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

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

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

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

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

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

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

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

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

MED4 has 364 genes without an orthologue in MIT9313, whereas MIT9313 has 923 that are not present in MED4. These strain-specific genes, which are dispersed throughout the chromosome (Fig. 2), clearly hold clues about the relative fitness of the two strains under different environmental conditions. Almost half of the 923 MIT9313-specific genes are in fact present in *Synechococcus* WH8102, suggesting that they have been lost from MED4 in the course of genome reduction. Lateral transfer events, perhaps mediated by phage[10], may also be a source of some of the strain-specific genes (Supplementary Figs 3–6).

Gene loss has played a major role in defining the *Prochlorococcus* photosynthetic apparatus. MED4 and MIT9313 are missing many of the genes encoding phycobilisome structural proteins and enzymes involved in phycobilin biosynthesis[15]. Although some of

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

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

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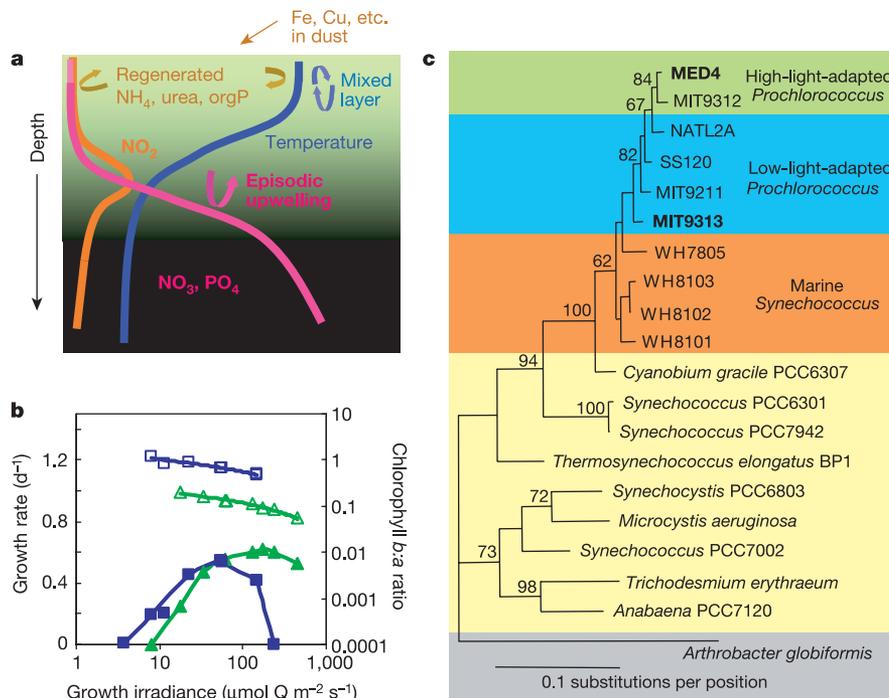
Figure 1 Ecology, physiology and phylogeny of *Prochlorococcus* ecotypes. **a**, Schematic stratified open-ocean water column illustrating vertical gradients allowing niche differentiation. Shading represents degree of light penetration. Temperature and salinity gradients provide a mixing barrier, isolating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

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

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

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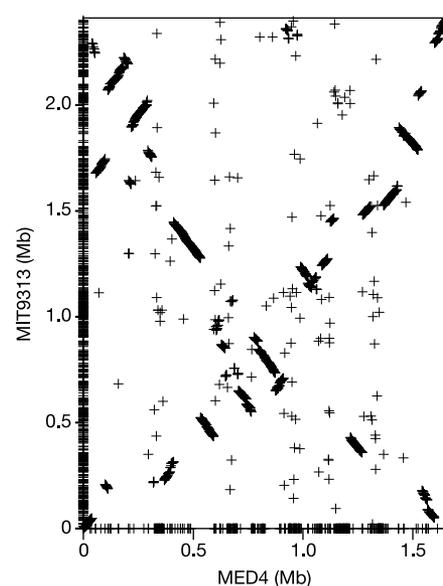

Figure 2 Global genome alignment as seen from start positions of orthologous genes. Genes present in one genome but not the other are shown on the axes. The 'broken X' pattern has been noted before for closely related bacterial genomes, and is probably due to multiple inversions centred around the origin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

Thus, loss of some circadian components may imply an alternative signalling pathway for circadian control.

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

Analysis of the nitrogen acquisition capabilities of the two strains points to a sequential decay in the capacity to use nitrate and nitrite during the evolution of the *Prochlorococcus* lineage (Fig. 3a). In *Synechococcus* WH8102—representing the presumed ancestral state—many nitrogen acquisition and assimilation genes are grouped together (Fig. 3a). MIT9313 has lost a 25-gene cluster, which includes genes encoding the nitrate/nitrite transporter and

nitrate reductase. The nitrite reductase gene has been retained in MIT9313, but it is flanked by a proteobacterial-like nitrite transporter rather than a typical cyanobacterial nitrate/nitrite permease (Supplementary Fig. 4), suggesting acquisition by lateral gene transfer. An additional deletion event occurred in MED4, in which the nitrite reductase gene was also lost (Fig. 3a). As a result of these serial deletion events MIT9313 cannot use nitrate, and MED4 cannot use nitrate or nitrite[9]. Thus each *Prochlorococcus* ecotype uses the N species that is most prevalent at the light levels to which they are best adapted: ammonium in the surface waters and nitrite at depth (Fig. 1a). *Synechococcus*, which is the only one of the three that has nitrate reductase, is able to bloom when nitrate is upwelled (Fig. 1a), as occurs in the spring in the North Atlantic[3] and the north Red Sea[26].

The two *Prochlorococcus* strains are also less versatile in their organic N usage capabilities than *Synechococcus* WH8102 (ref. 13). MED4 contains the genes necessary for usage of urea, cyanate and oligopeptides, but no monomeric amino acid transporters have been identified. In contrast, MIT9313 contains transporters for urea, amino acids and oligopeptides but lacks the genes necessary
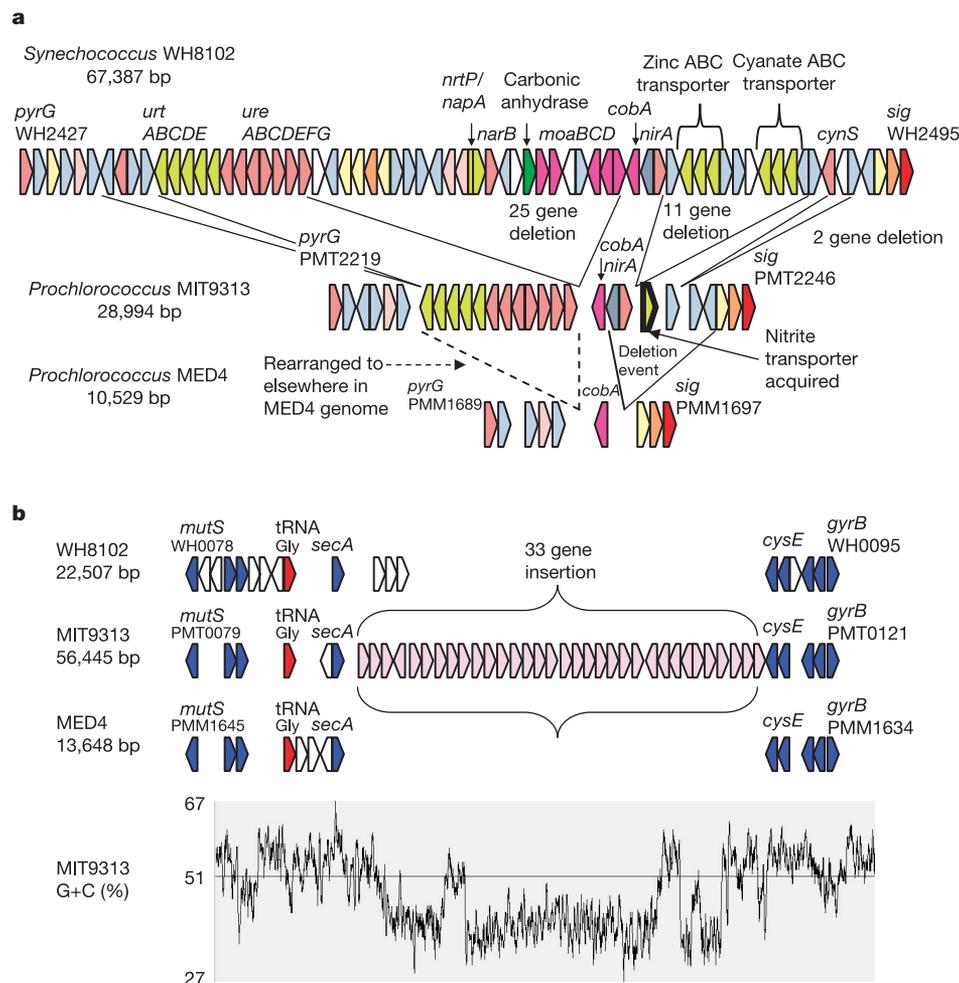

Figure 3 Dynamic architecture of marine cyanobacterial genomes. **a**, Deletion, acquisition and rearrangement of nitrogen usage genes. In MIT9313, 25 genes including the nitrate/nitrite transporter (*nrtP/napA*), nitrate reductase (*narB*) and carbonic anhydrase have been deleted. The cyanate transporter and cyanate lyase (*cynS*) were probably lost after the divergence of MIT9313 from the rest of the *Prochlorococcus* lineage, as MED4 possesses these genes. MIT9313 has retained nitrite reductase (*nirA*) and acquired a nitrite transporter. In MED4 *nirA* has been lost and the urea transporter (*urt*

cluster) and urease (*ure* cluster) genes have been rearranged (dotted line). Genes in different functional categories are colour-coded to guide the eye. **b**, Lateral transfer of genes involved in lipopolysaccharide biosynthesis including sugar transferases, sugar epimerases, modifying enzymes and two pairs of ABC-type transporters. Blue, genes in all three genomes; pink, genes hypothesized to have been laterally transferred; red, tRNAs; white, other genes. The percentage of G + C content in MIT9313 along this segment is lower (42%) than the whole-genome average (horizontal line).

3

for cyanate usage (cyanate transporter and cyanate lyase) (Fig. 3a). As expected, both genomes contain the high-affinity ammonium transporter *amt1* and both lack the nitrogenase genes essential for nitrogen fixation. Finally, both contain the nitrogen transcriptional regulator encoded by *ntcA* and there are numerous genes in both genomes, including *ntcA*, *amt1*, the urea transport and GS/GOGAT genes (glutamine synthetase and glutamate synthase, both involved in ammonia assimilation), with an upstream NtcA-binding-site consensus sequence.

The genomes also have differences in genes involved in phosphorus usage that have obvious ecological implications. MED4, but not MIT9313, is capable of growth on organic P sources (L. R. Moore and S.W.C., unpublished data), and organic P can be the prevalent form of P in high-light surface waters[27]. This difference may be due to the acquisition of an alkaline phosphatase-like gene in MED4 (Supplementary Fig. 5). Both genomes contain the high-affinity phosphate transport system encoded by *pstS* and *pstABC*[28], but MIT9313 contains an additional copy of the phosphate-binding component *pstS*, perhaps reflecting an increased reliance on orthophosphate in deeper waters. MED4 contains several P-related regulatory genes including the *phoB*, *phoR* two-component system and the transcriptional activator *ptrA*. In MIT9313, however, *phoR* is interrupted by two frameshifts and *ptrA* is further degenerated, suggesting that this strain has lost the ability to regulate gene expression in response to changing P levels.

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

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

Prochlorococcus has been identified in deep suboxic zones where it is unlikely that they can sustain themselves by photosynthesis alone[29], thus we looked for genomic evidence of heterotrophic capability. Indeed, the presence of oligopeptide transporters in both genomes, and the larger proportion of transporters (including some sugar transporters) in the MIT9313 strain-specific genes (Supplementary Fig. 2), suggests the potential for partial heterotrophy. However, neither genome contains known pathways that would allow for complete heterotrophy. They are both missing genes for steps in the tricarboxylic acid cycle, including 2-oxoglutarate dehydrogenase, succinyl-CoA synthetase and succinyl-CoA-acetoacetate-CoA transferase.

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

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

Methods

Genome sequencing and assembly

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

Genome annotation

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

Genome comparisons

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

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

1. Chisholm, S. W. *et al.* A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* **334,** 340–343 (1988).

2. Campbell, L., Liu, H., Nolla, H. & Vaulot, D. Annual variability of phytoplankton and bacteria in the subtropical North Pacific Ocean at Station ALOHA during the 1991–1994 ENSO event. *Deep Sea Res. I* **44,** 167–192 (1997).

3. DuRand, M. D., Olson, R. J. & Chisholm, S. W. Phytoplankton population dynamics at the Bermuda Atlantic Time-series Station in the Sargasso Sea. *Deep Sea Res. II* **48,** 1983–2003 (2001).

4. Rocap, G., Distel, D. L., Waterbury, J. B. & Chisholm, S. W. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S–23S rDNA internal transcribed spacer (ITS) sequences. *Appl. Environ. Microbiol.* **68,** 1180–1191 (2002).

5. Moore, L. R. & Chisholm, S. W. Photophysiology of the marine cyanobacterium *Prochlorococcus*: Ecotypic differences among cultured isolates. *Limnol. Oceanogr.* **44,** 628–638 (1999).

6. Moore, L. R., Rocap, G. & Chisholm, S. W. Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393,** 464–467 (1998).

7. Moore, L. R., Goericke, R. E. & Chisholm, S. W. Comparative physiology of *Synechococcus* and *Prochlorococcus*: influence of light and temperature on growth, pigments, fluorescence and absorptive properties. *Mar. Ecol. Prog. Ser.* **116,** 259–275 (1995).

8. Mann, E. L., Ahlgren, N., Moffett, J. W. & Chisholm, S. W. Copper toxicity and cyanobacteria ecology in the Sargasso Sea. *Limnol. Oceanogr.* **47,** 976–988 (2002).

9. Moore, L. R., Post, A. F., Rocap, G. & Chisholm, S. W. Utilization of different nitrogen sources by the marine cyanobacteria, *Prochlorococcus* and *Synechococcus*. *Limnol. Oceanogr.* **47,** 989–996 (2002).

10. Sullivan, M. B., Waterbury, J. B. & Chisholm, S. W. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* (in the press).

11. Hagström, Å. *et al.* Use of 16S ribosomal DNA for delineation of marine bacterioplankton species. *Appl. Environ. Microbiol.* **68,** 3628–3633 (2002).

12. West, N. J. *et al.* Closely related *Prochlorococcus* genotypes show remarkably different depth distributions in two oceanic regions as revealed by *in situ* hybridization using 16S rRNA-targeted oligonucleotides. *Microbiology* **147,** 1731–1744 (2001).

13. Palenik, B. *et al.* The genome of a motile marine *Synechococcus*. *Nature* (this issue).

14. La Roche, J. *et al.* Independent evolution of the prochlorophyte and green plant chlorophyll a/b light harvesting proteins. *Proc. Natl Acad. Sci. USA* **93,** 15244–15248 (1996).

15. Hess, W. *et al.* The photosynthetic apparatus of *Prochlorococcus*: Insights through comparative genomics. *Photosynth. Res.* **70,** 53–71 (2001).

16. Stickforth, P., Steiger, S., Hess, W. R. & Sandmann, G. A novel type of lycopene ε-cyclase in the marine cyanobacterium *Prochlorococcus marinus* MED4. *Arch. Microbiol.* **179,** 407–415 (2003).

17. Frankenberg, N., Mukougawa, K., Kohchi, T. & Lagarias, J. C. Functional genomic analysis of the HY2 family of ferredoxin-dependent bilin reductases from oxygenic photosynthetic organisms. *Plant Cell* **13,** 965–978 (2001).

18. Ting, C., Rocap, G., King, J. & Chisholm, S. W. Phycobiliprotein genes of the marine prokaryote *Prochlorococcus*: Evidence for rapid evolution of genetic heterogeneity. *Microbiology* **147,** 3171–3182 (2001).

19. Golden, S. S., Brusslan, J. & Haselkorn, R. Expression of a family of *psbA* genes encoding a photosystem II polypeptide in the cyanobacterium *Anacystis nidulans* R2. *EMBO J.* **5,** 2789–2798 (1986).

20. Shen, J. R., Qian, M., Inoue, Y. L. & Burnap, R. L. Functional characterization of *Synechocystis* sp. PCC 6803 ΔpsbU and ΔpsbV mutants reveals important roles of cytochrome c-550 in cyanobacterial

oxygen evolution. *Biochemistry* **37,** 1551–1558 (1998).

21. Garczarek, L., van der Staay, G. W. M., Hess, W. R., Le Gall, F. & Partensky, F. Expression and phylogeny of the multiple antenna genes of the low-light-adapted strain *Prochlorococcus marinus* SS120 (Oxyphotobacteria). *Plant Mol. Biol.* **46,** 683–693 (2001).

22. Nakamura, Y. *et al.* Complete genome structure of the thermophilic Cyanobacterium *Thermosynechococcus elongatus* BP-1. *DNA Res.* **9,** 123–130 (2002).

23. Ishiura, M. *et al.* Expression of a gene cluster *kaiABC* as a circadian feedback process in cyanobateria. *Science* **281,** 1519–1523 (1998).

24. Vaulot, D., Marie, D., Olson, R. J. & Chisholm, S. W. Growth of *Prochlorococcus*, a photosynthetic prokaryote, in the Equatorial Pacific Ocean. *Science* **268,** 1480–1482 (1995).

25. Michaels, M. L. & Miller, J. H. The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-Hydroxyguanine (7,8-Dihydro-8-Oxoguanine). *J. Bacteriol.* **174,** 6321–6325 (1992).

26. Lindell, D. & Post, A. F. Ultraphytoplankton succession is triggered by deep winter mixing in the Gulf of Aqaba (Eilat). *Red Sea. Limnol. Oceanogr.* **40,** 1130–1141 (1995).

27. Karl, D. M., Bidigare, R. R. & Letelier, R. M. Long-term changes in plankton community structure and productivity in the North Pacific Subtropical Gyre: The domain shift hypothesis. *Deep Sea Res. II* **48,** 1449–1470 (2001).

28. Scanlan, D. J., Mann, N. H. & Carr, N. G. The response of the picoplanktonic marine cyanobacterium *Synechococcus* species WH7803 to phosphate starvation involves a protein homologous to the periplasmic phosphate binding protein of *Escherichia coli*. *Mol. Microbiol.* **10,** 181–191 (1993).

29. Johnson, Z. *et al.* Energetics and growth kinetics of a deep *Prochlorococcus* spp. population in the Arabian Sea. *Deep Sea Res. II* **46,** 1719–1743 (1999).

30. Dufresne, A. *et al.* Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a near minimal oxyphototrophic genome. *Proc. Natl Acad. Sci. USA* (in the press).

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

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

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