



Biogeographical Patterns and Genomes of Aquatic Photoautotrophs

Juan José Pierella Karlusich, Charlotte Nef, Chris Bowler, Richard G Dorrell

► To cite this version:

Juan José Pierella Karlusich, Charlotte Nef, Chris Bowler, Richard G Dorrell. Biogeographical Patterns and Genomes of Aquatic Photoautotrophs. Stephen Christopher Maberly; Brigitte Gontero. Blue Planet, Red and Green Photosynthesis: Productivity and Carbon Cycling in Aquatic Ecosystems, Wiley, 2022, 9781789450828. 10.1002/9781119986782.ch3 . hal-03942752

HAL Id: hal-03942752

<https://cnrs.hal.science/hal-03942752>

Submitted on 17 Jan 2023

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

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

Chapter 5: Biogeographical patterns and genomes of aquatic photoautotrophs

Juan José Pierella Karlusich, Charlotte Nef, Chris Bowler and Richard G. Dorrell

Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Ecole Normale Supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France

Introduction- The Changing Face of Algal Genomes

Almost half of photosynthesis on Earth occurs in marine and freshwater environments, hence aquatic photosynthetic organisms have an enormous impact on food webs and on the biogeochemistry of the entire Earth System. Photosynthesis in marine and freshwater habitats is supported by cyanobacteria and a diverse range of eukaryotic algae that acquired photosynthesis through the endosymbiotic uptake of chloroplasts (Raven, Chapter 3). These eukaryotic algae are highly important for global environmental processes, accounting for over one-third of planetary primary production (Dorrell and Smith 2011), and may be particularly sensitive to environmental change (Li et al. 2009; Hinder et al. 2012). Eukaryotic algae, however, comprise huge functional diversity, ranging from tiny single-celled organisms of less than one micrometre diameter to giant seaweeds many hundreds of metres in length; and including obligate phototrophs, predators, and symbionts of other organisms (**Fig. 1**). The phylogenetic relationships and evolutionary histories of these groups are also extremely complex; with different algal groups separated by one another, and from plants, by hundreds of millions of years of evolution, and multiple horizontal and endosymbiotic gene exchanges (**Fig. 1**) (Dorrell and Smith 2011; Sibbald and Archibald 2017; Novák Vanclová et al. 2020). The diversity of eukaryotic algae renders understanding their biology extremely challenging; dependent on resolving the genomic content of multiple, distantly related lineages, even before this information can be placed into a functional or environmental context.

The advent of next-generation sequencing has revolutionised our understanding of microbial evolution, physiology and ecology. Within the context of microalgae, this commenced with the completion of chloroplast (Kowallik et al. 1995; Reith and Munholland 1995) and mitochondrial genomes (Wolff et al. 1994; Ohta, Sato, and Kuroiwa 1998) in the 1990s, followed by the complete nuclear genomes of the red alga *Cyanidioschyzon merolae* and the diatom *Thalassiosira pseudonana*, and the green algae *Chlamydomonas reinhardtii* and *Ostreococcus tauri* in the early 2000s (Armbrust et al. 2004; Matsuzaki et al. 2004; Derelle et al. 2006; Merchant et al. 2007). At the date of writing this publication, well over a hundred eukaryotic algal genomes have been completed (Grigoriev et al. 2020; Hanschen and Starkenburg 2020; Nelson et al. 2021). Alongside these, hundreds of transcriptome libraries, that is to say a survey of expressed genomic sequences through high-throughput sequencing of cellular RNA, have been completed for eukaryotic algae (**Fig. 2**), providing unprecedented insight into the macro- and micro-diversity, and physiological genetics of marine photoautotrophs.

Despite the impressive scale of sequence information assembled in the last three decades, the genomes and transcriptomes of cultured algae may provide somewhat limited insights into the functional biology of algae in their native environment. Many of the algal species in the world's ocean rest uncultured, and may remain unculturable (Mangot et al. 2017; Absolon, Smith, and Helliwell 2019). Here, environmental data, exemplified by ambitious projects such as the BioMarKs, Malaspina and Tara Oceans expeditions

(Karsenti et al. 2011; Dunthorn et al. 2014; Fernández-Castro et al. 2014; Bork et al. 2015) can provide valuable insights into this microbial dark matter, which can then be correlated to spatial information and environmental parameters. In this chapter, we provide an overview of the main groups of aquatic photoautotrophs; exploring their complex evolutionary histories, and the diversity and limitations of genome and transcriptome data from cultured species for understanding algal biology. Subsequently, we consider how meta-genomic and meta-transcriptomic data generated by the *Tara* Oceans expeditions (Carradec et al. 2018; Pierella Karlusich, Ibarbalz, and Bowler 2020a) has unveiled patterns of algal biogeography and mechanisms of adaptation and acclimation to the marine environment, changing our understanding of how the dominant groups of algae function in the open ocean.

1. Diversity of Algae and their Chloroplasts

The advent of genome sequencing has provided substantial insights into the taxonomy and diversity of eukaryotic algae (**Fig. 1**). The earliest-evolving eukaryotic algal groups are found within the archaeplastids (glaucophytes and green and red algae), which are unified by the presence of chloroplasts surrounded by two membranes. Sequencing of archaeplastid chloroplast genomes confirmed the long-established hypothesis that they derived from the primary endosymbiosis of a free-living beta-cyanobacterium (Margulis 1975; Reith and Munholland 1995).

Archaeplastids notably include multicellular red algae that form important crop species (e.g., *Pyropia yezoensis*, or nori seaweed; the model single-celled freshwater alga *Chlamydomonas* within the chlorophyte lineage; and land plants within the streptophyte lineage, both in the green algae (Merchant et al. 2006; Umen 2014; Skokan et al. 2019; Wang et al. 2020; **Fig. 1**). The most abundant of the archaeplastid groups in the contemporary ocean are the prasinophytes, tiny single-celled chlorophytes considered particularly to dominate photosynthetic biomass and production in oligotrophic open ocean systems. Well studied genera include *Ostreococcus*, considered to be the smallest (ca. 0.95 μm) free-living eukaryote (Derelle et al. 2006), and *Micromonas*, both of which are found worldwide.

The dominant eukaryotic algal groups in the oceans today, however, are distantly related to the archaeplastids; including groups such as cryptomonads, haptophytes, dinoflagellates, and the various algal groups that comprise the ochrophytes (**Fig. 1**). The majority of these algae contain chloroplasts that are red or brown in colour (dependent on the accessory pigments they contain; typically phycobilisomes in red-coloured cryptomonads, and fucoxanthin in brown-coloured dinoflagellates, haptophytes and ochrophytes) and are surrounded by three or four membranes, as compared to the two found in archaeplastids (Dorrell and Smith 2011). Consideration of the membrane ultrastructure of these chloroplasts (Cavalier-Smith 2003), alongside phylogenetic analysis of their genomes (Kowallik et al. 1995), suggests that they have arisen through the secondary or higher (e.g., tertiary) endosymbioses of eukaryotic algae, as opposed to the primary endosymbiosis underpinning archaeplastid chloroplasts; and that the majority of algal groups in the ocean today possess chloroplasts of red algal origin. The secondary red chloroplasts are closely related to one another, although their underlying hosts are not (Dorrell and Smith 2011; **Fig. 1**).

Specific groups of algae with secondary red chloroplasts have risen to positions of particular ecological importance. The diatoms, for example, which are members of the ochrophytes, are of unmatched

dominance in marine ecosystems, probably performing well over one-fifth of total planetary photosynthesis (Bowler, Vardi, and Allen 2010; Dorrell and Smith 2011). The characteristic highly elaborate siliceous cell walls of diatoms, known as frustules, are easily recognizable and are well preserved in the fossil record, and many species form chains of interlocking cells which are also easily recognizable (**Fig. 1**). Diatoms tend to thrive in nutrient-rich turbulent systems such as those found in coastal upwelling regions and at high latitudes, and their life histories are often characterized by rapid proliferations, followed by die-offs, in spring and autumn following nutrient injection events (Malviya et al. 2016; Tréguer et al. 2018).

The haptophytes are mainly composed of small planktonic unicellular species, occurring either as single cells or colonies (**Fig. 1**). This group includes the coccolithophores, which are covered by small regular calcareous plates (coccoliths) and are extremely important in biogeochemical cycles because they are responsible for about half of all modern precipitation of CaCO_3 in the ocean, forming the basis of calcareous rocks such as limestone (Milliman 1993). The most well-known representative is the coccolithophore *Emiliania huxleyi*, which is the most numerically abundant and widespread coccolithophore species, and can form giant blooms visible from space (Tyrrell and Merico 2004; Durak, Brownlee, and Wheeler 2017).

The dinoflagellates are a species rich group, with taxa inhabiting both planktonic and benthic environments, and with lifestyles ranging from photoautotrophy to heterotrophy, as well as symbionts and parasites (Steidinger and Jangen 1997; Simon et al. 2009). Some species can form highly resistant spores that are preserved in the geological record. Many others are endosymbionts of marine animals, the most well-known being *Symbiodinium*, the zooxanthellae of reef-building corals; while others can form toxic algal blooms, such as the red tide-causing *Alexandrium* (Cumbo et al. 2013; González-Pech et al. 2019). Dinoflagellates are known for their divergent and bizarre evolutionary trajectories, for example not using histones for packaging their nuclear DNA, in contrast to all other known eukaryotic lineages; and possessing dramatically expanded nuclear genomes and dramatically shrunken chloroplast and mitochondrial genomes (Dorrell and Howe 2015; Gornik et al. 2019).

Alongside these three major groups are many others; for example within the ochrophytes the pelagophytes (typified by the harmful algal bloom-causing *Aureococcus anophagefferens*), chrysophytes (golden algae, e.g. *Ochromonas*) and phaeophytes (containing giant kelps); the cryptomonads and chromerids, which also contain red-derived chloroplasts (Cumbo et al. 2013); and euglenids and chlorarachniophytes (**Fig. 1**). The euglenids and chlorarachniophytes differ from the other groups described, possessing chloroplasts derived from the secondary endosymbiosis of green algae, although analysis of euglenid and chlorarachniophyte chloroplast genomes indicate that they are distantly related to one another on the green algal tree (Kamikawa et al. 2015; Jackson et al. 2018).

Finally, alongside other aspects of their unusual evolutionary history, dinoflagellates have acquired chloroplasts on multiple occasions, from multiple sources. These include the red-derived chloroplasts found in most dinoflagellate species, but also alternatives derived from green algae, and explicitly from the tertiary endosymbiosis of diatoms and haptophytes (themselves containing secondary, red-type chloroplasts) (Dorrell and Howe 2015; Sarai et al. 2020). This extreme promiscuity in chloroplast diversity is the result of rampant serial endosymbiosis, in which the original dinoflagellate chloroplast has been lost and replaced multiple times; and may be linked to the extreme propensity of dinoflagellates for

phagotrophy and photo-mixotrophic life strategies (Dorrell and Howe 2015; Sarai et al. 2020; Cohen et al. 2021; Jeong et al. 2021) (**Fig. 1**).

2. Genomic Insights into Algal Evolution

The annotation of eukaryotic algal nuclear genome sequences, particularly within a phylogenetic context, has yielded an even more complex story. Large numbers of genes, which may be the footprints of endosymbiont-derived genes transferred to the nucleus (Henze, Martin, and Schnarrenberger 2002; Bock 2017), phylogenetically link together different groups of algae with secondary chloroplasts, suggesting that several of them have in fact originated via sequential secondary, tertiary and even quaternary endosymbiosis (Stiller et al. 2014; Dorrell et al. 2017). For example, different phylogenetic investigations of algal genomes have suggested that the chloroplasts of ochrophytes originate from cryptomonad algae (Stiller et al. 2014); the chloroplasts of haptophytes may originate from pelagophytes within the ochrophytes (Dorrell et al. 2017; Dorrell et al. 2021); the chloroplasts of dinoflagellates may originate from haptophytes (Yoon, Hackett, and Bhattacharya 2002; Dorrell et al. 2021); and the chloroplasts of chromerids may originate from chrysophytes within the ochrophytes (Ševčíková et al. 2015).

In addition, nuclear genome sequences have revealed cryptic genetic signals in many eukaryotic algae that are not related to their current chloroplasts. For instance, many of the algae with chloroplasts of purported red algal origin (e.g., cryptomonads, haptophytes, and ochrophytes) contain large numbers of genes of green algal origin (Moustafa et al. 2009; Dorrell et al. 2017); whereas chlorarachniophytes and euglenids, which possess chloroplasts derived from the secondary endosymbioses of green algae, also contain genes derived from red algae, or their endosymbiotic progeny (Ponce-Toledo et al. 2018; Novák Vanclová et al. 2020). Many of these genes encode chloroplast-associated proteins, suggesting that their origin is in some way connected to chloroplast endosymbiosis (Dorrell et al. 2017); although it is not known whether these signals are the relics of ancient endosymbioses that have subsequently been replaced, similarly to the serial endosymbiotic replacements known in dinoflagellates (Dorrell and Howe 2015; Sarai et al. 2020); or may have other origins entirely, e.g., the endosymbiotic acquisition of algae that themselves contained mixtures of red and green genes (Dorrell et al. 2017; Ponce-Toledo et al. 2018).

Alongside endosymbiotic signals, algal nuclear genomes can contain large numbers of genes derived from non-endosymbiotic horizontal gene transfer events, which are known to occur between eukaryotic algae (Raymond and Kim 2012; Kazamia et al. 2018); and from bacterial and viral donors (Rossoni et al. 2019; Vancaester et al. 2020; Dorrell et al. 2021; Nelson et al. 2021). These horizontally-acquired genes may dramatically affect the functions of eukaryotic algae in the environment. For example, inspection of diatom genomes has identified unique combinations of genes of bacterial and eukaryotic origin that collectively encode non-canonical pathways of nutrient assimilation and metabolite management (Wilhelm 2006), including a urea cycle that is integral to nitrogen metabolism (Allen et al. 2011; Smith et al. 2019; Dorrell et al. 2021), and high-affinity iron uptake systems not known in other eukaryotic groups (Kazamia et al. 2018; McQuaid et al. 2018).

Finally, the genomes of eukaryotic algae may also undergo complexification and simplification events pertaining to lifestyle transitions. Diatom genomes, for example, have been found to encode expanded sets of cyclins (Huysman et al. 2010), key regulators of cell division, as well as light-harvesting

chlorophyll binding proteins (Bailleul et al. 2010; Mock et al. 2017) that may be important in, respectively, driving proliferation and stress responses in the ocean environment. In contrast, many lineages of eukaryotic algae have secondarily reverted from photoautotrophic to exclusively heterotrophic lifestyles, accompanied by the loss of large numbers of genes associated with photosynthesis. This is particularly notable in the chrysophytes, within the ochrophytes; as well as the dinoflagellates, in which fewer than half of all documented species are photosynthetic (Sournia 1986; Gomez 2005; Simon et al. 2009; Dorrell et al. 2019).

3. Limitations of Cultured Algal Sequencing Projects

In **Figs 2** and **3**, we present the current (as of end-2020) phylogenetic and biogeographical diversity of available algal nuclear genomes and transcriptomes; based on the manually verified isolation sites of the source cultures sequenced (Vaulot et al. 2004; Gachon et al. 2013; Guiry et al. 2014; Boundy-Mills et al. 2015; Nelson et al. 2021). The greatest numbers of genomes have been sequenced from green algae and diatoms (**Fig. 2**), reflecting their interest as models for plant evolution (in the case of green algae), aquaculture (both lineages) and ecological prominence in the contemporary ocean (diatoms in particular) (Bowler, Vardi, and Allen 2010; Kumar, Singh, and Sharma 2017; One Thousand Plant Transcriptomes Initiative 2019); however, multiple genome sequences have now been completed for many other algal groups, including haptophytes, pelagophytes and dinoflagellates (Grigoriev et al. 2020; Nelson et al. 2021).

Alongside this genomic diversity, hundreds of algal transcriptomes have been completed, either through individual projects or consortium efforts, as described above (Keeling et al. 2014; One Thousand Plant Transcriptomes Initiative 2019). Transcriptome libraries carry some major advantages over genome assemblies, not least that coding sequences can be easily annotated without considering introns (which can be large and difficult to infer in some major algal groups, e.g., dinoflagellates (Gornik et al. 2019)); and that preliminary insights can be made into gene expression patterns. However, transcriptomes cannot provide insights into non-coding DNA, or genes typically removed by poly(A) selection (e.g., organelle transcripts, Smith and Sanitá Lima 2017); and may be more heavily impacted by biological contamination than genomes, in the absence of a possibility to assemble gene sequences into contigs (Carpenter et al. 2019). Nonetheless, transcriptomes used in partnership with genome sequences can radically expand our breadth of understanding of algal genome diversity, and allow exploration of algal lineages (e.g., dictyochophytes or silicoflagellates, bolidophytes; **Fig. 2**) and biogeographical regions (the Arctic and Pacific Oceans, **Fig. 3**) for which reference algal genomes remain in preparation.

Despite this vast amount of resources, some biases and limitations are still visible. For example, the overwhelming majority of complete algal genomes and transcriptomes are from species isolated from North American and European oceanic regions (**Fig. 3**), and many are from coastal and terrestrial environments, while open-ocean algae, which may be less easily accessible, are under-represented (Lommer et al. 2012) (**Fig. 3**). Moreover, many of the algal species for which we have genome sequences, even if biologically interesting, may not be particularly representative of those that grow in the oceans. For example, the first sequenced red alga, *C. merolae*, was an extremophile isolated from acidic hot sulphur springs, with a growth temperature (up to 56 °C) and pH (as low as 1.5) very different to most marine red algal relatives (Matsuzaki et al. 2004; Brawley et al. 2017), and may accordingly have very

different core metabolic pathways to other red algal species (e.g., Helliwell et al. 2011; Rossoni et al. 2019).

Even within well-sampled oceanic regions, there may be limitations in what genomic information is available. Comparisons between observational (e.g., microscopy) and culturable data under a variety of environments have long pointed to the presence of an uncultured majority of marine micro-organisms (Razumov 1932; Jannasch and Jones 1959; Zimmermann, Iturriaga, and Becker-Birck 1978); with only between 0.01 and 1% microbial taxa being cultivable (Amann, Ludwig, and Schleifer 1995; Rappé and Giovannoni 2003). Within the context of algae several major groups of organisms have been identified from environmental sampling, such as MOCHs, within the ochrophytes (Massana et al., 2014) and raphemonads, which are related to haptophytes (Kim et al. 2011); that have largely eluded successful culture in laboratory conditions (Absolon, Smith and Helliwell 2019).

Finally, even the laboratory conditions in which algal genomes and transcriptomes are realised may be highly non-naturalistic (**Fig. 4**). Laboratory growth chambers may be unable to approach the extremely low environmental temperatures encountered in polar oceanic environments (Raymond and Kim 2012; Terrado et al. 2015); and the growth media necessary to facilitate dense cultures for sequencing may contain nutrient concentrations many hundreds of times those experienced by algae in the open ocean (**Fig. 4**), although admittedly the concentrations of these nutrients that are biologically available to algae in the lab and in the wild remain poorly understood (Lommer et al. 2012; Kazamia et al. 2018). Reference strains may have been cultured for many years or decades under replete conditions in the laboratory since isolation, and the somatic mutations that have occurred naturally to algae during their time in laboratory culture may impact on our understanding of their physiology (Schaum and Collins 2014; Flowers et al. 2015; Rastogi et al. 2020). These biases, while necessary for the generation of algal sequence libraries in the first place, may limit our understanding of which genes truly underpin inducible responses to environmental stresses, or facilitate the acquisition of growth-limiting nutrients in the ocean (Browning et al. 2017).

4. History of Omics-Based Approaches Applied to Environmental Plankton Samples

As discussed above, the application of next-generation sequencing to cultured algae has allowed us to understand algal diversity more accurately, but with less insight into functional biology in the environment. ‘Omics data generated directly from the natural environment may provide an interesting perspective, offering genomic contexts for taxa not represented in laboratory collections, and direct insights into organisms’ experiences in the wild.

Community assessment of plankton samples has occurred for over 100 years (Karlusich, Ibarbalz, and Bowler 2020b), starting with data collection from the Western Channel Observatory, and the Continuous Plankton Recorder in the early 20th century (Reid et al. 2003; Southward et al. 2005); yielding large amounts of microscopic abundance and diversity data, that may even be transformable into genomic information via sequencing of preserved samples (Stern et al. 2018). Community assessment of plankton samples based on DNA sequencing itself started in the 1990s (Giovannoni et al. 1990; Fuhrman, McCallum, and Davis 1995), but bloomed with the advent of high-throughput sequencing in the 2000s. A global-scale metagenomic sampling of the ocean began with J. Craig Venter’s Global Ocean Sampling

expeditions between 2003 and 2008, which included the collection of bacterial size fractions from surface waters from the North–West Atlantic and Eastern Tropical Pacific Oceans from 2004 to 2006, generating a 6.1 million gene set using Sanger sequencing (Rusch et al. 2007). Later, in the context of the Census of Marine Life (<http://www.coml.org/>), a consortium of researchers deployed the International Census of Marine Microbes (Amaral-Zettler et al. 2010) to provide a detailed inventory of marine microbial diversity. Further projects of large spatial coverage came subsequently with the Malaspina expedition, led by Carlos Duarte, which targeted principally the deep ocean but also the epipelagic layers in a worldwide sampling campaign from 2010 to 2011 (Duarte 2015), and the Ocean Sampling Day initiative, which began with a simultaneous global sampling campaign on 21 June 2014 at 191 different sites, mostly in coastal areas (Kopf et al. 2015).

In 2008, Eric Karsenti led a consortium of scientists that organized a circumglobal expedition on board the 36-m-long schooner SV TARA (**Fig. 3**). The expedition was specifically designed for studying microscopic plankton ecosystems at global scale. Based on a holistic approach, the *Tara* Oceans pan-oceanic expedition sampled plankton ranging in size from viruses to small metazoans, coupled with comprehensive *in situ* biogeochemical measurements (Karsenti et al. 2011). A wide range of contrasting ecosystems were targeted, using sampling protocols that were highly standardized and consistent at each site. *Tara* Oceans is in fact derived from two research expeditions performed between 2009 and 2013. The first expedition (named *Tara* Oceans) lasted two years and eight months and sampled all of the principal ocean basins with the exception of the Arctic Ocean; the second (named *Tara* Oceans Polar Circle) lasted seven months and circumnavigated the Arctic Circle in 2013. An additional expedition denoted *Tara* Pacific, currently under analysis, retrieved sites across the Pacific Ocean, with a particular focus on the biology of coral ecosystems (Planes et al. 2019) (**Fig. 3**). The *Tara* Oceans expeditions consisted in the collection of >35 000 plankton samples from 210 sampling sites, which were used to generate >60 terabases of DNA and RNA sequences and ~7 million images (Sunagawa et al. 2020).

5. Biogeographical Insights of Algae from *Tara* Oceans Metabarcoding

Environmental sequencing initiatives such as the *Tara* Oceans expedition have provided unprecedented insight into the biological diversity of ocean communities, placed within a geographical context. This diversity has been most clearly resolved by amplicon sequencing or barcoding, in which a fragment of the small subunit of the rRNA gene (16S for prokaryotes, 18S for eukaryotes) is universally amplified and massively sequenced from an environmental sample (Le Bescot et al. 2016). Most studies of eukaryotes have so far focused on hypervariable regions of the 18S rRNA gene, which are universal, taxonomically informative, flanked by conserved regions for primer binding, short enough for current sequencing technologies, and are by far the most represented in reference databases (Pawlowski et al., 2018); although alternatives exist (e.g., chloroplast *rbcl*, 16S and 23S rDNA; Djemiel et al., 2020). The V9 and V4 regions are the most commonly used (Stoeck et al., 2010; Pernice et al., 2016; Tragin et al., 2018). This approach enabled a tremendous expansion of richness estimates in the microscopic realm, in spite of recurrent and unresolved problems (e.g., definition of species in the absence of data from sexual crosses, and artefacts introduced by the polymerase chain reaction and sequencing methods; Bhadury and Austen 2010; Berruti et al. 2017).

The first metabarcoding survey of eukaryotes from *Tara* Oceans consisted of the sequencing of >1.7 million reads belonging to the 18S rRNA gene (V9 region) from 334 samples from 47 different geographical sites and revealed the occurrence of approximately 110,000 distinct eukaryotic operational taxonomic units (OTUs, a proxy for species in molecular surveys), of which approximately 20% could be assigned to photosynthetic species (i.e., phytoplankton, and hosts of photosynthetic symbionts (de Vargas et al. 2015; Le Bescot et al. 2016)). Among phytoplankton, the largest numbers of OTUs were assigned to dinoflagellates and diatoms, followed by prasinophytes within the green algae and haptophytes (**Fig. 5**). Other photosynthetic groups, principally other members of the ochrophytes (dictyochophytes, chrysophytes) and cryptomonads were detected at >1% total abundance (**Fig. 5**). In contrast, many other lineages of algae that have formed important references for genomic study, for example red algae, glaucophytes and phaeophytes, represented relatively minor components of the marine microbial communities sampled (**Figs 2, 5**).

In addition to biodiversity, molecular methods can give information into the habitat preferences and morphologies of different algal groups. For example, metabarcoding libraries generated from seawater samples filtered with different exclusion filters can be used to gain insight into the repartition of different algal groups into different size fractions (de Vargas et al. 2015), each of which may have very different functional physiologies and respond differently to environmental perturbations (Li et al. 2009). The largest size fractions considered by *Tara* Oceans (between 180 and 2,000 μm) are dominated by diatoms, due to the presence of chain-forming (e.g., *Hyalosira* and *Fragilaria*) and epizoic (e.g., *Pseudohimantidium*) species; while the phytoplankton in the 5–20 μm and 20–180 μm fractions principally consist of diatoms and dinoflagellates (**Fig. 5**; Malviya et al. 2016). Abundance in the smaller 0.8–5- μm size fraction is much more heterogeneous, with haptophytes, dinoflagellates, diatoms (Leblanc et al. 2018), prasinophytes, pelagophytes, dictyochophytes, and cryptomonads all making substantial contributions.

In relation to their biogeographical distribution, in general terms picophytoplankton (0.2–2 μm , mainly the cyanobacteria *Prochlorococcus* and *Synechococcus*; but also eukaryotes such as pelagophytes) are found in warm, nutrient-poor waters in the tropical and subtropical ocean, while nanophytoplankton (2–20 μm) can be detected potentially year-round in temperate regions (**Fig. 6**). By contrast, microphytoplankton (20–200 μm) bloom mainly at higher latitudes at the beginning of spring and sometimes also in the late summer, as well as in upwelling regions (**Fig. 6**) (Li et al. 2009). Diatoms are particularly prevalent at high latitudes and in upwelling environments (**Fig. 6**), where they are able to bloom and outcompete other marine phytoplankton when nitrate and silicate are abundant (Malviya et al. 2016). The contrasting environmental distributions of diatoms and pelagophytes is particularly notable given their close relationship to one another within the ochrophytes (**Fig. 1**).

In contrast to the clear spatial trends shown by picophytoplankton and diatoms, the abundance patterns of prasinophytes, dinoflagellates, and haptophytes are less straightforward to contextualize, and may vary between different members of each group. For example, the haptophyte *Phaeocystis* is a dominant component of microalgal communities at high latitudes (shown in the corresponding map of **Fig. 6a** the larger circles with colors indicating lower diversity values), whereas other haptophytes (e.g., *Emiliania*) constituting diverse communities show predominantly temperate or even tropical distributions (Bendif et al. 2014; Pfaff et al. 2016). Within the prasinophytes, members of the Mamiellophyceae (e.g.,

Ostreococcus, *Micromonas*) are prominent in coastal waters (Monier, Worden, and Richards 2016; Vannier et al. 2016), while clade VII prasinophytes dominate oceanic waters (Lopes Dos Santos et al. 2017). The high diversity and broad biogeographical distribution of dinoflagellates, including coastal, oceanic, polar, temperate and tropical abundance (Le Bescot et al. 2016) (**Fig. 6**) is indicative of the broad ecological strategies employed by different genera.

6. Functional Studies of Algae from *Tara* Oceans Metagenomes and Metatranscriptomes

In addition to taxonomic analyses based on metabarcoding data, the functional analysis of natural phytoplankton communities can be carried out by metagenomics and metatranscriptomics. The current version of the *Tara* Oceans eukaryotic gene catalog consists of 116 million non-redundant transcribed sequences mainly from eukaryotic plankton ranging from 0.8 to 2000 μm (Carradec et al. 2018). It was generated by assembling 441 poly-A+ metatranscriptomes from samples across the main ocean basins and has no apparent saturation (Carradec et al. 2018). Among the most abundant transcripts found in the catalogue are the chlorophyll *a/b* binding proteins, involved in light harvesting and considered the most abundant membrane proteins on Earth (Dittami et al. 2010).

The variability of the gene and transcript levels of all these sequences across different environmental conditions can give clues about functional responses in phytoplankton communities. The environmental response of phytoplankton on short timescales (over the lifetime of the cell) is called acclimation and includes the modulation of gene expression. By contrast, adaptation to an environmental niche refers to genetic changes (e.g., in gene copy number) that accumulate over many generations. Responses to environmental cues also include shifts in the community structure, changes of broad taxonomic groups as well as changes of genotypes of the same species adapted to local conditions. These links were detected in diatoms, for example with nitrogen (Busseni et al. 2019) and iron uptake systems (Caputi et al. 2019).

Fig. 7 shows the correlations in *Tara* Oceans data between conditions of high productivity and the expression of gene families containing members involved in photosynthesis and carbon fixation. Among the displayed set of photosynthetic genes are those coding for ferredoxins, iron-containing proteins that can be replaced by flavodoxins, which are iron-free versions of less efficiency (Pierella Karlusich et al., 2014). Although diatoms can concentrate up to 30-40% of their cellular iron into ferredoxin (Pierella Karlusich et al., 2014), they show an extreme positive correlation between its expression and conditions of primary production (**Fig. 7**). Furthermore, while quantification of ferredoxin-encoding transcripts among poly-A+ metatranscriptomes can be affected by the fact that the gene is plastid encoded in many diatom species (Lommer et al. 2012; Dorrell et al. 2017), the strong correlation with primary production might reflect the iron uptake and usage efficiency in diatoms, avoiding high investment in flavodoxins due to the presence of other strategies (Carradec et al. 2018; Kazamia et al. 2018; Caputi et al. 2019).

Metatranscriptome data can even provide information regarding the broader life strategies used by different phytoplankton groups, for example in the context of mixotrophy (Burkholder, Gilbert, and Skelton 2008). **Fig. 7** shows that obligate photoautotrophs, such as diatoms and prasinophytes, exhibit high positive correlations between conditions of high productivity and the expression of gene families containing members involved in photosynthesis (i.e., light harvesting, photosynthetic electron transport, and carbon dioxide assimilation through the Calvin cycle); while haptophytes and dinoflagellates show

weak or no correlations (**Fig. 7**). Indeed, dinoflagellates show high positive correlations to primary production for the expression of gene families encoding cell lytic components, such as proteases and lipases, suggesting that dinoflagellates in the most productive marine communities function predominantly as heterotrophs and predators (Carradec et al. 2018). This might reflect switches in the trophic strategy of dinoflagellates under mixotrophic conditions, e.g., the replacement of autonomous amino acid assimilation pathways with scavenging from prey or dissolved external organic material (Johnson 2015; Li et al. 2020; Cohen et al. 2021). Similar positive correlations to peptidases for haptophytes might also reflect previously reported facultative mixotrophy in this group (Unrein et al. 2014; Anderson et al. 2018).

7. Applying Genome-Resolved Metagenomics to Phototrophic Eukaryotes

Initial studies of environmental datasets have focused either on the study of single marker genes (e.g., 18S rDNA) or functional genes, based on comparison to reference genomes and transcriptomes. While these data can provide deep insights into the diversity and functions of lineages with well annotated cultured representatives, they necessarily provide fewer insights into taxa for which reference libraries from cultured species do not exist, as discussed above (Absolon, Smith, and Helliwell 2019).

Recently, environmental sequence data has been harnessed to provide insights into the complete genome sequences of species and lineages that are not yet cultivable in the lab. Two major types of genomic data can be generated from natural samples: single amplified genomes (SAGs) and metagenome-assembled genomes (MAGs). The former are generated by isolating individual cells, amplifying their genomes using whole genome amplification (WGA) and sequencing their DNA, with the first environmental SAG generated in 2007 for microbes of the human mouth (Marcy et al. 2007). Conversely, MAGs are obtained by assembling metagenomic reads into contigs (continuous DNA segment) with the help of specific algorithms, then grouped using nucleotide sequence signatures such as GC content, tetra-nucleotide frequencies, marker gene phylogenies, depth of DNA sequence coverage, and abundance across the samples (Iverson et al. 2012). The grouped samples are then assigned to individual populations using both their coverage and composition as proxies.

The first MAGs were reconstructed from prokaryotic communities in an acid mine drainage (Tyson et al. 2004). This technique has subsequently been widely applied to marine viral, archaeal and bacterial populations (Parks et al. 2017; Tully et al. 2017; Delmont et al. 2018; Tully, Graham, and Heidelberg 2018; Gregory et al. 2019; Cao et al. 2020; Reji and Francis 2020); and for some heterotrophic eukaryotes (Yoon et al. 2011; Mangot et al. 2017; Seeleuthner et al. 2018); but very few examples have been available from eukaryotic algae. This bias reflects the size and complexity of eukaryotic genomes, which makes them difficult to assemble from limited environmental sequencing data (Tyson et al. 2004; Gregory 2005), particularly in the case of genomes without closely related reference genomes (Bowers et al. 2017; Nelson, Tully, and Mobberley 2020); and historical priority in funding biases for less easily cultivable heterotrophic protists (del Campo et al. 2014). Moreover, the absence of standardised meta-data, e.g., environmental parameters, for assembled metagenomes can limit their direct comparability to one another (Sibbald and Archibald 2017).

Nonetheless, new methods are emerging to improve the recovery of eukaryotic genomes from environmental data (West et al. 2018). For instance, among the MAGs generated for marine phototrophic eukaryotes are small species such as *Micromonas* (Delmont et al. 2015, 2018), *Bathycoccus*, *Ostreococcus* and *Emiliana* (Delmont et al. 2018), as well as other prasinophyte and diatom species (Duncan et al. 2020). Such studies allow us to gain insight into the diversity and ecology of these groups based on their natural populations.

Very recently, an important database has been released that represents 713 marine eukaryotic SMAGs (both MAGs and a few SAGs) from a wide range of organisms, generated from *Tara* Oceans samples, including from the Arctic Ocean (Delmont et al. 2020). In this, more than 270 SMAGs (~30% of the database) have been assigned as marine eukaryotic algae, dramatically expanding the number of algal genomes available (**Fig. 2A**). For instance, it increases by 18 times the number of available genomes for haptophytes, by 15 for cryptomonads, by 6 for prasinophytes, diatoms and chrysophytes, and enables access for the first time to genomic data for taxonomic groups (e.g., dictyochophytes and bolidophytes; **Fig. 2A**) for which no genomes were previously available. Moreover, this *Tara* Oceans SMAGs database gathers highly contextualized data, is freely available, and the genomes have relatively elevated completeness (~40% on average and up to 94% BUSCO completion, a measure of genome completeness based on the detection of broadly conserved marker proteins; Simão et al. 2015, Hanschen and Starkenburg 2020). This new database will be a very important resource to gain insight into the genomic diversity of previously under-sampled branches of the eukaryotic tree of life.

8. Perspectives

In this chapter, we have explored the dramatic changes in eukaryotic algal genome resources that have become available within one generation of scientific research; starting with the completion of the first chloroplast genomes of cultured algae in the mid-1990s (Kowallik et al. 1995; Reith and Munholland 1995) and currently extending to ecologically contextualised genomes and transcriptomes of uncultured algae from environmental sequence metadata (Carradec et al. 2018; Delmont et al. 2020). This is revolutionising our understanding of algal diversity, evolution, and functional biology, but challenges nonetheless remain.

The quality of the assembly and annotation of cultured algal genomes and transcriptomes may vary substantially (Hanschen and Starkenburg 2020), and indeed assembly is only the start of the much more complex process of functional annotation of genome content, which may depend on the development of transformation strategies enabling reverse genetics (Faktorová et al. 2020). However, an increasing number of genomes from transformable algal species (e.g., the diatom *Phaeodactylum tricornutum*; Rastogi et al. 2018; Yang et al. 2018) have undergone multiple rounds of revision, integrating transcriptomic and even proteomic data to completely annotate the encoded complement of genes. An additional question to consider is the genomic representativeness of individual isolates for an entire species. Recent projects, focusing on within-species genomic diversity, have shown that genomic content and organisation can vary dramatically even between strains of the same algal species, isolated from different environments (Read et al. 2013; Flowers et al. 2015; Rastogi et al. 2020). This micro-evolution may be as significant as the phylogenetic distances between major groups for underpinning genomic diversity in eukaryotic algae (Blanc-Mathieu et al. 2017).

The environmental sequence datasets available through projects such as *Tara* Oceans expand our insights into the biodiversity of previously unexplored oceanic regions, although with notable gaps (e.g., the Bay of Bengal; much of the southern Indian Ocean; and the Atlantic African coast; **Fig. 3**). We hope that future sequencing projects may be able to further redress some of the geographical discrepancies in what marine sequence data are available; and may involve more direct participation from researchers in transitional and developing economies, which may be highly phylogenetically diverse and disproportionately affected by anthropogenic pollution and climate change (Msuya and Hurtado 2017; Lee 2019). These projects may also form the starting point for groundbreaking explorations of algal diversity in freshwater and terrestrial environments, which have very different chemical properties (e.g., extreme nutrient limitations and enrichments, highly variant pH, lower salinity, and large diurnal and annual fluctuations in temperature), and in which very different patterns (e.g., a more dominant role for chrysophytes and cryptomonads, and of mixotrophic and obligately heterotrophic species) are observed (Dorrell et al. 2019; Djemiel et al. 2020; Singer et al. 2020; Nelson et al. 2021).

The identification and assembly of SMAGs raises new challenges concerning assembly and quality; for example concerning the minimum information about single-amplified and metagenome-assembled genomes (MISAG, MIMAG) acceptable for deposition in public repositories (Bowers et al. 2017). Certain lineages, particularly dinoflagellates, are currently not represented at all in SMAG data (**Fig. 2**); reflecting their large and highly complex nuclear genomes (Gornik et al. 2019). These may require deeper sequencing than currently feasible for genome-resolved metagenomics. Finally, the metagenomic identification of poorly culturable organisms, some of which may correspond to novel branches of the algal tree of life (Delmont et al. 2020), may open up new chapters in laboratory research. SMAGs may be particularly important in this regard, revealing possible undocumented auxotrophies (e.g., for vitamins; Helliwell et al. 2011; Absolon, Smith, and Helliwell 2019) that may eventually allow us to bring some of the previously uncultured lineages, such as MOCHs and rappemonads (Kim et al. 2011; Massana et al. 2014), into the laboratory for more detailed observation.

The genome revolution has already transformed our understanding of eukaryotic algal biology, but this revolution is ongoing. Understanding the genomics and biodiversity of eukaryotic algae will be essential in meeting conservation and biofuels targets agreed within the COP21 climate agreement (Pires 2017); and is a central objective of the upcoming UN Decade for Ocean Science (Lee 2019; Nash et al. 2020). Further exploration, and new sequence assembly and analysis, may help expose new patterns of diversity and functions in marine algal communities

Glossary Box

Genome: an assembled sequence of all DNA found in a genome, including both coding and non-coding regions. After sequencing and assembly, genomes have to be annotated to define where genes are located, and predicted probable coding content.

Transcriptome: a survey of all expressed sequences in a cell, based on next-generation sequencing of cellular RNA (via cDNA). Transcriptomes can provide direct insight into the sequences of coding regions

(i.e., genes) in eukaryotic genomes, and (by allowing consideration of the levels of different gene expression) reveal physiological responses to environmental stimuli.

Phytoplankton and Algae: typically refer to single-celled or colonial (chain-forming) photosynthetic organisms, which can either include bacteria (i.e., cyanobacteria) or refer only to eukaryotes (e.g. diatoms, dinoflagellates, and haptophytes). Algae are evolutionarily diverse, with many of the main eukaryotic groups distantly related to plants, to us, and even to one another, although all eukaryotic algae share chloroplasts, derived through endosymbiosis.

Endosymbiosis: the process by which mitochondria and chloroplasts originate, proposed in 1904 by Konstantin Mereschkowsky and revitalised in the 1960s by Lynn Margulis. Free-living organisms (photosynthetic in the case of chloroplasts; and either bacteria or single-celled eukaryotes) are swallowed up by a predatory eukaryote, and converted into permanent bioenergetic organelles. While only one mitochondrial endosymbiosis is known, at the origin of eukaryotes, chloroplasts have been acquired through endosymbiosis at least ten times across the eukaryotic tree.

Horizontal Gene Transfer: also referred to as HGT or Lateral Gene Transfer; the non-sexual exchange of genetic information between cells, typically of different species. Horizontal Gene Transfer may occur as a continuous process over evolution, and may involve both prokaryotic and eukaryotic partners; and may be a driving force in rapid environmental adaptations.

Photo-mixotrophy: a lifestyle in which an organism both derives energy through photosynthesis, and through the phagotrophic consumption of other organisms or the uptake of dissolved organic nutrients. This strategy, while rare in plants (c.f., carnivorous plants) is widespread in algae, and is particularly important for the biology of dinoflagellates.

Diatoms: “the rainforests of the oceans”, diatoms are typically rather large, sessile algae living inside silica shells, which contribute up to one-fifth of total planetary photosynthesis. Diatoms are particularly abundant at high latitudes (e.g. the Arctic, and the Antarctic), where they can exploit abundant seawater nutrients to underpin the marine food chain.

Haptophytes: also referred to as coccolithophorids; typically small, eukaryotic algae that live inside calcium carbonate shells, and can form highly abundant seasonal “blooms” visible from space. The haptophyte shells, if deposited on the seafloor and compressed, can be converted over geological timescales into calcareous rocks, e.g. limestone and chalk.

Dinoflagellates: a highly heterogeneous and diverse group of algae, encompassing photosynthesisers, mixotrophs and heterotrophs. Dinoflagellates include the principal photosynthetic symbiont of coral (*Symbiodinium*) alongside harmful “red tide” species, which can produce toxic compounds that result in mass shellfish and marine vertebrate die-offs.

Barcode: a short region of sequence (typically from the nuclear small subunit rDNA) that has been sequenced in many cultured organisms. Sequencing DNA barcodes amplified from environmental

566 samples, and comparing to databanks of equivalent sequences from cultured species, allows us to profile
567 the probable species diversity in the wild.

568
569 **Meta-genome** and **meta-transcriptome**: sequencing of all DNA or RNA in an environmental sample,
570 e.g. seawater. This will provide the sequences of not just one species, but a coexisting community of
571 different organisms; which allows insights into ecological interactions at the expense of resolving
572 sequence origin.

573
574 **SMAGs**: Assembled Genome sequences obtained from uncultured organisms, either by intensive
575 sequencing of Single isolated cells (SAGs) or the mathematical resolution of large amounts of sequence
576 data from multiple Metagenome datasets (MAGs).

577 578 **Figures** 579

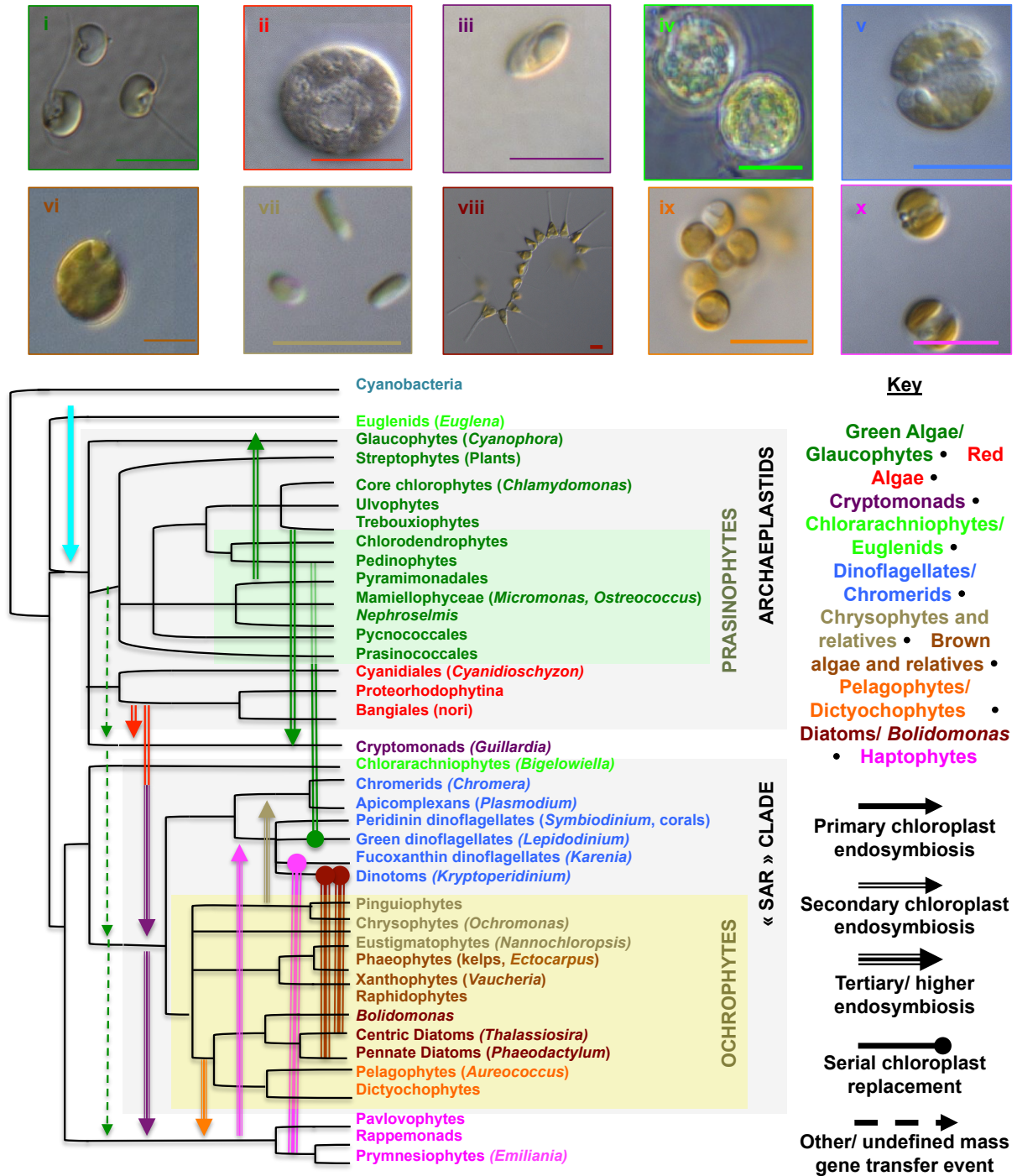


Fig. 1. Diversity of eukaryotic algae. Top: exemplar photos of eukaryotic algae. Images are (i) *Nephroselmis* sp. RCC1805 (prasinophyte); (ii) *Rhodella maculata* RCC655 (proteorhodophytinan red alga); (iii) *Hemiselmis* sp. RCC4215 (cryptomonad); (iv) *Lotharella reticulosa* RCC376 (chlorarachniophyte); (v) *Karlodinium veneficum* RCC2539 (fucoxanthin-containing dinoflagellate); (vi) *Heterosigma akashiwo* RCC1502 (raphidophyte); (vii) *Ochromonas* sp. RCC2350 (chrysophyte); (viii) *Asterionellopsis glacialis* RCC2636 (diatom); (ix) *Pelagococcus subviridis* RCC5644 (pelagophyte); and (x) *Isochrysis* sp. RCC1956 (haptophyte). Scale bars: 10 μ m. All images kindly supplied by the Roscoff

Culture Collection. **Bottom:** schematic tree of eukaryotic algae, summarising key phylogenetic relationships and endosymbiotic gene transfers that underpin the evolution of algal genomes. Branch colours correspond to the box colours in the photo above. Data in this figure is summarised from Stiller et al. 2014; Dorrell et al. 2017; and Dorrell et al. 2021.

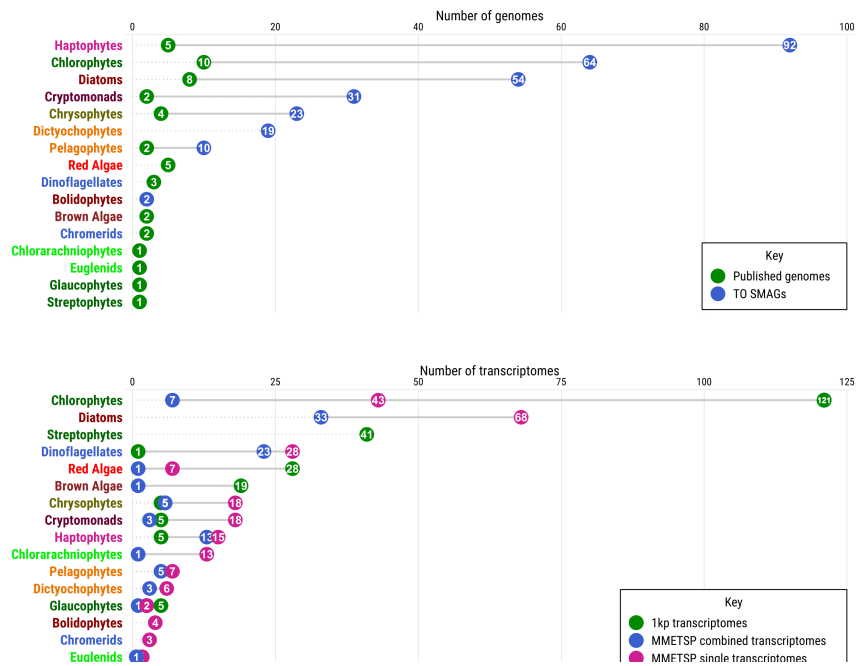


Fig. 2. Diversity of published algal sequence libraries. Top: number of published algal genomes (Grigoriev et al. 2020; Hanschen and Starkenburg. 2020; Nelson et al. 2021) and Tara Oceans single amplified genomes and metagenome-assembled genomes (SMAGs, Delmont et al. 2020); and **bottom:** number of equivalent from the MMETSP and OneKp sequencing projects (Keeling et al. 2014; Carpenter et al. 2019). Dots that shade each other (e.g. for chrysophyte and euglenid transcriptomes) denote a same value. Taxa are shaded by phylogenetic origin, using the same scheme as Fig. 1.

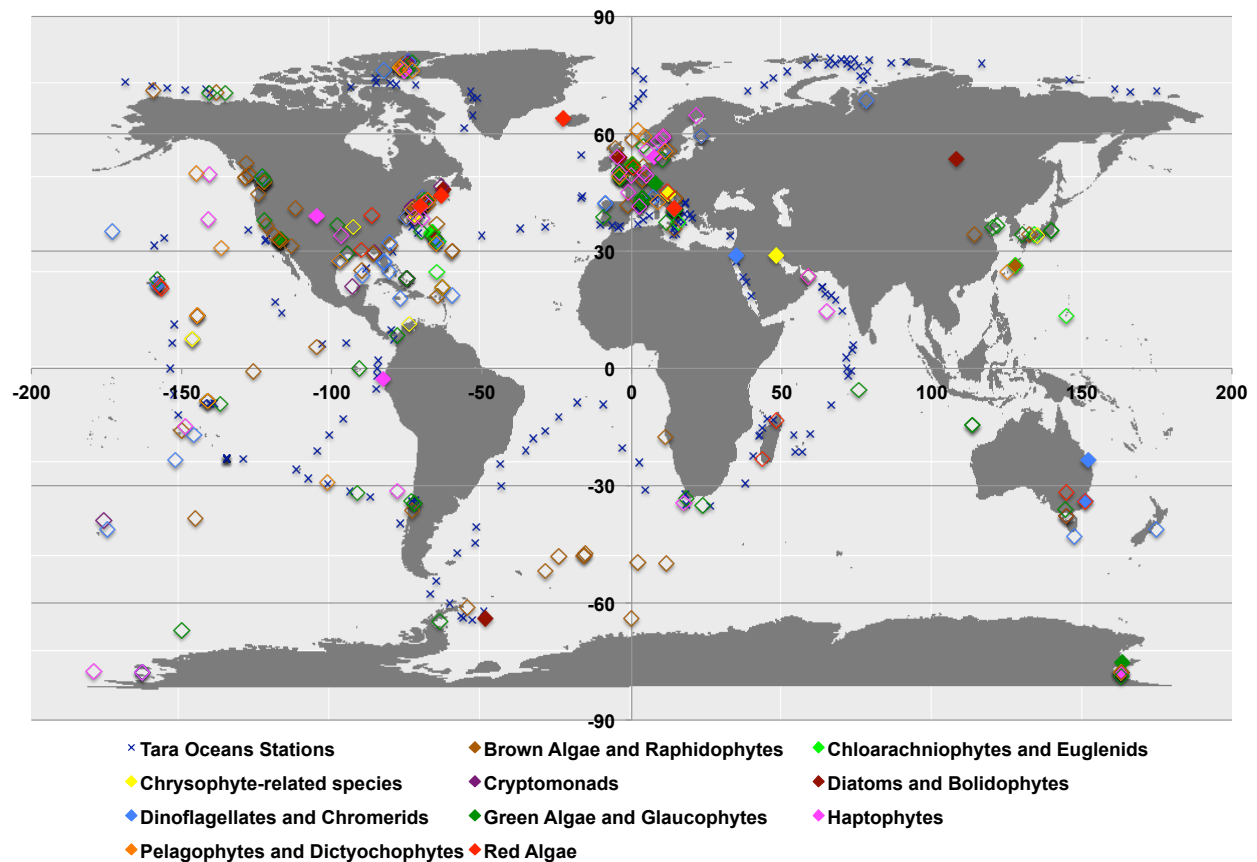


Fig. 3. Isolation sites of sequenced algal genomes (solid shapes) and MMETSP transcriptomes (open shapes), compared to sampling sites for the Tara Oceans and Tara Oceans Polar Circle expeditions. Isolation sites were manually verified from culture collection accession for each strain (Vaulot et al. 2004; Gachon et al. 2013; Guiry et al. 2014; Boundy-Mills et al. 2015). Algal sequence libraries are coloured by phylogenetic origin, using the same scheme as fig. 1.

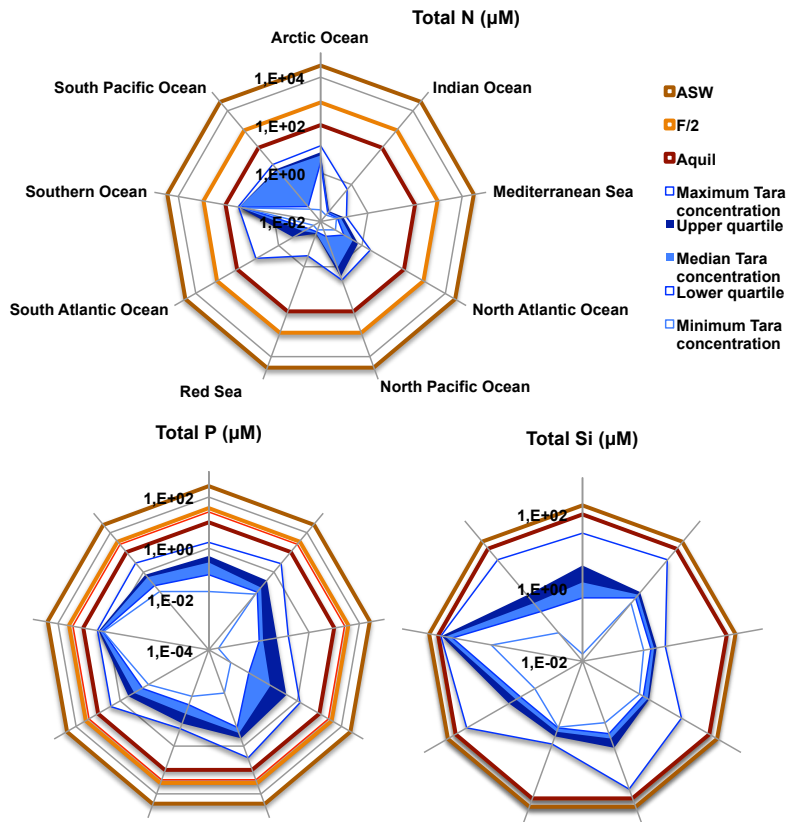


Fig. 4. Maximum and minimum values (blue lines) and interquartile ranges (blue sectors) of measured total concentrations at 5 m depth of nitrogen, phosphorus and silicon in different Tara Oceans regions; compared to supplemented concentrations of each nutrient in three growth media (ASW, f/2 and Aquil; red lines) commonly used for growth of laboratory algal cultures. The authors acknowledge Elena Kazamia for the calculation of growth media nutrient concentrations.

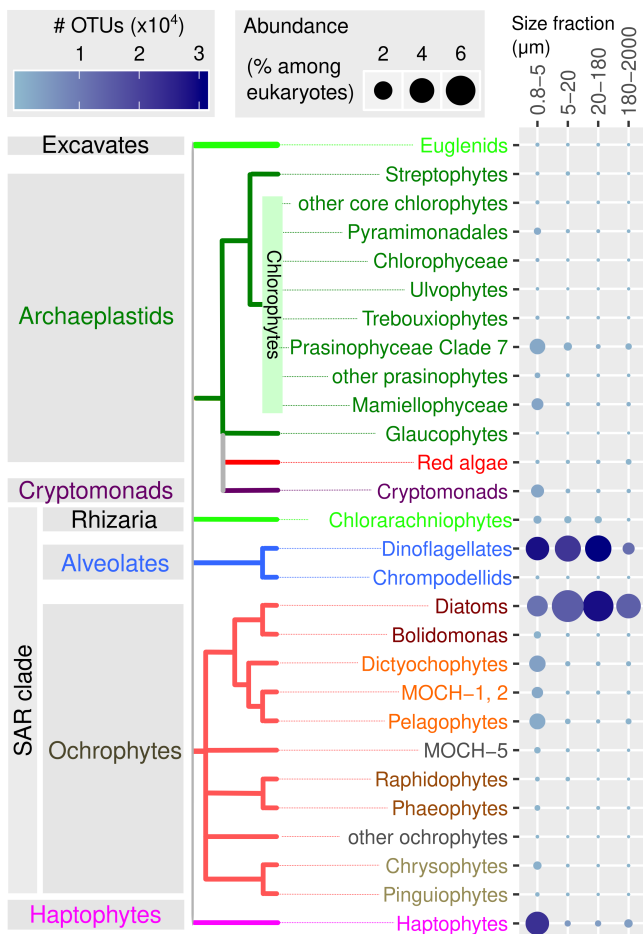


Fig 5: Taxonomic distribution, abundance, and richness of marine eukaryotic phytoplankton based on 18S rRNA gene (V9) metabarcoding across different size fractionated Tara Oceans samples. The area of each circle corresponds to the average abundance relative to total eukaryotic read counts, while the color gradient varies according to the number of operational taxonomic units (OTUs). Data are derived from the complete Tara Oceans data set for 18S rRNA gene (V9 region) metabarcoding (de Vargas et al. 2015, Ibarbalz et al. 2019). Size-fractionated samples were generated via membrane filtration with different pore sizes.

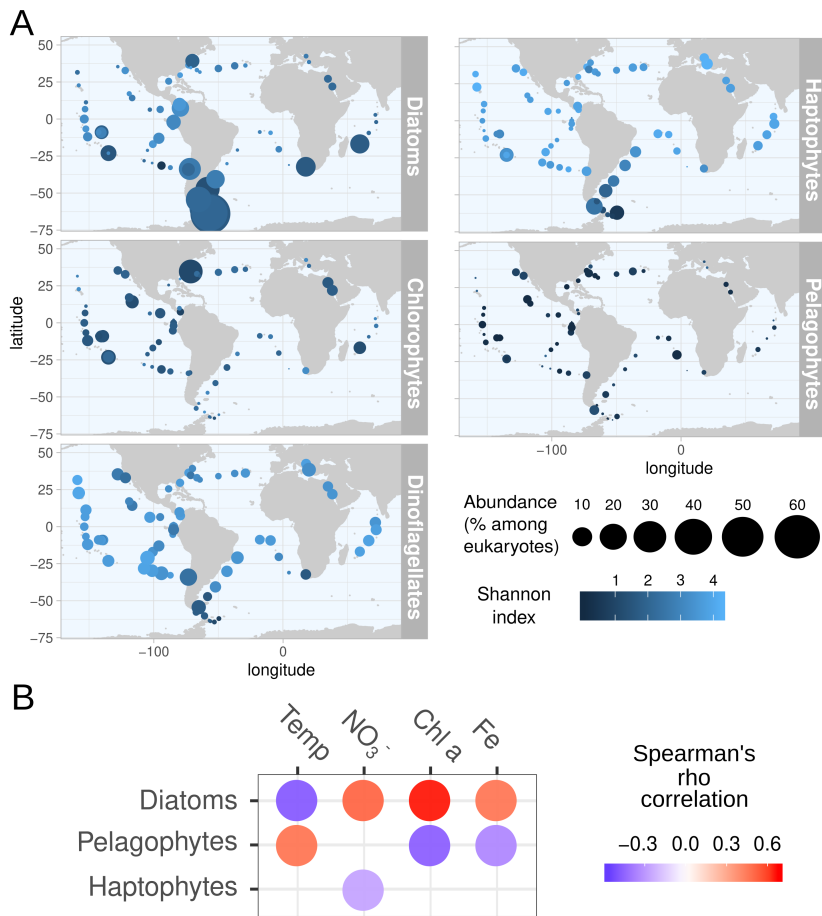


Fig 6: Biogeography and environmental distribution of photosynthetic algae inferred from Tara Oceans samples. A: Biogeography of the major photosynthetic protists based on 18S rRNA gene (V9 region) metabarcoding data from the 0.8–2,000- μm size fraction (de Vargas et al. 2015). The size of each circle corresponds to the abundance at each location, and the fill color refers to the Shannon diversity index (dark blue for low diversity and light blue for high diversity). **B:** Correlation analysis between relative abundances of the main phytoplankton groups and a selection of productivity-related parameters. Color represents Spearman's rho coefficients. Empty spaces refer to nonsignificant correlation values ($p > 0.05$). Chlorophytes and dinoflagellates are not shown due to absence of significant correlations.

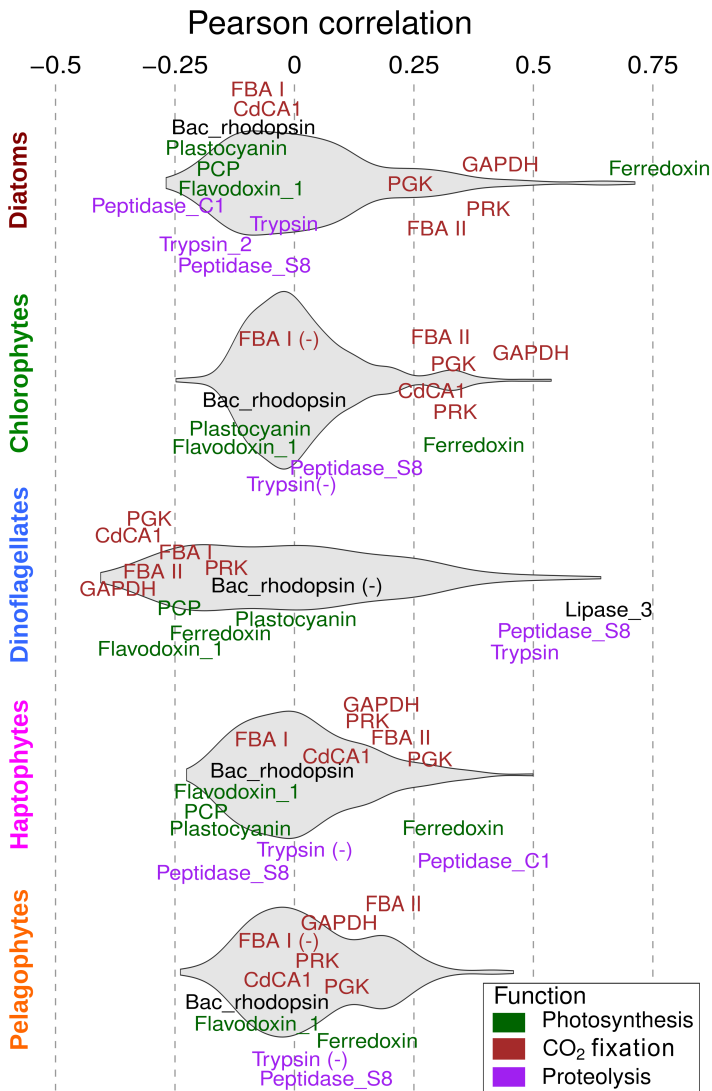


Fig. 7: Functional insights into global ocean communities of algae using Tara Oceans metatranscriptomes. This figure shows the distribution of Pearson correlation values of Pfam domains with measurements of net primary productivity in the five major photosynthetic planktonic groups. The profiles illustrate the density distribution of all the correlation values of the Pfams having a relative expression value at least equal to 0.05% (relative expression value is indicated by the size of the grey dots). Pfam representatives of three main functional classes (CO₂ fixation through the Calvin Cycle, photosynthetic light reactions/ light harvesting, and proteolysis) are coloured. Adapted from Carradec et al. (2018).

Acknowledgments

RGD acknowledges a CNRS Momentum Fellowship (awarded 2019-2021). CB has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Diatomic; grant agreement No. 835067). Additional funding is acknowledged from the FFEM - French Facility for Global Environment (Fonds Français pour l'Environnement Mondial), and the French Government 'Investissements d'Avenir' Programmes MEMO LIFE (Grant

ANR-10-LABX-54), Université de Recherche Paris Sciences et Lettres (PSL) (Grant ANR-1253 11-IDEX-0001-02), France Genomique (ANR-10-INBS-09), and OCEANOMICS (Grant ANR-11-BTBR-0008). JPK acknowledges postdoctoral funding from the Fonds Français pour l'Environnement Mondial. CN acknowledges postdoctoral funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme through the project DIATOMIC (grant agreement No. 835067) to C.B. The authors thank Daniel Vaultot and Ian Probert (Roscoff Culture Collection) for the reuse of images included in Fig. 1; Elena Kazamia for assistance with the generation of Fig. 4; and Quentin Carradec for assistance with the generation of Fig. 7.

References

- Absolon, D., Smith, A., and Helliwell, K. 2019. Shining a light on microbial dark matter: a role for the forgotten B vitamins in marine algal communities? *Access Microbiol.* 10: 1099. <https://doi.org/10.1099/acmi.ac2019.po0295>.
- Allen, A.E., Dupont, C.L., Oborník, M., Horák, A., Nunes-Nesi, A., McCrow, J.P., Zheng, H. et al. 2011. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* 473 (7346): 203–7. <https://doi.org/10.1038/nature10074>
- Amann, R. I., Ludwig, W., and Schleifer, K.H. 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev* 59 (1): 143–69.
- Amaral-Zettler, L. Artigas, L.F., Baross, J., Loka Bharathi, P. A., Boetius, A., Chandramohan, D., Herndl, G. et al. 2010. A global census of marine microbes. *Life in the World's Oceans*. <https://doi.org/10.1002/9781444325508.ch12>.
- Anderson, R., Charvet, S., and Hansen, P.J. 2018. Mixotrophy in chlorophytes and haptophytes—effect of irradiance, macronutrient, micronutrient and vitamin limitation. *Front. Microbiol.* 1704. <https://doi.org/10.3389/fmicb.2018.01704>
- Armbrust, E.V., Berges, J.A., Bowler, C., Green, B.R., Martinez, D., Putnam, N.H., Zhou, S., et al. 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* 306 (5693): 79–86.
- Bailleul, B., Rogato, A., de Martino, A., Coesel, S., Cardol, P., Bowler, C., Falciatore, A., and Finazzi, G. 2010. An atypical member of the light-harvesting complex stress-related protein family modulates diatom responses to light. *Proc. Natl. Acad. Sci. USA* 107 (42): 18214–19.
- Bendif, E.M., Probert, I., Carmichael, M., Romac, S., Hagino, K., and de Vargas, C. 2014. Genetic delineation between and within the widespread coccolithophore morpho-species *Emiliania huxleyi* and *Gephyrocapsa oceanica* (Haptophyta). *J. Phycol.* 50 (1): 140–48.
- Berruti, A., Desirò, A., Visentin, S., Zecca, O., and Bonfante, P. 2017. ITS fungal barcoding primers versus 18S AMF-specific primers reveal similar AMF-based diversity patterns in roots and soils of three mountain vineyards. *Envi. Microbiol. Rep.* <https://doi.org/10.1111/1758-2229.12574>.
- Bhadury, P. and Austen, M.C. 2010. Barcoding marine nematodes: an improved Set of Nematode 18S rRNA Primers to Overcome Eukaryotic Co-Interference. *Hydrobiologia*. <https://doi.org/10.1007/s10750-009-0088-z>.
- Blanc-Mathieu, R., Krasovec, M., Hebrard, M., Yau, S., Desgranges, E., Martin, J., Schackwitz, W. et al. 2017. Population genomics of picophytoplankton unveils novel chromosome hypervariability. *Science Advances* 3 (7): e1700239. <https://doi.org/10.1126/sciadv.1700239>
- Bock, R. 2017. Witnessing genome evolution: experimental reconstruction of endosymbiotic and horizontal gene transfer. *Ann. Rev. Genet.* 51 (11): 1–22.
- Bork, P., Bowler, C., de Vargas, C., Gorsky, G., Karsenti, E., and Wincker, P. 2015. Tara Oceans studies plankton at a planetary scale. *Science*. 348 (6237) 873. <https://doi.org/10.1126/science.aac5605>.
- Boundy-Mills, K., Hess, M., Bennett, A.R., Ryan, M., Kang, S., Nobles, D., Eisen, J.A. et al. 2015. The United States Culture Collection Network (USCCN): enhancing microbial genomics research

- through living microbe culture collections. *Appl. Env. Microbiol.* 81 (17): 5671–74.
- Bowers, R. M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., Schulz, F. et al. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of Bacteria and Archaea. *Nat. Biotechnol.* 35 (8): 725–31.
- Bowler, C., Vardi, A., and Allen, A.E. 2010. Oceanographic and biogeochemical insights from diatom genomes. *Ann. Rev. Mar. Sci.* 2: 333–65.
- Brawley, S. H., Blouin, N.A., Ficko-Blean, E., Wheeler, G.L., Lohr, M., Goodson, H.V., Jenkins, J.W. et al. 2017. Insights into the red algae and eukaryotic evolution from the genome of *Porphyra purpurea* (Bangiophyceae, Rhodophyta). *Proc. Natl. Acad. Sci. USA* 114 (31): 6361–70.
- Browning, T.J., Achterberg, E.P., Rapp, I. Engel, A., Bertrand, E.M., Tagliabue, A., and Moore, C.M. 2017. Nutrient co-limitation at the boundary of an oceanic gyre. *Nature* 551 (7679): 242–46.
- Burkholder, J.M., Glibert, P.M., and Skelton, H.M. 2008. Mixotrophy, a major mode of nutrition for harmful algal species in eutrophic waters. *Harmful Algae*. 8 (1): 77–93.
<https://doi.org/10.1016/j.hal.2008.08.010>.
- Bussené, G., Vieira, F.R.J., Amato, A., Pelletier, E., Pierella Karlusich, J.J., Ferrante, M.I., Wincker, P. et al. 2019. Meta-omics reveals genetic flexibility of diatom nitrogen transporters in response to environmental changes. *Mol. Biol. Evol.* 36 (11): 2522–2535.
<https://doi.org/10.1093/molbev/msz157>.
- del Campo, J., Sieracki, M.E., Molestina, R., Keeling, P., Massana, R., and Ruiz-Trillo, I. 2014. The others: our biased perspective of eukaryotic genomes. *Trends Ecol. Evol.* 29 (5): 252–59.
- Cao, S., Zhang, W., Ding, W., Wang, M., Fan, S., Yang, B., Mcminn, A. et al. 2020. Structure and function of the Arctic and Antarctic marine microbiota as revealed by metagenomics. *Microbiome* 8 (1): 47.
- Caputi, L., Carradec, Q., Eveillard, D. Kirilovsky, A., Pelletier, E., Pierella Karlusich, J.J. Vieira, F.R.J, et al. 2019. Community-level responses to iron availability in open ocean plankton ecosystems. *Global Biogeochemical Cycles* 33 (3): 391–419.
- Carpenter, E. J., Matasci, N., Ayyampalayam, S., Wu, S., Sun, J., Yu, J., Vieira, F.R.J., et al. 2019. Access to RNA-sequencing data from 1,173 plant species: the 1000 plant transcriptomes initiative (1KP). *GigaScience* 8 (10): 1093. <https://doi.org/10.1093/gigascience/giz126>.
- Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., Lima-Mendez, G., et al. 2018. A global ocean atlas of eukaryotic genes. *Nat. Commun.* 9 (1): 373.
- Cavalier-Smith, T. 2003. Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-Algae). *Phil. Trans. R. Soc. Biol.* 358 (1429): 109–34.
- Cohen, N.R., McIlvin, M.R., Moran, D.M., Held, N.A., Saunders, J.K., Hawco, N.J., Brosnahan, M. et al. 2021. Dinoflagellates alter their carbon and nutrient metabolic strategies across environmental gradients in the central Pacific Ocean. *Nat. Microbiol.* <https://doi.org/10.1038/s41564-020-00814-7>
- Cumbo, V.R., Baird, A.H., Moore, R.B., Negri, A.P., Neilan, B.A., Salih, A., van Oppen, M.J.H., Wang, Y., and Marquis, C.P. 2013. *Chromera velia* Is endosymbiotic in larvae of the reef corals *Acropora digitifera* and *A. tenuis*. *Protist* 164 (2): 237–44.
- Delmont, T.O., Eren, A.M., Vineis, J.H., and Post, A.F. 2015. Genome reconstructions indicate the partitioning of ecological functions inside a phytoplankton bloom in the Amundsen Sea, Antarctica. *Frontiers Microbiol.* 6 (10): 1090.
- Delmont, T.O., Gaia, M., Hinsinger, D.D. Fremont, P., Fernandez Guerra, A., Eren, A.M. Vanni, C. et al. 2020. Functional repertoire convergence of distantly related eukaryotic plankton lineages revealed by genome-resolved metagenomics. *Cold Spring Harbor Laboratory*.
<https://doi.org/10.1101/2020.10.15.341214>.
- Delmont, T.O., Quince, C., Shaiber, A., Esen, Ö.C., Lee, S.T., Rappé, M.S., McLellan, S.L., Lückner, S., and Eren, A.M. 2018. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat. Microbiol.* 3 (7): 804–13.
- Derelle, E., Ferraz, C., Rombauts, S., Rouzé, P., Worden, A.Z., Robbens, S., Partensky, F., et al. 2006.

Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proc. Natl. Acad. Sci USA* 103 (31): 11647–52.

Dittami, S.M., Michel, G., Collén, J., Boyen, C., and Tonon, T. 2010. Chlorophyll-binding proteins revisited—a multigenic family of light-harvesting and stress proteins from a brown algal perspective. *BMC Evol. Biol.* 10 (11): 365.

Djemiel, C., Plassard, D., Terrat, D., Crouzet, O., Sauze, J., Mondy, S., Nowak, V., Wingat, L., Ogée, J. and Maron, P.A. 2020. μ green-db: a reference database for the 23S rRNA gene of eukaryotic plastids and cyanobacteria. *Sci Rep* (2020) 10:5915, <https://doi.org/10.1038/s41598-020-62555-1>

Dorrell, R.G., Azuma, T., Nomura, M., Audren de Kerdrel, G., Paoli, L., Yang, S., Bowler, C. et al. 2019. Principles of plastid reductive evolution illuminated by nonphotosynthetic chrysophytes. *Proc. Natl. Acad. Sci. USA* 116 (14): 6914–23.

Dorrell, R.G., Gile, G., McCallum, G., Méheust, R., Baptiste, E.P., Klinger, C.M., Brillet-Guéguen, L., Freeman, K.D., Richter, D.J., and Bowler, C. 2017. Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. *eLife* 6 (5): 23717. <https://doi.org/10.7554/eLife.23717>.

Dorrell, R.G., Villain, A., Perez-Lamarque, B., McCallum, G., Watson, A.K., Ait-Mohamed, O., Alberti, A., et al. 2021. Phylogenomic fingerprinting of tempo and functions of horizontal gene transfer within ochrophytes. *Proc. Natl. Acad. Sci. USA* 118 (4): e2009974118. <https://doi.org/10.1073/pnas.2009974118>

Dorrell, R.G., and Howe, C.J. 2015. Integration of plastids with their hosts: lessons learned from dinoflagellates. *Proc. Natl. Acad. Sci USA* 112 (33): 10247–54.

Dorrell, R.G., and Smith, A.G. 2011. Do red and green make brown?: Perspectives on plastid acquisitions within chromalveolates. *Eukaryot. Cell* 10 (7): 856–68.

Duarte, C.M. 2015. Seafaring in the 21st Century: The Malaspina 2010 circumnavigation expedition. *Limnology and Oceanography Bulletin*. <https://doi.org/10.1002/lob.10008>.

Duncan, A., K. Barry, C. Daum, E. Eloie-Fadrosch, S. Roux, S. G. Tringe, K. Schmidt, et al. 2020. Metagenome-assembled genomes of phytoplankton communities across the Arctic Circle. *Cold Spring Harbor Laboratory*. <https://doi.org/10.1101/2020.06.16.154583>.

Dunthorn, M., Otto, J., Berger, S.A., Stamatakis, A., Mahé, F., Romac, S., de Vargas, C., et al. 2014. Placing environmental next-generation sequencing amplicons from microbial eukaryotes into a phylogenetic context. *Mol. Biol. Evol.* 31 (4): 993–1009.

Durak, G.M., Brownlee, C., and Wheeler, G.L. 2017. The role of the cytoskeleton in biomineralisation in haptophyte algae. *Sci. Rep.* 7 (1): 15409.

Faktorová, D., Nisbet, R.E.R., Fernández Robledo, J.A., Casacuberta, E., Sudek, L., Allen, A.E., Ares Jr, M. et al. 2020. Genetic tool development in marine protists: emerging model organisms for experimental cell biology. *Nature Methods* 17 (5): 481–94.

Fernández-Castro, B., Mouriño-Carballido, B., Benítez-Barrios, V. M., Chouciño, P., Fraile-Nuez, E., Graña, R., Piedeleu, M., and Rodríguez-Santana, A. 2014. Microstructure turbulence and diffusivity parameterization in the tropical and subtropical Atlantic, Pacific and Indian Oceans during the Malaspina 2010 Expedition. *Deep Sea Research Part I: Oceanographic Research Papers*. <https://doi.org/10.1016/j.dsr.2014.08.006>.

Flowers, J. M., Hazzouri, K. M., Pham, G. M., Rosas, U., Bahmani, T., Khraiwesh, B., Nelson, D. R. et al. 2015. Whole-genome resequencing reveals extensive natural variation in the model green alga *Chlamydomonas reinhardtii*. *Plant Cell*. 27(9):2353-69. <https://doi.org/10.1105/tpc.15.00492>.

Fuhrman, J. A., McCallum, K., and Davis, A. A. 1995. Phylogenetic diversity of subsurface marine microbial communities from the Atlantic and Pacific Oceans. *Appl. Env. Microbiol.* 59(5): 1294–1302. <https://doi.org/10.1128/aem.61.12.4517-4517>.

Gachon, C. M. M., Heesch, S., Küpper, F. C., Achilles-Day, U. E. M., Brennan, D., Campbell, C. N., Clarke, A. et al. 2013. The CCAP KnowledgeBase: linking protistan and cyanobacterial biological resources with taxonomic and molecular data. *Systemat. Biodiv.* 11 (4): 407-413. <https://doi.org/10.1080/14772000.2013.859641>.

- Giovannoni, S. J., Britschgi, T. B., Moyer, C. L. and Field, K. G. 1990. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* 345 (6270): 60–63.
- Gomez, F. 2005. A list of free-living dinoflagellates species in the world's ocean. *Acta Botanica Croatica* 64: 129–212.
- González-Pech, R. A., Bhattacharya, D., Ragan, M. A., and Chan, C. X. 2019. Genome Evolution of Coral Reef Symbionts as Intracellular Residents. *Trends in Ecology & Evolution* 34 (9): 799–806.
- Gornik, S. G., Hu, I., Lassadi, I., and Waller, R. F. 2019. The biochemistry and evolution of the dinoflagellate nucleus. *Microorganisms* 7 (8): 245. <https://doi.org/10.3390/microorganisms7080245>.
- Gregory, A. C., Zayed, A. A., Conceição-Neto, N., Temperton, B., Bolduc, B., Alberti, A., Ardyna, M., et al. 2019. Marine DNA viral macro- and microdiversity from pole to pole. *Cell* 177 (5): 1109–23.e14.
- Gregory, T. R. 2005. Synergy between sequence and size in large-scale genomics. *Nature Reviews. Genetics* 6 (9): 699–708.
- Grigoriev, I. V., Hayes, R. D., Calhoun, S., Kamel, B., Wang, A., Ahrendt, S., Dusheyko, S. et al. 2020. PhycoCosm, a comparative algal genomics resource. *Nucl. Acids Res.*, gkaa898. <https://doi.org/10.1093/nar/gkaa898>.
- Guiry, M. D., Guiry, G. M., Morrison, L., Rindi, F., Valenzuela Miranda, S., Mathieson, A. C., Parker, B. C. et al. 2014. AlgaeBase: an on-line resource for algae. *Cryptogamie, Algologie*. 35(2):105-115. <https://doi.org/10.7872/crya.v35.iss2.2014.105>.
- Hanschen, E. R., and Starkenburg, S.R. 2020. The state of algal genome quality and diversity. *Algal Res.* 101968. <https://doi.org/10.1016/j.algal.2020.101968>.
- Helliwell, K.E., Wheeler, G.L., Leptos, K.C., Goldstein, R.E., and Smith, A.G. 2011. Insights into the evolution of vitamin B₁₂ auxotrophy from sequenced algal genomes. *Mol Biol Evol* 28 (10): 2921–2933, <https://doi.org/10.1093/molbev/msr124>
- Henze, K., Martin, W., and Schnarrenberger, C. 2002. Endosymbiotic gene transfer. *Horizontal Gene Transfer*. <https://doi.org/10.1016/b978-012680126-2/50034-7>.
- Hinder, S. L., Hays, G.C., Edwards, M., Roberts, E. C., Walne, A. W., and Gravenor, M. B. 2012. Changes in marine dinoflagellate and diatom abundance under climate change. *Nature Climate Change*. 2: 271–275. <https://doi.org/10.1038/nclimate1388>.
- Huysman, M. J. J., Martens, C., Vandepoele, K., Gillard, J., Rayko, E., Heijde, M., Bowler, C. et al. 2010. Genome-wide analysis of the diatom cell cycle unveils a novel type of cyclins involved in environmental signaling. *Genom. Biol.* 11 (2): R17.
- Iverson, V., Morris, R.M., Frazar, C.D., Berhiau, C.T., Morales, R.L., and Armbrust, E.V. 2012. Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science*. 335 (6068): 587-590. <https://doi.org/10.1126/science.1212665>
- Jackson, C., Knoll, A.H., Chan, C.X., and Verbruggen, H. 2018. Plastid phylogenomics with broad taxon sampling further elucidates the distinct evolutionary origins and timing of Secondary Green Plastids. *Sci. Rep.* 8 (1): 1523.
- Jannasch, H. W., and Jones, G.E. 1959. Bacterial populations in sea water as determined by different methods of enumeration. *Limnology and Oceanography* 4 (2): 128–39.
- 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. and Kim, K.Y. Feeding diverse prey as an excellent strategy of mixotrophic dinoflagellates for global dominance. *Sci Adv.* 7 (2), eabe4214. <https://doi.org/10.1126/sciadv.abe4214>
- Johnson, M.D. 2015. Inducible Mixotrophy in the Dinoflagellate *Prorocentrum minimum*. *J. Euk. Microbiol.* 62 (4): 431-443. <https://doi.org/10.1111/jeu.12198>
- Kamikawa, R., Tanifuji, G., Kawachi, M., Miyashita, H., Hashimoto, T., and Inagaki, Y. 2015. Plastid genome-based phylogeny pinpointed the origin of the green-colored plastid in the dinoflagellate *Lepidodinium chlorophorum*. *Genom. Biol. Evol.* 7 (4): 1133–40.
- Karsenti, E., Acinas, S.G. Bork, P., Bowler, C., De Vargas, C., Rae, J., Sullivan, M. et al. 2011. A holistic approach to marine eco-systems biology. *PLoS Biology* 9 (10): e1001177.
- Kazamia, E., Sutak, R., Paz-Yepes, J., Dorrell, R. G., Vieira, F. R. J., Mach, J., Morrissey, J. et al. 2018.

- Endocytosis-mediated siderophore uptake as a strategy for Fe acquisition in diatoms. *Sci. Adv.* 4 (5): 4536.
- Keeling, P. J., Burki, F., Wilcox, H. M., Allam, B., Allen, E. E. Amaral-Zettler, L. A., Armbrust, E. V. et al. 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.
- Kim, E., Harrison, J. W., Sudek, S., Jones, M. D. M., Wilcox, H. M., Richards, T. A., Worden, A. Z., and Archibald, J. M. 2011. Newly identified and diverse plastid-bearing branch on the eukaryotic tree of life. *Proc. Natl. Acad. Sci. USA* 108 (4): 1496–1500.
- Kopf, A., Bick, M., Kottmann, R., Schnetzer, J., Kostadinov, I., Lehmann, K., Fernandez-Guerra, A. et al. 2015. The Ocean Sampling Day consortium. *GigaScience* 4 (6): 27.
- Kowallik, K. V., Stoebe, B., Schaffran, I., Kroth-Pancic, P., and Freier, U. 1995. The chloroplast genome of a chlorophylla c-containing alga, *Odontella sinensis*. *Plant Mol. Biol. Rep.* 13: 336–342 <https://doi.org/10.1007/bf02669188>.
- Kumar, D., Singh, B., and Sharma, Y. C. 2017. Challenges and opportunities in commercialization of algal biofuels. *Algal Biofuels*. https://doi.org/10.1007/978-3-319-51010-1_20.
- Le Bescot, N., Mahé, F., Audic, S., Dimier, C., Garet, M.J., Poulain, J., Wincker, P., de Vargas, C., and Siano, R. 2016. Global patterns of pelagic dinoflagellate diversity across protist size classes unveiled by metabarcoding. *Env. Microbiol.* 18 (2): 609–26.
- Leblanc, K., Quéguiner, B., Diaz, F., Cornet, V., Michel-Rodriguez, M., Durrieu de Madron, X., Bowler, C. et al. 2018. Nanoplanktonic diatoms are globally overlooked but play a role in spring blooms and carbon export. *Nat. Commun.* 9 (1): 953.
- Lee, YH. 2019. Ocean cultural heritage and ocean literacy programs in the UN Decade of Ocean Science for Sustainable Development (2021-2030). *J. Ocean Culture*. <https://doi.org/10.33522/joc.2019.2.136>.
- Li, H., Li, L., Yu, L., Yang, X., Shi, X., Wang, J., Li, J., and Lin, S. 2020. Transcriptome profiling reveals versatile dissolved organic nitrogen utilization, mixotrophy, and N conservation in the dinoflagellate *Prorocentrum shikokuense* under N deficiency. *Sci. Tot. Env.* 143103, <https://doi.org/10.1016/j.scitotenv.2020.143013>.
- Li, W. K. W., McLaughlin, F. A., Lovejoy, C., and Carmack, E.C. 2009. Smallest algae thrive as the Arctic Ocean freshens. *Science* 326 (5952): 539.
- Lommer, M., Specht, M., Roy, A.S., Kraemer, L., Andreson, R., Gutowska, M. A., Wolf, J. et al. 2012. Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genom. Biol.* 13 (7): R66.
- Lopes Dos Santos, A., Gourvil, P., Tragin, M., Noël, M.H., Decelle, J., Romac, S., and Vaultot, D. 2017. Diversity and oceanic distribution of prasinophytes Clade VII, the dominant group of green algae in oceanic waters. *ISME J.* 11 (2): 512–28.
- Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., Wincker, P. et al. 2016. Insights into global diatom distribution and diversity in the world's ocean. *Proc. Natl. Acad. Sci. USA* 113 (11): 1516–25.
- Mangot, J.F., Logares, R., Sánchez, P., Latorre, F., Seeleuthner, Y., Mondy, S., Sieracki, M.E. et al. 2017. Accessing the genomic information of unculturable oceanic picoeukaryotes by combining multiple single cells. *Sci. Rep.* 7 (1): 41498.
- Marcy, Y., Ouverney, C., Bik, E. M., Lösekann, T., Ivanova, N., Martin, H. G. S., Szeto, E. et al. 2007. Dissecting biological 'dark matter' with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proc. Natl. Acad. Sci. USA* 104 (29): 11889–94.
- Margulis, L. 1975. Symbiotic theory of the origin of eukaryotic organelles; criteria for proof. *Symposia Soc. Exp. Biol.* 29: 21–38.
- Massana, R., del Campo, J., Sieracki, M. E., Audic, S., and Logares, R. 2014. Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. *ISME J.* 8 (4): 854–66.

- Matsuzaki, M., Misumi, O., Shin-I, T., Maruyama, S., Takahara, M., Miyagishima, SY., Mori, T. et al. 2004. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* 428 (6983): 653–57.
- McQuaid, J. B., Kustka, A. B., Oborník, M., Horák, A., McCrow, J. P., Karas, B. J., Zheng, H. et al. 2018. Carbonate-sensitive phytotransferrin controls high-affinity iron uptake in diatoms. *Nature* 555 (7697): 534–37.
- Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., et al.. 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science*. 318(5848): 245–50. doi: 10.1126/science.1143609.
- Milliman, J. D. 1993. Production and accumulation of calcium carbonate in the ocean: budget of a nonsteady state. *Global Biogeochem. Cycles* 7 (4): 927–57.
- Mock, T., Otilar, R.P., Strauss, J., McMullan, M., Paaajanen, P., Schmutz, J., Salamov, A. et al. 2017. Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature* 541 (7638): 536–40.
- Monier, A., Worden, A. Z., and Richards, T. A. 2016. Phylogenetic diversity and biogeography of the Mamiellophyceae lineage of eukaryotic phytoplankton across the oceans. *Env. Microbiol. Rep.* 8(4): 461–9. <https://doi.org/10.1111/1758-2229.12390>.
- Moustafa, A., Beszteri, B., Maier, U. G., Bowler, C., Valentin, K., and Bhattacharya, D. 2009. Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* 324 (5935): 1724–26.
- Msuya, F. E., and 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. <https://doi.org/10.1080/09670262.2017.1357084>.
- Nash, K. L., Alexander, K. Melbourne Thomas, J., Novioaglio, C., Sbrocchi, C., Villanueva, C., and Pecl, G. T. 2020. Developing achievable alternate futures for key challenges during the UN Decade of Ocean Science for Sustainable Development. *Preprint*. <https://doi.org/10.22541/au.160166561.14686672>.
- Nelson, W. C., Tully, B. J. and Mobberley, J. M. 2020. Biases in genome reconstruction from metagenomic data. *PeerJ* 8 (10): e10119.
- Nelson, D.R., Hazzouri, K.M., Lauersen, K.J., Jaiswal, A., Chaiboonchoe, A., Mystikou, A., Fu, W. et al. 2021. Large-scale genome sequencing reveals the driving forces of viruses in microalgal evolution. *Cell Host and Microbe*. In press. <https://doi.org/10.1016/j.chom.2020.12.005>.
- Novák Vanclová, A. M. G., Zoltner, M., Kelly, S., Soukal, P., Záhonová, K., Füßsy, Z., Ebenezer, T.E. et al. 2020. Metabolic quirks and the colourful history of the *Euglena gracilis* secondary plastid. *New Phytol.* 225 (4): 1578–92.
- Ohta, N., Sato, N., and Kuroiwa, T. 1998. Structure and organization of the mitochondrial genome of the unicellular red alga *Cyanidioschyzon merolae* deduced from the complete nucleotide sequence. *Nucl. Acids Res.* 26 (22): 5190–98.
- One Thousand Plant Transcriptomes Initiative. 2019. One Thousand Plant Transcriptomes and the Phylogenomics of Green Plants. *Nature* 574 (7780): 679–85.
- Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, PA., Woodcroft, B. J., Evans, P. N., Hugenholtz, P. and Tyson, G. W. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* 2 (11): 1533–42.
- Pfaff, S., Gustavs, L., Reichardt, A., Jaskulke, R., Gäbler-Schwarz H. E., and Karsten, E. 2016. Ecophysiological plasticity in the Arctic phytoplankton species *Phaeocystis pouchetii* (Prymnesiophyceae, Haptophyta). *Algological Studies*. 151–152: 87–102. https://doi.org/10.1127/algol_stud/2016/0260.
- Pierella Karlusich, J.J., Lodeyro, A.F., and Carrillo, N. 2014. The long goodbye: the rise and fall of flavodoxin during plant evolution. *J. Exp. Bot.* 65 (18): 5161–5178. <https://doi.org/10.1093/jxb/eru273>
- Pierella Karlusich, J.J., Ibarbalz, F.M., and Bowler, C. 2020. Exploration of marine phytoplankton: from their historical appreciation to the ‘omics era. Ed: Campbell L. *J. Plankton Res.* 52 (10): 1989.

- Pierella Karlusich, J. J., Ibarbalz, F.M. and Bowler, C. 2020. Phytoplankton in the ocean. *Ann. Rev. Mar. Sci.* 12 (1): 233–65.
- Pires, J. C. M. 2017. COP21: the algae opportunity? *Renew. Sust. Energy Rev* 79: 867-877. <https://doi.org/10.1016/j.rser.2017.05.197>.
- Planes, S., Allemand, D., Agostini, S., Banaigs, B., Boissin, E., Boss, E., Bourdin, G. et al. 2019. The Tara Pacific Expedition-A Pan-Ecosystemic Approach of the ‘-Omics’ Complexity of Coral Reef Holobionts across the Pacific Ocean. *PLoS Biology* 17 (9): e3000483.
- Ponce-Toledo, R. I., Moreira, D., López-García, P., and Deschamps, P. 2018. Secondary plastids of euglenids and chlorarachniophytes function with a mix of genes of red and green algal ancestry. *Mol. Biol. Evol.* 35 (9): 2198–2204.
- Rappé, M. S., and Giovannoni, S. J. 2003. The uncultured microbial majority. *Ann. Rev. Microbiol.* 57: 369–94.
- Rastogi, A., Maheswari, U., Dorrell, R. G., Vieira, F. R. J., Maumus, F., Kustka, A., McCarthy, J. et al. 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.
- Rastogi, A., Vieira, F. R. J., Deton-Cabanillas, AF., Veluchamy, A., Cantrel, C., Wang, G., Vanormelingen, P. et al. 2020. A genomics approach reveals the global genetic polymorphism, structure, and functional diversity of ten accessions of the marine model diatom *Phaeodactylum tricornutum*. *ISME Journal*. 14: 347–363. <https://doi.org/10.1038/s41396-019-0528-3>.
- Raymond, J. A., and Kim H. J. 2012. Possible role of horizontal gene transfer in the colonization of sea ice by algae. *PLoS ONE*. 7 (5): 35968 <https://doi.org/10.1371/journal.pone.0035968>.
- Read, B. A., Kegel, J., Klute, M. J., Kuo, A., Lefebvre, S. C., Maumus, F., Mayer, C. et al. 2013. Pan genome of the phytoplankton *Emiliania* underpins its global distribution. *Nature* 499 (7457): 209–13.
- Reid, P. C., Colebrook, J. M., Matthews, J. B. L., and Aiken, J. 2003. The Continuous Plankton Recorder: concepts and history, from plankton indicator to undulating recorders. *Progress Oceanography*. 58 (2-4): 117-173. <https://doi.org/10.1016/j.pocean.2003.08.002>.
- Reith, M., and Munholland, J. 1995. Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. *Plant Mol. Biol. Reporter*. 13: 333–335. <https://doi.org/10.1007/bf02669187>.
- Reji, L., and Francis. C. A. 2020. Metagenome-Assembled Genomes reveal unique metabolic adaptations of a basal marine Thaumarchaeota lineage. *ISME J*. 14 (8): 2105–15.
- Rossoni, A.W., Price, D.C., Seger, M., Lyska, D., Lammers, P., Bhattacharya, D., and Weber, A.P. 2019. The genomes of polyextremophilic cyanidiales contain 1% horizontally transferred genes with diverse adaptive functions. *eLife* 10: 45017. <https://doi.org/10.7554/eLife.45017.001>
- Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., Wu, D., et al. 2007. The Sorcerer II global ocean sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol.* 5 (3): e77.
- Sarai, C., Tanifuji, G., Nakayama, T., Kamikawa, R., Takahashi, K., Yazaki, E., Matsuo, E. et al. 2020. Dinoflagellates with relic endosymbiont nuclei as models for elucidating organellogenesis. *Proc. Natl. Acad. Sci. USA* 117 (10): 5364–75.
- Schaum, C.E., and Collins, S. 2014. Plasticity predicts evolution in a marine alga. *Phil. Trans. R. Soc. Biol.* 281: 20141486. <https://doi.org/10.1098/rspb.2014.1486>
- Seeleuthner, Y., Mondy, S., Lombard, V., Carradec, Q., Pelletier, E., Wessner, M., Leconte, L. et al. 2018. Single-cell genomics of multiple uncultured stramenopiles reveals underestimated functional diversity across oceans. *Nat. Commun.* 9 (1): 310.
- Ševčíková, T., Horák, A., Klimeš, V., Zbránková, V., Demir-Hilton, E., Sudek, S., Jenkins, J. et al. 2015. Updating algal evolutionary relationships through plastid genome sequencing: did alveolate plastids emerge through endosymbiosis of an ochrophyte? *Sci. Rep.* 5 (5): 10134.
- Sibbald, S. J., and Archibald, J. M. 2017. More protist genomes needed. *Nat. Ecol. Evol.* 1 (5): 145.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov. E. M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.

- Bioinformatics*. 31(19): 3210-2. <https://doi.org/10.1093/bioinformatics/btv351>.
- Simon, N., Cras, A.L., Foulon, E. and Lemée, R. 2009. Diversity and evolution of marine phytoplankton. *Comptes Rendus Biologies*. 332 (2-3): 159-170. <https://doi.org/10.1016/j.crv.2008.09.009>.
- Singer, D., Seppey, C. V. W., Lentendu, G., Dunthorn, M., Bass, D., Belbahri, L., Blandenier, Q. et al. 2020. Protist taxonomic and functional diversity in soil, freshwater and marine ecosystems. *Env. Int.* 146 (11): 106262.
- Skokan, R., Medvecká, E., Viaene, T., Vosolsobě, S., Marta Zwiewka, Karel Müller, Petr Skůpa, et al. 2019. PIN-Driven Auxin Transport Emerged Early in Streptophyte Evolution. *Nature Plants* 5 (11): 1114–19.
- Smith, D. Roy, and Sanitá Lima, M. 2017. Unraveling chloroplast transcriptomes with ChloroSeq, an organelle RNA-Seq bioinformatics pipeline. *Briefings Bioinform.* 18 (6): 1012–16.
- Smith, S. R., Dupont, C. L., McCarthy, J. K., Broddrick, J. T., Obornik, M., Horák, A., Füssy, Z. et al. 2019. Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. *Nat. Commun.* 10 (1): 4552.
- Sournia, A. 1986. Atlas du phytoplancton marin: introduction, cyanophycées, dictyochophycées, dinophycées et raphidophycées. Centre National de la Recherche Scientifique.
- Southward, A. J., Langmead, O., Hardman-Mountford, N. J., Aiken, J., Boalch, G. T., Dando, P. R., Genner, M. J. et al. 2005. Long-term oceanographic and ecological research in the western English Channel. *Adv. Mar. Biol.* 47: 1–105.
- Steidinger, K. A., and Jangen, K. 1997. Dinoflagellates. In *Identifying Marine Phytoplankton*, 387–584.
- Stern, R., Kraberg, A., Bresnan, E., Kooistra, W. H. C., Lovejoy, C., Montresor, M. Morán, X. A. G. et al. 2018. Molecular analyses of protists in long-term observation programmes- current status and future perspectives. *J. Plankton Res.* <https://doi.org/10.1093/plankt/fby035>.
- Stiller, J. W., Schreiber, J., Yue, J., Guo, H., Ding, Q. and Huang, J. 2014. The evolution of photosynthesis in chromist algae through serial endosymbioses. *Nat. Commun.* 5 (12): 5764.
- Sunagawa, S., Tara Oceans Coordinators, Acinas, S.G., Bork, P., Bowler, C., Eveillard, D., Gorsky, G. et al. 2020. Tara Oceans: towards global ocean ecosystems biology. *Nat. Rev. Microbiol.* 18: 424-445. <https://doi.org/10.1038/s41579-020-0364-5>.
- Terrado, R., Monier, A., Edgar, R., and Lovejoy, C. 2015. Diversity of Nitrogen Assimilation Pathways among Microbial Photosynthetic Eukaryotes. *J. Phycology*. <https://doi.org/10.1111/jpy.12292>.
- Tréguer, P., Bowler, C., Moriceau, B., Dutkiewicz, S., Gehlen, M., Aumont, O., Bittner, L. et al. 2018. Influence of diatom diversity on the ocean biological carbon pump. *Nat. Geosci.* 11: 27-38. <https://doi.org/10.1038/s41561-017-0028-x>.
- Tully, B. J., Graham, E. D., and Heidelberg, J. F. 2018. The reconstruction of 2,631 Draft metagenome-assembled genomes from the global oceans. *Sci. Data* 5 (1): 170203.
- Tully, B. J., Sachdeva, R., Graham, E. D., and Heidelberg, J. F. 2017. 290 Metagenome-assembled genomes from the Mediterranean Sea: a resource for marine microbiology. *PeerJ* 5 (7): e3558.
- Tyrrell, T., and Merico, A. 2004. *Emiliania huxleyi*: bloom observations and the conditions that induce them. *Coccolithophores*. https://doi.org/10.1007/978-3-662-06278-4_4.
- Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., Solovyev, V. V., Rubin, E. M., Rokhsar, D. S. and Banfield, J. F. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428 (6978): 37–43.
- Umen, J. G. 2014. Green Algae and the origins of multicellularity in the plant kingdom. *Cold Spring Harbor Perspectives in Biology* 6 (11): 016170.
- Unrein, F., Gasol, J. M., Not, F., Forn, I., and Massana, R. (2014). Mixotrophic haptophytes are key bacterial grazes in oligotrophic coastal waters. *ISME J.*, 8 :164–176
- Vancaester, E., Depuydt, T., Osuna-Cruz, C. M., and Vandepoele, K. 2020. Comprehensive and functional analysis of Horizontal Gene Transfer Events in diatoms. *Mol. Biol. Evol.*, 37 (11): 3243-3257. <https://doi.org/10.1093/molbev/msaa182>.
- Vannier, T., Leconte, J., Seeleuthner, Y., Mondy, S., Pelletier, E., Aury JM., de Vargas, C. et al. 2016. Survey of the green picoalga *Bathycoccus* genomes in the global ocean. *Sci. Rep.* 6 (11): 37900.

- de Vargas, C., Audic, Henry, S., Decelle, H. J., Mahé, F., Logares, R., Lara, E. et al. H.Johan Decelle, Frédéric Mahé, Ramiro Logares, Enrique Lara, et al. 2015. Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. *Science* 348 (6237): 1261605.
- Vaulot, D., Le Gall, F., Marie, D., Guillou, L., and Partensky, F. 2004. The Roscoff Culture Collection (RCC): a collection dedicated to marine picoplankton. *Nova Hedwigia* 79(1-2): 49-70. <https://doi.org/10.1127/0029-5035/2004/0079-0049>.
- Wang, D., Yu, X., Xu, K., Bi, G., Cao, M., Zelzion, E., Fu, C. et al. 2020. *Pyropia yezoensis* genome reveals diverse mechanisms of carbon acquisition in the intertidal environment. *Nat. Commun.* 11 (1): 4028.
- West, P. T., Probst, A. J., Grigoriev, I. V., Thomas, B. C., and Banfield, J. F. 2018. Genome reconstruction for eukaryotes from complex natural microbial communities. *Genom. Res.* 28 (4): 569–80.
- Willhelm, C., Büchel, C., Fisahn, J., Goss, R., Jakob, T., Laroche, J., Lavaud, J., Lohr, M., Riebesell, U., Stehfest, K., Valentin, K., and Kroth, P.G., 2006. The regulation of carbon and nutrient assimilation in diatoms is significantly different from green algae. *Protist* 157(2):91-124. doi: 10.1016/j.protis.2006.02.003.
- Wolff, G., I. Plante, B. F. Lang, U. Kück, and G. Burger. 1994. Complete sequence of the mitochondrial DNA of the chlorophyte alga *Prototheca wickerhamii*. Gene content and genome organization. *J. Mol. Biol.* 237 (1): 75–86.
- Yang, M., Lin, X., Liu, X., Zhang, J. and Ge, F. 2018. Genome annotation of a model diatom *Phaeodactylum tricornutum* using an integrated proteogenomic pipeline. *Mol. Plant* 11 (10): 1292–1307.
- Yoon, H. S., Hackett, J. D., and 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. <https://doi.org/10.1073/pnas.172234799>
- Yoon, H. S., Price, D.C. Stepanauskas, R., Rajah, V. D., Sieracki, M. E., wilson, W. H., Yang, E. C., Duffy, S., and Bhattacharya, D. 2011. Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science* 332 (6030): 714–17.
- Zimmermann, R., Iturriaga, R. and Becker-Birck, J. 1978. Simultaneous determination of the total number of aquatic bacteria and the number thereof involved in respiration. *Appl. Env. Microbiol.* 36: 926–35.