The cyanobacterium *Prochlorococcus* has divergent light-harvesting antennae and may have evolved in a low-oxygen ocean

Osvaldo Ulloa,a,b,1 Carlos Henríquez-Castilla,a,b,2 Salvador Ramírez-Flandes,a,b,3 Alvaro M. Plominsky,a,b,c,3 Alejandro A. Murillo,a,b,4 Connor Morgan-Langd, Steven J. Hallamc,d,e,f,g, and Ramunas Stepanauskasb

aDepartamento de Oceanografía, Universidad de Concepción, 4070386 Concepción, Chile; bInstituto Milenio de Oceanografía, 4070386 Concepción, Chile; cDepartment of Microbiology and Immunology, University of British Columbia, Vancouver, BC V6T 1Z1, Canada; dGraduate Program in Bioinformatics, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; eGenome Science and Technology Program, University of British Columbia, Vancouver, BC V6T 124, Canada; fLife Sciences Institute, University of British Columbia, Vancouver, BC V6T 123, Canada; gECOSCOPE Training Program, University of British Columbia, Vancouver, BC V6T 1Z3, Canada; and hBigelow Laboratory for Ocean Sciences, East Boothbay, ME 04544-0380

Edited by Sallie W. Chisholm, Massachusetts Institute of Technology, Cambridge, MA, and approved February 4, 2021 (received for review December 28, 2020)

Marine picocyanobacteria of the genus *Prochlorococcus* are the most abundant photosynthetic organisms in the modern ocean, where they exert a profound influence on elemental cycling and energy flow. The use of transmembrane chlorophyll complexes instead of phycobilisomes as light-harvesting antennae is considered a defining attribute of *Prochlorococcus*. Its ecology and evolution are understood in terms of light, temperature, and nutrients. Here, we report single-cell genomic information on previously uncharacterized phylogenetic lineages of this genus from nutrient-rich anoxic waters of the eastern tropical North and South Pacific Ocean. The most basal lineages exhibit optical and genotypic properties of phycobilisome-containing cyanobacteria, indicating that the characteristic light-harvesting antenna of the group is not an ancestral attribute. Additionally, we found that all the Indigenous lineages analyzed encode genes for pigment biosynthesis under oxygen-limited conditions, a trait shared with other freshwater and coastal marine cyanobacteria. Our findings thus suggest that *Prochlorococcus* diverged from other cyanobacteria under low-oxygen conditions before transitioning from phycobilisomes to transmembrane chlorophyll complexes and may have contributed to the oxidation of the ancient ocean.

cyanobacteria | microbiology | genomics | oceanography | anoxia

The free-living, planktonic, unicellular cyanobacterium of the genus *Prochlorococcus* is the smallest and most abundant photosynthetic organism on Earth (1). Together with their closest phylogenetic relatives of the genus *Synechococcus*, they are collectively known as the marine picocyanobacteria (≤2 μm in cell diameter) and are significant contributors to primary production in the modern ocean (2). Distinguishing characteristics of *Prochlorococcus*, vis-a-vis other marine cyanobacteria, include an extremely small size (<1 μm in diameter) and the possession of transmembrane divinyl chlorophyll *a* and *b* complexes as their main photosynthetic light-harvesting antennae instead of the membrane-bound phycobilisomes that are present in most other cyanobacteria (1, 3). These features are readily discernible with flow cytometry and are widely utilized to selectively quantify *Prochlorococcus* in marine samples (1). The coding potential and diversity of *Prochlorococcus* have been studied in numerous phylogenetically distinct isolates and hundreds of single-cell–amplified genomes (SAGs) (4). Based on their physiology, genetic characteristics, and environmental niches, the genus is broadly divided into high-light (HL) and low-light (LL) adapted ecotypes (5). The current understanding is that light, temperature, and nutrients—including iron—play integral roles in the ecology and evolution of *Prochlorococcus* (6–10).

*Prochlorococcus* thrives in and is best known from sunlit and well-oxygenated, nutrient-poor, tropical and subtropical waters. However, *Prochlorococcus* is also the dominant photoautotroph in the nutrient-rich, oxygen-depleted waters of major anoxic marine zones (AMZs) in the Arabian Sea and the eastern tropical North (ETNP) and South Pacific (ETSP) (11, 12), where photosynthetic eukaryotes are not found. In these systems, *Prochlorococcus* can drive a cryptic oxygen cycle (13), by which aerobic metabolic processes can exist under apparent anoxic conditions due to the oxygen produced through oxygenic photosynthesis. Phylogenetic analysis of environmental sequences of the 16S to 23S ribosomal RNA (rRNA) internal transcribed spacer (ITS) region has revealed that the majority of the AMZ *Prochlorococcus* belong to two distinct phylogenetic lineages, originally termed LL V and LL VI (12) and renamed here as AMZ I and AMZ II. These lineages correspond to uncultured basal groups within the *Prochlorococcus* genus, for which there is no public genetic information except for their inferred capacity for nitrate utilization.

**Significance**

The marine unicellular cyanobacterium *Prochlorococcus* is the most abundant photosynthetic organism on Earth. Members of this genus are classically thought to be adapted to high-oxygen and nutrient-poor ocean conditions, with a principle divergence between high-light and low-light ecotypes. We show that the most basal *Prochlorococcus* lineages are adapted to the low-oxygen, low-light, and high-nutrient conditions found in the dimly illuminated waters of anoxic marine zones. The most basal lineages have retained phycobilisomes as light-harvesting antennae—a characteristic of most other cyanobacteria—whose loss was thought to define all *Prochlorococcus*. As oxygenic photosynthesis drove ocean oxidation in the ancient Earth, oxygen appears to have played as much a role as light and nutrients in driving *Prochlorococcus* evolution.


The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1To whom correspondence may be addressed. Email: oulloa@udec.cl.

2Present address: Laboratorio de Fisiología y Genética Marina, Centro de Estudios Avanzados en Zonas Áridas, 1781421 La Serena, Chile.

3Present address: Marine Biology Research Division, Scripps Institution of Oceanography, University of California, San Diego, CA 92093-0202.

4Present address: Centro de Estudios de Algas Nocivas, Instituto de Fomento Pesquero, 5480000 Puerto Montt, Chile.

*PNAS* 2021 Vol. 118 No. 11 e2025638118 https://doi.org/10.1073/pnas.2025638118 | 1 of 6
based on limited shotgun metagenomic sequencing data (14). Thus, further details on their coding potential and evidence for anaerobic or microaerobic metabolisms remain unconstrained for these AMZ lineages. Such information is important for understanding the origin and structure of the photosynthetic apparatus (15, 16), the evolution of marine pelagic picocyanobacteria (17), the functioning of ancient and modern AMZs (18), and the differential modes of metabolic coupling between Prochlorococcus and other AMZ microorganisms.

Results and Discussion

To explore the genetic makeup of these uncultured picocyanobacteria from AMZ waters, we analyzed 128 SAGs from water samples of the secondary fluorescence maximum in the eastern Pacific Ocean off northern Chile and Mexico (Fig. 1A, Table 1, and SI Appendix, Table S1). The geochemical characteristics of these AMZ waters (18) include undetectable levels of oxygen, no accumulation of sulfide or ammonium, the presence of significant levels of major nutrients, and, particularly of nitrite, a product of the anaerobic process of dissimilative nitrate reduction. Cells with optical characteristics typical of either Prochlorococcus (red fluorescence only) or Synechococcus (red and orange fluorescence) (Fig. 1B) were sorted for SAG generation and sequencing using previously described techniques (19). The AMZ SAG assemblies had an estimated completeness of up to 75%, an estimated genome size of ∼2.5 Mbp, and a guanine–cytosine content of ∼50% (SI Appendix, Table S1). These characteristics are more similar to those previously observed for the Prochlorococcus LL IV and marine Synechococcus lineages than for HL Prochlorococcus (6, 20).

Prochlorococcus SAGs affiliated with the uncharacterized AMZ I lineage were recovered with the traditional red-fluorescence-only cell sorting approach. However, SAGs affiliated with the previously uncharacterized AMZ II lineage were obtained only from cells that exhibited red and orange fluorescence, which is traditionally associated with Synechococcus. Moreover, we discovered a previously unknown Prochlorococcus lineage, termed here as AMZ III, among the cells with both red and orange fluorescence. According to phylogenetic analyses of individual and concatenated universal single-copy marker genes, AMZ I is a sister clade to LL IV and presents two subclades, termed here as AMZ IA and AMZ IB, with AMZ II occupying a basal position to the LL IV and AMZ I lineages (Fig. 1C and SI Appendix, Fig. S1). This tree topology is consistent with the established ITS phylogeny (8, 12). Furthermore, the newly discovered AMZ III lineage is located at the base of the Prochlorococcus radiation, thus representing the extant Prochlorococcus relatives closest to marine Synechococcus identified to date. The phylogenetic grouping of the AMZ Prochlorococcus SAGs was further corroborated by their genomic similarity (SI Appendix, Fig. S2).

An analysis of the individual SAGs (Fig. 2 and SI Appendix, Fig. S3 and Tables S2 and S3) revealed that all three AMZ lineages encode genes for chlorophyll b synthase pcOba (21) and lycopene ε-cyclase crtL (22) but lack the gene dvr encoding for the divinyl chlorophyll a reductase (23). This coding potential contributes to the characteristic pigment traits of all Prochlorococcus, such as divinyl chlorophyll a (instead of the monovinyl form present in most of the other cyanobacteria, including marine Synechococcus), divinyl chlorophyll b, and α-carotene (3, 24). This genomic information is consistent with the abundance of these diagnostic pigments in AMZ waters of the Arabian Sea and eastern North Pacific (11) as well as of the eastern South Pacific (SI Appendix, Table S4). However, the AMZ II and AMZ III SAGs were also found to contain genes encoding for complete synthesis of phycobilisomes, coinciding with the orange fluorescence of these two lineages (Fig. 1B). Specifically, AMZ II and AMZ III SAGs encoded genes for the synthesis of the phycobiliproteins allophycocyanin, phycocyanin, and phycoerythrin (I and II), as well as the corresponding linkers, and lack the chlorophyll light-harvesting pcb genes typical of Prochlorococcus (Fig. 2 and SI Appendix, Fig. S4 and Table S3). Furthermore, AMZ II and AMZ III SAGs encoded a chlorophyll-binding protein of the iron-stress–induced (IsiA) type, which is typical of marine Synechococcus and is phylogenetically distant from the Prochlorococcus-typical pcb genes (25), including those in AMZ I SAGs (SI Appendix, Fig. S5). Likewise, we found the kaiA gene in AMZ II and AMZ III (Fig. 2 and SI Appendix, Table S3), a critical component of the circadian clock KaiABC system (26).

Fig. 1. Representative biogeochemical and flow-cytometry profiles and phylogenetic tree of unicellular cyanobacteria with Prochlorococcus from AMZs. (A) Dissolved oxygen concentration (blue), in vivo chlorophyll a fluorescence in relative units (dark green), and Synechococcus-like (orange) and Prochlorococcus-like (light green) cell abundance at Station 6 in the AMZ of the ETNP (Table 1). The gray area corresponds to depths with <0.5 μmol kg⁻¹O₂. (B) Fluorescence versus light scatter cytograms (in relative units) of cyanobacteria-like particles in the secondary deep-chlorophyll maximum. The upper panel shows particles with red fluorescence, a proxy for chlorophyll a-containing cells, and the lower panel shows particles with orange fluorescence, a proxy for phycoerythrin-containing cells. (C) Maximum-likelihood tree determined by phylogenetic inference using 49 concatenated ribosomal proteins: Prochlorococcus (collapsed genomes of recognized ecotypes in green), Synechococcus (collapsed genomes of marine subclusters in orange), and AMZ Prochlorococcus (in bold). PCC 7002 is the marine onshore Synechococcus sp. strain PCC 7002, and PCC 6803 is the freshwater Synechocystis sp. strain PCC 6803. The circles on each node represent support values higher than 50% (n = 1,000 iterations). Characteristic traits of the different cyanobacteria lineages examined are given in colored lines. The dashed line indicates that only a few members of the Synechococcus subcluster have that trait.

Ulloa et al.
The cyanobacterium Prochlorococcus has divergent light-harvesting antennae and may have evolved in a low-oxygen ocean.
that has not been found in *Prochlorococcus* before (27, 28). Thus, our results indicate that the replacement of phycobilisomes for divinyl chlorophyll *a* and *b* complexes and the loss of the *kai*A gene occurred within the *Prochlorococcus* diversification and support previous observations regarding genetic remnants of such evolutionary events still being present in some *Prochlorococcus* lineages (28, 29). As shown above, these evolutionary steps appear to have occurred after *Prochlorococcus* acquired the capacity to synthesize divinyl chlorophyll *b* (Fig. 1C).

One of the major challenges modern cyanobacteria experience when living in anoxia and temporary darkness is the synthesis of tetrapyrrole pigments, as oxygen is required in several of their shared enzymatic steps. These biomolecules have important functions, including electron transfer, oxygen binding, and light absorption, with chlorophyll *a* being essential for photosynthesis (30). Certain cyanobacteria, however, have alternative or complementary enzymes for pigment biosynthesis that work under oxygen-limited conditions (31). In particular, three genes, *hemN*2, *acsF*2, and *ho*2, encode secondary versions of enzymes involved in heme, chlorophyll, and phycocyanobilin biosynthesis, respectively. The gene *hemN*2 encodes the oxygen-independent version of the enzyme that catalyzes the biosynthesis of protoporphyrinogen IX, the common precursor of Chl-*a*, hemes, and bilins; *acsF*2 encodes the cyclase that converts Mg-Proto monomethyl ester to divinyl prototchlorophyllide during the synthesis of Chl-*a*; and *ho*2 encodes a heme oxygenase involved in bilin synthesis. These genes are clustered together in the genomes of the freshwater cyanobacterium *Synechocystis* PCC 6803 and the euryhaline coastal marine cyanobacterium *Synechococcus* PCC 7002. In both isolates, such genes are highly expressed under oxygen-limited conditions but not under fully oxic conditions (32, 33). We identified the gene cluster *acsF*2-*ho*2-*hemN*2 in SAGs of the three AMZ lineages and—with a similar genomic configuration—also in marine *Synechococcus* sp. strains RS9917, WH 8101, and CB0101 (Figs. 2 and 3A). This gene cluster has not been previously reported in any of the prior studies of *Prochlorococcus* and marine *Synechococcus* (31). Phylogenetic analyses revealed that the gene encoding the *AcsF*2 cyclase branches closest to homologs in the LL *Prochlorococcus* and *Synechococcus* strains RS9917, WH 8101, and CB0101 but not with homologs in other marine *Synechococcus* nor with those in HL *Prochlorococcus* ecotypes (Fig. 3B). The cluster-forming *ho*2 and *hemN*2 genes branch in a similar way as *acsF*2, except that no homologs from LL *Prochlorococcus* are grouped closely with them (Fig. 3 C and D).

The phylogeny and conserved gene arrangement shared between freshwater and marine cyanobacteria suggests that *Prochlorococcus* may have evolved before the ocean became fully oxygenated. In this scenario, *ho*2 and *hemN*2 would have been lost as *Prochlorococcus* adapted to anoxic ocean, but *acsF*2 would have been retained in the LL lineages (Fig. 3A and B). The alternative scenario that the low-oxygen pigment biosynthesis trait was acquired after *Prochlorococcus* evolved in anoxic ocean cannot be totally discarded. However, a detailed analysis of contigs containing the genes for pigment biosynthesis (as well as those for the light-harvesting antenna) using previously described approaches (14, 34, 35) did not provide any evidence of recent horizontal gene transfer or of hallmark viral signatures. If *Prochlorococcus* evolved in an oxygen-depleted ocean before it could adapt to fully oxygenated oligotrophic surface waters, the replacement of phycobilisomes for transmembrane divinyl chlorophyll complexes may have been driven by the low levels and narrow spectral characteristics of the light available at depth (36). Moreover, continuous changes in the light field at depth due to variable optical conditions of surface waters or due to internal waves could have allowed for the coexistence of lineages with different light-harvesting antennae, as is currently observed. However, in the proposed scenario, only those *Prochlorococcus* with chlorophyll complexes would have been able to colonize the high-oxygen surface waters.

**Table 1.** Sampling location and hydrographic characteristics of waters from which samples for single-cell genomics were collected

<table>
<thead>
<tr>
<th>Cruise</th>
<th>Station</th>
<th>Latitude (°N)</th>
<th>Longitude (°W)</th>
<th>Date (dd/mm/yy)</th>
<th>Depth (m)</th>
<th>Temperature (°C)</th>
<th>Salinity (μmol/kg)</th>
<th>Phosphate (μmol/kg)</th>
<th>Nitrate + nitrite (μmol/kg)</th>
<th>Chl-α (μg/L)</th>
<th>Prochlorococcus-like (10⁴ cells/mL)</th>
<th>Synechococcus-like (10⁴ cells/mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MV1015</td>
<td>1</td>
<td>−20,083</td>
<td>70,800</td>
<td>19/11/10</td>
<td>53</td>
<td>12.874</td>
<td>34.731</td>
<td>3.3</td>
<td>2.72*</td>
<td>17.85*</td>
<td>1.04</td>
<td>0.37</td>
</tr>
<tr>
<td>NH1315</td>
<td>6</td>
<td>18,920</td>
<td>104,891</td>
<td>19/07/13</td>
<td>100</td>
<td>13.995</td>
<td>34.781</td>
<td>1.5</td>
<td>2.52</td>
<td>20.44</td>
<td>60.8</td>
<td>4.3</td>
</tr>
</tbody>
</table>

*Values from 50 m in separate cast.

**Fig. 2.** Presence–absence of selected genes in AMZ *Prochlorococcus* SAGs (see SI Appendix, Tables S2 and S3 for details).
Metagenomic fragment recruitment (SI Appendix, Fig. S7) showed that AMZ I, II, and III lineages are found only in oxygen-depleted oceanic waters. Moreover, AMZ I outnumbers II and III, which is consistent with the flow cytometry data (11, 12) indicating that red-only fluorescent cells dominate in AMZ waters. The presence of indigenous lineages with a specific gene cluster for pigment synthesis under oxygen-limited conditions (Fig. 3) is thus consistent with the designation of Prochlorococcus “AMZ” ecotypes. Moreover, these ecotypes, which inhabit nutrient-rich waters, have a complete gene repertoire for the use of ammonium, urea, nitrite, and nitrate (Fig. 2 and SI Appendix, Table S3) as nitrogen sources. On the other hand, according to the analysis of public metatranscriptomic data (SI Appendix, Fig. S8), at least some of the photosynthesis genes discussed above are being actively transcribed in AMZs. However, targeted gene expression studies over the light–dark periods will be required to address this issue fully.

Other genotypic properties that appear to be specific to AMZ ecotypes and that may confer them adaptations to an anoxic environment include the presence of a hybrid-cluster protein (hcp) gene and cognate fnr regulator of the anaerobic Cpr/Fnr family as well as a gene ytfE encoding an iron-sulfur cluster repair protein (37). Additionally, AMZ Prochlorococcus SAGs encode enzymes involved in glycerol oxidation (glpK and glpD) and alcohol fermentation (adh) (Fig. 2). These additional features indicate that AMZ Prochlorococcus may be capable of supplementing their carbon and energy requirements through the use of organic compounds, as well as a putative transcriptional regulator, and ytfE encodes an unknown protein. Maximum-likelihood phylogenetic trees of (B) acsF, (C) ho, and (D) hemN-predicted amino acid sequences. The colored boxes mark the gene versions that appear organized in a cluster and should work under oxygen-limited conditions; lineages appearing in A are in bold. Conserved residues for several functions are given in SI Appendix, Fig. S6. Filled circles on each node correspond to support values higher than 50% (n = 1,000 iterations).

Fig. 3. Synteny and phylogeny of three genes involved in tetrapyrrole biosynthesis under oxygen-limited conditions. (A) Synteny of three orthologous genes involved in chlorophyll (acsF2), phycocyanobilin (ho2), and heme (hemN2) biosynthesis across selected cyanobacterial lineages. The gene arsR encodes a putative transcriptional regulator, and ytfE encodes an unknown protein. Maximum-likelihood phylogenetic trees of (B) acsF, (C) ho, and (D) hemN-predicted amino acid sequences. The colored boxes mark the gene versions that appear organized in a cluster and should work under oxygen-limited conditions; lineages appearing in A are in bold. Conserved residues for several functions are given in SI Appendix, Fig. S6. Filled circles on each node correspond to support values higher than 50% (n = 1,000 iterations).
previously suggested for other lineages (39,40), to survive under dark anoxic conditions.

Marine planktonic picocyanobacteria likely emerged in the Neoproterozoic 1,000 to 542 Mya (17), roughly at the time of the oxygenation of the deep ocean (41). We have shown that extant lineages inhabiting AMZs are basal to the Prochlorococcus genus, can phycobilisomes as light-harvesting antenna, and carry ancestral adaptations to low-oxygen conditions. These findings suggest that Prochlorococcus emerged in low-oxygen environments and may have contributed to sustaining aerobic metabolisms and providing a significant carbon source through oxygenic photosynthesis, in the ancient, anoxic upper ocean, as they do today (13). The evolutionary history of Prochlorococcus thus appears intimately related to the availability of oxygen in the ocean.

Materials and Methods

Sampling and Hydrographic Data Collection. Sampling was carried out during cruise MV1015 (DOI: 10.7284/9303808), the “Biogeochemical Gradients: Role in Arranging Planktonic Assemblages (BIG RAPA) Expedition” (https://hahana.soest.hawaii.edu/moreserver/cruises/big_rapa平原.html), at Station 1 in the ETSP and during cruise NH1315 (DOI: 10.7284/9308369) at Station 1 in the ETNP (Table 1). For Cruise MV1015, hydrographic data and seawater samples were collected with a rosette/conductivity-temperature-depth (CTD) system, consisting of Niskin bottles and a CTD profiler (Sea-Bird SBE 911plus) with dissolved-oxygen, fluorometer, and transmissometer sensors. For Cruise NH1315, hydrographic data and seawater samples were collected with a pump profiling system, similar to the one previously used in AMZ waters (13).

Prochlorococcus Cell Count by Flow Cytometry. For each site, duplicate samples of 1 mL raw seawater were fixed in a final concentration of 1% (vol/vol) glutaraldehyde at room temperature, in the dark, for 10 min and then fast frozen in liquid nitrogen and stored at −80 °C until analysis. In the laboratory, samples were thawed on ice and passed through a 60-μm mesh net, and cell abundance was determined with a FACSCalibur flow cytometer (Becton Dickinson) by using red and orange fluorescence as proxies for cells containing chlorophyll a and phycobiliproteins, respectively (42).

AMZ Prochlorococcus Single-Cell Sample Collection. For the MV1015 (ETSP) cruise, water samples for single-cell sorting were collected from the Niskin bottles and prefilttered through a 60-μm mesh-size net. Then, ~4,000 cells of Prochlorococcus-like cells were sorted on board into 1 mL of sterile glycerol-Tris-EDTA (TE) buffer (43) with an Influx flow cytometer (BD Biosciences), followed by single-cell PCR amplification of the 16S rRNA gene (Table 1). For Cruise MV1015, hydrographic data and seawater samples were collected with a pump profiling system, similar to the one previously used in AMZ waters (13).

SAG Sequencing Assembly, Gene Prediction, and Annotation. Single-cell sorting, whole-genome amplification, 16S rRNA real-time PCR screening, and PCR product sequencing were performed at the Bigelow Laboratory for Ocean Sciences’ Single Cell Genomics Center (SCGC, http://scgc.bigelow.org/), as described previously (19). For the ETSP SAGs, an Illumina standard shotgun library was constructed, and samples were sequenced utilizing the HiSeq 2000 platform at Canada’s Michael Smith Genome Sciences Centre. Artificial sequences were filtered off the raw data and draft SAGs were assembled using SPAdes 3.5. SAG contigs of >1 kb were decontaminated using ProDeGe (44), and sequences with lengths between >200 bp and <2 kb were considered whenever they had >80% identity (>75% alignment coverage) with Prochlorococcus reference genomes (SI Appendix, Table S1). Quality filtered sequences of <200 bp were discarded. The decontaminated SAG sequences were uploaded for genefinding and annotation to the Integrated Microbial Genomes & Microbiomes (IMG/MER) system (45). Genome completeness was estimated by the presence of 114 single-copy conserved bacterial phylogenetic markers among the predicted protein-coding genes of each SAG (46). The ETNP SAGs were sequenced at SCGC as described previously (19). Gene predictions, functional annotation, manual curation, and pathway reconstruction for all individual SAGs were performed using IMG/MER platform (img.mer.upstate.edu) (19), supported by grants from the National Agency for Research and Development of Chile (Grants IC12_019-MA, 1130784, 1161483, and 316020724) and the Agouron Institute to O.U. Support for this research was also provided by the US NSF (Grants OCE-1335810 and OIA-1826734 to R.S.) and the Tula Foundation, the G. Unger Veteslen and Ambrose Monell Foundations, the Natural Sciences and Engineering Research council of Canada, and Compute/Canada to S.J.H. A.M.P. received postdoctoral support through the Tula Foundation.

Ulloa et al.

The cyanobacterium Prochlorococcus has divergent light-harvesting antennae and may have evolved in a low-oxygen ocean.

Retrieval of AMZ SAGs Phylogenetic Marker Sequences. Prochlorococcus ITS sequences were retrieved from reference genomes (SI Appendix, Table S1) and utilized as BLASTn queries to retrieve those from the Prochlorococcus AMZ SAGs. Then, transfer RNA genes were removed and the identity was determined for all ITS sequences that ended with >600 bp. The same approach was adopted with the rbcL gene, encoding the Rubisco large subunit and the petB gene, a component of the cytochrome b6f complex. Average nucleotide identity (ANI) analysis (SI Appendix, Fig. S2) was performed using FastANI (47).

Identification and Conserved Residues of Cluster-Forming Terapryrole Biosynthesis Genes. The gene cluster acaF2-hox-Hmn2, associated with pigment biosynthesis under oxygen-limited conditions (31), was identified in individual Prochlorococcus AMZ SAGs within IMG/MER. Conserved residues in AcaF2, HoxO2, and HmnN2 (SI Appendix, Fig. S5) were visually detected and manually annotated according to information provided in the literature (48–50). Though alignments were also analyzed with PROSITE (https://prosite.expasy.org), no additional relevant domains were found.

Phylogenetic and Phylogenomic Analyses. Predicted amino acid sequences were aligned using MAFFT (51) version 7.271. Maximum-likelihood phylogenetic and phylogenomic inferences were performed using IQ-TREE (52) version 1.6.12. The models of sequence evolution were selected with ModelFinder (53) according to the Bayesian information criterion. The best fit models for single-gene phylogenies are given in the respective figure legends. Branch support values were computed using the Shimodaira–Hasegawa approximate likelihood ratio test within IQ-TREE. For phylogenomic reconstructions, concatenated amino acid-aligned sequences of 49 ribosomal proteins (Fig. 1C and 17 single-copy genes (SI Appendix, Fig. S1C) were used (54). The best fit model for the tree in Fig. 1C was LG+R10, and LG+I+G4 was the best fit model for the tree in Fig. S1 Appendix, Fig. S6).

Fragment Recruitment. The recruitment of sequence fragments from global ocean metagenomes (SI Appendix, Table S5) was performed against the contigs associated with the AMZ Prochlorococcus ecotypes obtained in the ETNP and ETSP. Fragment recruitment onto Prochlorococcus LL genomes was included for comparison (SI Appendix, Fig. S7). Details about the procedure can be found in SI Appendix, Table S8.

Transcript Recruitment. The recruitment of transcripts encoding pigment-biosynthesis and photosynthetic proteins from public AMZ metatranscriptomes (SI Appendix, Table S5) was carried out against the AMZ SAGs and Prochlorococcus marine Synechococcus genomes (SI Appendix, Fig. S8). The details about the procedure can be found in SI Appendix, Table S9.


ACKNOWLEDGMENTS. We thank the chief scientists Dan Repeta (MV1015) and Frank J. Stewart (NH1315) and the captains, crews, and scientific support personnel of the R/V Melville and the R/V New Horizon. We also thank Gadiel Alarcon andMontserrat Aldaune for assistance in sample collection, Maria Lorena Gonzalez for cell sorting during the BIG RAPA Expedition, Cristian Venegas for assistance in flow cytometry analysis, Marcela Montoya for help with the SAG analyses, and the SCGC staff. This work was supported by grants from the National Agency for Research and Development of Chile (Grants IC12_019-MA, 1130784, 1161483, and 316020724) and the Agouron Institute to O.U. Support for this research was also provided by the US NSF (Grants OCE-1335810 and OIA-1826734 to R.S.) and the Tula Foundation, the G. Unger Veteslen and Ambrose Monell Foundations, the Natural Sciences and Engineering Research Council of Canada, and Compute/Canada to S.J.H. A.M.P. received postdoctoral support through the Tula Foundation.


