Emergence of *Prochlorococcus* in the Tonian oceans and the initiation of Neoproterozoic oxygenation

Hao Zhang[1,2^], Sishuo Wang[2^], Tianhua Liao[2], Sean A. Crowe[3], Haiwei Luo[2,4,5*]

[1]Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen 518000, China

[2]Simon F. S. Li Marine Science Laboratory, School of Life Sciences and State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Shatin, Hong Kong SAR

[3]Departments of Microbiology and Immunology, and Earth, Ocean, and Atmospheric Sciences, University of British Columbia, Vancouver, British Columbia, Canada

[4]Earth and Environmental Sciences Programme, The Chinese University of Hong Kong, Shatin, Hong Kong SAR

[5]Institute of Environment, Energy and Sustainability, The Chinese University of Hong Kong, Shatin, Hong Kong SAR

[^] These authors contributed equally to this study

[*]**Corresponding author:** Haiwei Luo

Email: haiweiluo@cuhk.edu.hk

Competing Interest Statement: The authors declare no competing interests.

Classification: Biological Sciences - Microbiology; Physical Sciences - Earth, Atmospheric, and Planetary Sciences

Key words: *Prochlorococcus*, Neoproterozoic oxygenation event, molecular dating, plastid, cyanobacteria

Abstract

Prochlorococcus are the smallest and most abundant photosynthetic organisms on Earth, contributing up to 50% of the chlorophyll in the oligotrophic oceans. Despite being important in regulating the carbon cycle in today's ocean, the ecological significance of *Prochlorococcus* in Earth's history remains elusive. Our new robustly calibrated molecular clock reveals that *Prochlorococcus* emerged in the deep photic zone of the Tonian oceans. The classical antenna for light harvesting in Cyanobacteria, the phycobilisome, was replaced in *Prochlorococcus* by the chlorophyll based light harvesting complex enabling more efficient use of blue light that penetrated to deeper water. Importantly, *Prochlorococcus* colonization of deep water enhanced access to phosphate, which was rich in upwelled seawater, but likely scarce in the Tonian surface ocean, promoting expansion of *Prochlorococcus* and associated photosynthetic oxygen production. Colonization of deeper waters would also have improved access to ammonium, leading to the neutral loss of nitrate utilization genes. Our research thus documents the conspicuous emergence of new photosynthetic bacterial lineages in the run-up to the Neoproterozoic oxygenation event, implying an additional layer of eco-evolutionary complexity during this pivotal interval in Earth's history.

Significance (max 120 words allowed)

Despite being important in today's oceans, the ecological significance of *Prochlorococcus* in the past remains elusive. By 'borrowing' fossils from photosynthetic eukaryotes and leveraging the sequential Bayesian dating approach, we have developed a new pipeline for dating the evolution of *Prochlorococcus*, which systematically quantifies the uncertainties in using different fossil calibrations, molecular datasets, species tree topologies and molecular clock models. Our analyses pin-pointed the origin of *Prochlorococcus* to the phosphorous-limited Tonian oceans. The changes in the photosynthetic apparatus in early *Prochlorococcus* enabled them to colonize the phosphorus-rich deep photic niches, which would have led to increased primary productivity and oxygen release, likely contributing to the Neoproterozoic Oxygenation Event.

Introduction

Prochlorococcus host more than half of the chlorophyll biomass in oligotrophic oceans (Partensky and Garczarek, 2010). As key primary producers in the modern ocean, *Prochlorococcus* fix as much as four gigatons of carbon each year and are the foundation of the marine carbon cycle and food web (Flombaum et al., 2013). The success of *Prochlorococcus* in today's oceans has been attributed to multiple physiological features. Importantly, *Prochlorococcus* use divinyl chlorophyll (DVChl) for harnessing light energy (Ralf and Repeta, 1992). DVChl harvests blue light much more efficiently than the more common monovinyl chlorophyll used by other cyanobacterial lineages and it thus enables *Prochlorococcus* to thrive in the deepest layers of the euphotic zone, where blue light

dominates (Ito and Tanaka, 2011). *Prochlorococcus* is also the smallest photosynthetic group on Earth (Partensky et al., 1999). As a result, their high surface-to-volume ratio provides enhanced nutrient acquisition efficiency, which together with effective blue-light absorption promotes photosynthesis by *Prochlorococcus* in oligotrophic tropical and subtropical oceans (Partensky et al., 1999). In today's ocean, *Prochlorococcus* is also the dominant phototroph in more nutrient-rich, oxygen-depleted anoxic marine zones (AMZ) in the eastern tropical North and South Pacific Oceans (Goericke et al., 2000; Lavin et al., 2010). Since the AMZ lineages represent the earliest-split branches of *Prochlorococcus*, it has been proposed that *Prochlorococcus* emerged in low-oxygen environments and, by extension, contributed to early ocean oxygenation (Ulloa et al., 2021).

Despite the ecological prominence of *Prochlorococcus* in both modern and likely the past oceans, the divergence time of *Prochlorococcus* remains contentious. Different relaxed clock analyses based on 16S and 23S rRNA genes yielded divergence times of the total *Prochlorococcus* group at ~200 Mya (Boden et al., 2021), ~500 Mya (Sánchez-Baracaldo, 2015) and ~800 Mya (Sánchez-Baracaldo et al., 2014). The use of protein-coding genes gave divergence times of the total *Prochlorococcus* group at ~400 Mya (Fournier et al., 2021), ~500 Mya (Sánchez-Baracaldo et al., 2017) and ~1,000 Mya (Schirrmeister et al., 2015). The discrepancies in *Prochlorococcus* divergence times are likely caused by variable use of fossils and other time calibrations and the use of alternative gene sets, clock models, and tree topologies (Warnock et al., 2012; Duchêne et al., 2014; dos Reis et al., 2015) (see Supplementary Discussion). Despite their importance, these factors were rarely tested rigorously thus making it difficult to evaluate the accuracy of age estimates in previous

studies. Importantly, even if all these factors are well tested and controlled, the rarity of cyanobacterial fossils and the absence of a maximum boundary of the available cyanobacterial fossils poses an appreciable challenge in determining the antiquity of *Prochlorococcus* (Zhang et al., 2021). Indeed, applying maximum age constraints at the calibration nodes strongly impacts posterior age estimates of uncalibrated lineages (Hedges et al., 2018; Morris et al., 2018; Wang and Luo, 2021). However, informative maximum age constraints are typically missing from microbial fossils and can only be found in some animal and plant fossils. Moreover, the recent availability of several genome sequences of uncultivated basal *Prochlorococcus* lineages (Ulloa et al., 2021) requires re-estimation of the antiquity of *Prochlorococcus*, which should also include an evaluation of the factors that influence the accuracy of posterior ages.

Here, we leverage the abundant plant and algal fossils and the well-established plant plastid endosymbiosis theory to develop a new pipeline that systematically tests the factors that influence the accuracy of posterior ages and thus yields robust estimates of *Prochlorococcus* antiquity. The plant plastid endosymbiosis theory states that the origin of all plastids in photosynthetic eukaryotes, except for the photosynthetic amoeba *Paulinella* (Marin et al., 2005), can be traced back to an ancient primary endosymbiosis involving a eukaryote and a cyanobacterium (Gray, 1992; Bhattacharya and Medlin, 1995; Keeling, 2013; Ponce-Toledo et al., 2017). Plant plastid endosymbiosis theory thus ties the evolutionary histories of cyanobacteria to those of photosynthetic eukaryotes. Eukaryotic fossils are indeed being increasingly used in dating the evolution of Cyanobacteria (Shih et al., 2016; Sánchez-Baracaldo et al., 2017; Fournier et al., 2021). Our strategy builds on these by: 1) using more

fossil-based age constraints (including maximum age constraints) on eukaryotic lineages with more complete taxonomic sampling including lineages that have undergone secondary endosymbiosis; and 2) applying a sequential Bayesian dating approach to better constrain the divergence time of eukaryotic lineages thereby propagating more accurate time information to Cyanobacteria, including *Prochlorococcus*. These joint analyses led us to conclude that *Prochlorococcus* arose in the Tonian ocean. Further, the use of an evolutionary genetic proxy implies that *Prochlorococcus* was born through a founder effect, and this strengthens the idea that *Prochlorococcus* emerged in the deep photic zone, a unique niche separated from upper waters where its *Synechococcus* ancestors thrived.

Results

Prochlorococcus originated in the Tonian ocean

We implemented a sequential Bayesian dating method that takes advantage of the abundant fossil records (Fig. 1a; Fig. S1a) available from photosynthetic eukaryotes to calibrate the evolution of Cyanobacteria. The last common ancestor (LCA) of *Prochlorococcus* (denoted as “Proch-AMZI/II/III-LCA”) clearly emerged within the Tonian period at 878 Mya (95% HPD: 987-767 Mya) (Fig. 1b). Meanwhile, the LCA of *Prochlorococcus* clades HL, LL and AMZI/II (denoted as “Proch-AMZI/II-LCA”) evolved at 787 Mya (95% HPD: 896-687 Mya), which coincided with the early stage of the Neoproterozoic Oxygenation Event (NOE) (Fig. 1b). In this case, the branch leading to the LCA of *Prochlorococcus* HL, LLI, and LLII/III clades (denoted as “SBE-LCA” to keep consistency with (Zhang et al., 2021)) spanned the time that encompassed the duration of the

Neoproterozoic Snowball Earth events (645-635 Mya for Marinoan glaciation and 717-659 Mya for Sturtian glaciation), supporting the main conclusion of the previous study, which used the traditional cyanobacterial fossil-based strategy (Zhang et al., 2021).

Note that the sequential Bayesian method used here is different from the commonly used secondary dating method (using time estimates from previous studies as calibrations) (Heckman et al., 2001; Aoki et al., 2013; Chriki-Adeeb and Chriki, 2016) for two reasons. First, the time prior used in sequential dating analysis follows a specific probability distribution, but secondary calibrations ignore the uncertainties on age estimates. Second, the sequential dating analysis contains two steps each based on a distinct set of molecular data, while secondary dating analysis largely relies on the same molecular dataset (gene alignments) (dos Reis et al., 2018).

In our implementation, the posterior ages of eukaryotic lineages derived from the first-step of the sequential analysis were used as the time priors to calibrate Cyanobacteria evolution where few time constraints are available. Since genome-scale dating of eukaryotic lineages is very slow with fully partitioned molecular data and is prone to have reduced coverage probability (the probability that the 95% HPD interval of posterior ages contains the true age) due to the large number of partitions (Angelis et al., 2018), we first clustered the eukaryotic gene families into 1-, 3-, 5-, 7-, and 9-partitions based on their evolutionary rates (see details in Supplementary Methods). In general, the posterior ages of eukaryotic lineages were consistent, except for the case where a single partition and a single substitution model was used, which did not fit the data well as indicated by its larger Akaike Information Criterion (AIC) value (Fig. S1b). In all other cases, the crown group of Archaeplastida was

estimated to evolve at ~1.60 Gya, which is consistent with the estimates from other studies (Parfrey et al., 2011; Betts et al., 2018; Wang and Luo, 2021) (Fig. S1b). By comparing the distributions of the posterior ages of eukaryotic nodes in the first-step of sequential analysis to the effective prior on the corresponding nodes in the second-step of analysis (see details in Supplementary Methods), we found that the two distributions in dating analyses were nearly identical when 5-, 7- and 9-partition data were used (Fig. S2), suggesting that the sequential Bayesian approach works well for these analyses. We therefore adopted the evolutionary timeline of eukaryotes based on the 5-partition data (lowest AIC value and high similarity between posterior ages and effective prior) and applied the posterior ages of eukaryotic lineages as time prior at the overlapping nodes in a larger phylogenomic tree that includes both photosynthetic eukaryotes and cyanobacteria (Fig. S1c).

To validate the evolutionary timeline of *Prochlorococcus*, we performed a series of tests by using different molecular data, clock models, fossil calibrations and species tree topologies. The posterior ages estimated with all these alternative settings are largely consistent with that estimated with the focal strategy (detailed above) and fully support the origin of *Prochlorococcus* (Proch-AMZI/II/III-LCA) in the Tonian period, followed by the emergence of Proch-AMZI/II-LCA in the early stages of the NOE (Fig. S3). Moreover, the branch leading to the SBE-LCA encompasses the Snowball Earth events in all the dating analyses (Fig. S4; see Supplementary Results). Our study also results in an estimate of Cyanobacteria age. While this is not the focus here, the antiquity of Cyanobacteria has been an important research topic and under heavy debate (Schirrmeister et al., 2015; Schirrmeister et al., 2016; Fournier et al., 2021; Sánchez-Baracaldo et al., 2022). We therefore included a

discussion (see Supplementary Discussion).

Prochlorococcus genomic changes

We inferred the genes gained and lost upon the emergence of *Prochlorococcus* first without using the SAGs to avoid the false prediction of genomic events due to the relatively low completeness of the SAGs (e.g., 81.9% for AMZI-B-ETNP). Our reconstructions rely on the gene tree-species tree reconciliation approach, in which a best-fit reconciliation tool was selected through a comprehensive simulation-based benchmarking analysis [see Fig. S4 in (Zhang et al., 2023)]. We then used the same reconstruction methods to validate these evolutionary events when SAGs were included.

As the only phytoplankton group using divinyl chlorophyll (DVChl) for harvesting light energy (Ralf and Repeta, 1992), *Prochlorococcus* lost the gene *bciB* that inhibits the conversion of DVChl to MVChl (monovinyl chlorophyll) at Proch-AMZI/II/III-LCA, thereby promoting the accumulation of DVChl in their membranes (Ito et al., 2008). Compared to the MVChl used by *Synechococcus*, DVChl more efficiently absorbs blue light that penetrates to deeper waters than other photosynthetically active radiation (Ito and Tanaka, 2011). Likewise, the gain of the *PcCao* gene for the synthesis of chlorophyll b (Satoh and Tanaka, 2006) in Proch-AMZI/II/III-LCA also enables *Prochlorococcus* ancestors to efficiently harvest blue light at exceedingly low intensities characteristic of deep waters (Hess et al., 2001) (Fig. 1b). These genomic changes would have enabled *Prochlorococcus* ancestors to explore the deep photic zone, a niche where its *Synechococcus* ancestor would not thrive.

Unlike most cyanobacteria, which use the phycobilisome as the photosynthetic antenna,

the main light-harvesting antenna of *Prochlorococcus* is made up of prochlorophyte chlorophyll-binding (Pcb) protein (Biller et al., 2015). By reconstructing the gene evolutionary paths with *Prochlorococcus* SAGs included, we found that the phycobilisome genes (*apcACDEF*, *cpcEFG*, and *cpeCES*) were present in both Proch-AMZI/II/III-LCA and Proch-AMZI/II-LCA and lost at SBE-LCA. Using the same approach, we found that the cholorophyll-binding protein (encoded by *pcbD*) was obtained at Proch-AMZI/II-LCA. The replacement of the photosynthetic antenna thus did not co-occur with the emergence of *Prochlorococcus*. We note that gene absence in SAGs could be due to the incomplete nature of the SAG genomes, however, in this case, simultaneous absence of the *pcbD* gene in all the five SAGs that are more than 80% complete seems unlikely.

Despite the absence of the nitrate utilization genes in most *Prochlorococcus* isolates, some uncultivated *Prochlorococcus* lineages contain these genes (Martiny et al., 2009; Berube et al., 2015; Berube et al., 2019). By including *Prochlorococcus* SAGs that contain the assimilatory nitrate transporter gene (*nrtB*) and the assimilatory nitrate reductase gene (*narB*) in our reconstructions, we inferred the presence of both genes in Proch-AMZI/II/III-LCA and Proch-AMZI/II-LCA and the losses of these genes at SBE-LCA (Fig. 1b). In addition to the nitrate utilization genes, we inferred the losses of molybdopterin biosynthesis genes at SBE-LCA. Since these genes are known to co-locate with the nitrate reductase genes in *Synechococcus* (Rubio et al., 1998; Palenik et al., 2003), they may function as the cofactor of nitrate reductase in *Prochlorococcus*. Therefore, the loss of nitrate reductase in *Prochlorococcus* likely rendered the molybdopterin dispensable and thus led to the losses of molybdopterin biosynthesis genes (*moaABCDE*, *mobA* and *moeA*) at SBE-LCA.

Selection efficiency was reduced upon the emergence of *Prochlorococcus*

The method for inferring selection efficiency in deep time was developed (Zhang, 2000) and recently improved (Luo et al., 2017), which involves nonsynonymous substitutions only and compares the rate of nonsynonymous substitutions leading to replacements of physicochemically dissimilar amino acids (i.e., radical changes; d_R) to the rate of nonsynonymous substitutions leading to replacements of physicochemically similar amino acids (i.e., conservative changes; d_C). Since radical changes are more likely to be deleterious than conservative changes (E. Zuckerkandl, 1965; Dayhoff, 1972), an excess of the radical changes in a deeply branching lineage compared to its sister lineage suggests random fixation of deleterious mutations by genetic drift in the former. Using this method, we found a significant increase of the d_R/d_C ratio across the genomic regions in the branch leading to Proch-AMZI/II/III-LCA relative to the branch leading to the LCA of *Synechococcus* clade 5.1 (Fig. 2a), indicating that the emergence of *Prochlorococcus* was accompanied by a significant reduction of the efficiency of purifying selection, allowing for an accelerated accumulation of deleterious mutations through genetic drift.

Discussion

Geochemical context that supports the emergence of *Prochlorococcus* in the Tonian period

At the 95% HPD, *Prochlorococcus* emerged between 989 to 770 Mya, encompassing almost the entire Tonian period. A Tonian, pre-Sturtian, age for the emergence of the *Prochlorococcus* lineage would have been set against the backdrop of dynamic ocean

chemistry characterized by a variably oxygenated surface ocean that was co-populated by diverse microbial eukaryotes, including phytoplankton and the earliest metazoans (Trower et al., 2023) (Fig. 1c). Multi-proxy data, collected from geographically diverse sites, converge on Tonian deep oceans that were pervasively anoxic across 60-80% of the ocean floor, or more (Tahata et al., 2015; Lau et al., 2017). The deep anoxic oceans were predominantly ferruginous in nature, albeit with evidence for transient, spatially restricted euxinia, and possibly even brief (<0.5 Myr) episodes of pervasive deep ocean oxygenation (Stolper and Keller, 2018; Tostevin and Mills, 2020). Shallow waters, by contrast, were modestly, though perhaps increasingly, oxygenated across the Tonian with evidence for the episodic and persistent intrusion of anoxic deep waters into the surface oceans (Zhang et al., 2022; Clarkson et al., 2023), with corresponding implications for nutrient cycling and availability and for marine life in the euphotic zone.

The strongest differentiators between the earliest *Prochlorococcus* (i.e., Proch-AMZI/II/III-LCA) and its *Synechococcus* ancestors are adaptation to low-light and loss of capacity for nitrate assimilation, which imply both a deeper photosynthetic niche and a likely shift to use ammonium as the primary nitrogen source. A key driver of *Prochlorococcus* emergence was thus likely a competitive advantage to dwelling in deeper waters. A well-known advantage emerges under conditions of strong P-limitation, whereby deep-dwelling low-light adapted phototrophs gain first access to upwelling phosphorus (Jones et al., 2015; Ozaki et al., 2019). The low-light advantage conferred on the earliest *Prochlorococcus* cells through the adoption of divinyl chlorophyll would have enhanced their access to P, relative to contemporary algae and their *Synechococcus* ancestors. It would also have increased access

to NH_4^+ upwelling from deeper anoxic waters obviating the need for nitrate assimilation (Michiels et al., 2017) (Fig. 1c).

In this way, the emergence of the *Prochlorococcus* would have restricted the flux of nutrients to shallower waters and limited the contributions of higher-light adapted phytoplankton like algae to primary production. This is supported through the biomarker record, which indicates a dominance of cyanobacterial primary production until after the Sturtian glaciation (Brocks, 2018). It is also supported by the Tonian N-isotope record, which implies a limited contribution from nitrate assimilation to primary production at this time (Kang et al., 2023). Importantly, the adaptation of the *Prochlorococcus* LCA to lower-light would also allow it to better compete with low-light adapted anoxygenic phototrophs that would likely have populated anoxic regions of the Tonian euphotic zone with strong potential to cause ocean oxygenation (Johnston et al., 2009a; Jones et al., 2015; Ozaki et al., 2019). Competition between oxygenic photosynthetic cyanobacteria and anoxygenic phototrophs that oxidize ferrous iron (photoferrotophs), is known to limit photosynthetic oxygen production (Jones et al., 2015; Ozaki et al., 2019). In this way, the capacity of the *Prochlorococcus* LCA to grow in deeper waters could have enhanced its ability to access upwelling phosphorus thereby enhancing the oxygenic fraction of total photosynthesis with potential to shift oxygen production towards runaway oxygenation (Ozaki et al., 2019).

<u>Evolutionary mechanisms underlying the origin of *Prochlorococcus*</u>

We propose that the colonization of the deep photic zone by *Prochlorococcus* started by a seed population with only a few individuals that obtained the ability to use divinyl

chlorophylls. Therefore, the reduced effective population size (N_e) upon the emergence of *Prochlorococcus* was likely caused by a "founder effect", which suggests that a new population was established by a few individuals from a larger ancestral population (Barton et al., 1984). Note that this mechanism is different from a population bottleneck, which has been used to explain the reduced N_e occurring on the branch leading to SBE-LCA (Zhang et al., 2021). The latter conclusion is confirmed here with the inclusion of new SAG genomes, as the d_R/d_C values in the branch leading to SBE-LCA were significantly elevated compared to the those in the branches leading to *Prochlorococcus* LLIV and AMZ clades (Fig. 2b)

Evolutionary drivers leading to the loss of phycobilisome and nitrate reductase in early *Prochlorococcus*

The phycobilisome constitutes as much as 50%-60% of the soluble proteins in *Synechococcus* (Grossman et al., 1993). The loss of the phycobilisome was thought to reduce nitrogen (N) investments by at least 40% in *Prochlorococcus* (Ting et al., 2002). Therefore, the losses of phycobilisome genes was once considered to be favored by *Prochlorococcus*, which conferred them advantages in inhabiting oligotrophic oceans (Ting et al., 2002). However, since the colonization of *Prochlorococcus* at the deep photic zone of the Tonian ocean conferred them advantages in acquiring the limiting nutrients like ammonium and phosphate upwelled from the deep ocean, the loss of the phycobilisome in *Prochlorococcus* was unlikely driven by selection for metabolic efficiency. Instead, since chlorophyll based light harvesting complex gradually became the primary photosynthetic antenna in *Prochlorococcus*, the phycobilisome genes were more likely subject to relaxed purifying

selection and thus neutral losses.

The absence of nitrate utilization genes (*narB* and *nrtB*) in cultured *Prochlorococcus* strains has been attributed to biased taxon sampling, since the *Prochlorococcus* were primarily cultured from ocean regions where P instead of N is the primarily limiting nutrient (Berube et al., 2015). However, phylogenetic analysis that included both *Prochlorococcus* and *Synechococcus* showed topological congruence of the *narB* gene tree with the species tree, suggesting vertical inheritance of the gene (Berube et al., 2019). By including *Prochlorococcus* SAGs in the present study, we confirmed the presence of the nitrate utilization genes in early ancestral *Prochlorococcus* (Proch-AMZI/II/III-LCA). The mechanism underlying the loss of nitrate utilization genes in *Prochlorococcus* was once attributed to relaxed selection efficiency due to the low level of nitrate caused by the intensive denitrification and anammox (anaerobic ammonium oxidation) activity (Canfield et al., 2008; Johnston et al., 2009b). However, since *Prochlorococcus* originated in the deep photic zone where ammonium would have been supplied through upwelling (Fig. 1c) (Michiels et al., 2017), the loss of the nitrate utilization genes in *Prochlorococcus* is more likely the result of superfluous nitrogen sources and a switch to uptake of ammonium instead of nitrate, which was again a neutral process.

<u>Implications for Neoproterozoic biogeochemical cycling</u>

In accordance with the previous study (Ulloa et al., 2021), our analysis suggests that *Prochlorococcus* diverged before transitioning from the use of the phycobilisome to the chlorophyll-based light-harvesting complex. Therefore, while the divinyl pigment synthesis

gene was obtained at the earliest *Prochlorococcus* (i.e., Proch-AMZI/II/III-LCA) to enable their exploration of the deep photic zone, our analyses imply that *Prochlorococcus* would not have strongly influenced the Neoproterozoic Earth system until the emergence of Proch-AMZI/II-LCA lineage during which the photosynthetic antenna was replaced. Our dating analysis shows that the emergence of this lineage indeed coincided with the early stages of Neoproterozoic Oxygenation Event (NOE).

The NOE is widely accepted as the second stage of Earth's protracted oxygenation, during which atmospheric oxygen rose to near modern levels (Och and Shields-Zhou, 2012; Lyons et al., 2014). This rise in atmospheric oxygen ultimately promoted the emergence of large and complex animal life (Knoll and Nowak, 2017). The initiation of the NOE has been previously attributed to the rise of eukaryotic algae (Lyons et al., 2014; Braakman et al., 2017; Brocks et al., 2017). Our results, instead, suggest that early *Prochlorococcus* may have played a role in enhancing oxygen production, with potential to initiate the NOE. This hypothesis could be further tested through geochemical analyses that help refine estimates for Tonian ocean oxygen contents and through biogeochemical models that can quantify the potential effects of *Prochlorococcus* emergence on the global carbon and oxygen cycles.

Concluding remarks

We have integrated molecular clocks, evolutionary genetic analyses, and comparative genomics along with knowledge of paleo-geochemistry to reconstruct the early evolution of *Prochlorococcus* and infer their possible impact on the evolution of the Neoproterozoic Earth system. The abundant fossil calibrations 'borrowed' from photosynthetic eukaryotes and the

new molecular dating pipeline provide robust dates for the origin of *Prochlorococcus* in the Tonian oceans. Comparative genomics suggests that gains and losses of a few important genes conferred the early ancestral *Prochlorococcus* with an unprecedented ability to absorb the blue light that effectively penetrates water, empowering early *Prochlorococcus* to colonize the deep photic zone, a niche distinctly different from the better illuminated surface waters supporting other phytoplankton groups including early Eukaryotes and its *Synechococcus* ancestor. The use of the evolutionary genetic proxy, d_R/d_C, implies that the earliest *Prochlorococcus* had a highly reduced effective population size compared to its *Synechococcus* ancestor, likely reflecting a founder effect and strengthening the idea that the deep photic zone was an ecological niche uniquely accessible to the earliest *Prochlorococcus*. This niche differentiation allowed the earliest *Prochlorococcus* to avoid fierce competition with its *Synechococcus* ancestor and other phytoplankton groups and gave them easy access to phosphate that otherwise strongly limited primary productivity in the Tonian oceans. Under these favorable conditions, the early *Prochlorococcus* would have grown in population size, promoting further evolution and the loss of the phycobilisome, which was replaced by a chlorophyll-based light-harvesting system ultimately leading to enhanced oxygen production and the possible initiation of the Neoproterozoic Oxygenation Event.

Materials and Methods

The nuclear-encoded and plastid-encoded protein sequences of eukaryotic representatives were obtained from Dicots PLAZA (5.0) database and NCBI Refseq release of plastid database and were used for the first-step and the second-step sequential Bayesian dating

analysis, respectively. The genomic sequences of Cyanobacteria (including *Prochlorococcus* SAGs), Vampirovibrionia and Sericytochromatia were also used for the second-step sequential dating analysis, which were obtained from GenBank and Integrated Microbial Genomes (IMG) database (Supplementary Data 1).

For the first-step sequential dating analysis, the orthologous gene families shared by eukaryotic species were identified by searching against a pre-compiled eukaryotic nuclear gene dataset (Strassert et al., 2021) and were partitioned by using Gaussian Mixture Model (GMM) clustering method based on their evolutionary rates. The posterior ages estimated with these partitioned data were approximated to one of the skew-normal, skew-t and Gamma distributions and then mapped back to the overlapping nodes in the species tree containing both eukaryotes and Cyanobacteria as the time prior (Table S2).

For the second-step sequential dating analysis, the orthologous gene families were firstly identified by searching against a pre-compiled plastid marker gene dataset (Ponce-Toledo et al., 2017) and then sorted by ΔLL (a proxy for the extent to which a marker gene rejects vertical evolution) and evolutionary rate difference (Supplementary Data 2) (Fig. S5). We employed the program “mcmc3r” to select the gene family-specific clock model by following published studies (McGowen et al., 2019; Wang and Luo, 2021). The calibrations and the species tree topology used in the present study were also adapted from publications (see details in Supplementary Methods). For those that remain disputed, we took all the possibilities in our analysis for comparison (Fig. S3). We also tested the necessity for using sequential Bayesian dating approach and hard bound calibration strategy in our analysis (Fig. S6) and tested the convergence of the posterior age estimates (Fig. S7).

The gains and losses of *Prochlorococcus* orthologous gene families were inferred with a gene tree-species tree reconciliation approach according to our recent studies (Zhang et al., 2021; Zhang et al., 2023). The reduction of *Prochlorococcus* selection efficiency in ancient time was inferred by comparing the genome-wide d_R/d_C values between the target lineage and its sister lineage based on our recent study (Luo et al., 2017) (see details in Supplementary Methods).

Code availability

The custom scripts used to analyze the data are available in the online GitHub repository (https://github.com/luolab-cuhk/Prochl-NOE).

Acknowledgements

This work is supported by the Natural Science Foundation of China (42293294), the Hong Kong Research Grants Council (RGC) General Research Fund (GRF) (14110820), the Natural Science Foundation of Guangdong Province, China (2022A1515010844 to HZ), the China Postdoctoral Science Foundation (2021M702296 to HZ), and the CUHK Direct Grant (4053551).

References

Angelis, K., Álvarez-Carretero, S., Dos Reis, M., and Yang, Z. (2018) An Evaluation of Different Partitioning Strategies for Bayesian Estimation of Species Divergence Times. *Syst Biol* **67**: 61-77.

Aoki, S., Ito, M., Iwasaki, W.J.M.b., and evolution (2013) From β-to α-proteobacteria: the origin and evolution of rhizobial nodulation genes nodIJ. **30**: 2494-2508.

Barton, N.H., Charlesworth, B.J.A.r.o.e., and systematics (1984) Genetic revolutions, founder effects, and speciation. **15**: 133-164.

Berube, P.M., Rasmussen, A., Braakman, R., Stepanauskas, R., and Chisholm, S.W. (2019) Emergence of trait variability through the lens of nitrogen assimilation in Prochlorococcus. *eLife* **8**: e41043.

Berube, P.M., Biller, S.J., Kent, A.G., Berta-Thompson, J.W., Roggensack, S.E., Roache-Johnson, K.H. et al. (2015) Physiology and evolution of nitrate acquisition in Prochlorococcus. *The ISME journal* **9**: 1195.

Betts, H.C., Puttick, M.N., Clark, J.W., Williams, T.A., Donoghue, P.C., Pisani, D.J.N.e., and evolution (2018) Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. **2**: 1556-1562.

Bhattacharya, D., and Medlin, L.J.J.P. (1995) Phylogeny of plastids A review based on comparisons of small subunit ribosomal RNA coding regions. **31**: 487-496.

Biller, S.J., Berube, P.M., Lindell, D., and Chisholm, S.W. (2015) Prochlorococcus: the structure and function of collective diversity. *Nature Reviews Microbiology* **13**: 13.

Boden, J.S., Konhauser, K.O., Robbins, L.J., and Sánchez-Baracaldo, P. (2021) Timing the evolution of antioxidant enzymes in cyanobacteria. *Nature Communications* **12**: 1-12.

Braakman, R., Follows, M.J., and Chisholm, S.W. (2017) Metabolic evolution and the self-organization of ecosystems. *Proceedings of the National Academy of Sciences* **114**: E3091-E3100.

Brocks, J.J. (2018) The transition from a cyanobacterial to algal world and the emergence of animals. *Emerging Topics in Life Sciences* **2**: 181-190.

Brocks, J.J., Jarrett, A.J., Sirantoine, E., Hallmann, C., Hoshino, Y., and Liyanage, T. (2017) The rise of algae in Cryogenian oceans and the emergence of animals. *Nature* **548**: 578.

Canfield, D.E., Poulton, S.W., Knoll, A.H., Narbonne, G.M., Ross, G., Goldberg, T., and Strauss, H. (2008) Ferruginous conditions dominated later Neoproterozoic deep-water chemistry. *Science* **321**: 949-952.

Chriki-Adeeb, R., and Chriki, A.J.E.B. (2016) Estimating divergence times and substitution rates in Rhizobia. **12**: EBO. S39070.

Clarkson, M.O., Sweere, T.C., Chiu, C.F., Hennekam, R., Bowyer, F., and Wood, R.A.J.E.-S.R. (2023) Environmental controls on very high δ238U values in reducing sediments: Implications for Neoproterozoic seawater records. 104306.

Dayhoff, M.O. (1972) A model of evolutionary change in proteins. *Atlas of protein sequence and structure* **5**: 89-99.

dos Reis, M., Gunnell, G.F., Barba-Montoya, J., Wilkins, A., Yang, Z., and Yoder, A.D. (2018) Using phylogenomic data to explore the effects of relaxed clocks and calibration strategies on divergence time estimation: primates as a test case. *Syst Biol* **67**: 594-615.

dos Reis, M., Thawornwattana, Y., Angelis, K., Telford, Maximilian J., Donoghue, Philip C.J., and Yang, Z. (2015) Uncertainty in the Timing of Origin of Animals and the Limits of Precision in Molecular Timescales. *Curr Biol* **25**: 2939-2950.

Duchêne, S., Lanfear, R., and Ho, S.Y.W. (2014) The impact of calibration and clock-model choice on molecular estimates of divergence times. *Mol Phylogen Evol* **78**: 277-289.

E. Zuckerkandl, L.P. (1965) Evolutionary divergence and convergence in proteins. In *Evolving genes and proteins*: Elsevier, pp. 97-166.

Flombaum, P., Gallegos, J.L., Gordillo, R.A., Rincón, J., Zabala, L.L., Jiao, N. et al. (2013) Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus. *Proceedings of the National Academy of Sciences* **110**: 9824-9829.

Fournier, G., Moore, K., Rangel, L., Payette, J., Momper, L., and Bosak, T. (2021) The Archean origin of oxygenic photosynthesis and extant cyanobacterial lineages. *Proceedings of the Royal Society B* **288**: 20210675.

Goericke, R., Olson, R., and Shalapyonok, A.J.D.S.R.P.I.O.R.P. (2000) A novel niche for Prochlorococcus sp. in low-light suboxic environments in the Arabian Sea and the Eastern Tropical North Pacific. **47**: 1183-1205.

Gray, M.W.J.I.r.o.c. (1992) The endosymbiont hypothesis revisited. **141**: 233-357.

Grossman, A.R., Schaefer, M.R., Chiang, G.G., and Collier, J.L. (1993) The phycobilisome, a light-harvesting complex responsive to environmental conditions. *Microbiol Mol Biol Rev* **57**: 725-749.

Heckman, D.S., Geiser, D.M., Eidell, B.R., Stauffer, R.L., Kardos, N.L., and Hedges, S.B.J.s. (2001) Molecular evidence for the early colonization of land by fungi and plants. **293**: 1129-1133.

Hedges, S.B., Tao, Q., Walker, M., and Kumar, S.J.P.o.t.N.A.o.S. (2018) Accurate timetrees require accurate calibrations. **115**: E9510-E9511.

Hess, W.R., Rocap, G., Ting, C.S., Larimer, F., Stilwagen, S., Lamerdin, J., and Chisholm, S.W. (2001) The photosynthetic apparatus of Prochlorococcus: insights through comparative genomics. *Photosynthesis Res* **70**: 53-71.

Ito, H., and Tanaka, A. (2011) Evolution of a divinyl chlorophyll-based photosystem in Prochlorococcus. *Proceedings of the National Academy of Sciences* **108**: 18014-18019.

Ito, H., Yokono, M., Tanaka, R., and Tanaka, A. (2008) Identification of a novel vinyl reductase gene essential for the biosynthesis of monovinyl chlorophyll in Synechocystis sp. PCC6803. *J Biol Chem* **283**: 9002-9011.

Johnston, D.T., Wolfe-Simon, F., Pearson, A., and Knoll, A.H. (2009a) Anoxygenic photosynthesis modulated Proterozoic oxygen and sustained Earth's middle age. *Proc Natl Acad Sci U S A* **106**: 16925-16929.

Johnston, D.T., Wolfe-Simon, F., Pearson, A., and Knoll, A.H. (2009b) Anoxygenic photosynthesis modulated Proterozoic oxygen and sustained Earth's middle age. *Proceedings of the National Academy of Sciences* **106**: 16925-16929.

Jones, C., Nomosatryo, S., Crowe, S.A., Bjerrum, C.J., and Canfield, D.E.J.G. (2015) Iron oxides, divalent cations, silica, and the early earth phosphorus crisis. **43**: 135-138.

Kang, J., Gill, B., Reid, R., Zhang, F., and Xiao, S. (2023) Nitrate limitation in early Neoproterozoic oceans delayed the ecological rise of eukaryotes. *Science Advances* **9**: eade9647.

Keeling, P.J. (2013) The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annu Rev Plant Biol* **64**: 583-607.

Knoll, A.H., and Nowak, M.A. (2017) The timetable of evolution. *Science Advances* **3**: e1603076.

Lau, K.V., Macdonald, F.A., Maher, K., and Payne, J.L. (2017) Uranium isotope evidence for temporary ocean oxygenation in the aftermath of the Sturtian Snowball Earth. *Earth and Planetary Science Letters* **458**: 282-292.

Lavin, P., González, B., Santibáñez, J.F., Scanlan, D.J., and Ulloa, O.J.E.m.r. (2010) Novel lineages of Prochlorococcus thrive within the oxygen minimum zone of the eastern tropical South Pacific. **2**: 728-738.

Luo, H., Huang, Y., Stepanauskas, R., and Tang, J. (2017) Excess of non-conservative amino acid changes in marine bacterioplankton lineages with reduced genomes. *Nature microbiology* **2**: 17091.

Lyons, T.W., Reinhard, C.T., and Planavsky, N.J. (2014) The rise of oxygen in Earth's early ocean and atmosphere. *Nature* **506**: 307.

Marin, B., M. Nowack, E.C., and Melkonian, M. (2005) A Plastid in the Making: Evidence for a Second Primary Endosymbiosis. *Protist* **156**: 425-432.

Martiny, A.C., Kathuria, S., and Berube, P.M. (2009) Widespread metabolic potential for nitrite and nitrate assimilation among Prochlorococcus ecotypes. *Proceedings of the National Academy of Sciences* **106**: 10787-10792.

McGowen, M.R., Tsagkogeorga, G., Álvarez-Carretero, S., dos Reis, M., Struebig, M., Deaville, R. et al. (2019) Phylogenomic Resolution of the Cetacean Tree of Life Using Target Sequence Capture. *Syst Biol* **69**: 479-501.

Michiels, C.C., Darchambeau, F., Roland, F.A., Morana, C., Llirós, M., García-Armisen, T. et al. (2017) Iron-dependent nitrogen cycling in a ferruginous lake and the nutrient status of Proterozoic oceans. *Nature Geoscience* **10**: 217-221.

Morris, J.L., Puttick, M.N., Clark, J.W., Edwards, D., Kenrick, P., Pressel, S. et al. (2018) The timescale of early land plant evolution. **115**: E2274-E2283.

Och, L.M., and Shields-Zhou, G.A. (2012) The Neoproterozoic oxygenation event: Environmental perturbations and biogeochemical cycling. *Earth-Sci Rev* **110**: 26-57.

Ozaki, K., Thompson, K.J., Simister, R.L., Crowe, S.A., and Reinhard, C.T.J.N.c. (2019) Anoxygenic photosynthesis and the delayed oxygenation of Earth's atmosphere. **10**: 3026.

Palenik, B., Brahamsha, B., Larimer, F., Land, M., Hauser, L., Chain, P. et al. (2003) The genome of a motile marine Synechococcus. **424**: 1037-1042.

Parfrey, L.W., Lahr, D.J., Knoll, A.H., and Katz, L.A.J.P.o.t.N.A.o.S. (2011) Estimating the timing of early eukaryotic diversification with multigene molecular clocks. **108**: 13624-13629.

Partensky, F., and Garczarek, L. (2010) Prochlorococcus: advantages and limits of minimalism. *Annual Review of Marine Science* **2**: 305-331.

Partensky, F., Hess, W.R., and Vaulot, D. (1999) Prochlorococcus, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* **63**: 106-127.

Ponce-Toledo, R.I., Deschamps, P., López-García, P., Zivanovic, Y., Benzerara, K., and Moreira, D. (2017) An Early-Branching Freshwater Cyanobacterium at the Origin of Plastids. *Curr Biol* **27**: 386-391.

Ralf, G., and Repeta, D.J. (1992) The pigments of Prochlorococcus marinus: The presence of divinylchlorophyll a and b in a marine procaryote. *Limnol Oceanogr* **37**: 425-433.

Rubio, L.M., Flores, E., and Herrero, A.J.J.o.b. (1998) The narA locus of Synechococcus sp. strain PCC 7942 consists of a cluster of molybdopterin biosynthesis genes. **180**: 1200-1206.

Sánchez-Baracaldo, P. (2015) Origin of marine planktonic cyanobacteria. *Scientific reports* **5**: 17418.

Sánchez-Baracaldo, P., Ridgwell, A., and Raven, J.A. (2014) A neoproterozoic transition in the marine nitrogen cycle. *Curr Biol* **24**: 652-657.

Sánchez-Baracaldo, P., Raven, J.A., Pisani, D., and Knoll, A.H. (2017) Early photosynthetic eukaryotes inhabited low-salinity habitats. *Proceedings of the National Academy of Sciences*: 201620089.

Sánchez-Baracaldo, P., Bianchini, G., Wilson, J.D., and Knoll, A.H. (2022) Cyanobacteria and biogeochemical cycles through Earth history. *Trends Microbiol* **30**: 143-157.

Satoh, S., and Tanaka, A. (2006) Identification of chlorophyllide a oxygenase in the Prochlorococcus genome by a comparative genomic approach. *Plant Cell Physiol* **47**: 1622-1629.

Schirrmeister, B.E., Gugger, M., and Donoghue, P.C. (2015) Cyanobacteria and the Great Oxidation Event: evidence from genes and fossils. *Palaeontology* **58**: 769-785.

Schirrmeister, B.E., Sanchez-Baracaldo, P., and Wacey, D. (2016) Cyanobacterial evolution during the Precambrian. *International Journal of Astrobiology* **15**: 187-204.

Shih, P.M., Hemp, J., Ward, L.M., Matzke, N.J., and Fischer, W.W. (2016) Crown group Oxyphotobacteria postdate the rise of oxygen. *Geobiology* **15**: 19-29.

Stolper, D.A., and Keller, C.B.J.N. (2018) A record of deep-ocean dissolved O2 from the oxidation state of iron in submarine basalts. **553**: 323-327.

Strassert, J.F.H., Irisarri, I., Williams, T.A., and Burki, F. (2021) A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. *Nat Commun* **12**: 1879.

Tahata, M., Sawaki, Y., Yoshiya, K., Nishizawa, M., Komiya, T., Hirata, T. et al. (2015) The marine environments encompassing the Neoproterozoic glaciations: evidence from C, Sr and Fe isotope ratios in the Hecla Hoek Supergroup in Svalbard. *Precambrian Research* **263**: 19-42.

Ting, C.S., Rocap, G., King, J., and Chisholm, S.W. (2002) Cyanobacterial photosynthesis in the oceans: the origins and significance of divergent light-harvesting strategies. *Trends Microbiol* **10**: 134-142.

Tostevin, R., and Mills, B.J.J.I.F. (2020) Reconciling proxy records and models of Earth's oxygenation during the Neoproterozoic and Palaeozoic. **10**: 20190137.

Trower, E.J., Gutoski, J.R., Wala, V.T., Mackey, T.J., and Simpson, C. (2023) Tonian Low-Latitude Marine Ecosystems Were Cold Before Snowball Earth. *Geophysical Research Letters* **50**: e2022GL101903.

Ulloa, O., Henríquez-Castillo, C., Ramírez-Flandes, S., Plominsky, A.M., Murillo, A.A., Morgan-Lang, C. et al. (2021) The cyanobacterium Prochlorococcus has divergent light-harvesting antennae and may have evolved in a low-oxygen ocean. *Proceedings of the National Academy of Sciences* **118**.

Wang, S., and Luo, H. (2021) Dating Alphaproteobacteria evolution with eukaryotic fossils. *Nature Communications* **12**: 3324.

Warnock, R.C., Yang, Z., and Donoghue, P.C.J.B.l. (2012) Exploring uncertainty in the calibration of the molecular clock. **8**: 156-159.

Zhang, F., Stockey, R.G., Xiao, S., Shen, S.-z., Dahl, T.W., Wei, G.-Y. et al. (2022) Uranium isotope evidence for extensive shallow water anoxia in the early Tonian oceans. *Earth and Planetary Science Letters* **583**: 117437.

Zhang, H., Hellweger, F.L., and Luo, H.J.b. (2023) Genome reduction occurred in early Prochlorococcus with an unusually low effective population size. 2023.2006. 2025.546417.

Zhang, H., Sun, Y., Zeng, Q., Crowe, S.A., and Luo, H. (2021) Snowball Earth, population bottleneck and Prochlorococcus evolution. *Proceedings of the Royal Society B: Biological Sciences* **288**: 20211956.

Zhang, J. (2000) Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *J Mol Evol* **50**: 56-68.

Figure Legends

Fig. 1 The evolutionary timeline of *Prochlorococcus* estimated with plastid-based strategy. (a) The diagram shows the phylogenetic relations among Cyanobacteria and eukaryotic species. The green solid line, black solid line and green dashed line in the cladogram leading to the tip of oxygenic Cyanobacteria, anoxygenic Vampirovibrionia and Sericytochromatia and eukaryotic species, respectively. The calibrated nodes in molecular clock analysis are marked with orange circle (see calibration justifications in Supplementary Methods). (b) (left) The *Prochlorococcus* evolutionary timeline estimated with the focal molecular dating strategy using sequential Bayesian dating approach based on the 5-partition eukaryotic genome-scale data (in the first step dating analysis) and fully partitioned 12-gene dataset "T30" (in the second step dating analysis) under independent rate clock model. The posterior age distribution is provided next to the ancestral node. The atmospheric oxygen level at the geological time scale is adapted from Lyons et al., 2014, which is represented by the percent of present atmospheric oxygen level (PAL). The vertical bars with green, orange and blue colors represent the time of Tonian, the time of NOE, and the time of Sturtian (left) and Marinoan (right) glaciation, respectively. (right) Phyletic pattern of key gene families. Solid square, solid circle and open circle next to each analyzed genome represent multi-copy gene family, single-copy gene family and absence of the gene family, respectively, in the extant genomes. Gene families marked with asterisk (*) were consistently estimated to be gained or lost by using both AnGST and GeneRax. (c) Diagram illustrating the biogeochemical environments when *Prochlorococcus* arose from its *Synechococcus* ancestor.

Fig. 2 Diagram illustrating the schemes and the results of d_R/d_C comparison. The genome-wide means of d_R/d_C values at the ancestral branches ('Target') leading to (a) Proch-AMZI/II/III-LCA and (b) SBE-LCA compared with that at the sister lineages ('Control'). The d_R/d_C values were classified based on the physicochemical classification of the amino acids by charge or by volume and polarity, and were either GC-corrected by codon frequency (blue), GC-corrected by amino acid (AA) frequency (red) or uncorrected (gray). Error bars of d_R/d_C represent the standard error of the mean.

Fig. 1

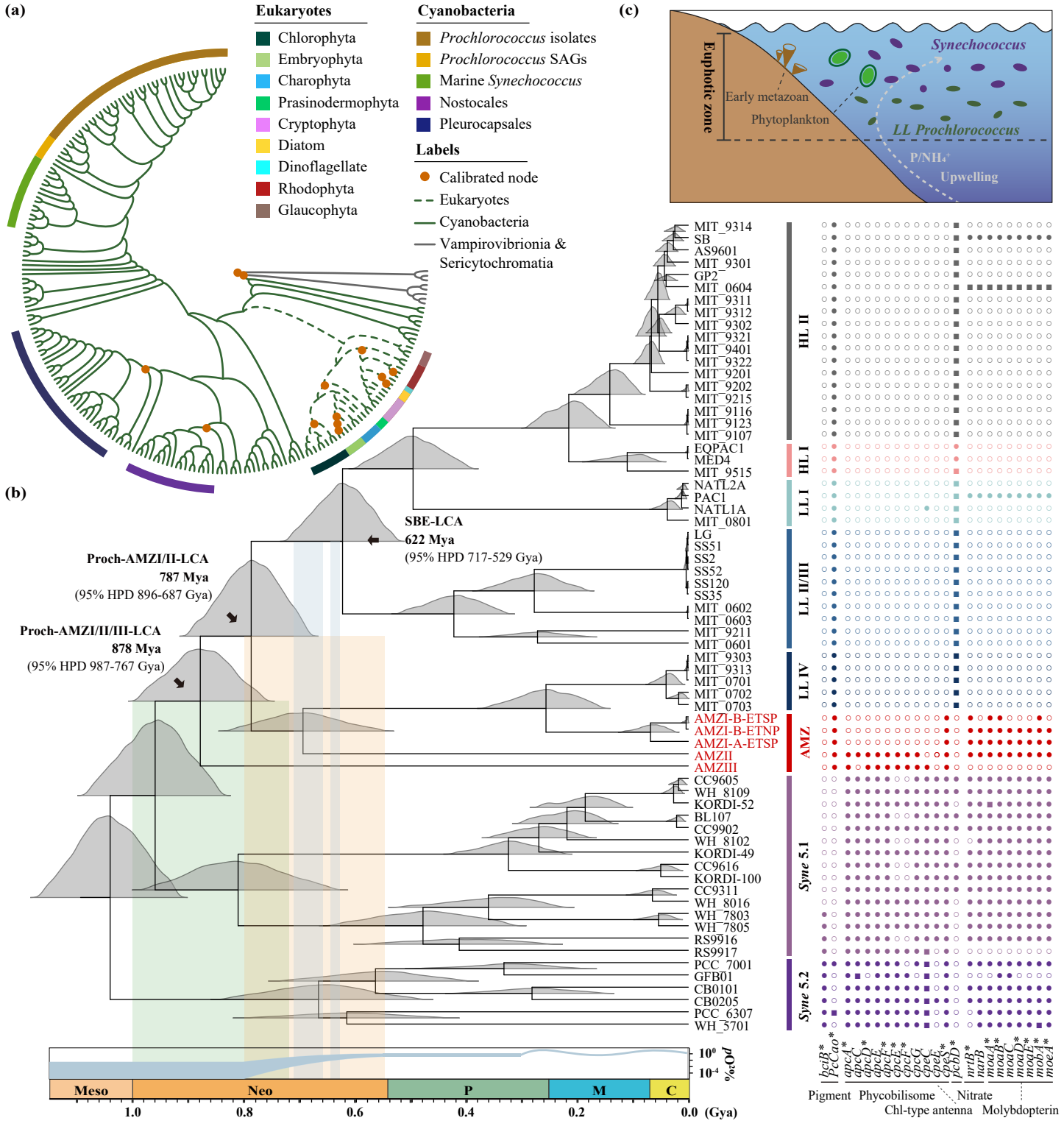
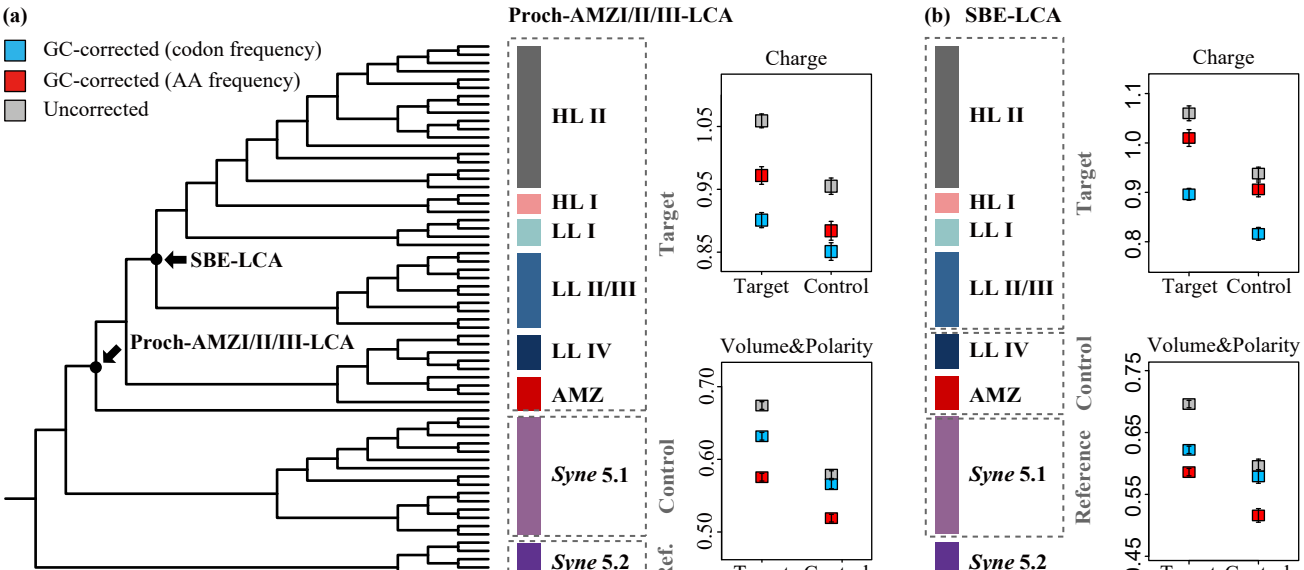

(a)
Eukaryotes
Chlorophyta
Embryophyta
Charophyta
Prasinodermophyta
Cryptophyta
Diatom
Dinoflagellate
Rhodophyta
Glaucophyta
Cyanobacteria
Prochlorococcus isolates
Prochlorococcus SAGs
Marine Synechococcus
Nostocales
Pleurocapsales
Labels
Calibrated node
Eukaryotes
Cyanobacteria
Vampirovibrionia & Sericytochromatia
(c)
Euphotic zone
Synechococcus
Early metazoan
Phytoplankton
LL Prochlorococcus
P/NH4+
Upwelling
(b)
SBE-LCA
622 Mya
(95% HPD 717-529 Gya)
Proch-AMZI/II-LCA
787 Mya
(95% HPD 896-687 Gya)
Proch-AMZI/II/III-LCA
878 Mya
(95% HPD 987-767 Gya)
MIT_9314
SB
AS9601
MIT_9301
GP2
MIT_0604
MIT_9311
MIT_9312
MIT_9302
MIT_9321
MIT_9401
MIT_9322
MIT_9201
MIT_9202
MIT_9215
MIT_9116
MIT_9123
MIT_9107
EQPAC1
MED4
MIT_9515
NATL2A
PAC1
NATL1A
MIT_0801
LG
SS51
SS2
SS52
SS120
SS35
MIT_0602
MIT_0603
MIT_9211
MIT_0601
MIT_9303
MIT_9313
MIT_0701
MIT_0702
MIT_0703
AMZI-B-ETSP
AMZI-B-ETNP
AMZI-A-ETSP
AMZII
AMZIII
CC9605
WH_8109
KORDI-52
BL107
CC9902
WH_8102
KORDI-49
CC9616
KORDI-100
CC9311
WH_8016
WH_7803
WH_7805
RS9916
RS9917
PCC_7001
GFB01
CB0101
CB0205
PCC_6307
WH_5701
HL II
HL I
LL I
LL II/III
LL IV
AMZ
Syne 5.1
Syne 5.2
bciB*
PcCao*
apcA*
apcC*
apcD*
apcE*
apcF*
cpcE*
cpcF*
cpcG*
cpeC*
cpeS*
pcbD*
nrtB*
narB*
moaA*
moaB*
moaC*
moaD*
moaE*
moeA*
Pigment
Phycobilisome
Nitrate
Chl-type antenna
Molybdopterin
10^0
10^-4
%Od
Meso
Neo
P
M
C
1.0
0.8
0.6
0.4
0.2
0.0 (Gya)

Fig. 2


(a)
GC-corrected (codon frequency)
GC-corrected (AA frequency)
Uncorrected
SBE-LCA
Proch-AMZI/II/III-LCA
Proch-AMZI/II/III-LCA
HL II
HL I
LL I
LL II/III
LL IV
AMZ
Syne 5.1
Syne 5.2
Target
Control
Ref.
Charge
Volume&Polarity
Target
Control
(b) SBE-LCA
HL II
HL I
LL I
LL II/III
LL IV
AMZ
Syne 5.1
Syne 5.2
Target
Control
Reference
Charge
Volume&Polarity
Target
Control