The ISME Journal (2013) 7, 281–298

www.nature.com/ismej

ORIGINAL ARTICLE

Expression patterns reveal niche diversification in a marine microbial assemblage

Scott M Gifford[1], Shalabh Sharma[2], Melissa Booth[3] and Mary Ann Moran[2]
[1]*Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA;* [2]*Department of Marine Sciences, University of Georgia, Athens, GA, USA and* [3]*Marine Institute, University of Georgia, Sapelo Island, GA, USA*

Resolving the ecological niches of coexisting marine microbial taxa is challenging due to the high species richness of microbial communities and the apparent functional redundancy in bacterial genomes and metagenomes. Here, we generated over 11 million Illumina reads of protein-encoding transcripts collected from well-mixed southeastern US coastal waters to characterize gene expression patterns distinguishing the ecological roles of hundreds of microbial taxa sharing the same environment. The taxa with highest *in situ* growth rates (based on relative abundance of ribosomal protein transcripts) were typically not the greatest contributors to community transcription, suggesting strong top-down ecological control, and their diverse transcriptomes indicated roles as metabolic generalists. The taxa with low *in situ* growth rates typically had low diversity transcriptomes dominated by specialized metabolisms. By identifying protein-encoding genes with atypically high expression for their level of conservation, unique functional roles of community members emerged related to substrate use (such as complex carbohydrates, fatty acids, methanesulfonate, taurine, tartrate, ectoine), alternative energy-conservation strategies (proteorhodopsin, AAnP, V-type pyrophosphatases, sulfur oxidation, hydrogen oxidation) and mechanisms for negotiating a heterogeneous environment (flagellar motility, gliding motility, adhesion strategies). On average, the heterotrophic bacterioplankton dedicated 7% of their transcriptomes to obtaining energy by non-heterotrophic means. This deep sequencing of a coastal bacterioplankton transcriptome provides the most highly resolved view of bacterioplankton niche dimensions yet available, uncovering a spectrum of unrecognized ecological strategies.
The ISME Journal (2013) **7,** 281–298; doi:10.1038/ismej.2012.96; published online 30 August 2012
Subject Category: integrated genomics and post-genomics approaches in microbial ecology
Keywords: bacterioplankton; gene expression; metatranscriptomics; niche; ribosomal proteins

Introduction

The competitive exclusion principle, originally developed to conceptualize the organization of macroorganism communities (Hardin, 1960), posits that species richness is maintained by niche differentiation. While competitive exclusion theory has been applied to planktonic microbial communities (Fuhrman *et al.*, 2006; Mou *et al.*, 2008), the validity of this framework for microbes has been challenged from its beginning because of apparent overlap in major niche dimensions among co-occurring species (Hutchinson, 1961). This puzzle has recently been reinforced by genomic and metagenomic data showing that many genes of known biogeochemical or ecological relevance are broadly shared across major marine bacterioplankton groups (Moran, 2008), making it difficult to establish whether there are clear and unique ecological roles for individual bacterial taxa. One potentially important component of ecological function that has not typically been considered is diversity in sensing and responding to environmental cues. Thus, information on dynamic gene expression of individual microbial taxa inhabiting the same environment (Poretsky *et al.*, 2005; Frias-Lopez *et al.*, 2008; Poretsky *et al.*, 2010) can address a key aspect of niche differentiation.

Here we examine taxon-specific gene expression by the microbial members of a well-mixed coastal ocean system. Sequencing with the Illumina GAIIx (San Diego, CA, USA) platform provided deep coverage of the community transcriptome, allowing comparisons of transcripts assigned to hundreds of different reference genomes. This analysis reveals novel details on the functional niches occupied by members of a marine bacterioplankton community and provides insights into the diversity of strategies that support this complex microbial assemblage.

Correspondence: MA Moran, Department of Marine Sciences, University of Georgia, Athens, GA 30602-3636, USA.
E-mail: mmoran@uga.edu
Received 12 March 2012; revised 3 July 2012; accepted 16 July 2012; published online 30 August 2012

Materials and methods

Sample collection

Quarterly samples were collected at Marsh Landing (31°25′4.08 N, 81°17′43.26 W), Sapelo Island, Georgia, USA as part of the Sapelo Island Microbial Observatory (SIMO, http://www.simo.marsci.uga.edu) and named as follows: FN96 (7 November 2008), FN116 (17 February 2009), FN125 (14 May 2009) and FN158 (14 August 2009). All samples were collected at night, 4–6 h after sunset and 1 h before high tide (see Supplementary Table S1 for times and environmental data). Cell collection for RNA extraction was conducted as described previously (Poretsky *et al.*, 2009; Gifford *et al.*, 2011). Briefly, 6–8 l of water was drawn from a depth of 1 m and passed through a 3-μm prefilter (Capsule Pleated Versapor Membrane; Pall Life Sciences, Ann Arbor, MI, USA) and a 0.22-μm collection filter (Supor polyethersulfone; Pall Life Sciences). The 0.22-μm filter was placed in a WhirlPak bag and flash frozen in liquid nitrogen. Total time from start of filtration to flash freezing was 11–14 min.

RNA processing and sequencing

RNA processing was carried out as described by Poretsky *et al.* (2009) and Gifford *et al.* (2011). The frozen 0.22-μm collection filters were shattered, placed into 50 ml Falcon tubes with 8 ml of RLT buffer (Qiagen, Valencia, CA, USA) and 2 ml of PowerSoil beads (MO BIO, Carlsbad, CA, USA) and vortexed for 10 min on a MO BIO vortex adapter. RNA was extracted from the 50 ml tubes using an RNeasy kit (Qiagen), and any contaminating DNA was digested using TurboDNAse (Applied Biosystems, Austin, TX, USA). Ribosomal RNA (rRNA) was reduced using a two-step approach in which the samples were first treated enzymatically with the mRNA-only isolation kit (Epicentre, Madison, WI, USA) and then used in subtractive hybridizations with MicrobeExpress and MicrobeEnrich kits (Applied Biosystems). The enriched mRNAs were linearly amplified using the Message Amp II-Bacteria kit (Applied Biosystems), reverse transcribed to complementary DNA (cDNA) with the Universal Riboclone cDNA synthesis system (Promega, Madison, WI, USA) and purified with the QIAQuick PCR purification kit (Qiagen). The four cDNA samples were sheared to ~300 bp, barcoded and sequenced in one lane of an Illumina GAIIx run (Supplementary Table S2). Sequences are deposited in the CAMERA database (http://camera.calit2.net) under accession name 'CAM_P_0000917'.

Bioinformatics pipeline

An initial BLASTn (bit score cutoff $\geqslant 50$) comparison of a 25 000 read subsample from the combined metatranscriptome library against the SILVA database was used to assemble a database of rRNA sequences. The full metatranscriptome libraries were then searched against this custom database to identify rRNA sequences for removal. All remaining, non-rRNA sequences were compared with National Center for Biotechnology Information's (NCBI; http://www.ncbi.nlm.nih.gov) RefSeq database (version 43, September 2010) using BLASTx with a bit score cutoff $\geqslant 40$ to identify protein-encoding sequences. Taxonomic affiliation was assigned based on the top RefSeq hit and included all three domains of life as well as viruses. Assignments to Clusters of Orthologous Groups (COGs) for select reference genomes were obtained from the Integrated Microbial Genomes system (http://img.jgi.doe.gov)

Ortholog identification

For a set of 16 genomes selected based on their representation in the transcriptome and taxonomic breadth, orthologous genes were identified in a two-step process. Each gene in a subject genome was reciprocally compared against the other 15 genomes using BLASTp, and reciprocal hits with an E-value $< 10^{-4}$ were considered orthologs. A complete list of reciprocal best-hit ortholog pairs among the 15 genomes was iteratively collapsed by combining rows with common orthologs.

Gene searches

Ribosomal protein (RP) transcripts were identified by a text-based query of the read annotations. Transcripts binning to broad ecological categories for motility, alternative energy generation, transporters and inorganic nutrient uptake and metabolism were identified by keyword searches of the 16 selected reference genomes (Supplementary Table S3), retrieving any gene with a matching annotation along with any orthologs of that gene. The selected genes' annotations were then visually inspected to confirm a match with the ecological category of interest. For organic compound transporters, the process was the same except genes were initially identified by their COG transporter classification as described in Poretsky *et al.* (2010).

Genome islands

Putative genome islands were defined as those regions containing a gene with the word 'phage' in the annotation and with flanking genes (10 up- or downstream of the phage gene) having a significantly lower mean ortholog count and mean hit count ($P < 0.05$, bootstrapping 95% confidence interval) than the average for the entire genome bin.

Statistical analysis

Statistically significant differences in relative gene expression within an ortholog group were

determined using the non-parametric Wilcoxon rank-sum test ($P<0.05$). Treating each genome bin separately, genes that had read counts >1.5 times the interquartile range of their ortholog count group were labeled as outliers. Rows in the ortholog master table (described above) that contained IDs of outlier genes were then retrieved to make a table of outlier orthologous relationships. The indicator species analysis approach of Dufrene and Legendre (1997) was applied to the ortholog master table to identify 'indicator genes': those genes that were both expression outliers and whose expression was biased toward certain genomes. The latter was determined by the indicator value (IV), calculated as the proportion of gene expression contributed by one genome to the total expression of that gene summed over all the genomes.

The mean percent transcriptome within a genome was calculated as:

$$X_{ji} = \frac{a_{ji}}{n_i}$$

where a_{ji} is the number of reads of gene j in genome i, and n_i is the total number of reads in genome i. The IV of gene j in genome i was then calculated as:

$$\mathrm{IV}_{ji} = \frac{X_{ji}}{\sum_{i=1}^{g} X_{ji}} \times 100$$

where g = number of genomes. Genes with an IV>50 (majority of expression) were considered indicator genes for a particular genome. Because this approach is based on binning of transcripts to reference genomes, genes present in the wild population but not in the best-matched reference strain are missed in the analysis.

Results and discussion

Samples and sequencing

The Sapelo Island Microbial Observatory site is characteristic of nearshore habitats of the southeastern US, with ecological influences from marsh, freshwater and coastal environments (Pomeroy and Wiegert, 1981; Cai, 2011). Four high-tide, nighttime samples representing the fall, winter, spring and summer seasons yielded 31 million 100 bp cDNA reads. Contaminating rRNAs were identified by a BLAST search against a subset of the SILVA database (see methods) and accounted for 62% of the reads (Supplementary Table S2). Of the remaining 11 million potential protein-encoding reads, 4.1 million had a significant hit (bit score>40) in a BLASTx analysis against NCBI's RefSeq database. The percentage of potential protein-encoding reads with a significant RefSeq hit (mean 34%) is lower than in previous analyses (52% in Poretsky *et al.*, 2010; 53% in Gifford *et al.*, 2011), likely due to the shorter Illumina read length.

Active community members

Based on the highest-scoring hit from the RefSeq BLAST, ~4000 taxa were represented in the community transcriptome (Table 1). The distribution of hits among the taxonomic bins was log normal, with the top 200 bins accounting for 75% of all hits, followed by a long tail of bins with very few hits. Sequences recruiting to the Alphaproteobacteria, Gammaproteobacteria, Betaproteobacteria and Bacteroidetes dominated the transcriptomes (Table 1). The recently sequenced genome of *Puniceispirillum marinum*, the only SAR116 representative, recruited the most reads of any genome. Small, streamlined genomes, such as those related to *Pelagibacter ubique*, Betaproteobacterium KB13 and Flavobacteria MS024-2A and MS024-3C, were well represented, as were some medium-to-large genomes, particularly those related to the Roseobacter clade and the 'oligotrophic marine gamma' group. Hits to the Betaproteobacteria were predominantly to genomes of methylotrophic taxa. Archaeal genomes generally recruited few transcripts except for *Nitrosopumilus maritimus*, which was the fifth highest transcript-recruiting genome. Transcripts assigned to the Verrucomicrobiales, a relatively recently described phylum identified in both terrestrial and aquatic habitats, were also well represented. Overall, the reference bins were indicative of a highly diverse active coastal bacterioplankton community.

RP expression as an in situ growth indicator

Ribosomal proteins (RPs) are essential for protein synthesis, and bacterial genomes typically contain 50–60 RP genes (http://ribosome.med.miyazaki-u.ac.jp). Over 218 000 reads were annotated as RPs in the metatranscriptomes (Table 1), recruiting to 1903 different taxonomic bins. As up to 40% of a bacterial cell's energy is allocated to protein synthesis, cells strictly regulate and coordinate synthesis of RPs to balance translation machinery with available resources (Wilson and Nierhaus, 2007; Maguire, 2009). Indeed, levels of RP transcripts have been shown previously to be well correlated with growth rates in yeast (Eisen *et al.*, 1998), Bacteria (Wei *et al.*, 2001) and Archaea (Hendrickson *et al.*, 2008). These characteristics of RPs in model organisms suggested to us that their abundance could be leveraged as an index of *in situ* growth rates.

We used three strategies to evaluate whether the %RP values for bacterioplankton transcriptomes tracked with *in situ* growth rates. First, %RP data was checked for temperature-related seasonal shifts, as temperature has previously been shown to positively correlate with bulk community growth rates (del Giorgio and Cole, 1998 and references within); these changes were observed (Figure 1a), with %RP maxima occurring in the summer for the majority of taxa (56%), followed by spring, winter and fall (20, 18, and 7% of taxa, respectively)

Table 1 Taxonomic binning of coastal metatranscriptomic reads based on the highest-scoring pair from the BLASTx search against RefSeq (after rRNA removal; see Materials and methods)

	Hits	*Rank*	*%AA*	*RPs*
Total (3902)	4 151 196			218 198
Alphaproteobacteria (249)	1 745 196			70 882
Roseobacter (38)[a]	536 207			24 429
Roseobacter sp. AzwK-3b	37 682	13	81	870
Rhodobacterales bacterium HTCC2083	30 145	19	82	1046
Silicibacter lacuscaerulensis ITI-1157	28 146	23	87	2523
Roseobacter litoralis Och 149	26 533	24	83	472
Citreicella sp. SE45	26 233	25	87	2424
SAR11(4)	434 663			9198
Pelagibacter sp. HTCC7211	253 217	2	91	7414
Pelagibacter ubique HTCC1002	88 822	3	87	300
Pelagibacter ubique HTCC1062	41 889	11	86	786
alpha proteobacterium HIMB114	50 735	9	85	698
Miscellaneous alphas (208)	774 326			37 255
Puniceispirillum marinum IMCC1322	259 512	1	83	18 059
alpha proteobacterium BAL199	57 261	7	76	1680
Labrenzia alexandrii DFL-11	15 151	41	81	160
Hoeflea phototrophica DFL-43	13 699	46	78	233
Gammaproteobacteria (592)	855 576			68 776
gamma proteobacterium HTCC2080	80 293	4	79	4716
gamma proteobacterium NOR51-B	51 404	8	81	6161
gamma proteobacterium HTCC2207	41 755	12	75	2065
gamma proteobacterium HTCC2143	35 745	15	74	1376
gamma proteobacterium HTCC2148	33 634	17	76	2460
Betaproteobacteria (173)	247 164			11 282
beta proteobacterium KB13	36 854	14	91	1395
Methylophilales bacteria HTCC2181	29 347	20	84	987
Methylotenera sp. 301	6031	135	85	325
Methylovorus sp. SIP3-4	5467	147	81	407
Methylibium petroleiphilum PM1	5039	155	72	71
Bacteriodetes (31)[a]	375 382			24 681
Flavobacteria bacteria MS024-2A	43 539	10	81	2004
Flavobacteria bacteria MS024-3C	23 785	27	81	1113
Zunongwangia profunda SM-A87	12 861	53	80	1419
Robiginitalea biform. HTCC2501	12 220	54	78	895
Kordia algicida OT-1	11 293	62	78	1098
Verrucomicrobia (9)	110 647			8638
Coral. akajimensis DSM 45221	29 145	21	76	3718
Pedosphaera parvula Ellin514	25 411	26	74	1765
Verrucomicrobiae bacteria DG1235	18 498	33	73	1002
Opitutus terrae PB90-1	9551	85	72	615
Chthoniobacter flavus Ellin428	8589	96	73	346
Cyanobacteria (60)	71 618			2643
Synechococcus sp. WH 8109	11 218	64	91	563
Synechococcus sp. CC9605	4139	183	90	171
Cyanobium sp. PCC 7001	4067	186	86	245
Synechococcus sp. RS9916	3813	202	82	172
Archaea (103)	91 028			4028
Nitrosopumilus maritimus SCM1	67 890	5	86	2162
Cenarchaeum symbiosum A	3485	224	72	27
Sulfolobus tokodaii str. 7	2956	246	65	8
Aciduliprofundum boonei T469	964	566	70	253
Pyrococcus furiosus DSM 3638	687	695	66	10
Eukaryota (15)	147 364			4402
Micromonas sp. RCC299	33 680	16	90	1021
Ostreococcus lucimarinus CCE9901	13 569	47	79	704
Floydiella terrestris	11 055	67	96	6

Table 1 (Continued)

	Hits	*Rank*	*%AA*	*RPs*
Rhodomonas salina	10 937	69	92	298
Ostreococcus tauri	8667	94	88	0
Miscellaneous (2435)	507 858			22 866

Abbreviations: %AA, mean percent amino-acid identity of reads to genes in the reference genome; Hits, number of reads binning to a taxon for all four seasonal datasets combined; Rank, Rank abundance of a bin based on the total number of reads recruited; RPs, total number of reads annotated as ribosomal proteins in a given bin.
For each major taxonomic group, the total number of bins recruiting transcripts is shown in parenthesis, as well as the top five recruiting reference genomes.
[a]Rhodobacterales bacterium HTCC2255 (Roseobacter) and *Psychroflexus torquis* ATCC 700755 (Bacteriodetes) were not considered in the analysis due to contaminating sequences in the database.

(Figure 1b). The higher frequency of maxima occurring in winter rather than fall is potentially driven by the enrichment in %RP values in the winter over all other seasons for several Bacteroidetes and Verrucomicrobia taxa. Second, %RP values were evaluated relative to bacterial community secondary production rates (based on ^{3}H-leucine incorporation) that were measured concurrently with RNA sample collection; comparisons showed that relative abundance of RP transcripts followed the observed differences in bacterial production (summer: 2.8×10^{-6} g C $l^{-1} h^{-1}$; spring: 1.8×10^{-6} g C $l^{-1} h^{-1}$; fall: 0.4×10^{-6} g C $l^{-1} h^{-1}$; and winter: 0.3×10^{-6} g C $l^{-1} h^{-1}$). Finally, %RP patterns were compared against available literature values for taxon-specific *in situ* growth rates (Yokakawa *et al.*, 2004; Malmstrom *et al.*, 2005; Allers *et al.*, 2007; Teira *et al.*, 2009; Ferrerra *et al.*, 2011), and results showed that rankings of taxa based on %RP tracked well with rankings based on measures of *in situ* growth rates (Figure 1d).

Among the 200 top-recruiting reference bins, the percent of sequences identified as ribosomal proteins (%RP hits) within a genome ranged from 0.05% (Candidatus *Pelagibacter ubique* HTCC1002) to 20.5% (*Chryseobacterium gleum* ATCC 35910), and showed distinct phylogenetic patterns (Figure 1c). Gammaproteobacteria transcriptomes were clearly enriched in RP genes (mean 8.8%), and represented many of the highest %RP bins, including the polysaccharide degraders *Teredinibacter turnerae* T7901 (16.2%; the reference bacterium was isolated from a wood-boring bivalve), *Saccharophagus degradans* 2–40 (14.3%), *Marinomonas* sp. *MWYL1* (13.9%), and *Cellvibriojaponicus* sp. Ueda107 (12.9%). Bacteroidetes transcriptomes ranged from 2.6 to 20.5%RP, and included the two bins with the highest %RP, *C. gleum* (the top symbol in Figure 1c) and *Capnocytophaga gingivalis* ATCC 33624 (18.8%). SAR116 *Puniceispirillum marinum* IMCC1322 had a mid-range %RP (6.9%). Reference genomes for the Roseobacters were dispersed throughout the %RP distribution, from 1.8% (*Roseobacter litoralis* Och 149) to 9.2% (*Citreicella sp.* SE45). Finally, despite their dominance in the total transcript pool (Table 1), members of the SAR11 clade had the lowest %RP of all the groups examined, ranging from 0.05 to 1.7%RP.

Transcriptome characteristics

We conducted detailed analyses on 16 reference genome bins with high coverage in the metatranscriptomes and spanning the range of %RP (Table 2). The 16 genomes accounted for 29% of combined library reads, with the number of reads recruited per genome ranging from 11 000 (*Synechococcus* sp. WH8109) to >259 000 (*P. marinum* IMCC1322; Table 1). While most genomes approached transcriptome saturation (Supplementary Figure S1), the percentage of genes hit in the reference genomes varied widely, from 28% in *P. parvula* Ellin514 to over 91% in SAR11 isolate HTCC7211 (Table 2). The average amino-acid sequence identity between the transcripts and the genes they were assigned to in the reference genomes ranged from 74% for *P. parvula* Ellin514 to 91% for *Synechococcus* sp. WH8109 (Table 1).

Functionally redundant genes among the 16 genomes were identified using reciprocal BLAST analysis (Figure 2 and Supplementary Figure S2). The overall commonness of genes present in a genome (that is, how often they were also found in other genomes) was inversely correlated to genome size ($R^2 = 0.63$, $P < 0.001$), with the highest overlap in gene content occurring in the small, streamlined SAR11 and Betaproteobacteria genomes, and the lowest in the large genomes of *Pedosphaera parvula* Ellin514 and alpha proteobacterium BAL199 (Table 2). Genome regions with low or non-detectable expression often had few orthologs in other genomes (Figure 2 and Supplementary Figure S2) and many were flanked on one side by phage elements, suggesting these are genomic islands (Coleman *et al.*, 2006) not present in the sampled coastal populations. These low-recruiting regions were enriched in the SAR116 and Roseobacter reference genomes and depleted in the streamlined SAR11, KB13 and Flavobacteria genomes (Table 2). In comparisons limited to reference genomes, these

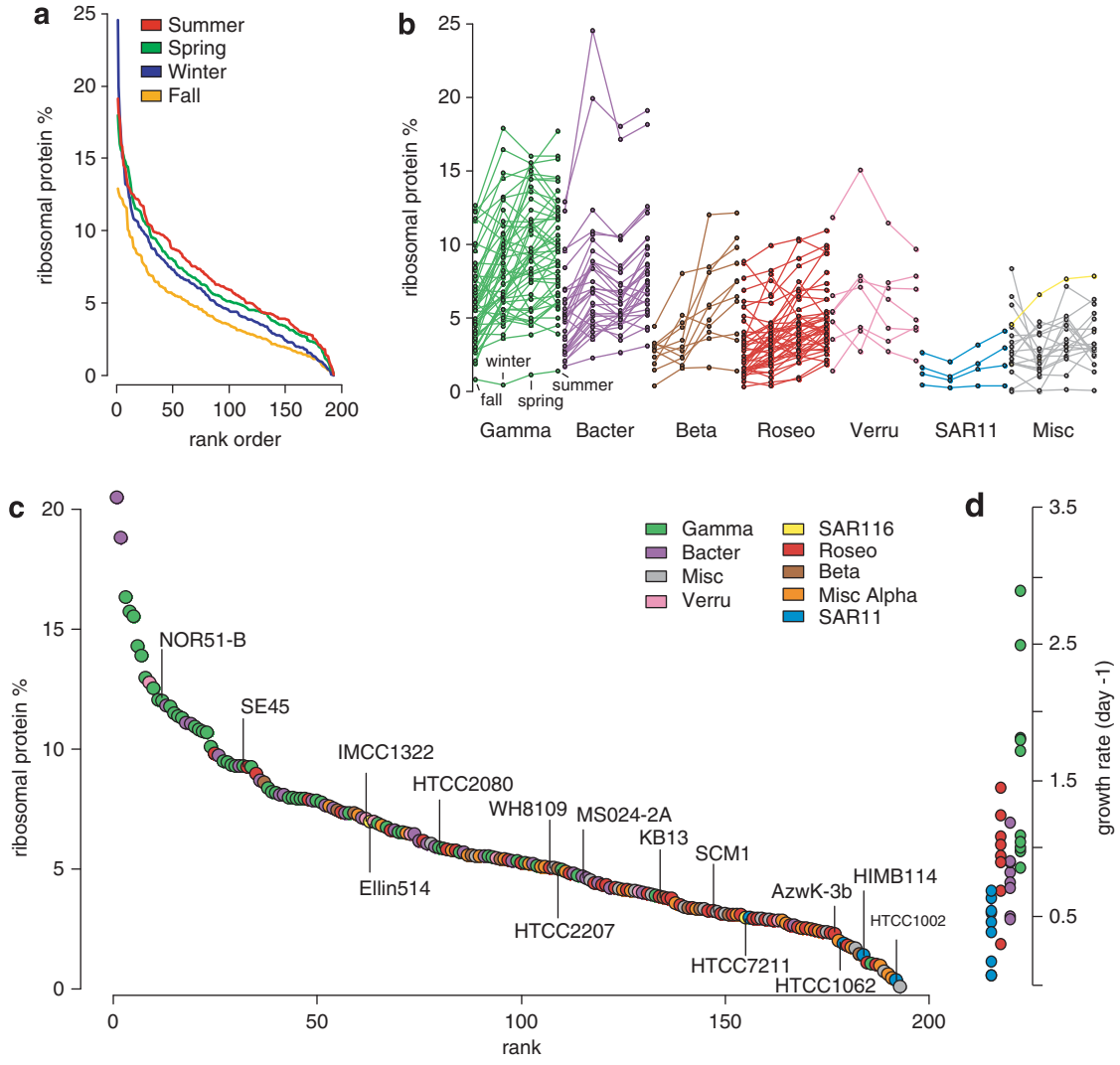


Figure 1 Relative abundance of ribosomal protein (RP) reads in the top 200 reference genomes (eukaryotic hits excluded). (**a**) Rank order for the 200 reference genomes separated by season. (**b**) Temporal trends in %RP for the four seasonal samples with individual taxa separated. Gamma = Gammaproteobacteria, Bacter = Bacteroidetes, Beta = Betaproteobacteria, Roseo = Roseobacter, Verru = Verrucomicrobia, Misc = Miscellaneous. (**c**) Same as (**a**), except the seasonal samples are combined for each genome bin. See Table 2 for full names and taxonomic affiliations. (**d**) Group-specific growth rate data from Yokakawa *et al.*, 2004; Malmstrom *et al.*, 2005; Allers *et al.*, 2007; Teira *et al.*, 2009; Ferrerra *et al.*, 2011.

islands could be mistakenly identified as regions of unique, niche-defining genes even if they are neither present nor expressed in the natural populations. However, the combination of expression data with ortholog analysis identifies actively expressed functional content representative of the sampled population.

Relationship between expression level and gene prevalence

We found a statistically significant positive relationship between the expression level of a gene (the number of reads recruited to a gene in a reference bin) and how commonly it was harbored in the other taxa (the number of genomes with an ortholog for that gene) (Wilcoxon rank-sum test, $P<0.05$; Figure 3, Supplementary Figure S3). Thus, most highly expressed genes were shared by multiple taxa, although the SAR11 members HTCC1002, HTCC1062 and HIMB114 as well as MS024-2A and *N. maritimus* SCM1 did not have as strong a pattern as the others (Supplementary Figure S3). Previous observations from marine metatranscriptomic data (Hewson *et al.*, 2009; Stewart *et al.*, 2011) support this conclusion, and suggest that an analysis restricted to only the more highly expressed genes is likely to miss unique functional capabilities that distinguish taxa.

Bacterioplankton niches

We therefore took a broader approach to identifying niche-defining features for each taxon that involved characterization of three classes of transcripts. (1) Highly expressed genes: the 10 most highly

Table 2 Summary statistics for reads binning to the select 16 genomes

Taxonomic affiliation	*Genome*	*Bin size*[a]	*%Genes hit*[b]	*%RP*[c]	*% Top10*[d]	*Mean orthologs*[e]	*Genome islands*[f]
SAR116	Candidatus *Puniceispirillum marinum* IMCC1322	2543	84	7.0	17	5.8	12
Alpha	alpha proteobacterium BAL199	6128	41	2.9	34	2.8	29
Roseobacter	*Roseobacter* sp. AzwK-3b	4145	45	2.3	42	4.0	26
Roseobacter	*Citreicella* sp. SE45	5427	34	9.2	24	3.1	18
SAR11	Candidatus *Pelagibacter* sp. HTCC7211	1447	91	2.9	38	8.0	1
SAR11	Candidatus *Pelagibacter ubique* HTCC1002	1393	61	0.3	64	8.4	0
SAR11	Candidatus *Pelagibacter ubique* HTCC1062	1354	66	1.9	43	8.7	0
SAR11	alpha proteobacterium HIMB114	1425	75	1.4	55	7.9	0
Gamma	marine gamma proteobacterium HTCC2080	3185	84	5.9	10	4.5	3
Gamma	gamma proteobacterium NOR51-B	2930	67	12.0	12	4.9	5
Gamma	marine gamma proteobacterium HTCC2207	2388	81	4.9	26	5.6	4
Beta	Betaproteobacterium KB13	1318	84	3.8	58	7.8	0
Bacteriodetes	Flavobacteria MS024-2 A	1772	87	4.6	24	5.1	1
Verrucomicrobia	*Pedosphaera parvula* Ellin514	6510	28	6.9	16	1.9	3
Cyanobacteria	*Synechococcus* sp. WH 8109	2577	61	5.0	10	3.7	4
Archaea	*Nitrosopumilus maritimus* SCM1	1797	73	3.1	45	3.0	1

[a]Total number of genes in reference genome.
[b]The percentage of total genes in the genome hit in the combined metatranscriptome.
[c]The percentage of bin hits annotated as ribosomal proteins.
[d]Percent of total bin hits by the top 10 most highly expressed genes.
[e]Average number of orthologs per gene.
[f]Number of regions (10 or more genes in length) of the genome with statistically lower number of hits and orthologs and flanked on one side by a phage-related protein.

expressed genes in each genome bin, representing the processes garnering the most transcriptional effort by that taxon (Table 3). (2) Ecological benchmark genes: selected biogeochemically relevant genes representing traits such as nutrient acquisition, substrate transport, energy acquisition and motility (Table 4). (3) Indicator genes: the genes whose expression was higher than expected based on commonness in the other genomes (>1.5 interquartile range of its ortholog group (Figure 3) and for which the majority of expression ($>50\%$) was found in that taxon (Table 5); see Materials and methods).

SAR116

The *P. marinum* highly expressed genes encoded a Na^+/solute symporter, and ABC sugar and TRAP dicarboxylate transporters, along with energy transduction and transcription/translation machinery (Table 3). The highly expressed V-type H^+-translocating pyrophosphatase establishes a proton gradient across the membrane and has been hypothesized to have a role in scavenging metabolic energy in substrate-limited cells (García-Contreras *et al*., 2004; Rinta-Kanto *et al*., 2012).

Expression of ecological benchmark genes showed that the SAR116 populations in this coastal ocean are likely motile and supplement heterotrophic growth with light-driven phototrophy (Table 4) via proteorhodopsin (Béjà *et al*., 2000). Transcripts indicate cells were actively transporting and metabolizing most inorganic nutrients, but the transcriptome was particularly enriched in sequences for processing polyphosphate and nitrate (Table 4). SAR116 populations also had sequences for amino acid and five-carbon carbohydrate uptake overrepresented compared with other bacterioplankton taxa (Table 4).

SAR116 indicator genes included those for the uptake and oxidation of methanesulfonate (Table 5), a compound generated from the oxidation of dimethyl sulfoxide (Kelly and Murrell, 1999). These included methanesulfonate monooxygenase (MSO) subunits, and a nitrate/sulfonate/bicarbonate permease that neighbors the SAR116 MSO genes and is homologous to a methanesulfonate transporter from the soil Alphaproteobacteria *Methylosulfonomonas methylovora* (Jamshed *et al*., 2006). Indeed, this entire SAR116 genome region (SAR116_2098-2109) has high homology and synteny with the *M. methylovora* methanesulfonate utilization operon. Two other SAR116 indicator genes fell in the degradation pathway for the aromatic compound protocatechuate (Table 5).

Roseobacters

The two selected roseobacter genome bins showed clear differences in their highly expressed genes (Table 3). For *Roseobacter* sp. AzwK-3b, these genes encoded aerobic anoxygenic photosynthesis (AAnP). For non-phototrophic C*itreicella* sp. SE45, which had the highest Roseobacter %RP, these genes encoded growth-related processes (transcription, translation and energy transduction). Formate dehydrogenase was expressed in both genomes, and made up $>11\%$ of all C*itreicella* sp. SE45 hits (Table 3).

The ecological gene analysis revealed that both roseobacter taxa were assimilating a variety of inorganic nutrients and obtaining energy via the *sox*-based sulfur oxidation pathway (Table 4). Cells

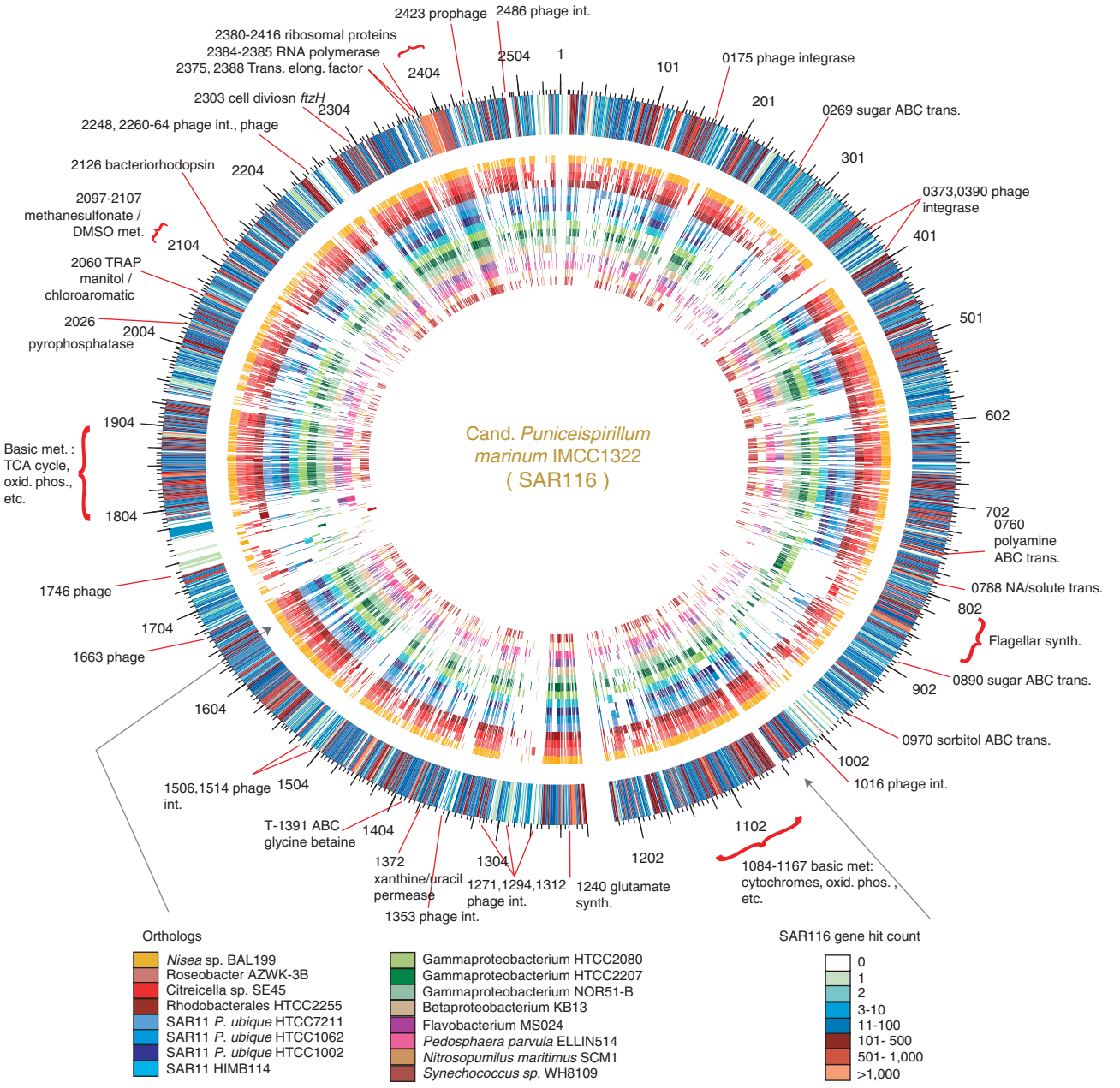


Figure 2 Transcriptome of SAR116 clade member *Puniceispirillum marinum* IMCC1322. The outer ring shows the 2543 genes in the IMCC1322 genome colored according to the number of RefSeq hits in the combined metatranscriptome. The inner rings denote the presence of orthologs to an IMC1322 gene in the other 15 genomes. Ring axis labels correspond to gene order as designated by NCBI; for SAR116 this corresponds with the order of locus tag numbers.

were expressing genes for the uptake of organic nitrogen compounds, such as amino acids and polyamines (Table 4), and there was an indicator gene for urea metabolism in SE45 (Table 5). Interestingly, while both reference genomes harbor genes for flagellar synthesis, expression of these genes was hardly detectable in the sampled populations (Table 4).

SAR11

For the four SAR11 genome bins, expression of a Na^+/solute symporter and proteorhodopsin accounted for 16–41% of transcripts (Table 3). The high expression of these two genes is likely linked, as establishing a sodium gradient across the cell membrane is important in proteorhodopsin-mediated growth stimulation in the flavobacterium *Dokdonia* MED134 (Kimura *et al.*, 2011), and this Na^+ gradient could subsequently drive substrate uptake via the highly expressed symporter. While the substrate of the SAR11 Na^+/solute symporter has not been experimentally identified, we hypothesize that it transports acetate based on sequence similarity to an acetate permease in *Escherichia coli* (HMPREF9346_04543) and proximity in all four

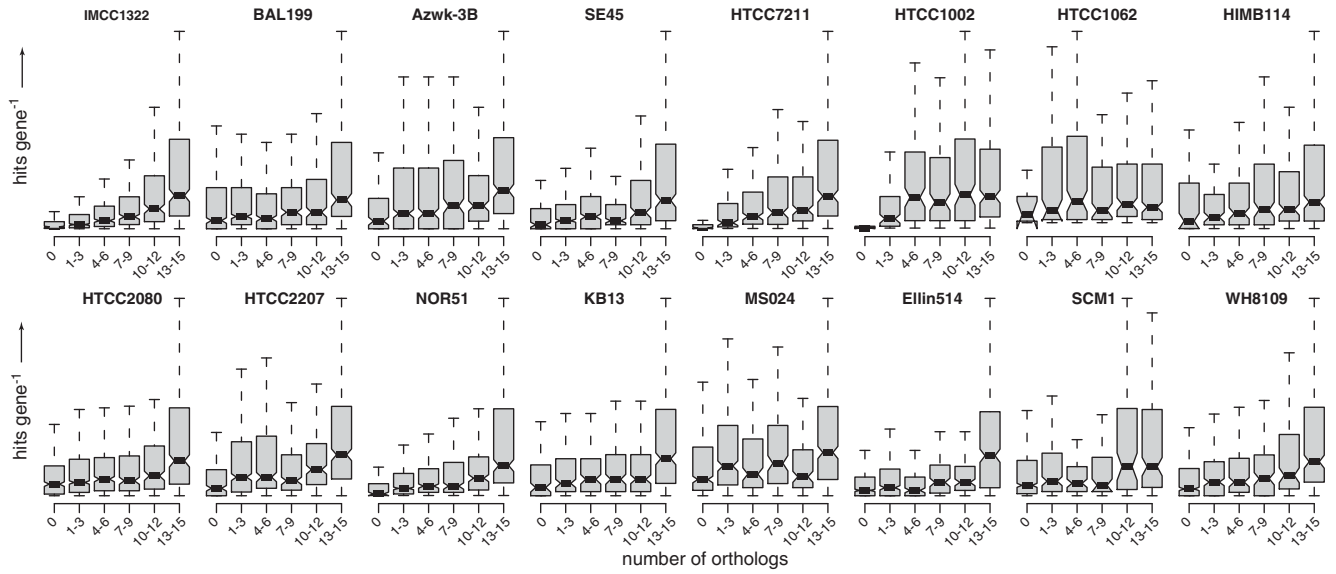


Figure 3 Expression level (hit count) as a function of ortholog number (representation in the other 15 genomes). Error bars denote 1.5 times the interquartile range. The magnitude of the *y*-axis varies for each genome and is not shown for clarity.

SAR11 reference genomes to an expressed acetyl-CoA synthetase (see Supplementary Information). Other highly expressed SAR11 genes included V-type H^+-translocating pyrophosphatases and taurine transporter protein.

The ecological benchmark gene set showed that ammonia transporters were among the most highly expressed genes (Table 3). Indeed, SAR11s were second only to the ammonia-oxidizing *N. maritimus* populations in the percent transcriptional effort devoted to ammonia processing (Table 4). Phosphate transporters were also enriched in several SAR11 bins, and HIMB114 devoted more transcriptional effort to phosphonate acquisition than any other taxon (Table 5). The SAR11 transcriptomes were enriched in transporters for amino acids, carboxylic acids and nucleotides (Table 4).

Indicator genes in HTCC7211 included two subunits of adenylyl-sulfate reductase (APS reductase; Table 5). These genes have close homologs in HTCC1002 and HTCC1062, which were also well expressed, and cluster in the Apr lineage I group of sulfur-oxidizing bacteria that are hypothesized to use these genes in the reverse direction to oxidize sulfur (Meyer and Kuever, 2007). The HIMB114 transcriptome was significantly enriched in carbon monoxide oxidation sequences (Table 4), though these genes are of the type II form whose function in CO oxidation has been recently questioned (Cunliffe, 2011). HIMB114's indicator genes included a dehydrogenase for tartrate (HIMB114_0953; Malik and Viola, 2010), a compound shown to be secreted by marine algae (Marsh *et al.*, 1992).

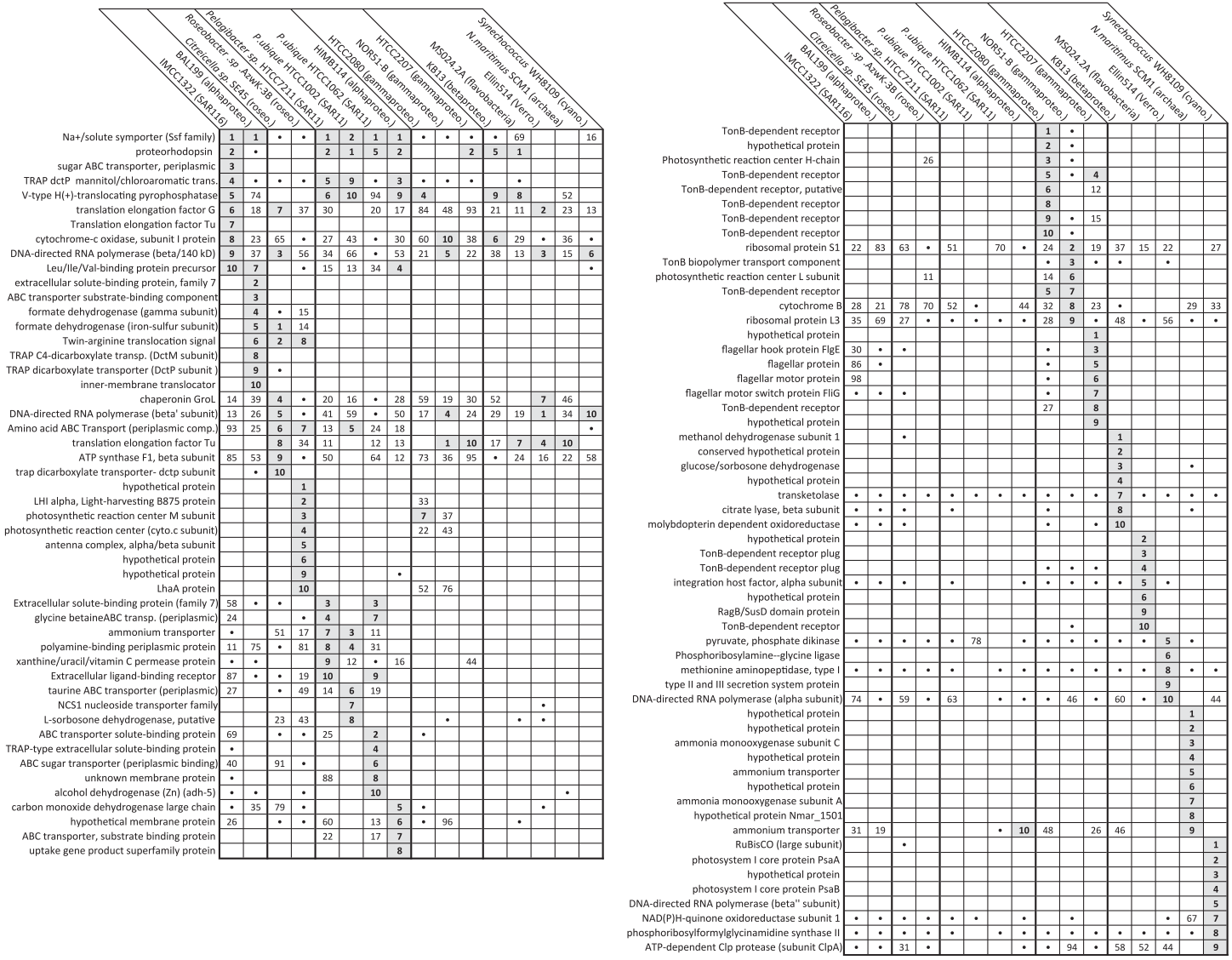
Three of the four SAR11 genomes were significantly enriched in proline/glycine betaine transporters (Table 4), and HTCC7211 had indicator genes for both transporting and degrading glycine betaine (Table 5). Tripp *et al.* (2008, 2009) have shown that the growth of SAR11 HTCC1062 in culture is significantly improved by glycine betaine. The HTCC7211 bin also had two indicator genes for uptake of ectoine/hydroxyectoine, which serve as compatible solutes during osmotic stress (Mulligan *et al.*, 2011) as well as a carbon and nitrogen source (Lecher *et al.*, 2009). Overall, the SAR11 bins indicate substantial transcriptional investment in the uptake and metabolism of compatible solutes.

Gammaproteobacteria

The three diverse Gammaproteobacteria lineages all showed high expression of TonB-dependent transporters associated with iron and vitamin uptake and, more recently, a range of substrates that includes metals, sugars and oligosaccharides (Schauer *et al.*, 2008) and are potentially involved in motility (Cursino *et al.*, 2009) (Tables 3, 4 and 5). All three bins also had abundant transcripts for light-driven energy generation via AAnP genes (HTCC2080 and NOR51-B) or proteorhodopsin (HTCC2207; Table 4). Members of the HTCC2080 and HTCC2207 populations were motile at the time of sampling, and HTCC2207 devoted more transcriptional effort to motility than any other taxonomic bin (Table 4).

Gammaproteobacteria transcript pools were variously enriched in ecological benchmark genes for acquisition of phosphorus (Table 4; inorganic phosphate transporters, alkaline phosphatases and polyphosphate metabolism), but depleted in ammonia transporter expression relative to the Alphaproteobacteria groups. The Gammaproteobacteria bins were enriched in Na^+/H^+ antiporter transcripts (Table 3), which work to maintain the membrane Na^+ gradient and, as in the SAR11s (see above) and

Table 3 Top 10 highest expressed genes for the select 16 reference genomes



The gene name is given at the front of each row (see Supplementary Table S4 for accession numbers). The rank (ordered from highest expressed gene to lowest within a genome) is shown for genomes that had the gene in its top 100 highest expressed. Cells containing a top 10 expressed gene are bolded and grayed. A dot (•) indicates that the gene was expressed but ranked >100. Empty cells indicate either the genome contained no ortholog to the gene or that no expression was detected.

Table 4 Transcriptional effort devoted to key ecological and niche-defining functions

		SAR116 *P. marinum* IMCC1322	*Nisaea* sp. BAL199	*Roseobacter* sp AzwK-3b	*Citreicella* sp. SE45	SAR11 *Pelagibacter* sp. HTCC7211	SAR11 *P. ubique* HTCC1002	SAR11 *P. ubique* HTCC1062	SAR11 HIMB114	gamma HTCC2080	gamma HTCC2207	gamma NOR51-B	beta KB13	Flavobacteria MS024-2A	*P. parvula* Ellin514	*Synechococcus* sp. WH 8109	*N. maritimus* SCM1
Motility	Flagella	2.00	0.34	0.01	0.05					1.11	13.55	0.01		0.01			
	Gliding motility		0.04										0.03	1.28	0.08		
	Secretion and Pilus	0.02		·		0.04	0.06	0.06	0.11	0.61	0.15	0.56	0.09		3.00	0.05	·
Alt. Energy Conservation	Proteorhodopsin	2.23	0.03			5.80	25.26	3.15	4.79		4.78		1.85	8.25			
	Phototrophy			25.93	·					6.10	0.01	4.22				5.83	
	Sulfur Oxidation		0.09	0.95	0.58					0.10		0.27	0.06			0.03	
	Hydrogen Oxidation				0.01									0.17			
	Ammonia oxidation	0.02	·	·	·				0.03								6.95
	Carbon Monoxide Oxidation	0.09	0.57	0.07	0.50				2.52	0.22	0.07	·		0.09	·		
Nutrients	Phosphate	0.06	0.02	·	0.08	0.05		0.68	0.30	0.12	0.07	0.29	0.10	0.14		0.15	·
	Alkaline phosphatase	·			0.04					0.03	0.04	·		0.05	·		
	Phosphonate		0.10	0.11	0.05	·			0.87		0.02			·		0.12	0.21
	Polyphosphate	0.05	0.02	0.01		·				0.04		0.06	0.02	0.04	0.13	·	
	Ammonium	0.61	0.66	0.93	1.30	2.28	7.81	2.24	1.74	0.30	0.54	0.02	0.23	·		0.51	3.69
	Nitrate	0.42	0.05	·	·					0.01						0.06	0.03
	Nitrite-Nitric-Nitrous	·		0.02						·	·	0.04	·			0.04	
	Sulfate	0.24	0.33	0.14	0.41	1.38	0.31	0.78	0.05	0.26	0.25	0.43	0.08	0.25	·	0.58	0.09
	Sulfite	0.03	0.09	0.04	0.06	·	·	0.02		0.09	·	0.16			0.11	0.19	0.09
	Sulfide	·	0.07		·	0.02	·	0.02	·	·		0.01		0.29	0.31		
Transport	Amino Acid	0.59	0.58	1.83	1.98	3.31	5.49	1.36	0.91	0.20	0.15	0.32	0.02	0.59	·	0.25	
	Amino Acid - branched chain	2.75	10.45	1.58	1.15	2.97	1.02	4.86	3.13	·	0.02	·			·	0.12	
	Amino Acid -olio/dipeptide	1.28	3.79	0.20	0.91	0.86				0.15	0.29	0.29	0.04	0.14	0.03	0.09	0.01
	Carbohydrate	3.33	0.40	0.46	1.32	0.69	·	10.62	0.01	0.57	0.35	0.07	·	0.39	0.22	0.17	·
	Carbohydrate - C5s	0.61			0.06										0.03		
	Carbohydrate - maltose																
	Carboxylic Acids	2.99	15.17	3.22	4.22	7.29	2.58	8.45	5.44	0.33	0.26	0.15	0.05	0.20		0.05	
	Proline/Glycine	0.75		0.29	0.09	4.89	0.16	3.86	1.64	0.43	0.14	0.02	0.04	0.52	0.01	0.72	
	Spermidine/putrescine	1.51	0.53	0.38	1.95	2.12	6.73	0.75									
	fatty acid/glycerol	0.02	0.02		0.01	0.01		0.02	0.01	·	0.01	0.20	·	0.13			0.26
	Nucleotide	0.43	0.01		0.28	0.06	2.38	0.02		0.09	0.01	0.18		0.07	·		
	TonB transport	0.01	·		·			0.05	·	8.43	7.41	4.31	0.13	7.24	0.02		
	Symporters (non AA)	5.56	8.81	0.03	0.01	14.32	9.64	12.98	35.93	0.60	0.31	0.60	0.05	1.00	0.04	0.94	
	Antiporters	0.14	0.07	0.05	0.05	0.06	0.03		0.09	0.20	0.09	0.21	·	0.27		0.37	0.13

Based on a keyword search of the 16 reference genomes, read counts for genes with matching annotations (Supplementary Table S2) were tallied and divided by the total number of reads in that bin. The cells are shaded according to the relative proportional magnitude across a row (i.e., across genomes). Dark gray cells with white font mark genomes significantly enriched in a category (% transcriptome >upper 95% confidence interval as determined by bootstrapping with 10 000 iterations). Cells with a black dot indicate expression was detected but made up <0.01% of hits to a genome.

Bacteroidetes (see below), may be coupled with light-driven proton pumping.

The indicator gene analysis for Gammaproteobacteria was biased toward genes for fatty acid metabolism. The HTCC2080 bin contained nine indicator genes for fatty acid metabolism, including four acyl-CoA dehydrogenases and a 3-ketoacyl-CoA thiolase involved in fatty acid β-oxidation (Table 5). Furthermore, HTCC2080 populations were expressing 153 genes involved in lipid transport and metabolism, three times the average for these functional gene categories in the other 15 genomes (Supplementary Table S5).

Twenty four of the Gammaproteobacterium HTCC2207 indicator genes were for motility, including those for flagellar assembly and chemotaxis (Table 5). Four HTCC2207 indicator genes contained cadherin domains involved in complex carbohydrate degradation via cell aggregation and direct binding to cellulose, xylan and related compounds

Table 5 Transcriptional effort devoted to select indicator genes of the 16 reference genomes, expressed as percent of a bin's transcriptome

	Locus tag	Indicator Gene	Cand. *P. marinum* IMCC1322	*Roseobacter* sp. BAL199	*Nisaea* sp. Azwk-3b	*Citreicella* sp. SE45	Cand. *P. sp.* HTCC7211	Cand. *P. ubique* HTCC1002	alpha proteo. HIMB114	gamma proteo. HIMB1062	gamma proteo. HTCC2080	gamma proteo. HTCC2207	gamma proteo. NOR51-B	beta proteo. KB13	Flavobacteria MS024-2A	bacterium Ellin514	*Synechococcus* sp. WH 8109	Cand. *N. maritimus* SCM1	Indicator mean orthologs
IMCC1322	1471	ABC-type nitrate/sulfonate/bicarbonate transport	0.07																0.0
	0930,1461	Branched-chain amino acid transport	0.56	0.12		0.05	0.03	0.03	0.01	0.05									4.5
	1528-1529	Peptide/nickel transport	0.27																0.0
	0970	Sorbitol/mannitol transport	0.61			0.01											0.01		2.0
	0936, 0940	Benzoate degradation	0.08																1.0
	0095	Carbon-monoxide dehydrogenase	0.05										0.00						1.0
	2101, 2109, 2112	Methanesulfonate trans. and met.	0.47															0.01	0.7
BAL199	27586	Formate dehydrogenase		2.84	0.76	0.05													2.0
	16503,26207-26227	Branched-chain amino acid ABC transport	0.03	6.45	0.00	0.03													1.5
	02669, 03564, 03574, 07173	C4-dicarboxylate-binding TRAP transport	11836, 12441, 27026, 28205, 28215	9.86	0.05	0.04				3.22									1.0
	21694-21699,27870	Iron(III) ABC transport		1.07															0.0
	06614-06624, 06634, 08938	peptide/nickel ABC transport	10592, 13683, 14060, 29315-29320, 30037-30047	3.43	0.03	0.03	0.01						0.05					0.01	1.4
	17983	Carbon-monoxide dehydrogenase		0.05															0.0
	02734,03729, 27256,28610	Glucose-methanol-choline oxidoreductase		0.29															0.0
	21114	Ammonium transporter		0.05															0.0
	19061-16066, 16158	Sulfonate/nitrate/taurine ABC transport		1.50		0.04													0.7
	19046	Alpha-ketoglutarate-dependent taurine dioxygenase		0.13			0.01			0.02	0.04								3.0
	04044, 27356	phytanoyl-CoA dioxygenase	0.01	0.17			0.00				0.02								2.5
	05144	Carboxymethylenebutenolidase		0.06															0.0
Azwk-3b	01370,19591	5-aminolevulinate synthase	0.04	0.00	0.51					0.18	0.04	0.00				0.03			4.0
	03805, 04215-04220,10096	Branched-chain amino acid ABC transport	0.00	0.10	0.63	0.18													1.8
	12819, 12839	Sulfur oxidation		0.07	0.72	0.02					0.03		0.07						1.5
	00830,00840,01240, 12507-12517	TRAP dicarboxylate transporter		0.05	0.84	0.00	0.02			0.24									2.3
	17073	Putative enoyl-CoA hydratase/isomerase		0.01	0.10														1.0
	15698	Beta-ketoacyl synthase family protein			0.09	0.02													1.0
	01665	Taurine--pyruvate aminotransferase			0.14	0.12						0.01							3.0
	19511,19551-19556,19586	Aerobic anoxygenic photosynthesis	19606-19611,19621-19626,19636-19341,19656		25.69						2.70		1.59				0.11		2.5
SE45	0340, 0344	Sulfur Oxidation			0.03	0.24					0.03		0.03	0.00					2.5
	5340-5341	Formate dehydrogenase		4.82	1.94	11.40													2.0
	0579	Urease		0.01	0.03	0.09						0.01					0.04		4.0
	3705	Taurine--pyruvate aminotransferase				0.17						0.01							1.0
	3824	Polyhydroxyalkonate synthesis repressor		0.00		0.10				0.02									4.0
	2435	Ammonia transport				0.82													0.0
	0253, 1127-1129, 1547-1549, 2091	C4-dicarboxylate TRAP transport	3674, 3678, 4678, 4876, 5363	0.64		3.14													1.1
	3669, 3671, 4427	Tricarboxylic TRAP transport				0.64													0.3
HTCC7211	147, 194	Glycine betaine/proline transport			0.01		0.21	0.11	0.00										1.5
	776, 1327	Ectoine/hydroxyectoine ABC transporter					0.57												0.0
	563, 1116	Adenylylsulfate reductase					1.16	0.16	0.59										2.0
HTCC1002	06696	Ammonium transporter	0.07		0.66	0.29	2.08	7.62	2.22										5.0
	02371	Taurine transport system periplasmic protein	0.49		0.25	0.13	1.28	4.07	1.05										5.0
	06546	Spermidine/putrescine-binding periplasmic protein	0.82	0.17	0.16	0.12	1.87	6.51	0.45										6.0
	02076	TRAP mannitol/chloroaromatic transport						0.22											2.0
HTCC1062	0269-271, 0769-0770	multiple sugar ABC transport	0.62	0.00	0.20	0.26	0.69		10.50								0.11		5.2
	1290	C4-dicarboxylate TRAP transporter	0.02						3.44										2.0
	1179	phosphate ABC transport	0.04	0.02	0.01	0.01	0.05		0.50	0.22		0.02		0.06					8.0
HIMB114	0364-0365	Carbon monoxide dehydrogenase subunits	0.08	0.42	0.03	0.42				2.34	0.19								5.5
	0953	Tartrate dehydrogenase	0.00	0.02			0.03			0.12									4.0
	1079	Phosphonate transport substrate-binding protein		0.01		0.03	0.00			0.84		0.01					0.08		6.0
	0552	Ferrous iron permease EfeU								0.53									0.0
	0852	Poly3-hydroxyalkanoate polymerase (PHAsynthase)		0.03	0.01	0.02				0.15						0.01		0.01	6.0
	1273	Putative Na+/solute symporter								0.11									0.0
	0332	Taurine dioxygenase		0.03						0.08	0.02								2.0
	0326, 0339, 0341, 1107	Putative tricarboxylic transport	0.19	0.06		0.15				2.79									0.8

Table 5 (Continued)

	Locus tag	Indicator Gene	Cand. P. marinum IMCC1322	Roseobacter sp. BAL199	Nisaea sp. Azwk-3b	Citreicella sp. SE45	Cand. P. ubique HTCC7211	Cand. P. ubique HTCC1002	alpha proteo. HTCC1062	gamma proteo. HIMB114	gamma proteo. HTCC2080	gamma proteo. HTCC2207	gamma proteo. NOR51-B	Flavobacteria MS024-2A	beta proteo. KB13	bacterium Ellin514	Synechococcus sp. WH 8109	Cand. N. maritimus SCM1	Indicator mean orthologs
HTCC2080	07579	Type 4 fimbrial biogenesis related protein									0.11								0.0
	01926, 02565, 03080, 08174	TonB-dependent receptors 10368, 10658, 12519, 12793, 13038, 13758									4.77	0.26	0.30	0.01	0.02				1.3
	01781, 12048	Na+:H+ antiporter									0.16		0.03						0.5
	00535	Fatty acid metabolism									0.16		0.06						1.0
	00625, 06687, 11573, 12683-12688, 15924	Fatty acid degradation		0.04							1.01	0.20							0.5
	11383, 14179	Fatty acid synthesis									0.13								0.0
	10153, 10208	Aerobic anoxygenic photosynthesis			0.02	0.01					0.21		0.02						2.5
	02471, 02750	Choline transporter			0.00	0.01					0.25	0.00			0.01				3.0
HTCC2207	08331-08336, 08351, 08361-08396, 08406, 08441	Flagellar synthesis 08451-08456, 08471, 08491, 08511-08516, 08551-08556, 08581, 08591, 08601	1.80	0.11	0.01	0.03					0.51	12.23			0.01		0.03		3.5
	00070, 03399, 03704, 03999	TonB dependant receptors 07223, 08191, 09321, 09881, 11273									2.17	5.17	0.13	0.01	0.02				2.8
	06084, 09816	Cellobiosidase										0.17							0.0
	00005-00010, 06108, 06128	Cadherin domain proteins	0.00			0.04						0.32							0.5
	09831, 09841, 06718, 10126,12017	Glucosidases		0.00							0.04	0.70			0.04	0.00	0.04		1.6
	10316	Betalactamase										0.23							0.0
NOR51-B	269, 1279, 2331, 2397, 2831	TonB-dependent receptors									0.56	0.67	2.00	0.09					1.4
	2735	Lipid A export ATP-binding/permease MsbA									0.20	0.24	0.67						2.0
	2258	biotin/lipoyl attachment containing protein									0.04		0.14						1.0
	2430	bacilysin biosynthesis oxidoreductase BacC											0.05						0.0
	2907	fatty acid transport salD											0.20						0.0
	2759, 2766	Bacterialchlorophyll			0.43						0.29		1.20				0.21		2.0
	1175	Nucleoside transport (NupC)		0.01							0.09		0.18		0.06				3.0
KB13	851	2-methylcitrate dehydratase					0.09	0.04	0.01		0.03			0.22		0.02			6.0
	622, 871	Citrate lyase	0.00	0.00			0.00				0.03		0.00	1.12				0.02	4.0
	1091	Bacterioferritin												0.08					3.0
	465	Ecotine synth. (diaminobutyrate-2-oxoglutarate AT)												0.08				0.00	4.0
	38	Methanol dehydrogenase			0.02	0.02					0.31	0.12	0.10	38.53					6.0
MS024-2A	1208	Citrate transporter													0.10				0.0
	0508, 1017, 1027-1028, 1172, 1423, 1425	TonB-dependent receptors									0.08	0.01	0.05		5.94				0.4
	0221, 0482	Na+/solute symporter													0.48	0.00	0.07		1.0
	1288	Na+-transporting NADH:ubiquinone oxid.red.				0.03					0.13	0.06	0.23	0.01	0.55				5.0
	0105, 0974	Glycosyl hydrolase		0.01											0.49				1.0
	0634	Starch synthase catalytic domain protein													0.18	0.03			1.0
	0725	Cadherin domain proteins													0.13				0.0
	1355, 1146-1148	Gliding motility proteins													0.99				0.0
	0303	Chemotaxis protein MotB													0.54				1.0
Ellin514	1326, 1983, 4765	Xylose isomerase														0.43			1.0
	2886	Capsular exopolysaccharide														0.21			0.0
	0067	Alpha-glucan phosphorylase														0.13	0.09		1.0
	3606, 3611,3988-3989	ABC-2 type transport		0.03												0.75			0.3
	2419, 4177	Type II and III secretion system proteins											0.09			1.26			1.0
	2420, 5481	Twitching motility protein												0.01		0.77			1.5
	2421, 4174	Type IV pilus assembly proteins							0.03		0.03	0.04	0.05			0.99			3.5
	5286	Secreted glycosyl hydrolase														0.11			0.0
	1073, 1254, 3059, 4724	Immunoglobulin domains (myrosinase)														0.44			0.0
WH 8109	0028, 0115, 0183, 0187-188, 220	Photosynthesis 0791, 0810, 0838, 0976, 0978, 1159, 1298, 2484															5.60		0.0
	1292	RUBISCO				0.00											1.47		1.0
	1863	2-Cys peroxiredoxin BAS1	0.00			0.00								0.05			0.15		6.0
	0458	UDP-sulfoquinovose synthase											0.02				0.27		2.0
	1514	Bicarbonate transporter															0.19		1.0
	1751	Sulfite reductase														0.09	0.19		1.0
	1424, 1960	Cell division protease FtsH															0.54		0.0
SCM1	0185, 0587, 1259, 1667	potential nitrite reductase blue copper domain																5.70	0.0
	1688, 1692-1693	H+transporting two-sector ATPase																0.63	0.0
	1479	Vitamin B6 biosynthesis protein																0.09	0.0
	0958	cobalamin B12-binding domain protein																0.10	0.0
	0694, 0811, 1314, 1728	Proteasome endopeptidase														0.01		0.45	0.3
	1500-1503, 1506-1507	Ammonia monooxygenase																9.68	0.0

The cells are shaded according to the relative proportional magnitude across a row (i.e. across genomes). The bolded, darkest shaded cells mark the genome where the indicator gene was identified. 'Indicator mean orthologs' represents the number of genome bins (of the remaining 15) that have an ortholog to the indicator gene. Non-integer numbers occur because some functions are represented by multiple genes.

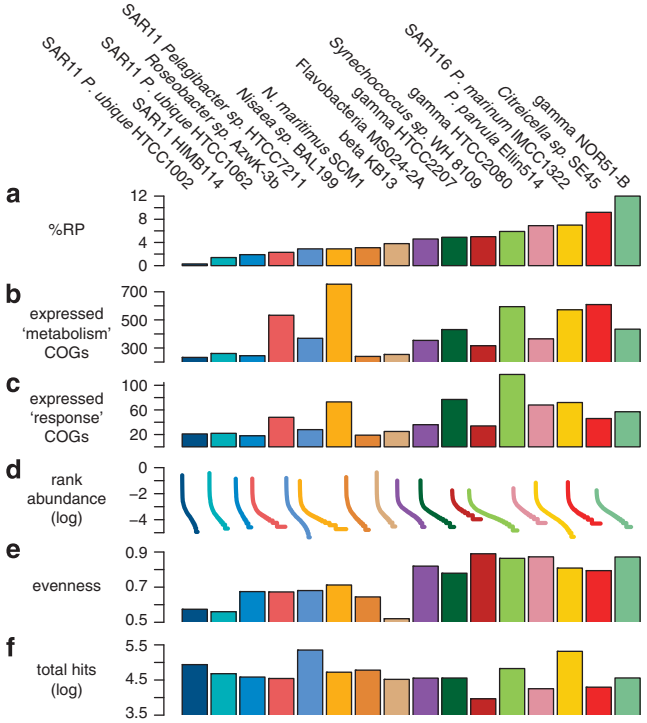


Figure 4 Potential growth rate and transcriptome composition for the 16 selected genome bins. (**a**) %RP. (**b**) Assignment to COG metabolic categories (excluding 'C' (energy production and conservation) and 'Q' (secondary metabolites)), representing metabolic diversity. (**c**) Assignment to COG cellular processes and signaling categories, 'V' (defense mechanisms),'T' (signal transduction mechanisms) and 'N' (cell motility), representing ability to sense and respond to the environment. (**d**) Rank abundance plots for expressed genes (RPs excluded). (**e**) Transcriptome evenness (Pielou, 1966) (RPs excluded). (**f**) Total reads binning to the 16 selected genomes.

(Fraiberg *et al.* 2010, 2011). There were six indicator genes for breaking glycosydic bonds, including four annotated as general glycosyl hydrolases, possibly targeting the β1–4 linkages found in cellulose, and two genes annotated as β1–3 glucanase and laminarninase, possibly targeting the β1–3 linkages of laminarin (a storage glucan found in brown algae) or chrysolaminarin (a storage glucan of diatoms), suggesting the HTCC2207 populations were binding to and degrading carbohydrate-rich particulate material.

Flavobacterium MS024-2A

Highly expressed genes in the MS024-2A bin included TonB-dependent transporters, proteorhodopsin, V-type H^+-translocating pyrophosphatase, translation elongation factor Tu and a histone family DNA-binding protein (Table 3). The ecological benchmark gene set for the MSO24-2A populations was unique in containing genes for the oxidation of hydrogen (Table 4) which, along with the expressed proteorhodopsin, can provide supplemental energy to bacteria (Woyke *et al.*, 2009).The MS024-2A populations were expressing genes for the transport of phosphorus and sulfur compounds, and were particularly enriched in transcripts for alkaline phosphatase, polyphosphate kinase (*ppk*) and polysulfide reductase (*nrfD*) (Table 4). Interestingly, there were relatively few N-related transcripts. MS024-2A populations were using a gliding motility system (*gldJMO*) likely involved in translocation across a solid surface and currently thought to be unique to the Flavobacteria (Table 4).

Similar to the Gammaproteobacteria, MS024-2A populations were also enriched in antiporters (two Na^+/H^+ exchangers, Table 4). Indicator genes included two Na^+/solute symporters and a subunit of NADH:ubiquinone oxidoreductase (Table 5). The latter gene was experimentally shown to be a component of flavobacterium MED134's light-driven growth (Kimura *et al.*, 2011), and further supports a transcriptional link between proteorhodopsin-based proton pumping and sodium-driven transport. Flavobacteria MS024-2A was similar to Gammaproteobacteria HTCC2207 in that the indicator genes included those for the attachment (cadherins) and breakdown (2 glycosyl hydrolases) of complex carbohydrates, and also included a carbohydrate synthesis gene (a glycogen synthase) (Table 5).

Betaproteobacteria

The genome bin for Betaproteobacterium strain KB13 portrayed a highly specialized chemoautotroph population. The most highly expressed gene was methanol dehydrogenase, accounting for 39% of all KB13 transcripts (Tables 3 and 5), in good

agreement with a proteomic study of coastal bacterioplankton (Sowell *et al.* 2011). Methylotrophy was likely supplemented by light-driven proton pumping, as suggested by the high expression of a xanthorhodopsin (which has not been experimentally differentiated from proteorhodopsin in KB13). Other highly expressed genes included V-type H^+-translocating pyrophosphatase (Rinta-Kanto *et al.*, 2012), a glucose/sorbosone dehydrogenase and a cytochrome *c* oxidase. An indicator gene for bacterioferritin suggests KB13 was storing iron (Table 5).

Archaea

Expression patterns in the Thaumarchaeota *Nitrosopumilus maritimus* bin was also indicative of an autotrophic specialist, with the most highly expressed genes including two subunits of ammonia monoxygenase and two ammonia transporters (Table 3). This genome bin contained the second highest proportion of transcripts devoted to phosphonate transport (Table 4; Urakawa *et al.*, 2011), as well as two *nirK*-like genes (though the exact function of the latter is uncertain, see Hollibaugh *et al.*, 2011; Table 5).

Synechococcus

The only photosynthetic organism in the 16 genomes was *Synechococcus* sp. WH8109 and, correspondingly, transcripts for light harvesting (photosystem core proteins; *psaAB*) and carbon fixation (RuBisCO) were found in both the highly expressed and indicator genes for this bin (Tables 3 and 5). Other highly expressed genes included three subunits of RNA polymerase (β, β′, γ), glutamate synthase and NADH dehydrogenase (ubiquinone). The ecological benchmark gene set revealed that WH8109 populations were expressing genes for nitrite assimilation (Table 4), and had the second highest relative expression of urease (Table 5).

Several *Synechococcus* indicator genes are involved in sulfur assimilation (Table 5), including the diagnostic gene for sulfolipid synthesis (UDP-sulfoquinovose synthase, *sqdB*). Sulfolipid-enriched membranes in open ocean Cyanobacteria have been hypothesized to decrease their phosphorus cell quota in P-limited environments (Van Mooy *et al.*, 2006, 2009), although the phosphorus concentrations in this coastal site were high for the four sampling dates (mean $0.72 \pm 0.19\,\mu M$), possibly suggesting sulfolipid synthesis is a widespread strategy.

Verrucomicrobia

Unlike all the other heterotrophs, the highly expressed genes in *Pedosphaera parvula* Ellin514 did not include any for alternative energy acquisition through phototrophy or chemolithotrophy. The most highly expressed genes included DNA-directed RNA polymerase (α, β and β′ subunits), translation elongation factors G and Tu, and chaperonin GroEL, which was similar to other high %RP taxa (SE45, IMCC1322, NOR51; Figure 1) for which many of the most highly expressed genes were related to protein synthesis (Table 3).

Few N or P transporter transcripts were caught in our ecological benchmark gene analysis, but interestingly, Ellin514 was enriched in genes for both polyphosphate (polyP kinase) and polysulphide (polyS reductase *NrfD*) metabolism.

P. parvula Ellin514 populations were highly enriched in transcripts for secretion systems (Table 4), with indicator genes including type II and III secretion system proteins and the type IV pilus assembly proteins used for gliding motility (Table 5). Three indicator genes were ABC-2 transporters, which transport polysaccharides to the outside of the cell. An indicator gene for capsular exopolysaccharide synthesis corroborates exopolymer formation activity. While there have been few phenotypic studies of Verrucomicrobia, particularly in aquatic environments, characterization of *Lentisphaera araneosa* from the sister phylum *Lentisphaerae* showed it too was an abundant producer of exopolysaccharides (Cho *et al.*, 2004; Thrash *et al.*, 2010). Indicator genes also encoded cell wall polymer degradation genes as well as homologs to myrosinases (Table 5), genes that cleave glucose from glucosinolate plant secondary metabolites. Together, the *P. parvula* indicator genes are suggestive of surface attachment and/or biofilm formation on cells or particles, possibly for a pathogenic or mutualistic lifestyle.

Synthesis

New aspects of niche differentiation based on preferential expression of genes for the transport and assimilation of organic compounds emerged from this analysis, including acetate in the SAR11s, fatty acids in Gammaproteobacteria HTCC2080 and NOR51, aromatic compounds in SAR116 and complex carbohydrates in Gammaproteobacteria HTCC2207 and Bacteroidetes MSO24-2A. Compounds not previously considered important substrates for heterotrophic bacteria but predicted here to be transported or metabolized in coastal seawater included tartrate (SAR11 HIMB114), taurine (several Alphaproteobacteria taxa), methanesulfonate (SAR116) and ectoine (SAR11 HTCC7211).

Although all selected taxa except *Synechococcus* WH8109 are heterotrophs, there was significant transcriptional effort devoted to obtaining energy by non-heterotrophic means. Expression of genes for phototrophy via AAnP and proteorhodopsin accounted for up to a quarter of some transcript bins (mean 6.5%; Table 3). A strong transcriptional pattern linked Na^+-driven transport with proteorhodopsin for multiple taxa. Chemolithotrophic energy acquisition through the *sox* system, APS reductase system and hydrogen oxidation accounted for up to 7% of bins (mean 0.5%; Table 3).

Variations in motility, adhesion and secretion systems suggested differentiation with regard to the extent of interactions with living cells or detrital particles, with the Verrucomicrobia transcriptomes, in particular, showing expression of genes associated with cell–cell interactions. *Synechococcus* WH8109 expressed genes to incorporate sulfur into membrane lipids despite the non-limiting concentrations of phosphorus. Betaproteobacterium KB13 and its relatives specialized in methanol-based C1 metabolism. Thus, the diverse assemblage of bacterioplankton is maintained to some degree by the presence of unrecognized niche dimensions involving differential gene content, differential gene regulation and transcriptional linkages between genes.

Marine bacterioplankton taxa are frequently partitioned into two super-niches believed to represent divergent adaptive strategies (Moran *et al.*, 2004; Polz *et al.*, 2006; Lauro *et al.*, 2009; Yooseph *et al.*, 2010); the first consists of cells that live singly in nutrient-poor seawater, while the second consists of cells that inhabit nutrient-rich particles, patches and eukaryotic cells suspended in the seawater matrix. Expression data from this study support the super-niche paradigms, and suggest a link between activity levels, substrate utilization and transcriptome diversity (as measured by Pielou's (1966) eveness index). Slowly growing (low %RP) taxa, such as SAR11 populations, *N. maritimus*, and betaproteobacterium KB13, exhibited low diversity transcriptomes, with the majority of transcriptional effort placed in a few key processes, such as light-mediated transport, ammonia oxidation and transport, or C1 metabolism, and relatively little investment in sensing and responding to the environment (Figure 4). These organisms may have evolved to focus on only a few metabolisms independent of environmental conditions, a strategy that potentially misses ephemeral substrates but maintains consistent growth. More rapidly growing (high %RP) taxa, such as Verrucomicrobium *P. parvula* Ellin514, SAR116 *P. marinum*, Roseobacter *Citreicella* sp. SE45 and Gammaproteobacteria NOR51 and HTCC2080 exhibited high diversity transcriptomes, with transcriptional effort spread across a variety of process such as motility, chemotaxis, defense systems and signal transduction (Figure 4). Relevant to the paradox of the plankton, some taxa showed intermediate characteristics or possessed traits from both canonical super-niches (for example, Roseobacter Azwk-3B, alpha proteobacterium BAL199 and flavobacterium MS024-2A), suggesting a continuum of ecological strategies between these two extremes (Figures 1 and 4).

Several rapidly growing taxa based on the %RP index had consistently low representation in the transcriptome relative to the other select genome bins (Figure 4f), indicating that the fastest growing bacterial groups were not necessarily the most abundant (for example, NOR51-B, *Citreicella* sp. SE45, *P. parvula* Ellin514) and that the most abundant groups were not the most active (for example, SAR11 clade members). This is consistent with recent data on estuarine bacterioplankton 16S rRNA:DNA ratios (Campbell *et al.*, 2011) and dilution experiments (Ferrera *et al.*, 2011). The assumption that bacteria capable of rapid growth show this ability only intermittently under conditions that are conducive for blooming (Yooseph *et al.*, 2010) may therefore underestimate the trophic and biogeochemical influence of these fast-growing taxa. Instead, numbers may be kept in check by top-down control within the microbial food web (Worden *et al.*, 2006; Ferrera *et al.*, 2011). Selective bacterial mortality mediated through protist grazing or viral lysis (Suttle, 2007) would lessen competitive exclusion between co-occurring heterotrophs.

These examinations of niche-defining gene expression patterns among major bacterioplankton lineages are based on four samples from southeastern US coastal waters, providing a time-averaged view of microbial activities that is not biased by one particular set of environmental conditions. However, seasonal variations in expression may offer a different, more dynamic view of a taxon's niche. Interestingly, though we detected taxon-specific seasonal shifts in %RP (Figure 1b), the most highly expressed genes were generally consistent across seasons for an individual taxon (Supplementary Table S6), suggesting that many of these genes are unlikely to resolve seasonal niche characteristics. Future work combining the indicator gene analysis with more highly resolved temporal information will provide better information on seasonal dynamics of niche differentiation in this coastal ocean.

While Hutchinson's paradox of the plankton (1961) was originally based on observations of functional overlap among tens of coexisting phytoplankton species, contemporary 16S rRNA gene surveys and metagenomic analyses suggest bacterioplankton assemblages contain orders of magnitude more coexisting taxa with many overlapping abilities (Moran, 2008). One resolution to the apparent paradox is that there is less functional overlap and more niche diversity in bacterioplankton communities than we have had the technical ability to observe. The increasing availability of reference genomes has enabled bacterial niche characterization based on unique genes, but such analyses define niches based on potential metabolic activities of reference organisms rather than on actual metabolic activities of natural populations. The 11 million transcripts sequenced in this study provide the most highly resolved catalog of realized niche dimensions yet available for hundreds of bacterioplankton taxa.

Acknowledgements

We thank S Holland for helpful discussions on statistical methods and M Moore, L Griffin, B Durham and R Newton

for assisting in sample collection. We also thank the three anonymous reviewers for their insightful and productive comments. This project was supported by funding from the Gordon and Betty Moore Foundation and the National Science Foundation Microbial Observatories Program (MCB-0702125).

References

Allers E, Gómez-Consarnau L, Pinhassi J, Gasol J, Simek K, Pernthaler J. (2007). Response of *Alteromonadaceae* and *Rhodobacteriaceae* to glucose and phosphorus manipulation in marine mesocosms. *Environ Microbiol* **9**: 2417–2429.

Béjà O, Aravind L, Koonin E, Suzuki M, Hadd A, Nguyen L *et al.* (2000). Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science* **289**: 1902–1906.

Cai WJ. (2011). Estuarine and coastal ocean carbon paradox: CO_2 sinks or sites of terrestrial carbon incineration? *Annu Rev Mar Sci* **3**: 123–145.

Campbell BJ, Yu L, Heidelberg JF, Kirchman DL. (2011). Activity of abundant and rare bacteria in a coastal ocean. *Proc Natl Acad Sci USA* **108**: 12776–12781.

Cho JC, Vergin KL, Morris RM, Giovannoni SJ. (2004). *Lentisphaera araneosa* gen. nov., sp. nov., a transparent exopolymer producing marine bacterium, and the description of a novel bacterial phylum, Lentisphaerae. *Environ Microbiol* **6**: 611–621.

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.

Cunliffe M. (2011). Correlating carbon monoxide oxidation with *cox* genes in the abundant Marine Roseobacter Clade. *ISME J* **5**: 685–691.

Cursino L, Li Y, Zaini PA, De La Fuente L, Hoch HC, Burr TJ. (2009). Twitching motility and biofilm formation are associated with tonB1 in *Xylella fastidiosa*. *FEMS Microbiol Lett* **299**: 193–199.

del Giorgio PA, Cole JJ. (1998). Bacterial growth efficiency in natural aquatic systems. *Annu Rev Ecol Syst* **29**: 503–541.

Dufrene M, Legendre P. (1997). Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecol Monogr* **67**: 345–366.

Eisen MB, Spellman PT, Brown PO, Botstein D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95**: 14863–14868.

Ferrera I, Gasol JM, Sebastian M, Hojerova E, Koblizek M. (2011). Comparison of growth rates of aerobic anoxygenic phototrophic bacteria and other bacterioplankton groups in coastal Mediterranean waters. *Appl Environ Microbiol* **77**: 7451–7458.

Fraiberg M, Borovok I, Bayer EA, Weiner RM, Lamed R. (2011). Cadherin domains in the polysaccharide-degrading marine bacterium *Saccharophagus degradans* 2-40 are carbohydrate-binding modules *J Bacteriol* **193**: 283–285.

Fraiberg M, Borovok I, Weiner RM, Lamed R. (2010). Discovery and characterization of cadherin domains in *Saccharophagus degradans* 2-40. *J Bacteriol* **192**: 1066–1074.

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, Hewson I, Schwalbach MS, Steele JA, Brown MV, Naeem S. (2006). Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci USA* **103**: 13104–13109.

García-Contreras R, Celis H, Romer I. (2004). Importance of *Rhodospirillum rubrum* H+-pyrophosphatase under low-energy conditions. *J Bacteriol* **186**: 6651–6655.

Gifford SM, Sharma S, Rinta-Kanto J, Moran MA. (2011). Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. *ISME J* **5**: 461–472.

Hardin G. (1960). Competitive exclusion principle. *Science* **131**: 1292–1297.

Hendrickson L, Liu Y, Rosas-Sandoval G, Porat I, Soll D, Whitman WB *et al.* (2008). Global responses of *Methanococcus maripaludis* to specific nutrient limitations and growth rate. *J Bacteriol* **190**: 2198–2205.

Hewson I, Poretsky RS, Dyhrman ST, Zielinski B, White AE, Tripp HJ *et al.* (2009). Microbial community gene expression within colonies of the diazotroph, *Trichodesmium*, from the Southwest Pacific Ocean. *ISME J* **3**: 1286–1300.

Hollibaugh J, Gifford SM, Sharma S, Bano N, Moran MA. (2011). Metatranscriptomic analysis of ammonia-oxidizing organisms in an estuarine bacterioplankton assemblage. *ISME J* **5**: 866–878.

Hutchinson GE. (1961). The paradox of the plankton. *Am Nat* **95**: 137–145.

Jamshad M, De Marco P, Pacheco CC, Hanczar T, Murrell JC. (2006). Identification, mutagenesis, and transcriptional analysis of the methanesulfonate transport operon of *Methylosulfonomonas methylovora*. *Appl Environ Microbiol* **72**: 276–283.

Kelly DP, Murrell JC. (1999). Microbial metabolism of methanesulfonic acid. *Arch Microbiol* **172**: 341–348.

Kimura H, Young CR, Martinez A, Delong EF. (2011). Light-induced transcriptional responses associated with proteorhodopsin-enhanced growth in a marine flavobacterium. *ISME J* **5**: 1641–1651.

Lauro FM, McDougald D, Thomas T, Williams T, Egan S, Rice S *et al.* (2009). The genomic basis of trophic strategy in marine bacteria. *Proc Natl Acad Sci USA* **106**: 15527–15533.

Lecher J, Pittelkow M, Zobel S, Bursy J, Bonig T, Smits SHJ *et al.* (2009). The crystal structure of UehA in complex with ectoine - A comparison with other TRAP-T binding proteins. *J Mol Biol* **389**: 58–73.

Maguire BA. (2009). Inhibition of bacterial ribosome assembly: a suitable drug target? *Microbiol Mol Biol Rev* **73**: 22–35.

Malik R, Viola RE. (2010). Structural characterization of tartrate dehydrogenase: a versatile enzyme catalyzing multiple reactions. *Acta Crystallogr D* **66**: 673–684.

Malmstrom RR, Cottrell MT, Elifantz H, Kirchman DL. (2005). Biomass production and assimilation of dissolved organic matter by SAR11 bacteria in the northwest Atlantic ocean. *Appl Environ Microbiol* **71**: 2979.

Marsh ME, Chang DK, King GC. (1992). Isolation and characterization of a novel acidic polysaccharide containing tartrate and glyoxylate residues from the mineralized scales of a unicellular

coccolithophorid alga *Pleurochrysis carterae*. *J Biol Chem* **267**: 20507–20512.
Meyer B, Kuever J. (2007). Molecular analysis of the distribution and phylogeny of dissimilatory adenosine-59-phosphosulfate reductase-encoding genes (aprBA) among sulfur-oxidizing prokaryotes. *Microbiology* **153**: 3478–3498.
Moran MA, Buchan A, Gonzalez JM, Heidelberg JF, Whitman WB, Kiene RP *et al.* (2004). Genome sequence of Silicibacter pomeroyi reveals adaptations to the marine environment. *Nature* **432**: 910–913.
Moran MA. (2008). Genomics and metagenomics of marine prokaryotes. In: Kirchman (ed), *Microbial Ecology of the Oceans*, 2nd edn. Wiley-Liss: New York, NY.
Mou XZ, Sun SL, Edwards RA, Hodson RE, Moran MA. (2008). Bacterial carbon processing by generalist species in the coastal ocean. *Nature* **451**: 708–711.
Mulligan C, Fischer M, Thomas GH. (2011). Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea. *FEMS Microbiol Rev* **35**: 68–86.
Pielou E. (1966). The measurement of diversity in different types of biological collections. *J Theor Biol* **13**: 131–144.
Polz MF, Hunt DE, Preheim SP, Weinreich DM. (2006). Patterns and mechanisms of genetic and phenotypic differentiation in marine microbes. *Philos Trans R Soc B* **361**: 2009–2021.
Pomeroy LJ, Wiegert RG (ed) (1981). *The Ecology of a Salt Marsh*. Springer: New York, NY.
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.
Poretsky RS, Gifford SM, Rinta-Kanto J, Vila-Costa M, Moran MA. (2009). Analyzing gene expression from marine microbial communities using environmental transcriptomics. *J Vis Exp* **24**: pii.1086.
Poretsky RS, Sun S, Mou X, Moran MA. (2010). Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. *Environ Microbiol* **12**: 616–627.
Rinta-Kanto JM, Sun S, Sharma S, Kiene RP, Moran MA. (2012). Bacterial community transcription patterns during a marine phytoplankton bloom. *Environ Microbiol* **14**: 228–239.
Schauer K, Rodionov DA, de Reuse H. (2008). New substrates for TonB-dependent transport: do we only see the 'tip of the iceberg'? *Trends Biochem Sci* **33**: 330–338.
Sowell SM, Abraham PE, Shah M, Verberkmoes NC, Smith DP, Barofsky DF *et al.* (2011). Environmental proteomics of microbial plankton in a highly productive coastal upwelling system. *ISME J* **5**: 856–865.
Stewart FJ, Sharma AK, Bryant JA, Eppley JM, DeLong EF. (2011). Community transcriptomics reveals universal patterns of protein sequence conservation in natural microbial communities. *Genome Biol* **12**: R26.
Suttle CA. (2007). Marine viruses - major players in the global ecosystem. *Nat Rev Microbiol* **5**: 801–812.
Teira E, Martinez-Garcia S, Lonberg C, Alvarez-Salgado XA. (2009). Growth rates of different phylogenetic bacterioplankton groups in a coastal upwelling system. *Environ Microbiol Rep* **1**: 545–554.
Thrash JC, Cho JC, Vergin KL, Morris RM, Giovannoni SJ. (2010). Genome sequence of *Lentisphaera araneosa* HTCC2155(T), the type species of the order Lentisphaerales in the phylum Lentisphaerae. *J Bacteriol* **192**: 2938–2939.
Tripp HJ, Kitner JB, Schwalbach MS, Dacey JWH, Wilhelm LJ, Giovannoni SJ. (2008). SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature* **452**: 741–744.
Tripp HJ, Schwalbach MS, Meyer MM, Kitner JB, Breaker RR, Giovannoni SJ. (2009). Unique glycine-activated riboswitch linked to glycine-serine auxotrophy in SAR11. *Environ Microbiol* **11**: 230–238.
Urakawa H, Martens-Habbena W, Stahl DA. (2011). Physiology and genomics of ammonia-oxidizing Archaea. In: Ward BB, Arp DJ, Klotz MG (eds.) *Nitrification*. ASM Press: Washington, DC.
Van Mooy BAS, Fredricks HF, Pedler BE, Dyhrman ST, Karl DM, Koblizek M *et al.* (2009). Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. *Nature* **458**: 69–72.
Van Mooy BAS, 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.
Wei Y, Lee JM, Richmond C, Blattner FR, Rafalski JA, LaRossa RA. (2001). High-density microarray-mediated gene expression profiling of *Escherichia coli*. *J Bacteriol* **183**: 545–556.
Wilson DN, Nierhaus KH. (2007). The weird and wonderful world of bacterial ribosome regulation. *Crit Rev Biochem Mol Biol* **42**: 187–219.
Worden AZ, Seidel M, Smriga S, Wick A, Malfatti F, Bartlett D *et al.* (2006). Trophic regulation of *Vibrio cholerae* in coastal marine waters. *Environ Microbiol* **8**: 21–29.
Woyke T, Xie G, Copeland A, González JM, Han C, Kiss H *et al.* (2009). Assembling the marine metagenome, one cell at a time. *PLoS One* **4**: e5299.
Yokokawa T, Nagata T, Cottrell MT, Kirchman DL. (2004). Growth rate of the major phylogenetic bacterial groups in the Delaware estuary. *Limnol Oceanogr* **49**: 1620–1629.
Yooseph S, Nealson KH, Rusch DB, McCrow JP, Dupont CL, Kim M *et al.* (2010). Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* **468**: 60–66.

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