PROCHLOROCOCCUS GENETIC TRANSFORMATION AND THE GENOMICS OF NITROGEN METABOLISM

by

Andrew Carl Tolonen

B.A. Biology
Dartmouth College, 1998

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

and the

WOODS HOLE OCEANOGRAPHIC INSTITUTION

August, 2005
[September 2005]

Signature of Author: _____
Joint Program in Biological Oceanography
Massachusetts Institute of Technology
and Woods Hole Oceanographic Institution

Certified by: _____
Sallie W. Chisholm
Professor of Civil and Environmental Engineering, and Biology
Massachusetts Institute of Technology
Thesis Advisor

Accepted by: _____
John Waterbury
Chair, Joint Committee for Biological Oceanography
Massachusetts Institute of Technology
and Woods Hole Oceanographic Institution

PROCHLOROCOCCUS GENETIC TRANSFORMATION AND THE GENOMICS OF NITROGEN METABOLISM

by

Andrew Carl Tolonen

Submitted to the Department of Biology, Massachusetts Institute of Technology and the Woods Hole Oceanographic Institution, August 2005 in partial fulfillment of the requirements for the degree of Doctor of Philosophy

ABSTRACT

Prochlorococcus, a unicellular cyanobacterium, is the most abundant phytoplankton in the oligotrophic, oceanic gyres where major plant nutrients such as nitrogen (N) and phosphorus (P) are at nanomolar concentrations. Nitrogen availability controls primary productivity in many of these regions. The cellular mechanisms that *Prochlorococcus* uses to acquire and metabolize nitrogen are thus central to its ecology. One of the goals of this thesis was to investigate how two *Prochlorococcus* strains responded on a physiological and genetic level to changes in ambient nitrogen. We characterized the N-starvation response of *Prochlorococcus* MED4 and MIT9313 by quantifying changes in global mRNA expression, chlorophyll fluorescence, and Fv/Fm along a time-series of increasing N starvation. In addition to efficiently scavenging ambient nitrogen, *Prochlorococcus* strains are hypothesized to niche-partition the water column by utilizing different N sources. We thus studied the global mRNA expression profiles of these two *Prochlorococcus* strains on different N sources.

The recent sequencing of a number of *Prochlorococcus* genomes has revealed that nearly half of *Prochlorococcus* genes are of unknown function. Genetic methods such as reporter gene assays and tagged mutagenesis are critical tools for unveiling the function of these genes. As the basis for such approaches, another goal of this thesis was to find conditions by which interspecific conjugation with *Escherichia coli* could be used to transfer plasmid DNA into *Prochlorococcus* MIT9313. Following conjugation, *E. coli* were removed from the *Prochlorococcus* cultures by infection with *E. coli* phage T7. We applied these methods to show that an RSF1010-derived plasmid will replicate in *Prochlorococcus MIT9313*. When this plasmid was modified to contain green fluorescent protein (GFP) we detected its expression in *Prochlorococcus* by Western blot and cellular fluorescence. Further, we applied these conjugation methods to show that Tn5 will transpose *in vivo* in *Prochlorococcus*. Collectively, these methods provide a means to experimentally alter the expression of genes in the *Prochlorococcus* cell.

ACKNOWLEDGEMENTS

"Do not worry. You have always written before and you will write now. All you have to do is write one true sentence. Write the truest sentence that you know". So finally I would write one true sentence, and then go on from there.

- Ernest Hemingway, "A Moveable Feast"

First and foremost, I would like to thank my thesis committee Penny Chisholm, George Church, Wolfgang Hess, Martin Polz, and Eric Webb. It was your diverse expertise and tireless enthusiasm that kept these experiments exciting.

On all the days when my experiments foundered, it was the fellowship of my co-workers in the Chisholm lab that kept me afloat. Zack Johnson: social chair, nerdy technical expert par-excellent. Erik Zinser: artist, fellow video projector afficionado, microbiology purist. Debbie Lindell: multi-faceted scientist, mother, and crazy disco danser in Roscoff. Adam Martiny: balancer of lab excellence with a being a connoisseur of the good life. Matt Sullivan: outdoorsman, defender of Ohio's glory, master of the phage. Maureen Coleman: Dartmouthian and rational advice giver. Luke Thompson: fellow reductionist, first MIT biologist to venture into the Chisholm lab.

When we could put lab behind us, it was my friends at MIT that made grad school fun. Greg Liszt: banjo bad-ass and all around first class act. Nick Bishop: adventurer/climber lost in translation. Mario Mikula: California lady-killer. Kimberly Hartwell: vestal scientist and ideological purist. Megan Higginsbotham: Southern Belle, lover of small dogs and big men. Anu Seshan: giggling Brahmin. Melissa Harrison: cycling/running machine and born-and-bred academic.

Charlotte Henson: If I discovered one thing at MIT, it was you.

A.C. Tolonen was funded by a Merck Graduate Fellowship in Computational Biology and an NSF Graduate Fellowship. This work was also funded by DOE Genomes to Life grant DE-FG02-02ER63445 and a grant from the Gordon and Betty Moore Foundation to S.W. Chisholm.

TABLE OF CONTENTS

INTRODUCTION

"How little we know is epitomized by bacteria of the genus *Prochlorococcus*, arguably the most abundant organisms on the planet and responsible for a large part of the organic production of the ocean--yet unknown to science until 1988. *Prochlorococcus* cells float passively in open water at 70,000 to 200,000 per milliliter, multiplying with energy captured by sunlight. They eluded recognition so long because of their extremely small size. Representing a special group called picoplankton, they are much smaller than conventional bacteria and barely visible at the highest optical magnification".

-E.O. Wilson, "The Future of Life" 2002

Prochlorococcus: an oxygenic phototroph of global ecological significance

Prochlorococcus was first observed just 20 years ago on a cruise from Barbados. A water sample was analyzed using flow cytometry which revealed a population of red-fluorescing particles (Olson, 1985). The first *Prochlorococcus* culture, called SARG, was isolated three years later by Brian Palenik from the base of the euphotic zone in the Sargasso Sea. *Prochlorococcus* has since been shown to be a unicellular, marine cyanobacterium approximately 0.5-0.8 microns in diameter. It is the smallest known photosynethetic organism (Partensky et al., 1999) and approaches the minimum predicted size for an oxygen evolving cell (Raven, 1994).

Prochlorococcus is distributed worldwide between 40° N and 40°S latitude and is likely the most abundant photosynthetic organism in the oceans (Partensky et al., 1999). A compilation of 8,400 flow cytometric field measurements showed that *Prochlorococcus* is abundant throughout the world's temperate ocean basins (Fig. 1). Measurements in the Arabian Sea quantified *Prochlorococcus* at densities up to 700,000 cells per milliliter of seawater (Campbell et al., 1998). *Prochlorococcus* is most abundant in oligotrophic central oceans, but it has also been found in coastal environments such as the outflow of the Rhone River in the Mediterranean Sea (Veldhuis et al., 1990) and the lagoons of a Pacific atoll (Charpy and Blanchot, 1996). In addition to growing in the oxygenated, euphotic zone, *Prochlorococcus* has been found to exploit a niche in the secondary chlorophyll maximum situated below the oxycline known as the oxygen minimum zone (OMZ) (Johnson et al., 1999). As a numerically dominant phototroph in many regions of the world's oceans, *Prochlorococcus* plays a critical role in the primary production of the oceans. Studies of photosynthetic rates estimate that the total phytoplankton production attributable to *Prochlorococcus* in many areas is between 11 and 57% (Li, 1994).

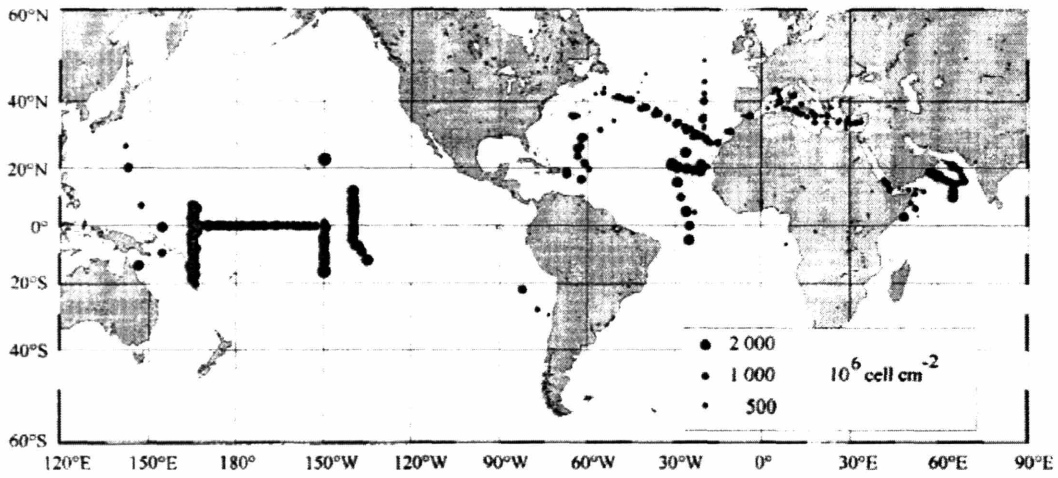


Fig. 1. *Prochlorococcus* cell concentrations integrated over the water column as measured by flow cytometry show that it is abundant in geographically diverse ocean basins. The diameter of the data points correlate to the abundance of *Prochlorococcus* (Partensky et al., 1999).

The vertical distribution of *Prochlorococcus* in the water column can extend from the surface to below the boundary of the euphotic zone. *Prochlorococcus* cells thus survive across a 10,000-fold variation in irradiance. This wide habitat range has been hypothesized to result from the coexistence of genetically and physiologically distinct populations adapted for growth at different light intensities. In fact, multiple *Prochlorococcus* strains with distinct light physiologies have been isolated from a single water sample (Moore et al., 1998). For example, the *Prochlorococcus* strains MIT9312 and MIT9313 were isolated from the same water sample in the Gulf Stream and differ remarkably in their growth rates as a function of light intensity (Fig. 2A). Similarly, the MIT9302 and MIT9303 strains came from the same Sargasso Sea sample but have different growth rates as a function of light intensity (Fig. 2B).

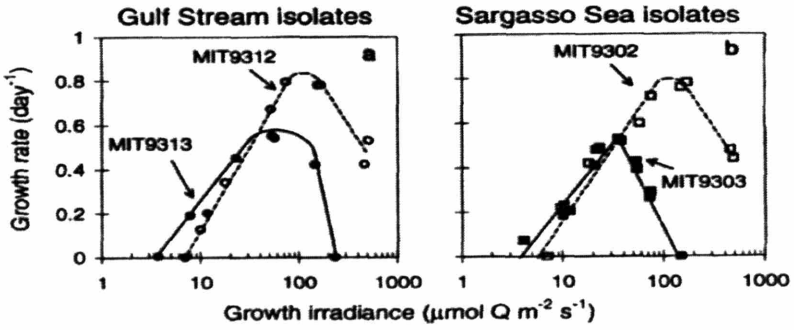


Fig. 2. Pairs of physiologically distinct *Prochlorococcus* strains were isolated from the same seawater sample. **A.** MIT9312 and MIT9313 are two isolates with different growth rates as a function of light intensity from the same Gulf Stream sample. **B.** MIT9302 and MIT9303 are two isolates with different growth rates as a function of light intensity from the same Sargasso Sea sample (Moore et al., 1998).

This co-occurrence of physiologically-distinct Prochlorococcus strains results in *Prochlorococcus* being able to exploit a wider niche than would be possible as a single strain.

Culture-based studies of *Prochlorococcus* light physiology have shown that *Prochlorococcus* isolates can be broadly be divided into two groups: high-light adapted strains (also called low chlorophyll B/A strains) and low-light adapted strains (also called high chlorophyll B/A strains). High-light adapted strains grow optimally at >100 micromoles photons $m^{-2} s^{-1}$ (Moore et al., 1995) and are most abundant in the surface waters (West et al., 2001). Low-light adapted strains grow best at 30-50 micromoles photons $m^{-2} s^{-1}$ (Moore et al., 1995) and are most abundant at greater depth (West et al., 2001). Molecular phylogenies based upon rDNA sequences correlate with groupings based on physiology (Fig. 3) (Urbach et al., 1998; Moore et al., 1998; Rocap et al., 2002). Because the DNA sequence phylogenies correspond to differences in physiology and distribution in the water column, the high-light adapted and low light adapted clades are referred to as "ecotypes".

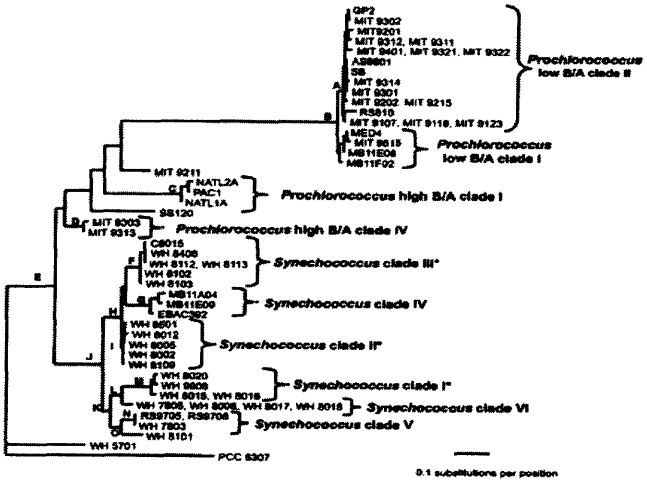


Fig.3. Phylogenetic relationship of *Prochlorococcus* strains as inferred by maximum likelihood using the 16S-23S rDNA spacer (Rocap et al., 2002). Low B/A strains are high-light adapted and high B/A strains are low light adapted.

Prochlorococcus ecological genomics

In addition to field and culture based studies, *Prochlorococcus* is emerging as a model system for ecological microbial genomics. To date, the complete genome sequences of three *Prochlorococcus* strains have been published (Rocap et al., 2003; Dufresne et al., 2003) and several more are currently being sequenced. The genomic diversity of *Prochlorococcus* is well illustrated by comparing the genomes of the high light-adapted MED4 and the low light-adapted MIT9313 which span the largest

evolutionary distance within the *Prochlorococcus* lineage (Rocap et al., 2003). *Prochlorococcus* MED4 has a smaller genome (1.66 Mb) consisting of 1,716 genes and is the smallest of any known oxygenic phototroph. MIT9313 has a relatively larger genome of 2.44 Mb with 2,275 genes. The two genomes have 1,350 genes in common and thus a significant fraction of the genes are ecotype-specific. These interstrain differences in genome content reveal differences in the ecological adaptation of the two strains (Rocap et al., 2003).

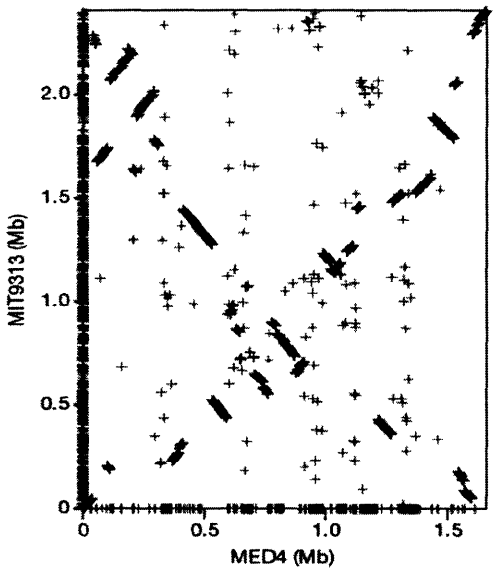


Fig. 4. Global genome alignment of MIT9313 and MED4 as seen from the amino acid start positions of orthologous genes. Genes present in one genome but not in the other are shown on the axes (Rocap et al., 2003). Contiguous blocks of conserved genes shown conserved operons.

Genome-wide alignments reveal the dynamic structure of *Prochlorococcus* genomes. Full genome nucleotide alignments comparing MED4 and MIT9313 genomes using the MUMmer program (Delcher et al., 1999) show that there are basically no large regions of conservation between the *Prochlorococcus* genomes. This may be largely be due to differences in GC content. MED4 is 31% GC while MIT9313 is 50.6% CG. Comparisons at the amino acid level are better able to identify regions of conservation between the *Prochlorococcus* genomes. The amino acid complement of the two *Prochlorococcus* genomes can be compared using BLASTp (Fig. 4). Amino acid aligments show that there are genomic regions where gene order is conserved between *Prochlorococcus* MED4 and MIT9313. These islands of conservation likely represent operons whose genes have been retained in order and function across evolutionary time. The borders of the orthlologous clusters are often flanked by transfer RNAs, suggesting that tRNAs genes serve as loci for

rearrangements.

By comparing *Prochlorococcus* photosynthetic genes with homologs in the NCBI database, one can find the genomic underpinnings for the differences in light-harvesting abilities of MED4 and MIT9313 (Hess et al., 2001). MED4 has many more genes encoding high-light inducible proteins and photolyases to repair UV damage, while MIT9313 has more genes associated with the photosynthetic apparatus. For example, MIT9313 has two genes for chlorophyll-binding proteins (*pcb* genes) and two genes for the Photosystem II reaction center protein (*psbA* gene), whereas MED4 has only one of each. MIT9313 may have evolved a more elaborate photosynthetic apparatus to enable it to efficiently harvest light at low intensities. rDNA phylogenies support that MED4 has evolved more recently than MIT9313 (Fig. 3). Genomic studies have also indicated that MED4 evolution resulted in a genome-wide winnowing of gene content. The *cpe* genes involved in phycoerythrin biosynthesis are an example of how this genomic reduction occurred. Comparing the *cpe* operons of the low light adapted strains, SS120 and MIT9313, to the high light adapted strain, MED4, shows a gradual loss of genes involved in phycoerythrin biosynthesis. For example, in both SS120 and MED4 the *cpe* genes are flanked by the unrelated genes *metK* and *uvrD*. In SS120 the *cpe* regions consists of 11.5 Kb containing 10 genes. MED4 has retained *cpeB*, the core gene involved in phycoerythrin biosynthesis. However, the *cpeB* region has been reduced to 4.5 Kb containing 7 genes. Moore et al. (2002) found similar gene loss in the *nirA* operon involved in nitrate reduction. These observations combined with the genome-wide blastP analyses (Fig. 4) support that MIT9313 and MED4 share a common genomic backbone and many conserved operons. However, the MED4 genome evolved by small-scale excision of non-essential genes.

<u>*Prochlorococcus* nitrogen metabolism</u>

Prochlorococcus dominates the phytoplankton community in the central ocean gyres where nutrients such as nitrogen (N) and phosphorus (P) are at nanomolar levels. The small size and resulting high surface area:volume ratio of the *Prochlorococcus* cell facilitates the uptake of ambient nutrients. However, survival in an oligotrophic environment likely requires additional adaptations such as low cellular nutrient requirements and higly efficient nutrient transport and assimilation systems. As such, the cellular mechanisms that *Prochlorococcus* uses to acquire and metabolize nitrogen are central to its ecology. **One of the goals of this thesis was to explore how two strains of *Prochlorococcus*, high light-adapted MED4 and low light-adapted MIT9313, respond genetically and physiologically to N**

starvation and different N sources. By comparing the nitrogen metabolism of MED4 and MIT9313, we hope to ultimately connect the cellular mechanism *Prochlorococcus* uses to respond to changes in ambient nitrogen to the environmental factors governing *Prochlorococcus* ecology. This section describes previous field and laboratory studies on the molecular biology of cyanobacterial N metabolism and how it relates to the *Prochlorococcus* ecology.

Cellular elemental stoichiometries relative to the ambient nutrient concentrations can elucidate the relationship of the *Prochlorococcus* cell to its environment. The C:N:P stoichiometry of *Prochlorococcus* MED4 have been characterized (Bertillsson et al., 2003). This study found that MED4 C:N:P cell quotas were 61:9.6:0.1 femtograms cell^{-1}, supporting that the small size of the *Prochlorococcus* cell manifests as low overall nutrient quotas. Interestingly, the C:N:P molar ratios of the cell differed significantly from 106C:16N:1P Redfield ratios classically believed to dictate the elemental composition of biomass in the sea (Redfield, 1958). Specifically, MED4 has elevated N requirements relative to phosphorus. *Prochlorococcus* quotas are >20N:1P (Bertilsson et al., 2003) and thus exceed the 16N:1P Redfield Ratio. If the nutrient ratios in the ambient seawater are 16N:1P and the MED4 cellular requirements are >20N:1P, then *Prochlorococcus* would have a propensity to become N limited relative to P. In support of this hypothesis, field studies have shown that nitrogen enrichment stimulated *Prochlorococcus* growth in the North Atlantic (Graziano et al., 1996) supporting that N availability can limit *Prochlorococcus* abundance.

Because of the important role nitrogen plays in the ecology of marine cyanobacteria, Lindell and Post (2001) developed a molecular assay of *ntcA* expression has been to monitor the N status of field populations (Fig. 5).

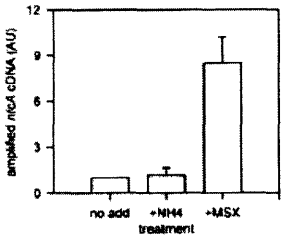


Fig. 5. An assay of *ntcA* expression in a *Synechococcus* population in the Red Sea shows that cells are not N stressed. 'no add' treatment show *ntcA* expression level of natural population. '+NH4' treatment shows ammonium addition did not decrease *ntcA* expression as expected if the cells were N stressed. '+MSX' shows maximum *ntcA* expression when ammonium assimilation is inhibited (Lindell and Post, 2001).

ntcA is a transcriptional activator that regulates many aspects of nitrogen metabolism in cyanobacteria. Marine cyanobacteria induce *ntcA* expression in

response to nitrogen stress, but not phosphorus or iron stress (Lindell and Post, 2001). As such, the level of *ntcA* expression can be used as a metric for N stress of field populations of marine cyanobacteria. This *ntcA* assay has thus far been applied to field *Synechococcus* populations in the Red Sea to show that these cells are not N stressed.

Another *Prochlorococcus* adaptation to efficiently scavenge ambient nitrogen is the ability to assimilate diverse nitrogen species. In fact, closely-related *Prochlorococcus* strains are hypothesized to niche partition the water column by utilizing different nitrogen sources. *Prochlorococcus* has discrete systems to transport and assimilate different N sources (Fig. 6). MED4 has been shown to exclusively utilize N sources such as ammonia and urea which are rapidly recycled in the nutrient-depleted surface waters (Moore et al., 2002). Genome sequencing revealed that MED4 also has genes putatively encoding a cyanate transporter and cyanate lyase (Rocap et al., 2003). Cyanate is a potential alternative N source that is in equilibrium in aqueous solution with urea (Hargel et al., 1971).

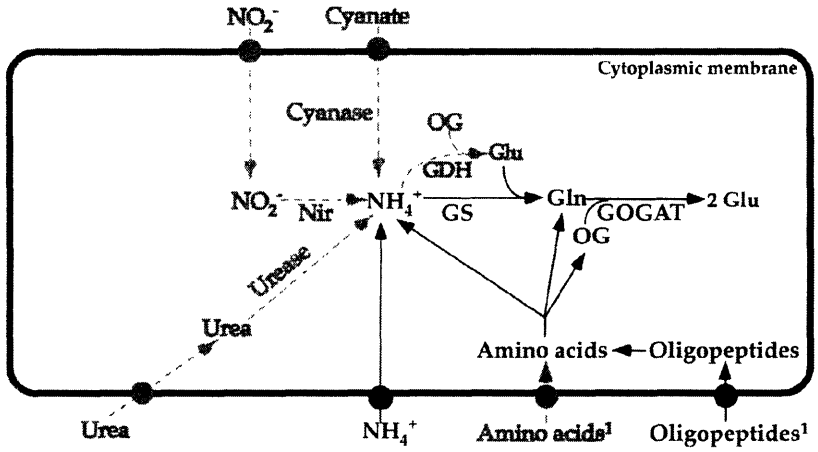


Fig. 6. Diagram of the *Prochlorococcus* cell showing discrete transport and assimilatory routes used for different N sources. Gray indicates N sources utilized by some, but not all, *Prochlorococcus* strains. Note that all N sources must first be reduced to ammonia before being assimilated as biomass (Garcia-Fernandez et al., 2004).

Preliminary studies supported that marine *Synechococcus* WH8102 (Palenik et al., 2003) and *Prochlorococcus* MED4 (Garcia-Fernandez et al., 2004) can grow on cyanate as a sole nitrogen source. In contrast, low light-adapted *Prochlorococcus* strains such as MIT9313 are most abundant in the deep euphotic zone (West et al., 2001) where nitrite levels are elevated (Olson, 1981). MIT9313 grows on ammonia, urea, and nitrite (Moore et al., 2002). Field studies using radio-labelled methionine

demonstrated that *Prochlorococcus* can also uptake amino acids (Zubkov et al., 2003). Unlike the closely-related *Synechococcus*, no *Prochlorococcus* strain has been shown to grown on nitrate and the gene for nitrate reduction, *narB*, is absent from *Prochlorococcus* genomes (Rocap et al., 2003). A number of molecular studies have investigated the expression and function of *Prochlorococcus* nitrogen-regulated genes. These studies have focused on *Prochlorococcus* PCC 9511, which has been shown to be genetically identical to MED4 in terms of the ITS (Laloui et al., 2002) and rDNA (Rippka et al., 2000). Much can also be learned about *Prochlorococcus* nitrogen metabolism by extrapolating from well-studied cyanobacteria such as *Synechococcus* PCC 7942 and *Synechocystis* PCC 6803.

Previous studies have shown that cyanobacterial nitrogen metabolism is governed by two master regulators, PII and NtcA (Fig. 7). The *glnB* gene encodes the PII protein (see Forchhammer, 2004 for a review). PII is a signal transducer that has been likened to the central processing using (CPU) of the cell for its role in coordinating carbon and nitrogen metabolism (Ninfa and Atkinson, 2000). PII monitors cellular nitrogen status by binding the metabolite 2-oxoglutarate (Forchhammer, 1999; Tandeau de Marsac and Lee, 1999) which, in turn, enhances PII phosphorlyation (Forchhammer and Hedler, 1997). PII monitors 2-oxoglutarate because it is the primary carbon-skeleton for ammonium incorporation. 2-oxoglutarate levels are low in ammonium-replete conditions and increase under N starvation (Muro-Pastor et al., 2001).

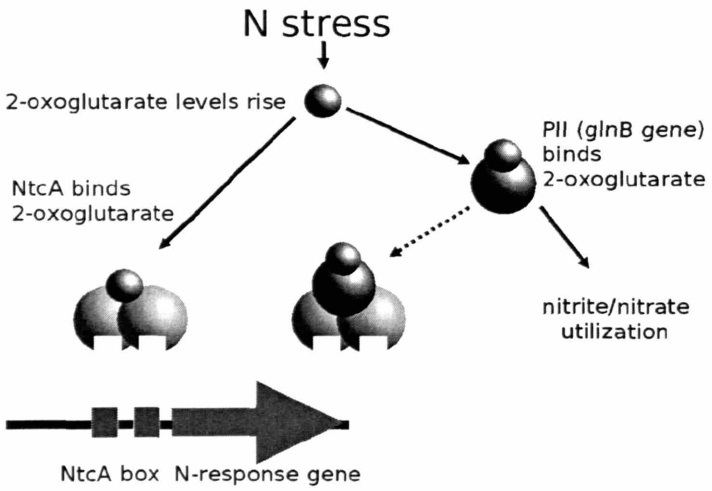


Fig. 7. Proposed mechanism for the interaction of PII, NtcA, and 2-oxoglutarate resulting in the activation of *ntcA*-regulated genes. 2-oxoglutarate levels increase under N deficiency. NtcA binds 2-oxolglutarate and activates the transcription of its targets. PII also binds 2-oxoglutarate and post-transcriptionally activates genes for utilization of oxidized N sources. In addition, there is evidence that NtcA interacts either directly or indirectly with PII.

It has been proposed that PII inhibits the activity of proteins for the uptake of oxidized N species as nitrate and nitrite when cells are in the presence of ammonium. Specifically, *Synechococcus* PCC7942 PII null mutants repress transcription of the *nir-nrtABCD-narB* genes for nitrite/nitrate uptake in the presence of ammonium similar to wild-type cells. The PII mutant, however, persists in the uptake of nitrite and nitrate in the presence of ammonium suggesting that PII acts to post-transcriptionally inhibit uptake of the N sources (Lee et al., 1998). The *Prochlorococcus* PII amino acid sequence contains the conserved cyanobacterial signatures, including the serine residue that is phosphorlyated in other cyanobacteria. However, phylogenetic analysis of PII has shown that the oceanic cyanobacteria form a separate subclade from other strains (Garcia-Fernandez et al., 2004). The *Prochlorococcus* PII protein also appears to function differently than other cyanobacteria in that it is not phosphorlyated in response to nitrogen deprivation (Palinska et al., 2002). It has thus been hypothesized that *Prochlorococcus* PII has a phosphorylation-independent means of regulation, perhaps mediated by the binding an allosteric effector such as 2-oxoglutarate (Forchhammer, 2004).

NtcA is a transcription factor in the CRP family that activates genes which are repressed in the presence of ammonium (Vega-Palas et al., 1990). Ammonium is the only nitrogen source utilized by all *Prochlorococcus* strains and is the preferred N source (Garcia-Fernandez et al., 2004). Oxidized forms of N such as nitrite must be reduced to ammonium for assimilation which is a significant expense with respect to the cellular energy budget (Garcia-Fernandez et al., 2004). The repression of genes for assimilation of alternate N sources in the presence of ammonia is common among cyanobacteria and is called N-control (Herrero et al., 2001). NtcA activates transcription of its targets by binding directly to their promoters with a conserved helix-turn-helix motif in the carboxy terminus. DNAse I footprinting (Luque, et al., 1994), *in vitro* oligonucleotide selection (Jiang et al., 2000), and sequence alignments (Herrero et al., 2001) indicate that *ntcA* binds as a dimer to the palindrome TGTA-N8-TACA. The expression of a number of nitrogen genes are known to be enhanced by *ntcA* including *amt1*, *glnA*, and *glnB* (see Herrero et al., 2001 for a review). A complex feedback exists between *glnB* and *ntcA* (Fig. 7). NtcA enhances the transcription of *glnB* (Lee et al., 1999). However, full activation of NtcA-regulated genes requires the PII protein (Paz-Yepes et al., 2003). NtcA can also act as a repressor for the photosynthetic gene *rbcL* (Ramasubramanian et al., 1994).

The primary avenue by which cyanobacteria assimilate ammonium into carbon skeletons is through its incorporation into glutamine by glutamine synthetase (Fig. 6) (Wolk et al., 1976). The *Prochlorococcus* PCC 9511 GS enzyme, encoded by the *glnA*

gene, is biochemically similar to other cyanobacteria in many respects (El Alaoui et al., 2003). However, the genetic regulation of *Prochlorococcus* glutamine synthetase has been shown to be quite novel. Unlike other cyanobacteria, studies have found that neither the *Prochlorococcus glnA* gene (Garcia-Fernandez et al., 2004) nor the GS protein (El Alaoui et al., 2001; El Alaoui et al., 2003) is upregulated in response to nitrogen starvation.

Prochlorococcus has discrete transport systems for the uptake of different N sources. *Prochlorococcus* takes up ammonia using the high-affinity transporter, *amt1*. *amt1* expression in other cyanobacteria is low in the presence of ammonium and enhanced in low N conditions (Montesinos et al., 1998; Vazquez-Bermudez et al., 2002). In contrast, *Prochlorococcus* PCC 9511 *amt1* expression is not regulated by ammonium availability and is proposed not to be *ntcA*-regulated (Lindell et al., 2002). *Prochlorococcus* also has several transporters for alternate N sources (Fig. 6). Urea is an important N source in many marine environments (DeManche et al., 1973) and both MIT9313 and MED4 have ABC-type urea transporters and urease genes. *Prochlorococcus* PCC 9511 urease activity is independent of the nitrogen source in the medium (Palinska et al., 2000), suggesting that the urease genes lack genetic regulation. MIT9313 has genes for nitrite transport and utilization whereas MED4 does not. The MIT9313 nitrite reductase (*nirA*) is adjacent to a proteobacterial-type nitrite transporter, suggesting that the genes for nitrite transport and utilization were acquired by horizontal gene transfer (Rocap et al., 2003).

In addition to genes involved in the acquisition and metabolism of nitrogen, cyanobacteria up-regulate general stress proteins under N-starvation. For example, cyanobacterial high light-inducible polypeptides (*hli*) are a family of genes that have recently been linked to survival under diverse conditions including nitrogen stress (He et al., 2001). Cyanobacterial *hli* genes were were identified by their similarity to Lhc polypeptides in plants (Dolganov et al., 1995). *Synechocystis* PCC6803 has five genes encoding *hli* polypeptides, all of which are induced during nitrogen starvation (He et al., 2001). Although the precise mechanism is yet unclear, it has been proposed that *hli* genes aid in the acclimation of cells to the absorption of excess light energy, perhaps by suppressing reactive oxygen species (He et al., 2001). The *hli* genes represent an extended gene family in *Prochlorococcus*, MED4 has 22 *hli* genes and MIT9313 has 9 genes (Rocap et al., 2003). By examining the expression patterns of *Prochlorococcus hli* genes, our goal was to learn more about their role in mediating the N-stress response.

Several of the studies described above suggest that regulation of nitrogen genes in *Prochlorococcus* is fundamentally different from other cyanobacteria:

glnA/GS is not changed in its abundance or activity under N-stress, *amt1* is not induced under N-stress, and PII is not phosphorylated under any tested conditions. These differences in the regulation of *Prochlorococcus* N metabolism genes relative to other cyanobacteria have been proposed as an adaptation to a homogenous, oligotrophic environment (Garcia-Fernandez et al., 2004). Global mRNA expression profiling combined with physiological measurements of N starvation provide an unprecedented opportunity to address questions about novel patterns of gene regulation in *Prochlorococcus*.

Prochlorococcus genetic transformation

In future studies, microarray data from multiple, independent experiments will be combined to determine a subset of genes that are altered in expression in a specific physiological state. For example, one will determine the subset of genes that are upregulated under N stress, but not P or Fe stress. In order to move beyond expression patterns and determine that a given gene is directly involved in mediating a physiological response, one needs methods to directly connect genotype to phenotype. Microarray experiments allow one to conclude that a given gene is elevated in expression under N stress, but how is the N stress response altered if this gene is disrupted?

Genetic methods provide an elegant means to directly connect genotype to phenotype by the introduction of foreign DNA into the target cell *in vivo*. Unfortunately, our direct knowledge of bacterial genetics relies upon a small number of well-studied model systems, most of which were chosen because of their clinical importance. Few genetic systems exist to study prokaryotes of ecological importance. *Prochlorococcus* represents a potential candidate for an ecologically relevent genetic system because many strains in are in culture and three (MED4, MIT9313, and MIT9312) have been rendered free of contaminants.

A goal of this thesis was to develop a system for the genetic transformation of *Prochlorococcus*. Prokaryotic genetic systems have three basic prerequisites. First, one must develop a means to deliver foreign DNA into the cell. The most common gene transfer system used in cyanobacteria is DNA-mediated transformation. Transformation methods have been clearly demonstrated in several strains of *Synechococcus* and *Synechocystis* (Porter, 1986). DNA-mediated transformation involves the direct uptake of naked DNA from the environment and thus requires conditions under which the recipient cell is competent to uptake DNA. Cell competence can be either natural or artificial. Natural competence describes the

condition when cells are able to naturally internalize exogenous DNA without special treatment. Cyanobacteria such as *ThermoSynechococcus elongatus* have been shown to be naturally competent (Onai et al., 2004). In contrast, artificial competence describes conditions whereby DNA uptake requires special treatment such as heat shock or electroporation. Electroporation has also been shown to be effective with certain freshwater cyanobacteria (Poo, 1997). However, cells cannot be electroporated in seawater because of its high conductivity. Cells must be instead be resuspended in a low electrical conductivity medium of the proper osmolarity. *Prochlorococcus* survives transfer to sorbitol-based media (Wolfgang Hess, personal communication) but cells have low survivorship following electroporation.

To date, there is no evidence for natural or artifical competence in *Prochlorococcus*. We therefore focused on conjugation-based methods because of their high efficiency and insensitivity to species barriers. Conjugation is a general means to introduce DNA from *E. coli* to diverse cyanobacteria (Wolk et al., 1984) using the broad host range conjugal apparatus of the RP4 plasmid. RP4, originally isolated from *Pseudomonas*, can mediate DNA transfer to a wide range of bacteria including myxobacteria (Breton et al., 1985), thiobacilli (Kulpa et al., 1983), and cyanobacteria (Wolk et al., 1984). These conjugation methods have even been extended to transfer DNA from *E. coli* to mammalian cells (Waters, 2001). Our initial challenge was to find a means by which conjugation methods could be adapted to *Prochlorococcus*.

The role of the conjugal plasmid is to construct an apparatus by which a second plasmid may be transferred into the recipient cell (Fig. 8). Conjugal plasmids are quite large (approximately 60 kb) because of the numerous genes required to build the pilus for DNA transfer.

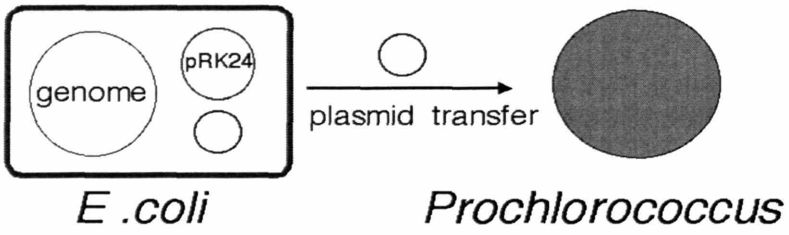


Fig. 8. Biparental mating strategy for the conjugal transfer of DNA from *E. coli* to *Prochlorococcus*. The *E. coli* cell contains two plasmids, the conjugal plasmid (here the RP4 derivative pRK24) and the transfer plasmid. The conjugal plasmid encodes genes for the pilus by which the transfer plasmid passes to the *Prochlorococcus* cell.

The transfer plasmid needs two features in order to be transferred by conjugation. First, the transfer plasmid much contain an origin of transfer (oriT)

which is cut when then plasmid is linearized during conjugation. Second, the transfer plasmid must encode, or be provided with, a nicking protein (mob gene) that recognizes and cuts at the oriT. In addition, the transfer plasmid should contain an origin of replication (oriV) and an antibiotic resistance marker. If the goal is to ectopically express a gene, then the transfer plasmid should have an oriV that replicates autonomously in the recipient cell. If the goal is targeted mutagenesis, then the origin can either replicate in the recipient (shuttle vector) or not (suicide vector). Suicide vectors are often preferable for targeted mutagenesis because the only means by which the recipient cell can continue to be be antibiotic resistant is if the plasmid integrates into the host chromosome. Finally, the transfer plasmid should contain an antibiotic resistance gene that allows exconjugants to be selected away from cells that did not receive the transfer plasmid. The transfer plasmid conjugated into *Prochlorococcus* in this thesis is shown in Fig. 9.

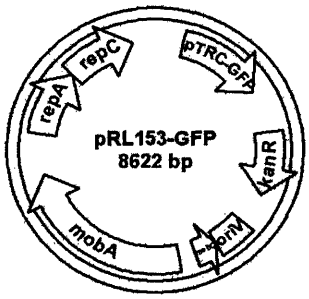


Fig. 9. Replicating plasmid for conjugal transfer to *Prochlorococcus*. pRL153 is a kanamycin-resistant derivative of the broad host range plasmid RSF1010. It contains an oriT, the requisite mob proteins, and an oriV that replicates in *E. coli* and diverse cyanobacteria. In addition, it has been modified to express GFP from the synthetic pTRC promoter.

Beyond the ability to transfer foreign DNA, the second prerequisite for a genetic system is the ability to express foreign proteins in the target cell. As described above, the expression of an antibiotic resistance gene is crucial to isolate exconjugants from their non-transformed brethren. The *nptI* gene derived from Tn903, (Oka et al., 1981) encoding the neomycin phosphotransferase conferring kanamycin resistance, is an effective selective marker in diverse cyanobacteria (Friedberg, 1988). However, different cyanobacteria taxa and even different strains of the same taxa (see Appendix IV of this thesis) have widely varying sensitivities to antibiotics such as kanamycin.

Reporter genes are another application requiring the expression of foreign proteins. Reporter genes fused to specfic promoters are often used for the analysis of the regulation of gene expression. The product of the reporter gene should be easily quantifiable and its synthesis should allow selection of cells expressing the

gene. Common reporter genes include chloramphenicol acetyltransferase (*cat*), beta-galactosidase (*lacZ*), luciferase (*lux*), and green fluorescent protein (*GFP*) genes. The *lux* genes have been used with great success in *Synechococcus* PCC7942 to show global circadian oscillation of gene expression (Ditty et al., 2003). A set of experiments in this thesis developed methods for the expression and quantification of the reporter gene GFP in *Prochlorococcus*.

Another application requiring the expression of foreign proteins in the recipient cell is transposon mutagenesis. A transposon is a DNA sequence that can move from one place in DNA to another with the aid of a transposase enzyme. Transposon mutagenesis is a technique by which a transposon is used to make random insertion mutations in the host chromosome. Transposon mutagenesis has been widely used in other cyanobacteria as a means to randomly inactivate gene function so as to study processes such as heterocyst formation (Cohen et al., 1998). Recently, The Tn5 transposon has been shown to transpose in the marine cyanobacterium *Synechococcus* (McCarren and Brahamsha, 2005) and permit the identification of genes required for mobility in *Synechococcus* WH8102. In this thesis, we show that Tn5 will also transpose *in vivo* in *Prochlorococcus*.

Once one has developed methods for DNA transfer and expression of foreign proteins, the final requirement for a genetic system is a means to isolate and identify isogenic mutants. Isolation of mutants is traditionally done by streaking cells on the surface of solid, agar-based media. However, oceanic cyanobacteria such as *Prochlorococcus* and *Synechococcus* are notoriously difficult to grow on the surface of plates perhaps because they are sensitive to dessication. An alternative plating protocol has been developed in which cells are embedded in low contentration agarose media (Brahamsha et al., 1996). This method has been applied with some success is certain *Prochlorococcus* strains and is the basis for isolating isogenic *Prochlorococcus* mutants in our experiments.

A *Prochlorococcus* genetic system thus has three requirements: introduction of foreign DNA to *Prochlorococcus* by interspecific conjugation with *E. coli*, discovery of plasmids for the expression of foreign genes in *Prochlorococcus*, and methods to isolate isogenic mutants. Many microarray and genomic studies will be completed in the next few years that will hypothesize cellular roles for *Prochlorococcus* genes based on sequence similarities and expression patterns. Genetic these methods can then be used to directly connect genotypic changes with a resulting *Prochlorococcus* phenotype.

Optimization strategies for microarray synthesis

Oligonucleotide microarrays, such as those developed for *Prochlorococcus*, are a primary tool in the field of genomics. These oligonucleotide arrays are synthesized using a modification of the photolithographic method developed in the semiconductor industry. In this method, the nucleotides A, C, G, and T are added to the appropriate positions in a series of cycles that construct the oligonucleotides *in situ* on the array surface. Each cycle requires a custom mask that permits light to penetrate at defined positions, thereby activitating the proper oligonucleotides for synthesis. The pattern in which light passes through a series of masks directs the base-by-base synthesis of oligonucleotides on the chip surface by repeated cycles of photodeprotection and nucleotide addition. Because of these custom masks and the photodeprotection reagents, the time and synthesis cost of an oligonucleotide array is largely a function of the number of cycles required to synthesize the oligonucleotides. Thus, it is of paramount importance to manufacture oligonucleotide arrays in as few cycles as possible. **The goal of this section of the thesis was to computationally model strategies to reduce the number of synthesis cycles required to fabricate oligonucleotide microarrays.** This area of research is called the s*ynthesis strategy optimization problem*.

The optimal synthesis strategy for a set of oligonucleotides is equivalent to the shortest common super-sequence problem (Kasif et al., 2002). The shortest common super-sequence (SCS) is a well-studied algorithmic problem in computer science (Jiang and Li, 1997) that is known to be NP-hard, meaning that the optimal solution cannot be found in polynomial time. The SCS problem is can also be thought of as a special case of the multiple sequence alignment problem (Kasif et al., 2002). As such, the discovery of an optimal strategy for a large set of oligonucleotides is computationally infeasible. Improvements for oligonucleotide synthesis are thus sought using heuristics.

The simplest method to construct a set of oligonucleotides is by adding A,C,G,T in series. If the oligonucleotides are of length K, then this strategy requires a maximum of 4K cycles. However, the optimal synthesis strategy requires many fewer than 4K cycles. One method to decrease the required number of cycles is to allow the oligonucleotides to be built at different rates (Fig. 10). Another way to reduce the required synthesis cycles is to skip a cycle if the nucleotide to be added is not needed by any of the oligonucleotides or if the set of oligonucleotides can still be synthesized when it is deposited later (Hubbell et al., 1996). In this thesis, we investigate several methods for further improving synthesis strategies. First, we focus on how to best find regions within each gene containing

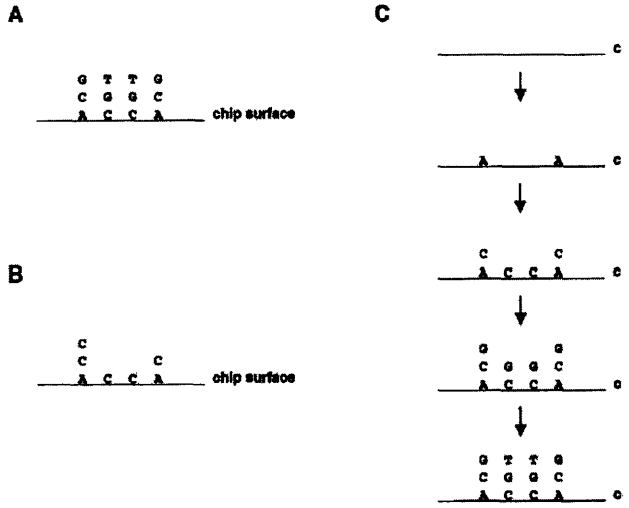


Fig. 10. *In situ* synthesis of an array of oligonucleotides on solid surface. The set of oligonucleotides shown in **A**. can be synthesized in 4 steps by allowing the oligonucleotides to grow at different rates using the strategy shown in **C**. (Kasif et al., 2002).

oligonucleotides that could be efficiently deposited. Second, we develop 'greedy approaches' that alter the nucleotide deposition order to maximize the number of nucleotides deposited at each step. By simultaneously improving oligonucleotide selection and deposition we significantly reduce the number of deposition cycles required to synthesize an oligonucleotide array.

REFERENCES

Bertilsson, S., O. Berglund, et al. (2003). "Elemental composition of marine *Prochlorococcus* and *Synechococcus*: Implications for the ecological stoichiometry of the sea." Limnology and Oceanography **48**(5): 1721-1731.

Brahamsha, B. (1996). "A genetic manipulation system for oceanic cyanobacteria of the genus *Synechococcus*". Applied and Environmental Microbiology. **62** (5): 1747-1751.

Breton, AM., Jaoua, S., Guespin-Michel, J. (1985). "Transfer of plasmid RP4 to *Myxococcus-xanthus* and evidence for its integration into the chromosome ". Journal of Bacteriology **161**: 523-

Campbell, L., Landry, MR., et al. (1998). "Response of microbial community structure to environmental forcing in the Arabian Sea." Deep-Sea Research Part II-Topical Studies in Oceanography **45**(10-11): 2301-2325.

Charpy, L, and Blanchot, J. (1996). "*Prochlorococcus*contribution to phytoplankton biomass and production of Takapoto atoll (Tuamotu archipelago). CR. Acad. Sci. (Paris). **319**: 131-137.

Cohen, MF., Meeks, JC., Cai, YA., Wolk, CP. (1998). "Transposon mutagenesis of heterocyst-forming filamentous cyanobacteria". Methods in Enzymology **297**: 3-17.

Delcher, AL., Kasif, S., Fleischmann, RD., Pterson, J., White, O., Salzberg, SL. (1999). "Alignment of whole genomes" Nucleic Acids Research 27 (11): 2369-2376

DeManche, JM., Curl, H, et al. (1973). "Automated Analysis for Urea in Seawater." Limnology and Oceanography **18**(4): 686-689.

Ditty, JL., Williams, SB., Golden, SS. (2003). "A cyanobacterial circadian timing mechanism". Annual Review of Genetics 37: 513-543.

Dolganov, NA., D. Bhaya, et al. (1995). "Cyanobacterial protein with similarity to the chlorophyll a/b binding proteins of higher plants: evolution and regulation." Proc Natl Acad Sci U S A **92**(2): 636-40.

Dufresne, A., M. Salanoubat, et al. (2003). "Genome sequence of the cyanobacterium *Prochlorococcus* marinus SS120, a nearly minimal oxyphototrophic genome." Proc Natl Acad Sci U S A **100**(17): 10020-5.

El Alaoui, S., J. Diez, et al. (2001). "In vivo regulation of glutamine synthetase activity in the marine chlorophyll b-containing cyanobacterium *Prochlorococcus* sp strain PCC 9511 (oxyphotobacteria)." Applied and Environmental Microbiology **67**(5): 2202-2207.

El Alaoui, S., J. Diez, et al. (2003). "Glutamine synthetase from the marine cyanobacteria *Prochlorococcus* spp.: characterization, phylogeny and response to nutrient limitation." Environmental Microbiology **5**(5): 412-423.

Forchhammer, K. and A. Hedler (1997). "Phosphoprotein PII from cyanobacteria--analysis of functional conservation with the PII signal-transduction protein from Escherichia coli." Eur J Biochem **244**(3): 869-75.

Forchhammer, K. (1999). The PII protein in *Synechococcus* PMM7942 senses and signals 2-oxoglutarate under ATP-replete conditions. The Phototrophic Prokaryots. W. Loeffelhardt, G. Schmetterer and G. A. Peschek. New York, Kluwer Academic / Plenum Publishers**:** 463-477.

Forchhammer, K. (2004). "Global carbon/nitrogen control by PII signal transduction in cyanobacteria: from signals to targets." FEMS Microbiol Rev **28**(3): 319-33.

Friedberg, D. (1988). "Use of reporter genes in cyanobacteria". Methods in Enzymology **167**: 736-747.

Garcia-Dominguez, M., J. C. Reyes, et al. (1997). "Purification and characterization of a new type of glutamine synthetase from cyanobacteria." Eur J Biochem **244** (1): 258-64.

Garcia-Fernandez, J. M., N. T. de Marsac, et al. (2004). "Streamlined regulation and gene loss as adaptive mechanisms in *Prochlorococcus* for optimized nitrogen utilization in oligotrophic environments." Microbiol Mol Biol Rev **68**(4): 630-8.

Graziano, L. M., R. J. Geider, et al. (1996). "Nitrogen limitation of North Atlantic phytoplankton: Analysis of physiological condition in nutrient enrichment experiments." Aquatic Microbial Ecology **11**(1): 53-64.

Hagel, P., J. J. Gerding, et al. (1971). "Cyanate formation in solutions of urea. I. Calculation of cyanate concentrations at different temperature and pH." Biochim Biophys Acta **243**(3): 366-73.

He, Q., N. Dolganov, et al. (2001). "The high light-inducible polypeptides in Synechocystis PCC6803. Expression and function in high light." J Biol Chem **276**(1): 306-14.

Herrero, A., A. M. Muro-Pastor, et al. (2001). "Nitrogen control in cyanobacteria." Journal of Bacteriology **183**(2): 411-425.

Hess, WR., Rocap, G., Ting, CS., Larimer, F., Stilwagen, S., Lamerdin, J., Chisholm, SW. (2001). "The photosynthetic apparatus of *Prochlorococcus*: Insights through comparative genomics"**.** Photosynthesis Research **70** (1): 53-71.

Hubbell, EA., Morris, MS., Winkler, JL. (1996). "Computer-aided engineering system for design of sequence arrays and lithographic masks. US Patent no. 5571639.

Jiang, T. and Li, M. (1997). "On the approximation of shortest common supersequences and longest common subsequences". SIAM J. Comput. **24**: 1122-1139.

Jiang, F. Y., S. Wisen, et al. (2000). "Examination of the transcription factor *ntcA*-

binding motif by in vitro selection of DNA sequences from a random library." Journal of Molecular Biology **301**(4): 783-793.

Johnson, Z., Landry, ML., Bidigare, RR., Brown, SL., Campbell, L., Gunderson, J., Marra, C., Trees, C. (1999). "Energetics and growth kinetics of a deep *Prochlorococcus* spp. population in the Arabian Sea". Deep-Sea Research II **46**, 1719-1743.

Kasif, S., Weng, Z., Derti, A., Beigel, R., DeLisi, C. (2002). "A computational framework for optimal masking in the synthesis of oligonucleotide microarrays". Nucleic Acids Research 30 (20): e106

Kudela, RM. and R. C. Dugdale (2000). "Nutrient regulation of phytoplankton productivity in Monterey Bay, California." Deep-Sea Research Part II-Topical Studies in Oceanography **47**(5-6): 1023-1053.

Kulpa, CF., Roskey, MT., Travis, MT. (1983). "Transfer of plasmid-RP1 into chemolithotrophic *Thiobacillis neapolitanus* ". Journal of Bacteriology **156** (1): 434-436.

Laloui, W., K. A. Palinska, et al. (2002). "Genotyping of axenic and non-axenic isolates of the genus *Prochlorococcus* and the OMF-'*Synechococcus*' clade by size, sequence analysis or RFLP of the Internal Transcribed Spacer of the ribosomal operon." Microbiology **148** (Pt 2): 453-65.

Lee, HM., Flores, E., Herrero, A., Houmard, J., Tandeau de Marsac, N. (1998). "A role for the signal transduction protein PII in the controle of nitrate/nitrite uptake in a cyanobacterium". FEBS Letters 427, 291-295.

Lee, H. M., M. F. Vazquez-Bermudez, et al. (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." Journal of Bacteriology **181**(9): 2697-2702.

Li, W. K. W. (1994). "Primary Production of Prochlorophytes, Cyanobacteria, and Eukaryotic Ultraphytoplankton - Measurements from Flow Cytometric Sorting." Limnology and Oceanography **39** (1): 169-175.

Lindell, D., and Post, AF. (2001). "Ecological aspects of *ntcA* gene expression and its use as an indicator of nitrogen status of marine *Synechococcus* spp." Applied and Environmental Microbiology. **67**: 8 3340-3349.

Lindell, D., D. Erdner, et al. (2002). "The nitrogen stress response of *Prochlorococcus* strain PCC 9511 (Oxyphotobacteria) involves contrasting regulation of *ntcA* and *amt1*." Journal of Phycology **38**(6): 1113-1124.

Luque, I., E. Flores, et al. (1994). "Molecular Mechanism for the Operation of Nitrogen Control in Cyanobacteria." Embo Journal **13**(12): 2862-2869.

Luque, I., Contreras, A., Zabulon, G., Herrero, A., Houmard, J. (2002). "Expression of the glutamyl-tRNA synthetase gene from the cyanobacterium *Synechococcus* sp. PCC 7942 depends on nitrogen availability and the global regulator *ntcA*". Molecular Microbiology **46** (4): 1157-1167.

McCarren, J. and Brahamsha, B. (2005). "Transposon mutagenesis in a marine *Synechococcus* strain: Isolation of swimming motility mutants". Journal of Bacteriology. 4457-4462.

Mérida, A., Candau, P., Florencio, FJ. (1991). "Regulation of glutamine synthetase activity in the unicellular cyanobacterium Synechocystis sp. strain PCC 6803 by the nitrogen source: effect of ammonium". Journal of Bacteriology **173**: 4095-4100.

Montesinos, M. L., A. M. Muro-Pastor, et al. (1998). "Ammonium/methylammonium permeases of a Cyanobacterium. Identification and analysis of three nitrogen-regulated amt genes in synechocystis sp. PCC 6803." Journal of Biol Chem **273** (47): 31463-70.

Moore, L. R., R. Goericke, et al. (1995). "Comparative Physiology of *Synechococcus* and *Prochlorococcus* - Influence of Light and Temperature on Growth,

Pigments, Fluorescence and Absorptive Properties." Marine Ecology-Progress Series **116**(1-3): 259-275.

Moore, L. R., G. Rocap, et al. (1998). "Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes." Nature **393** (6684): 464-7.

Moore, L. R., A. F. Post, et al. (2002). "Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*." Limnology and Oceanography **47** (4): 989-996.

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

Ninfa, A. J. and M. R. Atkinson (2000). "PII signal transduction proteins." Trends in Microbiology **8**(4): 172-179.

Oka, A., Sugisaki, H., Takanami, M. (1981). "Nucleotide-sequence of the kanamycin resistance transposon TN903". Journal of Mol. Bio. 147: 217-

Olson, RJ. (1981). "^{15}N tracer studies of the primary nitrite maximum." Journal of Marine Research **39**: 1033-1051.

Olson, RJ. (1985) unpublished data.

Onai, K., Morishita., M., Kaneko, T., Tabata, S., Ishiura, M. (2004). "Natural transformation of the thermophilic cyanobacterium *ThermoSynechococcus elongatus* BP-1: a simple and efficient methods for gene transfer". Molecular Genetics and Genomics **271** (1): 50-59.

Palenik, B., B. Brahamsha, et al. (2003). "The genome of a motile marine *Synechococcus*." Nature **424** (6952): 1037-42.

Palinska, K. A., T. Jahns, et al. (2000). "*Prochlorococcus* marinus strain PCC 9511, a picoplanktonic cyanobacterium, synthesizes the smallest urease." Microbiology **146 Pt 12**: 3099-107.

Palinska, K. A., W. Laloui, et al. (2002). "The signal transducer P-II and bicarbonate acquisition in *Prochlorococcus* marinus PCC 9511, a marine cyanobacterium naturally deficient in nitrate and nitrite assimilation." Microbiology-Sgm **148**: 2405-2412.

Partensky, F., Hess, WR., Vaulot, D. (1999). "*Prochlorococcus*, a marine photosynthetic prokaryote of global significance". Microbiology and Molecular Biology Reviews **63**: 1 106-127.

Paz-Yepes, J., E. Flores, et al. (2003). "Transcriptional effects of the signal transduction protein PII (*glnB* gene product) on *ntcA*-dependent genes in *Synechococcus* sp PCC 7942." FEBS Letters **543** (1-3): 42-46.

Poo, H. (1997). "Optimal condition for efficient DNA transfer in filamentous cyanobacteria". Journal of Microbiology **35** (3): 181-187.

Porter, RD. (1986). "Transformation in cyanobacteria ". CRC Crit. Rev. Microbiology. **13** (2): 111-132.

Ramasubramanian, T. S., T. F. Wei, et al. (1994). "2 Anabaena Sp Strain Pcc 7120 DNA-Binding Factors Interact with Vegetative Cell-Specific and Heterocyst-Specific Genes." Journal of Bacteriology **176** (5): 1214-1223.

Raven, JA. (1994). "Why are there no picoplankton O_2 evolvers with volumes less than 10^{-19} m^3?". J. Plankton Res. **16**: 565-580.

Redfield, AC. (1958). "The biological control of chemical factors in the environment." American Scientist **46**: 205-221.

Rippka, R., T. Coursin, et al. (2000). "*Prochlorococcus* marinus Chisholm et al. 1992 subsp. pastoris subsp. nov. strain PCC 9511, the first axenic chlorophyll a2/b2-containing cyanobacterium (Oxyphotobacteria)." Int J Syst Evol Microbiol **50 Pt 5**: 1833-47.

Rocap, G., Distel, D., Waterbury, JB., Chisholm, SW. (2002). "Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences". Applied and Environmental Microbiology **68** (3): 1180-1191.

Rocap, G., F. W. Larimer, et al. (2003). "Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation." Nature **424** (6952): 1042-7.

Tandeau de Marsac, N. and H. M. Lee (1999). Regulation of carbon and nitrogen metabolism in the unicellular cyaobacteria *Synechococcus* sp. The Phototrophic Prokaryots. W. Loeffelhardt, G. Schmetterer and G. A. Peschek. New York, Kluwer Academic / Plenum Publishers**:** 463-477.

Urbach, E., Scanlan, D., Distel, DL., Waterbury, JB., Chisholm, SW. (1998). "Rapid diversification of marine picophytoplankton with dissimilar light-harvesting structures inferred from sequences of *Prochlorococcus* and *Synechococcus* (Cyanobacteria)". Journal of Molecular Evolution **46** (2): 188-201.

Vazquez-Bermudez, M. F., J. Paz-Yepes, et al. (2002). "The *ntcA*-activated *amt1* gene encodes a permease required for uptake of low concentrations of ammonium in the cyanobacterium *Synechococcus* sp. PCC 7942." Microbiology **148**(Pt 3): 861-9.

Vega-Palas, MA., F. Madueño, et al. (1990). "Identification and Cloning of a Regulatory Gene for Nitrogen Assimilation in the Cyanobacterium *Synechococcus* Sp Strain Pcc-7942." Journal of Bacteriology **172** (2): 643-647.

Veldhuis, MJW., Kraay, GW. (1990). "Vertical distribution and pigments composition of a picoplankton prochlorophyte in the subtropical North-Atlantic: a combined study of HPLC-analysis of pigments and flow cytometry". Marine Ecol. Prog. Series **68**: 121-127.

Waters, VL. (2001). "Conjugation between bacterial and mammalian cells". Nature Genetics **29** (4): 375-376.

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-SGM **147**: 1731-1744.

Wolk, CP., J. Thomas, et al. (1976). "Pathway of Nitrogen-Metabolism after Fixation of N-13-Labeled Nitrogen Gas by Cyanobacterium, Anabaena-Cylindrica." Journal of Biological Chemistry **251** (16): 5027-5034.

Wolk, CP., Vonshak, A., Kehoe, P., Elhai, J. (1984). "Construction of shuttle vectors capable of conjugative transfer from Escherichia-coli to nitrogen-fixing filamentous cyanobacteria". PNAS. **81**: 1561.

Zubkov, M. V., B. M. Fuchs, et al. (2003). "High rate of uptake of organic nitrogen compounds by *Prochlorococcus* cyanobacteria as a key to their dominance in oligotrophic oceanic waters." Appl Environ Microbiol **69** (2): 129

"Global gene expression of *Prochlorococcus* ecotypes under nitrogen starvation and on different nitrogen sources"

This project was done with the following collaborators: John Aach[1], George Church[1], Zackary I. Johnson[2], Debbie Lindell[3], Anton Post[4], Trent Rector[1]

1. *Dept. of Genetics, Harvard Medical School. Boston, MA USA*
2. *Dept of Oceanography, University of Hawaii. Honolulu, Hawaii USA*
3. *Dept. of Civil and Environmental Engineering, MIT. Cambridge, MA USA*
4. *Interuniversity Institute of Eilat. Eilat, Israel*

ABSTRACT

Prochlorococcus is the most abundant phytoplankton in the oligotrophic, oceanic gyres where major plant nutrients such as N and P are at nanomolar concentrations. Nitrogen (N) availability controls primary productivity in many of these regions. The cellular mechanisms that *Prochlorococcus* uses to respond to changes in ambient nitrogen are thus central to its ecology. We characterized the N-stress response of two *Prochlorococcus* strains, MED4 and MIT9313, by measuring changes in global mRNA expression, chlorophyll fluorescence, and Fv/Fm along a time-series of increasing N starvation. Initially, both strains of *Prochlorococcus* responded to N-stress by inducing the expression of a set of genes which promoter analysis support are an *ntcA* regulon. The latter stages of N-stress involved genome-wide changes in gene expression such as repression of photosynthesis and translation. Comparison of MED4 and MIT9313 expression profiles revealed differences in the expression of central nitrogen metabolism genes such as *glnA*, *glnB*, and *amt1*. In addition, the two strains up-regulated different N transporters in response to N starvation. A subset of the high light-inducible genes (hli genes) responded to nitrogen starvation in both strains. In addition, we identified conserved genes of unknown function that were highly up-regulated under N starvation and may thus be suitable as novel field probes for *Prochlorococcus* N stress.

Numerous *Prochlorococcus* strains have been isolated that differ in their rDNA sequences and nutrient physiologies. For example, *Prochlorococcus* strains are hypothesized to niche-partition the water column by utilizing different N sources. MIT9313 is restricted to the deep euphotic zone near the nitracline and utilizes ammonia, urea, and nitrite. MED4 is most abundant in the surface waters and grows on ammonia, urea, and cyanate. In this study, we characterized the global mRNA expression profiles of the two strains on these alternative N sources relative to expression in ammonia. A subset of the hli genes were increased in both strains on alternative N sources along with a host of unknown proteins. MIT9313 induced nitrite and urea transporters and repressed *glnB* on both alternative N sources. MED4 repressed sigA on both alternative N sources. The MED4 cyanate transporters and *glnA* were increased in cyanate media. MED4 did not alter expression of urea transporter and utilization genes in urea media. We discuss novel findings about *Prochlorococcus* nitrogen metabolism and their implications for the ecology of this globally abundant phytoplankton.

INTRODUCTION

Prochlorococcus is the most abundant member of the oceanic phytoplankton community in diverse ocean regions (Partensky et al., 1999). Measurements in the Arabian Sea have quantified *Prochlorococcus* densities of 700,000 cells per milliliter of seawater (Campbell et al., 1998). As the numerically dominant phytoplankton, *Prochlorococcus* contributes significantly to global phytoplankton productivity. Phytoplankton productivity greatly influences global geochemical cycles and, ultimately, the composition of the Earth's atmosphere (Falkowski et al., 1998). Phytoplankton growth is regulated by the availability of fixed inorganic nitrogen (N) in

many areas of the coastal (Kudela and Dugdale, 2000) and open ocean (Capone, 2000). It is thus important to understand how *Prochlorococcus* responds to changes in ambient nitrogen. This study examines how two strains of *Prochlorococcus*, MED4 and MIT9313, respond genetically and physiologically to N starvation and different N sources.

Prochlorococcus thrives in oligotrophic waters that are depleted of the primary macronutrients nitrogen and phosphorus (Partensky et al., 1999), but the cells have elevated N requirements relative to P. *Prochlorococcus* cell quotas are >20N:1P (Bertilsson et al., 2003) and thus exceed the 16N:1P Redfield Ratio classically believed to dictate the elemental composition of biomass in the sea (Redfield, 1958). If the nutrient ratios in the ambient seawater are 16N:1P and the MED4 cellular requirements are >20N:1P, then *Prochlorococcus* would have a propensity to become N limited relative to P. In support of this hypothesis, field studies have shown that nitrogen enrichment stimulates *Prochlorococcus* growth in the North Atlantic (Graziano et al., 1996).

Prochlorococcus can be broadly divided into two "ecotypes" based upon growth physiology and rDNA sequence. High light-adapted ecotypes including MED4 are most abundant in the surface waters and low light-adapted ecotypes such as MIT9313 are confined to deeper in the euphotic zone near the nitracline (West et al., 2001. Closely-related strains of *Prochlorococcus* are hypothesized to niche-partition the water column by utilizing different nitrogen sources. MED4 utilitzes ammonia and urea (Moore et al., 2002) which are rapidly recycled in the nutrient-depleted surface waters. The MED4 genome also contains genes putatively encoding a cyanate transporter and cyanate lyase (Rocap et al., 2003). Cyanate is potential alternative N source that is in equilibrium in aqueous solution with urea (Hagel et al., 1971). Culture-based studies have reported that marine *Synechococcus* WH8102 (Palenik et al., 2003) and *Prochlorococcus* MED4 (Garcia-Fernandez et al., 2004) can grow on cyanate as a sole nitrogen source. Low light-adapted strains such as MIT9313 are most abundant in the deep euphotic zone (West et al., 2001) where nitrite levels are elevated (Olson, 1981). MIT9313 grows on ammonia, urea, and nitrite (Moore et al., 2002). Field studies using radio-labelled methionine demonstrated that *Prochlorococcus* can also uptake amino acids (Zubkov et al., 2003). Unlike the closely-related *Synechococcus*, no *Prochlorococcus* strain has been shown to grown on nitrate; the gene for nitrate reduction, narB, is absent from *Prochlorococcus* genomes (Rocap et al., 2003).

A primary goal of this study is to understand *Prochlorococcus* nitrogen metabolism from the perspective of two master nitrogen regulators, *ntcA* and *glnB.*

NtcA is a transcriptional activator of genes that are repressed in the presence of ammonia (Vega-Palas et al., 1990). *glnB* encodes the PII protein (see Forchhammer, 2004 for a review) which has been proposed to act post-transcriptionally to inhibit the activity of genes for the uptake of oxidized N species as nitrate and nitrite (Lee et al., 1999). Several studies have focused nitrogen-regulated genes in *Prochlorococcus*. In addition, much has been learned about *Prochlorococcus* N metabolism by extrapolating from more well-studied cyanobacteria such as *Synechococcus* PCC 7942 and *Synechocystis* PCC 6803. This introduction describes what was previously known about cyanobacterial nitrogen metabolism by highlighting several of these studies.

NtcA is one of the master regulators of cyanobacterial N metabolism. It is a transcription factor in the CRP family that activates the transcription of genes which are repressed in the presence of ammonium (Vega-Palas et al., 1990). Ammonium is the preferred N source because oxidized N species such as nitrite must first be reduced to ammonium for assimilation; reduction of alternative N sources is a significant expense with respect to the cellular energy budget (Garcia-Fernandez et al., 2004). The repression of genes for assimilation of alternate N sources in the presence of ammonia is common among cyanobacteria and is called N-control (Herrero et al., 2001). NtcA alters the transcription by binding the promoters of its targets at the site TGTA-N8-TACA (Luque et al., 1994; Jiang et al., 2000; Herrero et al., 2001). NtcA upregulates the transcription of many N-metabolism genes including *glnB* (see Herrero et al., 2001 for a review). A feedback exists between PII and NtcA. NtcA enhances the transcription of *glnB* (Lee et al., 1999). However, full activation of NtcA-regulated genes requires the *glnB* (Paz-Yepes et al., 2003).

PII is a signal transducer that has been likened to the central processing unit (CPU) of the cell for its role in coordinating carbon and nitrogen metabolism (Ninfa and Atkinson, 2000). PII monitors cellular nitrogen status by binding 2-oxoglutarate (Forchhammer, 1999; Tandeau de Marsac and Lee, 1999) which, in turn, enhances PII phosphorylation (Forchhammer and Hedler, 1997). PII monitors 2-oxoglutarate because it is the branch point between C and N assimilation. 2-oxoglutarate levels are low in ammonium-replete conditions and increase under N starvation (Muro-Pastor et al., 2005). The *Prochlorococcus* PCC9511 PII amino acid sequence contains the conserved cyanobacterial signatures, including the serine residue that is phosphorylated in other cyanobacteria. The *Prochlorococcus* PCC9511 PII protein, however, appears to function differently in that it is not phosphorylated in response to nitrogen deprivation (Palinska et al., 2000).

The primary avenue by which cyanobacteria assimilate ammonium into carbon

skeletons is through its incorporation into glutamine by glutamine synthetase (Wolk et al., 1976). The *Prochlorococcus* PCC 9511 GS enzyme, encoded by the *glnA* gene, is biochemically similar to other cyanobacteria in many respects (El Alaoui et al., 2003). However, the genetic regulation of *Prochlorococcus* glutamine synthetase has been shown to be quite novel. Unlike other cyanobacteria, neither the *Prochlorococcus glnA* gene (Garcia-Fernandez et al., 2004) nor the GS protein (El Alaoui et al., 2001; El Alaoui et al., 2003) is up-regulated in response to nitrogen starvation.

Prochlorococcus strains have discrete transport systems for several forms of nitrogen. Ammonia is transported by the high-affinity transporter, *amt1,* in all *Prochlorococcus* strains. In contrast to other cyanobacteria, *Prochlorococcus* PCC 9511 *amt1* expression is not regulated by ammonium availability and is proposed not to be NtcA-regulated (Lindell et al., 2002). Genome sequencing has revealed that *Prochlorococcus* has putative transporters for additional N sources. *Prochlorococcus* MED4 has transporters for urea, cyanate, and oligopeptides; MIT9313 has transporters for urea, amino acids, oligopeptides, and a nitrite permease (Rocap et al., 2003).

Although many nitrogen metabolism genes in other cyanobacteria are conserved in *Prochlorococcus*, several of the studies described above suggest that N-regulation is fundamentally different in *Prochlorococcus*: *glnA*/GS is not changed in its abundance or activity under N-stress, *amt1* is not induced under N-stress, and PII is not phosphorylated under any tested conditions. These differences in the regulation of *Prochlorococcus* N metabolism genes relative to other cyanobacteria have been proposed as an adaptation to a homogenous, oligotrophic environment (Garcia-Fernandez et al., 2004). In addition, many N-regulated genes in *Prochlorococcus* are yet to be discovered; the function of nearly half of the *Prochlorococcus* genes are unknown (Dufresne et al., 2003; Rocap et al., 2003). Global mRNA expression profiling is an unprecedented opportunity to further explore nitrogen-regulation in this experimental system for microbial ecology of the oceans.

MATERIALS AND METHODS

Cell culture. *Prochlorococcus* cultures were grown at 22°C with a continuous photon flux of either 10 µmol Q m^{-2} s^{-1} (MIT9313) or 50 µmol Q 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. All experiments were done using duplicate cultures. Log phase growth rates are reported both as doubling times and as the specific growth rate µ (day^{-1})

which represents the slope of the $\log_e$ of culture fluorescence versus time.

To examine the MED4 and MIT9313 cellular response to nitrogen stress, 2 liter cultures were grown through three successive 1/10 volume transfers to establish that the growth rate was constant under these conditions. To begin the experiment, the cells were concentrated in mid-log growth by centrifugation (15 minutes, 9000g, 22° C), washed once, and resuspended in Pro99 (+NH_4 medium) or Pro99 medium lacking any supplemented nitrogen (-N medium). Samples were taken at the following time points: 0 hrs, 3 hrs, 6 hrs, 12 hrs, 24 hrs, and 48 hrs, for fluorescence measurements, Fv/Fm, and RNA isolation. Culture fluorescence was measured using a Turner fluorometer (450 nm excitation; 680 nm absorbance). Fv/Fm was quantified using a single turnover fluorometer. Single turnover fluorescence measurements were made using a Background Irradiance Gradient – Single Turnover fluorometer (BIG-STf) to measure the photosynthetic conversion efficiency (Fv/Fm) of PSII (Johnson, 2004). Duplicate aqueous samples were dark acclimated for 15 minutes, after which, single turnover fluorescence induction curves were measured. Photosynthetic parameters (Fv/Fm) were estimated by fitting standard models to data to determine values of Fo (initial fluorescence), Fm (maximal fluorescence) and Fv (Fm-Fo) (Kolber et al., 1998).

To characterize the mRNA expression changes during growth on different N sources, two liter cultures of MIT9313 and MED4 cultures were grown to mid-log phase in Pro99 medium. These cultures were centrifuged and the cells were resuspended in Pro99 medium containing one of the following nitrogen sources: 800 µM ammonia (standard medium), 400 µM urea, 200 µM nitrite, or 800 µM cyanate. Urea was added at 400 µM because it has 2 nitrogen atoms per molecule. Nitrite was added at 200 µM because higher concentrations were found to be toxic to MIT9313. These cultures were monitored until they had reached balanced growth and RNA samples were taken for microarray analysis in mid-log phase.

RNA preparation. Samples were collected for RNA isolation by concentrating 150 mls of culture (15 minutes, 9000g, 22°C), resuspending in 1 ml of RNA storage buffer (200 mM sucrose, 10 mM sodium acetate pH 5.2, 5 mM EDTA) and storing at -80°C. RNA was isolated using the mirvana miRNA isolation kit (Ambion Inc., cat. #1560) according to the manufacturers instructions. Prior to RNA isolation, MIT9313 cells required an initial 60 minute 1 mg ml^{-1} lysozyme incubation at 37°C. DNA was removed using the Turbo DNase treatment (Ambion Inc., cat. # 2238) according to the manufacturers instructions. RNA was then ethanol precipitated and resuspended at a concentration of 100 ng μl^{-1}.

DNA microarray hybridizations. 2 µg total RNA was reverse transcribed, fragmented, and biotin labeled using the Affymetric prokaryotic RNA protocols

(http://www.affymetrix.com/support/technical). The BioArray™ Terminal labeling kit (Enzo. Cat. no. 42630) was used for terminal labeling. Gel shift assays using 1% TBE were included as quality controls to assure that at least 1 µg of cDNA was labeled for each array. We followed the ProkGE-WS2v3 fluidics protocols for microarray hybridization.

Data analysis. Expression summaries for each gene were computed from the probe intensities in Affymetrix .CEL files by RMA normalization using Genespring software (Silicon Genetics Corp.). Because of microarray hybridization problems with the $+NH_4$ samples at t=24 hrs., the -N expression sumaries at this time point were compared to the $+NH_4$ at t=12 hrs. instead. As the gene expression correlations between $+NH_4$ time points were as high as between replicates at a single time point, (Fig. 10, appendix VI) this had a minimal effect on our results. Normalized expression summaries were exported and all subsequent analyses were done using scripts written in Perl and Matlab. These scripts are available upon request.

Putative NtcA binding sites were identified by searching 100 base pairs upstream of the start codon of each gene with a position-specific scoring matrix derived from the nucleotide frequencies in the NtcA binding site alignment in (Herrero et al., 2001) (see Appendix VI, Fig. 6 for a description of the scoring matrix). Upstream regions with NtcA binding matrix scores in the top 5% of all genes represent positive hits. We assessed the significance of the NtcA binding site predictions by comparing the genes with putative NtcA binding sites to those induced in -N conditions at t=3hrs. The predictive capacity of the NtcA scoring matrix was quantified as the probability the observed number of genes up-regulated in -N conditions would putative NtcA binding sites due to chance alone (N = number of genes up-regulated in -N treatment. m = number of -N upregulated genes with putative NtcA binding sites. P_{hit} = fraction of total genes scored as putative NtcA targets (0.05)).

$$\text{probability of } \geq m \text{ genes up in -N with NtcA binding sites due to chance alone} = \sum_{i=m}^{N} \frac{N!}{(N-i)! * i!} * P_{hit}^{\,i} * (1-P_{hit})^{N-i}$$

The $\log_2$-transformed $-N/+NH_4$ expression summaries were clustered using the Matlab implementation of the k-means algorithm (k=30 clusters) to iteratively minimize the sum of the squared euclidian distance from each gene to the mean of the cluster (J) using the following formula (k = number of clusters. n = number of genes in a cluster. x_n = position of gene in expression space. μ_j = position of mean of

cluster in expression space). At the end of each iteration, each gene was assigned to the cluster with the nearest mean.

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} |x_n - \mu_j|^2$$

All genes were clustered and a complete list of the members of each cluster are available in appendix VI.

RESULTS AND DISCUSSION

Growth and physiology. Our experimental stategy for the N-starvation experiments was to compare a time course of log-phase cells ($+NH_4$ treatment) to in creasingly N-starved cells (-N treatment). Chlorophyll fluorescence measurements (Fig. 1A, 1B) over the time course showed that MED4 and MIT9313 cells grew with doubling times of 1.06 days (μ=0.65 day^{-1}) and 3 days (μ=0.23 day^{-1}), respectively. Chlorophyll fluorescence of $+NH_4$ treatments increased logarithmically for the duration of the experiment. The -N treatments decreased precipitously in chlorophyll fluorescence beginning at t=12hrs, supporting that these cells became increasingly nitrogen starved.

Fv/Fm is a biophysical metric for photochemical conversion efficiency (Kolber et al., 1998) with values of ~0.65 indicating a healthy population. Nitrogen starvation leads to the inability to repair and synthesize new proteins. Because photosystem II core proteins (PSII) turnover rapidly (Aro et al., 1993), nitrogen starvation quickly leads to an accumulation of inactive PSII and a decrease in Fv/Fm (Kolber et al., 1988). A decrease in Fv/Fm has been shown to be an indicator of N starvation in *Prochlorococcus* (Steglich et al., 2001). The Fv/Fm in the $+NH_4$ treatments remained constant during the experiment at levels consistent with healthy photosystems (Geider et al., 1993; Geider et al., 1998). In contrast, Fv/Fm in the -N treatments remained stable for the first 12 hours and then decreased (Fig. 1C, 1D).

Together, chlorophyll fluorescence and Fv/Fm are two distinct physiological metrics supporting that the expression profiles of the $+NH_4$ cultures reflect log-phase cells and that the -N treatments became progressively nitrogen starved during the experiment. It is also notable that differences in gene expression between the $+NH_4$ and –N treatments were observed by t=3 hrs. (Fig. 2) while differences in chlorophyll fluorescence and Fv/Fm were not evident until t=12 hrs. (Fig. 1). By t=12 hrs., the

gene expression measurements already indicated global changes in translation and photosynthesis (Fig. 4). Field assays that use gene expression to measure nutrient stress in phytolankton, such as *ntcA* expression to measure N stress in marine cyanobacteria (Lindell and Post, 2001), may be able to detect when cells are mildly stressed while physiological assays require cells to be at an advanced state of starvation.

We also examined differences in global mRNA expression changes during balanced growth in media containing different N sources. To this end, *Prochlorococcus* cells were resuspended in media containing different N sources and transferred until the cultures reached a constant log phase growth rate. MIT9313 cultures grew on ammonia, nitrite, and urea with mean division rates of 3.00 days (μ=0.22 day^{-1}), 3.22 days (μ=0.21 day^{-1}), and 3.12 days (μ=0.22 day^{-1}), respectively. MED4 cultures grew on ammonia, cyanate, and urea with mean division rates of 1.19 days (μ=0.58 day^{-1}), 1.96 days (μ=0.35 day^{-1}) and 1.36 days.(μ=0.51 day^{-1}).

Overview of microarray analysis methods. We analyzed the N-stress expression profiles using three approaches: identification of all genes elevated in expression at the second time point (t=3 hrs) (Fig. 2), interstrain comparison of the expression profiles of individual genes across all time points (Fig. 3), and K-means clustering of expression profiles (Fig. 4). The K-means algorithm was used to find co-expressed genes that may function together to mediate the cellular response to N starvation. K-means clustering of the log2(-N/+N) expression summaries revealed clusters of differentially expressed genes which are shown for MED4 (Fig. 4A) and MIT9313 genes (Fig. 4B).

In addition, we identified genes differentially expressed on alternative N sources (Fig. 5). The expression of a number of genes were changed on alternative N sources relative to ammonia in each strain (Fig. 5). In MIT9313, 26 genes were differentially expressed in nitrite-based medium and 38 genes were in changed in urea-based medium relative to ammonia (Fig. 5A, 5B). Nineteen of the differentially expressed MIT9313 genes were common to both nitrite and urea, suggesting that there is a large overlap in the cellular response to different alternative N sources. Twenty-three MED4 genes were differentially expressed in cyanate medium and 19 genes changed in urea medium (Fig 5C, 5D). Six of the differentially expressed genes were common to both cyanate and urea. In the following sections, we discuss *Prochlorococcus* N-regulation in the context of these N-stress and N-source gene expression results.

The role of NtcA in *Prochlorococcus* N-regulation

1. NtcA controls the initial N-stress response. The genes up-regulated in the -N treatment at the second timepoint (t=3 hrs) for MED4 (Fig. 2A) and MIT9313 (Fig. 2B) comprise the initial response to N stress. Several genes are known to be NtcA targets in other cyanobacteria such as *urtA*, *glnA*, *glnB*, *amt1*, and *nirA*. Others have no known function. We hypothesize that many of these N-responsive genes, both those of known and of unknown function, constitute a *Prochlorococcus* NtcA regulon. We found that 12 of 18 MED4 genes (Fig. 2A) and 8 of 15 MIT9313 genes (Fig. 2B) up-regulated in -N conditions at t=3hrs had putative NtcA binding sites. The probability that this many of -N up-regulated genes would have high-scoring NtcA binding sites due to chance alone is quite low (MED4 p=6e-11; MIT9313 p=9e-4).

The high number of -N up-regulated genes bearing NtcA binding sites supports that binding specificity of *Prochlorococcus* NtcA is similar to other cyanobacteria. The NtcA scoring matrix had a greater statistical capacity to predict -N induced MED4 genes than MIT9313 genes. In addition, the MED4 *ntcA* has a putative upstream binding site while MIT9313 *ntcA* does not (Fig. 2), which was unexpected because NtcA is autoregulatory in other cyanobacteria (Herrero, Muro-Pastor and Flores 2001). It is possible that the relatively lower percentage of -N up-regulated genes in MIT9313 indicates that NtcA plays a lesser role in mediating the response to N stress in this strain. Alternatively, our computational predictions may have been less accurate because of a substitution in the MIT9313 NtcA amino-acid sequence. NtcA activates transcription of its targets by binding directly to their promoters with a conserved helix-turn-helix motif in the carboxy terminus. MIT9313 NtcA has a serine for alanine substitution at position 199 in this helix-turn-helix, whereas the MED4 NtcA motif is the same as in other cyanobacteria. It would be interesting to biochemically determine if this amino acid substitution in MIT9313 NtcA has altered its DNA binding affinity.

2. Differential expression of known *ntcA* targets. *ntcA* was up-regulated in response to N-stress in both strains (Fig. 2A). In addition, we observed other genes elevated in expression in -N at t=3hrs (*glnA, amt1, urtA, nirA*) that are known to be involved in N metabolism and have been shown to be NtcA targets in other cyanobacteria. *glnA* encodes the glutamine synthetase enzyme (GS) which assimilates ammonium by incorporating it into glutamine. The expression of both MIT9313 and MED4 *glnA* genes were elevated upon N starvation (Fig. 3C). *Prochlorococcus glnA* upregulation was unexpected in light of previous studies that have found that its protein levels and protein activity are not changed in response to

N starvation (El Alaoui et al., 2001; El Alaoui et al., 2003; Garcia-Fernandez et al., 2004). MIT9313 *glnA* mRNA levels were no longer elevated by the final time point at t=48 hrs (fig. 3C). Previous studies that found no *glnA* (GS) changes under N starvation may have assayed *glnA* at an advanced state of N starvation where *glnA* expression was no longer up-regulated. Alternatively, *glnA* (GS) may have a dual-level regulation such that the mRNA levels are elevated in response to N-starvation but the protein levels and activity are not.

amt1 encodes a high-affinity ammonium transporter. *amt1* expression is low in the presence of ammonium and enhanced in low N conditions in *Synecocystis* PCC6803 (Montesinos et al., 1998) and *Synechococcus* PCC7942 (Vazquez-Bermudez et al., 2002). In contrast, *amt1* is constitutively expressed under N-deprivation in *Prochlorococcus* PCC9511 (Lindell et al., 2002). *Prochlorococcus* PCC 9511 has been shown to be genetically identical to MED4 in terms of the ITS (Laloui et al., 2002) and rDNA (Rippka et al., 2000). Our results show that *amt1* expression was elevated in -N conditions in both strains (Fig. 3A). Differences in *amt1* expression between MED4 and PCC9511 were unexpected because these strains have identical rDNA sequences. We did, however, find that *amt1* was more greatly up-regulated in MIT9313 than in MED4. Lindell et al., (2002) proposed that *amt1* expression is constitutive in a high light-adapted strain such as PCC 9511 because it lives in the surface waters where levels of recycled N sources such as ammonium are constant. In contrast, MIT9313 ecotypes are most abundant at greater depth. It is yet unknown if the greater range of differential expression of *amt1* in MIT9313 represents an adaptation to variations in ambient ammonium deeper in the water column.

In addition to *amt1*, *Prochlorococcus* has genes encoding transporters for alternative N sources which are NtcA-regulated in other cyanobacteria. MED4 K-means cluster 1 contains the most highly up-regulated genes under N starvation (Fig. 4A). Along with *ntcA* and *glnA*, this cluster contained two genes for the transport of alternative N-sources: *urtA* and cynA (a putative cyanate transporter). *urtA* encodes a sub-unit of an ABC-type urea transporter. Urea is an important N source in many marine environments (DeManche et al., 1973) and both MIT9313 and MED4 have a urea transporter and urease genes. MED4 and MIT9313 *urtA* genes were both up-regulated in response to N-deficiency and have putative NtcA binding sites (Fig. 2). MIT9313 also induced *urtA* expression in urea and nitrite media (Fig. 5A, 5B). Surprisingly, the MED4 *urtA* was not elevated in urea media (Fig. 5D). *Prochlorococcus* PCC 9511 urease activity is independent of the nitrogen source in the medium (Palinska et al., 2000), suggesting that the urease genes lack genetic regulation. It is thus possible that the MED4 urea transporter responds to N-

deficiency but not specifically to urea in the medium.

MED4 also has a putative cyanate transporter/lyase with an upstream NtcA binding site (Fig. 6C). As described above, *cynA* clustered among the most highly-elevated genes under N-starvation (Fig. 4A, cluster 1). In addition, *cynA* and *cynB* were up-regulated in cyanate media (Fig. 6C) supporting that these genes transport cyanate. We believe, however, that cyanate growth experiments are at least partially confounded by the hydrolysis rate of cyanate in aqueous media. The initial hydrolysis of cyanate with pure water has a first order rate constant $k=2.67x10^{-4}$ min^{-1} (Wen and Brooker, 1994) meaning that half the cyanate had hydrolyzed to ammonium within the first two days. RNA samples were taken 7 days after transfer fresh cyanate media (Appendix VI). We thus believe that it is unjustified to conclude that MED4 can grow in cyanate as a 'sole N source' based on culture-based experiments. On the other hand, the mRNA expression profiles support that the putative cyanate transporter is up-regulated under N-stress and in cyanate-based media.

MIT9313 also has nitrite reductase, *nirA*, which is an NtcA target in other cyanobacteria. Gene expression patterns on alternative N sources and the gene organization (Fig. 6C) suggest that the MIT9313 nitrite reductase (*nirA*) is co-expressed along with a nitrite permease and PMM2241, a gene of unknown function. However, these are not typical cyanobacterial nitrite utilization genes. The MIT9313 nitrite permease appears to have been horizontally transferred from protobacteria (Rocap et al., 2003). Further, the MIT9313 *nirA* lacks a putative NtcA binding site (Fig. 2B).

In addition to activating transcription, NtcA may act as a transcriptional repressor of genes such as *rbcL* (Ramasubramanian et al., 1994). The *rbc* genes encode the central carbon-fixing enzyme, Rubisco. The MED4 *rbc* genes clustered among the most repressed genes in the genome (Fig. 4A, cluster 6) whereas the MIT9313 *rbc* genes were not repressed at any time points. This difference in *rbc* gene expression may indicate global differences in the relationship between carbon and nitrogen metabolism in MED4 and MIT9313. It is yet unknown if MED4 *rbc* repression is mediated by NtcA. The *rbc* genes also showed interesting expression patterns on alternative N sources. MIT9313 *rbc*S/L were repressed in nitrite medium, while *rbc*S expression increased in urea medium. This opposing change in *rbc* gene expression may be because urea is a carbon-containing molecule while nitrite is not. If so, MIT9313 may be harvesting carbon in addition to nitrogen from growth on urea.

3. *ntcA* has novel putative targets in *Prochlorococcus*. In addition to genes known to be NtcA targets in other cyanobacteria. We identified genes of unknown function that are up-regulated in -N, have putative NtcA binding sites, and

share genomic proximity to known N metabolism genes. For example, PMM1462 was the second most enhanced MED4 gene in the -N treatment at t=3hrs (Fig. 2A) and remained elevated for the duration of the experiment (Fig. 3B). PMM1462 has no known function but has a putative NtcA binding site and is located directly upstream from *glnB* (Fig. 6A) suggesting it may be functionally related to *glnB*. PMM0374 also has no known function but is adjacent to the cynABDS cluster. Although it is divergently transcribed from cynABD (Fig. 6C), the presence of an NtcA binding site and its proximity to the cynanate transporter suggest that PMM0374 is also involved in N utilization.

The MED4 PMM0958 was most up-regulated gene at all time points. The only BLAST hits to PMM0958 in the NR database are to genes of unknown function in *Prochlorococcus* SS120 and *Synechococcus* WH8102. PMM0958 is not up-regulated in response to P starvation (Maureen Coleman, personal communication) and it has a putative *ntcA* binding site. Similarly, we found highly up-regulated putative *ntcA* targets of unknown function in MIT9313. MIT9313 cluster 2 consists of six genes: *ntcA*, *amt1*, *nirA*, the nitrite permease, *urtA*, and PMT0951 (Fig. 4B). PMT0951 has a putative NtcA binding site but no known function. Because of the high level of induction of these conserved hypothetical genes, mRNA profiling may be useful for identifying novel field indicators of N starvation that are more sensitive than current indicators.

4. N-regulated hli genes are putative NtcA targets. The hli genes represent an extended gene family in *Prochlorococcus*, MED4 has 22 hli genes and MIT9313 has 9 genes (Rocap et al., 2003). We found that hli genes were highly elevated in expression under N -starvation and on different N sources. Cyanobacterial high light-inducible polypeptides (Hli) are a family of genes that have recently been linked to survival under diverse conditions including nitrogen stress (He et al., 2001). Cyanobacterial hli genes were were identified by their similarity to Lhc polypeptides in plants (Dolganov et al., 1995). *Synechocystis* PCC6803 has five genes encoding hli polypeptides, all of which are up-regulated during nitrogen starvation (He et al., 2001). Although the precise mechanism is yet unclear, it has been proposed that hli genes aid in the acclimation of cells to the absorption of excess light energy, perhaps by suppressing reactive oxygen species (He et al., 2001).

Three MED4 *hli* genes (*hli*10, *hli*21, *hli*22) and two MIT9313 *hli* genes (*hli*5 and *hli*7) were up-regulated under N-stress. MED4 K-means cluster 2 contained 19 -N up-regulated genes (Fig 4A), including these three hli genes. Among these three genes, hli10 was the most highly up-regulated and the only one with putative *ntcA* binding

site. In MIT9313, hli5, the glutamine/glutamate tRNA synthetase, and hli7 clustered independently as by far the most up-regulated genes in the genome (approximately 70-fold at t=24hrs) (Fig. 4B, cluster 1) and both have putative *ntcA* binding sites. MIT9313 hli7 and MED4 hli10 are homologs, suggesting a conserved subset of the hli genes have evolved to respond to N stress.

MIT9313 *hli*5, the glutamyl tRNA-synthetase, and *hli*7 are adjacent in the MIT9313 genome (Fig. 6B). Transcript levels of the *Synechococcus* PCC7942 glutamyl tRNA-synthetase increase under nitrogen deficiency and this gene is believed to be *ntcA*-regulated (Luque et al., 2002). This tRNA synthetase charges its cognate tRNA with glutamate or glutamine. As the cell becomes progressively N starved, the intracellular levels of these two amino acids plummet (Mérida et al., 1991). MIT9313 may enhance levels of this tRNA synthetase to more efficiently scavenge glutamate and glutamine to facilitate continued of protein synthesis. It is unclear if there is a direct functional link between the *hli* genes and this tRNA synthetase or if they are simply co-expressed because they are both central to the N-stress response.

In addition, we found that *hli* proteins are differentially expressed on all alternative N sources in both strains (Fig. 5). Five MIT9313 *hli* genes were elevated on alternative N sources along with the tRNA synthetase located between *hli*5 and *hli*7 (Fig. 5A, 5B). These were among the most highly induced genes on alternate N sources. *hli* genes were the largest group of differentially expressed MED4 genes on alternative N sources (Fig. 5C, 5D). Six *hli* genes were induced in cyanate and 5 on urea. *hli*5 was the only *hli* gene up-regulated on both N alternative sources.

The specific role of *hli* genes in nitrogen stress is yet unknown. As *Prochlorococcus* becomes N starved, the photochemical efficiency (Fv/Fm) declines as PSII becomes damaged (Fig. 1C, 1D). Damage to PSII could result in an accumulation of potentially damaging, reactive species in the cell. We propose that a subset of the Hli proteins in *Prochlorococcus* are specialized to avoid damage due to the reactive species that accumulate as a result of N-stress. *Hli* genes are up-regulated on alternative N sources because these sources represent a mild N-stress relative to ammonium. A subset of the hli proteins may have evolved as NtcA targets to ensure that they are rapidly up-regulated in response to nitrogen stress.

The role of *glnB* in the *Prochlorococcus* N-regulation

***glnB* is expressed differently in *Prochlorococcus* strains.** We found striking interstrain differences in the *glnB* expression patterns during N starvation.

MED4 *glnB* expression was highly elevated in -N conditions whereas MIT9313 *glnB* expression was not changed (Fig. 3B). It was unexpected that MIT9313 *glnB* was not induced under N starvation; Synechocystis PCC 6803 *glnB* is an NtcA target (Garcia-Dominguez et al., 2000) whose transcription is enhanced 10-fold under nitrogen deprivation (Garcia-Dominguez et al., 1997).

It is possible that these interstrain differences in *glnB* expression are mediated by differences in the genes upstream of *glnB*. In MIT9313, there are two genes directly upstream of *glnB*: PMT1479 and PMT1480 (Fig. 6A), neither of which have any BLAST hits in the NR database. PMT1479 is the most repressed gene in the genome under N starvation (Fig. 3B) while PMT1480 and *glnB* were not altered in expression (Fig. 3B). MIT9313 *glnB* along with PMT1479 and PMT1480 were repressed to a similar degree in nitrite medium (Fig. 5A) and *glnB* was repressed in urea medium (Fig. 5B). In MED4, PMM1462 is the only gene directly upstream of *glnB* (Fig. 6A). PMM1462 also has no BLAST hits in the NR database. Both PMM1462 and MED4 *glnB* were up-regulated under N starvation (Fig. 2A, Fig. 3B).

These results support two novel findings regarding *Prochlorococcus glnB*. First, *glnB* expression patterns under N starvation differ between MED4 and MIT9313. Interstrain differences in nitrogen regulation are thus manifested even at the level of the central regulators. Second, the genome organization and expression patterns suggest that *glnB* is co-expressed with additional genes. As is shown with the *glnB* gene organization in marine *Synechococcus* (Fig. 6A), this is not generally the case. It would be interesting to know whether these genes upstream of *glnB* in *Prochlorococcus* encode proteins that are direct binding partners of PII.

Given these interstrain differences in *glnB* expression, one might ask "what is the role of PII in N-regulation in *Prochlorococcus*?". Characterization of *glnB* mutants has been used to disentangle the function of *glnB* in other cyanobacteria. For example, *Synechococcus* PCC7942 PII null mutants repress transcription of the nir-*nrtABCD-narB* genes for nitrite/nitrate uptake in the presence of ammonium similar to wild-type cells. Unlike wild-type, these PII mutants uptake nitrite and nitrate in the presence of ammonium (Lee et al., 1999), suggesting that PII acts post-transcriptionally to inhibit nitrite/nitrate uptake. As the cell becomes N-starved, PII binds 2-oxoglutarate (Forchhammer, 1999; Tandeau de Marsac and Lee, 1999) which enhances PII phosphorylation (Forchhammer and Hedler, 1997). Because *Prochlorococcus* PII is not phosphorylated in response to N-deficiency, it was proposed that it has a phosphorylation-independent means of N-regulation, perhaps mediated by the binding an allosteric effector such as 2-oxoglutarate (Forchhammer, 2004). Thus, *glnB* is an NtcA-target that is up-regulated in response to N-stress that

controls the activity of genes for the utlilization of nitrite and nitrate.

Amusingly, MED4 upregulates *glnB* under N-stress but lacks the genes for nitrite/nitrate utilization whereas MIT9313 does not upregulate *glnB* but has genes for nitrite utilization. If PII has a role in MED4 N metabolism, it is evidently independent of nitrite/nitrate utilization. MIT9313 upregulates genes for nitrite utilization under N-starvation (Fig. 2B) and on alternative N sources (Fig. 5A, 5B), but *glnB* is not changed in expression during N-starvation (Fig. 3B) and is actually repressed on alternative N sources (Fig. 5A, 5B). As described above, the MIT9313 nitrite permease appears to be horizontally transferred and the nir operon does not have a putative *ntcA* binding site, suggesting a novel form of regulation. It is, however, still possible that the activity of these proteins is still controlled by PII.

Additional insights into *Prochlorococcus* N-regulation

In addition to the expression changes related to *ntcA* and *glnB* described above, there were a few other gene expression changes worthy of discussion. Sigma factors are sub-units of RNA polymerase that modify its affinity to mediate global transcriptional changes in response to stress. In total, MED4 has 5 and MIT9313 has 7 sigma factors. Each sigma factors is differentiated to alter transcription under specific conditions. The types of conditions for which the sigma factors are specialized can reveal the forces governing *Prochlorococcous* ecology. We observed that two MED4 and two MIT9313 sigma factors were induced upon N starvation (Fig. 2F). MED4 PMM1289 was up-regulated before PMM1697, suggesting that it may be more directly involved in the N stress response. Two MIT9313 sigma factors, PMT0346 and PMT2246 increased in expression. As *Prochlorococcus* expression profiles for different environmental perturbations become available, it will be interesting to see if these sigma factors are nitrogen-specific. We also found that SigA, the principle sigma factor, was repressed on both cyanate and urea suggesting there was a general repression of transcription in alternative N sources.

Another interesting finding relates to the largest cluster of MIT9313 genes differentially expressed on alternative N sources. Subsets of this gene cluster, PMT1570-PMT1577, were repressed on both nitrite and urea (Fig. 5A, 5B). PMT1570 encodes the large subunit of carbamoyl phosphate synthase which is involved in arginine and pyramidine biosynthesis. PMT1573-1576 have significant sequence similarity to the devABC transporter whose transcription is induced under N deficiency and is *ntcA*-regulated in *Anabaena* (Fiedler et al., 2001). Interestingly, the *Anabaena* devABC transporter is proposed to be involved in heterocyst development

as an exporter of heterocyst-specific glycolipids (Fiedler et al., 1998). The *Prochlorococcus* homologs are evidently not involved in heterocyst formation, but appear to have another role related to nitrogen metabolism.

CONCLUSIONS

The majority of genes initially induced in -N conditions have putative *ntcA* binding sites, supporting that NtcA mediates the initial N stress response in *Prochlorococcus. GlnB*, encoding a signal transduction protein that coordinates carbon and nitrogen metabolism in other cyanobacteria, showed different expression patterns in the two *Prochlorococcus* strains here studied. MED4 *glnB* and its putative upstream partner PMM1462 were both elevated under N-deprivation. In constrast, MIT9313 *glnB* and the gene directly upstream, PMT1480, were not changed in expression in -N conditions and were repressed on alternative N sources. PMT1479, the gene upstream of PMT1480, was highly repressed under N deprivation and on alternative N sources. Based on the expression patterns of MIT9313 *glnB* and its putatively co-expressed partners, we propose that MIT9313 *glnB* functions in a novel manner relative to other cyanobacteria.

Prochlorococcus has an extended *hli* gene family, a subset of which appear to be NtcA targets that are N-regulated. The most highly up-regulated MIT9313 genes under ammonium deprivation were three adjacent genes: two *hli* genes and the tRNA synthetase for glutamine/glutamate. The specific cellular role of *hli* genes is yet unknown. They are hypothesized to aid cells in the absorption of excess light energy, perhaps by supressing reactive oxygen species. We propose that a subset of the Hli proteins have evolved to alleviate potentially damaging reactive species that accumulate during N-stress.

Collectively, these results give a portait of how two related strains of a globally abundant marine prokaryote respond to nutrient limitation. During N-starvation, both strains express transporters for ammonium and urea. In addition, each strain expresses an additional transporter that is specific to its ecology: MED4 up-regulates a cyanate transporter and MIT9313 up-regulates a nitrite transporter. These interspecific differences in gene expression during N-stress extend to genes involved in central metabolism such the *rbc* genes and the master regulator *glnB.* Previous studies focusing on rDNA sequences have shown that the *Prochlorococcus* community is composed of many-related strains (Rocap et al., 2002). This study shows that this microdiversity among *Prochlorococcus* strains is also manifested as global differences in gene expression patterns.

REFERENCES

Aro, E. M., Virgin, I. and Andersson, B. (1993). "Photoinhibition of Photosystem II. Inactivation, protein damage and turnover." Biochim Biophys Acta **1143**(2): 113-34.

Bertilsson, S., Berglund, O., Karl, D. M. and Chisholm, S. W. (2003). "Elemental composition of marine *Prochlorococcus* and Synechococcus: Implications for the ecological stoichiometry of the sea." Limnology and Oceanography **48**(5): 1721-1731.

Campbell, L., Landry, M. R., Constantinou, J., Nolla, H. A., Brown, S. L., Liu, H. and Caron, D. A. (1998). "Response of microbial community structure to environmental forcing in the Arabian Sea." Deep-Sea Research Part II-Topical Studies in Oceanography **45**(10-11): 2301-2325.

Capone, D. G. (2000). The marine microbial nitrogen cycle. Microbial Ecology of the Oceans. D. L. Kirchman. New York, Wiley-Liss**:** 455-494.

DeManche, J. M., Curl, H. and Cougheno, D. D. (1973). "Automated Analysis for Urea in Seawater." Limnology and Oceanography **18**(4): 686-689.

Dolganov, N. A., Bhaya, D. and Grossman, A. R. (1995). "Cyanobacterial protein with similarity to the chlorophyll a/b binding proteins of higher plants: evolution and regulation." Proc Natl Acad Sci U S A **92**(2): 636-40.

Dufresne, A., Salanoubat, M., Partensky, F., Artiguenave, F., Axmann, I. M., Barbe, V., Duprat, S., Galperin, M. Y., Koonin, E. V., et al. (2003). "Genome sequence of the cyanobacterium *Prochlorococcus* marinus SS120, a nearly minimal oxyphototrophic genome." Proc Natl Acad Sci U S A **100**(17): 10020-5.

El Alaoui, S., Diez, J., Humanes, L., Toribio, F., Partensky, F. and Garcia-Fernandez, J. M. (2001). "In vivo regulation of glutamine synthetase activity in the marine chlorophyll b-containing cyanobacterium *Prochlorococcus* sp strain PCC 9511 (oxyphotobacteria)." Applied and Environmental Microbiology **67**(5): 2202-2207.

El Alaoui, S., Diez, J., Toribio, F., Gomez-Baena, G., Dufresne, A. and Garcia-Fernandez, J. M. (2003). "Glutamine synthetase from the marine cyanobacteria *Prochlorococcus* spp.: characterization, phylogeny and response to nutrient limitation." Environmental Microbiology **5**(5): 412-423.

Falkowski, P. G., Barber, R. T. and Smetacek, V. V. (1998). "Biogeochemical Controls and Feedbacks on Ocean Primary Production." Science **281**(5374): 200-7.

Fiedler, G., Arnold, M., Hannus, S. and Maldener, I. (1998). "The DevBCA exporter is essential for envelope formation in heterocysts of the cyanobacterium Anabaena sp. strain PCC 7120." Molecular Microbiology **27**(6): 1193-1202.

Fiedler, G., Muro-Pastor, A. M., Flores, E. and Maldener, I. (2001). "*ntcA*-dependent expression of the devBCA operon, encoding a heterocyst-specific ATP-binding Cassette Transporter in Anabaena spp." Journal of Bacteriology **183**(12): 3795-3799.

Forchhammer, K. (1999). The PII protein in *Synechococcus* PMM7942 senses and signals 2-oxoglutarate under ATP-replete conditions. The Phototrophic Prokaryotes. W. Loeffelhardt, G. Schmetterer and G. A. Peschek. New York, Kluwer Academic / Plenum Publishers**:** 463-477.

Forchhammer, K. (2004). "Global carbon/nitrogen control by PII signal transduction in cyanobacteria: from signals to targets." FEMS Microbiol Rev **28**(3): 319-33.

Forchhammer, K. and Hedler, A. (1997). "Phosphoprotein PII from cyanobacteria--analysis of functional conservation with the PII signal-transduction protein from Escherichia coli." Eur J Biochem **244**(3): 869-75.

Garcia-Dominguez, M., Lopez-Maury, L., Florencio, F. J. and Reyes, J. C. (2000). "A gene cluster involved in metal homeostasis in the cyanobacterium Synechocystis sp strain PCC 6803." Journal of Bacteriology **182**(6): 1507-1514.

Garcia-Dominguez, M., Reyes, J. C. and Florencio, F. J. (1997). "Purification and

characterization of a new type of glutamine synthetase from cyanobacteria." Eur J Biochem **244**(1): 258-64.

Garcia-Dominguez, M., Reyes, J. C. and Florencio, F. J. (2000). "*ntcA* represses transcription of gifA and gifB, genes that encode inhibitors of glutamine synthetase type I from Synechocystis sp PCC 6803." Molecular Microbiology **35**(5): 1192-1201.

Garcia-Fernandez, J. M., de Marsac, N. T. and 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**(4): 630-8.

Geider, R. J., Laroche, J., Greene, R. M. and Olaizola, M. (1993). "Response of the Photosynthetic Apparatus of Phaeodactylum-Tricornutum (Bacillariophyceae) to Nitrate, Phosphate, or Iron Starvation." Journal of Phycology **29**(6): 755-766.

Geider, R. J., Macintyre, H. L., Graziano, L. M. and McKay, R. M. L. (1998). "Responses of the photosynthetic apparatus of Dunaliella tertiolecta (Chlorophyceae) to nitrogen and phosphorus limitation." European Journal of Phycology **33**(4): 315-332.

Graziano, L. M., Geider, R. J., Li, W. K. W. and Olaizola, M. (1996). "Nitrogen limitation of North Atlantic phytoplankton: Analysis of physiological condition in nutrient enrichment experiments." Aquatic Microbial Ecology **11**(1): 53-64.

Hagel, P., Gerding, J. J., Fieggen, W. and Bloemendal, H. (1971). "Cyanate formation in solutions of urea. I. Calculation of cyanate concentrations at different temperature and pH." Biochim Biophys Acta **243**(3): 366-73.

He, Q., Dolganov, N., Bjorkman, O. and Grossman, A. R. (2001). "The high light-inducible polypeptides in Synechocystis PCC6803. Expression and function in high light." J Biol Chem **276**(1): 306-14.

Herrero, A., Muro-Pastor, A. M. and Flores, E. (2001). "Nitrogen control in cyanobacteria." Journal of Bacteriology **183**(2): 411-425.

Jiang, F. Y., Wisen, S., Widersten, M., Bergman, B. and Mannervik, B. (2000). "Examination of the transcription factor *ntcA*-binding motif by in vitro selection of DNA sequences from a random library." Journal of Molecular Biology **301**(4): 783-793.

Kolber, Z., Zehr, J. and Falkowski, P. (1988). "Effects of Growth Irradiance and Nitrogen Limitation on Photosynthetic Energy-Conversion in Photosystem-II." Plant Physiology **88**(3): 923-929.

Kolber, Z. S., Prasil, O. and Falkowski, P. G. (1998). "Measurements of variable chlorophyll fluorescence using fast repetition rate techniques: defining methodology and experimental protocols." Biochim Biophys Acta **1367**(1-3): 88-106.

Kudela, R. M. and Dugdale, R. C. (2000). "Nutrient regulation of phytoplankton productivity in Monterey Bay, California." Deep-Sea Research Part II-Topical Studies in Oceanography **47**(5-6): 1023-1053.

Laloui, W., Palinska, K. A., Rippka, R., Partensky, F., Tandeau de Marsac, N., Herdman, M. and Iteman, I. (2002). "Genotyping of axenic and non-axenic isolates of the genus *Prochlorococcus* and the OMF-'Synechococcus' clade by size, sequence analysis or RFLP of the Internal Transcribed Spacer of the ribosomal operon." Microbiology **148**(Pt 2): 453-65.

Lee, H. M., Vazquez-Bermudez, M. F. and de Marsac, N. T. (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." Journal of Bacteriology **181**(9): 2697-2702.

Lindell, D., Erdner, D., Marie, D., Prasil, O., Koblizek, M., Le Gall, F., Rippka, R., Partensky, F., Scanlan, D. J., et al. (2002). "The nitrogen stress response of *Prochlorococcus* strain PCC 9511 (Oxyphotobacteria) involves contrasting

regulation of *ntcA* and *amt1*." Journal of Phycology **38**(6): 1113-1124.
Lindell, D. and Post, a. F. (2001). "Ecological aspects of *ntcA* gene expression and its use as an indicator of the nitrogen status of marine Synechococcus spp." Applied and Environmental Microbiology **67**(8): 3340-3349.
Luque, I., Contreras, A., Zabulon, G., Herrero, A. and Houmard, J. (2002). "Expression of the glutamyl-tRNA synthetase gene from the cyanobacterium Synechococcus sp PCC 7942 depends on nitrogen availability and the global regulator *ntcA*." Molecular Microbiology **46**(4): 1157-1167.
Luque, I., Flores, E. and Herrero, A. (1994). "Molecular Mechanism for the Operation of Nitrogen Control in Cyanobacteria." Embo Journal **13**(12): 2862-2869.
Montesinos, M. L., Muro-Pastor, A. M., Herrero, A. and Flores, E. (1998). "Ammonium/methylammonium permeases of a Cyanobacterium. Identification and analysis of three nitrogen-regulated amt genes in synechocystis sp. PCC 6803." J Biol Chem **273**(47): 31463-70.
Moore, L. R., Post, A. F., Rocap, G. and Chisholm, S. W. (2002). "Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and Synechococcus." Limnology and Oceanography **47**(4): 989-996.
Muro-Pastor, M. I., Reyes, J. C. and Florencio, F. J. (2005). "Ammonium assimilation in cyanobacteria." Photosynthesis Research **83**(2): 135-150.
Mérida, A., Candau, P. and Florencio, F. J. (1991). "Regulation of Glutamine-Synthetase Activity in the Unicellular Cyanobacterium Synechocystis Sp Strain Pcc-6803 by the Nitrogen-Source - Effect of Ammonium." Journal of Bacteriology **173**(13): 4095-4100.
Ninfa, A. J. and Atkinson, M. R. (2000). "PII signal transduction proteins." Trends in Microbiology **8**(4): 172-179.
Olson, R. J. (1981). "^{15}N tracer studies of the primary nitrite maximum." Journal of Marine Research **39**: 1033-1051.
Palenik, B., Brahamsha, B., Larimer, F. W., Land, M., Hauser, L., Chain, P., Lamerdin, J., Regala, W., Allen, E. E., et al. (2003). "The genome of a motile marine Synechococcus." Nature **424**(6952): 1037-42.
Palinska, K. A., Jahns, T., Rippka, R. and Tandeau De Marsac, N. (2000). "*Prochlorococcus* marinus strain PCC 9511, a picoplanktonic cyanobacterium, synthesizes the smallest urease." Microbiology **146 Pt 12**: 3099-107.
Partensky, F., Hess, W. R. and Vaulot, D. (1999). "*Prochlorococcus*, a marine photosynthetic prokaryote of global significance." Microbiology and Molecular Biology Reviews **63**(1): 106-+.
Paz-Yepes, J., Flores, E. and Herrero, A. (2003). "Transcriptional effects of the signal transduction protein PII (*glnB* gene product) on *ntcA*-dependent genes in Synechococcus sp PCC 7942." Febs Letters **543**(1-3): 42-46.
Ramasubramanian, T. S., Wei, T. F. and Golden, J. W. (1994). "2 Anabaena Sp Strain Pcc 7120 DNA-Binding Factors Interact with Vegetative Cell-Specific and Heterocyst-Specific Genes." Journal of Bacteriology **176**(5): 1214-1223.
Redfield, A. C. (1958). "The biological control of chemical factors in the environment." American Scientist **46**: 205-221.
Rippka, R., Coursin, T., Hess, W., Lichtle, C., Scanlan, D. J., Palinska, K. A., Iteman, I., Partensky, F., Houmard, J., et al. (2000). "*Prochlorococcus* marinus Chisholm et al. 1992 subsp. pastoris subsp. nov. strain PCC 9511, the first axenic chlorophyll a2/b2-containing cyanobacterium (Oxyphotobacteria)." Int J Syst Evol Microbiol **50 Pt 5**: 1833-47.
Rocap, G., Distel, D., Waterbury, JB., Chisholm, SW. (2002). "Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences". Applied and Environmental Microbiology **68** (3): 1180-1191.
Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A., Arellano, A., Coleman, M., Hauser, L., et al. (2003). "Genome divergence in two

Prochlorococcus ecotypes reflects oceanic niche differentiation." Nature **424** (6952): 1042-7.

Steglich, C., Behrenfeld, M., Koblizek, M., Claustre, H., Penno, S., Prasil, O., Partensky, F. and Hess, W. R. (2001). "Nitrogen deprivation strongly affects photosystem II but not phycoerythrin level in the divinyl-chlorophyll b-containing cyanobacterium *Prochlorococcus* marinus." Biochim Biophys Acta **1503**(3): 341-9.

Tandeau de Marsac, N. and Lee, H. M. (1999). Regulation of carbon and nitrogen metabolism in the unicellular cyaobacteria *Synechococcus* sp. The Phototrophic Prokaryots. W. Loeffelhardt, G. Schmetterer and G. A. Peschek. New York, Kluwer Academic / Plenum Publishers**:** 463-477.

Vazquez-Bermudez, M. F., Paz-Yepes, J., Herrero, A. and Flores, E. (2002). "The *ntcA*-activated *amt1* gene encodes a permease required for uptake of low concentrations of ammonium in the cyanobacterium Synechococcus sp. PCC 7942." Microbiology **148**(Pt 3): 861-9.

Vega-Palas, M. A., Madueño, F., Herrero, A. and Flores, E. (1990). "Identification and Cloning of a Regulatory Gene for Nitrogen Assimilation in the Cyanobacterium Synechococcus Sp Strain Pcc-7942." Journal of Bacteriology **172**(2): 643-647.

Wen, N. P. and Brooker, M. H. (1994). "Rate Constants for Cyanate Hydrolysis to Urea - a Raman-Study." Canadian Journal of Chemistry-Revue Canadienne De Chimie **72**(4): 1099-1106.

West, N. J., Schonhuber, W. A., Fuller, N. J., Amann, R. I., Rippka, R., Post, A. F. and Scanlan, D. J. (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**(Pt 7): 1731-44.

Wolk, C. P., Thomas, J., Shaffer, P. W., Austin, S. M. and Galonsky, A. (1976). "Pathway of Nitrogen-Metabolism after Fixation of N-13-Labeled Nitrogen Gas by Cyanobacterium, Anabaena-Cylindrica." Journal of Biological Chemistry **251** (16): 5027-5034.

Zubkov, M. V., Fuchs, B. M., Tarran, G. A., Burkill, P. H. and Amann, R. (2003). "High rate of uptake of organic nitrogen compounds by *Prochlorococcus* cyanobacteria as a key to their dominance in oligotrophic oceanic waters." Appl Environ Microbiol **69**(2): 1299-304.

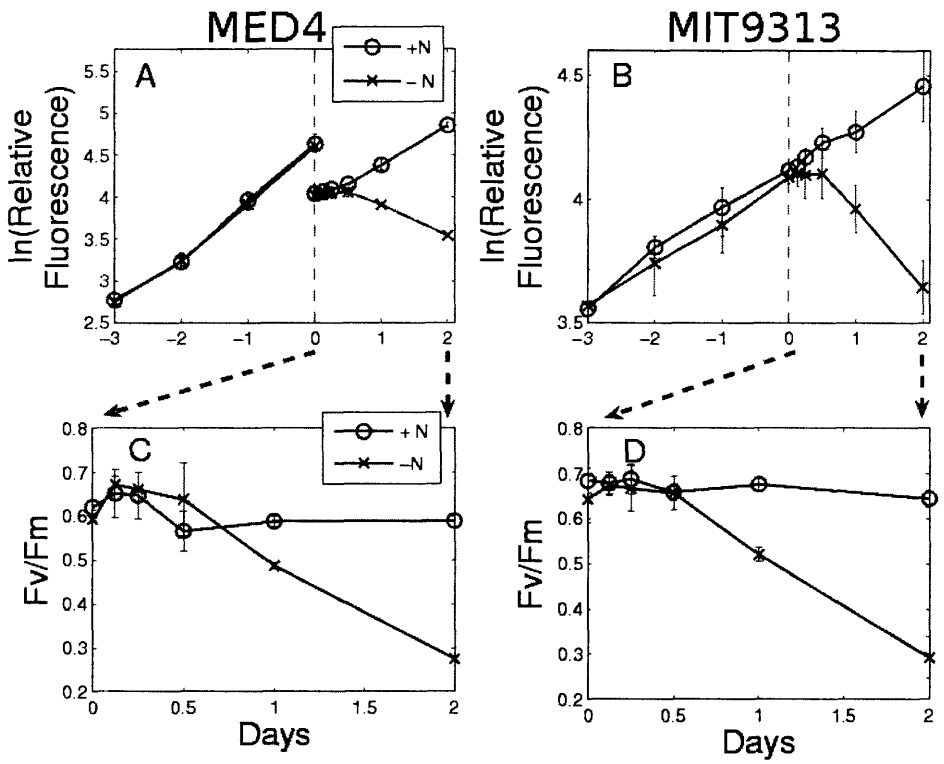
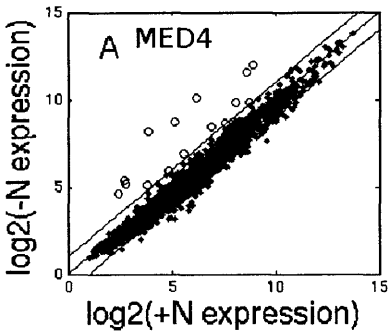
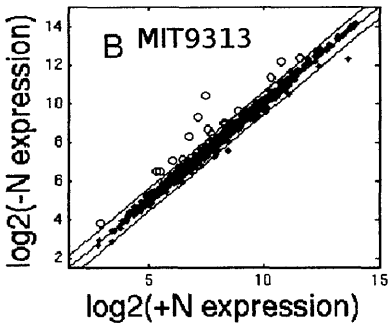


Fig. 1. MED4 (**A**) and MIT9313 (**B**) chlorophyll fluorescence during the experiment support that the cultures were in log phase growth. The vertical dashed line shows the start of the experiment when cultures were transferred to either $_{+NH4}$ media (o) or -N media (*). The discontinuity in MED4 chlorophyll fluorescence at the start of the experiment resulted from a fraction of the cells remaining in the supernatant following centrifugation. MIT9313 cells are larger than MED4 and are thus more efficiently concentrated by centrifugation at speeds not damaging to the cells. Changes in Fv/Fm of MED4 (**C**) and MIT9313 (**D**) during the experiment show that -N cultures (*) became increasingly N starved while $_{+NH4}$ cultures (o) remained N replete. All data points show means of duplicate cultures; error bars show the range.



GENE	FUNCTION	FOLD CHANGE	NtcA SITE
PMM0958	conserved	19.65	X
PMM1462	conserved	16.46	X
PMM0370	cyanate transport	13.04	X
PMM0970	**urtA**	9.23	X
PMM0920	**glnA**	8.64	X
PMM0246	**ntcA**	7.28	X
PMM1041	conserved	5.06	
PMM0374	mttA family	4.46	X
PMM1463	glnB	3.33	X
PMM0687	conserved	3.15	X
PMM1037	conserved	2.75	X
PMM1424	conserved	2.28	
PMM0971	**urtB**	2.24	X
PMM0365	conserved	2.19	
PMM1390	**hli10**	2.13	X
PMM0380	conserved	2.07	
PMM0359	conserved	2.03	
PMM0336	conserved	2.02	



GENE	FUNCTION	FOLD CHANGE	NtcA SITE
PMT2229	**urtA**	7.33	X
PMT1831	**ntcA**	4.33	
PMT2240	nitrite transport	2.65	
PMT1853	amt1	2.61	X
PMT2239	nirA	2.1	
PMT2228	**urtB**	2.1	X
PMT0601	**glnA**	2.01	X
PMT0939	conserved	1.92	
PMT0990	hli5	1.88	X
PMT0929	conserved	1.71	
PMT0951	conserved	1.68	X
PMT0992	**hli7**	1.65	X
PMT1577	conserved	1.62	X
PMT2117	conserved	1.51	
PMT0631	conserved	1.5	

Fig. 2. Comparison of MED4 (**A**) and MIT9313 (**B**) gene expression in -N and +NH_4 media at t = 3 hours. MED4 genes up-regulated >2 fold and MIT9313 genes up-regulated >1.5 fold in -N media are shown as circles. The gene name, function, fold induction, and presence of an *ntcA* binding site for each gene are shown in the tables at right. Gene names shown in bold have homologs which are also induced in the other strain.

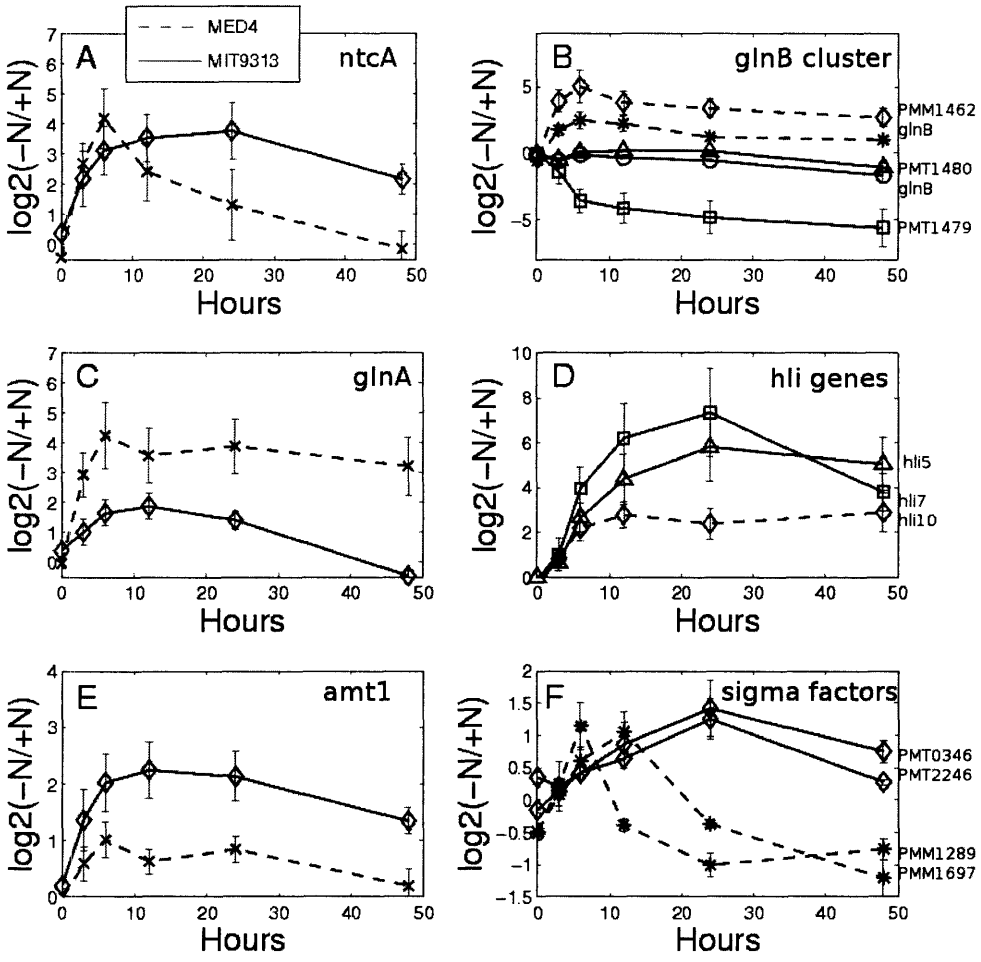


Fig. 3. Comparison of MED4 and MIT9313 expression patterns under NH_4-deprivation. **A.** *ntcA* is up-regulated in both strains. **B**. MED4 *glnB* and the upstream gene PMM1462 are up-regulated. MIT9313 *glnB* is directly downstream of PMT1480 and PMT1479. Expression of *glnB* and PMT1480 were not different between the ±N treatments. The upstream gene, PMT1479 is the most repressed gene in the genome under N stress. **C**. *glnA*, encoding glutamine synthase, is up-regulated in both strains. **D**. hli genes with putative *ntcA* binding sites are up-regulated in both strains. **E**. *amt1*, the ammonium transporter, is induced in MIT9313; MED4 *amt1* is up-regulated, but less than two fold. **F**. Sigma factors induced under N stress. Two MED4 and two MIT9313 sigma factors increased in expression under N stress. Datapoints show log2-transformed mean expression values of duplicate cultures; error bars show one standard deviation of the mean.

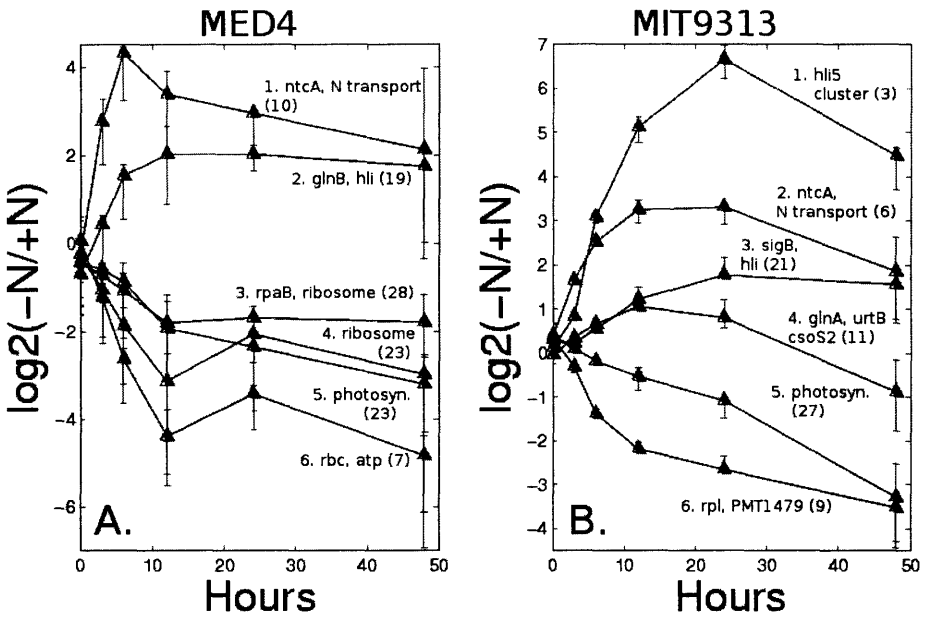


Fig. 4. Expression patterns of differentially expressed k-means cluster for MED4 (**A**) and MIT9313 (**B**). Each datapoint shows the log2-transformed mean expression of all genes in the cluster; bars show range from 25th to 75th percentile. Numbers in parentheses show number of genes in each cluster.

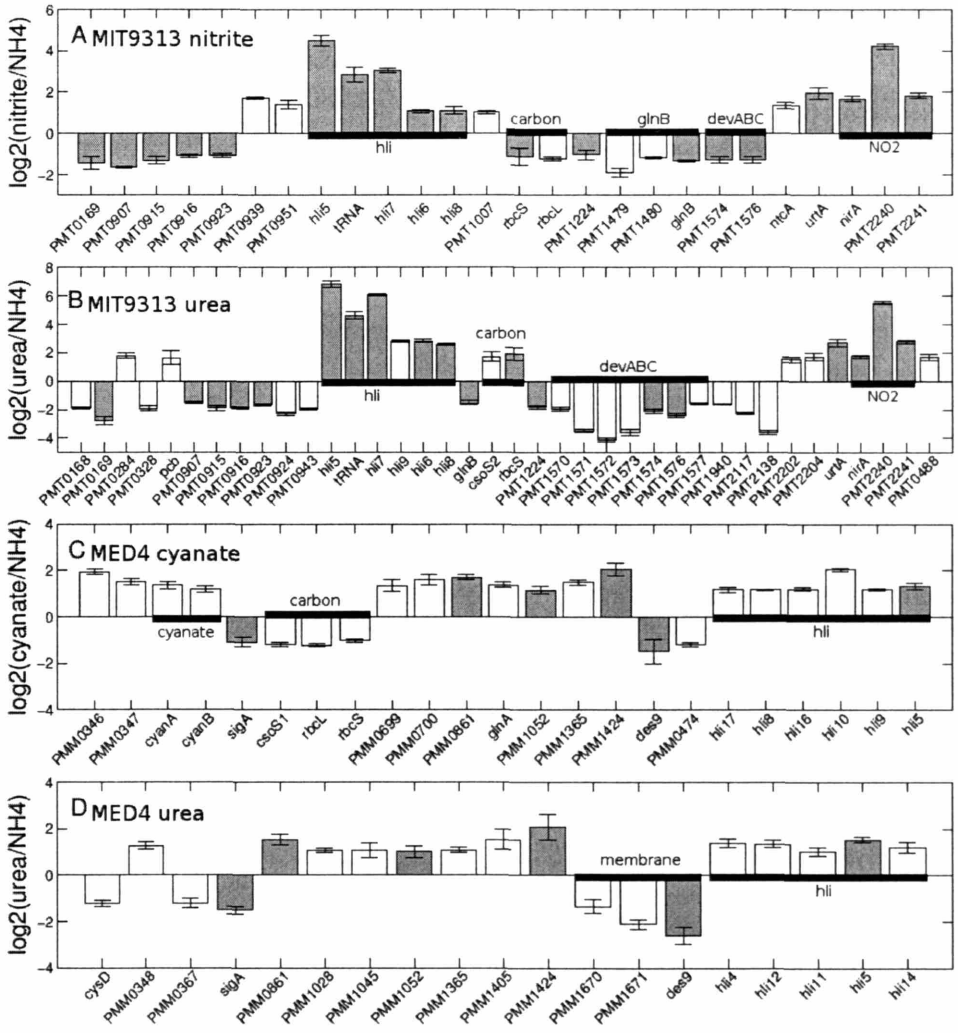


Fig. 5. MIT9313 (**A,B**) and MED4 (**C,D**) differentially expressed genes on alternative N sources relative to ammonia. MIT9313 plots show all genes differentially expressed >1 $\log_2$ unit on nitrite (**A**) or >1.5 $\log_2$ units on urea (**D**) relative to ammonia. MED4 plots show all genes differentially expressed >1 $\log_2$ unit on either cyanate (**C**) or urea (**D**) relative to ammonia. Datapoints show log2-transformed means of duplicate cultures; errorbars show one standard deviation. Colored bars show genes which are differentially expressed on both N-sources for a given strain.

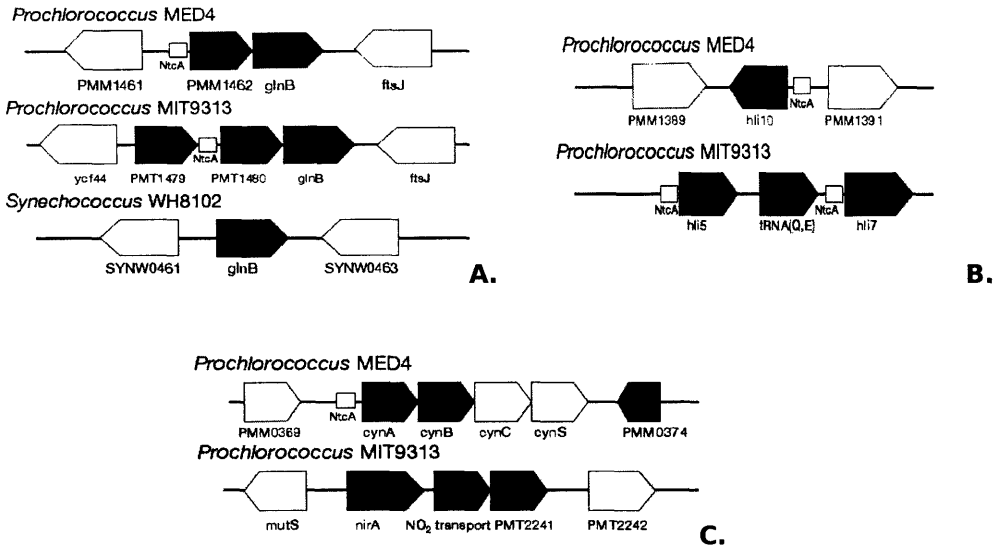


Fig. 6. Gene organization of N-responsive *Prochlorococcus* genes. **A**. Comparison of gene organization surrounding *glnB* in *Prochlorococcus* and marine *Synechococccus*. **B**. N-responsive hli genes in *Prochlorococcus*. **C.** Alternative N transporters. The MED4 cyanate transporters/lyase and the MIT9313 nitrite reductase, transporter. Boxes labelled '*ntcA*' denote putative *ntcA* binding sites. Black genes are differentially expresssed either under N-starvation or on alternative N sources.

Genetic manipulation of *Prochlorococcus* MIT9313: GFP expression on an RSF1010 plasmid and Tn5 transposition

Andrew C. Tolonen[1], Wolfgang R. Hess[2], Gregory B. Liszt[3], and Sallie W. Chisholm[4]

[1]*Dept. of Biology, MIT/WHOI Joint Program in Biological Oceanography, Cambridge, MA USA*
[2]*University Freiburg, Inst. Biology II/Experimental Bioinformatics, Schaenzlestr. 1, D-79014 Frieburg, Germany*
[3]*Dept of Biology, MIT. Cambridge, MA USA*
[4]*Dept of. Civil and Environmental Engineering, and Biology, MIT. Cambridge, MA USA*

ABSTACT

Prochlorococcus is the smallest yet described oxygenic phototroph. It numerically dominates the phytoplankton community in the mid-latitude oceanic basins where it plays an important role in the global carbon cycle. Recently the complete genomes of three *Prochlorococcus* strains have been sequenced (Rocap et al, 2003; Dufresne et al, 2003) and nearly half of the genes in the *Prochlorococcus* genomes are of unknown function. Genetic methods such as reporter gene assays and tagged mutagenesis are critical tools for unveiling the function of these genes. As the basis for such approaches, we describe conditions by which interspecific conjugation with *Escherichia coli* can be used to transfer plasmid DNA into *Prochlorococcus* MIT9313. Following conjugation, *E. coli* were removed from the *Prochlorococcus* cultures by infection with *E. coli* phage T7. We applied these methods to show that an RSF1010-derived plasmid will replicate in *Prochlorococcus MIT9313*. When this plasmid was modified to contain green fluorescent protein (GFP) we detected its expression in *Prochlorococcus* by Western blot and cellular fluorescence. Further, we applied these conjugation methods to show that a mini-Tn5 transposon will transpose *in vivo* in *Prochlorococcus*.

INTRODUCTION

Prochlorococcus, a unicellular, marine cyanobacterium, is distributed worldwide between 40 N and 40 S latitude. Measurements in the Arabian Sea have shown that *Prochlorococcus* can reach densities up to 700,000 cells ml^{-1} of seawater (Campbell et al., 1998) and it is likely the most numerically abundant photosynthetic organism in the oceans (Partensky et al., 1999). Culture-based studies support that *Prochlorococcus* isolates have different light and nutrient physiologies. *Prochlorococcus* isolates can be divided into high-light and low-light adapted strains. High-light adapted strains grow optimally near 200 micromoles photons $m^2 s^{-1}$ and are most abundant in the surface waters; low-light adapted strains such as MIT9313 grow best near 30 micromoles photons $m^2 s^{-1}$ and are most abundant in deeper waters (Moore and Chisholm, 1999). *Prochlorococcus* isolates also differ in their nutrient physiologies. For example, MIT9313 can grow on nitrate as a sole nitrogen source whereas the high-light adapted MED4 cannot (Moore et al, 2002). Molecular phylogenies based upon rDNA sequences correlate with groupings based on light and nutrient physiology (Urbach et al, 1998; Moore et al., 1998).

Many *Prochlorococcus* strains are in culture, but only three (MED4, MIT9313, and MIT9312) have been rendered free of contaminants and are thus suitable for genetic studies. The initial goal of this study was to find methods by which foreign DNA could be introduced and expressed in the *Prochlorococcus* cell. To date, we have no evidence for natural competence or susceptibility to electroporation in *Prochlorococcus*. We thus focused on conjugation-based methods because of their high efficiency and insensitivity to species barriers. For example, conjugation has been used to efficiently transfer DNA from *E. coli to* other cyanobacteria (Wolk et al, 1984) including marine *Synechococcus* (Brahamsha, 1996) and these methods have been extended to even transfer DNA to mammalian cells (Waters, 2001). Our initial challenge was to find a means by which conjugation methods could be adapted to *Prochlorococcus*.

We initially focused on the conjugal transfer of plasmids that are expected to replicate autonomously in *Prochlorococcus.* No endogenous plasmids have been isolated from *Prochlorococcus*, but broad host-range plasmids such as RSF1010 derivatives have been shown to replicate in other cyanobacteria (Mermetbouvier et al, 1993). pRL153, an RSF1010 derivative, has been shown to replicate in three strains of a related oceanic cyanobacterium, *Synechococcus* (Brahamsha, 1996). We modified pRL153 to express a variant of Green Fluorescent Protein (GFP) called GFPmut3.1 (Clontech, BD Biosciences) which is optimized for bacterial GFP expression. GFPmut3.1 expression was driven by the synthetic pTRC promoter which has been shown to be active in other cyanobacteria (Nakahira et al, 2004).

We describe conditions by which Tn5 will transpose and integrate into the *Prochlorococcus* chromosome. Transposon mutagenesis has been widely used in other cyanobacteria as a means to randomly inactivate gene function and study processes such as heterocyst formation (Cohen et al, 1998). Recently, Tn5 has been shown to transpose in the marine cyanobacterium *Synechococcus* (McCarren and Brahamsha, 2005). In total, these data provide new opportunitites to investigate *Prochlorococcus* genes *in situ* using reporter genes and tagged mutagenesis.

MATERIALS AND METHODS

Microbial growth conditions. The microbial stains used in this study are listed in table 1. *Prochlorococcus* MIT9313 was grown at 22°C in Pro99 medium (Moore et al, 1995) with a continuous photon flux of 10 µmols Q m^2 s^{-1} from white fluorescent bulbs. *Prochlorococcus* MIT9313 grew under these conditions with a doubling time of 3.3 days (µ=0.24 $days^{-1}$). Growth of cultures was monitored by chlorophyll

fluorescence using a Turner fluorometer (450 nm excitation; 680 nm absorbance). Chlorophyll measurements were correlated to cell counts by flow cytometry. *Prochlorococcus* was plated in seawater-agarose pour plates (Brahamsha, 1996). The plate medium consisted of Pro99 medium supplemented with 0.5% ultra-pure low melting point agarose (Invitrogen Corp., product 15517-014). *Prochlorococcus* cells were pipetted into the liquid agarose when it had cooled below 28°C. Plates subsequently solidified with cells embedded in the agarose.

E. coli strains were grown in Luria-Bertani (LB) medium supplemented with ampicillin (150 µg ml^{-1}), kanamycin (50 µg ml^{-1}), or tetracycline (15 µg ml^{-1}) as appropriate at at 37 °C. Cultures were continuously shaken except for cultures expressing the RP4 conjugal pilus which were not shaken to minimize the probability of shearing the conjugal pili.

Conjugation. pRL153 was conjugally transferred to *Prochlorococcus* from the *E. coli* host 1100-2 containing the conjugal plasmid pRK24. pRL27 was transferred from the *E. coli* conjugal donor strain BW19851. *E. coli* were mated with *Prochlorococcus* MIT9313 using the following method. A 100 ml culture of the *E. coli* donor strain containing the transfer plasmid was grown to mid-log phase OD_{600} 0.7-0.8. Parallel matings under the same conditions using *E. coli* lacking conjugal capabilites were done to confirm that they were not sufficient for *Prochlorococcus* to become kanamycin-resistant. The *E. coli* cultures were centrifuged three times for 10 minutes at 3000 g to remove antibiotics from the medium. After the first two spins, the cell pellet was resuspended in 15 mls LB medium. After the third spin, the pellet was resuspended in 1 ml Pro99 medium for mating with *Prochlorococcus*.

A 100 ml culture of *Prochlorococcus* MIT9313 was grown to late-log phase (10^8 cell ml^{-1}). The culture was concentrated by centrifugation for 15 minutes at 9000 g and resuspended in 1 ml Pro99 medium. The concentrated *E. coli* and *Prochlorococcus* cells were then mixed at a 1:1 volume ratio and aliquoted as a set of 20 µl spots onto HATF filters (Millipore Corp., product HATF08250) on Pro99 plates containing 0.5% ultra-pure agarose. The plates were then transferred to 10 µmol photons m^2s^{-1} continuous, white light at 22º C for 48 hours to facilitate mating. The cells were resuspended off the filters in Pro99 medium and transferred to 25 ml cultures at an initial cell density of 5 x 10^6 cells ml^{-1}. Kanamycin was added to the cultures after the *Prochlorococcus* cells had recovered from the mating procedure such that the chlorophyll fluorescence of the culture had increased two-fold. 50 µg ml^{-1} kanamycin was added to cultures mated with pRL153 and 25 µg ml^{-1} was added to those mated with pRL27.

Isolation of pure *Prochlorococcus* cultures after conjugation. Once the mated *Prochlorococcus* cultures had grown under kanamycin selection, cells were transferred to pour plates containing 25 µg ml^{-1} kanamycin to isolate colonies. *Prochlorococcus* colonies were excised using a sterile spatula and transferred back to liquid medium containing 50 µg ml^{-1} kanamycin. Once the MIT9313 cultures had reached late log-phase, a 100 µl aliquot of the culture was spread onto LB plates to titer the remaining *E. coli*. Unfortunately, 10^2 to 10^3 *E. coli* cells ml^{-1} often remained viable in the MIT9313 cultures even after isolating MIT9313 colonies on Pro99-agarose plates. To eliminate the remaining *E. coli*, the MIT9313 cultures were infected with *E. coli* phage T7(Demerec and Fano, 1945: Studier, 1969) at a multiplicity of infection (MOI) of 10^6 phage per *E. coli* host. The *E. coli* were again titered on LB plates the following day to show that no viable cells remained.

Plasmid isolation from *Prochlorococcus*. Plasmid DNA from MIT9313 cultures expressing pRL153 was isolated from 5 mls of stationary phase cultures using a Qiagen mini-prep spin column kit. As found by Brahamsha, 1996 with *Synechococcus*, the yield of pRL153 from *Prochlorococcus* was too low to visualize directly by gel electrophoresis. We thus electroporated competent *E. coli* with the plasmids isolated from *Prochlorococcus* in order to compare the structure of pRL153 from MIT9313 to the original plasmid. Following transformation into *E. coli*, pRL153 was isolated from kanamycin resistant *E. coli* transformants and digested with EcoRV and HindIII to compare its structure with the original plasmid. All restriction enzymes used in this study were purchased from New England Biolabs (Beverly, MA. USA) and were used according to the manufacturer's instructions.

pRL153-GFP Plasmid construction. pRL153 was modified to express GFPmut3.1 from the synthetic pTRC promoter to determine if GFP expression could be detected in *Prochlorococcus* (Fig. 1). pRL153 contains unique sites for HindIII and NheI in the Tn5 fragment that are outside the kanamycin resistance gene. pTRC-GFPmut3.1 was cloned into into the unique NheI site to create pRL153-GFP. To this end, pTRC-GFPmut3.1 was PCR amplified from pJRC03 using PFU polymerase (Invitrogen Corp., Carlesbad, CA. USA) using primers with 5' NheI sites: forward primer (pTRC): 5'-acgtac-gctagc-ctgaaatgagctgttgacaatt-3' and reverse primer (GFPmut3.1) 5'-cgtacc-gctagc-ttatttgtatagttcatccatgc-3'. pTRC-GFP PCR product was then NheI digest, CIP-treated, and ligated with NheI-digested pRL153. The ligation was electroporated into *E. coli* and the pTRC-GFP insertion was confirmed by DNA sequencing. GFP expression from pRL153-GFP in *E. coli* was visualized by

epifluorescence microscopy.

Western blot. Total protein extracts from *Prochlorococcus* were made by centrifuging 50 mls of cells, resuspending in 10 mM TrisCl with 0.1% SDS, and boiling at 95°C for 15 minutes. Samples were resolved by SDS-PAGE on a 4-15% Tris-HCl gradient gel (Bio-Rad Corp., Hercules, CA. USA), transferred to nitrocellulose membrane and blocked using 4% nonfat dry milk in PBS with 0.1% Tween-20 (PBS-T). GFP was detected by incubation with rabbit polyclonal anti-GFP (Abcam Corp., Cambridge, UK) antisera diluted 1: 5,000 in PBS-T. Peroxidase-conjugated donkey anti-rabbit IgG secondary antibody (Amersham Biosciences, Piscataway, NJ. USA) was used at a dilution of 1: 10,000. Chemiluminescent detection was achieved by incubation with the ECL reagent (Amersham Biosciences). Blots were stripped for 20 minutes in 50°C stripping buffer (62.5 mM Tris-HCl pH 7.5, 2% SDS, 100 mM beta-mercaptoethanol) and reprobed with polyclonal rabbit antisera specific to *Prochlorococcus* MED4 pcb protein as a loading control.

GFP detection. GFPmut3.1 has maximal excitation and emission wavelengths of 501 nm and 511 nm, respectively. The fluorescence emission spectra of MIT9313 cells expressing pRL153-GFP, and control cells of equal density expressing pRL153, were quantified using a Perkin Elmer Luminescence Spectrometer LS50B. The cells were excited at 490 nm and their cellular fluorescence was measured at 5 nm intervals from 510-700 nm. Cells from duplicate, independently mated +GFP and -GFP MIT9313 cultures were measured. We quantified fluorescence differences between +GFP cells as -GFP cells as mean of the +GFP measurements minus the mean of -GFP measurements.

Identification of transposon insertion sites in *Prochlorococcus*. The Tn5 delivery vector pRL27 carries Tn5 transposase that is expressed from broad host-range tetA promoter from RP4 (Larsen et al., 2002). The transposon itself contains a kanamycin resistance gene as a selectable marker and the origin of replication from plasmid R6K which requires that the pir protein be supplied *in trans* for the plasmid to replicate. Because the transposon contains an origin of replication, transposon insertions could be cloned and sequenced to determine the insertion site in the *Prochlorococcus* genome. Genomic DNA was isolated from Tn5-mated MIT9313 exconjugants using a Qiagen DNeasy Tissue kit (Qiagen Corp., Valencia, CA. USA). 1 µg of genomic DNA was digested with BamHI. The genomic DNA was ethanol

precipitated and religated using T4 DNA ligase (New England Biolabs, Beverly, MA. USA) overnight at 16°C. 20 ng of the ligated DNA was electroporated into *E. coli and* plasmids were isolated from 10 kanamycin-resistant *E. coli* transformants. EcoRI digestion of the plasmids revealed 3 distinct restriction patterns which were sequenced using an outward-facing primer from within the Tn5 casette (aacaagccagggatgtaacg).

RESULTS

pRL153 replication in *Prochlorococcus*. MIT9313 cultures mated with *E. coli* containing the plasmids pRK24 and pRL153 grew under kanamycin selection in liquid culture; control MIT9313 cultures mated with *E. coli* lacking the conjugal plasmid did not grow (Fig. 2), indicating that conjugation with *E. coli* was required for *Prochlorococcus* to become kanamycin resistant. Plating of *Prochlorococcus* is notoriously difficult; plating efficiencies are low and variable and not all strains have been successfully plated at all. We were unable to isolate kanamycin-resistant MIT9313 colonies when cells were plated directly after mating. We were, however, able to get kanamycin-resistant colonies to grow (plating efficiencies of 1 per 100-10,000 cells) after 6 weeks when the cells had grown in liquid medium for one transfer after mating. This suggests that initially growing MIT9313 in liquid after mating may allow the cells to physiologically recover from the mating procedure such that they survive then to form colonies.

We were unable to use standard plating methods to calculate mating efficiencies because we could only isolate *Prochlorococcus* colonies after the cells had first grown in liquid medium after mating. We thus estimated the conjugation efficiency using the following method that assumes that chlorophyll fluorescence correlates with cell counts for log-phase cells. Chlorophyll fluorescence values from the log-phase cells shown in Fig. 2 were correlated to cell abundances using flow cytometry. A linear regression correlating time to the number of transconjugant cells in culture was fit to the data points between days 25 and 55 of Fig. 2: ($R = 0.044*t + 4.82$ where R is $\log_{10}$(tranconjugant cells) and t is days). We calculated the number of transconjugant cells immediately after mating as the intersection of the regression line with the ordinate axis. Using this value, one can calculate the conjugation efficiency to be about 1% by dividing the initial number of transconjugants ($6.9x10^4$ cells) by the number of cells initial transferred into the culture ($6.5x10^6$ cells) .

We found that 10^2 to 10^3 *E. coli* cells ml^{-1} often persisted in the MIT9313 cultures even after the *Prochlorococcus* colonies had been excised from the Pro99-

agarose pour plates and transferred back into the liquid medium. Residual *E. coli* were removed by infecting the cultures with *E. coli* phage T7 at a multiplicity of infection (MOI) of 10^6 phage per host. T7 infection at any MOI resulted in no adverse effects on *Prochlorococcus* viability.

Once we had obtained axenic Prochlorococcus cultures, we examined the structure of pRL153 in *Prochlorococcus*. pRL153 must autonomously replicate in *Prochlorococcus* without suffering structural rearrangements in order to to stably express foreign proteins. We isolated plasmid DNA from MIT9313 cultures to compare the pRL153 structure from MIT9313 to the original plasmid. To this end, *E. coli* was transformed with plasmid DNA isolated from *Prochlorococcus*. We typically obtained approximately 100 *E. coli* transformants when DH5-alpha cells competent to 10^5 transformants μg^{-1} DNA were transformed with one-fifth of a plasmid DNA prep from an MIT9313 culture of $5x10^8$ cells. These efficiencies support that the total plasmid yield was 5 ng of pRL153. Based on the molecular weight of DNA (1bp = 660 daltons), one can calculate that 5 ng of plasmid DNA from $5x10^8$ cells constitutes a plasmid isolation efficiency of 1.06 plasmids per MIT9313 cell. Restriction digestion of the rescued plasmid DNA supports that the gross structure of pRL153 is generally conserved in *Prochlorococcus* (Fig. 3). In total, we examined the digestion patterns of 20 plasmids; 19 of the plasmids appeared identical to the original pRL153. The final plasmid (Fig. 3, lane 3) appears to have acquired an additional DNA segment. We did not further characterize this plasmid. It is most likely that this plasmid rearrangement occurred in either *Prochlorococcus* or in E. coli prior to conjugal transfer. It is, however, also possible that restriction digestion was incapable of cutting this plasmid.

Western blot of GFP protein. The GFP protein was detected in mated *Prochlorococcus* MIT9313 cells by Western blot. MIT9313 cells mated with pRL153-GFP expressed a protein recognized by the GFP antibody at the expected size of 27 kD (Fig. 4A). This band was absent in control preparations from MIT9313 cells lacking pRL-GFP. Blots were stripped and re-probed with an antibody to *Prochlorococcus* MED4 pcb protein to confirm that equal amounts of protein had been loaded in the +GFP and -GFP lanes (Fig. 4B).

GFP expression in *Prochlorococcus*. pRL153 was modified to express GFPmut3.1 from the pTRC promoter. We isolated MIT9313 cultures expressing pRL153-GFP and quantified GFP expression in these cultures (+GFP) by comparing their fluorescence spectra to MIT9313 cells expressing pRL153 (-GFP cells) (Fig. 5). Emission at 680 nm

corresponds to chlorophyll fluorescence. The observation that both the +GFP and -GFP cells had the same emission at 680 nm supports that both treatments had the same overall chlorophyll fluorescence. GFPmut3.1 has a maximum emission at 511 nm. We observed that +GFP cells fluoresced significantly brighter specifically in the wavelengths of GFP emission, supporting that MIT9313 cells containing pRL153-GFP were expressing measurable quantities of GFP.

Tn5 transposition in *Prochlorococcus*. Similar to the matings with pRL153, we found that MIT9313 mated with the *E. coli* conjugal donor strain BW19851 expressing pRL27 became kanamycin resistant. MIT9313 cultures in mock-matings with non-donor *E. coli* expressing pRL27 did not become kanamycin resistant. Because the Tn5 cassette in pRL27 contains an origin of replication, we could clone and sequence the insertion sites of the transposon in the *Prochlorococcus* genome. In total, we isolated 10 plasmids which represented 3 independent genomic insertions, the most common of which is shown in Fig. 6. The insertion shown in Fig. 6 in is in a phage-derived duplication fragment in the gene PMT0236 which encodes a putative serine/threonine protein phosphatase.

DISCUSSION

The primary contribution of this paper is to describe the foundations of a genetic system for *Prochlorococcus*. We found conditions under which an interspecific conjugation system based on the RP4 plasmid family can be used to transfer DNA into *Prochlorococcus* MIT9313. pRL153, an RSF1010-derived plasmid, replicates autonomously in MIT9313 conferring resistance to kanamycin and can be used to express stably foreign proteins such as those for kanamycin-resistance and GFP. In addition, we found that Tn5 will transpose i*n vivo* in *Prochlorococcus*. Once a liquid culture of kanamycin-resistant cells has been isolated, pour plating methods can be used to isolate individual colonies. These colonies can be transferred back to liquid medium for further characterization.

This study is the first report of GFP expression in oceanic cyanobacteria, which has a number of potential applications. For example, one could create transcriptional fusions between *Prochlorococcus* promoters and GFP to study the diel cycling of gene expression in *Prochlorococcus*. Rhythmicity of gene expression is particularly interesting because of results in other cyanobacteria supporting that the expression of all genes cycle daily and are controlled by a central oscillator (Golden, 2003).

Second, GFP expression could provide a means to sort transgenic from non-transgenic cells by flow cytometry. Faced with variable and overall low plating efficiencies, flow sorting cells is an attractive alternative in order to isolate mutants following conjugation. Alternatively, RSF1010-derived plasmids could be modified to cause *Prochlorococcus* to express other foreign proteins. For example, a His-tagged MIT9313 protein could be cloned into pRL153 and transferred into *Prochlorococcus* by conjugation. The ectopically expressed, tagged protein could then be purified to determine which proteins interact with it *in vivo*.

Tn5 transposition provides a means to make tagged mutations in the *Prochlorococcus* chromosome. The Tn5 transposon from pRL27 can be conjugally transferred to *Prochlorococcus* to generate a population of transposon mutants in liquid culture. In this study, we cloned and sequenced 10 Tn insertions and identified 3 independent insertion events. Because the tranconjugant culture represented a mixed-population of transposon mutants, some competitively dominant mutants likely increased in relative abundance and were among those that we identified. These mutants may have been relatively abundant in the culture because they had transposon insertions in selectively-neutral sites in the chromosome such as a phage-derived duplication segment (Fig. 6). Our results suggest that *Prochlorococcus* transconjugants do not survive to form colonies if they are plated directly after mating. It is, however, important to plate the transconjugants as early as possible to avoid certain mutants overtaking the culture, resulting in a low diversity of transposon mutants. The methods described in this study show that genetic methods including transposon mutagenesis are tractable in *Prochlorococcus*, thus providing a foundation for future genetic studies in this ecologically important microbe.

REFERENCES

Bandrin, SV, Rabinovich, PM, Stepanov, AI. (1983). "Three linkage groups of genes involved in riboflavin biosynthesis in *E. coli*." Genetika (Sov. Genet.) 19: 1419-1425.

Brahamsha, B. (1996). "A genetic manipulation system for oceanic cyanobacteria of the genus *Synechococcus*". Appl Environ Microbiol. May;62(5):1747-51.

Campbell, L., Landry, MR., Constantinou, J., Nolla, HA., Brown, SL., Lui, H. (1998). Response of microbial community structure to environmental forcing in the Arabian Sea. Deep Sea Research II 45, 2301-2325.

Cohen, MF., Meeks, JC., Cai, YA., Wolk, CP. (1998). "Transposon mutagenesis of heterocyst-forming filamentous cyanobacteria". Methods in Enzymology 297: 3-17.

Demerec, M. and Fano, U. (1945). "Bacteriophage-resistant mutants in Escherichia coli". Genetics 30: 119-136.

Dufrense A et al, (2003). "Genome sequence of the cyanobacterium *Prochlorococcus* marinus SS120, a nearly minimal oxyphototrophic genome. PNAS, vol. 100. no. 17, 10020-10025.

Elhai, J. and Wolk, CP. (1988). "Conjugal transfer of DNA to cyanobacteria". Methods in Enzymology. 167: 747-754.

Golden, SS. (2003). "Timekeeping in bacteria: the cyanobacterial circadian clock". Current Opinion in Microbiology (6): 535-540.

Koksharova, OA and Wolk, CP. (2002). "Genetic tools for cyanobacteria." Appl Micro and Biotech. 58 (2): 123-137 FEB 2002

Larsen RA, Wilson MM, Guss AM, Metcalf WW. (2002). "Genetic analysis of pigment biosynthesis in Xanthobacter autotrophicus Py2 using a new, highly efficient transposon mutagenesis system that is functional in a wide variety of bacteria." Archives of Microbiology 178 (3): 193-201.

Li, WKW. (1994). Primary productivity of prochlorophytes, cyanobacteria, and eucaryotic ultraphytoplankton: measurements from flow cytometric sorting. Limnology and Oceanography 39, 169-175.

McCarren, J. and Brahamsha, B. (2005). "Transposon mutagenesis in a marine Synechococcus strain: Isolation of swimming motility mutants". Journal of Bacteriology. 4457-4462.

Mermetbouvier, P., Cassierchauvat, C., Marraccini, P., Chauvat, F. (1993). "Transfer and replication of RSF1010 derived plasmids in severl of the general Synechocystis and *Synechococcus*. Current Microbiology. (6): 323-327.

Metcalf, WW., W. Jiang, LL. Daniels, SK. Kim, A. Haldimann, and Wanner. BL. (1996). "Conditionally replicative and conjugative plasmids carrying *lacZ*α for cloning, mutagenesis, and allele replacement in bacteria". Plasmid 35:1-13. Highwire ID="182:6:1671:27".

Moore, LR., Goericke, R., Chisholm, SW. (1995). "Comparative physiology of *Synechococcus* and *Prochlorococcus* – Influence of light and temperature on growth, pigments, fluorescence, and absorptive properties". Marine Ecology Progress Series 16 (1-3): 259-275.

Nakahira, Y., Katayama, M., Miyashita, H., Kutsuna, S., Iwasaki, H., Oyama, T., Kondo, T. (2004). "Global gene repression by KaiC as a master processor of prokaryotic circadian system". PNAS 101 (3): 881-885.

Palenik, B. (2001). "Chromatic adaptation in marine *Synechococcus* strains". Applied and Environmental Microbiology. 67: 991-994.

Partensky, F., Hess, WR., Vaulot, D. (1999). *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. Microbiology and Molecular Biology Reviews 63 (1), 106-127.

Rocap, G. et al. (2003). "Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation". Nature 424 (6952): 1042-1047

Scholz,P., Haring,V., Wittmann-Liebold,B., Ashman,K., Bagdasarian,M. and Scherzinger,E. (1989). "Complete nucleotide sequence and gene organization of the broad-host-range plasmid RSF1010". Gene 75 (2), 271-288.

Studier, FW. (1969). "The genetics and physiology of bacteriophage T7". Virology 39: 562-574.

Toledo, G., Palenik, B., and Brahamsha, B. (1999). "Swimming marine *Synechococcus* strains with widely different photosynthetic pigment ratios form a monophyletic group". Applied and Environmental Microbiology. 65: 5247-5251.

Waters, VL. (2001). "Conjugation between bacterial and mammalian cells". Nature, 29, 4. pp 375-376.

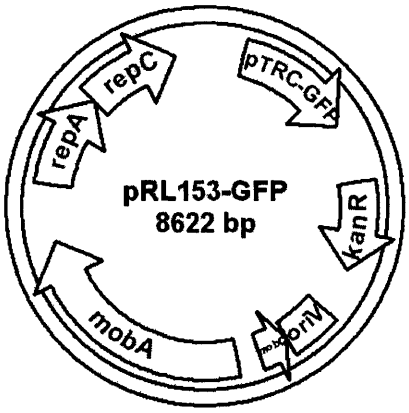


Fig. 1. Diagram of the RSF1010-derived plasmid pRL153 modified to contain pTRC-GFPmut3.1. pRL153 consists of bp 2118-7770 of RSF1010 ligated to bp 680-2516 thereby replacing the sulfonamide resistance gene of RSF1010 with the kanamycin resistance gene of Tn5. pRL153 was then further modified to express GFP mut3.1 from the pTRC promoter.

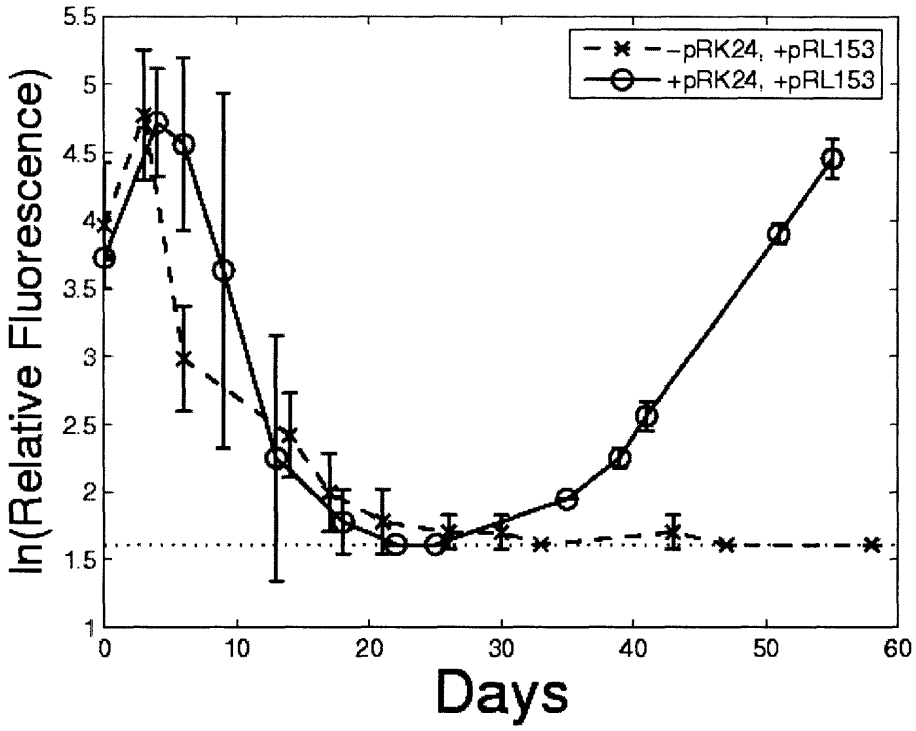
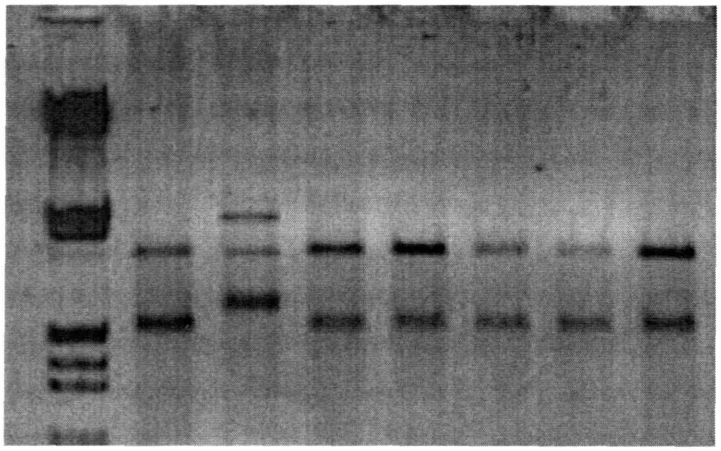


Fig. 2. MIT9313 cultures grown in medium containing 50 µg ml^{-1} kanamycin after mating with *E. coli* containing the conjugal plasmid pRK24 and pRL153 (solid line with diamonds). Control MIT9313 cultures mated with *E. coli* lacking pRK24 (dashed line with stars) did not grow under kanamycin selection. Curves are the average of duplicate cultures, error bars show one standard deviation from the mean. The horizontal dotted line shows the minimum limit of detection of the fluorometer.



1 2 3 4 5 6 7 8

Fig. 3. EcoRV/HindIII digestion of pRL153 plasmids isolated from MIT9313 cultures. Lane 1: EcoRI/HindII digested phage lambda DNA. 2: pRL153 prepared from *E. coli*. 3-8: pRL153 derived from MIT9313 cultures. The digestion pattern in lane 3 shows that the structure of pRL153 is not always retained in MIT9313. However, lanes 4-8 support that the pRL153 structure is generally conserved.

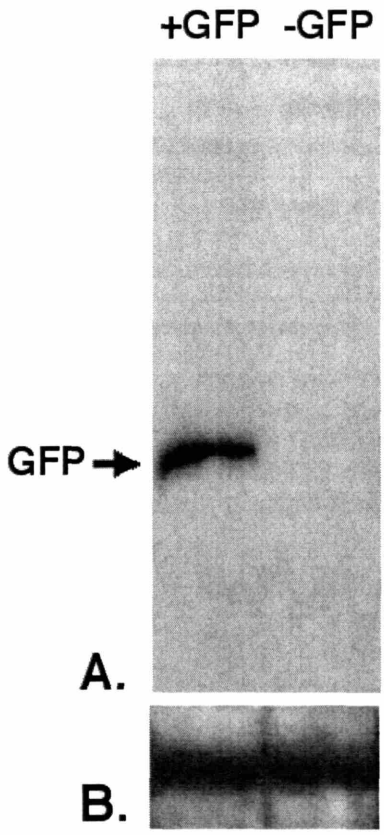


Fig. 4. Western blot comparing *Prochlorococcus* cells expressing GFP (+GFP) to -GFP *Prochlorococcus* controls. **A**. *Prochlorococcus* exconjugants express the GFP protein at the expected size of 27 kD whereas -GFP *Prochlorococcus* cells do not. **B**. To demonstrate that equal amounts of protein had been added in the +GFP and -GFP lanes, the blots were probed with an antibody to the *Prochlorococcus* pcb protein.

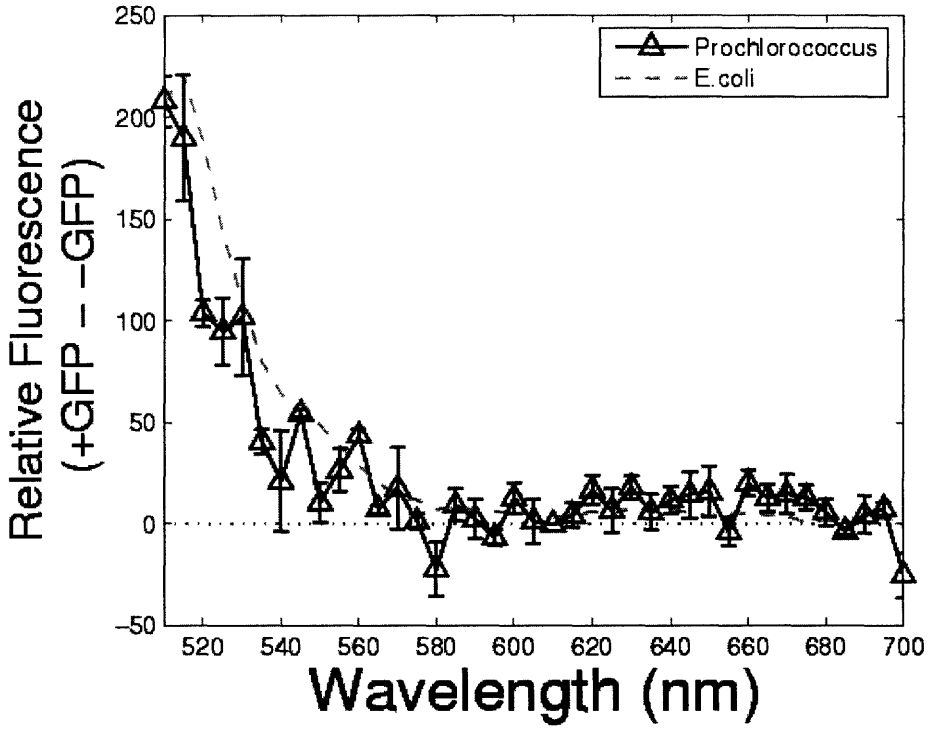


Fig. 5. MIT9313 cells expressing GFPmut3.1 have a higher cellular fluorescence in the GFP emission spectrum (maximum emission 511 nm) than cells lacking GFP. MIT9313 cells expressing pRL153-GFP and control cells lacking GFP were excited at 490 nm and their fluorescence spectrum from 510-700 nm was measured. The fluorescence of +GFP *Prochlorococcus* cells was measured relative to -GFP cells; the mean of duplicate -GFP measurements was subtracted from the mean duplicate +GFP fluorescences. The dashed line shows the relative fluorescence of +GFP to -GFP *E. coli* cells measured by the same method. The horizontal dotted line shows the zero line where the relative fluorescence of +GFP cells is equal to -GFP cells.

```
Tn Insertion    CCCCGAGCTCTTAATTAATTTAAATCTAGAGTCGACCTGCAGGCATGCAAGCTTCAGGGT
pRl27           CCCCGAGCTCTTAATTAATTTAAATCTAGAGTCGACCTGCAGGCATGCAAGCTTCAGGGT
9313 genome     ------------------------------------------------------------


Tn Insertion    TGAGATGTGTATAAGAGACAGCATTTCAGGTTCTAAGGCTTCTGCTTGTTTTCGTTGTTG
pRL27           TGAGATGTGTATAAGAGACAG---------------------------------------
9313 genome     ---------------------CATTTCAGGTTCTAAGGCTTCTGCTTGTTTTCGTTGTTG


Tn Insertion    CTCTTGTTGCCAGATCTCAGTTGCGAGCTGCTCATCCCAAATCTGGTAAGAGATCATGAT
pRL27           ------------------------------------------------------------
9313 genome     CTCTTGTTGCCAGATCTCAGTTGCGAGCTGCTCATCCCAAATCTGGTAAGAGATCATGAT
```

Fig. 6. Alignment of a cloned transposon insertion from MIT9313, the pRL27 plasmid, and the MIT9313 genome. The first 85 bp of the cloned insertion correspond to the transposon cassette from pRL27 and the following sequence shows the point of insertion of the transposon into the MIT9313 genome at bp 271,016 into PMT0236 encoding a serine/threonine protein phosphatase.

Strains	***Description***	***Source***
E. coli		
1100-2	conjugal donor	Bandrin et al, 1983 (obtained from Yale *E. coli* stock center)
DH5-alpha	cloning strain used for all transformations	Invitrogen Corp., Carlsbad, CA.
BW19851	host for pRL27	B. Metcalf, Univ. Illinois
Phage		
E. coli phage T7	phage to kill *E. coli* in Pro99 medium	D. Endy, MIT
Prochlorococcus		
MIT9313	conjugal recipient	Chisholm lab, MIT
Plasmids	**Description**	**Source**
pRL153	RSF1010-derivative	P. Wolk, MSU
pRK24	conjugal plasmid	D. Figurski, Columbia University
pJRC03	pTRC-GFPmut3.1	A. Van Oudenaarden, MIT
pRL27	Tn5 plasmid	B. Metcalf, Univ. Illinois

Table 1. Strains and plasmids used in this study

Nucleic Acids Research, 2002, Vol. 30 No. 20 e107

Optimized *in situ* construction of oligomers on an array surface

Andrew C. Tolonen*, Dinu F. Albeanu[1], Julia F. Corbett[2], Heather Handley, Charlotte Henson[3] and Pratap Malik[2]

Department of Biology, MIT/WHOI Joint Program, Cambridge, MA 02139, USA, [1]Department of Biology, MIT, Cambridge, MA 02139, USA, [2]Harvard University, Cambridge, MA 02115, USA and [3]Whitehead Institute/MIT Center for Genome Research, Cambridge, MA 02139, USA

Received April 8, 2002; Revised July 15, 2002; Accepted July 30, 2002

ABSTRACT

Oligonucleotide arrays are powerful tools to study changes in gene expression for whole genomes. These arrays can be synthesized by adapting photolithographic techniques used in microelectronics. Using this method, oligonucleotides are built base by base directly on the array surface by numerous cycles of photodeprotection and nucleotide addition. In this paper we examine strategies to reduce the number of synthesis cycles required to construct oligonucleotide arrays. By computer modeling oligonucleotide synthesis, we found that the number of required synthesis cycles could be significantly reduced by focusing upon how oligonucleotides are chosen from within genes and upon the order in which nucleotides are deposited on the array. The methods described here could provide a more efficient strategy to produce oligonucleotide arrays.

INTRODUCTION

The advent of genomics has facilitated a shift in molecular biology from studies of the expression of single genes to studies of whole-genome expression profiles. Genome-wide expression profiling is a powerful tool being applied in gene identification, drug discovery, pathological and toxicological mechanisms and clinical diagnosis. By simultaneously measuring the expression of thousands of genes, researchers can get a picture of the transcriptional profile of a whole genome in a given physiological condition. One of the leading technologies for expression profiling is oligo or gene chips. Oligo chips consist of oligonucleotides immobilized upon a support substrate, commonly silica. They have certain advantages over other technologies. Since all of the oligomers can be carefully designed, inter-feature variability is low. Also, oligo chips can be designed to contain several oligonucleotides representing each gene, allowing more quantitative analysis of expression levels.

One of the most successful methods used to make oligonucleotide chips is an adaptation of photolithographic techniques used in microelectronics (http://www.affymetrix.com). Initially, a specific mask is fabricated for each cycle of nucleotide addition that permits light to penetrate only at positions where nucleotides are to be added. A synthesis cycle consists of shining light through the mask onto the chip surface. The positions where light passes through the mask and reaches the chip are activated for synthesis by the removal of a photolabile protective group from the exposed end of the oligonucleotide. Thus, the pattern in which light penetrates the masks directs the base by base synthesis of oligonucleotides on a solid surface (1). After photodeprotection the chip is washed in a solution containing a single nucleotide (A, C, G or T) that binds to oligonucleotides at the deprotected positions. This method results in the *in situ* synthesis of oligonucleotides on an array surface. Light-directed chemical synthesis has been used to produce arrays with as many as 300 000 features (up to 1 000 000 on experimental products) with minimal cross-hybridization or inter-feature variability (2).

When using photolithography to make DNA arrays, the series of masks and the sequence in which nucleotides are added defines the oligonucleotide products and their locations. Because a separate photolithographic mask must be designed for each synthesis cycle it is advantageous to build oligo chips in as few deposition cycles as possible. To this end, we developed an algorithm to reduce the number of cycles required to build an array of oligonucleotides. If the length of the oligomer is N and the number of possible subunits of the oligomer is K, our goal was to build a set of oligomers in as many fewer than $N \times K$ steps as possible. The simplest strategy for the *in situ* synthesis of oligonucleotides upon an array surface is to first add A everywhere it is needed for the first base, then C, G and T. Using this strategy, a set of oligonucleotides of length N can be synthesized in a maximum of $4N$ steps (3). An array of 25mer oligonucleotides thus would take 100 cycles to build.

Our strategy reduced the number of required synthesis cycles by focusing upon two areas of improvement. First, we focused upon how to best select regions of each gene to be used for oligonucleotides. From within each gene we selected oligonucleotides that could be deposited most efficiently. Once the set of oligonucleotides had been selected they could be deposited on the array surface. The second part of our strategy was to determine a deposition order of nucleotide

*To whom correspondence should be addressed. Tel: +1 617 253 8686; Fax: +1 617 253 7475; Email: tolonen@mit.edu

bases on the array surface with a minimum number of steps. We allowed the deposition order to vary so as to add the most common base at each point in the deposition process. During deposition we added bases at every available position and thus allowed oligonucleotides to be built at different rates. Thus, after four cycles, a given oligonucleotide could theoretically have no bases added and another have four bases. By simultaneously optimizing oligonucleotide selection and deposition we significantly reduced the number of deposition cycles required to synthesize an oligonucleotide array.

MATERIALS AND METHODS

Our strategy consists of two basic parts. Initially, we focused upon selecting those oligonucleotides from each gene that could be most efficiently deposited upon the array. Second, we determined an order of oligonucleotide deposition that could efficiently deposit these oligonucleotides. The source code used in modeling is freely available and can be obtained by emailing tolonen@mit.edu.

Oligonucleotide selection

First, we determined a candidate set of unique 25mer oligonucleotides to be deposited on the array. As the input to our program, we arbitrarily selected the second chromosome of *Arabidopsis thaliana* (ftp://ncbi.nlm.nih.gov/genbank/genomes/A_thaliana/CHR_II/). This chromosome is 19.6 Mb and contains 4036 genes. In this paper we modeled the deposition of the first 1000 genes on the chromosome that were >300 bp. However, our strategy could be applied to any number of genes in any genome. For each gene we chose five non-overlapping 25mer oligonucleotides to be deposited on the array. To define the source for each oligonucleotide we parsed the 3′ 300 bp into five 60 bp regions. Thus, each 60 bp region consisted of a total of 35 potential 25mers. We subjected each potential oligonucleotide to a series of simple tests for biological suitability. The tests required that each oligonucleotide be unique in the genome, have a GC content between 25 and 75% and have no region of self-complementarity of five or more bases at either end. In our data set, 2.7% of the 60 bp gene regions contained no suitable oligonucleotides. From the set of oligonucleotides that passed the tests, we then selected one oligonucleotide from each region. Thus, for 1000 genes, we selected a total of 5000 oligonucleotides that were evenly distributed across the 3′ region of each gene.

Modeling oligonucleotide construction

Once we had selected a complete set of oligonucleotides, the next step in our method was to evaluate how many deposition cycles were required to build each oligonucleotide *in situ* on an array surface. Broadly, our deposition strategy was to maximize the number of bases added at each step of the oligonucleotide synthesis. A position was defined as available if it was the next undeposited base in the oligonucleotide sequence. During each deposition cycle, we assumed that a specific base could be added only once at an available position. For example, even if the next two bases to be added to an oligonucleotide were CC, we added only one C at a time.

For each step of oligonucleotide construction, we identified the first available base in each oligonucleotide in the data set. We calculated the frequency of each base at this position and selected the most common base for deposition. This base was deposited for each oligonucleotide in which this base occupied the first position. In each of these oligonucleotides, we then incremented the next available position by one base. One loop of our program was analogous to one cycle of oligonucleotide deposition. The deposition subroutine continued to loop until we had calculated the total number of steps required to synthesize each oligonucleotide.

Optimizing oligonucleotide selection

The goal of this section was to see if selecting alternative oligonucleotides from the same gene region could streamline the deposition process. We investigated two strategies to optimize oligonucleotide selection, iterative re-selection and pooling of candidate oligonucleotides. Our iterative re-selection strategy identified those oligonucleotides that took the most steps to build, replaced them with an equivalent oligonucleotide from the same section of the same gene and tested if the new set of oligonucleotides could be deposited more efficiently. We viewed this process as analogous to an 'oligonucleotide natural selection' to weed out unfit oligonucleotides and replace them with potentially more fit substitutes. After completing an iteration of the deposition process, we knew the number of steps required to deposit each oligonucleotide. We identified the 75th percentile as the number of steps to produce 75% of the oligonucleotides. For example, if 75% of the oligonucleotides were deposited in 50 steps, we focused upon all oligonucleotides that took 51 or more steps to deposit. We then replaced all oligonucleotides above the 75th percentile with alternative oligonucleotides from the same gene region. We replaced oligonucleotides by going back to the input sequence and re-selecting an oligonucleotide that started one position downstream. If that oligonucleotide passed our biological suitability criteria it was used instead of the original oligonucleotide in the next iteration of the deposition process. If the replacement failed our suitability criteria, then we again replaced this oligonucleotide with one from one base downstream. Our goal was to converge upon a set of oligonucleotides that could be most efficiently deposited by repeated oligonucleotide re-selection.

Our second method of oligonucleotide optimization was to initially include all possible 25mer oligonucleotides in the data set passed to the deposition subroutine and then to select the oligonucleotide that is deposited in the fewest steps for each gene region. Thus, all 35 25mers from each gene region were initially included in the data set. When a single oligonucleotide was completed from a given gene region it was selected and the remaining oligonucleotides were deleted from the data set. After completing the deposition subroutine we had selected the oligonucleotide from each 60 bp region that could be deposited in the fewest steps. This method circumvented the need to iterate the oligonucleotide selection process.

RESULTS

Our oligonucleotide selection and deposition strategy demonstrated that oligonucleotides can be synthesized *in situ* upon an array in many fewer than $4N$ steps. In our trial data set, we deposited all oligonucleotides in 83 steps. To further reduce

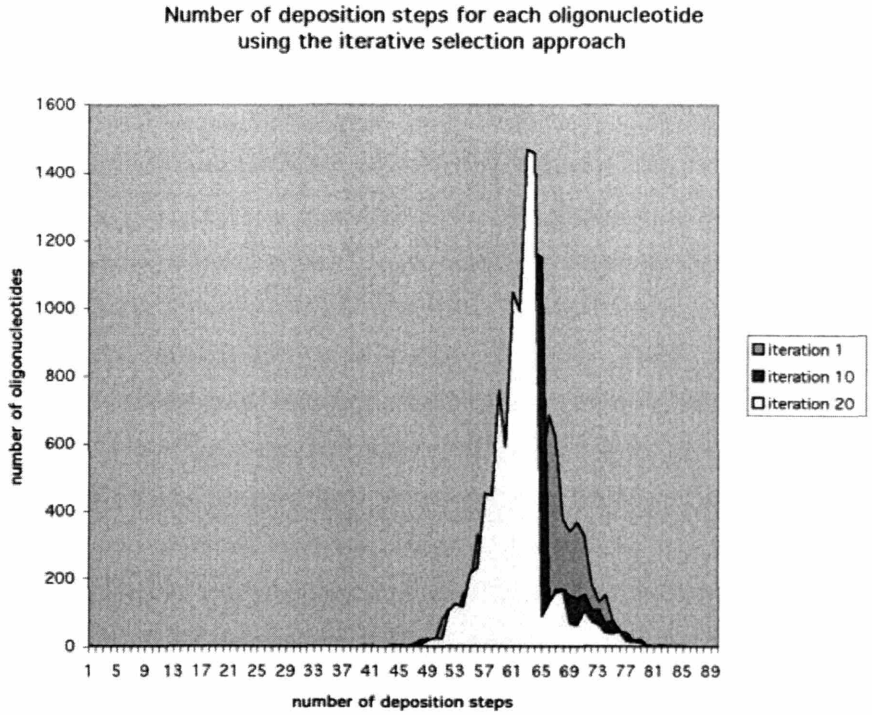


Figure 1. Distribution of the number of steps required to build each oligonucleotide across iterations. Data from iterations 1, 10 and 20 are shown. As the number of iterations increased, the upper tail of the distribution became compressed. However, the number of cycles required to build the entire oligonucleotide set did not decrease.

the number of required steps, we investigated the effect of iterative replacement of the most costly oligonucleotides. We observed that across iterations the distribution became compressed and the mean number of steps decreased (Fig. 1). However, even when the oligonucleotide selection process was iterated 20 times, the number of steps required to complete the deposition process was not reduced. In fact, it increased by two cycles. While in the upper tail the distribution became reduced in size, we were unable to eliminate those oligonucleotides that required the most steps to build from the data set. In light of this result, we identified the gene regions that contained oligonucleotides above the 75th percentile. Because in the upper tail the distribution diminished in successive iterations, the number of oligonucleotides above the 75th percentile became smaller. It became clear that the oligonucleotides above the 75th percentile were coming from the same gene regions across iterations. Figure 2 is a Venn diagram showing that the most costly oligonucleotides came from the same gene regions across iterations. For example, of the 353 oligonucleotides above the 75th percentile in iteration 20, 263 were from the same gene regions represented in iteration 1.

As an alternative means to select more efficient oligonucleotides, we investigated a pooling approach in which the initial data set consisted of all potential oligonucleotides from each gene region. We passed this complete data set to our deposition subroutine and when a single oligonucleotide from a given gene region was completed, it was selected and the remaining oligonucleotides from that gene region were deleted from the data set. We found that this strategy produced significant improvements (Fig. 3). Using this strategy, the entire set of oligonucleotides could be deposited in 73 steps. A summary comparing the results of these two strategies is shown in Table 1.

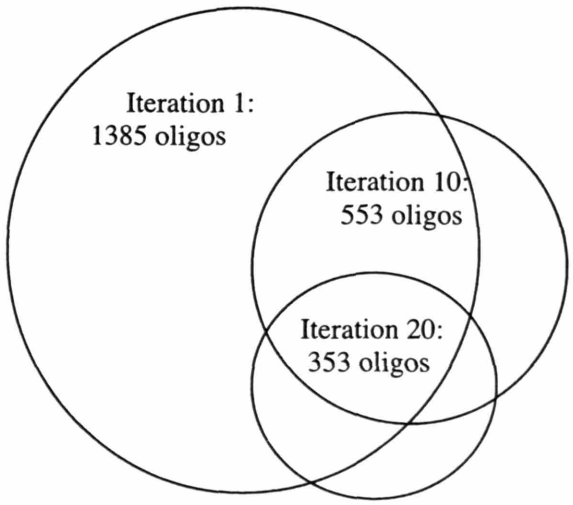


Figure 2. The oligonucleotides requiring the most deposition cycles were from the same gene regions across iterations. This diagram shows overlap in the gene regions that contained oligonucleotides above the 75% percentile. Common oligonucleotides: iterations 1 and 10 share 421 common gene regions; iterations 1 and 20 share 263 gene regions; iterations 10 and 20 share 241 gene regions.

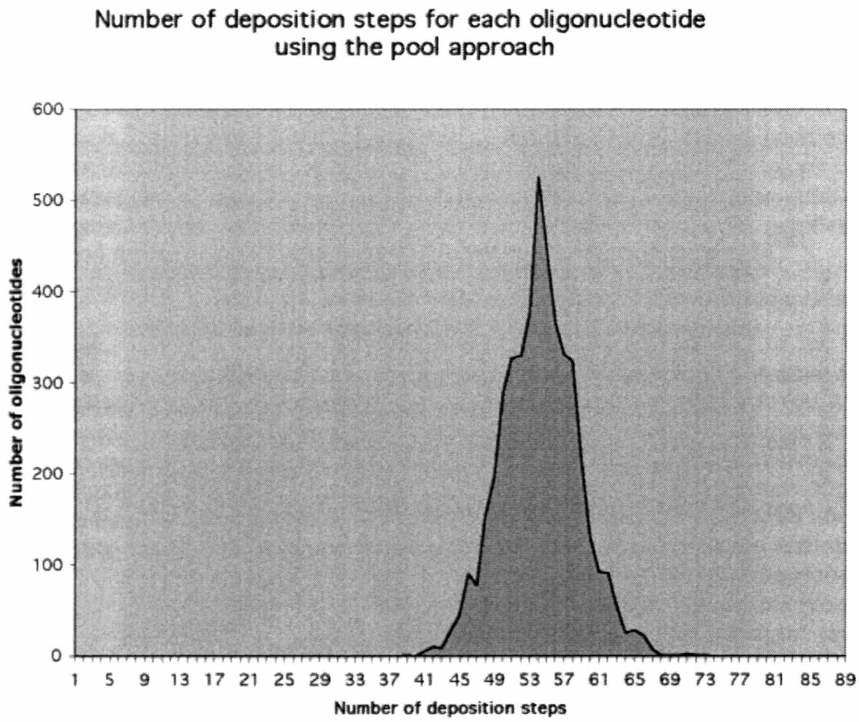


Figure 3. Distribution of the number of steps required to build each oligonucleotide using the oligonucleotide pooling strategy.

Table 1. Summary of the synthesis cycles required to deposit oligonucleotides using the iterative and pooling strategies

Deposition strategy	Median cycles	Maximum cycles
1 iteration	60	83
10 iterations	60	85
20 iterations	59	85
Pool	54	73

Iterative results are shown for the first, tenth and twentieth iterations. For each strategy, the number of cycles required to deposit 50% (median) of oligonucleotides and the number of cycles to deposit all the oligonucleotides (maximum) are shown.

DISCUSSION

Our results demonstrate that both oligonucleotide selection and nucleotide deposition order are important steps towards minimizing the number of steps required to construct oligonucleotides *in situ* upon an array surface. From within a specific gene region, selecting one oligonucleotide versus another can have a significant impact upon the number of deposition steps required. Further, the opportunistic deposition of bases in which the most common next base is added and oligonucleotides may grow at different rates will almost always result in fewer deposition steps than when all oligonucleotides are built at the same rate. Our strategy minimized the number of required deposition steps by attempting to simultaneously optimize oligonucleotide selection and deposition. Because the photolithographic synthesis of oligonucleotides requires expensive reagents and a custom mask for each step of synthesis, our methods could reduce the time and money required to synthesize these arrays.

Our oligonucleotide selection program required that each oligonucleotide pass a set of criteria for biological suitability before it was accepted into the data set. Our criteria included uniqueness in the genome, moderate CG content, no self-complementarity and availability of a unique mismatch oligonucleotide. However, our process of oligonucleotide selection was by no means rigorous. We did not explicitly test whether the melting temperatures of the oligonucleotides were similar. Also, cross-hybridization might be better prevented by searching the genome for regions of significant local alignment rather than perfect matches.

Our deposition strategy of adding the most common base at each position can be thought of as similar to a chess game. At each stage in the game we selected the move that provided the greatest marginal benefit. However, an algorithm that could predict a few steps into the future might be a more optimal deposition solution. It is easy to see that the number of pathways for N steps into the future increases at 4^N and rapidly becomes computationally prohibitive. However, we thought that if we calculated all the possibilities for a few steps ahead that this might yield some improvement. To this end, we tested two look-ahead strategies. First, we calculated all the possibilities for four moves ahead and chose the best path for these four moves. Second, we calculated the best path for the next four steps, executed a single move, and then re-evaluated the next move based upon the next four steps. Unfortunately, neither strategy yielded an improvement.

We found that strategies relating to oligonucleotide selection can result in a more efficient deposition. By replacing all the oligonucleotides above the 75th percentile, we hoped to gradually eliminate the most costly oligonucleotides from the data set. We examined how the distribution of synthesis steps

required for each oligonucleotide changed as the number of iterations increased (Fig. 1). We found that reiteration compressed the distribution and reduced the mean, but it did not reduce the number of cycles needed to deposit the entire data set. We believe that this is due to certain genes that have a small pool of available oligonucleotides. Thus, even if the process is reiterated, costly oligonucleotides from these genes cannot be removed from the data set. In light of these results, we investigated a different strategy in which all the available oligonucleotides were pooled into the initial data set and passed to the deposition subroutine. When a single oligonucleotide from a given gene region was completed, it was selected and the remaining oligonucleotides from that gene region were deleted. We found that this strategy significantly reduced the number of required deposition steps (Fig. 3). Perhaps this is because it is less constrained by those genes with fewer available oligonucleotides.

Our deposition strategy allowed the oligonucleotides to be built at different rates. Thus, at any point in the deposition process the length of an oligonucleotide could be different from that of its neighbors. Hubbell *et al.* (4) wrote that it is usually desirable for the synthesis of adjacent probes to vary in as few synthesis cycles as possible. They explained that an undesirable 'delta edge' is produced when a monomer is added to a synthesis region but not to an adjacent region. To avoid delta edges, it may be important to distribute the oligonucleotides on the chip surface so that adjacent probes are built at similar rates.

With regard to oligonucleotide selection, there might be an unavoidable conflict between choosing oligonucleotides to minimize cross-hybridization and to lower the number of steps required for deposition. Oligonucleotide probes will more efficiently hybridize with only a single mRNA transcript if they represent regions of the genome that are specific to that gene. On the other hand, a set of oligonucleotides can be built in fewer steps if the oligonucleotides are more similar to each other and thus represent areas that are more conserved among genes. In our oligonucleotide selection procedure, we tested to ensure that each oligonucleotide was unique in the genome. However, the re-selection of oligonucleotides likely selected for oligonucleotides that were more similar to the rest of the data set. Thus, our method might result in increased cross-hybridization on the chip.

In conclusion, the optimal set of oligonucleotides can be deposited on an array in a minimum number of steps while retaining the ability to quantify the abundance of each transcript. Our process produces a set of oligonucleotides that can be deposited in many fewer than $4N$ steps. In the future, we would like to explore whether this process builds a chip that can effectively monitor changes in global mRNA expression.

ACKNOWLEDGEMENTS

We thank G. M. Church and S. Kasif for their helpful suggestions. We would also like to thank A. Derti for his insights and programming expertise.

REFERENCES

1. Fodor,S.P., Read,J.L., Pirrung,M.C., Stryer,L., Lu,A.T. and Solas,D. (1991) Light-directed, spatially addressable parallel chemical synthesis. *Science*, **251**, 767–773.
2. Lipshutz,R.J., Fodor,S.P., Gingeras,T.R. and Lockhart,D.J. (1998) High density synthetic oligonucleotide arrays. *Nature Genet.*, **21** (suppl.), 20–24.
3. Chee,M., Yang,R., Hubbell,E., Berno,A., Huang,X.C., Stern,D., Winkler,J., Lockhart,D.J., Morris,M.S. and Fodor,S.P. (1996) Accessing genetic information with high-density DNA arrays. *Science*, **274**, 610–614.
4. Hubbell,E.A., Morris,M.S. and Winkler,J.L. (1999) Computer-aided engineering system for design of sequence arrays and lithographic masks. US patent 5,856,101.

Future Directions

As is so often the case, the experiments described in this thesis probably raise as many questions as they answer. The three chapters explore independent, but related, subject matters. The first chapter focuses on the microarray expression profiling of two *Prochlorococcus* strains in response to changes in ambient nitrogen. The second chapter describes methods for the genetic manipulation of *Prochlorococcus*. The third chapter describes computational approaches to streamline the synthesis of microarrays, such as those used in the first chapter. This discussion outlines a few of the most relevant future experiments that would help resolve some of the yet-unanswered questions relating to the experiments in this thesis.

Nitrogen-regulation of gene expression. Microarrays simultaneously measure the mRNA levels of all the genes in a cell at a specific point in time. The development of *Prochlorococcus* microarrays provided a deluge of mRNA expression data in an organism for which only a few genes had previously been characterized. With microarrays one can compile a list of the complete set of genes that are differentially expressed during a given environmental perturbation. Which of course begs the question "What do all these genes do?". Linking a genes mRNA expression profile to a function is a challenging prospect. First of all, nearly half of the *Prochlorococcus* genes are still annotated simply as 'conserved hypothetical' because they lack sequence similarity to anything in the NCBI database. Even after learning that a conserved hypothetical gene is differentially expressed in a specific condition, it is often difficult to think of an experiment that would elucidate the function of this gene. In addition, many of the laboratory tools used to determine gene function in other organisms are still in their infancy in *Prochlorococcus*. One of the objectives of this thesis was to develop genetic methods for *Prochlorococcus*. Methods for the complementation of mutants of a related organism such as *Synechococcus* PCC 7942 with *Prochlorococcus* proteins would be useful. In addition, the biochemical and high-throughput methods described below will hopefully aid to determine the function of *Prochlorococcus* genes.

The focus of this sub-section is to describe several experiments to further explore N-regulation of *Prochlorococcus* gene expression. We made a few main conclusions from our microarray experiments. First, the majority of genes initially elevated in expression in response to N-stress represent putative targets of the transcriptional factor NtcA; NtcA thus controls the initial *Prochlorococcus* N-stress

response. Second, *glnB* which encodes the PII protein, coordinates N and C metabolism in other cyanobacteria. The *glnB* expression patterns suggest that PII may function fundamentally differently in the two *Prochlorococcus* strains here studied. Third, a subset of the *hli* protein family has evolved to specifically respond to N-stress.

Additional experiments are needed to demonstrate that the genes with *ntcA* binding sites that were elevated in expression in response to N-deficiency are, in fact, *ntcA* targets. We defined the *ntcA* binding site based upon data from other cyanobacteria. The *Prochlorococcus ntcA* binding specificities should be defined. The *Prochlorococcus ntcA* binding specificities could be studied biochemically by *in vitro* selection of oligonucleotides (Jiang et al., 2000) or by DNase footprinting assays. Alternatively, microarrays have recently been adapted to characterize the *in vitro* DNA binding-site sequence specificity of transcription factors with a method called protein-binding microarrays (PBMs) (Mukherjee et al., 2004).

We found that *glnB* expression pattern differed remarkably in response to N-stress in *Prochlorococcus* MED4 and MIT9313. *glnB* is an NtcA-target that is transcriptionally up-regulated in response to N-stress (Garcia-Dominguez et al., 1997). The PII protein, encoded by *glnB*, post-transcriptionally controls the activity of genes for the utlilization of nitrite and nitrate (Lee et al., 1999). We found that MED4 upregulates *glnB* under N-stress but lacks the genes for nitrite/nitrate utilization whereas MIT9313 does not up-regulate *glnB* in response to changes in ambient nitrogen but has genes for nitrite utilization. What is the role of *Prochlorococcus glnB*? If *glnB* is regulates nitrite utilization in MIT9313, why is it not up-regulated on alternative N sources? Further, any role of the MED4 PII protein is evidently independent of nitrite utilization.

The *Prochlorococcus* expression profiles suggest that *glnB* is co-expressed with upstream genes in both strains. These upstream genes could be key to determining the function of PII in *Prochlorococcus*. In MIT9313, there are two genes directly upstream of *glnB*: PMT1479 and PMT1480, neither of which have any BLAST hits in the NR database. PMT1479 is the most repressed gene in the genome under N starvation while PMT1480 and *glnB* were not altered in expression. MIT9313 *glnB* along with PMT1479 and PMT1480 were repressed to a similar degree in nitrite medium and *glnB* was repressed in urea medium. In MED4, PMM1462 is the only gene directly upstream of *glnB*. PMM1462 also has no BLAST hits in the NR database. Both PMM1462 and MED4 *glnB* were upregulated under N starvation. A yeast 2-hybrid screen of *Prochlorococcus* PII could reveal if any of these of these putatively co-expressed genes are direct binding partners of PII. Alternatively, methods under

development in George Church's lab for *in vivo* crosslinking combined with mass spectrometry could be used to determine if any proteins are bound to PII *in vivo*.

Another confounding aspect of the *Prochlorococcus* PII protein is that it is not phosphorlyated in response to nitrogen deprivation (Palinska et al. 2000). PII monitors cellular nitrogen status by binding 2-oxoglutarate (Forchhammer 1999; Tandeau de Marsac and Lee 1999), which, in turn, enhances PII phosphorlyation (Forchhammer and Hedler 1997). Phosphorlylation thus is the mechanism by which PII activity is regulated in other cyanobacteria. If there are conditions under which PII is either phosphorlyated or binds a metabolite such as 2-oxoglutarate, this might shed light on the cellular role of *Prochlorococcus* PII.

Prochlorococcus expression profiling also revealed that a subset of the *hli* gene family is highly upregulated both under N-stress and on alternative N sources. For example, the most highly upregulated MIT9313 genes under N deprivation were three adjacent genes: two *hli* genes and the tRNA synthetase for glutamine/glutamate. Cyanobacterial *hli* genes were identified by their similarity to Lhc polypeptides in plants (Dolganov et al., 1995). Although the precise mechanism is yet unclear, it has been proposed that *hli* genes aid in the acclimation of cells to the absorption of excess light energy, perhaps by suppressing reactive oxygen species (He et al., 2001). We propose that a subset of the *hli* proteins have evolved to alleviate potentially damaging reactive species that accumulate during N-stress. In order to better define the role of *Prochlorococcus hli* proteins, one could localize the proteins in the cells. Are the *hli* proteins directly linked to the photosystems? Are they cytosolic proteins that bind chlorophyll? Traditional methods of protein localization such as GFP-tagging would be time-consuming, albeit possible. Alternatively, if *hli* proteins localize to the membranes, they could be separated in the membrane fraction and probed by Western blot.

A diversity of *Prochlorococcus* microarray experiments are currently in progress and the data they produce will further elucidate the genetic architecture of *Prochlorococcus*. In the future, it will be interesting to integrate the data from multiple microarray experiments and to look for both similarities and differences. For example, which genes are up-regulated under multiple nutrient stresses? These genes are more likely involved in central aspects of metabolism than those genes only elevated under a specific nutrient stress. In addition, future studies will combine data both on the abundances of mRNA and proteins. These studies will shed light on the interconnections between transcriptional and translational control. Is the slow growth rate of *Prochlorococcus* reflected in the sythesis rate of its proteins and the subsequent feedback on transcription? Rapidly growing cells may require forms of

genetic regulation that can respond more quickly to changes in the environment than slow growing cell such as *Prochlorococcus*.

***Prochlorococcus* genetic manipulation**. Chapter two of this thesis describes methods for the genetic manipulation of *Prochlorococcus*. Specifically, we determined how to introduce foreign DNA into *Prochlorococcus* such that foreign proteins such as antibiotic resistance markers, GFP, or a transposase can be expressed in *Prochlorococcus in vivo*. One of the main contributions of these experiments are simply to show that there are no technical barriers to applying the vast array of genetic methods developed for other prokaryotes to *Prochlorococcus*.

At this point, the main barrier to *Prochlorococcus* genetics is the growth rate of this organism. If *E. coli* doubles every 20 minutes and *Prochlorococcus* MIT9313 (the strain used for genetic methods in this thesis) doubles every 3 days, then E. coli doubles 216-times faster than *Prochlorococcus*. The importance of this distinction cannot be overstated. An experiment that requires 1 day in *E. coli* requires 7.1 months in *Prochlorococcus*. The slow rate of growth is certainly not the coup de grâce for *Prochlorococcus* genetics. Genetic studies in *Prochlorococcus* should, however, be confined to processes that are impossible to study in other faster-growing cyanobacteria such as *Synechococcus* PCC7942 and *Synechocystis* PCC6803.

The greatest contribution to facilitate genetic studies of *Prochlorococcus* would thus be the isolation of an axenic, fast-growing strain that yields colonies on plates with high frequency. Three separate approaches could be taken to this end. First, one could attempt to isolate a mutant of one of the current axenic strains. Such a mutant could be isolated either by successive rounds of plating, picking the first colony, and re-plating of the fastest growing cells. Alternatively, chemostats could be used to isolate a fast-growing strain by continually raising the dilution rate. This would select for fast-growing cells by washing out the slow growing member of the population. Alternatively, one could screen the existing culture collection for the strain that grows the fastest both in liquid and on plates. Erik Zinser has begun these experiments with promising preliminary results. He found that *Prochlorococcus* MIT9215 efficiently forms colonies within 1 month when streaked on the surface of plates (Fig. 1); *Prochlorococcus* colonies have never been seen before on the surface of a plate. Finally, one could attempt to isolate a fast-growing *Prochlorococcus* strain from the field by flow sorting *Prochlorococcus* cells away from contaminants and directly plating the sorted cells.

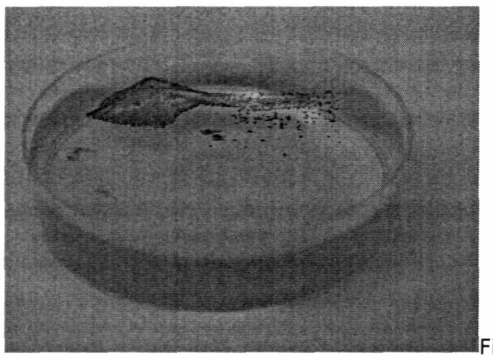


Fig. 1. *Prochlorococcus* MIT9215 colonies growing on the surface of an agarose-seawater plate. Cells were streaked on the surface using standard, microbiological methods (image courtesy of Erik Zinser).

The efficient synthesis of oligonucleotide microarrays. Microarrays are increasingly become standard tools in the molecular biology laboratory. As such, methods to streamline microarray fabrication will be in constant demand. Improvements to microarray fabrication will occur in two areas. First, new hardware-based methods will arise for the efficient fabrication of oligonucleotide microarrays. An example is the use of micro-mirrors to direct oligonucleotide synthesis in lieu of photolithography (Nuwaysir et al., 2002). Second, mathematical optimizations will improve the strategies used to direct the microarray fabrication process. Chapter three of this thesis describes a few such optimization strategies for the efficient *in situ* synthesis of an array of oligonucleotides on a solid surface. With respect to these methods, the most important area of future improvement will be to ensure that improving the efficiency of microarray fabrication does not reduce the ability of the array to detect changes in gene expression. For example, the array that could be most efficiently synthesized would be a set of identical oligonucleotides. This array would, of course, have no means to differentiate among genes. In the future, it is important to explore this trade-off between choosing a set of oligonucleotides that effectively differentiate among genes and a set that can be efficiently synthesized.

REFERENCES

Dolganov, N. A., Bhaya, D. and Grossman, A. R. (1995). "Cyanobacterial protein with similarity to the chlorophyll a/b binding proteins of higher plants: evolution and regulation." Proc Natl Acad Sci U S A **92**(2): 636-40.

Forchhammer, K. and Hedler, A. (1997). "Phosphoprotein PII from cyanobacteria--analysis of functional conservation with the PII signal-transduction protein from Escherichia coli." Eur J Biochem **244**(3): 869-75.

Forchhammer, K. (1999). The PII protein in *Synechococcus* PMM7942 senses and signals 2-oxoglutarate under ATP-replete conditions. The Phototrophic Prokaryots. W. Loeffelhardt, G. Schmetterer and G. A. Peschek. New York,

Kluwer Academic / Plenum Publishers: 463-477.

Garcia-Dominguez, M., Reyes, J. C. and Florencio, F. J. (1997). "Purification and characterization of a new type of glutamine synthetase from cyanobacteria." Eur J Biochem **244**(1): 258-64.

He, Q., Dolganov, N., Bjorkman, O. and Grossman, A. R. (2001). "The high light-inducible polypeptides in Synechocystis PCC6803. Expression and function in high light." J Biol Chem **276**(1): 306-14.

Jiang, F. Y., Wisen, S., Widersten, M., Bergman, B. and Mannervik, B. (2000). "Examination of the transcription factor *ntcA*-binding motif by in vitro selection of DNA sequences from a random library." Journal of Molecular Biology **301**(4): 783-793.

Lee, H. M., Vazquez-Bermudez, M. F. and de Marsac, N. T. (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." Journal of Bacteriology **181**(9): 2697-2702.

Mukherjee S, Berger MF, Jona G, Wang XS, Muzzey D, Snyder M, Young RA, Bulyk ML. (2004). "Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays". Nature Genetics. **36**(12):1331-9

Nuwaysir, E. et al., (2002). "Gene expression analysis using oligonucleotide arrays produced by maskless photolithography". Genome Research. **12**(11): 1749-1755.

Palinska, K. A., Jahns, T., Rippka, R. and Tandeau De Marsac, N. (2000). "*Prochlorococcus* marinus strain PCC 9511, a picoplanktonic cyanobacterium, synthesizes the smallest urease." Microbiology **146 Pt 12**: 3099-107.

Tandeau de Marsac, N. and Lee, H. M. (1999). Regulation of carbon and nitrogen metabolism in the unicellular cyaobacteria *Synechococcus* sp. The Phototrophic Prokaryots. W. Loeffelhardt, G. Schmetterer and G. A. Peschek. New York, Kluwer Academic / Plenum Publishers: 463-477.

4. Scanlan, D. J. & West, N. J. Molecular ecology of the marine cyanobacterial genera *Prochlorococcus* and *Synechococcus*. *FEMS Microbiol. Ecol.* **40,** 1–12 (2002).
5. Ferris, M. J. & Palenik, B. Niche adaptation in ocean cyanobacteria. *Nature* **396,** 226–228 (1998).
6. Toledo, G., Palenik, B. & Brahamsha, B. Swimming strains of marine *Synechococcus* with widely different photosynthetic pigment ratios form a monophyletic group. *Appl. Environ. Microbiol.* **65,** 5247–5251 (1999).
7. Bushman, F. *Lateral DNA Transfer Mechanisms and Consequences* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2002).
8. Brahamsha, B. An abundant cell-surface polypeptide is required for swimming by the nonflagellated marine cyanobacterium *Synechococcus*. *Proc. Natl Acad. Sci. USA* **93,** 6504–6509 (1996).
9. Monger, B. C., Landry, M. R. & Brown, S. L. Feeding selection of heterotrophic marine nanoflagellates based on the surface hydrophobicity of their picoplankton prey. *Limnol. Oceanogr.* **44,** 1917–1927 (1999).
10. Paulsen, I. T., Nguyen, L., Sliwinski, M. K., Rabus, R. & Saier, M. H. Jr Microbial genome analyses: comparative transport capabilities in eighteen prokaryotes. *J. Mol. Biol.* **301,** 75–101 (2000).
11. Willey, J. M. & Waterbury, J. B. Chemotaxis toward nitrogenous compounds by swimming strains of marine *Synechococcus* spp. *Appl. Environ. Microbiol.* **55,** 1888–1894 (1989).
12. Harano, Y. *et al.* Identification and nitrogen regulation of the cyanase gene from the cyanobacteria *Synechocystis* sp. strain PCC6803 and *Synechococcus* sp. strain PCC7942. *J. Bacteriol.* **179,** 5744–5750 (1997).
13. Clark, L. L., Ingall, E. D. & Benner, R. Marine phosphorus is selectively remineralized. *Nature* **393,** 426 (1998).
14. Zehr, J. P. & Ward, B. B. Nitrogen cycling in the ocean: new perspectives on processes and paradigms. *Appl. Environ. Microbiol.* **68,** 1015–1024 (2002).
15. Long, R. A. & Azam, F. Antagonistic interactions among marine pelagic bacteria. *Appl. Environ. Microbiol.* **67,** 4975–4983 (2001).
16. Mann, E. L., Ahlgren, N., Moffett, J. W. & Chisholm, S. W. Copper toxicity and cyanobacteria ecology in the Sargasso Sea. *Limnol. Oceanogr.* **47,** 976–988 (2002).
17. Palenik, B. & Dyhrman, S. T. in *Phosphorus in Plant Biology: Regulatory Roles in Molecular, Cellular, Organismic, and Ecosystem Processes* (eds Lynch, J. P. & Deikman, J.) 26–38 (American Society of Plant Physiologists, Rockville, MD, 1998).
18. Nyyssola, A., Kerovuo, J., Kaukinen, P., von Weymarn, N. & Reinikainen, T. Extreme halophiles synthesize betaine from glycine by methylation. *J. Biol. Chem.* **275,** 22196–22201 (2000).
19. Nomura, M., Ishitani, M., Takabe, T., Rai, A. K. & Takabe, T. *Synechococcus* sp. PCC7942 transformed with *Escherichia coli bet* genes produces glycine betaine from choline and acquires resistance to salt stress. *Plant Physiol.* **107,** 703–708 (1995).
20. Badger, M. R., Hanson, D. & Price, G. D. Evolution and diversity of CO2 concentrating mechanisms in cyanobacteria. *Funct. Plant Biol.* **29,** 161–173 (2002).
21. de Lorimier, R., Guglielmi, G., Bryant, D. A. & Stevens, S. E. J. Structure and mutation of a gene encoding a M, 33 000 phycocyanin-associated linker polypeptide. *Arch. Microbiol.* **153,** 541–549 (1990).
22. Collier, J. L. & Grossman, A. R. A small polypeptide triggers complete degradation of light-harvesting phycobiliproteins in nutrient-deprived cyanobacteria. *EMBO J.* **13,** 1039–1047 (1994).
23. Dolganov, N. & Grossman, A. R. A polypeptide with similarity to phycocyanin alpha-subunit phycocyanobilin lyase involved in degradation of phycobilisomes. *J. Bacteriol.* **181,** 610–617 (1999).
24. Volz, K. in *Two-Component Signal Transduction* (eds Hoch, J. A. & Silhavy, T. J.) 53–64 (American Society for Microbiology, Washington DC, 1995).
25. Goto-Seki, A., Shirokane, M., Masuda, S., Tanaka, K. & Takahashi, H. Specificity crosstalk among group 1 and group 2 sigma factors in the cyanobacterium *Synechococcus* sp PCC7942: *in vitro* specificity and a phylogenetic analysis. *Mol. Microbiol.* **34,** 473–484 (1999).
26. Brahamsha, B. A genetic manipulation system for oceanic cyanobacteria of the genus *Synechococcus*. *Appl. Environ. Microbiol.* **62,** 1747–1751 (1996).
27. Badger, J. H. & Olsen, G. J. CRITICA: Coding region identification tool invoking comparative analysis. *Mol. Biol. Evol.* **16,** 512–524 (1999).
28. Delcher, A. L., Harmon, D., Kasif, S., White, O. & Salzberg, S. L. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* **27,** 4636–4641 (1999).
29. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequences. *Nucleic Acids Res.* **25,** 955–964 (1997).
30. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W:. improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680 (1994).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank the members of the *Prochlorococcus* annotation team: G. Rocap, S. W. Chisholm, D. Lindell, N. Algren, M. Coleman, W. Hess, A. Post, S. Shaw, C. Steglich, C. Ting, M. Sullivan, A. Tolonen, Z. Johnson and E. Zinser. We also thank T. Lane for discussions about carbonic anhydrases. This research was funded by the Biological and Environmental Research Program and the US Department of Energy's Office of Science. Sequencing was carried out and managed at the Joint Genome Institute. Computational analysis was performed at Oak Ridge National Laboratory, managed by UT-BATTELLE for the US Department of Energy. Additional support was provided by a DOE grant to B.P, B.B. and I.P., and an NSF grant to B.B. F.P. and A.D. were supported by the EC program Margenes, and by the Region Bretagne.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.P. (bpalenik@ucsd.edu). The sequence for the chromosome of *Synechococcus* sp. strain WH8102 is deposited in GenBank under accession number BX548020.

Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation

Gabrielle Rocap[1], Frank W. Larimer[2,3], Jane Lamerdin[3], Stephanie Malfatti[3], Patrick Chain[3,4], Nathan A. Ahlgren[1], Andrae Arellano[3], Maureen Coleman[5], Loren Hauser[2,3], Wolfgang R. Hess[9]*, Zackary I. Johnson[5], Miriam Land[2,3], Debbie Lindell[5], Anton F. Post[10], Warren Regala[3], Manesh Shah[2,3], Stephanie L. Shaw[6]*, Claudia Steglich[9], Matthew B. Sullivan[7], Claire S. Ting[8], Andrew Tolonen[7], Eric A. Webb[11], Erik R. Zinser[5] & Sallie W. Chisholm[5,8]

[1]*School of Oceanography, University Of Washington, Seattle, Washington 98195, USA*
[2]*Computational Biology, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA*
[3]*Joint Genome Institute, Walnut Creek, California 94598, USA*
[4]*Lawrence Livermore National Laboratory, Livermore, California 94550, USA*
[5]*Department of Civil and Environmental Engineering,* [6]*Department of Earth, Atmospheric and Planetary Sciences,* [7]*Joint Program in Biological Oceanography, Woods Hole Oceanographic Institution, and* [8]*Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[9]*Institute of Biology, Humboldt-University, D-10115 Berlin, Germany*
[10]*Interuniversity Institute of Marine Science, 88103 Eilat, Israel*
[11]*Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA*

* Present addresses: Ocean Genome Legacy, Beverly, Massachusetts 01915, USA (W.R.H.); Department of Environmental Science Policy and Management, University of California, Berkeley, California 94720, USA (S.L.S.)

The marine unicellular cyanobacterium *Prochlorococcus* is the smallest-known oxygen-evolving autotroph[1]. It numerically dominates the phytoplankton in the tropical and subtropical oceans[2,3], and is responsible for a significant fraction of global photosynthesis. Here we compare the genomes of two *Prochlorococcus* strains that span the largest evolutionary distance within the *Prochlorococcus* lineage[4] and that have different minimum, maximum and optimal light intensities for growth[5]. The high-light-adapted ecotype has the smallest genome (1,657,990 base pairs, 1,716 genes) of any known oxygenic phototroph, whereas the genome of its low-light-adapted counterpart is significantly larger, at 2,410,873 base pairs (2,275 genes). The comparative architectures of these two strains reveal dynamic genomes that are constantly changing in response to myriad selection pressures. Although the two strains have 1,350 genes in common, a significant number are not shared, and these have been differentially retained from the common ancestor, or acquired through duplication or lateral transfer. Some of these genes have obvious roles in determining the relative fitness of the ecotypes in response to key environmental variables, and hence in regulating their distribution and abundance in the oceans.

As an oxyphototroph, *Prochlorococcus* requires only light, CO_2 and inorganic nutrients, thus the opportunities for extensive niche differentiation are not immediately obvious—particularly in view of the high mixing potential in the marine environment (Fig. 1a). Yet co-occurring *Prochlorococcus* cells that differ in their ribosomal DNA sequence by less than 3% have different optimal light intensities for growth[6], pigment contents[7], light-harvesting efficiencies[5], sensitivities to trace metals[8], nitrogen usage abilities[9] and cyanophage specificities[10] (Fig. 1b, c). These 'ecotypes'—distinct genetic lineages with ecologically relevant physiological differences—would be lumped together as a single species on the basis of their rDNA similarity[11], yet they have markedly different distributions within a stratified oceanic water column, with high-

Table 1 **General features of two *Prochlorococcus* genomes**

Genome feature	MED4	MIT9313
Length (bp)	1,657,990	2,410,873
G+C content (%)	30.8	50.7
Protein coding (%)	88	82
Protein coding genes	1,716	2,275
With assigned function	1,134	1,366
Conserved hypothetical	502	709
Hypothetical	80	197
Genes with orthologue in:		
Prochlorococcus MED4	–	1,352
Prochlorococcus MIT9313	1,352	–
Synechococcus WH8102	1,394	1,710
Genes without orthologue in:		
MED4 and WH8102	–	527
MIT9313 and WH8102	284	–
Transfer RNA	37	43
Ribosomal RNA operons	1	2
Other structural RNAs	3	3

light-adapted ecotypes most abundant in surface waters, and their low-light-adapted counterparts dominating deeper waters[12] (Fig. 1a). The detailed comparison between the genomes of two *Prochlorococcus* ecotypes we report here reveals many of the genetic foundations for the observed differences in their physiologies and vertical niche partitioning, and together with the genome of their close relative *Synechococcus*[13], helps to elucidate the key factors that regulate species diversity, and the resulting biogeochemical cycles, in today's oceans.

The genome of *Prochlorococcus* MED4, a high-light-adapted strain, is 1,657,990 base pairs (bp). This is the smallest of any oxygenic phototroph—significantly smaller than that of the low-light-adapted strain MIT9313 (2,410,873 bp; Table 1). The genomes of MED4 and MIT9313 consist of a single circular chromosome (Supplementary Fig. 1), and encode 1,716 and 2,275 genes respectively, roughly 65% of which can be assigned a functional category (Supplementary Fig. 2). Both genomes have undergone numerous large and small-scale rearrangements but they retain conservation of local gene order (Fig. 2). Break points between the orthologous gene clusters are commonly flanked by transfer RNAs, suggesting that these genes serve as loci for rearrangements caused by internal homologous recombination or phage integration events.

The strains have 1,352 genes in common, all but 38 of which are also shared with *Synechococcus* WH8102 (ref. 13). Many of the 38 '*Prochlorococcus* -specific' genes encode proteins involved in the atypical light-harvesting complex of *Prochlorococcus*, which contains divinyl chlorophylls *a* and *b* rather than the phycobilisomes that characterize most cyanobacteria. They include genes encoding the chlorophyll *a*/*b*-binding proteins (*pcb*)[14], a putative chlorophyll *a* oxygenase, which could synthesize (divinyl) chlorophyll *b* from (divinyl) chlorophyll *a*[15], and a lycopene epsilon cyclase involved in the synthesis of alpha carotene[16]. This remarkably low number of 'genera defining' genes illustrates how differences in a few gene families can translate into significant niche differentiation among closely related microbes.

MED4 has 364 genes without an orthologue in MIT9313, whereas MIT9313 has 923 that are not present in MED4. These strain-specific genes, which are dispersed throughout the chromosome (Fig. 2), clearly hold clues about the relative fitness of the two strains under different environmental conditions. Almost half of the 923 MIT9313-specific genes are in fact present in *Synechococcus* WH8102, suggesting that they have been lost from MED4 in the course of genome reduction. Lateral transfer events, perhaps

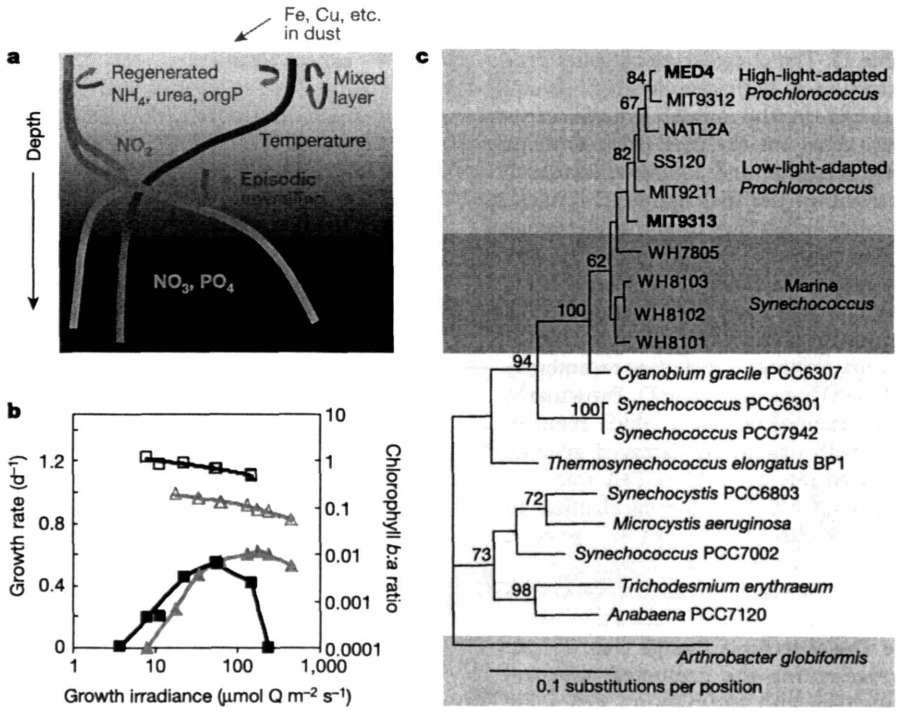


Figure 1 Ecology, physiology and phylogeny of *Prochlorococcus* ecotypes. **a**, Schematic stratified open-ocean water column illustrating vertical gradients allowing niche differentiation. Shading represents degree of light penetration. Temperature and salinity gradients provide a mixing barrier, isolating the low-nutrient/high-light surface layer from the high-nutrient/low-light deep waters. Photosynthesis in surface waters is driven primarily by rapidly regenerated nutrients, punctuated by episodic upwelling. **b**, Growth rate (filled symbols) and chlorophyll *b*:*a* ratio (open symbols) as a function of growth irradiance for MED4 (ref. 7) (green) and MIT9313 (ref. 6) (blue). **c**, Relationships between *Prochlorococcus* and other cyanobacteria inferred using 16S rDNA.

mediated by phage[10], may also be a source of some of the strain-specific genes (Supplementary Figs 3–6).

Gene loss has played a major role in defining the *Prochlorococcus* photosynthetic apparatus. MED4 and MIT9313 are missing many of the genes encoding phycobilisome structural proteins and enzymes involved in phycobilin biosynthesis[15]. Although some of these genes remain, and are functional[17], others seem to be evolving rapidly within the *Prochlorococcus* lineage[18]. Selective genome reduction can also be seen in the photosynthetic reaction centre of *Prochlorococcus*. Light acclimation in cyanobacteria often involves differential expression of multiple, but distinct, copies of genes encoding photosystem II D1 and D2 reaction centre proteins (*psbA* and *psbD* respectively)[19]. However, MED4 has a single *psbA* gene, MIT9313 has two that encode identical photosystem II D1 polypeptides, and both possess only one *psbD* gene, suggesting a diminished ability to photoacclimate. MED4 has also lost the gene encoding cytochrome *c*550 (*psbV*), which has a crucial role in the oxygen-evolving complex in *Synechocystis* PCC6803 (ref. 20).

There are several differences between the genomes that help account for the different light optima of the two strains. For example, the smaller MED4 genome has more than twice as many genes (22 compared with 9) encoding putative high-light-inducible proteins, which seem to have arisen at least in part through duplication events[15]. MED4 also possesses a photolyase gene that has been lost in MIT9313, probably because there is little selective pressure to retain ultraviolet damage repair in low light habitats. Regarding differences in light-harvesting efficiencies, it is noteworthy that MED4 contains only a single gene encoding the chlorophyll *a*/*b*-binding antenna protein Pcb, whereas MIT9313 possesses two copies. The second type has been found exclusively in low-light-adapted strains[21], and may form an antenna capable of binding more chlorophyll pigments.

Both strains have a low proportion of genes involved in regulatory functions. Compared with the freshwater cyanobacterium *Thermosynechococcus elongatus* (genome size <2.6 megabases)[22], MIT9313 has fewer sigma factors, transcriptional regulators and two-component sensor-kinase systems, and MED4 is even more reduced (Supplementary Table 1). The circadian clock genes provide an example of this reduction as both genomes lack several components (*pex*, *kaiA*) found in the model *Synechococcus* PCC7942 (ref. 23). However, genes for the core clock proteins (*kaiB*, *kaiC*) remain in both genomes, and *Prochlorococcus* cell division is tightly synchronized to the diel light/dark cycle[24]. Thus, loss of some circadian components may imply an alternative signalling pathway for circadian control.

Gene loss may also have a role in the lower percentage of G+C content of MED4 (30.8%) compared with that of MIT9313 (50.74%), which is more typical of marine *Synechococcus*. MED4 lacks genes for several DNA repair pathways including recombinational repair (*recJ*, *recQ*) and damage reversal (*mutT*). Particularly, the loss of the base excision repair gene *mutY*, which removes adenosines incorrectly paired with oxidatively damaged guanine residues, may imply an increased rate of G•C to T•A transversions[25]. The tRNA complement of MED4 is largely identical to MIT9313 and is not optimized for a low percentage G+C genome, suggesting that it is not evolving as fast as codon usage.

Analysis of the nitrogen acquisition capabilities of the two strains points to a sequential decay in the capacity to use nitrate and nitrite during the evolution of the *Prochlorococcus* lineage (Fig. 3a). In *Synechococcus* WH8102—representing the presumed ancestral state—many nitrogen acquisition and assimilation genes are grouped together (Fig. 3a). MIT9313 has lost a 25-gene cluster, which includes genes encoding the nitrate/nitrite transporter and nitrate reductase. The nitrite reductase gene has been retained in MIT9313, but it is flanked by a proteobacterial-like nitrite transporter rather than a typical cyanobacterial nitrate/nitrite permease (Supplementary Fig. 4), suggesting acquisition by lateral gene transfer. An additional deletion event occurred in MED4, in which the nitrite reductase gene was also lost (Fig. 3a). As a result of these serial deletion events MIT9313 cannot use nitrate, and MED4 cannot use nitrate or nitrite[9]. Thus each *Prochlorococcus* ecotype uses the N species that is most prevalent at the light levels to which they are best adapted: ammonium in the surface waters and nitrite at depth (Fig. 1a). *Synechococcus*, which is the only one of the three that has nitrate reductase, is able to bloom when nitrate is upwelled (Fig. 1a), as occurs in the spring in the North Atlantic[3] and the north Red Sea[26].

The two *Prochlorococcus* strains are also less versatile in their organic N usage capabilities than *Synechococcus* WH8102 (ref. 13). MED4 contains the genes necessary for usage of urea, cyanate and oligopeptides, but no monomeric amino acid transporters have been identified. In contrast, MIT9313 contains transporters for urea, amino acids and oligopeptides but lacks the genes necessary for cyanate usage (cyanate transporter and cyanate lyase) (Fig. 3a). As expected, both genomes contain the high-affinity ammonium transporter *amt1* and both lack the nitrogenase genes essential for nitrogen fixation. Finally, both contain the nitrogen transcriptional regulator encoded by *ntcA* and there are numerous genes in both genomes, including *ntcA*, *amt1*, the urea transport and GS/GOGAT genes (glutamine synthetase and glutamate synthase, both involved in ammonia assimilation), with an upstream NtcA-binding-site consensus sequence.

The genomes also have differences in genes involved in phosphorus usage that have obvious ecological implications. MED4, but not MIT9313, is capable of growth on organic P sources (L. R. Moore and S.W.C., unpublished data), and organic P can be the prevalent form of P in high-light surface waters[27]. This difference may be due to the acquisition of an alkaline phosphatase-like gene in MED4 (Supplementary Fig. 5). Both genomes contain the high-affinity phosphate transport system encoded by *pstS* and *pstABC*[28], but MIT9313 contains an additional copy of the phosphate-binding component *pstS*, perhaps reflecting an increased reliance on orthophosphate in deeper waters. MED4 contains

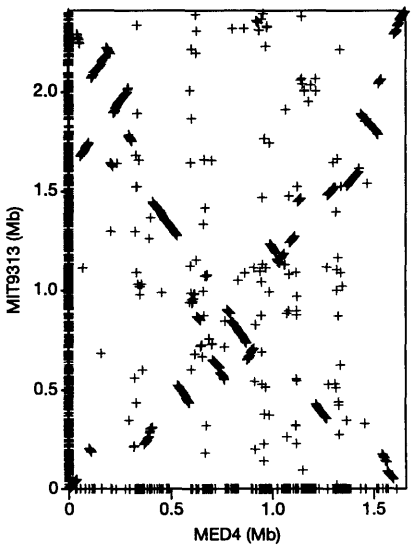


Figure 2 Global genome alignment as seen from start positions of orthologous genes. Genes present in one genome but not the other are shown on the axes. The 'broken X' pattern has been noted before for closely related bacterial genomes, and is probably due to multiple inversions centred around the origin of replication. Alternating slopes of many adjacent gene clusters indicate that multiple smaller-scale inversions have also occurred.

several P-related regulatory genes including the *phoB*, *phoR* two-component system and the transcriptional activator *ptrA*. In MIT9313, however, *phoR* is interrupted by two frameshifts and *ptrA* is further degenerated, suggesting that this strain has lost the ability to regulate gene expression in response to changing P levels.

Both *Prochlorococcus* strains have iron-related genes that are missing in *Synechococcus* WH8102, which may explain its dominance in the iron-limited equatorial Pacific[2]. These genes include flavodoxin (*isiB*), an Fe-free electron transfer protein capable of replacing ferredoxin, and ferritin (located with the ATPase component of an iron ABC transporter), an iron-binding molecule implicated in iron storage. Additional characteristics of the iron acquisition system in these genomes include: an Fe-induced transcriptional regulator (Fur) that represses iron uptake genes; numerous genes with an upstream putative *fur* box motif that are candidates for a high-affinity iron scavenging system; and absence of genes involved in Fe–siderophore complexes.

Prochlorococcus does not use typical cyanobacterial genes for inorganic carbon concentration or fixation. Both genomes contain a sodium/bicarbonate symporter but lack homologues to known families of carbonic anhydrases, suggesting that an as yet unidentified gene is fulfilling this function. One of the two carbonic anhydrases in *Synechococcus* WH8102 was lost in the deletion event that led to the loss of the nitrate reductase (Fig. 3a); the other is located next to a tRNA and seems to have been lost during a genome rearrangement event. Similar to other *Prochlorococcus* and marine *Synechococcus*, MED4 and MIT9313 possess a form IA ribulose-1,5-bisphosphate carboxylase/oxygenase, rather than the typical cyanobacterial form IB. The ribulose-1,5-bisphosphate carboxylase/oxygenase genes are adjacent to genes encoding structural carboxysome shell proteins and all have phylogenetic affinity to genes in the γ-proteobacterium *Acidithiobacillus ferroxidans*[15], suggesting lateral transfer of the extended operon.

Prochlorococcus has been identified in deep suboxic zones where it is unlikely that they can sustain themselves by photosynthesis alone[29], thus we looked for genomic evidence of heterotrophic capability. Indeed, the presence of oligopeptide transporters in both genomes, and the larger proportion of transporters (including some sugar transporters) in the MIT9313 strain-specific genes (Supplementary Fig. 2), suggests the potential for partial hetero-

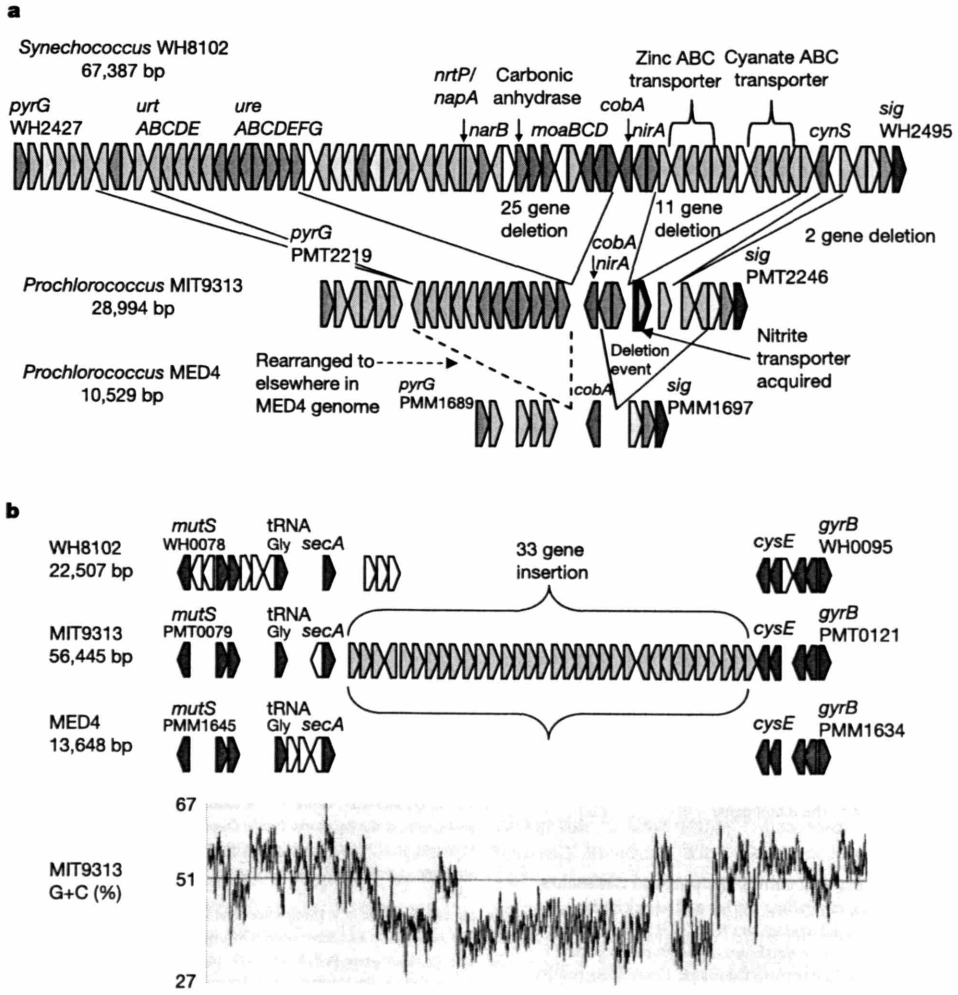


Figure 3 Dynamic architecture of marine cyanobacterial genomes. **a**, Deletion, acquisition and rearrangement of nitrogen usage genes. In MIT9313, 25 genes including the nitrate/nitrite transporter (*nrtP/napA*), nitrate reductase (*narB*) and carbonic anhydrase have been deleted. The cyanate transporter and cyanate lyase (*cynS*) were probably lost after the divergence of MIT9313 from the rest of the *Prochlorococcus* lineage, as MED4 possesses these genes. MIT9313 has retained nitrite reductase (*nirA*) and acquired a nitrite transporter. In MED4 *nirA* has been lost and the urea transporter (*urt* cluster) and urease (*ure* cluster) genes have been rearranged (dotted line). Genes in different functional categories are colour-coded to guide the eye. **b**, Lateral transfer of genes involved in lipopolysaccharide biosynthesis including sugar transferases, sugar epimerases, modifying enzymes and two pairs of ABC-type transporters. Blue, genes in all three genomes; pink, genes hypothesized to have been laterally transferred; red, tRNAs; white, other genes. The percentage of G + C content in MIT9313 along this segment is lower (42%) than the whole-genome average (horizontal line).

trophy. However, neither genome contains known pathways that would allow for complete heterotrophy. They are both missing genes for steps in the tricarboxylic acid cycle, including 2-oxoglutarate dehydrogenase, succinyl-CoA synthetase and succinyl-CoA-acetoacetate-CoA transferase.

Cell surface chemistry has a major role in phage recognition and grazing by protists and thus is probably under intense selective pressure in nature. The two *Prochlorococcus* genomes and the *Synechococcus* WH8102 genome show evidence of extensive lateral gene transfer and deletion events of genes involved in lipopolysaccharide and/or surface polysaccharide biosynthesis, reinforcing the role of predation pressures in the creation and maintenance of microdiversity. For example, MIT9313 has a 41.8-kilobase (kb) cluster of surface polysaccharide genes (Fig. 3b), which has a lower percentage G+C composition (42%) than the genome as a whole, implicating acquisition by lateral gene transfer. MED4 has acquired a 74.5-kb cluster consisting of 67 potential surface polysaccharide genes (Supplementary Fig. 6a) and has lost another cluster of surface polysaccharide biosynthesis genes shared between MIT9313 and *Synechococcus* WH8102 (Supplementary Fig. 6b).

The approach we have taken in describing these genomes highlights the known drivers of niche partitioning of these closely related organisms (Fig. 1). Detailed comparisons with the genomes of additional strains, such as *Prochlorococcus* SS120 (ref. 30), will enrich this story, and the analysis of whole genomes from *in situ* populations will be necessary to understand the full expanse of genomic diversity in this group. The genes of unknown function in all of these genomes hold important clues for undiscovered niche dimensions in the marine pelagic zone. As we unveil their function we will undoubtedly learn that the suite of selective pressures that shape these communities is much larger than we have imagined. Finally, it may be useful to view *Prochlorococcus* and *Synechococcus* as important 'minimal life units', as the information in their roughly 2,000 genes is sufficient to create globally abundant biomass from solar energy and inorganic compounds. □

Methods

Genome sequencing and assembly

DNA was isolated from the clonal, axenic strain MED4 and the clonal strain MIT9313 essentially as described previously[4]. The two whole-genome shotgun libraries were obtained by fragmenting genomic DNA using mechanical shearing and cloning 2–3-kb fragments into pUC18. Double-ended plasmid sequencing reactions were carried out using PE BigDye Terminator chemistry (Perkin Elmer) and sequencing ladders were resolved on PE 377 Automated DNA Sequencers (Perkin Elmer). The whole-genome sequence of *Prochlorococcus* MED4 was obtained from 27,065 end sequences (7.3-fold redundancy), whereas *Prochlorococcus* MIT9313 was sequenced to ×6.2 coverage (33,383 end sequences). For *Prochlorococcus* MIT9313, supplemental sequencing (×0.05 sequence coverage) of a pFos1 fosmid library was used as a scaffold. Sequence assembly was accomplished using PHRAP (P. Green). All gaps were closed by primer walking on gap-spanning library clones or PCR products. The final assembly of *Prochlorococcus* MED4 was verified by long-range genomic PCR reactions, whereas the assembly of *Prochlorococcus* MIT9313 was confirmed by comparison to the fosmid clones, which were fingerprinted with *Eco*RI. No plasmids were detected in the course of genome sequencing, and insertion sequences, repeated elements, transposons and prophages are notably absent from both genomes. The likely origin of replication in each genome was identified based on G+C skew, and base pair 1 was designated adjacent to the *dnaN* gene.

Genome annotation

The combination of three gene-modelling programs, Critica, Glimmer and Generation, were used in the determination of potential open reading frames and were checked manually. A revised gene/protein set was searched against the KEGG GENES, Pfam, PROSITE, PRINTS, ProDom, COGs and CyanoBase databases, in addition to BLASTP against the non-redundant peptide sequence database from GenBank. From these results, categorizations were developed using the KEGG and COGs hierarchies, as modified in CyanoBase. Manual annotation of open reading frames was done in conjunction with the *Synechococcus* team. The three-way genome comparison was used to refine predicted start sites, add additional open reading frames and standardize the annotation across the three genomes.

Genome comparisons

The comparative genome architecture of MED4 and MIT9313 was visualized using the Artemis Comparison Tool (http://www.sanger.ac.uk/Software/ACT/). Orthologues were determined by aligning the predicted coding sequences of each gene with the coding sequences of the other genome using BLASTP. Genes were considered orthologues if each was the best hit of the other one and both *e*-values were less than e^{-10}. In addition, bidirectional best hits with *e*-values less than e^{-6} and small proteins of conserved function were manually examined and added to the orthologue lists.

Phylogenetic analyses used PAUP*, logdet distances and minimum evolution as the objective function. The degree of support at each node was evaluated using 1,000 bootstrap resamplings. Ribosomal DNA analyses used 1,160 positions. The Gram-positive bacterium *Arthrobacter globiformis* was used to root the tree.

Received 11 May; accepted 25 July 2003; doi:10.1038/nature01947.
Published online 13 August 2003.

1. Chisholm, S. W. *et al.* A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* **334,** 340–343 (1988).
2. Campbell, L., Liu, H., Nolla, H. & Vaulot, D. Annual variability of phytoplankton and bacteria in the subtropical North Pacific Ocean at Station ALOHA during the 1991–1994 ENSO event. *Deep Sea Res. I* **44,** 167–192 (1997).
3. DuRand, M. D., Olson, R. J. & Chisholm, S. W. Phytoplankton population dynamics at the Bermuda Atlantic Time-series Station in the Sargasso Sea. *Deep Sea Res. II* **48,** 1983–2003 (2001).
4. Rocap, G., Distel, D. L., Waterbury, J. B. & Chisholm, S. W. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S–23S rDNA internal transcribed spacer (ITS) sequences. *Appl. Environ. Microbiol.* **68,** 1180–1191 (2002).
5. Moore, L. R. & Chisholm, S. W. Photophysiology of the marine cyanobacterium *Prochlorococcus*: Ecotypic differences among cultured isolates. *Limnol. Oceanogr.* **44,** 628–638 (1999).
6. Moore, L. R., Rocap, G. & Chisholm, S. W. Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393,** 464–467 (1998).
7. Moore, L. R., Goericke, R. E. & Chisholm, S. W. Comparative physiology of *Synechococcus* and *Prochlorococcus*: influence of light and temperature on growth, pigments, fluorescence and absorptive properties. *Mar. Ecol. Prog. Ser.* **116,** 259–275 (1995).
8. Mann, E. L., Ahlgren, N., Moffett, J. W. & Chisholm, S. W. Copper toxicity and cyanobacteria ecology in the Sargasso Sea. *Limnol. Oceanogr.* **47,** 976–988 (2002).
9. Moore, L. R., Post, A. F., Rocap, G. & Chisholm, S. W. Utilization of different nitrogen sources by the marine cyanobacteria, *Prochlorococcus* and *Synechococcus*. *Limnol. Oceanogr.* **47,** 989–996 (2002).
10. Sullivan, M. B., Waterbury, J. B. & Chisholm, S. W. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424,** 1047–1051 (2003).
11. Hagström, Å. *et al.* Use of 16S ribosomal DNA for delineation of marine bacterioplankton species. *Appl. Environ. Microbiol.* **68,** 3628–3633 (2002).
12. West, N. J. *et al.* 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,** 1731–1744 (2001).
13. Palenik, B. *et al.* The genome of a motile marine *Synechococcus*. *Nature* **424,** 1037–1042 (2003).
14. La Roche, J. *et al.* Independent evolution of the prochlorophyte and green plant chlorophyll a/b light harvesting proteins. *Proc. Natl Acad. Sci. USA* **93,** 15244–15248 (1996).
15. Hess, W. *et al.* The photosynthetic apparatus of *Prochlorococcus*: Insights through comparative genomics. *Photosynth. Res.* **70,** 53–71 (2001).
16. Stickforth, P., Steiger, S., Hess, W. R. & Sandmann, G. A novel type of lycopene ε-cyclase in the marine cyanobacterium *Prochlorococcus marinus* MED4. *Arch. Microbiol.* **179,** 407–415 (2003).
17. Frankenberg, N., Mukougawa, K., Kohchi, T. & Lagarias, J. C. Functional genomic analysis of the HY2 family of ferredoxin-dependent bilin reductases from oxygenic photosynthetic organisms. *Plant Cell* **13,** 965–978 (2001).
18. Ting, C., Rocap, G., King, J. & Chisholm, S. W. Phycobiliprotein genes of the marine prokaryote *Prochlorococcus*: Evidence for rapid evolution of genetic heterogeneity. *Microbiology* **147,** 3171–3182 (2001).
19. Golden, S. S., Brusslan, J. & Haselkorn, R. Expression of a family of *psbA* genes encoding a photosystem II polypeptide in the cyanobacterium *Anacystis nidulans* R2. *EMBO J.* **5,** 2789–2798 (1986).
20. Shen, J. R., Qian, M., Inoue, Y. L. & Burnap, R. L. Functional characterization of *Synechocystis* sp. PCC 6803 Δ*psbU* and Δ*psbV* mutants reveals important roles of cytochrome c-550 in cyanobacterial oxygen evolution. *Biochemistry* **37,** 1551–1558 (1998).
21. Garczarek, L., van der Staay, G. W. M., Hess, W. R., Le Gall, F. & Partensky, F. Expression and phylogeny of the multiple antenna genes of the low-light-adapted strain *Prochlorococcus marinus* SS120 (Oxyphotobacteria). *Plant Mol. Biol.* **46,** 683–693 (2001).
22. Nakamura, Y. *et al.* Complete genome structure of the thermophilic Cyanobacterium *Thermosynechococcus elongatus* BP-1. *DNA Res.* **9,** 123–130 (2002).
23. Ishiura, M. *et al.* Expression of a gene cluster *kaiABC* as a circadian feedback process in cyanobateria. *Science* **281,** 1519–1523 (1998).
24. Vaulot, D., Marie, D., Olson, R. J. & Chisholm, S. W. Growth of *Prochlorococcus*, a photosynthetic prokaryote, in the Equatorial Pacific Ocean. *Science* **268,** 1480–1482 (1995).
25. Michaels, M. L. & Miller, J. H. The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-Hydroxyguanine (7,8-Dihydro-8-Oxoguanine). *J. Bacteriol.* **174,** 6321–6325 (1992).
26. Lindell, D. & Post, A. F. Ultraphytoplankton succession is triggered by deep winter mixing in the Gulf of Aqaba (Eilat). *Red Sea. Limnol. Oceanogr.* **40,** 1130–1141 (1995).
27. Karl, D. M., Bidigare, R. R. & Letelier, R. M. Long-term changes in plankton community structure and productivity in the North Pacific Subtropical Gyre: The domain shift hypothesis. *Deep Sea Res. II* **48,** 1449–1470 (2001).
28. Scanlan, D. J., Mann, N. H. & Carr, N. G. The response of the picoplanktonic marine cyanobacterium *Synechococcus* species WH7803 to phosphate starvation involves a protein homologous to the periplasmic phosphate binding protein of *Escherichia coli*. *Mol. Microbiol.* **10,** 181–191 (1993).
29. Johnson, Z. *et al.* Energetics and growth kinetics of a deep *Prochlorococcus* spp. population in the Arabian Sea. *Deep Sea Res. II* **46,** 1719–1743 (1999).
30. Dufresne, A. *et al.* Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a near minimal oxyphototrophic genome. *Proc. Natl Acad. Sci. USA* (in the press).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements This research was funded by the Biological and Environmental Research Program of the US Department of Energy's Office of Science. The Joint Genome Institute managed the overall sequencing effort. Genome finishing was carried out under the auspices of the US Department of Energy by the University of California, Lawrence Livermore National Laboratory. Computational annotation was carried out at the Oak Ridge National Laboratory, managed by UT-BATTELLE for the US Department of Energy. Additional support was provided by the DOE, NSF and the Seaver Foundation to S.W.C., the Israel–US Binational Science Foundation to A.F.P. and S.W.C., and FP5-Margenes to W.R.H. and A.F.P. We thank the *Synechococcus* WH8102 annotators (B. Palenik, B. Brahamsha, J. McCarren, E. Allen, F. Partensky, A. Dufresne and I. Paulsen) for their help with curating the *Prochlorococcus* genomes and E. V. Armbrust and L. Moore for critical reading of the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.W.C. (chisholm@mit.edu). The complete nucleotide sequences and sequences of predicted open reading frames have been deposited in the EMBL/GenBank/DDBJ databases under accession numbers BX548174 (MED4) and BX548175 (MIT9313).

Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*

Matthew B. Sullivan[1], John B. Waterbury[3] & Sallie W. Chisholm[2]

[1]*MIT/Woods Hole Oceanographic Institution Joint Program in Biological Oceanography, and* [2]*Department of Civil and Environmental Engineering and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[3]*Woods Hole Oceanographic Institution, Department of Biology, Woods Hole, Massachusetts 02543, USA*

***Prochlorococcus* is the numerically dominant phototroph in the tropical and subtropical oceans, accounting for half of the photosynthetic biomass in some areas[1,2]. Here we report the isolation of cyanophages that infect *Prochlorococcus*, and show that although some are host-strain-specific, others cross-infect with closely related marine *Synechococcus* as well as between high-light- and low-light-adapted *Prochlorococcus* isolates, suggesting a mechanism for horizontal gene transfer. High-light-adapted *Prochlorococcus* hosts yielded *Podoviridae* exclusively, which were extremely host-specific, whereas low-light-adapted *Prochlorococcus* and all strains of *Synechococcus* yielded primarily *Myoviridae*, which has a broad host range. Finally, both *Prochlorococcus* and *Synechococcus* strain-specific cyanophage titres were low ($<10^3$ ml^{-1}) in stratified oligotrophic waters even where total cyanobacterial abundances were high ($>10^5$ cells ml^{-1}). These low titres in areas of high total host cell abundance seem to be a feature of open ocean ecosystems. We hypothesize that gradients in cyanobacterial population diversity, growth rates, and/or the incidence of lysogeny underlie these trends.**

Phages are thought to evolve by the exchange of genes drawn from a common gene pool through differential access imposed by host range limitations[3]. Similarly, horizontal gene transfer, important in microbial evolution[4,5], can be mediated by phages[6] and is probably responsible for many of the differences in the genomes of closely related microbes[5]. Recent detailed analyses of molecular phylogenies constructed for marine *Prochlorococcus* and *Synechococcus*[7,8] (Fig. 1) show that these genera form a single group within the marine picophytoplankton clade[9] (>96% identity in 16S ribosomal DNA sequences), yet display microdiversity in the form of ten well-defined subgroups[8]. We have used members of these two groups to study whether phage isolated on a particular host strain cross-infect other hosts, and if so, whether the probability of cross-infection is related to rDNA-based evolutionary distance between the hosts.

Analyses of host range were conducted (Fig. 1) with 44 cyanophages, isolated as previously described[10] from a variety of water depths and locations (see Supplementary Information) using 20 different host strains chosen to represent the genetic diversity of *Prochlorococcus* and *Synechococcus*[8]. Although we did not examine how these patterns would change if phage were propagated on different hosts, this would undoubtedly add another layer of complexity due to host range modifications as a result of methylation of phage DNA[6]. Similar to those that infect other marine bacteria[11] and *Synechococcus*[10–14], our *Prochlorococcus* cyanophage isolates fell into three morphological families: *Myoviridae, Siphoviridae* and *Podoviridae*[15].

As would be predicted[10–14], *Podoviridae* were extremely host specific with only two cross-infections out of a possible 300 (Fig. 1). Similarly, the two *Siphoviridae* isolated were specific to their hosts. In instances of extreme host specificity, *in situ* host abundance would need to be high enough to facilitate phage–host contact. It is noteworthy in this regard that members of the high-light-adapted *Prochlorococcus* cluster, which yielded the most host-specific cyanophage, have high relative abundances *in situ*[16]. The *Myoviridae* exhibited much broader host ranges, with 102 cross-infections out of a possible 539. They not only cross-infected among and between *Prochlorococcus* ecotypes but also between *Prochlorococcus* and *Synechococcus*. Those isolated with *Synechococcus* host strains have broader host ranges and are more likely to cross-infect low-light-adapted than high-light-adapted *Prochlorococcus* strains. The low-light-adapted *Prochlorococcus* are less diverged from *Synechococcus* than high-light-adapted *Prochlorococcus*[7,8], suggesting a relationship, in this instance, between the probability of cross-infection and rDNA relatedness of hosts. Finally, we tested the *Myoviridae* for cross-infection against marine bacterial isolates closely related to *Pseudoalteromonas*, which are known to be broadly susceptible to diverse bacteriophages (bacterial strains HER1320, HER1321, HER1327, HER1328)[11]. None of the *Myoviridae* cyanophages infected these bacteria.

Phage morphotypes isolated were determined, to some degree, by the host used for isolation (Fig. 1). For example, ten of ten cyanophages isolated using high-light-adapted *Prochlorococcus* strains were *Podoviridae*. In contrast, all but two cyanophages isolated on *Synechococcus* were *Myoviridae*, a bias that has been reported by others[14], and over half of those isolated on low-light-adapted *Prochlorococcus* belonged to this morphotype. We further substantiated these trends by examining lysates (as opposed to plaque-purified isolates) from a range of host strains, geographic locations and depths—of 58 *Synechococcus* lysates 93% contained *Myoviridae*, of 43 low-light-adapted *Prochlorococcus* lysates 65% contained *Myoviridae*, and of 107 high-light-adapted *Prochlorococcus* lysates 98% contained *Podoviridae* (see Supplementary Information).

Maximum cyanophage titres, using a variety of *Synechococcus* hosts, are usually found to be within an order of magnitude of the total *Synechococcus* abundance[10,14,17,18], and can be as high as 10^6 phage ml^{-1}. One study[17] has shown, for example, that along a transect in which total *Synechococcus* abundance decreased from 10^5 cells ml^{-1} to 250 cells ml^{-1}, maximum cyanophage titres remained at least as high as the total number of *Synechococcus*. We wondered whether titres of *Prochlorococcus* cyanophage in the Sargasso Sea, where *Prochlorococcus* cells are abundant (10^5 cells ml^{-1}), would be comparable to those measured in coastal oceans for *Synechococcus* where total *Synechococcus* host abundances are of similar magnitude. We assayed cyanophage titres in a depth profile in the Sargasso Sea at the end of seasonal stratification using 11 strains of *Prochlorococcus* (Fig. 2), choosing at least one host strain from each of the six phylogenetic clusters that span the rDNA-based genetic diversity of our culture collection[8].

Three *Prochlorococcus* host strains (MIT 9303, MIT 9313 and SS120) yielded low or no cyanophage. Other hosts yielded titres

Transfer of photosynthesis genes to and from *Prochlorococcus* viruses

Debbie Lindell*†, Matthew B. Sullivan†‡, Zackary I. Johnson*, Andrew C. Tolonen‡, Forest Rohwer§, and Sallie W. Chisholm*¶‖

Departments of *Civil and Environmental Engineering and ¶Biology, Massachusetts Institute of Technology, Cambridge, MA 02139; ‡Joint Program in Biological Oceanography, Woods Hole Oceanographic Institution and Massachusetts Institute of Technology, Cambridge, MA 02139; and §Department of Biology, San Diego State University, San Diego, CA 92182

Edited by W. Ford Doolittle, Dalhousie University, Halifax, Canada, and approved June 11, 2004 (received for review March 3, 2004)

Comparative genomics gives us a new window into phage–host interactions and their evolutionary implications. Here we report the presence of genes central to oxygenic photosynthesis in the genomes of three phages from two viral families (*Myoviridae* and *Podoviridae*) that infect the marine cyanobacterium *Prochlorococcus*. The genes that encode the photosystem II core reaction center protein D1 (*psbA*), and a high-light-inducible protein (HLIP) (*hli*) are present in all three genomes. Both myoviruses contain additional *hli* gene types, and one of them encodes the second photosystem II core reaction center protein D2 (*psbD*), whereas the other encodes the photosynthetic electron transport proteins plastocyanin (*petE*) and ferredoxin (*petF*). These uninterrupted, full-length genes are conserved in their amino acid sequence, suggesting that they encode functional proteins that may help maintain photosynthetic activity during infection. Phylogenetic analyses show that phage D1, D2, and HLIP proteins cluster with those from *Prochlorococcus*, indicating that they are of cyanobacterial origin. Their distribution among several *Prochlorococcus* clades further suggests that the genes encoding these proteins were transferred from host to phage multiple times. Phage HLIPs cluster with multicopy types found exclusively in *Prochlorococcus*, suggesting that phage may be mediating the expansion of the *hli* gene family by transferring these genes back to their hosts after a period of evolution in the phage. These gene transfers are likely to play a role in the fitness landscape of hosts and phages in the surface oceans.

The genomes of bacterial viruses (phages) contain a variety of genes homologous to those found in their hosts (1–5). Many encode functional proteins involved in processes of direct importance for the production of phage progeny. They include genes involved in DNA replication, nucleotide metabolism, and RNA transcription and are found in both lytic phage and prophage (3, 6). It is likely that many originated from their hosts (2, 4) and that some host genes that occur in multiple copies have been (re)acquired from phages (2, 7) either after a period of evolution in the phage or after acquisition of the gene from a different host.

Host genes that are not directly related to the production of new phages, such as genes involved in phosphate sensing and metabolism (8, 9), and the scavenging of oxygen radicals (10) are also found in phage genomes and may benefit phages by temporarily enhancing host functionality before lysis. In addition, prophages can provide their hosts with new functions by encoding genes, such as virulence factors, toxin production genes, and immune response genes (5, 6, 11).

Genes involved in photosynthesis have recently been found in a lytic phage isolated on *Synechococcus* WH7803 (12), a member of the marine cluster A unicellular cyanobacteria that is widespread in the oceans. A member of the *Myoviridae* family of double-stranded DNA viruses, this phage contains two photosynthetic genes (*psbD* and an interrupted *psbA* gene) that code for the two photosystem II (PSII) core reaction center proteins found in all oxygenic photosynthetic organisms. These genes were not found in a different phage (a member of the *Podoviridae* family) isolated on the same strain of *Synechococcus* (13). These observations lead one to wonder whether the presence of photosynthetic genes in phage is a rare phenomenon and to what extent it is specific for a particular phage or host type. If these genes are widespread in cyanophage, what is their origin? Were they acquired through a single ancestral transfer event?

The phage–host system for *Prochlorococcus* and *Synechococcus* (14, 15), which form a monophyletic clade within the cyanobacteria (16–19), is well suited to begin to answer these questions. Members of each genus form distinct subgenera clusters within this clade, which in *Prochlorococcus* also correspond to their efficiency of light utilization (17). Numerous phages have been isolated by using this diverse group, including members of the *Myoviridae*, *Podoviridae*, and *Siphoviridae* families, and the degree of cross-infection, a mechanism for horizontal gene transfer, has been analyzed (14, 15). The genomes of four host strains (20–22) and three phages (U.S. Department of Energy Joint Genome Institute; www.jgi.doe.gov) have been sequenced, providing a database to analyze the distribution and phylogenetic relationships of host genes among hosts and their phages.

Here we report that the genomes of three *Prochlorococcus* phages collectively contain a number of host-like photosynthetic genes. We further hypothesize from bioinformatic analyses that these genes likely play a functional role during infection and impact the evolutionary trajectory of both phages and hosts in the surface oceans.

Materials and Methods

Selection and Preparation of Cyanophage for Genome Sequencing. Three phages were chosen for sequencing with no prior knowledge of their gene content. P-SSP7, a T7-like podovirus characterized by a small capsid (≈50 nm), a noncontractile tail, and a 45-kb genome infects a single high-light-adapted (HL) *Prochlorococcus* strain. P-SSM2 and P-SSM4 are T4-like myoviruses characterized by larger capsids (≈85 nm and ≈80 nm respectively), long contractile tails, and larger genomes (252 kb and 178 kb, respectively). P-SSM2 infects three low-light-adapted (LL) *Prochlorococcus* strains, and P-SSM4 infects two HL and two LL *Prochlorococcus* strains (see Table 1) (15). None of the three phages infect *Synechococcus*. The vastly different protein com-

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

Abbreviations: PSII, photosystem II; HLIP, high-light-inducible protein; HL, high-light-adapted; LL, low-light-adapted.

Data deposition: The phage genome and *Prochlorococcus psbA* sequences reported in this article have been deposited in the GenBank database (accession nos. AY571331, AY575566, and AY575567 for phage genome sequences and AY599028–AY599035 for *Prochlorococcus psbA* sequences).

†D.L. and M.B.S. contributed equally to this work.

‖To whom correspondence should be addressed at: Department of Civil Environmental Engineering, Massachusetts Institute of Technology, Room 48-425, 77 Massachusetts Avenue, Cambridge, MA 02139. E-mail: chisholm@mit.edu.

Table 1. Phages used in this study and their photosynthesis-related genes

Phage	Family	Host strains infected	Gene products
P-SSP7	Podovirus	*Pro* **MED4** (HL)	D1 and one HLIP
P-SSM2	Myovirus	*Pro* **NATL1A**, NATL2A, and MIT9211 (LL)	D1, six HLIPs, ferredoxin, and plastocyanin
P-SSM4	Myovirus	*Pro* NATL1A, **NATL2A** (LL), *Pro* MED4, and MIT9215 (HL)	D1, D2, and four HLIPs
S-PM2*	Myovirus	*Syn* **WH7803** and WH8109	D1 and D2

Phage family and host-range information is per ref. 15. Boldface indicates the host on which the phage was isolated.
*From Mann *et al.* (12).

plements of the T7- and T4-like phages distinguishes them as distinctly different organisms in whole proteomic taxonomic reconstructions (23).

Phages were propagated on their *Prochlorococcus* hosts (P-SSP7 on MED4, P-SSM2 on NATL1A, and P-SSM4 on NATL2A) and were purified for DNA extraction and construction of clone libraries as described in ref. 8. Briefly, cell lysate was treated with nucleases to degrade host nucleic acids. Phages were precipitated by using polyethylene glycol 8000, purified on a cesium chloride step gradient (steps were ρ = 1.30, 1.40, 1.50, and 1.65) spun at 104,000 × g for 2 h at 4°C, and dialyzed against a buffer containing 100 mM Tris·HCl (pH 7.5), 100 mM $MgSO_4$, and 30 mM NaCl. Purified phages were burst by using SDS (0.5%) and proteinase K (50 μg/ml). DNA was extracted with phenol:chloroform and concentrated by ethanol precipitation. A custom Los Alamos Scientific Lab clone library was constructed by Lucigen (Middleton, WI) as described in ref. 24. Inserts were sequenced and genomes were assembled by the Department of Energy Joint Genome Institute. Analyses were conducted on the phage genomes as provided on October 17, 2003 (P-SSM2 and P-SSM4), and November 19, 2003 (P-SSP7). At that time, these genomes were in large high-quality contigs compiled from 26-fold (P-SSP7), 30-fold (P-SSM2), and 39-fold (P-SSM4) coverage, respectively.

PCR Amplification of *psbA*. Genomic DNA was isolated from *Prochlorococcus* cultures by using the DNeasy kit (Qiagen, Valencia, CA). Partial *psbA* sequences were amplified by using primers from (19) or for *Prochlorococcus* MIT9211 by using the following primers: 5′-AACATCATYTCWGGTGCWGT-3′ and 5′-TCGTGCATTACTTCCATACC-3′. Reactions (50 μl) consisted of 4 mM $MgCl_2$, 200 μM dNTP, 0.25 μM (each) primer, 2.5 units of *Taq*DNA polymerase (Invitrogen), and 4 ng of genomic DNA. Amplification conditions, which were run on a RoboCycler Gradient 96 thermocycler (Stratagene), comprised steps at 92°C for 4 min; 35 cycles at 92°C for 1 min, 50°C for 1 min, and 68°C for 1 min; followed by a final extension step at 68°C for 10 min. Fragments were gel-purified and sequenced in both forward and reverse directions (Davis Sequencing, Davis, CA).

Identification of Genes and Transcriptional Regulatory Elements. ORFs in the phage genomes were identified by using GENEMARK (25), and gene identifications were based on homology to known proteins by using the BLASTP program (ftp://ftp.ncbi.nih.gov/blast) with an *E*-value cutoff of 10^{-5}. Ferredoxin-encoding genes (*petF*) were included in our analyses if they encoded the 2Fe–2S iron–sulfur cluster-binding domain (fer2) (with an *E* value $<10^{-10}$ as determined by the BLAST tool RPSBLAST from the conserved domain database of the National Center for Biotechnology Information. High-light-inducible protein (HLIP)-encoding genes (*hli*) were identified as present if they encoded at least six of 10 amino acids in the motif AExxNGRxAMIGF (26). Bhaya *et al.* (27) report that many *Prochlorococcus hli* genes code for a conserved 9-aa C-terminal sequence with the consensus sequence TGQIIPGI/FF. Here this sequence was defined as present when at least six of the nine conserved amino acids were found.

ρ-Independent transcriptional terminators were identified by using the TRANSTERM program (28), and all had an energy score of <-10 and a tail score of <-5. Potential bacterial σ^{70} promoters were identified in intergenic regions by using the program BPROM (SoftBerry, Mount Kisco, NY). Promoter sequences had a linear discriminant function >2.5. Although identification of terminators is robust, promoter identification in cyanophage is presently more precarious.

Sequence Manipulation and Analyses. Sequences were aligned by using CLUSTALX and edited manually as necessary. Amino acid alignments served as the basis for the manual alignment of nucleotide sequences. Regions that could not be confidently aligned were excluded from analyses, as were gaps. The divergence estimator program K-ESTIMATOR 6.0 (29) was used to estimate the frequency of synonymous and nonsynonymous nucleotide substitutions and employs the Kimura 2p correction method for multiple hits.

PAUP Version 4.0b10 was used for the construction of distance and maximum parsimony trees. Amino acid distance trees were inferred by using minimum evolution as the objective function and mean distances. Heuristic searches were performed with 100 random addition-sequence replicates and the tree-bisection and reconnection branch-swapping algorithms. Starting trees were obtained by stepwise addition of sequences. Bootstrap analyses of 100 resamplings were carried out. Maximum likelihood trees were constructed by using TREE-PUZZLE 5.0. Evolutionary distances were calculated by using the JTT model of substitution (except for the highly divergent HLIPs, for which the VT model of substitution was used) assuming a γ-distributed model of rate heterogeneities with 16 γ-rate categories empirically estimated from the data. Quartet puzzling support was estimated from 10,000 replicates.

For cases in which phylogenetic analyses of small genes received low bootstrap support we used GENERAGE (30) to cluster proteins with significant relationships at user-defined *E*-value thresholds. The input to GENERAGE was an all-against-all table of BLAST comparisons of amino acid sequences. GENERAGE uses a Smith–Waterman dynamic programming alignment algorithm to correct for false positive linkages whenever pairwise relationships are not symmetrical. For HLIPs, an *E*-value cutoff of 10^{-14} was used. The clusters containing the phage HLIPs were preserved down to an *E*-value cutoff of 10^{-17}. For plastocyanin and ferredoxin respectively, *E*-value cutoffs of 10^{-26} and 10^{-34} linked the phage proteins with other proteins, whereas, at *E*-value cutoffs of 10^{-28} and 10^{-36}, the respective phage proteins did not cluster with other sequences.

Results

A suite of host photosynthesis genes was found in the three *Prochlorococcus* phage genomes (Fig. 1). The *psbA* gene, encoding the PSII core reaction center protein D1 (hereafter referred to as the D1-encoding gene) and one *hli* gene type encoding the HLIP cluster 14-type protein (*sensu*, see ref. 27)

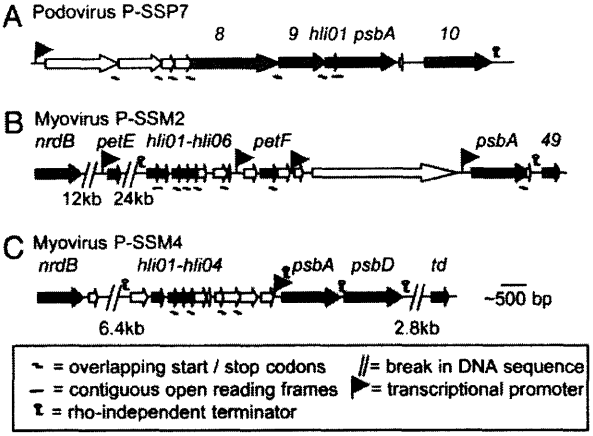


Fig. 1. Arrangement of photosynthesis genes in three *Prochlorococcus* phages. (*A*) Podovirus P-SSP7. (*B*) Myovirus P-SSM2. (*C*) Myovirus P-SSM4. Black bars indicate genes related to photosynthesis, gray bars indicate genes commonly found in phage, and white bars indicate predicted ORFs of unknown function. Genes and their protein designations are as follows: *psbA*, D1; *psbD*, D2; *hli*, HLIP; *petE*, plastocyanin; *petF*, ferredoxin; *8*, T7-like head-to-tail connector; *9*, T7-like capsid assembly protein; *10*, T7-like capsid protein; *nrdB*, T4-like ribonucleotide reductase β-subunit; *49*, T4-like restriction endonuclease VII; and *td*, T4-like thymidylate synthetase.

were present in all three phages. HLIPs are thought to protect the photosynthetic apparatus from excess excitation energy during stressful conditions in cyanobacteria (31). In addition, one of the myoviruses, P-SSM4, contains the *psbD* gene encoding the second PSII core reaction center protein, D2, (hereafter referred to as the D2-encoding gene), whereas the other myovirus, P-SSM2, contains two photosynthetic electron transport genes coding for plastocyanin (*petE*) and ferredoxin (*petF*) (Fig. 1 *B* and *C*). Both myoviruses contain additional gene types from the *hli* multigene family.

The deduced amino acid sequences of the phage photosynthesis genes are highly conserved and therefore have the potential to be functional proteins. The coding sequences of all of these genes are uninterrupted and show a high degree of identity to their host homologs (up to 85% and 95% nucleotide and amino acid identities, respectively; see Table 2 and Figs. 4–8, which are published as supporting information on the PNAS web site). The greatest amino acid divergence in D1 and D2 from all three phages is in the N-terminal leader sequences that do not form part of the functional protein. Furthermore, divergence analyses based on estimates of the frequency of nonsynonymous (K_a) and synonymous (K_s) nucleotide substitutions between phage- and host-encoded genes revealed that the phage genes have diverged relative to those from their hosts (K_s values range from 0.65 to 3.11 and are higher than for *Prochlorococcus* gene pairs; see Table 3, which is published as supporting information on the PNAS web site), but that the majority of nucleotide substitutions did not cause a change in amino acid sequence (K_a/K_s ratios <0.45 for all genes, with values of <0.1 for the D1 and D2 encoding genes; Table 3). Although we cannot rule out the possibility of a recent transfer of these genes from as yet unknown *Prochlorococcus* types with sequences nearly identical to those found in the phages, these findings suggest that the phage-encoded genes, particularly those encoding D1 and D2, have been subjected to strong selective pressure to conserve their amino acid sequences, which is consistent with the hypothesis that they are functional.

All of the photosynthesis genes (with the exception of plastocyanin) are arranged together in the phage genomes. Such gene clustering in phage often suggests that they are expressed at a similar stage of infection (3, 32). In addition, identification of potential promoter and terminator elements suggests that distinct transcriptional units are present. In the genome of P-SSP7, for example, the *hli* and D1-encoding gene may be cotranscribed with the adjacent phage structural genes in a single operon. Most of the genes in this region have overlapping start and stop codons and are flanked by a putative σ^{70} transcriptional promoter and ρ-independent transcriptional terminator (Fig. 1*A*). This arrangement further suggests that the photosynthesis genes are expressed in the latter portion of the lytic cycle, if indeed they are expressed, as is known for structural proteins in other T7-like podoviruses (32). In contrast, the presence of transcriptional terminators flanking the regions containing photosynthetic genes in the myoviruses suggests that they may be transcribed as discrete transcriptional units largely independent of the surrounding phage genes. These hypotheses require further testing by measuring phage gene expression over the infective cycle.

The cyanobacterial origin of the phage D1- and D2-encoding genes is suggested by the presence of certain features in both phage and host genes. Phage D1 proteins contain a 7-aa indel close to the C terminus of the protein (Fig. 4) which is found in all cyanobacterial D1 proteins as well as in nongreen algal plastids (33). Similarly, phage D2 contains a 7-aa indel in the center of the protein that is also found in *Prochlorococcus* MED4 and SS120 (but not in other cyanobacterial or eukaryotic D2 proteins) (Fig. 5). These additional amino acids are not found in the D2 proteins encoded by either *Synechococcus* WH8102 or the *Synechococcus* phage S-PM2 (Fig. 5), suggesting that *Prochlorococcus* phages acquired the D2-encoding gene from *Prochlorococcus* and that *Synechococcus* phages acquired it from *Synechococcus*.

Phylogenetic analyses of the PSII core reaction center proteins further supports the cyanobacterial origin of the phage genes and, along with knowledge of phage host ranges (15), suggests that they were acquired multiple times from their hosts. Phage D1 and D2 proteins clustered with marine cyanobacteria (Fig. 2). Proteins encoded by *Prochlorococcus* phages clustered with *Prochlorococcus*, whereas those from a phage that infects only *Synechococcus* (12) clustered with *Synechococcus*, as did an environmental sequence (BAC9D04) encoding both D1 and phage structural genes (34). Despite low bootstrap support for *Synechococcus* D1 clades in the distance tree, a similar tree topology also emerged from maximum likelihood and maximum parsimony reconstructions (data not shown). Moreover, D1 from two *Prochlorococcus* phages clustered within *Prochlorococcus* clades that match their host range (Fig. 2*A*). However, D1 from the third *Prochlorococcus* phage did not cluster within a specific *Prochlorococcus* clade, suggesting that its gene was acquired from an as yet uncultured *Prochlorococcus* type or has diverged to an extent that prevents identification of the common ancestor. The fact that the phage D1 and D2 proteins are distributed in both the *Prochlorococcus* and *Synechococcus* clades and are largely consistent with their host range suggests that the genes were acquired in independent transfer events from their cyanobacterial hosts (*sensu*; see refs. 2 and 4). These transfer events could have occurred *de novo* between distinct hosts and phages several times, or these genes may have been transferred from host to phage in a process akin to gene conversion subsequent to an ancestral transfer event (see *Discussion*). If host genes in phages resulted from a single ancestral event followed by subsequent vertical or lateral transfers from phage to phage, the phage- and host-encoded genes would have formed monophyletic clades distinct from each other.

Phylogenetic analyses of plastocyanin proteins also suggests that the phage *petE* gene is of cyanobacterial origin (Fig. 9, which is published as supporting information on the PNAS web site). However, the data are not conclusive as to the origin of the phage

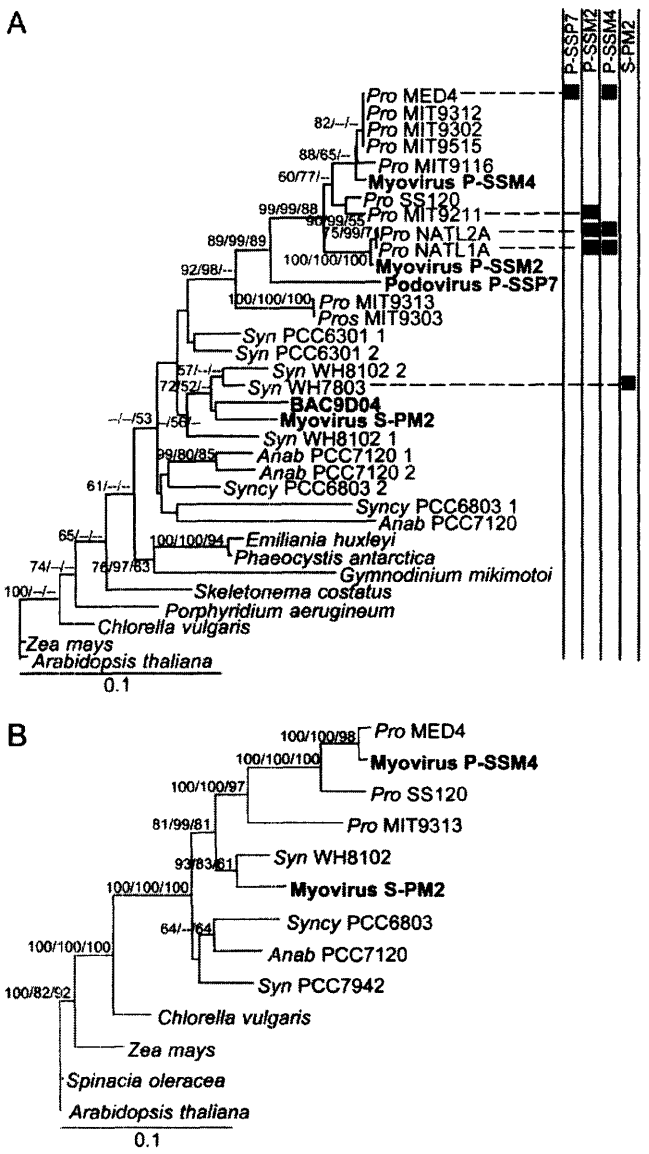


Fig. 2. Distance trees of PSII core reaction center proteins. (*A*) D1 (*psbA*). (*B*) D2 (*psbD*). Phage sequences are shown in bold. The host strains that each phage infects are indicated by black squares. Trees were generated from 244 and 336 amino acids for D1 and D2, respectively (see Figs. 4 and 5). Bootstrap values for distance and maximum parsimony analyses and quartet puzzling values for maximum likelihood analysis >50% are shown at the nodes (distance/maximum likelihood/maximum parsimony). Trees were rooted with *Arabidopsis thaliana* proteins. Essentially, the same topology was obtained when nucleotide trees (third position excluded) were constructed, except for *psbA* from P-SSP7, which clustered with HL *Prochlorococcus*, albeit with low bootstrap support. *Pro*, *Prochlorococcus*; *Syn*, *Synechococcus*; *Anab*, *Anabaena*; *Syncy*, *Synechocystis*.

gene from within the cyanobacteria. The phage protein clusters with filamentous cyanobacteria, but contains a 10-aa indel found only in unicellular cyanobacteria (Fig. 6). GENERAGE analysis did not resolve the phage plastocyanin clustering. Both phylogenetic and GENERAGE analyses of ferredoxin proteins were inconclusive as to the origin of the phage *petF* gene. These results, together with the greater divergence estimates (K_a/K_s) for the phage and *Prochlorococcus petE* and *petF* gene pairs (0.19–0.43) than among *Prochlorococcus* gene pairs (0.03–0.07) (Table 3), suggest that these phage genes either originated from a host for which a close relative does not currently exist in the database or have diverged to an extent that prevents inference as to their origin. The latter model may be due to either significant changes in gene sequence or through the formation of mosaic genes from more than one source. These may be new genes in the making.

Previous analyses of HLIPs in cyanobacterial genomes revealed the presence of genetically diverse types, with distinctly different clusters formed for single and multiple copy HLIPs (27). Genes found in a single copy in each of the four sequenced marine cyanobacterial genomes form four distinct clusters (GR C5, C6, C7, and C8 in Fig. 3) that are interspersed with HLIPs from freshwater cyanobacteria in a large cluster (Fig. 3), whereas multicopy *Prochlorococcus* HLIPs are in a separate cluster (Fig. 3). Although bootstrap support for these two broad clusters is low, all three phylogenetic reconstruction methods resulted in the same separation of the multicopy HLIPs from the other HLIPs (Fig. 3 and Figs. 10 and 11, which are published as supporting information on the PNAS web site), lending support to this tree architecture. When we add the phage HLIPs to this analysis, some interesting patterns appear. Ten of 11 phage HLIPs cluster with those that are encoded by multiple gene copies in *Prochlorococcus*, some with more bootstrap support than others. That these phage HLIPs do not group with those from freshwater cyanobacteria nor with the single-copy marine cyanobacterial HLIP types receives greater bootstrap support (Fig. 3). These results were obtained from four different analyses (distance, maximum parsimony and maximum likelihood phylogenetic analyses, and GENERAGE clustering). Indeed, GENERAGE clusters 7 of 11 phage HLIPs with the four HLIP types encoded by multicopy genes in *Prochlorococcus* genomes (GR 10, GR 12, GR 14, and GR 15), with the remaining four of indeterminate affiliation. As for nearly all of the multicopy HLIP sequences from *Prochlorococcus* (28 of 29), all but one of the phage HLIPs contain a 9-aa signature sequence at the C terminus of the protein that is absent from other cyanobacterial HLIPs (27), further supporting a connection between phage *hli* genes and multicopy *hli* genes in the host.

Although the lack of strong bootstrap support for most of the clustering patterns in Fig. 3 makes it impossible to draw definitive conclusions, the fact that both phage and *Prochlorococcus* HLIPs cooccur in four different clusters suggests that it is likely that *hli* genes have been transferred between hosts and their phages multiple times. Moreover, the clustering of phage HLIPs with a subset of the HLIPs that are found exclusively in *Prochlorococcus* suggests that these distinct *hli* gene types may have been reacquired from phage after a period of evolution, leading to the expansion of the *hli* multigene family in this genus.

Discussion

Our findings, along with those by Millard *et al.* (35), indicate that the presence of photosynthesis genes is common, although not universal (13), among phages that infect both HL and LL *Prochlorococcus* and *Synechococcus*. Photosynthesis genes are found in representatives of both the *Myoviridae*, which predominantly infect *Synechococcus* and LL *Prochlorococcus* ecotypes, and *Podoviridae*, which generally infect a single HL *Prochlorococcus* strain (15). The presence of these genes in the members of the latter viral family, which have greater constraints on carrying extra genetic material than members of the former, supports our suggestion that they play a functional role in the phage.

The gene encoding the PSII core reaction center protein, D1, has been found in all phages with photosynthesis genes, suggesting that it plays a particularly significant role. Other photosynthesis genes were more sporadically distributed among the phages. Genes encoding HLIPs were found in all three *Prochlorococcus* phages but in only one of five *Synechococcus* phages (35). In contrast, the gene encoding the second PSII core

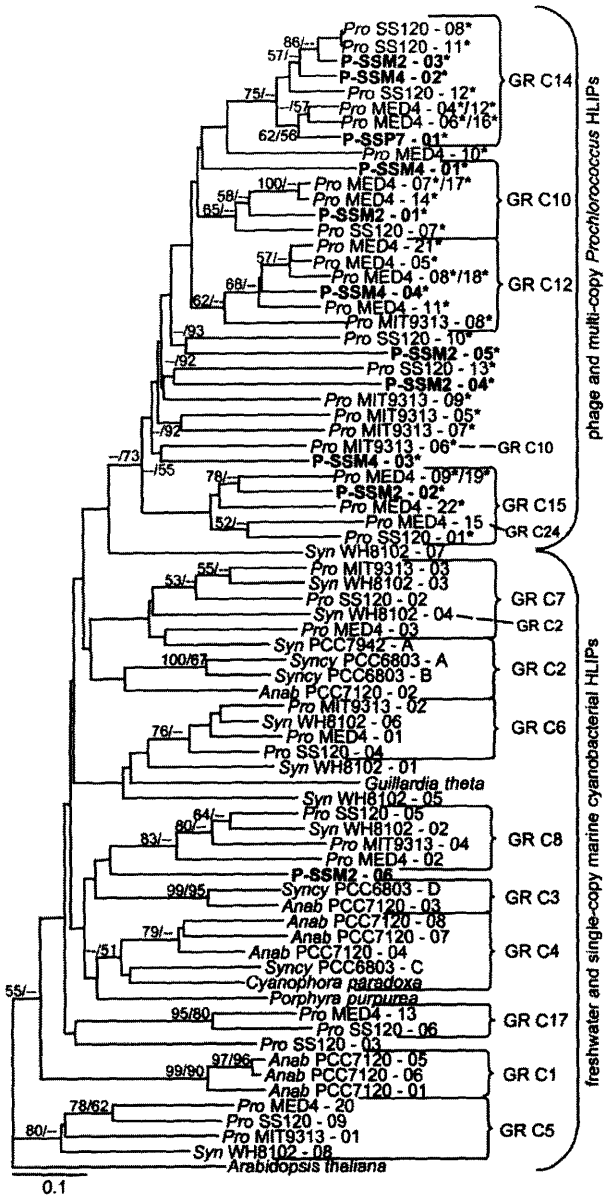


Fig. 3. Distance tree of HLIPs. Phage HLIPs appear in bold. The tree was generated from 36 amino acids (see Fig. 8), with gaps treated as missing data. GENERAGE clusters are indicated to the right of the tree, with cluster designations following ref. 27. Three discrepancies found between GENERAGE and distance tree clustering are indicated by the dashed line and their GR cluster designations. Asterisks denote proteins encoding at least six of the nine amino acids of the C-terminal 9-aa consensus sequence. Bootstrap and quartet puzzling values >50% are shown at the nodes for distance and maximum likelihood analyses, respectively. The tree was rooted with the single HLIP from *A. thaliana*. Abbreviations are as for Fig. 2.

reaction center protein, D2, was found in all *Synechococcus* phages but in only one *Prochlorococcus* phage. The small number of phage genomes presently available for analysis precludes making strong conclusions from this asymmetry, but if the trend holds up, it is likely that phages gain a differential benefit from these two genes that is influenced by genera-level attributes of their cyanobacterial hosts.

Photosynthetic electron transport genes were found in one *Prochlorococcus* phage and in none of the *Synechococcus* phages, whereas the transaldolase gene was found both in *Prochlorococcus* myoviruses (M.B.S., F.R., and S.W.C., unpublished data) and in one *Synechococcus* phage (35). Assuming that these genes are functional, this scattered distribution may have arisen from differential gain and loss resulting from tradeoffs between the burden of carrying such genes and their utility during infection. Alternatively, we may be observing the transient passage of host genes through the phage genome pool.

The arrangements of photosynthesis genes in both *Prochlorococcus* and *Synechococcus* phages have some similar properties (compare Fig. 1 of this study with figure 1 of ref. 35), including adjacent D1- and D2-encoding genes, adjacent HLIP- and D1-encoding genes, and the D1-encoding gene adjacent to a T4-like phage gene encoding gp49. These gene organizations are distinctly different from those in cyanobacterial genomes in which photosynthetic genes are spread throughout the chromosome (20–22, 36). Most noticeably, the D1- and D2-encoding genes are hundreds of thousands of kilobases apart in the hosts. Yet phylogenetic analyses show that the D1 and D2 proteins from *Prochlorococcus* phages cluster with those from *Prochlorococcus*, and, in at least the one *Synechococcus* phage available for analysis, these proteins cluster with those from *Synechococcus* (Fig. 2). Assuming that the ancestral cyanobacterial donors of these genes had a similar gene arrangement to extant cyanobacteria, one likely explanation for these findings is that the genes were acquired from their respective hosts in separate transfer events, integrating at recombination hot-spots within the phage genome and forming advantageous gene arrangements. Alternatively, one early transfer event may have occurred, and the observed gene organization patterns formed before the divergence of these phages. In this latter case, for gene sequences to be similar to that from their respective hosts, they would have to have been swapped between phage and host in a process similar to gene conversion, whereby one gene is replaced by another in a nonreciprocal fashion. The direction of this gene conversion for both the D1- and D2-encoding genes is most likely with the host gene replacing the phage gene, as cyanobacterial phylogenies inferred from these gene products are congruent with those from other genes (Fig. 2) (16–19). This latter scenario would suggest that encoding PSII reaction center genes similar to those from the host is advantageous.

The presence of highly conserved PSII reaction center and *hli* genes in the three *Prochlorococcus* phages suggests that selection pressure has driven their acquisition and retention. The presence of these genes is liable to have important implications for phage–host interactions during infection. It has been known for some time that viral infection of many photosynthetic organisms leads to a decline in photosynthetic rates soon after infection (37, 38). This decline is attributed to damage to the PSII membrane–protein complexes (39, 40) and may be due to oxidative stress caused by an increase in destructive reactive oxygen species subsequent to infection (40). Alternatively, the shut-down of host protein synthesis soon after infection (41) could lead to a reduced supply of the highly turned-over D1 and D2 proteins. However, in many phage-infected unicellular freshwater cyanobacteria, the production of phage progeny depends on photosynthetic activity continuing until just before lysis (42, 43). Phage PSII reaction center proteins may, if expressed, prevent photoinhibitory damage to PSII in *Synechococcus* (12). We further suggest that expression of phage PSII reaction center proteins and the photoprotective HLIPs may help maintain photosynthetic activity during infection of *Prochlorococcus*, leading to increased phage fitness and resulting in selection for cyanophages that encode functional photosynthetic genes. Comparing the fitness of a phage with inactivated photosynthetic genes with that of a wild-type phage would enable one to test this hypothesis.

Our analyses of host genes in phages have implications not only for phage fitness but also for the evolution of the hosts, because there is suggestive evidence that phages may have mediated horizontal gene transfer and, hence, expansion of the *hli* multigene family in the hosts. It has recently been suggested that widely distributed, single-copy genes are resistant to horizontal transfer (44), whereas sporadically distributed multicopy genes are those most likely to have been dispersed by this method (44, 45). The clustering patterns displayed by the *hli* genes in our analyses, although not statistically robust, are consistent with this tenant. Each of the single-copy *hli* gene types common to the four sequenced unicellular marine cyanobacteria (20–22) are likely to have been vertically inherited, as is evident from the conserved gene arrangement surrounding these *hli* types and from their clustering to those from the other marine unicellular cyanobacteria (Fig. 3) (27). In contrast, *hli* gene types present in multiple copies per genome are found in only some *Prochlorococcus* genomes. These latter *hli* gene types are those that are found in the *Prochlorococcus* phage, with at least one phage *hli* gene in each of the four clusters of multicopy *Prochlorococcus hli* gene types (Fig. 3). We therefore suggest that phages have mediated the horizontal dispersal of these multicopy genes among *Prochlorococcus*.

The presence of numerous *hli* genes in *Prochlorococcus* MED4, a HL ecotype, is likely to have influenced its fitness in the surface waters of the open oceans (20, 27, 36). Indeed, upon shifts to high light, cyanobacterial mutants with inactivated *hli* genes are competitively inferior to wild-type cells (31). Our hypothesized phage-mediated expansion of the *hli* multigene family may have contributed to the numerical dominance of the HL ecotype in many ocean ecosystems (46). Other photosynthetic genes found in phages are also present in multiple copies in many cyanobacteria, including the D1-, D2-, and ferredoxin-encoding genes (Table 4, which is published as supporting information on the PNAS web site). The importance of gene duplication in the evolution of new gene functions is well recognized in other systems (47, 48); thus, it would not be surprising if it were playing a role in the evolution of physiological variants within the *Prochlorococcus* cluster.

The exchange of photosynthetic genes between *Prochlorococcus* and their phages could have significant implications for the evolutionary trajectory of both hosts and phages and may represent a more general phenomenon of metabolic facilitation of key host processes. That is, host genes retained in a particular phage could reflect key selective forces in the host environment. Indeed, phosphate sensing and acquisition genes have been found in phages that infect organisms in low phosphate environments (8, 9). Might we also find salt tolerance genes in phages that infect halotolerant organisms and thermal tolerance genes in phages that infect thermophilic organisms? Such coupled evolutionary processes in hosts and phages, if widespread, may play a role in defining host ranges for phages and niche space for hosts, leading to specialization and even speciation.

We thank D. Veneziano, G. Rocap, D. Mead, M. Ermolaeva, and F. Chagnon for assistance with and discussions of various aspects of this work. Sequencing and assembly of the phage genomes were performed by the production sequencing group at the Department of Energy Joint Genome Institute through the Sequence-for-Others Program under the auspices of the Biological and Environmental Research Program from the Office of Science at the Department of Energy, the University of California, Lawrence Livermore National Laboratory (Contract W-7405-ENG-48), Lawrence Berkeley National Laboratory (Contract DE-AC03-76SF00098), Los Alamos National Laboratory (Contract W-7405-ENG-36), and Stanford University (Contract DE-FC02-99ER62873). This research was supported by U.S. Department of Energy Grants DE-FG02-99ER62814 and DE-FG02-02ER63445 and National Science Foundation Grant OCE-9820035 (to S.W.C.).

1. Hendrix, R. W., Lawrence, J. G., Hatfull, G. F. & Casjens, S. (2000) *Trends Microbiol.* **8,** 504–508.
2. Filee, J., Forterre, P. & Laurent, J. (2003) *Res. Microbiol.* **154,** 237–243.
3. Miller, E. S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T. & Ruger, W. (2003) *Microb. Mol. Biol. Rev.* **67,** 86–156.
4. Moreira, D. (2000) *Mol. Microbiol.* **35,** 1–5.
5. Wagner, P. L. & Waldor, M. K. (2002) *Infect. Immun.* **70,** 3985–3993.
6. Casjens, S. (2003) *Mol. Microbiol.* **49,** 277–300.
7. Forterre, P. (1999) *Mol. Microbiol.* **33,** 457–465.
8. Rohwer, F., Segall, A., Steward, G., Seguritan, V., Breitbart, M., Wolven, F. & Azam, F. (2000) *Limnol. Oceanogr.* **45,** 408–418.
9. Miller, E. S., Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Durkin, A. S., Ciecko, A., Feldblyum, T. V., White, O., Paulsen, I. T., Nierman, W. C., *et al.* (2003) *J. Bacteriol.* **185,** 5220–5233.
10. Figueroa-Bossi, N. & Bossi, L. (1999) *Mol. Microbiol.* **33,** 167–176.
11. Pedulla, M. L., Ford, M. E., Houtz, J. M., Karthikeyan, T., Wadsworth, C., Lewis, J. A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N. R., *et al.* (2003) *Cell* **113,** 171–182.
12. Mann, N. H., Cook, A., Millard, A., Bailey, S. & Clokie, M. (2003) *Nature* **424,** 741.
13. Chen, F. & Lu, J. (2002) *Appl. Environ. Microbiol.* **68,** 2589–2594.
14. Waterbury, J. B. & Valois, F. W. (1993) *Appl. Environ. Microbiol.* **59,** 3393–3399.
15. Sullivan, M. B., Waterbury, J. B. & Chisholm, S. W. (2003) *Nature* **424,** 1047–1051.
16. Ferris, M. J. & Palenik, B. (1998) *Nature* **396,** 226–228.
17. Moore, L. R., Rocap, G. & Chisholm, S. W. (1998) *Nature* **393,** 464–467.
18. Rocap, G., Distel, D., Waterbury, J. B. & Chisholm, S. W. (2002) *Appl. Environ. Microbiol.* **68,** 1180–1191.
19. Zeidner, G., Preston, C. M., Delong, E. F., Massana, R., Post, A. F., Scanlan, D. J. & Beja, O. (2003) *Environ. Microbiol.* **5,** 212–216.
20. Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N., Arellano, A., Coleman, M., Hauser, L., Hess, W. R., *et al.* (2003) *Nature* **424,** 1042–1047.
21. Palenik, B., Brahamsha, B., McCarren, J., Waterbury, J., Allen, E., Webb, E. A., Partensky, F. & Larimer, F. W. (2003) *Nature* **424,** 1037–1041.
22. Dufresne, A., Salanoubat, M., Partensky, F., Artiguenave, F., Axmann, I. M., Barbe, V., Duprat, S., Galperin, M. Y., Koonin, E. V., Le Gall, F., *et al.* (2003) *Proc. Natl. Acad. Sci. USA* **100,** 10020–10025.
23. Rohwer, F. & Edwards, R. (2002) *J. Bacteriol.* **184,** 4529–4535.
24. Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., Azam, F. & Rohwer, F. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 14250–14255.
25. Besemer, J., Lomsadze, A. & Borodovsky, M. (2001) *Nucleic Acids Res.* **29,** 2607–2618.
26. Jansson, S., Andersson, J., Kim, S. J. & Jackowski, G. (2000) *Plant Mol. Biol.* **42,** 345–351.
27. Bhaya, D., Dufresne, A., Vaulot, D. & Grossman, A. (2002) *FEMS Microbiol. Lett.* **215,** 209–219.
28. Ermolaeva, M. D., Khalak, H. G., White, O., Smith, H. O. & Salzberg, S. L. (2000) *J. Mol. Biol.* **301,** 27–33.
29. Comeron, J. M. (1999) *Bioinformatics* **15,** 763–764.
30. Enright, A. J. & Ouzounis, C. A. (2000) *Bioinformatics* **16,** 451–457.
31. He, Q., Dolganov, N., Bjorkman, O. & Grossman, A. R. (2001) *J. Biol. Chem.* **276,** 306–314.
32. Dunn, J. J. & Studier, F. W. (1983) *J. Mol. Biol.* **166,** 477–535.
33. Hess, W. R., Weihe, A., Loiseaux-de Goer, S., Partensky, F. & Vaulot, D. (1995) *Plant Mol. Biol.* **27,** 1189–1196.
34. Zeidner, G. & Beja, O. (2004) *Environ. Microbiol.* **6,** 528–534.
35. Millard, A., Clokie, M. R. J., Shub, D. A. & Mann, N. H. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 11007–11012.
36. Hess, W. R., Rocap, G., Ting, C. S., Larimer, F. W., Stilwagen, S., Lamerdin, J. & Chisholm, S. W. (2001) *Photosyn. Res.* **70,** 53–71.
37. Padan, E. & Shilo, M. (1973) *Bacteriol. Rev.* **37,** 343–370.
38. Zaitlin, M. & Hull, R. (1987) *Ann. Rev. Plant Physiol.* **38,** 291–315.
39. Rahoutei, J., Garcia-Luque, I. & Baron, M. (2000) *Physiol. Plant.* **110,** 286–292.
40. Arias, M. C., Lenardon, S. & Taleisnik, E. (2003) *J. Phytopathol.* **151,** 267–273.
41. Sherman, L. A. & Pauw, P. (1976) *Virology* **71,** 17–25.
42. Mackenzie, J. J. & Haselkorn, R. (1972) *Virology* **49,** 517–521.
43. Sherman, L. A. (1976) *Virol.* **71,** 199–206.
44. Lerat, E., Daubin, V. & Moran, N. A. (2003) *PLoS Biol.* **1,** e19.
45. Kurland, C. G., Canback, B. & Berg, O. G. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 9658–9662.
46. Partensky, F., Hess, W. R. & Vaulot, D. (1999) *Microbiol. Mol. Biol. Rev.* **63,** 106–127.
47. Zhang, J. (2003) *Trends Ecol. Evol.* **18,** 292–298.
48. Gu, Z., Steinmetz, L. M., Gu, X., Scharfe, C., Davis, R. W. & Li, W. (2003) *Nature* **421,** 63–66.

DAF-16 Target Genes That Control *C. elegans* Life-Span and Metabolism

Siu Sylvia Lee,[1] Scott Kennedy,[1] Andrew C. Tolonen,[2] Gary Ruvkun[1]*

Signaling from the DAF-2/insulin receptor to the DAF-16/FOXO transcription factor controls longevity, metabolism, and development in disparate phyla. To identify genes that mediate the conserved biological outputs of *daf-2*/insulin-like signaling, we used comparative genomics to identify 17 orthologous genes from *Caenorhabditis* and *Drosophila*, each of which bears a DAF-16 binding site in the promoter region. One-third of these DAF-16 downstream candidate genes were regulated by *daf-2*/insulin-like signaling in *C. elegans*, and RNA interference inactivation of the candidates showed that many of these genes mediate distinct aspects of *daf-16* function, including longevity, metabolism, and development.

The *C. elegans daf-2* pathway controls longevity, metabolism, and development and is orthologous to the mammalian insulin and insulin-like growth factor 1 signaling cascade (*1*). Decreased *daf-2* signaling causes up to threefold life-span extension, increased fat storage, and constitutive arrest at the dauer diapause stage (*2–4*). The *daf-2* mutant phenotypes are suppressed by mutations in *daf-16*, indicating that *daf-16* is negatively regulated by *daf-2* signaling and is the major downstream effector. *daf-16* encodes a forkhead transcription factor (*5*, *6*), which translocates into the nucleus (*7*) and modulates transcription when *daf-2* signaling is abrogated. Multiple *daf-16* transcriptional targets are likely to mediate the diverse functions of *daf-2*/insulin-like signaling. Candidate gene and biochemical approaches revealed that genes encoding superoxide dismutase (*sod-3*), an FK506 binding protein, and a nucleolar protein are regulated by *C. elegans daf-16* (*8*, *9*). The mammalian DAF-16 orthologs (FOXO1, FOXO3, and FOXO4) regulate genes involved in growth control, apoptosis, DNA repair, and oxidative stress (*10*).

Because the pathway from *DAF-2*/insulin receptor to *DAF-16*/FOXO regulates both longevity and metabolism in *C. elegans*, *Drosophila*, and mammals (*1*, *11–14*), DAF-16/FOXO might control homologous target genes in different species to mediate conserved functions. DAF-16 and its mammalian homologs bind to an identical consensus DNA sequence (TTGTTTAC) in vitro (*15*), FOXO3 binds to this consensus site in the *MnSod* promoter in mammalian cells, and binding to this consensus site is required for FOXO3 transactivation of *MnSod* (*16*). We sought to identify DAF-16 transcriptional targets by searching for DAF-16 binding sites in the regulatory regions of genes. Given the high expected rate of detecting a DAF-16 binding site by chance alone [3700 sites expected by chance (*17*)], the search for such a site upstream of a *C. elegans* gene and upstream of its ortholog in a divergent animal species would highlight functional DAF-16 sites in conserved components of the DAF-16 transcriptional cascade. Because the *Drosophila* genome is relatively small and well assembled, we searched for DAF-16 binding sites in *Drosophila* and *C. elegans* orthologous genes.

We surveyed 1 kb upstream of the predicted ATG of 17,085 *C. elegans* and 14,148 *Drosophila* genes and identified 947 *C. elegans* and 1760 *Drosophila* genes that contain at least one perfect-match consensus DAF-16 binding site within the 1-kb promoter region. We then compared these DAF-16 binding site–containing worm and fly genes with a list of 3283 *C. elegans* and *Drosophila* genes that are orthologous to each other (*17*), and identified 17 genes that are orthologous between *Drosophila* and *C. elegans* and bear a DAF-16 binding site within 1 kb of their start codons in both species (Table 1). One *Drosophila* and one *C. elegans* candidate target gene had more than one DAF-16 binding site within the 1-kb region (Table 1).

To examine whether the predicted DAF-16 downstream genes are regulated by insulin signaling through DAF-16, we compared the RNA expression level of each candidate in wild-type, *daf-2(e1370)*, and *daf-2(e1370); daf-16(mgDf47)* animals (Fig. 1). Under conditions in which *sod-3* was robustly induced in the *daf-2* mutant (*18*), we found that 6 of the 17 (~35%) predicted DAF-16 downstream genes were differentially expressed in *daf-2* and *daf-2;daf-16* mutant animals (Fig. 1), indicating that their expression was regulated by insulin signaling through DAF-16. Three of the six genes were expressed at levels three to seven times higher in a *daf-2* mutant than in the wild type or the *daf-2;daf-16* double mutant. This fraction of genes, robustly regulated by the *daf-2* pathway, is much higher than the fraction expected to occur by chance; data from a microarray analysis indicate that 1% of the 16,721 *C. elegans* genes tested were regulated by threefold or more (*19*).

The expression of ZK593.4, T21C12.2, and F43G9.5 was down-regulated and that of C10G11.5, F52H3.5, and C39F7.5 was up-regulated in the *daf-2* mutant in a *daf-16*–dependent manner (Fig. 1 and Table 1). Because the positively and negatively regulated genes bear conserved DAF-16 binding sites and are likely

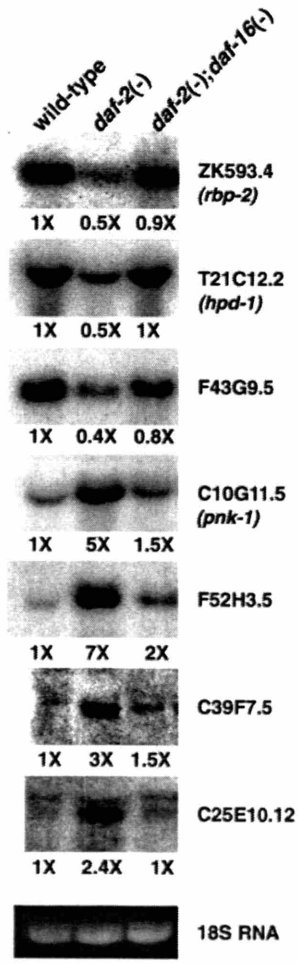


Fig. 1. The expression of seven DAF-16 target candidate genes is regulated by *daf-2*/insulin-like signaling in a *daf-16*–dependent manner. RNA from wild-type, *daf-2(e1370)*, and *daf-2(e1370);daf-16(mgDf47)* animals was tested. Fold differences in expression levels are shown below each band.

[1]Department of Molecular Biology, Massachusetts General Hospital, Department of Genetics, Harvard Medical School, 50 Blossom Street, Boston, MA 02114, USA. [2]Department of Biology, Massachusetts Institute of Technology, 15 Vassar Street, Building 48-424, Cambridge, MA 02139, USA.

*To whom correspondence should be addressed. E-mail: ruvkun@molbio.mgh.harvard.edu

to be direct targets of DAF-16, these results suggest that DAF-16 acts as both a transcriptional activator and a transcriptional repressor, depending on gene context, similar to the forkhead transcription factor LIN-31 (*20*). We failed to detect the expression of three of the DAF-16 downstream gene candidates (E04F6.2, F27C8.1, and T20B3.1), probably because of low endogenous expression. For the remaining eight candidates, we did not detect a noticeable change of expression under the conditions tested. These genes may represent false positives predicted by informatics. Alternatively, some of these genes may be regulated by *daf-2* signaling in a tissue- or stage-specific manner, so that their differential expression was not detected in RNA that was isolated from whole adult animals. Because neuronal *daf-2* signaling is sufficient to regulate *C. elegans* longevity (*21*), analysis based on changes of mRNA levels in whole animals might miss regulatory genes acting in particular tissues, such as neurons. Such regulatory genes would be identified by the informatic search for DAF-16 binding sites. Green fluorescent protein fusions to these candidate genes might reveal whether they are expressed in particular tissues and whether their expression is regulated by *daf-2* signaling.

To examine whether the candidate DAF-16 downstream genes are biologically important targets of *daf-2* signaling, we used RNA interference (RNAi) (*22*) in wild-type or *rrf-3(pk1426)* strains and *daf-2(e1370)* or *age-1(hx546)* strains to reduce the expression of each gene and to determine whether life-span, dauer arrest, and fat storage were affected. *rrf-3(pk1426)* animals are hypersensitive to RNAi (*23*) but are otherwise wild type in our functional assays (*18*). *age-1(hx546)* animals live long but do not arrest as dauer constitutively at 25°C (*24*), and they represent a sensitized genetic background with a slight reduction of *daf-2* pathway signaling. We expected RNAi inactivation of the genes that are down-regulated in the *daf-2* mutant to promote *daf-2* mutant phenotypes, including life-span extension, dauer arrest, and increased fat storage, and we expected RNAi inactivation of the genes up-regulated in the *daf-2* mutant to suppress the *daf-2* mutant phenotypes.

RNAi of ZK593.4 (*rbp-2*) and T21C12.2 (*hpd-1*), genes that are down-regulated in the *daf-2* mutant, caused *rrf-3(pk1426)* animals to live considerably longer than those undergoing control RNAi or RNAi of an unrelated gene (Fig. 2, A and B) (*18*). The life-span extension was modest compared to that of RNAi inactivation of *daf-2* (a 30% increase in mean life-span for *rbp-2* or *hpd-1* RNAi as compared with a 100% increase for *daf-2* RNAi). *rbp-2* and *hpd-1* might constitute a fraction of the DAF-16 transcriptional cascade. RNAi of *hpd-1* also promoted dauer arrest under sensitized conditions (Table 2), whereas RNAi of *rbp-2* did not. Although RNAi inactivation of *hpd-1* or *rbp-2* in wild-type animals did not induce dauer arrest, *hpd-1* RNAi inhibited dauer recovery of *daf-2(e1370)* at 22°C, compared with control or *rbp-2* RNAi (Table 2) (*18*). *rbp-2* might specifically regulate life-span, whereas *hpd-1* might have a broader role in *daf-16* regulation of both dauer arrest and longevity.

rbp-2 encodes a homolog of the mammalian RB binding protein 2 (RBP2), which is implicated in gene expression control and chromatin remodeling (*25*, *26*). *sir-2*, which modulates longevity in yeast and in *C. elegans* (*27*, *28*), encodes a histone deacetylase, also highlighting a role for chromatin remodeling in longevity control. *rbp-2* might be regulated by DAF-16 to further modify chromatin when *daf-2* signaling is decreased. *hpd-1* encodes the enzyme 4-hydroxyphenylpyruvate dioxygenase involved in the catabolism of phenylalanine and tyrosine to fumarate and acetoacetate. Insulin signaling might regulate amino acid degradation and contribute to the coupling of nutritional status and amino acid turnover. In *Drosophila*, reduced function of the *Indy* transporter, which carries metabolic intermediates including fumarate, markedly extends life-span (*29*, *30*). *hpd-1* might also affect the balance of metabolic intermediates such as fumarate and influence longevity through a mechanism similar to that of *Indy* in *Drosophila*. Alternatively, *hpd-1* encodes a dioxygenase in a degradation pathway from tyrosine; mutations in this dioxygenase could affect tyrosine pools and in turn affect dopaminergic signaling, or they could affect free radical production, an expected byproduct of dioxygenases.

pnk-1 (C10G11.5), a gene up-regulated in the *daf-2* mutant, encodes one of the two pantothenate kinases in *C. elegans*, the rate-limiting enzymes in coenzyme A synthesis. Because

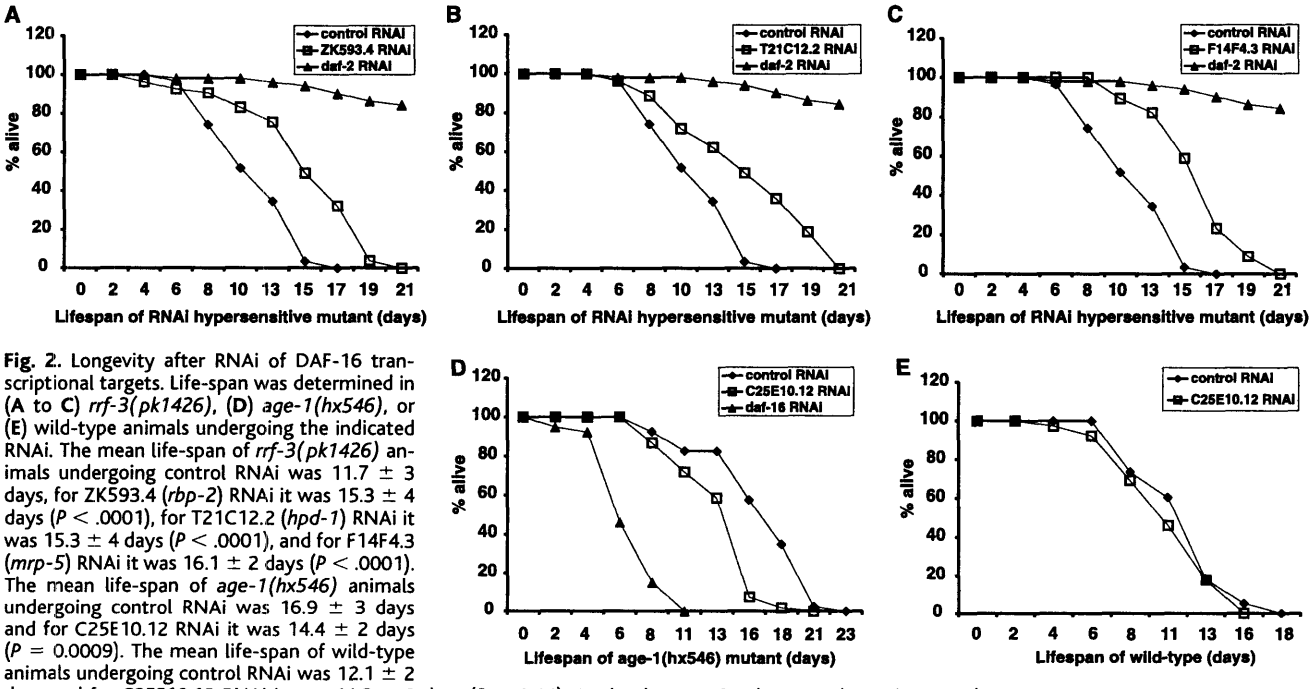


Fig. 2. Longevity after RNAi of DAF-16 transcriptional targets. Life-span was determined in (**A** to **C**) *rrf-3(pk1426)*, (**D**) *age-1(hx546)*, or (**E**) wild-type animals undergoing the indicated RNAi. The mean life-span of *rrf-3(pk1426)* animals undergoing control RNAi was 11.7 ± 3 days, for ZK593.4 (*rbp-2*) RNAi it was 15.3 ± 4 days ($P < .0001$), for T21C12.2 (*hpd-1*) RNAi it was 15.3 ± 4 days ($P < .0001$), and for F14F4.3 (*mrp-5*) RNAi it was 16.1 ± 2 days ($P < .0001$). The mean life-span of *age-1(hx546)* animals undergoing control RNAi was 16.9 ± 3 days and for C25E10.12 RNAi it was 14.4 ± 2 days ($P = 0.0009$). The mean life-span of wild-type animals undergoing control RNAi was 12.1 ± 2 days and for C25E10.12 RNAi it was 11.3 ± 3 days ($P = 0.24$). Student's *t* test *P* values are shown in parentheses.

Table 1. *daf-16* transcriptional target candidates predicted by the survey of 1 kb upstream of each ATG in both *C. elegans* and *Drosophila* genomes. n.c., no change from control; n.d., DAF-16 site was not searched for because no clear starting ATG of the *C. briggsae* homolog was determined; dash indicates no expression detected.

Gene	Homology	DAF-16 site* (*C. elegans*)	DAF-16 Site* (*Drosophila*)	DAF-16 site* (*C. briggsae*)	mRNA in *daf-2(−)*	RNAi inactivation phenotype		
						Life-span	Dauer	Fat storage
C08B11.8	Similar to yeast glucosyltransferase	48	324	~500†	n.c.	n.c.	n.c.	n.c.
C10G11.5 (*pnk-1*)	Pantothenate kinase	389	334	~300	5X	Shortened	n.c.	Reduced
C39F7.5	Cytochrome c heme binding site	375	299	~350†	3X	n.c.	n.c.	n.c.
E04F6.2	Unknown	240	150	~400†	–	n.c.	n.c.	n.c.
F14F4.3 (*mrp-5*)	ABC transporter	111, 920	567	~900†	n.c.	Extended	Enhanced	n.c.
F27C8.1	Amino acid transporter	915	828	~2500†	–	n.c.	n.c.	n.c.
F43G9.5	Subunit of pre-mRNA cleavage factor I	371	609	~350	0.4X	n.c.	n.c.	n.c.
F52H3.5	Similar to yeast stress-induced protein	763	982, 400	~2200†	7X	n.c.	n.c.	n.c.
F54D5.7	Acyl-CoA dehydrogenase	513	825	n.d.	n.c.	n.c.	n.c.	n.c.
K07B1.3	Mitochondrial carrier	895	69	n.d.	n.c.	n.c.	n.c.	n.c.
T20B3.1	Carnitate acyltransferase	536	96	n.d.	–	n.c.	n.c.	n.c.
T20B5.3	Hyaluronoglucosaminidase	588	507	~500, ~600	n.c.	n.c.	n.c.	n.c.
T21C12.2 (*hpd-1*)	Hydroxyphenylpyruvate dioxygenase	983	175	~1700†	0.5X	Extended	Enhanced	n.c.
T23B12.4	Similar to yeast glucose repressible protein MAK10	90	633	~100	n.c.	n.c.	n.c.	n.c.
Y106G6H.7	Mitochondrial energy transfer protein signature	71	343	~70	n.c.	n.c.	n.c.	n.c
ZC506.3	Phosphatidylserine synthase I	702	358	~630†	n.c.	n.c.	n.c.	n.c.
ZK593.4 (*rbp-2*)	Similar to retinoblastoma binding protein 2	27	716	~2700†	0.5X	Extended	n.c.	n.c.

*Nucleotide position upstream of the predicted ATG. †These binding sites contain one mismatch from the consensus that retains DAF-16 binding in vitro.

coenzyme A is key to fat metabolism, we examined fat storage in *pnk-1* RNAi animals, using Nile Red staining (*31*). RNAi of *pnk-1* caused dramatic reduction of fat storage in the intestine of wild-type or *daf-2* mutant animals (Fig. 3). Thus, increased fat storage in *daf-2* mutants might be partly a result of *pnk-1* upregulation. RNAi of *pnk-1* also dramatically shortened wild-type and *daf-2* mutant adult lifespan (*23*), suggesting that inactivation of *pnk-1* compromises the health of animals.

RNAi inactivation of F43G9.5, C39F7.5, and F52H3.5 did not affect dauer arrest, lifespan, or fat storage under the conditions tested (Table 1). It is possible that RNAi did not reduce their expression to a level necessary to produce a phenotype. Alternatively, these genes might have more subtle functions in *daf-2* regulation of metabolism or longevity, or other genes might provide redundant functions to compensate for their inhibition.

RNAi inactivation of F14F4.3 (*mrp-5*) promoted life-span extension and dauer arrest (Fig. 2C and Table 2). Although we did not detect differential expression of *mrp-5* in *daf-2* as compared with *daf-2;daf-16*, it is possible that *daf-2* signaling regulates *mrp-5* expression in specific tissues or at specific times, and this was not detected under our experimental conditions. *mrp-5* encodes an adenosine triphosphate–binding cassette, subfamily C transporter. Members of this subclass are implicated in modulating insulin secretion and in transport of nucleoside analogs and glutathione (*32*). *mrp-5* might act as a feedback regulator of insulin secretion to influence life-span and dauer arrest. Alternatively, *mrp-5* might also affect life-span by regulating glutathione transport and antioxidant defense.

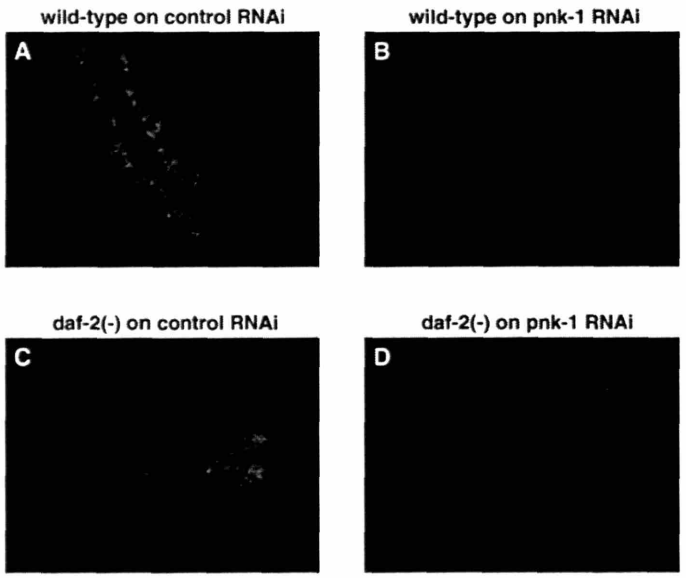


Fig. 3. RNAi of *pnk-1* reduced lipid storage. Nile Red staining of wild-type or *daf-2(e1370)* animals undergoing the indicated RNAi is shown. (**A** and **C**) Nile Red staining showing intestinal fat droplets in wild-type or *daf-2(e1370)* animals. (**B** and **D**) Reduced Nile Red staining in wild-type or *daf-2(e1370)* animals undergoing RNAi against *pnk-1*.

The genome of the nematode *C. briggsae* has been sequenced. Because *C. elegans* and *C. briggsae* are more closely related than *C. elegans* and *Drosophila* (*33*), we examined whether the DAF-16 binding site that is conserved between orthologous *C. elegans* and *Drosophila* genes is also conserved in the promoters of the *C. briggsae* homologs. Among the 14 *C. elegans* DAF-16 downstream gene candidates that have a close *C. briggsae* homolog, 5 genes have a DAF-16 binding site within 1 kb of the predicted ATG, and 5 genes have a DAF-16 binding site containing one mismatch, with specific substitutions that would retain DAF-16 binding (*15*) (Table 1). For the remaining four DAF-16 downstream gene candidates, we found DAF-16 binding sites only when intergenic regions further upstream were surveyed (up to 2.7 kb) (Table 1). It is possible that DAF-16 binding sites drift and relocate frequently, and for some of the *C. elegans* and *Drosophila* genes that bear DAF-16 binding sites within 1 kb of the ATG, the counterparts in *C. briggsae* might have relocated the binding site away from the 1-kb promoter region.

This informatic search for DAF-16 sites within the 1 kb upstream of the ATG is not

Table 2. Dauer formation of *daf-2(e1370)* animals at 22°C under the indicated RNAi conditions.

Day 4 at 22°C	Control RNAi	*daf-2* RNAi	T21C12.2 RNAi	F14F4.3 RNAi
daf-2(e1370) adult	100%	0	10%	2%
daf-2(e1370) dauer	0	100%	90%	98%

yet saturating. A more complete search would cover the intergenic regions that are located upstream of the worm and fly genes, as well as large introns near the ATG. This would make the *C. elegans* search space about five times larger and the *Drosophila* search space about six times larger (*34*). In addition, allowed mismatches in the consensus that retain DAF-16 binding could also be searched. However, because enhancer elements are highly enriched in the region proximal to the start codon, our 1-kb search is a reasonable first stage of the analysis.

We have thus far expanded the informatic search to cover 1.5 kb of the worm promoter and 5 kb of the fly promoter, and this yielded 66 additional DAF-16 downstream gene candidates (table S1). Inspection of the molecular identity of the predicted candidates led us to focus on candidate C25E10.12, which encodes a serine/threonine phosphatase. The expression of C25E10.12 was up-regulated in the *daf-2* mutant in a *daf-16*–dependent manner (Fig. 1). When C25E10.12 was RNAi-inactivated, it shortened the life-span of *age-1(hx546)* animals (Fig. 2D) but did not alter the life-span of wild-type animals (Fig. 2E), indicating that C25E10.12 RNAi specifically suppressed the life-span extension caused by reduced *daf-2*/insulin signaling.

Continued characterization of DAF-16 targets conserved between disparate animal taxa will identify additional key mediators of the conserved longevity and metabolism functions of insulin signaling.

Note added in proof: We searched *C. elegans* and *Drosophila* intergenic regions and detected 115 orthologous genes that each contain at least one DAF-16 site in the region between the start codon and the next gene upstream (table S3).

References and Notes

1. C. E. Finch, G. Ruvkun, *Annu. Rev. Genom. Hum. Genet.* **2**, 435 (2001).
2. C. Kenyon, J. Chang, E. Gensch, A. Rudner, R. Tabtiang, *Nature* **366**, 461 (1993).
3. K. D. Kimura, H. A. Tissenbaum, Y. Liu, G. Ruvkun, *Science* **277**, 942 (1997).
4. W. G. Wadsworth, D. L. Riddle, *Dev. Biol.* **132**, 167 (1989).
5. S. Ogg *et al.*, *Nature* **389**, 994 (1997).
6. K. Lin, J. B. Dorman, A. Rodan, C. Kenyon, *Science* **278**, 1319 (1997).
7. R. Y. Lee, J. Hench, G. Ruvkun, *Curr. Biol.* **11**, 1950 (2001).
8. Y. Honda, S. Honda, *FASEB J.* **13**, 1385 (1999).
9. H. Yu, P. L. Larsen, *J. Mol. Biol.* **314**, 1017 (2001).
10. K. U. Birkenkamp, P. J. Coffer, *Biochem. Soc. Trans.* **31**, 292 (2003).
11. R. S. Garofalo, *Trends Endocrinol. Metab.* **13**, 156 (2002).
12. J. Nakae *et al.*, *Nature Genet.* **32**, 245 (2002).
13. M. Holzenberger *et al.*, *Nature* **421**, 182 (2003).
14. M. Bluher, B. B. Kahn, C. R. Kahn, *Science* **299**, 572 (2003).
15. T. Furuyama, T. Nakazawa, I. Nakano, N. Mori, *Biochem. J.* **349**, 629 (2000).
16. G. J. Kops *et al.*, *Nature* **419**, 316 (2002).
17. Materials and methods are available as supporting material on *Science* Online.
18. S. S. Lee, S. Kennedy, A. C. Tolonen, G. Ruvkun, data not shown.
19. S. S. Lee, S. Kennedy, G. Ruvkun, unpublished results.
20. P. B. Tan, M. R. Lackner, S. K. Kim, *Cell* **93**, 569 (1998).
21. C. A. Wolkow, K. D. Kimura, M. S. Lee, G. Ruvkun, *Science* **290**, 147 (2000).
22. L. Timmons, A. Fire, *Nature* **395**, 854 (1998).
23. F. Simmer *et al.*, *Curr. Biol.* **12**, 1317 (2002).
24. D. B. Friedman, T. E. Johnson, *Genetics* **118**, 75 (1988).
25. S. W. Chan, W. Hong, *J. Biol. Chem.* **276**, 28402 (2001).
26. J. J. Gildea, R. Lopez, A. Shearn, *Genetics* **156**, 645 (2000).
27. L. Guarente, *Trends Genet.* **17**, 391 (2001).
28. H. A. Tissenbaum, L. Guarente, *Nature* **410**, 227 (2001).
29. F. Knauf, B. Rogina, Z. Jiang, P. S. Aronson, S. L. Helfand, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14315 (2002).
30. B. Rogina, R. A. Reenan, S. P. Nilsen, S. L. Helfand, *Science* **290**, 2137 (2000).
31. K. Ashrafi *et al.*, *Nature* **421**, 268 (2003).
32. M. Dean, *The Human ATP-Binding Cassette (ABC) Transporter Superfamily* (National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD, 2002).
33. A. Coghlan, K. H. Wolfe, *Genome Res.* **12**, 857 (2002).
34. M. D. Adams *et al.*, *Science* **287**, 2185 (2000).
35. We thank Y. Grad for the worm and fly ortholog list; R. Kamath, A. Fraser, and J. Ahringer for the RNAi library; K. Ashrafi for assistance with Nile Red; I. Ruvinsky, H. Mak, D. Kim, members of the Ruvkun lab, and B. Weiss for discussions; P. Chisholm for support and encouragement; and the *C. elegans* Genetics Center for providing strains. Supported by a Damon Runyon postdoctoral fellowship to S.S.L., an NIH postdoctoral fellowship to S.K., and grants from the Ellison Research Foundation and NIH to G.R.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1083614/DC1
Materials and Methods
Tables S1 to S3

18 February 2003; accepted 1 April 2003
Published online 10 April 2003;
10.1126/science.1083614
Include this information when citing this paper.

Requirement of Cks2 for the First Metaphase/Anaphase Transition of Mammalian Meiosis

Charles H. Spruck,[1]* Maria P. de Miguel,[2]* Adrian P. L. Smith,[1] Aimee Ryan,[3]† Paula Stein,[4] Richard M. Schultz,[4] A. Jeannine Lincoln,[2] Peter J. Donovan,[2] Steven I. Reed[1]‡

We generated mice lacking Cks2, one of two mammalian homologs of the yeast Cdk1-binding proteins, Suc1 and Cks1, and found them to be viable but sterile in both sexes. Sterility is due to failure of both male and female germ cells to progress past the first meiotic metaphase. The chromosomal events up through the end of prophase I are normal in both *CKS2*$^{-/-}$ males and females, suggesting that the phenotype is due directly to failure to enter anaphase and not a consequence of a checkpoint-mediated metaphase I arrest.

Like the mitotic cell cycle, the meiotic cell cycle is controlled by regulating the activity of maturation promoting factor (MPF), the complex of cyclin B and Cdk1. But the need to produce a haploid cell has necessitated unique modifications to the cell cycle. Two aspects of meiosis that distinguish it from mitosis are the behavior of sister chromatids in meiosis I and the transition from one metaphase (metaphase I) to a second (metaphase II) without intervening DNA synthesis (*1*–*3*). Changes in cell-cycle regulation observed in meiosis are brought about in part by modifications of MPF activity, most likely through interaction with regulatory proteins. Of the panoply of Cdk1-interacting proteins, among the least well understood are the Cks homologs. In both fission (*Schizosaccharomyces pombe*) and budding (*Saccharomyces cerevisiae*) yeast, depletion of Cks homologs leads to mitotic arrest (*4*, *5*). Immunodepletion of Xe-p9, a *Xenopus* Cks homolog, from egg extracts prevents both entry into and exit from mitosis, depending on the experimental design (*6*). However, no conclusive evidence

[1]Department of Molecular Biology, MB-7, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA. [2]Kimmel Cancer Center, Thomas Jefferson University, 233 South 10th Street, Philadelphia, PA 19107, USA. [3]Department of Medicine, University of California, San Diego, School of Medicine, La Jolla, CA 92093, USA. [4]Department of Biology, University of Pennsylvania, 415 South University Avenue, Philadelphia, PA 19104–6018, USA.

*These authors contributed equally to this work.
†Present address: Department of Human Genetics, McGill University, Montreal, QC H3H 1P3, Canada.
‡To whom correspondence should be addressed. E-mail: sreed@scripps.edu

Appendix IV: Antibiotic Sensitivities of *Prochlorococcus* MED4 and MIT9313

INTRODUCTION

Methods to transfer foreign DNA into prokaryotic cells such as interspecific conjugation and transformation are quite inefficient. Even under the best conditions with *E. coli*, only a tiny fraction of a cell population will be genetially transformed in a given experiment. Isolation of genetic mutants thus requires a means to select cells that received the foreign DNA away from those that did not. Typically, this selection is accomplished using antibiotics. The foreign DNA is engineered to contain antibiotic resistance genes that, when expressed in the host cell, allow them to survive under conditions where the wild-type cells will not. Two of the most commonly used antibiotic markers used in cyanobacteria are kanamycin and chloramphenicol (Elhai and Wolk, 1988; Tsinoremas et al, 1994). In order to use these antibiotics in genetic selections with *Prochlorococcus*, we needed to determine appropriate antibiotic concentrations. The ideal antibiotic concentration is high enough to kill wild-type cells without being so high as to overwhelm the level of resistance endowed by an antibiotic resistance gene. The goal of these experiments was to determine the sensitivity levels of two axenic Prochlorococcus strains, MED4, and MIT9313, to kanamycin and chloramphenicol.

METHODS AND MATERIALS

In order to determine appropriate antibiotic concentrations for genetic screening, we transferred late log-phase cells into fresh medium containing various concentrations of antibiotics. One ml of a late-log phase culture (approximately 10^8 cells) was transferred into 20 mls of fresh medium containing antibiotics. The experiments were designed this way so as to be as similar as possible to how an antibiotic selection would be conducted following conjugation. Following transfer into medium containing antibiotics, the growth of the cells was monitored by chlorophyll fluorescence using a Turner fluorometer.

In other cyanobacteria used in genetic studies, kanamycin is generally applied at either 25 or 50 μg ml^{-1} (Elhai and Wolk, 1988). We tested these levels in MED4 (Fig. 1A) and MIT9313 (Fig. 2). The kanamycin resistance gene from Tn5 also gives resistance to the related antibiotic neomycin. Because kanamycin did not prove to be a potent antibiotic for MED4, we also tested the efficacy of neomycin at level typically used with prokaryotes (Fig. 1A). In parallel, the sensitivies of

Prochlorococcus MED4 cultures to chloramphenicol were also tested. Wolfgang Hess's lab (Wolfgang Hess, pers. comm.) had found 0.5 μg ml^{-1} to be a strong selection against MED4. We thus tested chloramphenicol at this concentration (Fig. 1B). Because chloramphenicol is solubilized in ethanol, we independently tested the toxicity of two different concentrations of ethanol: 1/1000x and 1/30,000x. The 1/1000 x ethanol concentrations correspond to adding 20 μl ethanol to a 20 ml culture. In order to calculate the number of resistant cells in a given culture, we converted the chlorophyll fluorescence measurements to cell counts using flow cytometry (chlorophyll fluorescence of 500 units equals 10^8 cell ml^{-1}). Because the chlorophyll content of the cell can vary with growth phase, it is a simplification to convert between chlorophyll and cells concentration with a single constant. However, because all chlorophyll measurements were taken in log phase, these conversions provide a reasonable approximation. We calculated the growth rates both as the doublings day^{-1} and as μ (day^{-1}) where μ is doublings day^{-1} multiplied by ln(2).

RESULTS

We found that kanamycin was not an effective selective agent against MED4 at levels used with other cyanobacteria (Fig. 1A). MED4 growth was delayed relative to controls in both the 25 and 50 μg ml^{-1} treatments, but ultimately the cultures grew. We found that neomycin provided an even poorer selections against MED4 (Fig. 1A). We estimated the initial number of resistant cells in the MED4 cultures at each kanamycin concentration by fitting a linear regression to the cells numbers and extrapolating the number of resistant cells present at time zero (Fig. 3). For the 25 μg ml^{-1} kanamycin treatment, a linear regression was fit using the cell numbers at 14, 20, and 24 days ($R = 0.38*t + 13.81$ where R is the log(resistant cells ml^{-1}) and t is days). Based on the intercept with the ordinate axis (cells ml^{-1} at time zero) we estimated that there were initially $9.94x10^5$ cell ml^{-1} resistant cells. This supports that 14% of the cells were resistant to 25 μg ml^{-1} kanamycin. A linear regression was also fit to the data for 50 μg ml^{-1}kanamycin using the data at 14, 20, and 24 days ($R = 0.15*t + 12.06$). This equation supports that there were initially $1.72x10^5$ resistant cells ml^{-1}; 2% of the cells were resistant to 50 μg ml^{-1}kanamycin. It is also notable that resistant cells grow more slowly at higher kanamycin concentrations. Based on the slope of the linear regressions, we calculated that MED4 grew at 0.73 doublings day^{-1} ($\mu=0.51$ day^{-1}) in the absence of kanamycin whereas they grew at 0.26 doublings day^{-1} ($\mu=0.18$ day^{-1})and 0.22 doublings day^{-1} ($\mu=0.15$ day^{-1}) in kanamycin 25 μg ml^{-1} and 50 μg ml^{-1}, respectively.

In contrast to MED4, we found that 50 μg ml^{-1} kanamycin did provide a strong

selection against MIT9313 (Fig. 2). We observed that while cells grew more slowly in 15 μg ml^{-1} kanamycin relative to no-kanamycin controls (0.33 versus 0.17 doublings day^{-1}) , log phase growth began immediumtely in both treatments. MIT9313 cultures containing 25 μg ml^{-1} kanamycin initially declined in fluorescence, but ultimately grew under selection. We fit a linear regression using the data points once growth had begun (R = 0.06*t + 12.31) which revealed that cells grew at a rate of 0.08 doublings day^{-1} (μ=0.06 day^{-1}) in 25 μg ml^{-1} kanamycin. We used the linear regression to extrapolate the number of kanamycin-resistant cells at t=0, thereby calculating that 6% of the cells were kanamycin-resistant. Even after 90 days, we observed no growth in the 50 μg ml^{-1} kanamycin treatment. It is not feasible from these experiments to formally conclude that no MIT9313 cells were kanamycin resistant in the 50 μg ml^{-1} treatment. However, from a practical standpoint we can conclude that no growth was observed for 90 days in 50 μg ml^{-1} kanamycin.

Because kanamycin and neomycin failed to provide a strong selection against MED4, we also tested the chloramphenicol sensitivities of MED4. We confirmed the Hess lab's findings that 0.5 μg ml^{-1} chloramphenicol did provide a strong selection against MED4 (Fig. 1B). However, we also observed that as little as 20 μl of ethanol can reduce the growth rate of MED4 (Fig. 1B). It is thus possible that some of the toxicity resulting from adding chloramphenicol comes from the ethanol solvent. We were unable to estimate the number of resistant cells in the 0.5 μg ml^{-1} chloramphenicol treatment because no growth was observed. We are thus unable to formally rule out that spontaneous chloramphenicol resistence is possible. However, a spontaneous mutation rate this low would be expected to be much lower than the rate of conjugal transfer of a plasmid.

CONCLUSIONS

We can conclude from these experiments that kanamycin and neomycin are not viable selections to be used in genetic experiments with MED4 (Fig. 1A). Although they delayed the growth of cultures relative to no-antibiotic controls, MED4 cultures ultimately grew under kanamycin and neomycin selection for all levels tested. In contrast, 0.5 μg ml^{-1} chloramphenicol appears to be a viable means to select against MED4 cells (Fig. 1B). Thus, plasmids designed for MED4 genetics should contain the chloramphenicol acetyl-transferase gene. In contrast to MED4, 50 μg ml^{-1} kanamycin did provide a stong selection against MIT9313 (Fig. 2). Thus, it would be reasonable to use plasmids containing the kanamycin resistance gene in experiments to develop MIT9313 genetics.

REFERENCES

Elhai, J and Wolk, P. (1988). “Conjugal transfer of DNA to cyanobacteria”. Methods in Enzymology. 167: 747-754.

Tsinoremas, NF., Kutach, AK., Strayer, CA. (1994). “Efficient gene-transfer in Synechococcus sp. strains PCC7942 and PCC6301 by interspecies conjugation and chromosomal recombination”. Journal of Bacteriology 176 (21): 6764-6768.

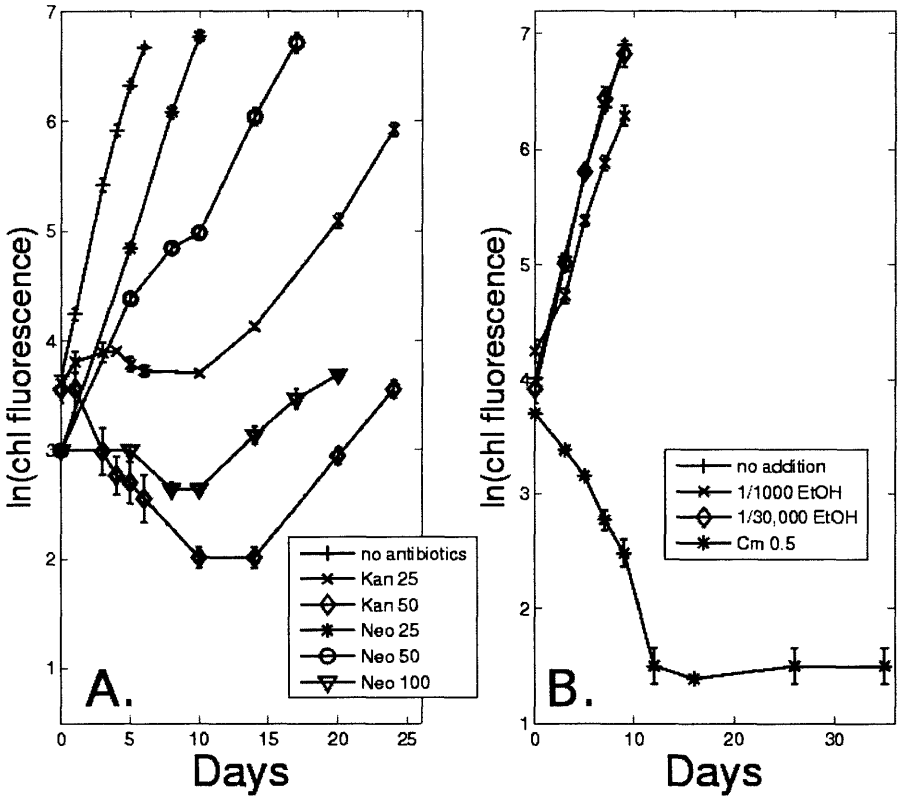


Fig. 1. MED4 sensitivities to kanamycin and neomycin (**A**) and to chloroamphenicol and ethanol (**B**). **A**. Growth of MED4 was monitored after addition of kanamycin and neomycin at concentrations typically used with other related cyanobacteria. Kanamycin was tested at 25 and 50 μg ml^{-1}. Neomycin was tested at 25, 50, and 100 μg ml^{-1}. **B**. MED4 sensitivies to chloramphenicol and ethanol. Chloramphenicol was added at the concentration of 0.5 μg ml^{-1}. Because chloramphenicol is solvated using ethanol, ethanol only controls were also included to examines its toxicity independently. Ethanol was added at two concentrations: 1/1000 (i.e. 20 μl added to 20 ml culture) and 1/30,000.

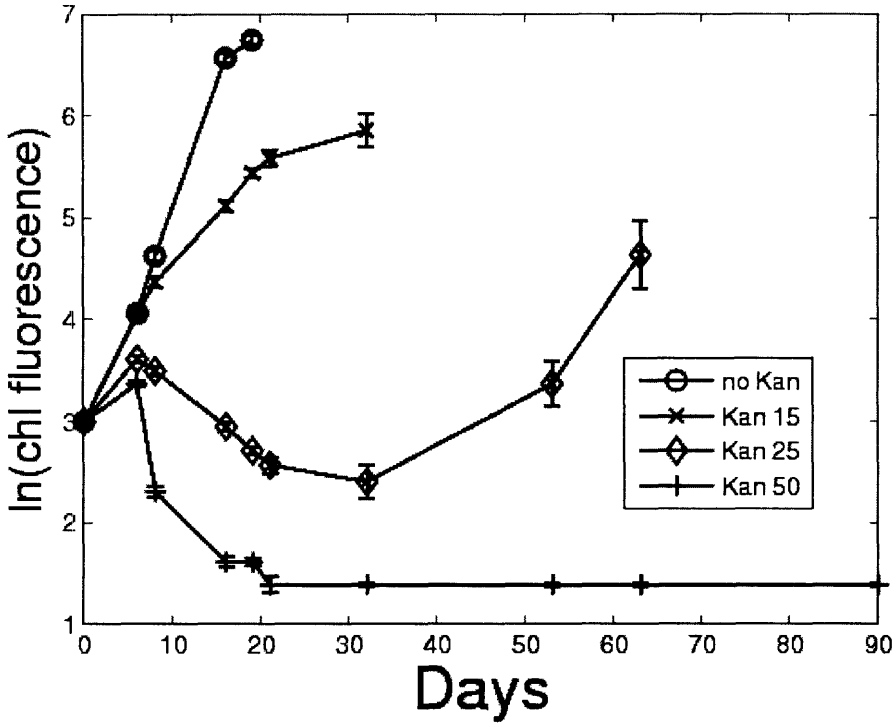


Fig. 2. MIT9313 sensitivity to kanamycin. Growth of MIT9313 was monitored after addition of kanamycin at concentrations typically used with other related cyanobacteria. Kanamycin was tested at 15, 25 and 50 μg ml^{-1}.

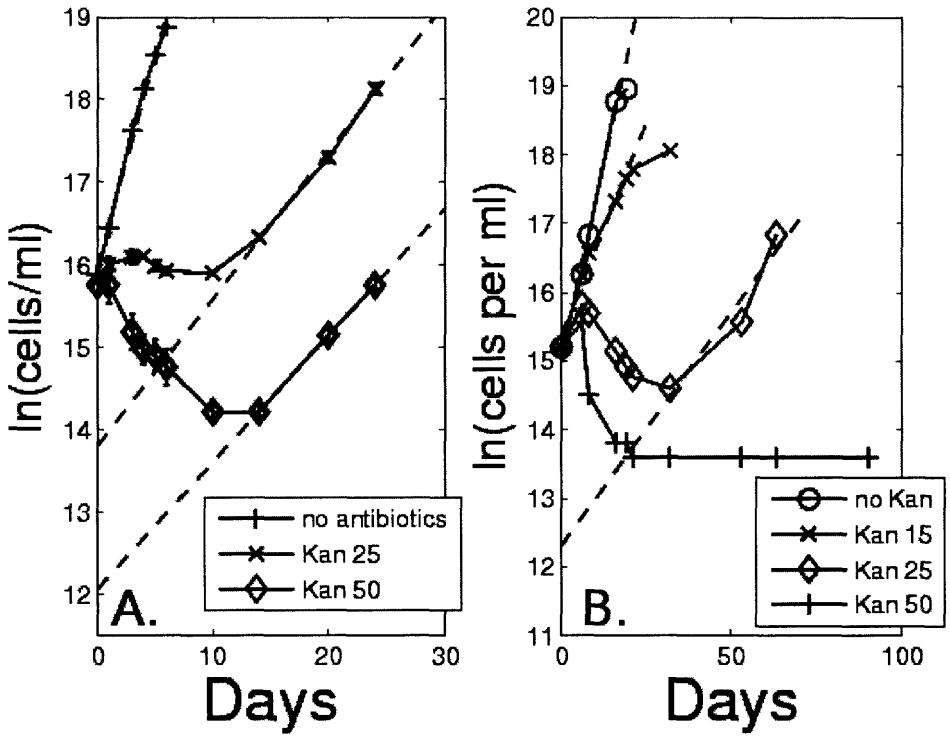


Fig. 3. Estimation of kanamycin resistence rates from MED4(**A**) and MIT9313 (**B**) growth rates under kanamycin selection. **A.** Kanamycin was added to MED4 cultures at either 25 or 50 μg ml^{-1}. Linear regressions were fit to the data points once the cells had resumed log phase growth. In the absence of antibiotics, a linear regression was fit to the data points at days 1 through 7 ($R = 0.51 * t + 15.96$, where R is the resistant cells ml^{-1} and t is days). In the 25 μg ml^{-1} treatments, a linear regression was fit using the data points at 14, 20, and 24 days ($R = 0.38* t + 13.81$) indicating that 14% of cells initally present were kanamycin resistant. In the 50 μg ml^{-1} treatment, a linear regression was fit using the data at 14, 20, and 24 days ($R = 0.15*t + 12.06$) indicating that 2% of cells were initially kanamycin resistant. **B.** MIT9313 kanamycin resistence rates from growth rates under kanamycin selection. Kanamycin was added to MIT9313 cultures at 15, 25 or 50 μg ml^{-1}. Linear regressions were fit to the data points once the cells had resumed log phase growth. In the absence of antibiotics, a linear regression was fit to the data points at days 1 through 12 ($R = 0.23*t + 15.08$). In the 15 μg ml^{-1} treatments, a linear regression was fit using the data points from day 1 to 20 ($R = 0.12 * t + 15.39$) indicating that nearly 100% of cells initally present were kanamycin resistant. In the 25 μg ml^{-1} treatment, a linear regression was fit using the data from days 32 to 63 ($R = 0.06* t + 12.31$) supporting that 6% of cells were kanamycin resistant.

Appendix V: Conjugal transfer of an RSF1010-derived plasmid to *Prochlorococcus*

INTRODUCTION

The initial goal of this study was to find methods by which foreign DNA could be introduced and expressed in the *Prochlorococcus* cell. To date, we have no evidence for natural competence or susceptibility to electroporation in *Prochlorococcus*. We thus focused on conjugation-based methods because of their high efficiency and insensitivity to species barriers. For example, conjugation has been used to efficiently transfer DNA from *E. coli to* other cyanobacterial taxa (Wolk et al, 1984) and these methods have been extended to even transfer DNA to mammalian cells (Waters, 2001). Our initial challenge was to find a means by which conjugation methods could be adapted to *Prochlorococcus*.

We focused on the conjugal transfer of plasmids that are expected to replicate autonomously in *Prochlorococcus.* No endogenous plasmids have been isolated from *Prochlorococcus*, but broad host-range plasmids such as RSF1010 derivatives have been shown to replicate in other cyanobacteria (Mermetbouvier et al, 1993). pRL153, an RSF1010 derivative, has been shown to replicate in three strains of a related oceanic cyanobacterium, *Synechococcus* (Brahamsha, 1996). We modified pRL153 to express a variant of Green Fluorescent Protein (GFP) called GFPmut3.1 which is optimized for bacterial GFP expression (Fig. 1). GFPmut3.1 expression was driven by the synthetic pTRC promoter which has been shown to be active in other cyanobacteria (Nakahira et al, 2004).

MATERIALS AND METHODS

Microbial growth conditions. The microbial stains used in this study are listed in table 1. *Prochlorococcus* was grown at 22°C in Pro99 medium (Moore et al, 1995) under continuous illumination from cool, white fluorescent lights at intensities of 50 µM Q m^{-2} s^{-1} and 10 µM Q m^{-2} s^{-1} for MED4 and MIT9313, respectively. *Prochlorococcus* was plated using the pour plating protocol from Brahamsha, 1996. These plates consisted of Pro99 medium supplemented with 0.5% ultra-pure low melting point agarose (Invitrogen Corp., product 15517-014). 1 ml of *Prochlorococcus* culture containing 10^5 cells ml^{-1} were added to the pour plates when the liquid agarose had cooled below 28°C.

E. coli stains were grown in Luria-Bertani (LB) medium supplemented with

ampicillin (150 μg ml^{-1}), kanamycin (50 μg ml^{-1}), or tetracycline (15 μg ml^{-1}) as appropriate. *E. coli* strains were grown at 37 °C. Cultures were continuously shaken except for cultures expressing the RP4 conjugal pilus which were not shaken to minimize the probability of shearing the conjugal pili.

Conjugation. pRL153 was conjugally transferred to *Prochlorococcus* from the *E. coli* host 1100-2 containing the conjugal plasmid pRK24. *E. coli* were mated with *Prochlorococcus* using the following method. A 100 ml culture of the *E. coli* donor strain containing the transfer plasmid was grown to mid-log phase OD 0.7-0.8. Parallel matings under the same conditions using *E. coli* lacking conjugal capabilites were done to confirm that non-donor *E. coli* were not sufficient for *Prochlorococcus* to become kanamycin-resistant. The *E. coli* cultures were centrifuged three times for 10 minutes at 3000 g. After the first two spins, the cell pellet was resuspended in 15 mls LB medium. After the third spin, the pellet was resuspended in 1 ml Pro99 medium for mating with *Prochlorococcus*.

A 100 ml culture of *Prochlorococcus* was grown to late-log phase (10^8 cell ml^{-1}). The culture was concentrated by centrifugation for 15 minutes at 9000 g and resuspended in 1 ml Pro99 medium. The concentrated *E. coli* and *Prochlorococcus* cells were then mixed at a 1:1 ratio and aliquoted as multiple 20 μl spots onto HATF filters (Millipore Corp., product HATF08250) on Pro99 plates containing 0.5% ultra-pure agarose. The plates were then transferred to 10μM Q m^{-2} s^{-1} continuous, white light at 22º C for 48 hours to facilitate mating. The cells were resuspended off the filters in Pro99 medium by pipetting and transferred to 25 ml cultures at an initial cell density of 5 x 10^6 cells ml^{-1}. Growth of the cultures was monitored by chlorophyll fluorescence using a Turner fluorometer (450 nm excitation; 680 nm excitation). 50 μg ml^{-1} kanamycin was added to the cultures after the *Prochlorococcus* cells had recovered from the mating procedure such that the chlorophyll fluorescence of the culture had increased two-fold.

Isolation of pure *Prochlorococcus* MIT9313 cultures after conjugation. Once the mated *Prochlorococcus* cultures had grown under kanamycin selection, cells were transferred to pour plates containing 25 μg ml^{-1} kanamycin to isolate colonies. Colonies generally formed in 6-10 weeks. *Prochlorococcus* colonies were excised using a sterile spatula and transferred back to liquid medium containing 50 μg ml^{-1} kanamycin. Once the MIT9313 cultures had reached late log-phase, a 100 μl aliquot of the culture was spread onto LB plates to titer the remaining *E. coli*. Unfortunately, 10^2 to 10^3 *E. coli* cells ml^{-1} often remained viable in the MIT9313 cultures even after

isolating MIT9313 colonies on Pro99-agarose plates. To eliminate the remaining *E. coli*, the MIT9313 cultures were infected with *E. coli* phage T7(Demerec and Fano, 1945: Studier, 1969) at a multiplicity of infection (MOI) of 10^6 phage per *E. coli* host. The *E. coli* were again titered on LB plates the following day to show that no viable cells remained.

Plasmid isolation from *Prochlorococcus* MIT9313. Plasmid DNA from MIT9313 cultures expressing pRL153 was isolated from 5 mls of stationary phase cultures using a Qiagen mini-prep spin column kit. As found by Brahamsha, 1996 with *Synechococcus*, the yield of pRL153 from *Prochlorococcus* was too low to visualize by gel electrophoresis; we thus transformed *E. coli* with the plasmids isolated from *Prochlorococcus* in order to compare the structure of pRL153 from MIT9313 to the original plasmid. Following transformation into *E. coli*, pRL153 was isolated from kanamycin resistant *E. coli* transformants and digested with EcoRV and HindIII to compare its structure with the original plasmid.

pRL153-GFP Plasmid construction. pRL153 was modified to express GFPmut3.1 from the synthetic pTRC promoter to determine if GFP expression could be detected in *Prochlorococcus*. pRL153 contains unique sites for HindIII and NheI in the Tn5 fragment that are outside the kanamycin resistance gene. pTRC-GFPmut3.1 was cloned into into the unique NheI site to create pRL153-GFP. To this end, pTRC-GFPmut3.1 was PCR amplified from pJRC03 using PFU polymerase using primers with 5' NheI sites: forward primer (pTRC): 5'-acgtac-gctagc-ctgaaatgagctgttgacaatt-3' and reverse primer (GFPmut3.1) 5'-cgtacc-gctagc-ttatttgtatagttcatccatgc-3'. pTRC-GFP PCR product was then NheI digest, CIP-treated, and ligated with NheI-digested pRL153. The ligation was transformed into DH5-alpha and the pTRC-GFP insertion was confirmed by restriction analysis. GFP expression from pRL153-GFP in *E. coli* was visualized by epifluorescence microscopy. A diagram of pRL153-GFP is shown in Figure 1.

GFP detection. GFPmut3.1 has maximal excitation and emission wavelengths of 501 nm and 511 nm, respectively (http://www.bdbiosciences.com/clontech/techinfo/vectors_dis/pGFPmut3.1.shtml). The fluorescence emission spectra of MIT9313 cells expressing pRL153-GFP and control cells of equal density expressing pRL153 were quantified using a Perkin Elmer Luminescence Spectrometer LS50B. The cells were excited at 490 nm and their

cellular fluorescence was measured at 5 nm intervals from 510-700 nm. Cells from duplicate, independently mated +GFP and -GFP MIT9313 cultures were measured. We quantified fluorescence differences between +GFP cells as -GFP cells as mean of the +GFP measurements minus the mean of -GFP measurements.

RESULTS

Conjugal transfer of pRL153 to *Prochlorococcus* MED4. Once the cells had acclimated to the growth conditions, we monitored the growth rate of the cells by chlorophyll fluorescence (Fig. 2). The MED4 growth rate under these conditions was 0.84 doublings day^{-1} ($\mu = 0.58$ day^{-1}) (Fig. 2A). The MIT9313 growth rate was 0.35 doublings day^{-1} ($\mu = 0.24$ day^{-1}) (Fig. 2B). Cultures for the matings were grown under these same conditions; matings were conducted when the cells reached late log phase. In all matings, we observed that MED4 grew under kanamycin selection when mated with *E. coli* containing the conjugal plasmid pRK24 and the transfer plasmid pRL153 (Fig. 3-5). In the first two matings, we observed that the control MED4 cultures mated with *E. coli* lacking the conjugal plasmid did not grow under kanamycin selection (Fig. 3-4). This suggests that pRL153 does replicate in MED4. However, previous data supported that MED4 can become resistant to kanamycin, even at 50 µg ml^{-1} as used in this study (see previous report). Thus, in the third experiment, we included an additional treatment in which the MED4 cultures were inoculated with at an initial concentration of 10^7 cells ml^{-1} instead of 10^6 cells ml^{-1} (Fig. 5). We found that, if the initial inoculum was sufficiently large, MED4 was able to overcome the kanamycin selection. This observation was consistent with previous data that MED4 can become spontaneously resistant to kanamycin. It is not known whether the the larger inoculum enabled MED4 to grow under kanamycin selection because a larger inoculum simply has a greater probability of containing a spontaneous mutant or because MED4 can detoxify the kanamycin when the cells are sufficiently dense.

Conjugal transfer of pRL153 to *Prochlorococcus* MIT9313. In the first two MIT9313 mating experiments, MIT9313 cultures mated with *E. coli* containing RK24 and pRL153 grew under kanamycin selection; control MIT9313 cultures mated with *E. coli* lacking the conjugal plasmid did not grow (Fig. 6 and 7). This growth data supported that conjugation with *E. coli* was required for *Prochlorococcus* to become kanamycin resistant. We did not find that mated MIT9313 grew under kanamycin selection in the subsequent matings (Fig. 8 and 9) even though the MIT9313 growth rates were the same in all four experiments. The only difference that we observed

between the first and second two matings was that the cells in the no-kanamycin treatments in the second two matings had a several day lag time before they began to grow in liquid immediately after matings. This difference is likely because we moved labs and the cultures had difficulty acclimating to different incubators. This difference can be observed by comparing growth of the -kan treatments in Fig. 6 and 7 versus Fig. 8 and 9. This lag in growth suggested that the MIT9313 cells had not recovered as well following the matings. To compensate for this potential stress increase, the mating procedure was modified so as to not add kanamycin to the cultures until the cells had resumed growth such that the chlorophyll fluoresence had doubled once, no matter how long that takes. In all previous matings, kanamycin was added to the +kan cultures 1 day after cells were transferred to liquid medium.

When the mating procedure was modified such that kanamycin was not added to the cultures until they had resumed growth, MIT9313 cultures grew under kanamycin selection if they had been mated with *E. coli* expressing pRK24 and pRL153 (Fig. 10 and 11). In contrast, MIT9313 cultures mated with *E. coli* lacking pRK24 did not grow under kanamycin selection even if they had resumed growth prior to kanamycin addition. These experiments support that pRL153 can be transferred to *Prochlorococcus* MIT9313 by conjugation and, if the cells had recovered from mating, they will express kanamycin resistence.

Isolation of MIT9313 expressing pRL153. We plated MIT9313 cells that had been mated with *E. coli* expressing pRK24 and pRL153 to isolate MIT9313 colonies. Plating efficiencies are generally between 0.01 to 1% and colonies were first observed 6 weeks after plating. Plating of *Prochlorococcus* is notoriously difficult. Plating efficiencies for *Prochlorococcus* are low and variable; not all strains have been successfully plated at all. While we were able to isolate MIT9313 colonies from cultures actively growing in liquid, no colonies were observed when cells were plated directly after mating. This suggests that initially growing MIT9313 in liquid may allow the cells to physiologically recover from the mating procedure such that they survive to form colonies in pour plates.

We were unable to use standard plating methods to calculate mating efficiencies because we could only isolate *Prochlorococcus* colonies after the cells had first been grown in liquid medium after mating. We estimated the conjugation efficiency using the following method. Chlorophyll fluorescence values from the cells shown in Fig. 2B were correlated to cell abundances using flow cytometery. A linear regression correlating time to the number of transconjugant cells in culture was fit to the data points between days 35 and 60 of Fig. 12 ($R = 0.044*t + 4.82$ where R is the

$\log_{10}$(tranconjugant cells) and t is days since mating). We calculated the number of transconjugant cells immediumtely after mating as the intersection of the regression line with the ordinate axis. Using this value, on can calculate the conjugation efficiency to be about 1% by dividing the initial number of transconjugants ($6.9x10^4$ cells) by the number of cells initial transferred into the culture ($6.5x10^6$ cells) .

We found that 10^2 to 10^3 *E. coli* cells ml^{-1} often persisted in the MIT9313 cultures even after colonies had been picked from Pro99-agarose plates. This is likely because *E. coli* cells were transferred back into the liquid medium along with the MIT9313 cells when the *Prochlorococcus* colonies were excised from the top agar. Residual *E. coli* were removed by infecting the cultures with *E. coli* phage T7 at a multiplicity of infection of 10^6 phage per host. T7 infection at any MOI resulted in no adverse effects on *Prochlorococcus* viability.

Plasmid DNA was then isolated from axenic MIT9313 cultures to compare the structure of pRL153 from MIT9313 to the original plasmid. To this end, *E. coli* was transformed with plasmid DNA isolated from *Prochlorococcus*. We typically obtained approximately 100 *E. coli* transformants when DH5-alpha cells competent to 10^5 transformants μg^{-1} DNA were transformed with one-fifth of a plasmid DNA prep from an MIT9313 culture of $5x10^8$ cells. These efficiencies support that the total plasmid yield was 5 ng of pRL153. Based on the molecular weight of DNA (1bp = 660 daltons), one can calculate that a 5 ng of plasmid DNA from $5x10^8$ cells constitutes a plasmid isolation efficiency of 1.06 plasmids per MIT9313 cell. Restriction fingerprinting of the rescued plasmid DNA supports that the gross structure of pRL153 is generally conserved in *Prochlorococcus* (Fig. 12). In total, we examined the fingerprints of 20 plasmids isolated from 4 independently mated cultures; 19 of the plasmids were identical to the original pRL153.

GFP expression in *Prochlorococcus*. pRL153 was modified to express GFPmut3.1 from the pTRC promoter. We isolated MIT9313 cultures expressing pRL153-GFP and quantified GFP expression in these cultures (+GFP cells) by comparing their fluorescence properties to MIT9313 cells expressing pRL153 lacking GFP (-GFP cells). GFPmut3.1 has an excitation maximum of 501 nm and a fluorescence maximum of 511 nm. Thus, to examine GFP fluorescence in Prochlorococcus, +GFP and -GFP MIT9313 cells were excited at 490 nm and their emission spectrum was measured from 510 to 700 nm using a spectrofluorometer (Fig. 13A). The increased cellular fluoresence of -GFP cells at lower wavelengths is presumably due to scattering of the 490 nm excitation wavelength. By comparing the means of +GFP cells to -GFP cells, we observed that +GFP cells had increased cellular fluorescence specifically in the

region of GFP fluorescence (Fig. 13B). We quantified GFP expression in Prochlorococcus by subtracting the mean -GFP signal from the mean +GFP signal (Fig. 13B). We observed that the mean fluorescence of +GFP cells was greater than in -GFP cells in the vicinity of GFP fluorescence.

DISCUSSION

The primary objective of these experiment was to investigate conditions by which a plasmid could be transferred to *Prochlorococcus* by conjugation with *E. coli*. Our data supports that an interspecific conjugation system based on the RP4 plasmid family can be used to transfer DNA into *Prochlorococcus* MED4 and MIT9313. A key factor in the mating procedure is to wait until the cells have recovered from the mating procedure before adding kanamycin to the medium. This wait period is presumably to allow the cells to begin expressing the kanamycin resistance gene. Although pRL153 appears to replicate in both strains, MIT9313 is preferable because MED4 has the potential to become spontaneously kanamycin resistant.

pRL153, an RSF1010-derived plasmid, replicates autonomously in MIT9313 conferring resistance to kanamycin and can be used to express foreign proteins such as those for kanamycin-resistance and GFP. Once a liquid culture of kanamycin-resistant cells has been isolated, pour plating methods can be used to isolate individual colonies. These colonies can be transferred back to liquid medium for further characterization. The transfer of replicating plasmids, especially those expressing GFP, will have myriad applications. For example, one could create transcriptional fusions between *Prochlorococcus* promoters and GFP to study the diel cycling of gene expression in *Prochlorococcus*. Rhythmicity of gene expression is particularly interesting because of results in other cyanobacteria supporting that the expression of all genes cycle daily and are controlled by a central oscillator (Golden, 2003). Second, GFP expression could provide a means to flow cytometrically sort transgenic from non-transgenic cells. Faced with variable and overall low plating efficiencies, flow sorting cells is an attractive alternative in order to isolate mutants following conjugation. Alternatively, RSF1010-derived plasmids could be modified to cause *Prochlorococcus* to express other foreign proteins. For example, a His-tagged MIT9313 protein could be cloned into pRL153 and transferred into *Prochlorococcus* by conjugation. The ectopically expressed, tagged protein could then be purified to determine which proteins interact with it *in vivo*.

REFERENCES

Bandrin, SV, Rabinovich, PM, Stepanov, AI. (1983). "Three linkage groups of genes involved in riboflavin biosynthesis in *E. coli*." Genetika (Sov. Genet.) 19: 1419-1425.

Brahamsha, B. (1996). "A genetic manipulation system for oceanic cyanobacteria of the genus *Synechococcus*". Appl Environ Microbiol. May;62(5):1747-51.

Campbell, L., Landry, MR., Constantinou, J., Nolla, HA., Brown, SL., Lui, H. (1998). Response of microbial community structure to environmental forcing in the Arabian Sea. Deep Sea Research II 45, 2301-2325.

Cohen, MF., Meeks, JC., Cai, YA., Wolk, CP. (1998). "Transposon mutagenesis of heterocyst-forming filamentous cyanobacteria". Methods in Enzymology 297: 3-17.

Demerec, M. and Fano, U. (1945). "Bacteriophage-resistant mutants in Escherichia coli". Genetics 30: 119-136.

Dufrense A et al, (2003). "Genome sequence of the cyanobacterium *Prochlorococcus* marinus SS120, a nearly minimal oxyphototrophic genome. PNAS, vol. 100. no. 17, 10020-10025.

Elhai, J. and Wolk, CP. (1988). "Conjugal transfer of DNA to cyanobacteria". Methods in Enzymology. 167: 747-754.

Golden, SS. (2003). "Timekeeping in bacteria: the cyanobacterial circadian clock". Current Opinion in Microbiology (6): 535-540.

Koksharova, OA and Wolk, CP. (2002). "Genetic tools for cyanobacteria." Appl Micro and Biotech. 58 (2): 123-137 FEB 2002

Larsen RA, Wilson MM, Guss AM, Metcalf WW. (2002). "Genetic analysis of pigment biosynthesis in Xanthobacter autotrophicus Py2 using a new, highly efficient transposon mutagenesis system that is functional in a wide variety of bacteria." Archives of Microbiology 178 (3): 193-201.

Li, WKW. (1994). Primary productivity of prochlorophytes, cyanobacteria, and eucaryotic ultraphytoplankton: measurements from flow cytometric sorting. Limnology and Oceanography 39, 169-175.

Mermetbouvier, P., Cassierchauvat, C., Marraccini, P., Chauvat, F. (1993). "Transfer and replication of RSF1010 derived plasmids in severl of the general Synechocystis and *Synechococcus*. Current Microbiology. (6): 323-327.

Metcalf, WW., W. Jiang, LL. Daniels, SK. Kim, A. Haldimann, and Wanner. BL. (1996). "Conditionally replicative and conjugative plasmids carrying *lacZ*α for cloning, mutagenesis, and allele replacement in bacteria". Plasmid 35:1-13. Highwire ID="182:6:1671:27".

Moore, LR., Goericke, R., Chisholm, SW. (1995). "Comparative physiology of *Synechococcus* and *Prochlorococcus* – Influence of light and temperature on growth, pigments, fluorescence, and absorptive properties". Marine Ecology Progress Series 16 (1-3): 259-275.

Nakahira, Y., Katayama, M., Miyashita, H., Kutsuna, S., Iwasaki, H., Oyama, T., Kondo, T. (2004). "Global gene repression by KaiC as a master processor of prokaryotic circadian system". PNAS 101 (3): 881-885.

Palenik, B. (2001). "Chromatic adaptation in marine *Synechococcus* strains". Applied and Environmental Microbiology. 67: 991-994.

Partensky, F., Hess, WR., Vaulot, D. (1999). *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. Microbiology and Molecular Biology Reviews 63 (1), 106-127.

Rocap, G. et al. (2003). "Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation". Nature 424 (6952): 1042-1047

Scholz,P., Haring,V., Wittmann-Liebold,B., Ashman,K., Bagdasarian,M. and Scherzinger,E. (1989). "Complete nucleotide sequence and gene organization of the broad-host-range plasmid RSF1010". Gene 75 (2), 271-288.

Studier, FW. (1969). "The genetics and physiology of bacteriophage T7". Virology 39: 562-574.

Toledo, G., Palenik, B., and Brahamsha, B. (1999). "Swimming marine *Synechococcus* strains with widely different photosynthetic pigment ratios form a monophyletic group". Applied and Environmental Microbiology. 65: 5247-5251.

Waters, VL. (2001). "Conjugation between bacterial and mammalian cells". Nature, 29, 4. pp 375-376.

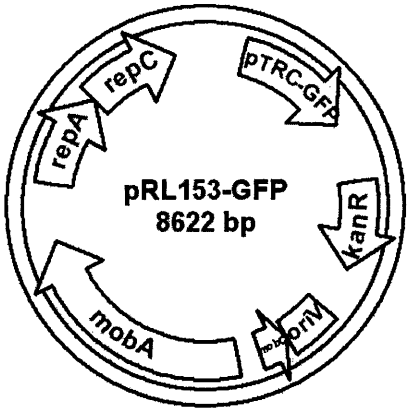


Fig.1. Diagram of the RSF1010-derived plasmid pRL153 modified to contain pTRC-GFPmut3.1. pRL153 consists of bp 2118-7770 of RSF1010 ligated to bp 680-2516 of Tn5 thereby replacing the sulfonamide resistance gene of RSF1010 with the kanamycin resistance gene of Tn5. pRL153 was modified to express GFP by cloning the pTRC-GFPmut3.1 fusion into the unique NheI site upstream of the kanamycin-resistance gene.

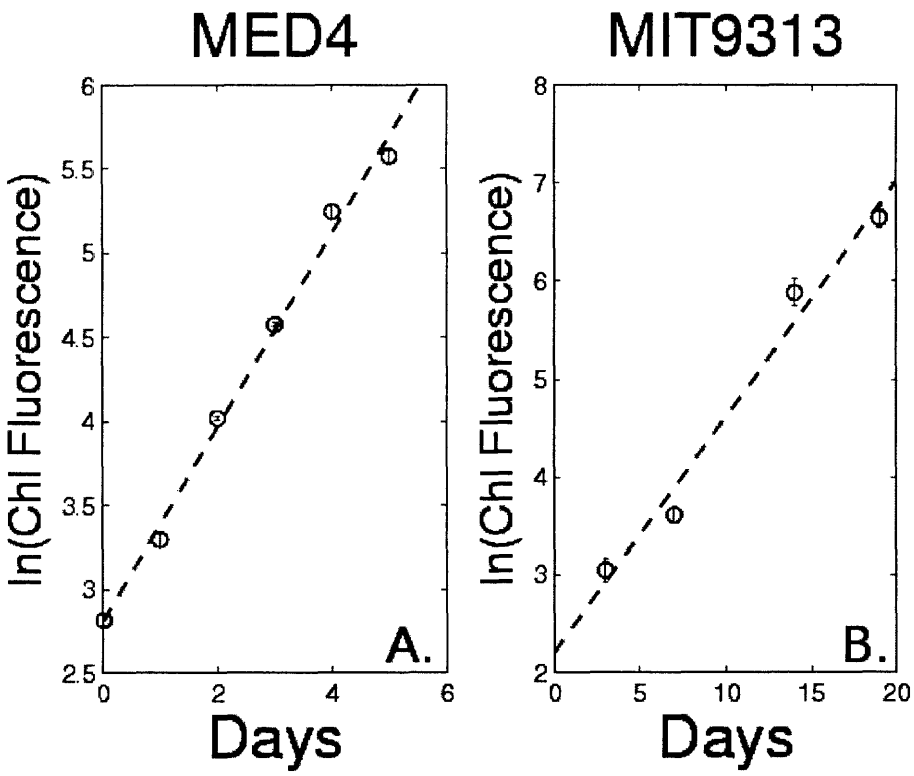


Fig. 2. Growth of MED4 (**A**) and MIT9313 (**B**) cells under conditions used in matings. **A.** MED4 grew at a rate of 0.84 doublings day^{-1} (μ = 0.58 day^{-1}). **B**. Growth rate of MIT9313 cells under conditions used in matings. MIT9313 grew at a rate of 0.35 doublings day^{-1} (μ = 0.24 day^{-1}).

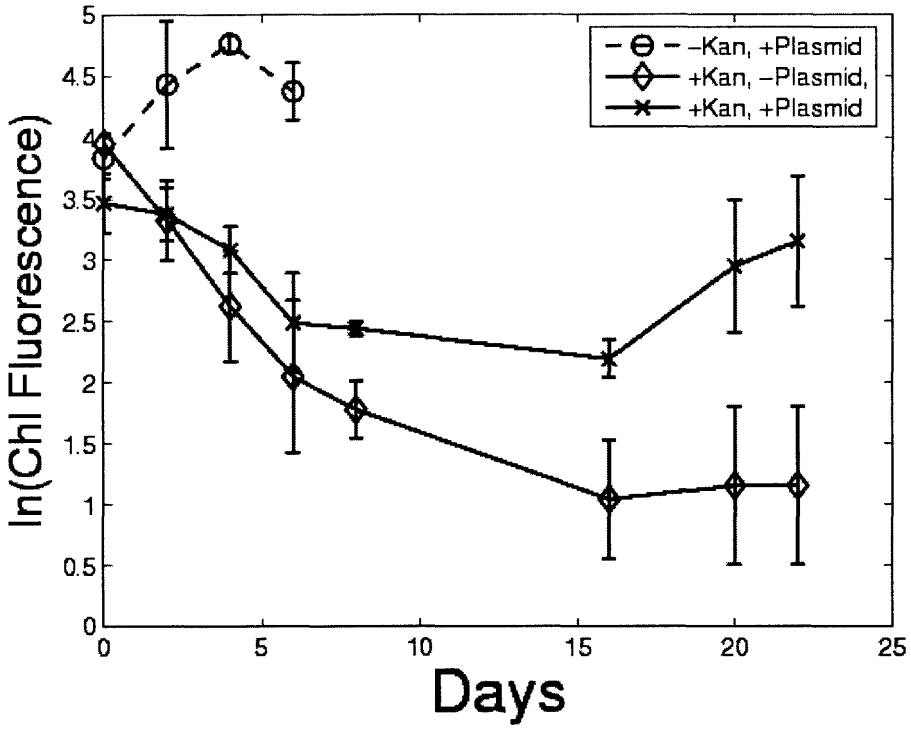


Fig. 3. MED4 cultures grow in medium containing 50 µg ml^{-1} kanamycin when mated with *E. coli* containing the conjugal plasmid pRK24 and pRL153 (+kan, +plasmid). Control MED4 cultures mated with *E. coli* lacking pRK24 (+kan, -plasmid) do not grow under kanamycin selection. Control cultures mated with *E. coli* containing pRK24 and pRL153 grow in medium lacking kanamycin (-kan, +plasmid). Curves are average of duplicate cultures; error bars show one standard deviation from the mean.

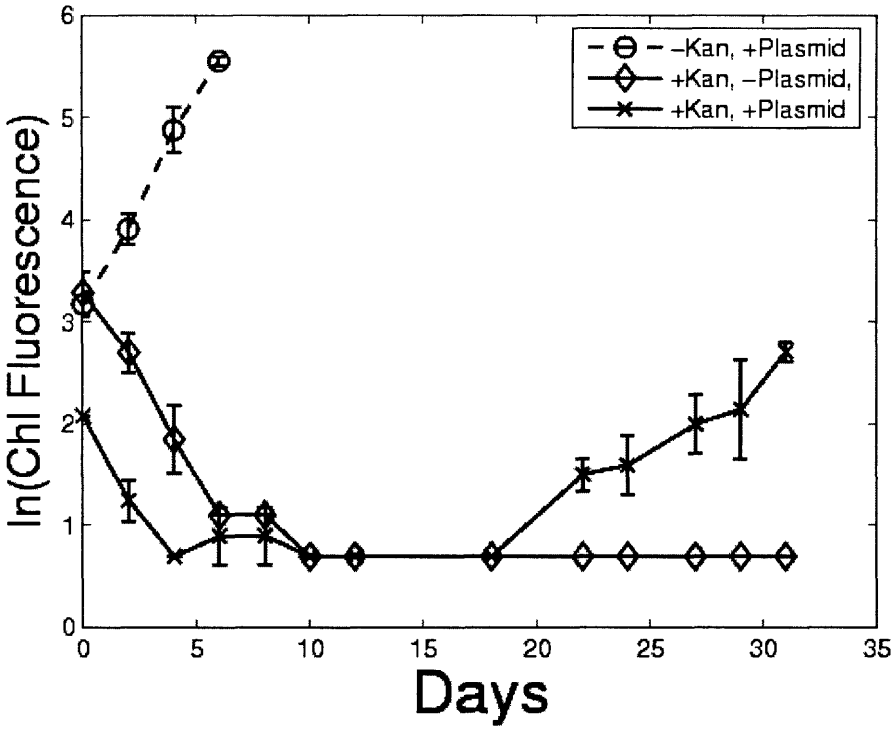


Fig. 4. MED4 cultures grow in medium containing 50 µg ml^{-1} kanamycin when mated with *E. coli* containing the conjugal plasmid pRK24 and pRL153 (+kan, +plasmid). Control MED4 cultures mated with *E. coli* lacking pRK24 (+kan, -plasmid) do not grow under kanamycin selection. Control cultures mated with *E. coli* containing pRK24 and pRL153 grow in medium lacking kanamycin (-kan, +plasmid). Curves are average of duplicate cultures; error bars show one standard deviation from the mean.

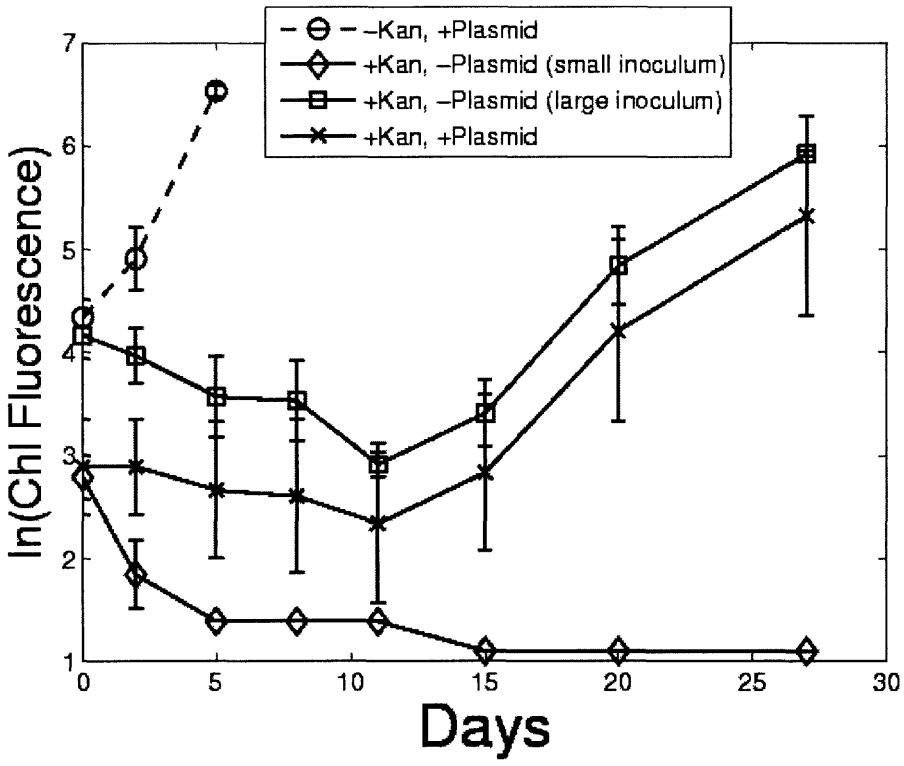


Fig. 5. MED4 lacking pRL153 grows under kanamycin selectin when the initial inoculum of cells into medium following mating if sufficiently large. MED4 cultures grow in medium containing 50 µg ml^{-1} kanamycin when mated with *E. coli* containing the conjugal plasmid pRK24 and pRL153 (+kan, +plasmid). However, control MED4 cultures mated with *E. coli* lacking pRK24 (+kan, -plasmid) also grow under kanamycin selection if the initial inoculum of $2x10^8$ cells (final concentration10^7 cells ml^{-1}). MED4 cultures mated with pRK24 lacking pRL153 (+kan, -plasmid) do not grow under kanamycin selection with a smaller inoculum (10^6 cell ml^{-1}) Control cultures mated with *E. coli* containing pRK24 and pRL153 grow in medium lacking kanamycin (-kan, +plasmid). Curves are average of duplicate cultures; error bars show one standard deviation from the mean.

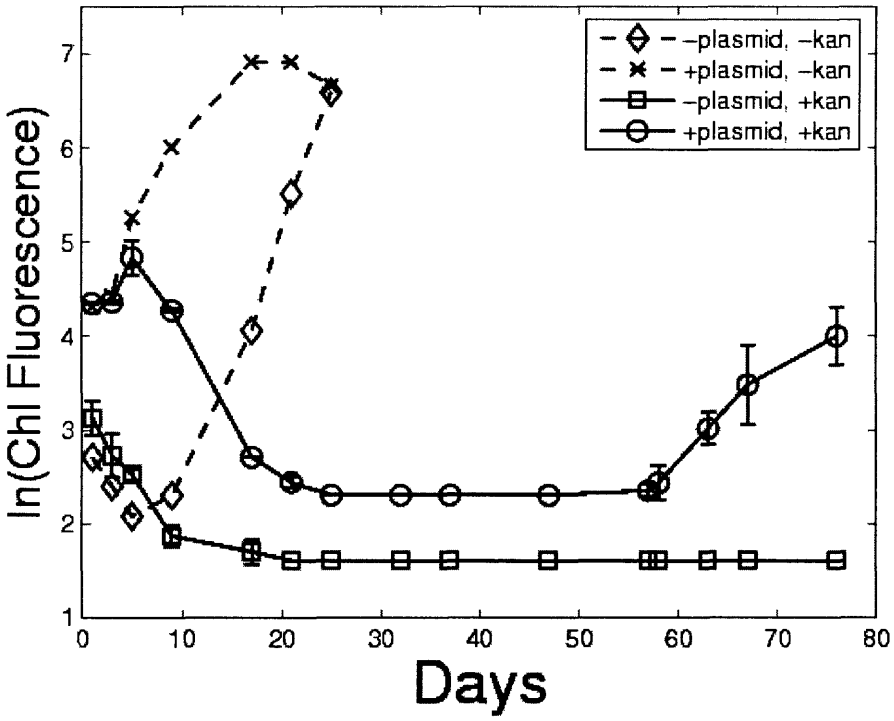


Fig. 6. MIT9313 cultures grow in medium containing 50 µg ml^{-1} kanamycin when mated with *E. coli* containing the conjugal plasmid pRK24 and pRL153 (+plasmid, +kan). Control MIT9313 cultures mated with *E. coli* lacking pRK24 (-plasmid, +kan) do not grow under kanamycin selection. Control cultures with and without plasmid grow in medium lacking kanamycin (+/-plasmid, -kan). +kan plots show mean of duplicate cultures; error bars show one standard deviation. -kan plots show individual cultures.

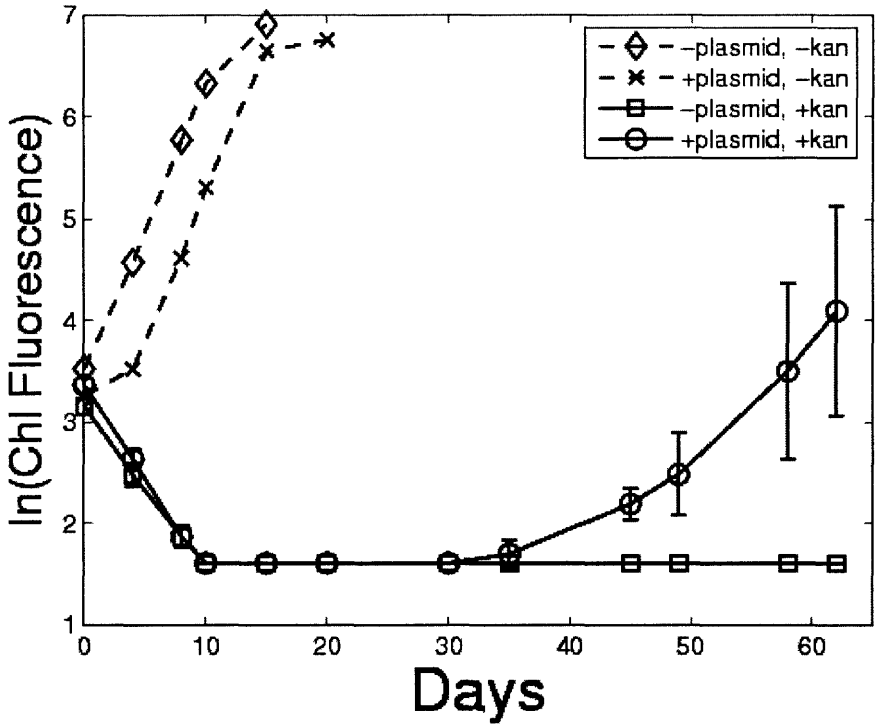


Fig. 7. MIT9313 cultures grow in medium containing 50 µg ml^{-1} kanamycin when mated with *E. coli* containing the conjugal plasmid pRK24 and pRL153 (+plasmid, +kan). Control MIT9313 cultures mated with *E. coli* lacking pRK24 (-plasmid, +kan) do not grow under kanamycin selection. Control cultures with and without plasmid grow in medium lacking kanamycin (+/-plasmid,-kan). +kan plots show mean of duplicate cultures; error bars show one standard deviation. -kan plots show individual cultures.

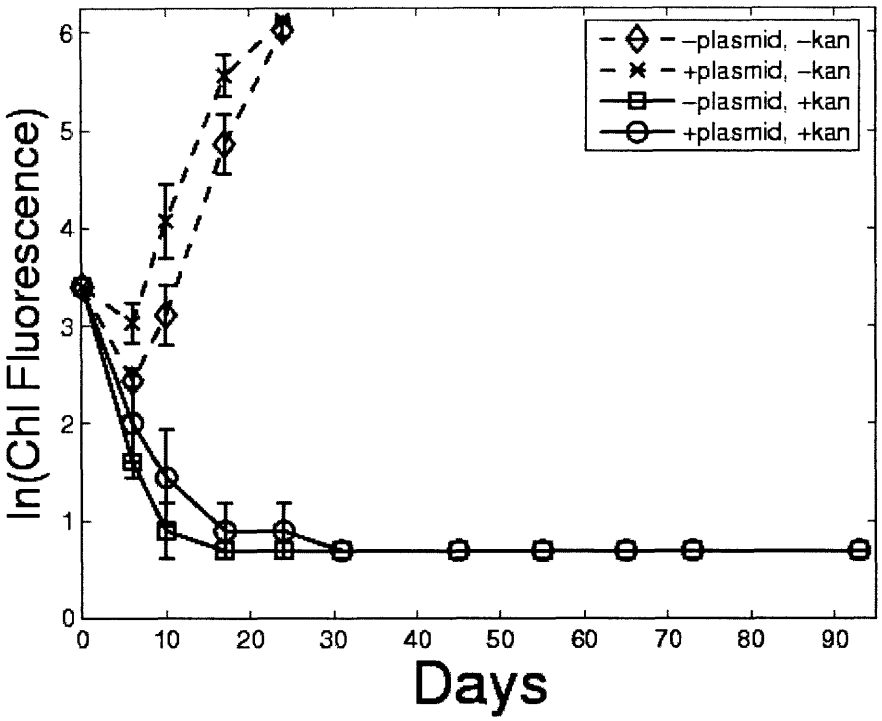


Fig. 8. MIT9313 cultures do not grow in medium containing 50 µg ml^{-1} kanamycin when mated with *E. coli* containing the conjugal plasmid pRK24 and pRL153 (+plasmid, +kan) if the cultures are not given sufficient time to recover prior to kanamycin additions. Kanamycin was added to all +kan cultures 1 day after transfer to liquid medium. Control MIT9313 cultures mated with *E. coli* lacking pRK24 (-plasmid, +kan) do not grow under kanamycin selection either. Control cultures mated with *E. coli* containing pRK24 and pRL153 grow in medium lacking kanamycin (+/-plasmid, -kan). Each curve represents the mean of duplicate cultures; error bars show one standard deviation.

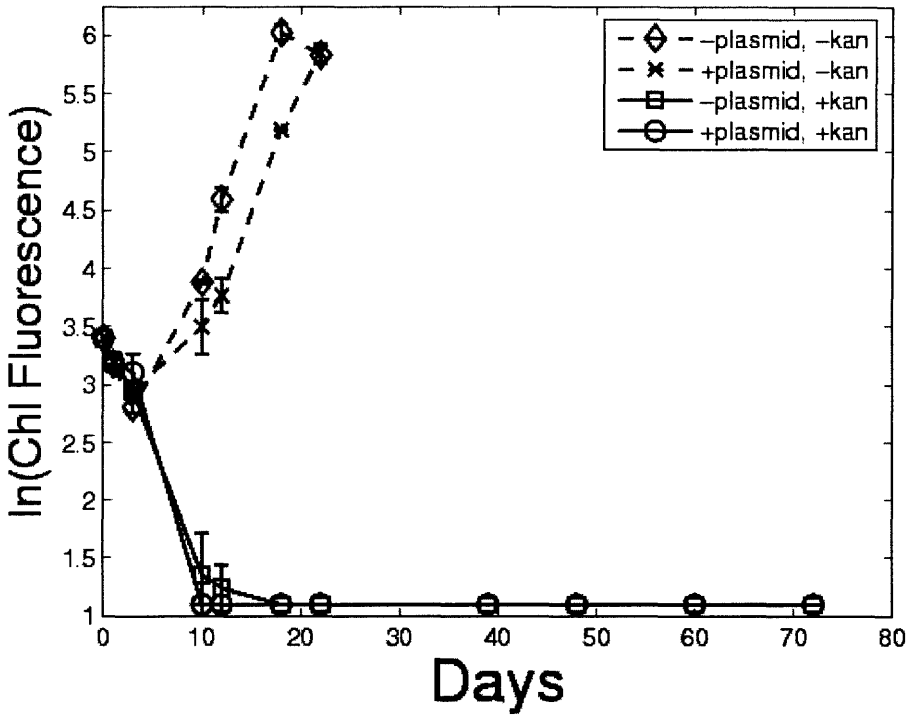


Fig. 9. MIT9313 cultures do not grow in medium containing 50 μg ml^{-1} kanamycin when mated with *E. coli* containing the conjugal plasmid pRK24 and pRL153 (+plasmid, K50) when not given sufficient time to recover prior to addition of kanamycin. Kanamycin was added to all +kan cultures 1 day after transfer to liquid medium. Control MIT9313 cultures mated with *E. coli* lacking pRK24 (-plasmid, +kan) do not grow under kanamycin selection either. Control cultures mated with *E. coli* with and without the conjugal plasmid grow in medium lacking kanamycin (+/- plasmid, -kan). Each curve represents the mean of duplicate cultures; error bars show one standard deviation.

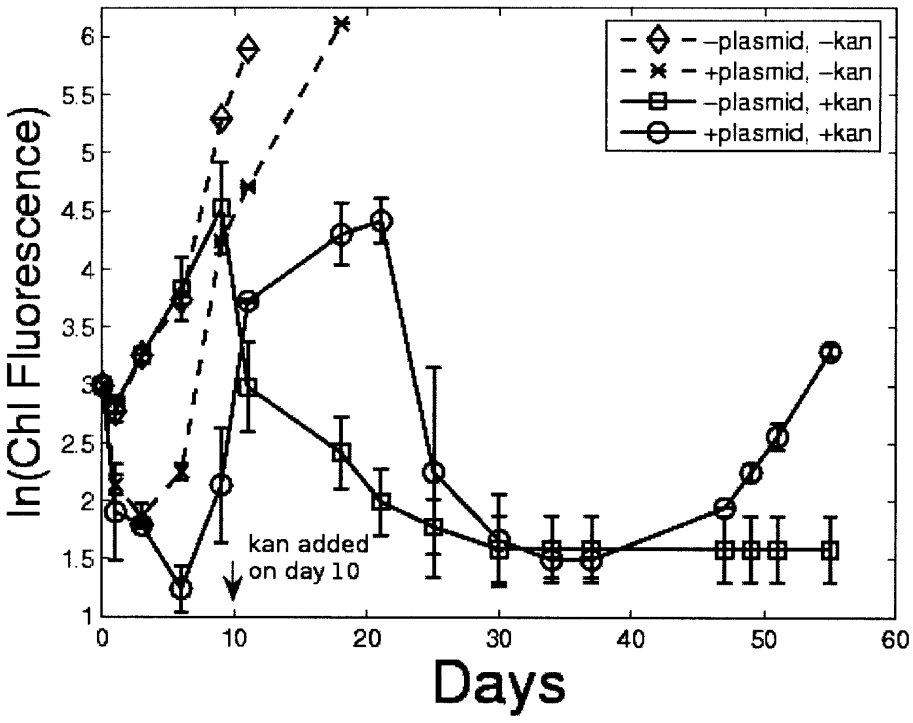


Fig. 10. When MIT9313 cultures are allowed to resume growth prior to addition of kanamycin, they grow in medium containing 50 μg ml^{-1} kanamycin when mated with *E. coli* containing the conjugal plasmid pRK24 and pRL153 (+plasmid, +kan). Control MIT9313 cultures mated with *E. coli* lacking pRK24 (-plasmid, +kan) do not grow under kanamycin selection. Control cultures mated with *E. coli* with and without the conjugal plasmid grow in medium lacking kanamycin (+/-plasmid, -kan). Each curve represents the mean of duplicate cultures, error bars show one standard deviation. The arrow shows that kanamycin was added to the +kan cultures 10 days after transfer to liquid medium.

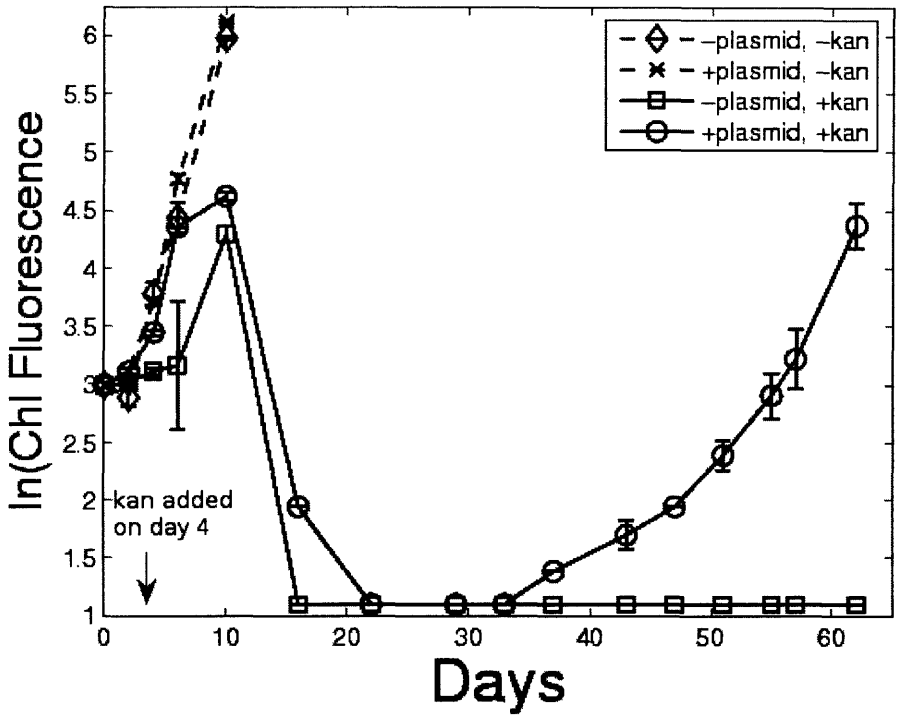


Fig. 11. When MIT9313 cultures are allowed to resume growth prior to addition of kanamycin, they grow in medium containing 50 μg ml^{-1} kanamycin when mated with *E. coli* containing the conjugal plasmid pRK24 and pRL153 (+plasmid, +kan). Control MIT9313 cultures mated with *E. coli* lacking pRK24 (-plasmid, +kan do not grow under kanamycin selection. Control cultures mated with *E. coli* with and without the conjugal plasmid grow in medium lacking kanamycin (+/-plasmid, -kan). Each curve represents the mean of duplicate cultures, error bars show one standard deviation. The arrow shows that kanamycin was added to +kan cultures 4 days after transfer to liquid medium.

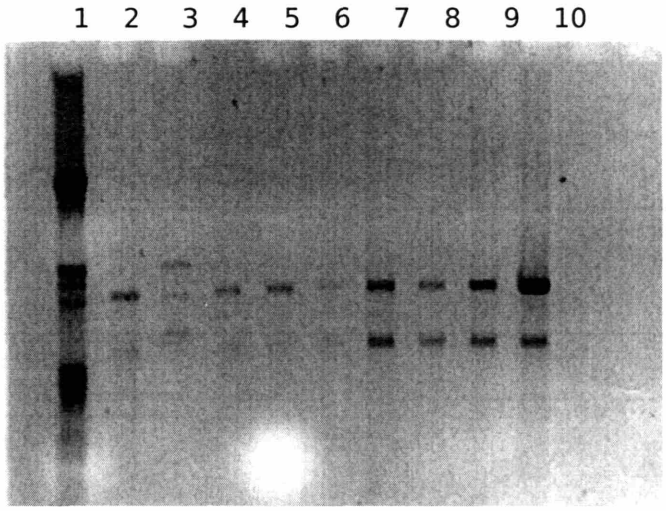


Fig. 12. EcoRV/HindIII digestion of pRL153 plasmids isolated from MIT9313 cultures. Lane 1: EcoRI/HindII digested phage lambda DNA. 2: pRL153 directly from *E. coli*. 3-10: pRL153 rescued from MIT9313 cultures. The digestion pattern in lane 3 shows that the structure of pRL153 is not always retained in MIT9313. However, lanes 4-10 support that the pRL153 structure is generally conserved.

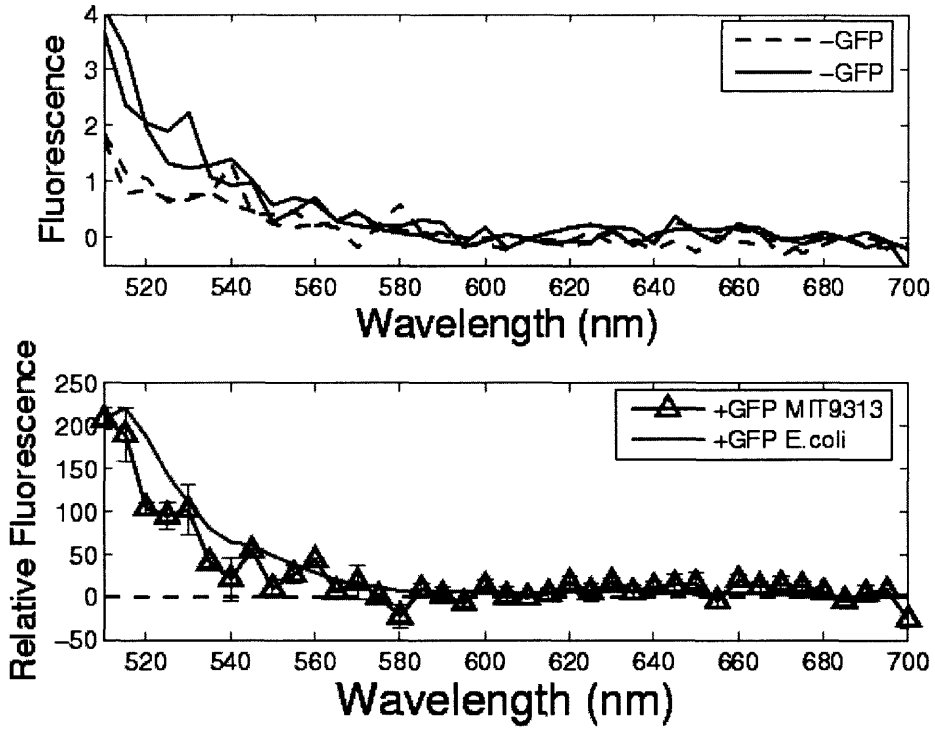


Fig. 13. MIT9313 cells expressing GFP have increased cellular fluorescence in the range of GFP fluorescence relative to -GFP cells. MIT9313 cells expressing pRL153-GFP and control cells lacking GFP were excited at 490 nm and their fluorescence spectrum from 510-700 nm was measured. **A**. Raw fluorescence measurements for ±GFP cultures. **B**. The fluorescence of +GFP cells relative to -GFP cells; the mean of duplicate -GFP measurements were subtracted from the mean duplicate +GFP fluorescences. The horizontal dashed line shows the zero line where the relative fluorescence of +GFP cells is equal to -GFP cells. Error bars show standard error of the mean.

Appendix VI: Supplemental figures for *Prochlorococcus* microarray analysis of gene expression.

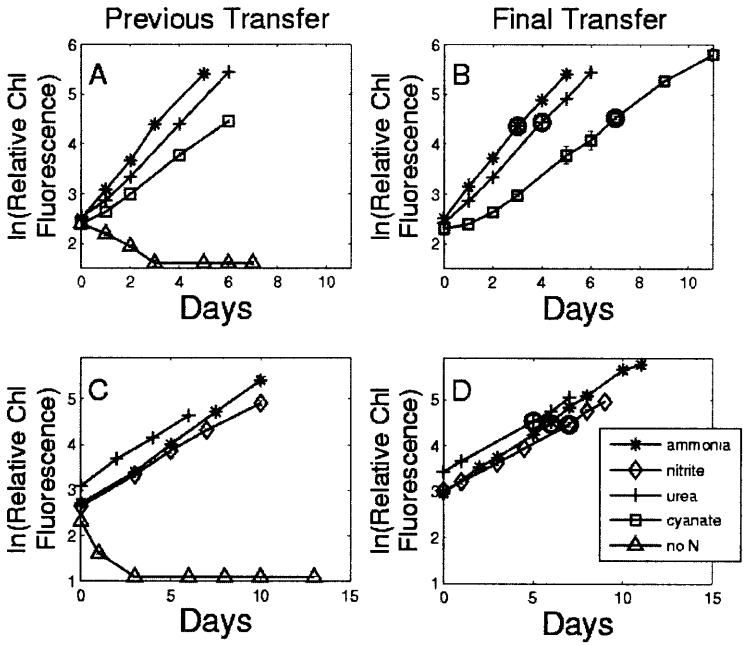


Fig. 1. Growth of *Prochlorococcus* MED4 (**A,B**) and MIT9313 (**C,D**) in media containing different nitrogen sources: 800 µmol ml^{-1} ammonia, 200 µmol ml^{-1}nitrite, 800 µmol ml^{-1} cyanate, 400 µmol ml^{-1} urea, or no added nitrogen. MED4 growth rates in the final two transfers (**A** and **B,** respectively) were calculated by linear regression: ammonia 0.58 day^{-1}, cyanate 0.35 day^{-1}, and urea 0.51 day^{-1}. MIT9313 growth rates in the final two transfers (**C** and **D**, respectively) were also calculated: ammonia 0.22 day^{-1}, nitrite 0.21 day^{-1}, urea day^{-1}. Neither strain grew when transferred into media lacking supplemental nitrogen. Circled data points in the second final transfer show when samples were taken for microarray analysis.

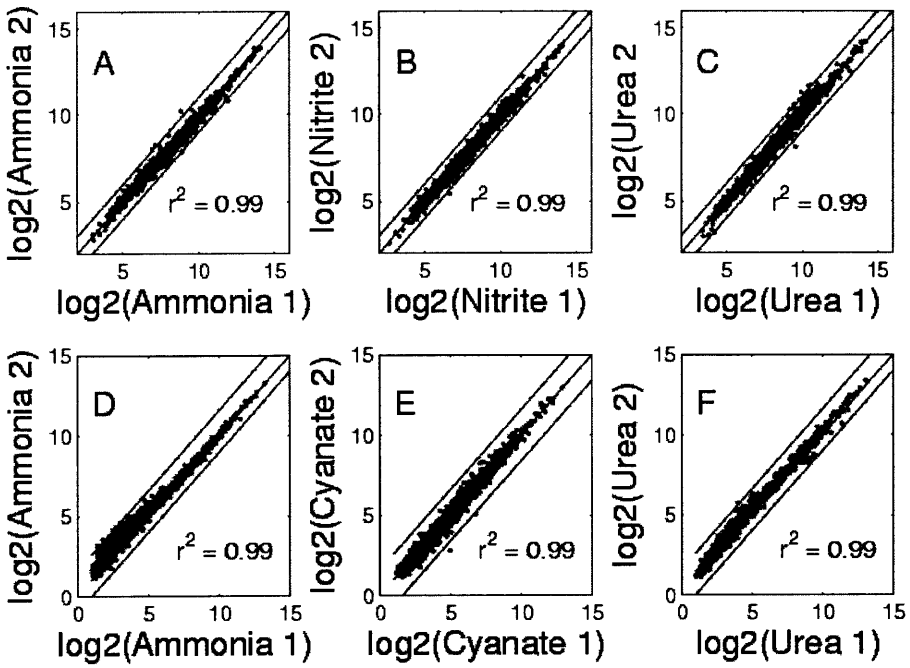


Fig. 2. Comparison of expression profiles from replicates cultures of *Prochlorococcus* MIT9313 (**A-C**) and MED4 (**D-F**) grown on different nitrogen sources. Correlation coefficients for expression profiles of replicate cultures are shown in each panel. Solid lines show 2-fold change in expression.

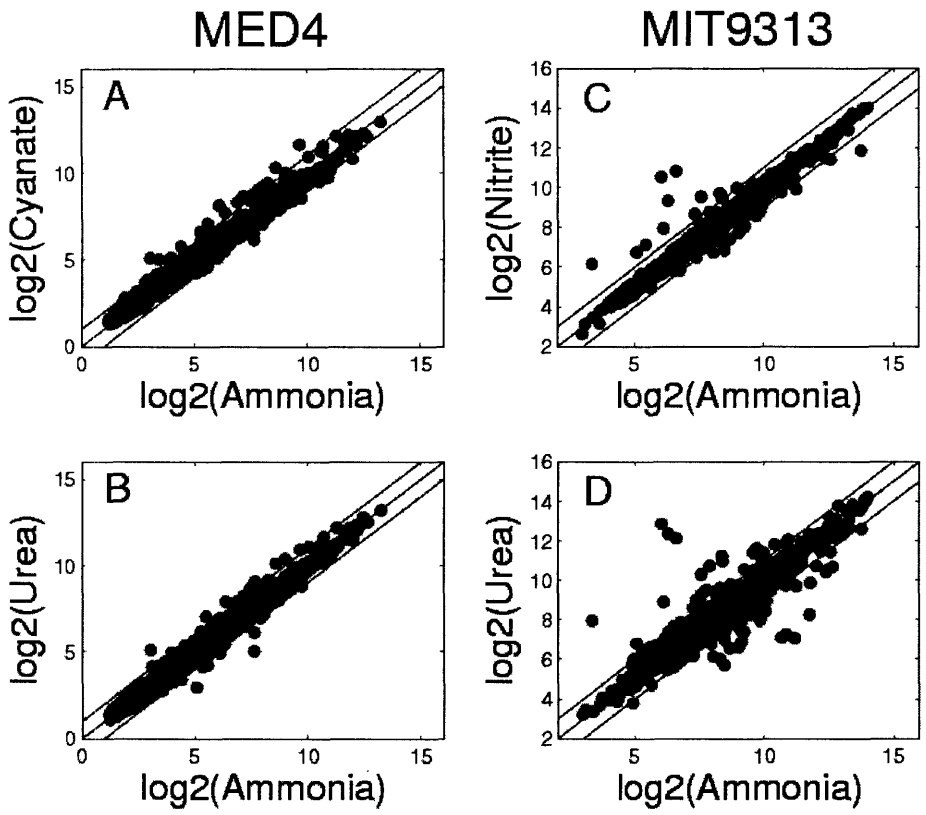


Fig. 3. Comparison of expression profiles of *Prochlorococcus* MED4 (**A-B**) and MIT9313 (**C-D**) grown on alternative nitrogen sources, relative to ammonium. Each data point represents the $\log_2$-transformed mean of duplicate cultures. Solid lines show 2-fold change in expression.

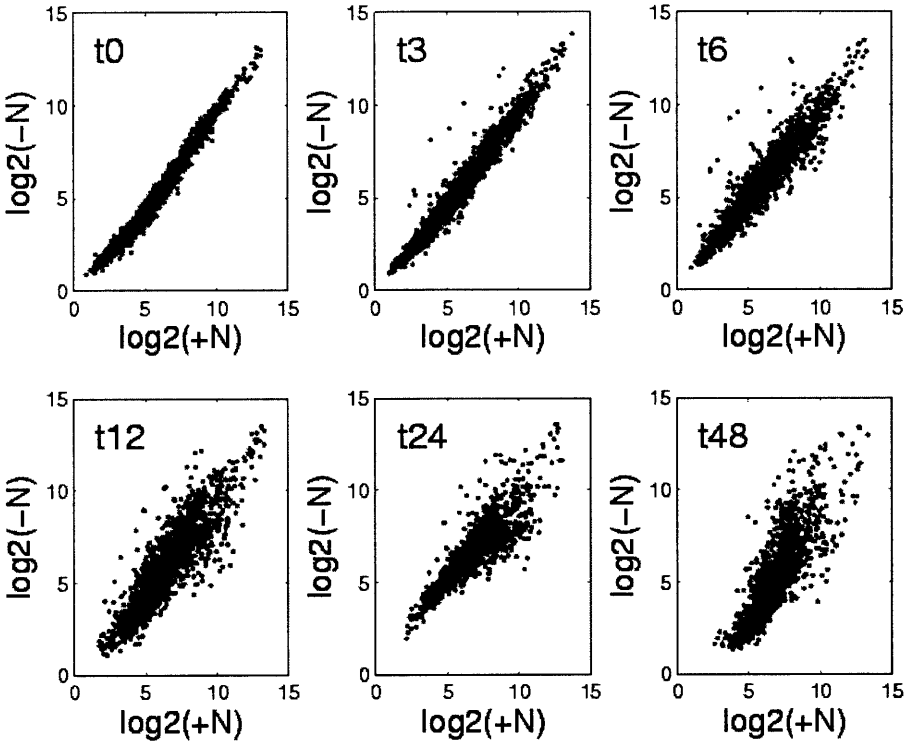


Fig. 4. Comparison of *Prochlorococcus* MED4 expression profiles from N-starvation time course. Each data point represents a $\log_2$-transformed mean of duplicate cultures in ±N media. Expression profiles are compared for each time point following transfer of the -N treatments to media lacking nitrogen: 0,3,6,12,24,48 hours.

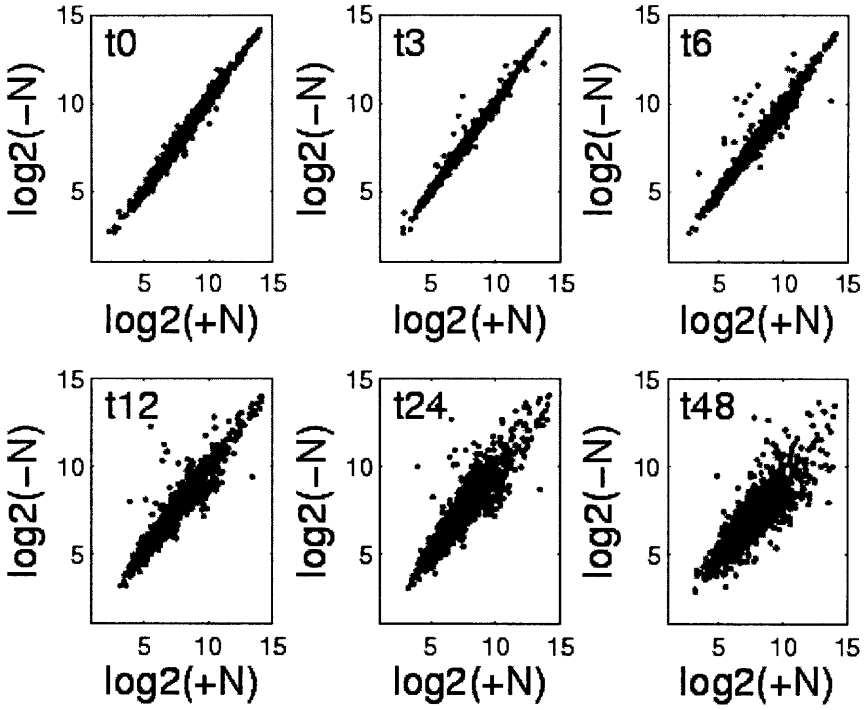


Fig. 5. Comparison of *Prochlorococcus* MIT9313 expression profiles from N-starvation time course. Each data point represents a $\log_2$-transformed mean of duplicate cultures in ±N media. Expression profiles are compared for each time point following transfer of the -N treatments to media lacking nitrogen: 0,3,6,12,24,48 hours.

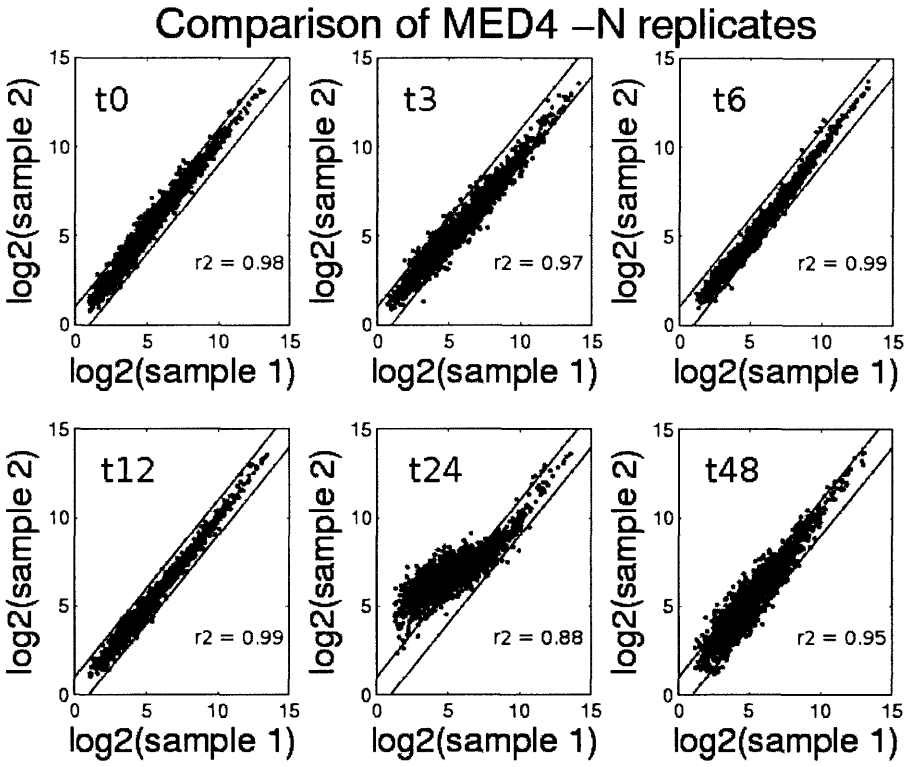


Fig. 6. Comparison of expression profiles from replicate *Prochlorococcus* MED4 cultures in the -N treatments for each time point. Correlation coefficients for expression profiles of replicate cultures are shown in each panel. Solid lines show 2-fold change in expression.

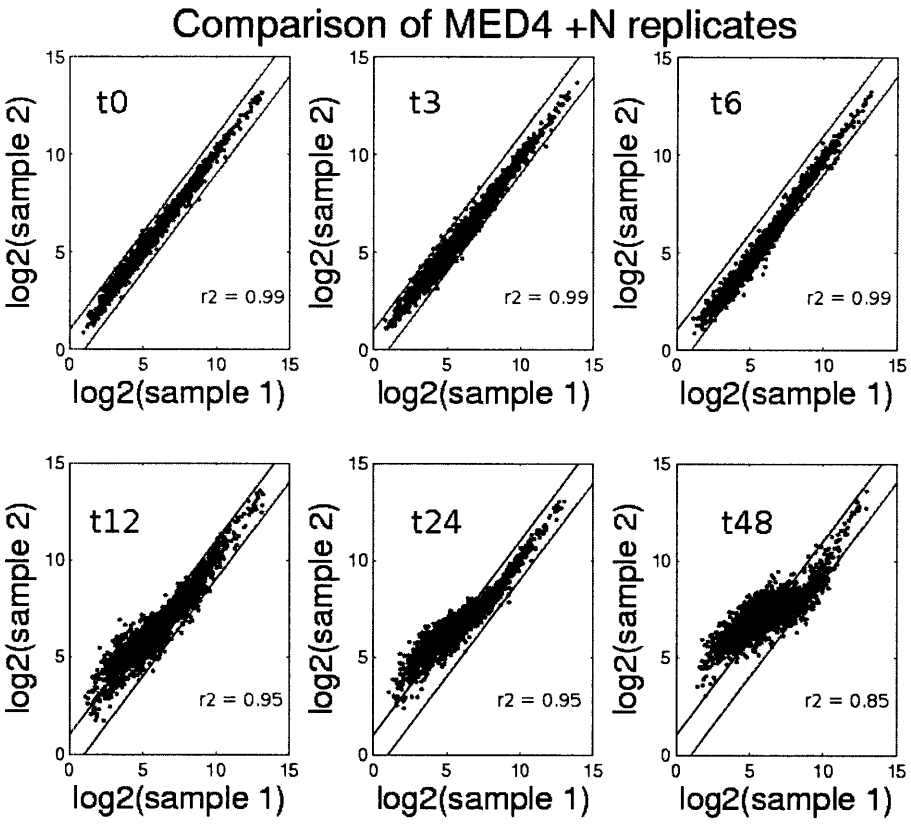


Fig. 7. Comparison of expression profiles from replicate *Prochlorococcus* MED4 cultures in the $+NH_4$ treatments for each time point. Correlation coefficients for expression profiles of replicate cultures are shown in each panel. Solid lines show 2-fold change in expression.

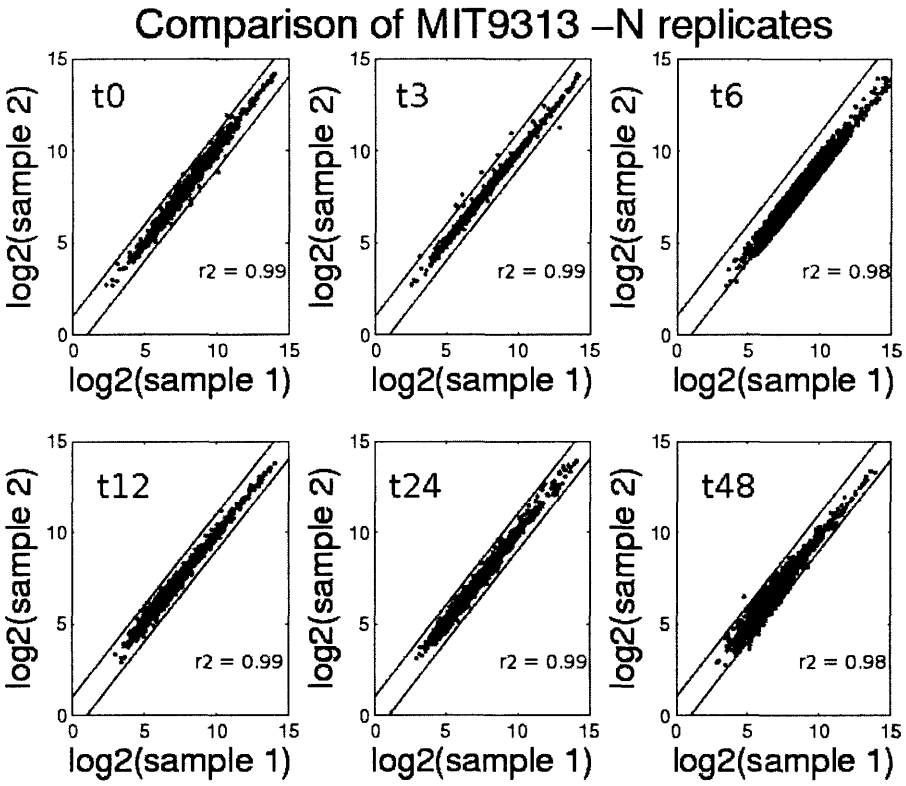


Fig. 8. Comparison of expression profiles from replicate *Prochlorococcus* MIT9313 cultures in the -N treatments for each time point. Correlation coefficients for expression profiles of replicate cultures are shown in each panel. Solid lines show 2-fold change in expression.

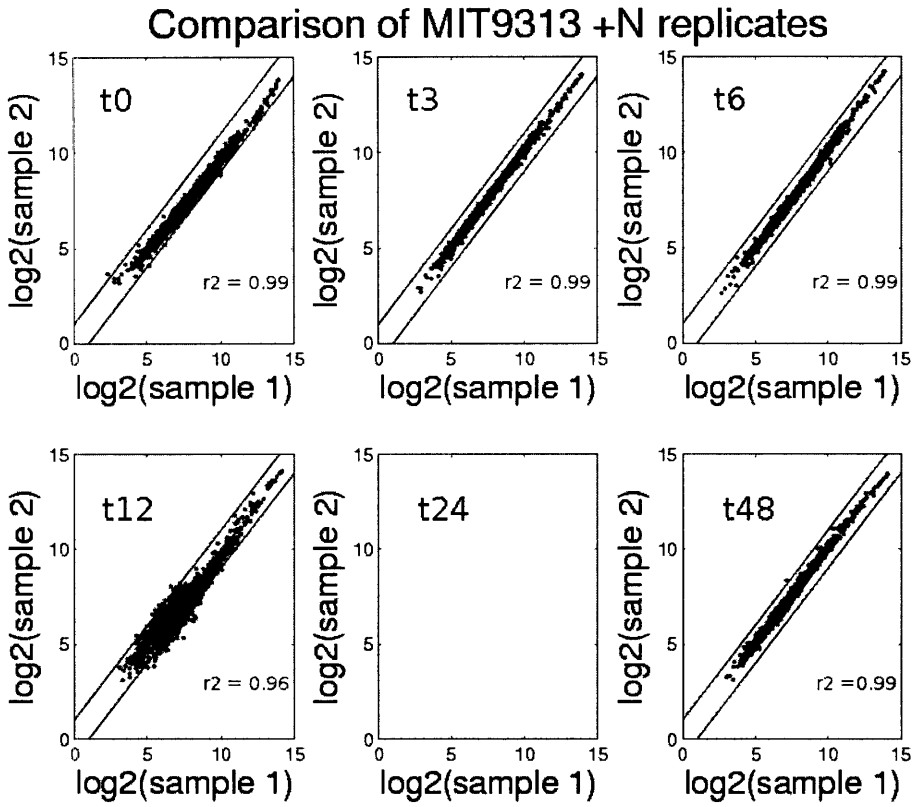


Fig. 9. Comparison of expression profiles from replicate *Prochlorococcus* MIT9313 cultures in the $+NH_4$ treatments for each time point. Correlation coefficients for expression profiles of replicate cultures are shown in each panel. Solid lines show 2-fold change in expression. No data is shown for the t=24 hr time point because these samples were lost during array hybridization.

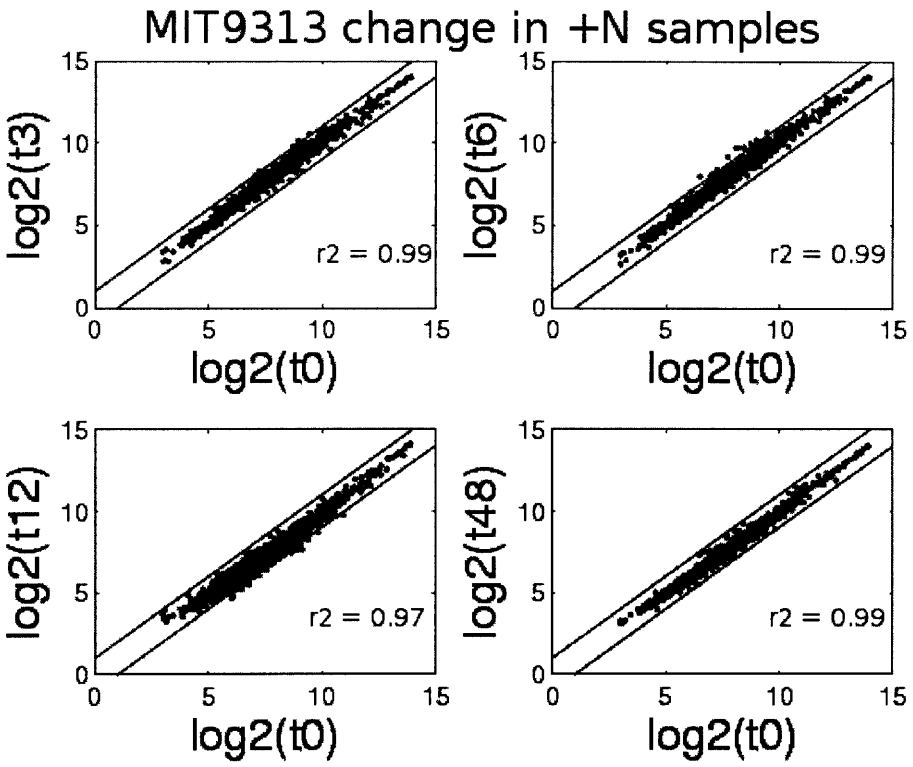


Fig. 10. Comparison of *Prochlorococcus* MIT9313 gene expression across time points in the $+NH_4$ treatments. Each datapoint represents the log-transformed mean of replicate cultures. Correlation coefficients for expression profiles between t=0 hrs. and later time points are shown in each panel. Solid lines show 2-fold change in expression.

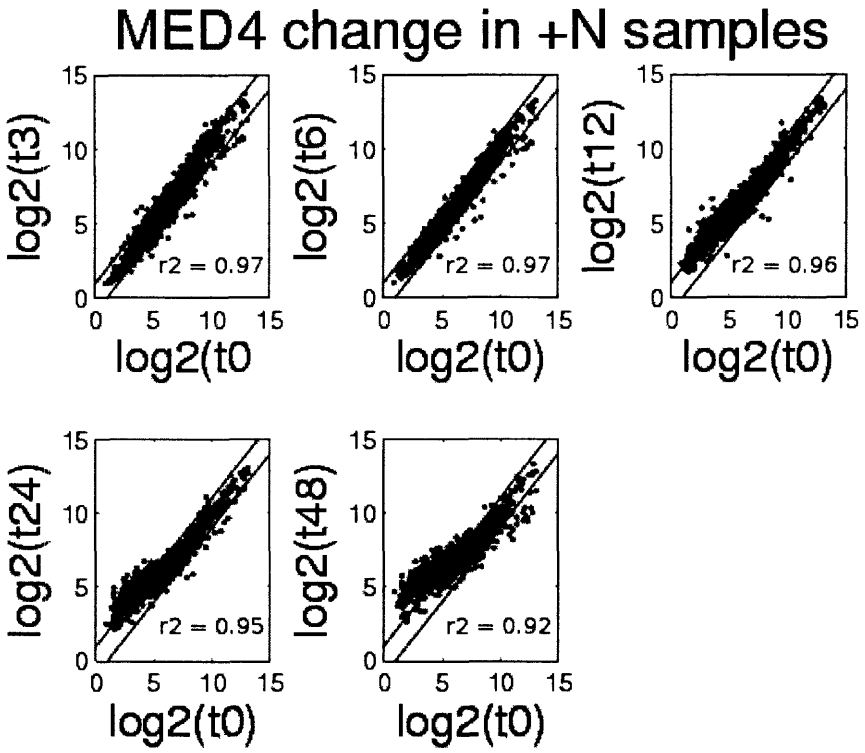


Fig. 11. Comparison of *Prochlorococcus* MED4 gene expression across time points in the $+NH_4$ treatments. Each datapoint represents the log-transformed mean of replicate cultures. Correlation coefficients for expression profiles between t=0 hrs. and later time points are shown in each panel. Solid lines show 2-fold change in expression.

Position:	1	2	3	4	5-12	13	14	15	16
A	0.24	-∞	-∞	0.86	0	0.05	0.95	-∞	0.71
C	0.1	-∞	-∞	0	0	0	0	1	0.19
G	0.1	1	-∞	0.1	0	0	0.05	-∞	0.05
T	0.56	-∞	1	0.05	0	0.95	0	-∞	0.05

Fig. 12. Scoring matrix used to detect putative NtcA-binding sites in the promoters of *Prochlorococcus* MED4 and MIT9313. Matrix elements were defined by the nucleotide frequencies of the consensus cyanobacterial NtcA binding site (Herrero et al., 2001).

K-means clustering of MED4 genes from N-starvation experiment

cluster 11:

PMM0064	PMM0072	PMM0109	PMM0119	PMM0140	PMM0141
PMM0148	PMM0152	PMM0212	PMM0274	PMM0280	PMM0284
PMM0286	PMM0308	PMM0333	PMM0349	PMM0415	PMM0420
PMM0478	PMM0515	PMM0525	PMM0561	PMM0562	PMM0563
PMM0585	PMM0590	PMM0594	PMM0608	PMM0628	PMM0633
PMM0645	PMM0652	PMM0663	PMM0680	PMM0696	PMM0701
PMM0716	PMM0749	PMM0791	PMM0834	PMM0839	PMM0840
PMM0877	PMM0898	PMM0913	PMM0925	PMM0955	PMM1021
PMM1026	PMM1034	PMM1040	PMM1085	PMM1087	PMM1089
PMM1095	PMM1099	PMM1116	PMM1187	PMM1209	PMM1216
PMM1218	PMM1237	PMM1242	PMM1243	PMM1255	PMM1266
PMM1290	PMM1303	PMM1343	PMM1356	PMM1574	PMM1598
PMM1599	PMM1716	PMM1244	PMM0094	PMM0820	PMM0820
PMM0820					

cluster 21:

PMM0550	PMM0551	PMM1450	PMM1452	PMM1453	PMM1454
PMM1455	PMM1455	PMM1455			

cluster 7:

PMM0134	PMM0142	PMM0201	PMM0202	PMM0203	PMM0204
PMM0296	PMM0312	PMM0410	PMM0510	PMM0541	PMM0560
PMM0609	PMM0622	PMM0754	PMM0769	PMM0902	PMM0907
PMM0992	PMM1191	PMM1264	PMM1285	PMM1345	PMM1440
PMM1489	PMM1506	PMM1509	PMM1510	PMM1511	PMM1532
PMM1534	PMM1536	PMM1537	PMM1542	PMM1543	PMM1544
PMM1545	PMM1546	PMM1548	PMM1608	PMM1609	PMM1644
PMM0253	PMM0299	PMM0870	PMM1183	PMM1538	PMM0347
PMM0869	PMM1375	PMM1547	PMM1547	PMM1547	

cluster 26:

PMM0007	PMM0045	PMM0057	PMM0127	PMM0131	PMM0133
PMM0167	PMM0182	PMM0277	PMM0289	PMM0290	PMM0319
PMM0330	PMM0344	PMM0345	PMM0351	PMM0354	PMM0388
PMM0394	PMM0400	PMM0442	PMM0444	PMM0449	PMM0456
PMM0458	PMM0460	PMM0479	PMM0490	PMM0492	PMM0538
PMM0653	PMM0668	PMM0693	PMM0705	PMM0727	PMM0734
PMM0746	PMM0776	PMM0792	PMM0822	PMM0831	PMM0888
PMM0890	PMM0891	PMM0911	PMM0917	PMM0919	PMM0927
PMM0938	PMM0939	PMM0948	PMM0968	PMM0980	PMM0986
PMM1012	PMM1014	PMM1025	PMM1031	PMM1049	PMM1068
PMM1081	PMM1120	PMM1124	PMM1163	PMM1174	PMM1198
PMM1213	PMM1241	PMM1253	PMM1287	PMM1296	PMM1300
PMM1380	PMM1429	PMM1466	PMM1473	PMM1646	PMM1647
PMM1658	PMM1679	PMM1684	PMM1694	PMM1699	PMM0381
PMM0006	PMM0165	PMM0517	PMM1319	PMM1641	PMM1641
PMM1641					

cluster 17:

PMM0032	PMM0035	PMM0055	PMM0063	PMM0086	PMM0088
PMM0091	PMM0116	PMM0145	PMM0150	PMM0159	PMM0172

PMM0224 PMM0235 PMM0243 PMM0244 PMM0263 PMM0294
PMM0313 PMM0334 PMM0342 PMM0343 PMM0363 PMM0364
PMM0366 PMM0368 PMM0372 PMM0378 PMM0446 PMM0448
PMM0519 PMM0546 PMM0651 PMM0661 PMM0726 PMM0742
PMM0795 PMM0804 PMM0883 PMM0936 PMM0941 PMM0965
PMM0975 PMM0996 PMM1013 PMM1045 PMM1097 PMM1119
PMM1121 PMM1123 PMM1128 PMM1149 PMM1157 PMM1169
PMM1312 PMM1313 PMM1314 PMM1409 PMM1412 PMM1424
PMM1434 PMM1478 PMM1479 PMM1514 PMM1601 PMM1602
PMM1616 PMM1625 PMM1667 PMM1678 PMM1689 PMM1700
PMM0103 PMM0714 PMM0714 PMM0714

cluster 2:
PMM0050 PMM0053 PMM0074 PMM0108 PMM0122 PMM0130
PMM0174 PMM0192 PMM0222 PMM0240 PMM0249 PMM0303
PMM0362 PMM0392 PMM0524 PMM0537 PMM0569 PMM0617
PMM0625 PMM0630 PMM0706 PMM0708 PMM0718 PMM0720
PMM0721 PMM0728 PMM0730 PMM0798 PMM0865 PMM0874
PMM0879 PMM0885 PMM0905 PMM0954 PMM0956 PMM0966
PMM0973 PMM1017 PMM1039 PMM1056 PMM1072 PMM1103
PMM1105 PMM1126 PMM1137 PMM1155 PMM1222 PMM1238
PMM1268 PMM1308 PMM1347 PMM1348 PMM1360 PMM1370
PMM1415 PMM1497 PMM1517 PMM1564 PMM1592 PMM1597
PMM1627 PMM1690 PMM0915 PMM1043 PMM1362 PMM1393
PMM1715 PMM1715 PMM1715

cluster 22:
PMM0010 PMM0079 PMM0080 PMM0097 PMM0279 PMM0393
PMM0430 PMM0454 PMM0481 PMM0484 PMM0506 PMM0557
PMM0574 PMM0576 PMM0578 PMM0588 PMM0601 PMM0602
PMM0636 PMM0639 PMM0640 PMM0665 PMM0669 PMM0735
PMM0752 PMM0764 PMM0789 PMM0821 PMM0826 PMM0836
PMM0837 PMM0899 PMM0900 PMM0933 PMM0949 PMM0952
PMM0967 PMM1000 PMM1004 PMM1033 PMM1139 PMM1167
PMM1208 PMM1219 PMM1224 PMM1258 PMM1278 PMM1282
PMM1674 PMM1675 PMM1711 PMM1714 PMM0437 PMM0575
PMM0624 PMM0846 PMM1374 PMM1374 PMM1374

cluster 1:
PMM0004 PMM0024 PMM0036 PMM0044 PMM0059 PMM0065
PMM0083 PMM0136 PMM0137 PMM0153 PMM0193 PMM0195
PMM0221 PMM0238 PMM0260 PMM0270 PMM0275 PMM0323
PMM0503 PMM0512 PMM0528 PMM0603 PMM0755 PMM0928
PMM1008 PMM1029 PMM1070 PMM1125 PMM1179 PMM1192
PMM1200 PMM1204 PMM1226 PMM1236 PMM1265 PMM1273
PMM1337 PMM1353 PMM1382 PMM1468 PMM1481 PMM1491
PMM1493 PMM1501 PMM1529 PMM1561 PMM1590 PMM1614
PMM1618 PMM1654 PMM0835 PMM0061 PMM1295 PMM1295
PMM1295

cluster 18:
PMM0005 PMM0037 PMM0054 PMM0058 PMM0066 PMM0068
PMM0110 PMM0117 PMM0132 PMM0158 PMM0162 PMM0173
PMM0191 PMM0196 PMM0218 PMM0234 PMM0241 PMM0241
PMM0341 PMM0355 PMM0403 PMM0431 PMM0450 PMM0451

PMM0463	PMM0493	PMM0545	PMM0565	PMM0570	PMM0571
PMM0572	PMM0577	PMM0579	PMM0600	PMM0629	PMM0646
PMM0650	PMM0655	PMM0677	PMM0678	PMM0683	PMM0704
PMM0707	PMM0713	PMM0748	PMM0765	PMM0806	PMM0823
PMM0824	PMM0825	PMM0845	PMM0847	PMM0854	PMM0855
PMM0873	PMM0884	PMM0887	PMM0897	PMM0909	PMM0922
PMM0931	PMM0942	PMM0947	PMM0953	PMM0960	PMM0984
PMM0985	PMM1019	PMM1057	PMM1096	PMM1106	PMM1107
PMM1108	PMM1133	PMM1153	PMM1165	PMM1170	PMM1193
PMM1196	PMM1211	PMM1239	PMM1256	PMM1301	PMM1305
PMM1322	PMM1329	PMM1330	PMM1331	PMM1333	PMM1351
PMM1366	PMM1386	PMM1388	PMM1392	PMM1401	PMM1414
PMM1419	PMM1444	PMM1477	PMM1528	PMM1563	PMM1565
PMM1569	PMM1571	PMM1575	PMM1626	PMM1633	PMM1648
PMM1656	PMM1668	PMM0715	PMM0828	PMM0476	PMM0732
PMM0950	PMM0979	PMM1275	PMM1275	PMM1275	

cluster 30:

PMM0049	PMM0078	PMM0107	PMM0129	PMM0155	PMM0156
PMM0271	PMM0358	PMM0382	PMM0390	PMM0408	PMM0440
PMM0498	PMM0516	PMM0711	PMM0724	PMM0729	PMM0841
PMM0849	PMM0904	PMM0978	PMM0994	PMM1027	PMM1047
PMM1144	PMM1328	PMM1389	PMM1410	PMM1417	PMM1420
PMM1426	PMM1469	PMM1476	PMM1502	PMM1503	PMM1533
PMM1573	PMM1620	PMM1698	PMM1698	PMM1698	

cluster 13:

PMM0017	PMM0026	PMM0060	PMM0163	PMM0168	PMM0206
PMM0215	PMM0281	PMM0292	PMM0305	PMM0346	PMM0417
PMM0533	PMM0534	PMM0564	PMM0580	PMM0595	PMM0619
PMM0641	PMM0644	PMM0657	PMM0759	PMM0786	PMM0843
PMM0851	PMM0893	PMM0923	PMM0924	PMM0929	PMM0962
PMM1042	PMM1051	PMM1093	PMM1094	PMM1110	PMM1113
PMM1181	PMM1229	PMM1230	PMM1231	PMM1299	PMM1483
PMM1582	PMM1665	PMM1666	PMM1688	PMM0857	PMM1279
PMM1279	PMM1279				

cluster 16:

PMM0047	PMM0126	PMM0170	PMM0179	PMM0186	PMM0197
PMM0198	PMM0207	PMM0233	PMM0247	PMM0304	PMM0339
PMM0361	PMM0375	PMM0376	PMM0401	PMM0414	PMM0419
PMM0422	PMM0424	PMM0438	PMM0459	PMM0464	PMM0470
PMM0472	PMM0520	PMM0523	PMM0539	PMM0566	PMM0589
PMM0620	PMM0631	PMM0634	PMM0647	PMM0681	PMM0703
PMM0761	PMM0772	PMM0811	PMM0860	PMM0862	PMM0875
PMM0882	PMM0896	PMM0914	PMM0916	PMM0921	PMM0932
PMM0937	PMM0972	PMM0974	PMM0976	PMM0981	PMM0990
PMM1001	PMM1006	PMM1023	PMM1077	PMM1104	PMM1115
PMM1136	PMM1156	PMM1162	PMM1164	PMM1168	PMM1178
PMM1202	PMM1261	PMM1269	PMM1270	PMM1280	PMM1318
PMM1358	PMM1361	PMM1367	PMM1372	PMM1383	PMM1395
PMM1416	PMM1421	PMM1423	PMM1464	PMM1474	PMM1496
PMM1513	PMM1526	PMM1566	PMM1583	PMM1593	PMM1606
PMM1649	PMM1692	PMM1709	PMM1712	PMM0586	PMM0695
PMM1114	PMM1130	PMM1130	PMM1130		

cluster 23:
PMM0338 PMM0548 PMM0684 PMM0817 PMM0988 PMM0997
PMM1135 PMM1262 PMM1397 PMM1562 PMM1672 PMM0252
PMM0818 PMM1118 PMM1118 PMM1384 PMM1385 PMM1396
PMM1404 PMM1404 PMM1404

cluster 29:
PMM0552 PMM0583 PMM1028 PMM1402 PMM1439 PMM1451
PMM1456 PMM1457 PMM1507 PMM1540 PMM1541 PMM1549
PMM1550 PMM1551 PMM1552 PMM1553 PMM1554 PMM1555
PMM1556 PMM1557 PMM1558 PMM1610 PMM1706 PMM1706
PMM1706

cluster 6:
PMM0014 PMM0052 PMM0070 PMM0076 PMM0077 PMM0081
PMM0095 PMM0102 PMM0104 PMM0139 PMM0151 PMM0171
PMM0188 PMM0216 PMM0254 PMM0261 PMM0264 PMM0269
PMM0276 PMM0278 PMM0318 PMM0331 PMM0350 PMM0357
PMM0389 PMM0399 PMM0402 PMM0406 PMM0455 PMM0457
PMM0466 PMM0473 PMM0487 PMM0587 PMM0591 PMM0597
PMM0598 PMM0658 PMM0674 PMM0694 PMM0773 PMM0808
PMM0809 PMM0850 PMM0871 PMM0886 PMM0895 PMM0940
PMM0951 PMM0977 PMM0995 PMM1018 PMM1065 PMM1102
PMM1112 PMM1159 PMM1172 PMM1173 PMM1177 PMM1195
PMM1201 PMM1206 PMM1212 PMM1233 PMM1257 PMM1311
PMM1364 PMM1381 PMM1425 PMM1447 PMM1470 PMM1480
PMM1505 PMM1518 PMM1527 PMM1576 PMM1579 PMM1635
PMM1645 PMM1651 PMM1682 PMM1686 PMM1705 PMM0181
PMM1100 PMM1448 PMM1448 PMM1448

cluster 25:
PMM0001 PMM0008 PMM0012 PMM0015 PMM0020 PMM0031
PMM0039 PMM0048 PMM0051 PMM0073 PMM0092 PMM0101
PMM0115 PMM0121 PMM0123 PMM0143 PMM0144 PMM0146
PMM0160 PMM0161 PMM0164 PMM0185 PMM0208 PMM0223
PMM0236 PMM0237 PMM0258 PMM0293 PMM0301 PMM0320
PMM0369 PMM0373 PMM0379 PMM0385 PMM0395 PMM0411
PMM0418 PMM0443 PMM0445 PMM0480 PMM0482 PMM0485
PMM0486 PMM0502 PMM0532 PMM0558 PMM0611 PMM0635
PMM0637 PMM0638 PMM0667 PMM0688 PMM0725 PMM0739
PMM0758 PMM0774 PMM0775 PMM0777 PMM0779 PMM0780
PMM0790 PMM0858 PMM0878 PMM0934 PMM0961 PMM0982
PMM0998 PMM1075 PMM1080 PMM1092 PMM1129 PMM1146
PMM1151 PMM1154 PMM1158 PMM1272 PMM1286 PMM1307
PMM1309 PMM1321 PMM1339 PMM1342 PMM1349 PMM1355
PMM1368 PMM1373 PMM1376 PMM1387 PMM1413 PMM1422
PMM1442 PMM1443 PMM1465 PMM1498 PMM1512 PMM1588
PMM1589 PMM1594 PMM1611 PMM1622 PMM1630 PMM1642
PMM1669 PMM1702 PMM1703 PMM1707 PMM0740 PMM0698
PMM0736 PMM1007 PMM1363 PMM1363 PMM1363

cluster 27:
PMM0013 PMM0043 PMM0084 PMM0106 PMM0124 PMM0128
PMM0149 PMM0154 PMM0210 PMM0226 PMM0232 PMM0259

PMM0268	PMM0316	PMM0326	PMM0327	PMM0335	PMM0377
PMM0405	PMM0407	PMM0435	PMM0465	PMM0467	PMM0477
PMM0499	PMM0526	PMM0554	PMM0555	PMM0581	PMM0593
PMM0618	PMM0743	PMM0744	PMM0856	PMM0861	PMM0872
PMM0910	PMM0943	PMM0963	PMM0989	PMM1005	PMM1015
PMM1030	PMM1032	PMM1061	PMM1066	PMM1088	PMM1090
PMM1098	PMM1145	PMM1186	PMM1190	PMM1245	PMM1252
PMM1283	PMM1288	PMM1293	PMM1294	PMM1332	PMM1340
PMM1352	PMM1354	PMM1359	PMM1365	PMM1378	PMM1428
PMM1431	PMM1435	PMM1459	PMM1484	PMM1492	PMM1499
PMM1568	PMM1581	PMM1624	PMM1652	PMM1653	PMM1670
PMM1671	PMM1673	PMM1708	PMM0034	PMM0093	PMM0231
PMM0853	PMM0697	PMM1408	PMM1408	PMM1408	

cluster 28:

PMM0018	PMM0041	PMM0135	PMM0200	PMM0214	PMM0295
PMM0409	PMM0494	PMM0511	PMM0535	PMM0542	PMM0673
PMM0741	PMM0747	PMM0797	PMM0829	PMM1071	PMM1101
PMM1184	PMM1251	PMM1297	PMM1298	PMM1325	PMM1335
PMM1336	PMM1522	PMM1539	PMM1572	PMM1585	PMM1643
PMM1663	PMM1664	PMM0783	PMM0783	PMM0783	

cluster 3:

PMM0033	PMM0075	PMM0090	PMM0120	PMM0169	PMM0266
PMM0321	PMM0340	PMM0360	PMM0380	PMM0387	PMM0491
PMM0509	PMM0544	PMM0626	PMM0656	PMM0660	PMM0666
PMM0719	PMM0731	PMM0733	PMM0737	PMM0757	PMM0770
PMM0908	PMM0918	PMM0945	PMM0957	PMM0964	PMM0969
PMM0971	PMM1016	PMM1024	PMM1037	PMM1038	PMM1050
PMM1069	PMM1073	PMM1078	PMM1197	PMM1461	PMM1490
PMM1584	PMM1595	PMM1596	PMM1615	PMM1677	PMM1687
PMM1697	PMM0432	PMM0738	PMM1036	PMM1036	PMM1036

cluster 9:

PMM0019	PMM0022	PMM0028	PMM0029	PMM0040	PMM0089
PMM0098	PMM0100	PMM0112	PMM0125	PMM0138	PMM0175
PMM0177	PMM0178	PMM0180	PMM0184	PMM0229	PMM0242
PMM0255	PMM0256	PMM0257	PMM0309	PMM0322	PMM0352
PMM0384	PMM0391	PMM0397	PMM0398	PMM0421	PMM0426
PMM0488	PMM0518	PMM0522	PMM0596	PMM0613	PMM0659
PMM0675	PMM0682	PMM0692	PMM0712	PMM0723	PMM0750
PMM0771	PMM0787	PMM0793	PMM0794	PMM0805	PMM0833
PMM0848	PMM0852	PMM0859	PMM0880	PMM0881	PMM1002
PMM1009	PMM1035	PMM1044	PMM1082	PMM1086	PMM1138
PMM1189	PMM1199	PMM1203	PMM1215	PMM1220	PMM1223
PMM1225	PMM1232	PMM1274	PMM1292	PMM1304	PMM1320
PMM1411	PMM1432	PMM1445	PMM1458	PMM1472	PMM1488
PMM1515	PMM1570	PMM1586	PMM1613	PMM1621	PMM1660
PMM1676	PMM1683	PMM1683	PMM1683		

cluster 12:

PMM0016	PMM0166	PMM0251	PMM0265	PMM0315	PMM0500
PMM0627	PMM0642	PMM0700	PMM0784	PMM0844	PMM0876
PMM0906	PMM0993	PMM1055	PMM1109	PMM1240	PMM1247
PMM1523	PMM1524	PMM1607	PMM1650	PMM1704	PMM0314

PMM0429 PMM1011 PMM1182 PMM1182 PMM1182

cluster 15:
PMM0025 PMM0027 PMM0085 PMM0105 PMM0205 PMM0282
PMM0297 PMM0324 PMM0367 PMM0423 PMM0436 PMM0462
PMM0495 PMM0496 PMM0536 PMM0543 PMM0547 PMM0556
PMM0615 PMM0722 PMM0762 PMM0866 PMM0987 PMM1054
PMM1059 PMM1060 PMM1067 PMM1079 PMM1152 PMM1185
PMM1205 PMM1250 PMM1323 PMM1326 PMM1377 PMM1407
PMM1485 PMM1487 PMM1494 PMM1500 PMM1530 PMM1531
PMM1535 PMM1617 PMM1619 PMM1636 PMM1639 PMM0474
PMM0751 PMM1058 PMM0812 PMM0812 PMM0812

cluster 20:
PMM0056 PMM0096 PMM0189 PMM0213 PMM0225 PMM0227
PMM0239 PMM0267 PMM0273 PMM0283 PMM0285 PMM0288
PMM0306 PMM0328 PMM0356 PMM0383 PMM0428 PMM0505
PMM0514 PMM0527 PMM0529 PMM0553 PMM0582 PMM0584
PMM0643 PMM0648 PMM0670 PMM0672 PMM0679 PMM0768
PMM0782 PMM0799 PMM0830 PMM0842 PMM0903 PMM1084
PMM1142 PMM1228 PMM1263 PMM1284 PMM1316 PMM1334
PMM1346 PMM1369 PMM1379 PMM1449 PMM1471 PMM1486
PMM1567 PMM1591 PMM1659 PMM0686 PMM1003 PMM1482
PMM1143 PMM1143 PMM1143

cluster 14:
PMM0246 PMM0336 PMM0370 PMM0687 PMM0920 PMM0958
PMM0970 PMM1041 PMM1462 PMM0374 PMM0374 PMM0374

cluster 8:
PMM0023 PMM0046 PMM0228 PMM0329 PMM0452 PMM0453
PMM0469 PMM0549 PMM0599 PMM0605 PMM0710 PMM0766
PMM0767 PMM0781 PMM0785 PMM0901 PMM1350 PMM1436
PMM1438 PMM1508 PMM1519 PMM1520 PMM1629 PMM1655
PMM1662 PMM0272 PMM0307 PMM0468 PMM0540 PMM1661
PMM0691 PMM1578 PMM1578 PMM1578

cluster 4:
PMM0021 PMM0042 PMM0069 PMM0071 PMM0082 PMM0113
PMM0118 PMM0176 PMM0183 PMM0190 PMM0217 PMM0248
PMM0302 PMM0310 PMM0332 PMM0404 PMM0413 PMM0425
PMM0427 PMM0433 PMM0439 PMM0489 PMM0497 PMM0568
PMM0592 PMM0604 PMM0607 PMM0610 PMM0612 PMM0616
PMM0654 PMM0671 PMM0756 PMM0803 PMM0868 PMM1046
PMM1048 PMM1053 PMM1064 PMM1127 PMM1140 PMM1161
PMM1207 PMM1221 PMM1248 PMM1302 PMM1306 PMM1357
PMM1371 PMM1403 PMM1418 PMM1460 PMM1467 PMM1521
PMM1525 PMM1560 PMM1580 PMM1628 PMM1681 PMM1693
PMM1695 PMM1696 PMM1701 PMM1631 PMM1632 PMM1632
PMM1632

cluster 24:
PMM0002 PMM0003 PMM0009 PMM0011 PMM0067 PMM0099
PMM0111 PMM0114 PMM0147 PMM0157 PMM0209 PMM0211
PMM0250 PMM0291 PMM0396 PMM0412 PMM0434 PMM0513

PMM0521	PMM0531	PMM0559	PMM0567	PMM0614	PMM0621
PMM0632	PMM0664	PMM0676	PMM0702	PMM0709	PMM0717
PMM0763	PMM0796	PMM0814	PMM0827	PMM0889	PMM0946
PMM0983	PMM1062	PMM1063	PMM1083	PMM1091	PMM1111
PMM1122	PMM1132	PMM1147	PMM1175	PMM1176	PMM1188
PMM1210	PMM1227	PMM1234	PMM1235	PMM1246	PMM1249
PMM1254	PMM1259	PMM1267	PMM1271	PMM1281	PMM1291
PMM1310	PMM1317	PMM1324	PMM1327	PMM1338	PMM1341
PMM1406	PMM1430	PMM1446	PMM1504	PMM1559	PMM1577
PMM1600	PMM1603	PMM1634	PMM1637	PMM1638	PMM1657
PMM0471	PMM0999	PMM0838	PMM0944	PMM1022	PMM1022
PMM1022					

cluster 19:

PMM0030	PMM0087	PMM0220	PMM0245	PMM0337	PMM0365
PMM0371	PMM0447	PMM0689	PMM0810	PMM0819	PMM1074
PMM1134	PMM1391	PMM1463	PMM1623	PMM0359	PMM0690
PMM1390	PMM1390	PMM1390			

cluster 10:

PMM0311	PMM0325	PMM0348	PMM0416	PMM0461	PMM0475
PMM0483	PMM0507	PMM0508	PMM0530	PMM0573	PMM0606
PMM0649	PMM0699	PMM0753	PMM0760	PMM0788	PMM0801
PMM0802	PMM0863	PMM0867	PMM0894	PMM0912	PMM0926
PMM0930	PMM1052	PMM1131	PMM1148	PMM1150	PMM1171
PMM1180	PMM1289	PMM1315	PMM1344	PMM1394	PMM1400
PMM1405	PMM1437	PMM0298	PMM0300	PMM0317	PMM1441
PMM1604	PMM0864	PMM0864	PMM0864		

cluster 5:

PMM0038	PMM0187	PMM0194	PMM0199	PMM0262	PMM0287
PMM0353	PMM0386	PMM0441	PMM0504	PMM0623	PMM0662
PMM0745	PMM0778	PMM0807	PMM0813	PMM0832	PMM0892
PMM0935	PMM0959	PMM0991	PMM1010	PMM1076	PMM1141
PMM1160	PMM1166	PMM1214	PMM1217	PMM1260	PMM1433
PMM1475	PMM1495	PMM1516	PMM1605	PMM1640	PMM1685
PMM1691	PMM1713	PMM1020	PMM1020	PMM1020	

K-means clustering of MIT9313 genes from N-starvation experiment

cluster 11:

PMT0010	PMT0013	PMT0046	PMT0049	PMT0085	PMT0087
PMT0093	PMT0103	PMT0112	PMT0171	PMT0210	PMT0217
PMT0335	PMT0390	PMT0397	PMT0402	PMT0407	PMT0426
PMT0435	PMT0455	PMT0460	PMT0469	PMT0471	PMT0474
PMT0492	PMT0494	PMT0505	PMT0531	PMT0543	PMT0555
PMT0557	PMT0558	PMT0578	PMT0579	PMT0659	PMT0704
PMT0705	PMT0707	PMT0714	PMT0718	PMT0736	PMT0766
PMT0777	PMT0795	PMT0799	PMT0800	PMT0809	PMT0863
PMT0872	PMT0974	PMT1032	PMT1033	PMT1062	PMT1077
PMT1108	PMT1177	PMT1191	PMT1206	PMT1225	PMT1229
PMT1239	PMT1256	PMT1268	PMT1292	PMT1301	PMT1302
PMT1330	PMT1338	PMT1349	PMT1353	PMT1399	PMT1435
PMT1456	PMT1476	PMT1477	PMT1491	PMT1501	PMT1511
PMT1547	PMT1594	PMT1605	PMT1606	PMT1614	PMT1624
PMT1648	PMT1649	PMT1725	PMT1774	PMT1785	PMT1805
PMT1809	PMT1865	PMT1881	PMT1936	PMT1968	PMT1973
PMT1984	PMT1997	PMT1998	PMT2021	PMT2024	PMT2032
PMT2063	PMT2081	PMT2191	PMT2192	PMT2225	PMT2253
PMT2264	PMT2264	PMT2264			

cluster 21:

PMT0023	PMT0045	PMT0065	PMT0121	PMT0139	PMT0208
PMT0220	PMT0250	PMT0252	PMT0257	PMT0266	PMT0269
PMT0275	PMT0287	PMT0416	PMT0417	PMT0461	PMT0472
PMT0582	PMT0602	PMT0617	PMT0650	PMT0654	PMT0674
PMT0680	PMT0686	PMT0745	PMT0756	PMT0762	PMT0780
PMT0781	PMT0808	PMT0812	PMT0821	PMT0848	PMT0884
PMT0896	PMT0966	PMT0978	PMT1026	PMT1049	PMT1065
PMT1087	PMT1181	PMT1251	PMT1252	PMT1287	PMT1351
PMT1358	PMT1411	PMT1431	PMT1441	PMT1555	PMT1575
PMT1596	PMT1603	PMT1623	PMT1629	PMT1662	PMT1669
PMT1701	PMT1711	PMT1728	PMT1786	PMT1807	PMT1811
PMT1857	PMT1885	PMT1888	PMT1903	PMT1904	PMT1905
PMT1928	PMT1942	PMT1960	PMT1976	PMT1977	PMT2016
PMT2017	PMT2033	PMT2082	PMT2151	PMT2157	PMT2181
PMT2261	PMT2261	PMT2261			

cluster 7:

PMT0024	PMT0029	PMT0037	PMT0047	PMT0060	PMT0073
PMT0113	PMT0114	PMT0123	PMT0143	PMT0165	PMT0175
PMT0185	PMT0196	PMT0200	PMT0206	PMT0209	PMT0212
PMT0222	PMT0225	PMT0226	PMT0229	PMT0234	PMT0264
PMT0267	PMT0281	PMT0290	PMT0291	PMT0317	PMT0318
PMT0326	PMT0330	PMT0338	PMT0343	PMT0351	PMT0354
PMT0358	PMT0362	PMT0364	PMT0373	PMT0379	PMT0405
PMT0412	PMT0413	PMT0414	PMT0415	PMT0425	PMT0429
PMT0438	PMT0442	PMT0481	PMT0504	PMT0524	PMT0525
PMT0529	PMT0530	PMT0533	PMT0561	PMT0580	PMT0587
PMT0590	PMT0603	PMT0608	PMT0611	PMT0624	PMT0641
PMT0643	PMT0644	PMT0651	PMT0658	PMT0660	PMT0669
PMT0671	PMT0684	PMT0690	PMT0703	PMT0711	PMT0716

PMT0720	PMT0737	PMT0759	PMT0765	PMT0767	PMT0772
PMT0778	PMT0782	PMT0784	PMT0786	PMT0787	PMT0816
PMT0819	PMT0841	PMT0842	PMT0865	PMT0903	PMT0921
PMT0931	PMT0933	PMT0958	PMT0972	PMT0975	PMT0984
PMT0995	PMT1002	PMT1020	PMT1030	PMT1070	PMT1085
PMT1086	PMT1091	PMT1103	PMT1104	PMT1107	PMT1112
PMT1128	PMT1129	PMT1132	PMT1133	PMT1134	PMT1138
PMT1162	PMT1168	PMT1184	PMT1236	PMT1240	PMT1249
PMT1266	PMT1270	PMT1280	PMT1303	PMT1319	PMT1327
PMT1339	PMT1359	PMT1380	PMT1388	PMT1390	PMT1397
PMT1401	PMT1412	PMT1428	PMT1430	PMT1433	PMT1448
PMT1464	PMT1482	PMT1502	PMT1508	PMT1530	PMT1545
PMT1546	PMT1589	PMT1615	PMT1616	PMT1630	PMT1641
PMT1655	PMT1671	PMT1687	PMT1698	PMT1705	PMT1709
PMT1712	PMT1720	PMT1723	PMT1763	PMT1764	PMT1772
PMT1773	PMT1802	PMT1814	PMT1823	PMT1834	PMT1845
PMT1861	PMT1871	PMT1872	PMT1901	PMT1907	PMT1938
PMT1974	PMT1980	PMT1983	PMT2001	PMT2011	PMT2015
PMT2038	PMT2045	PMT2048	PMT2061	PMT2065	PMT2067
PMT2071	PMT2074	PMT2097	PMT2098	PMT2108	PMT2112
PMT2142	PMT2145	PMT2150	PMT2160	PMT2162	PMT2171
PMT2184	PMT2186	PMT2200	PMT2215	PMT2230	PMT2231
PMT2244	PMT2245	PMT2251	PMT2270	PMT0084	PMT0084
PMT0084					

cluster 26:

PMT0990	PMT0991	PMT0992	PMT0992	PMT0992	

cluster 17:

PMT0019	PMT0067	PMT0190	PMT0261	PMT0303	PMT0348
PMT0457	PMT0513	PMT0542	PMT0559	PMT0609	PMT0635
PMT0672	PMT0710	PMT0779	PMT0823	PMT0824	PMT0858
PMT0870	PMT0969	PMT0976	PMT0987	PMT1009	PMT1017
PMT1039	PMT1057	PMT1072	PMT1113	PMT1115	PMT1141
PMT1172	PMT1182	PMT1210	PMT1235	PMT1241	PMT1265
PMT1313	PMT1375	PMT1385	PMT1432	PMT1439	PMT1445
PMT1457	PMT1490	PMT1496	PMT1534	PMT1539	PMT1552
PMT1581	PMT1607	PMT1613	PMT1631	PMT1632	PMT1652
PMT1658	PMT1678	PMT1718	PMT1726	PMT1803	PMT1826
PMT1841	PMT1842	PMT1883	PMT1949	PMT1962	PMT1965
PMT2014	PMT2092	PMT2114	PMT2115	PMT2128	PMT2154
PMT2164	PMT2168	PMT2185	PMT2190	PMT2211	PMT2212
PMT2234	PMT2235	PMT2236	PMT2265	PMT2266	PMT1634
PMT1634	PMT1634				

cluster 2:

PMT0061	PMT0099	PMT0102	PMT0142	PMT0265	PMT0422
PMT0427	PMT0431	PMT0454	PMT0498	PMT0528	PMT0537
PMT0539	PMT0556	PMT0571	PMT0572	PMT0634	PMT0664
PMT0679	PMT0791	PMT0817	PMT0899	PMT0980	PMT1064
PMT1101	PMT1126	PMT1136	PMT1167	PMT1212	PMT1224
PMT1227	PMT1272	PMT1324	PMT1336	PMT1337	PMT1363
PMT1371	PMT1372	PMT1414	PMT1452	PMT1498	PMT1517
PMT1519	PMT1528	PMT1588	PMT1619	PMT1686	PMT1791
PMT1799	PMT1891	PMT1919	PMT1967	PMT2003	PMT2003

PMT2003

cluster 22:

PMT0127	PMT0201	PMT0240	PMT0352	PMT0487	PMT0508
PMT0665	PMT0697	PMT0732	PMT0733	PMT0751	PMT0760
PMT0804	PMT0811	PMT0837	PMT0885	PMT0911	PMT0949
PMT0954	PMT0957	PMT0971	PMT0982	PMT0983	PMT0988
PMT0993	PMT0994	PMT1006	PMT1068	PMT1131	PMT1276
PMT1286	PMT1293	PMT1423	PMT1565	PMT1597	PMT1609
PMT1636	PMT1676	PMT1688	PMT1873	PMT2075	PMT2166
PMT2256	PMT2256	PMT2256			

cluster 1:

PMT0001	PMT0018	PMT0025	PMT0054	PMT0070	PMT0077
PMT0078	PMT0086	PMT0089	PMT0092	PMT0096	PMT0108
PMT0118	PMT0137	PMT0254	PMT0293	PMT0295	PMT0305
PMT0386	PMT0419	PMT0420	PMT0463	PMT0466	PMT0480
PMT0506	PMT0511	PMT0521	PMT0527	PMT0544	PMT0550
PMT0562	PMT0576	PMT0591	PMT0593	PMT0594	PMT0607
PMT0612	PMT0646	PMT0681	PMT0688	PMT0698	PMT0700
PMT0713	PMT0723	PMT0731	PMT0735	PMT0744	PMT0763
PMT0773	PMT0796	PMT0853	PMT0854	PMT0879	PMT0894
PMT0948	PMT1001	PMT1024	PMT1040	PMT1058	PMT1067
PMT1074	PMT1080	PMT1099	PMT1100	PMT1116	PMT1228
PMT1234	PMT1237	PMT1244	PMT1250	PMT1257	PMT1260
PMT1263	PMT1283	PMT1289	PMT1332	PMT1356	PMT1417
PMT1426	PMT1444	PMT1532	PMT1556	PMT1559	PMT1569
PMT1690	PMT1702	PMT1724	PMT1771	PMT1775	PMT1787
PMT1795	PMT1804	PMT1852	PMT1869	PMT1890	PMT1902
PMT1911	PMT1918	PMT1935	PMT1951	PMT1958	PMT1994
PMT2007	PMT2034	PMT2044	PMT2060	PMT2084	PMT2104
PMT2135	PMT2149	PMT2152	PMT2159	PMT2161	PMT2170
PMT2189	PMT2194	PMT2209	PMT2213	PMT2214	PMT2226
PMT2262	PMT2263	PMT2274	PMT0111	PMT0111	PMT0111

cluster 18:

PMT0346	PMT0565	PMT0805	PMT0908	PMT0912	PMT0939
PMT0963	PMT0964	PMT1007	PMT1144	PMT1153	PMT1154
PMT1314	PMT1341	PMT1608	PMT1640	PMT1874	PMT1946
PMT2174	PMT2241	PMT1152	PMT1152	PMT1152	

cluster 30:

PMT0169	PMT0227	PMT1081	PMT1377	PMT1466	PMT1467
PMT1468	PMT1470	PMT1471	PMT1472	PMT1473	PMT1574
PMT1697	PMT1732	PMT1734	PMT1735	PMT1736	PMT1738
PMT1739	PMT1741	PMT1745	PMT1746	PMT1747	PMT1748
PMT1749	PMT1753	PMT1755	PMT1758	PMT1759	PMT1956
PMT2088	PMT2090	PMT2090	PMT2090		

cluster 13:

PMT0027	PMT0033	PMT0034	PMT0035	PMT0068	PMT0074
PMT0119	PMT0120	PMT0134	PMT0150	PMT0151	PMT0184
PMT0194	PMT0197	PMT0202	PMT0218	PMT0223	PMT0233
PMT0235	PMT0236	PMT0237	PMT0298	PMT0316	PMT0324
PMT0334	PMT0340	PMT0377	PMT0380	PMT0381	PMT0384

PMT0449 PMT0451 PMT0458 PMT0489 PMT0596 PMT0622
PMT0638 PMT0692 PMT0719 PMT0742 PMT0830 PMT0873
PMT0880 PMT0889 PMT0890 PMT0891 PMT0901 PMT0905
PMT0919 PMT0920 PMT0932 PMT0936 PMT0938 PMT0940
PMT0942 PMT0945 PMT0947 PMT0952 PMT1005 PMT1016
PMT1023 PMT1044 PMT1121 PMT1137 PMT1139 PMT1140
PMT1145 PMT1183 PMT1194 PMT1221 PMT1226 PMT1281
PMT1284 PMT1298 PMT1321 PMT1328 PMT1347 PMT1362
PMT1405 PMT1406 PMT1422 PMT1425 PMT1429 PMT1434
PMT1443 PMT1460 PMT1462 PMT1488 PMT1516 PMT1543
PMT1549 PMT1561 PMT1563 PMT1578 PMT1583 PMT1592
PMT1600 PMT1699 PMT1798 PMT1801 PMT1806 PMT1820
PMT1822 PMT1833 PMT1855 PMT1860 PMT1916 PMT2000
PMT2025 PMT2028 PMT2050 PMT2055 PMT2069 PMT2113
PMT2126 PMT2127 PMT2140 PMT2187 PMT2205 PMT2218
PMT2221 PMT2222 PMT2247 PMT2248 PMT2248 PMT2248

cluster 16:
PMT0072 PMT0079 PMT0128 PMT0133 PMT0149 PMT0153
PMT0174 PMT0188 PMT0192 PMT0203 PMT0211 PMT0219
PMT0231 PMT0268 PMT0288 PMT0289 PMT0294 PMT0297
PMT0306 PMT0308 PMT0337 PMT0345 PMT0355 PMT0366
PMT0370 PMT0372 PMT0375 PMT0399 PMT0501 PMT0549
PMT0563 PMT0574 PMT0581 PMT0589 PMT0655 PMT0691
PMT0734 PMT0761 PMT0783 PMT0807 PMT0818 PMT0826
PMT0828 PMT0831 PMT0881 PMT0883 PMT0892 PMT0910
PMT0944 PMT0953 PMT0955 PMT0961 PMT0968 PMT1000
PMT1014 PMT1021 PMT1053 PMT1082 PMT1084 PMT1089
PMT1222 PMT1271 PMT1340 PMT1346 PMT1348 PMT1350
PMT1379 PMT1463 PMT1487 PMT1492 PMT1535 PMT1537
PMT1540 PMT1554 PMT1579 PMT1584 PMT1593 PMT1595
PMT1598 PMT1599 PMT1626 PMT1639 PMT1673 PMT1681
PMT1714 PMT1715 PMT1722 PMT1765 PMT1788 PMT1797
PMT1808 PMT1824 PMT1839 PMT1870 PMT1876 PMT1878
PMT1910 PMT1920 PMT1932 PMT2022 PMT2042 PMT2051
PMT2057 PMT2133 PMT2156 PMT2232 PMT2237 PMT2271
PMT0996 PMT0996 PMT0996

cluster 23:
PMT0167 PMT0246 PMT0456 PMT0483 PMT0923 PMT0925
PMT0929 PMT0943 PMT1223 PMT1577 PMT1610 PMT1940
PMT2117 PMT2118 PMT2137 PMT2137 PMT2137

cluster 29:
PMT0951 PMT1831 PMT1853 PMT2229 PMT2239 PMT2240
PMT2240 PMT2240

cluster 25:
PMT0032 PMT0081 PMT0101 PMT0138 PMT0162 PMT0249
PMT0445 PMT0497 PMT0535 PMT0568 PMT0583 PMT0584
PMT0642 PMT0724 PMT0740 PMT0754 PMT0832 PMT0915
PMT0924 PMT1035 PMT1213 PMT1219 PMT1323 PMT1345
PMT1450 PMT1451 PMT1454 PMT1505 PMT1506 PMT1740
PMT1751 PMT1752 PMT1754 PMT1756 PMT1760 PMT1779
PMT1780 PMT1783 PMT1836 PMT1854 PMT1859 PMT1899

PMT1955 PMT1957 PMT2089 PMT2091 PMT1307 PMT1590
PMT1666 PMT1666 PMT1666

cluster 6:
PMT0005 PMT0022 PMT0057 PMT0058 PMT0064 PMT0129
PMT0136 PMT0147 PMT0152 PMT0157 PMT0161 PMT0170
PMT0180 PMT0181 PMT0186 PMT0198 PMT0199 PMT0214
PMT0215 PMT0230 PMT0270 PMT0271 PMT0283 PMT0302
PMT0336 PMT0367 PMT0368 PMT0391 PMT0394 PMT0396
PMT0398 PMT0406 PMT0409 PMT0434 PMT0440 PMT0452
PMT0473 PMT0476 PMT0477 PMT0486 PMT0490 PMT0510
PMT0515 PMT0548 PMT0595 PMT0597 PMT0598 PMT0614
PMT0620 PMT0647 PMT0648 PMT0653 PMT0661 PMT0662
PMT0668 PMT0678 PMT0685 PMT0693 PMT0708 PMT0715
PMT0717 PMT0727 PMT0738 PMT0749 PMT0750 PMT0768
PMT0770 PMT0775 PMT0797 PMT0801 PMT0803 PMT0814
PMT0822 PMT0869 PMT0882 PMT0887 PMT0893 PMT0900
PMT0918 PMT0979 PMT0999 PMT1028 PMT1031 PMT1071
PMT1088 PMT1123 PMT1147 PMT1155 PMT1156 PMT1164
PMT1170 PMT1187 PMT1189 PMT1192 PMT1193 PMT1232
PMT1246 PMT1253 PMT1261 PMT1279 PMT1295 PMT1309
PMT1310 PMT1311 PMT1312 PMT1352 PMT1383 PMT1398
PMT1407 PMT1415 PMT1424 PMT1438 PMT1475 PMT1484
PMT1485 PMT1486 PMT1489 PMT1497 PMT1503 PMT1525
PMT1551 PMT1557 PMT1580 PMT1585 PMT1604 PMT1611
PMT1617 PMT1627 PMT1628 PMT1644 PMT1646 PMT1651
PMT1660 PMT1689 PMT1691 PMT1708 PMT1717 PMT1719
PMT1729 PMT1731 PMT1790 PMT1793 PMT1794 PMT1810
PMT1812 PMT1818 PMT1825 PMT1843 PMT1862 PMT1864
PMT1906 PMT1912 PMT1917 PMT1922 PMT1923 PMT1926
PMT1934 PMT1953 PMT1961 PMT1971 PMT1972 PMT1978
PMT1981 PMT1989 PMT1991 PMT1995 PMT1996 PMT2035
PMT2041 PMT2059 PMT2070 PMT2079 PMT2085 PMT2101
PMT2107 PMT2109 PMT2141 PMT2143 PMT2146 PMT2158
PMT2163 PMT2179 PMT2182 PMT2207 PMT2243 PMT2243
PMT2243

cluster 27:
PMT0008 PMT0009 PMT0031 PMT0091 PMT0098 PMT0148
PMT0164 PMT0178 PMT0239 PMT0273 PMT0313 PMT0328
PMT0408 PMT0421 PMT0439 PMT0453 PMT0465 PMT0540
PMT0541 PMT0566 PMT0569 PMT0570 PMT0613 PMT0627
PMT0632 PMT0676 PMT0728 PMT0748 PMT0769 PMT0792
PMT0815 PMT0820 PMT0827 PMT0836 PMT0849 PMT0860
PMT0874 PMT0895 PMT0897 PMT1015 PMT1025 PMT1043
PMT1093 PMT1106 PMT1127 PMT1176 PMT1178 PMT1233
PMT1245 PMT1258 PMT1273 PMT1275 PMT1305 PMT1333
PMT1357 PMT1365 PMT1382 PMT1391 PMT1420 PMT1499
PMT1509 PMT1523 PMT1638 PMT1643 PMT1653 PMT1670
PMT1696 PMT1781 PMT1784 PMT1792 PMT1850 PMT1889
PMT1893 PMT1908 PMT1947 PMT1959 PMT1985 PMT1988
PMT2008 PMT2012 PMT2029 PMT2036 PMT2043 PMT2054
PMT2122 PMT2130 PMT2139 PMT2193 PMT2249 PMT2250
PMT2260 PMT2268 PMT1130 PMT1837 PMT1950 PMT1950
PMT1950

cluster 28:
PMT0080 PMT0168 PMT0509 PMT0687 PMT0840 PMT0916
PMT0989 PMT1045 PMT1046 PMT1316 PMT1427 PMT1571
PMT1572 PMT1665 PMT1683 PMT1710 PMT1769 PMT1770
PMT1800 PMT1954 PMT2138 PMT0757 PMT1220 PMT1317
PMT1663 PMT1767 PMT1767 PMT1767

cluster 3:
PMT0069 PMT0088 PMT0110 PMT0187 PMT0242 PMT0243
PMT0247 PMT0278 PMT0411 PMT0443 PMT0464 PMT0484
PMT0485 PMT0499 PMT0532 PMT0577 PMT0645 PMT0657
PMT0730 PMT0850 PMT0926 PMT0927 PMT1047 PMT1094
PMT1095 PMT1125 PMT1186 PMT1204 PMT1205 PMT1217
PMT1231 PMT1306 PMT1322 PMT1384 PMT1481 PMT1570
PMT1576 PMT1768 PMT1782 PMT1816 PMT1856 PMT1896
PMT1915 PMT2019 PMT2121 PMT2123 PMT2124 PMT2125
PMT2129 PMT2134 PMT0739 PMT1022 PMT1840 PMT1863
PMT1898 PMT1898 PMT1898

cluster 9:
PMT0256 PMT0259 PMT0260 PMT0482 PMT0601 PMT0631
PMT1203 PMT1209 PMT2078 PMT2180 PMT2228 PMT2228
PMT2228

cluster 12:
PMT0007 PMT0012 PMT0052 PMT0063 PMT0071 PMT0076
PMT0097 PMT0126 PMT0130 PMT0159 PMT0179 PMT0195
PMT0228 PMT0238 PMT0263 PMT0280 PMT0286 PMT0296
PMT0299 PMT0300 PMT0301 PMT0310 PMT0320 PMT0331
PMT0344 PMT0353 PMT0356 PMT0371 PMT0376 PMT0383
PMT0392 PMT0404 PMT0428 PMT0436 PMT0462 PMT0467
PMT0470 PMT0479 PMT0514 PMT0522 PMT0538 PMT0547
PMT0553 PMT0560 PMT0604 PMT0637 PMT0639 PMT0640
PMT0695 PMT0712 PMT0722 PMT0788 PMT0789 PMT0794
PMT0802 PMT0834 PMT0835 PMT0851 PMT0877 PMT0878
PMT0886 PMT0888 PMT0909 PMT0941 PMT0946 PMT0985
PMT0997 PMT0998 PMT1011 PMT1012 PMT1018 PMT1019
PMT1034 PMT1036 PMT1037 PMT1041 PMT1042 PMT1056
PMT1066 PMT1076 PMT1090 PMT1114 PMT1122 PMT1143
PMT1159 PMT1160 PMT1169 PMT1190 PMT1207 PMT1208
PMT1242 PMT1267 PMT1297 PMT1299 PMT1300 PMT1304
PMT1315 PMT1320 PMT1393 PMT1400 PMT1447 PMT1455
PMT1458 PMT1465 PMT1493 PMT1494 PMT1510 PMT1538
PMT1562 PMT1586 PMT1601 PMT1620 PMT1677 PMT1694
PMT1704 PMT1713 PMT1713 PMT1757 PMT1766 PMT1776
PMT1776 PMT1813 PMT1821 PMT1832 PMT1846 PMT1868
PMT1882 PMT1895 PMT1913 PMT1924 PMT1937 PMT1939
PMT1941 PMT1963 PMT1970 PMT2026 PMT2027 PMT2039
PMT2053 PMT2062 PMT2064 PMT2072 PMT2094 PMT2095
PMT2103 PMT2105 PMT2110 PMT2131 PMT2136 PMT2144
PMT2155 PMT2165 PMT2220 PMT2224 PMT2227 PMT2233
PMT2238 PMT2238 PMT2238

cluster 20:

PMT0002 PMT0011 PMT0017 PMT0044 PMT0048 PMT0075
PMT0090 PMT0122 PMT0124 PMT0131 PMT0132 PMT0135
PMT0144 PMT0145 PMT0154 PMT0156 PMT0158 PMT0183
PMT0189 PMT0204 PMT0274 PMT0276 PMT0292 PMT0312
PMT0369 PMT0393 PMT0401 PMT0437 PMT0450 PMT0468
PMT0491 PMT0518 PMT0526 PMT0588 PMT0599 PMT0615
PMT0663 PMT0666 PMT0741 PMT0758 PMT0798 PMT0810
PMT0813 PMT0833 PMT0839 PMT0843 PMT0844 PMT0852
PMT0861 PMT0868 PMT0898 PMT0906 PMT0913 PMT0934
PMT0935 PMT0965 PMT0973 PMT1003 PMT1027 PMT1055
PMT1073 PMT1110 PMT1146 PMT1163 PMT1174 PMT1264
PMT1288 PMT1290 PMT1291 PMT1335 PMT1360 PMT1364
PMT1394 PMT1396 PMT1402 PMT1409 PMT1437 PMT1474
PMT1524 PMT1529 PMT1544 PMT1560 PMT1587 PMT1618
PMT1621 PMT1656 PMT1659 PMT1661 PMT1679 PMT1684
PMT1730 PMT1829 PMT1851 PMT1875 PMT1886 PMT1927
PMT1930 PMT1982 PMT1990 PMT1993 PMT1999 PMT2030
PMT2031 PMT2047 PMT2052 PMT2058 PMT2073 PMT2076
PMT2093 PMT2096 PMT2100 PMT2188 PMT2201 PMT2206
PMT2223 PMT2252 PMT2269 PMT0166 PMT0166 PMT0166

cluster 14:
PMT0213 PMT0221 PMT0244 PMT0245 PMT0251 PMT0258
PMT0321 PMT0493 PMT0623 PMT0626 PMT0629 PMT0630
PMT0752 PMT0753 PMT0774 PMT0829 PMT0902 PMT0904
PMT0922 PMT1149 PMT1202 PMT1218 PMT1440 PMT1480
PMT1531 PMT1558 PMT1633 PMT1657 PMT1695 PMT1707
PMT1716 PMT1815 PMT1848 PMT1925 PMT1986 PMT1992
PMT2040 PMT2046 PMT2132 PMT2147 PMT2195 PMT2196
PMT0586 PMT0586 PMT0586

cluster 15:
PMT0907 PMT1469 PMT1479 PMT1573 PMT1733 PMT1737
PMT1742 PMT1743 PMT1744 PMT1744 PMT1744

cluster 8:
PMT0006 PMT0016 PMT0021 PMT0030 PMT0040 PMT0117
PMT0163 PMT0172 PMT0176 PMT0205 PMT0207 PMT0224
PMT0285 PMT0309 PMT0311 PMT0319 PMT0333 PMT0341
PMT0342 PMT0347 PMT0410 PMT0433 PMT0447 PMT0459
PMT0475 PMT0500 PMT0545 PMT0546 PMT0573 PMT0585
PMT0619 PMT0625 PMT0689 PMT0702 PMT0746 PMT0747
PMT0776 PMT0838 PMT0847 PMT0857 PMT0864 PMT0930
PMT0950 PMT0970 PMT1029 PMT1051 PMT1078 PMT1079
PMT1096 PMT1117 PMT1142 PMT1148 PMT1161 PMT1195
PMT1196 PMT1211 PMT1215 PMT1278 PMT1344 PMT1355
PMT1367 PMT1387 PMT1404 PMT1418 PMT1419 PMT1436
PMT1521 PMT1566 PMT1612 PMT1622 PMT1637 PMT1650
PMT1675 PMT1685 PMT1700 PMT1703 PMT1721 PMT1778
PMT1828 PMT1830 PMT1849 PMT1884 PMT1894 PMT1948
PMT1975 PMT2005 PMT2006 PMT2056 PMT2080 PMT2153
PMT2167 PMT2178 PMT2183 PMT2197 PMT2199 PMT2210
PMT2219 PMT2258 PMT2267 PMT2267 PMT2267

cluster 4:

PMT0015 PMT0059 PMT0082 PMT0125 PMT0141 PMT0155
PMT0173 PMT0191 PMT0279 PMT0314 PMT0323 PMT0327
PMT0339 PMT0357 PMT0361 PMT0363 PMT0365 PMT0374
PMT0378 PMT0385 PMT0388 PMT0400 PMT0430 PMT0444
PMT0502 PMT0516 PMT0520 PMT0523 PMT0575 PMT0605
PMT0621 PMT0636 PMT0656 PMT0667 PMT0677 PMT0682
PMT0694 PMT0743 PMT0825 PMT0866 PMT0871 PMT0876
PMT0914 PMT0981 PMT1013 PMT1038 PMT1050 PMT1060
PMT1069 PMT1097 PMT1105 PMT1118 PMT1119 PMT1150
PMT1157 PMT1158 PMT1175 PMT1185 PMT1188 PMT1247
PMT1255 PMT1269 PMT1277 PMT1294 PMT1318 PMT1326
PMT1329 PMT1361 PMT1381 PMT1386 PMT1413 PMT1416
PMT1453 PMT1461 PMT1478 PMT1495 PMT1500 PMT1512
PMT1541 PMT1548 PMT1553 PMT1591 PMT1602 PMT1625
PMT1635 PMT1645 PMT1647 PMT1672 PMT1680 PMT1727
PMT1762 PMT1819 PMT1858 PMT1867 PMT1877 PMT1879
PMT1887 PMT1900 PMT1909 PMT1943 PMT1945 PMT1952
PMT1964 PMT1966 PMT1969 PMT1979 PMT2020 PMT2023
PMT2049 PMT2066 PMT2068 PMT2102 PMT2111 PMT2116
PMT2176 PMT2177 PMT2254 PMT2257 PMT2273 PMT2273
PMT2273

cluster 24:
PMT0036 PMT0038 PMT0039 PMT0262 PMT0307 PMT0315
PMT0329 PMT0382 PMT0387 PMT0423 PMT0441 PMT0564
PMT0649 PMT0683 PMT0806 PMT0845 PMT0937 PMT0986
PMT1008 PMT1054 PMT1075 PMT1102 PMT1151 PMT1248
PMT1274 PMT1296 PMT1342 PMT1343 PMT1446 PMT1483
PMT1536 PMT1542 PMT1564 PMT1642 PMT1664 PMT1667
PMT1817 PMT2010 PMT2173 PMT2175 PMT2217 PMT2242
PMT2246 PMT2272 PMT0488 PMT1838 PMT1838 PMT1838

cluster 19:
PMT0020 PMT0026 PMT0028 PMT0062 PMT0104 PMT0116
PMT0160 PMT0182 PMT0216 PMT0248 PMT0255 PMT0282
PMT0284 PMT0322 PMT0325 PMT0332 PMT0349 PMT0403
PMT0432 PMT0512 PMT0519 PMT0551 PMT0592 PMT0628
PMT0652 PMT0696 PMT0699 PMT0701 PMT0709 PMT0721
PMT0755 PMT0771 PMT0846 PMT0875 PMT0967 PMT1010
PMT1052 PMT1059 PMT1061 PMT1092 PMT1098 PMT1111
PMT1165 PMT1171 PMT1216 PMT1230 PMT1238 PMT1243
PMT1259 PMT1282 PMT1308 PMT1325 PMT1331 PMT1334
PMT1366 PMT1370 PMT1373 PMT1374 PMT1378 PMT1389
PMT1392 PMT1403 PMT1408 PMT1421 PMT1459 PMT1513
PMT1514 PMT1518 PMT1522 PMT1526 PMT1527 PMT1550
PMT1568 PMT1654 PMT1674 PMT1692 PMT1706 PMT1761
PMT1844 PMT1847 PMT1880 PMT1921 PMT1933 PMT1987
PMT2009 PMT2037 PMT2099 PMT2106 PMT2169 PMT2198
PMT2202 PMT2203 PMT2204 PMT2216 PMT2255 PMT2255
PMT2255

cluster 10:
PMT0004 PMT0014 PMT0043 PMT0050 PMT0055 PMT0056
PMT0094 PMT0095 PMT0100 PMT0107 PMT0109 PMT0115
PMT0193 PMT0241 PMT0272 PMT0277 PMT0350 PMT0389

PMT0395 PMT0418 PMT0478 PMT0495 PMT0503 PMT0517
PMT0534 PMT0552 PMT0554 PMT0600 PMT0606 PMT0618
PMT0633 PMT0675 PMT0729 PMT0793 PMT0856 PMT0862
PMT0917 PMT1109 PMT1124 PMT1135 PMT1179 PMT1180
PMT1201 PMT1214 PMT1254 PMT1262 PMT1354 PMT1369
PMT1376 PMT1395 PMT1442 PMT1533 PMT1567 PMT1668
PMT1682 PMT1777 PMT1827 PMT1892 PMT1914 PMT1929
PMT1944 PMT2002 PMT2004 PMT2013 PMT2018 PMT2077
PMT2083 PMT2086 PMT2087 PMT2119 PMT2120 PMT2148
PMT2172 PMT2208 PMT2259 PMT1368 PMT1897 PMT1897
PMT1897

cluster 5:
PMT0003 PMT0066 PMT0083 PMT0105 PMT0106 PMT0140
PMT0146 PMT0177 PMT0232 PMT0253 PMT0424 PMT0446
PMT0496 PMT0536 PMT0567 PMT0616 PMT0706 PMT0725
PMT0726 PMT0785 PMT0855 PMT0859 PMT0977 PMT1004
PMT1063 PMT1083 PMT1120 PMT1166 PMT1173 PMT1197
PMT1198 PMT1199 PMT1200 PMT1410 PMT1449 PMT1504
PMT1507 PMT1515 PMT1520 PMT1750 PMT1789 PMT1835
PMT1866 PMT1866 PMT1866