PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Zhang H, Sun Y, Zeng Q, Crowe SA, Luo H. 2021 Snowball Earth, population bottleneck and *Prochlorococcus* evolution. *Proc. R. Soc. B* **288**: 20211956. https://doi.org/10.1098/rspb.2021.1956

Received: 2 September 2021
Accepted: 26 October 2021

Subject Category:
Evolution

Subject Areas:
microbiology, genetics, evolution

Keywords:
Prochlorococcus, Neoproterozoic Snowball Earth, genome reduction, molecular dating

Author for correspondence:
Haiwei Luo
e-mail: hluo2006@gmail.com

†These authors contributed equally to this study.
‡Current affiliation: Department of Microbiology and Immunology, University of British Columbia, 2350 Health Sciences Mall, Vancouver, British Columbia V6T 1Z3, Canada.

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5700067.

THE ROYAL SOCIETY PUBLISHING

Snowball Earth, population bottleneck and *Prochlorococcus* evolution

Hao Zhang[1,2,†], Ying Sun[2,†], Qinglu Zeng[3], Sean A. Crowe[4,‡] and Haiwei Luo[1,2]

[1]Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen 518000, People's Republic of 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]Department of Ocean Science, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong SAR
[4]Department of Earth Sciences, School of Biological Sciences, and Swire Institute for Marine Science (SWIMS), University of Hong Kong, Pokfulam Road, Hong Kong SAR

HL, 0000-0001-8452-6066

Prochlorococcus are the most abundant photosynthetic organisms in the modern ocean. A massive DNA loss event occurred in their early evolutionary history, leading to highly reduced genomes in nearly all lineages, as well as enhanced efficiency in both nutrient uptake and light absorption. The environmental landscape that shaped this ancient genome reduction, however, remained unknown. Through careful molecular clock analyses, we established that this *Prochlorococcus* genome reduction occurred during the Neoproterozoic Snowball Earth climate catastrophe. The lethally low temperature and exceedingly dim light during the Snowball Earth event would have inhibited *Prochlorococcus* growth and proliferation, and caused severe population bottlenecks. These bottlenecks are recorded as an excess of deleterious mutations accumulated across genomic regions and inherited by descendant lineages. *Prochlorococcus* adaptation to extreme environmental conditions during Snowball Earth intervals can be inferred by tracing the evolutionary paths of genes that encode key metabolic potential. Key metabolic innovation includes modified lipopolysaccharide structure, strengthened peptidoglycan biosynthesis, the replacement of a sophisticated circadian clock with an hourglass-like mechanism that resets daily for dim light adaption and the adoption of ammonia diffusion as an efficient membrane transporter-independent mode of nitrogen acquisition. In this way, the Neoproterozoic Snowball Earth event may have altered the physiological characters of *Prochlorococcus*, shaping their ecologically vital role as the most abundant primary producers in the modern oceans.

1. Introduction

Prochlorococcus are the smallest and most abundant photosynthetic organisms on Earth [1]. They are prevalent throughout the photic zone of the oligotrophic oceans between 40° N and 40° S [1], where they account for more than 40% of the biomass and contribute almost half of the net primary production [2]. *Prochlorococcus* have diversified into two major phylogenetic groups with distinct ecology (ecotypes), with the high-light (HL) adapted monophyletic group imbedded in the low-light (LL) adapted paraphyletic group [3]. The distinct ecotypes of *Prochlorococcus* evolved different pigments, light-harvesting systems and phycobiliproteins, which allowed for efficient light absorption in the water column [4], and thus increased growth rates and primary production [5].

Prochlorococcus genomes have been shaped by stepwise downsizing, including a major genome reduction in their early evolution and a few minor modifications that followed [6]. To understand the evolutionary mechanisms underlying these genome reduction events, which have gained much interest [7], we developed a new method that allows calculating the rates of different types of non-synonymous substitutions leading to radical (d_R) versus

conservative (d_C) changes in amino acid sequences [8]. Since radical changes are more likely to be deleterious than conservative changes, an inflated ratio of d_R/d_C at the genome-wide scale indicates an enhanced role of genetic drift [9,10]. Using this approach reveals that the major genome reduction in *Prochlorococcus* took place under reduced selection efficiency [8] and implies that the ancient population went through severe bottlenecks as the likely result of environmental catastrophe.

The environmental context underlying *Prochlorococcus* genome reduction remains unknown, however, and careful molecular dating is needed to link this important evolutionary event to its possible environmental drivers. By implementing comprehensive molecular clock analyses, we now link the early major genome reduction event of *Prochlorococcus* to the Neoproterozoic Snowball Earth events. These catastrophic disruptions to the Earth system would probably have challenged warm-water-loving photosynthetic *Prochlorococcus*, with strong potential to cause the population bottlenecks inferred from the genome sequences described above. *Prochlorococcus* probably survived this catastrophe through gains and losses of key metabolic functions reconstructed from the same genome sequences, which have a far-reaching impact on their success in today's oceans.

2. Results and discussion

(a) Ancestral *Prochlorococcus* genome reduction and population bottlenecks coincided with Neoproterozoic Snowball Earth events

It was previously shown that *Prochlorococcus* experienced a massive gene loss event on the ancestral branch leading to the last common ancestor (LCA; corresponding to the 'SBE-LCA' in figure 1*a*) of clades HL, LLI and LLII/III [6,7,12]. On the same ancestral branch, it was also shown that genome-wide d_R/d_C is significantly elevated compared to its sister ancestral branch leading to the LCA of the LLIV [8]. Consistent with these, in the present study, we reconstructed 366 and 107 gene family losses and gains on the ancestral branch leading to SBE-LCA, respectively (figure 1*a*; electronic supplementary material, figure S1), and validated the significant inflation of genome-wide d_R/d_C on this branch (both sign test and paired *t*-test, $p < 0.001$; figure 1*b*). These results confirm that the major genome reduction event occurring on this branch was likely driven by genetic drift as a result of one or recurrent population bottlenecks [8].

To establish the environmental context for the large, ancient genome reduction, we estimated the timeline of *Prochlorococcus* evolution by implementing molecular clock analyses based on essential calibrations available in the cyanobacterial lineage. We recognize that the use of calibration sets adapted from previous studies (C1–C8; electronic supplementary material, figure S2) results in up to approximately 320 Ma disparity (electronic supplementary material, figure S2) in the estimated time for the LCA of *Prochlorococcus* HL, LLI and LLII/III clades that emerged with the major genome reduction. We note that the calibrations in previous studies were not properly used. For example, the akinete fossil identified to 2100 Ma was used as either the maximum bound or the minimum bound to calibrate the crown group of Nostocales [13,14]. However, given the fact that an apomorphic character must evolve earlier than the divergence of the crown group, morphological fossils can only serve as the minimum bounds on total groups of assigned lineages [15]. Thus, in the present study, we modified the calibration sets by constraining the lower bounds of the Nostocales (and the Pleurocapsales) total groups with morphological fossils and left their upper bounds unconstrained (C9–C14; electronic supplementary material, figure S2). Intriguingly, the variations between posterior age estimates are reduced to less than 10 Ma when these modified calibration sets are used (electronic supplementary material, figure S2).

Recent identification of non-oxygenic Cyanobacteria lineages such as Melainabacteria and Sericytochromatia as sister groups of oxygenic Cyanobacteria [16] provides an alternative way to constrain the evolution of oxygenic Cyanobacteria. Specifically, given that oxygenic photosynthesis evolved along the stem lineage of oxygenic Cyanobacteria, we constrained the minimum age of the total Cyanobacteria group at 3.0 Ga, which is supported by geochemical evidence as the time when atmospheric oxygen became available [17]. To avoid overly precise and potentially misleading age estimates, we calibrated the upper limit of the Cyanobacteria root using the ages when the planet Earth formed and became habitable (C15–C38; electronic supplementary material, figure S2). Using this strategy, we show that the age of *Prochlorococcus* major genome reduction remains stable when non-oxygenic Cyanobacteria outgroups were included (electronic supplementary material, figure S2). Since including the non-oxygenic Cyanobacteria has consistently reduced the precision of posterior age estimates, manifested as the higher slopes of the regression line between highest posterior density (HPD) width and the posterior age estimates compared to those without including these lineages (C15–C38 versus C1–C14; electronic supplementary material, figure S3), we focus on the crown oxygenic Cyanobacteria group dating (C9–C14) in the following discussions.

By comparing the width of the 95% HPD derived from each molecular clock analysis (electronic supplementary material, figure S3), we inferred the most precise timeline of *Prochlorococcus* evolution (electronic supplementary material, figure S4). A 682 Ma (95% HPD 732–632 Ma) date for the emergence of the LCA of *Prochlorococcus* HL, LLI and LLII/III clades places the large genome reduction that took place in this lineage firmly within the Cryogenian period (approx. 720–635 Ma; figure 1*a*) and implicates the Snowball Earth icehouse climate conditions eponymous with the period in the corresponding *Prochlorococcus* population bottleneck. We, therefore, refer to this ancestor as SBE-LCA, short for 'Snowball Earth' LCA (figure 1*a*). The Neoproterozoic climate catastrophe culminated in the Sturtian (approx. 717–659 Ma) and Marinoan (approx. 645–635 Ma) glaciations (figure 1*a*), which stretched the ice cover from the poles to sea level near the equators, possibly wrapping the entire Earth under a frozen skin [18]. This 'Snowball Earth' persisted with the freezing temperature of seawater below the ice sheet lowered to −3.5°C [19]. Since all assayed *Prochlorococcus* strains, including those affiliated with the basal LL ecotypes, reach maximum growth rates at approximately 25°C and rarely survive when the temperature drops to approximately 10°C (figure 1*c*) [11], we propose that extreme climate cooling during the Neoproterozoic Snowball Earth events was probably the major driver of severe bottlenecks in early *Prochlorococcus* populations.

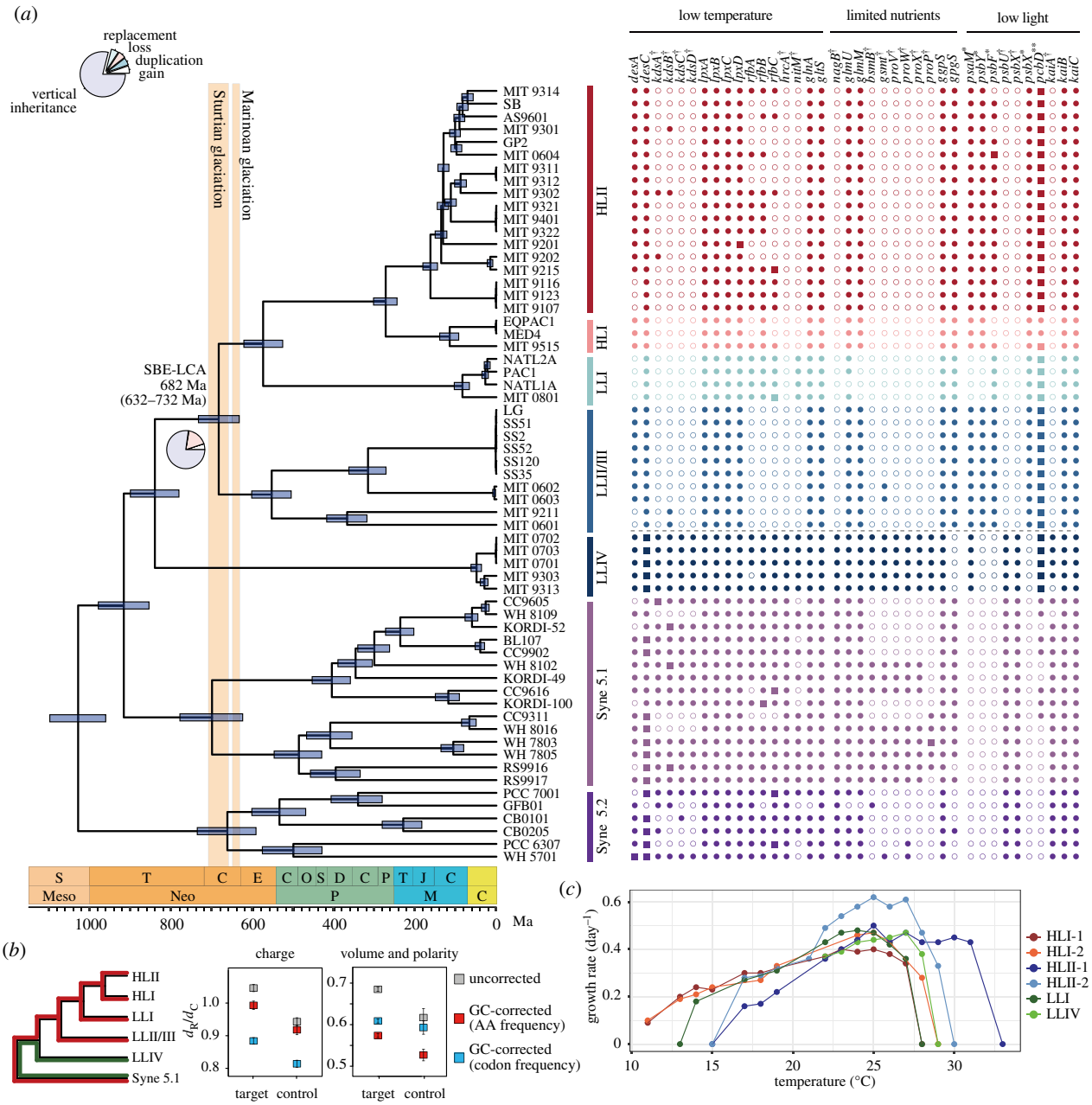
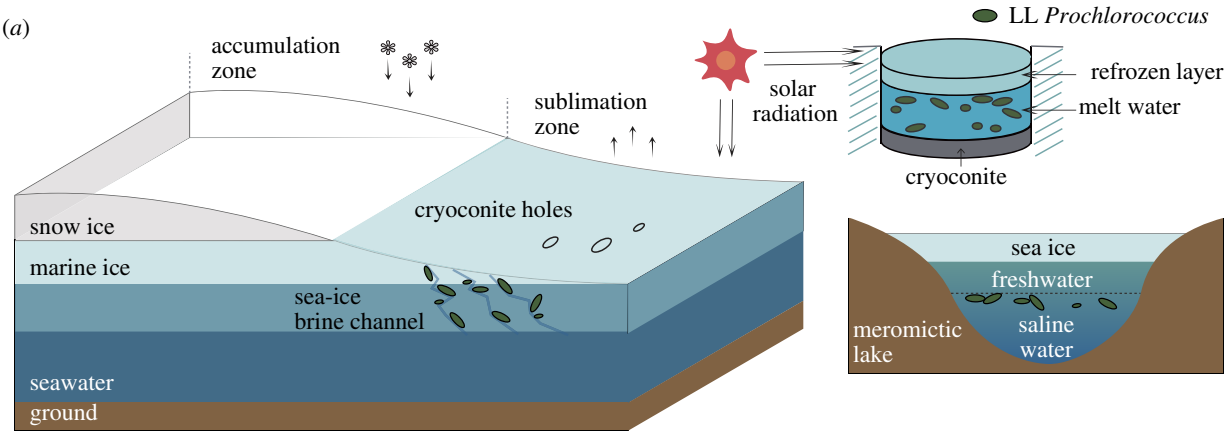


Figure 1. (*a*) (left) Chronogram of the evolutionary history of *Prochlorococcus* estimated by MCMCTree. The vertical bars represent the estimated time of the Neoproterozoic glaciation events. The flanking horizontal bars on ancestral nodes represent the posterior 95% highest probability density (HPD) interval of the estimated divergence time. The pie chart on the ancestral branches leading to the node SBE-LCA provides the proportion of reconstructed genomic events including gene gain, gene loss, gene replacement, gene duplication and gene vertical inheritance. (right) Phyletic pattern of key gene families that potentially enabled *Prochlorococcus* to survive harsh conditions during the Neoproterozoic Snowball Earth (at the ancestral node 'SBE-LCA'). Solid square, solid circle and open circle next to each extant taxon represent multi-copy gene family, single-copy gene family and absence of the gene family, respectively, in the genome. The genomic events that occurred on the branch leading to SBE-LCA including gene family loss, gene family gain and gene family expansion are marked with dagger, asterisk and double asterisks, respectively. (*b*) (left) The diagram helps understand how the d_R/d_C was calculated. In this context, the 'target' group includes all genomes of all HL clades, LLI and LLII/III, the 'control' group includes all genomes of LLIV and the 'reference' group includes all genomes of Syne 5.1. The d_R/d_C for the 'target' group (shown in middle and right) is calculated by comparing a genome from the 'target' group to a genome from the 'reference' group, followed by averaging the value across all possible genome pairs. Likewise, the d_R/d_C for the 'control' group (shown in middle and right) is calculated by comparing a genome from the 'control' group to a genome from the 'reference' group and then by averaging the value across all possible genome pairs. (middle and right) The genome-wide means of d_R/d_C values at the ancestral branch leading to SBE-LCA and that at its sister lineage. They were classified based on the physico-chemical classification of the amino acids by charge or by volume and polarity, and were either GC-corrected by codon frequency, GC-corrected by amino acid (AA) frequency or uncorrected. Error bars of d_R/d_C values represent the standard error of the mean. (*c*) The average growth rate of *Prochlorococcus* ecotypes at different temperatures. Replicate cell cultures were grown in a 14 : 10 light : dark cycle at 66 ± 1 µmol m^{-2} s^{-1}. The growth data used for plotting are collected from Johnson *et al*. [2] and Zinser *et al*. [11]. (Online version in colour)

(b) Biotic refugia, environmental stresses and *Prochlorococcus* adaptation during Snowball Earth

Survival of *Prochlorococcus* populations through the Cryogenian would have required refugia, the nature of which would have shaped continued *Prochlorococcus* evolution. A variety of biotic refugia have been identified during Snowball Earth intervals, including the sea-ice brine channels, the supraglacial cryoconite holes, the ice-covered meromictic lakes and the dry-valley hydrothermal systems [20–22] (figure 2). In analogy

(*a*)



(*b*)

refugia	temperature	illumination	nutrient supply
sea-ice brine channels	cold	weak	rich
cryoconite holes/ponds	cold	strong	poor
ice-covered meromictic deep lakes	warm	weak	rich
dry-valley hydrothermal systems	variable	variable	rich?

Figure 2. (*a*) Diagram of biotic refugia including sea-ice brine channels, cryoconite holes and ice-covered meromictic deep lakes. Dry-valley hydrothermal systems were not shown due to their temporal existence and uncertain physico-chemical conditions. (*b*) Summary of temperature, illumination and nutrient supply in biotic refugia during the Snowball Earth. (Online version in colour)

to modern ecological niches, these biotic refugia were proposed to have different properties. Sea-ice brine channels in deep cracks were rich in nutrients due to the near-zero organic uptake in the dark snowball ocean, but cold and weakly illuminated due to the high albedo of snow cover [20]. Cryoconite-based ecosystems widespread on sea glacier sublimation zones were oligotrophic owing to the limited mineral nutrients from dusts and were cold and highly illuminated [21]. Meromictic lakes during the Snowball Earth were ice-capped and stratified, and the deep layer with saline water was warm, nutrient-rich and weakly illuminated [22,23] (figure 2). Hydrothermal systems were short-lived on snowball time scales with variable temperature and illumination but uncertain nutrient conditions [24]. Despite providing the essential space for *Prochlorococcus* survival, these refugia together presented a number of environmental stresses to *Prochlorococcus* populations, including low temperature, dim light and limited nutrients. *Prochlorococcus* therefore probably evolved a number of adaptive mechanisms to cope with these stresses via gene gains and losses, which we assessed by reconstructing the evolutionary imprints that the Snowball Earth climate left in extant *Prochlorococcus* genomes. To minimize the potential bias in the process of reconstruction, we employed two independent approaches (i.e. a gene tree versus species tree reconciliation approach and a birth-and-death stochastic model-based approach that relies on the estimation of gene turnover rates) and discussed the genomic events that were consistently inferred.

Among the stresses imposed by biotic refugia, the most prominent was probably lethally low temperature. Maintaining membrane fluidity is of paramount importance under low-temperature conditions, which is largely achieved by the activities of fatty acid desaturase encoded by *desA* and *desC* [25]. Consistent with this hypothesis, we inferred that these genes were retained in SBE-LCA (figure 1*a*). Lipopolysaccharide in the outer membrane is known to provide the first line of defense against harsh environments [26], which contains the O-specific polysaccharide, the glycolipid anchor lipid A and the polysaccharide core region. Based on our analyses, genes encoding the polysaccharide core region (*kdsABCD* for 3-deoxy-D-manno-octulosonate biosynthesis; figure 1*a*) were likely lost at SBE-LCA, while those encoding the other components were retained (*lpxABCD* and *rfbABC* for Lipid A precursor and O-specific lipopolysaccharide precursor biosynthesis; figure 1*a*). This inference is consistent with a previous conclusion that the loss of the lipopolysaccharide core region would increase the hydrophobicity and permeability of the cell envelope [27] to protect against cold conditions [28]. Another metabolic modification in SBE-LCA was related to heat shock proteins (HSPs), which play crucial roles in tolerating environmental stresses including thermal shocks. Typically, HSPs are tightly regulated, as they respond quickly to stress and turn off rapidly once the stress disappears [29]. However, the HSP repressor protein encoded by *hrcA* was inferred to be lost at SBE-LCA, which likely allowed the organism to continuously express HSPs to cope with prolonged lethally low temperature. In fact, constitutive expression of HSPs occurs in polar organisms, such as the Antarctic ciliate *Euplotes focardii* [30]. Extremely low temperature also made substrate acquisition difficult due to increased lipid stiffness and decreased efficiency and affinity of membrane transporters [31]. Under such conditions, bacteria may increasingly rely on substrates whose uptake shows lower dependence on temperature. In sea-ice brines where less CO_2 is dissolved [32], elevated pH promotes the conversion of ammonium to ammonia, which diffuses directly into cells without the aid of transporters in the membrane. Accordingly, species of bacteria and microalgae show a greater dependence on ammonium and ammonia at low temperatures and high pH than nitrate [33], thereby reducing reliance on membrane transporters. In SBE-LCA, the potentially efficient utilization of ammonia made other N acquisition genes dispensable, leading to the neutral loss of nitrite transporter (*nitM*), whereas glutamine synthetase (*glnA*) and glutamate synthase (*gltS*) responsible for

the utilization of ammonia after its assimilation were conserved (figure 1*a*).

An additional stress to *Prochlorococcus* during Snowball Earth was probably the oligotrophic condition presented by bacteria refugia. Accordingly, SBE-LCA of *Prochlorococcus* evolved a few metabolic strategies for their survival. The amino sugar *N*-acetylglucosamine (GlcNAc) is used by bacteria such as *Corynebacterium glutamicum* as a carbon, energy and nitrogen source [34]. GlcNAc enters bacteria in the form of GlcNAc-6-phosphate (GlcNAc-6-P). However, instead of being metabolized, the loss of *nagB* for GlcNAc-6-P deamination at SBE-LCA suggests that GlcN6P is more likely to be involved in peptidoglycan recycling through the cascade catalysis by GlmM and GlmU (figure 1*a*) to generate UDP-GlcNAc, which is an essential precursor of both lipopolysaccharide and peptidoglycan in cell wall [35]. During cell turnovers, peptidoglycan is continuously broken down and reused through the peptidoglycan recycling pathway, and in some bacteria, peptidoglycan recycling is critical for their long-term survival when growth is stalled under nutrient limitation [36]. Thus, such a recycling mechanism seems to be key for the maintenance of cell integrity in SBE-LCA under oligotrophic conditions [21]. Glycine betaine is known to be a ubiquitous protein-stabilizing osmolyte in bacteria, in particular cyanobacteria [37]. However, genes involved in glycine betaine biosynthesis and transport were lost at SBE-LCA, including *bsmB* for dimethylglycine *N*-methyltransferase, *gsmt* for glycine/sarcosine *N*-methyltransferase, and *proVWXP* for glycine betaine/proline transport system (figure 1*a*). Instead, several other organic osmolytes might have been used during the Snowball Earth, as their biosynthetic genes were retained at SBE-LCA. The first examples are the *ggpS* gene encoding glucosylglycerol phosphate synthase for glucosylglycerol synthesis and the *gpgS* encoding glucosyl-phosphoglycerate synthase for glucosylglycerate synthesis (figure 1*a*). Since the biosynthesis of glucosylglycerol and glucosylglycerate requires less N compared to that of glycine betaine [38], the potential use of glucosylglycerol/glucosylglycerate instead of glycine betaine appeared favourable to SBE-LCA.

LL intensity presented by biotic refugia during Snowball Earth was another formidable challenge to phototrophs including *Prochlorococcus*. Moreover, *Prochlorococcus* may also need to compete with contemporary eukaryotic algae for the limited amount of light, because the latter became increasingly abundant during the Cryogenian glaciations and likely shared the same refugia with bacteria, like cryoconite holes [39]. Consequently, photosynthetic organisms trapped in bacterial refugia or inhabiting waters below ice need to be physiologically geared to cope with LL. It was proposed that modification of the photosystem structure enables adaptation to the LL condition [40]. We inferred a few changes in photosystem I and II (PSI/PSII) that occurred in SBE-LCA, including the gain of RC1 subunit PsaM, RC2 subunit PsbY, and an extra copy of the RC2 subunit PsbF, the loss of RC2 protein PsbU and the replacement of RC2 subunit PsbX (figure 1*a*), though the molecular mechanism of these changes underlying LL adaptation is poorly understood. We also inferred an expansion of the *Prochlorococcus* antenna gene (*pcbD*) from two to six copies during the Snowball Earth (figure 1*a*), which may boost the light-harvesting capacity under LL conditions [41].

Many cyanobacteria have a sophisticated circadian clock, which is essential in controlling the global diel transcriptional activities of the cells. This circadian oscillator system requires only three components: KaiA, KaiB and KaiC [42]. While all marine *Synechococcus* possess the three *kai* genes, most *Prochlorococcus* lack *kaiA* and, as a consequence, their circadian clocks rather behave like an 'hourglass' which is reset every morning [43]. Our analysis indicated that *kaiA* was lost at SBE-LCA (figure 1*a*). This is likely to be due to the prolonged darkness or LL conditions during the Snowball Earth, rendering the sophisticated circadian clock dispensable.

Among the 38 gene families discussed above, 5 were acquired or expanded at SBE-LCA and 13 existed at earlier nodes (electronic supplementary material, figure S5). Of the latter, six were lost at SBE-LCA (electronic supplementary material, figure S5) and the remaining seven were conserved at SBE-LCA. While only the gene families that were acquired/expanded and lost at SBE-LCA were likely driven by Snowball Earth events, those whose copy number remained constant may also have important roles in assisting *Prochlorococcus* to survive the Snowball Earth catastrophe. For example, although the genes encoding fatty acid desaturase (*desA* and *desC*) were obtained before the LCA of all extant *Prochlorococcus*, they were likely essential for SBE-LCA to maintain its membrane fluidity under low temperature [25]. Even in post-glacial oceans in which the seawater temperature rose to approximately 40°C [44], these genes could have remained important. Experimental analysis in *Synechococcus* revealed that both *desA* and *desC* have detectable transcripts when temperature reaches 34°C [45]. Other genes may have remained important because functionally replaceable genes got lost during Snowball Earth. For example, although glucosylglycerol and glucosylglycerate biosynthesis genes (*ggpS* and *gpgS*) were acquired by stem *Prochlorococcus* in pre-snowball oceans, they may have remained crucial during the Snowball Earth as glucosylglycerol and glucosylglycerate likely acted to balance the osmotic pressure, especially considering that the glycine betaine biosynthesis (*bsmB* and *gsmt*) and transport genes (*proVWXP*) were lost during the Snowball Earth. Likewise, although the genes *kaiB* and *kaiC* existed before the LCA of all extant *Prochlorococcus*, their functions were shown to be essential to dampened circadian oscillation in the absence of the gene *kaiA*, as evidenced by the experimental study of *Synechococcus elongatus* [46].

We note that adaptation of SBE-LCA to oligotrophic condition was achieved by gene loss rather than gene gain (electronic supplementary material, figure S5). In fact, most adaptive events during the Snowball Earth occurred by gene loss. The few exceptions (i.e. through gene gain) were all associated with adaptation to LL intensity, among which the gene *pcbD* duplicated from two to six copies during the Snowball Earth (electronic supplementary material, figure S5). *PcbD* encodes the antenna of photosystem and thus its duplication boosted the light-harvesting capacity under LL conditions. By contrast to oligotrophy, which had already existed before the Snowball Earth, illumination reaching to seawater was dramatically reduced owing to the high sea-ice albedo during Snowball Earth compared to pre-snowball. In post-glacial time when the illumination was recovered, the extra copies of the *pcbD* genes probably became dispensable, especially in *Prochlorococcus* lineages that inhabit the upper layer of the euphotic zone. Accordingly, we found that the duplicated copies of *pcbD* got lost in the ancestor of HL-adapted *Prochlorococcus* (electronic supplementary material, figure S5). This example well explains the evolutionary trajectory of an

adaptive gene in response to different climate and ocean conditions. If the underlying principle is generalizable to the evolution of other gene families, we may infer that oligotrophy may not be a key selective pressure during the Snowball Earth. Otherwise, we might see a signature of gene acquisitions for nutrient scavenging, which did not happen. We therefore conclude that lethally low temperature and exceedingly dim light were likely the two dominant stresses during Snowball Earth.

(c) Impacts of Snowball Earth events on modern *Prochlorococcus* physiology and geographical distribution

We argue that the metabolic adaptation in response to the harsh conditions during Snowball Earth and the resulting genome reduction were essential to shape the physiological characters and the biogeographic distribution of their descendants in the modern ocean. For example, the genome reduction that occurred in the early evolution of *Prochlorococcus* probably resulted in the reduced cell size and increased surface-to-volume ratio in their descendants, which may have enhanced their efficiency in nutrient acquisition [47] and eventually led them to dominate the photosynthetic communities in the most oligotrophic regions of today's oceans [2]. Physical structure of microfossils, if recognizable, is helpful to determine the cell size of microbial species. However, the only physical microfossil described from Cryogenian glaciomarine deposits is related to *Bavlinella faveolata* [48], whose contemporary analogues, known as *Xenococcus* and *Dermocarpa* [48], are phylogenetically distant to *Prochlorococcus* and thus cannot be used to infer the cell size of *Prochlorococcus* during that time. Likewise, new metabolic strategies that *Prochlorococcus* evolved to overcome the nutrient stresses during Snowball Earth, such as the recycling of cell wall components and the use of glucosylglycerol and glucosylglycerate instead of nitrogen-rich glycine betaine as the organic osmolytes, decreased the nutrient requirements of the descendants' cells and thus contributed to their success in the modern oligotrophic nitrogen-limited oceans.

Aside from the potentially beneficial outcomes, modifications of some important metabolic pathways may have imposed deleterious effects on *Prochlorococcus* descendants. For example, whereas the replacement of circadian clock with an hourglass-like mechanism might have facilitated the ancestral lineage to adapt to the prolonged dim light condition during the Snowball Earth catastrophe, it probably prevents the dispersal of *Prochlorococcus* to high latitude regions in the modern ocean, where the day length varies substantially across seasons. Normally, organisms with circadian rhythms deal with these changes by anticipating the changes of light intensity and promptly regulating cellular processes such as DNA transcription and recombination via chromosome compaction, a known mechanism to protect DNA from UV radiation [49]. In the absence of the circadian clock, however, species such as *Prochlorococcus* cannot synchronize the endogenous oscillation with the environmental cycles and thus are at high risks of cell damages [50].

3. Caveats and concluding remarks

The relaxed clock model implemented in the present study takes into account the rate variation among species and allows us to estimate a reliable timeline of *Prochlorococcus* evolution. On this basis, we link an ancestral phylogenetic branch that supported population bottlenecks and genome reduction of *Prochlorococcus* to the Neoproterozoic Snowball Earth. Using ancestral gene gain and loss analysis, we further identify potentially important metabolic strategies that the Cryogenian *Prochlorococcus* evolved to survive the glacial catastrophe. Despite these fascinating results, there are important caveats. We postulated that the icehouse conditions during the Cryogenian were lethal to the ancestral *Prochlorococcus* and thus induced population bottlenecks. This key assumption derives from our knowledge on the modern *Prochlorococcus* populations which do not grow below approximately 10°C (figure 1*c*), but this physiological character is not necessarily transmissible to the ancestral population. We also postulated that the metabolic traits we discussed earlier, which allowed *Prochlorococcus* to survive the lethally low temperature and exceedingly dim light, each must have conferred a strong fitness advantage to the Cryogenian *Prochlorococcus* population. The rationale is that for any advantageous trait that can be visible to and thus promoted by positive selection, the benefit it confers to the organism has to be sufficiently large to overcome the power of genetic drift [51]. While the metabolic traits we discussed fit the geochemical context well, they need additional evidence to support the hypothesis that they were subjected to positive selection and facilitated *Prochlorococcus* adaptation in those harsh conditions. From the perspective of the dating methodology, the uncertainty of our analysis largely comes from the use of calibrations. Molecular clock analysis requires at least one maximum age constraint [52]. However, available cyanobacterial fossils can only serve as the minimum bounds [15], and therefore our analysis has to rely on the maximum bound provided by the root. We took a conservative approach by successively increasing the maximum bound of the root from 3800 Ma to 4500 Ma, which showed that the posterior age of the ancestral node SBE-LCA increased only slightly by approximately 7% (electronic supplementary material, figure S2), and thus strengthened our conclusion of the coincidence between SBE-LCA and Cryogenian Snowball Earth.

The Neoproterozoic Snowball Earth hypothesis posits the extension of ice sheet to the sea level near the equator [44]. Under this circumstance, primary producers had to inhabit the biotic refugia instead of the ice-free marine areas [53], which likely resulted in the prolonged collapse in biological productivity in the surface ocean [44]. However, *Prochlorococcus* and other photoautotrophic lineages including *Synechococcus*, green and red algae, which evolved before the Neoproterozoic glaciations, must have survived the catastrophe [21]. Their survivorship was likely crucial in sustaining primary production, heterotrophy and carbon cycling, as well as broader ecosystem functioning during the Snowball Earth glaciations. It is worthy of note that the population bottlenecks that occurred in *Prochlorococcus* may not necessarily have occurred in other cooccurring photoautotrophic lineages. Take *Synechococcus*, which is evolutionarily most closely related to *Prochlorococcus*, as an example. Modern *Synechococcus* have wider geographical distributions than *Prochlorococcus*, and they even seasonally dominate the phototrophic community in polar oceans, where *Prochlorococcus* are absent [54]. Moreover, *Synechococcus* harbour more diverse pigments than *Prochlorococcus*, which allow them to live in a wider range of light conditions [55].

These unique traits may increase the survivorship of *Synechococcus* during Snowball Earth and thus reduce the chance to detect population bottlenecks, if any, in that difficult time.

By contrast to the hypothesis of productivity collapse, a few studies have proposed that microbial communities might have been only mildly affected by the Snowball Earth climate catastrophe [56]. However, these inferences can be problematic. Since glaciers or ice sheets entering the ocean from tropical continental shelves would have inevitably eroded the organic matter from early deposited sediments [57], the microfossil and biomarker identified in glaciomarine sediments may not represent the microbial community during the Snowball Earth. Even without considering the effect of glacial erosion, the lack of lineage-specific microfossil and biomarker impedes the reconstruction of the evolutionary history of many ecologically important lineages, such as the *Prochlorococcus* studied here. Instead, in the present study, we find that substantial disruptions to the Earth system, like the Neoproterozoic Snowball Earth, leave indelible signatures in microbial genomes, such that these heritable changes allow us to reconstruct interactions between environmental change and biological evolution deep in Earth's history. By employing the accelerated genome-wide accumulation of the deleterious mutations as a proxy for a rapid decrease in the population size of ancient lineages, we uncovered severe bottlenecks that shaped the early evolution of *Prochlorococcus* lineages. The careful molecular clock analyses as well as the ancestral genome reconstruction enabled us to link dynamics in ancestral population sizes to changes in metabolic potential and adaptation to icehouse climates through natural selection. Collectively, our findings demonstrate how paleomicrontological approaches can be used to connect large-scale dynamics in the Earth System to the genomic imprints left on extant microorganisms, which shape their ecological role and biogeographic distribution in the world today.

4. Material and methods

Genomic sequences of Cyanobacteria were downloaded from public databases and manually annotated (see 'Dataset_1.tbl' in online GitHub repository and §1 in electronic supplementary material, Methods). Divergence time of *Prochlorococcus* was estimated with MCMCTree v. 4.9e [58] on top of 27 genes (see 'Dataset_2.tbl' in online GitHub repository) previously proposed to be valuable to date bacterial divergence [59] and Cyanobacteria phylogenomic trees. In previous studies, the LPP (*Leptolyngbya*, *Plectonema* and *Phormidium*) group of Cyanobacteria located either at the basal of the Microcyanobacteria group [13] or at the basal of the Macrocyanobacteria group [60]. Our analysis showed that this controversy is probably caused by the inclusion of composition-heterogeneous proteins and that using composition-homogeneous proteins led to consistent support for the former hypothesis (electronic supplementary material, figure S6; see 'Dataset_3.tbl' in online GitHub repository and §2.b in electronic supplementary material, Methods). Since molecular dating analysis is known to be intrinsically associated with calibration points [61], we summarized the calibrations of Cyanobacteria used in previous studies and modified them for our analyses with caution. Moreover, we proposed a new strategy to use calibrations when non-oxygenic Cyanobacteria were used as outgroups (electronic supplementary material, figure S2; see electronic supplementary material, §2.c for justification). We further assessed the fitness of different molecular clock models implemented in MCMCTree by using the package 'mcmc3r' v. 0.3.2, based on which we decided to use the independent rates model for further molecular clock analyses. For each molecular clock analysis, the software ran twice with a burn-in of 50 000 and a total of 500 000 generations. The convergence was assessed based on the correlations of posterior mean time of all ancestral nodes between independent runs (electronic supplementary material, figure S7). By implementing statistical tests based on the 'infinite-site' theory (electronic supplementary material, figure S3 and §2.f). We were able to select the most precise estimates of *Prochlorococcus* evolutionary timeline for illustration (electronic supplementary material, figure S4) and further discussion.

Evolution of genome content via gene gains and losses was inferred using two independent methods, AnGST [62] and BadiRate v. 1.35 [63]. The former assumes that the statistically supported topological differences between a gene tree and the species tree result from evolutionary events (gene loss, gene duplication, HGT, gene birth and speciation), and infers these evolutionary events by reconciling the topological incongruences under a generalized parsimony framework by achieving a minimum number of the evolutionary events along the species tree, with penalties of an evolutionary event determined by the genome flux analysis [62]. The latter does not rely on the tree topological incongruence information, but instead uses a full maximum-likelihood approach to determine the gene family turnover rates that maximize the probability of observing the gene count patterns provided by the family size table. The BadiRate analyses were run using nine strategies each with a distinct turnover rate model and a distinct branch model. The likelihoods of different runs were compared, and three strategies with the highest likelihood values were used (electronic supplementary material, figure S8A). Further, results derived from AnGST (electronic supplementary material, figure S8B) and BadiRate were compared and summarized to determine the common patterns shared by the two software (electronic supplementary material, figure S8C), and important functional genes discussed were consistently inferred by these two methods. As the two methods inferred the qualitatively same pattern of genome size reduction on the branches leading to SBE-LCA, the number of gene gains and losses derived from the AnGST analysis was presented.

The inference of a potential change of selection efficiency on a given branch was performed by comparing the genome-wide d_R/d_C value across single-copy orthologous genes of the branch to that of the closest sister branch. The d_R/d_C value was calculated using RCCalculator (http://www.geneorder.org/RCCalculator/; see electronic supplementary material, §4) based on two independent amino acid classification schemes (electronic supplementary material, table S1).

Data accessibility. The custom scripts as well as the sequence datasets are available in the online GitHub repository (https://github.com/luolab-cuhk/Prochl-SBE). The data are provided in the electronic supplementary material [64].

Authors' contributions. H.Z.: Data curation, formal analysis, writing—original draft, writing—review and editing; Y.S.: data curation, formal analysis, writing—original draft; Q.Z.: supervision, writing—original draft; S.A.C.: supervision, writing—review and editing; H.L.: conceptualization, data curation, supervision, writing—review and editing. All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Competing interests. The authors declare no competing interests.

Funding. This work is supported by the National Natural Science Foundation of China (grant no. 92051113), the Hong Kong Research Grants Council General Research Fund (grant no. 14110820), the Hong Kong Research Grants Council Area of Excellence Scheme (grant no. AoE/M-403/16), HKU FoS funds to S.A.C., and the Direct Grant of CUHK (grant nos. 4053257 and 3132809).

Acknowledgements. We thank the two reviewers for providing constructive suggestions that significantly improved this manuscript. Part of

the narratives regarding the types of refugia and their physio-chemical conditions are from the comments provided by one of the reviewers, Prof. Paul Hoffman. We thank Allison Coe, Erik Zinser and Zackary Johnson for providing the data of *Prochlorococcus* growth rates, Sishuo Wang, Tianhua Liao and Xiaoyuan Feng for their suggestions on molecular dating analyses.

References

1. Biller SJ, Berube PM, Lindell D, Chisholm SW. 2015 *Prochlorococcus*: the structure and function of collective diversity. *Nat. Rev. Microbiol.* **13**, 13. (doi:10.1038/nrmicro3378)
2. Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EMS, Chisholm SW. 2006 Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311**, 1737–1740. (doi:10.1126/science.1118052)
3. West NJ, Scanlan DJ. 1999 Niche-partitioning of *Prochlorococcus* populations in a stratified water column in the eastern North Atlantic Ocean. *Appl. Environ. Microbiol.* **65**, 2585–2591. (doi:10.1128/AEM.65.6.2585-2591.1999)
4. Hess WR, Rocap G, Ting CS, Larimer F, Stilwagen S, Lamerdin J, Chisholm SW. 2001 The photosynthetic apparatus of *Prochlorococcus*: insights through comparative genomics. *Photosynth. Res.* **70**, 53–71. (doi:10.1023/A:1013835924610)
5. Moore LR, Rocap G, Chisholm SW. 1998 Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393**, 464. (doi:10.1038/30965)
6. Luo H, Friedman R, Tang J, Hughes AL. 2011 Genome reduction by deletion of paralogs in the marine cyanobacterium *Prochlorococcus*. *Mol. Biol. Evol.* **28**, 2751–2760. (doi:10.1093/molbev/msr081)
7. Batut B, Knibbe C, Marais G, Daubin V. 2014 Reductive genome evolution at both ends of the bacterial population size spectrum. *Nat. Rev. Microbiol.* **12**, 841. (doi:10.1038/nrmicro3331)
8. Luo H, Huang Y, Stepanauskas R, Tang J. 2017 Excess of non-conservative amino acid changes in marine bacterioplankton lineages with reduced genomes. *Nat. Microbiol.* **2**, 17091. (doi:10.1038/nmicrobiol.2017.91)
9. Zuckerkandl E, Pauling L. 1965 Evolutionary divergence and convergence in proteins. In *Evolving genes and proteins* (eds V Bryson, HJ Vogel), pp. 97–166. New York, NY: Academic Press.
10. Dayhoff MO. 1972 A model of evolutionary change in proteins. *Atlas Protein Seq. Struct.* **5**, 89–99.
11. Zinser ER, Johnson ZI, Coe A, Karaca E, Veneziano D, Chisholm SW. 2007 Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. *Limnol. Oceanogr.* **52**, 2205–2220. (doi:10.4319/lo.2007.52.5.2205)
12. Kettler GC *et al.* 2007 Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet.* **3**, e231. (doi:10.1371/journal.pgen.0030231)
13. Sánchez-Baracaldo P. 2015 Origin of marine planktonic cyanobacteria. *Sci. Rep.* **5**, 17418. (doi:10.1038/srep17418)
14. Sánchez-Baracaldo P, Ridgwell A, Raven JA. 2014 A neoproterozoic transition in the marine nitrogen cycle. *Curr. Biol.* **24**, 652–657. (doi:10.1016/j.cub.2014.01.041)
15. Marshall CR. 2019 Using the fossil record to evaluate timetree timescales. *Front. Genet.* **10**, 1049. (doi:10.3389/fgene.2019.01049)
16. Soo RM, Hemp J, Parks DH, Fischer WW, Hugenholtz P. 2017 On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. *Science* **355**, 1436–1440. (doi:10.1126/science.aal3794)
17. Crowe SA, Døssing LN, Beukes NJ, Bau M, Kruger SJ, Frei R, Canfield DE. 2013 Atmospheric oxygenation three billion years ago. *Nature* **501**, 535. (doi:10.1038/nature12426)
18. Hoffman PF, Kaufman AJ, Halverson GP, Schrag DP. 1998 A Neoproterozoic snowball earth. *Science* **281**, 1342–1346. (doi:10.1126/science.281.5381.1342)
19. Ashkenazy Y, Gildor H, Losch M, Macdonald FA, Schrag DP, Tziperman E. 2013 Dynamics of a Snowball Earth ocean. *Nature* **495**, 90. (doi:10.1038/nature11894)
20. Thomas D, Dieckmann G. 2002 Antarctic sea ice-a habitat for extremophiles. *Science* **295**, 641–644. (doi:10.1126/science.1063391)
21. Hoffman PF *et al.* 2017 Snowball Earth climate dynamics and Cryogenian geology-geobiology. *Sci. Adv.* **3**, e1600983. (doi:10.1126/sciadv.1600983)
22. Vincent WF, Vincent CL. 1982 Factors controlling phytoplankton production in Lake Vanda (77 S). *Can. J. Fish. Aquat. Sci.* **39**, 1602–1609. (doi:10.1139/f82-216)
23. Parker BC, Simmons Jr GM, Seaburg KG, Cathey DD, Allnutt F. 1982 Comparative ecology of plankton communities in seven Antarctic oasis lakes. *J. Plankton Res.* **4**, 271–286. (doi:10.1093/plankt/4.2.271)
24. Zakharov D, Bindeman I, Slabunov A, Ovtcharova M, Coble M, Serebryakov N, Schaltegger U. 2017 Dating the Paleoproterozoic Snowball Earth glaciations using contemporaneous subglacial hydrothermal systems. *Geology* **45**, 667–670. (doi:10.1130/G38759.1)
25. Los DA, Murata N. 2004 Membrane fluidity and its roles in the perception of environmental signals. *Biochim. Biophys. Acta Biomembr.* **1666**, 142–157. (doi:10.1016/j.bbamem.2004.08.002)
26. Benforte FC, Colonnella MA, Ricardi MM, Venero EC.S., Lizarraga L, López NI, Tribelli PM. 2018 Novel role of the LPS core glycosyltransferase WapH for cold adaptation in the Antarctic bacterium *Pseudomonas extremaustralis*. *PLoS ONE* **13**, e0192559. (doi:10.1371/journal.pone.0192559)
27. Wang Z, Wang J, Ren G, Li Y, Wang X. 2015 Influence of core oligosaccharide of lipopolysaccharide to outer membrane behavior of *Escherichia coli*. *Mar. Drugs* **13**, 3325–3339. (doi:10.3390/md13063325)
28. Feller G, Gerday C. 2003 Psychrophilic enzymes: hot topics in cold adaptation. *Nat. Rev. Microbiol.* **1**, 200. (doi:10.1038/nrmicro773)
29. Schumann W. 2016 Regulation of bacterial heat shock stimulons. *Cell Stress Chaperones* **21**, 959–968. (doi:10.1007/s12192-016-0727-z)
30. La Terza A, Papa G, Miceli C, Luporini P. 2001 Divergence between two Antarctic species of the ciliate *Euplotes E. focardii* and *E. nobilii*, in the expression of heat-shock protein 70 genes. *Mol. Ecol.* **10**, 1061–1067. (doi:10.1046/j.1365-294X.2001.01242.x)
31. Lawrence RP, William JW. 2001 Temperature and substrates as interactive limiting factors for marine heterotrophic bacteria. *Aquat. Microb. Ecol.* **23**, 187–204. (doi:10.3354/ame023187)
32. Gleitz M, Loeff MR, Thomas DN, Dieckmann GS, Millero FJ. 1995 Comparison of summer and winter inorganic carbon, oxygen and nutrient concentrations in Antarctic sea ice brine. *Mar. Chem.* **51**, 81–91. (doi:10.1016/0304-4203(95)00053-T)
33. Reay DS, Nedwell DB, Priddle J, Ellis-Evans JC. 1999 Temperature dependence of inorganic nitrogen uptake: reduced affinity for nitrate at suboptimal temperatures in both algae and bacteria. *Appl. Environ. Microbiol.* **65**, 2577–2584. (doi:10.1128/AEM.65.6.2577-2584.1999)
34. Uhde A, Youn JW, Maeda T, Clermont L, Matano C, Krämer R, Wendisch VF, Seibold GM, Marin K. 2013 Glucosamine as carbon source for amino acid-producing *Corynebacterium glutamicum*. *Appl. Microbiol. Biotechnol.* **97**, 1679–1687. (doi:10.1007/s00253-012-4313-8)
35. Park JT, Uehara T. 2008 How bacteria consume their own exoskeletons (turnover and recycling of cell wall peptidoglycan). *Microbiol. Mol. Biol. Rev.* **72**, 211–227. (doi:10.1128/MMBR.00027-07)
36. Borisova M *et al.* 2016 Peptidoglycan recycling in gram-positive bacteria is crucial for survival in stationary phase. *MBio* **7**, e00923-16. (doi:10.1128/mBio.00923-16)
37. Papageorgiou GC, Murata N. 1995 The unusually strong stabilizing effects of glycine betaine on the structure and function of the oxygen-evolving photosystem II complex. *Photosynth. Res.* **44**, 243–252. (doi:10.1007/BF00048597)
38. Scanlan DJ *et al.* 2009 Ecological genomics of marine picocyanobacteria. *Microbiol. Mol. Biol. Rev.* **73**, 249–299. (doi:10.1128/MMBR.00035-08)

39. Hoffman P. 2016 Cryoconite pans on Snowball Earth: supraglacial oases for Cryogenian eukaryotes? *Geobiology* **14**, 531–542. (doi:10.1111/gbi.12191)
40. Kouřil R, Wientjes E, Bultema JB, Croce R, Boekema EJ. 2013 High-light vs. low-light: effect of light acclimation on photosystem II composition and organization in *Arabidopsis thaliana*. *Biochim. Biophys. Acta Bioenerg.* **1827**, 411–419. (doi:10.1016/j.bbabio.2012.12.003)
41. Bibby T, Mary I, Nield J, Partensky F, Barber J. 2003 Low-light-adapted *Prochlorococcus* species possess specific antennae for each photosystem. *Nature* **424**, 1051. (doi:10.1038/nature01933)
42. Dong G, Golden SS. 2008 How a cyanobacterium tells time. *Curr. Opin Microbiol.* **11**, 541–546. (doi:10.1016/j.mib.2008.10.003)
43. Axmann IM, Dühring U, Seeliger L, Arnold A, Vanselow JT, Kramer A, Wilde A. 2009 Biochemical evidence for a timing mechanism in prochlorococcus. *J. Bacteriol.* **191**, 5342–5347. (doi:10.1128/JB.00419-09)
44. Hoffman PF, Schrag DP. 2002 The Snowball Earth hypothesis: testing the limits of global change. *Terra Nova* **14**, 129–155. (doi:10.1046/j.1365-3121.2002.00408.x)
45. Sakamoto T, Bryant DA. 1997 Temperature-regulated mRNA accumulation and stabilization for fatty acid desaturase genes in the cyanobacterium *Synechococcus* sp. strain PCC 7002. *Mol. Microbiol.* **23**, 1281–1292. (doi:10.1046/j.1365-2958.1997.3071676.x)
46. Kawamoto N, Ito H, Tokuda IT, Iwasaki H. 2020 Damped circadian oscillation in the absence of KaiA in *Synechococcus*. *Nat. Commun.* **11**, 2242. (doi:10.1038/s41467-020-16087-x)
47. Giovannoni SJ, Thrash JC, Temperton B. 2014 Implications of streamlining theory for microbial ecology. *ISME J.* **8**, 1553. (doi:10.1038/ismej.2014.60)
48. Knoll AH, Christie-Blick N, Awramik SM. 1981 Stratigraphic and ecologic implications of late Precambrian microfossils from Utah. *Am. J. Sci.* **281**, 247–263. (doi:10.2475/ajs.281.3.247)
49. Simons MJ. 2009 The evolution of the cyanobacterial posttranslational clock from a primitive 'phoscillator'. *J. Biol. Rhythms* **24**, 175–182. (doi:10.1177/0748730409333953)
50. Mullineaux CW, Stanewsky R. 2009 The rolex and the hourglass: a simplified circadian clock in *Prochlorococcus*? *J. Bacteriol.* **191**, 5333–5335. (doi:10.1128/JB.00719-09)
51. Luo H, Moran MA. 2015 How do divergent ecological strategies emerge among marine bacterioplankton lineages? *Trends Microbiol.* **23**, 577–584. (doi:10.1016/j.tim.2015.05.004)
52. Szollosi GJ, Höhna S, Williams TA, Schrempf D, Daubin V, Boussau B. 2020 Relative time constraints improve molecular dating. *bioRxiv*. (doi:10.1101/2020.10.17.343889)
53. Vincent W, Gibson J, Pienitz R, Villeneuve V, Broady P, Hamilton P, Howard-Williams C. 2000 Ice shelf microbial ecosystems in the high arctic and implications for life on Snowball Earth. *Naturwissenschaften* **87**, 137–141. (doi:10.1007/s001140050692)
54. Vincent WF, Quesada A. 2012 Cyanobacteria in high latitude lakes, rivers and seas. In *Ecology of cyanobacteria II* (ed. BA Whitton), pp. 371–385. Dordrecht, The Netherlands: Springer.
55. Six C, Thomas JC, Garczarek L, Ostrowski M, Dufresne A, Blot N, Scanlan DJ, Partensky F. 2007 Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative genomics study. *Genome Biol.* **8**, R259. (doi:10.1186/gb-2007-8-12-r259)
56. Olcott AN, Sessions AL, Corsetti FA, Kaufman AJ, De Oliviera TF. 2005 Biomarker evidence for photosynthesis during Neoproterozoic glaciation. *Science* **310**, 471–474. (doi:10.1126/science.1115769)
57. Wagner T, Hölemann JA. 1995 Deposition of organic matter in the Norwegian-Greenland sea during the past 2.7 million years. *Quat. Res.* **44**, 355–366. (doi:10.1006/qres.1995.1080)
58. Yang Z. 2007 PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591. (doi:10.1093/molbev/msm088)
59. Battistuzzi FU, Hedges SB. 2008 A major clade of prokaryotes with ancient adaptations to life on land. *Mol. Biol. Evol.* **26**, 335–343. (doi:10.1093/molbev/msn247)
60. Uyeda JC, Harmon LJ, Blank CE. 2016 A comprehensive study of cyanobacterial morphological and ecological evolutionary dynamics through deep geologic time. *PLoS ONE* **11**, e0162539. (doi:10.1371/journal.pone.0162539)
61. Schirrmeister BE, Sanchez-Baracaldo P, Wacey D. 2016 Cyanobacterial evolution during the Precambrian. *Int. J. Astrobiol.* **15**, 187–204. (doi:10.1017/S1473550415000579)
62. David LA, Alm EJ. 2011 Rapid evolutionary innovation during an Archaean genetic expansion. *Nature* **469**, 93. (doi:10.1038/nature09649)
63. Librado P, Vieira F, Rozas J. 2011 BadiRate: estimating family turnover rates by likelihood-based methods. *Bioinformatics* **28**, 279–281. (doi:10.1093/bioinformatics/btr623)
64. Zhang H, Sun Y, Zeng Q, Crowe SA, Luo H. 2021 Snowball Earth, population bottleneck and *Prochlorococcus* evolution. Figshare.