Physiology and evolution of nitrate acquisition in Prochlorococcus

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1038/ismej.2014.211">http://dx.doi.org/10.1038/ismej.2014.211</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Nature Publishing Group</td>
</tr>
<tr>
<td>Version</td>
<td>Author's final manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Sun Nov 04 22:23:56 EST 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/97195">http://hdl.handle.net/1721.1/97195</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Physiology and evolution of nitrate acquisition in *Prochlorococcus*

Paul M Berube¹,*, Steven J Biller¹, Alyssa G Kent², Jessie W Berta-Thompson¹,³, Sara E Roggensack¹, Kathryn H Roache-Johnson⁴,⁵, Marcia Ackerman⁵, Lisa R Moore⁵, Joshua D Meisel⁶, Daniel Sher⁷, Luke R Thompson⁸, Lisa Campbell⁹, Adam C Martiny²,¹⁰, and Sallie W Chisholm¹,⁶,*

¹ Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
² Department of Ecology and Evolutionary Biology, University of California, Irvine, CA, USA
³ Microbiology Graduate Program, Massachusetts Institute of Technology, Cambridge, MA, USA
⁴ Department of Molecular and Biomedical Sciences, University of Maine, Orono, ME, USA
⁵ Department of Biological Sciences, University of Southern Maine, Portland, ME, USA
⁶ Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA
⁷ Department of Marine Biology, University of Haifa, Haifa, Israel
⁸ Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO, USA
⁹ Department of Oceanography, Texas A&M University, College Station, TX, USA
¹⁰ Department of Earth System Science, University of California, Irvine, CA, USA

* Correspondence: SW Chisholm, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Bldg 48-419, Cambridge, MA 02139, USA. E-mail: chisholm@mit.edu; PM Berube, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Bldg 48-424, Cambridge, MA 02139, USA. E-mail: pmberube@gmail.com

Running Title: Genomics of *Prochlorococcus* nitrate utilization
Subject Category: Evolutionary genetics
Prochlorococcus is the numerically dominant phototroph in the oligotrophic subtropical ocean and carries out a significant fraction of marine primary productivity. While field studies have provided evidence for nitrate uptake by Prochlorococcus, little is known about this trait because axenic cultures capable of growth on nitrate have not been available. Additionally, all previously sequenced genomes lacked the genes necessary for nitrate assimilation. Here we introduce three Prochlorococcus strains capable of growth on nitrate and analyze their physiology and genome architecture. We show that the growth of high-light adapted strains on nitrate is approximately 17% slower than their growth on ammonium. By analyzing 41 Prochlorococcus genomes, we find that genes for nitrate assimilation have been gained multiple times during the evolution of this group, and can be found in at least three lineages. In low-light adapted strains, nitrate assimilation genes are located in the same genomic context as in marine Synechococcus. These genes are located elsewhere in high-light adapted strains and may often exist as a stable genetic acquisition as suggested by the striking degree of similarity in the order, phylogeny, and location of these genes in one high-light adapted strain and a consensus assembly of environmental Prochlorococcus metagenome sequences. In another high-light adapted strain, nitrate utilization genes may have been independently acquired as indicated by adjacent phage mobility elements; these genes are also duplicated with each copy detected in separate genomic islands. These results provide direct evidence for nitrate utilization by Prochlorococcus and illuminate the complex evolutionary history of this trait.

Keywords: cyanobacteria / genomics / narB / nitrate / Prochlorococcus / Synechococcus
INTRODUCTION

The unicellular cyanobacterium *Prochlorococcus* is the smallest known free-living oxygenic phototroph (Chisholm et al., 1992; Partensky et al., 1999; Partensky & Garczarek, 2010; Coleman & Chisholm, 2007). It is numerically dominant in the tropical and subtropical regions of the world’s oceans and responsible for 5-10% of marine primary productivity (Campbell et al., 1994; Partensky et al., 1999; Flombaum et al., 2013; Buitenhuis et al., 2012). *Prochlorococcus* has undergone a process of genome reduction following divergence from its closest relatives, the marine *Synechococcus* (Rocap et al., 2002; Kettler et al., 2007). These streamlined genomes are often considered an adaptation to the oligotrophic environments they occupy (Rocap et al., 2003; Dufresne et al., 2003). Even though individual genomes are small, the collective of all *Prochlorococcus* cells possesses a vast reservoir of genetic and physiological diversity (Kettler et al., 2007). *Prochlorococcus* is composed of a polyphyletic group of low-light (LL) adapted clades (LLI-LLVI and NC1), and a more recently diverged monophyletic group of high-light (HL) adapted clades (HLI-HLVI) (Malmstrom et al., 2013; Lavin et al., 2010; Huang et al., 2012; Moore et al., 1998; Moore & Chisholm, 1999; Rocap et al., 2002; Martiny et al., 2009c; Shi et al., 2011). Some of these clades are known to be differentially distributed along gradients of light intensity, temperature, and nutrient concentrations (Bouman et al., 2006; Johnson et al., 2006; Zinser et al., 2006; Zvirglmaier et al., 2007; Zvirglmaier et al., 2008; Malmstrom et al., 2010; Malmstrom et al., 2013).

Nitrogen availability often limits primary productivity in marine systems (Tyrrell, 1999), and organisms have evolved diverse mechanisms for uptake of various chemical forms of nitrogen. Nitrate is one of the more abundant sources of inorganic nitrogen available to phytoplankton (Gruber, 2008), and the majority of cyanobacteria possess pathways for the uptake and assimilation of nitrate (García-Fernández et al., 2004; Herrero et al., 2001; Ohashi...
et al., 2011). Early reports on the vertical distributions of Prochlorococcus noted a subsurface maximum in abundance at the base of the euphotic zone, which suggested Prochlorococcus was sensitive to nitrogen depletion and might be assimilating nitrate supplied from deep waters (Olson et al., 1990; Vaulot & Partensky, 1992). Therefore, it was surprising that nearly all isolates of Prochlorococcus could not use nitrate and lacked the genes required for this function (Kettler et al., 2007; Coleman & Chisholm, 2007; Moore et al., 2002) even though most isolates of Synechococcus are capable of using nitrate (Ahlgren & Rocap, 2006; Fuller et al., 2003). Only a single Prochlorococcus culture, PAC1 isolated in 1992, was reported to utilize nitrate (Williams et al., 1999), but due to the presence of other bacteria in that culture, direct nitrate uptake by Prochlorococcus could not be conclusively demonstrated.

Several pieces of evidence indicated that nitrate assimilation was a more common trait within Prochlorococcus populations than previously thought. Field experiments demonstrated the uptake of isotopically labeled nitrate by Prochlorococcus cells in the Sargasso Sea (Casey et al., 2007), and nitrate assimilation genes were found to be associated with uncultivated Prochlorococcus genomes from many regions of the subtropical oceans (Martiny et al., 2009b). A scaffold assembled from metagenomic data from the Global Ocean Sampling (GOS) expedition indicated that all the genes required for nitrate assimilation were co-localized in a specific region of the genomes of high-light adapted Prochlorococcus. The metagenomic data primarily identified nitrate utilization genes in the HLII clade of Prochlorococcus since sequences from this clade comprised the majority of Prochlorococcus-like sequences in the GOS dataset (Rusch et al., 2007).

These past observations raised two important questions about nitrate assimilation in Prochlorococcus. (1) Can axenic strains grow on nitrate as the sole nitrogen source? (2) What is the evolutionary history of nitrate assimilation genes in this group? To address these questions, we isolated and sequenced Prochlorococcus strains capable of nitrate assimilation.
and examined their growth on different nitrogen sources. We then used comparative genomics
to better understand how this trait had evolved in *Prochlorococcus*. 
MATERIALS AND METHODS

Strains and enrichments. Five strains of Prochlorococcus (SB, MIT0604, PAC1, MIT9301, and MED4), one strain of Synechococcus (WH8102), and two Prochlorococcus enrichment cultures (P0902-H212 and P0903-H212) were used in this study. MIT9301, MED4, and WH8102 have previously been rendered axenic (free of heterotrophic contaminants). All axenic cultures were routinely assessed for purity by confirming a lack of turbidity after inoculation into a panel of purity test broths: ProAC (Morris et al., 2008), MPTB (Saito et al., 2002), and ProMM (Pro99 medium (Moore et al., 2007) supplemented with 1x Va vitamin mix (Waterbury & Willey, 1988) and 0.05% w/v each of pyruvate, acetate, lactate, and glycerol). ProMM is a modified version of the PLAG medium (Morris et al., 2008), but uses 100% seawater as the base.

PAC1 was enriched from seawater collected from the deep chlorophyll maximum in the North Pacific Ocean at Station ALOHA (22.75°N, 158°W) on Hawai’i Ocean Time-series (HOT) cruise 36. Seawater was passed through a 0.6 μm Nucleopore filter twice, and the filtrate was serially diluted into K/10 medium (Chisholm et al., 1992), but with the following modifications for final nutrient concentrations: 5 μM urea, 5 μM ammonium, 1 μM β-glycerophosphate replacing inorganic phosphate, 0.01 μM Na2MoO4 and 0.05 μM NiCl2.

MIT0604 was derived from an enrichment culture initiated with Pro2 nutrient additions (Moore et al., 2007) to seawater obtained at Station ALOHA on HOT cruise 181, but with all nitrogen sources replaced by 0.217 mM sodium nitrate. The P0902-H212 and P0903-H212 enrichments were initiated with Pro2 nutrient additions (Moore et al., 2007) to seawater obtained from Station ALOHA on HOT cruise 212, but with all nitrogen sources replaced by 0.05 mM sodium nitrate.

Purification of Prochlorococcus strains. SB and MIT0604 were rendered axenic in this study using a modified dilution to extinction method. Prochlorococcus from exponential
phase cultures were enumerated using an Influx Cell Sorter (BD Biosciences, San Jose CA, USA) or a FACSCalibur flow cytometer (BD Biosciences) as previously described (Olson et al., 1985; Cavender-Bares et al., 1999). Cultures consisting of >80% Prochlorococcus cells were serially diluted into multiple multi-well plates at final concentrations of 1-10 cells per well in at least 200 µL of ProMM medium. Axenic Prochlorococcus do not grow from such low cell densities in Pro99 medium without “helper” heterotrophic bacteria (Morris et al., 2008; Morris et al., 2011), however, they do grow when diluted into ProMM. The main ingredient in ProMM which promotes the growth of cells from low densities is pyruvate, and we suspect that in this context pyruvate serves as a potent hydrogen peroxide scavenger (Giandomenico et al., 1997). Wells contaminated with heterotrophic bacteria were identified by the appearance of turbidity. The multi-well plates were monitored by eye and by fluorometry using a Synergy HT Microplate Reader (BioTek, Winooski, VT, USA), and non-turbid wells were monitored by flow cytometry using a FACSCalibur flow cytometer. Wells that appeared green or had Prochlorococcus cells as determined by flow cytometry were immediately transferred to Pro99 medium directly, or into fresh ProMM medium until consistent growth was observed, at which point the cultures were introduced back into Pro99 medium. Cultures were examined for heterotrophic bacteria contaminants by flow cytometry and by inoculation into the panel of purity test broths as described above.

**PCR screen for the nitrate reductase gene.** Based on an alignment of GOS reads coding for the Prochlorococcus narB sequence (Martiny et al., 2009b), degenerate primers 30narB175f (5’-TGYGTDAAAGGMGCAACAGTNTG-3’) and 30narB574r (5’-GACAYTCWGCBGTATTWGTHCC-3’) were designed specifically to amplify the narB gene from HLII clade Prochlorococcus, and degenerate primers 40narB1447f (5’-TATTGYCCAGCWTTYMGDCCDTG-3’) and 40narB1766r (5’-AKAGGWTGYTTWGTRTARAAYTG-3’) were designed specifically to amplify the narB
gene from LLI clade Prochlorococcus. Polymerase chain reactions (PCR) used annealing temperatures of 52.5°C for the HLII narB sequence and 56°C for the LLI narB sequence. Reactions contained 1x PCR buffer, 2.5 mM MgCl₂, 0.2 mM each of dATP, dTTP, dCTP, and dGTP, 0.2 µM of each primer, 1 unit of Platinum Taq DNA polymerase (Life Technologies, Grand Island, NY, USA), and 1 ng of genomic DNA prepared from Prochlorococcus cultures in the MIT Cyanobacteria Culture Collection (Chisholm Laboratory, MIT). DNA from Synechococcus WH8102, which contains a narB gene, was used as a negative control. Reactions were cycled 30 times at 94°C for 15 s, the primer specific annealing temperature for 15 s, and 72°C for 60 s. PCR products with the expected size were sequenced at the Dana-Farber/Harvard Cancer Center DNA Resource Core to confirm amplification of the narB gene.

Growth in the presence of alternative nitrogen sources. Axenic Prochlorococcus strains SB, MIT0604, MIT9301, and MED4, and axenic Synechococcus strain WH8102 were acclimated to Pro99 medium (Moore et al., 2007) prepared with seawater from the South Pacific Subtropical Gyre and grown at 24°C and 50 µmol photons m⁻² s⁻¹ continuous illumination for at least 10 generations or until growth rates were similar between successive transfers. Bulk culture fluorescence was measured as a proxy for biomass using a 10AU fluorometer (Turner Designs, Sunnyvale, CA, USA). Triplicate cultures of each strain were initiated in Pro99, which contained 0.8 mM ammonium chloride. Once cultures had reached mid-exponential phase, they were transferred into Pro99 medium containing 0.8 mM ammonium chloride, 0.8 mM sodium nitrate, 0.8 mM sodium cyanate, or no nitrogen additions as a control to monitor utilization of carry-over ammonium. Cultures were successively transferred at mid-exponential phase until growth in the cultures lacking nitrogen additions had arrested due to nitrogen limitation. Specific growth rates were estimated from the log-linear portion of the growth curve for the final transfer. Two tailed homoscedastic t-
tests were conducted in Microsoft Excel (Microsoft Corporation, Redmond, WA, USA) in order to evaluate the likelihood of significantly different growth rates in each strain for each pair of nitrogen sources and for strains grown on the same nitrogen source.

**Genome data.** 41 *Prochlorococcus* and 15 *Synechococcus* genomes (Biller et al., 2014), which include the genomes of the nitrate assimilating strains SB, MIT0604, and PAC1, were used in this study. Sequence data were also obtained for the P0902-H212 and P0903-H212 enrichment cultures as described in the Supplementary Methods. These enrichment assemblies had total sequence lengths approximately twice the size of previously sequenced *Prochlorococcus* genomes, suggesting the presence of at least two unique strains dominating each enrichment. Binning contigs based on average sequencing coverage yielded a subset of highly covered contigs in each assembly with a total sequence length similar to that of previously sequenced *Prochlorococcus* genomes. In the highly covered subsets for each assembly, the complete set of nitrate assimilation genes were only found on a single contig. For the purpose of this study, only these contigs were relevant and entered into our analysis.

All sequence data were annotated using the RAST server (Aziz et al., 2008) with FIGfam release 49 in order to facilitate comparison between genomes by ensuring a uniform methodology for gene calling and functional annotation. Clusters of orthologous groups of proteins (COGs) were identified as previously described (Kelly et al., 2012). These clusters are included in the “V4” CyCOGs on the ProPortal website (http://proportal.mit.edu) (Kelly et al., 2012; Biller et al., 2014).

**Genome phylogeny.** We translated 537 single-copy core genes to amino acid sequences, aligned each gene individually in protein space using ClustalW (Larkin et al., 2007), and then back-translated the sequences using TranslatorX (Abascal et al., 2010). Using the principle previously described (Kettler et al., 2007), we randomly concatenated 100 of these aligned genes and built maximum likelihood (ML) and neighbor joining (NJ)
phylogenies using PHYLIP v3.69 (Felsenstein, 2005). We repeated the random concatenation
and tree generation 100 times.

Estimation of gene gain and loss. Using a maximum parsimony approach (Mirkin et
al., 2003), the patterns of gene gain and loss were mapped onto the topology of the ML
nucleotide tree using WH5701 as an outgroup. Utilizing 13,590 non-core single-copy COGs,
we reconstructed ancestral character states of gene absence and presence on our guide tree
and minimized the cost of gains and losses given a gene gain equal to twice a gene loss. We
used the program DendroPy to implement the tree traversal portion of the algorithm
(Sukumaran & Holder, 2010).

Phylogenies of genes involved in the transport and reduction of nitrate and nitrite.
COGs corresponding to the nirA, narB, focA, and napA genes were aligned in protein space
using ClustalW. Phylogenetic trees were estimated with PHYLIP v3.69 (Felsenstein, 2005)
using the programs SEQBOOT, PROTDIST with the Jones-Taylor-Thornton matrix and a
constant rate of variability among sites, and NEIGHBOR on the aligned amino acid
sequences with Synechococcus WH5701 used as an outgroup for nirA and narB and
Synechococcus CB0101 used as an outgroup for focA and napA. We included GOS consensus
sequences: GOS nirA, GOS narB, and GOS napA (Martiny et al., 2009b).
RESULTS AND DISCUSSION

Isolates of Prochlorococcus are capable of nitrate assimilation. To identify possible cultures capable of nitrate assimilation, we screened existing Prochlorococcus cultures for the assimilatory nitrate reductase gene, *narB*, using PCR. We found that the low-light adapted PAC1 strain (Penno et al., 2000) and the high-light adapted SB strain (Shimada et al., 1995) each contained the gene. In search of additional strains capable of utilizing nitrate, we performed selective enrichments from seawater obtained from the subtropical North Pacific Ocean using nitrate as the sole added nitrogen source. This yielded one high-light adapted strain (Prochlorococcus MIT0604) and two mixed Prochlorococcus cultures (P0902-H212 and P0903-H212) with the *narB* gene (Table 1).

We then rendered SB and MIT0604 axenic and examined their growth in the presence of nitrate or ammonium. As hypothesized, both SB and MIT0604 can grow on nitrate as the sole source of nitrogen, but with a significant reduction in growth rate (18% and 17% respectively), compared to growth on ammonium (Figure 1 and Supplementary Figure S1). Although the slower growth on nitrate could be explained by the greater amount of reducing power required to assimilate more oxidized N sources (García-Fernández et al., 2004), we assume that these cultures were growing at saturating light intensities based on previous measurements of light saturating irradiances for the growth of Prochlorococcus (Moore and Chisholm, 1999); thus energy supply and reducing power were likely not limiting. Furthermore, recent work has shown that the growth rates and chemical composition of some marine cyanobacteria are not directly related to the oxidation state of the cells’ N source (Collier et al., 2012). Under light limiting conditions, for example, the growth rate and chemical composition of Synechococcus grown on ammonium was the same as that on nitrate; but, under light saturating conditions, cells grown on nitrate had a higher C:N ratio (Collier et al., 2012). This perhaps suggests a bottleneck in the uptake and conversion of
nitrate compared to ammonium when energy is sufficient (Collier et al., 2012), and may
explain the slower growth of *Prochlorococcus* on nitrate compared to ammonium.

In the early days of research on *Prochlorococcus*, the absence of cultures known to
utilize nitrate resulted in a distorted view of *Prochlorococcus*’ role in marine ecosystems;
ecosystem models and ecophysiological interpretations were guided by the assumption that
most, if not all, *Prochlorococcus* were incapable of nitrate assimilation (Follows et al., 2007;
Fuller et al., 2005; García-Fernández et al., 2004). Why have nitrate-utilizing
*Prochlorococcus* appeared so infrequently in culture collections in the past? Is it because we
were selecting against them in isolations using media containing ammonium but not nitrate
(Moore et al., 2007)? We think not because SB and MIT0604 – both narB containing strains –
grow at equal or better rates on ammonium compared to other high-light adapted
*Prochlorococcus* strains (Figure 1 and Supplementary Figure S1). An alternative explanation
is that most of the early cultures of *Prochlorococcus* were isolated from environments that are
relatively nitrogen replete – i.e. thought to be more limited by phosphorus or iron availability
(e.g. the Sargasso Sea, Mediterranean Sea, and the Equatorial Pacific) (Kettler et al., 2007;
Wu et al., 2000; Marty et al., 2002; Vaulot et al., 1996; Mann & Chisholm, 2000; Rusch et al.,
2010). We now know that *Prochlorococcus* cells capable of nitrate assimilation are more
likely to be found in ocean regions with lower average nitrate concentrations, such as the
Caribbean Sea and Indian Ocean (Martiny et al., 2009b). Indeed, PAC1 and SB (both narB
containing strains that were isolated on medium containing ammonium but lacking nitrate),
were isolated from N-poor regions (Penno et al., 2000; Wu et al., 2000; Shimada et al., 1995;
Iwata et al., 2005). Thus we believe that the probability of obtaining a *narB* containing strain
using medium containing ammonium is in large part a function of the particular water sample
used to start enrichment cultures.
Nitrate assimilation is found in diverse lineages of Prochlorococcus. What can the features of the nitrate assimilation genes in Prochlorococcus tell us about how they have been gained or lost during the evolution of this group? The genomes of PAC1, SB, and MIT0604, along with contigs containing nitrate assimilation genes from the P0902-H212 and P0903-H212 enrichment cultures, were informative in this regard. These Prochlorococcus belong to both the low-light adapted LLI clade (PAC1, P0902-H212, and P0903-H212) and the high-light adapted HLII clade (SB and MIT0604) (Figure 2 and Supplementary Figures S2 and S3), demonstrating that nitrate utilization is found in multiple and diverse lineages of Prochlorococcus and suggesting a complex evolutionary history. The presence of nitrite and nitrate metabolism in Prochlorococcus follows that of Synechococcus in that some strains are able to reduce nitrite and some are able to reduce both nitrite and nitrate. Because these traits are not monophyletic, a model of gene gain and loss events provides evidence for 3 gains and 2 losses for the \( \text{narB} \) nitrate reductase gene and 2 gains and 3 losses for the \( \text{nirA} \) nitrite reductase gene (Figure 2). With the limited number of genomes available, it appears that there is evidence for multiple gains and losses of nitrogen assimilation traits through the evolution of Prochlorococcus and Synechococcus, with \( \text{narB} \) found in at least three distinct Prochlorococcus lineages.

The genomic context of the nitrate assimilation gene cluster suggests a complex evolutionary history. To look for features that might help us interpret the gains and losses of nitrate and nitrite assimilation genes in Prochlorococcus we examined the local genomic context of these genes. While the full complement of nitrate assimilation genes was predicted to be localized in a single region of the highly syntenic HLII clade genomes from metagenomic assemblies (Martiny et al., 2009b), it was unclear whether this context would be found in any individual cell. Further, given that these genes were found in a different region
in *Prochlorococcus* compared to marine *Synechococcus*, we were curious as to whether we might find evidence for rearrangements or lateral gene transfer.

The nitrate assimilation genes in PAC1 and the P0902-H212 and P0903-H212 contigs are syntenic and also found in the same genomic region as the nitrite assimilation genes in NATL1A and the nitrate assimilation genes in *Synechococcus* WH8102 (Figure 3). This region is bounded by a pyrimidine biosynthesis gene (*pyrG*) and a polyphosphate kinase gene (*ppk*) between which many nitrogen assimilation genes are located in marine *Synechococcus*. While gene gains and losses have been observed in this region (Scanlan et al., 2009), our data indicate that the genomic location of the nitrate and nitrite assimilation genes is reasonably well fixed in LLI *Prochlorococcus* and closely related *Synechococcus*. Although our model of gene gain and loss events suggests the loss of nitrate assimilation genes early in the evolution of *Prochlorococcus* (Figure 2), the local genomic features of these genes are consistent with the interpretation that some lineages may have retained these genes following the divergence of *Prochlorococcus* from *Synechococcus*.

Analysis of metagenomic data from GOS (Martiny et al., 2009b) suggested that the nitrate utilization genes in HLII *Prochlorococcus* should be located in a different genomic region compared to LLI genomes, indicating an alternative evolutionary origin. Based on a scaffold of mate-paired metagenomic reads, it was inferred that this cluster should be located approximately 500 kb downstream of the *pyrG-ppk* region containing the nitrate assimilation genes in WH8102 and the nitrite assimilation genes in NATL1A (Martiny et al., 2009b). We found a high degree of similarity between the nitrate assimilation gene cluster in SB and the scaffold derived from GOS metagenome sequences obtained from multiple individual cells from multiple sampling stations. This similarity manifested itself not only in the gene order and chromosomal location, but also the phylogeny of the nitrate assimilation genes (Figures 3-5), placing the nitrate assimilation gene cluster in a genomic region that is syntenic with
other HLII genomes and adjacent to a known genomic island (ISL3) in this clade (Figure 4).

Further, a partial genome from a *Prochlorococcus* single-cell belonging to the HLII clade (B241-528J8; Genbank JFLE01000089.1) (Kashtan et al., 2014) also possesses a nitrate assimilation gene cluster in the same location and in the same order. The striking similarity between the nitrate assimilation gene clusters of these individual *Prochlorococcus* and the GOS consensus indicates that the order and location of nitrate assimilation genes are stable within HLII genomes.

The nitrate assimilation genes in strain MIT0604 had a different local genome structure compared to strain SB and the partial single-cell genome, B241-528J8. MIT0604 has duplicate clusters of these genes, which are inversely oriented and located upstream and downstream of the GOS-predicted location (Figure 3 and 4). A Southern blot confirmed that MIT0604 does indeed contain two copies of *narB* whereas SB contains only one (Supplementary Figure S4), and they are located within genomic islands ISL3 and ISL4 of HLII clade *Prochlorococcus* (Figure 4). Genomic islands are common features of *Prochlorococcus* genomes, particularly within the high-light adapted clades (Coleman et al., 2006; Kettler et al., 2007). They harbor much of the variability in gene content between members of the same clade and are hotspots for lateral gene transfer. Phage integrase genes are located proximal to both nitrate assimilation gene clusters in MIT0604, and a transfer RNA gene is adjacent to one of these clusters (Figure 3). The transfer RNA genes are known to serve as sites for insertion of phage DNA in bacteria (Williams, 2002), and thus the location of these phage integrase and transfer RNA genes suggests transduction as a possible mechanism by which MIT0604 has acquired the nitrate assimilation gene cluster. Notably, duplication of such a large region of the chromosome has not been observed previously in *Prochlorococcus*, and thus far, MIT0604 is the only *Prochlorococcus* or *Synechococcus* strain possessing two complete copies of the genes required for nitrate assimilation.
The phylogenies of nitrate assimilation genes are similar to the phylogeny of genomes. Given the evidence for both a stable arrangement of the nitrate assimilation genes in some Prochlorococcus and possible gene transfer leading to acquisition of the nitrate assimilation trait in MIT0604, we were curious to know whether the phylogenies of these genes were congruent with whole genome phylogenies (Figure 2 and Supplementary Figure S2), as well as the phylogeny of GyrB (Supplementary Figure S3) which has been identified as a useful phylogenetic marker for Prochlorococcus (Mühling, 2012). Thus, we reconstructed the amino acid phylogenies of the NirA and NarB reductases, the FocA nitrite transporter, and the NapA nitrite/nitrate transporter (Figure 5). The NirA phylogeny is largely consistent with our observations based on the GOS metagenome data (Martiny et al., 2009b), such that the NirA proteins from genomes in the LLIV clade are more closely associated with marine Synechococcus than with other Prochlorococcus sequences. In all phylogenetic trees, the PAC1, P0902-H212, and P0903-H212 sequences are in a separate clade distinct from that of the SB and MIT0604 sequences, reinforcing the HL versus LL differentiation (Figure 5). The NirA and NarB sequences from SB are consistently more closely affiliated with the GOS consensus sequence (Martiny et al., 2009b) than with the MIT0604 sequences. NapA sequences from SB and MIT0604 are also both closely related to the GOS NapA consensus sequence (Figure 5). Similar to the GyrB phylogeny (Supplementary Figure S3), the P0903-H212 sequences fall outside the clade containing the other LLI sequences. With the exception of the LLIV NirA sequences, the phylogenies of these nitrite and nitrate assimilation proteins (Figure 5) are congruent with whole genome and GyrB phylogenies (Figure 2 and Supplementary Figures S2-S3) at a resolution defining the major Prochlorococcus clades.

Nitrate assimilating Prochlorococcus possess a diverse set of nitrogen acquisition pathways. Gene content in Prochlorococcus has been shown, for several traits, to reflect the selective pressures in the specific environments from which they (or their genes) were
captured (Martiny et al., 2006; Coleman & Chisholm, 2010; Feingersch et al., 2012; Malmstrom et al., 2013; Rusch et al., 2007). Thus, we wondered if other nitrogen assimilation traits might co-occur with nitrate assimilation in Prochlorococcus, and examined the potential for PAC1, SB, and MIT0604 to access alternative sources of nitrogen based on their gene content (Supplementary Table S1 and Supplementary Figure S5).

Like other members of the LLI clade, PAC1 possesses genes for the assimilation of ammonium and urea, but lacks cyanate transporter genes. In addition to the napA nitrite/nitrate transporter, the focA nitrite transporter is found in both PAC1 and in the contig from P0902-H212. However, the focA gene is absent from high-light adapted strains SB and MIT0604, and most surface water metagenomic samples (Martiny et al., 2009b). Some Synechococcus strains (e.g. WH8102) (Supplementary Figure 5) also lack focA; thus, this gene is clearly subject to gain and loss. While focA is also similar to formate transporters, evidence implicates its role in nitrite uptake in Prochlorococcus; e.g. the gene is located near other nitrite assimilation genes (Figure 3), it's upregulated under nitrogen stress (Tolonen et al., 2006), and it's absent from Prochlorococcus that cannot grow on nitrite (Moore et al., 2002; Coleman & Chisholm, 2007; Kettler et al., 2007) (Supplementary Figure 5). Since PAC1 possesses both a nitrite transporter (focA) and the dual function nitrate/nitrite transporter (napA), it is possible that focA provides some advantage to low-light adapted cells which are often maximally abundant near the nitrite maxima in the oceans (Scanlan & West, 2002; Lomas & Lipschultz, 2006). Low-light adapted cells that possess the dual function nitrite/nitrate transporter may benefit from having an additional transporter for nitrite. Given that high-light adapted Prochlorococcus strains capable of nitrate utilization lack the focA gene, these cells may be less reliant on nitrite as a nitrogen source.

SB and MIT0604 possess urea assimilation genes and can utilize urea as a sole nitrogen source (Supplementary Figure S6). Further, SB possesses cyanate transporter genes,
which are rare in both *Prochlorococcus* and *Synechococcus* strains (Kamennaya et al., 2008), and it can indeed grow utilizing cyanate (Supplementary Figure S1) as the sole source of nitrogen. While very little is known about cyanate concentrations in marine systems, *cynA* genes (encoding the periplasmic component of the cyanate ABC-type transporter system) were relatively abundant in the seasonally stratified and nitrogen depleted waters of the northern Red Sea (Kamennaya et al., 2008). The *cynA* gene of SB clusters with clones obtained from the Red Sea (Supplementary Figure S7), supporting their origin in HLII clade genomes as hypothesized by Kamennaya et al.

SB contains the most extensive suite of nitrogen acquisition pathways of any cultured *Prochlorococcus* strain examined to date. Why might this be? A useful analogy can be drawn from our understanding of selection pressures that have shaped *Prochlorococcus* genomes with respect to adaptations involved in phosphorus assimilation. Individual cells and populations from phosphorus-limited environments possess accessory phosphorus acquisition genes, such as alkaline phosphatase (*phoA*) and phosphonate utilization (*phnYZ*) genes, at a higher frequency than *Prochlorococcus* from phosphorus-replete environments (Martiny et al., 2006; Martiny et al., 2009a; Coleman & Chisholm, 2010; Feingersch et al., 2012). Thus, we hypothesize that the nitrogen assimilation traits present in *Prochlorococcus* SB were likely shaped by frequent nitrogen limitation in its original habitat (Iwata et al., 2005); i.e. cells capable of accessing a wide pool of nitrogen compounds may be at a selective advantage in nitrogen-limited environments.
CONCLUSIONS

Given the large standing stock of *Prochlorococcus* in the subtropical oceans and the extent to which nitrogen limits primary production in these regions (Tyrrell, 1999; Moore et al., 2013), the absence of nitrate assimilation capabilities in cultured strains of *Prochlorococcus* has long puzzled biological oceanographers. This motivated field studies (Casey et al., 2007; Martiny et al., 2009b) and the use of models to help us understand the selection pressures driving the loss of nitrate assimilation genes in *Prochlorococcus* relative to *Synechococcus* (Bragg et al., 2010). In this study we show unequivocally that some strains of *Prochlorococcus* are indeed capable of growth using nitrate as the sole nitrogen source.

Future studies of these strains will help elucidate the physiological trade-offs of carrying these genes and help refine the nitrogen inventory in biogeochemical models of the global ocean (Follows et al., 2007). Correlations between environmental nitrate concentrations and ribotype phylogeny (Martiny et al., 2009c) and the striking similarity between *Prochlorococcus* SB and the GOS consensus sequence both suggest that the trait for nitrate assimilation could be tied to distinct ribotype lineages. Still, evolution has many ways of introducing genomic complexity: the MIT0604 genome suggests that these genes are also subject to horizontal gene transfer, allowing further diversification of this trait in other lineages. This is reminiscent of the phylogenetic characteristics of phosphorus acquisition traits, which are nearly independent of ribotype phylogeny (Martiny et al., 2009c) – with extensive diversity in the ‘leaves of the tree’. As we learn more about these layers of diversity it will inform parameterizations of the relationship between light, temperature, and nutrient acquisition traits for ocean simulation modeling.
ACKNOWLEDGEMENTS

We thank the captain and crew of the R/V Kilo Moana and members of the Hawai‘i Ocean Time-series program (HOT181 and HOT212) for technical support with field operations. We also thank Robert D. Harper and Hassan Shaleh (University of Southern Maine, Portland, ME) for culturing assistance as well as Libusha Kelly (Albert Einstein College of Medicine, Bronx, NY) for advice on bioinformatics analyses. This work was funded in part by the Gordon and Betty Moore Foundation through Grant GBMF495 to SWC and by the National Science Foundation (Grants OCE-1153588 and DBI-0424599 to SWC, OCE-0928544 to ACM, OCE-0851288 to LRM and OCE-9417071 to LC). AGK was supported by the NSF Graduate Research Fellowship Program (DGE-1321846). This article is a contribution from the NSF Center for Microbial Oceanography: Research and Education (C-MORE).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Supplementary information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej).
REFERENCES


TITLES AND LEGENDS TO FIGURES

Figure 1. Maximum specific growth rates ($\mu_{\text{max}}$) of Prochlorococcus strains SB, MIT0604, MIT9301, MED4, and Synechococcus WH8102 in the presence of ammonium or nitrate. Values represent the mean and standard deviation of 3 biological replicates. Growth rate differences for each strain grown on ammonium compared with nitrate as well as growth rate differences between strains on the same nitrogen source were significant (p < 0.05) in a two tailed homoscedastic t-test. n.g., no growth.

Figure 2. Maximum likelihood phylogeny of Prochlorococcus and Synechococcus based on the similarity of 100 randomly concatenated single-copy core genes. Nodes are marked by closed circles to indicate that the associated taxa clustered together in at least 75% of 100 replicate trees. Genes lost and gained in the evolution of Prochlorococcus and Synechococcus are indicated at each node by values representing losses followed by gains. Predicted losses (open circles) or gains (closed circles) of nirA (blue) or narB (orange) are labeled on their respective branches.

Figure 3. Architecture of the nitrite and nitrate assimilation genes in low-light adapted (LLI clade) and high-light adapted (HLII clade) Prochlorococcus relative to Synechococcus WH8102. Similar to Synechococcus, the nitrite and nitrate assimilation genes in the LLI clade of Prochlorococcus are found within the region between the pyrG (pyrimidine biosynthesis) and ppk (polyphosphate kinase) genes. Most LLI clade Prochlorococcus, with the exception of the P0903-H212 contig, possess a focA nitrite transporter in this region (possibly acquired from proteobacteria (Rocap et al., 2003)). Metagenome data (Martiny et al., 2009b), a partial genome from a single cell (B241-528J8) (Kashtan et al., 2014), and a culture genome (Prochlorococcus SB) indicate that the nitrate assimilation genes within HLII clade Prochlorococcus are commonly found in a syntenic region adjacent to genomic island ISL3.
Prochlorococcus MIT0604 is an exception in that it possesses duplicate nitrate assimilation gene clusters located within genomic islands ISL3 and ISL4 (see Figure 4), with phage integrase genes immediately adjacent to each copy of the nirA (nitrite reductase) gene.

Figure 4. Locations of nitrate and cyanate assimilation genes in strains of Prochlorococcus capable of nitrate assimilation relative to the known genomic islands (shaded regions) observed in the HLII and LLI clades of Prochlorococcus; plots modified from Kettler et al., 2007. Prochlorococcus genomes are highly syntenic and genomic islands have been identified in high-light adapted genomes (e.g. AS9601) by conserved breaks in gene synteny among strains (Coleman et al., 2006; Kettler et al., 2007). Genomic islands have also been identified (e.g. the large region within LLI clade genomes such as NATL1A) by predicted gene gain events along the chromosome (Kettler et al., 2007).

Figure 5. Neighbor joining phylogeny of 4 proteins involved in the transport and reduction of nitrate and nitrite in marine cyanobacteria: (a) NirA; nitrite reductase, (b) NarB; nitrate reductase, (c) FocA; nitrite transporter, and (d) NapA; nitrite/nitrate transporter. The percentage of 100 replicate trees in which the associated taxa clustered together is indicated at nodes by closed circles (>75%) or open circles (>50%). Scale bars represent substitutions per site.
FIGURE 2

substitutions per site

narB loss

narB gain

nirA loss

nirA gain

Synechococcus Prochlorococcus

substitutions per site

HLIHLIILLILLII,IIILLIV

FIGURE 2

ACCEPTED AUTHOR MANUSCRIPT
Synechococcus WH8102

**FIGURE 3**

*Urea Assimilation*

*Nitrate Assimilation*

Conserved genes in pyrG-ppk region

Nitrate transport and reduction
Nitrite transport and reduction
focA nitrite transporter
Cyanate transport and hydrolysis

Core HLII clade genes
tRNA
Putative phage integrase
Urea transport and hydrolysis

Predicted location of nitrate assimilation genes based on GOS consensus

**Synechococcus**

LL IV

HL I

**Prochlorococcus**

HL II

LL I

**HLII Prochlorococcus**

Synechococcus LL IV

LL II,III

LL I

HL II

HL I

**Core HLII clade genes**

**Nitrate transport and reduction**

**Nitrite transport and reduction**

**focA nitrite transporter**

**Cyanate transport and hydrolysis**

**GOS Consensus**

*COGs Matching GOS Mate Pair Reads*

Synechococcus

P0902-H212

P0903-H212

AS9601

SB

B241_528J8 Single Cell

Synechococcus

PAC1

HL I

HL II

HL III

HL IV

**Conserved genes in pyrG-ppk region**

**Nitrate transport and reduction**

**Nitrite transport and reduction**

**focA nitrite transporter**

**Cyanate transport and hydrolysis**

**Synechococcus**

LL IV

HL I

**Synthetic construction**

HL II

LL I

**Core HLII clade genes**

**Nitrate transport and reduction**

**Nitrite transport and reduction**

**focA nitrite transporter**

**Cyanate transport and hydrolysis**

**GOS Consensus**

*COGs Matching GOS Mate Pair Reads*
FIGURE 4

Number of genes gained

Location on chromosome (Mbp)

HLII (AS9601)

ISL1 ISL2 ISL3 ISL4 ISL5

SB cynABOS SB narB1 SB GOS narB2 MIT0604 narB2

LLI (NAL1A)

PAI1, P0902-H212, P0903-H212 narB

MIT0604

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk

pyrG ppk
Table 1. Prochlorococcus strains and enrichments capable of growth in the presence of nitrate as the sole nitrogen source.

<table>
<thead>
<tr>
<th>Name</th>
<th>Clade</th>
<th>Axenic</th>
<th>Isolation Depth (m)</th>
<th>Isolation Coordinates</th>
<th>Region</th>
<th>Isolation Date</th>
<th>Assembly Size (bp)</th>
<th>Contigs</th>
<th>% GC</th>
<th>Genbank Accession</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unialgal Cultures (complete genome sequences)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB</td>
<td>HL II</td>
<td>Yes</td>
<td>40</td>
<td>35°N, 138.3°E</td>
<td>Suruga Bay, Japan</td>
<td>October 1992</td>
<td>1 668 514</td>
<td>3</td>
<td>31.5</td>
<td>JNAS0000000000</td>
<td>Shimada et al, 1995; Biller et al, 2014</td>
</tr>
<tr>
<td>MIT0604</td>
<td>HL II</td>
<td>Yes</td>
<td>175</td>
<td>22.75°N, 158°W</td>
<td>North Pacific</td>
<td>May 2006</td>
<td>1 780 061</td>
<td>1</td>
<td>31.2</td>
<td>CP007753</td>
<td>This study</td>
</tr>
<tr>
<td>PAC1</td>
<td>LL I</td>
<td>No</td>
<td>100</td>
<td>22.75°N, 158°W</td>
<td>North Pacific</td>
<td>April 1992</td>
<td>1 825 493</td>
<td>15</td>
<td>35.1</td>
<td>JNAX0000000000</td>
<td>Penno et al, 2000; Biller et al, 2014</td>
</tr>
<tr>
<td>Mixed Enrichments (partial genome assemblies)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P0902-H212</td>
<td>LL I</td>
<td>No</td>
<td>175</td>
<td>22.75°N, 158°W</td>
<td>North Pacific</td>
<td>July 2009</td>
<td>501 825</td>
<td>1</td>
<td>35.4</td>
<td>KJ947870</td>
<td>This study</td>
</tr>
<tr>
<td>P0903-H212</td>
<td>LL I</td>
<td>No</td>
<td>200</td>
<td>22.75°N, 158°W</td>
<td>North Pacific</td>
<td>July 2009</td>
<td>291 739</td>
<td>1</td>
<td>35.2</td>
<td>KJ947871</td>
<td>This study</td>
</tr>
</tbody>
</table>
SUPPLEMENTARY METHODS
Berube et al. Physiology and evolution of nitrate acquisition in *Prochlorococcus*

**DNA sequencing and assembly for the P0902-H212 and P0903-H212 enrichment cultures.** Genomic DNA from the P0902-H212 and P0903-H212 cultures was isolated using the QIAamp DNA mini kit (Qiagen, Germantown, MD, USA). 2 µg of DNA was then used to construct Illumina sequencing libraries as previously described (Rodrigue et al., 2009); this protocol used a double solid phase reversible immobilization size-selection in which the bead:sample ratios were 0.9 followed by 0.21 in order to purify fragments with an average size of ~220 bp (range: 100-300 bp). DNA libraries were sequenced on an Illumina GAIIX, yielding 200+200 nt paired-end reads, at the MIT BioMicro Center.

Low quality regions of sequencing data were removed from the raw Illumina data using quality_trim (from the CLC Assembly Cell package, CLC bio, Cambridge, MA, USA) with default settings (at least 50% of the read must be of a minimum quality of 20). Paired-end reads were overlapped using the SHE-RA algorithm (Rodrigue et al., 2010), keeping any resulting overlapping sequences with an overlap score > 0.5. Both the overlapped reads, as well as the trimmed mate pair reads that did not overlap, were assembled using clc_novo_assemble (from the CLC Assembly Cell package, CLC bio) with a minimum contig length for output set at 500 bp and the wordsize automatically determined for the input data. We identified the most “*Prochlorococcus*-like” contigs by searching each resulting contig against a custom database of sequenced marine microbial genomes (Coleman & Chisholm, 2010) using BLAST (Camacho et al., 2009). Contigs with a best match to a non-*Prochlorococcus* genome were removed from the assembly and reads mapping to only the *Prochlorococcus* contigs were then re-assembled using clc_novo_assemble with the same parameters as above.

The P0902-H212 and P0903-H212 assemblies had total lengths (3.93 and 3.95 Mb, respectively) that were approximately twice the size of previously sequenced *Prochlorococcus* genomes (Kettler et al., 2007). The contigs in each assembly were binned based on average sequencing coverage. The subset of most highly covered contigs for the P0902-H212 assembly had a total length of 1.86 Mb, with 97% of the total sequence found in contigs > 10 kb with an average sequencing coverage of 105x (± 9x, standard deviation). The subset of most highly covered contigs for the P0903-H212 assembly had a total length of 1.93 Mb with 98% of the total sequence found in contigs > 10 kb with an average sequencing coverage of 339x (± 17x, standard deviation). The highly covered subsets from each assembly...
were annotated using the RAST server (Aziz et al., 2008) with FIGfam release 49. These annotated contigs were most similar to the Prochlorococcus NATL1A genome sequence. Aligning the highly covered subsets of contigs in each assembly against the Prochlorococcus NATL1A genome using the progressiveMAUVE algorithm in MAUVE v 2.3.1 (Darling et al., 2010) revealed that the majority of contigs mapped to Prochlorococcus NATL1A.

**Identification of genes related to nitrogen and phosphorus acquisition.** Genes encoding nitrogen and phosphorus metabolism proteins (Supplementary Table 1; Supplementary Figure S5) were identified primarily from COGs (clusters of orthologous groups of proteins). However, in some cases the clustering algorithm combined or split known COGs. We used three main methods to manually curate genes related to nitrogen and phosphorus acquisition: by adjacency to subunit counterparts, phylogeny, or comparison to previously published results (Martiny et al., 2006; Martiny et al., 2009; Scanlan et al., 2009).

**Phylogenetic analysis.** The amino acid phylogeny of 56 Prochlorococcus and Synechococcus strains (Supplementary Figure S2) was reconstructed using 537 single-copy core genes that were translated to amino acid sequences and aligned individually in protein space using ClustalW (Larkin et al., 2007). Using the principle previously described (Kettler et al., 2007), we randomly concatenated 100 of these aligned amino acid sequences and built maximum likelihood (ML) and neighbor joining (NJ) phylogenies using PHYLIP v3.69 (Felsenstein, 2005). We repeated the random concatenation and tree generation 100 times.

The phylogeny of the GyrB protein was used to reconstruct the phylogeny of incomplete genomes (e.g. P0902-H212 and P0903-H212) (Supplementary Figure S3). The gyrB gene has been found to be a useful phylogenetic marker that correlates well with 16S and rpoC phylogenies (Mühling, 2012). Phylogenetic trees were estimated with PHYLIP v3.69 using the programs SEQBOOT, PROTDIST with the Jones-Taylor-Thornton matrix and without a gamma distribution of rates among sites, and NEIGHBOR on the aligned amino acid sequences with WH5701 used as an outgroup. Maximum likelihood trees were estimated on the gyrB resampled datasets using the PROML program from PHYLIP v3.69 (Felsenstein, 2005). We included the W2, W4, W7, and W8 single-cell genomes (Malmstrom et al., 2013) as well as the HNLC1 and HNLC2 metagenome assemblies (Rusch et al., 2010) as representatives of lineages from the HLIII and HLIV clades of Prochlorococcus.

The phylogeny of the cynA gene (Supplementary Figure S7) was reconstructed using reference genomes and environmental clones from the Gulf of Aqaba, northern Red Sea (Kamennaya et al., 2008). Nucleotide sequences were aligned by codon using MACSE
(Ranwez et al., 2011) and the phylogenetic analysis was conducted in MEGA5 (Tamura et al., 2011) by using the maximum likelihood method based on the Jukes-Cantor model (Jukes & Cantor, 1969). There were a total of 652 positions in the final dataset after eliminating positions containing gaps and missing data.

**Southern blotting.** For detection of narB gene copies in HLII genomes, a digoxigenin (DIG) labeled RNA probe was constructed. The narB gene from MIT0604 was amplified using the primers narB34F (5’-TGCCCWTATTGYGGTGTWGGHTG-3’) and narB2099R (5’-ATBGGRCATGWYTKYTCTTGTC-3’) at an annealing temperature of 57°C. The narB amplicon was cloned into a pCR4 plasmid vector (Life Technologies, Grand Island, NY, USA), which was then linearized by digestion with BglII (New England Biolabs, Ipswitch, MA, USA). Antisense DIG labeled RNA complimentary to the 5’ end of the MIT0604 narB gene was synthesized by run off in *vitro* transcription at 37°C for 2 hours in a reaction containing 1 µg of the linearized plasmid, 1x DIG RNA Labeling Mix (Roche Applied Science, Indianapolis, IN, USA), 1x Transcription Buffer (Roche Applied Science), 40 U of T7 RNA Polymerase (Roche Applied Science), and 20 U SUPERase-In RNase Inhibitor (Life Technologies). Labeling efficiency was estimated in a spot hybridization assay using known concentrations of DIG labeled control RNA (Roche Applied Science) and detection of narB gene from MIT0604 and SB was confirmed in a dot blot using genomic DNA and PCR amplicons of narB from each strain. All hybridizations were conducted using positively charged nylon membranes with the DIG Luminescent Detection Kit (Roche Applied Science) according to the manufacturer’s recommendations. Blots were imaged using a ChemiDoc XRS+ System (Bio-Rad Laboratories, Hercules, CA, USA). Genomic DNA from axenic cultures of MED4, MIT9301, MIT0604, and SB was separated by pulse field gel electrophoresis using a CHEF-DR II electrophoresis system (Bio-Rad Laboratories) according to the manufacturer’s recommendations. Cells were embedded in 1% agarose at a concentration of 1.5 x 10⁹ cells/mL and lysed using proteinase K and lysozyme. Genomic DNA was digested with either ApaI or BsiWI (New England Biolabs) and separated by electrophoresis for 24 hours at 14°C, 6 V/cm, an initial switch time of 1 s, and a final switch time of 25 s. DNA was blotted to a positively charged nylon membrane, probed with the DIG labeled narB probe, and imaged as described above (Supplementary Figure S4).

**Growth in the presence of urea.** Axenic cultures of Prochlorococcus SB and Prochlorococcus MIT0604 were grown in modified PRO99 media in Sargasso seawater with 50 mM NaNO₃ as the sole N source at 24°C and 30 µmol photons m⁻² s⁻¹ on a 14 hours light
and 10 hours dark cycle. At late exponential phase, each culture was transferred to replicate tubes that contained modified PRO99 media with 50 mM NH₄Cl, 50 mM urea, or no N as a control. Growth was monitored by flow cytometry using a FACSCalibur (BD Biosciences, San Jose CA, USA) and specific growth rates were estimated from the log-linear portion of the growth curve (Supplementary Figure S6).

REFERENCES FOR SUPPLEMENTARY INFORMATION


Supplementary Figure S1. Growth of axenic *Prochlorococcus* strains SB, MIT0604, MIT9301, MED4, and axenic *Synechococcus* strain WH8102 in the presence of 800 µM ammonium, nitrate, or cyanate. Bulk culture fluorescence (y-axis) was used as a proxy for cell numbers during exponential growth. Data points for the growth of parent cultures in ammonium based medium are highlighted in purple. Dashed lines represent sequential transfers in the alternative nitrogen sources. A control without added nitrogen was used to estimate when carry-over ammonium from the parent culture was completely consumed. Exponential phase during the final growth curve is highlighted in pink with the data points used for calculating growth rates connected by a line. Values are mean ± 1 standard deviation of triplicate cultures. When error bars do not show, they are within the size of the symbol.
**Supplementary Figure S2.** Maximum likelihood phylogeny of *Prochlorococcus* and *Synechococcus* proteins based on 100 resamplings of 100 randomly concatenated single-copy core proteins. Bootstrap values (total 100) were calculated using maximum likelihood (first value at each node) and neighbor joining (second value at each node), with dashes representing maximum likelihood topology unsupported by most of the neighbor joining trees.
**Supplementary Figure S3.** Phylogeny of *Prochlorococcus* and *Synechococcus* GyrB proteins. Bootstrap values (total 100) were calculated using maximum likelihood (first value at each node) and neighbor joining (second value at each node), with dashes representing maximum likelihood topology unsupported by most of the neighbor joining trees.
**Supplementary Figure S4.** Southern blot analysis confirms that *Prochlorococcus* MIT0604 contains two copies of *narB*. The ethidium bromide stained gel is shown at left and the southern blot is shown at right. The *narB* gene in MIT0604 is found on two restriction fragments of the expected sizes (100kb/197kb when digested with Apal and 62kb/155kb when digested with BsiWI). SB contains a single copy of *narB*. Arrows mark DNA fragments hybridizing to the *narB* probe.
Supplementary Figure S5. Comparison of the distribution of nitrogen and phosphorus related genes within Prochlorococcus and Synechococcus genomes to explore the relationship between nitrogen and phosphorus acquisition traits within the streamlined genomes of Prochlorococcus. Genomes are ordered based on the phylogeny in Figure 2. Box color represents % GC content. The + or – above a gene cluster denotes whether it is composed of more than one cluster or if the cluster has been manually reduced. Gray strain labels denote if a strain has been found to assimilate nitrate from culture experiments.
Supplementary Figure S6. Growth curves of Prochlorococcus SB and Prochlorococcus MIT0604 in the presence of either ammonium or urea as the sole nitrogen source. Values are mean ± 1 standard deviation of duplicate cultures. When error bars do not show, they are within the size of the symbol. Both SB and MIT0604 have the ability to grow on urea at the same rate as growth on ammonium, consistent with the presence of urease genes. When grown on urea, both strains reach final cell yields that are near double that achieved when supplied with ammonium as the sole nitrogen source (SB: $1 \times 10^8 \pm 5 \times 10^5$ cells mL$^{-1}$ on ammonium vs. $1.8 \times 10^8 \pm 6 \times 10^6$ cells mL$^{-1}$ on urea; MIT0604: $8.6 \times 10^7 \pm 1 \times 10^6$ cells mL$^{-1}$ on ammonium vs. $2.2 \times 10^8 \pm 5 \times 10^6$ cells mL$^{-1}$ on urea), indicating that both amino functional groups are removed from the urea molecule, transported into the cell and utilized for growth. Specific growth rates for SB were $0.362 \pm 0.004$ d$^{-1}$ on ammonium and $0.36 \pm 0.01$ d$^{-1}$ on urea. Specific growth rates for MIT0604 were $0.304 \pm 0.003$ d$^{-1}$ on ammonium and $0.292 \pm 0.003$ d$^{-1}$ on urea.
Supplementary Figure S7. Phylogeny of the cynA gene from reference genomes and environmental clones. The cynX-X-XX sequences correspond to those obtained by Kamennaya et al. in the Gulf of Aqaba, northern Red Sea (Kamennaya et al., 2008). The percentage of 100 replicate trees in which the associated taxa clustered together is indicated on nodes by closed circles (>75%) or open circles (>50%). Prochlorococcus SB clusters with many of the cynA clones obtained from the Red Sea indicating that these sequences were derived from the HLII clade of Prochlorococcus.
Supplementary Table 1. Genes related to nitrogen and phosphorus assimilation examined in this study.

<table>
<thead>
<tr>
<th>Gene</th>
<th>ProPortal v4.0 COG</th>
<th>Product</th>
<th>Role</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nitrogen Genes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><code>amtB/amt1</code></td>
<td>1478</td>
<td>ammonium transporter protein</td>
<td>ammonium transport</td>
<td>García-Fernández et al., 2004</td>
</tr>
<tr>
<td><code>cynA</code></td>
<td>25277</td>
<td>cyanate ABC type transporter substrate binding protein</td>
<td>cyanate transport</td>
<td>Kamennaya et al., 2008</td>
</tr>
<tr>
<td><code>cynB</code></td>
<td>17453</td>
<td>cyanate ABC type transporter permease protein</td>
<td>cyanate transport</td>
<td>Kamennaya et al., 2008</td>
</tr>
<tr>
<td><code>cynS</code></td>
<td>16887</td>
<td>cyanate lyase</td>
<td>hydrolysis of cyanate to ammonium and carbon dioxide</td>
<td>Kamennaya et al., 2008</td>
</tr>
<tr>
<td><code>focA</code></td>
<td>10584</td>
<td>nitrite transporter from formate/nitrite family</td>
<td>nitrite transport</td>
<td>Rocap et al., 2003</td>
</tr>
<tr>
<td><code>moaA</code></td>
<td>8269</td>
<td>molybdenum cofactor biosynthesis protein A</td>
<td>molybdopterin biosynthesis</td>
<td>Martiny et al., 2009</td>
</tr>
<tr>
<td><code>moaB</code></td>
<td>9123</td>
<td>molybdenum cofactor biosynthesis protein B</td>
<td>molybdopterin biosynthesis</td>
<td>Martiny et al., 2009</td>
</tr>
<tr>
<td><code>moaC</code></td>
<td>12914</td>
<td>molybdenum cofactor biosynthesis protein C</td>
<td>molybdopterin biosynthesis</td>
<td>Martiny et al., 2009</td>
</tr>
<tr>
<td><code>moaD</code></td>
<td>7626</td>
<td>molybdenum cofactor biosynthesis protein D</td>
<td>molybdopterin biosynthesis</td>
<td>Martiny et al., 2009</td>
</tr>
<tr>
<td><code>moaE</code></td>
<td>20838</td>
<td>molybdenum cofactor biosynthesis protein E</td>
<td>molybdopterin biosynthesis</td>
<td>Martiny et al., 2009</td>
</tr>
<tr>
<td><code>moeA</code></td>
<td>6195</td>
<td>molybdopterin biosynthesis protein MoeA</td>
<td>molybdopterin biosynthesis</td>
<td>Martiny et al., 2009</td>
</tr>
<tr>
<td><code>nadB</code></td>
<td>253</td>
<td>L-aspartate oxidase</td>
<td>deamination of amino acids</td>
<td>Tedeschi et al., 1996</td>
</tr>
<tr>
<td><code>napA/nrtP</code></td>
<td>5121</td>
<td>nitrate/nitrite transporter</td>
<td>nitrate/nitrite transport</td>
<td>Martiny et al., 2009b; Wang et al., 2000; Bird &amp; Wyman, 2003</td>
</tr>
<tr>
<td><code>narB</code></td>
<td>3405</td>
<td>assimilatory nitrate reductase</td>
<td>nitrate reduction to nitrite</td>
<td>Martiny et al., 2009</td>
</tr>
<tr>
<td><code>narX1</code></td>
<td>12460</td>
<td>conserved hypothetical protein</td>
<td>unknown function</td>
<td>Martiny et al., 2009</td>
</tr>
<tr>
<td><code>narX2</code></td>
<td>30465, 26956, 33277</td>
<td>conserved hypothetical protein</td>
<td>unknown function</td>
<td>Martiny et al., 2009</td>
</tr>
<tr>
<td><code>nirA</code></td>
<td>5136</td>
<td>ferredoxin nitrite reductase</td>
<td>nitrite reduction to ammonium</td>
<td>Martiny et al., 2009</td>
</tr>
<tr>
<td><code>nirX</code></td>
<td>27176, 11823</td>
<td>conserved hypothetical protein</td>
<td>unknown function</td>
<td>Martiny et al., 2009</td>
</tr>
<tr>
<td><code>thiO</code></td>
<td>772</td>
<td>glycine oxidase</td>
<td>deamination of amino acids</td>
<td>Nishiya &amp; Imanaka, 1998</td>
</tr>
<tr>
<td><code>ureA</code></td>
<td>1864</td>
<td>urease subunit alpha</td>
<td>hydrolysis of urea to ammonium and carbon dioxide</td>
<td>Palinska et al., 2000</td>
</tr>
<tr>
<td>Gene</td>
<td>ProPortal v4.0 COG</td>
<td>Product</td>
<td>Role</td>
<td>Reference</td>
</tr>
<tr>
<td>----------</td>
<td>--------------------</td>
<td>----------------------------------------------</td>
<td>-----------------------------------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>PMM0707</td>
<td>30300, 31904</td>
<td>hypothetical protein</td>
<td>expressed in MED4 during phosphorus starvation</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>PMM0715</td>
<td>26328</td>
<td>hypothetical protein</td>
<td>expressed in MED4 during phosphorus starvation</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>PMM0717</td>
<td>32234</td>
<td>hypothetical protein</td>
<td>expressed in MED4 during phosphorus starvation</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>PMM0719</td>
<td>3650</td>
<td>hypothetical protein</td>
<td>expressed in MED4 during phosphorus starvation</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>PMM0720</td>
<td>28615</td>
<td>hypothetical protein</td>
<td>expressed in MED4 during phosphorus starvation</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>PMM0721</td>
<td>28631</td>
<td>hypothetical protein</td>
<td>expressed in MED4 during phosphorus starvation</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>PMM0722</td>
<td>2536</td>
<td>hypothetical protein</td>
<td>expressed in MED4 during phosphorus starvation</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>arsA</td>
<td>22394</td>
<td>arsenite efflux pump subunit</td>
<td>arsenate resistance</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>arsR</td>
<td>1361</td>
<td>arsenate reductase</td>
<td>arsenate resistance</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>carA</td>
<td>20</td>
<td>carbamoyl phosphate synthetase small subunit</td>
<td>carbamoyl phosphate synthesis</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>carB</td>
<td>346</td>
<td>carbamoyl phosphate synthetase large subunit</td>
<td>carbamoyl phosphate synthesis</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>chrA</td>
<td>13381</td>
<td>response regulator</td>
<td>chromate resistance</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>gap1</td>
<td>99</td>
<td>glyceraldehyde-3-phosphate dehydrogenase</td>
<td>expressed in MED4 during phosphorus starvation</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>mfs</td>
<td>817</td>
<td>major facilitator superfamily transporter</td>
<td>expressed in MED4 during phosphorus starvation</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>prpB</td>
<td>6142</td>
<td>phosphoenolpyruvate mutase</td>
<td>phosphonate biosynthesis</td>
<td>Yu et al., 2013</td>
</tr>
<tr>
<td>phnC</td>
<td>506</td>
<td>phosphonate ABC type transporter ATP binding protein</td>
<td>phosphonate transport</td>
<td>Feingersch et al., 2012; Martinez et al., 2010</td>
</tr>
<tr>
<td>phnD</td>
<td>4518</td>
<td>phosphonate ABC type transporter substrate binding protein</td>
<td>phosphonate transport</td>
<td>Feingersch et al., 2012; Martinez et al., 2010</td>
</tr>
<tr>
<td>phoA</td>
<td>15427, 26745</td>
<td>alkaline phosphatase</td>
<td>dephosphorylation</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>phoB</td>
<td>204</td>
<td>phosphate regulon response regulator</td>
<td>phosphate two component regulatory system</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>phoR</td>
<td>13582</td>
<td>phosphate regulon sensor histidine kinase</td>
<td>phosphate two component regulatory system</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>phoX</td>
<td>26697</td>
<td>alkaline phosphatase</td>
<td>dephosphorylation</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>pstA</td>
<td>3725</td>
<td>phosphate ABC type transporter permease protein</td>
<td>phosphate transport</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>pstB</td>
<td>88</td>
<td>phosphate ABC type transporter ATP binding protein</td>
<td>phosphate transport</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>pstC</td>
<td>4183, 30634</td>
<td>phosphate ABC type transporter permease protein</td>
<td>phosphate transport</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>pstS</td>
<td>1827</td>
<td>phosphate ABC type transporter substrate binding protein</td>
<td>phosphate transport</td>
<td>Martiny et al., 2006</td>
</tr>
<tr>
<td>ptrA</td>
<td>37989, 6860, 11384</td>
<td>transcriptional regulator</td>
<td>stress response to phosphorus starvation</td>
<td>Ostrowski et al., 2010</td>
</tr>
<tr>
<td>sphX</td>
<td>25109</td>
<td>phosphate binding protein</td>
<td>phosphate transport</td>
<td>Mann &amp; Scanlan, 1994</td>
</tr>
</tbody>
</table>
References for Supplementary Table 1:


