Supplementary Information for:

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

Contributors: Sangita Ganesh, Anthony D. Bertagnolli, Laura A. Bristow, Cory C. Padilla, Nigel Blackwood, Montserrat Aldunate, Annie Bourbonnais, Mark A. Altabet, Rex R. Malmstrom, Tanja Woyke, Osvaldo Ulloa, Konstantinos T. Konstantinidis, Bo Thamdrup, Frank J. Stewart

Supplementary Tables

Tables S1-S6 are appended as a separate Excel spreadsheet.

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.

Supplementary Figures

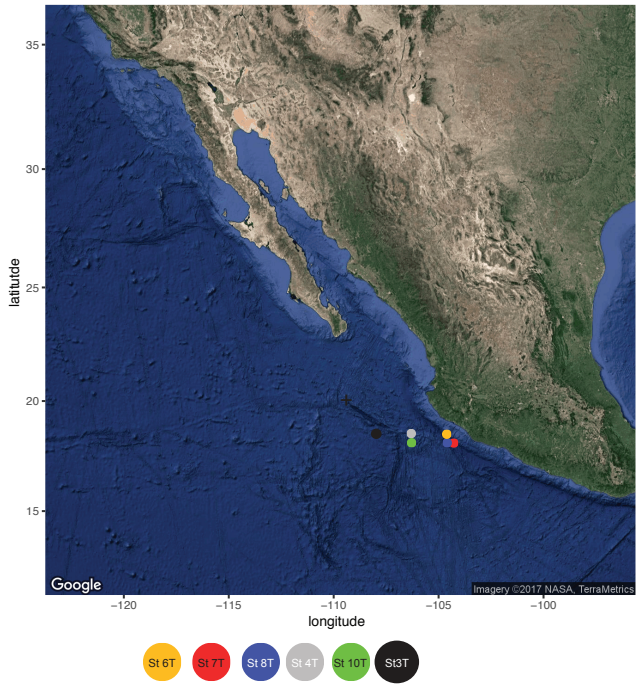

Figure S1. Map of study area, showing locations of stations sampled in 2014 and identified in Figure 1 (main text). Samples for SAG analysis were collected from station 6T in 2013. Exact coordinates can be found in Table S1.

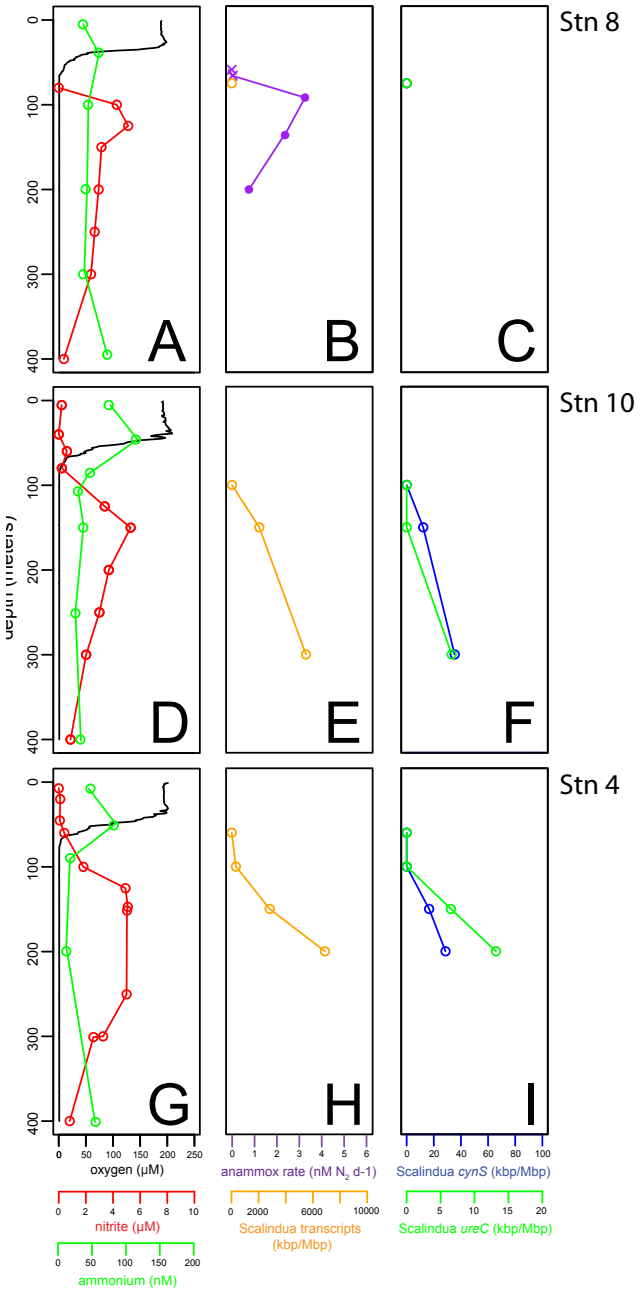

Figure S2. Anammox rates and representation of ETNP *Ca.* Scalindua transcripts relative to dissolved oxygen, ammonium, and nitrite concentrations at three ETNP stations: 8 (A-C), 10 (D-F), and 4 (G-I). The first column displays dissolved oxygen (black line, µM), nitrite (red circles and line, µM), and 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.

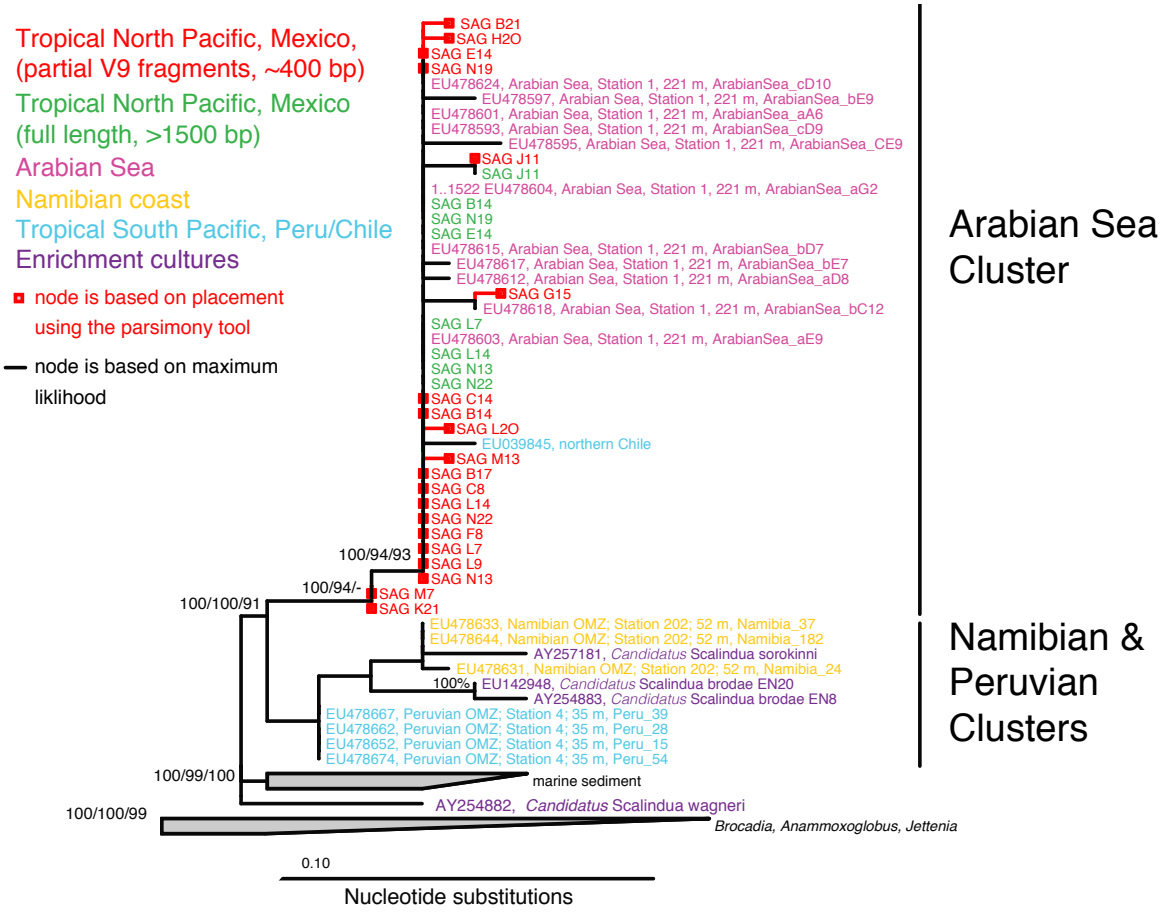

Figure S3. Phylogenetic approximation of PCR-amplified 16S rRNA genes generated from SAG template DNA (following multiple displacement amplification). Sequences were inserted into the backbone tree based on Figure 2 using the parsimony tool in ARB, and hence represent a phylogenetic approximation.

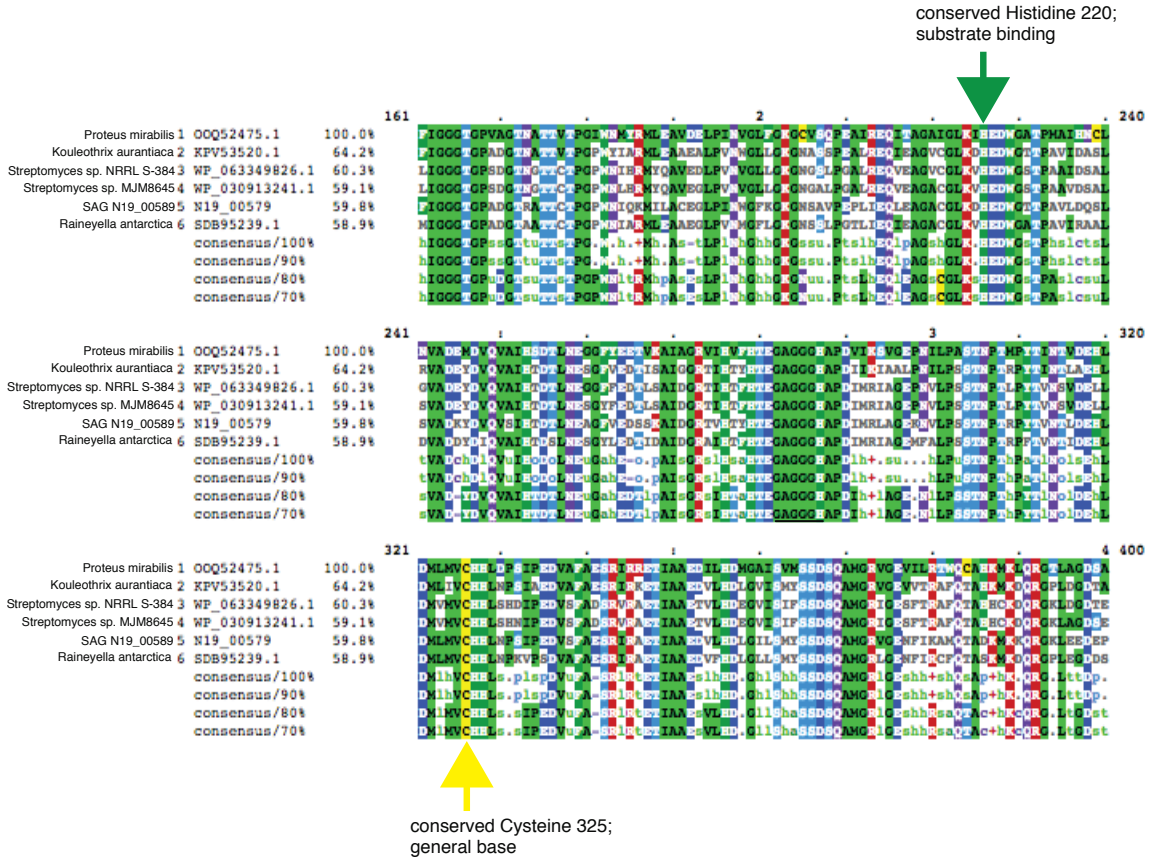

Figure S4. Alignment of UreC amino acid sequences from characterized urease-positive organisms *Proteus mirabilis*, *Streptomyces* sp. NRLL and MJM, and from two taxa (*Kouleothrix aurantiaca* and *Raineyella antarctica*) identified as best BLASTP matches to the ETNP *Ca*. Scalindua SAG UreC (UreC from 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).

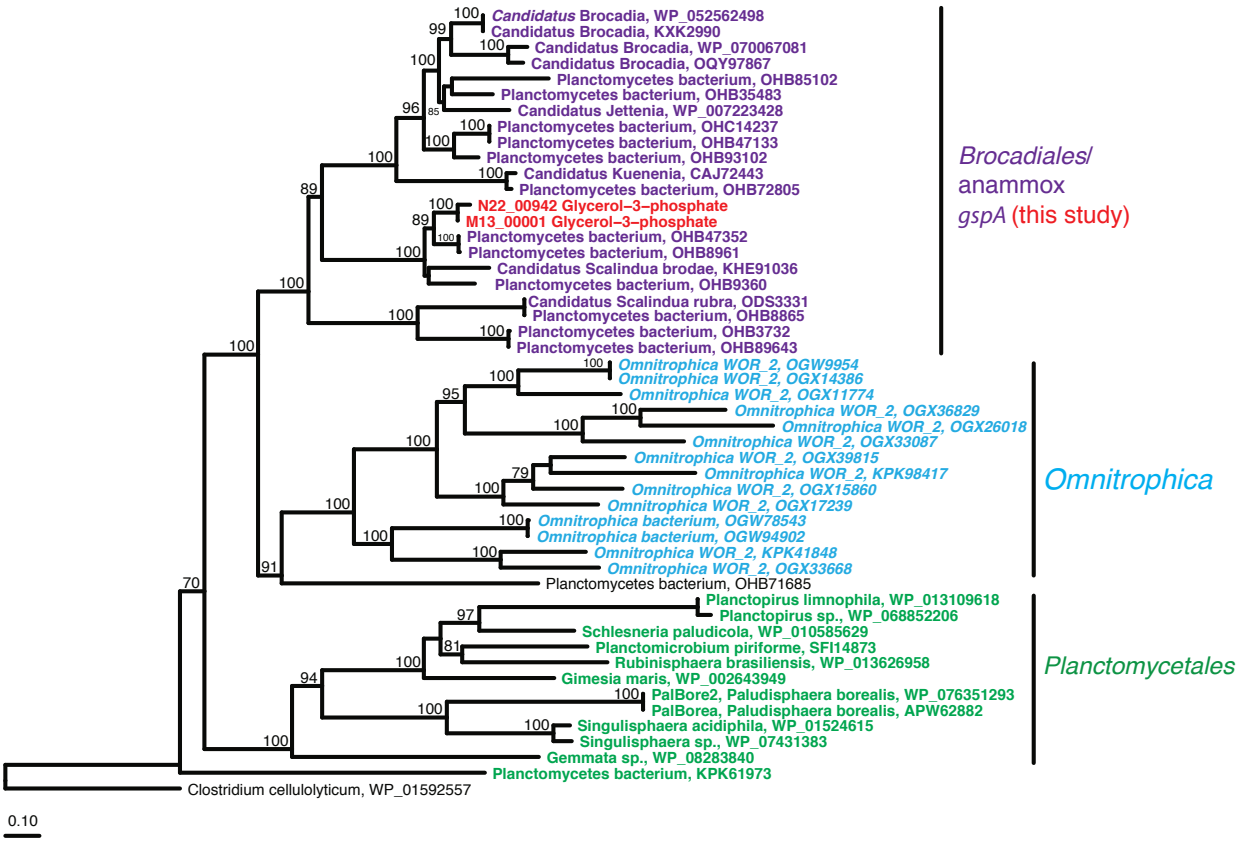

Figure S5. Phylogenetic approximation of SAG-associated glyceraldehyde 3-phosphate dehydrogenase, GspA (red). Sequences include GspA from a large *ure* and *urt*-containing contig from SAG M13 (see Figure 3A), and GspA from a smaller contig from SAG N22. Purple sequences were identified based on BLASTP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

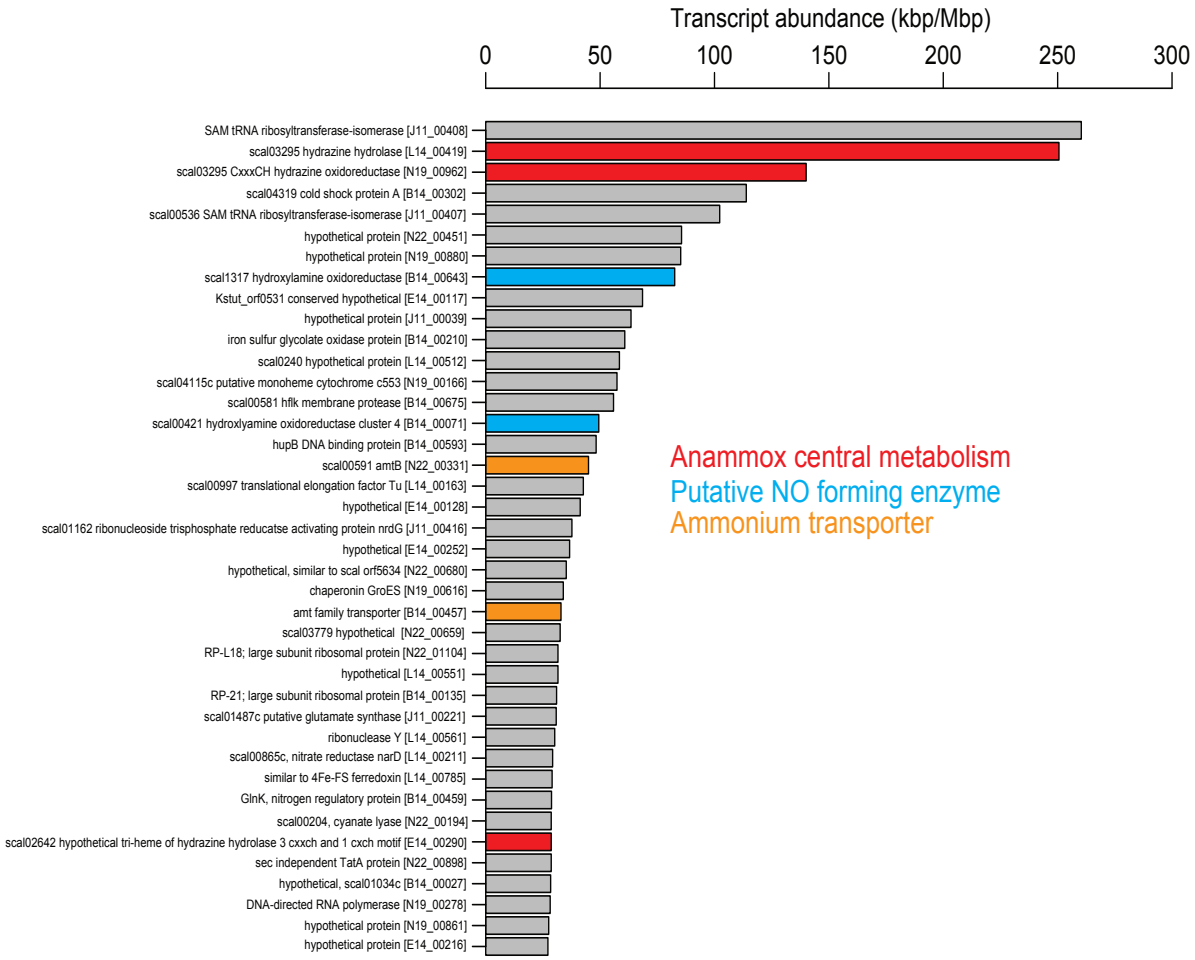

Figure S6. Top 40 most highly transcribed ETNP *Ca.* Scalindua genes observed at Station 6T (200 m, AMZ core).

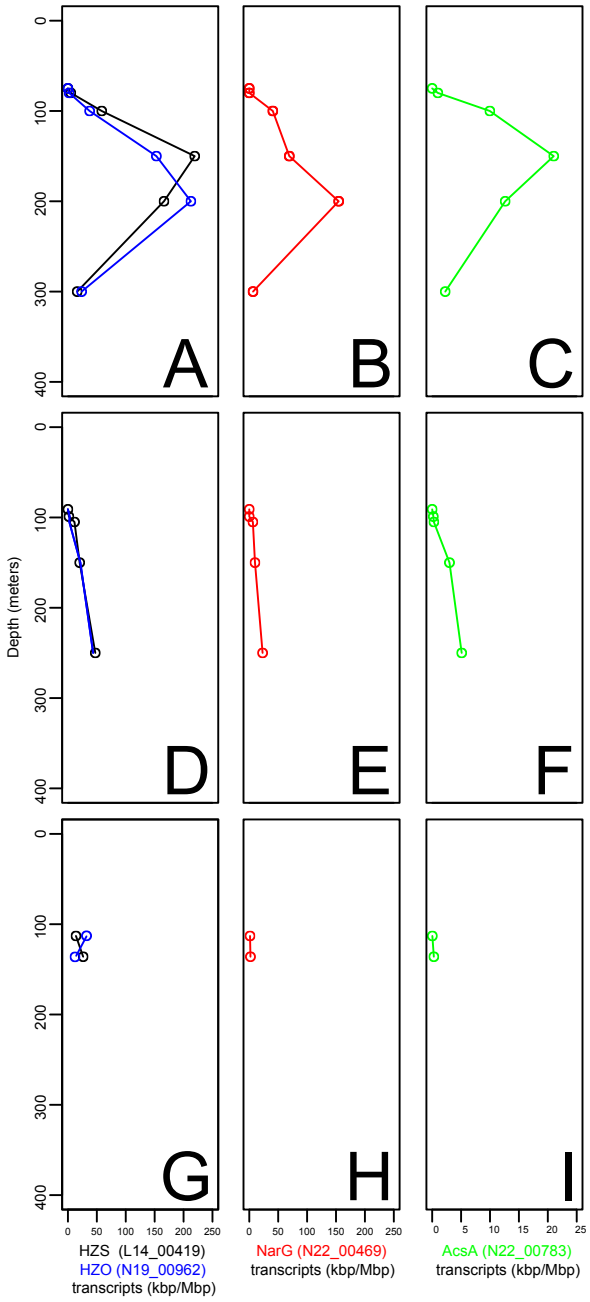

Figure S7. Proportional abundance of transcripts encoding the hydrazine-producing (hydrazine synthase) and consuming (hydrazine oxidoreductase) enzymes, nitrite/nitrate oxidoreductases likely involved in nitrite oxidation, and acetyl coA synthase involved in the Wood-Ljungdahl pathway at ETNP stations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