DOCTORAL DISSERTATION

Prochlorococcus Genetic Transformation and the Genomics of Nitrogen Metabolism

by

Andrew Carl Tolonen

September 2005

DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

20060119 011
Prochlorococcus Genetic Transformation and the Genomics of Nitrogen Metabolism

by

Andrew Carl Tolonen

Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

and

Woods Hole Oceanographic Institution
Woods Hole, Massachusetts 02543

September 2005

DOCTORAL DISSERTATION

Funding was provided by a Merck Graduate Fellowship in Computational Biology and a National Science Foundation Graduate Fellowship. This work was also funded by Department of Energy (DOE) Genomes to Life grant DE-FG02-02ER63445 and a grant from the Gordon and Betty Moore Foundation.


Approved for publication; distribution unlimited.

Approved for Distribution:

John J. Stegeman, Chair
Department of Biology

Paola Malanotte-Rizzoli
MIT Director of Joint Program

John W. Farrington
WHOI Dean of Graduate Studies
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

©2005 Massachusetts Institute of Technology, All Rights Reserved

Signature of Author: 

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 aficionado,
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 Higgsbotham: 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........................................................................................................................................1-20

Chapter 1  Global gene expression of *Prochlorococcus* ecotypes under nitrogen starvation and on different nitrogen sources.................................................................21-46

Chapter 2  Genetic manipulation of *Prochlorococcus* MIT9313: GFP expression on an RSF1010 plasmid and Tn5 transposition..........................................................47-64

Chapter 3  Optimized *in situ* construction of oligomers on an array surface.........................65-69

Future Directions................................................................................................................................70-75

Appendix I  Divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation..........................................................................................................................76-81

Appendix II  Transfer of photosynthesis genes to and from *Prochlorococcus* viruses........82-87

Appendix III  Daf-16 Target genes that control *C. elegans* life-span and metabolism........88-91

Appendix IV  Antibiotic Sensitivities of *Prochlorococcus* MED4 and MIT9313.................92-98

Appendix V  Conjugal transfer of an RSF1010-derived plasmid to *Prochlorococcus* MED4 and MIT9313........................................................................................................99-120

Appendix VI  Supplemental figures for *Prochlorococcus* microarray analysis of gene expression..........................................................................................................................121-148
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 photosynthetic 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.](image_url)

**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 alignments 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 orthologous 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.

*Prochlorococcus* nitrogen metabolism

*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 highly 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).

![Diagram of Prochlorococcus](image)

**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-oxoglutarate 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 *amtI*, *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 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:
**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 relevant genetic system because many strains 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 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 artificial 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.](image)

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 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.](image)

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 specific 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
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 in 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 activating 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 synthesis 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
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


---

**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).


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


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


Olson, RJ. (1985) unpublished data.


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

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

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 gInA 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 utilizes 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 loge 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 (+NH4 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.
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 +NH₄ samples at t=24 hrs., the -N expression summaries at this time point were compared to the +NH₄ at t=12 hrs. instead. As the gene expression correlations between +NH₄ 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. Phit = fraction of total genes scored as putative NtcA targets (0.05)).

\[
\text{probability of } >m \text{ genes up in -N with NtcA binding sites due to chance alone} = \sum_{i=m}^{N} \frac{N!}{(N-i)! \cdot i!} \cdot P_{\text{hit}}^i \cdot (1-P_{\text{hit}})^{N-i}
\]

The log₂-transformed -N/+NH₄ 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 (\(\mu\)) using the following formula (k = number of clusters. \(n\) = number of genes in a cluster. \(x_n\) = position of gene in expression space. \(\mu_i\) = 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.

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

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 strategy 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 ($\mu=0.65$ day$^{-1}$) and 3 days ($\mu=0.23$ 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 $\sim 0.65$ 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 *Synechocystis* 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 $k=2.67 \times 10^{-4}$ 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 rbcS/L were repressed in nitrite medium, while rbcS 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 cyanate 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 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 (hli10, hli21, hli22) and two MIT9313 hli genes (hli5 and hli7) 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
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 hli5, the glutamyl tRNA-synthetase, and hli7 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 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 hli5 and hli7 (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. hli5 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 utilization 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 Prochlorococcus 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 pyrimidine 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. GinB, 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 ginB and its putative upstream partner PMM1462 were both elevated under N-deprivation. In contrast, MIT9313 ginB 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 ginB and its putatively co-expressed partners, we propose that MIT9313 ginB 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 suppressing 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 portrait 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 ginB. 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.


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


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 (○) 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 (○) remained N replete. All data points show means of duplicate cultures; error bars show the range.
Fig. 2. Comparison of MED4 (A) and MIT9313 (B) gene expression in -N and +NH₄ 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 NH₄-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 log₂ unit on nitrite (A) or >1.5 log₂ units on urea (D) relative to ammonia. MED4 plots show all genes differentially expressed >1 log₂ 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 *Synechococcus*. **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 expressed 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. Tolonen¹, Wolfgang R. Hess², Gregory B. Liszt³, and Sallie W. Chisholm⁴

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

ABSTRACT

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 opportunities 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 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 OD₆₀₀ 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 ml s 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 (10⁸ 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 22°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 10⁸ 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 \( \mu \text{g} \text{ml}^{-1} \) kanamycin to isolate colonies. Prochlorococcus colonies were excised using a sterile spatula and transferred back to liquid medium containing 50 \( \mu \text{g} \text{ml}^{-1} \) kanamycin. Once the MIT9313 cultures had reached late log-phase, a 100 \( \mu \text{l} \) aliquot of the culture was spread onto LB plates to titer the remaining \( E. \text{coli} \). Unfortunately, \( 10^2 \) to \( 10^3 \) \( E. \text{coli} \) cells ml\(^{-1} \) often remained viable in the MIT9313 cultures even after isolating MIT9313 colonies on Pro99-agarose plates. To eliminate the remaining \( E. \text{coli} \), the MIT9313 cultures were infected with \( E. \text{coli} \) phage T7 (Demerec and Fano, 1945: Studier, 1969) at a multiplicity of infection (MOI) of \( 10^6 \) phage per \( E. \text{coli} \) host. The \( E. \text{coli} \) 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 \( E. \text{coli} \) with the plasmids isolated from Prochlorococcus in order to compare the structure of pRL153 from MIT9313 to the original plasmid. Following transformation into \( E. \text{coli} \), pRL153 was isolated from kanamycin resistant \( E. \text{coli} \) 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 Nhel in the Tn5 fragment that are outside the kanamycin resistance gene. pTRC-GFPmut3.1 was cloned into into the unique Nhel 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' Nhel sites: forward primer (pTRC): 5'-acgtac-gctagc-ctgaaatgagctgttgacaatt-3' and reverse primer (GFPmut3.1) 5'-cgacc-gctagc-ttattgtatagttcatccatgc-3'. pTRC-GFP PCR product was then Nhel digest, CIP-treated, and ligated with Nhel-digested pRL153. The ligation was electroporated into \( E. \text{coli} \) and the pTRC-GFP insertion was confirmed by DNA sequencing. GFP expression from pRL153-GFP in \( E. \text{coli} \) 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 cassette (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: \( R = 0.044t + 4.82 \) where \( R \) is \( \log_{10}(\text{transconjugant cells}) \) 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 \( (6.9 \times 10^4 \text{ cells}) \) by the number of cells initial transferred into the culture \( (6.5 \times 10^6 \text{ cells}) \).

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 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 lag-1 DNA were transformed with one-fifth of a plasmid DNA prep from an MIT9313 culture of $5 \times 10^8$ 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 $5 \times 10^8$ 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 *in 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


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.
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.
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.
<table>
<thead>
<tr>
<th>Strains</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1100-2</td>
<td>conjugal donor</td>
<td>Bandrin et al, 1983 (obtained from Yale <em>E. coli</em> stock center)</td>
</tr>
<tr>
<td>DH5-alpha</td>
<td>cloning strain used for all</td>
<td>Invitrogen Corp., Carlsbad, CA.</td>
</tr>
<tr>
<td></td>
<td>transformations</td>
<td></td>
</tr>
<tr>
<td>BW19851</td>
<td>host for pRL27</td>
<td>B. Metcalf, Univ. Illinois</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phage</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>E. coli</em> phage T7</td>
<td>phage to kill <em>E. coli</em> in Pro99</td>
<td>D. Endy, MIT</td>
</tr>
<tr>
<td></td>
<td>medium</td>
<td></td>
</tr>
<tr>
<td><em>Prochlorococcus</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIT9313</td>
<td>conjugal recipient</td>
<td>Chisholm lab, MIT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRL153</td>
<td>RSF1010-derivative</td>
<td>P. Wolk, MSU</td>
</tr>
<tr>
<td>pRK24</td>
<td>conjugal plasmid</td>
<td>D. Figurski, Columbia University</td>
</tr>
<tr>
<td>pjRC03</td>
<td>pTRC-GFPmut3.1</td>
<td>A. Van Oudenaarden, MIT</td>
</tr>
<tr>
<td>pRL27</td>
<td>Tn5 plasmid</td>
<td>B. Metcalf, Univ. Illinois</td>
</tr>
</tbody>
</table>

Table 1. Strains and plasmids used in this study
Optimized in situ construction of oligomers on an array surface

Andrew C. Tolonen*, Dinu F. Albeanu†, Julia F. Corbett‡, Heather Handley, Charlotte Henson§ and Pratap Malik¶

Department of Biology, MIT/WHOI Joint Program, Cambridge, MA 02139, USA, †Department of Biology, MIT, Cambridge, MA 02139, USA, ‡Harvard University, Cambridge, MA 02115, USA and §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 bonds to oligonucleotides at the depotted 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...
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
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 cycles 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 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

<table>
<thead>
<tr>
<th>Deposition strategy</th>
<th>Median cycles</th>
<th>Maximum cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 iteration</td>
<td>60</td>
<td>83</td>
</tr>
<tr>
<td>10 iterations</td>
<td>60</td>
<td>85</td>
</tr>
<tr>
<td>20 iterations</td>
<td>59</td>
<td>85</td>
</tr>
<tr>
<td>Pool</td>
<td>54</td>
<td>73</td>
</tr>
</tbody>
</table>

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

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 \textit{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 \textit{Prochlorococcus} \( ntcA \) binding specificities should be defined. The \textit{Prochlorococcus} \( ntcA \) binding specificities could be studied biochemically by \textit{in vitro} selection of oligonucleotides (Jiang et al., 2000) or by DNase footprinting assays. Alternatively, microarrays have recently been adapted to characterize the \textit{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 \textit{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 utilization 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 \textit{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 \textit{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 \textit{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 \textit{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 P11 in vivo.

Another confounding aspect of the Prochlorococcus PII protein is that it is not phosphorylated 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 phosphorylation (Forchhammer and Hedler 1997). Phosphorylation thus is the mechanism by which PII activity is regulated in other cyanobacteria. If there are conditions under which PII is either phosphorylated 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 synthesis 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.
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


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

Acknowledgements We thank the members of the Prochlorococcus annotation team: G. Rocq, S. W. Chisholm, D. Lindell, N. Algren, M. Colman, W. Hess, A. Post, S. Shaw, C. Steggil, C. Ting, M. Sullivan, A. Tolonen, Z. Johnson and E. Zinner. 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.B. B.B. and L.P., and an NSF grant to B.B. F.P. and A.D. were supported by the EC program Margines, 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 R.P. (hpalenik@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 evolutionary path of Prochlorococcus. It numerically dominates the phytoplankton in the low-light-adapted ecotype has the smallest genome (1,657,990 base pairs, 2,221 genes) and that have different minimum, maximum and optimal light intensities for growth. The high-light-adapted ecotype has the largest genome (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 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.
light-adapted ecotypes most abundant in surface waters, and their low-light-adapted counterparts dominating deeper waters (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, 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 oxygenc 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)4, a putative chlorophyll a oxygenase, which could synthesize (divinyl) chlorophyll b from (divinyl) chlorophyll a4, and a lycopene epsilon cyclase involved in the synthesis of alpha carotene4. 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. About 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

Table 1 General features of two Prochlorococcus genomes

<table>
<thead>
<tr>
<th>Genome feature</th>
<th>MED4</th>
<th>MIT9313</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length (bp)</td>
<td>1,657,990</td>
<td>2,410,873</td>
</tr>
<tr>
<td>G+C content (%)</td>
<td>30.3</td>
<td>50.7</td>
</tr>
<tr>
<td>Protein coding (%)</td>
<td>88</td>
<td>62</td>
</tr>
<tr>
<td>Protein coding genes</td>
<td>1,716</td>
<td>2,275</td>
</tr>
<tr>
<td>With assigned function</td>
<td>1,194</td>
<td>1,286</td>
</tr>
<tr>
<td>Conserved hypothetical</td>
<td>502</td>
<td>709</td>
</tr>
<tr>
<td>Hypothetical</td>
<td>80</td>
<td>197</td>
</tr>
<tr>
<td>Genes with orthologue in: Prochlorococcus MED4</td>
<td>–</td>
<td>1,352</td>
</tr>
<tr>
<td>Prochlorococcus MIT9313</td>
<td>1,352</td>
<td>–</td>
</tr>
<tr>
<td>Synechococcus WH8102</td>
<td>1,394</td>
<td>1,710</td>
</tr>
<tr>
<td>Genes without orthologue in: MED4 and WH8102</td>
<td>–</td>
<td>527</td>
</tr>
<tr>
<td>MIT9313 and WH8102</td>
<td>264</td>
<td>–</td>
</tr>
<tr>
<td>Transfer RNA</td>
<td>37</td>
<td>43</td>
</tr>
<tr>
<td>Ribosomal RNA</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Other structural RNAs</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

---

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 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.

letters to nature
mediated by phage\textsuperscript{46}, 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 photosystem II D1 and D2 reaction centre proteins (psbA and psbD respectively)\textsuperscript{15}. 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 c550 (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\textsuperscript{15}. MED4 also possesses a polypeptide 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\textsuperscript{31}, 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 Thermosynchyococcus elongatus (genome size <2.6 megabases)\textsuperscript{32}, 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\textsuperscript{44}. 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 (50.7\%), which is more typical of marine Synechococcus. MED4 lacks genes for several DNA repair pathways including recombinational repair (recA, recC) and damage reversal (mutT). Particularly, the loss of the base excision repair gene mutX, which removes adenosines incorrectly paired with oxidatively damaged guanine residues, may imply an increased rate of G→C to T→A transversions\textsuperscript{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 nitrate 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\textsuperscript{6}. 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\textsuperscript{9} and the north Red Sea\textsuperscript{96}.

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 amr1 and both lack the nitrogense 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, amr1, 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\textsuperscript{37}. 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\textsuperscript{38}, 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. 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, 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, 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-

![Diagram of marine cyanobacterial genomes](image)

**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 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).
trophic. 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-aceotacetate-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 study, 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 previously6. 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 373 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 X6.2 coverage (33,383 end sequences). For Prochlorococcus MIT9313, supplemental sequencing (X0.05 sequence coverage) of a pFosI 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 EcoRI. 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, Critical, Glimmer and Generation, were used in the determination of potential open reading frames and were checked manually. A revised gene/protein net was searched against the KEGG GENES, Plam, 26. Lindell, et al. 1998. and small proteins of conserved function were manually examined and added to the orthologous 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.

Cyanophages infecting the oceanic cyanobacterium Prochlorococcus

Matthew B. Sullivan1, John B. Waterbury2 & Sally W. Chisholm2

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

Prochlorococcus is the numerically dominant phototroph in the tropical and subtropical oceans, accounting for half of the photosynthetic biomass in some areas1,2. 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 (<1012 ml-1) in stratified oligotrophic waters even where total cyanobacterial abundances were high (>105 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 density, 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 limitations3. Similarly, horizontal gene transfer, important in microbial evolution4, can be mediated by phages5 and is probably responsible for many of the differences in the genomes of closely related microbes6. Recent detailed analyses of molecular phylogenies constructed for marine Prochlorococcus and Synechococcus7 (Fig. 1) show that these genera form a single group within the marine picophytoplankton clade (≥96% identity in 16S ribosomal DNA sequences), yet display microdiversity in the form of ten well-defined subgroups8. 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 described3 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 Synechococcus4. 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 DNA5. Similar to those that infect other marine bacteria11 and Synechococcus10-14, our Prochlorococcus cyanophage isolates fell into three morphological families: Myoviridae, Siphoviridae and Podoviridae15.

As would be predicted10-14, 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 situ5. The Myoviridae exhibited much broader host ranges, with 102 cross-infections out of a possible 539. They not only cross-infect 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 Prochlorococcus14,15, 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). 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 onSynechococcus were Myoviridae, a bias that has been reported by others16, 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 abundance11,14,17,18, and can be as high as 107 phage ml-1. One study19 has shown, for example, that along a transect in which total Synechococcus abundance decreased from 107 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 (107 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 collection4.

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 myovirids 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) 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 larger capsids (-85 nm and 82 kb, respectively). 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.

© 2004 by The National Academy of Sciences of the USA
Table 1. Phages used in this study and their photosynthesis-related genes

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

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).

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).
Synonymous gene clustering in phage often suggests that they are expressed, however, the data are not conclusive as to the origin of the phage amino acid sequences, which is consistent with the hypothesis formed monophyletic clades distinct from each other.

That they are functional, phylogenetic analyses of plastocyanin proteins also suggests to those found in the phages, these findings suggest that the unknown possibility of a recent transfer of these genes from as yet transferred from host to phage in a process akin to gene.

Table 3, which is published as supporting information on the clades and are largely consistent with their host range suggests that the phage genes have diverged to an extent that prevents identification of the common bases on estimates of the frequency of nonsynonymous (Ka) and specific based on estimates of the frequency of nonsynonymous (Ka) and specific.

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 (Ks) and synonymous (Ks) nucleotide substitutions between phage- and host-encoded genes revealed that the phage genes have diverged relative to those from their hosts (Ks 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 the majority of nucleotide substitutions did not cause a change in amino acid sequence (Ks/Ka 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 promotor 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 \(\sigma^9\) transcriptional promotor and \(\rho\)-independent transcriptional terminator (Fig. 1.4). 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 euphotic 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. 2.4). 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 parsimony/maximum likelihood/bootstrap support). 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. 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 supports 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
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 risen 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 \textit{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 \textit{hli} genes in our analyses, although not statistically robust, are consistent with this tenant. Each of the single-copy \textit{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 \textit{hli} types and from their clustering to those from the other marine unicellular cyanobacteria (Fig. 3) (27). In contrast, \textit{hli} gene types present in multiple copies per genome are found in only some \textit{Prochlorococcus} genomes. These latter \textit{hli} gene types are those that are found in the \textit{Prochlorococcus} phage, with at least one phage \textit{hli} gene in each of the four clusters of multicopy \textit{Prochlorococcus} \textit{hli} gene types (Fig. 3). We therefore suggest that phages have mediated the horizontal dispersal of these multicopy genes among \textit{Prochlorococcus}.

The presence of numerous \textit{hli} genes in \textit{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 \textit{hli} genes are competitively inferior to wild-type cells (31). Our hypothesized phage-mediated expansion of the \textit{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 DI-, D2-, and ferredoxin-containing strains.

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-FG02-99ER62873). This research was supported by U.S. Department of Energy Grants DE-FG02-99ER62814 and DE-FG02-02ER643445 and National Science Foundation Grant OCE-9820035 (to S.W.C.).
DAF-16 Target Genes That Control C. elegans Life-Span and Metabolism

Siu Sylvia Lee,1 Scott Kennedy,1 Andrew C. Tolonen,2 Gary Ruvkun1*

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 orthologs (FOXO1, FOXO3, and FOXO4) regulate genes encoding superoxide dismutase (sod-3) 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 sites—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(mgD47) 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. Few 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
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 daf-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 ZK934.4 (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 acetocacete. 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.

pdk-1 (C10G11.5), a gene up-regulated in the daf-2 mutant, encodes one of the two pantetheinyl kinases in C. elegans, the rate-limiting enzymes in coenzyme A synthesis. Because...
Alternatively, might act as a feedback regulator of insulin molog, 5 genes have a DAF-l 6 binding site er region. secretion to influence life-span and dauer arrest. nucleoside analogs and glutathione modulating insulin secretion and in transport of the 14 Members of this subclass are implicated in moters of the -5 encoded an adenosine triphosphate- served between orthologous -5 encodes an adenosine triphosphate- is key to fat metabolism, we examined fat storage in -1 RNAi animals, using Nile Red staining (31). RNAi of -1 caused dramatic reduction of fat storage in the intestine of wild-type or -2 mutant animals (Fig. 3). Thus, increased fat storage in -2 mutants might be partly a result of -1 up-regulation. RNAi of -1 also dramatically shortened wild-type and -2 mutant adult life-span (23), suggesting that inactivation of -1 compromises the health of animals.

RNAi inactivation of F43G9.5, C39F7.5, and F52H3.5 did not affect dauer arrest, life-span, 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 -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 -5 in -2 as compared with -2/-16, it is possible that -2 signaling regulates -5 expression in specific tissues or at specific times, and this was not detected under our experimental conditions. -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). -5 might act as a feedback regulator of insulin secretion to influence life-span and dauer arrest. Alternatively, -5 might also affect life-span by regulating glutathione transport and antioxidant defense.

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

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

<table>
<thead>
<tr>
<th>Gene</th>
<th>Homology</th>
<th>DAF-16 site* (C. elegans)</th>
<th>DAF-16 Site* (Drosophila)</th>
<th>DAF-16 site* (C. briggsae)</th>
<th>mRNA in def-2(-)</th>
<th>RNAi inactivation phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>C08B11.8</td>
<td>Similar to yeast glucosyltransferase</td>
<td>48</td>
<td>324</td>
<td>-500†</td>
<td>n.c.</td>
<td>n.c.</td>
</tr>
<tr>
<td>C10C11.5 (pnk-1)</td>
<td>Pantothenate kinase</td>
<td>389</td>
<td>354</td>
<td>~300†</td>
<td>5X</td>
<td>Shortened</td>
</tr>
<tr>
<td>C39F7.5</td>
<td>Cytochrome c heme binding site</td>
<td>375</td>
<td>299</td>
<td>~350†</td>
<td>3X</td>
<td>n.c.</td>
</tr>
<tr>
<td>E04F6.2</td>
<td>Unknown</td>
<td>240</td>
<td>150</td>
<td>~400†</td>
<td>-</td>
<td>n.c.</td>
</tr>
<tr>
<td>F14F4.3 (mrp-5)</td>
<td>ABC transporter</td>
<td>111,920</td>
<td>567</td>
<td>~900†</td>
<td>n.c.</td>
<td>Extended</td>
</tr>
<tr>
<td>F27CB1.8</td>
<td>Amino acid transporter</td>
<td>915</td>
<td>828</td>
<td>~2500†</td>
<td>-</td>
<td>n.c.</td>
</tr>
<tr>
<td>F43G9.5</td>
<td>Subunit of pre-mRNA cleavage factor I</td>
<td>371</td>
<td>609</td>
<td>~350†</td>
<td>0.4X</td>
<td>n.c.</td>
</tr>
<tr>
<td>F52H3.5</td>
<td>Similar to yeast stress-induced protein</td>
<td>763</td>
<td>982,400</td>
<td>~2200†</td>
<td>TX</td>
<td>n.c.</td>
</tr>
<tr>
<td>F54D5.7</td>
<td>Acyl-CoA dehydrogenase</td>
<td>513</td>
<td>825</td>
<td>n.d.</td>
<td>n.c.</td>
<td>n.c.</td>
</tr>
<tr>
<td>K07B1.3</td>
<td>Mitochondrial carrier</td>
<td>895</td>
<td>69</td>
<td>n.d.</td>
<td>n.c.</td>
<td>n.c.</td>
</tr>
<tr>
<td>T20B3.1</td>
<td>Carnitine acetyltransferase</td>
<td>536</td>
<td>96</td>
<td>-</td>
<td>n.c.</td>
<td>n.c.</td>
</tr>
<tr>
<td>T20B3.2</td>
<td>H pylori carcinogen</td>
<td>588</td>
<td>507</td>
<td>~500, ~600†</td>
<td>n.c.</td>
<td>n.c.</td>
</tr>
<tr>
<td>T21C12.2 (hpd-1)</td>
<td>Hydroxyphenylpyruvate dioxygenase</td>
<td>983</td>
<td>175</td>
<td>~1700†</td>
<td>0.5X</td>
<td>Extended</td>
</tr>
<tr>
<td>T23B12.4</td>
<td>Similar to yeast glucose repressible protein MAK10</td>
<td>90</td>
<td>633</td>
<td>~100</td>
<td>TX</td>
<td>n.c.</td>
</tr>
<tr>
<td>Y10G6H.7</td>
<td>Mitochondrial energy transfer protein signature</td>
<td>71</td>
<td>343</td>
<td>~70</td>
<td>n.c.</td>
<td>n.c.</td>
</tr>
<tr>
<td>ZC506.3</td>
<td>Phosphatidylserine synthase I</td>
<td>702</td>
<td>358</td>
<td>~630†</td>
<td>n.c.</td>
<td>n.c.</td>
</tr>
<tr>
<td>ZK593.4 (bpd-2)</td>
<td>Similar to retinoblastoma binding protein 2</td>
<td>27</td>
<td>716</td>
<td>~2700†</td>
<td>0.5X</td>
<td>Extended</td>
</tr>
</tbody>
</table>

Fig. 3. RNAi of -1 reduced lipid storage. Nile Red staining of wild-type or -2(e1370) animals undergoing the indicated RNAi is shown. (A and C) Nile Red staining showing intestinal fat droplets in wild-type or -2(e1370) animals. (B and D) Reduced Nile Red staining in wild-type or -2(e1370) animals undergoing RNAi against -1.
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 informative 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 daf-2(e1370) and 5 kb of the fly promoter, and this yielded a reasonable first stage of the analysis. 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 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**


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

<table>
<thead>
<tr>
<th>Day 4 at 22°C</th>
<th>Control RNAi</th>
<th>daf-2 RNAi</th>
<th>T21C12.2 RNAi</th>
<th>F14F4.3 RNAi</th>
</tr>
</thead>
<tbody>
<tr>
<td>daf-2(e1370) adult</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>daf-2(e1370) dauer</td>
<td>0</td>
<td>100%</td>
<td>100%</td>
<td>90%</td>
</tr>
</tbody>
</table>

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

Charles H. Spruck, Maria P. de Miguel, Adrian P. L. Smith, Aimee Ryan, Paula Stein, Richard M. Schultz, A. Jeaninne Lincoln, Peter J. Donovan, Steven I. Reed

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 upsets 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 II) without intervening as complex of cyclin B and Cdkl. But the need phase (metaphase

**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.
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 genetically 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 sensitivities 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 \(\mu\) (day\(^{-1}\)) where \(\mu\) 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 (\(R = 0.38t + 13.81\) 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 (\(R = 0.15t + 12.06\)). 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\(^{-1}\) (\(\mu=0.51 \text{ day}^{-1}\)) in the absence of kanamycin whereas they grew at 0.26 doublings day\(^{-1}\) (\(\mu=0.18 \text{ day}^{-1}\)) and 0.22 doublings day\(^{-1}\) (\(\mu=0.15 \text{ day}^{-1}\)) 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 \( \mu \text{g ml}^{-1} \) kanamycin relative to no-kanamycin controls (0.33 versus 0.17 doublings day\(^{-1}\)), log phase growth began immediately in both treatments. MIT9313 cultures containing 25 \( \mu \text{g ml}^{-1} \) kanamycin initially declined in fluorescence, but ultimately grew under selection. We fit a linear regression using the data points once growth had begun (\( R = 0.06t + 12.31 \)) which revealed that cells grew at a rate of 0.08 doublings day\(^{-1}\) (\( \mu = 0.06 \text{ day}^{-1} \)) in 25 \( \mu \text{g ml}^{-1} \) 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 \( \mu \text{g ml}^{-1} \) kanamycin treatment. It is not feasible from these experiments to formally conclude that no MIT9313 cells were kanamycin resistant in the 50 \( \mu \text{g ml}^{-1} \) treatment. However, from a practical standpoint we can conclude that no growth was observed for 90 days in 50 \( \mu \text{g ml}^{-1} \) 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 \( \mu \text{g ml}^{-1} \) chloramphenicol did provide a strong selection against MED4 (Fig. 1B). However, we also observed that as little as 20 \( \mu \text{l} \) 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 \( \mu \text{g ml}^{-1} \) chloramphenicol treatment because no growth was observed. We are thus unable to formally rule out that spontaneous chloramphenicol resistance 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 \( \mu \text{g ml}^{-1} \) 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 \( \mu \text{g ml}^{-1} \) kanamycin did provide a strong 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


Fig. 1. MED4 sensitivities to kanamycin and neomycin (A) and to chloramphenicol 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 \( \mu \text{g} \text{ ml}^{-1} \). Neomycin was tested at 25, 50, and 100 \( \mu \text{g} \text{ ml}^{-1} \). B. MED4 sensitivities to chloramphenicol and ethanol. Chloramphenicol was added at the concentration of 0.5 \( \mu \text{g} \text{ ml}^{-1} \). Because chloramphenicol is solvated using ethanol, ethanol only controls were also included to examine its toxicity independently. Ethanol was added at two concentrations: 1/1000 (i.e. 20 \( \mu \text{l} \) added to 20 \( \text{ml} \) 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 resistance 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 initially 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 resistance 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 initially 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 \( \mu \text{M Q m}^{-2} \text{s}^{-1} \) and 10 \( \mu \text{M Q m}^{-2} \text{s}^{-1} \) 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 pill.

**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 capabilities 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 ml 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 (10⁸ 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 22° 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 10⁶ 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, 10² to 10³ *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 Nhel in the Tn5 fragment that are outside the kanamycin resistance gene. pTRC-GFPmut3.1 was cloned into the unique Nhel site to create pRL153-GFP. To this end, pTRC-GFPmut3.1 was PCR amplified from pJRC03 using PFU polymerase using primers with 5' Nhel sites: forward primer (pTRC): 5'-acgtac-gctagc-ctgaaatgagctgttgacaatt-3' and reverse primer (GFPmut3.1) 5'-cgtacc-gctagc-ttattgtatagttcatccatgc-3'. pTRC-GFP PCR product was then Nhel digest, CIP-treated, and ligated with Nhel-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 (μ = 0.58 day⁻¹) (Fig. 2A). The MIT9313 growth rate was 0.35
doublings day⁻¹ (μ = 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 10⁷ cells ml⁻¹ instead of 10⁶ 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 fluorescence 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 pRK2 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 resistance.

**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 cytometry. 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
\[ \log_{10}(\text{transconjugant cells}) \text{ and } t \text{ is days since mating}. \] We calculated the number of transconjugant cells immediately 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 (6.9 \times 10^4 cells) by the number of cells initial transferred into the culture (6.5 \times 10^6 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 \( \mu \)g^{-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 (1 bp = 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 fluorescence 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**


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\(^{-1}\) (\(\mu = 0.58 \text{ day}^{-1}\)). B. Growth rate of MIT9313 cells under conditions used in matings. MIT9313 grew at a rate of 0.35 doublings day\(^{-1}\) (\(\mu = 0.24 \text{ day}^{-1}\)).
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 selection 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 $2 \times 10^8$ cells (final concentration $10^7$ 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 \( \mu \text{g ml}^{-1} \) kanamycin when mated with \textit{E. coli} containing the conjugal plasmid pRK24 and pRL153 (+plasmid, +kan). Control MIT9313 cultures mated with \textit{E. coli} lacking pRK24 (-plasmid, +kan) do not grow under kanamycin selection. Control cultures mated with \textit{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/HindIII 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 log₂-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.
Comparison of MED4 -N replicates

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.
Comparison of MED4 +N replicates

![Comparison of expression profiles from replicate Prochlorococcus MED4 cultures in the +NH₄ 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.](image-url)
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.
Comparison of MIT9313 +N replicates

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 +NH₄ 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 +NH₄ 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.
<table>
<thead>
<tr>
<th>Position:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5-12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.24</td>
<td>-∞</td>
<td>-∞</td>
<td>0.86</td>
<td>0</td>
<td>0.05</td>
<td>0.95</td>
<td>-∞</td>
<td>0.71</td>
</tr>
<tr>
<td>C</td>
<td>0.1</td>
<td>-∞</td>
<td>-∞</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.19</td>
</tr>
<tr>
<td>G</td>
<td>0.1</td>
<td>1</td>
<td>-∞</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
<td>-∞</td>
<td>0.05</td>
</tr>
<tr>
<td>T</td>
<td>0.56</td>
<td>-∞</td>
<td>1</td>
<td>0.05</td>
<td>0</td>
<td>0.95</td>
<td>0</td>
<td>-∞</td>
<td>0.05</td>
</tr>
</tbody>
</table>

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 PMM0274 PMM0280 PMM0284 PMM0286
PMM0308 PMM0333 PMM0349 PMM0415 PMM0420 PMM0563
PMM0478 PMM0515 PMM0525 PMM0561 PMM0562 PMM0563
PMM0585 PMM0590 PMM0688 PMM0688 PMM0628 PMM0633
PMM0645 PMM0652 PMM0663 PMM0680 PMM0696 PMM0701
PMM0716 PMM0749 PMM0834 PMM0839 PMM0840 PMM0877
PMM0898 PMM0913 PMM0925 PMM0955 PMM1021 PMM1026
PMM1095 PMM1099 PMM1116 PMM1187 PMM1209 PMM1216
PMM1218 PMM1237 PMM1243 PMM1255 PMM1266 PMM1290
PMM1308 PMM1343 PMM1356 PMM1574 PMM1598 PMM1599
PMM1716 PMM1244 PMM0894 PMM0820 PMM0820
PMM0820

cluster 21:

PMM0550 PMM0551 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 PMM0460 PMM0442 PMM0444 PMM0449 PMM0456
PMM0458 PMM0460 PMM0479 PMM0490 PMM0492 PMM0538
PMM0653 PMM0668 PMM0693 PMM0705 PMM0727 PMM0734
PMM0746 PMM0776 PMM0926 PMM0822 PMM0831 PMM0888
PMM0890 PMM0891 PMM0911 PMM0917 PMM0919 PMM0927
PMM0938 PMM0939 PMM0948 PMM0968 PMM0986 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
PMM0066 PMM0165 PMM0517 PMM1319 PMM1641 PMM1641
PMM1641

cluster 17:

PMM0032 PMM0035 PMM0055 PMM0063 PMM0086 PMM0088
PMM0091 PMM0116 PMM0145 PMM0150 PMM0159 PMM0172
<table>
<thead>
<tr>
<th>Cluster 2:</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PMM0050</td>
<td>PMM0053</td>
<td>PMM0074</td>
<td>PMM0108</td>
<td>PMM0122</td>
<td>PMM0130</td>
</tr>
<tr>
<td>PMM0174</td>
<td>PMM0192</td>
<td>PMM0222</td>
<td>PMM0240</td>
<td>PMM0249</td>
<td>PMM0303</td>
</tr>
<tr>
<td>PMM0362</td>
<td>PMM0392</td>
<td>PMM0524</td>
<td>PMM0537</td>
<td>PMM0569</td>
<td>PMM0617</td>
</tr>
<tr>
<td>PMM0625</td>
<td>PMM0630</td>
<td>PMM0706</td>
<td>PMM0708</td>
<td>PMM0718</td>
<td>PMM0720</td>
</tr>
<tr>
<td>PMM0721</td>
<td>PMM0728</td>
<td>PMM0730</td>
<td>PMM0798</td>
<td>PMM0865</td>
<td>PMM0874</td>
</tr>
<tr>
<td>PMM0879</td>
<td>PMM0885</td>
<td>PMM0905</td>
<td>PMM0954</td>
<td>PMM0956</td>
<td>PMM0966</td>
</tr>
<tr>
<td>PMM1105</td>
<td>PMM1126</td>
<td>PMM1137</td>
<td>PMM1155</td>
<td>PMM1222</td>
<td>PMM1238</td>
</tr>
<tr>
<td>PMM1268</td>
<td>PMM1368</td>
<td>PMM1347</td>
<td>PMM1348</td>
<td>PMM1360</td>
<td>PMM1370</td>
</tr>
<tr>
<td>PMM1415</td>
<td>PMM1497</td>
<td>PMM1517</td>
<td>PMM1564</td>
<td>PMM1592</td>
<td>PMM1597</td>
</tr>
<tr>
<td>PMM1627</td>
<td>PMM1690</td>
<td>PMM0915</td>
<td>PMM1043</td>
<td>PMM1362</td>
<td>PMM1393</td>
</tr>
<tr>
<td>PMM1715</td>
<td>PMM1715</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster 22:</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PMM0010</td>
<td>PMM0079</td>
<td>PMM0880</td>
<td>PMM0897</td>
<td>PMM0927</td>
<td>PMM0939</td>
</tr>
<tr>
<td>PMM0430</td>
<td>PMM0454</td>
<td>PMM0481</td>
<td>PMM0484</td>
<td>PMM0506</td>
<td>PMM0557</td>
</tr>
<tr>
<td>PMM0574</td>
<td>PMM0576</td>
<td>PMM0578</td>
<td>PMM0588</td>
<td>PMM0601</td>
<td>PMM0602</td>
</tr>
<tr>
<td>PMM0636</td>
<td>PMM0639</td>
<td>PMM0640</td>
<td>PMM0665</td>
<td>PMM0669</td>
<td>PMM0735</td>
</tr>
<tr>
<td>PMM0752</td>
<td>PMM0764</td>
<td>PMM0789</td>
<td>PMM0821</td>
<td>PMM0826</td>
<td>PMM0836</td>
</tr>
<tr>
<td>PMM0837</td>
<td>PMM0899</td>
<td>PMM0900</td>
<td>PMM0933</td>
<td>PMM0949</td>
<td>PMM0952</td>
</tr>
<tr>
<td>PMM0967</td>
<td>PMM1000</td>
<td>PMM1004</td>
<td>PMM1033</td>
<td>PMM1139</td>
<td>PMM1167</td>
</tr>
<tr>
<td>PMM1208</td>
<td>PMM1219</td>
<td>PMM1224</td>
<td>PMM1258</td>
<td>PMM1278</td>
<td>PMM1282</td>
</tr>
<tr>
<td>PMM1674</td>
<td>PMM1675</td>
<td>PMM1711</td>
<td>PMM1714</td>
<td>PMM0437</td>
<td>PMM0575</td>
</tr>
<tr>
<td>PMM0624</td>
<td>PMM0846</td>
<td>PMM1374</td>
<td>PMM1374</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster 1:</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PMM0004</td>
<td>PMM0024</td>
<td>PMM0036</td>
<td>PMM0044</td>
<td>PMM0059</td>
<td>PMM0065</td>
</tr>
<tr>
<td>PMM0083</td>
<td>PMM0136</td>
<td>PMM0137</td>
<td>PMM0153</td>
<td>PMM0193</td>
<td>PMM0195</td>
</tr>
<tr>
<td>PMM0221</td>
<td>PMM0238</td>
<td>PMM0260</td>
<td>PMM0270</td>
<td>PMM0275</td>
<td>PMM0323</td>
</tr>
<tr>
<td>PMM0503</td>
<td>PMM0512</td>
<td>PMM0528</td>
<td>PMM0603</td>
<td>PMM0755</td>
<td>PMM0928</td>
</tr>
<tr>
<td>PMM1008</td>
<td>PMM1029</td>
<td>PMM1070</td>
<td>PMM1125</td>
<td>PMM1179</td>
<td>PMM1192</td>
</tr>
<tr>
<td>PMM1200</td>
<td>PMM1204</td>
<td>PMM1226</td>
<td>PMM1236</td>
<td>PMM1265</td>
<td>PMM1273</td>
</tr>
<tr>
<td>PMM1337</td>
<td>PMM1353</td>
<td>PMM1382</td>
<td>PMM1468</td>
<td>PMM1481</td>
<td>PMM1491</td>
</tr>
<tr>
<td>PMM1493</td>
<td>PMM1501</td>
<td>PMM1529</td>
<td>PMM1561</td>
<td>PMM1590</td>
<td>PMM1614</td>
</tr>
<tr>
<td>PMM1618</td>
<td>PMM1654</td>
<td>PMM0835</td>
<td>PMM0861</td>
<td>PMM1295</td>
<td></td>
</tr>
<tr>
<td>PMM1295</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster 18:</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PMM0005</td>
<td>PMM0037</td>
<td>PMM0054</td>
<td>PMM0058</td>
<td>PMM0066</td>
<td>PMM0068</td>
</tr>
<tr>
<td>PMM0110</td>
<td>PMM0117</td>
<td>PMM0132</td>
<td>PMM0158</td>
<td>PMM0162</td>
<td>PMM0173</td>
</tr>
<tr>
<td>PMM0191</td>
<td>PMM0196</td>
<td>PMM0218</td>
<td>PMM0234</td>
<td>PMM0241</td>
<td>PMM0241</td>
</tr>
<tr>
<td>PMM0341</td>
<td>PMM0355</td>
<td>PMM0403</td>
<td>PMM0431</td>
<td>PMM0450</td>
<td>PMM0451</td>
</tr>
</tbody>
</table>
cluster 30:
PMM0049  PMM0078  PMM0107  PMM0129  PMM0155  PMM0156
PMM0271  PMM0358  PMM0382  PMM0390  PMM0408  PMM0440
PMM0498  PMM0516  PMM0711  PMM0724  PMM0729  PMM0841
PMM0849  PMM0904  PMM0978  PMM0994  PMM1627  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

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  PMM0763
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  PMM1278  PMM1280  PMM1318
PMM1358  PMM1361  PMM1367  PMM1372  PMM1383  PMM1395
PMM1416  PMM1421  PMM1423  PMM1464  PMM1474  PMM1496
PMM1513  PMM1526  PMM1566  PMM1583  PMM1593  PMM1606
PMM1649  PMM1692  PMM1769  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

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  PMM0407  PMM0455
PMM0466  PMM0473  PMM0487  PMM0587  PMM0591  PMM0597
PMM0598  PMM0658  PMM0674  PMM0694  PMM0737  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  PMM0080  PMM0081  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  PMM1807  PMM1363  PMM1363  PMM1363

cluster 27:
PMM0013  PMM0043  PMM0084  PMM0106  PMM0124  PMM0128
PMM0149  PMM0154  PMM0210  PMM0226  PMM0232  PMM0259
<table>
<thead>
<tr>
<th>Cluster 28:</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PMM0018</td>
<td>PMM0041</td>
<td>PMM0135</td>
<td>PMM0200</td>
<td>PMM0214</td>
<td>PMM0295</td>
</tr>
<tr>
<td>PMM0049</td>
<td>PMM0494</td>
<td>PMM0511</td>
<td>PMM0535</td>
<td>PMM0542</td>
<td>PMM0673</td>
</tr>
<tr>
<td>PMM0074</td>
<td>PMM0747</td>
<td>PMM0797</td>
<td>PMM0829</td>
<td>PMM1071</td>
<td>PMM1101</td>
</tr>
<tr>
<td>PMM0116</td>
<td>PMM1251</td>
<td>PMM1297</td>
<td>PMM1298</td>
<td>PMM1325</td>
<td>PMM1335</td>
</tr>
<tr>
<td>PMM0133</td>
<td>PMM1522</td>
<td>PMM1539</td>
<td>PMM1572</td>
<td>PMM1585</td>
<td>PMM1643</td>
</tr>
<tr>
<td>PMM0163</td>
<td>PMM1644</td>
<td>PMM0783</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster 3:</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PMM0033</td>
<td>PMM0075</td>
<td>PMM0090</td>
<td>PMM0120</td>
<td>PMM0169</td>
<td>PMM0266</td>
</tr>
<tr>
<td>PMM0032</td>
<td>PMM0340</td>
<td>PMM0360</td>
<td>PMM0380</td>
<td>PMM0387</td>
<td>PMM0491</td>
</tr>
<tr>
<td>PMM0059</td>
<td>PMM0544</td>
<td>PMM0626</td>
<td>PMM0656</td>
<td>PMM0660</td>
<td>PMM0666</td>
</tr>
<tr>
<td>PMM0071</td>
<td>PMM0731</td>
<td>PMM0733</td>
<td>PMM0737</td>
<td>PMM0757</td>
<td>PMM0770</td>
</tr>
<tr>
<td>PMM0096</td>
<td>PMM0918</td>
<td>PMM0945</td>
<td>PMM0957</td>
<td>PMM0964</td>
<td>PMM0969</td>
</tr>
<tr>
<td>PMM0097</td>
<td>PMM1016</td>
<td>PMM1024</td>
<td>PMM1037</td>
<td>PMM1038</td>
<td>PMM1050</td>
</tr>
<tr>
<td>PMM0106</td>
<td>PMM1073</td>
<td>PMM1078</td>
<td>PMM1197</td>
<td>PMM1461</td>
<td>PMM1490</td>
</tr>
<tr>
<td>PMM0158</td>
<td>PMM1595</td>
<td>PMM1596</td>
<td>PMM1615</td>
<td>PMM1677</td>
<td>PMM1687</td>
</tr>
<tr>
<td>PMM0169</td>
<td>PMM0432</td>
<td>PMM0738</td>
<td>PMM1036</td>
<td>PMM1036</td>
<td>PMM1036</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster 9:</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PMM0019</td>
<td>PMM0022</td>
<td>PMM0028</td>
<td>PMM0029</td>
<td>PMM0040</td>
<td>PMM0089</td>
</tr>
<tr>
<td>PMM0098</td>
<td>PMM0100</td>
<td>PMM0112</td>
<td>PMM0125</td>
<td>PMM0138</td>
<td>PMM0175</td>
</tr>
<tr>
<td>PMM0177</td>
<td>PMM0178</td>
<td>PMM0180</td>
<td>PMM0184</td>
<td>PMM0229</td>
<td>PMM0242</td>
</tr>
<tr>
<td>PMM0255</td>
<td>PMM0256</td>
<td>PMM0257</td>
<td>PMM0309</td>
<td>PMM0322</td>
<td>PMM0352</td>
</tr>
<tr>
<td>PMM0384</td>
<td>PMM0391</td>
<td>PMM0397</td>
<td>PMM0398</td>
<td>PMM0421</td>
<td>PMM0426</td>
</tr>
<tr>
<td>PMM0488</td>
<td>PMM0518</td>
<td>PMM0522</td>
<td>PMM0596</td>
<td>PMM0613</td>
<td>PMM0659</td>
</tr>
<tr>
<td>PMM0675</td>
<td>PMM0682</td>
<td>PMM0692</td>
<td>PMM0712</td>
<td>PMM0723</td>
<td>PMM0750</td>
</tr>
<tr>
<td>PMM0771</td>
<td>PMM0787</td>
<td>PMM0793</td>
<td>PMM0794</td>
<td>PMM0805</td>
<td>PMM0833</td>
</tr>
<tr>
<td>PMM0848</td>
<td>PMM0852</td>
<td>PMM0859</td>
<td>PMM0880</td>
<td>PMM0881</td>
<td>PMM1002</td>
</tr>
<tr>
<td>PMM1009</td>
<td>PMM1035</td>
<td>PMM1044</td>
<td>PMM1082</td>
<td>PMM1086</td>
<td>PMM1138</td>
</tr>
<tr>
<td>PMM1189</td>
<td>PMM1199</td>
<td>PMM1203</td>
<td>PMM1215</td>
<td>PMM1220</td>
<td>PMM1223</td>
</tr>
<tr>
<td>PMM1225</td>
<td>PMM1232</td>
<td>PMM1274</td>
<td>PMM1292</td>
<td>PMM1304</td>
<td>PMM1320</td>
</tr>
<tr>
<td>PMM1411</td>
<td>PMM1432</td>
<td>PMM1445</td>
<td>PMM1458</td>
<td>PMM1472</td>
<td>PMM1488</td>
</tr>
<tr>
<td>PMM1515</td>
<td>PMM1570</td>
<td>PMM1586</td>
<td>PMM1613</td>
<td>PMM1621</td>
<td>PMM1660</td>
</tr>
<tr>
<td>PMM1676</td>
<td>PMM1683</td>
<td>PMM1683</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster 12:</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PMM0016</td>
<td>PMM0166</td>
<td>PMM0251</td>
<td>PMM0265</td>
<td>PMM0315</td>
<td>PMM0500</td>
</tr>
<tr>
<td>PMM0062</td>
<td>PMM0642</td>
<td>PMM0700</td>
<td>PMM0784</td>
<td>PMM0844</td>
<td>PMM0876</td>
</tr>
<tr>
<td>PMM0096</td>
<td>PMM0993</td>
<td>PMM1055</td>
<td>PMM1109</td>
<td>PMM1249</td>
<td>PMM1247</td>
</tr>
<tr>
<td>PMM1523</td>
<td>PMM1524</td>
<td>PMM1607</td>
<td>PMM1650</td>
<td>PMM1704</td>
<td>PMM0314</td>
</tr>
<tr>
<td>Cluster 15:</td>
<td>PMM025</td>
<td>PMM027</td>
<td>PMM0285</td>
<td>PMM0105</td>
<td>PMM0205</td>
</tr>
<tr>
<td>------------</td>
<td>--------</td>
<td>--------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td></td>
<td>PMM0297</td>
<td>PMM0324</td>
<td>PMM0367</td>
<td>PMM0423</td>
<td>PMM0436</td>
</tr>
<tr>
<td></td>
<td>PMM0495</td>
<td>PMM0496</td>
<td>PMM0536</td>
<td>PMM0543</td>
<td>PMM0547</td>
</tr>
<tr>
<td></td>
<td>PMM0615</td>
<td>PMM0722</td>
<td>PMM0762</td>
<td>PMM0866</td>
<td>PMM0987</td>
</tr>
<tr>
<td></td>
<td>PMM1059</td>
<td>PMM1060</td>
<td>PMM1067</td>
<td>PMM1679</td>
<td>PMM1152</td>
</tr>
<tr>
<td></td>
<td>PMM1205</td>
<td>PMM1259</td>
<td>PMM1323</td>
<td>PMM1326</td>
<td>PMM1377</td>
</tr>
<tr>
<td></td>
<td>PMM1485</td>
<td>PMM1487</td>
<td>PMM1494</td>
<td>PMM1500</td>
<td>PMM1530</td>
</tr>
<tr>
<td></td>
<td>PMM1535</td>
<td>PMM1617</td>
<td>PMM1619</td>
<td>PMM1636</td>
<td>PMM1639</td>
</tr>
<tr>
<td></td>
<td>PMM0751</td>
<td>PMM1058</td>
<td>PMM0812</td>
<td>PMM0812</td>
<td>PMM0812</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster 20:</th>
<th>PMM0056</th>
<th>PMM0096</th>
<th>PMM0189</th>
<th>PMM0213</th>
<th>PMM0225</th>
<th>PMM0227</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PMM0239</td>
<td>PMM0267</td>
<td>PMM0273</td>
<td>PMM0283</td>
<td>PMM0285</td>
<td>PMM0288</td>
</tr>
<tr>
<td></td>
<td>PMM0306</td>
<td>PMM0328</td>
<td>PMM0356</td>
<td>PMM0383</td>
<td>PMM0428</td>
<td>PMM0505</td>
</tr>
<tr>
<td></td>
<td>PMM0514</td>
<td>PMM0527</td>
<td>PMM0529</td>
<td>PMM0553</td>
<td>PMM0582</td>
<td>PMM0584</td>
</tr>
<tr>
<td></td>
<td>PMM0643</td>
<td>PMM0648</td>
<td>PMM0670</td>
<td>PMM0672</td>
<td>PMM0679</td>
<td>PMM0768</td>
</tr>
<tr>
<td></td>
<td>PMM0782</td>
<td>PMM0799</td>
<td>PMM0830</td>
<td>PMM0842</td>
<td>PMM0903</td>
<td>PMM1084</td>
</tr>
<tr>
<td></td>
<td>PMM1142</td>
<td>PMM1228</td>
<td>PMM1263</td>
<td>PMM1284</td>
<td>PMM1316</td>
<td>PMM1334</td>
</tr>
<tr>
<td></td>
<td>PMM1346</td>
<td>PMM1369</td>
<td>PMM1379</td>
<td>PMM1449</td>
<td>PMM1471</td>
<td>PMM1486</td>
</tr>
<tr>
<td></td>
<td>PMM1507</td>
<td>PMM1591</td>
<td>PMM1659</td>
<td>PMM0686</td>
<td>PMM1003</td>
<td>PMM1482</td>
</tr>
<tr>
<td></td>
<td>PMM1143</td>
<td>PMM1143</td>
<td>PMM1143</td>
<td>PMM1143</td>
<td>PMM1143</td>
<td>PMM1143</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster 14:</th>
<th>PMM0246</th>
<th>PMM0336</th>
<th>PMM0370</th>
<th>PMM0689</th>
<th>PMM0920</th>
<th>PMM0958</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PMM0970</td>
<td>PMM1041</td>
<td>PMM1462</td>
<td>PMM0374</td>
<td>PMM0374</td>
<td>PMM0374</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster 8:</th>
<th>PMM0023</th>
<th>PMM0046</th>
<th>PMM0228</th>
<th>PMM0329</th>
<th>PMM0452</th>
<th>PMM0453</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PMM0469</td>
<td>PMM0549</td>
<td>PMM0599</td>
<td>PMM0605</td>
<td>PMM0710</td>
<td>PMM0766</td>
</tr>
<tr>
<td></td>
<td>PMM0767</td>
<td>PMM0781</td>
<td>PMM0785</td>
<td>PMM0901</td>
<td>PMM1350</td>
<td>PMM1436</td>
</tr>
<tr>
<td></td>
<td>PMM1438</td>
<td>PMM1508</td>
<td>PMM1519</td>
<td>PMM1520</td>
<td>PMM1629</td>
<td>PMM1655</td>
</tr>
<tr>
<td></td>
<td>PMM1662</td>
<td>PMM0272</td>
<td>PMM0397</td>
<td>PMM0468</td>
<td>PMM0540</td>
<td>PMM1661</td>
</tr>
<tr>
<td></td>
<td>PMM0691</td>
<td>PMM1578</td>
<td>PMM1578</td>
<td>PMM1578</td>
<td>PMM1578</td>
<td>PMM1578</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster 4:</th>
<th>PMM0021</th>
<th>PMM0042</th>
<th>PMM0069</th>
<th>PMM0071</th>
<th>PMM0082</th>
<th>PMM0113</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PMM0118</td>
<td>PMM0176</td>
<td>PMM0183</td>
<td>PMM0190</td>
<td>PMM0217</td>
<td>PMM0248</td>
</tr>
<tr>
<td></td>
<td>PMM0302</td>
<td>PMM0310</td>
<td>PMM0332</td>
<td>PMM0404</td>
<td>PMM0413</td>
<td>PMM0425</td>
</tr>
<tr>
<td></td>
<td>PMM0427</td>
<td>PMM0433</td>
<td>PMM0439</td>
<td>PMM0489</td>
<td>PMM0497</td>
<td>PMM0568</td>
</tr>
<tr>
<td></td>
<td>PMM0592</td>
<td>PMM0604</td>
<td>PMM0607</td>
<td>PMM0610</td>
<td>PMM0612</td>
<td>PMM0616</td>
</tr>
<tr>
<td></td>
<td>PMM0654</td>
<td>PMM0671</td>
<td>PMM0756</td>
<td>PMM0893</td>
<td>PMM0868</td>
<td>PMM1046</td>
</tr>
<tr>
<td></td>
<td>PMM1048</td>
<td>PMM1053</td>
<td>PMM1064</td>
<td>PMM1127</td>
<td>PMM1140</td>
<td>PMM1161</td>
</tr>
<tr>
<td></td>
<td>PMM1207</td>
<td>PMM1221</td>
<td>PMM1248</td>
<td>PMM1302</td>
<td>PMM1306</td>
<td>PMM1357</td>
</tr>
<tr>
<td></td>
<td>PMM1371</td>
<td>PMM1403</td>
<td>PMM1418</td>
<td>PMM1460</td>
<td>PMM1467</td>
<td>PMM1521</td>
</tr>
<tr>
<td></td>
<td>PMM1525</td>
<td>PMM1560</td>
<td>PMM1580</td>
<td>PMM1628</td>
<td>PMM1681</td>
<td>PMM1693</td>
</tr>
<tr>
<td></td>
<td>PMM1695</td>
<td>PMM1696</td>
<td>PMM1701</td>
<td>PMM1631</td>
<td>PMM1632</td>
<td>PMM1632</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster 24:</th>
<th>PMM0082</th>
<th>PMM0083</th>
<th>PMM0089</th>
<th>PMM0011</th>
<th>PMM0067</th>
<th>PMM0099</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PMM0111</td>
<td>PMM0114</td>
<td>PMM0147</td>
<td>PMM0157</td>
<td>PMM0209</td>
<td>PMM0211</td>
</tr>
<tr>
<td></td>
<td>PMM0250</td>
<td>PMM0291</td>
<td>PMM0396</td>
<td>PMM0412</td>
<td>PMM0434</td>
<td>PMM0513</td>
</tr>
</tbody>
</table>
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 PMM0818 PMM0819 PMM1074
PMM1134 PMM1391 PMM1463 PMM1623 PMM0359 PMM0690
PMM1390 PMM1390 PMM1390

cluster 10:
PMM0311 PMM0325 PMM0348 PMM0416 PMM0461 PMM0475
PMM0483 PMM0507 PMM0508 PMM0530 PMM0573 PMM0806
PMM0649 PMM0699 PMM0753 PMM0760 PMM0788 PMM0801
PMM0882 PMM0863 PMM0867 PMM0894 PMM0912 PMM0926
PMM0930 PMM1652 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:**

<table>
<thead>
<tr>
<th>Cluster Gene 1</th>
<th>Cluster Gene 2</th>
<th>Cluster Gene 3</th>
<th>Cluster Gene 4</th>
<th>Cluster Gene 5</th>
<th>Cluster Gene 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMT0010</td>
<td>PMT0013</td>
<td>PMT0046</td>
<td>PMT0049</td>
<td>PMT0085</td>
<td>PMT0087</td>
</tr>
<tr>
<td>PMT0093</td>
<td>PMT0103</td>
<td>PMT0112</td>
<td>PMT0171</td>
<td>PMT0210</td>
<td>PMT0217</td>
</tr>
<tr>
<td>PMT0335</td>
<td>PMT0390</td>
<td>PMT0397</td>
<td>PMT0402</td>
<td>PMT0407</td>
<td>PMT0426</td>
</tr>
<tr>
<td>PMT0435</td>
<td>PMT0455</td>
<td>PMT0460</td>
<td>PMT0469</td>
<td>PMT0471</td>
<td>PMT0474</td>
</tr>
<tr>
<td>PMT0492</td>
<td>PMT0494</td>
<td>PMT0565</td>
<td>PMT0531</td>
<td>PMT0543</td>
<td>PMT0555</td>
</tr>
<tr>
<td>PMT0557</td>
<td>PMT0558</td>
<td>PMT0578</td>
<td>PMT0579</td>
<td>PMT0659</td>
<td>PMT0704</td>
</tr>
<tr>
<td>PMT0705</td>
<td>PMT0707</td>
<td>PMT0714</td>
<td>PMT0718</td>
<td>PMT0736</td>
<td>PMT0766</td>
</tr>
<tr>
<td>PMT0777</td>
<td>PMT0795</td>
<td>PMT0799</td>
<td>PMT0809</td>
<td>PMT0809</td>
<td>PMT0863</td>
</tr>
<tr>
<td>PMT0872</td>
<td>PMT0974</td>
<td>PMT1032</td>
<td>PMT1033</td>
<td>PMT1062</td>
<td>PMT1077</td>
</tr>
<tr>
<td>PMT1108</td>
<td>PMT1177</td>
<td>PMT1191</td>
<td>PMT1266</td>
<td>PMT1225</td>
<td>PMT1229</td>
</tr>
<tr>
<td>PMT1239</td>
<td>PMT1256</td>
<td>PMT1268</td>
<td>PMT1292</td>
<td>PMT1301</td>
<td>PMT1302</td>
</tr>
<tr>
<td>PMT1330</td>
<td>PMT1338</td>
<td>PMT1349</td>
<td>PMT1353</td>
<td>PMT1399</td>
<td>PMT1435</td>
</tr>
<tr>
<td>PMT1456</td>
<td>PMT1476</td>
<td>PMT1477</td>
<td>PMT1491</td>
<td>PMT1501</td>
<td>PMT1511</td>
</tr>
<tr>
<td>PMT1547</td>
<td>PMT1594</td>
<td>PMT1605</td>
<td>PMT1666</td>
<td>PMT1614</td>
<td>PMT1624</td>
</tr>
<tr>
<td>PMT1648</td>
<td>PMT1649</td>
<td>PMT1725</td>
<td>PMT1774</td>
<td>PMT1785</td>
<td>PMT1805</td>
</tr>
<tr>
<td>PMT1809</td>
<td>PMT1865</td>
<td>PMT1881</td>
<td>PMT1936</td>
<td>PMT1968</td>
<td>PMT1973</td>
</tr>
<tr>
<td>PMT1984</td>
<td>PMT1997</td>
<td>PMT1998</td>
<td>PMT2021</td>
<td>PMT2024</td>
<td>PMT2032</td>
</tr>
<tr>
<td>PMT2063</td>
<td>PMT2081</td>
<td>PMT2191</td>
<td>PMT2192</td>
<td>PMT2225</td>
<td>PMT2253</td>
</tr>
<tr>
<td>PMT2264</td>
<td>PMT2264</td>
<td>PMT2264</td>
<td>PMT2264</td>
<td>PMT2264</td>
<td>PMT2264</td>
</tr>
</tbody>
</table>

**cluster 21:**

<table>
<thead>
<tr>
<th>Cluster Gene 1</th>
<th>Cluster Gene 2</th>
<th>Cluster Gene 3</th>
<th>Cluster Gene 4</th>
<th>Cluster Gene 5</th>
<th>Cluster Gene 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMT0023</td>
<td>PMT0045</td>
<td>PMT0065</td>
<td>PMT0121</td>
<td>PMT0139</td>
<td>PMT0208</td>
</tr>
<tr>
<td>PMT0220</td>
<td>PMT0250</td>
<td>PMT0252</td>
<td>PMT0257</td>
<td>PMT0266</td>
<td>PMT0269</td>
</tr>
<tr>
<td>PMT0275</td>
<td>PMT0287</td>
<td>PMT0416</td>
<td>PMT0417</td>
<td>PMT0461</td>
<td>PMT0472</td>
</tr>
<tr>
<td>PMT0582</td>
<td>PMT0602</td>
<td>PMT0617</td>
<td>PMT0650</td>
<td>PMT0654</td>
<td>PMT0674</td>
</tr>
<tr>
<td>PMT0680</td>
<td>PMT0686</td>
<td>PMT0745</td>
<td>PMT0756</td>
<td>PMT0762</td>
<td>PMT0780</td>
</tr>
<tr>
<td>PMT0781</td>
<td>PMT0808</td>
<td>PMT0812</td>
<td>PMT0821</td>
<td>PMT0848</td>
<td>PMT0884</td>
</tr>
<tr>
<td>PMT0896</td>
<td>PMT0966</td>
<td>PMT0978</td>
<td>PMT1026</td>
<td>PMT1049</td>
<td>PMT1065</td>
</tr>
<tr>
<td>PMT1087</td>
<td>PMT1181</td>
<td>PMT1251</td>
<td>PMT1252</td>
<td>PMT1287</td>
<td>PMT1351</td>
</tr>
<tr>
<td>PMT1358</td>
<td>PMT1411</td>
<td>PMT1431</td>
<td>PMT1441</td>
<td>PMT1555</td>
<td>PMT1575</td>
</tr>
<tr>
<td>PMT1596</td>
<td>PMT1603</td>
<td>PMT1623</td>
<td>PMT1629</td>
<td>PMT1662</td>
<td>PMT1669</td>
</tr>
<tr>
<td>PMT1701</td>
<td>PMT1711</td>
<td>PMT1728</td>
<td>PMT1786</td>
<td>PMT1807</td>
<td>PMT1811</td>
</tr>
<tr>
<td>PMT1857</td>
<td>PMT1885</td>
<td>PMT1888</td>
<td>PMT1903</td>
<td>PMT1904</td>
<td>PMT1905</td>
</tr>
<tr>
<td>PMT2017</td>
<td>PMT2083</td>
<td>PMT2082</td>
<td>PMT2151</td>
<td>PMT2157</td>
<td>PMT2181</td>
</tr>
<tr>
<td>PMT2261</td>
<td>PMT2261</td>
<td>PMT2261</td>
<td>PMT2261</td>
<td>PMT2261</td>
<td>PMT2261</td>
</tr>
</tbody>
</table>

**cluster 7:**

<table>
<thead>
<tr>
<th>Cluster Gene 1</th>
<th>Cluster Gene 2</th>
<th>Cluster Gene 3</th>
<th>Cluster Gene 4</th>
<th>Cluster Gene 5</th>
<th>Cluster Gene 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMT0024</td>
<td>PMT0029</td>
<td>PMT0037</td>
<td>PMT0047</td>
<td>PMT0060</td>
<td>PMT0073</td>
</tr>
<tr>
<td>PMT0113</td>
<td>PMT0114</td>
<td>PMT0123</td>
<td>PMT0143</td>
<td>PMT0165</td>
<td>PMT0175</td>
</tr>
<tr>
<td>PMT0185</td>
<td>PMT0196</td>
<td>PMT0200</td>
<td>PMT0206</td>
<td>PMT0209</td>
<td>PMT0212</td>
</tr>
<tr>
<td>PMT0222</td>
<td>PMT0225</td>
<td>PMT0226</td>
<td>PMT0229</td>
<td>PMT0234</td>
<td>PMT0264</td>
</tr>
<tr>
<td>PMT0267</td>
<td>PMT0281</td>
<td>PMT0290</td>
<td>PMT0291</td>
<td>PMT0317</td>
<td>PMT0318</td>
</tr>
<tr>
<td>PMT0326</td>
<td>PMT0330</td>
<td>PMT0338</td>
<td>PMT0343</td>
<td>PMT0351</td>
<td>PMT0354</td>
</tr>
<tr>
<td>PMT0358</td>
<td>PMT0362</td>
<td>PMT0364</td>
<td>PMT0373</td>
<td>PMT0379</td>
<td>PMT0405</td>
</tr>
<tr>
<td>PMT0412</td>
<td>PMT0413</td>
<td>PMT0414</td>
<td>PMT0415</td>
<td>PMT0425</td>
<td>PMT0429</td>
</tr>
<tr>
<td>PMT0438</td>
<td>PMT0442</td>
<td>PMT0481</td>
<td>PMT0504</td>
<td>PMT0524</td>
<td>PMT0525</td>
</tr>
<tr>
<td>PMT0529</td>
<td>PMT0530</td>
<td>PMT0533</td>
<td>PMT0561</td>
<td>PMT0580</td>
<td>PMT0587</td>
</tr>
<tr>
<td>PMT0590</td>
<td>PMT0663</td>
<td>PMT0668</td>
<td>PMT0611</td>
<td>PMT0624</td>
<td>PMT0641</td>
</tr>
<tr>
<td>PMT0643</td>
<td>PMT0644</td>
<td>PMT0651</td>
<td>PMT0658</td>
<td>PMT0660</td>
<td>PMT0669</td>
</tr>
<tr>
<td>PMT0671</td>
<td>PMT0684</td>
<td>PMT0690</td>
<td>PMT0703</td>
<td>PMT0711</td>
<td>PMT0716</td>
</tr>
</tbody>
</table>
PMT0720 PMT0737 PMT0759 PMT0765 PMT0767 PMT0772
PMT0778 PMT0782 PMT0784 PMT0786 PMT0787 PMT0816
PMT0819 PMT0841 PMT0842 PMT0865 PMT0903 PMT0921
PMT0931 PMT0933 PMT0958 PMT0972 PMT0975 PMT0984
PMT0985 PMT1002 PMT1030 PMT1070 PMT1085 PMT1112
PMT1128 PMT1132 PMT1133 PMT1134 PMT1138 PMT1140
PMT1162 PMT1184 PMT1236 PMT1240 PMT1249 PMT1266
PMT1268 PMT1280 PMT1303 PMT1319 PMT1327 PMT1338
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
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
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
PMT0878 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
cluster 2:
PMT0099 PMT0102 PMT0142 PMT0265 PMT0269 PMT0422
PMT0427 PMT0431 PMT0454 PMT0498 PMT0528 PMT0537
PMT0539 PMT0556 PMT0571 PMT0899 PMT0928 PMT0964
PMT0679 PMT0791 PMT0817 PMT0899 PMT0964 PMT0964
PMT1181 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
<table>
<thead>
<tr>
<th>PMT2003</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>cluster 22:</strong></td>
</tr>
<tr>
<td>PMT0127</td>
</tr>
<tr>
<td>PMT0665</td>
</tr>
<tr>
<td>PMT0884</td>
</tr>
<tr>
<td>PMT0954</td>
</tr>
<tr>
<td>PMT0993</td>
</tr>
<tr>
<td>PMT1286</td>
</tr>
<tr>
<td>PMT1636</td>
</tr>
<tr>
<td>PMT2256</td>
</tr>
<tr>
<td><strong>cluster 1:</strong></td>
</tr>
<tr>
<td>PMT0001</td>
</tr>
<tr>
<td>PMT0078</td>
</tr>
<tr>
<td>PMT0118</td>
</tr>
<tr>
<td>PMT0386</td>
</tr>
<tr>
<td>PMT0506</td>
</tr>
<tr>
<td>PMT0562</td>
</tr>
<tr>
<td>PMT0612</td>
</tr>
<tr>
<td>PMT0713</td>
</tr>
<tr>
<td>PMT0773</td>
</tr>
<tr>
<td>PMT0948</td>
</tr>
<tr>
<td>PMT1074</td>
</tr>
<tr>
<td>PMT1234</td>
</tr>
<tr>
<td>PMT1263</td>
</tr>
<tr>
<td>PMT1426</td>
</tr>
<tr>
<td>PMT1690</td>
</tr>
<tr>
<td>PMT1795</td>
</tr>
<tr>
<td>PMT1911</td>
</tr>
<tr>
<td>PMT2087</td>
</tr>
<tr>
<td>PMT2135</td>
</tr>
<tr>
<td>PMT2189</td>
</tr>
<tr>
<td>PMT2262</td>
</tr>
<tr>
<td><strong>cluster 18:</strong></td>
</tr>
<tr>
<td>PMT0346</td>
</tr>
<tr>
<td>PMT0963</td>
</tr>
<tr>
<td>PMT1314</td>
</tr>
<tr>
<td>PMT2174</td>
</tr>
<tr>
<td><strong>cluster 30:</strong></td>
</tr>
<tr>
<td>PMT0169</td>
</tr>
<tr>
<td>PMT1468</td>
</tr>
<tr>
<td>PMT1697</td>
</tr>
<tr>
<td>PMT1739</td>
</tr>
<tr>
<td>PMT1749</td>
</tr>
<tr>
<td>PMT2088</td>
</tr>
<tr>
<td><strong>cluster 13:</strong></td>
</tr>
<tr>
<td>PMT0027</td>
</tr>
<tr>
<td>PMT0119</td>
</tr>
<tr>
<td>PMT0194</td>
</tr>
<tr>
<td>PMT0235</td>
</tr>
<tr>
<td>PMT0334</td>
</tr>
</tbody>
</table>
PMT0449 PMT0451 PMT0458 PMT0489 PMT0596 PMT0622
PMT0638 PMT0692 PMT0719 PMT0890 PMT0901 PMT0919
PMT0920 PMT0945 PMT0947 PMT0952 PMT1005 PMT1023
PMT1145 PMT1183 PMT1194 PMT1221 PMT1226 PMT1281
PMT1284 PMT1298 PMT1321 PMT1328 PMT1342 PMT1405
PMT1406 PMT1460 PMT1462 PMT1488 PMT1516 PMT1549
PMT1561 PMT1563 PMT1578 PMT1583 PMT1592 PMT1600
PMT1699 PMT1798 PMT1799 PMT1860 PMT1880 PMT1890
PMT1892 PMT1901 PMT1905 PMT1960 PMT1968 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
PMT0366 PMT0369 PMT0374 PMT0378 PMT0389 PMT0399
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 PMT1593 PMT1595 PMT1598
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 PMT0748 PMT0754 PMT0832 PMT0915
PMT0924 PMT1035 PMT1213 PMT1219 PMT1232 PMT1345
PMT1450 PMT1451 PMT1454 PMT1505 PMT1506 PMT1740
PMT1751 PMT1752 PMT1754 PMT1756 PMT1760 PMT1779
PMT1780 PMT1783 PMT1836 PMT1854 PMT1859 PMT1899
<table>
<thead>
<tr>
<th>PMT1955</th>
<th>PMT1957</th>
<th>PMT2089</th>
<th>PMT2091</th>
<th>PMT1307</th>
<th>PMT1590</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMT1666</td>
<td>PMT1666</td>
<td>PMT1666</td>
<td>PMT1666</td>
<td>PMT1666</td>
<td>PMT1666</td>
</tr>
</tbody>
</table>

**cluster 6:**

<table>
<thead>
<tr>
<th>PMT0065</th>
<th>PMT0072</th>
<th>PMT0057</th>
<th>PMT0068</th>
<th>PMT0073</th>
<th>PMT0149</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMT0161</td>
<td>PMT0197</td>
<td>PMT0179</td>
<td>PMT0199</td>
<td>PMT0196</td>
<td>PMT0214</td>
</tr>
<tr>
<td>PMT0668</td>
<td>PMT0670</td>
<td>PMT0670</td>
<td>PMT0670</td>
<td>PMT0670</td>
<td>PMT0670</td>
</tr>
<tr>
<td>PMT0765</td>
<td>PMT0765</td>
<td>PMT0765</td>
<td>PMT0765</td>
<td>PMT0765</td>
<td>PMT0765</td>
</tr>
<tr>
<td>PMT0824</td>
<td>PMT0824</td>
<td>PMT0824</td>
<td>PMT0824</td>
<td>PMT0824</td>
<td>PMT0824</td>
</tr>
<tr>
<td>PMT0946</td>
<td>PMT0946</td>
<td>PMT0946</td>
<td>PMT0946</td>
<td>PMT0946</td>
<td>PMT0946</td>
</tr>
<tr>
<td>PMT1086</td>
<td>PMT1086</td>
<td>PMT1086</td>
<td>PMT1086</td>
<td>PMT1086</td>
<td>PMT1086</td>
</tr>
<tr>
<td>PMT1170</td>
<td>PMT1170</td>
<td>PMT1170</td>
<td>PMT1170</td>
<td>PMT1170</td>
<td>PMT1170</td>
</tr>
<tr>
<td>PMT1246</td>
<td>PMT1246</td>
<td>PMT1246</td>
<td>PMT1246</td>
<td>PMT1246</td>
<td>PMT1246</td>
</tr>
<tr>
<td>PMT1310</td>
<td>PMT1310</td>
<td>PMT1310</td>
<td>PMT1310</td>
<td>PMT1310</td>
<td>PMT1310</td>
</tr>
<tr>
<td>PMT1407</td>
<td>PMT1407</td>
<td>PMT1407</td>
<td>PMT1407</td>
<td>PMT1407</td>
<td>PMT1407</td>
</tr>
<tr>
<td>PMT1457</td>
<td>PMT1457</td>
<td>PMT1457</td>
<td>PMT1457</td>
<td>PMT1457</td>
<td>PMT1457</td>
</tr>
<tr>
<td>PMT1551</td>
<td>PMT1551</td>
<td>PMT1551</td>
<td>PMT1551</td>
<td>PMT1551</td>
<td>PMT1551</td>
</tr>
<tr>
<td>PMT1617</td>
<td>PMT1617</td>
<td>PMT1617</td>
<td>PMT1617</td>
<td>PMT1617</td>
<td>PMT1617</td>
</tr>
<tr>
<td>PMT1660</td>
<td>PMT1660</td>
<td>PMT1660</td>
<td>PMT1660</td>
<td>PMT1660</td>
<td>PMT1660</td>
</tr>
<tr>
<td>PMT1729</td>
<td>PMT1729</td>
<td>PMT1729</td>
<td>PMT1729</td>
<td>PMT1729</td>
<td>PMT1729</td>
</tr>
<tr>
<td>PMT1812</td>
<td>PMT1812</td>
<td>PMT1812</td>
<td>PMT1812</td>
<td>PMT1812</td>
<td>PMT1812</td>
</tr>
<tr>
<td>PMT1906</td>
<td>PMT1906</td>
<td>PMT1906</td>
<td>PMT1906</td>
<td>PMT1906</td>
<td>PMT1906</td>
</tr>
<tr>
<td>PMT1934</td>
<td>PMT1934</td>
<td>PMT1934</td>
<td>PMT1934</td>
<td>PMT1934</td>
<td>PMT1934</td>
</tr>
<tr>
<td>PMT2041</td>
<td>PMT2041</td>
<td>PMT2041</td>
<td>PMT2041</td>
<td>PMT2041</td>
<td>PMT2041</td>
</tr>
<tr>
<td>PMT2107</td>
<td>PMT2107</td>
<td>PMT2107</td>
<td>PMT2107</td>
<td>PMT2107</td>
<td>PMT2107</td>
</tr>
<tr>
<td>PMT2163</td>
<td>PMT2163</td>
<td>PMT2163</td>
<td>PMT2163</td>
<td>PMT2163</td>
<td>PMT2163</td>
</tr>
</tbody>
</table>

**cluster 27:**

<table>
<thead>
<tr>
<th>PMT0008</th>
<th>PMT0009</th>
<th>PMT0031</th>
<th>PMT0091</th>
<th>PMT0098</th>
<th>PMT0148</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMT0164</td>
<td>PMT0178</td>
<td>PMT0239</td>
<td>PMT0273</td>
<td>PMT0313</td>
<td>PMT0328</td>
</tr>
<tr>
<td>PMT0408</td>
<td>PMT0421</td>
<td>PMT0439</td>
<td>PMT0463</td>
<td>PMT0540</td>
<td>PMT0540</td>
</tr>
<tr>
<td>PMT0541</td>
<td>PMT0566</td>
<td>PMT0569</td>
<td>PMT0570</td>
<td>PMT0613</td>
<td>PMT0627</td>
</tr>
<tr>
<td>PMT0632</td>
<td>PMT0676</td>
<td>PMT0728</td>
<td>PMT0748</td>
<td>PMT0769</td>
<td>PMT0792</td>
</tr>
<tr>
<td>PMT0815</td>
<td>PMT0820</td>
<td>PMT0827</td>
<td>PMT0836</td>
<td>PMT0849</td>
<td>PMT0860</td>
</tr>
<tr>
<td>PMT0874</td>
<td>PMT0895</td>
<td>PMT0897</td>
<td>PMT1015</td>
<td>PMT1025</td>
<td>PMT1043</td>
</tr>
<tr>
<td>PMT1093</td>
<td>PMT1106</td>
<td>PMT1127</td>
<td>PMT1176</td>
<td>PMT1178</td>
<td>PMT1233</td>
</tr>
<tr>
<td>PMT1245</td>
<td>PMT1258</td>
<td>PMT1273</td>
<td>PMT1275</td>
<td>PMT1305</td>
<td>PMT1333</td>
</tr>
<tr>
<td>PMT1357</td>
<td>PMT1365</td>
<td>PMT1382</td>
<td>PMT1391</td>
<td>PMT1420</td>
<td>PMT1499</td>
</tr>
<tr>
<td>PMT1509</td>
<td>PMT1523</td>
<td>PMT1638</td>
<td>PMT1643</td>
<td>PMT1653</td>
<td>PMT1670</td>
</tr>
<tr>
<td>PMT1696</td>
<td>PMT1781</td>
<td>PMT1784</td>
<td>PMT1792</td>
<td>PMT1850</td>
<td>PMT1889</td>
</tr>
<tr>
<td>PMT1893</td>
<td>PMT1908</td>
<td>PMT1947</td>
<td>PMT1959</td>
<td>PMT1985</td>
<td>PMT1988</td>
</tr>
<tr>
<td>PMT2008</td>
<td>PMT2012</td>
<td>PMT2029</td>
<td>PMT2036</td>
<td>PMT2043</td>
<td>PMT2054</td>
</tr>
<tr>
<td>PMT2122</td>
<td>PMT2136</td>
<td>PMT2139</td>
<td>PMT2193</td>
<td>PMT2249</td>
<td>PMT2250</td>
</tr>
<tr>
<td>PMT2260</td>
<td>PMT2268</td>
<td>PMT1130</td>
<td>PMT1837</td>
<td>PMT1950</td>
<td>PMT1950</td>
</tr>
</tbody>
</table>
cluster 28:
PMT0080 PMT0168 PMT0509 PMT0687 PMT0840 PMT0916
PMT0989 PMT1045 PMT1046 PMT1316 PMT1427 PMT1571
PMT1572 PMT1665 PMT1683 PMT1710 PMT1769 PMT1770
PMT1808 PMT1954 PMT2138 PMT2057 PMT1220 PMT1317
PMT1663 PMT1767 PMT1767 PMT1767

cluster 3:
PMT0069 PMT0088 PMT0110 PMT0187 PMT0242 PMT0243
PMT0247 PMT0278 PMT0411 PMT0443 PMT0464 PMT0484
PMT0730 PMT0850 PMT0926 PMT0927 PMT1047 PMT1094
PMT1095 PMT1125 PMT1186 PMT1204 PMT1205 PMT1217
PMT1231 PMT1306 PMT1322 PMT1384 PMT1481 PMT1570
PMT1576 PMT1665 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 PMT0381 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 PMT1288
PMT1242 PMT1267 PMT1297 PMT1299 PMT1300 PMT1304
PMT1315 PMT1320 PMT1393 PMT1400 PMT1447 PMT1455
PMT1458 PMT1465 PMT1493 PMT1494 PMT1510 PMT1538
PMT1562 PMT1586 PMT1601 PMT1620 PMT1677 PMT1694
PMT1794 PMT1713 PMT1713 PMT1757 PMT1766 PMT1776
PMT1776 PMT1813 PMT1821 PMT1832 PMT1846 PMT1868
PMT1882 PMT1895 PMT1913 PMT1924 PMT1927 PMT1937
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 PMT0144
PMT0145 PMT0154 PMT0156 PMT0158 PMT0159 PMT0183
PMT0191 PMT0204 PMT0227 PMT0276 PMT0292 PMT0312
PMT0369 PMT0401 PMT0437 PMT0450 PMT0468 PMT0489
PMT0518 PMT0526 PMT0588 PMT0599 PMT0615 PMT0663
PMT0666 PMT0741 PMT0758 PMT0798 PMT0810 PMT0813
PMT0852 PMT0861 PMT0868 PMT0889 PMT0906 PMT0913
PMT0935 PMT0965 PMT0973 PMT1003 PMT1027 PMT1055
PMT1073 PMT1110 PMT1146 PMT1163 PMT1174 PMT1264
PMT1288 PMT1290 PMT1291 PMT1325 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
PMT2031 PMT2047 PMT2052 PMT2058 PMT2073 PMT2076
PMT2083 PMT2089 PMT2096 PMT2100 PMT2147 PMT2149
PMT2223 PMT2252 PMT2269 PMT0166 PMT0166 PMT0166
PMT0213 PMT0221 PMT0244 PMT0245 PMT0251 PMT0258
PMT0321 PMT0493 PMT0623 PMT0626 PMT0629 PMT0630
PMT0752 PMT0775 PMT0777 PMT0829 PMT0902 PMT0904
PMT0922 PMT1149 PMT1202 PMT1218 PMT1440 PMT1448
PMT1531 PMT1558 PMT1633 PMT1657 PMT1695 PMT1707
PMT1716 PMT1815 PMT1848 PMT1925 PMT1986 PMT1992
PMT2040 PMT2046 PMT2132 PMT2147 PMT2195 PMT2196
PMT0586 PMT0586 PMT0586
PMT0907 PMT1469 PMT1479 PMT1573 PMT1733 PMT1737
PMT1742 PMT1743 PMT1744 PMT1744 PMT1744
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 14:

cluster 15:

cluster 8:

cluster 4:
<table>
<thead>
<tr>
<th>PMT0395</th>
<th>PMT0418</th>
<th>PMT0478</th>
<th>PMT0495</th>
<th>PMT0503</th>
<th>PMT0517</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMT0534</td>
<td>PMT0552</td>
<td>PMT0554</td>
<td>PMT0600</td>
<td>PMT0606</td>
<td>PMT0618</td>
</tr>
<tr>
<td>PMT0633</td>
<td>PMT0675</td>
<td>PMT0729</td>
<td>PMT0793</td>
<td>PMT0856</td>
<td>PMT0862</td>
</tr>
<tr>
<td>PMT0917</td>
<td>PMT1109</td>
<td>PMT1124</td>
<td>PMT1135</td>
<td>PMT1179</td>
<td>PMT1180</td>
</tr>
<tr>
<td>PMT1201</td>
<td>PMT1214</td>
<td>PMT1254</td>
<td>PMT1262</td>
<td>PMT1354</td>
<td>PMT1369</td>
</tr>
<tr>
<td>PMT1376</td>
<td>PMT1395</td>
<td>PMT1442</td>
<td>PMT1533</td>
<td>PMT1567</td>
<td>PMT1668</td>
</tr>
<tr>
<td>PMT1682</td>
<td>PMT1777</td>
<td>PMT1827</td>
<td>PMT1892</td>
<td>PMT1914</td>
<td>PMT1929</td>
</tr>
<tr>
<td>PMT1944</td>
<td>PMT2002</td>
<td>PMT2004</td>
<td>PMT2013</td>
<td>PMT2018</td>
<td>PMT2077</td>
</tr>
<tr>
<td>PMT2083</td>
<td>PMT2086</td>
<td>PMT2087</td>
<td>PMT2119</td>
<td>PMT2120</td>
<td>PMT2148</td>
</tr>
<tr>
<td>PMT2172</td>
<td>PMT2208</td>
<td>PMT2259</td>
<td>PMT1368</td>
<td>PMT1897</td>
<td>PMT1897</td>
</tr>
</tbody>
</table>

**cluster 5:**

<table>
<thead>
<tr>
<th>PMT0003</th>
<th>PMT0066</th>
<th>PMT0083</th>
<th>PMT0105</th>
<th>PMT0106</th>
<th>PMT0140</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMT0146</td>
<td>PMT0177</td>
<td>PMT0232</td>
<td>PMT0253</td>
<td>PMT0424</td>
<td>PMT0446</td>
</tr>
<tr>
<td>PMT0496</td>
<td>PMT0536</td>
<td>PMT0567</td>
<td>PMT0616</td>
<td>PMT0706</td>
<td>PMT0725</td>
</tr>
<tr>
<td>PMT0726</td>
<td>PMT0785</td>
<td>PMT0855</td>
<td>PMT0859</td>
<td>PMT0897</td>
<td>PMT1094</td>
</tr>
<tr>
<td>PMT1063</td>
<td>PMT1083</td>
<td>PMT1120</td>
<td>PMT1166</td>
<td>PMT1173</td>
<td>PMT1197</td>
</tr>
<tr>
<td>PMT1198</td>
<td>PMT1199</td>
<td>PMT1200</td>
<td>PMT1410</td>
<td>PMT1449</td>
<td>PMT1504</td>
</tr>
<tr>
<td>PMT1507</td>
<td>PMT1515</td>
<td>PMT1520</td>
<td>PMT1750</td>
<td>PMT1789</td>
<td>PMT1835</td>
</tr>
<tr>
<td>PMT1866</td>
<td>PMT1866</td>
<td>PMT1866</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Title and Subtitle**

*Prochlorococcus* Genetic Transformation and the Genomics of Nitrogen Metabolism

**Author(s)**
Andrew Carl Tolonen

**Performing Organization Name and Address**

MIT/WHOI Joint Program in Oceanography/Applied Ocean Science & Engineering

**Sponsoring Organization Name and Address**

Merck & Co.
Department of Energy Genomes to Life
National Science Foundation
Gordon and Betty Moore Foundation

**Abstract (Limit: 200 words)**

*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.