyzon merolae</i>	Red algae	10D	Hot spring	40.9	14.3	
<i>Galdieria sulphuraria</i>	Red algae	074W	Fumarole	-7.6	11.2	Encodes MCM, CblA
<i>Geminigera cryophila</i>	Cryptomonads	CCMP2564	Antarctic	-77.8	-163	
<i>Ignatius tetrasporus</i>	Chlorophytes	UTEX2012	Freshwater	N/A	N/A	May encode RNR II
<i>Leptosira obovate</i>	Chlorophytes	SAG 445.1	Freshwater	47.5	7.9	
<i>Porphyridium cruentum</i>	Red algae	UTEX 161	Freshwater	47.6	7.6	Encodes CblB
<i>Poteriospumella JBC07</i>	Chrysophytes	JBC07	Freshwater	31.5	120.2	
<i>Pseudosourfieldia marina</i>	Chlorophytes	SCCAP K/0017	Estuarine	59.4	10.6	Encodes CblB
<i>Trentepohlia annulata</i>	Chlorophytes	SAG 20.94	Freshwater	49.7	16	Based on PFAM only; Trentepohliales
<i>Vitrella brassicaformis</i>	Alveolates	CCMP3155	Coral symbiont	-23.5	152	

Table 4.1. Phylogenetic and sampling contexts of species inferred to only utilise B₁₂-independent methionine synthesis pathways.

Considering the biotic context of these organisms, many are symbionts of other species, either plants or lichens (Trentepohliales) (Kosecka et al., 2020; Pitaloka et al., 2015), corals (*Vitrella brassicaformis*) (Oborník et al., 2012; Probert et al., 2014), or Foraminifera (*Brandtodinium nutriculum*) (Dittami et al., 2021; Probert et al., 2014). I wonder if this may relate to limited B₁₂ availability inside holobionts (Dittami et al., 2021). Finally, based on the METE tree generated at least three of these groups (plants, the red alga *Cyanidioschyzon merolae*, and the chrysophyte *Poteriospumella lacustris*) have independently reacquired B₁₂-independent methionine metabolism by horizontal transfers from bacteria (**Fig. 4.10**). It is possible that the horizontal acquisition of METE influences the loss of METH across the algal tree of life.

4.6. Concluding remarks and pedagogical insights

These projects concerning phenomena of genome evolution have been an education, both in algal biology and in the scientific process. While my initial interests were tilted towards the origins of algal chloroplasts (**Figs. 4.1, 4.2**), I am increasingly interested in the continuity of HGT and its importance for defining leaf nodes in the tree of life (**Figs. 4.4, 4.8**). I am equally interested in how HGT relates to phylum-level differences in algal biology (**Fig. 4.3, 4.9**), and in species-specific adaptations such as colonisation of polar ecosystems, or loss of B₁₂-dependent metabolism (**Fig. 4.7, Table 4.1**). I wish to further explore how HGT relates to interactions between algae and their environment, e.g. considering the secretory bias observed in bacterially derived genes in diatom genomes, and the probably impacts of IBP acquisition on the thermal niches of polar algae (**Figs 4.5, 4.6**). I hope that my future work will be able to keep a balance between macro-scale and micro-scale views of algal evolution.

Concerning future projects, there are several loose ends within these data that could benefit from more appropriately guided research. Our current views of the haptophyte chloroplast proteome (**Figs 4.2, 4.3**) is largely based on targeting predictors developed in diatoms (Gruber et al., 2015) and phylogenomic questions built around *Phaeodactylum* gene queries (Dorrell et al., 2021). An experimental chloroplast proteome for haptophytes, either via classical plastid isolation or next-generation proteomic strategies such as LOPIT (Huang et al., 2023; Mulvey et al., 2017) would be helpful to exploring its content, and phylogenetic origin. Understanding the precise endosymbiotic connections linking haptophytes to both pelagophytes and dictyochophytes within the ochrophytes, and to cryptomonads, could be best explored by targeted analyses of genome sequences from the respective sister-groups for each lineage. These would include centrohelids and/ or telonemids for haptophytes (Sørensen et al., 2023; Yazaki et al., 2022); *Goniomonas* for cryptomonads (Cenci et al., 2018; Schön et al., 2021); and potentially pinguiophytes for pelagophytes/ dictyochophytes (Di Franco et al., 2022) (**Fig. 4.1**).

The particular enrichment of predicted secretory localisations in bacterially-derived genes *Phaeodactylum* and other protists is interesting and I think merits deeper functional consideration (**Fig. 4.5**). Perhaps the proteins encoded by these horizontally acquired genes are cytotoxic, and secretion is an effective way to expel them from the cell, which might explain the high frequencies of HGT loss projected in our data (**Fig. 4.4**). That said, there are clear examples of secreted and bacterially derived proteins that confer adaptive advantages (e.g., ice-binding proteins; **Figs. 4.5, 4.8**), and I doubt that all of them are deleterious.

I wonder if some of these bacterially-derived proteins have specific functions in mediating diatom interactions with the environment, e.g. interacting with bacterial epibionts. This is roughly analogous to the idea that chloroplast targeting sequences originated from antimicrobial peptides to control the biology of intracellular parasites (Caspari & Lafontaine, 2021; Garrido et al., 2020). For this, we need to validate carefully the secretory localisations of bacterially derived diatom proteins, e.g., by experimental proteomics of cell culture supernatants and adhesive trails (Lachnit et al., 2019). We might also consider the functions of some of these proteins by reverse genetics, or even purely comparative gene expression approaches (Kazamia et al., 2018). The bacterially-derived genes in *Phaeodactylum* are typically rather lowly expressed, which might suggest dispensable functions (Dorrell et al., 2021). That said, most *Phaeodactylum* transcriptome data published to my knowledge are from axenic cultures, where the functions of genes e.g. involved in the regulation of bacterial epibionts may be obscured. Here, more sequence data from non-axenic cultures, including different *Phaeodactylum* ecotypes with their microbiomes may be useful to decode possible biotic functions for bacterial HGT (Tirichine & Piganeau, 2023).

I think we are only beginning to scratch the surface of Arctic-specific genome innovations. A key limitation of our initial study is its almost complete reliance on PFAM as a metric for protein orthology (**Fig. 4.7**), when in reality individual PFAMs may group together protein families with different functions (e.g., in the case of MFS transporters, see Chapter 3), and are likely to exclude protein families not found in cultivable laboratory models. Here, moving beyond PFAMs to consider gene families encoding proteins with structural but no primary sequence homology (Oteri et al., 2022; Zheng et al., 2021) will be important to identifying the full extent of the Arctic-associated genome.

We also do not know where within-Arctic HGT occurs. For example, Arctic-adaptive genes may be exchanged within the Arctic Ocean itself, or in other ocean regions by species that subsequently immigrate into the Arctic. Within the Arctic itself, we might expect HGT to be more likely in certain environments than others. Sea ice, which may act as an accumulator of cells and barrier to dispersal during seasonal freeze-up events, has been proposed to facilitate the horizontal transfer of Arctic-adaptive genes between distantly related species (Ardyna et al., 2014; Raymond & Kim, 2012). This might be considered as roughly analogous to the way that winter ice bridges have, historically, enabled the exchange of Arctic-adaptive technologies between the geographically distant peoples indigenous to this region (Egeesiak et al., 2017).

Understanding in what specific habitats within-Arctic HGT occurs might be best explored using ancestral niche reconstruction of species distributions from metagenomic data (Lambert et al., 2023), or by building a mathematical model for within-Arctic HGT, analogous to existing models for the horizontal spread of resistance plasmids across bacteria (Djidjou-Demasse et al., 2023).

We are also most likely missing some of the taxonomic subtlety of Arctic algal evolution by grouping all species together, instead of separating diatoms from flagellates, under strong temporal segregation respectively in Arctic spring and summer conditions (Leu et al., 2015), or by grouping Arctic endemic with cosmopolitan and invasive species that may show different thermal niches (Fig. 4.6) (Ibarbalz et al., 2023). I have recently begun to explore the genetic and physiological differences between different groups of Arctic algae, which has formed the projects of several undergraduate project students (Victoria Powell, Pierre der-Gazerian, Ema Siomeo, Ilona Dupuis, and Alice Bossard) and are now being developed by Louis Graf, an incoming postdoc in my group with expertise both in algal taxonomy and functional genomics (Andersen et al., 2017; Graf et al., 2021). I provide more detail of this in my Research Project.

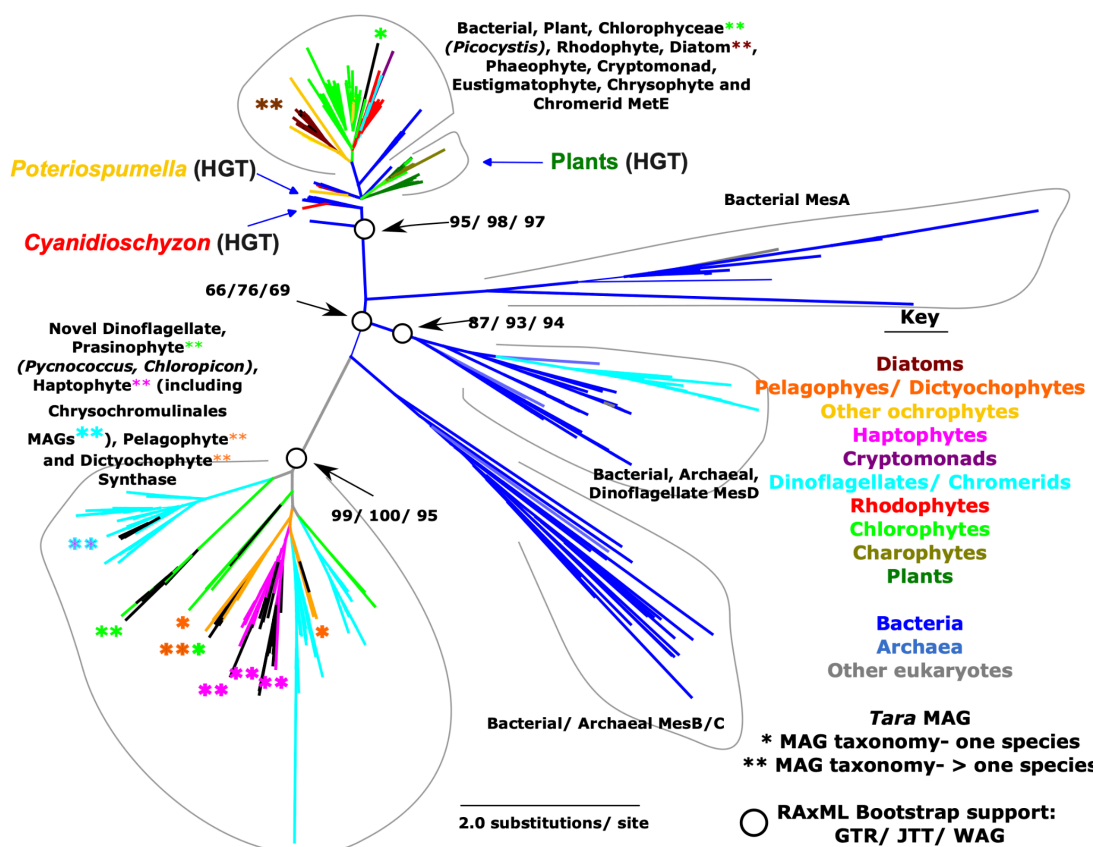


Fig. 4.10. A global tree of METE superfamily diversity involving Tara Oceans data. This figure shows a consensus RAXML tree of with three matrices (GTR, JTT, and WAG) of a 503 taxa x 589 aa alignment of the B₁₂-independent methionine synthase METE and its related bacterial proteins (MesA, MesB, MesC and MesD from

both cultured plant and algal genomes and transcriptomes, and *Tara* Oceans MAGs following (Dorrell, Nef, et al., 2024; Price et al., 2021). This tree reveals the presence of a novel and previously METE-related protein found both in cultured and uncultured haptophytes, pelagophytes, dinoflagellates and prasinophytes. It further suggests horizontal acquisitions from bacteria of METE in plants, *Cyanidioschyzon*, and *Potteriospumella*, who have all independently lost the B₁₂-dependent methionine synthase METH (**Table 4.1**).

Finally, across all of these different projects, I have predominantly worked with cultured genomes and transcriptomes (**Fig. 4.8**). A key concern for me is how to use meta-genomics appropriately when, in a cell-free context, we have even less ability to avoid impacts of contamination in our data. For this I am starting to work with *Tara* Oceans meta-genome associated genomes, for example with Charlotte Nef from the Bowler group building new phylogenies of Vitamin B₁₂-associated proteins (Dorrell, Nef, et al., 2024; Pierella Karlusich et al., 2022). Here, we have found the conserved presence of novel protein families across closely-related metagenome-assembled genomes, suggesting that they can form a complementary resource to cultured genomes for exploring protein family evolution (**Fig. 4.10**). I am particularly interested to explore phylogenetic signals in MAG families without cultured relatives (Delmont et al., 2022), and also if we can observe similar relationships between gene content and biogeographical in MAGs and cultured species collection sites.

On pedagogical grounds, these projects have been an instructive lesson in working within consortia with multiple PIs. The diatom HGT project developed organically around several different strands (manual and automated phylogenetics, mathematical reconciliation, protein localisation) and I am grateful to Guillaume Blanc and H el ene Morlon for letting me work with their students to build a more comprehensive, synthetic story than my data alone provided. The Arctic project was more complex, involving a mature consortium that I joined several years after the algae were collected, put into culture and sequenced, and involved multiple paper resubmissions and rewrites. Here I hope I can be more mindful in the future of the dynamic between collaborators, and also my role in projects relative to work I have not seen that preceded my arrival. I am grateful to see the vitamin B₁₂ project now accepted for publication, and the new collaborations it may allow me to create. Moving forward, I hope I will be able to maintain a more balanced view in my sense of obligation to collaborators, to my students, and to the science.

Chapter 5: Perspectives and future directions

I have had the privilege of being able to work on the same research theme, i.e. the horizontal evolution of eukaryotic algae and their chloroplasts (Chapter 1), since my undergraduate studies (Dorrell & Smith, 2011). Since joining the Bowler group as a postdoc in 2014, I have developed three complementary objectives: the computational and phylogenetic reconstruction of algal chloroplast proteomes (Dorrell, Azuma, et al., 2019; Novák Vanclová et al., 2024); the environmental and experimental characterisation of novel algal chloroplast proteins (Dorrell, Gile, et al., 2017; Dorrell, Zhang, et al., 2024); and most recently the evaluation of broader trends in inter-algal HGT over temporal and spatial scales (Dorrell et al., 2021, 2023). I am grateful to have been joined in this journey by brilliant students, who have helped take my research further than I was capable (S. Liu et al., 2022; Nonoyama et al., 2019; Penot et al., 2022). I would like to close this *mémoire* by reflecting on how this research has evolved from my initial vision, and its likely orientation in the future.

My interests have shifted somewhat from the large- to small-scale. I had initially planned, following initial reconstruction of the ochrophyte plastid proteome (Dorrell, Gile, et al., 2017; Dorrell & Bowler, 2017), to scale this work up to considering the pan-secondary red chloroplast proteome, incorporating cryptomonads, haptophytes, dinoflagellates and apicomplexans. I think this could still be very interesting, and I would be interested to more carefully explore the deep evolution of the haptophyte chloroplast, given its evident phylogenetic connections to several other algal groups (Dorrell et al., 2021; Dorrell, Gile, et al., 2017). At the same time, Anna's work on the chloroplast proteomes of just one algal group, the Kareniaceae, has surprised me by how much chloroplast proteome content and phylogenetic history may vary between species, which I would have missed with a more crude, "top down" approach (Novák Vanclová et al., 2024). Concerning horizontal gene transfer, I see much more interesting events happening at the level of individual species, where we may be able to pose questions about how biogeography influences gene acquisition and loss (Dorrell et al., 2023; Dorrell, Nef, et al., 2024) than in stem nodes where we have less visibility over the ancestral functions, and long-term retention of horizontally acquired genes (Dorrell et al., 2021; Rastogi et al., 2018). One challenge moving forward will be to strike a balance between macroevolutionary trends and deep signals, which may have had more profound roles in shaping algal origins but are less clearly definable; and microevolutionary examples where we

can use individual groups to explore principles of what can happen in evolution, e.g. in the case of the convergent losses of photosynthesis in chrysophytes (Dorrell, Azuma, et al., 2019)

A second point of evolution has been my understanding of how to equilibrate environmental and experimental data. I had initially planned to pinpoint interesting diatom chloroplast proteins from phylogenetic distributions, generate CRISPR knockout lines in *Phaeodactylum*, and then use *Tara* data to explore the significance of these functions in the wild (Kazamia et al., 2018; Nonoyama et al., 2019). I think we can actually go further than this, for example using meta-genomic data to predict appropriate conditions for experimental phenotyping (Dorrell, Zhang, et al., 2024). I think that *Tara* data may also be useful for understanding protein function using criteria that are simply not visible in the lab. For example, I am particularly intrigued by Shun's observation that the diatom chloroplast transporters specific to this lineage are strongly correlated to indices of primary production in *Tara* data (S. Liu et al., 2022).

Nonetheless, I think there are limitations to our use of metagenomic data to explore algal physiology. I am agnostic about whether the overexpression of diatom chloroplast glycolysis enzymes in high latitude *Tara* stations is truly an adaptive feature relating to daylength and temperature, or whether it is purely as diatoms are particularly metabolically active in these stations, and cpEnol and cPGM play central roles in their primary chloroplast metabolism (Ait-Mohamed et al., 2020; Pierella Karlusich et al., 2022). A secondary concern is whether particular proteins are used in physiologically consistent ways across different algae, and whether we can use *Tara* data for all diatoms to understand one species (*Phaeodactylum*). Moving forward, I am more interested in the phylogenetic applications of meta-genomics, as per the ice-binding proteins, where reconciling *Tara* and genome data reveals a biogeographical structure (Dorrell et al., 2023). Nonetheless, I think meta-genomics can be best used as a complement to experiments in cultured species for exploring protein function.

Finally, I think that my research still does not take enough account of interactions between proteins, or indeed between species. Within dinoflagellates, I would be very interested to see if the most fast-evolving regions of chloroplast genes interact with unusual nucleus-encoded proteins that might drive their divergent evolution (Dorrell, Klinger, et al., 2017; Klinger et al., 2018). At the level of chloroplast proteome evolution, we need to know more about the phylogenetic histories of both the host and symbiont genomes, e.g. whether green genes in diatoms come from the host, the endosymbiont or both (Dorrell, Gile, et al., 2017), to fully

interpret the different mosaic signals we observe. Finally, for the horizontally acquired bacterial genes in diatoms, I would be interested to explore whether any of these proteins are derived from episymbionts, or preferentially expressed in xenic conditions that may suggest roles in biotic interactions (Dorrell et al., 2021; Dorrell, Nef, et al., 2024). Here I think that meta-genomic data, and particularly inference of associations between species distributions, may take us somewhat further than experimental approaches (Novák Vanclová et al., 2024).

In the short term, my priority is to conclude projects underway in my group, particularly those of my former and current PhD students. For Shun's project on Trans3, we have to confirm its function as a molybdate transporter, find some biological explanation for why molybdate would be transported into the diatom PPC, and in the case that it does not function as such attempt to identify an alternative substrate and transport role (Gasber et al., 2011; Tejada-Jiménez et al., 2011). For Mathias' project on xATPA, I feel that we have explored as far as possible its functions by reverse genetics in diatoms, and we plan instead to focus on expressing the protein and characterising its possible functions (e.g., ATP-binding) *in vitro* (LaConte et al., 2017); and on localising its homologues in non-photosynthetic groups such as labyrinthulomycetes, to assess where it functions beyond the chloroplast (Rius et al., 2023). In both projects it is apparent that we should be working with transformable systems beyond *Phaeodactylum* to close these questions.

In terms of supervision, an ongoing challenge is to balance being hands-on and hands-off with my group. Both Anna and Shun developed their work beyond my vision when I did not interfere, considering species-specific variations in Kareniacean chloroplast proteomes which in turn revealed multiple endosymbiotic events, and Trans3 as a homologue of MOT2 involved in molybdate transport. For the current members of my group, I need to maintain enough structure and regularity in our interactions that they are guided scientifically and so that technical problems can be resolved. I take some degree of inspiration from the open and collaborative dynamic in the Bowler group during my postdoc, where I was able to reciprocally enrich the work of my colleagues (Bey et al., 2022; Kazamia et al., 2018; Pierella Karlusich et al., 2023; Shao et al., 2023), and the dinoflagellate "evening lab" that I established largely with other early career researchers (Dorrell, Klinger, et al., 2017; Klinger et al., 2018). I am pleased to see Louis, Mathias and Victoria from my group work together, support each other's work and create their own research ecosystem under and beyond my supervision.

Nonetheless, I would like to use the next five years to reorient the work of my group towards more urgent questions of environmental change. The years of my postdoc have been accompanied by visible and accelerating evidence of global heating as a result of anthropogenic emissions (IPCC, 2007)), and we can anticipate this to escalate in the next decade (Lenton, 2013). The concept of performing fundamental research in the era of climate collapse, particularly at the expense of fully recognising and taking action to address these threats, is a source of anxiety both for myself and for some of the junior members of my group (Barnacle & Cuthbert, 2021).

I am therefore, on reflection, proposing a slightly different future research project than that previously performed in my group, using cultured species and environmental data to build a genome-scale model of climate change fragility and/ or resilience for eukaryotic algae. I believe that this project is achievable, as it uses techniques benchmarked in my previous research, e.g. genome sequencing and annotation (Dorrell, Azuma, et al., 2019; Dorrell et al., 2021); automated and manual phylogenetic resolution (Dorrell, Gile, et al., 2017; Novák Vanclová et al., 2024); metabolic modelling (Ait-Mohamed et al., 2020; Dorrell, Zhang, et al., 2024); and meta-genomics (S. Liu et al., 2022). I am particularly keen to expand on Arctic algae and haptophytes as model systems (Dorrell et al., 2023; Penot et al., 2022), but hope eventually to build predictive tools that can be generalised across the diversity of eukaryotic algae. I outline both these short- and longer-term objectives within my Research Project.

References

(**bold** indicates post-PhD participation from my group, underlined = supervised group member)

- Aihara, Y., Takahashi, S., & Minagawa, J. (2016). Heat induction of cyclic electron flow around Photosystem I in the symbiotic dinoflagellate *Symbiodinium*. *Plant Physiol*, *171*(1), 522–529.
- Ait-Mohamed, O., Novák Vanclová, A. M. G., Joli, N., Liang, Y., Zhao, X., Genovesio, A., Tirichine, L., Bowler, C., & Dorrell, R. G. (2020). PhaeoNet: a holistic RNAseq-based portrait of transcriptional coordination in the model diatom *Phaeodactylum*. *Front Plant Sci*, *11*, 590949.**
- Allen, A. E., Dupont, C. L., Oborník, M., Horák, A., Nunes-Nesi, A., McCrow, J. P., Zheng, H., Johnson, D. A., Hu, H., Fernie, A. R., & Bowler, C. (2011). Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature*, *473*(7346), 203–207.
- Allen, A. E., Laroche, J., Maheswari, U., Lommer, M., Schauer, N., Lopez, P. J., Finazzi, G., Fernie, A. R., & Bowler, C. (2008). Whole-cell response of the pennate diatom *Phaeodactylum tricorutum* to iron starvation. *Proc Natl Acad Sci USA*, *105*(30), 10438–10443.
- Andersen, R. A. (2004). Biology and systematics of heterokont and haptophyte algae. *Am J Bot*, *91*(10), 1508–1522.
- Andersen, R. A., Graf, L., Malakhov, Y., & Yoon, H. S. (2017). Rediscovery of the *Ochromonas* type species *Ochromonas triangulata* (Chrysophyceae) from its type locality (Lake Veysove, Donetsk region, Ukraine). *Phycologia*, *56*(6), 591–604.
- Anderson, D. M., Fachon, E., Pickart, R. S., Lin, P., Fischer, A. D., Richlen, M. L., Uva, V., Brosnahan, M. L., McRaven, L., Bahr, F., Lefebvre, K., Grebmeier, J. M., Danielson, S. L., Lyu, Y., & Fukai, Y. (2021). Evidence for massive and recurrent toxic blooms of *Alexandrium catenella* in the Alaskan Arctic. *Proc Natl Acad Sci USA*, *118*(41). doi: 10.1073/pnas.2107387118
- Andriotis, V. M. E., Kruger, N. J., Pike, M. J., & Smith, A. M. (2010). Plastidial glycolysis in developing *Arabidopsis* embryos. *New Phytol*, *185*(3), 649–662.
- Annenkova, N. V., Hansen, G., Moestrup, Ø., & Rengefors, K. (2015). Recent radiation in a marine and freshwater dinoflagellate species flock. *ISME J*, *9*(8), 1821–1834.
- Arai, T., Fukami, D., Hoshino, T., Kondo, H., & Tsuda, S. (2019). Ice-binding proteins from the fungus *Antarctomyces psychrotrophicus* possibly originate from two different bacteria through horizontal gene transfer. *FEBS J*, *286*(5), 946–962.
- Ardyna, M., & Arrigo, K. R. (2020). Phytoplankton dynamics in a changing Arctic Ocean. *Nat Climate Change*, *10*(10), 892–903.
- Ardyna, M., Babin, M., Gosselin, M., Devred, E., Rainville, L., & Tremblay, J.-É. (2014). Recent Arctic Ocean sea ice loss triggers novel fall phytoplankton blooms. *Geophys Res Lett*, *41*(17), 6207–6212.
- Asch, R. G., Stock, C. A., & Sarmiento, J. L. (2019). Climate change impacts on mismatches between phytoplankton blooms and fish spawning phenology. *Global Change Biol*, *25*(8), 2544–2559.
- Ashworth, J., Turkarslan, S., Harris, M., Orellana, M. V., & Baliga, N. S. (2016). Pan-transcriptomic analysis identifies coordinated and orthologous functional modules in the diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricorutum*. *Mar Genom*, *26*, 21–28.
- Azuma, T., Pánek, T., Tice, A. K., Kayama, M., Kobayashi, M., Miyashita, H., Suzuki, T., Yabuki, A., Brown, M. W., & Kamikawa, R. (2022). An enigmatic stramenopile sheds light on early

evolution in Ochrophyta plastid organellenogenesis. *Mol Biol Evol*, 39(4). doi: 10.1093/molbev/msac065

- Bailey, S., Melis, A., Mackey, K. R. M., Cardol, P., Finazzi, G., van Dijken, G., Berg, G. M., Arrigo, K., Shrager, J., & Grossman, A. (2008). Alternative photosynthetic electron flow to oxygen in marine *Synechococcus*. *Biochim Biophys Acta*, 1777(3), 269–276.
- Bai, Y., Cao, T., Dautermann, O., Buschbeck, P., Cantrell, M. B., Chen, Y., Lein, C. D., Shi, X., Ware, M. A., Yang, F., Zhang, H., Zhang, L., Peers, G., Li, X., & Lohr, M. (2022). Green diatom mutants reveal an intricate biosynthetic pathway of fucoxanthin. *Proc Natl Acad Sci USA*, 119(38), e2203708119.
- Ban, H., Sato, S., Yoshikawa, S., Yamada, K., Nakamura, Y., Ichinomiya, M., Sato, N., Blanc-Mathieu, R., Endo, H., Kuwata, A., & Ogata, H. (2023). Genome analysis of Parmales, the sister group of diatoms, reveals the evolutionary specialization of diatoms from phago-mixotrophs to photoautotrophs. *Commun Biol*, 6(1), 697.
- Bannerman, B. P., Kramer, S., Dorrell, R. G., & Carrington, M. (2018). Multispecies reconstructions uncover widespread conservation, and lineage-specific elaborations in eukaryotic mRNA metabolism. *PLoS One*, 13(3), e0192633.**
- Barbrook, A. C., Dorrell, R. G., Burrows, J., Plenderleith, L. J., Nisbet, R. E. R., & Howe, C. J. (2012). Polyuridylylation and processing of transcripts from multiple gene minicircles in chloroplasts of the dinoflagellate *Amphidinium carterae*. *Plant Mol Biol*, 79(4-5), 347–357.
- Barbrook, A. C., & Howe, C. J. (2000). Minicircular plastid DNA in the dinoflagellate *Amphidinium operculatum*. *Molecular & General Genetics: MGG*, 263(1), 152–158.
- Bar Dolev, M., Braslavsky, I., & Davies, P. L. (2016). Ice-Binding Proteins and their function. *Ann Rev Biochem*, 85, 515–542.
- Barkan, A., & Small, I. (2014). Pentatricopeptide repeat proteins in plants. *Ann Rev Plant Biol*, 65, 415–442.
- Barnacle, R., & Cuthbert, D. (2021), eds. *The PhD at the End of the World: provocations for the doctorate and a future contested*. Springer Nature. doi: 10.1007/978-3-030-62219-0
- Barylyuk, K., Koreny, L., Ke, H., Butterworth, S., Crook, O. M., Lassadi, I., Gupta, V., Tromer, E., Mourier, T., Stevens, T. J., Breckels, L. M., Pain, A., Lilley, K. S., & Waller, R. F. (2020). A comprehensive subcellular atlas of the *Toxoplasma* proteome via hyperLOPIT provides spatial context for protein functions. *Cell Host & Microbe*, 28(5), 752–766.e9.
- Becker, B., Hoef-Emden, K., & Melkonian, M. (2008). Chlamydial genes shed light on the evolution of photoautotrophic eukaryotes. *BMC Evol Biol*, 8, 203.
- Beisser, D., Graupner, N., Bock, C., Wodniok, S., Grossmann, L., Vos, M., Sures, B., Rahmann, S., & Boenigk, J. (2017). Comprehensive transcriptome analysis provides new insights into nutritional strategies and phylogenetic relationships of chrysophytes. *PeerJ*, 5, e2832.
- Benoiston, A.-S., Ibarbalz, F. M., Bittner, L., Guidi, L., Jahn, O., Dutkiewicz, S., & Bowler, C. (2017). The evolution of diatoms and their biogeochemical functions. *Phil Trans R Soc Biol*, 372(1728).
- Berge, T., Hansen, P. J., & Moestrup, Ø. (2008). Feeding mechanism, prey specificity and growth in light and dark of the plastidic dinoflagellate *Karlodinium armiger*. *Aquat Microbial Ecol*, 50, 279–288.
- Bertrand, E. M., Saito, M. A., Rose, J. M., Riesselman, C. R., Lohan, M. C., Noble, A. E., Lee, P. A., & DiTullio, G. R. (2007). Vitamin B₁₂ and iron colimitation of phytoplankton growth in the Ross Sea. *Limnol Oceanograph*, 52(3), 1079–1093.
- Bey, H., Charton, F., Cruz de Carvalho, H., Liu, S., Dorrell, R. G., Bowler, C., Boccara, C., & Boccara, M. (2022). Dynamic Cell Imaging: application to the diatom *Phaeodactylum tricorutum* under environmental stresses. *Eur J Phycol*, 1–11.**
- Blanc, G., Ogata, H., Robert, C., Audic, S., Claverie, J.-M., & Raoult, D. (2007). Lateral gene

- transfer between obligate intracellular bacteria: evidence from the *Rickettsia massiliae* genome. *Genom Res*, 17(11), 1657–1664.
- Bock, N. A., Charvet, S., Burns, J., Gyaltschen, Y., Rozenberg, A., Duhamel, S., & Kim, E. (2021). Experimental identification and *in silico* prediction of bacterivory in green algae. *ISME J*, 15(7), 1987–2000.
- Bodył, A. (2018). Did some red alga-derived plastids evolve via kleptoplastidy? A hypothesis. *Biol Rev Cambridge Phil Soc*, 93(1), 201–222.
- Boenigk, J., Beisser, D., Franke, L., Klar, L., Ilić, M., & Fink, P. (2023). Differences in pigment composition and concentration between phototrophic, mixotrophic, and heterotrophic Chrysophyceae. *Fottea*, 23(2), 223–234.
- Bork, P., Bowler, C., de Vargas, C., Gorsky, G., Karsenti, E., & Wincker, P. (2015). Tara Oceans studies plankton at planetary scale. *Science*, 348(6237), 873.
- Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otilar, R. P., Rayko, E., Salamov, A., Vandepoele, K., Beszteri, B., Gruber, A., Heijde, M., Katinka, M., Mock, T., Valentin, K., ... Grigoriev, I. V. (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature*, 456(7219), 239–244.
- Bowler, C., Vardi, A., & Allen, A. E. (2010). Oceanographic and biogeochemical insights from diatom genomes. *Ann Rev Mar Sci*, 2, 333–365.
- Brawley, S. H., Blouin, N. A., Ficko-Blean, E., Wheeler, G. L., Lohr, M., Goodson, H. V., Jenkins, J. W., Blaby-Haas, C. E., Helliwell, K. E., Chan, C. X., Marriage, T. N., Bhattacharya, D., Klein, A. S., Badis, Y., Brodie, J., Cao, Y., Collén, J., Dittami, S. M., Gachon, C. M. M., ... Prochnik, S. E. (2017). Insights into the red algae and eukaryotic evolution from the genome of (Bangiophyceae, Rhodophyta). *Proc Natl Acad Sci USA*, 114(31), E6361–E6370.
- Brodrick, J. T., Du, N., Smith, S. R., Tsuji, Y., Jallet, D., Ware, M. A., Peers, G., Matsuda, Y., Dupont, C. L., Mitchell, B. G., Palsson, B. O., & Allen, A. E. (2019). Cross-compartment metabolic coupling enables flexible photoprotective mechanisms in the diatom *Phaeodactylum tricorutum*. *New Phytol*, 222(3), 1364–1379.
- Bromke, M. A. (2013). Amino Acid biosynthesis pathways in diatoms. *Metabolites*, 3(2), 294–311.
- Browning, T. J., Achterberg, E. P., Rapp, I., Engel, A., Bertrand, E. M., Tagliabue, A., & Moore, C. M. (2017). Nutrient co-limitation at the boundary of an oceanic gyre. *Nature*, 551(7679), 242–246.
- Burki, F., Flegontov, P., Oborník, M., Cihlar, J., Pain, A., Lukes, J., & Keeling, P. J. (2012). Re-evaluating the green versus red signal in eukaryotes with secondary plastid of red algal origin. *Genom Biol Evol*, 4(6), 626–635.
- Burki, F., Hirakawa, Y., & Keeling, P. J. (2012). Intragenomic spread of plastid-targeting presequences in the coccolithophore *Emiliania huxleyi*. *Mol Biol Evol*, 29(9), 2109–2112.
- Burki, F., Imanian, B., Hehenberger, E., Hirakawa, Y., Maruyama, S., & Keeling, P. J. (2014). Endosymbiotic gene transfer in tertiary plastid-containing dinoflagellates. *Eukaryot Cell*, 13(2), 246–255.
- Buschmann, H., & Holzinger, A. (2020). Understanding the algae to land plant transition. *J Exp Bot*, 71(11), 3241–3246.
- Butterfield, E. R., Howe, C. J., & Nisbet, R. E. R. (2013). An analysis of dinoflagellate metabolism using EST data. *Protist*, 164(2), 218–236.
- Butterfield, E. R., Howe, C. J., & Nisbet, R. E. R. (2016). Identification of sequences encoding *Symbiodinium minutum* mitochondrial proteins. *Genom Biol Evol*, 8(2), 439–445.
- Cardenas, J., & Mortenson, L. E. (1974). Determination of molybdenum and tungsten in biological materials. *Anal Biochem*, 60(2), 372–381.
- Carpenter, E. J., Matasci, N., Ayyampalayam, S., Wu, S., Sun, J., Yu, J., Jimenez Vieira, F. R.,**

Bowler, C., Dorrell, R. G., Gitzendanner, M. A., Li, L., Du, W., K Ullrich, K., Wickett, N. J., Barkmann, T. J., Barker, M. S., Leebens-Mack, J. H., & Wong, G. K.-S. (2019). Access to RNA-sequencing data from 1,173 plant species: the 1000 Plant Transcriptomes initiative (1KP). *GigaScience*, 8(10).

- Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., Lima-Mendez, G., Rocha, F., Tirichine, L., Labadie, K., Kirilovsky, A., Bertrand, A., Engelen, S., Madoui, M.-A., Méheust, R., Poulain, J., Romac, S., Richter, D. J., Yoshikawa, G., ... Wincker, P. (2018). A global ocean atlas of eukaryotic genes. *Nat Commun*, 9(1), 373.
- Caspari, O. D., & Lafontaine, I. (2021). The role of antimicrobial peptides in the evolution of endosymbiotic protein import. *PLoS Pathogens*, 17(4), e1009466.
- Cavalier-Smith, T. (1999). Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J Euk Microbiol*, 46(4), 347–366.
- Cavalier-Smith, T. (2010). Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. *Biol Lett*, 6(3), 342–345.
- Cenci, U., Sibbald, S. J., Curtis, B. A., Kamikawa, R., Eme, L., Moog, D., Henrissat, B., Maréchal, E., Chabi, M., Djemiel, C., Roger, A. J., Kim, E., & Archibald, J. M. (2018). Nuclear genome sequence of the plastid-lacking cryptomonad *Goniomonas avonlea* provides insights into the evolution of secondary plastids. *BMC Biol*, 16(1), 137.
- Chaffron, S., Delage, E., Budinich, M., Vintache, D., Henry, N., Nef, C., Ardyna, M., Zayed, A. A., Junger, P. C., Galand, P. E., Lovejoy, C., Murray, A. E., Sarmiento, H., Tara Oceans coordinators, Acinas, S. G., Babin, M., Iudicone, D., Jaillon, O., Karsenti, E., ... Eveillard, D. (2021). Environmental vulnerability of the global ocean epipelagic plankton community interactome. *Sci Adv*, 7(35).
- Chan, C. Q. H., Low, L. L., & Lee, K. H. (2016). Oral vitamin B₁₂ replacement for the treatment of pernicious anemia. *Front Med*, 3, 38.
- Cho, C. H., Park, S. I., Huang, T.-Y., Lee, Y., Ciniglia, C., Yadavalli, H. C., Yang, S. W., Bhattacharya, D., & Yoon, H. S. (2023). Genome-wide signatures of adaptation to extreme environments in red algae. *Nat Commun*, 14(1), 10.
- Choi, C. J., Bachy, C., Jaeger, G. S., Poirier, C., Sudek, L., Sarma, V. V. S. S., Mahadevan, A., Giovannoni, S. J., & Worden, A. Z. (2017). Newly discovered deep-branching marine plastid lineages are numerically rare but globally distributed. *Curr Biol*, 27(1), R15–R16.
- Cock, J. M., Sterck, L., Rouzé, P., Scornet, D., Allen, A. E., Amoutzias, G., Anthouard, V., Artiguenave, F., Aury, J.-M., Badger, J. H., Beszteri, B., Billiau, K., Bonnet, E., Bothwell, J. H., Bowler, C., Boyen, C., Brownlee, C., Carrano, C. J., Charrier, B., ... Wincker, P. (2010). The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature*, 465(7298), 617–621.
- Cooney, E. C., Holt, C. C., Jacko-Reynolds, V. K. L., Leander, B. S., & Keeling, P. J. (2023). Photosystems in the eye-like organelles of heterotrophic warnowiid dinoflagellates. *Curr Biol*, 33(19), 4252–4260.e3.
- Cooney, E. C., Leander, B. S., & Keeling, P. J. (2022). Phylogenomics shows unique traits in Noctilucales are derived rather than ancestral. *PNAS Nexus*, 1(4), gac202.
- Cornwallis, C. K., Svensson-Coelho, M., Lindh, M., Li, Q., Ståbile, F., Hansson, L.-A., & Rengefors, K. (2023). Single-cell adaptations shape evolutionary transitions to multicellularity in green algae. *Nat Ecol Evol*, 7(6), 889–902.
- Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J., & Smith, A. G. (2005). Algae acquire vitamin B₁₂ through a symbiotic relationship with bacteria. *Nature*, 438(7064), 90–93.
- Cruz de Carvalho, M. H., Sun, H.-X., Bowler, C., & Chua, N.-H. (2016). Noncoding and coding transcriptome responses of a marine diatom to phosphate fluctuations. *New Phytol*,

210(2), 497–510.

- Cruz, S., & Cartaxana, P. (2022). Kleptoplasty: getting away with stolen chloroplasts. *PLoS Biology*, 20(11), e3001857.
- Daebeler, A., Herbold, C. W., Vierheilig, J., Sedlacek, C. J., Pjevac, P., Albertsen, M., Kirkegaard, R. H., de la Torre, J. R., Daims, H., & Wagner, M. (2018). Cultivation and genomic analysis of *Nitrosocaldus islandicus*, an obligately thermophilic, ammonia-oxidizing thaumarchaeon from a hot spring biofilm in Graendalur Valley, Iceland. *Front Microbiol*, 9, 193.
- Dang, Y., & Green, B. R. (2009). Substitutional editing of *Heterocapsa triquetra* chloroplast transcripts and a folding model for its divergent chloroplast 16S rRNA. *Gene*, 442(1-2), 73–80.
- Dautermann, O., & Lohr, M. (2017). 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(5), 879–891.
- Decelle, J., Probert, I., Bittner, L., Desdevises, Y., Colin, S., de Vargas, C., Galí, M., Simó, R., & Not, F. (2012). An original mode of symbiosis in open ocean plankton. *Proc Natl Acad Sci USA*, 109(44), 18000–18005.
- Del Campo, J., Carlos-Oliveira, M., Čepička, I., Hehenberger, E., Horák, A., Karnkowska, A., Kolisko, M., Lara, E., Lukeš, J., Pánek, T., Piwosz, K., Richter, D. J., Škaloud, P., Sutak, R., Tachezy, J., & Hampl, V. (2023). The protist cultural renaissance. *Trends Microbiol*. doi: 10.1016/j.tim.2023.11.010
- Delmont, T. O., Gaia, M., Hinsinger, D. D., Frémont, P., Vanni, C., Fernandez-Guerra, A., Eren, A. M., Kourlaiev, A., d'Agata, L., Clayssen, Q., Villar, E., Labadie, K., Cruaud, C., Poulain, J., Da Silva, C., Wessner, M., Noel, B., Aury, J.-M., Tara Oceans Coordinators, ... Jaillon, O. (2022). Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean. *Cell Genom*, 2(5), 100123.
- De Riso, V., Raniello, R., Maumus, F., Rogato, A., Bowler, C., & Falciatore, A. (2009). Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucl Acids Res*, 37(14), e96.
- De Salas, M. F., Bolch, C. J. S., Botes, L., Nash, G., Wright, S. W., & Hallegraeff, G. M. (2003). *Takayama* gen. nov. (Gymnodiniales, Dinophyceae), a new genus of unarmored dinoflagellates with sigmoid apical grooves, including the description of two new species. *J Phycol*, 39(6), 1233–1246.
- De Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.-M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., ... Karsenti, E. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science*, 348(6237), 1261605.
- Di Franco, A., Baurain, D., Glöckner, G., Melkonian, M., & Philippe, H. (2022). Lower statistical support with larger data sets: insights from the Ochrophyta radiation. *Mol Biol Evol*, 39(1).
- Dittami, S. M., Arboleda, E., Auguet, J.-C., Bigalke, A., Briand, E., Cárdenas, P., Cardini, U., Decelle, J., Engelen, A. H., Eveillard, D., Gachon, C. M. M., Griffiths, S. M., Harder, T., Kayal, E., Kazamia, E., Lallier, F. H., Medina, M., Marzinelli, E. M., Morganti, T. M., ... Not, F. (2021). A community perspective on the concept of marine holobionts: current status, challenges, and future directions. *PeerJ*, 9, e10911.
- Djidjou-Demasse, R., Sofonea, M. T., Choisy, M., & Alizon, S. (2023). Within-host evolutionary dynamics of antimicrobial quantitative resistance. *Math Model Nat Phenomena*, 18, 24.
- Dodge, J. D. (1971). A dinoflagellate with both a mesocaryotic and a eucaryotic nucleus. I. Fine structure of the nuclei. *Protoplasma*, 73(2), 145–157.
- Dorrell, R. G. (2020). Convergence in the RNA processing of fractured algal organelle genomes. *J Phycol*, 56(5), 1121–1123.**
- Dorrell, R. G., Azuma, T., Nomura, M., Audren de Kerdrel, G., Paoli, L., Yang, S., Bowler,**

- C., Ishii, K.-I., Miyashita, H., Gile, G. H., & Kamikawa, R. (2019). Principles of plastid reductive evolution illuminated by nonphotosynthetic chrysophytes. *Proc Natl Acad Sci USA*, *116*(14), 6914–6923.
- Dorrell, R. G., & Bowler, C. (2017). Secondary plastids of stramenopiles. *Adv Bot Res* *84*, 57–103. Elsevier.
- Dorrell, R. G., Butterfield, E. R., Nisbet, R. E. R., & Howe, C. J. (2013). Evolution: unveiling early alveolates. *Curr Biol*, *23*(24), R1093–R1096.
- Dorrell, R. G., Drew, J., Nisbet, R. E. R., & Howe, C. J. (2014). Evolution of chloroplast transcript processing in *Plasmodium* and its chromerid algal relatives. *PLoS Genet*, *10*(1), e1004008.
- Dorrell, R. G., Gile, G., McCallum, G., Méheust, R., Bapteste, E. P., Klinger, C. M., Brillet-Guéguen, L., Freeman, K. D., Richter, D. J., & Bowler, C. (2017). Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. *eLife*, *6*.
- Dorrell, R. G., Hinksman, G. A., & Howe, C. J. (2016). Diversity of transcripts and transcript processing forms in plastids of the dinoflagellate alga *Karenia mikimotoi*. *Plant Mol Biol*, *90*(3), 233–247.
- Dorrell, R. G., & Howe, C. J. (2012a). Functional remodeling of RNA processing in replacement chloroplasts by pathways retained from their predecessors. *Proc Natl Acad Sci USA*, *109*(46), 18879–18884.
- Dorrell, R. G., & Howe, C. J. (2012b). What makes a chloroplast? Reconstructing the establishment of photosynthetic symbioses. *J Cell Sci*, *125*(8), 1865–1875.
- Dorrell, R. G., & Howe, C. J. (2015). Integration of plastids with their hosts: lessons learned from dinoflagellates. *Proc Natl Acad Sci USA*, *112*(33), 10247–10254.
- Dorrell, R. G., Klinger, C. M., Newby, R. J., Butterfield, E. R., Richardson, E., Dacks, J. B., Howe, C. J., Nisbet, R. E., & Bowler, C. (2017). Progressive and biased divergent evolution underpins the origin and diversification of peridinin dinoflagellate plastids. *Mol Biol Evol*, *34*(2), 361–379.
- Dorrell, R. G., Kuo, A., Füßy, Z., Richardson, E. H., Salamov, A., Zarevski, N., Freyria, N. J., Ibarbalz, F. M., Jenkins, J., Pierella Karlusich, J. J., Stecca Steindorff, A., Edgar, R. E., Handley, L., Lail, K., Lipzen, A., Lombard, V., McFarlane, J., Nef, C., Novák Vanclová, A. M., ... Lovejoy, C. (2023). Convergent evolution and horizontal gene transfer in Arctic Ocean microalgae. *Life Sci Alliance*, *6*(3), 26508.
- Dorrell, R. G., Liu, F., & Bowler, C. (2022). Reconstructing dynamic evolutionary events in diatom nuclear and organelle genomes. In *The Molecular Life of Diatoms*, eds. Falciatore and Mock. pp. 147–177. Springer International Publishing.
- Dorrell, R. G., Nef, C., Altan-Ochir, S., Bowler, C., & Smith, A. G. (2024). Presence of vitamin B₁₂ metabolism in the last common ancestor of land plants. *Proc R Soc Biol*, *in press*. doi: 10.1098/rstb.2023.0354
- Dorrell, R. G., Nisbet, R. E. R., Barbrook, A. C., Rowden, S. J. L., & Howe, C. J. (2019). Integrated genomic and transcriptomic analysis of the peridinin dinoflagellate *Amphidinium carterae* plastid. *Protist*, *170*(4), 358–373.
- Dorrell, R. G., & Smith, A. G. (2011). Do red and green make brown?: perspectives on plastid acquisitions within chromalveolates. *Eukaryot Cell*, *10*(7), 856–868.
- Dorrell, R. G., Villain, A., Perez-Lamarque, B., Audren de Kerdrel, G., McCallum, G., Watson, A. K., Ait-Mohamed, O., Alberti, A., Corre, E., Frischkorn, K. R., Pierella Karlusich, J. J., Pelletier, E., Morlon, H., Bowler, C., & Blanc, G. (2021). Phylogenomic fingerprinting of tempo and functions of horizontal gene transfer within ochrophytes. *Proc Natl Acad Sci USA*, *118*(4).
- Dorrell, R. G., Zhang, Y., Liang, Y., Gueguen, N., Nonoyama, T., Croteau, D., Penot, M.

Adiba, S., Bailleul, B., Gros, V., Karlusich, J. J. P., Zweig, N., Fernie, A. R., Jouhet, J., Maréchal, E., & Bowler, C. (2024). Environmentally-informed functional characterization of a plastid diatom metabolic bridge of mitochondrial origin. *Plant Cell*, in resubmission doi: 10.1101/2022.09.08.507166v2

- Duchêne, A.-M., Giritch, A., Hoffmann, B., Cognat, V., Lancelin, D., Peeters, N. M., Zaepfel, M., Maréchal-Drouard, L., & Small, I. D. (2005). Dual targeting is the rule for organellar aminoacyl-tRNA synthetases in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA*, 102(45), 16484–16489.
- Ebenezer, T. E., Zoltner, M., Burrell, A., Nenarokova, A., Novák Vanclová, A. M. G., Prasad, B., Soukal, P., Santana-Molina, C., O'Neill, E., Nankisoor, N. N., Vadakedath, N., Daiker, V., Obado, S., Silva-Pereira, S., Jackson, A. P., Devos, D. P., Lukeš, J., Lebert, M., Vaughan, S., ... Field, M. C. (2019). Transcriptome, proteome and draft genome of *Euglena gracilis*. *BMC Biol*, 17(1), 11.
- Eegeesiak, O., Aariak, E., & Kleist, K. V. (2017). People of the Ice Bridge: the future of the Pikialasorsuaq. Inuit Circumpolar Council Canada. doi: 10.25607/OBP-1693
- Eglit, Y., Shiratori, T., Jerlström-Hultqvist, J., Williamson, K., Roger, A. J., Ishida, K.-I., & Simpson, A. G. B. (2024). *Meteora sporadica*, a protist with incredible cell architecture, is related to Hemimastigophora. *Curr Biol*, 34(2), 451–459.e6.
- Eikrem, W., Medlin, L. K., Henderiks, J., Rokitta, S., Rost, B., Probert, I., Throndsen, J., & Edvardsen, B. (2016). Haptophyta. In *Handbook of the Protists* (pp. 1–61). Springer International Publishing.
- Ellis, R. J. (1977). Protein synthesis by isolated chloroplasts. *Biochim Biophys Acta*, 463(2), 185–215.
- Eme, L., Gentekaki, E., Curtis, B., Archibald, J. M., & Roger, A. J. (2017). Lateral gene transfer in the adaptation of the anaerobic parasite *Blastocystis* to the gut. *Curr Biol*, 27(6), 807–820.
- Epstein, S., & López-García, P. (2007). "Missing" protists: a molecular prospective. In *Protist Diversity and Geographical Distribution* (pp. 27–42). Springer Netherlands.
- Faktorová, D., Nisbet, R. E. R., Fernández Robledo, J. A., Casacuberta, E., Sudek, L., Allen, A. E., Ares, M., Jr, Aresté, C., Balestreri, C., Barbrook, A. C., Beardslee, P., Bender, S., Booth, D. S., Bouget, F.-Y., Bowler, C., Breglia, S. A., Brownlee, C., Burger, G., Cerutti, H., ... Lukeš, J. (2020). Genetic tool development in marine protists: emerging model organisms for experimental cell biology. *Nat Method*, 17(5), 481–494.
- Falciatore, A., Casotti, R., Leblanc, C., Abrescia, C., & Bowler, C. (1999). Transformation of nonselectable reporter genes in marine diatoms. *Mar Biotechnol*, 1(3), 239–251.
- Fan, X., Qiu, H., Han, W., Wang, Y., Xu, D., Zhang, X., Bhattacharya, D., & Ye, N. (2020). Phytoplankton pangenome reveals extensive prokaryotic horizontal gene transfer of diverse functions. *Sci Adv*, 6(18), eaba0111.
- Fausser, F., Vilarrasa-Blasi, J., Onishi, M., Ramundo, S., Patena, W., Millican, M., Osaki, J., Philp, C., Nemeth, M., Salomé, P. A., Li, X., Wakao, S., Kim, R. G., Kaye, Y., Grossman, A. R., Niyogi, K. K., Merchant, S. S., Cutler, S. R., Walter, P., ... Jinkerson, R. E. (2022). Systematic characterization of gene function in the photosynthetic alga *Chlamydomonas reinhardtii*. *Nature Genetics*, 54(5), 705–714.
- Flori, S., Jouneau, P. H., Finazzi, G., Maréchal, E., & Falconet, D. (2016). Ultrastructure of the periplastidial compartment of the diatom *Phaeodactylum tricorutum*. *Protist*, 167(3), 254–267.
- Foflonker, F., Mollegard, D., Ong, M., Yoon, H. S., & Bhattacharya, D. (2018). Genomic analysis of *Picochlorum* species reveals how microalgae may adapt to variable environments. *Mol Biol Evol*, 35(11), 2702–2711.
- Fowler, N., Tomas, C., Baden, D., Campbell, L., & Bourdelais, A. (2015). Chemical analysis of

Karenia papilionacea. *Toxicon*, 101, 85–91.

- Freitas, L., & Nery, M. F. (2020). Expansions and contractions in gene families of independently-evolved blood-feeding insects. *BMC Evol Biol*, 20(1), 87.
- Freyria, N. J., Joli, N., & Lovejoy, C. (2021). A decadal perspective on north water microbial eukaryotes as Arctic Ocean sentinels. *Sci Rep*, 11(1), 8413.
- Freyria, N. J., Kuo, A., Chovatia, M., Johnson, J., Lipzen, A., Barry, K. W., Grigoriev, I. V., & Lovejoy, C. (2022). Salinity tolerance mechanisms of an Arctic pelagophyte using comparative transcriptomic and gene expression analysis. *Commun Biol*, 5(1), 500.
- Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., Le Quéré, C., Luijkx, I. T., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Alkama, R., ... Zheng, B. (2022). Global carbon budget 2022. *Earth System Science Data*, 14(11), 4811–4900.
- Froese, D. S., Fowler, B., & Baumgartner, M. R. (2019). Vitamin B, folate, and the methionine remethylation cycle-biochemistry, pathways, and regulation. *J Inher Metab Dis*, 42(4), 673–685.
- Gabrielsen, T. M., Minge, M. A., Espelund, M., Tooming-Klunderud, A., Patil, V., Nederbragt, A. J., Otis, C., Turmel, M., Shalchian-Tabrizi, K., Lemieux, C., & Jakobsen, K. S. (2011). Genome evolution of a tertiary dinoflagellate plastid. *PLoS One*, 6(4), e19132.
- Gachon, C. M. M., Heesch, S., Küpper, F. C., Achilles-Day, U. E. M., Brennan, D., Campbell, C. N., Clarke, A., Dorrell, R. G., Field, J., Gontarek, S., Rad Menendez, C., Saxon, R. J., Veszelovszki, A., Guiry, M. D., Gharbi, K., Blaxter, M., & Day, J. G. (2013). The CCAP KnowledgeBase: linking protistan and cyanobacterial biological resources with taxonomic and molecular data. *Systemat Biodiv*, 11(4), 407–413.
- Gachon, C. M. M., Sime-Ngando, T., Strittmatter, M., Chambouvet, A., & Kim, G. H. (2010). Algal diseases: spotlight on a black box. *Trends Plant Sci*, 15(11), 633–640.
- Gao, S., Yang, W., Li, X., Zhou, L., Liu, X., Wu, S., Wang, L., & Wang, G. (2024). Cryptochrome PtCPF1 regulates high temperature acclimation of marine diatoms through coordination of iron and phosphorus uptake. *ISME J*. doi: 10.1093/ismejo/wrad019
- Garrido, C., Caspari, O. D., Choquet, Y., Wollman, F.-A., & Lafontaine, I. (2020). Evidence Supporting an antimicrobial origin of targeting peptides to endosymbiotic organelles. *Cells*, 9(8), 1795.
- Gasber, A., Klaumann, S., Trentmann, O., Trampczynska, A., Clemens, S., Schneider, S., Sauer, N., Feifer, I., Bittner, F., Mendel, R. R., & Neuhaus, H. E. (2011). Identification of an *Arabidopsis* solute carrier critical for intracellular transport and inter-organ allocation of molybdate. *Plant Biol*, 13(5), 710–718.
- Gatti, D. L., & Tzagoloff, A. (1991). Structure and evolution of a group of related aminoacyl-tRNA synthetases. *J Mol Biol*, 218(3), 557–568.
- Gebremedhin, S. (2021). Trends in vitamin B₁₂ supply and prevalence of inadequate intake in Africa: regional and country-specific estimates. *Food and Nutrition Bulletin*, 42(4), 467–479.
- Gérikas Ribeiro, C., dos Santos, A. L., Gourvil, P., Le Gall, F., Marie, D., Tragin, M., Probert, I., & Vaultot, D. (2020). Culturable diversity of Arctic phytoplankton during pack ice melting. *Elementa*, 8(1), 6.
- Gile, G. H., Moog, D., Slamovits, C. H., Maier, U.-G., & Archibald, J. M. (2015). Dual organellar targeting of aminoacyl-tRNA synthetases in diatoms and cryptophytes. *Genom Biol Evol*, 7(6), 1728–1742.
- Godrijan, J., Drapeau, D., & Balch, W. M. (2020). Mixotrophic uptake of organic compounds by coccolithophores. *Limnol Oceanograph*, 65(6), 1410–1421.
- Gomez Isaza, D. F., Cramp, R. L., & Franklin, C. E. (2020). Living in polluted waters: A meta-analysis of the effects of nitrate and interactions with other environmental stressors

- on freshwater taxa. *Env Pollution*, 261, 114091.
- Gooday, A. J., Schoenle, A., Dolan, J. R., & Arndt, H. (2020). Protist diversity and function in the dark ocean - Challenging the paradigms of deep-sea ecology with special emphasis on foraminiferans and naked protists. *Eur J Protistol*, 75, 125721.
- Gornik, S. G., Febrimarsa, Cassin, A. M., MacRae, J. I., Ramaprasad, A., Rchiad, Z., McConville, M. J., Bacic, A., McFadden, G. I., Pain, A., & Waller, R. F. (2015). Endosymbiosis undone by stepwise elimination of the plastid in a parasitic dinoflagellate. *Proc Natl Acad Sci USA*, 112(18), 5767–5772.
- Gould, S. B., Sommer, M. S., Kroth, P. G., Gile, G. H., Keeling, P. J., & Maier, U.-G. (2006). Nucleus-to-nucleus gene transfer and protein retargeting into a remnant cytoplasm of cryptophytes and diatoms. *Mol Biol Evol*, 23(12), 2413–2422.
- Graeff, J. E., & Leblond, J. D. (2023). Composition of galactolipids, betaine lipids and triglyceride-associated fatty acids of the symbiotic dinoflagellate *Zooxanthella (Brandtodinium) nutricula*: A glimpse into polyunsaturated fatty acids available to its polycystine radiolarian host. *Phycol Res*. doi: 10.1111/pre.12530
- Graf, L., Shin, Y., Yang, J. H., Choi, J. W., Hwang, I. K., Nelson, W., Bhattacharya, D., Viard, F., & Yoon, H. S. (2021). A genome-wide investigation of the effect of farming and human-mediated introduction on the ubiquitous seaweed *Undaria pinnatifida*. *Nat Ecol Evol*, 5(3), 360–368.
- Graupner, N., Jensen, M., Bock, C., Marks, S., Rahmann, S., Beisser, D., & Boenigk, J. (2018). Evolution of heterotrophy in chrysophytes as reflected by comparative transcriptomics. *FEMS Microbiology Ecology*, 94(4).
- Green, B. R. (2011a). After the primary endosymbiosis: an update on the chromalveolate hypothesis and the origins of algae with Chl c. *Photosynth Research*, 107(1), 103–115.
- Green, B. R. (2011b). Chloroplast genomes of photosynthetic eukaryotes. *Plant J*, 66(1), 34–44.
- Grigoriev, I. V., Hayes, R. D., Calhoun, S., Kamel, B., Wang, A., Ahrendt, S., Dusheyko, S., Nikitin, R., Mondo, S. J., Salamov, A., Shabalov, I., & Kuo, A. (2021). PhycoCosm, a comparative algal genomics resource. *Nucl Acids Res*, 49(D1), D1004–D1011.
- Grossmann, L., Bock, C., Schweikert, M., & Boenigk, J. (2016). Small but manifold - hidden diversity in "Spumella-like flagellates." *J Euk Microbiol*, 63(4), 419–439.
- Gruber, A., Roca, G., Kroth, P. G., Armbrust, E. V., & Mock, T. (2015). Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *Plant J*, 81(3), 519–528.
- Gruber, A., Vugrinec, S., Hempel, F., Gould, S. B., Maier, U.-G., & Kroth, P. G. (2007). Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. *Plant Mol Biol*, 64(5), 519–530.
- Gschloessl, B., Guermeur, Y., & Cock, J. M. (2008). HECTAR: a method to predict subcellular targeting in heterokonts. *BMC Bioinform*, 9, 393.
- Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., Darzi, Y., Audic, S., Berline, L., Brum, J., Coelho, L. P., Espinoza, J. C. I., Malviya, S., Sunagawa, S., Dimier, C., Kandels-Lewis, S., Picheral, M., Poulain, J., Searson, S., ... Gorsky. (2016). Plankton networks driving carbon export in the oligotrophic ocean. *Nature*, 532(7600), 465–470.
- Gwak, I. G., Jung, W. S., Kim, H. J., Kang, S.-H., & Jin, E. (2010). Antifreeze protein in Antarctic marine diatom, *Chaetoceros neogracile*. *Mar Biotechnol*, 12(6), 630–639.
- Hadariová, L., Vesteg, M., Hampl, V., & Krajčovič, J. (2018). Reductive evolution of chloroplasts in non-photosynthetic plants, algae and protists. *Curr Genet*, 64(2), 365–387.
- Hamilton, A. K., Lovejoy, C., Galand, P. E., & Ingram, R. G. (2008). Water masses and biogeography of picoeukaryote assemblages in a cold hydrographically complex system. *LimnolOceanograph*, 53(3), 922–935.

- Hammond, M., Dorrell, R. G., Speijer, D., & Lukeš, J. (2022). Eukaryotic cellular intricacies shape mitochondrial proteomic complexity. *BioEssays*, 44(5), e2100258.**
- Han, K. Y., Maciszewski, K., Graf, L., Yang, J. H., Andersen, R. A., Karnkowska, A., & Yoon, H. S. (2019). Dictyochophyceae plastid genomes reveal unusual variability in their organization. *J Phycol*, 55(5), 1166–1180.
- Haxo, F. T., Kycia, J. H., Somers, G. F., Bennett, A., & Siegelman, H. W. (1976). Peridinin-chlorophyll a proteins of the dinoflagellate *Amphidinium carterae*. *Plant Physiol*, 57(2), 297–303.
- Hehenberger, E., Gast, R. J., & Keeling, P. J. (2019). A kleptoplastidic dinoflagellate and the tipping point between transient and fully integrated plastid endosymbiosis. *Proc Natl Acad Sci USA*, 116(36), 17934–17942.
- He, J., Huang, Y., Li, L., Lin, S., Ma, M., Wang, Y., & Lin, S. (2024). Novel plastid genome characteristics in *Fugacium kawagutii* and the trend of accelerated evolution of plastid proteins in dinoflagellates. *Genom Biol Evol*, 16(1).
- Helliwell, K. E., Wheeler, G. L., Leptos, K. C., Goldstein, R. E., & Smith, A. G. (2011). Insights into the evolution of vitamin B₁₂ auxotrophy from sequenced algal genomes. *Mol Biol Evol*, 28(10), 2921–2933.
- Hicks, J. L., Lassadi, I., Carpenter, E. F., Eno, M., Vardakis, A., Waller, R. F., Howe, C. J., & Nisbet, R. E. R. (2019). An essential pentatricopeptide repeat protein in the apicomplexan remnant chloroplast. *Cell Microbiol*, 21(12), e13108.
- Hirt, H., Al-Babili, S., Almeida-Trapp, M., Martin, A., Aranda, M., Bartels, D., Bennett, M., Blilou, I., Boer, D., Boulouis, A., Bowler, C., Brunel-Muguet, S., Chardon, F., Colcombet, J., Colot, V., Daszkowska-Golec, A., Dinneny, J. R., Field, B., Froehlich, K., ... Young, I. M. (2023). PlantACT! - how to tackle the climate crisis. *Trends Plant Sci*, 28(5), 537–543.
- Huang, T., Pan, Y., Maréchal, E., & Hu, H. (2023). Proteomes reveal the lipid metabolic network in the complex plastid of *Phaeodactylum tricornutum*. *Plant J*. doi: 10.1111/tpj.16477
- Huesgen, P. F., Alami, M., Lange, P. F., Foster, L. J., Schröder, W. P., Overall, C. M., & Green, B. R. (2013). Proteomic amino-terminal profiling reveals targeting information for protein import into complex plastids. *PLoS One*, 8(9), e74483.
- Ibarbalz, F. M., Henry, N., Mahé, F., Ardyna, M., Zingone, A., Scalco, E., Lovejoy, C., Lombard, F., Jaillon, O., Iudicone, D., Malviya, S., Sullivan, M. B., Chaffron, S., Karsenti, E., Babin, M., Boss, E., Wincker, P., Zinger, L., de Vargas, C., ... Coordinators, T. O. (2023). Pan-Arctic plankton community structure and its global connectivity. *Elementa (Washington, D.C.)*, 11(1).
- Imanian, B., Pombert, J.-F., Dorrell, R. G., Burki, F., & Keeling, P. J. (2012). Tertiary endosymbiosis in two dinotoms has generated little change in the mitochondrial genomes of their dinoflagellate hosts and diatom endosymbionts. *PLoS One*, 7(8), e43763.
- IPCC. (2007). *Climate Change 2007 - Mitigation of Climate Change: Working Group III contribution to the Fourth Assessment Report of the IPCC*. Cambridge University Press.
- Irwin, N. A. T., Pittis, A. A., Richards, T. A., & Keeling, P. J. (2022). Systematic evaluation of horizontal gene transfer between eukaryotes and viruses. *Nat Microbiol*, 7(2), 327–336.
- Ishida, K.-I., & Green, B. R. (2002). Second- and third-hand chloroplasts in dinoflagellates: phylogeny of oxygen-evolving enhancer 1 (PsbO) protein reveals replacement of a nuclear-encoded plastid gene by that of a haptophyte tertiary endosymbiont. *Proc Natl Acad Sci USA*, 99(14), 9294–9299.
- Jackson, C. J., Gornik, S. G., & Waller, R. F. (2013). A tertiary plastid gains RNA editing in its new host. *Mol Biol Evol*, 30(4), 788–792.
- Jackson, C. L., Ménétrey, J., Sivia, M., Dacks, J. B., & Eliáš, M. (2023). An evolutionary perspective on Arf family GTPases. *Curr Opin Cell Biol*, 85(102268), 102268.

- Jamy, M., Biber, C., Vaulot, D., Obiol, A., Jing, H., Peura, S., Massana, R., & Burki, F. (2022). Global patterns and rates of habitat transitions across the eukaryotic tree of life. *Nat Ecol Evol*, 6(10), 1458–1470.
- Janouškovec, J., Gavelis, G. S., Burki, F., Dinh, D., Bachvaroff, T. R., Gornik, S. G., Bright, K. J., Imanian, B., Strom, S. L., Delwiche, C. F., Waller, R. F., Fensome, R. A., Leander, B. S., Rohwer, F. L., & Saldarriaga, J. F. (2017). Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. *Proc Natl Acad Sci USA*, 114(2), E171–E180.
- Janouškovec, J., Horák, A., Barott, K. L., Rohwer, F. L., & Keeling, P. J. (2012). Global analysis of plastid diversity reveals apicomplexan-related lineages in coral reefs. *Curr Biol*, 22(13), 518–519.
- Janouškovec, J., Horák, A., Oborník, M., Lukes, J., & Keeling, P. J. (2010). A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proc Natl Acad Sci USA*, 107(24), 10949–10954.
- Jarvis, P. (2008). Targeting of nucleus-encoded proteins to chloroplasts in plants. *New Phytol*, 179(2), 257–285.
- Jeong, H. J., Kang, H. C., Lim, A. S., Jang, S. H., Lee, K., Lee, S. Y., Ok, J. H., You, J. H., Kim, J. H., Lee, K. H., Park, S. A., Eom, S. H., Yoo, Y. D., & Kim, K. Y. (2021). Feeding diverse prey as an excellent strategy of mixotrophic dinoflagellates for global dominance. *Sci Adv*, 7(2).
- Jeong, H. J., Ok, J. H., Lim, A. S., Kwon, J. E., Kim, S. J., & Lee, S. Y. (2016). Mixotrophy in the phototrophic dinoflagellate *Takayama helix* (family Kareniaceae): predator of diverse toxic and harmful dinoflagellates. *Harmful Algae*, 60, 92–106.
- Jiang, Y., Cao, T., Yang, Y., Zhang, H., Zhang, J., & Li, X. (2023). A chlorophyll synthase widely co-opted by phytoplankton. *Science*, 382(6666), 92–98.
- Jirsová, D., Füssy, Z., Richtová, J., Gruber, A., & Oborník, M. (2019). Morphology, ultrastructure, and mitochondrial genome of the marine non-photosynthetic bicosoecid *Cafilera*, gen. et sp. nov. *Microorganisms*, 7(8).
- Joli, N., Concia, L., Mocaer, K., Guterman, J., Laude, J., Guerin, S., Sciandra, T., Bruyant, F., Ait-Mohamed, O., Beguin, M., Forget, M.-H., Bourbousse, C., Lacour, T., Bailleul, B., Nef, C., Savoie, M., Tremblay, J.-E., Campbell, D. A., Lavaud, J., ... Bowler, C. (2023). Hypometabolism to survive the long polar night and subsequent successful return to light in the diatom *Fragilariopsis cylindrus*. *New Phytol*. doi: 10.1111/nph.19387
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589.
- Jung, W., Campbell, R. L., Gwak, Y., Kim, J. I., Davies, P. L., & Jin, E. (2016). New cysteine-rich ice-binding protein secreted from the Antarctic microalga, *Chloromonas* sp. *PLoS One*, 11(4), e0154056.
- Kamikawa, R., Tanifuji, G., Ishikawa, S. A., Ishii, K.-I., Matsuno, Y., Onodera, N. T., Ishida, K.-I., Hashimoto, T., Miyashita, H., Mayama, S., & Inagaki, Y. (2015). Proposal of a Twin Arginine Translocator system-mediated constraint against loss of ATP synthase genes from nonphotosynthetic plastid genomes. *Mol Biol Evol*, 32(10), 2598–2604.
- Kamikawa, R., Yubuki, N., Yoshida, M., Taira, M., Nakamura, N., Ishida, K.-I., Leander, B. S., Miyashita, H., Hashimoto, T., Mayama, S., & Inagaki, Y. (2015). Multiple losses of photosynthesis in *Nitzschia*(Bacillariophyceae). *Phycol Res*, 63(1), 19–28.
- Kanehisa, M., & Sato, Y. (2020). KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci*, 29(1), 28–35.
- Kato, H., Tokutsu, R., Kubota-Kawai, H., Burton-Smith, R. N., Kim, E., & Minagawa, J. (2020). Characterization of a giant PSI supercomplex in the symbiotic dinoflagellate

Symbiodiniaceae. *Plant Physiol*, 183(4), 1725–1734.

Kawachi, M., Nakayama, T., Kayama, M., Nomura, M., Miyashita, H., Bojo, O., Rhodes, L., Sym, S., Pienaar, R. N., Probert, I., Inouye, I., & Kamikawa, R. (2021). Rappemonads are haptophyte phytoplankton. *Curr Biol*, 31(11), 2395–2403.e4.

Kazamia, E., Sutak, R., Paz-Yepes, J., Dorrell, R. G., Vieira, F. R. J., Mach, J., Morrissey, J., Leon, S., Lam, F., Pelletier, E., Camadro, J.-M., Bowler, C., & Lesuisse, E. (2018). Endocytosis-mediated siderophore uptake as a strategy for Fe acquisition in diatoms. *Sci Adv*, 4(5), eaar4536.

Keeling, P. J., Burki, F., Wilcox, H. M., Allam, B., Allen, E. E., Amaral-Zettler, L. A., Armbrust, E. V., Archibald, J. M., Bharti, A. K., Bell, C. J., Beszteri, B., Bidle, K. D., Cameron, C. T., Campbell, L., Caron, D. A., Cattolico, R. A., Collier, J. L., Coyne, K., Davy, S. K., ... Worden, A. Z. (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol*, 12(6), e1001889.

Keeling, P. J., & Eglit, Y. (2023). Openly available illustrations as tools to describe eukaryotic microbial diversity. *PLoS Biol*, 21(11), e3002395.

Khan, H., Parks, N., Kozera, C., Curtis, B. A., Parsons, B. J., Bowman, S., & Archibald, J. M. (2007). Plastid genome sequence of the cryptophyte alga *Rhodomonas salina* CCMP1319: lateral transfer of putative DNA replication machinery and a test of chromist plastid phylogeny. *Mol Biol Evol*, 24(8), 1832–1842.

Klinger, C. M., Paoli, L., Newby, R. J., Wang, M. Y.-W., Carroll, H. D., Leblond, J. D., Howe, C. J., Dacks, J. B., Bowler, C., Cahoon, A. B., Dorrell, R. G., & Richardson, E. (2018). Plastid transcript editing across dinoflagellate lineages shows lineage-specific application but conserved trends. *Genom Biol Evol*, 10(4), 1019–1038.

Kosecka, M., Jabłońska, A., Flakus, A., Rodriguez-Flakus, P., Kukwa, M., & Guzow-Krzemińska, B. (2020). Trentepohlialean Algae (Trentepohliales, Ulvophyceae) show preference to selected mycobiont lineages in lichen symbioses. *J Phycol*, 56(4), 979–993.

Krause, K. (2011). Piecing together the puzzle of parasitic plant plastome evolution. *Planta*, 234(4), 647–656.

Kroth, P. G., Chiovitti, A., Gruber, A., Martin-Jezequel, V., Mock, T., Parker, M. S., Stanley, M. S., Kaplan, A., Caron, L., Weber, T., Maheswari, U., Armbrust, E. V., & Bowler, C. (2008). A model for carbohydrate metabolism in the diatom *Phaeodactylum tricorutum* deduced from comparative whole genome analysis. *PLoS One*, 3(1), e1426.

Ku, C., Nelson-Sathi, S., Roettger, M., Sousa, F. L., Lockhart, P. J., Bryant, D., Hazkani-Covo, E., McInerney, J. O., Landan, G., & Martin, W. F. (2015). Endosymbiotic origin and differential loss of eukaryotic genes. *Nature*, 524(7566), 427–432.

Kwong, W. K., Del Campo, J., Mathur, V., Vermeij, M. J. A., & Keeling, P. J. (2019). A widespread coral-infecting apicomplexan with chlorophyll biosynthesis genes. *Nature*, 568(7750), 103–107.

Lachnit, M., Buhmann, M. T., Klemm, J., Kröger, N., & Poulsen, N. (2019). Identification of proteins in the adhesive trails of the diatom *Amphora coffeaeformis*. *Phil Trans R Soc Biol*, 374(1784), 20190196.

LaConte, L. E. W., Srivastava, S., & Mukherjee, K. (2017). Probing protein kinase-ATP interactions using a fluorescent ATP analog. *Methods Mol Biol*, 1647, 171–183.

Lajeunesse, T. C., Parkinson, J. E., Gabrielson, P. W., Jeong, H. J., Reimer, J. D., Voolstra, C. R., & Santos, S. R. (2018). Systematic revision of Symbiodiniaceae highlights the antiquity and diversity of coral endosymbionts. *Curr Biol*, 28(16), 2570–2580.e6.

Lambert, S., Voznica, J., & Morlon, H. (2023). Deep learning from phylogenies for diversification analyses. *Systematic Biology*. doi: 10.1093/sysbio/syad044

- Lampe, R. H., Mann, E. L., Cohen, N. R., Till, C. P., Thamatrakoln, K., Brzezinski, M. A., Bruland, K. W., Twining, B. S., & Marchetti, A. (2018). Different iron storage strategies among bloom-forming diatoms. *Proc Natl Acad Sci USA*, *115*(52), E12275–E12284.
- Lane, N. (2015). The unseen world: reflections on Leeuwenhoek (1677) "Concerning little animals." *Phil Trans R Soc Biol*, *370*(1666).
- Larkum, A. W. D., Lockhart, P. J., & Howe, C. J. (2007). Shopping for plastids. *Trends Plant Sci*, *12*(5), 189–195.
- Le Bescot, N., Mahé, F., Audic, S., Dimier, C., Garet, M.-J., Poulain, J., Wincker, P., de Vargas, C., & Siano, R. (2016). Global patterns of pelagic dinoflagellate diversity across protist size classes unveiled by metabarcoding. *Env Microbiol*, *18*(2), 609–626.
- Le Corguillé, G., Pearson, G., Valente, M., Viegas, C., Gschloessl, B., Corre, E., Bailly, X., Peters, A. F., Jubin, C., Vacherie, B., Cock, J. M., & Leblanc, C. (2009). 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.
- Leger, M. M., Eme, L., Stairs, C. W., & Roger, A. J. (2018). Demystifying Eukaryote Lateral Gene Transfer (Response to Martin 2017). *BioEssays*, *40*(5), e1700242.
- Lenton, T. M. (2013). Environmental tipping points. *Ann Rev Env Reso*, *38*(1), 1–29.
- Leonard, G., Labarre, A., Milner, D. S., Monier, A., Soanes, D., Wideman, J. G., Maguire, F., Stevens, S., Sain, D., Grau-Bové, X., Sebé-Pedrós, A., Stajich, J. E., Paszkiewicz, K., Brown, M. W., Hall, N., Wickstead, B., & Richards, T. A. (2018). Comparative genomic analysis of the "pseudofungus.", *Hyphochytrium catenoides* *Open Biol*, *8*(1).
- Leu, E., Mundy, C. J., Assmy, P., Campbell, K., Gabrielsen, T. M., Gosselin, M., Juul-Pedersen, T., & Gradinger, R. (2015). Arctic spring awakening – steering principles behind the phenology of vernal ice algal blooms. *Progress Oceanograph*, *139*, 151–170.
- Lévesque, C. A., Brouwer, H., Cano, L., Hamilton, J. P., Holt, C., Huitema, E., Raffaele, S., Robideau, G. P., Thines, M., Win, J., Zerillo, M. M., Beakes, G. W., Boore, J. L., Busam, D., Dumas, B., Ferriera, S., Fuerstenberg, S. I., Gachon, C. M. M., Gaulin, E., ... Buell, C. R. (2010). Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genom Biol*, *11*(7), R73.
- Levy, P., Dijk, N., Gray, A., Sutton, M., Jones, M., Leeson, S., Dise, N., Leith, I., & Sheppard, L. (2019). Response of a peat bog vegetation community to long-term experimental addition of nitrogen. *J Ecol*, *107*(3), 1167–1186.
- Lie, A. A. Y., Liu, Z., Terrado, R., Tatters, A. O., Heidelberg, K. B., & Caron, D. A. (2018). A tale of two mixotrophic chrysophytes: Insights into the metabolisms of two *Ochromonas* species (Chrysophyceae) through a comparison of gene expression. *PLoS One*, *13*(2), e0192439.
- Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., Ignacio-Espinosa, J. C., Roux, S., Vincent, F., Bittner, L., Darzi, Y., Wang, J., Audic, S., Berline, L., Bontempi, G., Cabello, A. M., Coppola, L., Cornejo-Castillo, F. M., ... Raes, J. (2015). Determinants of community structure in the global plankton interactome. *Science*, *348*(6237), 1262073.
- Lim, L., & McFadden, G. I. (2010). The evolution, metabolism and functions of the apicoplast. *Phil Trans R Soc Biol*, *365*(1541), 749–763.
- Liu, H., Probert, I., Uitz, J., Claustre, H., Aris-Brosou, S., Frada, M., Not, F., & de Vargas, C. (2009). Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *Proc Natl Acad Sci USA*, *106*(31), 12803–12808.
- Liu, S., Storti, M., Finazzi, G., Bowler, C., & Dorrell, R. G. (2022). A metabolic, phylogenomic and environmental atlas of diatom plastid transporters from the model species. *Front Plant Sci*, *13*, 950467.**
- Li, W. K. W., McLaughlin, F. A., Lovejoy, C., & Carmack, E. C. (2009). Smallest algae thrive as the

- Arctic Ocean freshens. *Science*, 326(5952), 539.
- Li, X., Yan, T., Yu, R., & Zhou, M. (2019). A review of *Karenia mikimotoi*: Bloom events, physiology, toxicity and toxic mechanism. *Harmful Algae*, 90, 101702.
- Li, X., Zhang, R., Patena, W., Gang, S. S., Blum, S. R., Ivanova, N., Yue, R., Robertson, J. M., Lefebvre, P. A., Fitz-Gibbon, S. T., Grossman, A. R., & Jonikas, M. C. (2016). An indexed, mapped mutant library enables reverse genetics studies of biological processes in *Chlamydomonas reinhardtii*. *Plant Cell*, 28(2), 367–387.
- Lovelock, J. (2000). *The Ages of Gaia: A Biography of Our Living Earth*. Oxford University Press.
- Lukes, J., Leander, B. S., & Keeling, P. J. (2009). Cascades of convergent evolution: the corresponding evolutionary histories of euglenozoans and dinoflagellates. *Proc Natl Acad Sci USA*, 106 (1), 9963–9970.
- Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., Wincker, P., Ludicone, D., de Vargas, C., Bittner, L., Zingone, A., & Bowler, C. (2016). Insights into global diatom distribution and diversity in the world's ocean. *Proc Natl Acad Sci USA*, 113(11), 1516–1525.
- Marchetti, A., Parker, M. S., Moccia, L. P., Lin, E. O., Arrieta, A. L., Ribalet, F., Murphy, M. E. P., Maldonado, M. T., & Armbrust, E. V. (2009). Ferritin is used for iron storage in bloom-forming marine pennate diatoms. *Nature*, 457(7228), 467–470.
- Margulis, L., & Dolan, M. (2002). *Early Life: Evolution on the Precambrian Earth*. Jones & Bartlett Learning.
- Marin, B., Palm, A., Klingberg, M., & Melkonian, M. (2003). Phylogeny and taxonomic revision of plastid-containing euglenophytes based on SSU rDNA sequence comparisons and synapomorphic signatures in the SSU rRNA secondary structure. *Protist*, 154(1), 99–145.
- Marinov, G. K., Trevino, A. E., Xiang, T., Kundaje, A., Grossman, A. R., & Greenleaf, W. J. (2021). Transcription-dependent domain-scale three-dimensional genome organization in the dinoflagellate *Breviolum minutum*. *Nat Genet*, 53(5), 613–617.
- Martin, W. F. (2017). Too much eukaryote LGT. *BioEssays*, 39(12).
- Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., & Penny, D. (2002). Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci USA*, 99(19), 12246–12251.
- Mathur, V., Del Campo, J., Kolisko, M., & Keeling, P. J. (2018). Global diversity and distribution of close relatives of apicomplexan parasites. *Env Microbiol*, 20(8), 2824–2833.
- Matsumoto, H., Takechi, K., Sato, H., Takio, S., & Takano, H. (2012). Treatment with antibiotics that interfere with peptidoglycan biosynthesis inhibits chloroplast division in the desmid *Closterium*. *PLoS One*, 7(7), e40734.
- Matsuo, E., & Inagaki, Y. (2018). Patterns in evolutionary origins of heme, chlorophyll and isopentenyl diphosphate biosynthetic pathways suggest non-photosynthetic periods prior to plastid replacements in dinoflagellates. *PeerJ*, 6, e5345.
- Maumus, F., & Blanc, G. (2016). Study of gene trafficking between *Acanthamoeba* and giant viruses suggests an undiscovered family of amoeba-infecting viruses. *Genom Biol Evol*, 8(11), 3351–3363.
- McCallum, G., & Potvin-Trottier, L. (2021). Using models to (re-)design synthetic circuits. *Methods Mol Biol*, 2229, 91–118.
- McCallum, G., & Tropini, C. (2023). The gut microbiota and its biogeography. *Nat Rev Microbiology*. doi: 10.1038/s41579-023-00969-0
- McCarthy, J. K., Smith, S. R., McCrow, J. P., Tan, M., Zheng, H., Beerli, K., Roth, R., Lichtle, C., Goodenough, U., Bowler, C. P., Dupont, C. L., & Allen, A. E. (2017). Nitrate Reductase knockout uncouples nitrate transport from nitrate assimilation and drives repartitioning of carbon flux in a model pennate diatom. *Plant Cell*, 29(8), 2047–2070.

- McKie-Krisberg, Z. M., Gast, R. J., & Sanders, R. W. (2015). Physiological responses of three species of Antarctic mixotrophic phytoflagellates to changes in light and dissolved nutrients. *Microb Ecol*, *70*(1), 21–29.
- Meade, M. J., Proulx, G. C. R., Manoylov, K. M., & Cahoon, A. B. (2020). Chloroplast mRNAs are 3' polyuridylylated in the green alga *Pithophora roettleri* (Cladophorales). *J Phycol*, *56*(5), 1124–1134.
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucl Acids Res*, *49*(D1), D412–D419.
- Molina, J., Hazzouri, K. M., Nickrent, D., Geisler, M., Meyer, R. S., Pentony, M. M., Flowers, J. M., Pelsler, P., Barcelona, J., Inovejas, S. A., Uy, I., Yuan, W., Wilkins, O., Michel, C.-I., Locklear, S., Concepcion, G. P., & Purugganan, M. D. (2014). Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). *Mol Biol Evol*, *31*(4), 793–803.
- Moog, D., Nozawa, A., Tozawa, Y., & Kamikawa, R. (2020). Substrate specificity of plastid phosphate transporters in a non-photosynthetic diatom and its implication in evolution of red alga-derived complex plastids. *Sci Rep*, *10*(1), 1167.
- Moreira, D., Tavera, R., Benzerara, K., Skouri-Panet, F., Couradeau, E., Gérard, E., Fonta, C. L., Novelo, E., Zivanovic, Y., & López-García, P. (2017). Description of *Gloeomargarita lithophora* gen. nov., sp. nov., a thylakoid-bearing, basal-branching cyanobacterium with intracellular carbonates, and proposal for Gloeomargaritales ord. nov. *Int J Syst Evol Microbiol*, *67*(3), 653–658.
- Morse, D., Salois, P., Markovic, P., & Hastings, J. W. (1995). A nuclear-encoded form II RuBisCO in dinoflagellates. *Science*, *268*(5217), 1622–1624.
- Moustafa, A., Beszteri, B., Maier, U. G., Bowler, C., Valentin, K., & Bhattacharya, D. (2009). Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science*, *324*(5935), 1724–1726.
- Moustafa, A., & Bhattacharya, D. (2008). PhyloSort: a user-friendly phylogenetic sorting tool and its application to estimating the cyanobacterial contribution to the nuclear genome of *Chlamydomonas*. *BMC Evol Biol*, *8*, 6.
- Msuya, F. E., & Hurtado, A. Q. (2017). The role of women in seaweed aquaculture in the Western Indian Ocean and South-East Asia. *Eur J Phycol*, *52*(4), 482–494.
- Mulvey, C. M., Breckels, L. M., Geladaki, A., Britovšek, N. K., Nightingale, D. J. H., Christoforou, A., Elzek, M., Deery, M. J., Gatto, L., & Lilley, K. S. (2017). Using hyperLOPIT to perform high-resolution mapping of the spatial proteome. *Nat Protocol*, *12*(6), 1110–1135.
- Mungpakdee, S., Shinzato, C., Takeuchi, T., Kawashima, T., Koyanagi, R., Hisata, K., Tanaka, M., Goto, H., Fujie, M., Lin, S., Satoh, N., & Shoguchi, E. (2014). Massive gene transfer and extensive RNA editing of a symbiotic dinoflagellate plastid genome. *Genom Biol Evol*, *6*(6), 1408–1422.
- Muñoz-Gómez, S. A., Mejía-Franco, F. G., Durnin, K., Colp, M., Grisdale, C. J., Archibald, J. M., & Slamovits, C. H. (2017). The new red algal subphylum Proteorhodophytina comprises the largest and most divergent plastid genomes known. *Curr Biol*, *27*(11), 1677–1684.e4.
- Mylnikov, A. P., Mylnikova, Z. M., & Tikhonenkov, D. V. (2008). The main cell morphology of the freshwater colorless chryomonad *Spumella* sp. (Ochromonadales, Chrysophyceae). *Inland Water Biol*, *1*(1), 32–36.
- Nand, A., Zhan, Y., Salazar, O. R., Aranda, M., Voolstra, C. R., & Dekker, J. (2021). Genetic and spatial organization of the unusual chromosomes of the dinoflagellate *Symbiodinium microadriaticum*. *Nat Genet*, *53*(5), 618–629.
- Nash, E. A., Barbrook, A. C., Edwards-Stuart, R. K., Bernhardt, K., Howe, C. J., & Nisbet, R. E. R.

- (2007). Organization of the mitochondrial genome in the dinoflagellate *Amphidinium carterae*. *Mol Biol Evol*, 24(7), 1528–1536.
- Nef, C., Jung, S., Mairet, F., Kaas, R., Grizeau, D., & Garnier, M. (2019). How haptophyte microalgae mitigate vitamin B limitation. *Sci Rep*, 9(1), 8417.
- Nef, C., Madoui, M.-A., Pelletier, É., & Bowler, C. (2022). Whole-genome scanning reveals environmental selection mechanisms that shape diversity in populations of the epipelagic diatom *Chaetoceros*. *PLoS Biol*, 20(11), e3001893.
- Nelson, D. R., Hazzouri, K. M., Lauersen, K. J., Jaiswal, A., Chaiboonchoe, A., Mystikou, A., Fu, W., Daakour, S., Dohai, B., Alzahmi, A., Nobles, D., Hurd, M., Sexton, J., Preston, M. J., Blanchette, J., Lomas, M. W., Amiri, K. M. A., & Salehi-Ashtiani, K. (2021). Large-scale genome sequencing reveals the driving forces of viruses in microalgal evolution. *Cell Host Microb*, 29(2), 250–266.e8.
- Nelson, M. J., Dang, Y., Filek, E., Zhang, Z., Yu, V. W. C., Ishida, K.-I., & Green, B. R. (2007). Identification and transcription of transfer RNA genes in dinoflagellate plastid minicircles. *Gene*, 392(1-2), 291–298.
- Neupert, W. (1997). Protein import into mitochondria. *Ann Rev Biochem*, 66, 863–917.
- Nhamo, G., Togo, M., & Dube, K. (2021). Sustainable Development Goals for Society Vol. 1: Selected topics of global relevance. Springer Nature.
- Nimmo, I. C., Barbrook, A. C., Lassadi, I., Chen, J. E., Geisler, K., Smith, A. G., Aranda, M., Purton, S., Waller, R. F., Nisbet, R. E. R., & Howe, C. J. (2019). Genetic transformation of the dinoflagellate chloroplast. *eLife*, 8, 45292.
- Nonoyama, T., Kazamia, E., Nawaly, H., Gao, X., Tsuji, Y., Matsuda, Y., Bowler, C., Tanaka, T., & Dorrell, R. G. (2019). Metabolic innovations underpinning the origin and diversification of the diatom chloroplast. *Biomolecules*, 9(8), 484.**
- Novak, L. V. F., Muñoz-Gómez, S. A., van Beveren, F., Ciobanu, M., Eme, L., López-García, P., & Moreira, D. (2024). Nucleomorph phylogenomics suggests a deep and ancient origin of cryptophyte plastids within Rhodophyta. *BioRxiv*. doi: 10.1101/2024.03.10.584144
- Novák Vanclová, A. M. G., Nef, C., Vancl, A., Liu, F., Füßsy, Z., Bowler, C., & Dorrell, R. G. (2024). New plastids, old proteins: repeated endosymbiotic acquisitions in Kareniacean dinoflagellates. *EMBO Reports*, in press. doi: 10.1038/s44319-024-00103-y**
- Novák Vanclová, A. M. G., Zoltner, M., Kelly, S., Soukal, P., Záhonová, K., Füßsy, Z., Ebenezer, T. E., Lacová Dobáková, E., Eliáš, M., Lukeš, J., Field, M. C., & Hampl, V. (2020). Metabolic quirks and the colourful history of the *Euglena gracilis* secondary plastid. *New Phytol*, 225(4), 1578–1592.
- Nowack, E. C. M., Vogel, H., Groth, M., Grossman, A. R., Melkonian, M., & Glöckner, G. (2011). Endosymbiotic gene transfer and transcriptional regulation of transferred genes in *Paulinella chromatophora*. *Mol Biol Evol*, 28(1), 407–422.
- Oborník, M., Modrý, D., Lukeš, M., Cernotíková-Stříbrná, E., Cihlář, J., Tesařová, M., Kotabová, E., Vancová, M., Prášil, O., & Lukeš, J. (2012). Morphology, ultrastructure and life cycle of *Vitrella brassicaformis* n. sp., n. gen., a novel chromerid from the Great Barrier Reef. *Protist*, 163(2), 306–323.
- One Thousand Plant Transcriptomes Initiative. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*, 574(7780), 679–685.
- Ong, H. C., Wilhelm, S. W., Gobler, C. J., Bullerjahn, G., Jacobs, M. A., McKay, J., Sims, E. H., Gillett, W. G., Zhou, Y., Haugen, E., Rocap, G., & Cattolico, R. A. (2010). Analyses of the complete chloroplast genome sequences of two members of the Pelagophyceae: *Aureococcus anophagefferens* CCMP1984 and *Aureoumbra lagunensis* CCMP1507. *J Phycol*, 46(3), 602–615.

- Oteri, F., Sarti, E., Nadalin, F., & Carbone, A. (2022). iBIS2Analyzer: a web server for a phylogeny-driven coevolution analysis of protein families. *Nucl Acids Res*, *50*(W1), W412–W419.
- Oude Vrielink, A. S., Aloij, A., Olijve, L. L. C., & Voets, I. K. (2016). Interaction of ice binding proteins with ice, water and ions. *Biointerphases*, *11*(1), 018906.
- Pánek, T., Barcytė, D., Treitli, S. C., Záhonová, K., Sokol, M., Ševčíková, T., Zadrobílková, E., Jaške, K., Yubuki, N., Čepička, I., & Eliáš, M. (2022). A new lineage of non-photosynthetic green algae with extreme organellar genomes. *BMC Biol*, *20*(1), 66.
- Patron, N. J., & Waller, R. F. (2007). Transit peptide diversity and divergence: a global analysis of plastid targeting signals. *BioEssays*, *29*(10), 1048–1058.
- Penot, M., Dacks, J. B., Read, B., & Dorrell, R. G. (2022). Genomic and meta-genomic insights into the functions, diversity and global distribution of haptophyte algae. *Appl Phycol*, *3*(1), 340–359.**
- Perez-Lamarque, B., & Morlon, H. (2019). Characterizing symbiont inheritance during host-microbiota evolution: application to the great apes gut microbiota. *Mol Ecol Res*, *19*(6), 1659–1671.
- Perez-Lamarque, B., Sommeria-Klein, G., Duret, L., & Morlon, H. (2023). Phylogenetic comparative approach reveals evolutionary conservatism, ancestral composition, and integration of vertebrate gut microbiota. *Mol Biol Evol*, *40*(7), 144.
- Petersen, J., Teich, R., Brinkmann, H., & Cerff, R. (2006). A “green” phosphoribulokinase in complex algae with red plastids: evidence for a single secondary endosymbiosis leading to haptophytes, cryptophytes, heterokonts, and dinoflagellates. *J Mol Evol*, *62*(2), 143–157.
- Pierella Karlusich, J. J., Nef, C., Bowler, C., & Dorrell, R. G. (2022). Biogeographical patterns and genomes of aquatic photoautotrophs. In *Blue Planet, Red and Green Photosynthesis*, eds. Gontero and Maberly, pp. 43–79. Wiley.**
- Pierella Karlusich, J. J., Pelletier, E., Zinger, L., Lombard, F., Zingone, A., Colin, S., Gasol, J. M., Dorrell, R. G., Henry, N., Scalco, E., Acinas, S. G., Wincker, P., de Vargas, C., & Bowler, C. (2023). A robust approach to estimate relative phytoplankton cell abundances from metagenomes. *Mol Ecol Res*, *23*(1), 16–40.**
- Pires, J. C. M. (2017). COP21: The algae opportunity? *Renew & Sust Energy Rev.*, *79*, 867–877.
- Pitaloka, M. K., Petcharat, V., Arikrit, S., & Sunpapao, A. (2015). *Cephaleuros virescens*, the cause of an algal leaf spot on para rubber in Thailand. *Australasian Plant Disease Notes*, *10*(1). doi: 10.1007/s13314-015-0158-1
- Place, A. R., Bowers, H. A., Bachvaroff, T. R., Adolf, J. E., Deeds, J. R., & Sheng, J. (2012). *Karlodinium veneficum*- the little dinoflagellate with a big bite. *Harmful Algae*, *14*, 179–195.
- Pochon, X., Montoya-Burgos, J. I., Stadelmann, B., & Pawlowski, J. (2006). Molecular phylogeny, evolutionary rates, and divergence timing of the symbiotic dinoflagellate genus *Symbiodinium*. *Mol Phylogenet Evol*, *38*(1), 20–30.
- Ponce-Toledo, R. I., Deschamps, P., López-García, P., Zivanovic, Y., Benzerara, K., & Moreira, D. (2017). An early-branching freshwater cyanobacterium at the origin of plastids. *Curr Biol*, *27*(3), 386–391.
- Ponce-Toledo, R. I., López-García, P., & Moreira, D. (2019). Horizontal and endosymbiotic gene transfer in early plastid evolution. *New Phytol*, *224*(2), 618–624.
- Prabhakar, V., Löttgert, T., Geimer, S., Dörmann, P., Krüger, S., Vijayakumar, V., Schreiber, L., Göbel, C., Feussner, K., Feussner, I., Marin, K., Staehr, P., Bell, K., Flügge, U.-I., & Häusler, R. E. (2010). Phosphoenolpyruvate provision to plastids is essential for gametophyte and sporophyte development in *Arabidopsis thaliana*. *Plant Cell*, *22*(8), 2594–2617.
- Price, M. N., Deutschbauer, A. M., & Arkin, A. P. (2021). Four families of folate-independent methionine synthases. *PLoS Genet*, *17*(2), e1009342.

- Prihoda, J., Tanaka, A., de Paula, W. B. M., Allen, J. F., Tirichine, L., & Bowler, C. (2012). Chloroplast-mitochondria cross-talk in diatoms. *J Exp Bot*, *63*(4), 1543–1557.
- Probert, I., Siano, R., Poirier, C., Decelle, J., Biard, T., Tuji, A., Suzuki, N., & Not, F. (2014). *Brandtodinium* gen. nov. and *B. nutricula* comb. Nov. (Dinophyceae), a dinoflagellate commonly found in symbiosis with polycystine radiolarians. *J Phycol*, *50*(2), 388–399.
- Rastogi, A., Maheswari, U., Dorrell, R. G., Vieira, F. R. J., Maumus, F., Kustka, A., McCarthy, J., Allen, A. E., Kersey, P., Bowler, C., & Tirichine, L. (2018). Integrative analysis of large scale transcriptome data draws a comprehensive landscape of *Phaeodactylum tricornerutum* genome and evolutionary origin of diatoms. *Scientific Reports*, *8*(1), 4834.**
- Raymond, J. A. (2014). The ice-binding proteins of a snow alga, *Chloromonas brevispina*: probable acquisition by horizontal gene transfer. *Extremophiles*, *18*(6), 987–994.
- Raymond, J. A., & Kim, H. J. (2012). Possible role of horizontal gene transfer in the colonization of sea ice by algae. *PLoS One*, *7*(5), e35968.
- Reddy, M. M., Goossens, C., Zhou, Y., Chaib, S., Raviglione, D., Nicolè, F., Hume, B. C. C., Forcioli, D., Agostini, S., Boissin, E., Boss, E., Bowler, C., de Vargas, C., Douville, E., Flores, M., Furla, P., Galand, P. E., Gilson, E., Lombard, F., ... Banaigs, B. (2023). Multi-omics determination of metabolome diversity in natural coral populations in the Pacific Ocean. *Commun Earth Env*, *4*(1)
- Reyes-Prieto, A., Hackett, J. D., Soares, M. B., Bonaldo, M. F., & Bhattacharya, D. (2006). Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. *Curr Biol*, *16*(23), 2320–2325.
- Rice, D. W., & Palmer, J. D. (2006). An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. *BMC Biology*, *4*, 31.
- Richardson, E., Dorrell, R. G., & Howe, C. J. (2014). Genome-wide transcript profiling reveals the coevolution of plastid gene sequences and transcript processing pathways in the fucoxanthin dinoflagellate *Karlodinium veneficum*. *Mol Biol Evol*, *31*(9), 2376–2386.
- Richardson, E., Zerr, K., Tsaousis, A., Dorrell, R. G., & Dacks, J. B. (2015). Evolutionary cell biology: functional insight from “endless forms most beautiful.” *Mol Biol Cell*, *26*(25), 4532–4538.**
- Richter, D. J., Watteaux, R., Vannier, T., Leconte, J., Frémont, P., Reygondeau, G., Maillet, N., Henry, N., Benoit, G., Da Silva, O., Delmont, T. O., Fernández-Guerra, A., Suweis, S., Narci, R., Berney, C., Eveillard, D., Gavory, F., Guidi, L., Labadie, K., ... Jaillon, O. (2022). Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems. *eLife*, *11*, 78129
- Richtová, J., Sheiner, L., Gruber, A., Yang, S.-M., Kořený, L., Striepen, B., & Oborník, M. (2021). Using diatom and apicomplexan models to study the heme pathway of *Chromera velia*. *Int J Mol Sci*, *22*(12).
- Rius, M., Rest, J. S., Filloramo, G. V., Novák Vanclová, A. M. G., Archibald, J. M., & Collier, J. L. (2023). Horizontal gene transfer and fusion spread carotenogenesis among diverse heterotrophic protists. *Genom Biol Evol*, *15*(3), 029
- Robbens, S., Derelle, E., Ferraz, C., Wuyts, J., Moreau, H., & Van de Peer, Y. (2007). The complete chloroplast and mitochondrial DNA sequence of *Ostreococcus tauri*: organelle genomes of the smallest eukaryote are examples of compaction. *Mol Biol Evol*, *24*(4), 956–968.
- Sagan, L. (1967). On the origin of mitosing cells. *J Theoret Biol*, *14*(3), 255–274.
- Sarai, C., Tanifuji, G., Nakayama, T., Kamikawa, R., Takahashi, K., Yazaki, E., Matsuo, E., Miyashita, H., Ishida, K.-I., Iwataki, M., & Inagaki, Y. (2020). Dinoflagellates with relic

endosymbiont nuclei as models for elucidating organellogenesis. *Proc Natl Acad Sci USA*, 117(10), 5364–5375.

- Sato, S., Nanjappa, D., Dorrell, R. G., Vieira, F. R. J., Kazamia, E., Tirichine, L., Veluchamy, A., Heilig, R., Aury, J.-M., Jaillon, O., Wincker, P., Fussy, Z., Obornik, M., Muñoz-Gómez, S. A., Mann, D. G., Bowler, C., & Zingone, A. (2020). Genome-enabled phylogenetic and functional reconstruction of an araphid pennate diatom *Plagiostriata* sp. CCMP470, previously assigned as a radial centric diatom, and its bacterial commensal. *Sci Rep*, 10(1), 9449.**
- Schön, M. E., Zlatogursky, V. V., Singh, R. P., Poirier, C., Wilken, S., Mathur, V., Strassert, J. F. H., Pinhassi, J., Worden, A. Z., Keeling, P. J., Ettema, T. J. G., Wideman, J. G., & Burki, F. (2021). Single cell genomics reveals plastid-lacking Picozoa are close relatives of red algae. *Nat Commun*, 12(1), 6651.
- Schwarz, G., & Mendel, R. R. (2006). Molybdenum cofactor biosynthesis and molybdenum enzymes. *Ann Rev Plant Biol*, 57, 623–647.
- Seeleuthner, Y., Mondy, S., Lombard, V., Carradec, Q., Pelletier, E., Wessner, M., Leconte, J., Mangot, J.-F., Poulain, J., Labadie, K., Logares, R., Sunagawa, S., de Berardinis, V., Salanoubat, M., Dimier, C., Kandels-Lewis, S., Picheral, M., Searson, S., Tara Oceans Coordinators, ... Wincker, P. (2018). Single-cell genomics of multiple uncultured stramenopiles reveals underestimated functional diversity across oceans. *Nat Commun*, 9(1), 310.
- Sekiguchi, H., Moriya, M., Nakayama, T., & Inouye, I. (2002). Vestigial chloroplasts in heterotrophic stramenopiles *Pteridomonas danica* and *Ciliophrys infusionum* (Dictyochophyceae). *Protist*, 153(2), 157–167.
- Ševčíková, T., Horák, A., Klimeš, V., Zbránková, V., Demir-Hilton, E., Sudek, S., Jenkins, J., Schmutz, J., Přibyl, P., Fousek, J., Vlček, Č., Lang, B. F., Oborník, M., Worden, A. Z., & Eliáš, M. (2015). Updating algal evolutionary relationships through plastid genome sequencing: did alveolate plastids emerge through endosymbiosis of an ochrophyte? *Sci Rep*, 5, 10134.
- Shao, Z., Ampomah, O., Vieira, F. R. J., Dorrell, R. G., Li, S., Tirichine, L., Bulone, V., Duan, D., & Bowler, C. (2023). Characterization of a marine diatom chitin synthase using a combination of meta-omics, genomics, and heterologous expression approaches. *mSystems*, 8(2), e0113122.**
- Shareefdeen, Z., Elkamel, A., & Babar, Z. B. (2023). Recent developments on the performance of algal bioreactors for CO removal: focusing on the light Intensity and photoperiods. *Biotech (Basel (Switzerland))*, 12(1), 10.
- Siano, R., Montresor, M., Probert, I., Not, F., & de Vargas, C. (2010). *Pelagodinium* gen. nov. and *P. béii* comb. nov., a dinoflagellate symbiont of planktonic foraminifera. *Protist*, 161(3), 385–399.
- Siaut, M., Heijde, M., Mangogna, M., Montsant, A., Coesel, S., Allen, A., Manfredonia, A., Falciatore, A., & Bowler, C. (2007). Molecular toolbox for studying diatom biology in *Phaeodactylum tricorutum*. *Gene*, 406(1-2), 23–35.
- Silke, J., O'Beirn, F., Cronin, M., & Marine Institute (Dublin, Ireland). (2005). *Karenia mikimotoi*: an exceptional dinoflagellate bloom in western Irish waters, summer 2005. Irish Marine Institute
- Simon, N., Foulon, E., Grulois, D., Six, C., Desdevises, Y., Latimier, M., Le Gall, F., Tragin, M., Houdan, A., Derelle, E., Jouenne, F., Marie, D., Le Panse, S., Vulot, D., & Marin, B. (2017). Revision of the genus *Micromonas* Manton et Parke (Chlorophyta, Mamiellophyceae), of the type species *M. pusilla* (Butcher) Manton & Parke and of the species *M. commoda* van Baren, Bachy and Worden and description of two new species based on the genetic and phenotypic characterization of cultured isolates. *Protist*, 168(5), 612–635.

- Singer, D., Seppely, C. V. W., Lentendu, G., Dunthorn, M., Bass, D., Belbahri, L., Blandenier, Q., Debroas, D., de Groot, G. A., de Vargas, C., Domaizon, I., Duckert, C., Izaguirre, I., Koenig, I., Mataloni, G., Schiaffino, M. R., Mitchell, E. A. D., Geisen, S., & Lara, E. (2021). Protist taxonomic and functional diversity in soil, freshwater and marine ecosystems. *Env Int* 146, 106262.
- Small, I. D., Rackham, O., & Filipovska, A. (2013). Organelle transcriptomes: products of a deconstructed genome. *Curr Opin Microbiol*, 16(5), 652–658.
- Smith, D. R., & Lee, R. W. (2014). A plastid without a genome: evidence from the nonphotosynthetic green algal genus *Polytomella*. *Plant Physiol*, 164(4), 1812–1819.
- Smith, D. R., & Sanitá Lima, M. (2017). Unraveling chloroplast transcriptomes with ChloroSeq, an organelle RNA-Seq bioinformatics pipeline. *Briefings Bioinform*, 18(6), 1012–1016.
- Smith, S. R., Abbriano, R. M., & Hildebrand, M. (2012). Comparative analysis of diatom genomes reveals substantial differences in the organization of carbon partitioning pathways. *Algal Res*, 1(1), 2–16.
- Smith, S. R., Dupont, C. L., McCarthy, J. K., Broddrick, J. T., Oborník, M., Horák, A., Füßy, Z., Cihlář, J., Kleessen, S., Zheng, H., McCrow, J. P., Hixson, K. K., Araújo, W. L., Nunes-Nesi, A., Fernie, A., Nikoloski, Z., Palsson, B. O., & Allen, A. E. (2019). Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. *Nat Commun*, 10(1), 4552.
- Smith, S. R., Gillard, J. T. F., Kustka, A. B., McCrow, J. P., Badger, J. H., Zheng, H., New, A. M., Dupont, C. L., Obata, T., Fernie, A. R., & Allen, A. E. (2016). Transcriptional orchestration of the global cellular response of a model pennate diatom to diel light cycling under Iron limitation. *PLoS Genet*, 12(12), e1006490.
- Sørensen, M. E. S., Zlatogursky, V. V., Onuț-Brännström, I., Walraven, A., Foster, R. A., & Burki, F. (2023). A novel kleptoplastidic symbiosis revealed in the marine centrohelid *Meringosphaera* with evidence of genetic integration. *Curr Biol*, 33(17), 3571–3584.e6.
- Stabler, S. P., & Allen, R. H. (2004). Vitamin B₁₂ deficiency as a worldwide problem. *Ann Rev Nutrition*, 24, 299–326.
- Stairs, C. W., Dharamshi, J. E., Tamarit, D., Eme, L., Jørgensen, S. L., Spang, A., & Ettema, T. J. G. (2020). Chlamydial contribution to anaerobic metabolism during eukaryotic evolution. *Science Advances*, 6(35), eabb7258.
- Stegemann, S., & Bock, R. (2006). Experimental reconstruction of functional gene transfer from the tobacco plastid genome to the nucleus. *Plant Cell*, 18(11), 2869–2878.
- Stephens, T. G., González-Pech, R. A., Cheng, Y., Mohamed, A. R., Burt, D. W., Bhattacharya, D., Ragan, M. A., & Chan, C. X. (2020). Genomes of the dinoflagellate *Polarella glacialis* encode tandemly repeated single-exon genes with adaptive functions. *BMC Biol*, 18(1), 56.
- Stiller, J. W., Huang, J., Ding, Q., Tian, J., & Goodwillie, C. (2009). Are algal genes in nonphotosynthetic protists evidence of historical plastid endosymbioses? *BMC Genom*, 10, 484.
- Stiller, J. W., Schreiber, J., Yue, J., Guo, H., Ding, Q., & Huang, J. (2014). The evolution of photosynthesis in chromist algae through serial endosymbioses. *Nat Commun*, 5, 5764.
- Stork, S., Moog, D., Przyborski, J. M., Wilhelmi, I., Zauner, S., & Maier, U. G. (2012). Distribution of the SELMA translocon in secondary plastids of red algal origin and predicted uncoupling of ubiquitin-dependent translocation from degradation. *Eukaryot Cell*, 11(12), 1472–1481.
- Strassert, J. F. H., Irisarri, I., Williams, T. A., & Burki, F. (2021). A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. *Nat Commun*, 12(1), 1879.
- Sutherland, E. W., Posternak, T., & Cori, C. F. (1949). Mechanism of the phosphoglyceric

- mutase reaction. *J Biol Chem*, 181(1), 153–159.
- Tabita, F. R., Hanson, T. E., Li, H., Satagopan, S., Singh, J., & Chan, S. (2007). Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs. *Microbiol Mol Biol Rev*, 71(4), 576–599.
- Takano, H., & Takechi, K. (2010). Plastid peptidoglycan. *Biochim Biophys Acta*, 1800(2), 144–151.
- Takenaka, M., Zehrmann, A., Verbitskiy, D., Kugelmann, M., Härtel, B., & Brennicke, A. (2012). Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proc Natl Acad Sci USA*, 109(13), 5104–5109.
- Takishita, K., Nakano, K., & Uchida, A. (1999). Preliminary phylogenetic analysis of plastid-encoded genes from an anomalously pigmented dinoflagellate *Gymnodinium mikimotoi* (Gymnodiniales, Dinophyta). *Phycol Res*, 47(4), 257–262.
- Tanaka, A., De Martino, A., Amato, A., Montsant, A., Mathieu, B., Rostaing, P., Tirichine, L., & Bowler, C. (2015). Ultrastructure and membrane traffic during cell division in the marine pennate diatom *Phaeodactylum tricornutum*. *Protist*, 166(5), 506–521.
- Tanaka, K., Oikawa, K., Ohta, N., Kuroiwa, H., Kuroiwa, T., & Takahashi, H. (1996). Nuclear encoding of a chloroplast RNA polymerase sigma subunit in a red alga. *Science*, 272(5270), 1932–1935.
- Tejada-Jiménez, M., Galván, A., & Fernández, E. (2011). Algae and humans share a molybdate transporter. *Proc Natl Acad Sci USA*, 108(16), 6420–6425.
- Terpis, K. X., Salomaki, E. D., Barcytė, D., Pánek, T., Verbruggen, H., Kolisko, M., Bailey, J. C., Eliáš, M., & Lane, C. E. (2024). Multiple plastid losses within photosynthetic stramenopiles revealed by comprehensive phylogenomics. *BioRxiv*. doi: 10.1101/2024.02.03.578753
- Tirichine, L., & Piganeau, G. (2023). Editorial: algal symbiotic relationships in freshwater and marine environments. *Front Plant Sci*, 14, 1155759.
- Unrein, F., Gasol, J. M., Not, F., Forn, I., & Massana, R. (2014). Mixotrophic haptophytes are key bacterial grazers in oligotrophic coastal waters. *ISME J*, 8(1), 164–176.
- van Beveren, F., Eme, L., López-García, P., Ciobanu, M., & Moreira, D. (2022). Independent size expansions and intron proliferation in red algal plastid and mitochondrial genomes. *Genom Biol Evol*, 14(4), 037.
- Vancaester, E., Depuydt, T., Osuna-Cruz, C. M., & Vandepoele, K. (2020). Comprehensive and functional analysis of horizontal gene transfer events in diatoms. *Mol Biol Evol*, 37(11), 3243–3257.
- Vance, T. D. R., Bayer-Giraldi, M., Davies, P. L., & Mangiagalli, M. (2019). Ice-binding proteins and the “domain of unknown function” 3494 family. *FEBS J*, 286(5), 855–873.
- Vanclová, A. M. G., Hadariová, L., Hrdá, Š., & Hampl, V. (2017). Secondary plastids of euglenophytes. *Adv Bot Res* 84, 321–358
- Vincent, F., Sheyn, U., Porat, Z., Schatz, D., & Vardi, A. (2021). Visualizing active viral infection reveals diverse cell fates in synchronized algal bloom demise. *Proc Natl Acad Sci USA*, 118(11).
- Walker, G., Dorrell, R. G., Schlacht, A., & Dacks, J. B. (2011). Eukaryotic systematics: a user’s guide for cell biologists and parasitologists. *Parasitol*, 138(13), 1638–1663.
- Wang, L., Patena, W., Van Baalen, K. A., Xie, Y., Singer, E. R., Gavrilenko, S., Warren-Williams, M., Han, L., Harrigan, H. R., Hartz, L. D., Chen, V., Ton, V. T. N. P., Kyin, S., Shwe, H. H., Cahn, M. H., Wilson, A. T., Onishi, M., Hu, J., Schnell, D. J., ... Jonikas, M. C. (2023). A chloroplast protein atlas reveals punctate structures and spatial organization of biosynthetic pathways. *Cell*, 186(16), 3499–3518.e14.
- Wang, Q., Sun, H., & Huang, J. (2017). Re-analyses of “algal” genes suggest a complex evolutionary history of oomycetes. *Front Plant Sci*, 8, 1540.
- Wang, Y., & Morse, D. (2006). Rampant polyuridylation of plastid gene transcripts in the

- dinoflagellate *Lingulodinium*. *Nucl Acids Res*, 34(2), 613–619.
- Wells, M. L., Trainer, V. L., Smayda, T. J., Karlson, B. S. O., Trick, C. G., Kudela, R. M., Ishikawa, A., Bernard, S., Wulff, A., Anderson, D. M., & Cochlan, W. P. (2015). Harmful algal blooms and climate change: learning from the past and present to forecast the future. *Harmful Algae*, 49, 68–93.
- Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M. S., Burleigh, J. G., Gitzendanner, M. A., Ruhfel, B. R., Wafula, E., Der, J. P., Graham, S. W., Mathews, S., Melkonian, M., Soltis, D. E., Soltis, P. S., Miles, N. W., ... Leebens-Mack, J. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc Natl Acad Sci USA*, 111(45), E4859–E4868.
- Woehle, C., Dagan, T., Martin, W. F., & Gould, S. B. (2011). Red and problematic green phylogenetic signals among thousands of nuclear genes from the photosynthetic and apicomplexa-related *Chromera velia*. *Genom Biol Evol*, 3, 1220–1230.
- Xiong, W., Song, Y., Yang, K., Gu, Y., Wei, Z., Kowalchuk, G. A., Xu, Y., Jousset, A., Shen, Q., & Geisen, S. (2020). Rhizosphere protists are key determinants of plant health. *Microbiome*, 8(1), 27.
- Yamada, N., Sym, S. D., & Horiguchi, T. (2017). Identification of highly divergent diatom-derived chloroplasts in dinoflagellates, including a description of *Durinskia kwazulunatalensis* sp. nov. (Peridinales, Dinophyceae). *Mol Biol Evol*, 34(6), 1335–1351.
- Yan, N. (2013). Structural advances for the major facilitator superfamily (MFS) transporters. *Trends Biochem Sci*, 38(3), 151–159.
- Yazaki, E., Yabuki, A., Imaizumi, A., Kume, K., Hashimoto, T., & Inagaki, Y. (2022). The closest lineage of Archaeplastida is revealed by phylogenomics analyses that include *Microheliella maris*. *Open Biol*, 12(4), 210376.
- Yokoyama, A., Takahashi, F., Kataoka, H., Hara, Y., & Nozaki, H. (2011). Evolutionary analyses of the the nuclear-encoded photosynthetic gene *psbO* from tertiary plastid-containing algae in Dinophyta. *J Phycol*, 47(2), 407–414.
- Yoon, H. S., Hackett, J. D., & Bhattacharya, D. (2002). A single origin of the peridinin- and fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis. *Proc Natl Acad Sci USA*, 99(18), 11724–11729.
- Yu, G., Nakajima, K., Gruber, A., Rio Bartulos, C., Schober, A. F., Lepetit, B., Yohannes, E., Matsuda, Y., & Kroth, P. G. (2022). Mitochondrial phosphoenolpyruvate carboxylase contributes to carbon fixation in the diatom *Phaeodactylum tricornutum* at low inorganic carbon concentrations. *New Phytol*, 235(4), 1379–1393.
- Záhonová, K., Füßy, Z., Birčák, E., Novák Vanclová, A. M. G., Klimeš, V., Vesteg, M., Krajčovič, J., Oborník, M., & Eliáš, M. (2018). Peculiar features of the plastids of the colourless alga *Euglena longa* and photosynthetic euglenophytes unveiled by transcriptome analyses. *Sci Rep*, 8(1), 17012.
- Zauner, S., Greilinger, D., Laatsch, T., Kowallik, K. V., & Maier, U.-G. (2004). Substitutional editing of transcripts from genes of cyanobacterial origin in the dinoflagellate *Ceratium horridum*. *FEBS Letters*, 577(3), 535–538.
- Zhang, Q.-C., Song, J.-J., Yu, R.-C., Yan, T., Wang, Y.-F., Kong, F.-Z., & Zhou, M.-J. (2013). Roles of mixotrophy in blooms of different dinoflagellates: Implications from the growth experiment. *Harmful Algae*, 30, 10–26.
- Zhang, Z., Green, B. R., & Cavalier-Smith, T. (1999). Single gene circles in dinoflagellate chloroplast genomes. *Nature*, 400(6740), 155–159.
- Zheng, W., Zhang, C., Li, Y., Pearce, R., Bell, E. W., & Zhang, Y. (2021). Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations. *Cell Rep Meth*, 1(3), 100014.

Towards a genome-scale model of algal resilience to climate change

Research Project
Habilitation à diriger des recherches
de l'Université Paris-Saclay

présentée et soutenue à Gif-sur-Yvette, le 7 Octobre
2024, par

Richard DORRELL

Composition du jury

Hervé PHILIPPE Directeur de Recherche (CNRS), SETE de Moulis/ Université de Toulouse	Rapporteur
Claire GACHON Professeure (MNHN), MCAM / Musée Nationale d'Histoire Naturelle	Rapportrice
Benjamin FIELD Directeur de Recherche (CNRS), BIAM / Université Aix-Marseille	Rapporteur
Catherine JACKSON Directrice de Recherche (CNRS), IJM / Université Paris-Cité	Examinatrice
Purificación LOPEZ-GARCIA Directrice de Recherche (CNRS), ESE / Université Paris-Saclay	Examinatrice
Olivier VALLON Directeur de Recherche (CNRS), LBCPLM / Institut de Biologie Physico-Chimique	Examineur

Titre : Vers un modèle à l'échelle du génome de la résilience des algues au changement climatique

Mots clés : Niche thermique ; métagénomes ; pan-génomique ; modélisation environnementale des caractères; mésocosmes ; évolution expérimentale

Résumé : L'activité anthropique fait évoluer rapidement et de manière irréversible l'écosystème global de la Terre vers un nouvel état, avec des conséquences biologiques encore méconnues pour les organismes qui l'habitent. Les raisons génomiques qui sous-tendent les températures de croissance des algues marines, et donc leur viabilité future, restent mal comprises. Cela est particulièrement vrai dans le contexte de conditions environnementales qui deviendront de plus en plus courantes dans le futur océan mondial, par exemple les impacts combinés du réchauffement et du rafraîchissement dans les écosystèmes polaires, ou des canicules prolongées dans les océans subtropicaux.

Au cours des cinq prochaines années, je propose de réorienter mon groupe afin d'identifier les caractéristiques des gènes qui permettent la thermotolérance des algues, en utilisant trois systèmes exemplaires : les algues de l' Océan Arctique, les haptophytes globalement répartis et l'évolution répétée de la cryophilie chez les Chlamydomonadales d'eau douce et d'eau marine. Cela impliquera des approches génomiques, environnementales, physiologiques et biochimiques combinées, établies dans mes études précédentes. L'objectif à long terme sera d'utiliser ces données préliminaires pour entraîner des modèles phylogénétiques pour l'identification des déterminants globaux de la résilience des algues face à un océan qui se réchauffe.

Title : Towards a genome-scale model of algal resilience to climate change

Keywords : Thermal niche; Meta-genomes; Pan-genomics; Environmentally-informed trait modelling; Mesocosms; Experimental evolution

Abstract : Anthropogenic activity is rapidly and irreversibly moving to the Global Earth ecosystem to a new state, with as yet unknown biological consequences to the organisms that inhabit it. The genomic reasons underpinning the growth temperatures of marine algae, and thus their future viability, remain poorly understood. This is particularly true in the context of environmental conditions that are rarely found today, but will become increasingly common in the heating global ocean, e.g. the combined impacts of heating and freshening in polar ecosystems, or of sustained heatwaves in subtropical oceans.

In the next five years, I am to reorient my group to identify gene features that enable algal thermotolerance, starting with three exemplar systems: distantly related algae native to the Arctic; globally distributed non-calcifying haptophytes; and the repeated evolution of cryophily across marine and freshwater Chlamydomonadales. This will involve combined genomic, environmental, physiological and biochemical approaches benchmarked in my previous studies. The long-term objective will be to scale these systems up with regard to phylogeny, to identify global determinants of algal resilience to a heating ocean.

1. Context of the research project

As a result of human activity, the Earth ecosystem is moving irreversibly to a new state, termed the Anthropocene (Beau & Larrère, 2018; Ruddiman, 2013). Release of greenhouse gases has caused global heating that is accelerating rapidly as feedback loops and tipping points are initiated (Intergovernmental Panel on Climate Change, 2007; Lenton, 2013). Further anthropogenic impacts include an enrichment in atmospheric CO₂ from pre-industrial levels of <280 to a current (2023) >420 parts per million (NOAA Global Monitoring Laboratory, 2024); and organic and inorganic pollution (e.g., nitrates, phosphates, particulates, plastics) across all planetary ecosystems (Suikkanen et al., 2013; Yang et al., 2014). A wider biospheric crisis (Díaz et al., 2019) relating to the erosion of wild habitats and biodiversity has top-down impacts on the functions and resilience of all planetary ecosystems (Murphy et al., 2020). As of 2023, six of the nine Stockholm Planetary Boundaries that define a safe operating space for humanity have been surpassed (Richardson et al., 2023; Rockström et al., 2009), with global but unequal impacts (Differbaugh & Burke, 2019).

As a biologist working on fundamental questions of algal evolution (Dorrell et al., 2019, 2021; Novák Vanclová et al., 2024); organelle metabolism (Ait-Mohamed et al., 2020; Dorrell et al., 2017; Nonoyama et al., 2019) and eco-physiology (Dorrell et al., 2023; Dorrell, Zhang, et al., 2024; Pierella Karlusich et al., 2023), the impacts of the climate crisis on algae interest me in both environmental and fundamental biology terms. We are creating a range of new aquatic ecosystems that have not existed in recent geological history. Beyond direct impacts of temperature on species viability (defined as the “thermal niche”) (Krinos et al., 2024), heating increases thermal stratification, which in turn reduces the available nutrients in the water column (Agusti et al., 2019), and can even change the marine light field due to deepening of the euphotic zone (Seifert et al., 2023). Changes to marine cycling, e.g. due to melting of polar ice caps (Ardyna & Arrigo, 2020; Duprat et al., 2016), can further alter the abundance and distribution of key nutrients in the ocean. The scale of heating will be particularly acute in certain habitats, e.g. the Arctic Ocean, and coastal ecosystems (Glibert, 2020; Griffith & Gobler, 2020; Hutchins et al., 2009).

Global systems models of future algal distributions, e.g. from *Tara* Oceans data, project a latitudinal flux of species from equatorial towards polar habitats (Chivers et al., 2017; Frémont et al., 2022). This is counterbalanced by biodiversity loss, with polar ecosystems and diatoms particularly threatened by extinction (Chaffron et al., 2021; Henson et al., 2021; Trubovitz et al., 2020). Nonetheless, little is known about what algae will proliferate in new environments not found in the current ocean. These include ecosystems that may heat beyond current ocean maxima, or faster than the migratory or adaptive capacities of the algae that inhabit them (Sauterey et al., 2023; Thomas et al., 2012; Trubovitz et al., 2020). Determining which algae may be most viable in these new conditions, and why, may give us insights into what biological features determine their ecological niche; and into the future composition and photosynthetic activity of the global ocean.

2. Research project objectives

In the next five years, I aim to move my research beyond questions of chloroplast evolution towards **creating a generalised genomic model for temperature preference across the algal tree of life**. This will involve the integration of phylogenomic, environmental, physiological and biochemical approaches to **identify genes whose presence, expression, and function can be used to reproducibly define the thermal niches** of distantly related algal species (Fig. 1). I am particularly interested to use culture-based approaches to explore **environmental conditions that are not yet observed in the wild**, e.g. the impacts of sustained heatwaves in tropical ecosystems,

or the combined impacts of mild heating and freshening (ice-melt) in polar ones. I would like to design projects that **close knowledge gaps between culture-based and meta-genomic studies** which can provide complementary insights respectively to the biological potential of algae in the lab, and their real-world functions in natural ecosystems. I can envision four major objectives:

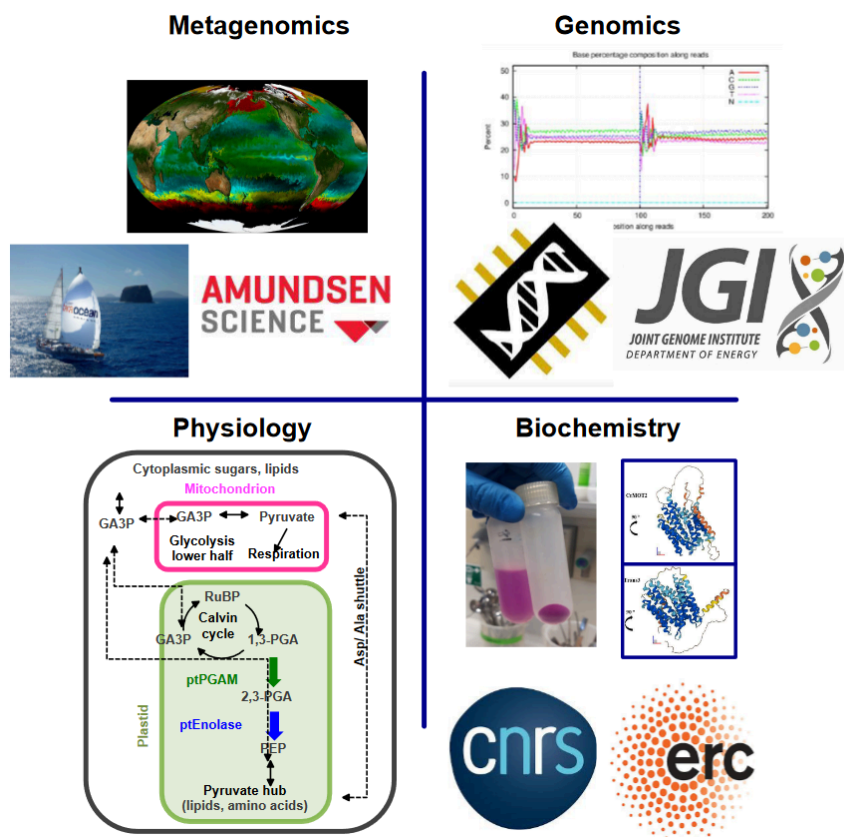


Fig. 1. Graphical abstract of different techniques and collaborating partners for the proposed work-programme.

1) Laboratory versus field estimates often produce different (5-10°C) estimates of microbial growth temperature optima (Brandenburg et al., 2019; Litchman & Thomas, 2022). This may be due to multiple factors, including interactive effects of nutrients and trophic constraints (Kendrick et al., 2014; Murphy et al., 2020); or adaptation of wild isolates to laboratory temperatures in continuous culture (Lehmann et al., 2023; Wells et al., 2015). A

first objective will be to develop an associative model between temperature dependence of laboratory growth rates and distributions of wild relatives from meta-genomic data.

- 2) A second objective will be to identify gene families whose presence, absence, expansion, or contraction impact on algal temperature preference. While in my previous research I have demonstrated e.g., that ice-binding proteins are restricted to polar algae; and plastidial glycolysis is more frequently associated with high latitudes (Dorrell et al., 2023; Dorrell, Zhang, et al., 2024), these are drawn from macro-scale datasets that may not detect genomic differences over strain, or population distances. Working at the level of pan-genomes, I will characterise temperature-associated genome differences between closely related strains of two algal groups (haptophytes, and Chlamydomonadales) that may be subsequently extrapolated to global phenomena.
- 3) To date, little is known about *how* genetic differences between closely related algal strains confer differences in thermal niche (Bendif et al., 2023; Krinos et al., 2024). A third objective will be to identify physiological reasons for different temperature optima in closely-related algae. This will involve integrative 'omics (growth, physiology, gene expression, metabolism) of cultured algae to identify which genes, and which biological processes, function at different temperatures; and correlating this to meta-gene expression to determine if the same trends are found in the wild. I am particularly interested to explore how algal gene expression and physiology change in response to thermal stress. For this, I am interested in two systems: the diverse responses of distantly related Arctic algae to

combined heating and freshening (Ardyna & Arrigo, 2020); and simulated impacts of heatwaves on thermotolerant marine haptophytes (Soulié et al., 2023).

- 4) Finally, I would like my research programme to maintain a basis in experimental cell biology, in particular characterising the biochemical functions of novel algal proteins implicated in thermotolerance. I can see three potential avenues for future research: protein localisation and reaction kinetics via heterologous expression in transformable models, e.g. *Chlamydomonas reinhardtii*; development of transformation strategies in new cryophilic and thermotolerant algae; or experimental evolution and sequencing of naïve field isolates under thermal stress conditions. I am particularly interested, given their implicated roles in thermal and salinity stress tolerance, to explore ice-binding proteins and plastid carbon metabolism as first model systems (Freyria et al., 2022; Treves et al., 2022).

I have recently begun a group leader position at the Laboratory of Computational and Quantitative Biology (LCQB, UMR7238), which is well positioned to support these objectives. With both experimental and computational facilities, research at the LCQB includes complementary interests in genome assembly (Gilles Fischer; Zhou Xu); domain annotation (Alessandra Carbone, Eric Bapteste, Philippe Lopez); phylogenetics (Elodie Laine, Martin Weigt); *Chlamydomonas* biology (Stéphane Lémaire, Julien Henri); microbial cell biology (Bianca Sclavi, Frédéric Devaux); and imaging (Judith Miné-Hattab). The LCQB is positioned within the Institut de Biologie Paris-Seine (IBPS, Sorbonne Université), with potential collaborations in neighbouring laboratories concerning plant physiology (Clara Richet-Bourbousse and Fredy Barneche, LBD); biodiversity (ISYEV, MCAM) and ecology and global systems biology (iEES, L'OCEAN). While I do not mention these explicitly here, I expect internal collaborations to form organically, with my group providing fundamental insights into algal taxonomy, systematics, and meta-genomics.

3. Proposed systems for research

While my previous research has focussed on the model diatom *Phaeodactylum* (Dorrell et al., 2017; Dorrell, Zhang, et al., 2024; S. Liu et al., 2022; Nonoyama et al., 2019), I am keen to explore alternative systems in my future research. We now know based on recent (e.g., meta-genomic) barcoding that flagellated algae are far more abundant in the oceans, compared to diatoms, than previously suspected (Pierella Karlusich et al., 2022, 2023). Flagellated algae are also functionally diverse, including multiple examples of mixotrophic and secondarily heterotrophic groups (e.g., in chrysophytes and dinoflagellates) alongside purely phototrophic members (Dorrell et al., 2019; Novák Vanclová et al., 2024). My interest is not only to project how these different algal groups are likely to change in abundance in the future ocean, but also to what extent their behavioural plasticity confers resilience in changing environments.

I plan to focus on two groups of algae already under study in my group: the heterogeneous but typically cryophilic algae indigenous to the Arctic Ocean (**Fig. 2**) (Dorrell et al., 2023); and globally distributed but taxonomically closely related members of the haptophytes, an ecologically prominent algal order with projected resilience to climate change (**Fig. 4**) (Penot et al., 2022). A final project, which will be established in relation to internal expertise at the LCQB, will focus on novel isolates of the model green algal order Chlamydomonadales, in particular from cryogenic (Arctic, Antarctic and montane) habitats (**Fig. 5**) (Cvetkovska et al., 2017; Ezzedine et al., 2023). Specific members of these algae are already under culture in my group (**Fig. 4**), and we have already in many cases completed genomes, transcriptomes, and established growth kinetics.

3.1. Heterogeneous responses to heating and freshening within Arctic Ocean algae

As a first system I am interested in exploring the diverse physiologies in Arctic algae, as established in my prior research (Dorrell et al., 2023). The photosynthetic activity of cryogenic (> 30°N, >50°S) marine habitats constitute up to a third of the global carbon budget (Hüner et al., 2023). Amongst these the Arctic is not only the most rapidly changing but likely the most sensitive in terms of its biota to climate change (Ardyna & Arrigo, 2020; Chaffron et al., 2021).

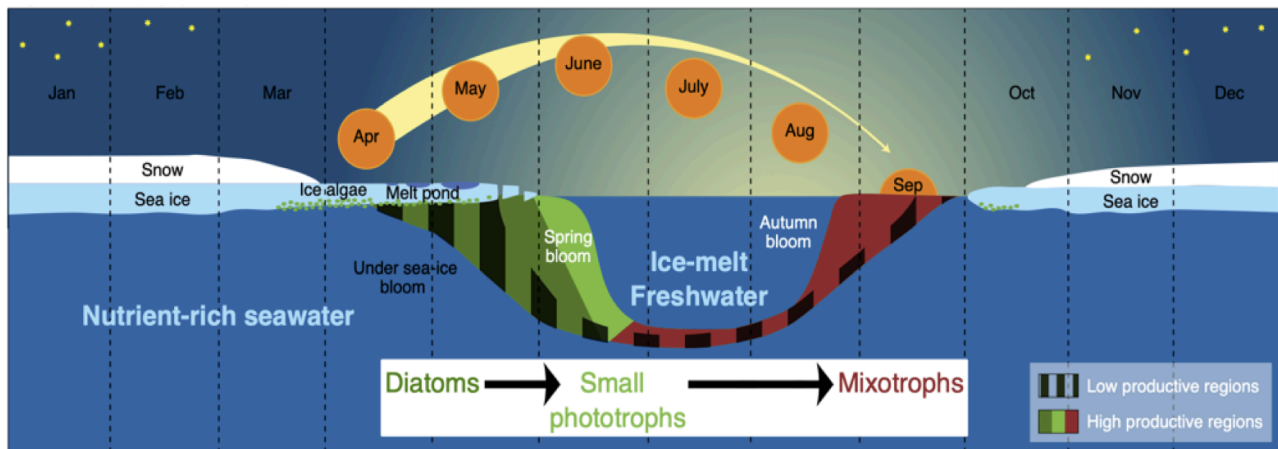


Fig. 2. Seasonal phenology of the Arctic Ocean. This figure was constructed by a postdoc in my group, Louis Graf (Graf et al., 2021), adapted from (Ardyna & Arrigo, 2020).

The seasonal phenology of the Arctic Ocean is defined by an initial spring pulse of diatoms, which proliferate in nutrient-rich and well-mixed water underneath floating sea ice (Freyria et al., 2021; Joli et al., 2024), into early summer. Over the course of the summer, alluvial and meltwater deposition depresses the nutrient-rich allocline, producing increasingly oligotrophic conditions that favour smaller flagellates (haptophytes, pelagophytes, chrysophytes) with lower nutrient requirements (W. K. W. Li et al., 2009). In the late summer, mixotrophs (e.g. dinoflagellates) predominate, with autumn production curtailed both by seasonal freeze-up and the onset of the polar night (Joli et al. 2024; Freyria, Joli, and Lovejoy 2021). These mixotrophic algae may further proliferate as heterotrophs in low-light conditions during the polar winter (Kameyama et al., 2020; Randelhoff et al., 2020; Søgaard et al., 2021).

Climate change is restructuring the composition of the Arctic Ocean, with complex interactions of heat, salinity and seasonality. As a result of heating, ice-melt and increased continental precipitation is creating a fresher, nutrient-scarce Arctic (W. K. W. Li et al., 2009). Retarded formation of floating autumn sea-ice may allow increased secondary autumn blooms, in particular favouring non-ice specialist algae such as flagellates over diatoms (Ardyna et al., 2014; Freyria et al., 2021). These community changes may further have complex impacts on the primary production of Arctic algae. For example, algal photosynthesis is viewed to be limited both by low temperatures (< -5°C) and high salinities (> 70 ‰) associated with the brine pockets of sea-ice (Kameyama et al., 2020), and heating and ice-loss is likely to be associated both with greater primary production *and* biodiversity loss. We thus need to know not only who will proliferate in a future Arctic Ocean, but why; and how the physiology of these species will change.

My group maintains a culture collection of approximately twenty Arctic species (**Fig. 3**), including previously isolated strains obtained from collaborators and culture collections (Connie Lovejoy, Roscoff Culture Collection) (Dorrell et al., 2023; Gérikas Ribeiro et al., 2020), and new ones isolated by from Arctic seawater as part of the Amundsen Darkedge mission (Ardyna et al., 2014; Denu et al., 2023). We are generating genomes and transcriptomes for these strains in-lab. We have also measured log phase growth velocities of these strains (**Fig. 4**) across seven temperatures (0°C, 4°C, 12°C, 19°C, 24°C, 28°C, 32°C) and four salinities, two of which

approximate to measured seawater salinities in *Tara* Oceans (35‰, mean global value; 14‰, minimal measured salinity of Arctic seawater); and two (7‰, 3.5‰) to hyposaline stress conditions (Ardyna et al., 2017). Finally, using phylogenetically reconciled 18S V4 ribotypes, we have mapped *Tara* Oceans distributions of each strain, following my previous work (Dorrell et al., 2023).

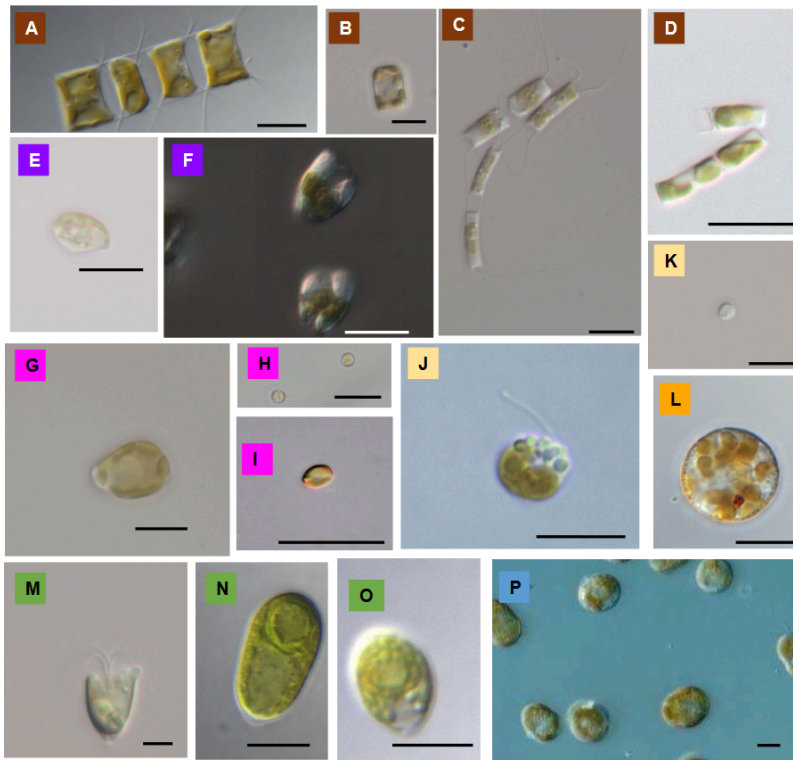


Fig. 3. Examples of Arctic species cultured in my group. Photos are from relatives deposited in the Roscoff Culture Collection (Gérikas Ribeiro et al., 2020) or NCMA (where strain named), and are shaded by taxonomy per Fig. 4. Scale bars: 10 μ m. **A-D:** diatoms; *Chaetoceros neogracile*, *Thalassiosira nordenskioldii*, *Attheya longicornis*, Naviculales sp. **E-F:** cryptomonads; *Baffinella frigidus*, *Rhodomonas* sp. **G-I:** haptophytes; *Pseudohaptolina sorokinii*, *Imantonia rotunda*, *Pavlova* (type CCMP2436). **J-K:** chrysophytes; *Ochromonas* (type CCMP1899), non-photosynthetic sp (e.g., *Paraphysomonas*, *Spumella*). **L:** pelagophyte, *Pelagococcus* sp. **M-O:** chlorophytes, *Pyramimonas discoicola*, *Chlamydomonas prisculii*, *Chlamydomonas* sp. ICE. **P:** dinoflagellate, *Polarella glacialis* (type CCMP2088).

Broadly, these cultures can be divided into three groups based on growth responses. Diatoms

(*Chaetoceros*, *Thalassiosira*, *Attheya*, Naviculales sp.) are tolerant of mild heating, with measurable growth at 12°C or even 19°C, but not of hyposalinity, with lower growth rate at salinities below 14‰. Surprisingly, *Tara* Oceans distributions of these species suggest (with the exception of *Thalassiosira*) endemism to the Arctic Ocean, raising the question of what factors constrain their abundance in the wild at higher temperatures (Fig. 4). Flagellates in the collection divide into two groups: Arctic endemic species, which are not viable at temperatures >4°C, but include species tolerant of extreme freshening (*Ochromonas* sp. RCC7855, *Pavlova* sp. CCMP2436; viable growth measured at 3.5‰ salinity); and species with more cosmopolitan distributions (*Pavlova* sp. CCMP3319; *Imantonia* sp.), which may be viable in both low salinities and high temperatures but not at temperatures below 4°C (Fig. 4) (Ibarbalz et al., 2023).

We are particularly interested in the chrysophytes, which appear to relate to derive from a deeply positioned marine clade (“*Ochromonas* clade II”) showing predominant presence in the Arctic (CCMP2296, RCC7855), and have known Antarctic relatives (CCMP1899). This clade further shows evolutionary proximity to the freshwater snow algal genus *Hydrurus*. These strains may thus link chrysophyte taxonomic diversity in marine and freshwater cryophilic habitats (Andersen et al., 2017; Delmont et al., 2022; Remias et al., 2013). In any case, future research questions will revolve around what gene features confer cryophily, thermotolerance, and tolerance of freshening stress in different Arctic algae- and how these adaptations may structure a future Arctic Ocean impacted by heating and salinity stress (Asch et al., 2019; Boras et al., 2010).

	Genus	Strain	Lineage	Habitat	35 ppt salinity, temperature (C):							4C, salinity (ppt):				
					0	4	12	19	24	28	32	14	7	3.5		
Arctic species	<i>Chaetoceros</i>	RCC7850	Diatom	Arctic												
	<i>Attheya</i>	RCC7856	Diatom	Arctic												
	<i>Thalassiosira</i>	RCC7842	Diatom	Arctic												
	<i>Thalassiosira</i>	RCC7852	Diatom	Arctic												
	Naviculales	RCC7840	Diatom	Arctic												
	<i>Ochromonas</i>	RCC7855	Chrysophyte	Arctic												
	<i>Ochromonas</i>	CCMP2296	Chrysophyte	Arctic												
	<i>Rhodomonas</i>	RCC2513	Cryptomonad	Arctic											NA	
	<i>Baffinella</i>	CCMP2293	Cryptomonad	Arctic												
	<i>Baffinella</i>	CCMP2045	Cryptomonad	Arctic												
	<i>Pedinella</i>	CCMP2098	Dictyochophyte	Arctic											NA	
	<i>Polarella</i>	CCMP2088	Dinoflagellate	Arctic											NA	
<i>Pyramimonas</i>	RCC7841	Prasinophyte	Arctic													
Pelagophyte	CCMP2097	Pelagophyte	Arctic											NA		
Pavlova	<i>Pavlova</i>	CCMP2436	Haptophyte	Arctic												
	<i>Pavlova</i>	CCMP3319	Haptophyte	Arctic												
	<i>Pavlova</i>	RCC1539	Haptophyte	Non-Arctic											NA	
	<i>Pavlova</i>	RCC2550	Haptophyte	Non-Arctic											NA	
	<i>Pavlova</i>	RCC5622	Haptophyte	Non-Arctic											NA	
	<i>Pavlova</i>	RCC5945	Haptophyte	Non-Arctic											NA	
Imantonia/ Pseudohaptolina	<i>Imantonia</i>	RCC4477	Haptophyte	Non-Arctic												
	<i>Imantonia</i>	RCC704	Haptophyte	Non-Arctic												
	<i>Imantonia</i>	RCC2298	Haptophyte	Arctic												
	<i>Imantonia</i>	RCC4501	Haptophyte	Arctic												
	<i>Imantonia</i>	RCC4780	Haptophyte	Arctic												
	<i>Imantonia</i>	RCC4500	Haptophyte	Arctic												
	<i>Pseudohaptolina</i>	RCC5270	Haptophyte	Arctic												
<i>Pseudohaptolina</i>	RCC5268	Haptophyte	Arctic													

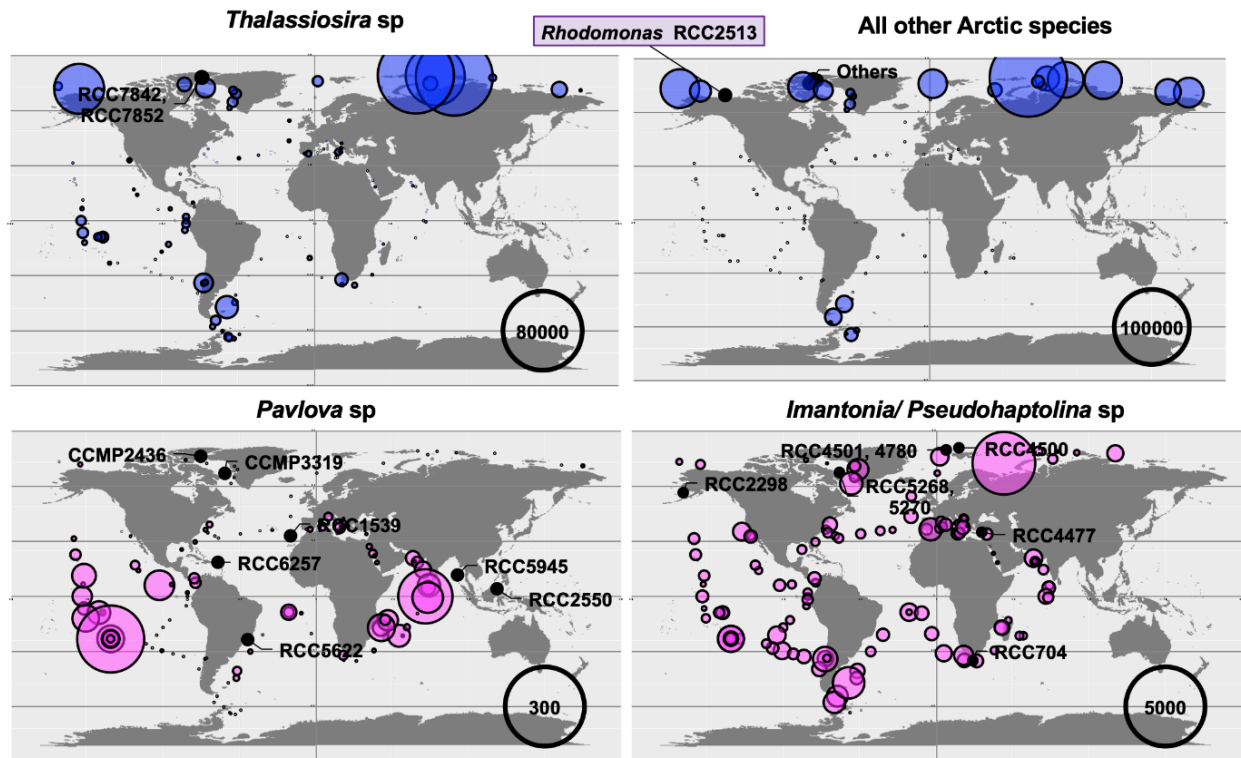


Fig. 4. Complex and conflicting distributions of Arctic algae and globally distributed haptophytes from experimental and field data. This figure shows 30 algal strains cultured in my group, with emphasis on distantly-related Arctic algae and broadly distributed representatives of the haptophytes *Pavlova*, *Imantonia* and *Pseudohaptolina*. These include strains from the NCMA, RCC and CCCM culture collections (Guillard, 1988; Vaultot et al., 2004), alongside new strains (RCC7840-RCC7856) isolated by serial dilution from seawater during the 2021 Amundsen Darkedge expedition.

Top: heatmap of exponential phase growth rates of each strain in enhanced seawater with k/2 amendment, under 50 $\mu\text{E m}^{-2} \text{ s}^{-1}$ cool white LED light, measured from 5×10^4 to 2×10^6 cells ml^{-1} with a ParTec CyFlow cell counter (Dorrell, Zhang, et al., 2024). Growth rates (mean divisions per day) were measured over seven constant temperatures (from 0 °C to 32 °C) under 35 ‰ salinity, and (for twelve Arctic species) over three additional salinities (from 14 ‰ to 3.5 ‰) at 4 °C. Species are shaded by taxonomic affiliation and collection site (Arctic, or other). Double-lined boxes group strains (*Ochromonas*, *Baffinella*, *Pavlova*, *Imantonia*, *Pseudohaptolina*, *Thalassiosira*) that show more than 99% 18S rRNA sequence identity to one another, and can be considered species. **Bottom:** collection sites (black circles, labelled) and Tara Oceans 18S v4 ribotype mean relative abundances (coloured circles) for Arctic strains (grouped together, except for *Thalassiosira* and haptophytes shown separately) and haptophyte strains (grouped by genus). V4 relative abundances were calculated from ribotypes reconciled to cultured strain 18S rDNA sequences of each strain by single-gene phylogeny, following (Dorrell et al., 2023), and are expressed as the average number of ribotypes per million total across all depths and size fractions from a given station. These data were produced with the aid of the following undergraduate students: Alice Bossard, Quentin Caris, Pierre Der-Gazerian, Ilona Dupuis, Victoria Powell, Ema Siomeo; and supervised research engineer: Marc Przcyszczkowski.

3.2. Variable thermal niche optima across closely related haptophytes

A second system of interest are haptophytes, a globally distributed and abundant algal order. Haptophytes are typically small, planktonic algae, characterized by a “haptonema,” or third flagellar protuberance, between two equally-lengthed flagella (Tsuji & Yoshida, 2017). Some haptophytes, including the well-known species *Emiliana* (*Gephyrocapsa*) *huxleyi* are calcifiers, forming extracellular “coccoliths” that constitute the major component of calcareous rocks such as limestone (Bendif et al., 2023; Skeffington et al., 2023). Most haptophytes are marine or euryhaline and photosynthetic (phototrophic, mixotrophic), and no known truly multicellular species are known, although colonial (*Phaeocystis globosa*), freshwater (*Chrysochromulina parva*) and rare, potentially obligately heterotrophic (*Paaposphaera*, *Trigonaspis*) members have been described (Deodato et al., 2019; Mars Brisbin & Mitarai, 2019; Thomsen et al., 1991).

While initial Tara Oceans estimates, based on 18S rDNA V9 barcodes, projected that haptophytes constitute about 10% of marine algal abundance, this is likely to be a substantial underestimate (e.g., due to primer amplification bias) (de Vargas et al., 2015; Edvardsen et al., 2016; H. Liu et al., 2009). Subsequent estimates based on 18S V4 and 16S plastid V4V5 rDNA barcoding markers; and assembled sequences from meta-transcriptome data (e.g., *psbO*), suggests haptophytes are nearly as abundant as diatoms (Penot et al., 2022; Pierella Karlusich et al., 2023). In contrast to diatoms, which are predominantly found in nutrient-rich, well-mixed waters at high latitudes, pigment and barcoding estimates suggest haptophytes have more homogeneous global distributions, with species complexes extending from polar to tropical oceans (Bendif et al., 2023; Krinos et al., 2024; H. Liu et al., 2009). More detailed appraisals of plankton vertical distributions, and distributions across size, suggest haptophytes are particularly abundant in small size fractions (0.8-3/ 5 μm), and in stratified surface layers where nutrients are depleted and cells are exposed to high irradiance (Lacour et al., 2023; Pierella Karlusich et al., 2023; Tsuji & Yoshida, 2017).

The global distributions, and adaptations of haptophytes to non-turbulent, stratified and oligotrophic waters may render them particularly resilient to ocean heating (Winder & Sommer, 2012). Multiple time-series studies over the last twenty years have noted increases in haptophyte abundance, e.g. in the Pacific Northwest (Nemcek et al., 2023); the equatorial Atlantic (Xi et al., 2023); the Baltic (Suikkanen et al., 2013) and the Barents Sea (Smyth et al., 2004). Concordant results have also been identified in experimental mesocosms, with haptophytes increasing in response to transient heatwaves and CO₂ enrichment (Soulié et al., 2023; Tortell et al., 2002). Considering modelled responses, haptophytes are projected to have lower biodiversity losses than high-latitude specialists (e.g. diatoms) in the next century (Chaffron et al., 2021). They show either changes in spatial distribution, or even moderate increases in abundance (e.g., in models incorporating interactions between nutrients, temperature, and light) (Henson et al., 2021; Seifert et al., 2023).

These heating resilience associated generally with haptophytes are likely to be counterbalanced by factors specific to individual groups. Some orders, e.g. Phaeocystales and Chrysochromulinales, are associated with high latitudes and may be more thermosensitive than others with more cosmopolitan distributions (Brandenburg et al., 2019; Smith & Trimborn, 2024; Q. Song et al., 2023; Sun et al., 2022). Calcifying haptophytes within the orders Coccolithales and Isochrysidales, and freshwater species, are sensitive to acidification (Fox et al., 2020; Prosser et al., 2012). Many of the non-calcifying orders (Pavlovophyceae, Prymnesiales) that are likely to be more abundant in the future ocean are mixotrophic (Godrijan et al., 2020). It is an open question to what extent haptophytes will utilise photosynthesis versus respiration in heating ocean scenarios (Hallegraeff, 2010; Wilken et al., 2013), with consequent impacts on global carbon budgets.

I am co-coordinator of a JGI CSP project to sequence and annotate the haptophyte pan-genome (Penot et al., 2022). This project will generate high-quality (chromosome scale with deep annotation) reference genomes for 50 haptophyte strains, with balanced taxonomic sampling. We aim to answer questions about haptophyte evolution and plastid metabolism by sampling basal haptophytes and their relatives (e.g., raphidomonads, centrohelids, telonemids) (Burki et al., 2012; Kawachi et al., 2021; Yazaki et al., 2022); haptophyte biochemistry (calcifying and alkenone-producing species) (Araie et al., 2018; Bendif et al., 2023; Plancq et al., 2019; Skeffington et al., 2023); and trophic interactions (mixotrophic and toxic species) (Godrijan et al., 2020; John et al., 2022; H. Liu et al., 2009; Tábora-Sarmiento et al., 2022). Alongside this, I am establishing some more biogeographically-aware projects, exploring how genome variation defines thermal niche occupancy of two non-calcifying haptophyte groups, *Pavlova* and *Imantonia* (**Fig. 4**).

The class Pavlovophyceae, and its only order Pavloales, comprise five genera, with *Pavlova* paraphyletic to *Chrysoculter*, *Diacronema*, *Exanthemachrysis*, and *Rebecca* (Bendif et al., 2011). Pavlovophyceae are projected from molecular clock dates to have radiated as much as 250 million years ago, with the separation of most species assigned to *Pavlova* (in particular from *Diacronema*) about 100 million years before present (Bendif et al., 2011). Additional meta-barcode sequences positioned in sister-group positions to the Pavlovophyceae, e.g. clades HAP1-5 (Edwardsen et al., 2016; Endo et al., 2023), may further expand the diversity of this order. Most cultured Pavlovophyceae are coastal or estuarine isolates (Bendif et al., 2011). However, they are pervasively detected in open-ocean environmental sequencing datasets (Endo et al., 2023), and *Tara* Oceans meta-barcodes assigned to Pavlovophyceae show no obvious abundance correlations to coastal distance (Penot et al., 2022). In particular, environmental sequences relating to Pavlovophyceae clade 3, which includes the type species *P. pinguis* (Véron et al., 2023), are widespread in the sub-tropical open ocean with preference for waters > 25°C (**Fig. 4**).

Imantonia is a non-calcifying haptophyte, which typically forms small (< 5 µm diameter), rounded cells with haptonemas reduced to a vestigial proboscis (Green & Hori, 1986). *Imantonia* is distinguished from other haptophytes via a novel series of fatty C₁₈–C₂₂ fatty acids containing an unique ω₃-*trans* double bond (Lee Chang et al., 2016; Marlowe et al., 1984). Single-gene and ribosomal DNA trees place *Imantonia* in the order Prymnesiales (Bendif et al., 2013; Edwardsen et al., 2000), specifically as a sister-group to the non-calcifying genus *Pseudohaptolina*, which has contrasting morphology: cells that are larger (20–30 µm length), are elliptical or oblong in shape, and retain haptonema (Edwardsen et al., 2011). *Imantonia* was first isolated from temperate and subpolar waters (the English Channel to Svalbard) although appears to be globally distributed both from subsequent collections and from *Tara* Oceans data (Lee Chang et al., 2016; Mas et al., 2008; Reynolds, 1974; Sukhanova et al., 2015) (**Fig. 4**). Nonetheless, *Imantonia* is principally only associated with open-water samples in the Arctic Ocean, and is unlikely to be a true ice specialist (Marchant & Thomsen, 1994). Cultured examples of *Pseudohaptolina* come exclusively from the

Arctic Ocean, including from sea- and pack-ice samples (Gérikas Ribeiro et al., 2020; Sun et al., 2022), and we and others have reconstructed effective Arctic endemism of this genus from environmental sequence data (Edwardsen et al., 2011; Penot et al., 2022) (**Fig. 4**).

Previous studies of other haptophytes (*Gephyrocapsa*, *Tisochrysis*) have established substantial differences in cell morphology, photosynthesis, and temperature growth optima of closely related strains (Bendif et al., 2023; Krinos et al., 2024; Pajot et al., 2023). Understanding how haptophyte genomic diversity impacts on thermal niche may provide us with clues as to their resilience to ocean heating. The haptophytes within my culture collection allow us to ask complementary questions over different evolutionary distances (**Fig. 4**). Regarding Pavlovophyceae, two Arctic strains (CCMP2436, CCMP3319) can grow at temperatures down to 0°C, with contrasting upper thermal limits to viability (Dorrell et al., 2023). At the same time, subtropical species from two different groups (RCC2550 and RCC5622; and RCC5945 and RCC6257) show optimal growth at 28°C and are viable up to 32°C, which reflects the maximum temperature in the contemporary ocean (e.g., the Persian Gulf) (Evans, 1966; Noori et al., 2019). These contrasting thermal niches may allow us to explore the divergent evolution of cryophilic and thermophilic life strategies within individual orders. In contrast, the *Imantonia* strains, of which six (RCC704, RCC2298, RCC4477, RCC4500, RCC4501, RCC4780) show > 99% internal rDNA identity to one another, have different growth temperature ranges from 4 to 24°C, allowing us to explore progressive evolution of temperature optima over microevolutionary scales (**Fig. 4**).

3.3. Freshwater versus marine cryotolerance across Chlamydomonadales

The green algal order Chlamydomonadales encompasses a broad range of species, including the freshwater model species *C. reinhardtii* but also marine ones from a wide range of habitats (**Fig. 5**). These include multiple cryotolerant marine clades, with distributions encompassing Arctic and Southern Oceans (E. J. Kim et al., 2019). Of particular interest to my group are the *C. sp. ICE* and *C. priscuii* clades, which are independently evolved from one another, and have colonised both polar ecosystems (**Fig. 5A**). Alongside these polar marine clades, Chlamydomonadales are known from terrestrial cryophilic habitats such as snow and glacier ice. These include *Sanguinea* (formerly) *Chlamydomonas nivalis* (Remias et al., 2010); and *S. nivaloides*, which is the predominant component of “red snow” algal blooms (Ezzedine et al., 2023).

Different polar marine Chlamydomonadales show variable tolerances of both high and low temperatures. Some (e.g. *Chlamydomonas malina* sp. RCC2488, *Chlamydomonas priscuii*) can be clearly defined as cryophilic, with mortality at temperatures > 15°C (Cvetkovska et al., 2022; Jung et al., 2016; Schulze et al., 2019). Others are more correctly defined as having cryotolerant behaviour (Barcytè et al., 2018; Chae et al., 2023; H. Song et al., 2020), with some even showing faster growth at moderate than low temperatures (Schulze et al., 2019). The salinity tolerance of polar Chlamydomonadales also varies substantially, with growth optima for *C. priscuii* measured at between 15-20‰ (Morales-Sánchez et al., 2020), while *C. latifrons* even shows viability on TAP media plates (Ahn et al., 2015; Chae et al., 2023; E. J. Kim et al., 2016, 2019).

The physiological adaptations of cryophilic Chlamydomonadales are comparatively well understood. For example, *C. priscuii* appears to constitutively express heat-shock factors which may protect against even moderate heating (Cvetkovska et al., 2022), and performs cyclic electron flow around photosystem I to limit low-temperature photodamage (Kalra et al., 2020). *S. nivalis* constitutively overexpresses all of its photosystem machinery, allowing constitutively high photosynthetic activity even under chilling stress (Peng et al., 2020). These may however be counterbalanced by abnormal responses to other stresses, with *C. priscuii* and *C. malina* not

modulating their pigment composition or performing state transitions in relation to the light field (Hüner et al., 2023), which may render them more vulnerable to photoinhibition.

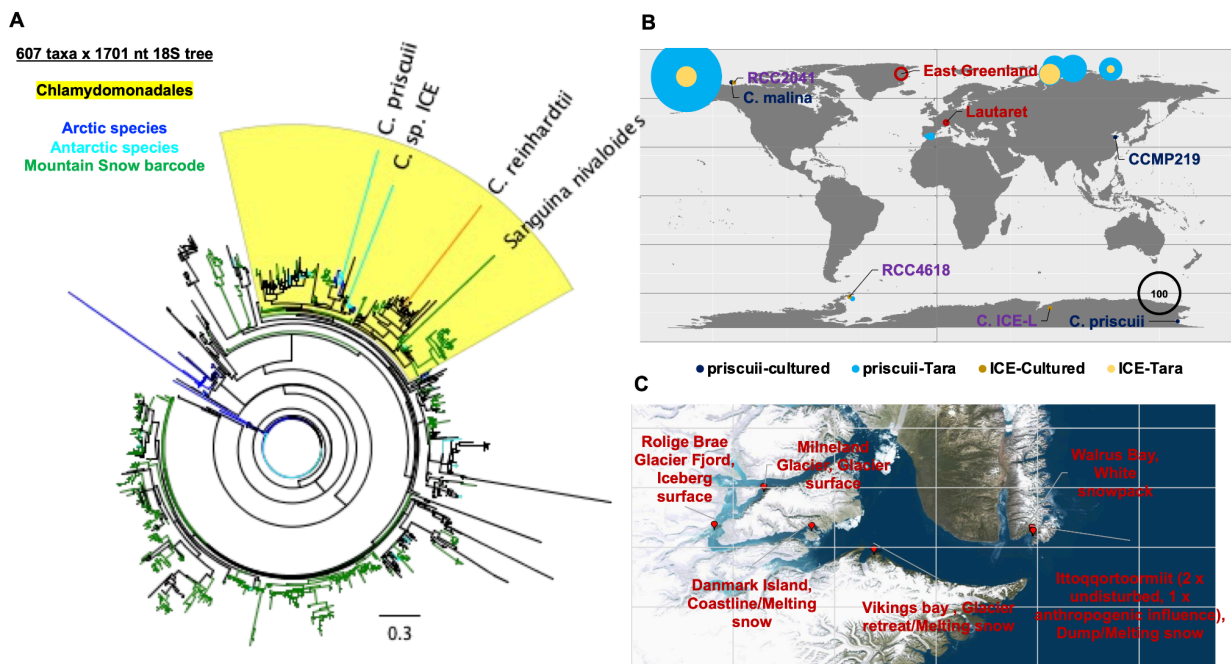


Fig. 5. Cryophily in Chlamydomonadales at taxonomic, global and regional scales. A: consensus RAxML phylogeny of cultured Chlorophyta 18S rDNA sequences, adapted from (Dorrell et al., 2023), and enriched with 18S V4 ribotypes sequenced in a transect of French Alpine mountain snow by (Stewart et al., 2021). Chlamydomonadales are shaded in yellow, and branches are coloured by collection site: Arctic (>60°N); Antarctic (<60°S); or mountain samples (V4 sequences only). Four families of interest (*C. reinhardtii*, *C. sp. ICE*, *C. priscuii* and *Sanguinea nivaloides*) are shown with artificially elongated branches (1.0 substitutions/ site). **B:** Tara Oceans mapped 18S V4 rDNA abundances (shown per Fig. 4) and collection sites of *C. sp. ICE* and *C. priscuii* strains maintained in my group and that of Marina Cvetkovska (University of Ottawa) (Cvetkovska et al., 2017). Both marine Chlamydomonadales show bipolar (Arctic and Southern Ocean) Tara distributions, and include representatives from Arctic and Southern Ocean marine and Antarctic continental masses. **C:** sites from a transect in East Greenland (highlighted in Panel B) performed by the Maréchal group (CEA Grenoble) that will be sampled for Chlamydomonadales diversity and abundance.

The variable physiology of these strains, particularly compared to the well-studied *C. reinhardtii*, provides an appealing system to identify genome features that enable cryophily over a smaller taxonomic scale than the pan-Arctic collection described above (Fig. 4). I am currently establishing two collaborations which I feel will be particularly appropriate for future student projects. One, with Marina Cvetkovska at the University of Ottawa, will be to sequence and understand the genome evolution of new Arctic (RCC2041, RCC2488) and Antarctic marine (RCC4618) strains of *C. priscuii* and *C. sp. ICE*, collected from continental Antarctic habitats (Cvetkovska et al., 2017, 2022). *C. priscuii* furthermore has mesophilic outgroups (*C. sp. CCMP219*, from China), allowing us to explore to what extent cryophilic adaptations are influenced by biogeography (Arctic versus Antarctic-) as well as taxonomy (Fig. 5B) (Morales-Sánchez et al., 2020). The second, with Eric Maréchal at the CEA Grenoble, will focus on convergent genomic adaptations of cryophilic freshwater versus marine Chlamydomonadales, using strains collected from his group in the French Alps (Ezzedine et al., 2023; Stewart et al., 2021) and from a novel transect in East Greenland that links snow, estuaries, sea-ice and polar oceans (Fig. 5C). I am particularly interested to characterise gene innovations that enable tolerance of mild salinities, as might be associated with a freshening Arctic, but might also be encountered in historical migrations of continental snow algae into the polar oceans (Dorrell et al., 2023; Jamy et al., 2022).

4. Work-plan and collaborations

All three projects (Arctic algae, globally distributed haptophytes, and cryophilic *Chlamydomonas*) will run in parallel, with complementary deliverables. For different projects we are varying sequencing partners to maximise the probability of being able to annotate genomes; and varying the physiological experiments in the context of relevant environmental questions. I therefore present a theoretical Gantt chart for one project only (**Fig. 6**), relating to Chlamydomonadales.

Year	2025				2026				2027				2028				
Project Month	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	
WP0 - Project Management (RD, EM)																	
[KO] Kickoff; [YM] yearly meeting; [IR] intermediate and [FR] final report	KO		YM						YM	IR			YM			YM	FR
WP1 - Environmental sequencing and life sample collection (EM)																	
Repeat sampling of Scoresby Sound/ East Greenland (D1.1)						1.1											
Amplification of meta-barcode data from concentrated cell mass (D1.2)				1.2 (2023, 2024 missions)				1.2 (2026 mission)									
Isolation of new species by serial dilution and picking (D1.3), barcoding and deposition in Culture Collection (D1.4), meta-genomic evaluation (D1.5) and publication	1.3 (2023, 2024 missions)		1.4	1.5				1.3 (2026 mission)	1.4	1.5			Publication 2- Culture Collection				
WP2 - Barcode analysis over glacier-to-ocean transects (RD)																	
Global evaluation of ice-binding protein diversity (D2.1) and development and validation of barcoding primers (D2.2)		2.1		2.2									Publication 1- IBPs				
Illumina sequencing of environmental samples from Scoreby Sound (D2.3); reception of visiting researcher (Elisabeth Richardson)						2.3 (2023, 2024 missions)			2.3 (2026 mission)		VR						
Evaluation of community composition over glacier ice to ocean transects (D2.4)												2.4					Publication 3- Environmental Diversity
Phylogenetic reconstruction of habitat transitions in Chlorophyta and Chrysochyta and publication (D2.5)												2.5					
WP3 - Comparative genomics and physiology of glacier ice and marine polar algae (RD, EM)																	
WP3.1 Genome assembly and annotation (RD, EM)																	
Final strain selection for genome sequencing (D3.1.1) , DNA and RNA isolation (3.1.2); exchange visit of PhD student to group of Marina Cvetkovska (University of Ottawa)			3.1.1	3.1.2		EV											
Short and long-read sequencing (D3.1.3), genome assembly (D3.1.4), transcriptome assembly (D3.1.5)							3.1.3	3.1.4	3.1.5								
WP3.2 Comparative genome evaluation (RD)																	
Gene model annotation from transcriptome data (D3.2.1)										3.2.1							
Identification of duplicated genes (D3.2.2), Horizontally acquired genes (D3.2.3), plastidial metabolism (D3.2.4)												3.2.2	3.2.3	3.2.4			Publication 4- Genomes and Physiology
Collaborative annotation projects (D3.2.5) and publication													3.2.5				
WP3.3 Physiological, transcriptomic and metabolic analysis (RD, EM)																	
Viability characterisation of selected strains over salinity transects (D3.3.1)				3.3.1													Publication 4- Genomes and Physiology
RNA sequencing under salinity stress conditions (D3.3.2) and DEG calculation (D3.3.3)								3.3.2		3.3.3							
Lipid profiling and carbon budget calculations (D3.3.4) and publication													3.3.4				
Salinity and temperature stress 'omics of <i>Chlamydomonas reinhardtii</i> (D3.3.5)		3.3.5															
WP4 – Functional characterisation of gene families implicated in polar- and glacier-specific adaptations in Chlamydomonadales (RD, EM)																	
WP4.1 Functional protein characterisation (RD)																	

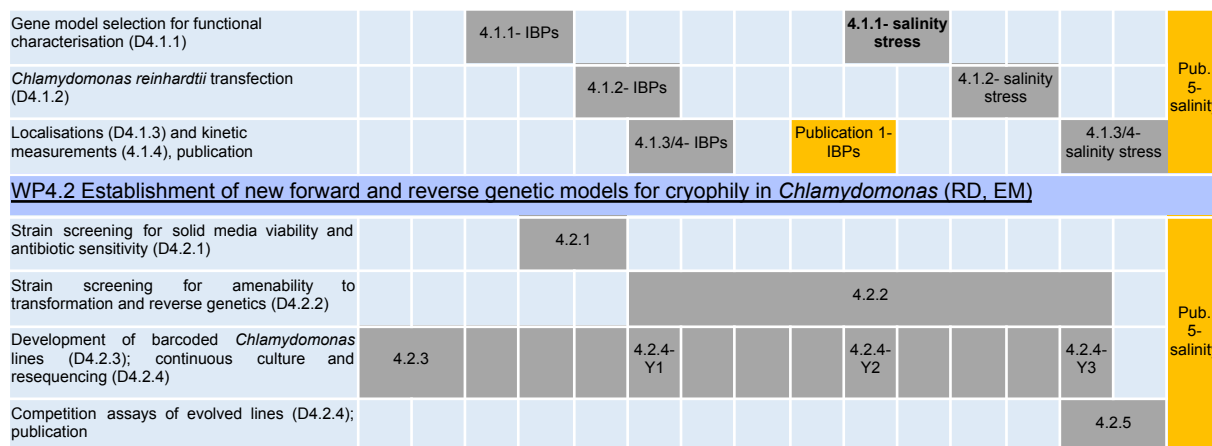


Fig. 6. Gantt Chart of project work-packages and deliverables for an exemplar project, on convergent adaptations in cryophilic marine and freshwater *Chlamydomonadales*. This project has been submitted as an ANR PRC between my group (RD) and that of Eric Maréchal (EM), CEA Grenoble (Dorrell, Zhang, et al., 2024).

Work on the Arctic algal and haptophyte pan-genomes is already supported, as a work-package of an ERC Starting Grant and via a JGI CSP Grant respectively. Funding is currently being sought to continue both projects, e.g. fellowships for the post-doc within the Arctic research programme (Louis Graf) and an HFSP collaborative grant (Alastair Skeffington, El Mahdi Bendif, Jelena Godrijan) on haptophytes. The research on polar *Chlamydomonadales* is currently with an ANR PRC pending (**Fig. 6**), and if funded is the most likely next doctoral project offered by my group.

Challenge 1: assembly of reference genomes and transcriptomes

An important starting point for all subsequent projects is the generation of high-quality reference algal genomes, for which we are using combined short- and long-read strategies. The haptophyte genomes are being sequenced as part of a JGI Community Sequencing Programme grant (Dorrell et al., 2023; Penot et al., 2022) using combined short-read (Illumina) and long-read (PacBio) sequencing. Other algae (i.e., Arctic species, *Chlamydomonadales*) are being sequenced with BGI (DBNSeq) and in-lab (Oxford Nanopore). This work is being led by a supervised postdoc (Louis Graf) with extensive expertise in genome assembly and annotation (Graf et al., 2021).

Genomes are typically being annotated using transcriptome data, with multiple physiological conditions to expose variably expressed gene models (Rastogi et al., 2018). For each strain sequenced by the JGI, we are performing Isoseq on RNA generated under light and dark conditions at the optimal growth temperature, and Illumina (2 x 100 bp) paired-end sequencing on RNA generated for all viable growth temperatures (**Fig. 4**). For Arctic and *Chlamydomonadalean* strains, we are performing strand-specific Illumina (2 x 150 bp, Genewiz) and paired-end DBNSeq (2 x 100 bp, Genewiz) under different temperatures (0, 4, 12°C) and salinities (35‰, 15‰, 7‰, 3.5‰). For the haptophyte CSP, we have also been asked to produce RNA from cells grown in 70‰ salinity, corresponding to some hypersaline terrestrial ecosystems in which they occur, e.g. Lake Deadmoose in the Canadian Rockies (Araie et al., 2018; Plancq et al., 2019).

Challenge 2: genome annotation and environmental mapping

A first question will be to assess global phylogenetic distributions of individual groups, in particular with regard to thermal niche. Previous studies of cryophilic *Chlamydomonadales* have suggested that those close to mesophilic strains such as *C. reinhardtii* may show viability > 20°C, while more distantly-related clades (e.g., *C. sp. ICE*) do not (Chae et al., 2023; E. J. Kim et al., 2016). However, the phylogenies in these studies are typically based on single marker gene (e.g., 18S rDNA, *rbcl*) trees, and a genome-scale (phylogenomic) tree will give us more information on how growth temperature varies with taxonomy (Dorrell et al., 2023; Novák Vanclová et al., 2024).

Similarly, for the haptophyte project, genome-scale phylogenies may reveal the internal structure of *Imantonia*, e.g. whether the polar strains sequenced (RCC4500, RCC4501, RCC4780) separate from their temperate relatives (RCC2298, RCC4477, RCC704; **Fig. 4**).

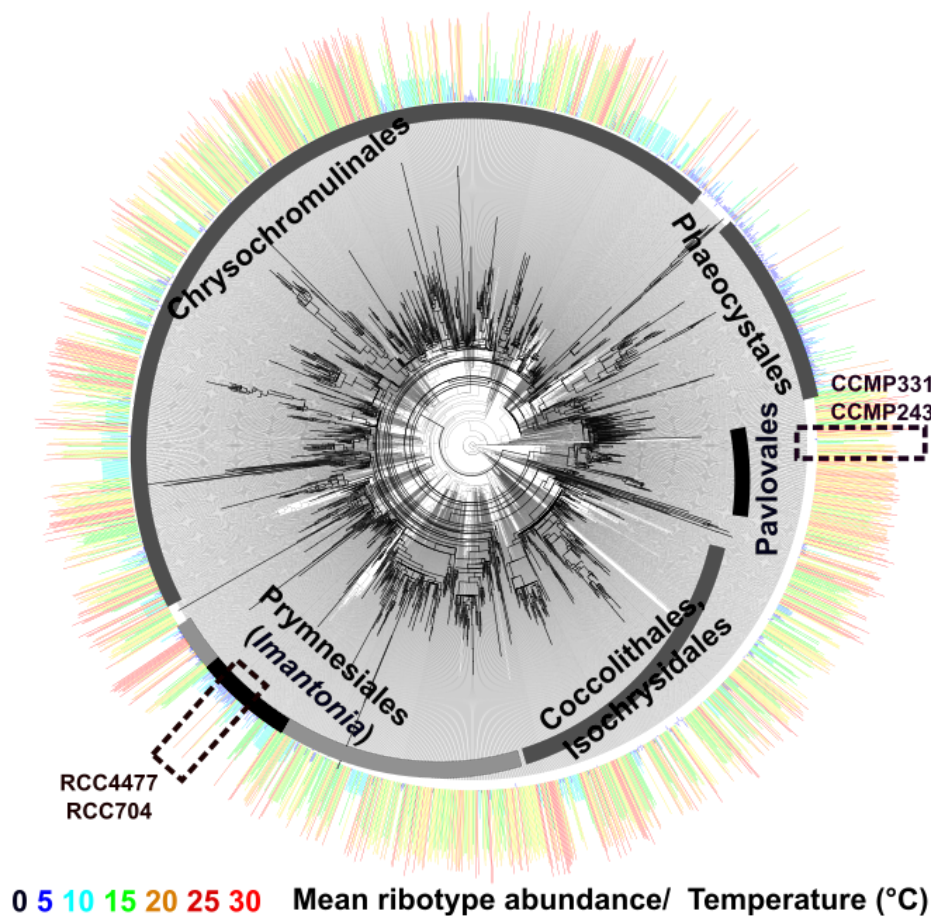


Fig. 7. Varying thermal niche of haptophytes assessed from environmental data. This figure shows the consensus (best-scoring) RAxML tree of a 2775 branch x 2624 nt tree of haptophyte 18S rDNA sequences, based on a previous alignment of cultured species enriched with *Tara* Oceans 18S V9 rDNA meta-barcodes (de Vargas et al., 2015; Dorrell et al., 2023). The tree is constrained based on a topology obtained with cultured species only, and five major haptophyte orders (Penot et al., 2022) are labelled. The outer wheel (coloured) infers the optimal temperature of each meta-barcode, based on the mean (temperature x relative abundance) calculated across all depths and size fractions. Two groups with atypical thermal niches for their lineage (cold-adapted

Pavlovophyceae, and warm-water adapted *Imantonia*) are shown with hatched boxes. This figure was generated with the aid of a short-term research engineer, Sophia Lambert, recruited to my group in 2022 (Lambert et al., 2023).

We will also explore the environmental distributions of sequenced strains, to determine whether the thermal niches observed in-lab (**Fig. 4**) can also be found in field data. This may be performed by mapping the ribotype distributions of sequenced strains in *Tara* Oceans (**Figs. 4, 5A**); or, if the species is abundant enough, to mapping its entire genome onto meta-genome data (Delmont et al., 2022). I am particular interested to identify strains that have recently moved across habitats from field data, so that we can identify genome adaptations specifically associated with changes in thermal niche. For haptophytes, using 18S V9 rDNA meta-barcodes (**Fig. 7**), we can see for example that the majority of Pavlovophyceae barcodes relate to high-temperature (20-30°C) oceans, while the *Imantonia/ Pseudohaptolina* complex is ancestrally from cooler regions (5-15°C). Within each clade, however, exceptions exist, and the cultured species to which they relate (cryotolerant Pavlovophyceae- CCMP2436, CCMP3319; thermotolerant *Imantonia* - RCC4477, RCC704; **Fig. 4, D**) may have traversed thermal niches in their recent evolution.

Next, we will aim to associate genes in each genome to the thermal niche of the strain, considering both lab and field data. Gene gains and losses associated will be determined using phylogenetically-aware techniques (e.g., phyl-PCA, CAFE, Functional Gene Fingerprinting) to allow us to separate taxonomic signals and environmental adaptations (Hennon & Dyhrman, 2020; Steindorff et al., 2021). A particular interest will be to go beyond conventional annotation techniques (e.g., PFAMs) to identify gene families that are unique to algae and have no obvious

homology amongst model species; which we may be able to explore considering expertise in AI-based annotation amongst my colleagues at the LCQB. The expression patterns of specific genes may be explored also using meta-transcriptome data from *Tara Oceans*, to identify if genes variably present across different strains in our collection show temperature dependent expression in the wild (Dorrell, Zhang, et al., 2024). This will be particularly useful to identify thermotolerance-associated genes from Arctic data, where we see greater overlap between cultured and wild species than in other ecosystems (Gérikas Ribeiro et al., 2020; Wutkowska et al., 2023).

Annotating horizontally acquired genes will remain of interest, as they have been implicated in temperature and salinity-tolerance adaptations in polar and cryophilic algae (Freyria et al., 2022; Peng et al., 2020). For example, one immediate question for study would be whether cryophilic Arctic algae possess expanded repertoires of ice-binding proteins, compared to species that cannot survive below 4°C; or whether haptophytes and chrysophytes, tolerant of hyposalinity, possess different ice-binding protein isoforms to diatoms, which are not (S. Kim et al., 2013; Lauritano & Coppola, 2023) (**Fig. 4**). We may also explore whether ancient inter-algal horizontal gene transfers inform the thermal niches of entire algal clades. Within the haptophytes, we might explore if Pavlovophyceae have exchanged genes with algal groups abundant in tropical oceans during their deep evolution (Pierella Karlusich et al., 2022). Likewise, we could determine if the *Pseudohaptolina/Imantonia* clade originates in polar oceans by determining if members of this clade, including temperate ones, possess Arctic-associated genes such as ice-binding proteins.

A second area of interest will be the reconstruction of the plastid proteomes of individual species, using *in silico* techniques developed with my former group members Tomomi Nonoyama (Nonoyama et al., 2019), Anna Novák Vanclová (Novák Vanclová et al., 2024) and Shun Liu (S. Liu et al., 2022). These will be used to construct metabolic models for plastidial amino acid and carbon metabolism (Dorrell et al., 2019; Dorrell, Zhang, et al., 2024), which may be used to infer trophic fluxes of different strains (i.e., phototrophy versus mixotrophy) using transcriptome and physiological data (Challenge 3). Where appropriate, trees will be generated for individual plastid proteins, e.g. light harvesting complex proteins, fatty acid biosynthesis enzymes and plastid transporters (Kalra et al., 2020; S. Liu et al., 2022; Remias et al., 2010; Wutkowska et al., 2023), to identify horizontal transfers, gene duplications or losses associated with thermal niche.

Finally, many of the algae cultured in my group grow with bacteria (Baker et al., 2022; Sato et al., 2020). While not intrinsically a focus of my research, bacterial coculture substantially impacts on algal primary production and metabolism, e.g. via exchange of vitamins (Croft et al., 2005; Dorrell, Nef, et al., 2024). In previous collaborations we have used *in silico* techniques (e.g., GC content and 5-mer distribution) to distinguish bacterial and algal sequences in mixed sequence libraries (Dorrell et al., 2021). We may be able to opportunistically assemble and annotate genomes from the bacterial commensals associated with some of our algae .

Challenge 3: comparative transcriptomics and stress physiology

To place genomic differences between different algal strains in a physiological context, we will need to evaluate viability of each strain as a function of temperature. Growth measurements are a crude approximation of this, and do not necessarily correspond to biogeography of collection sites (Krinis et al., 2024; Larsen et al., 1998). Alongside classical measurements of exponential phase single strain cultures (**Fig. 4**), I am interested to explore thermal niche in a biotic space via competition assay-type experiments, e.g. following the composition of mixed cultures at different temperatures by real-time sequencing or digital PCR (Shaw & Bourret, 2019; Tucker et al., 2018). This will require careful evaluation of the efficiency of barcoding primers used to follow individual

species, and the effects of starting cell concentrations on growth dynamics, but I feel is feasible with optimisation (Elbrecht & Leese, 2015; Emmerson, 1980; Schmidt & Hansen, 2001).

Beyond cell growth, photo-physiology (e.g. Fv/ Fm as a marker of photosynthetically active cells; non-photochemical quenching; lincomycin sensitivity) and mortality (e.g. SYBR staining, ROS production) may be assessed at different temperatures, particularly those observed to be outside viable growth ranges (Lacour et al., 2023; Seoane et al., 2023; H. Song et al., 2020). For example, my group is collaborating with that of Marina Cvetkovska (University of Ottawa), who has demonstrated that temperature and salinity have interactive effects on *C. priscuii*, with cultures grown under hyposaline conditions showing faster declines in viability in response to short-term heating > 19°C (Cvetkovska et al., 2022; Kalra et al., 2020). I feel that these experiments are best performed in collaboration with specialist researchers, and I would integrate an exchange visit to the Cvetkovska group for a PhD student assigned to the Chlamydomonadales project (**Fig. 6**).

Changes in cell morphology with temperature may also be considered. Cell diameter has been noted to both become more heterogeneous and on average decrease in haptophytes in response to heating and may affect biomass and cellular carbon budgets more than population size (Arnold et al., 2013). Similar techniques could be used to explore aggregation, which has been observed in heat-stressed Chlamydomonadales (Hüner et al., 2023; Zhang et al., 2023). My group may be able to leverage both conventional microscopy, and novel optometry ones (e.g., Dynamic Cell Imaging, in collaboration with Martine Boccara; MNHN) to explore these questions (Bey et al., 2022). Finally, although my laboratory does not have the capacity to assess growth at temperatures < 0°C, we may test survivability of short-term freezing (< -10°C) of specific polar species at multiple salinities (Eddie et al., 2008; S. Kim et al., 2013).

To decode the meaning of these physiological measurements, RNA sequencing will be performed of each strain under multiple temperatures to identify genes associated with heat stress responses. For species with large enough datasets (e.g., *Pavlova* sp. RCC1539, where we have performed RNAseq for six temperatures from 4°C to 32°C at which it is viable, **Fig. 4**), we may be able to attempt WGCNA (weighted gene network co-regulation analysis) to identify modules of genes that show conserved transcriptional signatures in response to temperature (Ait-Mohamed et al., 2020; Zhang et al., 2023). Particular interest will be the expression of genes encoding plastid-targeted photosynthesis and carbon metabolism proteins, as a complement to photo-physiological and lipid measurements, and to calculate trophic fluxes (Dorrell, Zhang, et al., 2024; H. Song et al., 2020; Treves et al., 2022). We may further collaborate with local researchers with expertise in plant histone dynamics and chromosome structure (e.g., Fredy Barneche and Clara Richet-Bourbousse, LBD (Teano et al., 2023)) to understand epigenetic reasons for these transcriptional changes.

Finally, we will support transcriptomic data with metabolite measurements in selected species. This will be oriented towards understanding the trophic strategies of mixotrophic groups such as haptophytes, in which increased photosynthetic capacity at warming temperatures (Schaum et al., 2017) may be counterbalanced by increased phagotrophy and respiration (Wilken et al., 2013). I am particularly interested to consider lipids, as there is already a substantial base of data available on algal lipid production in response to temperature (Ahn et al., 2015; Jung et al., 2016; E. J. Kim et al., 2019; Lauritano & Coppola, 2023). Changes in lipid chain length and unsaturation may mediate algal physiological responses to temperature stress via homeoviscous adaptation (Morales-Sánchez et al., 2020; Theroux et al., 2020). Lipid measurements are more likely to be appropriate to secondary chloroplast-containing species (e.g., haptophytes) where it forms the predominant storage pool, as opposed to in green algae (e.g. Chlamydomonadales) where carbon storage is biased towards plastidial starch (Ezzedine et al., 2023; Treves et al., 2022). Regarding my research, the interest nonetheless is to use lipids as a means for understanding total cellular

nutrient fluxes (i.e., glycolysis versus gluconeogenesis; and C: N: P ratios); that may allow us to readily trace the trophic poise of individual strains (Dorrell, Zhang, et al., 2024).

Should resources permit, we may attempt stable isotope labelling to reconstruct metabolic fluxes in individual species. We previously tested this for *Phaeodactylum* in the glycolysis project, in which we could detect the accumulation of $^{13}\text{C}_2$ -labelled glycerol from one day incubation (Dorrell, Zhang, et al., 2024; S. R. Smith et al., 2019). ^{14}C labelling has now been developed in other major algal groups including haptophytes (Godrijan et al., 2020), where it may have particular use for following flexible carbon concentrating mechanisms and membrane permeability to CO_2 that vary as a function of trophic strategy and temperature (Blanco-Ameijeiras et al., 2020). These experiments will involve first determining what metabolites to measure, given the multiplicity of carbon and nitrogen metabolism pathways in most marine algae (Dorrell, Zhang, et al., 2024; Kroth et al., 2008). Alongside optimising incubation conditions, I would only explore this in species for which we can first construct whole-cell metabolic models (S. R. Smith et al., 2019).

Challenge 4: cell biology in transformable model systems

Finally, we may try to characterise the functions of proteins whose presence, absence or expression contributes to algal thermal niche. The most likely strategy for this will be heterologous expression of these proteins in *E. coli*, yeast or *Chlamydomonas*. These techniques are currently being developed by Mathias Penot-Raquin and Victoria Powell within my group to explore the biochemical activities of xATPA and Trans3 respectively. I am interested to explore the kinetic activities of ice-binding proteins, for which well-defined assays have been developed to measure the variable capacities of different protein forms to prevent ice nucleation (“thermal hysteresis”) (Lauritano & Coppola, 2023; Vance et al., 2019). This could be combined with phylogenetic or *in silico* structural modelling, to determine if ice-binding proteins associated with cryophilic species show greater thermal hysteresis than those from eurythermal relatives.

A parallel question may be to develop reverse genetic strategies in one or two species in our collection. This is high-risk and potentially time-consuming, requiring integration of many factors: the ability to isolate clonal cultures on solid or liquid media; receptiveness to foreign DNA via different techniques (biolistic transformation, conjugation, or electroporation); and having a high quality reference genome (Endo et al., 2016; Skeffington et al., 2020). Within the haptophytes, I feel that *Pavlova* may be an interesting system to explore due to its soft structure (no extracellular scales or coccoliths), robustness to both heat and salinity stress, small (70-90 MB) genomes sequenced by us and other authors, and at least one prior report of its successful transformation (**Fig. 4**) (Hulatt et al., 2021; Velmurugan & Deka, 2018). Within the polar species in our collection, *Chlamydomonas* is the most appealing model for future experimental manipulation, as many marine polar strains can be grown on solid media, and may be amenable to transformation strategies established in *C. reinhardtii* (Fauser et al., 2022; L. Wang et al., 2023).

A final and very high-risk approach would be to combine sequencing with directed evolution of species to high or low temperatures, as has been performed for *C. reinhardtii*, to identify spontaneous mutations that allow adaptation to these conditions (Schaum et al., 2017; Valledor et al., 2013). I am hesitant to perform experimental evolution within my group as it requires precise understanding of algal generation time and life-cycle, has a high risk of failure (e.g., due to culture loss); and may not give informative results with laboratory strains that have been domesticated under non-environmental temperatures (Eydallin et al., 2014; Lehmann et al., 2023). I would nonetheless be interested to contribute sequencing and genome analysis expertise to collaborators working on the experimental evolution of freshly isolated algal strains, as detailed below.

5. Position related to current work and longer-term outcomes

Defining the thermal niches of algae is a question of fundamental biological importance, and is critical to modelling the future viability of the Earth ecosystem. It is unsurprisingly the focus of many groups, working from the levels of pan-genomes (Bendif et al., 2023; Smith & Trimborn, 2024); comparative physiology (Krinos et al., 2024); environmental genomics (Chaffron et al., 2021); and experimental evolution (Schaum et al., 2017). My work-programme is differentiated by trying to link all of these approaches together, e.g. defining gene innovations that relate to thermal niche both in pan-genomes and meta-genomic data, and characterising these innovations at the bench. I also hope to provide complementary insights through the phylogenetically aware framework of my research programme. On the one hand, I am interested in using different taxonomic scales (e.g., heating versus salinity tolerance in Arctic diatoms versus flagellates, different thermal niches in haptophytes over micro-evolutionary distances, **Fig. 4**) to expose trait variations for experimental study. On the other hand, I am particularly interested in evaluating how horizontally acquired genes (e.g., ice-binding proteins) and the metabolism of endosymbiotic organelles contribute to these niches. I hope that purely by approaching these questions from a macro-evolutionary background I may be able to provide complementary insights to those already explored within the field.

A long-term objective will be to move from laboratory culture up to field abundances. While I predominantly work with cultured species, experimental and environmental reconstructions produce different estimates of temperature optima (Litchman & Thomas, 2022). Within our culture collection, we observe Arctic endemic diatoms with viable laboratory growth temperatures up to 19°C (**Fig. 4**). In the long term, I feel it will be important to build general associative models that link lab and field thermal niches. This could be performed at a meta-analysis level, e.g. by identifying the known thermal niches of all algal strains with genomes and transcriptomes, which I have already assembled from culture collection records (Dorrell et al., 2023), and could further supplement with data from the literature (Brandenburg et al., 2019; Sheward et al., 2023). To validate this model, I would then use 18S, 16S and *psbO* sequences to calculate probable *Tara* Oceans distributions for each strain (Pierella Karlusich et al., 2022, 2023), and determine to what extent laboratory based predictions define wild thermal niche (Wells et al., 2015).

The data from this associative model could effectively be used to train a “temperature preference predictor” from genome data, in the same way as *in silico* predictors have illuminated the chloroplast proteomes and probable mixotrophy of non-model algal groups (Bock et al., 2021; Gruber et al., 2015). To validate this model, we could look for concordant gene patterns in meta-genome assembled genomes, for which *de facto* species ranges are already assessed (Delmont et al., 2022). A weakness of this approach is the validation of laboratory predictions from present-day meta-genomic data, which may not give us insights into which genes permit viability in temperature regimes beyond current maxima (Sauterey et al., 2023).

I am therefore more interested to collaborate with researchers working on time-series studies in ecosystems impacted by progressive heating. In one case, collaborating with Eric Maréchal at CEA Grenoble (Stewart et al., 2021), we are planning to explore ice-binding proteins as secondary meta-barcoding markers of heat and salinity stress across a freshwater to marine transect in East Greenland, which should allow us to identify IBP families that become more abundant in meltwater conditions (**Fig. 5C**). I am also interested in collaborating with researchers working with mesocosms, to determine if heating resilience trends in the lab apply in more unpredictable conditions in the field (Soulié et al., 2023; Trainer et al., 2020). For this, I am currently preparing an HFSP application to identify which haptophytes predominate in natural mesocosms in the Arctic, Canada and Croatia under both natural and artificial heatwaves, with Alastair Skeffington at Univ.

Stirling, El Mahdi Bendif, at Univ. Québec-Rimouski, and Jelena Godrijan at the Ruđer Bošković Institute (Bendif et al., 2011; Godrijan et al., 2020; Skeffington et al., 2020).

To date, most of my research has been based around sequencing, with some ancillary integration of physiology. It may be possible in the longer term to integrate ecosystem markers that go beyond meta-genomics, e.g. pigments such as 19'-butanoyl- and hexanoyl- oxyfucoxanthin for haptophytes (J.-X. Wang et al., 2021), and lipids such as *trans*- ω 3 PUFAs specific to *Imantonia* (Lee Chang et al., 2016). Projecting how these markers respond to heating, at the lab and field, may allow for more time-sensitive visualisation of algal abundance (e.g. by remote sensing) (Xi et al., 2023) or reconstructions of past community composition (e.g., from sediments) (Hou et al., 2014). A further question may be to reconstruct predator-prey cycles. These are likely to show complex responses to climate change, due to lower migratory velocities of large predators than small prey (Boras et al., 2010; Chivers et al., 2017; Frémont et al., 2022; Murphy et al., 2020; Sommeria-Klein et al., 2021). This could be explored via experimental grazing (Prosser et al., 2012), or fecal pellet barcoding in mesocosm communities, to determine how predation of algae changes with temperature.

An alternative approach would be to combine mesocosms and cultivation, e.g. isolating thermo-resilient algae from heated mesocosm water for subsequent genome profiling (Eydallin et al., 2014). This will be important to determine how thermal niches vary between closely related strains within the same population (Kling et al., 2023); and for experimental evolution, where we may explore thermal adaptation more informatively from naïve strains (O'Donnell et al., 2018; Schaum et al., 2017). This will be important for augmenting our understanding of diverse but poorly cultivated clades, e.g. the new marine chrysophyte clade II in study in my group (**Fig. 4**) (Andersen et al., 2017); and Arctic Kareniacean dinoflagellates, which are abundant in meta- genome data, but not to my knowledge yet in culture (Comeau et al., 2013; Wutkowska et al., 2023).

Finally, I note that temperature is only one axis of climate change, and it will be important to be aware of additional environmental variables in my future research. For example, changes in nutrient availability (e.g. due to stratification) may change the magnitude, or even direction of the relationship between cell size, growth and temperature (Skau et al., 2017). Alternative future projects may probe combined impacts of temperature and toxicity (e.g., eutrophication, heavy metal, and xenobiotic stressors), which have been shown to have interactive effects on temperature stress responses in Arctic algae (Du et al., 2023; Seoane et al., 2023; Zhang et al., 2023). Similarly, if my group moves into exploring coastal and terrestrial algae, then considering changes in rainfall and seasonality alongside temperature may be important to understanding their future biology and distributions (Hu et al., 2022; F. Li et al., 2023; Roelke et al., 2012; Siswanto et al., 2020; Suikkanen et al., 2013; Zhao et al., 2023). Nonetheless I feel that considering these combined interactions go beyond the current scope of my group to accurately study.

6. References

- Agusti, S., Lubián, L. M., Moreno-Ostos, E., Estrada, M., & Duarte, C. M. (2019). Projected changes in photosynthetic picoplankton in a warmer subtropical ocean. *Front Mar Sci*, *5*, 00506.
- Ahn, J.-W., Hwangbo, K., Yin, C. J., Lim, J.-M., Choi, H.-G., Park, Y.-I., & Jeong, W.-J. (2015). Salinity-dependent changes in growth and fatty acid composition of new Arctic *Chlamydomonas* species, ArM0029A. *Plant Cell, Tissue and Organ Culture*, *120*(3), 1015–1021.
- Ait-Mohamed, O., Novák Vanclová, A. M. G., Joli, N., Liang, Y., Zhao, X., Genovesio, A., Tirichine, L., Bowler, C., & Dorrell, R. G. (2020). PhaeoNet: A holistic RNAseq-based portrait of transcriptional coordination in the model diatom *Phaeodactylum tricornutum*. *Front Plant Sci*, *11*, 590949.

- Andersen, R. A., Graf, L., Malakhov, Y., & Yoon, H. S. (2017). Rediscovery of the *Ochromonas* type species *Ochromonas triangulata* (Chrysophyceae) from its type locality (Lake Veysove, Donetsk region, Ukraine). *Phycologia*, *56*(6), 591–604.
- Araie, H., Nakamura, H., Toney, J. L., Haig, H. A., Plancq, J., Shiratori, T., Leavitt, P. R., Seki, O., Ishida, K.-I., Sawada, K., Suzuki, I., & Shiraiwa, Y. (2018). Novel alkenone-producing strains of genus *Isochrysis* (Haptophyta) isolated from Canadian saline lakes show temperature sensitivity of alkenones and alkenoates. *Organic Geochemistry*, *121*, 89–103.
- Ardyna, M., & Arrigo, K. R. (2020). Phytoplankton dynamics in a changing Arctic Ocean. *Nat Climate Change*, *10*(10), 892–903.
- Ardyna, M., Babin, M., Gosselin, M., Devred, E., Rainville, L., & Tremblay, J.-É. (2014). Recent Arctic Ocean sea ice loss triggers novel fall phytoplankton blooms. *Geophys Res Lett*, *41*(17), 6207–6212.
- Ardyna, M., d'Ovidio, F., Speich, S., Leconte, J., Chaffron, S., Audic, S., Garczarek, L., Pesant, S., Tara Oceans Consortium Coordinators, & Tara Oceans Expedition Participants. (2017). PANGAEA. doi: 10.1594/PANGAEA.875577
- Arnold, H. E., Kerrison, P., & Steinke, M. (2013). Interacting effects of ocean acidification and warming on growth and DMS-production in the haptophyte coccolithophore *Emiliana huxleyi*. *Glob Change Biol*, *19*(4), 1007–1016.
- Asch, R. G., Stock, C. A., & Sarmiento, J. L. (2019). Climate change impacts on mismatches between phytoplankton blooms and fish spawning phenology. *Glob Change Biol*, *25*(8), 2544–2559.
- Baker, D., Lauer, J., Ortega, A., Jackrel, S. L., & Deneff, V. J. (2022). Effects of phycosphere bacteria on their algal host are host species-specific and not phylogenetically conserved. *Microorganisms*, *11*(1).
- Barcytė, D., Hodač, L., Nedbalová, L., & Elster, J. (2018). *Chloromonas arctica* sp. nov., a psychrotolerant alga from snow in the High Arctic (Chlamydomonadales, Chlorophyta). *Int J Syst Evol Microbiol*, *68*(3), 851–859.
- Beau, R., & Larrère, C. (2018). *Penser l'Anthropocène*. Presses de Sciences Po.
- Bendif, E. M., Probert, I., Archontikis, O. A., Young, J. R., Beaufort, L., Rickaby, R. E., & Filatov, D. (2023). Rapid diversification underlying the global dominance of a cosmopolitan phytoplankton. *ISME J*, *17*(4), 630–640.
- Bendif, E. M., Probert, I., Hervé, A., Billard, C., Goux, D., Lelong, C., Cadoret, J.-P., & Véron, B. (2011). Integrative taxonomy of the Pavlovophyceae (Haptophyta): a reassessment. *Protist*, *162*(5), 738–761.
- Bendif, E. M., Probert, I., Schroeder, D. C., & de Vargas, C. (2013). On the description of *Tisochrysis lutea* gen. nov. sp. nov. and *Isochrysis nuda* sp. nov. in the Isochrysidales, and the transfer of *Dicrateria* to the Prymnesiales (Haptophyta). *J Appl Phycol*, *25*(6), 1763–1776.
- Bey, H., Charton, F., Cruz de Carvalho, H., Liu, S., Dorrell, R. G., Bowler, C., Boccara, C., & Boccara, M. (2022). Dynamic Cell Imaging: application to the diatom *Phaeodactylum tricorutum* under environmental stresses. *Eur J Phycol*, 1–11.
- Blanco-Ameijeiras, S., Stoll, H. M., Zhang, H., & Hopkinson, B. M. (2020). Influence of temperature and CO₂ on plasma-membrane permeability to CO₂ and HCO₃⁻ in the marine haptophytes *Emiliana huxleyi* and *Calcidiscus leptoporus* (Prymnesiophyceae). *J Phycol*, *56*(5), 1283–1294.
- Bock, N. A., Charvet, S., Burns, J., Gyaltsen, Y., Rozenberg, A., Duhamel, S., & Kim, E. (2021). Experimental identification and in silico prediction of bacterivory in green algae. *ISME J*, *15*(7), 1987–2000.
- Boras, J. A., Sala, M. M., Arrieta, J. M., Sà, E. L., Felipe, J., Agustí, S., Duarte, C. M., & Vaqué, D. (2010). Effect of ice melting on bacterial carbon fluxes channelled by viruses and protists in the Arctic Ocean. *Polar Biol*, *33*(12), 1695–1707.
- Brandenburg, K. M., Velthuis, M., & Van de Waal, D. B. (2019). Meta-analysis reveals enhanced growth of marine harmful algae from temperate regions with warming and elevated CO levels. *Glob Change Biol*, *25*(8), 2607–2618.
- Burki, F., Okamoto, N., Pombert, J.-F., & Keeling, P. J. (2012). The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. *Proc R Soc Biol*, *279*(1736), 2246–2254.
- Chae, H., Kim, S., Choi, H.-G., Kim, J. H., Han, S. J., & Kim, E. J. (2023). Arctic sea ice microalga KNF0041: identification and statistical optimization of medium for enhanced biomass and omega-3/ omega-6. *Marine Drugs*, *21*(8).

- Chaffron, S., Delage, E., Budinich, M., Vintache, D., Henry, N., Nef, C., Ardyna, M., Zayed, A. A., Junger, P. C., Galand, P. E., Lovejoy, C., Murray, A. E., Sarmiento, H., Tara Oceans coordinators, Acinas, S. G., Babin, M., Iudicone, D., Jaillon, O., Karsenti, E., ... Eveillard, D. (2021). Environmental vulnerability of the global ocean epipelagic plankton community interactome. *Sci Adv*, 7(35), 1921
- Chivers, W. J., Walne, A. W., & Hays, G. C. (2017). Mismatch between marine plankton range movements and the velocity of climate change. *Nat Commun*, 8(1), 14434.
- Comeau, A. M., Philippe, B., Thaler, M., Gosselin, M., Poulin, M., & Lovejoy, C. (2013).. *Journal of Phycology*, 49(2), 229–240.
- Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J., & Smith, A. G. (2005). Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature*, 438(7064), 90–93.
- Cvetkovska, M., Hüner, N. P. A., & Smith, D. R. (2017). Chilling out: the evolution and diversification of psychrophilic algae with a focus on Chlamydomonadales. *Polar Biol*, 40(6), 1169–1184.
- Cvetkovska, M., Zhang, X., Vakulenko, G., Benzaquen, S., Szyszka-Mroz, B., Malczewski, N., Smith, D. R., & Hüner, N. P. A. (2022). A constitutive stress response is a result of low temperature growth in the Antarctic green alga *Chlamydomonas* sp. UWO241. *Plant, Cell Env*, 45(1), 156–177.
- Delmont, T. O., Gaia, M., Hinsinger, D. D., Frémont, P., Vanni, C., Fernandez-Guerra, A., Eren, A. M., Kourlaiev, A., d'Agata, L., Clayssen, Q., Villar, E., Labadie, K., Cruaud, C., Poulain, J., Da Silva, C., Wessner, M., Noel, B., Aury, J.-M., Tara Oceans Coordinators, ... Jaillon, O. (2022). Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean. *Cell Genom*, 2(5), 100123.
- Dennu, L., Devic, M., Rigonato, J., Falciatore, A., Lozano, J.-C., Vergé, V., Mariac, C., Jaillon, O., Sabot, F., Bouget, F.-Y., & The Dark Edge genomics sampling team. (2023). Biological and genomic resources for the cosmopolitan phytoplankton *Bathycoccus*: insights into genetic diversity and major structural variations. In *bioRxiv*. doi: 10.1101/2023.10.16.562038
- Deodato, C. R., Barlow, S. B., Hovde, B. T., & Cattolico, R. A. (2019). Naked *Chrysochromulina* (Haptophyta) isolates from lake and river ecosystems: an electron microscopic comparison including new observations on the type species of this taxon. *Algal Res*, 40(101492), 101492.
- De Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.-M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., ... Karsenti, E. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science*, 348(6237), 1261605.
- Díaz, S., Settele, J., Brondízio, E. S., Ngo, H. T., Agard, J., Arneeth, A., Balvanera, P., Brauman, K. A., Butchart, S. H. M., Chan, K. M. A., Garibaldi, L. A., Ichii, K., Liu, J., Subramanian, S. M., Midgley, G. F., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., ... Zayas, C. N. (2019). Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science*, 366(6471).
- Diffenbaugh, N. S., & Burke, M. (2019). Global warming has increased global economic inequality. *Proc Natl Acad Sci USA*, 116(20), 9808–9813.
- Dorrell, R. G., Azuma, T., Nomura, M., Audren de Kerdrel, G., Paoli, L., Yang, S., Bowler, C., Ishii, K.-I., Miyashita, H., Gile, G. H., & Kamikawa, R. (2019). Principles of plastid reductive evolution illuminated by nonphotosynthetic chrysophytes. *Proc Natl Acad Sci USA*, 116(14), 6914–6923.
- Dorrell, R. G., Gile, G., McCallum, G., Méheust, R., Bapteste, E. P., Klinger, C. M., Brillet-Guéguen, L., Freeman, K. D., Richter, D. J., & Bowler, C. (2017). Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. *eLife*, 6, 23717.
- Dorrell, R. G., Kuo, A., Füßy, Z., Richardson, E. H., Salamov, A., Zarevski, N., Freyria, N. J., Ibarbalz, F. M., Jenkins, J., Pierella Karlusich, J. J., Stecca Steindorff, A., Edgar, R. E., Handley, L., Lail, K., Lipzen, A., Lombard, V., McFarlane, J., Nef, C., Novák Vanclová, A. M., ... Lovejoy, C. (2023). Convergent evolution and horizontal gene transfer in Arctic Ocean microalgae. *Life Science Alliance*, 6(3), 26508.
- Dorrell, R. G., Nef, C., Altan-Ochir, S., Bowler, C., & Smith, A. G. (2024). Presence of vitamin B₁₂ metabolism in the last common ancestor of land plants. *Proc R Soc Biol*, in press. doi: 10.1098/rstb.2023.0354
- Dorrell, R. G., Villain, A., Perez-Lamarque, B., Audren de Kerdrel, G., McCallum, G., Watson, A. K., Ait-Mohamed, O., Alberti, A., Corre, E., Frischkorn, K. R., Pierella Karlusich, J. J., Pelletier, E., Morlon, H., Bowler, C., & Blanc, G. (2021). Phylogenomic fingerprinting of tempo and

- functions of horizontal gene transfer within ochrophytes. *Proc Natl Acad Sci USA*, 118(4), 2009974118
- Dorrell, R. G., Zhang, Y., Liang, Y., Gueguen, N., Nonoyama, T., Croteau, D., Penot, M., Adiba, S., Bailleul, B., Gros, V., Karlusich, J. J. P., Zweig, N., Fernie, A. R., Jouhet, J., Maréchal, E., & Bowler, C. (2024). Environmentally-informed functional characterization of a plastid diatom metabolic bridge of mitochondrial origin. *Plant Cell*, in resubmission. doi: 10.1101/2022.09.08.507166v2
- Du, J., Izquierdo, D., Xu, H.-F., Beisner, B., Lavaud, J., Ohlund, L., Sleno, L., & Juneau, P. (2023). Responses to herbicides of Arctic and temperate microalgae grown under different light intensities. *Env Pollution*, 333, 121985.
- Duprat, L. P. A. M., Bigg, G. R., & Wilton, D. J. (2016). Enhanced Southern Ocean marine productivity due to fertilization by giant icebergs. *Nat Geoscience*, 9(3), 219–221.
- Eddie, B., Krembs, C., & Neuer, S. (2008). Characterization and growth response to temperature and salinity of psychrophilic, halotolerant *Chlamydomonas* sp. ARC isolated from Chukchi Sea ice. *Mar Ecol Progress Series*, 354, 107–117.
- Edwardsen, B., Egge, E. S., & Vault, D. (2016). Diversity and distribution of haptophytes revealed by environmental sequencing and metabarcoding – a review. *Persp Phycol*, 3(2), 77–91.
- Edwardsen, B., Eikrem, W., Green, J. C., Andersen, R. A., Moon-van der Staay, S. Y., & Medlin, L. K. (2000). Phylogenetic reconstructions of the Haptophyta inferred from 18S ribosomal DNA sequences and available morphological data. *Phycologia*, 39(1), 19–35.
- Edwardsen, B., Eikrem, W., Thronsen, J., Sáez, A. G., Probert, I., & Medlin, L. K. (2011). Ribosomal DNA phylogenies and a morphological revision provide the basis for a revised taxonomy of the Prymnesiales (Haptophyta). *Eur J Phycol*, 46(3), 202–228.
- Elbrecht, V., & Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass–sequence relationships with an innovative metabarcoding protocol. *PLoS One*, 10(7), e0130324.
- Emmerson, W. D. (1980). Ingestion, growth and development of *Penaeus indicus* larvae as a function of *Thalassiosira weissflogii* cell concentration. *Mar Biol*, 58(1), 65–73.
- Endo, H., Umezawa, Y., Takeda, S., & Suzuki, K. (2023). Haptophyte communities along the Kuroshio current reveal their geographical sources and ecological traits. *Mol Ecol*, 32(1), 110–123.
- Endo, H., Yoshida, M., Uji, T., Saga, N., Inoue, K., & Nagasawa, H. (2016). Stable nuclear transformation system for the coccolithophorid alga *Pleurochrysis carterae*. *Sci Rep*, 6, 22252.
- Evans, G. (1966). A discussion concerning the floor of the northwest Indian Ocean - the recent sedimentary facies of the Persian Gulf region. *Phil Trans R Soc*, 259(1099), 291–298.
- Eydallin, G., Ryall, B., Maharjan, R., & Ferenci, T. (2014). The nature of laboratory domestication changes in freshly isolated *Escherichia coli* strains. *Env Microbiol*, 16(3), 813–828.
- Ezzedine, J. A., Uwizeye, C., Si Larbi, G., Villain, G., Louwagie, M., Schilling, M., Hagenmuller, P., Gallet, B., Stewart, A., Petroustos, D., Devime, F., Salze, P., Liger, L., Jouhet, J., Dumont, M., Ravanel, S., Amato, A., Valay, J.-G., Jouneau, P.-H., ... Maréchal, E. (2023). Adaptive traits of cysts of the snow alga *Sanguina nivaloides* unveiled by 3D subcellular imaging. *Nat Commun*, 14(1), 7500.
- Fausser, F., Vilarrasa-Blasi, J., Onishi, M., Ramundo, S., Patena, W., Millican, M., Osaki, J., Philp, C., Nemeth, M., Salomé, P. A., Li, X., Wakao, S., Kim, R. G., Kaye, Y., Grossman, A. R., Niyogi, K. K., Merchant, S. S., Cutler, S. R., Walter, P., ... Jinkerson, R. E. (2022). Systematic characterization of gene function in the photosynthetic alga *Chlamydomonas reinhardtii*. *Nat Genet*, 54(5), 705–714.
- Fox, L., Stukins, S., Hill, T., & Miller, C. G. (2020). Quantifying the effect of anthropogenic climate change on calcifying plankton. *Sc Rep*, 10(1), 1620.
- Frémont, P., Gehlen, M., Vrac, M., Leconte, J., Delmont, T. O., Wincker, P., Iudicone, D., & Jaillon, O. (2022). Restructuring of plankton genomic biogeography in the surface ocean under climate change. *Nat Climate Change*, 12(4), 393–401.
- Freyria, N. J., Joli, N., & Lovejoy, C. (2021). A decadal perspective on north water microbial eukaryotes as Arctic Ocean sentinels. *Sci Rep*, 11(1), 8413.
- Freyria, N. J., Kuo, A., Chovatia, M., Johnson, J., Lipzen, A., Barry, K. W., Grigoriev, I. V., & Lovejoy, C. (2022). Salinity tolerance mechanisms of an Arctic pelagophyte using comparative transcriptomic and gene expression analysis. *Commun Biol*, 5(1), 500.
- Gérikas Ribeiro, C., dos Santos, A. L., Gourvil, P., Le Gall, F., Marie, D., Tragin, M., Probert, I., &

- Vaulot, D. (2020). Culturable diversity of Arctic phytoplankton during pack ice melting. *Elementa*, 8(1), 6.
- Glibert, P. M. (2020). Harmful algae at the complex nexus of eutrophication and climate change. *Harmful Algae*, 91, 101583.
- Godrijan, J., Drapeau, D., & Balch, W. M. (2020). Mixotrophic uptake of organic compounds by coccolithophores. *Limnol Oceanograph*, 65(6), 1410–1421.
- Graf, L., Shin, Y., Yang, J. H., Choi, J. W., Hwang, I. K., Nelson, W., Bhattacharya, D., Viard, F., & Yoon, H. S. (2021). A genome-wide investigation of the effect of farming and human-mediated introduction on the ubiquitous seaweed *Undaria pinnatifida*. *Nat Ecol Evol*, 5(3), 360–368.
- Green, J. C., & Hori, T. (1986). The ultrastructure of the flagellar root system of *Imantonia rotunda* (Prymnesiophyceae). *Brit Phycol Jo*, 21(1), 5–18.
- Griffith, A. W., & Gobler, C. J. (2020). Harmful algal blooms: A climate change co-stressor in marine and freshwater ecosystems. *Harmful Algae*, 91, 101590.
- Gruber, A., Roco, G., Kroth, P. G., Armbrust, E. V., & Mock, T. (2015). Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *Plant J*, 81(3), 519–528.
- Guillard, R. R. L. (1988). The center for culture of marine phytoplankton: history, structure, function, and future¹. *J Protozool*, 35(2), 255–256.
- Hallegraeff, G. M. (2010). Ocean climate change, phytoplankton community responses, and harmful algal blooms: a formidable predictive challenge. *J Phycol*, 46(2), 220–235.
- Hennon, G. M. M., & Dyhrman, S. T. (2020). Progress and promise of omics for predicting the impacts of climate change on harmful algal blooms. *Harmful Algae*, 91, 101587.
- Henson, S. A., Cael, B. B., Allen, S. R., & Dutkiewicz, S. (2021). Future phytoplankton diversity in a changing climate. *Nat Commun*, 12(1), 5372.
- Hou, W., Dong, H., Li, G., Yang, J., Coolen, M. J. L., Liu, X., Wang, S., Jiang, H., Wu, X., Xiao, H., Lian, B., & Wan, Y. (2014). Identification of photosynthetic plankton communities using sedimentary ancient DNA and their response to late-Holocene climate change on the Tibetan Plateau. *Sci Rep*, 4, 6648.
- Hulatt, C. J., Wijffels, R. H., & Posewitz, M. C. (2021). The genome of the haptophyte *Diacronema lutheri* (*Pavlova lutheri*, Pavlovales): a model for lipid biosynthesis in eukaryotic algae. *Genom Biol Evol*, 13(8), 178
- Hüner, N. P. A., Szyszka-Mroz, B., Ivanov, A. G., Kata, V., Lye, H., & Smith, D. R. (2023). Photosynthetic adaptation and multicellularity in the Antarctic psychrophile, *Chlamydomonas priscuii*. *Algal Res*, 74, 103220.
- Hutchins, D., Mulholland, M., & Fu, F. (2009). Nutrient cycles and marine microbes in a CO₂-enriched ocean. *Oceanography*, 22(4), 128–145.
- Hu, Z., Yao, J., Chen, X., Gong, X., Zhang, Y., Zhou, X., Guo, H., & Liu, M. (2022). Precipitation changes, warming, and N input differentially affect microbial predators in an alpine meadow: evidence from soil phagotrophic protists. *Soil Biol Biochem*, 165(108521), 108521.
- Ibarbalz, F. M., Henry, N., Mahé, F., Ardyna, M., Zingone, A., Scalco, E., Lovejoy, C., Lombard, F., Jaillon, O., Iudicone, D., Malviya, S., Sullivan, M. B., Chaffron, S., Karsenti, E., Babin, M., Boss, E., Wincker, P., Zinger, L., de Vargas, C., ... Tara Oceans Coordinators. (2023). Pan-Arctic plankton community structure and its global connectivity. *Elementa*, 11(1), 60.
- Intergovernmental Panel on Climate Change. (2007). Climate Change 2007 - The Physical Science Basis: Working Group I Contribution to the Fourth Assessment Report of the IPCC. Cambridge University Press.
- Jamy, M., Biwer, C., Vaulot, D., Obiol, A., Jing, H., Peura, S., Massana, R., & Burki, F. (2022). Global patterns and rates of habitat transitions across the eukaryotic tree of life. *Nat Ecol Evol*, 6(10), 1458–1470.
- John, U., Šupraha, L., Gran-Stadniczeňko, S., Bunse, C., Cembella, A., Eikrem, W., Janouškovec, J., Klemm, K., Kühne, N., Naustvoll, L., Voss, D., Wohlrab, S., & Edvardsen, B. (2022). Spatial and biological oceanographic insights into the massive fish-killing bloom of the haptophyte *Chrysochromulina leadbeateri* in northern Norway. *Harmful Algae*, 118, 102287.
- Joli, N., Concia, L., Mocaer, K., Guterman, J., Laude, J., Guerin, S., Sciandra, T., Bruyant, F., Ait-Mohamed, O., Beguin, M., Forget, M.-H., Bourbousse, C., Lacour, T., Bailleul, B., Nef, C., Savoie, M., Tremblay, J.-E., Campbell, D. A., Lavaud, J., ... Bowler, C. (2024). Hypometabolism to survive the long polar night and subsequent successful return to light in the diatom *Fragilariopsis cylindrus*. *New Phytol*, 241(5), 2193–2208.
- Jung, W., Kim, E. J., Lim, S., Sim, H., Han, S. J., Kim, S., Kang, S.-H., & Choi, H.-G. (2016).

- Cellular growth and fatty acid content of Arctic chlamydomonadales. *Algae*, 31(1), 61–71.
- Kalra, I., Wang, X., Cvetkovska, M., Jeong, J., McHargue, W., Zhang, R., Hüner, N., Yuan, J. S., & Morgan-Kiss, R. (2020). *Chlamydomonas* sp. UWO 241 exhibits high cyclic electron flow and rewired metabolism under high salinity. *Plant Physiol*, 183(2), 588–601.
- Kameyama, S., Otomaru, M., McMinn, A., & Suzuki, K. (2020). Ice melting can change DMSP production and photosynthetic activity of the haptophyte *Phaeocystis antarctica*. *J Phycol*, 56(3), 761–774.
- Kawachi, M., Nakayama, T., Kayama, M., Nomura, M., Miyashita, H., Bojo, O., Rhodes, L., Sym, S., Pienaar, R. N., Probert, I., Inouye, I., & Kamikawa, R. (2021). Rappemonads are haptophyte phytoplankton. *Curr Biol*, 31(11), 2395–2403.e4.
- Kendrick, B. J., DiTullio, G. R., Cyronak, T. J., Fulton, J. M., Van Mooy, B. A. S., & Bidle, K. D. (2014). Temperature-induced viral resistance in *Emiliania huxleyi* (Prymnesiophyceae). *PLoS One*, 9(11), e112134.
- Kim, E. J., Jung, W., Lim, S., Kim, S., Han, S. J., & Choi, H.-G. (2016). Growth and lipid content at low temperature of Arctic alga *Chlamydomonas* sp. KNM0029C. *Bioproc Biosyst Eng*, 39(1), 151–157.
- Kim, E. J., Jung, W. S., Kim, S., Choi, H.-G., & Han, S. J. (2019). Enhancing lipid production in the arctic microalga *Chlamydomonas* sp. KNM0029C by using statistical optimization of medium components. *KSB J*, 34(3), 173–177.
- Kim, S., Kim, M. J., Jung, M. G., Lee, S., Baek, Y.-S., Kang, S.-H., & Choi, H.-G. (2013). De novo transcriptome analysis of an Arctic microalga, *Chlamydomonas* sp. *Genes & Genomics*, 35(2), 215–223.
- Kling, J. D., Lee, M. D., Walworth, N. G., Webb, E. A., Coelho, J. T., Wilburn, P., Anderson, S. I., Zhou, Q., Wang, C., Phan, M. D., Fu, F., Kremer, C. T., Litchman, E., Ryneerson, T. A., & Hutchins, D. A. (2023). Dual thermal ecotypes coexist within a nearly genetically identical population of the unicellular marine cyanobacterium. *Proc Natl Acad Sci USA*, 120(47), e2315701120.
- Krinos, A. I., Shapiro, S. K., Li, W., Haley, S. T., Dyhrman, S. T., Dutkiewicz, S., Follows, M. J., & Alexander, H. (2024). Intraspecific diversity in thermal performance determines phytoplankton ecological niche. In *bioRxiv*. doi: 10.1101/2024.02.14.580366
- Kroth, P. G., Chiovitti, A., Gruber, A., Martin-Jezequel, V., Mock, T., Parker, M. S., Stanley, M. S., Kaplan, A., Caron, L., Weber, T., Maheswari, U., Armbrust, E. V., & Bowler, C. (2008). A model for carbohydrate metabolism in the diatom *Phaeodactylum tricorutum* deduced from comparative whole genome analysis. *PLoS One*, 3(1), e1426.
- Lacour, T., Robert, E., & Lavaud, J. (2023). Sustained xanthophyll pigments-related photoprotective NPQ is involved in photoinhibition in the haptophyte *Tisochrysis lutea*. *Sci Rep*, 13(1), 14694.
- Lambert, S., Voznica, J., & Morlon, H. (2023). Deep learning from phylogenies for diversification analyses. *Syst Biol*, 72(6), 1262–1279.
- Larsen, A., Bryant, S., & Båmstedt, U. (1998). Growth rate and toxicity of *Prymnesium parvum* and *Prymnesium patelliferum* (Haptophyta) in response to changes in salinity, light and temperature. *Sarsia*, 83(5), 409–418.
- Lauritano, C., & Coppola, D. (2023). Biotechnological applications of products released by marine microorganisms for cold adaptation strategies: polyunsaturated fatty acids, antioxidants, and antifreeze proteins. *J Mar Sci Eng*, 11(7), 1399.
- Lee Chang, K. J., Dunstan, G. A., Mansour, M. P., Jameson, I. D., & Nichols, P. D. (2016). A novel series of C18–C22 trans ω 3 PUFA from Northern and Southern Hemisphere strains of the marine haptophyte *Imantonia rotunda*. *J Appl Phycol*, 28(6), 3363–3370.
- Lehmann, M., Prohaska, C., Zeldes, B., Poehlein, A., Daniel, R., & Basen, M. (2023). Adaptive laboratory evolution of a thermophile toward a reduced growth temperature optimum. *Front Microbiol*, 14, 1265216.
- Lenton, T. M. (2013). Environmental tipping points. *Ann Rev Env Res*, 38(1), 1–29.
- Li, F., Sun, A., Liu, X., Ren, P., Wu, B.-X., Shen, J.-P., Bi, L., He, J.-Z., Yang, Y., & Hu, H.-W. (2023). Seasonality regulates the taxonomic and functional compositions of protists responding to climate warming in forest ecosystems. *J Sust Agriculture Env*, 2(4), 529–540.
- Litchman, E., & Thomas, M. K. (2022). Are we underestimating the ecological and evolutionary effects of warming? Interactions with other environmental drivers may increase species vulnerability to high temperatures. *Oikos* 2, 09155.

- Liu, H., Probert, I., Uitz, J., Claustre, H., Aris-Brosou, S., Frada, M., Not, F., & de Vargas, C. (2009). Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *Proc Natl Acad Sci USA*, *106*(31), 12803–12808.
- Liu, S., Storti, M., Finazzi, G., Bowler, C., & Dorrell, R. G. (2022). A metabolic, phylogenomic and environmental atlas of diatom plastid transporters from the model species. *Front Plant Sci*, *13*, 950467.
- Li, W. K. W., McLaughlin, F. A., Lovejoy, C., & Carmack, E. C. (2009). Smallest algae thrive as the Arctic Ocean freshens. *Science*, *326*(5952), 539.
- Marchant, H. J., & Thomsen, H. A. (1994). Haptophytes in polar waters. In *The Haptophyte Algae* (pp. 209–228). Oxford University Press, Oxford.
- Marlowe, I. T., Green, J. C., Neal, A. C., Brassell, S. C., Eglinton, G., & Course, P. A. (1984). Long chain ($n\text{-C}_{37}\text{--C}_{39}$) alkenones in the Prymnesiophyceae. Distribution of alkenones and other lipids and their taxonomic significance. *Brit Phycol J*, *19*(3), 203–216.
- Mars Brisbin, M., & Mitarai, S. (2019). Differential gene expression supports a resource-intensive, defensive role for colony production in the bloom-forming haptophyte, *Phaeocystis globosa*. *J Euk Microbiol*, *66*(5), 788–801.
- Mas, S., Roy, S., Blouin, F., Mostajir, B., Therriault, J.-C., Nozais, C., & Demers, S. (2008). Diel variations in optical properties of *Imantonia rotunda* (Haptophytceae) and *Thalassiosira pseudonana* (Bacillariophyceae) exposed to different irradiance levels. *J Phycoly*, *44*(3), 551–563.
- Morales-Sánchez, D., Schulze, P. S. C., Kiron, V., & Wijffels, R. H. (2020). Production of carbohydrates, lipids and polyunsaturated fatty acids (PUFA) by the polar marine microalga *Chlamydomonas malina* RCC2488. *Algal Res*, *50*(102016), 102016.
- Murphy, G. E. P., Romanuk, T. N., & Worm, B. (2020). Cascading effects of climate change on plankton community structure. *Ecol Evol*, *10*(4), 2170–2181.
- Nemcek, N., Hennekes, M., Sastri, A., & Perry, R. I. (2023). Seasonal and spatial dynamics of the phytoplankton community in the Salish Sea, 2015–2019. *Progr Oceanograph*, *217*(103108), 103108.
- NOAA Global Monitoring Laboratory. (2024). <https://gml.noaa.gov/ccgg/trends/>
- Nonoyama, T., Kazamia, E., Nawaly, H., Gao, X., Tsuji, Y., Matsuda, Y., Bowler, C., Tanaka, T., & Dorrell, R. G. (2019). Metabolic innovations underpinning the origin and diversification of the diatom chloroplast. *Biomolecules*, *9*(8), 484.
- Noori, R., Tian, F., Berndtsson, R., Abbasi, M. R., Naseh, M. V., Modabberi, A., Soltani, A., & Kløve, B. (2019). Recent and future trends in sea surface temperature across the Persian Gulf and Gulf of Oman. *PLoS One*, *14*(2), e0212790.
- Novák Vanclová, A. M. G., Nef, C., Vancl, A., Liu, F., Füssy, Z., Bowler, C., & Dorrell, R. G. (2024). New plastids, old proteins: repeated endosymbiotic acquisitions in karenian dinoflagellates. *EMBO Rep*, *in press*. doi: 10.1038/s44319-024-00103-y
- O'Donnell, D. R., Hamman, C. R., Johnson, E. C., Kremer, C. T., Klausmeier, C. A., & Litchman, E. (2018). Rapid thermal adaptation in a marine diatom reveals constraints and trade-offs. *Glob Change Biol*, *24*(10), 4554–4565.
- Pajot, A., Lavaud, J., Carrier, G., Lacour, T., Marchal, L., & Nicolau, E. (2023). Light-response in two clonal strains of the haptophyte *Tisochrysis lutea*: evidence for different photoprotection strategies. *Algal Res*, *69*, 102915.
- Peng, Z., Liu, G., & Huang, K. (2020). Cold adaptation mechanisms of a snow alga during temperature fluctuations. *Front Microbiol*, *11*, 611080.
- Penot, M., Dacks, J. B., Read, B., & Dorrell, R. G. (2022). Genomic and meta-genomic insights into the functions, diversity and global distribution of haptophyte algae. *Appl Phycol*, *3*(1), 340–359.
- Pierella Karlusich, J. J., Nef, C., Bowler, C., & Dorrell, R. G. (2022). Biogeographical patterns and genomes of aquatic photoautotrophs. In *Blue Planet, Red and Green Photosynthesis*, eds. Gontero & Maberly, pp. 43–79. Wiley.
- Pierella Karlusich, J. J., Pelletier, E., Zinger, L., Lombard, F., Zingone, A., Colin, S., Gasol, J. M., Dorrell, R. G., Henry, N., Scalco, E., Acinas, S. G., Wincker, P., de Vargas, C., & Bowler, C. (2023). A robust approach to estimate relative phytoplankton cell abundances from metagenomes. *Mol Ecol Res*, *23*(1), 16–40.
- Plancq, J., Couto, J. M., Ijaz, U. Z., Leavitt, P. R., & Toney, J. L. (2019). Next-generation sequencing to identify lacustrine haptophytes in the Canadian prairies: Significance for temperature proxy applications. *J Geophys Res*, *124*(7), 2144–2158.

- Prosser, K. N., Valenti, T. W., Jr, Hayden, N. J., Neisch, M. T., Hewitt, N. C., Umphres, G. D., Gable, G. M., Grover, J. P., Roelke, D. L., & Brooks, B. W. (2012). Low pH preempts bloom development of a toxic haptophyte. *Harmful Algae*, 20, 156–164.
- Randelhoff, A., Lacour, L., Marec, C., Leymarie, E., Lagunas, J., Xing, X., Darnis, G., Penkerch, C., Sampei, M., Fortier, L., D'Ortenzio, F., Claustre, H., & Babin, M. (2020). Arctic mid-winter phytoplankton growth revealed by autonomous profilers. *Sci Advances*, 6(39).
- Rastogi, A., Maheswari, U., Dorrell, R. G., Vieira, F. R. J., Maumus, F., Kustka, A., McCarthy, J., Allen, A. E., Kersey, P., Bowler, C., & Tirichine, L. (2018). Integrative analysis of large scale transcriptome data draws a comprehensive landscape of *Phaeodactylum tricornutum* genome and evolutionary origin of diatoms. *Sci Rep*, 8(1), 4834.
- Remias, D., Albert, A., & Lütz, C. (2010). Effects of realistically simulated, elevated UV irradiation on photosynthesis and pigment composition of the alpine snow alga *Chlamydomonas nivalis* and the arctic soil alga *Tetracystis* sp. (Chlorophyceae). *Photosynthetica*, 48(2), 269–277.
- Remias, D., Boenigk, J. S., Wastian, J., & Lütz, C. (2013). *Hydrurus*-related golden algae (Chrysophyceae) cause yellow snow in polar summer snowfields. *Phycol Res*, 61(4), 277–285.
- Reynolds, N. (1974). *Imantonia rotunda* gen. et sp. nov., a new member of the Haptophyceae. *Brit Phycol J*, 9(4), 429–434.
- Richardson, K., Steffen, W., Lucht, W., Bendtsen, J., Cornell, S. E., Donges, J. F., Drüke, M., Fetzer, I., Bala, G., von Bloh, W., Feulner, G., Fiedler, S., Gerten, D., Gleeson, T., Hofmann, M., Huiskamp, W., Kummu, M., Mohan, C., Nogués-Bravo, D., ... Rockström, J. (2023). Earth beyond six of nine planetary boundaries. *Sci Adv*, 9(37), eadh2458.
- Rockström, J., Steffen, W., Noone, K., Persson, A., Chapin, F. S., 3rd, Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., Nykvist, B., de Wit, C. A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P. K., Costanza, R., Svedin, U., ... Foley, J. A. (2009). A safe operating space for humanity. *Nature*, 461(7263), 472–475.
- Roelke, D. L., Brooks, B. W., Grover, J. P., Gable, G. M., Schwierzke-Wade, L., & Hewitt, N. C. (2012). Anticipated human population and climate change effects on algal blooms of a toxic haptophyte in the south-central USA. *Can J Fisheries Aquat Sci*. 69(8), 1389–1404.
- Ruddiman, W. F. (2013). The Anthropocene. *Ann Rev Earth Planet Sci*, 41(1), 45–68.
- Sato, S., Nanjappa, D., Dorrell, R. G., Vieira, F. R. J., Kazamia, E., Tirichine, L., Veluchamy, A., Heilig, R., Aury, J.-M., Jaillon, O., Wincker, P., Fussy, Z., Obornik, M., Muñoz-Gómez, S. A., Mann, D. G., Bowler, C., & Zingone, A. (2020). Genome-enabled phylogenetic and functional reconstruction of an araphid pennate diatom *Plagiostrata* sp. CCMP470, previously assigned as a radial centric diatom, and its bacterial commensal. *Sci Rep*, 10(1), 9449.
- Sauterey, B., Gland, G. L., Cermeño, P., Aumont, O., Lévy, M., & Vallina, S. M. (2023). Phytoplankton adaptive resilience to climate change collapses in case of extreme events – A modeling study. *Ecol Modelling*, 483(110437), 110437.
- Schaum, C.-E., Barton, S., Bestion, E., Buckling, A., Garcia-Carreras, B., Lopez, P., Lowe, C., Pawar, S., Smirnov, N., Trimmer, M., & Yvon-Durocher, G. (2017). Adaptation of phytoplankton to a decade of experimental warming linked to increased photosynthesis. *Nat Ecol Evol*, 1(4), 94.
- Schmidt, L. E., & Hansen, P. J. (2001). Allelopathy in the prymnesiophyte *Chrysochromulina polylepis*: effect of cell concentration, growth phase and pH. *Mar Ecol Prog Series*, 216, 67–81.
- Schulze, P. S. C., Hulatt, C. J., Morales-Sánchez, D., Wijffels, R. H., & Kiron, V. (2019). Fatty acids and proteins from marine cold adapted microalgae for biotechnology. *Algal Res*, 42(101604), 101604.
- Seifert, M., Nissen, C., Rost, B., Vogt, M., Völker, C., & Hauck, J. (2023). Interaction matters: bottom-up driver interdependencies alter the projected response of phytoplankton communities to climate change. *Glob Change Biol*, 29(15), 4234–4258.
- Seoane, M., Conde-Pérez, K., Esperanza, M., Cid, Á., & Rioboo, C. (2023). Unravelling joint cytotoxicity of ibuprofen and oxytetracycline on *Chlamydomonas reinhardtii* using a programmed cell death-related biomarkers panel. *Aquat Toxicol*, 257, 106455.
- Shaw, J. A., & Bourret, T. J. (2019). Digital PCR-based competitive index for high-throughput analysis of fitness in *Salmonella*. *J Vis Exp*, 147, 59630.
- Sheward, R. M., Liefer, J. D., Irwin, A. J., & Finkel, Z. V. (2023). Elemental stoichiometry of the key calcifying marine phytoplankton *Emiliania huxleyi* under ocean climate change: a meta-analysis. *Glob Change Biol*, 29(15), 4259–4278.
- Siswanto, E., Horii, T., Iskandar, I., Gaol, J. L., Setiawan, R. Y., & Susanto, R. D. (2020). Impacts of

- climate changes on the phytoplankton biomass of the Indonesian Maritime Continent. *J Mar Syst*, 212(103451), 103451.
- Skau, L. F., Andersen, T., Thrane, J.-E., & Hessen, D. O. (2017). Growth, stoichiometry and cell size; temperature and nutrient responses in haptophytes. *PeerJ*, 5, e3743.
- Skeffington, A. W., Fischer, A., Sviben, S., Brzezinka, M., Górka, M., Bertinetti, L., Woehle, C., Huettel, B., Graf, A., & Scheffel, A. (2023). A joint proteomic and genomic investigation provides insights into the mechanism of calcification in coccolithophores. *Nat Commun*, 14(1), 3749.
- Skeffington, A. W., Grimm, A., Schönefeld, S., Petersen, K., & Scheffel, A. (2020). An efficient method for the plating of haploid and diploid *Emiliania huxleyi* on solid medium. *J Phycol*, 56(1), 238–242.
- Smith, S. R., Dupont, C. L., McCarthy, J. K., Brodrick, J. T., Oborník, M., Horák, A., Füßy, Z., Cihlár, J., Kleessen, S., Zheng, H., McCrow, J. P., Hixson, K. K., Araújo, W. L., Nunes-Nesi, A., Fernie, A., Nikoloski, Z., Palsson, B. O., & Allen, A. E. (2019). Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. *Nat Commun*, 10(1), 4552.
- Smith, W. O., Jr, & Trimborn, S. (2024). Phaeocystis: A Global Enigma. *Ann Rev Mar Sci*, 16, 417–441.
- Smyth, T. J., Tyrrell, T., & Tarrant, B. (2004). Time series of coccolithophore activity in the Barents Sea, from twenty years of satellite imagery. *Geophys Res Lett*, 31(11).
- Søgaard, D. H., Sorrell, B. K., Sejr, M. K., Andersen, P., Rysgaard, S., Hansen, P. J., Skyttä, A., Lemcke, S., & Lund-Hansen, L. C. (2021). An under-ice bloom of mixotrophic haptophytes in low nutrient and freshwater-influenced Arctic waters. *Sci Rep*, 11(1), 2915.
- Sommeria-Klein, G., Watteaux, R., Ibarbalz, F. M., Pierella Karlusich, J. J., Iudicone, D., Bowler, C., & Morlon, H. (2021). Global drivers of eukaryotic plankton biogeography in the sunlit ocean. *Science*, 374(6567), 594–599.
- Song, H., He, M., Wu, C., Gu, C., & Wang, C. (2020). Global transcriptomic analysis of an Arctic *Chlorella*-Arc reveals its eurythermal adaptivity mechanisms. *Algal Res*, 46(101792), 101792.
- Song, Q., Xue, Y., Zhang, Y., Yin, J., & Shen, P. (2023). Diversity and geographical distribution of haptophyte *Phaeocystis* in the Chinese seas based on metabarcoding analysis. *J Oceanol Limnol*, 41(6), 2197–2207.
- Soulié, T., Vidussi, F., Mas, S., & Mostajir, B. (2023). Functional and structural responses of plankton communities toward consecutive experimental heatwaves in Mediterranean coastal waters. *Sci Rep*, 13(1), 8050.
- Steindorff, A. S., Carver, A., Calhoun, S., Stillman, K., Liu, H., Lipzen, A., He, G., Yan, M., Pangilinan, J., LaButti, K., Ng, V., Bruns, T. D., & Grigoriev, I. V. (2021). Comparative genomics of pyrophilous fungi reveals a link between fire events and developmental genes. *Env Microbiol*, 23(1), 99–109.
- Stewart, A., Rioux, D., Boyer, F., Gielly, L., Pompanon, F., Saillard, A., Thuiller, W., Valay, J.-G., Maréchal, E., & Coissac, E. (2021). Altitudinal zonation of green algae biodiversity in the French Alps. *Front Plant Sci*, 12, 679428.
- Suikkanen, S., Pulina, S., Engström-Öst, J., Lehtiniemi, M., Lehtinen, S., & Brutemark, A. (2013). Climate change and eutrophication induced shifts in northern summer plankton communities. *PLoS One*, 8(6), e66475.
- Sukhanova, I. N., Flint, M. V., Druzhkova, E. I., Sazhin, A. F., & Sergeeva, V. M. (2015). Phytoplankton in the Northwestern Kara Sea. *Oceanol*, 55(4), 547–560.
- Sun, P., Liao, Y., Wang, Y., Yang, E.-J., Jiao, N., Lee, Y., Jung, J., Cho, K.-H., Moon, J.-K., & Xu, D. (2022). Contrasting community composition and co-occurrence relationships of the active pico-sized haptophytes in the surface and subsurface chlorophyll maximum layers of the Arctic Ocean in summer. *Microorganisms*, 10(2), 248.
- Táborá-Sarmiento, S., Patiño, R., Portillo-Quintero, C., & Coldren, C. (2022). Air, land, and water variables associated with the first appearance and current spatial distribution of toxic *Prymnesium parvum* blooms in reservoirs of the Southern Great Plains, USA. *Sci Total Env*, 836, 155567.
- Teano, G., Concia, L., Wolff, L., Carron, L., Biocanin, I., Adamusová, K., Fojtová, M., Bourge, M., Kramdi, A., Colot, V., Grossniklaus, U., Bowler, C., Baroux, C., Carbone, A., Probst, A. V., Schrupfová, P. P., Fajkus, J., Amiard, S., Grob, S., ... Barneche, F. (2023). Histone H1 protects telomeric repeats from H3K27me3 invasion in *Arabidopsis*. *Cell Rep*, 42(8), 112894.

- Theroux, S., Huang, Y., Toney, J. L., Andersen, R., Nyren, P., Bohn, R., Salacup, J., Murphy, L., & Amaral-Zettler, L. (2020). Successional blooms of alkenone-producing haptophytes in Lake George, North Dakota: implications for continental paleoclimate reconstructions. *Limnol Oceanograph* 65(2), 413–425.
- Thomas, M. K., Kremer, C. T., Klausmeier, C. A., & Litchman, E. (2012). A global pattern of thermal adaptation in marine phytoplankton. *Science*, 338(6110), 1085–1088.
- Thomsen, H. A., Østergaard, J. B., & Hansen, L. E. (1991). Heteromorphic life histories in Arctic coccolithophorids (Prymnesiophyceae). *Jo Phycol*, 27(5), 634–642.
- Tortell, P. D., DiTullio, G. R., Sigman, D. M., & Morel, F. M. M. (2002). CO₂ effects on taxonomic composition and nutrient utilization in an equatorial Pacific phytoplankton assemblage. *Mar Ecol Prog Series*, 236, 37–43.
- Trainer, V. L., Moore, S. K., Hallegraeff, G., Kudela, R. M., Clement, A., Mardones, J. I., & Cochlan, W. P. (2020). Pelagic harmful algal blooms and climate change: lessons from nature's experiments with extremes. *Harmful Algae*, 91, 101591.
- Treves, H., Küken, A., Arrivault, S., Ishihara, H., Hoppe, I., Erban, A., Höhne, M., Moraes, T. A., Kopka, J., Szymanski, J., Nikoloski, Z., & Stitt, M. (2022). Carbon flux through photosynthesis and central carbon metabolism show distinct patterns between algae, C and C plants. *Nat Plants*, 8(1), 78–91.
- Trubovitz, S., Lazarus, D., Renaudie, J., & Noble, P. J. (2020). Marine plankton show threshold extinction response to Neogene climate change. *Nat Commun*, 11(1), 5069.
- Tsuji, Y., & Yoshida, M. (2017). Biology of haptophytes: complicated cellular processes driving the global carbon cycle. In *Adv Bot Res*, 84, 219–261.
- Tucker, A. T., Leonard, S. P., DuBois, C. D., Knauf, G. A., Cunningham, A. L., Wilke, C. O., Trent, M. S., & Davies, B. W. (2018). Discovery of next-generation antimicrobials through bacterial self-screening of surface-displayed peptide libraries. *Cell*, 172(3), 618–628.e13.
- Valledor, L., Furuhashi, T., Hanak, A.-M., & Weckwerth, W. (2013). Systemic cold stress adaptation of *Chlamydomonas reinhardtii*. *Mol Cell Proteom*, 12(8), 2032–2047.
- Vance, T. D. R., Bayer-Giraldi, M., Davies, P. L., & Mangiagalli, M. (2019). Ice-binding proteins and the “Domain of Unknown Function” 3494 family. *FEBS J*, 286(5), 855–873.
- Vaulot, D., Gall, F., Le Marie, D., Guillou, L., & Partensky, F. (2004). The Roscoff Culture Collection (RCC): a collection dedicated to marine picoplankton. *Nova Hedwigia*, 79(1-2), 49–70.
- Velmurugan, N., & Deka, D. (2018). Transformation techniques for metabolic engineering of diatoms and haptophytes: current state and prospects. *Appl Microbiol Biotechnol*, 102(10), 4255–4267.
- Véron, B., Rougier, E., Taylor, A., & Goux, D. (2023). New species of Pavlovophyceae (Haptophyta) and revision of the genera *Exanthemachrysis*, *Rebecca* and *Pavlova*. *Eur J Taxon*, 861, 21–47.
- Wang, J.-X., Kong, F.-Z., Geng, H.-X., Zhang, Q.-C., Yuan, Y.-Q., & Yu, R.-C. (2021). CHEMTAX analysis of phytoplankton assemblages revealed potential indicators for blooms of haptophyte *Phaeocystis globosa*. *Ecol Ind*, 131(108177), 108177.
- Wang, L., Patena, W., Van Baalen, K. A., Xie, Y., Singer, E. R., Gavrilenko, S., Warren-Williams, M., Han, L., Harrigan, H. R., Hartz, L. D., Chen, V., Ton, V. T. N. P., Kyin, S., Shwe, H. H., Cahn, M. H., Wilson, A. T., Onishi, M., Hu, J., Schnell, D. J., ... Jonikas, M. C. (2023). A chloroplast protein atlas reveals punctate structures and spatial organization of biosynthetic pathways. *Cell*, 186(16), 3499–3518.e14.
- Wells, M. L., Trainer, V. L., Smayda, T. J., Karlson, B. S. O., Trick, C. G., Kudela, R. M., Ishikawa, A., Bernard, S., Wulff, A., Anderson, D. M., & Cochlan, W. P. (2015). Harmful algal blooms and climate change: learning from the past and present to forecast the future. *Harmful Algae*, 49, 68–93.
- Wilken, S., Huisman, J., Naus-Wiezer, S., & Van Donk, E. (2013). Mixotrophic organisms become more heterotrophic with rising temperature. *Ecology Letters*, 16(2), 225–233.
- Winder, M., & Sommer, U. (2012). Phytoplankton response to a changing climate. *Hydrobiologia*, 698(1), 5–16.
- Wutkowska, M., Vader, A., Logares, R., Pelletier, E., & Gabrielsen, T. M. (2023). Linking extreme seasonality and gene expression in Arctic marine protists. *Sci Rep*, 13(1), 14627.
- Xi, H., Bretagnon, M., Losa, S., Brotas, V., Gomes, M., Peeken, I., Alvarado, L. M. A., Mangin, A., & Bracher, A. (2023). Satellite monitoring of surface phytoplankton functional types in the Atlantic Ocean over 20 years (2002–2021). *State of the Planet*, 1-osr7(5).

- Yang, H., Huang, X., Thompson, J. R., & Flower, R. J. (2014). China's soil pollution: urban brownfields. *Science*, *344*(6185), 691–692.
- Yazaki, E., Yabuki, A., Imaizumi, A., Kume, K., Hashimoto, T., & Inagaki, Y. (2022). The closest lineage of Archaeplastida is revealed by phylogenomics analyses that include *Microheliella maris*. *Open Biol*, *12*(4), 210376.
- Zhang, X., Zhang, Y., Chen, Z., Gu, P., Li, X., & Wang, G. (2023). Exploring cell aggregation as a defense strategy against perchlorate stress in *Chlamydomonas reinhardtii* through multi-omics analysis. *Sci Tot Env*, *905*, 167045.
- Zhao, J., Fan, D., Guo, W., Wu, J., Zhang, X., Zhuang, X., & Kong, W. (2023). Precipitation drives soil protist diversity and community structure in dry grasslands. *Microb Ecol*, *86*(4), 2293–2304.

NAME: Richard George Dorrell

ORCID: 0000-0001-6263-9115

Scholar: <https://scholar.google.com/citations?user=6OCmXHEAAAAJ&hl=fr>

h-index: 22

POSITION TITLE: Group Leader, Laboratory of Computational and Quantitative Biology (LCQB), Sorbonne Université

<https://www.ibps.sorbonne-universite.fr/fr/Recherche/umr-7238/evolution-horizontale-des-modes-d-e-vie-des-algues>

EDUCATION/TRAINING

INSTITUTION	DEGREE/ POSITION	Completion Date	FIELD OF STUDY
University of Cambridge, UK	BA (First Class Hons)	06/2010	Natural Sciences
University of Cambridge, UK	PhD	06/2014	Biochemistry
IBENS, Paris, France	Postdoc (Mairie de Paris, EMBO, CNRS)	01/2024	Plant Sciences

POSITIONS AND APPOINTMENTS

Research and Teaching Positions:

2020-present Chargé de Recherche CNRS, Section 23, IBENS Paris
2013-present Lecturer, University of Cambridge (Department of Plant Sciences, L3)
2021-2023 Université PSL (module Sciences pour un Monde Durable, L2)

Current Supervision Activities:

2024-present Two additional group members, *to be recruited on ERC funding*
2023-present Victoria Powell, AI, ERC Funded
2023-present Louis Graf, Postdoc, ERC Funded
2021-present Mathias Penot-Raquin, M2/ PhD student (direction d'une première these sans HDR), IBENS Paris

Prior Supervision Activities:

2022 Fabian van Beveren, Sophia Lambert, Marc Przyszczykowski, IE, ANR funded
2020-2023 Anna Novák Vanclová, Postdoc, CNRS funded
2019-2023 Shun Liu, PhD student (co-encadrant), Univ. Paris-Saclay/ Chinese Academy of Science
2017-2018 Tomomi Nonoyama, visiting PhD student (day-to-day lab supervisor), Tokyo University of Agriculture and Technology
2013-present 3 x further visiting PhD; 6 x Masters; 15 x L2/ L3/ BTS students; IBENS, Univ. Cambridge; ENCPB Pierre Gilles de Gennes; Sup'Bio'Tech; Univ. Paris-Saclay; Université PSL

Training:

2024 Sensibilisation au Harcèlement Moral au Travail, ENS, completed
2023 Sensibilisation aux Violences Sexistes et Sexuelles, ENS, completed
2023 Formation à l'encadrement doctoral, Univ. Paris-Saclay, completed
2019 Laboratory Group Leadership, EMBO, completed

Current Funding:

2023-2027 ERC Starting Grant (value 1.48 m€), The Mosaic Chloroplast of Secondary Red Algae, PI
2022-2026 JGI CSP (value 200 k€), The Haptophyte Pan-Genome, co-PI

Prior Funding:

2022	ANR t-ERC (value 113 k€), Fluorescent Labelling of the Haptophyte Chloroplast, PI
2022	ANR JCJC (value 374 k€), The Pan-Arctic Pan-Algal Genome, PI
2019-2021	CNRS Momentum Fellowship (value 600 k€), The Marine Carbon Cycle Revisited, PI
2019-2023	ANR PRCe (value 425 k€), The Diatom Chloroplast Proteome, named collaborator

Invited presentations in international conferences (underlined indicates keynote speaker):

2024	<u>European Diatom Conference</u> , Ohrid, North Macedonia
2022	<u>Jírovec's Protozoological Days</u> , Czech Republic
2021	EFOR Network Annual Symposium
2018	<u>Oomycete Genomics Network Meeting</u> , SAMS, UK
2018	<u>ALFF Symposium</u> , SAMS
2017	International Phycological Congress, Szczecin, Poland
2016, 2019	Evolutionary Biology in Marseilles Meeting

2018-present Invited seminars: IFREMER, Nantes; Institut Jacques Monod; iEES, Sorbonne Université; University of South Bohemia, Czech Republic ; University of Ostrava, Czech Republic; University College London, UK ; Imperial College London, UK

Current collaborations and research thematics (student participation in italics) :

Giovanni Finazzi (CEA Grenoble): elucidation of diatom chloroplast-to-mitochondria metabolite transporters, *1 x co-supervised PhD student*

Alisdair Fernie (MPI Potsdam), **Juliette Jouhet** (CEA Grenoble), **Yue Liang** (Inst Oceanography Qingdao): metabolic functions of diatom chloroplast glycolysis, *1 x supervised PhD student*

Eric Maréchal (CEA Grenoble), **Marina Cvetkovska** (Univ Ottawa):molecular evolution of cryophily in Chlamydomonadales; **ANR PRC in preparation, funding sought for 1 x supervised Postdoc, 1x PhD and 2 x engineers**

Ellen Nisbet (Univ Nottingham) **Ansgar Gruber** (Univ South Bohemia): the dinoflagellate chloroplast proteome, *1 x engineer to be recruited on ERC funding*

Miroslav Obornik (Univ South Bohemia): environmental functions of diatom chlorophyll synthesis pathways, *1 x supervised visiting PhD student*

Betsy Read, Xiaoyu Zhang (UCSM) ; **Joel Dacks** (Univ Alberta) ; **Uwe John** (AWI Bremen): the haptophyte pan-genome, **JGI CSP grant, 1 x postdoc to be recruited on ERC funding**

Alastair Skeffington, (Univ Stirling) ; **Jelena Godrijan** (Ruđer Bošković Institute) ; **El Mahdi Bendif** (Université de Québec à Rimouski): molecular elucidation of heating stress resilience in natural mesocosms, **HFSP Collaborative Grant in preparation**

Review and service:

2018-present Grant reviewer: *European Research Council, UK Research and Innovation (UKRI), National Science Foundation (US), Human Frontiers Science Programme, Czech Science Foundation, DFG (Germany), Zukunftskolleg Fellowships (Germany)*

2017-2022 Associate editor: *European Journal of Phycology, Frontiers in Plant Science, Frontiers in Microbiology*

2021-present: Guest editor: *Applied Phycology (Vol 3, Issue 1; Towards a Global View of Algal Molecular Ecology; co-editors, Flora Vincent, Deo Florence Onda); Frontiers in Plant Science (Vol*

12, 2021; *Mixotrophic, Secondary Heterotrophic, and Parasitic Algae*; co-editors, Miroslav Obornik, Denis Tikhononkov)

2014-present Academic reviewer: *Cell, Science, PLoS Biology, Current Biology, PNAS, Nature Communications, ISME J, Plant J, Molecular Biology and Evolution, Communications Biology, New Phytologist, Molecular Ecology, Plant Physiology, BMC Biology*

STATEMENT ON PRIOR SUPERVISION ACTIVITIES

While I have a relatively short duration of experience as a research supervisor, many of the alumni of my group have continued to successful further positions in science and/ or academia. Concerning previously supervised PhD students and postdocs:

- Anna Novák Vanclová (postdoc, 2020-2023) is completing a second postdoc with Cathy Jackson (Institut Jacques Monod, Paris)
- Shun Liu (cosupervised PhD, 2019-2023) has found a research position with a government organisation (Guangzhou, China) focussing on marine geological surveys
- Tomomi Nonoyama (visiting PhD, 2017-2018) is on maternity leave from a permanent position in a small pharmaceutical company in Tokyo, Japan

Two of my former supervised students have now obtained their own positions. Lucas Paoli (L3, 2015) is completing a postdoc at the Institut Pasteur (Aude Bernheim team) following a PhD at ETH Zürich (Shinichi Sunagawa team) and will begin a first appointment at Université Lausanne this autumn. Elisabeth Richardson (L3 equivalent, Cambridge University, 2014) is now an associate professor at Mount Royal University (Alberta, Canada) following a PhD at the University of Alberta (Joel Dacks).

Other former students who I know to have continued to postdoctoral research or teaching include: Bridget Bannerman (visiting PhD, 2015), now a bye-fellow at Wolfson College, Cambridge; Giselle McCallum (M1, 2016), post-doc at the University of British Columbia (Carolina Tropini team); Fabian van Beveren (IE, 2022), post-doc at Université Toulouse (Pierre-Marc Delaux team); Sophia Lambert (IE, 2022), post-doc at the University of Oregon (Stilianos Louca team); and Dominic Absolon (visiting PhD, 2019), staff scientist at the Wellcome Sanger Institute.

Other supervised students are now completing PhD studies in other teams: Nikola Zarevski (Institute of Science and Technology Austria, Guet Group) and Shun-Min Yang (University of South Bohemia, Obornik Group); pursuing Masters levels studies: Setsen Altan-Ochir, Alice Bossard, Morenikedji Fafoumi, Ena Siomeo, Emily Zajac; in governmental and consultancy positions : Guillemette Audren de Kerdel, James Drew, George Hinksman; or working as research assistants : Quentin Caris; Pierre der Gazerian; Fuhai Liu; Marc Przystarczykowski.

Finally, I would like to note that three former supervised students have remained active collaborators of my group, or have returned to my group in contracted positions. Elisabeth Richardson has remained a collaborator, coauthoring four papers with my group independent to her initial undergraduate internship, and is a named collaborator on an ANR PRC currently under review with me as PI. Mathias Penot-Raquin has returned to my group as a full-time PhD student (starting September 2022), matriculated at the Université Paris-Saclay but with his own independent fellowship (CDSN, from the ENS), following his initial M2 internship in 2021. Finally, Victoria Powell now works in my group as an AI while studying for a Diplôme d'Ingenieur at the CNAM, following an initial BTS internship in 2022.

COMPLETE PUBLICATION LIST

(* indicates corresponding, underlined indicates supervised group member)

Primary publications in peer-reviewed journals

1. **R. G. Dorrell***, C. J. Howe, Functional remodeling of RNA processing in replacement chloroplasts by pathways retained from their predecessors. *Proc Natl Acad Sci USA*, **109** 18879-18884 (2012).
2. A.C. Barbrook, **R.G. Dorrell** ... C.J. Howe, Polyuridylylation and processing of transcripts from multiple gene minicircles in chloroplasts of the dinoflagellate *Amphidinium carterae*. *Plant Mol Biol* **79**, 347-357 (2012).
3. B. Imanian ... **R.G. Dorrell** ... 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).
4. C.M.M Gachon ... **R.G. Dorrell** ... J. Day, The CCAP KnowledgeBase: linking protistan and cyanobacterial biological resources with taxonomic and molecular data. *Systemat Biodiv* **11**, 407-413 (2013).
5. **R. G. Dorrell**, J. Drew ... C.J. Howe, Evolution of chloroplast transcript processing in *Plasmodium* and its chromerid algal relatives. *PLoS Genet* **10**, 1004008 (2014).
6. E.H. Richardson, **R.G. Dorrell** ... C.J. Howe, Genome-wide transcript profiling reveals the coevolution of plastid gene sequences and transcript processing pathways in the fucoxanthin dinoflagellate *Karlodinium veneficum*. *Mol Biol Evol* **31**, 237-2386 (2014).
7. **R.G. Dorrell**, G.A. Hinksman, C.J. Howe, Diversity of transcripts and transcript processing forms in plastids of the dinoflagellate alga *Karenia mikimotoi*. *Plant Mol Biol* **90**, 233-247 (2016).
8. **R.G. Dorrell*** ... E.H. Richardson ... C. Bowler. Progressive and biased divergent evolution underpins the origin and diversification of peridinin dinoflagellate plastids. *Mol Biol Evol* **34**, 361-379 (2016).
9. **R.G. Dorrell*** ... G. McCallum ... C. Bowler, Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. *ELife* **6**, 23717 (2017).
10. C.M. Klinger, L. Paoli ... **R.G. Dorrell**, E.H. Richardson, Plastid transcript editing across dinoflagellate lineages shows lineage-specific application but conserved trends. *Genom Biol Evol* **10**, 1019-1038 (2018).
11. B.P. Bannerman ... **R.G. Dorrell**, M.C. Carrington, Multispecies reconstructions uncover widespread conservation, and lineage-specific elaborations in eukaryotic mRNA metabolism. *PLoS One* **13**, 0192633 (2018).
12. A. Rastogi ... **R.G. Dorrell** ... L. Tirichine, Integrative analysis of large scale transcriptome data draws a comprehensive landscape of *Phaeodactylum tricornutum* genome and evolutionary origin of diatoms. *Sci Rep* **8**, 1-14 (2018).
13. E.J. Carpenter ... **R.G. Dorrell** ... GKS. Wong, Access to RNA-sequencing data from 1,173 plant species: The 1000 Plant transcriptomes initiative (1KP). *Gigascience* **8**, 126 (2019).
14. E. Kazamia ... **R.G. Dorrell** ... C. Bowler, Endocytosis-mediated siderophore uptake as a strategy for Fe acquisition in diatoms. *Sci Adv* **4**, 4536 (2018).
15. **R.G. Dorrell*** ... L. Paoli, G. Audren de Kerdrel ... R. Kamikawa, Principles of plastid reductive evolution illuminated by nonphotosynthetic chrysophytes. *Proc Natl Acad Sci USA* **116**, 6914-6923 (2019).
16. **R.G. Dorrell*** ... C.J. Howe, Integrated genomic and transcriptomic analysis of the peridinin dinoflagellate *Amphidinium carterae* plastid. *Protist* **170**, 358-373 (2019).
17. T. Nonoyama ... X. Gao ... **R.G. Dorrell***, Metabolic innovations underpinning the origin and diversification of the diatom chloroplast. *Biomolecules* **9**, 322 (2019).
18. O. Ait-Mohamed, A.M.G. Novák Vanclová ... **R.G. Dorrell**, PhaeoNet: a holistic RNAseq-based portrait of transcriptional coordination in the model diatom *Phaeodactylum tricornutum*. *Frontiers Plant Sci* **11**, 590949 (2020).

19. S. Sato ... **R.G. Dorrell** ... A. Zingone, 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**, 1-20.
20. **R. G. Dorrell*** ... G. Audren de Kerdrel, G. McCallum ... G. Blanc, Phylogenomic fingerprinting of tempo and functions of horizontal gene transfer within ochrophytes. *Proc Natl Acad Sci USA* **118**, 2009974118 (2020).
21. J.J. Pierella Karlusich ... **R.G. Dorrell** ... C. Bowler, A robust approach to estimate relative phytoplankton cell abundance from metagenomic data. *Mol Ecol Res* **23**, 16-40 (2023).
22. H. Bey ... S. Liu, **R.G. Dorrell** ... M. Boccara, Dynamic Cell Imaging: application to the diatom *Phaeodactylum tricornutum* under environmental stresses. *Europ J Phycol* **58**, 145-155 (2022).
23. S. Liu ... **R.G. Dorrell***. A metabolic, phylogenomic and environmental atlas of diatom plastid transporters from the model species *Phaeodactylum*. *Front Plant Sci* **3640**, 950467 (2022).
24. Z. Shao ... **R.G. Dorrell** ... D. Duan. Characterization of a marine diatom chitin synthase using a combination of meta-omics, genomics, and heterologous expression approaches, *MSystems* e01131-22 (2023).
25. **R. G. Dorrell** ... N. Zarevski, A.M.G. Novák Vanclová ... C. Lovejoy, Convergent evolution and horizontal gene transfer in Arctic Ocean microalgae. *Life Sci All* **6**, 3 (2023).
26. A. M. G. Novák Vanclová ... F. Liu ... **R.G. Dorrell***. New plastids, old proteins: repeated endosymbiotic acquisitions in kareniacean dinoflagellates. *EMBO Reports*, **in press** 10.1038/s44319-024-00103-y (2024).
27. **R. G. Dorrell*** ... S. Altan-Ochir ... A.G. Smith. Presence of vitamin B12-metabolism in the last common ancestor of land plants. *Proc R Soc Biol* **in press** 10.1098/rstb.2023.0354 (2024).

Review articles and book chapters

28. **R.G. Dorrell***, A.G. Smith., Do Red and Green Make Brown? Perspectives on plastid acquisitions within chromalveolates, *Eukaryot Cell* **10**, 856-868 (2011).
29. G. Walker, **R.G. Dorrell**, ... J.B. Dacks., Eukaryotic systematics: a user's guide for cell biologists and parasitologists, *Parasitol* **138**, 1638-1663 (2011).
30. **R.G. Dorrell***, C.J. Howe, What makes a chloroplast? Reconstructing the establishment of photosynthetic symbioses. *J. Cell Sci* **125**, 1865-1875 (2012).
31. **R.G. Dorrell*** ... C.J. Howe, Evolution: unveiling early alveolates. *Curr Biol* **23**, 1093-1096 (2013).
32. **R. G. Dorrell***, C. J. Howe, Integration of plastids with their hosts: lessons learned from dinoflagellates. *Proc Natl Acad Sci USA* **112**, 10247-10254 (2015).
33. E.H. Richardson ... **R.G. Dorrell**, J.B. Dacks, Evolutionary cell biology: functional insight from "endless forms most beautiful". *Mol Biol Cell* **26**, 4532-4538 (2015).
34. **R.G. Dorrell***, C. Bowler. Secondary plastids of stramenopiles. *Adv Bot Res* **84**, 57-103 (2017).
35. **R.G. Dorrell***. Convergence in the RNA processing of fractured algal organelle genomes. *J Phycol* **56**, 1121-1123 (2020).
36. M. Obornik, **R.G. Dorrell**, D. Tikhonenkov. Mixotrophic, secondary heterotrophic, and parasitic algae. *Front Plant Sci*, **12** (2021).
37. **R.G. Dorrell***, F. Liu, C. Bowler. Reconstructing dynamic evolutionary events in diatom nuclear and organelle genomes. In the *Molecular Life of Diatoms*, e.d. A. Falciatore and T. Mock, 147-177 (2022).
38. M. Hammond, **R.G. Dorrell** ... J. Lukeš. Eukaryotic cellular intricacies shape mitochondrial proteomic complexity. *Bioessays* **44**, 2100258 (2022).
39. M. Penot ... **R.G. Dorrell.*** Genomic and meta-genomic insights into the functions, diversity and global distribution of haptophyte algae, *Appl Phycol* **3**, 340-359 (2022).

40. J.J Pierella Karlusich ... **R.G. Dorrell**.* Modèles biogéographiques et génomes des photoautotrophes aquatiques. In *Planète bleue, photosynthèse rouge et verte: Productivité et cycle du carbone dans les écosystèmes marins*, e.d. B.A. Gontero and S. Maberly, 1-45 (2023).

Manuscripts under review

41. **R.G. Dorrell*** ... M. Penot ... C. Bowler. Complementary environmental data analysis and functional characterization of a plastid diatom lower-half glycolytic-gluconeogenesis pathway. *Plant Cell*, **resubmission**, *BioRxiv* 10.1101/2022.09.08.507166v3 (2024)
42. A.J. Perrin, **R.G. Dorrell**. Protists and Protistology in a climate and ecological crisis. *BMC Biology*, **under review** (2024)