

ORIGINAL ARTICLE

Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic analysis

Rex R Malmstrom^{1,2,5}, Sébastien Rodrigue^{1,3,5}, Katherine H Huang^{1,4}, Libusha Kelly¹, Suzanne E Kern¹, Anne Thompson¹, Sara Roggensack¹, Paul M Berube¹, Matthew R Henn⁴ and Sallie W Chisholm¹

¹Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA; ²Joint Genome Institute, US Department of Energy, Walnut Creek, CA, USA; ³Département de biologie, Université de Sherbrooke, Sherbrooke, Québec, Canada and ⁴Broad Institute of MIT and Harvard, Cambridge, MA, USA

Prochlorococcus is the numerically dominant photosynthetic organism throughout much of the world's oceans, yet little is known about the ecology and genetic diversity of populations inhabiting tropical waters. To help close this gap, we examined natural *Prochlorococcus* communities in the tropical Pacific Ocean using a single-cell whole-genome amplification and sequencing. Analysis of the gene content of just 10 single cells from these waters added 394 new genes to the *Prochlorococcus* pan-genome—that is, genes never before seen in a *Prochlorococcus* cell. Analysis of marker genes, including the ribosomal internal transcribed sequence, from dozens of individual cells revealed several representatives from two uncultivated clades of *Prochlorococcus* previously identified as HNLC1 and HNLC2. While the HNLC clades can dominate *Prochlorococcus* communities under certain conditions, their overall geographic distribution was highly restricted compared with other clades of *Prochlorococcus*. In the Atlantic and Pacific oceans, these clades were only found in warm waters with low Fe and high inorganic P levels. Genomic analysis suggests that at least one of these clades thrives in low Fe environments by scavenging organic-bound Fe, a process previously unknown in *Prochlorococcus*. Furthermore, the capacity to utilize organic-bound Fe appears to have been acquired horizontally and may be exchanged among other clades of *Prochlorococcus*. Finally, one of the single *Prochlorococcus* cells sequenced contained a partial genome of what appears to be a prophage integrated into the genome.

The ISME Journal (2013) 7, 184–198; doi:10.1038/ismej.2012.89; published online 16 August 2012

Subject Category: microbial ecology and functional diversity of natural habitats

Keywords: HNLC; *Prochlorococcus*; siderophore

Introduction

The unicellular cyanobacterium *Prochlorococcus* has a profound impact on marine carbon cycling because of its high abundance in tropical and subtropical oceans (Partensky *et al.*, 1999). To better understand its ecological role, an integrative approach has been taken to investigate *Prochlorococcus* in terms of its genome content (Dufresne *et al.*, 2003; Roco *et al.*, 2003; Kettler *et al.*, 2007), gene expression (Martiny *et al.*, 2006; Tolonen *et al.*, 2006; Zinser *et al.*, 2009; Thompson *et al.*, 2011a),

cell physiology (Moore and Chisholm, 1999; Moore *et al.*, 2002; Moore *et al.*, 2005), and geographic distribution (Bouman *et al.*, 2006; Johnson *et al.*, 2006; Zwirgmaier *et al.*, 2008; Malmstrom *et al.*, 2010). The emerging picture is of a diverse clade composed of phylogenetic sub-groups with distinct differences in light physiology (for example, high-light (HL) and low-light (LL)-adapted clades) and temperature optima (Johnson *et al.*, 2006; Zinser *et al.*, 2007). These connections between phylogeny and phenotype were uncovered through observations of natural communities as well as experimental manipulations of cultured isolates. Indeed, the ability to study *Prochlorococcus* in both the lab and the field has led to this organism becoming a model for exploring the links between diversity and ecological function in marine systems.

Analyses of environmental DNA sequences have recently revealed new clades of HL-adapted

Correspondence: SW Chisholm, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 15 Vassar Street, Cambridge, MA 02139, USA.
E-mail: chisholm@mit.edu

⁵These authors contributed equally to this work.

Received 29 March 2012; revised 28 June 2012; accepted 2 July 2012; published online 16 August 2012

Prochlorococcus that currently lack cultured representatives (Rusch *et al.*, 2010; West *et al.*, 2011; Huang *et al.*, 2012). More specifically, Rusch *et al.* (2010) identified two groups through phylogenetic analysis of several 'core' functional genes, that is, those common to all sequenced *Prochlorococcus* genomes, from the Global Ocean Survey (GOS) metagenomic data set, whereas West *et al.* (2011) recovered novel 16S rRNA sequences in a separate sampling expedition. In both cases, these new clades were abundant in high nutrient low chlorophyll (HNLC) regions of the ocean where low iron concentrations limit primary production (Martin *et al.*, 1994). It remains unclear, however, if the same clades were identified in both studies since connections cannot be drawn between the core functional genes from one study and the 16S rRNA sequences from the other. More importantly, our understanding of how these HNLC clades might differ genetically from cultured clades of *Prochlorococcus*, or which factors control their geographic distribution, remains incomplete.

Without cultivated representatives, the ecological roles of HNLC clades can only be uncovered through examination of natural assemblages. Metagenomics is a powerful, culture-independent approach for exploring the genetic diversity of marine microbial assemblages (Venter *et al.*, 2004; DeLong *et al.*, 2006), and has already provided new insights into the nitrogen and phosphorous metabolism of wild *Prochlorococcus* populations (Martiny *et al.*, 2009a, b; Coleman and Chisholm, 2010). In fact, assembly of community DNA sequences into composite genomes has provided the first look into the genetic composition of HNLC clades (Rusch *et al.*, 2010). However, these assemblies were composed entirely of genes previously observed in a variety of cultured *Prochlorococcus* strains, and no new genes were identified. This is surprising considering genomes of closely related *Prochlorococcus* isolates typically have hundreds of unique genes (Kettler *et al.*, 2007). It is possible, even likely, that the new clades harbor genes not previously seen in cultured strains, but the inherent ambiguities of metagenomic assembly make it difficult to confidently assign them to HNLC clades.

Single-cell genomics facilitates genomic analysis of uncultured microbes while maintaining unequivocal links between gene sequences and their sources (Stepanuskas and Sieracki, 2007; Woyke *et al.*, 2009), and thus is highly complementary to metagenomics. In this approach, individual cells are separated, often by fluorescence activated cell sorting, and their genomes amplified using a multiple displacement amplification (MDA) strategy that relies on Phi29 polymerase and random hexamers (Raghunathan *et al.*, 2005; Ishoey *et al.*, 2008; Rodrigue *et al.*, 2009). These MDA reactions produce enough DNA for genome sequencing, thus enabling genomic analysis of specific microbes without the need for cultivation. *Prochlorococcus* is ideally suited for examination with both

metagenomics and single-cell genomics as it is well represented in metagenomic surveys of marine communities (DeLong *et al.*, 2006; Rusch *et al.*, 2007), has 12 reference genomes from cultured strains that span the diversity of cultured clades to serve as templates for recruiting metagenomic reads (Kettler *et al.*, 2007), and finally, it is easily identified and sorted based on its unique auto-fluorescence signature (Olson *et al.*, 1990).

Here, we use a combination of single-cell genomics, metagenomics and environmental distribution data to explore the ecology of *Prochlorococcus* HNLC clades. More specifically, we examine natural communities from the Pacific Ocean along a transect from Fiji to Hawaii, and select individual cells from these uncultivated clades for whole-genome amplification and sequencing. Abundance levels of these clades are also determined by qPCR in both the Pacific and Atlantic oceans, revealing highly restricted geographic distribution patterns. Analysis of single-cell genome content and metagenomic composition provide insights into the genetic basis for their adaptation to low iron environments.

Materials and methods

Sample collection and single-cell whole-genome amplification

Samples for single-cell genomics, qPCR and flow cytometry were collected from four depths (15, 45, 100 and 150 m) at eight locations in April 2007 during the CMORE-BULA cruise from Fiji to Hawaii (Hewson *et al.*, 2009). Accompanying physical, chemical and biological measurements can be found at the cruise website (<http://hahana.soest.hawaii.edu/cmorbula>), and corresponding metagenomic data on the CAMERA 2.0 Portal (<http://portal.camera.calit2.net>; CAM_PROJ_Bacterioplankton). Samples for qPCR and flow cytometry were also collected previously on a cruise from the United Kingdom to the Falkland Islands as part of the Atlantic Meridional Transect Project (AMT-13) (Johnson *et al.*, 2006). Seawater samples for single-cell genomics were preserved with 10% glycerol, frozen in liquid nitrogen and kept at -80°C . Flow cytometry samples were immediately fixed with glutaraldehyde (final concentration 0.125% v/v) for 10 min, frozen in liquid nitrogen and stored at -80°C . Samples for qPCR were collected by filtering 100 ml of seawater onto 0.2 μm polycarbonate filters, rinsing in 3 ml wash buffer (10 mM Tris (pH 8.0), 100 mM EDTA, 0.5 M NaCl) and stored at -80°C (Zinser *et al.*, 2006).

Single-cell whole-genome amplification

Single-cell sorting and whole-genome amplification were performed as described previously (Rodrigue *et al.*, 2009) with cells collected at Station 2 (7.5167°S, 167.005°W) at a depth of 15 m. Briefly, *Prochlorococcus* cells were pre-sorted once to reduce contaminating free-DNA dissolved in the seawater

before individual cells were deposited into the wells of a 384-well plate. A mild alkaline lysis was used to release cellular DNA, which was then amplified by MDA using Phi29 DNA polymerase and random hexamers. Successful reactions were identified by PCR screening of the 16S–23S ITS (Martiny *et al.*, 2009c). Sanger sequences of the ITS regions from 94 individual cells were used to construct a neighbor-joining tree that included ITS regions from 13 previously sequenced *Prochlorococcus* genomes. Five cells from each of the two new clades were selected for an additional round of MDA amplification to generate sufficient DNA quantities for genome sequencing. Two microliters of the initial amplification were mixed with the appropriate buffer (40 mM Tris-HCl pH 7.5, 50 mM KCl, 10 mM MgCl₂, 5 mM (NH₄)₂SO₄ and 4 mM DTT), 50 μM phosphorothioate-protected random hexamers, 1 mM dNTPs and 2000 units of RepliPhi Phi29 DNA polymerase in a 50 μl reaction. The reaction was incubated at 30 °C for 10 h and heat-inactivated. The resulting DNA, ~25 μg, was purified with a Qiagen (Valencia, CA, USA) QiaAmp column and directly used to prepare sequencing libraries.

Library preparation, genome assembly and annotation
Roche (Branford, CT, USA) 454-FLX Titanium libraries were prepared as described previously (Margulies *et al.*, 2005; Rodrigue *et al.*, 2009) for 10 selected *Prochlorococcus* cells. Five single cells (three from HL IV and two from HL III) were selected for deeper sequencing using overlapping paired-end reads of 144 nucleotides on the Illumina GAIIx platform. Corresponding PE reads were joined using the SHE-RA software to produce single composite reads of ~180 bp (Rodrigue *et al.*, 2010). These reads were assembled, along with the 454-FLX Titanium reads, in successive batches of 500 000 reads with the Roche gsAssembler version 2.0. The contigs obtained from the first group of 500 000 reads were used to filter redundant reads that were already part of this assembly. The rest of the assembly was performed on batches of 500 000 filtered reads. The resulting contigs were extended using a custom script developed at the DOE Joint Genome Institute (Hess *et al.*, 2011). The extended contigs were next combined into a final assembly with Phrap (Gordon *et al.*, 1998). The initial reads were then aligned against the contigs for quality assurance and the JGI Polisher software was further used to detect and correct potential errors in the assembly. Each assembly was then searched for non-*Prochlorococcus* contigs by using BLASTn against the NCBI nt database (Altschul *et al.*, 1997). Contigs were considered from single cells if the best hit was *Prochlorococcus* at an *e*-value ($\leq 1e^{-5}$, or if they had no hit but their GC content < 40%.

Assembled contigs and raw sequencing reads from single-cell genomes are available through the ProPortal database (<http://proportal.mit.edu>). Contigs >200bp were deposited at DDBJ/EMBL/

GenBank under the accession numbers ALPB 00000000-ALPL00000000. The version described in this paper is the first version, ALB*01000000.

Gene predictions

The gene prediction methods used with the single-cell *Prochlorococcus* genomes have been described in Kettler *et al.* (2007). Briefly, CRITICA version 1.05 (Badger and Olsen, 1999) was first trained using all available sequenced prokaryotic sequences from NCBI Genbank, including all complete sequenced *Prochlorococcus* and marine *Synechococcus* genomes. GLIMMER version 3.02 (Delcher *et al.*, 1999) was then utilized to find additional coding genes that might have been missed. GLIMMER was trained using the predicted CRITICA gene models for the *Prochlorococcus* single cells and all coding genes found in the ProPortal database (Kelly *et al.*, 2012). We observed that both CRITICA and GLIMMER performed poorly when contigs were short (≤ 5 kbp). In these cases, we searched the NCBI nr database using BLASTx for potentially missed known genes (at *e*-value ($\leq 1e^{-5}$).

Counting HL III and HL IV clades and siderophore transporters in metagenomic data sets

Reads from the CMORE-BULA cruise and GOS were first matched against *Prochlorococcus* isolate genomes and *Prochlorococcus* single-cell contigs >500 bp using BLASTn. Those recruits with an *e*-value ($\leq 1e^{-5}$) were then compared against NCBI's nt database to confirm there were no better hits to other microbial sequences. Reads with identical bit scores to more than one genome were discarded. The number reads with top hits to HL III and HL IV were tallied, then normalized to the size of the single-cell genome recovered and the overall size of the metagenomic database from which the read came.

The abundance of siderophore transporter genes (tonB, tolQ, cirA, fepB, C, D) was determined by first matching metagenomic reads to transporter sequences from single-cell W2 using tBLASTx. Reads with an *e*-value ($\leq 1e^{-5}$) were then compared by BLASTx with NCBI's nr database. Reads with either a top hit to the W2 genes or to siderophore transport sequences in *Prochlorococcus* MIT9202 were identified as *Prochlorococcus*-like siderophore transporters. The small number of GOS reads that best matched siderophore transporters in other microbes but whose mate pair had a best hit to *Prochlorococcus* were also identified as *Prochlorococcus*-like siderophore transporters. Counts of transporters were normalized to a gene size of 1000 aa and the overall size of the metagenomic database from which the read came.

Phylogenetic analysis

Neighbor-joining trees of 16S–23S ribosomal ITS, 16S genes and functional gene *rptS* were

constructed from 500 bootstrap iterations using MEGA 5.0 (Tamura *et al.*, 2011). We used sequences from 12 previously described isolate genomes plus one undescribed isolate genome (MIT9202) in all trees. The ribosomal ITS tree also included ITS sequences from the 10 HL III/IV single-cell genomes that were sequenced as part of this study. When possible, we also included 16S and *rptS* sequences from single cells in the corresponding trees, but sequences were not available for every cell due to incomplete genome recovery. Finally, we included the same *rptS* sequences used by Rusch *et al.* (2010) and 16S rRNA sequenced used by West *et al.* (2011) in their phylogenetic analyses of the HL III and HL IV clades.

Quantitative PCR of HL III and HL IV

Primers targeting the ITS regions of HL III and HL IV were designed using ARB and a large database of environmental ITS sequences (Ludwig *et al.*, 2004; Martiny *et al.*, 2009c). Only one specific priming site was identified for HL IV, so this primer was used in conjunction with the more universal *Prochlorococcus* ITS-F primer (5'-CCGAAGTCGTTACTYYAACCC-3'). The specificity of the primers (HL IVr: 5'-CTCTTTGGA WAGGTGGTA-3'; HL IIIf: 5'-CGATCGGAACCTCTGAT TTTCCA-3'; HL IIIr: 5'-TAACAGGAAGCTAGATTCT CCCA-3') was tested against cultured isolates MIT9312, MED4, NATL2a and MIT9313. These tests confirmed specificity within the dynamic range of the assay ($\sim 5\text{--}5 \times 10^5$ cell ml⁻¹). Environmental DNA templates were collected and processed as described previously (Ahlgren *et al.*, 2006; Zinser *et al.*, 2006), and reactions were carried out in 384-well plates on a Roche Light Cycler 480 in 15 μ l volumes. Primers were annealed at 50 °C for 45 s, followed by extension at 72 °C for 20 s and melting at 95 °C for 45 s. Quantification standards for HL III and HL IV covering a range from ~ 5 to 5×10^5 cells ml⁻¹ were generated from 10-fold serial dilutions of representative ITS sequences initially amplified from single-cell MDA products. Values following below the standard curve were set to the theoretical detection limit of 0.65 cells ml⁻¹.

Expression of siderophore transporter genes during iron starvation

Prochlorococcus isolate MIT9202 was grown in PRO99 media (Moore *et al.*, 2007) at 20 °C under a constant illumination of 27 μ E m⁻² s⁻¹. The PRO99 medium was modified for trace-metal clean work through microwave sterilization of seawater, increased EDTA (11.7 μ M), Chelex-100 treatment of major nutrients and soaking polycarbonate culture vessels in 0.1% Citranox, 10% HCl and pH 2 H₂O for 24 h each (Saito *et al.*, 2002).

To initiate iron starvation, an iron-replete culture was grown to mid-log phase and centrifuged at 7500 r.p.m. for 10 m at 21 °C. The cell pellet was rinsed twice in iron-free PRO99 media before being

inoculated into duplicate vessels containing either iron-free or iron-replete media. With each rinse the iron concentration was reduced ~ 1000 -fold with the addition of 25 μ l cell pellet to 25 ml iron-free media, resulting in FeCl₃ concentrations of 1 μ M in the Fe-replete cultures and ~ 1 pM in the Fe-starved cultures. RNA and cell number samples were collected by centrifugation at various time points over 5 days following the initiation of iron stress. RNA samples were extracted using the mirVana miRNA kit (Ambion, Grand Island, NY, USA) and DNA removed with a 60-m Turbo DNase (Ambion) incubation as described previously (Lindell *et al.*, 2007).

Reverse transcription of *cirA*, *fepB*, *fepC*, *fepD*, *tolQ*, *tonB* and *mnpB* was carried out on 2 ng RNA using Super Script II (Invitrogen, Grand Island, NY, USA) following the manufacturer's protocol along with no reverse-transcriptase controls for each gene. Reverse transcription reactions were diluted 1:5 with 10 mM Tris-HCl pH 8.0 before performing triplicate qPCR for each sample and gene with Qiagen Quantitect Sybr Green PCR kit following manufacturer's protocol. An annealing temperature of 54 °C was used for all reactions. The delta-delta Ct method was used to calculate transcript abundances in each sample relative to *mnpB*, a housekeeping gene constitutively expressed in *Prochlorococcus* isolates MED4 and MIT9313 during iron starvation (Livak and Schmittgen, 2001; Thompson *et al.*, 2011a).

Flow cytometric analysis of environmental populations and cultured isolates

An Influx flow cytometer (Becton Dickinson, Franklin Lakes, NJ, USA) was employed to measure cell numbers for the determination of relative cell size and chlorophyll per cell based on the unique autofluorescence and scatter signals of *Prochlorococcus* (Olson *et al.*, 1990). When counting isolates, the relative cell size and chlorophyll per cell were approximated by normalizing forward-angle light scatter and red fluorescence per cell, respectively, to 2 μ m-diameter Fluoresbrite beads (Polysciences, Warrington, PA, USA).

Results and discussion

Identification and sequencing of novel *Prochlorococcus* single cells

To examine the phylogenetic diversity and genetic makeup of *Prochlorococcus* at a site in the equatorial Pacific, we flow sorted hundreds of individual *Prochlorococcus* cells from a 15 m sample at station 2 along the CMORE-BULA transect (Figure 1a) and subjected them to whole-genome amplification. The internal transcribed sequence (ITS) region between 16S and 23S rRNA was sequenced from 94 individually amplified cells. Phylogenetic analysis of these ITS sequences revealed two clades that were

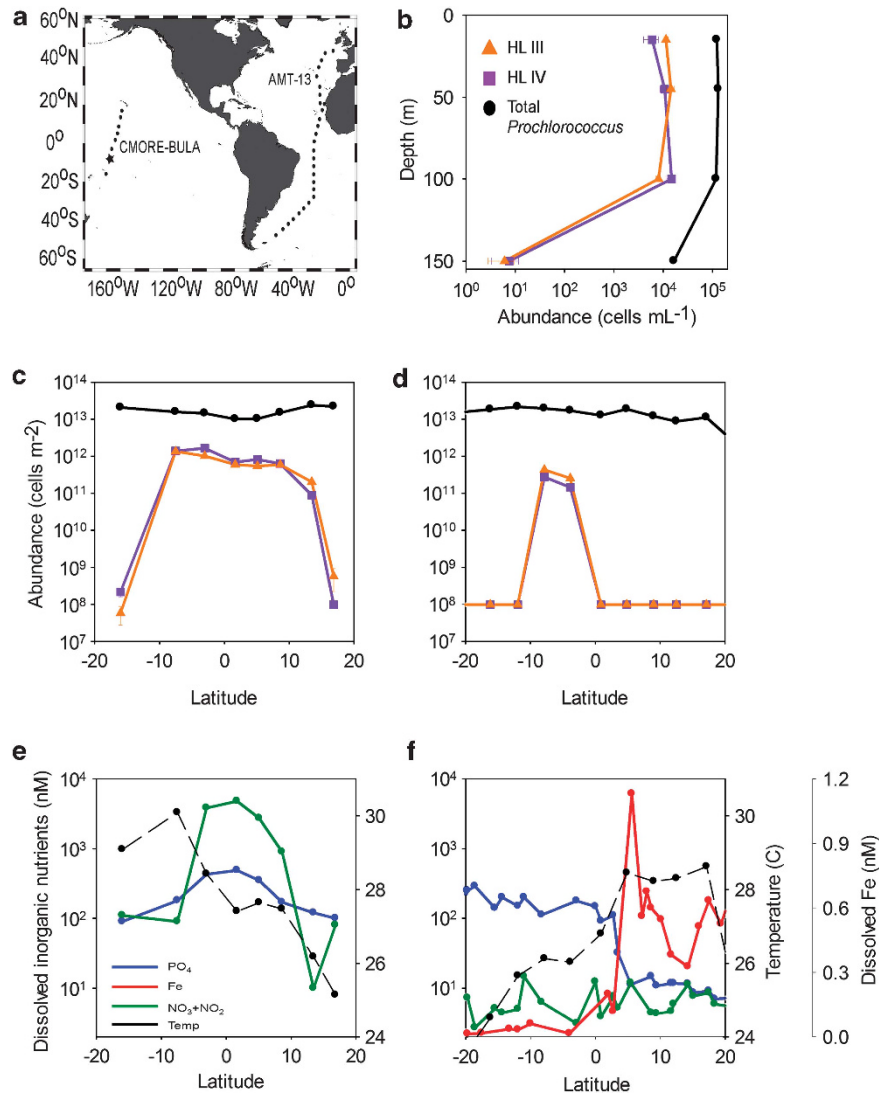


Figure 1 Geographic distribution of cells belonging to HL III and HL IV in the Atlantic and Pacific Oceans. Sampling locations of the CMORE-BULA (Pacific Ocean) and AMT-13 (Atlantic Ocean) expeditions (a). Individual *Prochlorococcus* cells where collected for single-cell whole-genome amplification at CMORE-BULA station 2 (7.5167 S, 167.005 W; identified by star). Depth distribution patterns of HL III and HL IV at CMORE-BULA station 2 (b). Error bars represent one s.d.. Integrated abundances (0–150 m) of total *Prochlorococcus* and HL III and HL IV along the Atlantic AMT-13 (c) and Pacific CMORE-BULA (d) transects were determined by flow cytometry and qPCR, respectively. Corresponding measurements of surface temperature and dissolved inorganic nutrient levels were collected as part the CMORE-BULA expedition (e), whereas dissolved inorganic nutrients and Fe levels in the Atlantic were collected during AMT-17, which followed a similar cruise track to AMT-13 at the latitudes provided (f).

distinct from cultured clades (Figure 2), and were later determined to belong to the uncultured HNLC clades (see below). Roughly half of the ITS sequences recovered in this sample belonged to these uncultured clades, while the rest matched the HL-adapted ecotype HL II; no sequences from HL I or the LL-adapted groups were detected.

New clades of *Prochlorococcus* were independently discovered in two recent studies of the Pacific and Indian Oceans. However, it was unclear whether these previously described clades are the same as connections could not be made between the core functional genes reported by Rusch *et al.* (2010) and the 16S rRNA sequences reported in West *et al.* (2011). The partial genomes assembled *de novo* by

Rusch *et al.* (2010) from metagenomic sequences were not suitable for phylogenetic analysis because the genomes, and the genes comprising them, were composites derived from multiple organisms. In contrast, the single-cell approach described here provides unequivocal linkages among functional genes, 16S rRNA and ribosomal ITS sequences. Thus, we were able to compare our single cells with the gene sequences of the previous studies and determine if the functional genes identified by Rusch *et al.* (2010) are associated with the 16S rRNA sequences identified by West *et al.* (2011). Phylogenetic analyses confirm the same novel clades of *Prochlorococcus* were identified independently in all three cases (Figure 2). However, the

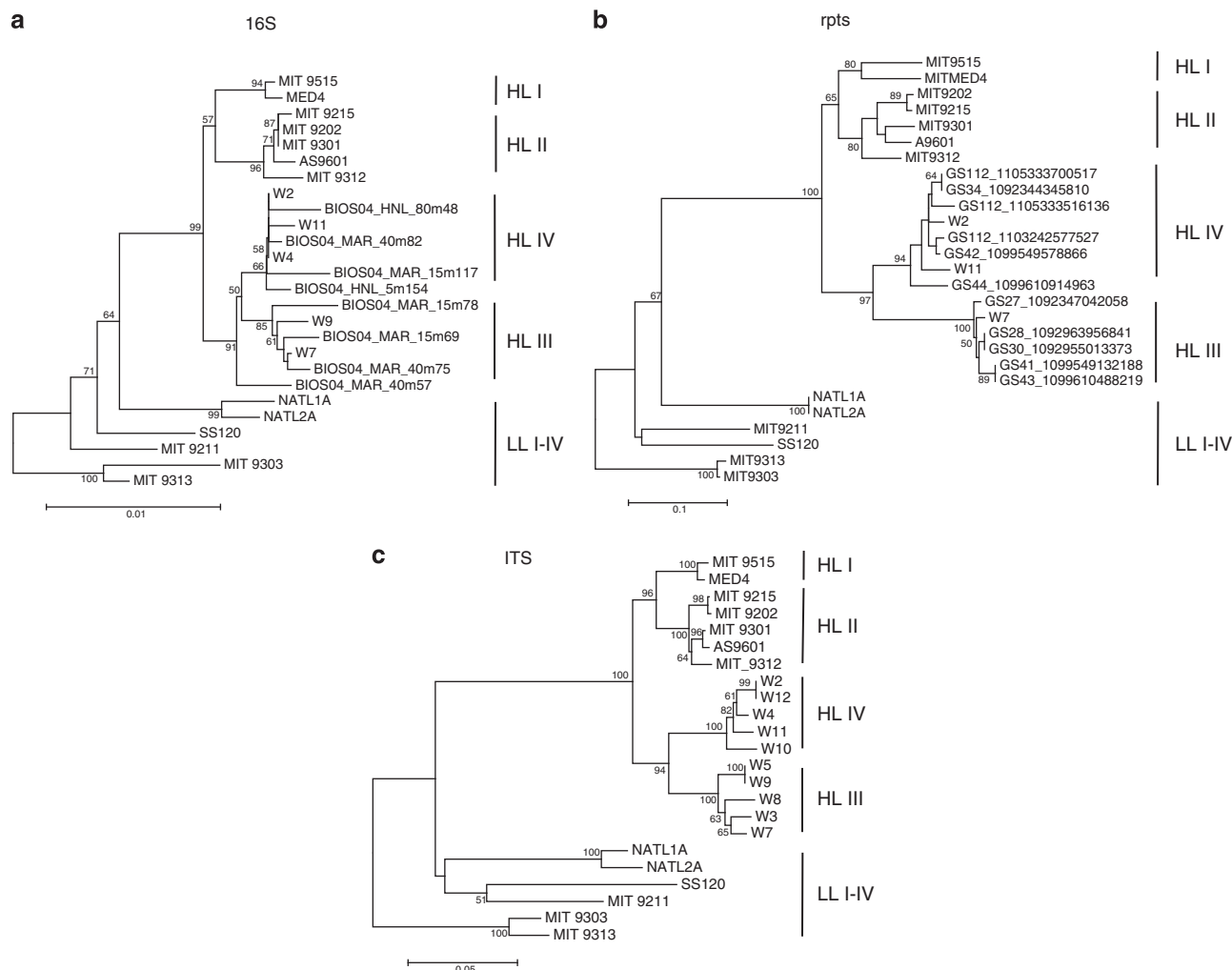


Figure 2 Phylogenetic affiliations of *Prochlorococcus* clades based on functional gene *rptS* (a), 16S rRNA (b), and ribosomal ITS (c) sequences as determined by neighbor-joining analysis (bootstrap = 500). The prefix ‘GS’ identifies sequences from the GOS database identified as HNLC *Prochlorococcus* clades by Rusch *et al.* (2010), whereas the prefix ‘BIOS’ indicates novel 16S rRNA sequences found by West *et al.* (2011). Sequences from uncultured single cells from this study are marked with ‘W’ prefix. Note, functional gene and 16S sequences were not available for all single cells as genome recovery was incomplete. The names of *Prochlorococcus* clades vary within the literature. Here, HL I = eMED4, HL II = eMIT9312 and LL I–IV = eNATL, eSS120, eMIT9211 and eMIT9313. See Huang *et al.* (2012) for additional information on *Prochlorococcus* nomenclature.

designations for HNLC1 and HNLC2 do not agree between previous studies (for example, HNLC1 in Rusch *et al.* (2010) was labeled HNLC2 by West *et al.* (2011), and vice versa). For clarity and conformity with existing nomenclature, we refer to these clades as HL III and HL IV following Huang *et al.* (2012).

To explore the genetic makeup of the HL III and HL IV identified in our sample, 10 single cells from these groups were subjected to whole-genome sequencing using 454-Titanium sequencing technology, of which 5 cells were selected for additional sequencing on the Illumina (San Diego, CA, USA) platform (GAIIx) to improve genome recovery and assembly (Table 1). The total base pairs assembled into contigs for the five deeply sequenced genomes ranged from ~0.77 to ~1.27 Mbp, representing roughly 45–74% of a typical HL-adapted *Prochlorococcus* genome (Rocap *et al.*, 2003; Kettler *et al.*,

2007). The genomes from the 10 single cells collectively contained total of 7535 predicted genes, of which 6131 had homologs in sequenced *Prochlorococcus* isolates. The single-cell genomes also revealed 394 new genes that had never before been seen in *Prochlorococcus* (Table 1), which stands in contrast to the consensus metagenomics assemblies reported previously (Rusch *et al.*, 2010). That is a 4.6% increase in the pan-genome of *Prochlorococcus* conferred by the analysis of only these 10 new partially assembled genomes. While most of these genes do not yet have a proposed function (Supplementary Table 1), their identification highlights the power of single-cell sequencing for revealing organismal diversity by identifying the source-organism of genes from environmental samples. Insights into the ecology of HL III and HL IV provided by a subset of these new genes are discussed below.

Table 1 Assembly statistics of *Prochlorococcus* single-cell genomes

Single-cell amplified genome	W2	W3	W4	W5	W7	W8	W9	W10	W11	W12	Total
Clade	HL IV	HL III	HL IV	HL III	HL III	HL III	HL III	HL IV	HL IV	HL IV	
Number of bp in contigs	1 267 128	3 424 465	7 657 786	1 009 999	9 053 355	8 422 234	4 237 377	5 664 522	7 681 112	4 258 771	6 408 139
Total Contigs	111	217	106	102	110	61	235	378	166	248	1734
Total Contigs > 500 bp	96	105	82	41	73	43	145	224	109	144	1062
Largest contig size	127 061	25 632	58 738	11 787	108 005	124 095	17 912	22 676	60 189	27 866	
N50	50 177	7273	22 887	2288	37 251	91 651	4485	3476	20 326	4307	
Number of predicted genes	1271	511	812	177	986	885	611	873	821	588	7535
Number of Pro core genes	935	383	555	123	755	745	476	673	607	385	5637
Number of Pro or Syn core genes	868	367	515	119	711	704	450	638	571	351	5294
Number of new genes	47	30	42	9	42	35	38	56	54	41	394
Approximate HL genome fraction assembled (%)	74	20	45	6	53	49	25	33	45	25	
Number of 454 reads	491 335	5551	1 210 395	15 554	1 251 850	646 072	83 943	97 383	685 576	88 901	4 576 560
Average length of 454 reads	163.2	668.8	569.4	738.4	531.1	539.4	551.6	569.7	540.8	546.2	
Number of paired-end 454 reads	N/A	45 055	25 128	48 051	22 112	59 024	52 486	33 457	42 718	52 416	380 447
Average length of paired-end 454 reads	N/A	308.9	301.8	307.8	280.9	291.6	306.3	289.9	280.9	293.7	
Number of SHERA-Illumina reads	47 814 244	N/A	6 052 664	N/A	6 460 833	5 627 943	N/A	N/A	10 074 406	N/A	32 997 270
Average length of SHERA-Illumina reads	175.9	N/A	178.7	N/A	167.8	171.9	N/A	N/A	172.9	N/A	

Abbreviations: HL, high-light; N/A, not applicable.

Abundance and distribution of HL III and HL IV

To explore the potential ecological differences between HL III, HL IV and known *Prochlorococcus* ecotypes, we analyzed the abundance and distribution of HL III and HL IV using qPCR in archived samples from the Pacific and Atlantic Oceans (CMORE-BULA and AMT-13, respectively). Depth profiles of the two clades in both the tropical Pacific and Atlantic oceans revealed the typical patterns associated with HL-adapted ecotypes (West *et al.*, 2001; Ahlgren *et al.*, 2006; Johnson *et al.*, 2006; Zinser *et al.*, 2007), that is, relatively high abundance in well-lit surface waters and low abundance below 100 m (Figure 1b), a result consistent with depth profiles in the southwestern Pacific (West *et al.*, 2011). Interestingly, HL III and HL IV always displayed strikingly similar depth distributions within the same location, and their abundances did not differ by >2.3-fold at any depth in the upper 100 m. HL III and HL IV also had identical geographic distributions. Both clades were restricted to equatorial regions in the Atlantic and Pacific (Figures 1c and d), and, when present, accounted for ~5–20% of all *Prochlorococcus* detected by flow cytometry. Their abundance along the CMORE-BULA transect in the tropical Pacific Ocean appears to be even higher based on the

number of metagenomic sequencing reads assigned to HL III and HL IV. Examining metagenomic libraries collected at the same stations where qPCR abundances were determined, we found 17–85% of total *Prochlorococcus* DNA sequences had best hits to HL III and HL IV (Supplementary Table 2), suggesting that abundances derived from qPCR are likely lower bound estimates.

The similar distribution patterns of HL III and HL IV stand in contrast to those of fellow HL-adapted clades HL I and HL II. It is unusual for HL I and HL II ecotypes to be equally abundant at the same location, except at ‘cross-over’ points in their distributions (for example, latitudes –30 S and 35 N in Supplementary Figure 1B), as they have evolved to thrive under different environmental conditions, particularly with regard to temperature (Johnson *et al.*, 2006; Zinser *et al.*, 2007). Interestingly, the evolutionary distance between HL III and HL IV is greater than that between HL II and HL I, as determined from ribosomal and protein coding sequences (Figure 2). We might expect this greater divergence to manifest clear phenotypic differences that would be revealed in their environmental distributions, as is seen with HL II and HL I, yet no clear differences between the distributions of HL III and HL IV were observed along the oceanic gradients

we examined. So while both HL III and HL IV differ from HL II and HL I ecotypes in terms of phylogeny and biogeography, it is not yet clear whether they each represent distinct ecotypes of *Prochlorococcus* or if they are sub-groups of a larger ecotype adapted to equatorial regions.

Although the ecological differences between HL III and HL IV eludes us because of their similar distributions along environmental gradients, comparing them with other HL-adapted clades reveals some interesting differences. First, their distribution in both the Atlantic and Pacific suggests HL III and HL IV are adapted to warm waters. That is, the clades were only abundant in locations with surface temperatures $\geq 26^\circ\text{C}$ (Figure 1), a pattern that also holds with observations from the southwestern Pacific (Supplementary Figure 2) and Indian Ocean (Rusch *et al.*, 2010; West *et al.*, 2011). The HL I clade, in contrast, is adapted to cold waters, while HL II thrives in both moderate and high temperature waters (Supplementary Figure 1) (Johnson *et al.*, 2006; Zinser *et al.*, 2007). Second, HL III and HL IV were only found where inorganic phosphorus levels were relatively high ($>100\text{ nM}$ in this study and West *et al.*, 2011) (Figure 1; Supplementary Figure 2), whereas HL II and HL I inhabit both high and low phosphorous environments, including the Sargasso and Mediterranean Seas where dissolved inorganic phosphorous levels are often $<1\text{ nM}$ (Wu *et al.*, 2000; Marty *et al.*, 2002). Finally, HL III and HL IV appear to be restricted to waters with low iron concentrations. In the Atlantic, HL III and HL IV were only found just south of the equator where measurements along a similar cruise track detected dissolved iron levels of $\sim 0.01\text{ nM}$ (Moore *et al.*, 2009), while HL II dominated waters from 30°N to 30°S where dissolved iron concentrations range from ~ 0.01 to 1 nM (Figure 1) (Johnson *et al.*, 2006; Moore *et al.*, 2009). HL III and HL IV were also only found in waters with $\sim 0.01\text{ nM}$ dissolved iron in the southeastern Pacific (Supplementary Figure 2) (Blain *et al.*, 2008; West *et al.*, 2011). Dissolved iron concentrations were not determined during our Pacific sampling expedition nor at GOS sample sites from which Rusch *et al.* (2010) initially hypothesized the HL III and HL IV clades inhabit low Fe waters, but simulated data from a global chemistry model presented by Rusch *et al.* (2010) suggest that our equatorial Pacific samples also came from low iron environments. Thus, inference from environmental distributions suggests that HL III and HL IV are most likely to appear in warm, low iron and high phosphorous environments, a combination that severely limits their geographic distribution.

The impact of inorganic nitrogen concentrations on HL III and HL IV distributions remains unclear. For example, nitrate + nitrite concentrations were typically high at locations dominated by HL III and HL IV along the CMORE-BULA transect (Figure 1e), whereas West *et al.* (2011) found no correlation

between HL III/IV abundance and nitrate + nitrite in the southeastern Pacific. Nitrite reductase genes were identified in the genomes of LL I and LL IV isolates (Rocap *et al.*, 2003; Kettler *et al.*, 2007), nitrate reductase genes were discovered in natural *Prochlorococcus* communities (Martiny *et al.*, 2009b), and nitrate uptake by *Prochlorococcus* was measured in the Sargasso Sea (Casey *et al.*, 2007). These data indicate some *Prochlorococcus* utilize nitrate and nitrite as N sources. However, we found no evidence of nitrate or nitrite transporters, reductases or molybdopterin biosynthesis genes within the HL III and HL IV single-cell genomes, a result consistent with metagenomic assemblies generated by Rusch *et al.* (2010). In addition, genes belonging to the *Prochlorococcus*-like nitrate assimilation gene cluster were not detected in the CMORE-BULA metagenomic data set at stations 2–6 where HL III and HL IV were most abundant (Figure 1c; Supplementary Table 2), suggesting these genes were not missed in the single cells because of incomplete assembly. Four reads with best hits to the *Prochlorococcus*-like nitrate assimilation gene cluster were found at station 7 where HL III/IV accounted for only $\sim 17\%$ *Prochlorococcus* based on metagenomic counts, and 68 *Prochlorococcus*-like nitrate assimilation genes were detected at station 1 where HL III/IV accounted for $\sim 1\%$ of *Prochlorococcus* (Supplementary Table 2). These data suggest nitrate and nitrite are unlikely to be important N sources to members of the HL III and HL IV clades in the environments we examined.

Adaptations to low iron environments

Iron is a critical micronutrient for photoautotrophic organisms like *Prochlorococcus* that use it primarily in the reaction centers of their photosynthetic and electron transport proteins (Raven *et al.*, 1999; Jordan *et al.*, 2001). Genomic analysis of *Prochlorococcus* isolates suggests they harvest trace amounts of free ferric iron (Fe^{3+}) from the environment to meet their needs (Dufresne *et al.*, 2003; Rocap *et al.*, 2003; Kettler *et al.*, 2007). However, scavenging free Fe^{3+} can present a challenge in waters far away from iron sources as most dissolved iron in the ocean is bound to organic ligands (Rue and Bruland, 1995; Van Den Berg, 1995; Wu and Luther, 1995), a form that *Prochlorococcus* is not known to exploit. In regions with low dissolved iron concentrations, such as the equatorial Pacific and southern equatorial Atlantic, the growth of *Prochlorococcus* can even be limited by Fe levels (Mann and Chisholm, 2000). Therefore, we might expect members of HL III and HL IV, which appear to be restricted to low iron environments, to have specific adaptations enabling them to cope with these conditions. Based on metagenomic reconstructions, Rusch *et al.* (2010) hypothesized that the loss of Fe-containing genes encoding PTOX (plastoquinol terminal oxidase) and cytochrome Cm from both HL III and HL IV reduced

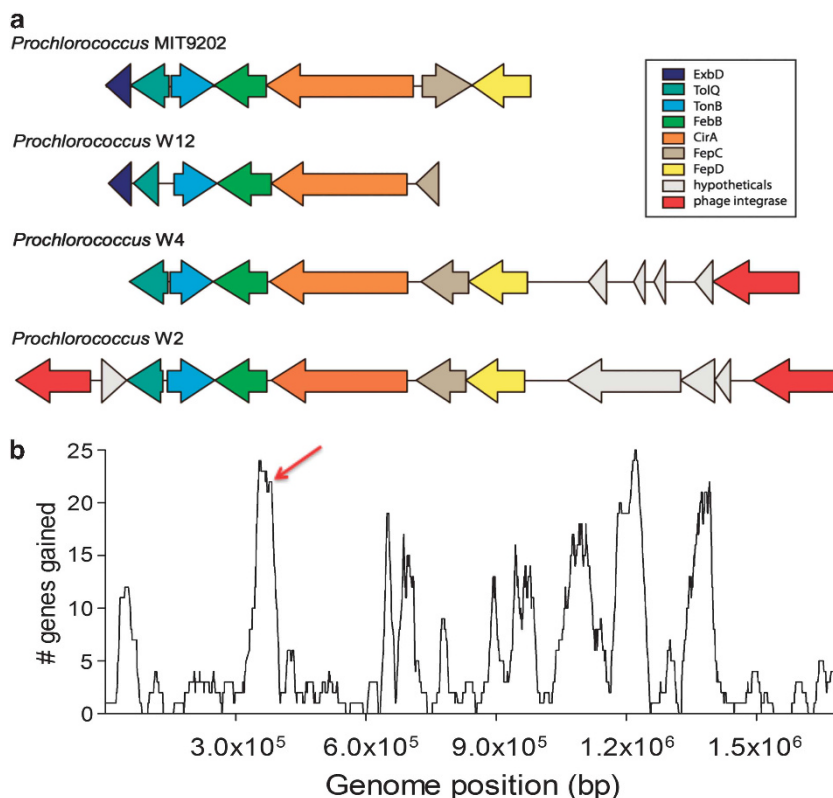


Figure 3 Ton-dependent siderophore acquisition genes observed in *Prochlorococcus* isolate MIT9202 and HL IV single-cell genomes W2, W4 and W12 (a). The contigs from single-cell W2 (contig 41; 69 073 bp) and W4 (contig 50; 27 014 bp) are longer than the region depicted in the panel, but only the regions of interest are presented. The entire single-cell W12 contig 44 (55 47 bp) is represented. A red arrow indicates the position of siderophore transport genes in a genomic island in *Prochlorococcus* MIT9202 (b). Following Kettler *et al.* (2007), genomic islands are identified as regions with high numbers of genes ‘gained’ as departure from the last common ancestor shared with *Synechococcus*.

their iron requirements and represented an adaptation to low Fe environments. However, it is unclear whether the loss of a few Fe-containing proteins could substantially lower Fe requirements as most Fe is used in photosynthetic and electron transporter proteins (Raven *et al.*, 1999; Jordan *et al.*, 2001), which were not missing in the metagenomic assemblies.

Analysis of our single-cell genomes revealed that members of the HL IV clade harbor genes for acquiring Fe bound to organic ligands. More specifically, three of the five partial HL IV genomes (W2, W4 and W12) contain genes for Ton-dependent siderophore acquisition (*tonB*, *tolQ*, *cirA* and *fepBCD*) (Figure 3a). In this pathway, proteins TolQ, R and TonB channel energy generated by proton-motive force to CirA, an Fe-ligand binding protein in the outer membrane, to power transport of the Fe-ligand into the periplasm (Ratledge and Dover, 2000; Postle and Kadner, 2003). The Fe-ligand complex is then transported across the cytoplasmic membrane by FebB, Cand D, where Fe can be stripped from the ligand and utilized. The presence of this transport system indicates that members of the HL IV clade can access the larger pool of organic-bound Fe to help satisfy their requirements, an adaptation that would be particularly useful in low iron

environments. No evidence of siderophore synthesis pathways was found in these genomes, suggesting that they do not make their own siderophores. However, it is possible that the genes for siderophore production could be found in the missing parts of the incomplete single-cell genomes, or that these cells use genes and pathways yet to be characterized.

Siderophore transporter genes were not identified in members of the HL III clade even though they were equally abundant to HL IV cells in iron poor waters. One interpretation is that the HL III clade uses an alternative strategy for coping with low iron concentrations. Partial genomes from HL III single cells lacked Fe-containing genes *PTOX* and cytochrome Cm, which would be consistent the hypothesis that this clade minimizes Fe demand by decreasing the number of Fe-containing proteins (Rusch *et al.*, 2010). However, gene absence must be interpreted cautiously when dealing with incomplete single-cell genomes and metagenomic reconstructions. Indeed, *PTOX*, cytochrome Cm and Ton-dependent siderophore transporters were missed in the metagenomic assembly of the HL IV clade (Rusch *et al.*, 2010), but were found in the single-cell genomes. Thus, it is quite possible that Ton-dependent siderophore transporters, or even

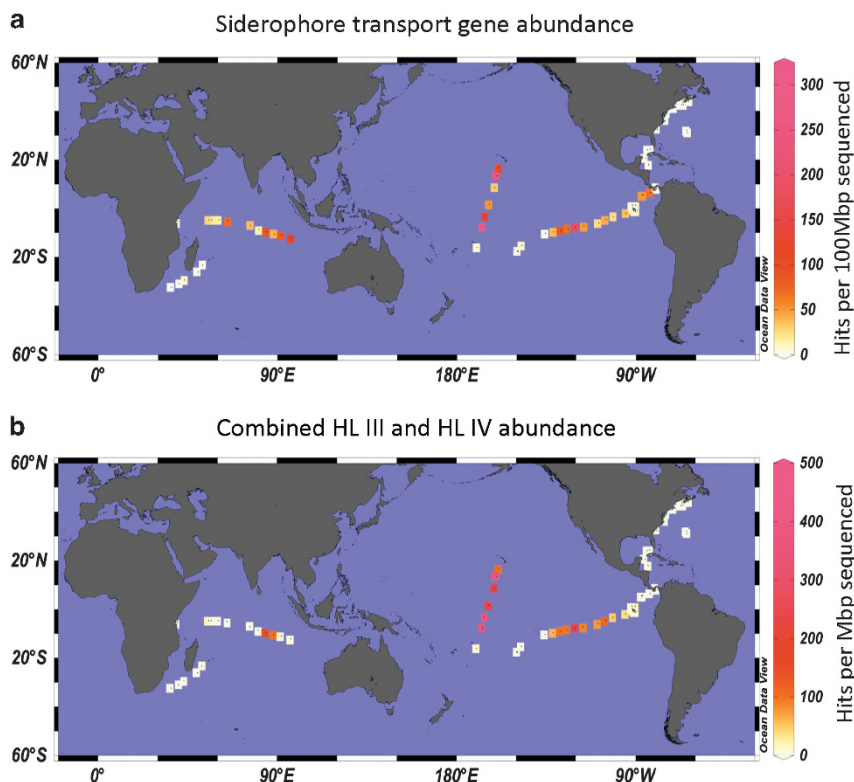


Figure 4 Abundance of *Prochlorococcus*-like siderophore transporters (a) and HL III and HL IV (b) in metagenomic data sets collected at the GOS and CMORE-BULA sites. Counts of siderophore transporter genes (*tonB*, *tolQ*, *cirA* and *fepBCD*) were normalized for gene length to 1000 aa, whereas the combined count of reads matching HL III and HL IV were normalized to the size of the recovered genomes. All data were normalized to the total database size at each location.

other Fe acquisition genes, are encoded in the missing portions of our HL III genomes. To address this point, we examined mate pair reads from the GOS data set where one read mapped to a *Prochlorococcus*-like siderophore transport gene and the other mapped outside this suite of siderophore genes. Focusing specifically on GOS sites 39–46, where HL III and HL IV dominate *Prochlorococcus* populations (Figure 4a), we found 5 of the 19 mates had best hits (blastn) to HL III single-cell genomes (Supplementary Table 3). This suggests that members of HL III clade may also access ligand-bound iron using siderophore transporters.

To better understand the importance of siderophore transporters to HL III and HL IV, we examined the abundance of these specific genes (that is, *Prochlorococcus*-like *tonB*, *tolQ*, *cirA* and *fepBCD*) in metagenomic samples collected along the BULA transect from Fiji to Hawaii, and at all GOS sampling locations. The abundance of *Prochlorococcus*-like siderophore transporters is clearly elevated at the specific locations in the Pacific and Indian Oceans where HL III and HL IV are also abundant (Figure 4). This suggests uptake of organic-bound iron may be a common feature to these clades regardless of geographic location. However, the abundance of *Prochlorococcus*-like siderophore transporters was also elevated at a few locations where HL III/IV abundance was low, for

example, other locations in the Indian Ocean and far eastern equatorial Pacific. This suggests Ton-dependent siderophore transport systems are not exclusive to HL III and HL IV, but instead are distributed among other *Prochlorococcus* ecotypes. In support of this observation, many of the GOS mate reads paired with *Prochlorococcus*-like siderophore transporters had top hits to the cultured HL-adapted clades (Supplementary Table 3). Finally, the same suite of transporters was also found in cultured strain *Prochlorococcus* MIT9202 (Figure 3), a previously undescribed member of the HL II ecotype isolated from the equatorial Pacific Ocean (an HNLC region). Thus, *Prochlorococcus*-like siderophore transport genes are not exclusive to HL III and HL IV. Like P and N-related genes in *Prochlorococcus* (Martiny *et al.*, 2006, 2009b; Coleman and Chisholm, 2010), their distribution in the oceans seem to be influenced by habitat and are not necessarily correlated with ecotype abundance.

Consistent with this image, the siderophore transporters in the cultured strain of *Prochlorococcus* MIT9202 from the Equatorial Pacific appear to be located in a genomic island, which constitutes a site of frequent horizontal gain and loss in *Prochlorococcus* (Figure 3b) (Coleman *et al.*, 2006; Kettler *et al.*, 2007). Similarly, two phage integrase genes flank the suite of siderophore transport genes in the W2 single-cell genome, while another phage integrase sits

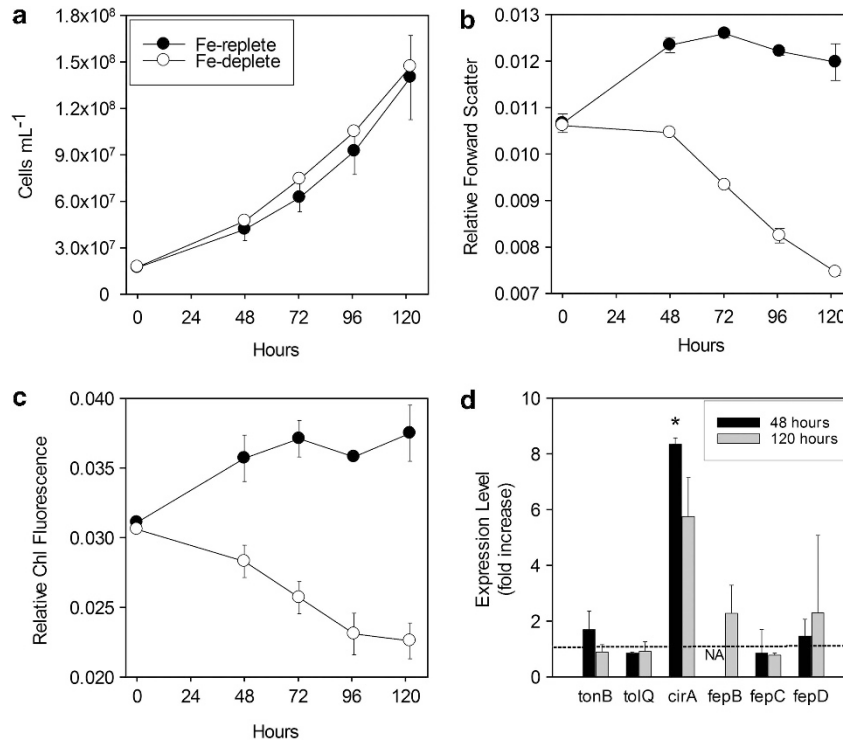


Figure 5 Abundance, mean forward light scatter per cell, and mean chlorophyll fluorescence per cell in *Prochlorococcus* MIT9202 under iron starvation conditions (a–c). Time 0 h indicates when cultures were transferred from Fe-replete (Fe +) conditions to Fe-starved (Fe –). Samples were taken at 48 and 120 h to quantify expression of siderophore acquisition genes *tonB*, *tolQ*, *cirA*, *fepB*, *C*, *D* by qPCR (d). Expression is reported as fold-change under Fe-starved conditions relative to expression under Fe-replete conditions. A reference line indicating no change in expression level (that is, fold-increase of 1) is provided. Error bars in all panels represent 1 s.d. Significantly different expression levels (P -value < 0.05) are indicated with an *.

adjacent to this suite of genes in the W4 single-cell genome (Figure 3a). Owing to the small size of W12 contig 44, we were unable to determine if phage integrase genes were adjacent to the suite of siderophore transporters. The location of these transport genes in MIT9202, and the close proximity of phage integrase genes in multiple HL IV genomes, suggest this pathway may have been acquired by virus-mediated horizontal gene transfer, and possibly exchanged among *Prochlorococcus* clades inhabiting low Fe environments.

Expression of siderophore acquisition genes

Although we could not directly test the role of the siderophore transporters in the Fe metabolism of HL III and HL IV, we were able to examine their response to Fe stress in laboratory experiments with *Prochlorococcus* isolate MIT9202. More specifically, we measured expression levels of the siderophore transport genes *tonB*, *tolQ*, *cirA* and *fepB*, *C*, *D* in cultures grown under Fe-replete and Fe-limited conditions. Although total cell numbers increased steadily over 120 h of Fe starvation, significant reductions in chlorophyll fluorescence and forward scatter, which is related to cell size, were observed (Figure 5), indicating the MIT9202 cells were experiencing significant physiological stress. The gene encoding the outer membrane Fe-binding

component (*cirA*) was upregulated after 48 h of Fe starvation (t -test; $P < 0.05$), whereas the expression levels of the other genes in the Ton-dependent system were not significantly different under iron starvation (Figure 5d). This suggests MIT9202 may respond to low Fe concentrations by increasing siderophore transport proteins in the outer membrane to facilitate transport of Fe-ligands into the periplasm, possibly because this step represents a bottleneck in Fe acquisition. Members of HL III and HL IV presumably do the same.

Detection of a potential phage in HL III and HL IV

Viruses are thought to have an important role in the exchange of genetic material between *Prochlorococcus* cells (Lindell *et al.*, 2004; Thompson *et al.*, 2011b), as highlighted above in the case of siderophore transporter genes. Thus far, only lytic phages infecting *Prochlorococcus* have been reported (Sullivan *et al.*, 2003, 2005), although a putative integrase gene along with a perfect 42 bp attP site corresponding to a *Prochlorococcus* tRNA-Leu have been observed in the genome of the podovirus P-SSP7 (Sullivan *et al.*, 2005), suggesting this phage may also be lysogenic. Several attempts to isolate P-SSP7 lysogens at this position in the genome of the *Prochlorococcus* MED4 strain under laboratory conditions have been unsuccessful

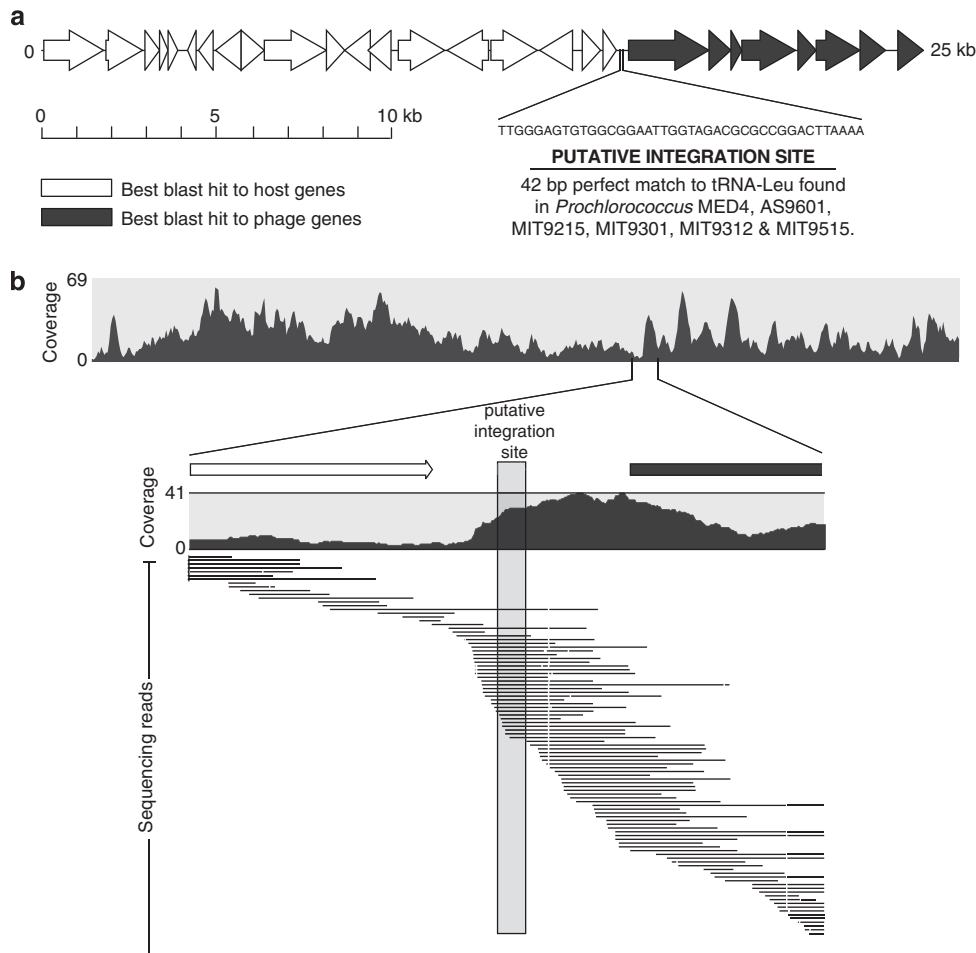


Figure 6 Cyanophage genes integrated into single-cell *Prochlorococcus* genome W8, a member of the HL III clade. The putative integration site of phage genes occurs at tRNA-Leu, identified in panel (a). Multiple reads span the integration site, supporting the assembly results (b). Uneven coverage along the assembled contig, including near the putative integration site, is characteristic of templates generated by MDA.

(unpublished results). However, we observed a partial P-SSP7-like prophage at the predicted locus in our single-cell genome W8, a cell that belongs to the HL III clade. The 25 kbp contig is composed of 21 predicted genes from the host followed by the tRNA-Leu and nine phage genes. The junction between the host and prophage gene is supported by several sequencing reads, and it is thus unlikely that the contig is chimeric (Figure 6). As predicted for P-SSP7, the first viral gene in the contig encodes an RNA polymerase, suggesting that the integrase and *attP* sequence would be located immediately upstream in the bacteriophage DNA molecule. The rest of the prophage genome is missing from this single-cell assembly, so it remains unclear if it encodes a functional virus. Still, the presence of a classically lytic phage integrated into its host genome is intriguing and could represent an adaptation to stressful conditions. For example, a mechanism could have evolved for the originally lytic phage to sense certain environmental parameter and become lysogenic until more favorable conditions are found. The presence of functional

integrase in phage genomes could contribute to lateral gene exchange such as the transfer of siderophore acquisition genes between *Prochlorococcus* cells as discussed above.

Conclusions

Single-cell genomics is a powerful approach that, in combination with genomics and metagenomics, enhances our ability to understand the structure and function of genetic diversity among microbes. Using this method, we were able to uncover siderophore transporters that could be unambiguously assigned to *Prochlorococcus*, providing insights into how *Prochlorococcus* is adapted to low Fe concentrations. Our work also further characterized two newly discovered *Prochlorococcus* clades, suggesting that dissolved phosphorus concentrations, in addition to Fe and temperature, may also have an important role in control the abundance of HL III and HL IV. In addition, the detection of a partial P-SSP7-like prophage in one of the single-cell genomes provides the most direct evidence to date

that marine cyanophage do have lysogeny as part of their life cycle. Although this study combines genome insights with data from quantitative distribution, further work will be needed to obtain cultured isolates and understand the molecular basis for the restricted distribution of HL III and HL IV.

Acknowledgements

We thank Doug Rusch and Nyree West for providing sequence data, and C Mark Moore and Malcolm Woodward for providing nutrient data for the Atlantic Meridional Transect 17. Greg Kettler assisted with identifying genomic islands in *Prochlorococcus* MIT9202. We thank the Broad Institute Genome Sequencing Platform for their contributions to the generation of genomic data. This work was supported by the Gordon and Betty Moore Foundation (SWC and MRH) and the NSF Center for Microbial Oceanography Research and Education (SWC). Sébastien Rodrigue was supported by postdoctoral fellowships from the Natural Sciences and Engineering Research Council of Canada and the 'Fonds québécois de la recherche sur la nature et les technologies.'

References

- Ahlgren NA, Rocop G, Chisholm SW. (2006). Measurement of *Prochlorococcus* ecotypes using real-time polymerase chain reaction reveals different abundances of genotypes with similar light physiologies. *Environ Microbiol* **8**: 441–454.
- Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.
- Badger JH, Olsen GJ. (1999). CRITICA: coding region identification tool invoking comparative analysis. *Mol Biol Evol* **16**: 512–524.
- Blain S, Bonnet S, Guieu C. (2008). Dissolved iron distribution in the tropical and sub tropical South Eastern Pacific. *Biogeosciences* **5**: 269–280.
- Bouman HA, Ulloa O, Scanlan DJ, Zwirgmaier K, Li WKW, Platt T *et al.* (2006). Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes. *Science* **312**: 918–921.
- Casey JR, Lomas MW, Mandecki J, Walker DE. (2007). *Prochlorococcus* contributes to new production in the Sargasso Sea deep chlorophyll maximum. *Geophys Res Lett* **34**: L10604.
- Coleman ML, Chisholm SW. (2010). Ecosystem-specific selection pressures revealed through comparative population genomics. *Proc Natl Acad Sci USA* **107**: 18634–18639.
- Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, DeLong EF *et al.* (2006). Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311**: 1768–1770.
- Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. (1999). Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* **27**: 4636–4641.
- DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU *et al.* (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496–503.
- Dufresne A, Salanoubat M, Partensky F, Artiguenave F, Axmann IM, Barbe V *et al.* (2003). Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proc Natl Acad Sci USA* **100**: 10020–10025.
- Gordon D, Abajian C, Green P. (1998). Consed: a graphical tool for sequence finishing. *Genome Res* **8**: 195–202.
- Hess M, Sczyrba A, Egan R, Kim TW, Chokhawa H, Schroth G *et al.* (2011). Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. *Science* **331**: 463–467.
- Hewson I, Paerl RW, Tripp HJ, Zehr JP, Karl DM. (2009). Metagenomic potential of microbial assemblages in the surface waters of the central Pacific Ocean tracks variability in oceanic habitat. *Limnol Oceanograph* **54**: 1981–1994.
- Huang SJ, Wilhelm SW, Harvey HR, Taylor K, Jiao NZ, Chen F. (2012). Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *Isme J* **6**: 285–297.
- Ishoey T, Woyke T, Stepanauskas R, Novotny M, Lasken RS. (2008). Genomic sequencing of single microbial cells from environmental samples. *Curr Opin Microbiol* **11**: 198–204.
- Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EMS, Chisholm SW. (2006). Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311**: 1737–1740.
- Jordan P, Fromme P, Witt HT, Klukas O, Saenger W, Krauss N. (2001). Three-dimensional structure of cyanobacterial photosystem I at 2.5 angstrom resolution. *Nature* **411**: 909–917.
- Kelly L, Huang KH, Ding HM, Chisholm SW. (2012). ProPortal: a resource for integrated systems biology of *Prochlorococcus* and its phage. *Nucleic Acids Res* **40**: D632–D640.
- Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, Rodrigue S *et al.* (2007). Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet* **3**: 2515–2528.
- Lindell D, Jaffe JD, Coleman ML, Futschik ME, Axmann IM, Rector T *et al.* (2007). Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449**: 83–86.
- Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, Chisholm SW. (2004). Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc Natl Acad Sci USA* **101**: 11013–11018.
- Livak KJ, Schmittgen TD. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2^(-T)(-Delta Delta C) method. *Methods* **25**: 402–408.
- Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar A *et al.* (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.
- Malmstrom RR, Coe A, Kettler GC, Martiny AC, Frias-Lopez J, Zinser ER *et al.* (2010). Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans. *Isme J* **4**: 1252–1264.
- Mann EL, Chisholm SW. (2000). Iron limits the cell division rate of *Prochlorococcus* in the eastern equatorial Pacific. *Limnol Oceanograph* **45**: 1067–1076.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bembem LA *et al.* (2005). Genome sequencing in

- microfabricated high-density picolitre reactors. *Nature* **437**: 376–380.
- Martin JH, Coale KH, Johnson KS, Fitzwater SE, Gordon RM, Tanner SJ *et al.* (1994). Testing the iron hypothesis in ecosystems of the equatorial Pacific-Ocean. *Nature* **371**: 123–129.
- Martiny AC, Coleman ML, Chisholm SW. (2006). Phosphate acquisition genes in *Prochlorococcus* ecotypes: Evidence for genome-wide adaptation. *Proc Natl Acad Sci USA* **103**: 12552–12557.
- Martiny AC, Huang Y, Li WZ. (2009a). Occurrence of phosphate acquisition genes in *Prochlorococcus* cells from different ocean regions. *Environ Microbiol* **11**: 1340–1347.
- Martiny AC, Kathuria S, Berube PM. (2009b). Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes. *Proc Natl Acad Sci USA* **106**: 10787–10792.
- Martiny AC, Tai APK, Veneziano D, Primeau F, Chisholm SW. (2009c). Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. *Environ Microbiol* **11**: 823–832.
- Marty JC, Chiaverini J, Pizay MD, Avril B. (2002). Seasonal and interannual dynamics of nutrients and phytoplankton pigments in the western Mediterranean Sea at the DYFAMED time-series station (1991–1999). *Deep-Sea Res Part II* **49**: 1965–1985.
- Moore LR, Chisholm SW. (1999). Photophysiology of the marine cyanobacterium *Prochlorococcus*: Ecotypic differences among cultured isolates. *Limnol Oceanograph* **44**: 628–638.
- Moore LR, Coe A, Zinser ER, Saito MA, Sullivan MB, Lindell D *et al.* (2007). Culturing the marine cyanobacterium *Prochlorococcus*. *Limnol Oceanograph* **5**: 353–362.
- Moore LR, Ostrowski M, Scanlan DJ, Feren K, Sweetsir T. (2005). Ecotypic variation in phosphorus acquisition mechanisms within marine picocyanobacteria. *Aquat Microb Ecol* **39**: 257–269.
- Moore LR, Post AF, Rocap G, Chisholm SW. (2002). Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Limnol Oceanograph* **47**: 989–996.
- Moore MC, Mills MM, Achterberg EP, Geider RJ, LaRoche J, Lucas MI *et al.* (2009). Large-scale distribution of Atlantic nitrogen fixation controlled by iron availability. *Nature Geosci* **2**: 867–871.
- Olson RJ, Chisholm SW, Zettler ER, Altabet MA, Dusenberry JA. (1990). Spatial and Temporal Distributions of *Prochlorophyte* Picoplankton in the North-Atlantic Ocean. *Deep-Sea Res Part a* **37**: 1033–1051.
- Partensky F, Hess WR, Vault D. (1999). *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* **63**: 106–127.
- Postle K, Kadner RJ. (2003). Touch and go: tying TonB to transport. *Mol Microbiol* **49**: 869–882.
- Raghunathan A, Ferguson HR, Bornarth CJ, Song WM, Driscoll M, Lasken RS. (2005). Genomic DNA amplification from a single bacterium. *Appl Environ Microbiol* **71**: 3342–3347.
- Ratledge C, Dover LG. (2000). Iron metabolism in pathogenic bacteria. *Annu Rev Microbiol* **54**: 881–941.
- Raven JA, Evans MCW, Korb RE. (1999). The role of trace metals in photosynthetic electron transport in O(2)-evolving organisms. *Photosynth Res* **60**: 111–149.
- Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA *et al.* (2003). Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**: 1042–1047.
- Rodrigue S, Malmstrom RR, Berlin AM, Birren BW, Henn MR, Chisholm SW. (2009). Whole Genome Amplification and *de novo* Assembly of Single Bacterial Cells. *Plos One* **4**: e6864.
- Rodrigue S, Materna AC, Timberlake SC, Blackburn MC, Malmstrom RR, Alm EJ *et al.* (2010). Unlocking Short Read Sequencing for Metagenomics. *Plos One* **5**: e11840.
- Rue EL, Bruland KW. (1995). Complexation of iron(III) by natural organic-ligands in the central north pacific as determined by a new competitive ligand equilibration adsorptive cathodic stripping voltammetric method. *Marine Chem* **50**: 117–138.
- Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S *et al.* (2007). The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *Plos Biol* **5**: 398–431.
- Rusch DB, Martiny AC, Dupont CL, Halpern AL, Venter JC. (2010). Characterization of *Prochlorococcus* clades from iron-depleted oceanic regions. *Proc Natl Acad Sci USA* **107**: 16184–16189.
- Saito MA, Moffett JW, Chisholm SW, Waterbury JB. (2002). Cobalt limitation and uptake in *Prochlorococcus*. *Limnol Oceanograph* **47**: 1629–1636.
- Stepanauskas R, Sieracki ME. (2007). Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc Natl Acad Sci USA* **104**: 9052–9057.
- Sullivan MB, Coleman ML, Weigle P, Rohwer F, Chisholm SW. (2005). Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations. *Plos Biol* **3**: 790–806.
- Sullivan MB, Waterbury JB, Chisholm SW. (2003). Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424**: 1047–1051.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2731–2739.
- Thompson AW, Huang K, Saito MA, Chisholm SW. (2011a). Transcriptome response of high- and low-light-adapted *Prochlorococcus* strains to changing iron availability. *ISME J* **5**: 1580–1594.
- Thompson LR, Zeng Q, Kelly L, Huang KH, Singer AU, Stubbe J *et al.* (2011b). Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc Natl Acad Sci USA* **108**: E757–E764.
- Tolonen AC, Aach J, Lindell D, Johnson ZI, Rector T, Steen R *et al.* (2006). Global gene expression of *Prochlorococcus* ecotypes in response to changes in nitrogen availability. *Mol Syst Biol* **2**: 53.
- Van Den Berg CMG. (1995). Evidence for Organic Complexation of Iron in Seawater. *Marine Chem* **50**: 139–157.
- Venter JC, Remington K., Heidelberg JF, Halpern AL, Rusch D, Eisen JA *et al.* (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66–74.
- West NJ, Lebaron P, Strutton PG, Suzuki MT. (2011). A novel clade of *Prochlorococcus* found in high nutrient low chlorophyll waters in the South and Equatorial Pacific Ocean. *ISME J* **5**: 933–944.

- West NJ, Schonhuber WA, Fuller NJ, Amann RI, Rippka R, Post AF *et al.* (2001). Closely related *Prochlorococcus* genotypes show remarkably different depth distributions in two oceanic regions as revealed by in situ hybridization using 16S rRNA-targeted oligonucleotides. *Microbiology* **147**: 1731–1744.
- Woyke T, Xie G, Copeland A, Gonzalez JM, Han C, Kiss H *et al.* (2009). Assembling the marine metagenome, one cell at a time. *Plos One* **4**: e5299.
- Wu JF, Luther GW. (1995). Complexation of Fe(III) by natural organic-ligands in the Northwest Atlantic-Ocean by a competitive ligand equilibration method and a kinetic approach. *Marine Chem* **50**: 159–177.
- Wu JF, Sunda W, Boyle EA, Karl DM. (2000). Phosphate depletion in the western North Atlantic Ocean. *Science* **289**: 759–762.
- Zinser ER, Coe A, Johnson ZI, Martiny AC, Fuller NJ, Scanlan DJ *et al.* (2006). *Prochlorococcus* ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method. *Appl Environ Microbiol* **72**: 723–732.
- Zinser ER, Johnson ZI, Coe A, Karaca E, Veneziano D, Chisholm SW. (2007). Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. *Limnol Oceanograph* **52**: 2205–2220.
- Zinser ER, Lindell D, Johnson ZI, Futschik ME, Steglich C, Coleman ML *et al.* (2009). Choreography of the Transcriptome, Photophysiology, and Cell Cycle of a Minimal Photoautotroph, *Prochlorococcus*. *Plos One* **4**: e5135.
- Zwirgmaier K, Jardillier L, Ostrowski M, Mazard S, Garczarek L, Vault D *et al.* (2008). Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* **10**: 147–161.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)