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.