



Diverse Subclade Differentiation Attributed to the Ubiquity of *Prochlorococcus* High-Light-Adapted Clade II

Wei Yan,^{a,b} Xuejin Feng,^b Ta-Hui Lin,^b Xingyu Huang,^b Le Xie,^b Shuzhen Wei,^b Kun Zhou,^{c,d} Yi-Lung Chen,^b Weicheng Luo,^b Wenqian Xu,^e Wei Zhang,^a Muhammad Zohaib Nawaz,^b Ya-Wei Luo,^b Qinglu Zeng,^{d,f,g} Rui Zhang,^b Nianzhi Jiao^b

^aCollege of Marine Science and Technology, China University of Geosciences, Wuhan, People's Republic of China

^bState Key Laboratory of Marine Environmental Science, Fujian Key Laboratory of Marine Carbon Sequestration, College of Ocean and Earth Sciences, Xiamen University, Xiamen, People's Republic of China

^cShenzhen University-HKUST Joint Marine Science PhD Program, Shenzhen University, Shenzhen, People's Republic of China

^dDepartment of Ocean Science, Hong Kong University of Science and Technology, Kowloon, Hong Kong SAR, People's Republic of China

^eDepartment of Statistics, China University of Geosciences, Wuhan, People's Republic of China

^fDivision of Life Science, Hong Kong University of Science and Technology, Kowloon, Hong Kong SAR, People's Republic of China

^gHong Kong Branch of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Hong Kong University of Science and Technology, Hong Kong SAR, People's Republic of China

Xuejin Feng and Ta-Hui Lin contributed equally to this work. Author order was determined alphabetically.

ABSTRACT *Prochlorococcus* is the key primary producer in marine ecosystems, and the high-light-adapted clade II (HLII) is the most abundant ecotype. However, the genomic and ecological basis of *Prochlorococcus* HLII in the marine environment has remained elusive. Here, we show that the ecologically coherent subclade differentiation of HLII corresponds to genomic and ecological characteristics on the basis of analyses of 31 different strains of HLII, including 12 novel isolates. Different subclades of HLII with different core and accessory genes were identified, and their distribution in the marine environment was explored using the TARA Oceans metagenome database. Three major subclade groups were identified, *viz.*, the surface group (HLII-SG), the transition group (HLII-TG), and the deep group (HLII-DG). These subclade groups showed different temperature ranges and optima for distribution. In regression analyses, temperature and nutrient availability were identified as key factors affecting the distribution of HLII subclades. A 35% increase in the relative abundance of HLII-SG by the end of the 21st century was predicted under the Representative Concentration Pathway 8.5 scenario. Our results show that the ubiquity and distribution of *Prochlorococcus* HLII in the marine environment are associated with the differentiation of diverse subclades. These findings provide insights into the large-scale shifts in the *Prochlorococcus* community in response to future climate change.

IMPORTANCE *Prochlorococcus* is the most abundant oxygenic photosynthetic microorganism on Earth, and high-light-adapted clade II (HLII) is the dominant ecotype. However, the factors behind the dominance of HLII in the vast oligotrophic oceans are still unknown. Here, we identified three distinct groups of HLII subclades, *viz.*, the surface group (HLII-SG), the transition group (HLII-TG), and the deep group (HLII-DG). We further demonstrated that the ecologically coherent subclade differentiation of HLII corresponds to genomic and ecological characteristics. Our study suggests that the differentiation of diverse subclades underlies the ubiquity and distribution of *Prochlorococcus* HLII in the marine environment and provides insights into the shifts in the *Prochlorococcus* community in response to future climate change.

KEYWORDS *Prochlorococcus* HLII, subclade differentiation, pangenome, metagenome

Editor Stephen J. Giovannoni, Oregon State University

Copyright © 2022 Yan et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Wei Yan, yanwei@cug.edu.cn, Rui Zhang, ruizhang@xmu.edu.cn, or Nianzhi Jiao, jiao@xmu.edu.cn.

The authors declare no conflict of interest.

This article is a direct contribution from Nianzhi Jiao, a Fellow of the American Academy of Microbiology, who arranged for and secured reviews by Fei-Xue Fu, University of Southern California, and Jinjun Kan, Stroud Water Research Center.

Received 3 February 2022

Accepted 23 February 2022

Published 14 March 2022

The marine cyanobacterium *Prochlorococcus* is the most abundant and smallest known oxygenic photosynthetic microorganism, and it is mainly distributed in the euphotic zone of tropical and subtropical oligotrophic oceans. Its abundance in marine environments typically ranges from 10^4 to 10^5 cells mL⁻¹, which accounts for 30% to 80% of the total photosynthetic biomass and net primary productivity in certain regions of the open ocean (1–8). According to recent estimates, the annual mean global abundance of *Prochlorococcus* in the ocean is approximately 3×10^{27} cells, and this value is expected to increase by one-third by the end of the 21st century in response to global warming under the Representative Concentration Pathway (RCP) 4.5 scenario (5).

Based on physiological characteristics, ecological distribution, and phylogeny, the 12 *Prochlorococcus* clades can generally be divided into two ecotypes: the high-light-adapted (HL) clades and low-light-adapted (LL) clades (9–20). High-light-adapted clade II (HLII) and its counterpart, HLI, which are distinguished by their optimal growth temperature (17), have a key evolutionary feature known as genome streamlining, which helps these cyanobacteria to reproduce rapidly under energy-limiting and nutrient-starved conditions in oligotrophic oceans (4, 21, 22). HLII account for more than 90% of total *Prochlorococcus* in tropical open oceans (17). Thus, bearing in mind the huge abundance of *Prochlorococcus*, HLII is one of the most abundant photosynthetic organisms on the planet if we consider it as a distinct cyanobacterial species.

The highly compact genome of individual HLII strains restricts their ability to adapt to a wide range of environmental conditions. Because HLII populations in the euphotic zone are continuously exposed to steep gradients of environmental factors, such as light intensity, temperature, and nutrient concentrations, they have evolved to respond and adapt to such conditions (20, 22–25). Additionally, the extremely large populations and relatively high growth rates of HL *Prochlorococcus* promote differentiation among closely related strains and impose strong selective pressures within the ecological niches in surface and subsurface waters (22). Several studies have reported a strong relationship between HLII gene content and environmental variables (26–28). Two environmental single-cell genomic studies revealed that subpopulations tend to coexist within HLII (23, 24). Investigations based on marker genes, including environmental ITS (16S–23S rRNA internal transcribed spacer) (20, 25), *petB* (29), and *rpoC1* sequences (30–32), revealed associations between HL subpopulations and environmental parameters. Overall, the results of those studies revealed the existence of diverse ecologically coherent subclades within HLII and identified potentially disparate properties that were previously masked by coarse phylogenetic resolution.

The genomic and physiological differences among HLII subclades and the influence of ecological niches on subclade differentiation remain largely unexplored, mainly due to the inadequate number of available whole genomes for detailed mining and the lack of large-scale environmental metagenomic studies. This gap, therefore, hampers our understanding of how HLII became the most abundant *Prochlorococcus* clade. In this study, we isolated 12 novel strains of *Prochlorococcus* HLII from the western Pacific Ocean and the South China Sea and performed a genomic analysis of 31 different *Prochlorococcus* HLII genomes. The environmental occurrence of HLII subclades was explored using data from the TARA Oceans metagenomic database (33). Finally, the environmental factors that drive HLII subclade distribution were identified using regression models, and their shifts in relative abundances in response to future global warming were predicted.

RESULTS

New HLII isolates from the western Pacific Ocean. Twelve new *Prochlorococcus* strains were isolated from depths of 50 to 300 m from six different sites in the western Pacific Ocean and the South China Sea (see Table S1 and Fig. S1). Only one ITS sequence was found in the contigs belonging to each genome, which confirmed the unialgal status of the HLII cultures. Their taxonomic positions were confirmed through

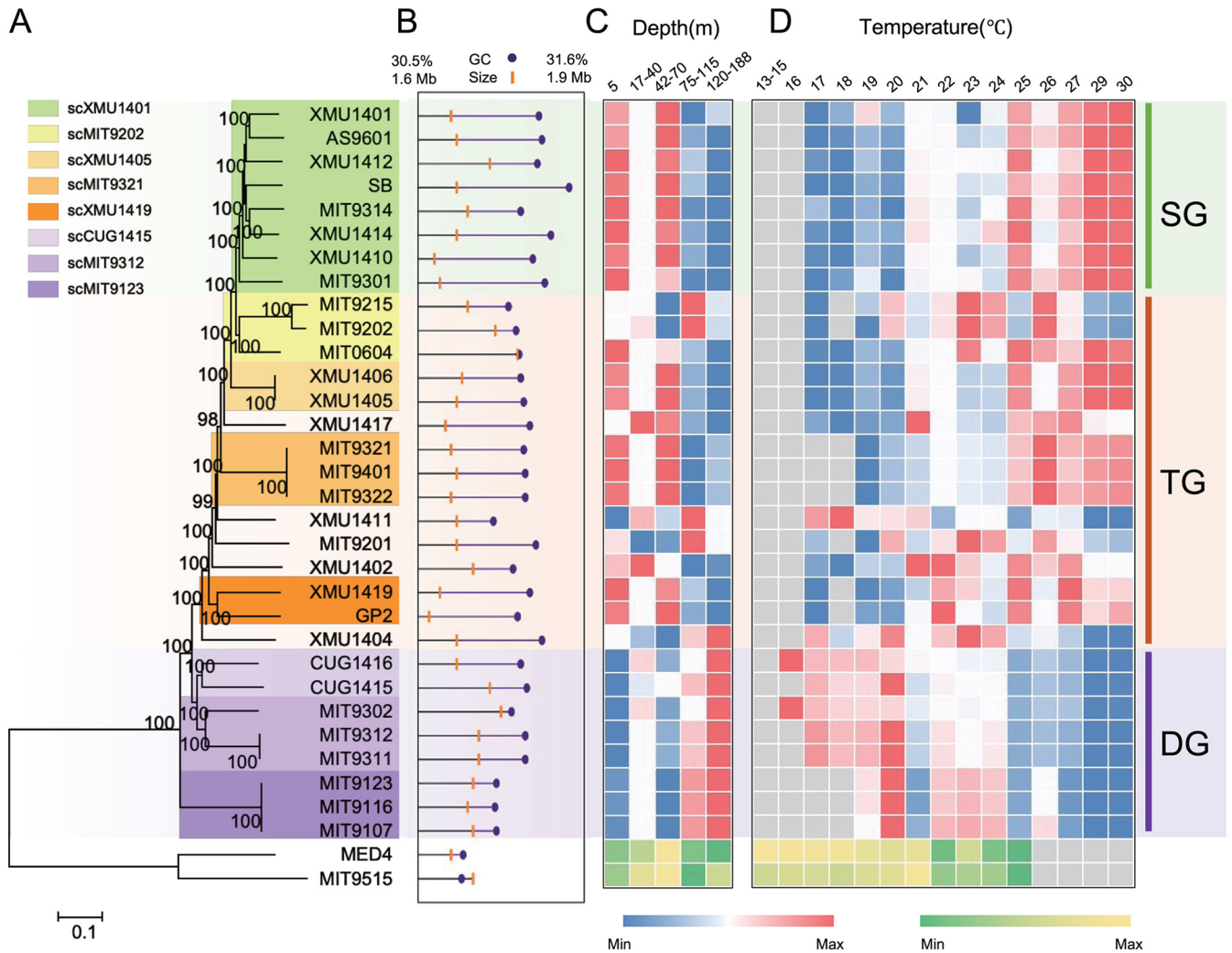


FIG 1 Phylogenomic diversity and environmental distribution of HLII strains. (A) Maximum likelihood phylogenetic tree constructed from the core genome single nucleotide polymorphisms (SNPs) of 31 *Prochlorococcus* HLII strains, with two HLII strains as an outgroup. Numbers at nodes are bootstrap values for 1000 resamplings (at least 80% support). (B) Genome characteristics and isolation depths of 33 HL strains. Heatmaps of log-transformed relative abundances in TARA Oceans metagenome libraries for each strain at different depths (C) and temperatures (D). Color gradient indicates increasing values of the relevant measurement from minimums (blue or green) to maximums within a row (red or yellow). Gray color indicates no detection. SG, TG, and DG indicate *Prochlorococcus* HLII surface group, transition group, and deep group, respectively. Relative abundances of HLII strains were normalized to the total abundance of all HLII reads. As an outgroup, relative abundances of HLI strains were normalized to the total abundance of HLI and HLII reads.

comprehensive phylogenetic analyses (see below and Fig. 1A and Fig. S2 and S3). The genome size of the 12 HLII strains ranged from 1,631,014 bp to 1,734,399 bp, with an average GC content of 31.38% (standard deviations [SD] of 0.09) and an average of 1,964 (SD of 37) coding sequences per genome (Fig. 1B and Table S1). Two strains, including CUG1417 and XMU1419, were isolated from a depth of 300 m, well below the euphotic zone and near the Luzon Strait in the South China Sea. These two strains represent *Prochlorococcus* of the deepest origin to date.

Phylogeny of HLII subclades. The phylogeny of the HLII subclades was inferred using ITS sequences and concatenated protein sequences from multiple core genes (defined at similarity cutoffs ranging from 50% to 95%). In most cases, the resulting phylogenetic trees were topologically similar and showed that *Prochlorococcus* HLII consists of several subclades (Fig. S2 and S3). Notably, the branch lengths of the core gene protein trees were relatively short, mainly because protein sequences encoded by core genes are highly conserved and all HLII strains are relatively closely related to each other (Fig. S3). Thus, to improve the resolution of the *Prochlorococcus* HLII subclade phylogeny, we constructed phylogenetic trees using core genome single

nucleotide polymorphisms (SNPs) at a range of similarity cutoffs (50% to 95%) (Fig. 1A and Fig. S3). Figure 1A shows the core genome SNP tree (at similarity cutoff of 60%) with high bootstrap support for its topology. The SNPs were distributed throughout the genome with high abundance and genetic stability, whereas the core genome contained essential genes that are often vertically transmitted and have a strong signal-to-noise ratio for phylogenetic inference, resulting in a high phylogenetic resolution (34). Thus, the tree constructed on the basis of core genome SNPs represented the clear phylogenetic relationships among *Prochlorococcus* HLII genomes and suggests that the 31 HLII strains could be divided into several subclades, including scXMU1401, scMIT9202, scXMU1405, scMIT9321, scXMU1419, scCUG1415, scMIT9312, and scMIT9123 (Fig. 1A).

Environmental occurrences of HLII subclades. To estimate the relative abundance and distribution patterns of the HLII subclades in marine environments, their genomes were mapped to 107 TARA Oceans metagenomic samples (0.2- to 1.6- μm - and 0.2- to 3- μm -size fractions) covering the surface/subsurface (5 to 188 m) layers of temperate to tropical oceans (Fig. S1 and Table S2). The results show that HL (including HLI and HLII) genomes were detected in 89 metagenomes, and the HLII subclades differed in their relative abundance at different depths and temperatures. Three ecological subclade groups of *Prochlorococcus* HLII strains, viz., the surface group (HLII-SG), the transition group (HLII-TG), and the deep group (HLII-DG), were defined based on their relative abundances at various depths initially, which interestingly showed an obvious pattern of temperature (Fig. 1D). The strains belonging to basal clades (highlighted with violet) demonstrated the maximum relative abundance of representatives dwelling in deep waters with a continuous and obvious pattern (Fig. 1C) and were regarded as HLII-DG. Similarly, representative strains of the terminal clade (shown in green) demonstrate a maximum abundance of strains inhabiting the surface water and were regarded as HLII-SG (Fig. 1C). Noticeably, HLII-SG mainly represents the high-temperature strains, whereas low-temperature strains mainly constitute the HLII-DG (Fig. 1C and D and Fig. S4). The strains representing the middle clades and showing dispersal patterns were named HLII-TG. HLII-SG, containing mainly scXMU1401, accounted for more than 52% of all HLII reads in our data set after normalizing the sizes of all libraries (Fig. S4). HLII-SG was most abundant in the upper euphotic layer (5 to 70 m) (Fig. 1C) and showed the highest relative abundance within the temperature range of 25°C to 30°C (Fig. 1D). In contrast, the HLII-DG group (scCUG1415, scMIT9312, and scMIT9123) was mainly restricted to mid-deep layers (75 to 188 m) (Fig. 1C) and had a lower temperature range of 16°C to 23°C (Fig. 1D). Notably, scMIT9123 was distantly related to all other HLII subclades in the phylogenetic tree. The strains in this subclade were isolated from a depth of 25 m (Table S1). However, analyses of its overall distribution pattern indicated that this subclade inhabits mid- or deep layers (Fig. 1C); therefore, it was grouped into HLII-DG. The scMIT9202, scXMU1405, scMIT9321, and scXMU1419 subclades in HLII-TG showed a complex distribution pattern with respect to depth and temperature. For example, HLII-TG subclades, including scXMU1405 and scMIT9321, demonstrated distribution patterns closer to that of HLII-SG, whereas the HLII-TG subclades, including scMIT9202 and scXMU1419, showed distribution patterns similar to that of HLII-DG (Fig. 1C and D). The outgroup, consisting of HLI strains, was dominant in the temperature range of 13°C to 23°C, in contrast to the temperature distribution patterns of the HLII subclades (Fig. 1D). Because depth is negatively correlated with light intensity in the water column, these observed distribution patterns reflect the significance of both light and temperature in shaping the distribution of the ecologically coherent subclades of HLII in the ocean.

Differentiation of the HLII core genome. On the whole-genome scale, we observed differences in GC content and genome size between HLII-SG and HLII-DG. The average GC content of HLII-SG ($31.32\% \pm 0.09\%$) was higher than that of HLII-DG ($31.12\% \pm 0.09\%$) ($P = 0.0008$, two-sample t test, $n = 8$) (Fig. 2A), and the average genome size of HLII-SG (1.670 ± 0.028) was smaller than that of HLII-DG (1.707 ± 0.023 Mbp) ($P = 0.0164$, two-sample t test, $n = 8$) (Fig. 2B). Additionally, the GC content of the core genome SNPs was higher in HLII-SG ($36.91\% \pm 0.16\%$) than in HLII-DG ($36.39\% \pm 0.31\%$)

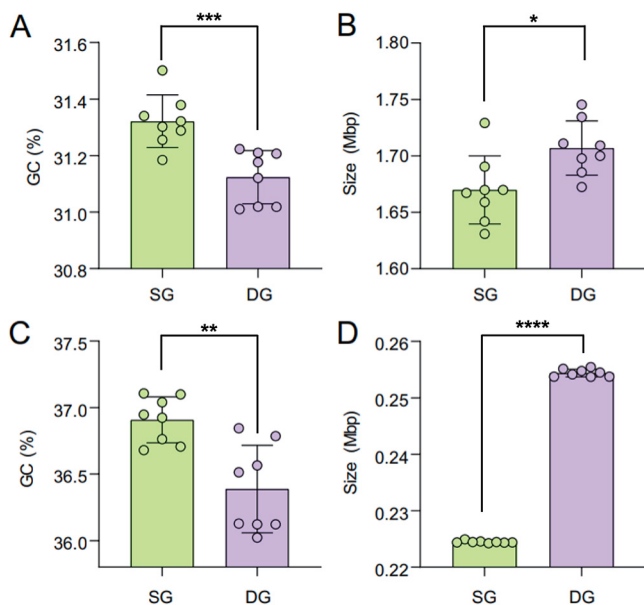


FIG 2 Contrasting *Prochlorococcus* HLII surface group and deep group. Panels A and B represent genome sizes and GC contents of whole genomes from *Prochlorococcus* HLII surface group (HLII-SG) and deep group (HLII-DG), respectively. Panels C and D represent genome sizes and GC contents of core genome single nucleotide polymorphisms (SNPs) from HLII-SG and HLII-DG, respectively. Colored circles represent individual genome values, and bars indicate one standard deviation. The number of colored circles denotes the number of statistical tests showing a significant difference (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$).

($P = 0.0014$, two sample t test, $n = 8$) (Fig. 2C). There were fewer core genome SNPs in HLII-SG ($224,472 \pm 190$ bp) than in HLII-DG ($254,402 \pm 631$ bp) ($P < 0.0001$, two sample t test, $n = 8$) (Fig. 2D). These differences in GC contents are similar to those reported in previous studies that detected variations in GC contents among different *Synechococcus* and *Prochlorococcus* clades and subclades in different ecological niches, especially thermal and variable salinity niches (35, 36).

In this study, the HLII pangenome was separated into three parts: the core genome (genome fragments/genes shared by all 31 HLII genomes), the accessory genome (genome fragments/genes shared by 3 to 30 HLII genomes, thereby targeting most genes potentially associated with subclade differentiation), and the strain-specific accessory genome (genome fragments that appeared in only one or two strains) (Fig. 3A). Noticeably, in this study, the pangenome analysis was performed among strains of the same ecotype, which are closely related to each other. For example, strains MIT9107, MIT9116, MIT9321, and MIT9401 share average nucleotide identity of $>99\%$ with MIT9116, MIT9123, MIT9322, and MIT9321, respectively. Considering such nearly identical genomes, genes limited to one/two strains were regarded as strain-specific genes and excluded from accessory gene analysis, since we tried to target most genes that might be related to subclade differentiation.

To identify core genes associated with subclade differentiation, we applied two similarity cutoffs to the 31 HLII genomes: a 50% similarity cutoff (amino acid level) resulted in 1,502 core genes, and a 70% cutoff resulted in 1,402 core genes (Fig. 3A). Therefore, 100 core genes with amino acid similarity between 50% and 70% were identified. These genes were considered core genes based on 50% amino acid similarity cutoff but fell in the category of accessory genes (resulting in 242 genes) while using a 70% cutoff (Fig. 3B). Hierarchical clustering based on Euclidean distance was performed using the binary matrix of these 242 genes (Fig. 3B), and the resulting dendrogram was generally congruent with the core genome SNP tree (Fig. 1A). Several of these genes with known functions were related to high-light protection, DNA repair, and nutrient

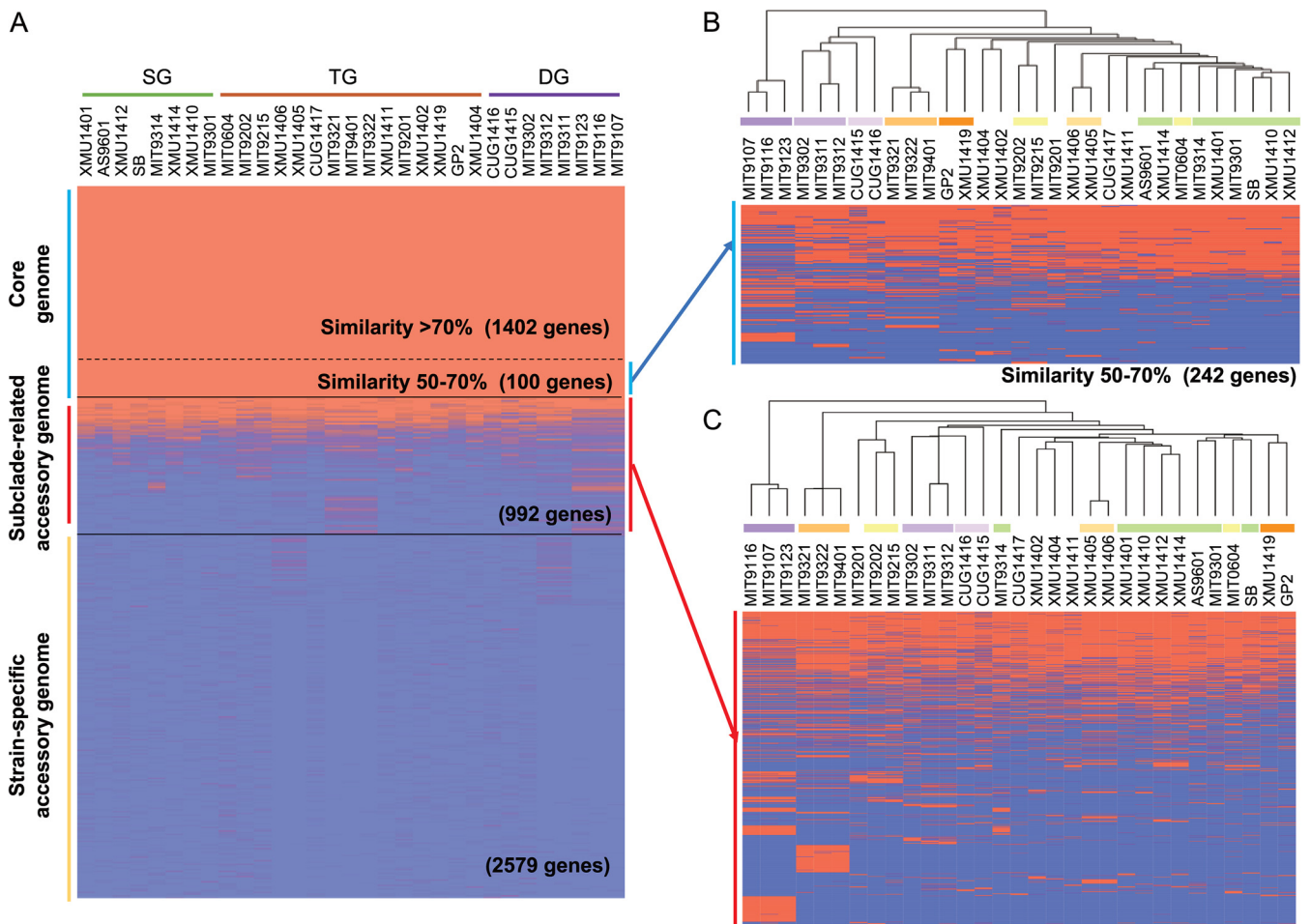


FIG 3 Pangenome of *Prochlorococcus* HLII. (A) Distribution of 5,073 genes in the pangenome of *Prochlorococcus* HLII. (B) Distribution of the genes which were considered core genes based on 50% amino acid similarity cutoff but fell in the category of accessory genes while using 70% cutoff. (C) Distribution of 992 accessory genes in the 31 genomes of *Prochlorococcus* HLII. Dendrograms above the two heatmaps show hierarchical clustering of the 31 HLII genomes based on gene distribution (Euclidian distance). SG, TG, and DG indicate *Prochlorococcus* HLII surface group, transition group, and deep group, respectively.

transport (Table S3). For example, the urea assimilation gene cluster that is present in most of the HLII genomes showed differences between HLII-SG and HLII-DG (Table S3).

At the SNP level, we used two filters (see Materials and Methods) to screen the 31 HLII core genomes. We found that most of the regions with substantial SNP variations were located between or within genes with unknown function (Fig. S5). Interestingly, three core genes (*chlB*, *nblS*, and *amtB*) were found to have relatively large SNP differences between HLII-SG and HLII-DG (Fig. S5). The *chlB* gene encodes a subunit of light-independent protochlorophyllide reductase, which is required for the light-independent accumulation of chlorophyll (37). The *nblS* gene encodes a protein that is involved in changes in the photosynthetic apparatus in response to both high light and nutrient limitation (38). The *amtB* gene encodes a low-affinity ammonium uptake protein that is a member of the Amt family of ammonium/ammonia transporters. The transporter is responsible for bidirectional ammonium diffusion across the membrane and is only necessary for growth at low ammonium concentrations (39, 40). These results suggest that the core genomes of HLII subclades have differentiated to adapt to the diverse environments in the stratified euphotic zone, e.g., to variations in light intensity and nutrient concentrations. A recent study also found that adaptation of the picocyanobacteria clade and subclades to ecological niches relies on amino acid substitutions in the core genome rather than on variations in gene content (35). Overall, our results

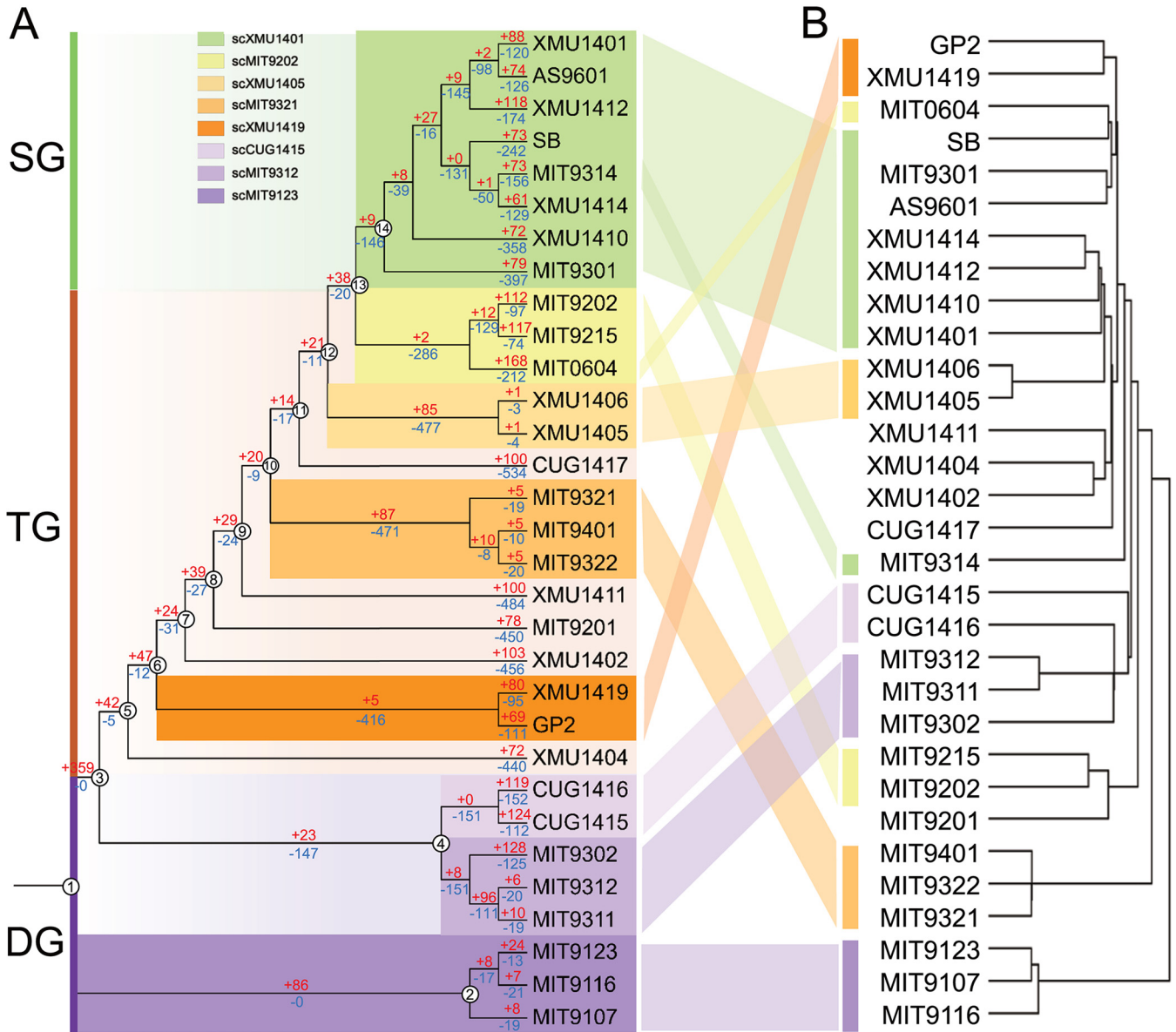


FIG 4 Relationships among *Prochlorococcus* HLII genomes based on phylogenomics (A) and on distribution (B) of 992 accessory genes. (A) Tree on the left shows the phylogeny constructed from core genome single-nucleotide polymorphisms (SNPs) of 31 *Prochlorococcus* HLII strains using maximum likelihood (same as Fig. 1A). Two HLI strains (MED4 and MIT9515) were used as an outgroup (not shown). Numbers on branches indicate estimated number of genes gained (red) and lost (blue) during evolution of *Prochlorococcus* HLII. (B) Tree on the right shows the clustering of the same HLII genomes based on the distribution of 992 accessory genes recovered from the pangenomic analysis (Euclidian distance).

show that HLII strains have a small, conserved core genome (1,402 genes, defined at 70% amino acid similarity), a subclade-specific core genome (100 genes defined at 50 to 70% amino acid similarity), and a subclade-specific accessory genome (992 genes) while having a large strain-specific genome (2,579 genes, although the actual size may be much larger).

Dynamics of evolution of HLII gene content. The evolution of the HLII subclades was further explored based on their accessory genome content. In this analysis, the similarity between subclades was defined on the basis of their shared genes, and the underlying gene content was interpreted in terms of subclade ecological divergence. The topologies of the accessory genome tree and the core genome SNP tree were largely congruent (Fig. 4A and B [same as Fig. 3C]), indicating that accessory genome differentiation echoes core genome evolution. We found several genes in the accessory genome that were associated with ecological differentiation. For example, most

genomes from HLII-DG lacked genes related to DNA repair that were present in HLII-SG genomes (e.g., DNA ligase, DEXH box helicase) (Table S3). Notably, all 31 HLII isolates carried *hli* genes encoding high-light-induced proteins (HLIPs) that are involved in protection against light-induced damage at the ocean surface (41). However, several *hli* genes were also found in the accessory genome, suggesting that different subclades carry different sets of *hli* genes (Table S3).

To get more insights into the evolutionary processes that led to the subclade differentiation of gene content within HLII genomes, we estimated the number of gene-gain and gene-loss events on each branch of the core genome SNP tree (Fig. 4A and Table S3). Branches 2 and 3, the first two branches after the divergence of the HLII ancestor (branch 1), gained 86 and 359 genes, respectively (Fig. 4A). Out of 359 genes gained on branch 3 in the core genome SNP tree, 15 were related to DNA replication and repair and 20 to transporters (Table S3). Seven genes were related to photosynthesis and light protection, three genes were related to nitrogen metabolism, and five genes were related to phosphorus metabolism (Table S3). Branches 2 and 3 exhibited large differences and were at the beginning of the evolutionary tree, indicating that scMIT9123 and the common ancestor of subclades other than scMIT9123 have been differentiated for a relatively long time after their divergence. Branch 3, containing the subclades other than scMIT9123, gained a large number of genes, indicating higher genetic diversity. This may have resulted from adaptation to the surface and subsurface ocean environments during evolutionary history. Branch 4, comprising HLII-DG in the deep euphotic zone, was subdivided into two distinct subclades, scMIT9312 and scCUG1415. One gene (*hli*) encoding HLIP has been lost from the strains positioned on branch 4 (Table S3), as they were not essential for adaptation in this environment due to its weak light intensity. Likewise, because the deeper layers of the euphotic zone have very little exposure to UV light, intracellular DNA is not damaged by UV radiation, and eight genes related to DNA repair functions have been lost from strains positioned on this branch.

In the core genome SNP tree, HLII-TG comprised a total of nine branches, from branches 5 to 13 (Fig. 4A) (Table S3). Overall, in this group, 14 transporter genes were acquired and nine were lost; nine DNA replication and repair genes were acquired and three were lost; two nitrogen metabolism genes were gained and three were lost; and one phosphorus metabolism gene was gained. Some HLII-TG subclades (including scXMU1419, scMIT9321, scXMU1405, and scMIT9202) might have been exposed to high light and high levels of UV radiation during the adaptation process in the upper and mid-euphotic zone. Therefore, constant acquisition of DNA repair-related genes was an evolutionary requirement to survive in this environment. Transporter genes were randomly gained and lost, which shows that adaptation strategies evolved in response to fluctuating nutrient conditions in the upper and mid-euphotic layer during evolution in HLII-TG. Branch 14 lost 146 genes, the highest number lost among these branches (Fig. 4A) (Table S3), implying that the HLII-SG (scXMU1401) lost genes related to adaptation to the nutrient-poor environment of the surface layer and the genome was effectively streamlined to lower their nutrient and energy requirements.

Drivers of global HLII subclade composition and distribution. Potential relationships between the relative abundances of HLII subclades (normalized to the total abundance of all HLII reads) and environmental variables, including latitude, depth, temperature, and nutrient concentrations, were identified using univariate regression analyses (Fig. 5 and Table S4). Specifically, the relative abundance of scXMU1401 (HLII-SG) showed a positive relationship with temperature and a quadratic relationship with latitude and NO_2/NO_3 concentration. In HLII-DG, scCUG1415 showed relationships with temperature, latitude, and NO_2/NO_3 and PO_4 concentrations. scMIT9312 showed relationships with latitude and temperature. scMIT9123 was affected by PO_4 and NO_2/NO_3 concentrations. The subclades of HLII-TG showed complicated relationships with environmental drivers. All the subclades in this group were affected by NO_2/NO_3 and PO_4 concentrations, except for scMIT9321, which was affected only by the NO_2/NO_3 concentration. scXMU1405 and

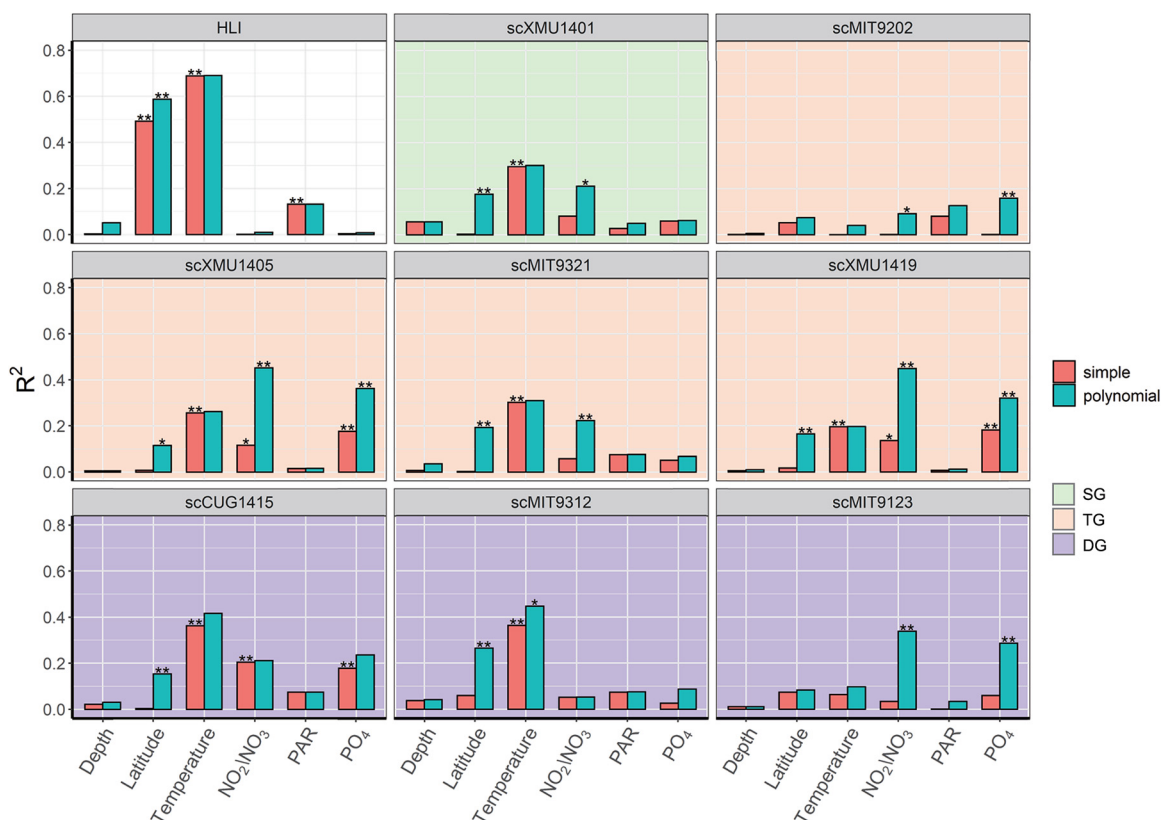


FIG 5 R -squared plots showing explanatory power of environmental variables for relative abundances of HLI and HLII subclades in univariate regression models. *, $P < 0.05$; **, $P < 0.01$. Red columns indicate linear models, and green columns indicate quadratic models. For statistical analyses, HLI relative abundance was normalized to all HL relative abundances (HLI and HLII). Relative abundances of HLII subclades were normalized to relative abundance of all HLII.

scXMU1419 showed relatively strong relationships with NO_2/NO_3 and PO_4 concentrations. All subclades in this group were affected by latitude and temperature, except for scMIT9202. In addition, as an outgroup, HLI abundance (normalized to the total abundance of HLI and HLII reads) exhibited a strong negative relationship with temperature and latitude and a relatively weak relationship with photosynthetically active radiation (PAR), indicating that temperature is a key feature that separates HLI and HLII. The results of multiple regression analyses also showed similar results (Table S4).

In this study, temperature was identified as a key factor associated with subclade distributions within HLII, with relatively high explanatory power for the relative abundances of scXMU1401, scCUG1415, and scMIT9312 (Fig. 5). Sea surface temperatures are predicted to increase under future global warming (42, 43). Therefore, we used univariate regression models to predict the relative abundances of HLII subclades and HLI in the global ocean for the 1990s and 2090s with simulated sea surface temperatures of historical and RCP 4.5 and 8.5 scenarios (Fig. S6) (42). HLI and the three HLII subclades for which temperature had relatively high explanatory power were selected for the model predictions (Fig. 5 and Fig. S6). To define the geographic limits of HLII and HLI in the models, the surface ocean temperature boundaries were set to 11°C to 29°C and 14°C to 33°C, respectively, based on the results of metagenomic detection. The results for the model with the present scenario (1990s) showed that HLI was abundantly present in mid-latitude oceans (20° to 45°, either N or S). Within HLII, scXMU1401 (HLII-SG) was the most abundant subclade in low-latitude oceans (30°N to 30°S) (Fig. 6A). The strains scCUG1415 and scMIT9312 in HLII-DG showed a pattern similar to that of HLI but with considerably lower abundance (Fig. 6A). The predicted results for the 2090s showed that the relative abundance of high-temperature-tolerant scXMU1401 (HLII-SG) is likely to

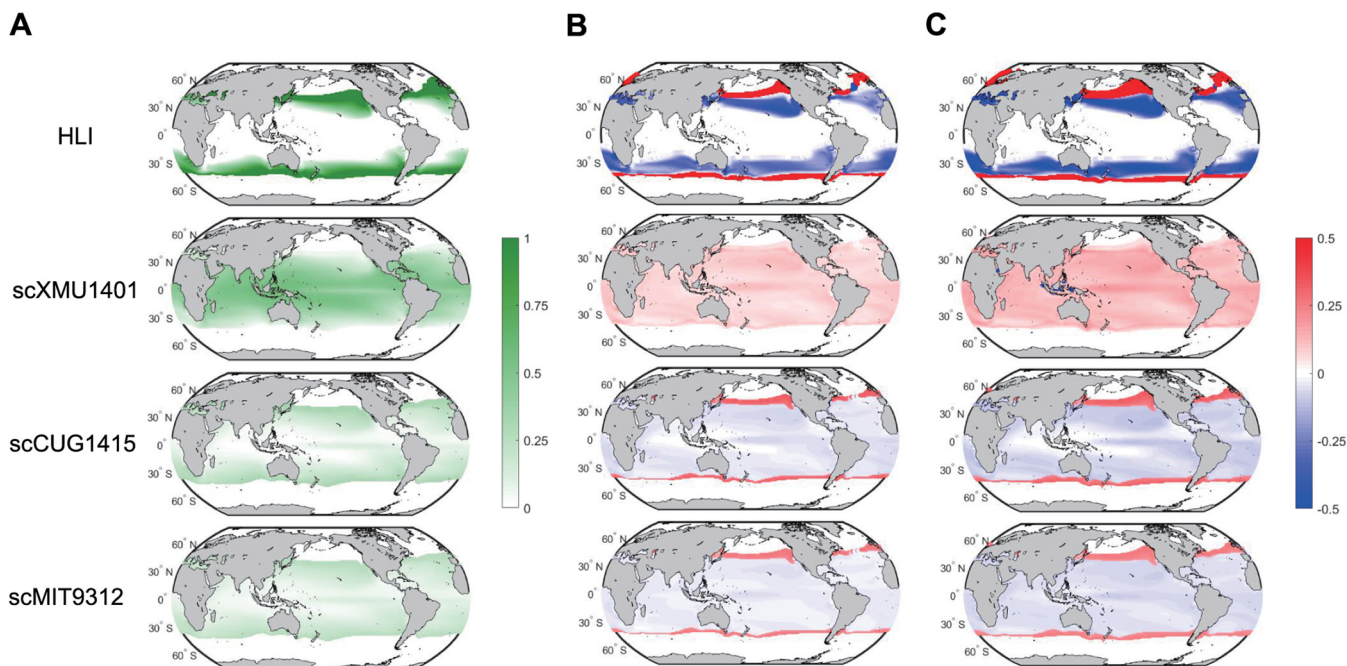


FIG 6 (A) Present global relative abundance and distribution of HLI and major HLII subclades at the sea surface (1990s). Changes in relative abundance between present and future climate (1990s and 2090s) at the sea surface for HLI and major HLII subclades under RCP 4.5 (B) and 8.5 scenarios (C). For statistical analyses, HLI relative abundance was normalized to relative abundance of all HL (HLI and HLII). Relative abundances of HLII subclades were normalized to relative abundance of all HLII.

increase by 21% and 35% under RCP 4.5 and 8.5, respectively, and its distribution limits will expand toward higher latitudes (Fig. 6B and C). In contrast, the relative abundance of HLI was predicted to decline by 36% and 55% in the 2090s under the RCP 4.5 and 8.5 scenarios, respectively. The predicted changes in HLI relative abundance were uneven at different latitudes, declining sharply at low latitudes (20° to 30°, either N or S) but increasing at higher latitudes (30° to 50°, either N or S), thereby moving its distribution further away from lower latitude oceans (Fig. 6B and C). The predicted patterns for scCUG1415 and scMIT9312 were similar to that of HLI (Fig. 6B and C).

DISCUSSION

Prochlorococcus has continuously evolved to adapt to the marine environment, and it can be considered the most recent evolutionary form of cyanobacteria (44). The diverse clades and subclades of *Prochlorococcus* increase its genomic diversity and environmental adaptability, allowing it to become the dominant species in its environment. *Prochlorococcus* HLII is one of the most recently evolved *Prochlorococcus* clades (22). In this study, we report that *Prochlorococcus* HLII consists of three broad subclade groups that can be further divided into ecologically coherent subclades based on their genomic, and ecological differences. Our results suggest that subclade genome differentiation has played an important role in the evolution and ecological niche partitioning of *Prochlorococcus* HLII. After its divergence from the common ancestor of all HLII subclades, HLII-SG may have adapted primarily (e.g., streamlined nonessential genes and increased GC content) to the upper mixed euphotic zone where it was exposed to high temperatures, high photon fluxes, UV radiation, and nutrient limitation. As a result, HLII-SG accounts for the majority of HLII in the upper euphotic zone. In contrast, HLII-DG may have acquired more accessory genes as it gradually adapted to the mid-deep euphotic layers, which feature strong environmental gradients. As a result, their overall genome sizes have increased and their GC contents have decreased. Therefore, gene gain and loss as well as changes in GC content appear to be ongoing and important adaptation strategies of HLII subclades to different environments. Our results also

indicate that *Prochlorococcus* HLII has evolved via long-term adaptation to the high-light and high-temperature environment of the upper surface ocean (4, 22, 26); however, its subclades retain the ability and tendency to adapt to different environments, even in the deep euphotic zone. We note that in this study, we use “transition group” (HLII-TG) to refer to several complex HLII subclades and strains with genomic, physiological, and ecological differences. Because HLII-TG appears to contribute to the evolutionary process of HLII-SG, more detailed studies should be conducted for this group in the future.

Rising ocean temperatures will change the productivity and composition of marine phytoplankton communities, affecting global biogeochemical cycles (45). Temperature has been suggested as an important ecological factor that separates HLII from HLI (17, 46). Our results indicate that temperature and nutrient concentrations drive the microdiversity among HLII subclades, and different subclades are associated with different combinations of environmental parameters. Similarly, Larkin et al. (31) found that differences in the relative abundance of HLII haplotypes in the Indian Ocean are driven by nutrient availability and temperature (31). Our model predictions show that, under future sea surface warming, the HLII subclades, especially HLII-SG, will be much more abundant than at present, and their distribution will expand toward mid-latitude regions, while the relative abundance of HLI and HLII-DG will decline sharply in low-latitude areas but their distribution will shift toward higher latitudes. Our results are consistent with those of a previous study that found that rising temperatures this century will cause poleward shifts in species' thermal niches and a sharp decline in tropical phytoplankton diversity (45). However, our findings represent a much finer-scale phylogenetic level. *Prochlorococcus* absolute abundance is expected to increase by one-third by the end of this century in response to global warming (5). In comparison, our results showed a change in relative abundance of *Prochlorococcus* ecotype and subclade level under global warming. In this study, temperature is found as a key driver controlling the relative abundances of *Prochlorococcus* HLII subclades; however, a few recent studies have shown that ocean warming may also interact with other environmental factors, like nutrients, Fe availability, pCO₂, and light, and have profound influences on phytoplankton community (47–49). It is necessary to consider the interactions between ocean warming and other primary factors to reveal a comprehensive picture of distributions and abundances of HLII subclades in the future ocean.

Using the strategy of diverse subclade differentiation, *Prochlorococcus* HLII enhances its ability to adapt to different environments while maintaining streamlined single-cell genomes that permit rapid reproduction with low energy and nutrient demands. These characteristics have allowed HLII to become the most abundant ecotype of all *Prochlorococcus* and even all marine phytoplankton. The HLII subclades have a long evolutionary history, and their subclade niche partitioning is complicated and ancient. In this study, we conducted hierarchical clustering analyses at two levels (clade and subclade) to reveal the detailed phylogenetic relationships in HLII. However, our results are still insufficient to clarify the relationships in this huge and complex system. Kashtan et al. (23) found that HLII is composed of hundreds of subpopulations with distinct “genomic backbones.” These closely related HLII subpopulations are estimated to have diverged at least several million years ago based on the comparison of environmental single-cell genomes (23). Fundamental differences in the diversity and genomic structure of HLII exist among different oceans (23, 24). Although many single-cell and metagenomic sequences have been obtained, limited isolates and whole-genome sequences are available, which hampers detailed investigation into the fundamental differences in the diversity and genomic structure of HLII strains. Therefore, more efforts should be made to isolate many new and representative HLII strains for analyses of whole-genome sequences and physiological characteristics. Such analyses may provide a more comprehensive picture of HLII subclade differentiation based on differences in genomic and phenotypic characteristics. Overall, our results shed light on the

divergence of *Prochlorococcus* HLII and provide a basis for predicting shifts in distributions of HLII subclades. Our analyses suggest that the expected increase in surface ocean temperature will cause large-scale changes in fine-scale *Prochlorococcus* community structures. Because of the extremely large population size of HLII, these changes are expected to have substantial impacts on ocean ecosystems and global biogeochemical cycles.

MATERIALS AND METHODS

Isolation of *Prochlorococcus* HLII strains. Twelve *Prochlorococcus* HLII strains were isolated from the western Pacific Ocean and the South China Sea in 2014 (see Fig. S1 and Table S1 for details of isolation locations). Seawater was collected with a Niskin bottle and subjected to gravity filtration through double polycarbonate filters (Millipore, USA) with a pore size of 0.6 μm (50, 51). A Pro2 medium nutrient stock solution was then added to the filtrate (50). To avoid cross-contamination, the filtration devices were acid-washed between filtering seawater from different sources. The filtrate was placed into an onboard incubator for initial enrichment. After confirmation by flow cytometry, enriched *Prochlorococcus* were purified using a modified dilution-to-extinction method (52), as described previously (53). The *Prochlorococcus* strains were maintained at a constant temperature of 22°C under a continuous light intensity of 10 to 20 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$.

Genome sequencing, assembly, and annotation. Genomic DNA was collected from 25-mL laboratory cultures by centrifugation (10,000 $\times g$, 30 min) and extracted using the QIAamp DNA minikit (Qiagen, Hilden, Germany). A 1- μg portion of the extracted DNA was fragmented using a Covaris ME220 focused-ultrasonicator (Covaris, Woburn, MA, USA). DNA libraries then were constructed using the NEBNext Ultra DNA library prep kit for Illumina according to the manufacturer's instructions (NEB, Beverly, MA, USA). Bidirectional sequencing was conducted using 10 ng library DNA on an Illumina HiSeq 4000 instrument with a read length of 150 bp or an Illumina MiSeq instrument with a read length of 250 bp (Table S1). All library construction and sequencing procedures were performed at the Shanghai Hanyu Biotechnology Co. (Shanghai, China). To recover *Prochlorococcus* genomes from nonaxenic cultures, genome assembly was performed as described in our previous study (52) using a binning workflow modified from a well-known method (54). After assembly, all contigs were manually checked and reordered using the "Move Contigs" tool in Mauve v2.4.0 (55). The 12 assembled genome sequences were annotated using tools at the RAST online server (Rapid Annotation using Subsystem Technology; FIGfam version release 70) (56) and the KEGG database (57). For comparison, we also reannotated 21 previously published *Prochlorococcus* HL genomes (19 HLII and 3 HLI) and 19 LL genomes using the same methods (Table S1).

Phylogenomic analysis of HLII subclades. Multiple methods were used to construct the evolutionary relationships among *Prochlorococcus* HLII strains. First, ITS sequences were aligned using MAFFT v7 (58) and a phylogenetic tree was constructed using the maximum-likelihood approach with 1,000 bootstrap replicates in RAxML v8.0 (59). Second, phylogenetic trees were constructed using concatenated protein sequences encoded by core genes. In brief, the protein sequences encoded by core genes defined at multiple amino acid similarity levels (ranging from 1,502 genes at 50% to 69 genes at 95%) were concatenated and then aligned with MAFFT v7 (58). The phylogenetic trees were then constructed using a maximum-likelihood approach with 1,000 bootstrap replicates in RAxML v8.0 (59). Finally, phylogenetic trees were constructed using single nucleotide polymorphisms (SNPs) from core genes defined at similarity levels ranging from 50% to 95%. Core genome polymorphic bases were identified and concatenated using the Panseq online server (60) with the following parameters: BLAST fragmentation size, 500 bp; homology fragment similarity percent identity cutoffs, 50 to 95%; BLAST size, 20. SNP-based phylogenomic trees with 1,000 bootstrap replicates were then constructed using RAxML v8.0 (59). All phylogenetic trees presented in this study were visualized using MEGA 7 (61).

Environmental distribution of HLII subclades. We searched the TARA Oceans metagenome database by following a reproducible workflow (http://merenlab.org/data/2018_Delmont_and_Eren_Metapangenomics/) that focused on linking *Prochlorococcus* pangenomes and metagenomes (27). In brief, 107 TARA Oceans metagenomes (fractions 0.2 to 1.6 μm and 0.2 to 3 μm in size) from the European Bioinformatics Institute repositories (Table S2) were downloaded and quality filtered. These samples covered low- and middle-latitude surface/subsurface waters (5- to 188-m depth) of major oceans. An *anvi'o* contig database describing 54 *Prochlorococcus* genomes was generated and mapped to quality-filtered metagenomic short reads from the TARA Oceans samples. To ensure that there was no substantive cross-recruitment from LL clades to HL clades, 20 LL genomes were added to the contig database. To investigate the influence of depth, we separated the 89 TARA Oceans metagenomes in which HL genomes were detected into the surface layer (5-m depth); the 17- to 40-m layer; the 42- to 70-m layer; the 75- to 115-m layer; and the 120- to 188-m layer. To investigate the influence of temperature, we separated the samples into 15 temperature ranges between 0°C and 31°C. Information on the reference genomes and the TARA Oceans metagenomes used in this study is presented in Table S2.

Pangenome analysis of HLII strains. Fragment-based pangenome analysis of the 31 HLII strains was performed using the Panseq online server (60) with parameters set as described above. Multiple BLAST fragmentation sizes (500 to 5,000 bp) and homology fragment similarity cutoffs (50 to 70%) were tested. RAST gene functional subsystem analysis was performed (56). Gene-based pangenome analysis

was conducted using the Bacterial Pangenome Analysis Pipeline (BPGA) v1.3.0 (62), which uses USEARCH algorithms to identify core genes. Hierarchical clustering based on Euclidean distance was performed using binary data, i.e., the presence or absence of accessory genes in each genome. Two amino acid similarity levels (70% and 50%) were used to identify core genes with significant amino acid variations. The Harvest suite (including Parsnp and Gingr) (34) was used to profile SNP positions across the 31 HLII genomes, with the “random reference,” “recombination detection,” and “inclusion of all genomes” options selected. Two criteria were used to identify core genes with substantial SNP variations: (i) the gene had SNPs in an aligned 100-bp window with <50% identity or (ii) the gene had SNPs in an aligned 100-bp window with >20 indels. The number of gene gains and losses were assessed using Count software (63), which uses a Dollo parsimony method for estimating the ancestral state of every gene in the data set using the core genome SNP tree as the reference. The Dollo parsimony method was used to infer the presence-absence of each gene cluster at each node of the core genome SNP tree and to calculate the number of gene gains and losses on each branch (63).

Modeling analysis. Statistical analyses were conducted, resulting in predictive models for the relative abundance of the HLII subclades and HLI at warmer temperatures. Statistical analyses were performed using the R statistical platform (R software, v. 3.6.1; Vienna, Austria). Model prediction and visualization were performed using MATLAB 9.5.0. Environmental factors were gathered from the environmental data set of TARA Oceans (64). To calculate the PAR at different depths, 8-day average diffuse attenuation coefficient at 490 nm (Kd490) values were derived from satellite data (Aqua satellite, 4 km). Detailed information on the environmental factors associated with the TARA Oceans metagenomes used in this study is provided in Table S2.

Correlation analyses were conducted to test for possible autocorrelations among the environmental variables. Univariate regression analyses (linear and quadratic) were used to identify relationships between the measured environmental variables and the relative abundances of the HLII subclades and HLI, using R^2 values of the regression model to evaluate the explanatory power of each environmental variable. The resulting univariate regression models were used to predict the annual average relative abundance of HLII subclades and HLI in the global surface ocean under different temperatures in 1990 to 2000 and 2090 to 2100. Simulated sea surface temperatures under historical and RCP 4.5 and 8.5 scenarios (42) were generated from three global circulation models (CSIRO Mk3 6 0, CanESM2, and HadGEM2 ES) in the Coupled Model Intercomparison Project (<https://esgf-node.llnl.gov/search/cmip5/>) (43). Multivariate regression analyses were also conducted to identify relationships between the environmental variables and the relative abundances of the HLII subclades and HLI, using Akaike information criterion (AIC) to add or remove variable during the stepwise procedure and R^2 values to evaluate the explanatory power of multiple environmental variables.

Data availability. The whole-genome sequence data of the 12 HLII strains reported in this study have been deposited in NCBI GenBank and the Genome Warehouse of the National Genomics Data Center under BioProject numbers PRJNA611498 and PRJCA002394, respectively. Accession numbers for the genomes of individual strains can be found in Table S1.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.4 MB.

FIG S2, PDF file, 0.4 MB.

FIG S3, PDF file, 0.4 MB.

FIG S4, PDF file, 0.4 MB.

FIG S5, PDF file, 0.4 MB.

FIG S6, PDF file, 0.5 MB.

TABLE S1, XLSX file, 0.02 MB.

TABLE S2, XLSX file, 0.04 MB.

TABLE S3, XLSX file, 0.1 MB.

TABLE S4, XLSX file, 0.03 MB.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (41906110 and 91751207 to W.Y., 41890802 to Y.L.), China Ocean Mineral Resources R&D Association (DY135-E2-1-04 to N.J.), the Fundamental Research Funds for the Central Universities (20720190141 to W.Y., 20720170107 to N.J.), and the Hong Kong Branch of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (SMSEGL20SC01 to Q.Z.).

W.Y., R.Z., and N.J. designed the study. X.F., T.L., and S.W. contributed to the physiological experiments. X.F., X.H., K.Z., and W.Z. participated in genome and metagenome analysis. L.X., W.L., W.X., and Y.L. contributed to the interpretation of the statistical results. Y.C., M.Z.N., and Q.Z. contributed to the interpretation of the results of

the genome analysis and physiological experiment. W.Y. and R.Z. wrote the manuscript.

All authors contributed to the final version of the manuscript.

We declare no competing interests.

REFERENCES

- Campbell L, Liu H, Nolla HA, Vault D. 1997. Annual variability of phytoplankton and bacteria in the subtropical North Pacific Ocean at Station ALOHA during the 1991–1994 ENSO event. *Deep Sea Res Part I Oceanogr Res Pap* 44:167–192. [https://doi.org/10.1016/S0967-0637\(96\)00102-1](https://doi.org/10.1016/S0967-0637(96)00102-1).
- Vault D, Marie D, Olson RJ, Chisholm SW. 1995. Growth of *Prochlorococcus*, a photosynthetic prokaryote, in the equatorial Pacific Ocean. *Science* 268:1480–1482. <https://doi.org/10.1126/science.268.5216.1480>.
- Liu H, Nolla HA, Campbell L. 1997. *Prochlorococcus* growth rate and contribution to primary production in the equatorial and subtropical North Pacific Ocean. *Aquat Microb Ecol* 12:39–47. <https://doi.org/10.3354/ame012039>.
- Partensky F, Garczarek L. 2010. *Prochlorococcus*: advantages and limits of minimalism. *Annu Rev Mar Sci* 2:305–331. <https://doi.org/10.1146/annurev-marine-120308-081034>.
- Flombaum P, Gallegos JL, Gordillo RA, Rincón J, Zabala LL, Jiao N, Karl DM, Li WKW, Lomas MW, Veneziano D, Vera CS, Vrugt JA, Martiny AC. 2013. Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci U S A* 110:9824–9829. <https://doi.org/10.1073/pnas.1307701110>.
- Jiao N, Yang Y. 2002. Ecological studies on *Prochlorococcus* in China seas. *Chinese Sci Bull* 47:1243–1250. <https://doi.org/10.1360/02tb9276>.
- Casey JR, Lomas MW, Mandecki J, Walker DE. 2007. *Prochlorococcus* contributes to new production in the Sargasso Sea deep chlorophyll maximum. *Geophys Res Lett* 34:L10604. <https://doi.org/10.1029/2006GL028725>.
- Partensky F, Hess WR, Vault D. 1999. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* 63:106–127. <https://doi.org/10.1128/MMBR.63.1.106-127.1999>.
- Rocap G, Distel DL, Waterbury JB, Chisholm SW. 2002. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol* 68:1180–1191. <https://doi.org/10.1128/AEM.68.3.1180-1191.2002>.
- Martiny AC, Tai APK, Veneziano D, Primeau F, Chisholm SW. 2009. Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. *Environ Microbiol* 11:823–832. <https://doi.org/10.1111/j.1462-2920.2008.01803.x>.
- Lavin P, González B, Santibáñez JF, Scanlan DJ, Ulloa O. 2010. Novel lineages of *Prochlorococcus* thrive within the oxygen minimum zone of the eastern tropical South Pacific. *Environ Microbiol Rep* 2:728–738. <https://doi.org/10.1111/j.1758-2229.2010.00167.x>.
- Huang S, Wilhelm SW, Harvey HR, Taylor K, Jiao N, Chen F. 2012. Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J* 6:285–297. <https://doi.org/10.1038/ismej.2011.106>.
- Garczarek L, Dufresne A, Rousvoal S, West NJ, Mazard S, Marie D, Claustre H, Raimbault P, Post AF, Scanlan DJ, Partensky F. 2007. High vertical and low horizontal diversity of *Prochlorococcus* ecotypes in the Mediterranean Sea in summer. *FEMS Microbiol Ecol* 60:189–206. <https://doi.org/10.1111/j.1574-6941.2007.00297.x>.
- Rusch DB, Martiny AC, Dupont CL, Halpern AL, Venter JC. 2010. Characterization of *Prochlorococcus* clades from iron-depleted oceanic regions. *Proc Natl Acad Sci U S A* 107:16184–16189. <https://doi.org/10.1073/pnas.1009513107>.
- West NJ, Lebaron P, Strutton PG, Suzuki MT. 2011. A novel clade of *Prochlorococcus* found in high nutrient low chlorophyll waters in the South and Equatorial Pacific Ocean. *ISME J* 5:933–944. <https://doi.org/10.1038/ismej.2010.186>.
- Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, DeLong EF, Chisholm SW. 2006. Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* 311:1768–1770. <https://doi.org/10.1126/science.1122050>.
- 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. <https://doi.org/10.1126/science.1118052>.
- Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA, Arellano A, Coleman M, Hauser L, Hess WR, Johnson ZI, Land M, Lindell D, Post AF, Regala W, Shah M, Shaw SL, Steglich C, Sullivan MB, Ting CS, Tolonen A, Webb EA, Zinser ER, Chisholm SW. 2003. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* 424:1042–1047. <https://doi.org/10.1038/nature01947>.
- Moore LR, Rocap G, Chisholm SW. 1998. Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* 393:464–467. <https://doi.org/10.1038/30965>.
- Thompson AW, Kouba K, Ahlgren NA. 2021. Niche partitioning of low-light adapted *Prochlorococcus* subecotypes across oceanographic gradients of the North Pacific Subtropical Front. *Limnol Oceanogr* 66:1548–1562. <https://doi.org/10.1002/lno.11703>.
- Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, Rodrigue S, Chen F, Lapidus A, Ferreira S, Johnson J, Steglich C, Church GM, Richardson P, Chisholm SW. 2007. Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet* 3:e231. <https://doi.org/10.1371/journal.pgen.0030231>.
- Biller SJ, Berube PM, Lindell D, Chisholm SW. 2015. *Prochlorococcus*: the structure and function of collective diversity. *Nat Rev Microbiol* 13:13–27. <https://doi.org/10.1038/nrmicro3378>.
- Kashtan N, Kashtan N, Roggensack SE, Roggensack SE, Rodrigue S, Rodrigue S, Thompson JW, Thompson JW, Biller SJ, Biller SJ, Coe A, Coe A, Ding H, Ding H, Marttinen P, Marttinen P, Malmstrom RR, Malmstrom RR, Stocker R, Stocker R, Follows MJ, Follows MJ, Stepanauskas R, Stepanauskas R, Chisholm SW, Chisholm SW. 2014. Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* 344:416–420. <https://doi.org/10.1126/science.1248575>.
- Kashtan N, Roggensack SE, Berta-Thompson JW, Grinberg M, Stepanauskas R, Chisholm SW. 2017. Fundamental differences in diversity and genomic population structure between Atlantic and Pacific *Prochlorococcus*. *ISME J* 11:1997–2011. <https://doi.org/10.1038/ismej.2017.64>.
- Larkin AA, Blinebry SK, Howes C, Lin Y, Loftus SE, Schmaus CA, Zinser ER, Johnson ZI. 2016. Niche partitioning and biogeography of high light adapted *Prochlorococcus* across taxonomic ranks in the North Pacific. *ISME J* 10:1555–1567. <https://doi.org/10.1038/ismej.2015.244>.
- Kent AG, Dupont CL, Yooshep S, Martiny AC. 2016. Global biogeography of *Prochlorococcus* genome diversity in the surface ocean. *ISME J* 10:1856–1865. <https://doi.org/10.1038/ismej.2015.265>.
- Delmont TO, Eren AM. 2018. Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ* 6:e4320. <https://doi.org/10.7717/peerj.4320>.
- Thompson LR, Haroon MF, Shibl AA, Cahill MJ, Ngugi DK, Williams GJ, Morton JT, Knight R, Goodwin KD, Stingl U. 2019. Red Sea SAR11 and *Prochlorococcus* single-cell genomes reflect globally distributed pangenomes. *Appl Environ Microbiol* 85:e00369-19. <https://doi.org/10.1128/AEM.00369-19>.
- Farrant GK, Doré H, Cornejo-Castillo FM, Partensky F, Ratn M, Ostrowski M, Pitt FD, Wincker P, Scanlan DJ, Iudicone D, Acinas SG, Garczarek L. 2016. Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc Natl Acad Sci U S A* 113:E3365–E3374. <https://doi.org/10.1073/pnas.1524865113>.
- Shibl AA, Haroon MF, Ngugi DK, Thompson LR, Stingl U. 2016. Distribution of *Prochlorococcus* ecotypes in the Red Sea basin based on analyses of *rpoC1* sequences. *Front Mar Sci* 3:69–14.
- Larkin AA, Garcia CA, Ingoglia KA, Garcia NS, Baer SE, Twining BS, Lomas MW, Martiny AC. 2020. Subtle biogeochemical regimes in the Indian Ocean revealed by spatial and diel frequency of *Prochlorococcus* haplotypes. *Limnol Oceanogr* 65:2629–2613. <https://doi.org/10.1002/lno.11251>.
- Kent AG, Baer SE, Mouginit C, Huang JS, Larkin AA, Lomas MW, Martiny AC. 2019. Parallel phylogeography of *Prochlorococcus* and *Synechococcus*. *ISME J* 13:430–412. <https://doi.org/10.1038/s41396-018-0287-6>.
- Pesant S, Picheral M, Kandels-Lewis S, Bescot NL, Iudicone D, Karsenti E, Speich S, Troublé R, Dimier C, Searson S, Coordinators TOC, Gorsky G, Not F, Tara Oceans Consortium Coordinators. 2015. Open science resources for the discovery and analysis of Tara Oceans data. *Sci Data* 2:150023. <https://doi.org/10.1038/sdata.2015.23>.
- Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The Harvest suite for rapid core-genome alignment and visualization of thousands of

- intraspecific microbial genomes. *Genome Biol* 15:524. <https://doi.org/10.1186/s13059-014-0524-x>.
35. Doré H, Farrant GK, Guyet U, Haguait J, Humily F, Ratin M, Pitt FD, Ostrowski M, Six C, Brillet-Guéguen L, Hoebeke M, Bisch A, Corguillé GL, Corre E, Labadie K, Aury J-M, Wincker P, Choi DH, Noh JH, Eveillard D, Scanlan DJ, Partensky F, Garczarek L. 2020. Evolutionary mechanisms of long-term genome diversification associated with niche partitioning in marine picocyanobacteria. *Front Microbiol* 11:567431. <https://doi.org/10.3389/fmicb.2020.567431>.
 36. Fuller NJ, Marie D, Partensky F, Vault D, Post AF, Scanlan DJ. 2003. Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Appl Environ Microbiol* 69:2430–2443. <https://doi.org/10.1128/AEM.69.5.2430-2443.2003>.
 37. Li J, Goldschmidt-Clermont M, Timko MP. 1993. Chloroplast-encoded *chlB* is required for light-independent protochlorophyllide reductase activity in *Chlamydomonas reinhardtii*. *Plant Cell* 5:1817–1829. <https://doi.org/10.1105/tpc.5.12.1817>.
 38. van Waasbergen LG, Dolganov N, Grossman AR. 2002. *nblS*, a gene involved in controlling photosynthesis-related gene expression during high light and nutrient stress in *Synechococcus elongatus* PCC 7942. *J Bacteriol* 184:2481–2490. <https://doi.org/10.1128/JB.184.9.2481-2490.2002>.
 39. Hall JA, Yan D. 2013. The molecular basis of K⁺ exclusion by the *Escherichia coli* ammonium channel AmtB. *J Biol Chem* 288:14080–14086. <https://doi.org/10.1074/jbc.M113.457952>.
 40. Paz-Yepes J, Herrero A, Flores E. 2007. The NtcA-regulated *amtB* gene is necessary for full methylammonium uptake activity in the cyanobacterium *Synechococcus elongatus*. *J Bacteriol* 189:7791–7798. <https://doi.org/10.1128/JB.00404-07>.
 41. Jeffrey W, Pledger R, Aas P, Hager S, Coffin R, Haven RV, Mitchell D. 1996. Diel and depth profiles of DNA photodamage in bacterioplankton exposed to ambient solar ultraviolet radiation. *Mar Ecol Prog Ser* 137:283–291. <https://doi.org/10.3354/meps137283>.
 42. Moss RH, Edmonds JA, Hibbard KA, Manning MR, Rose SK, van Vuuren DP, Carter TR, Emori S, Kainuma M, Kram T, Meehl GA, Mitchell JFB, Nakicenovic N, Riahi K, Smith SJ, Stouffer RJ, Thomson AM, Weyant JP, Wilbanks TJ. 2010. The next generation of scenarios for climate change research and assessment. *Nature* 463:747–756. <https://doi.org/10.1038/nature08823>.
 43. Taylor KE, Stouffer RJ, Meehl GA. 2012. An overview of CMIP5 and the experiment design. *Bull Am Meteorol Soc* 93:485–498. <https://doi.org/10.1175/BAMS-D-11-00094.1>.
 44. Yan W, Feng X, Zhang W, Zhang R, Jiao N. 2020. Research advances on ecotype and sub-ecotype differentiation of *Prochlorococcus* and its environmental adaptability. *Sci China Earth Sci* 63:1691–1700. <https://doi.org/10.1007/s11430-020-9651-0>.
 45. Thomas MK, Kremer CT, Klausmeier CA, Litchman E. 2012. A global pattern of thermal adaptation in marine phytoplankton. *Science* 338:1085–1088. <https://doi.org/10.1126/science.1224836>.
 46. Chandler JW, Lin Y, Gainer PJ, Post AF, Johnson ZI, Zinser ER. 2016. Variable but persistent coexistence of *Prochlorococcus* ecotypes along temperature gradients in the ocean's surface mixed layer. *Environ Microbiol Rep* 8:272–284. <https://doi.org/10.1111/1758-2229.12378>.
 47. Liu K, Suzuki K, Chen B, Liu H. 2021. Are temperature sensitivities of *Prochlorococcus* and *Synechococcus* impacted by nutrient availability in the subtropical northwest Pacific? *Limnol Oceanogr* 66:639–651. <https://doi.org/10.1002/lno.11629>.
 48. Hawco NJ, Fu F, Yang N, Hutchins DA, John SG. 2021. Independent iron and light limitation in a low-light-adapted *Prochlorococcus* from the deep chlorophyll maximum. *ISME J* 15:359–362. <https://doi.org/10.1038/s41396-020-00776-y>.
 49. Feng Y, Chai F, Wells ML, Liao Y, Li P, Cai T, Zhao T, Fu F, Hutchins DA. 2021. The combined effects of increased pCO₂ and warming on a coastal phytoplankton assemblage: from species composition to sinking rate. *Front Mar Sci* 8:622319. <https://doi.org/10.3389/fmars.2021.622319>.
 50. Moore LR, Coe A, Zinser ER, Saito MA, Sullivan MB, Lindell D, Frois-Moniz K, Waterbury J, Chisholm SW. 2007. Culturing the marine cyanobacterium *Prochlorococcus*. *Limnol Oceanogr Methods* 5:353–362. <https://doi.org/10.4319/lom.2007.5.353>.
 51. Chisholm SW, Frankel SL, Goericke R, Olson RJ, Palenik B, Waterbury JB, West-Johnsrud L, Zettler ER. 1992. *Prochlorococcus marinus* nov. gen. nov. sp.: an oxyphototrophic marine prokaryote containing divinyl chlorophyll a and b. *Arch Microbiol* 157:297–300. <https://doi.org/10.1007/BF00245165>.
 52. Yan W, Wei S, Wang Q, Xiao X, Zeng Q, Jiao N, Zhang R. 2018. Genome rearrangement shapes *Prochlorococcus* ecological adaptation. *Appl Environ Microbiol* 84:e01178-18. <https://doi.org/10.1128/AEM.01178-18>.
 53. Berube PM, Biller SJ, Kent AG, Berta-Thompson JW, Roggensack SE, Roache-Johnson KH, Ackerman M, Moore LR, Meisel JD, Sher D, Thompson LR, Campbell L, Martiny AC, Chisholm SW. 2015. Physiology and evolution of nitrate acquisition in *Prochlorococcus*. *ISME J* 9:1195–1207. <https://doi.org/10.1038/ismej.2014.211>.
 54. Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* 31:533–538. <https://doi.org/10.1038/nbt.2579>.
 55. Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with rearrangement. *PLoS One* 5:e11147. <https://doi.org/10.1371/journal.pone.0011147>.
 56. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RW, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75–15. <https://doi.org/10.1186/1471-2164-9-75>.
 57. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44:D457–D462. <https://doi.org/10.1093/nar/gkv1070>.
 58. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
 59. Stamatakis A. 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
 60. Laing C, Buchanan C, Taboada EN, Zhang Y, Kropinski A, Villegas A, Thomas JE, Gannon VPJ. 2010. Pan-genome sequence analysis using Panseq: an online tool for the rapid analysis of core and accessory genomic regions. *BMC Bioinformatics* 11:461–414. <https://doi.org/10.1186/1471-2105-11-461>.
 61. Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874. <https://doi.org/10.1093/molbev/msw054>.
 62. Chaudhari NM, Gupta VK, Dutta C. 2016. BPGA- an ultra-fast pan-genome analysis pipeline. *Sci Rep* 6:24373. <https://doi.org/10.1038/srep24373>.
 63. Csűös M. 2010. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* 26:1910–1912. <https://doi.org/10.1093/bioinformatics/btq315>.
 64. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, Cornejo-Castillo FM, Costea PI, Cruaud C, d'Ovidio F, Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C, Lima-Mendez G, Poulain J, Poulos BT, Royo-Llonch M, Sarmiento H, Vieira-Silva S, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Coordinators TO, Bowler C, de Vargas C, Gorsky G, Grimsley N, Hingamp P, Ludicone D, Jaillon O, Not F, Ogata H, Pesant S, Speich S, Stemann L, Sullivan MB, Weissenbach J, Wincker P, Karsenti E, Raes J, Acinas SG, Tara Oceans Coordinators, et al. 2015. Structure and function of the global ocean microbiome. *Science* 348:1261359. <https://doi.org/10.1126/science.1261359>.