

## ORIGINAL ARTICLE

# Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone

Sangita Ganesh<sup>1</sup>, Darren J Parris<sup>1</sup>, Edward F DeLong<sup>2,3</sup> and Frank J Stewart<sup>1</sup>

<sup>1</sup>*School of Biology, Georgia Institute of Technology, Atlanta, GA, USA;* <sup>2</sup>*Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Parsons Laboratory 48, Cambridge, MA, USA and* <sup>3</sup>*Center for Microbial Ecology: Research and Education, Honolulu, Hawaii, USA*

**Marine oxygen minimum zones (OMZs) support diverse microbial communities with roles in major elemental cycles. It is unclear how the taxonomic composition and metabolism of OMZ microorganisms vary between particle-associated and free-living size fractions. We used amplicon (16S rRNA gene) and shotgun metagenome sequencing to compare microbial communities from large (>1.6 μm) and small (0.2–1.6 μm) filter size fractions along a depth gradient in the OMZ off Chile. Despite steep vertical redox gradients, size fraction was a significantly stronger predictor of community composition compared to depth. Phylogenetic diversity showed contrasting patterns, decreasing towards the anoxic OMZ core in the small size fraction, but exhibiting maximal values at these depths within the larger size fraction. Fraction-specific distributions were evident for key OMZ taxa, including anammox planctomycetes, whose coding sequences were enriched up to threefold in the 0.2–1.6 μm community. Functional gene composition also differed between fractions, with the >1.6 μm community significantly enriched in genes mediating social interactions, including motility, adhesion, cell-to-cell transfer, antibiotic resistance and mobile element activity. Prokaryotic transposase genes were three to six fold more abundant in this fraction, comprising up to 2% of protein-coding sequences, suggesting that particle surfaces may act as hotbeds for transposition-based genome changes in marine microbes. Genes for nitric and nitrous oxide reduction were also more abundant (three to seven fold) in the larger size fraction, suggesting microniche partitioning of key denitrification steps. These results highlight an important role for surface attachment in shaping community metabolic potential and genome content in OMZ microorganisms.**

*The ISME Journal* (2014) 8, 187–211; doi:10.1038/ismej.2013.144; published online 12 September 2013

**Subject Category:** Microbial ecology and functional diversity of natural habitats

**Keywords:** Archaea; bacteria; microbial diversity; OMZ; oxycline; particle

## Introduction

Marine oxygen minimum zones (OMZs) contain diverse communities of Bacteria and Archaea whose metabolisms control key steps in marine biogeochemical cycling. Metagenome and single-gene surveys have identified marked transitions in microbial community composition and metabolism from oxygenated surface waters to the suboxic OMZ core (Stevens and Ulloa, 2008; Zaikova *et al.*, 2010; Bryant *et al.*, 2012; Stewart *et al.*, 2012b). These shifts have been linked to meter-scale vertical gradients in dissolved oxygen and organic and inorganic energy substrate availability. However, heterogeneity in the marine water column also potentially exists in the form of microscale chemical

gradients, microbial taxa and microbial processes associated with particles, including aggregates of decaying organic matter, as well as live phytoplankton or zooplankton cells (Karl *et al.*, 1984; DeLong *et al.*, 1993; Fenchel, 2002; Stocker, 2012). These particles support complex surface-attached microbial communities whose composition and life history strategies differ substantially from those of free-living microbes. Although taxonomic surveys have compared free-living and particle-attached communities in a variety of marine ecosystems, the differences in functional gene composition that distinguish free-living from surface-attached life histories have been explored only sparingly. For OMZs in particular, the microscale partitioning of microbial communities and metabolisms has not been explored, despite a potentially significant role for particle-associated microhabitats in these zones (Whitmire *et al.*, 2009).

OMZs occur where the aerobic respiration of organic matter combines with water column stabilization to form a persistent, low-oxygen layer. The largest and most oxygen-depleted OMZs are

Correspondence: F Stewart, School of Biology, Georgia Institute of Technology, Ford ES and T Building, Room 1242, 311 Ferst Drive, Atlanta, GA 30332, USA.

E-mail: frank.stewart@biology.gatech.edu

Received 19 April 2013; revised 21 July 2013; accepted 22 July 2013; published online 12 September 2013

found in regions of nutrient upwelling, as in the Eastern Tropical South Pacific off the coasts of Chile and Peru (ETSP OMZ; Karstensen *et al.*, 2008; Ulloa and Pantoja, 2009; Ulloa *et al.*, 2012). In the ETSP, dissolved oxygen declines from near saturation levels ( $>250\ \mu\text{M}$ ) at the surface to below the level of detection ( $<10\ \text{nm}$ ) at the OMZ core ( $\sim 200\text{--}300\ \text{m}$ ; Thamdrup *et al.*, 2012; Ulloa *et al.*, 2012). This steep drawdown drives changes in the water column microbial community. Notably, communities along the oxycline are phylogenetically and metabolically diverse relative to other depths (Bryant *et al.*, 2012), containing both microaerophilic assemblages, which include ammonia- and nitrite-oxidizing members, as well as microbes capable of anaerobic metabolism (Stewart *et al.*, 2012b). In contrast, community metabolism at the anoxic OMZ core is dominated by anaerobic autotrophic and heterotrophic processes that primarily utilize oxidized nitrogen compounds as terminal oxidants (Ulloa and Pantoja, 2009; Ulloa *et al.*, 2012; Wright *et al.*, 2012). Up to half of oceanic nitrogen loss occurs in OMZs through the anaerobic processes of denitrification and anaerobic ammonium oxidation (anammox; Thamdrup *et al.*, 2006; Lam *et al.*, 2009; Zehr and Kudela, 2011). Recent studies also have identified an important role in OMZs for chemoautotrophic bacteria that oxidize reduced sulfur compounds with nitrate, as well as for sulfate-reducing heterotrophs (Stevens and Ulloa, 2008; Lavik *et al.*, 2009; Walsh *et al.*, 2009; Canfield *et al.*, 2010; Stewart *et al.*, 2012b). It is unclear, however, whether these key biochemical processes are differentially partitioned between free-living versus particle-associated microbial communities in suboxic water columns.

In studies of non-OMZ systems, particle-association has been shown to be a significant component of microbial distributions, community composition and activity (DeLong *et al.*, 1993; Hollibaugh *et al.*, 2000; LaMontagne and Holden, 2003; Eloe *et al.*, 2010). Most analyses impose a prefiltering step to separate microbial communities according to particle size, with typical prefilter and microfilter (collection filter) pore sizes of  $0.8\text{--}30$  and  $0.2\ \mu\text{m}$ , respectively. Although the microfilter fraction (typically cells between  $0.2$  and  $1.6\text{--}3\ \mu\text{m}$ ) is presumably dominated by non-surface-attached (free-living) prokaryotes (Cho and Azam, 1988), the prefilter may retain a range of organisms, including larger free-living prokaryotes (for example, filamentous cyanobacteria), microbial eukaryotes and zooplankton, but presumably also captures particulate aggregates composed of organic debris and surface-attached microbial cells (marine snow). Compared with the open water column, these particles constitute unique microhabitats that are relatively enriched in nutrients and contain potentially steep microscale (microns) gradients in pH and redox substrates, including organic carbon (Alldredge and Cohen, 1987; Alldredge and Silver, 1988; Stocker,

2012). Relative to free-living bacteria, particle-associated bacteria are typically larger (Caron *et al.*, 1982; Lapoussiere *et al.*, 2011), occur at higher local densities (Simon *et al.*, 2002) and exhibit higher rates of substrate acquisition, enzymatic activity, protein production and respiration (Kirchman and Mitchell, 1982; Karner and Herndl, 1992; Grossart *et al.*, 2003, 2007). Not surprisingly, free-living and particle-associated communities can differ significantly in composition at multiple levels of phylogenetic resolution (DeLong *et al.*, 1993; Ploug *et al.*, 1999; Grossart *et al.*, 2006; Hunt *et al.*, 2008; Kellogg and Deming, 2009). Notably, phytoplankton-derived particle communities tend to be relatively enriched in members of the Bacteroidetes (notably, Cytophaga and Flavobacteria), Planctomycetes and Deltaproteobacteria (DeLong *et al.*, 1993; Crump *et al.*, 1999; Smith *et al.*, 2013), whereas planktonic communities are enriched in taxa adapted for oligotrophic or autotrophic free-living lifestyles (for example, *Pelagibacter*, picocyanobacteria; Lauro *et al.*, 2009).

Analyses of prokaryotic microbial diversity in OMZs have focused primarily on the microfilter size fraction ( $\sim 0.2\text{--}3\ \mu\text{m}$ ; Stevens and Ulloa, 2008; Canfield *et al.*, 2010; Stewart *et al.*, 2012b), with larger or particle-associated cells being excluded. However, certain taxonomic groups with important functional roles in OMZ elemental cycling may differ in abundance and activity between free-living versus particle-associated communities. Recent diversity surveys based on 16S rRNA gene sequences confirm distinct taxon compositions among size-fractionated samples ( $30\ \mu\text{m}$  cutoff) from the suboxic water column of the Black Sea (Fuchsman *et al.*, 2011, 2012). Notably, members of the Alphaproteobacteria (for example, SAR11-like sequences) and the marine anammox planctomycete *Candidatus* Scalindua were significantly enriched in the smaller size fraction, while key anaerobic lineages, including members of the sulfate-reducing Deltaproteobacteria were more abundant in the larger, particle-associated fraction ( $>30\ \mu\text{m}$ ). A recent study that demonstrated sulfate-reduction activity in the ETSP OMZ also described 16S rRNA and sulfur metabolism gene (for example, *aprA*, encoding adenosine-5'-phosphosulfate (APS) reductase) sequences matching sulfate-reducing clades of the Deltaproteobacteria, although these sequences were at relatively low abundance (Canfield *et al.*, 2010). However, metagenomic sequencing in this study focused only on the free-living community ( $0.2\text{--}1.6\ \mu\text{m}$  cell size), and it remains unclear to what extent OMZ processes such as sulfate reduction may instead be mediated by particle communities.

Shotgun metagenomics provides a snapshot of the metabolic functions available in a mixed community, while simultaneously allowing for taxonomic identification of community members (DeLong *et al.*, 2006). Surprisingly, metagenomics has been used only sparingly to identify differences between

marine microbial communities separated by size fraction. Seminal metagenomic studies by Venter *et al.* (2004) and Rusch *et al.* (2007) sampled different filter size fractions from diverse ocean sites visited on the *Global Ocean Survey* expeditions. This work identified important linkages between gene content and environmental variables, as well as patterns of metagenomic sequence similarity representing distinct community types. However, a direct comparison of protein-coding gene content between fractions was not presented, and high-resolution taxonomic surveys from these analyses are not available. Although such comparisons identified functional differences in size-fractionated picoeukaryote (Not *et al.*, 2009) and viral (Williamson *et al.*, 2012) communities, to the best of our knowledge only two other studies (Allen *et al.*, 2012; Smith *et al.*, 2013), focusing on temperate sites in the North Pacific, have directly compared microbial (Bacteria and Archaea) functional gene content between filter size fractions. Here, we use metagenome and 16S rRNA gene amplicon sequencing to compare microbial communities from two size classes in the ETSP OMZ. The results identify surface attachment as a major driver of community composition and genome diversity, and highlight the potential for key physiological processes to be partitioned between free-living and particle-associated OMZ microbes.

## Materials and methods

### Sample collection

Microbial community samples were collected from the ETSP OMZ as part of the Center for Microbial Ecology: Research and Education (C-MORE) BiG RAPA (Biogeochemical Gradients: Role in Arranging Planktonic Assemblages) cruise aboard the *R/V Melville* (18 November–14 December 2010). Seawater was sampled from 7 depths (5, 32, 70, 110, 200, 320 and 1000 m) at Station 1 (20° 04.999 S, 70° 48.001 W) off the coast of Iquique, Chile on November 19th (5 m), 20th (32 m), 21st (70, 110, 200 and 320 m) and 23rd (1000 m). Collections were made using Niskin bottles deployed on a rosette containing a Conductivity-Temperature-Depth profiler (Sea-Bird SBE 911plus) equipped with a dissolved Oxygen Sensor, fluorometer and transmissometer (see Supplementary Figure S1). Microbial biomass was collected by sequential in-line filtration of seawater samples (10 l) through a prefilter (GF/A, 1.6 µm pore-size, 47 mm diameter, Whatman, GE Healthcare, Piscataway, NJ, USA) and a primary collection filter (Sterivex, 0.22 µm pore-size, Millipore, Billerica, MA, USA) using a peristaltic pump. Prefilters were transferred to microcentrifuge tubes containing lysis buffer (~1.8 ml; 50 mM Tris-HCl, 40 mM EDTA and 0.73 M sucrose). Sterivex filters were filled with lysis buffer (~1.8 ml), and capped at both ends. Both filter types were stored at -80 °C until DNA extraction.

### DNA extraction

Genomic DNA was extracted from prefilter disc and Sterivex cartridge filters using a phenol:chloroform protocol modified from Frias-Lopez *et al.* (2008). Briefly, cells were lysed by adding lysozyme (2 mg in 40 µl of lysis buffer per filter) directly to the prefilter-containing microcentrifuge tube or to the Sterivex cartridge, sealing the caps/ends and incubating for 45 min at 37 °C. Proteinase K (1 mg in 100 µl lysis buffer, with 100 µl 20% SDS) was added, and the tubes and cartridges were resealed and further incubated for 2 h at 55 °C. Lysate was removed from each filter, and nucleic acids were extracted once with phenol:chloroform:isoamyl alcohol (25:24:1) and once with chloroform:isoamyl alcohol (24:1). The purified aqueous phase was concentrated by spin dialysis using Amicon Ultra-4w/100 kDa MWCO centrifugal filters (Millipore). Aliquots of purified DNA from each depth and filter size fraction were used for PCR. Additional aliquots (5 µg) were used to prepare libraries for shotgun pyrosequencing of microbial metagenomes.

### 16S rRNA gene PCR

Pyrosequencing of PCR amplicons encompassing hypervariable regions of the 16S rRNA gene was used to assess bacterial community composition in both filter types from all water column depths. Archaeal 16S rRNA gene diversity was not evaluated via amplicon analysis. PCR amplicons were synthesized according to established protocols. Briefly, a ~480-bp fragment of the bacterial 16S rRNA gene was amplified using barcoded universal primers targeting the V1–V3 region, as described in the protocol established for the Human Microbiome Project by the Broad Institute Jumpstart Consortium Human Microbiome Project Data Generation Working Group, 2012). Briefly, thermal cycling conditions were: initial denaturation at 95 °C (2 min), followed by 30 cycles of denaturation at 95 °C (20 s), primer annealing at 50 °C (30 s) and primer extension at 72 °C (5 min). 16S rRNA gene PCR was replicated using a second set of barcoded primers to assess potential variation introduced during PCR. Amplicons were analyzed by agarose gel electrophoresis to verify size, purified using the QIAquick PCR Clean-Up Kit, pooled (~200 ng per sample), and used as template for multiplex amplicon pyrosequencing.

### Pyrosequencing

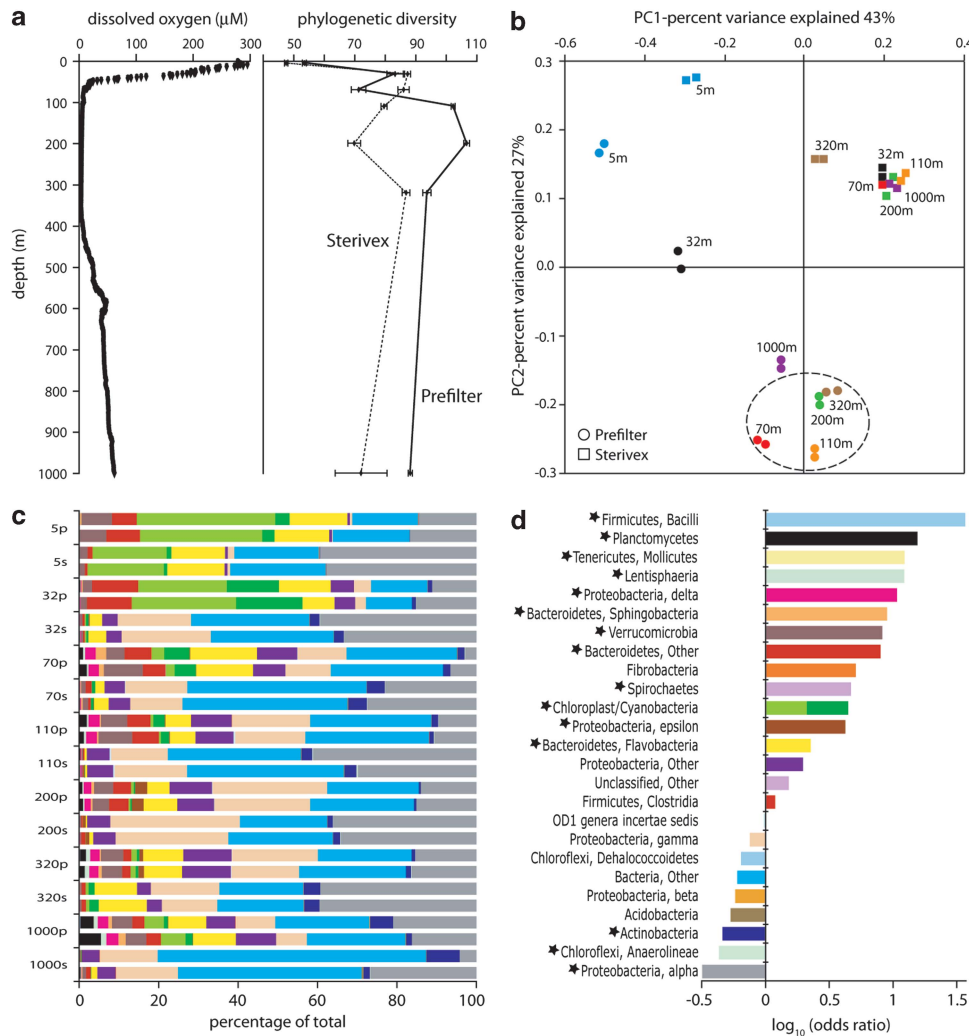
Shotgun pyrosequencing (454 Life Sciences, Roche Applied Science, Branford, CT, USA) was used to characterize the community DNA (metagenome) from two microbial size fractions (prefilter and Sterivex) from four depths (70, 110, 200 and 1000 m), as well as the multiplexed 16S amplicon samples (seven depths, two filter types per depth, two PCR replicates per filter). DNA templates were used to prepare single-stranded libraries for emulsion PCR using established protocols (454 Life

Sciences, Roche Applied Science). Each metagenome sample was sequenced with a half-plate run on a Roche Genome Sequencer FLX Instrument using Titanium chemistry. The multiplexed amplicon sample was sequenced using a single full-plate run.

#### Sequence analysis—16S rRNA gene amplicons

Amplicons were analyzed using the software pipeline QIIME (Caporaso *et al.*, 2010), according to standard protocols. Briefly, barcoded 16S data sets were de-multiplexed and filtered to remove

low quality sequences using default parameters (minimum quality score = 25, minimum sequence length = 200, no ambiguous bases allowed). De-multiplexed sequences were clustered into operational taxonomic units (OTUs) at 97% sequence similarity, with taxonomy assigned to representative OTUs from each cluster using the Ribosomal Database Project classifier in QIIME, trained on the Greengenes database. OTU counts were rarefied (10 iterations) and alpha diversity was quantified at a uniform sequencing depth across samples using the phylogenetic diversity (PD) metric as described by Faith (1992); (Figure 1a).



**Figure 1** OMZ bacterial community diversity revealed by 16S rRNA gene pyrosequencing. **(a)** PD as a function of water column depth and dissolved oxygen concentration (left). Data points are mean values based on rarefaction of OTU counts at a standardized sequence count ( $n = 4996$ ) per sample, with bars indicating 95% confidence intervals for the rarefied measurements. Data from both PCR duplicates are combined for averaging. **(b)** Principle component analysis of community taxonomic relatedness, as quantified by the weighted Unifrac metric. OMZ depths are circled. **(c)** Relative abundance of major bacterial divisions within the Ribosomal Database Project (RDP) classification, as a percentage of total identifiable bacterial sequences. Colors and ordering of taxa match those in D. Samples are labeled by depth and filter type, where  $p = \text{prefilter}$  ( $> 1.6 \mu\text{m}$ ),  $s = \text{Sterivex}$  ( $0.2\text{--}1.6 \mu\text{m}$ ). Duplicates in B and C reflect duplicate PCR reactions. **(d)** Variation in the relative abundance of bacterial divisions between filter size fractions. Values are the base-10 logarithm of the odds ratio: the ratio of the odds a taxon occurs on the prefilter to the odds it occurs on the Sterivex. Positive values indicate taxa that are more likely to occur on the prefilter. Values are based on counts pooled from all depths, with corrections for differences in data set size. Panels **c** and **d** exclude divisions present in only one filter type and occupying less than 0.01% of total bacterial sequences. Stars mark taxa whose relative abundance differed significantly between size fractions ( $P < 0.05$ ).

To compare community composition between samples, sequences were aligned using the PyNAST aligner in QIIME and beta diversity was calculated using the weighted Unifrac metric. This metric compares samples based on the phylogenetic relatedness (branch lengths) of OTUs in a community, while taking into account relative OTU abundance (Lozupone and Knight, 2005). Values range from 0 to 1, with 1 indicating the maximum distance between samples. Sample relatedness based on Unifrac was visualized using a two-dimensional Principal Coordinate Analysis (Figure 1b). For all pairs of samples, a Monte Carlo permutation test (1000 permutations) was used to determine if the Unifrac distance between samples was greater than expected by chance, with a false discovery rate correction ( $\alpha = 0.05$ ) imposed for multiple tests according to Benjamini and Hochberg (1995) (see Supplementary Table S2).

#### Sequence analysis—metagenomes

Analysis of protein-coding metagenome sequences followed that of Canfield *et al.* (2010) and Stewart *et al.* (2012b). Duplicate reads sharing 100% nucleotide similarity and identical lengths, which may represent artifacts of pyrosequencing, were identified by clustering in the program CD-HIT (Li and Godzik, 2006) and removed from each data set as in Stewart *et al.* (2010). Metagenome sequences were compared using BLASTX against the NCBI-nr database of nonredundant protein sequences (as of January 2012). BLASTX matches to prokaryote genes (Bacteria and Archaea) above a bit score of 50 were retained and classified according to functional category based on the SEED classification of functional roles and subsystems (Overbeek *et al.*, 2005), using the program MEtaGenome ANalyzer 4 (MEGAN Version 4; Huson *et al.*, 2011). The relative abundance of a SEED subsystem was calculated for each sample as the number of sequences per subsystem normalized by the total number of sequences matching subsystems; these values were then averaged across the four depths to obtain the SEED abundances shown in Figure 2. Normalized SEED counts for individual depths are available in Supplementary Table S4. The taxonomic composition of protein-coding sequences was determined based on the taxonomic annotation of each gene, according to the NCBI-nr taxonomy in MEGAN4.

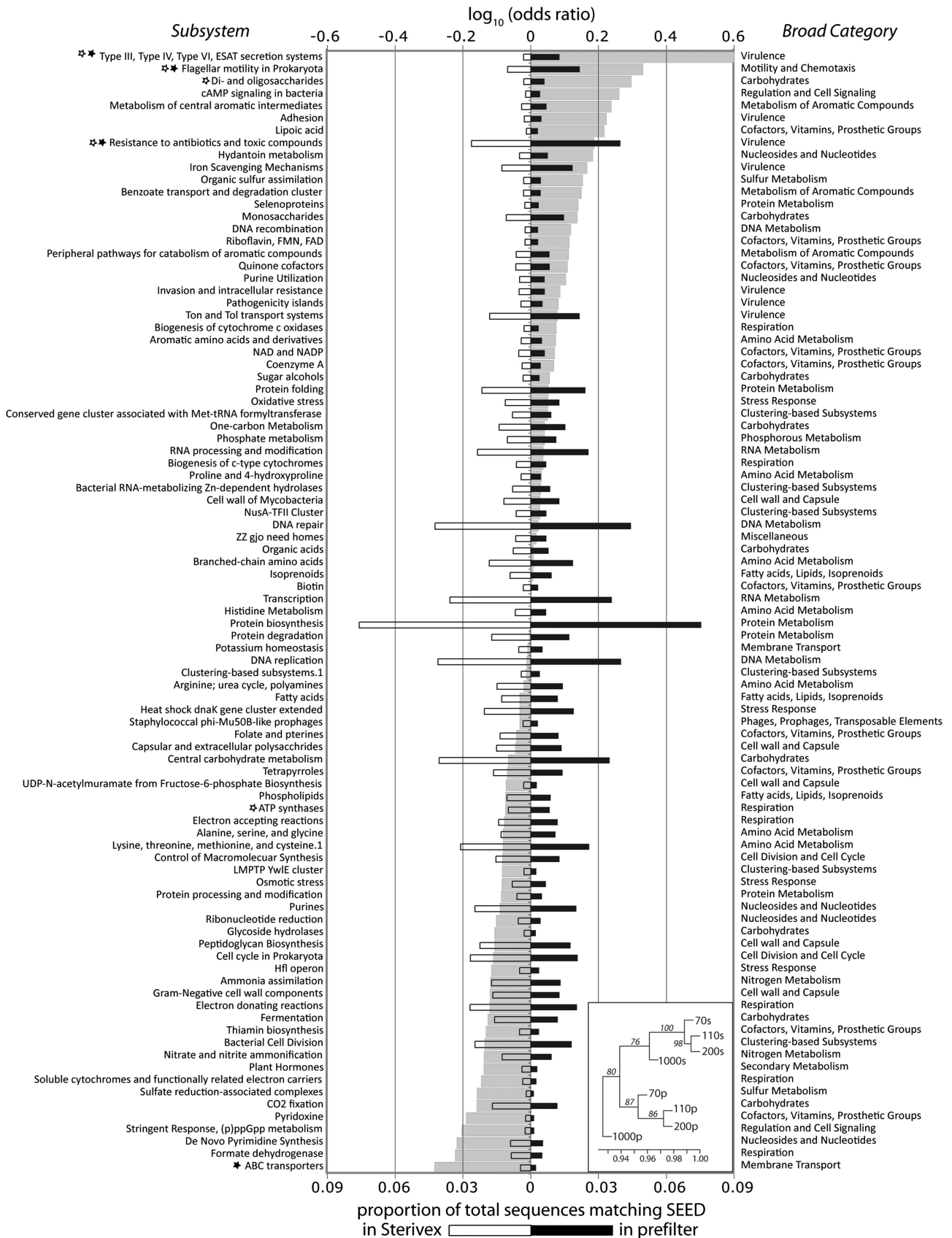
Samples were clustered based on SEED subsystem profiles (Figure 2, inset). For each sample, hit counts per subsystem were normalized to the proportion of total prokaryote reads matching SEED. An arcsine square root transformation was applied to proportions to stabilize variance relative to the mean. Pearson's correlation coefficients were calculated for each pair of transformed data sets and used as similarity metrics for hierarchical clustering using the complete-linkage method. The probability of

sample clusters was evaluated via multiscale bootstrapping (1000 replicates) based on the approximately unbiased method, implemented using the program pvclust in the R language. Bray–Curtis dissimilarities were also calculated from transformed count data to assess dissimilarity in functional gene composition (Supplementary Table S3).

To further evaluate genes or gene categories not represented in SEED, BLASTX results ( $>$ bit score 50) were manually parsed via keyword searches based on NCBI-nr annotations, as in Canfield *et al.* (2010). NCBI-nr genes representing top BLASTX matches were recovered from GenBank, and each database entry was examined manually to confirm gene identity. Entries with ambiguous annotations were further verified by BLASTX. Manual searches focused on key enzymes of dissimilatory nitrogen and sulfur metabolism (Lam *et al.*, 2009; Canfield *et al.*, 2010; Figure 3), including: ammonia monooxygenase (*amoC*), nitrite oxidoreductase (*nxrB*), hydrazine oxidoreductase (*hzo*), nitrate reductase (*narG*), nitrite reductase, nitric oxide reductase (*norB/norZ*) and nitrous oxide reductase (*nosZ*). Nitrite reductase genes were further characterized as either *nirK* or *nirS*, encoding the functionally equivalent copper-containing and cytochrome *cd*<sub>1</sub>-containing nitrite reductases, respectively, or as *nrfA*, encoding the cytochrome *c* nitrite reductase. Sulfur metabolism enzymes included dissimilatory sulfite reductase (*dsrA*), APS reductase (*aprA*) and sulfate thiol esterase (*soxB*). We also present results for key genes involved with mobile element activity, notably transposases and integrases, which were highly represented in the data sets. Genes encoding transposases and integrases from diverse families were combined into single categories for presentation.

When possible (Figure 3), gene abundances were normalized based on best approximate gene length (kb), estimated based on full-length open reading frames from sequenced genomes: *amoC* (750 bp); *nxrB* (1500 bp); *hzo* (1650 bp); *narG* (3600 bp); *nirK* (1140 bp); *nirS* (1620 bp); *nrfA* (1440 bp); *norB* (1410 bp); *nosZ* (1950 bp); *dsrA* (1200 bp); *aprA* (1860 bp); *soxB* (1680 bp). The length of genes encoding transposases and integrases varies among gene family type. Here, averages of 900 and 1150 bp were used for transposases and integrases, respectively, based on averaging randomly selected full-length transposase and integrase genes (100 each) identified in our data sets. Sequence counts per kilobase of target gene were then normalized to counts of sequences matching the universal, putatively single-copy gene encoding RNA polymerase subunit B (*rpoB*, 4020 bp). A value of 1 (Figure 3) indicates abundance in the metagenome equivalent to that of *rpoB*, assuming the gene lengths used in our calculations.

The proportion of differentially abundant genes (in metagenomes) or taxa (in 16S amplicon data) between Sterivex and prefilter samples was estimated via an empirical Bayesian approach in the



R program baySeq (Hardcastle and Kelly, 2010), as in the study of Stewart *et al.* (2012a). As true replicates at each depth were not available, statistical validation of depth-specific differences between filter fractions was not possible. Consequently, all samples (depths) belonging to each filter type were modeled as biological replicates. The baySeq method assumes a negative binomial distribution of the data with prior distributions derived empirically from the data (100 000 iterations). Dispersion was estimated via a quasi-likelihood method, with the sequence count data normalized by data set size (that is, total number of prokaryotic protein-coding genes and total number of bacterial 16S rRNA sequences for the metagenome and amplicon analyses, respectively). Posterior likelihoods per gene category or taxon were calculated for models (sample groupings) in which genes/taxa were either predicted to be equivalently abundant in both prefilter and Sterivex samples or differentially abundant between filter types. A false discovery rate threshold of 0.05 was used for detecting differentially abundant categories.

All sequence data generated in this study are publicly available in the NCBI database under BioSample numbers SAMN02317187-SAMN02317194 (metagenomes) and SAMN02339399-SAMN02339426 (amplicons).

## Results and discussion

Attachment to suspended or sinking particles is a major life history strategy for marine microorganisms (Smith *et al.*, 1992; Simon *et al.*, 2002; Grossart, 2010) and consequently has an important role in structuring community taxonomic composition and biochemical activity (DeLong *et al.*, 1993; Crump *et al.*, 1999). However, the patterns by which microbial physiological traits are distributed between particle-associated and free-living communities are not well understood for many ocean regions. Characterizing particle-associated microbes may be especially important in OMZs, where local particle maxima have been positively related to both oxygen depletion (Garfield *et al.*, 1983; Whitmire *et al.*, 2009) and microbial metabolic activity (Naqvi *et al.*, 1993).

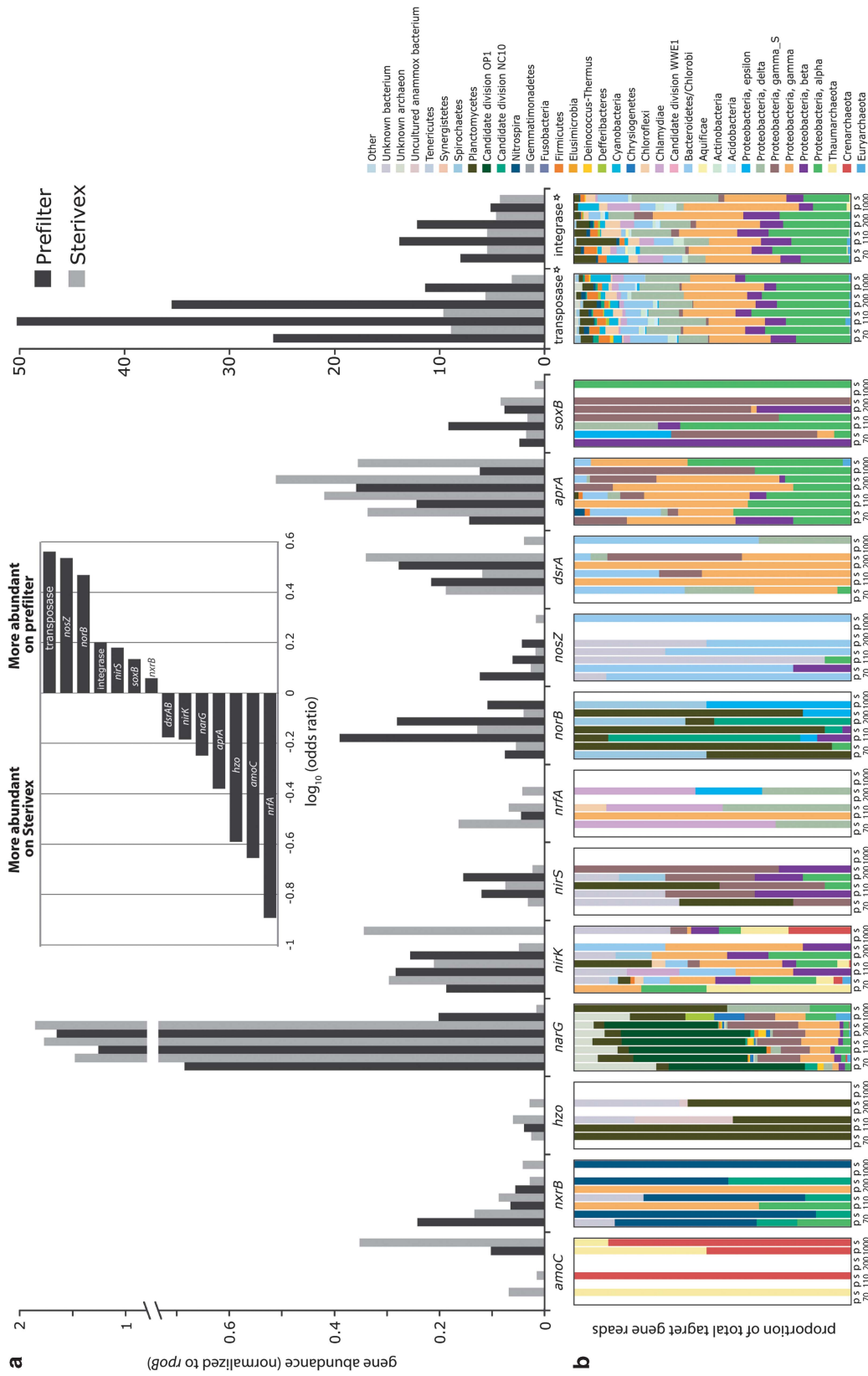
This study presents the first metagenomic comparison of microbial size fractions in a marine OMZ, and one of the first to examine community metabolic traits between size-fractionated marine

microbes in general. Microbial biomass from environmentally distinct depth zones spanning the permanent OMZ off Chile was separated by filtration into a large size fraction ( $>1.6\ \mu\text{m}$ ) and a small size fraction ( $0.2\text{--}1.6\ \mu\text{m}$ ), which for convenience are referred to here as 'prefilter' and 'Sterivex', respectively. Although the prefilter fraction is presumably enriched in particle-associated microorganisms and cell-cell aggregates, it may also contain larger free-living cells (for example, bacterial filaments and protists). Similarly, the small size fraction is likely dominated by free-living Bacteria and Archaea, but may also contain surface-associated microorganisms dislodged from particles during sampling (Hunt *et al.*, 2008). Consequently, filter size fraction is a potentially uncertain indicator of microbial life history strategy (that is, particle-attached versus free-living). Nonetheless, in comparing the taxonomic and functional gene compositions between fractions, the following sections highlight a significant role of size fraction in structuring microbial communities and identify physiological and genomic properties suggestive of a partitioning between surface-associated and free-living microbial strategies, as well as key OMZ processes of dissimilatory nitrogen and sulfur metabolism.

### Oxygen conditions

The ETSP OMZ sample site near Iquique, Chile was characterized by steep vertical gradients in dissolved oxygen (Figure 1a), similar to what has been reported previously for this region (Dalsgaard *et al.*, 2012; Stewart *et al.*, 2012b; Ulloa *et al.*, 2012). The base of the photic zone (1% surface PAR) occurred at  $\sim 40\ \text{m}$ , within the oxycline ( $\sim 30\text{--}70\ \text{m}$ ). Dissolved oxygen conditions at the time of sampling decreased from  $\sim 250\ \mu\text{M}$  at the surface to below  $5\ \mu\text{M}$  through the OMZ core ( $\sim 100\text{--}400\ \text{m}$ ), before gradually increasing below  $400\ \text{m}$  to  $\sim 60\ \mu\text{M}$  at  $1000\ \text{m}$ . The oxygen sensor used here has resolution in the micromolar oxygen range. However, recent measurements with high-resolution ( $10\ \text{nm}$ ) switchable trace oxygen sensors indicated that the ETSP OMZ core is anoxic, with oxygen below detection throughout the OMZ core (Thamdrup *et al.*, 2012). Our amplicon data sets therefore span the oxygenated photic zone and oxycline (5 and 32 m samples), the suboxic ( $<10\ \mu\text{M}$ ) upper OMZ (70 m), the anoxic OMZ core (110, 200, 320 m), and the oxic zone beneath the

**Figure 2** Differences in the relative abundance of functional gene categories between microbial size fractions (filter type), summarized across depths. Categories on the left are subsystems in the SEED classification, with the figure showing only subsystems comprising  $>0.1\%$  of the total sequences matching SEED. Higher level classifications of each subsystem are listed on the right. Filled and unfilled black bars reflect the relative abundance of prokaryotic sequence reads matching each category, normalized to the total number of prokaryotic sequences matching SEED. Light gray bars reflect the base-10 logarithm of the odds ratio: the ratio of the odds a gene category occurs on the prefilter to the odds it occurs on the Sterivex. Positive values indicate categories that are more likely to occur on the prefilter. Values are based on counts pooled from all depths sampled for metagenomics (70, 110, 200 and 1000 m), with corrections for differences in data set size. Categories whose relative abundance differed significantly between size fractions ( $P < 0.05$ ; baySeq) are starred for analyses based on all four depths (filled stars) or only the OMZ depths (70, 110 and 200 m; open stars). Dendrogram (inset) shows relatedness of individual samples based on SEED subsystem profiles, with samples labeled by depth and filter type ( $p$  = prefilter,  $s$  = Sterivex). Numbers at nodes are probabilities based on multiscale bootstrap resampling (1000 replicates). X axis = correlation coefficients.



**Figure 3** Relative abundance (a) and taxonomic representation (b) of sequences matching genes of key dissimilatory nitrogen and sulfur pathways. Abundance is calculated as read count per gene per kilobase of gene length, and shown as a proportion of the abundance of the universal single-copy gene encoding RNA polymerase subunit B (*rpoB*). A value of '1' indicates gene abundance equal to that of *rpoB*. Taxonomic identifications are based on annotations of NCBI reference sequences identified as top matches (above bit score 50) in BLASTX searches. See Methods for gene identifications. Stars mark genes whose abundance differed significantly between size fractions ( $P < 0.05$ ; baySeq) in an analysis of only the OMZ depths (70, 110 and 200 m). Samples are labeled by depth and filter type, where  $p$  = prefilter ( $> 1.6 \mu\text{m}$ ),  $s$  = Sterivex ( $0.2-1.6 \mu\text{m}$ ). Taxonomic group 'Proteobacteria, gamma\_S' indicates Gammaproteobacteria of the sulfur-oxidizing SUP05 clade (Walsh *et al.*, 2009). Inset shows the base-10 logarithm of the odds ratio for each gene category: the ratio of the odds a gene occurs on the prefilter to the odds it occurs on the Sterivex. Values are based on counts pooled across depths, with corrections for differences in data set size.



OMZ (1000 m). The metagenome samples focus on a subset of depths in the upper OMZ (70 m), OMZ core (110, 200 m) and beneath the OMZ (1000 m).

These data sets likely also span a gradient in bulk particle load. Consistent with prior studies reporting elevated particle concentrations within the ETSP OMZ (Pak *et al.*, 1980; Whitmire *et al.*, 2009), particulate load, inferred indirectly here from beam attenuation measurements, exhibited local maxima within the upper photic zone (~15 m) and then again within the OMZ core (~140 m), before declining to a consistent minimum below the OMZ (Supplementary Figure S1). However, the size distribution and composition of particles contributing to the beam attenuation signal are not characterized here. It therefore remains to be determined how changes in bulk particle load relate to changes in the abundance of the size-fractionated communities discussed in detail below.

#### *Taxonomic diversity—16S rRNA gene amplicons*

Pyrosequencing of 16S rRNA gene amplicons revealed a species-rich OMZ bacterial community whose composition varied over depth and between size fractions. A total of 17 014 bacterial OTUs (97% similarity clusters) were recovered across all samples, with per sample OTU counts ranging from 658 to 2484 based on data set size (Supplementary Table S1). Despite relatively high numbers of sequences per sample (mean: 14 527), this analysis did not capture the total OTU richness in each sample (that is, no rarefaction curves approached saturation; Supplementary Figure S2), as anticipated for marine bacterioplankton assemblages (Huber *et al.*, 2007).

*OTU diversity with depth.* OTU diversity patterns varied with both depth and filter size fraction (Figure 1, Supplementary Figures S2–S4). For both size fractions, PD, the total branch length connecting all OTUs in the 16S rRNA gene phylogeny (Faith, 1992), was shortest at the surface (5 m) and increased within the oxycline (~30–70 m) (Figure 1a). However, PD of the two size fractions differed within the OMZ. PD of Sterivex communities decreased from the oxycline to the anoxic OMZ core at 200 m, whereas PD of prefilter communities increased within the core (Figure 1a). PD trends closely paralleled those of other alpha diversity indices, including counts of observed OTUs and Chao1 estimates (Supplementary Figure S3).

Vertical patterns of OMZ microbial diversity are not consistent among studies. A decline in PD from the oxycline to anoxic depths, based on 16S rRNA gene fragments from metagenomes, was observed for the 0.2–1.6  $\mu\text{m}$  size fraction across years and seasons in the ETSP OMZ off Chile (Bryant *et al.*, 2012), suggesting temporal stability and a consistent decline in diversity within the OMZ free-living community. Low diversity associated with suboxia was also reported in a gene fingerprinting

study of the 0.2–2.7  $\mu\text{m}$  fraction from a seasonal OMZ off British Columbia (Zaikova *et al.*, 2010). In contrast, Stevens and Ulloa (2008), based on libraries of cloned 16S sequences, identified a peak in OTU diversity (multiple indices) at the ETSP OMZ within the 0.2–3  $\mu\text{m}$  fraction, a pattern consistent with that observed for the prefilters in our study. Similarly, elevated OTU richness in the 0.2–1.6  $\mu\text{m}$  fraction has been shown to coincide with the zone of minimum oxygen concentration at tropical non-OMZ sites (Brown *et al.*, 2009; Kembel *et al.*, 2011).

These studies consistently highlight a shift in microbial community complexity associated with zones of low oxygen. For the ETSP OMZ, where oxygen declines to the nanomolar range, increasing diversity in the OMZ core has been hypothesized to be linked to the use of a wider range of terminal oxidants, compared with non-OMZ depths where oxygen is the dominant electron acceptor (Stevens and Ulloa, 2008). Conversely, Bryant *et al.* (2012) argue that niche diversity declines within the anoxic OMZ as niches linked with light and labile organic matter utilization, which are more prevalent at the surface and oxycline, are lost. Our data confirm that taxonomic diversity varies between size fractions, with diversity elevated in larger size fractions within the OMZ. Similarly, elevated OTU richness in particle-associated compared with free-living communities has been reported for other marine habitats (for example, Eløe *et al.*, 2010), suggesting that higher diversity may be linked to an increase in niche richness associated with microgradients in substrate availability and composition on particles. However, this pattern is not observed across all depths or ocean sites (for example, Figure 1a; Moeseneder *et al.*, 2001; Ghiglione *et al.*, 2007), suggesting a need for quantifying niche availability in response to diverse parameters, notably the organic composition, abundance, and size distribution of particles, combined with physical and chemical gradients of the bulk water column.

*Taxonomic composition.* Bacterial taxonomic composition varied markedly among samples. Vertical trends in the community composition of free-living bacteria in the ETSP OMZ have been reported previously (Stevens and Ulloa, 2008; Bryant *et al.*, 2012) and agree broadly with those observed here. We instead focus primarily on comparisons between size fractions. Figure 1d shows the odds of a given taxonomic division occurring in the prefilter fraction relative to the Sterivex fraction, based on OTU counts pooled across depths. Of the 25 major bacterial divisions identified in the amplicon analysis, 15 were significantly over- or underrepresented in the prefilter fraction ( $P < 0.05$ , baySeq). A subset of these trends is discussed below.

Alphaproteobacteria sequences were abundant in both filter fractions but were consistently enriched

in the free-living community, where they constituted an average of 32% of all sequences (versus 13% in the prefilters) (Figure 1c). Enrichment was driven primarily by the SAR11 clade (*Pelagibacter* sp.), which represented 50–97% (mean: 84%) of Alphaproteobacteria sequences from Sterivex filters, and 25–72% (mean: 58%) of those from prefilters. SAR11 enrichment in the Sterivex fraction is consistent with these bacteria being free-living oligotrophs adapted for the efficient use of dissolved substrates (Giovannoni *et al.*, 2005). In contrast, SAR11 sequences in the prefilter fraction may represent unique surface-associated ecotypes, as proposed for SAR11 detected in larger size fractions (0.8–3.0 and 3.0–200.0  $\mu\text{m}$ ) from an oxic upwelling zone (Allen *et al.*, 2012). Diversity surveys that do not examine the prefilter fraction may be excluding important components of the SAR11 community. Here, SAR11 were abundant at both oxic and anoxic depths, as shown previously for the ETSP OMZ (Stevens and Ulloa, 2008; Stewart *et al.*, 2012a,b). However, the metabolic adaptations that enable these putatively aerobic bacteria to grow under low or no oxygen remain uncharacterized (Wright *et al.*, 2012). Future analyses at finer levels of taxonomic resolution may identify SAR11 subclades unique to both particle-associated and low-oxygen environments of the ETSP.

Sequences matching high GC gram-positive Actinobacteria (Actinomycetes) were a relatively minor component of the total amplicon pool (mean: 2% across all samples), but were significantly more abundant in Sterivex filters (excluding the 5 m sample where Actinobacteria were negligible in both size fractions; Figure 1c). A similar enrichment was observed recently in the 0.1–0.8  $\mu\text{m}$  bacterioplankton fraction from temperate coastal communities (Smith *et al.*, 2013). Actinobacteria are most commonly associated with terrestrial soil habitats but are also regularly cultivated from marine sediments and, less commonly, from suspended organic aggregates (Grossart *et al.*, 2004) and pelagic environments (Rappe *et al.*, 1999; Bull *et al.*, 2005), including OMZs (Fuchs *et al.*, 2005). Here, the majority of Actinobacterial sequences (76%) were unclassified (data not shown), suggesting the possibility of novel planktonic lineages, potentially distinct from those associated with particles (Jensen and Lauro, 2008; Prieto-Davo *et al.*, 2008).

Twelve major bacterial divisions were significantly enriched in prefilter communities (Figure 1d). Diverse clades of the Bacteroidetes, including the Flavobacteria and Sphingobacteria, were among the most overrepresented groups in this fraction. As suggested in prior reports from non-OMZ systems (Crump *et al.*, 1999; Simon *et al.*, 2002; Allen *et al.*, 2012), elevated numbers of Bacteroidetes on particles may be linked to the enhanced capacity of these bacteria to degrade high molecular weight biopolymers, such as chitin or proteins (Cottrell and Kirchman, 2000). Members of

the Spirochetes and Mollicutes, though at negligible overall abundance in our data sets (<0.1%), were also significantly enriched in the large size fraction. Spirochetes, which are traditionally associated with marine sediments or microbial mats, were only detected at the core OMZ depths (110, 200 m), consistent with this group being dominated by strictly or facultatively anaerobic members (Munn, 2011). Marine spirochetes have also been found in association with eukaryotes (Ruehland *et al.*, 2008; Demiri *et al.*, 2009), and may therefore be enriched in the particle fraction via attachment to larger organisms or sinking fecal matter. Similarly, sequences matching Mollicutes, which here were affiliated exclusively with the *Mycoplasma* (data not shown), may have originated from eukaryotic material, as mycoplasmas have been found in the larval stages of marine invertebrates and in the intestinal microflora of several fish species (Zimmer and Woollacott, 1983; Bano *et al.*, 2007).

Sequences matching eukaryotic chloroplasts or cyanobacteria were also more abundant on average on the prefilters. This sequence group was a substantial component (20–38%) of both filter fractions at 5 m (Figure 1c), but was confined primarily to the prefilter beginning at 32 m. This change in size fraction was accompanied by a shift in the structure of the eukaryotic phototroph community, from dominance by Chlorophyta and Cryptomonadaceae at 5 m to Bacillariophyta (diatoms) at 32 m and below (data not shown). Throughout the depth range, cyanobacterial sequences primarily matched clade GpIIa (for example, *Prochlorococcus* and *Synechococcus*), with abundance peaking in the prefilters at 32 and 70 m within the photic zone. The presence of cyanobacterial-like sequences below the photic zone (Figure 1c) has been reported previously for the ETSP OMZ (Bryant *et al.*, 2012) and in other deep-water habitats (Smith *et al.*, 2013) and may be due to aggregation onto or release from sinking particles, including fecal pellets. Here, the relative abundance of these sequences (namely chloroplasts) increased in the 1000 m prefilter, which could be due to changes in the turnover rates of different particle-associated cell fractions (that is, chloroplasts embedded in fecal particles increase in relative abundance as bacterial activity and cell numbers on particles decrease).

Consistent with several prior studies of particle-associated bacteria (DeLong *et al.*, 1993; Fuchsman *et al.*, 2012), Deltaproteobacteria were significantly enriched on prefilters. Although this group was a negligible component of both size fractions at the surface (<0.1% of total sequences at 5 m), deltaproteobacterial sequences increased in relative abundance in both fractions by 70 m, and were 8- to 28-fold more abundant in prefilters from 70 m down to 1000 m, representing up to 3% of total sequences in the larger size fraction (Supplementary Figure S8). Notably, the Myxococcales, a widely distributed

Order with both terrestrial and marine members that exist primarily in surface-attached swarms (Shimkets *et al.*, 2006; Jiang *et al.*, 2010), were up to 95-fold more abundant on prefilters (relative to Sterivex) from depths below the oxycline. Marine myxobacteria have been found in anoxic sediments, but are associated primarily with oxic habitats (Brinkhoff *et al.*, 2012), and the Order as a whole is dominated by strictly aerobic heterotrophs (Shimkets *et al.*, 2006). Myxobacteria have also been found in open ocean picoplankton (DeLong *et al.*, 2006; Pham *et al.*, 2008). It therefore is possible that OMZ myxobacteria, as well as other particle-associated taxa, have been transported to anoxic depths after attachment to sinking particles in the oxic zone, but are not metabolically active in the OMZ.

Deltaproteobacteria clades known to contain sulfate-reducing members (for example, Desulfobacterales) were generally more abundant at core OMZ depths (110, 200, and 320 m; Supplementary Figure S8), consistent with their low-oxygen requirements. However, the relative abundance of these groups did not differ appreciably between prefilter and Sterivex fractions, except within the 320 and 1000 m samples where these sequences were barely (or not) detectable in the free-living fraction (Supplementary Figure S8). Sulfate reduction in oxic water columns presumably is localized to reduced microzones on particles (Shanks and Reeder, 1993). Our data raise the possibility that sulfate reduction in the OMZ, which has been demonstrated recently using radiolabeling of bulk water samples (Canfield *et al.*, 2010), may not be confined to particle-associated microhabitats. However, the vast majority (mean: 72%) of deltaproteobacterial 16S sequences across both filter fractions were unclassified (Supplementary Figure S8). Classification at higher levels of phylogenetic resolution will be necessary to clarify how particle-attachment in the OMZ affects the distribution of deltaproteobacterial subclades, including those with sulfate reducers.

rRNA amplicons matching members of the superphylum comprising the Planctomycetes, Verrucomicrobia, Lentisphaerae and Chlamydiae (Wagner and Horn, 2006) were significantly overrepresented on prefilters (Figure 1d). Notably, on average across the depths, the relative abundance of Planctomycetes was 15-fold higher in prefilter compared with Sterivex samples. However, this enrichment was not uniform throughout the water column. Planctomycete sequences were either not detectable or a very minor percentage (0–0.3%) of total amplicons in the oxic 5 and 32 m samples, even within the prefilter fraction (Supplementary Figure S7). In contrast, Planctomycetes represented 1–2% of total sequences in the prefilter at 70 m and increased to a peak of ~5% at 1000 m (Supplementary Figure S7). A similar depth-specific increase in relative enrichment was observed in the Verrucomicrobia and Lentisphaerae. Presumably, the distribution of these groups, which predominantly contain anaerobic members, is tied to

the presence of anoxia, which may be scarce on newly formed particles in the oxic depths, but relatively common in older, deeper particles where microbial respiration has created local pockets of oxygen depletion. It is also possible that the sinking of particles into the suboxic OMZ facilitates particle-associated anaerobic metabolisms.

Amplicons matching Planctomycetes provided limited phylogenetic information, with 63% of all planctomycete amplicons (both filter fractions) identified only to the Family Planctomycetaceae. This pattern was most pronounced at 1000 m, where 84% of Planctomycetaceae sequences were unclassified. Of the sequences assignable to a Genus, 54% matched the Genus Planctomyces, with the vast majority of these being detected only in the prefilters where Planctomyces were enriched on average 75-fold compared with the Sterivex fraction. This pattern agrees with genetic and isolation-based studies identifying surface attachment as a key life history state for diverse Planctomycete genera, including Planctomyces (Bauld and Staley, 1976; Morris *et al.*, 2006; Bengtsson and Øvreås, 2010) and with a general Planctomycetes enrichment in particle-associated microbial cell fractions (DeLong *et al.*, 1993; Fuchsman *et al.*, 2011, 2012). Although the 16S amplicon pool provided limited phylogenetic resolution for some taxonomic groups discussed here, additional insight into the composition of key OMZ clades can be provided by analyzing the taxonomic identification of protein-coding genes (see Metagenome data below).

**Community relatedness.** Sample relatedness based on community phylogenetic composition (Unifrac metric) varied with depth. For both the prefilter and Sterivex sample sets, communities on the periphery (70 m) and within the OMZ (110, 200 and 320 m) clustered to the exclusion of those from the surface (5 and 32 m) and beneath the OMZ (1000 m), although this pattern was most pronounced for the prefilter communities (Figure 1b and Supplementary Figure S4B,C). Notably, the communities at 5 m (prefilter and Sterivex) were highly distinct from those at deeper depths, due primarily to the high abundance of cyanobacteria and eukaryotic chloroplasts at the surface, as well as a shift in the structure of the proteobacterial community (Figures 1b and c). Of the OMZ depths, samples from 70, 110 and 200 m were most closely related (Figure 1b and Supplementary Figure S4B,C). The 320 m sample was a relative outlier. Specifically, the 320 m Sterivex community was enriched in Flavobacteria (primarily ‘unclassified’ Flavobacteria) compared with the other OMZ depths (Figure 1c). This shift highlights the potential for community variation throughout the OMZ core, despite apparent uniformity in some environmental conditions, notably oxygen, across these depths (Figure 1a, Thamdrup *et al.*, 2012).

Filter size fraction also had a major role in determining community relatedness. Sterivex

communities clustered to the exclusion of the corresponding prefilter communities from the same depth (Figure 1b). Ninety-three percent (182/196) of the pairwise comparisons between filter types (*p* versus *s* in Supplementary Table S2, top) revealed significant differences in taxonomic composition based on phylogenetic relatedness ( $P < 0.05$ ; mean Unifrac: 0.53). At only one depth, 320 m, did prefilter and Sterivex communities not differ significantly ( $P = 0.11$ – $0.13$ ; note similarity in PC1 coordinates in Figure 1b), although even in this sample clear differences between filters were evident (Figure 1c). In contrast, in comparisons involving the same filter type, 54 (49/91) and 26% (24/91) of prefilter and Sterivex comparisons showed significant differences (mean Unifrac: 0.39 and 0.35 for prefilters and Sterivex, respectively). Of the 14 comparisons involving data from duplicate PCR reactions, two (the 32 m prefilter and 1000 m Sterivex samples) showed significant compositional differences, indicating a potential for PCR-induced variation to influence diversity comparisons.

Even when the outlier surface sample (5 m) was excluded from the analysis, clustering patterns indicated that microbial size fraction was a stronger predictor of community relatedness than was vertical position in the water column (Figure 1b, Supplementary Figure S4A). Many prior studies have confirmed fundamental differences in community composition between size fractions (DeLong *et al.*, 1993; Acinas *et al.*, 1999; Crump *et al.*, 1999; Ghiglione *et al.*, 2007; Parveen *et al.*, 2011). However, others show relative similarity between fractions (Hollibaugh *et al.*, 2000). For example, communities from three size fractions (3.0–20, 0.8–3.0 and 0.1–0.8  $\mu\text{m}$ ) from a surface ocean sample grouped together based on shared metagenome sequence, distinct from communities at distant oceanic sites (Rusch *et al.*, 2007). A similar pattern, whereby bacterial communities from distinct size fractions of the same sample are more similar than communities from other samples, has been shown for the deep ocean (Eloe *et al.*, 2010) and the anoxic Black Sea (Fuchsman *et al.*, 2011) based on 16S gene sequences, and for a coastal hypoxic layer based on the taxonomic annotations of coding genes from metagenomes (Smith *et al.*, 2013). Conflicting patterns of community relatedness are potentially due to differences in size-fractionation and taxonomic identification methods across studies, as well as variation in water column conditions among samples. Indeed, zonation in parameters such as light, temperature or nutrient availability significantly influences taxonomic diversity and metabolic function across diverse marine habitats (DeLong *et al.*, 2006; Qian *et al.*, 2011), including the ETSP OMZ (Stevens and Ulloa, 2008; Bryant *et al.*, 2012; Stewart *et al.*, 2012b), and is therefore a critical driver of niche differentiation in ocean microbes (Rocap *et al.*, 2003; Johnson *et al.*, 2006). Nonetheless, in our study, samples from depths

spanning the oxycline and OMZ core clustered by size fraction despite strong vertical stratification in environmental parameters such as oxygen (Figure 1).

These trends suggest that life history mode associated with size fraction (free-living versus particle-associated) has a greater role in structuring OMZ communities than water column oxygen levels. Although  $\text{O}_2$  and nutrient concentrations at the anoxic OMZ core are dramatically different from those of the overlying oxic layer, the taxonomic composition of OMZ particles broadly reflects that of particles from oxic marine habitats (DeLong *et al.*, 1993; Rath *et al.*, 1998), with a relative enrichment of the Bacteroidetes, Firmicutes, Planctomycetes and Deltaproteobacteria, and an underrepresentation of bacteria typically associated with oligotrophic conditions (for example, SAR11). Many of the clades enriched on prefilters also contain aerobic members (for example, Flavobacteria and Myxobacteria), raising the question of whether particle-associated bacteria are metabolically active within the OMZ, or are quiescent, having been transported to OMZ depths on particles originating in the oxic zone. Indeed, size fraction-specific clustering of samples suggests that passage through the anoxic OMZ may have a relatively minor effect on the composition of the particle-associated microbial communities. However, release of microbes from sinking particles may represent a valuable conduit of anaerobic bacteria to OMZ depths.

#### *Taxonomic and functional variation among size fractions—Metagenomics*

Shotgun metagenomics was used to examine differences in taxonomic composition and metabolic function between free-living and particle-associated size fractions. Pyrosequencing of eight metagenome samples (four depths, two filters per depth) generated 1 660 922 reads (range: 163 987–275 575 per sample; mean length: 305 bp; Table 1). Of these, 48% were designated as protein-coding based on BLASTX matches (bit score  $> 50$ ) to proteins in the NCBI-nr database, with 10% of these matching prokaryotic genes in SEED Subsystem categories. The percentage of identifiable protein-coding sequences was consistently higher in the Sterivex fraction (mean: 66% of total sequences matching NCBI-nr, compared with 24% on prefilters). This discrepancy was likely due in part to the enrichment (8–16-fold) of eukaryotic genes on prefilters (Table 1), which would have increased the proportion of non-coding DNA per metagenome. In addition, prefilters were enriched threefold in sequences matching genes annotated as viral in origin. Presumably, most free-living marine viruses are too small ( $< 0.2 \mu\text{m}$ ) to have been retained in either filter fraction. The viral reads in the data therefore likely originate either from extracellular viruses attached to the surfaces of cells or particles (for example, within a biofilm), or from prophage. As we did not

**Table 1** Metagenome sequence statistics and taxonomic (Domain) identities

Sample	Reads	Coding reads <sup>a</sup>	% Of Bacteria	% Of Archaea	% Of unidentified prokaryotes	% Of eukaryotes	% Of viruses	% Of other	SEED <sup>b</sup>
70 m pf	163 987	37 841	79.4	5.3	1.6	10.5	2.1	1.1	2823
70 m st	275 575	188 452	87.3	6.4	3.2	1.3	0.7	1.1	20 268
110 m pf	142 140	38 957	84.6	3.0	1.1	8.6	1.7	1.0	3190
110 m st	238 182	151 560	90.5	2.6	4.3	1.1	0.7	0.9	18 104
200 m pf	204 644	65 128	85.5	2.1	1.3	8.3	1.9	1.0	6081
200 m st	253 507	170 856	92.5	1.6	3.8	0.9	0.5	0.7	22 837
1000 m pf	209 778	31 332	72.1	3.4	1.1	21.3	1.4	0.7	2205
1000 m st	173 089	113 693	73.5	20.2	3.7	1.3	0.5	0.9	10 132

Abbreviations: pf, prefilter; st, sterivex.

<sup>a</sup>No. of reads matching coding genes in the NCBI-nr database (above bit score 50) % = percentage of coding reads matching reference genes from bacteria, archaea, unidentified prokaryotes, eukaryotes, viruses or other (genes lacking a taxonomic annotation, or annotated as 'unknown').

<sup>b</sup>No. of prokaryote reads assigned to SEED subsystem categories.

distinguish between prokaryote and eukaryote-derived viruses, it is possible that the prefilter viruses originated from eukaryotes. In the following sections, we focus on protein-coding sequences of prokaryotes. We first describe taxonomic patterns inferred from coding gene annotations in contrast to those based on amplicon data, and then highlight functional categories involved with key OMZ biogeochemical processes, as well as categories that differed notably between size fractions.

#### Taxonomic composition—protein-coding genes.

The taxonomic identities of bacterial protein-coding genes broadly reflected those inferred from 16S rRNA gene amplicons, with important caveats. For this comparison, certain phylogenetic groups detected in the amplicon data sets (Figure 1) were collapsed to higher taxonomic levels to match groupings based on protein-coding genes (Supplementary Figure S5). Analysis of 16S amplicons from the four depths where metagenomes were sampled (70, 110, 200 and 1000 m; all depths combined) revealed 10 major bacterial divisions (out of 20) that were at higher relative abundance on prefilters. Of these, nine also were enriched in prefilters based on protein-coding data sets (Supplementary Figure S5). However, the magnitude of this enrichment was markedly higher in comparisons using amplicon data (Supplementary Figure S5). For example, in the amplicon analysis, the odds of detecting a deltaproteobacterial sequence were 11-fold greater in the prefilter relative to the Sterivex fraction. In contrast, when inferred from protein-coding sequences, these odds were effectively equal between filter fractions (odds ratio: 1.1). Similar patterns were evident for the Planctomycetes, Spirochetes, Tenericutes and Epsilon- and Betaproteobacteria (Supplementary Figure S5C, F). On average, the major bacterial divisions were more evenly represented in the metagenome data (Simpson's *E*: 0.45 and 0.36 for prefilter and Sterivex, respectively) compared with the amplicon data (Simpson's *E*: 0.32 and 0.21 for prefilter and Sterivex) (Supplementary Figure S5). The cause of

this discrepancy between data types is unclear, but may involve differences in the representation of taxonomic groups between the Ribosomal Database Project and NCBI-nr databases, in rRNA operon copy number among taxa, and in the phylogenetic resolution between protein-coding and rRNA gene fragments. Indeed, both the 16S and coding gene data sets contained large numbers of sequences that could not be assigned to a bacterial phylum (Supplementary Figure S5). Additional reference sequences, including whole genomes of OMZ microorganisms, will likely increase the probability and accuracy of read assignment, and potentially resolve differences between rRNA and coding gene-based identifications. Until then, however, studies should account for the chance that community composition shifts can be underestimated or misinterpreted when based on metagenome data.

Protein-coding gene annotations can nonetheless provide taxonomic insight by identifying groups not well represented in 16S databases. For example, in contrast to the amplicon data, which revealed an overall enrichment of Deltaproteobacteria on prefilters, coding sequences matching the ubiquitous deltaproteobacterial SAR324 cluster were 2 to 12-fold more abundant on Sterivex filters compared with prefilters (Supplementary Figure S8). Notably, SAR324-like reads peaked at >4% of total bacterial coding reads in the Sterivex fraction at 1000 m, consistent with reports of the mesopelagic distribution of this group (Wright *et al.*, 1997). SAR324 enrichment in the free-living fraction contrasts with recent genomic evidence indicating that this group is adapted for a particle-associated lifestyle (Swan *et al.*, 2011). However, this lineage also contains chemoautotrophic members (Swan *et al.*, 2011), which presumably would have less of a need to attach to organic-rich particles.

Coding genes suggested that the archaeal community also differs markedly between size fractions (Supplementary Figure S6). The Sterivex fraction was enriched up to sixfold (mean 3.3) in reads matching aerobic ammonia-oxidizing autotrophs of the Thaumarchaeota, with the highest representation

of Thaumarchaea in the suboxic 70 m sample at the top of the OMZ. A similar enrichment in the small (0.1–0.8  $\mu\text{m}$ ) size fraction was reported for low-oxygen ( $\sim 20 \mu\text{M}$ ) waters in the North Pacific (Smith *et al.*, 2013), highlighting the free-living lifestyle of this group, as well as adaptation to low-oxygen conditions. Similarly, reads matching Crenarchaeota were 2 to 11-fold (mean 4.2) more abundant in the free-living size fraction in the ETSP OMZ, peaking at 15% of total coding reads at 1000 m. The increase in Crenarchaeota with depth agrees with patterns from other subtropical and tropical sites (DeLong, 2003). Indeed, the overwhelming majority of Crenarchaeota sequences (91%) from 1000 m matched fosmid clones representing Group I Crenarchaeota collected at 4000 m in the Pacific Ocean (Konstantinidis and DeLong, 2008). Functional genes on these fosmids indicate a role in aerobic ammonia oxidation, suggesting the potential for re-classifying these sequences as Thaumarchaeota (Brochier-Armanet *et al.*, 2008). In contrast, sequences matching Euryarchaeota, which peaked in abundance in the upper OMZ samples (70 m), were on average 50% more abundant on prefilters. Of these sequences, those matching uncultured marine group II (MG-II) constituted the single largest fraction (mean: 40% of total Euryarchaeota sequences; Supplementary Figure S6). Recently, sequencing of a MG-II euryarchaeote genome indicated a motile heterotrophic lifestyle, with genes for proteins mediating adhesion, fatty acid metabolism and protein degradation suggesting adaptations to growth on marine particles (Iverson *et al.*, 2012). Consistent with this prediction, MG-II-like sequences were at 26 to 82% higher relative abundance in the prefilter fraction in the OMZ and oxycline depths (70, 110 and 200 m). Other Euryarchaeota sequences matched diverse clades, including those with methanogenic members, which were marginally enriched (15–44%) on prefilters (Supplementary Figure S6). Methanogenic and potentially hydrogenotrophic Euryarchaeota have previously been detected on marine particles (van der Maarel *et al.*, 1999; Ditchfield *et al.*, 2012), suggesting that anoxic microzones on particles create conditions favorable for these anaerobic taxa. It also has been hypothesized that methanogens in OMZs may occur in symbiotic associations with anaerobic protists (Orsi *et al.*, 2012), as observed in other diverse reducing habitats (Nowack and Melkonian, 2010; Edgcomb *et al.*, 2011). Overall, however, data directly comparing marine Archaea between free-living and particle-associated niches remain limited, and conflicting. Moderate differences in archaeal community composition have been observed between size fractions at coastal sites (Crump and Baross, 2000; Galand *et al.*, 2008, Smith *et al.*, 2013), but not at an open ocean site (Galand *et al.*, 2008). Our data indicate size fraction-specific variation, suggesting the need for more targeted studies exploring potential archaeal genotype and ecotype variation between microniches.

Protein-coding sequences provided additional insight into the taxonomic composition of the OMZ Planctomycete community. In contrast to the amplicon data, sequences matching Planctomycete genes were a large component of metagenomes from the free-living fraction, peaking at a high of 9% of identifiable coding reads at the 110 m OMZ depth before declining to 2% beneath the OMZ (Supplementary Figure S7). This pattern agrees with prior metagenomic data from the ETSP OMZ (Stewart *et al.*, 2012b). The taxonomic identities of Planctomycete-like coding genes differed significantly between size fractions and between OMZ and non-OMZ depths. Notably, Planctomycete sequences from Sterivex filters predominantly matched the marine anammox genus *Candidatus Scalindua*, whose sequences constituted 59–81% of Planctomycete reads from OMZ depths (Supplementary Figure S7). In comparison, *Candidatus Scalindua* represented 14–20% of Planctomycete reads in the prefilters at these depths, which were instead enriched in the non-anammox genus *Planctomyces*, as also indicated by the amplicon data. Analyses with genus-specific FISH probes previously showed that a minor fraction of total *Candidatus Scalindua* cells in the Namibian and ETSP OMZs associates directly with particles (Woebken *et al.*, 2007), consistent with our data. In contrast, marker gene surveys from the suboxic Black Sea detected this genus in the 0.2–30  $\mu\text{m}$  fraction but not in the larger particle-associated fraction above 30  $\mu\text{m}$  (Fuchsman *et al.*, 2012). As *Candidatus Scalindua* is presumed to be the primary lineage responsible for anammox in OMZs (Woebken *et al.*, 2008; Galan *et al.*, 2012), these patterns suggest that the bulk of anammox-capable cells in OMZs may be spatially separated at the microscale from potentially linked metabolic transformations on particles, for example, nitrite production and ammonia remineralization by heterotrophic denitrifiers. A shift to a higher proportion of free-living *Candidatus Scalindua* may be facilitated in OMZs where suboxia extends beyond particle microniches and also by the autotrophic metabolism of this organism, which may eliminate pressure to attach to carbon-rich particles.

*Trends in SEED subsystems.* Classification of sequences into SEED subsystems highlighted variation in functional content between filter fractions. Subsystem abundances were correlated ( $R > 0.94$ ) between samples (Figure 2, inset), and Bray–Curtis distances between prefilter and Sterivex SEED profiles (mean: 0.11) did not differ appreciably from those between samples of the same filter type (mean: 0.10 and 0.07 for Prefilter-only and Sterivex-only comparisons, respectively; Supplementary Table S3). This similarity reflects redundancy in house-keeping gene categories, notably protein biosynthesis, DNA repair and central carbohydrate metabolism (Figure 2), which are ubiquitous and abundant across

even highly divergent taxa (Burke *et al.*, 2011). However, despite broad similarity in subsystem profiles, filter fractions clustered independently of one another based on SEED content (Figure 2, inset), suggesting functional differences separating free-living from surface-attached life history modes.

The prefilter fraction was enriched in genes for navigating and persisting within a spatially and chemically heterogeneous environment. Figure 2 (main) shows the odds of a SEED subsystem occurring in the prefilter relative to the Sterivex fraction. These trends are based on data pooled across depths (that is, the four depths are treated as replicates) and therefore only identify categories that consistently differed between filter fractions. Genes mediating motility, chemotaxis and adhesion were among the most overrepresented on prefilters (Figure 2, Supplementary Table S4). These functions are presumably critical for detecting local patches of nutrients and energy substrates (Fenchel, 2002; Stocker *et al.*, 2008), colonizing surfaces (Fenchel, 2001), and potentially also for navigating substrate gradients on particles themselves. Notably, prefilters contained significantly higher counts of genes involved in bacterial secretion. Of these, 97% encoded elements of Type IV secretion systems, notably Type IV pili (84%) and mannose-sensitive hemagglutinin Type IV pili (8%). These cell surface structures mediate diverse functions, including gene exchange, transfer of effector proteins between cells, twitching motility and adherence (Christie *et al.*, 2005; Burrows, 2012), and have been shown to promote attachment to algal surfaces by marine bacteria (Dalisay *et al.*, 2006).

Prefilters also were enriched in genes encoding virulence and antibiotic resistance functions. Life in particle-associated biofilms would presumably increase the frequency of cell-to-cell contact, and therefore the likelihood of antagonistic interactions. Indeed, the production of antibacterial compounds is more common in surface-attached bacteria relative to planktonic cells (Long and Azam, 2001; Long *et al.*, 2005; Gram *et al.*, 2010) and has been shown to affect the colonization dynamics of marine particles (Grossart *et al.*, 2003). Prefilter metagenomes also contained high abundances of genes mediating signaling via the universal secondary messenger cAMP, which in prokaryotes regulates functions ranging from virulence, stress response and energy and carbon metabolism. A high abundance of cAMP signaling genes was shown recently for soil metagenomes (Delmont *et al.*, 2012) and may be a general feature of bacterial communities on surfaces with fluctuating (spatially, temporally) substrate conditions and potentially high cell densities.

Compared with prefilter metagenomes, Sterivex metagenomes were proportionally enriched in genes for substrate acquisition, energy and nutrient metabolism, and cell growth. Notably, genes encoding ATP-binding cassette transporters were at

significantly higher proportions, with enrichment driven by genes for branched chain amino-acid transporters (Figure 2), which accounted for 65% (average) of all sequences in this category and were two to threefold more abundant in the Sterivex fraction across depths (Supplementary Table S4). Consistent with this pattern, Sterivex metagenomes contained higher fractions of genes encoding the guanosine 5',3' bispyrophosphate (ppGpp)-controlled stringent response. Induced under diverse stress conditions, notably amino-acid starvation, the stringent response regulates a shift from growth-supporting functions (for example, stable RNA synthesis, translation and cell division) to those enabling survival under growth limitation, such as amino-acid biosynthesis and DNA replication (Cashel *et al.*, 1996; Durfee *et al.*, 2008; Traxler *et al.*, 2008). The free-living fraction was also overrepresented in genes for electron transport and energy generation, notably components of fermentation and respiration pathways. These included formate dehydrogenase, which in some bacteria is used for respiratory nitrate reduction (Sawers, 1994), as well as genes for assimilatory and respiratory nitrogen and sulfur metabolism (for example, ammonia assimilation, nitrate and nitrite ammonification; discussed in more detail below). Genes for biosynthesis and cell division were also at higher abundance, as were genes for CO<sub>2</sub> fixation, suggesting a propensity for autotrophic cells to be decoupled from organic-rich particles.

Together, these patterns indicate distinct microbial life history strategies. Free-living bacteria exhibit an overall greater investment in genes mediating core cellular functions and growth. These include adaptations for metabolic regulation under potential substrate limitation and a significant investment in mechanisms for the uptake of low-molecular weight compounds (that is, dissolved organic carbon), a pattern consistent with metagenomes from free-living bacteria at other ocean sites, including members of the SAR11 clade (Kirchman, 2003; Malmstrom *et al.*, 2005; Poretsky *et al.*, 2010), which were well represented in our OMZ samples. In contrast, prefilter-associated cells are more likely to encode functions for signal recognition and cell-to-cell interactions, presumably key adaptations for detecting and colonizing organic-rich particles and for life in close proximity to neighbors. Several of these trends are broadly consistent with functional genomic differences separating copiotrophic and oligotrophic life history strategies (Lauro *et al.*, 2009). These trends are detected here at the community-level across multiple depth zones. These patterns indicate that metagenome-based inferences about the relative importance of microbial traits in marine environments will vary depending on the ratio of particle-associated to free-living bacteria in a sample.

*Marker genes of nitrogen and sulfur metabolism.* Analysis of marker genes suggests that key OMZ metabolic processes may be partitioned between

particle-associated and free-living microbial communities. Results of BLASTX against the NCBI-nr database were manually queried to determine the relative abundances of target genes of nitrogen and sulfur energy metabolism, some of which are not well represented in the SEED hierarchy. Abundances are shown in Figure 3, normalized to gene length and to the abundance of a universal, single-copy gene (*rpoB*).

Genes of dissimilatory nitrogen oxidation (ammonia and nitrite oxidation) exhibited variable abundance but were generally overrepresented in the small size fraction (Figure 3, inset). Both *hzo* and *amoC*, markers for anammox and aerobic ammonia oxidation respectively, were enriched on average approximately fourfold in Sterivex metagenomes. *Hzo* sequences were detected only in the OMZ depths and were closely related to *hzo* of *Candidatus Scalindua*, consistent with the anaerobic nature of anammox and the overall distribution of *Scalindua*-matching protein-coding reads (Figure 3 and Supplementary Figure S7). Sequences matching *amoC* were affiliated exclusively with the Thaumarchaeota and Crenarchaeota, supporting recent evidence that nitrification in the OMZ is mediated primarily by Archaea (Stewart *et al.*, 2012b).

In contrast, the relative abundance of *nxB*, a marker for aerobic nitrite oxidation, did not vary substantially between size fractions (Figure 3, inset). However, fraction-specific patterns were evident when the *nxB* pool was evaluated according to taxonomic affiliation. The majority (55%) of all *nxB* sequences were most closely related to *nxB* of *Ca. Nitrospira defluvii*, a nitrite oxidizer isolated from activated sludge (Spieck *et al.*, 2006). *Nitrospira nxB* abundance peaked at the oxycline base (70 m) in both filter fractions. At the lower depths, including at the anoxic OMZ core, *Nitrospira nxB* was found exclusively in the free-living fraction (Figure 3 bottom). Subsequent to our analysis, the genome of the nitrite oxidizer *Nitrospira gracilis* was published (Lücker *et al.*, 2013). *Nitrospira* is the dominant nitrite-oxidizer genus in the oceans, and the *N. gracilis* genome is closely related evolutionarily to that of *Ca. Nitrospira defluvii* (Lücker *et al.*, 2013), raising the possibility that re-analysis of our data may instead classify OMZ *Nitrospira*-like sequences as belonging to *Nitrospira*. Both *Nitrospira* and *Ca. Nitrospira* genomes show adaptations to low-oxygen environments (Lücker *et al.*, 2010, 2013). This is consistent with the distribution of related sequences in the ETSP and other low-oxygen zones (Labrenz *et al.*, 2007; Jørgensen *et al.*, 2012), and with the recent detection of nitrite oxidation (by *Nitrospira* or *Nitrococcus* bacteria) under low oxygen ( $O_2 < 1 \mu M$ ) in the OMZ off Namibia (Fussel *et al.*, 2012). Together, these studies indicate a role for nitrite oxidation in the suboxic zones of the upper OMZ, potentially by diverse bacteria with distributions varying vertically and at the microscale.

Denitrification genes were also differentially partitioned between fractions. Sequences encoding *NarG*, catalyzing nitrate reduction to nitrite, were enriched in both fractions at OMZ depths compared with the oxic 1000 m sample, but were consistently, albeit marginally, overrepresented in the free-living fraction (Figure 3). This pattern was reversed at the oxic 1000 m depth, where *narG* was 13-fold more abundant in the prefilter compared with the Sterivex. A similar pattern was observed in metagenomes from the coastal North Pacific, where *narG* at a high oxygen site was enriched in larger size fractions (0.8–300  $\mu m$ ) compared with a 0.1–0.8  $\mu m$  fraction, but occurred at relatively uniform abundance across fractions at low-oxygen ( $\sim 20 \mu M$ ) sites (Smith *et al.*, 2013). These studies suggest that nitrate respiration in oxic water columns is confined primarily to suboxic microniches on particles, but under low-oxygen conditions is utilized by both free-living and surface-attached cells. In both filter fractions within the ETSP OMZ, a substantial proportion (42–50%) of the *narG* pool matched sequences from an uncultivated bacterium in candidate division OP1 isolated from a subsurface thermophilic microbial mat community (Takami *et al.*, 2012). Although relatives of the OP1 division have been detected in OMZs (Stevens and Ulloa, 2008; Wright *et al.*, 2012), their contributions to OMZ biogeochemistry remain uncharacterized.

Nitrite reductase genes (*nirK*, *nirS* and *nrfA*), involved in both denitrification and anammox, exhibited contrasting distributions between size fractions. In the OMZ depths, *nirK* genes encoding the copper-containing enzyme were distributed relatively evenly among filter fractions at all depths, excluding the 1000 m sample (Figure 3), whereas *nirS* genes for the cytochrome  $cd_1$ -containing nitrite reductase were most abundant on prefilters from the OMZ core (110, 200 m). In contrast, *nrfA* sequences, indicators of dissimilatory nitrate reduction to ammonium (Simon, 2002; Jensen *et al.*, 2011), were confined almost exclusively to the free-living size fraction and were affiliated predominantly with members of the Chlamydiae and Deltaproteobacteria.

Genes involved in the two terminal steps of denitrification, the reduction of nitric oxide to nitrous oxide (*norB/norZ*) and nitrous oxide to dinitrogen (*nosZ*), exhibited among the strongest size fraction-specific patterns, being on average fourfold more abundant on prefilters. Prefilter enrichment of these genes is consistent with recent metagenome data from an estuarine site (Smith *et al.*, 2013) and with studies linking  $N_2O$  and  $N_2$  production or *nor/nos* expression with diverse surface-attached environments (for example, sediments, soils, algal epibiont communities; Scala and Kerkhof, 1998; Rösch *et al.*, 2002; Long *et al.*, 2013), including suspended particles and the epibiotic communities of marine algae (Michotey and Bonin, 1997; Wyman *et al.*, 2013). Here, taxonomic partitioning between filter fractions was evident,



notably for the *norB/Z* pool. The majority (83–93%) of *norB/Z* sequences on Sterivex filters were most closely related to nitric oxide reductase from the anammox planctomycete *Candidatus Scalindua profunda* (scal02135). It has been hypothesized that in this bacterium NorB may act to relieve oxidative stress (van de Vossenberg *et al.*, 2012), as opposed to functioning in energy metabolism. In contrast, large proportions (47–71%) of particle-associated *norB/Z* sequences at the OMZ core (110 and 200 m) were most closely related to nitric oxide reductases of *Ca. Methyloirabilis oxyfera* (CBE69496.1 and CBE69502.1), a member of the NC10 candidate division originally enriched from sediment (Figure 3). *Ca. M. oxyfera* oxidizes methane under anaerobic conditions using O<sub>2</sub> generated intracellularly through an alternative denitrification pathway involving the dismutation of nitric oxide into dinitrogen and O<sub>2</sub>, potentially via a Nor enzyme acting as a dismutase (Raghoebarsing *et al.*, 2006; Ettwig *et al.*, 2010). Here, sequences matching *Ca. M. oxyfera* genes were recovered across all four depths (31–212 distinct genes per sample; 0.11–0.23% of total prokaryotic coding reads), raising the possibility that *Methyloirabilis*-like organisms may contribute to methane cycling in these waters.

Genes of dissimilatory sulfur metabolism also had variable distribution patterns. The gene (*aprA*) encoding APS reductase, which controls the AMP-dependent oxidation of sulfite to APS but also acts reversibly during sulfate reduction, was consistently enriched (twofold) in the free-living fraction. Other genes of sulfur metabolism did not show strong fraction-specific trends. These included the sulfur oxidation gene *soxB*, an indicator of thiosulfate oxidation, as well as the dissimilatory sulfite reductase gene *dsrA*, which is present in diverse sulfate reducers (notably Deltaproteobacteria), as well as in chemolithotrophic sulfide oxidizers (Dhillon *et al.*, 2005), in which Dsr operates in the reverse direction. Here, *aprA* and *dsrA* sequences matching known sulfate-reducing taxa (determined based on groupings by Canfield *et al.* (2010) constituted minor fractions of the total *aprA* and *dsrA* pools (3% and 0%, respectively) and did not exhibit clear size fraction-specific distributions, although such patterns could be obscured by the low representation of these sequences. The majority of *aprA* and *dsrA* sequences instead matched either known sulfur-oxidizing taxa, or groups for which the functional role of these genes is uncertain (for example, *aprA* in the Alphaproteobacterium *Pelagibacter*). However, taxonomic composition differed between sulfur metabolism genes. For example, sequences matching the SUP05 clade, containing free-living sulfur oxidizers from low-oxygen pelagic environments as well as thiotrophic deep-sea vent symbionts (see Gamma\_S in Figure 3; Walsh *et al.*, 2009), were relatively abundant among *soxB* sequences (49% of total), but made relatively minor

contributions to the *aprA* and *dsrA* pools (14% and 20%, respectively). These patterns suggest a complex sulfur-oxidizing community, with distinct pathways of sulfur oxidation mediated by different taxa.

Together, indicator gene distributions highlight the potential that key metabolic processes are partitioned spatially between free-living and surface-associated microbial communities. We anticipated that genes traditionally associated with autotrophic processes (for example, anammox, aerobic nitrification and sulfur oxidation) would be more prevalent in the free-living fraction, presuming that other metabolic substrates are not limiting in the open water column. In contrast, surface attachment would potentially benefit heterotrophic taxa (for example, heterotrophic denitrifiers) by providing a localized source of organic substrates. A subset of the data support this prediction (for example, *hzo*, *amoC*, *aprA*, *norB* and *nosZ*), whereas the distribution of other genes is less uniform. In one of the only studies to compare bacterioplankton metagenomes across size fractions (Smith *et al.*, 2013), genes of anaerobic metabolism, including *nar*, *nir*, *nor* and the reductive variant of *dsr*, were enriched in particle-associated cell fractions (relative to free-living fractions) at an oxic site, but not at hypoxic sites. These data, interpreted alongside our results, confirm that the distribution of metabolic functions between particle-associated and free-living niches varies among sites and is likely driven by the oxygen and substrate conditions of the surrounding water column, combined with the composition of the particles themselves.

*Mobile element genes.* Prefilter communities were significantly enriched in genes mediating mobile element activity via transposition and phage integration. Many of these genes were not recovered during functional analysis of BLASTX results using MEGAN, presumably due to limited representation in the SEED classification. However, manual parsing of significant BLASTX matches revealed that genes encoding transposases, which catalyze the movement of DNA segments (transposons) within genomes, were three to sixfold more abundant in prefilter compared with Sterivex metagenomes, with peak abundances at the OMZ core (110 and 200 m; Figure 3). At these depths, transposase genes were among the most abundant in the data sets, being up to 50-fold more abundant than the reference *rpoB* and constituting up to 2% of identifiable coding genes. By 1000 m, transposase abundance had declined by ~80% of peak values. A similar enrichment on prefilters and at OMZ depths was observed for genes encoding integrases that mediate site-specific recombination of phage DNA, typically via an RNA intermediate. This pattern is consistent with the observed prefilter enrichment of viral DNA (Table 1), as well as a 2.5-fold enrichment of restriction modification systems and restriction enzymes in prefilters at OMZ depths (70, 110 and

200 m; Supplementary Table S4), suggesting an increased need for protection against foreign DNA. The transposase and integrase gene pools matched database genes belonging to taxonomically diverse Bacteria and Archaea (Figure 3, bottom), with compositions resembling those inferred from all coding sequences and from 16S gene amplicons (Figure 1 and Supplementary Figure S5), notably with large contributions from groups abundant in both filter fractions (for example, Alpha- and Gammaproteobacteria, Bacteroidetes). The majority of prefilter transposable element genes therefore appear to originate from the bulk community and not disproportionately from those groups that appear most strongly overrepresented in this fraction (for example, Planctomycetes), suggesting that the observed enrichment is a community-wide phenomenon.

These results reinforce emerging evidence for the global significance of transposition in diverse microorganisms. In a survey of over 10 million annotated genes or gene fragments, transposases were determined to be the single most abundant protein-coding genes in sequenced genomes and environmental metagenomes, and to be ubiquitous across all domains of life (Aziz *et al.*, 2010). Notably, transposases constituted up to 8% of all sequences in a deep-sea hydrothermal vent biofilm (Brazelton and Baross, 2009) and were overrepresented in an algal-associated community compared with planktonic bacteria (Burke *et al.*, 2011). In addition, transposase gene transcription in oral spirochete bacteria was upregulated during growth as a biofilm (Mitchell *et al.*, 2010). These patterns, interpreted alongside our data, indicate a significant potential for transposase-mediated gene transfer in the genomes of surface-attached microbial communities.

It is unclear why transposases are enriched in surface-attached communities. Transposable element activity is often considered harmful for genomes, as transposition can inactivate genes or alter chromosome structure (Mahillon *et al.*, 1999; Kidwell and Lisch, 2001). Under this assumption, prefilter enrichment of transposases suggests a reduced capacity of particle-associated microbes to selectively purge mobile elements. This could occur in taxa with reduced effective population sizes, as has been hypothesized to explain an increase of transposase abundance with depth in free-living marine microbes (Konstantinidis *et al.*, 2009) and to partly explain mobile element accumulation in symbiont genomes (McCutcheon and Moran, 2011). As of yet, however, there is not strong evidence that particle-associated microorganisms have reduced population sizes compared with free-living taxa, although the prefilter fraction presumably is more likely to contain symbionts (for example, of marine protists). Here, the taxonomic composition of the transposase pool was broadly similar between filter fractions (Figure 3) and included groups both with and without strong size fraction-specific tendencies.

This raises the possibility that transposase enrichment on prefilters is independent of taxonomic identity and may instead be a consequence of the microenvironment itself. (This is similar to the enrichment of transposases as a function of depth in the water column, where deep-sea conditions seem to favor the prevalence of transposons—see below (DeLong *et al.*, 2006; Konstantinidis *et al.*, 2009)). Presumably, the spread of transposons and other mobile elements to new host genomes could be facilitated by growth in particle-associated biofilms where cells, relative to those in a planktonic form, occur in close spatial proximity and have an increased likelihood of encountering foreign DNA (for example, via conjugation or via phage or naked DNA caught in the biofilm matrix; Madsen *et al.*, 2012). In this scenario, the ubiquity of transposases in prefilter communities would not preclude the possibility that transposition is selectively deleterious, and may instead simply reflect an increased rate of mobile element transfer among and within genomes (Werren, 2011).

Alternatively, transposition may be beneficial for particle-associated microbes if it generates genetic variants that enhance host fitness, or provides raw genetic material for evolving new functions (Aertsen and Michiels, 2005; Chou *et al.*, 2009). Marine particles can contain steep gradients in nutrient or redox substrate availability, locally high concentrations of bacteriocidal chemicals produced by neighboring microbes, and increased incidences of antagonism (Alldredge and Cohen, 1987; Long and Azam, 2001; Long *et al.*, 2005), consistent with the enrichment of virulence, motility and chemotaxis genes observed in OMZ prefilter metagenomes (Figure 2). Unlike free-living bacterial cells, particles may sink through the water column, exposing attached microbes to changing physical and biological conditions along the depth gradient. Zooplankton feeding on particles also has the potential to rapidly shift the local environment for particle-associated microbes. Such spatial and temporal variability may select for genomes with an elevated potential for acquiring new functions via lateral gene movement. Indeed, transposition in bacteria has been shown to be induced by environmental change (Ohtsubo *et al.*, 2005; Twiss *et al.*, 2005) and to facilitate adaptive responses to repeated cycles of growth and stress (Sleight *et al.*, 2008).

The vertical distribution of transposase genes in the OMZ, while limited to only four depths, suggests interesting contrasts to patterns observed in other oceanic regions. Metagenome analysis of the free-living microbial fraction revealed a steady increase in the relative abundance of transposase genes with depth down to 4000 m in the oligotrophic North Pacific off Hawaii (DeLong *et al.*, 2006; Konstantinidis *et al.*, 2009). An increase in both relative transposase abundance and transcription, similar in magnitude to that recorded off Hawaii over a 4000 m gradient, has been observed

previously in the free-living microbial community of the ETSP OMZ (Stewart *et al.*, 2012b), but over a much narrower depth range. Indeed, the latter study sampled to a depth of only 200 m (OMZ core). In the current study, transposase abundance in both filter fractions spiked within the OMZ, but then declined substantially by 1000 m. Given the patterns we observe in our size fraction analysis, transposase abundance in the free-living fraction may be influenced by the relative load of particulate material in the water column (and therefore the proportion of the total community that is surface-attached). This may be possible if sampling causes the detachment of particle-associated cells, which are then sampled as part of the free-living fraction. Indeed, the vertical distribution of transposase abundance in ETSP metagenomes roughly parallels that of particulate backscattering, with peaks in the upper oxycline and suboxic core and a decline with depth out of the OMZ (Van Mooy *et al.*, 2002; Whitmire *et al.*, 2009; Supplementary Figure S1).

The enrichment of transposable element genes within OMZ metagenomes may coincide with distributions of ETSP OMZ viruses (Table 1). Cassman *et al.* (2012) described atypically low ratios of free-living (non-surface-attached) viruses to microbes in the Chilean OMZ, compared with surface waters and to other marine regions. The authors hypothesized a transition from lytic to lysogenic viral lifestyles within the anoxic zone, potentially linked to low microbial division rates. An enhanced role for lysogeny in the OMZ may be consistent with the vertical distribution of transposases and integrases in the particle fraction, notably if these genes originate from prophage. Alternatively, it is possible that viral and mobile element genes detected in the prefilter metagenomes originated from lysed viral particles entrained within surface-associated biofilms. Collectively, our data indicate an enhanced potential for mobile element activity in surface-associated OMZ bacterioplankton—it remains to be clarified how this potential is linked to phage activity on particles.

## Conclusions

Metagenomic comparisons of different bacterioplankton size fractions are (surprisingly) rare but are necessary for clarifying the distribution of functional traits in marine microbes. OMZ microbial communities from the particle-associated size fraction were consistently overrepresented in functions suggesting adaptation to surface attachment and growth in a biofilm, notably genes mediating colonization and cell–cell interactions. This pattern highlights fundamental divisions between particle-associated and free-living life history modes and correlates with a strong partitioning of taxonomic groups between fractions. Indeed, association with prefilters was a stronger predictor of community

structure than depth-specific environmental variation (Figure 1). This was unexpected given studies showing the strong influence of vertical environmental gradients (for example, oxygen, nutrient availability) in shaping OMZ community structure (Bryant *et al.*, 2012), but agrees with taxon-specific (for example, *Vibrio*) analyses identifying particle-association as a major driver of ecological niche diversification in marine bacteria (Hunt *et al.*, 2008). However, in other ocean regions, microbial biomass per unit volume has been shown to be substantially higher for the smallest size fraction (Rosso and Azam, 1987; Cho and Azam, 1988; Karl *et al.*, 1988), suggesting that analysis of the bulk OMZ community (all size fractions combined) may show patterns of community clustering resembling those observed for the small size fraction. Measurements of live cell abundance and activity among different size fractions in the OMZ are clearly needed.

Particle-associated microbial communities also were markedly enriched in functions controlling genetic exchange. This is consistent with experimental analyses of cultured bacteria showing that life on surfaces (in biofilms) is an important route for diversification (Boles *et al.*, 2004; Boles and Singh, 2008). Although the adaptive value for biofilm-mediated diversification is unclear, our data suggest that particle-associated communities may be hotbeds for genome reshuffling, with a strong role for horizontal gene transfer via transposition. Future studies should test the hypothesis that rates of genetic exchange differ between particle-associated and free-living marine microbes.

Key metabolic pathways mediating OMZ elemental cycling also were differentially partitioned between size fractions. This implies that particle abundance, and by extension the percentage of the total microbial community attached to surfaces, may have an important impact on community biogeochemical transformations. For example, genes facilitating the last two steps of denitrification were markedly enriched on particles, raising the possibility that bulk denitrification rates could be correlated with particle load. However, the relative abundance of genes and taxa is an imperfect predictor of functional significance. For example, Deltaproteobacteria of the genus *Desulfosporosinus* constituted 0.006% of microbial cell counts in a peatland ecosystem but were responsible for the majority of sulfate reduction in the community (Pester *et al.*, 2010). Such results highlight the need to directly couple genetic and process rate characterizations across multiple size fractions to determine how the activity of particle-associated communities contributes to OMZ biochemical cycling.

Resolving questions about the role of particle-associated microbes in OMZs should be a research priority, but will require standardizations of sampling design and methodology. As our study focused on a single site, conclusions about partitioning between fractions are most robust for genes and taxa

with consistent patterns across depths. However, these data, and data from other recent metagenome studies (Smith *et al.*, 2013), also suggest that the level of partitioning varies with depth and environmental (for example, oxygen) conditions—additional vertical profiles from diverse low-oxygen sites will be required to statistically identify such trends. The potential for sampling to physically disrupt particles should also be considered. It is possible that some of the taxa identified in the free-living fraction originate from particles that are broken apart during water collection (for example, during rosette bottle-firing, or pumping during the filtration step). If this occurs, inferences of community metabolism in the free-living fraction may be especially prone to bias in regions of high particle load. There is also a critical need to standardize sample treatment between process rate measurements and genetic characterizations. Rate measurements are often based on incubations (for example, in exetainers) of the bulk water (for example, Dalsgaard *et al.*, 2012), whereas the majority of OMZ molecular studies have focused on microbes retained after prefiltration (for example, Stevens and Ulloa, 2008; Stewart *et al.*, 2012a,b). Furthermore, different studies use different filter pore-size cutoffs, making direct comparisons challenging. As marine particles differ significantly in age, size, organic substrate contents and redox state, conclusions about particle-associated communities should involve uniform comparisons across multiple particle size fractions. These comparisons should be directly coupled to measurements of bulk particle load and to gene expression and process rate characterizations in order to fully understand how particle-associated communities contribute to OMZ biochemical cycling.

## Conflict of Interest

The authors declare no conflict of interest.

## Acknowledgements

We thank Jess Bryant, Adrian Sharma, Jaime Becker and the captain and crew of the *R/V Melville* for help in sample collection, and Tsultrim Palden and Raghav Sharma for help in DNA library preparation and pyrosequencing. This work is a contribution of the Center for Microbial Oceanography: Research and Education (C-MORE) and was made possible by generous support from the National Science Foundation (1151698 to FJS and EF0424599 to EFD), the Gordon and Betty Moore Foundation (EFD), and the Agouron Institute (EFD).

## References

Acinas SG, Anton J, Rodriguez-Valera F. (1999). Diversity of free-living and attached bacteria in offshore western Mediterranean waters as depicted by analysis

of genes encoding 16S rRNA. *Appl Environ Microb* **65**: 514–522.

- Aertsen A, Michiels CW. (2005). Diversify or die: generation of diversity in response to stress. *Crit Rev Microbiol* **31**: 69–78.
- Allredge AL, Cohen Y. (1987). Can microscale chemical patches persist in the sea? Microelectrode study of marine snow, fecal pellets. *Science* **235**: 689–691.
- Allredge AL, Silver MW. (1988). Characteristics, dynamics and significance of marine snow. *Prog Oceanogr* **20**: 41–82.
- Allen LZ, Allen EE, Badger JH, McCrow JP, Paulsen IT, Elbourne LDH *et al.* (2012). Influence of nutrients and currents on the genomic composition of microbes across an upwelling mosaic. *ISME J* **6**: 1403–1414.
- Aziz RK, Breitbart M, Edwards RA. (2010). Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Res* **38**: 4207–4217.
- Bano N, DeRae Smith A, Bennett W, Vasquez L, Hollibaugh JT. (2007). Dominance of mycoplasma in the guts of the long-jawed mudsucker, *Gillichthys mirabilis*, from five California salt marshes. *Environ Microbiol* **9**: 2636–2641.
- Bauld J, Staley JT. (1976). *Planctomyces maris* sp. nov., marine isolate of *Planctomyces blastocaulis* group of budding bacteria. *J R Gen Microbiol* **97**: 45–55.
- Bengtsson MM, Øvreås L. (2010). Planctomycetes dominate biofilms on surfaces of the kelp *Laminaria hyperborea*. *BMC Microbiol* **10**: 261.
- Benjamini Y, Hochberg Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* **57**: 289–300.
- Boles BB, Singh PK. (2008). Endogenous oxidative stress produces diversity and adaptability in biofilm communities. *Proc Natl Acad Sci USA* **105**: 12503–12508.
- Boles BR, Thoendel M, Singh PK. (2004). Self-generated diversity produces “insurance effects” in biofilm communities. *Proc Natl Acad Sci USA* **101**: 16630–16635.
- Brazelton WJ, Baross JA. (2009). Abundant transposases encoded by the metagenome of a hydrothermal chimney biofilm. *ISME J* **3**: 1420–1424.
- Brinkhoff T, Fisher D, Vollmers J, Voget S, Beardsley C, Thole S *et al.* (2012). Biogeography and phylogenetic diversity of a cluster of exclusively marine myxobacteria. *ISME J* **6**: 1260–1272.
- Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P. (2008). Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol* **6**: 245–252.
- Brown MV, Philip GK, Bunge JA, Smith MC, Bissett A, Lauro FM *et al.* (2009). Microbial community structure in the North Pacific Ocean. *ISME J* **3**: 1374–1386.
- Bryant JA, Stewart FJ, Eppley JM, DeLong EF. (2012). Microbial community phylogenetic and trait diversity decline steeply with depth in a marine oxygen minimum zone. *Ecology* **93**: 1659–1673.
- Bull AT, Stach JEM, Ward AC, Goodfellow M. (2005). Marine actinobacteria: perspectives, challenges, future directions. *Anton Van Lee J M S* **87**: 65–79.
- Burke C, Steinberg P, Rusch D, Kjelleberg S, Thomas T. (2011). Bacterial community assembly based on functional genes rather than species. *Proc Natl Acad Sci USA* **108**: 14288–14293.

- Burrows LL. (2012). *Pseudomonas aeruginosa* twitching motility: type IV pili in action. *Annu Rev Microbiol* **66**: 493–520.
- Canfield DE, Stewart FJ, Thamdrup B, De Brabandere L, Dalsgaard T, DeLong EF *et al.* (2010). A cryptic sulfur cycle in oxygen-minimum-zone waters off the Chilean Coast. *Science* **330**: 1375–1378.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.
- Caron DA, Davis PG, Madin LP, Sieburth JM. (1982). Heterotrophic bacteria and bacterivorous protozoa in oceanic macroaggregates. *Science* **218**: 795–797.
- Cashel M, Gentry D, Hernandez VJ, Vinella D. (1996). The stringent response. In: Neidhardt FC (eds). *Escherichia coli and Salmonella: cellular and molecular biology*. American Society for Microbiology Press: Washington, DC, USA, pp 1458–1496.
- Cassman N, Prieto-Davó A, Walsh K, Silva GG, Angly F, Akhter S *et al.* (2012). Oxygen minimum zones harbour novel viral communities with low diversity. *Environ Microbiol* **14**: 3043–3065.
- Cho BC, Azam F. (1988). Major role of bacteria in biogeochemical fluxes in the ocean's interior. *Nature* **332**: 441–443.
- Chou HH, Berthet J, Marx CJ. (2009). Fast growth increases the selective advantage of a mutation arising recurrently during evolution under metal limitation. *PLoS Genet* **5**: e1000652.
- Christie PJ, Atmakuri K, Krishnamoorthy V, Jakubowski S, Cascales E. (2005). Biogenesis, architecture, and function of bacterial type IV secretion systems. *Annu Rev Microbiol* **59**: 451–485.
- Cottrell MT, Kirchman DL. (2000). Natural assemblages of marine proteobacteria and members of the *Cytophaga-Flavobacter* cluster consuming low- and high-molecular-weight dissolved organic matter. *Appl Environ Microb* **66**: 1692–1697.
- Crump BC, Armbrust EV, Baross JA. (1999). Phylogenetic analysis of particle-attached and free-living bacterial communities in the Columbia river, its estuary, and the adjacent coastal ocean. *Appl Environ Microb* **65**: 3192–3204.
- Crump BC, Baross JA. (2000). Archaeoplankton in the Columbia River, its estuary and the adjacent coastal ocean, USA. *FEMS Microbiol Ecol* **31**: 231–239.
- Dalisay DS, Webb JS, Scheffel A, Svenson C, James S, Holmström C *et al.* (2006). A mannose-sensitive haemagglutinin (MSHA)-like pilus promotes attachment of *Pseudoalteromonas tunicata* cells to the surface of the green alga *Ulva australis*. *Microbiology* **152**: 2875–2883.
- Dalsgaard T, Thamdrup B, Farias L, Revsbech NP. (2012). Anammox and denitrification in the oxygen minimum zone of the eastern South Pacific. *Limnol Oceanogr* **57**: 1331–1346.
- Delmont TO, Prestat E, Keegan KP, Faubladiere M, Robe P, Clark IM *et al.* (2012). Structure, fluctuation and magnitude of a natural grassland soil metagenome. *ISME J* **6**: 1677–1687.
- DeLong EF, Franks DG, Alldredge AL. (1993). Phylogenetic diversity of aggregate-attached vs free-living marine bacterial assemblages. *Limnol Oceanogr* **38**: 924–934.
- DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU *et al.* (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496–503.
- DeLong EF. (2003). Oceans of archaea. *ASM News* **69**: 503–511.
- Demiri A, Meziti A, Papaspyrou S, Thessalou-Legaki M, Kormas K. (2009). Abdominal setae and midgut Bacteria of the mudshrimp *Pestarella tyrrhena*. *Cent Eur J Biol* **4**: 558–566.
- Dhillon A, Goswami S, Riley M, Teske A, Sogin M. (2005). Domain evolution and functional diversification of sulfite reductases. *Astrobiology* **5**: 18–29.
- Ditchfield AK, Wilson ST, Hart MC, Purdy KJ, Green DH, Hatton AD. (2012). Identification of putative methylotrophic and hydrogenotrophic methanogens within sedimenting material and copepod faecal pellets. *Aquat Microb Ecol* **67**: 151–160.
- Durfee T, Hansen AM, Zhi H, Blattner FR, Jin DJ. (2008). Transcription profiling of the stringent response in *Escherichia coli*. *J Bacteriol* **190**: 1084–1096.
- Edgcomb V, Leadbetter ER, Bourland W, Beaudoin D, Bernhard JM. (2011). Structured multiple endosymbiosis of bacteria and archaea in a ciliate from marine sulfidic sediments: a survival mechanism in low oxygen, sulfidic sediments. *Front Microbiol* **2**: 55.
- Eloe EA, Shulse CN, Fadrosch DW, Williamson SJ, Allen EE, Bartlett DH. (2010). Compositional differences in particle-associated and free-living microbial assemblages from an extreme deep-ocean environment. *Environ Microbiol Rep* **3**: 449–458.
- Ettwig KF, Butler MK, Le Paslier D, Pelletier E, Mangenot S, Kuypers MM *et al.* (2010). Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* **464**: 543–548.
- Faith DP. (1992). Conservation evaluation and phylogenetic diversity. *Biol Conserv* **61**: 1–10.
- Fenchel T. (2001). Eppu si muove: many water column bacteria are motile. *Aquat Microb Ecol* **24**: 197–201.
- Fenchel T. (2002). Microbial behavior in a heterogeneous world. *Science* **296**: 1068–1071.
- Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW *et al.* (2008). Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci USA* **105**: 3805–3810.
- Fuchs BM, Woeckel D, Zubkov MV, Burkill P, Amann R. (2005). Molecular identification of picoplankton populations in contrasting waters of the Arabian Sea. *Aquat Microb Ecol* **39**: 145–157.
- Fuchsman CA, Kirkpatrick JB, Brazelton WJ, Murray JW, Staley JT. (2011). Metabolic strategies of free-living and aggregate-associated bacterial communities inferred from biologic and chemical profiles in the Black Sea suboxic zone. *FEMS Microbiol Ecol* **78**: 586–603.
- Fuchsman CA, Staley JT, Oakley BB, Kirkpatrick JB, Murray JW. (2012). Free-living and aggregate-associated Planctomycetes in the Black Sea. *FEMS Microbiol Ecol* **80**: 402–416.
- Fussel J, Lam P, Lavik G, Jensen MM, Holtappels M, Gunter M *et al.* (2012). Nitrite oxidation in the Namibian oxygen minimum zone. *ISME J* **6**: 1200–1209.
- Galan A, Molina V, Belmar L, Ulloa O. (2012). Temporal variability and phylogenetic characterization of planktonic anammox bacteria in the coastal upwelling ecosystem off central Chile. *Prog Oceanogr* **92–95**: 110–120.

- Galand PE, Lovejoy C, Pouliot J, Vincent WF. (2008). Heterogeneous archaeal communities in the particle-rich environment of an arctic shelf ecosystem. *J Marine Syst* **74**: 774–782.
- Garfield PC, Packard TT, Friedrich GE, Codispoti LA. (1983). A subsurface particle maximum layer and enhanced microbial activity in the secondary nitrite maximum of the northeastern tropical Pacific Ocean. *J Mar Res* **41**: 747–768.
- Ghiglione JF, Mevel G, Pujo-Pay M, Mousseau L, Lebaron P, Goutx M. (2007). Diel and seasonal variation in abundance, activity, and community structure of particle-attached and free-living bacteria in NW Mediterranean Sea. *Microb Ecol* **54**: 217–231.
- Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D *et al.* (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**: 1242–1245.
- Gram L, Melchiorson J, Bruhn JB. (2010). Antibacterial activity of marine culturable bacteria collected from a global sampling of ocean surface waters and surface swabs of marine organisms. *Mari Biotechnol* **12**: 439–451.
- Grossart HP, Hietanen S, Ploug H. (2003). Microbial dynamics on diatom aggregates in Øresund, Denmark. *Mar Ecol Prog Ser* **249**: 69–78.
- Grossart HP, Kjørboe T, Tang KW, Allgaier M, Yam EM, Ploug H. (2006). Interactions between marine snow and heterotrophic bacteria: aggregate formation, bacterial activities and phylogenetic composition. *Aquat Microb Ecol* **42**: 19–26.
- Grossart HP, Schlingloff A, Bernhard M, Simon M, Brinkhoff T. (2004). Antagonistic activity of bacteria isolated from organic aggregates of the German Wadden Sea. *FEMS Microbiol Ecol* **47**: 387–396.
- Grossart HP, Tang KW, Kjørboe T, Ploug H. (2007). Comparison of cell-specific activity between free-living and attached bacteria using isolates and natural assemblages. *FEMS Microbiol Lett* **266**: 194–200.
- Grossart HP. (2010). Ecological consequences of bacterioplankton lifestyles: changes in concepts are needed. *Environ Microbiol Rep* **2**: 706–714.
- Hardcastle TJ, Kelly KA. (2010). BaySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics* **11**: 422.
- Hollibaugh T, Wong PS, Murrell MC. (2000). Similarity of particle-associated and free-living bacterial communities in northern San Francisco Bay, California. *Aquat Microb Ecol* **21**: 102–114.
- Huber JA, Mark Welch DB, Morrison HG, Huse SM, Neal PR, Butterfield DA *et al.* (2007). Microbial population structures in the deep marine biosphere. *Science* **318**: 97–100.
- Hunt DE, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF. (2008). Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* **320**: 1081–1085.
- Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC. (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome Res* **21**: 1552–1560.
- Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV. (2012). Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science* **335**: 587–590.
- Jensen PR, Lauro FM. (2008). An assessment of actinobacterial diversity in the marine environment. *Anton Van Lee J M S* **94**: 51–62.
- Jensen MM, Lam P, Revsbech NP, Nagel B, Gaye B, Jetten MS *et al.* (2011). Intensive nitrogen loss over the Omani Shelf due to anammox coupled with dissimilatory nitrite reduction to ammonium. *ISME J* **5**: 1660–1670.
- Jiang DM, Kato C, Zhou XW, Wu ZH, Sato T, Li YZ. (2010). Phylogeographic separation of marine and soil myxobacteria at high levels of classification. *ISME J* **4**: 1520–1530.
- Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EMS, Chisholm SW. (2006). Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311**: 1737–1740.
- Jorgensen SL, Hannisdal B, Lanzen A, Baumberg T, Flesland K, Fonseca R *et al.* (2012). Correlating microbial community profiles with geochemical data in highly stratified sediments from the Arctic Mid-Ocean Ridge. *Proc Natl Acad Sci USA* **109**: E2846–E2855.
- Jumpstart Consortium Human Microbiome Project Data Generation Working Group (2012). Evaluation of 16S rDNA-based community profiling for human microbiome research. *PLoS ONE* **7**: e39315.
- Karl DM, Knauer GA, Martin JH, Ward BB. (1984). Bacterial chemolithotrophy in the ocean is associated with sinking particles. *Nature* **309**: 54–56.
- Karl DM, Knauer GA, Martin JH. (1988). Downward flux of particulate organic matter in the ocean: a particle decomposition paradox. *Nature* **332**: 438–441.
- Karner M, Herndl G. (1992). Extracellular enzymatic activity and secondary production in free-living and marine-snow-associated bacteria. *Mar Biol* **113**: 341–347.
- Karstensen J, Stramma L, Visbeck M. (2008). Oxygen minimum zones in the eastern tropical Atlantic and Pacific oceans. *Prog Oceanogr* **77**: 331–350.
- Kellogg CTE, Deming JW. (2009). Comparison of free-living, suspended particle, and aggregate-associated bacterial and archaeal communities in the Laptev Sea. *Aquat Microb Ecol* **57**: 1–18.
- Kembel SW, Eisen JA, Pollard KS, Green JL. (2011). The phylogenetic diversity of metagenomes. *PLoS ONE* **6**: e23214.
- Kidwell MG, Lisch DR. (2001). Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution* **55**: 1–24.
- Kirchman DL. (2003). The contribution of monomers and other low-molecular weight compounds to the flux of dissolved organic material in aquatic ecosystems. In: Findlay SEG, Sinsabaugh RL (eds). *Aquatic Ecosystems: Interactivity of Dissolved Organic Matter*. Academic Press: San Diego, CA, USA, pp 218–237.
- Kirchman D, Mitchell R. (1982). Contribution of particle-bound bacteria to total microheterotrophic activity in five ponds and two marshes. *Appl Environ Microb* **43**: 200–209.
- Konstantinidis KT, DeLong EF. (2008). Genomic patterns of recombination, clonal divergence and environment in marine microbial populations. *ISME J* **2**: 1052–1065.
- Konstantinidis KT, Braff J, Karl DM, DeLong EF. (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. *Appl Environ Microb* **75**: 5345–5355.
- Labrenz M, Jost G, Jurgens K. (2007). Distribution of abundant prokaryotic organisms in the water column of the central Baltic Sea with an oxic-anoxic interface. *Aquat Microb Ecol* **46**: 177–190.
- Lam P, Lavik G, Jensen MM, van de Vossenberg J, Schmid M, Woebken D *et al.* (2009). Revising the nitrogen cycle in the Peruvian oxygen minimum zone. *Proc Natl Acad Sci USA* **106**: 4752–4757.
- LaMontagne MG, Holden PA. (2003). Comparison of free-living and particle-associated bacterial communities in a coastal lagoon. *Microb Ecol* **46**: 228–237.
- Lapoussiere A, Michel C, Starr M, Gosselin M, Poulin M. (2011). Role of free-living and particle-attached bacteria in the recycling and export of organic material in the Hudson Bay system. *J Marine Syst* **88**: 434–445.
- Lauro FM, McDougald D, Thomas T, Williams TJ, Egan S, Rice S *et al.* (2009). The genomic basis of trophic strategy in marine bacteria. *Proc Natl Acad Sci USA* **106**: 15527–15533.
- Lavik G, Stuhmann T, Bruchert V, Van der Plas A, Mohrholz V, Lam P *et al.* (2009). Detoxification of sulphidic African shelf waters by blooming chemolithotrophs. *Nature* **457**: 581–586.
- Li WZ, Godzik A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659.
- Long A, Heitman J, Tobias C, Philips R, Song B. (2013). Co-occurring anammox, denitrification, and codenitrification in agricultural soils. *Appl Environ Microb* **79**: 168–176.
- Long R, Azam F. (2001). Antagonistic interactions among marine pelagic bacteria. *Appl Environ Microbiol* **67**: 4975–4983.
- Long RA, Rowley DC, Zamora E, Liu JY, Bartlett DH, Azam F. (2005). Antagonistic interactions among marine bacteria impede the proliferation of *Vibrio cholerae*. *Appl Environ Microb* **71**: 8531–8536.
- Lozupone C, Knight R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microb* **71**: 8228–8235.
- Lücker S, Nowka B, Rattei T, Spieck E, Daims H. (2013). The genome of *Nitrospina gracilis* illuminates the metabolism and evolution of the major marine nitrite oxidizer. *Front Microbiol* **4**: 27.
- Lücker S, Wagner M, Maixner F, Pelletier E, Koch H, Vacherie B *et al.* (2010). A *Nitrospira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proc Natl Acad Sci USA* **107**: 12479–13484.
- Madsen JS, Burmølle M, Hansen LH, Sørensen SJ. (2012). The interconnection between biofilm formation and horizontal gene transfer. *FEMS Immunol Med Mic* **65**: 183–195.
- Mahillon J, Léonard C, Chandler M. (1999). IS elements as constituents of bacterial genomes. *Res Microbiol* **150**: 675–687.
- Malmstrom RR, Cottrell MT, Elifantz H, Kirchman DL. (2005). Biomass production and assimilation of dissolved organic matter by SAR11 bacteria in the Northwest Atlantic Ocean. *Appl Environ Microb* **71**: 2979–2986.
- McCutcheon JP, Moran NA. (2011). Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* **10**: 13–26.
- Michotey V, Bonin P. (1997). Evidence for anaerobic bacterial processes in the water column: denitrification and dissimilatory nitrate ammonification in the northwestern Mediterranean Sea. *Mar Ecol Prog Ser* **160**: 47–56.
- Mitchell HL, Dashper SG, Catmull DV, Paolini RA, Cleal SM, Slakeski N *et al.* (2010). *Treponema denticola* biofilm-induced expression of a bacteriophage, toxin-antitoxin systems and transposases. *Microbiology* **156**: 774–788.
- Moeseneder MM, Winter C, Herndl GJ. (2001). Horizontal and vertical complexity of attached and free-living bacteria of the eastern Mediterranean Sea, determined by 16S rDNA and 16S rRNA fingerprints. *Limnol Oceanogr* **46**: 95–107.
- Morris RM, Longnecker K, Giovannoni SJ. (2006). Pirellula and OM43 are among the dominant lineages identified in an Oregon coast diatom bloom. *Environ Microbiol* **8**: 1361–1370.
- Munn C. (2011). *Marine Microbiology: Ecology and Applications*, 2nd edn. Garland Science: New York, NY.
- Naqvi SWA, Kumar MD, Narvekar PV, De Sousa SN, George MD, Silva CD. (1993). An intermediate nepheloid layer associated with high microbial rates and denitrification in the northwest Indian Ocean. *J Geophysical Res* **98**: 16469–16479.
- Not F, del Campo J, Balagué V, de Vargas C, Massana R. (2009). New insights into the diversity of marine picoeukaryotes. *PLoS One* **4**: e7143.
- Nowack ECM, Melkonian M. (2010). Endosymbiotic associations within protists. *Philos T Roy Soc B* **365**: 699–712.
- Ohtsubo Y, Genka H, Komatsu H, Nagata Y, Tsuda M. (2005). High-temperature-induced transposition of insertion elements in *Burkholderia multivorans* ATCC 17616. *Appl Environ Microbiol* **71**: 1822–1828.
- Orsi W, Song YC, Hallam S, Edgcomb V. (2012). Effect of oxygen minimum zone formation on communities of marine protists. *ISME J* **6**: 1586–1601.
- Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M *et al.* (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33**: 5691–5702.
- Pak H, Codispoti LA, Zaneveld JRV. (1980). On the intermediate particle maxima associated with oxygen-poor water off Western South America. *Deep Sea Res* **27**: 783–797.
- Parveen B, Reveilliez JP, Mary I, Ravet V, Bronner G, Mangot JF *et al.* (2011). Diversity and dynamics of free-living and particle-associated Betaproteobacteria and Actinobacteria in relation to phytoplankton and zooplankton communities. *FEMS Microbiol Ecol* **77**: 461–476.
- Pester M, Bittner N, Deevong P, Wagner M, Loy A. (2010). A ‘rare biosphere’ microorganism contributes to sulfate reduction in a peatland. *ISME J* **4**: 1591–1602.
- Pham VD, Konstantinidis KT, Palden T, DeLong EF. (2008). Phylogenomic assessment of bacterioplankton distribution in a 4000 metre vertical profile of the North Pacific Subtropical Gyre. *Environ Microbiol* **10**: 2313–2330.

- Ploug H, Grossart HP, Azam F, Jorgensen BB. (1999). Photosynthesis, respiration, and carbon turnover in sinking marine snow from surface waters of Southern California Bight: implications for the carbon cycle in the ocean. *Mar Ecol Prog Ser* **179**: 1–11.
- Poretsky RS, Sun S, Mou X, Moran MA. (2010). Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. *Environ Microbiol* **12**: 616–627.
- Prieto-Davo A, Fenical W, Jensen PR. (2008). Comparative actinomycete diversity in marine sediments. *Aquat Microb Ecol* **52**: 1–11.
- Qian PY, Wang Y, Lee OO, Lau SCK, Yang J, Lafi FF *et al*. (2011). Vertical stratification of microbial communities in the Red Sea revealed by 16S rDNA pyrosequencing. *ISME J* **5**: 507–518.
- Raghoebarsing AA, Pol A, van de Pas-Schoonen KT, Smolders AJ, Ettwig KF, Rijpstra WI *et al*. (2006). A microbial consortium couples anaerobic methane oxidation to denitrification. *Nature* **440**: 918–921.
- Rappe MS, Gordon DA, Vergin K, Giovannoni SJ. (1999). Phylogeny of *Actinobacteria* small subunit rRNA (SSU rRNA) gene clones recovered from diverse sea water samples. *Syst Appl Microbiol* **22**: 106–112.
- Rath J, Wu KY, Herndl GJ, DeLong EF. (1998). High phylogenetic diversity in a marine-snow-associated bacterial assemblage. *Aquat Microb Ecol* **14**: 261–269.
- Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA *et al*. (2003). Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**: 1042–1047.
- Rösch C, Mergel A, Bothe H. (2002). Biodiversity of denitrifying and dinitrogen-fixing bacteria in an acid forest soil. *Appl Environ Microb* **68**: 3818–3829.
- Rosso AL, Azam F. (1987). Proteolytic activity in coastal oceanic waters: depth distribution and relationship to bacterial populations. *Mar Ecol Prog Ser* **41**: 231–240.
- Ruehland C, Blazejak A, Lott C, Loy A, Erséus C, Dubilier N. (2008). Multiple bacterial symbionts in two species of co-occurring gutless oligochaete worms from Mediterranean sea grass sediments. *Environ Microbiol* **10**: 3404–3416.
- Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooshep S *et al*. (2007). The sorcerer II global ocean sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5**: e77.
- Sawers G. (1994). The hydrogenases and formate dehydrogenases of *Escherichia coli*. *Anton Van Lee J M S* **66**: 57–88.
- Scala DJ, Kerkhof LJ. (1998). Nitrous oxide reductase (*nosZ*) gene-specific PCR primers for detection of denitrifiers and three *nosZ* genes from marine sediments. *FEMS Microbiol Lett* **162**: 61–68.
- Shanks AL, Reeder ML. (1993). Reducing microzones and sulfide production in marine snow. *Mar Ecol Prog Ser* **96**: 43–47.
- Shimkets LJ, Dworkin M, Reichenbach H. (2006). The myxobacteria. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (eds). *The Prokaryotes*, 3rd edn. Springer: Heidelberg, Germany, pp 31–115.
- Simon J. (2002). Enzymology and bioenergetics of respiratory nitrate ammonification. *FEMS Microbiol Rev* **26**: 285–309.
- Simon M, Grossart HP, Schweitzer B, Ploug H. (2002). Microbial ecology of organic aggregates in aquatic ecosystems. *Aquat Microb Ecol* **28**: 175–211.
- Sleight SC, Orlic C, Schneider D, Lenski RE. (2008). Genetic basis of evolutionary adaptation by *Escherichia coli* to stressful cycles of freezing, thawing and growth. *Genetics* **180**: 431–443.
- Smith DC, Simon M, Alldredge AL, Azam F. (1992). Intense hydrolytic enzyme activity on marine aggregates and implications for rapid particle dissolution. *Nature* **359**: 139–142.
- Smith MW, Allen LZ, Allen AE, Herfort L, Simon HM. (2013). Contrasting genomic properties of free-living and particle-attached microbial assemblages within a coastal ecosystem. *Front Microbiol* **4**: 120.
- Spieck E, Hartwig C, McCormack I, Maixner F, Wagner M, Lipski A *et al*. (2006). Selective enrichment and molecular characterization of a previously uncultured *Nitrospira*-like bacterium from activated sludge. *Environ Microbiol* **8**: 405–415.
- Stevens H, Ulloa O. (2008). Bacterial diversity in the oxygen minimum zone of the eastern tropical South Pacific. *Environ Microbiol* **10**: 1244–1259.
- Stewart FJ, Dalsgaard T, Thamdrup B, Revsbech NP, Ulloa O, Canfield DE *et al*. (2012a). Experimental perturbation and oxygen addition elicit profound changes in community transcription in OMZ bacterioplankton. *PLoS One* **7**: e37118.
- Stewart FJ, Ottesen EA, DeLong EF. (2010). Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *ISME J* **4**: 896–907.
- Stewart FJ, Ulloa O, DeLong EF. (2012b). Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environ Microbiol* **14**: 23–40.
- Stocker R. (2012). Marine microbes see a sea of gradients. *Science* **338**: 628–633.
- Stocker R, Seymour JR, Samadani A, Hunt DE, Polz MF. (2008). Rapid chemotactic response enables marine bacteria to exploit ephemeral microscale nutrient patches. *Proc Natl Acad Sci USA* **105**: 4209–4214.
- Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D *et al*. (2011). Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* **333**: 1296–1300.
- Takami H, Noguchi H, Takaki Y, Uchiyama I, Toyoda A, Nishi S *et al*. (2012). A deeply branching thermophilic bacterium with an ancient acetyl-CoA pathway dominates a subsurface ecosystem. *PLoS One* **7**: e30559.
- Thamdrup B, Dalsgaard T, Jensen MM, Ulloa O, Farias L, Escobedo R. (2006). Anaerobic ammonium oxidation in the oxygen-deficient waters off northern Chile. *Limnol Oceanogr* **51**: 2145–2156.
- Thamdrup B, Dalsgaard T, Revsbech NP. (2012). Widespread functional anoxia in the oxygen minimum zone of the eastern South Pacific. *Deep Sea Res I* **65**: 36–45.
- Traxler MF, Summers SM, Nguyen HT, Zacharia VM, Hightower GA, Smith JT *et al*. (2008). The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Mol Microbiol* **68**: 1128–1148.
- Twiss E, Coros AM, Tavakoli NP, Derbyshire KM. (2005). Transposition is modulated by a diverse set of host factors in *Escherichia coli* and is stimulated by environmental stress. *Mol Microbiol* **57**: 1593–1607.
- Ulloa O, Canfield DE, DeLong EF, Letelier RM, Stewart FJ. (2012). Perspective: microbial oceanography of anoxic



- oxygen minimum zones. *Proc Natl Acad Sci USA* **109**: 15996–16003.
- Ulloa O, Pantoja S. (2009). The oxygen minimum zone of the eastern South Pacific. *Deep Sea Res Pt II* **56**: 987–991.
- van de Vossen J, Woebken D, Maalcke WJ, Wessels HJ, Dutilh BE, Kartal B *et al.* (2012). The metagenome of the marine anammox bacterium '*Candidatus Scalindua profunda*' illustrates the versatility of this globally important nitrogen cycle bacterium. *Environ Microbiol* **15**: 1275–1289.
- van der Maarel MJ, Sprenger W, Haanstra R, Forney LJ. (1999). Detection of methanogenic archaea in seawater particles and the digestive tract of a marine fish species. *FEMS Microbiol Lett* **173**: 189–194.
- Van Mooy BAS, Keil RG, Devol AH. (2002). Impact of suboxia on sinking particulate organic carbon: enhanced carbon flux and preferential degradation of amino acids via denitrification. *Geochim Cosmochim Acta* **66**: 457–465.
- Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA *et al.* (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66–74.
- Wagner M, Horn M. (2006). The Planctomycetes, Verrucomicrobia, Chlamydiae and sister phyla comprise a superphylum with biotechnological and medical relevance. *Curr Opin Biotechnol* **17**: 241–249.
- Walsh DA, Zaikova E, Howes CG, Song YC, Wright JJ, Tringe SG *et al.* (2009). Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science* **326**: 578–582.
- Werren JH. (2011). Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc Natl Acad Sci USA* **108**: 10863–10870.
- Whitmire AL, Letelier RM, Villagran V, Ulloa O. (2009). Autonomous observations of in vivo fluorescence and particle backscattering in an oceanic oxygen minimum zone. *Optics Exp* **17**: 21992–22004.
- Williamson SJ, Allen LZ, Lorenzi HA, Fadrosch DW, Brami D, Thiagarajan M *et al.* (2012). Metagenomic exploration of viruses throughout the Indian Ocean. *PLoS ONE* **7**: e42047.
- Woebken D, Fuchs BM, Kuypers MM, Amann R. (2007). Potential interactions of particle-associated anammox bacteria with bacterial and archaeal partners in the Namibian upwelling system. *Appl Environ Microb* **73**: 4648–4657.
- Woebken D, Lam P, Kuypers MM, Naqvi SW, Kartal B, Strous M *et al.* (2008). A microdiversity study of anammox bacteria reveals a novel *Candidatus Scalindua* phylotype in marine oxygen minimum zones. *Environ Microbiol* **10**: 3106–3119.
- Wright JJ, Konwar KM, Hallam SJ. (2012). Microbial ecology of expanding oxygen minimum zones. *Nat Rev Microbiol* **10**: 381–394.
- Wright TD, Vergin KL, Boyd PW, Giovannoni SJ. (1997). A novel delta-subdivision proteobacterial lineage from the lower ocean surface layer. *Appl Environ Microb* **63**: 1441–1448.
- Wyman M, Hodgson S, Bird C. (2013). Denitrifying Alphaproteobacteria from the Arabian Sea that express the gene (*nosZ*) encoding nitrous oxide reductase in oxic and sub-oxic waters. *Appl Environ Microbiol* **79**: 2670–2681.
- Zaikova E, Walsh DA, Stilwell CP, Mohn WW, Tortell PD, Hallam SJ. (2010). Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. *Environ Microbiol* **12**: 172–191.
- Zehr JP, Kudela RM. (2011). Nitrogen cycle of the open ocean: from genes to ecosystems. *Ann Rev Mar Sci* **3**: 197–225.
- Zimmer RL, Woollacott RM. (1983). Mycoplasma-like organisms: occurrence with the larvae and adults of a marine bryozoan. *Science* **220**: 208–210.

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