Title
Characterization of Prochlorococcus clades from iron-depleted oceanic regions

Permalink
https://escholarship.org/uc/item/43w9h08g

Journal
Proceedings of the National Academy of Sciences, 107(37)

ISSN
0027-8424 1091-6490

Authors
Rusch, D. B
Martiny, A. C
Dupont, C. L
et al.

Publication Date
2010-08-23

DOI
10.1073/pnas.1009513107

Supplemental Material
https://escholarship.org/uc/item/43w9h08g#supplemental

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed
Characterization of Prochlorococcus clades from iron-depleted oceanic regions

Douglas B. Rusch a, Adam C. Martiny b,c, Christopher L. Dupont d, Aaron L. Halpern b,1, and J. Craig Venter d,2

a J. Craig Venter Institute, Rockville, MD 20855; Departments of b Earth System Science and Ecology and c Evolutionary Biology, University of California, Irvine, CA 92697; and d J. Craig Venter Institute, San Diego, CA 92121

Contributed by J. Craig Venter, July 29, 2010 (sent for review February 21, 2010)

Prochlorococcus describes a diverse and abundant genus of marine photosynthetic microbes. It is primarily found in oligotrophic waters across the globe and plays a crucial role in energy and nutrient cycling in the ocean ecosystem. The abundance, global distribution, and availability of isolates make Prochlorococcus a model system for understanding marine microbial diversity and biogeochemical cycling. Analysis of 73 metagenomic samples from the Global Ocean Sampling expedition acquired in the Atlantic, Pacific, and Indian Oceans revealed the presence of two uncharacterized Prochlorococcus clades. A phylogenetic analysis using six different genetic markers places the clades close to known lineages adapted to high-light environments. The two uncharacterized clades consist of Synechococcus and dominate the surface waters of high-temperature, macronutrient-replete, and low-iron regions of the Eastern Equatorial Pacific upwelling and the tropical Indian Ocean. They are genetically distinct from each other and other high-light Prochlorococcus isolates and likely define a previously unrecognized ecotype. Our detailed genomic analysis indicates that these clades comprise organisms that are adapted to iron-depleted environments by reducing their iron quota through the loss of several iron-containing proteins that likely function as electron sinks in the photosynthetic pathway in other Prochlorococcus clades from high-light environments. The presence and inferred physiology of these clades may explain why Prochlorococcus populations from iron-depleted regions do not respond to iron fertilization experiments and further expand our understanding of how phytoplankton adapt to variations in nutrient availability in the ocean.

One of the most abundant photoautotrophs in tropical and subtropical oligotrophic marine environments is Prochlorococcus (1). More than a dozen genetically diverse isolates have been cultivated and sequenced since Prochlorococcus was first discovered (2–4). These isolates can be subdivided into taxonomically and ecologically distinct lineages according to their adaptation to general environmental conditions [e.g., high-light (HL) vs. low-light (LL) (5)] or temperature regimes (6). Across the globe, warm surface waters are dominated by a single clade of Prochlorococcus (eMIT9312) (7, 8). This clade possesses high rates of sequence divergence while maintaining remarkable conservation both in gene content and synteny, despite a global distribution that spans large gradients in the bioavailable concentrations of the macronutrients nitrogen (N) and phosphorus (P).

However, evidence is mounting that adaptation of Prochlorococcus to local nutrient conditions is often mediated by gene cassettes that enable transport and assimilation of the available nutrients (4, 9, 10). Although these cassettes impart a strong selective advantage under the appropriate conditions, they have not been associated with phylogenetically distinct clades (9). This suggests that horizontal gene transfer and strong selective pressures to maintain a minimal genome allows cells to acquire or discard genes as required by local conditions and enables Prochlorococcus to proliferate under a wide range of environmental conditions (3, 11–14).

It is likely that our picture of the genetic diversity of the Prochlorococcus community is incomplete. For one, the number of Prochlorococcus cells counted by flow cytometry often exceeds that detected with PCR using primers based on our current knowledge of genetic variation (15, 16). In addition, most sequencing surveys have focused on Mediterranean (17), Atlantic (18–20), and the Hawaii Ocean time-series program sites (21), leaving the possibility that undescribed clades exist in other regions of the ocean.

To address the possibility that there are uncharacterized Prochlorococcus phylogenotypes, we queried samples from the Global Ocean Sampling (GOS) expedition. GOS represents the largest metagenomic study to date and includes samples collected from surface waters from the Atlantic, Pacific, and Indian Oceans during a 3-y circumnavigation of the world. In these samples, we discovered two previously undescribed Prochlorococcus clades in high-nutrient, low-chlorophyll (HNLC), iron-depleted waters in the Pacific and Indian Oceans. Reconstruction of consensus genomes and comparative approaches revealed that the gene repertoires of these clades reflect a streamlining of their iron requirements. These results are consistent with the presence of a HL Prochlorococcus ecotype adapted to low iron bioavailability.

Results

Identification of Prochlorococcus Diversity. The first evidence for uncharacterized Prochlorococcus lineages was observed as a distinctive pattern of fragment recruitment (Methods) with substantial divergence from HL Prochlorococcus genomes arising from a small number of GOS samples. To further investigate this pattern of Prochlorococcus diversity, we developed phylogenetic markers for Prochlorococcus suitable for analyzing metagenomic libraries. To find markers that are not prone to lateral gene transfer and able to distinguish between closely related phylogenotypes in metagenomic libraries, we looked for markers that satisfied five criteria: (i) present as a single copy in all Prochlorococcus and marine Synechococcus genomes, (ii) similar phylogeny as the 13 sequenced Prochlorococcus genomes, (iii) high rates of sequence divergence, (iv) variation consisting of substitutions rather than insertions and deletions, and (v) lengths <800 bp so that the entire gene could be identified within a single GOS sequencing read. We identified six genetic markers (psbT, psl27, purC, nadD, mstA, and PMSR) adhering to these criteria. These genetic markers could resolve most nodes including branch points between strains within the HL-adapted eMIT9312 clade. The only exception was the branching point of LL strains SS120 and MIT9211, which varied among markers. However, other studies have shown that this branch point is difficult to resolve, and the position is still debated.

Author contributions: D.B.R. designed research; D.B.R., A.C.M., and C.L.D. performed research; J.C.V. contributed new reagents/analytic tools; D.B.R., A.C.M., C.L.D., and A.L.H. analyzed data; and D.B.R., A.C.M., and C.L.D. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: This Whole Genome Shotgun project has been deposited at DNA Data Base in Japan/European Molecular Biology Laboratory/GenBank under accession no. ADHP00000000. The version described in this article is the first version, ADHP00000000.

1Present address: Complete Genomics, Inc., Mountainview, CA 94043.
2To whom correspondence should be addressed. E-mail: jcventer@jcvi.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1009513107/-/DCSupplemental.
Using these genetic markers, we screened the unassembled GOS metagenomic libraries (Table S1). All genetic markers revealed an undescribed Prochlorococcus clade branching between the LL eNATL and the HL ecotypes (Fig. 1 and Fig. S1). This topology was highly supported by both neighbor-joining and maximum likelihood phylogenetic approaches. The branch point is much closer to the HL clades than eNATL. The clade can be further separated with some confidence into two distinct subclades.

A detailed analysis of the geographic distribution of the newly identified clades shows that they were associated with the Eastern Equatorial Pacific and tropical Indian Ocean (Fig. 2), with relative abundances comparable to eMIT9312 in other GOS samples (Table S1). We then looked at the distribution of different Prochlorococcus clades in the context of various physicochemical parameters, such as light, temperature, nitrate, and iron, to determine which parameters correlate with the distribution of these newly detected clades (Fig. S2 A–D and Table S2). Iron concentrations were simulated using the ocean biogeochemical element cycling (BEC) model because we were not properly equipped to measure trace metal values (22). Nitrate and phosphate were monthly averages derived from the World Ocean Atlas (23). Because all GOS samples were from surface waters, only the HL ecotypes eMED4, eMIT9312, and the uncharacterized clades were present in noteworthy numbers according to fragment recruitment rates; thus, we restricted the analysis to these lineages. Previous studies have demonstrated that the distribution of eMED4 and eMIT9312 can be distinguished according to temperature (6). Supporting this, we only detected eMED4 at sites with a temperature <21 °C (Fig. S2A). In contrast, both eMIT9312 and the uncharacterized clades are detected in warmer waters and, in particular, the uncharacterized clades are only seen in samples with a temperature above 25 °C. However, the distribution of eMIT9312 and the uncharacterized clades is very different across nitrate concentrations (Fig. S2B). eMIT9312 is found in samples with <1 μM nitrate, whereas the uncharacterized clades are detected in high-nitrate waters. Despite the high N and P concentration, only modest chlorophyll concentrations are observed in these sites; such oceanic regions are historically denoted HLNC. Thus, we denoted the uncharacterized clades HLNC1 and HLNC2.

Most of the HLNC regions are depleted in dissolved iron relative to other nutrients, specifically N or P, resulting in a unique nutrient limitation scenario (24). Thus, we speculated that the distribution of the HLNC clades is linked to iron availability. On the basis of the simulated iron values, the HLNC clades are mainly found in waters with <0.5 nm dissolved iron, whereas eMIT9312 and eMED4 are found in regions with higher iron concentrations (Fig. S2C). Two samples from the tropical Indian Ocean (GS111 and 112) dominated by HLNC clades presumably had a higher concentration of iron. However, satellite measurements of the nonphotochemical quenching-corrected fluorescence quantum yields suggest that the phytoplankton found in this region are indeed iron-stressed (Fig. S2D) (25). Thus, the BEC model may overestimate the iron concentration in the sparsely studied Indian Ocean samples. The HLNC clades are missing from Eastern Pacific samples GS23 and GS25, where the model indicates that iron is low and N is high. However, the model may underestimate the concentration because of coastal influences on iron availability. Using canonical correspondence analysis to summarize the distribution pattern, we observe that the primary habitat of the HLNC clades is low-iron, high-nitrate, and higher-temperature waters, with iron being the predominant factor distinguishing these strains from the HL-adapted eMIT9312 (Fig. 3). However, other unaccounted-for factors may also influence their distribution.

A recent study produced several metagenomic datasets from the central Pacific Ocean, including both the N-limited North Pacific and the iron-limited Equatorial Pacific Ocean (26, 27). On the basis of the best BLASTX hits to known Prochlorococcus genomes, these samples were reported to contain an unusually high abundance of the cold water–adapted eMED4 ecotype, despite very warm water temperatures (>27 °C). In light of this information, we reexamined these stations and found that the majority of previously designated eMED4 and eMIT9312 sequences instead grouped with the HLNC clades (Fig. 2 and Table S1).

**Genome Organization of HLNC Lineages.** For a more detailed analysis of the genome content of HLNC1 and HLNC2 cells, we assembled a consensus genome for each lineage using the Celera Assembler (27). These consensus genomes represent an ensemble view of the genome organization of populations and are not equivalent to a complete or even draft genome assemblies from a clonal isolate. The resulting assembly required minor manual curation to break two chimeric scaffolds and ultimately produced a set of 24 scaffolds corresponding to two HLNC clades. The assemblies had an average depth of coverage of 52 and 21 reads, respectively, for the HLNC1 and HLNC2. Average percentage GC for all scaffolds is ~30%. The individual scaffolds were separated into HLNC1 and HLNC2 bins according to differences in depth of coverage across sampling sites (Fig. S3). The scaffolds were ordered and oriented so that they best matched the layout found in the closest relatives, eMED4 and eMIT9312, both of which are highly syntenic with each other. This resulted in consensus genomes for HLNC1 and HLNC2 of 1.56 Mb.
and 1.48 Mb, respectively. The lengths of the assemblies are roughly equivalent in size to other HL Prochlorococcus genomes when one allows for the presence of hypervariable genomic islands that we would not expect to assemble (12, 13). Synteny within each scaffold is highly consistent with the eMIT9312 and eMED4 sequenced genomes (Fig. S4). We also observe that many of the break points are syntenic with the locations of hypervariable and variable islands. Thus, we conclude that the fragmentary nature of the assembly is largely the result of variation present within the HNLC lineages. Overall these results suggest that isolates from these clades, should they become available, will have genomes that are structurally highly syntenic to that of other HL isolates.

Genomic Adaptation. To elucidate whether and how these HNLC clades might be physiologically distinct, we identified genes that seem to be gained or lost. For this comparative analysis, we checked for the presence of individual genes both in the assemblies and directly in the environmental data (to circumvent potential assembly errors; Dataset S1). Given that the abundance of HNLC clades is strongly associated with predicted low iron availability (Fig. 3), we first examined genes related to iron uptake and stress response. Like most other HL Prochlorococcus strains, the HNLC clades seem to possess the same genes for iron acquisition and response to iron stress (28). Like other HL Prochlorococcus, the HNLC clades seem to lack TonB-like receptors and a siderophore biosynthesis pathway, so there is no evidence that the HNLC lineages possess a high-affinity siderophore-mediated iron uptake system.

Faced with the absence of any new systems for acquiring iron, we instead find at least six iron-containing proteins that seem to be both statistically underrepresented in samples dominated by HNLC1 and HNLC2 and missing in the genome assemblies (Table 1). This is a high number, considering that cyanobacterial genomes of similar size—including those of HL Prochlorococcus—contain between 60 and 65 iron-binding protein domains (29). Thus, the HNLC clade has the lowest number of iron-binding proteins of any cyanobacteria (Fig. S5). Taken together, these results suggest that the HNLC clades have adapted to low-iron environments not by becoming more efficient at acquiring iron but by eliminating a number of their iron-requiring proteins.

The lost iron-containing proteins include plastoquinol terminal oxidase (PTOX), two ferredoxins, and cytochrome Cm (Table 1). Each of these losses would remove a sink for electrons within the photosynthetic pathway. PTOX is hypothesized to catalyze cyclic electron flow around photosystem (PS) II in Synechococcus and Ostreococcus (30, 31), providing an outlet for electron flow when acceptors downstream of PSII are already reduced or unavailable. In both of these organisms, this “electron safety valve” allows for a high ratio of PSII to PSI by alleviating the potential for photodamage with increases in light. PTOX-like activity has also been detected in phytoplankton assemblages from the predominantly P-limited Sargasso Sea and N-limited North Pacific (32). However, it was not detected at the subsurface chlorophyll maximum, consistent with the lack of PTOX in the genomes of the LL Prochlorococcus strains (33). Recently, it has been suggested that the usage of PTOX may be an adaptation to chronic iron limitation (34), although the usage of PTOX in response to iron limitation has not been examined. Although PTOX does contain a paucity of Fe (2 molecules per protein) relative to the cytochrome B_{6f} complex and PSI (18 molecules total), the true determinants for metal requirements, the relative protein concentrations, and protein turnover times are not known. Instead, the biogeography of PTOX in Prochlorococcus genomes suggests that, at least in this genus, PTOX usage may be more related to N, P, or light stress than to adaptation to iron limitation.

Although PTOX is not found in the consensus genomes, it is found in a short 10-kb scaffold from the assembly. The distribution of reads associated with this scaffold is different from the HNLC consensus assemblies, having an association with only a subset of GOS samples in the South Pacific Gyre (Table S1). The percentage GC for this scaffold is 25%, which is significantly lower than the mean of 32.5% for similarly sized subsequences from the consensus HNLC genomes (SD 1.27%). The scaffold contains a number of HL-inducible genes, a phage-associated gene, and a PSI gene (psaC). Potentially, this PTOX-containing scaffold belongs to a Prochlorococcus phage, a possibility supported by the observation of PTOX in other cyanophages (35–38). Alternatively
this scaffold may be part of a hypervariable region found within a subset of the HNLC populations (Table S1).

The loss of other iron-containing proteins removes more potential outlets from the photosynthetic electron chain. Cytochrome C_m is thought to be involved in electron transfer from PSI to the cytochrome c oxidase complex during periods of stress (39). Although there are some clear exceptions, it is likely that the loss of two ferredoxins removes several pathways for electrons to move away from or to photosynthetic reaction centers. Finally, the HNLC genomes do not contain genes for the iron-containing enzyme nitrate reductase (10), which is consistent with a physiological strategy of iron requirement minimization. This adaptation removes a potential physiological sink for reducing equivalents and indicates that this organism does not use the abundant compound in the exogenous N pool.

In summary, the HNLC clade has lost proteins implicated in iron-requiring processes and therefore may be less able to cope with environmental changes, such as shifting light levels. We hypothesize that physiologically the loss of PTOX almost certainly prevents HNLC1 and HNLC2 from maintaining a high ratio of PSII to PSI in the HL environment of the tropical surface ocean. Given the low bioavailability of iron, HNLC1 and HNLC2 may maintain low absolute levels of PSII and PSI at a near even ratio, which would place a substantial limit on the maximal growth rate of these organisms and increase susceptibility to shifts in light intensity. Given that the growth of all other photoautotrophs in these systems is severely limited by iron, such tradeoffs may be warranted yet provide a ready explanation for the lack of the HNLC clades in other ocean regions.

**Genomic Diversity in the HNLC Clade.** Although there are some clear cases of genes being absent in the HNLC lineages, there are a number of genes that are present at lower or higher abundances than we would anticipate if the genes were present in every cell. These examples probably indicate further diversity within the accessory proteins of HNLC1 and HNLC2 lineages. One example includes the urease and urea transporter operons that include genes for acquiring and metabolizing urea. Two distinct urease operons were assembled from the metagenomic data and could be tentatively assigned to the HNLC1 and HNLC2 consensus genomes on the basis of their abundances. Both operons occur at significantly lower-than-expected coverage, which averages 24.2 and 8.6 reads per base, respectively. Numerous other genes are less abundant than would be expected. Some of these genes will be part of the accessory genome and thus present in only a subset of each lineage. Others may be present in only one of the two lineages. In total, 45 HNLC1 and 20 HNLC2 genes were found in only one of the consensus genomes, suggesting that they may be present in only one of the lineages (Dataset S1), although this is likely an overestimate. Phage-associated genes may be responsible for some of the genes that are less abundant than expected, or they may lead to higher-than-expected frequencies for genes that are also part of the core genome. We also identified 225 clusters of potentially novel peptides that are abundant only in the waters containing the HNLC clades and that could be linked to the HNLC assemblies (Dataset S2). All of these peptides were uncharacterized, so their functions and potential roles are unknown.

**Discussion**

**Biogeochemical Considerations.** The Equatorial Pacific Upwelling Region sampled here has been shown to have extremely low available iron (≤0.1 nM) and is thought to limit phytoplankton growth (40, 41). Part of the Indian Ocean gyre may also be iron

### Table 1. Selected genes from Prochlorococcus strain AS9601 that are present or absent from the HNLC consensus genomes and metagenomic samples

<table>
<thead>
<tr>
<th>GI no.*</th>
<th>Description</th>
<th>Counts in region A</th>
<th>Counts in region B</th>
<th>Counts in region C</th>
<th>Counts in region D</th>
<th>Peptide length</th>
<th>No. of HL genomes containing</th>
<th>Total cyanobacterial genomes containing</th>
</tr>
</thead>
<tbody>
<tr>
<td>123968043</td>
<td>Glycoprotein</td>
<td>27</td>
<td>0</td>
<td>14</td>
<td>10</td>
<td>171</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>123968658</td>
<td>Putative stress-induced protein OsmC</td>
<td>27</td>
<td>0</td>
<td>33</td>
<td>15</td>
<td>139</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>123968599</td>
<td>Putative purine phosphoribosyltransferase-related protein</td>
<td>25</td>
<td>0</td>
<td>27</td>
<td>3</td>
<td>131</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>123968811</td>
<td>Fumarate reductase subunit D</td>
<td>20</td>
<td>0</td>
<td>10</td>
<td>8</td>
<td>86</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>123968756</td>
<td>Photosystem II protein Y</td>
<td>19</td>
<td>0</td>
<td>7</td>
<td>4</td>
<td>38</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>123968104</td>
<td>Putative arsenate reductase</td>
<td>15</td>
<td>0</td>
<td>20</td>
<td>9</td>
<td>118</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>123967877</td>
<td>Ferredoxin</td>
<td>14</td>
<td>0</td>
<td>3</td>
<td>7</td>
<td>120</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>123968809</td>
<td>Metal-binding protein</td>
<td>13</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>172</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>123968377</td>
<td>Fructose-1,6-bisphosphate aldolase</td>
<td>32</td>
<td>1</td>
<td>17</td>
<td>16</td>
<td>357</td>
<td>5</td>
<td>18</td>
</tr>
<tr>
<td>123968851</td>
<td>Ferredoxin</td>
<td>22</td>
<td>1</td>
<td>7</td>
<td>5</td>
<td>113</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>123968799</td>
<td>Chorismate binding enzyme</td>
<td>14</td>
<td>1</td>
<td>12</td>
<td>4</td>
<td>59</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>123968834</td>
<td>Heat-labile enterotoxin alpha chain</td>
<td>22</td>
<td>2</td>
<td>7</td>
<td>8</td>
<td>98</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>123969152</td>
<td>Hemagglutinin-neuraminidase</td>
<td>128</td>
<td>3</td>
<td>146</td>
<td>31</td>
<td>172</td>
<td>48</td>
<td>9</td>
</tr>
<tr>
<td>123968577</td>
<td>Transglutaminase-like superfamily protein</td>
<td>41</td>
<td>3</td>
<td>67</td>
<td>7</td>
<td>285</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>123969233</td>
<td>Fe-S oxidoreductase</td>
<td>36</td>
<td>3</td>
<td>68</td>
<td>20</td>
<td>539</td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>123967902</td>
<td>Plastoquinol terminal oxidase (PTOX)</td>
<td>164</td>
<td>23</td>
<td>232</td>
<td>27</td>
<td>169</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>123967937</td>
<td>DNA polymerase III subunit beta (rpoC2)</td>
<td>66</td>
<td>58</td>
<td>130</td>
<td>14</td>
<td>385</td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>123969192</td>
<td>Cell division protein FtsW</td>
<td>58</td>
<td>90</td>
<td>111</td>
<td>16</td>
<td>411</td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>123969221</td>
<td>DNA-directed RNA polymerase subunit beta (rpoB)</td>
<td>98</td>
<td>116</td>
<td>318</td>
<td>50</td>
<td>1097</td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>123969301</td>
<td>Recombinase A (recA)</td>
<td>37</td>
<td>58</td>
<td>105</td>
<td>20</td>
<td>365</td>
<td>6</td>
<td>22</td>
</tr>
</tbody>
</table>

*Genes in white are present from the assemblies, in blue are suspected of being viral, and in purple are core genes present in all cyanobacteria.

1Sites dominated by eMIT9312 before HNLC abundant waters, including GOS sampling sites GS52–53.

2Sites dominated by HNLC clades near South Pacific Gyre, including GOS sampling sites GS40–47.

3Sites dominated by eMIT9312 post Pacific Gyre, including GOS sampling sites GS22–23 and GS26.

4Sites dominated by HNLC clades near South Pacific Gyre, including GOS sampling sites GS40–47.

5Sites in Indian Ocean containing a mixture of HNLC and eMIT9312 clades. Consists of GOS sampling sites GS111–112, including 454 titanium data, resulting in higher counts.
limited (42, 43), although the BEC model suggests otherwise (22).

However, the biogeochemistry is less defined here compared with the Pacific and Atlantic Oceans. Satellite measurements of fluorescence quantum yield of phytoplankton has been proposed as a measure of iron stress (44), and using this, Behrenfeld et al. (25) observed extended iron stress in part of the Indian Ocean gyre. Thus, it is conceivable that the iron concentration is low in the Indian Ocean sites, where we detect the HNLC clades of *Prochlorococcus*. The detection of these clades also supports the observation from satellite data that iron may be low in this region.

Iron addition experiments carried out in the Eastern Equatorial Pacific found that most phytoplankton, including marine *Synechococcus*, underwent rapid growth upon addition of iron (41, 45–47). The most striking exception was *Prochlorococcus*, which showed only a minimal response both in growth and fluorescence to iron enrichment. This suggests that *Prochlorococcus* cells in this region were not iron stressed to the same extent as other phytoplankton lineages. Our data suggest that the HNLC ecotype may have a reduced iron quota at the expense of a variety of pathways for energy generation and stress tolerance, which would explain their lack of growth response to iron addition. However, the tradeoff could be an upper limit to their photosynthetic capacity and cap their growth rate under elevated iron conditions. Thus, we hypothesize that cells from the eMIT9312 clade may have a higher growth rate when iron is more available, allowing them to outcompete HNLC cells in iron replete environments.

Adaptation in *Prochlorococcus*. Genomic islands conferring fitness under particular environmental conditions have been previously identified for phosphate and N metabolism in *Prochlorococcus* (9, 10). In this case, ecotypes adapted to high or low nutrients seem to be associated with microdiverse clusters (14). In contrast, adaptations to high light and low light or to specific water temperature have resulted in phylogenetically distinct clades and thus define the major lineages of *Prochlorococcus* (5, 6). Our genomic analysis indicates a loss of $\approx 10\%$ of the iron-containing genes, thus suggesting a reduction of the iron quota in the HNLC clades compared with other H. *Prochlorococcus*. It is important to note that we do not know the protein expression level of any of these lost proteins, so it is not possible to translate the genomic changes into a specific iron quota reduction. Future research of *Prochlorococcus* iron quota in conjunction with more precise iron measurements in this and other areas will reveal the effect of these genome variations. The phylogenetic distinctiveness of the HNLC clade suggests that adaptation to low iron availability could involve systemic changes beyond the gain or loss of a few genes. Given that gene content is not dramatically different from other clades, some of these changes may be regulatory in nature. This could include an adjustment of both the absolute expression and the ratio of PSI and -II complexes, which predominantly dictate iron requirements (48).

We are puzzled to observe not one but two coexisting HNLC lineages—each nearly as divergent from each other as they are from eMIT9312 and eMED4. Comparisons of the two genomes have not uncovered any significant gene content differences that would indicate unique physiologies. Perhaps the two lineages carry different adaptations to iron stress that each result in low iron quota. Alternatively, they have slightly different light optima and resulting depth distribution, and mixing would then allow for coexistence. However, more research is needed to identify the underlying factors for this coexistence.

Overall, our data provide evidence for a previously undescribed HL ecotype of *Prochlorococcus* proliferating in low-iron environments like the Equatorial Pacific and tropical Indian Ocean. These results are likely to influence our understanding of how iron availability affects phytoplankton physiology under light and nutrient stress. ocean biogeochemistry, and diversity now and in a future warmer and acidiac ocean.

Methods

Fragment Recruitment. Fragment recruitment was performed as described previously (13). Recruitment plots to the HNLC assemblies and PTOX scaffold can be accessed at http://gos.jcvi.org/users/gosPhaseII/advancedReferenceViewer.html.

Phylogenetic Analysis. First, we aligned three different protein families from *Prochlorococcus* and Synechococcus genomes using ClustalW (49). The peptide ClustalW alignments were used to train HMMER2 hidden Markov models (HMMs) (50). These models were used to search every GOS sequencing read (1e-2 e-value cutoff) in all six reading frames using Logical Depth (HMMs)’s accelerated version of HMMER2. Reads with positive hits were filtered to exclude partial hits ($< 95\%$ of full length) and to remove sequences with better matches to noncyanobacterial genomes according to BLASTN (51). The nucleotide sequences corresponding to these peptide matches were then extracted. The metagenomic nucleotide sequences were combined with the orthologous genes from the sequences *Prochlorococcus* and Synechococcus genomes. Phylogenetic trees with 100 bootstrap were built with PHYLIP (ver. 3.65) using neighbor-joining and maximum likelihood (52).

Environmental Data. Temperature was measured during sampling. We retrieved chlorophyll concentration data from MODISAQUA satellite data (modis.gsfc.nasa.gov). Quenching-corrected fluorescence quantum yields were estimated as described by Behrenfeld et al. (25). We retrieved monthly average values for phosphate and nitrate for each location and depth from the World Ocean Atlas (23). A caveat is that this approach may introduce additional variance in the analysis of correspondence between genome content and phosphate availability, because these values do not represent the exact nutrient concentration at the time of sampling.

Determination of *Prochlorococcus* Abundance. The abundance of environmental populations closely related to *Prochlorococcus* strains AS9601, MED4, MIT9215, and NALT1A, as well as the HNLC1 and HNLC2 genomes, was derived from fragment recruitment data (13). Large segments of each genome were selected because they had normal recruitment patterns and were not associated with variable or hypervariable segments of the genome. Only reads that recruited best to the given reference genome relative to all of the complete and draft genomes available at the National Center for Biotechnology Information as of September 18, 2009 were used. Valid reads also had to recruit to the reference genome at 80% or better identity over their entire length. The relative abundance per sample was derived from the count of base pairs per sample after being normalized for the length of the genomic segment on the reference that was used and normalized for the amount of sequencing that was collected per sample. Values are provided in terms of percentage of reads recruited per megabase pair.

Canonical Correspondence Analysis. The variance contribution of individual environmental factors on *Prochlorococcus* clade distribution was determined with partial canonical correspondence analysis (forward selection, $\alpha = 0.05$ and 9,999 perturbations) using CANOCO (ver. 4.0, Microcomputer Power) (53). Ordination plots were visualized in CanoDraw.

Census of Iron-Binding Proteins. HMMs from the Superfamily (54) database were used to identify protein structural domains within the fold family (FF) category of the Structural Classification of Proteins Database (SCOP) (55), which categorizes all of the protein structures in the Protein Data Bank (56). The SCOP was manually annotated according to metal-binding and specificity (29). At the FF hierarchy, 90% of the metal-binding FFs bind a single metal (57). In Fig. 5S, only the counts of the known Fe-binding FFs are presented. These methods of metal-binding protein identification have been vetted through manual curation (29, 57–59).

ACKNOWLEDGMENTS. We thank Dr. Keith Moore (University of California, Irvine, CA) for his biogeochemical element cycling model data; Dr. Michael Behrenfeld (Oregon State University, Corvallis, OR) for his iron stress data; Charles H. Howard, captain of the Sorcerer II, and fellow crew members for their time and effort in support of this research; numerous individuals at the J. Craig Venter Institute for their support in processing and sequencing the libraries used in this analysis; and Drs. Nyree West and Marcelino Suzuki for their comments and suggestions on the paper. Funding for this study was supplied by the Department of Energy, Office of Science, and Office of Biological and Environmental Research (DE-FG02-02ER63453), the National Science Foundation, the Gordon and Betty Moore Foundation, and the J. Craig Venter Institute.
Prochlorococcus populations along a North-South Atlantic Ocean transect. Science 312:1768–1771.


