ENVIRONMENTAL NICHE PARTITIONING OF MICROBIAL COMMUNITY GENOMIC DIVERSITY, GENE EXPRESSION, AND METABOLISM IN A MARINE OXYGEN MINIMUM ZONE

A Dissertation
Presented to
The Academic Faculty

by

Sangita Ganesh

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Biology

Georgia Institute of Technology
December 2016

COPYRIGHT 2016 BY SANGITA GANESH

ENVIRONMENTAL NICHE PARTITIONING OF MICROBIAL COMMUNITY GENOMIC DIVERSITY, GENE EXPRESSION, AND METABOLISM IN A MARINE OXYGEN MINIMUM ZONE

Approved by:

Dr. Frank J Stewart, Advisor
School of Biology
Georgia Institute of Technology

Dr. Fredrik O Vannberg
School of Biology
Georgia Institute of Technology

Dr. Thomas J DiChristina
School of Biology
Georgia Institute of Technology

Dr. Konstantinos Konstantinidis
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Joel E Kostka
School of Biology
Georgia Institute of Technology

Date Approved: 11th November 2016

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to my advisor, Dr. Frank Stewart for his immense support and guidance through every step of the way. He has been the greatest inspiration for any graduate student, and has motivated me to be a better scientist and instilled in me a great work ethic, for which I will forever be indebted to him. He has continually inspired me and made me realize my passion for this subject, which I hope to continually be working toward. I would also like to thank each and every member of my Graduate Committee for their feedback, guidance and support through all of my years at Georgia Tech.

I am grateful to my Mother, Grandmother, and Brother for having stood by me and supported me through all my years in graduate school. I thank them for believing in me and I hope my work makes them proud.

I am also grateful to all my friends at the Stewart lab, in Atlanta and elsewhere who have always had my back and made me believe that things are not as tough as they seem, and for enriching my graduate school experience.

I would like to express my overwhelming gratitude to Tushar Bhatia, for being my source of support, my sounding board, my friend, and guide, every step of the way, and for making me believe in myself and constantly encouraging me to reach greater heights.

Finally, I recognize that this research would not have been possible without all the support offered by Georgia Tech and the School of Biology, and I express my sincerest gratitude to the school for helping shape my future.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

pH	Hydrogen ion concentration
µM	Micromolar
µm	Micrometer
N_2	Dinitrogen Gas
NO_3^-	Nitrate
NH_4^+/NH_3	Ammonium
N_2O	Nitrous oxide
$S_2O_3^{2-}$	Thiosulfate
H_2S	Hydrogen sulfide
O_2	Oxygen gas
OMZ	Oxygen Minimum Zone
ETNP	Eastern Tropical North Pacific
ETSP	Eastern Tropical South Pacific
R/V	Research Vessel
Ca.	*Candidatus*
sp.	species
PF	Prefilter
PA	Particle-associated
DO	Dissolved oxygen
DNRA	Dissimilatory Nitrate Reduction to Ammonium
Anammox	Anaerobic Ammonia Oxidation
PCR	Polymerase Chain Reaction
qPCR	Quantitative Polymerase Chain Reaction

DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
rRNA	Ribosomal RNA
EDTA	Ethylene diamine tetraacetic acid
Tris-HCl	2-Amino-2-hydroxymethyl-1,3-propanediol hydrochloride
BLAST	Basic Local Alignment Search Tool
NCBI	National Center for Biotechnology Information
OTU	Operational Taxonomic Unit

SUMMARY

Oxygen Minimum Zones (OMZs) serve as habitats to diverse assemblages of microorganisms that play an important role in mediating global biogeochemical cycles. OMZ microbial communities have not been extensively characterized, and the linkages between microbial community structure, and ecological and biogeochemical processes are still unclear. OMZs act as model systems to study partitioning of microbial niches and biogeochemical transformations owing to their steep vertical gradients of oxygen, nutrient, and redox substrates. This thesis combined genomic tools with environmental measurements of nitrogen transformation rates to characterize how microbial community structure, function and ecological diversity vary at the microscale between free-living (planktonic) and particle-associated microbial communities and over vertical and longitudinal gradients in two of the world's largest permanent OMZs. The results show an important role for particle-association as a major driver of OMZ microbial community metabolic potential and genome content, and identify wide variation in nitrogen transformation rates in the presence versus absence of particles. These results highlight the dependence of free-living microorganisms on particles for substrates and nutrients, as well as selective partitioning of genes facilitating key steps of an important nitrogen loss pathway, denitrification, in the particle-associated microbial fraction. Finally, this thesis describes the genomic composition and gene expression variation of an important OMZ bacterium, *Candidatus* Scalindua sp., responsible for anaerobic ammonia oxidation (anammox), the second major nitrogen loss pathway in OMZs. Combining single cell genomics, transcriptome profiling, and rate measurements, this study identifies high metabolic plasticity of OMZ anammox bacteria in different niches along the OMZ

redoxcline, including a potential for use of diverse nitrogen substrates to drive anammox. Collectively, these studies enhance our understanding of the environmental determinants of microbial diversity and biochemical activity in low oxygen marine systems.

CHAPTER 1

INTRODUCTION

Microorganisms represent the most abundant, biogeochemically important biological component of the oceans. These microorganisms are also the greatest potential reservoir of useful genes for medicine and biotechnology (Pedros-Alio, 2006). Understanding marine microbial diversity patterns is crucial towards understanding mechanisms that help structure microbial communities, and the processes that shape global diversity. This knowledge of microbial diversity would further our understanding of global biogeochemical cycles in which microbes play central roles (Martiny et al., 2006), and informing ecosystem-level conservation and management decisions (Richardson and Whittaker, 2010).

1.1 Environmental Factors shaping Marine Microbial Diversity

Microbial habitats in the oceans are influenced by numerous forces and factors including, but not limited to light, salinity, ocean currents, terrestrial inputs, nutrients, dissolved oxygen, and climate. Ocean and seafloor currents lead to water column mixing and drive nutrient turnover, impacting transport processes. These ocean currents also expose the water column microbes to rapidly fluctuating environmental conditions like temperature, leading to selection of communities with greatest plasticity and evolvability (Doblin et al., 2016). Terrestrial inputs from streams, rivers and winds can create gradients of nutrients, pollutants, and organic matter within and across the ocean water column. Together, these factors help establish vertical and longitudinal nutrient, temperature, and redox gradients within the water column.

Climate exerts one of the largest influences on microbial habitats. Increase in global temperatures is predicted to cause a decrease in dissolved oxygen content in seawater, thereby causing an expansion of low oxygen or dead zones in the global oceans (Stramma et al., 2010). Dissolved oxygen is a crucial organizing determinant in marine ecosystems. Being the most energetically favorable terminal electron acceptor, its depletion in the water column, leads to a restructered microbial community community specialized in anaerobic metabolism, with the capability to utilize other electron acceptors like nitrate, nitrite, and sulfate.

Interfaces within the water column are also considered hotbeds for microbial activity. Interfaces provide opposing gradients of substrates and shape a unique ecological niche that allows microorganisms to actively metabolize an exclusive combination of nutrients (Brune et al., 2000). Interfaces for microbial activity in the marine environment may be sediment-water interfaces (Novitsky, 1983), oxic-anoxic boundaries (Brune et al., 2000), and the boundaries of the photic zone in the water column. Microorganisms at interfaces may rapidly evole adaptations or may exhibit genomic plasticity owing to their dynamic and rapidly evolving environmental conditions.

In addition to vertical and longitudinal gradients, microscale heterogeneities also exist within the marine water column. These heterogeneities ranging from a few microns to a few centimeters in size, may be in the form of small inorganic or organic particles, detritus, or live phytoplankton and zooplankton cells (Karl et al., 1984; DeLong et al., 1993; Fenchel, 2002; Stocker et al., 2012). They can provide a surface for microbial growth within the water column. Compared with the open water column, these particles

form unique microhabitats that are relatively enriched in nutrients and contain steep microscale (microns) gradients in pH and redox substrates, including organic carbon potentially supporting both aerobic and anaerobic respiration (Alldredge and Cohen, 1987; Alldredge and Silver, 1988; Stocker, 2012).

Marine organic particles are thought to be important agents in the flux of biogenic carbon from the surface to the deep ocean (Karl et al., 1988; Azam et al., 1998) and have been studied extensively in oxygenated open ocean environments. Marine snow can form localized highly reducing microzones capable of producing and acting as a source of sulfide as they sink through the water column (Shanks and Reeder, 1993), serving as a conduit of nutrients to free-living microbes as they travel through the water column. Thus, the microbes themselves exert an influence on their habitats by consuming, producing, and sequestering a variety of compounds.

Marine particles support complex surface-attached microbial communities whose composition and life history strategies differ significantly from those of free-living microbes. Particle-associated heterotrophic communities have been shown to be able to enzymatically metabolize high molecular weight substrates to provide hydrolysate to the particle-associated community as well as to the surrounding free-living community, (Smith et al., 1992; Simon et al., 2002; Grossart, 2010; Arnosti et al., 2012), which are potentially enriched in taxa adapted for oligotrophic or autotrophic free-living lifestyles (Lauro et al., 2009).

The activities of particle-associated communities likely directly and indirectly affect free-living microbes, potentially via metabolic transformations of particulate

material (e.g., denitrification) that release inorganic and organic substrates. Knowledge of such linkages in marine systems is critical for understanding the role of particulate flux in regulating bulk elemental transformations.

Together, these environmental factors establish a stratification of ocean microbiota. However, the factors structuring linkages between diverse microbial metabolisms in the ocean are unclear. Notably, the extent to which metabolisms are partitioned at the microscale remains unresolved, and the patterns of gene and protein expression underlying microbial responses to environmental change are not well understood in these complex natural communities.

Understanding the linkages between factors driving community structure would facilitate the identification and characterization of genotypic and phenotypic variants of dominant marine taxa that form a part of the stratified microbial assemblages in the marine water column.

1.2 Oxygen Minimum Zones as model systems to understand drivers of microbial diversity

Oxygen Minimum Zones (OMZs) are an intrinsic feature formed in regions of the ocean where aerobic respiration of organic matter combined with little or no mixing of the water column, leads to persistent low or no oxygen conditions, with dissolved O_2 concentrations falling to near or below the detection limit (<1 μM) at mid-water depths (Ulloa et al. 2012). OMZs are a prominent feature in the Eastern Tropical North Pacific (ETNP), Eastern Tropical South Pacific (ETSP), and the Arabian Sea, and are formed as

a result of coastal upwelling which brings nutrient rich deep waters to the surface along the continental shelf.

OMZs exhibit steep gradients of light, oxygen, chlorophyll, and electron donors and acceptors, with depth. OMZs are also typically characterized by an accumulation of nitrite (NO_2^-) (Kamykowski and Zentara, 1991). OMZ systems also possess a characteristic intermediate depth particle maximum (Pak et al, 1980, Kullenberg, 1981, Garfield et al., 1983), and this particle maximum has been associated with a peak in bacterial abundance and heterotrophic microbial metabolism (Whitmire et al., 2009; Naqvi et al., 1993). This particle maximum is confined to the suboxic region within the OMZ where a shift to anaerobic microbial metabolism occurs i.e. denitrification, suggesting a niche partitioning of an important OMZ metabolic process. These unique gradients of nutrients and redox substrates makes OMZs excellent model systems for investigating the effect of dramatic environmental variation over relatively small spatial scales in shaping marine microbial diversity.

Although OMZs account for only 0.1% of the global ocean volume (Paulmier and Ruiz-Pino, 2009), large fixed nitrogen deficits occur in these OMZs as a result of the conversion of large amounts of bioavailable nitrate (NO_3^-) to dinitrogen gas (N_2) by the process of denitrification and anaerobic oxidation of ammonium using nitrite (Anammox) to N_2. This accounts for 30-50% of the total fixed nitrogen lost to the atmosphere as N_2 from oceans (Codispoti et al., 2001; Gruber, 2004). Thus, OMZs are considered to be global sinks for fixed nitrogen. The process of nitrifier-denitrification carried out by nitrifying bacteria at low oxygen levels occurring at the boundaries of OMZs, also

contribute to the production and release of a potent greenhouse gas N_2O which contributes to the destruction of stratospheric ozone (Codispoti et al., 2005).

Recent studies have also linked the oxidation of reduced sulfur-compounds including thiosulfate ($S_2O_3^{2-}$) and hydrogen sulfide (H_2S) to nitrogen transformations in OMZs, providing evidence for the existence of a cryptic sulfur cycle with the potential to drive inorganic carbon fixation (Canfield et al., 2010).

1.3 Microbial Diversity in Oxygen Minimum Zones

Oxygen drawdown in OMZs significantly restructures the pelagic ecosystem. Higher eukaryotic organisms like fish are not usually detected in OMZ waters and anaerobic microbial processes dominate in these zones, resulting in a community dominated by Bacteria and Archaea with diverse anaerobic metabolisms.

Communities along the oxycline are 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., 2012). In contrast, community metabolism at the 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).

OMZs play an essential role in the global nitrogen cycle. The important microbial processes contributing to the global nitrogen cycle are: Nitrification (in the upper OMZ and oxycline), Denitrification (OMZ core), and the more recently discovered process of

Anammox (OMZ core) (Paulmier and Ruiz-Pino, 2009). Each of these nitrogen transformation processes is carried out by different lineages of microbes potentially interacting with one another. Under oxic and suboxic conditions, the process of nitrification transforms NH_4 into NO_3 resulting in the accumulation of bioavailable nitrogen in the OMZ (Ward, 2008). This process is mediated by ammonia oxidizing archaea/bacteria and nitrite oxidizing bacteria from diverse taxonomic lineages. Another important process associated with OMZs is denitrification, which is a bacterial process occurring only in oxygen deficient regions (Codispoti et al., 2001). Denitrification converts NO_3, one of the main limiting nutrients in the ocean, into gaseous nitrogen (N_2O, N_2) which is lost to the atmosphere, and contributes to the oceanic nitrate deficit (Tyrrell, 1999). Partial denitrification also leads to the release of the greenhouse gas N_2O. Diverse assemblages of α, β, γ - proteobacteria make up the denitrifying microbial community in OMZs. A previously unknown process in the ocean was observed in the water column in the OMZs (Kuypers et al., 2003): anaerobic oxidation of ammonium (NH_3) using nitrite (NO_2) (anammox) (Arrigo, 2005), leading to a complete revision of the global nitrogen cycle. The process of anammox is carried out by microorganisms possessing highly specialized cellular compartments called the anammoxosome, which retains the cellular machinery to carry out the process of anammox (Strous et al., 1999). Thus far, the only known bacteria carrying out this process form a monophyletic branch within the phylum Planctomycetes. A single known clade of anammox planctomycete, *Candidatus* Scalindua dominates the anammox process in OMZs (Thamdrup et al., 2006; Galan et al., 2009; Lam et al., 2009). Research in the large anoxic OMZs associated with upwelling zones of the Eastern Pacific indicates that anammox and denitrification are

linked with a diverse set of other microbial nitrogen transformations, including the production of nitrite and nitrate via nitrification along the oxycline, and dissimilatory nitrate reduction to ammonium (DNRA) within the OMZ (Lam et al., 2009; Kalvelage et al., 2013). Taken together, these microbial nitrogen transformations constitute a distributed metabolic network linking the metabolic potentials of different taxonomic groups to higher-order biogeochemical cycling of nitrogen in the environment.

Warming of the upper ocean waters as a consequence of climate change has contributed to deoxygenation of the global oceans and expansion of OMZ waters (Ulloa et al., 2012). Continued OMZ expansion due to climate change is an emerging environmental concern as it will potentially exacerbate the loss of fixed nitrogen from the ocean, in addition to increasing N_2O production (Codispoti, 2010) and a resulting feedback on marine ecosystem function by limiting primary production in the overlying oxygenated waters. Despite the importance of these effects, little is known about the metabolism of OMZ microbes and their response to environmental changes. Determining how microbial interaction networks form, function and change over time may reveal potential links between microbial community structure and ecological and biogeochemical processes.

1.4 Objectives and Experimental Overview

In order to broadly characterize the microbial community in OMZs and understand the factors shaping the community genetic diversity, we employ a molecular and experimental approach to critically examine how vertical and micro-scale gradients shape the functional diversity of OMZ microbiota. The work described here was carried out in the Eastern Tropical North and South Pacific OMZs. They are permanent OMZ features occuring in the coastal upwelling zones off the coast of Mexico and Chile, respectively. They are effectively anoxic and exhibit a large nitrite accumulation. They make up the two largest low oxygen zones in the world, and the microbial ecology of these regions, particularly in the ETNP OMZ, remains understudied. Characterizing the distribution of microbial taxa and metabolisms across vertical gradients and across micro-niches will help clarify the role of these important driving forces in shaping OMZ biogeochemistry.

The research outlined here describes the role of micro- and macro-scale gradients in shaping community genomic and metabolic diversity in the OMZ by focusing on the following objectives:

OBJECTIVE 1 (Chapter 2): Elucidate the community genomic structure (metagenome) of particle-associated microorganisms, and how it differs from that of free-living microorganisms in the OMZ water column. We used filter size fractionation as a means to separate large particles and aggregates from the small, potentially free-living microbes in the water column. For the purpose of our study, we use the term “particle” to refer to material excluding non-aggregated or individual bacterial cells, so as to draw a

distinction between surface-associated and free-living microbial niches. Metagenomics and 16S rRNA gene amplicon sequencing of size-fractionated microbial biomass was used to compare the community genetic potential and taxonomic composition between particle-associated and free-living communities relative to depth and redox gradients in the OMZ. A primary focus was placed on the partitioning of major metabolic processes of dissimilatory nitrogen and sulfur cycles. Particle-association was identified 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.

OBJECTIVE 2 (Chapter 3): Experimentally demonstrate how microbial gene expression and major steps of the OMZ nitrogen cycle are partitioned between the particle-associated and free-living microbial communities. Metatranscriptomic analysis was used to predict the metabolic activity of particle-associated and free-living fractions, and was analyzed in tandem with process rate measurements of dissimilatory OMZ N-cycle processes, carried out in collaboration with Dr. Bo Thamdrup. Resulting meta-omic data was analyzed statistically to assess variation in functional gene content, population/community structure, single-taxon abundance, and differential transcript representation, thereby identifying the relative activity of diverse community members and metabolic pathways. These data were interpreted relative to qPCR-based measurements of bacterial abundance in particle-associated and free-living fractions and relative to gradients in oxygen and nutrient concentrations. The results indicate a significant contribution of particles and particle-associated bacteria to nitrogen cycling in OMZs.

OBJECTIVE 3 (Chapter 4): Describe the effects of filtration method on quantitative measurements of community genetic diversity. This study was carried out in collaboration with another graduate student, Cory Padilla, who performed the ship-board filtration experiments. Experiments were conducted to test whether variation in the volume of filtered seawater biases estimates of marine microbial diversity. Size-fractionated biomass representing two size fractions (0.2–1.6μm, >1.6μm) was collected from a marine oxygen minimum zone following a sequential inline filtration protocol. Sequencing of community 16S rRNA gene amplicons was used to study shifts in community structure and richness over volumes ranging from 0.05 to 5 L. We were able to highlight the potential of sample volume variation in confounding comparisons of aquatic microbial diversity estimates and inferring the structure of particle-associated communities.

OBJECTIVE 4 (Chapter 5): Demonstrate the existence of genetically and functionally divergent ecotypes of a major OMZ planctomycete clade catalyzing anaerobic ammonium oxidation, at larger macroscale gradients. Single Cell Genomics was used as a method to understand the genomic potential of an important marine lineage *Candidatus* Scalindua and its functional dynamics across OMZ vertical gradients. Scalindua Single Amplified Genomes (SAGs) were used to enable transcriptome mapping and target-gene phylogenetic analyses to identify differences in gene content, Scalindua-specific expression, and its metabolic diversity across larger scale OMZ gradients. Our analyses indicate a transcriptional coupling of gene expression to substrate availability, and the genomic potential of these bacteria to utilize alternate nitrogen sources within the OMZ.

1.5 References

Alldredge AL, Silver MW. (1988). Characteristics, dynamics and significance of marine snow. Prog Oceanogr 20: 41–82.

Arnosti, C., Fuchs, B.M., Amann, R. and Passow, U. (2012). Contrasting extracellular enzyme activities of particle-associated bacteria from distinct provinces of the North Atlantic Ocean. Front Microbiol 3: 425.

Arrigo, K.R. (2005). Marine microorganisms and global nutrient cycles. Nature 437:349-355.

Azam, F. (1998). Microbial control of oceanic carbon flux: the plot thickens. Science *280*:694-696.

Brune, A., Frenzel, P. and Cypionka, H. (2000). Life at the oxic–anoxic interface: microbial activities and adaptations. FEMS Microbiology Reviews, 24:691-710.

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.

Canfield, D.E., Stewart, F.J., Thamdrup, B., De Brabandere, L., Dalsgaard, T., Delong, E.F., Revsbech, N.P. and Ulloa, O. (2010). A cryptic sulfur cycle in oxygen-minimum–zone waters off the Chilean coast. Science 330:1375-1378.

Codispoti, L.A., Brandes, J.A., Christensen, J.P., Devol, A.H., Naqvi, S.W.A., Paerl, H.W. and Yoshinari, T. (2001). The oceanic fixed nitrogen and nitrous oxide budgets: Moving targets as we enter the anthropocene? Scientia Marina 65:85-105.

DeLong EF, Franks DG, Alldredge AL. (1993). Phylogenetic diversity of aggregate-attached vs free-living marine bacterial assemblages. Limnol Oceanogr 38: 924–934.

Doblin, M.A. and van Sebille, E. (2016). Drift in ocean currents impacts intergenerational microbial exposure to temperature. Proc Natl Acad Sci USA 113:5700-5705.

Fenchel T. (2002). Microbial behavior in a heterogenous world. Science 296: 1068–1071.

Galán, A., Molina, V., Belmar, L. and Ulloa, O. (2012). Temporal variability and phylogenetic characterization of planktonic anammox bacteria in the coastal upwelling ecosystem off central Chile. Prog. Oceanogr. 92:110-120.

Grossart HP. (2010). Ecological consequences of bacterioplankton lifestyles: changes in concepts are needed. Environ Microbiol Rep 2: 706–714.

Kalvelage T, Lavik G, Lam P, Contreras S, Arteaga L, Loscher CR et al. (2013). Nitrogen cycling driven by organic matter export in the South Pacific oxygen minimum zone. Nat Geosci 6: 228–234.

Kamykowski, D. and Zentara, S.J. (1991). Spatio-temporal and process-oriented views of nitrite in the world ocean as recorded in the historical data set. Deep Sea Research Part A. Oceanographic Research Papers 38:445-464.

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.

Kullenberg, G. (1982). A comparison of distributions of suspended matter in the Peru and northwest African upwelling areas. Coastal upwelling 282-290.

Kuypers, M.M., Sliekers, A.O., Lavik, G., Schmid, M., Jørgensen, B.B., Kuenen, J.G., Damsté, J.S.S., Strous, M. and Jetten, M.S. (2003). Anaerobic ammonium oxidation by anammox bacteria in the Black Sea. Nature 422:608-611.

Lam, P., Lavik, G., Jensen, M.M., van de Vossenberg, J., Schmid, M., Woebken, D., Gutiérrez, D., Amann, R., Jetten, M.S. and Kuypers, M.M. (2009). Revising the nitrogen cycle in the Peruvian oxygen minimum zone. Proc Natl Acad Sci USA 106:4752-4757.

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

Martiny, J.B.H., Bohannan, B.J., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., Horner-Devine, M.C., Kane, M., Krumins, J.A., Kuske, C.R. and Morin, P.J. (2006). Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol, 4:102-112.

Naqvi, S.W.A., Kumar, M.D., Narvekar, P.V., De Sousa, S.N., George, M.D. and D'Silva, C. (1993). An intermediate nepheloid layer associated with high microbial metabolic rates and denitrification in the northwest Indian Ocean. Journal of Geophysical Research: Oceans 98:16469-16479.

Novitsky, J.A. (1983). Microbial activity at the sediment-water interface in Halifax Harbor, Canada. Appl Environ Microbiol 45:1761-1766.

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.

Paulmier, A. and Ruiz-Pino, D., (2009). Oxygen minimum zones (OMZs) in the modern ocean. Prog. Oceanogr. 80:113-128.

Pedrós-Alió, C., (2006). Marine microbial diversity: can it be determined? Trends Microbiol 14:257-263.

Richardson, D.M. and Whittaker, R.J. (2010). Conservation biogeography–foundations, concepts and challenges. Diversity and Distributions, 16:313-320.

Shanks, A.L. and Reeder, M.L., (1993). Reducing microzones and sulfide production in marine snow. Mar Ecol Prog Ser 96:43-47.

Simon M, Grossart HP, Schweitzer B, Ploug H. (2002). Microbial ecology of organic aggregates in aquatic ecosystems. Aquat Microb Ecol 28: 175–211.

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.

Stewart FJ, Ulloa O, DeLong EF. (2012). 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.

Stramma, L., Schmidtko, S., Levin, L.A. and Johnson, G.C. (2010). Ocean oxygen minima expansions and their biological impacts. Deep Sea Res Pt I: Oceanographic Research Papers 57:587-595.

Strous, M., Kuenen, J.G. and Jetten, M.S. (1999). Key physiology of anaerobic ammonium oxidation. Appl Environ Microbiol 65:3248-3250.

Thamdrup, B., Dalsgaard, T., Jensen, M.M., Ulloa, O., Farías, L. and Escribano, R., (2006). Anaerobic ammonium oxidation in the oxygen-deficient waters off northern Chile. Limnol Oceanogr 51:2145-2156.

Tyrrell, T. (1999). The relative influences of nitrogen and phosphorus on oceanic primary production. Nature 400:525-531.

Ulloa O, Pantoja S. (2009). The oxygen minimum zone of the eastern South Pacific. Deep Sea Res Pt II 56: 987–991.

Ulloa, O., Canfield, D.E., DeLong, E.F., Letelier, R.M. and Stewart, F.J. (2012). Microbial oceanography of anoxic oxygen minimum zones. Proc Natl Acad Sci USA 109:15996-16003.

Whitmire, A.L., Letelier, R.M., Villagrán, V. and Ulloa, O. (2009). Autonomous observations of in vivo fluorescence and particle backscattering in an oceanic oxygen minimum zone. Optics express 17:21992-22004.

Wright JJ, Konwar KM, Hallam SJ. (2012). Microbial ecology of expanding oxygen minimum zones. Nat Rev Microbiol 10: 381–394.

CHAPTER 2

METAGENOMIC ANALYSIS OF SIZE-FRACTIONATED PICOPLANKTON IN A MARINE OXYGEN MINIMUM ZONE

2.1 Abstract

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 upto threefold in the 0.2–1.6μm community. Functional gene composition also differed between fractions, with the >1.6μm 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.

2.2 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 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 (~250mM) at the surface to below the level of detection (<10nM) at the OMZ core (~200–300m; 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 ammoniaand 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 0.8–30 and 0.2μm, respectively. Although the microfilter fraction (typically cells between 0.2 and 1.6–3 μm) 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 surfaceattached 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, particleassociated 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 (~0.2–3 μm; 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 μm 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 μm). 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 (0.2–1.6μm 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 pico-eukaryote (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.

2.3 Materials and Methods

2.3.1 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 (20º 04.999 S, 70º 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-TemperatureDepth profiler (Sea-Bird SBE 911plus) equipped with a dissolved Oxygen Sensor, fluorometer and transmissometer (see Supplementary Figure A.1). Microbial biomass was collected by sequential in-line filtration of seawater samples (10l) 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 mm 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 -80ºC until DNA extraction.

2.3.2 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 ml 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 1C. Proteinase K (1 mg in 100 ml lysis buffer, with 100 ml 20% SDS) was added, and the tubes and cartridges were resealed and further incubated for 2 h at 55ºC. 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 mg) were used to prepare libraries for shotgun pyrosequencing of microbial metagenomes.

2.3.3 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 B480-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 95ºC (2 min), followed by 30 cycles of denaturation at 95ºC (20 s), primer annealing at 50ºC (30 s) and primer extension at 72ºC (5min). 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 QIA Quick PCR Clean-Up Kit, pooled (~200 ng per sample), and used as template for multiplex amplicon pyrosequencing.

2.3.4 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 1000m), 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 fullplate run.

2.3.5 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 2.1a). 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 relatednesss based on Unifrac was visualized using a two-dimensional Principal Coordinate Analysis (Figure 2.1b). 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 (a 1⁄4 0.05) imposed for multiple tests according to Benjamini and Hochberg (1995) (see Supplementary Table A.2).

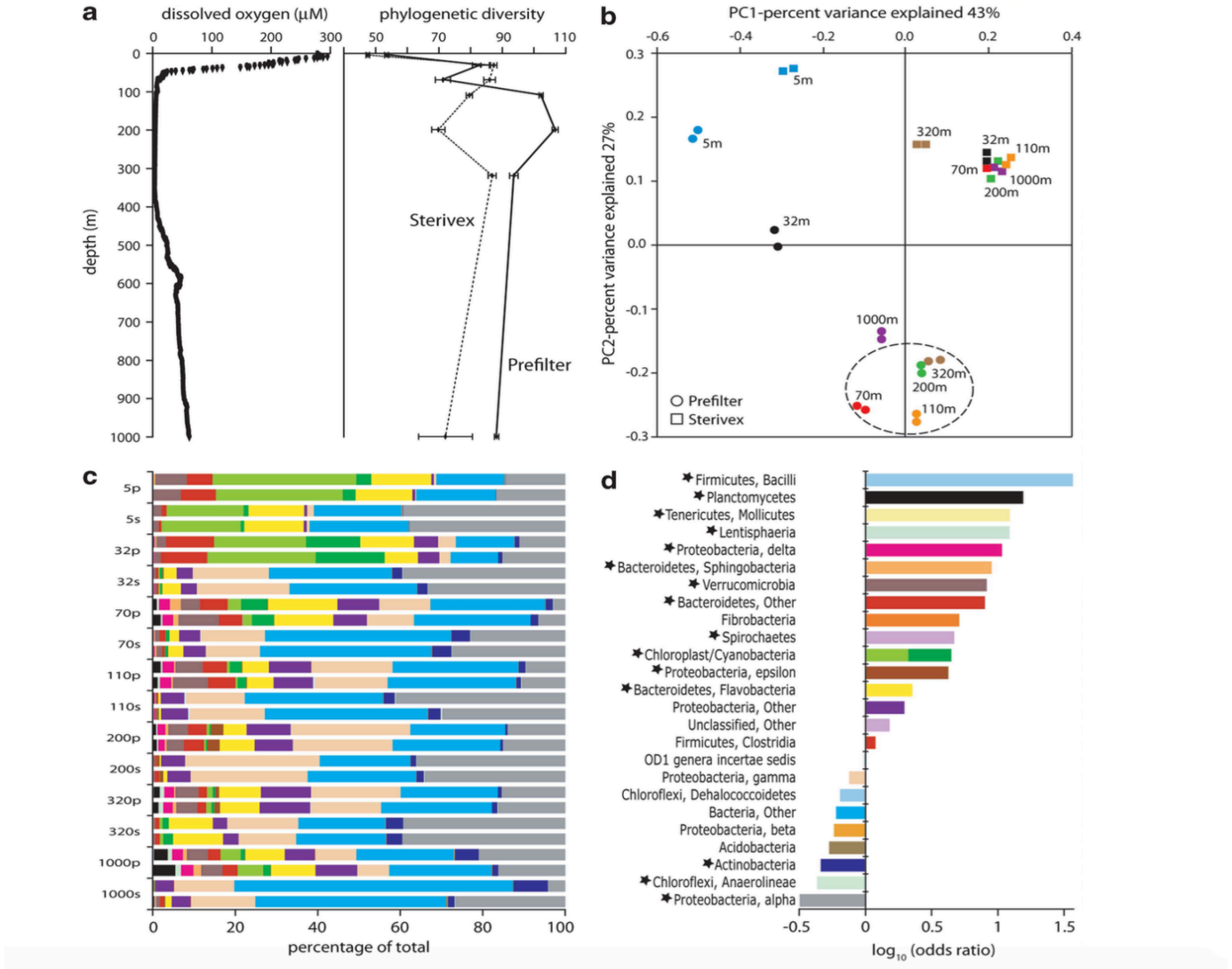

a
dissolved oxygen (µM)
phylogenetic diversity
depth (m)
Sterivex
Prefilter
b
PC1-percent variance explained 43%
PC2-percent variance explained 27%
5m
32m
70m
110m
200m
320m
1000m
Prefilter
Sterivex
c
percentage of total
d
Firmicutes, Bacilli
Planctomycetes
Tenericutes, Mollicutes
Lentisphaeria
Proteobacteria, delta
Bacteroidetes, Sphingobacteria
Verrucomicrobia
Bacteroidetes, Other
Fibrobacteria
Spirochaetes
Chloroplast/Cyanobacteria
Proteobacteria, epsilon
Bacteroidetes, Flavobacteria
Proteobacteria, Other
Unclassified, Other
Firmicutes, Clostridia
OD1 genera incertae sedis
Proteobacteria, gamma
Chloroflexi, Dehalococcoidetes
Bacteria, Other
Proteobacteria, beta
Acidobacteria
Actinobacteria
Chloroflexi, Anaerolineae
Proteobacteria, alpha
log10 (odds ratio)

Figure 2.1. 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=prefilter (>1.6μm), s=Sterivex (0.2–1.6μm). 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 ($P<0.05$).

2.3.6 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 non-redundant 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 2.2. Normalized SEED counts for individual depths are available in Supplementary Table A.4. The taxonomic composition of proteincoding 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 2.2, 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 A.3).

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 2.3), 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 cd1-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 2.3), 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* (1410 bp); *nrfA* (1440 bp); *norB* (1410 bp); *nosZ* (1950 bp); *dsrA* (1200bp); *aprA* (1860 bp); *soxB* (1680bp). 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 2.3) 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 (100000 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).

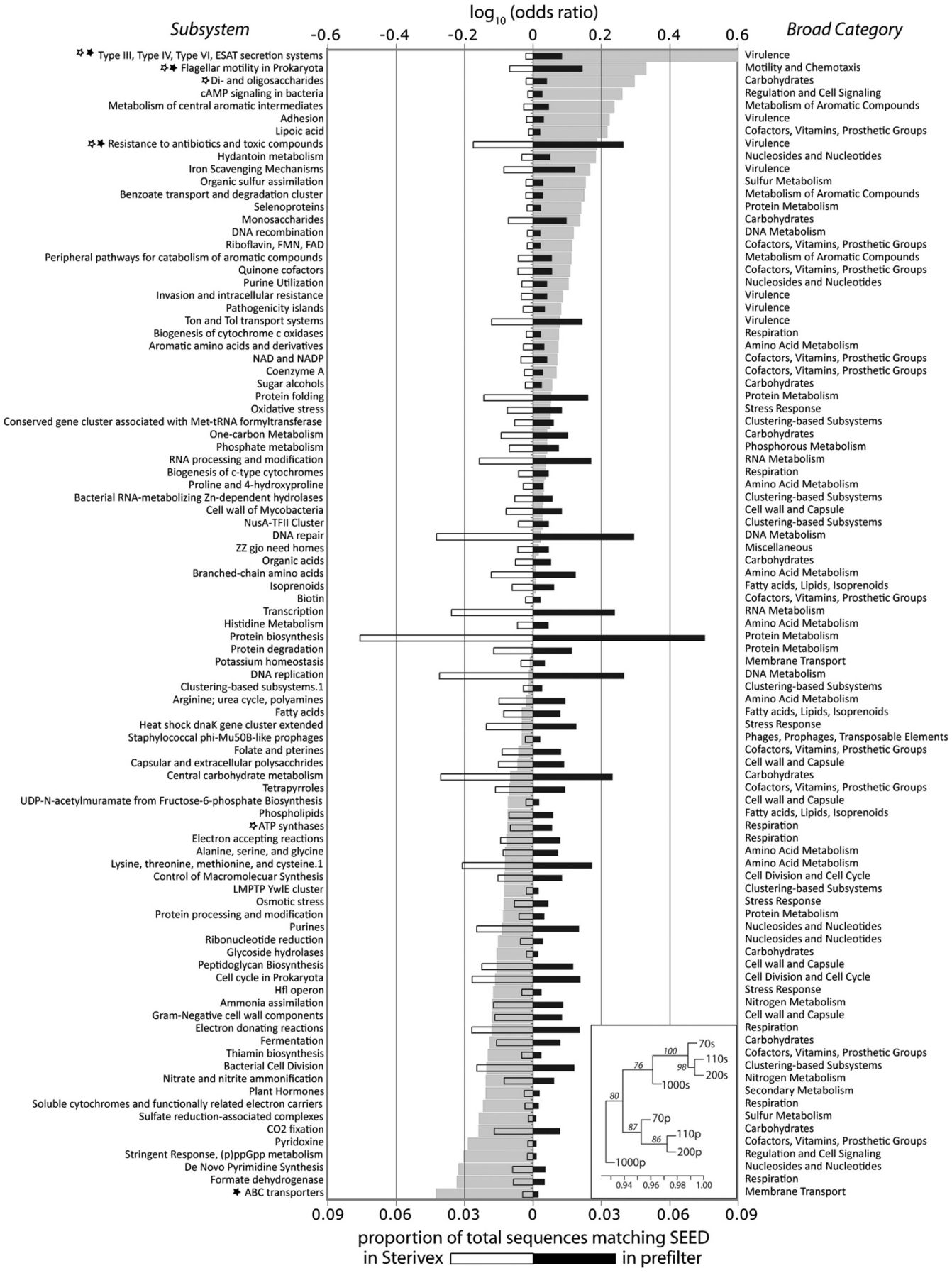

log10 (odds ratio)
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
Subsystem
Broad Category
Type III, Type IV, Type VI, ESAT secretion systems
Flagellar motility in Prokaryota
Di- and oligosaccharides
cAMP signaling in bacteria
Metabolism of central aromatic intermediates
Adhesion
Lipoic acid
Resistance to antibiotics and toxic compounds
Hydantoin metabolism
Iron Scavenging Mechanisms
Organic sulfur assimilation
Benzoate transport and degradation cluster
Selenoproteins
Monosaccharides
DNA recombination
Riboflavin, FMN, FAD
Peripheral pathways for catabolism of aromatic compounds
Quinone cofactors
Purine Utilization
Invasion and intracellular resistance
Pathogenicity islands
Ton and Tol transport systems
Biogenesis of cytochrome c oxidases
Aromatic amino acids and derivatives
NAD and NADP
Coenzyme A
Sugar alcohols
Protein folding
Oxidative stress
Conserved gene cluster associated with Met-tRNA formyltransferase
One-carbon Metabolism
Phosphate metabolism
RNA processing and modification
Biogenesis of c-type cytochromes
Proline and 4-hydroxyproline
Bacterial RNA-metabolizing Zn-dependent hydrolases
Cell wall of Mycobacteria
NusA-TFII Cluster
DNA repair
ZZ gjo need homes
Organic acids
Branched-chain amino acids
Isoprenoids
Biotin
Transcription
Histidine Metabolism
Protein biosynthesis
Protein degradation
Potassium homeostasis
DNA replication
Clustering-based subsystems.1
Arginine; urea cycle, polyamines
Fatty acids
Heat shock dnaK gene cluster extended
Staphylococcal phi-Mu50B-like prophages
Folate and pterines
Capsular and extracellular polysaccharides
Central carbohydrate metabolism
Tetrapyrroles
UDP-N-acetylmuramate from Fructose-6-phosphate Biosynthesis
Phospholipids
ATP synthases
Electron accepting reactions
Alanine, serine, and glycine
Lysine, threonine, methionine, and cysteine.1
Control of Macromolecuar Synthesis
LMPTP YwlE cluster
Osmotic stress
Protein processing and modification
Purines
Ribonucleotide reduction
Glycoside hydrolases
Peptidoglycan Biosynthesis
Cell cycle in Prokaryota
Hfl operon
Ammonia assimilation
Gram-Negative cell wall components
Electron donating reactions
Fermentation
Thiamin biosynthesis
Bacterial Cell Division
Nitrate and nitrite ammonification
Plant Hormones
Soluble cytochromes and functionally related electron carriers
Sulfate reduction-associated complexes
CO2 fixation
Pyridoxine
Stringent Response, (p)ppGpp metabolism
De Novo Pyrimidine Synthesis
Formate dehydrogenase
ABC transporters
Virulence
Motility and Chemotaxis
Carbohydrates
Regulation and Cell Signaling
Metabolism of Aromatic Compounds
Virulence
Cofactors, Vitamins, Prosthetic Groups
Virulence
Nucleosides and Nucleotides
Virulence
Sulfur Metabolism
Metabolism of Aromatic Compounds
Protein Metabolism
Carbohydrates
DNA Metabolism
Cofactors, Vitamins, Prosthetic Groups
Metabolism of Aromatic Compounds
Cofactors, Vitamins, Prosthetic Groups
Nucleosides and Nucleotides
Virulence
Virulence
Virulence
Respiration
Amino Acid Metabolism
Cofactors, Vitamins, Prosthetic Groups
Cofactors, Vitamins, Prosthetic Groups
Carbohydrates
Protein Metabolism
Stress Response
Clustering-based Subsystems
Carbohydrates
Phosphorous Metabolism
RNA Metabolism
Respiration
Amino Acid Metabolism
Clustering-based Subsystems
Cell wall and Capsule
Clustering-based Subsystems
DNA Metabolism
Miscellaneous
Carbohydrates
Amino Acid Metabolism
Fatty acids, Lipids, Isoprenoids
Cofactors, Vitamins, Prosthetic Groups
RNA Metabolism
Amino Acid Metabolism
Protein Metabolism
Protein Metabolism
Membrane Transport
DNA Metabolism
Clustering-based Subsystems
Amino Acid Metabolism
Fatty acids, Lipids, Isoprenoids
Stress Response
Phages, Prophages, Transposable Elements
Cofactors, Vitamins, Prosthetic Groups
Cell wall and Capsule
Carbohydrates
Cofactors, Vitamins, Prosthetic Groups
Cell wall and Capsule
Fatty acids, Lipids, Isoprenoids
Respiration
Respiration
Amino Acid Metabolism
Amino Acid Metabolism
Cell Division and Cell Cycle
Clustering-based Subsystems
Stress Response
Protein Metabolism
Nucleosides and Nucleotides
Nucleosides and Nucleotides
Carbohydrates
Cell wall and Capsule
Cell Division and Cell Cycle
Stress Response
Nitrogen Metabolism
Cell wall and Capsule
Respiration
Carbohydrates
Cofactors, Vitamins, Prosthetic Groups
Clustering-based Subsystems
Nitrogen Metabolism
Secondary Metabolism
Respiration
Sulfur Metabolism
Carbohydrates
Cofactors, Vitamins, Prosthetic Groups
Regulation and Cell Signaling
Nucleosides and Nucleotides
Respiration
Membrane Transport
100
76
98
80
87
86
70s
110s
200s
1000s
70p
110p
200p
1000p
0.94
0.96
0.98
1.00
0.09
0.06
0.03
0
0.03
0.06
0.09
proportion of total sequences matching SEED
in Sterivex
in prefilter

Figure 2.2 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 40.1% 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.

2.4 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 particleassociated 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 μm) and a small size fraction (0.2–1.6 μm), 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.

2.4.1 Oxygen conditions

The ETSP OMZ sample site near Iquique, Chile was characterized by steep vertical gradients in dissolved oxygen (Figure 2.1a), 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 ~40 m, within the oxycline (~30–70 m). Dissolved oxygen conditions at the time of sampling decreased from ~250 μM at the surface to below 5 μM through the OMZ core (~100–400 m), before gradually increasing below 400 m to ~60 μM at 1000m. The oxygen sensor used here has resolution in the micromolar oxygen range. However, recent measurements with high-resolution (10 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 32m samples), the suboxic (< 10 μM) upper OMZ (70 m), the anoxic OMZ core (110, 200, 320 m), and the oxic zone beneath the 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 (~15m) and then again within the OMZ core (~140m), before declining to a consistent minimum below the OMZ (Supplementary Figure A.1). 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.

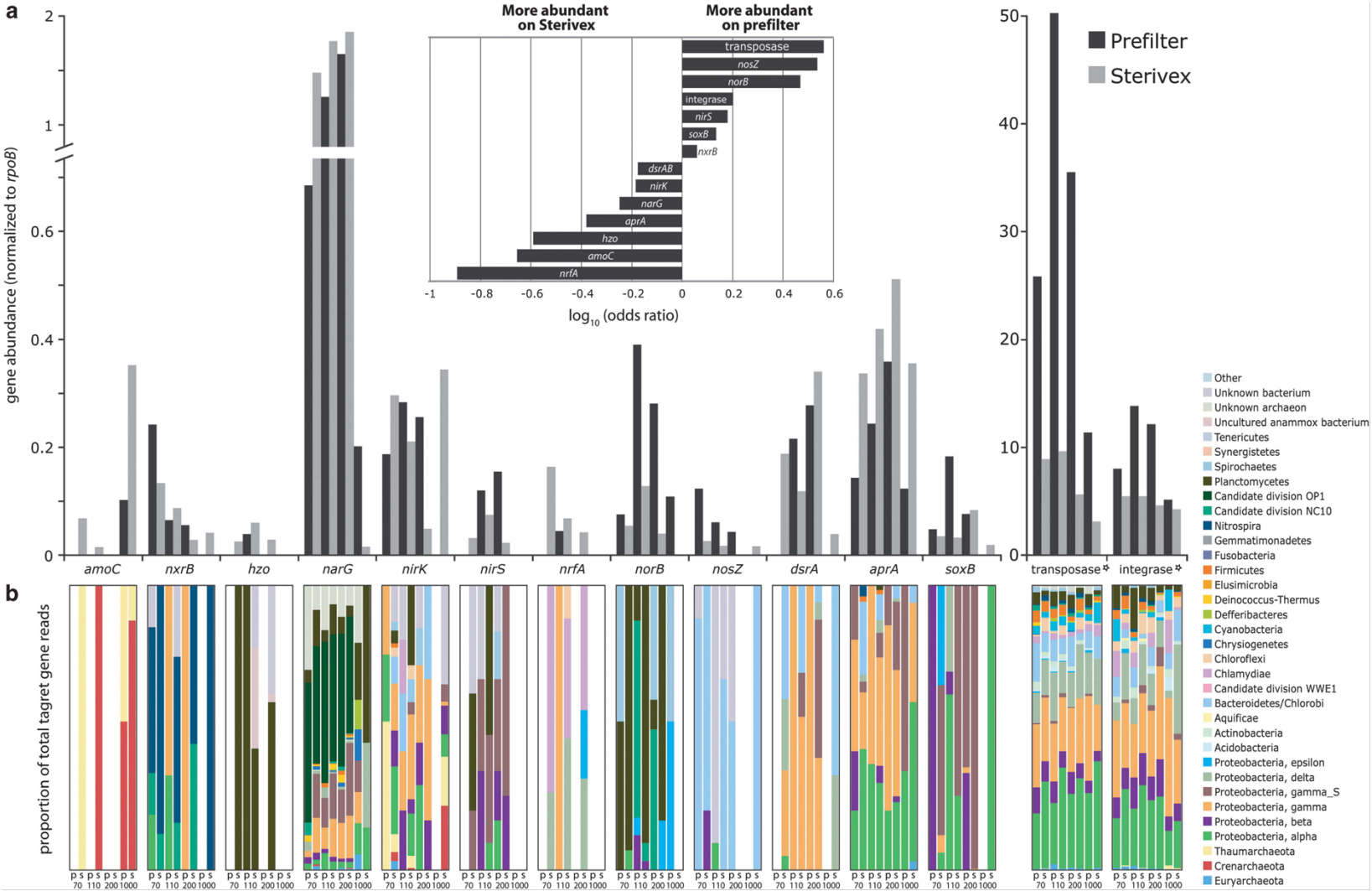

a
b
gene abundance (normalized to rpoB)
More abundant on Sterivex
More abundant on prefilter
transposase
nosZ
norB
integrase
nirS
soxB
nxrB
dsrAB
nirK
narG
aprA
hzo
amoC
nrfA
log10 (odds ratio)
Prefilter
Sterivex
amoC
nxrB
hzo
narG
nirK
nirS
nrfA
norB
nosZ
dsrA
aprA
soxB
transposase
integrase
proportion of total tagret gene reads
p s p s p s p s p s p s
70 110 200 1000
Other
Unknown bacterium
Unknown archaeon
Uncultured anammox bacterium
Tenericutes
Synergistetes
Spirochaetes
Planctomycetes
Candidate division OP1
Candidate division NC10
Nitrospira
Gemmatimonadetes
Fusobacteria
Firmicutes
Elusimicrobia
Deinococcus-Thermus
Defferibacteres
Cyanobacteria
Chrysiogenetes
Chloroflexi
Chlamydiae
Candidate division WWE1
Bacteroidetes/Chlorobi
Aquificae
Actinobacteria
Acidobacteria
Proteobacteria, epsilon
Proteobacteria, delta
Proteobacteria, gamma_S
Proteobacteria, gamma
Proteobacteria, beta
Proteobacteria, alpha
Thaumarchaeota
Crenarchaeota
Euryarchaeota

Figure 2.3 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 (Po0.05; 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 μm), s=Sterivex (0.2–1.6 μm). 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.

2.4.2 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 A.1). Despite relatively high numbers of sequences per sample (mean: 14527), this analysis did not capture the total OTU richness in each sample (that is, no rarefaction curves approached saturation; Supplementary Figure A.2), 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 2.1, Supplementary Figures A.2–A.4). 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 (5m) and increased within the oxycline (~30–70 m) (Figure 2.1a). 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 200m, whereas PD of prefilter communities increased within the core (Figure 2.1a). PD trends closely paralleled those of other alpha diversity indices, including counts of observed OTUs and Chao1 estimates (Supplementary Figure A.3).

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 0.2–1.6μm 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 0.2–2.7μm 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 μm fraction, a pattern consistent with that observed for the prefilters in our study. Similarly, elevated OTU richness in the 0.2–1.6μm 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, Eloe et al., 2010), suggesting that higher diversity may be linked to an increase in niche richness associated with micro-gradients in substrate availability and composition on particles. However, this pattern is not observed across all depths or ocean sites (for example, Figure 2.1a; 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 2.1d 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 overor under-represented 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 2.1c). 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 μm) 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., 2012 a,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 5m sample where Actinobacteria were negligible in both size fractions; Figure 2.1c). A similar enrichment was observed recently in the 0.1–0.8μm 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 2.1d). 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 nonOMZ 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, 200m), 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 5m (Figure 2.1c), but was confined primarily to the prefilter beginning at 32m. 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 5m to Bacillariophyta (diatoms) at 32m 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 2.1c) 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 1000m prefilter, which could be due to changes in the turnover rates of different particle-associated cell fractions (that is, choroplasts 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 70m, and were 8to 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 A.8). 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 A.8), 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 A.8). 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 A.8). 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 2.1d). 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 A.7). In contrast, Planctomycetes represented 1–2% of total sequences in the prefilter at 70m and increased to a peak of B5% at 1000 m (Supplementary Figure A.7). 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 particleassociated 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 2.1b and Supplementary Figure A.4 B,C). Notably, the communities at 5m (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 2.1b and c). Of the OMZ depths, samples from 70, 110 and 200 m were most closely related (Figure 2.1b and Supplementary Figure A.4 B,C). The 320m 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 2.1c). 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 2.1a, 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 2.1b). Ninety-three percent (182/196) of the pairwise comparisons between filter types (p versus s in Supplementary Table A.2, top)

revealed significant differences in taxonomic composition based on phylogenetic relatedness ($P<0.05$; mean Unifrac: 0.53). At only one depth, 320m, did prefilter and Sterivex communities not differ significantly ($P=0.11–0.13$; note similarity in PC1 coordinates in Figure 2.1b), although even in this sample clear differences between filters were evident (Figure 2.1c). 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 (5m) 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 2.1b, Supplementary Figure A.4A). 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 μm) 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 2.1).

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 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.

2.4.3. 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 2.1). 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 2.1), 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 (o0.2 mm) 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 distinguish between prokaryote and eukaryotederived 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.

Table 2.1 Metagenome sequence statistics and taxonomic (Domain) identities

Sample	*Reads*	*Coding reads*[a]	*% Of Bacteria*	*% Of* Archaea	*% Of* unidentified prokaryotes	*% Of* eukaryotes	*% Of* viruses	*% Of other*	*SEED*[b]
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.
[a]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').
[b]No. of prokaryote reads assigned to SEED subsystem categories.

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 2.1) were collapsed to higher taxonomic levels to match groupings based on protein-coding genes (Supplementary Figure A.5). Analysis of 16S amplicons from the four depths where metagenomes were sampled (70, 110, 200 and 1000m; 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 A.5). However, the magnitude of this enrichment was markedly higher in comparisons using amplicon data (Supplementary Figure A.5). 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 A.5 C, 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 A.5). 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 A.5). 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 genebased 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 A.8). Notably, SAR324-like reads peaked at 44% of total bacterial coding reads in the Sterivex fraction at 1000m, 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 A.6). 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 70m

sample at the top of the OMZ. A similar enrichment in the small (0.1–0.8 mm) size fraction was reported for low-oxygen (~20μM) 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 (70m), 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 A.6). 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

A.6). 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 particleassociated 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 110m OMZ depth before declining to 2% beneath the OMZ (Supplementary Figure A.7). 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 A.7). 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 0.2–30μm fraction but not in the larger particle-associated fraction above 30mm (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 2.2, 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 A.3). This similarity reflects redundancy in housekeeping gene categories, notably protein

biosynthesis, DNA repair and central carbohydrate metabolism (Figure 2.2), 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 2.2, 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 2.2 (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 2.2, Supplementary Table A.4). 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 atsignificantly higher proportions, with enrichment driven by genes for branched chain amino-acid transporters (Figure 2.2), 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 A.4). 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 CO2 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 2.3, 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 2.3, 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 2.3 and Supplementary Figure A.7). Sequences matching amoC were affiliated exclusively with the Thaumarchaeota and Crenarcheaota, supporting recent evidence that nitrification in the OMZ is mediated primarily by Archaea (Stewart et al., 2012b).

In contrast, the relative abundance of nxrB, a marker for aerobic nitrite oxidation, did not vary substantially between size fractions (Figure 2.3, inset). However, fraction-

specific patterns were evident when the nxrB pool was evaluated according to taxonomic affiliation. The majority (55%) of all nxrB sequences were most closely related to nxrB of Ca. Nitrospira defluvii, a nitrite oxidizer isolated from activated sludge (Spieck et al., 2006). Nitrospira nxrB abundance peaked at the oxycline base (70 m) in both filter fractions. At the lower depths, including at the anoxic OMZ core, Nitrospira nxrB was found exclusively in the free-living fraction (Figure 2.3 bottom). Subsequent to our analysis, the genome of the nitrite oxidizer Nitrospina gracilis was published (Lücker et al., 2013). Nitrospina 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 reanalysis of our data may instead classify OMZ Nitrospira-like sequences as belonging to Nitrospina. Both Nitrospina 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; Jorgensen et al., 2012), and with the recent detection of nitrite oxidation (by Nitrospina or Nitrococcus bacteria) under low oxygen ($O2 < 1$ μM) 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 1000m sample, but were consistently, albeit marginally, overrepresented in the free-living fraction (Figure 2.3). This pattern was reversed at the oxic 1000m 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 mm) compared with a 0.1–0.8 mm fraction, but occurred at relatively uniform abundance across fractions at low-oxygen (~20 μM) 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 1000m sample (Figure 2.3), whereas nirS genes for the cytochrome cd1-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 N2O and N2 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. Methylomirabilis *oxyfera* (CBE69496.1 and CBE69502.1), a member of the NC10 candidate division originally enriched from sediment (Figure 2.3). Ca. M. *oxyfera* oxidizes methane under anaerobic conditions using O2 generated intracellularly through an alternative denitrification pathway involving the dismutation of nitric oxide into dinitrogen and O2, 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 Methylomirabilis-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 AMPdependent oxidation of sulfite to APS but also acts reversibly during sulfate reduction, was consistently enriched (two-fold) 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 2.3; 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 six-fold more abundant in prefilter compared with Sterivex metagenomes, with

peak abundances at the OMZ core (110 and 200 m; Figure 2.3). 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 B80% 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 2.1), as well as a 2.5-fold enrichment of restriction modification systems and restriction enzymes in prefilters at OMZ depths (70, 110 and 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.

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).

Disclaimer: This chapter was published with the same title along with the supplementary material in Appendix A, in ISME J on 12 September 2013.

Citation: **Ganesh, S., Parris, D. J., DeLong, E. F., & Stewart, F. J.** (2014). Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. *ISME J*, *8*(1), 187-211.

1.5 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.

Alldredge AL, Cohen Y. (1987). Can microscale chemical patches persist in the sea? Microelectrode study of marine snow, fecal pellets. Science 235: 689–691.

Alldredge 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 Gen Microbiol 97: 45–55.

Bengtsson MM, Øvreaș 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-Davo´ 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 CytophagaFlavobacter cluster consuming lowand 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). Archaeaplankton in the Columbia River, its estuary and the adjacent coastal ocean, USA. FEMS Microbiol Ecol 31: 231–239.

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, Faubladier 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, Fadrosh 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 heterogenous 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, Woebken 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 aggregateassociated 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). Heterogenous archaeal communities in the particlerich 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, Melchiorsen 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, Kiørboe 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, Kiørboe 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, Baumberger 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 materal 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, Stuhrmann 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, Balague´ 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.

Ro¨sch 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, Erse´us 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, Yooseph 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 or 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, Escribano 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 II56: 987–991.

van de Vossenberg 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 Ac 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, Fadrosh 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.

Zimmer RL, Woollacott RM. (1983). Mycoplasma-like organisms: occurrence with the larvae and adults of a marine bryozoan. Science 220: 208–210

Zehr JP, Kudela RM. (2011). Nitrogen cycle of the open ocean: from genes to ecosystems. Ann Rev Mar Sci 3: 197–225.

CHAPTER 3

SIZE-FRACTION PARTITIONING OF COMMUNITY GENE TRANSCRIPTION AND NITROGEN METABOLISM IN A MARINE OXYGEN MINIMUM ZONE

3.1 Abstract

The genetic composition of marine microbial communities varies at the microscale between particle- associated (PA; >1.6 μm) and free-living (FL; 0.2–1.6 μm) niches. It remains unclear, however, how metabolic activities differ between PA and FL fractions. We combined rate measurements with metatranscriptomics to quantify PA and FL microbial activity in the oxygen minimum zone (OMZ) of the Eastern Tropical North Pacific, focusing on dissimilatory processes of the nitrogen (N) cycle. Bacterial gene counts were 8- to 15-fold higher in the FL compared with the PA fraction. However, rates of all measured N cycle processes, excluding ammonia oxidation, declined significantly following particle (>1.6μm) removal. Without particles, rates of nitrate reduction to nitrite (1.5–9.4 nM N_2 d^{-1}) fell to zero and N_2 production by denitrification (0.5–1.7nM N_2 d^{-1}) and anammox (0.3–1.9 nM N_2 d^{-1}) declined by 53–85%. The proportional representation of major microbial taxa and N cycle gene transcripts in metatranscriptomes followed fraction-specific trends. Transcripts encoding nitrate reductase were uniform among PA and FL fractions, whereas anammox-associated transcripts were proportionately enriched up to 15-fold in the FL fraction. In contrast, transcripts encoding enzymes for N_2O and N_2 production by denitrification were enriched up to 28-fold in PA samples. These patterns suggest that the majority of N cycle activity, excluding N_2O and N_2 production by denitrification, is confined to a FL majority that is critically dependent on access to particles, likely as a source of organic carbon and

inorganic N. Variable particle distributions may drive heterogeneity in N cycle activity and gene expression in OMZs.

3.2 Introduction

The genetic diversity and activity of pelagic marine microorganisms vary in complex patterns in response to environmental conditions and interspecies interactions. Such variation is best understood at the scale of meters, for example, over depth gradients of light or over regional (kilometer) gradients in nutrient availability. Microbial communities are also structured over much smaller spatial scales, notably the micron-scale distances separating free-living (FL) cells from those on the surfaces or within the diffusive boundary layer of suspended or sinking organic particles (Stocker, 2012). Extensive research has identified consistent taxonomic compositional differences between FL and particle-associated (PA) bacterial and archaeal communities (DeLong et al., 1993; Hollibaugh et al., 2000; Ganesh et al., 2014), as well as differences in total microbial abundance, production and enzyme activity (Simon et al., 2002; Grossart et al., 2003, 2007). Surprisingly, most studies analyzing microbial genomic or metagenomic composition or rates of specific metabolic processes have either focused on microbes within a single biomass size fraction (typically the 0.2–3.0-μm range) or on bulk water samples. Such studies risk excluding important information about contributions from PA cells, larger non-PA cells or metabolic substrates in PA fractions. Furthermore, only a single study (Satinsky et al., 2014) has examined community gene expression (metatranscriptomes) in PA versus FL communities. When coupled to measurements of

metabolic rates, size-fractionated metatranscriptomics can help identify the drivers, magnitudes and spatial scales of biogeochemistry in the oceans.

The extent to which microbial metabolisms are partitioned between PA and FL communities remains uncharacterized for many ocean regions of importance to global biogeochemical cycles. These include the marine oxygen minimum zones (OMZs) where microbial communities control key steps in nitrogen (N), carbon and sulfur transformations. OMZs form in poorly ventilated regions where microbial respiration fueled by high productivity in the overlying waters creates a layer of oxygen depletion. The largest OMZs occur in areas of persistent nutrient upwelling, as in the Eastern Tropical North Pacific (ETNP) off Mexico (Karstensen et al., 2008). Encompassing ~ 12 $\times 10^6$ km^2 of shelf and off-shelf waters south of Baja California, the ETNP OMZ is the largest of the major permanent OMZs (41% of total OMZ area; Paulmier and Ruiz, 2009), with dissolved O2 concentrations falling below the detection limit (<0.1 μM) at mid-water depths (~150–750 m; Cline and Richards, 1974; Tiano et al., 2014). Oxygen drawdown in OMZs significantly restructures the pelagic ecosystem, resulting in microbedominated communities expressing diverse microaerophilic or anaerobic metabolisms. These include the reductive processes of denitrification and anaerobic ammonia oxidation with nitrite (anammox), whose combined activity in OMZs contribute up to half of oceanic N loss as N_2 or N_2O gas (Codispoti et al., 2001; Gruber, 2004). Research in the eastern Pacific OMZs, notably in the anoxic OMZ off Chile and Peru, indicates that anammox and denitrification are linked to varying degrees with a diverse set of other microbial N transformations, including the production of nitrite and nitrate via nitrification along the oxycline and dissimilatory nitrate reduction to ammonium

within the OMZ (Lam et al., 2009; Kalvelage et al., 2013). OMZ N cycle processes also mediate key steps in other elemental cycles, for example, the fixation of carbon by anammox and nitrifying autotrophs, the remineralization of organic carbon and nutrients by denitrifying heterotrophs, and the oxidation of reduced sulfur compounds by chemoautotrophic denitrifiers (Ulloa et al., 2012).

The distributions of these networked metabolic interactions at multiple spatial scales are becoming understandable for some OMZs. Surveys combining marker gene counts and rate measurements have suggested mesoscale patterns in N cycling along coastal to offshore gradients in the Eastern Tropical South Pacific (ETSP) OMZ off Chile and Peru (Kalvelage et al., 2013), as well as clear depth-specific variation in N cycle genes, transcripts, and rates (Thamdrup et al., 2006; Dalsgaard et al., 2012; Stewart et al., 2012). Recently, analysis of microbial metagenomes in the ETSP OMZ also identified a non-uniform distribution of N cycle genes among filter size fractions (Ganesh et al., 2014). Genes for aerobic ammonia oxidation at OMZ boundaries and for anammox at the OMZ core were consistently overrepresented in the FL fraction (0.2–1.6 μm). In contrast, genes mediating different steps of denitrification showed variable distributions between size fractions. Genes encoding NarG, part of the enzyme catalyzing nitrate reduction to nitrite, were abundant and evenly distributed between PA (>1.6 μm) and FL fractions within the OMZ, whereas genes mediating the two terminal steps of denitrification, the reduction of nitric oxide to nitrous oxide (norB) and nitrous oxide to dinitrogen (nosZ), exhibited a four-fold increase in relative abundance in PA compared with FL metagenomes. These data highlight a potential separation of linked metabolic processes

at varying spatial scales, potentially over micron gradients separating PA and FL communities.

The proportional contribution of PA N metabolism to bulk process rates in OMZs remains unclear, notably for the larger ETNP OMZ. Despite recent studies describing the diversity and activity of key functional taxonomic groups in the ETNP (Podlaska et al., 2012; Rush et al., 2012; Beman et al., 2012, 2013), this region remains under-explored from a microbiological and molecular perspective. Assessing the actual biochemical significance of microspatial separation of microbial processes in OMZs requires quantification of metabolic rates and gene expression among size fractions from diverse OMZ regions. We combine community transcription profiling and rate measurements to quantify size fraction partitioning of key steps of the marine N cycle. These data, coupled with taxonomic analysis of genes and transcripts and estimates of bacteria abundance among size fractions, indicate a significant contribution of particles and PA bacteria to OMZ N cycling.

3.3 Materials and Methods

3.3.1 Molecular analysis

Sample collection. Samples were collected from the ETNP OMZ during the OMZ Microbial Biogeochemistry Expedition cruise (R/V New Horizon, 13–28 June 2013). Seawater was sampled from six depths spanning the upper oxycline (30m), lower oxycline (85m) oxic–nitrite interface (91m), secondary chlorophyll maximum (100 m), secondary nitrite maximum (125 m) and OMZ core (300 m) at Station 6 (18° 54.0′ N, 104° 54.0′ W) on 19 June. Collections were made using Niskin bottles on a rosette

containing a Conductivity Temperature Depth profiler (Sea-Bird SBE 911plus, Sea-Bird Electronics Inc., Bellevue, WA, USA) equipped with a Seapoint fluorometer (Seapoint Sensors Inc., Exeter, NH, USA) and SBE43dissolved oxygen sensor (Sea-Bird Electronics Inc.). High-resolution Switchable Trace amount OXygen sensors (STOX) were also mounted to the rosette to quantify oxygen at nanomolar concentrations (Revsbech et al., 2009; Supplementary Methods). Samples were also collected from 20, 40, 90, 100, 125 and 300 m during a replicate cast on the same day and used to count bacterial 16S rRNA genes.

Size-fractionated biomass was collected for RNA analysis by sequential inline filtration of seawater (~10–15 l) through a nylon disk filter (47 mm, 30μm pore size, Millipore, Bellerica, MA, USA), a glass fiber disc filter (GF/A, 47 mm, 1.6μm pore size, Whatman, GE Healthcare Bio-sciences, Pittsburgh, PA, USA) and a primary collection filter (Sterivex, 0.22μm pore size, Millipore) via peristaltic pump. Replicate filters for DNA were collected from equivalent water volumes following RNA collection. Water samples (45ml each) for 16S rRNA gene counts were filtered through 0.2μm disc filters (25 mm, cellulose nitrate, Whatman). Disc filters were transferred to cryovials containing RNA stabilizing buffer (25 mM sodium citrate, 10 mM EDTA and 70 g ammonium sulfate per 100 ml solution, pH 5.2) for RNA samples or lysis buffer (50 mM TrisHCl, 40mM EDTA and 0.73 M sucrose) for DNA samples. Sterivex filters were filled with either RNA stabilizing buffer or lysis buffer, capped and flashfrozen (RNA samples). Filters were stored at − 80 °C. Less than 20 min elapsed between sample collection (water on deck) and fixation in buffer.

DNA and RNA extraction. DNA was extracted from disc and Sterivex filters using

a phenol/chloroform protocol as in Ganesh et al. (2014). Cells were lysed by adding lysozyme (2 mg in 40 µl of lysis buffer per filter) directly to the disc filter containing cryovial or to the Sterivex cartridge, sealing the caps/ends and incubating for 45 min at 37 °C. Proteinase K (1 mg in 100µl lysis buffer with 100µl 20% SDS) was added, and cryovials and cartridges were resealed and incubated for 2 h at 55 °C. The lysate was removed and DNA was extracted once with phenol:chloroform:isoamyl alcohol (25:24:1) and once with chloroform:isoamyl alcohol (24:1), and then concentrated by spin dialysis using Ultra-4 (100kDa, Amicon, Millipore) centrifugal filters.

RNA was extracted from disc and Sterivex filters using a modification of the mirVana miRNA Isolation kit (Ambion, Life Technologies, Carlsbad, CA, USA) as in Stewart et al. (2012). Filters were thawed on ice and RNA stabilizing buffer was removed by pipette from cryovials or expelled via syringe from Sterivex cartridges and discarded. Cells were lysed by adding lysis buffer and miRNA homogenate additive (Ambion) directly to the cryovial or cartridge. Following vortexing and incubation on ice, lysates were transferred to RNase-free tubes and processed via acid–phenol/chloroform extraction according to the kit protocol. The TURBO DNAfree kit (Ambion) was used to remove DNA and the extract was purified using the RNeasy MinElute Cleanup Kit (Qiagen, Hilden, Germany).

16S rRNA gene counts. Quantitative PCR was used to count bacterial 16S rRNA genes in bulk seawater (biomass 40.2µm) and in 0.2–1.6 and 1.6–30µm size fractions. Extracts from the >30 µm fraction did not contain measureable DNA (via Qubit) and failed to yield signals during quantitative PCR with diluted and undiluted extracts. We were not able to sum counts from size-fractionated samples to obtain counts per volume

because exact volumes of filtered seawater were not recorded. We instead used samples of fixed volume from a replicate cast ('Sample collection' section above) to quantify total gene copies per ml of seawater. Because equal water volumes passed through sequential filters during size fractionation, counts from size-fractionated samples were used to quantify the ratio of copies in the 0.2–1.6-μm fraction to those in the 1.6–30-μm fraction (FL/PA ratio); counts from the >30μm fraction are assumed to be negligible.

Quantitative PCR used TaqMan-based reagents and universal bacterial 16S primers 1055f and 1392r, as in Ritalahti et al. (2006) and Hatt et al. (2013). Tenfold serial dilutions of DNA from a plasmid carrying a single copy of the 16S rRNA gene (from Dehalococcoides mccartyi; Ritalahti et al., 2006) were included on each quantitative PCR plate and used to generate standard curves. Assays were run on a 7500 Fast PCR System and a StepOnePlus Real-Time PCR System (Applied Biosystems, Life Technologies). All samples were run in triplicate (20 μl each) and included 1 × TaqMan Universal PCR Master Mix (Life Technologies), 300 nM of primers, 300 nM of TaqMan MGB probe (Life Technologies) and 2μl of template DNA. Thermal cycling involved incubation at 50°C for 2 min to activate uracil-N-glycosylase, followed by 95 °C for 10 min to inactivate uracil-N-glycosylase, denature template DNA and activate the AmpliTaq Gold polymerase, followed by 40 cycles of denaturation at 95 °C (15 s) and annealing at 60 °C (1 min).

16S rRNA gene amplicon sequencing. High-throughput sequencing of dual-indexed PCR amplicons encompassing the V4 region of the 16S rRNA gene was used to assess bacterial community composition in all filter fractions. Despite low DNA yields (above), 16S gene fragments were amplifiable from the 430-μm fraction. The diversity of

amplicons generated using Archaea-specific primers was not evaluated. Amplicons were synthesized using Platinum PCR SuperMix (Life Technologies) with primers F515 and R806 (Caporaso et al., 2011). Both forward and reverse primers were barcoded and appended with Illumina-specific adapters according to Kozich et al. (2013). Thermal cycling involved: denaturation at 94°C (3 min), followed by 30 cycles of denaturation at 94°C (45s), primer annealing at 55°C (45s) and primer extension at 72°C (90s), followed by extension at 72°C for 10min. Amplicons were analyzed by gel electrophoresis to verify size (~400 bp) and purified using the QIAQuick PCR Clean-Up Kit (Qiagen). Amplicons from different samples were pooled at equimolar concentrations and sequenced on an Illumina MiSeq (Illumina, San Diego, CA, USA) using a 500 cycle kit with 5% PhiX as a control.

cDNA synthesis and metatranscriptome sequencing. Shotgun Illumina sequencing of community cDNA was used to characterize gene expression in 0.2– 1.6μm and 1.6–30μm biomass fractions. Most samples of the 430-μm fraction did not yield sufficient RNA for sequencing; this fraction was excluded from meta-omic analysis. Community RNA was prepared for sequencing using the ScriptSeq v2 RNA-Seq Library preparation kit (Epicenter, Madison, WI, USA). Briefly, cDNA was synthesized from fragmented total RNA (rRNA was not removed) using reverse transcriptase and amplified and barcoded using ScriptSeq Index PCR Primers (Epicenter) to generate single-indexed cDNA libraries. cDNA libraries were pooled and sequenced on an Illumina MiSeq using a 500 cycle kit (Supplementary Table B.1). Metagenomes from coupled DNA samples were analyzed to allow standardization of taxon transcript levels relative to abundance in the DNA pool. These data sets were generated in a separate

study (NCBI accession SRP044185) using the Nextera XT DNA Sample Prep kit and paired-end MiSeq sequencing as above for cDNA samples.

Sequence analysis—16S rRNA gene amplicons. Amplicons were analyzed using QIIME (Caporaso et al., 2010) following standard protocols. Barcoded sequences were de-multiplexed and filtered to remove low-quality reads (Phred score < 25). Pairedend sequences were merged using custom scripts incorporating the FASTX toolkit (http://hannonlab. cshl.edu/fastx_toolkit/index.html) and USEARCH algorithm (Edgar, 2010), with criteria of minimum 10% overlap and 95% nucleotide identity within the overlapping region. Merged sequences were clustered into operational taxonomic units at 97% sequence similarity using open-reference picking with the UCLUST algorithm (Edgar, 2010), with taxonomy assigned to representative operational taxonomic units from each cluster using the Greengenes database (DeSantis et al., 2006). Operational taxonomic unit counts were rarefied (10 iterations) and alpha diversity was quantified at a uniform sequence depth (n=6506) using the phylogenetic diversity metric of Faith (1992). To compare community composition between samples, sequences were aligned using the PyNAST aligner (Caporaso et al., 2010) and beta diversity was calculated using the weighted Unifrac metric (Lozupone and Knight, 2005). Sample relatedness based on Unifrac was visualized with a two-dimensional principal coordinate analysis.

Sequence analysis—metatranscriptomes and metagenomes. Analysis of protein-coding genes and transcripts followed that of Stewart et al. (2012) and Ganesh et al. (2014). Reads were filtered by quality and merged as described above. rRNA transcripts were identified using riboPicker (Schmieder et al., 2012) and removed. Merged non-

rRNA sequences were queried via BLASTX against the NCBI-nr database (November 2013). BLASTX matches to bacterial and archaeal genes (>bit score 50) were retained to evaluate marker genes of dissimilatory N metabolism: ammonia monooxygenase (amoC), nitrite oxidoreductase (nxrB), hydrazine oxidoreductase (hzo), nitrate reductase (narG), nitrite reductase (nirK+nirS), nitric oxide reductase (norB) and nitrous oxide reductase (nosZ). N gene transcript abundances were normalized based on gene length and expressed as a proportion of the abundance of transcripts matching the gene encoding RNA polymerase subunit B (rpoB), as has been done in studies of diverse bacteria (for example, Schumann et al., 2010; Ceja-Navarro et al., 2014; Dalsgaard et al., 2014; Eldholm et al., 2014). Although rpoB expression can vary (Vandecasteele et al., 2001), rpoB appears to be one of the more stably expressed housekeeping genes (Sue et al., 2004; Sihto et al., 2014). Furthermore, it has been shown that rpoB can be a proxy of bulk mRNA transcription level for bacteria (Milohanic et al., 2003; Sue et al., 2004)—rpoB-normalized values therefore reflect transcription of a target gene relative to a housekeeping gene under a given condition/sample.

The taxonomic composition of protein-coding genes and transcripts was determined using MEtaGenome ANalyzer 5 (Huson et al., 2011) based on the annotations of BLASTX-identified genes, according to the NCBI taxonomy. The proportional contribution of a taxon (Phyla and Class levels) to the RNA pool was calculated as an RNA:DNA ratio, calculated as in Frias-Lopez et al. (2008) and Stewart et al. (2012) to account for variation in taxon abundance in the DNA pool: (protein-coding RNA reads per taxon/total protein-coding RNA reads)/(protein-coding DNA reads per taxon/total protein-coding DNA reads). For subsets of transcripts matching specific functional taxa

(anammox genera, Nitrospina), genes differentially expressed between FL and PA fractions were identified using baySeq (Hardcastle and Kelly, 2010) as in Ganesh et al. (2014). Full details regarding sequence analysis are in the Supplementary Methods.

Amplicon and metatranscriptome sequences are available through NCBI under BioProject ID PRJNA263621

3.3.2 Biochemical analysis

Water for rate measurements was taken from Niskin bottles immediately after arrival on deck and transferred to 2l glass bottles. Bottles were overflowed (three volume equivalents) and sealed without bubbles using deoxygenated butyl rubber stoppers (De Brabandere et al., 2014). Bottles were stored in the dark at in situ temperature until experimentation (o6 h). Each bottle was purged with a gas mixture of helium and carbon dioxide (800 p.p.m. carbon dioxide) for ~ 20 min. Under a slight overpressure, water was dispensed into 12 ml exetainers (Labco, Lampeter, Ceredigion, UK) either directly or through helium-flushed inline filter holders, and immediately capped with deoxygenated lids.

Rates were measured for two biomass fractions: bulk water containing all particles (no filtration) and water without particles 41.6 μm. Particles were excluded via filtration as described above. Exetainers were subsequently amended with ^{15}N-labeled substrate by injection through the septum. Three ^{15}N amendments were carried out for all samples: $^{15}NH^{+}_{4}$ addition (to 5 μM concentration), $^{15}NO2^{-}$ addition (5 μM) and $^{15}NO^{-}_{3}$ addition (15 μM). Headspaces of 2ml were introduced into each exetainer and

flushed twice with the helium–carbon dioxide mixture as in De Brabandere et al. (2014). For each experiment, triplicate exetainers were preserved with 100 µl of 50% (w/v) $ZnCl_2$ at the start of the incubation and again at the end point (27 h). Oxygen concentration was o80 nM in all treatments. The act of filtering did not affect exetainer oxygen content (Supplementary Methods).

The production of $^{14}N^{15}N$ and $^{15}N^{15}N$ was determined on a gas chromatography isotope ratio mass spectrometer as in Dalsgaard et al. (2012). Rates of N_2 production by anammox and denitrification were calculated using the equations in Thamdrup and Dalsgaard (2002). Nitrite oxidation was determined from the production of $^{15}NO^-_3$ from incubations with $^{15}NO^-_2$. After removal of unused $^{15}NO^-_2$ using sulfamic acid, $^{15}NO^-_3$ was converted to $^{15}NO^-_2$ with cadmium and then to N_2 with sulfamic acid (Füssel et al., 2012; McIlvin and Altabet 2005). Ammonia oxidation and nitrate reduction rates were determined as production of $^{15}NO^-_2$ from incubations with $^{15}NH^+_4$ or $^{15}NO^-_3$. $^{15}NO^-_2$ produced during incubations was converted to N_2 with sulfamic acid. N_2 was then analyzed on the gas chromatography isotope ratio mass spectrometer. Rates for all processes were calculated from the slope of the linear regression of $^{14}N^{15}N$ or $^{15}N^{15}N$ with time. T-tests were applied to determine whether rates were significantly different from zero (Po0.05). Nonsignificant rates are presented as not detected.

Nitrate, nitrite and ammonium concentrations were determined using standard protocols, as described in the Supplementary Methods. The N deficit estimates the

amount of fixed N loss and was calculated as: N deficit = (DIN) – (16 × [PO_4^{3-}] × 0.86), where DIN = [NO_3^-]+ [NO_2^-]+[NH_4^+]. The value of 0.86 accounts for the release of PO_4^{3-} from organic N remineralized during denitrification (Codispoti et al., 2001).

3.4 Results and Discussion

Particle removal significantly altered N cycle rates and gene expression, suggesting physiological variation at the microscale between water column and PA niches. These patterns are discussed below relative to oxygen and N concentrations, bacterial abundance and community taxonomic composition.

Oxygen and nitrogen. Oxygen concentrations recorded by the SBE43 sensor were highest (~200μM) in surface waters and declined steeply in the oxycline between 30 and 85 m (Figure 3.1a). From 90 to 800 m, concentrations measured by STOX sensors were < 50 nM, with the majority close to or below the detection limit (Figure 3.1b) representing the OMZ core. Oxygen concentrations at the six depths sampled for size fractionation work varied markedly, from 200μM at 30 m to 229 nM at 85 m, representing the upper and lower limits of the oxycline, respectively, 43 nM at the oxic–nitrite interface (91 m), 21 nM at the secondary chlorophyll maximum (100 m) and 14 nM at the secondary nitrite maximum (125 m), before reaching the detection limit (9 nM) at 150 m and staying undetectable to 300m in the OMZ core (Figure 3.1b). Oxygen profiles coincided with patterns in nitrate, nitrite and the N deficit. As commonly observed in anoxic OMZs (Ulloa et al., 2012), a prominent secondary nitrite maximum (up to 6.1 μM) was observed as oxygen declined, coinciding with local nitrate and N deficit minima (Figures 1a and c).

Ammonium concentrations remained below 155 nM throughout the profile. Chlorophyll peaked at 0.6 and 0.8 $\mu g\ l^{-1}$ at 50 and 100 m, respectively, indicating the primary and secondary chlorophyll maxima characteristic of OMZs (Figure 3.1c; Ulloa et al., 2012).

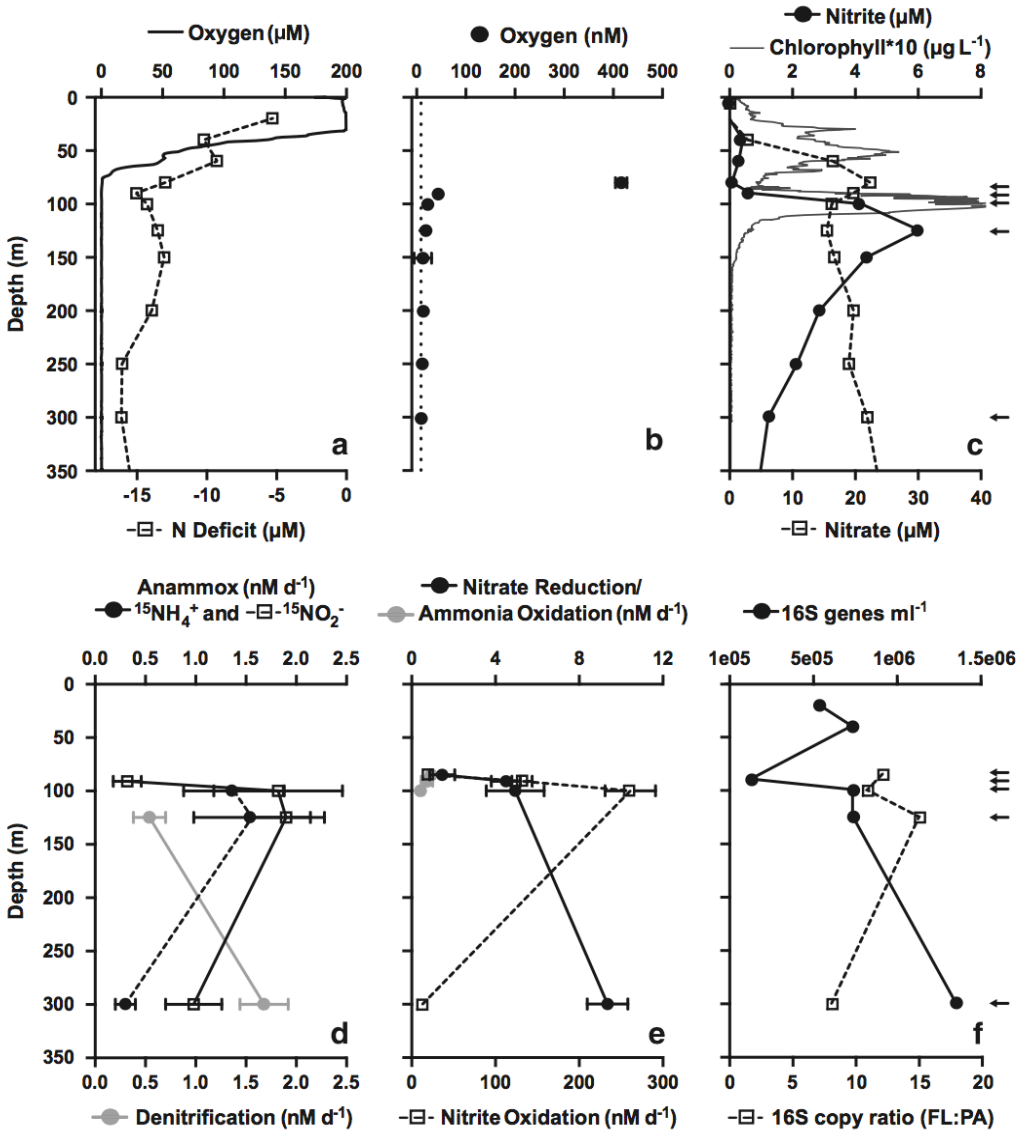

Oxygen (µM)
Oxygen (nM)
Nitrite (µM)
Chlorophyll*10 (µg L-1)
Depth (m)
a
b
c
N Deficit (µM)
Nitrate (µM)
Anammox (nM d-1)
15NH4+ and 15NO2-
Nitrate Reduction/
Ammonia Oxidation (nM d-1)
16S genes ml-1
d
e
f
Denitrification (nM d-1)
Nitrite Oxidation (nM d-1)
16S copy ratio (FL:PA)

Figure 3.1 Vertical profiles of hydrochemical parameters, bulk N cycling process rates and bacterial 16S rRNA gene counts at Station 6 off Manzanillo, Mexico on 19 June 2013. The N deficit (a, open squares) is shown relative to oxygen concentrations, measured during the upcast profile with both a SBE43 sensor (a; solid line) and STOX sensors (b), and to profiles of nitrite, nitrate and chlorophyll concentration (c). The dashed line in b indicates the detection limit of the STOX sensors (~9 nM). N cycle process rates (d, e) are bulk rates (no size fractionation) in nM N per day. (f) Quantitative PCR-based counts of 16S rRNA gene copies per ml of bulk seawater (all biomass >0.2 μm) and the ratio of gene counts in the FL (0.2–1.6 μm) biomass fraction to counts in the PA (1.6–30 μm) fraction. Arrows (c, f) indicate the depths sampled for rate measurements. Error bars represent the standard error.

16S rRNA gene abundance. Counts of bacterial 16S rRNA genes varied by an order of magnitude with depth (Figure 3.1f). Bulk seawater counts declined from ~ 6×10^5 per ml in the primary chlorophyll maximum and upper oxycline to 1.3×10^5 at the oxic–nitrite interface (90 m), before increasing to a maximum of 1.4×10^6 at the OMZ core. This pattern agrees with trends from other OMZ sites, where particulate backscattering and bacterial load show local maxima in the photic zone and again in the OMZ (Spinrad et al., 1989; Naqvi et al., 1993; Whitmire et al., 2009). At all depths, gene numbers were substantially higher in the 0.2– 1.6 μm (FL) fraction compared with the 1.6–30 μm (PA) fraction, with FL/PA ratios from 8 to 15 (Figure 3.1f). Elevated microbial biomass in FL compared with PA fractions has been observed in diverse ocean regions (Cho and Azam 1988; Karl et al., 1988; Turley and Stutt 2000; Ghiglione et al., 2009). Here, FL/PA was lowest at 300 m, suggesting an increased contribution of PA or large bacteria within the OMZ core, possibly related to increases in particle load or size.

Community composition—16S rRNA gene amplicons. Community 16S rRNA gene diversity varied substantially among biomass fractions and depths. Unifrac-based clustering indicated that size class was a stronger predictor of community relatedness compared with depth, notably for samples from the OMZ layer (100, 125 and 300 m) (Figure 3.2). Phylogenetic diversity differed widely among fractions, with values in the two PA fractions up to 50% greater than those for the corresponding FL fraction (Supplementary Figure B.1). Elevated diversity in PA communities has been reported for the ETSP OMZ (Ganesh et al., 2014) and may be driven by increased niche richness due to substrate heterogeneity on particles, or to the exposure of particles during sinking to

diverse communities in different biogeochemical zones. Here, phylogenetic diversity in all size fractions peaked at the secondary chlorophyll maximum, similar to trends reported for non-fractionated samples (biomass >0.2 μm) from ETNP sites north of the study area, where taxon richness peaked just below the oxycline, before declining with depth (Beman and Carolan, 2013). That study identified deoxygenation as a driver of bulk community diversity, whereas our results suggest that diversity estimates vary depending on the filter fraction examined.

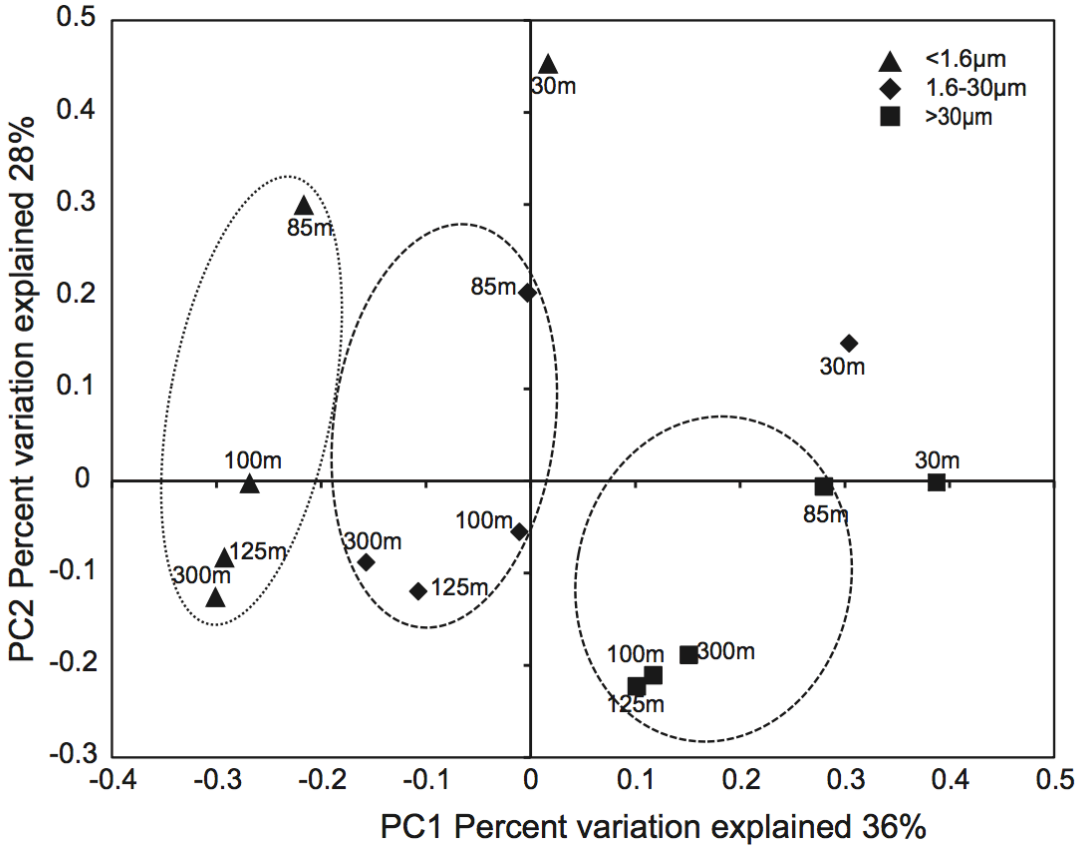


Figure 3.2 Principle coordinate analysis of community taxonomic relatedness based on 16S rRNA gene amplicons, as quantified by the weighted Unifrac metric. Samples from different filter size fractions are circled (excluding outliers from the 30m depth).

Several microbial clades with roles in N cycling were proportionally enriched in FL relative to PA fractions (Figure 3.3; Supplementary Figure B.2).

Sequences matching ammonia-oxidizing Thaumarchaeota and nitrite-oxidizing Nitrospina (classified as Deltaproteobacteria in Supplementary Figure B.3; Teske et al., 1994) peaked in relative abundance (2.2–2.9% of total) in the FL fraction from the lower oxycline before declining with depth. Abundances of both groups decreased by 50–60% in the 1.6–30-μm PA fraction relative to the FL fraction and were negligible in the 430-μm PA fraction. Sequences matching the anammox Planctomycete Candidatus Scalindua (Order Brocadiales) were also confined primarily to the FL fraction, peaking at 14.1% of total FL amplicons at 300 m, compared with 0.3–1.5% in PA fractions. Anammox bacteria were previously shown to be enriched in the smallest size fractions in studies of the Chilean OMZ (Ganesh et al., 2014), the Black Sea (Fuchsman et al., 2012), and an OMZ off Costa Rica (Kong et al., 2013), but have also been detected in direct contact with particles in oxygen-depleted waters over the Namibian shelf (Woebken et al., 2007).

The composition of the ETNP bacterial community with potential roles in coupled N and sulfur cycles differed from that of other OMZs and varied among size fractions. The sulfur-oxidizing and denitrifying Gammaproteobacterial SUP05 clade, a ubiquitous inhabitant of low-oxygen waters (Wright et al., 2012), accounted for ~0.2% of amplicons in ETNP samples. ETNP samples were instead enriched in members of the uncultured SAR406 and Deltaproteobacterial SAR324 lineages (Figure 3.3), as well as Gammaproteobacteria of the Thiohalorhabdales, Chromatiales and Thiotrichales (Supplementary Figure B.3 and Supplementary Information), all of which may contain sulfur-oxidizing members. The SAR406 cluster was particularly abundant in FL samples,

contributing 22% (average) of FL sequences, compared with 2.8% in PA fractions (Figure 3.3). SAR406 may participate in sulfur cycling via dissimilatory polysulfide reduction to sulfide or sulfide oxidation (Wright et al., 2014), although pathways for denitrification and carbon fixation have not been unambiguously identified. The SAR324 lineage, which includes members with genes for sulfur chemolithoautotrophy and nitrite reduction (Swan et al., 2011; Sheik et al., 2014), was also enriched in FL samples, representing 5–8% of sequences at OMZ depths, compared with o0.5% in PA fractions (Supplementary Figure B.3). FL enrichment of SAR406 and SAR324 is consistent with an autotrophic lifestyle, as autotrophs would not require organic particles for carbon acquisition. Whereas sulfide-oxidizing autotrophs in other OMZs are capable of using oxidized N compounds as terminal oxidants (Walsh et al., 2009; Canfield et al., 2010), the extent to which ETNP clades participate in sulfur-driven denitrification remains unclear.

Many microbial groups common to low-oxygen waters, including those discussed above, were either at low abundance or absent from the >30-μm PA fraction. This fraction was instead enriched in taxa known to associate with the surfaces or guts of plankton or polymer aggregates (for example, fecal pellets), notably Gammaproteobacteria related to Vibrio, Pseudomonas and Alteromonas (Supplementary Figure B.3; Rao et al., 2005; Hunt et al., 2008; Ivars-Martinez et al., 2008). Microorganisms on particles >30 μm are likely a minor component of the bulk OMZ microbial community, and may be relatively transient members, passing through low-oxygen depths attached to sinking aggregates or zooplankton.

RNA:DNA ratios. The taxonomic identities of protein-coding genes and transcripts suggest differences in the proportional contributions of specific taxa to bulk transcript pools in FL versus PA (1.6–30 μm) communities. RNA/DNA ratios for several groups with roles in N cycling, including anammox bacteria (Brocadiales), ammonia-oxidizing Thaumarchaeota and nitrite-oxidizing Nitrospina bacteria, were consistently elevated in FL relative to PA communities (Figure 3.4). These groups were also proportionately enriched in 16S gene pools from FL samples. In contrast, groups such as the Actinobacteria and Gammaproteobacteria exhibited greater RNA/DNA ratios in the PA fraction. Similar patterns were observed for ratios based on 16S rRNA gene amplicons and transcripts from metatranscriptomes (rRNA:rDNA; Supplementary Figure B.4). Although RNA:DNA ratios, based either on target genes or community gene pools, have been used as proxies for metabolic activity (Campbell and Yu, 2011; Hunt et al., 2013; Satinsky et al., 2014), linking ratios to activity is tenuous. Notably, as RNA/DNA ratios are based on proportional abundances, increases in taxon representation in the transcriptome can reflect either increases in the activity of the target taxon or decreases in the activity of other organisms. Furthermore, using ratios to infer activity is challenging for broad divisions such as the Gammaproteobacteria that contain taxonomically and functionally diverse members, as the relationship between cell RNA content and activity is unlikely to be consistent across members or activity gradients (Blazewicz et al., 2013). Here, the composition of Gammaproteobacteria at the Order level varied widely (Supplementary Figure B.3), highlighting the possibility that taxon-specific differences confounds interpretation of Phylum-level ratios. Studies at finer levels of taxonomic resolution utilizing transcriptome mapping to reference genomes coupled with

measurements of absolute transcript numbers are necessary to confirm shifts in taxon-specific activity between FL and PA niches. The latter could reflect differences in substrate conditions or the proportions of live versus dead cells between fractions.

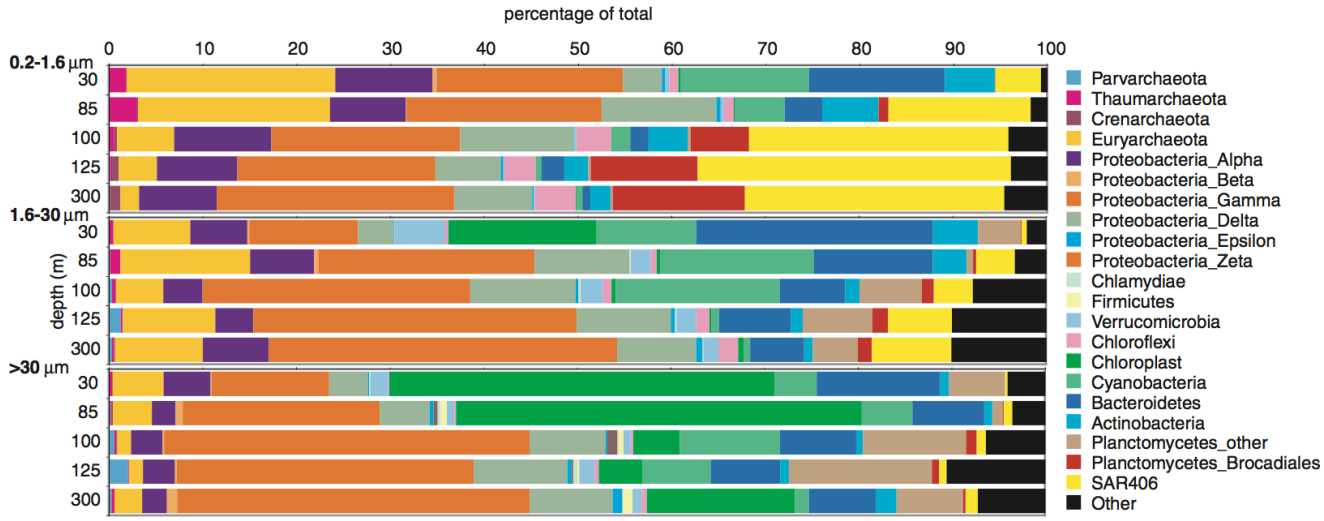


Figure 3.3 ETNP OMZ bacterial community diversity revealed by 16S rRNA gene Illumina sequencing. The abundances of major bacterial divisions are shown as a percentage of total identifiable 16S rRNA gene sequences. 'Other' includes 34 divisions, including unassigned sequences.

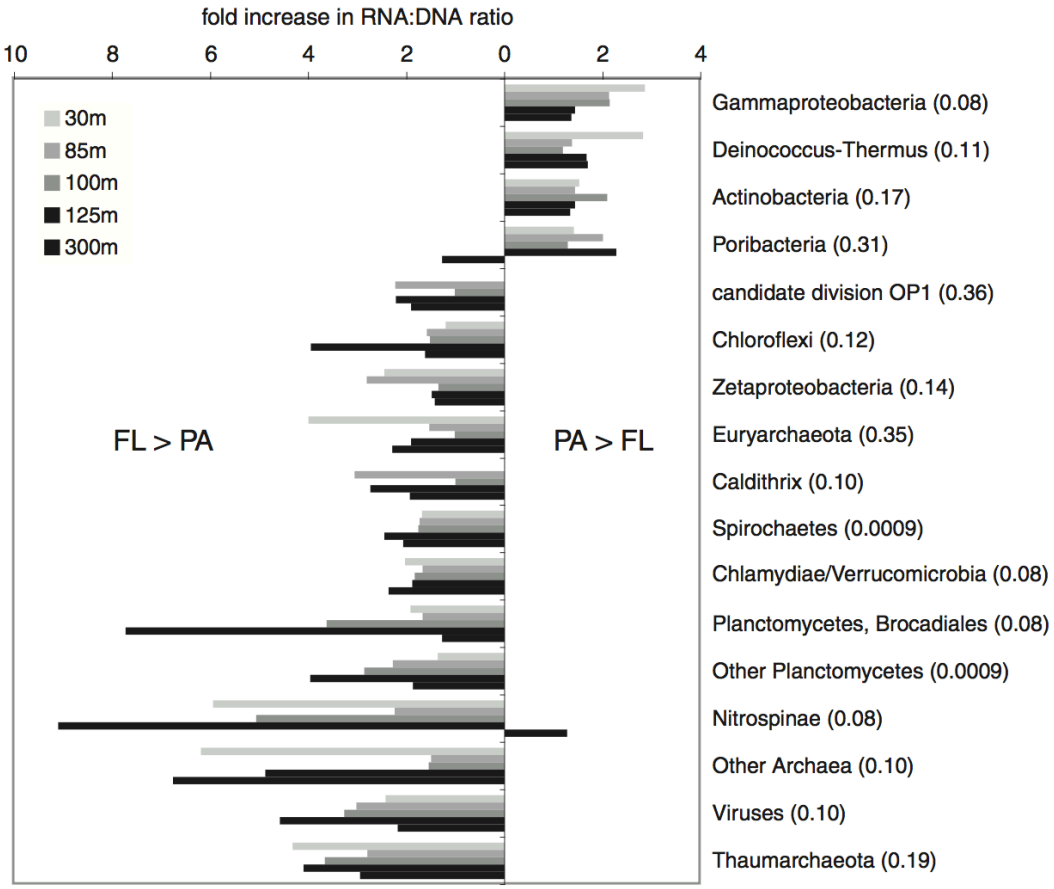


Figure 3.4 Fold increase in taxon RNA/DNA ratios between PA (1.6–30 μm) and FL (0.2–1.6 μm) communities. Plot shows the fold increase in ratios for major taxonomic groups identified by the NCBI annotations of protein-coding genes and transcripts in coupled metatranscriptome and metagenome data sets. Values right of zero indicate higher ratios in PA communities (PA/FL>1). Values left of zero indicate higher ratios in FL communities (FL/PA>1). The plot shows only groups with relative abundance in metagenome data sets greater than 0.2% of total sequences, and only groups with an average (across all five depths) fold increase >1.5. Numbers in parentheses are false discovery rate q-values for t-test comparisons of PL and FL ratios, calculated as in Storey and Tibshirani (2003).

Bulk biochemical rates. N transformation rates in bulk water (no filter fractionation) varied among processes and over depth gradients. Rates of nitrate reduction to nitrite increased with depth and diminishing in situ oxygen concentration, with measured rates from 1.5 to 9.4 nM N d^{-1} (Figure 3.1e). The reduction of nitrite to N_2 via denitrification was slower and measurable only below 125m, with a maximum of 1.7nM N d^{-1} at 300 m (Figure 3.1d). Anammox was measurable between 91and 300m at rates from 0.3 to1.9 nMNd^{-1},with the highest rates in the secondary nitrite maximum. Anammox rates after $^{15}NO_2^-$ addition were consistently higher than those after $^{15}NH^+_4$ addition, similar to trends in De Brabandere et al. (2014) working in the ETSP OMZ. The largest discrepancy in $^{15}NO^-_2$ versus $^{15}NH^+_4$ -based anammox rates occurred at 300 m, the depth where maximum denitrification rates were measured (Figure 3.1d), suggesting that nitrite shunting by denitrifiers could explain the offset in anammox rates, as proposed by De Brabandere et al. (2014). In this interpretation, incubations with $^{15}NH^+_4$ reflect true anammox rates, while incubations with $^{15}NO^-_2$ overestimate the process. Our measured N loss rates are similar to those reported by Babbin et al. (2014) for stations close to ours (0.8–4.7 and 0.9–3.3 nM N d^{-1} for anammox and denitrification, respectively), and are generally consistent with those recorded in other OMZs (Lam and Kuypers, 2011). Dissimilatory nitrate reduction to ammonium was not detected in any of the samples. Ammonia oxidation was detected only at 100 m or shallower, with maximum rates at 85 m (0.92 nM N d^{-1}), the shallowest and most oxygenated of the depths where rates were measured (Figure 3.1e). These rates are at the lower end of those

previously observed in the Gulf of California and ETNP (0–348 nM d^{-1}; Beman et al., 2012) and in the ETSP OMZ (0.2–89 nM d^{-1}; Kalvelage et al., 2013). Nitrite oxidation was detected at all depths except 125 m, at rates (13–261 nM d^{-1}; Figure 3.1e) up to an order of magnitude higher than those for other processes, with rates peaking in the secondary chlorophyll maximum, suggesting a decoupling between ammonia and nitrite oxidation. These values are consistent with nitrite oxidation rates measured in the Gulf of California and ETNP (0– 213nM d^{-1}; Beman et al., 2013), the ETSP (5–928 nM d^{-1}; Kalvelage et al., 2013) and the Benguela upwelling (14–372 nM d^{-1}; Füssel et al., 2012), further confirming the activity of nitrite oxidizers at nanomolar oxygen levels (Füssel et al., 2012).

Size-fractionated biochemical rates and gene expression. Nitrate reduction and denitrification. Rates of nitrate reduction to nitrite were undetectable after removal of particles (>1.6 μm) by filtration (Figure 3.5a). Filtration therefore removed either a significant proportion of total nitrate-reducing cells or a critical substrate source fueling FL nitrate reducers. Although the proportional abundances of transcripts encoding nitrate reductase (narG) were uniform among filter fractions (Figure 3.5b), suggesting roughly equivalent contributions of nitrate reduction to fraction-specific metabolism in PA and FL niches, total transcript counts were still considerably higher in the FL fraction (Figure 3.5c). The taxonomic identities of narG transcripts spanned a wide phylogenetic range (Figure 5d), with sequences related to anammox bacteria and candidate divisions OP3 and OP1 particularly highly represented in both fractions, which is consistent with results from ETSP OMZ metagenomes (Ganesh et al., 2014). The distribution of narG suggests

that declines in nitrate reduction after filtration were likely driven primarily by the removal of particulate organic matter (OM) for use by diverse FL nitrate reducers and to a minor extent by the removal of particle-attached nitrate reducers.

Rates of N_2 production by denitrification also declined significantly (by 85%) following particle removal. Transcripts encoding enzymes for nitrite reduction (nirK/S) and N_2O and N_2 gas production (nosZ and norB, respectively) were at low abundance compared with those of nitrate reduction and anammox (Figure 3.5c), and showed contrasting size fraction distributions. Transcripts for NirK/S were at relatively uniform proportional abundances in PA and FL data sets. These transcripts reflected a diverse assemblage, but were notably enriched at the lower oxycline and oxic–nitrite interface by Thaumarchaeota nirK sequences (Figure 3.5d). NO-forming Nir activity in ammonia-oxidizing Thaumarchaeota is not well understood (Lund et al., 2012), but may have roles in both energy generation and nitrite detoxification, with the latter of potential relevance in nitrite-replete OMZs. In contrast, nosZ and norB transcripts were enriched up to 28-fold in PA fractions, particularly at upper OMZ depths (Figure 3.5b), and affiliated with a community taxonomically distinct from that catalyzing upstream steps of denitrification (Figure 3.5d), as also reported in Dalsgaard et al. (2014). These results support metagenome data from the Chilean OMZ (Ganesh et al., 2014) and from an estuarine site (Smith et al., 2013) showing nor/nos enrichment in PA fractions, suggesting a conserved PA niche. After scaling based on 16S gene counts (a proxy for cell counts), absolute counts of nosZ and norB transcripts were estimated to be comparable between size fractions (Figure 3.5c), suggesting that filtration removes a significant proportion of the N_2O and N_2-forming denitrifier community.

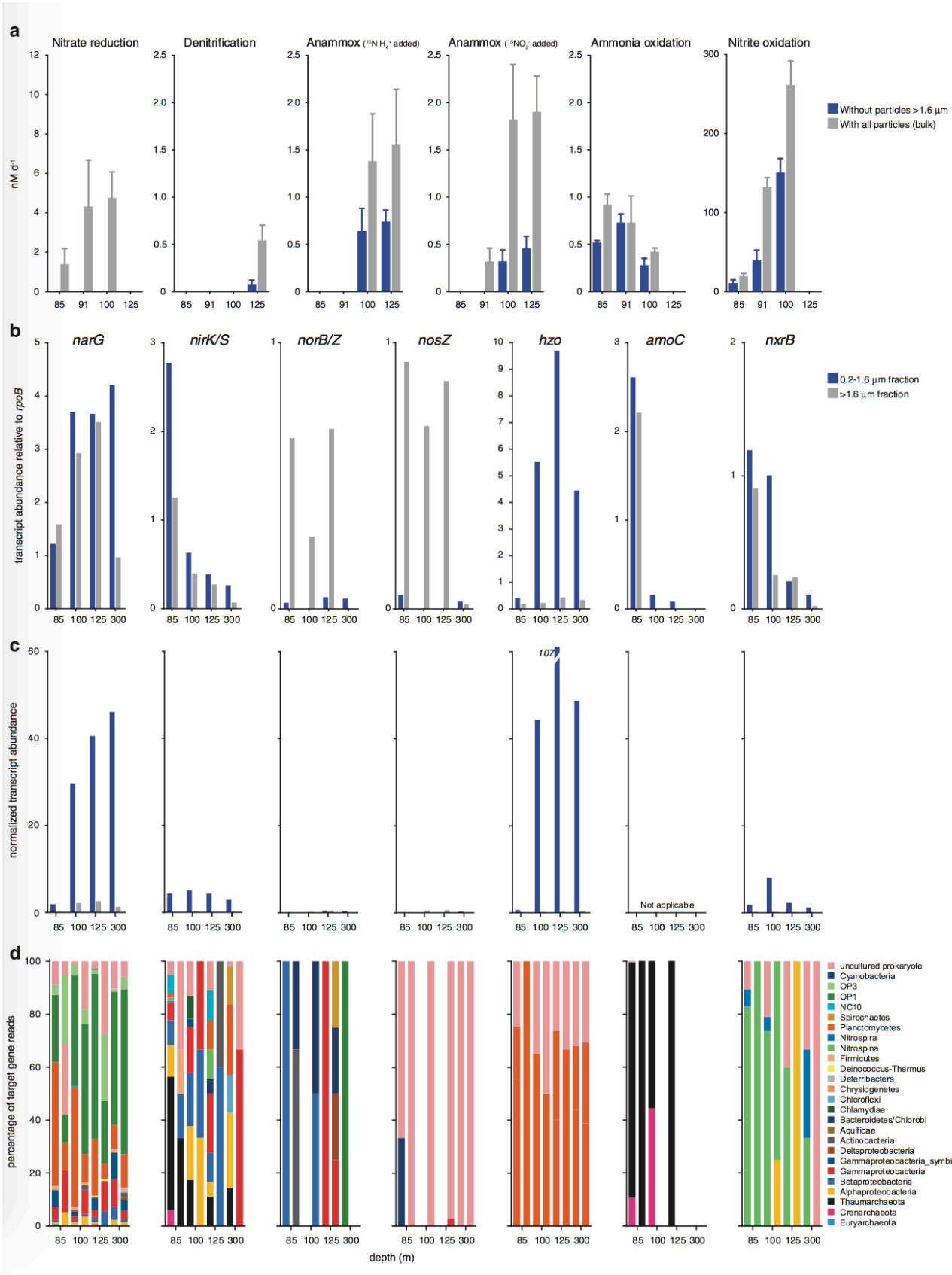

a
Nitrate reduction
Denitrification
Anammox (15N H4+ added)
Anammox (15NO2- added)
Ammonia oxidation
Nitrite oxidation
nM d-1
Without particles >1.6 µm
With all particles (bulk)
b
narG
nirK/S
norB/Z
nosZ
hzo
amoC
nxrB
transcript abundance relative to rpoB
0.2-1.6 µm fraction
>1.6 µm fraction
c
normalized transcript abundance
107
Not applicable
d
percentage of target gene reads
depth (m)
uncultured prokaryote
Cyanobacteria
OP3
OP1
NC10
Spirochaetes
Planctomycetes
Nitrospira
Nitrospina
Firmicutes
Deinococcus-Thermus
Deferribacters
Chrysiogenetes
Chloroflexi
Chlamydiae
Bacteroidetes/Chlorobi
Aquificae
Actinobacteria
Deltaproteobacteria
Gammaproteobacteria_symbionts
Gammaproteobacteria
Betaproteobacteria
Alphaproteobacteria
Thaumarchaeota
Crenarchaeota
Euryarchaeota

Figure 3.5 Measured rates and marker gene transcript levels for major dissimilatory N cycle processes. Process rates (a) are shown for incubations with particles (no filtration) and without particles measuring 1.6 μm or larger. Zero values indicate non-significant rates and error bars represent the standard errors. Note variation in y axis scales. (b) Marker gene transcript abundances in FL (0.2–1.6 μm) and PA (1.6–30 μm) filter fractions. Abundances are calculated as read count per gene per kilobase of gene length, and shown as a proportion of the abundance of transcripts matching the universal, single-copy gene rpoB. A value of 1 indicates abundance equal to that of rpoB. Note variation in y axis scales. (c) Multiplies the values in b by the counts of bacterial 16S genes in Figure 3.3.1f and the FL/PA ratio (for FL counts only), with the values then divided by 1,000,000 for presentation. The resulting values do not reflect absolute counts, but provide an approximation of how variation in bacterial abundance (inferred by proxy from 16S counts) is predicted to affect comparisons of absolute transcript counts across samples, assuming that absolute transcript counts scale with bacterial load. Scaled amoC values are not provided in c, as FL:PA ratios were determined using bacterial-specific primers, and amoC transcripts were dominated exclusively by archaeal sequences; FL/PA ratios do not reflect archaeal contributions. Patterns in c should be interpreted cautiously, as they could change if RNA content per unit biomass varies substantially among samples. (d) The taxonomic affiliations of marker gene transcripts based on NCBI annotations of genes identified as top matches (> bit score 50) via BLASTX.

Anammox. Anammox rates declined 53–100% without particles. These declines may indicate that particle removal eliminates an important substrate source for FL anammox bacteria. Particles could facilitate anammox by providing (1) ammonium remineralized by heterotrophic PA bacteria consuming particulate OM (Shanks and Trent, 1979), (2) ammonium produced by anammox bacteria via oxidation of particulate OM with nitrate or nitrite (Kartal et al., 2007) or (3) nitrite produced by other (non-anammox) nitrate-reducing heterotrophs consuming particulate OM. Explanations (1) and (3) are consistent with the observed reduction in rates of nitrate reduction to nitrite, the most active respiratory process detected, in the absence of particles. We therefore hypothesize that particledependent nitrate or nitrite reduction by both FL and PA communities likely provides a critical supply of ammonium or nitrite for anammox, although it remains unclear whether anammox bacteria are ammoniumor nitrite-limited in OMZs (Lam and Kuypers, 2011; De Brabandere et al., 2014).

Two alternate hypotheses could explain the decline in anammox after filtration. First, filtration removed a significant fraction of total anammoxcapable cells. This hypothesis is dismissed given the substantial enrichments of total 16S rRNA gene copies and Brocadiales 16S copies in the FL fraction.

Second, anammox-capable cells in PA fractions, although a minor component of the total population, contributed disproportionately to bulk anammox rates, potentially overexpressing anammox enzymes compared with FL cells. The active transcription of anammox genes in the FL fraction argues against this hypothesis. Transcripts encoding hydrazine oxidoreductase, catalyzing hydrazine oxidation to N_2 during anammox, were up to 15-fold higher in proportional abundance in FL samples (Figure 3.5b), consistent

with the 14-fold enrichment of Brocadiales 16S genes in the FL fraction.

To further examine the potential that anammox bacteria alter their metabolism between FL and PA niches, we tested for differential gene expression between PA and FL data sets, focusing only on genes from known anammox genera (Supplementary Methods). After normalizing for variation in total transcripts matching anammox genera, PA and FL transcript profiles were well correlated ($R^2 = 0.81$) and no genes (of 1392 total) were differentially expressed between fractions ($P<0.05$). However, testing for DE required grouping data sets from different depths as replicates, raising the possibility that depth-specific variation confounded the detection of DE between fractions. Despite this apparent uniformity in expression, marker genes for the anammox process itself (hzo and hzs encoding hydrazine synthase) were consistently, although not statistically, enriched (2–3 fold) in the FL fraction, suggesting an upregulation of anammox in the FL niche (Supplementary Figure B.5). In contrast, transcripts encoding nitrate reductase NarGH were among those most abundant and enriched in PA profiles. Nar likely acts in reverse in anammox bacteria, converting nitrite to nitrate and helping replenish electrons lost during CO_2 fixation via the acetyl-CoA pathway (Strous et al., 2006; Jetten et al., 2009). Here transcripts encoding key acetyl-CoA pathway genes (formate dehydrogenase and carbon monoxide dehydrogenase) were ~ 10-fold less abundant than Nar transcripts, and did not differ appreciably between PA and FL fractions. Alternatively, Nar in anammox bacteria may function as a traditional nitrate reductase during anaerobic growth on organic compounds (Güven et al., 2005), and it is plausible that oxidation of organics is enhanced in anammox bacteria on particles. The observed fold change differences in hzo, hzs and nar raise the possibility of microniche-specific gene expression by anammox-

capable bacteria, but do not support the hypothesis that the anammox process itself is upregulated in PA compared with FL communities. Thus, the effect of filtration on anammox rates is best explained by dependence of FL anammox bacteria on the PA fraction for substrate.

Nitrification. Compared with denitrification and anammox, ammonia and nitrite oxidation were less affected by particle removal. Filtration decreased ammonia oxidation rates by ~25% in the lower oxycline and upper OMZ. At the oxic–nitrite interface, rates did not change following filtration. Nitrite oxidation decreased by ~ 50% without particles, although rates after filtration were still several orders of magnitude higher than those of other N processes. Transcripts encoding ammonia monooxygenase (amoC) were proportionately enriched in the FL fraction (excluding at 85m) and affiliated predominantly with sequences annotated as Thaumarchaeota or as 'uncultured crenarchaeote' (the latter were entered into NCBI before formal recognition of the 'Thaumarchaeota' designation and likely should be re-annotated as Thaumarchaeota; Figure 3.5d). Transcripts encoding nitrite oxidoreductase (nxrB) did not show a strong size fraction-specific enrichment, consistent with nxrB gene distributions in metagenomes from the Chilean OMZ (Ganesh et al., 2014). NxrB transcripts affiliated primarily with sequences from the genome of Nitrospina gracilis (Lücker et al., 2013), supporting reports of high Nitrospina activity in the ETNP region (Beman et al., 2013). Analysis of only N. gracilis-affiliated transcripts did not reveal any genes that were differentially expressed between FL and PA fractions ($P<0.05$), although low N. gracilis read counts in PA data sets (5% those in FL data sets) likely confounded statistical detection. Indeed, nxr genes (all subunits combined) represented similar proportions of

total N. gracilis counts in both fractions (2.5% and 3.2% in FL and PA, respectively), suggesting that expression of nitrite oxidation genes was not substantially altered between PA and FL niches.

3.5 Conclusions

Filter fractionation significantly altered N metabolism rates, confirming a critical role for particles and PA microorganisms in OMZ N cycling. A strong dependence on particles may help explain heterogeneity of denitrification and anammox rates among OMZ sites and time points (Dalsgaard et al., 2012; De Brabandere et al., 2014; Kalvelage et al., 2013), notably if particle load varies over depth and distance from shore or between times of high versus low export from the photic zone. Our data also suggest differences in how OMZ microbes interact with particles. A significant fraction of some functional groups, notably N_2O and N_2-producing denitrifiers, are recovered in PA fractions, suggesting direct adherence to particles, or potentially larger cell sizes. A tight coupling to particle surfaces could facilitate access to organic carbon for respiration, as well as oxidized N compounds produced by upstream steps of the denitrification pathway. Indeed, enhanced hydrolytic activity of particulate organic substrates has been demonstrated in diverse marine waters (Hoppe et al., 1993; Taylor et al., 2009; Yamada et al., 2012). Denitrifiers are generally known as facultative anaerobes (Zumft, 1997) and populations of denitrifiers on OMZ particles may develop during sinking through the oxycline, switching to denitrification upon transition to the OMZ core.

In contrast, groups such as obligately anaerobic anammox bacteria are recovered almost exclusively within the FL fraction, but appear no less dependent on particles.

Anammox cells presumably rely on particles indirectly as sources of nitrite and ammonium, produced by nitrate or nitrite reducers consuming particulate OM. Although anammox cells have been recovered in direct contact with particles in other systems (Woebken et al., 2007), enrichment of anammox genes and transcripts in OMZ FL fractions suggests a relatively loose physical association with particles, although it is possible that some particle-adhered cells are dislodged into the FL fraction during filtration. It is plausible that the proportion of the anammox community adhered to particles varies depending on local substrate conditions and oxygen content. Direct adherence may enable access to anoxic microniches and be favored at upper OMZ depths where water column oxygen content is closer to the threshold for anammox inhibition (~900 nM; Dalsgaard et al., 2014).

The potential for OMZ microbes to alter their physiology between FL and PA niches remains unresolved. Anammox Planctomycetes and nitriteoxidizing Nitrospina, two groups recovered in both FL and PA fractions and occupying relatively welldefined functional roles, did not exhibit significantly different expression profiles between fractions. This suggests that the physical act of filtration had a relatively minor effect on cell state. Nonetheless, certain genes exhibited large fold changes in relative abundance between fractions, raising the possibility that additional analyzes with higher replication per depth may reveal size fraction-specific transcriptional differences in certain taxa. Clarifying such patterns, potentially by sequencing deeply across biological replicates coupled with transcriptome mapping to reference genomes, may help estimate per-cell differences in biochemical processing at the microscale. However, the use of biomass fractionation to characterize the microspatial architecture of microbial activity remains

challenging. As shown here, particle removal significantly disrupts bulk N cycle processes mediated by a FL majority, likely by severing important chemical and biological linkages between FL and PA niches.

Acknowledgements

We thank the crew of the R/V New Horizon for help in sample collection, Neha Sarode and Josh Parris for help in sequencing, Anni Glud for technical assistance and for preparing the STOX, and Peter Girguis, Kathy Barbeau and Forest Rohwer for providing essential equipment. This work was supported by the National Science Foundation (1151698 to FJS), the Sloan Foundation (RC944 to FJS), the Danish National Research Foundation DNRF53, the Danish Council of Independent Research and the European Research Council 'Oxygen' grant (267233 supporting LAB, ML and BT).

Disclaimer: This chapter was published with the same title, along with the supplementary material in Appendix B, in ISME J on 7 April 2015.

Citation: **Ganesh, S., Bristow, L. A., Larsen, M., Sarode, N., Thamdrup, B., & Stewart, F. J.** (2015). Size-fraction partitioning of community gene transcription and nitrogen metabolism in a marine oxygen minimum zone. *ISME J*, *9*(12), 2682-2696.

3.6 References

Babbin AR, Keil RG, Devol AH, Ward BB. (2014). Organic matter stoichiometry, flux, and oxygen control nitrogen loss in the ocean. Science 344: 406–408.

Beman JM, Popp BN, Alford SE. (2012). Quantification of ammonia oxidation rates and ammonia-oxidizing archaea and bacteria at high resolution in the Gulf of California and eastern tropical North Pacific Ocean. Limnol Oceanogr 57: 711–726.

Beman JM, Carolan MT. (2013). Deoxygenation alters bacterial diversity and community composition in the ocean's largest oxygen minimum zone. Nat Commun 4: 2705.

Beman JM, Leilei Shih J, Popp BN. (2013). Nitrite oxidation in the upper water column and oxygen minimum zone of the eastern tropical North Pacific Ocean. ISME J 7: 2192–2205.

Blazewicz SJ, Barnard RL, Daly RA, Firestone MK. (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. ISME J 7: 2061–2208.

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

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.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ et al. (2011). Global patterns of 16 S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci USA 15: 108.

Ceja-Navarro JA, Nguyen NH, Karaoz U, Gross SR, Herman DJ, Andersen GL et al. (2014). Compartmentalized microbial composition, oxygen gradients and nitrogen fixation in the gut of Odontotaenius disjunctus. ISME J 8: 6–18.

Cho BC, Azam F. (1988). Major role of bacteria in biogeochemical fluxes in the ocean's interior. Nature 332: 441–443.

Cline JD, Richards FA. (1974). Oxygen deficient conditions and nitrate reduction in the eastern tropical North Pacific ocean. Limnol Oceanogr 17: 885–900.

Codispoti LA, Brandes JA, Christensen JP, Devol AH, Naqvi SWA, Paerl HW et al. (2001). The oceanic fixed nitrogen and nitrous oxide budgets: moving targets as we enter the anthropocene? Sci Mar 65: 85–105.

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.

Dalsgaard T, Stewart FJ, Thamdrup B, De Brabandere L, Revsbech NP, Ulloa O et al. (2014). Oxygen at nanomolar levels reversibly suppresses process rates and gene expression of anammox and denitrification in the oxygen minimum zone off northern Chile. mBio 5: e01966–14.

De Brabandere L, Canfield DE, Dalsgaard T, Friederich GE, Revsbech NP, Ulloa O et al. (2014). Vertical partitioning of nitrogen-loss processes across the oxic–anoxic interface of an oceanic oxygen minimum zone. Environ Microbiol 16: 3041–3054.

DeLong EF, Franks DG, Alldredge AL. (1993). Phylogenetic diversity of aggregate-attached vs free-living marine bacterial assemblages. Limnol Oceanogr 38: 924–934.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K et al. (2006). Greengenes, a chimera-checked 16 S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72: 5069–5072.

Edgar RC. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26: 2460–2461. Eldholm V, Norheim G, von der Lippe B, Kinander W, Dahle UR, Caugant DA et al. (2014). Evolution of extensively drug-resistant Mycobacterium tuberculosis from a susceptible ancestor in a single patient. Genome Biol 15: 490.

Faith DP. (1992). Conservation evaluation and phylogenetic diversity. Biol Conserv 61: 1–10.

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.

Fuchsman CA, Staley JT, Oakley BB, Kirkpatrick JB, Murray JW. (2012). Free-living and aggregateassociated Planctomycetes in the Black Sea. FEMS Microbiol Ecol 80: 402–416.

Füssel 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.

Ganesh S, Parris DJ, DeLong EF, Stewart FJ. (2014). Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. ISME J 8: 187–211.

Ghiglione JF, Conan P, Pujo-Pay M. (2009). Diversity of total and active free-living vs particle-attached bacteria in the euphotic zone of the NW Mediterranean Sea. FEMS Microbiol Lett 299: 9–21.

Grossart HP, Hietanen S, Ploug H. (2003). Microbial dynamics on diatom aggregates in Øresund, Denmark. Mar Ecol Prog Ser 249: 69–78.

Grossart HP, Tang KW, Kiørboe 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.

Gruber N. The dynamics of the marine nitrogen cycle and its influence on atmospheric CO_2 variations. Follows M, Oguz T . The Ocean Carbon Cycle and Climate, NATO ASI Series. Kluwer: Dordrecht, Germany, 2004; 97–148.

Güven D, Dapena A, Kartal B, Schmid MC, Maas B, van de Pas-Schoonen K et al. (2005). Propionate oxidation by and methanol inhibition of anaerobic ammonium-oxidizing bacteria. Appl Environ Microbiol 71: 1066–1071.

Hardcastle TJ, Kelly KA. (2010). baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinform 11: 422.

Hatt JK, Ritalahti KM, Ogles DM, Lebrón CA, Löffler FE. (2013). Design and application of an internal amplification control to improve Dehalococcoides mccartyi 16S rRNA gene enumeration by qPCR. Environ Sci Technol 47: 11131–11138.

Hollibaugh T, Wong PS, Murrell MC. (2000). Similarity of particleassociated and free-living bacterial communities in northern San Francisco Bay, California. Aquat Microb Ecol 21: 102–114.

Hoppe HG, Ducklow H, Karrasch B. (1993). Evidence for dependency of bacterial growth on enzymatic hydrolysis of particulate organic matter in the mesopelagic ocean. Mar Ecol Prog Ser 93: 277–283.

Hunt DE, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF et al. (2008). Resource partitioning and sympatric differentiation among closely related bacterioplankton. Science 320: 1081–1085.

Hunt DE, Lin Y, Church MJ, Karl DM, Tringe SG, Izzo LK et al. (2013). Relationship between abundance and specific activity of bacterioplankton in open ocean surface waters. Appl Environ Microbiol 79: 177–184.

Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC. (2011). Integrative analysis of environmental sequences using MEGAN4. Genome Res 21: 1552–1560.

Ivars-Martinez E, Martin-Cuadrado AB, D'Auria G, Mira A, Ferriera S, Johnson J et al. (2008). Comparative genomics of two ecotypes of the marine planktonic copiotroph Alteromonas macleodii suggests alternative lifestyles associated with different kinds of particulate organic matter. ISME J 2: 1194–1212.

Jetten MS, Niftrik LV, Strous M, Kartal B, Keltjens JT, Op den Camp HJ et al. (2009). Biochemistry and molecular biology of anammox bacteria. Crit Rev Biochem Mol Biol 44: 65–84.

Kalvelage T, Lavik G, Lam P, Contreras S, Arteaga L, Loscher CR et al. (2013). Nitrogen cycling driven by organic matter export in the South Pacific oxygen minimum zone. Nat Geosci 6: 228–234.

Karl DM, Knauer GA, Martin JH. (1988). Downward flux of particulate organic matter in the ocean: a particle decomposition paradox. Nature 332: 438–441.

Karstensen J, Stramma L, Visbeck M. (2008). Oxygen minimum zones in the eastern tropical Atlantic and Pacific oceans. Prog Oceanogr 77: 331–350.

Kartal B, Kuypers MMM, Lavik G, Schalk J, Op den Camp HJM, Jetten MSM et al. (2007). Anammox bacteria disguised as denitrifiers: nitrate reduction to dinitrogen gas via nitrite and ammonium. Environ Microbiol 9: 635–642.

Kong L, Jing H, Kataoka T, Buchwald C, Liu H. (2013). Diversity and spatial distribution of hydrazine oxidoreductase (hzo) gene in the oxygen minimum zone off Costa Rica. PLoS One 8: e78275.

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol 79: 5112–5120.

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.

Lam P, Kuypers MMM. (2011). Microbial nitrogen cycling processes in oxygen minimum zones. Annu Rev Mar Sci 3: 317–345.

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.

Lund MB, Smith JM, Francis CA. (2012). Diversity, abundance and expression of nitrite reductase (nirK)-like genes in marine thaumarchaea. ISME J 6: 1966–1977.

McIlvin MR, Altabet MA. (2005). Chemical conversion of nitrate and nitrite to nitrous oxide for nitrogen and oxygen isotopic analysis in freshwater and seawater. Anal Chem 77: 5589–5595.

Milohanic E, Glaser P, Coppée JY, Frangeul L, Vega Y, Vázquez-Boland JA et al. (2003). Transcriptome analysis of Listeria monocytogenes identifies three groups of genes differently regulated by PrfA. Mol Microbiol 47: 1613–1625.

Naqvi SWA, Kumar MD, Narvekar PV, De Sousa SN, George MD, Silva CD et al. (1993). An intermediate nepheloid layer associated with high microbial rates and denitrification in the northwest Indian Ocean. J Geophys Res 98: 16469–16479.

Paulmier A, Ruiz Pino. (2009). Oxygen minimum zones (OMZs) in the modern ocean. Prog Oceanogr 80: 113–128.

Podlaska A, Wakeham SG, Fanning KA, Taylor GT. Microbial community structure and productivity in the oxygen minimum zone of the eastern tropical North Pacific. Deep Sea Res Part I 2012; 66: 77–89.

Rao D, Webb JS, Kjelleberg S. (2005). Competitive interactions in mixed-species biofilms containing the marine bacterium Pseudoalteromonas tunicata. Appl Environ Microbiol 71: 1729–1736.

Revsbech NP, Larsen LH, Gundersen J, Dalsgaard T, Ulloa O, Thamdrup B et al. (2009). Determination of ultra-low oxygen concentrations in oxygen minimum zones by the STOX sensor. Limnol Oceanogr Methods 7: 371–381.

Ritalahti KM, Amos BK, Sung Y, Wu Q, Koenigsberg SS, Löffler FE et al. (2006). Quantitative PCR targeting 16 S rRNA and reductive dehalogenase genes simultaneously monitors multiple Dehalococcoides strains. Appl Environ Microbiol 72: 2765–2774.

Rush D, Wakeham SG, Hopmans EC, Schouten S, Sinninghe Damsté JS. (2012). Biomarker evidence for anammox in the oxygen minimum zone of the Eastern Tropical North Pacific. Org Geochem 53: 80–87.

Satinsky BM, Crump BC, Smith CB, Sharma S, Zielinski BL, Doherty M et al. (2014). Microspatial gene expression patterns in the Amazon River Plume. Proc Natl Acad Sci USA 111: 11085–11090.

Schmieder R, Lim YW, Edwards R. (2012). Identification and removal of ribosomal RNA sequences from metatranscriptomes. Bioinformatics 28: 433–435.

Schumann U, Edwards MD, Rasmussen T, Bartlett W, van West P, Booth IR et al. (2010). YbdG in Escherichia coli is a threshold-setting mechanosensitive channel with MscM activity. Proc Natl Acad Sci USA 107: 12664–12669.

Shanks AL, Trent JD. (1979). Marine snow: microscale nutrient patches. Limnol Oceanogr 24: 850–854.

Sheik CS, Jain S, Dick GJ. (2014). Metabolic flexibility of enigmatic SAR324 revealed through metagenomics and metatranscriptomics. Environ Microbiol 6: 304–317.

Sihto HM, Tasara T, Stephan R, Johler S. (2014). Validation of reference genes for normalization of qPCR mRNA expression levels in Staphylococcus aureus exposed to osmotic and lactic acid stress conditions encountered during food production and preservation. FEMS Microbiol Lett 356: 134–140.

Simon M, Grossart HP, Schweitzer B, Ploug H. (2002). Microbial Ecology of organic aggregates in aquatic ecosystems. Aquat Microb Ecol 28: 175–211.

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.

Spinrad RW, Glover H, Ward BB, Codispoti LA, Kullenberg G. (1989). Suspended particle and bacterial maxima in Peruvian coastal waters during a cold water anomaly. Deep-Sea Res 36: 715–733.

Stewart FJ, Ulloa O, DeLong EF. (2012). 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.

Storey JD, Tibshirani R. (2003). Statistical significance for genomewide studies. Proc Natl Acad Sci USA 100: 9440–9445.

Strous M, Pelletier E, Mangenot S, Rattei T, Lehner A, Taylor MW et al. (2006). Deciphering the evolution and metabolism of an anammox bacterium from a community genome. Nature 440: 790–794.

Sue D, Fink D, Wiedmann M, Boor KJ. (2004). Sigma(B)dependent gene induction and expression in Listeria monocytogenes during osmotic and acid stress conditions simulating the intestinal environment. Microbiology 150: 3843–3855.

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.

Taylor GT, Thunell R, Varela R, Benitez-Nelson C, Scranton MI. (2009). Hydrolytic ectoenzyme activity associated with suspended and sinking organic particles within the anoxic Cariaco Basin. Deep Sea Res I 56: 1266–1283.

Teske A, Alm E, Regan JM, Toze S, Rittmann BE, Stahl DA et al. (1994). Evolutionary relationships among ammoniaand nitrite-oxidizing bacteria. J Bacteriol 176: 6623–6630.

Thamdrup B, Dalsgaard T. (2002). Production of N_2 through anaerobic ammonium oxidation coupled to nitrate reduction in marine sediments. Appl Environ Microbiol 68: 1312–1318.

Thamdrup B, Dalsgaard T, Jensen MM, Ulloa O, Farias L, Escribano R et al. (2006). Anaerobic ammonium oxidation in the oxygen-deficient waters off northern Chile. Limnol Oceanogr 51: 2145–2156.

Tiano L, Garcia Robledo E, Dalsgaard T, Devol A, Ward B, Uloa O et al. (2014). Oxygen distribution and aerobic respiration in the north and southeastern tropical Pacific oxygen minimum zones. Deep-Sea Res I 194: 173–183.

Turley CM, Stutt ED. (2000). Depth-related cellspecific bacterial leucine incorporation rates on particles and its biogeochemical significance in the Northwest Mediterranean. Limnol Oceanogr 45: 419–425.

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.

Van de casteele SJ, Peetermans WE, Merckx R, Van Eldere J. (2001). Quantification of expression of Staphylococcus epidermidis housekeeping genes with Taqman quantitative PCR during in vitro growth and under different conditions. J Bacteriol 183: 7094–7101.

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.

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 Express 17: 21992–22004.

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 Microbiol 73: 4648–4657.

Wright JJ, Konwar KM, Hallam SJ. (2012). Microbial ecology of expanding oxygen minimum zones. Nat Rev Microbiol 10: 381–394.

Wright JJ, Mewis K, Hanson NW, Konwar KM, Maas KR, Hallam SJ et al. (2014). Genomic properties of Marine Group A bacteria indicate a role in the marine sulfur cycle. ISME J 8: 455–468.

Yamada N, Fukuda H, Ogawa H, Saito H, Suzumura M. (2012). Heterotrophic bacterial production and extracellular enzymatic activity in sinking particulate matter in the western North Pacific Ocean. Front Microbiol 3: 379.

Zumft WG. (1997). Cell biology and molecular basis of denitrification. Microbiol Mol Biol Rev 61: 533–616.

CHAPTER 4

STANDARD FILTRATION PRACTICES MAY SIGNIFICANTLY DISTORT PLANKTONIC MICROBIAL DIVERSITY ESTIMATES

4.1 Abstract

Fractionation of biomass by filtration is a standard method for sampling planktonic microbes. It is unclear how the taxonomic composition of filtered biomass changes depending on sample volume. Using seawater from a marine oxygen minimum zone, we quantified the 16S rRNA gene composition of biomass on a prefilter (1.6 μm pore-size) and a downstream 0.2μm filter over sample volumes from 0.05 to 5 L. Significant community shifts occurred in both filter fractions, and were most dramatic in the prefilter community. Sequences matching Vibrionales decreased from ~40 to 60% of prefilter datasets at low volumes (0.05–0.5 L) to less than 5% at higher volumes, while groups such at the Chromatiales and Thiohalorhabdales followed opposite trends, increasing from minor representation to become the dominant taxa at higher volumes. Groups often associated with marine particles, including members of the Deltaproteobacteria, Planctomycetes, and Bacteroidetes, were among those showing the greatest increase with volume (4 to 27-fold). Taxon richness (97% similarity clusters) also varied significantly with volume, and in opposing directions depending on filter fraction, highlighting potential biases in community complexity estimates. These data raise concerns for studies using filter fractionation for quantitative comparisons of aquatic microbial diversity, for example between free-living and particle-associated communities.

4.2 Introduction

Most studies of planktonic microbes begin by isolating organisms from the environment. This task usually involves collection of suspended biomass by passing water through a filter. For studies of planktonic bacteria and archaea, it is routine to use an inline series of filters during the collection step, often with a prefilter of larger pore size (typically 1.0–30 μm) upstream of a primary collection filter of smaller pore size (0.1–0.2 μm; Fuhrman et al., 1988; Venter et al., 2004; Walsh et al., 2009; Morris and Nunn, 2013; Stewart, 2013). Both filters retain microbial biomass, and the larger and smaller biomass size fractions are often used to approximate divisions between particle-associated and free-living microbes, respectively (e.g., Acinas et al., 1999; Crump et al., 1999; Hollibaugh et al., 2000; Moeseneder et al., 2001; LaMontagne and Holden, 2003; Ghiglione et al., 2007, 2009; Rusch et al., 2007; Kellogg and Deming, 2009; Smith et al., 2013; D'Ambrosio et al., 2014; Ganesh et al., 2014; Mohit et al., 2014; Orsi et al., 2015, among others). Retained biomass can be used toward a range of analytical goals, including quantitative measurements of community taxonomic and metabolic diversity, based for example on 16S rRNA gene or metagenome sequencing. Such analyses shape our understanding of microbial distributions and functions in the environment.

The effects of filtration method on quantitative measurements of community genetic diversity are poorly understood, and rarely accounted for in study design. Notably, it is unclear how prefiltration and variation in filtered water volume influence the relative abundances of taxa retained in different filter fractions. This is surprising given an extensive literature demonstrating variation in particle retention and clogging rates among filter types and in the size spectra of retained particles due to clogging

(Nagata, 1986; Taguchi and Laws, 1988; Knefelkamp et al., 2007), and a smaller number of studies showing differential selection of target taxa based on filter type (Gasol and Moran, 1999), or variation in DNA extraction efficiency depending on filtered water volume (Boström et al., 2004). Indeed, sample volume varies widely across studies, from less than 1 L for coastal or eutrophic environments (Hollibaugh et al., 2000) to greater than 100 L for open ocean settings (Rusch et al., 2007). Fewer studies sample at very low volumes (microliters), in part to avoid under-sampling community richness, potentially due to microscale heterogeneity in biomass distributions (Long and Azam, 2001). Sample volume is often also variable among samples within a study, differing for instance between samples collected for DNA vs. RNA analysis (Frias-Lopez et al., 2008; Hunt et al., 2013). While it has been assumed that the proportion of free-living bacteria retained in the prefilter fraction increases with the volume of filtered water (Lee et al., 1995), the effects of such retention on community composition measurements are unknown. Advances in high-throughput 16S rRNA gene sequencing now enable rapid, low-cost quantification of microbial community structure (Caporaso et al., 2011). These methods can be used to help identify biases associated with common sample collection practices.

We conducted experiments to test whether variation in the volume of filtered seawater biases estimates of marine microbial diversity. Biomass representing two size fractions (0.2–1.6μm, >1.6μm) was collected from a marine oxygen minimum zone (OMZ) following a sequential inline filtration protocol used routinely to study planktonic microbes (Fuhrman et al., 1988; Frias-Lopez et al., 2008; Stewart, 2013). Sequencing of community 16S rRNA gene amplicons revealed significant shifts in community structure and richness over volumes ranging from 0.05 to 5 L. We discuss these results as evidence

that sample volume should be critically examined as a potentially confounding variable in comparing estimates of aquatic microbial diversity.

4.3 Materials and Methods

4.3.1 Sample Collection and Filtration

Seawater was collected from the OMZ off Manzanillo, Mexico during the Oxygen Minimum Zone Microbial Biogeochemistry Expedition 2 (OMZoMBiE2) cruise (R/V New Horizon; cruise NH1410; May 10-June 8, 2014). Collections were made using 30 L Niskin bottles attached to a CTD-Rosette. Seawater was emptied into 20 L cubitainers upon retrieval on deck and stored in the dark at 4◦C until filtration (<1 h).

Two experiments were conducted to assess the impact of sample volume on microbial diversity. Experiment 1 collected biomass after filtration of 0.1, 1, and 5 L from the sample pool, with 3–5 replicate filtration lines per volume. Experiment 2 examined volumes of 0.05, 0.1, 0.5, 1, 2, and 5 L, with 2–3 replicates per volume. These ranges encompass volumes typical for studies of both coastal and open ocean environments, although fewer studies collect volumes at the lower end of the range. Water for experiment 1 was collected from depth 150 m near the secondary nitrite maximum at a site ~95 km offshore from Manzanillo (18^{o} 12″ N 104^{o} 12″ W). Water for experiment 2 was collected from depth 400 m at a site ~300 km offshore from Manzanillo (20^{o} 00″ N, 107^{o} 00″ W). Collections depths at both sites were beneath the photic zone. Total water depth at both sites was >1000 m.

For each experiment, discrete water volumes were measured from the stored sample water, transferred into pre-washed secondary containers (graduated cylinders or

glass bottles), and filtered by sequential in-line filtration at room temperature through a glass fiber disc pre-filter (GF/A, 47 mm, 1.6 μm pore- size, Whatman) and a primary collection filter (Sterivex-GP filter cartridge, polyethersulfone membrane, 0.22 μm pore-size, Millipore) using a peristaltic pump (Masterflex L/S modular drive pump, 1–100 rpm, Cole-Parmer ®) at constant speed (60 rpm). Replicate filters were collected in parallel (separate filter lines) for each discrete volume, with a maximum of 5 lines (replicates) running in parallel at once. After filtration, prefilters were transferred to cryovials containing lysis buffer (~1.8 ml; 50 mM Tris-HCl, 40 mM EDTA, and 0.73 M Sucrose), Sterivex cartridges were filled with lysis buffer (~1.8 ml) and capped at both ends, and both filter types were stored at −80◦C until DNA extraction. After filtration of one set of replicates, filter lines were rinsed, and clean filters were added. Aliquots representing the next sample in the volume series were then measured from the source water, and the filtration sequence was repeated.

4.3.2 DNA Extraction

DNA was extracted from prefilters and Sterivex cartridges using a phenol:chloroform protocol as in Ganesh et al. (2014). Cells were lysed by adding lysozyme (2 mg in 40 μl of lysis buffer per filter) directly to the prefilter-containing cryovials and the Sterivex cartridges, 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 cryovials and cartridges were resealed and incubated for 2 h at 55◦C. The lysate was removed, and DNA was extracted once with Phenol:Chloroform:Isoamyl Alcohol (25:24:1) and once with Chloroform:Isoamyl Alcohol (24:1) and then concentrated by

spin dialysis using Ultra-4 (100 kDa, Amicon) centrifugal filters. Double stranded DNA was quantified on a Qubit Fluorometer using the dsDNA BR Assay Kit.

4.2.3 16S rRNA Gene Quantitative PCR

Quantitative PCR (qPCR) was used to count total bacteria 16S rRNA gene copies in DNA extracted from both prefilter and Sterivex size fractions. Total 16S counts were obtained using SYBR Green-based qPCR and universal bacterial 16S primers 1055f and 1392r, as in Hatt et al. (2013). Ten-fold serial dilutions of DNA from a plasmid carrying a single copy of the 16S rRNA gene (from Dehalococcoides mccartyi) were included on each qPCR plate and used to generate standard curves. AssaTMys were run on a 7500 Fast PCR System and a StepOnePlus Real- Time PCR System (Applied Biosystems). All samples were run in triplicate (20 µL each) and included 1X SYBR Green Supermix (BIO-RAD), 300 nM of primers, and 2 µL of template DNA (diluted 1:100). Thermal cycling involved: incubation at 50◦C for 2 min to activate uracil-N-glycosylase, followed by 95◦ C for 10 min to inactivate UNG, denature template DNA, and activate the polymerase, followed by 40 cycles of denaturation at 95◦C (15 s) and annealing at 60◦C (1 min).

4.2.4 Diversity Analyses

High-throughput sequencing of dual-indexed PCR amplicons encompassing the V4 region of the 16S rRNA gene was used to assess microbial community composition. Briefly, amplicons were synthesized using Platinum◯R PCR SuperMix (Life Technologies) with primers F515 and R806, encompassing the V4 region of the 16S

rRNA gene (Caporaso et al., 2011). These primers are used primarily for bacterial 16S rRNA genes analysis, but also amplify archaeal sequences. Both forward and reverse primers were barcoded and appended with Illumina- specific adapters according to Kozich et al. (2013). Equal amounts of starting DNA (0.5 ng) were used for each PCR reaction to avoid potential PCR biases due to variable template concentrations (Kennedy et al., 2014). Thermal cycling involved: denaturation at 94◦ C (3 min), followed by 30 cycles of denaturation at 94◦ C (45 s), primer annealing at 55◦C (45 s) and primer extension at 72◦C (90 s), followed by extension at 72◦C for 10 min. Amplicons were analyzed by gel electrophoresis to verify size (~400 bp, including barcodes and adaptor sequences) and purified using the Diffinity RapidTip2 PCR purification tips (Diffinity Genomics, NY). Amplicons from different samples were pooled at equimolar concentrations and sequenced on an Illumina MiSeq using a 500 cycle kit with 5% PhiX to increase read diversity.

Amplicons were analyzed using QIIME (Caporaso et al., 2010) following standard protocols. Barcoded sequences were de-multiplexed and trimmed (length cutoff 100 bp) and filtered to remove low quality reads (average Phred score <25) using Trim Galore! (http://www.bioinformatics.babraham. ac.uk/projects/trim_galore/). Paired-end reads were then merged using FLASH (Magocˇ and Salzberg, 2011), with criteria of average read length 250, fragment length 300, and fragment standard deviation 30. Chimeric sequences were detected by reference-based searches using USEARCH (Edgar, 2010). Identified chimeras were filtered from the input dataset, and merged non-chimeric sequences were clustered into Operational Taxonomic Units (OTUs) at 97% sequence similarity using open-reference picking with the UCLUST algorithm (Edgar, 2010) in

QIIME. The number of sequences assigned to OTUs averaged 22,693 (range: 2255–182,051) and 20,638 (range: 5152–46,634) per sample for experiments 1 and 2, respectively.

Taxonomy was assigned to representative OTUs from each cluster using the Greengenes database (DeSantis et al., 2006). OTU counts were rarefied (10 iterations) and alpha diversity was quantified at a uniform sequence depth (n = 2255 for exp 1 and 5152 for exp 2) using the Chao1 estimator of richness. Significant differences in Chao1 estimates between volume groupings were detected using pairwise two-sample t- tests with a Bonferroni correction for multiple comparisons. To compare community composition between samples, sequences were aligned in QIIME using PyNAST, and beta diversity was calculated using the weighted Unifrac metric (Lozupone and Knight, 2005). Sample relatedness based on Unifrac was visualized with two-dimensional Principal Coordinate Analyses. The statistical significance of sample grouping according to filtered volume was assessed using the nonparametric adonis method in QIIME based on Unifrac distances run with 999 permutations and Bonferroni corrections for multiple tests (Table C.1).

Taxa (Order level) differing significantly in proportional abundance in pairwise comparisons of filter volumes (lowest vs. highest volumes) were identified via an empirical Bayesian approach using the program baySeq (Hardcastle and Kelly, 2010), as in Stewart et al. (2012) and Ganesh et al. (2014). The baySeq method assumes a negative binomial distribution with prior distributions derived empirically from the data (100,000 iterations). Dispersion was estimated via a quasi-likelihood method, with the count data normalized by dataset size (total number of identifiable 16S rRNA gene amplicons per

dataset). Posterior likelihoods per Order were calculated for models (volume groupings) in which Orders were either predicted to be equivalently abundant between lowest vs. highest filter volumes or differentially abundant. A false discovery rate threshold of 0.05 was used for detecting differentially abundant categories.

All sequence data have been submitted to the Sequence Read Archive at NCBI under BioProject ID PRJNA275901.

4.4 Results

Total bacterial 16S rRNA gene counts (prefilter + Sterivex combined), a rough proxy for microbial abundance, averaged 4 × 105 and 7 × 104 (averaged across all volumes and replicates) for experiment 1 and 2, respectively, and were 10 to 11-fold higher in the Sterivex size fraction (0.2–1.6 μm; Table C.1). These counts are within the range of or slightly lower than prior bacterial 16S gene counts at OMZ and oxycline depths in the study area (~5 × 104 to 1.3 × 106; Ganesh et al., 2015; Stewart, unpublished), values typical of open ocean waters, and consistent with prior reports of elevated microbial biomass in the smallest ("free-living") size fraction compared to particulate size fractions (Cho and Azam, 1988; Ghiglione et al., 2009). DNA yields, representing both prokaryote and eukaryote sources, were ~2 to 4-fold higher in Sterivex filter fractions compared to prefilters (Figure 4.1A). Both DNA yields and bacterial 16S counts increased linearly with volume in both fractions, and were most variable (among replicates) for the 5 L samples, suggesting differences in extraction efficiency at high volumes (Figure 4.1A, Figure C.1). Similarly, filtration time increased linearly as a function of sample volume (R2 = 0.99; Figure 4.1B), ranging from <40 s (0.05 L) to 43

min (5 L). Average flow rate, calculated from filtration times from all samples, was 0.12 L min−1. A minor increase of filtration times per unit volume was observed for the 5 L samples compared to 1 and 2 L samples, highlighting a very slight decline in flow rate with volume, potentially due to filter clogging. No filters visibly ruptured during the experiments.

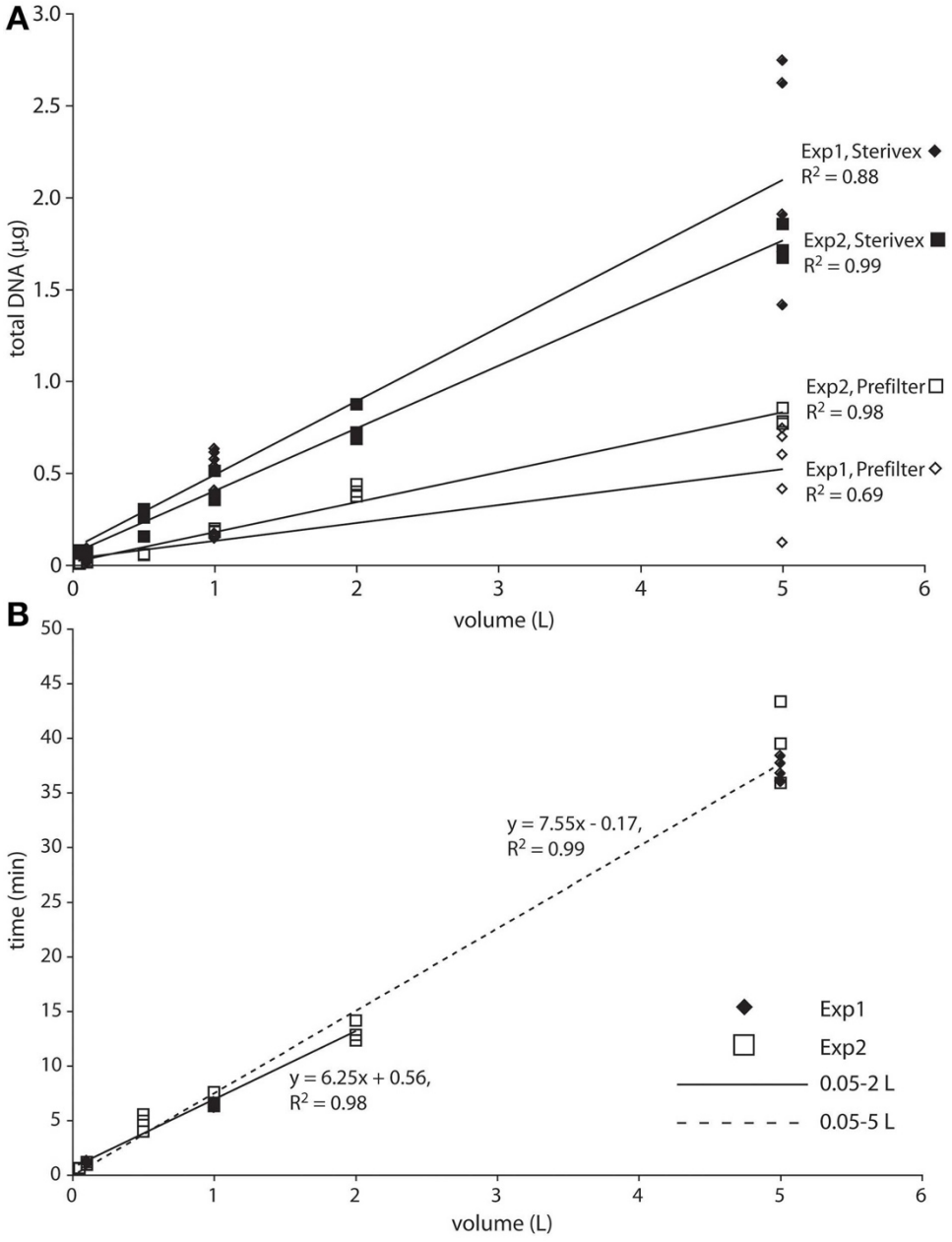


Figure 4.1 Total DNA yield (A) and filtration time (B) as a function of filtered water volume. The solid regression line in B spans the 0.05–2 L range. The dashed regression spans the full dataset, with the increased slope suggesting a slight decrease in flow rate at the highest volumes.

Community structure varied significantly with sample volume in both biomass fractions. Principal coordinates analysis based on the weighted Unifrac metric showed significant differentiation of samples based on filtered volume for both size fractions in both experiments (R2: 0.50–0.87; $P < 0.05$), with the strongest separation occurring between samples of 1 L or more and those of lower volumes (Figures 4.2, 4.3; Table C.2). Chao1 estimates of operational taxonomic unit (OTU; 97% similarity cluster) richness varied with filter volume in both size fractions (Figure 4.4). In both experiments, median richness in the >1.6 μm fraction increased (by up to 68%) with volume, reaching peak values at 1 L, before declining again at the highest volumes (2–5 L). However, differences in prefilter richness between volume groupings were not significant, due partly to limited replication per volume. In contrast, in both experiments richness in the downstream 0.2–1.6 μm "Sterivex" size fraction declined by ~25–30% over the volume range, with statistically significant declines observed between lower volumes and the 5 L samples in experiment 2 ($P < 0.05$).

The most dramatic shifts in taxonomic representation with volume occurred in the prefilter fraction (Figures2, 3). Of the major Prokaryotic Orders detected in experiment 1, where "major" here indicates an average percent abundance >0.5% across filter type replicates, 59% (19 of 32) varied significantly in abundance from lowest to highest volumes (0.1 vs. 5 L; $P < 0.05$), with 88% (28) undergoing a fold-change of 2 or greater (positive or negative; Table C.3). In experiment 2, 96% of major Orders underwent a fold-change of 2 or greater over the volume range (0.05 vs. 5 L), although fewer (17%) of these changes were statistically significant (likely due to fewer replicates per volume in experiment 2; Table C.4). In both experiments, sequences matching Vibrionales

decreased dramatically (21–33 fold) with filter volume, from >50% (average) of the lowest volume datasets, to <3% at 5 L. In contrast, other groups increased to dominate the dataset over the volume range. In experiment 1, sequences related to the gammaproteobacterial Thiohalorhabdales increased nearly 4-fold, becoming the single most abundant Order at 5 L with 17% of total sequences. In experiment 2, the Order Chromatiales increased from an average of 2% at lower volumes (range: 0.2– 5.7% over 0.05, 0.1, 0.5 L datasets) to become the most abundant taxon at higher volumes, representing an average of 42% of total sequences at volumes of 1 L or greater (range: 15–77%; Figure 4.3, Table C.4). In both experiments, Euryarchaeota of the Thermoplasmata increased 10 to 20-fold to become the second most abundant Order (7–10%) at 5 L. Diverse members of the Deltaproteobacteria, Planctomycetes, and Bacteroidetes also underwent substantial increases (4 to 27-fold) in the prefilter fraction in both experiments (Tables S3, S4).

Large changes in taxon representation also occurred in the Sterivex fraction, but were less dramatic that those in the prefilter community. In both experiments, 14% of the major prokaryotic Orders detected in this fraction changed significantly in representation over the volume range ($P < 0.05$, Tables S3, S4). As in the prefilter datasets, the most substantial shifts occurred in the Vibrionales, which decreased 51 to 97-fold with volume. Other groups undergoing significant declines included the Alteromonadales, Rhodobacterales, and Kiloniellales, notably with these groups declining 16 to 94-fold in experiment 2. Less dramatic, but still substantial, shifts occurred in the opposite direction, notably with the abundant SAR406 lineage increasing ~2-fold in representation in both experiments to constitute over one quarter of all sequences in the 5L datasets.

4.5 Discussion

Studies of diverse aquatic habitats consistently show that microbial community composition in the prefilter fraction differs from that of the smaller size fraction (e.g., Simon et al., 2002, 2014, and references therein). These differences have been used to suggest taxonomic, chemical, and metabolic partitioning between particle-associated vs. free-living microbial communities (see references in Introduction). While the pore size of the primary collection filter used in such studies is almost always 0.2μm, the pore size of the prefilter varies, typically from 0.8 to 1.0 (e.g., Hollibaugh et al., 2000; Allen et al., 2012) to 30μm (e.g., Fuchsman et al., 2011, 2012), but is most often in the range of 1 to 3μm, encompassing the 1.6 μm GF/A prefilter used in this study. Here, sample volume had a much stronger effect on prefilter community composition compared to that of the downstream Sterivex fraction. Taxa such as the Thiohalorhabdales and Chromatiales in experiments 1 and 2, respectively, were evenly distributed between filter fractions at low volumes but dramatically enriched in the prefilter at higher volumes. Furthermore, major taxonomic divisions often described as being enriched in particulate fractions, including the Flavobacteriales (Bacteroidetes), Myxococcales (Deltaproteobacteria), and Planctomycetes (Crump et al., 1999; Simon et al., 2002; Eloe et al., 2011; Allen et al., 2012; Fuchsman et al., 2012), were among those whose proportional abundances in the prefilter fraction increased the most with filter volume, although these groups have also been detected on particles sampled directly by syringe (DeLong et al., 1993; Rath et al., 1998). Such changes were not exclusive to the prefilter fraction. SAR406, an uncultured lineage commonly affiliated with low- oxygen waters (Wright et al., 2012, 2014), increased steadily with volume to become the single most abundant taxonomic group in

the Sterivex fraction in both experiments. These patterns suggest that differences in sample volume can bias multiple sequential filter fractions, but may be particularly problematic for studies inferring the structure of the particle-associated community, or the association of certain taxa with a particle-attached niche.

Community diversity differences between particle-associated and free-living biomass fractions are often variable across diverse ocean habitats. This has been attributed to environmental variation associated with geographic and vertical zonation, as well as flexibility in microbial lifestyles allowing alternation between particle-associated and free-living lifestyles (Grossart, 2010). Surprisingly, only a small number of studies have evaluated the potential for collection methods to bias community-level diversity measurements (Kirchman et al., 2001; Long and Azam, 2001; Ghiglione et al., 2007). Focusing on the 0.2–3.0μm filter fraction, Long and Azam (2001) recovered qualitatively similar community 16S rRNA gene fingerprints (denaturing gradient gel electrophoresis; DGGE) from coastal seawater volumes ranging from 1μl to 1 L. Similarly, Ghiglione et al. (2007) reported no visible effect of volume (0.1 to 5L) on fingerprints (capillary electrophoresis- single strand conformation polymorphism) from Mediterranean Sea bacterioplankton on 0.2 and 0.8 μm filters. High throughout 16S amplicon sequencing now provides a cost-competitive alternative to fingerprinting with higher sensitivity for detecting community structural shifts, including those involving rare taxa. Using this method, our results suggest that biases due to variation in filtered volume may also confound comparisons of size-fractionated community structure across habitats and studies.

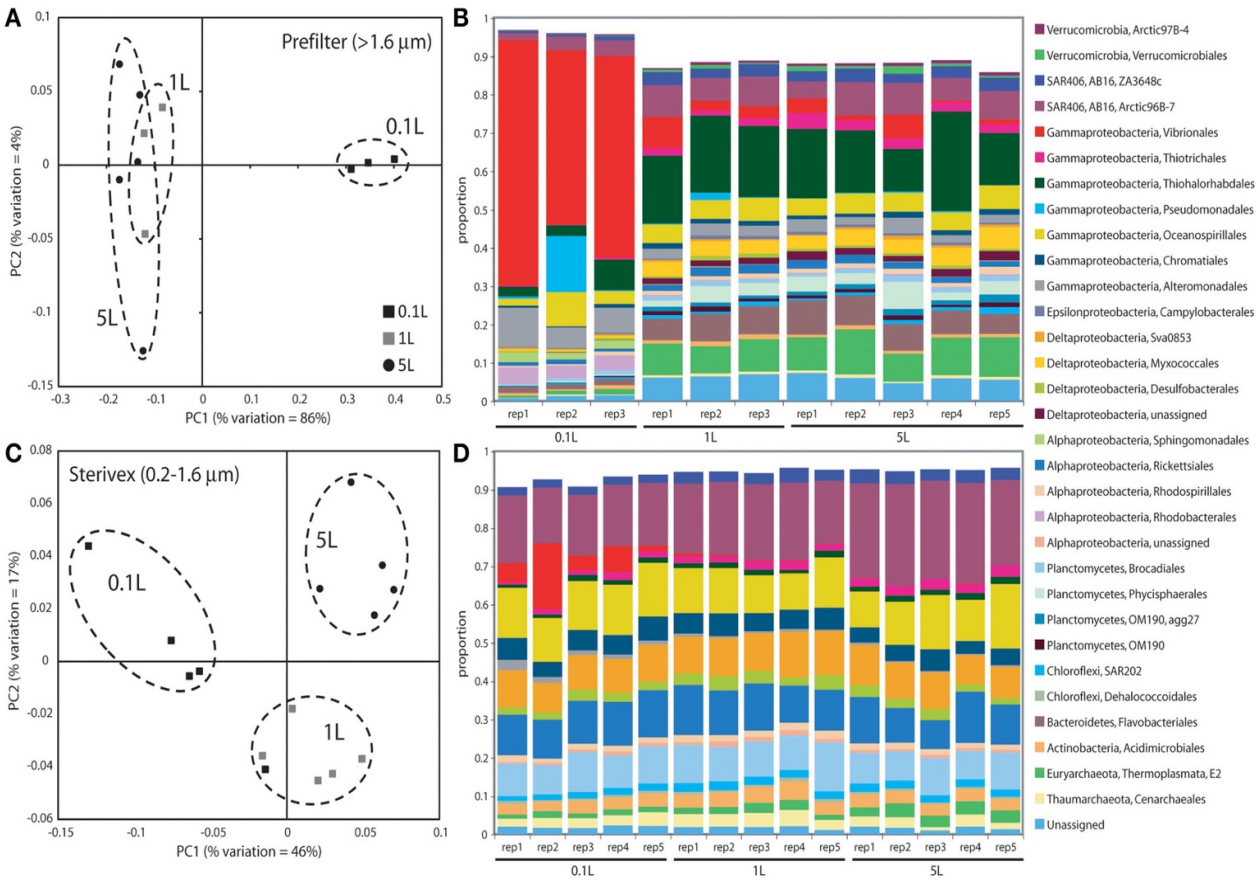


Figure 4.2 Microbial community relatedness (A,C) and taxon abundances (B,D) in experiment 1 prefilter (>1.6 µm; A,B) and Sterivex (0.2–1.6 µm; C,D) samples. Relatedness based on 16S rRNA gene amplicon sequencing, as quantified by the weighted Unifrac metric. Samples representing different filtered water volumes are circled. Abundances in (B,D) are percentage abundances of major bacterial Orders. Only Orders with abundance >0.5% (averaged across all replicates) are shown

.

Our data also suggest that metrics of community (OTU) richness may be sensitive to filtered water volume. In both experiments, richness of the prefilter and Sterivex communities followed opposing trends with volume, increasing in the larger size fraction and decreasing in the smaller. Declines in diversity with filter volume (0.01 vs. 2L) were previously observed in a study of prefiltered bacterioplankton (0.2–3.0μm fraction) using DGGE, although determinants of that decline could not be related directly to volume due to variations in the DNA extraction procedure (Kirchman et al., 2001). Here, in experiment 1, richness of the prefilter fraction was lower than that of the Sterivex fraction at 0.1 L, but this pattern was reversed at higher volumes. Enhanced richness in prefilter vs. 0.2μm fractions has been reported from diverse systems (Eloe et al., 2011; Bižic´-Ionescu et al.,2014; Ganesh et al.,2014),leading to speculation that the complexity of the microbial community varies differentially between free-living and particle-associated niches in response to fluctuation in water column conditions or niche-availability on particles. Bulk estimates of community complexity may instead be driven by variation in filtered water volume.

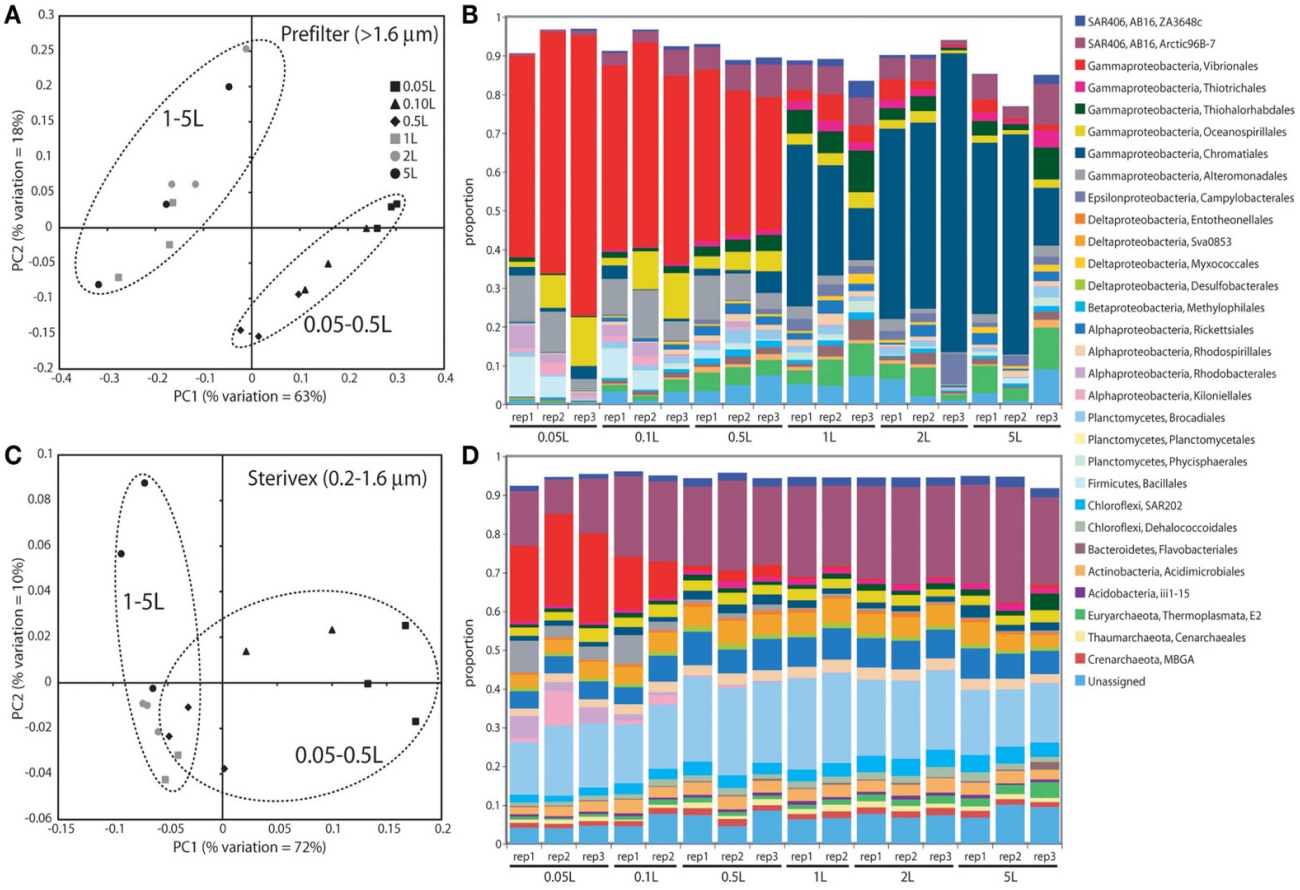


Figure 4.3 Microbial community relatedness (A,C) and taxon abundances (B,D) in experiment 2 prefilter (>1.6 µm; A,B) and Sterivex (0.2–1.6 µm; C,D) samples. Relatedness based on 16S rRNA gene amplicon sequencing, as quantified by the weighted Unifrac metric. Samples representing different filtered water volumes are circled. Abundances in (B,D) are percentage abundances of major bacterial Orders. Only Orders with abundance >0.5% (averaged across all replicates) are shown.

The observed compositional shifts with sample volume are potentially driven by a range of mechanisms. First, the filters may be clogging. A minor increase in filtration time per unit sample volume for the 5 L samples compared to the lower volume samples is consistent with filter clogging, although this trend is not supported by the near linear increase in total DNA yield and bacterial 16S gene counts per volume (Figure 4.1A, Figure C.1). If the filters are indeed clogging, clogging would presumably lead to a progressive increase in the size range and diversity of the retained particles, with smaller (potentially free-living) cells below the size threshold of the unclogged filter increasingly retained as volumes increase. We speculate, for example, that Chromatiales and Thiohalorhabdales cells or cell aggregates are potentially either near the size threshold for passage through the prefilter, or are otherwise easily entrained in the filter matrix, and are therefore preferentially retained if the prefilter clogs.

Clogging could also affect diversity estimates for the downstream Sterivex filter fraction. In this fraction, however, effects due to clogging and increased retention of cells <0.2 μm are presumably minimal, as few planktonic marine microbes likely pass through the 0.2 μm filter to begin with (Fuhrman et al., 1988), although the abundance of very small cells in the open ocean remains an active area of research. Rather, the community shifts in the Sterivex fraction could be driven by clogging of the upstream prefilter, with the decline in OTU richness with filter volume observed in both experiments potentially reflecting a selective narrowing of the range of taxa making it through the prefilter. Taxa that increase in abundance with volume in the 0.2μm fraction, such as SAR406, may be relatively small cells, which presumably would be increasingly selected for as the prefilter clogs.

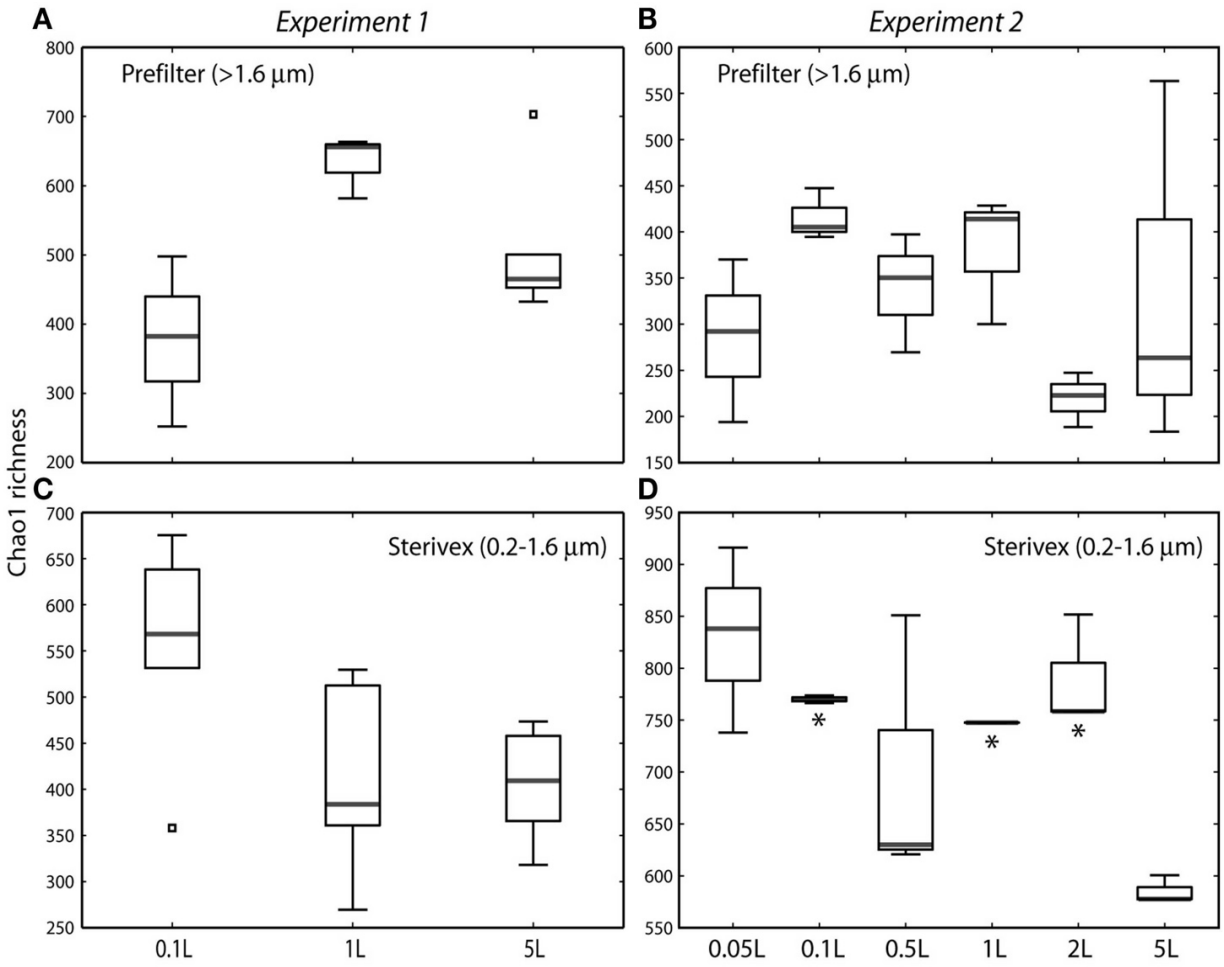


Figure 4.4 Chao1 estimates of operational taxonomic unit (97% similarity cluster) richness in prefilter (>1.6 µm; A,B) and Sterivex (0.2–1.6 µm; C,D) samples in experiments 1 and 2. Boxplots show medians within first and third quartiles, with whiskers indicating the lowest and highest values within 1.5 times the interquartile range of the lower and upper quartiles, respectively. Note variation in y-axis scales. Asterisks in (D) indicate volume groupings with richness estimates significantly different from those of the 5 L sample ($P < 0.05$). All other pairwise differences between volume groupings were not statistically significant.

Second, filtration of small volumes, by chance, may fail to capture a representative subset of the particulate fraction, due to microscale patchiness of bacterioplankton and particle distributions (Azam and Hodson, 1981; Azam and Ammerman, 1984; Mitchell and Fuhrman, 1989). Marine particles span gradients of age, size, and source material, and potentially microbial composition (Simon et al., 2002). It is possible that certain particles, due to their low abundance, are unlikely to be captured at low water volumes, but would be sampled at higher volumes, and potentially contribute significantly to bulk microbial counts depending on the local density and richness of microbes on the particle. Changes in taxon representation due to increases in water volume may not be easily distinguished from changes due to filter clogging.

Third, adsorption of free dissolved DNA to filters may contribute disproportionately to diversity estimates at smaller volumes (Turk et al., 1992; Boström et al., 2004). While we cannot rule out this possibility, proportional decreases of filter-bound DNA with volume seems an unlikely explanation for the observed taxonomic shifts, as the dissolved DNA pool would need to be dominated almost exclusively by Vibrionales for this to be true. Also, DNA extraction efficiency has been shown to decrease with filtered water volume (Boström et al., 2004), suggesting that variable extraction efficiencies may affect community composition estimates. A consistent decrease in extraction efficiency with volume was not clearly evident in our samples, although the range in DNA recovered from 5 L samples indicates that efficiency was variable (Figure 4.1A). Although the potential for extraction efficiency to change with sample volume deserves more attention, and is likely variable among filter types and

extraction methods, decreasing efficiencies would only alter composition estimates if the change in efficiency is non-uniform among taxa. It is unknown if this is true.

Finally, it is possible that population turnover (growth or lysis) in the source water during filtering may also have contributed to compositional shifts with volume. We consider this possibility to be unlikely. Growth in the source water was likely negligible given characteristic doubling times of marine microbes (hours to days; Ducklow, 2000) and the low temperature of the stored source water (stored at 4◦C, but filtered at room temperature). In both experiments, the total processing time of 0.1 and 1 L samples was less than 15 min. Growth by even the most active microbes under optimal conditions would be unlikely to result in the dramatic community structural shifts observed between these sample volumes (Figures 4.2, 4.3). Furthermore, the observed taxonomic shifts are inconsistent with changes due to differential turnover of microbial groups. Taxa typically associated with rapid growth during bottle incubations and in response to organic matter enrichment, notably members of the Alteromonadales, Pseudomonadales, and Vibrionales (Pinhassi and Berman, 2003; Allers et al., 2007, 2008; Stewart et al., 2012), decreased in abundance with increasing volume in both experiments (Tables S3, S4), inconsistent with results due to rapid growth during filtering. Furthermore, had the observed changes been due to growth, or to differential lysis of cells during sampling, we might predict changes of similar magnitude in both biomass fractions. This was not the case, as the most dramatic changes were observed in the prefilter biomass, consistent with effects due to filter clogging.

4.6 Conclusion

These results, based on biomass from an open ocean environment and collected following a widely used filtration scheme, highlight a potential for sample volume variation to confound community diversity estimates. However, it is unlikely that our results extrapolate evenly to all systems. The magnitude and exact mechanism of sample volume-based biases likely differ depending on filter type and pore size, pumping rate, community composition, bulk microbial abundance, particle load, and potentially other variables. Such biases may be even greater in waters more eutrophic than those tested here, and may not easily be eliminated by frequent replacement of prefilters. For studies comparing particle-associated vs. free-living communities, elimination of the prefiltration step may be necessary, with collection of particle communities based instead on direct sampling of particles without filtration (DeLong et al., 1993), potentially via gravity or flow-cytometric separation. Such alternative methods should be explored. Our data suggest that measurements of community structure for biomass separated by filter fractionation should only be considered accurate and used for quantitative comparisons when effects of sample volume variation are shown to be negligible.

Acknowledgments

We thank the crew of the R/V New Horizon for help in sample collection. This work was supported by the National Science Foundation (1151698 to FS) and the Sloan Foundation (RC944 to FS). The authors declare that we have no competing commercial interests that conflict with publication of this manuscript.

Disclaimer: This chapter was published with the same title, along with the supplementary material in Appendix C, in the journal Frontiers in Microbiology on 2 June 2015.

Citation: **Padilla, C. C., Ganesh, S., Gantt, S., Huhman, A., Parris, D. J., Sarode, N., & Stewart, F. J.** (2015). Standard filtration practices may significantly distort planktonic microbial diversity estimates. *Front Microbiol*, *6*, 547-547.

4.7 References

Acinas, S. G., Antón, J., and Rodríguez-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. Microbiol. 65, 514–522.

Allen, L. Z., Allen, E. E., Badger, J. H., McCrow, J. P., Paulsen, I. T., Elbourne, L. D., et al. (2012). Influence of nutrients and currents on the genomic composition of microbes across an upwelling mosaic. ISME J. 6, 1403–1414. doi: 10.1038/ismej.2011.201

Allers, E., Gomez-Consarnau, L., Pinhassi, J., Gasol, J. M., Simek, K., and Pernthaler, J. (2007). Response of Alteromonadaceae and Rhodobacteriaceae to glucose and phosphorus manipulation in marine mesocosms. Environ. Microbiol. 9, 2417–2429. doi: 10.1111/j.1462-2920.2007.01360.x

Allers, E., Niesner, C., Wild, C., and Pernthaler, J. (2008). Microbes enriched in seawater after the addition of coral mucus. Appl. Environ. Microbiol. 74, 3274–3278. doi: 10.1128/AEM.01870-07

Azam, F., and Ammerman, J. W. (1984). "Cycling of organic matter by bacterioplankton in pelagic marine ecosystems: microenvironmental considerations," in Flows of Energy and Materials in Marine Ecosystems, ed M. J. R. Fasham (New York, NY: Plenum Press), 345–360.

Azam, F., and Hodson, R. E. (1981). Multiphasic kinetics for D-glucose uptake by assemblages of natural marine bacteria. Mar. Ecol. Prog. Ser. 6, 213–222. doi: 10.3354/meps006213

Bižic´-Ionescu, M., Zeder, M., Ionescu, D., Orlic´, S., Fuchs, B. M., Grossart, H. P., et al. (2014). Comparison of bacterial communities on limnic versus coastal marine particles reveals profound differences in colonization. Environ. Microbiol. doi: 10.1111/1462-2920.12466. [Epub ahead of print].

Boström, K. H., Simu, K., Hagström, Å., and Riemann, L. (2004). Optimization of DNA extraction for quantitative marine bacterioplankton community analysis. Limnol. Oceanogr. Meth. 2, 365–373. doi: 10.4319/lom.2004.2.365

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high- throughput community sequencing data. Nat. Methods. 7, 335–336. doi: 10.1038/nmeth.f.303

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. U.S.A. 108(Suppl. 1), 4516–4522. doi: 10.1073/pnas.1000080107

Cho, B. C., and Azam, F. (1988). Major role of bacteria in biogeochemical fluxes in the ocean's interior. Nature. 332, 441–443. doi: 10.1038/332441a0

Crump, B., Armbrust, E., and Baross, J. (1999). Phylogenetic analysis of particle-attached and free-living bacterial communities in the Columbia river, its estuary, and the adjacent coastal ocean. Appl. Environ. Microbiol. 65, 3192–3204.

D'Ambrosio, L., Ziervogel, K., MacGregor, B., Teske, A., and Arnosti, C. (2014). Composition and enzymatic function of particle-associated and free- living bacteria: a coastal/offshore comparison. ISME J. 8, 2167–2179. doi: 10.1038/ismej.2014.67

DeLong, E. F., Franks, D. G., and Alldredge, A. L. (1993). Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. Limnol. Oceanogr. 38, 924–934. doi: 10.4319/lo.1993.38.5.0924

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 72, 5069–5072. doi: 10.1128/AEM.03006-05

Ducklow, H. W. (2000). "Bacterial production and biomass in the oceans," in Microbial Ecology of the Oceans, 2nd Edn., ed D. L. Kirchman (Hoboken, NJ: John Wiley & Sons Inc.), 85–120.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Eloe, E. A., Shulse, C. N., Fadrosh, D. W., Williamson, S. J., Allen, E. E., and Bartlett, D. H. (2011). Compositional differences in particle-associated and free-living microbial assemblages from an extreme deep-ocean environment. Environ. Microbiol. Rep. 3, 449–458. doi: 10.1111/j.1758-2229.2010. 00223.x

Frias-Lopez, J., Shi, Y., Tyson, G. W., Coleman, M. L., Schuster, S. C., Chisholm, S. W., et al. (2008). Microbial community gene expression in ocean surface waters. Proc. Natl. Acad. Sci. U.S.A. 105, 3805–3810. doi: 10.1073/pnas.0708897105

Fuchsman, C. A., Kirkpatrick, J. B., Brazelton, W. J., Murray, J. W., and Staley, J. T. (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. doi: 10.1111/j.1574-6941.2011.01189.x

Fuchsman, C. A., Staley, J. T., Oakley, B. B., Kirkpatrick, J. B., and Murray, J. W. (2012). Free-living and aggregate-associated Planctomycetes in the Black Sea. FEMS Microbiol. Ecol. 80, 402–416. doi: 10.1111/j.1574-6941.2012. 01306.x

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

Ganesh, S., Bristow, L. A., Larsen, M., Sarode, N., Thamdrup, B., and Stewart, F. J. (2015). Size-fraction partitioning of community gene transcription and nitrogen metabolism in a marine oxygen minimum zone. ISME J. doi: 10.1038/ ismej.2015.44. [Epub ahead of print].

Ganesh, S., Parris, D. J., DeLong, E. F., and Stewart, F. J. (2014). Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. ISME J. 8, 187–211. doi: 10.1038/ismej.2013.144

Gasol, J. M., and Moran, A. G. (1999). Effects of filtration on bacterial activity and picoplankton community structure as assessed by flow cytometry. Aquat. Microb. Ecol. 16, 251–264. doi: 10.3354/ame016251

Ghiglione, J. F., Conan, P., and Pujo-Pay, M. (2009). Diversity of total and active free-living vs. particle-attached bacteria in the eutrophic zone of the NW Mediterranean Sea. FEMS Microbiol. Lett. 299, 9–21. doi: 10.1111/j.1574-6968.2009.01694.x

Ghiglione, J. F., Mevel, G., Pujo-Pay, M., Mousseau, L., Lebaron, P., and Goutx, M. (2007). Diel and seasonal variations on abundance, activity and community structure of particle-attached and free-living bacteria in NW Mediterranean Sea. Microb. Ecol. 54, 217–231. doi: 10.1007/s00248-006-9189-7

Grossart, H. P. (2010). Ecological consequences of bacterioplankton lifestyles: changes in concepts are needed. Environ. Microbiol. Rep. 2, 706–714. doi: 10.1111/j.1758-2229.2010.00179.x

Hardcastle, T. J., and Kelly, K. A. (2010). baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics 11:422. doi: 10.1186/1471-2105-11-422

Hatt, J. K., Ritalahti, K. M., Ogles, D. M., Lebron, C. A., and Loffler, F. E. (2013). Design and application of an internal amplification control to improve Dehalococcoides mccartyi 165 rRNA gene enumeration by qPCR. Environ. Sci. Tech. 47, 11131–11138. doi: 10.1021/es4019817

Hollibaugh, J. T., Wong, P. S., and Murrell, M. C. (2000). Similarity of particle-associated and free-living bacterial communities in northern San Francisco Bay, California. Aquat. Microb. Ecol. 21, 103–114. doi: 10.3354/ame 021103

Hunt, D. E., Lin, Y., Church, M. J., Karl, D. M., Tringe, S. G., Izzo, L. K., et al. (2013). Relationship between abundance and specific activity of bacterioplankton in open ocean surface waters. Appl. Environ. Microbiol. 79, 177–184. doi: 10.1128/AEM.02155-12

Kellogg, C. T. E., and Deming, J. W. (2009). Comparison of free-living, suspended particle, and aggregate-associated bacterial and archaeal communities in the Laptev Sea. Aquat. Microb. Ecol. 57, 1–18. doi: 10.3354/ame01317

Kennedy, K., Hall, M. W., Lynch, M. D., Moreno-Hagelsieb, G., and Neufeld, J. D. (2014). Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles. Appl. Environ. Microbiol. 80, 5717–5722. doi: 10.1128/AEM.01451-14

Kirchman, D. L., Yu, L. Y., Fuchs, B. M., and Amann, R. (2001). Structure of bacterial communities in aquatic systems as revealed by filter PCR. Aquat. Microb. Ecol. 26, 13–22. doi: 10.3354/ame026013

Knefelkamp, B., Carstens, K., and Wiltshire, K. H. (2007). Comparison of different filter types on chlorophyll-a retention and nutrient measurements. J. Exp. Mar. Biol. Ecol. 345, 61–70. doi: 10.1016/j.jembe.2007.01.008

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., and Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl. Environ. Microbiol. 79, 5112–5120. doi: 10.1128/AEM. 01043-13

LaMontagne, M. G., and Holden, P. A. (2003). Comparison of free-living and particle-associated bacterial communities in a coastal lagoon. Microb. Ecol. 46, 228–237. doi: 10.1007/s00248-001-1072-y

Lee, S., Kang, Y. C., and Fuhrman, J. A. (1995). Imperfect retention of natural bacterioplankton cells by glass fiber filters. Mar. Ecol. Prog. Ser. 119, 285–290. doi: 10.3354/meps119285

Long, R. A., and Azam, F. (2001). Microscale patchiness of bacterioplankton assemblage richness in seawater. Aquat. Microb. Ecol. 26, 103–113. doi: 10.3354/ame026103

Lozupone, C., and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. Appl. Environ. Microb. 71, 8228–8235. doi: 10.1128/AEM.71.12.8228-8235.2005

Magocˇ, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27, 2957–2963. doi: 10.1093/bioinformatics/btr507

Mitchell, J. G., and Fuhrman, J. A. (1989). Centimeter scale vertical heterogeneity in bacteria and chlorophyll a. Mar. Ecol. Prog. Ser. 54, 141–148. doi: 10.3354/meps054141

Moeseneder, M. M., Winter, C., and Herndl, G. J. (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. doi: 10.4319/lo.2001.46.1.0095

Mohit, V., Archambault, P., Toupoint, N., and Lovejoy, C. (2014). Phylogenetic differences in attached and free-living bacterial communities in a temperate coastal lagoon during summer, revealed via high-throughput 16S rRNA gene sequencing. Appl. Environ. Microbiol. 80, 2071–2083. doi: 10.1128/AEM.02916-13

Morris, R. M., and Nunn, B. L. (2013). Sample preparation and processing for planktonic microbial community proteomics. Method. Enzymol. 531, 271–287. doi: 10.1016/b978-0-12-407863-5.00014-9

Nagata, T. (1986). Carbon and nitrogen content of natural planktonic bacteria. Appl. Environ. Microbiol. 52, 28–32.

Orsi, W. D., Smith, J. M., Wilcox, H. M., Swalwell, J. E., Carini, P., Worden, A. Z., et al. (2015). Ecophysiology of uncultivated marine euryarchaea is linked to particulate organic matter. ISME J. doi: 10.1038/ismej.2014.260. [Epub ahead of print].

Pinhassi, J., and Berman, T. (2003). Differential growth response of colony- forming α- and γ-proteobacteria in dilution culture and nutrient addition experiments from Lake Kinneret (Israel), the eastern Mediterranean Sea, and the Gulf of Eilat. Appl. Environ. Microbiol. 69, 199–211. doi: 10.1128/AEM.69.1.199-211.2003

Rath, J., Wu, K. Y., Herndl, G. J., and and, E. F., DeLong (1998). High phylogenetic diversity in a marine-snow-associated bacterial assemblage. Aquat. Microb. Ecol. 14, 261–269. doi: 10.3354/ame014261

Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., et al. (2007). The Sorcerer II global ocean sampling expedition: northwest Atlantic through Eastern Tropical Pacific. PLoS Biol. 5:e77. doi: 10.1371/journal.pbio.0050077

Simon, H. M., Smith, M. W., and Herfort, L. (2014). Metagenomic insights into particles and their associated microbiota in a coastal margin ecosystem. Front. Microbiol. 5:466. doi: 10.3389/fmicb.2014.00466

Simon, M., Grossart, H. P., Schweitzer, B., and Ploug, H. (2002). Microbial Ecology of organic aggregates in aquatic ecosystems. Aquat. Microb. Ecol. 28, 175–211. doi: 10.3354/ame028175

Smith, M. W., Allen, L. Z., Allen, A. E., Herfort, L., and Simon, H. M. (2013). Contrasting genomic properties of free-living and particle-attached microbial assemblages within a coastal ecosystem. Front. Microbiol. 4:120. doi: 10.3389/fmicb.2013.00120

Stewart, F. J. (2013). Preparation of microbial community cDNA for metatranscriptomic analysis in marine plankton. Method. Enzymol. 531, 187–218. doi: 10.1016/b978-0-12-407863-5.00010-1

Stewart, F. J., Dalsgaard, T., Thamdrup, B., Revsbech, N. P., Ulloa, O., Canfield, D. E., et al. (2012). Experimental perturbation and oxygen addition elicit profound changes in community transcription in OMZ bacterioplankton. PLoS ONE 7:e37118. doi: 10.1371/journal.pone.0037118

Taguchi, S., and Laws, E. A. (1988). On the microparticles which pass through glass fiber type GF/F in coastal and open waters. J. Plankton. Res. 10, 999–1008. doi: 10.1093/plankt/10.5.999

Turk, V., Rehnstam, A. S., Lundberg, E., and Hagström, Å. (1992). Release of bacterial DNA by marine nanoflagellates, an intermediate step in phosphorus regeneration. Appl. Environ. Microbiol. 58, 3744–3750.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. Science 304, 66–74. doi: 10.1126/science.1093857

Walsh, D. A., Zaikova, E., and Hallam, S. J. (2009). Large Volume (20L+) Filtration of coastal seawater samples. J. Vis. Exp. 28:e1161. doi: 10.3791/1161

Wright, J. J., Konwar, K. M., and Hallam, S. J. (2012). Microbial ecology of expanding oxygen minimum zones. Nat. Rev. Microbiol. 10, 381–394. doi: 10.1038/nrmicro2778

Wright, J. J., Mewis, K., Hanson, N. W., Konwar, K. M., Maas, K. R., and Hallam, S. J. (2014). Genomic properties of Marine Group A bacteria indicate a role in the marine sulfur cycle. ISME J. 8, 455–468. doi: 10.1038/ismej.2013.152

CHAPTER 5

METABOLIC PLASTICITY OF A PELAGIC ANAMMOX BACTERIUM OVER REDOX GRADIENTS IN A MARINE OXYGEN MINIMUM ZONE

5.1 Abstract

The process of anaerobic ammonia oxidation (anammox) accounts for up to half of the total fixed nitrogen loss from the world's oceans, and is a primary nitrogen removal pathway in marine Oxygen Minimum Zones (OMZs). The anammox process appears to be mediated exclusively by bacteria from the phylum *Planctomycetes*, with members of the genus *Candidatus* Scalindua being the dominant anammox contributors in marine settings. Prior genomic characterizations of marine *Candidatus* Scalindua from non-OMZ environments indicate metabolic versatility, suggesting the potential for diverse biochemical contributions depending on environmental conditions. However, the environmental partitioning of genomic variation and metabolic activity by pelagic *Ca.* Scalindua in OMZs remains unclear. Here, we sequence and analyze single amplified genomes (SAGs) of 20 individual *Ca.* Scalindua cells from the Eastern Tropical North Pacific OMZ. The analysis revealed genes for cyanate and urea degradation by OMZ *Ca*. Scalindua, suggesting the potential use of these substrates as sources of ammonia for anammox. Genome content overall appeared relatively uniform across individuals spanning both the nitrite maximum and OMZ core zones, suggesting a relative lack of partitioning of genetic variation among ecotypes. Rather, potential metabolic variation was manifest in the differential expression of gene content, inferred by mapping of metatranscriptome sequences originating from samples that represented gradients in

nitrogen substrate availability and organic matter supply on SAGs genomes. In general, transcriptome patterns for genes in the anammox pathway were consistent with variation in anammox rates, which increased steeply from the base of the oxycline to the nitrite maximum in the upper OMZ, before declining with depth and with distance from shore. These data represent the first genomic sequences from *Ca.* Scalindua from an OMZ, and suggest a potential for previously unrecognized contributions by anammox-capable bacteria to other biochemical pathways and utilization of nitrogen compounds other than ammonia in OMZs, highlighting the spatial variability in metabolic activity by this ubiquitous OMZ clade.

5.2 Introduction

Ocean oxygen minimum zones (OMZs) are formed in regions of the ocean where large nutrient input and oceanic upwelling, combined with little to no mixing and low ventilation, leads to anoxic conditions in the mid water depths (Ulloa et al., 2012). The largest permanent OMZs in the world's oceans occur in the Eastern Tropical North Pacific (ETNP), off the coast of Mexico and in the Eastern Tropical South Pacific (ETSP), off the coast of Chile. These OMZs are characterized by little to no oxygen (<100nM) and an accumulation of nitrite at mid-water depths (~150-750m) (Kamykowski et al; 1991). Serving as habitats to taxonomically and functionally diverse Bacteria and Archaea, OMZs are sites of globally important biogeochemical cycling. In the absence of oxygen the microbial communities mediate diverse microaerophilic and anaerobic metabolic processes including sulfate reduction, denitrification, sulfur oxidation and anaerobic ammonia oxidation. Among the microbially mediated

biogeochemical transformations are the processes of denitrification and anaerobic ammonia oxidation (anammox), which contribute up to half of the total fixed nitrogen loss from oceans as N_2 and N_2O (Codispoti et al., 2004; Gruber, 2001), generating large nitrogen deficits, and making OMZs global sinks for nitrogen. Prior research in the Eastern Pacific OMZs suggests that anammox and denitrification are linked to other microbial N transformations to varying degrees, including the production of nitrite and nitrate via nitrification along the oxycline, and dissimilatory nitrate reduction to ammonium within the OMZ (Lam et al., 2009; Kalvelage et al., 2013). OMZ N cycle processes also mediate key steps in other elemental cycles, for example, the fixation of carbon by anammox and nitrifying autotrophs, the remineralization of organic carbon and nutrients by denitrifying heterotrophs, and the oxidation of reduced sulfur compounds by chemoautotrophic denitrifiers (Ulloa et al., 2012). However, the potential for these diverse OMZ microbes to alter their physiology among different ecological niches along the OMZ redoxcline remains unclear. Here, we explore this potential within a ubiquitous clade of marine bacteria that mediate the anammox process.

The anammox process, which involves the conversion of ammonium and nitrite to dinitrogen gas, is carried out by a deep branching monophyletic group of bacteria within the phylum Planctomycetes. The identified anammox-specific Planctomycete lineages include *Ca.* Brocadia (Kartal et al., 2008), *Ca*. Kuenenia (Schmid et al., 2000), *Ca*. Anammoxoglobus (Kartal et al., 2007), *Ca*. Jettenia (Quan et al., 2008), and *Ca*. Scalindua (Schmid et al., 2003). Anammox capable organisms have been detected globally in natural and man-made environments, and in freshwater and marine ecosystems (Kartal et al., 2008), with *Ca*. Scalindua constituting the dominant genera of

anammox bacteria in marine habitats, including OMZs (Villanueva et al., 2014). Thus far, none of these bacteria have been obtained in a pure culture (Speth et al., 2012) and genomic data exist for only a few anammox bacteria (Strous et al., 2006; Gori et al., 2011; Hira et al., 2012; Hu et al., 2012; Speth et al., 2012; van de Vossenberg et al., 2012). These data include a single metagenome from a marine sediment enrichment of *Ca.* Scalindua profunda and a near complete draft genome from *Ca.* Scalindua brodae (Speth et al., 2015). However, genomic insight into OMZ *Ca.* Scalindua is limited to fragments of sequence information obtained from environmental metagenomes or metatranscriptomes (e.g., Stewart et al., 2012; Ganesh et al., 2014, 2015), thereby constraining conclusions regarding the extent of ecotype or metabolic diversity in pelagic marine settings.

Several studies have provided evidence of genetic and metabolic diversity in anammox bacteria. Genetically distinct subclades (based on 16S -23S rRNA gene internal transcribed spacer sequences) of marine *Ca.* Scalindua were identified among sites with varying geochemical conditions (Woebken et al., 2008). This pattern suggests the potential for biogeographic separation or for variable ecotypes of Scalindua based on environmental niche, with the latter potentially suggestive of metabolic diversity. Indeed, in addition to occurring in distinct low oxygen zones (Woebken et al., 2008; Galan et al., 2012), *Ca.* Scalindua can be found both attached to marine particles (Woebken et al., 2007) and in planktonic free-living fractions in the water column (Ganesh et al., 2014; Ganesh et al., 2015), albeit varying significantly in abundance between fractions. In our previous studies, we observed variation in the representation of Scalindua-associated genes in metatranscriptomes from particle-associated fractions and non-particle-

associated microbial fractions within the anoxic ETNP OMZ (Ganesh et al., 2015), suggesting the potential for these bacteria to carry out distinct functions based on micro-niche partitioning.

Such partitioning may be driven by the availability of nutrient and energy substrates. Enrichment-based studies have confirmed that *Ca*. Scalindua bacteria, although likely autotrophic while conducting anammox, also possess the ability to oxidize small organic acids like formate, acetate and propionate, and use other electron acceptors like Fe (III) and Mn (IV) (van de Vossenberg et al., 2008). In our prior work in the ETNP OMZ, genes associated with a potential for heterotrophic nitrate reduction (e.g., encoding dissimilatory nitrate reductase) were shown to be generally enriched in OMZ Ca. Scalindua communities associated with particles, whereas genes of the anammox process were more abundant in the transcriptomes of non-particle associated communities (Ganesh et al. 2015). Such plasticity may allow *Ca*. Scalindua to capitalize on patches of organic matter enrichment (e.g., on particles) and to persistent in regions of the OMZ where anammox substrates are less abundant, potentially including lower in the OMZ core beneath the nitrite maximum.

Metabolic plasticity in OMZ anammox bacteria may also manifest in the ability to use diverse nitrogen-containing compounds as sources for anammox substrates. For example, the reduced compound cyanate (OCN^-), which typically occurs at nanomolar concentrations in estuarine and marine systems (Rees et al, 2006; Widner et al., 2013), may contribute to the nitrogen and carbon requirements of marine microbial communities (Palenik et al., 2003; Rocap et al, 2003; Kamennaya et al, 2011, Widner et al, 2013). Although cyanate is generally toxic to organisms because of its reactivity with

nucleophilic groups in proteins (Palatinszsky et al., 2015; Kamennaya et al., 2011), many microorganisms possess genes encoding cyanases that catalyze cyanate breakdown, via reaction with bicarbonate, to carbon dioxide and ammonia (Luque-Almagro et al., 2008, Kamennaya et al., 2008). Indeed, a single copy of a cyanase gene is present in the metagenome of *Ca.* Scalindua *profunda* enriched from marine sediment. Cyanases likely function predominantly in detoxification but may also benefit cells if the ammonia resulting from cyanase activity can be used either for assimilation or as an energy source. Indeed, recent work shows that *Nitrososphaera gargensis*, an ammonia-oxidizing thaumarcheote, can grow on cyanate as an energy source (Palatinszky et al., 2015). A similar use for cyanate as an ammonia source for marine anammox bacteria remains unproven, but may be important in OMZs if ammonia is limiting. Urea, another organic nitrogen substrate and precursor to ammonia, may also be utilized as an energy source by anammox bacteria possessing genes for ureases, which breakdown urea to ammonia. Compared to cyanate, urea may be more readily available in marine systems from phytoplankton release of dissolved organic nitrogen (Bronk and Gilbert, 1991; Berman et al., 1999) or, remineralization of organic matter or anthropogenic inputs. Assessing the genomic potential for use of diverse nitrogen sources by anammox bacteria would lend support to the ability of these bacteria to occupy niches with variable substrate regimes. These trends lead us to hypothesize that 1) *Ca*. Scalindua bacteria in OMZs possess the genomic potential for diverse biogeochemical transformations in addition to canonical anammox, potentially including heterotrophy or alternative nitrogen substrate utilization for anammox, and 2) the pathways mediating these metabolisms are differentially expressed over vertical and spatial gradients in the OMZ depending on substrate

(ammonia and nitrite) availability. Alternatively, anammox-capable bacteria in OMZs may represent distinct ecotypes that differ in genome content, in addition to gene expression. We test these hypotheses using single cell genomics coupled with metatranscriptomics and high-resolution measurements of anammox rates in the ETNP OMZ off northern Mexico. The results provide an understanding of the metabolic diversity and activity of anammox bacteria operating in different environmental niches, thereby providing an understanding of per-cell differences in biochemical processing and the genetic basis for variation and adaptation in this important OMZ lineage.

5.3 Materials and Methods

5.3.1 Sample collection

Sample collection in the ETNP. Samples for Single Cell Genomes were collected from the ETNP OMZ during the Oxygen Minimum Zone Microbial Biogeochemistry Expedition (OMZoMBiE) cruise (*R/V New Horizon*, 13-28 June, 2013) from Station 6 (18◦ 54.0N, 104◦ 54.0W). Seawater for single cell sorting and single amplified genome (SAG) analysis was collected from two depths within the OMZ, the secondary nitrite maximum OMZ (125 m) and OMZ core (300 m). Collections were made using Niskin bottles on a rosette containing a Conductivity-Temperature-Depth profiler (Sea-Bird SBE 911plus). Water samples for were preserved by cryopreservation according to the protocol recommended by the Bigelow Single Cell Genomics Center. Briefly, triplicate 1 ml samples of bulk seawater (no pre-filtration) were aliquoted into sterile cryovials and 100 µl of a glycerol TE stock solution (20 ml 100X TE pH 8.0, 60 ml sterile deionized

water, 100 ml molecular grade glycerol) was added to each. The vials were gently mixed and then frozen at -80ºC.

Biomass samples for RNA processing were collected from the ETNP OMZ during the Oxygen Minimum Zone Microbial Biogeochemistry Expedition 2 (OMZoMBiE2) cruise (*R/V New Horizon*; May 10 - June 8, 2014). Seawater samples were collected at six stations along an east-west transect spanning a coastal to off-shore gradient (Figure 1). Stations and water column depths sampled for RNA processing are summarized in Table 1. Seawater collection strategy was the same as described above. Seawater samples were collected from discrete depths and microbial biomass was collected by sequential in-line filtration of seawater (~2-4 L) through a glass fiber disc prefilter (GF/A, 47 mm, 1.6 μm pore-size, Whatman) and a primary collection filter (Sterivex™, 0.22 μm pore-size, Millipore) using a peristaltic pump. Sterivex filters were filled with RNA stabilizing buffer (25 mM Sodium Citrate, 10 mM EDTA, 5.3M Ammonium sulfate, pH 5.2), flash-frozen in liquid nitrogen, and stored at -80°C. Less than 20 min elapsed between sample collection (water on deck) and fixation in buffer.

Collection of samples and measurements of process rates. Seawater for anammox process rate measurements was collected from stations corresponding to the RNA samples (Table 5.1). Water for rate determinations was sampled directly from the Niskin bottle immediately after its arrival onboard and transferred to 250 ml glass bottles. Bottles were overflowed (three volume equivalents) and sealed bubble free with deoxygenated butyl rubber stoppers (De Brabandere et al, 2014). Within 6 hours of

collection, each bottle was amended with 5 μM $^{15}NH_4^+$, and subsequently purged with helium for ~ 20 min. With a slight overpressure, water was dispensed into 12 ml exetainers (Labco, Lampeter, Ceredigion, UK) and immediately capped with deoxygenated lids. Headspaces of 2ml were introduced into each exetainer and flushed twice with helium and incubated in the dark at *in situ* temperature. For each experiment, triplicate exetainers were preserved with 100 μl of 50% (w/v) $ZnCl_2$ at the start of the incubation and again after 24 hours at each depth.

Production of $^{14}N^{15}N$ and $^{15}N^{15}N$ was determined on a gas chromatography isotope ratio mass spectrometer (GC-IRMS; Dalsgaard et al, 2012). Rates of N_2 production by anammox were calculated based on the equations presented in Thamdrup and Dalsgaard, (2002). Anammox rates were calculated from the slope of the linear regression of $^{14}N^{15}N$ with time. T-tests were applied in all cases to determine whether rates were significantly different from zero ($p<0.05$).

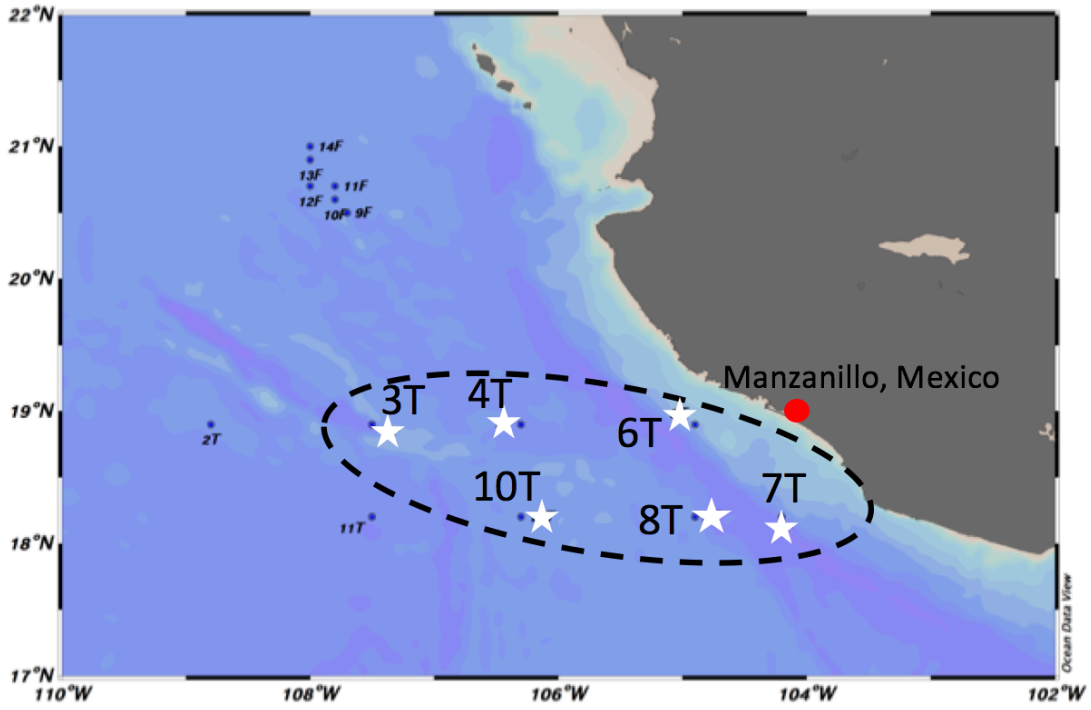


Figure 5.1. Map of sampling sites in the ETNP OMZ during OMZoMBiE2 cruise (*R/V New Horizon*; May 10 - June 8, 2014). Map showing the location of sampling sites (off the coast of Manzanillo, Mexico) was generated in Ocean Data View. Stations marked T were located along an east-west transect of coastal to off-shore gradient.

Table 5.1 Summary of Stations and Depths sampled for RNA processing and Anammox rate measurements in the ETNP OMZ

Station	Latitude N	Longitude W	Depth (m)	Study generated in	Samples for Anammox Rate measurements
6T	18 54.001	104 54.040	75	Padilla et al.,2016	x
6T	18 54.001	104 54.040	80	Garcia-Robledo et al. (In prep)	x
6T	18 54.001	104 54.040	100	Garcia-Robledo et al. (In prep)	x
6T	18 54.001	104 54.040	150	Garcia-Robledo et al. (In prep)	x
6T	18 54.001	104 54.040	200	Garcia-Robledo et al. (In prep)	x
6T	18 54.001	104 54.040	300	Padilla et al.,2016	
3T	18 54.201	107 29.922	113	This study	x
3T	18 54.201	107 29.922	136	This study	x
7T	18 12.023	104 12.160	91	This study	x
7T	18 12.023	104 12.160	99	This study	x
7T	18 12.023	104 12.160	105	This study	x
7T	18 12.023	104 12.160	150	This study	x
7T	18 12.023	104 12.160	250	This study	x
4T	18 53.873	106 17.977	60	Garcia-Robledo et al. (In prep)	
4T	18 53.873	106 17.977	100	Garcia-Robledo et al. (In prep)	
4T	18 53.873	106 17.977	150	Garcia-Robledo et al. (In prep)	
4T	18 53.873	106 17.977	200	Garcia-Robledo et al. (In prep)	
8T	18 11.885	104 53.804	75	This study	x
10T	18 12.047	106 17.797	100	Padilla et al.,2016	
10T	18 12.047	106 17.797	150	Padilla et al.,2016	
10T	18 12.047	106 17.797	300	Padilla et al.,2016	

x represents depths sampled for anammox rate measurements

5.3.2 Single cell sorting, multiple displacement amplification, and SAG 16S rRNA gene screening

Single amplified genomes (SAGs) were generated according to standard procedures in the Department of Energy Joint Genome Institute workflow involving single cell sorting, whole genome amplification, and 16S rRNA screening following the procedures outlined previously (Rinke et al., 2013) with minor modifications. Briefly, sorted cells were treated with Ready-Lyse lysozyme (Epicentre; 5U/μL final conc.) for 15 min at room temperature prior to the addition of lysis solution. Whole genome multiple displacement amplification was performed with the REPLI-g Single Cell Kit (Qiagen) with final reaction volumes of 2 uL. Amplification reactions were terminated after 6 hr. Preliminary taxonomic identity was assigned to each of the SAGs by PCR amplification and Sanger sequencing of a ~470 bp region of the 16S rRNA gene (amplified using primers 926wF (5′-AAACTYAAAKGAATTGRCGG3′) and 1392R (5′-ACGGGCGGTGTGTRC3′) for archaea and bacteria), followed by comparisons to the Greengenes rRNA database.

5.3.3 Single Amplified Genome sequencing

A total of 20 SAGs identified by 16S rRNA screening as belonging to *Ca.* Scalindua were selected for Illumina MiSeq sequencing. These included 9 and 11 SAGs from 125m and 300m in the ETNP, respectively. SAG DNA sequencing libraries were prepared using the Nextera XT DNA Library Prep kit (Illumina, San Diego, CA, USA) to generate indexed libraries following manufacturer instructions. Libraries were pooled and

sequenced at Georgia Tech on a single run of an Illumina MiSeq using a v2-500 cycle (paired end 250 x 250bp) kit.

5.3.4 RNA extraction and cDNA sequencing

RNA was extracted following the protocol in Ganesh et al. (2015). Briefly, RNA was extracted from Sterivex filters using a modification of the mirVana™ miRNA Isolation kit (Ambion). Filters were thawed on ice, and RNA stabilizing buffer was expelled via syringe from Sterivex cartridges and discarded. Cells were lysed by adding Lysis buffer and miRNA Homogenate Additive (Ambion) directly to the cartridge. Following vortexing and incubation on ice, lysates were transferred to RNAase-free tubes and processed via acid phenol:chloroform extraction according to the kit protocol. The TURBO DNA-free™ kit (Ambion) was used to remove DNA, and the extract was purified using the RNeasy MinElute Cleanup Kit (Qiagen).

Community cDNA sequencing was used to characterize microbial gene transcription in biomass (Sterivex filter fraction). Community RNA was prepared for sequencing using the ScriptSeq™ v2 RNA-Seq Library preparation kit (Epicentre). cDNA was synthesized from fragmented total RNA (rRNA was not removed) using reverse transcriptase and amplified and barcoded using ScriptSeq™ Index PCR Primers (Epicenter) to generate single-indexed cDNA libraries. cDNA libraries were pooled and sequenced on an Illumina MiSeq using a v2-500 cycle (250x250bp) kit.

5.3.5 SAG sequence analysis

Illumina reads were filtered by quality with a Phred score cutoff of 25 and trimmed using TrimGalore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). High-quality paired reads were merged using FLASH (Magoc and Salzberg, 2011). Quality-trimmed merged, and unmerged reads were assembled using the SPAdes assembler (Bankevich, 2012) with k-mer sizes of 21,33,55,77,99,127, and the single-cell (-sc) option. Percentage of contamination and genome completeness were assessed based on recovery of lineage-specific marker gene sets using CheckM (Parks et al., 2015). Coding sequences were predicted on contigs using GeneMark.hmm (Lukashin and Borodovsky, 1998) and 16S rRNA gene sequences were identified using RNAmmer (Lagesen et al., 2007), both using default parameters. Only SAGs that had greater than 1000 predicted genes were used for further analysis.

For the assemblies that included a 16S rRNA sequence (n=8 SAGs), the assembled 16S sequence was compared to the 470 bp 16S fragment obtained during the initial SAG screening at the DOE JGI and confirmed to be identical. The full-length 16S rRNA gene sequences obtained from SAGs were aligned with reference sequences from environmental clones and closely related sequenced representatives obtained from GenBank, using Clustal Omega (Sievers and Higgins, 2014). The alignments were trimmed and a maximum likelihood analysis of the sequence alignment was carried out in MEGA7 (Kumar et al., 2015) based on the Kimura 2- parameter model (Kimura M,

1980) and 100 bootstraps. A discrete gamma distribution was used to model evolutionary rate differences among sites (+G; parameter=0.05).

Predicted protein coding genes were functionally annotated using the blast2go pipeline (Conesa et al., 2005) and the BLASTKOALA pipeline (Kanehisa et al., 2016) for assignment of EC numbers and mapping to metabolic pathways. Functional assignments were manually screened via text searches to identify genes for anammox metabolism *(hydrazine synthase, hydrazine/hydroxylamine oxidoreductase)*, central carbon metabolism *(carbon monoxide dehydrogenase, formate dehydrogenase)* and other nitrogen metabolism associated genes *(nitrate reductase/nitrite oxidoreductase, ammonium transporters, cyanate metabolism genes, ureases and urea transporters)*.

To determine relatedness among the SAGs, Average Amino Acid Identity (AAI) was calculated between each pair of SAGs using the aai.rb script from the enveomics toolkit (Rodriguez-R and Konstantinidis, 2016). The get_homologues software package (Contreras-Moreira, 2013), which uses reciprocal best BLAST matches followed by identification of shared gene clusters with the OrthoMCL (Li L et al., 2003) algorithm, was used to identify genes shared between each SAG and reference Ca. Scalindua *brodae* protein coding sequences. Ca. Scalindua brodae was selected as a reference as it was the closest relative to the OMZ SAGs (based on 16S similarity) with an almost complete (92%) draft genome available.

A custom cyanate hydratase database was created comprising of genes found in IMG genomes and IMG metagenomes annotated as cyanate lyase/hydratase, as well as Uniprot entries for cyanate lyase. Phylogenetic analysis of putative cyanate hydratase sequences identified in the SAGs was carried out with 100 reference sequences from this

custom database comprising of sequences from ammonia oxidizing archaea including *Nitrosospheaera gargensis*, nitrite oxidizing bacteria and several broad level taxa. Sequences were aligned using MAFFT (Katoh, 2013) and a phylogeny was inferred using RAxML implemented in the program Practical Alignment using SATé and TrAnsitivity (PASTA; Mirarab, S. et al., 2014).

5.3.6 Metatranscriptome analysis

Non-rRNA reads from the metatranscriptome datasets from stations and depths listed in Table 5.1 were queried against databases created from individual SAG amino acid sequences for a best-BLAST match. Only protein coding sequences from SAGs that had >1000 predicted genes were used for this analysis (n=8). The BLAST output was parsed to identify transcripts recruiting to SAGs with a bit score > 50 and >95% amino acid identity. Counts of mapped transcripts for each unique protein coding gene were normalized to a uniform sequence depth of 2500 reads matching the SAGs to facilitate comparison between datasets. Scalindua transcriptional abundance is expressed as the total number of mRNA transcripts that found a significant match against a SAG protein database, normalized to the total dataset size (total non-rRNA reads).

For each sample, read counts per protein-coding gene were normalized to a total of 2500 sequences matching the SAG database, for comparison between samples. Normalized read counts were used for a Canonical Correspondence Analysis (CCA) alongside environmental measurements of water column nitrite, oxygen, anammox rates, and depth, using the program Past3 (Hammer and Harper, 2001).

Tests for differential expression of functional genes between two distinct depth

zones, the nitrite maximum (125-150m) in the upper OMZ and the OMZ core (200m-300m) where nitrite levels are lower, were carried out in DESeq2 (Love et al., 2014). Due to lack of biological replicates from each sampling site, samples from different sites were grouped as replicates for each depth zone.

5.4 Results and Discussion

OMZ chemical conditions and spatial heterogeneity in anammox rates. Dissolved oxygen concentrations recorded by the SBE43 sensors were highest in the surface waters (~200μM) with oxygen (O_2) concentrations dropping rapidly between 50m and 80m (oxycline). Between 85m and 870m (OMZ core), oxygen concentration was <50 nM, representing the core of the OMZ. As observed previously, a characteristic secondary nitrite maximum (with nitrite concentrations ranging from 3-5 μM) was identified within the anoxic core, spanning the depths between 125m and 150m, with nitrite concentrations then declining at deeper core depths. Samples for RNA analysis and anammox rate measurements spanned this oxygen and nitrite gradient, including samples from the oxycline and OMZ core from six sites on a coastal to off-shore transect in the ETNP (Figure 5.1).

Anammox rates varied substantially over depth gradients and with proximity to shore (Figure 5.2). The highest rates were observed in the upper OMZ with the sharpest rate increases observed at the base of the oxycline upon transition into the anoxic zone and nitrite-enriched layer. For example, rates at station 7T increased from effectively zero at 91 m to 3.1 nM N_2 d^{-1} at 99 m, coincident with an increase in nitrite concentrations (from 0.01μM to 0.39μM). Ammonium concentrations are generally low

in OMZs and decrease substantially with depth (0.01-0.03 μM in the ETSP OMZ; Galan et al., 2009). Elevated anammox rates in the upper OMZ is likely driven by co-availability of ammonium and nitrite in this zone (Ulloa et al., 2012). These rates are consistent with those measured by Babbin et al., (2014) and Ganesh et al., (2015), at sites close to ours in the ETNP, and are also generally consistent with anammox rates recorded in other open-ocean OMZs (Lam and Kuypers, 2011). Rates were also consistently highest at stations nearest to shore (6T, 7T), compared to off-shore sites (Figure 5.2; ~5 fold greater rates as coastal sites), consistent with analyses in the ETSP showing an overall increase in bulk N cycling rates driven by increased organic matter supply close to shore (Kalvelage et al. 2013).

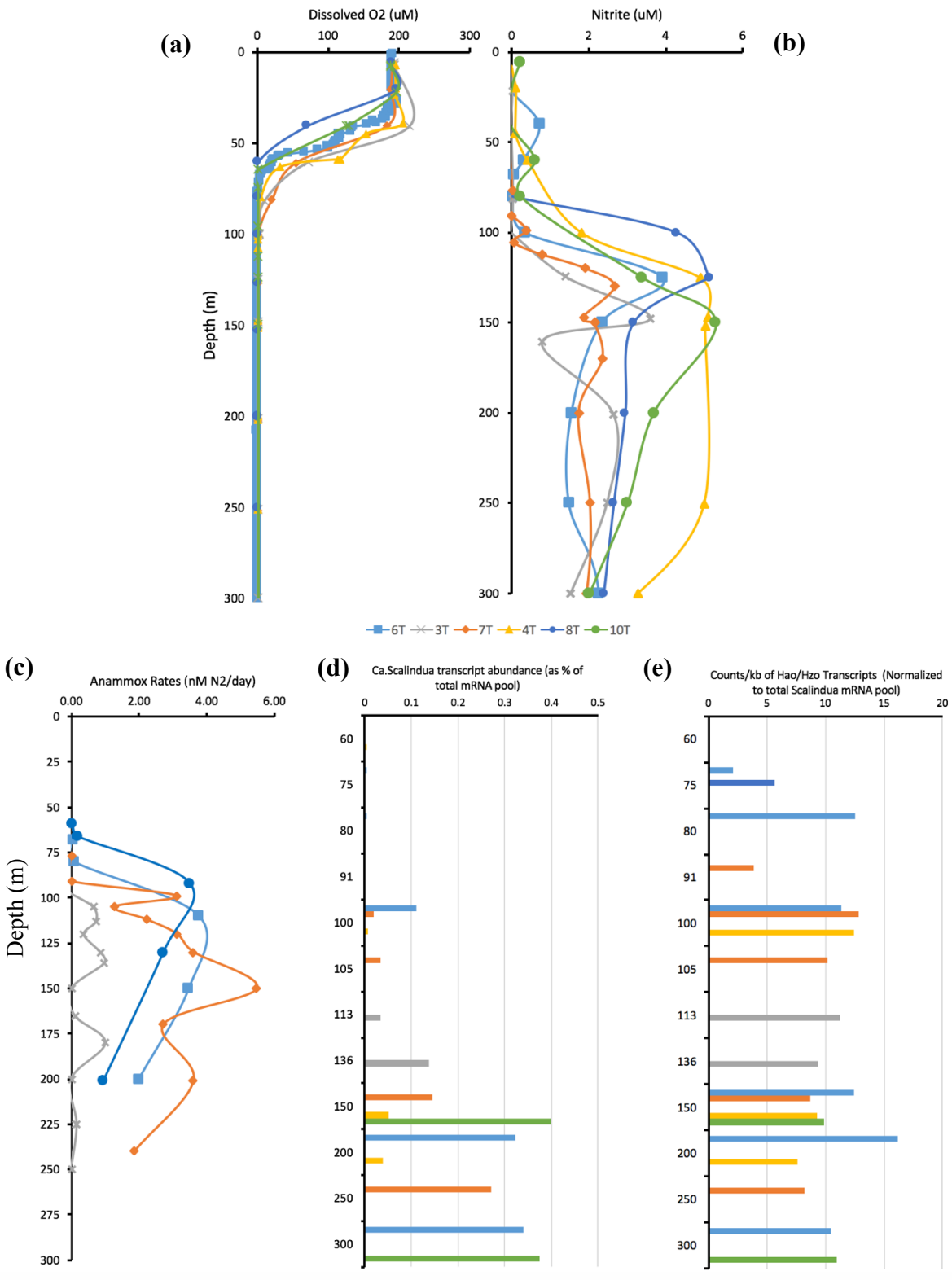

(a)
Dissolved O2 (uM)
0
100
200
300
Depth (m)
0
50
100
150
200
250
300
(b)
Nitrite (uM)
0
2
4
6
6T
3T
7T
4T
8T
10T
(c)
Anammox Rates (nM N2/day)
0.00
2.00
4.00
6.00
Depth (m)
0
25
50
75
100
125
150
175
200
225
250
275
300
(d)
Ca.Scalindua transcript abundance (as % of total mRNA pool)
0
0.1
0.2
0.3
0.4
0.5
60
75
80
91
100
105
113
136
150
200
250
300
(e)
Counts/kb of Hao/Hzo Transcripts (Normalized to total Scalindua mRNA pool)
0
5
10
15
20
60
75
80
91
100
105
113
136
150
200
250
300

Figure 5.2 Water column chemistry and *Ca*. Scalindua transcriptional abundance in the ETNP OMZ. (a) Dissolved oxygen profile based on measurements by SBE43 sensor. (b) Profiles of Nitrite concentration. (c) Anammox rates measured in nM N_2 per day. (d) Scalindua transcriptional abundance (as a % of total mRNA transcripts) obtained by mapping community mRNA transcriptomes to 8 different SAGs. (e) *Hao/Hzo* transcript relative abundance normalized to gene length and expressed as a proportion of total mRNA transcripts matching *Ca.* Scalindua. Colors represent the different sampling sites.

Genomic features of Ca. Scalindua SAGs

General features and phylogenetic placement. General sequence features of the assembled SAGs are summarized in Table 3. Genome completeness estimates, based on the detection of phylogenetic marker genes (as computed by CheckM; Parks et al., 2015) ranged from 0-50% for the 20 sequenced SAGs. Ribosomal RNA (rRNA) operons were identified in 8/20 SAGs. SAG 16S rRNA sequences were closely affiliated (>95% nucleotide identity) with the 16S rRNA of the anammox bacterium *Ca.* Scalindua *brodae* from enrichment cultures, as well as several uncultured environmental planctomycetes. Indeed, phylogenetic analysis of the SAG 16S sequences revealed clustering into two very closely related clades that are highly similar to environmental anammox planctomycetes from the Arabian Sea and the Red Sea (Woebken et al., 2008; Galan et al., 2009; Guan et al., 2015) (Figure 5.3). SAG sequences from these two clades showed a greater nucleotide similarity among themselves (99-100% nucleotide identity) than to the 16S rRNA from sequenced representatives in the database, suggesting limited species-level heterogeneity in the OMZ.

SAG genomes were compared against the most closely related reference Scalindua genome, that of *Ca.* S. brodae (Speth et al., 2015), to identify genes unique to OMZ representatives. The analysis revealed that each of the SAGs (n=8) share ~47% of their total protein coding genes (total protein coding genes in each SAG > 1000) with Scalindua *brodae*, with an average amino acid identity (AAI) of 75%, suggesting the SAGs are distantly related species of the same genus (Konstantinidis and Tiedje, 2007). Among the shared genes are the genes for the anammox metabolism, essential cellular machinery like protein and nucleotide metabolism and a large number of hypothetical

proteins. Unique genes present in the OMZ SAGs not present in Scalindua *brodae* include genes encoding ureases, urea transporters, some transposases and integrases.

Table 5.2 *Candidatus* Scalindua Single Amplified Genome assembly statistics and genome features

Single Amplified Genome	Sample Depth (m)	Total Reads used for Assembly	Genome Completeness Estimate (%)*	%GC	N50	Contigs >=500bp	tRNA#	rRNA+	Protein coding Genes†
B17	125	616,661	31.58	39.93	6087	288	23	-	1068
B21	125	813,282	23.1	39.71	4477	157	1	-	550
C14	125	388,439	12.28	39.23	10322	149	24	-	704
G15	125	526,599	28.55	39.89	6794	306	14	-	1140
K21	125	116,016	27.83	40.19	7602	288	29	-	1190
L20	125	390,204	26.56	38.92	8213	146	12	-	665
L7	125	335,299	15.79	39.98	7249	173	36	5S, 16S, 23S	664
L9	125	568,887	39.47	40.14	8220	300	26	23S	1282
M7	125	701,750	33.28	39.57	7952	203	34	-	907
B14	300	533,544	30.99	39.62	12944	161	27	16S, 23S	805
C8	300	756,070	19.65	40.47	3854	220	4	-	647
E14	300	362,296	28.23	39.78	7038	194	10	16S	748
F8	300	253,457	36.53	39.6	12506	225	11	-	1068
H20	300	616,767	43.09	39.97	11063	192	38	5S, 23S	1052
J11	300	385,564	20.51	39.69	4973	148	14	5S, 16S, 23S	524
L14	300	549,440	30.97	40.06	9634	217	19	16S	925
M13	300	291,425	4.17	39.58	6825	126	6	5S, 23S	561
N13	300	262,065	0	40.37	5465	62	11	16S	217
N19	300	203,823	46.19	39.95	12813	226	35	5S, 16S, 23S	1212
N22	300	393,120	49.97	40.04	6417	405	37	5S, 16S, 23S	1528

* Completeness estimates generated by CheckM based on identification of lineage-specific marker genes
# tRNAs identified by tRNA-Scan SE
+ rRNAs identified by RNAmmer
† Protein coding genes predicted by GeneMark.hmm
Shaded rows represent SAGs with > 1000 predicted genes, used for further analysis

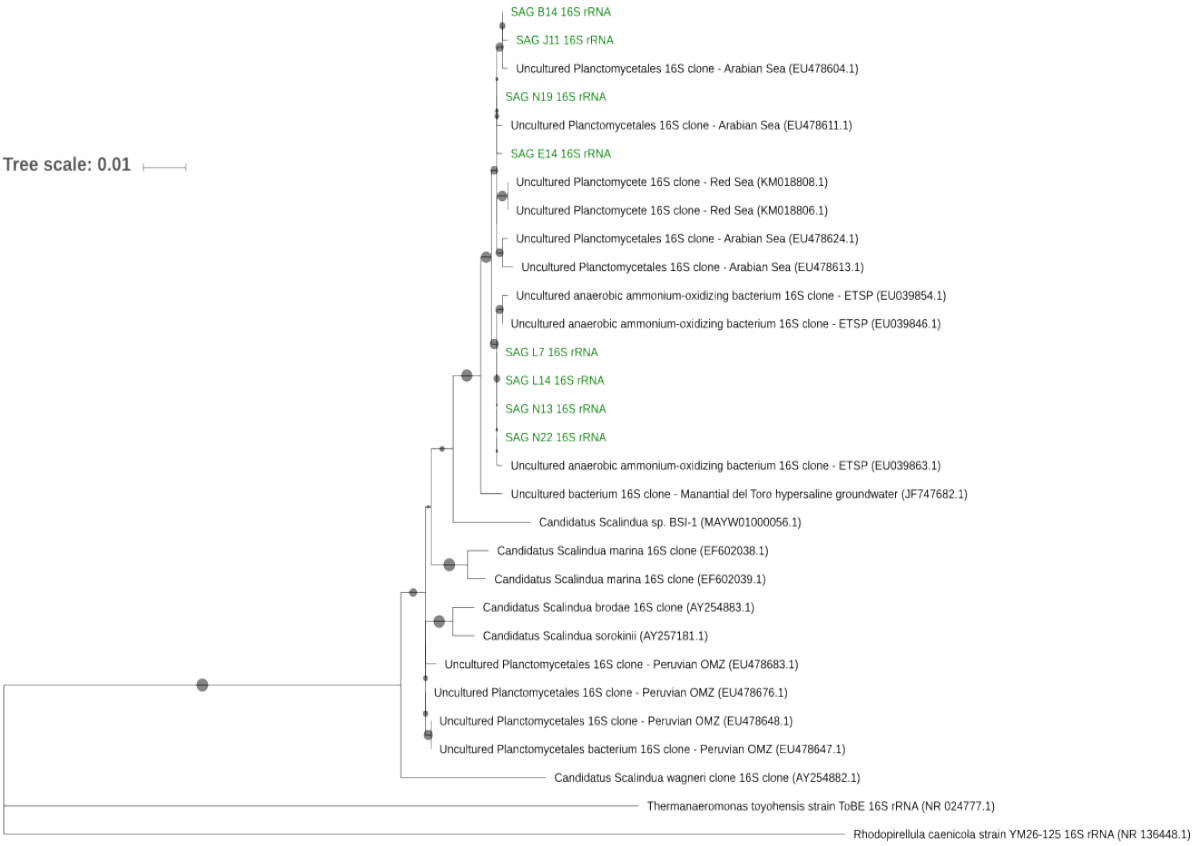


Figure 5.3 16S rRNA gene phylogeny. Full length 16S rRNA sequences identified in the *Ca.* Scalindua SAGs (n=8) are placed in two closely related clades of anaerobic ammonium oxidizing bacterium/planctomycete obtained from the Arabian Sea and the Red Sea (obtained in other studies, Woebken et al., 2008; Galan et al., 2009; Guan et al., 2015). Tree was constructed based on Maximum Likelihood analysis using the Kimura-2 parameter model with Gamma distribution (+G; parameter=0.05; 100 bootstraps) in MEGA7. Bootstrap values >50 are represented by filled circles (gray).

Nitrogen utilization genes

Anammox. Genes encoding enzymes of the anammox process were identified in all assembled SAGs. These included hydrazine synthase (*hzs*), which catalyzes the condensation of ammonium and nitric oxide into hydrazine, and hydrazine/hydroxylamine oxidoreductase (*hzo/hao)* which function to transform hydrazine to N_2 during anammox (Kuenen, 2008), which were (in four cases) co-localized on the same contig. The *hzsBC* genes are present in a single copy while the *hao* genes are present in up to 6 copies per SAG, which is consistent with prior reports of multiple copies of this gene in Scalindua genomes (van de Vossenberg et al., 2012; Speth et al., 2015). Multiple genes (1-4 copies depending on the SAG) encoding ammonium transporters (Amt), which facilitate acquisition of this substrate for the anammox process, were detected on OMZ SAGs, with multiple copies together on the same contig in most of the SAGs.

Growth of anammox bacteria is fueled by the reduction of nitrite and transfer of electrons to drive autotrophy, with nitrite reduction catalyzed by nitrate:nitrite oxidoreductases *(Nar/Nxr)*. Genes matching *Nar/Nxr* of anammox bacteria were identified in 6/8 SAGs. The *NirS* gene for the synthesis of nitric oxide from nitrite, was present in 4/8 SAGs.

Alternate nitrogen substrates

Cyanases. Putative cyanate metabolism genes (*CynS*: cyanate hydratase/lyase; *CynT*: carbonic anhydrase) were identified in 5 SAGs. Cyanate hydratase or cyanase is used to convert cyanate and bicarbonate to produce ammonia and carbon dioxide. The

ammonia produced may be used to fuel the anammox process. The carbonic anhydrase encoded by the CynT gene prevents depletion of bicarbonate during cyanate decomposition due to loss of CO_2 by diffusion out of the cell (Guilloton et al., 1992).

Specific transporters for cyanate were not identified in the SAGs, but it has been suggested that cyanate may be transported into the cell by transporters of nitrite *(*NirC*)* or formate *(FocA)* (Spang et al., 2012), genes for which were identified in the SAG genomes The genes identified to be putative cyanate hydratases exhibited a 71% amino acid similarity to a cyanate hydratase gene from *Nitrospina* (top BLASTP match) and 44% amino acid similarity to an experimentally verified cyanase from the ammonia oxidizing archaea *Nitrososphaera gargensis* (Palatinzsky et al., 2015), with 100% coverage across the 150 aa sequence, and possesses a conserved catalytic Cyanase C domain. A phylogenetic analysis of the SAG cyanase sequences, along with reference cyanate hydratase sequences from environmental and sequenced representatives, confirmed a close clustering of the SAG genes to those of *Nitrospina* suggesting a close relationship that may be attributed to horizontal gene transfer events (Figure 5.4).

Indeed, horizontal gene transfer between anammox bacteria and *Nitrospina*, both of which are abundant in the upper OMZ*,* has been suggested previously based on a large number of shared homologous genes (Lucker et al., 2013).

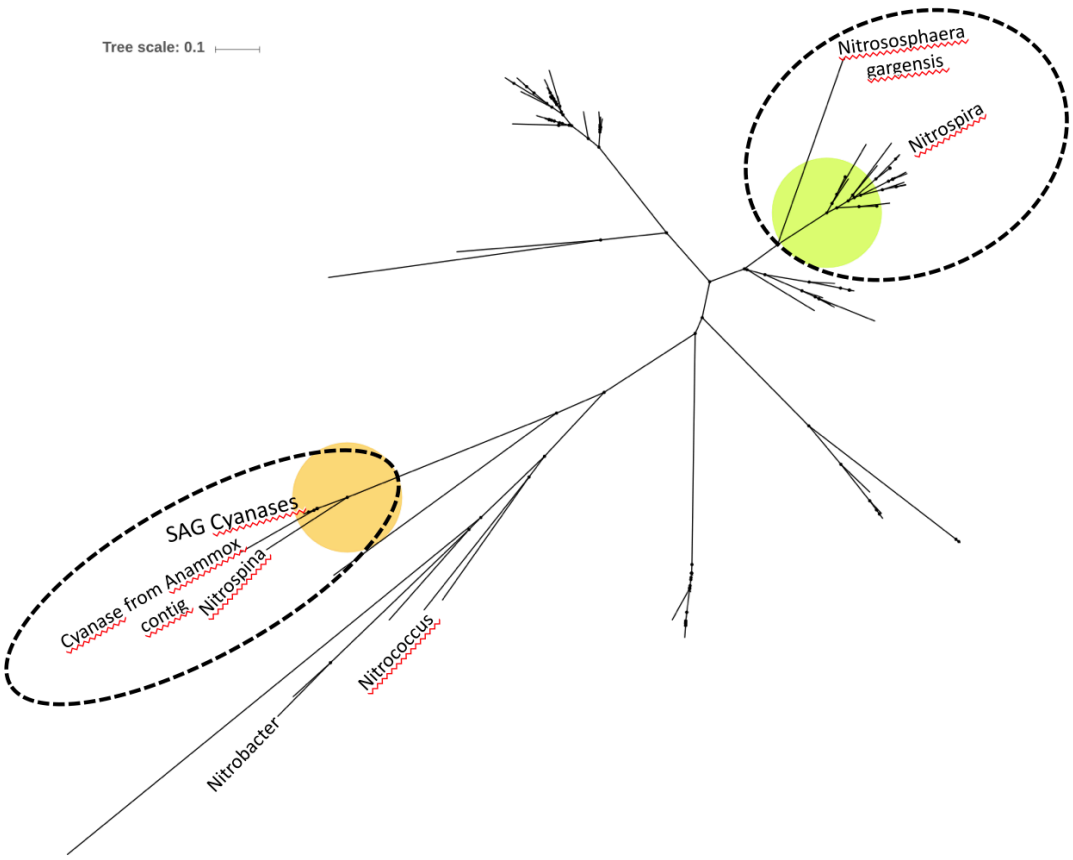


Figure 5.4 Phylogenetic reconstruction of Cyanate hydratase sequences from SAGs. Maximum Likelihood tree of SAG cyanate hydratase sequences along with reference cyanate hydratase sequences from nitrite oxidizing bacteria and archaea, and environmental metagenomes was constructed using RAxML implemented in the program Practical Alignment using SATé and TrAnsitivity, with an alignment generated using MAFFT. SAG cyanate hydratases cluster in a single clade with cyanase from Nitrospina and a putative cyanase from an anammox enrichment.

We searched publicly available anammox bacterial genomes and metagenomes against a curated custom database of cyanate hydratase protein sequences obtained from 100 sequenced representative species and environmental samples across a broad range of taxa, to identify the presence of the cyanase gene in other anammox bacteria (using BLASTP with a bitscore cutoff of 50 and e-value 0.0001). We were able to identify a putative cyanate hydratase gene in two metagenomic assemblies of *Ca*. Scalindua *profunda* from a marine sediment enrichment (van de Vossenberg et al., 2012; JGI taxon ID 2017108002). This gene had previously been annotated as a putative cyanase based on its limited homology to a cyanase from a plant (*Vitis vinifera*; van de Vossenberg et al., 2012). Our analysis shows its greater homology to cyanate hydratase from nitrite oxidizing bacteria and anammox bacteria, suggesting a closer evolutionary relationship. In addition, this metagenome also possesses a gene encoding a putative cyanate permease. This suggests a previously unidentified potential of anammox bacteria to degrade cyanate in the environment. Cyanate hydratase genes were not identified in protein coding sequences from the genomes of *Ca.* Scalindua brodae, *Ca*. Kuenenia stuttgartiensis, *Ca*. Jettenia, *Ca*. Brocadia *fulgida*, or other planctomycetes (Rhodopirrellula *baltica* and Blastopirellula *marina*).

These data raise the hypothesis that OMZ *Ca*. Scalindua sp. metabolize cyanate, potentially as a source of ammonium, when ammonium is limiting in the environment. Potential sources of cyanate in marine systems include *in situ* release of cyanate by the microbial community as a byproduct of cellular metabolism, (Sorokin et al., 2001), cyanate release through cell lysis or by grazers, spontaneous decomposition of the metabolic intermediate carbamoyl phosphate (Allen and Jones, 1964) and other organic

cellular metabolites, and urban agricultural runoff containing urea (Gilbert et al., 2006) followed by spontaneous aqueous decomposition to cyanate. Although data for cyanate distribution within the OMZ water column is currently unavailable, coupling cyanate utilization rates with anammox rates, along with measurements of cyanate concentrations in OMZs, would provide further support to the ability of these bacteria to metabolize cyanate as an energy or nitrogen source (ammonium).

Ureases. In addition to ammonium and cyanate, urea may also be an ammonia source for anammox bacteria in marine systems. Urease enzymes catalyze the hydrolysis of urea to ammonia and carbon dioxide, and are found ubiquitously in diverse species of plants, bacteria, and fungi.

Genes coding for the entire set of ureases and accessory proteins *(ureABCDEFG)* were identified in two of the Scalindua SAGs, in addition to several genes encoding urea transport proteins *(urtABCDE*). The entire set of *ure* genes was detected on a single contig with the genes encoding the transport proteins on a separate contig. The arrangement of the *ure* genes in gene cluster is similar to that observed in other bacteria, such as members of the *Enterobacteriaceae* (D'orazia and Collins, 1993). Urease genes may be constitutively expressed in some bacteria (Collins and Falcow, 1990), or expressed only during nitrogen-limiting growth, as observed in some *Klebsiella* species (Collins et al., 1993), or in the presence of urea in the surrounding environment (D'orazia and Collins, 1993). Autotrophic ammonia oxidizing archaea from acidic soils have also been shown to possess ureolytic activity (Lu and Jia, 2013), and ammonia oxidizing bacteria in an enrichment reactor were shown to utilize urea as an energy source for

nitrification, with anammox bacteria depending on them for ammonium (Sliekers et al., 2004). The ability of anammox bacteria to utilize urea over ammonium would depend on the relative availabilities of both substrates. The oxic-anoxic interface in the upper OMZ water column may be where urea is most available (Sambrotto, 2000), owing to the peak in particles and organic matter remineralization in this zone (Kalvalege et al., 2013), and potentially where anammox bacteria might be predicted to expressing genes for urea utilization. The presence of genes for the uptake and utilization of diverse nitrogen sources provides support for the environmental plasticity of anammox bacteria along OMZ gradients.

Carbon utilization genes. Anammox bacteria use the energy from their catabolism for autotrophic growth. The acetyl-coA pathway (Wood-Ljungdahl pathway) has been demonstrated to be active in anammox bacteria through the expression of two key enzymes in this pathway – carbon monoxide dehydrogenase and formate dehydrogenase (Strous et al., 2007). Marker genes for the acetyl-coA pathway of carbon fixation– carbon monoxide dehydrogenase and formate dehydrogenase were identified in all the SAGs, and were found to be transcriptionally active within the core of the OMZ (100 – 300m), indicative of active autotrophic carbon fixation by Scalindua in the OMZ core, as reported earlier (Ganesh et al., 2015). The enzyme formate dehydrogenase can also be used for the heterotrophic growth of these organisms on formate as a carbon source (Kartal et al., 2014), potentially lending metabolic flexibility if anammox substrates are limiting. Genes for intermediary metabolism of gluconeogenesis and the TCA cycle were present in the SAGs and transcriptionally active. A gene encoding putative citrate

synthase, which was reported to be missing in *Ca.* Kuenenia *stuttgartiensis* but detected in *Ca*. Scalindua *profunda* (Kartal et al., 2013; Van de Vossenberg et al., 2012), as well as those for 2-oxoglutarate ferredoxin oxidoreductase (marker genes for the reverse TCA cycle), are present in the SAGs.

Ca. Scalindua gene transcription

Protein-coding transcripts matching *Ca*. Scalindua SAGs (at 95% aa identity and bit score >50) formed a relatively small fraction of the total mRNA transcript pool in the ETNP OMZ, from 0.0015% at the oxycline and upper OMZ depths, then increasing with decreasing oxygen to 0.5% in the OMZ core (Fig 1). A similar vertical distribution of anammox bacterial genes and transcripts, with peaks in abundance at core OMZ depths below the nitrite maximum, was previously observed in the Chilean OMZ (Stewart et al. 2012) and in a prior study in the ETNP OMZ (Ganesh et al. 2015), and may be linked to both substrate and oxygen availability (Dalsgaard et al., 2014). For example, ammonium in the OMZ water column is dependent on the remineralization of organic matter from the surface (Galan et al., 2009) thereby leading to a vertical gradient of decreasing ammonium with depth. This suggests intensity of upwelling/water circulation and availability of ammonium as other important factors that determine the vertical partitioning of anammox rates at any time (Ulloa et al., 2012). Scalindua transcript abundance, however, did not parallel the measured anammox rates within the OMZ, which peaked at the top of the OMZ (100m) near the nitrite maximum and not in the core, which is in line with previous reports (Thamdrup et al., 2006; Lam et al., 2009; Galan et al., 2009; Lam et al., 2011; Daalsgard et al., 2012). This potential decoupling of

anammox rates and anammox bacterial abundances suggests that the total contribution of Scalindua to the community is not exclusively driven by the anammox process.

To examine site-specific variation in Scalindua gene transcription, normalized transcript abundances of 960 Scalindua-specific genes (matching the SAG N19), along with measurements of nitrite and anammox rates from each OMZ depth, were subjected to a Canonical Correspondence Analysis (CCA) (Figure 5.5). This analysis revealed gene expression profiles of Scalindua are strongly correlated to depth, nitrite availability and anammox rates across different sites. Functional gene expression at the nitrite maximum (between 125m-150m) is largely driven by nitrite availability, while deeper in the OMZ, depth plays a greater role in structuring gene expression. Gene expression also appeared to vary with proximity to shore, as evidenced by a partitioning of profiles between offshore sites vs those closer to shore where anammox rates were higher, suggesting the role of proximity to the coast and organic matter input in driving overall activity.

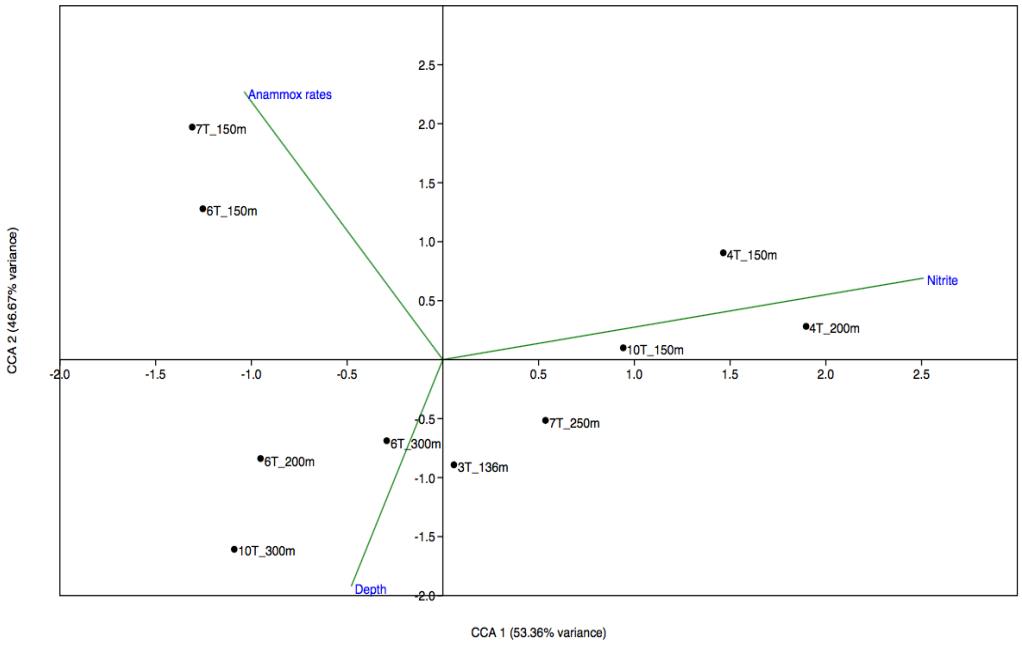


Figure 5.5 Canonical Correspondence Analysis (CCA) plot of Scalindua functional gene expression profile. Normalized transcript abundance of 960 protein-coding genes (matching SAG N19) across OMZ depths and sites, alongside environmental measurements of water column Nitrite, Anammox Rates and Depths was subjected to a CCA analysis. Clustering patterns reveal a site-specific gene expression within the OMZ driven by environmental gradients – gene expression profiles are similar at the nitrite max across off-shore sites, coastal sites with high anammox rates have similar gene expression profiles, and expression profiles are similar across sites at the core OMZ depths.

Anammox: The transcription of a key gene for the anammox process, hydrazine synthase (*hzsBC)* accounted for up to 16% of all Scalindua-like transcripts in the OMZ. The *hzs* transcription showed two peaks – at 80m (oxycline) and at 150m (nitrite max) suggesting an increase in anammox activity with the distribution of nitrite. Hydrazine synthase genes were also the most transcriptionally active Scalindua gene in the OMZ. Hydrazine/hydroxylamine oxidoreductase *(Hao)* genes peaked in transcriptional activity at 200m in the core of the OMZ, and its transcription increased ~1-3 fold from 75m to 200m. *Hao* transcripts accounted for up to 15% of the total Scalindua-like transcript pool within the ETNP OMZ. Of the multiple *hao* variants identified in the SAGs, only one variant was highly transcribed within the OMZ. It has been suggested that paralogs of *hao* may have alternative functions in Scalindua, including detoxification of potentially hazardous nitrogen compounds (Kartal et al., 2011), or even an ammonia-forming capacity as in the epsilon proteobacterium *Nautilia* (Campbell et al., 2009).

Nitrite oxidation/Nitrate reduction: Transcripts matching the *NarGH/Nxr* genes were present at a higher proportional abundance in the upper depths (80m), where these transcripts accounted for up to ~2% of the Scalindua-like transcript pool. Higher transcriptional abundance of nitrate reductases associated with anammox bacteria at upper OMZ depths has previously been observed in the ETSP OMZ, and their proportional abundance was reported to be negatively correlated with the abundance of *hao* transcripts (Stewart et al., 2012). We observe a similar trend here. *Nar/Nxr* in anammox bacteria traditionally function to oxidize nitrite to nitrate anaerobically while pumping electrons into transport systems to fuel autotrophy (Lucker et al., 2010). They

can also function as a traditional nitrate reductase when growing heterotrophically on organic matter with nitrate as electron acceptor (Kartal et al., 2007). NirS genes were transcribed in very low abundances in the metatranscriptomes. A similar low expression was observed in K. *stuttgartiensis* (Kartal et al., 2011). However, in the Ca. Scalindua *profunda* proteome, *NirS* was one of the most abundant proteins (Van de Vossenberg et al., 2012). This difference in expression of an important intermediate suggests variation in response to ecological conditions.

Ammonium Transport: Transcripts encoding genes for ammonium transporters *(Amt)* formed a significant proportion of the total Scalindua-like transcripts within the OMZ (~3% of the transcript pool). Ammonium transporter genes are transcriptionally active in the anoxic depths of the OMZ but are not transcribed at the upper oxycline (75, 80m). The transcription of genes for ammonium transporters (*amt*) peaked at the 200m depth. Of the multiple ammonium transporter genes present in the SAGs, one variant was transcribed at a higher proportional abundance. As anammox bacteria may be ammonium-limited in the OMZs, due to low environmental concentrations like other marine environments (Schmidt et al., 2004; Weidinger et al., 2007), the higher transcription of these high-affinity transporter genes is suggestive of their adaptation to nutrient limitation.

Organic Nitrogen Utilization: Cyanate hydratase genes were found to be transcriptionally active within the OMZ core. Transcripts encoding cyanate metabolism genes were detected only below the oxycline (depths below 100m) with a peak in

transcription at 200m. Preliminary studies in open ocean systems suggests that cyanate concentrations are highest just below the euphotic zone and decrease with depth (Widner et al., 2013). The close proximity of anammox bacteria to the nitrite maximum coupled to a limitation in ammonium at these depths and potential availability of cyanate may make these intermediate depths, important sites for active cyanate utilization. Transcripts for cyanate transporters were not detected, however like discussed previously, transporters of nitrite (NirK) and formate (FocA) which were transcriptionally active within the OMZ, may function in transporting cyanate from the environment. Additionally, the detection of a putative cyanate permease in an anammox bacteria contig from an enrichment metagenome (see Genomic features of Scalindua SAGs above) suggests that Scalindua may possess this gene, but due to the incomplete nature of the SAGs, was likely not detected in our analyses.

Relative transcript abundance of urease genes increases with depth, with a transcription peak at 150m. Transcripts for urea transporters *(urt)* are also enriched at 100m and 150m within the OMZ. It is interesting to note that transcripts for the gene *urtA*, which encodes a urea transport system substrate binding protein, is on average 36 fold higher than transcripts for other urea transport proteins, and the relative transcript abundance for *urtA* increases with depth. The *urt* transporters for urea are high affinity transporters, expressed only during nitrogen limitation, and may be an adaptation to low urea concentrations (Koch et al., 2015). Additional information on urea availability and distribution in OMZ water columns would be very valuable in deciphering the potential of OMZ anammox bacteria to utilize urea to drive the anammox process.

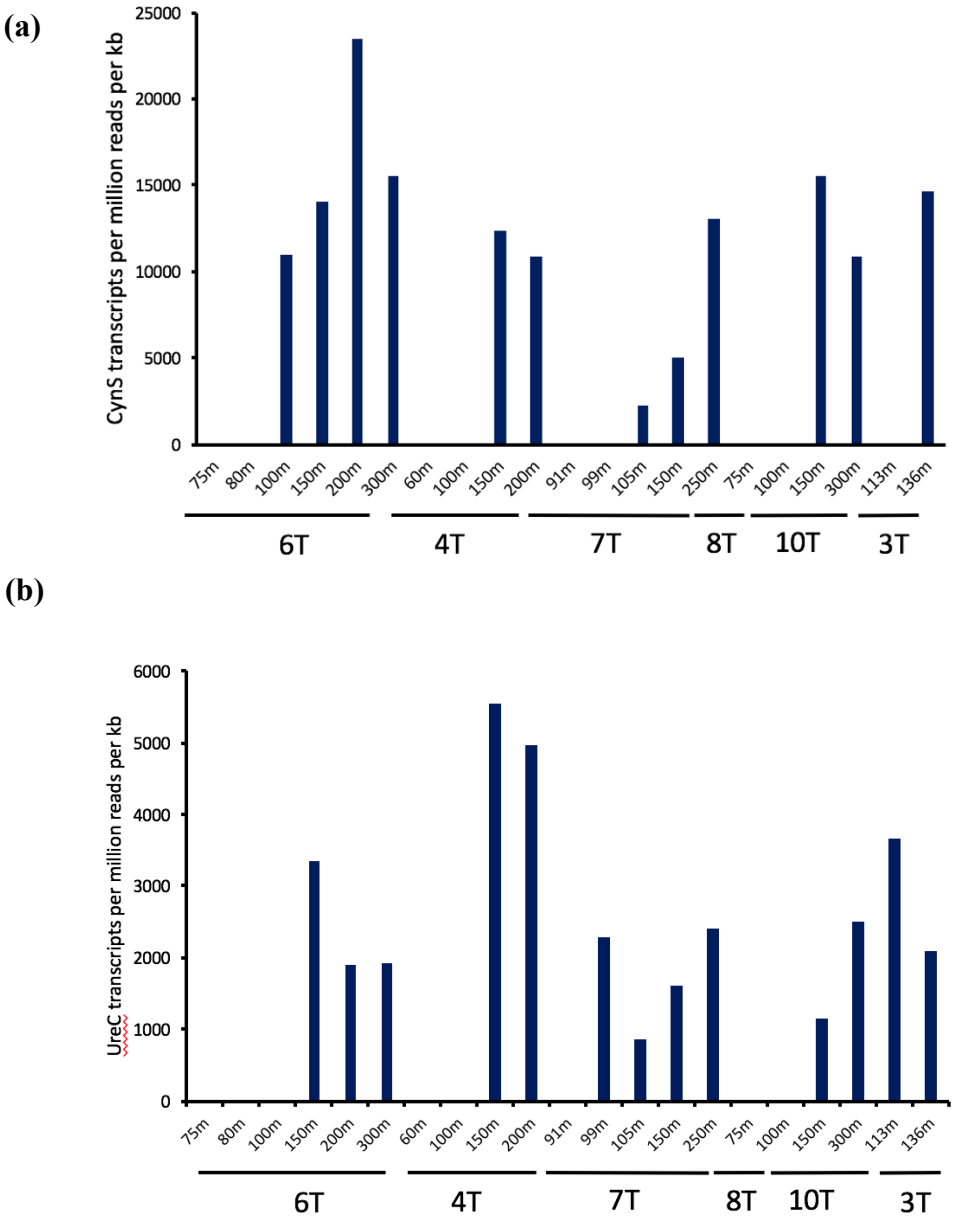


Figure 5.6. Relative abundance of transcripts matching the (a) Cyanate hydratase (CynS) gene and (b) Urease (UreC) gene from the Scalindua SAGs in the ETNP OMZ sampling sites.

5.5 Conclusions

Ca. Scalindua in OMZs are metabolically diverse and are distinct from representative anammox bacteria previously obtained from enrichment cultures. OMZ Scalindua SAGs revealed low level of genetic variation among each other based on 16S rRNA gene sequence identity suggesting only strain level heterogeneities in the population. This relative lack of diversity within Scalindua has been reported previously (Villanueva et al., 2014), and may be attributed to the evolutionary pressures associated with developing the mechanism to carry out the highly specialized process of anammox. Analysis of functional gene expression patterns of Scalindua within the OMZ, however, suggest that they exhibit metabolic variation within the OMZ by modifying gene expression at various ecological niches. This variation is driven by environmental factors such as nitrite availability, oxygen, and anammox rates within the OMZ depth gradient. However, testing for differentially expressed Scalindua genes between transcriptomes from the nitrite maximum (125m-150m) versus the OMZ core (200m-300m) did not reveal any statistically significant expression of functional genes ($p<0.05$). Due to the lack of biological replicates from each of the sites, we combined transcriptomes from nitrite maximum across different sites and OMZ depths across different sites (coastal and off-shore). Thus, individual site-specific trends in differential expression of functional genes could be masked, leading to conservative estimates of differential gene expression between these two ecological niches in the OMZ.

Scalindua vary in transcriptional abundance distinctly along the depth of the water column, and exhibit a maximum at the core of the OMZ (200m). The measured anammox process rates however do not follow this pattern, with rates being highest at the top of the

OMZ (100m). This uncoupling of anammox rates and anammox bacteria abundance suggests that the contribution of Scalindua to the total microbial community is not exclusively driven by the anammox process.

In addition to the anammox process, OMZ Scalindua are potentially capable of participating in other processes including dissimilatory nitrate reduction for heterotrophy (Kartal et al., 2006), as well as conversion of cyanate and urea to ammonia, for downstream use in the anammox process. Their contribution to other metabolic pathways remains to be understood, and may provide further support to their metabolic versatility. Measurements of anammox rates in communities amended with cyanate will help clarify the contribution of cyanate-driven anammox to total anammox rates, and provide a better understanding of the role of cyanate within OMZs.

Due to the fragmentary nature of the SAGs, a complete genome reconstruction has not been possible. But based on the fragmentary genomic information available from these SAGs, we see limited protein-coding genes being shared with other closely related Scalindua genomes. Sequencing at greater depth across these SAGs coupled with transcriptome mapping would in exhaustively determining the transcriptional patterns, which would likely provide greater support to their genomic plasticity within the unique OMZ environment that is characterized by steep chemical gradients.

Disclaimer: This chapter is currently being prepared as a manuscript for submission to a journal for publication.

Citation: **Ganesh S, Bristow LA, Padilla CC, Thamdrup B, Stewart FJ**. *(In prep)*. Metabolic plasticity of a pelagic anammox bacterium over redox gradients in a marine oxygen minimum zone.

5.6 References

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D. and Pyshkin, A.V. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol, 19:455-477.

Codispoti, L.A., Brandes, J.A., Christensen, J.P., Devol, A.H., Naqvi, S.W.A., Paerl, H.W. and Yoshinari, T. (2001). The oceanic fixed nitrogen and nitrous oxide budgets: Moving targets as we enter the anthropocene? Scientia Marina, 65:85-105.

Collins, C.M. and D'Orazio, S.E., 1993. Bacterial ureases: structure, regulation of expression and role in pathogenesis. Mol Microbiol, 9:.907-913.

Collins, C.M. and Falkow, S., 1990. Genetic analysis of Escherichia coli urease genes: evidence for two distinct loci. J Bacteriol, 172:7138-7144.

Collins, C.M., Gutman, D.M. and Laman, H. (1993). Identification of a nitrogen-regulated promoter controlling expression of Klebsiella pneumoniae urease genes. Mol Microbiol, 8:187-198.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M. and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics, 21:3674-3676.

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.

Dalsgaard, T., Thamdrup, B., Farías, L. and Revsbech, N.P. (2012). Anammox and denitrification in the oxygen minimum zone of the eastern South Pacific.Limnol and Oceanogr, 57:1331-1346.

De Brabandere, L., D. E. Canfield, T. Dalsgaard, G. E. Friederich, N. P. Revsbech, O. Ulloa, and B. Thamdrup. (2014). Vertical partitioning of nitrogen-loss processes across the oxic-anoxic interface of an oceanic oxygen minimum zone. Environ. Microbiol. 16:3041-3054

de Hoon, M.J., Imoto, S., Nolan, J. and Miyano, S. (2004). Open source clustering software. Bioinformatics, 20:1453-1454.

Galán, A., Molina, V., Belmar, L. and Ulloa, O. (2012). Temporal variability and phylogenetic characterization of planktonic anammox bacteria in the coastal upwelling ecosystem off central Chile. Prog. Oceanogr., 92:110-120.

Ganesh, S., Bristow, L.A., Larsen, M., Sarode, N., Thamdrup, B. and Stewart, F.J. (2015). Size-fraction partitioning of community gene transcription and nitrogen metabolism in a marine oxygen minimum zone. ISME J, 9:2682-2696.

Ganesh, S., Parris, D.J., DeLong, E.F. and Stewart, F.J. (2014). Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. ISME J, 8:187-211.

Garcia-Robledo, E., Padilla C.C., Aldunate, M., Stewart, F.J., Ulloa, O, Paulmier, A., Gregori, G., Revsbech, N.P. Cryptic Oxygen Cycling in Anoxic Oceanic Zones. (*In Preparation)*

Gori, F., Tringe, S.G., Kartal, B., Machiori, E. and Jetten, M.S. (2011). The metagenomic basis of anammox metabolism in Candidatus 'Brocadia fulgida'. Biochemical Society Transactions, 39:1799-1804.

Gruber, N. (2004). The dynamics of the marine nitrogen cycle and its influence on atmospheric CO2 variations. In *The ocean carbon cycle and climate* (pp. 97-148). Springer Netherlands.

Guilloton, M.B., Lamblin, A.F., Kozliak, E.I., Gerami-Nejad, M., Tu, C.H.I.N.G.K.U.A.N.G., Silverman, D., Anderson, P.M. and Fuchs, J.A. (1993). A physiological role for cyanate-induced carbonic anhydrase in Escherichia coli. J Bacteriol, 175:1443-1451.

Hammer, Ø., Harper, D.A.T. and Ryan, P.D. (2005). PAST-PAlaeontological Statistics, ver. 1.35. *Users manual*.

Hira, D., Toh, H., Migita, C.T., Okubo, H., Nishiyama, T., Hattori, M., Furukawa, K. and Fujii, T. (2012). Anammox organism KSU-1 expresses a NirK-type copper-containing nitrite reductase instead of a NirS-type with cytochrome cd 1. FEBS letters, 586:1658-1663.

Hu, Z., Speth, D.R., Francoijs, K.J., Quan, Z.X. and Jetten, M. (2012). Metagenome analysis of a complex community reveals the metabolic blueprint of anammox bacterium "Candidatus Jettenia asiatica". Front Microbiol, 3:366.

Kalvelage, T., Lavik, G., Lam, P., Contreras, S., Arteaga, L., Löscher, C.R., Oschlies, A., Paulmier, A., Stramma, L. and Kuypers, M.M. (2013). Nitrogen cycling driven by organic matter export in the South Pacific oxygen minimum zone. Nature geoscience, 6:228-234.

Kamennaya, N.A. and Post, A.F. (2011). Characterization of cyanate metabolism in marine Synechococcus and Prochlorococcus spp. Appl Environ Microbiol, 77:291-301.

Kamennaya, N.A., Chernihovsky, M. and Post, A.F. (2008). The cyanate utilization capacity of marine unicellular Cyanobacteria. Limnol Oceanogr, 53:2485.

Kamykowski, D. and Zentara, S.J. (1991). Spatio-temporal and process-oriented views of nitrite in the world ocean as recorded in the historical data set. Deep Sea Research Part A. Oceanographic Research Papers, 38:445-464.

Kanehisa, M., Sato, Y. and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. Journal of molecular biology, 428:726-731.

Kartal, B., de Almeida, N.M., Maalcke, W.J., den Camp, H.J.O., Jetten, M.S. and Keltjens, J.T.(2013). How to make a living from anaerobic ammonium oxidation. FEMS microbiology reviews, 37:428-461.

Kartal, B., Kuypers, M.M., Lavik, G., Schalk, J., Op den Camp, H.J., Jetten, M.S. and Strous, M. (2007). Anammox bacteria disguised as denitrifiers: nitrate reduction to dinitrogen gas via nitrite and ammonium. Environ Microbiol, 9:635-642.

Kartal, B., Van Niftrik, L., Rattray, J., Van De Vossenberg, J.L., Schmid, M.C., Damsté, J.S., Jetten, M.S. and Strous, M. (2008). Candidatus 'Brocadia fulgida': an autofluorescent anaerobic ammonium oxidizing bacterium. FEMS microbiology ecology, 63:46-55.

Konstantinidis, K.T. and Tiedje, J.M. (2005). Towards a genome-based taxonomy for prokaryotes. J Bacteriol, 187:6258-6264.

Konstantinidis, K.T., and Tiedje, J.M. (2007). Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. Curr Opin Microbiol 10: 504–509.

Lagesen, K., Hallin, P., Rødland, E.A., Stærfeldt, H.H., Rognes, T. and Ussery, D.W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res, 35:3100-3108.

Lam, P. and Kuypers, M.M. (2011). Microbial nitrogen cycling processes in oxygen minimum zones. Ann Rev of Mar Sci, 3:317-345.

Lam, P., Lavik, G., Jensen, M.M., van de Vossenberg, J., Schmid, M., Woebken, D., Gutiérrez, D., Amann, R., Jetten, M.S. and Kuypers, M.M. (2009). Revising the nitrogen cycle in the Peruvian oxygen minimum zone. Proc Natl Acad Sci USA, 106:4752-4757.

Lu, L. and Jia, Z. (2013). Urease gene-containing Archaea dominate autotrophic ammonia oxidation in two acid soils. Environ Microbiol,15:1795-1809.

Lücker, S., Nowka, B., Rattei, T., Spieck, E. and Daims, H. (2013). The genome of Nitrospina gracilis illuminates the metabolism and evolution of the major marine nitrite oxidizer. Front Microbiol, 4:27.

Lukashin, A.V. and Borodovsky, M. (1998). GeneMark. hmm: new solutions for gene finding. Nucleic Acids Res, 26:1107-1115.

Luque-Almagro, V.M., Huertas, M.J., Sáez, L.P., Luque-Romero, M.M., Moreno-Vivián, C., Castillo, F., Roldán, M.D. and Blasco, R. (2008). Characterization of the Pseudomonas pseudoalcaligenes CECT5344 cyanase, an enzyme that is not essential for cyanide assimilation. Appl Environ Microbiol, 74:6280-6288.

Magoč, T. and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics, 27:2957-2963.

Mirarab, S., Nguyen, N., Guo, S., Wang, L. S., Kim, J., & Warnow, T. (2015). PASTA: ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. J Comput Biol, 22:377-386.

Padilla, C.C., Bristow, L.A., Sarode, N., Garcia-Robledo, E., Ramírez, E.G., Benson, C.R., Bourbonnais, A., Altabet, M.A., Girguis, P.R., Thamdrup, B., Stewart, F.J. (2016). NC10 bacteria in marine oxygen minimum zones. ISME J, 10:2067-2071.

Palatinszky, M., Herbold, C., Jehmlich, N., Pogoda, M., Han, P., von Bergen, M., Lagkouvardos, I., Karst, S.M., Galushko, A., Koch, H. and Berry, D. (2015). Cyanate as an energy source for nitrifiers. Nature, 524:105-108.

Palenik, B., Brahamsha, B., Larimer, F.W., Land, M., Hauser, L., Chain, P., Lamerdin, J., Regala, W., Allen, E.E., McCarren, J. and Paulsen, I. (2003). The genome of a motile marine Synechococcus. Nature, 424:1037-1042.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P. and Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome research, 25:1043-1055.

Quan, Z.X., Rhee, S.K., Zuo, J.E., Yang, Y., Bae, J.W., Park, J.R., Lee, S.T. and Park, Y.H. (2008). Diversity of ammonium-oxidizing bacteria in a granular sludge anaerobic ammonium-oxidizing (anammox) reactor. Environ Microbiol, 10:3130-3139.

Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N.A., Arellano, A., Coleman, M., Hauser, L., Hess, W.R. and Johnson, Z.I. (2003). Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation. Nature, 424:1042-1047.

Rodriguez-R, L.M. and Konstantinidis, K.T. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes (No. e1900v1). PeerJ Preprints.

Sambrotto, R.N. (2001). Nitrogen production in the northern Arabian Sea during the Spring Intermonsoon and Southwest Monsoon seasons. Deep Sea Research Part II: Topical Studies in Oceanography, 48:1173-1198.

Schmid, M., Walsh, K., Webb, R., Rijpstra, W.I., van de Pas-Schoonen, K., Verbruggen, M.J., Hill, T., Moffett, B., Fuerst, J., Schouten, S. and Damsté, J.S.S. (2003). Candidatus “Scalindua brodae”, sp. nov., Candidatus “Scalindua wagneri”, sp. nov., two new species of anaerobic ammonium oxidizing bacteria. Syst Appl Microbiol, 26:529-538.

Schmid, M.C., Maas, B., Dapena, A., van de Pas-Schoonen, K., van de Vossenberg, J., Kartal, B., Van Niftrik, L., Schmidt, I., Cirpus, I., Kuenen, J.G. and Wagner, M. (2005). Biomarkers for in situ detection of anaerobic ammonium-oxidizing (anammox) bacteria. Appl Environ Microbiol, 71:1677-1684.

Sievers, F. and Higgins, D.G. (2014). Clustal omega. Current protocols in bioinformatics, 3-13.

Sliekers, A.O., Haaijer, S., Schmid, M. and Harhangi, H. (2004). Nitrification and anammox with urea as the energy source. Sys Appl Microbiol, 27:271.

Spang, A., Poehlein, A., Offre, P., Zumbrägel, S., Haider, S., Rychlik, N., Böhm, C. (2012). The genome of the ammonia-oxidizing Candidatus Nitrososphaera gargensis: insights into metabolic versatility and environmental adaptations. Environ Microbiol, 14:3122-3145.

Speth, D.R., Hu, B., Bosch, N., Keltjens, J., Stunnenberg, H. and Jetten, M. (2012). Comparative genomics of two independently enriched "Candidatus Kuenenia stuttgartiensis" anammox bacteria. Front Microbiol, 3:307.

Speth, D.R., Russ, L., Kartal, B., Op den Camp, H.J., Dutilh, B.E. and Jetten, M.S. (2015). Draft genome sequence of anammox bacterium "Candidatus Scalindua brodae," obtained using differential coverage binning of sequencing data from two reactor enrichments. Genome announcements, 3: e01415-14.

Strous, M., Pelletier, E., Mangenot, S., Rattei, T., Lehner, A., Taylor, M.W., Horn, M., Daims, H., Bartol-Mavel, D., Wincker, P. and Barbe, V. (2006). Deciphering the evolution and metabolism of an anammox bacterium from a community genome. Nature, 440:790-794.

Kumar, S., Stecher, G., and Tamura, K. (2015) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0. Molecular Biology and Evolution (submitted)

Thamdrup B, Dalsgaard T. (2002). Production of N_2 through anaerobic ammonium oxidation coupled to nitrate reduction in marine sediments. Appl Environ Microbiol. 68:1312-1318.

Thamdrup, B., Dalsgaard, T., Jensen, M.M., Ulloa, O., Farías, L. and Escribano, R. (2006). Anaerobic ammonium oxidation in the oxygen-deficient waters off northern Chile. Limnol Oceanogr, 51(5):2145-2156.

Ulloa, O., Canfield, D.E., DeLong, E.F., Letelier, R.M. and Stewart, F.J. (2012). Microbial oceanography of anoxic oxygen minimum zones. Proc Natl Acad Sci USA, 109:15996-16003.

van de Vossenberg, J., Woebken, D., Maalcke, W.J., Wessels, H.J., Dutilh, B.E., Kartal, B., Janssen-Megens, E.M., Roeselers, G., Yan, J., Speth, D. and Gloerich, J. (2013). 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.

Villanueva, L., Speth, D. R., Van Alen, T., Hoischen, A., & Jetten, M. S. (2014). Shotgun metagenomic data reveals significant abundance but low diversity of "Candidatus Scalindua" marine anammox bacteria in the Arabian Sea oxygen minimum zone. Front Microbiol, 5:2014.

Widner, B., Mulholland, M.R. and Mopper, K. (2013). Chromatographic determination of nanomolar cyanate concentrations in estuarine and sea waters by precolumn fluorescence derivatization. Analytical chemistry,85:6661-6666.

Woebken, D., Fuchs, B.M., Kuypers, M.M. and Amann, R. (2007). Potential interactions of particle-associated anammox bacteria with bacterial and archaeal partners in the Namibian upwelling system. Appl Environ Microbiol, 73:4648-4657.

Woebken, D., Lam, P., Kuypers, M. M. M., Naqvi, S. W. A., Kartal, B., Strous, M., Jetten, M. S. M., Fuchs, B. M. and Amann, R. (2008). A microdiversity study of anammox bacteria reveals a novel *Candidatus* Scalindua phylotype in marine oxygen minimum zones. Environ Microbiol, 10: 3106–3119.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

The thesis couples molecular genomic and biochemical approaches to investigate how environmental variation over multiple spatial scales shapes microbial community diversity, genomic potential, and biochemical cycling in marine oxygen minimum zones. Specifically, we utilized metagenomic and metatranscriptomic approaches to characterize microbial communities attached to particles and those that occur free-living in the OMZ water column, and their function and contribution to OMZ metabolic processes. Measurements of biochemical process rates were coupled to infer the role of particles in mediating these processes. Single cell genomics was further used as a method to characterize an important functional clade and identify potential genotypic variants occupying distinct ecological niches within the OMZ. Additionally, we highlight how larger scale nutrient and redox gradients in marine systems can affect gene expression in a bacterial lineage playing a critical role in ocean nitrogen cycling.

Metagenomic comparisons of different bacterioplankton size fractions are surprisingly rare but are necessary for clarifying the distribution of functional traits in marine microbes and assessing the role of surface attachment in driving microbial ecology in the oceans. We provided the first description of microbial community genomic partitioning between two size fractions in the Eastern Tropical South Pacific OMZ (Chapter 2). We described the enrichment in the larger (particle-associated) fraction of functional genes mediating microbial interactions and exchange of genetic material, suggesting adaptations to surface attachment and growth in a biofilm. Spatial

clustering of microbes on particles facilitates DNA transfer, potentially due to increased physical contact between microbes or selective ecological pressures, and this may help explain differences in genome content and sizes between microbes in different microhabitats. We were able to highlight that particle association was a stronger predictor of community structure than depth-specific environmental variation. We also described how metabolic pathways mediating OMZ elemental cycling were differentially partitioned between size fractions, implying that particle abundance, and by extension the percentage of the total microbial community attached to surfaces, may have important impacts on community biogeochemical transformations. Inferences of the drivers of functional and taxonomic diversity in OMZs requires an understanding of the distribution of particles and particle-associated communities along with predictions of how vertical gradients in oxygen and nutrients affect OMZ communities.

To further explore how particle association drives microbial community function, we combined community transcription profiling and rate measurements to quantify size fraction partitioning of key steps of the marine nitrogen cycle in the Eastern Tropical North Pacific OMZ (Chapter 3). We showed that filter fractionation significantly altered nitrogen metabolism rates, confirming a critical role for particles and particle-associated microorganisms in OMZ nitrogen cycling. We also highlighted differences in how diverse OMZ microorganisms may interact with particles – from direct adherence to particles by some denitrifiers, to an indirect dependence on particles for substrates by free-living anammox bacteria. Thus, particle removal significantly disrupted bulk nitrogen cycle processes mediated by a free-living majority, possibly by severing important chemical and biological linkages between free-living and particle-associated

niches. This work showcases our ability to identify the magnitude and spatial scales of biogeochemical processes in the OMZ by coupling metatranscriptomics of size-fractionated biomass and metabolic process rate measurements.

Experimental approaches may be incorporated into future research focused on identifying niche-specific gene expression of microorganisms. Allowing free-living microorganisms to colonize artificial particles would allow identification of changes in gene expression after surface colonization of particles, and help quantify patterns in gene expression change with changing growth strategies. It also remains to be understood how the chemical environment of a marine particle changes with microbial colonization on particles. This would require direct sampling of individual particles and measurements of chemical gradients within these particles using microsensors or a technique like NanoSIMS. Coupling this with tracking changes in microbial abundances on particles using taxon-specific probes would help decipher the chemical basis of these complex interactions that help drive important metabolic processes.

Filter fractionation is a common method used for size fractionating biomass. Sampling size fractionated microbial biomass requires several standardizations of sampling design and methodology. Because marine particles can vary largely in size and composition, a sampling strategy that involves a uniform comparison across multiple size fractions should be used to make inferences about particle-associated microorganisms, while keeping in mind potential biases associated with sampling particles like particle rupturing and unknown biases due to sample volume. We described potential biases in microbial community composition estimates associated with using filter fractionation to size-fractionate biomass along with variation in sample volume (Chapter 4). Our results

suggest that biases due to variation in filtered volume may confound comparisons of size-fractionated community structure across habitats and studies, and that particles may be better collected by direct sampling without filtration to infer community composition. Filter fractionation of biomass should be used cautiously and the effect of variation in sample volume in measurements of community structure should be taken into consideration when estimating community richness. Estimates of functional gene partitioning between size fractions still provide useful albeit conservative estimates of functional differences, as truly particle-associated microbes (which are detected by direct sampling of particles) are still retained in the larger size fraction. However, conclusions about exclusive niche partitioning of specific taxonomic groups and functions may not be made with confidence and would require direct sampling of particles to make such inferences. Direct sampling of particles also poses several problems as common particle capture methods only capture sinking particles, while direct sampling of marine particles is challenging owing to their fragile nature and tendency to disaggregate.

Finally, progressing from microscale variation to larger meso-scale gradients we described the genomic composition and gene expression variation of an important OMZ bacterium, *Candidatus* Scalindua sp., responsible for anaerobic ammonia oxidation (anammox), the second major nitrogen loss pathway in OMZs (Chapter 5). *Ca.* Scalindua is the only known marine planctomycete clade and its complete genomic potential remains unknown. We showed that *Ca.* Scalindua in OMZs exhibit ecological niche specific gene expression which is largely driven by the availability of nutrients. We also described its genomic potential to utilize diverse nitrogen substrates for driving the anammox process. This work advances our understanding of the metabolic versatility of

an important OMZ functional clade and environmental factors that determine its activity within the OMZ. These are also the first genomic sequences of this clade from an OMZ that will serve as important sources of genetic information for future studies of OMZ anammox bacteria, to potentially determine special adaptive mechanisms that allow *Ca.* Scalindua to occupy diverse ecological niches. In order to more completely evaluate the unique adaptations in OMZ Scalindua single amplified genomes may sequenced in greater depth to improve coverage of existing draft assemblies along with the use of metagenomes to close gaps. Additionally, this data can guide more intensive physiological characterizations of ecologically relevant functional genes.

One of the potentially interesting results was the identification of genes for cyanate metabolism in OMZ anammox bacteria. A potentially toxic compound, cyanate can serve as a source of ammonium for microbes possessing the molecular machinery to breakdown this compound. Anammox bacteria may be capable of utilizing cyanate to fuel the anammox process, the total contribution of which remains uncharacterized. In addition to functionally verifying the activity of the cyanase gene from *Ca*. Scalindua, quantifying the contribution of cyanate-driven anammox to bulk anammox process rates in OMZs should be a research priority to evaluate its significance and contribution to this global nitrogen loss pathway.

Microbial communities within OMZs play central roles in ocean and global biogeochemical cycles, yet we still lack a holistic understanding of how microbial biodiversity is distributed across OMZs. There may be more diverse assemblages of indigenous microbial species that remain ecologically and genetically uncharacterized. The potential impacts of OMZs on marine ecosystem structure and global

biogeochemical cycling highlights the need to accelerate the rapid pace of exploration and discovery of OMZ microbial diversity.

APPENDIX A

SUPPLEMENTARY MATERIAL FOR CHAPTER 2

METAGENOMIC ANALYSIS OF SIZE-FRACTIONATED PICOPLANKTON IN A MARINE OXYGEN MINIMUM ZONE

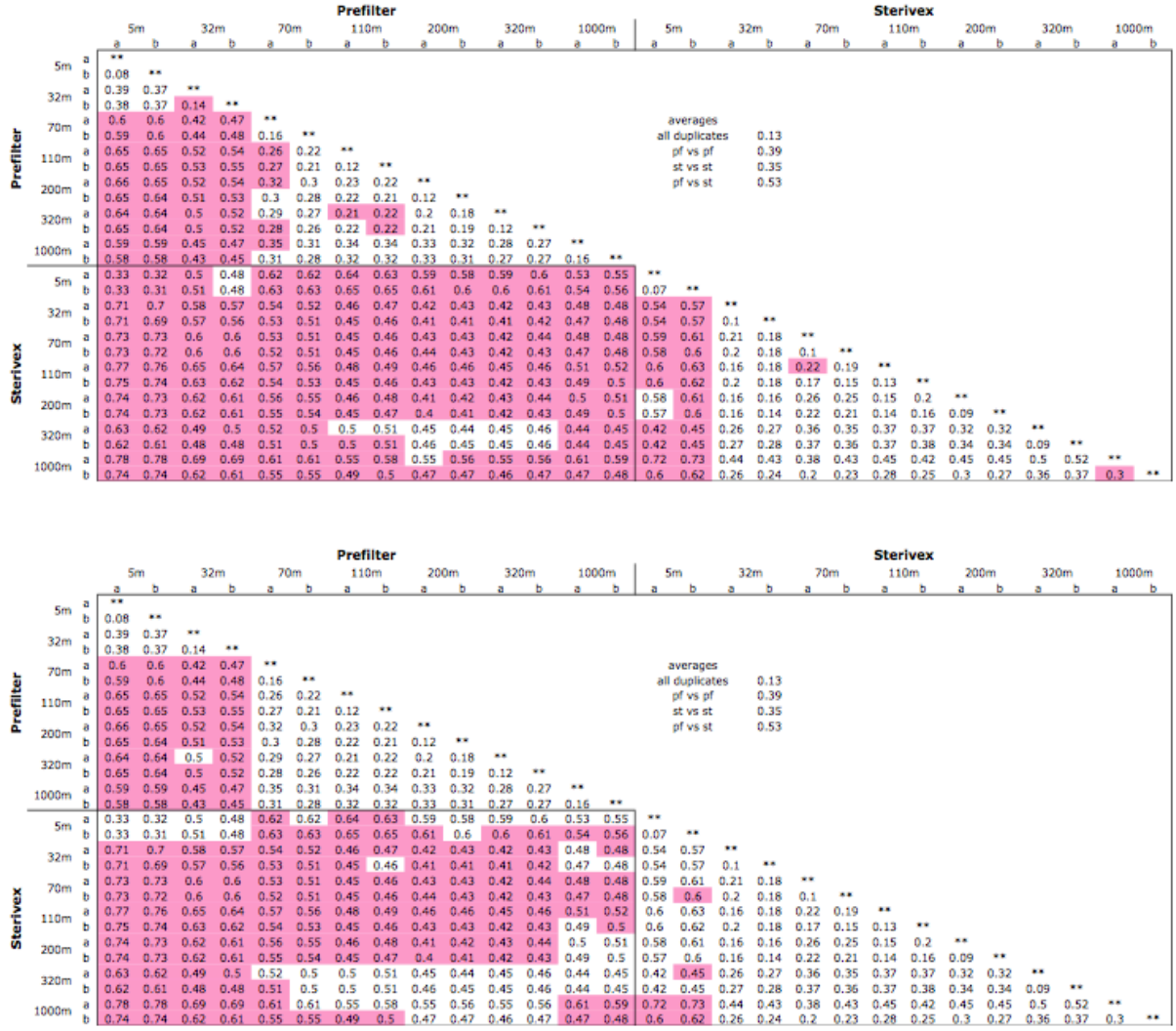
Supplementary Tables

Table A.1 16S rRNA gene amplicon sequencing statistics

Sample[1]	Count	OTU[2]
TOTAL sequences[3]	742,553	
TOTAL following parsing[4]	406,768	
Mean length[5]	409 bp	
TOTAL unique OTUs		17,014
5p,*a*	18,837	1,298
5p,*b*	15,803	1,229
5s,*a*	20,506	1,225
5s,*b*	21,240	1,344
32p,*a*	13,540	1,818
32p,*b*	5,619	956
32s,*a*	18,681	2,354
32s,*b*	14,314	1,902
70p,*a*	6,281	760
70p,*b*	17,412	1,298
70s,*a*	15,929	1,991
70s,*b*	18,917	2,484
110p,*a*	11,336	1,717
110p,*b*	12,059	1,786
110s,*a*	19,057	2,066
110s,*b*	15,300	1,830
200p,*a*	12,026	1,903
200p,*b*	12,142	1,929
200s,*a*	17,436	1,564
200s,*b*	18,572	1,864
320p,*a*	9,647	1,493
320p,*b*	12,490	1,553
320s,*a*	19,410	2,163
320s,*b*	15,796	2,135
1000p,*a*	18,739	1,839
1000p,*b*	5,268	1,080
1000s,*a*	5,256	658
1000s,*b*	15,155	2,205

[1] Samples are labeled by depth and filter type, where p = prefilter (>1.6 µm), s = Sterivex (0.2-1.6 µm), and *a,b* indicate datasets generated by duplicate PCR reactions using DNA from the same sample
[2] operational taxonomic unit (unique cluster of sequences sharing 97% nucleotide similarity)
[3] generated on one full plate 454 run (Titanium chemistry); this plate also included 18S rRNA amplicon samples, which are not included in this analysis
[4] sequences sorted by barcode (sample ID) and parsed based on quality; deleted sequences included those with length <200 or >1000 bp (51,915), with 1 or more ambiguous bases (50,866), missing a quality score (0), with mean quality score <25 (741), with a homopolymer run exceeding 6 bp (4,864), or with mismatches in the primer sequence (177,586)
[5] mean length of sequences remaining after parsing

Table A.2 Weighted Unifrac values comparing community phylogenetic diversity between sample pairs, based on 16S rRNA gene amplicon datasets.



a,b indicate datasets generated by duplicate PCR reactions using DNA from the same sample; shading highlights comparisons showing significant differences between samples following a false discovery rate correction for multiple tests using the method of Benjamini and Hochberg (TOP), or a Bonferroni correction ($\alpha = 0.05$; BOTTOM)

Table A.3 Bray-Curtis distances between metagenome samples based on sequence counts to SEED Subsystems.

		70		110		200		1000	
		p	s	p	s	p	s	p	s
70	p	**							
	s	0.09	**						
110	p	0.09	0.09	**					
	s	0.11	0.05	0.10	**				
200	p	0.09	0.07	0.08	0.08	**			
	s	0.12	0.06	0.12	0.04	0.09	**		
1000	p	0.11	0.11	0.11	0.13	0.10	0.14	**	
	s	0.10	0.07	0.11	0.08	0.09	0.09	0.12	**

p,s indicate prefilter and Sterivex samples, respectively

Table A.4 Normalized[1] metagenome sequence counts per SEED Subsystem[2]

SEED Subsystem	70p	70s	110p	110s	200p	200s	1000p	1000s
Amino Acid Metabolism[3]	**2125**	**2373**	**2320**	**2305**	**2142**	**2540**	**2460**	**2307**
Alanine, serine, and glycine	319	360	321	322	222	316	238	323
Arginine; urea cycle, polyamines	275	319	337	400	382	451	442	331
Aromatic amino acids and derivatives	106	115	118	133	185	89	91	86
Branched-chain amino acids	363	463	502	421	428	458	590	506
Glutamine, glutamate, aspartate, asparagine; ammonia assimilation	106	42	63	26	70	20	102	101
Histidine Metabolism	213	205	212	177	127	176	136	128
Lysine, threonine, methionine, and cysteine	9	0	8	1	12	0	0	0
Lysine, threonine, methionine, and cysteine.1	602	767	666	733	629	906	714	720
Proline and 4-hydroxyproline	133	102	94	91	86	124	147	111
Carbohydrates	**2878**	**3022**	**2876**	**3001**	**2919**	**3025**	**2710**	**2756**
Aminosugars	35	26	47	21	53	22	147	27
Central carbohydrate metabolism	708	987	925	1056	970	1049	918	982
CO2 fixation	345	384	345	412	275	450	227	466
Di- and oligosaccharides	213	96	133	52	132	46	136	116
Fermentation	275	434	306	439	345	460	283	276
Glycoside hydrolases	106	63	39	90	49	70	34	69
Lacto-N-Biose I and Galacto-N-Biose Metabolic Pathway	0	0	0	0	0	1	0	0
Monosaccharides	390	311	408	269	354	300	340	212
One-carbon Metabolism	381	377	353	347	423	336	385	345
Organic acids	204	217	204	214	230	205	159	146
Polysaccharides	53	38	8	14	12	7	45	32
Sugar alcohols	168	89	110	87	74	80	34	84
Cell Division and Cell Cycle	**912**	**1100**	**940**	**1066**	**929**	**1179**	**862**	**1049**
Cell cycle in Prokaryota	629	641	408	620	551	738	510	686
Control of Macromolecuar Synthesis	204	400	462	412	321	406	295	331
YgjD and YeaZ	80	59	71	35	58	35	57	32
Cell wall and Capsule	**1718**	**1679**	**1520**	**1708**	**1694**	**1778**	**1417**	**2028**
Capsular and extracellular polysacchrides	416	368	361	360	329	346	272	449
Cell wall of Mycobacteria	319	289	384	261	300	240	272	407
Gram-Negative cell wall components	257	386	313	431	399	462	317	407
Gram-Positive cell wall components	62	26	24	8	41	16	34	5
Peptidoglycan Biosynthesis	505	503	361	533	506	579	408	646
UDP-N-acetylmuramate from Fructose-6-phosphate Biosynthesis	89	69	47	80	62	95	68	69
YjeE	71	38	31	35	58	39	45	44
Clustering-based Subsystems	**1284**	**1353**	**1285**	**1377**	**1377**	**1472**	**1224**	**1453**
Bacterial Cell Division	531	580	415	529	539	656	340	713
Bacterial RNA-metabolizing Zn-dependent hydrolases	186	194	274	210	247	218	159	188
Carbohydrates.1	0	1	0	3	0	4	0	2

Catabolism of an unknown compound	0	0	0	1	0	1	11	2
CBSS-261594.1.peg.2640	0	0	0	1	0	0	0	0
CBSS-562.2.peg.5158 SK3 including	0	6	0	0	0	0	0	0
Clustering-based subsystems.1	106	84	86	157	95	108	125	74
Conserved gene cluster associated with Met-tRNA formyltransferase	239	216	227	200	247	209	204	192
LMPTP YfkJ cluster	9	9	16	6	8	8	0	0
LMPTP YwlE cluster	35	91	86	66	45	65	79	76
NusA-TFII Cluster	133	150	125	174	177	178	261	151
proteosome related	44	19	47	21	16	25	34	52
Putative hemin transporter	0	1	0	1	0	0	11	0
Putative sulfate assimilation cluster	0	0	0	1	0	0	0	0
Type III secretion system related	0	2	0	3	0	0	0	2
Type III secretion system, extended	0	0	8	3	4	0	0	0
Cofactors, Vitamins, Prosthetic Groups	**1621**	**1654**	**1489**	**1435**	**1517**	**1557**	**1565**	**1532**
Biotin	71	75	63	76	82	92	113	84
Coenzyme A	97	97	157	110	103	106	91	69
Coenzyme B	0	0	0	1	4	2	0	0
Coenzyme F420	44	21	24	15	12	8	23	17
Folate and pterines	390	359	219	340	304	309	329	358
Lipoic acid	53	51	110	55	86	55	68	32
NAD and NADP	133	162	172	133	177	125	147	116
Pyridoxine	27	67	47	70	41	63	34	30
Quinone cofactors	248	171	188	148	197	174	215	170
Riboflavin, FMN, FAD	97	78	94	58	62	51	79	69
Tetrapyrroles	381	461	337	337	317	406	385	456
Thiamin biosynthesis	80	112	78	91	132	165	79	131
DNA Metabolism	**2754**	**2226**	**2132**	**2223**	**2257**	**2252**	**2268**	**2322**
CRISPRs	0	10	31	1	4	4	11	42
DNA recombination	124	63	78	65	78	54	45	67
DNA repair	1293	1064	854	1033	1069	1073	1259	1091
DNA replication	1196	992	1050	1029	1015	1055	771	1059
DNA structural proteins, bacterial	27	19	24	36	21	27	57	15
Restriction-Modification System	97	35	71	35	58	22	23	32
YcfH	18	44	24	25	12	16	102	17
Dormancy and Sporulation	**18**	**2**	**8**	**6**	**4**	**5**	**0**	**5**
Spore Coat	18	0	0	1	0	0	0	0
Spore germination	0	1	0	0	0	0	0	0
Sporulation gene orphans	0	1	8	4	4	5	0	5
Fatty acids, Lipids, Isoprenoids	**682**	**817**	**878**	**903**	**892**	**828**	**714**	**832**
Fatty acids	204	287	329	373	333	301	340	336
Isoprenoids	248	237	290	220	271	213	125	249
Phospholipids	186	247	235	289	255	282	215	237
Polyhydroxybutyrate metabolism	44	44	24	22	33	31	11	5
Triacylglycerols	0	1	0	0	0	0	23	5
Membrane Transport	**221**	**306**	**329**	**311**	**271**	**338**	**317**	**363**

ABC transporters	62	122	63	112	70	115	45	111
peptide_(TC_3.A.1.5.5)	*0*	*0*	*0*	*1*	*4*	*0*	*0*	*0*
branched-chain_amino_acid_(TC_3.A.1.4.1)	*26*	*77*	*41*	*95*	*40*	*89*	*30*	*64*
dipeptide_(TC_3.A.1.5.2)	*9*	*24*	*11*	*3*	*0*	*11*	*15*	*8*
alkylphosphonate_(TC_3.A.1.9.1)	*9*	*11*	*11*	*4*	*13*	*9*	*0*	*22*
tungstate_(TC_3.A.1.6.2)	*18*	*10*	*0*	*9*	*13*	*6*	*0*	*17*
Choline Transport	0	0	8	0	0	0	0	0
ECF class transporters	18	16	0	17	0	13	0	15
Potassium homeostasis	80	112	196	139	127	143	125	143
Sugar Phosphotransferase Systems, PTS	0	2	0	0	4	0	34	7
Transport of Manganese	0	4	31	4	8	1	11	0
Transport of Molybdenum	0	0	0	0	4	0	0	0
Transport of Nickel and Cobalt	9	4	0	0	4	1	0	10
Transport of Zinc	35	35	31	22	41	48	45	54
Uni- Sym- and Antiporters	18	11	0	17	12	16	57	22
Metabolism of Aromatic Compounds	**584**	**432**	**447**	**392**	**604**	**347**	**760**	**688**
Anaerobic degradation of aromatic compounds	18	32	24	32	0	14	0	7
Aromatic Amin Catabolism	0	4	8	4	0	2	0	0
Benzoate transport and degradation cluster	89	69	94	88	107	84	159	76
carbazol degradation cluster	18	7	0	4	16	3	11	5
Gentisare degradation	53	32	8	44	29	45	57	54
Metabolism of central aromatic intermediates	159	116	133	64	197	74	215	155
p-cymene degradation	0	0	0	1	0	0	0	0
Peripheral pathways for catabolism of aromatic compounds	204	126	172	108	222	81	238	333
Salicylate and gentisate catabolism	44	46	8	47	33	43	79	57
Miscellaneous	**221**	**189**	**243**	**213**	**197**	**194**	**283**	**205**
Archease	0	4	0	0	4	1	0	0
Luciferases	0	0	0	1	0	1	0	0
Muconate lactonizing enzyme family	9	1	8	4	0	0	11	7
YaaA	18	1	8	1	8	1	23	12
YbbK	35	27	24	26	29	19	68	17
ZZ gjo need homes	159	155	204	180	156	172	181	168
Motility and Chemotaxis	**558**	**324**	**682**	**249**	**493**	**219**	**703**	**375**
Bacterial Chemotaxis	71	43	63	30	37	28	79	32
Flagellar motility in Prokaryota	487	281	619	218	456	190	624	343
Nitrogen Metabolism	**576**	**817**	**846**	**1011**	**773**	**981**	**476**	**476**
Allantoin Utilization	18	5	16	6	8	3	0	27
Amidase clustered with urea and nitrile hydratase functions	0	2	0	0	4	0	0	5
Ammonia assimilation	213	433	337	452	407	476	374	380
Denitrification	53	23	31	23	21	25	0	10
Dissimilatory nitrite reductase	0	4	24	4	8	7	0	0
Nitrate and nitrite ammonification	239	322	345	497	300	440	45	12
Nitric oxide synthase	18	12	24	4	8	2	23	35
Nitrogen fixation	9	7	8	6	4	3	23	0

Nitrosative stress	27	7	63	19	12	24	11	7
Nucleosides and Nucleotides	**1311**	**1331**	**1207**	**1323**	**1361**	**1239**	**850**	**1308**
Adenosyl nucleosidases	9	4	8	3	16	2	23	7
AMP to 3-phosphoglycerate	0	2	0	0	0	0	0	0
De Novo Pyrimidine Synthesis	115	195	118	162	201	256	113	289
Detoxification.1	35	32	31	33	53	16	11	39
Hydantoin metabolism	213	117	259	167	247	112	45	106
Purine Utilization	213	148	219	135	156	108	34	99
Purines	602	667	415	627	514	571	499	624
Pyrimidines	35	48	16	22	41	16	45	35
Ribonucleotide reduction	89	117	141	174	132	157	79	109
Phages, Prophages, Transposable Elements	**115**	**80**	**71**	**112**	**78**	**85**	**91**	**84**
IbrA and IbrB: co-activators of prophage gene expression	0	0	0	3	4	0	0	0
Listeria phi-A118-like prophages	0	0	0	0	4	0	0	0
Staphylococcal phi-Mu50B-like prophages	115	74	55	108	70	85	79	76
Tn552	0	6	16	1	0	0	11	7
Phosphorous Metabolism	**407**	**252**	**251**	**305**	**288**	**256**	**442**	**338**
Alkylphosphonate utilization	35	12	8	14	4	8	68	25
High affinity phosphate transporter and control of PHO regulon	53	5	8	1	25	2	23	17
Phosphate metabolism	301	233	235	287	255	242	351	279
Phosphonate metabolism	18	1	0	3	4	4	0	17
Photosynthesis	**115**	**25**	**31**	**3**	**0**	**3**	**0**	**2**
Electron transport and photophosphorylation	106	21	31	3	0	3	0	2
Light-harvesting complexes	9	4	0	0	0	0	0	0
Potassium Metabolism	**44**	**12**	**24**	**7**	**16**	**4**	**45**	**10**
Glutathione-regulated potassium-efflux system and associated functions	44	12	24	7	16	4	45	10
Protein Metabolism	**3232**	**3133**	**3135**	**3158**	**3100**	**2929**	**3413**	**3316**
Protein biosynthesis	1877	1945	1920	1924	1813	1755	1995	2008
Protein degradation	514	429	400	412	423	366	385	533
Protein folding	602	548	596	539	555	560	692	528
Protein processing and modification	124	122	110	195	177	169	91	126
Secretion	62	16	47	28	21	28	113	44
Selenoproteins	53	73	63	62	111	51	136	76
Regulation and Cell Signaling	**248**	**231**	**376**	**244**	**234**	**213**	**351**	**210**
cAMP signaling in bacteria	142	68	102	55	62	41	113	67
CytR regulation	0	1	31	0	0	0	0	0
DNA-binding regulatory proteins, strays	0	10	0	8	0	4	11	7
Orphan regulatory proteins	18	16	78	12	45	12	34	0
Programmed Cell Death and Toxin-antitoxin Systems	0	7	0	6	16	12	11	2
Proteolytic pathway	0	0	24	0	0	0	0	5
Sex pheromones in Enterococcus faecalis and other Firmicutes	53	28	0	36	37	44	45	30
Stringent Response, (p)ppGpp metabolism	0	64	86	75	25	54	45	57
Two-component regulatory systems in Campylobacter	35	36	55	52	49	47	91	42

Virulence	**2754**	**2093**	**2751**	**1924**	**2713**	**1708**	**3197**	**1902**
Adhesion	133	67	78	97	132	57	136	67
Detection	44	32	39	41	45	41	79	5
Invasion and intracellular resistance	151	134	102	128	173	129	204	123
Iron Scavenging Mechanisms	496	386	470	334	419	228	488	338
Pathogenicity islands	124	130	118	101	107	117	181	91
Pseudomonas quinolone signal PQS	0	0	0	1	0	0	0	0
Quorum sensing and biofilm formation	9	11	8	11	12	4	11	2
Regulation of virulence	27	32	31	43	25	44	68	39
Resistance to antibiotics and toxic compounds	992	699	1074	631	917	626	1020	681
Streptococcus pyogenes Virulome	0	0	8	0	0	0	11	0
Ton and Tol transport systems	487	495	486	454	563	403	646	479
Toxins and superantigens	0	1	8	0	8	0	0	0
Type III, Type IV, Type VI, ESAT secretion systems	292	106	329	81	312	59	351	76
TOTAL non-normalized reads matching SEED[2]	2823	20268	3190	18104	6081	22837	2205	10132

[1] Counts are normalized to a standard dataset size (number of prokaryote reads matching SEED) of 25,000 to allow comparison across columns. To determine true read counts, divide values by the sample-specific scaling factor: 8.86 (70p), 1.23 (70s), 7.84 (110p), 1.38 (110s), 4.11 (200p), 1.09 (200s), 11.34 (1000p), 2.47 (1000s).

[2] NOTE: Some gene sequences are classified into multiple Subsystems. The sum of read counts across all Subsystems exceeds the total number of SEED-assigned reads by 12-15%.

[3] Bolded values are sums of Subsystem (Level 2) counts per broader functional category (Level 1).

Supplementary Figures

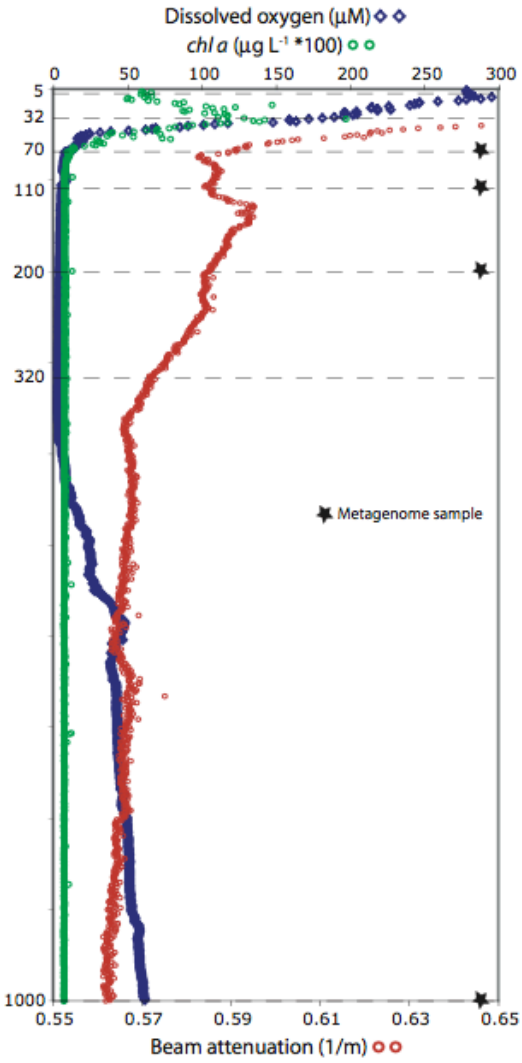


Figure A.1 Dissolved oxygen, chlorophyll a, and particulate matter (beam attenuation coefficient) concentrations at Station 1 (20o 04.999S, 70o 48.001W). Data for DO and chl a are based on CTD-based DO sensor and fluorometer measurements from Cast #10 at ~2200 hrs on November 21, 2010. Data for beam attenuation are averages of transmissometer measurements from 20 casts at station #1 from November 19-24. Attenuation coefficients are not shown for depths less than 40 m, where values peaked at ~0.8 m^{-1} close to the surface. 16S rRNA gene amplicon diversity was analyzed for DNA collected at 7 depths (dashed lines). Metagenome analyses were conducted for 4 of these depths (stars). Note that chl a values (in ug L-1) have been multiplied by 100 to match axis units. Oxygen profiles from other days during the 5-day sampling period (Nov. 19-23) closely resembled that shown in the Figure.

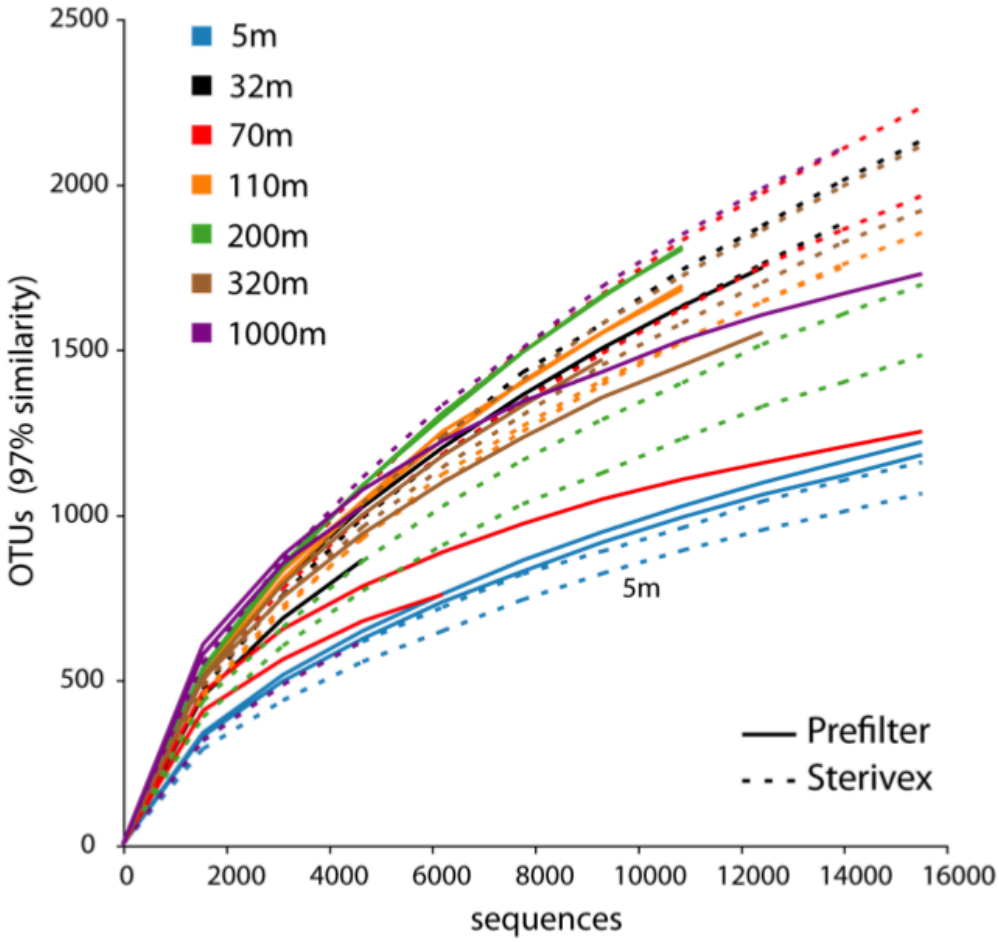


Figure A.2 Number of observed OTUs (97% similarity clusters) as a function of sequencing depth, based on rarefaction of OTU counts. Duplicate samples reflect duplicate PCR reactions.

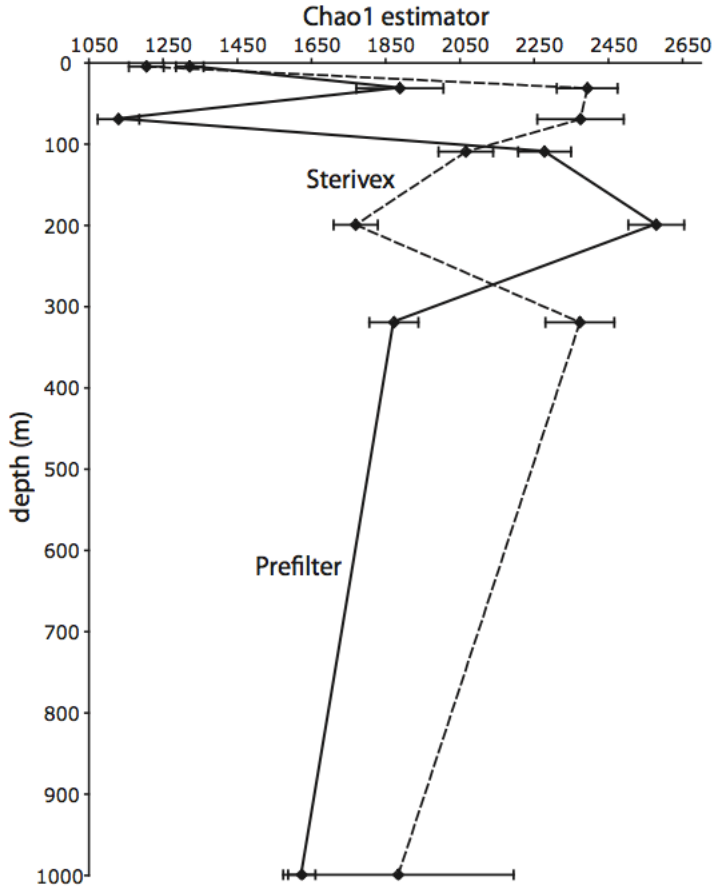


Figure A.3 Chao1 estimator as a function of water column depth. Data points are mean values based on rarefaction at a standardized sequence count (n = 4996) per sample, with bars indicating 95% confidence intervals for the rarefied measurements. Data from both PCR duplicates are combined for averaging.

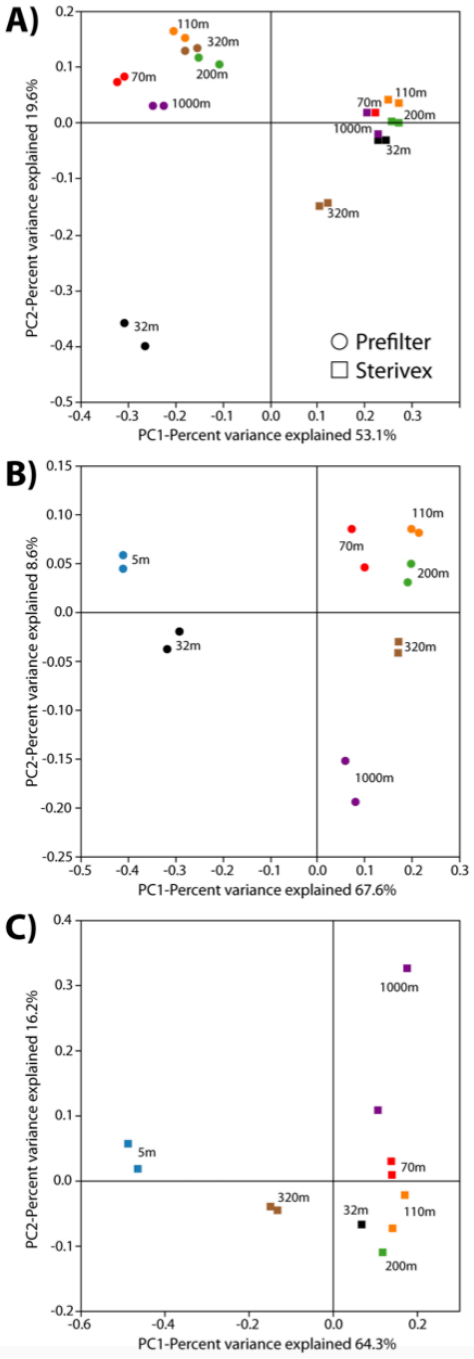


Figure A.4 Principle component analysis of community relatedness based on relative bacterial taxon abundance in the 16S rRNA gene amplicon pool, as quantified by the weighted Unifrac metric. A) Excludes samples from the 5m depth. B) Prefilter only samples. C) Sterivex only samples

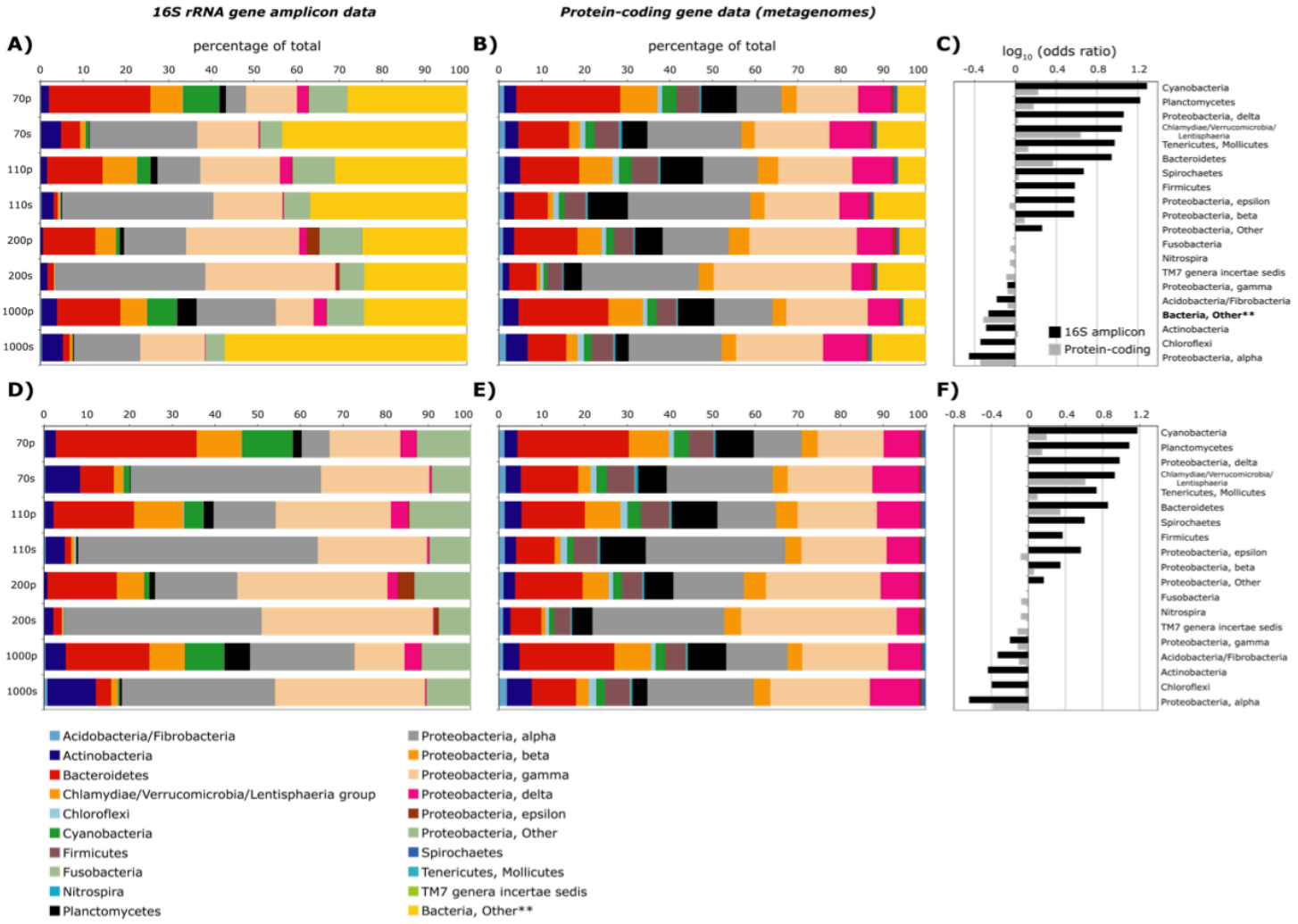

16S rRNA gene amplicon data
Protein-coding gene data (metagenomes)
A)
B)
C)
D)
E)
F)
percentage of total
log10 (odds ratio)
70p
70s
110p
110s
200p
200s
1000p
1000s
16S amplicon
Protein-coding
Acidobacteria/Fibrobacteria
Actinobacteria
Bacteroidetes
Chlamydiae/Verrucomicrobia/Lentisphaeria group
Chloroflexi
Cyanobacteria
Firmicutes
Fusobacteria
Nitrospira
Planctomycetes
Proteobacteria, alpha
Proteobacteria, beta
Proteobacteria, gamma
Proteobacteria, delta
Proteobacteria, epsilon
Proteobacteria, Other
Spirochaetes
Tenericutes, Mollicutes
TM7 genera incertae sedis
Bacteria, Other**

Figure A.5 Relative abundance of major bacterial divisions at four depths in the OMZ based on the taxonomic identification of 16S rRNA gene fragments (A,D) and protein-coding genes identified in metagenomes (B,E).
16S rRNA data are those from Figure 2.1 (main text), with certain groups collapsed to higher taxonomic levels to match taxonomic groupings based on protein-coding genes. Panels A and B include sequences in the category "Bacteria, Other**", which contains sequences unidentifiable to Phylum level, as well as sequences in bacterial divisions that were detected in one analysis (16S or metagenome) but not the other. These include 1) 130 candidate divisions BRC1, OD1, SR1 and WS3, which were identifiable only in the 16S amplicon analysis, and 2) the Aquificae, Caldiserica, Deferribacteres, Deinococcus-Thermus, Elusimicrobia, Gemmatimonadetes, Synergistetes, Thermotogae, and candidate divisions NC10 and OP1, all of which were identifiable only based on protein-coding gene annotations. Protein-coding genes annotated as "unknown prokaryote" were also included in the "Bacteria, Other" category, although a proportion of these could belong to Archaea. Panels C and D exclude the "Bacteria, Other" category. The category "Proteobacteria, Other" was present only in the 16S amplicon analysis (based on the RDP classification in QIIME). Panels C and F show variation in the relative abundance of bacterial divisions between filter size fractions, as inferred from both 16S and protein-coding gene data. Values are the base-10 logarithm of the odds ratio: the ratio of the odds a taxon occurs in the prefilter fraction to the odds it occurs in the Sterivex fraction. Positive values indicate taxa that are more likely to occur in the prefilter. Values are based on counts pooled from only the four depths shown, with corrections for differences in dataset size. Relative abundances are calculated as a percentage of total identifiable bacterial sequences in C, and total identifiable bacterial sequences excluding "Bacteria, Other" in F. Samples are labeled by depth and filter type, where p = prefilter (>1.6 μm), s = Sterivex (0.2-1.6 μm).

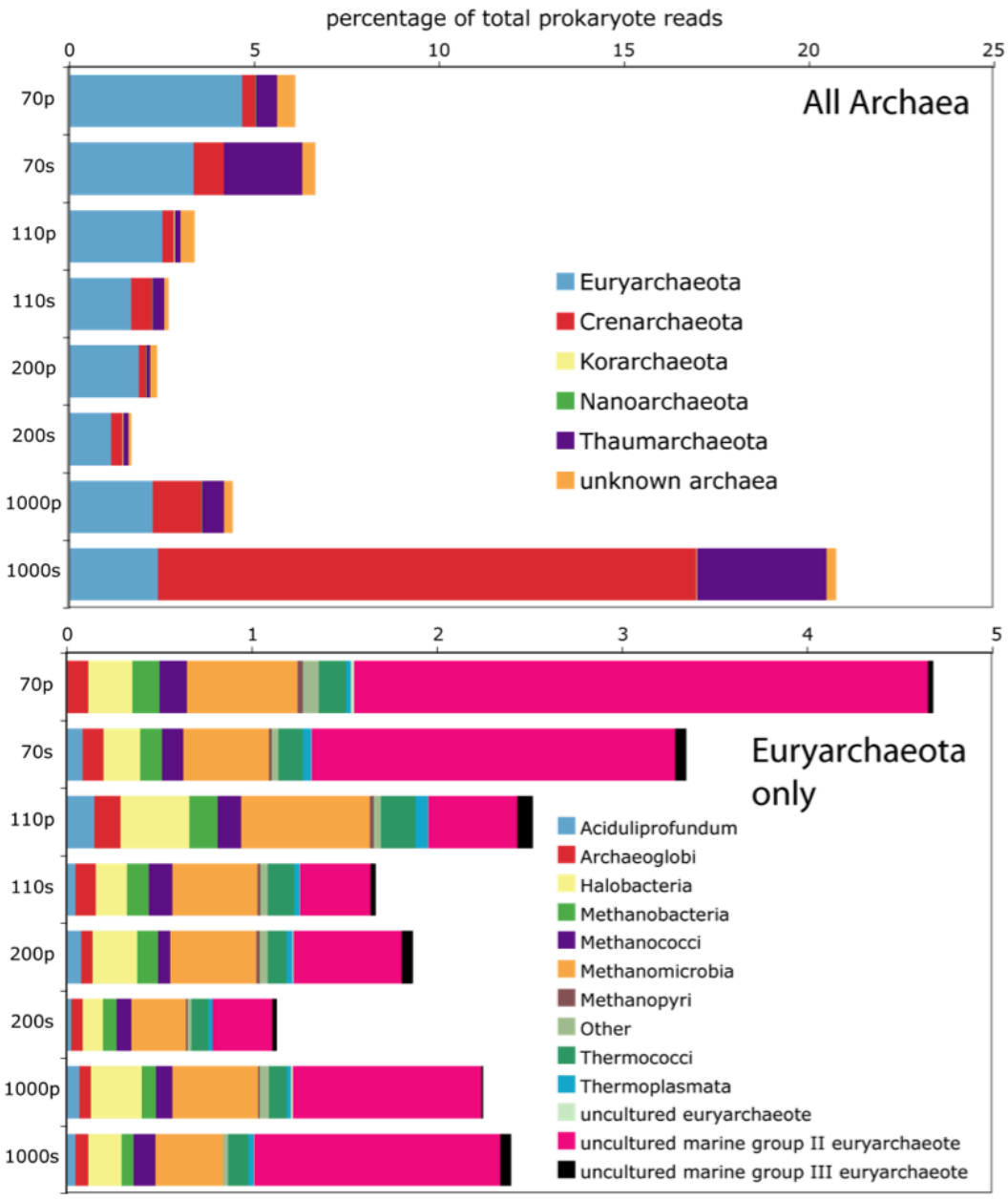


Figure A.6 Relative abundance and phylum-level taxonomic composition of protein-coding reads matching the domain Archaea. Relative abundance is expressed as a percentage of total prokaryote protein-coding reads. Phylum designations are based on annotations of NCBI-nr genes identified as top matches via BLASTX (bit score > 50). Samples are labeled by depth and filter type, where p = prefilter (>1.6 µm), s = Sterivex (0.2-1.6 µm)

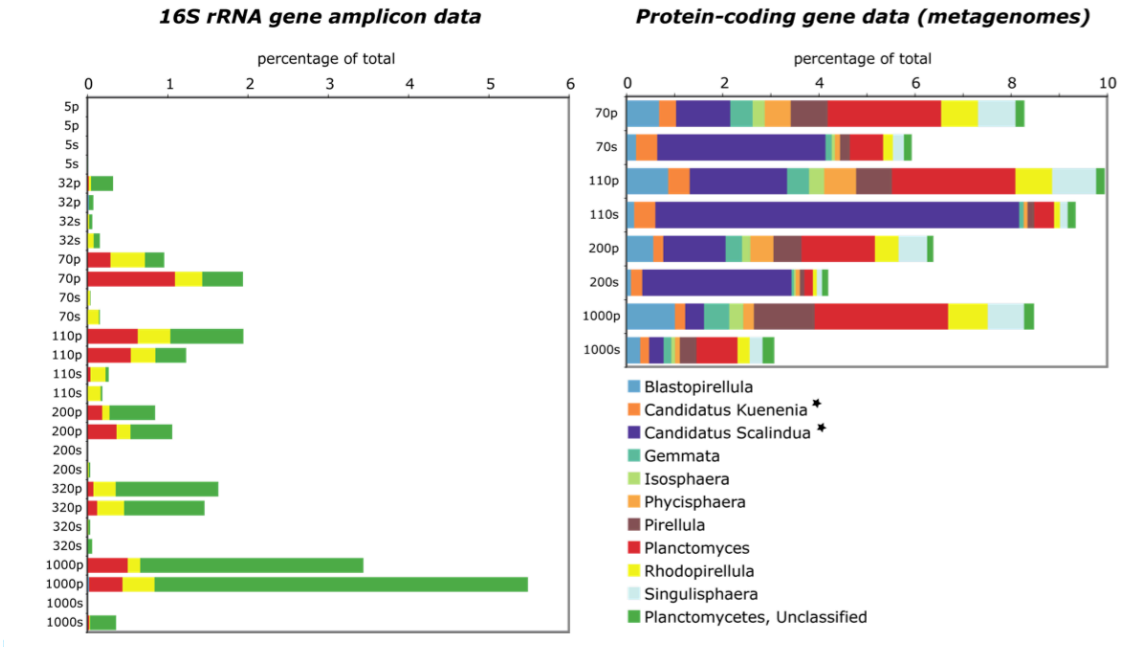


Figure A.7 Relative abundance and genus-level taxonomic composition of 16S rRNA (left) and protein-coding (right) reads matching Planctomycetes. Relative abundance is expressed as a percentage of total bacteria protein-coding reads (Protein-coding genes annotated as "unknown prokaryote" were also included in the total bacteria count, although a proportion of these could belong to Archaea). Taxon designations are based on RDP classifications of 16S rRNA gene fragments (left) and the annotations of NCBI-nr genes identified as top matches via BLASTX (bit score > 50; right). Samples are labeled by depth and filter type, where p = prefilter (>1.6 μm), s = Sterivex (0.2-1.6 μm). Stars mark taxa known to conduct anammox

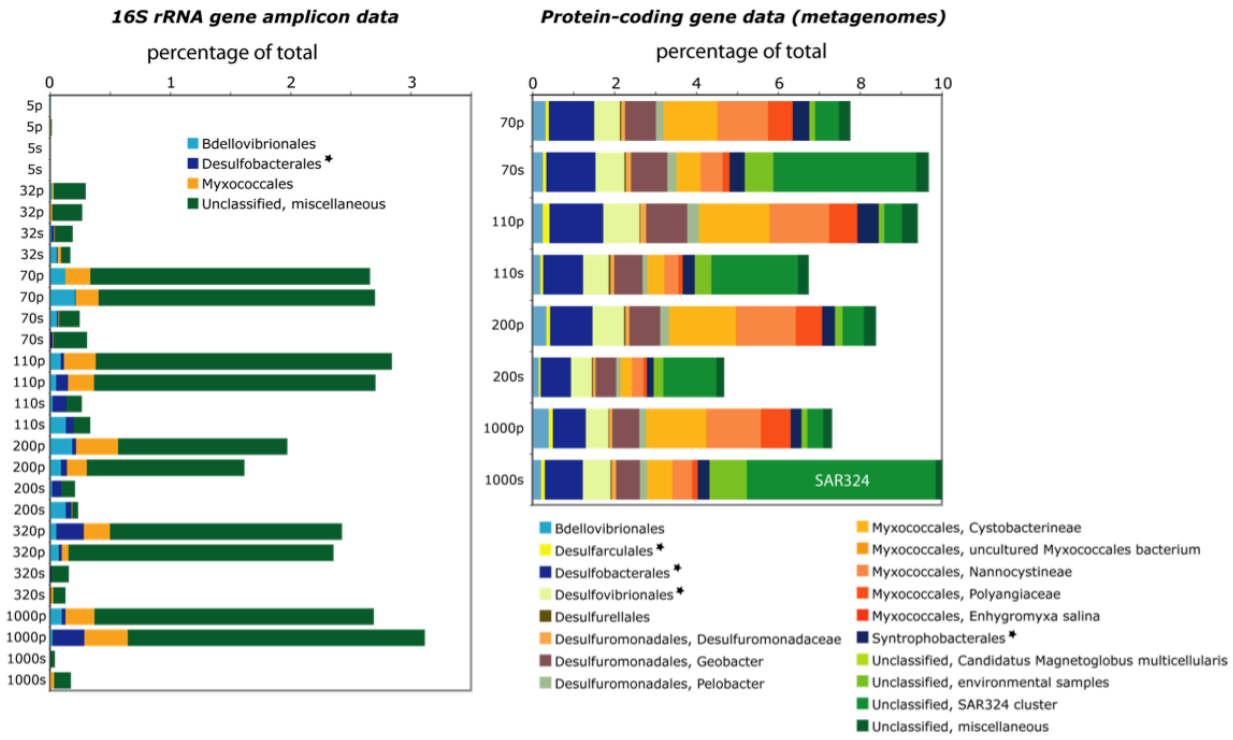


Figure A.8 Relative abundance and taxonomic composition of 16S rRNA (left) and protein- coding (right) reads matching Deltaproteobacteria. Relative abundance is expressed as a percentage of total bacterial reads (Protein-coding genes annotated as "unknown prokaryote" were also included in the total bacteria count, although a proportion of these could belong to Archaea). Taxon designations are based on RDP classifications of 16S rRNA gene fragments (left) and the annotations of NCBI-nr genes identified as top matches via BLASTX (bit score > 50; right). Samples are labeled by depth and filter type, where p = prefilter (>1.6 µm), s = Sterivex (0.2-1.6 µm). Stars mark groups known to contain sulfate-reducing members.

APPENDIX B

SUPPLEMENTARY MATERIAL FOR CHAPTER 3

SIZE-FRACTION PARTITIONING OF COMMUNITY GENE TRANSCRIPTION AND NITROGEN METABOLISM IN A MARINE OXYGEN MINIMUM ZONE

Supplementary Methods

Oxygen measurements

In addition to the SBE 43 oxygen sensor, Switchable Trace amount OXygen (STOX)sensors were mounted to the rosette and used to quantify in situ trace oxygen concentrations (Revsbech et al., 2009). Signals from two STOX sensors were simultaneously recorded by a custom standalone data logging unit, consisting of a 16-bit A/D converter (DT9816, Data translation) controlled by a single board computer (fit-PC2i, CompuLab) housed in a titanium cylinder mounted on the rosette. Amplification of the sensor signals was performed using a custom made amplifier. Sensor switching was controlled by a cyclic switch operating with a 60 sec on/off cycle. The signal was sampled at 120 Hz and subsequently smoothed using a 10 sec moving average and binned in 1 sec intervals. The detection limit of the STOX sensors was estimated to be ~9 nM, based on three times the standard deviation of the zero signal. The precision of the measurements was typically ~5-8 nM in the core of the OMZ. The concentrations presented in Figure 3.1 are based on 1-3 measurement cycles during periods where the rosette was stationary in the water column. Calibration and calculations were performed as in Revsbech et al., (2009) and Thamdrup et al., (2012).

Marker gene analysis – Metatranscriptomes and metagenomes

Transcript and gene sequences with significant matches to references in the NCBI-nr database (above bit score 50) were used to analyze marker genes of dissimilatory nitrogen metabolism: ammonia monooxygenase (amoC), nitrite oxidoreductase (nxrB), hydrazine oxidoreductase (hzo), nitrate reductase (narG), nitrite reductase (nirK + nirS), nitric oxide reductase (norB), and nitrous oxide reductase (nosZ). Nitrite reductase genes were characterized as nirK or nirS variants, encoding the copper-containing and cytochrome cd1-containing nitrite reductases, respectively, but pooled for representation in Figure 3.5. Transcripts matching nrfA, encoding the cytochrome c nitrite reductase, were at negligible abundance or not detected and were excluded from analysis. BLASTX results were parsed via keyword queries based on NCBI-nr annotations, as in Canfield et al. (2010) and Ganesh et al. (2014). Genes recovered as top BLASTX matches were parsed from GenBank, and each GenBank annotation was examined manually to confirm gene identity based on length and conserved protein domains. Genes with ambiguous annotations were verified by BLASTX. Abundances were normalized based on best approximate gene length (kb), estimated as in Ganesh et al. (2014) 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); norB (1410 bp); nosZ (1950 bp). Transcript counts per kilobase of gene were normalized to counts of transcripts matching the universal, putatively single-copy gene encoding RNA polymerase subunit B (rpoB, 4020 bp), where a value of 1 (Figure 3.5) indicates abundance in the metatranscriptome equivalent to that of rpoB, assuming the gene lengths above

Differential expression of taxon-specific genes

An empirical Bayesian approach using the program baySeq (Hardcastle and Kelly, 2010) was used to identify taxon-specific genes differentially expressed between FL and PA fractions. These analyses focused on taxon-specific subsets of genes with top BLASTX matches to 1) genes of all known anammox-capable bacterial genera (Scalindua, Kuenenia, Brocadia, Jettenia, Anammoxoglobus), and 2) the genome of the nitrite-oxidizer Nitrospina gracilis. These taxon gene subsets were targeted based on their representation in both FL and PA fractions, and because the functional niches (anammox and aerobic nitrite oxidation) of these groups are known OMZ-associated processes. As replicates at each depth were not available, all samples (depths) belonging to each size fraction were modeled as biological replicates. Dispersion was estimated via a quasi-likelihood method, with the count data normalized by data subset size (e.g., total number of sequences with top matches to anammox genera). Posterior likelihoods per gene were calculated for models (sample groupings) in which genes were predicted to be either non-DE or DE between FL and PA fractions. A false discovery rate threshold of 0.05 was used for detecting DE categories. These analyses evaluate whether a taxon's transcriptome profile (i.e., the relative abundances of genes within the taxon-specific subset) varies between FL and PA niches, irrespective of that taxon's contribution to the metatranscriptome.

Evaluation of filtration on oxygen content in exetainer incubations

We repeated all incubation steps as described in the main text, from the filling of

exetainers with and without filtration of degassed water to their incubation, while monitoring oxygen in the exetainers using custom made highly sensitive optical trace oxygen sensors (Borisov et al., 2011). Oxygen concentrations remained below 80 nM in all treatments, demonstrating that the filtration step did not introduce any additional oxygen contamination.

Nutrient measurements

Nitrite concentrations were determined spectrophotometrically aboard ship using the Griess method (Grasshoff et al., 1983). Ammonium concentrations were determined fluorometrically aboard ship using the orthophthaldialdehyde method (Holmes et al., 1999). Samples for nitrate and phosphate concentrations were filtered (0.45μm cellulose acetate) and frozen until analysis. Concentrations of nitrate + nitrite were determined using chemiluminescence after reduction to nitric oxide with acidic vanadium (III) (Braman and Hendrix, 1989). Phosphate was analyzed spectrophotometrically according to Grasshoff et al.(1983).

Supplementary Results and Discussion

ETNP OMZ community composition

Two groups with potential roles in OMZ sulfur and nitrogen cycling were also abundant in the free-living microbial community. Of the major bacterial divisions, the uncultured SAR406 cluster represented the single largest proportion of the FL amplicon dataset, contributing an average of 22% of all FL amplicons, compared to 2.8% in particulate fractions (Figure 3.3). SAR406 sequences in the FL fraction grouped

exclusively with subclades Arctic96B-7 (90%) and ZA3648c (10%) and increased with depth into the OMZ, peaking at 33% of sequences at the secondary nitrite maximum (125 m). Increased SAR406 abundance with deoxygenation has been reported for less oxygen-depleted waters in the North Pacific (Allers et al., 2013), but also in the ETNP (Beman and Carolan, 2013). Recent genomic evidence suggests that SAR406, also known as 'Marine Group A', may participate in OMZ sulfur cycling, potentially via dissimilatory polysulfide reduction to sulfide or dissimilatory sulfide oxidation (Wright et al. 2014). The ubiquitous SAR324 lineage of Deltaproteobacteria, which in other regions has been shown to contain genes for sulfur chemolithoautotrophy (Swan et al., 2011, Sheik et al., 2014), is also abundant in the ETNP FL community, representing 5-8% of total sequences in the 0.2-1.6 μm fraction from OMZ depths, compared to less than <0.5% in particulate fractions (Figure B.3). The enrichment of SAR406 and SAR324 in the FL fraction is consistent with an autotrophic lifestyle, as autotrophs presumably would not depend on attachment to organic particles for carbon acquisition. In other OMZs, sulfide-oxidizing autotrophs have been shown to use oxidized nitrogen species (e.g., nitrate) as terminal oxidants (Walsh et al., 2009, Canfield et al., 2010), thereby coupling dissimilatory OMZ sulfur and nitrogen cycles. Metagenome analyses of hydrothermal plume communities have identified SAR324 genes mediating dissimilatory nitrite reduction (Sheik et al., 2014). However, pathways for denitrification and carbon fixation have not been unambiguously identified in SAR406. It remains unclear whether either group participates in sulfur-driven denitrification in the ETNP.

The composition of the ETNP bacterial community with roles in OMZ sulfur cycling was distinct from that of other OMZs, and also varied among size fractions. Sulfur-oxidizing

Gammaproteobacteria, notably those of the SUP05 clade (Oceanospirillales), have been observed worldwide as abundant community members in both seasonally and permanently oxygen-depleted waters (Stevens and Ulloa, 2008; Zaikova et al., 2010), representing as much as 30% of total prokaryotes in some systems (Glaubitz et al., 2013). Surprisingly, SUP05-affiliated sequences were rare in the ETNP dataset, accounting for <0.2% of total sequences in all samples, consistent with results of Beman and Carolan (2013). In contrast, Gammaproteobacteria of the Thiohalorhabdales, Chromatiales, and Thiotrichales, all of which contain putative sulfur-oxidizing members, represented 7-23% of all sequences at the deeper OMZ depths (100, 125, 300m; Figure B.3). The proportional abundance of these groups was highest in the intermediate size fraction (1.6-30 μm), in which the Thiohalorhabdales represented 16% of total sequences. Thiohalorhabdales are chemolithoautotrophs that have been shown to oxidize thiosulfate under anaerobic conditions using nitrate as an electron acceptor (Sorokin et al., 2008). Relatives of this group have been isolated from hypersaline lake sediments (Sorokin et al., 2008) and microbial mats (Isenbarger et al., 2008). However, the group remains relatively under-characterized and, to our knowledge, has not been reported from OMZs, although Thiohalorhabdales-like sequences have been detected in mucus from diseased corals (Roder et al., 2014). Together, these results suggest that Thiohalorhabdales are either larger cells (and therefore retained preferentially in PA fractions) or are directly attached to particles, and raise the possibility for contributions by this group to sulfur-driven nitrate reduction in the ETNP OMZ.

Size fraction-specific rates

For a subset of depths and 15N amendments, an additional size fraction, water without particles > 30 μm, was analyzed following an identical protocol to that outlined in the Methods. I.e., filtration through a nylon net disc filter (30 μm pore-size, 47 mm dia., Millipore) using a peristaltic pump. For these analyses, rates for each fraction were determined subtractively. For example, rates for the 1.6-30 μm size fraction were calculated by subtracting rates measured in the <1.6 μm fraction (obtained after pre-filtration through 30 μm and 1.6 μm filters), from bulk rates measured following pre-filtration through only a 30 μm filter. For most processes, rates were shown to be confined primarily to the 1.6-30 μm fraction, with little activity in the >30 μm fraction. Notably, in samples for which nitrate reduction was measured in all three fractions, activity was almost completely confined to the 1.6 to 30 μm fraction (up to 97% of activity; Figure B.6). Similarly, N_2 production by denitrification was proportionally enriched in the 1.6 to 30 μm PA fraction (55% of activity) compared to the larger fraction (Figure B.6). These values suggest that the bulk of particles may be below 30 μm in size. The exception involved rates of nitrite oxidation at 91 m (oxic-nitrite interface) and 100 m (secondary chlorophyll maximum), where rates in the largest fraction were relatively enriched (Figure B.6), suggesting a potential role for larger particles as substrate or cell sources in the upper OMZ.

Supplementary Tables and Figures

Table B.1 16S rRNA gene amplicon, metatranscriptome, and metagenome** sequencing statistics

	16S rRNA gene		Metatranscriptome							Metagenome						
sample	Reads[1]	OTU[2]	Reads[3]	%bact	%arch	%euk	%vir	%other	SEED[4]	Reads[3]	%bact	%arch	%euk	%vir	%other	SEED[4]
0.2-1.6 µm																
30m	20,222	595	87,916	71.5	6.3	17.6	0.8	3.4	2678	1,223,450	80	11.4	1.7	3.1	3.7	45,893
85m	19,725	594	139,160	83.5	3.8	8.7	0.5	3.3	2577	1,654,938	83.6	9.2	0.9	1.1	4.9	73,656
100m	18,277	657	129,225	91.8	1.7	4	0.4	2.1	1057	1,055,121	89.6	3.7	1	0.9	4.5	40,027
125m	19,376	635	197,419	90.7	2.6	3.6	0.4	2.6	1582	721,275	90.7	3.3	0.9	0.5	4.4	25,511
300m	20,049	627	87,610	90.5	3.1	2.3	0.4	3.6	1230	3,161,196	91.5	2.4	0.8	0.4	4.7	111,942
1.6-30 µm																
30m	32,800	739	99,217	20.5	0.2	75.2	0.3	3.5	465	6,110	56.8	5.8	27.2	7.8	2.1	138
85m	25,068	687	99,759	27.8	0.4	68.5	0.6	2.5	458	111,725	65.8	6.2	15.9	9.3	2.5	4,869
100m	19,718	817	132,783	43.3	0.5	52.8	0.4	2.7	541	72,951	76.2	2.5	12.6	6.6	1.7	3,046
125m	20,945	830	177,355	35.8	0.5	61.4	0.2	1.8	451	30,401	76.6	3.4	15.5	2.5	1.8	897
300m	21,673	843	124,766	70.7	1.3	24.4	1.1	2.4	1482	69,399	81.5	3	10.1	2.5	2.7	2,248
>30 µm																
30m	39,503	727	-	-	-	-	-	-	-	-	-	-	-	-	-	-
85m	8,665	473	-	-	-	-	-	-	-	-	-	-	-	-	-	-
100m	15,186	702	-	-	-	-	-	-	-	-	-	-	-	-	-	-
125m	12,276	785	-	-	-	-	-	-	-	-	-	-	-	-	-	-
300m	6,708	601	-	-	-	-	-	-	-	-	-	-	-	-	-	-

[1] no. merged reads following QC filtering
[2] no. estimated OTUs (97% similarity clusters) based on rarified counts (n = 6506 sequences)
[3] no. merged reads matching coding genes in the NCBI-nr database (above bit score 50), following QC filtering
% = percentage of coding reads matching reference genes from bacteria (bact), archaea (arch), eukaryotes (euk), viruses (vir), or other (genes lacking a taxonomic annotation, or annotated as "unknown")
[4] no. of reads assigned to SEED subsystem categories
**Metagenome datasets were generated in a prior study (NCBI Sequence Read Archive accession SRP044185

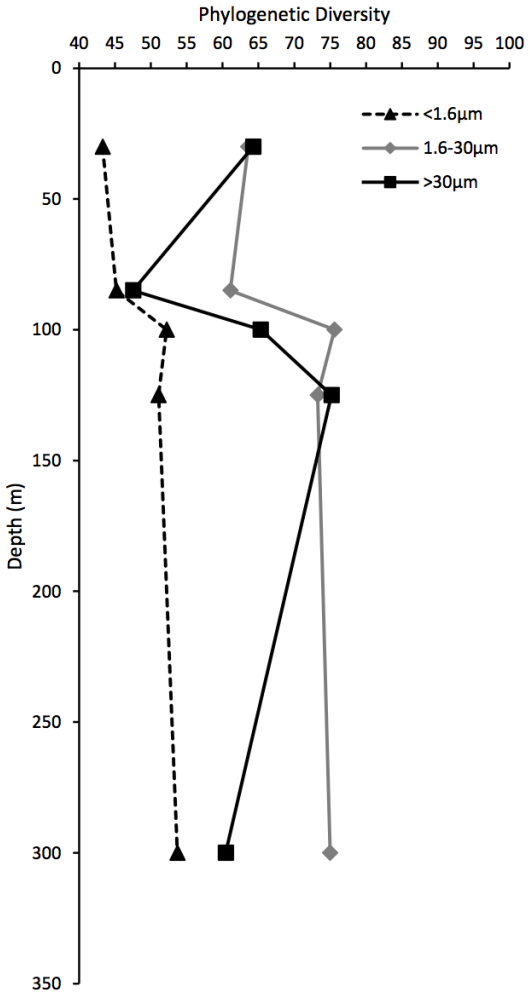


Figure B.1 Phylogenetic diversity as a function of water column depth. Data points are mean values based on rarefaction of OTU (97% similarity) counts at a standardized sequence count (n = 6506) per sample.

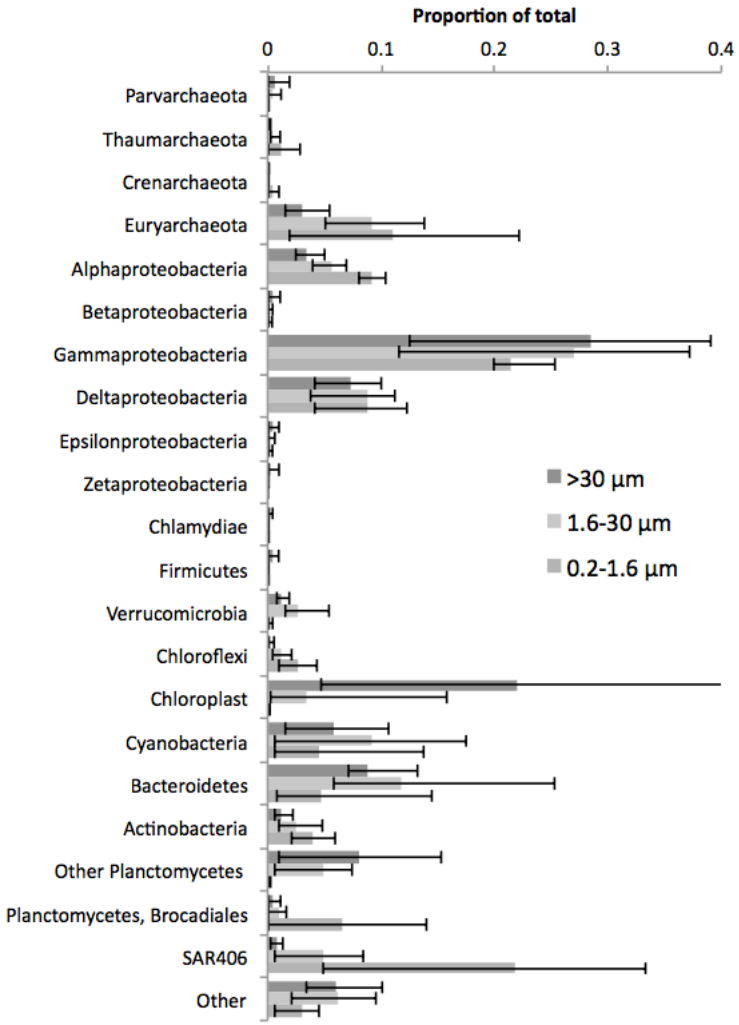


Figure B.2 Average proportional abundances of 16S rRNA gene amplicons affiliated with major microbial taxa. Bars are averages of five depths, partitioned according to filter size fraction. Thin bars reflect ranges. "Other" includes 34 taxonomic divisions, as well as unassigned sequences.

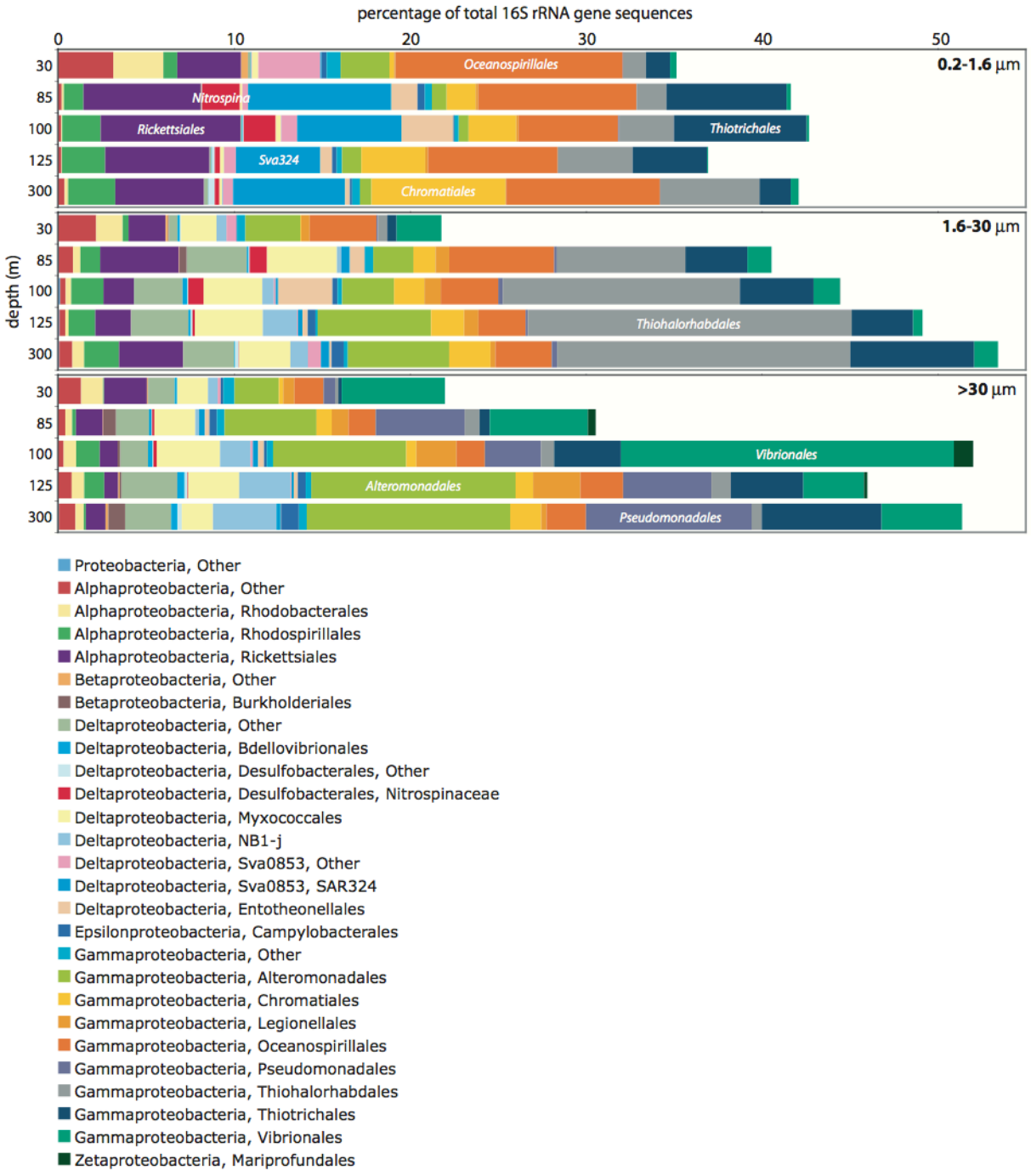


Figure B.3 Taxonomic composition of 16S rRNA gene amplicons within the Phylum Proteobacteria. The abundances of major divisions are shown as a percentage of total identifiable 16S rRNA gene sequences. Taxonomic identifications are based on the Greengenes taxonomy

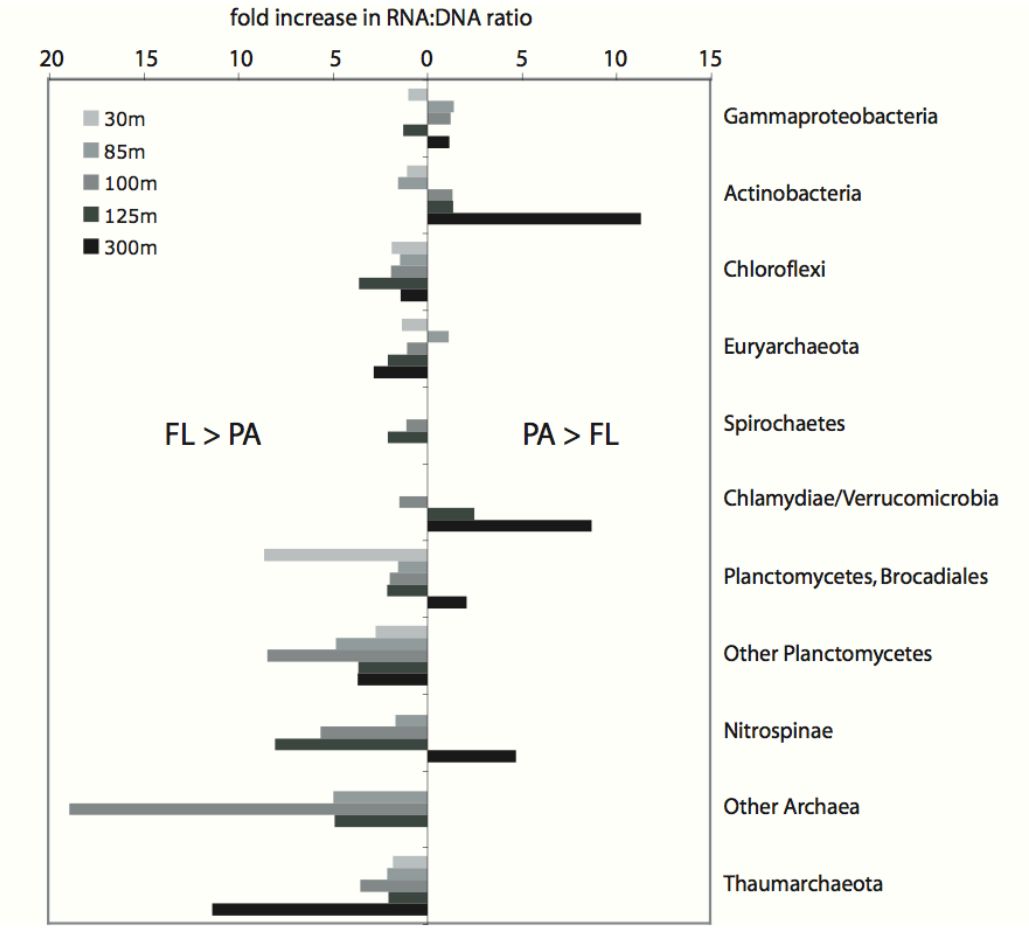


Figure B.4 Fold increase in taxon RNA:DNA between PA (1.6-30 μm) and FL (0.2-1.6 μm) communities. Plot shows the fold increase in ratios for major taxonomic groups identified by the Greengenes identities of 16S rRNA gene amplicons (DNA) and 16S rRNA transcript (RNA) fragments in metatranscriptome datasets. These patterns are similar to those reported in the main text based on the identities of protein-coding sequences, but show higher levels of variation among depths. RNA sequences in metatranscriptomes were identified using riboPicker and analyzed in QIIME as described in the main text. Values right of zero indicate higher ratios in PA communities (PA/FL > 1). Values left of zero indicate higher ratios in FL communities (FL/PA > 1). The plot shows only those groups shown in Figure 3.4 in the main text, excluding those groups that were not detected in the 16S rRNA gene amplicon dataset. Numbers in parentheses are false discovery rate q-values for T-test comparisons of PL and FL ratios, calculated as in Storey and Tibshirani (2003) (Shown if < 0.1).

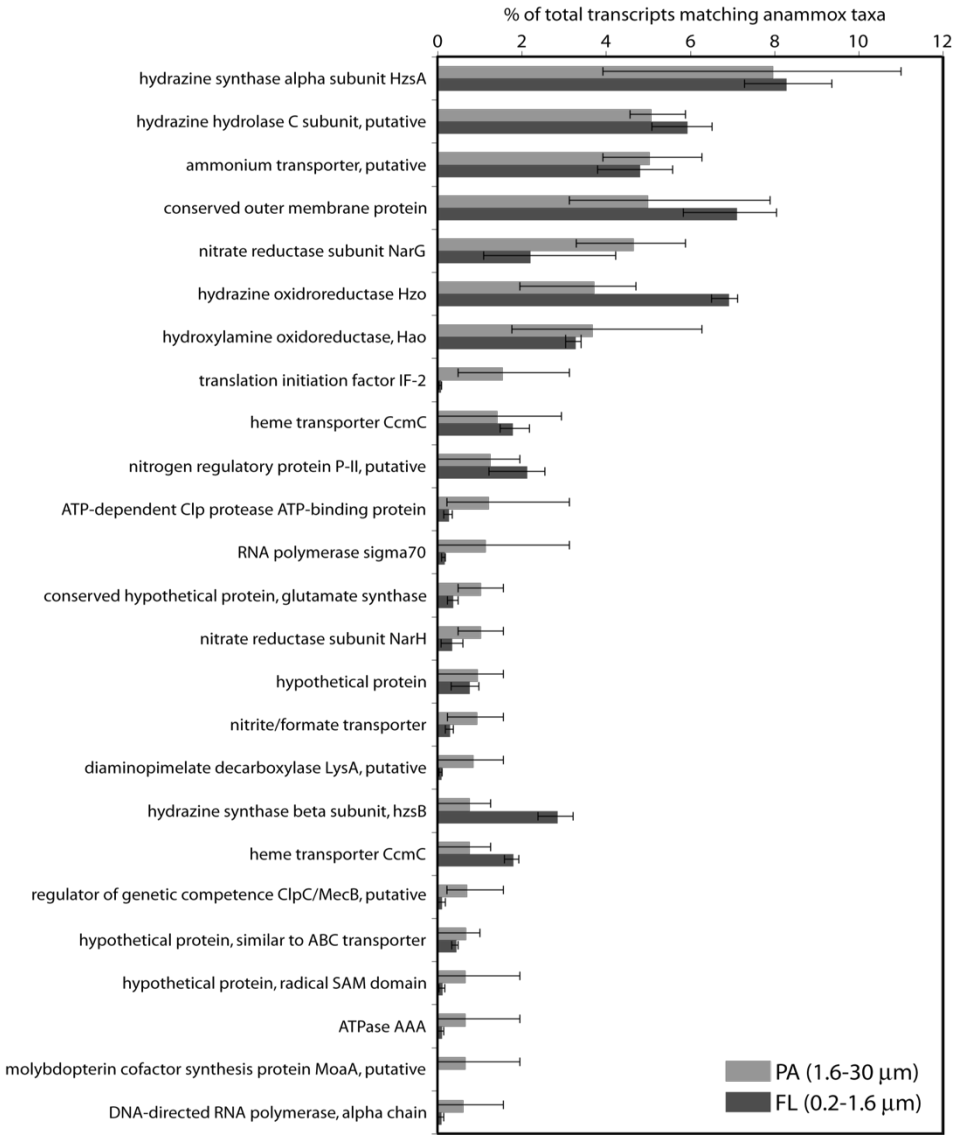


Figure B.5 Size fraction-specific differences in transcripts affiliated with anammox bacteria. Bars are abundances shown as a percentage of the total number of transcripts with top matches (bit score > 50) to genes of known anammox genera (Scalindua, Kuenenia, Brocadia, Jettenia, Anammoxoglobus), as identified by BLASTX against NCBI-nr. Values are averages of the three OMZ depths at which anammox bacteria were most abundant (100, 125, 300m), partitioned according to filter size fraction (FL vs. PA). Thin bars reflect ranges. Only the top 25 most abundant genes are shown, ordered by rank abundance in the PA fraction. PA and FL transcript abundance profiles were generally well correlated (R2 = 0.80), and no genes (out of 1392 total) were significantly differentially expressed (DE) being fractions (P>0.05; FDR- correction; baySeq). Statistical evaluation required a conservative approach of grouping datasets from different depths as replicates; depth-specific variation in expression could therefore confound the detection of DE between PA and FL fractions. Detection likely was also confounded by an overall low representation of anammox- affiliated transcripts in the PA fraction (proportional abundances of transcripts matching anammox taxa were 7- to 58- higher in the FL fraction).

References

Allers E, Wright JJ, Konwar KM, Howes CG, Beneze E, Hallam SJ, Sullivan MB. (2013). Diversity and population structure of Marine Group A bacteria in the Northeast sub-arctic Pacific Ocean. ISME J 7: 256-268.

Beman JM, Carolan MT. (2013). Deoxygenation alters bacterial diversity and community composition in the ocean's largest oxygen minimum zone. Nat Commun 4: 2705.

Borisov SM, Lehner P, Klimant I. (2011). Novel optical trace oxygen sensors based on platinum(II) and palladium(II) complexes with 5,10,15,20-meso-tetrakis-(2,3,4,5,6-pentafluorphenyl)-porphyrin covalently immobilized on silica-gel particles. Analy Chim Acta 690: 108-115.

Braman RS, Hendrix SA. (1989). Nanogram nitrite and nitrate determination in environmental and biological materials by vanadium (III) reduction with chemiluminescence detection. Anal Chem 61: 2715-2718.

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.

Ganesh S, Parris DJ, DeLong EF, Stewart FJ. (2014). Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. ISME J 8: 187-211.

Glaubitz S, Kießlich K, Meeske C, Labrenz M, Jürgens K. (2013). SUP05 dominates the Gammaproteobacterial sulfur oxidizer assemblages in pelagic redoxclines of the central Baltic and Black Seas. Appl Environ Microbiol 79: 2767-2776.

Grasshoff K, Ehrhardt M, Kremling K. (1983). Methods of seawater analysis, 2nd ed. Verlag Chemie.

Hardcastle TJ, Kelly KA. (2010). baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics 11: 422.

Holmes RM, Aminot A, Kerouel R, Hooker A, Peterson BJ. (1999). A simple and precise method for measuring ammonium in marine and freshwater ecosystems. Can J Fish Aquat Sci 56: 1801-1808.

Isenbarger TA, Finney M, Ríos-Velázquez C, Handelsman J, Ruvkun G. (2008). Miniprimer PCR, a new lens for viewing the microbial world. Appl Environ Microbiol 74: 840-849.

Revsbech NP, Larsen LH, Gundersen J, Dalsgaard T, Ulloa O, Thamdrup B. (2009). Determination of ultra-low oxygen concentrations in oxygen minimum zones by the STOX sensor. Limnol Oceanogr-Meth 7: 371-381.

Roder C, Arif C, Bayer T, Aranda M, Daniels C, Shibl A et al. (2014). Bacterial profiling of White Plague Disease in a comparative coral species framework. ISME J 8: 31-39.

Sheik CS, Jain S, Dick GJ. (2014). Metabolic flexibility of enigmatic SAR324 revealed through metagenomics and metatranscriptomics. Environ Microbiol 6: 304-317.

Sorokin DY, Tourova TP, Galinski EA, Muyzer G, Kuenen JG. (2008). Thiohalorhabdus denitrificans gen. nov., sp. nov., an extremely halophilic, sulfur-oxidizing, deep-lineage gammaproteobacterium from hypersaline habitats. Int J Syst Evol Microbiol 58: 2890-2897.

Stevens H, Ulloa O. (2008) Bacterial diversity in the oxygen minimum zone of the eastern tropical South Pacific. Environ Microbiol 10: 1244-1259.

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.

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. 210

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.

Wright JJ, Mewis K, Hanson NW, Konwar KM, Maas KR, Hallam SJ. (2014). Genomic properties of Marine Group A bacteria indicate a role in the marine sulfur cycle. ISME J 8: 455-68.

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.

APPENDIX C

SUPPLEMENTARY MATERIAL FOR CHAPTER 4

STANDARD FILTRATION PRACTICES MAY SIGNIFICANTLY DISTORT PLANKTONIC MICROBIAL DIVERSITY ESTIMATES

Supplementary Tables

Table C.1 Bacterial 16S rRNA gene copies per mL in sample water from experiments 1 and 2. Values are averages across all volume replicates, with standard deviation in parentheses. The ratio of Sterivex to prefilter counts is shown in the last column.

exp	Sterivex (0.2-1.6 µm)	prefilter (>1.6 µm)	combined	ratio (Sterivex/prefilter)
1-150m	3.7×10^5 (1.1×10^5)	3.3×10^4 (1.4×10^4)	4.0×10^5 (1.2×10^5)	11.1
2-400m	6.4×10^4 (2.2×10^4)	6.3×10^3 (1.4×10^3)	7.0×10^4 (2.4×10^4)	10.3

Table C.2 Percentage variation (R2) in weighted UniFrac distances explained by filtered water volume differences, based on adonis tests in QIIME. All P-values are significant following Bonferroni correction for multiple tests.

dataset	R^2, *P-value*
Exp1, prefilter	0.87, *0.01*
Exp1, Sterivex	0.50, *0.001*
Exp2, prefilter	0.74, *0.001*
Exp2, Sterivex	0.81, *0.001*

Table C.3. Abundance of microbial orders in experiment 1, expressed as a % of total 16S rRNA gene amplicons

* average % of replicates
**positive values: fold increase in abundance from the lowest to the highest filtered volumes;
**negative values: fold increase from the highest to the lowest filtered volumes, multiplied by -1
Shading = significant difference in abundance between lowest vs highest volumes (P<0.05; FDR; baySeq)

Sterivex filter fraction (0.2-1.6 μm)

	Filtered water volume*				
Taxon	0.1L	1L	5L	Average (all)	Fold Change**
SAR406, Arctic96B-7	16.2	18.8	25.1	20.0	1.6
Gammaproteobacteria, Oceanospirillales	12.9	11.2	12.5	12.2	-1.0
Alphaproteobacteria, Rickettsiales	11.2	11.5	10.6	11.1	-1.1
Deltaproteobacteria, Sva0853	9.0	10.6	9.1	9.6	1.0
Planctomycetes, Brocadiales	9.0	10.0	8.5	9.2	-1.1
Gammaproteobacteria, Chromatiales	5.3	5.4	4.3	5.0	-1.2
Actinobacteria, Acidimicrobiales	3.0	3.9	3.2	3.4	1.0
SAR406, AB16, ZA3648c	2.1	3.0	3.3	2.8	1.6
Thaumarchaeota, Cenarchaeales	2.7	3.4	2.2	2.8	-1.2
Deltaproteobacteria, Desulfobacterales	2.3	3.2	2.4	2.6	1.0
Gammaproteobacteria, Vibrionales	6.7	0.3	0.1	2.4	-96.5
Euryarchaeota, Thermoplasmata, E2	1.5	2.0	3.2	2.2	2.2
Unassigned	2.1	2.0	1.8	2.0	-1.2
Gammaproteobacteria, Thiotrichales	1.5	1.9	2.5	1.9	1.7
Chloroflexi, SAR202	1.6	2.1	2.0	1.9	1.3
Alphaproteobacteria, Rhodospirillales	1.4	1.8	1.5	1.6	1.0
Gammaproteobacteria, Thiohalorhabdales	1.1	1.4	1.6	1.4	1.4
Alphaproteobacteria, unassigned	0.8	1.2	0.8	0.9	-1.0
Gammaproteobacteria, Alteromonadales	1.5	0.6	0.4	0.8	-3.6
Chloroflexi, Dehalococcoidales	0.6	0.7	0.5	0.6	-1.1
Bacteroidetes, Flavobacteriales	0.4	0.4	0.6	0.5	1.5

Table S3 continued

Prefilter fraction (>1.6 μm)

Taxon	Filtered water volume* 0.1L	1L	5L	Average (all)	Fold Change
Gammaproteobacteria, Vibrionales	54.1	4.6	2.6	17.2	-20.5
Gammaproteobacteria, Thiohalorhabdales	4.4	18.8	17.0	14.1	3.8
Euryarchaeota, Thermoplasmata, E2	1.0	7.9	9.6	6.8	9.8
SAR406, Arctic96B-7	3.1	7.3	6.9	6.0	2.2
Bacteroidetes, Flavobacteriales	1.5	6.7	7.0	5.4	4.7
Gammaproteobacteria, Oceanospirillales	4.7	5.2	5.1	5.0	1.1
Unassigned	1.3	6.7	6.1	4.9	4.6
Gammaproteobacteria, Alteromonadales	7.4	2.8	2.6	4.0	-2.8
Deltaproteobacteria, Myxococcales	0.5	3.6	4.2	3.1	8.4
Planctomycetes, Phycisphaerales	0.2	3.0	3.9	2.7	15.9
SAR406, ZA3648c	0.7	3.1	3.0	2.4	4.1
Gammaproteobacteria, Thiotrichales	0.3	1.8	2.9	1.9	8.7
Gammaproteobacteria, Pseudomonadales	5.2	0.8	0.2	1.7	-21.6
Alphaproteobacteria, Rickettsiales	0.9	2.1	1.9	1.7	2.0
Deltaproteobacteria, unassigned	0.1	1.6	2.0	1.4	27.4
Alphaproteobacteria, Rhodobacterales	3.9	0.2	0.3	1.2	-12.9
Gammaproteobacteria, Chromatiales	0.6	1.3	1.3	1.1	2.0
Planctomycetes, Brocadiales	0.7	1.2	1.2	1.1	1.6
Alphaproteobacteria, Rhodospirillales	0.6	1.1	1.3	1.1	2.2
Planctomycetes, OM190, agg27	0.1	1.0	1.4	0.9	15.5
Actinobacteria, Acidimicrobiales	0.7	1.1	0.8	0.9	1.2
Chloroflexi, SAR202	0.3	0.9	1.0	0.8	3.6
Planctomycetes, OM190,	0.2	1.1	0.9	0.8	4.9
Verrucomicrobia, Verrucomicrobiales	0.1	0.6	1.1	0.7	7.9
Deltaproteobacteria, Desulfobacterales	0.4	1.0	0.7	0.7	2.0
Alphaproteobacteria, Sphingomonadales	2.1	0.2	0.1	0.7	-30.4
Thaumarchaeota, Cenarchaeales	0.3	0.7	0.7	0.6	2.6
Epsilonproteobacteria, Campylobacterales	0.2	0.7	0.7	0.6	3.1
Verrucomicrobia, Arctic97B-4	0.2	0.5	0.7	0.5	3.6
Deltaproteobacteria, Sva0853	0.4	0.4	0.7	0.5	1.8
Deltaproteobacteria, PB19	0.1	0.8	0.5	0.5	6.9
Gammaproteobacteria, Legionellales	0.2	0.6	0.5	0.5	2.8

Table C.4 Abundance of microbial orders in experiment 2, expressed as a % of total 16S rRNA gene amplicons.

* average % of replicates
**positive values: fold increase in abundance from the lowest to the highest filtered volumes;
**negative values: fold increase from the highest to the lowest filtered volumes, multiplied by -1
Shading = significant difference in abundance between lowest vs highest volumes (P<0.05; FDR; baySeq)

Sterivex filter fraction (0.2-1.6 μm)

	Filtered water volume*							
Taxon	0.05L	0.1L	0.5L	1L	2L	5L	Average (all)	Fold Change**
SAR406, Arctic96B-7	12.3	14.4	21.2	19.7	24.2	26.0	19.7	2.1
Planctomycetes, Brocadiales	16.0	11.3	21.7	19.7	20.0	15.6	17.4	-1.0
Unassigned	4.4	4.6	6.9	6.5	7.4	8.8	6.4	2.0
Alphaproteobacteria, Rickettsiales	4.5	3.9	7.6	7.0	7.5	6.7	6.2	1.5
Gammaproteobacteria, Vibrionales	22.0	7.8	2.4	0.6	0.2	0.4	5.6	-51.0
Gammaproteobacteria, Chromatiales	1.3	7.0	2.3	12.3	1.4	2.3	4.4	1.8
Deltaproteobacteria, Sva0853	3.6	3.1	5.4	4.9	5.1	4.3	4.4	1.2
Chloroflexi, SAR202	2.0	1.9	3.2	3.0	4.4	4.2	3.1	2.1
Actinobacteria, Acidimicrobiales	2.4	2.4	3.2	2.8	3.2	3.0	2.8	1.2
Gammaproteobacteria, Oceanospirillales	2.6	2.4	2.7	2.4	2.6	2.7	2.6	1.0
Alphaproteobacteria, Rhodospirillales	2.1	1.9	2.8	2.6	3.0	2.6	2.5	1.3
Gammaproteobacteria, Alteromonadales	4.9	4.6	1.2	0.5	0.4	0.3	2.0	-15.6
SAR406, ZA3648c	1.0	1.1	2.1	2.0	2.3	2.4	1.8	2.3
Chloroflexi, Dehalococcoidales	1.0	1.0	1.5	1.4	2.1	1.8	1.5	1.7
Euryarchaeota, Thermoplasmata, E2	0.6	1.1	1.1	1.1	1.9	2.8	1.4	4.4
Crenarchaeota, MBGA	1.2	1.0	1.7	1.4	1.7	1.5	1.4	1.3
Gammaproteobacteria, Thiohalorhabdales	0.9	1.2	1.2	1.0	1.4	2.4	1.4	2.8
Alphaproteobacteria, Kiloniellales	3.3	4.0	0.2	0.1	0.0	0.0	1.3	-93.6
Gammaproteobacteria, Thiotrichales	0.7	0.8	1.3	1.1	1.3	1.5	1.1	2.0
Thaumarchaeota, Cenarchaeales	1.0	0.5	1.3	1.3	1.3	1.2	1.1	1.2
Deltaproteobacteria, Desulfobacterales	0.9	1.1	1.4	1.1	1.1	1.0	1.1	1.2
Alphaproteobacteria, Rhodobacterales	4.1	1.4	0.3	0.1	0.1	0.0	1.0	-85.7
Bacteroidetes, Flavobacteriales	0.2	2.8	0.4	0.4	0.3	0.9	0.8	3.7
Planctomycetes, Planctomycetales	0.0	4.5	0.0	0.1	0.0	0.0	0.8	-11.4
Deltaproteobacteria, Entotheonellales	0.6	0.6	1.0	0.7	0.9	0.8	0.8	1.2
Acidobacteria, iii1-15	0.6	0.4	0.7	0.6	0.8	0.7	0.6	1.2
Betaproteobacteria, Methylophilales	0.0	3.8	0.0	0.0	0.0	0.0	0.6	-1.2
Chloroflexi, H39	0.2	0.3	0.5	0.6	0.6	0.7	0.5	3.5

Table S4 continued

Prefilter fraction (>1.6 μm)

Taxon	Filtered water volume*						Average (all)	Fold Change
	0.05L	0.1L	0.5L	1L	2L	5L		
Gammaproteobacteria, Vibrionales	62.2	49.6	38.5	4.7	2.6	1.9	26.6	-33.4
Gammaproteobacteria, Chromatiales	2.2	1.4	2.8	27.9	58.1	38.7	21.8	17.6
Gammaproteobacteria, Alteromonadales	8.1	8.8	7.5	2.3	1.6	1.8	5.0	-4.5
SAR406, Arctic96B-7	0.9	4.2	7.0	7.0	4.2	6.6	5.0	7.7
Gammaproteobacteria, Oceanospirillales	7.5	7.9	4.4	3.4	2.0	1.7	4.5	-4.4
Euryarchaeota, Thermoplasmata, E2	0.4	2.0	4.5	6.2	4.2	7.1	4.1	19.5
Unassigned	0.6	2.6	5.3	5.9	3.3	4.4	3.7	7.1
Gammaproteobacteria, Thiohalorhabdales	0.6	1.4	3.3	7.5	2.6	4.6	3.3	7.6
Firmicutes, Bacillales	5.1	4.3	1.0	0.3	0.0	0.5	1.9	-10.2
Epsilonproteobacteria, Campylobacterales	0.1	0.2	0.9	2.7	4.2	1.9	1.7	20.0
Alphaproteobacteria, Rickettsiales	0.4	1.1	2.5	2.3	1.4	2.0	1.6	4.7
Bacteroidetes, Flavobacteriales	0.3	1.0	1.5	3.0	1.5	1.4	1.5	4.2
Planctomycetes, Brocadiales	0.6	1.6	2.4	1.6	1.2	1.3	1.4	2.3
Gammaproteobacteria, Thiotrichales	0.0	0.5	1.3	2.4	1.4	2.5	1.3	65.0
Alphaproteobacteria, Rhodobacterales	3.2	2.4	0.7	0.1	0.2	0.1	1.1	-49.0
Alphaproteobacteria, Rhodospirillales	0.3	0.8	1.5	1.9	0.6	1.2	1.0	4.5
SAR406, ZA3648c	0.2	0.6	1.3	2.4	0.6	0.9	1.0	5.2
Chloroflexi, SAR202,	0.2	0.7	1.6	1.3	0.5	0.9	0.9	4.1
Deltaproteobacteria, Myxococcales	0.2	0.4	0.5	1.4	0.5	1.4	0.8	8.5
Planctomycetes, Phycisphaerales	0.2	0.5	0.4	1.9	0.2	1.1	0.7	6.3
Actinobacteria, Acidimicrobiales	0.2	0.3	1.2	0.8	0.7	0.8	0.7	4.8
Alphaproteobacteria, Kiloniellales	1.6	1.3	0.3	0.0	0.0	0.1	0.6	-23.2
Alphaproteobacteria, Sphingomonadales	1.0	0.7	0.0	0.0	0.0	1.2	0.5	1.2
Bacteroidetes, Bacteroidales	0.1	0.1	0.4	0.4	0.8	0.9	0.5	10.9

Supplementary Figures

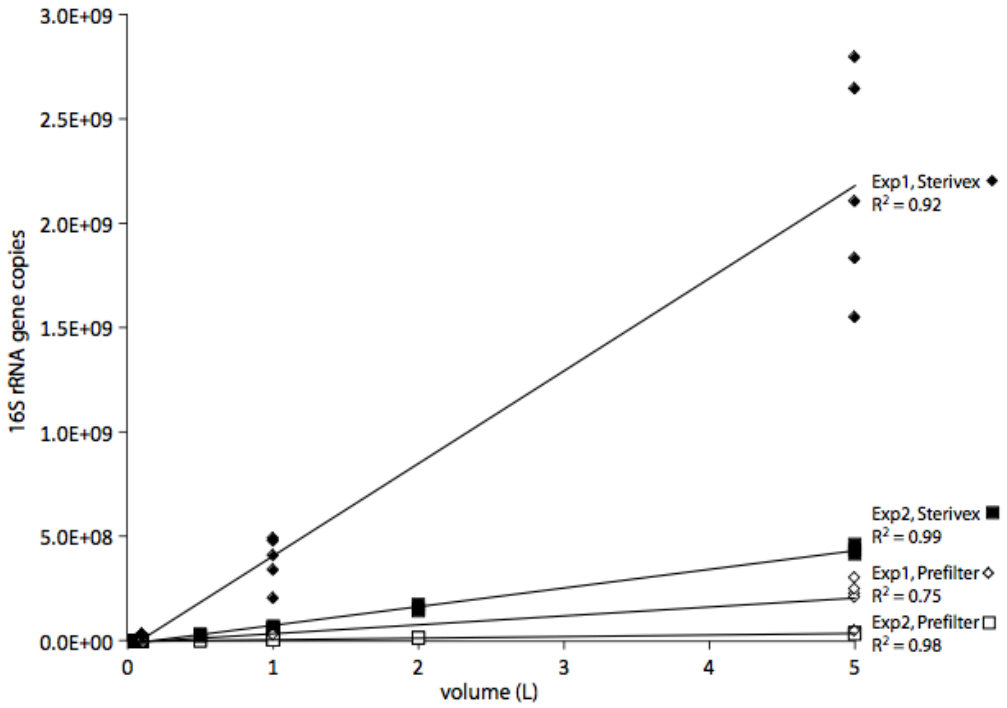


Figure C.1 Total bacterial 16S rRNA gene counts as a function of filtered water volume.

VITA

SANGITA GANESH

Originally from Chennnai, India, Sangita earned her Bachelor's degree in Biotechnology from SASTRA University in Tanjore, India, in 2006. Her passion for Molecular Biology research led her to the Ph.D program at Georgia Institute of Technology, where she pursued research in the fields of Molecular Biology and Genomics in Dr. Frank Stewart's lab.
Below is a list of all of Sangita's publications:

- **Ganesh S**, Bristow LA, Larsen M, Sarode N, Thamdrup B, Stewart FJ, Size-fraction partitioning of community gene transcription and rates of nitrogen metabolism in a marine oxygen minimum zone. *ISME J*. 9: 2682-2696, **2015**
- **Ganesh S**, Parris DJ, DeLong EF, Stewart FJ, Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. *ISME J*. 8: 187-211, **2014**
- Padilla C*, **Ganesh S***, Gantt S, Huhman A, Parris DJ, Sarode N, Stewart FJ, Standard filtration practices significantly distort planktonic microbial diversity estimates. (*Equal contribution) *Frontiers in Microbiology*. 6:547, **2015**
- Parris DJ, **Ganesh S**, Edgcomb V, DeLong EF, Stewart FJ, Microbial eukaryote diversity in the marine oxygen minimum zone off northern Chile. *Frontiers in Microbiology*. 5:543, **2014**
- Sarode N, Parris DJ, **Ganesh S**, Seston SL, Stewart FJ, Generation and analysis of microbial metatranscriptomes. *In Yates M, Nakatsu C, Miller R, Pillai S (ed) Manual of Environmental Microbiology*, Fourth Edition. ASM Press, Washington, DC, **2015**
- Glass JB, Kretz CB, **Ganesh S**, Ranjan P, Seston SL, Buck KN, Landing WM, Morton PL, Moffett JW, Giovannoni SJ, Vergin KL and Stewart FJ, Meta-omic signatures of microbial metal and nitrogen cycling in marine oxygen minimum zones. *Frontiers in Microbiology*. 6:998, **2015**
- Seston, SL. Beinart RA, Sarode N, Shockey AC, Ranjan P, **Ganesh S**, Girguis PR, Stewart FJ. Metatranscriptional response of chemoautotrophic Ifremeria nautilei endosymbionts to differing sulfur regimes. *Frontiers in Microbiology* 7, **2016**