The ISME Journal (2009) 3, 93–105
© 2009 International Society for Microbial Ecology All rights reserved 1751-7362/09 $32.00
www.nature.com/ismej

ORIGINAL ARTICLE

Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea

Sarah M Sowell[1], Larry J Wilhelm[2], Angela D Norbeck[3], Mary S Lipton[3], Carrie D Nicora[3], Douglas F Barofsky[4], Craig A Carlson[5], Richard D Smith[3] and Stephen J Giovanonni[2]
[1]*Molecular and Cellular Biology Program, Oregon State University, Corvallis, OR, USA;* [2]*Department of Microbiology, Oregon State University, Corvallis, OR, USA;* [3]*Pacific Northwest National Laboratory, Richland, WA, USA;* [4]*Department of Chemistry, Oregon State University, Corvallis, OR, USA and* [5]*Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, CA, USA*

The northwestern Sargasso Sea undergoes annual cycles of productivity with increased production in spring corresponding to periods of upwelling, and oligotrophy in summer and autumn, when the water column becomes highly stratified. The biological productivity of this region is reduced during stratified periods as a result of low concentrations of phosphorus and nitrogen in the euphotic zone. To better understand the mechanisms of microbial survival in this oligotrophic environment, we used capillary liquid chromatography (LC)-tandem mass spectrometry to detect microbial proteins in surface samples collected in September 2005. A total of 2215 peptides that mapped to 236 SAR11 proteins, 1911 peptides that mapped to 402 *Prochlorococcus* proteins and 2407 peptides that mapped to 404 *Synechococcus* proteins were detected. Mass spectra from SAR11 periplasmic substrate-binding proteins accounted for a disproportionately large fraction of the peptides detected, consistent with observations that these extremely small cells devote a large proportion of their volume to periplasm. Abundances were highest for periplasmic substrate-binding proteins for phosphate, amino acids, phosphonate, sugars and spermidine. Proteins implicated in the prevention of oxidative damage and protein refolding were also abundant. Our findings support the view that competition for multiple nutrients in oligotrophic systems is extreme, but nutrient flux is sufficient to sustain microbial community activity.
The ISME Journal (2009) **3,** 93–105; doi:10.1038/ismej.2008.83; published online 4 September 2008
Subject Category: integrated genomics and post-genomics approaches in microbial ecology
Keywords: metaproteomics; oligotrophy; SAR11; Sargasso Sea; transport

Introduction

Metaproteomics is a new approach that enables the direct observation of proteins expressed by mixed microbial assemblages. Mass spectrometry coupled with an ever increasing bank of genomic and metagenomic sequences can be used effectively for protein identification, thus providing insight into microbial community functions. To date, proteomic analyses have been used to identify proteorhodopsin (Giovannoni *et al.*, 2005a), an oxidoreductase, a predicted aminopeptidase and a conserved hypothetical protein (Kan *et al.*, 2005) in marine environments, but a comprehensive analysis of a marine metaproteome has yet to be performed. Comprehensive metaproteomic analyses have been applied to other microbial communities including the natural microbial biofilm of an acid mine drainage site (Ram *et al.*, 2005; Lo *et al.*, 2007), soil (Schulze *et al.*, 2005), activated sludge (Ehlers and Cloette, 1999; Wilmes and Bond, 2004) and waste water (Lacerda *et al.*, 2007).

The Sargasso Sea is an oligotrophic subtropical gyre where key nutrients, particularly nitrogen and phosphorus, are often depleted. A shortage of nutrients can limit microbial productivity (Cotner *et al.*, 1997; Rivkin and Anderson, 1997) and affect community composition (Lomas *et al.*, 2004). Levels of dissolved inorganic phosphorus and dissolved nitrate + nitrite in the Sargasso Sea are often below concentrations of 5 and 10 nM, respectively (Wu *et al.*, 2000; Lipschultz, 2001; Steinberg *et al.*, 2001). At the Bermuda Atlantic Time-series Study site, prokaryotes have a major function in nutrient cycling and can reach densities of 5–10 $\times$ 10^5 cells

Correspondence: SJ Giovanonni, Department of Microbiology, Oregon State University, 220 Nash Hall, Corvallis, OR 97331, USA.
E-mail: steve.giovannoni@oregonstate.edu
Received 26 May 2008; revised 4 August 2008; accepted 4 August 2008; published online 4 September 2008

per ml in the photic zone during summer (Carlson *et al.*, 1996). For an organism to survive in such oligotrophic conditions, it has been suggested that efficient competition for nutrients that are at low levels in the environment would require the expression of high affinity, multisubstrate transporters (Hirsch *et al.*, 1979).

The SAR11 clade is a group of alpha-proteobacteria that is abundant throughout the oceans and always present in the Sargasso Sea. SAR11 bacteria are comprised of several ecotypes that vary temporally and over depth (Morris *et al.*, 2005; Carlson *et al.*, unpublished), and reach their greatest abundance in surface waters during the summer where 35% of all prokaryotic cells and 18% of the biomass are SAR11 cells (Morris *et al.*, 2002). This clade is highly active, accounting for 50% of amino-acid assimilation and 30% of dimethylsulfoniopropionate assimilation among microbes in one tracer study (Malmstrom *et al.*, 2004), and is thought to have a significant function in nutrient cycling in the oceans.

Marine *Prochlorococcus* and *Synechococcus* are the dominant oxygenic phototrophs in subtropical oceans, accounting for up to 15% and 5% of phytoplankton cell counts, respectively (Sieracki *et al.*, 1995; DuRand *et al.*, 2001), and the majority of oceanic primary production. In the Sargasso Sea, *Prochlorococcus* cell numbers peak during the summer at depths between 80 and 200 m, whereas *Synechococcus* cells reach their maximum density during spring blooms in surface waters. *Synechococcus* cells are slightly larger (0.62–1.56 μm^3 compared to 0.077–0.22 μm^3 for *Prochlorococcus* (Heldal *et al.*, 2003)), and thus account for a higher proportion of primary producer biomass than cell numbers would suggest.

The goal of this study was to identify proteins expressed by SAR11 in the Sargasso Sea during a season when nutrients were highly depleted. We employed the method of Wilhelm *et al.* (2007) to build a database of SAR11 environmental protein-coding sequences (eCDSs) from the Sargasso Sea, using the genome of a Sargasso Sea SAR11 isolate, HTCC7211 (Stingl *et al.*, unpublished), as a query. Peptides detected in microbial samples collected from the Sargasso Sea were then mapped to the SAR11 environmental sequences. To test the specificity of our approach, we repeated the analysis using the high-light adapted *Prochlorococcus marinus* str. MIT 9312 and marine *Synechococcus* sp. strain CC9605.

Materials and methods

Database construction

The SAR11 eCDS database used in this study was constructed by binning SAR11 genomic DNA fragments from the Sargasso Sea (Venter *et al.*, 2004) with the reciprocal best-hit procedure that is fully described by Wilhelm *et al.* (2007). eCDSs that were similar to the protein-coding sequences of a query genome were collected using TBLASTN with an expect score cutoff of 1×10^{-20} and complexity filtering turned off. Each hit returned from this search was then translated and queried using BLASTP with default parameters against the bacterial portion of the National Center for Biotechnology Information nonredundant proteins (NCBI-nr) database. If the best hit in the second search was the original query gene, that is, if the eCDS had higher sequence similarity to the query gene than any other known gene, it was scored as a best hit. The entire nonredundant collection of such sequences obtained using the genomes of SAR11 strains HTCC7211 and HTCC1062 as queries (Stingl *et al.*, unpublished; Giovannoni *et al.*, 2005b) were included in this study, as well as the sequences from the query genomes. The total number of sequences in this database is 257836.

The *Prochlorococcus* environmental sequence (*Prochlorococcus* eCDS) and *Synechococcus* environmental sequence (*Synechococcus* eCDS) databases were created in an identical manner using *Prochlorococcus marinus* str. MIT 9312 and *Synechococcus* sp. strain CC9605 as the query genomes. There are 21477 sequences in the *Prochlorococcus* eCDS database and 23347 sequences in the *Synechococcus* eCDS database.

Sample collection

Microbial cells were collected from hydrostation S off the Bermuda coast (32° 10′ N, 64° 30′ W) on 19 and 20 September 2005. Cells were concentrated from 230 l (19 September) or 240 l (20 September) of surface water (5 m) by tangential flow filtration using tandem Millipore Pellicon systems with 30 kDa regenerated cellulose filters at a rate of $\sim 1\ l^{-1}\ min^{-1}$ per system. A Beckman J2-21 centrifuge with a JA-20 rotor was used at 48 400 g for 1 h at 4°C to pellet the concentrated cells. Each of the resulting pellets was split into two equal samples (four samples total) and stored at −80°C until analysis.

Reagents

All reagents were obtained from Sigma Aldrich (St Louis, MO, USA) unless otherwise specified. Nanopure or Milli-Q quality water ($\sim$18 MΩ cm or better), ammonium bicarbonate (NH_4HCO_3), bicinchoninic acid (Pierce, Rockford, IL, USA), urea, thiourea, tetrafluoroethylene, dithiothreitol, 3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate (CHAPS) sequencing grade modified trypsin (Promega, Madison, WI, USA), HPLC grade methanol, trifluoroacetic acid, acetonitrile, ammonium formate, formic acid and ammonium hydroxide (NH_4OH).

Sample preparation

Each of the four sample pellets was divided into two equal subsamples; one for a global preparation and one for a soluble/insoluble preparation. For two of

the samples (1 and 2), bead beating in NH_4HCO_3 buffer was used to physically lyse the cells. The remaining two samples (3 and 4) were chemically lysed using tetrafluoroethylene.

Global and soluble/insoluble protein preparation with bead beating

Two sample pellets (1 and 2) were prepared as outlined by Adkins *et al*. (2006). Briefly, pellets were resuspended in 100 mM NH_4HCO_3 buffer and lysed by bead beating. Proteins from subsamples designated for global analysis were denatured, reduced with dithiothreitol and digested using trypsin. Subsamples for soluble/insoluble analysis were ultracentrifuged and the supernatant was treated as above, whereas the pellet was resuspended in denaturing buffer containing 1% CHAPS to solubilize hydrophobic proteins before enzymatic digestion.

Global and soluble/insoluble protein preparation with tetrafluoroethylene

Samples (3 and 4) were prepared as described in the previous section, except that lyses was achieved chemically by the addition of tetrafluoroethylene to the resuspended pellet at a concentration of 50% (v:v) followed by sonication for one minute in an ice bath and incubation at 60°C for 2 h with gentle shaking (300 r.p.m.).

Sample cleanup

Peptides were cleaned using either a 1 ml/100 mg C18 solid phase extraction (SPE) column (samples 1 and 2) or a strong cation exchange (SCX) SPE column (samples 3 and 4) following the manufacturer's (Supelco/Sigma Aldrich, St Louis, MO, USA) instructions. Peptides were eluted with 80:20 acetonitrile: H_2O, 0.1% trifluoroacetic acid (C18 columns) or 80:15:5 methanol:H_2O:NH_4OH (SCX columns), and were concentrated in a speed-vac to a volume of 50–100 μl. Final peptide concentration was determined by bicinchoninic acid protein assay and samples were quick frozen in liquid nitrogen and stored at −80 °C until needed for analysis.

SCX sample fractionation

Strong cation exchange fractionation following the methods outlined by Adkins *et al*. (2002, 2006) was performed only on sample 2 using a Polysulfoethyl 2.1 × 35 mm, 3 μm particle size column (PolyLC, Columbia, MD, USA) on an Agilent 1100 series high performance liquid chromatography system. Mobile phases consisted of 10 mM ammonium formate (pH 3.0)/25% acetonitrile (A) and 500 mM ammonium formate (pH 6.8)/25% acetonitrile (B) at a flow rate of 200 μl min^{-1}. Peptides were separated using a gradient from 0% to 50% B over 40 min, followed by a gradient of 50%–100% B over 10 min. The mobile phase was then held at 100% B for 10 min. A total of 23 fractions were collected and eight were chosen from the SCX chromatogram for analysis. Each fraction was dried under vacuum and redissolved in 30 μl of 25 mM NH_4HCO_3, and 10 μl of each fraction were analyzed by capillary LC-MS/MS.

Sample loading and capillary LC separation

The online capillary liquid chromatography (LC) setup with an ion trap mass spectrometry (MS) has been described elsewhere (Sowell *et al*., 2008). Briefly, peptide samples were loaded onto a 60 cm hand-packed C18 column before electrospray into a Thermo Fisher Scientific LTQ linear ion trap mass spectrometer set to perform tandem MS (MS/MS) on the top 10 ions using data-dependent settings and a dynamic exclusion window of 1 min. The aqueous to organic (mobile phase A: 0.2% acetic acid and 0.05% trifluoroacetic acid in water, and mobile phase B: 0.1% trifluoroacetic acid in 90% acetonitrile/10% water) gradient flowed exponentially over 100 min.

Peptide identifications

A total of 247 LC-MS/MS analyses were performed to produce MS/MS spectra. Each data set was searched, using the SEQUEST algorithm (Eng *et al*., 1994), against the protein databases mentioned above for peptide/protein identification. A standard parameter file with no modifications to amino-acid residues and a mass error window of 3 *m/z* units for precursor mass and 0 *m/z* units for fragmentation mass was used. The searches were allowed for all possible peptide termini, that is, not limited to tryptic only termini. Peptide identifications were considered acceptable if they passed the thresholds determined by Washburn *et al*. (2001) and the additional filter of having a Peptide Prophet score of at least 0.5 (Keller *et al*., 2002).

Phylogenetic tree construction

A neighbor-joining phylogenetic tree of nearly full length SAR11 PstS eCDS amino-acid sequences was generated in the sequence environment ARB (Ludwig *et al*., 2004) using the Dayhoff PAM model (Dayhoff, 1978). The *Synechococcus* sequence of PstS (gi|15618) was used as the outgroup. Short sequences were inserted using the ARB parsimony insertion tool (Ludwig *et al*., 1998).

Results

A total of 236 SAR11 proteins were identified from 2215 unique peptides that collectively matched 13040 spectra (Supplementary Table S1). The criterion for protein identification was that at least one of its peptides be identified in three or more spectra or that each of at least two of its peptides be identified in one or more spectra. These peptide

sequences typically matched multiple SAR11 eCDSs and therefore were mapped to 9380 SAR11 eCDSs from the Sargasso Sea.

The most frequently detected proteins included periplasmic transport proteins for substrates such as phosphate, amino acids, phosphonate and spermidine/putrescine (Table 1). The most frequently detected proteins with functions other than nutrient transport included glutamine synthetase, the chaperones DnaK and GroEL and F_0F_1-ATP synthase subunits (Table 2).

Using the *Prochlorococcus* eCDS database, 1911 unique peptides from 6270 spectra that mapped to 402 proteins and 1681 eCDSs were detected (Sup-

Table 1 SAR11 proteins involved in nutrient uptake detected by capillary LC MS/MS in Sargasso Sea surface water in September 2005

ID[a]	*Product*	*Gene*	*Number of spectra*[b]	*Number of peptides*[c]	*Number of eCDS*[d]
PB7211_1190	Phosphate ABC transporter	*pstS*	2949	238	169
PB7211_1204	ABC transporter, amino acids	*yhdW*	1479	193	252
PB7211_298	TRAP dicarboxylate transporter—DctP subunit	*dctP*	697	87	99
PB7211_926	Phosphonate ABC transporter, periplasmic phosphonate-binding protein	*phnD*	532	44	14
PB7211_697	Spermidine/putrescine-binding periplasmic protein	*potD*	497	104	204
PB7211_687	Substrate-binding region of ABC type glycine betaine transport system	*opuAC*	367	56	117
PB7211_576	Na+/solute symporter (Ssf family)	*yjcG*	361	44	286
PB7211_1324	Leu/Ile/Val-binding protein precursor	*livJ2*	360	53	249
PB7211_890	Hypothetical protein		197	36	33
PB7211_601	Taurine transport system periplasmic protein	*tauA*	184	39	142
PB7211_925	Extracellular solute-binding protein		170	55	79
PB7211_463	Hypothetical protein		159	41	78
PB7211_130	Hypothetical protein	*livJ*	148	40	225
YP_266190	Probable binding protein component of ABC sugar transporter		84	22	41
PB7211_194	Putative glycine betaine-binding lipoprotein		74	17	11
PB7211_965	Putative tricarboxylic transport TctC	*tctC*	65	22	133
PB7211_1089	Putative extracellular solute-binding protein, amino acids		61	16	19
PB7211_522	Hypothetical	*yiaO*	60	13	5
PB7211_972	Bacterial extracellular solute-binding protein		39	17	60
PB7211_620	Phosphonate ABC transporter, ATP-binding protein	*phoC*	38	5	5
PB7211_781	Related to peptide ABC-transporter, periplasmic substrate binding protein		30	9	14
YP_266078	Probable Leu/Ile/Val-binding protein	*braC*	28	8	4
PB7211_689	ABC transporter, amino acids	*occT*	26	11	151
PB7211_754	Hypothetical amino-acid transporter	*proX*	22	10	55
PB7211_412	Phosphate ABC transporter	*pstB*	20	7	82
PB7211_866	Potassium uptake protein	*trkA*	18	1	1
PB7211_644	Substrate-binding region of ABC-type glycine betaine transport system		17	9	18
PB7211_1408	Probable taurine uptake ABC transporter permease protein	*tauC*	13	1	2
PB7211_704	Cyclohexadienyl dehydratase	*pheC*	11	9	62
PB7211_679	Phosphate transport regulon regulator	*phoU*	10	7	39
PB7211_347	Hypothetical transport protein	*yrbC*	10	4	10
PB7211_402	Sodium:solute transporter family		8	1	1
PB7211_647	ABC sugar transporter, nucleotide binding/ATPase protein		7	2	25
PB7211_911	ABC transporter related	*ugpC*	7	2	15
PB7211_586	ABC transporter, phosphate	*pstA*	5	4	13
PB7211_1146	Xanthine/uracil/vitamin C permease family protein		5	2	16
YP_266193	Sugar ABC transporter, ATP-binding protein	*malK*	5	1	7
PB7211_1290	TRAP dicarboxylate transporter, DctP subunit		4	2	2
YP_266698	TRAP type bacterial extracellular solute-binding protein, family 7		4	2	7
PB7211_974	Ammonium transporter	*amtB*	3	3	39
PB7211_934	Magnesium transport protein	*mgtE*	3	2	30
PB7211_1074	Ferric uptake regulation protein	*fur*	3	1	58
PB7211_760	Potassium uptake protein	*trkH*	3	1	1
PB7211_725	Phosphonate ABC transporter, permease protein	*phnE*	3	1	5
PB7211_1080	Phosphate regulon	*phoR*	3	1	1
PB7211_565	Possible AbrB protein	*abrB*	3	1	1

Abbreviation: eCDSs, environmental protein-coding sequences.
[a]Sequences for IDs beginning with PB7211 can be found at https://moore.jcvi.org/moore/
[b]The number of mass spectra that matched to peptides of given SAR11 strain HTCC7211 protein and its homologous eCDSs.
[c]The number of unique peptide sequences from SAR11 strain HTCC7211 proteins and their eCDSs that were detected.
[d]The number of eCDSs for each protein that contained the detected peptide sequence(s).

Table 2 Proportion of transporters detected in a variety of Gram-negative bacteria using techniques similar to those employed in this study

Dataset type	*Bacterium*	*Number of proteins*			*Number of spectra*			*Reference*
		Transporters	*All*	*%*	*Transporters*	*All*	*%*	
Proteomic library	*Candidatus* Pelagibacter ubique	76	1170	6.50	48 200	168 398	28.62	Sowell *et al.* (2008)
	Rhodobacter sphaeroides	440	4251	10.35	102 153	943 125	10.83	Callister *et al.* (2006)
	Salmonella typhimurium	428	4446	9.63	581 664	8 242 534	7.06	Adkins *et al.* (2006)
	Yersinia pestis	347	4382	7.92	11 677	237 480	4.92	Hixson *et al.* (2006)
	Shewanella oneidensis	215	4198	5.12	567 440	14 402 503	3.94	Elias *et al.* (2008)
	Geobacter sulfurreducens	149	3449	4.32	118 655	3 180 336	3.73	Ding *et al.* (2006)
Culture comparison	*Candidatus Pelagibacter ubique*_exponential	42	571	7.36	5476	15 783	34.70	Sowell *et al.* (2008)
	Candidatus Pelagibacter ubique_stationary	42	657	6.39	7707	25 646	30.05	Sowell *et al.* (2008)
Environmental	SAR11	41	236	17.37	8755	13 040	67.14	This study
	Synechococcus sp. CC9605	12	404	2.97	3018	12 092	24.96	This study
	Prochlorococcus marinus str. MIT 9312	16	402	3.98	2019	6270	32.20	This study

plementary Table S2). For *Synechococcus,* 2407 unique peptides from 12092 spectra mapping to 404 proteins and 2632 eCDSs were identified. (Supplementary Table S3). Frequently detected *Prochlorococcus* and *Synechococcus* proteins were similar to those of SAR11 including transport proteins (substrate specificity for urea and phosphate), glutamine synthetase, and chaperones (DnaK, Cpn60 and GroEL). Proteins involved in CO_2 fixation and photosynthesis were also prevalent.

To determine the specificity of our results, peptide hits to the SAR11, *Prochlorococcus* and *Synechococcus* eCDS databases were compared to each other and to the remainder of the Sargasso Sea database (Sargasso Sea eCDSs minus SAR11, *Prochlorococcus* and *Synechococcus* eCDSs) using BLASTP. Surprisingly, only 177 spectra corresponding to 24 peptides from 11 proteins were common between the SAR11 and the cyanobacterial eCDS databases (Figure 1a). As would be expected from their more recent evolutionary divergence, there were more peptides (1062) in common between the *Prochlororoccus* and *Synechococcus* eCDS databases. Slightly more than half of the peptides (1226 peptides (55%)) from the SAR11 data set matched eCDSs from the remainder of the Sargasso Sea metagenomic data (Figure 1b). In the cases of *Prochlorococcus* and *Synechococcus,* only 41% (784 peptides) and 36% (860 peptides) of the peptides, respectively, matched eCDSs from the remainder of the Sargasso Sea database. Even after the removal of all peptides that were not unique to each data set, 84% of the detected SAR11 proteins (49% of eCDSs), 71% of the detected *Prochlorococcus* proteins (83% of eCDSs) and 76% of the detected *Synechococcus* proteins (36% of eCDSs) were hit by at least one unique peptide. The distribution of the peptides between the databases

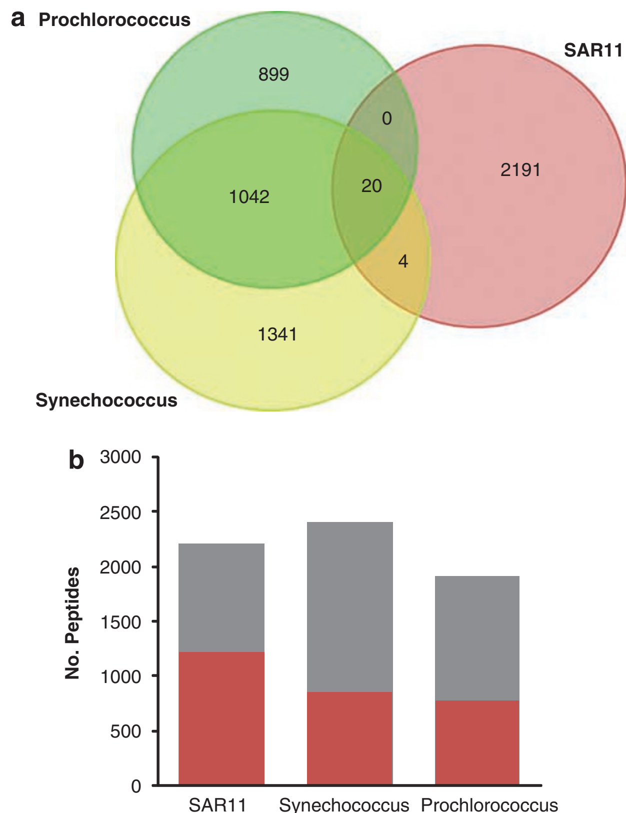

Figure 1 A large proportion of the detected peptides were unique to either SAR11, *Prochlorococcus* or *Synechococcus* (**a**). SAR11, *Prochlorococcus* and *Synechococcus* peptides detected in a Sargasso Sea protein sample and their distribution between the SAR11 eCDS, the *Prochlorococcus* eCDS and the *Synechococcus* eCDS databases. (**b**) Total number of peptides detected from each database and the fraction (red) that was also seen in the remainder of the Sargasso Sea eCDS database (Sargasso Sea eCDS database−(SAR11 eCDS database + *Prochlorococcus* eCDS database + *Synechococcus* eCDS database)) determined using BLASTP with a requirement of 100% identity over the length of the query peptide. eCDS,environmental protein-coding sequence.

can be found in Figure 1 and Supplementary Tables S1, S2 and S3.

Discussion

The prevalence of periplasmic substrate-binding proteins

Our most striking finding is the very high proportion of spectra matching SAR11 transport proteins (Table 3). In particular, spectra for periplasmic substrate-binding subunits of ABC transport systems were frequently observed (Table 1). Spectra corresponding to peptides that matched SAR11 eCDSs associated with transport functions accounted for 67% of the total spectra for SAR11. Using similar methods with cultured SAR11 cells, 28%–35% of all spectra matched transport proteins (Sowell *et al*., 2008), whereas when the same approach was applied to other Gram-negative bacteria, only 4%–11% of spectra matched to transport proteins (Adkins *et al*., 2006; Callister *et al*., 2006; Ding *et al*., 2006; Hixson *et al*., 2006; Elias *et al*., 2008).

The prevalence of SAR11 periplasmic substrate-binding proteins in the Sargasso Sea metaproteome was not unanticipated. SAR11 cells are among the smallest known free-living bacteria (Rappe *et al*., 2002) and cryo-electron tomography images have shown that 25%–35% of cell volume is devoted to periplasmic space (Nicastro *et al*., 2006). As proposed by Hirsch *et al*. (1979) and later developed by Button (1993), optimal metabolic designs for survival in oligotrophic environments would likely devote a large proportion of resources to the expression of high affinity uptake systems that target multiple substrates. Button and Robertson (2000) also suggested that the surface area-to-volume ratio of a cell is determined by equilibrium between the cytoplasmic volume and the membrane's capacity to import nutrients for metabolism. According to this model, small cells with large membranes and dilute cytoplasms are better able to compete in oligotrophic environments, because their capacity to metabolize available substrates is matched to the membrane's ability to import them.

Two of the most abundant proteins we observed were periplasmic substrate-binding proteins involved in phosphorus acquisition, suggesting that these cells were maximizing phosphorus transport to cope with the low concentrations found in Sargasso Sea surface waters in the late summer (Wu *et al*., 2000; Steinberg *et al*., 2001). The periplasmic substrate-binding protein involved in the uptake of phosphate (PstS) had the highest spectral count (Table 2 and Supplementary Table S1), but peptides from the ATPase and permease subunits (PstB and PstA) as well as from the proteins involved in the regulation of the *pst* operon (PhoU, PhoR and PhoB) were also detected (Figure 2). Expression of PstS and other members of the *pst* operon is known to be induced in many bacteria when phosphorus is depleted in the medium (Scanlan *et al*., 1997; Dyhrman and Haley, 2006; Martiny *et al*., 2006). Concentrations of phosphorus are generally $<5\,\text{nM}$ during summer stratified conditions in the Sargasso Sea (Wu *et al*., 2000; Mills *et al*., 2008), where SAR11 cells have been reported to contribute substantially to inorganic phosphate uptake (Zubkov *et al*., 2007), indicating that they can effectively compete for inorganic nutrients at extremely low concentrations.

Figure 3 shows that multiple variants of the SAR11 PstS protein were detected, with as many as six amino-acid substitutions observed at one sequence position. Marine bacteria often have highly variable gene sequences in nature. Much of this appears to be neutral variation (Wilhelm *et al*., 2007), but some may be associated with the evolutionary divergence of ecotypes. The SAR11 clade consists of multiple ecotypes that vary in abundance with depth and season (Field *et al*., 1997; Morris *et al*., 2002, 2005; Carlson *et al*., unpublished). The data in Figure 3 show that some subclades of the SAR11 *pstS* genes found in the Sargasso Sea metagnomic data were highly translated in summer surface samples from the Sargasso Sea, but others were not. Notably, the PstS subclade that includes Sargasso Sea SAR11 strain HTCC7211 was highly detected, but the subclade of PstS proteins that included coastal strain HTCC1062 was infrequently detected.

The periplasmic substrate-binding protein for phosphonate acquisition (PhnD) was also among the top frequently detected SAR11 proteins, and the membrane-spanning ATPase (PhnE) for this transport system was also observed (Table 2 and Supplementary Table S1). Phosphonates have recently been recognized as a potential source of phosphorus in oligotrophic systems (Kolowith *et al*., 2001). These compounds, which contain a C–P bond, are particularly stable and are thought to mainly originate from eukaryotic lipids. Genes for phosphonate metabolism were not found in coastal SAR11 isolates, but appear in the Sargasso Sea isolate, HTCC 7211. Pho proteins that regulate the *pst* operon also regulate the *phn* operon in *Escherichia coli* and have been shown to be induced by phosphorus limitation (Wanner, 1996).

Our findings provide the first direct evidence of SAR11 cells in open ocean gyres devoting significant resources to the acquisition of phosphorus from phosphonates. From Figure 2, it is apparent that while proteins involved in phosphonate transport were frequently detected, the cytoplasmic proteins involved in the degradation of phosphonate were not, even though the genetic potential for expression of these proteins was present in the environment. This observation supports the theory of Button and Robertson and suggests that phosphonate processing enzymes are rare, relative to transporters, because the total flux of phosphonates through the pathway never requires more than one or a few of each

Table 3 Detected *Prochlorococcus* and *Synechococcus* proteins involved in photosynthesis, CO_2 fixation,and nitrogen utilization

Product name	*Common peptides*	*Prochlorococcus*		*Synechococcus*	
		Unique peptides	*Accession*	*Unique peptides*	*Accession*
Proteins involved in photosynthesis					
Plastocyanin	4	1	YP_397078	1	YP_381328
Photosystem I subunit VII	3		YP_398197		YP_380458
Photosystem II reaction center Psb27 protein-like	2		YP_397005		YP_381021
Photosystem II reaction center Psb28 protein-like	2	1	YP_397370	1	YP_381507
Putative chaperon-like protein for quinone binding in photosystem II	2		YP_397745		YP_382282
Photosystem II 44 kDa subunit reaction center protein	2	1	YP_397752	2	YP_382289
Photosystem q(b) protein	1	1	YP_396723	5	YP_380638 YP_380848 YP_381355
Carotene 7,8-desaturase		4	YP_396644		
Light-independent protochlorophyllide reductase subunit N	2	3	YP_397042		YP_381071
Carotene 7,8-desaturase		3	YP_396616		
Protochlorophyllide oxidoreductase		3	YP_397039		
Photosystem I PsaF protein (subunit III)		2	YP_396966		
Photosystem I assembly protein Ycf3	1	1	YP_396632		YP_382669
C-phycoerythrin class I β-chain				34	YP_380757
C-phycoerythrin class II β-chain				33	YP_380766
C-phycoerythrin class I α-chain				32	YP_380758
C-phycoerythrin class II α-chain				31	YP_380765
Phycocyanin β-subunit				11	YP_380752
Photosystem II complex extrinsic protein precursor, PsuB				9	YP_382621
Phycocyanin, α-subunit				8	YP_380751
Photosystem II manganese-stabilizing polypeptide				6	YP_380630
Allophycocyanin α-chain				6	YP_382491
Allophycocyanin, β-subunit				5	YP_382492
Photosystem II chlorophyll-binding protein CP47	1		YP_396818	4	YP_380799
Photosystem I core protein PsaA				4	YP_380662
Phycobilisome linker polypeptide				4	YP_380774 YP_380775
High light inducible protein-like	1		YP_396594	3	YP_382859 YP_382610
Photosystem I reaction center subunit II (PsaD)	5		YP_398167	3	YP_380730
Photosystem I reaction center subunit III (PsaF)				3	YP_380963
Photosystem II D2 protein (photosystem q(a) protein)	1		YP_397751	2	YP_382288
Photosystem II D2 protein (photosystem q(a) protein)	1		YP_396672	2	YP_382662
Photosystem I core protein PsaB				2	YP_380663
Phycobilisome core component-allophycocyanin β-18 subunit				2	YP_381516
Photosystem II reaction center protein PsbH				1	YP_380594
CO_2 fixation					
Carboxysome shell protein CsoS1	31	1	YP_397046	3	YP_381075 YP_381082
Ribulose-bisphosphate carboxylase	1		YP_397048		YP_381077
Carboxysome shell protein CsoS3		2	YP_397050		
Ribulose bisophosphate carboxylase				25	YP_381076
Putative carboxysome shell polypeptide CsoS3				2	YP_381079
Putative carboxysome structural peptide CsoS2				2	YP_381078
Nitrogen acquisition					
Putative urea ABC transporter, substrate-binding protein	86	5	YP_397325	28	YP_382900
Glutamine synthetase, type I	55	3	YP_397376	10	YP_381515
ABC transporter, substrate-binding protein, possibly oligopeptides		4	YP_397555		
Ammonium transporter		2	YP_396762		
Cyanate hydratase				11	YP_382939
Nitrogen regulatory protein P-II (GlnB, GlnK)	11		YP_398052	8	YP_382500
Urease, α-subunit				4	YP_382907
Formate and nitrite transporters				3	YP_382937
Putative ATP-binding subunit of urea ABC transport system	4		YP_397322	2	YP_382897
Ferredoxin–nitrite reductase				2	YP_382936
Nitrate transporter				1	YP_382922
Urease, γ-subunit				1	YP_382905

When *Prochlorococcus* ortholog is more prevalent, the protein is highlighted in light gray. When the *Synechococcus* ortholog is more prevalent, the protein is highlighted in dark gray.

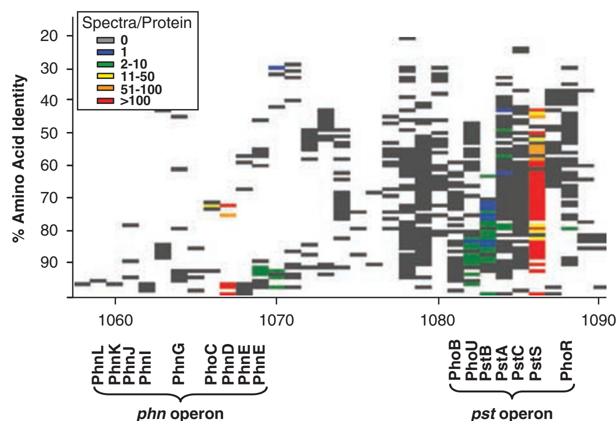

Figure 2 Proteins involved in phosphate and phosphonate uptake were among the most frequently detected in a mass spectrometric (MS/MS) analysis of Sargasso Sea surface water. The figure shows spectral matches to eCDSs in the region of the HTCC7211 genome from gene 1058 to gene 1090, which includes genes for phosphate (*pst*) and phosphonate (*phn*) transport and metabolism. The SAR11 eCDSs are arranged vertically by their percent amino-acid identity to the homologous HTCC7211 gene (Wilhelm *et al.*, 2007) and are heat mapped according to their MS/MS spectral count. Periplasmic transporters (PstS and PhnD) were the most highly detected subunits, but the ATPase and permease subunits (PstA, PstB and PstE) and regulatory proteins (PhoU, PhoR and PhoB) were also frequently observed. eCDSs,environmental protein-coding sequences.

phosphonate processing enzyme per cell and expressing fewer cytoplasmic proteins allows the cells to maintain a large surface area-to-volume ratio.

We found evidence that SAR11 cells express transporters involved in the acquisition of organic compounds that contain reduced sulfur, such as methionine and dimethylsulfoniopropionate (DMSP). Tripp *et al.* (2008) reported that the SAR11 clade lacks the genes for assimilatory sulfate reduction and therefore relies solely on organic molecules for sulfur. While there are no known transporters that specifically import methionine or DMSP, it has been predicted that the high-affinity amino-acid transporter, YhdWXYZ, could be involved in methionine uptake, and the OpuA uptake system, which is annotated as an ABC-type glycine betaine transporter, has also been shown to import DMSP (Holtmann and Bremer, 2004). The observation of multiple peptides from the periplasmic components of each of these transport systems (YhdW and OpuAC), but not sulfate transport proteins, is consistent with the finding that SAR11 relies on reduced sulfur compounds as a source of sulfur.

Transporters for other organic compounds, such as amino acids and sugars, were also frequently detected. DctP is the periplasmic-binding subunit of a tripartite ATP-independent periplasmic (TRAP) dicarboxylate transporter for mannitol and/or chloroaromatic compounds. This high-affinity transporter relies on proton motive force to drive uptake (Forward *et al.*, 1997). The periplasmic substrate-binding protein PotD is involved in spermidine uptake in *E. coli*, but is also reported to import putrescine with a lower affinity. Both of these compounds are nitrogen-containing polyamines that have been detected in many organisms (Igarashi and Kashiwagi, 1999) including diatoms, where they are a component of the silica wall (Sumper *et al.*, 2005). Diatoms are a likely source of polyamines in the Sargasso Sea.

All of the proteins involved in the uptake of ammonium and its incorporation into the amino acids, glutamine and glutamate, were detected (AmtB, GlnA, GlnT and GltB; Table 2), suggesting that the SAR11 population was metabolically poised to assimilate inorganic nitrogen.

The absence of evidence for iron transport functions was conspicuous, suggesting that under the prevailing conditions, this part of the SAR11 metabolic repertoire may not have been highly expressed. Consistent with our observations, Sedwick *et al.* (2005) reported higher levels of iron in Sargasso Sea summer surface waters compared to the spring, presumably from eolian input from the Saharan desert. The periplasmic iron-binding transporter was one of only two periplasmic substrate-binding proteins not observed in the metaproteomic data, the other being a protein of uncertain substrate specificity. A number of different factors (e.g., peptide ionization efficiency or sample preparation biases) might explain the absence of mass spectral data for a given protein. However, it is unlikely that these factors would lead to the absence of peptide-spectra for the iron transport proteins, when so many other transport proteins were detected at such high abundance.

The membrane protease HflKC and the chaperone proteins GroEL, GroES and DnaK were among the most prevalent proteins detected in this study; ATP-dependent proteases FtsH, ClpX and HslU were also detected (Supplementary Table S1). The detection of these chaperone and protease proteins suggests that protein refolding and proteolysis may be integral to bacterial survival in ocean surface water, because proteins are continually being damaged as a result of exposure to environmental stresses.

The approach we employed in this study has previously been applied to many different organisms and was chosen here to maximize the number of proteins detected. However, it was not designed to detect integral membrane proteins like proteorhodopsin. In previous work (Giovannoni *et al.*, 2005a), we employed procedures that were optimized for proteorhodopsin detection by (1) focusing on small cells, (2) isolating the membrane fraction, (3) focusing on SDS-polyacrylamide gel electrophoresis gel sections that contained proteins of the desired size and (4) solubilizing the protein with dodecylmaltoside, a detergent that was known to be compatible with proteorhodopsin detection. Proteorhodopsin requires specialized methods for detection, because it contains seven transmembrane

Figure 3 Peptide coverage and phylogenetic position of PstS eCDSs. Amino-acid coverage of the PstS eCDSs reveals that some subclades of SAR11 PstS genes represented in the Sargasso Sea metagenomic data were frequently detected in these summer surface samples, and others were not. The neighbor-joining phylogenetic tree was inferred from nearly full-length SAR11 PstS eCDS amino-acid sequences with ARB, using the Dayhoff PAM model. Short sequences (designated with an asterisk (*)) were inserted using the ARB parsimony insertion tool. For each sequence in the tree, the colored bars show which amino-acid positions were detected. Amino acids are heat mapped according to the prevalence of the given residue at that position in the consensus sequence, with red indicating the most common and green indicating the least common amino acid. Gray indicates undetected residues. In some cases, as many as six amino-acid substitutions were observed. eCDSs, environmental protein-coding sequences.

helices and is imbedded in the plasma membrane, making it difficult to solubilize and digest. In addition, few of the tryptic peptides in proteorhodopsin are within the mass range commonly scanned in proteomics experiments. Like most analytical procedures, proteomics is sensitive to

bias, and the absence of detection of proteorhodopsin, which is thought to be abundant in marine environments (Fuhrman *et al.*, 2008), is a good reminder that the lack of detection of a protein is not a conclusive proof of the absence of that protein.

Abundant cyanobacterial proteins

To provide a point of reference for comparison, we used the same method to examine the distribution of *Prochlorococcus* and *Synechococcus* peptides in the mass spectrometric data. *Prochlorococcus* is the dominant phototroph in the Sargasso Sea, reaching maximum cell abundances of $\sim 10^5$ cells per ml in deeper (80–200 m) summer waters (DuRand *et al.*, 2001). *Synechococcus* generally blooms in spring surface water, but can range in abundance from 10^3 to 10^4 cells per ml during stratified periods and can comprise a significant fraction of the picoplankton in the 1-m surface of the Sargasso Sea (DuRand *et al.*, 2001). Approximately twice as many *Synechoccocus* spectra and peptides were detected compared with *Prochlorococcus*, even though *Prochlorococcus* strain MIT9312 (used as the query genome for database construction) is thought to be more abundant that *Synechococcus* sp. in summer surface water in the Sargasso Sea. We attribute this to the larger size of *Synechococcus* cells and natural variation in microbial community composition.

Like SAR11, the most frequently detected cyanobacterial proteins were involved in transport functions (Supplementary Tables S2 and S3). However, the overall number and variety of transporters detected was much lower than in SAR11, suggesting that *Prochlorococcus* and *Synechococcus* devote a lower percentage of their overall energy expenditure to nutrient acquisition.

One of the most abundant transporters detected in both of the cyanobacteria was involved in urea uptake. *Synechococcus* proteins for two urease subunits, which are necessary for urea degradation, were also detected (Collier *et al.*, 1999). This suggests that urea, the most common form of dissolved organic nitrogen in the open ocean (Anita *et al.*, 1991) is an important nitrogen source for *Synechococcus*. Other *Synechococcus* proteins involved in nitrogen metabolism that were detected include nitrate and nitrite transporters, cyanate hydratase, and the nitrogen regulatory protein P-II (Table 3 and Supplementary Table S3). *Prochlorococcus* proteins for urea degradation were not detected even though they are present in the genome, but proteins for ammonium and oligopetide transport were detected (Table 3 and Supplementary Table S2), supporting the previous data suggesting that *Synechococcus* can grow on a wide range of nitrogen sources, whereas high-light adapted *Prochlorococcus* strains preferentially use ammonium and organic sources of nitrogen (Moore *et al.*, 2002; Zubkov *et al.*, 2003).

Both *Prochlorococcus* and *Synechococcus* also expressed proteins known to be induced during phosphate depletion. These included phosphate and phosphonate uptake proteins and the sulfolipid (UDP-sulfoquinovose) biosynthesis protein (Guler *et al.*, 1996; Van Mooy *et al.*, 2006). In addition, a PhoH-like phosphate starvation-inducible protein encoded by *Prochlorococcus,* but not *Synechococcus*, was also detected.

Proteins involved in photosynthesis and carbon fixation were frequently detected (Table 3 and Supplementary Tables S2 and S3). In addition to general photosystem proteins, proteomic evidence of the pigment differences between *Prochlorococcus* and *Synechococcus* were observed with chlorophyll and carotenoid synthesis proteins detected for *Prochlorococcus* and phycobilisome proteins detected for *Synechococcus*. Carboxysome proteins, which are induced by inorganic carbon limitation and are thought to be signaled by high light (McKay *et al.*, 1993; Woodger *et al.*, 2003) were also observed, with peptides for the *Synechococcus* orthologs appearing more frequently. The carboxysome allows cyanobacteria to overcome inorganic carbon limitation by concentrating CO_2 for its conversion to organic carbon by the enzyme ribulose 1, 5 bisphosphate carboxylase/oxygenase (RuBisCO), which was also detected.

The observation of abundant damage-control systems in the cyanobacteria paralleled similar findings in SAR11. However, in addition to chaperones that refold damaged proteins, proteins focused on the prevention of oxidative damage from photosynthesis by-products were also highly represented by peptide spectra (i.e., relatively abundant). In particular, thioredoxin, thioredoxin peroxidase and superoxide dismutase were among the most frequently detected cyanobacterial proteins.

Proteomic analyses of microbial communities

Post-transcriptional mechanisms, such as transcript degradation or inactivation by riboswitches, can lead to differences between transcript levels and actual metabolic activity. Proteomic analyses examine proteins, the final product of all levels of gene expression, and thus provide information that is often different from the information provided by the measurements of gene expression (Gygi *et al.*, 1999). Although our experiments were not designed to detect post-translational modifications, certain modifications that affect protein function can also be detected using proteomic analysis. Genomic, transcriptomic and proteomic analyses are complementary, each contributing to our understanding of how microbial communities respond to their environment. Transcriptomic and proteomic analyses build on genomics by describing different levels of regulation for microbial responses to environmental perturbations. Recent transcriptomic studies have been effective at illuminating the transcriptional level of gene expression by identifying gene

transcripts in marine and freshwater ecosystems, (Poretsky *et al.*, 2005; Frias-Lopez *et al.*, 2008).

Conclusion

The metaproteomes of SAR11, *Prochlorococcus* and *Synechococcus* provided ample evidence of cellular adaptations to an extreme environment in which cells are subjected to damage by light and oxidative stresses, while competing for essential nutrients that are at extremely low concentrations. Multiple chaperones and proteins involved in mediating oxidative stress were found in conjunction with overwhelmingly abundant proteins for nutrient acquisition in SAR11 and proteins for photosynthesis and carbon fixation in *Prochlorococcus* and *Synechococcus*. In the case of SAR11, the very high spectral counts for periplasmic substrate-binding proteins suggests that in conjunction with maintaining a small cell size and large periplasmic space, the cells express abundant transport proteins to maximize nutrient uptake activity. These adaptations are likely the means by which SAR11 organisms have evolved to become dominant prokaryotes in nutrient-depleted environments. Our findings may focus attention on transport systems as the boundary that defines the interaction of cells with their chemical environment.

Acknowledgements

We thank the scientists and staff of the Bermuda Atlantic Time-series Study Site for their assistance and support; specifically, the crew of the Weatherbird II for their help in sample collection and Rachel Parsons for her technical and administrative assistance. We also thank Dr Allen Milligan for sharing his expertize. This work was supported by a Marine Microbiology Initiative Investigator Award from the Gordon and Betty Moore Foundation. Portions of this research were also supported by the US Department of Energy (DOE) Office of Biological and Environmental Research and performed at the Environmental Molecular Science Laboratory, a DOE national scientific user facility located at Pacific Northwest National Laboratory.

References

Adkins JN, Mottaz HM, Norbeck AD, Gustin JK, Rue J, Clauss TR *et al.* (2006). Analysis of the *Salmonella typhimurium* proteome through environmental response toward infectious conditions. *Mol Cell Proteomics* **5**: 1450–1461.

Adkins JN, Varnum SM, Auberry KJ, Moore RJ, Angell NH, Smith RD *et al.* (2002). Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry. *Mol Cell Proteomics* **1**: 947–955.

Anita NJ, Harrison PJ, Oliveira L. (1991). The role of dissolved organic nitrogen in phytoplankton nutrition, cell biology and ecology. *Phycologia* **30**: 1–89.

Button DK. (1993). Nutrient-limited microbial growth kinetics: overview and recent advances. *Antonie Van Leeuwenhoek* **63**: 225–235.

Button DK, Robertson B. (2000). Effect of nutrient kinetics and cytoarchitecture on bacterioplankter size. *Limnol Oceanogr* **45**: 499–505.

Callister SJ, Dominguez MA, Nicora CD, Zeng X, Tavano CL, Kaplan S *et al.* (2006). Application of the accurate mass and time tag approach to the proteome analysis of sub-cellular fractions obtained from *Rhodobacter sphaeroides* 2.4.1. Aerobic and photosynthetic cell cultures. *J Proteome Res* **5**: 1940–1947.

Carlson CA, Ducklow HW, Sleeter TD. (1996). Stocks and dynamics of bacterioplankton in the northwestern Sargasso Sea. *Deep Sea Res II* **43**: 491–515.

Collier JL, Brahamsha B, Palenik B. (1999). The marine cyanobacterium *Synechococcus* sp. WH7805 requires urease (urea amidohydrolase, EC 3.5.1.5) to utilize urea as a nitrogen source: molecular-genetic and biochemical analysis of the enzyme. *Microbiology* **145**: 447–459.

Cotner JB, Ammerman JW, Peele ER, Bentzen E. (1997). Phosphorus-limited bacterioplankton growth in the Sargasso Sea. *Aquat Microb Ecol* **13**: 141–149.

Dayhoff MO. (1978). *Atlas of Protein Sequence and Structure*, Vol. 5, Supplement 3. National Biomedical Research Foundation: Washington, DC. p 348.

Ding YH, Hixson KK, Giometti CS, Stanley A, Esteve-Nunez A, Khare T *et al.* (2006). The proteome of dissimilatory metal-reducing microorganism *Geobacter sulfurreducens* under various growth conditions. *Biochim Biophys Acta* **1764**: 1198–1206.

DuRand MD, Olson RJ, Chisholm SW. (2001). Phytoplankton population dynamics at the Bermuda Atlantic Time-series station in the Sargasso Sea. *Deep Sea Res II* **48**: 1983–2003.

Dyhrman ST, Haley ST. (2006). Phosphorus scavenging in the unicellular marine diazotroph *Crocosphaera watsonii*. *Appl Environ Microbiol* **72**: 1452–1458.

Ehlers MM, Cloette TE. (1999). Comparing the protein profiles of 21 different activated sludge systems after SDS-PAGE. *Wat Res* **33**: 1181–1186.

Elias DA, Tollaksen SL, Kennedy DW, Mottaz HM, Giometti CS, McLean JS *et al.* (2008). The influence of cultivation methods on *Shewanella oneidensis* physiology and proteome expression. *Arch Microbiol* **189**: 313–324.

Eng J, McCormack AL, Yates III JR. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5**: 976–989.

Field KG, Gordon D, Wright T, Rappe M, Urback E, Vergin K *et al.* (1997). Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. *Appl Environ Microbiol* **63**: 63–70.

Forward JA, Behrendt MC, Wyborn NR, Cross R, Kelly DJ. (1997). TRAP transporters: a new family of periplasmic solute transport systems encoded by the dctPQM genes of *Rhodobacter capsulatus* and by homologs in diverse gram-negative bacteria. *J Bacteriol* **179**: 5482–5493.

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW *et al.* (2008). Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci USA* **105**: 3805–3810.

Fuhrman JA, Schwalbach MS, Stingl U. (2008). Proteorhodopsins: an array of physiological roles? *Nat Rev Microbiol* **6**: 488–494.

Giovannoni SJ, Bibbs L, Cho JC, Stapels MD, Desiderio R, Vergin KL *et al.* (2005a). Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438**: 82–85.

Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D *et al.* (2005b). Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**: 1242–1245.

Guler S, Seeliger A, Hartel H, Renger G, Benning C. (1996). A null mutant of *Synechococcus* sp. PCC7942 deficient in the sulfolipid sulfoquinovosyl diacylglycerol. *J Biol Chem* **271**: 7501–7507.

Gygi SP, Rochon Y, Franza BR, Aebersold R. (1999). Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19**: 1720–1730.

Heldal M, Scanlan DJ, Norland S, Thingstad F, Mann NH. (2003). Elemental composition of single cells of various strains of marine *prochlorococcus* and *synechococcus* using X-Ray microanalysis. *Limnol Oceanogr* **48**: 1732–1743.

Hirsch P, Bernhard M, Cohen SS, Ensign JC, Jannasch HW, Koch AL *et al.* (1979). Life under conditions of low nutrient concentrations. In: Shilo M (ed). *Strategies of Microbial Life in Extreme Environments: report of the Dahlem Workshop on Strategy of Life in Extreme Environments, Berlin, 1978, November 20–24* Weinheim Verlag Chemie: New York. p 513.

Hixson KK, Adkins JN, Baker SE, Moore RJ, Chromy BA, Smith RD *et al.* (2006). Biomarker candidate identification in *Yersinia pestis* using organism-wide semiquantitative proteomics. *J Proteome Res* **5**: 3008–3017.

Holtmann G, Bremer E. (2004). Thermoprotection of *Bacillus subtilis* by exogenously provided glycine betaine and structurally related compatible solutes: Involvement of Opu transporters. *J Bacteriol* **186**: 1683–1693.

Igarashi K, Kashiwagi K. (1999). Polyamine transport in bacteria and yeast. *Biochem J* **344**: 633–642.

Kan J, Hanson TE, Ginter JM, Wang K, Chen F. (2005). Metaproteomic analysis of Chesapeake Bay microbial communities. *Saline Systems* **1**: 7.

Keller A, Nesvizhskii AI, Kolker E, Aebersold R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **74**: 5383–5392.

Kolowith LC, Ingall ED, Benner R. (2001). Composition and cycling of marine organic phosphorus. *Limnol Oceanogr* **46**: 309–320.

Lacerda CM, Choe LH, Reardon KF. (2007). Metaproteomic analysis of a bacterial community response to cadmium exposure. *J Proteome Res* **6**: 1145–1152.

Lipschultz F. (2001). A time-series assessment of the nitrogen cycle at BATS. *Deep Sea Res II* **48**: 1897–1924.

Lo I, Denef VJ, Verberkmoes NC, Shah MB, Goltsman D, DiBartolo G *et al.* (2007). Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. *Nature* **446**: 537–541.

Lomas MW, Swain A, Shelton R, Ammerman JW. (2004). Taxonomic variability of phosphorus stress in Sargasso Sea phytoplankton. *Limnol Oceanogr* **49**: 2303–2309.

Ludwig W, Strunk O, Klugbauer S, Klugbauer N, Weizenegger M, Neumaier J *et al.* (1998). Bacterial phylogeny based on comparative sequence analysis. *Electrophoresis* **19**: 554–568.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar *et al.* (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.

Malmstrom RR, Kiene RP, Cottrell MT, Kirchman DL. (2004). Contribution of SAR11 bacteria to dissolved dimethylsulfoniopropionate and amino acid uptake in the North Atlantic ocean. *Appl Environ Microbiol* **70**: 4129–4135.

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.

McKay RML, Gibbs SP, Espie GS. (1993). Effect of dissolved inorganic carbon on the expression of carboxysomes, localization of Rubisco and the mode of inorganic carbon transport in cells of the cyanobacterium *Synechococcus* UTEX. *Arch Microbiol* **159**: 21–29.

Mills MM, Moore CM, Langlois R, Milne A, Achterberg E, Nachtigall K *et al.* (2008). Nitrogen and phosphorus co-limitation of bacterial productivity and growth in the oligotrophic subtropical North Atlantic. *Limnol Oceanogr* **53**: 824–834.

Moore LR, Post AF, Rocap G, Chisholm SW. (2002). Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Limnol Oceanogr* **47**: 989–996.

Morris RM, Rappe MS, Connon SA, Vergin KL, Siebold WA, Carlson CA *et al.* (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420**: 806–810.

Morris RM, Vergin KL, Cho JC, Rappe MS, Carlson CA, Giovannoni SJ. (2005). Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic Time-series Study site. *Limnol Oceanogr* **50**: 1687–1696.

Nicastro D, Schwartz C, Pierson J, Cho J-C, Giovannoni SJ, McIntosh JR. (2006). Three-dimensional structure of the tiny bacterium *Pelagibacter ubique* studied by cryo-electron tomography. *Micros Microanal* **12**: 180–181.

Poretsky RS, Bano N, Buchan A, LeCleir G, Kleikemper J, Pickering M *et al.* (2005). Analysis of microbial gene transcripts in environmental samples. *Appl Environ Microbiol* **71**: 4121–4126.

Ram RJ, Verberkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake II RC *et al.* (2005). Community proteomics of a natural microbial biofilm. *Science* **308**: 1915–1920.

Rappe MS, Connon SA, Vergin KL, Giovannoni SJ. (2002). Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418**: 630–633.

Rivkin RB, Anderson MR. (1997). Inorganic nutrient limitation of oceanic bacterioplankton. *Limnol Oceanogr* **42**: 730–740.

Scanlan DJ, Silman NJ, Donald KM, Wilson WH, Carr NG, Joint I *et al.* (1997). An immunological approach to detect phosphate stress in populations and single cells of photosynthetic picoplankton. *Appl Environ Microbiol* **63**: 2411–2420.

Schulze WX, Gleixner G, Kaiser K, Guggenberger G, Mann M, Schulze ED. (2005). A proteomic fingerprint of dissolved organic carbon and of soil particles. *Oecologia* **142**: 335–343.

Sedwick PN, Church TM, Bowie AR, Marsay CM, Ussher SJ, Achilles KM *et al.* (2005). Iron in the Sargasso Sea (Bermuda Atlantic Time-series Study region) during summer: Eolian imprint, spatiotemporal variability, and ecological implications. *Global Biogeochem Cycles* **19**: 1–11.
Sieracki ME, Haugen EM, Cucci TL. (1995). Overestimation of heterotrophic bacteria in the Sargasso Sea: direct evidence by flow and imaging cytometry. *Deep Sea Res I* **42**: 1399–1409.
Sowell SM, Norbeck AD, Lipton MS, Nicora CD, Callister SJ, Smith RD *et al.* (2008). Proteomic analysis of stationary phase in the marine bacterium 'Candidatus Pelagibacter ubique'. *Appl Environ Microbiol* **74**: 4091–4100.
Steinberg DK, Carlson CA, Bates NR, Johnson RJ, Michaels AF, Knap AH. (2001). Overview of the US JGOFS Bermuda Atlantic Time-series Study (BATS): a decade-scal look at ocean biology and biogeochemistry. *Deep Sea Res II* **48**: 1405–1447.
Sumper M, Brunner E, Lehmann G. (2005). Biomineralization in diatoms: characterization of novel polyamines associated with silica. *FEBS Lett* **579**: 3765.
Tripp HJ, Kitner JB, Schwalbach MS, Dacey JWH, Wilhelm LJ, Giovannoni SJ. (2008). SAR11 marine bacteria require exogenous reduced sulfur for growth. *Nature* **452**: 741–744.
Van Mooy BA, Rocap G, Fredricks HF, Evans CT, Devol AH. (2006). Sulfolipids dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine environments. *Proc Natl Acad Sci USA* **103**: 8607–8612.
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.
Wanner BL. (1996). Phosphorus assimilation and control of the phosphate regulon. In: Neidhardt FC (ed). *Escherichia coli and Salmonella Typhimurium: Cellular and Molecular Biology*. ASM Press: Washington, DC. pp 1357–1381.
Washburn MP, Wolters D, Yates III JR. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**: 242–247.
Wilhelm LJ, Tripp HJ, Givan SA, Smith DP, Giovannoni SJ. (2007). Natural variation in SAR11 marine bacterioplankton genomes inferred from metagenomic data. *Biol Direct* **2**: 27.
Wilmes P, Bond PL. (2004). The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environ Microbiol* **6**: 911–920.
Woodger FJ, Badger MR, Price GD. (2003). Inorganic carbon limitation induces transcripts encoding components of the CO(2)-concentrating mechanism in *Synechococcus* sp. PCC7942 through a redox-independent pathway. *Plant Physiol* **133**: 2069–2080.
Wu J, Sunda W, Boyle EA, Karl DM. (2000). Phosphate depletion in the western North Atlantic Ocean. *Science* **289**: 759–762.
Zubkov MV, Fuchs BM, Tarran GA, Burkill PH, Amann R. (2003). High rate of uptake of organic nitrogen compounds by *Prochlorococcus* cyanobacteria as a key to their dominance in oligotrophic oceanic waters. *Appl Environ Microbiol* **69**: 1299–1304.
Zubkov MV, Mary I, Woodward EMS, Warwick PE, Fuchs BM, Scanlan DJ *et al.* (2007). Microbial control of phosphate in the nutrient-depleted North Atlantic subtropical gyre. *Environ Microbiol* **9**: 2079–2089.

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