Phosphate acquisition genes in *Prochlorococcus* ecotypes: Evidence for genome-wide adaptation

Adam C. Martiny*, Maureen L. Coleman*, and Sallie W. Chisholm†

Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by Rita R. Colwell, University of Maryland, College Park, MD, and approved June 28, 2006 (received for review February 20, 2006)

The cyanobacterium *Prochlorococcus* is the numerically dominant phototroph in the oligotrophic oceans. This group consists of multiple ecotypes that are physiologically and phylogenetically distinct and occur in different abundances along environmental gradients. Here we examine adaptations to phosphate (P) limitation among ecotypes. First, we used DNA microarrays to identify genes involved in the P-starvation response in two strains belonging to different ecotypes, MED4 (high-light-adapted) and MIT9313 (low-light-adapted). Most of the up-regulated genes under P starvation were unique to one strain. In MIT9313, many ribosomal genes were down-regulated, suggesting a general stress response in this strain. We also observed major differences in regulation. The P-starvation-induced genes comprise two clusters on the chromosome, the first containing the P master regulator *phoB* and most known P-acquisition genes and the second, absent in MIT9313, containing genes of unknown function. We examined the organization of the *phoB* gene cluster in 11 *Prochlorococcus* strains belonging to diverse ecotypes and found high variability in gene content that was not congruent with rRNA phylogeny. We hypothesize that this genome variability is related to differences in P availability in the oceans from which the strains were isolated. Analysis of a metagenomic library from the Sargasso Sea supports this hypothesis; most *Prochlorococcus* cells in this low-P environment contain the P-acquisition genes seen in MED4, although a number of previously undescribed gene combinations were observed.

genome evolution | microarrays | phoB

The oceans play a key role in global nutrient cycling and climate regulation. The unicellular cyanobacterium *Prochlorococcus* is a significant contributor to these processes, because it accounts for ≈30% of primary productivity in midlatitude oceans (1). *Prochlorococcus* is composed of closely related physiologically distinct cells, enabling proliferation of the group as a whole over a broad range of environmental conditions (2). Early observations revealed that there are two genetically and physiologically distinct types of *Prochlorococcus*, high-light (HL) and low-light (LL)-adapted (2), which are distributed differently in the water column (3, 4). Cells belonging to these two groups differ not only in light optima and pigmentation (5) but also in nitrogen (6) and phosphorus (7) utilization capabilities, presumably adaptations that are related to depth-dependent nutrient concentrations.

The HL and LL groups can be further divided into at least six clades (two HL- and four LL-adapted) based on the phylogeny of the 16S/23S rRNA internal transcribed spacer region (8). The relative abundance of cells belonging to these clades has been measured in several ocean regions, revealing patterns that agree, for the most part, with their HL/LL phenotype: HL-adapted cells dominate the surface mixed layer, and LL-adapted cells most often dominate in deeper waters (3, 9–12). By combining physiological studies of isolates and clade abundance in the ocean, it was recently shown that temperature, in addition to light, is an important determinant of the ocean-scale abundance of these six phylogenetic clades (12). Based on the observed correlations between phylogenetic origin, physiological properties, and environmental distributions, these six clades are considered ecotypes, i.e., distinct phylogenetic clades with ecologically relevant physiological differences (2, 13).

A closer examination of physiological properties among cultured isolates reveals variability that is not consistent with their phylogenetic relationships. For example, some LL-adapted strains can use nitrite as sole nitrogen source, whereas others require ammonium (6). Moreover, one HL-adapted strain (MED4) can grow on organic phosphates as a sole phosphorus source, whereas another (MIT9312) and a LL-adapted strain (MIT9313) cannot (7). Thus strains with similar temperature and light optima for growth can vary in nutrient assimilation capabilities. This implies that nutrient adaptation has occurred more recently than adaptation to light and temperature gradients. One mechanism for rapid adaptation to a specific environment is the acquisition of genes by lateral transfer. Indeed, several key genes involved in nutrient assimilation in *Prochlorococcus* are thought to be of foreign origin (13), and we have recently identified variable genomic islands in *Prochlorococcus*, thought to have arisen by lateral gene transfer (14), that contain a number of genes involved in nutrient assimilation.

To better understand the relationship between variability in nutrient acquisition mechanisms, phylogeny, and light adaptation, we undertook a detailed analysis of phosphate (P) acquisition in *Prochlorococcus*. We first identified P-starvation-induced genes in HL- and LL-adapted isolates using DNA microarrays. Having identified these genes, we then analyzed their distribution among the genomes of 11 phylogenetically diverse *Prochlorococcus* strains. Finally, we compared these findings with the collective P-acquisition gene content of a natural *Prochlorococcus* population from the surface waters of the Sargasso Sea, which is periodically P-limited.

Results and Discussion

Identification of Differentially Expressed Genes Under P Starvation. To determine genes involved in the P-starvation response in *Prochlorococcus*, we subjected strains MED4 (HL-adapted) and MIT9313 (LL-adapted) to abrupt P limitation and monitored changes in gene expression. To initially map the time course of the response, we used quantitative RT-PCR to measure expression levels of *pstS*, which encodes a periplasmic P-binding protein known to be induced under P-limiting conditions in many cyanobacteria, including MED4 (15). The temporal profile of the P-starvation response differed significantly between the two strains. In MED4, the transcript level of *pstS* began to increase 12 h after cells were resuspended in P-free medium (Fig. 1*A*) and

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: HL, high-light; LL, low-light; P, phosphate.

Data deposition: Orthologs to genes in the MED4 *phoB* region reported in this paper have been deposited in the GenBank database (accession nos. DQ786954–DQ787011 and DQ856305–DQ856313).

*A.C.M. and M.L.C. contributed equally to this work.

†To whom correspondence should be addressed. E-mail: chisholm@mit.edu.

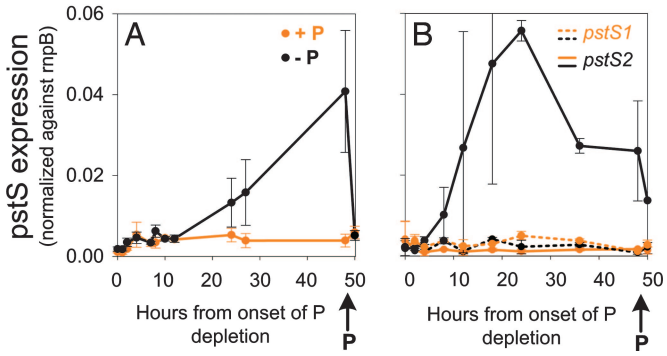

Fig. 1. Time course of expression of *pstS* in *Prochlorococcus* cells resuspended in medium with no added P at 0 h (black lines), compared to cells resuspended in P-replete medium (orange lines). Arrows indicate P addition after 48 h. (*A*) MED4: *pstS* is ORF PMM0710. (*B*) MIT9313: *pstS1* is ORF PMT0508 (dashed lines), and *pstS2* is ORF PMT0993 (solid lines).

increased steadily until P was added at 48 h. This release from P starvation caused a rapid decline in transcript level, which reached the control value within 2 h. In MIT9313, which has two copies of *pstS*, the expression of one (*pstS1*) was unresponsive to P starvation, whereas that of the other (*pstS2*) was elevated 50-fold by 24 h (Fig. 1*B*), followed by a decline. The addition of P to the medium after 48 h appeared to accelerate this decrease. Despite 94% amino acid sequence identity between the two copies of *pstS* in MIT9313, the genes responded very differently to P starvation. The function of *pstS1* is unknown.

We next examined genome-wide differences in gene expression in response to P starvation between the two strains. In MED4, a progressive induction of genes was observed over 48 h after the cells were resuspended in P-free medium. Thirty genes were significantly up-regulated, and four were down-regulated, by 48 h (Fig. 2*A*; Table 1, which is published as supporting information on the PNAS web site). The general response was different in MIT9313, where 176 genes were differentially expressed after 24 h, but most (143) were down-regulated (Fig. 2*B* and Table 2, which is published as supporting information on the PNAS web site). The high fraction of down-regulated genes, including many ribosomal proteins, could indicate a general reduction in the metabolic rate of MIT9313 cells (16).

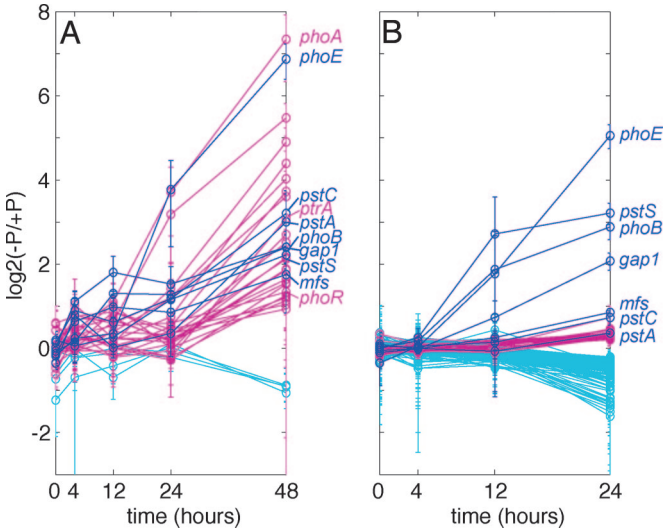

Fig. 2. Time course of gene expression in P-starved *Prochlorococcus* cultures. Differentially expressed genes ($q < 0.05$) in MED4 (*A*) and MIT9313 (*B*). Dark-blue lines indicate genes that were up-regulated in both strains, magenta lines are genes up-regulated in only one strain, and light-blue lines are genes down-regulated in only one strain. Error bars represent one standard deviation of fold change.

Only seven up-regulated genes were common to both strains (blue lines with gene names in Fig. 2). Most are orthologs to *Escherichia coli* genes implicated in P scavenging, such as the response regulator (*phoB*) and the transport system for orthophosphate (*pstABCS*). A porin gene located just downstream from *phoB* (PMM0709 in MED4 and PMT0998 in MIT9313) was also induced in both strains, and we propose that this gene encodes *phoE*, which is known to facilitate transport of orthophosphate across the outer membrane in other organisms. In addition to known P-starvation genes, genes previously unassociated with P starvation were up-regulated in both strains (Fig. 2 and Tables 1 and 2). Only two of these genes were common to both MED4 and MIT9313: *gap1*, which encodes glyceraldehyde-3-phosphate dehydrogenase, and *mfs*, which encodes a major facilitator superfamily transporter. Both genes are located just downstream from *phoB*, suggesting they play an important but unknown role in the P-starvation response, as has been suggested (17).

A number of orthologs to genes involved in the P-starvation response in other bacteria (18) were not induced in either *Prochlorococcus* strain, including *phoH* (whose function is unknown) and phosphonate transport genes (*phnCDE*). The lack of an identifiable phosphonatase or C-P lyase gene suggests that *phnCDE* encode a transport system for a different substrate in *Prochlorococcus* or may be nonfunctional. Also, genes encoding polyphosphate utilization (*ppK* and *ppX*) did not respond to P starvation in either strain of *Prochlorococcus*, although they are known to respond in some bacteria (19).

Despite similarities between the responses of MED4 and MIT9313, there were also important differences. MIT9313 lacks an ortholog to the most highly up-regulated gene in MED4, *phoA*, encoding alkaline phosphatase, which cleaves P from organic compounds. *ptrA*, which encodes a transcription factor thought to be involved in the P-starvation response (17), is up-regulated 8-fold in MED4 (PMM0718), whereas MIT9313 carries only a remnant of this gene (between PMT0998 and -999) that is not expressed. Similarly, MIT9313 carries a pseudogene of the sensor kinase *phoR* (17), which was not up-regulated, whereas the intact version of this gene was up-regulated in MED4. Despite the absence of *phoR* expression, both *phoB* and *pstABCS*, which normally depend on *phoR*, were induced under P starvation in MIT9313. Several regulatory genes that do not have orthologs in MED4 (PMT0265, PMT1357, and PMT2151) were differentially expressed in MIT9313 (Table 2), and these may be involved in activating *phoB* and in turn *pstABCS*. The remaining differentially expressed genes are unique to either strain and are primarily of unknown function. They should be further examined as potentially important for shaping the ecotype-specific response to P starvation.

The genes that are differentially expressed under P starvation are not distributed randomly along the chromosomes of the two strains (Fig. 3, $P < 0.0001$). Fifteen are located in a 21-gene stretch of the genome in MED4 (PMM0705–PMM0725), which includes *phoB*, most of the known P-acquisition genes, and several transporters. MIT9313 lacks intact orthologs to eight of these 15 genes, but most of the remaining seven are similarly located in the "*phoB* region." In addition, MED4 contains a second cluster of up-regulated genes located between PMM1403 and PMM1416, which is part of a variable genomic island (14). This organization suggests that the gene cluster around *phoB* is involved in the uptake of various forms of P, whereas the second cluster encodes an unknown component of the P-starvation response.

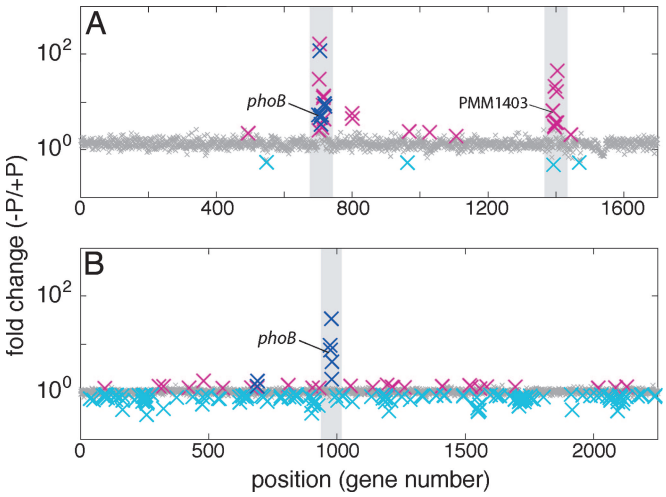

Fig. 3. Genome position of genes that were differentially expressed under P starvation in MED4 (*A*) and MIT9313 (*B*). The color code is the same as for Fig. 2 for differentially expressed genes; gray indicates genes with no significant ($q < 0.05$) change. The data plotted are from the 48-h time point in MED4 and the 24-h time point in MIT9313, the time of maximal *pstS* expression in each strain.

Genome Content and Organization of P-Acquisition Genes. Genes that are differentially expressed in response to P starvation in MED4 and MIT9313 were more likely to be lost or gained than randomly selected genes ($P < 0.0001$) in the genomes of 11 *Prochlorococcus* strains. In particular, genes found in the *phoB* region in MED4 are often missing or rearranged in the other genomes (Fig. 4*A*). Some strains (MED4, NATL1A, NATL2A, MIT9312, and MIT9301) share many orthologs with MED4, similarly grouped in a large cluster. In contrast, MIT9303, MIT9313, SS120, MIT9211, MIT9515, and AS9601 harbor fewer than half the *phoB* region genes found in MED4, and many of these are scattered throughout the genome.

This variability in genome content and architecture of P-acquisition genes is not related to phylogeny, as defined by rRNA sequence divergence (Fig. 4 *A* and *B*). Two HL-adapted strains belonging to the eMED4 clade (MIT9515 and MED4) share 99.9% 16S rRNA sequence identity, yet MIT9515 lacks orthologs to 15 MED4 genes from the *phoB* region. Similarly, three strains belonging to the eMIT9312 clade (MIT9312, MIT9301, and AS9601; 99.9% 16S rRNA identity) differ in gene content and organization relative to the MED4 *phoB* region. In fact, MIT9312 is more similar to MED4 and AS9601 to MIT9515 in terms of P-acquisition gene content (Fig. 4*A*), which is the inverse of their rRNA similarity. Thus it is reasonably clear, even from this limited data set, that the organization of P-acquisition genes in *Prochlorococcus* strains is not dictated by phylogenetic origin.

Ordering the genomes by gene content and organization relative to the MED4 *phoB* region, as depicted in Fig. 4*A*, reveals patterns that suggest that P availability in the waters from which these strains were isolated could influence genome content. MED4, the strain with the most-expansive *phoB* region, was isolated from surface waters in the northwest Mediterranean Sea, where the P concentration is typically <100 nM and has been shown to limit growth of cyanobacteria (20, 21). NATL1A and NATL2A, which possess orthologs to most of the MED4 *phoB* region genes, came from surface waters in the central North Atlantic Ocean, where surface P levels were between 50 and 150 nM (22) at the time these strains were isolated. Conversely, the strains with the fewest orthologs to the *phoB* region in MED4 (AS9601, MIT9515, and MIT9211) were isolated from ocean regions with high surface P levels (>600 nM; refs. 23 and 24). The remaining five strains in Fig. 4*A* contain an intermediate number of orthologs relative to the *phoB* region in MED4. Although they were isolated from regions where P concentrations are either low (<100 nM throughout the euphotic zone in Sargasso Sea) or variable (Gulf Stream; refs. 25 and 26), all came from deep in the euphotic zone (between 90 and 135 m). Light is likely the primary limiting factor for growth at this depth, perhaps relaxing selective pressure on the P-acquisition system. Thus, we predict that in P-limited environments, cells will contain many P-acquisition genes, primarily in a cluster around *phoB*.

Frequency of *Prochlorococcus* P-Acquisition Genes in the Sargasso Sea. To test this hypothesis, we examined gene stoichiometries in surface waters of the Sargasso Sea (27), where the P concentration is extremely low (25, 26). Indeed, all genes from the MED4 *phoB* region were present at roughly one copy per *Prochlorococcus* genome in this population (Fig. 4*C*). This includes genes between PMM0717 and PMM0722, which are largely absent from the other genomes, including ones affiliated with eMIT9312, the ecotype dominating this wild *Prochlorococcus* population (based on internal transcribed spacer sequence analysis from this data set). The abundance of P-acquisition genes similar to those found in MED4, in a population dominated by eMIT9312 cells, further supports our hypothesis that the regional environment influences the P-acquisition gene content of *Prochlorococcus* cells.

We also analyzed the frequency of occurrence of orthologs to the second up-regulated cluster in MED4 (spanning PMM1403 to -1416; Fig. 3*A*) in the Sargasso Sea population. As mentioned previously, the cluster is present only in MED4 and is located in a variable genomic island. In the Sargasso Sea, most genes from this cluster were present in a ratio close to 0.5 compared to core genes (data not shown), indicating that some, but not all, *Prochlorococcus* genomes contained these genes (see also ref 14). We discovered genome fragments containing genes from this island in proximity to known P-acquisition genes commonly found around *phoB* (Fig. 4*D*). These fragments demonstrated physical linkage between PMM1406 and *phoBR*, PMM1416 and *phoA* and several other combinations. This association of genes from two separate P-starvation-induced clusters in the MED4 genome supports the importance of these genes in responding to P limitation.

In MIT9301 and in several genome fragments from the Sargasso Sea, we also saw an intriguing linkage between genes found in the *phoB* region of MED4 and phosphonate uptake genes (*phnCDE*; Fig. 4 *A* and *D*). It has been proposed that phosphonates are an important phosphorus resource in marine ecosystems (28), but efforts to grow *Prochlorococcus* on phosphonates as a sole P source have been unsuccessful thus far. The clustering of phosphonate uptake genes and genes up-regulated under P starvation suggests that some *Prochlorococcus* lineages may be capable of using this organic phosphorus source.

Adaptation to P Limitation in *Prochlorococcus*. Our analysis revealed genomic variation among *Prochlorococcus* isolates that is not consistent with their rRNA-based phylogenetic relationships. We propose that these differences are related, in part, to the nutrient regime from which the cells were isolated. However, other forces are likely shaping genome content as well, such as phages using outer membrane proteins (e.g., PhoE) as receptors (29), crosstalk between regulatory circuits (e.g., PhoBR; ref. 30), and limitation by other factors (e.g., light). Stochastic variation may also play a role.

Lateral gene transfer may explain the lack of correspondence between the gene complements of the strains and their phylogenetic relationships. The *pstS* gene is encoded in the genomes

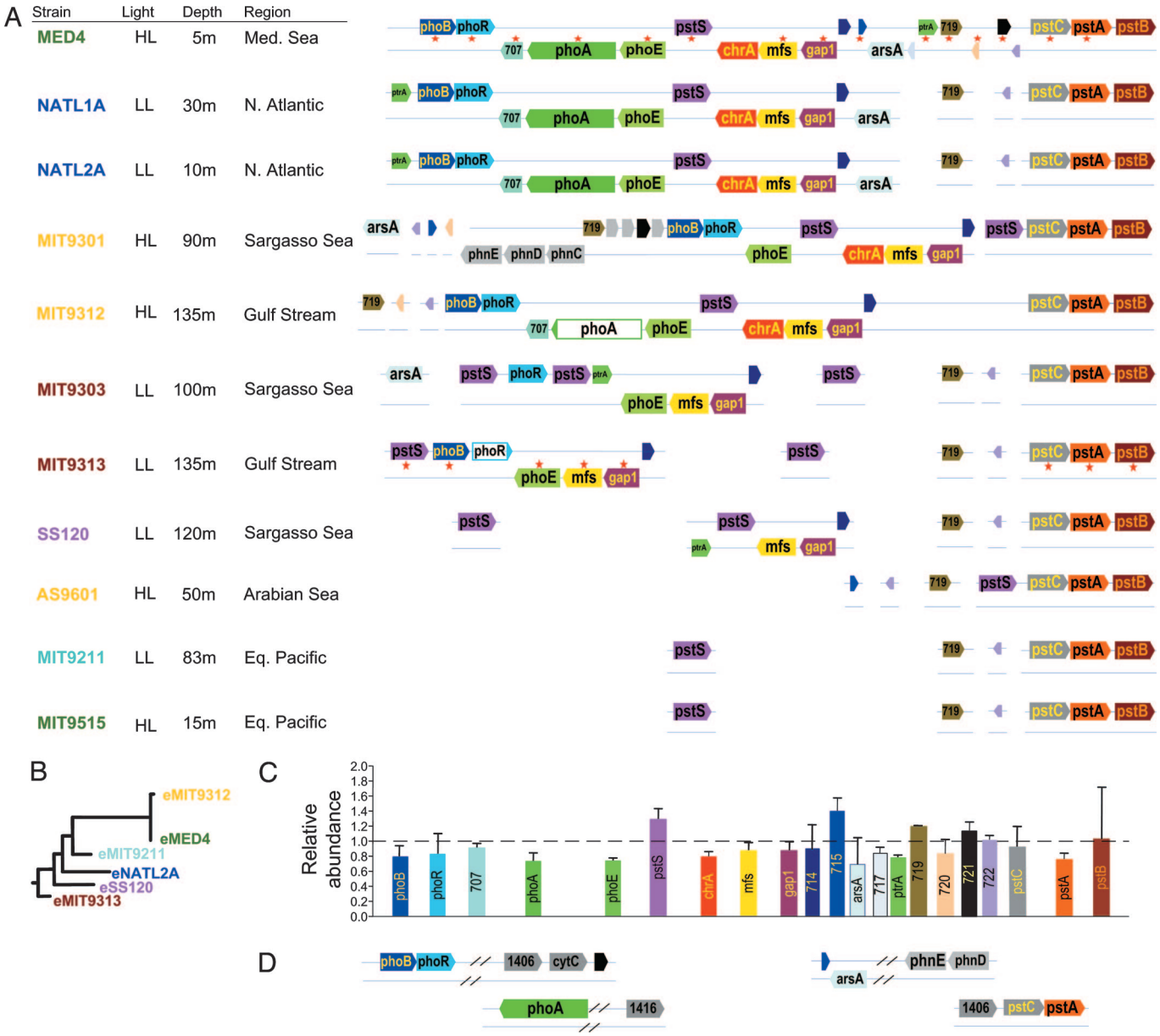

Fig. 4. P-acquisition genes in *Prochlorococcus*. (*A*) Genes located in proximity to *phoB* in MED4 (at the top) and the presence of their orthologs in the genomes of 11 *Prochlorococcus* strains. A red star indicates a gene that was significantly up-regulated in MED4 or MIT9313 from the microarray analyses. Gene numbers refer to PMM0XXX in MED4. Unfilled genes are likely pseudogenes. Color coding of strain names reflects ecotype affiliation shown in *B* (2). (*B*) Schematic of the phylogenetic relationship among different *Prochlorococcus* ecotypes (9). (*C*) Gene frequency in small insert libraries from the surface waters of the Sargasso Sea (27). Error bars indicate standard deviation of abundance based on all 150-bp fragments covering a gene. (*D*) Examples of genomic variants in the Sargasso Sea, showing linkage between genes found in the *phoB* region of MED4 and genes found elsewhere in the MED4 genome. Diagonal lines represent unknown sequence between two end reads of a clone in the data set.

of cyanophages that infect *Prochlorococcus* (31), suggesting a mechanism for moving genes across phylogenetic clades, and there is evidence that *phoA* and other genes involved in nutrient assimilation have been acquired laterally in some *Prochlorococcus* lineages (13). Furthermore, we observed that genes clustered in a variable genomic island in MED4 are up-regulated during P starvation (14). We were unable to detect any other obvious events of lateral gene transfer in the *phoB* region using phylogenetic analysis, but we anticipate that these events will become apparent as the sequences of more genomes from marine environments become available.

Unlike the P-starvation response, some traits, such as adaptations to light and temperature, are consistent with the phylogeny of *Prochlorococcus* (2, 12). One explanation for this difference is that photosynthesis requires a large protein complex that does not readily incorporate whole genes from foreign organisms (32, 33), and temperature adaptation can occur through genome-wide changes in amino acid and membrane lipid composition (34, 35). In contrast, the acquisition of a few key genes can rapidly change the spectrum of nutrient sources for a cell (e.g., nitrite reductase and alkaline phosphatase). A simplified calculation (see *Materials and Methods*) shows that if a *Prochlorococcus* cell acquires genes that improve growth rate by 1%, its progeny will dominate the entire population in an ocean basin in a few decades. This time scale is comparable to the observed domain shift in the North Pacific Ocean gyre from

a nitrogen- to a P-controlled state, purportedly fueled by increased nitrogen fixation in this region (36). Considering the strong feedback between the metabolic activity of *Prochlorococcus* (and all phytoplankton) and the local nutrient regime (37), understanding this type of genomic adaptation may be crucial for understanding shifts in biogeochemical processes in the oceans.

Materials and Methods

Culture Conditions. *Prochlorococcus* strains were grown at 22°C in Pro99 medium (6). Before the experiment, cultures were maintained in continuous light in log-phase growth at an irradiance of 12 μE m^{-2}·s^{-1} [E, einstein (1 mol of photons)] for MIT9313 (growth rate = 0.18 d^{-1}), and 30 μE·m^{-2}·s^{-1} for MED4 (growth rate = 0.27 d^{-1}) for >30 generations. Chlorophyll fluorescence was monitored on a Synergy HT fluorometer (BioTek, Burlington, VT).

P-Starvation Time Series. To induce P starvation, triplicate 4-liter cultures were harvested by centrifugation (10,000 × *g*), split in two, and washed twice in either P-replete (Pro99 with 50 μM PO_4) or -depleted (Pro99 with no added PO_4) medium and resuspended in 2 liters of the same medium. Samples were taken for RNA extraction, microarray hybridization, and quantitative RT-PCR (qRT-PCR) analysis at 0, 4, 12, 24, and 48 h after resuspension. Additional samples were taken for qRT-PCR at selected time points. After 48 h, 50 μM P was added to the P-depleted cultures to monitor the recovery response.

RNA Extraction. RNA was isolated according to ref. 38. In brief, cells were harvested by centrifugation (10,000 × *g*), resuspended in storage buffer (200 mM sucrose/10 mM NaOAc, pH 5.2/5 mM EDTA) and stored at −80°C. Before RNA extraction, MIT9313 cells were treated with 10 μg/μl lysozyme (Sigma, St. Louis, MO) for 1 h at 37°C (39). Total RNA was extracted by using the mirVana miRNA kit (Ambion, Austin, TX). DNA was removed by using Turbo DNase (Ambion). RNA was concentrated by ethanol precipitation and resuspended in milli-Q water.

Quantitative RT-PCR. RNA (2–10 ng of total RNA) was reverse-transcribed by using 100 units of SuperScript II (Invitrogen, Carlsbad, CA) in the presence of 200 units of SuperaseIN (Ambion). Primers are described in Table 3, which is published as supporting information on the PNAS web site. The resulting cDNA was diluted 5-fold in 10 mM Tris, pH 8. Triplicate real-time PCRs were performed by using the Qiagen (Valencia, CA) SYBR green kit and the diluted cDNA as template. The following program was run on an MJ Research (Cambridge, MA) Opticon DNA engine: 15 min at 95°C, followed by 40 cycles of denaturation (95°C, 15 s), annealing (56°C, 30 s), and extension (72°C, 30 s), followed by 5 min at 72°C. cDNA for *pstS* was quantified relative to *rnpB* by using the Δ-Δ C_T method (40).

Array Analysis. cDNA synthesis, labeling, and hybridization onto custom MD4–9313 Affymetrix (Santa Clara, CA) microarrays was done following the standard Affymetrix protocol. The probe arrays were scanned, and data visualization was done with GeneSpring software (Version 7.1; Silicon Genetics, Palo Alto, CA). Normalization was done by using the Robust Multichip Average algorithm (41) implemented in GeneSpring. Bayesian statistical analysis was applied to identify differentially expressed genes using Cyber-T (42). The Bayesian estimate of variance, which incorporates both the experimental variance for a given gene and variance of genes with similar expression levels (42), was calculated by using window sizes of 81 for MED4 and 101 for MIT9313 and a confidence value of 10 for both strains. A *t* test was then performed on log-transformed expression values by using the Bayesian variance estimate. To account for the multiple *t* tests performed, we used the program QVALUE, which measures significance in terms of the false discovery rate (43). A gene was identified as differentially expressed if the *q* value was <0.05. Signal intensities of individual probes targeting intergenic regions and potential miscalled ORFs were extracted by using Intensity Mapper (Affymetrix).

Tests for Clustering and Selective Loss/Gain of Induced Genes. We tested whether differentially expressed genes were distributed randomly along the genome by comparing the gene distance (in base pairs) against a simulated random distribution of genes. The weighted gene distance (*d*) was calculated by using the following decay function (adjusted for a circular genome):

$$d = \sum_i \text{sort} \sum_j \frac{1}{j} (n_i - n_{j+l}), \qquad [1]$$

where $i, j = 1, 2, \ldots,$ number of expressed genes, and n = position in genome. The second summation is based on a sorted array to nearest neighbor of n_i (i.e., $n_i - n_1 = 0$). The physical distance between differentially expressed genes was then compared to the *d* value of *i* randomly selected genes (10,000 permutations). We also tried other decay functions (e.g., different log bases of $n_i - n_j$) as well as using gene order as a measure for distance instead of actual base-pair difference, but all summations yielded the same result.

We also tested whether differentially expressed genes in MED4 (34 genes) and MIT9313 (176 genes) were more commonly lost or gained compared to randomly selected genes in the other *Prochlorococcus* genomes. We randomly chose 34 genes in the MED4 genome, counted the total number of orthologs to these 34 genes in the other 10 genomes, and repeated this process 10,000 times to generate a distribution. We then tested whether the total number of orthologs of the 34 differentially expressed genes in MED4 fell significantly outside this distribution. We repeated the test using the 176 differentially expressed MIT9313 genes. Orthologs were identified as pairwise best blastp hits. To further support the ortholog assignments, we constructed phylogenetic trees (maximum parsimony) for each gene in the MED4 *phoB* region and its putative orthologs.

Blast Analysis of Sargasso Sea Shotgun Library. We examined the occurrence of genes found in the *phoB* region of MED4 (between PMM0705 and PMM0725), in the Sargasso Sea environmental sequence data set sampled in February 2003 (excluding samples 5, 6, and 7; ref. 27). We used MED4 as the template for PMM0715 to PMM0722 and MIT9312 for the remaining genes. A sliding window of 150-bp fragments (step length = 50 bp) from the *phoB* region was first searched (blastn or tblastx; ref. 44) against the environmental sequence data set. A positive hit was scored if the environmental sequence and the paired end recovered *Prochlorococcus* as best hit when searched against a database consisting of *Prochlorococcus*, marine *Synechococcus* (WH8102, CC9905, and CC9902), *Pelagibacter ubique*, and *Silicibacter pomeroyi*. The number of copies of a particular *phoB*-region gene in the Sargasso Sea data set was estimated by averaging the number of hits for 150-bp segments comprising that gene and normalized against the average occurrence of known single-copy genes in all sequenced Cyanobacteria: *cpeA*, *glnA*, *gyrB*, *hemA*, 16S/23S internal transcribed spacer region (single copy in HL *Prochlorococcus* clades), *recA*, *rpl10*, *rpoB*, *rpsD*, and *tyrS*.

Changes in Genotype Frequency as a Function of Relative Fitness. To calculate how long it might take a new genotype with slightly improved fitness to overtake a population of *Prochlorococcus* cells in an ocean, we used equation **11** from ref. 45:

$$\ln[x_1(t)/x_2(t)] = \ln[x_1(0)/x_2(0)] + st, \quad [2]$$

where $x_1(t)$ is the fraction of the new genotype, and $x_2(t)$ is the fraction of the ancestral genotype at time t (days). At $t = 0$, x_1 was set to 10^{-24}, and x_2 was set at 1, assuming 10^{24} cells in an ocean basin such as the Sargasso Sea (46). We assumed a growth rate of 0.5 per day^{-1} (47) for the ancestral genotype and an increase in growth rate (or relative fitness) of new genotype (s) of 1%, so $s = 0.005$ d^{-1}.

We thank Debbie Lindell for many helpful discussions and Robert Steen and Trent Rector at Harvard Biopolymer Facility for labeling RNA and hybridizing the microarrays. We also thank numerous members of the Chisholm and DeLong labs for helpful comments on the manuscript. This work was supported in part by a fellowship from the Danish National Science Foundation (to A.C.M.); a National Science Foundation Graduate Fellowship (to M.L.C.); and grants from the National Science Foundation, the Gordon and Betty Moore Foundation, and the U.S. Department of Energy GTL Program (to S.W.C.).

1. Goericke, R. & Welschmeyer, N. A. (1993) *Deep-Sea Res.* **40,** 2283–2294.
2. Moore, L. R., Rocap, G. & Chisholm, S. W. (1998) *Nature* **393,** 464–467.
3. West, N. J. & Scanlan, D. J. (1999) *Appl. Environ. Microbiol.* **65,** 2585–2591.
4. Urbach, E. & Chisholm, S. W. (1998) *Limnol. Oceanogr.* **43,** 1615–1630.
5. Moore, L. R. & Chisholm, S. W. (1999) *Limnol. Oceanogr.* **44,** 628–638.
6. Moore, L. R., Post, A. F., Rocap, G. & Chisholm, S. W. (2002) *Limnol. Oceanogr.* **47,** 989–996.
7. Moore, L. R., Ostrowski, M., Scanlan, D. J., Feren, K. & Sweetsir, T. (2005) *Aquat. Microbial Ecol.* **39,** 257–269.
8. Rocap, G., Distel, D. L., Waterbury, J. B. & Chisholm, S. W. (2002) *Appl. Environ. Microbiol.* **68,** 1180–1191.
9. Ahlgren, N. A., Rocap, G. & Chisholm, S. W. (2006) *Environ. Microbiol.* **8,** 441–454.
10. Zinser, E. R., Coe, A., Johnson, Z. I., Martiny, A. C., Fuller, N. J., Scanlan, D. J. & Chisholm, S. W. (2006) *Appl. Environ. Microbiol.* **72,** 723–732.
11. West, N. J., Schonhuber, W. A., Fuller, N. J., Amann, R. I., Rippka, R., Post, A. F. & Scanlan, D. J. (2001) *Microbiology* **147,** 1731–1744.
12. Johnson, Z. I., Zinser, E. R., Coe, A., McNulty, N. P., Woodward, E. M. & Chisholm, S. W. (2006) *Science* **311,** 1737–1740.
13. Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A., Arellano, A., Coleman, M., Hauser, L., Hess, W. R., *et al.* (2003) *Nature* **424,** 1042–1047.
14. Coleman, M. L., Sullivan, M. B., Martiny, A. C., Steglich, C., Barry, K., Delong, E. F. & Chisholm, S. W. (2006) *Science* **311,** 1768–1770.
15. Scanlan, D. J., Silman, N. J., Donald, K. M., Wilson, W. H., Carr, N. G., Joint, I. & Mann, N. H. (1997) *Appl. Environ. Microbiol.* **63,** 2411–2420.
16. Nomura, M. (1999) *J. Bacteriol.* **181,** 6857–6864.
17. Scanlan, D. J. & West, N. J. (2002) *FEMS Microbiol. Ecol.* **40,** 1–12.
18. Wanner, B. L. (1996) in *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ed. Neidhardt, F. C. (Am. Soc. Microbiol., Washington, DC), pp. 1357–1381.
19. Kornberg, A., Rao, N. N. & Ault-Riche, D. (1999) *Annu. Rev. Biochem.* **68,** 89–125.
20. Marty, J. C., Chiaverini, J., Pizay, M. D. & Avril, B. (2002) *Deep-Sea Res.* **49,** 1965–1985.
21. Vaulot, D., LeBot, N., Marie, D. & Fukai, E. (1996) *Appl. Environ. Microbiol.* **62,** 2527–2533.
22. Irwin, B. (2000) *Nutrient Data from the Atlantic, JGOFS Canada Data Sets 1989–1998* (Marine Enviromental Data Service, Department of Fisheries and Oceans, Canada). CD-ROM Version 1.0.
23. Coale, K. H., Johnson, K. S., Fitzwater, S. E., Gordon, R. M., Tanner, S., Chavez, F. P., Ferioli, L., Sakamoto, C., Rogers, P., Millero, F., *et al.* (1996) *Nature* **383,** 495–501.
24. Morrison, J. M., Codispoti, L. A., Gaurin, S., Jones, B., Manghani, V. & Zheng, Z. (1998) *Deep-Sea Res.* **45,** 2053–2101.
25. Cavender-Bares, K. K., Karl, D. M. & Chisholm, S. W. (2001) *Deep-Sea Res.* **48,** 2373–2395.
26. Wu, J., Sunda, W., Boyle, E. A. & Karl, D. M. (2000) *Science* **289,** 759–762.
27. Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., *et al.* (2004) *Science* **304,** 66–74.
28. Karl, D. M. & Björkman, K. M. (2002) in *Biogeochemistry of Marine Dissolved Organic Matter*, eds. Hansell, D. A. & Carlson, C. A. (Academic, London), pp. 250–366.
29. Ho, T. D. & Slauch, J. M. (2001) *J. Bacteriol.* **183,** 1495–1498.
30. Fisher, S. L., Jiang, W., Wanner, B. L. & Walsh, C. T. (1995) *J. Biol. Chem.* **270,** 23143–23149.
31. Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F. & Chisholm, S. W. (2005) *PLoS Biol.* **3,** e144.
32. Blankenship, R. E. (1992) *Photosynth. Res.* **33,** 91–111.
33. Shi, T., Bibby, T. S., Jiang, L., Irwin, A. J. & Falkowski, P. G. (2005) *Mol. Biol. Evol.* **22,** 2179–2189.
34. Morgan-Kiss, R. M., Priscu, J. C., Pocock, T., Gudynaite-Savitch, L. & Huner, N. P. A. (2006) *Microbiol. Mol. Biol. Rev.* **70,** 222–252.
35. Tekaia, F., Yeramian, E. & Dujon, B. (2002) *Gene* **297,** 51–60.
36. Karl, D. M., Letelier, R., Tupas, L., Dore, J. E., Christian, J. & Hebel, D. V. (1997) *Nature* **388,** 533–538.
37. Redfield, A. C. (1934) in *James Johnstone Memorial Volume*, ed. Daniel, R. J. (Univ. Press of Liverpool, Liverpool, U.K.).
38. Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M. & Chisholm, S. W. (2005) *Nature* **438,** 86–89.
39. Tolonen, A. C., Aach, J., Lindell, D., Johnson, Z. I., Rector, T., Steen, R., Church, G. M. & Chisholm, S. W. (2006) *Mol. Syst. Biol.*, in press.
40. Livak, K. J. & Schmittgen, T. D. (2001) *Methods* **25,** 402–408.
41. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. (2003) *Bioinformatics* **19,** 185–193.
42. Baldi, P. & Long, A. D. (2001) *Bioinformatics* **17,** 509–519.
43. Storey, J. D. & Tibshirani, R. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 9440–9445.
44. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402.
45. Dykhuizen, D. E. & Hartl, D. L. (1983) *Microbiol. Rev.* **47,** 150–168.
46. Partensky, F., Hess, W. R. & Vaulot, D. (1999) *Microbiol. Mol. Biol. Rev.* **63,** 106–127.
47. Liu, H., Landry, M. R., Vaulot, D. & Campbell, L. (1999) *J. Geophys. Res.* **104,** 3391–3399.