Metagenomic analysis of nitrogen and methane cycling in the Arabian Sea oxygen minimum zone

Claudia Lüke, Daan R. Speth, Martine A.R. Kox, Laura Villanueva, Mike S.M. Jetten

Oxygen minimum zones (OMZ) are areas in the global ocean where oxygen concentrations drop to below one percent. Low oxygen concentrations allow alternative respiration with nitrate and nitrite as electron acceptor to become prevalent in these areas, making them main contributors to oceanic nitrogen loss. The contribution of anammox and denitrification to nitrogen loss seems to vary in different OMZs. In the Arabian Sea, both processes were reported. Here, we performed a metagenomics study of the upper and core zone of the Arabian Sea OMZ, to provide a comprehensive overview of the genetic potential for nitrogen and methane cycling. We propose that aerobic ammonium oxidation is carried out by a diverse community of *Thaumarchaeota* in the upper zone of the OMZ, whereas a low diversity of *Scalindua*-like anammox bacteria contribute significantly to nitrogen loss in the core zone. Aerobic nitrite oxidation in the OMZ seems to be performed by *Nitrospina spp*. and a novel lineage of nitrite oxidizing organisms that is present in roughly equal abundance as *Nitrospina*. Dissimilatory nitrate reduction to ammonia (DNRA) can be carried out by yet unknown microorganisms harbouring a divergent *nrfA* gene. The metagenomes do not provide conclusive evidence for active methane cycling, however, a low abundance of novel alkane monooxygenase diversity was detected. Taken together our approach confirmed the genomic potential for an active nitrogen cycle in the Arabian Sea and allowed detection of hitherto overlooked lineages of carbon and nitrogen cycle bacteria.

Metagenomic analysis of nitrogen and methane cycling in the Arabian Sea oxygen minimum zone

Claudia Lüke[1]*, Daan R. Speth[1]*, Martine A.R. Kox[1], Laura Villanueva[3], Mike S.M. Jetten[1,2,4]

* These authors contributed equally to this work

[1]Department of Microbiology, IWWR, Radboud University Nijmegen, Nijmegen, the Netherlands

[2]Department of Biotechnology, Delft University of Technology, Delft, the Netherlands.

[3]Department of Marine Organic Biogeochemistry, Royal Netherlands Institute for Sea Research (NIOZ), 't Horntje (Texel), the Netherlands

[4]Soehngen Institute of Anaerobic Microbiology, Nijmegen, the Netherlands

Corresponding Author addresses (clueke@science.ru.nl, D.Speth@science.ru.nl)

Abstract

Oxygen minimum zones (OMZ) are areas in the global ocean where oxygen concentrations drop to below one percent. Low oxygen concentrations allow alternative respiration with nitrate and nitrite as electron acceptor to become prevalent in these areas, making them main contributors to oceanic nitrogen loss. The contribution of anammox and denitrification to nitrogen loss seems to vary in different OMZs. In the Arabian Sea, both processes were reported. Here, we performed a metagenomics study of the upper and core zone of the Arabian Sea OMZ, to provide a comprehensive overview of the genetic potential for nitrogen and methane cycling. We propose that aerobic ammonium oxidation is carried out by a diverse community of *Thaumarchaeota* in the upper zone of the OMZ, whereas a low diversity of *Scalindua*-like anammox bacteria contribute significantly to nitrogen loss in the core zone. Aerobic nitrite oxidation in the OMZ seems to be performed by *Nitrospina spp*. and a novel lineage of nitrite oxidizing organisms that is present in roughly equal abundance as *Nitrospina*. Dissimilatory nitrate reduction to ammonia (DNRA) can be carried out by yet unknown microorganisms harbouring a divergent *nrfA* gene. The metagenomes do not provide conclusive evidence for active methane cycling, however, a low abundance of novel alkane monooxygenase diversity was detected. Taken together our approach confirmed the genomic potential for an active nitrogen cycle in the Arabian Sea and allowed detection of hitherto overlooked lineages of carbon and nitrogen cycle bacteria.

Introduction

Oxygen is a key parameter for biogeochemical cycling and has major impact on the marine nitrogen and carbon turnover. The vast majority of the global ocean waters is well oxygenated, allowing aerobic micro- and macro-organisms to thrive. However, in several areas, underlying regions of high productivity, dissolved oxygen concentrations drop to very low levels. These regions are referred to as oxygen minimum zones (OMZ). There is not a general agreement on the definition of an OMZ, however, an oxygen concentration of ≤ 20 µM was proposed (Lam & Kuypers, 2011). Using this threshold, approximately 1% of the total ocean volume can be defined as an OMZ (Lam & Kuypers, 2011). In the eastern tropical North Pacific (ETNP), the eastern tropical South Pacific (ETSP) and the Arabian Sea, the three prominent OMZs, oxygen concentration can even drop below levels detectable by sensitive modern techniques (Revsbech et al., 2009; Thamdrup, Dalsgaard & Revsbech, 2012).

Despite comprising only a small fraction of the total ocean volume, OMZs contribute 30-50% of the nitrogen loss from the ocean (Gruber & Sarmiento, 1997; Codispoti et al., 2001). This can be attributed to a highly active nitrogen cycle in these systems (Lam & Kuypers, 2011). After depletion of oxygen, nitrate is the next most energetically favourable terminal electron acceptor and is present in micro-molar concentrations in OMZs (e.g. Pitcher et al., 2011). Nitrate reduction coupled to the oxidation of organic matter releases 16 mole ammonium per mole organic matter oxidized (Redfield et al, 1963). In addition to this, ammonium can be produced by dissimilatory nitrite reduction to ammonium (DNRA). Despite being aerobic processes, ammonium and nitrite oxidation occur in OMZs, partially converting the ammonium back to nitrite and nitrate (Kalvelage et al., 2011, 2015; Füssel et al., 2012). Eventually, nitrogen is lost

67 from the system due to denitrification (Groffman et al., 2006; Ward et al., 2009) or anaerobic
68 ammonium oxidation (anammox) (Kuypers et al., 2003; Kalvelage et al., 2015).
69 The relative contribution of anammox and denitrification to nitrogen loss from OMZs has been
70 the subject of debate (Lam & Kuypers, 2011). Before the discovery of anammox, denitrification
71 was thought to be the only contributor (Codispoti & Richards, 1976; Lipschultz et al., 1990;
72 Devol et al., 2006), but the detected ammonium concentrations were lower than expected based
73 on just denitrification. After the discovery of the anammox process in the Black Sea (Kuypers et
74 al., 2003), anammox bacteria were shown to occur in all the major OMZ using marker genes,
75 lipid analysis, FISH and stable isotope pairing (e.g. Kuypers et al., 2005; Thamdrup et al., 2006;
76 Hamersley et al., 2007). In the ETSP OMZ, anammox was the dominant process involved in
77 nitrogen loss (Lam & Kuypers, 2011; Stewart, Ulloa & DeLong, 2012). For the Arabian Sea
78 OMZ, evidence for both denitrification and anammox as the dominant cause of nitrogen loss
79 exists (Ward et al., 2009; Jensen et al., 2011; Pitcher et al., 2011) and the contribution of either
80 process likely varies with season and location.
81 Previous studies on nitrogen cycling the Arabian Sea OMZ focused on one or (a comparison of)
82 a few processes such as ammonia oxidation (Newell et al., 2011; Schouten et al., 2012),
83 denitrification (Jayakumar et al., 2004; Ward et al., 2009), and anammox (Jensen et al., 2011;
84 Pitcher et al., 2011; Villanueva et al., 2014). A recent study used metagenomes obtained from
85 samples across the ETSP, ETNP and the Bermuda Atlantic Time-series Station to access the
86 distribution of iron and copper containing nitrogen cycle enzymes in these systems (Glass et al.,
87 2015), but a comprehensive study of the nitrogen cycling potential in the Arabian Sea OMZ is
88 lacking.

89 Furthermore, only very little is known on methane turnover in oxygen minimum zones. Low
90 concentrations of methane have been reported at sites of the ETNP and the ETSP suggesting the
91 possibility of methane cycling in these ecosystems (Sansone & Popp, 2001, Padilla et al., 2016).
92 In addition, *pmoA* genes (encoding the particulate methane monooxygenase, see Table 1) have
93 been detected in the water column of the ETSP OMZ indicating the presence of aerobic
94 methanotrophs (Tavormina et al., 2013). In the Arabian Sea, high methane concentrations (up to
95 227% saturation compared to atmospheric levels) have been measured in the surface waters
96 (Owens et al., 1991; Bange et al., 1998; Upstill-Goddard, Barnes & Owens, 1999) and elevated
97 concentrations (up to approximately 8 nM) were found at 150-200m depths (Jayakumar et al.,
98 2001). Not much is known about the processes and the organisms involved in production or
99 turnover, but recently *Thaumarchaeota* cleaving methylphosphonates have been proposed as a
100 potential source of methane in the oceans (Metcalf et al., 2012).
101 Metagenomics is a powerful tool to provide an all-inclusive picture of the functional potential of
102 an ecosystem. As sequencing is becoming easier and less expensive, the bottleneck in
103 metagenomics is shifting from data generation to sequence analysis strategies. Here, we
104 developed a new analysis strategy for mining metagenome data, based on curated databases of
105 marker genes for nitrogen and methane cycle processes (see Table 1 for an overview of the used
106 marker genes). We applied this strategy to metagenome data retrieved from two depths along the
107 oxygen gradient of the Arabian Sea OMZ.
108

109 **Materials and methods**

110 Sampling and sample preparation

Samples used in this manuscript were collected in the Arabian Sea OMZ during the PASOM cruise funded by the Netherlands Organization for Scientific Research (NWO) under number 817.01.015. The sampling site and sampling procedure are described in detail in a previous study (Pitcher et al., 2011). Briefly, the site is located within the Arabian Sea (lat. 21°55.6', long. 63°10.6') and influenced by Persian Gulf Water and Red Sea Water. It is just outside the region characterized by a quasi-permanent secondary nitrite maximum (Revsbech et al., 2009). Twelve samples were taken along a depth profile (stations PA1-PA12) in the water column ranging from the ocean surface (0 m) down to 2000 m below the surface (total water depth at this site is 3010 m). Dissolved oxygen concentrations decreased from fully saturated concentrations at the surface to 3.2 μM at the core of the OMZ. The latter value reflects the detection limit of the CTD oxygen sensor, the real values are probably substantially lower as suggested by STOX (Switchable Trace amount OXygen; Revsbech et al., 2009) sensor measurements in the Arabian Sea OMZ, where reported values were ≤0.09 μM which was the detection limit of the sensor in that sampling campaign (Jensen et al., 2011).

In our sampling dataset, the depth of 170 m below ocean surface was defined as OMZ transition zone whereas the depth of 600 m belongs to the OMZ core zone at this sampling site (Pitcher et al., 2011). Oxygen concentrations increased again at 1050 m (Pitcher et al, 2011). At 170 m depth, the ammonium concentration showed a peak (0.14 μM), nitrite concentrations had peaks at 170 m and 600 m below surface (Pitcher et al., 2011). Furthermore, biomarker analysis based on the distribution of lipids and copy numbers of the 16S rRNA / *amoA* (encoding the ammonia monooxygenase) / *hdh* (encoding the hydrazine dehydrogenase) genes and/or transcripts indicated highest numbers of *Thaumarchaeota* at 170 m and highest numbers of anammox at 600 m depth (Pitcher et al., 2011). Thus, in this study we focused on two different depths in the water

column: 170 m below ocean surface (station PA2) and 600 m below ocean surface (station PA5). Main physiochemical parameters were taken from Pitcher et al. (2011) and are reported in Table S1. From these depths, large-volumes of seawater (200−1700 L) were filtered through 142-mm diameter 0.2 μm polycarbonate filters (Millipore, Billerica, MA). Filters were cut into fragments and DNA extraction was performed as described by Pitcher and co-workers (Pitcher et al., 2011). After extraction, DNA was precipitated using ice-cold ethanol, dried, and re-dissolved in 100 μl of 10 mM Tris-HCl, pH 8. Total nucleic acid concentrations were quantified spectrophotometrically (Nanodrop, Thermo Scientific, Wilmington, DE, USA) and checked by agarose gel electrophoresis for quality. Extracts were kept frozen at -80°C.

Ion Torrent library preparation and sequencing

All kits used in this section were obtained from Life technologies (Life technologies, Carlsbad, CA, USA). For both samples an identical library preparation was performed. Genomic DNA was sheared for 7 minutes using the Ion Xpress™ Plus Fragment Library Kit following the manufacturer's instructions. Further library preparation was performed using the Ion Plus Fragment Library Kit following manufacturer's instructions. Size selection of the library was performed using an E-gel® 2% agarose gel, resulting in a median fragment size of approximately 330 bp. Emulsion PCR was performed using the OneTouch 200 bp kit and sequencing was performed on an Ion Torrent PGM using the Ion PGM 200 bp sequencing kit and an Ion 318 chip, resulting in 4.9 million reads for station PA2 and 2.6 million reads for station PA5. The raw reads were submitted to the NCBI Sequence Read Archive under the accession number SRA304624.

157 Bioinformatics

158 *Quality assessment and assembly feasibility*

159 Raw sequence reads were imported into the CLC Genomics Workbench (v7.0.3, CLCbio Arhus,
160 Denmark) and end-trimmed on quality using the CLC genomics default settings (quality limit
161 0.05 and two ambiguous nucleotides allowed) and length (≥ 100 bp) resulting in 3.3 million
162 reads of station PA2 and 1.6 million reads from station PA5, which were used for subsequent
163 analyses.

164 To assess the feasibility of assembly for the analysis of the read data, the datasets were
165 assembled *de novo* using the CLC genomics workbench with word size 35 and bubble size 5000.
166 Using the data from both sampling sites combined, this resulted in 5105 contigs longer than
167 1000 bp, incorporating only 7.6% of the reads. Based on the results of the metagenome
168 assembly, we decided to employ a read based analysis strategy.

169 *16S rRNA gene analysis*

170 To extract reads matching the ribosomal small subunit for taxonomic classification, the SILVA
171 SSU RefNR99 dataset (version 115, Quast et al., 2013) was used as reference. First, the
172 metagenome reads were mapped against the SILVA dataset using the CLC Genomics
173 Workbench (mismatch penalty 2, In/Del penalty 3, 50% identity over 70% of the sequence).
174 Mapped reads were extracted and used for BLAST searches against the identical SILVA dataset
175 (E-value cut-off 10^{-6}). Positive hits were aligned using the SINA aligner (Pruesse, Peplies &
176 Glockner, 2012) and imported into the SILVA refNR99 version 115 ARB database (Quast et al.,
177 2013). Sequences were added to the existing Neighbor Joining 16S rRNA tree (including nearly
178 500,000 16S rRNA sequences) using maximum parsimony criteria without optimization of the
179 tree topology (ARB parsimony Quick add marked). All phylogenetic clusters containing reads

from our metagenomes were visually inspected and the number of reads in the respective cluster was recorded. For reads that did not cluster unambiguously to a specific group, the phylogeny of the higher taxonomic level was recorded.

Functional gene analysis

Construction of reference datasets

To screen the metagenomes for potential function, reference datasets of marker genes (amino acid sequences) were manually curated. Marker genes for nitrogen cycle processes (nitrogen fixation, nitrification, denitrification, dissimilatory nitrite reduction to ammonia (DNRA) and anammox) and marker genes for methane cycle processes (methanogenesis, methane oxidation, methylphosphonate turnover) were selected. Table 1 gives an overview of genes and gene products used in this study. Reference databases were curated for all genes individually. Where available, existing databases were used as a reference, as for the *nifH* gene (Gaby & Buckley, 2014), the bacterial *amoA*/*pmoA* gene (Dumont et al, 2014), the archaeal *amoA* gene (Pester et al., 2012) and the *mcrA* gene (Angel et al, 2011). From these ARB databases, a subset (490 *nifH* sequences, 167 bacterial *pmoA*/*amoA* sequences also including novel clades from a manuscript in revision (all sequences are provided in the supplement), 134 archaeal *amoA* sequences, 124 methanogen *mcrA* sequences) covering the main described clusters was selected from the phylogentic trees in ARB and extracted for the analysis in this study. For the *mcrA* gene, 78 sequences from described methanotrophic ANME clusters (Knittel & Boetius, 2009) were added to the selection. In addition, *mcrA* sequences from *Bathyarchaeota* (Evans et al., 2015) were downloaded from NCBI. For the remaining genes (*hao*/*hdh*, *narG*/*nxrA*, *nirS*, *nirK*, *nosZ*, *nrfA*, *hzsA*, *mmoX*, *mpnS*, *phnGHI*), sequences were collected from public databases (proteins that

203 matched to the respective Interpro family and keyword/accession number search in NCBI). The
204 amino acid sequences were downloaded and aligned using Clustal Omega (Sievers et al., 2011).
205 Alignments were manually inspected and for sequences with large gaps or insertions, literature
206 was queried to verify the function of the respective enzyme. In cases where the function of a
207 divergent sequence has not been described or is different than the target database, the sequence
208 was discarded. The remaining aligned sequences were imported into ARB. Sequences with
209 ambiguous characters or obvious shifts in reading frame were removed. The remaining
210 sequences were used for phylogenetic tree construction. Phylogenies were compared to literature
211 to verify that the known and described phylogenetic groups are covered and that the dataset is
212 representative. For the *norB/norZ* database, all sequences in the cytochrome oxidase PFAM
213 family (PF00115) were downloaded, replicates were removed using UClust (80% identity)
214 (Edgar, 2010) and the clustered sequences were screened against characterized cNOR
215 (cytochrome c dependent nitric oxide reductase) and qNOR (quinol dependent nitric oxide
216 reductase) sequences using a BLAST score ratio (BSR) (Rasko, Myers & Ravel, 2005).
217 Sequences with a BSR over 0.4 were included in the database. Representatives of phylogenetic
218 clusters (distance level of 0.1) were exported in fasta format and imported into ARB where
219 sequences with ambiguous characters or obvious shifts in reading frame were removed. The
220 remaining sequences were used for phylogenetic tree construction and exported as reference set
221 for subsequent BLAST analysis. All reference sets are provided as supplementary material.
222
223 *BLAST analysis using the reference datasets*
224 The reads of the metagenomes described above were used in BLASTx searches against the
225 reference sets (E-value cutoff 10^{-6}). The E-value was established after first test analyses with

different E-values (10^{-4} to 10^{-6}). Positive hits were extracted using a custom perl script available at www.github.com/dspeth. In a second step, positive reads were used in BLASTx searches against the NCBI NR database. For reference datasets resulting in low number of target sequences (approximately n<100), false positive reads were removed by manual inspection of the BLAST results. For large number of target reads (approx. n>100), a modified BLAST score ratio (BSR) approach was used to remove false positive hits while keeping divergent sequences. In this approach, rather than calculating the ratio between the score of a hit against our database and the maximum score (of a self-hit) for every read (Rasko, Myers & Ravel, 2005), we calculated the ratio between a hit against our database and a hit against the NCBI-NR database (NCBI non-redundant protein sequence database, see Figure S2 for the procedure overview). This guaranteed that sequences with low similarity to both NR and our database were kept as true positives, whereas reads with a much better hit to the NR than to our database were discarded as false positives. We determined a separate threshold value for a positive hit for each specific gene database (Table 1). This threshold was selected based on a first manual inspection of reads with different bit-score ratios (Figures S3-S7). To distinguish between the closely related *hao* (encoding the hydroxylamine oxidoreductase) and *hdh* (encoding the hydrazine dehydrogenase) sequences and the *narG* (encoding the nitrate reductase) and *nxrA* (encoding the nitrite oxidoreductase) sequences respectively, a second step was applied. Positive *hao* reads obtained after the BSR analysis were mapped in CLC genomics (mismatch penalty 2, In/Del penalty 3, minimum 50% identity over 50% of the read) against both copies of the Scalindua brodae *hdh* (accession numbers: KHE92657.1, KHE91265.1). Mapped reads were extracted and classified as *hdh*, non-mapped reads were classified as *hao*. To distinguish between *narG* and *nxrA*, positive *narG* reads obtained after the first BSR analysis were used in a second round of

BLASTx against two *nxrA* gene subsets, one containing the *Nitrobacter/Nitrococcus/Nitrolancea* *nxrA* sequences and the other containing the *Nitrospira/Nitrospina/*anammox *nxrA* sequences. A BSR cut-off of 0.85 was used to separate *nxrA* from *narG*.

For all genes, after removal of false positive targets using the BSR analysis, reads were assigned to taxonomy using MEGAN (Huson et al., 2007, Huson et al., 2011) and the taxonomy of maximum 5 hits with over 90% score from the top hit.

Normalization of metagenomic read counts

After phylogenetic assignment, the positive read counts were normalized according to gene length and metagenome size according to the following procedure: (Number of positive reads) / ((length of gene in bp) * (number of total reads in the metagenome)). The *rpoB* gene (encoding the RNA polymerase) was used as a single copy gene reference. To estimate the abundance of microorganisms encoding nitrogen or methane cycling marker genes in their genomes, the normalized read counts were discussed as fraction of the normalized total read counts of the *rpoB* gene.

Assembly of selected reads

For selected genes (*hdh, hzsA, nirS, nrfA, nxrA*), the reads were extracted and imported in the CLC genomics workbench to reconstruct (near) full-length genes. In all cases, reconstruction through direct assembly (word size 35, bubble size 5000) was attempted. This was successful for *hzsA, nirS* and *nrfA,* where diversity was limited. For the more diverse data (*nxrA* and *hdh*) iterative mapping (Dutilh, Huynen & Strous, 2009) was used. *Hdh* could be reconstructed by mapping against both copies of *Scalindua brodae hdh* (KHE92657.1 & KHE91265.1; mismatch

penalty 2, In/Del penalty 3, minimum 50% identity over 50% of the read). The *nxrA* sequences of *Nitrospina sp.* and *Scalindua sp.* were reconstructed by iterative mapping on the sequences of *Nitrospina gracilis* and *Scalindua brodae* (WP_042250442.1 & KHE93157.1; mismatch penalty 2, In/Del penalty 3, minimum 80% identity over 50% of the read). Novel *nxrA* sequences were retrieved using iterative mapping of the reads that could not be assigned to either *Nitrospina* or *Scalindua* on the *nxrA* sequence of *Kuenenia stuttgartiensis* (CAJ72445.1; mismatch penalty 1, In/Del penalty 3, minimum 30% identity over 50% of the read).

Results and discussion

Metagenomics can be used as powerful tool to gain insights into the functional potential of an ecosystem. As the sequencing procedure itself is becoming easier and less expensive resulting in generation of large amounts of data, sequence analysis strategies are becoming a bottleneck in time and resources. Depending on the diversity and complexity of the dataset, different analysis approaches are needed. In this study, we provide a strategy for the systematic screening of metagenomes for nitrogen and methane cycling potential using curated functional gene reference databases. We applied our strategy to a dataset from the Arabian Sea oxygen minimum zone (OMZ) analysing the genetic potential for nitrogen and methane turnover in the upper limit (station PA2) characterized by low oxygen (approximately 5 μM, Table S1) and the core zone (station PA5) in which the oxygen concentration drops below the detection limit (3.4 μM) (Table S1).

Taxonomy based on 16S rRNA gene analysis

To get insight into the overall microbial community in the Arabian Sea OMZ, we analysed 16S rRNA gene reads retrieved from the metagenomes at station PA2 and PA5 (Figure 1). At the upper limit of the OMZ (station PA2, Table S1), the SAR11 clade (*Alphaproteobacteria*) formed one of the most abundant microbial groups (14%). This clade represents in general one of the most abundant microorganisms in seawater, contributing up to 30% of all bacterioplankton (Morris et al., 2005). Sub-clusters within the SAR11 clade have been linked to ecotypes occupying different niches in the ocean water column (Field et al., 1997; Vergin et al., 2013). In the PA2 dataset, most sequence reads clustered within SAR11 subgroup 1, closely related to cultivated strains of '*Candidatus* Pelagibacter ubique'. '*Ca.* P. ubique' has a small, streamlined genome adapted to rapid heterotrophic growth (Rappe et al., 2002; Giovannoni et al., 2005) and

is unlikely to directly contribute to nitrogen cycling in the Arabian Sea OMZ. The remainder of the SAR11 reads were distributed across the entire SAR11 clade. Ammonium oxidizing *Archaea* (AOA) of the Marine Group I (MG-I) *Thaumarchaeota* were as abundant as SAR11 at station PA2 (14%), confirming previous PCR- and lipid-based analyses showing that *Thaumarchaeota* were abundant at this location (Pitcher et al., 2011). Both SAR11 and MG-I were less abundant in the OMZ core, but still have a substantial presence of 8.1% and 3.3% respectively (Figure 1). Reads affiliated with the bacterial SAR86 clade (*Gammaproteobacteria*) and archaeal Marine Group II (MG-II) made up 8% and 7% of the 16S rRNA gene reads at station PA2, but were only marginally present in the OMZ core (1.5 % and 0.4 % of the reads, respectively). This is consistent with an aerobic heterotrophic lifestyle predicted from previously obtained genomes of organisms of both lineages (Dupont et al., 2012; Iverson et al., 2012). So far, no metabolic adaptations of these organisms to an anaerobic lifestyle have been characterized. Nevertheless, the persistence, albeit in low abundance, of aerobic organisms in the anoxic OMZ core might be partially explained by the attachment to slowly sinking organic particles, also referred to as marine snow (Wright, Konwar & Hallam, 2012). Formed in the metabolically active photic zone, these particles continuously sink through the water column to the floor of the ocean and thereby also passing the anoxic core of the OMZ. Association to particles has been previously proposed for SAR11 (Zeigler Allen et al., 2012) and demonstrated for the MG-II *Archaea* (Orsi et al., 2015).

The deep-branching bacterial phylum *Marinimicrobia* (formerly SAR406) (Fuhrman, Mccallum & Davis, 1993; Gordon & Giovannoni, 1996; Rinke et al., 2013) comprised 9% of all 16S rRNA affiliated sequence reads at station PA2 and 20% at PA5. A recent transcriptome study indicates the involvement of these bacteria in extracellular proteolysis and fermentative amino acid

325 degradation in a methanogenic environment (Nobu et al., 2015). These findings of an anaerobic
326 lifestyle agree well with their high abundance in a low oxygen environment (upper OMZ) and
327 their dominance in the OMZ core. The detected *Rhodospirillales* (*Alphaproteobacteria*) might
328 also contribute to fermentation of organic matter, as this order comprises known acetic acid
329 bacteria. *Rhodospirillales* were barely detectable at the OMZ upper limit, but comprise 4.1% of
330 the community at the OMZ core. Other abundant lineages include the *Deltaproteobacteria*
331 SAR324. These bacteria are frequently found in the ocean and seem to be correlated with low
332 oxygen concentrations (Wright, Konwar & Hallam, 2012; Sheik, Jain & Dick, 2014). SAR324
333 representatives have been shown to be able to fix CO_2 (Swan et al., 2011) and are predicted to be
334 capable of autotrophic denitrification with various electron donors (Sheik, Jain & Dick, 2014).
335 Likewise, *Gammaproteobacteria* of the SUP05 lineage have been implied in autotrophic
336 denitrification with sulphur as electron donor (Lavik et al., 2009; Walsh et al., 2009; Russ et al.,
337 2014). They furthermore have been associated with cryptic sulfur cycling in the ETSP OMZ
338 (Canfield et al., 2010). Members of both SAR324 and SUP05 have a similar presence at both
339 stations (5% in PA2 and 3.1% in PA5 for the SAR324 lineage, 2.9% and 2.7% for the SUP05
340 lineage), with SAR324 slightly decreasing in the OMZ core (Figure 1).
341 Anammox 16S rRNA genes are barely detected at the upper limit of the OMZ, consistent with
342 previous analyses (Pitcher et al., 2011; Villanueva et al., 2014), but comprise almost 5% of the
343 community at the OMZ core. Interestingly, a recent study showed that autotrophic denitrifiers
344 oxidizing hydrogen sulphide could form a stable community with anammox bacteria in a reactor
345 system (Russ et al., 2014). Whether this is also the case at the OMZ core remains to be
346 investigated. Although some of the lineages discussed above are abundant, assembly of the

complete metagenome reads did not yield any contigs with high sequencing depth, indicating the diversity within each lineage is substantial.

To assess the diversity and phylogeny of the detected AOA and anammox bacteria in more detail, reads matching the 16S rRNA gene of either group were extracted and assembled into contig sequences for phylogenetic tree construction. Reads affiliated with MG-I AOA from the OMZ upper limit (PA2) could be assembled into two representative contigs: Contig-1 comprises 41% of all extracted MG-I reads and contig-2 was built from 15% of the reads. Thus, the two contigs represent the majority of the thaumarchaeotal community. Nevertheless, these sequences do not represent a single genotype but represent a hybrid of 16S rRNA reads from multiple closely related organisms (Figure S1). Contig-1 shows 98% identity to the very recently described '*Candidatus* Nitrosopelagicus brevis' (Figure 2) (Santoro et al., 2015), contig-2 has only moderate identities to isolated or enriched ammonia oxidizers (93% identity to *Nitrosoarchaeum limnia* SFB1) (Figure 2).

Reads from station PA5 clustering with the *Brocadiaceae,* the family comprising all known anammox bacteria, were extracted and could be assembled into one representative sequence. Here, diversity was considerably less than for the *Thaumarchaeota* and the contig represents a single genotype that is 96% identical to *Scalindua brodae,* the closest sequenced relative (Speth et al., 2015), and 99% identical to '*Candidatus* Scalindua arabica' clones previously obtained from the Arabian sea OMZ (Figure 3) (Woebken et al., 2008). Interestingly, sequences obtained by Ward and co-workers from the Arabian Sea OMZ share only 97% sequence identity to the extracted contig (Ward et al., 2009) indicating a spatial or temporal niche differentiation of different *Scalindua*-like ecotypes in this system.

371

372 Genetic potential for nitrogen and methane cycling

373 To assess the nitrogen and methane cycling potential in the metagenome of both the upper limit
374 and the core of the OMZ, we performed BLASTx searches of the metagenomic reads against
375 curated databases of key genes (Table 1) involved in nitrogen and methane cycle processes. To
376 remove false positive hits while keeping divergent sequences, we used a modified BLAST score
377 ratio (BSR) approach (see methods section).

378

379 Nitrogen cycling potential

380 *Nitrification*

381 Marker genes indicative for the first step in nitrification, the conversion from ammonia to nitrite,
382 are the *amoA* (encoding a subunit of the membrane bound ammonium monooxygenase) and the
383 *hao* gene (encoding the hydroxylamine oxidoreductase). We found 228 reads matching *amoA* in
384 the PA2 dataset, 227 of which could be assigned to ammonia oxidizing Archaea (AOA),
385 indicating that they contribute approximately 25% to the total microbial community in this
386 sample (Figure 4). This estimate exceeds the estimated abundance based on 16S rRNA genes
387 (Figure 1). Our analysis strategy includes the correction for sequencing depth and gene length,
388 however, no correction for gene copy numbers was applied. This information can only be
389 deduced from genomes and is not known for the vast majority of microorganisms. In the
390 presence of genomes harbouring multiple rRNA operons, the total number of detected organisms
391 is artificially inflated, leading to an underestimation of organisms with a single rRNA operon
392 which would explain the lower 16S rRNA estimates in this dataset. An alternative, albeit less
393 likely, explanation for higher *amoA* estimates can be the presence of multiple copies of the *amoA*

394 gene in the detected AOA genomes, but this has not been observed in any previously sequenced
395 AOA species.
396 The *amoA* reads could be assembled into 5 major contigs (4 contigs from site PA2 and 1 contig
397 from PA5) that were compared to 16S rRNA phylogeny (Figure 1). As for the 16S rRNA contig,
398 *amoA* sequences could not be affiliated with a single MG-I species, but showed a diversity of at
399 least 2 major genotypes. Unlike the 16S rRNA gene analysis, none of the contigs clustered with
400 *Nitrosopelagicus*-like *amoA* sequences. Instead two out of the four obtained contigs from PA2
401 did not cluster with, but between the *Nitrosopelagicus* cluster and *Nitrosoarchaeum limnia*
402 SFB1. The two other contigs clustered within an environmental group only distantly related to
403 described MG-I AOA. This environmental cluster also contained the only contig that could be
404 assembled from the OMZ core. A niche differentiation between shallow and deep-water clades
405 of MG-I *Thaumarchaeota* has also been described before in the Arabian Sea OMZ (Villanueva,
406 Schouten & Sinninghe Damsté, 2014), but also in other marine environments (Beman, Popp &
407 Francis, 2008; Santoro, Casciotti & Francis, 2010). Besides archaeal *amoA*, only a single read
408 with low identity (<50% on the amino acid level) to known bacterial *amoA* reads was detected,
409 indicating ammonium oxidizing bacteria (AOB) likely play only a small role in the Arabian sea
410 OMZ although they were detected in other OMZs (Molina et al., 2007; Lam et al., 2009).
411 Consistent with the absence of AOB *amoA,* only five reads matching *hao* were detected in the
412 upper limit of the OMZ. In the core, 475 reads matched the *hao* database, but even after removal
413 of the bonafide anammox hydrazine dehydrogenase hits (Figure 3), over 90% of the *hao* matches
414 were affiliated with *Scalindua,* which is known to encode up to ten paralogs of this protein (van
415 de Vossenberg et al., 2013; Speth et al., 2015).

The second step of complete nitrification, nitrite oxidation, is challenging to study using a marker gene approach as nitrite oxidoreductase (*NxrA*) and nitrate reductase (*NarG*) are homologous enzymes. A further complicating factor is the polyphyletic nature of the *nxrA* gene (Lücker et al., 2010). To account for this, we first extracted all the reads matching a combined *narG* / *nxrA* reference set, and then used a second round of BLASTx and BLAST score ratio separation (threshold 0.85-0.95) to distinguish between *narG* and *nxrA*. No reads could be confidently assigned to *nxrA* of the *Nitrobacter / Nitrococcus / Nitrolancetus* group in either station. Conversely, 33% and 23% of the reads matching *narG / nxrA* were assigned to *nxrA* of the *Nitrospira / Nitrospina /* anammox group in PA2 and PA5, respectively. Further separation between *nxrA* matches, to distinguish the nitrifier *nxrA* from anammox, was achieved using iterative mapping. Classification using MEGAN indicated that anammox made up 44% of the reads matching *nxrA* at station PA2. As all other analyses indicated anammox bacteria were virtually absent from this station (Figure 1 and Table S2), we explored these reads in detail using iterative mapping and manual curation of the sequences. This led to the identification of a novel lineage of *nxrA* clustering between anammox and *Nitrospina* sequences (Figure 5), which was slightly more abundant than the retrieved *Nitrospina sp.* at both station PA2 (approx. 2 % and 1.5 % of the population respectively) and PA5 (approx. 3 % and 2 % of the population respectively). Interestingly, distinct lineages of both the putative novel nitrite oxidizer and *Nitrospina* seemed to occupy either station (Figure 5). The abundance of anammox *nxrA* in the OMZ core sample correlates well with the abundance as assessed using the 16S rRNA gene (Table 1) and other anammox markers (Table S3, discussed below).

437 No other nitrite oxidizing organisms were detected. The detection of a significant abundance of
438 nitrite oxidizers in OMZ ecosystems is consistent with a previous study showing that nitrite
439 oxidation was an active process in the Namibian OMZ (Füssel et al., 2012).
440
441
442 *Genetic potential for processes contributing to nitrogen loss*
443 Nitrite and nitrate resulting from nitrification can be readily used in denitrification, anammox
444 and DNRA. To date, it is still unclear which process is dominating or if a combination of various
445 processes occurs. Most of the recent studies indicated anammox instead of denitrification as
446 prevalent pathway in OMZs (Kuypers et al., 2005; Thamdrup et al., 2006; Hamersley et al.,
447 2007; Lam et al., 2009). However, two reports described high and active denitrification rates in
448 the Arabian Sea OMZ (Ward et al., 2009; Bulow et al., 2010). In yet another study, Jensen and
449 co-workers found that anammox coupled to DNRA was the prevalent process in this system
450 (Jensen et al., 2011) and further studies confirmed a high abundance of anammox bacteria at the
451 core of the Arabian Sea OMZ (Pitcher et al., 2011; Villanueva et al., 2014).
452 Here, we found that the nitrate reductase is by far the most dominant nitrogen cycle enzyme
453 encoded in the Arabian Sea OMZ core (78% of normalized *rpoB* gene abundance, Figure 4).
454 This is consistent with observations in the Peruvian oxygen minimum zone (Lam et al., 2009;
455 Glass et al., 2015) and in the Eastern Tropical South Pacific OMZ (Stewart, Ulloa & DeLong,
456 2012). The reduction of nitrate to nitrite is a crucial step as the nitrite forms the starting point for
457 many subsequent processes: nitrite reduction in denitrification, in DNRA and in anammox.
458 Additionally, nitrite can also be re-oxidized to nitrate. The genetic potential for all these
459 processes is substantially encoded in the OMZ core. Consistent with Jensen and co-workers

(Jensen et al., 2011), we find the potential for DNRA (*nrfA* encoding the penta-heme nitrite reductase as marker gene), approximately in equal abundance to anammox. Upon close inspection 60% of the reads matching *nrfA* originate from two closely related strains of an unknown organism. The 812 bp hybrid sequence obtained after assembly has 73% identity (AA level) to only two sequences in the database: *Coraliomargerita akajimensis* (*Verrucomicrobia*) and *Pelobacter carbinolicus* (*Deltaproteobacteria*). Although the phylogeny of the organisms most likely responsible for DNRA in the Arabian Sea OMZ remains unclear, retrieval of the divergent *nrfA* emphasizes the potential of our approach for novel microbiological gene discovery.

Only few sequences indicative for the process of denitrification were detected. Although many reads matched nitrite reductase encoded by either *nirS or nirK* (Table S3), 50% of all *nirS* reads could be classified as *Scalindua*-related. The *nirS* contig obtained after assembly of the reads matching *Scalindua* showed 77% nucleotide sequence identity to *Scalindua brodae* and 99% identity to unpublished sequences from the Gulf of California and Eastern Tropical North Pacific OMZs (Genbank accession: KC596869). It corroborates the 16S rRNA gene analysis indicating the presence of one dominant uncultivated anammox ecotype. Of all detected *nirK* reads, 70 % originated from AOA. These *Thaumarchaeota* are known to harbour multiple copies of *NirK*-like copper oxidases encoding genes (Stahl & de la Torre, 2012) explaining their high abundance in the datasets (Figure 4). It has been previously suggested that nitrifier denitrification accounts for the majority of nitrous oxide observed in the ocean (Babbin et al., 2015, Kozlowski et al., 2016). Besides playing a role in nitrifier denitrification, these enzymes are also hypothesized to be the missing hydroxylamine oxidoreductase equivalent in AOA catalysing the oxidation of hydroxylamine to nitrite (Stahl & de la Torre, 2012).

Similar to the *nirS*-type nitrite reductase genes, we found that 40% of all reads matching the *norB / norZ* reference set (encoding the nitric oxide reductase) at both stations combined were affiliated with *Scalindua* (Table S3). Potential for the final step of denitrification via nitrous oxide reduction (encoded by the *nos* genes) was limited with 92 reads at station PA2 and only 27 reads at station PA5 (Figure 4, Table S3). We could not find genomic indication for nitrogen loss via the nitric oxide dismutase of NC10 phylum bacteria as recently described for the Eastern Pacific OMZ (Padilla et al., 2016).

The abundance of anammox marker genes *hzsA* (encoding the hydrazine synthase) and *hdh* (encoding the hydrazine dehydrogenase) coincides well with the abundance estimates based on 16S rRNA, *nxrA*, and *nirS* genes consistently indicating that anammox bacteria of the *Scalindua* genus are present at approximately 5% abundance.

The *nifH* gene, encoding a subunit of the nitrogenase, was used as marker in screening for nitrogen fixation potential in the Arabian Sea datasets. Ocean circulation models have predicted highest nitrogen fixation rates close to zones of nitrogen loss (Deutsch et al., 2007) and were supported by recent studies reporting the presence and transcription of *nifH* genes in sub-oxic waters of the Arabian Sea (Jayakumar et al., 2012; Bird & Wyman, 2013). However, we detected no *nifH* hits in the PA2 metagenome and only 3 reads matched *nifH* in the PA5 dataset (Figure 3). The *nifH* genes in the above mentioned studies were obtained using PCR amplification with specific primer sets able to detect a much lower abundance of diazotrophs in the environment. Our results indicate a low abundance of diazotrophs that nevertheless would easily be detected by PCR amplification. We can estimate the abundance of diazotrophs from our dataset, assuming an average microbial genome size of 3 Mbp and the *nifH* gene length of 900 bp. If present in all genomes, 3 reads per 10.000 should contain part of a *nifH*. However, only 3 reads were detected

506 in the PA5 metagenome (1,6 million reads), which is 160 times lower and thus results in a
507 diazotroph abundance estimate of 0.6% based on the dataset. PCR primers should be able to
508 amplify genes present in organisms with this abundance. Alternatively, the diazotrophic
509 community in our dataset was, for unknown reasons, lower than in other studies from similar
510 ecosystems.
511
512 Methane cycling potential
513 To examine the methane cycling potential in the Arabian Sea, we selected marker genes
514 targeting methane production and methane oxidation (Table 1). The *mcrA* gene (encoding the
515 methyl-Coenzyme M reductase) was used as functional marker for methane production and
516 anaerobic methane oxidation. We could not detect any *mcrA*-like sequences of canonical
517 methanogens or sequences of the very recently described *mcrA* homologues from
518 *Bathyarchaeota* (Evans et al., 2015). Also no sequences of anaerobic methane oxidizing archaea
519 (ANME clades) were found. This is consistent with the absence of methanogen 16S rRNA genes
520 and very low abundance (3 reads) of *Bathyarchaeota* 16S rRNA gene sequences. Thus,
521 methanogenesis does not seem to play a major role in the Arabian Sea OMZ water column.
522 Nevertheless, in marine ecosystems, a second pathway for methane production apart from
523 methanogenesis was proposed (Karl et al., 2008). This aerobic pathway includes the cleavage of
524 methylphosphonate (Mpn). Mpn was shown to be synthesized by the AOA *Nitrosopumilus*
525 *maritimus* using the Mpn synthase as key enzyme and a homologue of this enzyme was also
526 found to be encoded in SAR11 genomes (Metcalf et al., 2012). The cleavage of Mpn that results
527 in the release of methane, is catalysed by the C-P lyase multi-enzyme complex (Daughton, Cook
528 & Alexander, 1979). In *Escherichia coli*, the C-P cleavage is encoded in a 14 gene operon

(Metcalf & Wanner, 1993). The composition of the gene cluster is variable among bacteria, however, the *phnGHIJKM* genes seem to be conserved and essential for activity (Huang, Su & Xu, 2005). We selected the *mpnS* gene (encoding the Mpn synthase) and the *phnGHI* genes (encoding components of the C-P lyase pathway) as functional marker genes for aerobic methane production in our datasets. In the PA2 dataset, 45 reads matching the *mpnS* gene were retrieved, most of which could be affiliated with the MG-I *Thaumarchaeota* (Figure 6, Table S4). Comparing the *mpnS* read numbers with corresponding 16S rRNA gene and *amoA* read numbers from *Thaumarchaeota* results in a far lower abundance of *mpnS* reads in the dataset. However, this is in good agreement with the observation that not all MG-I *Thaumarchaeota* encode *mpnS* in their genome. The dominant MG-I ammonia oxidizer in the Arabian Sea datasets is closely related to *Nitrosopelagicus brevis* (Figure 2), which does not possess the *mpnS* gene (Santoro et al., 2015). In the metagenome from the OMZ core zone, 7 *mpnS* reads (3 reads matching to MG-I) were detected.

The C-P lyase gene cluster has been found in many genomes of marine bacteria and could also be associated to growth with Mpn as sole phosphorous source (Dyhrman et al., 2006; Martinez, Tyson & DeLong, 2010; White et al., 2010). In the Arabian Sea datasets, no reads matching the *phnGHI* genes were detected in PA2 and only 3 *phnI* sequences affiliated to *Rhodobacteriales* (*Alphaproteobacteria*) were retrieved from PA5 (Figure 6, Table S4). Depending on the dissolved inorganic phosphorous availability, the abundance of bacteria harbouring the gene cluster can vary between as much as 20% and below 1% of all bacteria (Martinez, Tyson & Delong, 2010), indicating that the presence of the C-P lyase provide an advantage in phosphorous limited environments. We could not find a high abundance of this protein complex

551 in the Arabian Sea OMZ, hence the ability for acquiring phosphorous from Mpn might be less
552 important in this system.
553 In accordance with the absence of marker genes indicative for methane production, only few
554 reads were retrieved that could be affiliated with methane oxidation. For aerobic methane
555 oxidation, we used the marker genes *pmoA* and *mmoX* encoding the particulate and soluble
556 methane monooxygenase. In the PA2 dataset, 15 *mmoX* and 3 *pmoA*-like reads were detected.
557 However, after closer inspection, all *mmoX* reads showed moderate sequence identity (up to 71%
558 on amino acid level) to homologous toluene monoxygenases. For the *pmoA* reads, highest
559 identity (68-97% on amino acid level) was found to the monooxygenase of the SAR324 clade
560 (*deltaproteobacteria*). Based on 16S rRNA gene abundance, the SAR324 clade was found to
561 belong, besides others, to the dominant bacteria in the Arabian Sea dataset (Figure 1). So far, no
562 enrichment of SAR324 clade bacteria is available and it is not known if this monooxygenase is
563 used for methane or higher alkane oxidation. Whereas the *pmoA* phylogeny indicates relation to
564 C2-C4 alkane monooxygenases (Li et al., 2014), the genomes of SAR324 members suggest the
565 potential for both, C1 and higher alkane utilization (Sheik, Jain & Dick, 2014). Nevertheless, not
566 all SAR324 genomes contain the alkane monooxygenase gene cluster (Swan et al., 2011). The
567 PA5 dataset revealed 2 *pmoA* and 2 *mmoX*-like sequences, again only with moderate sequence
568 identities to known alkane monooxygenases. Consistent with the absence of genes encoding the
569 putative nitric oxide dismutase of NC10 phylum bacteria, we did not find any *pmoA* genes of
570 these bacteria in the Arabian sea OMZ. Nevertheless, they might be present in a low abundance
571 that could not be retrieved by our approach. Thus, if these bacteria play a role in methane and
572 nitrogen cycling in OMZs, as suggested for the Eastern Pacific OMZ, needs to be investigated in
573 future research (Padilla et al., 2016). In addition, 16S rRNA gene sequences of known aerobic

574 methanotrophs were nearly absent. Only 3 reads clustering within gammaproteobacterial
575 methanotrophs were present (1 read from PA2 and 2 reads from PA5). Although the overall
576 abundance of hydrocarbon monooxygenase encoding reads is low in the Arabian Sea OMZ
577 dataset, our analysis shows the existence of novel sequence diversity only moderately related to
578 known sequences, that is not captured by currently used PCR primers.
579
580
581
582 Conclusion
583 In this study, we compared the functional diversity in two metagenomes retrieved from the
584 Arabian Sea oxygen minimum zone. Using manually curated reference databases, we screened
585 the datasets for homologues indicative for nitrogen and methane turnover in this ecosystem. We
586 are aware that the presence of genetic potential alone cannot be used to draw conclusions on
587 activity of various processes. However, despite this limitation, the picture that emerges from our
588 analysis is that the vast majority of organisms can contribute to nitrate reduction, probably
589 coupled to degradation of organic matter and release of ammonium (Kalvelage et al., 2015). The
590 nitrite formed by nitrate reduction can be re-oxidized to nitrate in a 'nitrite loop' ultimately
591 resulting in removal of additional organic matter and release of more ammonium. The *nrf*-like
592 nitrite reductase assembled from our dataset is only distantly related to described *nrf* sequences
593 indicating that a yet unknown organism is responsible for this reaction in the Arabian Sea. The
594 released ammonium can partially be oxidized by a diverse community of microaerophilic
595 archaeal ammonium oxidizers, but in the core of the OMZ, the majority is likely converted by
596 anammox bacteria, which contribute about 5% to the total abundance at the station PA5. Here,

we could confirm the presence of a low diversity dominated by a Candidatus '*Scalindua arabica*'-like anammox species as observed in previous studies (Villanueva et al., 2014, Woebken et al., 2010). Although denitrification was observed as the dominant process in another Arabian Sea study (Ward et al., 2009), our analysis, albeit only based on the genetic potential, does not support this point. Rather, an intricate nitrogen cycle involving many organisms and the exchange of intermediates and connection to other processes, as recently hypothesized for estuary and an aquifer (Hug et al, 2015, Baker et al., 2015) seems likely. The ultimate removal of nitrogen is most likely mediated by anammox. We found no evidence for methane turnover in the Arabian Sea OMZ, however, the metagenomes revealed the presence of new alkane monooxygenase diversity in this ecosystem.

Acknowledgements

We thank the department of marine organic biogeochemistry and molecular lab facilities of the Royal NIOZ for support.

610 **References**

611 Angel R, Matthies D, Conrad R. 2011. Activation of Methanogenesis in Arid Biological Soil
612 Crusts Despite the Presence of Oxygen. *Plos One* 6: e20453.
613 Babbin AR, Bianchi D, Jayakumar A, Ward BB. 2015. Rapid nitrous oxide cycling in the
614 suboxic ocean. *Science* 348:1127–1129.
615 Bange HW, Ramesh R, Rapsomanikis S, Andreae MO. 1998. Methane in surface waters of the
616 Arabian Sea. *Geophysical Research Letters* 25:3547–3550.
617 Beman JM, Popp BN, Francis C a. 2008. Molecular and biogeochemical evidence for ammonia
618 oxidation by marine Crenarchaeota in the Gulf of California. *The ISME Journal* 2:429–441.
619 Bird C, Wyman M. 2013. Transcriptionally active heterotrophic diazotrophs are widespread in
620 the upper water column of the Arabian Sea. *FEMS Microbiology Ecology* 84:189–200.
621 Bulow SE, Rich JJ, Naik HS, Pratihary AK, Ward BB. 2010. Denitrification exceeds anammox
622 as a nitrogen loss pathway in the Arabian Sea oxygen minimum zone. *Deep-Sea Research*
623 *Part I: Oceanographic Research Papers* 57:384–393.
624 Canfield DE, Stewart FJ, Thamdrup B, De Brabandere L, Dalsgaard T, Delong EF, Revsbech
625 NP, Ulloa O. 2010. A cryptic sulfur cycle in oxygen-minimum-zone waters off the Chilean
626 coast. *Science (New York, N.Y.)* 330:1375–1378.
627 Codispoti L a, Brandes J a, Christensen JP, Devol a H, Naqvi SW a, Paerl HW, Yoshinari T.
628 2001. The oceanic fixed nitrogen and nitrous oxide budgets : Moving targets as we enter the
629 anthropocene? *Scientia Marina* 65:85–105.
630 Codispoti L a., Richards F a. 1976. An analysis of the horizontal regime of denitrification in the
631 eastern tropical North Pacific. *Limnology and Oceanography* 21:379–388.
632 Daughton CG, Cook AM, Alexander M. 1979. by using standard methods [3]. The enrichment
633 and isolation procedures have been described [4]. 5:91–93.
634 Deutsch C, Sarmiento JL, Sigman DM, Gruber N, Dunne JP. 2007. Spatial coupling of nitrogen
635 inputs and losses in the ocean. *Nature* 445:163–167.
636 Devol a. H, Uhlenhopp a. G, Naqvi SW a., Brandes J a., Jayakumar D a., Naik H, Gaurin S,
637 Codispoti L a., Yoshinari T. 2006. Denitrification rates and excess nitrogen gas
638 concentrations in the Arabian Sea oxygen deficient zone. *Deep Sea Research Part I:*
639 *Oceanographic Research Papers* 53:1533–1547.
640 Dumont MG, Lüke C, Deng Y, Frenzel P. 2014. Classification of pmoA amplicon pyrosequences
641 using BLAST and the lowest common ancestor method in MEGAN . *Frontiers in*
642 *Microbiology* **5**:34.
643 Dupont CL, Rusch DB, Yooseph S, Lombardo M-J, Alexander Richter R, Valas R, Novotny M,
644 Yee-Greenbaum J, Selengut JD, Haft DH, Halpern AL, Lasken RS, Nealson K, Friedman R,
645 Craig Venter J. 2012. Genomic insights to SAR86, an abundant and uncultivated marine
646 bacterial lineage. *The ISME Journal* 6:1186–1199.
647 Dutilh BE, Huynen MA, Strous M. 2009. Increasing the coverage of a metapopulation consensus
648 genome by iterative read mapping and assembly. *Bioinformatics* 25:2878–2881.
649 Dyhrman ST, Chappell PD, Haley ST, Moffett JW, Orchard ED, Waterbury JB, Webb EA. 2006.
650 Phosphonate utilization by the globally important marine diazotroph Trichodesmium.
651 *Nature* 439:68–71.
652 Edgar RC. 2010. Supplementary Material - Search and clustering orders of magnitude faster than
653 BLAST. *Bioinformatics* 26:2460–2461.

654 Evans PN, Parks DH, Chadwick GL, Robbins SJ, Orphan VJ, Golding SD, Tyson GW. 2015.
655 Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric
656 metagenomics. 350.
657 Field KG, Gordon D, Wright T, Rappé M, Urbach E, Vergin K, Giovannoni SJ. 1997. Diversity
658 and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic
659 bacteria. *Applied and Environmental Microbiology* 63:63–70.
660 Frigaard N-U, Martinez A, Mincer TJ, DeLong EF. 2006. Proteorhodopsin lateral gene transfer
661 between marine planktonic Bacteria and Archaea. *Nature* 439:847–850.
662 Fuhrman JA, McCallum K, Davis AA. 1993. Phylogenetic diversity of subsurface marine
663 microbial communities from the Atlantic and Pacific Oceans. *Applied and Environmental*
664 *Microbiology* 9:1294–1302.
665 Füssel J, Lam P, Lavik G, Jensen MM, Holtappels M, Günter M, Kuypers MM. 2012. Nitrite
666 oxidation in the Namibian oxygen minimum zone. *The ISME Journal* 6:1200–1209.
667 Gaby JC, Buckley DHA. 2014. A comprehensive aligned *nifH* gene database: a multipurpose
668 tool for studies of nitrogen-fixing bacteria. *Database-Oxford* bau001.
669 Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, Bibbs L, Eads J,
670 Richardson TH, Noordewier M, Rappé MS, Short JM, Carrington JC, Mathur EJ. 2005.
671 Genome Streamlining in a Cosmopolitan Oceanic Bacterium. *Science* 309 :1242–1245.
672 Glass JB, Kretz CB, Ganesh S, Ranjan P, Seston SL, Buck KN, Landing WM, Morton PL,
673 Moffett JW, Giovannoni SJ, Vergin KL, Stewart FJ. 2015. Meta-omic signatures of
674 microbial metal and nitrogen cycling in marine oxygen minimum zones. *Frontiers in*
675 *Microbiology* 6:1–13.
676 Gordon D a, Giovannoni SJ. 1996. Stratified microbial populations related to Chlorobium and
677 Fibrobacter detected in the Atlantic and Pacific Oceans. *Appl. Environ. Microbiol.* 62:1171–
678 1177.
679 Groffman PM, Altabet M a., Bohlke JK, Butterbach-Bahl K, David MB, Firestone MK, Giblin
680 AE, Kana TM, Nielsen LP, Voyteck M a. 2006. Methods for Measuring Denitrification :
681 *Ecological Applications* 16:2091–2122.
682 Gruber N, Sarmiento JL. 1997. Global patterns of marine nitrogen fixation and denitrification.
683 *Global Biogeochemical Cycles* 11:235.
684 Hamersley MR, Lavik G, Woebken D, Rattray JE, Lam P, Hopmans EC, Damsté JSS, Krüger S,
685 Graco M, Gutiérrez D, Kuypers MMM. 2007. Anaerobic ammonium oxidation in the
686 Peruvian oxygen minimum zone. *Limnol. Oceanogr.* 52:923–933.
687 Huang J, Su Z, Xu Y. 2005. The Evolution of Microbial Phosphonate Degradative Pathways.
688 *Journal of Molecular Evolution* 61:682–690.
689 Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. *Genome*
690 *Research* 17:377-386.
691 Huson DH, Mitra S, Ruscheweyh H-J, Wever N, Schuster SC. 2011. Integrative analysis of
692 environmental sequences using MEGAN4. *Genome Research* 21:1552-1560.
693 Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV. 2012.
694 Untangling Genomes from Metagenomes: Revealing an Uncultured Class of Marine
695 Euryarchaeota. *Science* 335 :587–590.
696 Jayakumar D a., Naqvi SW a, Narvekar P V., George MD. 2001. Methane in coastal and
697 offshore waters of the Arabian Sea. *Marine Chemistry* 74:1–13.

698 Jayakumar DA, Francis C a., Naqvi SW a., Ward BB. 2004. Diversity of nitrite reductase genes
699 (nirS) in the denitrifying water column of the coastal Arabian Sea. *Aquatic Microbial*
700 *Ecology* 34:69–78.
701 Jayakumar A, Al-Rshaidat MMD, Ward BB, Mulholland MR. 2012. Diversity, distribution, and
702 expression of diazotroph *nifH* genes in oxygen-deficient waters of the Arabian Sea. *FEMS*
703 *Microbiology Ecology* 82:597–606.
704 Jensen MM, Lam P, Revsbech NP, Nagel B, Gaye B, Jetten MS, Kuypers MM. 2011. Intensive
705 nitrogen loss over the Omani Shelf due to anammox coupled with dissimilatory nitrite
706 reduction to ammonium. *The ISME Journal* 5:1660–1670.
707 Kalvelage T, Jensen MM, Contreras S, Revsbech NP, Lam P, Günter M, LaRoche J, Lavik G,
708 Kuypers MMM. 2011. Oxygen sensitivity of anammox and coupled N-cycle processes in
709 oxygen minimum zones. *PLoS ONE* 6:e29299.
710 Kalvelage T, Lavik G, Jensen MM, Revsbech NP, Löscher C, Schunck H, Desai DK, Hauss H,
711 Kiko R, Holtappels M, LaRoche J, Schmitz R a, Graco MI, Kuypers MMM. 2015. Aerobic
712 Microbial Respiration In Oceanic Oxygen Minimum Zones. *PloS one* 10:e0133526.
713 Karl DM, Beversdorf L, Bjorkman KM, Church MJ, Martinez A, Delong EF. 2008. Aerobic
714 production of methane in the sea. *Nature Geosci* 1:473–478.
715 Knittel K, Boetius A. 2009. Anaerobic Oxidation of Methane: Progress with an Unknown
716 Process. *Annual Review of Microbiology* 63:311-334.
717 Kozlowski JA, Stieglmeier M, Schleper C, Klotz MG, Stein LY. 2016. Pathways and key
718 intermediates required for obligate aerobic ammonia-dependent chemolithotrophy in
719 bacteria and Thaumarchaeota. *The ISME Journal* doi:10.1038/ismej.2016.2
720 Kuypers MMM, Sliekers AO, Lavik G, Schmid M, Jorgensen BB, Kuenen JG, Sinninghe
721 Damste JS, Strous M, Jetten MSM. 2003. Anaerobic ammonium oxidation by anammox
722 bacteria in the Black Sea. *Nature* 422:608–611.
723 Kuypers MMM, Kuypers MMM, Lavik G, Lavik G, Woebken D, Woebken D, Schmid M,
724 Schmid M, Fuchs BM, Fuchs BM, Amann R, Amann R, Jørgensen BB, Jørgensen BB,
725 Jetten MSM, Jetten MSM. 2005. Massive nitrogen loss from the Benguela upwelling
726 system through anaerobic ammonium oxidation. *Proceedings of the National Academy of*
727 *Sciences of the United States of America* 102:6478–83.
728 Lam P, Lavik G, Jensen MM, van de Vossenberg J, Schmid M, Woebken D, Gutiérrez D,
729 Amann R, Jetten MSM, Kuypers MMM. 2009. Revising the nitrogen cycle in the Peruvian
730 oxygen minimum zone. *Proceedings of the National Academy of Sciences of the United*
731 *States of America* 106:4752–4757.
732 Lam P, Kuypers MMM. 2011. Microbial Nitrogen Cycling Processes in Oxygen Minimum
733 Zones. *Annual Review of Marine Science* 3:317–345.
734 Lavik G, Stuhrmann T, Bruchert V, Van der Plas A, Mohrholz V, Lam P, Muszmann M, Fuchs
735 BM, Amann R, Lass U, Kuypers MMM. 2009. Detoxification of sulphidic African shelf
736 waters by blooming chemolithotrophs. *Nature* 457:581–584.
737 Li M, Jain S, Baker BJ, Taylor C, Dick GJ. 2014. Novel hydrocarbon monooxygenase genes in
738 the metatranscriptome of a natural deep-sea hydrocarbon plume. *Environmental*
739 *Microbiology* 16:60–71.
740 Lipschultz F, Wofsy SC, Ward BB, Codispoti L a, Friedrich G, Elkins JW. 1990. Bacterial
741 transformations of inorganic nitrogen in the oxygen-deficient waters of the Eastern Tropical

South Pacific Ocean. *Deep Sea Research Part A. Oceanographic Research Papers* 37:1513–1541.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T, Steppi S, Jobb G, Förster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, König A, Liss T, Lüßmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A, Schleifer K-H. 2004. ARB: a software environment for sequence data. *Nucleic Acids Research* 4:1363-1371.

Lücker S, Wagner M, Maixner F, Pelletier E, Koch H, Vacherie B, Rattei T, Damste JSS, Spieck E, Le Paslier D, Daims H. 2010. A Nitrospira metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proceedings of the National Academy of Sciences* 107:13479–13484.

Martinez A, Tyson GW, DeLong EF. 2010. Widespread known and novel phosphonate utilization pathways in marine bacteria revealed by functional screening and metagenomic analyses. *Environmental Microbiology* 12:222–238.

Metcalf WW, Griffin BM, Cicchillo RM, Gao J, Janga SC, Cooke HA, Circello BT, Evans BS, Martens-Habbena W, Stahl DA, van der Donk WA. 2012. Synthesis of Methylphosphonic Acid by Marine Microbes: A Source for Methane in the Aerobic Ocean. *Science* 337:1104–1107.

Metcalf WW, Wanner BL. 1993. Mutational analysis of an Escherichia coli fourteen-gene operon for phosphonate degradation, using TnphoA' elements. *Journal of Bacteriology* 175:3430–3442.

Molina V, Ulloa O, Farías L, Urrutia H, Ramírez S, Junier P, Witzel K-P. 2007. Ammonia-oxidizing beta-proteobacteria from the oxygen minimum zone off northern Chile. *Applied and environmental microbiology* 73:3547–55.

Morris RM, Vergin KL, Cho J-C, Rappé MS, Carlson C a., Giovannoni SJ. 2005. Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic Time-series Study site. *Limnology and Oceanography* 50:1687–1696.

Newell SE, Babbin AR, Jayakumar A, Ward BB. 2011. Ammonia oxidation rates and nitrification in the Arabian Sea. *Global Biogeochemical Cycles* 25:1–10.

Nobu MK, Narihiro T, Rinke C, Kamagata Y, Tringe SG, Woyke T, Liu W-T. 2015. Microbial dark matter ecogenomics reveals complex synergistic networks in a methanogenic bioreactor. *The ISME Journal* 9:1710–1722.

Orsi WD, Smith JM, Wilcox HM, Swalwell JE, Carini P, Worden AZ, Santoro AE. 2015. Ecophysiology of uncultivated marine euryarchaea is linked to particulate organic matter. *The ISME Journal* 9:1747-1763.

Owens NJP, Law CS, Mantoura RFC, Burkill PH, Llewellyn CA. 1991. Methane flux to the atmosphere from the Arabian Sea. *Nature* 354:293–296.

Pester M, Rattei T, Flechl S, Gröngröft A, Richter A, Overmann J, Reinhold-Hurek B, Loy A, Wagner M. 2012. amoA-based consensus phylogeny of ammonia-oxidizing archaea and deep sequencing of amoA genes from soils of four different geographic regions . *Environmental Microbiology* 14:525–539 .

783 Padilla CC, Bristow LA, Sarode N, Garcia-Robledo E, Gómez Ramírez E, Benson CR,
784 Bourbonnais A, Altabet MA, Girguis PR, Thamdrup B, Stewart FJ. 2016. NC10 bacteria in
785 marine oxygen minimum zones. *The ISME Journal*. doi:10.1038/ismej.2015.262.
786 Pitcher A, Villanueva L, Hopmans EC, Schouten S, Reichart G-J, Sinninghe Damsté JS. 2011.
787 Niche segregation of ammonia-oxidizing archaea and anammox bacteria in the Arabian Sea
788 oxygen minimum zone. *The ISME Journal* 5:1896–1904.
789 Pruesse E, Peplies J, Glockner FO. 2012. SINA: Accurate high-throughput multiple sequence
790 alignment of ribosomal RNA genes. *Bioinformatics* 28:1823–1829.
791 Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. 2013. The
792 SILVA ribosomal RNA gene database project: improved data processing and web-based
793 tools. *Nucleic Acids Research* 41:D590–D596.
794 Rappe MS, Connon SA, Vergin KL, Giovannoni SJ. 2002. Cultivation of the ubiquitous SAR11
795 marine bacterioplankton clade. *Nature* 418:630–633.
796 Rasko D a, Myers GS a, Ravel J. 2005. Visualization of comparative genomic analyses by
797 BLAST score ratio. *BMC bioinformatics* 6:2.
798 Redfield AC, Ketchum BH, Richards FA. 1963. The influence of organisms on the composition
799 of sea-water. p. 26– 77. In M. N. Hill [ed.], The sea, v. 2. Interscience.
800 Revsbech NP, Larsen LH, Gundersen J, Dalsgaard T, Ulloa O, Thamdrup B. 2009.
801 Determination of ultra-low oxygen concentrations in oxygen minimum zones by the STOX
802 sensor. *Limnology and Oceanography: Methods* 7:371–381.
803 Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling AE, Malfatti
804 S, Swan BK, Gies E a, Dodsworth J a, Hedlund BP, Tsiamis G, Sievert SM, Liu W-T, Eisen
805 J a, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T. 2013.
806 Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499:431–
807 437.
808 Russ L, Speth DR, Jetten MSM, Op den Camp HJM, Kartal B. 2014. Interactions between
809 anaerobic ammonium and sulfur-oxidizing bacteria in a laboratory scale model system.
810 *Environmental Microbiology* 16:3487–3498.
811 Sansone F, Popp B. 2001. Highly elevated methane in the eastern tropical North Pacific and
812 associated isotopically enriched fluxes to the atmosphere. *Geophysical ...* 28:4567–4570.
813 Santoro AE, Dupont CL, Richter RA, Craig MT, Carini P, McIlvin MR, Yang Y, Orsi WD,
814 Moran DM, Saito MA. 2015. Genomic and proteomic characterization of " *Candidatus*
815 Nitrosopelagicus brevis": An ammonia-oxidizing archaeon from the open ocean.
816 *Proceedings of the National Academy of Sciences* 112:1173–1178.
817 Santoro a E, Casciotti KL, Francis C a. 2010. Activity, abundance and diversity of nitrifying
818 archaea and bacteria in the central California Current. *Environmental Microbiology*
819 12:1989–2006.
820 Schouten S, Pitcher A, Hopmans EC, Villanueva L, van Bleijswijk J, Sinninghe Damsté JS.
821 2012. Intact polar and core glycerol dibiphytanyl glycerol tetraether lipids in the Arabian
822 Sea oxygen minimum zone: I. Selective preservation and degradation in the water column
823 and consequences for the TEX86. *Geochimica et Cosmochimica Acta* 98:228–243.
824 Sheik CS, Jain S, Dick GJ. 2014. Metabolic flexibility of enigmatic SAR324 revealed through
825 metagenomics and metatranscriptomics. *Environmental Microbiology* 16:304–317.
826 Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert
827 M, Soding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality

protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7:539.

Speth DR, Russ L, Kartal B, Camp JM Den, Dutilh BE, Jetten SM. 2015. Draft Genome Sequence of Anammox Bacterium "Candidatus Scalindua brodae", Obtained Using Differential Coverage Binning of Sequencing Data from Two Reactor Enrichments. *Genome announcements* 4:1–2.

Stahl DA, de la Torre JR. 2012. Physiology and Diversity of Ammonia-Oxidizing Archaea. *Annual Review of Microbiology* 66:83–101.

Stewart FJ, Ulloa O, DeLong EF. 2012. Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environmental microbiology* 14:23–40.

Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D, Reinthaler T, Poulton NJ, Masland EDP, Gomez ML, Sieracki ME, DeLong EF, Herndl GJ, Stepanauskas R. 2011. Potential for Chemolithoautotrophy Among Ubiquitous Bacteria Lineages in the Dark Ocean. *Science* 333:1296–1300.

Tavormina PL, Ussler W, Steele J a., Connon S a., Klotz MG, Orphan VJ. 2013. Abundance and distribution of diverse membrane-bound monooxygenase (Cu-MMO) genes within the Costa Rica oxygen minimum zone. *Environmental Microbiology Reports* 5:414–423.

Thamdrup B, Dalsgaard T, Jensen MM, Ulloa O, Farías L, Escribano R. 2006. Anaerobic ammonium oxidation in the oxygen-deficient waters off northern Chile. *Limnology and Oceanography* 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 Research Part I: Oceanographic Research Papers* 65:36–45.

Upstill-Goddard RC, Barnes J, Owens NJP. 1999. Nitrous oxide and methane during the 1994 SW monsoon in the Arabian Sea/northwestern Indian Ocean. *Journal of Geophysical Research* 104:30067.

Vergin KL, Beszteri B, Monier A, Thrash JC, Temperton B, Treusch AH, Kilpert F, Worden AZ, Giovannoni SJ. 2013. High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time-series Study site by phylogenetic placement of pyrosequences. *The ISME journal* 7:1322–32.

Villanueva L, Speth DR, van Alen T, Hoischen A, Jetten MSM. 2014. Shotgun metagenomic data reveals significant abundance but low diversity of "Candidatus Scalindua" marine anammox bacteria in the Arabian Sea oxygen minimum zone. *Frontiers in Microbiology* 5:1–9.

Villanueva L, Schouten S, Sinninghe Damsté JS. 2014. Depth-related distribution of a key gene of the tetraether lipid biosynthetic pathway in marine Thaumarchaeota. *Environmental microbiology* 6:1–32.

Van de Vossenberg J, Woebken D, Maalcke WJ, Wessels HJCT, Dutilh BE, Kartal B, Janssen-Megens EM, Roeselers G, Yan J, Speth D, Gloerich J, Geerts W, van der Biezen E, Pluk W, Francoijs K-J, Russ L, Lam P, Malfatti SA, Tringe SG, Haaijer SCM, Op den Camp HJM, Stunnenberg HG, Amann R, Kuypers MMM, Jetten MSM. 2013. The metagenome of the marine anammox bacterium " *Candidatus* Scalindua profunda" illustrates the versatility of this globally important nitrogen cycle bacterium. *Environmental Microbiology* 15:1275–1289.

872 Walsh DA, Zaikova E, Howes CG, Song YC, Wright JJ, Tringe SG, Tortell PD, Hallam SJ.
873 2009. Metagenome of a Versatile Chemolithoautotroph from Expanding Oceanic Dead
874 Zones. *Science* 326 :578–582.
875 Ward BB, Devol AH, Rich JJ, Chang BX, Bulow SE, Naik H, Pratihary A, Jayakumar A. 2009.
876 Denitrification as the dominant nitrogen loss process in the Arabian Sea. *Nature* 461:78–81.
877 White a. E, Karl DM, Björkman K, Beversdorf LJ, Letelier RM. 2010. Production of organic
878 matter by Trichodesmium IMS101 as a function of phosphorus source. *Limnology and*
879 *Oceanography* 55:1755–1767.
880 Woebken D, Lam P, Kuypers MMM, Naqvi SW a, Kartal B, Strous M, Jetten MSM, Fuchs BM,
881 Amann RI. 2008. A microdiversity study of anammox bacteria reveals a novel Candidatus
882 Scalindua phylotype in marine oxygen minimum zones. *Environmental Microbiology*
883 10:3106–3119.
884 Wright JJ, Konwar KM, Hallam SJ. 2012. Microbial ecology of expanding oxygen minimum
885 zones. *Nature Reviews Microbiology* 10:381–394.
886 Zeigler Allen L, Allen EE, Badger JH, McCrow JP, Paulsen IT, Elbourne LDH, Thiagarajan M,
887 Rusch DB, Nealson KH, Williamson S, Venter JC, Allen AE. 2012. Influence of nutrients
888 and currents on the genomic composition of microbes across an upwelling mosaic . *The*
889 *ISME Journal* 6:1403-1414.

Figure legends

Figure 1. Overview of microbial 16S rRNA gene diversity detected in the upper limit (station PA2) and the anoxic core (station PA5) of the Arabian Sea oxygen minimum zone. Sequence reads are shown as percentage of total 16S rRNA read counts. Only phylogenetic groups accounting for more than 1% of the total community in at least one of the two datasets are listed.

Figure 2. Phylogenetic inference of thaumarchaeal 16S rRNA and ammonia monooxygenase (*amoA)* gene contigs assembled from the OMZ metagenomes. The trees were calculated using the Neighbor Joining algorithm and based on 1225 nucleotide positions for the 16S rRNA gene (**left**) and 144 deduced amino acid positions for the *amoA* gene (**right**).

Figure 3. Phylogenetic inference of *Scalindua*-related 16S rRNA and hydrazine dehydrogenase *hdh* gene contigs assembled from the OMZ metagenomes. The tree were calculated using the Neighbor Joining algorithm and based on 1388 nucleotide positions for the 16S rRNA gene (**left**) and 180 deduced amino acid positions for the *hdh* gene **(right**).

Figure 4. Nitrogen cycling potential in the Arabian Sea oxygen minimum zone. Read abundances were normalized according to gene length and total read abundance in the metagenome dataset. Normalized abundances are shown as proportion (**blue**) of total normalized *rpoB* (RNA polymerase) gene abundance (**grey**). The description of all marker genes and methane and nitrogen cycling processes is given in Table 1.

Figure 5. Phylogenetic inference of nitrite oxidoreductase (*nxrA)* gene sequences from the OMZ metagenomes. The tree was calculated using the Neighbor Joining algorithm and based on 3209 nucleotide positions. Bootstrap values represent 1000 replicates.

Figure 6. Methane cycling potential in the Arabian Sea oxygen minimum zone. Read abundances were normalized according to gene length and total read abundance in the metagenome dataset. Normalized abundances are shown as proportion (**blue**) of total normalized *rpoB* (RNA polymerase) gene abundance (**grey**). Original read abundances are given in addition below or above the pie charts. HEP: 2-hydroxyethylphosphonate. Mpn: Methylphosphonate. All nitrogen and methane cycling marker genes are described in Table 1.

Figure 1(on next page)

Overview of microbial 16S rRNA gene diversity in the Arabian Sea oxygen minimum zone

Overview of microbial 16S rRNA gene diversity detected in the suboxic zone (station PA2) and the anoxic core (station PA5) of the Arabian Sea oxygen minimum zone. Sequence reads are shown as percentage of total 16S rRNA read counts. Only phylogenetic groups accounting for more than 1% of the total community in at least one of the two datasets are listed.

Taxonomy				PA2	PA5
Bacteria	Proteobacteria	Alpha	SAR11	14.3	8.1
			Rhodobacteriacaea	1.5	0.5
			Rhodospirillales	1.4	4.2
		Gamma	Alteromonadales	1.9	1.1
			SAR86	7.9	1.5
			SUP05 and alike	2.9	2.7
			Salinisphaeraceae	1.6	1.9
			E01-9C-26	0.5	1.6
		Delta	SAR324	5.0	3.1
	Bacteriodetes		Flavobacteriales	2.0	0.5
	Planctomycetes		Brocadiaceae	< 0.1	5.2
			Phycisphaerae	< 0.1	1.1
	Verrucomicrobia		Arctic97B-4	1.1	0.2
	Deferribacteres		SAR406	8.5	19.9
	Nitrospinae		Nitrospina	1.4	2.7
	Actinobacteria		Acidimicrobia	3.2	2.4
	Chloroflexi		SAR202	0.8	2.4
			Dehalococcoidia	0	1.6
Archaea	Thaumarchaeota		Marine Group I	14.0	3.3
	Euryarchaeota		Marine Group II	7.5	0.4
			Marine Group III	1.4	0.5
			Haloarchaea	0	3.8
Sum of remaining rare taxa				23.0	27.6

Figure 2(on next page)

Phylogenetic inference of thaumarchaeal contigs assembled from the OMZ metagenomes.

The trees were calculated using the Neighbor Joining algorithm and based on 1225 nucleotide positions for the 16S rRNA gene (**left**) and 144 deduced amino acid positions for the ammonia monooxygenase encoding gene (*amoA*) (right).

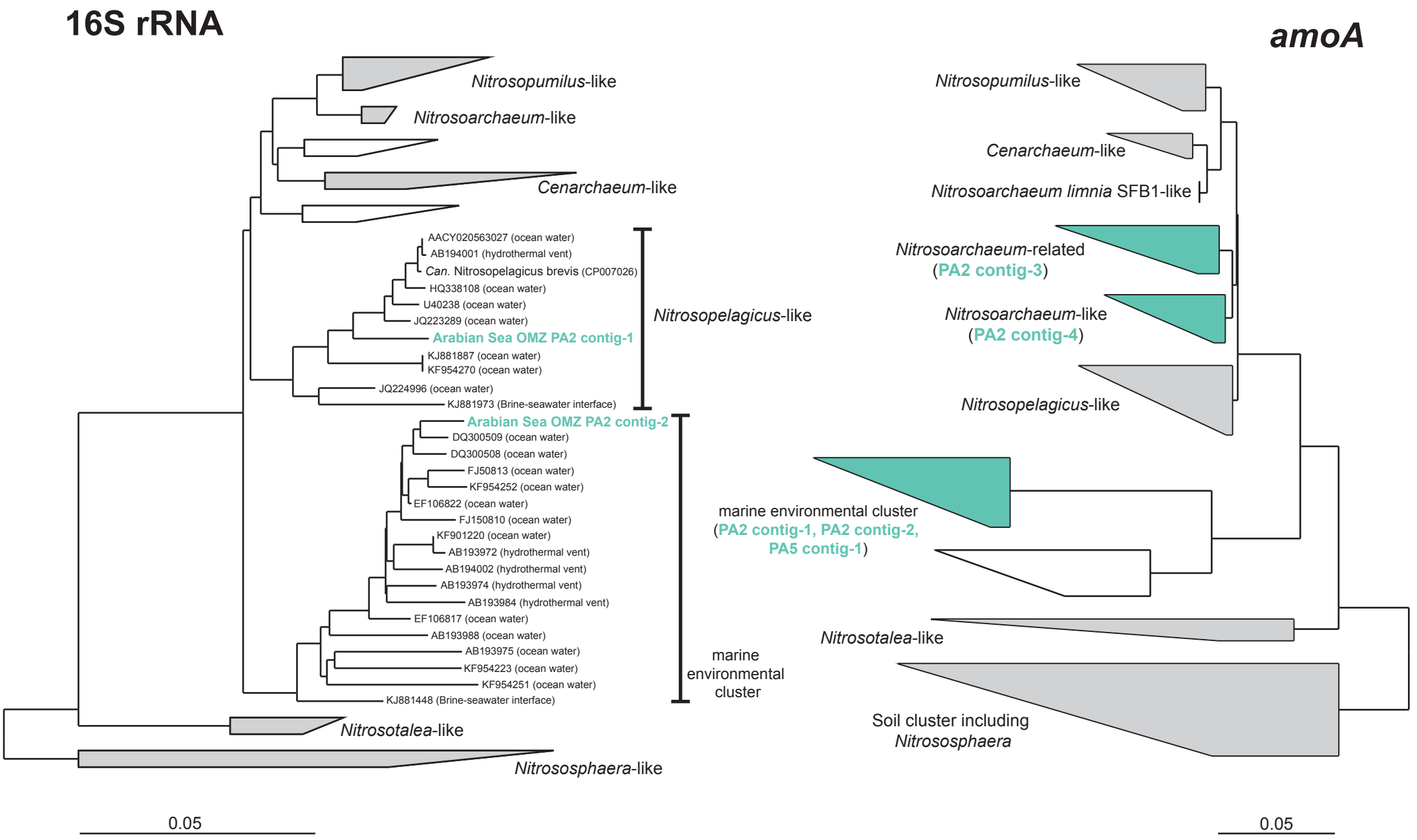
16S rRNA
Nitrosopumilus-like
Nitrosoarchaeum-like
Cenarchaeum-like
AACY020563027 (ocean water)
AB194001 (hydrothermal vent)
Can. Nitrosopelagicus brevis (CP007026)
HQ338108 (ocean water)
U40238 (ocean water)
JQ223289 (ocean water)
Arabian Sea OMZ PA2 contig-1
KJ881887 (ocean water)
KF954270 (ocean water)
JQ224996 (ocean water)
KJ881973 (Brine-seawater interface)
Nitrosopelagicus-like
Arabian Sea OMZ PA2 contig-2
DQ300509 (ocean water)
DQ300508 (ocean water)
FJ50813 (ocean water)
KF954252 (ocean water)
EF106822 (ocean water)
FJ150810 (ocean water)
KF901220 (ocean water)
AB193972 (hydrothermal vent)
AB194002 (hydrothermal vent)
AB193974 (hydrothermal vent)
AB193984 (hydrothermal vent)
EF106817 (ocean water)
AB193988 (ocean water)
AB193975 (ocean water)
KF954223 (ocean water)
KF954251 (ocean water)
KJ881448 (Brine-seawater interface)
marine environmental cluster
Nitrosotalea-like
Nitrososphaera-like
0.05
amoA
Nitrosopumilus-like
Cenarchaeum-like
Nitrosoarchaeum limnia SFB1-like
Nitrosoarchaeum-related (PA2 contig-3)
Nitrosoarchaeum-like (PA2 contig-4)
Nitrosopelagicus-like
marine environmental cluster (PA2 contig-1, PA2 contig-2, PA5 contig-1)
Nitrosotalea-like
Soil cluster including Nitrososphaera
0.05

Figure 3(on next page)

Phylogenetic inference of Scalindua-related contigs assembled from the OMZ metagenomes.

The tree were calculated using the Neighbor Joining algorithm and based on 1388 nucleotide positions for the 16S rRNA gene (**left**) and 180 deduced amino acid positions for the hydrazine dehydrogenase encoding gene (*hdh*) (**right**).

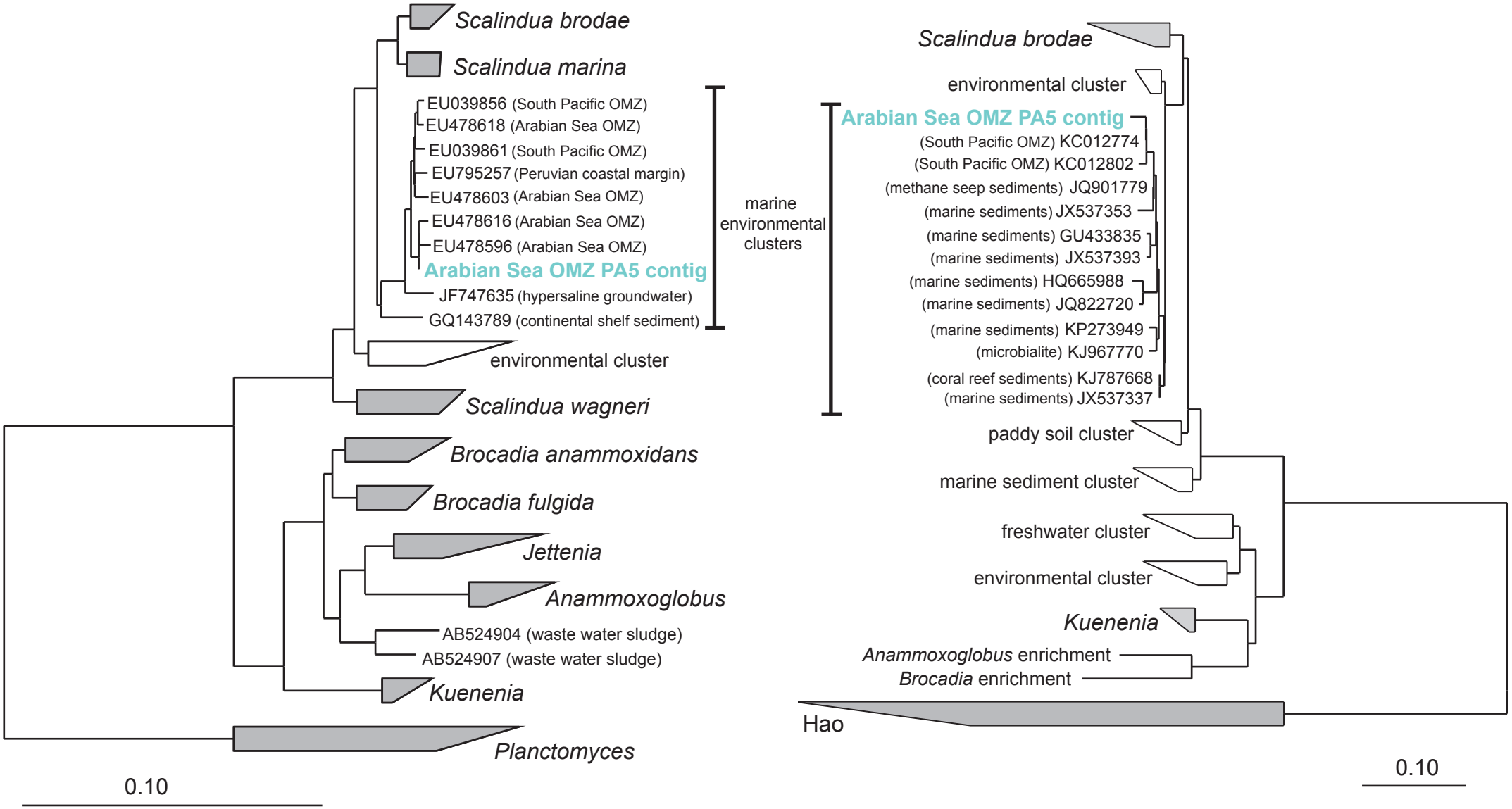
16S rRNA
Scalindua brodae
Scalindua marina
EU039856 (South Pacific OMZ)
EU478618 (Arabian Sea OMZ)
EU039861 (South Pacific OMZ)
EU795257 (Peruvian coastal margin)
EU478603 (Arabian Sea OMZ)
EU478616 (Arabian Sea OMZ)
EU478596 (Arabian Sea OMZ)
Arabian Sea OMZ PA5 contig
JF747635 (hypersaline groundwater)
GQ143789 (continental shelf sediment)
marine environmental clusters
environmental cluster
Scalindua wagneri
Brocadia anammoxidans
Brocadia fulgida
Jettenia
Anammoxoglobus
AB524904 (waste water sludge)
AB524907 (waste water sludge)
Kuenenia
Planctomyces
0.10
hdh
Scalindua brodae
environmental cluster
Arabian Sea OMZ PA5 contig
(South Pacific OMZ) KC012774
(South Pacific OMZ) KC012802
(methane seep sediments) JQ901779
(marine sediments) JX537353
(marine sediments) GU433835
(marine sediments) JX537393
(marine sediments) HQ665988
(marine sediments) JQ822720
(marine sediments) KP273949
(microbialite) KJ967770
(coral reef sediments) KJ787668
(marine sediments) JX537337
paddy soil cluster
marine sediment cluster
freshwater cluster
environmental cluster
Kuenenia
Anammoxoglobus enrichment
Brocadia enrichment
Hao
0.10

Figure 4(on next page)

Nitrogen cycling potential in the Arabian Sea oxygen minimum zone.

Read abundances were normalized according to gene length and total read abundance in the metagenome dataset. Normalized abundances are shown as proportion (blue) of total normalized *rpoB* (RNA polymerase) gene abundance (**grey**). The description of all marker genes and methane and nitrogen cycling processes is given in Table 1.

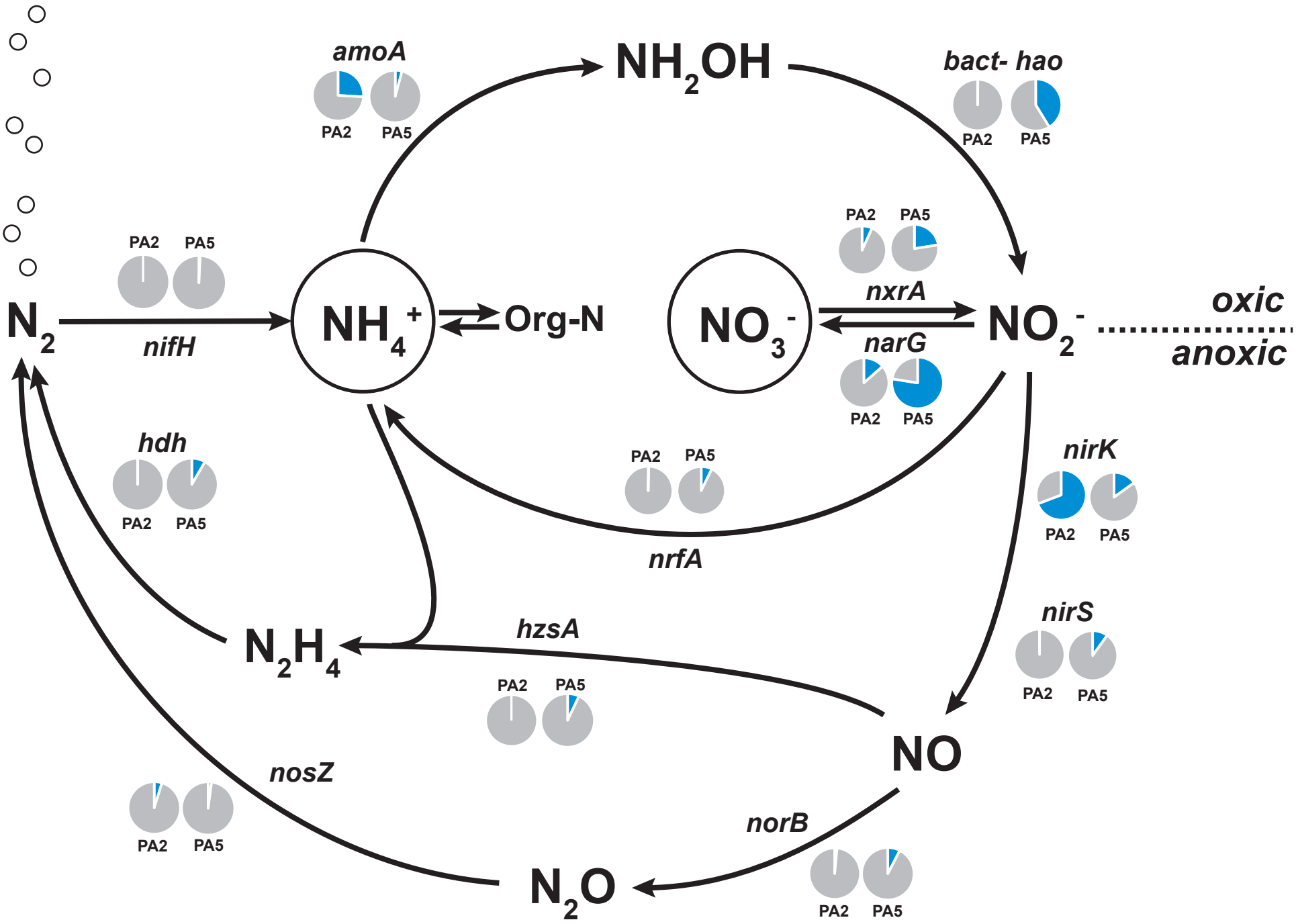
amoA
PA2
PA5
NH2OH
bact- hao
PA2
PA5
PA2
PA5
PA2
PA5
nxrA
oxic
N2
nifH
NH4+
Org-N
NO3-
NO2-
narG
anoxic
PA2
PA5
PA2
PA5
hdh
PA2
PA5
PA2
PA5
nirK
nrfA
PA2
PA5
nirS
hzsA
N2H4
PA2
PA5
NO
nosZ
PA2
PA5
norB
N2O
PA2
PA5

Figure 5(on next page)

Phylogenetic inference of *nxrA* sequences from the OMZ metagenomes.

The tree was calculated using the Neighbor Joining algorithm and based on 3209 nucleotide positions. Bootstrap values represent 1000 replicates.

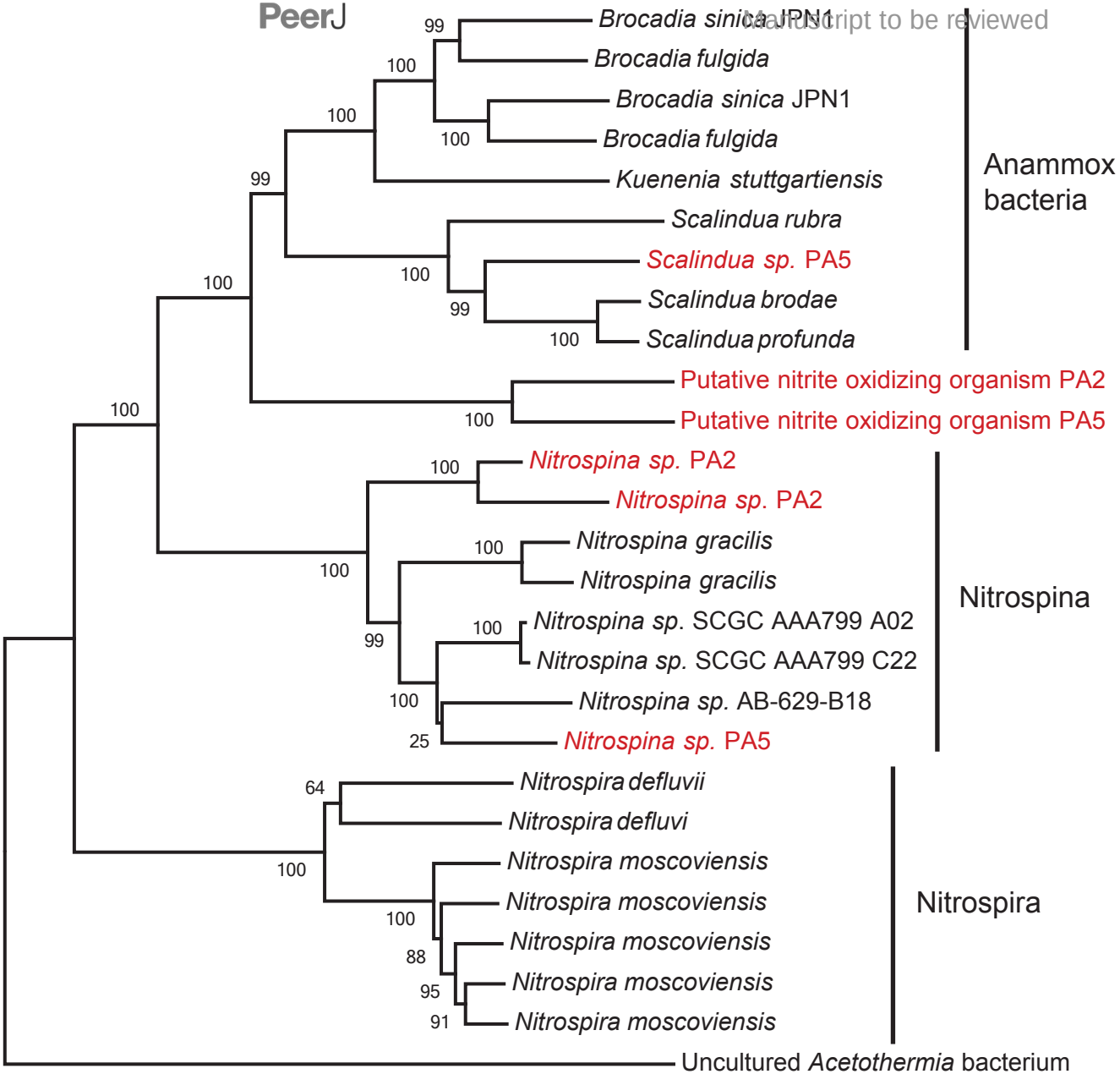
Brocadia sinica JPN1
Brocadia fulgida
Brocadia sinica JPN1
Brocadia fulgida
Kuenenia stuttgartiensis
Scalindua rubra
Scalindua sp. PA5
Scalindua brodae
Scalindua profunda
Anammox bacteria
Putative nitrite oxidizing organism PA2
Putative nitrite oxidizing organism PA5
Nitrospina sp. PA2
Nitrospina sp. PA2
Nitrospina gracilis
Nitrospina gracilis
Nitrospina sp. SCGC AAA799 A02
Nitrospina sp. SCGC AAA799 C22
Nitrospina sp. AB-629-B18
Nitrospina sp. PA5
Nitrospina
Nitrospira defluvii
Nitrospira defluvi
Nitrospira moscoviensis
Nitrospira moscoviensis
Nitrospira moscoviensis
Nitrospira moscoviensis
Nitrospira moscoviensis
Nitrospira
Uncultured Acetothermia bacterium
99
100
100
100
99
100
100
99
100
100
100
100
100
100
99
100
100
25
64
100
100
88
95
91
0.1

Figure 6(on next page)

Methane cycling potential in the Arabian Sea oxygen minimum zone.

Read abundances were normalized according to gene length and total read abundance in the metagenome dataset. Normalized abundances are shown as proportion (blue) of total normalized *rpoB* (RNA polymerase) gene abundance (grey). Original read abundances are given in addition below or above the pie charts. HEP: 2-hydroxyethylphosphonate. Mpn: Methylphosphonate. The description of all nitrogen and methane cycling marker genes is given in Table 1.

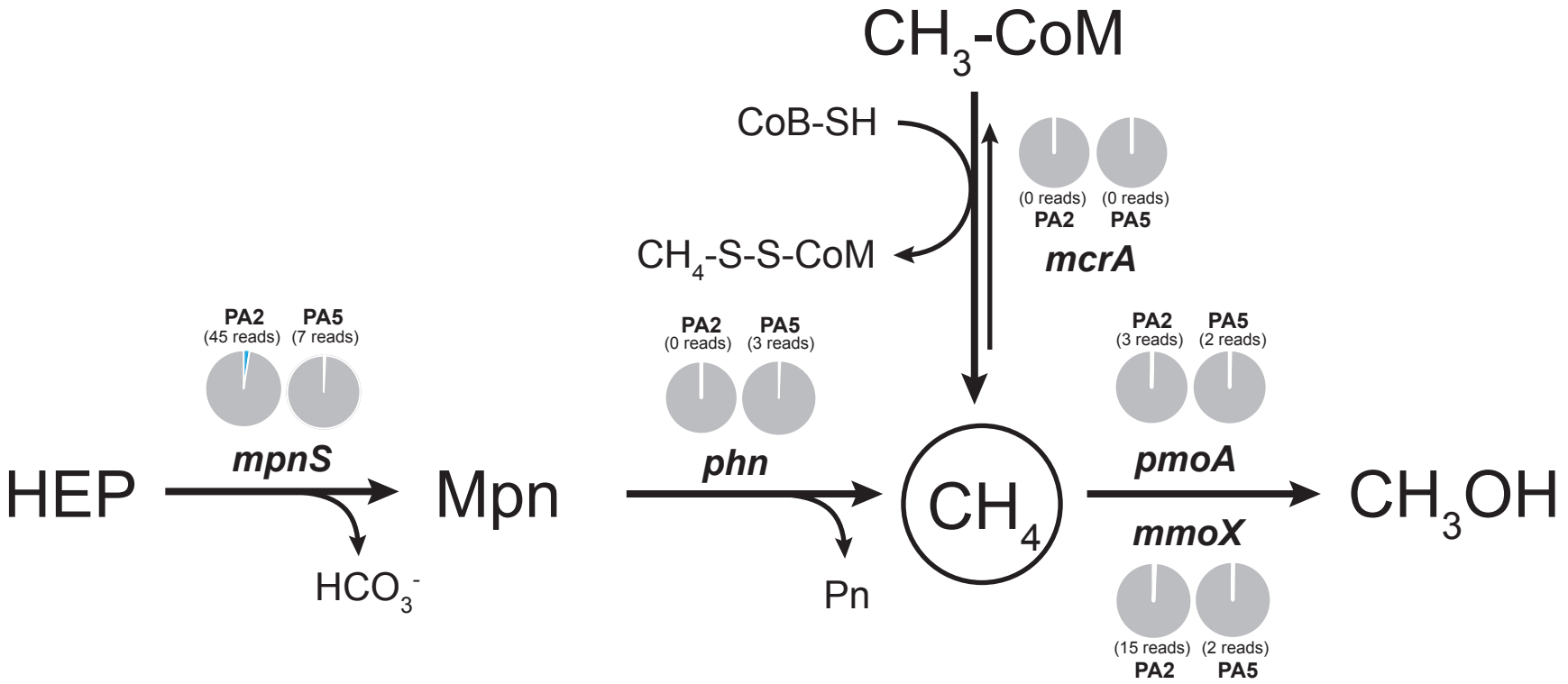
CH3-CoM
CoB-SH
CH4-S-S-CoM
(0 reads) (0 reads)
PA2 PA5
mcrA
PA2 PA5
(45 reads) (7 reads)
mpnS
HEP
Mpn
HCO3-
PA2 PA5
(0 reads) (3 reads)
phn
Pn
CH4
PA2 PA5
(3 reads) (2 reads)
pmoA
mmoX
CH3OH
(15 reads) (2 reads)
PA2 PA5

Table 1(on next page)

Overview of nitrogen and methane cycle marker genes and BLAST score ratio cut-off value used for removal of false positive BLAST hits.

See Figure S2 for a graphical overview of the bit-score ratio analysis pipeline. DNRA: dissimilatory nitrite reduction to ammonia. Anammox: anaerobic ammonia oxidation.

Table 1. Overview of nitrogen and methane cycle marker genes and BLAST score ratio cut-off value used for removal of false positive BLAST hits. See Figure S2 for a graphical overview of the bit-score ratio analysis pipeline. DNRA: dissimilatory nitrite reduction to ammonia. Anammox: anaerobic ammonia oxidation.

Process	Enzyme name	Gene abbreviation	BLAST score ratio cutoff (PA2 / PA5)
Nitrogen fixation	nitrogenase	*nifH*	# / #
Nitrification	ammonium monooxygenase	*amoA*	# / #
	hydroxylamine oxidoreductase	*hao*	# / 0.75
	nitrate:nitrite oxidoreductase	*nxrA*	0.85* / 0.85*
Denitrification/DNRA/ anammox	nitrate reductase	*narG*	0.5 / 0.5
Denitrification	copper nitrite reductase	*nirK*	0.55 / 0.55
	heme *cd1* nitrite reductase	*nirS*	# / 0.6
	nitric oxide reductase	*norB / norZ*	0.8 / 0.8
	nitrous oxide reductase	*nosZ*	0.8 / 0.75
DNRA	cytochrome c nitrite reductase	*nrfA*	# / #
anammox	hydrazine synthase	*hzsA*	# / 0.75
	hydrazine dehydrogenase	*hdh*	** / **
Methanogenesis/anaerobic methane oxidation	methyl-coenzyme M reductase	*mcrA*	# / #
Aerobic methane oxidation	soluble methane monooxygenase	*mmoX*	# / #
	particulate methane monooxygenase	*pmoA*	# / #
Methylphosphonate production	methylphosphonate synthase	*mpnS*	# / #

Methylphosphonate cleavage	C-P lyase	*phnGHI*	# / #

 # manually checked

** subset of narG hits, ** subset of hao hits after mapping against Scalindua hdh*