MIT Open Access Articles

Microbial community phylogenetic and trait diversity declines with depth in a marine oxygen minimum zone

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Bryant, Jessica A., Frank J. Stewart, John M. Eppley, and Edward F. DeLong. "Microbial Community Phylogenetic and Trait Diversity Declines with Depth in a Marine Oxygen Minimum Zone." *Ecology* 93, no. 7 (July 2012): 1659–1673. © 2012 Ecological Society of America

As Published: <http://dx.doi.org/10.1890/11-1204.1>

Publisher: Ecological Society of America

Persistent URL: <http://hdl.handle.net/1721.1/101417>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.

Microbial community phylogenetic and trait diversity declines with depth in a marine oxygen minimum zone

JESSICA A. BRYANT,¹ FRANK J. STEWART,² JOHN M. EPPLEY,¹ AND EDWARD F. DELONG^{1,3}

¹*Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Parsons Laboratory 48, 15 Vassar Street, Cambridge, Massachusetts 02139 USA*

²*School of Biology, Georgia Institute of Technology, 311 Ferst Drive, Atlanta, Georgia 30332 USA*

Abstract. Oxygen minimum zones (OMZs) are natural physical features of the world's oceans. They create steep physiochemical gradients in the water column, which most notably include a dramatic draw down in oxygen concentrations over small vertical distances (<100 m). Microbial communities within OMZs play central roles in ocean and global biogeochemical cycles, yet we still lack a fundamental understanding of how microbial biodiversity is distributed across OMZs. Here, we used metagenomic sequencing to investigate microbial diversity across a vertical gradient in the water column during three seasons in the Eastern Tropical South Pacific (ETSP) OMZ. Based on analysis of small subunit ribosomal RNA (SSU rRNA) gene fragments, we found that both taxonomic and phylogenetic diversity declined steeply along the transition from oxygen-rich surface water to the permanent OMZ. We observed similar declines in the diversity of protein-coding gene categories, suggesting a decrease in functional (trait) diversity with depth. Metrics of functional and trait dispersion indicated that microbial communities are phylogenetically and functionally more over-dispersed in oxic waters, but clustered within the OMZ. These dispersion patterns suggest that community assembly drivers (e.g., competition, environmental filtering) vary strikingly across the oxygen gradient. To understand the generality of our findings, we compared OMZ results to two marine depth gradients in subtropical oligotrophic sites and found that the oligotrophic sites did not display similar patterns, likely reflecting unique features found in the OMZ. Finally, we discuss how our results may relate to niche theory, diversity–energy relationships and stress gradients.

Key words: community ecology; functional diversity; microbial diversity; oxygen minimum zones; phylogenetic diversity; trait diversity.

INTRODUCTION

Microorganisms compose the vast majority of biodiversity in the oceans and play crucial roles in global biogeochemical cycles, yet the factors shaping and constraining marine microbial diversity are complex and poorly understood. The first step in demystifying these processes is documenting biodiversity. There is a rich tradition of documenting plant and animal diversity patterns (Lomolino et al. 2006), but quantitative documentation of microorganismal patterns is in its infancy. Studies have shown that different depths in the ocean harbor distinct microbial communities and suggest that general ecological patterns well documented in macro-organisms, such as the relationship between species diversity and productivity or species diversity and latitude, may hold true for microorganisms (DeLong 2006, Smith 2007, Fuhrman et al. 2008).

Marine oxygen minimum zones (OMZs) create a steep physiochemical gradient in the water column (Galán et al. 2009), and so are excellent model systems for

investigating how dramatic environmental variation over relatively small spatial scales shapes marine microbial diversity. OMZs form where nutrient-rich water supports high quantities of primary production. As this biomass sinks, it is respired by heterotrophs and the oxygen provided by photosynthesis is rapidly depleted. This biological activity, coupled with local current conditions that prohibit oxygen resupply from lateral or vertical advection and diffusion, maintain anoxic conditions at intermediate depths in the water column over large spatial areas. OMZs harbor unique microbial communities with important biogeochemical roles. For example, microbial activity in OMZs is responsible for 30–50% of nitrogen loss in the ocean (Lam et al. 2011). OMZs are inhospitable to many higher organisms, and are additionally of pressing interest because they are expanding due to anthropogenic effects, including climate change (Stramma et al. 2008). Although several metabolic types have been investigated in targeted studies (e.g., Canfield et al. 2010, Lam and Kuypers 2011), quantitative measurements of overall microbial diversity along this physiochemical gradient are not yet well described.

Manuscript received 3 July 2011; revised 5 December 2011; accepted 5 January 2012. Corresponding Editor: J. B. Yavitt.

³ E-mail: delong@mit.edu

Microbial community diversity studies are challenging in part because quantifying a parameter as seemingly simple as “species” richness is not straightforward for microbial communities. Studies have relied on genetic proxies to enumerate diversity, such as fingerprinting approaches, or the binning of marker genes into taxonomic groups based on genetic similarity (e.g., 97% small subunit ribosomal RNA [SSU rRNA] sequence similarity as a “species” threshold). Incorporating phylogenetic information into ecological analyses facilitates the identification of drivers of macro- and microorganism biodiversity (Webb et al. 2002, Chave et al. 2007, Lozupone and Knight 2007). Phylogenetic diversity (PD), defined here as the total branch length connecting all taxa in a phylogeny, is a clear phylogenetic equivalent to species richness (Faith 1992). Each taxon’s contribution to the overall PD of a community is based on its relatedness to other organisms in that community, unlike taxonomic richness where all taxa contribute equally (May 1990, Faith 1994). PD is a more ecologically relevant measure of diversity than simple taxonomic richness, as PD is believed to reflect trait diversity.

There has been a growing interest in quantifying functional (or trait) diversity directly in metazoan or plant biodiversity studies. Functional units are defined by either measuring the variation within a set of relevant phenotypic traits or by binning organisms into guilds (reviewed in Schleuter et al. 2010). These approaches are challenging to apply because they require extensive ecological and morphological data (Kraft et al. 2008, Kraft and Ackerly 2009, Lake and Ostling 2009). Currently, gathering such data for diverse communities of microorganisms would be extremely challenging, as most microorganisms in natural communities defy cultivation and characterizations using standard methods.

As an alternative, metagenomic (community genome sequencing) approaches allow significant proportions of the genomes originating from microbial communities (the metagenome) to be sequenced (DeLong et al. 2006). Metagenomes provide a “parts list” of the traits microbes can utilize to persist in their environment (Tringe and Rubin 2005). Parallel to species richness and phylogenetic diversity, the number of unique gene categories identified in a community metagenome equates to that community’s functional richness (FR) (Raes et al. 2011). This approach to quantifying functional diversity circumvents many drawbacks of traditional marker gene (e.g., SSU) methods that can potentially decouple phylogeny from function (lateral gene transfer, multiple gene copies per genome). However, unless lateral gene transfer or the occurrence of multiple marker gene copy numbers predominates, taxonomic richness, PD, and FR should correlate with one another (Petchey and Gaston 2002, Rodrigues and Gaston 2002, Morlon et al. 2011). FR is a promising metric for understanding community responses to environmental gradients or ecological pressures, as it is

the most direct link to the mechanisms by which organisms in a community interact with each other and their environment.

It is difficult to predict how microbial diversity will be distributed across the water column. Parallels have been made between marine depth gradients and the more commonly studied latitudinal and elevational gradients because all three involve complex changes in a variety of environmental attributes, including pressure, temperature, nutrient availability, or productivity (Smith and Brown 2002). Microbes have shown both decreasing and hump-shaped diversity patterns across latitudinal and elevation gradients and therefore could demonstrate similar trends across depth gradients (Bryant et al. 2008, Fuhrman et al. 2008, but see Fierer et al. 2011).

Measures of trait richness and phylogenetic dispersion, when combined with null models based on regional taxon pools, provide insight into the potential dominant forces structuring biodiversity (Webb et al. 2002, Cornwell et al. 2006). The most commonly used measure of phylogenetic dispersion, the extent to which community members are clustered (or spread evenly) throughout a phylogeny, is the mean phylogenetic distance between all pairs of taxon (MPD) in a community (Webb et al. 2002). To test if the MPD of a local assemblage is larger (phylogenetically overdispersed) or smaller (phylogenetically clustered) than expected by chance, null communities are generated by randomly resampling taxa from a regional taxa pool. The distribution of null community MPDs is then compared to the focal community MPD (Webb et al. 2002).

Similarly, FR may be larger (functional overdispersion) or smaller (functional clustering) than expected by chance. There are an absolute number of genomes in a given sample, each with a set of traits (gene categories). If genomes within the local sample are functionally very similar (trait clustering) or very different (trait overdispersion), then the sample will display a lower or higher number of total unique traits (trait richness) respectively, relative to a distribution of null community FRs generated by resampling a regional genome pool.

MPD, assuming that niches are phylogenetically conserved (more closely related organisms have more similar niche preferences), and FR can provide insight into the dominant community assembly processes. Phylogenetic clustering within communities, as well as trait clustering, supports the hypothesis that environmental filtering is the dominant force structuring community assembly (Weiher et al. 1998, Webb et al. 2002; but see Mayfield and Levine 2010). Environmental filtering is a process where abiotic conditions select for the organisms with similar traits that are best adapted to survive prevailing conditions. Alternatively, phylogenetic or trait overdispersion supports the hypothesis that competitive exclusion is the dominant force structuring community assembly, whereby greater competition between similar species leads to co-occurrence of

predominately distantly related species with different niches (Tofts and Silvertown 2000, Webb et al. 2002).

Many forces act in concert to assemble communities, but the relative importance of each can vary depending on environmental conditions (Bryant et al. 2008). Environmental gradients provide natural experiments to test the relative importance of community assembly processes under varying conditions. For example, Horner-Devine and Bohannan (2006) found that betaproteobacteria became increasingly phylogenetically dispersed with increasing total organic carbon in a set of mesocosm experiments, suggesting that competition has a greater role in structuring betaproteobacteria communities in more productive environments. We predict that signals for environmental filtering will predominate in locations where abiotic conditions place the largest constraint on organism growth and that signals for competition will predominate in places where the environment is more conducive to growth.

Here, we analyzed 17 metagenomic data sets to quantify depth-specific phylogenetic and functional diversity and dispersion patterns along a depth gradient encompassing the OMZ in the Eastern Tropical South Pacific (ETSP) off the coast of Iquique, Chile. Our samples span three seasons (fall, winter, summer), as well as dramatic gradients in light, oxygen, and nutrient availability. To put our results in a broader context, we compared ETSP results to two depth gradients sampled in oligotrophic regions of the ocean.

METHODS

Site description and sample collection

Microbial community DNA samples previously published in Canfield et al. (2010) and Stewart et al. (2011), along with additional previously unpublished samples (see Table 1 for details) were collected from Station #3 (20°07' S, 70°23' W) off the coast of Iquique, Chile in austral fall (June 2008), winter (August 2009), and summer (January 2010) as part of three Microbial Oceanography of Oxygen Minimum Zone (MOOMZ) cruises aboard the R/V *Vidal Gormaz*. Sample collection procedures have been described in detail (Canfield et al. 2010, Stewart et al. 2011a). In brief, seawater samples were collected using either 10-L Niskin bottles deployed on a rosette system (all 2008 samples; 35-m, 50-m, 110-m, and 200-m 2009 samples) or from an in situ pump-cast system (70-m 2009 sample and all 2010 samples). Samples were pre-filtered through 1.6 µm GF/A filters (125 mm diameter, Whatman) and collected on either 0.22 Steripak-GP20 (2008 samples) or 0.22 Sterivex filters (2009 and 2010 samples; Millipore, Billerica, Massachusetts, USA). Previously published oligotrophic ocean samples of the same size fraction were collected in October 2006 from the Hawaiian Ocean Time Series Station (HOT) ALOHA (22°44' N, 158°2' W) and the Bermuda Atlantic Time Series (BATS) (31°40' N, 64°10' W) as described previously (Frias-Lopez et al. 2008, Coleman and Chisholm 2010). The volume of

water filtered for each oligotrophic sample is reported in Appendix A: Table A1.

DNA extraction and pyrosequencing

Genomic DNA extractions were conducted as previously described (ETSP 2008, BATS, and HOT samples [Frias-Lopez et al. 2008], ETSP 2009 and 2010 samples [Canfield et al. 2010]). Briefly, cells were lysed by adding lysis buffer (50 mmol/L Tris-HCl, 40 mmol/L EDTA, and 0.75 mol/L sucrose) and lysozyme directly to filters, followed by a proteinase K digestion. DNA was then purified from ETSP 2008, BATS, and HOT lysate using column purification and phenol chloroform extraction. ETSP 2009 and 2010 lysate was purified using the QuickGene-610L system (Fujifilm, Tokyo, Japan). The extracted genomic DNA was used for the generation of single-stranded DNA libraries and emulsion PCR according to standard protocols (454 Life Sciences, Roche, Branford, Connecticut, USA). Clonally amplified library fragments were sequenced on a Roche Genome Sequencer FLX instrument using FLX (ETSP 2008, HOTS, and BATS samples) or Titanium (ETSP 2009 and 2010) series reagents. Artificially created identical duplicate sequences were removed from data sets prior to analyses using the sequence-clustering package CD-HIT (Li and Godzik 2006, Stewart et al. 2010). Titanium series reagents and sequencing protocols yield a greater number of sequences and longer read lengths than do FLX series reagents and protocols. The FLX and Titanium reagents and protocols however are not expected to selectively target different genes or taxa, since the major difference between these two methods is simply the length of DNA fragments captured for sequencing.

SSU rRNA and protein-coding gene assignments

MegaBLAST searches against the ARB-SILVA non-redundant (99% identity criterion) SSU rRNA reference database were used to identify SSU genes within each genomic DNA data set (Altschuel et al. 1999, Pruesse et al. 2007; release 102). Reads identified as SSU genes were assigned to their closest relative (top BLAST hit with a bit-score of 50 or above) in the database and the corresponding leaf on the ARB-SILVA non-redundant SSU reference guide phylogeny (Hamady et al. 2010). For the purpose of this study, ARB-SILVA SSU reference sequences were considered to be taxa. The ARB-SILVA non-redundant SSU reference guide tree was pruned to include only sequences that contained taxa with assigned reads from each data set using the ARB software program (Ludwig et al. 2004). Taxa were also binned into phylum- or class-level groups according to ARB-SILVA taxonomy.

Non-rRNA sequences within each DNA data set were compared to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database using BLASTX homology searches (Kanehisa and Groto 2000; as of 30 May 2010). Reads were annotated as their top hit KEGG homologue

TABLE 1. Eastern Tropical South Pacific (ETSP) sample and sequence library characteristics.

Depth series	Depth of sample (m)	Volume of water filtered (L)	No. sequenced reads†	Average read length (bps)	No. SSU reads	SSU identity to ARB-SILVA (%)‡	SSU alignment length§	KEGG hits to total reads (%)	NCBI accession number
2008	15	20	771 623	238	1020	97.7 (3.6)	185 (87)	40.7	SRX080962
	50	15	341 163	256	485	97.7 (3.6)	191 (90)	43.0	SRX025906¶
	65	15	382 821	251	500	98.0 (3.2)	191 (87)	42.2	SRX080938
	85	17.5	569 046	253	740	97.6 (3.6)	197 (87)	43.2	SRX025908¶
	110	30	380 764	243	383	97.6 (3.2)	190 (87)	41.3	SRX025910¶
	200	16	485 911	249	540	96.5 (4.0)	187 (87)	42.1	SRX025912¶
	500	20	515 676	248	657	97.6 (3.4)	199 (83)	42.0	SRX080950
	800	15	173 051	242	338	97.6 (4.0)	203 (77)	39.8	SRX080951
2009	35	16	937 420	333	1257	97.3 (3.6)	236 (151)	45.9	SRX080953
	50	16	1 042 057	339	1370	97.6 (3.5)	242 (153)	38.5	SRX080954
	70	15	1 147 856	385	1403	96.9 (3.7)	275 (172)	47.5	SRX029170#
	110	16	905 059	403	1244	96.8 (3.7)	257 (178)	54.5	SRX080955
	200	13.5	930 359	246	1014	96.9 (3.6)	192 (117)	40.4	SRX029171#
2010	50	10	1 530 891	386	1955	96.7 (4.0)	267 (172)	51.8	SRX080956
	80	10	1 359 823	428	1918	96.4 (4.3)	291 (190)	53.3	SRX029172#
	110	11	1 456 854	409	2003	96.1 (4.3)	282 (179)	52.7	SRX080960
	150	11	1 301 664	431	1944	96.4 (4.1)	283 (190)	36.2	SRX029173#

† Duplicate reads not included.

‡ Average percentage sequence identity of each small subunit ribosomal RNA (SSU rRNA) read to its closest hit in the ARB-SILVA database. Standard deviation is in parentheses.

§ Average alignment length of each SSU read to its top hit in the ARB-SILVA database. Standard deviation is in parentheses.

¶ Previously published in Stewart et al. (2011).

Previously published in Canfield et al. (2010).

with a bit-score of 50 or greater and assigned to the top hit's corresponding KEGG ortholog (ko) gene category and then KEGG pathway. To investigate database effects, we chose the 2009 depth series as a representative series and classified each 2009 data set using the same approach according to annotations in the eggNOG database (Muller et al. 2010). EggNOG is a fusion of the clusters of orthologous groups (COG) data set and non-supervised orthologous groups (NOG) data sets. We also used a database-independent approach to assign protein-coding genes in ETSP samples to protein clusters. We first used Metagene, an ab initio gene-finding program, to identify amino acid sequences in each 454 library (Noguchi et al. 2006). We then pooled the amino acid sequences that originated from the same depth series, and clustered amino acid sequences with a 60% identity threshold using cd-hit (Li and Godzik 2006).

Diversity measures

We used the SSU data to quantify community diversity based on two metrics: (1) taxonomic richness (TR), the number of unique taxa identified in each data set, and (2) phylogenetic diversity (PD), as described in *Introduction* and calculated using the pd function in the R package Picante (Faith 1992, Kembel et al. 2010, R Development Core Team 2011). We did not include the root from a larger regional phylogeny while calculating PD (Cadotte et al. 2010). We quantified functional richness (FR) within each community as the total number of unique KEGG ko gene classifications, protein clusters or eggNOG orthologous groups, referred to as FR_K, FR_{PC}, and FR_{EN}, respectively.

To accurately compare TR, PD, and FR between data sets with differing numbers of reads within the same depth series, we normalized each data set to the smallest data set in its respective depth series using rarefaction. For example, all 2009 KEGG TR data sets were normalized to the 2009 sample with the fewest total number of hits to the ARB-SILVA database. We then used these values in calculating model statistics (see *Correlation with depth*). However, for display and comparative purposes (Fig. 2; Appendix D: Figs. D2–4), we normalized all TR and PD ETSP data sets to the 800-m 2008 sample, which had the fewest number of reads from all three depth series (years). For making comparisons between the ETSP and the oligotrophic sites (Appendix E: Fig. E2), we normalized samples from the 2009 ETSP depth series, and the HOT and BATS data sets to the BATS 500-m sample (the smallest data set in the study). Rarefaction was conducted for TR and FR using the rarefy function in the R package vegan and for PD by taking the average PD of 1000 random subsamples of each community (R Development Core Team 2011; vegan package *available online*).⁴

Dispersion metrics

Mean phylogenetic distance (MPD) and abundance-weighted MPD (aMPD) along with corresponding null expectations were calculated with the ses.mpd function in the R package Picante (Kembel et al. 2010) using the ARB-SILVA-based phylogenies. Null distributions for each data set (community) were obtained by randomly

⁴ <http://CRAN.R-project.org/package=vegan>

FIG. 1. Photosynthetically available radiation (PAR, $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$), fluorescence ($\mu\text{g/L}$), and oxygen ($\mu\text{mol/L}$) levels in the Eastern Tropical South Pacific (ETSP) during sampling periods. PAR data were not available in 2009. Black dots signify sampling depths.

sampling a pool of all taxa found in the respective depth series 999 times while constraining both the number of taxa in the community and taxa occurrence across all samples (Webb et al. 2002, Kembel and Hubbell 2006, Kembel et al. 2010). Observed MPD and aMPD values were considered significantly clustered or overdispersed ($P < 0.05$; Webb et al. 2002) if they fell within the lower or upper 2.5% of the null distribution.

Because of the fragmentary nature of metagenomic data, we could not resample whole genomes to build null distributions for FR values. Instead, we created regional trait pools for each depth series by combining all traits (KEGG ko genes) identified in all the samples within each depth series, while maintaining observed trait abundances. We constructed a distribution of null community FR values for each sample by resampling without replacement the total number of traits identified in each sample and calculating FR 999 times. This is equivalent to

resampling simulated whole genomes that were constructed using ko genes randomly selected from the entire depth series. Each randomly constructed genome had the same average genome size (number of KEGG ko genes) as the genomes within the focal sample and the null communities each had the same total number of genomes as the sample. Observed FR values were considered significantly clustered or overdispersed ($P < 0.05$) if they fell within the lower or upper 2.5% of the null distribution. Null expectations were also recalculated for the 2009 ETSP eggNOG-based FR values.

To ensure that differences in sequencing depth between data sets within the same depth series did not bias MPD, aMPD, or FR null expectation results (e.g., over-representation of larger data sets in regional pools), we subsampled each depth series, as described above for PD 100 times, recalculated MPD, aMPD, FR, and the null expectations for each subsample, and averaged

FIG. 2. Phylogenetic diversity across depths at the ETSP: orange squares represent 2008 samples, red circles represent 2009 samples, and blue triangles represent 2010 samples. Phylogenetic diversity was calculated as the sum of the branch lengths (nucleotide substitutions per site) linking all taxa in each sample across the ARB-SILVA reference phylogeny. Rarefaction analysis was used to account for differences in sequencing depth between samples by normalizing all samples in the figure to the smallest ETSP sample in the figure (ETSP 2008 800-m sample; see *Methods*).

index values and null model ranking results. R codes to perform these tasks, as well as PD rarefaction are available in the Supplement.

Correlation with depth

We used linear regression analysis to examine the correlation between rarefied diversity (TR, PD, FR_K , FR_{EN} , FR_{PC}) and dispersion (MPD, aMPD) measures with depth as well as the relationship between TR and PD with FR_K . Linear, logarithmic, and power-law models were all fit for each model of diversity or dispersion as a function of depth. Final model choice was based on meeting regression model assumptions and Akaike information criterion. We normalized each sample to the smallest sample in its respective depth series (year) and included a “sampling year” term in ETSP models of TR, PF, and FR as a function of depth to account for differences in sequencing depth (see *Diversity measures*). This was not necessary for models of MPD and aMPD as a function of depth because MPD and aMPD are mean values rather than additive indices. In the initial ETSP models, we included a “volume of water filtered” variable because water volumes did vary between samples (see Table 1). The water volume term was never significant ($P < 0.05$) and therefore left out of the final models.

RESULTS

We analyzed over 14 million sequence reads from 17 microbial community gDNA samples collected from the Eastern Tropical South Pacific (ETSP) OMZ (Table 1) in austral fall (June 2008), winter (August 2009), and

summer (January 2010). Samples spanned a depth gradient encompassing distinct physiochemical zones in the water column, including the photic zone, the oxycline, the suboxic core of the OMZ, and the oxygenated waters beneath the OMZ (Fig. 1). Notably, oxygen concentrations along the vertical sampling gradient span three orders of magnitude, peaking at $>200 \mu\text{mol/L}$ in the upper photic zone, declining steeply along the oxycline ($\sim 30\text{--}70$ m depth), and dropping well below $5 \mu\text{mol/L}$ throughout the core of the OMZ ($\sim 100\text{--}450$ m depth; Stewart et al. 2012).

Phylogenetic diversity

The distribution of broad taxonomic groups was similar to that found in other microbial diversity studies at the same station off the coast of Iquique (Appendix D: Fig. D1; see also Stevens and Ulloa 2008, Stewart et al. 2012). After statistically accounting for variation in data set size within each depth series, we found that taxonomic richness (TR) and phylogenetic diversity (PD) both decreased steeply with ocean depth ($P < 0.001$; Fig. 2, Table 2, Appendix D: Fig. D2). The decrease was observed in each sampling year. Within each year, TR decreased from the shallowest to the deepest sample by 27–32%, and PD decreased by 14–23%. A closer look at the 2009 depth series revealed that TR within over half of the broad taxonomic groups declined between the shallowest and deepest sample (Appendix B: Table B1). The largest declines in TR were observed in Flavobacteria, Alphaproteobacteria, Cyanobacteria, and Euryarchaeota.

Microbial communities in our study are defined as all microbes between the sizes of $0.22\text{--}1.6 \mu\text{m}$ (see *Methods*). Therefore, the majority of SSU genes we uncovered were of bacterial or archaeal origin. However, we also observed a small percentage of eukaryotic SSU (18S rRNA) genes (average across all data sets: 3.8%; Fig. D1), with the 2009 50-m data set containing the largest percentage (9.2%). Many of the eukaryotic SSU sequences matched picoeukaryotes that fit within the $0.22\text{--}1.6 \mu\text{m}$ size class (e.g., Alveolata, Stamenopiles), but a small number of others likely originated from fragmented tissue or the nuclei of lysed cells from larger organisms. We chose not to exclude the eukaryotic sequences from the analyses described in the main text, as we cannot confirm their origin. However, depth-specific trends in TR and PD did not change when all eukaryotic sequences were excluded from the analyses ($P < 0.002$; Table 2, Fig. D3).

Functional diversity

After statistically accounting for differences in data set size (see *Methods*), we found that functional richness as defined using the KEGG database (FR_K) or protein clusters (FR_{PC}), declined with ocean depth ($P < 0.03$, Fig. 3, Table 2, Appendix D: Fig. D4). This is consistent with measures of taxonomic and phylogenetic diversity. Within each depth series, FR_K decreased across the gradient

TABLE 2. Models predicting taxonomic richness (TR), functional richness (FR), phylogenetic diversity (PD), or phylogenetic structure (mean phylogenetic distance, MPD).

Site and models	<i>P</i>	Multiple <i>R</i> ²
Eastern Tropical South Pacific (ETSP)		
log(TR)~year + log(depth)	9.4×10^{-4}	0.96
log(PD)~year + log(depth)	1.5×10^{-4}	0.98
log(non-eukaryotic TR)~year + log(depth)	1.8×10^{-3}	0.96
log(non-eukaryotic PD)~year + log(depth)	6.0×10^{-4}	0.98
log(FR _K)~log(TR)	1.8×10^{-6}	0.79
log(FR _K)~log(PD)	9.4×10^{-9}	0.90
log(FR _K)~year + log(depth)	1.8×10^{-2}	0.90
log(FR _{PC})~year + log(depth)	1.1×10^{-2}	0.80
MPD~depth†	1.3×10^{-3}	0.56
aMPD~depth†	1.4×10^{-2}	0.38
non-eukaryotic MPD~depth	2.6×10^{-4}	0.65
non-eukaryotic aMPD~depth	4.4×10^{-3}	0.48
Bermuda Atlantic Time-Series (BATS)		
log(PD)~log(depth)	5.2×10^{-2}	0.90
MPD~log(depth)	4.8×10^{-3}	0.99
aMPD~log(depth)	6.8×10^{-2}	0.87

Notes: All logs are base 10. FR_K and FR_{PC} represent the number of unique KEGG ko gene categories and protein clusters, respectively, identified in each sample. The parameter aMPD stands for abundance-weighted MPD.
† Only included samples above 500 m.

FIG. 3. Functional richness (FR) across depths in the ETSP. Colored shapes (by sample year, as in Fig. 2) represent FR_K, the number of unique KEGG ko gene categories identified in each sample. Black bars outline 95% confidence intervals expected from randomly generated communities. Blue shading outlines regions of the water column with oxygen concentrations below 5 μmol/L. Rarefaction analysis was used to account for differences in sequencing depth between samples within the same depth series (year; see *Methods*).

FIG. 4. Phylogenetic dispersion across depths in the ETSP. Colored shapes (by sample year, as in Fig. 2) represent mean phylogenetic distance (MPD) between all pairs of taxa identified in each sample. Black bars outline 95% confidence intervals expected from randomly generated communities. Blue shading outlines regions of the water column with oxygen concentrations below 5 $\mu\text{mol/L}$. Rarefaction analysis was used to account for differences in sequencing depth between samples within the same depth series (year; see *Methods*).

between 8.9% and 24.8%. A closer look at the 2009 depth series, obtained by binning reads into KEGG pathways, revealed that the number of unique KEGG ko categories within KEGG pathways decreased from the shallowest to deepest samples in 263 out of 306 KEGG pathways identified at the ETSP. The largest decreases in FR_K were observed in pathways specific to eukaryotic organisms and pathways related to protein production and growth including the purine metabolism, pyrimidine metabolism (both largely comprised of polymerase genes) and cell cycle pathways (Appendix C: Table C1).

Phylogenetic and trait dispersion

Phylogenetic dispersion varied significantly across depths (Fig. 4). MPDs and aMPDs became increasingly clustered (smaller) from the mixed layer to the heart of the OMZ (all depths above 500 m; $P < 0.02$; Table 2, Fig. 4, Appendix D: Fig. D5). Some of the shallower samples showed significant phylogenetic overdispersion. All of the samples within the core of the OMZ (100–200 m) had significantly clustered MPDs and all but one sample had significantly clustered aMPDs. Increased

clustering with depth was also observed when all eukaryotic taxa were removed from the analysis ($P < 0.001$; Table 2, Appendix D: Figs. D5 and D6).

We also created a KEGG-based null model to test if FR is lower (trait clustering) or higher (trait overdispersion) than expected by chance given the total number of KEGG ko genes found in each sample. The 2009 and 2010 KEGG dispersion data were nearly identical to the 2009 and 2010 phylogenetic dispersion results (Figs. 3 and 4). In each depth series, only one point differed as to whether it was significantly clustered or within the bounds of the null expectation. In the 2008 depth series, both phylogenetic and trait dispersion results showed high levels of clustering in the OMZ, but in general were not as well coupled as in the 2009–2010 data.

Comparison to Bermuda Atlantic Time Series (BATS) and Hawaiian Oceanographic Time Series (HOT)

To determine the generality of the patterns observed in the OMZ, we compared the diversity patterns across the ETSP samples to diversity along depth gradients in

the oligotrophic, open ocean water column at the Hawaii Ocean Time Series (HOT) and Bermuda Atlantic Time Series (BATS) study sites. Nutrient concentrations (e.g., inorganic nitrogen) at these open ocean sites can be several orders of magnitude lower than those in and above the ETSP OMZ (Cavender-Bares et al. 2001, Stewart et al. 2012). Also, fluorescence peaks at a much deeper depth at BATS and HOT relative to the ETSP (Appendix E: Fig. E1), highlighting a much deeper photic zone at these sites. The oxygen concentrations at BATS and HOT stay relatively high (~ 100 – 250 $\mu\text{mol/L}$) from the surface to a depth of ~ 200 m, before decreasing gradually to a minimum at approximately 700–800 m depth (below the depth range sampled in this study).

We did not observe a consistent decrease in PD, TR, or FR with ocean depth as in the ETSP (Figs. E2 and E3). Phylogenetic diversity at BATS actually increases as a function of depth with marginal significance ($P < 0.052$; Table 2). We also did not observe a decrease in phylogenetic or trait dispersion with depth (Figs. E3 and E4). Rather, BATS appears to become more dispersed with depth ($P < 0.001$; Fig. E4, Table 2) and HOT shows no clear trends.

Changes in TR within most broad taxonomic groups followed similar trends at all sites (Appendix B: Tables B1, B3, and B5). At BATS, HOT, and the ETSP, large declines in TR between the shallowest to deepest sample were observed for Alphaproteobacteria, Flavobacteria, and Cyanobacteria clades. In addition, at all sampling sites, TR within several groups, including Chloroflexi, Deltaproteobacteria, and Deferribacteres, increased from the shallowest to deepest samples. However, at HOT and BATS, increases in richness within some taxonomic groups at deeper depths effectively canceled out richness decreases within other groups. This resulted in a very small net change in TR across the entire water column at the oligotrophic sites, whereas TR sharply declined with depth in the ETSP.

In contrast to the ETSP, where the vast majority of KEGG pathways declined in richness (FR_K) with depth, at both BATS and HOT the number of pathways with an increasing FR_K roughly equaled those with a decreasing FR_K (BATS 144 vs. 150; HOT: 124 vs. 160, respectively). As in the ETSP, at both oligotrophic sites FR_K within the photosynthesis pathway declined substantially with depth, while two-component signal transduction systems were among the pathways with the greatest increases with depth (Appendix C: Tables C1, C3, and C5).

DISCUSSION

The advent of high throughput metagenomics has opened up a new world to microbial ecologists, providing an unprecedented view of the functional and genetic diversity of microorganisms in natural environments. Here we applied metrics commonly used in macroorganism ecology to metagenomic data from oxygen minimum zones (OMZs) in the Eastern Tropical South Pacific Ocean

FIG. 5. EggNOG functional richness (FR_{EN}) across depths in the 2009 ETSP depth series. Red symbols represent the number of unique eggNOG orthologous groups identified in each sample. Black bars outline 95% confidence intervals expected from randomly generated communities. Blue shading outlines regions of the water column with oxygen concentrations below $5 \mu\text{mol/L}$. Rarefaction analysis was used to account for differences in sequencing depth between samples (see *Methods*).

(ETSP). We were able to find evidence of some of the broad forces shaping the ecology of microbial communities in these unique marine habitats.

Effects of database selection

Metagenomic studies are potentially hampered by the biases and limitations of the sequence databases used to infer the identity of environmental sequence fragments. As is true for most environmental metagenomic studies, the relative level of representation among organisms in sequence databases is a potential source of bias that is difficult to assess. Some microbial communities sampled along the depth gradients could be better represented in the ARB-SILVA, KEGG, and eggNOG databases than other communities. To minimize this potential bias in the SSU-based analyses (TR, PD, and MPD/aMPD), we used ARB-SILVA's non-redundant database, in which no two sequences are more than 99% identical, rather than using the full ARB-SILVA reference database to limit the influence of highly sampled environments. We did observe differences between data sets in the percentage of total non-duplicated reads that were assigned to KEGG ko categories, the largest difference being 16% between two 2009 data sets (Table 1). However, the percentage of reads assigned to KEGG ko categories did not consistently decrease with depth, suggesting that uneven genome representation in the database is not driving FR_K observed depth patterns.

Functional richness (FR) and dispersion trends differed when the eggNOG database was used to define traits rather than the KEGG database, although FR_{EN}

was still lowest in the OMZ samples (Fig. 5). The KEGG database consists of genes from 1293 genomes with known and homology-based functional assignments, while the eggNOG database was built by clustering all genes with known and unknown functions from 629 fully sequenced genomes (Kanehisa and Goto 2000, Muller et al. 2010). Because KEGG includes more genomes, this database likely encompassed a wider breadth of traits from different taxonomic groups. But because genes with unknown functions are included in the eggNOG database, searches against eggNOG produced more traits within each genome that was present in the database. Overall, these discrepancies illustrate the potential database-dependency of FR analyses. As our understanding of genes and gene functions grow in the future, databases will become more complete and database biases should decrease.

Database-independent clustering of protein-coding sequences can circumvent database biases (Stewart et al. 2011). However, database-independent methods have caveats as well. Notably, data sets composed of smaller sequences typically produce more clusters composed of fewer reads. This was evident in our results. The 2008 data sets contain fewer and on average shorter sequencing reads than the 2009 and 2010 data sets but produced more clusters (Table 1, Appendix D: Fig. D4). The sampling year term in the model accounted for the differences in average read lengths between years (adding an additional average read length term did not improve the model). The clustering analysis, supports our KEGG based results in suggesting that FR decreases with depth across the water column in the ETSP.

Diversity decreases with depth

As predicted, phylogenetic diversity (PD) correlated better with functional richness (FR_K) than it did with taxonomic richness (TR, R^2 of 0.90 vs. 0.79; Table 2). PD also had a slightly stronger correlation with ocean depth than TR in the ETSP (R^2 of 0.98 vs. 0.96). Other studies have observed that PD has a stronger correlation with environmental variables than TR (Chave et al. 2007). FR had a lower correlation with ocean depth, which may reflect real ecological differences between TR, PD, and FR, or more likely the incompleteness of the KEGG database. Regardless, the TR, PD, and FR results all demonstrate a distinct decrease in microbial diversity with depth across three seasons, suggesting that this pattern is relatively stable. Similarly, analysis of a seasonal OMZ off British Columbia (Saanich Inlet) showed SSU rRNA sequence richness peaking at the hypoxic transition zone before declining drastically within the OMZ (Zaikova et al. 2010). In contrast, after sequencing PCR amplified-SSU rRNA genes, Stevens and Ulloa (2008) found that bacterial OTU diversity (binned at 99–90% sequence similarity) was higher in the ETSP OMZ core than in the upper OMZ and surface waters in the summer of 2008. These divergent findings

may be due to methodological differences. For example, Stevens and Ulloa targeted a larger microbial size fraction (<3 μ m). Also, PCR can preferentially amplify some microbial groups over others (Kanagawa 2003).

Similar to our findings in the ETSP, Brown and colleagues (2009), using deep-coverage sequencing of the V9 hyper-variable region of the SSU gene and binning sequences at fine taxonomic resolutions (98% sequence similarity and greater), reported that bacterial OTU richness in the mixed surface layer at HOT station ALOHA was higher than in the mesopelagic OMZ at 800 m depth. In contrast, at broader taxonomic resolutions (95–80% sequence similarity), which are likely more comparable to TR thresholds used in our study, Brown et al. observed that archaeal and bacterial OTU richness was greater in the OMZ than in the surface layer. Additionally, Kembel et al. (2011) observed at HOT station ALOHA that OTU richness was higher in the mesopelagic OMZ than in surface water, at both 95% and 99% sequence similarity. Environmental differences may explain why the HOT OMZ appears species-rich relative to surface waters, whereas we observed lower richness within the ETSP OMZ compared to surface waters. Notably, dissolved oxygen concentrations within the HOT OMZ are orders of magnitude higher than those in the ETSP OMZ (tens of μ mol/L vs. tens of nmol/L; DeLong et al. 2006, Revsbech et al. 2009, Canfield et al. 2010). In addition, the OMZ layer at HOT, which begins much deeper in the water column (~700 m), is much thinner (<100 m) than at the ETSP and therefore more susceptible to edge effects, which may boost diversity (Harris 1988).

Our sampling at the oligotrophic HOT and BATS sites did not span the OMZ layer, which is located at ~770 m depth at that site. Though the vertical distances over which samples were collected were similar across all sites, environmental heterogeneity was much greater across the ETSP profile (Fig. 1 and Appendix E: Fig. E1). The upwelling of nutrient-rich water coupled with exceptionally high surface productivity above the ETSP OMZ likely create much steeper physical and biological gradients than in the oligotrophic open ocean. These dramatic gradients are likely the major drivers of the observed diversity patterns at the ETSP.

Factors shaping diversity

Though our data do not conclusively determine the mechanisms shaping community diversity, they do suggest some hypotheses. The steep decline in diversity within the OMZ may reflect a decrease in niche availability. High primary productivity generated in the photic zone through photoautotrophy creates a highly concentrated, potentially diverse pool of organic matter, which supports a very active heterotrophic microbial community in surface waters (Pantoja et al. 2004, Sempéré et al. 2008). Concurrent with a transition from surface waters into the OMZ, sinking rich organic carbon is degraded and becomes increasingly refractory,

and possibly more homogeneous (Pantoja et al. 2004, 2009, Brown et al. 2009). Additionally, the transition into the OMZ generally co-occurs with reduced light levels. We hypothesize that this reduction and potential homogenization of labile organic matter and loss of solar radiation may reduce the number of niches available to both heterotrophic microorganisms and microorganisms that interact with light. This hypothetical rise in niche scarcity would drive the observed decline in TR, PD, and trait richness with depth.

The taxonomic (phylum-level) and KEGG pathway data suggest that niches associated with light and labile organic matter utilization are more prevalent in surface water than in the OMZ at the ETSP. Notably, although ABC transporter abundance was higher in deep samples at the ETSP, transporter richness was highest in shallow samples, suggesting greater variation in the types of substrates used near the surface (Appendix C: Table C1 and C2). Also, the richness and abundance of Cyanobacteria and Flavobacteria and of genes within the photosynthesis pathway substantially declined between 50 m and 70 m in the 2009 ETSP depth series (Appendix B: Table B1 and B2, Appendix C: Table C1). Cyanobacteria are photoautotrophs and are anticipated to be at high abundance in surface layers. In contrast, Flavobacteria are primarily heterotrophs, though some strains may use the retinal-containing proteorhodopsin as a light-driven proton pump to support energy production. Furthermore, Flavobacteria tend to co-occur with phytoplankton and are believed to specialize in degrading proteins and other polymers (González et al. 2008, Gómez-Pereira et al. 2010), potentially explaining their higher abundance near or in the photic depths.

Similar declines in photosynthesis pathway richness and Cyanobacteria and Flavobacteria richness also occurred at BATS and HOT (Tables B3, B5, C3, and C5). However, the declines occurred at deeper depths compared to the ETSP (between 100 and 500 m at oligotrophic sites vs. 50 and 70 m at the ETSP), consistent with the much deeper photic zone in these oligotrophic waters (approximately 120 m depth at BATS/HOT vs. 50 m at ETSP). Also in contrast to the ETSP, ABC transporter richness at BATS and HOT stayed consistent across samples above 500 m, suggesting that substrate diversity across these sites may not have changed. Overall, these trends suggest that depths of 110 m and above are more homogenous in the oligotrophic sites than at the ETSP.

The reduction in TR, PD, and FR in the OMZ at the ETSP suggests that the number of functional niches may be reduced in the OMZ by the requirement for anaerobic or microaerophilic lifestyles. Alternatively, it has been hypothesized that more niches may be available in OMZ waters because suboxic conditions allow microbes to use multiple electron acceptors for respiration, while in surface waters, oxygen is preferentially used and not limiting (Stevens and Ulloa 2008).

Nitrate, manganese (IV), iodate, and sulfate, in addition to nitrite (via the anammox reaction and denitrification) have all been reported as being utilized during respiration in OMZs (Canfield et al. 2010, Lam and Kuypers 2011). We observed an increase in the relative abundance of sequences matching dissimilatory nitrogen pathways, as well as those mediating sulfur and methane metabolism, in deeper samples in the ETSP (Table C2). Also, the 2009 ETSP samples collected within the OMZ had lower TR and PD values than the aphotic HOT and BATS 500-m samples (Appendix E: Fig. E2). It is possible our results reflect a net decline in the number of niches in the OMZ relative to oxic water that outweigh an increased number of niches in the OMZ due to electron acceptor partitioning.

We speculate that, overall, lower energy availability in the core of the OMZ (due to lower redox potentials, less readily available organic material, and lack of incoming solar radiation) reduces the total number of possible niches compared to shallower depths. This hypothesis is consistent with general species-energy theories, which predict that richness is limited by the amount of available energy in a system (Wright 1983, Currie 1991). Solar energy (light) also creates an energy gradient at the oligotrophic BATS and HOT sites, though this gradient is distributed over a much greater depth range (Fig. E1). Relative to the ETSP, lower nutrient concentrations in the water column at these sites restrict primary production, limiting the amount of energy made accessible to other organisms through photosynthesis (Corno et al. 2008, Stewart et al. 2012). Consistent with this hypothesis, even at shallow depths in the oxygenated photic zone, TR and PD were higher at the ETSP than at the oligotrophic sites (Fig. E2). This pattern is at least partially due to the presence in the ETSP of Crenarchaea and Chloroflexi at shallow depths.

At the ETSP, we observed a monotonic decrease in diversity across depths, rather than a hump-shaped pattern. Both patterns are commonly observed across energy gradients in nature (Mittelbach et al. 2001). Our study may have captured the “true” underlying diversity pattern in our system. Alternatively, the scale at which our study was conducted may explain our results. For example, the sampling extent (overall area encompassed by the study) may only span the decreasing region of a larger hump-shaped pattern. Also, theoretical and empirical work has demonstrated that smaller sample grains (area of each sample unit) yield hump-shaped patterns while increasing sampling grain transforms the pattern into a monotonic decrease in diversity with decreasing energy (Chase and Leibold 2002). Considering the small size of microbes compared to our sample volumes, we are likely sampling at large grains.

Drivers of clustering and overdispersion

Our data show that microbial communities are more phylogenetically and functionally dispersed in the oxic zone and oxycline, whereas they were more clustered in

- Cavender-Bares, J., A. Keen, and B. Miles. 2006. Phylogenetic structure of Floridian plant communities depends on taxonomic and spatial scale. *Ecology* 87:109–122.
- Cavender-Bares, K. K., D. M. Karl, and S. W. Chisholm. 2001. Nutrient gradients in the western North Atlantic Ocean: relationship to microbial community structure and comparison to patterns in the Pacific Ocean. *Deep-Sea Research* 48:2373–2395.
- Chase, J. M., and M. A. Leibold. 2002. Spatial scale dictates the productivity-biodiversity relationship. *Nature* 416:427–430.
- Chave, J., G. Chust, and C. Thebaud. 2007. The importance of phylogenetic structure in biodiversity studies. Page 151–167 in D. Storch, P. Marquet, and J. Brown, editors. *Scaling biodiversity*. Cambridge University Press, Cambridge, UK.
- Coleman, M. L., and S. W. Chisholm. 2010. Ecosystem-specific selection pressures revealed through comparative population genomics. *Proceedings of the National Academy of Sciences USA* 107:18634–18639.
- Corno, G., R. M. Letelier, M. R. Abbott, and D. M. Karl. 2008. Temporal and vertical variability in photosynthesis in the North Pacific Subtropical Gyre. *Limnology and Oceanography* 53:1252–1265.
- Cornwell, W. K., D. W. Schilck, and D. D. Ackerly. 2006. A trait-based test for habitat filtering: convex hull volume. *Ecological Monographs* 76:1465–1471.
- Currie, D. J. 1991. Energy and large-scale patterns of animal- and plant-species richness. *American Naturalist* 137:27–49.
- DeLong, E. F., C. M. Preston, T. Mincer, V. Rich, S. J. Hallam, N. U. Frigaard, A. Martinez, M. B. Sullivan, R. Edwards, B. R. Brito, S. W. Chisholm, and D. M. Karl. 2006. Community genomics among stratified microbial assemblages in the ocean's interior. *Science* 311:496–503.
- Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. *Biological Conservation* 61:1–10.
- Faith, D. P. 1994. Phylogenetic pattern and the quantification of organismal biodiversity. *Philosophical Transactions: Biological Science* 345:45–58.
- Fierer, N., C. M. McCain, P. Meir, M. Zimmermann, J. M. Rapp, M. R. Silman, and R. Knight. 2011. Microbes do not follow the elevational diversity patterns of plants and animals. *Ecology* 92:797–804.
- 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. *Proceedings of the National Academy of Sciences USA* 105:3805–3810.
- Fuhrman, J. A., 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. *Proceedings of the National Academy of Sciences USA* 105:7774–7778.
- Galán, A., V. Molina, B. Thamdrup, D. Woebken, G. Lavik, M. M. M. Kuypers, and O. Ulloa. 2009. Anammox bacteria and the anaerobic oxidation of ammonium in the oxygen minimum zone off northern Chile. *Deep Sea Research Part II: Topical Studies in Oceanography* 56:1021–1031.
- Gómez-Pereira, P. R., B. M. Fuchs, C. Alonso, M. J. Oliver, J. E. E. van Beusekom, and R. Amann. 2010. Distinct flavobacterial communities in contrasting water masses of the North Atlantic Ocean. *ISME Journal* 4:472–487.
- González, J. M., et al. 2008. Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp. MED152 (Flavobacteria). *Proceedings of the National Academy of Sciences USA* 105:8724–8729.
- Hamady, M., C. Lozupone, and R. Knight. 2010. Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME Journal* 4:17–27.
- Harris, L. D. 1988. Edge effects and conservation of biotic diversity. *Conservation Biology* 2:330–332.
- Horner-Devine, M. C., and B. J. M. Hohannan. 2006. Phylogenetic clustering and overdispersion in bacterial communities. *Ecology* 86:S100–S108.
- Kanagawa, T. 2003. Bias and artifacts in multitemplate polymerase chain reactions (PCR). *Journal of Bioscience and Bioengineering* 96:317–323.
- Kanehisa, M., and S. Goto. 2000. Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* 28:27–30.
- Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26:1463–1464.
- Kembel, S. W., J. A. Eisen, K. S. Pollard, and J. L. Green. 2011. The phylogenetic diversity of metagenomes. *PLoS ONE* 6: e23214.
- Kembel, S. W., and S. P. Hubbell. 2006. The phylogenetic structure of a neotropical forest tree community. *Ecology* 87:86–99.
- Konstantinidis, K. T., J. Bragg, D. M. Karl, and E. F. DeLong. 2009. Comparative metagenomic analysis of a microbial community residing at a depth of 4,000 meters at station ALOHA in the North Pacific subtropical gyre. *Applied and Environmental Microbiology* 75:5345–5355.
- Kraft, N. J. B., and D. D. Ackerly. 2009. Response to Comment on “Functional traits and niche-based tree community assembly in an Amazonian forest.” *Science* 324:1015.
- Kraft, N. J. B., R. Valencia, and D. D. Ackerly. 2008. Functional traits and niche-based tree community assembly in an Amazonian forest. *Science* 322:580–582.
- Lake, J. K., and A. Ostling. 2009. Comment on “Functional traits and niche-based tree community assembly in an Amazonian forest.” *Science* 324:1015.
- Lam, P., and M. M. M. Kuypers. 2011. Microbial nitrogen cycling processes in oxygen minimum zones. *Annual Review of Marine Science* 3:317–345.
- Li, W., and A. Godzik. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–1659.
- Lomolino, M. V., B. R. Riddle, and J. H. Brown. 2006. *Biogeography*. Third edition. Sinauer, Sunderland, Massachusetts, USA.
- Lozupone, C. A., and R. Knight. 2007. Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences USA* 104:11436–11440.
- Ludwig, W., et al. 2004. ARB: a software environment for sequence data. *Nucleic Acids Research* 32:1363–1371.
- May, R. M. 1990. Taxonomy as destiny. *Nature* 347:129–130.
- Mayfield, M. M., and J. M. Levine. 2010. Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology Letters* 13:1085–1093.
- Mittelbach, G. G., C. F. Steiner, S. M. Scheiner, K. L. Gross, H. L. Reynolds, R. B. Waide, M. R. Willig, S. I. Dodson, and L. Gough. 2001. What is the observed relationship between species richness and productivity? *Ecology* 82:2381–2396.
- Morlon, H., D. W. Schilck, J. A. Bryant, P. A. Marquet, A. G. Rebelo, C. Tauss, B. J. M. Bohannan, and J. L. Green. 2011. Spatial patterns of phylogenetic diversity. *Ecology Letters* 14:141–149.
- Muller, J., D. Szklarczyk, P. Julien, I. Letunic, A. Roth, M. Kuhn, S. Powell, C. von Mering, T. Doerks, L. J. Jensen, and P. Bork. 2010. eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic Acids Research* 38:D190–195.
- Noguchi, H., J. Park, and T. Takagi. 2006. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Research* 34:5623–5630.
- Pantoja, S., P. Rossel, R. Castro, L. A. Cuevas, G. Daneri, and C. Córdova. 2009. Microbial degradation rates of small peptides and amino acids in the oxygen minimum zone of

- Chilean coastal waters. *Deep Sea Research Part II: Topical Studies in Oceanography* 56:1055–1062.
- Pantoja, S., J. Sepúlveda, and H. González. 2004. Decomposition of sinking proteinaceous material during fall in the oxygen minimum zone off northern Chile. *Deep Sea Research Part I: Oceanographic Research Papers* 51:55–70.
- Petchey, O. L., and K. J. Gaston. 2002. Functional diversity (FD), species richness and community composition. *Ecology Letters* 5:402–411.
- Pruesse, E., C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies, and F. O. Glöckner. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research* 35:7188–7196.
- R Development Core Team. 2011. A language and environment for statistical computing. R Project for Statistical Computing, Vienna, Austria. www.r-project.org
- Raes, J., I. Letunic, T. Yamada, L. J. Jensen, and P. Bork. 2011. Toward molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data. *Molecular Systems Biology* 7:1–9.
- Revsbech, N. P., L. H. Larsen, J. Gundersen, T. Dalsgaard, O. Ulloa, and B. Thamdrup. 2009. Determination of ultra-low oxygen concentrations in oxygen minimum zones by the STOX sensor. *Limnology and Oceanography: Methods* 7:371–381.
- Rodrigues, A. S. L., and K. J. Gaston. 2002. Maximising phylogenetic diversity in the selection of networks of conservation areas. *Biological Conservation* 105:103–111.
- Schleuter, D., M. Daufresne, F. Massol, and C. Argillier. 2010. A user's guide to functional diversity indices. *Ecological Monographs* 80:469–484.
- Sempéré, R., M. Tedetti, C. Panagiotopoulos, B. Charrière, and F. Van Wambeke. 2008. Distribution and bacterial availability of dissolved neutral sugars in the South East Pacific. *Biogeosciences* 5:1165–1173.
- Smith, K. F., and J. H. Brown. 2002. Patterns of diversity, depth range and body size among pelagic fishes along a gradient of depth. *Global Ecology and Biogeography* 11:313–322.
- Smith, V. H. 2007. Microbial diversity-productivity relationships in aquatic ecosystems. *FEMS Microbiology Ecology* 62:181–186.
- Stevens, H., and O. Ulloa. 2008. Bacterial diversity in the oxygen minimum zone of the eastern tropical South Pacific. *Environmental Microbiology* 10:1244–1259.
- Stewart, F. J., E. A. Ottesen, and E. F. DeLong. 2010. Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *ISME Journal* 4:896–907.
- Stewart, F. J., A. K. Sharma, J. A. Bryant, J. M. Eppley, and E. F. DeLong. 2011. Community transcriptomics reveals universal patterns of protein sequence conservation in microbial communities. *Genome Biology* 12:R26.
- Stewart, F. J., O. Ulloa, and E. F. DeLong. 2012. Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environmental Microbiology* 14:23–40.
- Stramma, L., G. C. Johnson, J. Sprintall, and V. Mohrholz. 2008. Expanding oxygen-minimum zones in the tropical oceans. *Science* 320:655–658.
- Swenson, N. G. 2011. The role of evolutionary processes in producing biodiversity patterns, and the interrelationships between taxonomic, functional and phylogenetic biodiversity. *American Journal of Botany* 98:472–480.
- Swenson, N. G., B. J. Enquist, J. Pither, J. Thompson, and J. Zimmerman. 2006. The problem and promise of scale dependency in community phylogenetics. *Ecology* 87:2418–2424.
- Swenson, N. G., B. J. Enquist, J. Thompson, and J. Zimmerman. 2007. The influence of spatial and size scale on phylogenetic relatedness in tropical forest communities. *Ecology* 88:1770–1780.
- Tofts, R., and J. Silvertown. 2000. A phylogenetic approach to community assembly from a local species pool. *Proceedings of the Royal Society B* 267:363–369.
- Tringe, S. G., and E. M. Rubin. 2005. Metagenomics: DNA sequencing of environmental samples. *Nature Review Genetics* 6:805–815.
- Webb, C. O., D. D. Ackerly, M. A. McPeck, and M. J. Donoghue. 2002. Phylogenies and community ecology. *Annual Review of Ecology and Systematics* 33:475–505.
- Weiher, E., G. D. P. Clarke, and P. A. Keddy. 1998. Community assembly rules, morphological dispersion, and the coexistence of plant species. *Oikos* 81:309–322.
- Weiher, E., and P. A. Keddy. 1995. Assembly rules, null models, and trait dispersion: new questions from old patterns. *Oikos* 74:159–164.
- Wright, D. H. 1983. Species-energy theory: an extension of species-area theory. *Oikos* 41:496–506.
- Zaikova, E., D. A. Walsh, C. P. Stilwell, W. W. Mohn, P. D. Tortell, and S. J. Hallam. 2010. Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. *Environmental Microbiology* 12:172–191.
- Zhou, J., S. Kang, C. W. Schadt, and C. T. Garten, Jr. 2008. Spatial scaling of functional gene diversity across various microbial taxa. *Proceedings of the National Academy of Sciences USA* 105:7768–7773.

SUPPLEMENTAL MATERIAL

Appendix A

Bermuda Atlantic Time-Series (BATS) and Hawaii Ocean Time-Series (HOT) sample and sequence library characteristics (*Ecological Archives* E093-145-A1).

Appendix B

Abundance and richness within broad taxonomic groups (*Ecological Archives* E093-145-A2).

Appendix C

Abundance and richness within KEGG Pathways (*Ecological Archives* E093-145-A3).

Appendix D

Additional Eastern Tropical South Pacific (ETSP) diversity figures (*Ecological Archives* E093-145-A4).

Appendix E

Figures related to BATS and HOT sites (*Ecological Archives* E093-145-A5).

Supplement

R code used to normalize diversity measures (phylogenetic diversity [PD] and mean phylogenetic distance [MPD]) between samples with different sequencing depths and R code used to generate functional richness (FR) null models (*Ecological Archives* E093-145-S1).