Title

Single cell genomic and transcriptomic evidence for the use of alternative nitrogen substrates by anammox bacteria

Journal

The ISME Journal: Multidisciplinary Journal of Microbial Ecology, 12(11)

ISSN

1751-7362

Authors

Ganesh, Sangita

Bertagnolli, Anthony D

Bristow, Laura A

et al.

Publication Date

2018-11-01

DOI

10.1038/s41396-018-0223-9

Peer reviewed

Single cell genomic and transcriptomic evidence for the use of alternative nitrogen substrates by anammox bacteria

Contributors: Sangita Ganesh[1,2]*, Anthony D. Bertagnolli[1]*, Laura A. Bristow[3], Cory C. Padilla[1], Nigel Blackwood[4], Montserrat Aldunate[5,6], Annie Bourbonnais[7,8], Mark A. Altabet[8], Rex R. Malmstrom[9], Tanja Woyke[9], Osvaldo Ulloa[6], Konstantinos T. Konstantinidis[10], Bo Thamdrup[11], Frank J. Stewart[1#]

Affiliations and footnotes:

* These authors contributed equally to this work.
# Corresponding author**.** Mailing address: School of Biological Sciences, Georgia Institute of Technology, Ford ES&T Building, Rm 1242, 311 Ferst Drive, Atlanta GA 30332
Email: frank.stewart@biology.gatech.edu

[1] School of Biological Sciences, Georgia Institute of Technology, Atlanta GA, USA, 30332
[2] Radiant Genomics, Emeryville CA, USA, 94608
[3] Biogeochemistry Group, Max Planck Institute for Marine Microbiology, Bremen, Germany
[4] Department of Biology, University of Pennsylvania, Philadelphia PA, USA, 19104
[5] Graduate Program in Oceanography, Department of Oceanography, Faculty of Natural Sciences and Oceanography, University of Concepción, Casilla 160-C, Concepción, Chile
[6] Departamento de Oceanografía, Universidad de Concepción, Casilla 160-C, 4070386, Concepción, Chile
[7] Marine Chemistry & Geochemistry, Woods Hole Oceanographic Institution, 266 Woods Hole Road, Woods Hole MA, USA, 02543
[8] School for Marine Science and Technology, University of Massachusetts Dartmouth, 706 Rodney French Blvd, New Bedford MA, USA, 02744
[9] Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek CA, USA, 94598
[10] School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta GA USA, 30332
[11] Department of Biology and Nordic Center for Earth Evolution (NordCEE), University of Southern Denmark, Odense, Denmark

Running title: Alternative N use potential in anammox bacteria

Subject category: Microbial ecology and functional diversity of natural habitats

Conflict of Interest: The authors declare no financial, commercial, or personal conflict of interest involving the publication of this work.

Abstract

Anaerobic ammonium oxidation (anammox) contributes substantially to ocean nitrogen loss, particularly in anoxic marine zones (AMZs). Ammonium is scarce in AMZs, raising the hypothesis that organic nitrogen compounds may be ammonium sources for anammox. Biochemical measurements suggest that the organic compounds urea and cyanate can support anammox in AMZs. However, it is unclear if anammox bacteria degrade these compounds to ammonium themselves, or rely on other organisms for this process. Genes for urea degradation have not been found in anammox bacteria, and genomic evidence for cyanate use for anammox is limited to a cyanase gene recovered from the sediment bacterium *Candidatus* Scalindua profunda. Here, analysis of *Ca*. Scalindua single amplified genomes from the Eastern Tropical North Pacific AMZ revealed genes for urea degradation and transport, as well as for cyanate degradation. Urease and cyanase genes were transcribed, along with anammox genes, in the AMZ core where anammox rates peaked. Homologs of these genes were also detected in meta-omic datasets from major AMZs in the Eastern Tropical South Pacific and Arabian Sea. These results suggest that anammox bacteria from different ocean regions can directly access organic nitrogen substrates. Future studies should assess if and under what environmental conditions these substrates contribute to the ammonium budget for anammox.

Keywords: Candidatus Scalindua, urease, cyanase, ammonium, oxygen minimum zone, nitrogen cycle

Introduction

Anaerobic ammonium oxidation (anammox) plays a major role in aquatic nutrient cycling. In this microbial process, autotrophic bacteria oxidize ammonium with nitrite, producing energy for CO_2 fixation and cellular growth, and N_2 as an end product. Anammox is therefore a nitrogen sink, converting bioavailable nitrogen to a gaseous form unavailable to most organisms. Anammox is particularly important in anoxic marine zones (AMZs) (Thamdrup *et al.*, 2006; Dalsgaard *et al.*, 2012; Ulloa *et al*. 2012). In the major AMZs of the Eastern Tropical Pacific and Arabian Sea, dissolved oxygen is below detection (<10 nM), nitrite is abundant (often > 5 μM), and anammox combined with heterotrophic denitrification drive nitrogen loss, with ~20-40% of ocean nitrogen loss occurring in AMZs (Codispoti *et al.*, 2001; Sabine *et al.*, 2004; Thamdrup *et al.*, 2012; Tiano *et al.*, 2014; Ganesh *et al.*, 2015). In these systems, anammox bacteria increase in abundance at anoxic depths, where ammonium is supplied by the mineralization of organic matter, yet concentrations are low and turnover and competition for this resource are high (Woebken *et al*., 2008; Kalvelage *et al.*, 2013). Under these conditions, anammox bacteria may be under pressure to use alternative substrates as ammonium sources, or potentially to use energy and biomass production pathways other than anammox. Indeed, in experimental studies of anammox in AMZs, the direct use of organics as a source of ammonium by anammox bacteria was proposed as an explanation for higher anammox rates in incubations with ^{15}N-nitrite compared to those with ^{15}N-ammonium, as the former would integrate N_2 production from anammox based on all ammonium sources (Nicholls *et al.*, 2007; De Brabandere *et al.*, 2014). However, the metabolic versatility of marine anammox bacteria remains largely unknown. This is due in part to limited genomic characterizations of anammox bacteria from diverse marine habitats, including AMZs.

Anammox has thus far been described only in bacteria of the Order Brocadiales in the phylum Planctomycetes. This Order occurs globally in natural and man-made environments in both fresh and saltwater. No Brocadiales bacteria have yet been isolated in pure culture. The known diversity of this group is distributed across the *Candidatus* genera Brocadia, Kuenenia, Anammoxoglobus, Jettenia, and Scalindua. Of these, *Ca*. Scalindua is the dominant genus in ocean habitats, including sediments and AMZs (Woebken *et al.*, 2008; Villanueva *et al.*, 2014), but has also been found in freshwater (Sonthiphand *et al.*, 2014). Insight into the genomic potential of *Ca*. Scalindua is based on metagenomic contigs of *Ca*. S. profunda from marine sediment (van de Vossenberg *et al.*, 2013), *Ca*. S. brodae from a wastewater plant (Speth *et al.*, 2015), and *Ca*. S. rubra from a marine brine pool (Speth *et al.*, 2017). These large genomes (>4000 genes; ~4-5.2 Mbp) contain many genes absent from characterized genomes of other anammox genera, but also vary in gene content among species. For example, of these three species, only *Ca*. S. rubra contains genes for gas vesicle biosynthesis, presumably as an adaptation for regulating position in brine pools. However, genomes of *Ca*. Scalindua cells from AMZs have not yet been reported. Our knowledge of gene content in *Ca*. Scalindua from AMZs is based on recruitment of meta-omic sequences to non-AMZ genomes, e.g., of *Ca*. Scalindua profunda (van de Vossenberg *et al.*, 2013; Ganesh *et al.*, 2015; Luke *et al.*, 2016). We therefore have limited understanding of how anammox bacteria may be adapted to AMZ conditions.

The waste product urea ($CO(NH_2)_2$) and its breakdown product cyanate (OCN^-) are potential alternative substrates for anammox bacteria. Urea is ubiquitous in ocean waters, originating from microbial degradation of dissolved organic matter and nitrogenous waste from microbes and

111 animals (Zehr and Ward, 2002). Diverse microorganisms produce urease enzymes that
112 hydrolyze urea to ammonia and CO_2, potentially to aid pH regulation or to acquire ammonia for
113 biomass production or energy generation (Konieczna *et al.*, 2012). Ureases have even been
114 found in aerobic ammonia-oxidizing bacteria (Burton and Prosser, 2001) and archaea (Hallam *et*
115 *al.*, 2006; Qin *et al.*, 2014) as well as nitrite-oxidizing bacteria (Koch *et al.*, 2015), suggesting
116 that organic nitrogen plays a role in nitrification. Indeed, certain aerobic ammonia-oxidizing
117 bacteria not only oxidize ammonia derived from urea, but also assimilate the CO_2 resulting from
118 urease activity (Marsh *et al.*, 2005). Recently, anammox bacteria from the Eastern Tropical
119 South Pacific (ETSP) AMZ were shown to produce N_2 from added urea, but only after a lag of
120 1.5 days (Babbin *et al.*, 2017). This was interpreted as evidence that anammox bacteria do not
121 degrade urea directly but instead rely on the urealytic activity of other organisms or on abiotic
122 urea degradation to supply ammonium. An inability of anammox bacteria to directly degrade
123 urea is supported by the absence of urease-encoding genes (*ure*) in available anammox genomes.
124
125 In contrast to urea, cyanate addition stimulated N_2 production by anammox without a lag phase
126 in incubations of AMZ water (Babbin *et al.*, 2017). This suggests that AMZ anammox bacteria
127 might use cyanate directly, presumably though conversion to ammonia and CO_2 by a cyanase
128 enzyme, with the resulting ammonium used for anammox. Although absent from draft genomes
129 of other *Ca*. Scalindua species (Speth *et al.*, 2017), a putative cyanase-encoding gene (*cynS*) is
130 present in the metagenome of *Ca*. S. profunda from sediment (van de Vossenberg *et al.*, 2013),
131 and sequences related to this gene were detected in AMZ metagenomes and metatranscriptomes
132 (Babbin *et al.*, 2017). However, data conclusively linking cyanases to anammox bacterial
133 genomes from AMZs are not yet available, and it is therefore unknown if these bacteria might

also rely on other microbes for cyanate degradation, as has been shown for certain aerobic ammonia oxidizers (Palatinszky *et al.*, 2015).

Here, we explored the metabolic properties of *Ca*. Scalindua sp. from a marine AMZ, testing the hypothesis that these bacteria have the potential for directly catabolizing organic nitrogen substrates as ammonium sources for anammox. We explored this hypothesis using genomes of *Ca*. Scalindua cells from the Eastern Tropical North Pacific (ETNP) AMZ off Mexico. These genomes were then analyzed in conjunction with ETNP chemical concentration, anammox rate, and metatranscriptome data from a cruise in 2014. The results provide insight into the genetic basis for environmental variation and adaptation in this globally important lineage.

Materials and Methods

Sample collection

Samples for single amplified genome (SAG) analysis were collected in 2013 from station 6T (18° 54.0N, 104° 54.0W; Figure S1) in the ETNP AMZ during the Oxygen Minimum Zone Microbial Biogeochemistry Expedition (OMZoMBiE) cruise (*R/V New Horizon*; 13-28 June). Seawater for cell sorting and SAG sequencing was collected from the secondary nitrite maximum (125 m) and AMZ core (300 m) using Niskin bottles on a rosette containing a Conductivity-Temperature-Depth profiler (Sea-Bird SBE 911plus). From each depth, triplicate 1 ml samples of bulk seawater (no pre-filtration) were aliquoted into sterile cryovials and 100 µl of a glycerol TE stock solution (20 ml 100X TE pH 8.0, 60 ml deionized water, 100 ml glycerol) was added to each vial. The vials were then mixed and frozen at -80ºC.

157 Samples for metatranscriptome analysis and measurements of anammox rates were collected
158 from the ETNP AMZ during a second OMZoMBiE cruise, in 2014 (*R/V New Horizon*; 10 May –
159 8 June, 2014). Water was collected at six stations spanning a coastal to offshore gradient (Figure
160 S1). Stations and depths sampled for metatranscriptomics (n=21) are in Table S1. Eight of the
161 21 metatranscriptome datasets were generated in this study; the remainder were generated in two
162 prior studies (Padilla *et al.,* 2016; Garcia-Robledo *et al.,* 2017) and re-analyzed here. Seawater
163 was collected by Niskin with microbial biomass then collected by in-line filtration of seawater
164 (~1.5-2.5 L) through a glass fiber disc prefilter (GF/A, 47 mm, 1.6 μm pore-size, Whatman) and
165 a primary collection filter (Sterivex™, 0.22 μm pore-size, Millipore) using a peristaltic pump.
166 Sterivex™ filters were filled with RNA stabilizing buffer (25 mM sodium citrate, 10 mM EDTA,
167 5.3M ammonium sulfate, pH 5.2), flash-frozen in liquid nitrogen, and stored at -80°C.
168 Approximately 15-45 min elapsed (depending on depth) between capture in the Niskin and
169 arrival on deck; approximately 20 min elapsed between water retrieval from the Niskin and
170 fixation of filters in buffer.
171
172 Ammonium concentrations were determined fluorometrically aboard ship using the
173 orthophaldialdehyde method (Holmes *et al.*, 1999), with a detection limit of 10 nM. Samples for
174 measuring nitrite concentrations were collected in acid-cleaned HDPE bottles and stored frozen
175 until spectrophotometric measurement using the Griess method (Grasshoff *et al.*, 1983) with a
176 Westco SmartChem 200 (Unity Scientific). On a cruise to the study area in 2017 (*R/V Oceanus*
177 cruise OC1705), urea concentrations were determined fluorometrically following Mulvenna and
178 Savidge (1992) with a 5 cm cuvette. This method has a detection limit of 45 nM.
179

Anammox rate measurements

Anammox rates were measured for 14 of the 21 water samples from which metatranscriptomes were generated (Figures 1, S2, Table S1). Water was sampled directly from the Niskin and transferred to 250 ml glass bottles without pre-filtration. Bottles were overflowed (three volume equivalents) and sealed without bubbles with deoxygenated butyl rubber stoppers to minimize oxygen contamination (following De Brabandere *et al.,* 2014). Within 6 hours of collection, each bottle was amended with 5 μM $^{15}NH_4+$, and purged with helium for ~20 min. With a slight overpressure, water was dispensed into 12 ml exetainers (Labco, Lampeter, Ceredigion, UK), which were immediately capped with deoxygenated lids. Headspaces of 2 ml were introduced into each exetainer and flushed twice with helium, with shaking between flushings. Exetainers were then incubated in the dark at *in situ* temperature (13□C) for 24 hours. For each sample, triplicate exetainers were preserved with 100 μl of 50% (w/v) $ZnCl_2$ at the start of the incubation and again after 24 hours.

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

SAG generation and taxonomic screening

SAGs were generated from individual bacterial cells according to the Department of Energy Joint Genome Institute workflow following Rinke *et al.* (2013, 2014) with minor modifications

(as in Tsementzi *et al.*, 2016). Cells were sorted on a BD Influx (BD Biosciences) and treated with Ready-Lyse lysozyme (Epicentre; 5U/μL final conc.) for 15 min at room temperature prior to adding lysis solution. Whole genomes were amplified by multiple displacement amplification (MDA) using the REPLI-g Single Cell Kit (Qiagen), with final reaction volumes of 2 □L and termination after 6 hours. The taxonomic identity of each SAG was determined by PCR amplification and Sanger sequencing of a ~470 bp region of the 16S rRNA gene using primers 926wF (5'-AAACTYAAAKGAATTGRCGG- 3') and 1392R (5'-ACGGGCGGTGTGTRC- 3') for archaea and bacteria. Recovered sequences (average length: 423 bp) were classified using MOTHUR's 'classify_seq' against the Greengenes database, with the probability of correct assignment to a taxonomic group calculated using the naïve Bayesian classifier method (Wang *et al.*, 2007).

SAG sequencing

20 SAGs classified with high confidence as belonging to the genus *Ca*. Scalindua were selected for genome sequencing. These included 9 and 11 SAGs from 125 m and 300 m, respectively. Indexed DNA sequencing libraries were prepared using the Nextera XT DNA Library Prep kit (Illumina, San Diego, CA, USA) following manufacturer instructions, pooled, and sequenced on an Illumina MiSeq using a v2-500 cycle (paired end 250x250 bp) kit.

RNA extraction and cDNA sequencing

RNA was extracted from Sterivex™ filters as in Ganesh *et al.* (2015) using a modification of the mirVana^{TM} miRNA Isolation kit (Ambion). Filter cartridges were thawed on ice, and RNA stabilizing buffer was expelled by syringe from each cartridge and discarded. Cells were lysed by

adding Lysis buffer and miRNA Homogenate Additive (Ambion). Following vortexing and incubation on ice, lysates were transferred to RNAase-free tubes and RNA extracted by acid phenol:chloroform according to the kit. The TURBO DNA-free™ kit (Ambion) was used to remove DNA and the extract purified using the RNeasy MinElute Cleanup Kit (Qiagen). RNA was prepared for sequencing using the ScriptSeqTM v2 RNA-Seq Library preparation kit (Epicentre). cDNA was synthesized from fragmented total RNA (rRNA not removed) using reverse transcriptase and amplified and barcoded using ScriptSeq™ Index PCR Primers (Epicenter) to generate single-indexed libraries. cDNA libraries were pooled and sequenced on an Illumina MiSeq using a v2-500 cycle (paired end 250x250 bp) kit.

SAG assembly, quality control, and sequence analysis

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

Full-length (>1500 bp) 16S rRNA gene sequences were detected on 8 SAGs. These sequences were imported into the ARB environment (Ludwig *et al.*, 2004) and placed within the ARB

backbone tree using the parsimony tool. Brocadiales-associated 16S rRNA genes from Woebken *et al.* (2008) and Galan *et al.* (2009) were imported for comparative purposes to assign SAG 16S rRNA genes to previously reported sub-clades of marine *Ca*. Scalindua. Additional sequences from Schmid *et al*. (2003), representing species-level *Candidatus* Scalindua designations, were also included as outgroups to the 'Arabian Sea' sequence cluster (see Results below). Sequence alignments were created using the automated aligner, then manually curated when needed. To assess the 16S rRNA gene phylogeny using only informative positions, a mask was created based on the curated alignment, and used for construction of Neighbor-Joining (with Feldstein correction), Maximum Likelihood (with LG substitution model), and Parsimony trees with 1000 bootstraps for all models.

All SAG-associated assemblies generated from MDA products were analyzed using Prokka (Seemann, 2014). The 'faa' files from this pipeline were used as queries for BLASTP searches against public and custom databases (described below). For visualization of gene order and synteny, contigs with features of interest were extracted from the 'gbk' files from Prokka. Contigs of interest were then imported into 'EasyFig' and compared to one another using BLASTN. The associated output figures were manually curated in Adobe Illustrator. The package T-REKs (Jorda and Kajava, 2009) was used to identify tandem repeats on contigs of interest.

Predicted amino acid sequences from 6 SAGs were used to create a composite SAG database for comparison against public databases using BLASTP and for use as a reference database for BLASTX-based analyses of metagenomes and metatranscriptomes (described below). Our goal

272 in creating this database was to capture the majority of functional gene content across the SAGs
273 (related to one another at roughly the species level; see Results), rather than to resolve
274 population-level variation among the SAGs. The 6 SAGs were chosen because they had
275 relatively high completeness (28.2-50.0%), minimal contamination (<5%), and full-length 16S
276 rRNA gene sequences. SAG sequences were clustered (using UCLUST and USEARCH
277 commands) at 50% amino acid identity (AAI). Clustering yielded 2703 proteins, including 49 of
278 the 54 universally conserved single-copy ribosomal proteins (Yutin *et al.*, 2012). This result
279 suggested that the composite database was ~91% complete (based only on universal marker gene
280 recovery) and that a complete composite database would contain ~3000 protein-coding genes
281 (50% AAI clusters), roughly consistent with genome size estimates based on individual SAGs
282 (extrapolated using estimated genome completeness and counts of detected genes per SAG;
283 Table 1).
284
285 A custom database of protein sequences from anammox taxa was created for comparison to the
286 SAG gene set. Anammox-associated genomes were identified in NCBI using the search term
287 'Brocadiales'. All amino acid sequences (n=50,272) from these genomes were downloaded and
288 combined with amino acid sequences (n=4330) from *Ca*. Scalindua profunda (obtained
289 according to (van de Vossenberg *et al.*, 2013)), yielding a database of 54,602 proteins. SAG
290 proteins were then queried against this database via BLASTP. A sequence was considered
291 unique to the SAG set if this query did not return a significant match (bit score > 50) to a
292 database sequence. Using this method, all urea-associated proteins (urease and accessory
293 proteins) were identified as unique to the ETNP SAG set, along with 1803 other genes (1811

294 total unique). Of the other genes unique to ETNP SAGs, 1604 encoded hypothetical proteins.
295 Non-hypothetical, unique protein-coding genes are listed in Table S2.
296
297 To further assess relatedness among SAGs, average nucleotide identity (ANI) and average amino
298 acid identity (AAI) between each pair of SAGs was calculated using the ani.rb and aai.rb scripts
299 from the enve-omics toolkit (Rodriguez-R and Konstantinidis, 2016) with ani.rb cutoffs of 700
300 bp minimum alignment length and 70% minimum identity, and aai.rb cutoffs of bit score > 50
301 and 90% minimum alignment length (as a fraction of the shorter sequence). The
302 get_homologues package (Contreras-Moreira and Vinuesa, 2013) was used to identify genes
303 shared between each SAG and protein-coding sequences in the *Ca*. Scalindua profunda genome.
304
305 *Metatranscriptome analysis*
306 Metatranscriptomic reads were trimmed and merged as above for SAG sequences. Merged reads
307 were compared to the SILVA rRNA database using BLASTN, and sequences with significant
308 matches to rRNA genes were identified and removed. Non-rRNA reads were queried
309 (BLASTX) against the composite database (see above). BLAST output was parsed to identify
310 transcripts recruiting to SAGs with bit score >50 and AAI >95%. Counts of mapped transcripts
311 per gene were normalized by gene length and sequencing depth, with final counts expressed as
312 kilobase pairs of mapped transcripts per Megabase pairs sequenced.
313
314 *Screening of AMZ metagenomes and metatranscriptomes*
315 Publically available metagenomes and metatranscriptomes from the ETSP AMZ and
316 metagenomes from the Arabian Sea and Saanich Inlet AMZs were screened for close homologs

of SAG *ure* and *cyn* genes (Table 2). Sequences first were compared to the SILVA rRNA database using BLASTN, and sequences identified as rRNA genes/transcripts were removed. Remaining reads were compared using BLASTX against a database of the urease and cyanase-encoding genes recovered from the ETNP SAGs, using match thresholds of bit score >50 and AAI >95%. To test for the presence of SAG urease genes in the ETNP community, we screened a deeply sequenced (HiSeq) metagenome from 200 meters at station 6T against a 39,476 bp *ure*-containing SAG contig using BLASTN. This metagenome was trimmed and cleaned as in (Padilla *et al.*, 2017) and binned using MetaBat (Kang *et al.*, 2015). Genes recovered on the assembled contigs of the aforementioned metagenome were also compared by BLASTX against a custom database containing the SAG urease genes and 2870 *ure*C genes available in NCBI's protein database (as of 01-03-2018), using the match parameters described above. The taxonomic identities of top matching database entries were used to estimate the taxonomic richness of ureases in the OMZ community. We did not screen for *Ca*. Scalindua-like *cynS* genes, as these have been previously reported in Eastern Pacific AMZ meta-omic datasets (Babbin *et al*. 2017).

All sequence data generated in this study are in NCBI under BioProject PRJNA407229.

Results and Discussion

Hydrographic conditions and anammox activity in the ETNP AMZ

Figure 1 shows data from the three most extensively sampled ETNP stations, including two near-shore stations (6T, 7T) and a station (3T) farther offshore; data from three additional stations are in Figure S2, with station coordinates in Table S1. At all stations, anoxic conditions were

observed from depths of ~70-100 meters. Nitrite concentration was near the detection limit in the surface layer, but increased directly below the oxic-anoxic interface to 3-5 μM, a characteristic feature of AMZs. Ammonium concentrations were generally in the low nanomolar range (20-50 nM) (Figures 1, S2). Neither cyanate nor urea was measured on the 2014 cruise. However, in samples from a 2017 cruise along the same transect, urea concentration was below the detection limit within the AMZ, but above detection in the mixed layer, reaching maxima of ~100 nM (data not shown).

Anammox rates varied over depth and with proximity to shore. Measured rates were highest in the upper AMZ, increasing sharply below the oxic-anoxic interface with near-maximal rates reached only 10–20 m deeper (Figures 1, S2). Rates were lowest and confined to a narrow depth range at station 3T farthest from shore (up to 1 nM N_2 d^{-1}), with rates increasing to 5.5 nM N_2 d^{-1} at the near-shore sites (6T, 7T; Figure 1). This is consistent with analyses in the ETSP AMZ off Chile and Peru, which showed anammox rates to be highest in shelf waters and to diminish with distance from shore, strongly correlating with organic matter export (Kalvelage *et al.*, 2013). Our measured rates are similar to those measured previously at sites close to ours in the ETNP (Babbin *et al.*, 2014, Ganesh *et al.*, 2015) and at other open ocean AMZ sites (Lam and Kuypers, 2011).

Genomic evidence for organic nitrogen utilization in ETNP Ca. *Scalindua*

Ca. Scalindua bacteria in AMZs may contain metabolic features distinct from those of anammox bacteria in other environments. We explored this potential by analyzing 20 SAGs from two anoxic depths at station 6T in the ETNP. All 20 SAGs were classified with high confidence

(Probability score = 100.0, Naïve Bayesian classifier) as *Ca*. Scalindua sp. based on PCR-amplified 16S rRNA gene fragments.

As is common in single-cell analysis (Rinke *et al.*, 2013, Thrash *et al.*, 2014), estimated genome completeness and strain heterogeneity varied considerably among the SAGs (range: 0-50% and 0-100%, respectively; average: 27.4% and 33.0%). However, contamination was consistently low (average: 2.0%) and in some cases non-detectable, with moderate levels (>5%) in only two SAGs (Table 1). Based on genome completeness and total recovered sequence length, estimated genome size averaged 2.6 Mbp (range: 2.0-3.8 Mbp). The SAG with the highest estimated completeness (50%, SAG N22) contained 1637 protein-coding genes, suggesting a total gene count (~3300) smaller than that of *Ca.* Scalindua genomes of other species (>4000 genes). On average, regions homologous between SAGs shared 94.1% ANI (standard deviation: 2.7) and homologous open reading frames shared 85.3% AAI (standard deviation: 3.8) among SAGs and 73% AAI with homologs from *Ca.* Scalindua profunda, a sediment anammox bacterium with a near complete (>90%) genome. Analysis of diverse bacteria shows that strains of the same species generally share >94% ANI (Konstantinidis and Tiedje, 2005), whereas the AAI value observed here falls at the lower end of the estimated species boundary (Rodríguez and Konstantinidis, 2014). Full-length (>1500 bp) 16S rRNA genes were identified in 8 SAGs (1 from 125 m, 7 from 300 m), shared 98-100% ANI, and were identical or nearly identical to the 16S rRNA gene fragments obtained from all SAGs by PCR-based screening (Figure S3). These 8 full-length 16S rRNA sequences clustered in a monophyletic sub-clade of *Ca.* Scalindua referred to as the Arabian Sea cluster and were nearly identical to clones primarily from the AMZ in the Arabian Sea (Woebken *et al.,* 2008), but more distantly related to a cluster of *Ca*. Scalindua

386 sequences from the ETSP AMZ off Peru and Chile (Figure 2). Together, these data suggest high
387 relatedness among the analyzed cells, which cluster within a *Ca*. Scalindua clade distributed
388 widely across diverse AMZs.
389
390 Protein-coding genes in the SAGs, hereafter referred to as ETNP *Ca*. Scalindua, were compared
391 against a database of amino acid sequences from all available anammox-associated genomes.
392 Genes diagnostic of the anammox process encoding hydrazine synthase (HZS), hydrazine
393 oxidase/dehydrogenase (HZO), and cd_1 nitrite:nitric oxide oxidoreductase (NirS) were found in
394 5, 8, and 6 of the 20 SAGs, respectively (Table 1), confirming the metabolic role of these
395 bacteria. Genes encoding octahaem hydroxylamine oxidoreductases (HAO) and ammonium
396 transporters (Amt), both of which are observed in multiple copies in anammox genomes, were
397 identified in 16 and 15 of the SAGs, respectively. Amino acid sequences of HZS, HZO, NirS,
398 HAO, and two of the three Amt proteins displayed highest scoring matches to homologs from
399 other anammox bacteria when queried (BLASTP) against the NCBI nr database. Together, the
400 recovery of multiple genes of anammox central metabolism and the shared ancestry of these
401 genes with other Brocadiales identifies the SAGs as members of the AMZ anammox community.
402
403 Comparative analyses revealed 1811 non-redundant genes (out of 14,610 total (redundant/non-
404 redundant) across the SAGs) that did not have a significant (bit score > 50, BLASTP) match to a
405 protein sequence in the custom database, and therefore may be unique to the ETNP *Ca*.
406 Scalindua group. This "unique" gene set is dominated by uncharacterized hypothetical proteins
407 (1569 of 1811, 86%), consistent with high proportions of uncharacterized lineage-specific genes
408 in other anammox genomes (Speth *et al.*, 2017). A total of 206 non-redundant proteins displayed

409 significant matches to the COG database via BLASTP, while 36 displayed an identifiable protein
410 domain structure but did not display significant similarity to the COG database (Table S2).
411
412 Of the unique classifiable sequences, we focused on those that allowed us to explore the
413 hypothesis of alternative nitrogen substrate use. In contrast to all characterized genomes of
414 anammox bacteria, ETNP *Ca*. Scalindua SAGs contain genes for hydrolysis and transport of urea
415 (Table 1, Figure 3A). In three SAGs (G15, M13, N19), we identified contigs containing *ureC*
416 encoding the alpha subunit of urease, the nickel (Ni)-containing enzyme that facilitates cleavage
417 of urea into ammonia and carbamate (Mobley *et al.*, 1995), with the carbamate then
418 spontaneously forming ammonia and carbon dioxide. The ETNP *Ca*. Scalindua *ureC* encodes
419 conserved catalytic site residues present in enzymatically verified UreC of urease-positive
420 bacteria (Figure S4) and is directly downstream of genes for the non-catalytic gamma and beta
421 urease subunits and directly upstream of genes encoding urease accessory proteins UreEFG
422 required for assembly and activation of the apoprotein (Figure 3). This gene order is nearly
423 identical to that observed in enzymatically verified urease-positive bacteria (e.g., *Proteus*
424 *mirabilis*; (Pearson *et al.*, 2008). Studies using *ure* knockout mutants indicate that UreE is likely
425 the Ni donor, while UreF and UreG are chaperones enabling Ni donation from UreE (Mobley *et*
426 *al.*, 1995). The *ureD* gene, which encodes a fourth subunit whose function is unclear but is
427 required for urease assembly in *P. mirabilis*, was identified downstream of *ureG* on one of the
428 *ureC*-containing contigs, and on separate, smaller contigs in other SAGs. Genes encoding high
429 affinity ABC-type urea transporters (*urtCDE*, Figure 3) are also present on the *ure*-containing
430 contigs of SAGs N19 and M13. The urease-associated genes (*ure* and *urt*) show ≥98% ANI
431 among SAGs, with the vast majority of mutations at the third codon position. The sequences

from one SAG (G15) are nearly identical to those from an assembled metagenome contig from the AMZ core (JGI Scaffold in Figure 3A), confirming the presence of these genes in community data from the site. However, SAG-affiliated *ure* genes (*ureC*) were at low proportional abundance (1 of 42, ~3%) in the total pool of *ure* genes present in the metagenome (Table S3). Overall, the dominant *ureC* variants were most closely related (based on BLASTX) to those of an alphaproteobacterium (*Sphingorhabdus flavimaris*; 24 of 42 *ureC* fragments), suggesting that other organisms in the OMZ may compete with ETNP *Ca*. Scalindua for urea.

UreC of ETNP *Ca.* Scalindua does not display a close phylogenetic affiliation with that of other lithotrophic organisms, including the ammonia-oxidizing Thaumarchaea and nitrite-oxidizing bacteria (e.g., *Nitrospira*). Rather, *Ca*. Scalindua UreC is most closely related (70% AAI) to UreC of a facultatively anaerobic marine *Bacteroidetes* bacterium (*Raineyella antarctica*; Pikuta *et al.*, 2016) (Figures 4A, S4). None of the urease-encoding contigs recovered from the SAGs contain marker genes typically used to assess phylogenies (e.g., 16S rRNA gene). However, the largest of these contigs (contig 1 from SAG M13; Figure 3A) contains a gene encoding the glycolysis protein glyceraldehyde 3-phosphate dehydrogenase (GspA). GspA is conserved among representative genomes from most anammox genera, and phylogenetic analysis placed the M13 GspA in a highly supported clade with that of other Brocadiales, including other *Ca*. Scalindua species (Figure S5). Several other genes on this contig also display highest similarity (BLASTP) to anammox-associated Brocadiales. These include two genes encoding XerC, an enzyme mediating site-specific recombination, a process potentially associated with horizontal gene transfer. Tandem repeat sequences, which are often affiliated with recombination, were identified on this contig in five protein-coding genes, all >1000 bp from the *ure* genes. Taken

455 together, these data link a potential for urea utilization to ETNP *Ca.* Scalindua and, given the
456 absence of these genes from other *Ca*. Scalindua species, raise the possibility that this function
457 was acquired horizontally from a non-anammox organism. The potential for horizontal transfer
458 is supported by the high ANI (≥98%) among *ure* genes from different SAGs, potentially
459 reflecting recent transfer or strong selection pressure.
460
461 The SAG data also support the hypothesis, proposed by Babbin *et al.* (2017), that *Ca*. Scalindua
462 in AMZs can use cyanate as an ammonium source. Five of the SAGs (Table 1) contain the *cynS*
463 gene putatively encoding cyanate hydratase (cyanase). Cyanases cleave cyanate to carbamate
464 (H_2NCOO^-) and carbon dioxide and occur in diverse non-anammox bacterial and eukaryotic
465 lineages (Rocap *et al.*, 2003; Kamennaya *et al.*, 2008; Kamennaya and Post, 2011), as cyanate is
466 a common by-product of urea degradation and amino acid metabolism. The SAG *cynS* sequences
467 share 94.5-98.9% ANI and, in two of the SAGs, are present on >9 kbp contigs with conserved
468 synteny (Figure 3B). The SAG *cynS* is most closely affiliated (82% AAI) with that of the only
469 other *cynS* sequence linked to an anammox bacterium, *Ca*. Scalindua profunda from sediment.
470 Both *Ca*. Scalindua CynS sequences cluster in a monophyletic clade with those of aerobic nitrite-
471 oxidizing bacteria (*Nitrospina*) common in the oxycline of AMZ regions (Figure 4B) (Zaikova *et*
472 *al.*, 2010, Fussel *et al.*, 2012). These SAG results link CynS to *Ca*. Scalindua in AMZs,
473 supporting the work of Babbin *et al.* (2017) showing that cyanate stimulates AMZ anammox.
474
475 *Transcription of alternative nitrogen acquisition pathways by ETNP* Ca*. Scalindua*
476 Metatranscriptomics confirmed the activity of key genes of *Ca.* Scalindua nitrogen-based energy
477 metabolism (Tables S4-S6). We first estimated ETNP *Ca.* Scalindua's contribution to

478 community transcription by querying metatranscriptome datasets from 5 ETNP sites and
479 multiple depths against a composite SAG amino acid database using a match threshold of >95%
480 AAI (above the average between-SAG AAI of 85%). The composite database contained
481 sequences from 6 of the most complete SAGs, representing an estimated ~90% of all homologs
482 shared among the SAG set. This analysis showed that the representation of ETNP *Ca.* Scalindua
483 transcripts increased dramatically from the base of the oxycline into the AMZ (Figures
484 1B,1E,1H, S2), likely due to an increase in the number of *Ca*. Scalindua bacteria along this
485 gradient (Ganesh *et al.,* 2015). This is consistent with the increase in anammox rates into the
486 AMZ, although the depth of highest transcript representation was below that of highest rates and
487 below the nitrite maximum (Figures 1C-D).
488
489 Transcripts encoding the putative urease and cyanase of ETNP *Ca*. Scalindua were detected
490 throughout the study area, increasing in representation with depth in a pattern roughly paralleling
491 that of the total ETNP *Ca*. Scalindua transcript pool (Figures 1C,1F,1I, S2). An analysis of
492 metatranscriptome data from the AMZ core (200 m) at station 6 indicated that over half of all
493 detected *ureC* transcripts in the AMZ community were most similar to those recovered on ETNP
494 *Ca*. Scalindua SAGs (Table S6). Transcripts encoding ABC-type urea transporters (Urt) were
495 also detected (Tables S4-S5), peaking in proportional representation at 200 meters at Station 6
496 where urease transcripts were also most abundant (data not shown). In general, urease
497 transcripts were less common than cyanase transcripts. Indeed, cyanases were among the top 40
498 most transcribed *Ca*. Scalindua genes at stations with high anammox rates (e.g., station 6T,
499 Figure S6). However, both urease and cyanase transcripts were far less abundant than those
500 encoding genes diagnostic of anammox, mainly HZS and HZO, which were commonly among

501 the top 5 most transcribed ETNP *Ca*. Scalindua proteins (Figures S6, S7). CO_2 fixation in
502 anammox bacteria occurs through the Wood-Ljungdahl pathway, with electrons donated from
503 the oxidation of nitrite to nitrate. Genes diagnostic of this pathway and of nitrite oxidation,
504 notably *acsA* encoding acetyl coA synthase and *narG* encoding nitrate/nitrite oxidoreductase
505 respectively, were consistently observed among the *Ca*. Scalindua transcripts, albeit at low levels
506 (Figure S7).
507
508 The transcript data also provide insight into the importance of other nitrogen-containing
509 compounds in AMZ anammox. Nitric oxide (NO) is a key intermediate in anammox, having
510 been shown in *Ca.* Kuenenia stuttgartiensis to be condensed with ammonium to form hydrazine
511 (Kartal *et al.,* 2011). Under non-limiting nitrogen conditions in batch reactors, NirS-type
512 cytochrome cd-1 containing nitrite reductase is implicated as the major route to NO and is
513 among the most highly expressed proteins. However, while we identified NirS homologs in the
514 SAGs, NirS transcripts were far less abundant than those encoding other anammox proteins, such
515 as HZS and HZO (Figures 5, S7). A similar result was observed in a prior study that used the *Ca.*
516 S. profunda genome to recruit metatranscriptome data from the ETSP AMZ (van de Vossenberg
517 *et al.*, 2013). Interestingly, biochemical investigations of octahaem cytochrome c-containing
518 hydroxylamine/hydrazine oxidoreductase (OCC-HAO) proteins from *Ca.* Kuenenia
519 stuttgartiensis revealed that one protein (kustc1061) produced NO through the oxidation of
520 hydroxylamine (Maalcke *et al.*, 2014). This protein belongs to a subclade of OCC-HAO proteins
521 named 'HZO cluster 2a' and is phylogenetically distinguishable from other OCC-HAO/HZO
522 homologs (Schmid *et al.*, 2008). ETNP SAGs also contain a homolog belonging to HZO cluster
523 2a (B14_Prokka_00643 in Figure 5A) and transcripts encoding this protein were consistently

more abundant (up to 10-50 times) than those encoding NirS (Figure 5C). A second HAO homolog, belonging to 'cluster 3', has also been predicted to produce NO from nitrite, although this function is not experimentally verified. This homolog (B14_00071 in Figure 5) was detected in the SAGs and transcribed at a level comparable to that of 'HZO cluster 2a' (Figure 5C). Interestingly, recent work on an anammox bacterium from activated sludge (*Ca.* Brocadia sinica) demonstrated that in the absence of canonical enzymes of NO production (*nirS*, *nirK*), nitrite was reduced to hydroxylamine, potentially by an OCC-HAO protein (although this remains untested), and the resulting hydroxylamine was coupled with ammonium for hydrazine and ultimately N_2 production (Oshiki *et al.*, 2016). Our results raise the possibility that hydroxylamine is also a critical intermediate in anammox bacteria from open ocean AMZs. However, the exact role of this substrate and that of associated OCC-HAO proteins in both NO-dependent and independent pathways of ETNP *Ca*. Scalindua remains speculative.

Evidence for alternative nitrogen use pathways by anammox bacteria in other AMZs

We screened other AMZ datasets to determine if the potential for urea and cyanase use by anammox bacteria is widespread (Table 2). All screened AMZ datasets contain high numbers of sequences matching (>95% AAI) genes in our composite SAG database (data not shown). Of these, sequences closely related to the urease and cyanase-encoding genes from ETNP *Ca*. Scalindua were identified in metagenomes and metatranscriptomes from the ETSP AMZ off Chile. These genes were not detected in the ETSP oxycline (Table 2), consistent with the low abundance of *Ca*. Scalindua at non-AMZ depths in this region (Stewart *et al.,* 2012). ETNP *Ca*. Scalindua-like urease genes were also detected in a metagenome from the core of the Arabian Sea AMZ (Luke *et al.*, 2016), although cyanase genes were not found in this dataset. Close

547 relatives of urease and cyanase genes were not detected in a metatranscriptome from a seasonally
548 anoxic coastal AMZ in Saanich Inlet. These results suggest that the potential for urea and
549 cyanate use for anammox exists in some AMZs, including the major AMZs of the ETSP and
550 Arabian Sea where high anammox rates have been recorded (Galan *et al.*, 2009, Ward *et al.*,
551 2009, Bulow *et al.*, 2010).. Also, the failure to detect these genes in some sites (e.g., Saanich
552 Inlet) is not definitive evidence of their absence, as several factors may preclude detection,
553 including the depth of sequence coverage, the proportional abundance of anammox-cells,
554 variability in the timing and depths of sample collection, and potential sequence divergence
555 across systems. Further genomic analyses of anammox populations across systems, and perhaps
556 at finer spatial and temporal scales of resolution, will help identify the evolutionary and
557 environmental controls determining the distribution of urease and cyanase genes in marine
558 anammox bacteria.
559
560
561 *Conclusions*
562 This study provides evidence that anammox bacteria have the potential to degrade both urea and
563 cyanate. At this time, however, it is not possible to determine the exact biochemical role that
564 ureases and cyanases play in ETNP *Ca*. Scalindua. It is possible that the ammonium liberated by
565 these enzymes is used as an energy substrate for anammox, and therefore contributes to nitrogen
566 loss through N_2 production. Alternatively, it is possible that these enzymes serve other
567 functions. Urea and cyanate are common products of protein degradation, and high intracellular
568 concentrations of these substrates may be detrimental to cellular processes. Ureases and
569 cyanases may therefore serve in detoxification, or potentially to cycle ammonium into anabolic

570 pathways. Thus, these enzymes may play a role in the biological retention or recycling, rather
571 than loss, of valuable nitrogen. However, the detection and transcription of urea transporter
572 genes indicates that ETNP *Ca*. Scalindua likely consumes this organic substrate from the
573 environment, suggesting that urease activity in this organism is not linked exclusively to nitrogen
574 recycling. Furthermore, even recycling will decrease the assimilatory demand for exogenous
575 ammonium and thereby indirectly increase its availability for dinitrogen production.
576
577 The consumption of organic nitrogen by aerobic ammonia-oxidizing microorganisms has gained
578 much attention, notably as urease potential occurs in ubiquitous ammonia-oxidizing
579 Thaumarchaeota (Alonso-Saez *et al.*, 2012). Aerobic nitrite-oxidizing bacteria of the genus
580 *Nitrospira* have also been shown to produce ammonia from urea, thereby sustaining co-occurring
581 ammonia-oxidizers that provide nitrite to *Nitrospira* (Koch *et al*., 2015). However, the
582 contribution of urea to anaerobic pathways of lithotrophic ammonium consumption remains
583 unclear. Indeed, these results are the first report of urease genes in any anammox-capable
584 lineage. Furthermore, while cyanase genes had been reported in anammox bacteria from non-
585 AMZ environments and cyanate shown to support anammox activity in AMZs (Babbin *et al.*,
586 2017), cyanase genes had not yet been definitively linked to genomes of AMZ anammox
587 bacteria, for example by being found co-localized on a metagenome contig with definitive *Ca*.
588 Scalindua signatures, or in a genome from a *Ca*. Scalindua cell/culture. The extent to which
589 ammonium limitation in AMZs selects for organic nitrogen consumption by anammox bacteria
590 remains uncertain, although our results suggest that urea and cyanate use may occur in
591 populations from different AMZs.
592

Overall, the results expand our knowledge about the metabolic capacity of anammox bacteria and predict mechanisms by which these widespread organisms might supplement direct consumption of free ammonium. Few urea and cyanate measurements have been made for AMZs and data from other regions suggest that concentrations rarely exceed nanomolar levels in the open ocean (Solomon *et al.*, 2010; Widner *et al*., 2017). While urea and cyanate concentrations were not comprehensively surveyed in this study, our preliminary measurements, coupled with previous measurements of cyanate in Eastern Pacific AMZs (Widner *et al.*, 2016, 2017), suggest levels comparable (urea) or likely lower (cyanate) than those of ammonium. However, it is possible that turnover of these substrates is rapid, particularly at times or depths of lower ammonium levels. Counts of cyanase transcripts in this study were higher than those for urease, suggesting a potential greater role for cyanate as an alternative ammonium source. We caution, however, that the transcript data provide no temporal resolution and may be a poor proxy for actual substrate turnover. Indeed, our knowledge of the temporal variability in AMZ inorganic and organic substrate availability, and in rates of coupled microbial metabolisms, remains limited. Our data suggest only that the potential for direct cyanate and urease use exists for AMZ anammox bacteria. Future experiments should assess the environmental conditions that constrain the use of different ammonium sources. Such experiments should also determine what proportion of the urea and cyanate pools consumed by anammox bacteria are indeed lost through anammox, versus lost through detoxification or incorporation into new biomass. Urea and cyanate have gathered increasing attention within the context of aerobic nitrification, and recent data have indicated that marine populations of aerobic nitrifiers can oxidize ammonia at vanishingly low oxygen levels (nM) that are well within the range under which anammox occurs

(Bristow *et al.*, 2016). Thus, an important question becomes to what extent periodic oxygenation of the anoxic core, and shoaling of the oxycline, change the dynamics of thaumarchaeal ammonium consumption and anammox, and the extent that use of alternative substrates by either group is stimulated or inhibited by these events. Resolving such questions may improve models estimating the role of diverse nitrogen consumption pathways in bulk nitrogen and carbon budgets under AMZ expansion, providing refinements for global marine nutrient cycling.

Acknowledgments

This work was supported by the National Science Foundation (1151698, 1558916, 1564559 to FJS and 1416673 to KTK), a European Research Council Advanced grant (OXYGEN, 267233 to BT), the Danish National Research Foundation (DNRF53 to BT), and a Community Science Program grant from the U.S. Department of Energy (to FJS and KTK). The work conducted by the DOE Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231. We are grateful to Neha Sarode for help in sequencing analysis, Philipp Hach for help with sample collection, and the captain and crew of the *R/V New Horizon* for enabling sample collection.

The authors declare no financial, commercial, or personal conflict of interest involving the publication of this work.

Supplementary Information

Supplementary information is available at *The ISME Journal* website.

References

Alonso-Saez L, Waller AS, Mende DR, Bakker K, Farnelid H, Yager PL *et al.* (2012). Role for urea in nitrification by polar marine Archaea. *P Natl Acad Sci USA* **109**: 17989-17994.

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

Babbin AR, Peters BD, Mordy CW, Widner B, Casciotti KL, Ward BB (2017) Multiple metabolisms constrain the anaerobic nitrite budget in the Eastern Tropical South Pacific. *Global Biogeochem Cy* **31**: 258-271.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. **19:** 455-77.

Bristow LA, Dalsgaard T, Tiano L, Mills DB, Bertagnolli AD, Wright JJ, et al. (2016). Ammonium and nitrite oxidation at nanomolar oxygen concentrations in oxygen minimum zone waters. *Proc Natl Acad Sci USA* **113**: 10601-10606.

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 Res Pt I* **57**: 384-393.

Burton SAQ, Prosser JI. (2001). Autotrophic ammonium oxidation at low pH through urea hydrolysis. *Appl Environ Microb* **67**: 2952-2957.

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

Contreras-Moreira B, Vinuesa P. (2013). GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol* **79**: 7696-7701.

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

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

Fussel J, Lam P, Lavik G, Jensen MM, Holtappels M, Gunter M *et al.* (2012). Nitrite oxidation in the Namibian oxygen minimum zone. *ISME J* **6**: 1200-1209.

Galan A, Molina V, Thamdrup B, Woebken D, Lavik G, Kuypers MMM *et al.* (2009). Anammox bacteria and the anaerobic oxidation of ammonium in the oxygen minimum zone off northern Chile. *Deep-Sea Res Pt Ii* **56**: 1125-1135.

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

Garcia-Robledo E, Padilla CC, Aldunate M, Stewart FJ, Ulloa O, Paulmier A, et al. (2017). Cryptic oxygen cycling in anoxic marine zones. *Proc Natl Acad Sci USA* **114**: 8319-8324.

679 Grasshoff K, Ehrhardt M, Kremling K, Almgren T. (1983). *Methods of seawater analysis*.
680 Verlag Chemie, Weinheim.

681 Hallam SJ, Mincer TJ, Schleper C, Preston CM, Roberts K, Richardson PM *et al.* (2006).
682 Pathways of carbon assimilation and ammonium oxidation suggested by environmental genomic
683 analyses of marine Crenarchaeota. *Plos Biol* **4**: 520-536.

684 Holmes RM, Aminot A, Kerouel R, Hooker BA, Peterson BJ. (1999). A simple and precise
685 method for measuring ammonium in marine and freshwater ecosystems. *Can J Fish Aquat Sci*
686 **56**: 1801-1808.

687 Jorda J, Kajava AV. (2009). T-REKS: identification of Tandem REpeats in sequences with a K-
688 meanS based algorithm. *Bioinformatics* **25**: 2632-2638.

689 Kalvelage T, Lavik G, Lam P, Contreras S, Arteaga L, Loscher CR, *et al.* (2013). Nitrogen
690 cycling driven by organic matter export in the South Pacific oxygen minimum zone. *Nat Geosci*
691 **6**: 228-234.

692 Kamennaya NA, Chernihovsky M, Post AF. (2008). The cyanate utilization capacity of marine
693 unicellular cyanobacteria. *Limnol Oceanogr* **53:** 2485-2494.

694 Kamennaya NA, Post AF. (2011). Characterization of cyanate metabolism in marine
695 *Synechococcus* and *Prochlorococcus* spp. *Appl Environ Microbiol* **77**: 291-301.

696 Kang DWD, Froula J, Egan R, Wang Z. (2015). MetaBAT, an efficient tool for accurately
697 reconstructing single genomes from complex microbial communities. *Peerj* **3**: e1165.

698 Kartal B, Keltjens JT. (2016) Anammox biochemistry: a tale of heme c proteins. *Trends Biochem*
699 *Sci* **41**: 998-1011.

700 Kartal B, Maalcke WJ, de Almeida NM, Cirpus I, Gloerich J, Geerts W *et al.* (2011). Molecular
701 mechanism of anaerobic ammonium oxidation. *Nature* **479**: 127-130.

702 Koch H, Lucker S, Albertsen M, Kitzinger K, Herbold C, Spieck E, *et al.* (2015). Expanded
703 metabolic versatility of ubiquitous nitrite-oxidizing bacteria from the genus Nitrospira. *P Natl*
704 *Acad Sci USA* **112**: 11371-11376.

705 Konieczna I, Zarnowiec P, Kwinkowski M, Kolesinska B, Fraczyk J, Kaminski Z *et al.* (2012).
706 Bacterial urease and its role in long-lasting human diseases. *Curr Protein Pept Sc* **13**: 789-806.

707 Konstantinidis KT, Tiedje JM. (2005). Genomic insights that advance the species definition for
708 prokaryotes. *Proc Natl Acad Sci U S A* **102:** 2567-2572.

709 Kuypers MM, Lavik G, Woebken D, Schmid M, Fuchs BM, Amann R *et al.* (2005). Massive
710 nitrogen loss from the Benguela upwelling system through anaerobic ammonium oxidation. *Proc*
711 *Natl Acad Sci U S A.* **102:** 6478-6483.

712 Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. (2007). RNAmmer:
713 consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res **35**: 3100-3108.

714 Lam P, Lavik G, Jensen MM, van de Vossenberg J, Schmid M, Woebken D *et al.* (2009).
715 Revising the nitrogen cycle in the Peruvian oxygen minimum zone. *Proc Natl Acad Sci USA*
716 **106**: 4752-4757.

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

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363-1371.

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

Luke C, Speth DR, Kox MAR, Villanueva L, Jetten MSM. (2016). Metagenomic analysis of nitrogen and methane cycling in the Arabian Sea oxygen minimum zone. *Peerj* **4**: e1924.

Maalcke WJ, Dietl A, Marritt SJ, Butt JN, Jetten MSM, Keltjens JT *et al*. (2014). Structural basis of biological NO generation by octaheme oxidoreductases. *J Biol Chem* **289**: 1228-1242.

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

Marsh KL, Sims GK, Mulvaney RL. (2005). Availability of urea to autotrophic ammonium-oxidizing bacteria as related to the fate of C-14- and N-15-labeled urea added to soil. *Biol Fert Soils* **42**: 137-145.

Mobley HLT, Island MD, Hausinger RP. (1995). Molecular biology of microbial ureases. *Microbiol Rev* **59**: 451-480.

Mulvenna PF, Savidge G. (1992). A modified manual method for the determination of urea in seawater using diacetylmonoxime reagent. *Estuar Coast Shelf S* **34**: 429-438.

Nicholls JC, Davies CA, Trimmer M. (2007). High-resolution profiles and nitrogen isotope tracing reveal a dominant source of nitrous oxide and multiple pathways of nitrogen gas formation in the central Arabian Sea. *Limnol Oceanogr* **52**: 156-168.

Oshiki M, Ali M, Shinyako-Hata K, Satoh H, and Okabe S. (2016). Hydroxylamine-dependent anaerobic ammonium oxidation (anammox) by "Candidatus Brocadia sinica". *Environ Microbiol* **18**: 3133-3143.

Padilla CC, Bristow LA, Sarode N, Garcia-Robledo E, Gómez Ramírez E, Benson CR *et al*. (2016). NC10 bacteria in marine oxygen minimum zones. *ISME J* **10**: 2067-2071.

Padilla CC, Bertagnolli A, Bristow LA, Sarode N, Glass J, Thamdrup B *et al*. (2017). Metagenomic binning recovers a transcriptionally active Gammaproteobacterium linking methanotrophy to partial denitrification in an anoxic oxygen minimum zone. *Front Mar Sci* **4**: 23.

Palatinszky M, Herbold C, Jehmlich N, Pogoda M, Han P, von Bergen M *et al.* (2015) Cyanate as an energy source for nitrifiers. *Nature* **524**: 105-108.

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043-1055.

Pearson MM, Sebaihia M, Churcher C, Quail MA, Seshasayee AS, Luscombe NM *et al.* (2008). Complete genome sequence of uropathogenic *Proteus mirabilis*, a master of both adherence and motility. *J Bacteriol* **190**: 4027-4037.

Pikuta EV, Menes RJ, Bruce AM, Lyu Z, Patel NB, Liu Y *et al.* (2016). *Raineyella antarctica* gen. nov., sp. nov., a psychrotolerant, d-amino-acid-utilizing anaerobe isolated from two geographic locations of the Southern Hemisphere. *Int J Syst Evol Microbiol* **66**: 5529-5536.

Qin W, Amin SA, Martens-Habbena W, Walker CB, Urakawa H, Devol AH *et al.* (2014). Marine ammonium-oxidizing archaeal isolates display obligate mixotrophy and wide ecotypic variation. *P Natl Acad Sci USA* **111**: 12504-12509.

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF *et al.* (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**: 431-437.

Rinke C, Lee J, Nath N, Goudeau D, Thompson B, Poulton N *et al.* (2014). Obtaining genomes from uncultivated environmental microorganisms using FACS–based single-cell genomics. *Nature Protocols* **9**: 1038-1048.

Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA *et al.* (2003). Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation. *Nature* **424**: 1042-1047.

Rodríguez-R LM, Konstantinidis KT. (2014) Bypassing cultivation to identify bacterial species. *Microbe* **9**: 111–118.

Rodriguez-R LM, Konstantinidis KT. (2016) The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ* **4**: e1900v1

Sabine CL, Feely RA, Gruber N, Key RM, Lee K, Bullister JL *et al.* (2004). The oceanic sink for anthropogenic CO_2. *Science* **305**: 367-371.

Schmid M, Walsh K, Webb R, Rijpstra WI, van de Pas-Schoonen K, Verbruggen MJ *et al*. (2003). *Candidatus* "Scalindua brodae", sp. nov., *Candidatus* "Scalindua wagneri", sp. nov., two new species of anaerobic ammonium oxidizing bacteria. *Syst Appl Microbiol* **26**: 529-538.

Schmid MC, Hooper AB, Klotz MG, Woebken D, Lam P, Kuypers MMM *et al*. (2008). Environmental detection of octahaem cytochrome c hydroxylamine/hydrazine oxidoreductase genes of aerobic and anaerobic ammonium-oxidizing bacteria. *Environ Microbiol* **10**: 3140-3149.

Seemann T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068-2069.

Solomon CM, Collier JL, Berg GM, Glibert PM. (2010) Role of urea in microbial metabolism in aquatic systems: a biochemical and molecular review. *Aquat Microb Ecol* **59**: 67–88.

Sonthiphand P, Hall MW, Neufeld JD. (2014). Biogeography of anaerobic ammonium-oxidizing (anammox) bacteria. *Front Microbiol* **5**: 399

Speth DR, Russ L, Kartal B, Op den Camp HJ, Dutilh BE, Jetten MS. (2015). Draft genome sequence of anammox bacterium "*Candidatus* Scalindua brodae," obtained using differential coverage binning of sequencing data from two reactor enrichments. *Genome Announce* **3**: e01415-14.

Speth DR, Lagkouvardos I, Wang Y, Qian PY, Dutilh BE, Jetten MSM. (2017). Draft genome of Scalindua rubra, obtained from the interface above the Discovery Deep Brine in the Red Sea, sheds light on potential salt adaptation strategies in anammox bacteria. *Microbial Ecol* **74**: 1-5.

Stewart FJ, Ulloa O, DeLong EF. (2012). Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environ Microbiol* **14**: 23-40.

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

Thamdrup B, Dalsgaard T, Revsbech NP. (2012). Widespread functional anoxia in the oxygen minimum zone of the Eastern South Pacific. *Deep-Sea Res Pt I* **65**: 36-45.

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

Thrash JC, Temperton B, Swan BK, Landry ZC, Woyke T, DeLong EF *et al*. (2014). Single-cell enabled comparative genomics of a deep ocean SAR11 bathytype. *ISME J* **8**: 1440-1451.

Tiano L, Garcia-Robledo E, Dalsgaard T, Devol AH, Ward BB, Ulloa O *et al.* (2014). Oxygen distribution and aerobic respiration in the north and south eastern tropical Pacific oxygen minimum zones. *Deep-Sea Res Pt I* **94**: 173-183.

Tsementzi D, Wu JY, Deutsch S, Nath S, Rodriguez-R LM, Burns AS *et al.* (2016). SAR11 bacteria linked to ocean anoxia and nitrogen loss. *Nature* **536**: 179-183.

van de Vossenberg J, Woebken D, Maalcke WJ, Wessels HJ, Dutilh BE, Kartal B *et al.* (2013). The metagenome of the marine anammox bacterium *'Candidatus* Scalindua profunda' illustrates the versatility of this globally important nitrogen cycle bacterium. *Environ Microbiol* **15**: 1275-1289.

Villanueva L, Speth 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. *Front Microbiol* **5**: 31.

Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microb* **73**: 5261-5267.

Ward BB, Devol AH, Rich JJ, Chang BX, Bulow SE, Naik H *et al.* (2009). Denitrification as the dominant nitrogen loss process in the Arabian Sea. *Nature* **461**: 78-81.

Widner B, Mulholland MR, Mopper K. (2016). Cyanate distribution and uptake in North Atlantic coastal waters. *Environ Sci Technol Lett* **3**: 297-302.

Widner B, Mordy CW, Mulholland MR. (2017) Cyanate distribution and uptake above and within the Eastern Tropical South Pacific oxygen deficient zone. *Limnol Oceangr*

Woebken D, Lam P, Kuypers MMM, Naqvi SWA, Kartal B, Strous M *et al*. (2008). A microdiversity study of anammox bacteria reveals a novel *Candidatus* Scalindua phylotype in marine oxygen minimum zones. *Environ Microbiol* **10**: 3106-3119.

Yutin N, Puigbo P, Koonin EV, Wolf YI. (2012). Phylogenomics of prokaryotic ribosomal proteins. *PLoS ONE* **7**: e36972.

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

Zehr JP, Ward BB. (2002). Nitrogen cycling in the ocean: new perspectives on processes and paradigms. *Appl Environ Microbiol* **68**: 1015-1024.

839

Table Legends

Tables 1 and 2 are uploaded as Excel documents.

Table 1. ETNP *Ca*. Scalindua single amplified genome (SAG) statistics.

Table 2. Detection of ETNP *Ca.* Scalindua urease and cyanase genes in diverse AMZ datasets.

Figure Legends

Figure 1. Anammox rates and representation of ETNP *Ca.* Scalindua transcripts relative to dissolved oxygen, ammonium, and nitrite concentrations at three ETNP stations: 6 (A-C), 7 (D-F), and 3 (G-I). The first column displays dissolved oxygen (black line, μM), nitrite (red circles and line, μM), and ammonium (green circle and line, nM). An ammonium profile at station 6 showed concentrations consistently above 140 nM, which is inconsistent with all other stations sampled and with AMZ literature to date. We therefore interpret this signal as potential contamination and have excluded these data. The second column displays anammox rates (purple line and circles) and the cumulative contribution of all transcripts recruiting to ETNP *Ca.* Scalindua (orange circle and lines, kbp/Mbp). Purple crosses denote non-significant rates. Transcript representation is calculated as length-corrected kilobase pairs of transcripts mapping (via BLASTX, with bit score > 50 and AAI > 95%) to a composite ETNP *Ca.* Scalindua SAG database, per Megabase pairs sequenced. The third column designates the activity and distribution of cyanate hydratase (*cynS*) and urease (*ureC*) transcripts associated with ETNP *Ca.* Scalindua. For all rows, the y-axis indicates water column depth. SAG samples were collected from 125 and 300 m at station 6 in 2013.

Figure 2. 16S rRNA gene-based phylogenetic placement of 8 ETNP *Ca.* Scalindua-related SAGs from two anoxic depths in the ETNP AMZ. Full length (>1500 bp) 16S rRNA genes were identified in SAGs using RNAmmer, and characterized phylogentically relative to database sequences. The phylogeny was estimated using maximum likelihood in ARB with boot strap values based on Neighbor joining, maximum likelihood, and maximum parsimony using 1000, 100, and 1000 boot-strap re-samplings, respectively. Maximum likelihood support values are based on approximate likelihood Bayesian ratios ('abayes').

Figure 3. Gene order and synteny of putative urease (A) and cyanase (B)-encoding contigs identified in SAGs and a metagenomic assembly (JGI Scaffold). Grey shading indicates nucleotide similarity based on BLASTN using the default settings in EasyFig.

Figure 4. Maximum likelihood-based phylogeny of the *ureC* gene encoding the urease alpha subunit (A) and the *cynS* gene encoding cyanase (B). Representative sequences recovered from ETNP *Ca*. Scalindua SAGs are highlighted in red, relative to homologs identified as best matches in BLASTP queries of the SAG sequences against the NCBI nr database (black). For both trees, support values are based on approximate likelihood Bayesian ratios ('abayes'), and trees were constructed using the maximum likelihood method.

Figure 5. Phylogeny and transcription of genes potentially mediating nitric oxide formation in anammox. (A) Octahem cytochrome c hydroxylamine/hydrazine oxidoreductase (HAO/HZO) phylogeny. Clade nomenclatures are based on a previous phylogenetic assessment of HAO/HZO proteins (Schmid et al. 2008). Clades with an asterisk indicate subgroups that are new based on the current work. (B) Cytochrome cd-1 containing nitrite reductase (NirS) phylogeny. Both trees were constructed using maximum likelihood, with support values based on approximate

likelihood Bayesian ratios ('abayes'). Proportional abundance of transcripts encoding the hypothesized hydrazine-oxidizing HZO (Panels 1, 4, 7), the hypothesized nitrite-reducing Hao (Panels 2, 5, 8), and the hydroxylamine oxidizing/NO forming HAO and cytochrome cd-1 containing NirS (panels 3, 6, and 9) at ETNP stations 6 (top row), 7 (middle row), and 3 (bottom row). All scales display activity in kbp/Mbp sequenced.

Supplementary Table and Figure Legends

Table S1. ETNP stations and depths sampled for metatranscriptome sequencing.

Table S2. Annotated, non-hypothetical proteins detected in ETNP *Ca*. Scalindua SAGs but not in any other available anammox-associated genomes (i.e., unique proteins). The table only shows unique proteins with an assigned annotation (based on similarity to homologs in the COG database).

Table S3. Taxonomic affiliation of *ureC* gene fragments recovered in AMZ metagenomes from station 6. Taxonomy is estimated by the identity of top BLASTX matches in a composite database of *ureC* genes from NCBI-nr and ETNP *Ca.* Scalindua SAGS.

Table S4. Length- and sequencing depth-normalized transcript distributions for genes associated with ETNP *Ca.* Scalindua.

Table S5. Annotations of genes associated with ETNP *Ca.* Scalindua (same genes as in Table S4).

Table S6. Taxonomic affiliation of *ureC* gene transcript fragments recovered in an AMZ metatranscriptome from 200 meters depth at station 6. Taxonomy is estimated by the identity of top BLASTX matches in a composite database of *ureC* genes from NCBI-nr and ETNP *Ca.* Scalindua SAGS.

Figure S1. Map of study area, showing locations of stations sampled in 2014 and identified in Figure 1 (main text). Samples for SAG analysis were collected from station 6T in 2013. Exact coordinates can be found in Table S1.

Figure S2. Anammox rates and representation of ETNP *Ca.* Scalindua transcripts relative to dissolved oxygen, ammonium, and nitrite concentrations at three ETNP stations: 8 (A-C), 10 (D-F), and 4 (G-I). The first column displays dissolved oxygen (black line, μM), nitrite (red circles and line, μM), and ammonium (green circle and line, nM). The second column displays anammox rates (purple line and circles) and the cumulative contribution of all transcripts recruiting to ETNP *Ca.* Scalindua (orange circle and lines, kbp/Mbp). Purple crosses denote non-significant rates. Transcript representation is calculated as length-corrected kilobase pairs of transcripts mapping (via BLASTX, with bit score > 50 and AAI > 95%) to a composite ETNP *Ca.* Scalindua SAG database, per Megabase pairs sequenced. The third column designates the activity and distribution of cyanate hydratase (*cynS*) and urease (*ureC*) transcripts associated with ETNP *Ca.* Scalindua. For all rows, the y-axis indicates water column depth.

Figure S3. Phylogenetic approximation of PCR-amplified 16S rRNA genes generated from SAG template DNA (following multiple displacement amplification). Sequences were inserted into the

backbone tree based on Figure 2 using the parsimony tool in ARB, and hence represent a phylogenetic approximation.

Figure S4. Alignment of UreC amino acid sequences from characterized urease-positive organisms *Proteus mirabilis*, *Streptomyces* sp. NRLL and MJM, and from two taxa (*Kouleothrix aurantiaca* and *Raineyella antarctica*) identified as best BLASTP matches to the ETNP *Ca*. Scalindua SAG UreC (UreC from SAG N19_00589 as a representative). Conserved catalytic site histidine and cysteine residues are noted by green and yellow arrows, respectively. The alignment was produced using clustalW and visualized with mView (https://www.ebi.ac.uk).

Figure S5. Phylogenetic approximation of SAG-associated glyceraldehyde 3-phosphate dehydrogenase, GspA (red). Sequences include GspA from a large *ure* and *urt*-containing contig from SAG M13 (see Figure 3A), and GspA from a smaller contig from SAG N22. Purple sequences were identified based on BLASTP against NCBI-nr. The phylogeny was estimated by Maximum likelihood with bootstrap support values based on the approximate Bayes method.

Figure S6. Top 40 most highly transcribed ETNP *Ca.* Scalindua genes observed at Station 6T (200 m, AMZ core).

Figure S7. Proportional abundance of transcripts encoding the hydrazine-producing (hydrazine synthase) and consuming (hydrazine oxidoreductase) enzymes, nitrite/nitrate oxidoreductases likely involved in nitrite oxidation, and acetyl coA synthase involved in the Wood-Ljungdahl pathway at ETNP stations 6 (A, D, G), 7 (B, E, H), and 3 (C, F, I).

Table 1. ETNP *Ca.* Scalindua single-cell amplified genome (SAG) statistics

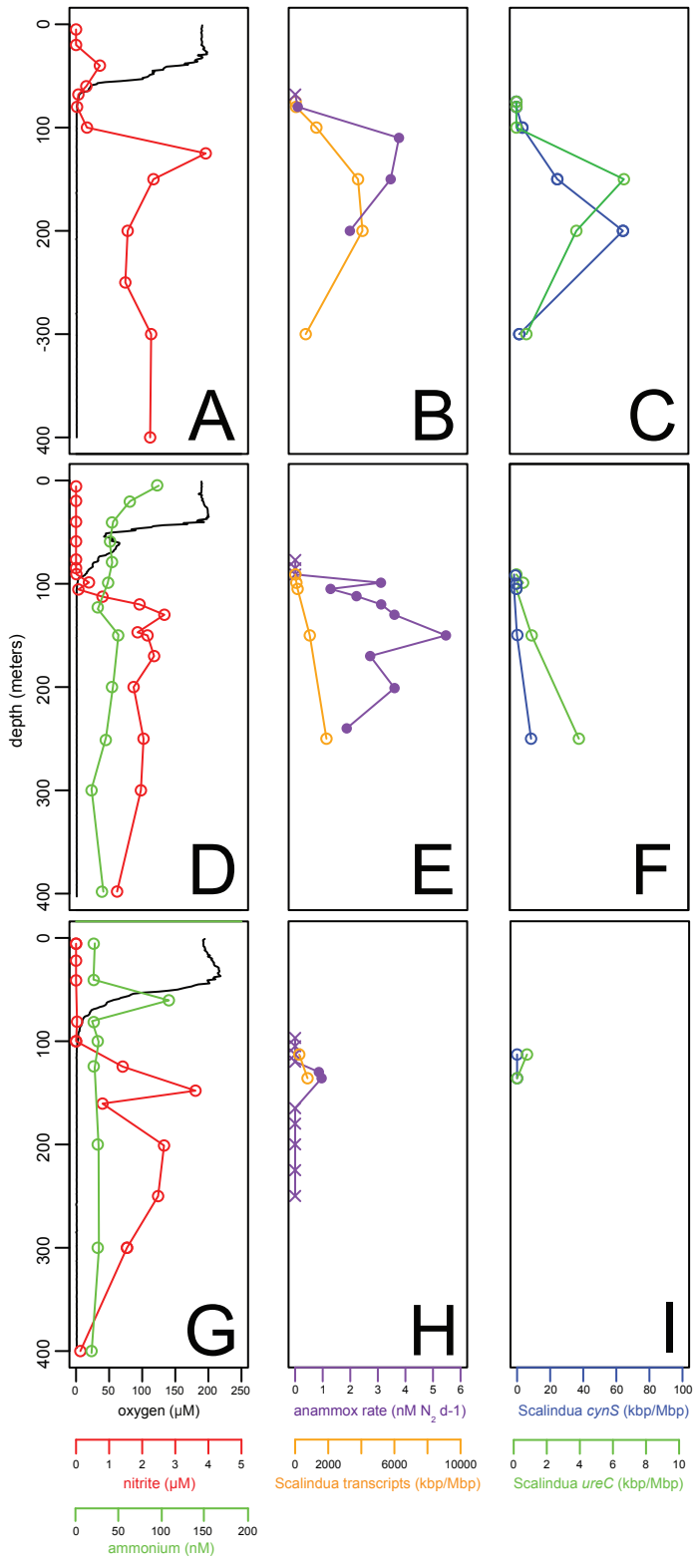
							rRNA genes			Urease and cyanase-associated								Marker proteins - anammox metabolism								
SAG	Comp	Cont	Heter	ORFs	Mbp	Est Gen (Mbp)	[1]5S	[1]16S	[1]23S	[3]UreA	[3]UreB	[3]UreC	[3]UreD	[3]UreE	[3]UreF	[3]UreG	[2]CynS	[2]Octa HZO	[2]HzsA	[2]Octa Hao1	[2]Octa Hao2	[2]Octa Hao3	[2]Octa Hao4	[2]NirS	[2]Amt B1	[3]Amt B2	[2]Amt B3
B14	31.0	3.5	50.0	700	0.74	2.40	-	+	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	+	-
B17	31.6	0.0	0.0	862	0.91	2.88	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	+	-	+	+	-
B21	23.1	1.7	50.0	451	0.45	1.95	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	+	+
C14	12.3	1.8	100.0	745	0.62	NA	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	+	+	-	-	-	-
C8	19.7	0.0	0.0	773	0.51	2.62	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
E14	28.2	0.0	0.0	800	0.62	2.18	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
F8	36.5	0.0	0.0	1101	0.92	2.53	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	+	+	+	+	-
G15	28.6	3.9	40.0	1251	1.04	3.65	-	-	-	+	+	+	-	+	+	+	+	+	+	-	+	-	-	-	6	-	+
H20	43.1	6.7	70.0	1104	0.98	2.28	+	-	+	-	-	-	+	-	+	+	-	+	2	+	+	2	-	-	+	-	+
J11	20.5	1.8	33.3	581	0.43	2.11	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
K21	27.8	1.9	100.0	1230	1.06	3.80	-	-	-	-	-	-	+	-	-	-	+	-	+	-	+	-	+	+	+	+	+
L14	31.0	1.1	0.0	951	0.79	2.56	-	+	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
L20	26.6	5.6	100.0	697	0.58	2.17	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	+	-	-
L7	15.8	1.8	0.0	695	0.58	NA	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-
L9	39.5	4.4	33.3	1321	1.12	2.83	-	-	+	+	+	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-
M13	4.2	0.0	0.0	580	0.48	NA	-	-	-	+	+	+	+	+	+	+	-	+	-	+	-	-	-	-	+	-	+
M7	33.3	2.8	50.0	978	0.79	2.36	-	-	-	-	-	-	-	-	-	-	-	+	+	-	+	-	+	+	-	+	+
N13	0.0	0.0	0.0	234	0.19	NA	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-
N19	46.2	1.1	0.0	1243	1.11	2.41	+	+	+	+	+	+	+	+	+	+	-	+	-	-	-	+	-	-	-	-	+
N22	50.0	2.8	33.3	1637	1.25	2.50	+	+	+	-	-	-	-	-	-	-	+	+	-	-	-	-	-	+	-	+	+

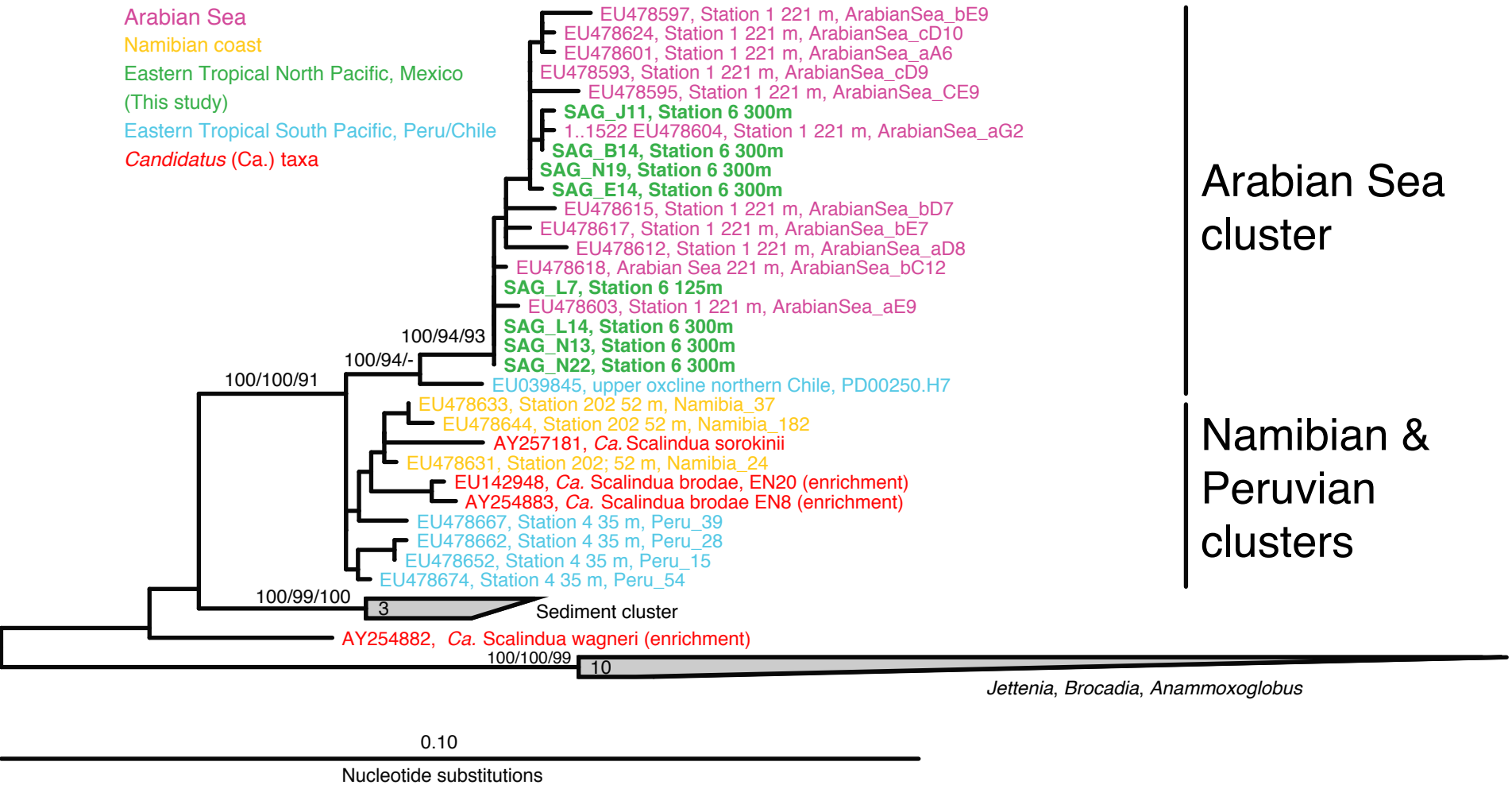
Comp; genome completeness (%) estimated by CheckM
Cont; genome contamination (%) estimated by CheckM
Heter; strain heterogeneity index (range 1-100) estimated by CheckM
ORFs; number of protein-coding open reading frames recovered per SAG
Mbp; Mbp of contigs recovered per SAG
Est Gen; estimated genome size based on estimated completeness and Mbp of SAG; done only for SAGs with completeness >20% and contamination <5%
Ure; urease. UreA; gamma subunit, UreB; beta subunit, ureC; alpha subunit
Cyn; cyanate hydratase
Ure; urease accesory proteins
Octa; Octahaem
Hzo; hydrazine oxidoreductase/dehydrogenase, S. profunda homolog scal03295
Hzs; hydrazine synthase, S. profunda homolog scal01318
Hao; hydroxylamine oxidoreductase. S. produnda Hao1 homolog 00421, S. profunda Hao2 homolog 01317, S. profunda Hao 3 homolog 02116, S. profunda homolog 4 04164
Amt; ammonia transporter. S.profunda Amt1 homolog 00587, S. profunda Amt2 homolog 00591, S. profunda Amt3 homolog 00594
NirS; cytochrome c nitrite reductase. S. profunda nirS homolog 02098
[1]top BLASTN match to Brocadiales-affiliated homolog via queries against the SILVA database
[2]top BLASTP match to Brocadiales-affiliated homolog via queries against the NCBI-nr database + publically available Brocadiales databases
[3]top BLASTP match to non-Brocadiales via queries against the NCBI-nr database (i.e., homologs not detected in other Brocadiales)

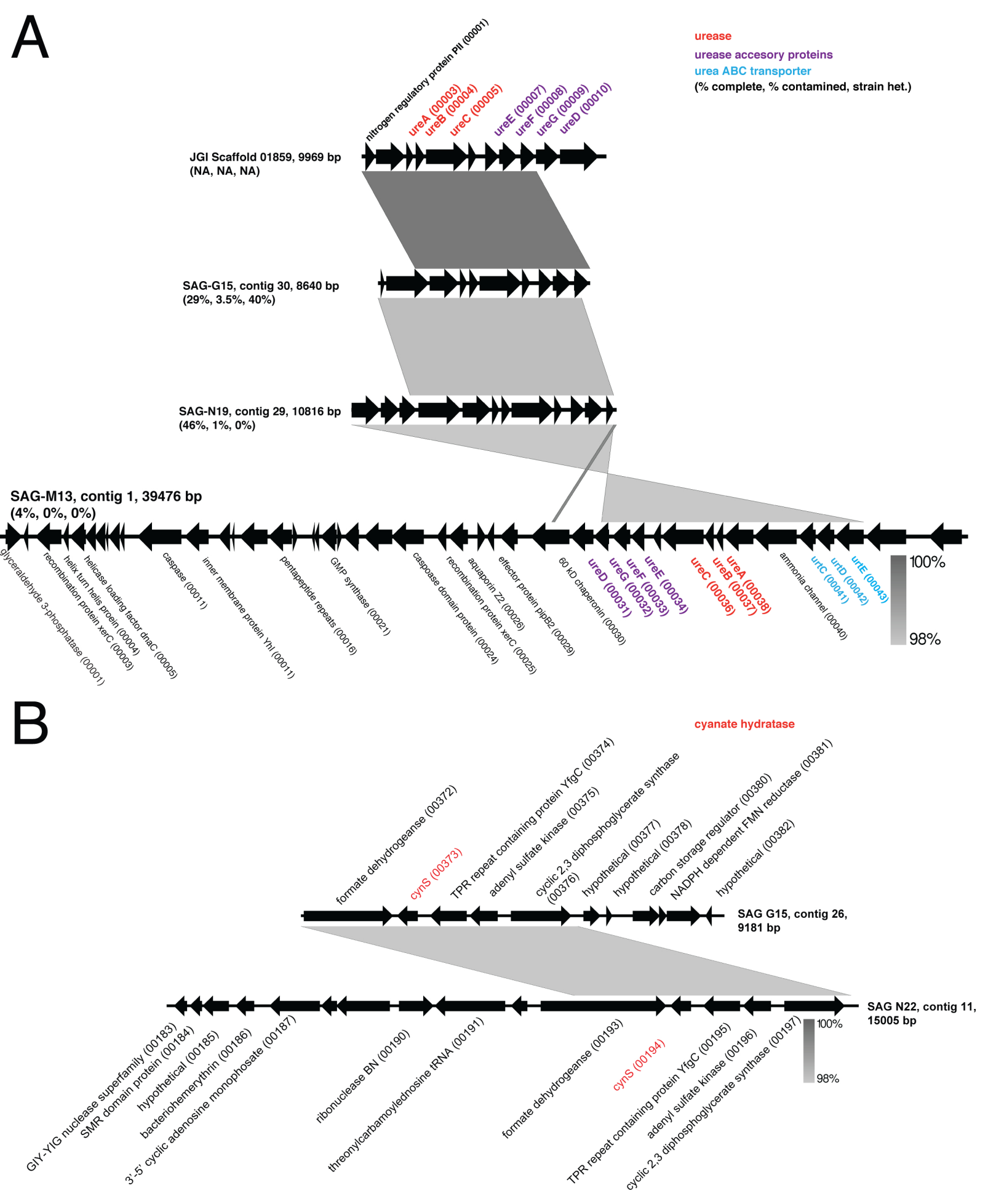
Table 2. Detection of ETNP *Ca.* Scalindua urease and cyanase genes in diverse AMZ datasets.

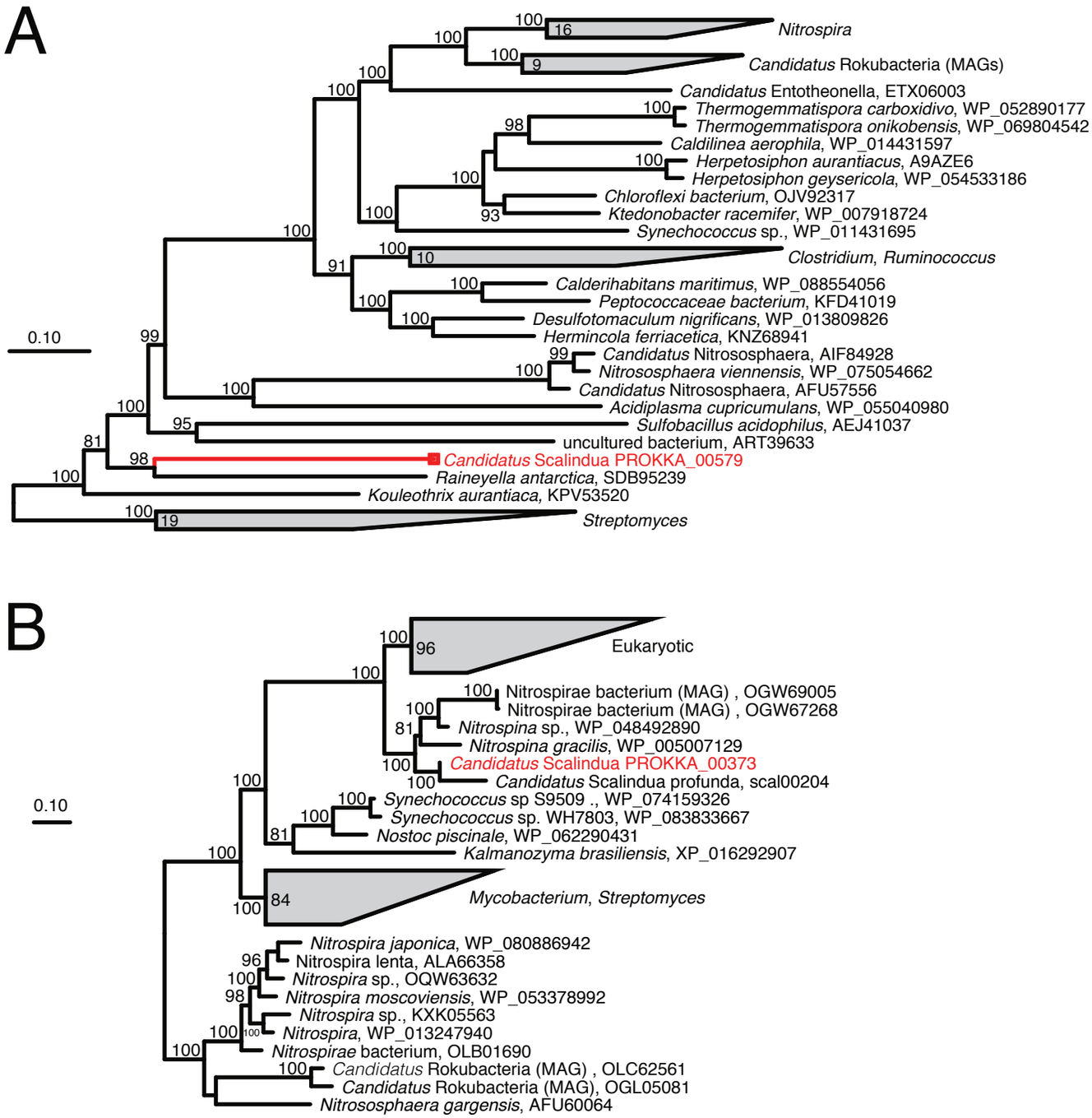
SRA identifier	DNA/RNA	Environment	reads[1]	*ureA*	*ureB*	*ureC*	*ureD*	*ureE*	*ureF*	*ureG*	*cynS*
SRX025906	DNA	50m, oxycline, Chile, 2008	393403	-	-	-	-	-	-	-	-
SRX025907	RNA	50m, oxycline, Chile, 2008	379333	-	-	-	-	-	-	-	-
SRX025908	DNA	85m, base of oxycline, Chile, 2008	595662	-	-	-	-	-	-	-	-
SRX025909	RNA	85m, base of oxycline, OMZ, Chile, 2008	142020	-	-	-	-	-	-	-	-
SRX025910	DNA	110m, upper anoxic OMZ, Chile, 2008	197843	-	-	-	1	-	-	1	-
SRX025911	RNA	110m, upper anoxic OMZ, Chile, 2008	514076	1	-	2	1	-	-	-	-
SRX025912	DNA	200m, anoxic OMZ core, Chile, 2008	516426	-	4	8	10	6	5	6	2
SRX025913	RNA	200m, anoxic OMZ core, Chile, 2008	441273	-	1	2	-	2	-	2	4
SRX080956	DNA	50m, oxycline, Chile, 2010	1534798	-	-	-	-	-	-	-	-
SRX080960	DNA	110m, upper anoxic OMZ, Chile, 2010	1459563	1	-	1	2	2	-	2	
SRX080955	DNA	110m, upper anoxic OMZ, Chile, 2009	917531	1	-	-	1	-	-	-	1
SRR070082	DNA	200m, anoxic OMZ core, Chile, 2009	930936	1	1	1	2	2	1	1	1
SRR2657589	DNA	600m, anoxic OMZ core, Arabian Sea, 200	2588720	2	0	2	3	1	1	0	-
SRX1878045	RNA	200m, anoxic seasonal OMZ, Saanich Inle	69907523	-	-	-	-	-	-	-	-

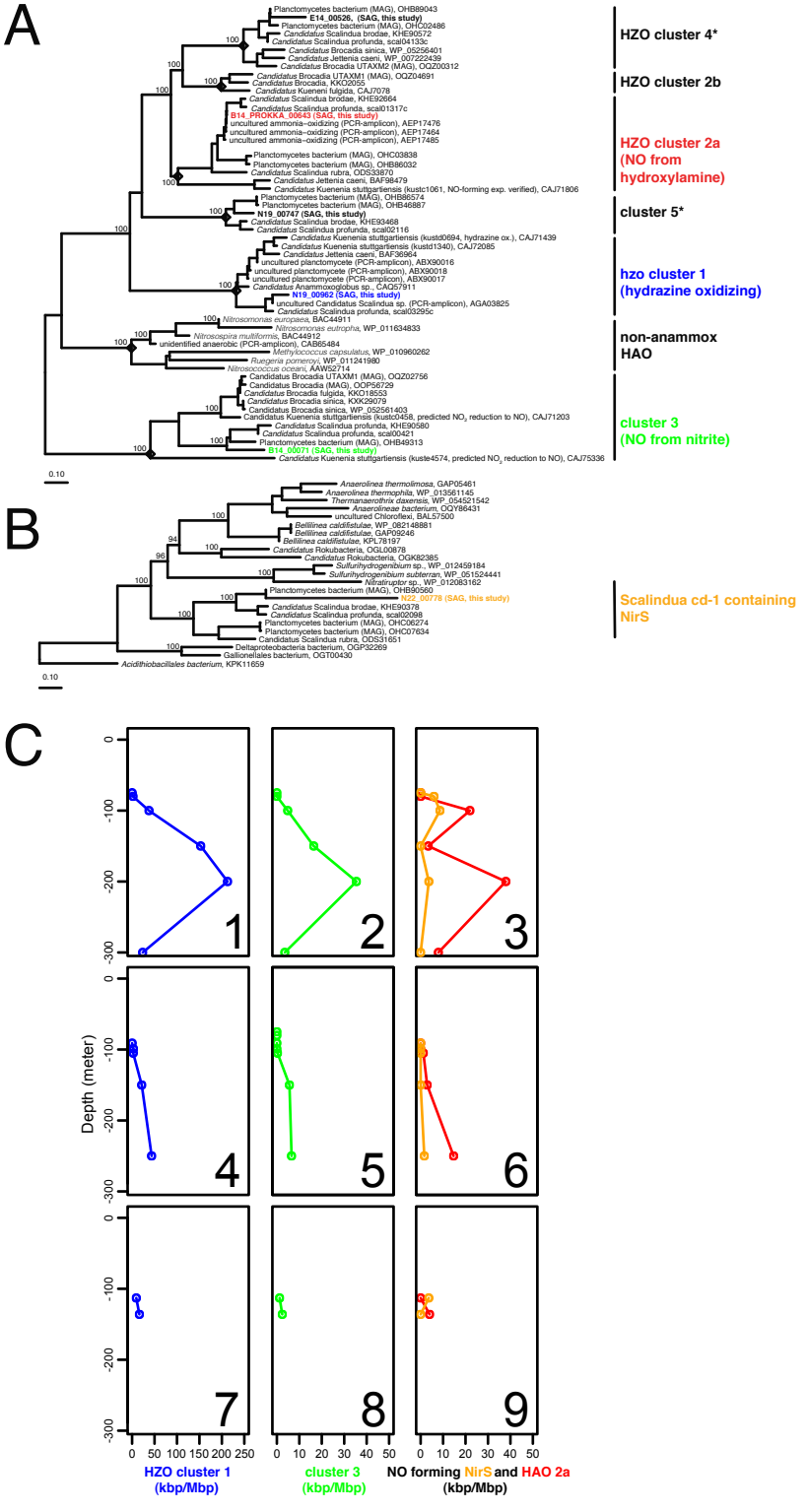
[1] Total number of sequence fragments in dataset


depth (meters)
0
100
200
300
400
-300
A
B
C
D
E
F
G
H
I
0 50 100 150 200 250
oxygen (µM)
0 1 2 3 4 5 6
anammox rate (nM N2 d-1)
0 20 40 60 80 100
Scalindua cynS (kbp/Mbp)
0 1 2 3 4 5
nitrite (µM)
0 2000 6000 10000
Scalindua transcripts (kbp/Mbp)
0 2 4 6 8 10
Scalindua ureC (kbp/Mbp)
0 50 100 150 200
ammonium (nM)


Arabian Sea
Namibian coast
Eastern Tropical North Pacific, Mexico
(This study)
Eastern Tropical South Pacific, Peru/Chile
Candidatus (Ca.) taxa
EU478597, Station 1 221 m, ArabianSea_bE9
EU478624, Station 1 221 m, ArabianSea_cD10
EU478601, Station 1 221 m, ArabianSea_aA6
EU478593, Station 1 221 m, ArabianSea_cD9
EU478595, Station 1 221 m, ArabianSea_CE9
SAG_J11, Station 6 300m
1..1522 EU478604, Station 1 221 m, ArabianSea_aG2
SAG_B14, Station 6 300m
SAG_N19, Station 6 300m
SAG_E14, Station 6 300m
EU478615, Station 1 221 m, ArabianSea_bD7
EU478617, Station 1 221 m, ArabianSea_bE7
EU478612, Station 1 221 m, ArabianSea_aD8
EU478618, Arabian Sea 221 m, ArabianSea_bC12
SAG_L7, Station 6 125m
EU478603, Station 1 221 m, ArabianSea_aE9
SAG_L14, Station 6 300m
SAG_N13, Station 6 300m
SAG_N22, Station 6 300m
EU039845, upper oxcline northern Chile, PD00250.H7
EU478633, Station 202 52 m, Namibia_37
EU478644, Station 202 52 m, Namibia_182
AY257181, *Ca.* Scalindua sorokinii
EU478631, Station 202; 52 m, Namibia_24
EU142948, *Ca.* Scalindua brodae, EN20 (enrichment)
AY254883, *Ca.* Scalindua brodae EN8 (enrichment)
EU478667, Station 4 35 m, Peru_39
EU478662, Station 4 35 m, Peru_28
EU478652, Station 4 35 m, Peru_15
EU478674, Station 4 35 m, Peru_54
100/94/93
100/94/-
100/100/91
100/99/100
3
Sediment cluster
AY254882, *Ca.* Scalindua wagneri (enrichment)
100/100/99
10
Jettenia, *Brocadia*, *Anammoxoglobus*
Arabian Sea cluster
Namibian & Peruvian clusters
0.10
Nucleotide substitutions


A
urease
urease accesory proteins
urea ABC transporter
(% complete, % contamined, strain het.)
nitrogen regulatory protein PII (00001)
ureA (00003)
ureB (00004)
ureC (00005)
ureE (00007)
ureF (00008)
ureG (00009)
ureD (00010)
JGI Scaffold 01859, 9969 bp
(NA, NA, NA)
SAG-G15, contig 30, 8640 bp
(29%, 3.5%, 40%)
SAG-N19, contig 29, 10816 bp
(46%, 1%, 0%)
SAG-M13, contig 1, 39476 bp
(4%, 0%, 0%)
glyceraldehyde 3-phosphatase (00001)
recombination protein xerC (00003)
helix turn helix protein (00004)
helicase loading factor dnaC (00005)
caspase (00011)
inner membrane protein YhI (00011)
pentapeptide repeats (00016)
GMP synthase (00021)
caspase domain protein (00024)
recombination protein xerC (00025)
aquaporin Z2 (00026)
effector protein pipB2 (00029)
60 kD chaperonin (00030)
ureD (00031)
ureG (00032)
ureF (00033)
ureE (00034)
ureC (00036)
ureB (00037)
ureA (00038)
ammonia channel (00040)
urtC (00041)
urtD (00042)
urtE (00043)
100%
98%
B
cyanate hydratase
formate dehydrogeanse (00372)
cynS (00373)
TPR repeat containing protein YfgC (00374)
adenyl sulfate kinase (00375)
cyclic 2,3 diphosphoglycerate synthase (00376)
hypothetical (00377)
hypothetical (00378)
carbon storage regulator (00380)
NADPH dependent FMN reductase (00381)
hypothetical (00382)
SAG G15, contig 26, 9181 bp
SAG N22, contig 11, 15005 bp
GIY-YIG nuclease superfamily (00183)
SMR domain protein (00184)
hypothetical (00185)
bacteriohemerythrin (00186)
3'-5' cyclic adenosine monophosate (00187)
ribonuclease BN (00190)
threonylcarbamoyladenosine tRNA (00191)
formate dehydrogeanse (00193)
cynS (00194)
TPR repeat containing protein YfgC (00195)
adenyl sulfate kinase (00196)
cyclic 2,3 diphosphoglycerate synthase (00197)
100%
98%


A
0.10
Nitrospira
16
Candidatus Rokubacteria (MAGs)
9
Candidatus Entotheonella, ETX06003
Thermogemmatispora carboxidivo, WP_052890177
Thermogemmatispora onikobensis, WP_069804542
Caldilinea aerophila, WP_014431597
Herpetosiphon aurantiacus, A9AZE6
Herpetosiphon geysericola, WP_054533186
Chloroflexi bacterium, OJV92317
Ktedonobacter racemifer, WP_007918724
Synechococcus sp., WP_011431695
Clostridium, Ruminococcus
10
Calderihabitans maritimus, WP_088554056
Peptococcaceae bacterium, KFD41019
Desulfotomaculum nigrificans, WP_013809826
Hermincola ferriacetica, KNZ68941
Candidatus Nitrososphaera, AIF84928
Nitrososphaera viennensis, WP_075054662
Candidatus Nitrososphaera, AFU57556
Acidiplasma cupricumulans, WP_055040980
Sulfobacillus acidophilus, AEJ41037
uncultured bacterium, ART39633
Candidatus Scalindua PROKKA_00579
Raineyella antarctica, SDB95239
Kouleothrix aurantiaca, KPV53520
Streptomyces
19
B
0.10
Eukaryotic
96
Nitrospirae bacterium (MAG) , OGW69005
Nitrospirae bacterium (MAG) , OGW67268
Nitrospina sp., WP_048492890
Nitrospina gracilis, WP_005007129
Candidatus Scalindua PROKKA_00373
Candidatus Scalindua profunda, scal00204
Synechococcus sp S9509 ., WP_074159326
Synechococcus sp. WH7803, WP_083833667
Nostoc piscinale, WP_062290431
Kalmanozyma brasiliensis, XP_016292907
Mycobacterium, Streptomyces
84
Nitrospira japonica, WP_080886942
Nitrospira lenta, ALA66358
Nitrospira sp., OQW63632
Nitrospira moscoviensis, WP_053378992
Nitrospira sp., KXK05563
Nitrospira, WP_013247940
Nitrospirae bacterium, OLB01690
Candidatus Rokubacteria (MAG) , OLC62561
Candidatus Rokubacteria (MAG), OGL05081
Nitrososphaera gargensis, AFU60064


A
Planctomycetes bacterium (MAG), OHB89043
E14_00526, (SAG, this study)
Planctomycetes bacterium (MAG), OHC02486
Candidatus Scalindua brodae, KHE90572
Candidatus Scalindua profunda, scal04133c
Candidatus Brocadia sinica, WP_05256401
Candidatus Jettenia caeni, WP_007222439
Candidatus Brocadia UTAXM2 (MAG), OQZ00312
Candidatus Brocadia UTAXM1 (MAG), OQZ04691
Candidatus Brocadia, KKO2055
Candidatus Kuenenii fulgida, CAJ7078
Candidatus Scalindua brodae, KHE92664
Candidatus Scalindua profunda, scal01317c
B14_PROKKA_00643 (SAG, this study)
uncultured ammonia-oxidizing (PCR-amplicon), AEP17476
uncultured ammonia-oxidizing (PCR-amplicon), AEP17464
uncultured ammonia-oxidizing (PCR-amplicon), AEP17485
Planctomycetes bacterium (MAG), OHC03838
Planctomycetes bacterium (MAG), OHB86032
Candidatus Scalindua rubra, ODS33870
Candidatus Jettenia caeni, BAF98479
Candidatus Kuenenia stuttgartiensis (kustc1061, NO-forming exp. verified), CAJ71806
Planctomycetes bacterium (MAG), OHB86574
Planctomycetes bacterium (MAG), OHB46887
N19_00747 (SAG, this study)
Candidatus Scalindua brodae, KHE93468
Candidatus Scalindua profunda, scal02116
Candidatus Kuenenia stuttgartiensis (kustd0694, hydrazine ox.), CAJ71439
Candidatus Kuenenia stuttgartiensis (kustd1340), CAJ72085
Candidatus Jettenia caeni, BAF36964
uncultured planctomycete (PCR-amplicon), ABX90016
uncultured planctomycete (PCR-amplicon), ABX90018
uncultured planctomycete (PCR-amplicon), ABX90017
Candidatus Anammoxoglobus sp., CAO57911
N19_00962 (SAG, this study)
uncultured Candidatus Scalindua sp. (PCR-amplicon), AGA03825
Candidatus Scalindua profunda, scal03295c
Nitrosomonas europaea, BAC44911
Nitrosomonas eutropha, WP_011634833
Nitrosospira multiformis, BAC44912
unidentified anaerobic (PCR-amplicon), CAB65484
Methylococcus capsulatus, WP_010960262
Ruegeria pomeroyi, WP_011241990
Nitrosococcus oceani, AAW52714
Candidatus Brocadia UTAXM1 (MAG), OQZ02756
Candidatus Brocadia (MAG), OOP56729
Candidatus Brocadia fulgida, KKO18553
Candidatus Brocadia sinica, KKK29079
Candidatus Brocadia sinica, WP_052561403
Candidatus Kuenenia stuttgartiensis (kustc0458, predicted NO2 reduction to NO), CAJ71203
Candidatus Scalindua profunda, KHE90580
Candidatus Scalindua profunda, scal00421
Planctomycetes bacterium (MAG), OHB49313
B14_00071 (SAG, this study)
Candidatus Kuenenia stuttgartiensis (kuste4574, predicted NO2 reduction to NO), CAJ75336
0.10
HZO cluster 4*
HZO cluster 2b
HZO cluster 2a (NO from hydroxylamine)
cluster 5*
hzo cluster 1 (hydrazine oxidizing)
non-anammox HAO
cluster 3 (NO from nitrite)
B
Anaerolinea thermolimosa, GAP05461
Anaerolinea thermophila, WP_013561145
Thermanaerothrix daxensis, WP_054521542
Anaerolineae bacterium, OQY86431
uncultured Chloroflexi, BAL57500
Bellilinea caldifistulae, WP_082148881
Bellilinea caldifistulae, GAP09246
Bellilinea caldifistulae, KPL78197
Candidatus Rokubacteria, OGL00878
Candidatus Rokubacteria, OGK82385
Sulfurihydrogenibium sp., WP_012459184
Sulfurihydrogenibium subterran, WP_051524441
Nitratiruptor sp., WP_012083162
Planctomycetes bacterium (MAG), OHB90560
G22_00778 (SAG, this study)
Candidatus Scalindua brodae, KHE90378
Candidatus Scalindua profunda, scal02098
Planctomycetes bacterium (MAG), OHC06274
Planctomycetes bacterium (MAG), OHC07634
Candidatus Scalindua rubra, ODS31651
Deltaproteobacteria bacterium, OGP32269
Gallionellales bacterium, OGT00430
Acidithiobacillales bacterium, KPK11659
0.10
Scalindua cd-1 containing NirS
C
Depth (meter)
0
-100
-200
-300
1
2
3
4
5
6
7
8
9
0 50 100 150 200 250
0 10 20 30 40 50
0 10 20 30 40 50
HZO cluster 1 (kbp/Mbp)
cluster 3 (kbp/Mbp)
NO forming NirS and HAO 2a (kbp/Mbp)