

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

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

**DOI**

[10.7717/peerj.1924](https://doi.org/10.7717/peerj.1924)

**Publication date**

2016

**Document Version**

Final published version

**Published in**

PeerJ

**Citation (APA)**

Lüke, C., Speth, D. R., Kox, M. A. R., Villanueva, L., & Jetten, M. S. M. (2016). Metagenomic analysis of nitrogen and methane cycling in the Arabian Sea oxygen minimum zone. *PeerJ*, 4, Article e1924. <https://doi.org/10.7717/peerj.1924>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.



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

Claudia Lüke<sup>1,\*</sup>, Daan R. Speth<sup>1,\*</sup>, Martine A.R. Kox<sup>1</sup>, Laura Villanueva<sup>2</sup> and Mike S.M. Jetten<sup>1,3,4</sup>

<sup>1</sup> Department of Microbiology, IWW, Radboud University Nijmegen, Nijmegen, Netherlands

<sup>2</sup> Department of Marine Organic Biogeochemistry, Royal Netherlands Institute for Sea Research (NIOZ), 't Horntje (Texel), Netherlands

<sup>3</sup> Department of Biotechnology, Delft University of Technology, Delft, Netherlands

<sup>4</sup> Soehngen Institute of Anaerobic Microbiology, Nijmegen, Netherlands

\*These authors contributed equally to this work.

## 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.

Submitted 28 November 2015

Accepted 21 March 2016

Published 7 April 2016

Corresponding authors

Claudia Lüke, clueke@science.ru.nl

Daan R. Speth, D.speth@science.ru.nl

Academic editor

Ludmila Chistoserdova

Additional Information and  
Declarations can be found on  
page 19

DOI 10.7717/peerj.1924

© Copyright  
2016 Lüke et al.

Distributed under  
Creative Commons CC-BY 4.0

OPEN ACCESS

**Subjects** Biogeography, Bioinformatics, Genomics, Marine Biology, Microbiology

**Keywords** Metagenomics, Oxygen minimum zone, Functional marker genes, Nitrogen cycle, Methane cycle, Anammox, Microbial, *Thaumarchaeota*

## 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  $\leq 20 \mu\text{M}$  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, Ketchum & Richards, 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; Kalvelage et al., 2015; Füssel et al., 2012). Eventually, nitrogen is lost from the system due to denitrification (Groffman et al., 2006; Ward et al., 2009) or anaerobic ammonium oxidation (anammox) (Kuypers et al., 2003; Kalvelage et al., 2015).

The relative contribution of anammox and denitrification to nitrogen loss from OMZs has been the subject of debate (Lam & Kuypers, 2011). Before the discovery of anammox, denitrification was thought to be the only contributor (Codispoti & Richards, 1976; Lipschultz et al., 1990; Devol et al., 2006), but the detected ammonium concentrations were lower than expected based on just denitrification. After the discovery of the anammox process in the Black Sea (Kuypers et al., 2003), anammox bacteria were shown to occur in all the major OMZ using marker genes, lipid analysis, FISH and stable isotope pairing (e.g., Kuypers et al., 2005; Thamdrup et al., 2006; Hamersley et al., 2007). In the ETSP OMZ, anammox was the dominant process involved in nitrogen loss (Lam & Kuypers, 2011; Stewart, Ulloa & DeLong, 2012). For the Arabian Sea OMZ, evidence for both denitrification and anammox as the dominant cause of nitrogen loss exists (Ward et al., 2009; Jensen et al., 2011; Pitcher et al., 2011) and the contribution of either process likely varies with season and location.

Previous studies on nitrogen cycling the Arabian Sea OMZ focused on one or (a comparison of) a few processes such as ammonia oxidation (Newell et al., 2011; Schouten et al., 2012), denitrification (Jayakumar et al., 2004; Ward et al., 2009), and anammox (Jensen et al., 2011; Pitcher et al., 2011; Villanueva et al., 2014). A recent study used metagenomes obtained from samples across the ETSP, ETNP and the Bermuda Atlantic Time-series Station to assess the distribution of iron and copper containing nitrogen cycle enzymes in these systems (Glass et al., 2015), but a comprehensive study of the nitrogen cycling potential in the Arabian Sea OMZ is lacking.

Furthermore, only very little is known on methane turnover in oxygen minimum zones. Low concentrations of methane have been reported at sites of the ETNP and the

**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 Fig. 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	<i>nifH</i>	##
Nitrification	Ammonium monooxygenase	<i>amoA</i>	##
	Hydroxylamine oxidoreductase	<i>hao</i>	# / 0.75
	Nitrate:nitrite oxidoreductase	<i>nxrA</i>	0.85*/0.85*
Denitrification/DNRA/anammox	Nitrate reductase	<i>narG</i>	0.5/0.5
Denitrification	Copper nitrite reductase	<i>nirK</i>	0.55/0.55
	Heme <i>cd1</i> nitrite reductase	<i>nirS</i>	#/0.6
	Nitric oxide reductase	<i>norB/norZ</i>	0.8/0.8
	Nitrous oxide reductase	<i>nosZ</i>	0.8/0.75
DNRA	Cytochrome c nitrite reductase	<i>nrjA</i>	##
Anammox	Hydrazine synthase	<i>hzsA</i>	#/0.75
	Hydrazine dehydrogenase	<i>hdh</i>	**/**
Methanogenesis/anaerobic methane oxidation	Methyl-coenzyme M reductase	<i>mcrA</i>	##
Aerobic methane oxidation	Soluble methane monooxygenase	<i>mmoX</i>	##
	Particulate methane monooxygenase	<i>pmoA</i>	##
Methylphosphonate production	Methylphosphonate synthase	<i>mpnS</i>	##
Methylphosphonate cleavage	C-P lyase	<i>phnGHI</i>	##

**Notes.**

# Manually checked.

\* Subset of *narG* hits.

\*\* Subset of *hao* hits after mapping against *Scalindua hdh*.

ETSP suggesting the possibility of methane cycling in these ecosystems (Sansone & Popp, 2001; Padilla et al., 2016). In addition, *pmoA* genes (encoding the particulate methane monooxygenase, see Table 1) have been detected in the water column of the ETSP OMZ indicating the presence of aerobic methanotrophs (Tavormina et al., 2013). In the Arabian Sea, high methane concentrations (up to 227% saturation compared to atmospheric levels) have been measured in the surface waters (Owens et al., 1991; Bange et al., 1998; Upstill-Goddard, Barnes & Owens, 1999) and elevated concentrations (up to approximately 8 nM) were found at 150–200 m depths (Jayakumar et al., 2001). Not much is known about the processes and the organisms involved in production or turnover, but recently *Thaumarchaeota* cleaving methylphosphonates have been proposed as a potential source of methane in the oceans (Metcalf et al., 2012).

Metagenomics is a powerful tool to provide an all-inclusive picture of the functional potential of an ecosystem. As sequencing is becoming easier and less expensive, the bottleneck in metagenomics is shifting from data generation to sequence analysis strategies. Here, we developed a new analysis strategy for mining metagenome data, based on curated databases of marker genes for nitrogen and methane cycle processes (see Table 1 for an overview of the used marker genes). We applied this strategy to metagenome data retrieved from two depths along the oxygen gradient of the Arabian Sea OMZ.

## MATERIALS AND METHODS

### 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 2,000 m below the surface (total water depth at this site is 3,010 m). Dissolved oxygen concentrations decreased from fully saturated concentrations at the surface to 3.2  $\mu\text{M}$  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  $\leq 0.09 \mu\text{M}$  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 1,050 m ([Pitcher et al., 2011](#)). At 170 m depth, the ammonium concentration showed a peak (0.14  $\mu\text{M}$ ), 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 (200–1,700 L) were filtered through 142-mm diameter 0.2  $\mu\text{m}$  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  $\mu\text{l}$  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  $-80^\circ\text{C}$ .

### 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 min using the Ion Xpress<sup>TM</sup> 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<sup>®</sup> 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](https://www.ncbi.nlm.nih.gov/sra/SRA304624).

## Bioinformatics

### ***Quality assessment and assembly feasibility***

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

To assess the feasibility of assembly for the analysis of the read data, the datasets were assembled *de novo* using the CLC genomics workbench with word size 35 and bubble size 5,000. Using the data from both sampling sites combined, this resulted in 5,105 contigs longer than 1,000 bp, incorporating only 7.6% of the reads. Based on the results of the metagenome assembly, we decided to employ a read based analysis strategy.

### ***16S rRNA gene analysis***

To extract reads matching the ribosomal small subunit for taxonomic classification, the SILVA SSU RefNR99 dataset (version 115, [Quast et al., 2013](#)) was used as reference. First, the metagenome reads were mapped against the SILVA dataset using the CLC Genomics Workbench (mismatch penalty 2, In/Del penalty 3, 50% identity over 70% of the sequence). Mapped reads were extracted and used for BLAST searches against the identical SILVA dataset (*E*-value cut-off  $10^{-6}$ ). Positive hits were aligned using the SINA aligner ([Pruesse, Peplies & Glockner, 2012](#)) and imported into the SILVA refNR99 version 115 ARB database ([Ludwig et al., 2004](#); [Quast et al., 2013](#)). Sequences were added to the existing Neighbor Joining 16S rRNA tree (including nearly 500,000 16S rRNA sequences) using maximum parsimony criteria without optimization of the 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, Matthies & Conrad, 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 [Supplemental Information](#)), 134 archaeal *amoA* sequences, 124 methanogen *mcrA* sequences) covering the main described clusters was selected from the phylogenetic 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 matched to the respective Interpro family and keyword/accession number search in NCBI). The amino acid sequences were downloaded and aligned using Clustal Omega (Sievers et al., 2011). Alignments were manually inspected and for sequences with large gaps or insertions, literature was queried to verify the function of the respective enzyme. In cases where the function of a divergent sequence has not been described or is different than the target database, the sequence was discarded. The remaining aligned sequences were imported into ARB (Ludwig et al., 2004). Sequences with ambiguous characters or obvious shifts in reading frame were removed. The remaining sequences were used for phylogenetic tree construction. Phylogenies were compared to literature to verify that the known and described phylogenetic groups are covered and that the dataset is representative. For the *norB/norZ* database, all sequences in the cytochrome oxidase PFAM family (PF00115) were downloaded, replicates were removed using UClust (80% identity) (Edgar, 2010) and the clustered sequences were screened against characterized cNOR (cytochrome c dependent nitric oxide reductase) and qNOR (quinol dependent nitric oxide reductase) sequences using a BLAST score ratio (BSR) (Rasko, Myers & Ravel, 2005). Sequences with a BSR over 0.4 were included in the database. Representatives of phylogenetic clusters (distance level of 0.1) were exported in fasta format and imported into ARB where sequences with ambiguous characters or obvious shifts in reading frame were removed. The remaining sequences were used for phylogenetic tree construction and exported as reference set for subsequent BLAST analysis. All reference sets are provided as [Supplemental Information](#).

*BLAST analysis using the reference datasets.* The reads of the metagenomes described above were used in BLASTx searches against the reference sets ( $E$ -value cutoff  $10^{-6}$ ). The  $E$ -value was established after first test analyses with different  $E$ -values ( $10^{-4}$ – $10^{-6}$ ). Positive hits were extracted using a custom perl script available at [www.github.com/dspeth](http://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 Fig. 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 (Figs. 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 5,000) 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  $\mu\text{M}$ , Table S1) and the core zone (station PA5) in which the oxygen concentration drops below the detection limit (3.4  $\mu\text{M}$ ) (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 (Fig. 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 (Fig. 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

Taxonomy				PA2	PA5
Bacteria	Proteobacteria	Alpha	SAR11	14.3	8.1
			Rhodobacteriaceae	1.5	0.5
			Rhodospirillales	1.4	4.2
	Gamma	Alteromonadales	SAR86	1.9	1.1
			SUP05 and alike	7.9	1.5
			Salinisphaeraceae	2.9	2.7
			E01-9C-26	1.6	1.9
				0.5	1.6
	Delta	SAR324	5.0	3.1	
	Bacteroidetes	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

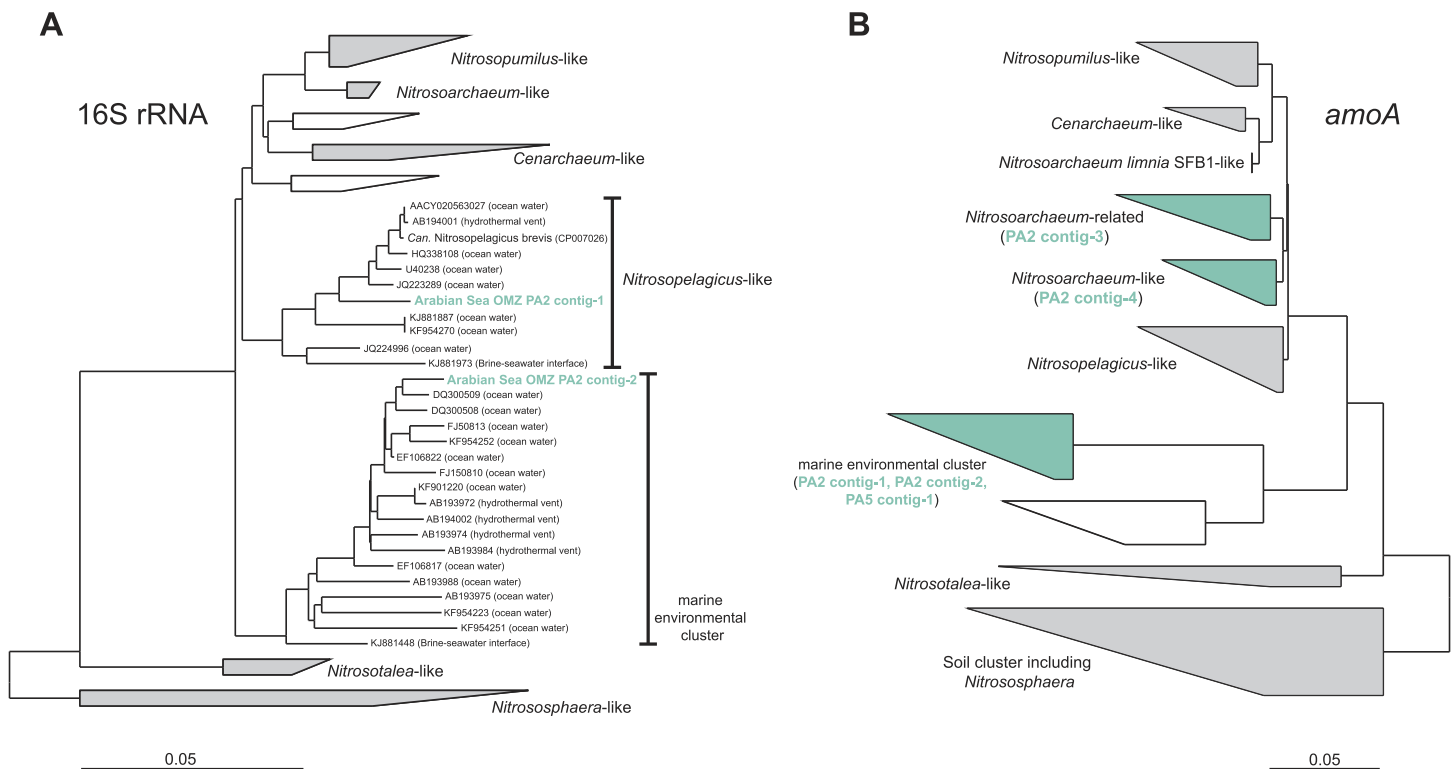
**Figure 1** Overview of microbial 16S rRNA gene diversity in the Arabian Sea oxygen minimum zone. Overview of microbial 16S rRNA gene diversity detected in the suboxic zone (station PA2) and the anoxic core (station PA5) of the Arabian Sea oxygen minimum zone. Sequence reads are shown as percentage 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.

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 degradation in a methanogenic environment ([Nobu et al., 2015](#)). These findings of an anaerobic lifestyle agree well with their high abundance in a low oxygen environment (upper OMZ) and their dominance in the OMZ core. The detected *Rhodospirillales* (*Alphaproteobacteria*) might also contribute to fermentation of organic matter, as this order comprises known acetic acid bacteria. *Rhodospirillales* were barely detectable at the OMZ upper limit, but comprise 4.1% of the community at the OMZ core. Other abundant lineages include the *Deltaproteobacteria* SAR324. These bacteria are frequently found in the ocean and seem to be correlated with low oxygen concentrations ([Wright, Konwar & Hallam, 2012](#); [Sheik, Jain & Dick, 2014](#)). SAR324 representatives have been shown to be able to fix CO<sub>2</sub> ([Swan et al., 2011](#)) and are predicted to be capable of autotrophic denitrification with various electron donors ([Sheik, Jain & Dick, 2014](#)). Likewise, *Gammaproteobacteria* of the SUP05 lineage have been implied in autotrophic denitrification with sulphur as electron donor ([Lavik et al., 2009](#); [Walsh et al., 2009](#); [Russ et al., 2014](#)). They furthermore have been associated with cryptic sulfur cycling in the ETSP OMZ ([Canfield et al., 2010](#)). Members of both SAR324 and SUP05 have a similar presence at both stations (5% in PA2 and 3.1% in PA5 for the SAR324 lineage, 2.9% and 2.7% for the SUP05 lineage), with SAR324 slightly decreasing in the OMZ core ([Fig. 1](#)).

Anammox 16S rRNA genes are barely detected at the upper limit of the OMZ, consistent with previous analyses ([Pitcher et al., 2011](#); [Villanueva et al., 2014](#)), but comprise almost 5% of the community at the OMZ core. Interestingly, a recent study showed that autotrophic denitrifiers oxidizing hydrogen sulphide could form a stable community with anammox bacteria in a reactor system ([Russ et al., 2014](#)). Whether this is also the case at the OMZ core remains to be 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 ([Fig. S1](#)). Contig-1



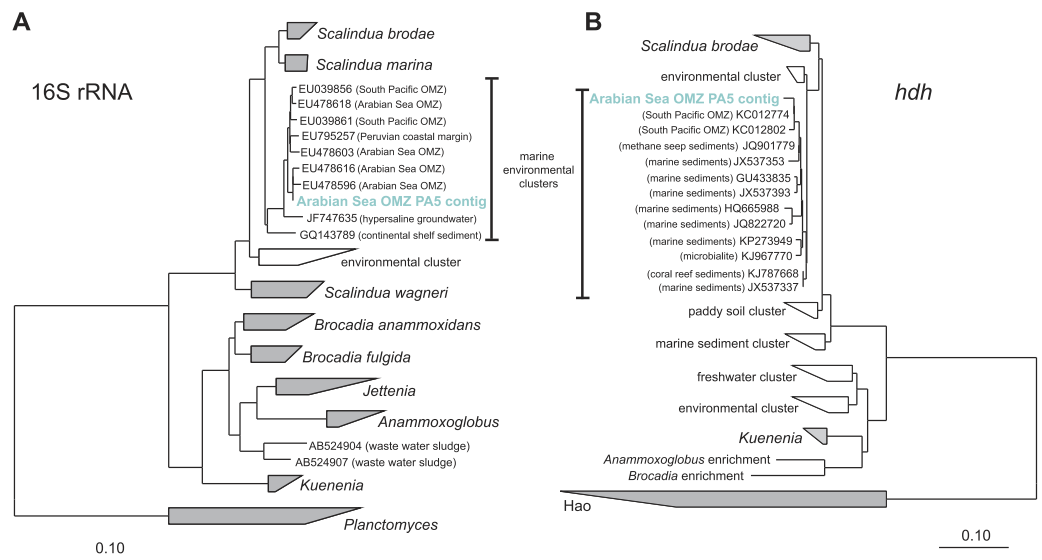
**Figure 2** Phylogenetic inference of thaumarchaeal contigs assembled from the OMZ metagenomes. The trees were calculated using the Neighbor Joining algorithm and based on 1,225 nucleotide positions for the 16S rRNA gene (A) and 144 deduced amino acid positions for the ammonia monooxygenase encoding gene (*amoA*) (B).

shows 98% identity to the very recently described ‘*Candidatus Nitrosopelagicus brevis*’ (Fig. 2) (Santoro *et al.*, 2015), contig-2 has only moderate identities to isolated or enriched ammonia oxidizers (93% identity to *Nitrosoarchaeum limnia* SFB1) (Fig. 2).

Reads from station PA5 clustering with the *Brocadia* family, 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 (Fig. 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.

### Genetic potential for nitrogen and methane cycling

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



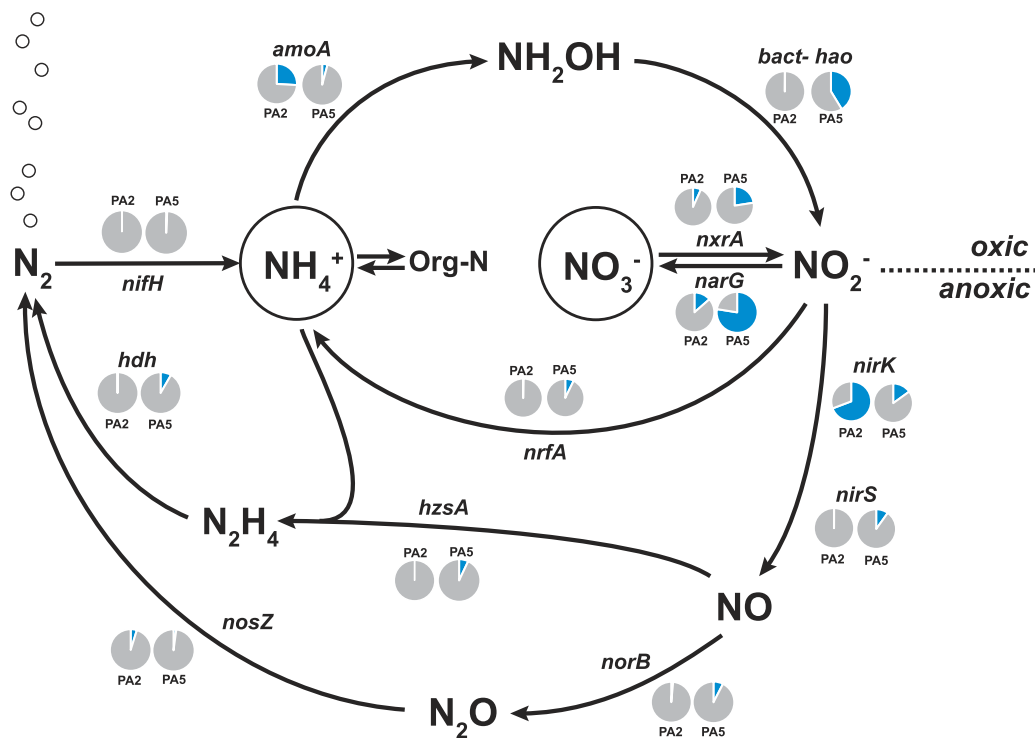
**Figure 3** Phylogenetic inference of *Scalindua*-related contigs assembled from the OMZ metagenomes. The tree was calculated using the Neighbor Joining algorithm and based on 1,388 nucleotide positions for the 16S rRNA gene (A) and 180 deduced amino acid positions for the hydrazine dehydrogenase encoding gene (*hdh*) (B).

## Nitrogen cycling potential

### Nitrification

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

The *amoA* reads could be assembled into 5 major contigs (4 contigs from site PA2 and 1 contig from PA5) that were compared to 16S rRNA phylogeny (Fig. 1). As for the 16S rRNA contig, *amoA* sequences could not be affiliated with a single MG-I species, but showed a diversity of at least 2 major genotypes. Unlike the 16S rRNA gene analysis, none of the contigs clustered with *Nitrosopelagicus*-like *amoA* sequences. Instead two out of the four obtained contigs from PA2 did not cluster with, but between the *Nitrosopelagicus*



**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.

cluster and *Nitrosoarchaeum limnia* SFB1. The two other contigs clustered within an environmental group only distantly related to described MG-I AOA. This environmental cluster also contained the only contig that could be assembled from the OMZ core. A niche differentiation between shallow and deep-water clades of MG-I *Thaumarchaeota* has also been described before in the Arabian Sea OMZ Villanueva, Schouten & Sinninghe Damsté (2014), but also in other marine environments (Beman, Popp & Francis, 2008; Santoro, Casciotti & Francis, 2010). Besides archaeal *amoA*, only a single read with low identity (<50% on the amino acid level) to known bacterial *amoA* reads was detected, indicating ammonium oxidizing bacteria (AOB) likely play only a small role in the Arabian sea OMZ although they were detected in other OMZs (Molina et al., 2007; Lam et al., 2009). Consistent with the absence of AOB *amoA*, only five reads matching *hao* were detected in the upper limit of the OMZ. In the core, 475 reads matched the *hao* database, but even after removal of the bonafide anammox hydrazine dehydrogenase hits (Fig. 3), over 90% of the *hao* matches were affiliated with *Scalindua*, which is known to encode up to ten paralogs of this protein (Van 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

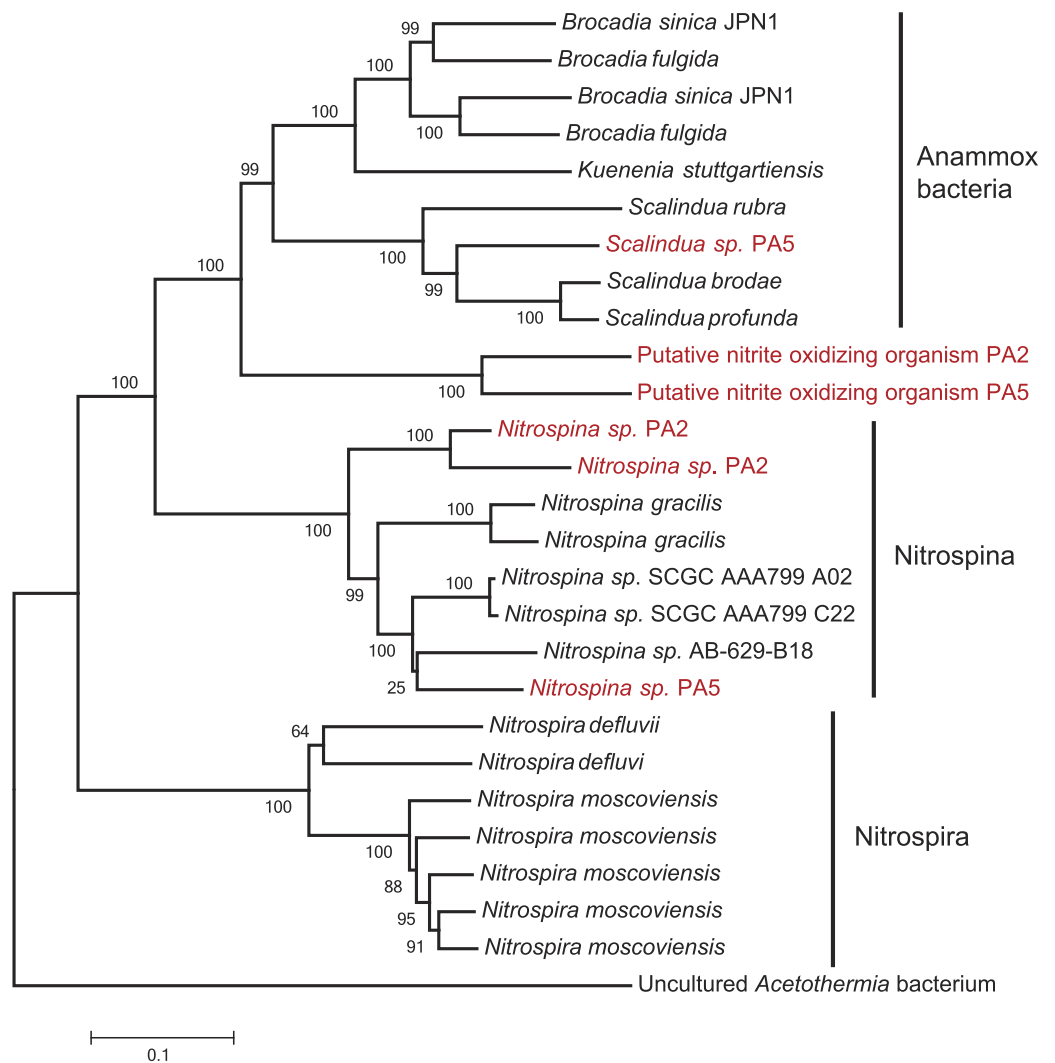
combined *narG/nxrA* reference set, and then used a second round of BLASTx and BLAST score ratio separation (threshold 0.85–0.95) to distinguish between *narG* and *nxrA*. No reads could be confidently assigned to *nxrA* of the *Nitrobacter/Nitrococcus/Nitrolanceus* group in either station. Conversely, 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 (Fig. 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 (Fig. 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 (Fig. 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).

No other nitrite oxidizing organisms were detected. The detection of a significant abundance of nitrite oxidizers in OMZ ecosystems is consistent with a previous study showing that nitrite oxidation was an active process in the Namibian OMZ (Füssel et al., 2012).

### **Genetic potential for processes contributing to nitrogen loss**

Nitrite and nitrate resulting from nitrification can be readily used in denitrification, anammox and DNRA. To date, it is still unclear which process is dominating or if a combination of various processes occurs. Most of the recent studies indicated anammox instead of denitrification as prevalent pathway in OMZs (Kuypers et al., 2005; Thamdrup et al., 2006; Hamersley et al., 2007; Lam et al., 2009). However, two reports described high and active denitrification rates in the Arabian Sea OMZ (Ward et al., 2009; Bulow et al., 2010). In yet another study, Jensen and co-workers found that anammox coupled to DNRA was the prevalent process in this system (Jensen et al., 2011) and further studies confirmed a high abundance of anammox bacteria at the core of the Arabian Sea OMZ (Pitcher et al., 2011; Villanueva et al., 2014).

Here, we found that the nitrate reductase is by far the most dominant nitrogen cycle enzyme encoded in the Arabian Sea OMZ core (78% of normalized *rpoB* gene abundance, Fig. 4). This is consistent with observations in the Peruvian oxygen minimum zone (Lam et al., 2009; Glass et al., 2015) and in the Eastern Tropical South Pacific OMZ (Stewart, Ulloa & DeLong, 2012). The reduction of nitrate to nitrite is a crucial step as the nitrite forms the starting point for many subsequent processes: nitrite reduction in denitrification, in DNRA and in anammox. Additionally, nitrite can also be re-oxidized to nitrate. The genetic potential for all these 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



**Figure 5** Phylogenetic inference of *nxrA* sequences from the OMZ metagenomes. The tree was calculated using the Neighbor Joining algorithm and based on 3,209 nucleotide positions. Bootstrap values represent 1,000 replicates.

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: *Coralimargarita 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 ([Fig. 4](#)). It has been previously suggested that nitrifier denitrification accounts for the majority of nitrous oxide observed in the ocean ([Babbin et al., 2015](#); [Kozłowski 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 ([Fig. 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, *nxA*, 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 ([Fig. 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 in the PA5 metagenome (1,6 million reads), which is 160 times lower and thus results in a diazotroph abundance estimate of 0.6% based on the dataset. PCR primers should be able to amplify genes present in organisms with this abundance. Alternatively, the diazotrophic community in our dataset was, for unknown reasons, lower than in other studies from similar ecosystems.

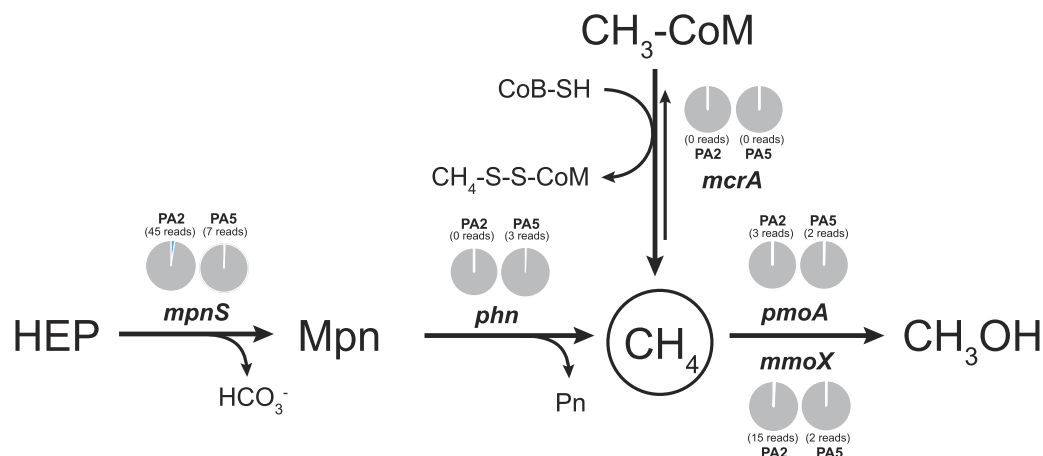
### Methane cycling potential

To examine the methane cycling potential in the Arabian Sea, we selected marker genes targeting methane production and methane oxidation ([Table 1](#)). The *mcrA* gene (encoding

the methyl-Coenzyme M reductase) was used as functional marker for methane production and anaerobic methane oxidation. We could not detect any *mcrA*-like sequences of canonical methanogens or sequences of the very recently described *mcrA* homologues from *Bathyarchaeota* (Evans et al., 2015). Also no sequences of anaerobic methane oxidizing archaea (ANME clades) were found. This is consistent with the absence of methanogen 16S rRNA genes and very low abundance (3 reads) of *Bathyarchaeota* 16S rRNA gene sequences. Thus, methanogenesis does not seem to play a major role in the Arabian Sea OMZ water column. Nevertheless, in marine ecosystems, a second pathway for methane production apart from methanogenesis was proposed (Karl et al., 2008). This aerobic pathway includes the cleavage of methylphosphonate (Mpn). Mpn was shown to be synthesized by the AOA *Nitrosopumilus maritimus* using the Mpn synthase as key enzyme and a homologue of this enzyme was also found to be encoded in SAR11 genomes (Metcalf et al., 2012). The cleavage of Mpn that results in the release of methane, is catalysed by the C-P lyase multi-enzyme complex (Daughton, Cook & 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* (Fig. 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* (Fig. 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 (Fig. 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 in the Arabian Sea OMZ, hence the ability for acquiring phosphorous from Mpn might be less important in this system.

In accordance with the absence of marker genes indicative for methane production, only few reads were retrieved that could be affiliated with methane oxidation. For aerobic methane oxidation, we used the marker genes *pmoA* and *mmoX* encoding the particulate and soluble methane monooxygenase. In the PA2 dataset, 15 *mmoX* and 3



**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. The description of all nitrogen and methane cycling marker genes is given in Table 1.

*pmoA*-like reads were detected. However, after closer inspection, all *mmoX* reads showed moderate sequence identity (up to 71% on amino acid level) to homologous toluene monooxygenases. For the *pmoA* reads, highest identity (68–97% on amino acid level) was found to the monooxygenase of the SAR324 clade (*deltaproteobacteria*). Based on 16S rRNA gene abundance, the SAR324 clade was found to belong, besides others, to the dominant bacteria in the Arabian Sea dataset (Fig. 1). So far, no enrichment of SAR324 clade bacteria is available and it is not known if this monooxygenase is used for methane or higher alkane oxidation. Whereas the *pmoA* phylogeny indicates relation to C2-C4 alkane monooxygenases (Li et al., 2014), the genomes of SAR324 members suggest the potential for both, C1 and higher alkane utilization (Sheik, Jain & Dick, 2014). Nevertheless, not all SAR324 genomes contain the alkane monooxygenase gene cluster (Swan et al., 2011). The PA5 dataset revealed 2 *pmoA* and 2 *mmoX*-like sequences, again only with moderate sequence identities to known alkane monooxygenases. Consistent with the absence of genes encoding the putative nitric oxide dismutase of NC10 phylum bacteria, we did not find any *pmoA* genes of these bacteria in the Arabian sea OMZ. Nevertheless, they might be present in a low abundance that could not be retrieved by our approach. Thus, if these bacteria play a role in methane and nitrogen cycling in OMZs, as suggested for the Eastern Pacific OMZ, needs to be investigated in future research (Padilla et al., 2016). In addition, 16S rRNA gene sequences of known aerobic methanotrophs were nearly absent. Only 3 reads clustering within gammaproteobacterial methanotrophs were present (1 read from PA2 and 2 reads from PA5). Although the overall abundance of hydrocarbon monooxygenase encoding reads is low in the Arabian Sea OMZ dataset, our analysis shows the existence of novel sequence diversity only moderately related to known sequences, that is not captured by currently used PCR primers.

## Conclusion

In this study, we compared the functional diversity in two metagenomes retrieved from the Arabian Sea oxygen minimum zone. Using manually curated reference databases, we screened the datasets for homologues indicative for nitrogen and methane turnover in this ecosystem. We are aware that the presence of genetic potential alone cannot be used to draw conclusions on activity of various processes. However, despite this limitation, the picture that emerges from our analysis is that the vast majority of organisms can contribute to nitrate reduction, probably coupled to degradation of organic matter and release of ammonium (*Kalvelage et al., 2015*). The nitrite formed by nitrate reduction can be re-oxidized to nitrate in a ‘nitrite loop’ ultimately resulting in removal of additional organic matter and release of more ammonium. The *nrf*-like nitrite reductase assembled from our dataset is only distantly related to described *nrf* sequences indicating that a yet unknown organism is responsible for this reaction in the Arabian Sea. The released ammonium can partially be oxidized by a diverse community of microaerophilic archaeal ammonium oxidizers, but in the core of the OMZ, the majority is likely converted by 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., 2008*). 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., 2016, 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.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

Samples used in this manuscript were collected during the PASOM cruise funded by the Netherlands Organization for Scientific Research (NWO 817.01.015). This work was further supported by the European Research Council (ERC AG 339880 Eco\_MoM and ERC AG 232937 anammox to MSMJ), the Dutch governments Gravitation grant (024002002 to MSMJ), and the Spinoza prize to MSMJ. CL and DRS were funded by BE-Basic FP 07.002.01. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

Netherlands Organization for Scientific Research: (NWO 817.01.015).

European Research Council: ERC AG 339880 Eco\_MoM, ERC AG 232937.

Gravitation grant: 024002002.

BE-Basic grant: FP 07.002.01.

### Competing Interests

The authors declare that they have no competing interests.

### Author Contributions

- Claudia Lüke and Daan R. Speth conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Martine A.R. Kox analyzed the data, contributed reagents/materials/analysis tools, reviewed drafts of the paper.
- Laura Villanueva conceived and designed the experiments, reviewed drafts of the paper, sampling.
- Mike S.M. Jetten conceived and designed the experiments, analyzed the data, reviewed drafts of the paper.

### Field Study Permissions

The following information was supplied relating to field study approvals (i.e., approving body and any reference numbers):

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.

### DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

NCBI Sequence Read Archive: [SRA304624](https://www.ncbi.nlm.nih.gov/sra/SRA304624).

### Data Availability

The following information was supplied regarding data availability:

Github: [https://github.com/dspeth/bioinfo\\_scripts/tree/master/proteins](https://github.com/dspeth/bioinfo_scripts/tree/master/proteins).

### Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.1924#supplemental-information>.

## REFERENCES

- Angel R, Matthies D, Conrad R. 2011. Activation of methanogenesis in arid biological soil crusts despite the presence of oxygen. *PLoS ONE* 6:e20453  
[DOI 10.1371/journal.pone.0020453](https://doi.org/10.1371/journal.pone.0020453).

- Babbin AR, Bianchi D, Jayakumar A, Ward BB. 2015.** Rapid nitrous oxide cycling in the suboxic ocean. *Science* **348**:1127–1129 DOI [10.1126/science.aaa8380](https://doi.org/10.1126/science.aaa8380).
- Baker B, Lazar CS, Teske AP, Dick GJ. 2015.** Genomic resolution of linkages in carbon, nitrogen, and sulfur cycling among widespread estuary sediment bacteria. *Microbiome* **3**:14 DOI [10.1186/s40168-015-0077-6](https://doi.org/10.1186/s40168-015-0077-6).
- Bange HW, Ramesh R, Rapsomanikis S, Andreae MO. 1998.** Methane in surface waters of the Arabian Sea. *Geophysical Research Letters* **25**:3547–3550 DOI [10.1029/98GL02710](https://doi.org/10.1029/98GL02710).
- Beman JM, Popp BN, Francis CA. 2008.** Molecular and biogeochemical evidence for ammonia oxidation by marine Crenarchaeota in the Gulf of California. *The ISME Journal* **2**:429–441 DOI [10.1038/ismej.2007.118](https://doi.org/10.1038/ismej.2007.118).
- Bird C, Wyman M. 2013.** Transcriptionally active heterotrophic diazotrophs are widespread in the upper water column of the Arabian Sea. *FEMS Microbiology Ecology* **84**:189–200 DOI [10.1111/1574-6941.12049](https://doi.org/10.1111/1574-6941.12049).
- Bulow SE, Rich JJ, Naik HS, Pratihary AK, Ward BB. 2010.** Denitrification exceeds anammox as a nitrogen loss pathway in the Arabian Sea oxygen minimum zone. *Deep-Sea Research Part I: Oceanographic Research Papers* **57**:384–393 DOI [10.1016/j.dsr.2009.10.014](https://doi.org/10.1016/j.dsr.2009.10.014).
- Canfield DE, Stewart FJ, Thamdrup B, De Brabandere L, Dalsgaard T, Delong EF, Revsbech NP, Ulloa O. 2010.** A cryptic sulfur cycle in oxygen-minimum-zone waters off the Chilean coast. *Science* **330**:1375–1378 DOI [10.1126/science.1196889](https://doi.org/10.1126/science.1196889).
- Codispoti LA, Brandes JA, Christensen JP, Devol AH, Naqvi SWA, Paerl HW, Yoshinari T. 2001.** The oceanic fixed nitrogen and nitrous oxide budgets: moving targets as we enter the anthropocene? *Scientia Marina* **65**:85–105 DOI [10.3989/scimar.2001.65s285](https://doi.org/10.3989/scimar.2001.65s285).
- Codispoti LA, Richards FA. 1976.** An analysis of the horizontal regime of denitrification in the eastern tropical North Pacific. *Limnology and Oceanography* **21**:379–388 DOI [10.4319/lo.1976.21.3.0379](https://doi.org/10.4319/lo.1976.21.3.0379).
- Daughton CG, Cook AM, Alexande M. 1979.** Biodegradation of phosphonate toxicants yields methane or ethane on cleavage of the C-P bond. *FEMS Microbiology Letters* **5**:91–93.
- Deutsch C, Sarmiento JL, Sigman DM, Gruber N, Dunne JP. 2007.** Spatial coupling of nitrogen inputs and losses in the ocean. *Nature* **445**:163–167 DOI [10.1038/nature05392](https://doi.org/10.1038/nature05392).
- Devol AH, Uhlenhopp AG, Naqvi SWA, Brandes JA, Jayakumar DA, Naik H, Gaurin S, Codispoti LA, Yoshinari T. 2006.** Denitrification rates and excess nitrogen gas concentrations in the Arabian Sea oxygen deficient zone. *Deep Sea Research Part I: Oceanographic Research Papers* **53**:1533–1547 DOI [10.1016/j.dsr.2006.07.005](https://doi.org/10.1016/j.dsr.2006.07.005).
- Dumont MG, Lüke C, Deng Y, Frenzel P. 2014.** Classification of pmoA amplicon pyrosequences using BLAST and the lowest common ancestor method in MEGAN. *Frontiers in Microbiology* **5**:Article 34 DOI [10.3389/fmicb.2014.00034](https://doi.org/10.3389/fmicb.2014.00034).
- Dupont CL, Rusch DB, Yooseph S, Lombardo M-J, Alexander Richter R, Valas R, Novotny M, Yee-Greenbaum J, Selengut JD, Haft DH, Halpern AL, Lasken RS,**

- Nealson K, Friedman R, Craig Venter J. 2012. Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *The ISME Journal* 6:1186–1199 DOI 10.1038/ismej.2011.189.
- Dutilh BE, Huynen MA, Strous M. 2009. Increasing the coverage of a metapopulation consensus genome by iterative read mapping and assembly. *Bioinformatics* 25:2878–2881 DOI 10.1093/bioinformatics/btp377.
- Dyhrman ST, Chappell PD, Haley ST, Moffett JW, Orchard ED, Waterbury JB, Webb EA. 2006. Phosphonate utilization by the globally important marine diazotroph *Trichodesmium*. *Nature* 439:68–71 DOI 10.1038/nature04203.
- Edgar RC. 2010. Supplementary Material—Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461 DOI 10.1093/bioinformatics/btq461.
- Evans PN, Parks DH, Chadwick GL, Robbins SJ, Orphan VJ, Golding SD, Tyson GW. 2015. Methane metabolism in the archaeal phylum *Bathyarchaeota* revealed by genome-centric metagenomics. *Science* 350:434–438.
- Field KG, Gordon D, Wright T, Rappé M, Urbach E, Vergin K, Giovannoni SJ. 1997. Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. *Applied and Environmental Microbiology* 63:63–70.
- Fuhrman JA, McCallum K, Davis AA. 1993. Phylogenetic diversity of subsurface marine microbial communities from the Atlantic and Pacific Oceans. *Applied and Environmental Microbiology* 9:1294–1302.
- Füssel J, Lam P, Lavik G, Jensen MM, Holtappels M, Günter M, Kuypers MM. 2012. Nitrite oxidation in the Namibian oxygen minimum zone. *The ISME Journal* 6:1200–1209 DOI 10.1038/ismej.2011.178.
- Gaby JC, Buckley DHA. 2014. A comprehensive aligned *nifH* gene database: a multipurpose tool for studies of nitrogen-fixing bacteria. *Database* 2014:Article bau001.
- Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, Bibbs L, Eads J, Richardson TH, Noordewier M, Rappé MS, Short JM, Carrington JC, Mathur EJ. 2005. Genome streamlining in a cosmopolitan oceanic bacterium. *Science* 309:1242–1245 DOI 10.1126/science.1114057.
- Glass JB, Kretz CB, Ganesh S, Ranjan P, Seston SL, Buck KN, Landing WM, Morton PL, Moffett JW, Giovannoni SJ, Vergin KL, Stewart FJ. 2015. Meta-omic signatures of microbial metal and nitrogen cycling in marine oxygen minimum zones. *Frontiers in Microbiology* 6:1–13.
- Gordon DA, Giovannoni SJ. 1996. Stratified microbial populations related to Chlorobium and Fibrobacter detected in the Atlantic and Pacific Oceans. *Applied and Environmental Microbiology* 62:1171–1177.
- Groffman PM, Altabet MA, Bohlke JK, Butterbach-Bahl K, David MB, Firestone MK, Giblin AE, Kana TM, Nielsen LP, Voyteck MA. 2006. Methods for measuring denitrification? *Ecological Applications* 16:2091–2122 DOI 10.1890/1051-0761(2006)016[2091:MFMDDA]2.0.CO;2.
- Gruber N, Sarmiento JL. 1997. Global patterns of marine nitrogen fixation and denitrification. *Global Biogeochemical Cycles* 11:235–266 DOI 10.1029/97GB00077.

- Hamersley MR, Lavik G, Woebken D, Rattray JE, Lam P, Hopmans EC, Damsté JSS, Krüger S, Graco M, Gutiérrez D, Kuypers MMM. 2007. Anaerobic ammonium oxidation in the Peruvian oxygen minimum zone. *Limnology and Oceanography* 52:923–933 DOI 10.4319/lo.2007.52.3.0923.
- Huang J, Su Z, Xu Y. 2005. The evolution of microbial phosphonate degradative pathways. *Journal of Molecular Evolution* 61:682–690 DOI 10.1007/s00239-004-0349-4.
- Hug LA, Thomas BC, Sharon I, Brown CT, Sharma R, Hettich RL, Wilkins MJ, Williams KH, Singh A, Banfield JF. 2016. Critical biogeochemical functions in the subsurface are associated with bacteria from new phyla and little studied lineages. *Environmental Microbiology* 18:159–173 DOI 10.1111/1462-2920.12930.
- Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. *Genome Research* 17:377–386 DOI 10.1101/gr.5969107.
- Huson DH, Mitra S, Ruscheweyh H-J, Wever N, Schuster SC. 2011. Integrative analysis of environmental sequences using MEGAN4. *Genome Research* 21:1552–1560 DOI 10.1101/gr.120618.111.
- Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV. 2012. Untangling genomes from metagenomes: revealing an uncultured class of marine euryarchaeota. *Science* 335:587–590 DOI 10.1126/science.1212665.
- Jayakumar A, Al-Rshaidat MMD, Ward BB, Mulholland MR. 2012. Diversity, distribution, and expression of diazotroph *nifH* genes in oxygen-deficient waters of the Arabian Sea. *FEMS Microbiology Ecology* 82:597–606 DOI 10.1111/j.1574-6941.2012.01430.x.
- Jayakumar DA, Francis CA, Naqvi SWA, Ward BB. 2004. Diversity of nitrite reductase genes (*nirS*) in the denitrifying water column of the coastal Arabian Sea. *Aquatic Microbial Ecology* 34:69–78 DOI 10.3354/ame034069.
- Jayakumar DA, Naqvi SWA, Narvekar PV, George MD. 2001. Methane in coastal and offshore waters of the Arabian Sea. *Marine Chemistry* 74:1–13 DOI 10.1016/S0304-4203(00)00089-X.
- Jensen MM, Lam P, Revsbech NP, Nagel B, Gaye B, Jetten MS, Kuypers MM. 2011. Intensive nitrogen loss over the Omani Shelf due to anammox coupled with dissimilatory nitrite reduction to ammonium. *The ISME Journal* 5:1660–1670 DOI 10.1038/ismej.2011.44.
- Kalvelage T, Jensen MM, Contreras S, Revsbech NP, Lam P, Günter M, LaRoche J, Lavik G, Kuypers MMM. 2011. Oxygen sensitivity of anammox and coupled N-cycle processes in oxygen minimum zones. *PLoS ONE* 6:e29299 DOI 10.1371/journal.pone.0029299.
- Kalvelage T, Lavik G, Jensen MM, Revsbech NP, Löscher C, Schunck H, Desai DK, Hauss H, Kiko R, Holtappels M, LaRoche J, Schmitz RA, Graco MI, Kuypers MMM. 2015. Aerobic microbial respiration in oceanic oxygen minimum zones. *PLoS ONE* 10:e0133526 DOI 10.1371/journal.pone.0133526.
- Karl DM, Beversdorf L, Bjorkman KM, Church MJ, Martinez A, Delong EF. 2008. Aerobic production of methane in the sea. *Nature Geoscience* 1:473–478 DOI 10.1038/ngeo234.



- Knittel K, Boetius A. 2009.** Anaerobic oxidation of methane: progress with an unknown process. *Annual Review of Microbiology* **63**:311–334  
[DOI 10.1146/annurev.micro.61.080706.093130](https://doi.org/10.1146/annurev.micro.61.080706.093130).
- Kozłowski JA, Stieglmeier M, Schleper C, Klotz MG, Stein LY. 2016.** Pathways and key intermediates required for obligate aerobic ammonia-dependent chemolithotrophy in bacteria and *Thaumarchaeota*. *The ISME Journal* Epub ahead of print Feb 16 2016  
[DOI 10.1038/ismej.2016.2](https://doi.org/10.1038/ismej.2016.2).
- Kuypers MMM, Kuypers MMM, Lavik G, Lavik G, Woebken D, Woebken D, Schmid M, Schmid M, Fuchs BM, Fuchs BM, Amann R, Amann R, Jørgensen BB, Jørgensen BB, Jetten MSM, Jetten MSM. 2005.** Massive nitrogen loss from the Benguela upwelling system through anaerobic ammonium oxidation. *Proceedings of the National Academy of Sciences of the United States of America* **102**:6478–6483  
[DOI 10.1073/pnas.0502088102](https://doi.org/10.1073/pnas.0502088102).
- Kuypers MMM, Sliemers AO, Lavik G, Schmid M, Jørgensen BB, Kuenen JG, Sinninghe Damste JS, Strous M, Jetten MSM. 2003.** Anaerobic ammonium oxidation by anammox bacteria in the Black Sea. *Nature* **422**:608–611 [DOI 10.1038/nature01472](https://doi.org/10.1038/nature01472).
- Lam P, Kuypers MMM. 2011.** Microbial nitrogen cycling processes in oxygen minimum zones. *Annual Review of Marine Science* **3**:317–345  
[DOI 10.1146/annurev-marine-120709-142814](https://doi.org/10.1146/annurev-marine-120709-142814).
- Lam P, Lavik G, Jensen MM, Van De Vossenberg J, Schmid M, Woebken D, Gutiérrez D, Amann R, Jetten MSM, Kuypers MMM. 2009.** Revising the nitrogen cycle in the Peruvian oxygen minimum zone. *Proceedings of the National Academy of Sciences of the United States of America* **106**:4752–4757 [DOI 10.1073/pnas.0812444106](https://doi.org/10.1073/pnas.0812444106).
- Lavik G, Stührmann T, Bruchert V, Van Der Plas A, Mohrholz V, Lam P, Muszmann M, Fuchs BM, Amann R, Lass U, Kuypers MMM. 2009.** Detoxification of sulphidic African shelf waters by blooming chemolithotrophs. *Nature* **457**:581–584  
[DOI 10.1038/nature07588](https://doi.org/10.1038/nature07588).
- Li M, Jain S, Baker BJ, Taylor C, Dick GJ. 2014.** Novel hydrocarbon monooxygenase genes in the metatranscriptome of a natural deep-sea hydrocarbon plume. *Environmental Microbiology* **16**:60–71 [DOI 10.1111/1462-2920.12182](https://doi.org/10.1111/1462-2920.12182).
- Lipschultz F, Wofsy SC, Ward BB, Codispoti LA, Friedrich G, Elkins JW. 1990.** Bacterial 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 [DOI 10.1016/0198-0149\(90\)90060-9](https://doi.org/10.1016/0198-0149(90)90060-9).
- 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 of the United States of America **107**:13479–13484  
DOI [10.1073/pnas.1003860107](https://doi.org/10.1073/pnas.1003860107).
- 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  
DOI [10.1111/j.1462-2920.2009.02062.x](https://doi.org/10.1111/j.1462-2920.2009.02062.x).
- 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 DOI [10.1126/science.1219875](https://doi.org/10.1126/science.1219875).
- Metcalf WW, Wanner BL. 1993.** Mutational analysis of an *Escherichia coli* fourteen-gene operon for phosphonate degradation, using TnpA 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**:3547–3555  
DOI [10.1128/AEM.02275-06](https://doi.org/10.1128/AEM.02275-06).
- Morris RM, Vergin KL, Cho J-C, Rappé MS, Carlson CA, 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 DOI [10.4319/lo.2005.50.5.1687](https://doi.org/10.4319/lo.2005.50.5.1687).
- Newell SE, Babbín AR, Jayakumar A, Ward BB. 2011.** Ammonia oxidation rates and nitrification in the Arabian Sea. *Global Biogeochemical Cycles* **25**:1–10  
DOI [10.1029/2010GB003940](https://doi.org/10.1029/2010GB003940).
- 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  
DOI [10.1038/ismej.2014.256](https://doi.org/10.1038/ismej.2014.256).
- 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 DOI [10.1038/ismej.2014.260](https://doi.org/10.1038/ismej.2014.260).
- Owens NJP, Law CS, Mantoura RFC, Burkill PH, Llewellyn CA. 1991.** Methane flux to the atmosphere from the Arabian Sea. *Nature* **354**:293–296 DOI [10.1038/354293a0](https://doi.org/10.1038/354293a0).
- Padilla CC, Bristow LA, Sarode N, Garcia-Robledo E, Gómez Ramírez E, Benson CR, Bourbonnais A, Altabet MA, Girguis PR, Thamdrup B, Stewart FJ. 2016.** NC10 bacteria in marine oxygen minimum zones. *The ISME Journal* Epub ahead of print Feb 26 2016 DOI [10.1038/ismej.2015.262](https://doi.org/10.1038/ismej.2015.262).
- 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  
DOI [10.1111/j.1462-2920.2011.02666.x](https://doi.org/10.1111/j.1462-2920.2011.02666.x).

- Pitcher A, Villanueva L, Hopmans EC, Schouten S, Reichart G-J, Sinninghe Damsté JS. 2011. Niche segregation of ammonia-oxidizing archaea and anammox bacteria in the Arabian Sea oxygen minimum zone. *The ISME Journal* 5:1896–1904 DOI 10.1038/ismej.2011.60.
- Pruesse E, Peplies J, Glockner FO. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28:1823–1829 DOI 10.1093/bioinformatics/bts252.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41:D590–D596 DOI 10.1093/nar/gks1219.
- Rappe MS, Connon SA, Vergin KL, Giovannoni SJ. 2002. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* 418:630–633 DOI 10.1038/nature00917.
- Rasko DA, Myers GSA, Ravel J. 2005. Visualization of comparative genomic analyses by BLAST score ratio. *BMC Bioinformatics* 6:2 DOI 10.1186/1471-2105-6-2.
- Redfield AC, Ketchum BH, Richards FA. 1963. The influence of organisms on the composition of sea-water. In: Hill MN, ed. *The sea*. Vol. 2. New York: Wiley Interscience, 26–77.
- 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. *Limnology and Oceanography: Methods* 7:371–381 DOI 10.4319/lom.2009.7.371.
- Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling AE, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu W-T, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T. 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499:431–437 DOI 10.1038/nature12352.
- Russ L, Speth DR, Jetten MSM, Op Den Camp HJM, Kartal B. 2014. Interactions between anaerobic ammonium and sulfur-oxidizing bacteria in a laboratory scale model system. *Environmental Microbiology* 16:3487–3498 DOI 10.1111/1462-2920.12487.
- Sansone F, Popp B. 2001. Highly elevated methane in the eastern tropical North Pacific and associated isotopically enriched fluxes to the atmosphere. *Geophysical Research Letters* 28:4567–4570 DOI 10.1029/2001GL013460.
- Santoro AE, Casciotti KL, Francis CA. 2010. Activity, abundance and diversity of nitrifying archaea and bacteria in the central California Current. *Environmental Microbiology* 12:1989–2006 DOI 10.1111/j.1462-2920.2010.02205.x.
- Santoro AE, Dupont CL, Richter RA, Craig MT, Carini P, McIlvin MR, Yang Y, Orsi WD, Moran DM, Saito MA. 2015. Genomic and proteomic characterization of “*Candidatus Nitrosopelagicus brevis*”: an ammonia-oxidizing archaeon from the open ocean. *Proceedings of the National Academy of Sciences of the United States of America* 112:1173–1178 DOI 10.1073/pnas.1416223112.

- Schouten S, Pitcher A, Hopmans EC, Villanueva L, Van Bleijswijk J, Sinninghe Damsté JS. 2012. Intact polar and core glycerol dibiphytanyl glycerol tetraether lipids in the Arabian Sea oxygen minimum zone: I. Selective preservation and degradation in the water column and consequences for the TEX86. *Geochimica et Cosmochimica Acta* 98:228–243 DOI 10.1016/j.gca.2012.05.002.
- Sheik CS, Jain S, Dick GJ. 2014. Metabolic flexibility of enigmatic SAR324 revealed through metagenomics and metatranscriptomics. *Environmental Microbiology* 16:304–317 DOI 10.1111/1462-2920.12165.
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert 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:Article 539 DOI 10.1038/msb.2011.75.
- 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 DOI 10.1146/annurev-micro-092611-150128.
- Stewart FJ, Ulloa O, DeLong EF. 2012. Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environmental Microbiology* 14:23–40 DOI 10.1111/j.1462-2920.2010.02400.x.
- 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 DOI 10.1126/science.1203690.
- Tavormina PL, Ussler W, Steele JA, Cannon SA, 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 DOI 10.1111/1758-2229.12025.
- Thamdrup B, Dalsgaard T, Jensen MM, Ulloa O, Farías L, Escobedo R. 2006. Anaerobic ammonium oxidation in the oxygen-deficient waters off northern Chile. *Limnology and Oceanography* 51:2145–2156 DOI 10.4319/lo.2006.51.5.2145.
- 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 DOI 10.1016/j.dsr.2012.03.001.
- 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 DOI 10.1029/1999JC900232.
- 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 DOI 10.1111/j.1462-2920.2012.02774.x.
- 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* 7:1322–1332 DOI 10.1038/ismej.2013.32.
- 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 DOI 10.3389/fmicb.2014.00031.
- 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.
- Walsh DA, Zaikova E, Howes CG, Song YC, Wright JJ, Tringe SG, Tortell PD, Hallam SJ. 2009. Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science* 326:578–582 DOI 10.1126/science.1175309.
- Ward BB, Devol AH, Rich JJ, Chang BX, Bulow SE, Naik H, Pratihary A, Jayakumar A. 2009. Denitrification as the dominant nitrogen loss process in the Arabian Sea. *Nature* 461:78–81 DOI 10.1038/nature08276.
- White AE, Karl DM, Björkman K, Beversdorf LJ, Letelier RM. 2010. Production of organic matter by *Trichodesmium* IMS101 as a function of phosphorus source. *Limnology and Oceanography* 55:1755–1767 DOI 10.4319/lo.2010.55.4.1755.
- Woeckel D, Lam P, Kuypers MMM, Naqvi SWA, Kartal B, Strous M, Jetten MSM, Fuchs BM, Amann RI. 2008. A microdiversity study of anammox bacteria reveals a novel *Candidatus Scalindua* phylotype in marine oxygen minimum zones. *Environmental Microbiology* 10:3106–3119 DOI 10.1111/j.1462-2920.2008.01640.x.
- Wright JJ, Konwar KM, Hallam SJ. 2012. Microbial ecology of expanding oxygen minimum zones. *Nature Reviews Microbiology* 10:381–394 DOI 10.1038/nrmicro2778.
- Zeigler Allen L, Allen EE, Badger JH, McCrow JP, Paulsen IT, Elbourne LDH, Thiagarajan M, Rusch DB, Nealson KH, Williamson S, Venter JC, Allen AE. 2012. Influence of nutrients and currents on the genomic composition of microbes across an upwelling mosaic. *The ISME Journal* 6:1403–1414 DOI 10.1038/ismej.2011.201.