The following resources related to this article are available online at
www.sciencemag.org (this information is current as of February 23, 2007):

Updated information and services, including high-resolution figures, can be found in the online
version of this article at:
http://www.sciencemag.org/cgi/content/full/311/5760/496

Supporting Online Material can be found at:
http://www.sciencemag.org/cgi/content/full/311/5760/496/DC1

This article **cites 35 articles**, 19 of which can be accessed for free:
http://www.sciencemag.org/cgi/content/full/311/5760/496#otherarticles

This article has been **cited by** 21 article(s) on the ISI Web of Science.

This article has been **cited by** 9 articles hosted by HighWire Press; see:
http://www.sciencemag.org/cgi/content/full/311/5760/496#otherarticles

This article appears in the following **subject collections**:
Microbiology
http://www.sciencemag.org/cgi/collection/microbio

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce**
this article in whole or in part can be found at:
http://www.sciencemag.org/help/about/permissions.dtl

32. Materials and methods are available as supporting material on *Science* Online.
33. J. R. Abo-Shaeer, C. Raman, W. Ketterle, *Phys. Rev. Lett.* **88**, 070409 (2002).
34. D. Guéry-Odelin, *Phys. Rev. A.* **62**, 033607 (2000).
35. Y. Kagan, E. L. Surkov, G. V. Shlyapnikov, *Phys. Rev. A.* **55**, R18 (1997).
36. C. Menotti, P. Pedri, S. Stringari, *Phys. Rev. Lett.* **89**, 250402 (2002).
37. G. B. Partridge, W. Li, R. I. Kamar, Y.-a. Liao, R. G. Hulet, *Science*, **311**, 503 (2006); published online 22 December 2005 (10.1126/science.1122876).

38. T. Mizushima, K. Machida, M. Ichioka, *Phys. Rev. Lett.* **94**, 060404 (2005).
39. P. Castorina, M. Grasso, M. Oertel, M. Urban, D. Zappalà, *Phys. Rev. A.* **72**, 025601 (2005).
40. A. A. Abrikosov, L. P. Gorkov, I. E. Dzyaloshinski, *Methods of Quantum Field Theory in Statistical Physics* (Dover, New York, 1975).
41. G. Bertsch, INT Workshop on Effective Field Theory in Nuclear Physics (Seattle, WA, February 1999).
42. T. D. Cohen, *Phys. Rev. Lett.* **95**, 120403 (2005).
43. We thank G. Campbell for critical reading of the manuscript and X.-G. Wen, E. Demler, and S. Sachdev for stimulating

Community Genomics Among Stratified Microbial Assemblages in the Ocean's Interior

Edward F. DeLong,[1]* Christina M. Preston,[2] Tracy Mincer,[1] Virginia Rich,[1] Steven J. Hallam,[1] Niels-Ulrik Frigaard,[1] Asuncion Martinez,[1] Matthew B. Sullivan,[1] Robert Edwards,[3] Beltran Rodriguez Brito,[3] Sallie W. Chisholm,[1] David M. Karl[4]

Microbial life predominates in the ocean, yet little is known about its genomic variability, especially along the depth continuum. We report here genomic analyses of planktonic microbial communities in the North Pacific Subtropical Gyre, from the ocean's surface to near–sea floor depths. Sequence variation in microbial community genes reflected vertical zonation of taxonomic groups, functional gene repertoires, and metabolic potential. The distributional patterns of microbial genes suggested depth-variable community trends in carbon and energy metabolism, attachment and motility, gene mobility, and host-viral interactions. Comparative genomic analyses of stratified microbial communities have the potential to provide significant insight into higher-order community organization and dynamics.

Microbial plankton are centrally involved in fluxes of energy and matter in the sea, yet their vertical distribution and functional variability in the ocean's interior is still only poorly known. In contrast, the vertical zonation of eukaryotic phytoplankton and zooplankton in the ocean's water column has been well documented for over a century (*1*). In the photic zone, steep gradients of light quality and intensity, temperature, and macronutrient and trace-metal concentrations all influence species distributions in the water column (*2*). At greater depths, low temperature, increasing hydrostatic pressure, the disappearance of light, and dwindling energy supplies largely determine vertical stratification of oceanic biota.

For a few prokaryotic groups, vertical distributions and depth-variable physiological properties are becoming known. Genotypic and phenotypic properties of stratified *Prochlorococcus* "ecotypes" for example, are suggestive of depth-variable adaptation to light intensity and nutrient availability (*3–5*). In the abyss, the vertical zonation of deep-sea piezophilic bacteria can be explained in

part by their obligate growth requirement for elevated hydrostatic pressures (*6*). In addition, recent cultivation-independent (*7–15*) surveys have shown vertical zonation patterns among specific groups of planktonic *Bacteria*, *Archaea*, and *Eukarya*. Despite recent progress however, a comprehensive description of the biological properties and vertical distributions of planktonic microbial species is far from complete.

Cultivation-independent genomic surveys represent a potentially useful approach for characterizing natural microbial assemblages (*16, 17*). "Shotgun" sequencing and whole genome assembly from mixed microbial assemblages has been attempted in several environments, with varying success (*18, 19*). In addition, Tringe *et al.* (*20*) compared shotgun sequences of several disparate microbial assemblages to identify community-specific patterns in gene distributions. Metabolic reconstruction has also been attempted with environmental genomic approaches (*21*). Nevertheless, integrated genomic surveys of microbial communities along well-defined environmental gradients (such as the ocean's water column) have not been reported.

To provide genomic perspective on microbial biology in the ocean's vertical dimension, we cloned large [~36 kilobase pairs (kbp)] DNA fragments from microbial communities at different depths in the North Pacific Subtropical Gyre

(NPSG) at the open-ocean time-series station ALOHA (*22*). The vertical distribution of microbial genes from the ocean's surface to abyssal depths was determined by shotgun sequencing of fosmid clone termini. Applying identical collection, cloning, and sequencing strategies at seven depths (ranging from 10 m to 4000 m), we archived large-insert genomic libraries from each depth-stratified microbial community. Bidirectional DNA sequencing of fosmid clones (~10,000 sequences per depth) and comparative sequence analyses were used to identify taxa, genes, and metabolic pathways that characterized vertically stratified microbial assemblages in the water column.

Study Site and Sampling Strategy

Our sampling site, Hawaii Ocean Time-series (HOT) station ALOHA (22°45' N, 158°W), represents one of the most comprehensively characterized sites in the global ocean and has been a focal point for time series–oriented oceanographic studies since 1988 (*22*). HOT investigators have produced high-quality spatial and time-series measurements of the defining physical, chemical, and biological oceanographic parameters from surface waters to the seafloor. These detailed spatial and temporal datasets present unique opportunities for placing microbial genomic depth profiles into appropriate oceanographic context (*22–24*) and leverage these data to formulate meaningful ecological hypotheses. Sample depths were selected, on the basis of well-defined physical, chemical, and biotic characteristics, to represent discrete zones in the water column (Tables 1 and 2, Fig. 1; figs. S1 and S2). Specifically, seawater samples from the upper euphotic zone (10 m and 70 m), the base of the chlorophyll maximum (130 m), below the base of the euphotic zone (200 m), well below the upper mesopelagic (500 m), in the core of the dissolved oxygen minimum layer (770 m), and in the deep abyss, 750 m above the seafloor (4000 m), were collected for preparing microbial community DNA libraries (Tables 1 and 2, Fig. 1; figs. S1 and S2).

The depth variability of gene distributions was examined by random, bidirectional end-sequencing of ~5000 fosmids from each depth, yielding ~64 Mbp of DNA sequence total from the 4.5 Gbp archive (Table 1). This represents raw sequence coverage of about 5 (1.8 Mbp sized) genome equivalents per depth. Because we surveyed ~180 Mbp of cloned DNA (5000 clones by

[1]Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]Monterey Bay Aquarium Research Institute, Moss Landing, CA 95064, USA. [3]San Diego State University, San Diego, CA 92182, USA. [4]University of Hawaii Honolulu, HI 96822, USA.

*To whom correspondence should be addressed. E-mail: delong@mit.edu

~36 kbp/clone per depth), however, we directly sampled ~100 genome equivalents at each depth. We did not sequence as deeply in each sample as a recent Sargasso Sea survey (19), where from 90,000 to 600,000 sequences were obtained from small DNA insert clones, from each of seven different surface-water samples. We hypothesized, however, that our comparison of microbial communities collected along well-defined environmental gradients (using large-insert DNA clones), would facilitate detection of ecologically meaningful taxonomic, functional, and community trends.

Vertical Profiles of Microbial Taxa

Vertical distributions of bacterial groups were assessed by amplifying and sequencing small subunit (SSU) ribosomal RNA (rRNA) genes from complete fosmid library pools at each depth (Fig. 2; fig. S3). Bacterial phylogenetic distributions were generally consistent with previous polymerase chain reaction–based cultivation-independent rRNA surveys of marine picoplankton (8, 15, 25). In surface-water samples, rRNA-containing fosmids included those from *Prochlorococcus*; *Verrucomicrobiales*; *Flexibacteraceae*; Gammaproteobacteria (SAR92, OM60, SAR86 clades); Alphaproteobacteria (SAR116, OM75 clades); and Deltaproteobacteria (OM27 clade) (Fig. 2). Bacterial groups from deeper waters included members of *Deferribacteres*; *Planctomycetaceae*; *Acidobacteriales*; *Gemmatamonadaceae*; *Nitrospina*; *Alteromonadaeceae*; and SAR202, SAR11, and Agg47 planktonic bacterial clades (Fig. 2; fig. S2). Large-insert DNA clones previously recovered from the marine environment (9, 10) also provide a good metric for taxonomic assessment of indigenous microbes. Accordingly, a relatively large proportion of our shotgun fosmid sequences most closely matched rRNA-containing bacterioplankton artificial clones previously recovered from the marine environment (fig. S3).

Taxonomic bins of bacterial protein homologs found in randomly sequenced fosmid ends (Fig. 2; fig. S4) also reflected distributional patterns generally consistent with previous surveys in the water column (8, 15). Unexpectedly large amounts of phage DNA were recovered in clones, particularly in the photic zone. Also unexpected was a relatively high proportion of Betaproteobacteria-like sequences recovered at 130 m, most sharing highest similarity to protein homologs from *Rhodoferax ferrireducens*. As expected, representation of *Prochlorococcus*-like and *Pelagibacter*-like genomic sequences was high in the photic zone. At greater depths, higher proportions of *Chloroflexi*-like sequences, perhaps corresponding to the co-occurring SAR202 clade, were observed (Fig. 2). *Planctomycetales*-like genomic DNA sequences were also highly represented at greater depths.

All archaeal SSU rRNA–containing fosmids were identified at each depth, quantified by macroarray hybridization, and their rRNAs sequenced

Table 1. HOT samples and fosmid libraries. Sample site, 22°45' N, 158°W. All seawater samples were pre-filtered through a 1.6-μm glass fiber filter, and collected on a 0.22-μm filter. See (35) for methods.

Depth (m)	Sample date	Volume filtered (liters)	Total fosmid clones	Total DNA (Mbp) Archived	Total DNA (Mbp) Sequenced
10	10/7/02	40	12,288	442	7.54
70	10/7/02	40	12,672	456	11.03
130	10/6/02	40	13,536	487	6.28
200	10/6/02	40	19,008	684	7.96
500	10/6/02	80	15,264	550	8.86
770	12/21/03	240	11,520	415	11.18
4,000	12/21/03	670	41,472	1,493	11.10

Table 2. HOT sample oceanographic data. Samples described in Table 1. Oceanographic parameters were measured as specified at (49); values shown are those from the same CTD casts as the samples, where available. Values in parentheses are the mean ± 1 SD of each core parameter during the period October 1988 to December 2004, with the total number of measurements collected for each parameter shown in brackets. The parameter abbreviations are Temp., Temperature; Chl a, chlorophyll a; DOC, dissolved organic carbon; N+N, nitrate plus nitrite; DIP, dissolved inorganic phosphate; and DIC, dissolved inorganic carbon. The estimated photon fluxes for upper water column samples (assuming a surface irradiance of 32 mol quanta m^{-2} d^{-1} and a light extinction coefficient of 0.0425 m^{-1}) were: 10 m = 20.92 (65% of surface), 70 m = 1.63 (5% of surface), 130 m = 0.128 (0.4% of surface), 200 m = 0.07 (0.02% of surface). The mean surface mixed-layer during the October 2002 sampling was 61 m. Data are available at (50). *Biomass derived from particulate adenosine triphosphate (ATP) measurements assuming a carbon:ATP ratio of 250. ND, Not determined.

Depth (m)	Temp. (°C)	Salinity	Chl a (μg/kg)	Biomass* (μg/kg)	DOC (μmol/kg)	N + N (nmol/kg)	DIP (nmol/kg)	Oxygen (μmol/kg)	DIC (μmol/kg)
10	26.40 (24.83 ± 1.27) [2,104]	35.08 (35.05 ± 0.21) [1,611]	0.08 (0.08 ± 0.03) [320]	7.21 ± 2.68 [78]	78 (90.6 ± 14.3) [140]	1.0 (2.6 ± 3.7) [126]	41.0 (56.0 ± 33.7) [146]	204.6 (209.3 ± 4.5) [348]	1,967.6 (1,972.1 ± 16.4) [107]
70	24.93 (23.58 ± 1.00) [1,202]	35.21 (35.17 ± 0.16) [1,084]	0.18 (0.15 ± 0.05) [363]	8.51 ± 3.22 [86]	79 (81.4 ± 11.3) [79]	1.3 (14.7 ± 60.3) [78]	16.0 (43.1 ± 25.1) [104]	217.4 (215.8 ± 5.4) [144]	1,981.8 (1,986.9 ± 15.4) [84]
130	22.19 (21.37 ± 0.96) [1,139]	35.31 (35.20 ± 0.10) [980]	0.10 (0.15 ± 0.06) [350]	5.03 ± 2.30 [90]	69 (75.2 ± 9.1) [86]	284.8 (282.9 ± 270.2) [78]	66.2 (106.0 ± 49.7) [68]	204.9 (206.6 ± 6.2) [173]	2,026.5 (2,013.4 ± 13.4) [69]
200	18.53 (18.39 ± 1.29) [662]	35.04 (34.96 ± 0.18) [576]	0.02 (0.02 ± 0.02) [97]	1.66 ± 0.24 [2]	63 (64.0 ± 9.8) [113]	1,161.9 ± 762.5 [7]	274.2 ± 109.1 [84]	198.8 (197.6 ± 7.1) [190]	2,047.7 (2,042.8 ± 10.5) [125]
500	7.25 (7.22 ± 0.44) [1,969]	34.07 (34.06 ± 0.03) [1,769]	ND	0.48 ± 0.23 [107]	47 (47.8 ± 6.3) [112]	28,850 (28,460 ± 2210) [326]	2,153 (2,051 ± 175.7) [322]	118.0 (120.5 ± 18.3) [505]	2197.3 (2,200.2 ± 17.8) [134]
770	4.78 (4.86 ± 0.21) [888]	34.32 (34.32 ± 0.04) [773]	ND	0.29 ± 0.16 [107]	39.9 (41.5 ± 4.4) [34]	41,890 (40,940 ± 500) [137]	3,070 (3,000 ± 47.1) [135]	32.3 (27.9 ± 4.1) [275]	2323.8 (2,324.3 ± 6.1) [34]
4,000	1.46 (1.46 ± 0.01) [262]	34.69 (34.69 ± 0.00) [245]	ND	ND	37.5 (42.3 ± 4.9) [83]	36,560 (35,970 ± 290) [108]	2,558 (2,507 ± 19) [104]	147.8 (147.8 ± 1.3) [210]	2325.5 (2,329.1 ± 4.8) [28]

(figs. S5 and S6). The general patterns of archaeal distribution we observed were consistent with previous field surveys (*15, 25, 26*). Recovery of "group II" planktonic *Euryarchaeota* genomic DNA was greatest in the upper water column and declined below the photic zone. This distribution corroborates recent observations of ion-translocating photoproteins (called proteorhodopsins), now known to occur in group II *Euryarchaeota* inhabiting the photic zone (*27*). "Group III" *Euryarchaeota* DNA was recovered at all depths, but at a much lower frequency (figs. S5 and S6). A novel crenarchaeal group, closely related to a putatively thermophilic *Crenarchaeota* (*28*), was observed at the greatest depths (fig. S6).

Vertically Distributed Genes and Metabolic Pathways

The depths sampled were specifically chosen to capture microbial sequences at discrete biogeochemical zones in the water column encompassing key physicochemical features (Tables 1 and 2, Fig. 1; figs. S1 and S2). To evaluate sequences from each depth, fosmid end sequences were compared against different databases including the Kyoto Encyclopedia of Genes and Genomes (KEGG) (*29*), National Center for Biotechnology Information (NCBI)'s Clusters of Orthologous Groups (COG) (*30*), and SEED subsystems (*31*). After categorizing sequences from each depth in BLAST searches (*32*) against each database, we identified protein categories that were more or less well represented in one sample versus another, using cluster analysis (*33, 34*) and bootstrap resampling methodologies (*35*).

Cluster analyses of predicted protein sequence representation identified specific genes and metabolic traits that were differentially distributed in the water column (fig. S7). In the photic zone (10, 70, and 130 m), these included a greater representation in sequences associated with photosynthesis; porphyrin and chlorophyll metabolism; type III secretion systems; and aminosugars, purine, proponoate, and vitamin B6 metabolism, relative to deep-water samples (fig. S7). Independent comparisons with well-annotated subsystems in the SEED database (*31*) also showed similar and overlapping trends (table S1), including greater representation in photic zone sequences associated with alanine and aspartate; metabolism of aminosugars; chlorophyll and carotenoid biosynthesis; maltose transport; lactose degradation; and heavy metal ion sensors and exporters. In contrast, samples from depths of 200 m and below (where there is no photosynthesis) were enriched in different sequences, including those associated with protein folding; processing and export; methionine metabolism; glyoxylate, dicarboxylate, and methane metabolism; thiamine metabolism; and type II secretion systems, relative to surface-water samples (fig. S7).

COG categories also provided insight into differentially distributed protein functions and categories. COGs more highly represented in photic zone included iron-transport membrane receptors,
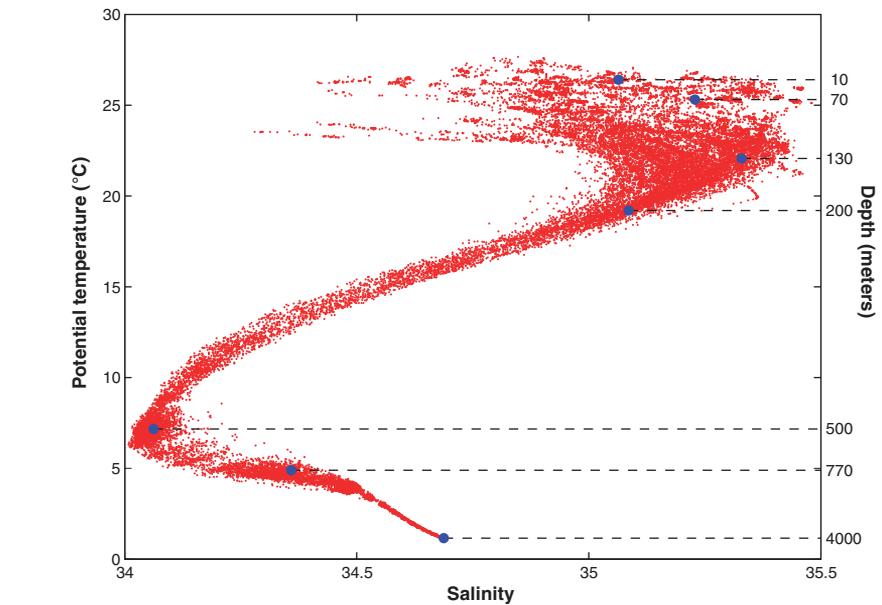
Fig. 1. Temperature versus salinity (T-S) relations for the North Pacific Subtropical Gyre at station ALOHA (22°45'N, 158°W). The blue circles indicate the positions, in T-S "hydrospace" of the seven water samples analyzed in this study. The data envelope shows the temperature and salinity conditions observed during the period October 1988 to December 2004 emphasizing both the temporal variability of near-surface waters and the relative constancy of deep waters.

deoxyribopyrimidine photolyase, diaminopimelate decarboxylase, membrane guanosine triphosphatase (GTPase) with the lysyl endopeptidase gene product LepA, and branched-chain amino acid–transport system components (fig. S8). In contrast, COGs with greater representation in deep-water samples included transposases, several dehydrogenase categories, and integrases (fig. S8). Sequences more highly represented in the deep-water samples in SEED subsystem (*31*) comparisons included those associated with respiratory dehydrogenases, polyamine adenosine triphosphate (ATP)–binding cassette (ABC) transporters, polyamine metabolism, and alkylphosphonate transporters (table S1).

Habitat-enriched sequences. We estimated average protein sequence similarities between all depth bins from cumulative TBLASTX high-scoring sequence pair (HSP) bitscores, derived from BLAST searches of each depth against every other (Fig. 3). Neighbor-joining analyses of a normalized, distance matrix derived from these cumulative bitscores joined photic zone and deeper samples together in separate clusters (Fig. 3). When we compared our HOT sequence datasets to previously reported Sargasso Sea microbial sequences (*19*), these datasets also clustered according to their depth and size fraction of origin (fig. S9). The clustering pattern in Fig. 3 is consistent with the expectation that randomly sampled photic zone microbial sequences will tend on average to be more similar to one another, than to those from the deep-sea, and vice-versa.

We also identified those sequences (some of which have no homologs in annotated databases)

that track major depth-variable environmental features. Specifically, sequence homologs found only in the photic zone unique sequences (from 10, 70, and 130 m), or deepwater unique sequences (from 500, 770, and 4000 m) were identified (Fig. 3). To categorize potential functions encoded in these photic zone unique (PZ) or deep-water unique (DW) sequence bins, each was compared with KEGG, COG, and NCBI protein databases in separate analyses (*29, 30, 36*).

Some KEGG metabolic pathways appeared more highly represented in the PZ than in DW sequence bins, including those associated with photosynthesis; porphyrin and chlorophyll metabolism; propanoate, purine, and glycerphospholipid metabolism; bacterial chemotaxis; flagellar assembly; and type III secretion systems (Fig. 4A). All proteorhodopsin sequences (except one) were captured in the PZ bin. Well-represented photic zone KEGG pathway categories appeared to reflect potential pathway interdependencies. For example the PZ photosynthesis bin [3% of the total (Fig. 4A)] contained *Prochlorococcus*-like and *Synechococcus*-like photosystem I, photosystem II, and cytochrome genes. In tandem, PZ porphyrin and chlorophyll biosynthesis sequence bins [~3.9% of the total (Fig. 4A)] contained high representation of cyanobacteria-like cobalamin and chlorophyll biosynthesis genes, as well as photoheterotroph-like bacteriochlorophyll biosynthetic genes. Other probable functional interdependencies appear reflected in the corecovery of sequences associated with chemotaxis (mostly methyl-accepting chemotaxis proteins), flagellar biosynthesis (predominant-
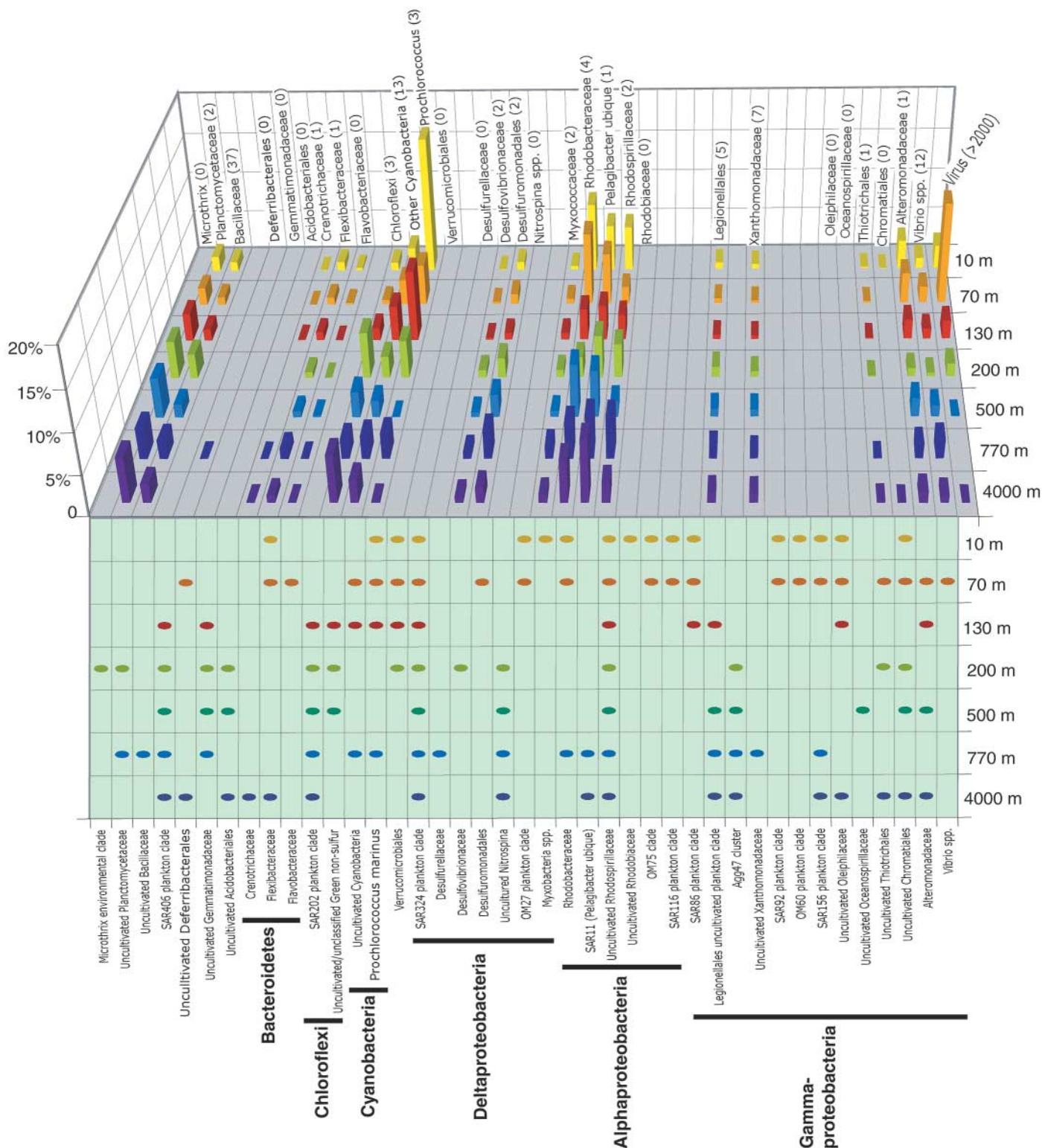
Fig. 2. Taxon distributions of top HSPs. The percent top HSPs that match the taxon categories shown at expectation values of $\leq 1 \times 10^{-60}$. Values in parentheses indicate number of genomes in each category, complete or draft, that were in the database at the time of analysis. The dots in the lower panel tabulate the SSU rRNAs detected in fosmid libraries from each taxonomic group at each depth (35) (figs. S3 and S6).

ly flagellar motor and hook protein-encoding genes), and type III secretory pathways (all associated with flagellar biosynthesis) in PZ (Fig. 4A).

DW sequences were enriched in several KEGG categories, including glyoxylate and dicarboxylate metabolism (with high representation of isocitrate lyase– and formate dehydrogenase–

like genes); protein folding and processing (predominantly chaperone and protease like genes); type II secretory genes (~40% were most similar to pilin biosynthesis genes); aminophospho-

Fig. 3. Habitat-specific sequences in photic zone versus deep-water communities. The dendrogram shows a cluster analysis based on cumulative bitscores derived from reciprocal TBLASTX comparisons between all depths. Only the branching pattern resulting from neighbor-joining analyses (not branch-lengths) are shown in the dendrogram. The Venn diagrams depict the percentage of sequences that were present only in PZ sequences (n = 12,713) or DW sequences (n = 14,132), as determined in reciprocal BLAST searches of all sequences in each depth versus every other. The percentage out of the total PZ or DW sequence bins represented in each subset is shown. See SOM for methods (35).
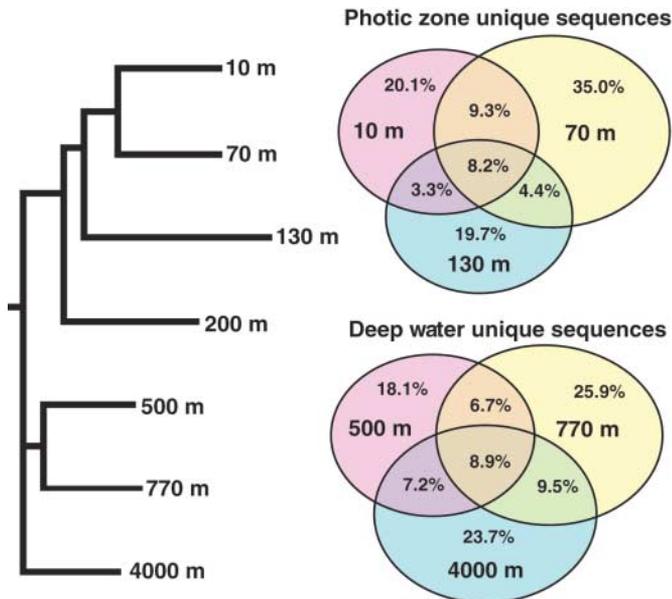
nate, methionine, and sulfur metabolism; butanoate metabolism; ion-coupled transporters; and other ABC transporter variants (Fig. 4B). The high representation in DW sequences of type II secretion system and pilin biosynthesis genes, polysaccharide, and antibiotic synthesis suggest a potentially greater role for surface-associated microbial processes in the deeper-water communities. Conversely, enrichment of bacterial motility and chemotaxis sequences in the photic zone indicates a potentially greater importance for mobility and response in these assemblages.

Similar differential patterns of sequence distribution were seen in COG categories (Fig. 4B). COGs enriched in the PZ sequence bin included photolyases, iron-transport outer membrane proteins, Na$^+$-driven efflux pumps, ABC-type sugar-transport systems, hydrolases and acyl transferases, and transaldolases. In deeper waters, transposases were the most enriched COG category (~4.5% of the COG-categorized DW), increasing steadily in representation with depth from 500 m to their observed maximum at 4000 m (Fig. 4B; fig. S9). Transposases represented one of the single-most overrepresented COG categories in deep waters, accounting for 1.2% of all fosmids sequenced from 4000 m (fig. S8). Preliminary analyses of the transposase variants and mate-pair sequences indicate that they represent a wide variety of different transposase families and originate from diverse microbial taxa. In contrast, other highly represented COG categories appeared to reflect specific taxon distribution and abundances. For example, the enrichment of transaldolases at 70 m (Fig. 4B; fig. S9) were mostly derived from abundant cyanophage DNA that was recovered at that depth (see discussion below).

Sargasso Sea surface-water microbial sequences (19) shared, as expected, many more homologous sequences with our photic zone sequences than those from the deep sea (fig. S10). There were 10 times as many PZ than DW sequences shared in common with Sargasso Sea samples 5 through 7 (19) (fig. S10). In contrast, PZ-like sequences were only three times higher in DW when compared with sequences from Sargasso Sea sample 3 (fig. S10). The fact that Sargasso sample 3 was collected during a period of winter deep-water mixing likely contributes to this higher representation of DW-like homologs. Sargasso Sea homologs of our PZ sequence bin included, as expected, sequences associated with photosynthesis; amino acid transport; purine, pyrimidine and nitrogen metabolism; porphyrin and chlorophyll metabolism; oxidative phosphorylation; glycolysis; and starch and sucrose metabolism (fig. S10).

Tentative taxonomic assignments of PZ or DW sequences (top HSPs from NCBI's nonredundant protein database) were also tabulated (fig. S11). As expected, a high percentage of *Prochlorococcus*-like sequences was found in PZ (~5% of the total), and a greater representation of Deltaproteobacteria-like, *Actinobacteria*-like and *Planctomycete*-like sequences were recovered in DW. Unexpectedly, the single most highly represented taxon category in PZ (~21% of all identified sequences in PZ) was derived from viral sequences that were captured in fosmid clones (fig. S11).

Community Genomics and Host-Virus Interactions

Viruses are ubiquitous and abundant components of marine plankton, and influence lateral gene transfer, genetic diversity, and bacterial

mortality in the water column (37–40). The large number of viral DNA sequences in our dataset was unexpected (Fig. 5; fig. S12), because we expected planktonic viruses to pass through our collection filters. Previous studies using a similar approach found only minimal contributions from viral sources (19, 40). The majority of viral DNA we captured in fosmid clone libraries apparently originates from replicating viruses within infected host cells (35). Viral DNA recovery was highest in the photic zone, with cyanophage-like sequences representing 1 to 10% of all fosmid sequences (Fig. 5), and 60 to 80% of total virus sequences there. Below 200 m, viral DNA made up no more than 0.3% of all sequences at each depth. Most photic zone viral sequences shared highest similarity to T7-like and T4-like cyanophage of the Podoviridae and Myoviridae. This is consistent with previous studies (40–42), suggesting a widespread distribution of these phage in the ocean.

Analyses of 1107 fosmid mate pairs provided further insight into the origins of the viral sequences. About 67% of the viruslike clones were most similar to cyanophage on at least one end, and half of these were highly similar to cyanophage at both termini. Many of the cyanophage clones showed apparent synteny with previously sequenced cyanophage genomes (fig. S12). About 11% of the cyanophage paired-ends contained a host-derived cyanophage "signature" gene (43) on one terminus. The frequency and genetic-linkage of phage-encoded (but host-derived) genes we observed, including virus-derived genes involved in photosynthesis (*psbA*, *psbD*, *hli*), phosphate-scavenging genes (*phoH*, *pstS*), a cobalamin biosynthesis gene (*cobS*), and carbon metabolism (*transaldolase*) supports their widespread distribution in natural viral populations and their probable functional importance to cyanophage replication (43, 44).

If we assume that the cyanophages' DNA was derived from infected host cells in which phage were replicating, the percentage of cyanophage-infected cells was estimated to range between 1 and 12% (35). An apparent cyanophage infection maxima was observed at 70 m, coinciding with the peak virus:host ratio (Fig. 5). Although these estimates are tentative, they are consistent with previously reported ranges of phage-infected picoplankton cells in situ (38, 45).

About 0.5% of all sequences were likely prophage, as inferred from high sequence similarity to phage-related integrases and known prophage genes (35). Paired-end analyses of viral fosmids indicated that ~2.5% may be derived from prophage integrated into a variety of host taxa. A few clones also appear to be derived from temperate siphoviruses, and a number of putative eukaryotic paired-end viral sequences shared highest sequence identity with homologs from herpes viruses, mimiviruses, and algal viruses.

A

B
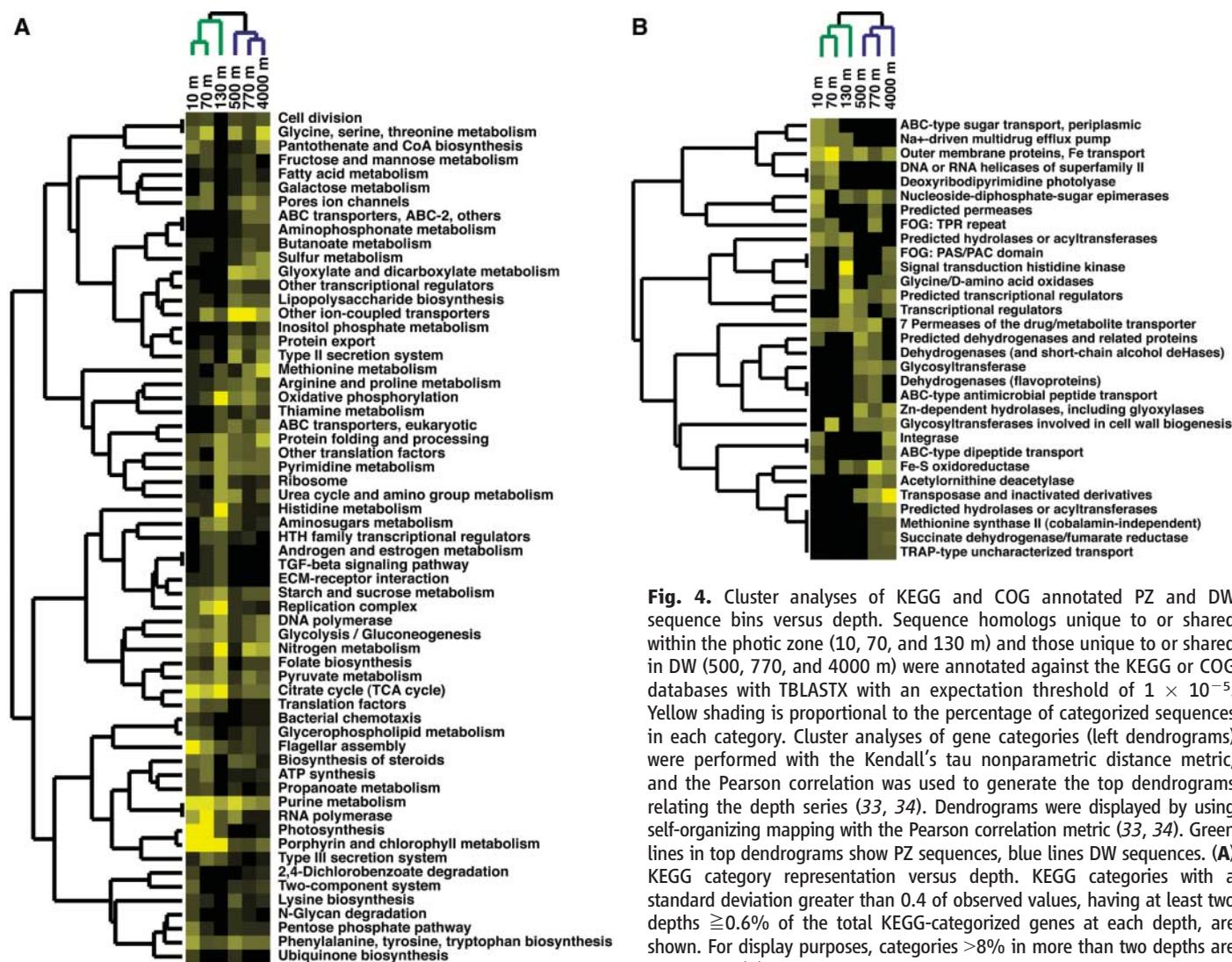
Fig. 4. Cluster analyses of KEGG and COG annotated PZ and DW sequence bins versus depth. Sequence homologs unique to or shared within the photic zone (10, 70, and 130 m) and those unique to or shared in DW (500, 770, and 4000 m) were annotated against the KEGG or COG databases with TBLASTX with an expectation threshold of 1×10^{-5}. Yellow shading is proportional to the percentage of categorized sequences in each category. Cluster analyses of gene categories (left dendrograms) were performed with the Kendall's tau nonparametric distance metric, and the Pearson correlation was used to generate the top dendrograms relating the depth series (*33*, *34*). Dendrograms were displayed by using self-organizing mapping with the Pearson correlation metric (*33*, *34*). Green lines in top dendrograms show PZ sequences, blue lines DW sequences. (**A**) KEGG category representation versus depth. KEGG categories with a standard deviation greater than 0.4 of observed values, having at least two depths ≧0.6% of the total KEGG-categorized genes at each depth, are shown. For display purposes, categories >8% in more than two depths are not shown. (**B**) COG category representation versus depth. COG categories with standard deviations greater than 0.2 of observed values, having at least two depths ≧0.3% of the total COG-categorized genes at each depth, are shown.

Ecological Implications and Future Prospects

Microbial community sampling along well-characterized depth strata allowed us to identify significant depth-variable trends in gene content and metabolic pathway components of oceanic microbial communities. The gene repertoire of surface waters reflected some of the mechanisms and modes of light-driven processes and primary productivity. Environmentally diagnostic sequences in surface waters included predicted proteins associated with cyanophage, motility, chemotaxis, photosynthesis, proteorhodopsins, photolyases, carotenoid biosynthesis, iron-transport systems, and host restriction-modification systems. The importance of light energy to these communities as reflected in their gene content was obvious. More subtle ecophysiological trends can be seen in iron transport, vitamin synthesis, flagella synthesis and secretion, and chemotaxis gene distributions. These data support hypothe-

ses about potential adaptive strategies of heterotrophic bacteria in the photic zone that may actively compete for nutrients by swimming toward nutrient-rich particles and algae (*46*). In contrast to surface-water assemblages, deep-water microbial communities appeared more enriched in transposases, pilus synthesis, protein export, polysaccharide and antibiotic synthesis, the glyoxylate cycle, and urea metabolism gene sequences. The observed enrichment in pilus, polysaccharide, and antibiotic synthesis genes in deeper-water samples suggests a potentially greater role for a surface-attached life style in deeper-water microbial communities. Finally, the apparent enrichment of phage genes and restriction-modification systems observed in the photic zone may indicate a greater role for phage parasites in the more productive upper water column, relative to deeper waters.

At finer scales, sequence distributions we observed also reflected genomic "microvari-

ability" along environmental gradients, as evidenced by the partitioning of high- and low-light *Prochlorococcus* ecotype genes observed in different regions of the photic zone (Fig. 5). Higher-order biological interactions were also evident, for example in the negative correlation of cyanophage versus *Prochlorococcus* host gene sequence recovery (Fig. 5). This relation between the abundance of host and cyanophage DNA probably reflects specific mechanisms of cyanophage replication in situ. These host-parasite sequence correlations we saw demonstrate the potential for observing community-level interspecies interactions through environmental genomic datasets.

Obviously, the abundance of specific taxa will greatly influence the gene distributions observed, as we saw, for example, in *Prochlorococcus* gene distribution in the photic zone. Gene sequence distributions can reflect more than just relative abundance of specific taxa, however.

Fig. 5. Cyanophage and cyanobacteria distributions in microbial community DNA. The percentage of total sequences derived from cyanophage, total cyanobacteria, total *Prochlorococcus* spp., high-light *Prochlorococcus*, low-light *Prochlorococcus* spp., or *Synechococcus* spp., from each depth. Taxa were tentatively assigned according to the origin of top HSPs in TBLASTX searches, followed by subsequent manual inspection and curation.

Percent total sequences per depth

enabled ecological studies matures, it should become possible to model microbial community genomic, temporal, and spatial variability with other environmental features. Significant future attention will no doubt focus on interpreting the complex interplay between genes, organisms, communities and the environment, as well as the properties revealed that regulate global biogeochemical cycles. Future efforts in this area will advance our general perspective on microbial ecology and evolution and elucidate the biological dynamics that mediate the flux of matter and energy in the world's oceans.

References and Notes

1. E. Forbes, in *Physical Atlas of Natural Phenomena*, A. K. Johnston, Ed. (William Blackwood & Sons, London and Edinburgh, 1856).
2. P. W. Hochachka, G. N. Somero, *Biochemical Adaptation* (Princeton Univ. Press, Princeton, NJ, 1984), pp. 450–495.
3. G. Rocap *et al.*, *Nature* **424**, 1042 (2003).
4. Z. I. Johnson *et al.*, manuscript submitted.
5. N. J. West *et al.*, *Microbiology* **147**, 1731 (2001).
6. A. A. Yayanos, *Annu. Rev. Microbiol.* **49**, 777 (1995).
7. N. R. Pace, *Science* **276**, 734 (1997).
8. M. S. Rappé, S. J. Giovannoni, *Annu. Rev. Microbiol.* **57**, 369 (2003).
9. M. T. Suzuki *et al.*, *Microb Ecol* (2005).
10. O. Béjà *et al.*, *Environ. Microbiol.* **2**, 516 (2000).
11. R. M. Morris, M. S. Rappé, E. Urbach, S. A. Connon, S. J. Giovannoni, *Appl. Environ. Microbiol.* **70**, 2836 (2004).
12. S. Y. Moon-van der Staay *et al.*, *Nature* **409**, 607 (2001).
13. J. A. Fuhrman, K. McCallum, A. A. Davis, *Nature* **356**, 148 (1992).
14. E. F. DeLong, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5685 (1992).
15. M. B. Karner *et al.*, *Nature* **409**, 507 (2001).
16. J. Handelsman, *Microbiol. Mol. Biol. Rev.* **68**, 669 (2004).
17. E. F. Delong, *Nat Rev Microbiol* (2005).
18. G. W. Tyson *et al.*, *Nature* **428**, 37 (2004).
19. J. C. Venter *et al.*, *Science* **304**, 66 (2004).
20. S. G. Tringe *et al.*, *Science* **308**, 554 (2005).
21. S. J. Hallam *et al.*, *Science* **305**, 1457 (2004).
22. D. M. Karl, R. Lukas, *Deep-Sea Res. II* **43**, 129 (1996).
23. D. M. Karl *et al.*, *Deep-Sea Res. II* **48**, 1449 (2001).
24. R. M. Letelier *et al.*, *Limnol. Oceanogr.* **49**, 508 (2004).
25. E. F. DeLong *et al.*, *Appl. Environ. Microbiol.* **65**, 5554 (1999).
26. A. Pernthaler *et al.*, *Appl. Environ. Microbiol.* **68**, 661 (2002).
27. N. U. Frigaard *et al.*, *Nature,* in press.
28. S. M. Barns, C. F. Delwiche, J. D. Palmer, N. R. Pace, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9188 (1996).
29. M. Kanehisa *et al.*, *Nucleic Acids Res.* **32**, D277 (2004).
30. R. L. Tatusov *et al.*, *BMC Bioinformatics* **4**, 41 (2003).
31. R. Overbeek *et al.*, *Nucleic Acids Res.* **33**, 5691 (2005).
32. S. F. Altschul *et al.*, *J. Mol. Biol.* **215**, 403 (1990).
33. M. J. L. de Hoon *et al.*, *Bioinformatics* **12**, 1453 (2004).
34. M. B. Eisen, P. T. Spellman, P. O. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 14863 (1998).
35. Materials and methods are available as supporting material on *Science* Online.
36. K. D. Pruitt, T. Tatusova, D. R. Maglott, *Nucleic Acids Res.* **33**, D501 (2005).
37. M. G. Weinbauer, *FEMS Microbiol. Rev.* **28**, 127 (2004).

Some depth-specific gene distributions we observed [e.g., transposases found predominantly at greater depths (Fig. 4B; fig. S8)], appear to originate from a wide variety of gene families and genomic sources. These gene distributional patterns seem more indicative of habitat-specific genetic or physiological trends that have spread through different members of the community. Community gene distributions and stoichiometries are differentially propagated by vertical and horizontal genetic mechanisms, dynamic physiological responses, or interspecies interactions like competition. The overrepresentation of certain sequence types may sometimes reflect their horizontal transmission and propagation within a given community. In our datasets, the relative abundance of cyanobacteria-like *psbA*, *psbD*, and transaldolase genes were largely a consequence of their horizontal transfer and subsequent amplification in the viruses that were captured in our samples. In contrast, the increase of transposases from 500 to 4000 m, regardless of community composition, reflected a different mode of gene propagation, likely related to the slower growth, lower

productivity, and lower effective population sizes of deep-sea microbial communities. In future comparative studies, similar deviations in environmental gene stoichiometries might be expected to provide even further insight into habitat-specific modes and mechanisms of gene propagation, distribution, and mobility (*27*, *47*). These "gene ecologies" could readily be mapped directly on organismal distributions and interactions, environmental variability, and taxonomic distributions.

The study of environmental adaptation and variability is not new, but our technical capabilities for identifying and tracking sequences, genes, and metabolic pathways in microbial communities is. The study of gene ecology and its relation to community metabolism, interspecies interactions, and habitat-specific signatures is nascent. More extensive sequencing efforts are certainly required to more thoroughly describe natural microbial communities. Additionally, more concerted efforts to integrate these new data into studies of oceanographic, biogeochemical, and environmental processes are necessary (*48*). As the scope and scale of genome-

38. J. Waterbury, F. Valois, *Appl. Environ. Microbiol.* **59**, 3393 (1993).
39. M. B. Sullivan *et al.*, *Nature* **424**, 1047 (2003).
40. R. A. Edwards, F. Rohwer, *Nat. Rev. Microbiol.* **3**, 504 (2005).
41. J. Filee, F. Tetart, C. A. Suttle, H. M. Krisch, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12471 (2005).
42. M. Breitbart, J. H. Miyake, F. Rohwer, *FEMS Microbiol. Lett.* **236**, 249 (2004).
43. M. B. Sullivan *et al.*, *PLoS Biol.* **3**, e144 (2005).
44. D. Lindell *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11013 (2004).
45. M. G. Weinbauer, I. Brettar, M. G. Hofle, *Limnol. Oceanogr.* **48**, 1457 (2003).
46. F. Azam, *Science* **280**, 694 (1998).
47. C. R. Woese, *Microbiol. Mol. Biol. Rev.* **68**, 173 (2004).
48. E. F. DeLong, D. M. Karl, *Nature* **437**, 336 (2005).
49. ⟨http://hahana.soest.hawaii.edu/hot/parameters.html⟩.
50. ⟨http://hahana.soest.hawaii.edu/hot/hot-dogs⟩.
51. This work was supported by a grant from the Gordon and Betty Moore Foundation, NSF grants MCB-0084211, MCB-0348001, and MCB-0509923 to E.F.D., and OCE-0326616 to D.M.K., and sequencing support from the U.S. Department of Energy Microbial Genomics Program. We thank Dennis Ryan for help with scripting and sequence analyses, the officers and crew of the R/V Ka'imikai-O-Kanaloa and the HOT team for assistance at sea, and L. Fujieki for help with oceanographic data display. J. Chapman, A. Salamov, and P. Richardson of the DOE Joint Genome Institute provided advice and assistance in DNA sequencing and analyses. Sequences have been deposited in GenBank with accession numbers DU731018-DU796676 and DU800850-DU800864 corresponding to fosmid end sequences, and accession numbers DQ300508-DQ300926 corresponding to SSU rRNA gene sequences.

REPORTS

Pairing and Phase Separation in a Polarized Fermi Gas

Guthrie B. Partridge, Wenhui Li, Ramsey I. Kamar, Yean-an Liao, Randall G. Hulet*

We report the observation of pairing in a gas of atomic fermions with unequal numbers of two components. Beyond a critical polarization, the gas separates into a phase that is consistent with a superfluid paired core surrounded by a shell of normal unpaired fermions. The critical polarization diminishes with decreasing attractive interaction. For near-zero polarization, we measured the parameter $\beta = -0.54 \pm 0.05$, describing the universal energy of a strongly interacting paired Fermi gas, and found good agreement with recent theory. These results are relevant to predictions of exotic new phases of quark matter and of strongly magnetized superconductors.

Fermion pairing is the essential ingredient in the Bardeen, Cooper, and Schrieffer (BCS) theory of superconductivity. In conventional superconductors, the chemical potentials of the two spin states are equal. There has been great interest, however, in the consequences of mismatched chemical potentials that may arise in several important situations, including, for example, magnetized superconductors (*1–3*) and cold dense quark matter at the core of neutron stars (*4*). A chemical potential imbalance may be produced by several mechanisms, including magnetization in the case of superconductors, mass asymmetry, or unequal numbers. Pairing is qualitatively altered by the Fermi energy mismatch, and there has been considerable speculation regarding the nature and relative stability of various proposed exotic phases. In the Fulde-Ferrel-Larkin-Ovchinnikov (FFLO) phase (*2, 3*), pairs possess a nonzero center-of-mass momentum that breaks translational invariance, whereas the Sarma (*1*), or the breached pair (*5*), phase is speculated to have gapless excitations. A mixed phase has also been proposed (*6–8*) in which regions of a paired BCS superfluid are surrounded by an unpaired normal phase. Little is known experimentally, however, because of the difficulty in creating magnetized superconductors. Initial evidence for an FFLO phase in a heavy-fermion superconductor has only recently been reported (*9, 10*). Opportunities for experimental investigation of exotic pairing states have expanded dramatically with the recent realization of the Bose-Einstein condensate (BEC)–BCS crossover in a two spin state mixture of ultracold atomic gases. Recent experiments have demonstrated both superfluidity (*11–13*) and pairing (*14–17*) in atomic Fermi gases. We report the observation of pairing in a polarized gas of ^{6}Li atoms. Above an interaction-dependent critical polarization, we observed a phase separation that is consistent with a uniformly paired superfluid core surrounded by an unpaired shell of the excess spin state. Below the critical polarization, the spatial size of the gas was in agreement with expectations for a universal, strongly interacting paired Fermi gas.

Our methods for producing a degenerate gas of fermionic ^{6}Li atoms (*18, 19*) and the realization of the BEC-BCS crossover at a Feshbach resonance (*17*) have been described previously (*20*). An incoherent spin mixture of the $F = \frac{1}{2}$, $m_F = \frac{1}{2}$ (state $|1\rangle$) and the $F = \frac{1}{2}$, $m_F = -\frac{1}{2}$ (state $|2\rangle$) sublevels (where F is the total spin quantum number and m_F is its projection) is created by radio frequency (rf) sweeps, where the relative number of the two states can be controlled by the rf power (*20*). The spin mixture is created at a magnetic field of 754 G, which is within the broad Feshbach resonance located near 834 G (*21, 22*). The spin mixture is evaporatively cooled by reducing the depth of the optical trap that confines it, and the magnetic field is ramped adiabatically to a desired field within the crossover region. States $|1\rangle$ and $|2\rangle$ are sequentially and independently imaged in the trap by absorption (*20*). Analysis of these images provides measurement of N_i and polarization $P = (N_1 - N_2)/(N_1 + N_2)$, where N_i is the number of atoms in state $|i\rangle$. We express the Fermi temperature, T_F, in terms of the majority spin state, state $|1\rangle$, as $k_B T_F = \hbar\overline{\omega} (6N_1)^{1/3}$, where $\overline{\omega} = 2\pi (v_r^2 v_z)^{1/3}$ is the mean harmonic frequency of the cylindrically symmetric confining potential with radial and axial frequencies v_r and v_z, respectively. For $P \approx 0$, we find that $N_1 \approx N_2 \approx 10^5$, giving $T_F \approx 400$ nK for our trap frequencies. Because of decreasing evaporation efficiency with increasing polarization, there is a correlation between P and total atom number (fig. S1).

For fields on the low-field (BEC) side of resonance, real two-body bound states exist, and molecules are readily formed by three-body recombination. For the case of $P = 0$, a molecular Bose-Einstein condensate (MBEC) is observed to form with no detectable thermal molecules (*17*). On the basis of an estimated MBEC condensate fraction of >90%, we place an upper limit on the temperature $T < 0.1 T_F$ at a field of 754 G (*17*). However, the gas is expected to be cooled further during the adiabatic ramp for final fields greater than 754 G (*17*). By using similar experimental methods, we previously measured the order parameter of the gas in the BCS regime and found good agreement with $T = 0$ BCS theory (*17*), indicating that the gas was well below the critical temperature for pairing.

Images of states $|1\rangle$ and $|2\rangle$ at a field of 830 G are shown (Fig. 1) for relative numbers

Department of Physics and Astronomy and Rice Quantum Institute, Rice University, Houston, TX 77251, USA.

*To whom correspondence should be addressed. E-mail: randy@rice.edu