

Review



Cite this article: Benoiston A-S, Ibarbalz FM, Bittner L, Guidi L, Jahn O, Dutkiewicz S, Bowler C. 2017 The evolution of diatoms and their biogeochemical functions. *Phil. Trans. R. Soc. B* **372**: 20160397.
<http://dx.doi.org/10.1098/rstb.2016.0397>

Accepted: 24 March 2017

One contribution of 16 to a theme issue
'The peculiar carbon metabolism in diatoms'.

Subject Areas:

ecology, evolution, genomics

Keywords:

biogeochemistry, carbon export, diatom,
geological record, genomics, photosynthesis

Author for correspondence:

Chris Bowler
e-mail: cbowler@biologie.ens.fr

[†]These authors contributed equally to this study.

The evolution of diatoms and their biogeochemical functions

Anne-Sophie Benoiston^{1,†}, Federico M. Ibarbalz^{2,†}, Lucie Bittner¹,
Lionel Guidi^{3,4}, Oliver Jahn⁵, Stephanie Dutkiewicz⁵ and Chris Bowler²

¹Sorbonne Universités, UPMC Univ. Paris 06, Univ. Antilles, Univ. Nice Sophia Antipolis, CNRS, Evolution Paris Seine - Institut de Biologie Paris Seine (EPS - IBPS), 75005 Paris, France

²Ecole Normale Supérieure, PSL Research University, Institut de Biologie de l'École Normale Supérieure (IBENS), CNRS UMR8197, INSERM U1024, 46 rue d'Ulm, 75005 Paris, France

³Sorbonne Universités, UPMC Univ. Paris 06, CNRS, Laboratoire d'Océanographie de Villefranche (LOV) UMR7093, Observatoire Océanologique, 06230 Villefranche-sur-Mer, France

⁴Department of Oceanography, University of Hawaii, Honolulu, HI 96822, USA

⁵Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, 54-1514 MIT, Cambridge, MA 02139, USA

CB, 0000-0003-3835-6187

In contemporary oceans diatoms are an important group of eukaryotic phytoplankton that typically dominate in upwelling regions and at high latitudes. They also make significant contributions to sporadic blooms that often occur in springtime. Recent surveys have revealed global information about their abundance and diversity, as well as their contributions to biogeochemical cycles, both as primary producers of organic material and as conduits facilitating the export of carbon and silicon to the ocean interior. Sequencing of diatom genomes is revealing the evolutionary underpinnings of their ecological success by examination of their gene repertoires and the mechanisms they use to adapt to environmental changes. The rise of the diatoms over the last hundred million years is similarly being explored through analysis of microfossils and biomarkers that can be traced through geological time, as well as their contributions to seafloor sediments and fossil fuel reserves. The current review aims to synthesize current information about the evolution and biogeochemical functions of diatoms as they rose to prominence in the global ocean.

This article is part of the themed issue 'The peculiar carbon metabolism in diatoms'.

1. Introduction

Microscopic photosynthetic plankton (phytoplankton) provide the organic biomass on which almost all ocean life depends and fuel a range of essential biogeochemical processes, ranging from the generation of oxygen, the recycling of elemental nutrients, and the removal of carbon dioxide from the atmosphere. They are responsible for around 45% of global primary production and yet represent only 1% of Earth's photosynthetic biomass [1], due to their rapid proliferation times and because all cells are photosynthetically active, unlike multicellular plants. Our appreciation of the roles of these microscopic organisms in the ocean has been transformed over the last decades by improved methods to explore the chequered history of life on Earth and by new DNA sequencing technologies. Scientists are using these resources to address the feedbacks between plankton and the climate system, because planktonic organisms can both influence climate and be affected by climate change [2]. As a major component of plankton communities in today's oceans diatoms are now key to their functioning, yet they rose to prominence only quite recently. Through photosynthesis they provide large amounts of organic material that sustains marine ecosystems as well as contributing to Earth's carbon cycle, and play major roles in the biogeochemical cycling of other nutrients such as nitrogen and silicon [3–5]. Their evolution can be traced back to the origin of photosynthesis.

2. Photosynthesis as the engine of life

Oxygenic photosynthesis is arguably the most important process in nature. It boosted the remarkable history of life on Earth following its appearance at least 2.4 billion years ago [6] (figure 1*a*). In spite of its early evolution it represents the most complex energy transduction system known; its water oxidizing machine has no analogues elsewhere and its functioning is still poorly understood [16]. The oxidizing or ‘splitting’ of water was made possible by the coupling of two photosystems that enabled oxygenic photosynthetic bacteria to use light energy to generate oxygen from water and reducing power in the form of NADPH. The oxygen generated from the process subsequently accumulated in the atmosphere and is one of Earth’s distinguishing features, because molecular oxygen is extremely rare in the Universe [17]. The utilization of light energy to split water in oxygenic photosynthesis also allows the fixation of CO₂ into organic matter that fuels the food chain.

Oxygenic photosynthesis first evolved in the cyanobacteria, which remain the only prokaryotes capable of performing it. Oxygen initially began to accumulate only slowly in the atmosphere because it was first consumed in oxidation reactions with abundant compounds that contained reduced forms of iron, sulfur, carbon, nitrogen, and other abundant materials, and because it was consumed in the biological process of respiration, which evolved after photosynthesis [8].

Following the evolution of oxygenic cyanobacteria it took around 2 billion years before complex multicellular animal life evolved (figure 1*a*). During this time, eukaryotic organisms appeared bearing the first mitochondria derived from the endosymbiosis of a proteobacterium in an Archaeal-like cell, in which respiratory processes could occur [18]. Unambiguous fossils of eukaryotes have been found in shales as old as 1.65–1.85 billion years [19]. Subsequently, chloroplasts evolved following the invasion or engulfment of a cyanobacterium into the prototypic eukaryote. Photosynthetic eukaryotes are considered to have evolved around 1.2 billion years ago [12] although the forms that dominate today’s ocean are predominantly derived from additional or ‘secondary’ endosymbiotic events in which eukaryotic green or red algae were incorporated a second time into a eukaryotic cell [20]. The timing of these events is not well resolved but it certainly happened prior to the appearance of multicellular lifeforms during the Cambrian explosion and preceded a major increase in atmospheric oxygen to levels similar to those found today, from around 1–5% to about 20% (figure 1*a*). The reason why the rise of photosynthetic eukaryotes stimulated such a dramatic increase in oxygen may be a consequence of carbon export to the seafloor, because their larger cells were more strongly ballasted and therefore more likely to sink than cyanobacteria [2] (figure 2). The consequent burial of carbon sequestered it away from the carbon cycle and so it could not be remineralized back to CO₂ by oxidative respiration. Alternatively (or additionally), photosynthetic activity may have increased significantly following the evolution of extensive planktonic ecosystems, e.g. fuelled by increased nutrient availability during this period. Regardless of the cause, atmospheric CO₂ levels dropped significantly during this period, which may have contributed to one or more of the Snowball Earth events that have been documented to have occurred [13]

(figure 1*a,b*), because CO₂ is a powerful greenhouse gas. Furthermore, the increase in molecular oxygen was probably instrumental in permitting multicellular life to evolve during more temperate periods because it allowed the development of more complex organisms less constrained by oxygen acquisition from a low oxygen environment.

Atmospheric oxygen concentrations have remained relatively stable at around 20% since the early Cambrian period. The emergence of land plants during the Devonian around 400 million years ago (Ma) likely led however to a further large increase in oxygen concomitant with CO₂ draw-down from the atmosphere (figure 1*b*) [23]. Although it did not persist, the elevated oxygen concentrations may have led to the evolution of giant insects and other large animals. Atmospheric oxygen in today’s world, while being similar to concentrations prior to the evolution of land plants, is now likely to be maintained principally by terrestrial plants that release oxygen directly to the atmosphere rather than by photosynthetic plankton because the oxygen generated within the water column is likely to be consumed by other organisms rather than being outgassed [24,25]. The release of biogenic oxygen from the ocean may nonetheless be significant in some regions and is likely to be sensitive to temperature changes [26].

The detailed analysis of the geological record left by dead eukaryotic plankton sinking to the seafloor over the last hundreds of millions of years, either based on biomarker molecules or microfossils, has revealed their history during major Earth transitions [27]. By likely underpinning the rise of oxygen that led to the evolution of multicellular organisms, they may have promoted the development of ever more complex lifeforms, not only in the ocean, but also on land as well. Besides the process of photosynthesis, the later appearance of calcification and silicification in some phytoplanktonic organisms (e.g. in coccolithophorids and diatoms, respectively), in addition to more ancient organisms such as foraminifers and radiolarians, permitted the precipitation of hard materials to the ocean interior, as well as organic carbon (figure 1*a,b*). A rich amount of data from microfossils, biomarkers, and molecular clocks using conserved marker genes indicate that these processes appeared in photosynthetic organisms around 200 Ma and permitted atmospheric CO₂ to be further sequestered into the deep ocean in the form of organic carbon and calcium carbonate, which over time contributed to the formation of sedimentary rocks such as limestones and cherts, as well as our oil and gas reserves [12,28–30] (figures 1 and 2). This, together with increased weathering and changes in ocean circulation, is believed to have initiated a period of declining atmospheric CO₂ concentrations, contributing to the switch from a greenhouse climate in the Mesozoic to an icehouse climate in the Cenozoic [31]. The concomitant increase in atmospheric O₂ (figure 1*a,b*) almost certainly contributed to the evolution of large animals, including placental mammals that have very high metabolic demands [29,32,33].

3. The rise of the diatoms

The composition of eukaryotic phytoplankton in the modern ocean is dominated by diatoms, dinoflagellates and coccolithophores [12]. Through photosynthesis and calcification these organisms make a small but significant contribution

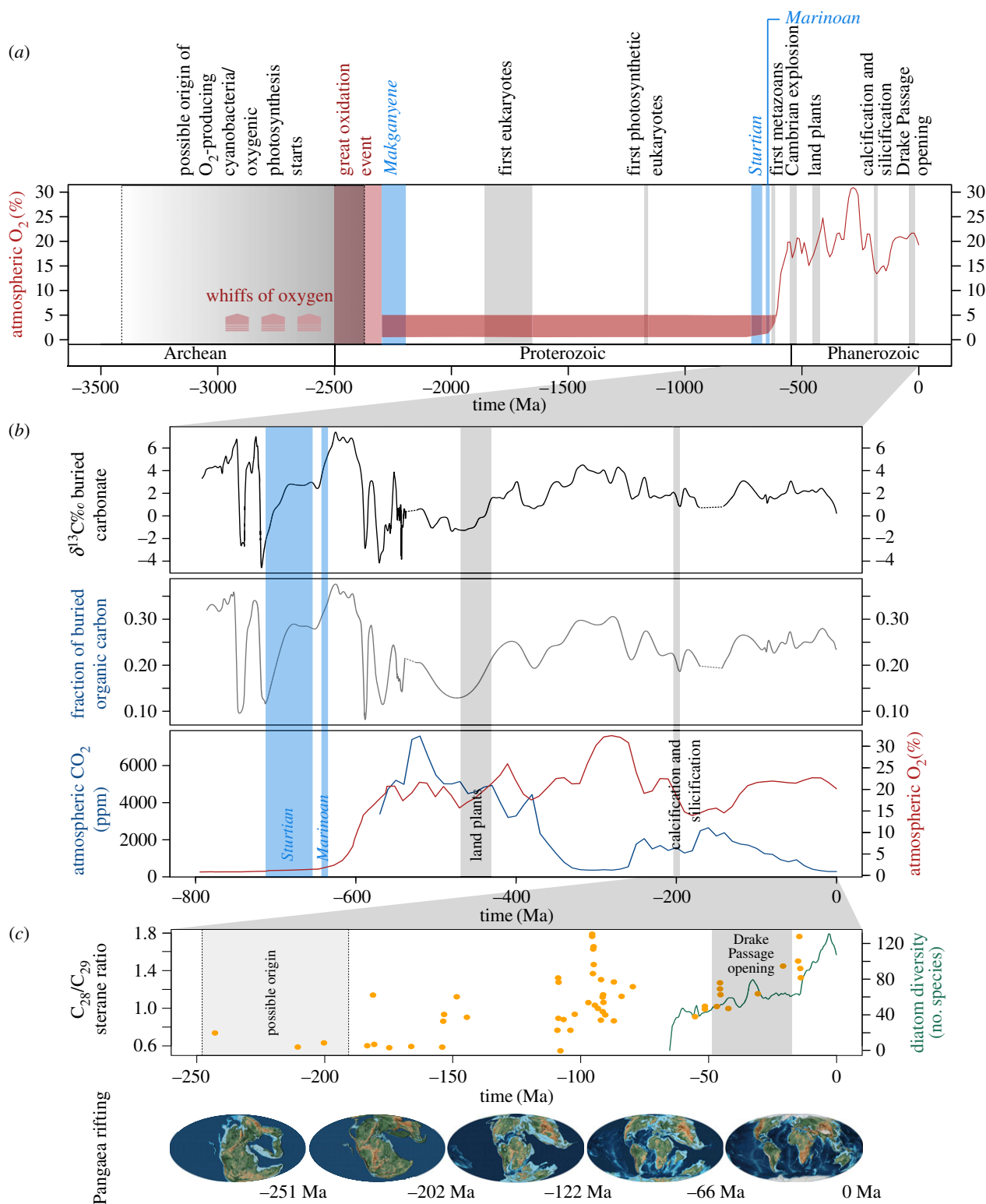


Figure 1. Major evolutionary and biogeochemical events during the history of life on Earth. (a) Trends since the evolution of oxygenic photosynthesis, (b) trends during the last 800 million years, and (c) diatom diversity and abundance data with respect to Pangaea rifting during the last 260 million years. Atmospheric O_2 was modified from Holland [7] according to Lyons *et al.* [8]; it is compared to $\delta^{13}C$ of carbonates [9], fraction of buried organic carbon [9], atmospheric CO_2 [10], diatom diversity [11] and C_{28}/C_{29} sterane ratios [12], which is a geochemical proxy for diatom abundance. Snowball Earth events are shown in light blue and were taken from Hoffman and Kopp *et al.* [13,14]. Pangaea rifting is illustrated with maps taken from the PALEOMAP Project [15]. The grey ranges on the plots represent the estimated span of the events cited in the text. Note that because the oldest direct measurements of atmospheric O_2 come from Pleistocene ice cores, all the detail in the Phanerozoic curve is based on models. Prior to that we have represented the views of Lyons *et al.* [8]: no stable O_2 trends before the Great Oxidation Event, some atmospheric O_2 (1–5%) through most of the Proterozoic, and then a rise to more or less modern values from the Ediacaran to the Silurian. The case is strong that pO_2 during the Carboniferous was higher than today's, but other details in the Phanerozoic curve are conjectural.

(probably around 10%) to the regulation of the partial pressure of carbon dioxide in the upper ocean [34,35]. The other 90% of oceanic carbon is derived from the physico-chemically regulated solubility of CO_2 , which generates

carbonate ions in the upper ocean [36]. The biological draw-down of atmospheric CO_2 through the activity of photosynthetic organisms in the ocean is known as the biological carbon pump which results in the generation of

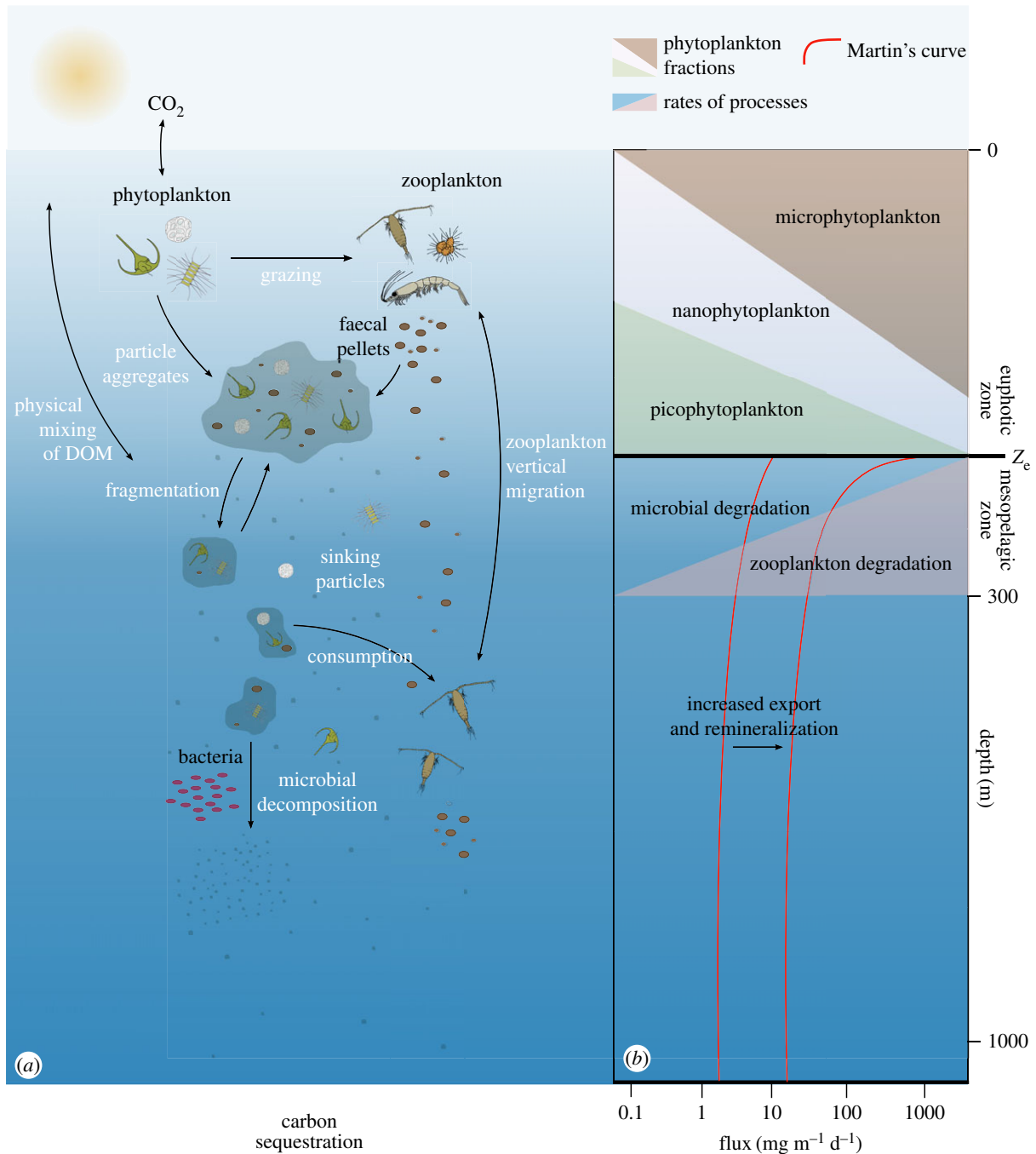


Figure 2. The biological carbon pump in the ocean. (a) Inorganic atmospheric carbon (CO_2) is transformed into organic carbon by phytoplankton in the euphotic zone. This carbon is then grazed on by heterotrophic organisms. A fraction of it is exported out of the surface layer as particulate organic carbon (such as dead organic material, faecal pellets produced by the zooplankton and aggregates of these materials) that sinks in the water column. Two other major processes help with the transfer of carbon below the surface layer: physical mixing of dissolved organic material (DOM) and transport by zooplankton vertical migration. (b) Different processes that can affect the decrease in the flux of particles in the ocean (adapted from [21]). The dimensions of the different areas represent the relative importance of phytoplankton fractions or rates of processes. The variation of the estimated flux with depth was modelled by fitting the Martin power relationship [22]. Carbon export is influenced by the phytoplankton composition in the euphotic zone: export is high when microphytoplankton (including diatoms) dominate the plankton community in the euphotic zone, while low export values correspond to systems dominated by picophytoplankton. Z_e = depth of the euphotic zone. Note that depth, organisms and particle sizes are not to scale.

organic matter that can be consumed by other organisms, as well as calcium carbonate (figure 2). The biological carbon pump exports approximately $5\text{--}12 \text{ PgC yr}^{-1}$ from the surface to the mesopelagic layer, from which approximately 0.2 PgC yr^{-1} is stored in sediment for millennia [34,35], thus contributing to the vertical gradient of carbon in the ocean. The process also results in biological feedback on atmospheric CO_2 and thus the Earth's climate [37,38]. This

structuring of the carbon cycle in the ocean appears to have been established as the three phytoplankton groups rose to prominence in the Mesozoic Era, perhaps as a consequence of the availability of ecological space populated previously by taxa that did not survive the Permian–Triassic mass extinction event, which was Earth's most severe extinction event (resulting in the loss of around 96% of all marine species) [39].

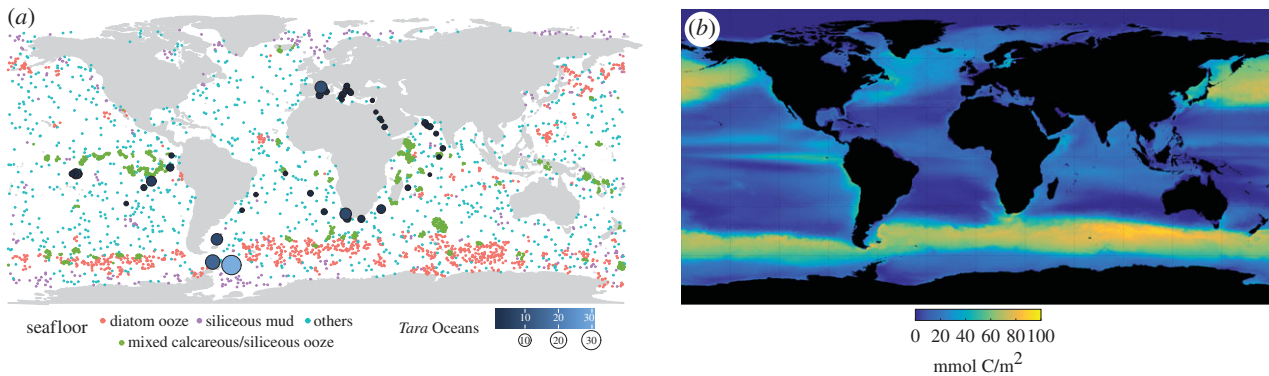


Figure 3. Extent of diatom-rich sediments compared with the distribution of modern diatoms in the ocean. Biosiliceous oozes are present in regions that, still today, are largely dominated by diatoms, in particular the Southern Ocean. (a) Small dots represent seafloor sediment samples defined as containing predominantly diatom ooze, siliceous mud, mixed calcareous/siliceous ooze, or others. Circles of varying size and blue colour correspond to diatom relative abundances determined by the *Tara Oceans* survey (modified from Dutkiewicz *et al.* and Malviya *et al.* [47,48]). (b) Water column inventory of diatom biomass (mmol C/m²) from a biogeochemical/ecosystem simulation (modified version of Dutkiewicz *et al.* [49]).

The fossil record left behind by the elaborate siliceous shells of diatoms indicates that they remained minor components in the ocean until the Cretaceous [31,40], when the supercontinent Pangaea began to break apart into the continents we know today and the major ocean basins were formed (figure 1c). As well as creating more space in marine ecosystems, the rifting of Pangaea was accompanied by the delivery of more nutrients to the oceans because it was concomitant with continental elevation. The increase in nutrients favoured the selection of large-celled phytoplankton that lived along the continental margins such as diatoms [31,41–43]. Following the mass extinction event at the Cretaceous/Paleogene boundary (65 Ma), the diatoms continued to expand and further populate the oceans. In contrast to dinoflagellates and coccolithophores, diatom diversity continued to increase through the Cenozoic; in particular two pulses of diversification occurred at the Eocene/Oligocene boundary interval (33.9 Ma) and the middle to late Miocene (5–20 Ma) [44] (figure 1c). Environmental changes such as sea-level rise, silicate bioavailability, predation, ocean chemistry, increased latitudinal thermal gradients and circulation all likely played a role in driving such diversification [31,41,42]. As one case in point, correlations between increased diatom abundance and carbon export to the deep ocean with reductions in atmospheric CO₂ and reduced temperatures during the opening of the Drake Passage between 19 and 49 Ma suggest that the resulting Antarctic Circumpolar Current may have generated a highly favourable environment for diatom proliferation in the Southern Ocean, that today is still characterized by diatom-rich plankton communities [45,46].

Diatoms today are found throughout the world's oceans, wherever there is sufficient light and nutrients (figure 3). They typically dominate well-mixed coastal and upwelling regions, where the organic carbon they generate supports productive fisheries such as in the Peruvian and Benguela upwelling systems. They appear well adapted to surviving long periods of nutrient and light limitation and often dominate oceanic spring blooms because they can divide more rapidly than other phytoplankton when conditions become favourable for growth, at least as long as silicon is not limiting [50]. They also dominate at high latitudes and in polar environments, in particular along the sea-ice edge where

other photosynthetic organisms are rare, making the Arctic and Southern Ocean ecosystems especially dependent on them [3,4] (figure 3). Their importance for the biogeochemistry of these regions over geological time periods is evidenced by the enormous deposits of siliceous mud and oozes more than 1 km thick in places [47] (figure 3a). The rise of diatoms in the last few millions of years is accompanied by the establishment of the main petroleum source rocks, derived from carbon export. The often spatial coincidence of silica and fossil fuels, together with the worldwide survey of biomarkers (such as 24-norcholestane or C₂₈–C₂₉ steranes) in sediments and source rocks, indicate a crucial role of diatoms in the formation of today's reserves [44]. Moreover, several petroleum basins overlap with regions where diatoms thrive, such as oceanic coastal environments and the Arctic Ocean [51]. Although previous assessments suggest that petroleum source rocks are relatively low in abundance in the Southern Ocean [52], this region may hold significant resources as well.

4. Diatom evolution through the lens of genomics

While sedimentary rocks and the biomarkers within them provide a coarse-grained record of the intertwined histories of life, geology and climate, the evolutionary trajectories of different organisms can best be found by finding remnants of them in their genome sequences (for example, see [53]).

Already prior to the advent of genome sequencing, biochemical and ultrastructural data had provided persuasive evidence that diatoms were derived from a secondary endosymbiotic event involving a red alga that had occurred sometime between 1200 and 700 Ma (figures 1a and 4) and that was common to all stramenopiles, the phylum in which diatoms sit, as well as the chromalveolate supergroup of eukaryotes that includes dinoflagellates and coccolithophores [55–57]. The diatom genome sequences analysed to date do provide support for a red algal endosymbiont [58,59], but the abundance of genes apparently derived from a green algal source has led to the controversial hypothesis that a green algal endosymbiont preceded the red alga and that many of its genes were retained prior to the arrival

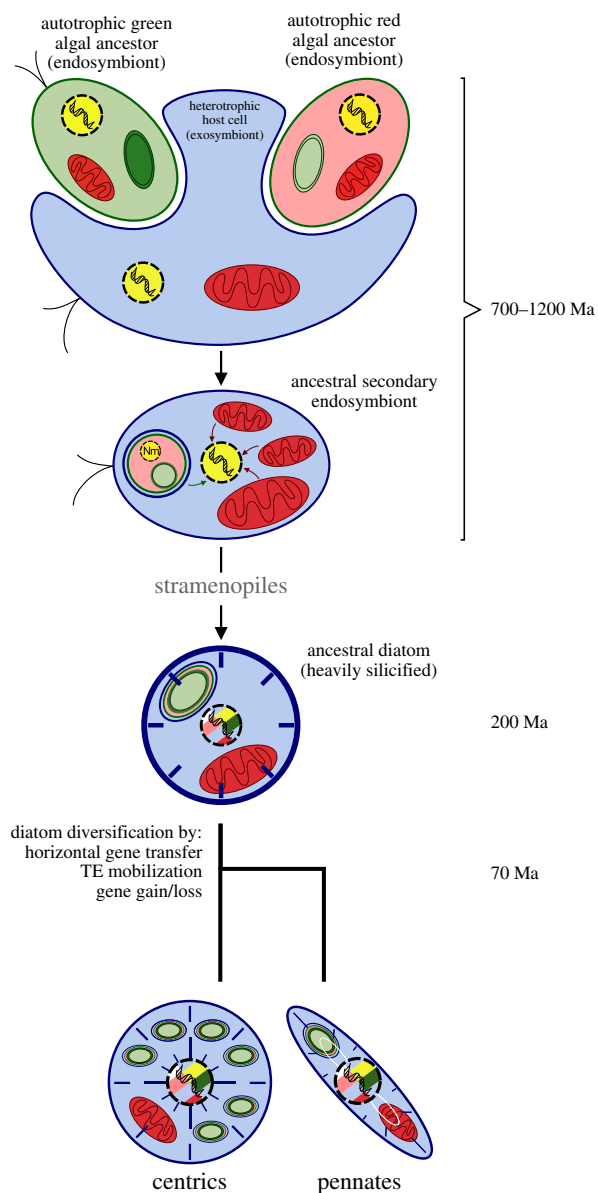


Figure 4. Diatom evolution through the lens of genomics. Major events during the evolution of diatoms from secondary endosymbionts are shown, together with approximate dates. TE, transposable element. The estimated time of separation between pennate and centric diatoms at around 70 million years (70 Ma) is based on Chacón-Baca *et al.* [54]. Figure modified from Bowler *et al.* [3].

of the red alga, whereas the red algal genes that were acquired later were not [60,61] (figure 4). In such a scenario, diatoms (and other photosynthetic chromalveolates) bear red algal-derived chloroplasts driven to a significant extent by green algal genes encoded in the nucleus, which may have provided a selective advantage in ocean environments and thus underlie why such organisms have come to dominate in the ocean whereas photosynthetic organisms harbouring green algal-derived plastids dominate terrestrial habitats [62,63].

An additional feature is the presence of several hundreds of bacterial genes scattered throughout diatom genomes [59], representing around 5% of total gene content. Many such genes appear to have ancient origins because they are shared among several diatoms, and encode functions essential for diatom biology [3]. Diatom-specific transposable elements additionally appear to have been instrumental in generating the rich diversity of species found today [3,64] (figure 4).

The chimeric nature of diatom genomes has brought together unique combinations of genes that collectively encode non-canonical pathways of nutrient assimilation and metabolite management, including for a urea cycle that is integral to nitrogen metabolism [65], and a novel configuration coupling photosynthesis and respiration between diatom chloroplasts and mitochondria [66]. The combined findings have profound and unanticipated implications for our understanding of the role of diatoms in biogeochemical cycles, and highlight the utility of genome sequences for revealing an organisms' metabolic potential. Diatom genomes have furthermore been found to encode large numbers of cyclins [67], key regulators of cell division, that may underlie their impressive proliferative capacity during oceanic blooms, as well as specialized stress-responsive light-harvesting chlorophyll-binding proteins that may be of particular importance for survival in polar-adapted diatoms [68,69].

The peculiarities of the diatom toolbox used to manage silicon metabolism and to generate their silicified cell walls are also being revealed (e.g. [70]), and it is emerging that such processes are deeply integrated within diatom primary metabolism, e.g. for the generation of frustule-localized long chain polyamines as offshoots of the urea cycle [65,71,72]. Notwithstanding, genomics has yet to reveal anything about what ecological or physiological advantages are associated with frustule biogenesis.

The extension of findings from genomics to natural environments will likely reveal further innovations [73]. Evidence is already emerging that some diatoms may have evolved permanent genome-level adaptations to certain conditions (e.g. related to iron bioavailability [74]) whereas others have retained the ability to acclimate to a wider range of conditions through more flexible responses at the transcriptional level [75]. The recent evaluation of the importance of epigenetic processes mediated at the level of DNA methylation or chromatin structural changes [76,77] will reveal whether diatoms have retained or acquired other features from their ancestors that permit additional opportunities for responding to a fluctuating environment over shorter timescales than are operative over macroevolutionary timescales [78].

5. Diatoms in the contemporary oceans

For decades, morphological studies have revealed diatoms to be one of the most ecologically important groups of phytoplankton in the modern oceans and one of the largest components of marine biomass [30,79,80]. More recent environmental omics studies have confirmed this. In particular, in the metabarcoding survey based on the V9 hypervariable region of 18S rDNA performed as part of the *Tara* Oceans global plankton sampling campaign, diatoms are the most abundant group of obligate photosynthetic eukaryotes and the fifth most abundant group of marine eukaryotes [48,81]. Moreover, in some Antarctic stations they represent more than 25% of the sequenced metabarcodes. Metabarcoding studies have allowed a refinement of the diversity estimation and the biogeographic distribution of diatoms even at the genus and species level. Meanwhile, metagenomics and metatranscriptomics data (unpublished results from the *Tara* Oceans consortium) will deepen our knowledge about the role of diatoms in the modern ocean.

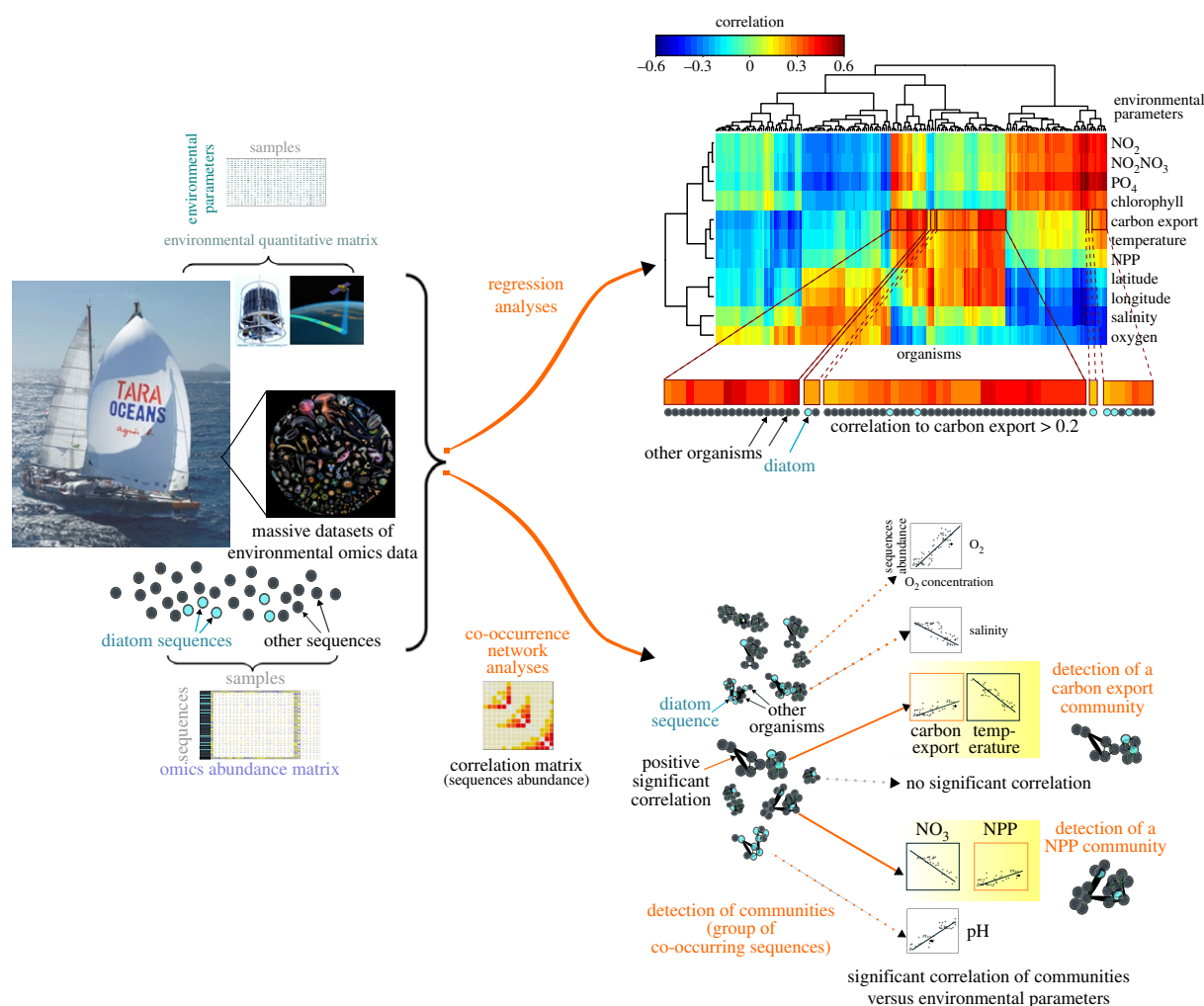


Figure 5. A new context for marine ecosystems biology. High-throughput sequencing technologies nowadays allow the production of massive omics datasets (metagenomics, meta-transcriptomics, meta-barcoding datasets) from microbial planktonic communities. Such massive datasets can be analysed in parallel with large quantitative matrices of environmental data, and have notably produced a first global picture of microbial organisms correlating with carbon export in oligotrophic oceans [83]. In the framework of the biological carbon study, bioinformatics analyses thus help to establish on one hand the list of the most correlated lineages to NPP (Net Primary Production) or carbon export (e.g. using regression analyses such as sparse partial least square (sPLS) analysis; upper part of the figure), or to detect the communities linked to NPP and carbon export (e.g. using co-occurrence network analyses such as weighted correlation network analyses; lower part of the figure; see more details on methods in Guidi *et al.* [83]). Today, the roles of diatoms in the oceans can thus be considered in a global and integrative context.

In terms of their biogeochemical roles, diatoms are believed to be the principal contributors of primary production and carbon export among all photosynthetic organisms in the modern oceans, in particular because of their dominance in highly productive regions [1,5]. Estimates based on time-series of surface chlorophyll from the SeaWiFS Project indicate that microphytoplankton (mostly diatoms) may contribute up to 70% of the net primary production in coastal upwelling systems and 50% in temperate and sub-polar regions during the spring-summer seasons [82]. Overall, diatoms are estimated to contribute around 40% of the total primary production in the oceans, and therefore around one fifth of all the photosynthesis on Earth, similar to all terrestrial rainforests combined [1]. Similarly, both carbon export and remineralisation variations at global scale seem to be partially explained by the phytoplankton community where diatoms and their resting spores may play critical roles [21,83,84]. Diatoms are also a key component in the biogeochemical cycling of silicon (reviewed extensively in [85]).

The combination of genomics data collected during the *Tara Oceans* expedition with ancillary environmental data allows a new framework, summarized in figure 5, to pinpoint

the importance of individual planktonic groups in specific processes, in a holistic context of the entire plankton communities that they are part of. Such network-based methods have already been used to disentangle the key players in euphotic zone communities related to carbon export to deeper layers in the oligotrophic ocean [83] (figure 5). Regression-based analyses on the entire eukaryotic metabarcoding dataset currently available from *Tara Oceans* [81] reveal the dominant roles of diatoms in contributing to net primary production and carbon export, in particular in areas characterized by low temperatures, high oxygen and nutrient concentrations (figure 6). It should therefore be possible to test the robustness of these results by bioinformatic analysis and to further disentangle the roles of diatoms in marine ecosystems using more extended datasets. Such studies could also be performed in the context of different climate simulations to better understand how diatoms affect the carbon cycle and climate regulation.

6. Perspectives

Incontrovertible evidence shows that the Earth's climate has begun to change markedly over the last decades as a

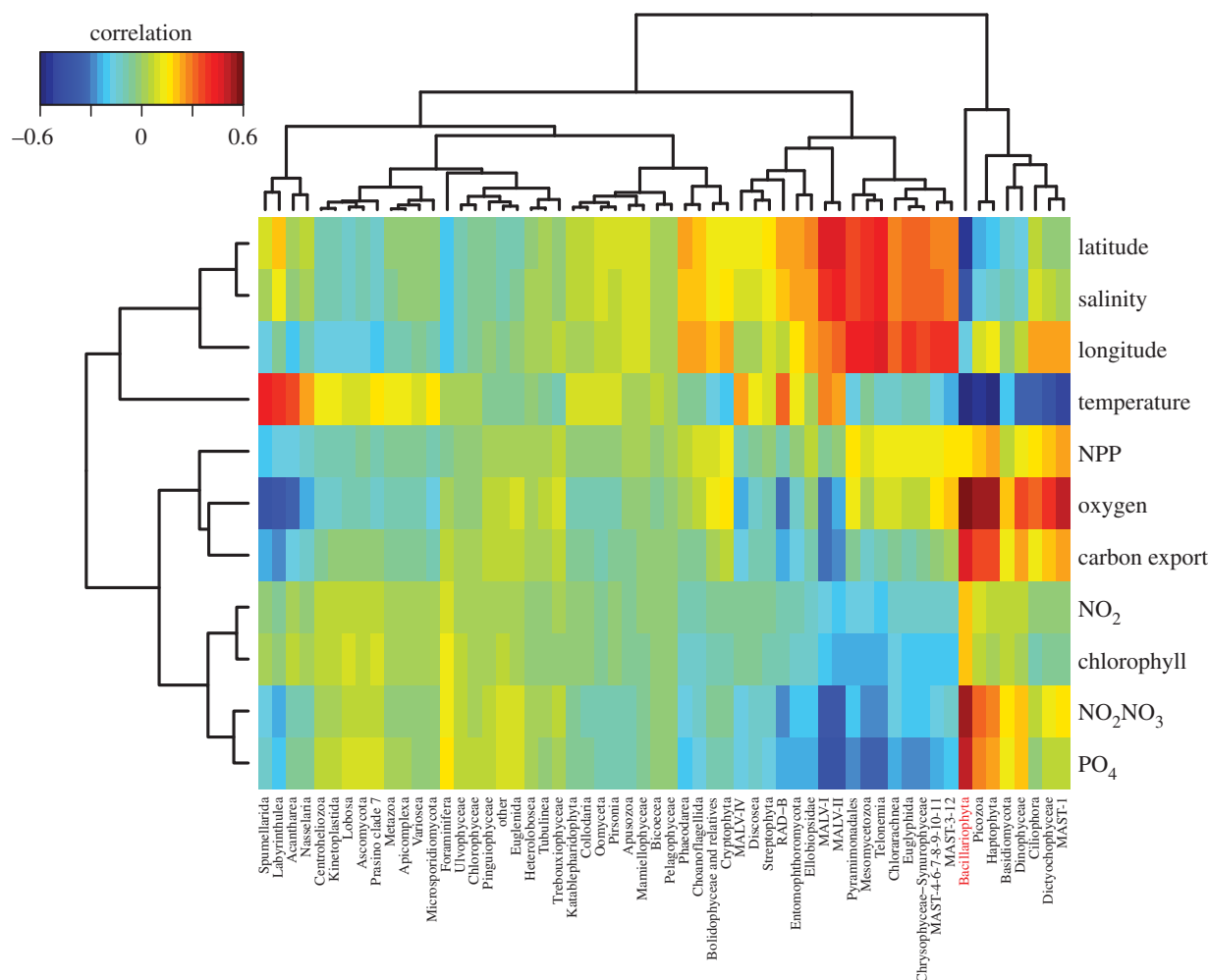


Figure 6. Diatoms and their role in the biological carbon pump as revealed by high-throughput DNA sequence-based datasets. Eukaryotic lineages associated to environmental parameters assessed by standard methods for regression-based modelling (sPLS analysis). Correlations between lineages and environmental parameters are depicted as a clustered heat map. This plot has been created using the *Tara* Oceans metabarcoding dataset (providing an abundance matrix based on 18s rDNA ribotypes (V9 region) from oligotrophic stations as well as a few Southern Ocean stations [48,81] and the associated environmental parameters [86]. With respect to other eukaryotic lineages, diatoms (*Bacillariophyta*) show significant correlations with NPP and chlorophyll, and the highest positive correlation to carbon flux (more than 0.46), supporting the hypothesis that diatoms play a major role in the biological carbon pump at a global scale.

consequence of CO₂ released into the atmosphere from the burning of fossil fuels. The overprint of human activities on Earth's biogeochemical cycles is evident from the simple fact that we are currently burning the equivalent of around 1 million years of buried carbon derived from diatoms and other plankton each year [2]. While we can be confident that the oceans will continue for some time to be the major sink absorbing excess heat and CO₂, and will consequently warm, acidify, and deoxygenate in the coming centuries [87,88], we have no consistent view about how the life support system of the oceans, the plankton, will fare. Regarding diatoms, we can expect shifts in several aspects of diatom diversity and biogeography, which could not only affect biogeochemical cycles but may also pose a challenge for the functioning of marine food webs, in which diatoms are intensely grazed. Given the rise of diatoms to global importance in marine ecosystems over the last tens of millions of years it is crucial that future research addresses their capabilities to adapt to changing environments, both by investigation of the geological record and by the exploration of diatom genomes.

Authors' contributions. C.B. conceived and designed the manuscript, and all authors contributed significantly to the writing and the creation of the figures. A.-S.B., F.M.I. and C.B. worked on the sections of evolutionary and biogeochemical events and diatom genomics. A.-S.B., L.B. and L.G. presented the biological carbon pump. F.M.I., O.J. and S.D. compared the marine sediments with modern distributions. L.B. and L.G. were in charge of the eco-systems biology and high-throughput sequencing sections.

Competing interests. We have no competing interests.

Funding. C.B. acknowledges funding from the ERC Advanced Award 'Diatomite', the Louis D Foundation, the Gordon and Betty Moore Foundation, and the French Government 'Investissements d'Avenir' programmes MEMO LIFE (ANR-10-LABX-54), PSL* Research University (ANR-1253 11-IDEX-0001-02), and OCEANOMICS (ANR-11-BTBR-0008). C.B. also thanks the Radcliffe Institute of Advanced Study at Harvard University for a scholars fellowship during the 2016-2017 academic year. A.-S.B. is funded by the PhD programme from the Ecole Doctorale Complexité du Vivant. L.B. and L.G. acknowledge funding from the French national programme EC2CO-LEFE (FunOmics project). S.D. and O.J. acknowledge funding from the Gordon and Betty Moore Foundation, and from NASA (NNX16AR47G).

Acknowledgements. The authors thank Andrew H. Knoll for his valuable contributions to this manuscript. This article is contribution number 55 of *Tara* Oceans.

1. Field CB, Behrenfeld MJ, Randerson JT, Falkowski PG. 1998 Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237–240. (doi:10.1126/science.281.5374.237)
2. Falkowski PG. 2015 *Life's engines: how microbes made Earth habitable*. Princeton, NJ: Princeton University Press.
3. Bowler C, Vardi A, Allen AE. 2010 Oceanographic and biogeochemical insights from diatom genomes. *Ann. Rev. Mar. Sci.* **2**, 333–365. (doi:10.1146/annurev-marine-120308-081051)
4. Armbrust EV. 2009 The life of diatoms in the world's oceans. *Nature* **459**, 185–192. (doi:10.1038/nature08057)
5. Nelson DM, Tréguer P, Brzezinski MA, Leynaert A, Quéguiner B. 1995 Production and dissolution of biogenic silica in the ocean: revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochem. Cycles* **9**, 359–372. (doi:10.1029/95GB01070)
6. Fischer WW, Hemp J, Johnson JE. 2016 Evolution of oxygenic photosynthesis. *Annu. Rev. Earth Planet. Sci.* **44**, 647–683. (doi:10.1146/annurev-earth-060313-054810)
7. Holland HD. 2006 The oxygenation of the atmosphere and oceans. *Phil. Trans. R. Soc. B* **361**, 903–915. (doi:10.1098/rstb.2006.1838)
8. Lyons TW, Reinhard CT, Planavsky NJ. 2014 The rise of oxygen in Earth's early ocean and atmosphere. *Nature* **506**, 307–315. (doi:10.1038/nature13068)
9. Hayes JM, Strauss H, Kaufman AJ. 1999 The abundance of ¹³C in marine organic matter and isotopic fractionation in the global biogeochemical cycle of carbon during the past 800 Ma. *Chem. Geol.* **161**, 103–125. (doi:10.1016/S0009-2541(99)00083-2)
10. Berner RA, Kothavala Z. 2001 GEOCARB III: a revised model of atmospheric CO₂ over Phanerozoic time. *Am. J. Sci.* **301**, 182–204. (doi:10.2475/ajs.301.2.182)
11. Lazarus D, Barron J, Renaudie J, Diver P, Türke A. 2014 Cenozoic planktonic marine diatom diversity and correlation to climate change. *PLoS ONE* **9**, e84857. (doi:10.1371/journal.pone.0084857)
12. Knoll AH, Summons RE, Waldbauer JR, Zumberge JE. 2007 The geological succession of primary producers in the oceans. In *Evolution of primary producers in the sea* (eds PG Falkowski, AH Knoll), pp. 133–163. Burlington, MA: Elsevier.
13. Hoffman PF. 2016 Cryoconite pans on Snowball Earth: supraglacial oases for Cryogenian eukaryotes? *Geobiology* **14**, 531–542. (doi:10.1111/gbi.12191)
14. Kopp RE, Kirschvink JL, Hilburn IA, Nash CZ. 2005 The Paleoproterozoic snowball Earth: a climate disaster triggered by the evolution of oxygenic photosynthesis. *Proc. Natl Acad. Sci. USA* **102**, 11 131–11 136. (doi:10.1073/pnas.0504878102)
15. Scotese CR. 2014 PALEOMAP Atlas for ArcGIS, PALEOMAP Project. See <http://www.scotese.com/>.
16. Nelson N, Ben-Shem A. 2004 The complex architecture of oxygenic photosynthesis. *Nat. Rev. Mol. Cell Biol.* **5**, 971–982. (doi:10.1038/nrm1525)
17. Goldsmith PF *et al.* 2011 HERSCHEL measurements of molecular oxygen in Orion. *Astrophys. J.* **737**, 96. (doi:10.1088/0004-637X/737/2/96)
18. Martin W, Russell MJ. 2003 On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Phil. Trans. R. Soc. Lond. B* **358**, 59–85. (doi:10.1098/rstb.2002.1183)
19. Knoll AH, Javaux EJ, Hewitt D, Cohen P. 2006 Eukaryotic organisms in Proterozoic oceans. *Phil. Trans. R. Soc. B* **361**, 1023–1038. (doi:10.1098/rstb.2006.1843)
20. Reyes-Prieto A, Weber APM, Bhattacharya D. 2007 The origin and establishment of the plastid in algae and plants. *Annu. Rev. Genet.* **41**, 147–168. (doi:10.1146/annurev.genet.41.110306.130134)
21. Guidi L, Stemmann L, Jackson GA, Ibanez F, Claustre H, Legendre L, Picheral M, Gorsky G. 2009 Effects of phytoplankton community on production, size, and export of large aggregates: a world-ocean analysis. *Limnol. Oceanogr.* **54**, 1951–1963. (doi:10.4319/lo.2009.54.6.1951)
22. Martin JH, Knauer GA, Karl DM, Broenkow WW. 1987 VERTEX: carbon cycling in the northeast Pacific. *Deep Sea Res. Part A Oceanogr. Res. Pap.* **34**, 267–285. (doi:10.1016/0198-0149(87)90086-0)
23. Dahl TW *et al.* 2010 Devonian rise in atmospheric oxygen correlated to the radiations of terrestrial plants and large predatory fish. *Proc. Natl Acad. Sci. USA* **107**, 17 911–17 915. (doi:10.1073/pnas.1011287107)
24. Plattner G-K, Joos F, Stocker TF. 2002 Revision of the global carbon budget due to changing air-sea oxygen fluxes. *Global Biogeochem. Cycles* **16**, 12–43. (doi:10.1029/2001GB001746)
25. Keeling RF, Shertz SR. 1992 Seasonal and interannual variations in atmospheric oxygen and implications for the global carbon cycle. *Nature* **358**, 723–727. (doi:10.1038/358723a0)
26. Keeling RF, Garcia HE. 2002 The change in oceanic O₂ inventory associated with recent global warming. *Proc. Natl Acad. Sci. USA* **99**, 7848–7853. (doi:10.1073/pnas.122154899)
27. Miller KG *et al.* 2005 The Phanerozoic record of global sea-level change. *Science* **310**, 1293–1298. (doi:10.1126/science.1116412)
28. Falkowski PG, Fenchel T, Delong EF. 2008 The microbial engines that drive Earth's biogeochemical cycles. *Science* **320**, 1034–1039. (doi:10.1126/science.1153213)
29. Falkowski PG, Katz ME, Milligan AJ, Fennel K, Cramer BS, Aubry MP, Berner RA, Novacek MJ, Zapol WM. 2005 The rise of oxygen over the past 205 million years and the evolution of large placental mammals. *Science* **309**, 2202–2204. (doi:10.1126/science.1116047)
30. Smetacek V. 1999 Diatoms and the ocean carbon cycle. *Protist* **150**, 25–32. (doi:10.1016/S1434-4610(99)70006-4)
31. Katz ME, Finkel ZV, Grzebyk D, Knoll AH, Falkowski PG. 2004 Evolutionary trajectories and biogeochemical impacts of marine eukaryotic phytoplankton. *Annu. Rev. Ecol. Evol. Syst.* **35**, 523–556. (doi:10.1146/annurev.ecolsys.35.112202.130137)
32. Knoll AH, Carroll SB. 1999 Early animal evolution: emerging views from comparative biology and geology. *Science* **284**, 2129–2137. (doi:10.1126/science.284.5423.2129)
33. Sperling EA, Frieder CA, Raman AV, Girguis PR, Levin LA, Knoll AH. 2013 Oxygen, ecology, and the Cambrian radiation of animals. *Proc. Natl Acad. Sci. USA* **110**, 13 446–13 451. (doi:10.1073/pnas.1312778110)
34. Bopp L, Bowler C, Guidi L, Karsenti E, de Vargas C. 2015 The ocean: a carbon pump. See <http://www.ocean-climate.org>.
35. Ciais P *et al.* 2013 Carbon and other biogeochemical cycles. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds TF Stocker *et al.*), pp. 465–570. Cambridge, UK: Cambridge University Press.
36. Siegenthaler U, Sarmiento JL. 1993 Atmospheric carbon dioxide and the ocean. *Nature* **365**, 119–125. (doi:10.1038/365119a0)
37. Volk T, Hoffert M. 1985 Ocean carbon pumps: analysis of relative strengths and efficiencies in ocean-driven atmospheric CO₂ changes. In *The carbon cycle and atmospheric CO₂: natural variations Archean to Present* (eds ET Sundquist, WS Broecker), pp. 99–110. American Geophysical Union.
38. Frankignoulle M, Canon C, Gattuso J-P. 1994 Marine calcification as a source of carbon dioxide: positive feedback of increasing atmospheric CO₂. *Limnol. Oceanogr.* **39**, 458–462. (doi:10.4319/lo.1994.39.2.0458)
39. Benton M. 2005 *When life nearly died: the greatest mass extinction of all time*. London, UK: Thames & Hudson.
40. Sims PA, Mann DG, Medlin LK. 2006 Evolution of the diatoms: insights from fossil, biological and molecular data. *Phycologia* **45**, 361–402. (doi:10.2216/05-22.1)
41. Kooistra WHCF, Gersonde R, Medlin LK, Mann DG. 2007 The origin and evolution of the diatoms: their adaptation to a planktonic existence. In *Evolution of primary producers in the sea* (eds PG Falkowski, AH Knoll), pp. 207–249. Burlington, MA: Elsevier.
42. Katz ME, Fennel K, Falkowski PG. 2007 Geochemical and biological consequences of phytoplankton evolution. In *Evolution of primary producers in the sea* (eds PG Falkowski, AH Knoll), pp. 405–430. Burlington, MA: Elsevier.
43. Cermeño P, Falkowski PG, Romero OE, Schaller MF, Vallina SM. 2015 Continental erosion and the Cenozoic rise of marine diatoms. *Proc. Natl Acad. Sci. USA* **112**, 4239–4244. (doi:10.1073/pnas.1412883112)

44. Cermeño P. 2016 The geological story of marine diatoms and the last generation of fossil fuels. *Perspect. Phycol.* **3**, 53–60. (doi:10.1127/pip/2016/0050)
45. Siegenthaler U *et al.* 2005 Stable carbon cycle—climate relationship during the late Pleistocene. *Science* **2**, 1313–1317. (doi:10.1126/science.1120130)
46. Rabosky DL, Sorhannus U. 2009 Diversity dynamics of marine planktonic diatoms across the Cenozoic. *Nature* **457**, 183–186. (doi:10.1038/nature07435)
47. Dutkiewicz A, Müller RD, O'Callaghan S, Jónasson H. 2015 Census of seafloor sediments in the world's ocean. *Geology* **43**, 795–798. (doi:10.1130/G36883.1)
48. Malviya S *et al.* 2016 Insights into global diatom distribution and diversity in the world's ocean. *Proc. Natl Acad. Sci. USA* **113**, E1516–E1525. (doi:10.1073/pnas.1509523113)
49. Dutkiewicz S, Hickman AE, Jahn O, Gregg WW, Mouw CB, Follows MJ. 2015 Capturing optically important constituents and properties in a marine biogeochemical and ecosystem model. *Biogeosciences* **12**, 4447–4481. (doi:10.5194/bg-12-4447-2015)
50. Winder M, Cloern JE. 2010 The annual cycles of phytoplankton biomass. *Phil. Trans. R. Soc. B* **365**, 3215–3226. (doi:10.1098/rstb.2010.0125)
51. Gautier DL *et al.* 2009 Assessment of undiscovered oil and gas in the Arctic. *Science* **324**, 1175–1179. (doi:10.1126/science.1169467)
52. Masters CD, Root DH, Dietzman WD. 1983 *Distribution and quantitative assessment of world crude oil reserves and resources*. Open-File Report 83-728, US Geological Survey, <http://pubs.er.usgs.gov/publication/ofr83728>.
53. Edwards D, Kenrick P. 2015 The early evolution of land plants, from fossils to genomics: a commentary on Lang (1937) 'On the plant-remains from the Downtonian of England and Wales'. *Phil. Trans. R. Soc. B* **370**, 20140343. (doi:10.1098/rstb.2014.0343)
54. Chacón-Baca E, Beraldi-Campesi H, Cevallos-Ferriz SRS, Knoll AH, Golubic S. 2002 70 Ma nonmarine diatoms from northern Mexico. *Geology* **30**, 279–281. (doi:10.1130/0091-7613)
55. Gibbs SP. 1981 The chloroplasts of some algal groups may have evolved from endosymbiotic eukaryotic algae. *Ann. N Y Acad. Sci.* **361**, 193–208. (doi:10.1111/j.1749-6632.1981.tb54365.x)
56. Parker MS, Mock T, Armbrust EV. 2008 Genomic insights into marine microalgae. *Annu. Rev. Genet.* **42**, 619–645. (doi:10.1146/annurev.genet.42.110807.091417)
57. Cavalier-Smith T. 1981 Eukaryote kingdoms: seven or nine? *Biosystems* **14**, 461–481. (doi:10.1016/0303-2647(81)90050-2)
58. Armbrust EV *et al.* 2004 The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**, 79–86. (doi:10.1126/science.1101156)
59. Bowler C *et al.* 2008 The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239–244. (doi:10.1038/nature07410)
60. Moustafa A, Beszteri BB, Maier UG, Bowler C, Valentin K, Bhattacharya D. 2009 Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* **324**, 1724–1726. (doi:10.1126/science.1172983)
61. Deschamps P, Moreira D. 2012 Reevaluating the green contribution to diatom genomes. *Genome Biol. Evol.* **4**, 683–688. (doi:10.1093/gbe/evs053)
62. Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield O, Taylor FJR. 2004 The evolution of modern eukaryotic phytoplankton. *Science* **305**, 354–360. (doi:10.1126/science.1095964)
63. Falkowski PG, Oliver MJ. 2007 Mix and match: how climate selects phytoplankton. *Nat. Rev. Microbiol.* **5**, 813–819. (doi:10.1038/nrmicro1792)
64. Maumus F, Allen AE, Mhiri C, Hu H, Jabbari K, Vardi A, Grandbastien M-A, Bowler C. 2009 Potential impact of stress activated retrotransposons on genome evolution in a marine diatom. *BMC Genomics* **10**, 624. (doi:10.1186/1471-2164-10-624)
65. Allen AE *et al.* 2011 Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* **473**, 203–207. (doi:10.1038/nature10074)
66. Bailleul B *et al.* 2015 Energetic coupling between plastids and mitochondria drives CO₂ assimilation in diatoms. *Nature* **524**, 366–369. (doi:10.1038/nature14599)
67. Huysman MJJ *et al.* 2010 Genome-wide analysis of the diatom cell cycle unveils a novel type of cyclins involved in environmental signaling. *Genome Biol.* **11**, R17. (doi:10.1186/gb-2010-11-2-r17)
68. Bailleul B, Rogato A, de Martino A, Coesel S, Cardol P, Bowler C, Falcatore A, 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**, 18 214–18 219. (doi:10.1073/pnas.1007703107)
69. Mock T *et al.* 2017 Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature* **541**, 536–540. (doi:10.1038/nature20803)
70. Kotzsch A, Pawolski D, Milentyev A, Shevchenko A, Scheffel A, Poulsen N, Shevchenko A, Kröger N. 2016 Biochemical composition and assembly of biosilica-associated insoluble organic matrices from the diatom *Thalassiosira pseudonana*. *J. Biol. Chem.* **291**, 4982–4997. (doi:10.1074/jbc.M115.706440)
71. Kröger N, Poulsen N. 2008 Diatoms—from cell wall biogenesis to nanotechnology. *Annu. Rev. Genet.* **42**, 83–107. (doi:10.1146/annurev.genet.41.110306.130109)
72. Prihoda J, Tanaka A, de Paula WBM, Allen JF, Tirichine L, Bowler C. 2012 Chloroplast-mitochondria cross-talk in diatoms. *J. Exp. Bot.* **63**, 1543–1557. (doi:10.1093/jxb/err441)
73. Caron DA *et al.* 2016 Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nat. Rev. Microbiol.* **15**, 6–20. (doi:10.1038/nrmicro.2016.160)
74. Peers G, Price NM. 2006 Copper-containing plastocyanin used for electron transport by an oceanic diatom. *Nature* **441**, 341–344. (doi:10.1038/nature04630)
75. Marchetti A, Schrueth DM, Durkin CA, Parker MS, Kodner RB, Berthiaume CT, Morales R, Allen AE, Armbrust EV. 2012 Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proc. Natl Acad. Sci. USA* **109**, E317–E325. (doi:10.1073/pnas.1118408109)
76. Veluchamy A *et al.* 2013 Insights into the role of DNA methylation in diatoms by genome-wide profiling in *Phaeodactylum tricornutum*. *Nat. Commun.* **4**, 2091. (doi:10.1038/ncomms3091)
77. Veluchamy A *et al.* 2015 An integrative analysis of post-translational histone modifications in the marine diatom *Phaeodactylum tricornutum*. *Genome Biol.* **16**, 102. (doi:10.1186/s13059-015-0671-8)
78. Tirichine L, Bowler C. 2011 Decoding algal genomes: tracing back the history of photosynthetic life on Earth. *Plant J.* **66**, 45–57. (doi:10.1111/j.1365-3113X.2011.04540.x)
79. Leblanc K *et al.* 2012 A global diatom database—abundance, biovolume and biomass in the world ocean. *Earth Syst. Sci. Data Discuss.* **5**, 147–185. (doi:10.5194/essdd-5-147-2012)
80. Brun P, Vogt M, Payne MR, Gruber N, O'Brien CJ, Buitenhuis ET, Le Quéré C, Leblanc K, Luo Y-W. 2015 Ecological niches of open ocean phytoplankton taxa. *Limnol. Oceanogr.* **60**, 1020–1038. (doi:10.1002/lno.10074)
81. de Vargas C *et al.* 2015 Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**, 1261605. (doi:10.1126/science.1261605)
82. Uitz J, Claustre H, Gentili B, Stramski D. 2010 Phytoplankton class-specific primary production in the world's oceans: seasonal and interannual variability from satellite observations. *Global Biogeochem. Cycles* **24**, GB3016. (doi:10.1029/2009GB003680)
83. Guidi L *et al.* 2016 Plankton networks driving carbon export in the oligotrophic ocean. *Nature* **532**, 465–470. (doi:10.1038/nature16942)
84. Rembauville M, Manno C, Tarling GA, Blain S, Salter I. 2016 Strong contribution of diatom resting spores to deep-sea carbon transfer in naturally iron-fertilized waters downstream of South Georgia. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **115**, 22–35. (doi:10.1016/j.dsr.2016.05.002)
85. Tréguer PJ, De La Rocha CL. 2013 The world ocean silica cycle. *Ann. Rev. Mar. Sci.* **5**, 477–501. (doi:10.1146/annurev-marine-121211-172346)
86. Pesant S *et al.* 2015 Open science resources for the discovery and analysis of Tara Oceans data. *Sci. Data* **2**, 150023. (doi:10.1038/sdata.2015.23)
87. Pörtner H-O, Karl DM, Boyd PW, Cheung WWL, Lluch-Cota SE, Nojiri Y, Schmidt DN, Zavialov PO. 2014 Ocean systems. In *Climate change 2014: impacts, adaptation, and vulnerability. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds CB Field *et al.*), pp. 411–484. Cambridge, UK: Cambridge University Press.
88. Allison EH, Bassett HR. 2015 Climate change in the oceans: human impacts and responses. *Science* **350**, 778–782. (doi:10.1126/science.aac8721)