Correlated expression of archaeal ammonia oxidation machinery across disparate environmental and culture conditions

Authors:
Paul Carini[1,3], Christopher L. Dupont[2], Alyson E. Santoro[1,4]

[1]Horn Point Laboratory, University of Maryland Center for Environmental Science, Cambridge, MD 21613
[2]J. Craig Venter Institute, San Diego, CA 92037
[3]Present Address: Department of Soil, Water and Environmental Science, University of Arizona, Tucson, AZ 85721
[4]Present address: Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, CA 93106

Keywords: Thaumarchaea, nitrification, metatranscriptome, NirK, Nitrogen cycling

Corresponding Authors: Alyson Santoro (alyson.santoro@lifesci.ucsb.edu) and Paul Carini (paulcarini@email.arizona.edu)

Abstract

Although marine metatranscriptomes suggest planktonic ammonia-oxidizing thaumarchaea are among the most active microbes in marine waters[1-4], we understand little about how thaumarchaeal expression patterns relate to substrate utilization and activity. Here, we characterize the global transcriptional response of a marine ammonia-oxidizing thaumarchaeon, '*Candidatus* Nitrosopelagicus brevis' str. CN25, to ammonia limitation. We further describe the genome and transcriptome of *Ca*. N. brevis str. U25, a new thaumarchaeal strain capable of oxidizing ammonia derived from urea. Ammonium limitation in CN25 resulted in reduced expression of transcripts coding for core ammonia oxidation proteins, and increased expression of a gene coding an Hsp20-like chaperone. Despite significantly different transcript abundances across treatments, *amoAB*, *nirK* and both ammonium transporter genes were always among the most abundant transcripts, regardless of growth state. *Ca*. N. brevis str. U25 cells growing on urea expressed a urea transporter 139-fold more than the urease catalytic subunit *ureC*, indicating that the expression of urea acquisition machinery is favored over urease genes during exponential growth. Gene co-expression networks derived from transcriptomes from CN25 and U25 cultures and ten thaumarchaea-

enriched metatranscriptomes revealed a high degree of correlated gene expression across disparate environmental conditions. We show *nirK* is tightly co-expressed with *amoABC*, suggesting a central role for NirK in ammonia oxidation. These findings demonstrate how transcriptomes from microbial cultures can be used to contextualize and identify gene expression relationships that are otherwise enigmatic.

Introduction:

Ammonia-oxidizing thaumarchaea are ubiquitous and abundant in the oceans, accounting for >30% of all cells below the thermocline[5,6] and are integral organisms in oxygen minimum zones[4,7-11]. In many marine environments, thaumarchaeal transcripts are among the most abundant that can be mapped to available prokaryotic genomes[1-4]. The most frequently detected thaumarchaeal transcripts encode for proteins involved in ammonia oxidation and acquisition, including ammonia monooxygenase subunits (*amoABC*), ammonium transporters (*amtB*), a putative Cu-containing nitrite reductase (*nirK*), and structural cellular components (for example, S-layer proteins[12]). In addition to dissolved ammonia, some marine nitrifying archaea utilize ammonia derived from urease-catalyzed urea hydrolysis as a chemolithoautotrophic growth substrate[13,14] and thaumarchaeal urease genes and transcripts have been detected in marine environments[15-18]. Despite the abundance of thaumarchaeal transcripts in natural assemblages, we still have a poor understanding of how the relative abundance of thaumarchaeal transcript markers such as *amoA*, *nirK* and *ureC* relate to nutrient availability across environmental gradients.

Thaumarchaeal ammonia oxidation is initiated by the oxidation of ammonia to hydroxylamine (NH_2OH) by the ammonia monooxygenase enzyme complex (Amo)[19], but the enzyme(s) catalyzing the oxidation of NH_2OH to nitrite (NO_2^-) have not been confirmed in thaumarchaea[20]. Orthologs of the bacterial hydroxylamine oxidoreductase (Hao) or c-type cytochrome synthesis and assembly machinery, thought to be required for NH_2OH oxidation and electron transfer in ammonia-oxidizing bacteria (AOB)[21], are absent from all sequenced thaumarchaeal genomes[22-24]. Instead, unidentified Cu-containing metalloenzymes or F_{420}-dependent monooxygenases are speculated to be involved in NH_2OH oxidation and electron transfer to archaeal terminal oxidases[20,25]. and may involve nitric oxide (NO) as either a direct intermediate or an electron shuttle[26-28]. While the precursor to NO has not yet been elucidated, all free-living thaumarchaea with complete genomes encode a Cu-containing

multicopper oxidase with homology to Cu nitrite reductases (NirK) and may be responsible for the reduction of NO_2^- to NO[25].

'*Candidatus* Nitrosopelagicus brevis' str. CN25 is a cultured representative of ubiquitous and abundant pelagic thaumarchaeal populations in the shallow oligotrophic ocean[24,29]. Here, we describe the genome and transcriptome during urea based growth of a *Ca*. N. brevis strain that can utilize ammonia cleaved from urea as a sole chemolithoautotrophic growth substrate. Additionally, we use *Ca*. N. brevis str. CN25 to investigate the transcriptional response to ammonium limitation in laboratory culture. These transcriptomes are further analyzed in the context of several marine metatranscriptomes and used to identify conserved gene co-expression networks.

Results and Discussion

'*Candidatus* Nitrosopelagicus brevis' strain U25 genome and transcriptome: *Ca*. N. brevis str. CN25 was originally enriched using ammonium as the sole chemolithoautotrophic growth substrate[29]. A urea-utilizing enrichment was obtained from this original CN25 enrichment via subculturing with urea as a sole nitrogen and energy source. This second lineage was used for the shotgun metagenome sequencing and experiments described here. After assembly and contig binning based on nucleotide frequencies and coverage, we obtained a three contig genome of this urea-utilizing thaumarchaeon (Supplementary Figure 1a). This genome is nearly identical to the *Ca*. N. brevis CN25 genome with regards to *i)* gene content; *ii)* genome organization (Supplementary Figure 1b); and *iii)* genome wide average nucleotide identity (99.99%; Supplementary Figure 2). We found 19 additional genes at four distinct genomic loci in this strain, relative to CN25 (Supplementary Table 1). The largest of these insertions (15 contiguous genes) includes 11 genes coding urea utilization machinery, including *ureABCDEFG*, which codes for urease and its chaperones, two urea sodium:solute symporter family (SSSF) transporters, a transcriptional regulator and several hypothetical proteins (Fig. 1a). We designate this urea-utilizing thaumarchaeon '*Candidatus* Nitrosopelagicus brevis' strain U25.

We sequenced a transcriptome from *Ca*. N. brevis str. U25 growing exponentially with urea as the growth substrate. Only one transcript from the chromosomal insertion containing the urea transport and metabolism

genes (Fig. 1a) was among the top 50 transcripts detected: A7X95_00990, coding for a putative urea SSSF transporter (ranked 13.7 ± 0.33; mean ± SE, *n* = 3). Surprisingly, transcripts coding for catalytic urease components, or the second putative urea SSSF transporter (A7X95_00985) located immediately adjacent to A7X95_00990, were not nearly as abundant as A7X95_00990. For example, the mean expression levels of *ureC*, coding for the fused catalytic αβ-subunit of urease, and *ureA* (γ urease subunit) were 139 and 784-fold less abundant than that of A7X95_00990, respectively (ranked 390 ± 19.7 and 950 ± 50.8, respectively; mean ± SD, *n*=3). A7X95_00985, coding for the second urea SSSF transporter, was also expressed at a low level, comparable to *ureA*, ranked 940 ± 92.0. During growth on urea, transcripts for genes coding for an ammonium transporter (AMT1), ammonia monooxygenase subunits (*amoAB*) and nitrite reductase (*nirK*) were within the top ten most abundant transcripts detected in strain U25.

Although urea utilization genes have been detected in wild thaumarchaeal populations, we have a poor understanding of how the abundances of urease transcripts relate to growth and activity. To contextualize the expression patterns observed in U25, we compared the relative rank of the transcript abundances for only those genes coding for urea uptake and catabolism (Fig. 1a) under laboratory growth conditions to the relative rank of the transcript abundances of the same genes from several deeply sequenced marine metatranscriptomes. The SSSF urea transporter (A7X95_00990) was the most abundant urea-related gene transcript in 38% of the environmental datasets (n=8) we investigated (Fig. 1b). In contrast to culture conditions, where the SSSF urea transporter was the most abundant urea-related gene transcript, *ureC* was the most abundant transcript in 25% of the environmental datasets (Fig. 1b). This shows that variability in the relative transcriptional activity of urea transport and catabolism genes is not unusual. Our finding that *ureC* was not highly expressed in exponentially growing cells also helps to explain previous field observations of low *ureC* expression, and suggests the abundance of *ureC* transcripts may be a poor molecular biomarker of active urea-based nitrification. For example, *ureC* expression and urea-based nitrification were found to be only weakly correlated across several environments[30], in contrast to high correlation between *amoA* expression and ammonia oxidation rates[31]. Similarly, in Arctic samples collected across seasons, *ureC* genes were detected, yet *ureC* transcripts were only sporadically detected and at low abundances[17].

Transcriptional response to ammonium limitation in '*Ca*. N. brevis' strain CN25*:* To understand adaptive mechanisms during ammonium starvation, we explored the transcriptional response of *Ca*. N. brevis CN25 to ammonium limitation. A total of 51 gene transcripts were differentially abundant when comparing the exponential growth phase of CN25 to ammonium-limited stationary phase (generalized linear model likelihood ratio test FDR ≤ 0.01 and ≥ 2-fold change in abundance; Supplementary Table 2). The gene transcripts that were significantly less abundant in stationary phase included *amoA*, *amoB*, *nirK*, both *amtB*-like ammonium transporters (AMT1=T478_1378; AMT2=T478_1350), several additional Cu-containing metalloenzymes, and two ferredoxin-like 4Fe-4S binding domain proteins (Fd1=T478_1472 and Fd2=T478_1259) (Fig. 2, Supplementary Table 2). These results are similar to previous studies where ammonium limitation of the ammonia-oxidizing thaumarchaeon *Nitrosopumilus maritimus* was characterized by a reduction in transcript abundances related to energy metabolism and ammonium acquisition[12]. Only two genes, T478_1481, coding an Hsp20/α-crystallin domain small heat shock protein, and T478_1394, annotated as a hypothetical protein, were significantly more abundant (~10-fold) in ammonium-limited stationary phase (Fig. 2, Supplementary Table 2). Hsp20 is a molecular chaperone that enhances thermotolerance and binds to unfolded proteins to prevent aggregation[32].

The upregulation of Hsp20 transcripts and concomitant downregulation of transcripts coding enzymes integral to energy production suggests that one adaptation to ammonium-limited stationary phase is to protect existing proteins from degradation. While it is important to note that our experimental design cannot distinguish between responses to energy limitation versus anabolic nitrogen limitation, a similar lack of transcriptional response to ammonium limitation was previously observed in both AOB and oligotrophic marine bacteria. For example, our finding that transcripts of only a few genes are more abundant during ammonium starvation parallels studies of ammonium limitation in the bacterium *Nitrosomonas europaea*[33]. Interestingly, despite *amoC* being abundant in both exponential and stationary phases (the 13th and 7th most abundant transcript, respectively) we did not observe a significant difference in the abundance of *amoC* between growth phases, which has been implicated in stress response in *N. europaea*[34]. Like our results with *Ca*. N. brevis CN25, the marine chemoorganoheterotroph '*Candidatus* Pelagibacter ubique,' did not

differentially regulate any genes in response to ammonium limitation[35]. Despite the lack of transcriptional differences in *Ca*. P. ubique, the abundances of several peptides related to nitrogen metabolism did change significantly upon nitrogen limitation, including the molecular chaperone protein GroEL[35]. The similar transcriptional responses to nitrogen starvation by *Ca*. N. brevis and *Ca*. P. ubique are consistent with the limited capacity of marine oligotrophs to respond rapidly to environmental change[36,37].

Despite significantly different transcript abundances of essential ammonia oxidation and transport genes across growth conditions, the rank order of these transcripts within a given treatment were similar, irrespective of growth condition. That is, the most abundant gene transcripts in exponential phase were generally still the most abundant transcripts in ammonium-limited stationary phase (Fig. 2b). For example, transcripts for 32 genes (64%) were in the top 50 most abundant transcripts in both exponential and stationary phase (Fig. 2b). Interestingly, the abundances of 18 of these transcripts were also significantly different across treatments (red points in Fig. 2b), illustrating that although transcripts can be differentially abundant across paired treatments, the changes in their relative cellular abundance may be subtler. Several of these consistently abundant but differentially expressed transcripts are common molecular markers predicted to be essential for ammonia oxidation and transport, including the ammonium transporters AMT1 and AMT2, ammonia monooxygenase subunits (*amoAB*) and nitrite reductase (*nirK*).

Correlated gene expression across disparate environments: Controlled laboratory experiments such as those described above help us to understand gene regulation by isolating one experimental variable at a time. However, gene expression patterns observed in natural thaumarchaeal populations are the result of cells responding to complex environmental conditions that can be cryptic and therefore difficult to mimic in the laboratory. To identify clusters of co-expressed genes across disparate environmental conditions, and relate them to our laboratory findings, we constructed and analyzed a gene expression correlation network constructed from transcriptomes of exponentially growing CN25 and U25 cultures and ten marine metatranscriptomes. The transcription of 1,407 of the 1,464 non-redundant genes in the two *Ca*. N. brevis genomes were significantly positively correlated with at least one other gene (Pearson's $r \geq 0.80$, $q \leq 0.025$; Fig. 3, Supplementary Table 3). Network modularity is a measure of the group connectivity within a network, where

connections contained within a module are denser than connections between modules. Modularity values range from -0.5 to 1, where 1 describes a highly modular system. The modularity of this positive correlation network was 0.71, indicating a high degree of modularity. Genes with positively correlated expression organized into 38 groups (modules), ranging in membership size from 2 to 236 genes (mean module size = 38.0 genes).

Genes encoding putative components of the core ammonia oxidation and transport machinery are significantly co-expressed across distinct environmental and laboratory conditions. A single 15-gene module (module 11 in Fig. 3) contained: *amoABCX*, AMT1, *nirK*, two PEFG-CTERM domain proteins, Fd1 and Fd2, an Fe-S cluster assembly protein, a membrane bound cupredoxin-containing protein, and three hypothetical proteins. Further investigation of these hypothetical proteins suggests that (T478_0057) is a putative archaeal cell division protein related to the endosomal sorting complexes involved in membrane trafficking (ESCRT)-III[38,39]. Our finding that *amoA*, *amoB* and *nirK* transcripts are abundant and co-expressed with other genes coding for membrane-bound Cu-containing metalloproteins (T478_1362 and T478_0895) implies the products of these genes may participate in ammonia oxidation. In contrast to this, previous speculation implicated multicopper oxidases[20,23,28] or novel F_{420}-dependent monooxygenases[25] in ammonia oxidation chemolithotrophy (specifically NH_2OH oxidation) based on Cu redox chemistry or ortholog conservation across thaumarchaeal genomes. However, the genes put forth in those studies were not present in module 11 (referred to as the AMO module, herein), suggesting they may not be involved in core energy metabolism (Fig. 3, Supplementary Table 3). Moreover, we show that, on average, the AMO module is expressed at a higher level than other modules, and was more abundant in conditions where ammonia oxidation rates and thaumarchaeal abundances would be predicted to be high (Fig. 4). For example, consistent with previous reports of higher ammonia oxidation rates within hydrothermal plumes[40], we show that the AMO module is expressed highly within the Guyamas Deep hydrothermal plume, relative to background samples (Fig. 4). Moreover, similar to reports of increased thaumarchaeal gene expression in the mesopelagic[41], the AMO module is less abundant in the surface waters of Landsort Deep (0 and 5 m), relative to deeper waters (90 and 200 m) (Fig. 4).

A new proposed model for thaumarchaeal chemolithotrophy via ammonia oxidation: A previous model of thaumarchaeal ammonia oxidation proposed NO is derived from the reduction of NO_2^- by NirK, and that this NO is subsequently used to oxidize NH_2OH[28]. However, this model does not agree with tracer experiments using ^{18}O labeled water, which show that only one O atom from water is incorporated into NO_2^- produced by thaumarchaea[42,43]. If NO_2^- was reduced to NO and used to produce additional NO_2^-, the resulting NO_2^- would retain an average of more than one O atom from water. Based on the co-expression data presented here, we propose two alternative models of thaumarchaeal ammonia oxidation that are consistent with previous isotope tracer data regarding the source of O atoms in NO_2^-. In both models, NirK and two membrane-anchored cupredoxins (T478_1362 and T478_0895) act in concert to oxidize NH_2OH to NO_2^- in two steps: a three-electron oxidation of NH_2OH to NO, followed by a one-electron oxidation of NO to NO_2^- (Supplementary Figure 3). Both proposed pathways are consistent with the observed co-expression and high abundance of these transcripts across distinct environmental conditions (refs 1-4 and Figs 2-4). The predicted localization of ammonia oxidation in the pseudoperiplasm[20] is also a key aspect of the proposed models, as protein domain analysis with InterPro[44] indicates all proteins in these models, and pertinent cupredoxin domains, are likely localized in the pseudoperiplasmic space. While we cannot determine which reaction is conducted by which Cu metalloenzyme, both scenarios are more parsimonious than existing models and plausible based on existing bioinorganic chemistry literature.

Surprisingly, two proteins coding for ferredoxin-type 4Fe-4S-domains (Fd1 and Fd2) and an Fe-S cluster assembly protein were present in the AMO module, suggesting a central role for Fe-S cluster-containing proteins in the electron transport chain of thaumarchaea. Both Fd1 and Fd2 lack discernable signal sequences or PEFG domains, suggesting that they are localized in the cytoplasm. Ferredoxin-containing DNA binding transcriptional regulators have been implicated as NO sensors[45]. However, there are no predicted DNA-binding domains in Fd1 or Fd2. Sequence structure threading of Fd1 and Fd2 with phyre2[46] returned best structural matches to NADH dehydrogenase (ubiquinone) iron-sulfur protein 8 (threading confidence score = 99.8% for both Fd1 and Fd2; coverage of Fd1 was 99% and 71% for Fd2). Thus, we speculate that Fd1 and Fd2 participate in supplying electrons to the ubiquinone pool. This proposed

role suggests Fd1 and Fd2 may be involved in supplying the electrons necessary to initiate ammonia oxidation via Amo.

Gene membership of the remaining significantly co-expressed modules did not reveal a clear pattern of how the genes contained within a given module are functionally related (Supplementary Table 3). In other words, except for the AMO module, the putative functions of genes contained within each module did not necessarily belong to the same metabolic pathways or cellular process. There are several reasons this might be the case. First, our methods likely underestimate the true modularity of *Ca*. N. brevis gene expression. Some of the larger expression modules may be composed of distinct modules that we could not resolve because we did not sample an environment with physicochemical parameters necessary to resolve subtle gene expression patterns. Second, although our goal was to be conservative in our network construction, we may be missing important network structural components because of the thresholding parameters or our analysis technique. Further resolution of such gene expression patterns would require deeply sequenced metatranscriptomes from additional distinct environments and transcriptome analysis of additional thaumarchaea grown under diverse culture conditions. Moreover, we do not yet understand why certain genes or pathways are co-expressed with one another, and how transcript abundances manifest into functional potential in each environment or culture setting.

Conclusions:

Metatranscriptome analyses often use ranked lists of genes or relative transcript abundances as a proxy for cellular activity and substrate utilization capacity, with the implicit assumption that relative transcript abundances are directly and universally related to these processes. Here we show the rank of most thaumarchaeal transcripts reported as being abundant in the environment (*amoABC*, both *amtB* genes and *nirK*, for example) are relatively rank invariant across growth phases and environmental conditions. However, consistent with other studies of thaumarchaeal transcription[12], some of these genes were indeed significantly differentially abundant across paired treatments, showing ammonium availability did affect the abundances of ammonia oxidation and transport transcripts. One explanation for this observation is that although these transcripts are not ‘constitutive’ in a classic sense (that is, they are differentially abundant across paired experiments), they are instead

'affluent,' in that they make up a large part of the total transcript pool, irrespective of growth condition.

Discovering gene function in fastidious or uncultivated lineages remains one of the biggest challenges in environmental microbiology. Narrowing the scope of targets for detailed biochemical investigation is difficult because manipulative experiments are limited in their ability to identify networks of co-regulated genes by the number of environmental parameters we can recreate in a laboratory. The approach used here leverages a series of 'natural experiments' - in which the environmental conditions are incompletely characterized - to identify genes that share similar transcription patterns. In addition to our putative models of ammonia oxidation in thaumarchaea, this approach shows that 4Fe-4S cluster-containing proteins likely have an important role in ammonia oxidation, indicating a role for iron in archaeal nitrification, which has been previously under appreciated. Detailed biochemical characterization of NirK, other cupredoxin-containing proteins, Fd1 and Fd2 is the next step in understanding their specific role in core thaumarchaeal energy metabolism.

Methods:

Organism sources and cultivation conditions: '*Candidatus* Nitrosopelagicus brevis' str. CN25[24], was propagated in ONP medium[29], consisting of aged natural seawater (collected from 10 m depth at 15°S, 173°W on 23 October 2011; 0.2 µm pore size filtered at sea) amended with a chemolithoautotrophic nitrogen source (NH_4Cl or urea, described in 'Experimental design'), ampicillin (10.8 µM), streptomycin (68.6 µM), potassium phosphate (29.4 µM), and a chelated trace metal mix consisting of disodium ethylenediaminetetraacetic acid (14 µM), $FeCl_2$ (7.25 µM), $ZnCl_2$ (0.5 µM), $MnCl_2$ (0.5 µM), H_3BO_3 (1 µM), $CoCl_2$-$6H_2O$ (0.8 µM), $CuCl_2$-$2H_2O$ (0.1 µM), $NiCl_2$-H_2O (0.1 µM), Na_2MoO_4-$2H_2O$ (0.15 µM). '*Ca.* N. brevis' str. U25 was enriched from the original CN25 enrichment culture[29] using sequential transfers of the initial enrichment into ONP medium amended with 50-100 µM urea, instead of NH_4Cl, over a period of ~48 months. All enrichments were propagated in 250 mL polycarbonate flasks at 22°C in the dark and monitored for NO_2^- production using the Griess reagent colorimetric method[47]. Cell counts were obtained with a Millipore Guava EasyCyte 5HT flow cytometer as described previously[48].

Cell harvesting for and genome sequencing of 'Ca. N. brevis' str. U25: A *Ca*. N. brevis U25 enrichment culture that was grown exclusively with urea

as the sole chemolithoautotrophic growth substrate for >50 generations, was harvested by filtration on to 25 mm diameter, 0.22 µm pore-size Supor-200 filters and frozen at -80°C. DNA was extracted using a DNeasy blood & Tissue DNA extraction kit (Qiagen, Valencia, CA, USA), following the manufacturer's instructions. The DNA was treated with RNAse and examined using a Bioanlayzer 2100 (Agilent) with 500 ng serving as the input for library construction (NEBNext paired-end DNA Library Prep kit, New England Biolabs). The sample was sequenced on an Illumina MiSEQ (v2 chemistry, paired 150 bp reads). Reads were quality trimmed and served as the inputs to assembly with metaSPAdes (v 0.5, 70mer)[49]. The K-mer usage of and phylogenetic annotation of the assembled contigs were then used to visually identify a putative thaumarchaeal bin (Supplementary Figure 1a)[50]. The 3 contig genome was annotated using the JGI IMG pipeline and the PGAP pipeline at NCBI.

Experimental design, cell harvesting and RNA extraction for culture transcriptomes: All cultures were grown in ONP medium prepared with aged surface seawater in acid-washed polycarbonate bottles as described above. For experiments investigating the effects of ammonia limitation, ONP medium was amended with 50 µM NH_4Cl as the chemolithoautotrophic growth substrate. In this experiment, six replicates were prepared, three of which were harvested in late exponential phase (Exponential phase) and three of which were harvested in late NH_4Cl-limited stationary phase (Stationary phase) (Supplementary Fig. 4). We deliberately harvested the exponential phase cultures in late exponential phase to ensure maximal cell biomass for transcriptome analysis. NH_4Cl-limitation was verified by demonstrating a linear the dose response in maximal cell density to NH_4Cl additions (Supplementary Fig. 5).

Transcriptomes that were included in the network analysis were obtained from mid-exponential phase *Ca*. N. brevis strains CN25 and U25 that were growing on ONP medium amended with either 100 µM NH_4Cl (str. CN25; n=3) or 100 µM urea (str. U25; n=3) as growth substrates, respectively (Supplementary Fig. 6). Cells were harvested by filtration on to 25 mm diameter, 0.22 µm pore-size Supor-200 filters and frozen at -80°C. For RNA extraction, cells were disrupted as described in[51]. RNA was extracted using trizol-LS per the manufacturer's instructions and stored in nuclease free water at -80°C. Urea consumption by str. U25 was determined colorimetrically using the diacetyl monoxime method[52] (Supplementary Fig. 7).

Transcriptome sequencing and mapping for culture experiments: Transcriptome samples were prepared for sequencing according the Epicenter TotalScript protocol, which biases against rRNA. Libraries were trial sequenced on an Illumina MiSEQ to determine uniformity between barcodes and then fully sequenced in one 300 cycle NextSEQ run which generated 246.6 million paired-end 150 base pair reads. Raw Illumina reads in fastq format are interleaved to match paired ends. Sequencing primers and barcode indexes were identified by BLAST against the NCBI vector database and trimmed along with regions with Q scores < 30. Reads mapping to ribosomal RNAs were identified and removed using ribopicker. Reads were mapped (clc_ref_assemble_long –s 0.9, CLC genomics) to the *Ca*. N. brevis strains CN25 or U25 genome sequences. Raw read counts per open reading frame (ORF) were compiled.

Analysis of differentially abundant gene transcripts: Differential gene abundance analysis was performed using a generalized linear model likelihood ratio test in the edgeR software package (v 3.8.5) [53]. We defined significant differential abundance as those genes with a false discovery rate (FDR) ≤ 0.01 and greater than 2-fold abundance change across treatments.

Rank Analyses: Raw read counts per ORF were and scaled to expression units of reads per base per million reads mapped ($RPKM=(10^6 * C)/(NL/10^3)$) where C is the number of transcript reads mapped to an ORF; N is total reads mapped to all ORFs in the genome; and L is the ORF length in base pairs[54]. RPKM values were subsequently ranked with a rank of 1 depicting the most abundant transcript within a given treatment. Rank ties within a treatment were averaged.

Metatranscriptome mapping to genomes of Ca. N. brevis strains: Sequence reads from 68 metatranscriptomes were mapped to the *Ca*. N. brevis genomes at 50% nucleotide identity using CLC workbench (command: clc_ref_assemble –s 0.5) (Supplementary Table 4). The number of metatranscriptome reads that mapped to the *Ca*. N. brevis genomes were variable and ranged from 10 reads to 236,954 reads and mapped to 0.5-89% of the unique genes in the *Ca*. N. brevis genomes (Supplementary Table 4). Raw read counts per ORF were compiled (ORF n=1445 for str. CN25 and n=1461 for U25).

Network construction: Only those metatranscriptomes that mapped to ≥45% of the ORFs in the *Ca*. N. brevis genomes, along with the transcriptomes from *Ca*. N. brevis strains CN25 and U25 growing in exponential phase initiated with 100 μM NH_4Cl or urea, respectively, were included for network analysis. Of the 68 metatranscriptomes mapped to the *Ca*. N. brevis genomes, only 10 passed this filtering step and were used for network analysis (Supplementary Table 4). Read counts were scaled to RPKM expression units. RPKM scores calculated for individual culture transcriptome replicates were averaged (n=3) to avoid pseudo-replication effects in the network. The resulting RPKM expression values were rank-normalized to Van der Waerden (VdW) scores using the formula ($s= \Phi^{-1}(r/(n+1))$), where s is the VdW score for a gene, r is the rank for that observation, n is the sample size and Φ is the Φ^{th} quantile from the standard normal distribution using *tRank* in the multic R package. Pearson correlation coefficients and *P* value estimates were calculated for all gene:gene pairs across the VdW-normalized metatranscriptomes and culture experiments (n=10 and 2, respectively) with the *rcorr* command in the Hmisc R package[55]. To correct for multiple hypothesis testing, *q* values were computed from *P* value estimates using the qvalue R package [56]. Correlations with a *q* value ≤ 0.025 were used for network analysis. All correlations at this threshold were strongly correlated (Pearson's r ≥ 0.8).

Network Statistics: Network modularity and module membership were calculated in Gephi (0.8.2 beta) with the following settings: resolution 1.0, randomized and unweighted[57]. Network was visualized using the Fruchterman Reingold algorithm in Gephi.

Data Availability/Sources:

Transcriptomes from *Ca*. N. brevis str. CN25 and U25 can be found in the NCBI BioSample archive under accession numbers SAMN6290440-6290457. The U25 genome has been deposited at DDBJ/ENA/GenBank under the accession LXWN00000000. The version described in this paper is version LXWN01000000. The metatranscriptomic data from Landsort Deep in the Baltic is available from the Sequence Read Archive under numbers SAMN04943349-SAMN04943415.

Other metatranscriptomes analyzed in the network are publically available through iMicrobe (https://imicrobe.us) or NCBI's Short Read Archive through the following accession numbers: CAM_PROJ_Sapelo2008, CAM_PROJ_AmazonRiverPlume, CAM_PROJ_PacificOcean, CAM_P_0000545, SRA023632.1.

References:

1. Hollibaugh, J. T., Gifford, S., Sharma, S., Bano, N. & Moran, M. A. Metatranscriptomic analysis of ammonia-oxidizing organisms in an estuarine bacterioplankton assemblage. *The ISME Journal* **5,** 866–878 (2011).
2. Baker, B. J., Lesniewski, R. A. & Dick, G. J. Genome-enabled transcriptomics reveals archaeal populations that drive nitrification in a deep-sea hydrothermal plume. *The ISME Journal* **6,** 2269–2279 (2012).
3. Gifford, S. M., Sharma, S., Booth, M. & Moran, M. A. Expression patterns reveal niche diversification in a marine microbial assemblage. *The ISME Journal* **7,** 281–298 (2013).
4. Stewart, F. J., Ulloa, O. & DeLong, E. F. Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environmental Microbiology* **14,** 23–40 (2012).
5. Karner, M. B., Delong, E. F. & Karl, D. M. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409,** 507–510 (2001).
6. Schattenhofer, M. *et al.* Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environmental Microbiology* **11,** 2078–2093 (2009).
7. Francis, C. A., Roberts, K. J., Beman, J. M., Santoro, A. E. & Oakley, B. B. Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 14683–14688 (2005).
8. Coolen, M. J. L. *et al.* Putative ammonia-oxidizing Crenarchaeota in suboxic waters of the Black Sea: a basin-wide ecological study using 16S ribosomal and functional genes and membrane lipids. *Environmental Microbiology* **9,** 1001–1016 (2007).
9. Lam, P. *et al.* Revising the nitrogen cycle in the Peruvian oxygen minimum zone. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 4752–4757 (2009).
10. Pitcher, A. *et al.* Niche segregation of ammonia-oxidizing archaea and anammox bacteria in the Arabian Sea oxygen minimum zone. *The ISME Journal* **5,** 1896–1904 (2011).
11. Beman, J. M., Popp, B. N. & Alford, S. E. Quantification of ammonia oxidation rates and ammonia-oxidizing archaea and bacteria at high resolution in the Gulf of California and eastern tropical North Pacific Ocean. *Limnol. Oceangr.* **57,** 711–726 (2012).
12. Nakagawa, T. & Stahl, D. A. Transcriptional response of the archaeal ammonia oxidizer *Nitrosopumilus maritimus* to low and

environmentally relevant ammonia concentrations. *Applied and Environmental Microbiology* **79,** 6911–6916 (2013).

13. Qin, W. *et al.* Marine ammonia-oxidizing archaeal isolates display obligate mixotrophy and wide ecotypic variation. *Proc. Natl. Acad. Sci. U.S.A.* **111,** 12504–12509 (2014).
14. Bayer, B. *et al.* Physiological and genomic characterization of two novel marine thaumarchaeal strains indicates niche differentiation. *The ISME Journal* **10,** 1051–1063 (2015).
15. Alonso-Sáez, L. *et al.* Role for urea in nitrification by polar marine Archaea. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 17989–17994 (2012).
16. Tolar, B. B. *et al.* Contribution of ammonia oxidation to chemoautotrophy in Antarctic coastal waters. *The ISME Journal* **10,** 2605–2619 (2016).
17. Pedneault, E., Galand, P. E., Potvin, M., Tremblay, J.-É. & Lovejoy, C. Archaeal *amoA* and *ureC* genes and their transcriptional activity in the Arctic Ocean. *Sci. Rep.* **4,** 1–11 (2014).
18. Shi, Y., Tyson, G. W., Eppley, J. M. & DeLong, E. F. Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *The ISME Journal* **5,** 999–1013 (2010).
19. Vajrala, N. *et al.* Hydroxylamine as an intermediate in ammonia oxidation by globally abundant marine archaea. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 1006–1011 (2013).
20. Walker, C. B. *et al. Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 8818–8823 (2010).
21. Arp, D. J., Chain, P. S. G. & Klotz, M. G. The impact of genome analyses on our understanding of ammonia-oxidizing bacteria. *Annu. Rev. Microbiol.* **61,** 503–528 (2007).
22. Hallam, S. J. *et al.* Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 18296–18301 (2006).
23. Stahl, D. A. & la Torre, de, J. R. Physiology and diversity of ammonia-oxidizing archaea. *Annu. Rev. Microbiol.* **66,** 83–101 (2012).
24. Santoro, A. E. *et al.* Genomic and proteomic characterization of '*Candidatus* Nitrosopelagicus brevis': an ammonia-oxidizing archaeon from the open ocean. *Proc. Natl. Acad. Sci. U.S.A.* **112,** 1173–1178 (2015).
25. Kerou, M. *et al.* Proteomics and comparative genomics of

Nitrososphaera viennensis reveal the core genome and adaptations of archaeal ammonia oxidizers. *Proc. Natl. Acad. Sci. U.S.A.* 201601212–19 (2016). doi:10.1073/pnas.1601212113

26. Martens-Habbena, W. *et al.* The production of nitric oxide by marine ammonia-oxidizing archaea and inhibition of archaeal ammonia oxidation by a nitric oxide scavenger. *Environmental Microbiology* **17,** 2261–2274 (2015).
27. Stieglmeier, M. *et al.* Aerobic nitrous oxide production through N-nitrosating hybrid formation in ammonia-oxidizing archaea. *The ISME Journal* **8,** 1135–1146 (2014).
28. Kozlowski, J. A., Stieglmeier, M., Schleper, C., Klotz, M. G. & Stein, L. Y. Pathways and key intermediates required for obligate aerobic ammonia-dependent chemolithotrophy in bacteria and Thaumarchaeota. *The ISME Journal* **10,** 1836–1845 (2016).
29. Santoro, A. E. & Casciotti, K. L. Enrichment and characterization of ammonia-oxidizing archaea from the open ocean: phylogeny, physiology and stable isotope fractionation. *The ISME Journal* **5,** 1796–1808 (2011).
30. Tolar, B. B., Wallsgrove, N. J., Popp, B. N. & Hollibaugh, J. T. Oxidation of urea-derived nitrogen by thaumarchaeota-dominated marine nitrifying communities. *Environmental Microbiology* (2016). doi:10.1111/1462-2920.13457
31. Smith, J. M., Casciotti, K. L., Chavez, F. P. & Francis, C. A. Differential contributions of archaeal ammonia oxidizer ecotypes to nitrification in coastal surface waters. *The ISME Journal* **8,** 1704–1714 (2014).
32. Li, D.-C., Yang, F., Lu, B., Chen, D.-F. & Yang, W.-J. Thermotolerance and molecular chaperone function of the small heat shock protein HSP20 from hyperthermophilic archaeon, *Sulfolobus solfataricus P2*. *Cell Stress and Chaperones* **17,** 103–108 (2011).
33. Wei, X. *et al.* Transcript profiles of *Nitrosomonas europaea* during growth and upon deprivation of ammonia and carbonate. *FEMS Microbiology Letters* **257,** 76–83 (2006).
34. Berube, P. M. & Stahl, D. A. The divergent AmoC3 subunit of ammonia monooxygenase functions as part of a stress response system in *Nitrosomonas europaea*. *Journal of Bacteriology* **194,** 3448–3456 (2012).
35. Smith, D. P. *et al.* Proteomic and transcriptomic analyses of '*Candidatus* Pelagibacter ubique' describe the first PII-independent response to nitrogen limitation in a free-living Alphaproteobacterium.

mBio **4,** e00133–12 (2013).

36. Giovannoni, S. J., Cameron Thrash, J. & Temperton, B. Implications of streamlining theory for microbial ecology. *The ISME Journal* **8,** 1553–1565 (2014).
37. Cottrell, M. T. & Kirchman, D. L. Transcriptional control in marine copiotrophic and oligotrophic bacteria with streamlined genomes. *Applied and Environmental Microbiology* **82,** 6010–6018 (2016).
38. Lindås, A.-C., Karlsson, E. A., Lindgren, M. T., Ettema, T. J. G. & Bernander, R. A unique cell division machinery in the Archaea. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 18942–18946 (2008).
39. Spang, A. *et al.* Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521,** 173–179 (2015).
40. Lam, P., Cowen, J. P. & Jones, R. D. Autotrophic ammonia oxidation in a deep-sea hydrothermal plume. *FEMS Microbiology Ecology* **47,** 191–206 (2004).
41. Church, M. J., Wai, B., Karl, D. M. & DeLong, E. F. Abundances of crenarchaeal *amoA* genes and transcripts in the Pacific Ocean. *Environmental Microbiology* **12,** 679–688 (2010).
42. Santoro, A. E., Buchwald, C., McIlvin, M. R. & Casciotti, K. L. Isotopic signature of N_2O produced by marine ammonia-oxidizing archaea. *Science* **333,** 1282–1285 (2011).
43. Buchwald, C., Santoro, A. E., McIlvin, M. R. & Casciotti, K. L. Oxygen isotopic composition of nitrate and nitrite produced by nitrifying cocultures and natural marine assemblages. *Limnol. Oceangr.* **57,** 1361–1375 (2012).
44. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30,** 1236–1240 (2014).
45. Kiley, P. J. & Beinert, H. The role of Fe–S proteins in sensing and regulation in bacteria. *Current Opinion in Microbiology* **6,** 181–185 (2003).
46. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10,** 845–858 (2015).
47. Strickland, J. D. H. & Parsons, T. R. *A Practical Handbook of Seawater Analysis. 2.ed*. (1972).
48. Tripp, H. J. Counting marine microbes with Guava Easy-Cyte 96 well plate reading flow cytometer. *Protocol Exchange* (2008). doi:10.1038/nprot.2008.29
49. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome*

Research **27,** 824–834 (2017).
50. Laczny, C. C. *et al.* VizBin - an application for reference-independent visualization and human-augmented binning of metagenomic data. *Microbiome* **3,** 1–7 (2015).
51. Santoro, A. E., Casciotti, K. L. & Francis, C. A. Activity, abundance and diversity of nitrifying archaea and bacteria in the central California Current. *Environmental Microbiology* **12,** 1989–2006 (2010).
52. Price, N. M. & Harrison, P. J. Comparison of methods for the analysis of dissolved urea in seawater. *Marine Biology* **94,** 307–317 (1987).
53. Robinson, M. D. & Smyth, G. K. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* **9,** 321–332 (2008).
54. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Meth* **5,** 621–628 (2008).
55. Harrell, F. E., Jr & Dupont, C. Hmisc: Harrell Miscellaneous.
56. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U.S.A.* **100,** 9440–9445 (2003).
57. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **10,** 10008– (2008).
58. Goris, J. *et al.* DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* **57,** 81–91 (2007).
59. Hosseinzadeh, P. *et al.* A purple cupredoxin from *Nitrosopumilus maritimus* containing a mononuclear yype 1 copper center with an open binding site. *J. Am. Chem. Soc.* **138,** 6324–6327 (2016).
60. Ritchie, G. & Nicholas, D. Identification of the sources of nitrous oxide produced by oxidative and reductive processes in Nitrosomonas europaea. *Biochem. J.* (1972).
61. Ritchie, G. A. & Nicholas, D. J. The partial characterization of purified nitrite reductase and hydroxylamine oxidase from Nitrosomonas europaea. *Biochem. J.* **138,** 471–480 (1974).
62. Wijma, H. J., Canters, G. W., de Vries, S. & Verbeet, M. P. Bidirectional catalysis by copper-containing nitrite reductase. *Biochemistry* **43,** 10467–10474 (2004).
63. Caranto, J. D. & Lancaster, K. M. Nitric oxide is an obligate bacterial nitrification intermediate produced by hydroxylamine oxidoreductase. *Proc. Natl. Acad. Sci. U.S.A.* **Vol 16,** 201704504–13 (2017).

718
719

Figures:

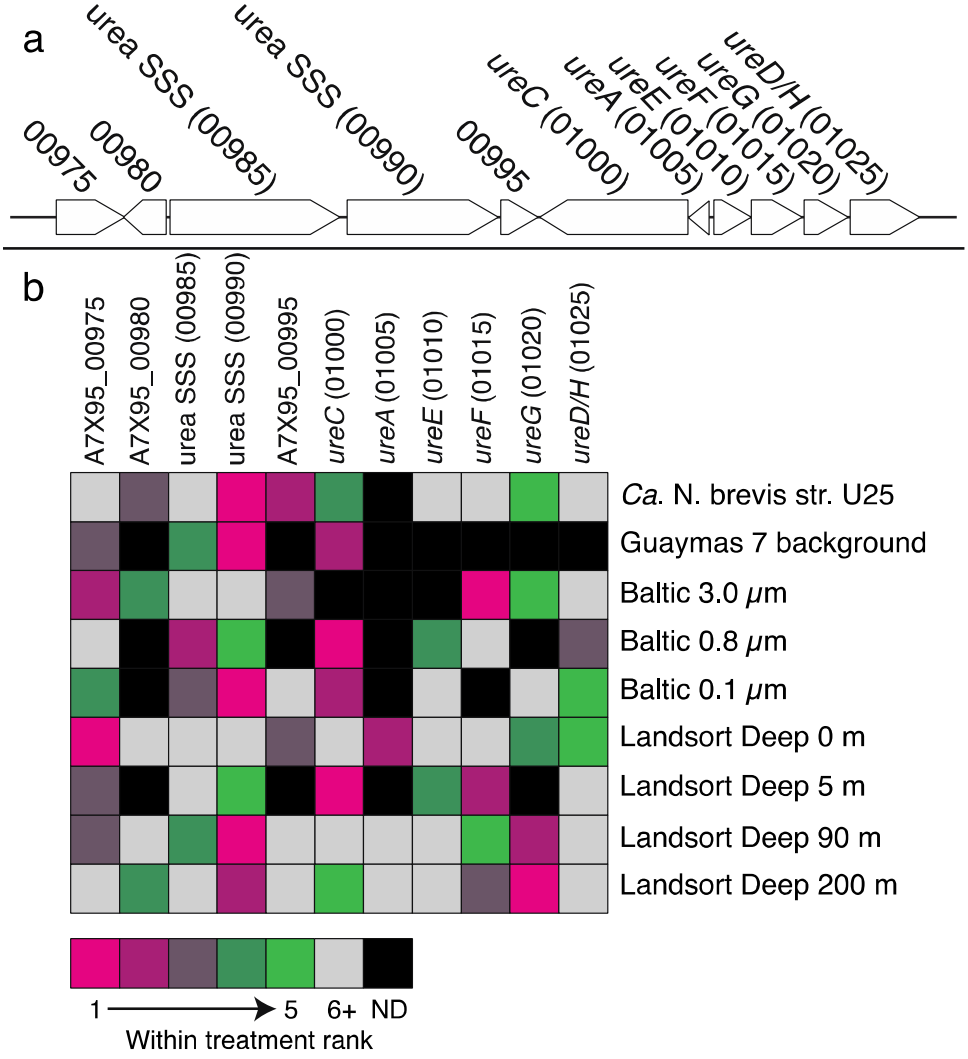


Figure 1: Urea transport and catalysis genes are frequently the most highly expressed *Ca*. N. brevis U25-specific genes. (a) Chromosomal orientation of the *Ca*. N. brevis str. U25 indel conferring urea transport and catalysis capability. (b) Heatmap illustrating the relative rank expression level of the 5 most abundant genes in the urea indel region for culture experiments and environmental metatranscriptomes. Rank was calculated within a given treatment from RPKM normalized expression values, where a rank of 1 is the most abundant transcript. The mean expression (n=3 replicates) value was used for ranking the culture treatment. ND=Not detected. Some metatranscriptomes were excluded because they did not have sufficient coverage of the urea transport and catabolism machinery. Numbers in parentheses after gene names refer to A7X95 locus tags.

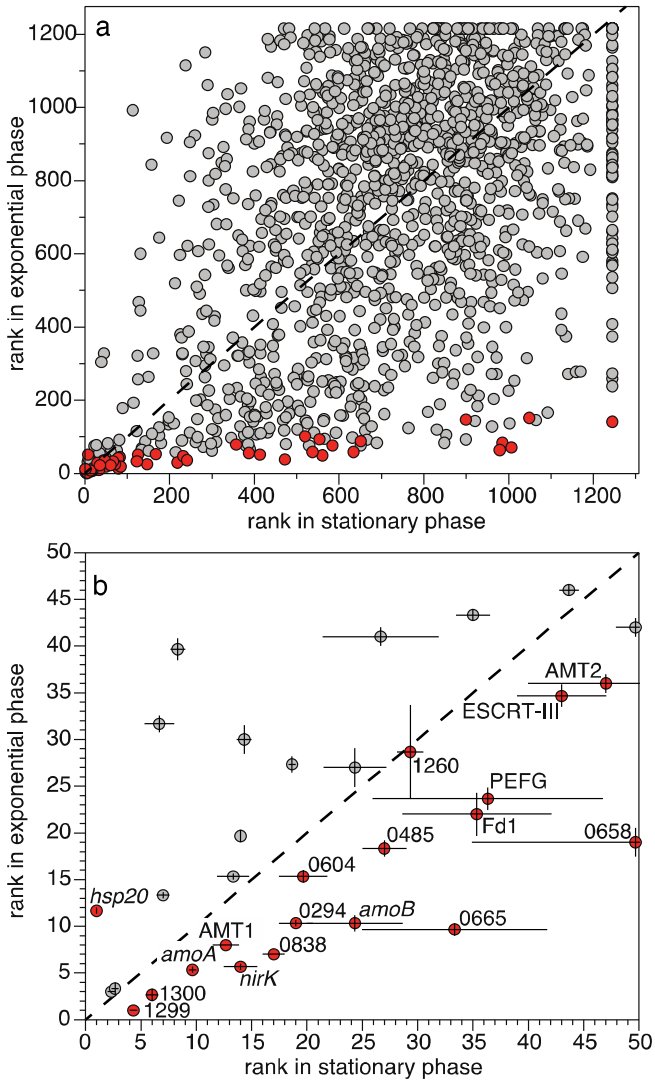
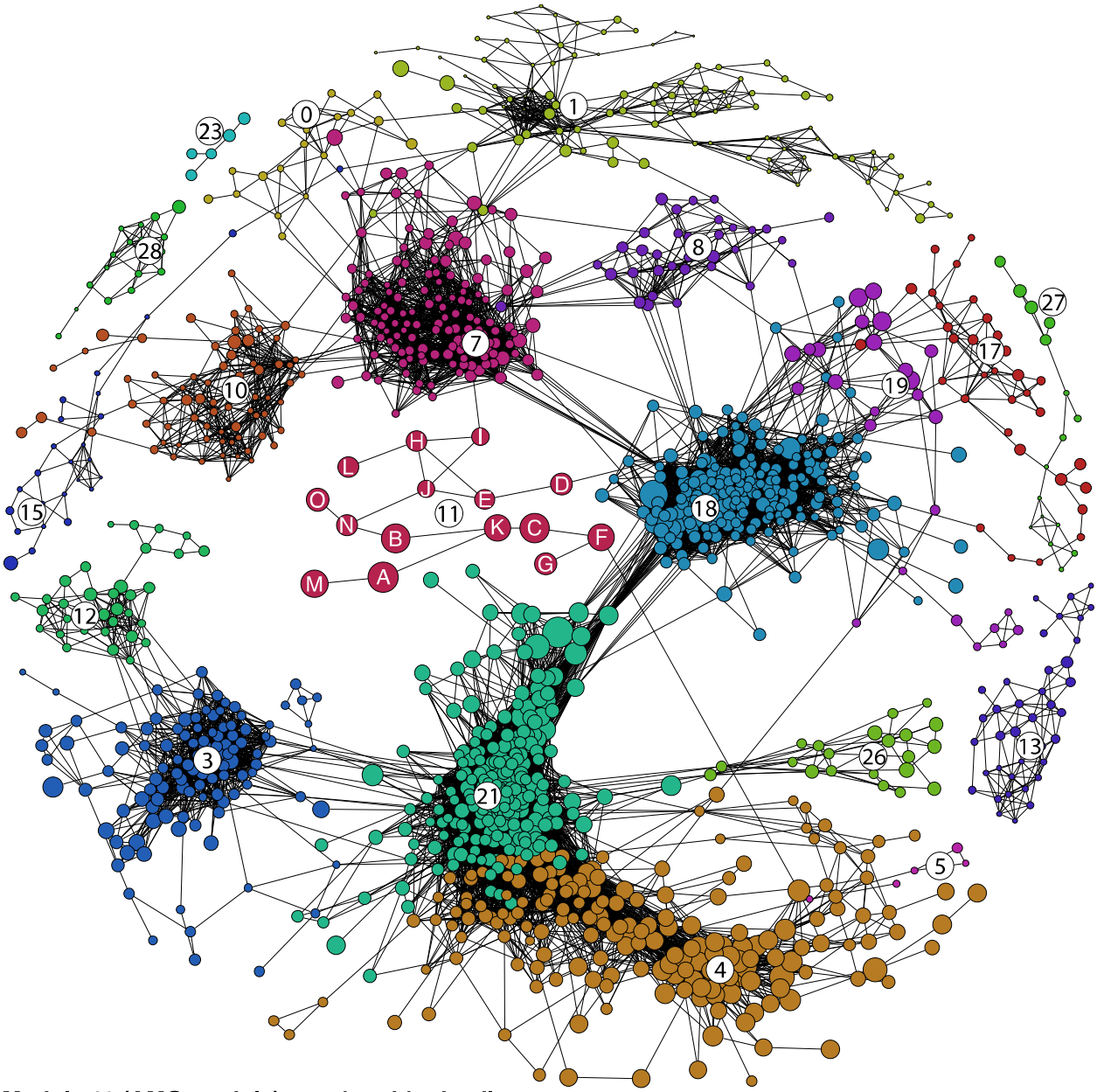


Figure 2: Highly expressed *Ca*. N. brevis str. CN25 gene transcripts in exponential phase are also highly expressed in stationary phase, despite significant differences in abundance. (a) Points are the mean rank (n=3) of RPKM normalized expression values for all genes in exponential and stationary growth phases. Red colored genes were significantly differentially transcribed across treatments (Supplementary Table 2). The abundances of grey points were not significantly different across treatments. Dashed line is 1:1 line indicating no change in rank. (b) Subset of panel (a), illustrating the rank of transcripts that are in the top 50 most abundant transcripts in both exponential and stationary phase. Points in (b) are the mean rank and error bars represent ± SE (n=3). Gene transcript abundances in (b) that were significantly different across treatments are labeled and colored red. The 'T478_' prefix is omitted from labels of genes annotated as 'hypothetical'. PEFG corresponds to T478_0596.



Module 11 (AMO module) membership details

A: *amoA* (0302)	I: ESCRT-III (0057)
B: *amoB* (0298)	J: PEFG-CTERM (0270)
C: *amoC* (0300)	K: *nirK* (1026)
D: Membrane-bound cupredoxin (0895)	L: *amoX* (0301)
E: 0487	M: AMT1 (1378)
F: Fd1 (1472)	N: Cupredoxin PEFG-CTERM (1362)
G: Fd2 (1259)	O: Fe-S cluster assembly protien (1056)
H: 1317	

Figure 3: Thaumarchaeal gene expression is highly modular and ammonia oxidation genes are co-expressed. Network diagram of strong and significant (Pearson's r ≥ 0.8, *q* value ≤ 0.025) positive correlations across ten environmental metatranscriptomes and two transcriptomes from

laboratory cultures (12 conditions total). Individual nodes are genes. Nodes are sized by the mean normalized rank abundance (VdW scores), whereby larger nodes are more abundant transcripts. Nodes are colored by module membership. Circled numbers are the module identity (Supplementary Table 3). The module 11 (the AMO module) genes are identified by letters A-O; their annotations are provided below network. Numbers in parentheses refer to T478 locus tags. Only modules with five or more nodes are shown for clarity; see Supplementary Table 3 for all module membership.

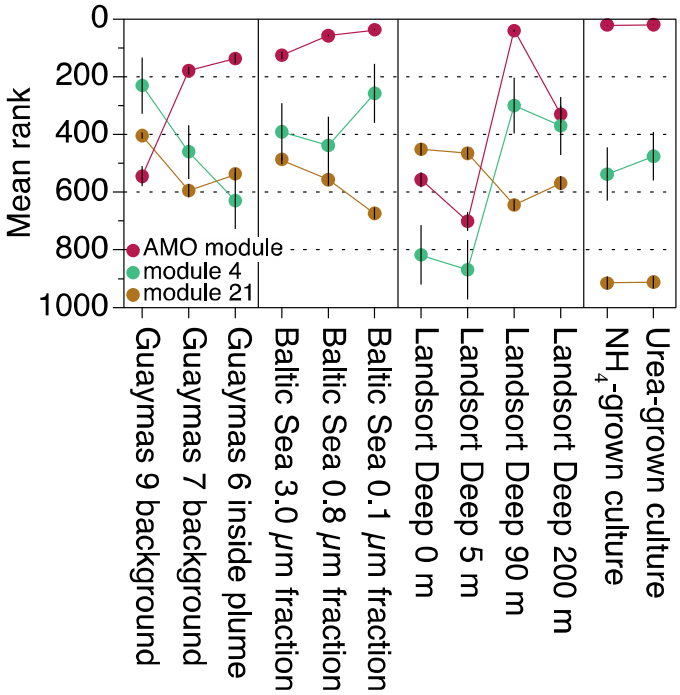
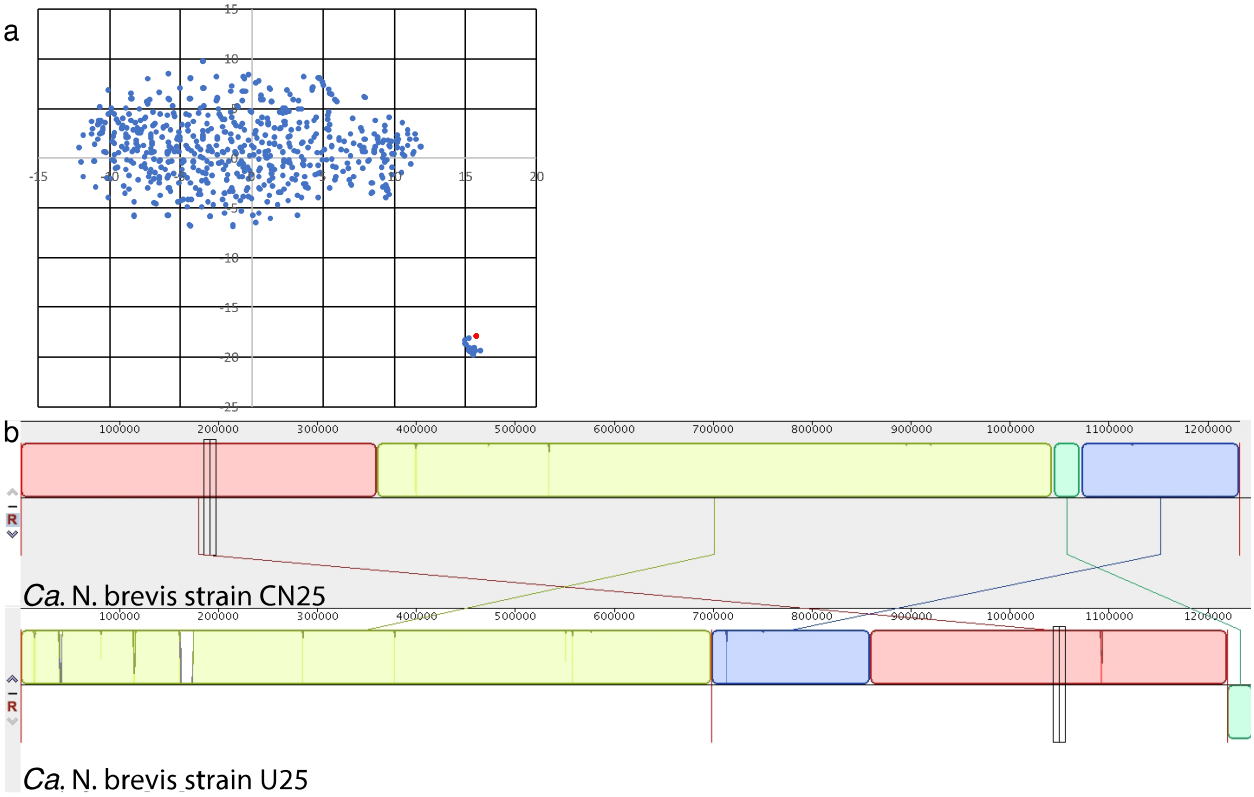


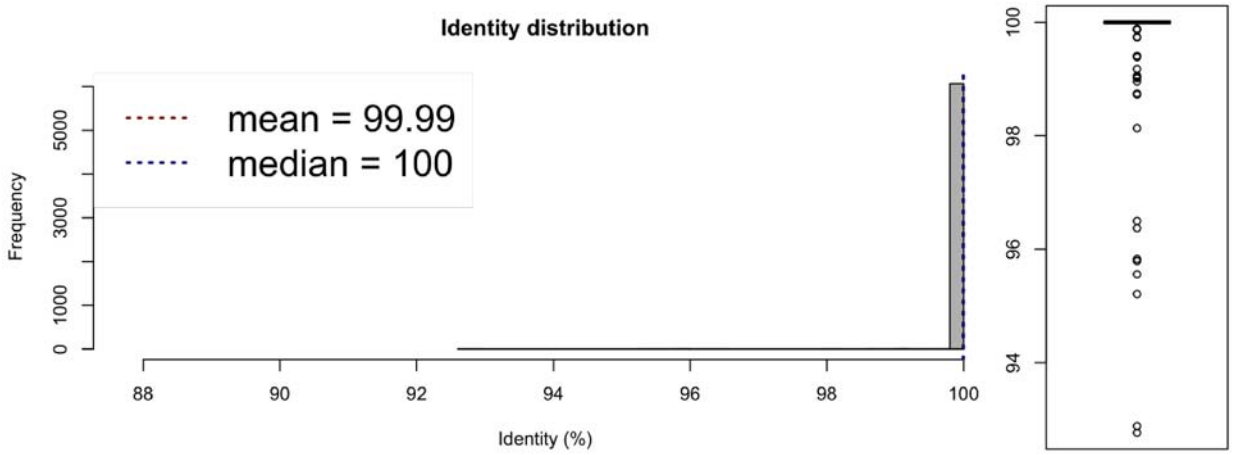
Fig. 4: Natural variability in the mean rank of expression modules across environments and culture conditions. Points are the average rank of all genes contained within a given module at each site; error bars are ± SE. On average, the AMO module and modules 36, 4 and 21 rank the highest (that is, they are the most abundant) across all sites. Module 36 is not displayed because it is only comprised of two genes, which prevents statistical analysis, and these genes are absent in the U25 genome.

Supplemental Information for: Correlated expression of archaeal ammonia oxidation machinery across disparate environmental and culture conditions

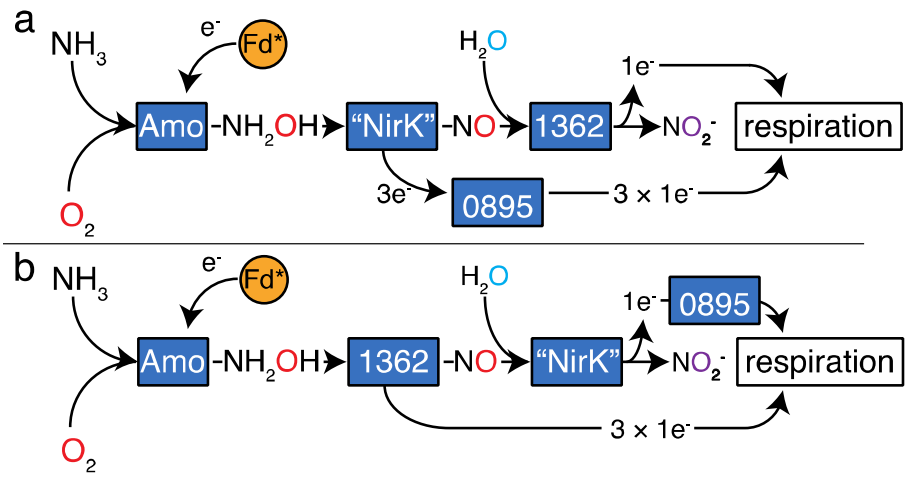
Paul Carini[1,3], Christopher L. Dupont[2], Alyson E. Santoro[1,4]



Supplementary Figure 1: *Ca*. N. brevis str. U25 metagenome binning and alignment to strain *Ca*. N. brevis str. CN25. (a) Vizbin analysis of 5 mer utilization of the assembly from an thaumarchaea enrichment culture maintained on urea. Contigs containing any thaumarchaea predicted genes are in red and code for the *Ca*. N. brevis str. U25 genome. (b) Mauve alignment of *Ca*. N. brevis strains CN25 (top) and U25 (bottom). The genomes are syntenic with only a few indels between the two strains (listed in Supplementary Table 1) (illustrated as white vertical bands). Red vertical lines in the U25 genome in (b) are contig breaks.

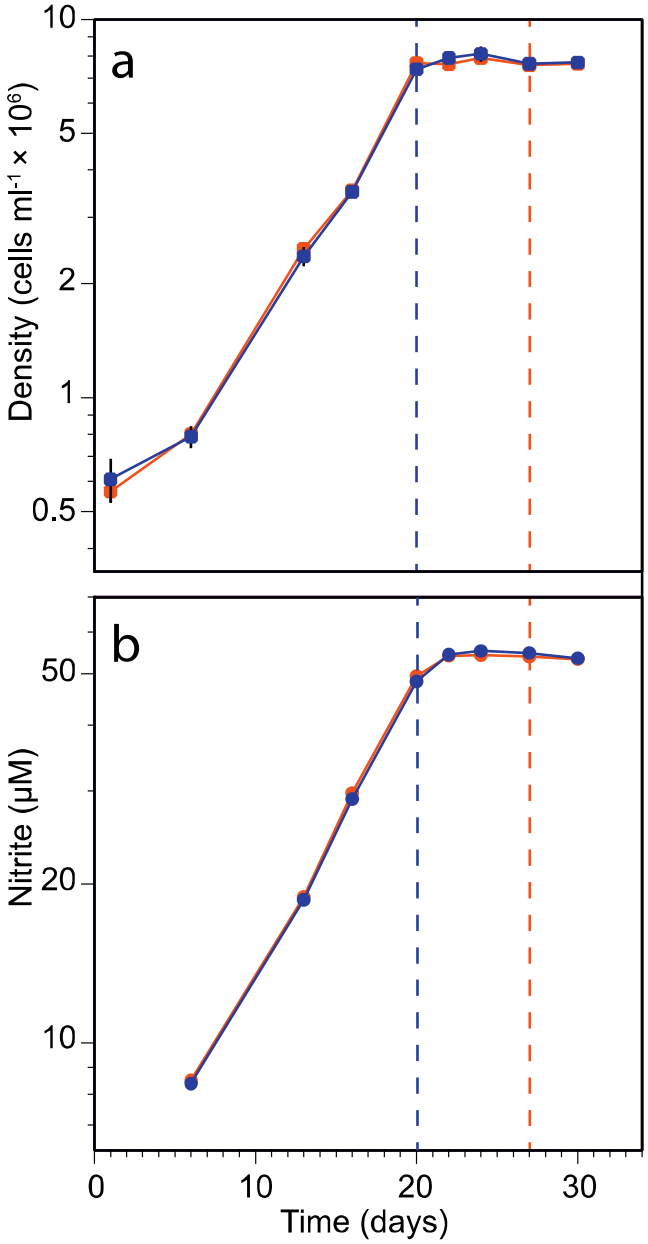


Supplementary Figure 2: Simulated DNA-DNA hybridization analysis for the CN25 and U25 genomes using the methods of ref[58]. The mean nucleotide identity of 99.99% is for a two-way average nucleotide identity analysis. The most divergent conserved genomic segment is 92% nucleotide identity.

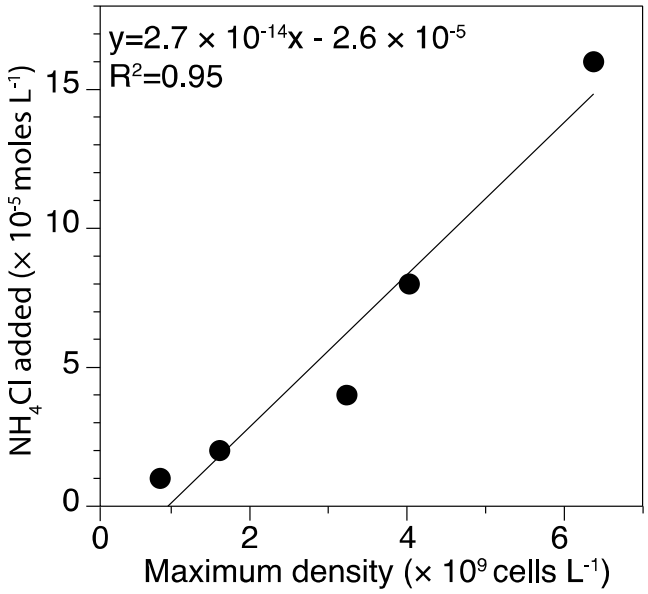


Supplementary Figure 3: Proposed ammonia (NH_3) oxidation pathways for thaumarchaea based on co-expression of pseudoperiplasmic-localized Cu metalloenzymes. Amo, T478_1362, Fd1, Fd2, "NirK" and T478_0895 are co-expressed in a conserved module (the AMO module) across different environments as illustrated in Fig. 3 in the main text. In these models, a membrane-anchored PEFG-CTERM domain-containing Cu metalloenzyme (T478_1362) acts in concert with "NirK" and T478_0895 to catalyze a two-step reaction: the 3-electron oxidation of hydroxylamine (NH_2OH) to nitric oxide (NO) followed by the one-electron oxidation of NO to NO_2^-. In (a) "NirK" acts as the hydroxylamine oxidase. The model proposed in (a) is supported by sequence structure threading of T478_1362, which predicts the cupredoxin domain has an open configuration, as in a recently characterized purple cupredoxin from *N. maritimus* that oxidizes NO to NO_2^-[59]. Additionally, the model in (a) is consistent with the potential for NirK to carry three electrons by way of three Cu atoms[20]; and the potential for NH_2OH to be oxidized to NO by an uncharacterized nitrite reductase [60,61]. In the model depicted in (b), T478_1362 acts as the hydroxylamine oxidase. Model (b) is consistent with the ability of purified bacterial NirK to favor the formation of NO_2^- from NO and water at biological pH[62]. Moreover, NirK may catalyze the final step in a three-step ammonia oxidation pathway in ammonia-oxidizing bacteria (AOB)[63]. In both model (a) and (b), electrons are putatively shuttled from NirK to respiratory complexes by membrane-anchored cupredoxin-containing T478_0895. Two ferredoxins (Fd* = Fd1 and Fd2) were also co-expressed with the core ammonia oxidation machinery. These ferredoxins are predicted to be cytoplasmic and may play a role in supplying the electrons required for the initial oxidation of NH_3 to NH_2OH by Amo (see main text). Enzyme complexes colored blue are Cu metalloenzymes.

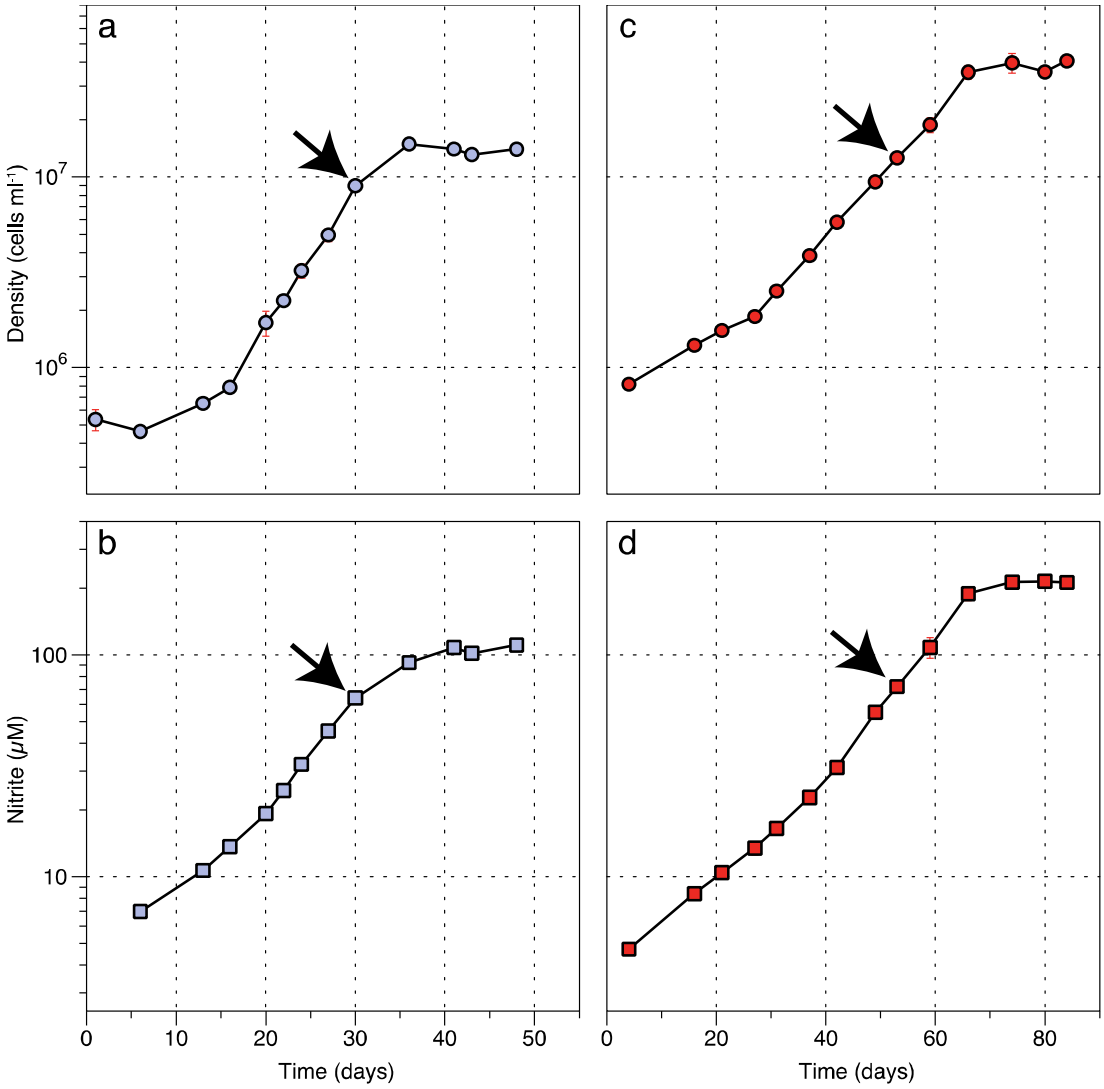
829 Enzymes colored orange are iron-containing. Color of the oxygen atom
830 depicts the source: red = molecular oxygen; Cyan=water; purple=one atom
831 from oxygen and one atom from water. Numbers are locus tags with
832 'T478_' omitted.
833



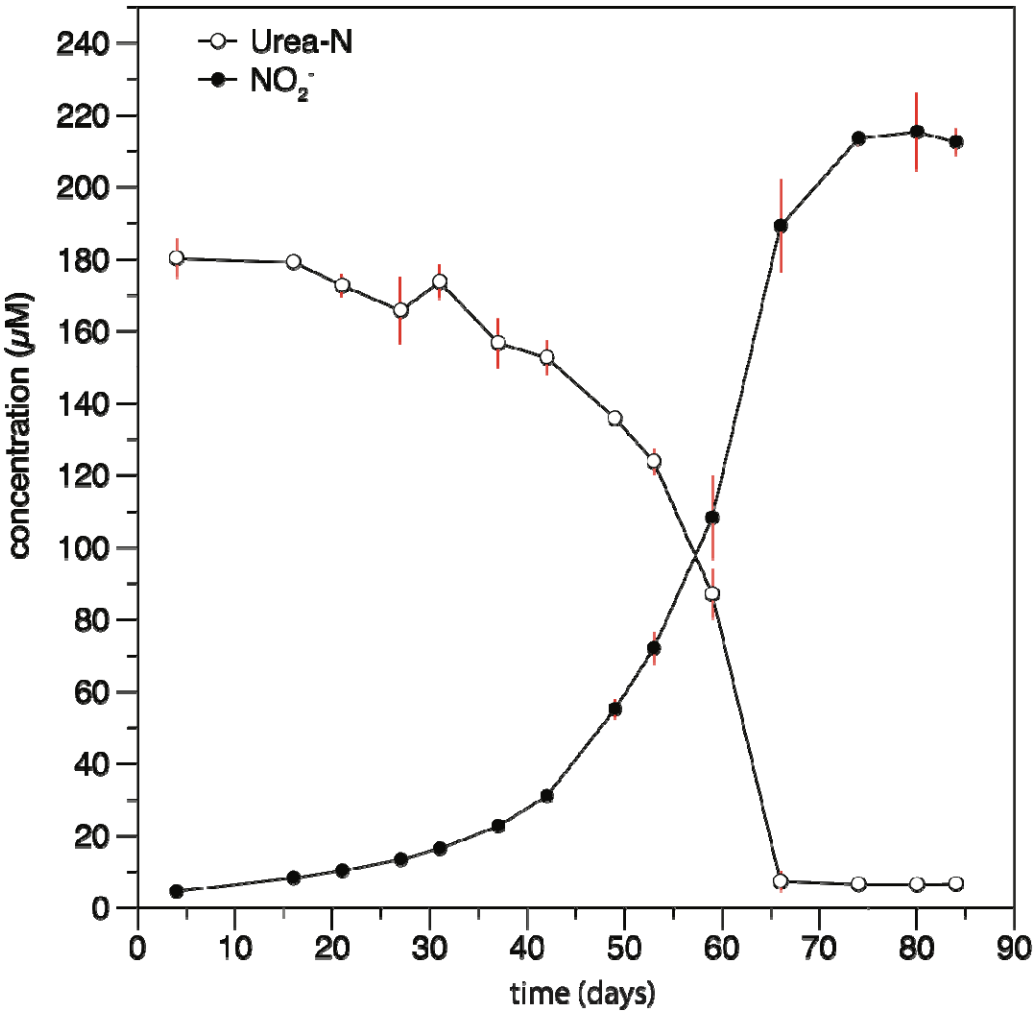
Supplementary Fig. 4: *Ca*. N. brevis CN25 growth curves illustrating the average cell density (a) and NO_2^- concentration (b) for cultures harvested for transcriptomes used to investigate the effects of ammonia limitation. Dashed vertical lines are sample time points for exponential phase (blue line) and ammonium-limited stationary phase (orange line). Points are the mean of biological triplicates ± SD. When error bars are not visible, they are smaller than the size of the symbols. Cells were grown in ONP medium[29] with 50 µM NH_4Cl, as described in materials and methods.



Supplementary Fig. 5: *Ca*. N. brevis CN25 molar growth yield in response to ammonium (NH_4^+) additions, illustrating that NH_4^+ was limiting growth in the transcriptome experiments. Points are the maximum density achieved by *Ca*. N. brevis str. CN25 as a function of NH_4Cl additions. Linear regression through all five points is shown, with equation and R^2. Cells were grown in ONP medium[29] as described in materials and methods.



Supplementary Fig. 6: *Ca*. N. brevis CN25 (a,b) and U25 (c,d) growth curves illustrating the average cell density (a,c) and NO_2^- concentration (b,d) for cultures harvested for transcriptomes. Arrowheads indicate the cell density and nitrite concentration at the time of sampling. Points are the mean of biological triplicates ± SD. When error bars are not visible, they are smaller than the size of the symbols. Cells were grown in ONP medium[29] with 100 μM NH_4Cl (a,b) or 100 μM urea (c,d), as described in materials and methods.



Supplementary Figure 7: *Ca*. N. brevis strain U25 oxidizes urea-N to NO_2^-. Points are the average ± SD (error bars) concentration of urea-N (open circles) and NO_2^- (filled circles) of triplicate *Ca*. N. brevis str. U25 cultures. Cells were grown in ONP medium[29] with 100 µM urea, as described in materials and methods.

Acknowledgements:

This research was supported by NSF awards OCE-1260006, OCE-1437310, and DBI-1318455 to AES. CLD was supported by NSF OCE-1259994. We thank Matt Rawls for urea measurements, Mike Stukel for obtaining oligotrophic seawater, and Albert Barberán and Jason Corwin for discussions regarding network construction.