Snowball Earths, population bottlenecks, and the evolution of marine photosynthetic bacteria

Hao Zhang[1^], Ying Sun[1^], Qinglu Zeng[2], Sean A. Crowe[3], Haiwei Luo[1*]

[1]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

[2]Department of Ocean Science, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong SAR

[3]Department of Earth Sciences, School of Biological Sciences, and Swire Institute for Marine Science (SWIMS), University of Hong Kong, Pokfulam Road, Hong Kong SAR

[^]These authors contributed equally to this study.

***Corresponding author:**
Haiwei Luo
The Chinese University of Hong Kong
Shatin, Hong Kong SAR
Phone: (+852) 39436121
E-mail: hluo2006@gmail.com

Running Title: *Prochlorococcus* and Snowball Earths

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

Abstract

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 that accumulated across genomic regions in the 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. This metabolic potential 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 altered the physiological characters of *Prochlorococcus*, shaping their ecologically vital role as the most abundant primary producers in the modern oceans.

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 the 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 streamlining, including a major genome reduction in their early evolution and a few minor modifications that followed[6,7]. Modern marine *Prochlorococcus* lineages, in particular those with small genomes, show very low ratios of nonsynonymous (d_N) to synonymous (d_S) nucleotide substitution rates, suggesting that natural selection is a highly efficient throttle on the accumulation of deleterious mutations (i.e., nonsynonymous mutations) in *Prochlorococcus*[6,8]. On long time scales, however, nucleotide substitutions at synonymous sites become saturated, invalidating the use of d_N/d_S to infer selection efficiency in deep time[9]. Thus, an alternative approach focuses instead on different types of nonsynonymous substitutions leading to radical versus conservative changes in amino acid sequences, with the former more likely to be deleterious[10,11]. Excess radical mutations accumulate from random fixation of deleterious mutations by genetic drift (i.e., reduced efficiency of selection). Using this approach reveals that the major genome

reduction in *Prochlorococcus* took place under reduced selection efficiency[9] 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 precise molecular dating is needed to link this important evolutionary event to its possible environmental drivers. 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 likely have challenged warm-water-loving photosynthetic *Prochlorococcus*, with strong potential to cause the population bottlenecks inferred from the genome sequences described above. *Prochlorococcus* survived this catastrophe through likely gains and losses of key metabolic functions reconstructed from the same genome sequences, which have far-reaching impact on their success in today's oceans. *Prochlorococcus* are thus vital "guardians of metabolism"[12], shepherding genes critical to the functioning of the biosphere across environmental catastrophes, including global glaciations.

Results and Discussion

Prochlorococcus experienced a massive gene loss event on the ancestral branch leading to the last common ancestor (LCA) of clades HL, LLI and LLII/III[6,7,13]. This is confirmed by our analysis, which reconstructed 366 and 107 gene family losses and gains on this branch, respectively (Fig. 1A & S1), among which 163 and 84 encode unknown proteins. On the same ancestral branch, it was shown that d_R/d_C is significantly elevated compared to the sister ancestral branch leading to the LCA of the LLIV[9], which was

validated here (both sign test and paired *t*-test, $p < 0.001$; Fig. 1B). 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[9]. Given the global distribution and abundance of *Prochlorococcus*, and cyanobacteria more generally, such a bottleneck would likely require a global-scale event, like an environmental or climate catastrophe (e.g. meteorite impact, large igneous province emplacement, or glaciation).

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 (under calibrations C1-C6 in Table S1 with related references included there) results in up to ~320 Ma disparity (Fig. 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 2,100 Mya was used as either the maximum bound or the minimum bound to calibrate the crown group of *Nostocales*[14,15]. However, given the fact that apomorphic character must evolve earlier than the divergence of crown group, morphological fossils can only serve as the minimum bounds on total groups of assigned lineages[16] (see Section 2.3 in Supplemental Methods for details). Thus, in the present study, we modified the calibration sets by constraining the lower bounds of the *Nostocales* (and also *Pleurocapsales*) total groups with morphological fossils and by leaving their upper bounds open (C9-C14; Table S1). Intriguingly, the variation is reduced to less than 10 Ma when these modified calibration sets are used (Fig. S2). By comparing the width of the

95% highest posterior density (HPD) derived from each molecular clock analysis (Fig. S3), we inferred the most precise timeline of *Prochlorococcus* evolution (corresponding to calibration set C14 in Table S1; see Section 2.3 in Supplemental Methods).

Our time estimates revealed that the LCA of *Prochlorococcus* HL, LLI and LLII/III clades diversified at 682 Mya [95% HPD (=highest posterior density) 632-732 Mya], precisely dating the genome reduction event to this time. A 682 Mya 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 (~720 to 635 Mya; Fig. 1A) 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 (see Fig. 1A), short for "Snowball Earth" LCA. The Neoproterozoic climate catastrophe culminated in the Sturtian (~717 to 659 Mya) and Marinoan (~645 to 635 Mya) glaciations (Fig. 1A), which stretched from the poles to sea level near the equators, possibly wrapping the entire Earth under a frozen skin[17]. This "Snowball Earth" persisted with the freezing temperature of seawater below the ice sheet lowered to -3.5 ℃[18]. Since all assayed *Prochlorococcus* strains, including those affiliated with the basal LL ecotypes, reach maximum growth rates at approximately 25 ℃ and rarely survive when temperature drops to ~10 ℃ (Fig. 1C)[19], we propose that extreme climate cooling during the Neoproterozoic Snowball Earth events was likely the major driver of severe bottlenecks in early *Prochlorococcus* populations.

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 within ice grounding-line crack systems[20] and cryoconite holes/ponds on the surface of the sublimation zone, which may have represented ~12% of the global sea glacier surface[21] (Fig. 1D). Despite providing the essential space for *Prochlorococcus* survival, these refugia would have presented a number of environmental stresses to *Prochlorococcus* populations including: low temperature, low light, high and variable osmotic pressure, and limited nutrients[20-22]. *Prochlorococcus* thus evolved a number of adaptive mechanisms to cope with these stresses via gene gains and losses, which we assessed by reconstructing the evolutionary paths of imprints that the Snowball Earth climate left in extant *Prochlorococcus* genomes.

Among these stresses, the most prominent was likely 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*. As a result, we inferred that these genes were retained in SBE-LCA (Fig. 1A). Lipopolysaccharide (LPS) in the outer membrane is known to provide the first line of defense against harsh environments[23], 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; Fig. 1A) were likely lost at SBE-LCA, while those encoding the other components were retained (*lpxABCD* and *rfbABC* for Lipid A precursor and O-specific LPS precursor biosynthesis; Fig. 1A). This inference is consistent with a previous conclusion that the loss of the LPS core region would increase the

hydrophobicity and permeability of the cell envelope[24] to protect against the cold conditions[25]. The amino sugar N-acetylglucosamine (GlcNAc) is used by bacteria such as *Corynebacterium glutamicum* as a carbon, energy and nitrogen source[26]. 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 (PG) recycling through the cascade catalysis by GlmM and GlmU (Fig. 1A) to generate UDP-GlcNAc, which is an essential precursor of cell wall PG and LPS[27]. During cell turnovers, PG is continuously broken down and reused through the PG recycling pathway to produce new PG, and in some bacteria, PG recycling is critical for their long-term survival when growth is stalled under nutrient limitation[28]. Thus, such a recycling mechanism seems to be key for maintenance of cell integrity in SBE-LCA under oligotrophic and lethally freezing conditions[21]. 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 thus 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] and the polar insect *Belgica antarctica* larvae[31]. Extremely low temperature also made substrate acquisition difficult due to increased lipid stiffness and decreased efficiency and affinity of membrane transporters[32]. 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[33], 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[34,35], 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 (Fig. 1A).

An additional stress to *Prochlorococcus* during Snowball Earth would have been variable osmotic pressure. Although multiple refugia might have supported bacterial survival during Snowball Earth, the schizohaline nature of these refugia must have imposed strong osmotic pressure on ice-trapped bacteria[21]. As temperature dropped, salts would have become increasingly concentrated in brine channels[20], whereas cryoconite ponds would have remained hyposaline, similar to modern Arctic and Antarctic cryoconite ecosystems[36]. Glycine betaine (GB) is among the most important organic osmolytes in halophilic cyanobacteria, and is used as the major osmolyte in *Synechococcus* sp. WH8102[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 (Fig. 1A). 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 (GG) synthesis and the *gpgS* encoding glucosyl-phosphoglycerate synthase for glucosylglycerate (GGA) synthesis (Fig. 1A). GG and GGA may be sufficient to provide osmotic tolerance under moderately saline conditions[38]. Interestingly, the biosynthesis of GG and GGA requires less N compared to that of GB[38,39], and thus the potential use of GG/GGA instead of GB appeared favorable to SBE-LCA, which had a reduced efficiency of membrane transporters and a low affinity for external nutrients at exceedingly low temperature.

Low light intensity during Snowball Earth was another formidable challenge to phototrophs including *Prochlorococcus*. In the Neoproterozoic, the Sun was still at least 6% dimmer than that at present[40]. Moreover, sea ice, especially when covered with snow, is an effective barrier to light transmission[41]. This is in analogy to the deeper layers of today's polar snow and glacier ice where irradiation is reduced and photosynthetic organisms and activities are scarcely detectable[42]. Consequently, photosynthetic organisms trapped in the brine channels or inhabiting waters below ice need to be physiologically geared to cope with low light. It was proposed that modification of the photosystem structure enables adaptation to the low light condition[43]. 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 (Fig. 1A), but the molecular mechanism of these changes underlying low light adaptation is poorly understood. We also inferred an expansion of the *Prochlorococcus* antenna Pcb from two to six copies during the Snowball Earth (Fig. 1A), which may boost the light-harvesting capacity under low-light conditions[44].

Many cyanobacteria have a sophisticated circadian clock, which is essential in controlling global diel transcriptional activities of the cells. This circadian oscillator system requires only three components: KaiA, KaiB and KaiC[45]. 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[46-48]. Our analysis indicated that *kaiA* was lost at SBE-LCA (Fig. 1A). This is likely due to the prolonged darkness or low light conditions during the Snowball Earth, rendering the sophisticated circadian clock dispensable.

We argue that the genome reduction and metabolic adaptation events discussed above not only enabled *Prochlorococcus* to survive the Snowball Earth climate catastrophe, but also shaped 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* likely resulted in the reduced cell size and increased surface-to-volume ratio in their descendants, which may have enhanced their efficiency in nutrient acquisition[49] and eventually led them to dominate the photosynthetic communities in the most oligotrophic regions of today's oceans[2]. 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 GG and GGA instead of nitrogen-rich GB 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. On the other hand, modifications of some important metabolic pathways may also 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 likely 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[50,51]. In the absence of the circadian clock, however, species such as *Prochlorococcus* cannot synchronize the endogenous oscillation with the environmental cycles and thus are under high risks of cell damages[52].

Concluding Remarks

The Neoproterozoic Snowball Earth hypothesis was proposed decades ago, which claimed the entire extinction of the photosynthetic organisms[17]. In contrast to the original "hard" version of the hypothesis, a modified "soft" version of the Snowball Earth hypothesis was later proposed to include the likely persistence of refugia across the Cryogenian Period, which allowed for the survival of bacterial and simple eukaryotic lineages[53,54]. Survivors of the Snowball Earth included photosynthetic microorganisms[55,56], which enabled continuous primary production across the interval[53,57]. Like other autotrophic organisms at the base of a food web, the survival of *Prochlorococcus* was likely important in sustaining primary production, heterotrophy and carbon cycling, as well as broader ecosystem functioning during the Snowball Earth glaciations[53].

On the other extreme, a few studies have proposed that microbial communities might have been only mildly affected by the Snowball Earth climate catastrophe[56,57]. These inferences were based on the microfossil and biomarker records, which, due to the lack of lineage-specificity, did not capture the nuances required to reconstruct effects on many ecologically important lineages like the *Prochlorococcus* studied here. Instead, 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 type 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 precise molecular clock analyses as well as the ancestral genome reconstruction, furthermore, 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. They also illustrate how *Prochlorococci* acted as important "guardians of metabolism"[12], safeguarding photosynthetic metabolic potential across the Snowball Earth climate catastrophe.

Materials and Methods

Genomic sequences of Cyanobacteria were downloaded from public databases and manually annotated (Table S2; see Section 1 in Supplemental Methods). Divergence time of *Prochlorococcus* was estimated with MCMCTree v4.9e[58] on top of 27 genes (Table S3) 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[14,60] or at the basal of the Macrocyanobacteria group[61,62]. Our analysis showed that this controversy is likely caused by the inclusion of composition-heterogeneous proteins, and that using composition-homogeneous proteins led to consistent support for the former hypothesis (Fig. S4; Table S4; see Section 2.2 in Supplemental Methods). Since molecular dating analysis is known to be intrinsically associated with calibration points[63], we summarized the calibrations of Cyanobacteria used in previous studies, and modified them for our own analyses with caution (Table S1; see Section 2.3 in Supplemental Methods for justification). We further assessed the fitness of different molecular clock models implemented in MCMCTree by using the package "mcmc3r" v0.3.2, 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 (Fig. S5). By implementing statistical tests based on the "infinite-site" theory (Fig. S3; see Section 2.6 in Supplemental Methods) we were able to select the most precise estimates of *Prochlorococcus* evolutionary timeline for illustration (Fig. S6) and further discussion.

Evolution of genome content via gene gains and losses was inferred using two independent methods, AnGST[64] and BadiRate v1.35[65]. 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 (Fig. S7)[64]. 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 maximizes 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 (Fig. S8), and three strategies with highest likelihood values were used (Fig. S9). Further, results derived from AnGST (Fig. S10) and BadiRate were compared and summarized to determine the common patterns shared by the two software (Fig. S11), 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 Section 4 in Supplemental Methods) based on two independent amino acid classification schemes (Table S5).

References

1 Biller, S. J., Berube, P. M., Lindell, D. & Chisholm, S. W. Prochlorococcus: the structure and function of collective diversity. *Nat Rev Microbiol* **13**, 13 (2015).

2 Johnson, Z. I. *et al.* Niche partitioning among Prochlorococcus ecotypes along ocean-scale environmental gradients. *Science (80-)* **311**, 1737-1740 (2006).

3 West, N. J. & Scanlan, D. J. 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 (1999).

4 Hess, W. R. *et al.* The photosynthetic apparatus of Prochlorococcus: insights through comparative genomics. *Photosynth Res* **70**, 53-71 (2001).

5 Moore, L. R., Rocap, G. & Chisholm, S. W. Physiology and molecular phylogeny of coexisting Prochlorococcus ecotypes. *Nature* **393**, 464 (1998).

6 Batut, B., Knibbe, C., Marais, G. & Daubin, V. Reductive genome evolution at both ends of the bacterial population size spectrum. *Nat Rev Microbiol* **12**, 841 (2014).

7 Luo, H., Friedman, R., Tang, J. & Hughes, A. L. Genome reduction by deletion of paralogs in the marine cyanobacterium Prochlorococcus. *Mol Biol Evol* **28**, 2751-2760 (2011).

8 Hu, J. & Blanchard, J. L. Environmental sequence data from the Sargasso Sea reveal that the characteristics of genome reduction in Prochlorococcus are not a harbinger for an escalation in genetic drift. *Mol Biol Evol* **26**, 5-13 (2008).

9 Luo, H., Huang, Y., Stepanauskas, R. & Tang, J. Excess of non-conservative amino acid changes in marine bacterioplankton lineages with reduced genomes. *Nat Microbiol* **2**, 17091 (2017).

10 Zuckerkandl, E. & Pauling, L. in *Evolving genes and proteins* 97-166 (Elsevier, 1965).

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

12 Falkowski, P. G., Fenchel, T. & Delong, E. F. The microbial engines that drive Earth's biogeochemical cycles. *Science (80-)* **320**, 1034-1039 (2008).

13 Kettler, G. C. *et al.* Patterns and implications of gene gain and loss in the evolution of Prochlorococcus. *PLoS Genet* **3**, e231 (2007).

14 Sánchez-Baracaldo, P. Origin of marine planktonic cyanobacteria. *Sci Rep* **5**, 17418 (2015).

15 Sánchez-Baracaldo, P., Ridgwell, A. & Raven, J. A. A neoproterozoic transition in the marine nitrogen cycle. *Current Biology* **24**, 652-657 (2014).

16 Marshall, C. R. Using the Fossil Record to Evaluate Timetree Timescales. *Front Genet* **10**, 1049, doi:10.3389/fgene.2019.01049 (2019).

17 Hoffman, P. F., Kaufman, A. J., Halverson, G. P. & Schrag, D. P. A Neoproterozoic snowball earth. *Science (80-)* **281**, 1342-1346 (1998).

18 Ashkenazy, Y. *et al.* Dynamics of a Snowball Earth ocean. *Nature* **495**, 90 (2013).

19 Zinser, E. R. *et al.* Influence of light and temperature on Prochlorococcus ecotype distributions in the Atlantic Ocean. *Limnol Oceanogr* **52**, 2205-2220 (2007).

20 Thomas, D. & Dieckmann, G. Antarctic sea ice--a habitat for extremophiles.

Science (80-) **295**, 641-644 (2002).

21 Hoffman, P. F. *et al.* Snowball Earth climate dynamics and Cryogenian geology-geobiology. *Sci Adv* **3**, e1600983 (2017).

22 Takeuchi, N. Optical characteristics of cryoconite (surface dust) on glaciers: the relationship between light absorbency and the property of organic matter contained in the cryoconite. *Annals of Glaciology* **34**, 409-414 (2002).

23 Benforte, F. C. *et al.* Novel role of the LPS core glycosyltransferase WapH for cold adaptation in the Antarctic bacterium Pseudomonas extremaustralis. *PLoS One* **13**, e0192559 (2018).

24 Wang, Z., Wang, J., Ren, G., Li, Y. & Wang, X. Influence of core oligosaccharide of lipopolysaccharide to outer membrane behavior of Escherichia coli. *Mar Drugs* **13**, 3325-3339 (2015).

25 Feller, G. & Gerday, C. Psychrophilic enzymes: hot topics in cold adaptation. *Nat Rev Microbiol* **1**, 200 (2003).

26 Uhde, A. *et al.* Glucosamine as carbon source for amino acid-producing Corynebacterium glutamicum. *Appl Microbiol Biotechnol* **97**, 1679-1687 (2013).

27 Park, J. T. & Uehara, T. How Bacteria Consume Their Own Exoskeletons (Turnover and Recycling of Cell Wall Peptidoglycan). *Microbiology and Molecular Biology Reviews* **72**, 211-227, doi:10.1128/mmbr.00027-07 (2008).

28 Borisova, M. *et al.* Peptidoglycan Recycling in Gram-Positive Bacteria Is Crucial for Survival in Stationary Phase. *MBio* **7**, doi:10.1128/mBio.00923-16 (2016).

29 Schumann, W. Regulation of bacterial heat shock stimulons. *Cell Stress and Chaperones* **21**, 959-968 (2016).

30 La Terza, A., Papa, G., Miceli, C. & Luporini, P. 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 (2001).

31 Rinehart, J. P. *et al.* Continuous up-regulation of heat shock proteins in larvae, but not adults, of a polar insect. *Proc Natl Acad Sci U S A* **103**, 14223-14227, doi:10.1073/pnas.0606840103 (2006).

32 Lawrence, R. P. & William, J. W. Temperature and substrates as interactive limiting factors for marine heterotrophic bacteria. *Aquatic Microbial Ecology* **23**, 187-204 (2001).

33 Gleitz, M., v.d. Loeff, M. R., Thomas, D. N., Dieckmann, G. S. & Millero, F. J. Comparison of summer and winter inorganic carbon, oxygen and nutrient concentrations in Antarctic sea ice brine. *Mar Chem* **51**, 81-91, (1995).

34 Reay, D. S., Nedwell, D. B., Priddle, J. & Ellis-Evans, J. C. Temperature dependence of inorganic nitrogen uptake: reduced affinity for nitrate at suboptimal temperatures in both algae and bacteria. *Appl. Environ. Microbiol.* **65**, 2577-2584 (1999).

35 Raven, J. A., Wollenweber, B. & Handley, L. L. A comparison of ammonium and nitrate as nitrogen sources for photolithotrophs. *New Phytologist* **121**, 19-32 (1992).

36 Webster-Brown, J. G., Hawes, I., Jungblut, A. D., Wood, S. A. & Christenson, H. K. The effects of entombment on water chemistry and bacterial assemblages in

436 closed cryoconite holes on Antarctic glaciers. *FEMS Microbiol Ecol* **91** (2015).
437 37 Mao, X. *et al.* Computational prediction of the osmoregulation network in
438 Synechococcus sp. WH8102. *BMC Genomics* **11**, 291, doi:10.1186/1471-2164-
439 11-291 (2010).
440 38 Scanlan, D. J. *et al.* Ecological genomics of marine picocyanobacteria. *Microbiol.*
441 *Mol. Biol. Rev.* **73**, 249-299 (2009).
442 39 Empadinhas, N. & da Costa, M. S. To be or not to be a compatible solute:
443 bioversatility of mannosylglycerate and glucosylglycerate. *Syst Appl Microbiol*
444 **31**, 159-168 (2008).
445 40 Carver, J. H. & Vardavas, I. M. Precambrian glaciations and the evolution of the
446 atmosphere. *Ann Geophys* **12**, 674-682, doi:10.1007/s00585-994-0674-3 (1994).
447 41 Raven, J. A., Kübler, J. & Beardall, J. Put out the light, and then put out the light.
448 *Journal of the Marine Biological Association of the United Kingdom* **80**, 1-25,
449 doi:10.1017/s0025315499001526 (2000).
450 42 Simon, C., Wiezer, A., Strittmatter, A. W. & Daniel, R. Phylogenetic diversity and
451 metabolic potential revealed in a glacier ice metagenome. *Appl Environ Microbiol*
452 **75**, 7519-7526, doi:10.1128/aem.00946-09 (2009).
453 43 Kouřil, R., Wientjes, E., Bultema, J. B., Croce, R. & Boekema, E. J. High-light vs.
454 low-light: effect of light acclimation on photosystem II composition and
455 organization in Arabidopsis thaliana. *Biochimica et Biophysica Acta (BBA)-*
456 *Bioenergetics* **1827**, 411-419 (2013).
457 44 Bibby, T., Mary, I., Nield, J., Partensky, F. & Barber, J. Low-light-adapted
458 Prochlorococcus species possess specific antennae for each photosystem. *Nature*
459 **424**, 1051 (2003).
460 45 Dong, G. & Golden, S. S. How a cyanobacterium tells time. *Curr Opin Microbiol*
461 **11**, 541-546 (2008).
462 46 Holtzendorff, J. *et al.* Genome streamlining results in loss of robustness of the
463 circadian clock in the marine cyanobacterium Prochlorococcus marinus PCC
464 9511. *J Biol Rhythms* **23**, 187-199, doi:10.1177/0748730408316040 (2008).
465 47 Axmann, I. M. *et al.* Biochemical evidence for a timing mechanism in
466 prochlorococcus. *J Bacteriol* **191**, 5342-5347, doi:10.1128/jb.00419-09 (2009).
467 48 Mella-Flores, D. *et al.* Prochlorococcus and Synechococcus have Evolved
468 Different Adaptive Mechanisms to Cope with Light and UV Stress. *Front*
469 *Microbiol* **3**, 285, doi:10.3389/fmicb.2012.00285 (2012).
470 49 Giovannoni, S. J., Thrash, J. C. & Temperton, B. Implications of streamlining
471 theory for microbial ecology. *ISME J* **8**, 1553 (2014).
472 50 Warters, R. L. & Lyons, B. W. Variation in radiation-induced formation of DNA
473 double-strand breaks as a function of chromatin structure. *Radiat Res* **130**, 309-
474 318 (1992).
475 51 Simons, M. J. The evolution of the cyanobacterial posttranslational clock from a
476 primitive "phoscillator". *J Biol Rhythms* **24**, 175-182 (2009).
477 52 Mullineaux, C. W. & Stanewsky, R. The rolex and the hourglass: a simplified
478 circadian clock in Prochlorococcus? *J Bacteriol* **191**, 5333-5335 (2009).
479 53 Moczydłowska, M. The Ediacaran microbiota and the survival of Snowball Earth
480 conditions. *Precambrian Res* **167**, 1-15 (2008).

54 Hoffman, P. F. & Schrag, D. P. The snowball Earth hypothesis: testing the limits of global change. *Terra nova* **14**, 129-155 (2002).

55 Allison, C. W. & Awramik, S. M. Organic-walled microfossils from earliest Cambrian or latest Proterozoic Tindir Group rocks, northwest Canada. *Precambrian Res* **43**, 253-294 (1989).

56 Corsetti, F. A., Awramik, S. M. & Pierce, D. A complex microbiota from snowball Earth times: microfossils from the Neoproterozoic Kingston Peak Formation, Death Valley, USA. *Proc Natl Acad Sci USA* **100**, 4399-4404 (2003).

57 Olcott, A. N., Sessions, A. L., Corsetti, F. A., Kaufman, A. J. & De Oliviera, T. F. Biomarker evidence for photosynthesis during Neoproterozoic glaciation. *Science (80-)* **310**, 471-474 (2005).

58 Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-1591 (2007).

59 Battistuzzi, F. U. & Hedges, S. B. A major clade of prokaryotes with ancient adaptations to life on land. *Mol Biol Evol* **26**, 335-343 (2008).

60 Shih, P. M. *et al.* Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci USA* **110**, 1053-1058 (2013).

61 Blank, C. & Sanchez-Baracaldo, P. Timing of morphological and ecological innovations in the cyanobacteria–a key to understanding the rise in atmospheric oxygen. *Geobiology* **8**, 1-23 (2010).

62 Uyeda, J. C., Harmon, L. J. & Blank, C. E. A comprehensive study of cyanobacterial morphological and ecological evolutionary dynamics through deep geologic time. *PLoS One* **11**, e0162539 (2016).

63 Schirrmeister, B. E., Sanchez-Baracaldo, P. & Wacey, D. Cyanobacterial evolution during the Precambrian. *Int J Astrobiol* **15**, 187-204 (2016).

64 David, L. A. & Alm, E. J. Rapid evolutionary innovation during an Archaean genetic expansion. *Nature* **469**, 93 (2011).

65 Librado, P., Vieira, F. & Rozas, J. BadiRate: estimating family turnover rates by likelihood-based methods. *Bioinformatics* **28**, 279-281 (2011).

Acknowledgments

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. This work is supported by the National Natural Science Foundation of China (92051113), the Hong Kong Research Grants Council General Research Fund (14110820), the Hong Kong Research Grants Council Area of Excellence Scheme (AoE/M-403/16), HKU FoS funds to SAC, and the Direct Grant of CUHK (4053257 and 3132809).

Author Contributions

H.L. conceived and directed the study. H.Z. and Y.S. performed the bioinformatics. All authors contribute to the interpretation of the results. H.Z., Y.S., S.A.C and H.L. wrote the paper.

Conflict of interests

The author declares no competing interests.

Fig. 1

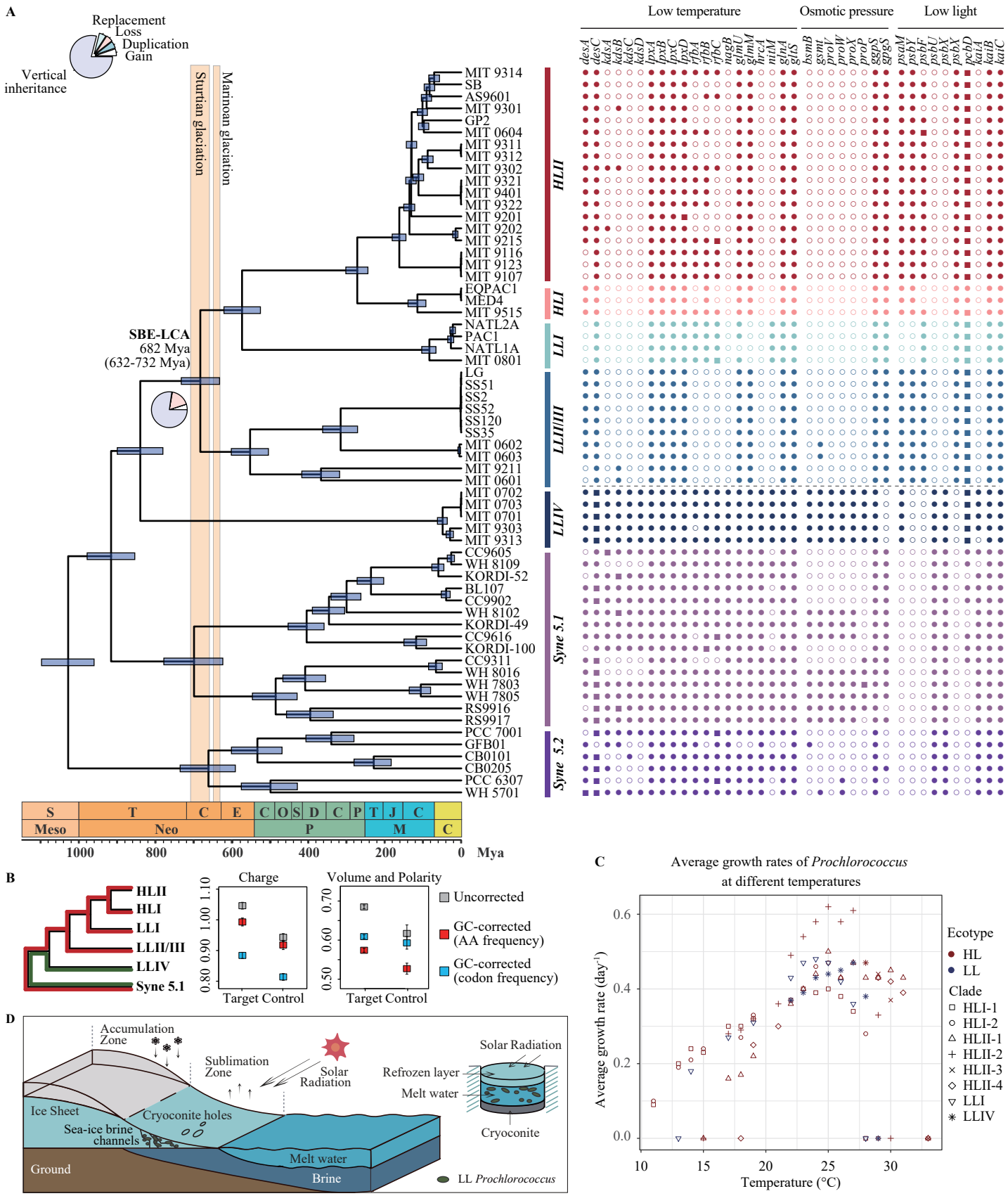
A
Replacement
Loss
Duplication
Gain
Vertical inheritance
Sturtian glaciation
Marinoan glaciation
SBE-LCA
682 Mya
(632-732 Mya)
MIT 9314
SB
AS9601
MIT 9301
GP2
MIT 0604
MIT 9311
MIT 9312
MIT 9302
MIT 9321
MIT 9401
MIT 9322
MIT 9201
MIT 9202
MIT 9215
MIT 9116
MIT 9123
MIT 9107
EQPAC1
MED4
MIT 9515
NATL2A
PAC1
NATL1A
MIT 0801
LG
SS51
SS2
SS52
SS120
SS35
MIT 0602
MIT 0603
MIT 9211
MIT 0601
MIT 0702
MIT 0703
MIT 0701
MIT 9303
MIT 9313
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
HLII
HLI
LLI
LLII/III
LLIV
Syne 5.1
Syne 5.2
S T C E C O S D C P T J C
Meso Neo P M C
1000 800 600 400 200 0 Mya
Low temperature
Osmotic pressure
Low light
desA desC kdsA kdsB kdsC kdsD lpxA lpxB lpxC lpxD rfbA rfbB rfbC nagB glmU glmM hrcA niiM glnA gltS
bsmB gsmt proV proW proX proP proQ ggpS gpgS
psaM psbY psbF psbU psbX psbX pcbD katA katB katC
B
HLII
HLI
LLI
LLII/III
LLIV
Syne 5.1
Charge
Volume and Polarity
Target Control
Target Control
Uncorrected
GC-corrected (AA frequency)
GC-corrected (codon frequency)
C
Average growth rates of Prochlorococcus at different temperatures
Average growth rate (day⁻¹)
Temperature (°C)
Ecotype
HL
LL
Clade
HLI-1
HLI-2
HLII-1
HLII-2
HLII-3
HLII-4
LLI
LLIV
D
Accumulation Zone
Sublimation Zone
Solar Radiation
Solar Radiation
Refrozen layer
Melt water
Cryoconite
Ice Sheet
Cryoconite holes
Sea-ice brine channels
Melt water
Brine
Ground
LL Prochlorococcus

Fig. 1 Evolution of *Prochlorococcus* during the Neoproterozoic Snowball Earth events. (A) (Left) Chronogram of the evolutionary history of *Prochlorococcus* estimated by MCMCTree. The evolutionary tree shown here is part of the species tree constructed with MrBayes based on 90 compositionally homogenous gene families shared by 159 cyanobacterial genomes (Fig. S4 D). Divergence time is estimated based on 27 gene sequences under calibration set C9 (Table S1). The vertical bars represent the estimated time of the Neoproterozoic glaciation events. The flanking horizontal blue 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. (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 & Right) is calculated by comparing a genome from the 'Target' group to a genome from the 'reference' group (marked with red), followed by averaging the value across all possible genome pairs. Likewise, the d_R/d_C for the 'Control' group (shown in Middle & Right) is calculated by comparing a genome from the 'Control' group to a

genome from the ‘reference’ group (marked with green) and then by averaging the value across all possible genome pairs. (Middle & 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 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 values represent the standard error of the mean. (C) Diagram of putative bacterial refugia including cryoconite holes and sea ice brine channels in Neoproterozoic Snowball Earth, which were featured with a number of stresses such as low temperature, low light, and variable osmotic pressure. (D) 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. 2006 and Zinser et al. 2007.