Limnol. Oceanogr., 54(6), 2009, 1981–1994

Metagenomic potential of microbial assemblages in the surface waters of the central Pacific Ocean tracks variability in oceanic habitat

Ian Hewson,[a,1,*] Ryan W. Paerl,[a] H. James Tripp,[a] Jonathan P. Zehr,[a] and David M. Karl[b]

[a] Department of Ocean Sciences, University of California Santa Cruz, Santa Cruz, California
[b] School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, Hawaii

Abstract

Oceanic habitats may select for different organisms, thereby tuning genomic capabilities to local environmental conditions. To understand the relationship between microbial assemblage composition, functional capability, and habitat, a random genome shotgun sequencing (metagenomic) survey was conducted with surface-water microbial assemblages (0.2–5-μm size fraction) collected at seven locations along a meridional transect from the northern edge of the South Pacific subtropical gyre to the southern edge of the North Pacific subtropical gyre (16°S–13.5°N). A total of 1.1 million unique sequence reads were obtained, of which ~45% could be annotated to metabolic category. Microbial assemblages in equatorial divergence and countercurrent habitats were distinct phylogenetically from those in gyre waters. Ecotypes of dominant Cyanobacteria (*Prochlorococcus* and *Synechococcus*) had distinct distributions congruent with their physiological characteristics in cultivation. The metagenomic distribution of genes among metabolic pathways was very similar at all stations despite phylogenetic differences, but was unrelated to physicochemical habitat, suggesting that dominant microorganisms have a core suite of genes necessary for life in the open ocean. Among metabolic genes that varied across the transect, several patterns were observed. For example, phosphate (PO_4^{3-}) stress response genes were more common in gyre waters than at the equator. The variability in frequency of several metabolic pathways (e.g., chlorophyll biosynthesis, PO_4^{3-} metabolism, and transcription initiation bacterial sigma factors) was related to physicochemical conditions, most of which were related to taxonomic differences among habitats. Microbial communities in the central Pacific Ocean are phylogenetically distinct to the oceanic provinces which they inhabit.

Past research has elucidated the critical roles of marine microorganisms in primary production, in net remineralization of organic into inorganic compounds, and in the global carbon cycle (Williams 1981; Azam et al. 1983). Microorganisms in surface waters of the open ocean are limited primarily by the availability of inorganic nutrients (Wells et al. 1994; Karl et al. 1996; Sañudo-Wilhelmy et al. 2001). In open-ocean waters away from coastal influence, N_2 fixation and advective upwelling support the N and P demands of phytoplankton, and result in higher biomass than waters lacking these processes (Carpenter 1983; Capone 2000; Karl et al. 2002). The enhanced primary productivity in these waters fuels microbial food webs, resulting in enhanced heterotrophic activity, top-down processes from higher trophic levels, and particle flux (Karl et al. 1996).

The central Pacific Ocean contains several provinces, including the North Pacific subtropical gyre, the South Pacific subtropical gyre, the Pacific equatorial divergence, and the North Pacific equatorial countercurrent provinces (Longhurst 2006). The equatorial divergence and countercurrent zones represent warm, nutrient-rich environments with seasonally invariant high light intensity, leading to high rates of primary and secondary production (Kirchman and Rich 1997). In contrast, the North and South Pacific tropical and subtropical gyres contain a paucity of inorganic nutrients (particularly N and P species), which leads to low abundance and production by phytoplankton (Karl et al. 1996). Activities of N_2-fixing microorganisms within these habitats support the demands of new primary production (Carpenter 1983; Capone et al. 1997; Zehr et al. 2007). Additionally, nutrition of microbial assemblages is variable between equatorial and gyre waters, with higher utilization of glucose and ammonium (NH_4^+) at the equator but more utilization of proteins in more subtropical waters, reflecting de novo synthesis of substrates (Christian and Karl 1995). Moreover, the extent to which the ecological geography of the ocean (Longhurst 2006) is reflected in the assemblage genomic capabilities for biogeochemical processes remains largely unknown.

Previous studies examining community variability using conserved genes or intergenic regions indicated correspondence between habitat and microbial assemblage composition in open-ocean plankton (Moeseneder et al. 2001; Hewson et al. 2006*a*,*b*), closely following seasonal cycles of solar irradiation (Fuhrman et al. 2006) and diversity of assemblages following global latitudinal gradients (Fuhrman et al. 2008). However, organisms with similar conserved gene sequences (Rocap et al. 2003) can have variable biological and ecological characteristics, particularly closely related "microdiverse" clades of microorganisms where diversification is driven by neutral selection (Acinas et al. 2004). It is unclear how diversity within communities corresponds with functional diversity, or what factors drive habitat specificity of microbial communities (Moeseneder et al. 2001).

Studies employing metagenomic approaches (i.e., whole community genome shotgun sequencing) have documented

* Corresponding author: hewson@cornell.edu
[1] Present address: Department of Microbiology, Cornell University, Ithaca, New York

the extraordinary diversity of microorganisms in the ocean (Breitbart et al. 2002; Venter et al. 2004; Rusch et al. 2007). In addition to revealing the presence and diversity of genetic capabilities not previously reported to occur in the oceans, e.g., rhodopsins (Beja et al. 2000), metagenomic analyses applied to samples collected from geographically diverse locations have elucidated "signature genes" for disparate habitats (Tringe et al. 2005; Angly et al. 2006). The genomic complement of microbial assemblages presents a valuable tool for interrogating assemblages for their function. Selection in the ocean can lead to highly streamlined genomes (Rocap et al. 2003; Giovannoni et al. 2005) where less- or underutilized genes are lost over evolutionary timescales. Although not indicating the activity of a detected gene, the presence of genes, by proxy, demonstrates their potential use in an environment (Tyson et al. 2004).

The aim of this study was to examine whether habitat variability across ecological provinces corresponds to differences in the functional genomic capabilities of microbial assemblages in surface waters of the open ocean. We examined the variability in assemblage gene content along a meridional transect from the northern edge of the South Pacific subtropical gyre to the southern edge of the North Pacific subtropical gyre (16°S–13.5°N), crossing the high-productivity equatorial divergence and countercurrent zones. We examined metagenomes for the presence of functional genes and genes from isolated taxa with known geochemical capabilities that corresponded to ecological provinces and their associated biological, physical, and chemical features. We found that the overall metabolic potential of metagenomes across the transect was homogeneous. However, the geographical patterns of whole microbial assemblage composition and distribution of individual ecotypes were congruent with ecological provinces and reflect variability in microbial nutritional sources.

Methods

Sample collection—Samples for metagenomic analyses were collected at seven stations between Fiji and Hawaii on board the R/V *Kilo Moana* during the Center for Microbial Oceanography: Research and Education "Biogeochemistry of the Upper Ocean: Latitudinal Assessment" (CMORE-BULA) cruise KM0704 in April 2007 (Fig. 1). Seawater (120 liters) was obtained by a pneumatic pump, with an intake hose lowered to ~5-m depth off the stern of the research vessel. The pump tubing was rinsed with deionized water between collections and was purged with approximately 40 L min^{-1} for 5 min at each station prior to sample collection. Seawater was then pumped into two sample-rinsed 60-liter polypropylene carboys for onboard processing.

The carboys were immediately filtered using positive air pressure (~10 mm Hg) through a 5-μm Isopore (Whatman) filter and then through a 0.22-μm Durapore (Whatman) filter. Processing of each 120-liter sample through a single filter took approximately 1 h. At the conclusion of processing, the filters were folded and placed into sterile Whirlpak bags, then frozen in liquid nitrogen for transport to the laboratory at the University of California Santa Cruz.

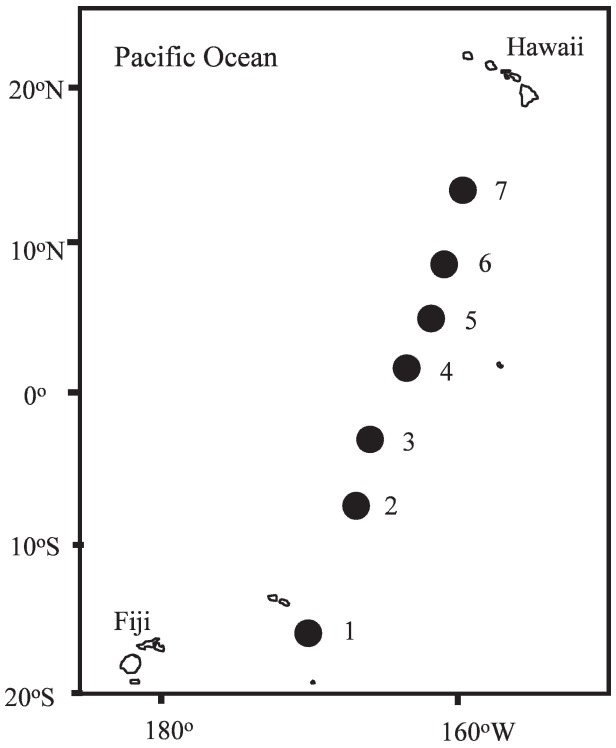

Fig. 1. Sampling locations on cruise KM0704 (CMORE-BULA), 21–27 April 2007 between the Fiji Islands and the Hawaiian Archipelago.

Habitat characteristics—Physicochemical conditions (temperature, salinity, and oxygen concentration) were measured using a conductivity-temperature-depth (CTD) sensor array mounted on a Niskin bottle rosette. Vertical profiles from 0 to 200 m were obtained at each station within 2 h of sample collection. In vivo chloropigment fluorescence measured using a CTD-mounted fluorometer was calibrated at each station using values from discrete water sample–extracted chlorophyll *a* (Chl *a*) and contoured along the transect. High-performance liquid chromatography–measured phytopigments, inorganic and organic nutrient concentrations, and the abundance of picophytoplankton via flow cytometry were examined by standard methods as described on the cruise website (http://hahana.soest.hawaii.edu/cmorebula/cmorebula.html), where all ancillary data can be obtained.

Deoxyribonucleic acid (DNA) extraction and preparation—DNA was extracted following the phenol- and chloroform-based protocol of Fuhrman et al. (1988). Frozen 0.22-μm filters were first manually broken up in the Whirlpak bags, then filter pieces transferred into sterile 50-mL conical tubes (Falcon). Ten milliliters extraction buffer (10% [w/v] sodium dodecyl sulfate) : STE (0.1 mol L^{-1} NaCl, 10 mmol L^{-1} Tris-HCl, 1 mmol L^{-1} ethylenediaminetetraacetic acid), 9 : 1 was added to the tubes and boiled in a water bath for 5 min. The extraction buffer was

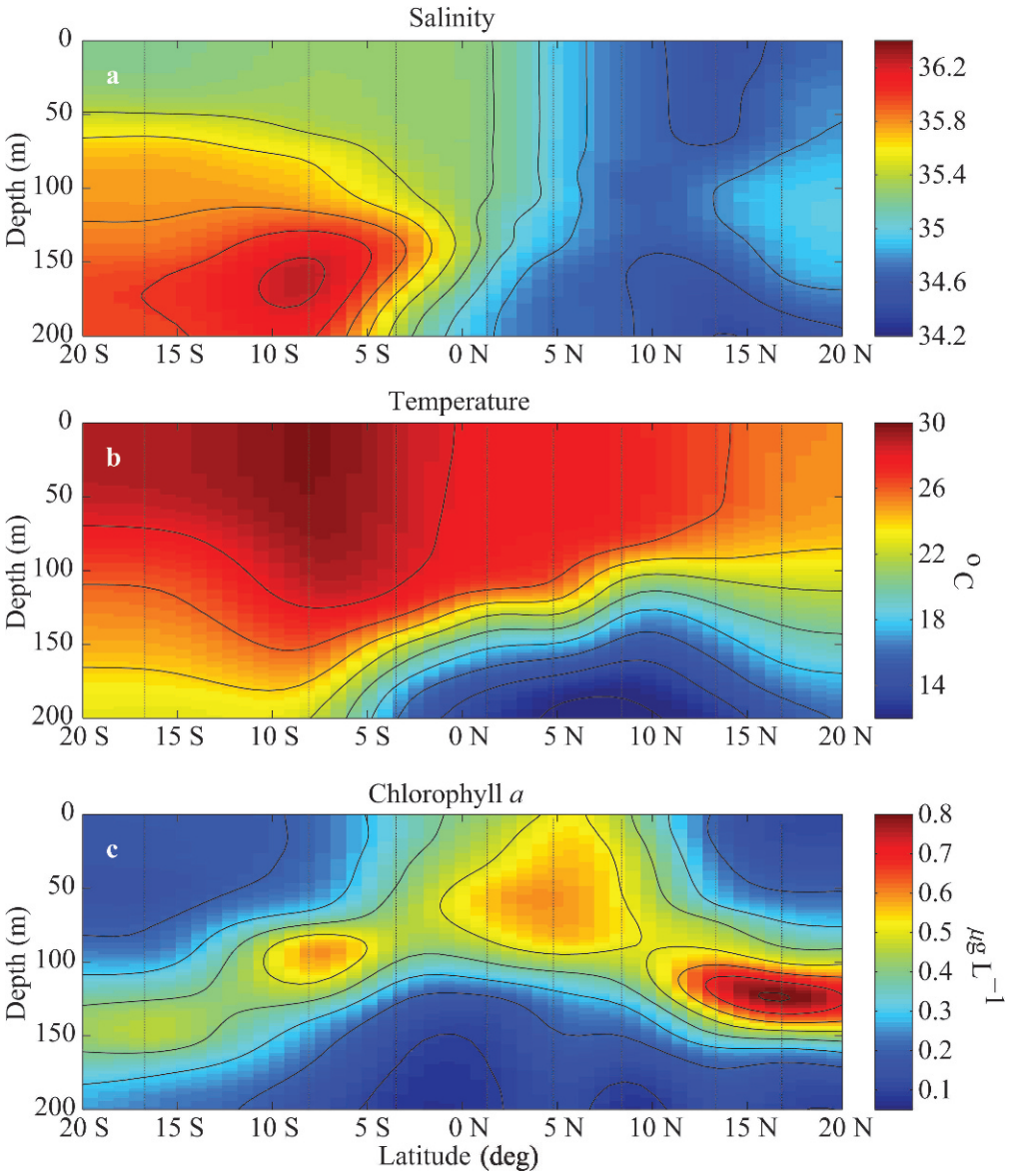

Fig. 2. Contour plot of (a) salinity, (b) temperature, and (c) Chl *a* from 0 to 200 m across the transect. Parameters were measured by CTD and are freely available at http://hahana.soest.hawaii.edu/cmorebula/cmorebula.html.

then removed from the tubes, placed into round-bottom centrifuge tubes (Oak Ridge); to that, 3 mL 10.5 mol L^{-1} NaOAc and 28 mL 100% EtOH were added. Organic macromolecules were precipitated overnight at −20°C before the tubes were centrifuged at 15,000 × *g*. The EtOH + NaOAc was decanted, and pellets dried for 30 min in the air. The pellets were then resuspended in 500 μL TE (Tris-EDTA) buffer, and sequentially extracted in microcentrifuge tubes with 500 μL phenol, 500 μL phenol : chloroform : i-soamyl alcohol (24 : 1 : 0.1), and 500 μL chloroform : isoamyl alcohol (9 : 1), each time the bottom layer was removed and discarded. The aqueous layer was removed into a new tube at the end of the last extraction. This was amended with 150 μL 10.5 mol L^{-1} NaOAc and 1.2 mL 100% EtOH and precipitated overnight. The tubes were then centrifuged at 15,000 × *g* and the supernatant decanted, and pellets dried in a speed vacuum dryer for 10 min. The DNA pellets were resuspended in 150 μL deoxyribonuclease- and ribonuclease-free deionized water (Ambion).

DNA was initially quantified using a NanoDrop spectrophotometer, then 50 ng run on Agilent Bioanalyzer DNA-1000 chips to size DNA. A total of 5 μg of the DNA extract from the 0.2–5-μm size fraction at each station was then sequenced using picoliter reactor pyrosequencing technology (Margulies et al. 2005) on a GS-FLX (Roche) instrument. Additionally, 2 μg of DNA extract from the >5-μm size fraction was sequenced at Sta. 5.

Bioinformatic analyses—Sequences <100 base pairs (bp), sequences containing >60% of each nucleotide, and reads

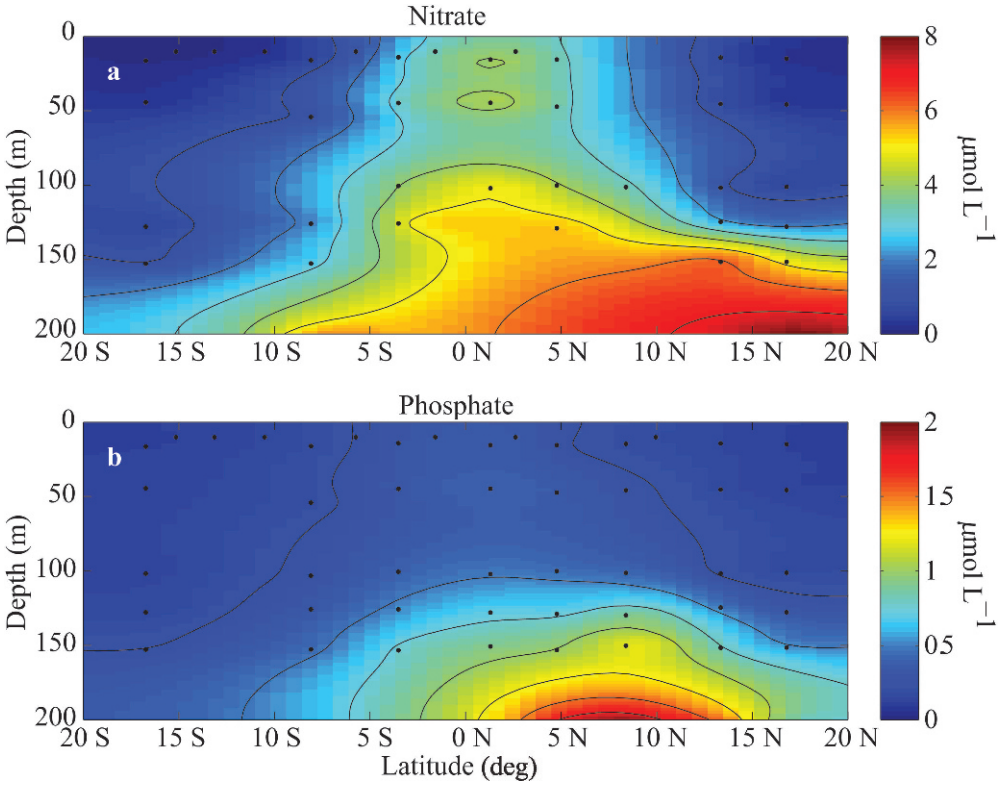

Fig. 3. Contour plot of (a) nitrate and (b) phosphate across the transect. Data and methods used are available at http://hahana.soest.hawaii.edu/cmorebula/cmorebula.html.

that were exactly the same in the same library over the first 100 bp of sequence were removed from metagenomic libraries (in the latter case leaving one sequence to represent all replicates). Resulting edited libraries were compared using the BLASTx program (translated nucleotide query–protein database) against the All Prokaryotic Proteins, Eukaryotic Microbial Proteins, and Viral Proteins databases at the Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis (CAMERA) using an E-value cutoff of $E < 0.001$ to identify putative protein-encoding reads. Where the same sequence matched different kingdoms, the highest bit-score match was used. A relational database of the BLASTx annotation was constructed in Microsoft Access. This database was queried (keyword search of subject line) for genes involved in biogeochemical processes: P substitution S-lipid genes (*sqdX*, uracil diphosphate-sulfoquinovose biosynthesis protein, glycosyl transferase protein); alkaline phosphatases (alkaline phosphatase, *DedA*, phosphodiesterase) ; ammonium transporters; nitrate reduction genes (nitrate reductase, glycerol trinitrate reductase); nitrogenases; phosphonate utilization genes (phosphonate lyases, phosphonate binding genes, phosphonate transport genes); nucleotidases (nucleotidases, acid phosphatases); and high-affinity phosphate transporters (*PstS*, adenosine triphosphate [ATP]–binding cassette phosphate transporters). Gene annotation lists for these processes were built based upon the Kyoto Encyclopedia for Genes and Genomes metabolic maps (Kanehisa et al. 2004), and keyword searches of BLASTx annotations were conducted using Microsoft Access. Because of differences in number of sequences per library, and differences in length between genes, gene abundance was standardized for this analysis by first dividing the number of matches by the length of each gene, then standardizing to the number of sequences per library. Strain and species variation among the dominant Cyanobacteria *Prochlorococcus marinus* and *Synechococcus* sp., respectively, were investigated by enumerating sequence reads matching each strain as a best BLASTx hit (i.e., that recruited to the genomes).

Subsystem technology was used to assign the phylogenetic identification of sequence reads at all locations using the Meta Genome Rapid Annotation using Subsystems Technology (MG-RAST) (Meyer et al. 2008). Phylogeny was assigned by comparing the sequences against the SEED database (Overbeek et al. 2005) and using any alignment length and maximum E-value of 0.01. Phylogeny was also compared by comparison against the Ribosomal Database Project (RDP) using MG-RAST using a E-value cutoff of 0.01. Finally, phylogeny of eukaryotic small subunit sequences was further resolved by comparison against the Silva database using MG-RAST using an E-value cutoff of 0.01. Sequence reads were assigned to metabolic and physiological subsystem pathways by MG-RAST using a maximum E-value cutoff of 0.01. The phylogenetic and subsystem patterns between stations were investigated by calculation of the Whittaker Index of similarity (Whittaker 1952) as described elsewhere (Hewson et al. 2006*b*). Metabolic and phylogenetic distribution between metage-

Table 1. Compositional characteristics of metagenomic libraries prepared from Stas. 1–7. Dereplicated reads were defined as those >100 bp, not containing >60% of any base, and unique over the first 100 bp of sequence.

Station	Size fraction (μm)	No. reads	No. dereplicated reads	Total nucleotide length (bp)	Mean read length	SD	Mean % G + C	SD
1	0.2–5	212,751	173,472	47,028,434	221	50	47.1	14.6
2	0.2–5	90,953	79,445	23,488,767	258	31	43.3	12.7
3	0.2–5	71,430	61,452	18,648,851	261	30	42.0	12.6
4	0.2–5	292,907	226,972	71,868,035	245	45	45.3	14.2
5	0.2–5	283,485	240,048	68,060,952	240	44	49.5	13.6
6	0.2–5	60,834	53,074	14,784,612	243	46	42.8	13.2
7	0.2–5	227,182	180,274	54,601,210	240	44	44.5	13.5
5	>5	75,048	55,040	17,032,429	227	50	53.5	12.1

nomes were compared by constructing dendrograms of similarity using unweighted pair-group mean average (UPGMA) using the XLStat program (Addinsoft Software Architecture Representation Language).

Statistical analysis of pathways—The pathway frequencies were compared to physicochemical data (temperature, salinity, oxygen concentration, PO_4^{3-} concentration, NO_3^- concentration, PO_4^{3-} and NO_3- concentrations measured by high-sensitivity protocols, particulate C, particulate N, particulate Si, Chl *a* fluorescence, extracted Chl *a* concentration, ATP concentration, CH_4, and N_2O concentrations) using two approaches. The first was comparison of pathways comprising >1% of all messenger ribonucleic acids (mRNAs) against physicochemical parameters using multiple regression. The second approach was first transforming physicochemical variables using the equation

$$z_n = (n - n_{ave})/p \quad (1)$$

where z is the transformed variable, n is the variable, and p is the range of n across all observations. The similarity between stations was then calculated by Pearson correlation coefficient, and compared to the Whittaker similarity index between stations based on pathway distribution by Mantel test using XLStat (Addinsoft SARL).

Results

Oceanographic conditions—Stations had heterogeneous salinity and temperature, with highest temperature and salinity south of the equator (16°S and 8°S) and lowest closer to Hawaii (13°N; Fig. 2). Vertical profiles from 0 to

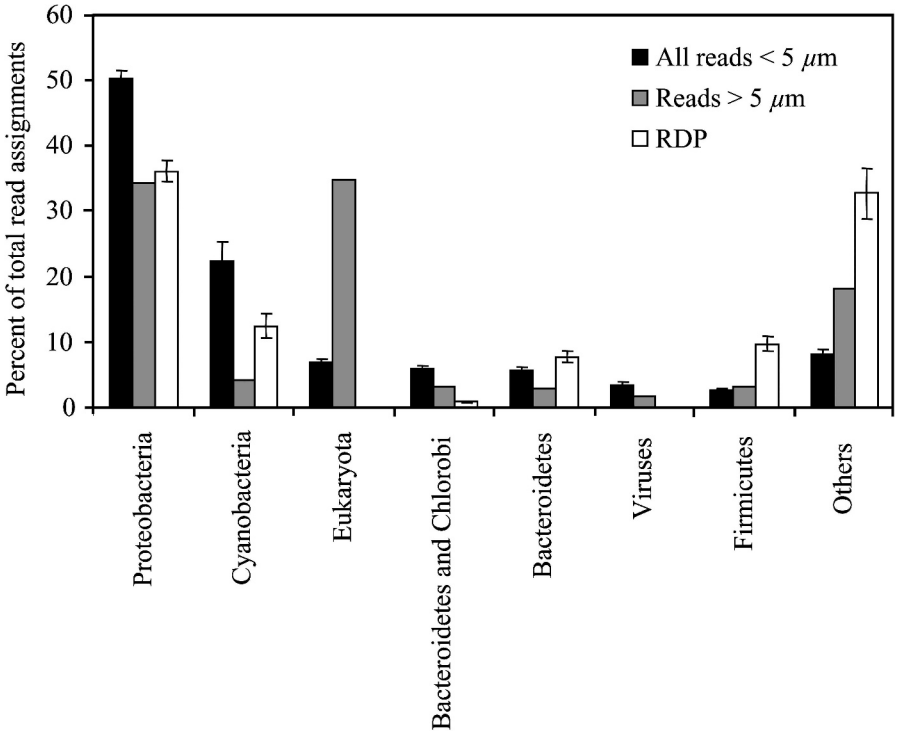

Fig. 4. Phylogenetic affiliation of all sequence reads assigned using subsystem technology comparisons against SEED (Overbeek et al. 2005), and based upon BLASTn comparison against the Ribosomal Database Project II with an E-value cutoff of 0.01. Error bars = SE.

Table 2. Summary of metagenomic annotation: based on BLASTx comparison against protein databases at the Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis using an E-value cutoff of $E < 0.001$; assigned to phylogeny and subsystems using MG-RAST, and matching the RDP database at an E-value of <0.01.

Station	Size fraction (μm)	Prokaryotic protein matches (% matches)	Eukaryotic protein matches (% matches)	Viral protein matches (% matches)	Reads not matching proteins (%)	Reads matching subsystems (%)	Reads matching RDP (%)*
1	0.2–5	90	7	3	55	36	0.38
2	0.2–5	90	5	5	58	42	0.41
3	0.2–5	92	4	4	33	43	0.39
4	0.2–5	89	9	2	56	36	0.33
5	0.2–5	88	10	2	61	32	0.38
6	0.2–5	88	7	5	52	37	0.28
7	0.2–5	84	13	3	46	39	0.48
5	>5	74	22	4	83	11	0.18

* RDP, Ribosomal Database Project.

200 m indicated a shoaling of the 18°C isotherm at 2°N and 5°N; however, there was no strong evidence for advective upwelling along the cruise track. The concentrations of inorganic nutrients (NO_3^- and PO_4^{3-}) were lowest at 16°S and 13°N, and highest at 3°S and 2°N at the equator (Fig. 3). Chl *a* in surface waters was elevated at stations immediately north–south of the equator (Pacific equatorial divergence and countercurrent provinces), with greatest concentration at 5°N in the divergence zone, and least at the distal ends of the transect in gyre waters (16°S and 13°N). The deep chlorophyll maximum shoaled from 150 m to 68 m between gyre and equatorial waters. Based on physicochemical features, stations at 16°S, 8°S, and 13°N were considered gyre stations (in the North and South Pacific Subtropical Gyres), stations at 2°N and 5°N considered equatorial stations (Pacific equatorial divergence and countercurrent, respectively) and stations at 3°S and 9°N intermediates between these provinces.

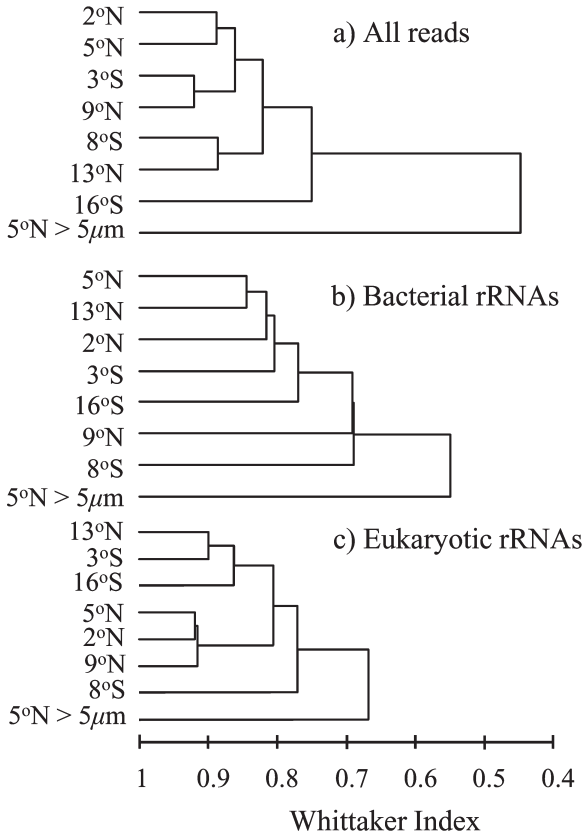

Fig. 5. Similarity between metagenome phylogenies based upon subsystems technology assignment of (a) all sequences and (b, c) 16S rRNA comparison only. Clustering was performed by unweighted pair-group mean average (UPGMA) and is based upon the Whittaker Index (Whittaker 1952). For both comparisons, organism-level phylogeny was used (i.e., species annotation).

Library characteristics—The sequencing effort yielded 1,314,950 reads representing 315,513,288 bp of community nucleotide sequence information. Of total reads, 1,069,777 (81%) were unique and >100 bp. Edited libraries from the 0.2–5-μm size fraction samples had average read lengths of 221–262 bp and average guanine + cytosine contents of 42–50%, whereas the library prepared from the >5-μm size fraction had an average G + C content of 53% (Table 1).

BLASTx comparison against prokaryotic, eukaryotic, and viral proteins revealed that 33–61% (mean 45%) of all edited reads in the 0.2–5 μm size fraction matched proteins in sequenced microbial genomes (Table 2). Of these, the majority of matches were to prokaryotic proteins (84–92%), with fewer eukaryotic (4–13%) and viral (2–5%) proteins. In the library of >5-μm–size fraction DNA, few reads matched proteins (17%). Eukaryotic matches comprised a larger fraction of total hits (22%) than at other stations.

Phylogeny of metagenomes—Phylogeny of metagenome libraries was examined by two approaches (Fig. 4). First, comparison against the Ribosomal Database Project II (RDP II) using the BLASTn program (nucleotide query–nucleotide database) with an E-value cutoff of 0.01 was used to assign phylogeny of reads matching bacterial and archaeal 16S ribosomal ribonucleic acids (rRNAs). Between 98 and 917 reads per library matched databased 16S rRNAs (Fig. 4). Of these, most were proteobacterial (34%) and cyanobacterial (12%) including a large proportion of unclassified bacteria (25%), which presumably include *Candidatus Pelagibacter ubique* because it has not been

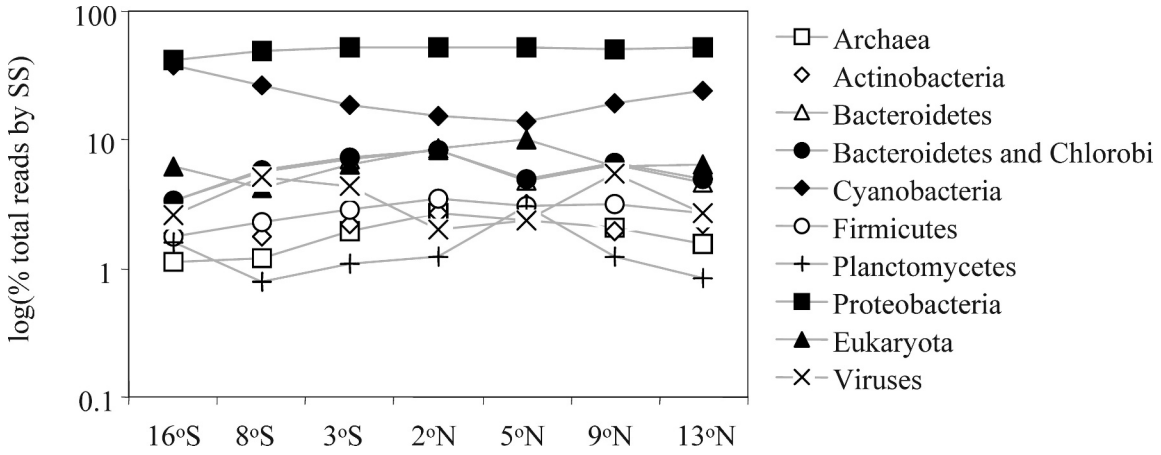

Fig. 6. Percentage of subsystems associated with phyla over the transect. Only phyla comprising >1% of all sequences are shown.

fully characterized taxonomically. No archaeal rRNAs were detected by RDP comparison. Eukaryotic 18S rRNAs were classified by comparison by BLASTn against the Silva database with an E-value cutoff of 0.01. Most eukaryotic 18S rRNAs in the 0.2–5-μm size fraction were fungal (15%), metazoan (7%), or alveolates (1.6%).

The second approach to assign phylogeny was annotation of all sequence reads based on subsystem technology against SEED using MG-RAST. Between 32% and 43% of edited libraries in the 0.2–5-μm size fraction could be assigned phylogenetic affiliation using subsystem technology, whereas only 11% of edited sequences in the >5-μm size fraction matched SEED sequences (Table 2). The largest components of SEED-assigned matches were to Proteobacteria (50%) and Cyanobacteria (22%). Eukaryotic sequences comprised 7% of sequence reads in the 0.2–5-μm size fraction and 34% of sequence reads in the >5-μm library (Fig. 4). Unlike RDP comparisons, SEED matches included 1.9% to Archaea. Notably, Sta. 1 had a very large fraction of matches (1007 sequence reads or 1.2% of all sequences) to the diazotrophic unicellular cyanobacterium *Crocosphaera watsonii* WH8501.

Clustering of metagenomes based upon bacterial and archaeal RDP, eukaryotic Silva, and SEED comparisons indicated that the >5-μm size sample shared the least similarity with other libraries (Fig. 5). The 0.2–5-μm size fraction libraries shared on average 74% and 82% of the taxonomic distribution based on RDP and SEED-assigned phylogeny, respectively. Clustering based on bacterial and archaeal RDP and Silva compositions demonstrated little pattern with habitat. However, clustering based on phylogeny of all sequence reads showed that stations with similar habitat shared similar types of organisms. Equatorial stations were more similar based on phylogeny than they were to gyre or intermediate stations. The variability in metagenome phylogenetic composition mostly reflected the variable abundance of Cyanobacteria and eukaryotes (Fig. 6).

The number of sequence reads recruiting by BLASTx to the proteins of dominant cyanobacterial (*Prochlorococcus* and *Synechococcus*) strains was compared between stations (Fig. 7). Eight strains of *Prochlorococcus* and two strains of *Synechococcus* had greater frequency at the equator than in gyre waters, whereas three strains of *Prochlorococcus* and four strains of *Synechococcus* had more sequences in gyre waters than at the equator. The remaining reads recruiting on picocyanobacterial genomes (28 strains of *Synechococcus* and 12 strains of *Prochlorococcus*) either had no discernible distribution along the transect or comprised <0.5% of all hits to the respective genera.

Metabolic classification—Sequence reads were assigned using subsystem analysis to metabolic categories (Overbeek et al. 2005). Metagenomes were dominated by carbohydrate, protein metabolism, and amino acids and derivative reads as dominant features (Table 3). The metagenomes shared a very high similarity of metabolic distribution (95% ± 1%; Whittaker Index) among subsystems between locations in the 0.2–5-μm size fraction (Fig. 8). The >5-μm size fraction was least similar in subsystem distribution to the metagenomes from the 0.2–5-μm fraction, whereas the pattern of clustering among metagenomes based on metabolic distribution in the smaller size fraction did not demonstrate any clear pattern with geographic location.

Biogeochemically relevant genes—The frequencies of genes encoding for several biogeochemically important processes within the nitrogen, phosphorus, and carbon cycles were determined based on annotation in the proteins database. Phosphorus cycle genes encoding for sulfolipid formation, high-affinity phosphate transporter, phosphonate utilization, and alkaline phosphatase activity had a higher abundance away from the equator, whereas nitrogen cycle genes encoding for ammonium uptake had lower abundance at the equatorial stations. However, nucleotidases and nitrate uptake genes had no discernible geographic pattern (Fig. 9).

Subsystems involved in uptake and utilization of carbohydrates and amino acids and proteins had also had no discernible pattern along the transect.

Relationship between pathways and physicochemical parameters—The similarity between stations based upon pathway distribution and similarity based on physicochem-

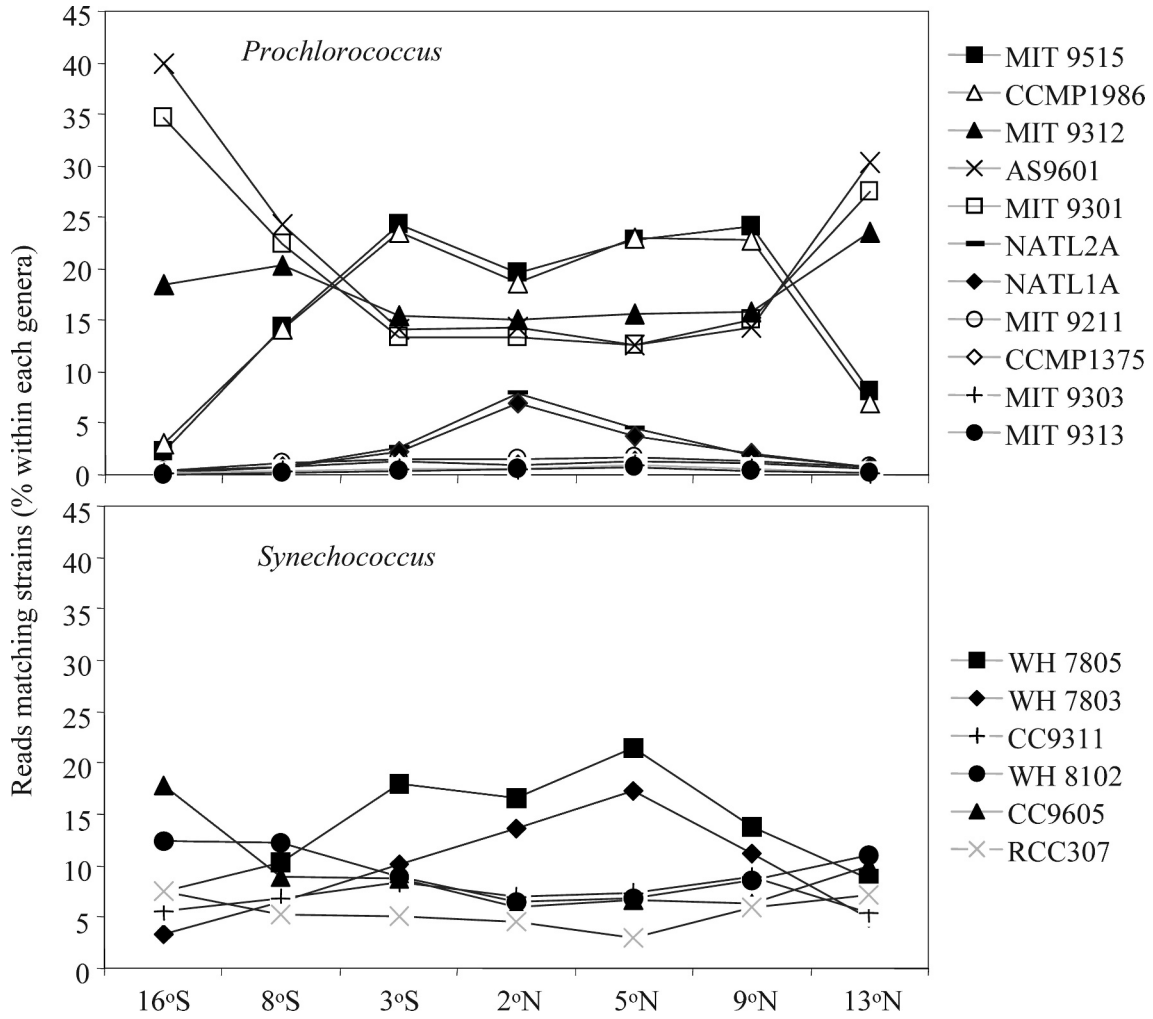

Fig. 7. Distribution of reads matching by BLASTx ($E < 0.001$) proteins of *Synechococcus* and *Prochlorococcus* strains across the transect as a percentage of matches to each genus.

ical conditions were not significantly related (Mantel test). However, a total of 16 pathways, each comprising >1% of all pathways (30 total), were significantly ($p < 0.05$) related to either single physicochemical parameters or combinations of multiple parameters (Table 4). Particulate silica concentration explained the majority of pathway distribution (six pathways). Particulate silica was significantly ($p < 0.05$) correlated with PO_4^{3-} ($r^2 = 0.79$), NO_3^- ($r^2 = 0.76$), particulate C ($r^2 = 0.80$) and particulate N ($r^2 = 0.87$).

Discussion

Our results demonstrate that the variability of oceanic habitats is reflected in the pattern of genes and genomes of the microorganisms that inhabit them, supporting the concept of geographically defined ecological provinces in the region (Longhurst 2006). The taxonomic composition of microbial assemblages, examined by community gene complement and the distribution of picocyanobacterial ecotypes, demonstrated biogeographical patterns across the equatorial divergence and countercurrent waters and within gyre waters. Metagenomes shared similar distributions of randomly sequenced genome fragments among metabolic and physiological pathways with most genes involved in carbohydrate, protein, and amino acid metabolism. Some key genes involved in nitrogen and phosphorus cycles in the open ocean varied in a pattern consistent with habitat and the distribution of inorganic and organic nutrients, whereas others had no discernible pattern. Finally, the library prepared from the >5-μm size fraction was distinct in both taxonomic composition and metabolic distribution from the 0.2–5-μm size fraction from the same location. Moreover, our results indicate selection for assemblages of microorganisms by the environmental conditions they experience, assemblage metagenome-level tuning to the environment, and that the oceanic provinces defined based upon physicochemical features represent distinct ecological habitats.

Phylogenetic affiliation—The composition of microbial assemblages across the transect based upon random sequencing of genome fragments demonstrated the utility of the approach for examining dynamics across different regimes. The phylogeny of metagenomes was consistent

Table 3. Metabolic distribution of all subsystems in the 0.2–5-μm size fraction from all stations.*

Subsystem category	Mean % reads
Carbohydrates	11.8
Protein metabolism	10.3
Amino acids and derivatives	9.9
Cofactors, vitamins, prosthetic groups, pigments	7.5
RNA metabolism	4.8
Respiration	4.4
Virulence	4.2
DNA metabolism	4.1
Nucleosides and nucleotides	3.7
Cell wall and capsule	3.7
Other subsystems	21.1
Clustering-based subsystems	14.6

* RNA, ribonucleic acid; DNA, deoxyribonucleic acid.

with studies of open-ocean microbial assemblage composition based on 16S rRNA (Fuhrman and Ouverney 1998; Giovannoni and Stingl 2005), 16S–23S rRNA intergenic spacer (Brown and Fuhrman 2005) sequencing studies, and previous metagenomic efforts (Venter et al. 2004; Delong et al. 2006; Rusch et al. 2007). Consistent phylogeny was observed between the whole metagenome and only reads matching RDP II, except for the omission of eukaryotes (which are not represented in RDP II), and the larger proportion of putative Firmicute 16S rRNA in the RDP II taxonomic assignment.

Cyanobacteria comprised a larger proportion of shotgun reads in gyre waters than at the equator, whereas eukaryotes were more represented at the equator than in gyre waters, concomitant with Chl *a* concentrations, fewer *Prochlorococcus* cells in the equatorial zone, and higher cell abundances of picoeukaryotes examined by flow cytometry (http://hahana.soest.hawaii.edu/cmorebula/cmorebula.html). Additionally, there were fewer *Prochlorococcus* reads at the equator relative to eukaryotic reads (Fig. 4). The eukaryotic reads at the equator were mostly fungal proteins. Fungi are not common components of marine open-ocean eukaryotic communities (Countway et al. 2005), and it is likely that the high number of fungal annotations is a consequence of the much larger number of sequenced fungal genomes relative to other types of microeukaryotes. Similarly, in the >5-μm size fraction at the equator, zmost matches were to bivalve (*Septifer virgatus*) (18%) and insect (12%) proteins, which likely reflects the dearth of sequenced planktonic metazoan genomes.

Both cren- and euryarchaeotes comprised a larger proportion of reads at the equator than in gyre waters, and Archaea comprised 1.1–2.7% of all reads, in line with reports of archaeal abundance in surface waters of the open ocean (Ouverney and Fuhrman 1999; Karner et al. 2001). Viral protein reads, which were almost exclusively double stranded DNA viruses with no RNA stage, comprised only a small fraction of total reads (2.4–5.6%) and demonstrated no pattern in frequency across the transect. These likely represent viruses present within cells during the infection cycle rather than trapped by the filter, and are in similar magnitude to the frequency of visually infected cells (Weinbauer and Peduzzi 1994).

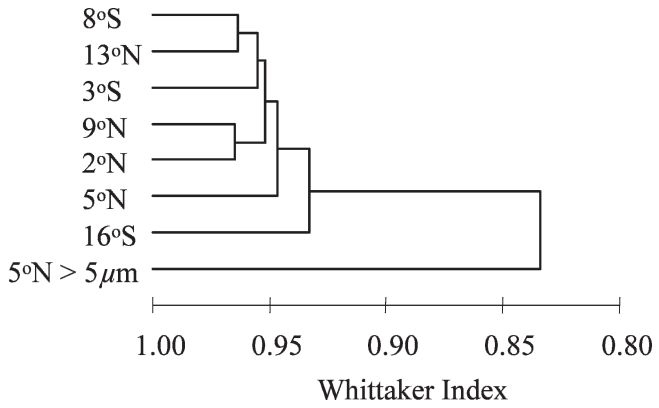

Fig. 8. Similarity between the metabolic distribution at each station. Clustering was performed by unweighted pair-group mean average (UPGMA) and is based upon the Whittaker Index (Whittaker 1952).

The phylogenetic clustering pattern of assemblages along the transect, which showed phylogenetic similarity between assemblages from the same oceanic ecological province, may demonstrate that the taxonomic composition of assemblages at different locations is comprised of organisms selected by local environmental conditions. Previous community fingerprinting (Moeseneder et al. 2001; Hewson et al. 2006*a*) and metagenomic (Tringe et al. 2005; Angly et al. 2006) comparisons between habitats have elucidated that assemblages are specific to habitat. High variability in surface-water bacterioplankton assemblages in the North Pacific Ocean, in the Amazon River Plume, and on a transect between the Arafura and Coral Seas was congruent in scale with physicochemical variation and water masses (Hewson et al. 2006*c*). Our study demonstrates that selection for individual taxa that comprise assemblages occurs at the scale of physicochemically defined oceanic provinces, and that variability in assemblages noted in previous sampling efforts over shorter spatial scales (Hewson et al. 2006*c*) based on intergenic regions may not be reflected in the functional gene complement.

The recruitment of metagenomic reads on the genomes of different *Prochlorococcus* and *Synechococcus* strains provided novel insight into the biogeographic patterns of dominant Cyanobacteria across a productivity gradient. *Synechococcus* had higher cell densities as measured by flow cytometry, and comprised a larger fraction of cyanobacterial reads at the equator than in gyre waters, which is consistent with observations of *Prochlorococcus* and *Synechococcus* distribution across productivity gradients in offshore waters of the Amazon River Plume (Hewson et al. 2006*a*), and cloning results from open-ocean and coastal waters (Fuhrman and Ouverney 1998; Giovannoni and Stingl 2005). Previous hybridization-based analyses of phylogeographic variation in picocyanobacterial strains indicated that *Prochlorococcus* was sensitive mostly to temperature and depth, and *Synechococcus* more by the availability of inorganic nutrients, despite the presence of temperature-optimized ecotypes of *Synechococcus* (Zwirglmaier et al. 2008).

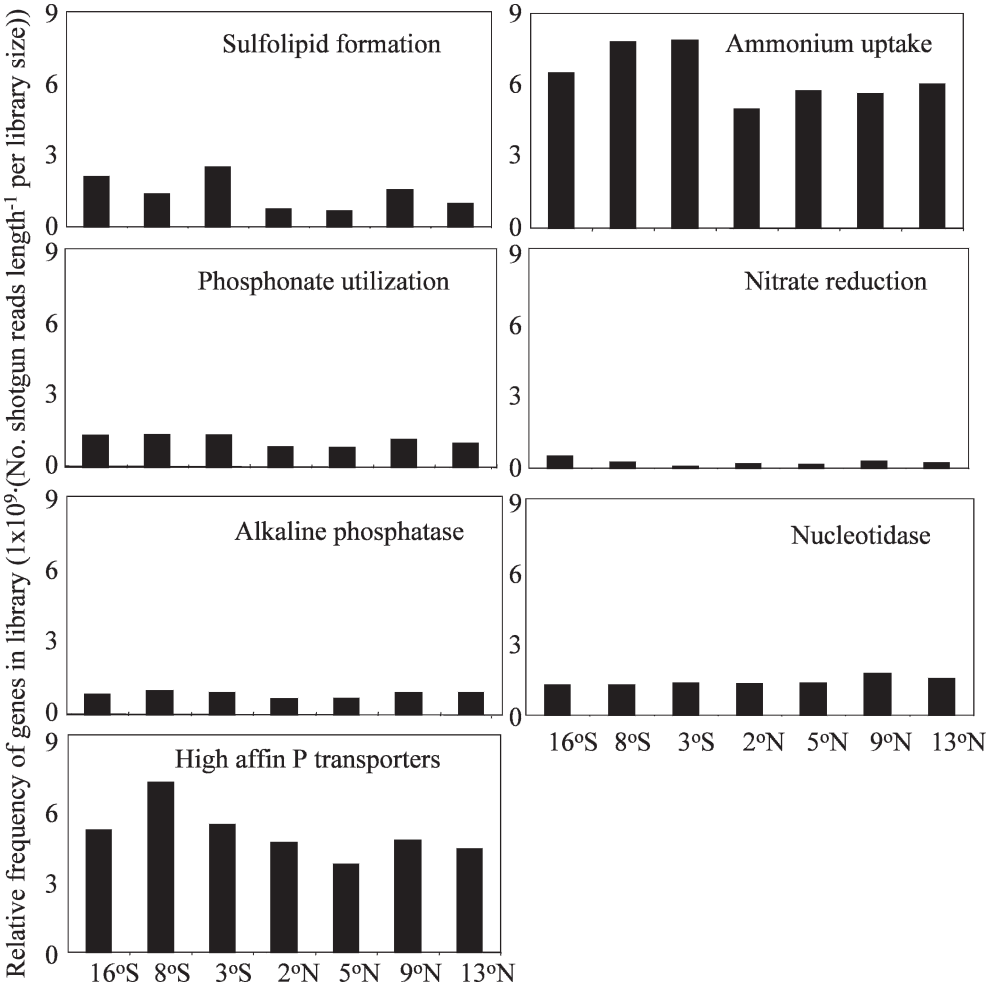

Fig. 9. Gene length and library-size standardized abundance of genes involved in key P and N biogeochemical processes along the transect, based upon annotations of BLASTx results ($E < 0.001$). Gene abundances were first normalized to the length of the closest matched genes in the protein database, then to the total library size (Frias-Lopez et al. 2008).

Synechococcus strains WH7803 and WH7805, which comprised a greater proportion of total *Synechococcus* reads at the equator than in gyre waters, were both isolated from mesotrophic waters by enrichment, and belong to marine clades V and VI, respectively (Rocap et al. 2002; Fuller et al. 2003). *Synechococcus* strains WH8102 and CC9605, which were depressed in representation at the equator relative to gyres, belong to clades III and II, respectively, and are typical of open-ocean conditions (Fuller et al. 2003). Strain RCC307 is a mesotroph isolated from the Mediterranean Sea, described as a "opportunist or coastal" strain (Dufresne et al. 2008; Zwirglmaier et al. 2008). The patterns of strain-level *Synechococcus* recruitment are congruent with geographic observations on strain distribution based on hybridization analyses (Zwirglmaier et al. 2008).

Prochlorococcus gene sequences demonstrated variability in ecotypes related primarily to light. All five dominant strains (AS9601, MIT9301, MIT9312, CCMP1986, and MIT9515) are high-light–adapted ecotypes (Rocap et al. 2002, 2003), whereas low-light ecotypes NATL1A, MIT9211, CCMP1375, MIT9303, and MIT9313 each comprised <2% of all *Prochlorococcus* genes detected. All metagenomes were prepared from surface waters where high light intensities may select for high-light–adapted strains. Strains AS9601, MIT9312, and MIT9301 comprised a larger component of total *Prochlorococcus* hits when abundances of *Prochlorococcus* were high; therefore, the higher abundance of strains MIT9515, CCMP1986, NATL1A, and NATL2A at the equator, when the abundance of *Prochlorococcus* was low, indicates that these strains are common components of bacterioplankton in the warm ocean but AS9601, MIT9312, and MIT9301 are obligate oligotrophic strains.

The large number of sequence reads matching a diazotroph, *C. watsonii* WH8501, is unusual in metagenomic surveys. In a study of surface waters of the Sargasso Sea, no nitrogenase genes were detected among several million metagenomic reads prepared from the 0.1–1-μm size fraction, which was interpreted to mean that nitrogen-fixing microorganisms were not present (Venter et al. 2004). However, diazotrophs are rare components of plankton,

Table 4. Results of multiple regression analysis of subsystems each comprising >1% of the total metagenome. Multiple regression was calculated using the stepwise forward selection procedure with change in r^2 as entry criteria. Model is indicated as Pathway = $\beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3 + \infty$, where β = variables and ∞ = intercept.†

Pathway	β_1	α_1	β_2	α_2	β_3	α_3	∞	r^2	p
Chlorophyll biosynthesis	temperature	0.077	part C	0.376	part Si	−21.733	0.140	0.998	***
Phosphate metabolism	part N	−1.012	part P	−0.030	part Si	2.400	2.021	0.995	***
Transcription initiation, bacterial sigma factors	PO_4^{3-}	1.143	ATP	−0.003	N_2O	−0.459	5.469	0.986	**
tRNA modification *E. coli*	part Si	−2.953	Chl *a*	2.220			2.093	0.976	***
Resistance to fluoroquinolones	PO_4^{3-}	2.717	NO_3^-	−0.194			1.183	0.818	*
Lipid A-Ara4N pathway (polymyxin resistance)	CH_4	1.433	N_2O	0.507			−6.485	0.803	*
Formaldehyde assimilation: ribulose monophosphate pathway	part Si	3.819					1.114	0.710	*
Bacterial cytoskeleton	temperature	0.079					−0.334	0.660	*
Deoxyribose and deoxynucleoside catabolism	part Si	4.232					0.974	0.634	*
ABC transporter branched-chain amino acid (TC 3.A.1.4.1)	part P	−0.036					1.554	0.712	*
Methionine biosynthesis	Chl *a* fluor	−1.075					3.085	0.828	**
Isoprenoid biosynthesis	part N	−0.496					1.338	0.840	**
Arginine biosynthesis extended	part Si	−5.996					1.635	0.866	**
Methylglyoxal metabolism	part C	0.533	part N	−1.676			3.763	0.887	*
FOS and raffinose utilization	part C	−0.120					1.457	0.961	***
Experimental subsystem FO	oxygen	0.029					−1.326	0.576	*

† ATP, adenosine triphosphate; tRNA, transfer RNA; Chl *a*, chlorophyll *a*; ABC, adenosine triphosphate–binding cassette; FOS, fructooligosaccharides.
* $p<0.05$.
** $p<0.01$.
*** $p<0.001$.

typically comprising <1% of all cells, and are mostly larger than 1 μm; therefore, diazotrophs and their nitrogenase genes would not be observed in even high-coverage libraries (Johnston et al. 2005). Therefore, the observation of metagenomic reads matching the *C. watsonii* WH8501 genome suggests their high abundance at Sta. 1. This is the first report of diazotrophs and components of the N_2 fixation gene cluster in a metagenomic survey. The presence of diazotrophic microorganisms similar to *C. watsonii* WH8501 was previously observed nearby (Church et al. 2008), and blooms of unicellular diazotrophs were observed in the region and in the same season (Hewson et al. 2009).

Metabolic insights—Our results provide novel information on the variability in genomic potential of microorganisms across a wide swath of the Pacific Ocean. Assemblages shared a large proportion of the read distribution between subsystems, which was unrelated to habitat physicochemical conditions, suggesting that dominant organisms have a core suite of genes necessary for life in the open ocean. The distribution of sequence reads among subsystem categories contained a large number of genes associated with cellular metabolism, which is consistent with previous reports of gene orthology in open-ocean metagenomes (Venter et al. 2004; Rusch et al. 2007). Previous metagenomic investigations demonstrated similarity between metagenomes based upon binned functional categories, but differences in particular groups of genes (notably high-affinity phosphate transporters) were variable between habitats (Rusch et al. 2007). The strong similarity in subsystem distribution between stations indicates that there is a core of genetic machinery necessary for microbial physiology in the open ocean, where differences between stations are driven by less-abundant genes. The remaining 56–68% of sequence reads not assignable to subsystems may comprise non–protein-encoding regions of sequenced genomes or organisms for which genome information is not available, and is in line with previous shotgun sequencing studies (Venter et al. 2004; Delong et al. 2006; Rusch et al. 2007).

Despite the relatively invariant gross genomic potential of assemblages across the transect, several less abundant genes encoding for biogeochemically relevant processes demonstrated heterogeneity between stations. Phosphate (PO_4^{3-}) availability is frequently a strong limiting factor in open-ocean conditions (Karl 2007). Microorganisms employ various strategies to overcome this limitation, including phosphonate utilization (Rocap et al. 2003; Dyhrman et al. 2006), use of nucleotidases to cleave PO_4^{3-} groups from nucleotides, PO_4^{3-} from a variety of dissolved organic compounds, and selective substitution of sulfolipids for phospholipids (Van Mooy et al. 2006, 2009). Because PO_4^{3-} availability is higher within equatorial waters compared to gyre waters, it was predicted that the distribution of genes involved in the P cycle would reflect the gradients in P availability. Our results support the hypothesis that selection for genes in P-poor waters leads to enhanced genomic capabilities for these processes, with a greater frequency of phosphonate utilization, alkaline phosphatase, and sulfolipid genes in waters away from the divergence and countercurrent zones. However, nucle-

otidases, which perform a wide variety of potential functions (including scavenging of nucleic acid bases) not limited to phosphate cleavage, do not show an obvious pattern from equatorial to gyre waters.

In addition to P availability, N is also limiting of primary production in gyre waters, but upwelled waters at the equator alleviate N limitation, causing enhanced abundance of microorganisms relative to gyre waters (Kirchman et al. 1995). Ammonium uptake and assimilatory nitrate reduction genes were hypothesized to be more abundant at the equator than in gyre waters, which is a consequence of selection for organisms with high affinity for upwelled N. However, genes involved in ammonium uptake and nitrate reduction did not have a strong trend between gyre and upwelling zone waters. Only four genes involved in nitrogen fixation were detected, all at Sta. 1, which is consistent with the low abundance of *nifH* gene copies in most waters relative to total gene inventories (Church et al. 2005).

The invariant frequency of genes involved in carbohydrate and protein metabolism stands in contrast to observed variation in ectoenzyme activities between North Pacific subtropical waters and those elsewhere (Christian and Karl 1995). Because elevated primary production releases a larger amount of carbohydrates that are available to heterotrophic microorganisms, whereas in gyre waters the ratio of carbohydrate to proteins is lower and hence more metabolism is potentially derived from protein, we expected to observe variable genomic capabilities concordant with carbohydrate to protein ratios. The observed lack of variation in protein and carbohydrate gene frequency may be a consequence of constitutive selection for genes involved in these uptake processes.

Several dominant gene subsystems were strongly correlated with physicochemical conditions; notably, >99% of variability in chlorophyll biosynthesis, phosphate metabolism, and transcription initiation bacterial sigma factors could be explained by combinations of chemical parameters. The strong relationship of chlorophyll metabolism to temperature, particulate carbon, and silicate (the latter two autocorrelate) reflects the distribution of photosynthetic microorganisms and probably the higher abundance of diatoms in equatorial than in gyre waters. Similarly, the relationship between phosphate metabolic gene frequency and particulate nitrogen, phosphate, and silicate concentrations is concomitant with higher phosphate uptake where the concentrations of cell-bound nutrients is high (Dyhrman and Haley 2006). The significant relationship between transcription initiation bacterial sigma factor gene frequency and phosphate, ATP, and N_2O concentrations may reflect the regional distribution of bacterial activity, because these growth-related genes correspond to the immediate by-product of microbial activity in the ocean (Karl 1979). Methionine biosynthesis was significantly and positively correlated with Chl *a* fluorescence (and hence less frequent in gyre than in equatorial waters), which is congruent with observations of reduced methionine use in nutrient-limited waters (Mary et al. 2008) and consistent with analysis of Global Ocean Survey metagenomes (Gianoulis et al. 2009). Interestingly, the frequency of methylglyoxal metabolic genes (methylglyoxal is a major product of glycolysis) was significantly and negatively related to particulate carbon and nitrogen concentrations, which may indicate a greater proportion of heterotrophic metabolism in gyre waters than in equatorial waters with higher particulate inventories. Moreover, these comparisons demonstrate habitat specificity of genomic capabilities within some metabolic pathways.

Comparison between >5-μm and 0.2–5-μm size fractions—The metagenome prepared from the >5-μm size fraction at Sta. 5 was distinct from the 0.2–5-μm metagenome prepared at the same station, and from other stations, in both subsystem pathway distribution and phylogenetic affiliation. Unlike metagenomes prepared from the 0.2–5-μm size fraction, for which 49–67% of reads could be annotated, only 11% and 18% could be annotated from the >5-μm size fraction by BLASTx and subsystem technology, respectively (Table 2). The >5-μm size fraction likely included a large number of eukaryotic organisms, including both unicellular organisms and metazoa. There are currently very few typical open-ocean eukaryotic microbial genomes that have been sequenced. Of annotated reads in the >5-μm size fraction, the taxonomic distribution was different than for the 0.2–5-μm size fraction and least similar to the phylogenies of other metagenomic libraries, driven primarily by the large number of eukaryotic protein reads. The distribution of sequences among metabolic subsystems was also different from metagenomes prepared from the 0.2–5-μm size fraction, primarily in the number of sequence reads of unclassified function. The low percentage of annotatable genes (based on current databases) is in a similar magnitude to early metagenomic (Venter et al. 2004) and viriomic (Breitbart et al. 2002) efforts. Future sequencing efforts targeting genomes of metazoa and eukaryotic microbial taxa typical of open-ocean plankton will permit wider annotation of this metagenome, as was the case with more recent metagenomic efforts.

The results of this study demonstrate that surface-water microbial assemblages are distinct to oceanic habitat. Changes in biogeochemical function arise as a consequence of selection of specific genes comprising only a small proportion of the total assemblage gene pool, and upon particular metabolic pathways. Despite differences in taxonomic composition and frequency of functional biogeochemical genes congruent with habitat, microbial assemblages share a common distribution of genomic potential. Moreover, habitat specificity of ecotypes and assemblages on the whole-genome level supports the concept of ecological provinces where factors causing selection are distinct from other regions. Finally, all our samples were taken from surface waters, yet enhanced productivity at the equator results from subsurface physical processes. Our data therefore emphasize the importance of measuring environmental parameters over vertical gradients, because these are at the scale relevant to community selection.

Acknowledgments

We thank chief scientist M. Church; also M. Nieto-Cid, J. Landrum, A. White, D. Hayakawa, B. Appelgate, K. Vellalos,

and the crew of the R/V *Kilo Moana* for assistance with sampling; K. Björkman, C. Mahaffey, S. Curless, B. Watkins, D. Viviani, and L. Fujieki for ancillary data publically available on the BULA-CMORE website (http://hahana.soest.hawaii.edu/cmorebula/cmorebula.html); and two anonymous reviewers for their comments on the manuscript draft. Sequencing was performed by S. Schuster (Pennsylvania State University) with bioinformatic assistance from P. Gilna and the Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis (CAMERA). This work was supported by the Gordon and Betty Moore Foundation Marine Microbiology Initiative and by National Science Foundation grant EF0424599 for the Center for Microbial Oceanography: Research and Education (CMORE).

References

Acinas, S. G., V. Klepac-Ceraj, D. E. Hunt, C. Pharino, I. Ceraj, D. L. Distel, and M. F. Polz. 2004. Fine-scale phylogenetic architecture of a complex bacterial community. Nature **430:** 551–554.

Angly, F., and others. 2006. The marine viromes of four oceanic regions. PLoS Biol. **4:** 2121–2131.

Azam, F., T. Fenchel, J. G. Field, J. S. Gray, L. A. Meyer-Reil, and F. Thingstad. 1983. The ecological role of water-column microbes in the sea. Mar. Ecol. Prog. Ser. **10:** 257–263.

Beja, O., and others. 2000. Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. Science **289:** 1902–1906.

Breitbart, M., A. Segall, D. Mead, F. Azam, and F. Rohwer. 2002. Genomic analysis of uncultured marine viral communities. Proc. Natl. Acad. Sci. USA **99:** 14250–14255.

Brown, M. V., and J. A. Fuhrman. 2005. Marine bacterial microdiversity as revealed by internal transcribed spacer analysis. Aquat. Microb. Ecol. **41:** 15–23.

Capone, D. G. 2000. The marine microbial nitrogen cycle, p. 455–493. *In* D. L. Kirchman [ed.], Microbial ecology of the oceans. Wiley-Liss.

———, J. P. Zehr, H. W. Paerl, B. Bergman, and E. J. Carpenter. 1997. *Trichodesmium*, a globally significant marine cyanobacterium. Science **276:** 1221–1229.

Carpenter, E. J. 1983. Nitrogen fixation by *Oscillatoria* (*Trichodesmium*) in the world's oceans, p. 95–104. *In* E. J. Carpenter and D. G. Capone [eds.], Nitrogen in the marine environment. Academic.

Christian, J. R., and D. M. Karl. 1995. Bacterial ectoenzymes in marine waters: Activity ratios and temperature responses in three oceanographic provinces. Limnol. Oceanogr. **40:** 1042–1049.

Church, M. J., K. M. Bjorkman, D. M. Karl, M. A. Saito, and J. P. Zehr. 2008. Regional distributions of nitrogen-fixing bacteria in the Pacific Ocean. Limnol. Oceanogr. **53:** 63–77.

———, B. D. Jenkins, D. M. Karl, and J. P. Zehr. 2005. Vertical distributions of nitrogen-fixing phylotypes at Stn ALOHA in the oligotrophic North Pacific Ocean. Aquat. Microb. Ecol. **38:** 3–14.

Countway, P. D., R. J. Gast, P. Savai, and D. A. Caron. 2005. Protistan diversity estimates based on 18S rDNA from seawater incubations in the western North Atlantic. J. Eukaryot. Microbiol. **52:** 95–106.

Delong, E., and others. 2006. Community genomics among stratified microbial assemblages in the ocean's interior. Science **311:** 496–503.

Dufresne, A., and others. 2008. Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. Genome Biol. **9:** R90, doi: 10.1186/gb-2008-9-5-r90.

Dyhrman, S. T., P. D. Chappell, S. T. Haley, J. W. Moffett, E. D. Orchard, J. B. Waterbury, and E. A. Webb. 2006. Phosphonate utilization by the globally important marine diazotroph *Trichodesmium*. Nature **439:** 68–71.

———, and S. T. Haley. 2006. Phosphorus scavenging in the unicellular marine diazotroph *Crocosphaera watsonii*. Appl. Environ. Microbiol. **72:** 1452–1458.

Frias-Lopez, J., Y. Shi, G. W. Tyson, M. L. Coleman, S. C. Schuster, S. W. Chisholm, and E. F. DeLong. 2008. Microbial community gene expression in ocean surface waters. Proc. Nat. Acad. Sci. USA **105:** 3805–3810.

Fuhrman, J. A., D. E. Comeau, A. Hagstrom, and A. M. Chan. 1988. Extraction of DNA suitable for molecular biological studies from natural planktonic microorganisms. Appl. Environ. Microbiol. **54:** 1426–1429.

———, I. Hewson, M. Schwalbach, J. Steele, M. Brown, and S. Naeem. 2006. Annually reoccurring bacterial communities are predictable from ocean conditions. Proc. Natl. Acad. Sci. USA **103:** 13104–13109.

———, and C. C. Ouverney. 1998. Marine microbial diversity studied via 16S rRNA sequences: Cloning results from coastal waters and counting of native archaea with fluorescent single cell probes. Aquat. Ecol. **32:** 3–15.

———, J. A. Steele, I. Hewson, M. S. Schwalbach, M. V. Brown, J. L. Green, and J. H. Brown. 2008. A latitudinal diversity gradient in planktonic marine bacteria. Proc. Natl. Acad. Sci. USA **105:** 7774–7778.

Fuller, N., D. Marie, F. Partensky, D. Vaulot, A. F. Post, and D. J. Scanlan. 2003. Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. Appl. Environ. Microbiol. **69:** 2430–2443.

Gianoulis, T. A., and others. 2009. Quantifying environmental adaptation of metabolic pathways in metagenomics. Proc. Natl. Acad. Sci. USA **106:** 1374–1379.

Giovannoni, S., and U. Stingl. 2005. Molecular diversity and ecology of microbial plankton. Nature **437:** 343–348.

———, and others. 2005. Genome streamlining in a cosmopolitan oceanic bacterium. Science **309:** 1242–1245.

Hewson, I., D. G. Capone, J. A. Steele, and J. A. Fuhrman. 2006*a*. Influence of the Amazon and Orinoco offshore surface water plumes on oligotrophic bacterioplankton diversity in the West Tropical Atlantic. Aquat. Microb. Ecol. **43:** 11–22.

———, J. A. Steele, D. G. Capone, and J. A. Fuhrman. 2006*b*. Remarkable heterogeneity in meso- to bathypelagic bacterioplankton assemblage composition. Limnol. Oceanogr. **51:** 1274–1283.

———, ———, ———, and, ———. 2006*c*. Temporal and spatial scales of variation in bacterioplankton assemblages of oligotrophic surface waters. Mar. Ecol. Prog. Ser. **311:** 67–77.

———, and others. 2009. *In situ* transcriptomic analysis of the globally important keystone N_2-fixing taxon *Crocosphaera watsonii*. ISME J. **3:** 618–631.

Johnston, A. W. B., Y. Li, and L. Ogilvie. 2005. Metagenomic marine nitrogen fixation—feast or famine? Trends Microbiol. **13:** 416–420.

Kanehisa, M., S. Goto, S. Kawashima, Y. Okuno, and M. Hattori. 2004. The KEGG resource for deciphering the genome. Nucleic Acids Res. **32:** D277–D280.

Karl, D. M. 1979. Measurement of microbial activity and growth in the ocean by rates of stable ribonucleic acid synthesis. Appl. Environ. Microbiol. **38:** 850–860.

———. 2007. Microbial oceanography: Paradigms, processes and promise. Nature Rev. Microbiol. **5:** 759–769.

———, J. R. Christian, J. E. Dore, D. V. Hebel, R. M. Letelier, L. M. Tupas, and C. D. Winn. 1996. Seasonal and interannual variability in primary production and particle flux at Station ALOHA. Deep-Sea Res. II **43:** 539–568.

———, and G. Tien. 1992. MAGIC—a sensitive and precise method for measuring dissolved phosphorus in aquatic environments. Limnol. Oceanogr. **37:** 105–116.

———, and others. 2002. Dinitrogen fixation in the world's oceans. Biogeochemistry **57/58:** 47–98.

Karner, M. B., E. F. Delong, and D. M. Karl. 2001. Archeal dominance in the mesopelagic zone of the Pacific Ocean. Nature **409:** 507–510.

Kirchman, D. L., and J. H. Rich. 1997. Regulation of bacterial growth by dissolved organic carbon and temperature in the equatorial Pacific Ocean. Microb. Ecol. **33:** 11–20.

———, ———, and R. T. Barber. 1995. Biomass and biomass production of heterotrophic bacteria along 140°W in the equatorial Pacific: Effect of temperature on the microbial loop. Deep-Sea Res. II **42:** 603–619.

Longhurst, A. R. 2006. Ecological geography of the sea, 2nd ed. Academic.

Margulies, M., and others. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:** 376–380.

Mary, I., G. A. Tarran, P. E. Warwick, M. J. Terry, D. J. Scanlan, P. H. Burkill, and M. V. Zubkov. 2008. Light enhanced amino acid uptake by dominant bacterioplankton groups in surface waters of the Atlantic Ocean. FEMS Microbiol. Ecol. **63:** 36–45.

Meyer, F., and others. 2008. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics **9:** 386.

Moeseneder, M. M., C. Winter, and G. J. Herndl. 2001. Horizontal and vertical complexity of attached and free-living bacteria of the eastern Mediterranean Sea, determined by 16S rDNA and 16S rRNA fingerprints. Limnol. Oceanogr. **46:** 95–107.

Ouverney, C. C., and J. A. Fuhrman. 1999. Marine planktonic Archaea take up amino acids. Appl. Environ. Microbiol. **66:** 4829–4833.

Overbeek, R., and others. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. **33:** 5691–5702.

Rocap, G., D. L. Distel, J. B. Waterbury, and S. W. Chisolm. 2002. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. Appl. Environ. Microbiol. **68:** 1180–1191.

———, and others. 2003. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. Nature **424:** 1042–1047.

Rusch, D. B., and others. 2007. The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific. PLoS Biol. **5:** 398–431.

Sañudo-Wilhelmy, S. A., and others. 2001. Phosphorus limitation of nitrogen fixation by *Trichodesmium* in the central Atlantic Ocean. Nature **411:** 66–69.

Tringe, S. G., and others. 2005. Comparative metagenomics of microbial communities. Science **308:** 554–557.

Tyson, G. W., and others. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature **428:** 37–43.

Van Mooy, B. A. S., G. Rocap, H. F. Fredricks, C. T. Evans, and A. H. Devol. 2006. Sulfolipids dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine environments. Proc. Natl. Acad. Sci. USA **103:** 8607–8612.

———, and others. 2009. Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. Nature **458:** 69–72.

Venter, J. C., and others. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science **304:** 66–74.

Weinbauer, M. G., and P. Peduzzi. 1994. Frequency, size and distribution of bacteriophages in different marine bacterial morphotypes. Mar. Ecol. Prog. Ser. **108:** 11–20.

Wells, M. L., N. M. Price, and K. W. Bruland. 1994. Iron limitation and the cyanobacterium *Synechococcus* in equatorial Pacific waters. Limnol. Oceanogr. **39:** 1481–1486.

Whittaker, R. H. 1952. A study of summer foliage insect communities in the Great Smoky Mountains. Ecol. Monogr. **22:** 1–44.

Williams, P. J. le B. 1981. Incorporation of microheterotrophic processes into the classical paradigm of the planktonic food web. Kiel. Meeresforsch. (Sonderh.) **5:** 1–28.

Zehr, J. P., and others. 2007. Experiments linking nitrogenase gene expression and N_2 fixation in the North Pacific Subtropical Gyre. Limnol. Oceanogr. **52:** 169–183.

Zwirglmaier, K., and others. 2008. Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. Environ. Microbiol. **10:** 147–161.

Associate editor: Wade H. Jeffrey

Received: 26 February 2009
Accepted: 18 June 2009
Amended: 25 June 2009