Cultivation and genomic analysis of *Candidatus* Nitrosocaldus islandicus, a novel obligately thermophilic ammonia-oxidizing *Thaumarchaeon*

Anne Daebeler[1], Craig Herbold[1], Julia Vierheilig[1], Christopher J. Sedlacek[1], Petra Pjevac[1], Mads Albertsen[2], Rasmus H. Kirkegaard[2], José R. de la Torre[3], Holger Daims[1], Michael Wagner[1]

[1]Division of Microbial Ecology, Department of Microbiology and Ecosystem Science, Research Network "Chemistry meets Microbiology", University of Vienna, Austria
[2]Center for Microbial Communities, Department of Chemistry and Bioscience, Aalborg University, Denmark
[3]Department of Biology, San Francisco State University, San Francisco, CA, USA

Running title: Cultivation and genomics of a novel *Nitrosocaldus* sp.

Keywords: AOA, thaumarchaea, thermophile, nitrification, hot spring, nirK, polymerase, anaerobic metabolism

Corresponding authors:
Anne Daebeler, University of Vienna, Division of Microbial Ecology, Althanstrasse 14, 1090 Vienna, Austria. Tel: +43 1 4277 76603, Email: daebeler@microbial-ecology.net

Michael Wagner, University of Vienna, Division of Microbial Ecology, Althanstrasse 14, 1090 Vienna, Austria. Tel: +43 1 4277 76600, Email: wagner@microbial-ecology.net

Abstract

Ammonia-oxidizing archaea (AOA) within the phylum *Thaumarchaea* are the only known aerobic ammonia oxidizers in geothermal environments. Although molecular data indicate the presence of phylogenetically diverse AOA from the *Nitrosocaldus* clade, group 1.1b and group 1.1a *Thaumarchaea* in terrestrial high-temperature habitats, only one enrichment culture of an AOA thriving above 50 °C has been reported and functionally analyzed. In this study, we physiologically and genomically characterized a novel *Thaumarchaeon* from the deep-branching *Nitrosocaldaceae* family of which we have obtained a high (~85 %) enrichment from biofilm of an Icelandic hot spring (73 °C). This AOA, which we provisionally refer to as "*Candidatus* Nitrosocaldus islandicus", is an obligately thermophilic, aerobic chemolithoautotrophic ammonia oxidizer, which stoichiometrically converts ammonia to nitrite at temperatures between 50 °C and 70 °C. *Ca.* N. islandicus encodes the expected repertoire of enzymes proposed to be required for archaeal ammonia oxidation, but unexpectedly lacks a *nirK* gene and also possesses no identifiable other enzyme for nitric oxide (NO) generation. Nevertheless, ammonia oxidation by this AOA appears to be NO-dependent as *Ca.* N. islandicus is, like all other tested AOA, inhibited by the addition of an NO scavenger. Furthermore, comparative genomics revealed that *Ca.* N. islandicus has the potential for aromatic amino acid fermentation as its genome encodes an indolepyruvate oxidoreductase (*iorAB*) as well as a type 3b hydrogenase, which are not present in any other sequenced AOA. A further surprising genomic feature of this thermophilic ammonia oxidizer is the absence of DNA polymerase D genes – one of the predominant replicative DNA polymerases in all other ammonia-oxidizing *Thaumarchaea*. Collectively, our findings suggest that metabolic versatility and DNA replication might differ substantially between obligately thermophilic and other AOA.

Introduction

Thaumarchaea (Brochier-Armanet *et al.,* 2008) are among the most abundant archaeal organisms on Earth, and thrive in most oxic environments (Francis *et al.,* 2007; Erguder *et al.*, 2009; Schleper and Nicol, 2010; Bouskill et al. 2012; Prosser and Nicol, 2012; Stahl and de la Torre, 2012; Stieglmeier *et al,* 2014a), but have also been detected in anoxic systems (Molina *et al.,* 2010; Bouskill *et al.,* 2012; Buckles *et al.,* 2013; Beam *et al.,* 2014; Lin *et al.,* 2015). This phylum comprises ammonia-oxidizing archaea (AOA) and other archaeal taxa in which ammonia oxidation has not been demonstrated. All cultured members of the *Thaumarchaea* are AOA and grow by using ammonia, urea or cyanate as substrate (Palatinszky *et al.* 2015; Bayer *et al.* 2016; Sauder *et al.* 2017; Qin *et al.* 2017a), although *in situ* experiments suggest that certain members of this phylum capable of ammonia oxidation also possess other lifestyles (Mußmann *et al.* 2011; Sauder *et al.* 2017). In aquatic and terrestrial environments *Thaumarchaea* often co-occur with ammonia-oxidizing bacteria (AOB), and frequently outnumber them by orders of magnitude (Francis *et al.*, 2005; Leininger *et al.*, 2006; Mincer *et al.,* 2007; Adair and Schwarz, 2008; Abell *et al.,* 2010; Mußmann *et al.,* 2011; Zeglin *et al.,* 2011; Daebeler *et al.*, 2012). *Thaumarchaea* also inhabit extreme environments like terrestrial hot springs and other high temperature habitats, where AOB are not detectable (Weidler *et al.*, 2007; Reigstad *et al.*, 2008; Wang *et al.*, 2009; Zhao *et al.*, 2011; Chen *et al.*, 2016). In addition to the presence of *Thaumarchaea* in hot environments, high *in situ* nitrification rates (Reigstad *et al.*, 2008; Dodsworth *et al.*, 2011; Chen *et al.*, 2016) and transcription of genes involved in archaeal ammonia oxidation in several hot springs over 74 °C (Zhang *et al.*, 2008; Jiang *et al.*, 2010) support an important role of thermophilic AOA in these systems.

Despite their apparent importance for nitrogen cycling in a wide range of thermal habitats, only one thermophilic [on the basis of the definition by Stetter (1998) that thermophiles grow optimally above 50 °C] AOA species from an enrichment culture has been reported to date (de la Torre *et al.,* 2008; Qin *et al.*, 2017b) and was named *Candidatus (Ca.)* Nitrosocaldus yellowstonensis. In addition, several enrichment cultures and one pure culture of moderately thermophilic AOA, which are able to grow at 50 °C, but grow optimally only at temperatures below 50 °C, have been described (Hatzenpichler *et al.*, 2008; Lebedeva *et al.*, 2013; Palatinszky *et al.*, 2015). Therefore, our current knowledge on specific adaptations or metabolic capabilities of thermophilic AOA growing preferably at temperatures above 50 °C is very limited (Spang *et al.*, 2012).

In 16S rRNA and ammonia monooxygenase subunit A (*amoA*) gene trees *Ca.* Nitrosocaldus yellowstonensis branches most deeply among *Thaumarchaea* that possess ammonia monooxygenase (AMO) genes. In consequence, the *Nitrosocaldales* clade has been considered as being close to the evolutionary origin of *Thaumarchaea* encoding the genetic repertoire for ammonia oxidation (Spang *et al.* 2017, de la Torre *et al.*, 2008). However, since the genome sequence of *Ca.* N. yellowstonensis is not yet published, phylogenomic analysis to confirm an ancestral position of the *Nitrosocaldales* relative to other *Thaumarchaea* have been pending.

Here we report on the enrichment, phylogenomic analyses, and selected (putative) metabolic features of a novel, obligately thermophilic, AOA from the *Nitrosocaldales* clade obtained from a biofilm collected from an Icelandic hot (73 °C) spring. This organism, provisionally referred to as *Ca.* Nitrosocaldus islandicus, occupies a fundamentally different niche compared to other genomically characterized AOA as its ammonia-oxidizing activity is restricted to temperatures ranging from 50 °C to 70 °C.

Materials and Methods

Enrichment, cultivation, and physiological experiments

The enrichment of *Ca.* N. islandicus was initiated by inoculation of 40 ml sterile mineral medium (Koch *et al.*, 2015) containing 0.5 mM filter-sterilized NH_4Cl with approximately 0.1 g of hot spring biofilm, which had been submerged in running water at the sampling site in a geothermal area in Graendalur valley, (64° 1’ 7” N, 21° 11’ 20” W) Iceland. At the sampling site, the spring had

a pH of 6.5 and a temperature of 73 °C. The culture was initially incubated without agitation in 100 ml glass bottles in the dark at 60 °C and checked weekly for ammonium and nitrite content of the medium by using Nessler´s reagent (K_2HgI_4 – KOH solution; Sigma-Aldrich) and nitrite/nitrate test stripes (Merkoquant; Merck). Ammonium (1 mM NH_4Cl) was replenished when completely consumed. At the same time pH was monitored by using pH test stripes (Machery-Nagel) and kept at pH 7–8 by titration with $NaHCO_3$. When the pH dropped below 6 the enrichment culture ceased to oxidize ammonia, but activity was restored by readjusting the pH to between 7 and 8. The ammonium and nitrite concentrations were quantified photometrically (Kandeler and Gerber, 1988; Miranda *et al.*, 2001) using an Infinite 200 Pro spectrophotometer (Tecan Group AG). The microbial community composition of the enrichment was regularly monitored by fluorescence *in situ* hybridization (FISH) with 16S rRNA-targeted probes labeled with dyes Cy3, Cy5, or Fluos as described elsewhere (Daims *et al.*, 2005). Probes targeting most bacteria (EUB338 probe mix; Amann et al., 1990; Daims et al., 1999), most archaea (Arch915, Stahl and Amann, 1991) and most *Thaumarchaea* (Thaum726, Beam, 2015) were applied. All positive results were verified using the nonsense probe nonEUB338 (Wallner *et al.* 1993) labeled with the same dyes. Treatments with the macrolide antibiotic spiramycin (15 mg l^{-1}), which partly retains its antibacterial activity at 60 °C (Zorraquinio *et al.,* 2011), were performed as described in Zhang *et al.* (2015) together with serial dilutions ranging from 10^{-5} to 10^{-8} to obtain a highly enriched (~ 85 %) AOA culture that was used for further characterization.

Growth rates of *Ca.* N. islandicus were determined across a range of incubation temperatures (50 °C to 70 °C). Triplicate cultures (25 ml) and negative controls (cultures not supplied with ammonium or inoculated with autoclaved biomass) were incubated for ten days in 100 ml Schott bottles without agitation in the dark at the respective temperature. Samples from these experiments were either stored at -20 °C for subsequent qPCR analyses (150 µl) or centrifuged (21,000 x *g*, 15min, 18 °C) to remove cells and the supernatant was stored at -20 °C for chemical analysis (600 µl). qPCR analysis with primers CrenamoA19F (Leininger *et al,* 2006) and CrenamoA616R (Tourna *et al.,* 2008) targeting the archaeal *amoA* gene was otherwise performed as described in Pjevac *et al.* (2017) before the genome sequence of *Ca.* N. islandicus was available. However, subsequent analysis demonstrated that the employed qPCR primers contain mismatches to the *amoA* sequence of this AOA in the middle of the forward and reverse primer. The specific growth rate was calculated from log-linear plots of *amoA* gene abundance in individual cultures. In this analysis, three out of seven time points were interpolated through linear regression.

To test whether the NO-scavenger 2-phenyl-4,4,5,5,-tetramethylimidazoline-3-oxide-1-oxyl (PTIO; TCI, Germany) inhibits ammonia oxidation by *Ca.* N. islandicus, 40 ml aliquots of mineral medium containing 1 mM ammonium were inoculated with 10 % (v/v) of an exponential-phase culture and incubated in duplicates in the presence of 0, 33, and 100 µM PTIO, respectively. PTIO was dissolved in sterile mineral medium before addition to the cultures. The cultures not exposed to PTIO were supplemented with the same volume of sterile medium. The cultures were sampled (2 ml) at the beginning of the experiment and after 15 days of incubation. Nitrite and ammonium concentrations were measured as described above.

DNA extraction, genome sequencing, and annotation

DNA from three replicate enrichment cultures containing *Ca.* N. islandicus as the only detectable ammonia oxidizer was extracted as described by Angel and Conrad (2013) and sequenced by Illumina HiSeq next generation sequencing (250 bp paired end reads). Since we did not obtain a complete genome with this approach we re-extracted genomic DNA from the enrichment at a later stage according to Zhou *et al.* (1996) yielding high molecular weight DNA. Genomic DNA was then sheared in a Covaris g-TUBE (Covaris, USA) at 9000 RPM for 2x 1 min. in an Eppendorf mini spin plus centrifuge (Eppendorf, DE). The DNA was run on a E-Gel™ EX 1 % agarose gel (ThemoFisher, USA) and small DNA fragments were removed by excising a band with a length of

~8 kb. The DNA was purified from the gel cut using the ultraClean 15 DNA Purification Kit (Qiagen, USA). The DNA was prepared for sequencing using the "1D Low Input gDNA with PCR SQK-LSK108" protocol (Oxford Nanopore Technologies, UK) and sequenced on a FLO-MIN106 flowcell using the MinION MK1b (Oxford Nanopore Technologies, UK) following the manufacturers protocol using MinKNOW (v. 1.7.14). The nanopore reads were basecalled using Albacore (V. 2.0.1) (Oxford Nanopore Technologies, UK). The complete genome was assembled using a hybrid approach combining the data from the Illumina and nanopore sequencing with the hybrid assembler Unicycler (v. 0.4.1, Wick et al., 2017). The genome bins of the two contaminating organisms were assembled from the Nanopore reads using Miniasm (Li *et al.,* 2016) and polished twice with the Nanopore reads using Racon (Vaser *et al.,* 2017). No other microbe encoding genes indicative for ammonia-oxidation was identified in either of the two the metagenomes.

The complete genome of *Ca.* N. islandicus was uploaded to the MicroScope platform (Vallenet *et al.*, 2013) for automatic annotation, which was amended manually where necessary. The full genome sequence of *Ca.* N. islandicus has been deposited in GenBank (accession CP024014) and associated annotations are publicly available in MicroScope (*Candidatus* Nitrosocaldus islandicus strain 3F).

Protein-coding genes from the novel *Thaumarchaeon* were compared to those from 30 *Thaumarchaea* with available genomic data (Table S1) downloaded from NCBI. The coding sequences (CDS) with accession numbers from each genome, as downloaded from NCBI, were combined with additional CDS predictions made by Prodigal (Hyatt *et al.*, 2010) to account for variability in CDS predictions from different primary data providers and platforms. Predicted CDS from the novel *Thaumarchaeon* were aligned to CDS from reference genomes using blastp (Word_size=2, substitution matrix BLOSUM45). Genes were considered homologous only if the blastp alignment exceeded 50 % of the length of both query and subject sequences. CDS of *Ca.* N. islandicus that lacked any homologs in other *Thaumarchaea* were considered "unique". Unique CDS of unknown function were searched for secretion signals and for predicted membrane-spanning domains of the encoded proteins using the Phobius web server (Käll *et al.*, 2007) and putative structures were determined using the Phyre2 web server (Kelley *et al.*, 2015). Homology to "*Thaumarchaea*-core" proteins was assessed by cross-referencing the blastp homology search to the proteins defined for *Ca.* Nitrosotalea devanaterra by Herbold *et al.* (2017).

Phylogenetic analysis and habitat preference

For 16S rRNA and *amoA* gene-based phylogenetic analysis, the full-length 16S rRNA and *amoA* gene sequences of *Ca.* N. islandicus retrieved from the genome assembly were imported into the ARB software package (Ludwig *et al.*, 2004) together with other full length 16S rRNA or *amoA* gene sequences from cultivated AOA strains and aligned with the integrated ARB aligner with manual curation. 171 sequences from the *Aigarchaea* were included in the alignment and used as outgroup in the 16S rRNA gene phylogenetic analyses. For the *amoA* gene phylogenetic analyses no outgroup was selected. The 16S rRNA and *amoA* gene consensus trees were reconstructed using Maximum-Likelihood (ML; using the GTRGAMMA evolution model), Neighbour Joining (NJ) and Maximum Parsimony (MP) methods. For all calculations, a sequence filter considering only positions conserved in ≥50 % of all *thaumarchaeal* and *aigarchaeal* sequences was used, resulting in 2444 and 488 alignment positions for the 16S rRNA and *amoA* genes, respectively.

A Bayesian-inference phylogenomic tree was obtained using the automatically generated alignment of 34 concatenated universal marker genes (Table S2), which were identified by CheckM in Parks *et al.* (2015). This alignment was used as input for PhyloBayes (Lartillot *et al.*, 2009) with ten independent chains of 4,000 generations using the CAT-GTR model; 2,000 generations of each chain were discarded as burn-in, the remainder were subsampled every second tree (bpcomp -x 2000 2 4000) and pooled together for calculation of posterior probabilities.

Whole-genome based average nucleotide identity (gANI, Varghese *et al.*, 2015) and average amino acid identity values (AAI, Konstantinidis and Tiedje, 2005) were calculated between the genomes of *Ca.* N. islandicus and *Ca.* N. yellowstonensis using sets of annotated genes supplemented with additional gene calls predicted by Prodigal (Hyatt *et al.*, 2010). gANI was calculated using the Microbial Species Identifier (MiSI) method (Varghese *et al.,* 2015). For AAI, bidirectional best hits were identified using blastp, requiring that query genes aligned over at least 70 % of their length to target genes (in each unidirectional blastp search). Query gene length was used to calculate a weighted average % identity over all best hit pairs and the calculations were repeated using each genome as query and target.

The occurrence of organisms closely related to *Ca.* N. islandicus and *Ca.* N. yellowstonensis in publicly deposited amplicon sequencing data sets was assessed using IMNGS (Lagkouvardos *et al.*, 2016) with the full-length 16S rRNA gene sequences of both organisms as query and a nucleotide identity threshold of 97 %.

PolB amino acid sequences were extracted from the arCOG database (arCOG14 ftp://ftp.ncbi.nih.gov/pub/wolf/COGs/arCOG/ (arCOG15272, arCOG00329, arCOG00328, arCOG04926, arCOG15270). Additional *thaumarchaeal polB* sequences were identified using *Ca.* N. islandicus as a query in a blastp search against the nr protein database. These additional *thaumarchaeal* sequences, the *polB* sequence from *Ca.* N. islandicus and arCOG database sequences were de-replicated using usearch (Edgar, 2010) with -sortbylength and -cluster_smallmem (-id 0.99 -query_cov 0.9), aligned using default settings in mafft (Katoh and Standley, 2013) and a phylogenetic tree was calculated using FastTree (Price *et al.,* 2010).

Nitrilase superfamily amino acid sequences were obtained from Pace and Brenner (2001). Alignment and phylogenetic reconstruction was carried out with Bali-Phy (Suchard and Redelings, 2006; randomize alignment, iterations=11000, burnin=6000). Posterior tree pools from 10 independent runs were combined to generate a 50 % PP consensus tree and to assess bipartition support.

A dataset for assessing the phylogenetic relationship of the alpha subunit of 2-oxoacid:ferredoxin oxidoreductases (OFORs) was based on Gibson *et al.* (2016) and supplemented with additional indolepyruvate oxidoreductase (*ior*) sequences. Genomes available (as of October 30, 2017) from the NCBI genomes database were downloaded, genes were predicted using Prodigal V2.6.3 (Hyatt *et al.*, 2010) and predicted genes were screened for *iorA* (TIGRFAM03336) using hmmsearch v3.1b2 (hmmer.org) with an e-value cutoff of 0.001. Genes meeting the search criteria were used as queries against the complete TIGRFAM database to ensure that the extracted *iorA* sequences matched the *iorA* model as the best-hit model with an e-value cutoff of 0.001. Reciprocal best-hit genes were required to align to the hmm over at least 500 contiguous bases. Amino acid sequences were then clustered into centroids using usearch v8.0.1517 (sortbylength and cluster_smallmem -id 0.8 -query_cov 0.9; Edgar, 2010). Centroids were aligned using mafft v7.245 (Katoh and Standley, 2013) and trees were constructed using FastTree 2.1.4 (Price *et al.,* 2010). The initial phylogenetic placement of *Ca.* N. islandicus *iorA* in the resulting large tree (3,179 sequences) was used to choose a small set of bacterial *iorA* sequences to include in the final tree. The final dataset was aligned using mafft v7.245 (Katoh and Standley, 2013) and trees were constructed using FastTree 2.1.4 (Price *et al.,* 2010).

Electron microscopy

For scanning electron microscopy, *Ca.* N. islandicus cells were harvested by centrifugation (4,500 x *g*, 15 min, 25 °C) and fixed on poly-L-lysine coated slides with a filter-sterilized 2.5 % glutaraldehyde fixation solution in phosphate buffered saline (PBS; 130 mM NaCl in 5 % [v/v] phosphate buffer mixture [20 to 80 v/v] of 200 mM NaH_2PO_4 and 200 mM Na_2HPO_4). Subsequently, fixed cells were washed three times for 10 min in PBS and post-fixed with a 1 % OsO_4 solution in PBS for 40 min. The fixed cells were again washed three times in PBS, dehydrated

in a 30 to 100 % (v/v) ethanol series, washed in acetone, and critical point dried with a CPD 300 unit (Leica). Samples were mounted on stubs, sputter coated with gold using a sputter coater JFC-2300HR (JEOL), and images were obtained with a JSM-IT300 scanning electron microscope (JEOL).

Results and Discussion

Enrichment and basic physiology of Ca. N. islandicus

An ammonia-oxidizing enrichment culture was established from biofilm material sampled from a hot spring located in the geothermal valley Graendalur of South-Western Iceland. Temperature tests for optimal activity and growth were performed at different time points during the enrichment period and showed varying results, but below 50 °C and above 75 °C activity and growth was never observed. Only during the initial enrichment phase did ammonia oxidation occur at 75 °C. At 65 °C the highest ammonia oxidation rates and the shortest lag phases were usually measured (data not shown), however in a single experiment the optimal temperature was 70 °C (Fig. S1). Likely, these variations reflect varying abundance ratios of *Ca.* N. islandicus and accompanying bacteria over time as described in Lebedeva *et al.* (2008). A high enrichment level of a single AOA phylotype (see below) was achieved by applying the antibiotic spiramycin (15 mg l^{-1}) followed by biomass transfers into fresh medium using serial dilutions. This enrichment culture showed near stoichiometric conversion of ammonium to nitrite when incubated at 65 °C (Fig. 1). This was accompanied by growth of the AOA with a specific growth rate of 0.128 ± 0.011 d^{-1} (mean generation time of 2.32 ± 0.24 d), which is substantially slower than those reported for *Ca.* Nitrosocaldus yellowstonensis HL72, *N. viennensis* EN76, or *N. maritimus* SCM1 (de al Tore *et al.,* 2008; Könneke *et al.,* 2005; Martens-Habbena *et al.,* 2009; Stieglmeier *et al.,* 2014b; Table 1), but faster than a marine enrichment culture (Berg *et al.,* 2015).

Genome reconstruction, phylogeny, and environmental distribution

Metagenomic sequencing of the enrichment culture with Illumina and Nanopore demonstrated that the current culture contained an AOA as the only taxon encoding the repertoire genes required for ammonia oxidation. Hybrid assembly allowed reconstruction of the complete genome of this AOA as one circular contiguous sequence of 1.62 Mbps length (Table 1). The 16S rRNA gene and *amoA* gene of the newly enriched AOA are 96 and 85 % identical respectively to the genes of *Ca.* Nitrosocaldus yellowstonensis, the only other cultured obligately thermophilic AOA. The average amino acid sequence identity (AAI) and the genomic average nucleotide identity (gANI) between the genome and the one of *Ca.* N. yellowstonensis are 65.4 % (alignment fraction: 0.86) and 75.8 % (alignment fraction: 0.59), respectively, which is above the proposed genus and below the proposed species boundary thresholds (Qin *et al.*, 2014; Varghese *et al.,* 2015). Consequently, the enriched obligately thermophilic AOA was assigned to the same genus and referred to as *Ca.* Nitrosocaldus islandicus. According to 16S rRNA gene- and *amoA* gene-based phylogenies, *Ca.* N. islandicus is a novel member of the *Nitrosocaldales* clade, which seems to predominantly encompass AOA from thermal environments (Fig. 2). An extended phylogenomic analysis using a concatenated alignment of 34 proteins (Table S2) identified by CheckM (Parks *et al.,* 2015) confirmed that *Ca.* N. islandicus represents a basal lineage within the known ammonia-oxidizing *Thaumarchaea* (Fig. 3). This result lends strong support to the earlier notion, which was based on single-gene 16S rRNA and *amoA* phylogenies (de la Torre *et al.*, 2008), that the thermophilic *Nitrosocaldales* clade is an early diverging group of the ammonia-oxidizing *Thaumarchaea*. It would also be compatible with the possibility that archaeal ammonia oxidation originated in thermal environments (de la Torre *et al.,* 2008; Hatzenpichler *et al.,* 2008; Groussin *et al.,* 2011; Brochier-Armanet *et al.,* 2012).

Metagenomic sequencing revealed that in addition to *Ca.* N. islandicus the culture also contained two heterotrophic bacterial contaminants, which were identified as a *Thermus sp.* and a member of

the *Chloroflexi* phylum (Fig. 4). The enrichment level of *Ca.* N. islandicus was approximately 85 % based on read counts from the Nanopore sequencing, whereas the *Thermus sp.* and *Chloroflexi* accounted for 12 % and 3 %, respectively. FISH-analysis of the enrichment culture confirmed the dominance of *Ca.* N. islandicus and showed that the AOA grew mainly in aggregates, whereas the bacterial cells grew either co-localized with the archaeal flocs or planktonic (Fig. 5A). Electron microscopy demonstrated that the cells of *Ca.* N. islandicus are small (with a diameter of approximately 0.5 to 0.7 µm) and have an irregular coccoid shape (Fig. 5B). Morphologically they resemble the cells of *Ca.* N. yellowstonensis (Qin *et al.*, 2017b).

The environmental distribution of the two cultured *Nitrosocaldales* members and closely related AOA was assessed by screening all publicly available 16S rRNA gene amplicon datasets (n=93,045) for sequences highly similar (97 %) to the 16S rRNA genes of *Ca.* N. islandicus and *Ca.* N. yellowstonensis using the pipeline described by Lagkouvardos *et al.* (2016). This analysis revealed that these taxa are highly confined in their distribution and occur predominantly in terrestrial hot springs where they can reach high relative abundances between 11.4 % and 86 % (*Ca.* N. islandicus and *Ca.* N. yellowstonensis, respectively) of the total microbial community (Fig. 6). Interestingly, *Ca.* N. yellowstonensis-related organisms seem to occur mainly in hot springs described as alkaline with a pH of around 8.5, but were also detected in a sample from a Tibetan wastewater treatment plant (Niu *et al.,* 2017). The unexpected detection of members of the *Nitrosocaldales* in the latter sample was confirmed by 16S rRNA-based phylogenetic analyses (data not shown) and it would be interesting to know whether this wastewater treatment plant is in some way connected to water from a close-by hot spring.

Genome features

Addition of the complete genome of *Ca.* N. islandicus to the set of available thaumarchaeal genome sequences (n=30) reduced the number of gene families identified as representing the "*Thaumarchaea*-core" (Herbold *et al.,* 2017) from 743 to 669 (reduction by 9.96 %; Table S3). In a few cases, genes with low sequence homology to apparently absent core gene families are actually present in the genome of *Ca.* N. islandicus, but were not scored as they did not match the alignment length criterion. For example, *Ca.* N. islandicus, like all other AOA sequenced to date, has a gene encoding the K-subunit of RNA polymerase class I, but with a low sequence similarity to the respective orthologous genes in other AOA. In a few other cases, enzymes found in all other AOA genomes are absent but functionally replaced by members of different enzyme families. For example, all other genome-sequenced AOA contain a cobalamin-dependent methionine synthase. In contrast, *Ca.* N. islandicus possesses only an unrelated cobalamin-independent methionine synthase, which is also found in some other AOAs.

In addition to updating the *thaumarchaeal* core genome we also specifically looked for genes that are present in *Ca.* N. islandicus, but were not reported for other AOA before. In the following sections, the most interesting findings from these analyses are reported and put in context.

Like all other AOA, the *Ca.* N. islandicus genome encodes the typical repertoire for CO_2 fixation via the modified 3-hydroxypropionate/4-hydroxybutyrate (3HP/4HB) cycle and for archaeal ammonia oxidation (Fig. 7; Fig. S2; Table 1) (Walker *et al.*, 2010; Spang *et al.*, 2012; Könneke *et al.,* 2014; Otte *et al.,* 2015; Kerou *et al.*, 2016). Unexpectedly however, the gene *nirK* encoding an NO-forming nitrite reductase is absent. NirK has been suggested to play an essential role for ammonia oxidation in AOA by providing NO for the NO-dependent dehydrogenation of hydroxylamine to nitrite (Kozlowski *et al.*, 2016). Interestingly, ammonia oxidation by *Ca.* N. islandicus was completely inhibited after the addition of ≥33 µM of the NO-scavenger PTIO (Fig. S3), a concentration that is lower or in the same range as previously reported to be inhibitory for other AOA (Shen et al., 2013; Jung et al., 2014; Martens-Habbena et al, 2015; Sauder et al., 2016).

This finding suggests that NO is required for ammonia oxidation in *Ca.* N. islandicus despite the absence of NirK. The only other known AOA without a *nirK* gene are the sponge symbiont *Cenarchaeum symbiosum* (Hallam *et al.,* 2006; Bartossek *et al.,* 2010) and *Ca.* N. yellowstonensis (Stahl and de la Torre, 2012). For the uncultured *C. symbiosum* ammonia-oxidizing activity has not been demonstrated and the absence of *nirK* might have resulted from gene loss during adaptation to a life-style as symbiont. *Ca.* N. yellowstonensis is the closest cultured representative of *Ca.* N. islandicus, and the lack of *nirK* may thus be a common feature of the *Nitrosocaldales*. These AOA might produce NO by a yet unknown mechanism. In this context it is interesting to note that the hydroxylamine dehydrogenase of AOB, of which the functional homolog in archaea has not been identified yet, has recently been reported to produce NO instead of nitrite (Caranto and Lancaster, 2017). Alternatively, NO could be provided by accompanying organisms such as the *Thermus* and *Chloroflexi*-like bacteria that remain in the enrichment. Indeed, the genome bins obtained for these organisms both encode a *nirK* gene. The *Thermus sp.* genome bin further contains a *norBC* and *narGH* genes, in line with described denitrification capabilities for the genus *Thermus* (Alvarez *et al.,* 2014). A dependence of *Nitrosocaldales* on NO production by other microorganisms could explain why no pure culture from this lineage has been obtained yet.

Ca. N. islandicus possesses genes coding for urease that are present in some but not all AOA (Walker *et al.,* 2010; Spang *et al.,* 2012; Kerou *et al.,* 2016; Lehtovirta-Morley *et al.,* 2016; Sauder *et al.,* 2017) (Fig. 7; Table 1), but lacks a cyanase that is used by *Nitrososphaera gargensis* for cyanate-based growth (Palatinszky *et al.,* 2015). Additionally, the genome encodes an enzyme that either belongs to a novel class of the nitrilase superfamily or to the cyanide hydratase family (Fig. 7; Fig. S4). Nitrilases catalyze the direct cleavage of a nitrile to the corresponding acid while forming ammonia (Pace and Brenner, 2001) and cyanide hydratases convert HCN to formamide. Both substrates are relatively thermostable (Miyakawa *et al.,* 2002; Isidorov *et al.,* 1992). Nitriles occur as intermediates of microbial metabolism (Kobayashi *et al.,* 1993) and nitrile hydratases have previously been isolated from several thermophiles (Cramp *et al.,* 1997; Almatawah *et al.,* 1999; Kabaivanova *et al.,* 2008). Furthermore, both compounds are intermediates of the proposed abiotic synthesis of organics at hydrothermal sites (Miller and Urey, 1959; Schulte and Shock, 2005) and could thus be available in the hot spring habitat of *Ca.* N. islandicus. Similar genes have been found in the genomes of several other AOA from the *Nitrosopumilus* and *Nitrosotenuis* genera (Walker *et al.,* 2010; Mosier *et al.,* 2012; Lebedeva *et al.,* 2013; Park *et al.,* 2014; Bayer *et al.*, 2016) (Table 1) and it will be interesting to find out for which metabolism they may be used in AOA.

Intriguingly, *Ca.* N. islandicus might be able to ferment amino acids under anaerobic conditions as it contains the entire pathway used by some hyperthermophilic archaea for ATP generation from aromatic amino acids (Mai and Adams, 1994; Adams *et al.,* 2001; Ozawa *et al.,* 2012) (Fig. 7). In this pathway arylpyruvates are formed from aromatic amino acids by the activity of amino acid aminotransferases using 2-oxoglutarate as the amine group acceptor. The glutamate produced by this transamination can be recycled back to 2-oxoglutarate via glutamate dehydrogenase (*gdhA*) with the concomitant reduction of $NADP^+$. With *ilvE* and *aspC* genes present, *Ca.* N. islandicus encodes at least two enzymes for which an aminotransferase activity specific for tyrosine, phenylalanine and aspartate has been demonstrated (Gelfand and Steinberg, 1977). Subsequently, these 2-ketoacids could be oxidatively decarboxylated and converted to aryl-CoAs by the oxygen-sensitive enzyme indolepyruvate oxidoreductase (Ozawa *et al.,* 2012) encoded by *iorAB* using oxidized ferredoxin as electron acceptor. *IorAB* is absent from all other genome-sequenced AOA and the *ior* genes present in *Ca.* N. islandicus have the highest similarity to and cluster together with *iorAB*-genes found in *Kyrpidia tusciae* and *Dadabacteria* (Fig. S5). Finally, transformation of aryl-CoAs to aryl acids catalyzed by ADP-dependent acetyl-CoA/acyl-CoA synthetase (Glasemacher *et al.,* 1997) leads to ATP formation via substrate-level phosphorylation (Fig. 7). *Ca.* N. islandicus encodes four acetyl-CoA/ acyl-CoA synthetases, two of which are most similar to

non-syntenous homologs of acetyl-CoA/ acyl-CoA synthetases found in other AOA. However, the third gene is absent in all other AOA to date and its most similar homologs are encoded by species of the peptidolysing thermophilic archaea *Thermoproteus* and *Sulfolobus* and the fourth is most similar to an acetyl-/ acyl-CoA synthetase found in members of the thermophilic *Bathyarchaea* and *Hadesarchaea*.

The fermentation of aromatic amino acids also requires regeneration of oxidized ferredoxin (reduced by IorAB) and $NADP^+$ (reduced by glutamate dehydrogenase). However, no canonical ferredoxin:$NADP^+$ oxidoreductase, or other enzymes (Buckel and Thauer, 2013) described to regenerate oxidized ferredoxin, are encoded in the genome of *Ca.* N. islandicus. It seems unlikely that the amount of ferredoxin oxidized by an encoded ferredoxin-dependent assimilatory sulfite/nitrite reductase (Fig. 7) would be sufficient to compensate for all ferredoxin reduced in the dissimilatory fermentation pathway. However, Ca. N. islandicus can also oxidize reduced ferredoxin with a 2:oxoglutarate-ferredoxin oxidoreductase (Fig. S5). NAD(P)H can be re-oxidized by a cytosolic, bidirectional, NAD(P)-coupled type 3b [NiFe] –hydrogenase that is encoded by *Ca.* N. islandicus in contrast to all other genomically characterized AOA (Fig. 7; Table 1). NAD(P)H oxidation by this hydrogenase could lead to hydrogen generation, or the enzyme could act as a sulfhydrogenase that reduces zero valent sulfur compounds (produced by other organisms or present in the environment) to hydrogen sulfide (Ma *et al.*, 1993, Adams *et al.,* 2001). The hydrogenase genes are clustered at a single locus and code for the four subunits of the holoenzyme and accessory proteins (Fig. S6). This hydrogenase might also allow *Ca.* N. islandicus to use hydrogen as energy source providing reduced NAD(P)H under oxic conditions as this type of hydrogenase has been shown to tolerate exposure to oxygen (Bryant and Adams, 1989; Berney *et al.*, 2014; Kwan *et al.*, 2015).

Surprisingly, the genome of *Ca.* N. islandicus lacks genes for both subunits of the DNA polymerase D (PolD), which is present in all other AOA and most archaeal lineages (including thermophiles) with the exception of the Crenarchaea (Cann *et al.*, 1998; Makarova *et al.*, 2014, Saw *et al.,* 2015) (Table 1). It is assumed that either PolD alone or together with DNA polymerases of the B family (PolB) is required for DNA synthesis and elongation in these archaea (Čuboňová *et al.,* 2013; Ishino and Ishino, 2013; Makarova *et al.,* 2014). The *Ca.* N. islandicus genome encodes only one B-type DNA polymerase (PolB1, Fig. S7) and one DNA polymerase of the Y family (PolY), generally considered to be involved in the rescue of stalled replication forks and enhancement of cell survival upon DNA damage (Friedberg *et al.,* 2002). Recently, it has been demonstrated for the PolD-lacking Crenarchaeon *Sulfolobus acidocaldarius* that both its PolB1 and PolY have polymerase activities *in vitro* (Peng *et al.,* 2016). However, *Ca.* N. islandicus (like other AOA) does not encode the PolB1-binding proteins PBP1 and PBP2, which are required to form a multisubunit DNA polymerase holoenzyme together with PolB in the Crenarchaeon *Sulfolobus solfataricus* P2 (Yan *et al.,* 2017). We hypothesize that *Ca.* N. islandicus may utilize one or both of the present polymerases for DNA replication, possibly in combination with its heterodimer PriSL, which has been demonstrated to function as a primase, a terminal transferase and a polymerase capable of polymerizing RNA or DNA chains of up to 7,000 nucleotides (Lao-Sirieix and Bell, 2004).

It is also interesting to note that the obligate thermophile *Ca.* N. islandicus like all genome-sequenced *Thaumarchaea* (Spang *et al.*, 2017) does not encode a reverse gyrase, which is widespread in hyperthermophilic microbes including other archaea of the TACK superphylum (Heine and Chandra, 2009; Makarova *et al.*, 2007; López-García *et al.*, 2015), but is not essential for growth under these conditions (Atomi *et al.,* 2004).

In conclusion, we have obtained a highly enriched (~ 85 %) culture of an obligately thermophilic AOA from a hot spring in Iceland. Despite the impressive diversity of AOA in high temperature environments as revealed by molecular tools (Zhang et al., 2008; Wang et al., 2009; Zhao et al., 2011; Nishizawa *et al.,* 2013; Li *et al.,* 2015; Chen *et al.,* 2016), cultivation of only a single

obligately thermophilic AOA species – *Ca.* Nitrosocaldus yellowstonensis - was reported before (de la Torre *et al.,* 2008). The newly enriched AOA represents a new species of the genus *Nitrosocaldus* and was named *Ca.* N. islandicus. Comparative analysis of its closed genome revealed several surprising features like the absence of DNA polymerase D and the lack of canonical NO-generating enzymes, although physiological experiments with a NO-scavenger demonstrated NO-dependent ammonia-oxidation, as described for other AOA (Shen *et al.,* 2013; Jung *et al.,* 2014; Martens-Habbena *et al.,* 2015; Sauder *et al.,* 2016). Furthermore, *Ca.* N. islandicus encodes the enzymatic repertoire for fermentation of aromatic amino acids that is, so far, unique among sequenced AOA. A pure culture of *Ca.* N. islandicus will be required to physiologically verify this genome-based hypothesis. Peptide or aromatic amino acid fermentation would enable an anaerobic lifestyle of *Ca.* N. islandicus and, if more widespread among *Thaumarchaea* not yet characterized (including mesophiles), might help explain their sometimes surprisingly high abundance in anaerobic ecosystems (Molina *et al.,* 2010; Bouskill *et al.,* 2012; Buckles *et al.,* 2013; Beam *et al.,* 2014; Lin *et al.,* 2015).

Based on the data presented here, we propose the following provisional taxonomic assignment for the novel *Thaumarchaeon* in our enrichment culture.

Nitrosocaldales order
Nitrosocaldaceae fam.
'*Candidatus* Nitrosocaldus islandicus' sp. nov.

Etymology. Nitrosus (Latin masculine adjective): nitrous; caldus (Latin masculine adjective): hot; islandicus (Latin masculine genitive name): from Iceland. The name alludes to the physiology of the organism (ammonia oxidizer, thermophilic) and the habitat from which it was recovered.
Locality. The biofilm of a terrestrial hot spring in Graendalur geothermal valley, Iceland (64° 1'7" N, 21° 11'20" W)
Diagnosis. An obligately thermophilic, aerobic chemolithoautotrophic ammonia oxidizer from the phylum *Thaumarchaea* growing as small irregular shaped cocci.

Author contributions

AD, JV and CS cultivated and enriched the culture; AD, CS and PP performed growth and activity experiments; AD performed FISH and SEM analysis; CH, AD, PP, MA and RK performed bioinformatic analysis; JT kindly provided access to the *Ca.* N. yellowstonensis genome; AD, JV, CS, PP, MW and HD manually curated the annotation of the genome and interpreted the genome data; AD and MW wrote the manuscript with help from all co-authors.

Conflict of interest statement

MA and RK own and run DNASense, the sequencing center at which the metagenomes were sequenced and the bins assembled. The authors declare no further conflict of interest.

Acknowledgements

We thank Bjarni Sigurdson and the Agricultural University of Iceland for permission to access and sample in Graendalur valley, Iceland. Daniela Gruber from the Core Facility of Cell Imaging and Ultrastructure Research at the University of Vienna and Stefano Romano are gratefully acknowledged for help with sample preparation for electron microscopy. The authors are grateful to Anja Spang for fruitful discussions. JV, CS, CH, and MW were supported by the ERC Advanced Grant NITRICARE (294343) to MW. AD and HD were supported by the Austrian Science Fund (FWF) grants P25231-B21 and T938.

References

Abell GCJ, Revill AT, Smith C, Bissett AP, Volkman JK, Robert SS. (2010). Archaeal ammonia oxidizers and nirS-type denitrifiers dominate sediment nitrifying and denitrifying populations in a subtropical macrotidal estuary. *ISME J* **4**: 286–300.

Adair KL, Schwartz E. (2008). Evidence that ammonia-oxidizing archaea are more abundant than ammonia-oxidizing bacteria in semiarid soils of northern arizona, USA. *Microb Ecol* **56**: 420–426.

Adams MWW, Holden JF, Menon AL, Schut GJ, Grunden AM, Hou C, *et al.* (2001). Key role for sulfur in peptide metabolism and in regulation of three hydrogenases in the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Bacteriol* **183**: 716–724.

Almatawah QA, Cramp R, Cowan DA. (1999). Characterization of an inducible nitrilase from a thermophilic bacillus. *Extremophiles* **3**: 283–291.

Alvarez L, Bricio C, Blesa A, Hidalgo A, Berenguer J. (2014). Transferable denitrification capability of *Thermus thermophilus*. *Appl Environ Microbiol* **80**: 19–28.

Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA. (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* **56**: 1919–1925.

Angel R, Conrad R. (2013). Elucidating the microbial resuscitation cascade in biological soil crusts following a simulated rain event. *Environ Microbiol* **15**: 2799–2815.

Atomi H, Matsumi R, Imanaka T. (2004). Reverse gyrase is not a prerequisite for hyperthermophilic Life. *J Bacteriol* **186**: 4829–4833.

Bartossek R, Nicol GW, Lanzen A, Klenk HP, Schleper C. (2010). Homologues of nitrite reductases in ammonia-oxidizing archaea: diversity and genomic context. *Environ Microbiol* **12**: 1075–1088.

Bayer B, Vojvoda J, Offre P, Alves RJE, Elisabeth NH, Garcia JA, *et al.* (2016). Physiological and genomic characterization of two novel marine *thaumarchaeal* strains indicates niche differentiation. *ISME J* **10**: 1051–1063.

Beam JP, Jay ZJ, Kozubal MA, Inskeep WP. (2014). Niche specialization of novel *Thaumarchaeota* to oxic and hypoxic acidic geothermal springs of Yellowstone National Park. *ISME J* **8**: 938–951.

Beam JP. (2015). Geobiological interactions of archaeal populations in acidic and alkaline geothermal springs of Yellowstone National Park, WY, USA. Montana State University, ProQuest Dissertations Publishing, 3708756.

Berg C, Listmann L, Vandieken V, Vogts A, Jürgens K. (2015). Chemoautotrophic growth of ammonia-oxidizing *Thaumarchaeota* enriched from a pelagic redox gradient in the Baltic Sea. *Front Microbiol* **5**. e-pub ahead of print, doi: 10.3389/fmicb.2014.00786.

Berney M, Greening C, Conrad R, Jacobs WR, Cook GM. (2014). An obligately aerobic soil bacterium activates fermentative hydrogen production to survive reductive stress during hypoxia. *Proc Natl Acad Sci* **111**: 11479–11484.

Bouskill NJ, Eveillard D, Chien D, Jayakumar A, Ward BB. (2012). Environmental factors determining ammonia-oxidizing organism distribution and diversity in marine environments. *Environ Microbiol* **14**: 714–729.

Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P. (2008). Mesophilic *Crenarchaeota*: proposal for a third archaeal phylum, the *Thaumarchaeota*. *Nat Rev Micro* **6**: 245–252.

Brochier-Armanet C, Gribaldo S, Forterre P. (2012). Spotlight on the *Thaumarchaeota*. *ISME J* **6**: 227–230.

Bryant FO, Adams MW. (1989). Characterization of hydrogenase from the hyperthermophilic archaebacterium *Pyrococcus furiosus*. *J Biol Chem* **264**: 5070–5079.

Buckel W, Thauer RK. (2013). Energy conservation via electron bifurcating ferredoxin reduction and proton/Na^+ translocating ferredoxin oxidation. *Biochim Biophys Acta BBA - Bioenerg* **1827**: 94–113.

Buckles LK, Villanueva L, Weijers JWH, Verschuren D, Damsté JSS. (2013). Linking isoprenoidal GDGT membrane lipid distributions with gene abundances of ammonia-oxidizing *Thaumarchaeota* and uncultured *crenarchaeotal* groups in the water column of a tropical lake (Lake Challa, East Africa). *Environ Microbiol* **15**: 2445–2462.

Cann IKO, Komori K, Toh H, Kanai S, Ishino Y. (1998). A heterodimeric DNA polymerase: Evidence that members of *Euryarchaeota* possess a distinct DNA polymerase. *Proc Natl Acad Sci* **95**: 14250–14255.

Caranto JD, Lancaster KM. (2017). Nitric oxide is an obligate bacterial nitrification intermediate produced by hydroxylamine oxidoreductase. *Proc Natl Acad Sci* **114**: 8217–8222.

Chen S, Peng X, Xu H, Ta K. (2016). Nitrification of archaeal ammonia oxidizers in a high-temperature hot spring. *Biogeosciences* **13**: 2051–2060.

Cramp R, Gilmour M, Cowan DA. (1997). Novel thermophilic bacteria producing nitrile-degrading enzymes. *Microbiology* **143**: 2313–2320.

Čuboňová L, Richardson T, Burkhart BW, Kelman Z, Connolly BA, Reeve JN, *et al.* (2013). Archaeal DNA polymerase D but not DNA polymerase B is required for genome replication in *Thermococcus kodakarensis*. *J Bacteriol* **195**: 2322–2328.

Daebeler A, Abell GCJ, Bodelier PLE, Bodrossy L, Frampton DMF, Hefting MM, *et al.* (2012). Archaeal dominated ammonia-oxidizing communities in Icelandic grassland soils are moderately affected by long-term N fertilization and geothermal heating. *Front Microbiol* **3**. e-pub ahead of print, doi: 10.3389/fmicb.2012.00352.

Daims H, Brühl A, Amann R, Schleifer K-H, Wagner M. (1999). The domain-specific probe EUB338 is insufficient for the detection of all bacteria: Development and evaluation of a more comprehensive probe Set. *Syst Appl Microbiol* **22**: 434–444.

Daims H, Stoecker K, Wagner M. (2005). Fluorescence in situ hybridisation for the detection of prokaryotes. In: Osborn AM, Smith CJ (eds). *Molecular Microbial Ecology*. Taylor & Francis: Abington, U.K., pp 213–239.

Demple B, Harrison L. (1994). Repair of oxidative damage to DNA: Enzymology and biology. *Annu Rev Biochem* **63**: 915–948.

Dodsworth JA, Hungate BA, Hedlund BP. (2011). Ammonia oxidation, denitrification and dissimilatory nitrate reduction to ammonium in two US Great Basin hot springs with abundant ammonia-oxidizing archaea. *Environ Microbiol* **13**: 2371–2386.

Edgar RC. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.

Erguder TH, Nico B, Lieven W, Massimo M, Willy V. (2009). Environmental factors shaping the ecological niches of ammonia-oxidizing archaea. *FEMS Microbiol Rev* **33**: 855–869.

Francis CA, Roberts KJ, Beman JM, Santoro AE, Oakley BB. (2005). 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.

Francis CA, Beman JM, Kuypers MMM. (2007). New processes and players in the nitrogen cycle: The microbial ecology of anaerobic and archaeal ammonia oxidation. *ISME J* **1**: 19–27.

Friedberg EC, Wagner R, Radman M. (2002). Specialized DNA polymerases, cellular survival, and the genesis of mutations. *Science* **296**: 1627–1630.

Gelfand DH, Steinberg RA. (1977). *Escherichia coli* mutants deficient in the aspartate and aromatic amino acid aminotransferases. *J Bacteriol* **130**: 429–440.

Gibson M, Chen P-T, Drennan CL. (2016). A structural phylogeny for understanding 2-oxoacid oxidoreductase function. *Curr Opin Struct Biol* **41**: 54–61.

Glasemacher J, Bock A-K, Schmid R, Schönheit P. (1997). Purification and properties of acetyl-CoA synthetase (ADP-forming), an archaeal enzyme of acetate formation and ATP Synthesis, from the hyperthermophile *Pyrococcus furiosus*. *Eur J Biochem* **244**: 561–567.

Groussin M, Gouy M. (2011). Adaptation to environmental temperature is a major determinant of molecular evolutionary rates in Archaea. *Mol Biol Evol* **28**: 2661–2674.

Guy L, Ettema TJG. (2011). The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol* **19**: 580–587.

Haas BJ, Sandigursky M, Tainer JA, Franklin WA, Cunningham RP. (1999). Purification and characterization of *Thermotoga maritima* endonuclease IV, a thermostable apurinic/ apyrimidinic endonuclease and 3′-repair diesterase. *J Bacteriol* **181**: 2834–2839.

Hallam SJ, Konstantinidis KT, Putnam N, Schleper C, Watanabe Y, Sugahara J, *et al.* (2006). Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc Natl Acad Sci* **103**: 18296–18301.

Hatzenpichler R, Lebedeva EV, Spieck E, Stoecker K, Richter A, Daims H, *et al.* (2008). A moderately thermophilic ammonia-oxidizing crenarchaeote from a hot spring. *Proc Natl Acad Sci* **105**: 2134–2139.

Heine M, Chandra SBC. (2009). The linkage between reverse gyrase and hyperthermophiles: A review of their invariable association. *J Microbiol* **47**: 229.

Herbold CW, Lehtovirta-Morley LE, Jung M-Y, Jehmlich N, Hausmann B, Han P, *et al.* (2017). Ammonia-oxidising archaea living at low pH: Insights from comparative genomics. *Environ Microbiol,* in press.

Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.

Ishino, Yoshizumi and Ishino, Sonoko (2013). DNA Replication in Archaea, the Third Domain of Life. In: Stuart, David (ed) The Mechanisms of DNA Replication; *InTech*, DOI: 10.5772/53986.

Isidorov V, Zenkevich I, Karpov G. (1992). Volatile organic compounds in steam-gas outflows of several volcanoes and hydrothermal systems in Kamchatka. *J Volc Seis* **13**: 287–293.

Jiang HC, Huang QY, Dong HL, Wang P, Wang FP, Li WJ, *et al.* (2010). RNA-based investigation of ammonia-oxidizing archaea in hot springs of Yunnan Province, China. *Appl Environ Microbiol* **76**: 4538–4541.

Jung M-Y, Kim J-G, Sinninghe Damsté JS, Rijpstra WIC, Madsen EL, Kim S-J, *et al.* (2016). A hydrophobic ammonia-oxidizing archaeon of the *Nitrosocosmicus* clade isolated from coal tar-contaminated sediment. *Environ Microbiol Rep* **8**: 983–992.

Jung M-Y, Park S-J, Kim S-J, Kim J-G, Damsté JSS, Jeon CO, *et al.* (2014). A Mesophilic, autotrophic, ammonia-oxidizing archaeon of *thaumarchaeal* Group I.1a cultivated from a deep oligotrophic soil horizon. *Appl Environ Microbiol* **80**: 3645–3655.

Jung MY, Park S-J, Min D, Kim J-S, Rijpstra WIC, Sinninghe Damsté JS, *et al.* (2011). Enrichment and characterization of an autotrophic ammonia-oxidizing archaeon of mesophilic crenarchaeal Group I.1a from an agricultural soil. *Appl Environ Microbiol* **77**: 8635–8647.

Kabaivanova L, Dimitrov P, Boyadzhieva I, Engibarov S, Dobreva E, Emanuilova E. (2008). Nitrile degradation by free and immobilized cells of the thermophile *Bacillus* sp. UG-5B, isolated from polluted industrial waters. *World J Microbiol Biotechnol* **24**: 2383–2388.

Käll L, Krogh A, Sonnhammer ELL. (2007). Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res* **35**: W429–W432.

Kandeler E, Gerber H. (1988). Short-term assay of soil urease activity using colorimetric determination of ammonium. *Biol Fertil Soils* **6**: 68–72.

Katoh K, Standley DM. (2013). MAFFT Multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* **30**: 772–780.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**: 845–858.

Kerou M, Offre P, Valledor L, Abby SS, Melcher M, Nagler M, *et al.* (2016). Proteomics and comparative genomics of *Nitrososphaera viennensis* reveal the core genome and adaptations of archaeal ammonia oxidizers. *Proc Natl Acad Sci* **113**: E7937–E7946.

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

Könneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**: 543–546.

Könneke M, Schubert DM, Brown PC, Hügler M, Standfest S, Schwander T, *et al.* (2014). Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO2 fixation. *Proc Natl Acad Sci* **111**: 8239–8244.

Konstantinidis KT, Tiedje JM. (2005). Towards a genome-based taxonomy for prokaryotes. *J Bacteriol* **187**: 6258–6264.

Kostan J, Sjöblom B, Maixner F, Mlynek G, Furtmüller PG, Obinger C, *et al.* (2010). Structural and functional characterisation of the chlorite dismutase from the nitrite-oxidizing bacterium "*Candidatus* Nitrospira defluvii": Identification of a catalytically important amino acid residue. *J Struct Biol* **172**: 331–342.

Kozlowski JA, Stieglmeier M, Schleper C, Klotz MG, Stein LY. (2016). Pathways and key intermediates required for obligate aerobic ammonia-dependent chemolithotrophy in bacteria and *Thaumarchaeota*. *ISME J* **10**: 1836–1845.

Kwan P, McIntosh CL, Jennings DP, Hopkins RC, Chandrayan SK, Wu C-H, *et al.* (2015). The [NiFe]-hydrogenase of *Pyrococcus furiosus* exhibits a new type of oxygen tolerance. *J Am Chem Soc* **137**: 13556–13565.

Lagkouvardos I, Joseph D, Kapfhammer M, Giritli S, Horn M, Haller D, *et al.* (2016). IMNGS: A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. *Sci Rep* **6**: 33721.

Lao-Sirieix S, Bell SD. (2004). The heterodimeric primase of the hyperthermophilic archaeon *Sulfolobus solfataricus* possesses DNA and RNA primase, polymerase and 3′-terminal nucleotidyl transferase activities. *J Mol Biol* **344**: 1251–1263.

Lartillot N, Lepage T, Blanquart S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinforma Oxf Engl* **25**: 2286–2288.

Lebedeva EV, Alawi M, Maixner F, Jozsa P-G, Daims H, Spieck E. (2008). Physiological and phylogenetic characterization of a novel lithoautotrophic nitrite-oxidizing bacterium, '*Candidatus* Nitrospira bockiana'. *Int J Syst Evol Microbiol* **58**: 242–250.

Lebedeva EV, Hatzenpichler R, Pelletier E, Schuster N, Hauzmayer S, Bulaev A, *et al.* (2013). Enrichment and genome sequence of the group I. 1a ammonia-oxidizing archaeon '*Ca.* Nitrosotenuis uzonensis' representing a clade globally distributed in thermal habitats. *Plos One* **8**. e-pub ahead of print, doi: 10.1371/journal.pone.0080835.

Lehtovirta-Morley LE, Sayavedra-Soto LA, Gallois N, Schouten S, Stein LY, Prosser JI, *et al.* (2016). Identifying potential mechanisms enabling acidophily in the ammonia-oxidising archaeon '*Candidatus* Nitrosotalea devanaterra'. *Appl Environ Microbiol* **82**: 2608-2619.

Leininger S, Urich T, Schloter M, Schwark L, Qi J, Nicol GW, *et al.* (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442**: 806–809.

Li H. (2016). Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32**: 2103–2110.

Li H, Yang Q, Li J, Gao H, Li P, Zhou H. (2015). The impact of temperature on microbial diversity and AOA activity in the Tengchong geothermal field, China. *Sci Rep* **5**: 17056.

Lin X, Handley KM, Gilbert JA, Kostka JE. (2015). Metabolic potential of fatty acid oxidation and anaerobic respiration by abundant members of *Thaumarchaeota* and *Thermoplasmata* in deep anoxic peat. *ISME J* **9**: 2740–2744.

López-García P, Zivanovic Y, Deschamps P, Moreira D. (2015). Bacterial gene import and mesophilic adaptation in archaea. *Nat Rev Microbiol* **13**: 447–456.

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

Ma K, Schicho RN, Kelly RM, Adams MW. (1993). Hydrogenase of the hyperthermophile *Pyrococcus furiosus* is an elemental sulfur reductase or sulfhydrogenase: Evidence for a sulfur-reducing hydrogenase ancestor. *Proc Natl Acad Sci U S A* **90**: 5341–5344.

Mai X, Adams MW. (1994). Indolepyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*. A new enzyme involved in peptide fermentation. *J Biol Chem* **269**: 16726–16732.

Maixner F, Wagner M, Lücker S, Pelletier E, Schmitz-Esser S, Hace K, *et al.* (2008). Environmental genomics reveals a functional chlorite dismutase in the nitrite-oxidizing bacterium '*Candidatus* Nitrospira defluvii'. *Environ Microbiol* **10**: 3043–3056.

Makarova KS, Krupovic M, Koonin EV. (2014). Evolution of replicative DNA polymerases in archaea and their contributions to the eukaryotic replication machinery. *Front Microbiol* **5**, doi: 10.3389/fmicb.2014.00354.

Makarova KS, Sorokin AV, Novichkov PS, Wolf YI, Koonin EV. (2007). Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea. *Biol Direct* **2**: 33.

Martens-Habbena W, Qin W, Horak REA, Urakawa H, Schauer AJ, Moffett JW, *et al.* (2015). The production of nitric oxide by marine ammonia-oxidizing archaea and inhibition of archaeal ammonia oxidation by a nitric oxide scavenger. *Environ Microbiol* **17**: 2261–2274.

Martens-Habbena W, Berube PM, Urakawa H, de la Torre JR, Stahl DA. (2009). Ammonia oxidation kinetics determine niche separation of nitrifying archaea and bacteria. *Nature* **461**: 976–979.

Miller SL, Urey HC. (1959). Organic compound synthesis on the primitive earth. *Science* **130**: 245–251.

Mincer TJ, Church MJ, Taylor LT, Preston C, Kar DM, DeLong EF. (2007). Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. *Environ Microbiol* **9**: 1162–1175.

Miranda KM, Espey MG, Wink DA. (2001). A rapid, simple spectrophotometric method for simultaneous detection of nitrate and nitrite. *Nitric Oxide* **5**: 62–71.

Miyakawa S, Cleaves HJ, Miller SL. (2002). The cold origin of life: Implications based on the hydrolytic stabilities of hydrogen, cyanide and formamide. *Orig Life Evol Biosph* **32**: 195–208.

Molina V, Belmar L, Ulloa O. (2010). High diversity of ammonia-oxidizing archaea in permanent and seasonal oxygen-deficient waters of the eastern South Pacific. *Environ Microbiol* **12**: 2450–2465.

Mussmann M, Brito I, Pitcher A, Damste JSS, Hatzenpichler R, Richter A, *et al.* (2011). *Thaumarchaeotes* abundant in refinery nitrifying sludges express *amoA* but are not obligate autotrophic ammonia oxidizers. *Proc Natl Acad Sci U S A* **108**: 16771–16776.

Nishizawa M, Koba K, Makabe A, Yoshida N, Kaneko M, Hirao S, *et al.* (2013). Nitrification-driven forms of nitrogen metabolism in microbial mat communities thriving along an ammonium-enriched subsurface geothermal stream. *Geochim Cosmochim Acta* **113**: 152–173.

Niu L, Zhang X, Li Y, Wang P, Zhang W, Wang C, *et al.* (2017). Elevational characteristics of the archaeal community in full-scale activated sludge wastewater treatment plants at a 3660-meter elevational scale. *Water Sci Technol* wst2017215.

Otte J, Mall A, Schubert DM, Könneke M, Berg IA. (2015). Malonic semialdehyde reductase from the archaeon *Nitrosopumilus maritimus* is involved in the autotrophic 3-hydroxypropionate/4-hydroxybutyrate cycle. *Appl Environ Microbiol* **81**: 1700–1707.

Ozawa Y, Siddiqui MA, Takahashi Y, Urushiyama A, Ohmori D, Yamakura F, *et al.* (2012). Indolepyruvate ferredoxin oxidoreductase: An oxygen-sensitive iron–sulfur enzyme from the hyperthermophilic archaeon *Thermococcus profundus*. *J Biosci Bioeng* **114**: 23–27.

Pace HC, Brenner C. (2001). The nitrilase superfamily: classification, structure and function. *Genome Biol* **2**: reviews0001.1-reviews0001.9.

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.

Park S-J, Ghai R, Martín-Cuadrado A-B, Rodríguez-Valera F, Chung W-H, Kwon K, *et al.* (2014). Genomes of two new ammonia-oxidizing archaea enriched from deep marine sediments. *PLoS ONE* **9**: e96449.

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.*

Peng L, Xia X, Liu X. (2016). Biochemical characterization of translesion synthesis by *Sulfolobus acidocaldarius* DNA polymerases. *Chem Res Chin Univ* **32**: 226–233.

Pjevac P, Schauberger C, Poghosyan L, Herbold CW, van Kessel MAHJ, Daebeler A, *et al.* (2017). *AmoA*-targeted polymerase chain reaction primers for the specific detection and quantification of Comammox *Nitrospira* in the environment. *Front Microbiol* **8**. doi: 10.3389/fmicb.2017.01508.

Price MN, Dehal PS, Arkin AP. (2010). FastTree 2 – Approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**: e9490.

Prosser JI, Nicol GW. (2012). Archaeal and bacterial ammonia-oxidisers in soil: the quest for niche specialisation and differentiation. *Trends Microbiol* **20**: 523–531.

Qin Q-L, Xie B-B, Zhang X-Y, Chen X-L, Zhou B-C, Zhou J, *et al.* (2014). A proposed genus boundary for the prokaryotes based on genomic insights. *J Bacteriol* **196**: 2210–2215.

Qin W, Heal KR, Ramdasi R, Kobelt JN, Martens-Habbena W, Bertagnolli AD, *et al.* (2017a). *Nitrosopumilus maritimus* gen. nov., sp. nov., *Nitrosopumilus cobalaminigenes* sp. nov., *Nitrosopumilus oxyclinae* sp. nov., and *Nitrosopumilus ureiphilus* sp. nov., four marine ammonia-oxidizing archaea of the phylum *Thaumarchaeota*. *Int J Syst Evol Microbiol*. e-pub ahead of print, doi: 10.1099/ijsem.0.002416.

Qin W, Jewell TNM, Russell VV, Hedlund BP, de la Torre JR, Stahl DA. (2017b). *Candidatus* Nitrosocaldus. In: *Bergey's Manual of Systematics of Archaea and Bacteria*. John Wiley & Sons, Ltd. e-pub ahead of print, doi: 10.1002/9781118960608.gbm01287.

Reigstad LJ, Richter A, Daims H, Urich T, Schwark L, Schleper C. (2008). Nitrification in terrestrial hot springs of Iceland and Kamchatka. *FEMS Microbiol Ecol* **64**: 167–174.

Sauder LA, Albertsen M, Engel K, Schwarz J, Nielsen PH, Wagner M, *et al.* (2017). Cultivation and characterization of *Candidatus* Nitrosocosmicus exaquare, an ammonia-oxidizing archaeon from a municipal wastewater treatment system. *ISME J* **11**: 1142–1157.

Saw JH, Spang A, Zaremba-Niedzwiedzka K, Juzokaite L, Dodsworth JA, Murugapiran SK, *et al.* (2015). Exploring microbial dark matter to resolve the deep archaeal ancestry of eukaryotes. *Phil Trans R Soc B* ***370***: 20140328.

Schleper C, Nicol GW. (2010). Ammonia-oxidizing archaea - physiology, ecology, evolution. In: Poole RK (ed) Vol. 57. *Adv in Microb physiol.* Elsevier Ltd.: London.

Schulte M, Shock E. (1995). Thermodynamics of strecker synthesis in hydrothermal systems. *Orig Life Evol Biosph* **25**: 161–173.

Shen T, Stieglmeier M, Dai J, Urich T, Schleper C. (2013). Responses of the terrestrial ammonia oxidizing archaeon *Ca.* Nitrososphaera viennensis and the ammonia oxidizing bacterium Nitrosospira multiformis to nitrification inhibitors. *FEMS Microbiol Lett* **344**: 121–129.

Spang A, Caceres EF, Ettema TJG. (2017). Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* **357**: eaaf3883.

Spang A, Poehlein A, Offre P, Zumbragel S, Haider S, Rychlik N, *et al.* (2012). The genome of the ammonia-oxidizing *Candidatus* Nitrososphaera gargensis: insights into metabolic versatility and environmental adaptations. *Environ Microbiol* **14**: 3122–3145.

Stahl DA, Amann R. (1991). Development and application of nucleic acid probes in bacterial systematics. In: Stackebrandt E, Goodfellow M (eds). *Nucleic Acid Techniques in Bacterial Systematics*. Chichester, England: John Wiley & Sons Ltd., pp 205–248.

Stahl DA, de la Torre JR. (2012). Physiology and diversity of ammonia-oxidizing archaea. In: Gottesman S, Harwood CS, Schneewind O (eds) Annual Review of Microbiology Vol. 66. *Annual Review of Microbiology, Vol 66*. pp 83–101.

Stetter KO. (1998). Hyperthermophiles: isolation, classification, and properties. In: Horikoshi K, Grant WD (eds). *Extremophiles: microbial life in extreme environments*. Wiley-Liss: New York, N.Y., pp 1–24.

Stieglmeier M, Alves RJE, Schleper C. (2014a). The phylum *Thaumarchaeota*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F (eds). *The Prokaryotes*. Springer Berlin Heidelberg, pp 347–362.

Stieglmeier M, Klingl A, Alves RJE, Rittmann SK-MR, Melcher M, Leisch N, *et al.* (2014b). *Nitrososphaera viennensis* sp. nov., an aerobic and mesophilic ammonia-oxidizing archaeon from soil and member of the archaeal phylum *Thaumarchaeota*. *Int J Syst Evol Microbiol* **64**: 2738–2752.

Suchard MA, Redelings BD. (2006). BAli-Phy: simultaneous Bayesian inference of alignment and phylogeny. *Bioinformatics* **22**: 2047–2048.

de la Torre JR de la, Christopher BW, Anitra EI, Martin K, David AS. (2008). Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. *Environ Microbiol* **10**: 810–818.

Tourna M, Freitag TE, Nicol GW, Prosser JI. (2008). Growth, activity and temperature responses of ammonia-oxidizing archaea and bacteria in soil microcosms. *Environ Microbiol* **10**: 1357–1364.

Vallenet D, Belda E, Calteau A, Cruveiller S, Engelen S, Lajus A, *et al.* (2013). MicroScope—an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Res* **41**: D636–D647.

Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, *et al.* (2015). Microbial species delineation using whole genome sequences. *Nucleic Acids Res* **43**: 6761–6771.

Vaser R, Sović I, Nagarajan N, Šikić M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res*. e-pub ahead of print, doi: 10.1101/gr.214270.116.

Walker CB, Torre JR de la, Klotz MG, Urakawa H, Pinel N, Arp DJ, *et al.* (2010). *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.

Wallner G, Amann R, Beisker W. (1993). Optimizing fluorescent in situ hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. *Cytometry* **14**: 136–143.

Wang S, Xiao X, Jiang L, Peng X, Zhou H, Meng J, *et al.* (2009). Diversity and abundance of ammonia-oxidizing archaea in hydrothermal vent chimneys of the Juan de Fuca Ridge. *Appl Environ Microbiol* **75**: 4216–4220.

Weidler GW, Dornmayr-Pfaffenhuemer M, Gerbl FW, Heinen W, Stan-Lotter H. (2007). Communities of archaea and bacteria in a subsurface radioactive thermal spring in the Austrian central Alps, and evidence of ammonia-oxidizing *Crenarchaeota*. *Appl Environ Microbiol* **73**: 259–270.

Wick RR, Judd LM, Gorrie CL, Holt KE. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Comput Biol* **13**: e1005595.

Williams TA, Szöllősi GJ, Spang A, Foster PG, Heaps SE, Boussau B, *et al.* (2017). Integrative modeling of gene and genome evolution roots the archaeal tree of life. *Proc Natl Acad Sci* **114**: E4602–E4611.

Yan J, Beattie TR, Rojas AL, Schermerhorn K, Gristwood T, Trinidad JC, *et al.* (2017). Identification and characterization of a heterotrimeric archaeal DNA polymerase holoenzyme. *Nat Commun* **8**: 15075.

Zeglin LH, Taylor AE, Myrold DD, Bottomley PJ. (2011). Bacterial and archaeal *amoA* gene distribution covaries with soil nitrification properties across a range of land uses. *Environ Microbiol Rep* **3**: 717–726.

Zhalnina KV, Dias R, Leonard MT, Dorr de Quadros P, Camargo FAO, Drew JC, *et al.* (2014). Genome sequence of *Candidatus* Nitrososphaera evergladensis from group I.1b enriched from Everglades soil reveals novel genomic features of the ammonia-oxidizing archaea. *PLoS ONE* **9**: e101648.

Zhang CL, Ye Q, Huang Z, Li W, Chen J, Song Z, *et al.* (2008). Global occurrence of archaeal *amoA* genes in terrestrial hot springs. *Appl Environ Microbiol* **74**: 6417–6426.

Zhao W, Song Z, Jiang H, Li W, Mou X, Romanek CS, *et al.* (2011). Ammonia-oxidizing archaea in Kamchatka hot springs. *Geomicrobiol J* **28**: 149–159.

Zhang Y, Tian Z, Liu M, Shi ZJ, Hale L, Zhou J, *et al.* (2015). High concentrations of the antibiotic spiramycin in wastewater lead to high abundance of ammonia-oxidizing archaea in nitrifying populations. *Environ Sci Technol* **49**: 9124–9132.

Zhou J, Bruns MA, Tiedje JM. (1996). DNA recovery from soils of diverse composition. *Appl Environ Microbiol* **62**: 316–322.

Zorraquino MA, Althaus RL, Roca M, Molina MP. (2011). Heat treatment effects on the antimicrobial activity of macrolide and lincosamide antibiotics in milk. *J Food Prot* **74**: 311–315.

525

530

Tables and figures

Table 1. Genome features and growth rates of Ca. N. islandicus and of selected reference AOA

Genome features	*Ca.* N. islandicus 3F	*N. gargensis* Ga9-2	*N. viennensis* EN76	*Ca.* N. exaquare G61	*N. devanaterra* Nd1	*Ca.* N. uzonensis N4	*N. maritimus* SCM1
Genome size [Mb]	1.62	2.83	2.53	2.99	1.81	1.65	1.60
Number of scaffolds	1	1	1	1	1	1	1
Number of contigs	1	1	1	1	1	14	1
Average G+C content [%]	41.54	48.35	52.72	33.94	37.07	42.25	34.17
Protein coding density [%]	87.85	83.37	86.43	77.14	90.55	90.42	91.65
Number of genomic objects (CDS, fragment CDS, r/tRNA)	1851	4037	3266	3394	2145	2001	2012
Number of coding sequences (CDS)	1824	3999	3277	3358	2106	1960	1969
Motility/ chemotaxis	+	+	+	-	+	+	-
Carbon fixation	3HP/4HB	3HP/4HB	3HP/4HB	3HP/4HB	3HP/4HB	3HP/4HB	3HP/4HB
Ammonium transporters	3	3	3	1	3	2	2
NirK	0	1	1	1	1	1	1
MCO1 + ZIP/ MCO1[a]	1/ 0	1/ 1	2/ 0	1/ 1	0/ 0	1/ 0	2/ 0
Urease and urea transport	+	+	+	+	-	-	-
Cyanate lyase	-	+	-	-	-	-	-
Nitrilase/ Cyanide hydratase	1	0	0	0	0	1	1
Aromatic amino acid fermentation	+	-	-	-	-	-	-
Hydrogenase	3b	-	4a	-	4a	-	-
Coenzyme F420	+	+	+	+	+	+	+
Vitamin B12	+	+	+	+	+	+	+
Catalase	0	(1)[b]	0	1	0	0	0
Peroxidase	0	0	0	1	0	0	0
Superoxide dismutase	1	1	1	1	1	1	2
Chlorite dismutase-like enzyme[c]	1	1	1	1	1	1	1
DNA polymerases	B1, Y	B1, D, Y	B1, D, Y	B1, D, Y	B1, D, Y	B1, D, Y	B1, D, Y
Generation time [d]	2.32 ± 0.24[d]	NA	1.25 ± 0.03	NA	NA	NA	0.88 – 1.08

a, MCO1+ZIP/ MCO1, multicopper oxidase 1 (as defined in Kerou et al., 2016) with adjacent zinc permease/ multicopper oxidase 1 without an adjacent zinc permease; b, The gene is truncated; c , chlorite dismutases are of interest in other nitrifiers (Maixner et al., 2008; Kostan et al., 2010), but it is not known what their function is in archaea; d, determined at 65 °C; NA, not available.

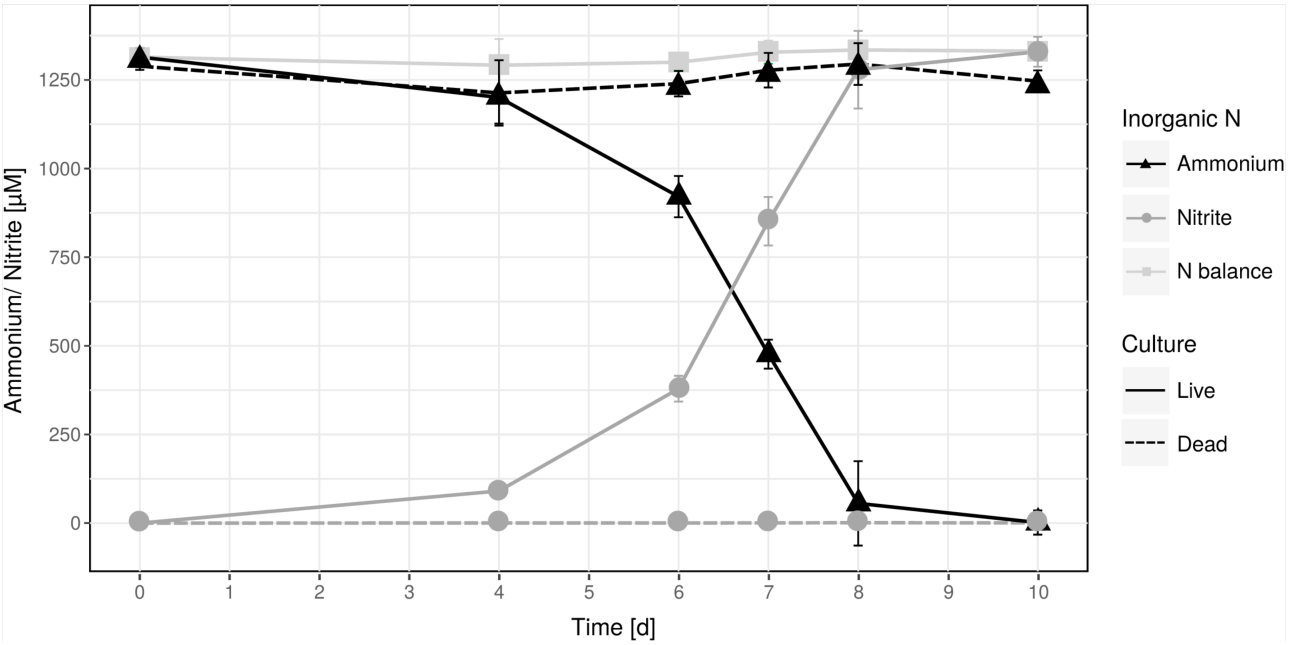

Figure 1. Near-stoichiometric oxidation of 1.25 mM ammonium to nitrite by the *Ca.* N. islandicus enrichment culture at 65 °C. Data points show means, error bars show 1 s.d. of n = 3 biological replicates. Solid and dashed lines denote live and dead culture incubations, respectively. If not visible, error bars are smaller than symbols.

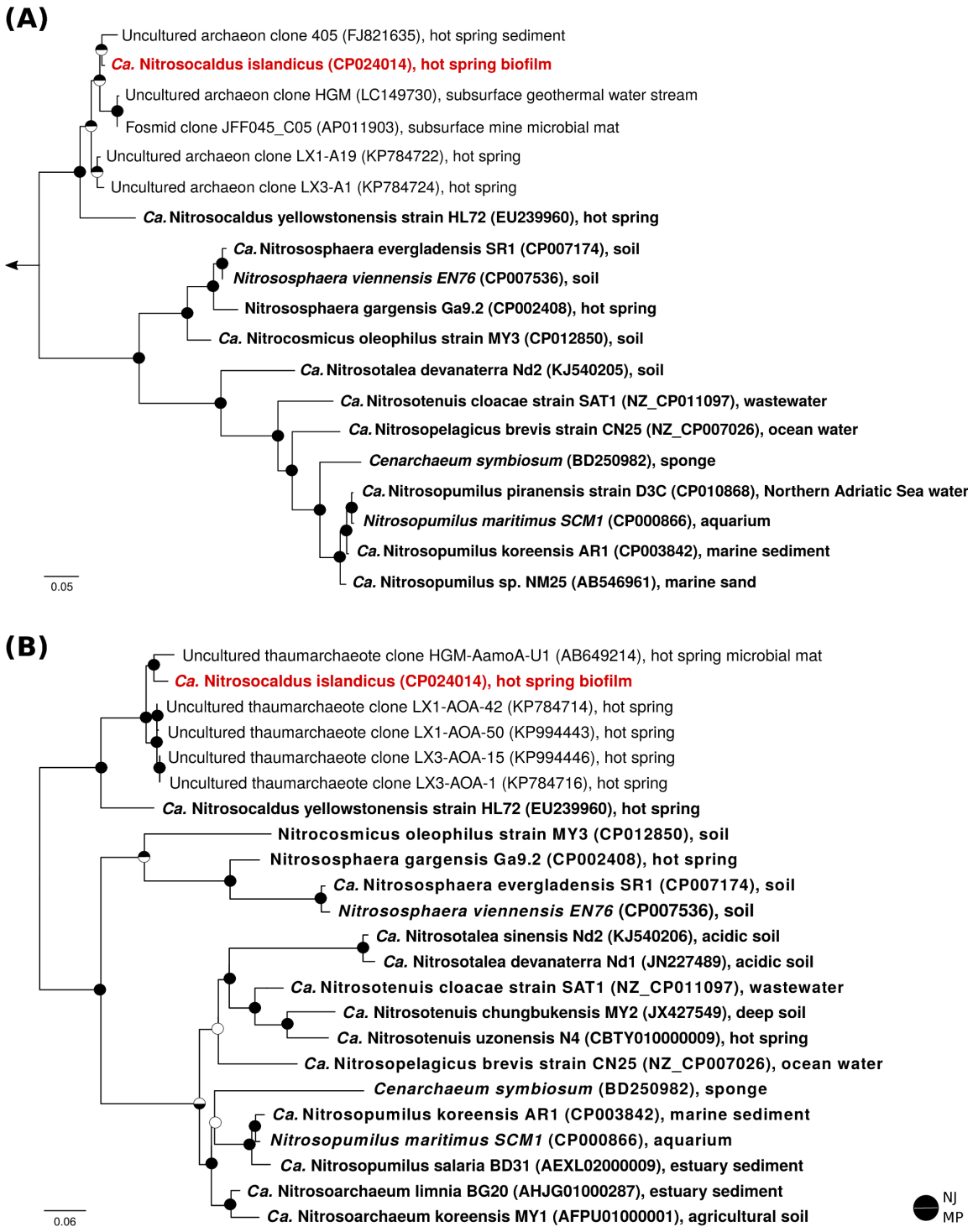

Figure 2. 16S rRNA gene-based (A) and *amoA* gene-based (B) maximum likelihood phylogenies of representative *thaumarchaeal* sequences. For each sequence, the accession number and environmental source are indicated. Sequences from pure and enrichment cultures are depicted in bold, and *Ca*. N. islandicus is highlighted in red. The outgroup for the 16S rRNA tree were *aigarchaeal* sequences; the *amoA* phylogeny was calculated unrooted, but artificially rooted to the *Nitrosocaldales* afterwards. Circles at nodes denote support (filled) or no support (open) from Neighbour Joining (NJ, top half) and Maximum Parsimony (MP, bottom half) trees. The scale bars in panels (A) and (B) indicate 9 and 6 % sequence divergence, respectively.

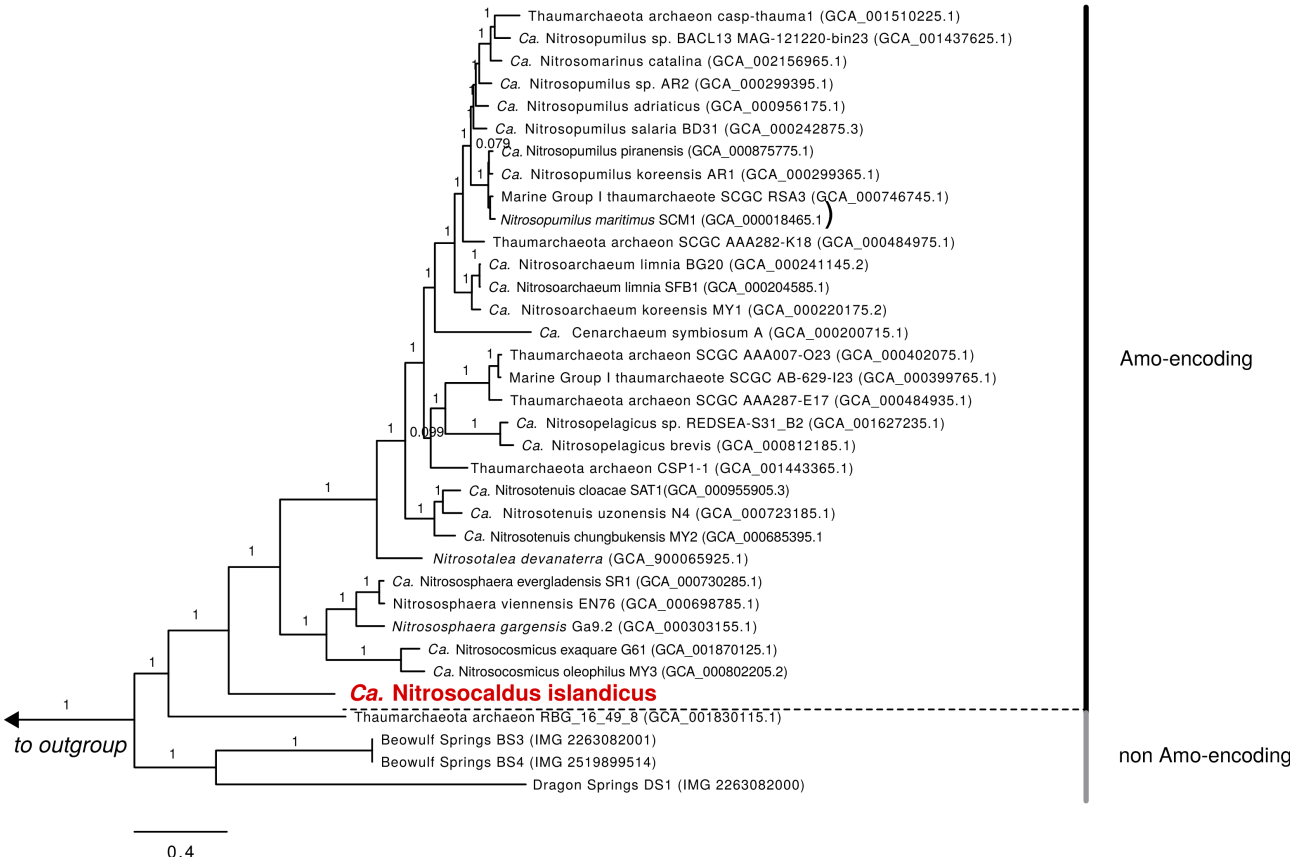

Figure 3. Bayesian inference tree of 34 concatenated universal marker proteins from 31 *amoA*-encoding *Thaumarchaea* including the *Nitrosocaldus*-like AOA and 4 non *amoA*-encoding *Thaumarchaea*-like Archaea. Nineteen additional TACK-superphylum (Guy and Ettema, 2011) members (not shown) were used as an outgroup: *Aigarchaea* (assemblies GCA_000494145.1, GCA_000270325.1), *Bathyarchaea* (GCA_001399805.1, GCA_001399795.1, GCA_001593865.1, GCA_001593855.1, GCA_001593935.1, GCA_002011035.1, GCA_001273385.1), *Crenarchaea* (GCA_000011205.1, GCA_000591035.1, GCA_000253055.1, GCA_000813245.1), *Geothermarchaea* (GCA_002011075.1), *Korarchaea* (GCA_000019605.1), *Thorarchaea* (GCA_001563335.1, GCA_001563325.1), and *Verstraetearchaea* (GCA_001717035.1, GCA_001717015.1). Branches are labelled with average Bayesian posterior probability support over ten independent chains and the scale bar indicates 0.4 amino acid substitutions per site.

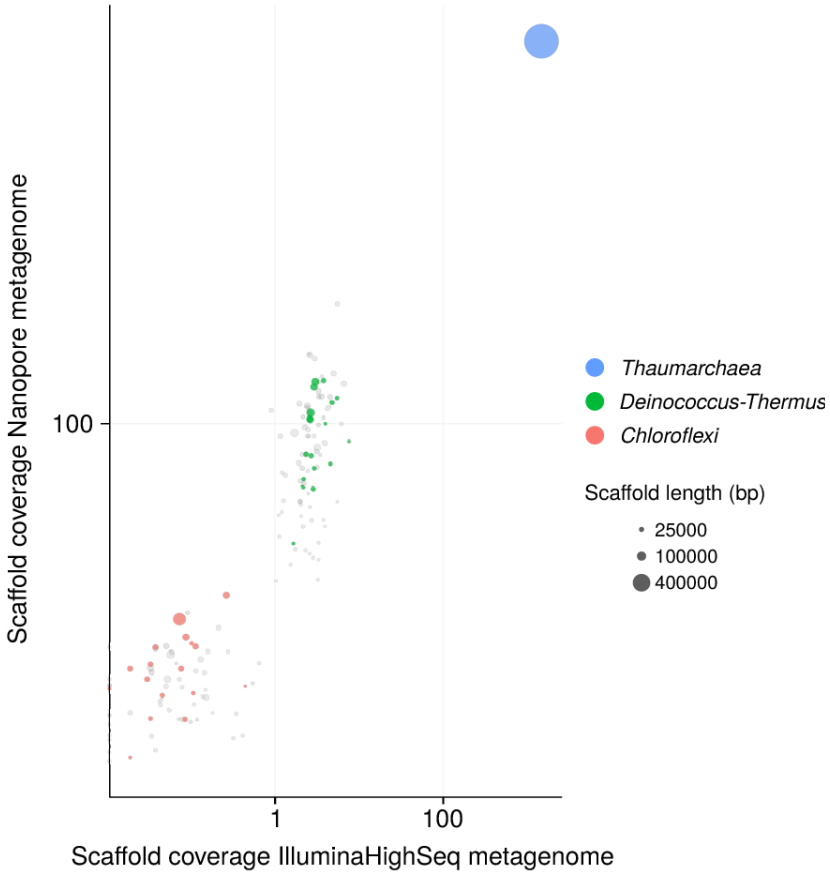

Figure 4. Sequence composition-independent binning of the metagenome scaffolds from two ammonia-oxidizing enrichment cultures. Circles represent scaffolds, scaled by the square root of their length. Clusters of similarly coloured circles represent potential genome bins. The x-axis shows binning of the scaffolds from an early enrichment culture, which still included other genera as well (not shown). The y-axis shows binning of the scaffolds from the latest enrichment culture containing only *Ca.* N. islandicus and the two remaining accompanying organisms. Genome bins for the *Thermus* (34 % complete) and the *Chloroflexi* (56 % complete) organism were obtained. The genome bin of the *Chloroflexi* organism contains genes that cluster within a clade of *Nitrobacter*/*Nitrolancea* nitrite oxidoreductase *(nxrAB)* genes (data not shown). Since we did not observe nitrate production by the enrichment culture, the function of these genes remains unknown.

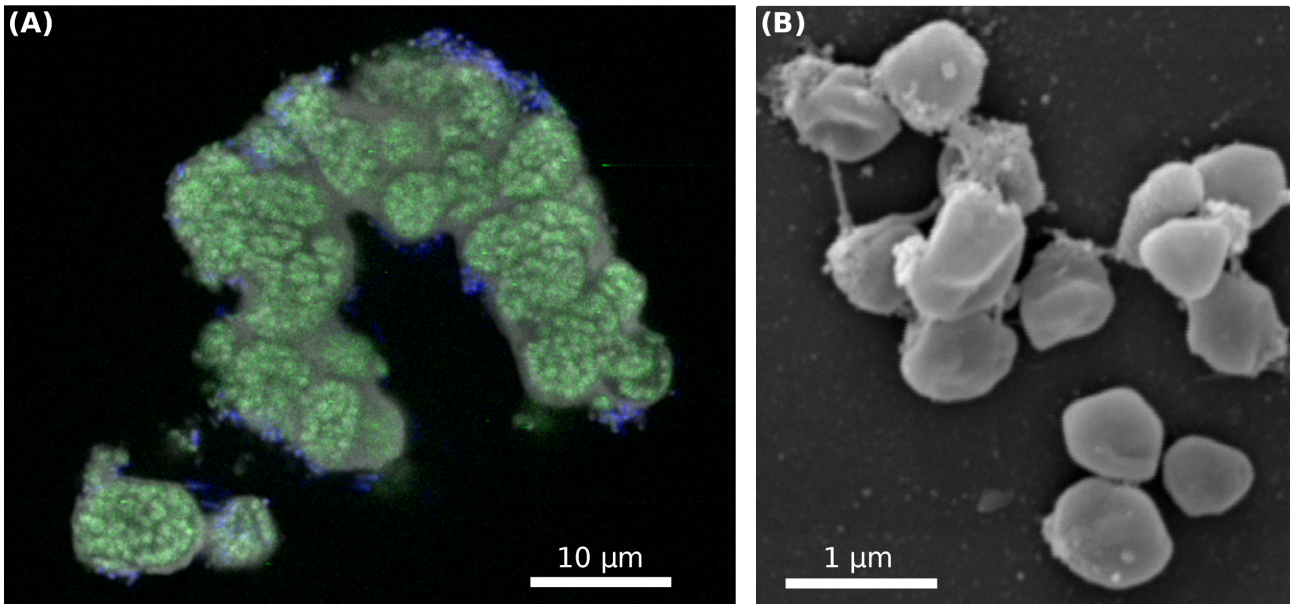

Figure 5. (A) FISH analysis of the enrichment culture illustrating the growth in microcolonies and the high relative abundance of *Ca.* N. islandicus. *Ca.* N. islandicus cells appear in green (stained by probe Thaum726 targeting most *Thaumarchaea*) and the bacterial contaminants in blue (labelled by probe EUB338). (B) Scanning electron micrograph of spherically shaped *Ca.* Nitrosocaldus islandicus cells. The cells have a diameter of 0.5 to 0.7 µm. *Ca.* N. islandicus cells were distinguishable from the rod-shaped bacterial contaminants by their smaller size and unique, 'dented' spherical shape.

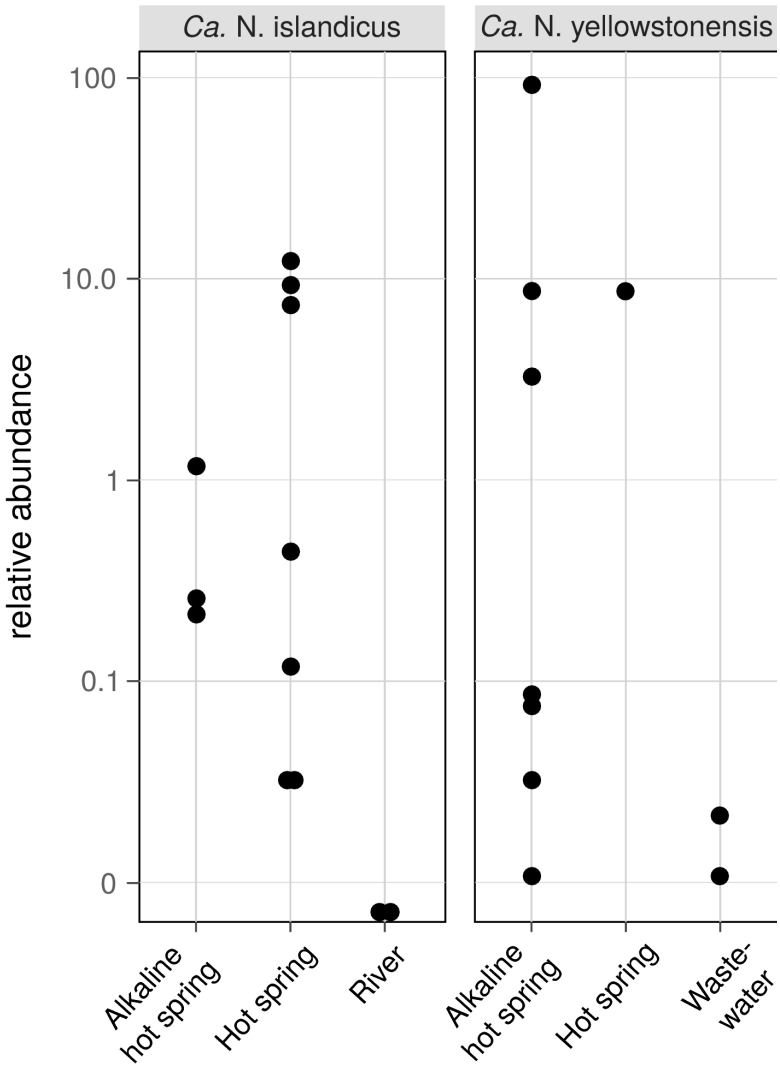

Figure 6. Occurrence and abundance of AOA related to *Ca.* Nitrosocaldus islandicus and *Ca.* Nitrosocaldus yellowstonensis in different habitats based on the presence of closely related 16S rRNA gene sequences in all public 16S rRNA gene amplicon datasets. Data shown are log-scale relative abundances of 16S rRNA gene sequences with a minimum similarity of 97 % in a sample (n=12 and n=10 out of 93,045 total datasets for *Ca.* N. islandicus and *Ca.* N. yellowstonensis, respectively). Sequences of the Tibetean waste water data set where retrieved from BioSample SAMN03464927 of Niu *et al.* (2017).

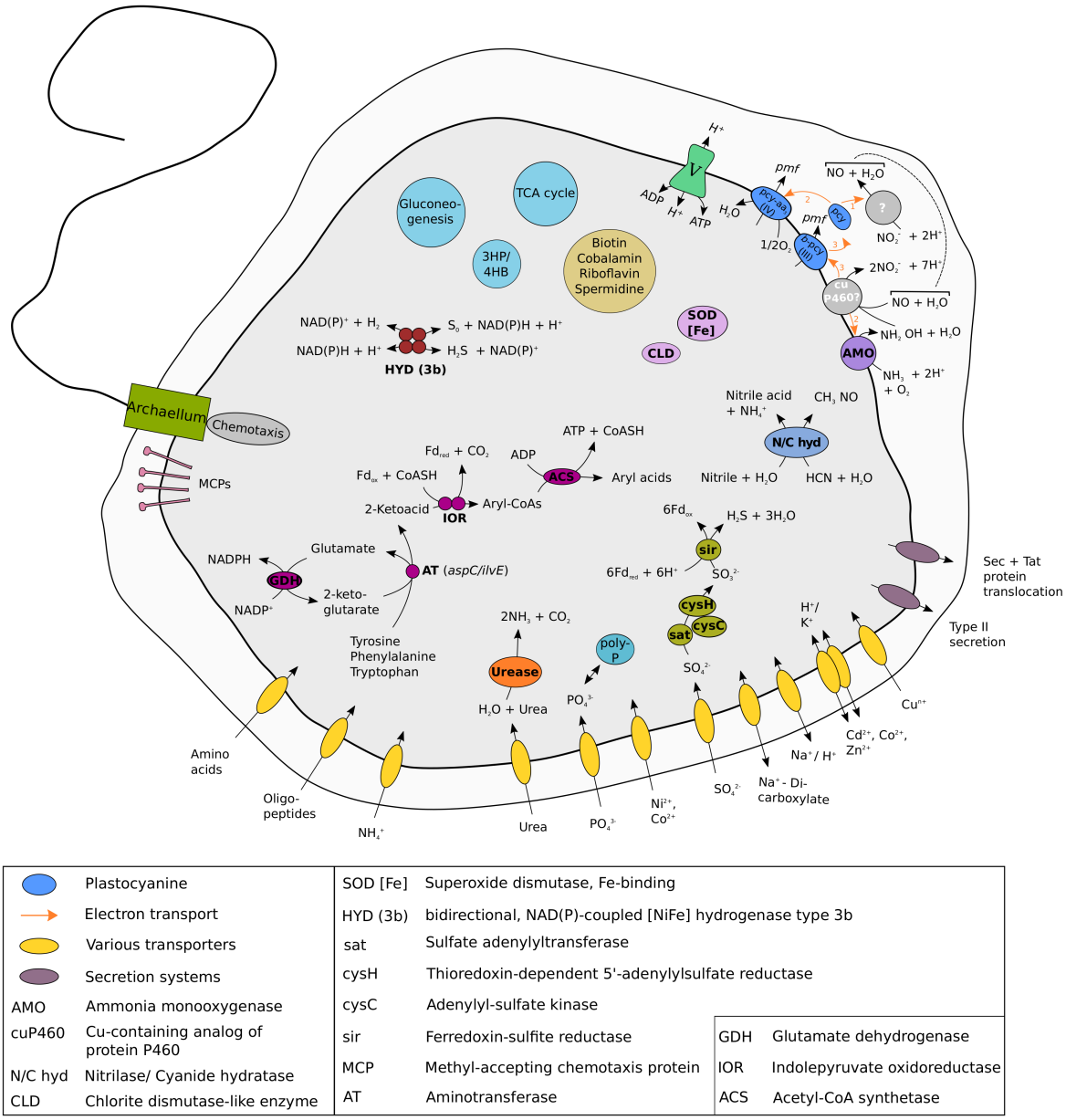

Figure 7. Cell metabolic cartoon constructed from the annotation of the *Ca.* N. islandicus genome. Enzyme complexes of the electron transport chain are labelled by Roman numerals. Most metabolic features displayed are discussed in the main text. The model of ammonia oxidation and electron transfer is depicted as proposed by Kozlowski *et al.* (2016). Locus tags of all genes discussed in the main text are given in table S4. Like some but not all AOA, *Ca.* N. islandicus encodes all genes required to assemble an archaeal flagellar apparatus that is composed of the flagellar filament, the motor, and its switch (Spang *et al.,* 2012; Lehtovirta-Morley *et al.,* 2016; Lebedeva *et al.,* 2013; Mosier *et al.,* 2012; Qin *et al.,* 2017; Li *et al.,* 2016), although no archaellum could be detected in our electron microscopic analysis. The *fla* gene cluster of *Ca.* N. islandicus shows a similar arrangement to *Nitrososphaera gargensis* and contains six genes including one gene for structural flagellin subunit FlaB/FlaA as well as the flagellar accessory genes *flaG*, *flaF*, *flaH*, *flaJ* and *flaI* (Spang *et al.,* 2012).

Supplemental information

Table S1. *Thaumarchaea* used for comparison of protein coding genes

Organism Name	Publication	BioSample	BioProject	Assembly
Thaumarchaeota archaeon casp-thauma1 (Caspean Sea)	10.7717/peerj.2687	SAMN03733542	PRJNA279271	GCA_001510225.1
Nitrosopumilus sp. BACL13 MAG-121220-bin23	10.1186/s13059-015-0834-7	SAMN03741946	PRJNA273799	GCA_001437625.1
Ca. Nitrosomarinus catalina	10.1111/1462-2920.13768	SAMN05730076	PRJNA341864	GCA_002156965.1
Ca. Nitrosopumilus sp. AR2	10.1128/JB.01869-12	SAMN02603138	PRJNA174388	GCA_000299395.1
Ca. Nitrosopumilus adriaticus	10.1038/ismej.2015.200	SAMN03253153	PRJNA269341	GCA_000956175.1
Ca. Nitrosopumilus salaria BD31	10.1128/JB.00013-12.	SAMN00016669	PRJNA50075	GCA_000242875.3
Ca. Nitrosopumilus piranensis	10.1038/ismej.2015.200	SAMN03257648	PRJNA269924	GCA_000875775.1
Ca. Nitrosopumilus koreensis AR1	10.1128/JB.01857-12	SAMN02603137	PRJNA174387	GCA_000299365.1
Marine Group I thaumarchaeote SCGC RSA3 (Red Sea)	10.1038/ismej.2014.137	SAMN02869648	PRJNA248555	GCA_000746745.1
Nitrosopumilus maritimus SCM1	10.1073/pnas.0913533107	SAMN00000032	PRJNA19265	GCA_000018465.1
Thaumarchaeota archaeon SCGC AAA282-K18	10.3389/fmicb.2016.00143	SAMN02440765	PRJNA190793	GCA_000484975.1
Ca. Nitrosoarchaeum limnia BG20	10.1128/JB.00007-12.	SAMN00016663	PRJNA50027	GCA_000241145.2
Ca. Nitrosoarchaeum limnia SFB1	10.1371/journal.pone.0016626.	SAMN02471010	PRJNA52465	GCA_000204585.1
Ca. Nitrosoarchaeum koreensis MY1	10.1128/JB.05717-11.	SAMN02470178	PRJNA67913	GCA_000220175.2
Ca. Cenarchaeum symbiosum A	10.1073/pnas.0608549103	SAMN02744041	PRJNA202	GCA_000200715.1
Thaumarchaeota archaeon SCGC AAA007-O23	10.3389/fmicb.2016.00143	SAMN02440520	PRJNA66857	GCA_000402075.1
Marine Group I thaumarchaeote SCGC AB-629-I23	10.1038/ismej.2014.137	SAMN02441296	PRJNA165501	GCA_000399765.1
Thaumarchaeota archaeon SCGC AAA287-E17	10.3389/fmicb.2016.00143	SAMN02441105	PRJNA190806	GCA_000484935.1
Nitrosopelagicus sp. REDSEA-S31_B2	10.1038/sdata.2016.50, PMC: 4932879	SAMN04534603	PRJNA289734	GCA_001627235.1
Ca. Nitrosopelagicus brevis	10.1073/pnas.1416223112	SAMN03273964	PRJNA223412	GCA_000812185.1
Thaumarchaeota archaeon CSP1-1 (sediment)	10.1111/1462-2920.12930	SAMN03462092	PRJNA262935	GCA_001443365.1
Ca. Nitrosotenuis cloacae	10.1038/srep23747	SAMN03286947	PRJNA272771	GCA_000955905.3
Ca. Nitrosotenuis uzonensis N4	10.1371/journal.pone.0080835	SAMEA3139018	PRJEB4650	GCA_000723185.1
Ca. Nitrosotenuis chungbukensis MY2	10.1128/AEM.03730-13	SAMN02767256	PRJNA210247	GCA_000685395.1
Nitrosotalea devanaterra	10.1128/AEM.04031-15	SAMEA3577360	PRJEB10948	GCA_900065925.1
Ca. Nitrososphaera evergladensis SR1	10.1371/journal.pone.0101648	SAMN03081530	PRJNA235208	GCA_000730285.1
Nitrososphaera viennensis EN76	10.1073/pnas.1601212113	SAMN02721150	PRJEA60103	GCA_000698785.1
Nitrososphaera gargensis Ga9.2	10.1111/j.1462-2920.2012.02893.x	SAMN02603264	PRJNA60505	GCA_000303155.1
Ca. Nitrosocosmicus exaquare G61	10.1038/ismej.2016.192	SAMN04606696	PRJNA317395	GCA_001870125.1
Ca. Nitrosocosmicus oleophilus MY3	10.1111/1758-2229.12477	SAMN03074222	PRJNA210256	GCA_000802205.2

Table S2. Marker genes used for phylogenomic tree

Gene	Pfam Id	Length	Description
Alanine – tRNA ligase	TIGR00344	847	Alanine – tRNA ligase
Ribosomal protein L10	PF00466	100	Ribosomal protein L10
Ribosomal protein L11	PF03946	60	Ribosomal protein L11, N-terminal domain
Ribosomal protein L11	PF00298	69	Ribosomal protein L11, RNA binding domain
Ribosomal protein L13	PF00572	128	Ribosomal protein L13
Ribosomal protein L14p/L23e	PF00238	122	Ribosomal protein L14p/L23e
Ribosomal protein L16p/L10e	PF00252	133	Ribosomal protein L16p/L10e
Ribosomal protein L18p/L5e	PF00861	119	Ribosomal protein L18p/L5e
Ribosomal protein L1p/L10e	PF00687	220	Ribosomal protein L1p/L10e
Ribosomal protein L22p/L17e	PF00237	105	Ribosomal protein L22p/L17e
Ribosomal protein L23	PF00276	92	Ribosomal protein L23
Ribosomal protein L29	PF00831	58	Ribosomal protein L29
Ribosomal protein L3	PF00297	263	Ribosomal protein L3
Ribosomal protein L4/L1	PF00573	192	Ribosomal protein L4/L1
Ribosomal protein L5	PF00281	56	Ribosomal protein L5
Ribosomal protein L5	PF00673	95	Ribosomal protein L5P, C-terminus
Ribosomal protein S11	PF00411	110	Ribosomal protein S11
Ribosomal protein S12/S23	PF00164	122	Ribosomal protein S12/S23
Ribosomal protein S15	PF00312	83	Ribosomal protein S15
Ribosomal protein S17	PF00366	69	Ribosomal protein S17
Ribosomal protein S19	PF00203	81	Ribosomal protein S19
Ribosomal protein S2	PF00318	211	Ribosomal protein S2
Ribosomal protein S3	PF00189	85	Ribosomal protein S3, C-terminal domain
Ribosomal protein S5	PF03719	74	Ribosomal protein S5, C-terminal domain
Ribosomal protein S5	PF00333	67	Ribosomal protein S5, N-terminal domain
Ribosomal protein S7p/S5e	PF00177	148	Ribosomal protein S7p/S5e
Ribosomal protein S8	PF00410	129	Ribosomal protein S8
Ribosomal protein S9/S16	PF00380	121	Ribosomal protein S9/S16
Ribosomal Protein L2	PF03947	130	Ribosomal Proteins L2, C-terminal domain
Ribosomal protein L2	PF00181	77	Ribosomal proteins L2, RNA binding domain
RNA polymerase beta subunit	PF04563	203	RNA polymerase beta subunit
RNA polymerase Rpb1	PF04997	337	RNA polymerase Rpb1, domain 1
RNA polymerase Rpb1	PF00623	166	RNA polymerase Rpb1, domain 2
RNA polymerase Rpb1	PF05000	108	RNA polymerase Rpb1, domain 4
RNA polymerase Rpb2	PF04561	190	RNA polymerase Rpb2, domain 2
RNA polymerase Rpb2	PF04565	68	RNA polymerase Rpb2, domain 3
RNA polymerase Rpb2	PF00562	386	RNA polymerase Rpb2, domain 6
RNA polymerase Rpb2	PF04560	82	RNA polymerase Rpb2, domain 7
RNA polymerase Rpb6	PF01192	57	RNA polymerase Rpb6
Signal peptide binding domain	PF02978	104	Signal peptide binding domain
Translation-initiation factor 2	PF11987	109	Translation-initiation factor 2
TruB family pseudouridylate synthase	PF01509	149	TruB family pseudouridylate synthase
Valine – tRNA ligase	TIGR00422	863	Valine – tRNA ligase

Table S3. Genes not present in *Ca*. N. islandicus, but previously present in the "*Thaumarchaea*-core" as defined by[27]

Locus tag of "core" gene in *N. devanaterra*	Annotation
NDEV_v3_0007|ID:23877774|	DNA polymerase II large subunit
NDEV_v3_0191|ID:23877958|	Protein pelota homolog
NDEV_v3_0273|ID:23878040|	SAM-dependent methyltransferase
NDEV_v3_0276|ID:23878043|	5-carboxymethyl-2-hydroxymuconate Delta-isomerase
NDEV_v3_0416|ID:23878183|	methionine sulfoxide reductase B
NDEV_v3_0872|ID:23878639|	UvrABC system protein C
NDEV_v3_0873|ID:23878640|	ATPase and DNA damage recognition protein of nucleotide excision repair excinuclease UvrABC
NDEV_v3_0874|ID:23878641|	excinulease of nucleotide excision repair, DNA damage recognition component
NDEV_v3_0889|ID:23878656|	Iron-containing alcohol dehydrogenase
NDEV_v3_0897|ID:23878664|	Endoribonuclease L-PSP
NDEV_v3_0915|ID:23878682|	Double-stranded beta-helix fold enzyme
NDEV_v3_0916|ID:23878683|	putative methyltransferase type 11
NDEV_v3_0925|ID:23878692|	Transcriptional regulator, ArsR family
NDEV_v3_0952|ID:23878719|	Disulfide Bond oxidoreductase D family protein
NDEV_v3_1056|ID:23878823|	Rossmann fold nucleotide-binding protein
NDEV_v3_1068|ID:23878835|	CMP/dCMP deaminase zinc-binding
NDEV_v3_1143|ID:23878910|	Oligoendopeptidase, PepF/M3 family
NDEV_v3_1165|ID:23878932|	Pyruvoyl-dependent arginine decarboxylase
NDEV_v3_1181|ID:23878948|	putative SMC domain protein
NDEV_v3_1198|ID:23878965|	Elongation factor Tu domain 2 protein
NDEV_v3_1303|ID:23879070|	Peptide methionine sulfoxide reductase MsrA
NDEV_v3_1775|ID:23879542|	Modification methylase LlaDCHIA
NDEV_v3_1856|ID:23879623|	putative bacterial transferase hexapeptide (Three repeats)
NDEV_v3_1858|ID:23879625|	Glycosyl transferase family protein
NDEV_v3_2071|ID:23879838|	Peptidyl-prolyl cis-trans isomerase
NDEV_v3_2075|ID:23879842|	PfkB domain protein
NDEV_v3_2078|ID:23879845|	Putative pyridoxal phosphate-dependent aminotransferase
NDEV_v3_2106|ID:23879873|	DNA-directed DNA polymerase
NDEV_v3_0018|ID:23877785|	conserved protein of unknown function
NDEV_v3_0103|ID:23877870|	Uncharacterized membrane protein required for N-linked glycosylation (Modular protein)
NDEV_v3_0319|ID:23878086|	conserved protein of unknown function
NDEV_v3_0351|ID:23878118|	Membrane protein-like protein
NDEV_v3_0419|ID:23878186|	conserved protein of unknown function
NDEV_v3_0557|ID:23878324|	conserved protein of unknown function
NDEV_v3_0732|ID:23878499|	conserved protein of unknown function
NDEV_v3_0831|ID:23878598|	conserved protein of unknown function
NDEV_v3_0848|ID:23878615|	conserved protein of unknown function
NDEV_v3_1197|ID:23878964|	conserved protein of unknown function
NDEV_v3_1246|ID:23879013|	conserved membrane protein of unknown function
NDEV_v3_1619|ID:23879386|	protein of unknown function
NDEV_v3_1649|ID:23879416|	protein of unknown function
NDEV_v3_1654|ID:23879421|	conserved protein of unknown function
NDEV_v3_1755|ID:23879522|	protein of unknown function
NDEV_v3_1759|ID:23879526|	conserved protein of unknown function
NDEV_v3_1776|ID:23879543|	protein of unknown function
NDEV_v3_1831|ID:23879598|	protein of unknown function
NDEV_v3_1852|ID:23879619|	conserved protein of unknown function
NDEV_v3_1910|ID:23879677|	Conserved protein of unknown function
NDEV_v3_1936|ID:23879703|	conserved exported protein of unknown function
NDEV_v3_2039|ID:23879806|	protein of unknown function
NDEV_v3_0100|ID:23877867|	Putative nucleic acid binding protein
NDEV_v3_0101|ID:23877868|	Putative nucleic acid binding protein
NDEV_v3_0267|ID:23878034|	protein of unknown function
NDEV_v3_0428|ID:23878195|	exported protein of unknown function
NDEV_v3_0951|ID:23878718|	Redoxin domain protein
NDEV_v3_1248|ID:23879015|	Methyltransferase type 11
NDEV_v3_1377|ID:23879144|	conserved membrane protein of unknown function
NDEV_v3_1561|ID:23879328|	conserved protein of unknown function
NDEV_v3_1613|ID:23879380|	Methionine synthase
NDEV_v3_1614|ID:23879381|	Homocysteine S-methyltransferase
NDEV_v3_1985|ID:23879752|	DNA topoisomerase type IA zn finger domain protein
NDEV_v3_0126|ID:23877893|	ABC transporter, permease component
NDEV_v3_1837|ID:23879604|	conserved protein of unknown function
NDEV_v3_0184|ID:23877951|	DEAD/DEAH box helicase domain protein
NDEV_v3_0023|ID:23877790|	CCA-adding enzyme
NDEV_v3_0924|ID:23878691|	Phosphoribosylaminoimidazole-succinocarboxamide synthase
NDEV_v3_0004|ID:23877771|	conserved protein of unknown function
NDEV_v3_0522|ID:23878289|	Molecular chaperone
NDEV_v3_1051|ID:23878818|	conserved protein of unknown function
NDEV_v3_0413|ID:23878180|	4Fe-4S ferredoxin iron-sulfur binding domain protein
NDEV_v3_1807|ID:23879574|	conserved protein of unknown function
NDEV_v3_0792|ID:23878559|	RNA polymerase Rpb6
NDEV_v3_0699|ID:23878466|	Uncharacterized Zn-finger containing protein
NDEV_v3_1725|ID:23879492|	ThiamineS protein

Table S4. Genome locus tags and annotations of genes discussed in the main text.

Gene	Product	Locus tag
3-hydroxypropionate-4-hydroxybutyrate pathway		
accB/pccB	Acetyl-CoA/ propionyl-CoA carboxylase, carboxyltransferase subunit	AOA3F1_v2_0346
accC/pccC	acetyl-CoA carboxylase, biotin carboxylase subunit	AOA3F1_v2_0345
accA/pccA	biotin carboxyl carrier protein of put. acetyl-CoA/ propionyl-CoA carboxylase	AOA3F1_v2_0344
	putative methylmalonyl-CoA epimerase	AOA3F1_v2_0551
	methylmalonyl-CoA mutase, large subunit	AOA3F1_v2_0552
	methylmalonyl-CoA mutase, small subunit, C-terminus	AOA3F1_v2_1302
	4-hydroxybutyryl-CoA dehydratase	AOA3F1_v2_0024
	4-hydroxybutyryl-CoA dehydratase	AOA3F1_v2_0025
	3-hydroxybutyryl-CoA dehydratase	AOA3F1_v2_0462
	(S)-3-hydroxybutyryl-CoA dehydrogenase (NAD+)	
	Acetoacetyl-CoA thiolase or ketoacyl-CoA thiolase	
Candidate genes for 3-hydroxypropionate-4-hydroxybutyrate carbon fixation pathway including alcohol dehydrogenases		
asd	aspartate-semialdehyde dehydrogenase	AOA3F1_v2_0008
	3-hydroxyacyl-CoA dehydrogenase	AOA3F1_v2_1520
acs	Acetyl-coenzyme A synthetase	AOA3F1_v2_0573
acsA-1	Acetyl-coenzyme A synthetase	AOA3F1_v2_0981
acsA-2	Acetyl-coenzyme A synthetase	AOA3F1_v2_1057
gabD	succinate-semialdehyde dehydrogenase	AOA3F1_v2_0424
	Succinate-semialdehyde dehydrogenase (acetylating)	AOA3F1_v2_0975
	acetoacetyl-CoA thiolase or ketoacyl-CoA thiolase	AOA3F1_v2_0535
	acetoacetyl-CoA thiolase	AOA3F1_v2_0064
	putative CoA-binding protein	AOA3F1_v2_0829
	Protein with CoA-binding domain	AOA3F1_v2_0830
Tricarboxylic acid cycle		
gltA	citrate synthase	AOA3F1_v2_0976
acnA	aconitate hydratase	AOA3F1_v2_0442
	putative isocitrate/isopropylmalate dehydrogenase	AOA3F1_v2_1453
(o)orAC	2:oxoacid-ferredoxin oxidoreductase, fused alpha and gamma subunit	AOA3F1_v2_1720
(o)orB	2:oxoacid-ferredoxin oxidoreductase, beta subunit	AOA3F1_v2_1719
sucD	succinyl-CoA ligase, subunit alpha	AOA3F1_v2_1193
sucC	succinyl-CoA ligase, subunit beta	AOA3F1_v2_1192
sdhA	succinate dehydrogenase flavoprotein subunit/fumarate reductase	AOA3F1_v2_1407
(sdhC)	putative succinate dehydrogenase/fumarate reductase	AOA3F1_v2_1406
(sdhD)	putative succinate dehydrogenase/fumarate reductase	AOA3F1_v2_1405
sdhB	FeS-center protein of succinate dehydrogenase/fumarate reductase	AOA3F1_v2_1404
fumC	fumarate hydratase	AOA3F1_v2_0771
mdh	malate dehydrogenase	AOA3F1_v2_0339
ytsJ	NAD(P)-dependent malic enzyme	AOA3F1_v2_1553
Gluconeogenesis		
pckA	ATP-dependent phosphoenolpyruvate carboxykinase	AOA3F1_v2_0396
ppdk	pyruvate, phosphate dikinase	AOA3F1_v2_1141
eno	enolase	AOA3F1_v2_0437
apgM	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	AOA3F1_v2_0723
(gpmB)	putative 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	AOA3F1_v2_0364
pgk	phosphoglycerate kinase	AOA3F1_v2_1746
(gap)	putative glyceraldehyde-3-phosphate dehydrogenase, phosphorylating	AOA3F1_v2_0467
tpiA	triosephosphate isomerase	AOA3F1_v2_1135
fbp	Inositol-1-monophosphatase/ bifunctional fructose-1,6-bisphosphatase	AOA3F1_v2_1480
11 subunit version of Complex I: type I NADH dehydrogenase (nuoEFG absent)		
nuoA	NADH-quinone oxidoreductase, subunit A	AOA3F1_v2_1711
nuoB	NADH-quinone oxidoreductase, subunit B	AOA3F1_v2_1710
nuoC	NADH-quinone oxidoreductase, subunit C	AOA3F1_v2_1709
nuoD	NADH-quinone oxidoreductase, subunit D	AOA3F1_v2_1708
nuoH	NADH-quinone oxidoreductase, subunit H	AOA3F1_v2_1707
nuoI	NADH-quinone oxidoreductase, subunit I	AOA3F1_v2_1706
nuoJ	NADH-quinone oxidoreductase, subunit J	AOA3F1_v2_1705
nuoK	NADH-quinone oxidoreductase, subunit K	AOA3F1_v2_1703
nuoM	NADH-quinone oxidoreductase, subunit M	AOA3F1_v2_1702
nuoL	NADH-quinone oxidoreductase, subunit L	AOA3F1_v2_1701
nuoN	NADH-quinone oxidoreductase, subunit N	AOA3F1_v2_1699

Table S4 continued

alternative NADH dehydrogenases, type-II NADH dehydrogenase, Coenzyme F420 dependent NADP dehydrogenase, nitroreductases, NADH-dependent FMN reductases, FAD dependent oxidoreductases		
	putative FAD-dependent pyridine nucleotide-disulphide oxidoreductase	AOA3F1_v2_0069
	putative NADH dehydrogenase/NAD(P)H nitroreductase AF_0226	AOA3F1_v2_1059
	FAD/NAD(P)-binding oxidoreductase	AOA3F1_v2_1072
	Geranylgeranyl reductase family protein	AOA3F1_v2_1423
Complex II		
sdhA	succinate dehydrogenase flavoprotein subunit/fumarate reductase	AOA3F1_v2_1407
(sdhC)	putative succinate dehydrogenase/fumarate reductase	AOA3F1_v2_1406
(sdhD)	putative succinate dehydrogenase/fumarate reductase	AOA3F1_v2_1405
sdhB	FeS-center protein of succinate dehydrogenase/fumarate reductase	AOA3F1_v2_1404
Complex III: 1.10.2.2		
petB	cytochrome b/b6 domain	AOA3F1_v2_1784
	Rieske iron sulfur protein	AOA3F1_v2_1785
	putative Rieske (2Fe-2S) domain protein	AOA3F1_v2_1749
Complex IV		
	conserved hypothetical protein	AOA3F1_v2_0365
(coxB)	putative heme-copper oxidase subunit II	AOA3F1_v2_0366
coxA	Cytochrome c oxidase polypeptide 1	AOA3F1_v2_0367
	blue (type 1) copper domain protein	AOA3F1_v2_0368
	putative Blue (Type 1) copper domain protein	AOA3F1_v2_0302
	putative cytochrome oxidase assembly protein	AOA3F1_v2_0369
Complex V: A1A0-type ATPase		
atpE	archaeal A1A0-type ATP synthase, subunit E	AOA3F1_v2_0087
atpA	archaeal A1A0-type ATP synthase, subunit A	AOA3F1_v2_0086
atpB	archaeal A1A0-type ATP synthase, subunit B	AOA3F1_v2_0085
atpD	archaeal A1A0-type ATP synthase, subunit D	AOA3F1_v2_0084
atpK	archaeal A1A0-type ATP synthase, subunit K	AOA3F1_v2_0082
atpI	archaeal A1A0-type ATP synthase, subunit I	AOA3F1_v2_0081
atpC	archaeal A1A0-type ATP synthase, subunit C	AOA3F1_v2_0080
atpF	archaeal A1A0-type ATP synthase, subunit F	AOA3F1_v2_0836
Electron carriers, plastocyanines		
	blue (type 1) copper domain protein	AOA3F1_v2_0368
	blue (type 1) copper domain protein	AOA3F1_v2_1174
	blue (type 1) copper domain protein	AOA3F1_v2_0302
	putative blue (type 1) copper domain protein	AOA3F1_v2_0302
	putative blue (type 1) copper domain protein	AOA3F1_v2_1780
	putative blue (type 1) copper domain protein (fragment)	AOA3F1_v2_1088
Multicopper oxigenases		
	multicopper oxidase type 3 (MCO1)	AOA3F1_v2_1374
Ammonia-monooxygenase		
amoB	putative archaeal ammonia monooxygenase subunit B	AOA3F1_v2_1440
amoC	Ammonia monooxygenase/methane monooxygenase, subunit C	AOA3F1_v2_1439
amoX	conserved hypothetical protein	AOA3F1_v2_1438
amoA	putative archaeal ammonia monooxygenase subunit A	AOA3F1_v2_1437
Urease		
ureD	Urease accessory protein UreD	AOA3F1_v2_0945
ureG	Urease accessory protein UreG	AOA3F1_v2_0944
ureF	Urease accessory protein UreF	AOA3F1_v2_0943
ureE	Urease accessory protein UreE	AOA3F1_v2_0942
ureC	Urease subunit alpha	AOA3F1_v2_0939
ureB	Urease subunit beta	AOA3F1_v2_0938
ureA	Urease gamma subunit	AOA3F1_v2_0937
Nitrilase/cyanide hydratase		
	putative Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase	AOA3F1_v2_0558
Detoxification		
(cld)	putative chlorite dismutase	AOA3F1_v2_1762
sod	Superoxide dismutase [Fe]	AOA3F1_v2_0392

Table S4 continued

Flagellum and Chemotaxis		
(flaK)	putative archaeal preflagellin peptidase FlaK	AOA3F1_v2_0613
flaB	archaeal flagellin	AOA3F1_v2_0578
	hypothetical protein	AOA3F1_v2_0579
flaG	flagellar protein FlaG	AOA3F1_v2_0580
flaF	flagellar protein FlaF	AOA3F1_v2_0581
flaH	flagella protein FlaH	AOA3F1_v2_0582
(flaJ)	putative flagella assembly protein FlaJ	AOA3F1_v2_0583
flaI	archaeal flagella protein FlaI	AOA3F1_v2_0584
(cheR)	putative chemotaxis MCP methyltransferase CheR	AOA3F1_v2_0585
cheD	chemoreceptor glutamine deamidase CheD	AOA3F1_v2_0586
(cheC)	putative chemotaxis protein CheC	AOA3F1_v2_0587
cheA	chemotactic sensor histidine kinase CheA	AOA3F1_v2_0594
cheB	chemotaxis response regulator methylesterase CheB	AOA3F1_v2_0593
cheY	chemotaxis response regulator CheY	AOA3F1_v2_0592
cheW	chemotaxis protein CheW	AOA3F1_v2_0591
(MCP)	putative methyl-accepting chemotaxis protein	AOA3F1_v2_0590
(MCP)	putative methyl-accepting chemotaxis protein	AOA3F1_v2_0589
	putative HEAT repeat-containing PBS lyase	AOA3F1_v2_0153
S assimilation		
(cysA)	putative thiosulfate sulfurtransferase	AOA3F1_v2_1414
sat	Sulfate adenylyltransferase	AOA3F1_v2_0054
(cysC)	putative adenylyl-sulfate kinase	AOA3F1_v2_1343
cysH	Thioredoxin-dependent 5'-adenylylsulfate reductase	AOA3F1_v2_0053
sir	Ferredoxin-sulfite/nitrite reductase	AOA3F1_v2_0949
	Rhodenase-Sulfurtransferase	AOA3F1_v2_0950
Amino acid fermentation		
ilvE	putative branched-chain-amino-acid aminotransferase	AOA3F1_v2_1069
ilvE-2	putative branched-chain-amino-acid aminotransferase	AOA3F1_v2_1492
aspC	Aspartate/ tyrosine/ aromtatic aminotransferase	AOA3F1_v2_1546
gdhA	Glutamate dehydrogenase	AOA3F1_v2_0468
iorB	Indolepyruvate oxidoreductase, beta subunit	AOA3F1_v2_1273
iorA	Indolepyruvate ferredoxin oxidoreductase, alpha subunit	AOA3F1_v2_1272
acs	Acetyl-coenzyme A synthetase	AOA3F1_v2_0573
acsA	Acetyl-coenzyme A synthetase	AOA3F1_v2_0981
acsA	Acetyl-coenzyme A synthetase	AOA3F1_v2_1057
	Acyl-CoA synthetase (NDP forming)	AOA3F1_v2_1415
sir	Ferredoxin-sulfite/nitrite reductase	AOA3F1_v2_0949
korB	2-oxoglutarate synthase subunit beta	AOA3F1_v2_1719
korA	2-oxoglutarate synthase subunit alpha	AOA3F1_v2_1720
Hydrogenase		
(hypA)	putative Hydrogenase nickel incorporation protein HypA	AOA3F1_v2_1388
	putative Hydrogenase 2 maturation protease	AOA3F1_v2_1387
(hydA)	Nickel-dependent hydrogenase alpha subunit	AOA3F1_v2_1386
(hydD)	putative Nickel-dependent hydrogenase delta subunit	AOA3F1_v2_1385
(hydG)	putative Nickel-dependent hydrogenase gamma subunit	AOA3F1_v2_1384
	putative Cyclic nucleotide-binding protein, hydrogenase accessory protein	AOA3F1_v2_1383
(hydB)	putative Nickel-dependent hydrogenase beta subunit	AOA3F1_v2_1382
hypF	putative carbamoyltransferase HypF	AOA3F1_v2_1380
(hypC/hupF)	putative Hydrogenase assembly chaperone HypC/HupF	AOA3F1_v2_1255
DNA replication and repair		
dbh2	DNA polymerase IV	AOA3F1_v2_0742
dpo	DNA polymerase B1	AOA3F1_v2_0645
	putative primase / polymerase	AOA3F1_v2_0930
priB	DNA primase large subunit PriL	AOA3F1_v2_0032
priA	putative DNA primase small subunit PriS	AOA3F1_v2_0030
dnaG	DNA primase DnaG	AOA3F1_v2_0952

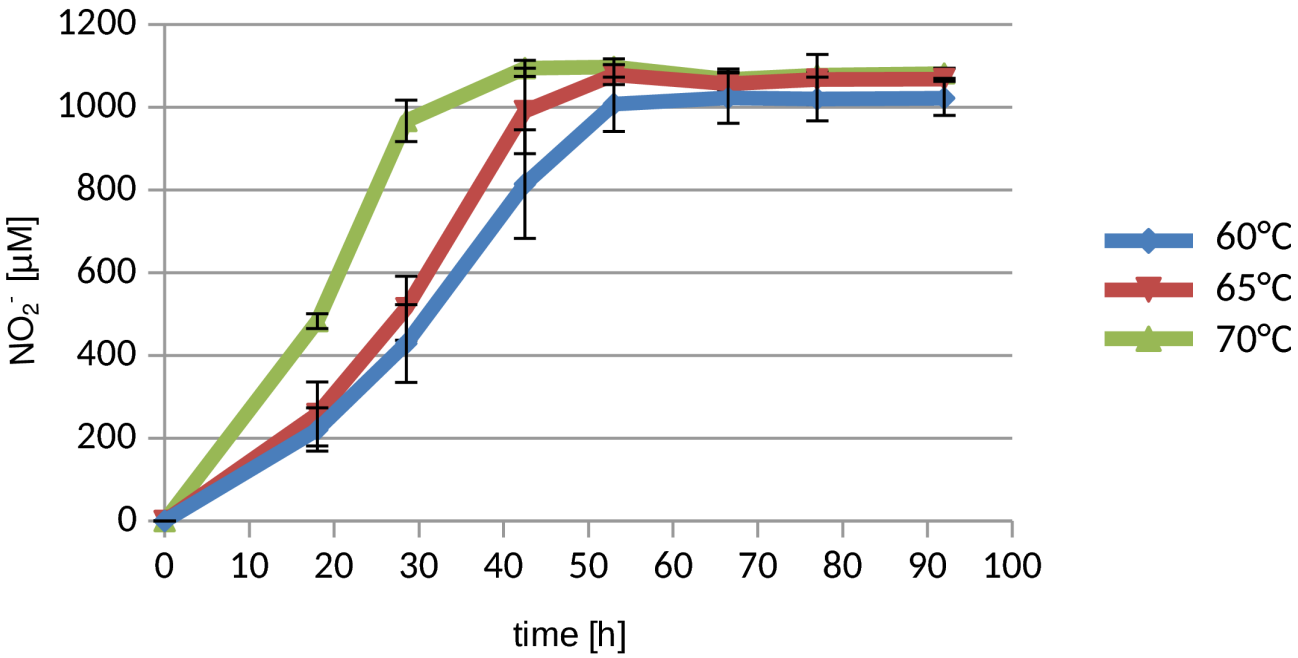

Figure S1. Nitrite accumulation through oxidation of ammonia at three different temperatures. Data points show means, error bars show standard errors of n = 3 biological replicates.

625

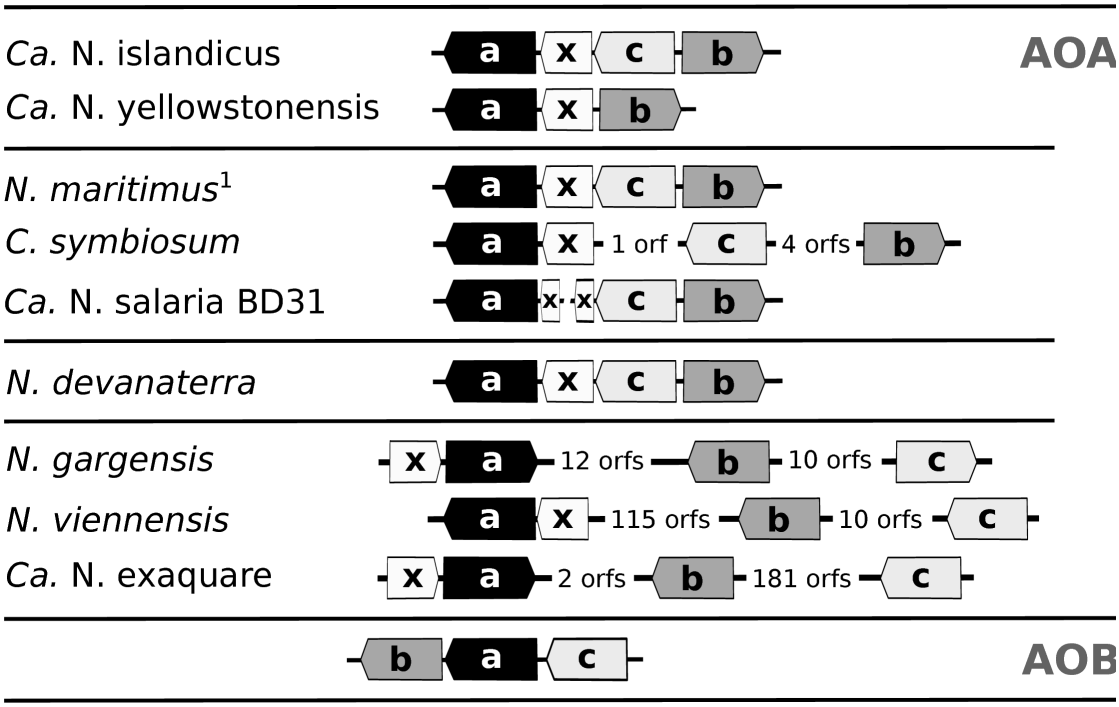

Figure S2. Gene order and orientation of the ammonia monooxygenase subunits (*amoA*, *amoB*, *amoC*, and the putative ‘‘*amoX*’’) in *Ca.* N. islandicus and other Thaumarchaea. The gene order of ammonia-oxidizing bacteria (AOB) is given on the bottom as a reference. [1] also represents the gene arrangement in *Ca.* N. limnia, *Ca.* N. koreensis and *Ca.* N. uzonensis. The figure is a modified version of the figure 26.3 in (*28*)

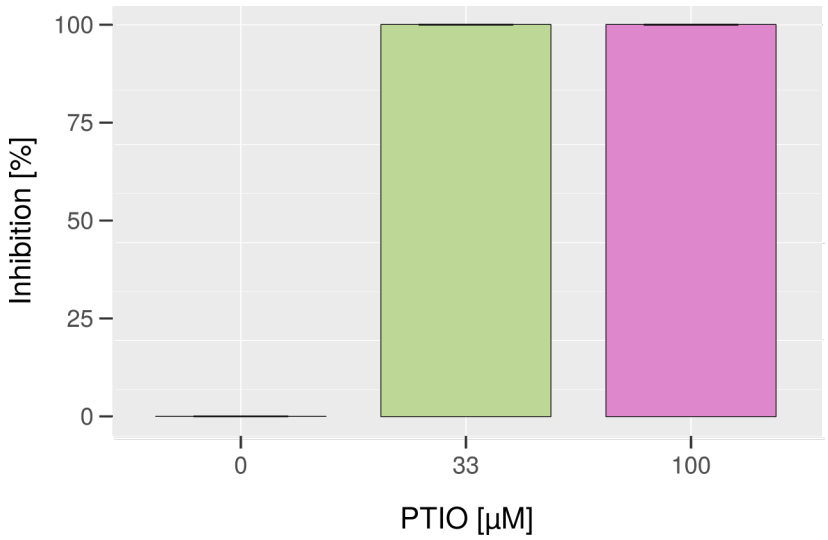

Figure S3. Inhibition of ammonia oxidation by *Ca.* N. islandicus caused by different concentrations of the NO-scavenger 2-phenyl-4,4,5,5-tetramethylimidazoline-1-oxyl 3-oxide (PTIO). Error bars show the standard error of two replicates.

635

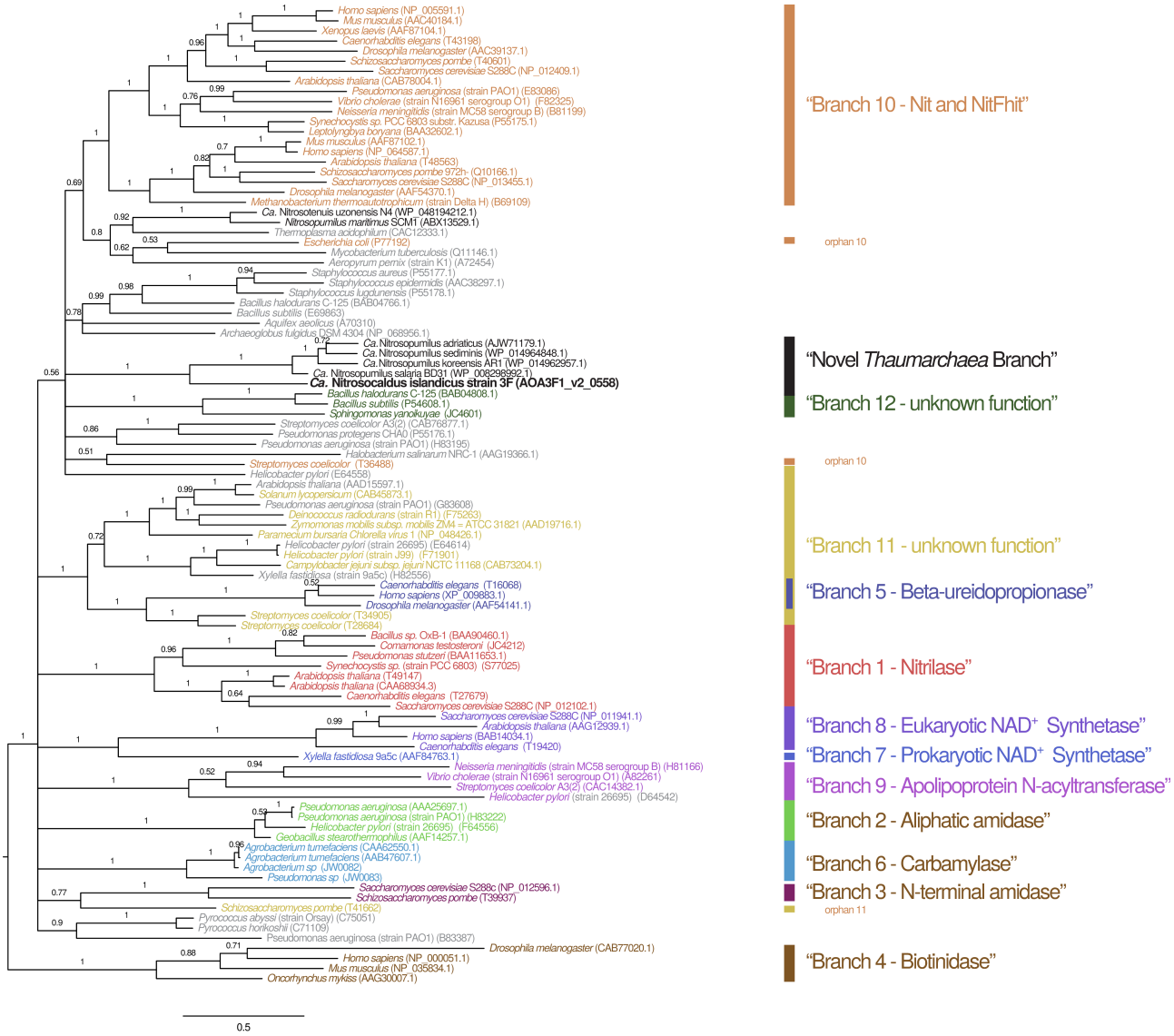

Figure S4. Phylogenetic tree of nitrilase superfamily members showing placement of *Ca.* Nitrosocaldus islandicus strain 3F in a novel branch consisting of *Thaumarchaea* (Novel *Thaumarchaea* branch). Sequences and branch labels are from (*29*). Grey labels indicate "nonfused outliers" as indicated by (*29*) which were not assigned to any of the 12 named branches. Black labels indicate sequences obtained from the *thaumarchaeal* genomes. Sequences assigned to branches 10 and 11 in (*29*) that do not clade with other members of those branches in this phylogenetic tree are labelled "orphan".

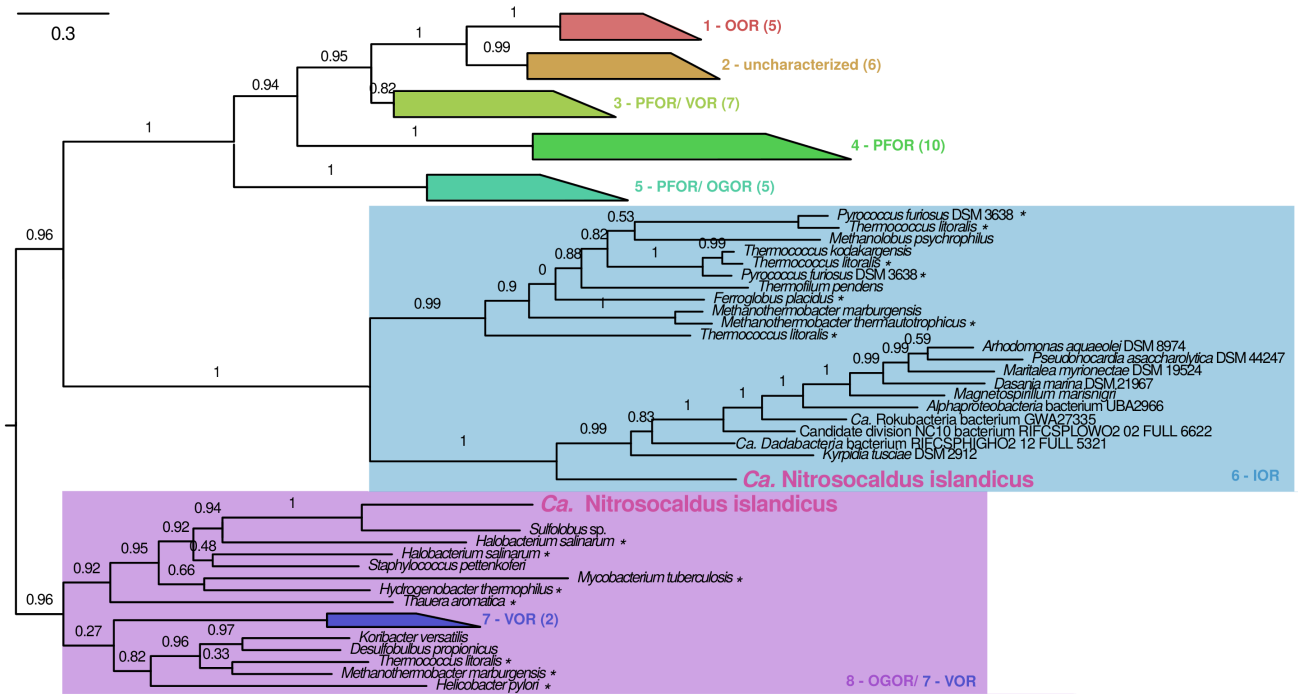

Fig. S5. Unrooted phylogenetic tree of 2-oxoacid oxidoreductases (OFORs) showing placement of the two OFORs present in *Ca.* Nitrosocaldus islandicus strain 3F. Clade labels and most sequences are from (*30*). Numbers in brackets show the number of sequences in collapsed clades. Functionally characterized enzymes are indicated with a "*". OOR, oxalate oxidoreductase; PFOR, pyruvate:ferredoxin oxidoreductase; VOR, 2-ketoisovalerate oxidoreductase; OGOR, 2-oxoglutarate oxidoreductase; IOR, indolepyruvate oxidoreductase.

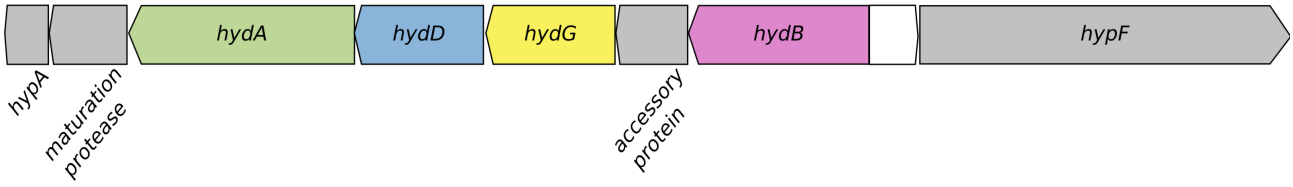

Figure S6. Schematic illustration of the genomic locus in *Ca.* Nitrosocaldus islandicus that encodes a bidirectional, NADP(H)-coupled type 3b [NiFe] hydrogenase. The locus contains the genes of the hydrogenase subunits hydADGB and of accessory proteins involved in enzyme maturation. Genes are drawn to scale. Locus tags (as found on MaGe) from left to right are as follows: AOA3F1_v2_1388, AOA3F1_v2_1387, AOA3F1_v2_1386, AOA3F1_v2_1385, AOA3F1_v2_1384, AOA3F1_v2_1383, AOA3F1_v2_1382, AOA3F1_v2_1381, AOA3F1_v2_1380.

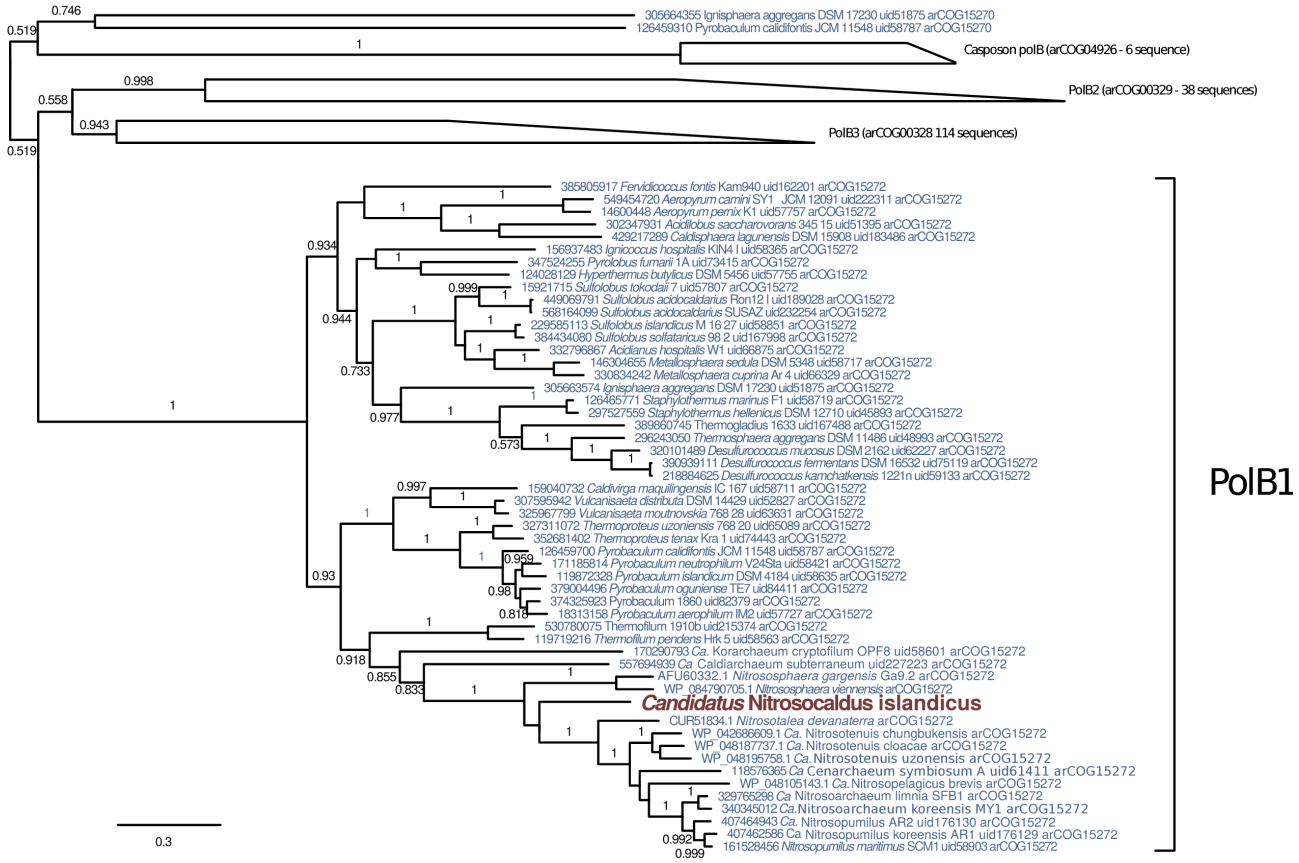

Fig S7. Unrooted approximate maximum likelihood tree showing placement of the PolB from *Ca.* Nitrosocaldus islandicus as a member of the PolB1 clade. The tree was calculated using FastTree[31] on 213 sequences aligned with mafft[32] (3058 aligned positions). Branch support greater than 0.5 is indicated on internal branches. PolB2, PolB3 and Casposon-related PolB (named according to *33*) have been collapsed into right trapezoids in which the bases indicate shortest and longest terminal branches within each clade.

References

1. Hallam, S. J. *et al.* Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc. Natl. Acad. Sci.* **103,** 18296–18301 (2006).

2. Lehtovirta-Morley, L. E. *et al.* Identifying potential mechanisms enabling acidophily in the ammonia-oxidising archaeon '*Candidatus* Nitrosotalea devanaterra'. *Appl. Environ. Microbiol.* AEM.04031-15 (2016). doi:10.1128/AEM.04031-15

3. Blainey, P. C., Mosier, A. C., Potanina, A., Francis, C. A. & Quake, S. R. Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. *PLoS ONE* **6,** (2011).

4. Zhalnina, K. V. *et al.* Genome sequence of *Candidatus* Nitrososphaera evergladensis from group I.1b enriched from Everglades soil reveals novel genomic features of the ammonia-oxidizing archaea. *PLoS ONE* **9,** e101648 (2014).

5. Butterfield, C. N. *et al.* Proteogenomic analyses indicate bacterial methylotrophy and archaeal heterotrophy are prevalent below the grass root zone. *PeerJ* **4,** e2687 (2016).

6. Kamanda Ngugi, D. *et al.* Comparative genomics reveals adaptations of a halotolerant *thaumarchaeon* in the interfaces of brine pools in the Red Sea. *ISME J.* **9,** 396–411 (2015).

7. Haroon, M. F., Thompson, L. R., Parks, D. H., Hugenholtz, P. & Stingl, U. A catalogue of 136 microbial draft genomes from Red Sea metagenomes. *Sci. Data* **3,** (2016).

8. 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).

9. 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.* **112,** 1173–1178 (2015).

10. 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.* **113,** E7937–E7946 (2016).

11. Hug, L. A. *et al.* Critical biogeochemical functions in the subsurface are associated with bacteria from new phyla and little studied lineages. *Environ. Microbiol.* **18,** 159–173 (2016).

12. Ahlgren, N. A. *et al.* Genome and epigenome of a novel marine *Thaumarchaeota* strain suggest viral infection, phosphorothioation DNA modification and multiple restriction systems. *Environ. Microbiol.* n/a-n/a (2017). doi:10.1111/1462-2920.13768

13. Jung, M.-Y. *et al.* A hydrophobic ammonia-oxidizing archaeon of the *Nitrosocosmicus* clade isolated from coal tar-contaminated sediment. *Environ. Microbiol. Rep.* **8,** 983–992 (2016).

14. Mosier, A. C., Allen, E. E., Kim, M., Ferriera, S. & Francis, C. A. Genome sequence of "*Candidatus* Nitrosoarchaeum limnia" BG20, a low-salinity ammonia-oxidizing archaeon from the San Francisco Bay estuary. *J Bacteriol* **194,** 2119–2120 (2012).

15. Mosier, A. C., Allen, E. E., Kim, M., Ferriera, S. & Francis, C. A. Genome sequence of "*Candidatus* Nitrosopumilus salaria" BD31, an ammonia-oxidizing archaeon from the San Francisco Bay estuary. *J. Bacteriol.* **194,** 2121–2122 (2012).

16. Park, S.-J. *et al.* Draft genome sequence of an ammonia-oxidizing archaeon, "*Candidatus* Nitrosopumilus koreensis" AR1, from marine sediment. *J. Bacteriol.* **194,** 6940–6941 (2012).

17. Park, S. J. *et al.* Draft genome sequence of an ammonia-oxidizing archaeon, '*Candidatus* Nitrosopumilus sediminis' AR2, from Svalbard in the arctic circle. *J. Bacteriol.* **194,** 6948–6949 (2012).

18. Kim, B. K. *et al.* Genome sequence of an ammonia-oxidizing soil archaeon, "*Candidatus* Nitrosoarchaeum koreensis" MY1. *J. Bacteriol.* **193,** 5539–5540 (2011).

19. Bayer, B. *et al.* Physiological and genomic characterization of two novel marine *thaumarchaeal* strains indicates niche differentiation. *ISME J.* **10,** 1051–1063 (2016).

20. Sauder, L. A. *et al.* Cultivation and characterization of *Candidatus* Nitrosocosmicus exaquare, an ammonia-oxidizing archaeon from a municipal wastewater treatment system. *ISME J.* **11,** 1142–1157 (2017).

21. Li, Y. *et al.* A novel ammonia-oxidizing archaeon from wastewater treatment plant: Its enrichment, physiological and genomic characteristics. *Sci. Rep.* **6,** (2016).

22. Jung, M.-Y. *et al.* A mesophilic, autotrophic, ammonia-oxidizing archaeon of *thaumarchaeal* group I.1a cultivated from a deep oligotrophic soil horizon. *Appl. Environ. Microbiol.* **80,** 3645–3655 (2014).

23. Spang, A. *et al.* The genome of the ammonia-oxidizing *Candidatus* Nitrososphaera gargensis: insights into metabolic versatility and environmental adaptations. *Environ. Microbiol.* **14,** 3122–3145 (2012).

24. Hugerth, L. W. *et al.* Metagenome-assembled genomes uncover a global brackish microbiome. *Genome Biol.* **16,** 279 (2015).

25. Lebedeva, E. V. *et al.* Enrichment and genome sequence of the group I. 1a ammonia-oxidizing archaeon '*Ca.* Nitrosotenuis uzonensis' representing a clade globally distributed in thermal habitats. *Plos One* **8,** (2013).

26. Mende, D. R., Aylward, F. O., Eppley, J. M., Nielsen, T. N. & DeLong, E. F. Improved environmental genomes via integration of metagenomic and single-cell assemblies. *Front. Microbiol.* **7,** (2016).

27. Herbold CW, Lehtovirta-Morley LE, Jung M-Y, Jehmlich N, Hausmann B, Han P, *et al.* Ammonia-oxidising archaea living at low pH: Insights from comparative genomics. *Environ Microbiol*, in press (2017).

28. Stieglmeier M, Alves RJE, Schleper C. The Phylum Thaumarchaeota. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F (eds). *The Prokaryotes*. Springer Berlin Heidelberg, pp 347–362 (2014).

29. Pace HC and Brenner C. (2001). The nitrilase superfamily: classification, structure and function. *Genome Biol* **2**: reviews0001.1-reviews0001.9.

30. Gibson M, Chen P-T, Drennan CL. (2016). A structural phylogeny for understanding 2-oxoacid oxidoreductase function - ScienceDirect. *Curr Opin Struct Biol* **41**: 54–61.

31. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26,** 1641–50 (2009).

32. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30,** 3059–3066 (2002).

33. Makarova, K. S., Krupovic, M. & Koonin, E. V. Evolution of replicative DNA polymerases in archaea and their contributions to the eukaryotic replication machinery. *Front. Microbiol.* **5,** (2014).

665