ACCEPTED AUTHOR MANUSCRIPT doi:10.1038/ismej.2014.211

Physiology and evolution of nitrate acquisition in *Prochlorococcus*

Paul M Berube[1,*], Steven J Biller[1], Alyssa G Kent[2], Jessie W Berta-Thompson[1,3], Sara E Roggensack[1], Kathryn H Roache-Johnson[4,5], Marcia Ackerman[5], Lisa R Moore[5], Joshua D Meisel[6], Daniel Sher[7], Luke R Thompson[8], Lisa Campbell[9], Adam C Martiny[2,10], and Sallie W Chisholm[1,6,*]

[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

ABSTRACT

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; Zwirglmaier et al., 2007; Zwirglmaier 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

102 and examined their growth on different nitrogen sources. We then used comparative genomics
103 to better understand how this trait had evolved in *Prochlorococcus*.
104

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 ß-glycerophosphate replacing inorganic phosphate, 0.01 µM Na_2MoO_4 and 0.05 µM $NiCl_2$. 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

130 phase cultures were enumerated using an Influx Cell Sorter (BD Biosciences, San Jose CA,
131 USA) or a FACSCalibur flow cytometer (BD Biosciences) as previously described (Olson et
132 al., 1985; Cavender-Bares et al., 1999). Cultures consisting of >80% *Prochlorococcus* cells
133 were serially diluted into multiple multi-well plates at final concentrations of 1-10 cells per
134 well in at least 200 μL of ProMM medium. Axenic *Prochlorococcus* do not grow from such
135 low cell densities in Pro99 medium without "helper" heterotrophic bacteria (Morris et al.,
136 2008; Morris et al., 2011), however, they do grow when diluted into ProMM. The main
137 ingredient in ProMM which promotes the growth of cells from low densities is pyruvate, and
138 we suspect that in this context pyruvate serves as a potent hydrogen peroxide scavenger
139 (Giandomenico et al., 1997). Wells contaminated with heterotrophic bacteria were identified
140 by the appearance of turbidity. The multi-well plates were monitored by eye and by
141 fluorometry using a Synergy HT Microplate Reader (BioTek, Winooski, VT, USA), and non-
142 turbid wells were monitored by flow cytometry using a FACSCalibur flow cytometer. Wells
143 that appeared green or had *Prochlorococcus* cells as determined by flow cytometry were
144 immediately transferred to Pro99 medium directly, or into fresh ProMM medium until
145 consistent growth was observed, at which point the cultures were introduced back into Pro99
146 medium. Cultures were examined for heterotrophic bacteria contaminants by flow cytometry
147 and by inoculation into the panel of purity test broths as described above.

148 ***PCR screen for the nitrate reductase gene.*** Based on an alignment of GOS reads
149 coding for the *Prochlorococcus narB* sequence (Martiny et al., 2009b), degenerate primers
150 30narB175f (5'-TGYGTDAAAGGMGCAACAGTNTG-3') and 30narB574r (5'-
151 GACAYTCWGCBGTATTWGTHCC-3') were designed specifically to amplify the *narB*
152 gene from HLII clade *Prochlorococcus*, and degenerate primers 40narB1447f (5'-
153 TATTGYCCAGCWTTYMGDCCDTG-3') and 40narB1766r (5'-
154 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 $MgCl_2$, 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 *narB* nitrate reductase gene and 2 gains and 3 losses for the *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 *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

296 in *Prochlorococcus* compared to marine *Synechococcus*, we were curious as to whether we
297 might find evidence for rearrangements or lateral gene transfer.

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

310 Analysis of metagenomic data from GOS (Martiny et al., 2009b) suggested that the
311 nitrate utilization genes in HLII *Prochlorococcus* should be located in a different genomic
312 region compared to LLI genomes, indicating an alternative evolutionary origin. Based on a
313 scaffold of mate-paired metagenomic reads, it was inferred that this cluster should be located
314 approximately 500 kb downstream of the *pyrG-ppk* region containing the nitrate assimilation
315 genes in WH8102 and the nitrite assimilation genes in NATL1A (Martiny et al., 2009b). We
316 found a high degree of similarity between the nitrate assimilation gene cluster in SB and the
317 scaffold derived from GOS metagenome sequences obtained from multiple individual cells
318 from multiple sampling stations. This similarity manifested itself not only in the gene order
319 and chromosomal location, but also the phylogeny of the nitrate assimilation genes (Figures
320 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,

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

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

416

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).

459 REFERENCES

460 Abascal F, Zardoya R, Telford MJ. (2010). TranslatorX: multiple alignment of nucleotide
461 sequences guided by amino acid translations. Nucleic Acids Res 38: W7-13.

462 Ahlgren NA, Rocap G. (2006). Culture isolation and culture-independent clone libraries
463 reveal new marine *Synechococcus* ecotypes with distinctive light and N physiologies. Appl
464 Environ Microbiol 72: 7193-7204.

465 Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, *et al.* (2008). The RAST
466 Server: rapid annotations using subsystems technology. BMC Genomics 9: 75.

467 Biller SJ, Berube PM, Berta-Thompson JW, Kelly L, Roggensack SE, Awad L, *et al.* (2014).
468 Genomes of diverse isolates of the marine cyanobacterium *Prochlorococcus.* Scientific Data
469 1: 140034.

470 Bouman HA, Ulloa O, Scanlan DJ, Zwirglmaier K, Li WK, Platt T, *et al.* (2006).
471 Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes. Science
472 312: 918-921.

473 Bragg JG, Dutkiewicz S, Jahn O, Follows MJ, Chisholm SW. (2010). Modeling selective
474 pressures on phytoplankton in the global ocean. PLoS ONE 5: e9569.

475 Buitenhuis ET, Li WKW, Vaulot D, Lomas MW, Landry MR, Partensky F, *et al.* (2012).
476 Picophytoplankton biomass distribution in the global ocean. Earth Syst Sci Data 4: 37-46.

477 Campbell L, Nolla HA, Vaulot D. (1994). The importance of *Prochlorococcus* to community
478 structure in the central North Pacific Ocean. Limnol Oceanogr 39: 954-961.

479 Casey JR, Lomas MW, Mandecki J, Walker DE. (2007). *Prochlorococcus* contributes to new
480 production in the Sargasso Sea deep chlorophyll maximum. Geophys Res Lett 34: L10604.

481 Cavender-Bares KK, Mann EL, Chisholm SW, Ondrusek ME, Bidigare RR. (1999).
482 Differential response of equatorial Pacific phytoplankton to iron fertilization. Limnol
483 Oceanogr 44: 237-246.

484 Chisholm SW, Frankel SL, Goericke R, Olson RJ, Palenik B, Waterbury JB, *et al.* (1992).
485 *Prochlorococcus marinus* nov. gen. nov. sp.: an oxyphototrophic marine prokaryote
486 containing divinyl chlorophyll *a* and *b*. Arch Microbiol 157: 297-300.

487 Coleman ML, Chisholm SW. (2007). Code and context: *Prochlorococcus* as a model for
488 cross-scale biology. Trends Microbiol 15: 398-407.

489 Coleman ML, Chisholm SW. (2010). Ecosystem-specific selection pressures revealed through
490 comparative population genomics. Proc Natl Acad Sci USA 107: 18634-18639.

491 Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, DeLong EF, Chisholm SW.
492 (2006). Genomic islands and the ecology and evolution of *Prochlorococcus*. Science 311:
493 1768-1770.

494 Collier JL, Lovindeer R, Xi Y, Radway JC, Armstrong RA. (2012). Differences in growth and
495 physiology of marine *Synechococcus* (Cyanobacteria) on nitrate versus ammonium are not
496 determined solely by nitrogen source redox state. J Phycol 48: 106-116.

497 Dufresne A, Salanoubat M, Partensky F, Artiguenave F, Axmann IM, Barbe V, *et al.* (2003).
498 Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal
499 oxyphototrophic genome. Proc Natl Acad Sci USA 100: 10020-10025.

500 Feingersch R, Philosof A, Mejuch T, Glaser F, Alalouf O, Shoham Y, Béjà O. (2012).
501 Potential for phosphite and phosphonate utilization by *Prochlorococcus*. ISME J 6: 827-834.

502 Felsenstein J. (2005). PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the
503 Author. Department of Genome Sciences, University of Washington, Seattle.

504 Flombaum P, Gallegos JL, Gordillo RA, Rincón J, Zabala LL, Jiao N, *et al.* (2013). Present
505 and future global distributions of the marine Cyanobacteria *Prochlorococcus* and
506 *Synechococcus*. Proc Natl Acad Sci USA. 110: 9824-9829.

507 Follows MJ, Dutkiewicz S, Grant S, Chisholm SW. (2007). Emergent biogeography of
508 microbial communities in a model ocean. Science 315: 1843-1846.

509 Fuller NJ, Marie D, Partensky F, Vaulot D, Post AF, Scanlan DJ. (2003). Clade-specific 16S
510 ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus*
511 clade throughout a stratified water column in the Red Sea. Appl Environ Microbiol 69: 2430-
512 2443.

513 Fuller NJ, West NJ, Marie D, Yallop M, Rivlin T, Post AF, Scanlan DJ. (2005). Dynamics of
514 community structure and phosphate status of picocyanobacterial populations in the Gulf of
515 Aqaba, Red Sea. Limnol Oceanogr 50: 363-375.

516 García-Fernández JM, de Marsac NT, Diez J. (2004). Streamlined regulation and gene loss as
517 adaptive mechanisms in *Prochlorococcus* for optimized nitrogen utilization in oligotrophic
518 environments. Microbiol Mol Biol Rev 68: 630-638.

519 Giandomenico AR, Cerniglia GE, Biaglow JE, Stevens CW, Koch CJ. (1997). The
520 importance of sodium pyruvate in assessing damage produced by hydrogen peroxide. Free
521 Radical Bio Med 23: 426-434.

522 Gruber N. (2008). The marine nitrogen cycle: overview and challenges. In: Nitrogen in the
523 Marine Environment. Capone DG, Bronk DA, Mulholland MR, Carpenter EJ (eds).
524 Academic Press: Burlington, MA, pp 1-50.

525 Herrero A, Muro-Pastor AM, Flores E. (2001). Nitrogen control in Cyanobacteria. Biochim
526 Biophys Acta 183: 411-425.

527 Huang S, Wilhelm SW, Harvey HR, Taylor K, Jiao N, Chen F. (2012). Novel lineages of
528 *Prochlorococcus* and *Synechococcus* in the global oceans. ISME J 6: 285-297.

529 Iwata T, Shinomura Y, Natori Y, Igarashi Y, Sohrin R, Suzuki Y. (2005). Relationship
530 between salinity and nutrients in the subsurface layer in the Suruga Bay. J Oceanogr 61: 721-
531 732.

532 Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EMS, Chisholm SW. (2006). Niche
533 partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients.
534 Science 311: 1737-1740.

535 Kamennaya NA, Chernihovsky M, Post AF. (2008). The cyanate utilization capacity of
536 marine unicellular Cyanobacteria. Limnol Oceanogr 53: 2485-2494.

537 Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, *et al.* (2014).
538 Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*.
539 Science 344: 416-420.

540 Kelly L, Huang KH, Ding H, Chisholm SW. (2012). ProPortal: a resource for integrated
541 systems biology of *Prochlorococcus* and its phage. Nucleic Acids Res 40: D632-D640.

542 Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, Rodrigue S, *et al.* (2007).
543 Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus.* PLoS
544 Genet 3: e231.

545 Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, *et al.*
546 (2007). Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948.

547 Lavin P, González B, Santibáñez JF, Scanlan DJ, Ulloa O. (2010). Novel lineages of
548 *Prochlorococcus* thrive within the oxygen minimum zone of the eastern tropical South
549 Pacific. Environ Microbiol Rep 2: 728-738.

550 Lomas MW, Lipschultz F. (2006). Forming the primary nitrite maximum: nitrifiers or
551 phytoplankton? Limnol Oceanogr 51: 2453-2467.

552 Malmstrom RR, Coe A, Kettler GC, Martiny AC, Frias-Lopez J, Zinser ER, Chisholm SW.
553 (2010). Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans.
554 ISME J 4: 1252–1264.

555 Malmstrom RR, Rodrigue S, Huang KH, Kelly L, Kern SE, Thompson A, *et al.* (2013).
556 Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and
557 biogeographic analysis. ISME J 7: 184-198.

558 Mann EL, Chisholm SW. (2000). Iron limits the cell division rate of *Prochlorococcus* in the
559 eastern equatorial Pacific. Limnol Oceanogr 45: 1067-1076.

560 Martiny AC, Coleman ML, Chisholm SW. (2006). Phosphate acquisition genes in
561 *Prochlorococcus* ecotypes: evidence for genome-wide adaptation. Proc Natl Acad Sci USA
562 103: 12552-12557.

563 Martiny AC, Huang Y, Li W. (2009a). Occurrence of phosphate acquisition genes in
564 *Prochlorococcus* cells from different ocean regions. Environ Microbiol 11: 1340–1347.

565 Martiny AC, Kathuria S, Berube PM. (2009b). Widespread metabolic potential for nitrite and
566 nitrate assimilation among *Prochlorococcus* ecotypes. Proc Natl Acad Sci USA 106: 10787-
567 10792.

568 Martiny AC, Tai AP, Veneziano D, Primeau F, Chisholm SW. (2009c). Taxonomic
569 resolution, ecotypes and the biogeography of *Prochlorococcus*. Environ Microbiol 11: 823-
570 832.

571 Marty J-C, Chiavérini J, Pizay M-D, Avril B. (2002). Seasonal and interannual dynamics of
572 nutrients and phytoplankton pigments in the western Mediterranean Sea at the DYFAMED
573 time-series station (1991-1999). Deep-Sea Res Part II-Top Stud Oceanogr

574 49: 1965-1985.

575 Mirkin BG, Fenner TI, Galperin MY, Koonin EV. (2003). Algorithms for computing
576 parsimonious evolutionary scenarios for genome evolution, the last universal common
577 ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. BMC
578 Evol Biol 3: 2.

579 Moore LR, Chisholm SW. (1999). Photophysiology of the marine cyanobacterium
580 *Prochlorococcus*: ecotypic differences among cultured isolates. Limnol Oceanogr 44: 628-
581 638.

582 Moore LR, Coe A, Zinser ER, Saito MA, Sullivan MB, Lindell D, *et al.* (2007). Culturing the
583 marine cyanobacterium *Prochlorococcus*. Limnol Oceanogr Meth 5: 353-362.

584 Moore LR, Post AF, Rocap G, Chisholm SW. (2002). Utilization of different nitrogen sources
585 by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. Limnol Oceanogr 47: 989-
586 996.

587 Moore LR, Rocap G, Chisholm SW. (1998). Physiology and molecular phylogeny of
588 coexisting *Prochlorococcus* ecotypes. Nature 393: 464-467.

589 Moore CM, Mills MM, Arrigo KR, Berman-Frank I, Bopp L, Boyd PW, *et al.* (2013).
590 Processes and patterns of oceanic nutrient limitation. Nat Geosci 6: 701-710.

591 Morris JJ, Johnson ZI, Szul MJ, Keller M, Zinser ER. (2011). Dependence of the
592 cyanobacterium *Prochlorococcus* on hydrogen peroxide scavenging microbes for growth at
593 the ocean's surface. PLoS ONE 6: e16805.

594 Morris JJ, Kirkegaard R, Szul MJ, Johnson ZI, Zinser ER. (2008). Facilitation of robust
595 growth of *Prochlorococcus* colonies and dilute liquid cultures by "helper" heterotrophic
596 bacteria. Appl Environ Microbiol 74: 4530-4534.

597 Mühling M. (2012). On the culture-independent assessment of the diversity and distribution
598 of *Prochlorococcus*. Environ Microbiol 14: 567-579.

599 Ohashi Y, Shi W, Takatani N, Aichi M, Maeda SI, Watanabe S, *et al.* (2011). Regulation of
600 nitrate assimilation in cyanobacteria. J Exp Bot 62: 1411-1424.

601 Olson RJ, Chisholm SW, Zettler ER, Altabet MA, Dusenberry JA. (1990). Spatial and
602 temporal distributions of prochlorophyte picoplankton in the North Atlantic Ocean. Deep-Sea
603 Res Part I-Oceanogr Res Pap 37: 1033-1051.

604 Olson RJ, Vaulot D, Chisholm SW. (1985). Marine phytoplankton distributions measured
605 using shipboard flow cytometry. Deep-Sea Res Part I-Oceanogr Res Pap 32: 1273-1280.

606 Partensky F, Blanchot J, Vaulot D. (1999). Differential distribution and ecology of
607 *Prochlorococcus* and *Synechococcus* in oceanic waters: a review. In: Charpy L and Larkum
608 AWD (eds). *Marine Cyanobacteria*. Bulletin de l'Institut oceanographique de Monaco, No.
609 spécial 19, Musée océanographique: Monaco, pp 457-475.

610 Partensky F, Garczarek L. (2010). *Prochlorococcus*: advantages and limits of minimalism.
611 Annu Rev Mar Sci 2: 305-331.

612 Penno S, Campbell L, Hess WR. (2000). Presence of phycoerythrin in two strains of
613 *Prochlorococcus* (Cyanobacteria) isolated from the subtropical north Pacific Ocean. J Phycol
614 36: 723-729.

615 Rocap G, Distel DL, Waterbury JB, Chisholm SW. (2002). Resolution of *Prochlorococcus*
616 and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer
617 sequences. Appl Environ Microbiol 68: 1180-1191.

618 Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA, *et al.* (2003). Genome
619 divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. Nature
620 424: 1042-1047.

621 Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, *et al.* (2007).
622 The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern
623 tropical Pacific. PLoS Biol 5: e77.

624 Rusch DB, Martiny AC, Dupont CL, Halpern AL, Venter JC. (2010). Characterization of
625 *Prochlorococcus* clades from iron-depleted oceanic regions. Proc Natl Acad Sci USA 107:
626 16184-16189.

627 Saito MA, Moffett JW, Chisholm SW, Waterbury JB. (2002). Cobalt limitation and uptake in
628 *Prochlorococcus*. Limnol Oceanogr 47: 1629-1636.

629 Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR, *et al.* (2009).
630 Ecological genomics of marine picocyanobacteria. Microbiol Mol Biol Rev 73: 249-299.

631 Scanlan DJ, West NJ. (2002). Molecular ecology of the marine cyanobacterial genera
632 *Prochlorococcus* and *Synechococcus*. FEMS Microbiol Ecol 40: 1-12.

633 Shimada A, Nishijima M, Maruyama T. (1995). Seasonal appearance of *Prochlorococcus* in
634 Suruga Bay, Japan. J Oceanogr 51: 289-300.

635 Shi Y, Tyson GW, Eppley JM, Delong EF. (2011). Integrated metatranscriptomic and
636 metagenomic analyses of stratified microbial assemblages in the open ocean. ISME J 5: 999-
637 1013.

638 Sukumaran J, Holder MT. (2010). DendroPy: a Python library for phylogenetic computing.
639 Bioinformatics 26: 1569-1571.

640 Tolonen AC, Aach J, Lindell D, Johnson ZI, Rector T, Steen R, *et al.* (2006). Global gene
641 expression of Prochlorococcus ecotypes in response to changes in nitrogen availability. Mol
642 Syst Biol 2: 53.

643 Tyrrell T. (1999). The relative influences of nitrogen and phosphorus on oceanic primary
644 production. Nature 400: 525-531.

645 Vaulot D, Lebot N, Marie D, Fukai E. (1996). Effect of phosphorus on the *Synechococcus*
646 cell cycle in surface Mediterranean waters during summer. Appl Environ Microbiol 62: 2527-
647 2533.

648 Vaulot D, Partensky F. (1992). Cell cycle distributions of prochlorophytes in the north
649 western Mediterranean Sea. Deep-Sea Res Part I-Oceanogr Res Pap 39: 727-742.

650 Waterbury JB, Willey JM. (1988). Isolation and growth of marine planktonic cyanobacteria.
651 Methods Enzymol 167: 100-105.

652 Williams EZ, Campbell L, DiTullio G. (1999). The nitrogen specific uptake of three strains of
653 *Prochlorococcus.* Presented at the American Society of Limnology and Oceanography
654 Aquatic Sciences Meeting. February 4, 1999, Santa Fe, NM.

655 Williams KP. (2002). Integration sites for genetic elements in prokaryotic tRNA and tmRNA
656 genes: sublocation preference of integrase subfamilies. Nucleic Acids Res 30: 866-875.

657 Wu J, Sunda W, Boyle EA, Karl DM. (2000). Phosphate depletion in the western North
658 Atlantic Ocean. Science 289: 759-762.

659 Zinser ER, Coe A, Johnson ZI, Martiny AC, Fuller NJ, Scanlan DJ, Chisholm SW. (2006).
660 *Prochlorococcus* ecotype abundances in the North Atlantic Ocean as revealed by an improved
661 quantitative PCR method. Appl Environ Microbiol 72: 723-732.

662 Zwirglmaier K, Heywood JL, Chamberlain K, Woodward EM, Zubkov MV, Scanlan DJ.
663 (2007). Basin-scale distribution patterns of picocyanobacterial lineages in the Atlantic Ocean.
664 Environ Microbiol 9: 1278-1290.

665 Zwirglmaier K, Jardillier L, Ostrowski M, Mazard S, Garczarek L, Vaulot D, *et al.* (2008).
666 Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct
667 partitioning of lineages among oceanic biomes. Environ Microbiol 10: 147-161.

668

669 TITLES AND LEGENDS TO FIGURES

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

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

683 Figure 3. Architecture of the nitrite and nitrate assimilation genes in low-light adapted (LLI
684 clade) and high-light adapted (HLII clade) *Prochlorococcus* relative to *Synechococcus*
685 WH8102. Similar to *Synechococcus*, the nitrite and nitrate assimilation genes in the LLI clade
686 of *Prochlorococcus* are found within the region between the *pyrG* (pyrimidine biosynthesis)
687 and *ppk* (polyphosphate kinase) genes. Most LLI clade *Prochlorococcus*, with the exception
688 of the P0903-H212 contig, possess a *focA* nitrite transporter in this region (possibly acquired
689 from proteobacteria (Rocap et al., 2003)). Metagenome data (Martiny et al., 2009b), a partial
690 genome from a single cell (B241-528J8) (Kashtan et al., 2014), and a culture genome
691 (*Prochlorococcus* SB) indicate that the nitrate assimilation genes within HLII clade
692 *Prochlorococcus* are commonly found in a syntenic region adjacent to genomic island ISL3

693 (see Figure 4). *Prochlorococcus* MIT0604 is an exception in that it possesses duplicate nitrate
694 assimilation gene clusters located within genomic islands ISL3 and ISL4 (see Figure 4), with
695 phage integrase genes immediately adjacent to each copy of the *nirA* (nitrite reductase) gene.

696 Figure 4. Locations of nitrate and cyanate assimilation genes in strains of *Prochlorococcus*
697 capable of nitrate assimilation relative to the known genomic islands (shaded regions)
698 observed in the HLII and LLI clades of *Prochlorococcus*; plots modified from Kettler et al.,
699 2007. *Prochlorococcus* genomes are highly syntenic and genomic islands have been
700 identified in high-light adapted genomes (e.g. AS9601) by conserved breaks in gene synteny
701 among strains (Coleman et al., 2006; Kettler et al., 2007). Genomic islands have also been
702 identified (e.g. the large region within LLI clade genomes such as NATL1A) by predicted
703 gene gain events along the chromosome (Kettler et al., 2007).

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

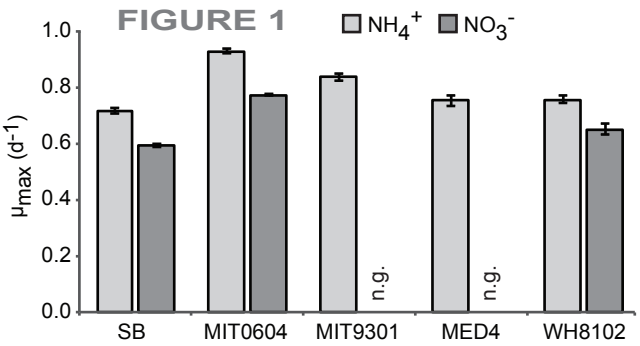


FIGURE 1
NH4+
NO3-
μmax (d-1)
1.0
0.8
0.6
0.4
0.2
0.0
n.g.
n.g.
SB
MIT0604
MIT9301
MED4
WH8102

FIGURE 2

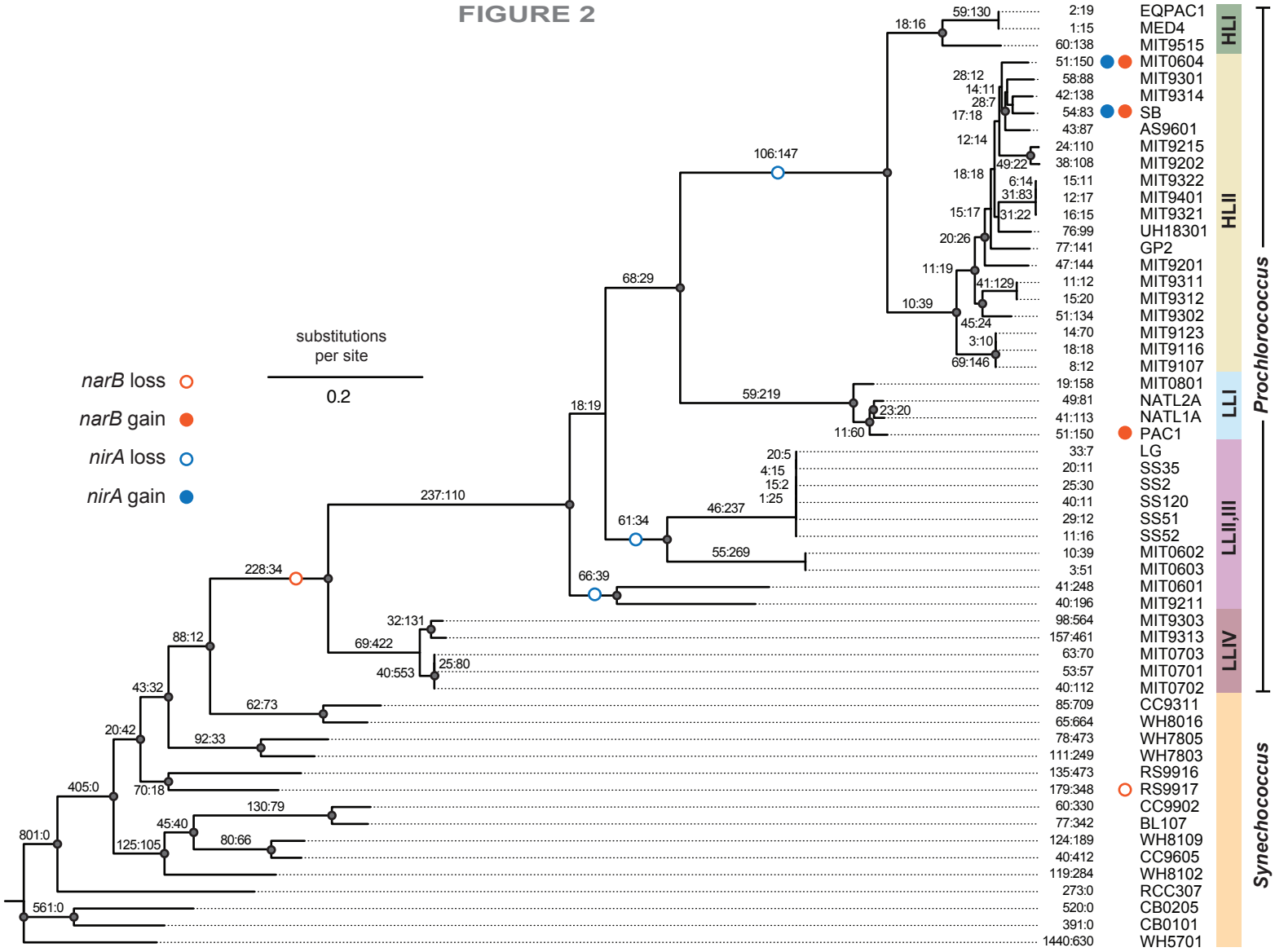

narB loss
narB gain
nirA loss
nirA gain
substitutions per site
0.2
EQPAC1 2:19
MED4 1:15
MIT9515 60:138
MIT0604 51:150
MIT9301 58:88
MIT9314 42:138
SB 54:83
AS9601 43:87
MIT9215 24:110
MIT9202 38:108
MIT9322 15:11
MIT9401 12:17
MIT9321 16:15
UH18301 76:99
GP2 77:141
MIT9201 47:144
MIT9311 11:12
MIT9312 15:20
MIT9302 51:134
MIT9123 14:70
MIT9116 18:18
MIT9107 8:12
MIT0801 19:158
NATL2A 49:81
NATL1A 41:113
PAC1 51:150
LG 33:7
SS35 20:11
SS2 25:30
SS120 40:11
SS51 29:12
SS52 11:16
MIT0602 10:39
MIT0603 3:51
MIT0601 41:248
MIT9211 40:196
MIT9303 98:564
MIT9313 157:461
MIT0703 63:70
MIT0701 53:57
MIT0702 40:112
CC9311 85:709
WH8016 65:664
WH7805 78:473
WH7803 111:249
RS9916 135:473
RS9917 179:348
CC9902 60:330
BL107 77:342
WH8109 124:189
CC9605 40:412
WH8102 119:284
RCC307 273:0
CB0205 520:0
CB0101 391:0
WH5701 1440:630
HLI
HLII
LLI
LLII,III
LLIV
Prochlorococcus
Synechococcus

FIGURE 3

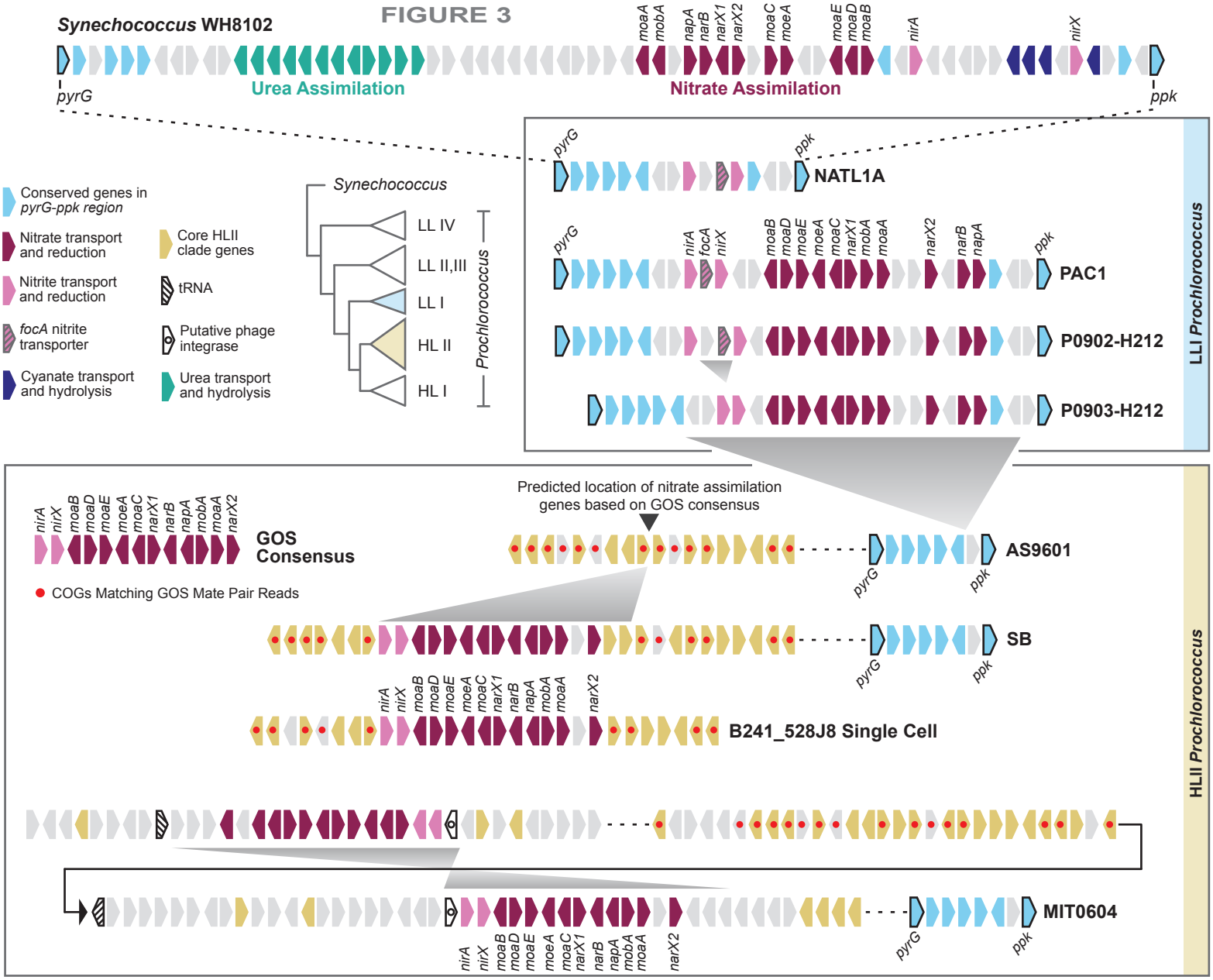

Synechococcus WH8102
pyrG
Urea Assimilation
moaA mobA napA narB narX1 narX2 moaC moeA moaE moaD moaB nirA
Nitrate Assimilation
nirX
ppk
Conserved genes in pyrG-ppk region
Nitrate transport and reduction
Nitrite transport and reduction
focA nitrite transporter
Cyanate transport and hydrolysis
Core HLII clade genes
tRNA
Putative phage integrase
Urea transport and hydrolysis
Synechococcus
LL IV
LL II,III
LL I
HL II
HL I
Prochlorococcus
NATL1A
PAC1
P0902-H212
P0903-H212
LLI Prochlorococcus
Predicted location of nitrate assimilation genes based on GOS consensus
GOS Consensus
COGs Matching GOS Mate Pair Reads
AS9601
SB
B241_528J8 Single Cell
MIT0604
HLII Prochlorococcus

FIGURE 4

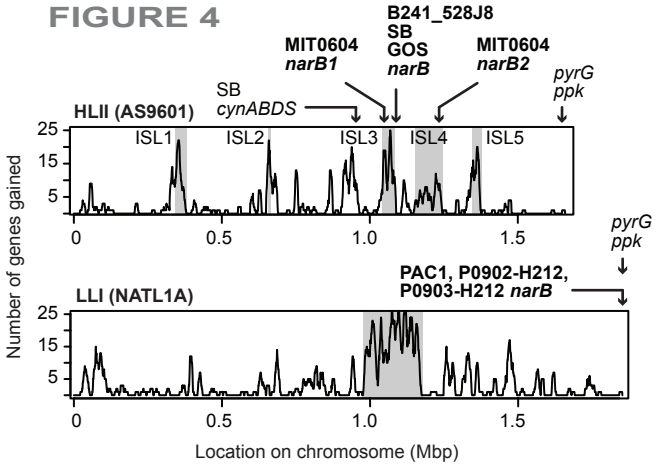

B241_528J8
SB
GOS
narB
MIT0604
narB1
MIT0604
narB2
pyrG
ppk
SB
cynABDS
HLII (AS9601)
ISL1
ISL2
ISL3
ISL4
ISL5
25
15
5
0
0.5
1.0
1.5
pyrG
ppk
PAC1, P0902-H212,
P0903-H212 narB
LLI (NATL1A)
Number of genes gained
Location on chromosome (Mbp)

FIGURE 5

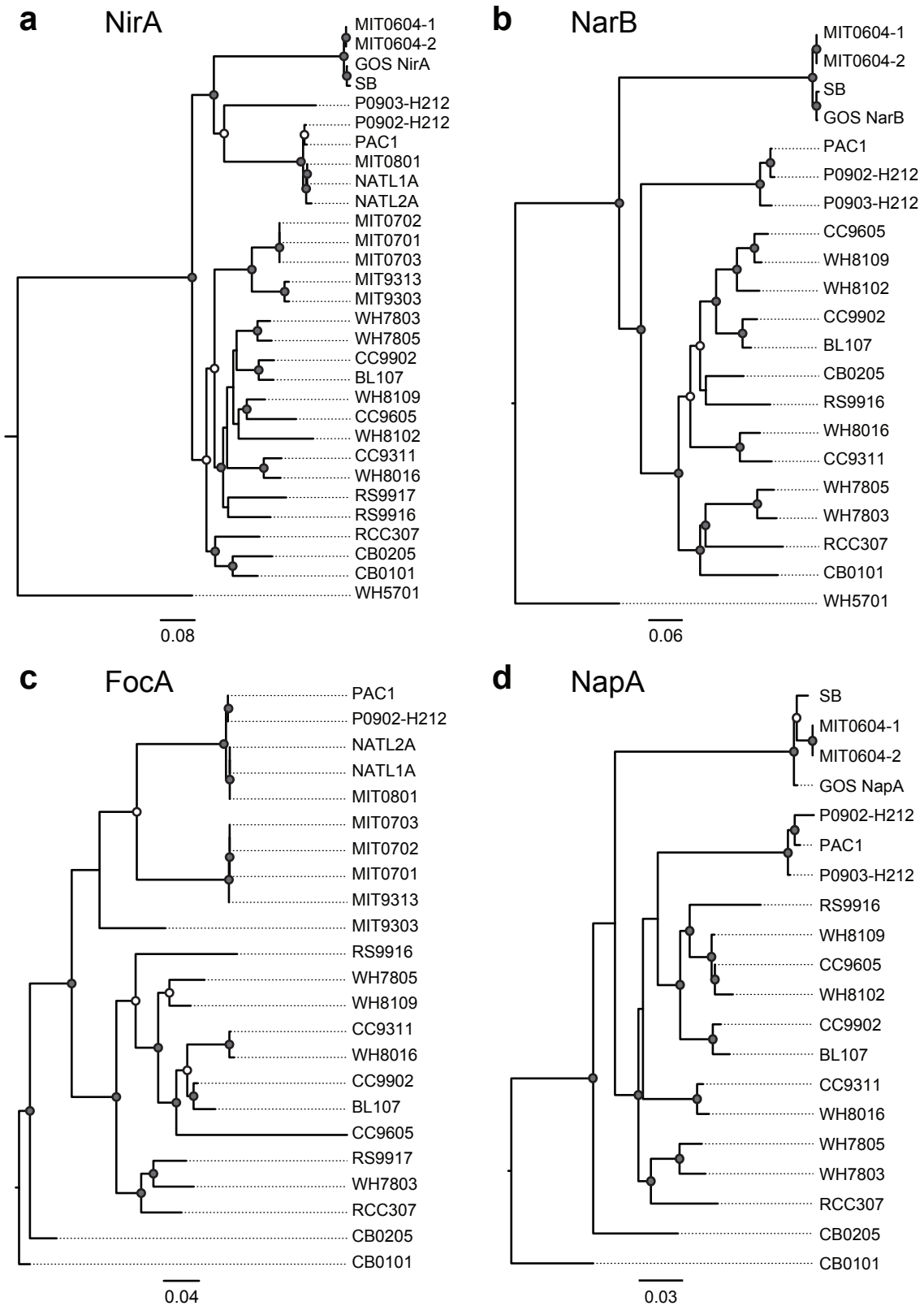

a
NirA
MIT0604-1
MIT0604-2
GOS NirA
SB
P0903-H212
P0902-H212
PAC1
MIT0801
NATL1A
NATL2A
MIT0702
MIT0701
MIT0703
MIT9313
MIT9303
WH7803
WH7805
CC9902
BL107
WH8109
CC9605
WH8102
CC9311
WH8016
RS9917
RS9916
RCC307
CB0205
CB0101
WH5701
0.08
b
NarB
MIT0604-1
MIT0604-2
SB
GOS NarB
PAC1
P0902-H212
P0903-H212
CC9605
WH8109
WH8102
CC9902
BL107
CB0205
RS9916
WH8016
CC9311
WH7805
WH7803
RCC307
CB0101
WH5701
0.06
c
FocA
PAC1
P0902-H212
NATL2A
NATL1A
MIT0801
MIT0703
MIT0702
MIT0701
MIT9313
MIT9303
RS9916
WH7805
WH8109
CC9311
WH8016
CC9902
BL107
CC9605
RS9917
WH7803
RCC307
CB0205
CB0101
0.04
d
NapA
SB
MIT0604-1
MIT0604-2
GOS NapA
P0902-H212
PAC1
P0903-H212
RS9916
WH8109
CC9605
WH8102
CC9902
BL107
CC9311
WH8016
WH7805
WH7803
RCC307
CB0205
CB0101
0.03

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

Name	Clade	Axenic	Isolation Depth (m)	Isolation Coordinates	Region	Isolation Date	Assembly Size (bp)	Contigs	% GC	Genbank Accession	Reference
Unialgal Cultures (complete genome sequences)											
SB	HL II	Yes	40	35°N, 138.3°E	Suruga Bay, Japan	October 1992	1 668 514	3	31.5	JNAS00000000	Shimada et al, 1995; Biller et al, 2014
MIT0604	HL II	Yes	175	22.75°N, 158°W	North Pacific	May 2006	1 780 061	1	31.2	CP007753	This study
PAC1	LL I	No	100	22.75°N, 158°W	North Pacific	April 1992	1 825 493	15	35.1	JNAX00000000	Penno et al, 2000; Biller et al, 2014
Mixed Enrichments (partial genome assemblies)											
P0902-H212	LL I	No	175	22.75°N, 158°W	North Pacific	July 2009	501 825	1	35.4	KJ947870	This study
P0903-H212	LL I	No	200	22.75°N, 158°W	North Pacific	July 2009	291 739	1	35.2	KJ947871	This study

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'-ATBGGRCATGWYTKYTCRTGC-3') at an annealing temperature of 57^{o}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^{o}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 10^{9} 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^{o}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^{o}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 NH_4Cl, 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

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, *et al.* (2008). The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9: 75.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. (2009). BLAST+: architecture and applications. BMC Bioinformatics 10: 421.

Coleman ML, Chisholm SW. (2010). Ecosystem-specific selection pressures revealed through comparative population genomics. Proc Natl Acad Sci USA 107: 18634-18639.

Darling AE, Mau B, Perna NT. (2010). progressiveMauve: multiple genome alignment with gene gain, loss, and rearrangement. PLoS One. 5: e11147.

Felsenstein J. (2005). PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the Author. Department of Genome Sciences, University of Washington, Seattle.

Jukes TH, Cantor CR. (1969). Evolution of protein molecules. In: Mammalian Protein Metabolism. Munro HN (ed). Academic Press: New York, NY, pp 21-132.

Kamennaya NA, Chernihovsky M, Post AF. (2008). The cyanate utilization capacity of marine unicellular Cyanobacteria. Limnol Oceanogr 53: 2485-2494.

Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, Rodrigue S, *et al.* (2007). Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus.* PLoS Genet 3: e231.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, *et al.* (2007). Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948.

Malmstrom RR, Rodrigue S, Huang KH, Kelly L, Kern SE, Thompson A, *et al.* (2013). Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic analysis. ISME J 7: 184-198.

Martiny AC, Coleman ML, Chisholm SW. (2006). Phosphate acquisition genes in *Prochlorococcus* ecotypes: evidence for genome-wide adaptation. Proc Natl Acad Sci USA 103: 12552-12557.

Martiny AC, Kathuria S, Berube PM. (2009). Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes. Proc Natl Acad Sci USA 106: 10787-10792.

Mühling M. (2012). On the culture-independent assessment of the diversity and distribution of *Prochlorococcus*. Environ Microbiol 14: 567-579.

Ranwez V, Harispe S, Delsuc F, Douzery EJ. (2011). MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. PLoS ONE 6: e22594.

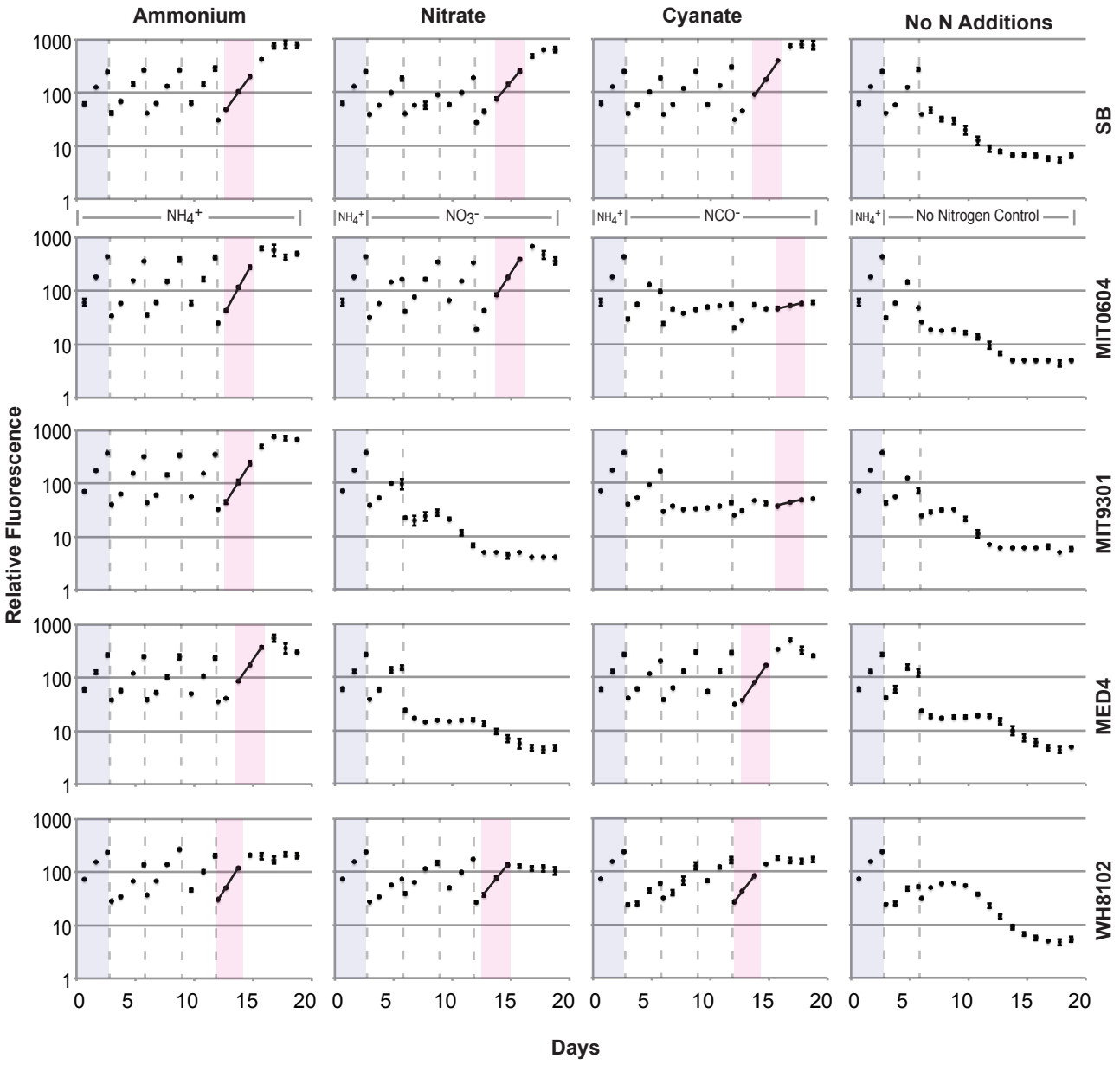
Rodrigue S, Malmstrom RR, Berlin AM, Birren BW, Henn MR, Chisholm SW. (2009). Whole genome amplification and de novo assembly of single bacterial cells. PLoS ONE 4: e6864.

Rodrigue S, Materna AC, Timberlake SC, Blackburn MC, Malmstrom RR, Alm EJ, Chisholm SW. (2010). Unlocking short read sequencing for metagenomics. PLoS ONE 5: e11840.

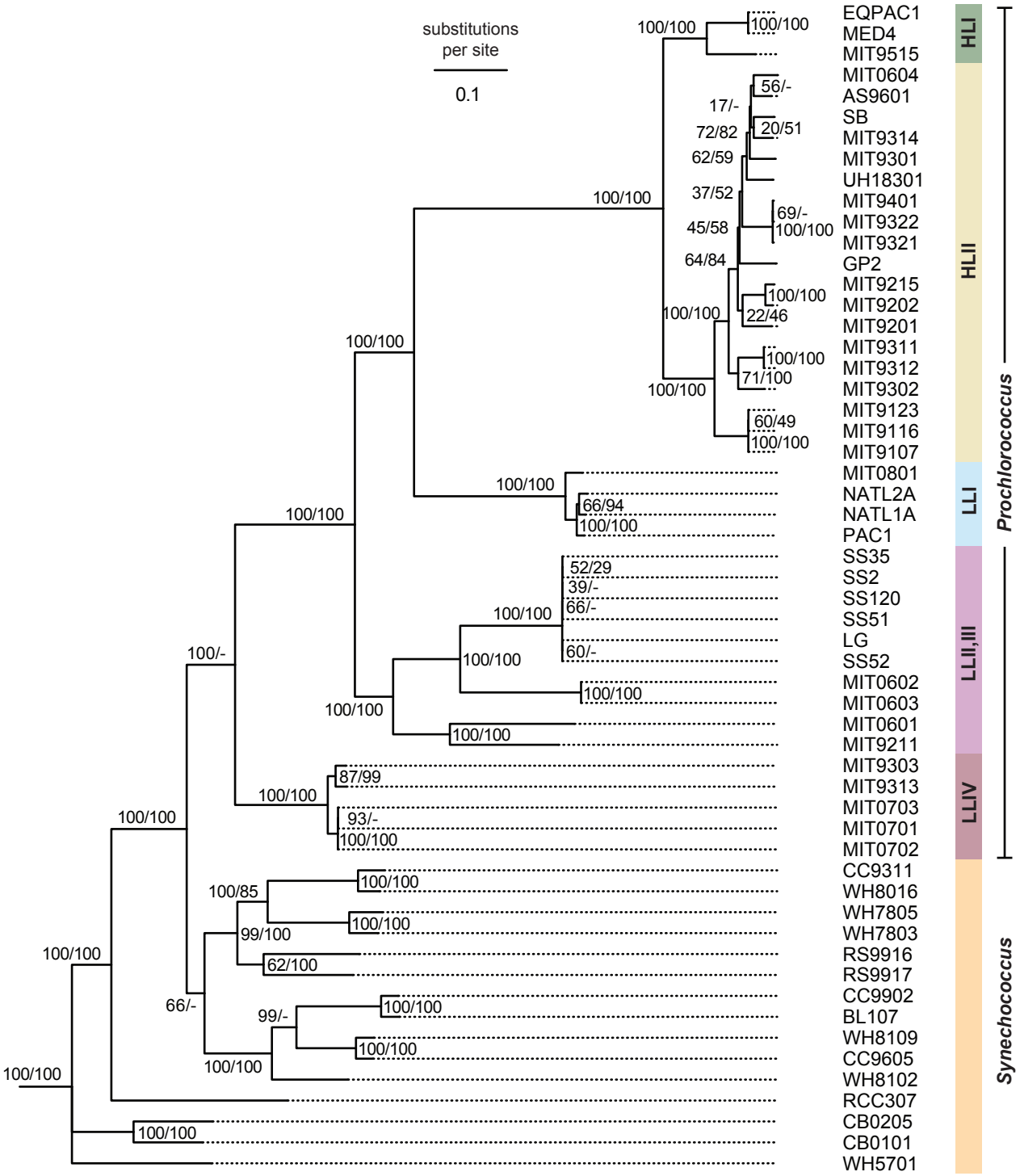
Rusch DB, Martiny AC, Dupont CL, Halpern AL, Venter JC. (2010). Characterization of *Prochlorococcus* clades from iron-depleted oceanic regions. Proc Natl Acad Sci USA 107: 16184-16189.

Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR, *et al.* (2009). Ecological genomics of marine picocyanobacteria. Microbiol Mol Biol Rev 73: 249-299.

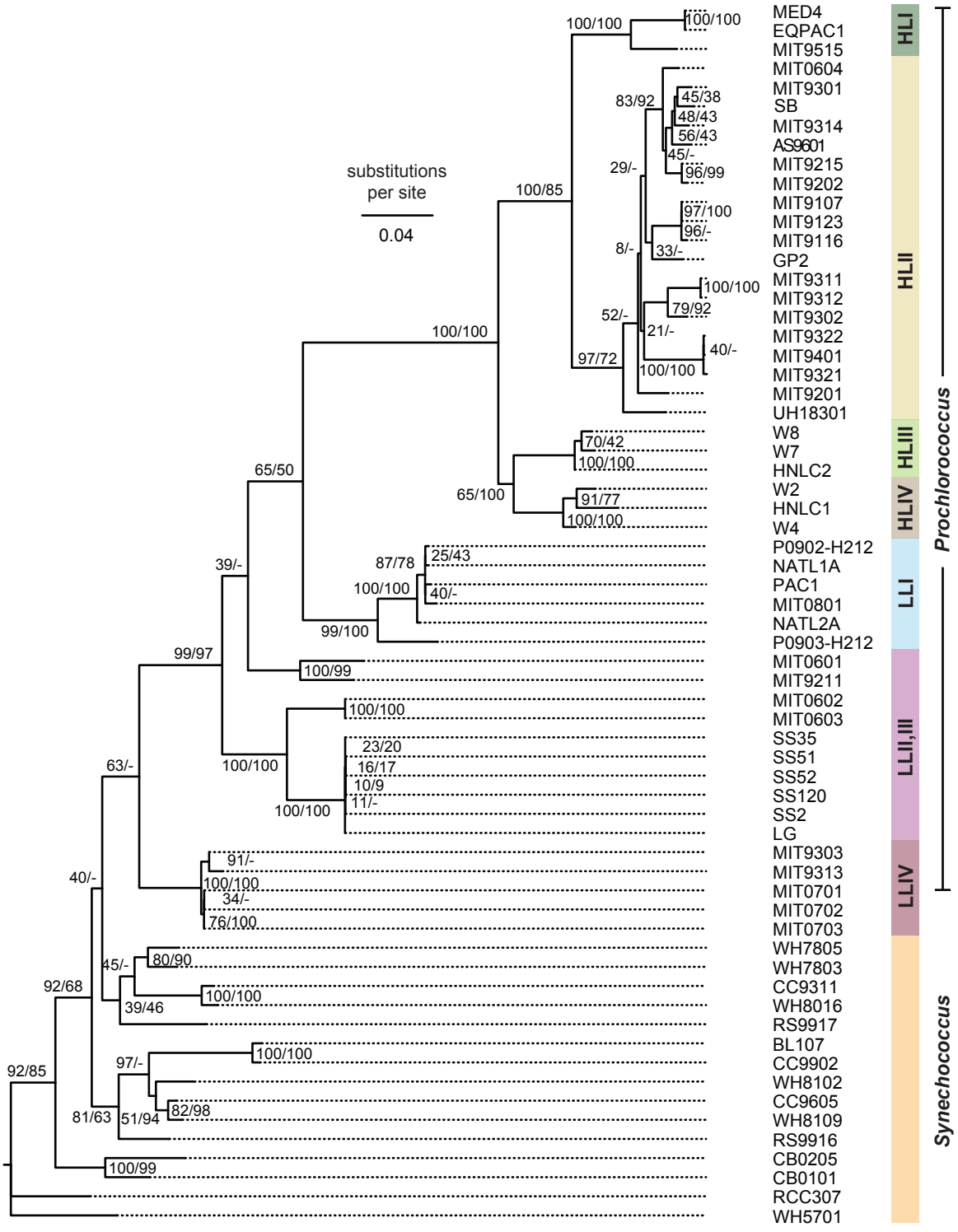
Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731-2739.



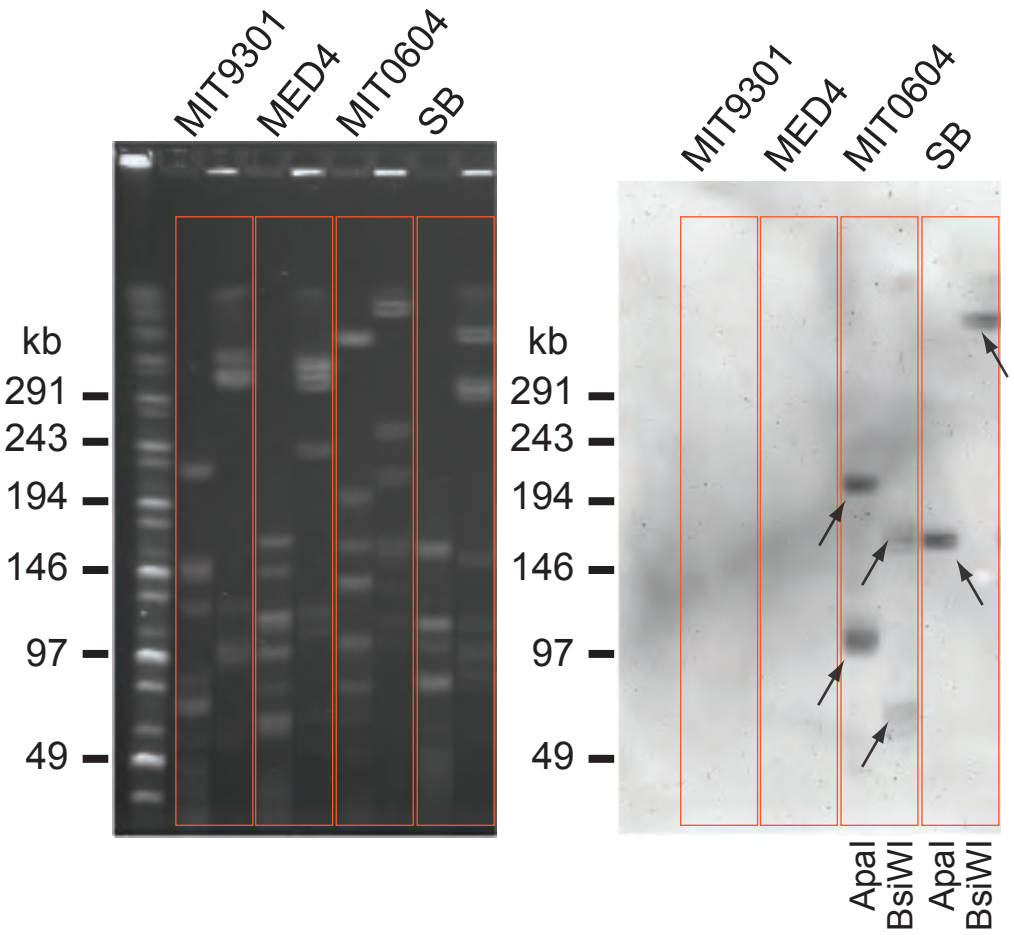
Supplementary Figure S1. Growth of axenic *Prochlorococcus* strains SB, MIT0604, MIT9301, MED4, and axenic *Synechococcus* strain WH8102 in the presence of 800 µM ammonium, nitrate, or cyanate. Bulk culture fluorescence (y-axis) was used as a proxy for cell numbers during exponential growth. Data points 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.



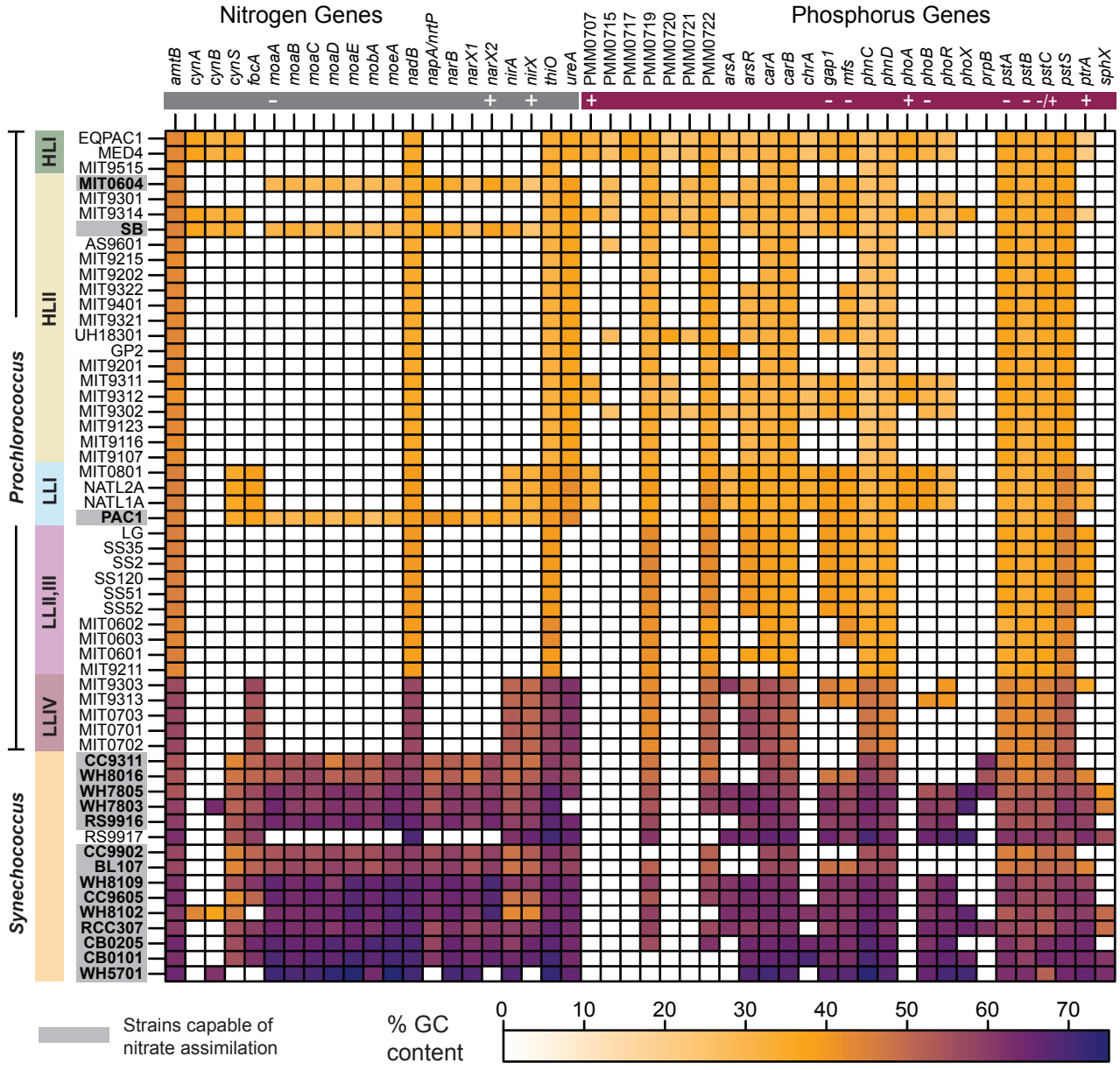
Supplementary Figure S2. Maximum likelihood phylogeny of *Prochlorococcus* and *Synechococcus* proteins based on 100 resamplings of 100 randomly concatenated single-copy core proteins. Bootstrap values (total 100) were calculated using maximum likelihood (first value at each node) and neighbor joining (second value at each node), with dashes representing maximum likelihood topology unsupported by most of the neighbor joining trees.



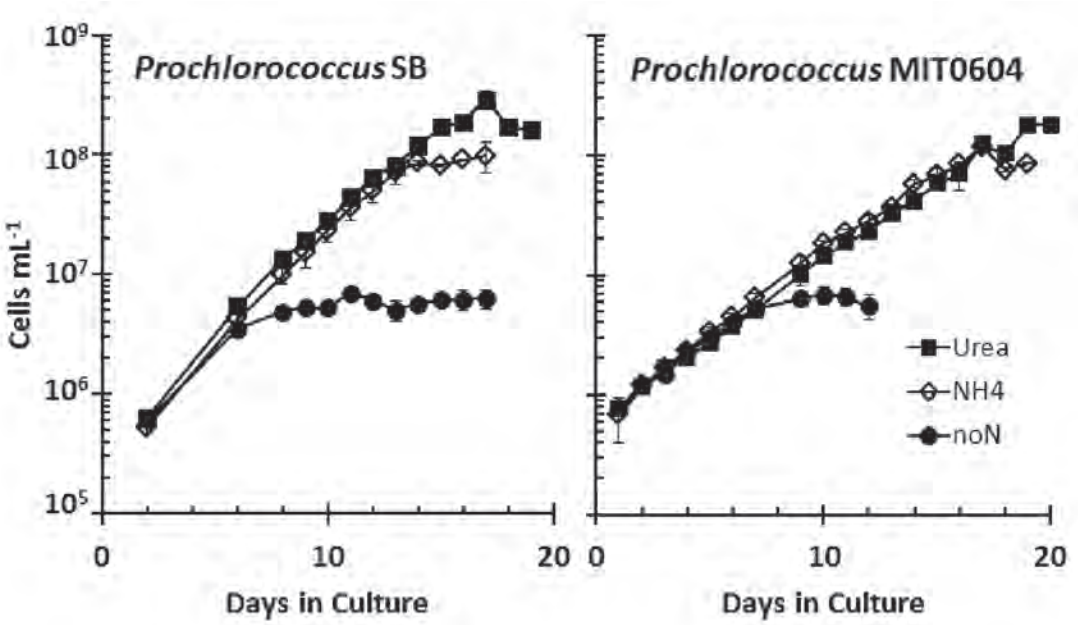
Supplementary Figure S3. Phylogeny of *Prochlorococcus* and *Synechococcus* GyrB proteins. Bootstrap values (total 100) were calculated using maximum likelihood (first value at each node) and neighbor joining (second value at each node), with dashes representing maximum likelihood topology unsupported by most of the neighbor joining trees.



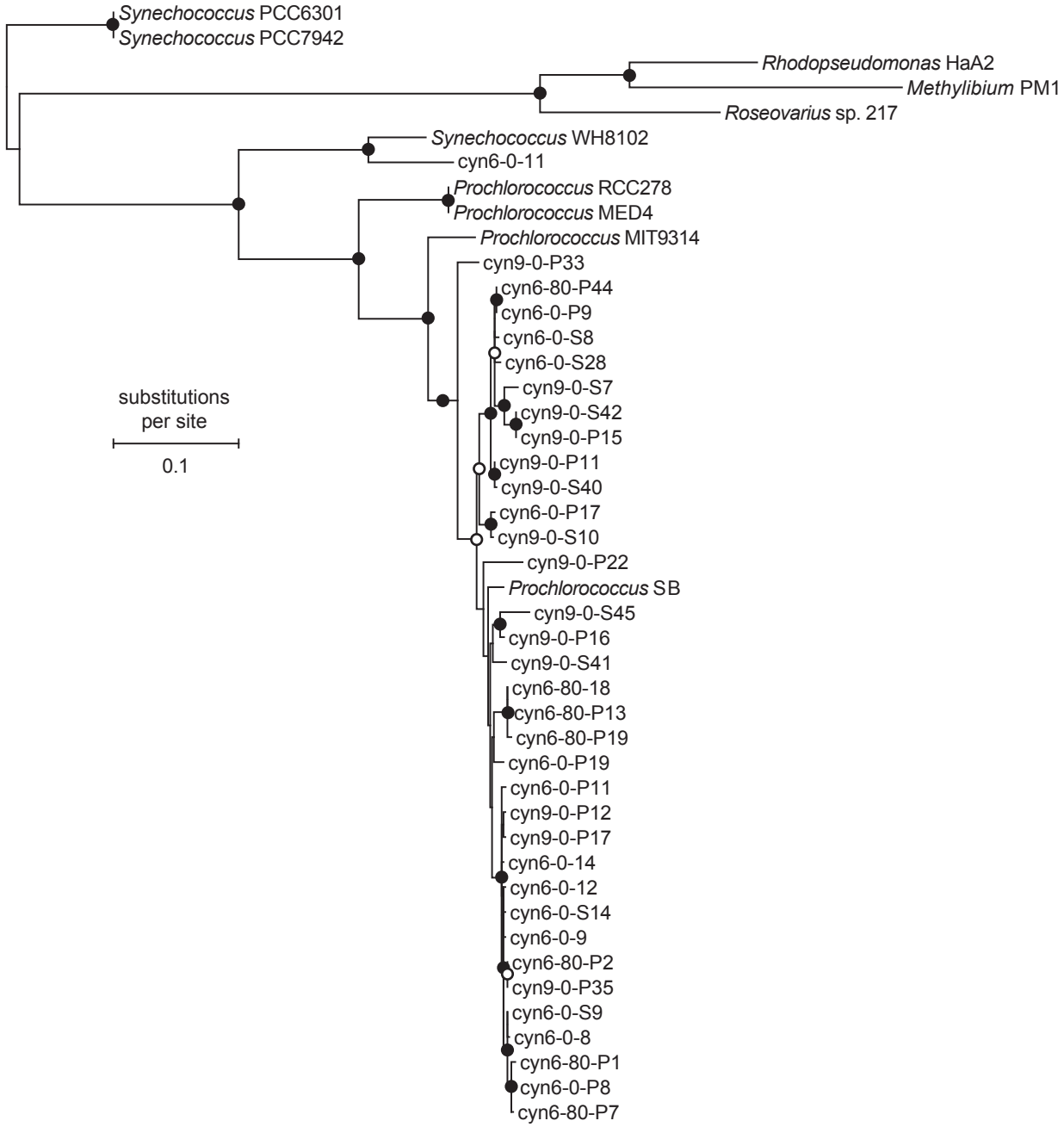
Supplementary Figure S4. Southern blot analysis confirms that *Prochlorococcus* MIT0604 contains two copies of *narB*. The ethidium bromide stained gel is shown at left and the southern blot is shown at right. The *narB* gene in MIT0604 is found on two restriction fragments of the expected sizes (100kb/197kb when digested with ApaI 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 growth 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 x 10^8 ± 5 x 10^5 cells mL^{-1} on ammonium vs. 1.8 x 10^8 ± 6 x 10^6 cells mL^{-1} on urea; MIT0604: 8.6 x 10^7 ± 1 x 10^6 cells mL^{-1} on ammonium vs. 2.2 x 10^8 ± 5 x 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 ± 0.004 d^{-1} on ammonium and 0.36 ± 0.01 d^{-1} on urea. Specific growth rates for MIT0604 were 0.304 ± 0.003 d^{-1} on ammonium and 0.292 ± 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.

Gene	ProPortal v4.0 COG	Product	Role	Reference
Nitrogen Genes				
amtB/amt1	1478	ammonium transporter protein	ammonium transport	García-Fernández et al., 2004
cynA	25277	cyanate ABC type transporter substrate binding protein	cyanate transport	Kamennaya et al., 2008
cynB	17453	cyanate ABC type transporter permease protein	cyanate transport	Kamennaya et al., 2008
cynS	16887	cyanate lyase	hydrolysis of cyanate to ammonium and carbon dioxide	Kamennaya et al., 2008
focA	10584	nitrite transporter from formate/nitrite family	nitrite transport	Rocap et al., 2003
moaA	8269	molybdenum cofactor biosynthesis protein A	molybdopterin biosynthesis	Martiny et al., 2009
moaB	9123	molybdenum cofactor biosynthesis protein B	molybdopterin biosynthesis	Martiny et al., 2009
moaC	12914	molybdenum cofactor biosynthesis protein C	molybdopterin biosynthesis	Martiny et al., 2009
moaD	7626	molybdenum cofactor biosynthesis protein D	molybdopterin biosynthesis	Martiny et al., 2009
moaE	20838	molybdenum cofactor biosynthesis protein E	molybdopterin biosynthesis	Martiny et al., 2009
mobA	7553	molybdopterin-guanine dinucleotide biosynthesis protein MobA	molybdopterin biosynthesis	Martiny et al., 2009
moeA	6195	molybdopterin biosynthesis protein MoeA	molybdopterin biosynthesis	Martiny et al., 2009
nadB	253	L-aspartate oxidase	deamination of amino acids	Tedeschi et al., 1996
napA/nrtP	5121	nitrate/nitrite transporter	nitrate/nitrite transport	Martiny et al., 2009b; Wang et al., 2000; Bird & Wyman, 2003
narB	3405	assimilatory nitrate reductase	nitrate reduction to nitrite	Martiny et al., 2009
narX1	12460	conserved hypothetical protein	unknown function	Martiny et al., 2009
narX2	30465, 26956, 33277	conserved hypothetical protein	unknown function	Martiny et al., 2009
nirA	5136	ferredoxin nitrite reductase	nitrite reduction to ammonium	Martiny et al., 2009
nirX	27176, 11823	conserved hypothetical protein	unknown function	Martiny et al., 2009
thiO	772	glycine oxidase	deamination of amino acids	Nishiya & Imanaka, 1998
ureA	1864	urease subunit alpha	hydrolysis of urea to ammonium and carbon dioxide	Palinska et al., 2000

Gene	ProPortal v4.0 COG	Product	Role	Reference
Phosphorus Genes				
PMM0707	30300, 31904	hypothetical protein	expressed in MED4 during phosphorus starvation	Martiny et al., 2006
PMM0715	26328	hypothetical protein	expressed in MED4 during phosphorus starvation	Martiny et al., 2006
PMM0717	32234	hypothetical protein	expressed in MED4 during phosphorus starvation	Martiny et al., 2006
PMM0719	3650	hypothetical protein	expressed in MED4 during phosphorus starvation	Martiny et al., 2006
PMM0720	28615	hypothetical protein	expressed in MED4 during phosphorus starvation	Martiny et al., 2006
PMM0721	28631	hypothetical protein	expressed in MED4 during phosphorus starvation	Martiny et al., 2006
PMM0722	2536	hypothetical protein	expressed in MED4 during phosphorus starvation	Martiny et al., 2006
arsA	22394	arsenite efflux pump subunit	arsenate resistance	Martiny et al., 2006
arsR	1361	arsenate reductase	arsenate resistance	Martiny et al., 2006
carA	20	carbamoyl phosphate synthetase small subunit	carbamoyl phosphate synthesis	Martiny et al., 2006
carB	346	carbamoyl phosphate synthetase large subunit	carbamoyl phosphate synthesis	Martiny et al., 2006
chrA	13381	response regulator	chromate resistance	Martiny et al., 2006
gap1	99	glyceraldehyde-3-phosphate dehydrogenase	expressed in MED4 during phosphorus starvation	Martiny et al., 2006
mfs	817	major facilitator superfamily transporter	expressed in MED4 during phosphorus starvation	Martiny et al., 2006
prpB	6142	phosphoenolpyruvate mutase	phosphonate biosynthesis	Yu et al., 2013
phnC	506	phosphonate ABC type transporter ATP binding protein	phosphonate transport	Feingersch et al., 2012; Martinez et al., 2010
phnD	4518	phosphonate ABC type transporter substrate binding protein	phosphonate transport	Feingersch et al., 2012; Martinez et al., 2010
phoA	15427, 26745	alkaline phosphatase	dephosphorylation	Martiny et al., 2006
phoB	204	phosphate regulon response regulator	phosphate two component regulatory system	Martiny et al., 2006
phoR	13582	phosphate regulon sensor histidine kinase	phosphate two component regulatory system	Martiny et al., 2006
phoX	26697	alkaline phosphatase	dephosphorylation	Martiny et al., 2006
pstA	3725	phosphate ABC type transporter permease protein	phosphate transport	Martiny et al., 2006
pstB	88	phosphate ABC type transporter ATP binding protein	phosphate transport	Martiny et al., 2006
pstC	4183, 30634	phosphate ABC type transporter permease protein	phosphate transport	Martiny et al., 2006
pstS	1827	phosphate ABC type transporter substrate binding protein	phosphate transport	Martiny et al., 2006
ptrA	37989, 6860, 11384	transcriptional regulator	stress response to phosphorus starvation	Ostrowski et al., 2010
sphX	25109	phosphate binding protein	phosphate transport	Mann & Scanlan, 1994

References for Supplementary Table 1:

Bird C, Wyman M. (2003). Nitrate/nitrite assimilation system of the marine picoplanktonic cyanobacterium *Synechococcus* sp. strain WH 8103: effect of nitrogen source and availability on gene expression. Appl Environ Microbiol 69: 7009-7018.

Feingersch R, Philosof A, Mejuch T, Glaser F, Alalouf O, Shoham Y, Béjà O. (2012). Potential for phosphite and phosphonate utilization by *Prochlorococcus*. ISME J 6: 827-834.

García-Fernández JM, de Marsac NT, Diez J. (2004). Streamlined regulation and gene loss as adaptive mechanisms in *Prochlorococcus* for optimized nitrogen utilization in oligotrophic environments. Microbiol Mol Biol Rev 68: 630-638.

Kamennaya NA, Chernihovsky M, Post AF. (2008). The cyanate utilization capacity of marine unicellular Cyanobacteria. Limnol Oceanogr 53: 2485-2494.

Mann NH, Scanlan DJ. (1994). The SphX protein of *Synechococcus* species PCC 7942 belongs to a family of phosphate-binding proteins. Mol Microbiol 14: 595-596.

Martinez A, Tyson GW, Delong EF. (2010). Widespread known and novel phosphonate utilization pathways in marine bacteria revealed by functional screening and metagenomic analyses. Environ Microbiol 12: 222-238.

Martiny AC, Coleman ML, Chisholm SW. (2006). Phosphate acquisition genes in *Prochlorococcus* ecotypes: evidence for genome-wide adaptation. Proc Natl Acad Sci USA 103: 12552-12557.

Martiny AC, Kathuria S, Berube PM. (2009). Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes. Proc Natl Acad Sci USA 106: 10787-10792.

Nishiya Y, Imanaka T. (1998). Purification and characterization of a novel glycine oxidase from *Bacillus subtilis.* FEBS Lett 438: 263-266.

Ostrowski M, Mazard S, Tetu SG, Phillippy K, Johnson A, Palenik B, *et al.* (2010). PtrA is required for coordinate regulation of gene expression during phosphate stress in a marine *Synechococcus*. ISME J 4: 908-921.

Palinska KA, Jahns T, Rippka R, de Marsac NT. (2000). *Prochlorococcus marinus* strain PCC 9511, a picoplanktonic cyanobacterium, synthesizes the smallest urease. Microbiology 146: 3099-3107.

Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA, *et al.* (2003). Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. Nature 424: 1042-1047.

Tedeschi G, Negri A, Mortarino M, Ceciliani F, Simonic T, Faotto L, Ronchi S. (1996). L-aspartate oxidase from *Escherichia coli*. II. Interaction with C4 dicarboxylic acids and identification of a novel L-aspartate: fumarate oxidoreductase activity. Eur J Biochem 239: 427-433.

Wang Q, Li H, Post AF. (2000). Nitrate assimilation genes of the marine diazotrophic, filamentous cyanobacterium *Trichodesmium* sp. strain WH9601. J Bacteriol 182: 1764-1767.

Yu X, Doroghazi JR, Janga SC, Zhang JK, Circello B, Griffin BM, *et al.* (2013). Diversity and abundance of phosphonate biosynthetic genes in nature. Proc Natl Acad Sci USA 110: 20759-20764.