processes because of their inherent noisiness. However, simulations of the neutral theory are no longer necessary, and all problems with simulations are moot, because an analytical solution is now available.

The lognormal distribution is biologically less informative and mathematically less acceptable as a dynamical null hypothesis for the distribution of RSA than the neutral theory. The parameters of the neutral theory or RSA are directly interpretable in terms of birth and death rates, immigration rates, size of the metacommunity, and speciation rates. A dynamical model of a community cannot yield a lognormal distribution with finite variance because in its time evolution, the variance increases through time without bound. However, as shown in ref. 18, the lognormal distribution can arise in static models, such as those based on niche hierarchy.

The steady-state deficit in the number of rare species compared to that expected under the log series can also occur because rare species grow differentially faster than common species and therefore move up and out of the rarest abundance categories owing to their rare-species advantage[19]. Indeed, it is likely that several different models (such as an empirical lognormal distribution, niche hierarchy models[18] or the theory presented here) might provide comparable fits to the RSA data (we have found that the lognormal does slightly better than the neutral theory for the Pasoh data set[20], obtained in a tropical tree community in Malaysia). Such fitting exercises in and of themselves, however, do not constitute an adequate test of the underlying theory. Neutral theory predicts that the degree of skewing of the RSA distribution ought to increase as the rate of immigration into the local community decreases. Dynamic data on rates of birth, death, dispersal and immigration are needed to evaluate the assumptions of neutral theory and determine the role played by niche differentiation in the assembly of ecological communities.

Our analysis should also apply to the field of population genetics in which the mutation-extinction equilibrium of neutral allele frequencies at a given locus has been studied for several decades[21–26]. □

Received 7 May; accepted 9 June 2003; doi:10.1038/nature01883.

1. MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton Univ. Press, Princeton, NJ, 1967).
2. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, NJ, 2001).
3. McGill, B. J. A test of the unified neutral theory of biodiversity. *Nature* **422,** 881–885 (2003).
4. Condit, R. *et al.* Beta-diversity in tropical forest trees. *Science* **295,** 666–669 (2002).
5. Preston, F. W. The commonness and rarity of species. *Ecology* **29,** 254–283 (1948).
6. May, R. M. *Ecology and Evolution of Communities* 81–120 (Harvard Univ. Press, Cambridge, MA, 1975).
7. Bell, G. Neutral macroecology. *Science* **293,** 2413–2418 (2001).
8. Diamond, J. & Case, T. J. (eds) *Community Ecology* (Harper and Row, New York, NY, 1986).
9. Tilman, D. *Plant Strategies and the Dynamics and Structure of Plant Communities* (Princeton Univ. Press, Princeton, NJ, 1988).
10. Weiher, E. & Keddy, P. *Ecological Assembly Rules: Perspectives, Advances, and Retreats* (Cambridge Univ. Press, Cambridge, UK, 1999).
11. Boswell, M. T., Ord, J. K. & Patil, G. P. *Statistical Distributions in Ecological Work* 3–157 (International Co-operative Publishing, Fairland, MD, 1979).
12. Caraco, T. *Statistical Distributions in Ecological Work* 371–387 (International Co-operative Publishing, Fairland, MD, 1979).
13. Feller, W. *An Introduction to Probability Theory and Its Applications* Vol. 1 (John Wiley & Sons, Hoboken, NJ, 1968).
14. Van Kampen, N. G. *Stochastic Processes in Physics and Chemistry* (Amsterdam, North-Holland, 2001).
15. Fisher, R. A., Corbet, A. S. & Williams, C. B. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12,** 42–58 (1943).
16. Morse, P. M. & Feshbach, H. *Methods of Theoretical Physics* Part 1 (McGraw-Hill, New York, NY, 1953).
17. Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge Univ. Press, Cambridge, UK, 1993).
18. Sugihara, G., Bersier, L., Southwood, T. R. E., Pimm, S. L. & May, R. M. Predicted correspondence between species abundances and dendrograms of niche similarities. *Proc. Natl Acad. Sci. USA* **100,** 5246–5251 (2003).
19. Chave, J., Muller-Landau, H. C. & Levin, S. A. Comparing classical community models: Theoretical consequences for patterns of diversity. *Am. Nat.* **159,** 1–23 (2002).
20. Manokaran, N. *et al. Stand Tables and Species Distributions in the Fifty Hectare Plot at Pasoh Forest Reserve* (Forest Research Institute Malaysia, Kuala Lumpur, 1992).
21. Kimura, M. Evolutionary rate at the molecular level. *Nature* **217,** 624–626 (1968).
22. Ewens, W. J. The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* **3,** 87–112 (1972).
23. Karlin, J. & McGregor, J. Addendum to a paper of W. Ewens. *Theor. Pop. Biol.* **3,** 113–116 (1972).
24. Watterson, G. A. Models for the logarithmic species abundance distributions. *Theor. Pop. Biol.* **6,** 217–250 (1975).
25. Kimura, M. & Ohta, T. *Theoretical Aspects of Population Genetics* (Princeton Univ. Press, Princeton, NJ, 1971).
26. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, UK, 1983).
27. McKane, A., Alonso, D. & Solé, R. V. Mean-field stochastic theory for species-rich assembled communities. *Phys. Rev. E* **62,** 8466–8484 (2000).
28. Rao, C. R. *Statistical Ecology* Vol. 1 *Spatial Patterns and Statistical Distributions* 131–142 (The Penn. State Univ. Press, University Park, PA, 1971).

Acknowledgements We are grateful to O. Kargaltsev for a careful reading of the manuscript. This work was supported by NASA, by grants from the NSF, and by the Department of Plant Biology, University of Georgia.

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

Correspondence and requests for materials should be addressed to J.R.B. (banavar@psu.edu) or S.P.H. (shubbell@dogwood.botany.uga.edu).

The genome of a motile marine *Synechococcus*

B. Palenik[1], B. Brahamsha[1], F. W. Larimer[2,3], M. Land[2,3], L. Hauser[2,3], P. Chain[3,4], J. Lamerdin[3,4], W. Regala[3,4], E. E. Allen[1]*, J. McCarren[1], I. Paulsen[5], A. Dufresne[6], F. Partensky[6], E. A. Webb[7] & J. Waterbury[7]

[1]*Marine Biology Research Division, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0202, USA*
[2]*Computational Biology, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6480, USA*
[3]*Joint Genome Institute, Walnut Creek, California 94598, USA*
[4]*Lawrence Livermore National Laboratory, Livermore, California 94550-9234, USA*
[5]*TIGR, 9712 Medical Center Drive, Rockville, Maryland 20850, USA*
[6]*UMR 7127 CNRS Station Biologique de Roscoff, 29682 Roscoff, France*
[7]*Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA*

* Present address: Department of Earth and Planetary Science, University of California, Berkeley, California 94720, USA

Marine unicellular cyanobacteria are responsible for an estimated 20–40% of chlorophyll biomass and carbon fixation in the oceans[1]. Here we have sequenced and analysed the 2.4-megabase genome of *Synechococcus* sp. strain WH8102, revealing some of the ways that these organisms have adapted to their largely oligotrophic environment. WH8102 uses organic nitrogen and phosphorus sources and more sodium-dependent transporters than a model freshwater cyanobacterium. Furthermore, it seems to have adopted strategies for conserving limited iron stores by using nickel and cobalt in some enzymes, has reduced its regulatory machinery (consistent with the fact that the open ocean constitutes a far more constant and buffered environment than fresh water), and has evolved a unique type of swimming motility. The genome of WH8102 seems to have been greatly influenced by horizontal gene transfer, partially through phages. The genetic material contributed by horizontal gene transfer includes genes involved in the modification of the cell surface and in swimming motility. On the basis of its genome, WH8102 is more of a generalist than two related marine cyanobacteria[2].

Most species of picoplanktonic marine cyanobacteria currently known belong to two genera: *Synechococcus* and *Prochlorococcus*. Members must have the ability to acquire major nutrients and trace metals at the submicromolar concentrations found in the oligotrophic open seas. Their light-harvesting apparatus is uniquely

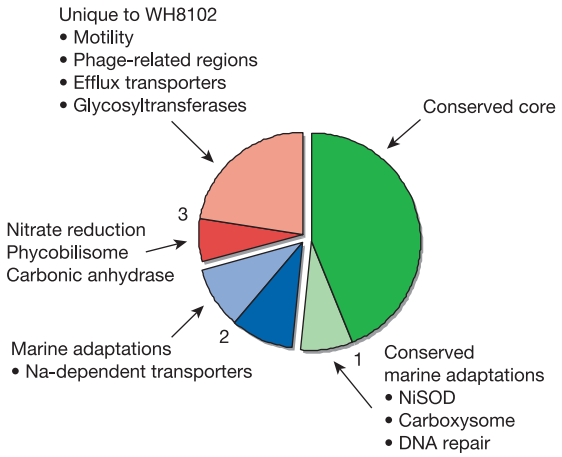

Figure 1 The genome of WH8102 can be divided into three categories: ORFs found in WH8102 and both related *Prochlorococcus* genomes (region 1); ORFs found in WH8102 and only one *Prochlorococcus* genome (region 2); and ORFs not found in the two *Prochlorococcus* genomes (region 3). These regions can be subdivided based on whether or not an ORF has a BLAST hit to freshwater *Synechocystis* sp. strain PCC6803. The lighter shading of each region represents the fraction not in PCC6803. As shown, examination of ORFs in each subcategory has provided insights into the evolutionary adaptation of marine *Synechococcus*.

adapted to the spectral quality of light in the ocean[3,4]. Of the two major marine unicellular genera, *Synechococcus* is usually less abundant in very oligotrophic environments, but has a broader global distribution[1,4]. A great deal of genetic diversity exists within the genus *Synechococcus*, with strains probably adapted to specific ecological niches[3,5]. Furthermore, one group of strains seems to have adapted to the oligotrophic marine environment by developing a new form of swimming motility not seen so far in any other prokaryotic group[6].

On the basis of the genome of *Synechococcus* sp. strain WH8102 (hereafter referred to as WH8102) and comparing it to the related genomes of two *Prochlorococcus* strains[2] we were able to define 1,314 open reading frames (ORFs) common to all three genomes (about half of the genome) and 736 ORFs found only in WH8102 (about a third of the genome) that indicate the specific ecological strategies of WH8102 relative to coexisting *Prochlorococcus* (see Methods, Fig. 1 and Supplementary Table 1). Here we show how the genome reveals some of the interactions of WH8102 with its environment (nutrients, light, toxins) and with other organisms, especially phages.

The WH8102 genome contains 16 probable or possible phage integrases—enzymes that function as site-specific DNA recombinases[7] (Supplementary Table 2). In WH8102, many of these occur adjacent to or near transfer RNAs and in regions with an anomalously low percentage of G + C content (Table 1a and Fig. 2). These regions of low G + C percentage also show atypical trinucleotide composition (data not shown). In addition, possible phage integrase regulators (SYNW2105, SYNW1660, SYNW1665) are also found. Thus, *Synechococcus* has regions that greatly resemble pathogenicity islands—regions that are often mobilized between strains of pathogenic bacteria[7]. Hence, although the genome of WH8102 does not contain prophages or plasmids, it does seem to have been, in its evolutionary past, extensively altered through horizontal gene transfer, possibly due to phages or plasmids. In contrast, far fewer potential phage integrases are found in *Prochlorococcus* (four in MIT9313 and one in MED4).

Other regions of low G + C percentage not associated with phage integrases also show atypical trinucleotide composition, suggestive of recent horizontal gene transfer (Table 1b). These regions encode genes involved in broadening the range of nitrogen substrates that can be used by WH8102, as well as some encoding transport capabilities. Furthermore, several of the genes found in these regions are homologues of ORFs involved in the carbohydrate modification of the cell envelope (including glycosyltransferases, and homologues of genes involved in the synthesis of sialic acid). One hypothesis for the function of these glycosyltransferases is that they may be required in constructing the motility apparatus of this organism, as at least one of its components is glycosylated[8]. Another possibility is that WH8102 may use these envelope-modifying genes to change its cell surface characteristics to help it evade grazers and other predators such as phages. Cell surface properties are known to affect grazing rates in the marine environment[9].

An examination of the genome of WH8102 provides an indication of the uniqueness of *Synechococcus* swimming motility. None of the proteins (motor, flagellar) associated with other forms of prokaryotic motility was found, with the exception of six ORFs associated with type IV-pilus-dependent motility (homologues of *pilB*, *-C*, *-D*, *-Q* and *-T*). Orthologues of these are also present in MIT9313, but not MED4. Nevertheless, these ORFs in WH8102 do not encode the full complement of genes required for pilus assembly and function, and pilin subunit homologues are absent. Pili or surface-associated twitching have not been observed in WH8102.

Recent studies using transposon mutagenesis (J.M. and B.B., manuscript in preparation) coupled with that of motility mutant *swmA*[8] indicate that genes required for motility are found in at least

Table 1 **Atypical regions of G + C per cent in the WH8102 genome**

Region	Size (kb)	No. ORFs	G + C (%)	Acquisition
(a) Low G + C per cent regions associated with predicted phage integrases				
1124989–1158432	33.4	27	53.6	–
1982294–1996886	14.6	14	50.7	Efflux
1606831–1591392	15.4	12	47.8	–
1335428–1344827	9.4	7	49.9	–
2312441–2322542	10.1	12	50.7	–
1488335–1524808	36.0	37	49.8	–
1183175–1171388	11.8	13	51.7	–
857915–842611	15.3	14	49.3	Na/Glut. symporter
2138960–2144868	5.9	4	40.9	–
353145–383689	30.0	29	52.0	–
(b) Low G + C per cent regions not associated with predicted phage integrases				
427233–465883	38.7	37	38.8	Modification of cell envelope, multidrug efflux
622199–633146	10.9	7	43.96	Modification of cell envelope
2379778–2394189	14.4	15	39.0	Cyanate usage, metal uptake
912098–954990	42.9	12	42.2	Motility

two widely separated regions (Fig. 3). The second region contains SwmB (SYNW0953), a very large ORF (10,791 amino acids) that constitutes more than 1% of the genome size and is currently one of the longest bacterial ORFs ever reported. Notably, it is found in one of the unusually low G + C percentage regions (Table 1b).

Transport in WH8102 accounts for about 5–6% of the predicted ORFs, similar to most other bacterial genomes[10]. Compared with other genomes[10], transporter capability is heavily biased towards the use of ABC transporters with about 60% of the ORFs encoding ABC transporter components. A distinct bias against P-type ATPase transporters is found, with only one such transporter, for copper, compared with nine in PCC6803, a model unicellular freshwater cyanobacterium. This one P-type ATPase may be conserved due to the use of copper in plastocyanin, an electron transfer protein that can substitute for an iron-containing cytochrome in photosynthesis.

Notably, WH8102 has multiple channels for transporting major seawater ions and multiple transporters or ABC-type solute-binding proteins for several major nutrients; for example, there are multiple solute-binding proteins for phosphate and two for urea. WH8102 seems to have an independent transporter for urea (SYNW2455), reinforcing its importance as a nitrogen source for cyanobacterial growth in oligotrophic environments. The multiple transporters in WH8102 may have different affinities and be regulated differently depending on nutrient concentrations.

One of the surprises from our analyses of the genome of WH8102 is the prediction that *Synechococcus* can use some new organic compounds as nitrogen and phosphorus sources. As inorganic nitrogen and phosphorus are often thought to be limiting in the marine environment, these potential sources are of particular interest. Amino acid and oligopeptide transporters are found, suggesting that *Synechococcus* may have the ability to use these ubiquitous compounds in sea water; transport of a few amino acids has also been demonstrated[11]. In addition, genes for the transport of cyanate and its breakdown by cyanase appear to be present in WH8102 and in *Prochlorococcus* MED4—in fact, WH8102 grows on

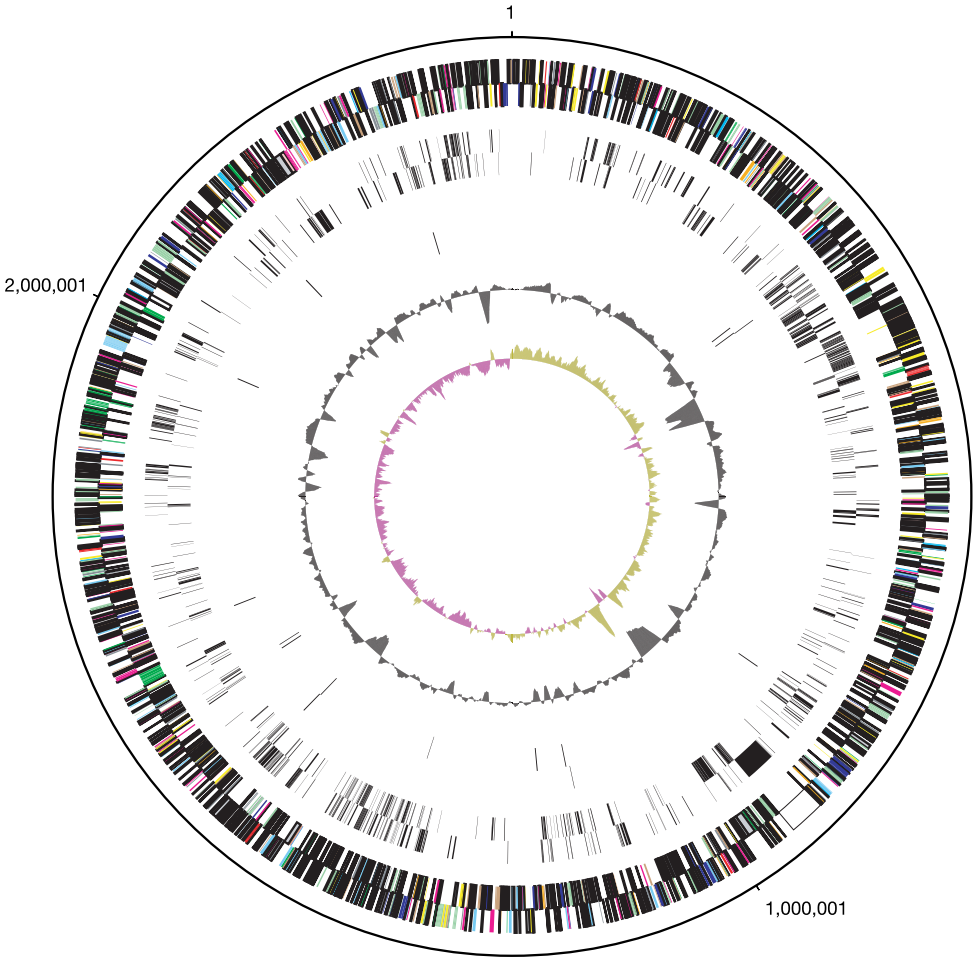

Figure 2 The chromosome of *Synechococcus* sp. strain WH8102. The following description is of the eight circles, with the first circle being the outer circle, and the others progressing inwards. First circle, predicted coding regions on the plus strand coloured by functional category: white, hypothetical; black, unassigned and other; dark red, energy metabolism; green, photosynthesis; blue, DNA replication and repair; cyan, fatty acid metabolism; magenta, biosynthesis of cofactors; yellow, cellular processes; pale green, transport and binding; sky blue, translation; orange, regulatory functions; brown, amino acid biosynthesis; pink, cell envelope; grey, conserved hypothetical; medium red, transcription; light red, purines and pyrimidines; pale pink, central metabolism. Second circle, predicted coding regions on minus strand (same colour scheme as in the first circle). Third and fourth circles, 736 'characteristic' genes not found in two *Prochlorococcus* strains (region 3 in Fig. 1), plus and minus strands, respectively. Fifth and sixth circles, predicted phage integrases on plus and minus strands, respectively. Seventh circle, G + C content (deviation from average); eighth circle, G + C skew curve in purple and olive. Phage integrases are often associated with low G + C regions. Low G + C regions often contain WH8102 characteristic ORFs. Scale (in bp) is indicated along the outside of the circle.

cyanate as a sole nitrogen source (B.P., unpublished data). The genes for cyanate use have been characterized for the freshwater cyanobacterium *Synechococcus elongatus* PCC7942 but cyanate remains uncharacterized as a nitrogen source in aquatic environments[12].

Similarly, genes for the transport of phosphonates (compounds with C–P bonds) are present in WH8102 and are found in the two sequenced *Prochlorococcus* genomes as well. WH8102 grows on phosphonate as a sole phosphorous source (B.P., unpublished data). Phosphonates are known to be produced by some major eukaryotic phytoplankton groups such as the coccolithophorids, and were recently reported to be an important fraction of total phosphate in sea water[13]. WH8102 also has multiple alkaline phosphatases (SYNW0120, SYNW0196, SYNW2391 and SYNW2390) that could be used to obtain phosphate from other organic phosphorus sources in its environment. Thus genome analyses are further dispelling the classical concept of cyanobacteria as being plant-like and dependent solely on inorganic forms of nutrients[14].

In addition, a number of conserved systems for exporting compounds (for example, multidrug efflux systems) are found both in the ABC transporter family and the MFS transporter family. WH8102 has a larger number of efflux transporters in the ABC family compared with *Prochlorococcus*. These results suggest that marine cyanobacteria, despite living in extremely oligotrophic conditions, may still find themselves in the position of needing to export 'toxins' produced by other microorganisms. Antagonistic interactions between pelagic bacteria have been reported recently[15]. Exposure to toxins may be greater for motile *Synechococcus* than for other marine cyanobacteria, as they may be chemotactic towards marine particles where higher localized concentrations of heterotrophic bacteria release nitrogenous coumpounds[11].

WH8102 also has efflux pumps for metals (SYNW1472 and SYNW0900) that are lacking in both *Prochlorococcus* strains. Characterizing these further may put a mechanistic basis behind previous observations[16] that *Synechococcus* seems to be more resistant to copper compared with *Prochlorococcus*, and that this resistance may help explain the seasonal cycles of these organisms in the Sargasso Sea. WH8102 also has predicted genes for the reduction of arsenate to arsenite (SYNW1767) and its efflux (SYNW1039). It has been hypothesized that arsenate is a competitor for phosphate and that systems would be needed to deal with this compound, especially in low-phosphate waters[17].

WH8102 has more capacity for sodium-driven transport than freshwater cyanobacteria such as PCC6803 (see Methods and Fig. 1), with transporters of the alanine/glycine:cation (sodium) symporter family (SYNW0828) and of the neurotransmitter:sodium symporter family (SYNW0699). It also has two transporters from the solute:sodium symporter family (SYNW2455, SYNW0619) compared with one in PCC6803.

In contrast to freshwater cyanobacteria, WH8102 has two potential transporters (SYNW1915, SYNW1916 and SYNW1917, and SYNW0229) for glycine betaine and related compounds found in marine waters. Adjacent to the ABC transporter but on the opposite strand are enzymes predicted to synthesize glycine betaine from glycine (SYNW1914, SYNW1913) using a pathway only reported before from an extremely halophilic proteobacterium[18]. When a freshwater *Synechococcus*, strain PCC7942, was genetically engineered to make glycine betaine, it became more halotolerant[19].

Despite the importance of iron as a limiting nutrient in the oceans, WH8102 does not have a detectable system for siderophore synthesis and uptake. However, it does have strategies for iron conservation such as using plastocyanin (copper) for electron transport and a cobalt-dependent ribonucleotide reductase (SYNW1692) rather than the Fe-containing one found in many other cyanobacteria. Another example of iron conservation in WH8102 is its predicted nickel superoxide dismutase (SOD). Multiple SOD types exist for removing photosynthetically produced superoxide radicals including ones using iron, manganese or copper-zinc as metal cofactors. Unlike the freshwater PCC6803, the marine cyanobacteria WH8102, both *Prochlorococcus* species and *Trichodesmium*, a marine N_2-fixing cyanobacterium (http://www.jgi.doe.gov/JGI_microbial/html/index.html), are predicted to use a new nickel SOD—seen recently in *Streptomyces*—as their only SOD, thus saving iron and manganese for other uses (see Supplementary Fig. 1).

In comparison with the sequenced *Prochlorococcus* strains, WH8102 is a transport generalist. It has predicted transporters for the efflux of chromate (SYNW1323) and arsenite (SYNW1039) that are found in MED4 but not MIT9313. It shares with MIT9313, but not MED4, the ability to use the sodium symporters mentioned above. In addition, WH8102 has transporters that are not found in *Prochlorococcus*, and that are predicted to be involved in the uptake of nitrate, a quaternary ammonium group ($R-N+(CH_3)3$) compound such as sarcosine, another nitrate-like compound, metals (magnesium/cobalt/nickel), and in cation efflux. This may be a characteristic of marine *Synechococcus* in general or it may be a characteristic of motile *Synechococcus*.

In WH8102 the major components of photosynthesis and respiration are well conserved and are usually most closely related to those of other cyanobacteria. Notable exceptions are genes in WH8102 and *Prochlorococcus* implicated in carboxysome structure and assembly, including those encoding the subunits of ribulose-1,5-bisphosphate carboxylase. This is thought to be due to a horizontal gene transfer event[20].

As in most cyanobacteria (other than marine *Prochlorococcus* and two other known prochlorophytes), *Synechococcus* harvests light using phycobilisomes, which are multisubunit complexes binding different types of phycobilins. Two interesting observations differentiate the use of phycobilisomes by WH8102 from those of other cyanobacteria analysed so far. Nowhere in the WH8102 genome are there homologues of the *cpcC* and *cpcD* genes, which in freshwater cyanobacteria are known to encode two types of phycocyanin-

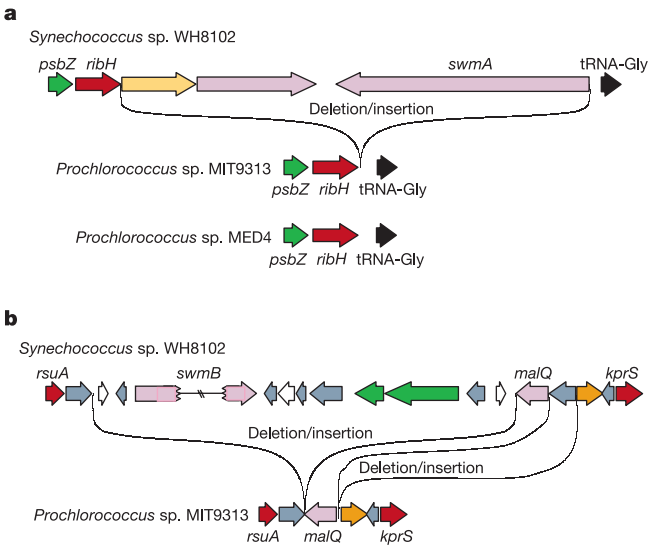

Figure 3 Organization of two chromosomal regions in WH8102 that contain motility genes. **a**, The region containing *swmA* and two other ORFs (a predicted glycosyltransferase and a sulphotransferase) appears to have been inserted in the *ribH*/tRNA-Gly region, or conversely, deleted in the two *Prochlorococcus* genomes. **b**, *swmB* and several other ORFs have been inserted between the *rsuA* and *malQ* regions. The double lines in *swmB* indicate that it is not drawn to scale, as it is approximately 20 times larger than *malQ*. The ORFs are colour-coded by predicted function as described in Fig. 2.

associated L_R linker polypeptides. These linkers are necessary for the correct assembly of phycocyanin discs in the phycobilisome rods[21]. Their absence in WH8102 suggests that there is a single disc of phycocyanin, as is the case in mutants of *Synechococcus* PCC7002 in which the *cpcC* gene has been inactivated[21]. The genome thus provides a basis for the interpretation of absorbance spectra, where reduced phycocyanin (orange-light absorbing) relative to phycoerythrin (blue-light absorbing) probably represents an adaptation to the oligotrophic marine environment where blue light is particularly important[3]. In addition, the genome of WH8102 also lacks homologues of *nblA* and *nblB*, two genes implicated in the degradation of phycobilisomes during nutrient stress in cyanobacteria[22,23]. Thus, phycobilisome degradation may not occur or may be under the control of other genes in WH8102.

Whereas *Synechococcus* possesses homologues of the low-affinity bicarbonate transport mechanism in PCC6803, it lacks homologues of *ndhD3*, *ndhF3* and *chpY*, genes implicated in high-affinity transport in the same organism[20]. Their absence might be an adaptation to the marine habitat, where bicarbonate (2 mM) is probably rarely limiting. Notably, WH8102 has two predicted carbonic anhydrases (SYNW0897 and SYNW2467) whereas *Prochlorococcus* has none, although these genes can be highly divergent and difficult to predict. SYNW2467 is adjacent to the genes encoding nitrate reductase. *Synechococcus* WH8102 can use nitrate for growth in contrast to the two *Prochlorococcus* strains[2], and this carbonic anhydrase may have been lost with the loss of nitrate usage. Although intriguing, a specific connection between nitrate usage and carbonic anhydrase has not been shown.

One way that bacteria sense and respond to their environment is by using two-component regulatory systems consisting of a sensor kinase and a response regulator. In PCC6803 there are nearly 40 sensor kinase and response regulator pairs (http://www.kazusa.or.jp/cyano/index.html). In contrast, WH8102 has only five sensor histidine kinases and nine response regulators, of which one, SYNW1598, may be a pseudogene as it is missing conserved functional residues[24]. Even accounting for a smaller genome size, WH8102 as well as the two *Prochlorococcus* species have fewer systems for responding to environmental changes using these gene families. Furthermore, as there are fewer sensors than response regulators, there seems to be an economy of regulation in which some sensors may transmit information to more than one response regulator.

In addition to the principal RNA polymerase sigma factor *sigA* (SYNW1783), WH8102 encodes five type II sigma factors, typical of cyanobacteria in general[25]. WH8102 however has only one homologue of the type III sigma factor (SYNW1232). This is a low number compared with the three to five seen in other sequenced cyanobacteria (PCC6803, PCC7120 and *Thermosynechococcus*; http://www.kazusa.or.jp/cyano/index.html). One hypothesis for the minimal regulatory machinery (two-component systems and sigma factors) in *Synechococcus* and *Prochlorococcus* is that they have evolved in an open ocean environment that is relatively constant, thus they do not need a regulatory system that could modulate their gene expression to a more variable environment. Alternatively, a minimal regulatory system could be the result of an ecological strategy of only some marine cyanobacteria.

On the basis of its genome, *Synechococcus* WH8102 is clearly more nutritionally versatile and a 'generalist' compared with its *Prochlorococcus* relatives. As the genus *Prochlorococcus* seems to have evolved only once, it may have gone through an evolutionary 'bottleneck' in which its capabilities were originally limited to those of a particular strain followed by subsequent acquisition of new abilities. Alternatively, *Synechococcus* may be more subject than *Prochlorococcus* to horizontal gene transfer from phages, as seen by the presence of more phage integrases. It is possible that not all *Synechococcus* are more versatile in their transport abilities, just the strains that are motile. Partial or complete genomes of additional marine cyanobacteria from this group will help answer these questions. □

Methods

Genome sequencing

Genomic DNA was isolated from WH8102 as reported previously[26]. Whole-genome shotgun libraries were obtained by fragmenting genomic DNA using mechanical shearing and cloning 2–3-kilobase 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). As the first genome drafted during the start-up of the microbial sequencing effort at the J.G.I. Production Sequencing Facility in Walnut Creek, California, this genome was sequenced to unusually high coverage. The whole-genome sequence of WH8102 was obtained from 66,550 reads with an average read length for this project of >575 base pairs (bp) per read for 16-fold redundancy. Sequence assembly was accomplished using PHRAP (P. Green). All gaps were closed by primer walking on gap-spanning library clones or PCR products. The overall genome structure was verified by long-range genomic PCR reactions. The two tandem repeats were resolved by combining information from individual clones, single-nucleotide polymorphism analysis and PCR. Only after this region was finished was it discovered that a single, long ORF was preserved.

Genome analysis

For genome analyses, the combination of three gene modelling programs—Critica[27], Glimmer[28] and Generation (http://compbio.ornl.gov/generation/index.shtml)—was used in the determination of potential coding sequences. These assignments were further checked manually. A revised gene/protein set was searched against the KEGG GENES, Pfam, PROSITE, PRINTS, ProDom and COGs databases, in addition to BLASTP against the NCBI non-redundant database. From these results, categorizations were developed using the KEGG and COGs hierarchies. Transfer RNAs were identified using tRNAscan-SE[29]. To identify regions of atypical nucleotide composition, the trinucleotide composition was determined.

Manual annotation of ORFs was carried out using Artemis, the Artemis Comparison Tool (http://www.sanger.ac.uk/Software/ACT/) and Clustal W[30]. The results of the KEGG and other comparisons described above were examined manually to check automated product assignments and make additional assignments. The proteome sequences of WH8102 and *Prochlorococcus* (MED4 and MIT9313)[2] were compared using the Artemis Comparison Tool. This program, in conjunction with Clustal W, was used for refining predicted start sites, adding ORFs not predicted by the gene modelling programs, and obtaining consistent annotation across three genomes. Manual annotation was done in conjunction with the *Prochlorococcus* annotation team. Transporters were analysed and annotated using methods described in ref. 10.

Pairwise BLAST analyses of three marine cyanobacterial genomes (WH8102 and *Prochlorococcus marinus* strains MED4 and MIT9313 (ref. 2)) against each other and a cut-off *e*-value of e^{-6}, followed by additional manual curation including examination of gene context, were used to partition the genome of WH8102 into three categories: 1,314 ORFs found in all three genomes and predicted to be orthologues; 476 predicted orthologous ORFs found in WH8102 and one other *Prochlorococcus* genome; and 736 ORFs characteristic of WH8102 (not found in either of the other *Prochlorococcus* genomes). The latter category partially represents the ecological capabilities of this organism compared with *Prochlorococcus* (Supplementary Table 3).

Using pairwise BLAST analyses, the three categories of WH8102 ORFs were further subdivided based on whether or not an ORF was found in a model freshwater cyanobacterium *Synechocystis* PCC6803 (hereafter termed as PCC6803; see http://www.kazusa.or.jp/cyano/index.html). After examination of different cut-offs, BLAST analyses with a cut-off *e*-value of e^{-10} were used for this assignment. For the 'core' marine cyanobacterial genome of 1,314 ORFs, 1,112 (85%) are also found in PCC6803 (Fig. 1). This provides an estimate of the portion of the WH8102 genome that has been conserved in all cyanobacterial genomes so far from a primal cyanobacterial ancestor (and includes ORFs conserved in all bacterial taxa). This portion is different from a minimal bacterial genome or a minimal cyanobacterial genome, as horizontally acquired genes could carry out functions required for cell viability. The 15% of the marine cyanobacterial core not in PCC6803 was found to include some of the adaptations and evolutionary events that distinguish the marine *Synechococcus*/*Prochlorococcus* cyanobacterial lineage from other cyanobacterial lineages.

Of the 736 WH8102 characteristic ORFs not found in the *Prochlorococcus* genomes, 23% have related ORFs in PCC6803, using a BLAST cut-off *e*-value of e^{-10}. This is not surprising as these partly represent the ability of both PCC6803 and WH8102, but not the *Prochlorococcus* strains, to create a functional phycobilisome for harvesting light, and a functional nitrate reductase with molybdenum cofactors for using nitrate as a nitrogen source. Forty-five per cent of these characteristic ORFs are hypothetical.

Received 9 May; accepted 28 July 2003; doi:10.1038/nature01943.
Published online 13 August 2003.

1. Partensky, F., Hess, W. R. & Vaulot, D. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol. Mol. Biol. Rev.* **63,** 106–127 (1999).
2. Rocap, G. *et al.* Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424,** 1042–1047 (2003).
3. Waterbury, J. B., Watson, F. W., Valois, F. W. & Franks, D. G. in *Photosynthetic Picoplankton* (eds Platt, T. & Li, W. K. W.) 71–120 (Canadian Department of Fisheries and Oceans, Ottawa, 1986).

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

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

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

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

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

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

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

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

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

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

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

