UNIVERSITY OF CALIFORNIA, MERCED

Quantifying distributions of and modeling interactions among sulfur- and nitrogen-cycling chemolithoautotrophs in the largest oxygen minimum zone of the global ocean

A thesis submitted in partial satisfaction for the requirements for the degree of Masters of Science

in

Quantitative and System Biology

by

Molly Theresa Carolan

Committee in charge:

Professor J. Michael Beman, Chair
Professor Michael Colvin
Professor Andy LiWang

The Thesis of Molly Theresa Carolan is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Dr. Andy LiWang

Dr. Michael Colvin

Dr. J. Michael Beman Chair

University of California, Merced
2014

Table of Contents

Acknowledgements

Figure 1 is used with permission by J. Michael Beman.

I would like to express my deep gratitude to my thesis advisor, J. Michael Beman, for his support and advice. Dr. Michael Colvin provided me with valuable advice and technical expertise, in addition to serving on this thesis committee. I would also like to express my great appreciation for Dr. Andy LiWang's service on this committee. Special thanks to Jesse Wilson and Curtis Hayden for their collegiality and comraderie. And, finally, I would like to thank Jesse Edelstein for his generous assistance and quick wit.

Figures and Tables

An abstract of

Quantifying distributions of and modeling interactions among sulfur- and nitrogen-cycling chemolithoautotrophs in the largest oxygen minimum zone of the global ocean

by

Molly Theresa Carolan

A thesis submitted in partial satisfaction for the requirements for the degree of Masters of Science

Anoxic marine microbe communities from the Eastern Tropical North Pacific were analyzed using culture-independent molecular methods. Community composition and associations were determined with pyrosequencing and phylotype analysis, followed by species distribution modeling and exploratory statistics. Real-time quantitative polymerase chain reaction (PCR) assays were used to quantify activity and distribution of key sulfur- and nitrogen-cycling groups within the oxygen minimum zone (OMZ). Strong correlations were found between putative sulfur-oxidizing groups and known nitrite-oxidizing bacteria. Quantitative PCR assays confirmed the activity of sulfate-reducing genes, strong evidence for cryptic sulfur cycling.

Introduction

Large areas of the ocean are characterized by extremely low oxygen concentrations. These naturally-occurring oxygen minimum zones (OMZs) are defined as regions of pelagic ocean where oxygen concentrations are depleted below 20 µmol kg^{-1} due to microbial respiration (Paulmier & Ruiz-Pino 2009, Gilly *et al*. 2013, Helly & Levin 2004). OMZs are often associated with upwelling along the west coasts of continents and near the equator, and the attendant high levels of microbial respiration that occur in these regions. OMZs are dominated by microbial communities, as much of the macrofauna of the pelagic ocean is unable to live in hypoxic conditions (Diaz and Rosenberg 2008, Gilly *et al*. 2008). At the same time, the total volume of OMZ waters is growing, their upper boundaries are vertically shoaling, and the degree of anoxia is intensifying within the core of the OMZs (Gilly *et al*. 2013, Stramma *et al*. 2008). These changes are collectively referred to as 'ocean deoxygenation' and are driven primarily by rising ocean temperatures that reduce oxygen solubility and increase water column stratification—both of which reduce oxygen at depth within the ocean (Keeling *et al*. 2010). Nitrogen deposition and warming may also stimulate primary production in surface waters, further decreasing available oxygen (Gilly *et al*. 2013).

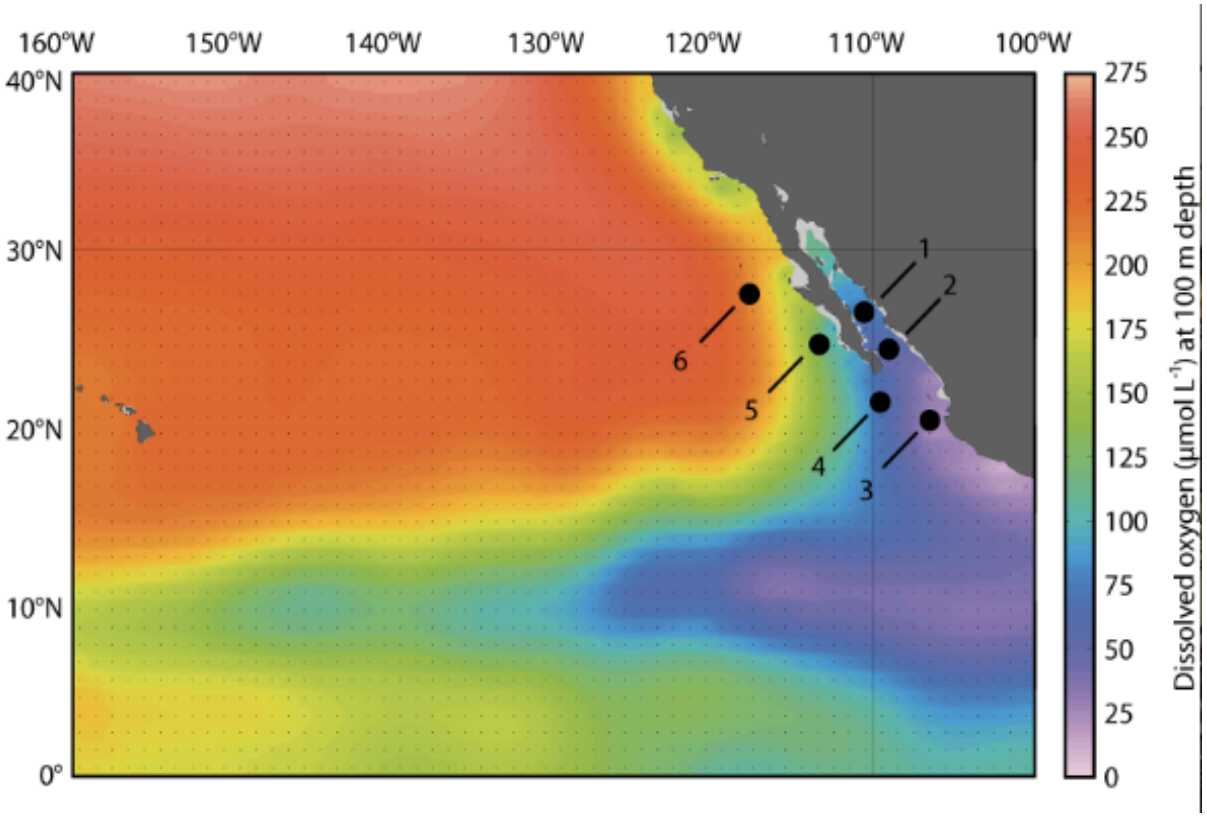

Figure 1: Sampling locations within the Gulf of California and the greater Eastern Tropical North Pacific. The scale on the right indicates oxygen concentration at 100m depth. Adapted from Beman and Carolan 2013.

Warming ocean temperatures and nutrient run-off are expected to expand OMZs world-wide over the next century, but the effects on microbial processes and communities are difficult to forecast. OMZs are of particular concern in quantifying the role of microbial assemblages in global elemental cycles, as they provide for a distinct redox environment for microbial metabolism. This leads to active cycling of carbon, nitrogen, and sulfur by aerobic and anaerobic microbial processes. This study is primarily concerned with identifying key members of the microbial assemblages involved in chemolithoautotrophy utilizing sulfur and nitrogen compounds within the Eastern Tropical North Pacific OMZ, the largest low oxygen water mass in the global ocean.

Sulfur and Nitrogen Cycling within OMZs

Microbes are the vanguards of metabolic diversity on earth, controlling global cycling of essential elements of life (C, H, O, N, and S) (Falkowski 2008). Sulfur metabolism is profoundly ancient, and genes related to sulfur oxidation and reduction are dispersed throughout the Bacterial and Archaeal domains (Ghosh and Dam 2009). These lineages are distributed throughout a wide range of habitats, from marine sediments to eutrophic ponds, animal guts and volcanic hot springs, where sulfur exists in oxidation states from -2 to +6, producing complex cycling interactions within the S cycle and with the carbon and nitrogen cycles (Muyzer and Stams 2008). Thiosulfate ($S_2O_3^{2-}$) is an abundant substrate for oxidation by sulfur lithotrophs, and *soxB*, a gene that encodes for part of a thiosulfate-oxidizing enzyme complex, is found in *Chlorobi, Chloroflexi, Spirochaeta*, and *Alpha-, Beta-, Gamma-,* and *Epsilonproteobacteria* lineages (Meyer *et al*. 2007, Ghosh and Dam 2009). In contrast to the very broad distribution of the *sox* multi-enzyme complex, dissimilatory sulfite reductase (*dsr*) is found within select lineages within the *Firmicutes*, *Deltaproteobacteria*, and *Archaea*. This gene is involved in S reduction conducted by these groups under anaerobic conditions. The related, but reverse-acting, dissimilatory sulfite reductase (*rdsr*) is found within *Chlorobi*, *Alpha-*, *Beta-*, and *Gammaproteobacteria* (Loy *et al*. 2009), and *rdsr* acts to oxidize sulfur species. The dissimilatory sulfite reductases are known to act on sulfite, sulfate, and elemental sulfur, allowing for diverse sulfur metabolism (Ghosh and Dam 2009). Adding another layer of complexity, sulfur-oxidizing bacteria can also be photo- or chemolithotrophic. *PufM* is a gene that encodes for a protein in the light-harvesting reaction center of anoxygenic photosynthetic bacteria, a group of photolithotrophic and photoheterotrophic *Alpha-*, *Beta-*, and *Gammaproteobacteria,* that are found in low-oxygen, euphotic, estuarine and marine habitats (Waidner and Kirchman 2008). This study uses these genes, as well as select 16S rRNA genes, to determine the distribution and relative abundance of sulfur-utilizing bacteria across the oxycline of the Eastern Tropical North Pacific oxygen minimum zone by quantitative PCR (qPCR).

Until recently, S-cycling was thought to occur in a relatively limited capacity in the ocean. Sulfate reduction was known to occur in anoxic basins such as the Black Sea and within ocean sediments (Jorgensen *et al*. 1991, Canfield 1989), but recent work indicates that a 'cryptic S cycle' is active in the ETSP OMZ (Canfield *et al*. 2010). Here S is rapidly reduced and oxidized by multiple bacterial groups; the cycling is so rapid that it is not easily detectable based on chemical measurements, but measuring gene expression

provides a way to track S cycling. Whether this occurs in other OMZs is unknown, but active S-cycling has several implications for ocean biology and biogeochemistry, including the fact that hydrogen sulfide is toxic to large organisms (Lavik *et al*. 2009), and that this alters competitive and symbiotic interactions among microbial groups and processes.

Interactions between S cycling and oceanic nitrogen cycling are of particular importance: as the fourth most-abundant element in biological macromolecules, nitrogen is an essential nutrient, and in its inert diatomic molecular form it makes up 78% of Earth's atmosphere. Biologically active nitrogen exists at oxidation states of +5 (nitrate) to -3 (ammonium) and is actively cycled by microorganisms (Francis *et al*. 2007). Prior to the invention of the Haber-Bosch process in 1913, biological nitrogen fixation was the major source of nitrogen to ecosystems, but anthropogenically produced reactive nitrogen has more than doubled the overall amount of available nitrogen (Gruber & Galloway 2008). However, marine productivity in large regions of the ocean is nitrogen-limited through the complex interplay with phosphorous, trace metals, and active N removal under low oxygen conditions found in OMZs (Zehr and Kudela 2011, Deutsch *et al*. 2007, Francis *et al*. 2005). Since the 1990s, canonical denitrification and anaerobic ammonium oxidation (anammox) in low-oxygen marine waters have received particular attention for two reasons. First, these processes represent the major sink for reactive nitrogen as they are the only ways it is returned to the atmosphere, and secondly, denitrification can produce and release intermediates such as nitrous oxide, a greenhouse gas that also destroys stratospheric ozone. Within an OMZ, canonical denitrification, anammox, nitrification (ammonium ->nitrite->nitrate), and dissimilatory nitrate reduction to ammonium (DNRA) can all occur simultaneously (Zehr and Kudela 2011). At present, the relative dominance of these different processes and their responses to changes in N forms and flux, carbon supply, deoxygenation, and other forms of environmental variability are poorly constrained (Francis *et al*. 2007, Lam *et al*. 2009, Ward *et al*. 2009, Zehr and Kudela 2011, Lam and Kuypers *et al*. 2012).

This effort to quantify the nitrogen and carbon budgets of the dark ocean has led to increased awareness of the microbial communities that drive these processes. Recent studies have discovered abundant Bacterial lineages (*Gammaproteobacteria SUP05* and *Arctic96BD-19* and *Deltaproteobacterium SAR324*) in suboxic and anoxic waters that are now believed to have chemolithoautotrophic metabolisms that utilize sulfide and thiosulfate, as well as oxidized nitrogen species (Walsh *et al*. 2009, Swan *et al*. 2011, Canfield *et al*. 2010). Despite making up large percentages of OMZ microbial communities, the biogeochemical roles of these groups have only been identified in the past few years, and their ecology and biogeochemical activity are still ill-defined. The oxidation of sulfur using nitrogen (nitrate or nitrite) is known as 'chemoautotrophic denitrification' and is distinct from conventional, heterotrophic denitrification (oxidation of organic carbon using nitrate or nitrite). *SUP05*, *Arctic96BD-19*, and *SAR324* are known to have genes for *RuBisCo* and *rdsr*, and in some cases may have *dsr* genes as well (Swan *et al*. 2011). Yet the degree to which these organisms use reactive nitrogen species as part of their metabolisms and their relative roles in the sulfur cycle are not well

known. These broader groups are each made up many operational taxonomic units (OTUs)—or bacterial 'species'—that are heterogeneously distributed within OMZs (Wright *et al*. 2012). Whether these represent different ecotypes with distinct biogeochemical roles and ecological niches is also unknown.

The flexible metabolism of the putative chemolithoautotrophic sulfide-oxidizers requires that the functional metabolism approach using genes of interest be paired with a broad survey of Bacterial diversity within the OMZ. This was accomplished using next-gen sequencing of 16S rRNA genes, where diversity and relative abundance of bacterial groups can be compared to the qPCR abundances and analyzed using exploratory statistics tools, in this case Maximal Information-based Nonparametric Exploration (MINE) (Reshef *et al*. 2011). To model the distribution of these organisms based on environmental data—particular oxygen concentrations—I also used species distribution models (SDMs) adapted from the ecological literature (Elith and Leathwick 2009). SDMs have only recently been applied to surface microbial communities (Larsen *et al*. 2012, Ladau *et al*. 2013), and this is their first application to OMZ bacterial communities.

Materials and Methods

Sample Collection

Samples were collected aboard the *R/V New Horizon* in the Gulf of California (GOC) and eastern tropical North Pacific (ETNP) in July and August of 2008. Six stations were occupied for several days each, and temperature, conductivity, cholorphyll, and dissolved oxygen concentrations were measured using a Seabird SBE 9 conductivity-temperature-depth (CTD) sensor package equipped with a Seapoint fluorometer and SBE oxygen sensor. Following initial test casts, samples were collected using 10L bottles deployed on the CTD rosette; samples for dissolved oxygen and nutrient measurements were analyzed within hours of collection to characterize the water column structure. Oxygen concentrations measured using the SBE oxygen sensor were corrected based on Winkler titrations ($r^2 = 0.997$, $n = 187$ for the cruise). DNA/RNA sample collection depths were selected primarily based on oxygen and nutrient profiles: one to three depths were selected straddling the base of the euphotic zone, followed by five to seven depths spaced every 5m, and then four to six depths spaced every 10-20 m across the transition from hypoxic to suboxic conditions in the OMZ.

DNA and RNA extraction

At each depth, 4L of seawater were collected from CTD rosette and divided into two sets of 2L samples, which were then filtered through separate 25 mm diameter 0.2 μm Suppor filters (Pall Corporation, Port Washington, NY, USA) using a peristaltic pump. Filters were flash frozen in liquid nitrogen and stored at -80°C until DNA and RNA extraction, with one set of filters dedicated to each. Details of DNA extraction were reported in Beman et al.[22]. Filters for RNA extraction were frozen in RLT buffer with beta-mercaptoethanol added (Qiagen, Valencia, CA, USA), and RNA was extracted using the Qiagen RNeasy kit, following Church et al.[42], with the following modifications: tubes

containing filters, glass beads, and buffer were first agitated for 80 seconds on a FastPrep machine (MP Biomedicals, Solon, OH, USA) at setting 5.5; each sample and an equal volume of 70% ethanol solution was then bound to the RNeasy spin column, and purified and eluted following the manufacturer's instructions. We used Turbo DNase (Ambion, Life Technologies Corporation, Carlsbad, CA, USA) to remove carry-over DNA; samples were treated with 10% DNase buffer and 5% DNase for 20 min at 37°C.

cDNA was generated from extracted RNA using the Invitrogen SuperScript III Reverse Transcriptase kit (Life Technologies Corporation, Carlsbad, CA, USA) following the manufacturer's instructions. 10µL of RNA extract, 1µL each of random hexamers and 1µL 10mM dNTPs were incubated at 65°C for 5 min, and then placed on ice for 1 min; samples were then incubated with 10µL of cDNA synthesis mix (2µL 10x RT buffer, 4µL 25mM $MgCl_2$, 2µL 0.1M DTT, 1µL RNase OUT (40 U μL^{-1}), and 1µL SuperScript III RT) for 25°C for 10 min, then 55°C for 50 min, and the RT reaction was terminated at 85°C for 5 min. Remaining RNA was removed through incubation with RNase H (Invitrogen) at 37°C for 20 min.

Quantitative PCR

Quantitative PCR (qPCR) assays were performed on a Stragene MX3005P (Agilent Technologies, LA Jolla, CA, USA) using the following reaction chemistry: 12.5µL SYBR Premix F (Epicenter Biotechnologies, Madison, WI, USA), 2 µL $MgCl_2$, 1 µL of each primer, 0.25 µL AmpliTaq polymerase (applied Biosystems, Life Technologies Corporation, Carlsbad, CA, USA), 40 ng μL^{-1} BSA, and 1 ng DNA in a final volume of 25 µL.

16S rRNA genes from *SUP05* were amplified with the primers ba519F, a universal bacterial primer (5'-CAGCMGCCGCGGTAANWC-3'), and SUP051048R (5'-CCATCTCTGGAAAGTTCCGTCT-3')(Zaikova *et al*. 2010). 16S rRNA genes from *Chromatiales* were amplified with the primers CHR986f (5'-AGCCCTTGACATCCTCGGAA-3') from Coolen and Overmann and universal bacterial primer 1392R (5'-ACGGGCGGTGTGTAC-3'). Samples were screened for the presence of purple sulfur and nonsulfur bacteria with primers that target M subunit of their photosynthetic reaction center, *pufM*557F (5'-CGCACCTGGACTGGAC-3') and *pufM*750R (5'-CCCATGGTCCAGCGCCAGAA-3') following Achenbach *et al*. 2001. *Dissimilatory sulfite reductase* subunit A was amplified using the dsr1F (5'-ACSCACTGGAAGCACGGCGG-3') and dsrR (5'-GTGGMRCCGTGCAKRTTGG-3') following Kondo *et al*. 2004.

Table 1: Reaction conditions for qPCR detection of chemolithotroph genes

Primer/GOI	Denaturation	Annealing	Extension	Source
SUP05	30sec @ 95°C	30sec @ 55°C	60sec @ 72°C	Zaikova *et al* 2010
dsrA	30sec @ 95°C	60sec @ 53°C	120sec @ 72°C	Kondo *et al* 2004
Chromatiales	30sec @ 95°C	60sec @ 53°C	120sec @ 72°C	Coolen & Overmann 1999
pufM	30sec @ 95°C	60sec @ 53°C	120sec @ 72°C	Achenbach *et al* 2001

Source denotes the origin of the primer set, but not necessarily the reaction conditions as those have been altered to optimize product efficiency.

For *SUP05, Chromatiales*, and *dsrA*, gene copy numbers were quantified using DNA standards for each primer set. A serial dilution of a short DNA sequences specific to the primer set was used to calculate a standard curve and PCR efficiency. For the *SUP05* assay, PCR efficiency was 99.17% and the r^2 of the standard curve was 0.947. The *dsrA* assay's efficiency was 121% and the r^2 was 0.96, while the *Chromatiales* assay's efficiency was 134% with an r^2 of 0.997. Efficiency ensures that the PCR reaction was reliably doubling the sequence of interest, and for quantification, efficiencies between 95% and 125% are considered ideal. The unusually high efficiency for the *Chromatiales* assay is likely due to the use of a universal bacterial primer, increasing the likelihood of non-specific amplification.

Cycling conditions for the *Chromatiales*, *pufM*, and *dsrA* were as follows: 4 min at 95°C, followed by 40 cycles starting with 30 seconds at 95°C, 1 min at 53°C, 2 min at 72°C, and a 7 second detection step at 77°C. For *SUP05*, cycle conditions were modified from Zaikova *et al* as follows: 3 min at 95°C before 40 cycles of 95°C for 30 seconds, 30 seconds at 55°C, and 1 min at 72°C. The annealing temperature was lowered from 63°C because those reaction conditions produced significant non-specific binding and PCR-products. See Table 1, above, for a summary.

Mothur methods

Sequences were checked for quality and length, removing sequences that did not match the primer or barcode precisely or had long (>8) homopolymer runs. At this stage, sequence libraries were normalized to 3000 sequences for each sample. Sequences were then screened to remove those less than 350bp long and over 500bp long. Sequences were then aligned using the silva SEED database, optimized so that 90% of sequences begin at the same position within the alignment, filtered to remove uninformative columns, and then classified again using the silva SEED database. A pre-clustering algorithm was then used to remove possible bias from sequencing errors by joining sequences with only 1bp difference. A distance matrix was computed using the filtered fasta file with a cutoff for nearest neighbor joining of 13% sequence identity. The sequences were then clustered into operational taxonomic units (OTUs) again using a nearest neighbor algorithm. This produced 6986 OTUs, which were then phylogentically classified again using Silva's SEED database and Mothur.

MINE

The absence, presence, and abundance of each OTU for every sample was then calculated and associated in a data matrix with the following environmental data: photosynthetically active radiation, water density, dissolved oxygen, ammonium concentration, chlorophyll concentration, nitrite concentration, salinity, temperature, radiation transmission, ammonium oxidation rate, and nitrite oxidation rate. To reduce the number of correlation calculations, only OTUs that represent more than 2 sequences were included in the matrix. The average number of members per OTU was 19.7, with a range of 1609 to 0, and a mode of 3. Estimated p-values for Maximal Information Criteria scores were provided by the authors (http://www.exploredata.net/Technical-information). To correct for multiple comparisons, fdr tool (v. 1.2.0, http://cran.r-project.org/web/packages/fdrtool/index.html) was used in the R statistics package (http://cran.r-project.org/) (Strimmer 2008). This resulted in a false discovery rate of 25.8%, and approximately 3500 significant MIC scores.

Microbial Community Modeling

Since the advent of metagenomic and other culture independent methods for studying microbial life, a new uncultureable microbial majority has emerged (Rappe & Giovannoni 2003). Community structure and microbial assemblages have gained traction as a way to understand their ecological impact, functional roles, and biogeography (Fuhrman 2009, Martiny *et al*. 2006, Faust & Raes 2012). In ocean systems, rare microbial groups tend to be endemic, and a few highly abundant groups appear to be cosmopolitan (Fuhrman 2009). What remains unknown, then, is whether the differences in community composition result in significantly different metabolic function. In soil communities, natural gradients, common garden, transplant, and manipulation experiments have been used to explore functional significance (Strickland *et al* 2009, Hansel *et al*. 2008, Lagomarsino *et al*. 2007). Controlled experiments and transplants are not generally viable in the ocean, so natural gradients and time series data predominate.

High-throughput sequencing, like the pyrosequencing described in the methods above, has become increasingly common, resulting in an abundance of microbial sequence data from select regions of the global ocean. OMZs are under-represented, particularly given their importance. Parsing this information to model complex systems, particularly to understand how community structure relates to functional metabolism, has also lagged data generation (Fuhrman 2009). In this work, we first scored pairwise relationships between microbial taxa and environmental variables in order to determine co-occurrence and exclusion patterns for both abiotic and biotic interactions. Then, network inference is used to find the relationships between the components of the system (taxa and the environment). Most pairwise, or similarity-based, measures are limited in the complexity of the relationships they can describe; however, mutual information criteria (MIC), described above, has the benefit of detecting more than just linear relationships and takes into account mutual absences (Faust & Raes 2012).

Results and Discussion

Biogeochemical patterns in the ETNP OMZ

Six stations spanning a range of oceanographic conditions were sampled in the Gulf of California and ETNP in 2008 (Beman et al. 2012, 2013, Beman and Carolan 2013, White et al. 2012). Here we focus on 4 stations that have a well-developed OMZ*:* station 2 is at the edge of the OMZ at the mouth of the Gulf of California, while station 3 is located south of station 2 and was the most southerly station sampled; stations 4 and 5 are further west and north, off the coast of Baja California. At stations 2 and 3, low (<20uM) oxygen concentrations were found at depths of 120m-300m. Nitrite levels were also very high, ranging from 102 to 4175 nM at station 2, and between 299 and 3634nM at station 3. In contrast, at stations 4 and 5, the suboxic waters had a smaller depth range, and lower dissolved nitrite compared with stations 2 and 3. Station 4 had low oxygen at depths of 140 to 200m, and the nitrite concentration did not rise above 47 nM. At station 5, suboxic waters were only found starting at 200m, and the nitrite concentration did not rise above 25 nM.

The accumulation of dissolved nitrite is commonly observed in OMZ regions and is a likely result of microbial nitrate reduction to nitrite under anaerobic conditions. However, a recent review by Ulloa *et al* distinguishes between typical OMZs and what they refer to as anoxic marine zones (AMZs), based on increased nitrite concentrations found in the latter. As discussed above, low oxygen waters allow for denitrification and anammox to proceed, but while it was previously thought that sulfate would only be exploited when oxidized nitrogen species were depleted, recent evidence from the Eastern Tropical South Pacific OMZ indicates that a cryptic sulfur cycle may occur within the nitrite maxima (Canfield 2006, Ulloa *et al*. 2012). Put another way, the presence of nitrite is now thought to be diagnostic of active S cycling in AMZs. If this is the case, our biogeochemical data indicate that S cycling would be most active at stations 2 and 8, and we tested for microbial activity using molecular techniques.

Quantitative PCR of S-cycling groups

Quantitative PCR allows for assays that target one gene of interest (GOI), so we targeted functional genes (*dsrA, rdsrA, soxB,* and *pufM*) and 16S ribosomal RNA genes from sulfur-metabolizing organisms. We applied qPCR both to DNA extractions and cDNA generated from extracted RNA. Amplification of particular genes in DNA indicates the presence of a particular groups or metabolism, whereas amplification in cDNA indicates that a particular group is actively growing or a particular gene is actively expressed. Our previous sequencing data indicated relatively high abundances of *Chromatiales* (Beman and Carolan 2013), or purple-sulfur bacteria, a group capable of chemotrophic growth or photosynthetic growth under anoxic conditions (Imhoff 2005). *Chromatiales* are typically found in a narrow band where both sufficient amounts of sulfide and light are found in other anoxic marine systems, as sulfide and light tend to occur at opposing gradients (Imhoff 2001, Overmann 2001). We used the *CHR986* assay to amplify *Chromatiales* 16S rRNA and examine the distribution of these organisms with depth and from station to

station. We also used the *pufM* assay to target the photosynthetic reaction center of anoxygenic aerobic bacteria, including the *Chromatiales*. Our expectation was that expression would be highest at the upper depths of the OMZ where light is available. We expected similar but not necessarily identical patterns for the two genes given that (1) *Chromatiale*s have a diverse metabolism, and (2) the *pufM* assay may amplify genes from other *Alpha*- and *Gammaproteobacteria*, such as *Rhodobacter*-related organisms or the OM60 clade.

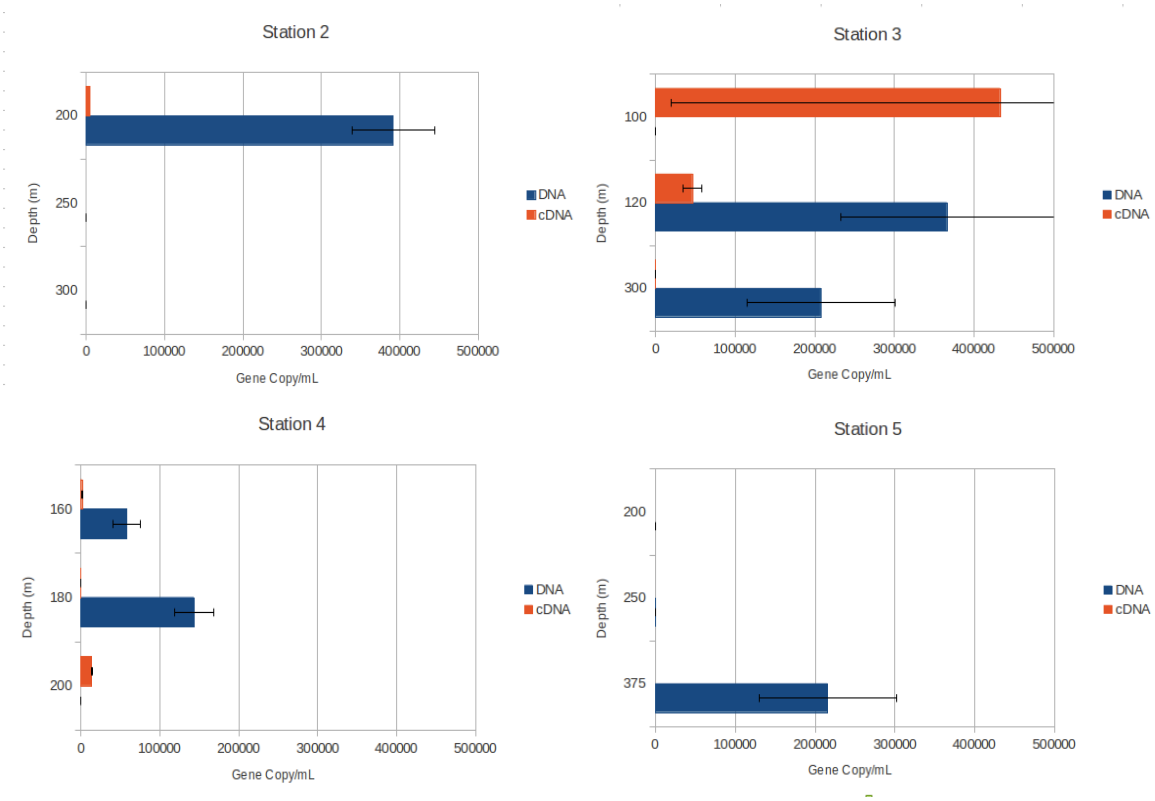

Figure 2: *Chromatiales* 16S rRNA qPCR assays. Error bars represent standard error of replicates. Note that some error bars at Station 3 exceed the scale of the graph.

Chromatiales 16S rRNA was most abundant in the DNA fraction at stations 3 and 2, while 16S rRNA was most abundant in cDNA at station 8 (Figure 2). The use of the universal bacterial 16S reverse primer was implicated earlier to explain the high efficiency of the qPCR reaction. This is also the likely source of the inconsistencies between replicates, resulting in very high standard error. Despite these uncertainties, the broad pattern in these data is clear: Chromatiales are readily detected in DNA and are present in the ETNP at all four stations, but do not appear to be very active. 16S rRNA was present in low amounts at a few depths at a few stations, and Chromatiales appear to be most active at station 3 in the shallower samples. However, we cannot reliably quantify the gene copy numbers due to the aforementioned inconsistency. In contrast, *pufM* expression was relatively homogenous among samples at station 2, low at station 3, highest in deeper samples at station 4, and highest in shallower samples at station 5

(Figure 5). This indicates that anoxygenic phototrophs other than the purple sulfur bacteria (*Chromatiales*) are active in the ETNP.

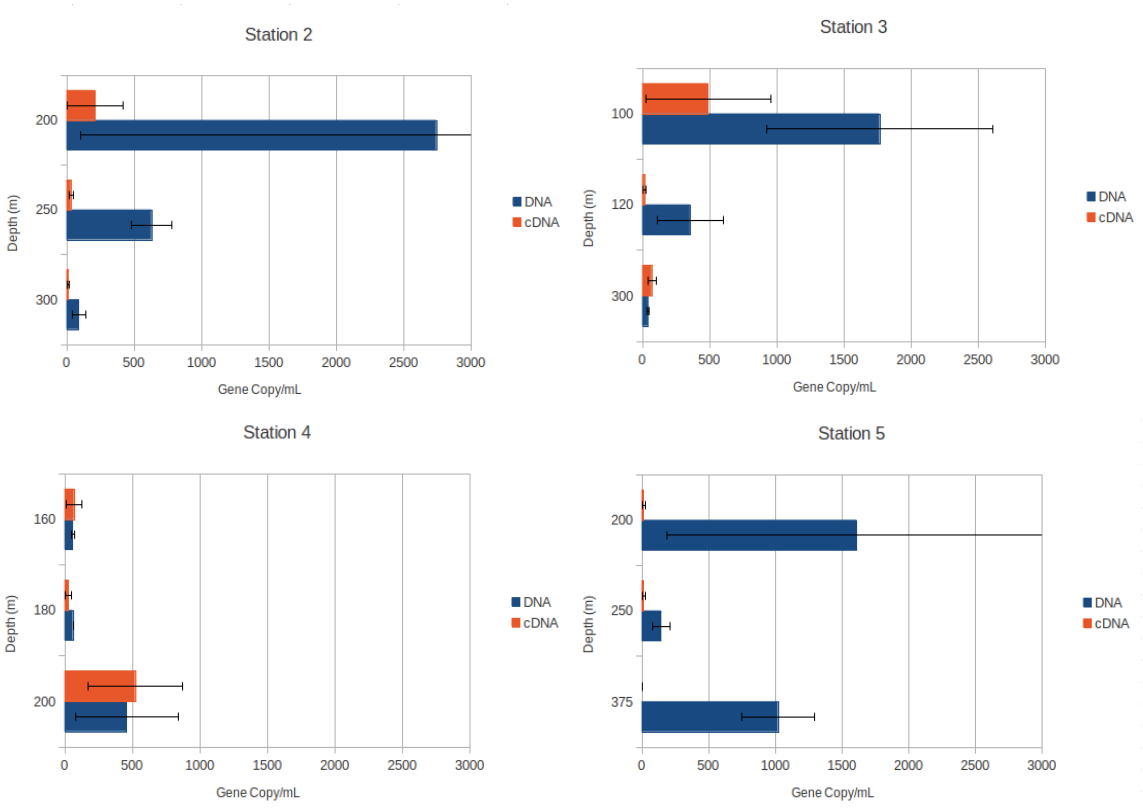

Figure 3: *SUP05* 16S rRNA qPCR assays. Error bars indicate standard error between replicates. At stations 2 and 5, DNA assay error bars exceed the scale of the graphs.

SUP05 Gammaproteobacteria are abundant in hypoxic and anoxic waters world-wide, and, based primarily on genomic data, have been characterized as chemoautotrophic denitrifiers that oxidize sulfur using nitrate/nitrite (Walsh *et al*. 2009). This process would play a key dual role in OMZs/AMZs through removing toxic sulfide (Lavik *et al*. 2009) and converting N to gaseous forms that may be lost from the system. Here we are able to examine both the abundance and potential activity of *SUP05* and related gammaproteobacterial sulfur oxidizers in the ETNP, by comparing the relative amplification of their 16S genes using extracted DNA and cDNA (produced from extracted RNA, and directly indicative of activity). SUP05 were detected at all stations but were present in low abundance (< 3000 16 rRNA genes/mL). SUP05 were least abundant at station 9, and present in comparable abundances at stations 2 and 5—however, SUP05 were much less active at Station 5 than their numbers would suggest, and 16S rRNA was barely detectable in cDNA. At station 4, 16S rRNA profiles closely tracked 16S rRNA genes, showing comparable numbers and an increase with depth.

Stations 2 and 3 were predicted to have the highest abundances since they have higher nitrite concentrations than stations 4 and 5; SUP05 were most abundant at these stations but were comparatively less active, with levels of 16S rRNA (in cDNA) that are comparable to station 4. At both station 2 and station 3, SUP05 were most abundant and active at shallower depths (200 m at station 2, 100 m at station 3). This is consistent with previous work and our hypotheses: in surveys of the Black Sea and Baltic Sea, as well as the Eastern Tropical South Pacific, *SUP05*-related organisms occur at highest abundance at the interface of oxic and anoxic waters (Glaubitz *et al*. 2013). There was a surprising amount of variation in DO and nitrite concentrations where SUP05 were most abundant and active at 200 m at station 2, 100 m at station 3, and 200 m at station 4. Stations 2 and 4 had low DO (<4 uM) at 200 m, whereas DO concentrations where higher at 100 m at station 3; nitrite concentations were elevated at 200 m at station 4 (> 4 uM), and were lower at 200 m at station 4 and 100 m at station 3. This is consistent with the idea that SUP05 have considerable metabolic diversity and may tolerate a range of geochemical conditions (Walsh *et al*. 2009).

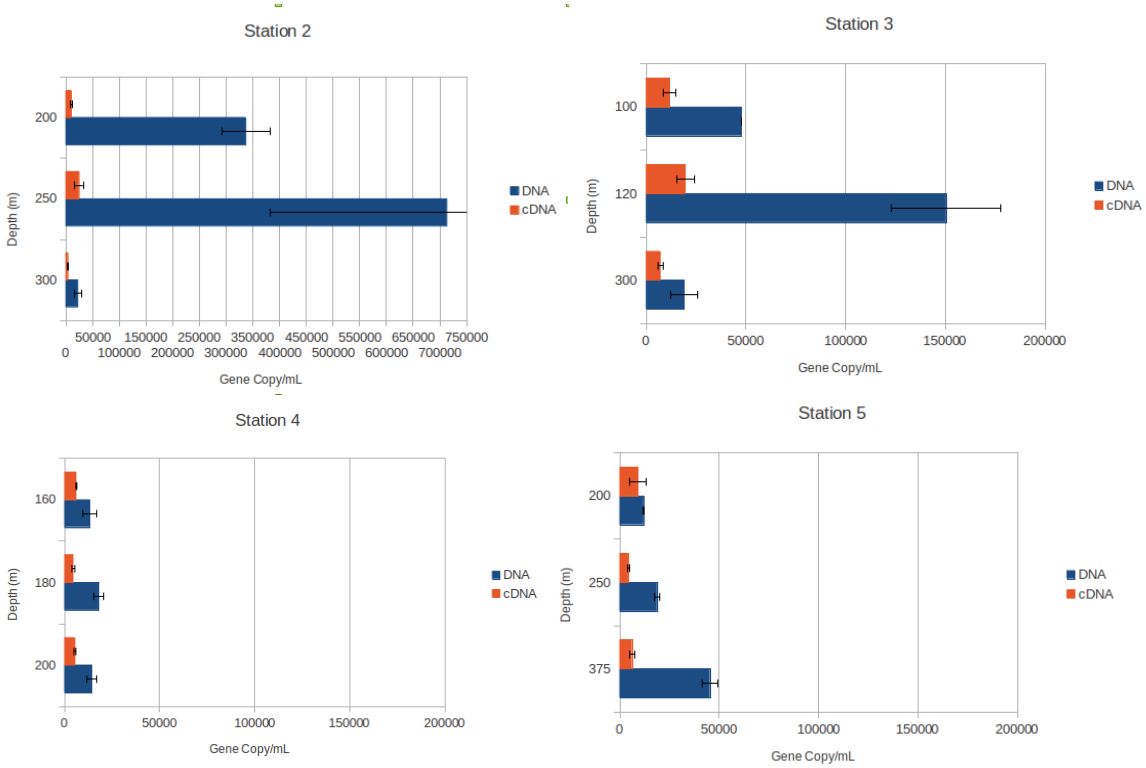

Figure 4: *DsrA* qPCR assay. Note that the scale is not consistent as amplification was much higher at station 2. Error bars represent standard error between replicates.

To determine if a cryptic sulfur cycle was possible in the ETNP OMZ, we quantified the DNA and cDNA gene copies for the gene *dsrA*, a hallmark of sulfate-reducing bacteria. Given the need for anaerobic conditions, we expected that *dsrA* expression would be

greatest in OMZ waters with high nitrite concentrations at stations 2 and 3, and amplification was in fact strongest at stations 2 and 3; weakest amplification occurred at stations 4 and 5. Gene copy counts were considerably higher at station 2, yet amplification for both DNA and cDNA was consistent for all samples. Again, stations 2 and 3 have similar profiles for dissolved oxygen and nitrite, so other factors are at work to produce such a pronounced difference in gene abundance. Notably, the nitrite oxidation rate at station 2 (200 and 250m depth) is 4 times greater than any of the measurements from the other 3 stations included in this assay. These data strongly suggest that S is being reduced in the ETNP, and that there may be syntrophic linkages between nitrite oxidation and sulfate reduction.

We also used *rdsrA* genes to examine S-cycling activity: whereas 16S rRNA genes target specific groups, *rdsrA* primers target *Alpha-* and *Gammaproteobacterial* sulfur-oxidizers—such as the *SUP05* lineage and the related *ArcticBD96-19* clade. *rdsrA* expression patterns are more difficult to interpret: if *rdsrA* is primarily expressed by anaerobic sulfur oxidizers, its expression could closely track *dsrA* expression; conversely, aerobic sulfur oxidation using *rdsrA* would be highest at shallower, more oxygenated depths. We observed a mixture of these patterns two: high expression in the OMZ at stations 2 and 3, as well as some expression at shallower depths at station 5. We also screened DNA samples for the ubiquitous *soxB* sulfur-oxidation gene, which can serve as a proxy for overall sulfur-oxidation within the water column. Here, again, the amplification signal was strongest at stations 3 and 2, but samples from 300m depth demonstrated the strongest amplification. This suggests significant diversity within S cycling pathways in the ETNP OMZ.

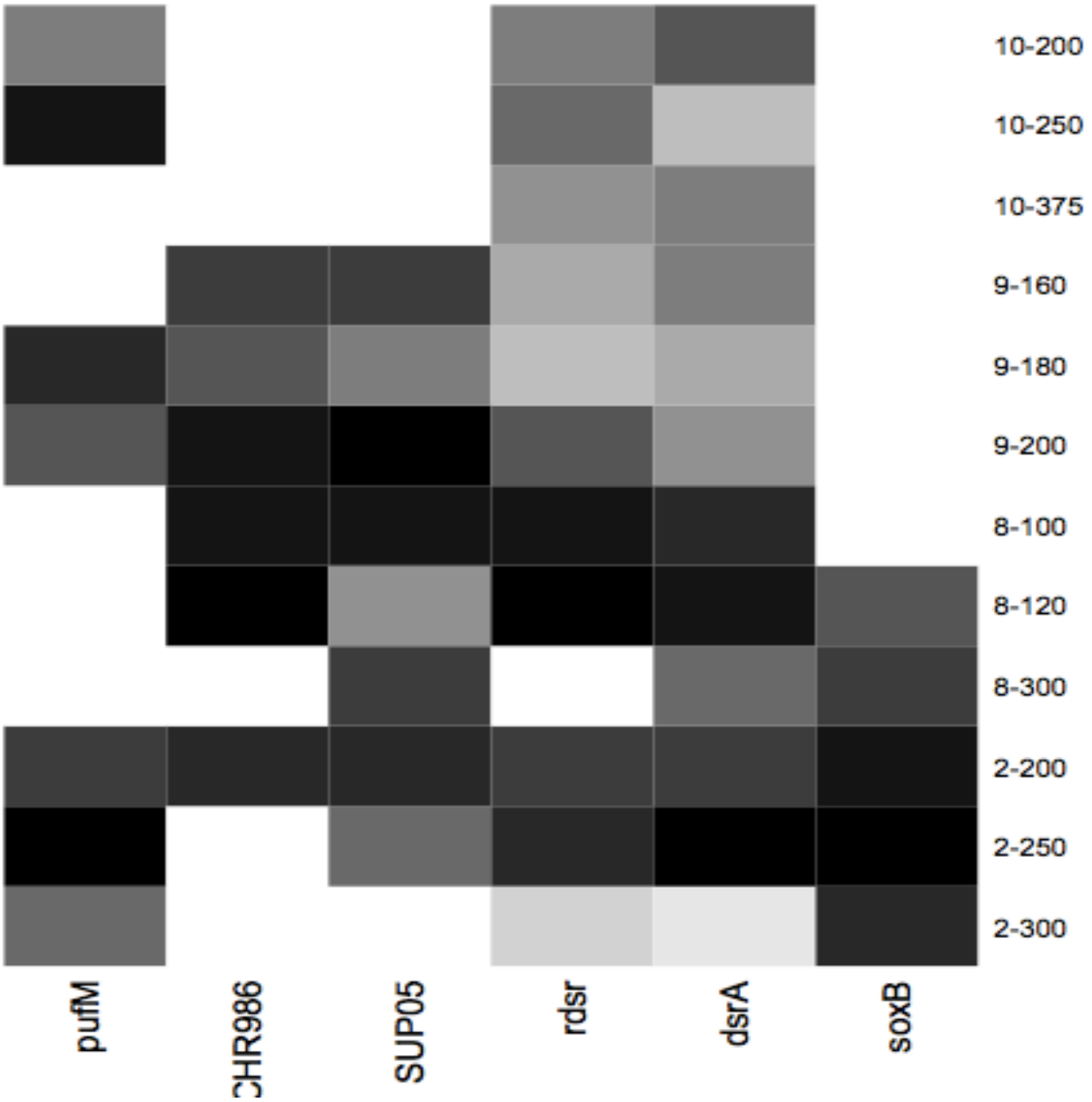

Figure 5: Ranked heatmap of qPCR amplification of GOIs. Black indicates relatively more amplification and white indicates no amplification. SoxB amplification used extracted DNA rather than cDNA as with the other GOIs.

Anaerobic S oxidation in the OMZ is consistent with the activity of the *Chromatiales* and SUP05 bacteria observed above, and along with *dsrA* expression, indicates that a cryptic S cycle operates in the ETNP. These data expand the known distribution of S cycling in the ocean, and raise several important issues. Most significant is determining the extent of S cycling in terms of absolute rates, lateral extent, and depth distribution in the

ETNP—as the largest OMZ in the ocean, this will be key for understanding biogeochemical and ecological implications. Interestingly, amplification was low at 300m depth at both stations 3 and 2 *for dsrA and rdsrA* genes even though the nitrite concentrations were at 2008 uM and 656 uM, while oxygen concentrations for these samples were measured as 3.0 umol kg^{-1} at station 3 and 3.7 umol kg^{-1} at station 2. In the ETSP, *dsr* and *rdsr* were also most prevalent at intermediate depth (150m) with low concentrations of dissolved oxygen but lower nitrite than in the ETNP (Canfield *et al*. 2010). The reasons for these patterns are unknown, but indicate that S cycling may not extend through the entire depth of the OMZ.

Identifying microbial interactions using MINE

Once the distribution patterns and activity of these known sulfur-metabolizing organisms had been quantified using pyrosequencing and qPCR, we could use MINE to find correlations between these groups of organisms and other members of their potential community. Following the screening and quality control outlined in the methods above, approximately 130,000 unique 16S rRNA sequences were clustered into OTUs. Then, for each of the 60 sample locations, the abundance of each OTU was listed along with environmental data, and analyzed using MINE. The redox chemistry of suboxic and anoxic waters lends itself to syntrophic interactions, and exploratory statistics like MINE offer the ability to detect many different types of correlations and possibly complex relationships between these microbial groups (Wright *et al*. 2012).

Seven OTUs were classified as belonging to the group *ArcticBD96*-19 (also known as *ZD0405*), but only one had more than 3 sequences across all 60 samples. Otu3059, comprised of 2399 sequences, was the second largest Gammaproteobacterial group behind *ZD0417*. The MIC scores demonstrated strong exclusion patterns between *ArcticBD96-19* and surface-dwelling OTUs such as *SAR86*, *Rhodobacterales*, and *Acidimicrobiales*. Co-occurrence patterns were seen between Otu3059 and the most abundant *Nitrospina* OTU, as well as the two most abundant *SAR324* OTUs and a *SAR406* OTU. It should also be noted that Otu3059 is excluded by high dissolved oxygen concentration.

The largest *Nitrospina* group, Otu3739, comprised of 2198 sequences, also has *ArcticBD96-19* as its reciprocal best MIC-score for co-occurrence. Additionally, it shows co-occurrence patterns with the same *SAR324* OTUs as Otu3059, and similar exclusion patterns with *Acidimicrobiales*, *Rhodobacterales*, *SAR86*, and higher dissolved oxygen concentrations. Otu2951 (6348 sequences) and Otu2952 (3409 sequences) represent the largest *SAR324* OTUs. Besides the relationships outlined above with *Nitrospina* and *ArcticBD96-19*, Otu2951 co-occurs with a *SAR406* OTU, a member of the *Deferribacteres* phylum. The smaller of the two OTUs, Otu2952, showed strong exclusion patterns with chlorophyll concentration and temperature, indicating that this group may be deeper-dwelling than the larger OTU. Additionally, it showed strong co-occurrence patterns with many unidentified *Gammaproteobacterial* OTUs. A search for similar sequences from these groups using public databases revealed that they were

distantly related to tube-worm and nematode endosymbionts collected from hydrothermal vents.

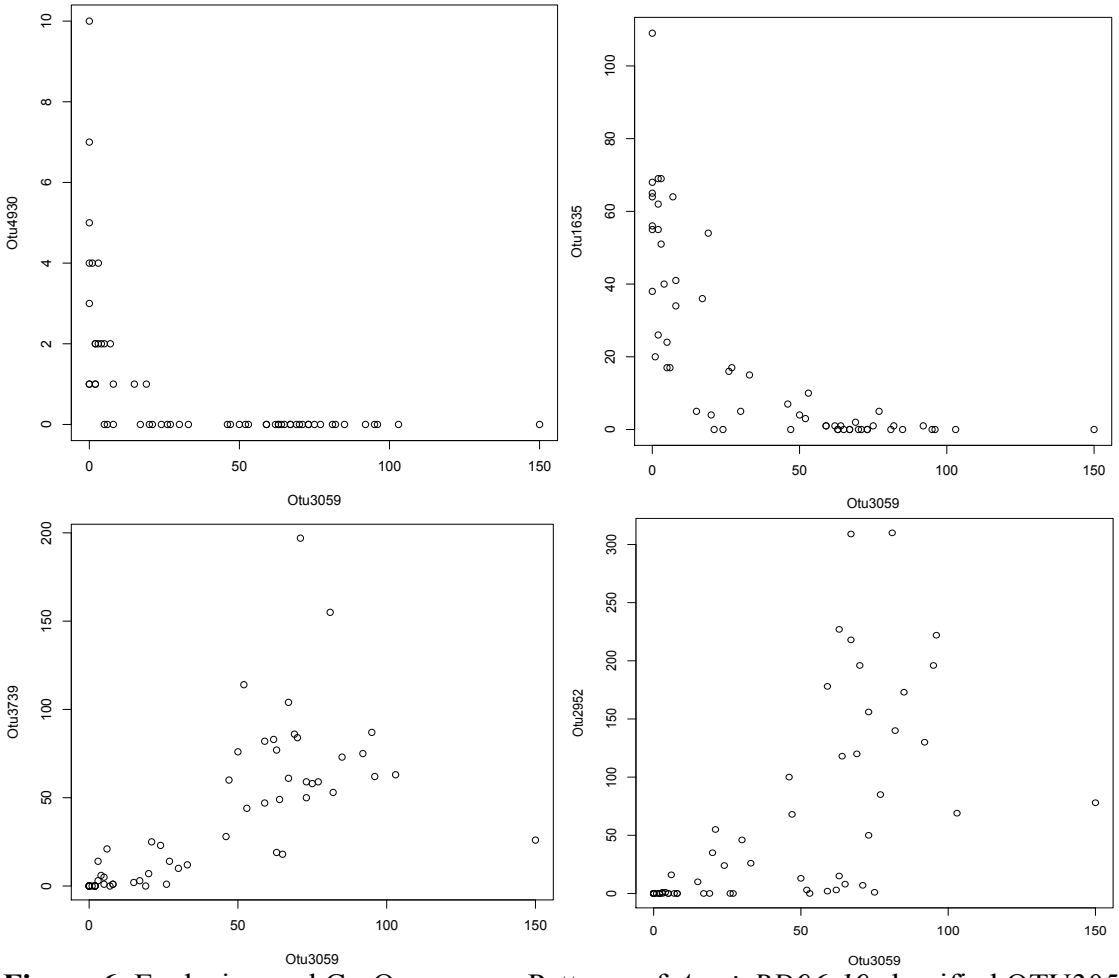

Figure 6: Exclusion and Co-Occurrence Patterns of *ArcticBD96-19* classified OTU3059. Top row features abundance patterns of OTU4930 (*SAR86*) and OTU1635 (most abundant *Rhodobacterales* group). Bottom row demonstrates abundance of the most abundant *Nitrospina* grouping (OTU3739) and the second most abundant SAR324 cluster (OTU2952).

MaxEnt

MaxEnt is a tool for deriving species distribution models from presence-only data using the principle of maximum entropy. Each OTU was associated with environmental data from the samples in which it was found, and MaxEnt produced a model for predicting OTU presence based on said environmental data. The goal of species distribution models like those produced with MaxEnt is to determine which OTUs distributions' are defined by low oxygen tension, high available ammonium, nitrite, or sunlight (photosynthetically

active radiation, or PAR). The main method of evaluating the strength of said model is the area under the receiver operating criteria (ROC) curve, or AUC. An AUC value of 1 indicates perfect model fitting, while a value of 0.5 means that the model was no better than a random distribution at predicting presence. A comparison of different methods of species distribution modeling found MaxEnt to be particularly strong when dealing with small and incomplete data sets, some as small as only 5 records of presence (Hernandez *et al*. 2006). Elith *et al* 2011 suggested a baseline AUC value of 0.75 to begin examining model strength and properties, so from the initial dataset of over 6000 OTUs, only 611 OTU models remained. These had an average AUC of 0.867 and a median of 0.863. Variable contribution is measured as a percentage, and of these OTU-models, dissolved oxygen (DO) was by far the most important variable, as seen in the table below.

Table 2: Variable contributions for MaxEnt species distribution models

	Ammonium	Chlorophyll	Dissolved Oxygen	Density	Nitrite	PAR	Salinity	Temperature	Transmission
Average	3.609	7.420	31.99	4.954	9.581	8.67	11.08	11.07	11.63
Median	0	0	15.16	0	2.699	0.313	0	0	0
Sum	2205	4534	19550	30267	5854	5298	6768	6762	7103

SUM is the total variable contribution across all 611 OTU-models.

These 611 well-described OTUs have a median size of 9 sequences, and so are smaller on average compared to the whole dataset. 143 of these OTUs were initially listed as unclassified bacteria using Mothur and the SILVA-SEED database as described above. However, the majority of these are plastid sequences, so combined with the 32 OTUs explicitly classified as Chloroplasts, this makes just under 1/3 of these OTUs microbial eukaryotes. The remaining OTUs are a diverse group including *Flavobacteria*, *Verrucomicrobia*, *Acidobacteria*, *Planctomycetes*, *Lentisphaerae*, and *Alpha-, Gamma-,* and *Deltaproteobacteria*.

Dissolved oxygen contributed 40% or more to the MaxEnt model for 229 OTUs. Approximately a quarter of these sequences were classified as Cyanobacteria or Chloroplasts. In addition to numerous small OTUs from the Alphaproteobacteria, Gammaproteobacteria, and Flavobacteria, a few familiar, OMZ-community large OTUs were modeled. These included OTU3068 (Chromatiales), OTUs 3740/3743 (Nitrospina), and OTUs representing SAR406/MGA and the Oceanospirialles (OM182).

Far fewer OTUs had 40% or more contribution from nitrite towards their models. Only 32 of the 611 fit into this category, and of those only OTU3073, classified as Alteromonas, represented 100 or more sequences. Otherwise, Flavobacteria and the aforementioned Altermonadales predominated.

Conclusions

Nitrospina, a genus formerly classified as part of the *Deltaproteobacteria*, was identified as a highly abundant taxon from the initial classification of the greater ETNP pyrosequencing library. MINE analysis revealed that it had many highly significant

interactions with other known OMZ bacteria, particularly sulfur-oxidzing bacteria such as the SUP05/ARCTICBD96-19 clade, and MaxEnt modeling identified it as a resident of the anoxic OMZ waters. Abundant world-wide in marine environments, *Nitrospina* has few cultured representatives, but a group at the University of Vienna has published a genome for *N. gracile* (Lücker *et al*. 2013). Lücker *et al* confirmed this representative as an obligate chemolithautotrophic nitrite-oxidizing bacterium. It utilizes sodium-dependent proton pumping to directly power a TCA cycle, relies on hydrogen, polysulfides, or thiosulfite as electron donors, and strangely lacks genes to protect it from reactive oxygen species. Representatives of the *Nitrospina* OTUs 3739, 3740, and 3743 shared 90% sequence identity with *N. gracile* 16S rRNA sequences available in GenBank. *Nitrospina* likely constitutes a major nitrate source for the OMZ community.

Highly abundant *Gammaproteobacterial* sulfur-oxidizers, as well as mixotrophs like *SAR324*, clearly take advantage of nitrate as an electron source in the anoxic waters, so what remains is identifying the sulfate-reducing bacteria present. Our *dsrA* assays indicate that sulfate reduction is occurring within the OMZ. Single amplified genomes (SAGs) and now metagenomes have been published for *MGA/SAR406* and *SAR324*, but none were able to settle the question as to whether these organisms are capable of sulfate-reduction (Swan *et al* 2011,Wright *et al* 2013, Shiek *et al*. 2013). Frequently *rdsr* and *dsr* are used interchangeably to refer to the reverse-acting (sulfur-oxidizing) gene, and the decision to classify a gene as one or the other depends on whether it shares similarity to genes found in known sulfur-oxidizers or sulfur-reducers. Also, *dsrAB*, while highly conserved and found in all known sulfate-reducing prokaryotes, has transferred horizontally between bacterial lineages and even to *Archaeans* (Klein *et al*. 2001).

Culture-independent technique has revealed the vast world of microbial diversity, but has not resolved long-standing problems in identifying and defining bacterial species (Fraser *et al*. 2009). The ecotype model, along with the use of DNA and RNA similarity measurements, has allowed microbial ecology to forge ahead in modeling community interactions. However, the difficult work of culturing these organisms may be necessary in many cases to assess the level of similarity within and between populations, as well as to verify their metabolic potential and preferences. Defined species or species-groups will allow for robust comparisons between OMZ waters, and will even allow for comparisons to seasonally anoxic waters or “dead-zones” created by eutrophication.

References

Beman, JM, BN Popp, and SE Alford. Quanitifcaiton of ammonia oxidation rates and ommnia-oxidizing archaea and bacteria at high resolution in the Gulf of California and the eastern tropical North Pacific. *Limnology and Oceanography* 57(3) (2012)

Beman, JM and MT Carolan. Deoxygenation alters bacterial diversity and community composition in the ocean's largest oxygen minimum zone. *Nature Communications* 4 (2013).

Canfield, DE. Sulfate reduction and oxic respiration in marine sediments: implications for organic carbon preservation in euxinic environments. *Deep Sea Research Part A: Oceanographic Research Papers* 38 (1989).

Canfield, DE, FJ Stewart, B Thamdrup, LD Brabandere, T Dalsgaard, EF Delong, NP Revsbech and O Ulloa. A Cryptic Sulfur Cycle in the Oxygen Minimum Zone Waters off the Chilean Coast. *Science* 330 (2010).

Deutsch C, JL Sarmiento, DM Sigman, N Gruber and JP Dunne. Spatial coupling of nitrogen inputs and losses in the ocean. *Nature* 445 (2007).

Diaz, RJ and R Rosenberg. Spreading dead zones and consequences for marine ecosystems. *Science* 321 (2008).

Elith, J. and JR Leathwick. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* 40 (2009).

Falkowski, PG, T Fenchel, EF Delong. The Microbial Engines that Drive Earth's Biogeochemical Cycles. *Science* 320 (2008).

Francis, CA, JM Beman and MMM Kuypers. New processes and players in the nitrogen cycle: the microbial ecology of anaerobic and archaeal ammonia oxidation. *The ISME Journal* 1 (2007).

Francis CA, KJ Roberts, JM Beman, AE Santoro, and BB Oakley. Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. *Proceedings of the National Academy of Sciences* 102: 41 (2005).

Fraser, C, EJ Alm, MF Polz, BG Spratt, WP Hanage. The Bacterial Species Challenge: Making Sense of Genetic and Ecological Diversity. *Science* 323 (2009).

Ghosh, W and B Dam. Biochemistry and Molecular Biology of Lithotrophic Sulfur Oxidation by Taxonomically and Ecologically Diverse Bacteria and Archaea. *FEMS Microbiology Review* 33:6 (2009).

Gilbert, JA and CL Dupont. Microbial Metagenomics: Beyond the Genome. *Annual Review of Marine Science* 3 (2011).

Gilly, WF, JM Beman, SY Litvin, BM Robinson. Oceanographic and biological effects of shoaling of the oxygen minimum zone. *Annual Review of Marine Science* 5 (2013).
Gruber, N and JN Galloway. An Earth-Systems perspective of the global nitrogen cycle. *Nature* 451 (2008).

Glaubitz, S, K Kießlich, C Meeske, M Labrenz, and K Jürgens. SUP05 dominates the gammaproteobacterial sulfur oxidizer assemblages in pelagic redoxclines of the central Baltic and Black Seas. *Applied Environmental Microbiology* 79(8) (2013).

Helly, JJ and LA Levin. Global distribution of naturally occurring marine hypoxia on continental margins. *Deep Sea Research Part I: Oceanographic Research Papers* 51:9 (2004).

Imhoff, JF. True marine and halophilic anoxygenic phototrophic bacteria. *Archives of Microbiology* 178(4):243-54 (2001).

Imhoff, JF. Chromatiales *ord. nov. Bergey's Manual of Systematic Bacteriology*. Springer US 2005.

Jørgensen, BB, H Fusing, CO Wirsen, HW Jannasen. Sulfide oxidation in the anoxic Black Sea chemocline. *Deep Sea Research Part A: Oceanographic Research Papers* 38 (1991).

Keeling, RF, A Kurtzinger, N Gruber. Ocean Deoxygenation in a Warming World. *Annual Review of Marine Science* 2 (2010).

Ladau, J, TJ Sharpton, MM Finucane, G Jospin, SW Kemel, J O'Dwyer, AF Koeppel, Jl Green, and KS Pollard. Global marine bacterial diversity peaks at high latitudes. *The International Society for Microbial Ecology Journal* 7 (2013).

Larsen, PE, D Field, and JA Gilbert. Predicting bacterial community assemblages using an artificial neural network approach. *Nature Methods* 9:6 (2012).

Lavik, G, T Stührmann, V Brüchert, A Van der Plas, V Mohrholz, P Lam, M Mußmann, BM Fuchs, R Amann, V Lass, and MMM Kuypers. Detoxification of sulphidic African shelf waters by blooming chemolithotrophs. *Nature* 457 (2009).

Loy, A, S Duller, C Baranyi, M Mushmann, J Ott, J Shanon, O Beja, DL Palier, C Dahl, and M Wagner. Reverse dissimilatory sulfite reductase as a phylogenetic marker for a subgroup of sulfur-oxidizing prokaryotes. *Environmental Microbiology* 11:2 (2009).

Lücker, S, B Nowka, T Rattei, E Spieck, and H Daims. The genome of *Nitrospina gracilis* illuminate the metabolism and evolution of the major marine nitrite oxidizer. *Frontiers in Microbiology* 4 (2013).

Marshall, KT and RM Morris. Isolation of an aerobic sulfur oxidizer from the SUP05/Arctic96BD-19 clade. *The ISME Journal* 7 (2013).

Meyer, B, JF Imhoff, and J Kuever. Molecular analysis of the distribution and phylogeny of the *soxB* gene among sulfur-oxidizing bacteria. *Environmental Microbiology* 11:2 (2009).

Morris RM, MS Rappe, SA Connon, KL Vergin, WA Siebold, CA Carlson, SJ Giovannoni. SAR11 Clade dominates ocean surface bacterioplankton communities. *Nature* 420 (2002)
Muyzer, G, and AJM Stams. The Ecology and Biotechnology of Sulphate-Reducing Bacteria. *Nature Review Microbiology* 6 (2008).

Overmann, J. Diversity and ecology of phototrophic sulfur bacteria. *Microbiology Today* 28 (2001).

Paulmier, A and D Ruiz-Pino. Oxygen minimum zones in the modern ocean. *Progress in Oceanography* 80 (2009).

Stramma, L, GC Johnson, J Sprintall, V Mohrholz. Expanding oxygen minimum zones in the tropical oceans. *Science* 320 (2008).

Strimmer, K. fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics* 24:12 (2008).

Swan BK, M Martinze-Garcia, CM Preston, AT Woyke, D Lamy, T Reinthaler, NJ Poulton, EDP Masland, ML Gomez, ME Sieracki, EF Delong, GJ Herndl and R Stepanauskas. Potential for Chemolithoautotrophy Among Ubiquitous Bacteria Lineages in the Dark Ocean. *Science* 333 (2011).

Waidner, LA and DL Kirchman. Diversity and distribution of ecotypes of the aerobic anoxygenic phototrophy gene *pufM* in the Delaware Estuary. *Applied and Environmental Microbiology* 74:13 (2008).

Walsh, DA, E Zaikova, CG Howes, YC Sung, JJ Wright, SG Tringe, PD Tortell, SJ Hallam. Metagenome of a Versatile Chemolithoautotroph from Expanding Oceanic Dead Zones. *Science* 326 (2009).

White, AE, RA Foster, CR Benitez-Nelson, P Masqué, E Verdeny, BN Popp, KE Arthur and FG Prahl. Nitrogen fixation in the Gulf of California and the Eastern Tropical North Pacific. *Progress in Oceanography* 209 (2012).

Wright, JJ, KM Knowar, and JJ Hallam. Microbial ecology of expanding oxygen minimum zones. *Nature Reviews: Microbiology* 10 (2012).

Zaikova, E, DA Walsh, CP Stilwell, WW Mohn, PD Tortell, and SJ Hallam. Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. *Environmental Microbiology* 12(1): 172-192 (2010).

Zehr JP and RM Kudela. Nitrogen Cycle of the Open Ocean: From Genes to Ecosystems. *Annual Review of Marine Science* 3 (2011).