Supplementary Online Material:

Size-fraction partitioning of community gene transcription and nitrogen metabolism in a marine oxygen minimum zone

Contributors: Sangita Ganesh, Laura A. Bristow, Morten Larsen, Neha Sarode, Bo Thamdrup, Frank J. Stewart

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 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 cd_1-containing nitrite reductases, respectively, but pooled for representation in Figure 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 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). 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 S3). 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 S3). 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 ^{15}N 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 S6). 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 S6). 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 S6), suggesting a potential role for larger particles as substrate or cell sources in the upper OMZ.

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 subarctic 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.

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.

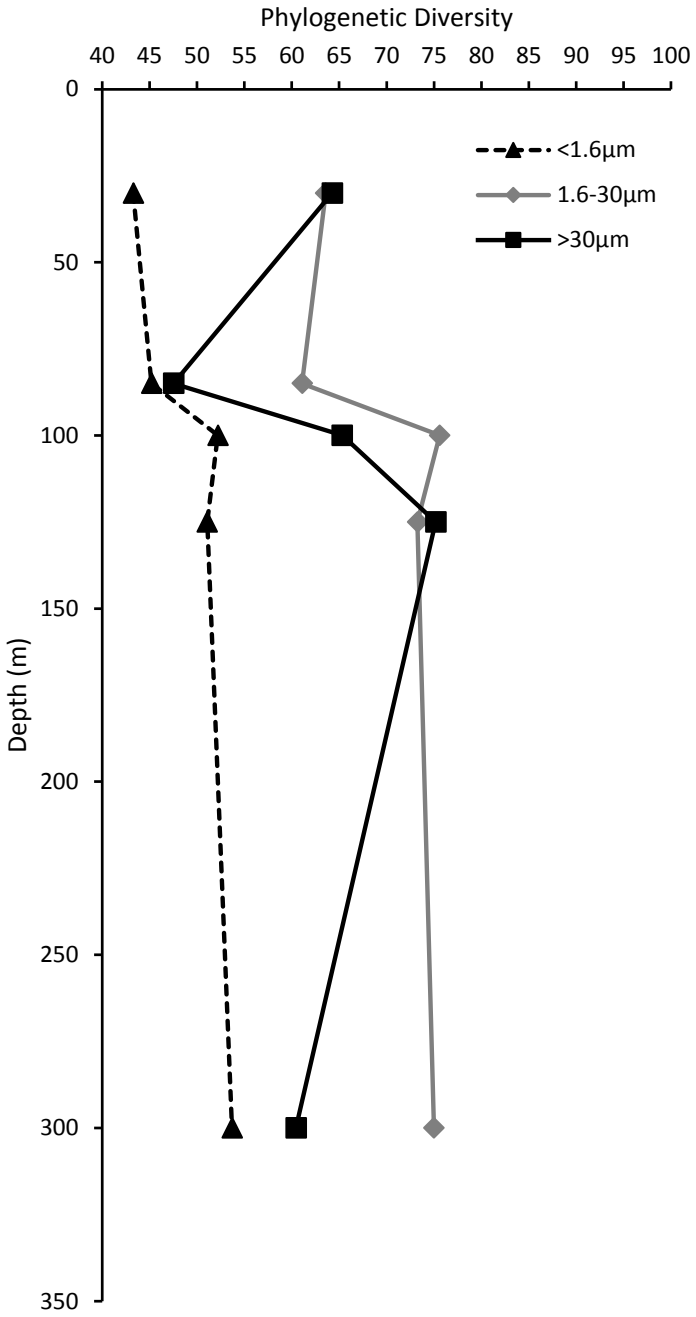
Zaikova E, Walsh DA, Stilwell CP, Mohn WW, Tortell PD, Hallam SJ. (2010). Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. *Environ Microbiol* **12**: 172-191.

Table S1. 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 S1. 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.



240

Figure S2. 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.

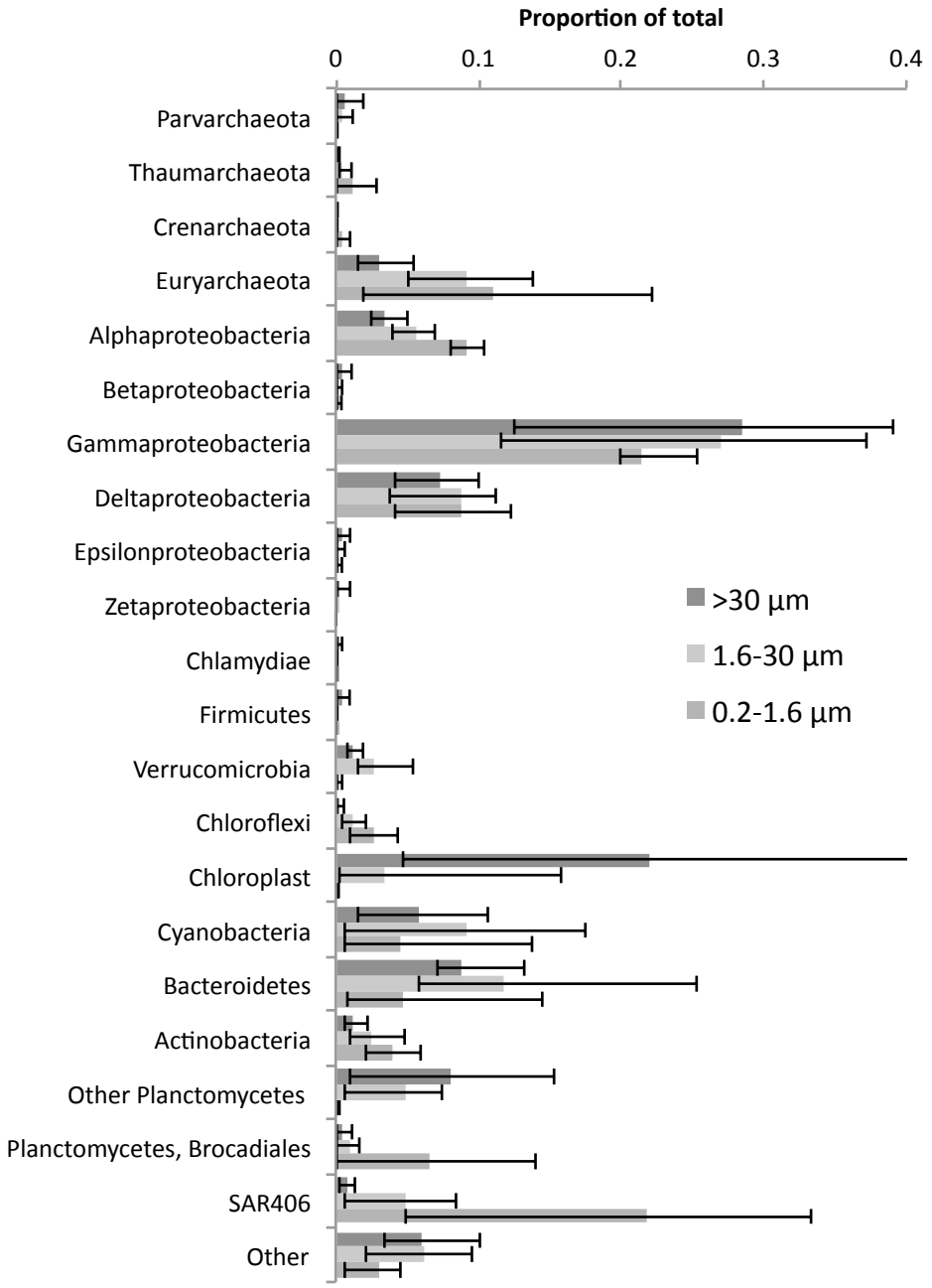


Figure S3. Taxonomic composition of 16S rRNA gene amplicons within the Phylum Proteobacteria. The abundances of major divisions are shown as a percentage of total identifiable 16S rRNA gene sequences. Taxonomic identifications are based on the Greengenes taxonomy.

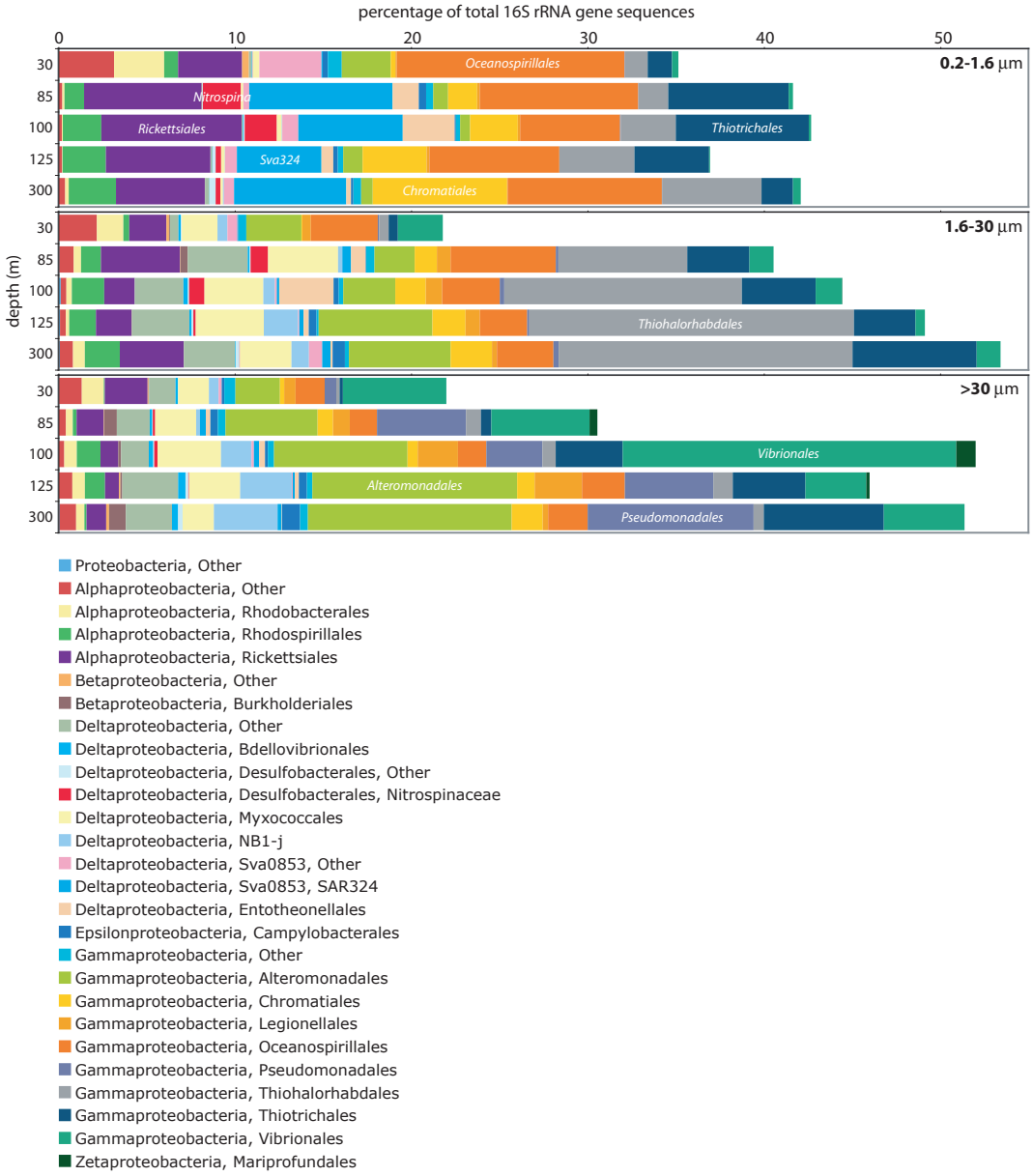


Figure S4. Fold increase in taxon RNA:DNA between PA (1.6-30 µm) and FL (0.2-1.6 µm) communities. Plot shows the fold increase in ratios for major taxonomic groups identified by the Greengenes identities of 16S rRNA gene amplicons (DNA) and 16S rRNA transcript (RNA) fragments in metatranscriptome datasets. 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 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).

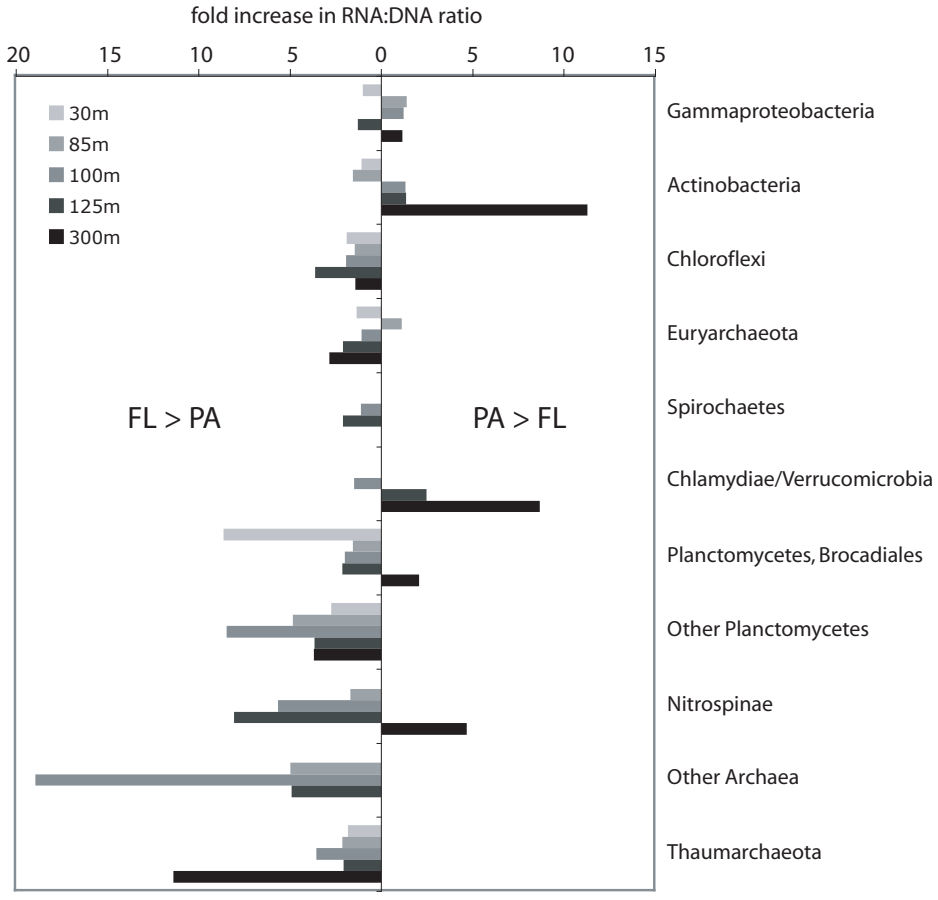


Figure S5. Size fraction-specific differences in transcripts affiliated with anammox bacteria. Bars are abundances shown as a percentage of the total number of transcripts with top matches (bit score > 50) to genes of known anammox genera (*Scalindua, Kuenenia, Brocadia, Jettenia, Anammoxoglobus*), as 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 (R^2 = 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).