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.

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 analysis reveals that *Prochlorococcus* emerged in the deep photic zone of the Tonian (1,000-720 Mya) oceans. The classical light-harvesting antenna complex in Cyanobacteria, i.e., the phycobilisome, was replaced in *Prochlorococcus* by the chlorophyll□based antenna, enabling more efficient use of blue light that penetrates into deeper water. Importantly, *Prochlorococcus* colonization of deep water enhanced access to phosphate, which was abundant in upwelled seawater, but likely scarce in the Tonian surface ocean, promoting expansion of *Prochlorococcus*, displacement of incumbent low-light adapted anoxygenic photoferrotrophs, and associated increases in 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.

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 organism 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).

The emergence of an early branch of *Prochlorococcus* (named SBE-LCA) during the Cryogenian based on our recent study (Zhang et al. 2021) implies the origin of the total

Prochlorococcus group earlier in the Proterozoic Eon. Such an earlier origin then suggests that *Prochlorococcus* might have contributed to the cyanobacterial dominated primary production that supported marine biogeochemical cycles and underpinned dynamic ocean chemistry in the run-up to the Sturtian Snowball Earth glaciation. This was an important period in Earth's history that ultimately led to an Earth system succession and the rise of eukaryotes with algae emerging as key primary producers before the end of the Cryogenian (Brocks et al. 2017). Deciphering the potential role, if any, that *Prochlorococcus* might have played depends critically on an accurate estimate of the origin time for *Prochlorococcus*.

By far, however, the divergence time of *Prochlorococcus* remains contentious from ~200 Mya to ~1,000 Mya (Sánchez-Baracaldo et al., 2014; Sánchez-Baracaldo, 2015; Schirrmeister et al., 2015; Sánchez-Baracaldo et al., 2017; Boden et al., 2021; Fournier et al., 2021; Martinez-Gutierrez et al., 2023). 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 poses a notable challenge in determining the antiquity of *Prochlorococcus*, particularly considering the lack of maximum age information when only cyanobacterial fossils were used (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 Bayesian sequential dating approach to better constrain the divergence time of eukaryotic lineages thereby propagating more accurate time information to Cyanobacteria, including *Prochlorococcus*. Note that the Bayesian sequential 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). These joint analyses led us to conclude that *Prochlorococcus* arose in the Tonian (1,000-720 Mya) oceans. Further, population genetic analysis 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 ancestors (i.e., the last common ancestor of *Prochlorococcus* and its sister clade in the genus *Synechococcus*) thrived.

Results and Discussion

Prochlorococcus originated in the Tonian ocean

We implemented a Bayesian sequential dating method that takes advantage of the abundant fossil records (Fig. 1a; Fig. S1a) available from photosynthetic eukaryotes to calibrate the evolution of Cyanobacteria. In our implementations, 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 only a few time constraints are available. To achieve this goal, we first implemented the genome-scale dating analysis of the eukaryotic lineages and confirmed that the posterior ages of the crown eukaryotic group were consistent with the previous estimates at ~1.6 Gya (Parfrey et al., 2011; Betts et al., 2018;

Wang and Luo, 2021) (Fig. S1b). We then compared the posterior age distributions of eukaryotic nodes from the first-step analysis to the distributions of the effective time prior on the corresponding nodes in the second-step analysis. The nearly identical distributions found in the comparisons suggest that the Bayesian sequential dating approach works well in these cases (Fig. S2).

Our dating analysis implies that the last common ancestor (LCA) of *Prochlorococcus* (denoted as "Proch-AMZI/II/III-LCA") 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; 800-550 Mya) (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).

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 (1,000-720 Mya), even when the maximum root age changed from 3.8 Gya to 4.5 Gya (Fig. S3). Moreover, the

151 emergence of Proch-AMZI/II-LCA was consistently estimated to occur in the early stages of
152 the NOE (800-550 Mya) (Fig. S4), and that the branch leading to the SBE-LCA encompasses
153 the Snowball Earth events in all the dating analyses (Fig. S5; see Supplementary Results).
154
155 The emergence of *Prochlorococcus* is associated with a founder effect
156 The method for inferring selection efficiency in deep time was developed (Zhang, 2000)
157 and recently improved (Luo et al., 2017), which involves nonsynonymous substitutions only
158 and compares the rate of nonsynonymous substitutions leading to replacements of
159 physicochemically dissimilar amino acids (i.e., radical changes; d_R) to the rate of
160 nonsynonymous substitutions leading to replacements of physicochemically similar amino
161 acids (i.e., conservative changes; d_C). Since radical changes are more likely to be deleterious
162 than conservative changes (E. Zuckerkandl, 1965; Dayhoff, 1972), an excess of the radical
163 changes in a deeply branching lineage compared to its sister lineage suggests random fixation
164 of deleterious mutations by genetic drift in the former. Using this method, we found a
165 significant increase of the d_R/d_C ratio across the genomic regions in the branch leading to
166 Proch-AMZI/II/III-LCA relative to the branch leading to the LCA of *Synechococcus* clade 5.1
167 (Fig. 2a), indicating that the emergence of *Prochlorococcus* was accompanied by a significant
168 reduction of the efficiency of purifying selection and a potentially severe reduction in the
169 effective population size (N_e), allowing for an accelerated accumulation of deleterious
170 mutations through genetic drift.
171 We propose that the colonization of the deep photic zone by *Prochlorococcus* started by a
172 seed population that obtained the ability to use divinyl chlorophylls. Therefore, the reduced

173 N_e upon the emergence of *Prochlorococcus* was likely caused by a "founder effect", which
174 suggests that a new population was established by a few individuals from a larger ancestral
175 population (Barton et al., 1984). Note that this mechanism is different from population
176 bottleneck, which is often caused by the environmental disasters. The latter has been used to
177 explain the reduced N_e occurring on the branch leading to SBE-LCA (Zhang et al., 2021).
178 That conclusion is confirmed here with the inclusion of new SAG genomes, as the d_R/d_C
179 values in the branch leading to SBE-LCA were significantly elevated compared to the those
180 in the branches leading to *Prochlorococcus* LLIV and AMZ clades (Fig. 2b)
181
182 <u>*Prochlorococcus* genomic changes and niche adaptation</u>
183 As the only phytoplankton group using divinyl chlorophyll (DVChl) for harvesting light
184 energy (Ralf and Repeta, 1992), *Prochlorococcus* lost the gene *bciB* that performs the
185 conversion of DVChl to MVChl (monovinyl chlorophyll) at Proch-AMZI/II/III-LCA, thereby
186 promoting the accumulation of DVChl in their membranes (Ito et al., 2008). Compared to the
187 MVChl used by *Synechococcus*, DVChl more efficiently absorbs blue light that penetrates to
188 deeper waters than other photosynthetically active radiation (Ito and Tanaka, 2011). Likewise,
189 the gain of the *PcCao* gene for the synthesis of chlorophyll b (Satoh and Tanaka, 2006) in
190 Proch-AMZI/II/III-LCA also enables *Prochlorococcus* ancestors to efficiently harvest blue
191 light at exceedingly low intensities characteristic of deep waters (Hess et al., 2001) (Fig. 1b).
192 These genomic changes would have enabled *Prochlorococcus* ancestors to explore the deep
193 photic zone, a niche where its *Synechococcus* ancestor would not thrive.
194 We infer that dwelling in deeper waters confers competitive advantages upon

195 *Prochlorococcus*. A well-known advantage emerges under conditions of strong phosphorus (P)
196 limitation, whereby deep-dwelling low-light adapted phototrophs gain first access to
197 upwelling phosphorus (Jones et al., 2015; Ozaki et al., 2019). The low-light advantage
198 conferred on the earliest *Prochlorococcus* cells through the adoption of divinyl chlorophyll
199 would have enhanced their access to P, relative to contemporary algae and its own ancestors.
200 It would also have increased access to NH_4^+ upwelling from deeper anoxic waters obviating
201 the need for nitrate assimilation (Michiels et al., 2017) (Fig. 1c).
202 Unlike most cyanobacteria, which use the phycobilisome as the photosynthetic antenna,
203 the main light-harvesting antenna of *Prochlorococcus* is made up of prochlorophyte
204 chlorophyll-binding (Pcb) protein (Biller et al., 2015). By reconstructing the gene
205 evolutionary paths with *Prochlorococcus* SAGs included, we found that the phycobilisome
206 genes (*apcACDEF*, *cpcEFG*, and *cpeCES*) were present in both Proch-AMZI/II/III-LCA and
207 Proch-AMZI/II-LCA and lost at SBE-LCA. Using the same approach, we found that the
208 cholorophyll-binding protein (encoded by *pcbD*) was obtained at Proch-AMZI/II-LCA. The
209 replacement of the photosynthetic antenna thus did not co-occur with the emergence of
210 *Prochlorococcus*. We note that gene absence in SAGs could result from the incomplete nature
211 of the SAG genomes, however, in this case, simultaneous absence of the *pcbD* gene in all the
212 five SAGs that are more than 80% complete seems unlikely.
213 The phycobilisome constitutes as much as 50%-60% of the soluble proteins in
214 *Synechococcus* (Grossman et al., 1993). The loss of the phycobilisome was thought to reduce
215 nitrogen (N) investments by at least 40% in *Prochlorococcus* (Ting et al., 2002). Therefore,
216 the losses of phycobilisome genes was once considered to be favored by *Prochlorococcus*

conferred advantages to inhabiting oligotrophic oceans (Ting et al., 2002). However, since the colonization of *Prochlorococcus* in 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 (Jones et al., 2015; Michiels et al., 2017; Ozaki et al., 2019), the loss of the phycobilisome in *Prochlorococcus* was unlikely driven by selection for metabolic efficiency. Instead, as the 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.

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). 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 *narB* gene inheritance (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

239 both genes in Proch-AMZI/II/III-LCA and Proch-AMZI/II-LCA and the losses of these genes
240 at SBE-LCA (Fig. 1b). The mechanism underlying the loss of nitrate utilization genes in
241 *Prochlorococcus* was once attributed to relaxed selection efficiency due to the low level of
242 nitrate caused by intensive denitrification and anammox (anaerobic ammonium oxidation)
243 activity (Canfield et al., 2008; Johnston et al., 2009). However, since *Prochlorococcus*
244 originated in the deep photic zone where ammonium would likely have been supplied
245 through upwelling (Fig. 1c) (Michiels et al., 2017), the loss of the nitrate utilization genes in
246 early *Prochlorococcus* is more likely to be the result of a switch of its N source from nitrate
247 to ammonium, which was again a neutral process.

248 In addition to the nitrate utilization genes, we inferred the losses of molybdopterin
249 biosynthesis genes at SBE-LCA. Since these genes are known to co-locate with the nitrate
250 reductase genes in *Synechococcus* (Rubio et al., 1998; Palenik et al., 2003), they may
251 function as the cofactor of nitrate reductase in *Prochlorococcus*. Therefore, the loss of nitrate
252 reductase in *Prochlorococcus* likely rendered the molybdopterin dispensable and thus led to
253 the losses of molybdopterin biosynthesis genes (*moaABCDE*, *mobA* and *moeA*) at SBE-LCA.
254

255 Geochemical context that supports the emergence of *Prochlorococcus* in Tonian ocean

256 A Tonian, pre-Sturtian, age for the emergence of the *Prochlorococcus* lineage would
257 have been set against the backdrop of dynamic ocean chemistry characterized by a variably
258 oxygenated surface ocean that was co-populated by diverse microbial eukaryotes, including
259 phytoplankton and the earliest metazoans (Erwin et al., 2011) (Fig. 1c). Multi-proxy data,
260 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). Surface 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.

In this way, the emergence of the *Prochlorococcus* would have restricted the flux of nutrients to surface 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., 2009; 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). Whereas the accumulation of O_2 produced through photosynthesis ultimately depends on organic matter burial (Berner, 1991), oxygen

283 production fluxes through organic matter burial scale with the fraction of total photosynthetic
284 production that is oxygenic rather than anoxygenic (Johnston et al., 2009; Ozaki et al., 2019).
285 In this way, the capacity of the *Prochlorococcus* LCA to grow in deeper waters would have
286 enhanced its ability to access upwelling phosphorus thereby increasing the oxygenic fraction
287 of total photosynthesis with potential to initiate a positive feedback on oxygenation (Ozaki et
288 al., 2019).
289
290 Implications for Neoproterozoic biogeochemical cycling
291 In accordance with previous research (Ulloa et al., 2021), our analysis suggests that
292 *Prochlorococcus* diverged before transitioning from the use of the phycobilisome to the
293 chlorophyll□based light□harvesting complex. Therefore, while the divinyl pigment synthesis
294 gene was obtained at the earliest *Prochlorococcus* (i.e., Proch-AMZI/II/III-LCA) to enable
295 their exploration of the deep photic zone, our analyses imply that *Prochlorococcus* would not
296 have strongly influenced the Neoproterozoic Earth system until the emergence of Proch-
297 AMZI/II-LCA lineage during which the photosynthetic antenna was replaced. Our dating
298 analysis shows that the emergence of this lineage indeed broadly coincides with geochemical
299 signals for the early stages of Neoproterozoic Oxygenation Event (NOE).
300 The NOE is widely considered the second stage of Earth's protracted oxygenation history,
301 during which atmospheric and marine oxygen concentrations rose to levels exceeding those
302 that characterized the proceeding mid-Proterozoic (Och and Shields-Zhou, 2012; Lyons et al.,
303 2014). This rise in atmospheric oxygen ultimately promoted the emergence of large and
304 complex animal life (Knoll and Nowak, 2017). The initiation of the NOE was previously

linked to eukaryotic algae (Lyons et al., 2014; Brocks et al., 2017) or marine picocyanobacteria (Sánchez-Baracaldo et al., 2014; Braakman et al., 2017). Our results show support for the latter assumption and indicate that early *Prochlorococcus* may have played a role in enhancing oxygen production, with potential to initiate the NOE, which was likely ultimately accelerated through increasing efficiency in the biological carbon pump driven by a progressively a larger role for eukaryotes. Such an elevated role for eukaryotes is clearly marked in the geologic record through an increase in sterane-hopane ratios following the Sturtian glaciation (Brocks et al., 2017). Pre-Sturtian, Tonian oxygenation, was thus likely driven by bacterially dominated carbon cycling, the timing of which strongly coincides with the emergence of low-light adapted *Prochlorococcus*, their ensuing colonization of the deep photic zone, and capacity to displace incumbent anoxygenic photoferrotrophs thereby increasing oxygen production. 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. Note that the inference of the roles of *Prochlorococcus* in facilitating Neoproterozoic biogeochemical cycle relies on the accurate molecular dating analysis. Given the heavy debate about the antiquity of Cyanobacteria, including *Prochlorococcus*, we have extensively discussed the previous dating analyses and pointed out their methodological deficiencies (see details in Supplementary Discussion).

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 and robust molecular dating pipeline pinpoint the origin of *Prochlorococcus* to 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 its *Synechococcus* ancestor and early Eukaryotes. 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. While all of these co-existing planktonic groups are likely to have contributed to the NOE, the emergence of *Prochlorococcus* in a previously uncolonized habitat with a unique light regime and a higher phosphate accessibility, followed by the acquisition of a more efficient light-harvesting system, may have facilitated a potentially rapid population expansion, which ultimately enhanced the oxygen production in Tonian oceans and facilitated the 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 Bayesian sequential 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). 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 clustered the eukaryotic gene families into 1-, 3-, 5-, 7-, and 9-partitions by using Gaussian Mixture Model (GMM) clustering method based on their estimated evolutionary rates (see details in Supplementary Methods). The 5-partition data was finally adopted for the first-step dating analysis because of the lowest Akaike Information Criterion (AIC) value (Fig. S1b) and the high similarity between the derived distributions of posterior ages and effective time prior (Fig. S2). The eukaryotic fossil-based calibrations used in this first-step dating analysis were adapted from published studies (Fig. S1a; Table S1; See details in Supplementary Methods).

371 For the second-step sequential dating analysis, the orthologous gene families were firstly
372 identified by searching against a pre-compiled plastid marker gene dataset (Ponce-Toledo et
373 al., 2017) and then sorted by ΔLL (i.e., the difference in log-likelihood values of the gene tree
374 constructed with and without the backbone species tree topology) and evolutionary rate
375 difference (Supplementary Data 2) (Fig. S6). In addition to the fossil-based calibrations
376 adapted from published studies, we used the posterior age distributions derived from the first-
377 step dating analysis to calibrate the overlapping nodes in the species tree containing both
378 eukaryotes and Cyanobacteria (Fig. S1c; Table S1; Table S2). Note that we placed the
379 cyanobacterial-fossil based calibrations at the total group of Nostocales and Pleurocapsales
380 by taking into account the incomplete taxon sampling of their early-split lineages, as well as
381 the possibility that their fossils belong to the stem lineages rather than the crown groups. In
382 this way, we allow for the posterior ages of crown Nostocales and Pleurocapsales groups to
383 be younger than the known fossil records.
384 Our molecular dating analysis was conducted using MCMCTree with the best-fit clock
385 model, which was selected using the program "mcmc3r" following published studies
386 (McGowen et al., 2019; Wang and Luo, 2021). For calibrations and species tree topology that
387 remain disputed, we took all the possibilities in our analysis for comparison (Fig. S4). We
388 also tested the necessity for using Bayesian sequential dating approach and hard bound
389 calibration strategy in our analysis (Fig. S7; see details in Supplementary Discussion) and
390 tested the convergence of the posterior age estimates (Fig. S8).
391 The gains and losses of *Prochlorococcus* orthologous gene families were inferred with a
392 gene tree-species tree reconciliation approach according to our recent studies (Zhang et al.,

2024). The best-fit reconciliation tool was selected through a comprehensive simulation-based benchmarking analysis [see Fig. S5 in (Zhang et al., 2024)]. In our implementations, we performed the reconstruction 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) and then used the same reconstruction methods to validate these evolutionary events when SAGs were included.

The reduction of *Prochlorococcus* selection efficiency in ancient time was inferred by comparing the genome-wide d_R/d_C (i.e., the relative rate of radical versus conservative nonsynonymous substitutions) 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).

415 **References**

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

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

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

423 Berner, R.A. (1991) A model for atmospheric CO 2 over Phanerozoic time. *American Journal*
424 *of Science* **291**: 339-376.

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

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

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

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

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

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

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

443 Brocks, J.J. (2018) The transition from a cyanobacterial to algal world and the emergence of

444 animals. *Emerging topics in life sciences* **2**: 181-190.

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

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

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

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

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

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

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

463 Duchêne, S., Lanfear, R., and Ho, S.Y.W. (2014) The impact of calibration and clock-model
464 choice on molecular estimates of divergence times. *Molecular Phylogenetics and Evolution*
465 **78**: 277-289.

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

468 Erwin, D.H., Laflamme, M., Tweedt, S.M., Sperling, E.A., Pisani, D., and Peterson, K.J.
469 (2011) The Cambrian conundrum: early divergence and later ecological success in the early
470 history of animals. *Science* **334**: 1091-1097.

471 Flombaum, P., Gallegos, J.L., Gordillo, R.A., Rincón, J., Zabala, L.L., Jiao, N. et al. (2013)
472 Present and future global distributions of the marine Cyanobacteria Prochlorococcus and

473 Synechococcus. *Proceedings of the National Academy of Sciences* **110**: 9824-9829.
474 Fournier, G., Moore, K., Rangel, L., Payette, J., Momper, L., and Bosak, T. (2021) The
475 Archean origin of oxygenic photosynthesis and extant cyanobacterial lineages. *Proceedings*
476 *of the Royal Society B* **288**: 20210675.
477 Goericke, R., Olson, R., and Shalapyonok, A. (2000) A novel niche for Prochlorococcus sp.
478 in low-light suboxic environments in the Arabian Sea and the Eastern Tropical North Pacific.
479 *Deep Sea Research Part I: Oceanographic Research Papers* **47**: 1183-1205.
480 Gray, M.W.J.I.r.o.c. (1992) The endosymbiont hypothesis revisited. **141**: 233-357.
481 Grossman, A.R., Schaefer, M.R., Chiang, G.G., and Collier, J.L. (1993) The phycobilisome, a
482 light-harvesting complex responsive to environmental conditions. *Microbiology and*
483 *Molecular Biology Reviews* **57**: 725-749.
484 Heckman, D.S., Geiser, D.M., Eidell, B.R., Stauffer, R.L., Kardos, N.L., and Hedges, S.B.J.s.
485 (2001) Molecular evidence for the early colonization of land by fungi and plants. **293**: 1129-
486 1133.
487 Hedges, S.B., Tao, Q., Walker, M., and Kumar, S.J.P.o.t.N.A.o.S. (2018) Accurate timetrees
488 require accurate calibrations. **115**: E9510-E9511.
489 Hess, W.R., Rocap, G., Ting, C.S., Larimer, F., Stilwagen, S., Lamerdin, J., and Chisholm,
490 S.W. (2001) The photosynthetic apparatus of Prochlorococcus: insights through comparative
491 genomics. *Photosynthesis Research* **70**: 53-71.
492 Ito, H., and Tanaka, A. (2011) Evolution of a divinyl chlorophyll-based photosystem in
493 Prochlorococcus. *Proceedings of the National Academy of Sciences* **108**: 18014-18019.
494 Ito, H., Yokono, M., Tanaka, R., and Tanaka, A. (2008) Identification of a novel vinyl
495 reductase gene essential for the biosynthesis of monovinyl chlorophyll in Synechocystis sp.
496 PCC6803. *Journal of Biological Chemistry* **283**: 9002-9011.
497 Johnston, D.T., Wolfe-Simon, F., Pearson, A., and Knoll, A.H. (2009) Anoxygenic
498 photosynthesis modulated Proterozoic oxygen and sustained Earth's middle age. *Proceedings*
499 *of the National Academy of Sciences* **106**: 16925-16929.
500 Jones, C., Nomosatryo, S., Crowe, S.A., Bjerrum, C.J., and Canfield, D.E.J.G. (2015) Iron
501 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. (2010) Novel lineages of Prochlorococcus thrive within the oxygen minimum zone of the eastern tropical South Pacific. *Environmental microbiology reports* **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.

Martinez-Gutierrez, C.A., Uyeda, J.C., and Aylward, F.O.J.E. (2023) A timeline of bacterial and archaeal diversification in the ocean. **12**: RP88268.

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. *Systematic Biology* **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-Science Reviews* **110**: 26-57.

Ozaki, K., Thompson, K.J., Simister, R.L., Crowe, S.A., and Reinhard, C.T. (2019) Anoxygenic photosynthesis and the delayed oxygenation of Earth's atmosphere. *Nature communications* **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. *Limnology and Oceanography* **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. *Current Biology* **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.

Satoh, S., and Tanaka, A. (2006) Identification of chlorophyllide a oxygenase in the Prochlorococcus genome by a comparative genomic approach. *Plant and cell physiology* **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.

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 in microbiology* **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.

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**.

589 Wang, S., and Luo, H. (2021) Dating Alphaproteobacteria evolution with eukaryotic fossils.
590 *Nature Communications* **12**: 3324.
591 Warnock, R.C., Yang, Z., and Donoghue, P.C.J.B.l. (2012) Exploring uncertainty in the
592 calibration of the molecular clock. **8**: 156-159.
593 Zhang, F., Stockey, R.G., Xiao, S., Shen, S.-z., Dahl, T.W., Wei, G.-Y. et al. (2022) Uranium
594 isotope evidence for extensive shallow water anoxia in the early Tonian oceans. *Earth and*
595 *planetary science letters* **583**: 117437.
596 Zhang, H., Hellweger, F.L., and Luo, H. (2024) Genome reduction occurred in early
597 Prochlorococcus with an unusually low effective population size. *Isme j* **18**.
598 Zhang, H., Sun, Y., Zeng, Q., Crowe, S.A., and Luo, H. (2021) Snowball Earth, population
599 bottleneck and Prochlorococcus evolution. *Proceedings of the Royal Society B: Biological*
600 *Sciences* **288**: 20211956.
601 Zhang, J. (2000) Rates of conservative and radical nonsynonymous nucleotide substitutions
602 in mammalian nuclear genes. *J Mol Evol* **50**: 56-68.
603

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 Bayesian sequential 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

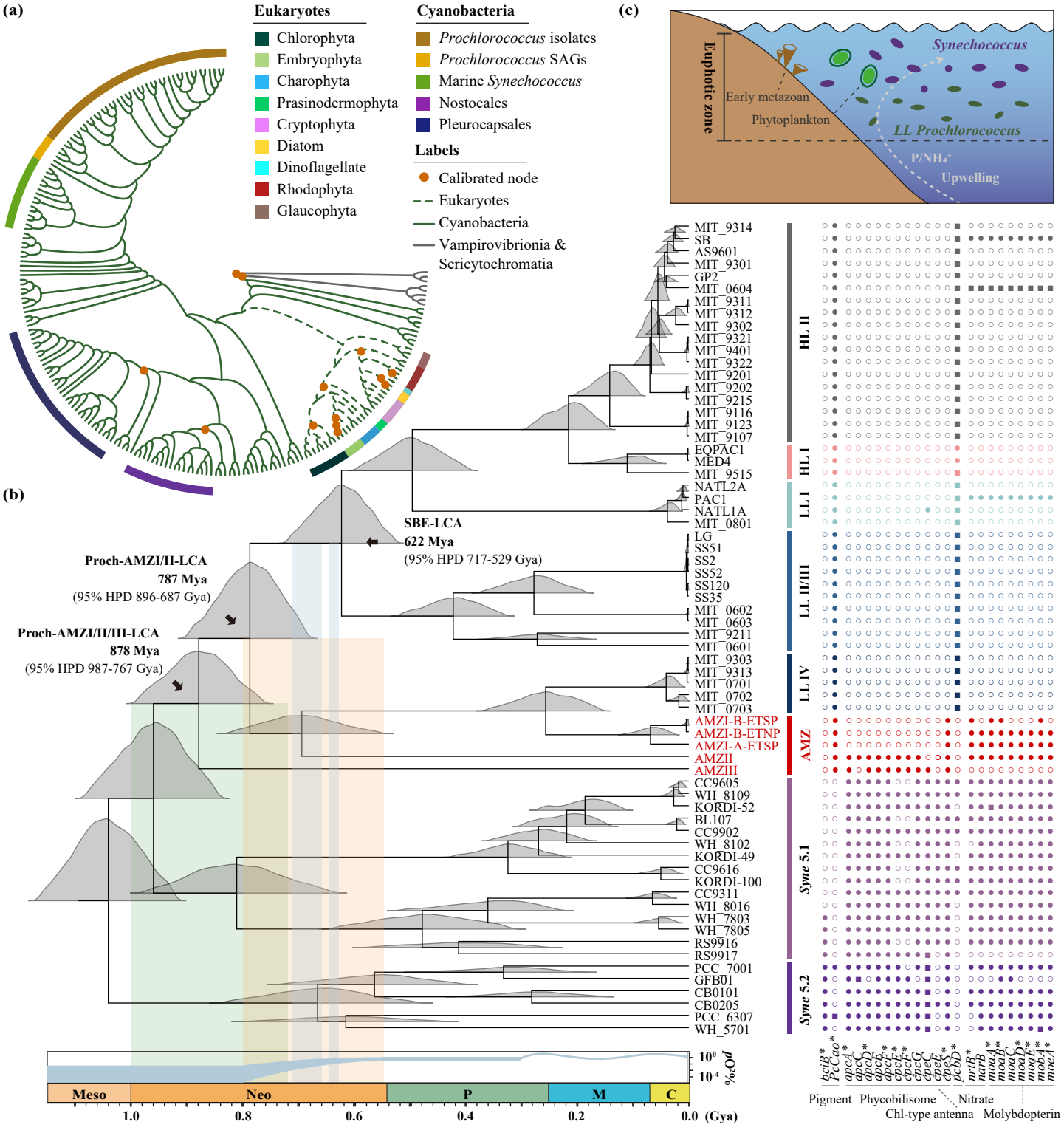
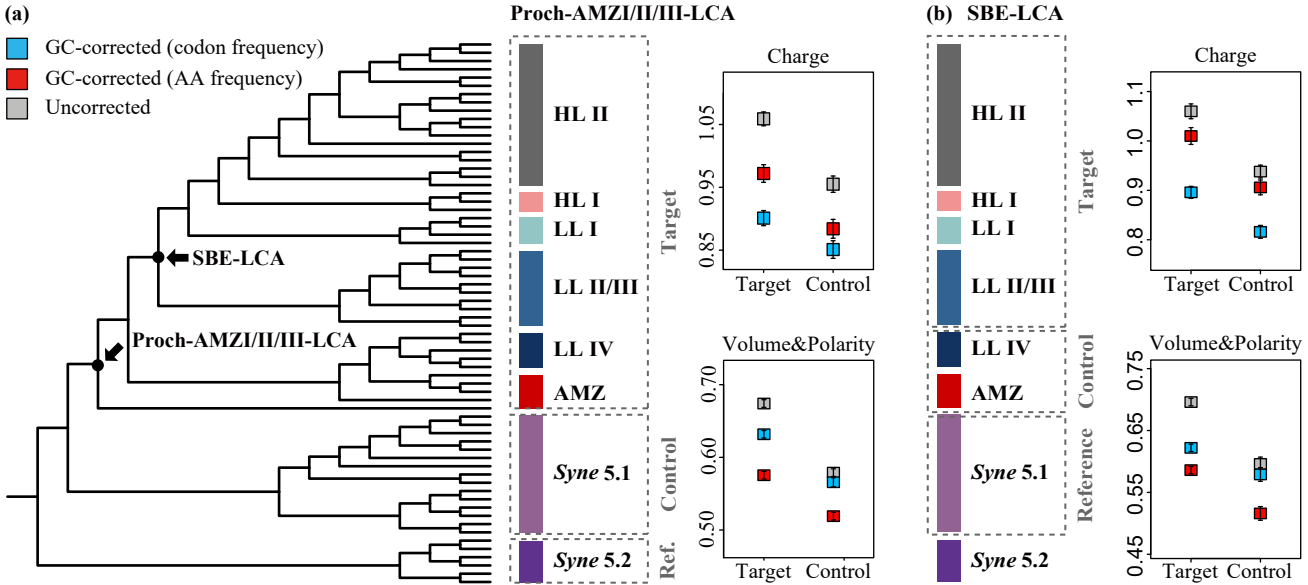

(a)
Eukaryotes
Chlorophyta
Embryophyta
Charophyta
Prasinodermophyta
Cryptophyta
Diatom
Dinoflagellate
Rhodophyta
Glaucophyta
Cyanobacteria
Prochlorococcus isolates
Prochlorococcus SAGs
Marine Synechococcus
Nostocales
Pleurocapsales
Labels
Calibrated node
Eukaryotes
Cyanobacteria
Vampirovibrionia & Sericytochromatia
(c)
Euphotic zone
Early metazoan
Phytoplankton
Synechococcus
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
Meso
Neo
P
M
C
1.0
0.8
0.6
0.4
0.2
0.0 (Gya)
10^0
10^-4
%Opd
bciB*
PcCao*
apcA*
apcC*
apcD*
apcE*
apcF*
cpcB*
cpcC*
cpeE*
cpeS*
pcbD*
nrtB*
narB*
moaA**
moaB*
moaC*
moaE**
moaD**
moeA*
Pigment
Phycobilisome
Nitrate
Chl-type antenna
Molybdopterin

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
Ref. Control
Charge
1.05
0.95
0.85
Target Control
Volume&Polarity
0.70
0.60
0.50
Target Control
(b) SBE-LCA
HL II
HL I
LL I
LL II/III
LL IV
AMZ
Syne 5.1
Syne 5.2
Target
Reference Control
Charge
1.1
1.0
0.9
0.8
Target Control
Volume&Polarity
0.75
0.65
0.55
0.45
Target Control