Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes

Adam C. Martiny[a,b,1], Satish Kathuria[a], and Paul M. Berube[c]

Departments of [a]Earth System Science and [b]Ecology and Evolutionary Biology, University of California, Irvine, CA 92697; and [c]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by David M. Karl, University of Hawaii, Honolulu, HI, and approved May 1, 2009 (received for review March 8, 2009)

The marine cyanobacterium *Prochlorococcus* is the most abundant photosynthetic organism in oligotrophic regions of the oceans. The inability to assimilate nitrate is considered an important factor underlying the distribution of *Prochlorococcus*, and thought to explain, in part, low abundance of *Prochlorococcus* in coastal, temperate, and upwelling zones. Here, we describe the widespread occurrence of a genomic island containing nitrite and nitrate assimilation genes in uncultured *Prochlorococcus* cells from marine surface waters. These genes are characterized by low GC content, form a separate phylogenetic clade most closely related to marine *Synechococcus*, and are located in a different genomic region compared with an orthologous cluster found in marine *Synechococcus* strains. This sequence distinction suggests that these genes were not transferred recently from *Synechococcus*. We demonstrate that the nitrogen assimilation genes encode functional proteins and are expressed in the ocean. Also, we find that their relative occurrence is higher in the Caribbean Sea and Indian Ocean compared with the Sargasso Sea and Eastern Pacific Ocean, which may be related to the nitrogen availability in each region. Our data suggest that the ability to assimilate nitrite and nitrate is associated with microdiverse lineages within high- and low-light (LL) adapted *Prochlorococcus* ecotypes. It challenges 2 long-held assumptions that (*i*) *Prochlorococcus* cannot assimilate nitrate, and (*ii*) only LL adapted ecotypes can use nitrite. The potential for previously unrecognized productivity by *Prochlorococcus* in the presence of oxidized nitrogen species has implications for understanding the biogeography of *Prochlorococcus* and its role in the oceanic carbon and nitrogen cycles.

metagenomics | cyanobacteria | nitrogen cycle | *nar*B | *nir*A

Marine phytoplankton are responsible for ≈50% of global net primary productivity (1). In many regions of the oligotrophic subtropical and tropical oceans, field studies and biogeochemical models suggest that nitrogen availability has an important role in regulating primary productivity (2, 3). In these regions, the marine cyanobacterium *Prochlorococcus* is often the numerically dominant phytoplankton group and an important contributor to primary production (4). *Prochlorococcus* strains can be divided into 2 physiologically and genetically distinct ecotypes, high-light (HL) and low-light (LL) adapted (5). Whereas most phytoplankton can use ammonia, nitrite, and nitrate as sole nitrogen sources, nearly all *Prochlorococcus* isolates are limited to ammonia as their source of nitrogen. The only known exceptions are 2 LL *Prochlorococcus* clades (eNATL and eMIT9313), which can assimilate nitrite, whereas no cultured strains are capable of nitrate assimilation (6). This inability to use nitrate is considered an important factor underlying the distribution of *Prochlorococcus*, and it is thought to explain, in part, the low abundance in coastal, temperate and upwelling zones (7–10).

In marine *Synechococcus*, genes responsible for nitrite and nitrate assimilation are located in one genomic region that includes nitrite reductase, nitrate reductase, a transporter, and several genes responsible for molybdopterin cofactor biosynthesis (11). A comparison of this genomic region between *Synechococcus* and *Prochlorococcus* suggests that the inability to use certain nitrogen species in *Prochlorococcus* is the result of past sequential genomic decay (Fig. 1*A*) (12). That is, during the divergence of *Prochlorococcus* from *Synechococcus*, LL *Prochlorococcus* first lost the nitrate assimilation genes, followed by HL *Prochlorococcus* losing genes for nitrite assimilation. It is hypothesized that these losses are an adaptation to the distribution of nitrate and nitrite in the water column. In contrast to the typically high concentrations of nitrite found in the deeper waters where LL *Prochlorococcus* is abundant, nitrite and nitrate are depleted in the oligotrophic surface waters where HL *Prochlorococcus* is abundant.

However, some physiological and field observations suggest a more complex interpretation. First, not all LL strains are capable of nitrite assimilation, implying that either the ability to use nitrite has been lost multiple times, or that these genes have been acquired by lateral gene transfer in certain lineages of *Prochlorococcus*. Second, Casey et al. (13) observed low nitrate assimilation rates in a subpopulation of *Prochlorococcus* in the deep chlorophyll maximum of the Western Atlantic Ocean; thus, demonstrating that some *Prochlorococcus* can use nitrate. Third, Vaulot and Partensky (14) detected increased *Prochlorococcus* cell division rates after nitrate additions to water collected from several depths of the Northwestern Mediterranean Sea. Fourth, the distribution of *Prochlorococcus* microdiversity is significantly correlated with variation in nitrate concentration (15). These observations point toward the existence of yet-to-be discovered lineages that are capable of nitrite and nitrate assimilation, possibly as a result of reacquisition of these genes by lateral gene transfer.

Results

To test this hypothesis, we examined metagenomic libraries for nitrite and nitrate reductase genes associated with *Prochlorococcus*. The samples were part of the Global Ocean Survey (GOS) and covered sites in the Atlantic, Pacific, and Indian Oceans (Table S1) (16). Specifically, we screened for genes encoding proteins similar to nitrite reductases (NirA) from *Synechococcus* and LL *Prochlorococcus*. Phylogenetic analysis of the translated GOS sequences revealed that these sequences fall within 2 large clades (Fig. 2*A*; Fig. S1). One clade includes sequences related to marine *Synechococcus* and LL *Prochlorococcus* eMIT9313 strains, and the second clade includes sequences related to *Prochlorococcus* eNATL strains. The GOS *nir*A sequences clustering with the *Synechococcus* strains have an average GC content of 55.1% (n = 121), whereas those cluster-

Author contributions: A.C.M., S.K., and P.M.B. designed research; A.C.M., S.K., and P.M.B. performed research; A.C.M. and P.M.B. contributed new reagents/analytic tools; A.C.M. and P.M.B. analyzed data; and A.C.M. and P.M.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

See Commentary on page 10400.

[1]To whom correspondence should be addressed. E-mail: amartiny@uci.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0902532106/DCSupplemental.

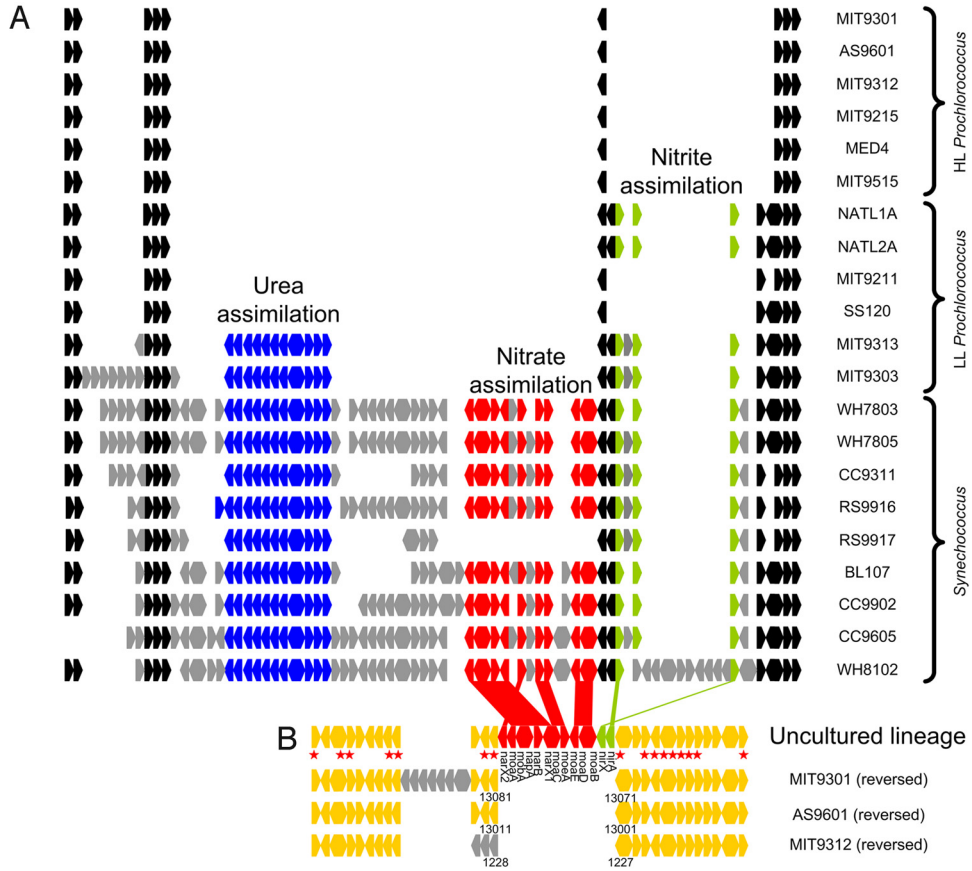

Fig. 1. Nitrogen acquisition genes in *Prochlorococcus* and *Synechococccus*. (*A*) Nitrogen genes in the genomic region between *pyr*G and ppk. (*B*) Suggested consensus sequence of nitrogen acquisition genes acquired in an uncultured lineage of *Prochlorococcus* and orthologous regions in HL *Prochlorococcus*. Green represents genes associated with nitrite assimilation, red represents genes associated nitrate assimilation, blue represents genes associated with urea assimilation, black represents other genes conserved across *Prochlorococcus* and *Synechococcus*, yellow represents conserved genes in *Prochlorococcus*, and the rest are gray. The red star represents genes that have paired end mates matching the region around the nitrate genomic island. It is important to note that all genes in this insertion region were detected on DNA fragments that also contained genes matching *Prochlorococcus* (i.e., the paired end sequence). We only used a scaffold to estimate the gene order. Note also that urea assimilation genes are located in a different genomic region in many *Prochlorococcus* strains.

ing with the eNATL strains have an average GC content of 32.2% ($n = 76$). This GC pattern is consistent with the genomic GC content of cultured isolates; marine *Synechococcus* strains have an average GC content of >50%, whereas *Prochlorococcus* strains (with the exception of strains from the LL clade eMIT9313) have a GC content of <40% (Fig. S2). Thus, based on the phylogenetic clustering and GC content, it appears that there is a large group of *nir*A sequences affiliated with *Prochlorococcus* among the GOS samples.

The putative *Prochlorococcus nir*A clade consists of 1 subclade with the eNATL sequences and a few GOS sequences (low GC II), and another subclade with no cultured representatives (low GC I; Fig. 2*A*). This division was supported by bootstrap values of at least 98% for neighbor-joining and maximum likelihood analyses and 59% for parsimony analysis. Sequences from low GC I with no cultured representatives were 65 to 69% similar to the NirA amino acid sequence in strain NATL1A. This sequence divergence is in the same range of similarity as between eNATL and eMIT9313 NirA sequences (66%), suggesting that these uncultured GOS sequences might originate from a different ecotype of *Prochlorococcus*.

Significantly, several of the paired end sequences associated with the *nir*A sequences from the low GC I clade had best blastn hits (average 92% nucleotide similarity) to HL *Prochlorococcus* genomes from the eMIT9312 ecotype (AS9601, MIT9312, and MIT9301) (Fig. 2*A*; Table S2). Because the paired end sequence is located on the same fragment of DNA as the *nir*A sequence, these data strongly support the presence of *nir*A in some HL *Prochlorococcus* genomes. The paired end sequences match genes on the edge of the genomic island ISL4 in *Prochlorococcus* (Fig. 1*B*; Fig. S3) (17). A nucleotide alignment shows that some GOS sequences are highly similar to gene A9601_13001 in *Prochlorococcus* AS9601 (and orthologs in other HL strains) and part of the downstream intergenic region (within ≈100 bp) (Fig. S3). Going further downstream, the GOS sequences are very different from known genomes of *Prochlorococcus*. On the same DNA fragment, some GOS sequences contain a region with a putative nitrogen stress regulator NtcA binding site followed by a low GC version of the *nir*A gene. This genetic organization suggests that *nir*A is inserted next to A9601_13001 in some HL adapted *Prochlorococcus* lineages (Fig. 1*B*). In contrast, the *nir*A gene known to be found in LL *Prochlorococcus* strains is located in a different section of the genome (Fig. 1*A*).

In addition to nitrite reductase, 2 other genes are present in *Synechococcus* and *Prochlorococcus* LL strains capable of nitrite assimilation: the nitrite transporter *foc*A and a gene orthologous to NATL1_21731, which we will refer to as *nir*X (Fig. 1*A*). Although we could not detect any low GC orthologs to *foc*A in the GOS dataset, we found many similar to *nir*X (Fig. 2*B*). Again, we observed a large clade of low GC sequences most similar to NATL1A, but forming an independent clade (low GC I). Also, several paired end sequences from the low GC I clade had best matches to A9601_13001 and upstream genes; thus,

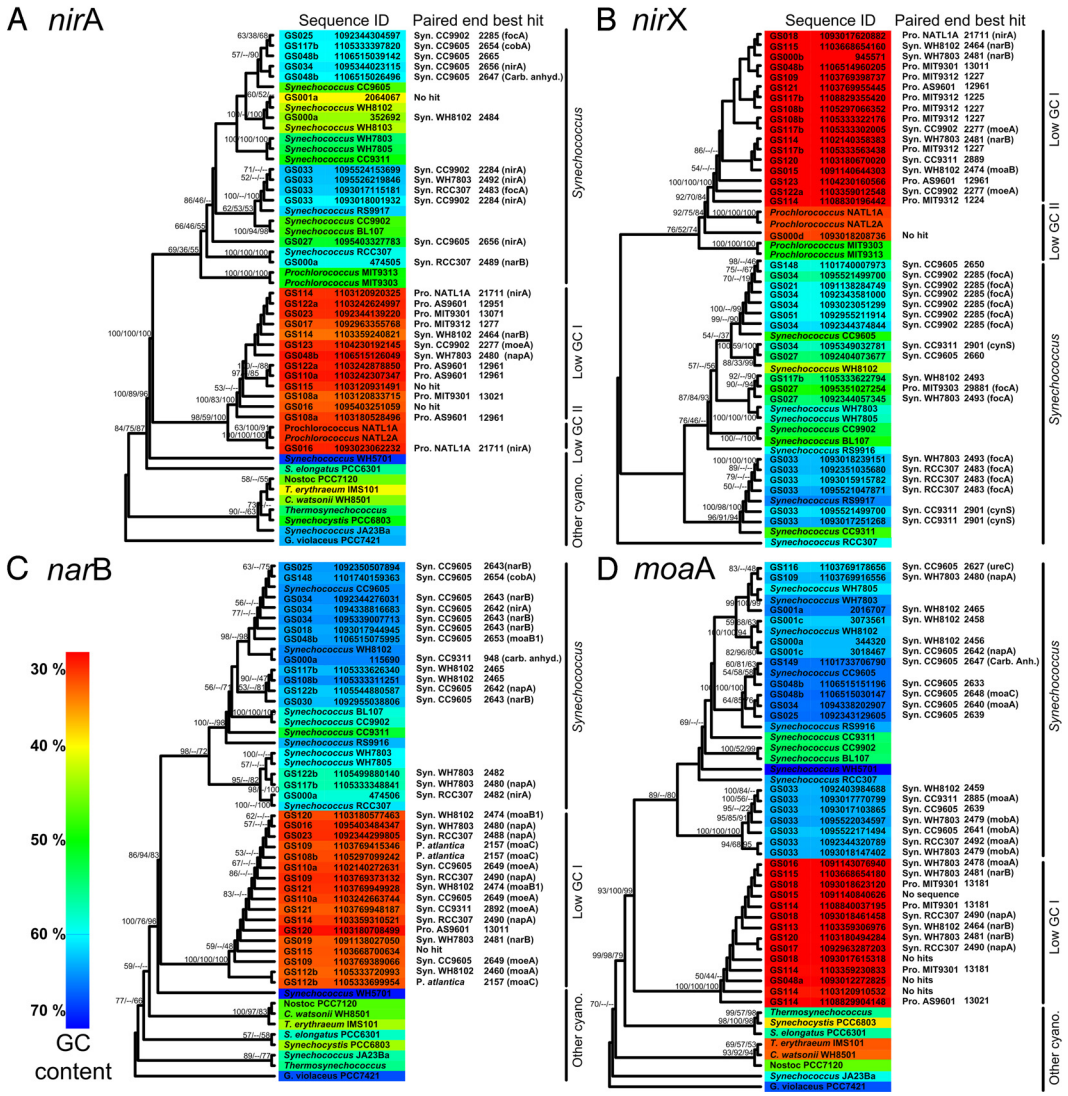

Fig. 2. Phylogeny of putative nitrite and nitrate assimilation protein sequences inferred from GOS expedition samples and Cyanobacteria strains. Next to the sequence ID is listed the organisms and locus most similar to the paired end sequence mate. (*A*) Nitrite reductase (NirA) (from position 221 to 513 in *Synechococcus* WH8102). (*B*) Conserved hypothetical protein (NirX) (all positions). (*C*) Nitrate reductase (NarB) (from position 1 to 280 in *Synechococcus* WH8102). (*D*) Molybdopterin biosynthesis protein A (MoaA) (from position 10 to 175 in *Synechococcus* WH8102). The phylogenetic trees are based on protein sequence similarity using neighbor-joining. Bootstrap values (total 100) are calculated using neighbor-joining, maximum parsimony, and maximum likelihood. Sequence ID refers to sample number and JCVI_READ_XXX. The best hit to the paired end mate of each sequence is found by blastx searching against all sequenced genomes in GenBank as of 01/28/08. Additional regions for each protein are detailed in the supplemental figures.

placing *nir*X in the vicinity of *nir*A in HL *Prochlorococcus* ecotypes.

We also noticed that several paired end mates to putative *Prochlorococcus nir*A and *nir*X sequences had best hits to *Synechococcus* genomes (Fig. 2 *A* and *B*; Fig. S1). These hits included matches to nitrate reductase (*nar*B), a nitrate/nitrite transporter (*nap*A), genes encoding molybdopterin biosynthesis, and 2 hypothetical genes. Together, these genes are necessary for nitrate assimilation, and have been previously found to be located in a cluster together with *nir*A in many Cyanobacteria including *Synechococcus* (Fig. 1*A*) (18). Similar to *nir*A and *nir*X, all sequences had low GC contents that would be very unusual for marine *Synechococcus*. Thus, we speculated that some *Prochlorococcus* lineages might also have genes encoding for nitrate assimilation.

To investigate further, we searched the GOS dataset for genes encoding proteins similar to *Synechococcus* WH8102 NarB and subjected the reciprocal best hits to a phylogenetic analysis. As with NirA, we found 2 separate clades of NarB that included sequences from the GOS dataset (Fig. 2*C*; Fig. S4). One clade consisted of high GC sequences (63.2%, $n = 100$), as well as the marine *Synechococcus* isolates. The second clade consisted of low GC sequences (32.9%, $n = 98$), and contained no cultured representatives (low GC I). Again, several paired end mate sequences from the low GC clade were very similar to HL *Prochlorococcus* isolates on a nucleotide level; thus, placing these putative nitrate reductase encoding genes on the same fragment of DNA as HL *Prochlorococcus* genomic sequences (Fig. 2*C*; Fig. S4 and Table S2).

When analyzing other genes responsible for nitrate assimilation including the nitrate/nitrite transporter (*nap*A), molybdopterin biosynthesis (*moa*ABCDE, *mob*A, and *moe*A), and 2 hypothetical genes orthologous to WH8102 gene 2465 (*nar*X1) and 2466 (*nar*X2), we saw similar patterns (Fig. 2*D*; Fig. S5). These results suggest the presence of putative *Prochlorococcus* lineages containing all of the genes necessary for nitrate assimilation.

An assembly of all these sequences suggests one genomic island with an average GC content of 30.7% inserted between orthologs to gene A9601_13001 and A9601_13011 (Fig. 1*B*). Fig. S6 shows a detailed alignment of the final part of the gene cluster including *nar*X2 and A9601_13011. It is noteworthy that the order of genes is different compared with marine *Synechococcus* (Fig. 1*B*). Combined with the low GC content, this arrangement suggests that these genes have been present in *Prochlorococcus* for a long time and not recently transferred from *Synechococcus*.

We also examined metagenomic libraries from the North Pacific Subtropical Gyre (19, 20), and detected copies of all genes belonging to the nitrate assimilation gene cluster. In particular, we found 1 clone taken at 130-m depth that contains a *nar*B copy (GenBank accession no. DU755900). This *nar*B copy has a GC content of 39.7%, and shares 59% nucleotide similarity to the low GC version described in our study. The paired end sequence for this read is highly similar to the *gidA* gene in the LL *Prochlorococcus* strain NATL1A (NATL1_21451), and is located close to the original genomic region surrounding *pyr*G where nitrate assimilation genes are found in *Synechococcus*. In the GOS dataset, there was also one *nar*B sequence (JCVI_READ_1105430353883) with a paired end mate matching LL *Prochlorococcus* NATL1A. This fragment containing *nar*B was also located in the original genomic region next to *pyr*G, and had a slightly higher GC content of 40.7% that is very similar to the whole genome average GC content of the eNATL ecotype (Fig. S2). Thus, it appears that the nitrate gene cluster is located in a different region in HL and LL *Prochlorococcus*. Jenkins et al. (21) also observed several novel lineages of *nar*B in the North Pacific Subtropical Gyre, and speculated that one of these clusters might be associated with *Prochlorococcus*. However, these sequences did not match the putative *Prochlorococcus nar*B sequences identified in this study (Fig. S5*K*).

The translation products of the putative *Prochlorococcus* low GC *nir*A *and nar*B consensus sequences have expected protein masses of 58 and 79 kDa, respectively, which is within the normal range of these proteins in Cyanobacteria (22). To confirm that these genes indeed encoded functional versions of nitrite and nitrate reductase, we expressed both genes and their orthologs in *Synechococcus* WH8102 heterologously in *Escherichia coli* strain BL21 (Fig. S7). Because no GOS fragments covered the entire reading frame of either gene, we used the consensus sequences for *nir*A and *nar*B, and synthesized them in vitro. The observed masses of the expressed proteins were consistent with the above predictions. Using dithionite reduced methyl-viologen as electron donor, we demonstrated the reduction of nitrite by NirA and reduction of nitrate to nitrite by NarB. The biochemical data confirm that these GOS sequences indeed encode functional reductases (Fig. S7).

Next, we determined whether these low GC *nirA* and *nar*B genes are expressed in the environment. Using RT-PCR on previously extracted RNA from cells at 75-m depth in the oligotrophic waters of the North Pacific as template, we detected transcription of the putative *Prochlorococcus nirA* and *nar*B genes (Fig. S8). Also, the sequences of the RT-PCR products shared >98% nucleotide similarity to the putative *Prochlorococcus nirA* and *nar*B consensus sequences (Fig. S8).

Last, we examined the field distribution of conserved *Prochlorococcus* and *Synechococcus* genes in addition to the nitrite and nitrate assimilation genes discussed above (Fig. 3; Fig. S9). Based on the abundance of 8 common single copy household genes, we confirmed that HL *Prochlorococcus* are most abundant among surface water samples from the Sargasso Sea, Caribbean Sea, Eastern Pacific Ocean, and Indian Ocean (16). In contrast, we rarely detected any sequences associated with LL *Prochlorococcus*. *Synechococcus* sequences are most prevalent in samples from the Sargasso Sea and coastal waters around Galapagos and Zanzibar. When we examine the abundance of nitrite and nitrate assimilation genes, we find that the low GC versions of these genes including *nar*B are solely found in samples containing *Prochlorococcus*. In particular, we commonly find low GC copies in the Caribbean Sea and Indian Ocean. Thus, the

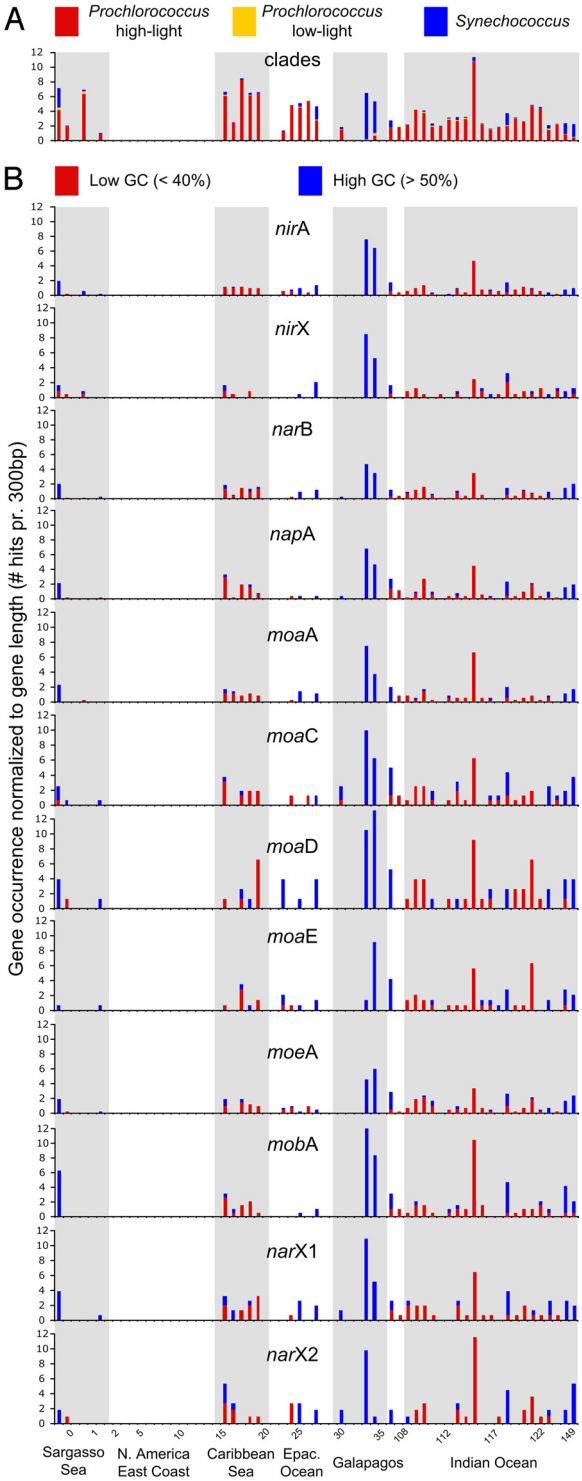

Fig. 3. Distribution of *Prochlorococcus* and *Synechococcus* and associated nitrite and nitrate assimilation genes across ocean regions. (*A*) Distribution of *Synechococcus*, LL, and HL *Prochlorococcus* based the average abundance of 8 single copy core genes. (*B*) Distribution of high and low GC copies of genes orthologous to nitrite and nitrate assimilation genes in *Synechococcus*. Abundance is normalized according to gene length.

distribution of HL adapted *Prochlorococcus* and low GC nitrate assimilation genes corresponds well for samples in the Caribbean Sea and Indian Ocean, whereas we find a lower abundance of low GC nitrate assimilation genes than HL *Prochlorococcus* in samples from the Sargasso Sea and Pacific Ocean (although not absent). The high GC copies are most abundant in samples from the Sargasso Sea, Galapagos, and Indian Ocean including around Zanzibar. This pattern matches the distribution of *Synechococcus*. Overall, the distribution of low and high GC copies is correlated with the abundance of *Prochlorococcus* and *Synechococcus*, respectively, and provides further support for the existence of *Prochlorococcus* lineages capable of nitrate assimilation. Also, it suggests that many *Prochlorococcus* cells contain these genes in the Caribbean Sea and Indian Ocean, whereas these genes are less common among *Prochlorococcus* cells from the Sargasso Sea and regions in the Pacific Ocean at the time of sampling (Fig. S9).

Discussion

Metagenomic analyses are not biased by cultivation media or design of PCR primers, so they have proven important for finding previously undiscovered functional traits in microorganisms (23). Here, we have used this approach to demonstrate that microdiverse lineages within *Prochlorococcus* have previously unrecognized physiological traits for nitrogen assimilation. In support of this conclusion, we found many genes required for nitrite and nitrate assimilation that form an independent cluster of sequences most closely related to *Prochlorococcus* and *Synechococcus*, but with GC contents matching that of *Prochlorococcus* whole genome sequences. Significantly, many paired end sequence mates are very similar on a nucleotide level to *Prochlorococcus* genomes (but not to any other lineages); thus, placing these nitrite and nitrate assimilation genes on fragments of *Prochlorococcus* genomic DNA. Although we did not find any orthologs to the nitrite transporter *foc*A among the GOS samples, this gene is also absent in some marine *Synechococcus* (e.g., WH8102), and the *nap*A transporter may cover both nitrite and nitrate uptake (24). Thus, this gene may not be essential for nitrite and nitrate uptake. Also, we have shown that these genes encode functional nitrite and nitrate reductases, which are transcribed in the oligotrophic open ocean waters of the North Pacific. This result suggests that nitrate assimilation by *Prochlorococcus* is a component of the biogeochemistry of these waters. Last, the high and low GC types of nitrate assimilation genes were only present in samples containing *Synechococcus* and HL *Prochlorococcus*, respectively. Thus, our results provide genetic and enzymatic evidence that *Prochlorococcus* can assimilate nitrate. Although a recent study showed that nitrate uptake by *Prochlorococcus* within the deep chlorophyll maximum of the North Atlantic Ocean could constitute 5–10% of the overall nitrogen assimilation of this group (13), our study demonstrates that this trait is also present in the surface mixed-layer in most ocean regions where *Prochlorococcus* is most abundant. It was previously believed that only LL *Prochlorococcus* could assimilate nitrite, but our results indicate that some HL *Prochlorococcus* are capable of using both nitrite and nitrate.

We find the genes for nitrate assimilation in both HL and LL ecotypes including the abundant HL ecotype eMIT9312. However, cultured members of these ecotypes do not have these genes. Thus, we propose that the ability to assimilate nitrite and nitrate is associated with microdiverse lineages within several HL and LL *Prochlorococcus* ecotypes, rather than one unique type. In marine *Synechococcus* and LL *Prochlorococcus*, *nir*A and *nir*X are located in the proximity of *pyr*G (orthologous to A9601_18991) (Fig. 1*A*) (12). A few fragments show that nitrate assimilation genes in LL *Prochlorococcus* may also be located in this genomic region. In contrast, our data suggest that HL *Prochlorococcus* have regained the ability to assimilate nitrite and nitrate in a new genomic region (Fig. 1*B*). Thus, similar to adaptation to phosphate limitation (25), it appears that genomic islands have a role in adaptation to nitrogen limitation in *Prochlorococcus*.

Our data also indicate that nitrate assimilation genes are more prevalent in some regions compared with others. We speculate that this distribution is related to nitrogen availability. The genes are present in most cells in the Caribbean Sea and Indian Ocean, and these regions are characterized by a low concentration of nitrate at this time of year and depth (Fig. S9) (26). In contrast, the genes are found in low occurrence at this time in the Sargasso Sea and Eastern Pacific, where the nitrate concentration commonly is higher or the cells are limited by other nutrients (e.g., phosphate). Thus, cells proliferating in regions with low nitrogen may contain genes to access this pool of nitrogen, whereas in regions with high nitrogen, it may be more advantageous to have a smaller genome. The occurrence of phosphate genes in *Prochlorococcus* show a similar trend (27). We think that the biogeography of nitrate genes in *Prochlorococcus* is not a result of cells actively losing or gaining genes but rather selection of lineages with the optimal genome.

The inability to assimilate nitrate has been used to explain in part the low abundance of *Prochlorococcus* relative to *Synechococcus* in coastal, temperate, and upwelling regions (7–10). We speculate that our observations of the widespread presence of potentially nitrate assimilating *Prochlorococcus* can be reconciled with the observed field distribution and model results. *Synechococcus* likely has a higher growth rate (μ_{max}) at elevated nitrate concentrations and, thus, dominates in nitrate rich waters. In contrast, *Prochlorococcus* cells probably have a higher uptake rate at very low nitrate concentrations due to their small size (i.e., lower K_s) and, thus, dominate in nitrate poor regions like the surface mixed-layer at lower latitudes.

These different growth parameters of *Synechococcus* and *Prochlorococcus* may also explain why no isolated *Prochlorococcus* strains possess the ability to assimilate nitrate, and why many are not capable of assimilating nitrite. Because *Synechococcus* likely grows faster at nitrogen concentrations normally provided for isolation (>10 μM), they will out compete *Prochlorococcus*. Thus, new cultivation strategies are probably required to culture nitrate assimilating *Prochlorococcus*. This conclusion also demonstrates how metagenomics can be used to guide future isolation studies.

Our overall findings have advanced our understanding of the environmental pressures that drive evolution of microdiverse lineages of *Prochlorococcus*. Also, the potential for previously unrecognized widespread productivity by *Prochlorococcus* in the presence of nitrite and nitrate has significant implications for understanding the biogeography of *Prochlorococcus* and its role in the oceanic carbon and nitrogen cycles.

Materials and Methods

Identification of Nitrogen Assimilation Genes in GOS Samples. Initially, we searched the GOS sequence database for nitrogen assimilation genes *nir*A, *nir*X, *nar*B, *nap*A, *moa*ABCDE, *moe*A, *mob*A, *nar*X1 (SYNW2465), and *nar*X2 (SYNW2466) matching protein sequences from *Prochlorococcus* NATL1A and MIT9303 (for NirA and NirX) and *Synechococcus* WH8102 (rest) using tblastn (e value = 1E-5). Sample location and environment conditions are listed in Table S1. Next, each GOS hit and its paired end sequence mate were compared with a database consisting of all sequenced genomes (as of 01/14/08) using both blastx and blastn (e value = 1E−30). Paired end mates without any hits to genome sequences were further compared with sequenced *Prochlorococcus* fosmids (17). GOS sequences matching nitrogen assimilation genes in *Prochlorococcus* or marine *Synechococcus* genomes were exported, and matching regions were translated. We aligned the protein sequences from GOS and Cyanobacteria genomes using the software Geneious (Biomatters) and carefully curated the alignments manually. Because GOS sequences were <1000-bp long, we split alignments of >800 bp to ensure that GOS sequences covered at least 75% of the region. Thus, we have included multiple phylo-

genetic trees for long genes. Phylogenetic trees (100 bootstraps) were estimated with Phylip v.3.66 using neighbor-joining (Dayhoff PAM matrix), maximum parsimony and maximum likelihood (JTT model, 10 jumbles) (28). Phylogenetic trees based on nucleotide alignments gave similar tree topology between *Synechococcus* and the putative *Prochlorococcus* sequences. The scaffold in Fig. 1*B* is assembled using Geneious and manually curated, and is based on 269 sequences from GOS samples from all regions and had an average coverage of 20× (Fig. S10).

Heterologous Cloning and Expression of *nirA* and *nar*B. We verified the function of the *nir*A and *nar*B by expressing each protein in *E. coli* and testing for enzymatic reduction of nitrite and nitrate, respectively (29). First, we estimated the consensus sequences based on GOS sequences from the low GC clusters discussed above and then synthesized each gene in vitro (Genscript Corporation). We cloned the synthesized putative *Prochlorococcus* as well as *Synechococcus* WH8102 *nir*A and *nar*B genes into pEcoli-Nterm 6xHN (Clontech). After verifying the inserts by sequencing, we transformed *E. coli* BL21 cells (Invitrogen) with each plasmid. We initiated expression by adding 1 mM IPTG to *E. coli* BL21 (including 1 of the 4 plasmids) growing in LB with 100 mg/L ampicillin at 22 °C. After harvesting BL21 cells containing NirA, we added alkyltrimethylammonium bromide 50 mg/L, 200 μM nitrite, 50 mM phosphate buffer (pH 8), and 4 mM methyl-viologen (Sigma-Aldrich). BL21 cells containing NarB were mixed with 50 mg/L alkyltrimethylammonium bromide, 20 mM nitrate, 50 mM carbonate (pH 10.5), and 4 mM methyl-viologen. After 10-min incubation at 30 °C, both reactions were started by adding freshly prepared 2g/L sodium dithionite in 30 mM bicarbonate. The reactions were stopped by vigorous vortexing, followed by adding 0.1 M NaOH and $ZnSO_4$. To verify the enzymatic function of NirA and NarB, we sampled the reaction mixture at t = 0, 5, and 15 min, and monitored the decrease or increase in nitrite, respectively. We used the reaction mixture with no cells and cells containing NarB as control for NirA and reversely for NarB. Nitrite was measured colormetrically at 540 nm by adding 20 mM sulfanilamide and 0.25 mM *N*-(1-naphthyl)-ethylenediamine and incubated for 15 min at 22 °C. All concentrations are final.

Field Expression of *Prochlorococcus nirA* and *nar*B. Primers specific to the *Prochlorococcus nir*A (position 874 to 1292) and *nar*B (position 1405 to 1942) were designed based on sequences identified in a recently published genomic DNA library from the North Pacific Subtropical Gyre (20). Single step RT-PCR using SuperScript III One-Step RT-PCR System with Platinum Taq (Invitrogen) was used to assay for expression of putative *Prochlorococcus nir*A and *narB* genes in linear amplified RNA prepared from the same microbial community. The primers HOTnirA874f (GAT CCA GGC TCA RTT TTC AAC G) and HOTnirA1292r (CCT GTC CAG TGG ATT TTA ACT TCC) were used to amplify *nirA* with an annealing temperature of 50 °C, and the primers HOTnarB1405F (TGC CCA TCT TTC AGA GAA TCG) and HOTnarB1942R (TTG CTT GAA CTT CCC CTC TTT TG) were used to amplify *narB* with an annealing temperature of 51 °C. PCR (omitting the reverse transcriptase) was performed as a control to confirm that trace amounts of genomic DNA were not amplified. RT-PCR products were cloned into pCR4 using TOPO TA cloning kit (Invitrogen), sequenced, and aligned against the GOS *Prochlorococcus nirA* and *nar*B consensus sequences.

Distribution of Nitrogen Assimilation Genes in GOS Samples. In each GOS sample, we estimated the abundance of marine *Synechococcus*, HL, and LL *Prochlorococcus* based on the number of reciprocal best hits to 8 core single copy genes from marine *Synechococcus* and *Prochlorococcus* genomes (*gln*A, *gyr*B, *hem*A, *rec*A, *rpl*10, *rpo*B, *rps*D, and *tyr*S). We searched all GOS samples for sequences similar to these core genes using input sequences from all *Prochlorococcus* and marine *Synechococcus* genomes (tblastn, e value = 1E−2). Second, we found the reciprocal match of each unique GOS hit in a database with all sequenced genomes (as of 01/14/08) using blastx (e value = 1E−30), and scored a hit when the best match was marine *Synechococcus*, HL or LL *Prochlorococcus*. For each core gene, we normalized the abundance by average sequence length. We then estimated the occurrence of high (> 50%) and low GC (< 40%) copies of the nitrogen assimilation genes *nir*A, *nir*X, *nar*B, *nap*A, *moa*ABCDE, *moe*A, *mob*A, *nar*X1, and *nar*X2 most similar to marine *Synechococcus*. Again, we searched all GOS samples using input sequences from all marine *Synechococcus* as well as the translated consensus sequence from the low GC clade to ensure that we found all hits in the GOS samples. Then, we found the reciprocal match of each unique GOS hit and scored a hit when the best match was marine *Synechococcus.* We normalized the abundance by average sequence length.

ACKNOWLEDGMENTS. We thank Antonia Herrero for assistance with the nitrate reductase assay; Penny Chisholm, Debbie Lindell, and Jennifer Martiny for many helpful comments on the manuscript; and the J. Craig Venter Institute (San Diego, CA) and Gordon and Betty Moore Foundation (GBMF, Palo Alto, CA) for allowing early access to the GOS samples from the Indian Ocean. RNA from the North Pacific Subtropical Gyre was provided by the DeLong and Chisholm laboratories at Massachusetts Institute of Technology. This work was supported in part by the University of California, Irvine; Paul Berube's contribution was supported by a grant from GBMF to S. Chisholm.

1. Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998) Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* 281:237–240.
2. Dugdale RC, Goering JJ (1967) Uptake of new and regenerated forms of nitrogen in primary productivity. *Limnol Oceanogr* 12:196–206.
3. Moore JK, Doney SC, Kleypas JA, Glover DM, Fung IY (2002) An intermediate complexity marine ecosystem model for the global domain. *Deep Sea Res (II Top Stud Oceanogr)* 49:403–462.
4. Liu H, et al. (1998) *Prochlorococcus* and *Synechococcus* growth rates and contributions to production in the Arabian Sea during the 1995 Southwest and Northeast Monsoons. *Deep Sea Res (II Top Stud Oceanogr)* 45:2327–2352.
5. Moore LR, Rocap G, Chisholm SW (1998) Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* 393:464–467.
6. Moore LR, Post AF, Rocap G, Chisholm SW (2002) Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus. Limnol Oceanogr* 47:989–996.
7. DuRand MD, Olson RJ, Chisholm SW (2001) Phytoplankton population dynamics at the Bermuda Atlantic Time-series Station in the Sargasso Sea. *Deep Sea Res (II Top Stud Oceanogr)* 48:1983–2003.
8. Blanchot J, Rodier M (1996) Picophytoplankton abundance and biomass in the western tropical Pacific ocean during the 1992 El Nino year: Results from flow cytometry. *Deep Sea Res* 43:877–895.
9. Partensky F, Blanchot J, Lantoine F, Neveux J, Marie D (1996) Vertical structure of picophytoplankton at different trophic site of the subtropical northeastern Atlantic. *Deep Sea Res (I Oceanogr Res Pap)* 43:1191–1213.
10. Bouman HA, et al. (2006) Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes. *Science* 312:918–921.
11. Palenik B, et al. (2003) The genome of a motile marine *Synechococcus*. *Nature* 424:1037–1042.
12. Rocap G, et al. (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* 424:1042–1047.
13. Casey JR, Lomas MW, Mandecki J, Walker DE (2007) *Prochlorococcus* contributes to new production in the Sargasso Sea deep chlorophyll maximum. *Geophys Res Lett* 34:L10604.
14. Vaulot D, Partensky F (1992) Cell-cycle distributions of prochlorophytes in the North-Western Mediterranean-Sea. *Deep Sea Res* 39:727–742.
15. Martiny AC, Tai APK, Veneziano D, Primeau F, Chisholm SW (2009) Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. *Environ Microbiol* 11:823–832.
16. Rusch DB, et al. (2007) The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol* 5:e77.
17. Coleman ML, et al. (2006) Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* 311:1768–1770.
18. Luque I, Herrero A, Flores E, Madueno F (1992) Clustering of genes involved in nitrate assimilation in the cyanobacterium *Synechococcus*. *Mol Gen Genet* 232:7–11.
19. DeLong EF, et al. (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* 311:496–503.
20. Frias-Lopez J, et al. (2008) Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci USA* 105:3805–3810.
21. Jenkins BD, Zehr JP, Gibson A, Campbell L (2006) Cyanobacterial assimilatory nitrate reductase gene diversity in coastal and oligotrophic marine environments. *Environ Microbiol* 8:2083–2095.
22. Rubio L, Flores E, Herrero A (1998) The *nar*A locus of *Synechococcus* sp. strain PCC 7942 consists of a cluster of molybdopterin biosynthesis genes. *J Bacteriol* 180:1200–1206.
23. DeLong EF (2005) Microbial community genomics in the ocean. *Nat Rev Microbiol* 3:459–469.
24. Galvan A, Quesada A, Fernandez E (1996) Nitrate and nitrite are transported by different specific transport systems and by a bispecific transporter in *Chlamydomonas reinhardtii*. *J Biol Chem* 271:2088–2092.
25. 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.
26. Boyer TP, et al. (2006) *World Ocean Database 2005* (U.S. Government Printing Office, Washington, DC).
27. Martiny AC, Huang Y, Li W (2009) Occurrence of phosphate acquisition genes in *Prochlorococcus* cells from different ocean regions. *Environ Microbiol*, 11:1340–1347.
28. Felsenstein J (2006) PHYLIP (Department of Genome Sciences, University of Washington, Seattle), Version 3.66.
29. Herrero A, Flores E, Guerrero MG (1985) Regulation of nitrate reductase cellular levels in the cyanobacteria *Anabaena variabilis* and *Synechocystis* sp. *FEMS Microbiol Lett* 26:21–25.