Global Gene Expression of Prochlorococcus ecotypes in Response to Changes in Nitrogen Availability

Molecular Systems Biology (2006) doi:10.1038/msb4100087

www.molecularsystemsbiology.com
Article number: 53

molecular systems biology

Global gene expression of *Prochlorococcus* ecotypes in response to changes in nitrogen availability

Andrew C Tolonen[1], John Aach[2], Debbie Lindell[3], Zackary I Johnson[4], Trent Rector[2], Robert Steen[2], George M Church[2] and Sallie W Chisholm[3,*]

[1] Department of Biology, MIT/WHOI Joint Program in Oceanography, Cambridge, MA, USA, [2] Department of Genetics, Harvard Medical School, Boston, MA, USA, [3] Department of Civil and Environmental Engineering, MIT, Cambridge, MA, USA and [4] Department of Oceanography, University of Hawaii, Honolulu, HI, USA
* Corresponding author. Department of Civil and Environmental Engineering, MIT, MIT 48-419, 77 Mass. Ave., Cambridge, MA 02139, USA.
Tel.: +1 617 253 1771; Fax: +1 617 324 0336; E-mail: chisholm@mit.edu

Received 16.2.06; accepted 3.7.06

Nitrogen (N) often limits biological productivity in the oceanic gyres where *Prochlorococcus* is the most abundant photosynthetic organism. The *Prochlorococcus* community is composed of strains, such as MED4 and MIT9313, that have different N utilization capabilities and that belong to ecotypes with different depth distributions. An interstrain comparison of how *Prochlorococcus* responds to changes in ambient nitrogen is thus central to understanding its ecology. We quantified changes in MED4 and MIT9313 global mRNA expression, chlorophyll fluorescence, and photosystem II photochemical efficiency (F_v/F_m) along a time series of increasing N starvation. In addition, the global expression of both strains growing in ammonium-replete medium was compared to expression during growth on alternative N sources. There were interstrain similarities in N regulation such as the activation of a putative NtcA regulon during N stress. There were also important differences between the strains such as in the expression patterns of carbon metabolism genes, suggesting that the two strains integrate N and C metabolism in fundamentally different ways.
Molecular Systems Biology 3 October 2006; doi:10.1038/msb4100087
Subject Categories: functional genomics; microbiology & pathogens
Keywords: cyanobacteria; interstrain; nitrogen; *Prochlorococcus*; transcription

Introduction

The cyanobacterium *Prochlorococcus*, an oxygenic phototroph, is the most abundant member of the oceanic phytoplankton community in the tropical and subtropical ocean basins (Partensky *et al*, 1999) where it reaches densities as high as 7×10^5 cells ml^{-1} sea water (Campbell *et al*, 1998). Field studies have reported that up to 79% of primary productivity in the North Atlantic is due to *Prochlorococcus* (Li, 1994), showing that this organism plays a key role in the global carbon cycle. Nitrogen (N) concentrations in the oligotrophic ocean are extremely low and sometimes limit phytoplankton growth. For example, ammonium concentrations at the Bermuda Atlantic Time-Series Study (BATS) reach only 20–200 nM during bloom periods (Lipschultz, 2001). *Prochlorococcus* cells may have a particular propensity to become N deficient relative to phosphorus (P) because its cellular requirements for N relative to P are >20N:1P (Bertilsson *et al*, 2003) and thus exceed the 16N:1P Redfield ratio classically believed to describe the elemental composition of biomass in the sea (Redfield, 1958). In support of this hypothesis, field studies have shown that nitrogen enrichment stimulates *Prochlorococcus* growth in the North Atlantic (Graziano *et al*, 1996).

The *Prochlorococcus* strains used in this study, MED4 and MIT9313, represent clades that have different depth distributions in the ocean and thus occupy distinct ecological niches with respect to light and nitrogen. As such, MED4 and MIT9313 belong to ecologically distinct groups, called ecotypes (Moore *et al*, 1998). Field studies in the Atlantic revealed that the MED4 ecotype is relatively most abundant in the upper euphotic zone, whereas the MIT9313 ecotype is primarily confined to the base of the euphotic zone at depths around 100 m (West *et al*, 2001; Johnson *et al*, 2006). In accordance with their different depth distributions, MED4 cultures grow optimally at higher light intensities than MIT9313 (Moore and Chisholm, 1999).

Nitrogen appears to be an important selective agent driving niche differentiation of strains such as MED4 and MIT9313, as revealed by their distinct nitrogen utilization capabilities (Moore *et al*, 2002) and by the N metabolism genes in the MED4 (1.7 Mb) and MIT9313 (2.4 Mb) genomes (Rocap *et al*, 2003). MIT9313, for example, has nitrite transport and reduction genes (Rocap *et al*, 2003) and grows on nitrite,

whereas MED4 does not (Moore *et al*, 2002). This interstrain difference correlates with the depth distribution of the MIT9313 ecotype, as a well-defined nitrite maximum is often found in the lower euphotic zone (Olson, 1981) where this ecotype is most abundant. The MED4 genome contains genes for cyanate transport and reduction and this strain grows on cyanate (Garcia-Fernandez *et al*, 2004), whereas the MIT9313 genome lacks these genes (Rocap *et al*, 2003). In addition to the above differences, there are similarities in N utilization between MED4 and MIT9313. Both MED4 and MIT9313 have genes for the transport and utilization of ammonium and urea (Rocap *et al*, 2003) and grow on them as the sole N source (Moore *et al*, 2002). These N sources are rapidly recycled in the nutrient-depleted surface waters. Unlike other phytoplankton, including most marine *Synechococcus*, both MED4 and MIT9313 do not grow on nitrate and the gene for nitrate reduction, *narB*, is absent from both genomes (Rocap *et al*, 2003).

Because of the important role of nitrogen in *Prochlorococcus* ecology, the molecular mechanisms regulating the response to changes in ambient nitrogen are of particular interest. Previous studies in other cyanobacteria have shown that the transcription factor NtcA governs widespread transcriptional changes to enable survival during N starvation (Sauer *et al*, 1999). NtcA responds to the N status of the cell through changes in the level of the metabolite 2-oxoglutarate, the carbon skeleton used for the assimilation of nitrogen. Cellular levels of 2-oxoglutarate rise during N starvation (Muro-Pastor *et al*, 2005), resulting in an increased affinity of NtcA for binding DNA (Muro-Pastor *et al*, 2001; Tanigawa *et al*, 2002; Vazquez-Bermudez *et al*, 2002). NtcA alters transcription by binding the DNA sequence GTA-N8-TAC (Luque *et al*, 1994; Jiang *et al*, 2000; Herrero *et al*, 2001) in the promoters of its targets, which include genes for N assimilation and for the utilization of nitrogen sources other than ammonium (Vega-Palas *et al*, 1990; Luque *et al*, 1994).

The maintenance of cellular homeostasis requires that other biochemical pathways, such as those for carbon metabolism, also respond to N stress. Likened to the central processing unit (CPU) of the cell (Ninfa and Atkinson, 2000), the signal transducer P_{II} has been shown in other cyanobacteria to coordinate the cellular carbon and nitrogen balance. Similar to NtcA, P_{II} responds to levels of 2-oxoglutarate (Forchhammer, 1999; Tandeau de Marsac and Lee, 1999). Increases in 2-oxoglutarate levels enhance P_{II} phosphorylation (Forchhammer and Hedler, 1997). P_{II} controls the activity of transporters for nitrite/nitrate and for bicarbonate (reviewed by Forchhammer, 2004). The *glnB* gene, encoding the P_{II} protein, is itself transcriptionally activated by NtcA (Lee *et al*, 1999). Conversely, full activation of NtcA-regulated genes under N stress requires the *glnB* gene product, P_{II} (Paz-Yepes *et al*, 2003), suggesting that P_{II} and NtcA are functionally interdependent.

Many nitrogen metabolism genes in other cyanobacteria, including *ntcA* and *glnB*, are conserved in *Prochlorococcus*. Although *Prochlorococcus ntcA* expression is enhanced under N stress similar to other cyanobacteria (Lindell *et al*, 2002), several studies have concluded that aspects of *Prochlorococcus* N regulation are different from that of other cyanobacteria. Neither the abundance nor the activity of glutamine synthetase (GS) is changed under N stress (El Alaoui *et al*, 2001, 2003). The ammonium transporter *amt1* was found not to be transcriptionally activated during N stress and is proposed not to be NtcA-regulated (Lindell *et al*, 2002). The P_{II} protein is not phosphorylated under any tested conditions (Palinska *et al*, 2000). These differences have been proposed to be examples of a general streamlining of genetic regulation, which may represent an adaptation to a homogenous, oligotrophic environment (Garcia-Fernandez *et al*, 2004).

We asked how *Prochlorococcus* MED4 and MIT9313, which occupy different niches with respect to nutrients and light availability, respond to N stress and growth on different N sources. Our approach combined whole genome microarray analyses for both strains, physiological measurements, sequence analysis, and incorporated comparative genomics information from a previous study (Su *et al*, 2005). We identified clusters of genes that were coexpressed along an N starvation time series. Functional associations among genes within certain clusters were found using gene function categories based on homology with other cyanobacteria. Because of the important role of NtcA in governing the N stress response in other cyanobacteria, we examined its role in *Prochlorococcus* by correlating the genes differentially expressed during N starvation with NtcA binding site predictions. The lack of *Prochlorococcus* P_{II} phosphorylation has led previous studies to conclude that this organism does not coordinate C and N metabolism in response to changes in nitrogen (Giordano *et al*, 2005). We thus explored how changes in ambient N were propagated from N stress sensors to changes in genes controlling carbon metabolism. These findings are assimilated into a systems-level model comparing the transcriptional response to N stress in MED4 and MIT9313. Finally, we discuss how interstrain similarities and differences in N regulation give insight into the mechanisms by which niche partitioning occurs among *Prochlorococcus* strains in the ocean.

Results and discussion

The response to N starvation

Physiological response

The N starvation experiments compared changes in physiology and gene expression in cells following abrupt N deprivation (−N treatment) to cells growing in N-replete conditions ($+NH_4$ treatment). The growth rates of the $+NH_4$ cultures were 0.65 and 0.23 day^{-1} for MED4 and MIT9313, respectively (Figure 1A and B). Bulk chlorophyll fluorescence, a proxy for biomass in the log-phase cultures, began to decrease between 12 and 24 h in −N medium for both strains (Figure 1A and B). Photochemical conversion efficiency of photosystem II, as determined by F_v/F_m, dropped below optimal values of ∼0.65 on similar time frames to decreases in chlorophyll fluorescence (Figure 1C and D). A constant growth rate and photosystem II conversion efficiency was maintained in the $+NH_4$ cultures of both strains for the duration of the experiment (Figure 1).

Global transcriptional responses

The transcriptional response to N deprivation occurred more rapidly than the physiological response with differential

expression first appearing within 6 h for both strains (Figure 2). The number of differentially expressed genes at an individual time point reached a maximum at t=12 h for MED4 and then declined, whereas that of MIT9313 increased throughout the 48 h experiment (Figure 2). A cumulative total of 131 distinct MED4 genes from all time points were upregulated (7.4% of genes in the genome) and 168 were downregulated (9.5%) in response to N starvation. In MIT9313, a total of 120 distinct genes were upregulated across all time points and 251 were downregulated, representing 5.1 and 10.8% of the genes in the genome, respectively. Thus, despite MED4 having a smaller genome (1.7 Mb with 1716 genes) than MIT9313 (2.4 Mb with 2275 genes), MED4 responded to N stress by upregulating more genes. In contrast, MIT9313 downregulated more genes than MED4.

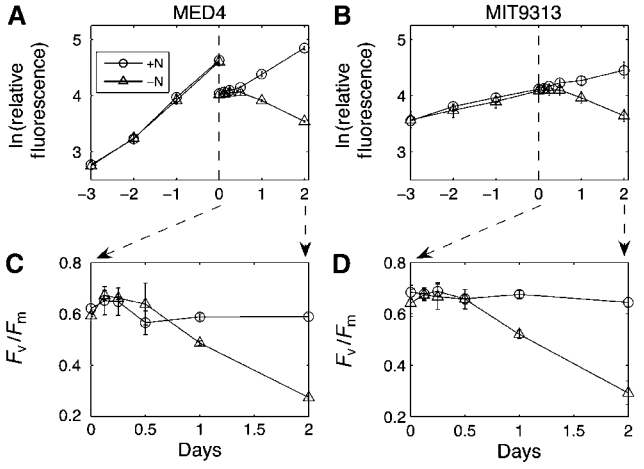

Figure 1 Effect of N starvation on culture fluorescence (**A, B**) and photochemical energy conversion efficiency F_v/F_m (**C, D**) in *Prochlorococcus* MED4 (**A, C**) and MIT9313 (**B, D**). Days −3, −2, and −1 correspond to culture growth prior to start of experiment. N-replete log-phase cells were spun down and resuspended in N-replete (+NH4, circles) and N-deficient (−N, triangles) media at t=0. The discontinuity in MED4 chlorophyll fluorescence at t=0 resulted from incomplete harvest of the cells before resuspension. Data points are means of replicates. Error bars show the range and are smaller than the symbols when not apparent.

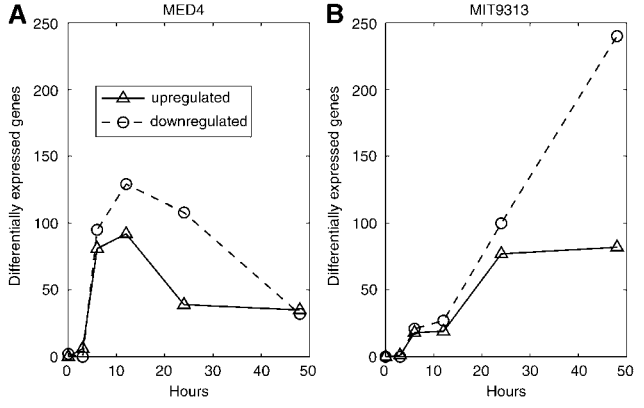

Figure 2 Number of differentially expressed genes (q<0.01) in *Prochlorococcus* MED4 (**A**) and MIT9313 (**B**) genes after the onset of N deprivation at t=0 for the cultures shown in Figure 1. Solid lines with triangles show upregulated genes and dashed lines with circles show repressed genes.

Clusters of differentially expressed genes that responded similarly over the N starvation time course, and may thus be functionally related, were identified using K-means clustering (Figure 3). MED4 has five clusters of upregulated genes and four clusters of downregulated genes, whereas MIT9313 genes were parsed into three upregulated clusters and four downregulated clusters. The dynamics of the global transcriptional response (Figure 2) is reflected in the patterns of the clusters. Both the upregulated and the downregulated MED4 clusters responded rapidly to N stress, but did not maintain an equally high level of differential expression. Overall, MIT9313 clusters responded more slowly but the level of differential expression increased throughout the experiment.

The number of clusters was chosen to maximize the mutual information between the clusters and CyanoBase functional categories (Nakamura *et al*, 1998) (see Materials and methods). For each cluster, we show the functional category with the greatest enrichment along with the statistical significance of this enrichment (Figure 3). In both strains, upregulated clusters have only weak functional category associations, suggesting that upregulated genes with similar

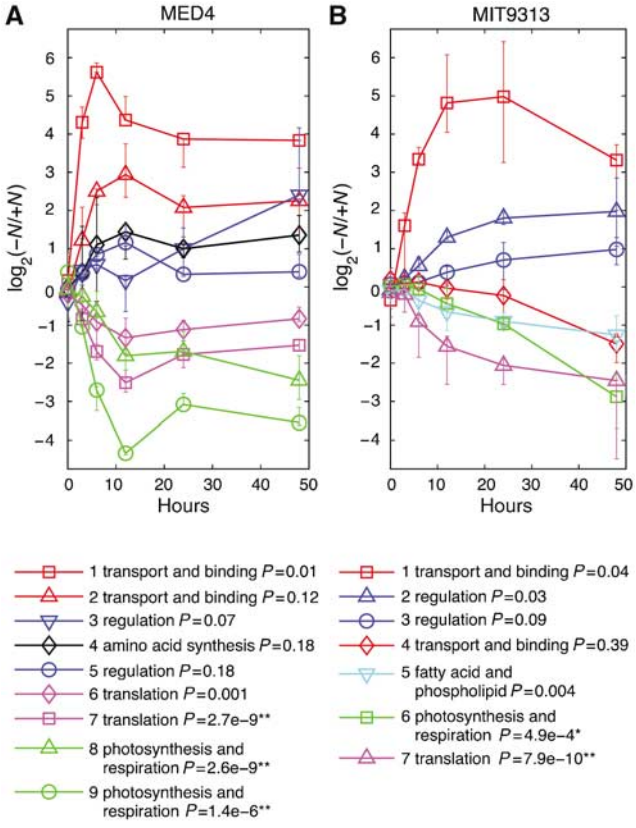

Figure 3 Temporal patterns of differentially expressed MED4 (**A**) and MIT9313 (**B**) genes clustered using the K-means algorithm. Each data point is the $\log_2$-transformed median expression of all genes in the cluster; bars show range from 25th to 75th percentile. Plots are colored to show the gene function category with the greatest enrichment: red=transport and binding, blue=regulatory, black=amino-acid synthesis, magenta=translation, green=photosynthesis and respiration, cyan=fatty acid/phospholipid/sterol metabolism. Legend shows P-values associated with functional enrichment of each cluster. Asterisks in legend denote significant enrichment for functional categories according to 'stringent' (**) or 'permissive' (*) significance thresholds (see Materials and methods).

expression profiles may be involved in diverse cellular functions. In contrast, highly downregulated clusters show strong enrichment for specific functional categories, namely Translation (MED4 cluster 7, MIT9313 cluster 7) and Photosynthesis and Respiration (MED4 clusters 8 and 9, MIT9313 cluster 6) (Figure 3).

Even though the functional enrichment in the upregulated clusters did not attain statistical significance, these clusters do contain noteworthy sets of functionally related genes. Cluster 1, the most rapidly and highly upregulated genes in each strain, contains N transport genes such as MED4 and MIT9313 *urtA*, MED4 *cynA*, and the MIT9313 nitrite permease. In both strains, the clusters revealed two distinct subsets of the *hli* genes: those that responded rapidly and highly (MED4 cluster 2 and MIT9313 cluster 1) and those that responded later and to a lesser degree (MED4 cluster 3 and MIT9313 cluster 2). Both strains also have an upregulated cluster containing two sigma factors, MED4 cluster 5 and MIT9313 cluster 2.

Examination of the individual genes composing the repressed clusters also reveals intercluster differences, even between clusters enriched for the same CyanoBase functional category. Although both MED4 clusters 8 and 9 are enriched for Photosynthesis and Respiration genes, MED4 cluster 8 contains numerous genes for photosystem I (*psaBDEIJKLM*) and cluster 9 contains ATP synthase subunits and the carbon fixation genes (*rbcLS*). MIT9313 cluster 6, the only cluster enriched for Photosynthesis and Respiration in this strain, contains genes for diverse aspects of photosystem I and II along with the phycoerythrin gene, *cpeB*. The repressed clusters that were significantly enriched for single functional categories (MED4 clusters 7–9 and MIT9313 clusters 6 and 7) also shed light on the role of a number of genes of unknown function. For example, these clusters contain seven MED4 genes and 42 MIT9313 genes categorized simply as 'hypothetical' or 'other'. The association of these genes with clusters highly enriched for a single functional category suggests that these genes may also be involved in the dominant cellular process of that cluster.

MIT9313 cluster 4 contains a number of genes that were unchanged until being repressed only at the final time point. The physiological measurements F_v/F_m and chlorophyll fluorescence show that the cells were in a severe state of starvation by this time. The genes in MIT9313 cluster 4 may thus represent those genes that are repressed as part of a general shutdown in transcription, rather than a specific N stress response. Interestingly, this MIT9313 cluster contains a number of genes linking N and C metabolism (*glnB*, *icd*, *acnB*, *rbcLS*). In MED4, some of these genes were upregulated (*glnB*, *icd*, *acnB*), whereas others were repressed (*rbcLS*).

Operon predictions

The expression of genes predicted to be in the same operon is often correlated, such as for *glnB* and its upstream partner in both strains (Figure 4C). We used the expression patterns during the N starvation time series to see if genes in predicted operons were correlated in their expression throughout the MED4 and MIT9313 genomes (see Materials and methods). For MED4, predicted operon positions 1 and n (n=2, 3, 4, 5) were all significantly more correlated than random gene pairs ($P<0.0004$ for each n) (Operon section of Supplementary information). For MIT9313, predicted operon positions 1 and 2 were significantly more correlated than random pairs ($P<0.0004$), whereas position pairs 1 and 3, and 1 and 4

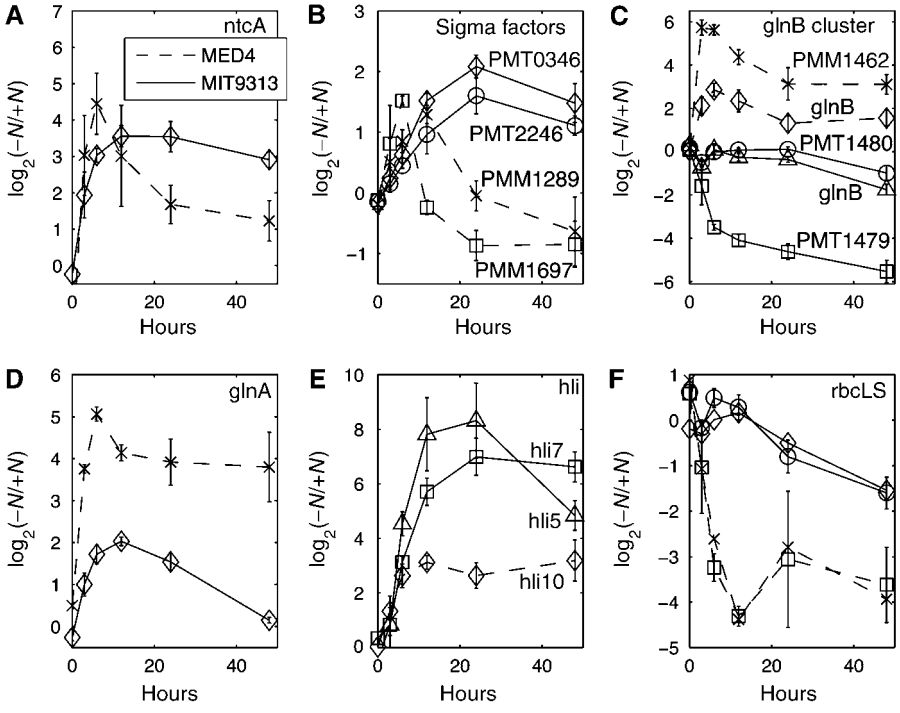

Figure 4 Comparison of MED4 and MIT9313 expression patterns under N stress for selected genes: *ntca* (**A**), sigma factors (**B**), *glnB* cluster (**C**), *glnA* (**D**), *hli* genes (**E**), and *rbcLS* (**F**). Data points show $\log_2$-transformed mean expression values of duplicate cultures; error bars show one standard deviation.

had elevated correlations that did not achieve statistical significance. Many predicted operons thus represent actual cotranscribed units. We, however, also found that the expression of tandem genes not in predicted operons was significantly correlated relative to random gene pairs ($P<0.0004$ for both MED4 and MIT9313), suggesting that many real operons were also missed by the predictions (Operon section of Supplementary information). Although the array data do not provide direct evidence whether tandem, coexpressed genes are indeed transcribed by the same mRNA, this analysis shows how array data help validate, and could potentially be used to optimize, *Prochlorococcus* operon predictions. More accurately identified operons will improve binding site predictions for transcription factors such as NtcA by better defining the regions upstream of operons where transcription factors bind.

The role of NtcA

NtcA was rapidly upregulated in response to N deprivation in both strains (Figure 4A) and clustered with genes known to be NtcA targets in other cyanobacteria (MED4 cluster 2 and MIT9313 cluster 1). We assessed the likely role of *Prochlorococcus* NtcA as a global transcriptional regulator during N starvation by comparing gene expression patterns with NtcA binding site predictions from Su *et al* (2005). High-ranking NtcA sites are abundant among the initially upregulated genes (t=6 h) in both strains (Table I). Because the algorithm to rank NtcA sites gives bonus points for conserved sites, it is useful to also rank sites separately for *Prochlorococcus* genes lacking orthologs in other cyanobacteria (Table I). For example, MED4 *hli10* had an overall NtcA rank score of 553 among 1087 genes. However, when the 258 MED4 genes lacking orthologs were

Table I *Prochlorococcus* MED4 (**A**) and MIT9313 (**B**) genes upregulated ($q<0.01$) at t=6 h and putative NtcA targets predicted by Su *et al* (2005)

(A) MED4 Gene name	Function	Fold change	NtcA site	NtcA site position	NtcA site ranking	
					All genes (1087)	Genes lacking orthologs (258)
PMM0958	Unknown	64.45	GTTgctttttaTAC	−53	1	
PMM0370	cynA	52.71	GTAtttgttaaTAC	−55	7	
PMM1462	Unknown	48.84	GTAtcaatgatAAC	−52	27	
PMM0970	**urtA**	**41.36**	GTTacctatgcTAC	**−48**	**4**	
PMM0920	**glnA**	**33.36**	GTTacctttgaTAC	**−52**	**2**	
PMM0246	**ntcA**	**21.86**	GTTactgttgaTAC	**−50**	**3**	
PMM0687	Unknown	14.42	GTAcatatctgTAC	−455	633	19
PMM0374	Unknown	13.36	GTTactttgatTAC	−50	601	8
PMM1041	Unknown	12.04	GTCtttacaatTAT	−62	940	177
PMM0336	Unknown	11.08	GAAtaaataaaCAC	−533	547	
PMM1463	glnB	7.26	see PMM1462	−52	27	
PMM1390	**hil10**	**6.15**	GTAggaaacacTAC	**−55**	**553**	**3**
PMM1037	Unknown	6.02	GTAtttaaaacTAC	−76	617	10
PMM0359	Unknown	5.91	GTCacgagataTAC	−463	667	31
PMM0337	Unknown	5.43	GTAtcatcaaaGAC	−136	813	109
PMM0971	**urtb**	**4.86**	GGCttaatgccTCC	**−39**	**518**	
PMM1074	zwf	4.76	GTAgctaattaGTC	−41	370	
PMM0365	Unknown	4.47	GTGgtaagggtTTC	−67	736	64

(B) MIT9313 Gene name	Function	Fold change	NtcA site	Position of NtcA site	NtcA site ranking	
					All genes (1563)	Genes lacking orthologs (431)
PMT0990	hli5	23.11	GTAacggtaaaTAC	−130	978	36
PMT2239	nirA	20.11	GTTcaatctgaTAC	−470	5	
PMT2240	NO_2 trans.	11.88	GCAacagcggtCAC	−144	434	
PMT0991	Unknown	10.13	GTCttgattctTTC	−227	1112	118
PMT2229	**urtA**	**9.45**	GTAtcattcacTAC	**−80**	**1**	
PMT0992	**hli7**	**8.75**	GTGaggttttaTAC	**−58**	**831**	**3**
PMT1831	**ntcA**	**8.17**	GTCaccattgcTAC	**−123**	**3**	
PMT2241	Unknown	5.78	see PMT2240	−144	434	
PMT0939	Unknown	5.58	GTAataatcaaAAC	−657	1056	73
PMT1853	**amt1**	**4.56**	GTAacaaaatgAAC	**−147**	**28**	
PMT0601	**glnA**	**3.29**	GTAcctgttgcTAC	**−66**	**2**	
PMT2228	**urtB**	**3.03**	GTTtgggaggaGTC	**−50**	**796**	
PMT1203	csoS2	2.82	GTTtcggtgtcTAT	−93	606	
PMT0951	Unknown	2.81	GTTacttgggtGAC	−34	31	
PMT1144	Unknown	2.27	GAAgagcccccTAC	−70	1361	265
PMT2234	ureA	2.08	GTGagcaccggAAC	−118	223	
PMT0631	Unknown	2.07	GTGagcaccggAAC	−496	1162	135
PMT0908	Unknown	1.93	GTTaggtctacTGC	−25	944	

Although a total of 81 MED4 genes were upregulated at t=6 h, only those greater than four-fold increased (18 in total) are compared to all MIT9313 upregulated genes (18 in total). Genes in bold have orthologs in the other strain that were also upregulated at t=6 h. The sequence of the NtcA sites and their positions relative to the translation start points are shown. The NtcA sites were predicted only for the first gene in each operon and thus ranks are relative to a total of 1087 MED4 and 1563 MIT9313 sites. Genes lacking orthologs (a total of 258 MED4 genes and 431 MIT9313 genes) were evaluated independently because NtcA rank scores are biased against these genes (see Materials and methods).

Table II *P*-values generated by GSEA (see Materials and methods; Subramanian *et al*, 2005), designed to test whether genes upregulated under N starvation are enriched for putative NtcA targets

	Time (h)					
Strain	0	3	6	12	24	48
MED4	0.2240	**0.0002***	**0.0002***	**0.0002***	**0.0002***	**0.0002***
MIT9313	0.1852	0.1040	**0.0016***	0.1194	0.0780	0.5940

The analysis was carried out using *Prochlorococcus* genes with the top 15 NtcA rank scores, all of which had orthologs in other cyanobacteria, plus the genes with top 15 rank scores among genes lacking orthologs in other cyanobacteria. This set was compared to all genes ordered by significant upregulation under N starvation. Asterisks/bold *P*-values indicate statistically significant enrichment as judged by $P<0.05/30$ (see Materials and methods and GSEA section of Supplementary information).

considered separately, *hli10* has the third highest score. Examination of the putative NtcA sites upstream of the initially upregulated genes suggests that *Prochlorococcus* NtcA targets include genes for N transport (*amt1*, *urtAB*, *cynA*), N assimilation (*glnA*, *ureA*, *nirA*), *hli* genes (MED4 *hli10* and MIT9313 *hli5*, *hli7*), and genes of unknown function.

To systematically test the extent to which genes upregulated under N starvation are regulated by NtcA, we applied gene set enrichment analysis (GSEA; Subramanian *et al*, 2005) to all genes at all time points. The GSEA results support that genes with predicted NtcA sites are significantly enriched among upregulated MED4 genes at all time points after the onset of N deprivation ($t=0$ h) and among upregulated MIT9313 genes at $t=6$ h only (Table II). We also explored whether sequence motifs similar to the NtcA binding site could be reconstructed by unsupervised motif searches using AlignACE (Roth *et al*, 1998; Hughes *et al*, 2000) from upstream sequences of upregulated clusters 1 and 2 for each of MED4 and MIT9313. For MED4, the motif with the top MAP and Group Specificity Score (Hughes *et al*, 2000) was highly similar to the NtcA binding site (Motif section of Supplementary information). In contrast, no NtcA-like motifs were found for MIT9313. MIT9313 upregulated genes did, however, have motifs with high MAP and specificity scores, some of which might represent binding sites for other activators (Motif section of Supplementary information).

In addition to activating transcription, NtcA has been shown to repress the transcription of genes in other cyanobacteria by binding near their transcriptional start sites (Ramasubramanian *et al*, 1994; Jiang *et al*, 1997). GSEA comparing the NtcA binding sites to the repressed genes produced no statistical evidence that NtcA represses transcription during N starvation in *Prochlorococcus*. Further, AlignACE motif discovery searches using the upstream sequences of MED4 repressed clusters found no motifs similar to known NtcA binding sites. We did, however, find other significant motifs that appear to be specific to the repressed clusters (Motif section of Supplementary information).

Our results correlating expression patterns during N starvation with binding site predictions support that NtcA influenced the upregulation of MED4 genes throughout the N starvation time course, whereas NtcA only influenced the initial upregulation of genes in MIT9313. This interstrain difference suggests one or both of the following scenarios: NtcA plays a lesser role in coordinating the MIT9313 N stress response, or MIT9313 NtcA binding site predictions are less accurate than those of MED4. Whereas the NtcA helix–turn–helix motif that binds DNA is the same in MED4 as in other cyanobacteria, the MIT9313 motif has a serine-for-alanine substitution (Su *et al*, 2005). If this amino-acid substitution altered the DNA binding specificity of NtcA in MIT9313, binding site predictions based on homology with other cyanobacteria may be less accurate in this strain. However, as we could not reconstruct a motif similar to known NtcA binding sites for the upregulated MIT9313 clusters, this amino acid change would need to have pervasively changed the binding specificity of NtcA in order for it to have a broad role in regulating these genes. These results suggest that future experiments such as *in vitro* selection of oligonucleotides (Jiang *et al*, 2000) would be particularly useful to characterize the DNA binding specificities of NtcA in these strains. Further, because the MD4-9313 array contains upstream sequences, microarray methods using chromatin immunoprecipitation (Lee *et al*, 2002) or phage display (Bulyk *et al*, 2001) could give a genome-wide picture of the MED4 and MIT9313 NtcA DNA binding specificities.

Sigma factors

Although NtcA clearly plays an important role in the N stress response in both strains, the lack of NtcA binding sites upstream of many of the differentially expressed genes suggests that other regulators are also involved. In other cyanobacteria, specific sigma factors are induced in response to N starvation (Brahamsha and Haselkorn, 1992; Caslake *et al*, 1997) and are required for long-term survival during N stress (Muro-Pastor *et al*, 2005). We found that two out of five MED4 sigma factors (PMM1289 and PMM1697) and two out of seven MIT9313 sigma factors (PMT2246 and PMT0346) were upregulated (Figure 4B), and may therefore play a role in the upregulation of gene expression during N stress in *Prochlorococcus*. Interestingly, one of the upregulated sigma factors in MIT9313 (PMT2246) has a strong NtcA binding site, suggesting that NtcA may indirectly act upon additional genes by activating this sigma factor. In contrast to MIT9313, an additional MED4 sigma factor (PMM1629) was repressed (cluster 7), suggesting that some repression of MED4 gene expression may be mediated by the downregulation of this sigma factor.

glnB

The *glnB* gene encodes the P_{II} post-translational regulator that coordinates carbon and nitrogen metabolism. Expression patterns of *glnB* during N stress were strikingly different between MED4 and MIT9313. As in other cyanobacteria, MED4 *glnB* expression was highly elevated, whereas MIT9313 *glnB* expression was unaffected (Figure 4C). A role of P_{II} in other cyanobacteria is to control the nitrate/nitrite and the bicarbonate transporters (Hisbergues *et al*, 1999; Lee *et al*, 1999). Unexpectedly, MED4 upregulates *glnB* under N stress but lacks genes for nitrite/nitrate utilization, whereas MIT9313 utilizes nitrite but does not upregulate *glnB*. Evidently, the

upregulation of *glnB* in MED4 during N starvation is independent of nitrite/nitrate utilization and is perhaps related to maintaining the cellular C–N balance.

Interstrain expression differences in *glnB* were reflected in the putatively cotranscribed genes directly upstream of *glnB*, MED4 PMM1462 and MIT9313 PMT1480 (Figure 4C). MED4 PMM1462 was highly upregulated following N deprivation, whereas MIT9313 PMT1480 expression was unchanged. These genes are orthologs that both have high-scoring NtcA sites, but no other BLAST hits in the NR database. The *glnB* gene is an NtcA target in other cyanobacteria (Garcia-Dominguez and Florencio, 1997). As MED4 PMM1462 and *glnB* are highly upregulated during N stress and have an NtcA binding site, they are likely NtcA-regulated. The lack of *glnB* upregulation in MIT9313 during N stress casts doubt on whether *glnB* is an NtcA target in this strain, even though there is an NtcA binding site upstream of PMT1480. PMT1480 and *glnB* may either not be regulated by NtcA or have an additional regulatory mechanism that can negate NtcA activation.

N transport and assimilation

Many *Prochlorococcus* genes for the transport and assimilation of nitrogen were activated during N starvation. The transporters for ammonium (*amt1*) and urea (*urt* genes) were upregulated in both strains. Strain-specific N transporters such as the MED4 cyanate transporter (*cyn* genes) and a nitrite permease in MIT9313 (PMT2240) were also enhanced. In contrast, none of the putative oligopeptide transporters were upregulated in either strain. Once in the cell, alternative N sources are converted to ammonium before being assimilated. Genes for the conversion of alternative N sources to ammonium were upregulated, such as urease genes (*ure* genes) in both strains and nitrite reductase in MIT9313 (*nirA*). In contrast, the MED4 cyanate lyase (*cynS*) was not differentially expressed.

Ammonium is assimilated into amino acids via the GS–glutamate synthase (GOGAT) pathway. The *glnA* gene, encoding GS, was upregulated in both strains (Figure 4D), similar to other cyanobacteria. The activation of *glnA* was unexpected in light of previous findings that neither the abundance nor the activity of the *Prochlorococcus* GS protein changes during N stress (El Alaoui *et al*, 2001, 2003). The different results between these studies may be due to experimental conditions, inter-laboratory variation in strains, or post-transcriptional regulation of *glnA*. The glutamate synthase (GOGAT) gene (*glsF*) was not differentially expressed during N starvation, as found for other cyanobacteria (Herrero *et al*, 2001). The glutamate dehydrogenase gene (*gdhA*) provides an alternative route to GS–GOGAT for ammonium assimilation and is found in MIT9313 but not in MED4. The MIT9313 *gdhA* gene was not, however, differentially expressed.

hli genes

The *hli* genes are a family of cyanobacterial genes proposed to protect the photosystems by dissipating excess absorbed light energy (Havaux *et al*, 2003). *Prochlorococcus* has the greatest number of *hli* genes among cyanobacteria examined to date: MED4 has 22 *hli* genes and MIT9313 has nine. Three of the *hli* genes in each strain (MED4 *hli*10, *hli*21, *hli*22 and MIT9313 *hli*5, *hli*7, *hli1*) were among the most upregulated genes in the genome during N stress. MED4 *hli*10 was the first and most highly upregulated among the *hli* genes in this strain (Table IA and Figure 4E). In MIT9313, *hli5* and *hli*7 genes were by far the most upregulated of all genes in the genome (approximately 70-fold) (Figure 4E). The most highly upregulated *hli* genes in each strain also have the strongest NtcA sites (Table IB). We thus propose that these *hli* genes evolved and specialized as NtcA targets to ensure their rapid upregulation to protect the photosystems during N stress. The *hli* genes do not appear to be NtcA targets in all other cyanobacteria. For example, none of the four *hli* genes in *Synechocystis* PCC 6803 have high-ranking NtcA binding sites (Su *et al*, 2005), even though they are all elevated in response to N stress (He *et al*, 2001).

Novel N-responsive genes

Genes of unknown function are among the most highly upregulated genes during N starvation in *Prochlorococcus*. Many of these have conserved NtcA binding sequences, suggesting that they are NtcA targets with specific roles in the N stress response. For example, MED4 PMM0958 was the most upregulated gene at all time points and has the top-ranking NtcA binding site in the genome (Table IA). This gene encodes a small protein of 75 amino acids with no conserved domains, but it has orthologs in other *Prochlorococcus* strains such as MIT9312, MIT9211, SS120, and NATL2A, as well as marine *Synechococcus* WH8102. PMM0958 is not upregulated in response to phosphorus starvation or phage infection (A Martiny, M Coleman, and D Lindell, unpublished data). The high level of upregulation of this gene and its apparent specificity to N stress suggests that it may, along with *ntcA* (Lindell and Post, 2001), be a sensitive indicator of N limitation in field populations of marine cyanobacteria.

Other differentially expressed genes of unknown function are proximal in the genome to genes with known roles in N metabolism. PMT2241 in MIT9313, for example, is found downstream of *nirA* and the putative nitrite transporter and these three genes were coexpressed. PMT1479, the gene directly upstream of PMT1480 and *glnB* in MIT9313, was the most repressed gene in the genome under N starvation (Figure 4C). PMM0374 in MED4 is upstream of the cyanate transporter genes and has a strong NtcA binding site, suggesting that it may be involved in cyanate utilization even though it is divergently transcribed from *cynABD*. As the number of sequenced cyanobacterial genomes rises (currently, 17 complete and 30 in progress according to the NCBI genomes database), comparative genomic methods such as phylogenetic profiling (Pellegrini *et al*, 1999), protein fusion analysis (Marcotte *et al*, 1999), and systematic orthology resources such as COGS (Tatusov *et al*, 2003) will become increasingly informative to elucidate the function of these genes in *Prochlorococcus*.

Carbon metabolism genes

The expression of genes for the transport and fixation of carbon differed between MED4 and MIT9313 over the course

of N starvation. The *rbcLS* genes, encoding the large and small subunits of the carbon-fixing enzyme Rubisco, were highly repressed in MED4 (Figure 4F). Both genes are members of K-means cluster 9, which contains the most downregulated genes in the genome (Figure 3A). In contrast, the expression of the *rbcLS* genes in MIT9313 was unchanged until the final time point, when they were mildly repressed (Figure 4F). As such, MIT9313 *rbcLS* are members of K-means cluster 4 (Figure 3B). These interstrain differences were also reflected in other carbon metabolism genes such as the bicarbonate transporter, *sbtA*, as well as the *csoS12* genes encoding the carboxysome shell proteins. MED4 thus responds to reduced N availability by repressing the expression of carbon transport and fixation genes to a much greater degree than MIT9313, perhaps as a means to conserve energy during N stress.

Regulation of glycogen, a carbon storage molecule, during N stress appears to be different in *Prochlorococcus* than in freshwater cyanobacteria. Freshwater cyanobacteria accumulate glycogen as cellular inclusions during N starvation (Allen, 1984; Schwarz and Forchhammer, 2005). MED4 and MIT9313, in contrast, enhanced transcription of the glycogen phosphorylase for glycogen degradation (*glgP*) and MED4 also repressed genes for glycogen synthesis (*glgABC)*. Perhaps freshwater cyanobacteria respond to N starvation by storing C in preparation for a future influx of nitrogen. *Prochlorococcus*, which lives in a comparatively homogenous, ocean environment, responds to N stress by expending C reserves.

Why, however, would MED4 respond to N stress by liberating carbon stored as glycogen (suggesting an increased C demand) while simultaneously repressing the fixation of new carbon? Cyanobacteria have neither a complete glycolytic pathway nor the Entner–Doudorhoff pathway. They therefore use the oxidative pentose phosphate pathway (PPP) to derive pyruvate and, ultimately, 2-oxoglutarate, the carbon skeleton used for N assimilation. Two key genes involved in the PPP are upregulated during N stress. The *zwf* gene, whose product drives the first step in the PPP, was upregulated in MED4. The *tal* gene, encoding the transaldolase that rearranges the carbon skeletons in the PPP, was upregulated in both strains. Further, the *acnB* and *icd* genes, whose products catalyze the final steps in the synthesis of 2-oxoglutarate, are upregulated in MED4. We thus propose that MED4 and MIT9313 respond to N starvation by liberating carbon from glycogen. This carbon is subsequently funneled through the PPP towards the synthesis of 2-oxoglutarate, as a means to more efficiently assimilate intracellular N.

The *glnB* mutant in *Synechococcus* PCC 7942 accumulates glycogen (Forchhammer and Tandeau de Marsac, 1995), indicating that P_{II} influences the regulation of glycogen levels in this organism. Glycogen regulation, as well as other aspects of carbon metabolism, may be influenced by P_{II} in *Prochlorococcus*. The *glnB* gene was differentially expressed in MED4, but remained unchanged in MIT9313. Similarly, a number of genes that provide carbon skeletons for N assimilation (*sbtA*, *rbcLS*, *csoS1*, *glgABC*, *acnB*, *icd*) were rapidly and highly differentially expressed in MED4, but not in MIT9313. The mechanism of P_{II} regulation in *Prochlorococcus* remains unclear. However, the correlated differential expression of *glnB* and these C metabolism genes each strain suggests that *glnB* in *Prochlorococcus* may have a role in regulating the C–N balance during N stress that extends beyond bicarbonate and nitrate/nitrite transport to include others aspects of carbon metabolism.

Steady-state expression on alternative N sources

Global transcriptional responses

Ammonium is the most energetically favorable N source for microbial growth because alternative sources such as urea, nitrite, and cyanate must be converted to ammonium before being assimilated through the GS–GOGAT pathway (reviewed by Herrero *et al*, 2001). Because reduction of alternative N sources requires additional energy, the repression of genes for the utilization of alternative N sources in ammonium-replete conditions is common among cyanobacteria. Growth on nitrogen sources other than ammonium thus requires the activation of genes for the transport and assimilation of that N source. We identified genes that were differentially expressed ($q < 0.01$) in each strain during log-phase, N-replete growth on alternative N sources relative to growth on ammonium (Figure 5) (see Goldenspike section of Supplementary information for a list of all differentially expressed genes). Further, we compared the expression changes of a number of N-regulated genes on alternative N sources to changes during N starvation (Table III).

Twenty MED4 genes were differentially expressed (relative to ammonium) when grown on urea or cyanate, 11 of which were common to both conditions. MIT9313 displayed considerably more differentially expressed genes (69 genes) when grown on urea than when grown on nitrite (14 genes). However, a comparison of the 14 most differentially expressed MIT9313 genes on urea to the 14 genes changed on nitrite revealed seven genes in common. Both *Prochlorococcus* strains thus displayed considerable overlap in the gene expression when grown on each alternative nitrogen source.

N transporters

MIT9313 upregulated the transporters for urea and nitrite on both alternative N sources (Table III). MED4 upregulated both the urea and cyanate transporter on cyanate, but neither transporter on urea (Table III). *Prochlorococcus* transporters for alternative N sources are thus upregulated in an all-or-none manner, suggesting that the cell lacks the ability to specifically identify ambient N sources. As with other cyanobacteria, *Prochlorococcus* likely perceives alternative N sources as a reduction in the rate of N assimilation, perhaps via 2-oxoglutarate (Forchhammer, 1999; Tandeau de Marsac and Lee, 1999), and thus responds by activating the transport of all N sources simultaneously.

Unexpectedly, expression of the MED4 urea transporter and urease complex was not upregulated on urea relative to ammonium, even though these genes were upregulated during N deprivation and are clearly required for urea metabolism. Similar results showing lack of changes in urea medium relative to ammonium have been found in other cyanobacteria. *Anabaena* PCC 7120, for example, takes up urea at similar rates in urea- and ammonium-based media (Valladares *et al*, 2002). The activity of the *Prochlorococcus* PCC 9511 urease complex is similar in both ammonium- and urea-based media

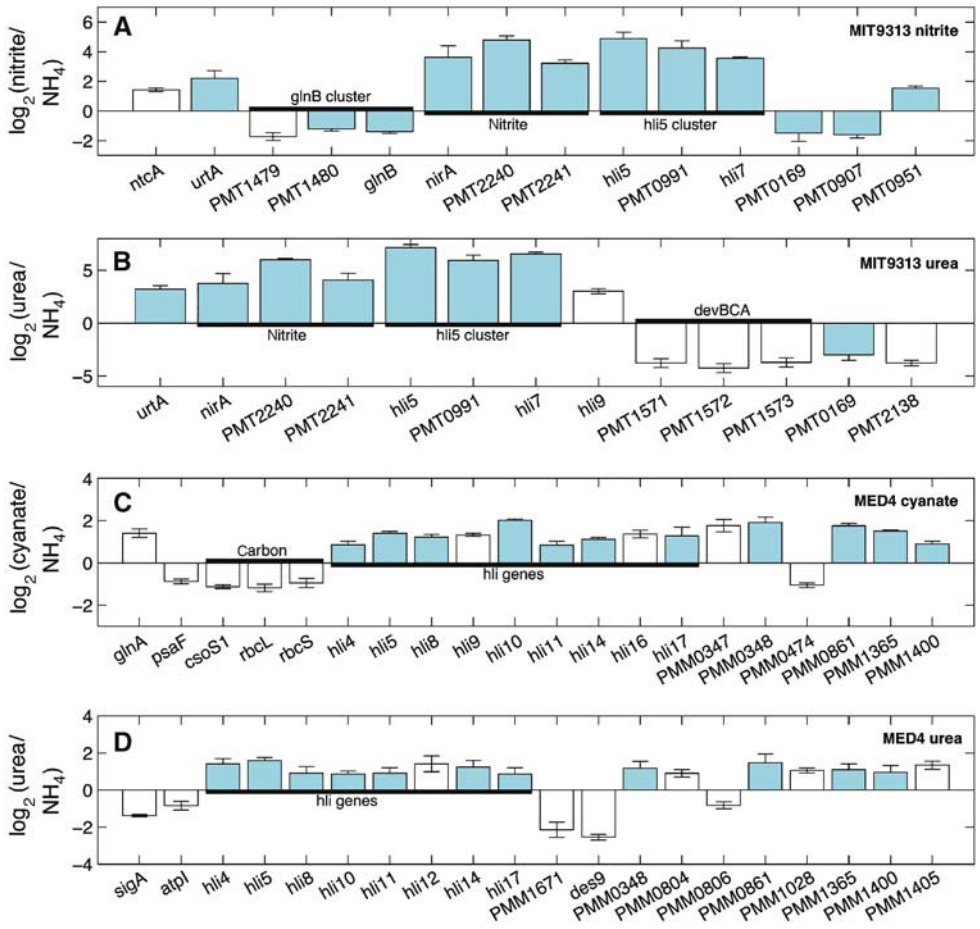

Figure 5 MIT9313 (A, B) and MED4 (C, D) differentially expressed genes on alternative N sources relative to ammonium. Plots show all genes differentially expressed ($q < 0.01$) for MIT9313 on nitrite (**A**), MIT9313 on urea (**B**), MED4 on cyanate (**C**), or MED4 on urea (**D**) relative to expression on ammonium. A total of 69 MIT9313 genes were differentially expressed on urea; only those >4-fold changed are shown. Data points show $\log_2$-transformed means of duplicate cultures; error bars show one standard deviation. Colored bars show genes that are differentially expressed on both N sources for a given strain.

(Palinska *et al*, 2000). It thus appears that the expression level of MED4 urea metabolism genes in ammonium medium is sufficient for growth on urea as well.

The role of NtcA

MIT9313 upregulated more putative NtcA targets on alternative N sources than did MED4. Even in MIT9313, many fewer putative NtcA targets were upregulated on alternative N sources than during N deprivation (Table III). Certain NtcA targets may require a high level of NtcA in the cell, which is reached under N starvation but not on alternative N sources (Muro-Pastor *et al*, 1996). NtcA is believed to monitor 2-oxoglutarate, which increases in concentration during N stress. The degree to which the response to an alternative N source is mediated by NtcA may thus be proportional to how much that N source represents an N stress relative to ammonium. It is also possible that NtcA targets that were upregulated during N deprivation were transiently activated when the cultures were first transferred to an alternative N source, but were not enhanced at steady-state growth relative to growth on ammonium. These hypotheses have yet to be tested.

Additional differentially expressed genes

Similar to N starvation, the *hli* genes were among the most highly induced genes in both strains on alternative N sources. Six MIT9313 *hli* genes were elevated in steady-state cultures grown on alternative N sources: *hli5* and *hli7* were highly upregulated on nitrite and urea and four additional *hli* genes were upregulated only on urea. MED4 upregulated 10 *hli* genes, seven of which were common to both N sources (Figure 5C and D). If *hli* genes indeed function to dissipate excess absorbed light energy during stress (Havaux *et al*, 2003), then these genes might be upregulated on alternative N sources to protect the photosystems, thus allowing the cell to sustain growth rates similar to those on ammonium.

Different ambient N sources resulted in contrasting expression patterns of the carbon fixation (*rbcLS*) genes, suggesting that alternative N sources affect the C–N balance differently. Expression of MIT9313 *rbc* genes was, for example, elevated in cells grown on urea but repressed in cells grown on nitrite (Table III). MED4 *rbcLS* were significantly repressed in cells grown on cyanate (Figure 5C) but not on urea. Several genes of unknown function were differentially expressed on both alternative N sources. For example, MIT9313 PMT1479 and

Table III Comparison of gene expression of selected N metabolism genes during N starvation and during N-replete growth on different N sources

		MED4				MIT9313			
Function	Gene	N stress	K-means cluster	Urea	Cyanate	N stress	K-means cluster	Urea	Nitrite
Regulation									
N regulator	ntcA	21.86** (6 h)	2	NS	NS	11.71** (12 h)	1	NS	2.69**
C–N balance	glnB	7.26** (6 h)	2	NS	NS	NS	4	−3.07**	−2.62**
glnB cluster	PMM1462	48.84** (6 h)	1	NS	NS	–	–	–	–
	PMT1480	–	–	–	–	NS	4	−2.32**	−2.31**
	PMT1479	–	–	–	–	−46.21** (48 h)	7	−2.07*	−3.32**
Sigma factors	PMM1289	2.85** (6 h)	5	NS	NS	–	–	–	–
	PMM1697	2.84** (12 h)	5	NS	NS	–	–	–	–
	PMM1629	−6.73** (24 h)	7	NS	NS	–	–	–	–
	PMT0346	–	–	–	–	4.20** (24 h)	2	NS	NS
	PMT2246	–	–	–	–	3.01** (24 h)	2	NS	NS
Transport									
Ammonium	amt1	2.12** (6 h)	5	NS	NS	4.89** (12 h)	2	NS	NS
Urea	urtA	41.36** (6 h)	1	NS	1.86*	16.56** (12 h)	1	9.19**	4.59**
Cyanate	cynA	52.71** (6 h)	1	NS	2.25*	–	–	–	–
NO_2	PMT2240	–	–	–	–	19.43** (12 h)	1	63.56**	27.47**
NO_2 transport partner	PMT2241	–	–	–	–	9.58** (12 h)	2	16.80**	9.25**
Bicarbonate	sbtA	−2.51** (12 h)	6	NS	NS	−2.79** (48 h)	5	NS	NS
Nitrogen metabolism									
NO_2 reduction	nirA	–	–	–	–	31.56** (24 h)	1	13.55**	12.30**
Urease	ureA	3.74** (6 h)	4	NS	NS	2.08** (6 h)	3	NS	NS
Glutamine syn.	glnA	33.36** (6 h)	1	NS	2.62**	4.08** (12 h)	2	NS	NS
GOGAT	glsF	2.36* (12 h)	4	NS	NS	−3.29** (48 h)	5	NS	NS
Carbon metabolism									
G-6-P dehydrogenase	zwf	4.76** (6 h)	4	NS	NS	2.23* (24 h)	3	NS	NS
Transaldolase	tal	3.25** (6 h)	4	NS	NS	2.69** (24 h)	3	NS	NS
2-Oxoglutarate syn.	icd	2.46** (6 h)	5	NS	NS	−3.03** (48 h)	4	NS	NS
Anonitate hydratase	acnB	2.84** (12 h)	4	NS	NS	−1.94* (48 h)	4	NS	NS
Rubisco	rbcL	−19.83** (12 h)	9	NS	−2.26**	−2.91** (48 h)	4	2.48*	−2.27*
Rubisco	rbcS	−20.82** (12 h)	9	NS	−1.92**	−3.10** (48 h)	4	4.23**	NS
Carboxysome	csoS1	−10.56** (48 h)	8	NS	−2.17**	NS	None	NS	NS
Carboxysome	csoS2	−5.02** (12 h)	7	NS	NS	2.82** (6 h)	4	4.37**	NS
Glycogen syn.	glgA	−2.53** (12 h)	6	NS	NS	NS	None	NS	NS

All expression changes are relative to N-replete growth on NH_4. N stress columns show maximum fold change over the time course and the associated time point. Two asterisks (**) denote $q<0.01$, single asterisk (*) shows $0.01<q<0.05$, 'NS' signifies no significant differential expression ($q>0.05$) and '–' means the gene is not present in that strain.

PMT1480 were repressed similar to *glnB* on both nitrite and urea, suggesting that the function of all three of these genes may be related. MIT9313 repressed PMT0169 and PMT0907 and upregulated PMT0951 on both nitrite and urea (Figure 5); these changes also occurred during N starvation. Similarly, MED4 upregulated PMM0861, PMM1365, PMM0348, and PMM1400 on both cyanate and urea (Figure 5), and the latter two were also differentially expressed under N starvation.

Comparison with *Synechococcus*

The transcriptional response of *Prochlorococcus* to alternative N sources can be compared to that of marine *Synechococcus* WH8102, for which expression in nitrate- and ammonium-based media was examined using microarrays (Su *et al*, 2006). Because *Prochlorococcus* does not grow on nitrate, a direct comparison with this *Synechococcus* data is, however, not possible. A total of 338 *Synechococcus* genes were differentially expressed in nitrate, suggesting that the response to alternative N sources requires many more genes in *Synechococcus* than in *Prochlorococcus*. Similar to *Prochlorococcus* during N deprivation, binding site predictions support that NtcA has a governing role in upregulating *Synechococcus* genes in nitrate medium. NtcA binding site predictions also have significant predictive capacity with respect to repression of *Synechococcus* genes, unlike our findings for *Prochlorococcus*. Additional expression changes were similar between *Synechococcus* and *Prochlorococcus*, such as upregulation of two *Synechococcus hli* genes and two sigma factors. Of particular interest, *Synechococcus ntcA* was upregulated whereas *glnB* was repressed. Opposing expression of these two nitrogen regulators is thus not specific to *Prochlorococcus* MIT9313 and may be a more general regulatory phenomenon in marine cyanobacteria.

Inter-ecotype comparison of N regulation

This study is a portrait of N-regulated gene expression in two *Prochlorococcus* strains, providing systems level insight into their transcriptional regulatory mechanisms through the integration of expression data, genome sequences, and comparative genomics (Su *et al*, 2005). We synthesize our findings into a model comparing the N stress response of

MED4 and MIT9313 (Figure 6). This model summarizes the transcriptional changes of N-responsive genes described in the previous sections along with the proposed interactions of the proteins encoded by these genes. The expression changes of these genes during N stress and on alternative N sources are shown in Table III.

The transcriptional regulator NtcA plays a significant role in the N stress response in both strains (Table II) and many of its putative targets are similarly upregulated in both strains (Figure 6 and Table III). These genes include N transporters (*amt1*, *urt*, *cyn* genes in MED4, and the nitrite permease in MIT9313), genes for the reduction of alternative N sources to ammonium (urease genes and the nitrite reductase in MIT9313), and *glnA* for the assimilation of ammonium into amino acids. NtcA also likely influenced the activation of a subset of the *hli* genes, which were highly upregulated and contain NtcA binding sites. As NtcA has a global role in upregulating genes during N stress in both strains (Table II), there are likely many additional genes activated by NtcA as well. Although studies in other cyanobacteria have shown that NtcA can also act as a repressor (Ramasubramanian *et al*, 1994; Jiang *et al*, 1997), we have no evidence that *Prochlorococcus* NtcA represses gene expression on a widespread basis. We thus propose that one or more yet unidentified repressors control the downregulation of ribosomal and photosynthetic genes observed during N starvation. In addition to NtcA, sigma factors likely influence widespread transcriptional changes during N stress. Two sigma factors were upregulated in each strain (Figure 4B) and an additional sigma factor was repressed in MED4. At least in MIT9313, one of the sigma factors activated during N stress appears to be an NtcA target, suggesting that these two regulators act in concert.

Expression of the *glnB* gene, encoding the regulator P_{II}, differed between the strains. MED4 *glnB* was upregulated in response to N starvation (typical for cyanobacteria), but MIT9313 *glnB* expression was unchanged (Figures 4C and 6). One of the major roles of the P_{II} protein is regulating the balance between C and N metabolism. This interstrain difference in *glnB* expression was mirrored in a number of genes linking N and C metabolism that were rapidly and highly differentially expressed in MED4 but not in MIT9313 (Figure 6 and Table III). For example, *rbcLS* genes for carbon fixation were highly repressed only in MED4 (Figure 4F). In addition, MED4 repressed genes for bicarbonate transport (*sbtA*), glycogen storage (*glgABC*), as well as upregulating genes for 2-oxoglutatate synthesis (*acnB*, *icd*). In MIT9313, the expression of all these genes remained unchanged until the final time point, when they were mildly repressed along with greater than 10% of the genome. The physiological measurements of F_v/F_m and chlorophyll fluorescence support that the cells were in an advanced state of starvation by the final time point. The downregulation of these MIT9313 carbon metabolism genes may have been part of a general transcriptional shutdown, and not specifically part of the N stress response. As such, their repression is not reflected in Figure 6.

An interstrain comparison of *Prochlorococcus* N regulation helps to elucidate the mechanisms underlying niche differentiation between ecotypes with different depth distributions in the ocean. Overall, the MED4 transcriptional response to N deprivation was rapid and transient, whereas the MIT9313 response was slower and sustained (Figure 2). MED4 belongs to one of the high-light-adapted *Prochlorococcus* ecotypes that dominate the surface mixed layer in the oceans where regenerated ammonium and urea are the predominant N sources. Because these N sources are patchily distributed (Valera and Harrison, 1999), MED4 likely experiences significant local fluctuations in N availability in the surface waters, which is consistent with its rapid and transient response to N starvation. This rapid response would also serve to protect its photosystems from rapid photodamage in higher light under N limited conditions. MIT9313 is a low-light-adapted strain that is most abundant in deeper waters with lower photon fluxes and higher nutrient levels. The slow and sustained transcriptional response observed in MIT9313 is consistent with this relatively more static and higher nutrient environment.

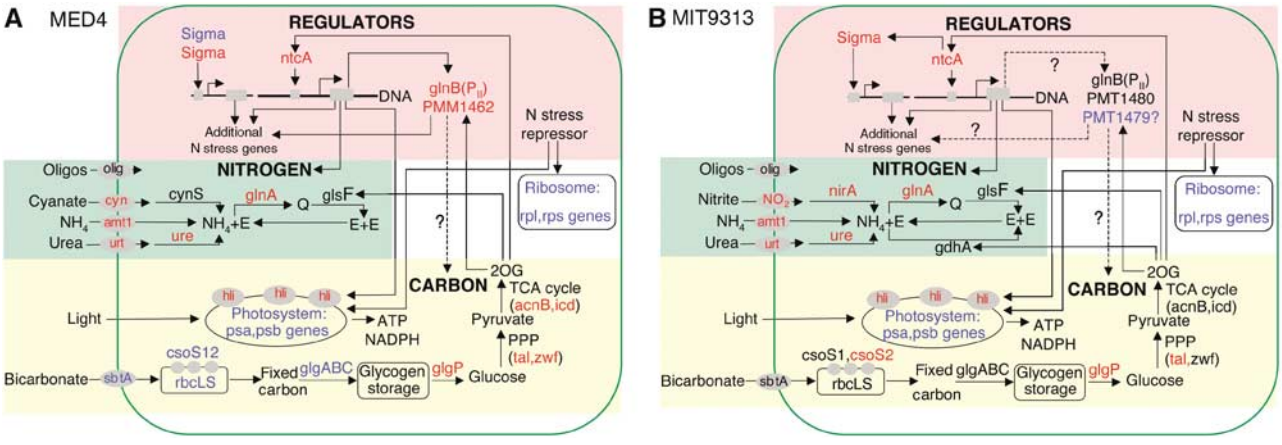

Figure 6 Model of the MED4 (**A**) and MIT9313 (**B**) transcriptional response to N starvation focusing on the regulation of nitrogen metabolism and its integration with carbon metabolism. Shaded regions show general cell processes: regulators (pink), nitrogen transport/assimilation (green), carbon metabolism (yellow). Text for genes upregulated in response to N starvation are in red, repressed genes are in blue, and unchanged genes are in black. Genes unchanged in expression except at the final time point (t=48 h) are not labeled as differentially expressed. Arrows show proposed interactions between the proteins encoded by these genes. Amino-acid abbreviations: glutamate=E, glutamine=Q.

Differences in the coordination of N and C metabolism may also reflect general physiological and, ultimately, ecological differentiation. The lack of repression of MIT9313 bicarbonate transport and carbon fixation genes may indicate that this strain has a higher C requirement than MED4 under the conditions tested and thus could not afford to repress C metabolism. Alternatively, if MED4 naturally experiences transient N depletion in the surface waters, it may have evolved more efficient mechanisms to shutdown other metabolic activities during N stress. In contrast, MIT9313, a strain adapted to deeper water conditions where photons are limiting, may continue inorganic C fixation in the face of N stress to maintain sufficient energy production.

The *Prochlorococcus* community is composed of many related strains (Rocap *et al*, 2002), which are proposed to niche partition the water column with respect to nitrogen (Moore *et al*, 2002). This study supports that niche partitioning between surface-water- and deep-water-adapted *Prochlorococcus* strains occurs, at least in part, through differences such as the dynamics of the transcriptional response to N stress and the maintenance of the C–N balance.

Materials and methods

Cell culture

Prochlorococcus cultures were grown at 22°C with a continuous photon flux of either 10 μmol quanta m^{-2} s^{-1} (MIT9313) or 50 μmol quanta m^{-2} s^{-1} (MED4) from cool white, fluorescent bulbs. Cultures were grown in Pro99 medium (Moore *et al*, 2002) supplemented to a final concentration of 1 mM Hepes pH 7.5 and 6 mM sodium bicarbonate. Total nitrogen in standard Pro99 medium was 800 μM ammonium. Although this nitrogen concentration is significantly higher than in the oligotrophic ocean, it was necessary to obtain sufficient biomass for microarray analysis. Further, *Prochlorococcus* cultures rapidly deplete N concentrations to below 100 nM during N stress (Lindell *et al*, 2002). All experiments were carried out using duplicate cultures.

To examine the MED4 and MIT9313 cellular response to nitrogen deprivation, 2 l cultures were grown through three successive transfers to establish that the growth rate was constant. They were then concentrated in mid-log growth by centrifugation (15 min, 9000 *g*, 22°C), washed once, and resuspended in Pro99 (' + NH4' treatment) or Pro99 medium lacking any supplemented nitrogen ('−N' treatment). Samples were taken at 0, 3, 6, 12, 24, and 48 h for chlorophyll fluorescence measurements, analysis of photosystem II photochemical conversion efficiency (F_v/F_m), and isolation of RNA. Culture fluorescence, a proxy for biomass, was measured using a Turner fluorometer (450 nm excitation; 680 nm absorbance). F_v/F_m was quantified using a Background Irradiance Gradient-Single Turnover fluorometer (Johnson, 2004). For this measurement, cells were dark acclimated for 15 min before single turnover fluorescence induction curves were measured. F_v/F_m was calculated by fitting standard models to the data to determine values of F_o (initial fluorescence), F_m (maximal fluorescence), and F_v ($F_m - F_o$) (Kolber *et al*, 1998).

To characterize mRNA expression in cultures grown on different N sources, log-phase cultures of MED4 and/or MIT9313 were established in Pro99 medium containing one of the following nitrogen sources: 800 μM ammonium (MED4 and MIT9313), 400 μM urea (MED4 and MIT9313), 200 μM nitrite (MIT9313), or 800 μM cyanate (MED4). Urea was added at 400 μM because it has two nitrogen atoms per molecule. Nitrite was added at 200 μM because higher concentrations were found to be toxic to MIT9313 (data not shown). MIT9313 growth rates were 0.23 day^{-1} on ammonium, 0.21 day^{-1} on nitrite, and 0.22 day^{-1} on urea. MED4 growth rates were 0.58 day^{-1} on ammonium, 0.35 day^{-1} on cyanate, and 0.56 day^{-1} on urea. Note that the lack of symmetry in the experimental design results from the fact that MED4 cannot grow on nitrite (Moore *et al*, 2002) and MIT9313 cannot grow on cyanate (data not shown).

RNA preparation

Cells were collected by centrifugation (15 min, 9000 *g*, 22°C), resuspended in 1 ml of RNA storage buffer (200 mM sucrose, 10 mM sodium acetate pH 5.2, 5 mM EDTA), frozen in liquid nitrogen, and stored at −80°C. RNA was isolated using the mirvana miRNA isolation kit (Ambion Inc., Austin, TX, USA) according to the manufacturer's instructions. Prior to RNA isolation, MIT9313 cells required an initial 60 min lysozyme (1 mg ml^{-1}) incubation at 37°C. DNA was removed using Turbo DNase treatment (Ambion Inc.) according to the manufacturer's instructions and DNA removal was confirmed by gel electrophoresis. RNA was then ethanol precipitated and resuspended at a concentration of 100 ng μl^{-1}.

Microarray labeling, hybridization, and scanning

Two micrograms of total RNA was incubated for 10 min at 70°C and annealed to random hexamer primers (25 ng μl^{-1}) for 10 min at 25°C. The RNA was reverse transcribed to produce complementary DNA (cDNA) by successive incubations for 10 min at 25°C, 60 min at 37°C, and 60 min at 42°C using 25 U μl^{-1} Superscript II (Invitrogen Life Technologies) with 0.5 mM dNTPs and 1 U μl^{-1} RNase Out RNase Inhibitor (Invitrogen). Superscript II was inactivated by a 10 min incubation at 70°C and RNA was removed by incubating the reaction mix for 30 min at 65°C with 0.25 N NaOH. The cDNA product was purified with MinElute PCR purification columns (Qiagen). The full yield of cDNA (1.5–2 μg) was digested with DNase I (0.6 U μg^{-1} cDNA) for 10 min at 37°C to obtain 50–200 bp fragments. DNase I was inactivated by a 10 min incubation at 98°C. The fragmented cDNA was biotin end-labeled with the BioArray Terminal Labeling Kit (Enzo Biochem.) by a 60 min incubation at 37°C. The reaction was stopped by freezing at −20°C overnight. The quality of end-labeling was verified by gel-shift assays with NeutrAvidin (Pierce Chemicals) on 1% TBE agarose gels. Biotin-labeled cDNA (1–1.65 μg) was hybridized to the array at 45°C in the presence of 100 mM MES, 1 M NaCl, 20 mM EDTA, 0.01% Tween 20, 0.1 mg ml^{-1} herring sperm DNA, 0.5 mg ml^{-1} BSA, 7.8% DMSO, and 3 nM of Affymetrix hybridization B2 oligo control probe for 16 h, rotating at 60 r.p.m. Microarrays were hybridized for each of the duplicate N starvation and alternative N source cultures. Following hybridization to the array, washes and stains were conducted on a GeneChip Fluidics Station 450 (Affymetrix) following the ProkGE_WS2v3 Affymetrix protocol. Arrays were scanned with the GeneChip Scanner (Affymetrix) using factory settings with excitation set for 570 nm and a 2.5 μm resolution.

A custom-ordered MD4-9313 Affymetrix array was used in this study. The MD4-9313 array has 25-mer oligonucleotide probe sets for predicted open reading frames (ORFs) and intergenic regions in each of the MED4 and MIT9313 genomes, as well as for two phage that infect MED4. To the extent possible, MED4 and MIT9313 probes are spaced approximately every 80 bases within each ORF and every 45 bases in the intergenic regions, with probe spacing reduced for short ORFs and intergenic regions. Only ORF probe set data were analyzed in this study. A detailed definition of MD4-9313 is available on ArrayExpress (accession number A-AFFY-58).

Data analysis

Expression summaries for each gene were computed from the probe intensities in Affymetrix.CEL files by using the Goldenspike R package (Choe *et al*, 2005) freely available at http://www.elwood9.net/spike. Genes differentially expressed between treatments were identified using *q*-values computed by Goldenspike. The *q*-value represents the false discovery rate of differentially expressed genes as the fraction of false positives in a group of genes exceeding a statistical cutoff. Because of microarray hybridization problems with the MIT9313 control ($+NH_4$) samples at t=24 h, the −N expression values at this time point were compared to the $+NH_4$ at t=12 h. Similarly, a MIT9313 $+NH_4$ array at t=3 h was substituted for one of the $+NH_4$

arrays at $t=0$ h in the analyses. Because the gene expression correlations between $+NH_4$ time points were as high as the correlations between replicates at a single time point, these substitutions should have had limited impact on our results. Raw and Goldenspike-normalized expression data for all experiments are available on ArrayExpress (accession=E-TABM-91). Our discussion in the text and most of our analyses define differential expression as $q<0.01$. However, we used the more relaxed cutoff $q<0.05$ for the clustering analysis to increase the number of genes being clustered.

The $\log_2$-transformed $-N/+NH_4$ expression summaries were clustered using the K-means algorithm. Clustering was performed in MATLAB (The Mathworks) using the squared Euclidian distance metric with the 'replicates' option set to 500. Only genes that were differentially expressed ($q<0.05$) at one or more time point after $t=0$ h were clustered, representing a total of 410 MED4 and 559 MIT9313 genes (a list of the gene members of each cluster is in the K-means section of Supplementary information). The number of clusters (K) was determined by using the mutual information Z-score of Gibbons and Roth (2002) to identify the K between 2 and 20 yielding clusters with maximal intra-cluster enrichment for CyanoBase gene function categories (Nakamura *et al*, 1998), resulting in $K=9$ for MED4 and $K=7$ for MIT9313. Once K was determined, the CyanoBase category with the greatest enrichment in each cluster was identified using P-values based on standard hypergeometric statistics (K-means section of Supplementary information). We evaluated the P-values for each cluster using two significance thresholds, a 'permissive' threshold that corrects for multiple hypotheses within the chosen K clustering, and a 'stringent' threshold that adjusts the permissive threshold to account for possible bias induced by using category distribution information to select K from between 2 and 20 (K-means section of Supplementary information).

We tested genome-wide operon predictions in both strains by evaluating if genes within putative operons were more correlated in their expression patterns across the N starvation time series than genes in different operons. To maintain consistency with the NtcA analyses, operons were defined as groups of ORFs that are transcribed in the same direction and are separated by <45 bp (Su *et al*, 2005). We checked the accuracy of these operon predictions by computing the average Pearson correlation coefficients of the expression levels of the first gene in each putative operon with that of gene n ($n=2, 3, 4, 5$) in the same operon. The correlations of genes within operons were compared to 2500 average correlations of same-sized sets of randomly chosen pairs of genes. We checked average correlations of tandem genes not in the same predicted operons by similar means (Operon section of Supplementary information).

Positions, scores, and orthologous relationships of NtcA binding sites were obtained from Su *et al* (2005). A total of 1087 putative MED4 operons and 1563 putative MIT9313 operons were ranked for candidate NtcA sites, with the highest scores being the strongest predictions. The NtcA site scoring metric reduces false positives by giving bonus points for NtcA sites that appear to be conserved in multiple cyanobacteria, but this also effectively penalizes scores for sites in *Prochlorococcus*-specific genes. In some analyses, we thus evaluate putative NtcA sites separately for the genes with orthologs and for the 258 MED4 and 431 MIT9313 genes without orthologs using the published Su *et al* (2005) score ranks for the former and the relative score rank among genes without orthologs for the latter. In Results and Discussion, we refer to a gene as having putative NtcA binding site if its ranks score is in the top 5% among all sites predicted in the genome.

We evaluated the enrichment of genes with NtcA sites among all upregulated genes under N starvation using a version of GSEA (Subramanian *et al*, 2005). In place of the correlations used in Subramanian *et al* (2005), we used 'signed $1-P$ values' $s(1-P)$. The 'P' is the CyberT P-value (Baldi and Long, 2001) of the gene for the comparison of the $-N$ and $+NH_4$ treatment as generated by Goldenspike. The 's' is $+1$ when the mean expression level for the gene in the $-N$ treatment is greater than or equal to that of the $+NH_4$ treatment, and -1 otherwise. The GSEA P-values for enrichment were estimated by comparing the enrichment score (ES) generated for the actual data against those generated for the data in 5000 random re-assignments of NtcA sites to genes. The ES graphs for each analysis and the MATLAB script used to implement GSEA are available in the GSEA section of Supplementary information.

The gene set examined by GSEA for NtcA site enrichment consisted of the genes with the top 15 ranking NtcA binding sites that have orthologs in other cyanobacteria plus the top 15 ranking genes lacking orthologs. Owing to operons, this set comprised 41 MED4 genes and 49 MIT9313 genes. We also examined two other gene sets by GSEA, the top 20 genes with orthologs plus the top 20 without orthologs, and the top 15 overall ranking genes, all of which had orthologs (GSEA section of Supplementary information). P-values for enrichment are corrected for multiple hypotheses for the full set of 30 comparisons covering all three sets. We found no significant enrichment for the set composed of just the top 15 NtcA rank scores. In contrast, we found highly significant enrichment when genes with and without orthologs are considered separately. This is evidence that the Su *et al* (2005) NtcA site scorings are overly conservative for genes without orthologs with respect to *Prochlorococcus*.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We thank S Choe for discussions and technical assistance with Goldenspike, Z Su and Y Xu for sharing their results on the prediction of NtcA targets, and E Zinser and N Tandeau de Marsac for their insights into bacterial metabolism. We also acknowledge the four anonymous referees for their helpful comments. This work was supported by a DOE GtL Program Grant (to GC and SWC), NSF (to SWC), and a grant from the Gordon and Betty Moore Foundation (to SWC). ACT was also supported by an NSF Graduate Fellowship.

References

Allen MM (1984) Cyanobacterial cell inclusions. *Annu Rev Microbiol* **38:** 1–25

Baldi P, Long AD (2001) A Bayesian framework for the analysis of microarray expression data: regularized *t*-test and statistical inferences of gene changes. *Bioinformatics* **17:** 509–519

Bertilsson S, Berglund O, Karl DM, Chisholm SW (2003) Elemental composition of marine *Prochlorococcus* and *Synechococcus*: implications for the ecological stoichiometry of the sea. *Limnol Oceanogr* **48:** 1721–1731

Brahamsha B, Haselkorn R (1992) Identification of multiple RNA polymerase sigma factor homologs in the cyanobacterium *Anabaena sp*. strain PCC 7120: cloning, expression, and inactivation of the *sigB* and *sigC* genes. *J Bacteriol* **174:** 7273–7282

Bulyk ML, Huang X, Choo Y, Church GM (2001) Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. *Proc Natl Acad Sci USA* **98:** 7158–7163

Campbell L, Landry MR, Constantinou J, Nolla HA, Brown SL, Liu H, Caron DA (1998) Response of microbial community structure to environmental forcing in the Arabian Sea. *Deep-Sea Res II* **45:** 2301–2325

Caslake LF, Gruber TM, Bryant DA (1997) Expression of two alternative sigma factors of *Synechococcus sp*.strain PCC 7002 is modulated by carbon and nitrogen stress. *Microbiology* **143** (Part 12)**:** 3807–3818

Choe SE, Boutros M, Michelson AM, Church GM, Halfon MS (2005) Preferred analysis methods for Affymetrix GeneChips revealed by a wholly defined control dataset. *Genome Biol* **6:** R16

El Alaoui S, Diez J, Humanes L, Toribio F, Partensky F, Garcia-Fernandez JM (2001) *In vivo* regulation of glutamine synthetase activity in the marine chlorophyll *b*-containing cyanobacterium *Prochlorococcus sp* strain PCC 9511 (oxyphotobacteria). *Appl Environ Microbiol* **67:** 2202–2207

El Alaoui S, Diez J, Toribio F, Gomez-Baena G, Dufresne A, Garcia-Fernandez JM (2003) Glutamine synthetase from the marine

cyanobacteria *Prochlorococcus spp.*: characterization, phylogeny and response to nutrient limitation. *Environ Microbiol* **5:** 412–423

Forchhammer K (1999) The P_{II} protein in *Synechococcus* PCC 7942 senses and signals 2-oxoglutarate under ATP-replete conditions. In *The Phototrophic Prokaryotes*, Peschek GA, Löeffelhardt W, Schmetterer G (eds) pp 549–553. New York, NY: Kluwer Academic

Forchhammer K (2004) Global carbon/nitrogen control by P_{II} signal transduction in cyanobacteria: from signals to targets. *FEMS Microbiol Rev* **28:** 319–333

Forchhammer K, Hedler A (1997) Phosphoprotein P_{II} from cyanobacteria—analysis of functional conservation with the P_{II} signal-transduction protein from *Escherichia coli*. *Eur J Biochem* **244:** 869–875

Forchhammer K, Tandeau de Marsac N (1995) Functional analysis of the phosphoprotein P_{II} (*glnB* gene product) in the cyanobacterium *Synechococcus sp.* strain PCC 7942. *J Bacteriol* **177:** 2033–2040

Garcia-Dominguez M, Florencio FJ (1997) Nitrogen availability and electron transport control the expression of *glnB* gene (encoding P_{II} protein) in the cyanobacterium *Synechocystis sp.* PCC 6803. *Plant Mol Biol* **35:** 723–734

Garcia-Fernandez 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

Gibbons FD, Roth FP (2002) Judging the quality of gene expression-based clustering methods using gene annotation. *Genome Res* **12:** 1574–1581

Giordano M, Beardall J, Raven JA (2005) CO_2 concentrating mechanisms in algae: mechanisms, environmental modulation, and evolution. *Annu Rev Plant Biol* **56:** 99–131

Graziano LM, Geider RJ, Li WKW, Olaizola M (1996) Nitrogen limitation of North Atlantic phytoplankton: analysis of physiological condition in nutrient enrichment experiments. *Aquatic Microb Ecol* **11:** 53–64

Havaux M, Guedeney G, He Q, Grossman AR (2003) Elimination of high-light-inducible polypeptides related to eukaryotic chlorophyll *a*/*b*-binding proteins results in aberrant photoacclimation in *Synechocystis* PCC6803. *Biochim Biophys Acta* **1557:** 21–33

He Q, Dolganov N, Bjorkman O, Grossman AR (2001) The high light-inducible polypeptides in *Synechocystis* PCC6803. Expression and function in high light. *J Biol Chem* **276:** 306–314

Herrero A, Muro-Pastor AM, Flores E (2001) Nitrogen control in cyanobacteria. *J Bacteriol* **183:** 411–425

Hisbergues M, Jeanjean R, Joset F, Tandeau de Marsac N, Bedu S (1999) Protein P_{II} regulates both inorganic carbon and nitrate uptake and is modified by a redox signal in *Synechocystis* PCC 6803. *FEBS Lett* **463:** 216–220

Hughes JD, Estep PW, Tavazoie S, Church GM (2000) Computational identification of *cis*-regulatory elements associated with functionally coherent groups of genes in *Saccharomyces cerevisiae*. *J Mol Biol* **296:** 1205–1214

Jiang F, Mannervik B, Bergman B (1997) Evidence for redox regulation of the transcription factor NtcA, acting both as an activator and a repressor, in the cyanobacterium *Anabaena* PCC 7120. *Biochem J* **327** (Part 2)**:** 513–517

Jiang FY, Wisen S, Widersten M, Bergman B, Mannervik B (2000) Examination of the transcription factor NtcA-binding motif by *in vitro* selection of DNA sequences from a random library. *J Mol Biol* **301:** 783–793

Johnson ZI (2004) Description and application of the background irradiance gradient-single turnover fluorometer (BIG-STf). *Mar Ecol Prog Ser* **283:** 73–80

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

Kolber ZS, Prasil O, Falkowski PG (1998) Measurements of variable chlorophyll fluorescence using fast repetition rate techniques: defining methodology and experimental protocols. *Biochim Biophys Acta* **1367:** 88–106

Lee HM, Vazquez-Bermudez MF, de Marsac NT (1999) The global nitrogen regulator NtcA regulates transcription of the signal transducer P_{II} (*glnB*) and influences its phosphorylation level in response to nitrogen and carbon supplies in the cyanobacterium *Synechococcus sp.* strain PCC 7942. *J Bacteriol* **181:** 2697–2702

Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne JB, Volkert TL, Fraenkel E, Gifford DK, Young RA (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298:** 799–804

Li WKW (1994) Primary production of prochlorophytes, cyanobacteria, and eucaryotic ultraphytoplankton: measurements from flow cytometric sorting. *Limnol Oceanogr* **39:** 169–175

Lindell D, Erdner D, Marie D, Prasil O, Koblizek M, Le Gall F, Rippka R, Partensky F, Scanlan DJ, Post AF (2002) The nitrogen stress response of *Prochlorococcus* strain PCC 9511 (Oxyphotobacteria) involves contrasting regulation of *ntcA* and *amt1*. *J Phycol* **38:** 1113–1124

Lindell D, Post AF (2001) Ecological aspects of *ntcA* gene expression and its use as an indicator of the nitrogen status of marine *Synechococcus spp*. *Appl Environ Microbiol* **67:** 3340–3349

Lipschultz F (2001) A time-series assessment of the nitrogen cycle at BATS. *Deep-Sea Res II* **48:** 1897–1924

Luque I, Flores E, Herrero A (1994) Molecular mechanism for the operation of nitrogen control in cyanobacteria. *EMBO J* **13:** 2862–2869

Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D (1999) Detecting protein function and protein–protein interactions from genome sequences. *Science* **285:** 751–753

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

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

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

Muro-Pastor MI, Reyes JC, Florencio FJ (1996) The NADP + -isocitrate dehydrogenase gene (*icd*) is nitrogen regulated in cyanobacteria. *J Bacteriol* **178:** 4070–4076

Muro-Pastor MI, Reyes JC, Florencio FJ (2001) Cyanobacteria perceive nitrogen status by sensing intracellular 2-oxoglutarate levels. *J Biol Chem* **276:** 38320–38328

Muro-Pastor MI, Reyes JC, Florencio FJ (2005) Ammonium assimilation in cyanobacteria. *Photosynth Res* **83:** 135–150

Nakamura Y, Kaneko T, Hirosawa M, Miyajima N, Tabata S (1998) CyanoBase, a www database containing the complete nucleotide sequence of the genome of *Synechocystis sp.* strain PCC6803. *Nucleic Acids Res* **26:** 63–67

Ninfa AJ, Atkinson MR (2000) P_{II} signal transduction proteins. *Trends Microbiol* **8:** 172–179

Olson RJ (1981) ^{15}N tracer studies of the primary nitrite maximum. *J Mar Res* **39:** 1033–1051

Palinska KA, Jahns T, Rippka R, Tandeau De Marsac N (2000) *Prochlorococcus marinus* strain PCC 9511, a picoplanktonic cyanobacterium, synthesizes the smallest urease. *Microbiology* **146** (Part 12)**:** 3099–3107

Partensky F, Hess WR, Vaulot D (1999) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* **63:** 106–127

Paz-Yepes J, Flores E, Herrero A (2003) Transcriptional effects of the signal transduction protein P_{II} (*glnB* gene product) on NtcA-

dependent genes in *Synechococcus sp.* PCC 7942. *FEBS Lett* **543:** 42–46

Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci USA* **96:** 4285–4288

Ramasubramanian TS, Wei TF, Golden JW (1994) Two *Anabaena Sp* strain Pcc 7120 DNA-binding factors interact with vegetative cell-specific and heterocyst-specific genes. *J Bacteriol* **176:** 1214–1223

Redfield AC (1958) The biological control of chemical factors in the environment. *Am Sci* **46:** 205–221

Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA, Arellano A, Coleman M, Hauser L, Hess WR, Johnson ZI, Land M, Lindell D, Post AF, Regala W, Shah M, Shaw SL, Steglich C, Sullivan MB, Ting CS, Tolonen A, Webb EA, Zinser ER, Chisholm SW (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424:** 1042–1047

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

Roth FP, Hughes JD, Estep PW, Church GM (1998) Finding DNA regulatory motifs within unaligned non-coding sequences by whole-genome mRNA quantitation. *Nat Biotechnol* **16:** 939–945

Sauer J, Gorl M, Forchhammer K (1999) Nitrogen starvation in *Synechococcus* PCC 7942: involvement of glutamine synthetase and NtcA in phycobiliprotein degradation and survival. *Arch Microbiol* **172:** 247–255

Schwarz R, Forchhammer K (2005) Acclimation of unicellular cyanobacteria to macronutrient deficiency: emergence of a complex network of cellular responses. *Microbiology* **151** (Part 8)**:** 2503–2514

Su Z, Mao F, Dam P, Wu H, Olman V, Paulson IT, Palenik B, Xu Y (2006) Computational inference and experimental validation of the nitrogen assimilation regulatory network in cyanobacterium *Synechococcus sp.* WH 8102. *Nucleic Acids Res* **34:** 1050–1065

Su Z, Olman V, Mao F, Xu Y (2005) Comparative genomics analysis of NtcA regulons in cyanobacteria: regulation of nitrogen assimilation and its coupling to photosynthesis. *Nucleic Acids Res* **33:** 5156–5171

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102:** 15545–15550

Tandeau de Marsac N, Lee HM (1999) Regulation of carbon and nitrogen metabolism in the unicellular cyanobacterium *Synechococcus* spp. In *The Phototrophic Prokaryotes*, Peschek GA, Löeffelhardt W, Schmetterer G (eds) pp 539–548. New York, NY: Kluwer Academic

Tanigawa R, Shirokane M, Maeda Si S, Omata T, Tanaka K, Takahashi H (2002) Transcriptional activation of NtcA-dependent promoters of *Synechococcus sp.* PCC 7942 by 2-oxoglutarate *in vitro*. *Proc Natl Acad Sci USA* **99:** 4251–4255

Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinform* **4:** 41

Valera DE, Harrison PJ (1999) Seasonal variability in nitrogenous nutrition of phytoplankton assemblages in the northeastern subarctic Pacific Ocean. *Deep-Sea Res II* **46:** 2505–2538

Valladares A, Montesinos ML, Herrero A, Flores E (2002) An ABC-type, high-affinity urea permease identified in cyanobacteria. *Mol Microbiol* **43:** 703–715

Vazquez-Bermudez MF, Herrero A, Flores E (2002) 2-Oxoglutarate increases the binding affinity of the NtcA (nitrogen control) transcription factor for the *Synechococcus glnA* promoter. *FEBS Lett* **512:** 71–74

Vega-Palas MA, Madueño F, Herrero A, Flores E (1990) Identification and cloning of a regulatory gene for nitrogen assimilation in the cyanobacterium *synechococcus sp.* strain PCC 7942. *J Bacteriol* **172:** 643–647

West NJ, Schonhuber WA, Fuller NJ, Amann RI, Rippka R, Post AF, Scanlan DJ (2001) Closely related *Prochlorococcus* genotypes show remarkably different depth distributions in two oceanic regions as revealed by *in situ* hybridization using 16S rRNA-targeted oligonucleotides. *Microbiology* **147** (Part 7)**:** 1731–1744