bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Candidatus Nitrosocaldus cavascurensis, an ammonia oxidizing, extremely thermophilic archaeon with a highly mobile genome

Sophie S. Abby[1,2&], Michael Melcher[1&], Melina Kerou[1], Mart Krupovic[3], Michaela Stieglmeier[1], Claudia Rossel[1], Kevin Pfeifer[1], Christa Schleper[1,*]

[1] Archaea Biology and Ecogenomics Division, Dep. of Ecogenomics and Systems Biology, University of Vienna, Vienna, Austria.

[2] Univ. Grenoble Alpes, Centre National de la Recherche Scientifique (CNRS), Laboratoire Techniques de l'Ingénierie Médicale et de la Complexité - Informatique, Mathématiques et Applications, Grenoble (TIMC-IMAG), Grenoble, France.

[3] Unité Biologie Moléculaire du Gène chez les Extrêmophiles, Institut Pasteur, Paris, France.

[&]These authors contributed equally

CORRESPONDENCE:
Prof. Christa Schleper
christa.schleper@univie.ac.at

KEYWORDS: ammonia oxidation; archaea; thaumarchaeota; nitrification; mobile genetic elements; thermophily; adaptations

RUNNING TITLE: Thermophilic Thaumarchaeon cultivation and genome

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Ammonia oxidizing archaea (AOA) of the phylum Thaumarchaeota are widespread in moderate environments but their occurrence and activity has also been demonstrated in hot springs. Here we present the first cultivated thermophilic representative with a sequenced genome, which allows to search for adaptive strategies and for traits that shape the evolution of Thaumarchaeota. *Candidatus* Nitrosocaldus cavascurensis has been cultivated from a hot spring in Ischia, Italy. It grows optimally at 68°C under chemolithoautotrophic conditions on ammonia or urea converting ammonia stoichiometrically into nitrite with a generation time of approximately 25h. Phylogenetic analyses based on ribosomal proteins place the organism as a sister group to all known mesophilic AOA. The 1.58 Mb genome of *Ca.* N. cavascurensis harbors an amoAXCB gene cluster encoding ammonia monooxygenase, genes for a 3-hydroxypropionate/4-hydroxybutyrate pathway for autotrophic carbon fixation, but also genes that indicate potential alternative energy metabolisms. Although a bona fide gene for nitrite reductase is missing, the organism is sensitive to NO-scavenging, underlining the importance of this compound for AOA metabolism. *Ca.* N. cavascurensis is distinct from all other AOA in its gene repertoire for replication, cell division and repair. Its genome has an impressive array of mobile genetic elements and other recently acquired gene sets, including conjugative systems, a provirus, transposons and cell appendages. Some of these elements indicate recent exchange with the environment, whereas others seem to have been domesticated and might convey crucial metabolic traits.

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Ammonia oxidizing archaea (AOA) now collectively classified as Nitrososphaeria within the phylum Thaumarchaeota (Brochier-Armanet et al., 2008;Spang et al., 2010;Kerou, 2016) represent the sole archaeal group that is globally distributed in oxic environments efficiently competing with aerobic bacteria. Because of their large numbers in the ocean plankton, in marine sediments, in lakes and in soils, AOA are considered one of the most abundant groups of prokaryotes on this planet (Schleper et al., 2005;Prosser and Nicol, 2008;Erguder et al., 2009;Schleper and Nicol, 2010;Pester et al., 2011;Hatzenpichler, 2012;Stahl and de la Torre, 2012;Offre et al., 2013). All currently cultivated AOA strains gain energy exclusively through the oxidation of ammonia to nitrite, i.e. they perform the first step in nitrification. They grow chemolithoautotrophically from inorganic carbon supply. Some strains show growth only in the presence of small organic acids (Tourna et al., 2011;Qin et al., 2014) which seem to catalyze degradation of reactive oxygen species from the medium (Kim et al., 2016). Different from their bacterial ammonia oxidizing counterparts, AOA are often adapted to rather low levels of ammonia for growth, which seems to favor their activities in oligotrophic environments, such as the marine pelagic ocean and marine sediments, but also in acidic environments, where the concentration of ammonia decreases in favor of ammonium (Nicol et al., 2008). However, AOA occur also in large numbers in terrestrial environments, including fertilized soils and some waste water treatment plants, and several studies indicate alternative energy metabolisms (Mussmann et al., 2011;Alves et al., 2013). Although ammonia oxidation in archaea has not been biochemically resolved, the presence of genes for an ammonia monooxygenase in all AOA with remote similarity to methane and ammonia monooxygenases of bacteria implies involvement of this complex in the process (Konneke et al., 2005;Treusch et al., 2005;Nicol and Schleper, 2006). Hydroxylamine has been suggested to be the first product of ammonia oxidation (Vajrala et al., 2013), but further conversion to nitrite is performed in an unknown process, as no homolog of the bacterial hydroxylamine dehydrogenase has been found in the genomes of AOA. However, nitrous oxide (NO) has been suggested to be involved in the process, because NO production and re-consumption have been observed (Kozlowski et al., 2016b) and the NO scavenger PTIO was shown to inhibit AOA at very low concentrations (Yan et al., 2012;Shen et al., 2013;Martens-Habbena et al., 2015).

Ammonia oxidation by AOA has also been documented to occur in hot springs. With the help of molecular marker genes, diverse ammonia oxidizing thaumarchaea were found in hot terrestrial and marine environments (Reigstad et al., 2008;Zhang et al., 2008;Jiang et al., 2010;Cole et al., 2013). Furthermore, *in situ* nitrification activities up to 84°C were measured using isotopic techniques in AOA-containing habitats in Iceland (Reigstad et al., 2008), in Yellowstone National park (Dodsworth et al., 2011), in a Japanese geothermal water stream (Nishizawa et al., 2016), and in enrichment cultures from Tengchong Geothermal Field in China (Li et al., 2015). In addition, an enrichment of an ammonia oxidizing thaumarchaeon, *Ca.* Nitrosocaldus yellowstonii, from a hot spring in Yellowstone National park grew up to 74 °C confirming that archaeal ammonia oxidation, different from bacterial ammonia oxidation, indeed occurs at high temperatures (de la Torre et al., 2008).

A remarkable diversity of archaea with different metabolisms has been described from terrestrial and marine hot springs and hyperthermophilic organisms tend to be found at the base of almost all lineages of archaea. Therefore, it has often been proposed that the ancestor of all archaea and the ancestors of mesophilic lineages of archaea were hyperthermophilic or at least thermophiles (Barns et al., 1996;Lopez-Garcia et al., 2004;Gribaldo and Brochier-

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Armanet, 2006;Brochier-Armanet et al., 2012;Eme et al., 2013). This has been supported by sequence analyses which suggested a parallel adaptation from hot to moderate temperatures in several lineages of archaea (Groussin and Gouy, 2011;Williams et al., 2017). Indeed, the thermophilic strain *Ca.* Nitrosocaldus yellowstonii emerged in 16S rRNA phylogeny as a sister group of all known mesophilic AOA (de la Torre et al., 2008) indicating that ammonia oxidizing archaea also evolved in hot environments.

However, until now, no genome of an obligate thermophilic AOA was available to analyze in depth the phylogenetic placement of thermophilic AOA and to investigate adaptive features of these ammonia oxidizers that have a pivotal position in understanding the evolution of Thaumarchaeota.

In this work we present the physiology and first genome of an extremely thermophilic AOA of the *Nitrosocaldus* lineage that we cultivated from a hot spring in Southern Italy. We analyze its phylogeny based on ribosomal proteins and reconstruct metabolic and genomic features and adaptations.

Material and Methods

Sampling and enrichment

About 500 mL of mud were sampled from a terrestrial hot spring on the Italian island Ischia at "Terme di Cavascura" and stored at 4 °C until it arrived at the laboratory in Vienna (after about one week, October 2013). The temperature of 77 °C and pH 7-8 were measured *in situ* using a portable thermometer (HANNA HI935005) and pH stripes. Initial enrichments (20mL final volume) were set up in 120 mL serum flasks (two times autoclaved with MilliQ water) containing 14 mL of autoclaved freshwater medium (FWM) (de la Torre et al., 2008;Tourna et al., 2011) amended with non-chelated trace element solution (MTE) (Konneke et al., 2005), vitamin solution, FeNaEDTA (20 µL each), 2 mM bicarbonate, 1 mM ammonium chloride and 2 mL of 0.2 µm filtrated hot spring water. Serum flasks were inoculated with 4 mL (20%) of sampled mud, sealed with grey rubber stoppers (6x boiled and autoclaved) and incubated aerobically at 78 °C while rotating at 100 rpm.

Enrichment Strategies

The temperature was changed after one week to 70 °C and medium amendment with filtrated hot spring water was stopped after four transfers as there was no effect observed. Growth was monitored by microscopy, nitrite production and ammonia consumption. Cultures were transferred at a nitrite concentration of ~700 µM and in case ammonia was depleted before the desired nitrite concentration was reached, cultures were fed with 1 mM ammonium chloride. The antibiotics Sulfamethazine, Rifampicin and Novobiocin (100 $\mu g mL^{-1}$) were used alone and in combination with puruvate, glyoxylate and oxalacetate (0.1 mM), but with little effect on enrichment. Filtration of enrichments with 1.2 µm filters had no or even detrimental effect on AOA abundance when done repeatedly and 0.45 µm filtration sterilized the cultures.

As nitrite production rate increased over time, inoculum size was decreased from 20% to 10% and finally to 5%. Omitting vitamin solution from the medium led to an increase in

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

nitrite production rate and AOA abundance. Most crucial for increasing thaumarchaeal abundance was keeping an exact timing on passaging cultures in late exponential phase (after 4 days) and setting up multiple cultures from which the best (based on microscopic observations) was used for further propagation.

Cultivation

Cultures are routinely grown at 68 °C using 5% inoculum in 20 mL FWM amended with MTE and FeNaEDTA solutions, 1 mM NH_4Cl, 2 mM $NaHCO_3$ and are transferred every 4 days once they reach about 700 µM NO_2^-.

Inhibition test with 2-phenyl-4, 4, 5, 5,-tetramethylimidazoline-1-oxyl 3-oxide (PTIO)

Cultures were grown in 20 mL FWM with 5% inoculum under standard conditions and different amounts of an aqueous PTIO stock solution were added when early to mid-exponential growth phase was reached (ca. 63 h after inoculation; 20, 100 and 400 µM final PTIO concentration). Ammonia oxidation ceased immediately at all applied PTIO concentrations, but cultures with 20 µM PTIO were able to resume nitrite production 72 h after the addition of PTIO.

DNA Extraction

DNA was extracted from cell pellets by bead beating in pre-warmed (65 °C) extraction buffer [40.9 gL^{-1} NaCl, 12.6 gL^{-1} Na_2SO_3, 0.1 M Tris/HCl, 50 mM EDTA, 1% sodium dodecyl sulfate (SDS)] and phenol/ chloroform/ isoamylalcohol [25:24:1 (v/v/v), Fisher BioReagents], in a FastPrep-24 (MP) for 30 s. After centrifugation (10 min, 4 °C, 16000 g) the aqueous phase was extracted with chloroform/ isoamylalcohol [24:1 (v/v)], prior to DNA precipitation with 1 µL of glycogen (20 $mg.mL^{-1}$) and 1 mL of PEG-solution (93.5 gL^{-1} NaCl, 300 gL^{-1} polyethylene glycol 6000) overnight at 4 °C. Nucleic acid pellets were washed, dried, resuspended in nuclease-free water and stored at -20 °C until further use (adapted from (Zhou et al., 1996;Griffiths et al., 2000)).

Amplification of amoA gene

Specific primers were designed for the amplification of the *amo*A gene from *Candidatus* N. cavascurensis: ThAOA-277F: CCA TAY GAC TTC ATG ATA GTC and ThAOA-632R: GCA GCC CAT CKR TAN GTC CA (R. Alves and C. Schleper, unpublished). PCR conditions were 95 °C for 10 min as initialization, followed by 35 cycles of 30 s denaturing at 95 °C, 45 s primer annealing at 55 °C, and elongation at 72 °C for 45 s, finishing with 10 min final elongation at 72 °C.

Quantitative PCR

Archaeal and bacterial 16s rRNA genes were quantified in triplicate 20 µL reactions containing 10 µL GoTaq® qPCR Master Mix 2x (Promega), 0.2 mg mL^{-1} BSA and 1 µM of the appropriate primers : Arch931F [5'-AGG AAT TGG CGG GGG AGC A-3' (Jackson et al., 2001) and Arch1100R [5'-BGG GTC TCG CTC GTT RCC-3' (Ovreas et al., 1997)] for the archaeal 16S rRNA gene, P1 (5'-CCT ACG GGA GGC AGC AG-3') and P2 (5'-ATT

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ACC GCG GCT GCT GG-3') (Muyzer et al., 1993) for the bacterial 16S rRNA gene. Reactions were performed in a realpex^2 (Mastercycler ep gradient S, Eppendorf AG) with the following cycling conditions: 95 °C for 2 min, followed by 40 cycles of 30 s denaturing at 95 °C, 30 s joint annealing-extension at 60 °C, and extension with fluorescence measurement at 60 °C for 30 s. Specificity of qPCR products was confirmed by melting curve analysis. Standards were prepared by cloning 16S rRNA genes from *Ca.* N. cavascurensis and *E. coli* into pGEM®-T Easy Vectors (Promega). These were amplified using M13 primers and concentration was determined with Qubit™ dsDNA BR Assay Kit (Thermo Fisher Scientific) before preparing serial dilutions. The efficiencies of archaeal and bacterial 16S rRNA standards were 95% and 99%, respectively.

Fluorescence in situ hybridization

2 mL samples were centrifuged at 4°C, 16000 g for 40 min, washed in PBS-buffer and fixed with 4% paraformaldehyde for 3 h using standard protocols (Amann et al., 1990). Cells were washed two times with 1 mL PBS and finally resuspended in 200 µL of 1:1 PBS:ethanol mix before storing them at -20 °C. After dehydration in ethanol cells were hybridized overnight in hybridisation buffer with 20% formamide using probes Eub338 (Flous) 5'-GCT GCC TCC CGT AGG AGT-3' (Amann et al., 1990) and Arch915 (dCy3) 5'-GTG CTC CCC CGC CAA TTC CT-3' (Stahl and Amann, 1991).

Scanning Electron Microscopy

For scanning electron microscopy, late exponential cells were harvested from 40 ml of culture by centrifugation (16000 g, 4 °C, 30 min). The cells were resuspended and washed three times (0.02 mM sodium cacodylate) prior to prefixation (0.5% glutaraldehyde in 0.02 mM sodium cacodylate) overnight, after which the concentration of glutaraldehyde was increased to 2.5% for 2h. Fixed cells were spotted onto 0.01%-poly-l-lysine coated glass slides (5 mm) and allowed to sediment for 15 min. Samples were then dehydrated using an ethanol series (30%-100%; 15 min per step) and dried by acetone (2x 100%) and HMDS (2x50%; 2x100%) substitution. The slides were subsequently placed on conductive stubs, gold coated for 30 seconds (JEOL JFC-2300HR) and analyzed (JEOL JSM-IT200).

Genome assembly

DNA was prepared from 480 mL culture using standard procedures and sequenced using a PacBio Sequel sequencer at the VBCF (Vienna BioCenter Core Facilities GmbH). Insert size was 10 kb. Around 500000 reads were obtained with an average size of ~5 kb (N50: 6866 nt, max read length: 84504 nt).

The obtained PacBio reads were assembled using the CANU program (version 1.4, parameters "genomeSize=20m, corMhapSensitivity=normal, corOutCoverage=1000, errorRate=0.013") (Koren et al., 2017), and then "polished" with the arrow program from the SMRT analysis software (Pacific Biosciences, USA). The *Ca.* N. cavascurensis genome consisted of two contiguous, overlapping contigs of 1580543 kb and 15533 kb. A repeated region between both extremities of the longest contig, and the 2nd small contig obtained was found, and analysed using the nucmer program from the MUMMER package (see Figure Sup. 1) (Kurtz et al., 2004). This region was merged using nucmer results, and the longest contig obtained was “circularized” using the information of the nucmer program, as sequence information from the 2nd contig was nearly identical (>99.5%) to that of the extremities of the long contig. The origin of replication was predicted using Ori-Finder 2 webserver (Luo et al.,

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2014). By analogy with *N. maritimus*, it was placed after the last annotated genomic object, before the ORB repeats and the *cdc* gene. The annotated genome sequence has been deposited into GenBank under accession number XXXXX.

Genome dataset

Thaumarchaeota and Aigarchaeaota complete or nearly-complete genomes were downloaded from the NCBI, Genbank database, or IMG database when not available in Genbank. We included the genomes of 27 Thaumarchaeaota: Thaumarchaeota archaeon BS4 (IMG_2519899514); Thaumarchaeota group 1.1c (bin Fn1) (IMG_2558309099); Thaumarchaeota NESA-775 (Offre and Schleper, unpublished); Thaumarchaeota archaeon DS1 (IMG_2263082000); *Cenarchaeum symbiosum* A (DP000238); Thaumarchaeota archaeon CSP1 (LDXL00000000); *Nitrosopumilus adriaticus* NF5 (CP011070); Nitrosocosmicus arcticus Kfb (Alves and Schleper unpublished); *Nitrosopelagicus brevis* CN25 (CP007026); *Nitrosotenuis chungbukensis* MY2 (AVSQ00000000); *Nitrosotalea devanaterra* Nd1 (LN890280); *Nitrososphaera evergladensis* SR1 (CP007174); *Nitrosocosmicus exaquare* G61 (CP017922); *Nitrososphaera gargensis* Ga9.2 (CP002408); *Nitrosoarchaeum koreensis* MY1 (GCF_000220175); *Nitrosopumilus salaria* BD31 (GCF_000242875); *Nitrosopumilus koreensis* AR1 (CP003842); *Nitrosoarchaeum limnia* BG20 (GCF_000241145); *Nitrosoarchaeum limnia* SFB1 (CM001158); *Nitrosopumilus maritimus* SCM1 (CP000866); *Nitrosocosmicus oleophilus* MY3 (CP012850); *Nitrosopumilus piranensis* D3C (CP010868); *Nitrosopumilus sediminis* AR2 (CP003843); *Nitrosotenuis uzonensis* N4 (GCF_000723185); *Nitrososphaera viennensis* EN76 (CP007536); *Nitrosotenuis cloacae* SAT1 (CP011097); *Nitrosocaldus cavascurensis* SCU2 (this paper).

Two Aigarchaeota genomes were included: *Calditenuis aerorheumensis* (IMG_2545555825) and *Caldiarchaeum subterraneum* (BA000048). Additionally, we selected 11 Crenarchaeota genomes from Genbank to serve as an outgroup in our analysis: *Sulfolobus solfataricus* P2 (AE006641); *Pyrobaculum aerophilum* IM2 (AE009441); *Hyperthermus butylicus* DSM 5456 (CP000493); *Thermofilum pendens* Hrk 5 (CP000505); *Acidilobus saccharovorans* 345-15 (CP001742); *Desulfurococcus fermentans* DSM 16532 (CP003321); *Caldisphaera lagunensis* DSM 15908 (CP003378); *Fervidicoccus fontis* Kam940 (CP003423); *Metallosphaera sedula* DSM 5348 (CP000682); *Aeropyrum pernix* K1 (BA000002) and *Thermoproteus tenax* Kra 1 (FN869859).

Genome annotation

The first annotation of the *Ca.* N. cavascurensis genome was obtained by the automatic annotation pipeline MicroScope (Vallenet et al., 2009;Medigue et al., 2017;Vallenet et al., 2017). Annotations from *Nitrososphaera viennensis* EN76 were carried over when aligned sequences were 35% identical (or 30% in syntenic regions) and covered 70% of the sequence. All subsequent manual annotations were held on the platform. Putative transporters were classified by screening against the Transporter Classification Database (Saier et al., 2014). To assist the annotation process, several sets of proteins of interest from Kerou et al. 2016 (EPS, metabolism), Offre et al. 2014 (transporters), Makarova et al. 2016 (archaeal pili and flagella) and ribosomal proteins (Offre et al., 2014;Kerou et al., 2016;Makarova et al., 2016) were screened specifically within our set of genomes using HMMER and HMM protein profiles obtained from databases (TIGRFAM release 15, PFAM release 28, SUPERFAMILY version 1.75), or built from arCOG families (Gough et al., 2001;Haft et al., 2003;Finn et al., 2014;Makarova et al., 2015). In the former case, the trusted cutoff "—cut_tc" was used if

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

defined in the profiles. In the latter case, for each family of interest, sequences were extracted from the arCOG2014 database, aligned using MAFFT ("linsi" algorithm) (Katoh and Standley, 2014), and automatically trimmed at both extremities of conserved sequence blocks using a home-made script relying on blocks built by BMGE v1.12 (BLOSUM 40) (Criscuolo and Gribaldo, 2010). HMMER was then used to screen genomes for different sets of specialized families (Eddy, 2011).

In order to visualize and compare the genetic organization of sets of genes (AMO, Ure, type IV pili), we built HMM profiles (as above except alignments were manually curated for AMO and Ure), based on the corresponding arCOG families and integrated the respective sets of profiles in the MacSyFinder framework that enables the detection of sets of co-localized genes in the genome (parameters inter_gene_max_space = 10, min_nb_mandatory_genes = 1 for AMO and Ure, inter_gene_max_space = 10, min_nb_mandatory_genes = 4 for type IV pili) (Abby et al., 2014). MacSyFinder was then run with the models and profiles on the genome dataset, and the resulting JSON files were used for visualization in the MacSyView web-browser application (Abby et al., 2014). The SVG files generated by MacSyView from the predicted operons were downloaded to create figures.

Reference phylogeny

Ribosomal proteins were identified in genomes using a set of 73 HMM profiles built for 73 arCOG families of ribosomal proteins (see Genome annotation section). When multiple hits were obtained in a genome for a ribosomal protein, the best was selected. The sequences were extracted and each family was aligned using the MAFFT program ("linsi" algorithm) (Katoh and Standley, 2014). Alignments were curated with BMGE (version 1.12, default parameters and BLOSUM 45) (Criscuolo and Gribaldo, 2010) and then checked manually. In several cases, sequences identified as part of a given ribosomal protein family did not align well and were excluded from the analysis. In the end, 59 families with more than 35 genomes represented were selected to build a phylogeny. The corresponding family alignments were concatenated, and a tree was built using the IQ-Tree program version 1.5.5 (Nguyen et al., 2015) (model test, partitioned analysis with the best model selected per gene, 1000 UF-Boot and 1000 aLRT replicates). A tree drawing was obtained with the Scriptree program (Chevenet et al., 2010), and then modified with Inkscape.

Protein families reconstruction and phyletic patterns analysis

Homologous protein families were built for the set of genomes selected. A blast all sequences against all was performed, and the results used as input for the Silix and Hifix programs to cluster the sets of similar sequences into families (Miele et al., 2011;Miele et al., 2012). For sequences to be clustered in a same Silix family, they had to share at least 30% of identity and the blast alignment cover at least 70% of the two sequences lengths. The distribution of protein families of interest was analyzed across the genome dataset.

Phylogenetic analyses

Sequences from Silix families of interest were extracted in fasta files, and a similarity search was performed against a large database of sequences which consisted of 5750 bacterial and archaeal genomes (NCBI Refseq, last accessed in November 2016), and the genomes from our genome dataset. Sequences with a hit having an e-value below 10^{-10} were extracted, and

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the dataset was then de-replicated to remove identical sequences. An alignment was then obtained for the family using MAFFT (linsi algorithm), filtered with BMGE (BLOSUM 30) and a phylogenetic tree was reconstructed by maximum-likelihood using the IQ-Tree program (version 1.5.5, best evolutionary model selected, 1000 UF-Boot and 1000 aLRT replicates). A tree drawing was obtained with the Scriptree program (Chevenet et al., 2010), and modified with Inkscape.

Identification and annotation of integrated MGEs

Integrated mobile genetic elements were identified as described previously (Kazlauskas et al., 2017). Briefly, the integrated MGEs were identified by searching the *Ca.* N. cavascurensis genome for the presence of the hallmark genes specific to mobile genetic elements, such as integrases, casposases, VirB4 ATPases, large subunit of the terminase, major capsid proteins of achaeal viruses, etc. In the case of a positive hit, the corresponding genomic neighborhoods were analysed further. IS elements were searched for using ISfinder (Siguier et al., 2006). The precise borders of integration were defined based on the presence of direct repeats corresponding to attachment sites. The repeats were searched for using Unipro UGENE (Okonechnikov et al., 2012). Genes of integrated MGE were annotated based on PSI-BLAST searches (Altschul et al., 1997) against the non-redundant protein database at NCBI and HHpred searches against CDD, Pfam, SCOPe and PDB databases (Soding, 2005).

Results and Discussion

Enrichment of Ca. Nitrosocaldus cavascurensis

An ammonium oxidizing enrichment culture was obtained from thermal water outflow of the *Terme di Cavascura* on the Italian island Ischia. After repeated transfers into artificial freshwater medium supplemented with 1 mM ammonium and 2 mM bicarbonate, highly enriched cultures were obtained that exhibited almost stoichiometric conversion of ammonium into nitrite within 4 to 5 days (Fig. 1A). A single phylotype of AOA but no AOB or Nitrospira/commamox was identified in the enrichment via amplification and sequencing of *amo*A and 16S rRNA gene fragments (data not shown) and the presence of a single AOA phylotype was confirmed by metagenomic sequencing (see below). Its closest relative in the 16S rRNA database was fosmid 45H12 isolated from a gold mine metagenome with 99% identity and the next closest cultivated relative was *Ca.* Nitrosocaldus yellowstonii with 96% identity (Nunoura et al., 2005;de la Torre et al., 2008). The AOA has been propagated for four years within a stable enrichment culture and its relative enrichment usually ranged between 75 and 90% based on qPCR or cell counts respectively. Diverse attempts to obtain a pure culture (heat treatment, antibiotics, dilution to extinction, filtration) resulted in higher enrichments of up to 92% (based on cell counts) but not in a pure culture. In high throughput sequencing analyses using general prokaryotic primers to amplify the V2/V4 region of the 16S rRNA gene we identified phylotypes of the Deinococcus/Thermus group at up to 97% of the bacterial contaminating sequences (with 96% identity to sequences from the *Thermus* genus) and to minor extent sequences related to the Chloroflexi and Armatimonadetes (not shown). Although Aigarchaeota had been detected in earlier enrichments, the AOA was the only archaeon discovered in the high-level enrichments used in this study.

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The shortest generation time of the thermophilic AOA was approximately 25 h, as measured by nitrite production rates and in qPCR, which is comparable to the generation time of *Ca.* N. yellowstonii (de la Torre et al., 2008). It grew in a temperature range from about 55 to 74°C with an apparent (but relatively wide) temperature optimum around 68°C (Fig. 2).

In fluorescence *in situ* hybridizations (FISH) with archaea- and bacteria-specific probes, all coccoid-shaped cells of less than 1 µm in diameter were assigned to the AOA, while all shorter and longer rod-shaped morphotypes were clearly assigned to bacteria (Fig. 3A, B). Scanning electron microscopy revealed the typical spherical and irregular shape of cocci, as seen e.g., for the ammonia oxidizing archaeon *Nitrososphaera viennensis* (Tourna et al., 2011) or hyperthermophilic and halophilic coccoid archaea strains with a diameter of around 0.7 µm (Fig. 3C, D). Based on its relationship with *Ca.* Nitrosocaldus yellowstonii (de la Torre et al., 2008), and the location it was sampled from (*Terme di cavascura*, Italy) we named this organism provisionally *Candidatus* Nitrosocaldus cavascurensis.

Ca. *Nitrosocaldus cavascurensis represents a deeply branching lineage of AOA*

We gathered all complete or nearly-complete 27 genome sequences available for Thaumarchaeota. Those included 23 genomes of closely-related cultivated or uncultivated AOA all harboring *amo* genes, and four genomes obtained from metagenomes that do not have genes for ammonia oxidation. Among the latter were two assembled genomes from moderate environments (Fn1 and NESA-775) and two from hot environments (BS4 and DS1) (Beam et al., 2014;Lin et al., 2015). In addition, two Aigarchaeota genomes, and a selection of representative genomes of Crenarchaeota (11) were included to serve as outgroups (Petitjean et al., 2014;Raymann et al., 2015;Adam et al., 2017;Williams et al., 2017). This resulted in a 40 genomes dataset (Material and Methods). A maximum likelihood analysis based on 59 concatenated ribosomal proteins (in one copy in at least 35 of the 40 genomes) resulted in a highly-supported tree (UF-boot and aLRT support) (Fig. 3). It confirms the monophyly of Thaumarchaeota and Nitrososphaeria (AOA), and the deeply-branching position of *Ca.* N. cavascurensis as a sister-group of all other (mesophilic) AOA, as initially indicated with a single gene (16S rRNA) phylogeny of *Ca.* Nitrosocaldus yellowstonii (de la Torre et al., 2008).

Genome and energy metabolism of Ca. *Nitrosocaldus cavascurensis*

The genome of *Ca*. N. cavascurensis contains 1.58 Mbp, with 1748 predicted coding sequences, one 16S/23S rRNA operon and 29 tRNA genes. It has a G+C content of 41.6%.

Similar to the genomes of most marine strains (Nitrosopumilales) and to *Ca*. Nitrosocaldus yellowstonii, but different from those of the terrestrial organisms (Nitrososphaerales) it encodes all putative subunits of the ammonia monooxygenase in a single gene cluster of the order *amoA*, *amoX*, *amoC*, *amoB* (Fig. 4) indicating that this might represent an ancestral gene order. Different from several other AOA it has a single copy of *amoC*. The genome contains a cluster of genes for the degradation of urea, including the urease subunits and two urea transporters (Fig. 4 and 5) with a similar structure to the urease locus of

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Nitrosophaerales. Accordingly, growth on urea, albeit slower than on ammonia, could be demonstrated (Fig. 1B). Urease loci were so far found in all Nitrososphearales genomes, and in some Nitrosopumilales (Hallam et al., 2006;Park et al., 2012;Spang et al., 2012;Bayer et al., 2016). No urease cluster could be found in non-AOA Thaumarchaeota, nor in Aigarchaeota. In Crenarchaeota, only one Sulfolobales genome harbored a urease (*Metallosphaera sedula* DSM 5348). A specific protein family of a putative chaperonin (Hsp60 homolog, arCOG01257) was found to be conserved within the urease loci of *Ca.* N. cavascurensis and the Nitrososphaerales ("soil-group" of AOA) (Fig. 4) as also observed in some bacteria (e.g., in *Haemophilus influenzae*). Unlike the conserved "thermosome" Hsp60 (also part of arCOG01257), this particular Hsp60 homolog was not found in any Nitrosopumilales genome. Given the conservation of the urease loci between *Ca.* Nitrosocaldus and Nitrosophaerales, it is possible that these genes have a common origin, and were acquired *en bloc* from bacteria by the ancestor of AOA. It is likely that the Hsp60 homolog is involved in stabilization of the urease, as demonstrated in the deltaproteobacterium *Helicobacter pylori* (Evans et al., 1992).

Intriguingly, a gene encoding a nitrite reductase that is present in all genomes of AOA (except the sponge symbiont, *Ca.* Cenarchaeum symbiosum, (Bartossek et al., 2010)) and which is highly transcribed during ammonia oxidation in *N. viennensis* (Kerou et al., 2016) and in meta-transcriptomic datasets (Shi et al., 2011) is missing from the genome. Nitrite reductases have been postulated to be involved in ammonia oxidation by providing nitric oxide (NO) for the oxidation step to nitrite (Kozlowski et al., 2016a). We therefore tested, if the organism was affected by the NO-scavenger PTIO which inhibited at low concentrations all AOA tested so far (Martens-Habbena et al., 2015;Kozlowski et al., 2016a). Ammonia oxidation of *Ca.* N. cavascurensis was fully inhibited at concentrations as low as 20 µM, similar to those affecting *N. viennensis* and *N. maritimus* (Fig. 6). This indicates that NO is produced by an unknown nitrite reductase or perhaps by an unidentified hydroxylamine dehydrogenase (HAO), as recently shown for the HAO of the ammonia oxidizing bacterium *Nitrosomonas europaea* (Caranto and Lancaster, 2017). Alternatively, NO might be supplied through the activity of other organisms in the enrichment culture. Indeed we found a *nirK* gene in the genome of the Deinococcus species of our enrichment culture. The sensitivity to PTIO reinforces the earlier raised hypothesis that NO represents an important intermediate or cofactor in ammonia oxidation in archaea (Schleper and Nicol, 2010;Simon and Klotz, 2013;Kozlowski et al., 2016b).

Pathways of central carbon metabolism and -fixation of *Ca.* N. cavascurensis were found to be generally similar to those of other AOA, underlining the strikingly high similarity and conservation of metabolism within the Nitrososphaeria (Fig. 5) (Kerou et al., 2016)).

Key enzymes of the 3-hydroxypropionate/4-hydroxybutyrate carbon fixation pathway were identified in the genome of *Ca.* N. cavascurensis (supplementary table 1, (Konneke et al., 2014)). Intriguingly, and similar to *Sulfolobales*, we find two copies of the gene encoding for 4-hydoxybutyryl-CoA dehydratase located in tandem. The first copy (NCAV_0127) is homologous to the other AOA sequences which are more closely related to fermenting Clostridia rather than their crenarchaeal counterparts ("Cren Type-1", Fig. S2), while the

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

second copy (NCAV_0126) exhibits high similarity to genes from bacterial candidate division NC10 and Handelsmanbacteria (49% identity). In a phylogenetic tree of the protein family, it clusters together with the second group of crenarchaeal genes ("Cren Type-2", Fig. S2, Material and Methods) which lack essential catalytic residues and which function remains unknown (Ramos-Vera et al., 2011). Könneke et al. 2014 suggested an independent emergence of the cycle in Thaumarchaeota and autotrophic Crenarchaeota based on the unrelatedness of their respective enzymes (Konneke et al., 2014). But the existence of the second gene in this deep-branching lineage rather indicates that both genes could have been present in a common ancestor of Crenarchaeota and Thaumarchaeota and that their CO_2 fixation pathways could have a common origin. No homolog of the experimentally characterized malonic semialdehyde reductase from *N. maritimus* (Otte et al., 2015), or any other iron-containing alcohol dehydrogenase protein family member, was found in the *Ca.* N. cavascurensis genome, indicating potentially the later recruitment of this enzyme, or a case of non-orthologous gene replacement. A full oxidative TCA cycle is present in *Ca.* N. cavascurensis, including a malic enzyme and a pyruvate/phosphate dikinase connecting the cycle to gluconeogenesis. Additionally, and like most mesophilic AOA, *Ca.* N. cavascurensis encodes a class III polyhydroxyalkanoate synthase (phaEC) allowing for the production of polyhydroxyalkanoates, carbon polyester compounds that form during unbalanced growth and serve as carbon and energy reserves (Poli et al., 2011;Spang et al., 2012) (Supplementary Table 1 and Fig. 5).

The presence of a full set of genes encoding for the four subunits (and maturation factors) of a soluble type 3b [NiFe]-hydrogenase, uniquely in *Ca.* N. cavascurensis, indicates the ability to catalyze hydrogen oxidation potentially as part of the energy metabolism. This group of hydrogenases is typically found among thermophilic archaea, is oxygen-tolerant and bidirectional, and can couple oxidation of H_2 to reduction of NAD(P) in order to provide reducing equivalents for biosynthesis, while some have been proposed to have sulfhydrogenase activity (Kanai et al., 2011;Peters et al., 2015;Greening et al., 2016) (and references therein). Although classified by sequence analysis (Sondergaard et al., 2016) and subunit composition as a type 3b hydrogenase, the alpha and delta subunits belong to the arCOG01549 and arCOG02472 families respectively, which contain coenzyme F_{420}-reducing hydrogenase subunits, so far exclusively found in methanogenic archaea. Given the fact that Thaumarchaeota can synthesize this cofactor and encode a number of F_{420}-dependent oxidoreductases with a yet unknown function, it is interesting to speculate whether oxidized F_{420} could also be a potential substrate for the hydrogenase. Expression of the hydrogenase is likely regulated through a cAMP-activated transcriptional regulator encoded within the hydrogenase gene cluster (Supplementary Table 1).

Adaptations to thermophilic life

The molecular adaptations that enable survival and the maintenance of cell integrity at high temperatures have been the subject of intense studies since the discovery of thermophilic organisms. The issue of extensive DNA damage occurring at high temperatures has led to the study of systems of DNA stabilization and repair in thermophilic and hyperthermophilic archaea. Among them, the reverse gyrase, a type IA DNA topoisomerase shown to stabilize and protect DNA from heat-induced damage, is often (but not always (Brochier-Armanet and Forterre, 2007)) found in thermophiles, and is even considered a hallmark of hyperthermophilic organisms growing optimally above 80°C (Bouthier de la Tour et al., 1990;Forterre, 2002). However, the gene might not always be essential for survival at high temperature in the laboratory (Atomi et al., 2004;Brochier-Armanet and Forterre, 2007).

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Interestingly, we could not identify a gene encoding for reverse gyrase in the genome of *Ca.* N. cavascurensis. This might either reflect that its growth optimum is at the lower end of extreme thermophiles or that there is a separate evolutionary line of adaptation to thermophily possible without reverse gyrase.

The following DNA repair mechanisms were identified in the genome of *Ca.* N. cavascurensis, in agreement with the general distribution of these systems in thermophiles (for reviews see (Rouillon and White, 2011;Grasso and Tell, 2014;Ishino and Narumi, 2015)):

a) Homologous recombination repair (HR): Homologs of RadA and RadB recombinase, Mre11, Rad50, the HerA-NurA helicase/nuclease pair, and Holliday junction resolvase Hjc are encoded in the genome. These genes have been shown to be essential in other archaeal (hyper)thermophiles, leading to hypotheses regarding their putative role in replication and more generally the tight integration of repair, recombination and replication processes in (hyper)thermophilic archaea (Grogan, 2015).

b) Base excision repair (BER): the machinery responsible for the repair of deaminated bases was identified in the genome, including uracil DNA glycosylases (UDG) and putative apurinic/apyrimidinic lyases. Deletion of UDGs was shown to impair growth of (hyper)thermophilic archaea (Grogan, 2015).

c) Nucleotide excision repair (NER): Homologs of the putative DNA repair helicase: nuclease pair XPB-Bax1 and nuclease XPF were found in the genome, but repair nuclease XPD could not be identified. XPD is present in all so far analyzed mesophilic AOA, but it is also absent in other thermophiles (Kelman and White, 2005). It should be noted that NER functionality in archaea is still unclear, and deletions of the respective genes were shown to have no observable phenotype (Grogan, 2015).

d) Translesion polymerase: A Y-family polymerase with low-fidelity able to perform translesion DNA synthesis is encoded in *Ca.* N. cavascurensis.

Key enzymes of all the above-mentioned systems are also found in mesophilic AOA. Given their extensive study in the crenarchaeal (hyper)thermophiles, it would be interesting to characterize their respective functions and regulation in both mesophilic and thermophilic Thaumarchaeota.

e) Bacterial-type UvrABC excision repair: In contrast to mesophilic AOA and in agreement with the known distribution of the system among mesophilic archaea and bacteria but its absence in (hyper)thermophiles, *Ca.* N. cavascurensis does not encode homologs of this repair machinery.

Ca. N. cavascurensis could potentially produce the polyamines putrescine and spermidine, which have been shown to bind and stabilize compacted DNA from thermal denaturation, acting synergistically with histone molecules, also present in AOA (Higashibata et al., 2000;Oshima, 2007). Although a putative spermidine synthase is also found in mesophilic AOA, the gene encoding for the previous step, S-adenosylmethionine decarboxylase (NCAV_0959), is only found in *Ca.* N. cavascurensis, located in tandem with the former. The biosynthesis of putrescine (a substrate for spermidine synthase) is unclear, since we could not identify a pyruvoyl-dependent arginine decarboxylase (ADC, catalyzing the first of the two-step biosynthesis of putrescine). However, it was shown that the crenarchaeal arginine decarboxylase evolved from an *S*-adenosylmethionine decarboxylase enzyme, raising the possibility of a promiscuous enzyme (Giles and Graham, 2008).

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The production of thermoprotectant compounds with a role in stabilizing proteins from heat denaturation seems to be a preferred strategy of heat adaptation in *Ca.* N. cavascurensis. Firstly, the presence of mannosyl-3-phosphoglycerate synthase (NCAV_1295) indicates the ability to synthesize this compatible solute, shown to be involved in heat stress response and protect proteins from heat denaturation in Thermococcales (Neves et al., 2005). Homologous genes are also present in mesophilic members of the order Nitrososphaerales and other (hyper)thermophiles (Spang et al., 2012;Kerou et al., 2016). Secondly, only *Ca.* N. cavascurensis, but no other AOA encodes a cyclic 2, 3-diphosphoglycerate (cDPG) synthetase (NCAV_0908), an ATP-dependent enzyme which can synthesize cDPG from 2,3-biphosphoglycerate, an intermediate in gluconeogenesis. High intracellular concentrations of cDPG accumulate in hyperthermophilic methanogens, where it is required for the activity and thermostability of important metabolic enzymes (Shima et al., 1998).

Notable features of the DNA replication and cell division systems in Ca. *N. cavascurensis*

Strikingly, only one family B replicative polymerase PolB was identified in the genome of *Ca.* N. cavascurensis (NCAV_1300), making it the only archaeon known so far to encode a single subunit of the replicative family B polymerase, as in Crenarchaeota multiple paralogs with distinct functions coexist (Makarova et al., 2014). Both subunits of the D-family polymerases PolD present in all other AOA and shown to be responsible for DNA replication in *T. kodakarensis* (also encoding both polD and polB families) (Cubonova et al., 2013) were absent from the genome, raising intriguing questions about the role of the polB family homolog in mesophilic Thaumarchaeota. Sequence analysis indicated that the *Ca.* N. cavascurensis homolog belongs to the polB1 group present exclusively in the TACK superphylum and shown recently by Yan and colleagues to be responsible for both leading and lagging strand synthesis in the crenarchaeon *Sulfolobus solfataricus* (Yan et al., 2017). However, the activity of *Sulfolobus solfataricus* PolB1 is determined based on the presence and binding of two additional proteins, PBP1 and PBP2, mitigating the strand-displacement activity during lagging strand synthesis and enhancing DNA synthesis and thermal stability of the holoenzyme, respectively. No homologs of these two additional subunits were identified in *Ca.* N. cavascurensis, raising the question of enzymatic thermal stability and efficiency of DNA synthesis on both strands.

Homologs of genes for the Cdv cell division system proteins CdvB (3 paralogues) and CdvC, but not CdvA, were identified in *Ca.* N. cavascurensis. This is surprising given that all three proteins were detected by specific immuno-labelling in the mesophilic AOA *N. maritimus*, where CdvA, CdvC and two CdvB paralogues were shown to localize mid-cell in cells with segregated nucleoids, indicating that this system mediates cell division in thaumarchaeota (Pelve et al., 2011). The *Ca.* N. cavascurensis CdvB paralogues (as all other thaumarchaeal CdvBs) all share the core ESCRT-III with the crenarchaeal CdvB sequences, contain a putative (but rather unconvincing) MIM2 motif necessary for interacting with CdvC in Crenarchaeota, while they all lack the C-terminal winged-helix domain responsible for interacting with CdvA (Samson et al., 2011;Ng et al., 2013). Interestingly, one of the *Ca.* N. cavascurensis paralogs (NCAV_0805) possesses a 40 amino-acids serine-rich C-terminal extension right after the putative MIM2 motif absent from other thaumarchaea. It is worth noting that CdvA is also absent from the published Aigarchaeota genomes *Ca.* Caldiarchaeum subterraneum (Nunoura et al., 2011) and *Ca.* Calditenuis aerorheumensis (Beam et al., 2016), while both phyla (Thaumarchaeota and Aigarchaeota) encode an atypical FtsZ homolog. Thermococcales, albeit they presumably divide with the FtsZ system, also

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

encode CdvB and CdvC homologs, while no CdvA homolog is detectable (Makarova et al., 2010). Given the emerging differences in the molecular and regulatory aspects of the Cdv system between Crenarchaeota and Thaumarchaeota (Pelve et al., 2011;Ng et al., 2013), the intriguing additional roles of the system (Samson et al., 2017) and the fact that only CdvB and CdvC are homologous to the eukaryotic ESCRT-III system and therefore seem to have fixed roles in evolutionary terms, this observation raises interesting questions regarding the versatility of different players of the cell division apparatus.

Ca. N. cavascurensis has a dynamic genome

The *Ca.* N. cavascurensis genome contains large clusters of genes that showed deviations from the average G+C content of the genome (*i.e.*, 41.6%), indicating that these regions might have been acquired by lateral gene transfer (Fig. 7).

Two of the larger regions were integrative and conjugative elements (ICE-1 and ICE-2 in Figures 7 and 8) of 65.5 kb and 43.6 kb, respectively. Both are integrated into tRNA genes and flanked by characteristic 21 bp and 24 bp-long direct repeats corresponding to the attachment sites, a typical sequence feature generated upon site-specific recombination between the cellular chromosome and a mobile genetic element (Grindley et al., 2006). ICE-1 and ICE-2 encode major proteins required for conjugation (colored in red in Fig. 8A), including VirD4-like and VirB4-like ATPases, VirB6-like transfer protein and VirB2-like pilus protein (in ICE-2). The two elements also share homologs of CopG-like proteins containing the ribbon-helix-helix DNA-binding motifs and beta-propeller-fold sialidases (the latter appears to be truncated in ICE-1). The sialidases of ICE-1 and ICE-2 are most closely related to each other (37% identity over 273 aa alignment). It is not excluded that ICE-1 and ICE-2 have inherited the conjugation machinery as well as the genes for the sialidase and the CopG-like protein from a common ancestor, but have subsequently diversified by accreting functionally diverse gene complements. Indeed, most of the genes carried by ICE-1 and ICE-2 are unrelated. ICE-1, besides encoding the conjugation proteins, carries many genes for proteins involved in DNA metabolism, including an archaeo-eukaryotic primase-superfamily 3 helicase fusion protein, Cdc6-like ATPase, various nucleases (HEPN, PD-(D/E)XK, PIN), DNA methyltransferases, diverse DNA-binding proteins and an integrase of the tyrosine recombinase superfamily (yellow in Fig. 8A). The conservation of the attachment sites and the integrase gene as well as of the conjugative genes indicates that this element is likely to be a still-active Integrative and Conjugative Element (ICE) able to integrate the chromosome and excise from it as a conjugative plasmid.

ICE-2 is shorter and encodes a distinct set of DNA metabolism proteins, including topoisomerase IA, superfamily 1 helicase, ParB-like partitioning protein, GIY-YIG and FLAP nucleases as well as several DNA-binding proteins. More importantly, this element also encodes a range of proteins involved in various metabolic activities as well as a Co/Zn/Cd efflux system that might provide relevant functions to the host (green in Fig. 8A). In bacteria, ICE elements often contain cargo genes that are not related to the ICE life cycle and that confer novel phenotypes to host cells (Johnson and Grossman, 2015). It is possible that under certain conditions, genes carried by ICE-1 and ICE-2, and in particular the metabolic genes of ICE-2, improve the fitness of *Ca.* N. cavascurensis. Notably, only ICE-1 encodes an integrase, whereas ICE-2 does not, suggesting that ICE-2 is an immobilized conjugative element that can be vertically inherited in AOA. Given that the attachment sites of the two elements do not share significant sequence similarity, the possibility that ICE-2 is mobilized *in trans* by the integrase of ICE-1 appears unlikely.

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The third mobile genetic element is derived from a virus, related to members of the viral order Caudovirales (Fig. 8B) (Prangishvili et al., 2017). It encodes several signature proteins of this virus group, most notably the large subunit of the terminase, the portal protein and the HK97-like major capsid protein (and several other viral homologs). All these proteins with homologs in viruses are involved in virion assembly and morphogenesis. However, no proteins involved in genome replication seem to be present. The element does not contain an integrase gene, nor is it flanked by attachment sites, which indicates that it is immobilized. Interestingly, a similar observation has been made with the potential provirus-derived element in *N. viennensis* (Krupovic et al., 2011). Given that the morphogenesis genes of the virus appear to be intact, one could speculate that these elements represent domesticated viruses akin to gene transfer agents, as observed in certain methanogenic euryarchaea (Eiserling et al., 1999;Krupovic et al., 2010), or killer particles (Bobay et al., 2014), rather than deteriorating proviruses. Notably, ICE-1, ICE-2 and the provirus-derived element all encode divergent homologs of TFIIB, a transcription factor that could alter the promoter specificity to the RNA polymerase.

A fourth set of unique, potentially transferred genes encodes a putative pilus (Fig. 8C). All genes required for the assembly of a type IV pilus (T4P) are present, including an ATPase and an ATPase binding protein, a membrane platform protein, a prepilin peptidase, and several prepilins (Makarova et al., 2016). Based on the arCOG family of the broadly conserved ATPase, ATPase-binding protein, membrane-platform protein, and prepilin peptidase (arCOG001817, -4148, -1808, and -2298), this pilus seems unique in family composition, but more similar to the archetypes of pili defined as clades 4 (A, H, I, J) by Makarova and colleagues, which are mostly found in Sulfolobales and Desulfurococcales (Fig. 4 from (Makarova et al., 2016)). Yet, the genes associated to this putative pilus seem to be more numerous, as the locus consists of approximately 16 genes (versus ~5 genes for the aap and ups in *Sulfolobus*). Such a combination of families is not found either in the T4P or flagellar loci found in analysed Thaumarchaeota genomes. Prepilins are part of the core machinery of T4P, but display a high level of sequence diversity. The three prepilins found in *Ca*. N. cavascurensis T4P locus correspond to families that are not found in any other genomes from our dataset (arCOG003872, -5987, -7276). We found a putative adhesin right in between the genes encoding the prepilins, and therefore propose that this *Ca*. N. cavascurensis-specific type IV pilus is involved in adhesion (Fig. 8C). This pilus thus appears to be unique in protein families composition when compared to experimentally validated T4P homologs in archaea (flagellum, ups, aap, bindosome (Szabo et al., 2007;Zolghadr et al., 2007;Frols et al., 2008;Tripepi et al., 2010;Henche et al., 2012)), bioinformatically predicted ones (Makarova et al., 2016), and the pili found in other Thaumarchaeota, which correspond to different types. Interestingly, this pilus gene cluster lies directly next to a conserved chemotaxis/archaellum cluster as the one found in *N. gargensis* or *N. limnia* (four predicted genes separate the T4P genes and *cheA*, Fig. 8C) (Spang et al., 2012). This suggests that this pilus might be controlled by exterior stimuli through chemotaxis. The interplay between the archaellum and pilus expression would be interesting to study in order to comprehend their respective roles.

The genome of *Ca.* N. cavascurensis also carries traces of inactivated integrase genes as well as transposons related to bacterial and archaeal IS elements, suggesting that several other types of mobile genetic elements have been colonizing the genome of *Ca*. N. cavascurensis.

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Collectively, these observations illuminate the flexibility of the *Ca.* N. cavascurensis genome, prone to lateral gene transfer and invasion by alien elements. Accordingly, we found a CRISPR-Cas adaptive immunity system among the sets of genes specific to *Ca*. N. cavascurensis that we could assign to the subtype I-B (Abby et al., 2014). We detected using the CRISPRFinder website (Grissa et al., 2007) at least three CRISPR arrays containing between 4 and 101 spacers presumably targeting mobile genetic elements associated with *Ca.* N. cavascurensis, reinforcing the idea of a very dynamic genome. Interestingly, the second biggest CRISPR array (96 spacers) lies within the integrated conjugative element ICE-1, which we hypothesize to be still active. This suggests that ICE-1 may serve as a vehicle for the horizontal transfer of the CRISPR spacers between *Ca*. N. cavascurensis and other organisms present in the same environment through conjugation, thus spreading the acquired immunity conferred by these spacers against common enemies.

Conclusions

We present an obligately thermophilic ammonia oxidizing archaeon from a hot spring in the Italian island of Ischia that is related to, but also clearly distinct from *Ca.* Nitrosocaldus yellowstonii. It contains most of the genes that have been found to be conserved among AOA and are implicated in energy and central carbon metabolism, except *nirK* encoding a nitrite reductase. Its genome gives indications for alternative energy metabolism and exhibits adaptations to the extreme environment. However, it lacks an identifiable reverse gyrase, which is found in most thermophiles with optimal growth temperatures above 65°C and apparently harbors a provirus of head-and-tail structure that is usually not found at high temperature. *Ca.* N. cavascurensis differs also in its gene sets for replication and cell division, which has implications for function and evolution of these systems in archaea. In addition, its extensive mobilome and the defense system indicate that thermophilic AOA are in constant exchange with the environment and with neighboring organisms as discussed for other thermophiles (van Wolferen et al., 2013). This might have shaped and continues in shaping the evolution of thaumarchaeota in hot springs. The pivotal phylogenetic position of *Ca*. N. cavascurensis will allow reconstructing the last common ancestor of AOA and provide further insights into the evolution of this ecologically widespread group of archaea.

For our enriched strain we propose a candidate status with the following taxonomic assignment:

Nitrosocaldales order
Nitrosocaldaceae fam. and
Candidatus Nitrosocaldus cavascurensis sp. nov.
Etymology: L. adj. nitrosus, “full of natron,” here intended to
mean nitrous (nitrite producer); L. masc.n. caldus, hot;
cavascurensis (L.masc. gen) describes origin of sample (terme di cavascura, Ischia)
Locality: hot mud, outflow from hot underground spring, 78 °C
Diagnosis: an ammonia oxidizing archaeon growing optimally around 68 °C at neutral pH under chemolithoautotrophic conditions with ammonia or urea,
spherically shaped with a diameter of approximately 0.6 to 0.8 µm, 4% sequence divergence in 16S rRNA gene from its next cultivated relative *Ca.* Nitrosocaldus yellowstonii.

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Authors contribution

CS conceived the study, MS did first enrichments, MM and CR made growth characterizations, KP did electron microscopy, SSA assembled genome and performed phylogenetic and genomic analyses for annotation, MKe annotated all metabolic and information processing genes, MKr analysed mobile genetic elements; CS, SSA, MKe wrote manuscript with contributions from MKr and MM.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

We are grateful to Lucia Monti for sharing her geological knowledge and for guiding us to the hot springs on Ischia in spring and fall 2013 and to Silvia Bulgheresi for crucial help in organising the fall expedition. We also thank Romana Bittner for excellent technical assistance and continuous cultivation and the 12 participants of the Bachelor practical class (University of Vienna) for help in sampling hot springs on Ischia and for setting up initial enrichments. We are grateful to Thomas Rattei and Florian Goldenberg for help in running computations and for providing access to the CUBE servers in Vienna, Austria. Special thanks go to LABGeM, the National Infrastructure « France Genomique » and the MicroScope team of the Genoscope (Evry, France) for providing the pipeline of annotation, and for their help during this study. We thank Eduardo Rocha and Marie Touchon for kindly providing access to their curated genome database.

Funding

This project was supported by ERC Advanced Grant TACKLE (695192) and Doktoratskolleg W1257 of the Austrian Science fund (FWF). SSA was funded by a “Marie-Curie Action” fellowship, grant number THAUMECOPHYL 701981.

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Abby, S.S., Neron, B., Menager, H., Touchon, M., and Rocha, E.P. (2014). MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. *PLoS One* 9**,** e110726.

Adam, P.S., Borrel, G., Brochier-Armanet, C., and Gribaldo, S. (2017). The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *ISME J* 11**,** 2407-2425.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25**,** 3389-3402.

Alves, R.J., Wanek, W., Zappe, A., Richter, A., Svenning, M.M., Schleper, C., and Urich, T. (2013). Nitrification rates in Arctic soils are associated with functionally distinct populations of ammonia-oxidizing archaea. *ISME J* 7**,** 1620-1631.

Amann, R.I., Krumholz, L., and Stahl, D.A. (1990). Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *J Bacteriol* 172**,** 762-770.

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

Barns, S.M., Delwiche, C.F., Palmer, J.D., and Pace, N.R. (1996). Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc Natl Acad Sci U S A* 93**,** 9188-9193.

Bartossek, R., Nicol, G.W., Lanzen, A., Klenk, H.P., and 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, R.J., Elisabeth, N.H., Garcia, J.A., Volland, J.M., Srivastava, A., Schleper, C., and Herndl, G.J. (2016). Physiological and genomic characterization of two novel marine thaumarchaeal strains indicates niche differentiation. *ISME J* 10**,** 1051-1063.

Beam, J.P., Jay, Z.J., Kozubal, M.A., and Inskeep, W.P. (2014). Niche specialization of novel Thaumarchaeota to oxic and hypoxic acidic geothermal springs of Yellowstone National Park. *ISME J* 8**,** 938-951.

Beam, J.P., Jay, Z.J., Schmid, M.C., Rusch, D.B., Romine, M.F., Jennings Rde, M., Kozubal, M.A., Tringe, S.G., Wagner, M., and Inskeep, W.P. (2016). Ecophysiology of an uncultivated lineage of Aigarchaeota from an oxic, hot spring filamentous 'streamer' community. *ISME J* 10**,** 210-224.

Bobay, L.M., Touchon, M., and Rocha, E.P. (2014). Pervasive domestication of defective prophages by bacteria. *Proc Natl Acad Sci U S A* 111**,** 12127-12132.

Bouthier De La Tour, C., Portemer, C., Nadal, M., Stetter, K.O., Forterre, P., and Duguet, M. (1990). Reverse gyrase, a hallmark of the hyperthermophilic archaebacteria. *J Bacteriol* 172**,** 6803-6808.

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

Brochier-Armanet, C., and Forterre, P. (2007). Widespread distribution of archaeal reverse gyrase in thermophilic bacteria suggests a complex history of vertical inheritance and lateral gene transfers. *Archaea* 2**,** 83-93.

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

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Caranto, J.D., and Lancaster, K.M. (2017). Nitric oxide is an obligate bacterial nitrification intermediate produced by hydroxylamine oxidoreductase. *Proc Natl Acad Sci U S A* 114**,** 8217-8222.

Chevenet, F., Croce, O., Hebrard, M., Christen, R., and Berry, V. (2010). ScripTree: scripting phylogenetic graphics. *Bioinformatics* 26**,** 1125-1126.

Cole, J.K., Peacock, J.P., Dodsworth, J.A., Williams, A.J., Thompson, D.B., Dong, H., Wu, G., and Hedlund, B.P. (2013). Sediment microbial communities in Great Boiling Spring are controlled by temperature and distinct from water communities. *ISME J* 7**,** 718-729.

Criscuolo, A., and Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* 10**,** 210.

Cubonova, L., Richardson, T., Burkhart, B.W., Kelman, Z., Connolly, B.A., Reeve, J.N., and Santangelo, T.J. (2013). Archaeal DNA polymerase D but not DNA polymerase B is required for genome replication in Thermococcus kodakarensis. *J Bacteriol* 195**,** 2322-2328.

De La Torre, J.R., Walker, C.B., Ingalls, A.E., Konneke, M., and Stahl, D.A. (2008). Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. *Environmental Microbiology* 10**,** 810-818.

Dodsworth, J.A., Hungate, B.A., and Hedlund, B.P. (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.

Eddy, S.R. (2011). Accelerated Profile HMM Searches. *PLoS Comput Biol* 7**,** e1002195.

Eiserling, F., Pushkin, A., Gingery, M., and Bertani, G. (1999). Bacteriophage-like particles associated with the gene transfer agent of methanococcus voltae PS. *J Gen Virol* 80 (Pt 12)**,** 3305-3308.

Eme, L., Reigstad, L.J., Spang, A., Lanzen, A., Weinmaier, T., Rattei, T., Schleper, C., and Brochier-Armanet, C. (2013). Metagenomics of Kamchatkan hot spring filaments reveal two new major (hyper)thermophilic lineages related to Thaumarchaeota. *Res Microbiol* 164**,** 425-438.

Erguder, T.H., Boon, N., Wittebolle, L., Marzorati, M., and Verstraete, W. (2009). Environmental factors shaping the ecological niches of ammonia-oxidizing archaea. *FEMS Microbiol Rev* 33**,** 855-869.

Evans, D.J., Jr., Evans, D.G., Engstrand, L., and Graham, D.Y. (1992). Urease-associated heat shock protein of Helicobacter pylori. *Infect Immun* 60**,** 2125-2127.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L., Tate, J., and Punta, M. (2014). Pfam: the protein families database. *Nucleic Acids Res* 42**,** D222-230.

Forterre, P. (2002). A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein. *Trends Genet* 18**,** 236-237.

Frols, S., Ajon, M., Wagner, M., Teichmann, D., Zolghadr, B., Folea, M., Boekema, E.J., Driessen, A.J., Schleper, C., and Albers, S.V. (2008). UV-inducible cellular aggregation of the hyperthermophilic archaeon Sulfolobus solfataricus is mediated by pili formation. *Mol Microbiol* 70**,** 938-952.

Giles, T.N., and Graham, D.E. (2008). Crenarchaeal arginine decarboxylase evolved from an S-adenosylmethionine decarboxylase enzyme. *J Biol Chem* 283**,** 25829-25838.

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

857 Gough, J., Karplus, K., Hughey, R., and Chothia, C. (2001). Assignment of homology to
858 genome sequences using a library of hidden Markov models that represent all proteins
859 of known structure. *J Mol Biol* 313**,** 903-919.
860 Grasso, S., and Tell, G. (2014). Base excision repair in Archaea: back to the future in DNA
861 repair. *DNA Repair (Amst)* 21**,** 148-157.
862 Greening, C., Biswas, A., Carere, C.R., Jackson, C.J., Taylor, M.C., Stott, M.B., Cook, G.M.,
863 and Morales, S.E. (2016). Genomic and metagenomic surveys of hydrogenase
864 distribution indicate H2 is a widely utilised energy source for microbial growth and
865 survival. *ISME J* 10**,** 761-777.
866 Gribaldo, S., and Brochier-Armanet, C. (2006). The origin and evolution of Archaea: a state
867 of the art. *Philos Trans R Soc Lond B Biol Sci* 361**,** 1007-1022.
868 Griffiths, R.I., Whiteley, A.S., O'donnell, A.G., and Bailey, M.J. (2000). Rapid method for
869 coextraction of DNA and RNA from natural environments for analysis of ribosomal
870 DNA- and rRNA-based microbial community composition. *Appl Environ Microbiol*
871 66**,** 5488-5491.
872 Grindley, N.D., Whiteson, K.L., and Rice, P.A. (2006). Mechanisms of site-specific
873 recombination. *Annu Rev Biochem* 75**,** 567-605.
874 Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify
875 clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 35**,** W52-
876 57.
877 Grogan, D.W. (2015). Understanding DNA Repair in Hyperthermophilic Archaea: Persistent
878 Gaps and Other Reasons to Focus on the Fork. *Archaea* 2015**,** 942605.
879 Groussin, M., and Gouy, M. (2011). Adaptation to environmental temperature is a major
880 determinant of molecular evolutionary rates in archaea. *Mol Biol Evol* 28**,** 2661-2674.
881 Haft, D.H., Selengut, J.D., and White, O. (2003). The TIGRFAMs database of protein
882 families. *Nucleic Acids Res* 31**,** 371-373.
883 Hallam, S.J., Mincer, T.J., Schleper, C., Preston, C.M., Roberts, K., Richardson, P.M., and
884 Delong, E.F. (2006). Pathways of carbon assimilation and ammonia oxidation
885 suggested by environmental genomic analyses of marine Crenarchaeota. *PLoS Biol* 4**,**
886 e95.
887 Hatzenpichler, R. (2012). Diversity, physiology, and niche differentiation of ammonia-
888 oxidizing archaea. *Appl Environ Microbiol* 78**,** 7501-7510.
889 Henche, A.L., Ghosh, A., Yu, X., Jeske, T., Egelman, E., and Albers, S.V. (2012). Structure
890 and function of the adhesive type IV pilus of Sulfolobus acidocaldarius. *Environ*
891 *Microbiol* 14**,** 3188-3202.
892 Higashibata, H., Fujiwara, S., Ezaki, S., Takagi, M., Fukui, K., and Imanaka, T. (2000).
893 Effect of polyamines on histone-induced DNA compaction of hyperthermophilic
894 archaea. *J Biosci Bioeng* 89**,** 103-106.
895 Ishino, Y., and Narumi, I. (2015). DNA repair in hyperthermophilic and hyperradioresistant
896 microorganisms. *Curr Opin Microbiol* 25**,** 103-112.
897 Jackson, C.R., Langner, H.W., Donahoe-Christiansen, J., Inskeep, W.P., and Mcdermott, T.R.
898 (2001). Molecular analysis of microbial community structure in an arsenite-oxidizing
899 acidic thermal spring. *Environ Microbiol* 3**,** 532-542.
900 Jiang, H., Huang, Q., Dong, H., Wang, P., Wang, F., Li, W., and Zhang, C. (2010). RNA-
901 based investigation of ammonia-oxidizing archaea in hot springs of Yunnan Province,
902 China. *Appl Environ Microbiol* 76**,** 4538-4541.
903 Johnson, C.M., and Grossman, A.D. (2015). Integrative and Conjugative Elements (ICEs):
904 What They Do and How They Work. *Annu Rev Genet* 49**,** 577-601.

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

905 Kanai, T., Matsuoka, R., Beppu, H., Nakajima, A., Okada, Y., Atomi, H., and Imanaka, T.
906 (2011). Distinct physiological roles of the three [NiFe]-hydrogenase orthologs in the
907 hyperthermophilic archaeon Thermococcus kodakarensis. *J Bacteriol* 193**,** 3109-3116.
908 Katoh, K., and Standley, D.M. (2014). MAFFT: iterative refinement and additional methods.
909 *Methods Mol Biol* 1079**,** 131-146.
910 Kazlauskas, D., Sezonov, G., Charpin, N., Venclovas, C., Forterre, P., and Krupovic, M.
911 (2017). Novel Families of Archaeo-Eukaryotic Primases Associated with Mobile
912 Genetic Elements of Bacteria and Archaea. *J Mol Biol*.
913 Kelman, Z., and White, M.F. (2005). Archaeal DNA replication and repair. *Curr Opin*
914 *Microbiol* 8**,** 669-676.
915 Kerou, M., Eloy Alves, R. J. And Schleper, C. (2016). "Nitrososphaeria," in *Bergey's Manual*
916 *of Systematics of Archaea and Bacteria.*, ed. B.S.M. Trust.: John Wiley & Sons, Inc.),
917 1-8.
918 Kerou, M., Offre, P., Valledor, L., Abby, S.S., Melcher, M., Nagler, M., Weckwerth, W., and
919 Schleper, C. (2016). Proteomics and comparative genomics of Nitrososphaera
920 viennensis reveal the core genome and adaptations of archaeal ammonia oxidizers.
921 *Proc Natl Acad Sci U S A* 113**,** E7937-E7946.
922 Kim, J.G., Park, S.J., Sinninghe Damste, J.S., Schouten, S., Rijpstra, W.I., Jung, M.Y., Kim,
923 S.J., Gwak, J.H., Hong, H., Si, O.J., Lee, S., Madsen, E.L., and Rhee, S.K. (2016).
924 Hydrogen peroxide detoxification is a key mechanism for growth of ammonia-
925 oxidizing archaea. *Proc Natl Acad Sci U S A* 113**,** 7888-7893.
926 Konneke, M., Bernhard, A.E., De La Torre, J.R., Walker, C.B., Waterbury, J.B., and Stahl,
927 D.A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature*
928 437**,** 543-546.
929 Konneke, M., Schubert, D.M., Brown, P.C., Hugler, M., Standfest, S., Schwander, T., Schada
930 Von Borzyskowski, L., Erb, T.J., Stahl, D.A., and Berg, I.A. (2014). Ammonia-
931 oxidizing archaea use the most energy-efficient aerobic pathway for CO2 fixation.
932 *Proc Natl Acad Sci U S A* 111**,** 8239-8244.
933 Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017).
934 Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and
935 repeat separation. *Genome Research*.
936 Kozlowski, J.A., Kits, K.D., and Stein, L.Y. (2016a). Comparison of Nitrogen Oxide
937 Metabolism among Diverse Ammonia-Oxidizing Bacteria. *Front Microbiol* 7**,** 1090.
938 Kozlowski, J.A., Stieglmeier, M., Schleper, C., Klotz, M.G., and Stein, L.Y. (2016b).
939 Pathways and key intermediates required for obligate aerobic ammonia-dependent
940 chemolithotrophy in bacteria and Thaumarchaeota. *ISME J* 10**,** 1836-1845.
941 Krupovic, M., Forterre, P., and Bamford, D.H. (2010). Comparative analysis of the mosaic
942 genomes of tailed archaeal viruses and proviruses suggests common themes for virion
943 architecture and assembly with tailed viruses of bacteria. *J Mol Biol* 397**,** 144-160.
944 Krupovic, M., Spang, A., Gribaldo, S., Forterre, P., and Schleper, C. (2011). A thaumarchaeal
945 provirus testifies for an ancient association of tailed viruses with archaea. *Biochem*
946 *Soc Trans* 39**,** 82-88.
947 Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and
948 Salzberg, S.L. (2004). Versatile and open software for comparing large genomes.
949 *Genome Biology* 5**,** R12.
950 Li, H., Yang, Q., Li, J., Gao, H., Li, P., and Zhou, H. (2015). The impact of temperature on
951 microbial diversity and AOA activity in the Tengchong Geothermal Field, China. *Sci*
952 *Rep* 5**,** 17056.

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

953 Lin, X., Handley, K.M., Gilbert, J.A., and Kostka, J.E. (2015). Metabolic potential of fatty
954 acid oxidation and anaerobic respiration by abundant members of Thaumarchaeota
955 and Thermoplasmata in deep anoxic peat. *ISME J* 9**,** 2740-2744.
956 Lopez-Garcia, P., Brochier, C., Moreira, D., and Rodriguez-Valera, F. (2004). Comparative
957 analysis of a genome fragment of an uncultivated mesopelagic crenarchaeote reveals
958 multiple horizontal gene transfers. *Environ Microbiol* 6**,** 19-34.
959 Luo, H., Zhang, C.T., and Gao, F. (2014). Ori-Finder 2, an integrated tool to predict
960 replication origins in the archaeal genomes. *Front Microbiol* 5**,** 482.
961 Makarova, K.S., Koonin, E.V., and Albers, S.V. (2016). Diversity and Evolution of Type IV
962 pili Systems in Archaea. *Front Microbiol* 7**,** 667.
963 Makarova, K.S., Krupovic, M., and Koonin, E.V. (2014). Evolution of replicative DNA
964 polymerases in archaea and their contributions to the eukaryotic replication
965 machinery. *Front Microbiol* 5**,** 354.
966 Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2015). Archaeal Clusters of Orthologous
967 Genes (arCOGs): An Update and Application for Analysis of Shared Features
968 between Thermococcales, Methanococcales, and Methanobacteriales. *Life* 5**,** 818-840.
969 Makarova, K.S., Yutin, N., Bell, S.D., and Koonin, E.V. (2010). Evolution of diverse cell
970 division and vesicle formation systems in Archaea. *Nat Rev Microbiol* 8**,** 731-741.
971 Martens-Habbena, W., Qin, W., Horak, R.E., Urakawa, H., Schauer, A.J., Moffett, J.W.,
972 Armbrust, E.V., Ingalls, A.E., Devol, A.H., and Stahl, D.A. (2015). The production of
973 nitric oxide by marine ammonia-oxidizing archaea and inhibition of archaeal
974 ammonia oxidation by a nitric oxide scavenger. *Environ Microbiol* 17**,** 2261-2274.
975 Medigue, C., Calteau, A., Cruveiller, S., Gachet, M., Gautreau, G., Josso, A., Lajus, A.,
976 Langlois, J., Pereira, H., Planel, R., Roche, D., Rollin, J., Rouy, Z., and Vallenet, D.
977 (2017). MicroScope-an integrated resource for community expertise of gene functions
978 and comparative analysis of microbial genomic and metabolic data. *Brief Bioinform*.
979 Miele, V., Penel, S., Daubin, V., Picard, F., Kahn, D., and Duret, L. (2012). High-quality
980 sequence clustering guided by network topology and multiple alignment likelihood.
981 *Bioinformatics* 28**,** 1078-1085.
982 Miele, V., Penel, S., and Duret, L. (2011). Ultra-fast sequence clustering from similarity
983 networks with SiLiX. *BMC Bioinformatics* 12**,** 116.
984 Mussmann, M., Brito, I., Pitcher, A., Sinninghe Damste, J.S., Hatzenpichler, R., Richter, A.,
985 Nielsen, J.L., Nielsen, P.H., Muller, A., Daims, H., Wagner, M., and Head, I.M.
986 (2011). Thaumarchaeotes abundant in refinery nitrifying sludges express amoA but
987 are not obligate autotrophic ammonia oxidizers. *Proc Natl Acad Sci U S A* 108**,**
988 16771-16776.
989 Muyzer, G., De Waal, E.C., and Uitterlinden, A.G. (1993). Profiling of complex microbial
990 populations by denaturing gradient gel electrophoresis analysis of polymerase chain
991 reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* 59**,** 695-700.
992 Neves, C., Da Costa, M.S., and Santos, H. (2005). Compatible solutes of the
993 hyperthermophile Palaeococcus ferrophilus: osmoadaptation and thermoadaptation in
994 the order thermococcales. *Appl Environ Microbiol* 71**,** 8091-8098.
995 Ng, K.H., Srinivas, V., Srinivasan, R., and Balasubramanian, M. (2013). The Nitrosopumilus
996 maritimus CdvB, but not FtsZ, assembles into polymers. *Archaea* 2013**,** 104147.
997 Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and
998 effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol*
999 *Biol Evol* 32**,** 268-274.

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1000 Nicol, G.W., Leininger, S., Schleper, C., and Prosser, J.I. (2008). The influence of soil pH on
1001 the diversity, abundance and transcriptional activity of ammonia oxidizing archaea
1002 and bacteria. *Environ Microbiol* 10**,** 2966-2978.
1003 Nicol, G.W., and Schleper, C. (2006). Ammonia-oxidising Crenarchaeota: important players
1004 in the nitrogen cycle? *Trends Microbiol* 14**,** 207-212.
1005 Nishizawa, M., Sakai, S., Konno, U., Nakahara, N., Takaki, Y., Saito, Y., Imachi, H.,
1006 Tasumi, E., Makabe, A., Koba, K., and Takai, K. (2016). Nitrogen and Oxygen
1007 Isotope Effects of Ammonia Oxidation by Thermophilic Thaumarchaeota from a
1008 Geothermal Water Stream. *Appl Environ Microbiol* 82**,** 4492-4504.
1009 Nunoura, T., Hirayama, H., Takami, H., Oida, H., Nishi, S., Shimamura, S., Suzuki, Y.,
1010 Inagaki, F., Takai, K., Nealson, K.H., and Horikoshi, K. (2005). Genetic and
1011 functional properties of uncultivated thermophilic crenarchaeotes from a subsurface
1012 gold mine as revealed by analysis of genome fragments. *Environ Microbiol* 7**,** 1967-
1013 1984.
1014 Nunoura, T., Takaki, Y., Kakuta, J., Nishi, S., Sugahara, J., Kazama, H., Chee, G.J., Hattori,
1015 M., Kanai, A., Atomi, H., Takai, K., and Takami, H. (2011). Insights into the
1016 evolution of Archaea and eukaryotic protein modifier systems revealed by the genome
1017 of a novel archaeal group. *Nucleic Acids Res* 39**,** 3204-3223.
1018 Offre, P., Kerou, M., Spang, A., and Schleper, C. (2014). Variability of the transporter gene
1019 complement in ammonia-oxidizing archaea. *Trends Microbiol* 22**,** 665-675.
1020 Offre, P., Spang, A., and Schleper, C. (2013). Archaea in biogeochemical cycles. *Annu Rev*
1021 *Microbiol* 67**,** 437-457.
1022 Okonechnikov, K., Golosova, O., Fursov, M., and Team, U. (2012). Unipro UGENE: a
1023 unified bioinformatics toolkit. *Bioinformatics* 28**,** 1166-1167.
1024 Oshima, T. (2007). Unique polyamines produced by an extreme thermophile, Thermus
1025 thermophilus. *Amino Acids* 33**,** 367-372.
1026 Otte, J., Mall, A., Schubert, D.M., Konneke, M., and Berg, I.A. (2015). Malonic
1027 semialdehyde reductase from the archaeon Nitrosopumilus maritimus is involved in
1028 the autotrophic 3-hydroxypropionate/4-hydroxybutyrate cycle. *Appl Environ*
1029 *Microbiol* 81**,** 1700-1707.
1030 Ovreas, L., Forney, L., Daae, F.L., and Torsvik, V. (1997). Distribution of bacterioplankton
1031 in meromictic Lake Saelenvannet, as determined by denaturing gradient gel
1032 electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. *Appl Environ*
1033 *Microbiol* 63**,** 3367-3373.
1034 Park, S.J., Kim, J.G., Jung, M.Y., Kim, S.J., Cha, I.T., Ghai, R., Martin-Cuadrado, A.B.,
1035 Rodriguez-Valera, F., and Rhee, S.K. (2012). Draft genome sequence of an ammonia-
1036 oxidizing archaeon, "Candidatus Nitrosopumilus sediminis" AR2, from Svalbard in
1037 the Arctic Circle. *Journal of bacteriology* 194**,** 6948-6949.
1038 Pelve, E.A., Lindas, A.C., Martens-Habbena, W., De La Torre, J.R., Stahl, D.A., and
1039 Bernander, R. (2011). Cdv-based cell division and cell cycle organization in the
1040 thaumarchaeon Nitrosopumilus maritimus. *Mol Microbiol* 82**,** 555-566.
1041 Pester, M., Schleper, C., and Wagner, M. (2011). The Thaumarchaeota: an emerging view of
1042 their phylogeny and ecophysiology. *Curr Opin Microbiol* 14**,** 300-306.
1043 Peters, J.W., Schut, G.J., Boyd, E.S., Mulder, D.W., Shepard, E.M., Broderick, J.B., King,
1044 P.W., and Adams, M.W. (2015). [FeFe]- and [NiFe]-hydrogenase diversity,
1045 mechanism, and maturation. *Biochim Biophys Acta* 1853**,** 1350-1369.

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Petitjean, C., Deschamps, P., Lopez-Garcia, P., and Moreira, D. (2014). Rooting the domain archaea by phylogenomic analysis supports the foundation of the new kingdom Proteoarchaeota. *Genome Biol Evol* 7**,** 191-204.

Poli, A., Di Donato, P., Abbamondi, G.R., and Nicolaus, B. (2011). Synthesis, production, and biotechnological applications of exopolysaccharides and polyhydroxyalkanoates by archaea. *Archaea* 2011**,** 693253.

Prangishvili, D., Bamford, D.H., Forterre, P., Iranzo, J., Koonin, E.V., and Krupovic, M. (2017). The enigmatic archaeal virosphere. *Nat Rev Microbiol* 15**,** 724-739.

Prosser, J.I., and Nicol, G.W. (2008). Relative contributions of archaea and bacteria to aerobic ammonia oxidation in the environment. *Environ Microbiol* 10**,** 2931-2941.

Qin, W., Amin, S.A., Martens-Habbena, W., Walker, C.B., Urakawa, H., Devol, A.H., Ingalls, A.E., Moffett, J.W., Armbrust, E.V., and Stahl, D.A. (2014). Marine ammonia-oxidizing archaeal isolates display obligate mixotrophy and wide ecotypic variation. *Proc Natl Acad Sci U S A* 111**,** 12504-12509.

Ramos-Vera, W.H., Weiss, M., Strittmatter, E., Kockelkorn, D., and Fuchs, G. (2011). Identification of missing genes and enzymes for autotrophic carbon fixation in crenarchaeota. *J Bacteriol* 193**,** 1201-1211.

Raymann, K., Brochier-Armanet, C., and Gribaldo, S. (2015). The two-domain tree of life is linked to a new root for the Archaea. *Proc Natl Acad Sci U S A* 112**,** 6670-6675.

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

Rouillon, C., and White, M.F. (2011). The evolution and mechanisms of nucleotide excision repair proteins. *Res Microbiol* 162**,** 19-26.

Saier, M.H., Jr., Reddy, V.S., Tamang, D.G., and Vastermark, A. (2014). The transporter classification database. *Nucleic Acids Res* 42**,** D251-258.

Samson, R.Y., Dobro, M.J., Jensen, G.J., and Bell, S.D. (2017). The Structure, Function and Roles of the Archaeal ESCRT Apparatus. *Subcell Biochem* 84**,** 357-377.

Samson, R.Y., Obita, T., Hodgson, B., Shaw, M.K., Chong, P.L., Williams, R.L., and Bell, S.D. (2011). Molecular and structural basis of ESCRT-III recruitment to membranes during archaeal cell division. *Mol Cell* 41**,** 186-196.

Schleper, C., Jurgens, G., and Jonuscheit, M. (2005). Genomic studies of uncultivated archaea. *Nat Rev Microbiol* 3**,** 479-488.

Schleper, C., and Nicol, G.W. (2010). Ammonia-oxidising archaea--physiology, ecology and evolution. *Adv Microb Physiol* 57**,** 1-41.

Shen, T., Stieglmeier, M., Dai, J., Urich, T., and 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.

Shi, Y., Tyson, G.W., Eppley, J.M., and Delong, E.F. (2011). Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *ISME J* 5**,** 999-1013.

Shima, S., Herault, D.A., Berkessel, A., and Thauer, R.K. (1998). Activation and thermostabilization effects of cyclic 2, 3-diphosphoglycerate on enzymes from the hyperthermophilic Methanopyrus kandleri. *Arch Microbiol* 170**,** 469-472.

Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* 34**,** D32-36.

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1093 Simon, J., and Klotz, M.G. (2013). Diversity and evolution of bioenergetic systems involved
1094 in microbial nitrogen compound transformations. *Biochim Biophys Acta* 1827**,** 114-
1095 135.
1096 Soding, J. (2005). Protein homology detection by HMM-HMM comparison. *Bioinformatics*
1097 21**,** 951-960.
1098 Sondergaard, D., Pedersen, C.N., and Greening, C. (2016). HydDB: A web tool for
1099 hydrogenase classification and analysis. *Sci Rep* 6**,** 34212.
1100 Spang, A., Hatzenpichler, R., Brochier-Armanet, C., Rattei, T., Tischler, P., Spieck, E., Streit,
1101 W., Stahl, D.A., Wagner, M., and Schleper, C. (2010). Distinct gene set in two
1102 different lineages of ammonia-oxidizing archaea supports the phylum
1103 Thaumarchaeota. *Trends Microbiol* 18**,** 331-340.
1104 Spang, A., Poehlein, A., Offre, P., Zumbragel, S., Haider, S., Rychlik, N., Nowka, B.,
1105 Schmeisser, C., Lebedeva, E.V., Rattei, T., Bohm, C., Schmid, M., Galushko, A.,
1106 Hatzenpichler, R., Weinmaier, T., Daniel, R., Schleper, C., Spieck, E., Streit, W., and
1107 Wagner, M. (2012). The genome of the ammonia-oxidizing Candidatus
1108 Nitrososphaera gargensis: insights into metabolic versatility and environmental
1109 adaptations. *Environ Microbiol* 14**,** 3122-3145.
1110 Stahl, D.A., and Amann, R. (1991). "Development and application of nucleic acid probes.,"
1111 in *Nucleic acid techniques in bacterial systematics.*, eds. E. Stackebrandt & M.
1112 Goodfellow. (New York, N.Y: John Wiley & Sons, Inc.), 205-248.
1113 Stahl, D.A., and De La Torre, J.R. (2012). Physiology and diversity of ammonia-oxidizing
1114 archaea. *Annu Rev Microbiol* 66**,** 83-101.
1115 Szabo, Z., Sani, M., Groeneveld, M., Zolghadr, B., Schelert, J., Albers, S.V., Blum, P.,
1116 Boekema, E.J., and Driessen, A.J. (2007). Flagellar motility and structure in the
1117 hyperthermoacidophilic archaeon Sulfolobus solfataricus. *J Bacteriol* 189**,** 4305-4309.
1118 Tourna, M., Stieglmeier, M., Spang, A., Konneke, M., Schintlmeister, A., Urich, T., Engel,
1119 M., Schloter, M., Wagner, M., Richter, A., and Schleper, C. (2011). Nitrososphaera
1120 viennensis, an ammonia oxidizing archaeon from soil. *Proc Natl Acad Sci U S A* 108**,**
1121 8420-8425.
1122 Treusch, A.H., Leininger, S., Kletzin, A., Schuster, S.C., Klenk, H.P., and Schleper, C.
1123 (2005). Novel genes for nitrite reductase and Amo-related proteins indicate a role of
1124 uncultivated mesophilic crenarchaeota in nitrogen cycling. *Environ Microbiol* 7**,**
1125 1985-1995.
1126 Tripepi, M., Imam, S., and Pohlschroder, M. (2010). Haloferax volcanii flagella are required
1127 for motility but are not involved in PibD-dependent surface adhesion. *J Bacteriol* 192**,**
1128 3093-3102.
1129 Vajrala, N., Martens-Habbena, W., Sayavedra-Soto, L.A., Schauer, A., Bottomley, P.J., Stahl,
1130 D.A., and Arp, D.J. (2013). Hydroxylamine as an intermediate in ammonia oxidation
1131 by globally abundant marine archaea. *Proc Natl Acad Sci U S A* 110**,** 1006-1011.
1132 Vallenet, D., Calteau, A., Cruveiller, S., Gachet, M., Lajus, A., Josso, A., Mercier, J.,
1133 Renaux, A., Rollin, J., Rouy, Z., Roche, D., Scarpelli, C., and Medigue, C. (2017).
1134 MicroScope in 2017: an expanding and evolving integrated resource for community
1135 expertise of microbial genomes. *Nucleic Acids Res* 45**,** D517-D528.
1136 Vallenet, D., Engelen, S., Mornico, D., Cruveiller, S., Fleury, L., Lajus, A., Rouy, Z., Roche,
1137 D., Salvignol, G., Scarpelli, C., and Medigue, C. (2009). MicroScope: a platform for
1138 microbial genome annotation and comparative genomics. *Database (Oxford)* 2009**,**
1139 bap021.

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Van Wolferen, M., Ajon, M., Driessen, A.J., and Albers, S.V. (2013). How hyperthermophiles adapt to change their lives: DNA exchange in extreme conditions. *Extremophiles* 17**,** 545-563.

Williams, T.A., Szollosi, G.J., Spang, A., Foster, P.G., Heaps, S.E., Boussau, B., Ettema, T.J.G., and Embley, T.M. (2017). Integrative modeling of gene and genome evolution roots the archaeal tree of life. *Proc Natl Acad Sci U S A* 114**,** E4602-E4611.

Yan, J., Beattie, T.R., Rojas, A.L., Schermerhorn, K., Gristwood, T., Trinidad, J.C., Albers, S.V., Roversi, P., Gardner, A.F., Abrescia, N.G.A., and Bell, S.D. (2017). Identification and characterization of a heterotrimeric archaeal DNA polymerase holoenzyme. *Nat Commun* 8**,** 15075.

Yan, J., Haaijer, S.C., Op Den Camp, H.J., Van Niftrik, L., Stahl, D.A., Konneke, M., Rush, D., Sinninghe Damste, J.S., Hu, Y.Y., and Jetten, M.S. (2012). Mimicking the oxygen minimum zones: stimulating interaction of aerobic archaeal and anaerobic bacterial ammonia oxidizers in a laboratory-scale model system. *Environ Microbiol* 14**,** 3146-3158.

Zhang, C.L., Ye, Q., Huang, Z., Li, W., Chen, J., Song, Z., Zhao, W., Bagwell, C., Inskeep, W.P., Ross, C., Gao, L., Wiegel, J., Romanek, C.S., Shock, E.L., and Hedlund, B.P. (2008). Global occurrence of archaeal amoA genes in terrestrial hot springs. *Appl Environ Microbiol* 74**,** 6417-6426.

Zhou, J., Bruns, M.A., and Tiedje, J.M. (1996). DNA recovery from soils of diverse composition. *Appl Environ Microbiol* 62**,** 316-322.

Zolghadr, B., Weber, S., Szabo, Z., Driessen, A.J., and Albers, S.V. (2007). Identification of a system required for the functional surface localization of sugar binding proteins with class III signal peptides in Sulfolobus solfataricus. *Mol Microbiol* 64**,** 795-806.

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure Legends

Figure 1. Growth of *Ca.* N. cavascurensis. (A) Enrichment cultures with near stoichiometric conversion of ammonia to nitrite were paralleled by cell growth, as estimated by archaeal 16S rRNA gene copies in qPCR. Gene copies declined once ammonia oxidation ceased. Data points for ammonia and nitrite are averages of biological duplicates. Technical triplicates of each biological duplicate were done to determine the gene copy number. Error bars represent the standard error of the mean. **(B)** Ammonia and successive nitrite production of enrichment cultures with 0.5 mM urea as nitrogen source. On urea as substrate ammonia accumulated during the early growth phase and the cultures showed decreasing nitrite production rate indicating growth limitation. Ammonia cultures served as control and all data points show the mean of biological triplicates with error bars representing the standard error of the mean.

Figure 2. Optimal growth temperature of *Ca.* N. cavascurensis. (A) Nitrite production occurs within the tested temperature range (65 – 78 °C) up to 74°C, with an optimum at 68 °C. Culture quadruplets were incubated for each tested temperature and the error bars represent the standard error of the mean which increases with temperature, indicating an increase of the stochastic element within the microbial community. **(B)** Temperature dependance of growth rates showing a maximum of 0.98 d^{-1} (24.6 h generation time) at 68 °C and the highest growth temperature at 74 °C. Growth rates were calculated by linear regression of semi-logarithmically plotted nitrite values during the exponential growth phase (min. five different time points).

Figure 3. Micrographs of *Ca.* N. cavascurensis. Light **(A)** and epifluorescence micrographs **(B)** of a late exponential enrichment culture analysed with FISH. *Ca.* N. cavascurensis cells (red) and bacterial cells (green) were labeled with ARCH 915 and EUB 338 probes respectively, to give a representative picture of the enrichment state. Scanning electron micrographs **(C)** and **(D)** show the spherical nature of *Ca.* N. cavascurensis cells, having a diameter of 0.6 – 0.8 µm, small rods are visible that belong to the remaining bacterial consortium.

Figure 4. Phylogenetic position of *Ca.* N. cavascurensis, and distribution of the ammonia monooxygenase (AMO) and urease gene clusters in Thaumarchaeota genomes.
This maximum-likelihood tree built with IQ-tree is based on 59 ribosomal proteins found in at least 35 genomes out of the 40 genomes dataset. AMO and Urease genes clusters found are displayed along the tree. Colors of gene boxes indicate the gene families involved in AMO or urease (*amoABCX* and *UreABCDEFG* respectively). Grey boxes correspond to other genes. Gene clusters are displayed in the orientation which maximizes the vertical alignment of the conserved genes. Genes are represented on an un-interrupted cluster when less than five genes separated them, otherwise, genes apart from more than five gene positions are displayed on another cluster. For AMO clusters, the number of different *amoC* homologs found in the genome outside of the main AMO cluster is indicated in the *amoC* box. Putative urea transporters are indicated by grey boxes outlined by blue for SSS-type urea transporter, and pink for UT urea transporter. The chaperonin Hsp60 homolog found in the urease cluster and discussed in the main text is indicated with a red box underlining its conserved position over multiple loci. Note that *N. evergladensis* and *N. viennensis* both have an extra set of three urease genes (*UreABC*) not displayed here.

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 5. Metabolic reconstruction of *Ca*. N. cavascurensis
Schematic reconstruction of the predicted metabolic modules and other genome features of *Ca.* N. cavascurensis, as discussed in the text. Gene accession numbers and transporter classes are listed in Supplementary Table 1. HBCD: 4-hydoxybutyryl-CoA dehydratase; cDPG: cyclic 2, 3-diphosphoglycerate; MCP: methyl-accepting chemotaxis proteins; Fla: archaellum; Hyd3b: type 3b hydrogenase complex; polY: Y-family translesion polymerase; HR: homologous recombination; NER: nucleotide excision repair; BER: base excision repair; S*: organosulfur compounds, sulfite, sulfate.

Figure 6. Effect of the NO scavenger 2-phenyl-4, 4, 5, 5,-tetramethylimidazoline-1-oxyl 3-oxide (PTIO) on nitrite production. Different concentrations of PTIO were added after 63 h in early to mid-exponential growth phase indicated by the arrow. Cultures with 20 µM PTIO resumed nitrite production 72 h after the addition of PTIO, while higher concentrations resulted in a complete arrest. Nitrite curves of control and 20 µM PTIO cultures show mean values from duplicates, whereas 100 and 400 µM curves are from triplicates and error bars represent the standard error of the mean.

Figure 7. Genomic regions of atypical GC% content contain mobile genetic elements in *Ca.* N. cavascurensis. The genome is displayed in the form of a ring where genes are colored by functional categories (arCOG and COG), with the mean GC% content (blue circle) and local GC% as computed using 5kb sliding windows (black graph) in inner rings. The minimal and maximal values of local GC% are displayed as dashed grey concentric rings. The predicted origin of replication is indicated with "Ori". Several large regions show a deviation to the average GC% that correspond to mobile genetic elements, i.e. ICEs (integrated conjugated elements), a specific pilus (see Fig. 8), and a defense system (CRISPR-Cas Type IB and three CRISPR arrays (see text).

Figure 8. Mobile genetic elements and cell appendages encoded in the genome of *Ca.* N. cavascurensis.
(A) Two putative Integrative Conjugative Elements (ICE) with sequence similarities shown by connexions between the two. Genes are colored by function (see legend). CHLP: geranylgeranyl diphosphate reductase; GGPS: geranylgeranyl pyrophosphate synthetase; UbiC: chorismate lyase; UbiA: 4-hydroxybenzoate octaprenyltransferase; RmlB: dTDP-glucose 4,6-dehydratase; UbiE: Ubiquinone/menaquinone biosynthesis C-methylase; HslU: ATP-dependent HslU protease; PflA: Pyruvate-formate lyase-activating enzyme; SuhB: fructose-1,6-bisphosphatase; TrxB: Thioredoxin reductase; YyaL: thoiredoxin and six-hairpin glycosidase-like domains; BCAT: branched-chain amino acid aminotransferase; PncA: Pyrazinamidase/nicotinamidase; SseA: 3-mercaptopyruvate sulfurtransferase; ThyX: Thymidylate synthase; CzcD: Co/Zn/Cd efflux system. **(B)** A head-tail like provirus from the Caudovirales order with genes in blue indicating putative viral functions. **(C)** *Ca.* N. cavascurensis specific type IV pilus locus next to a set of chemotaxis and flagellum (archaellum) genes. Homologs shared between the flagellum and the T4P are displayed with the same color. A scheme of the putative type IV pilus is shown on the left with the same color code. (inspired from drawings in (Makarova et al., 2016). mb: membrane; cyto: cytoplasm; MCP: methyl-accepting chemotaxis proteins.

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

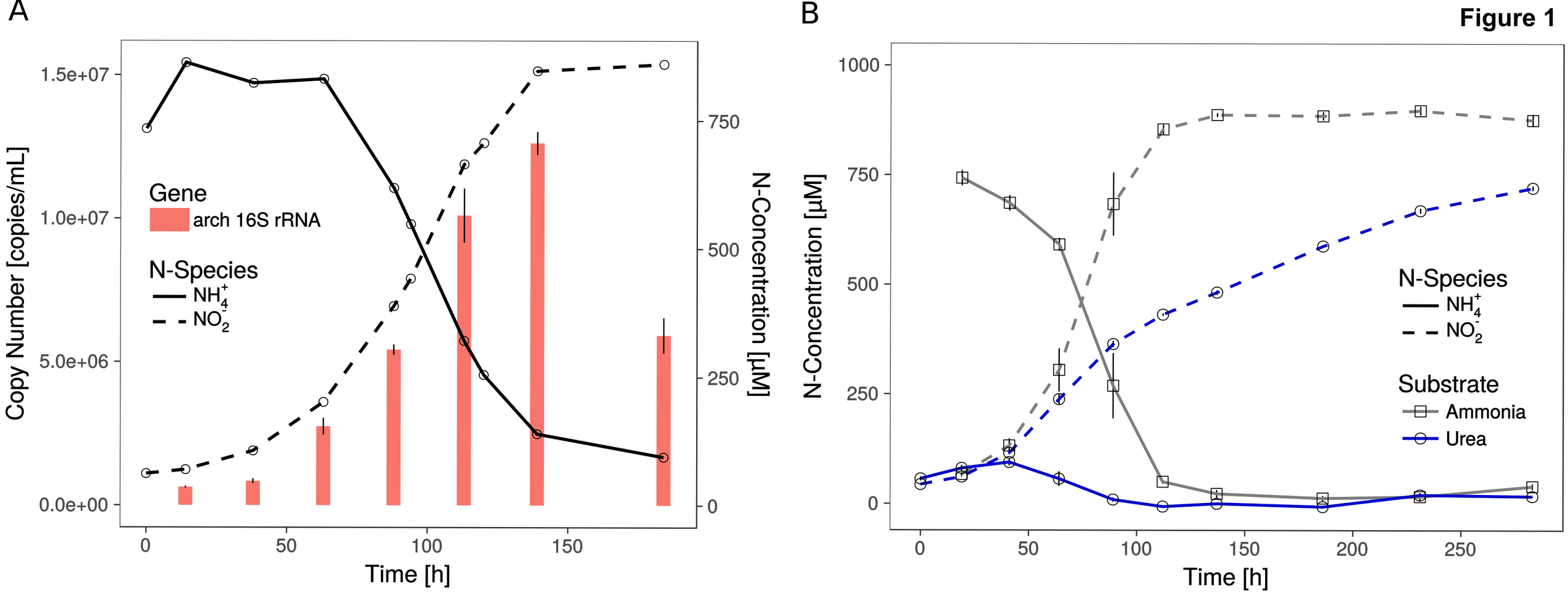
Supplementary Material

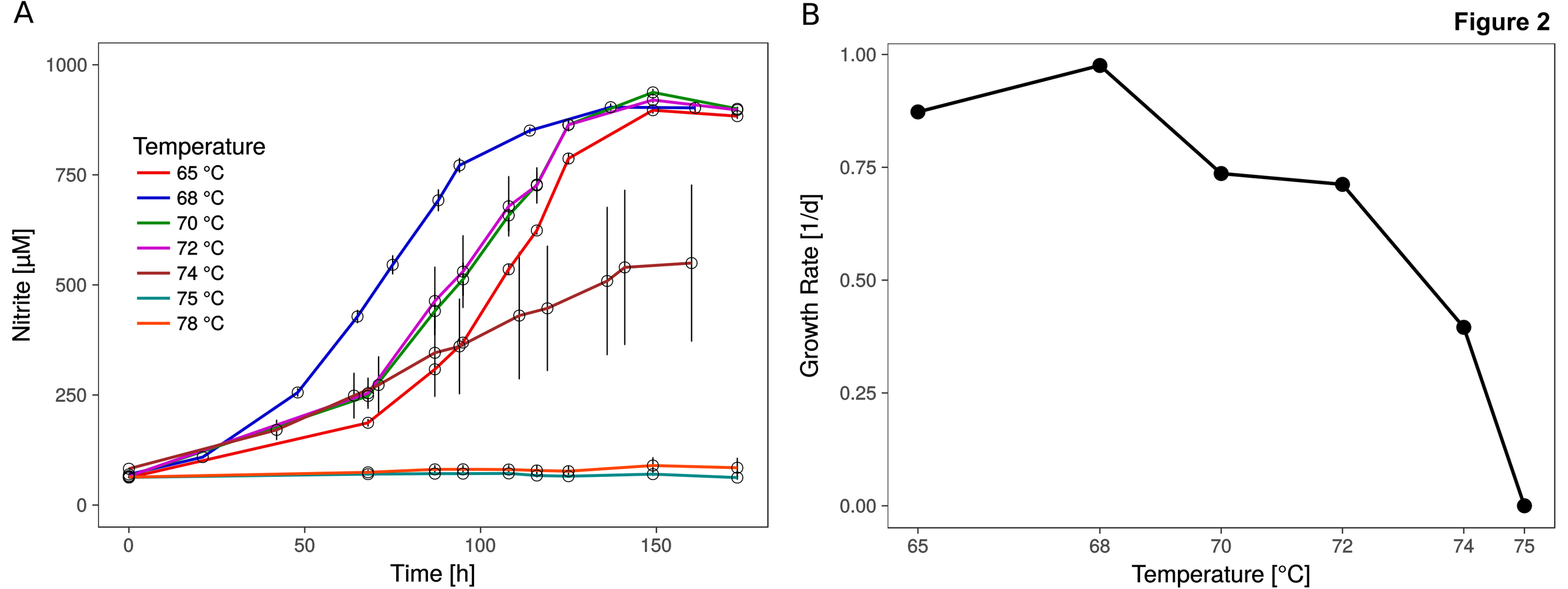
Supplementary Figure 1. Bioinformatic genome closure in a repeat-rich region encoding a hypothetical adhesin. (A) The two initially assembled contigs were aligned using MUMMER (Kurtz et al., 2004), revealing a sequence overlap with very high sequence identity (>99.5%) that could be resolved and resulted in the genome closure: contig 2 was a redundancy, and contig 1 was circularized through a long gene carrying repeated sequences **(B)** Analysis of this long gene showed that it encodes for an adhesin (a likely surface protein involved in adhesion). As is typical for such proteins, it contains several repeats including LamG/concavalin A domains interspersed with Ca2+-binding cadherin domains (both involved in adhesion) and terminates with two phospholipase D domains (PLD).

Supplementary Figure 2. Phylogenetic tree of the 4-hydoxybutyryl-CoA dehydratase protein family. A maximum likelihood phylogenetic tree was obtained with IQ-Tree v 1.5.5. The groups previously defined in (Konneke et al., 2014) "CREN type-1", "CREN type-2", "Anaerobe cluster", and "AOA" are indicated by red boxes. The organisms' class and phylum are indicated for each sequence along brackets on the right. The two homologs of this enzyme found in *Ca*. N. cavascurensis genome are indicated by a red arrow. This figure was generated using the Scriptree program (Chevenet et al., 2010), and the resulting SVG file modified using Inkscape.

Supplementary Table 1 (supplied as an excel file). Locus tags and annotations of the main metabolic and biosynthetic pathways, information processing genes and transporters of *Ca.* Nitrosocaldus cavascurensis.

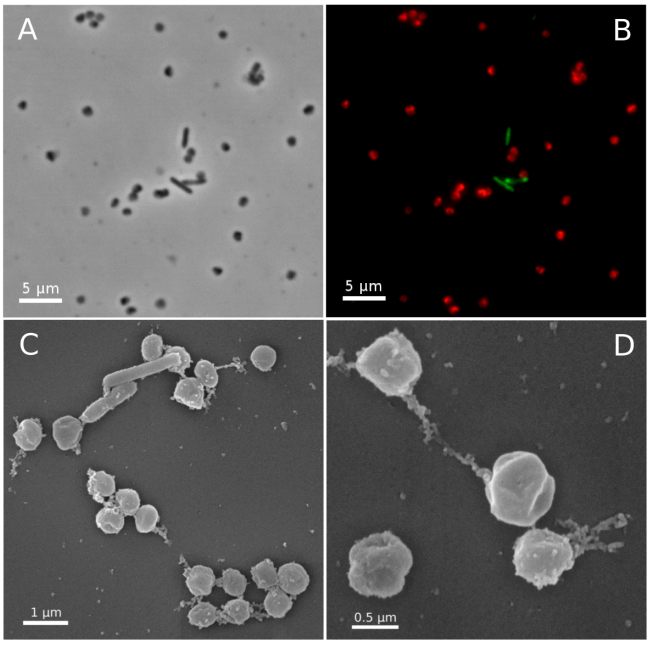
Figure 1


A
1.5e+07
1.0e+07
5.0e+06
0.0e+00
Copy Number [copies/mL]
750
500
250
0
N-Concentration [µM]
Gene
arch 16S rRNA
N-Species
NH4+
NO2-
0
50
100
150
Time [h]
B
1000
750
500
250
0
N-Concentration [µM]
N-Species
NH4+
NO2-
Substrate
Ammonia
Urea
0
50
100
150
200
250
Time [h]


A
B
Figure 2
Nitrite [µM]
Time [h]
Temperature
65 °C
68 °C
70 °C
72 °C
74 °C
75 °C
78 °C
0
250
500
750
1000
0
50
100
150
Growth Rate [1/d]
Temperature [°C]
0.00
0.25
0.50
0.75
1.00
65
68
70
72
74
75

bioRxiv preprint doi: https://doi.org/10.1101/240879; this version posted December 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3



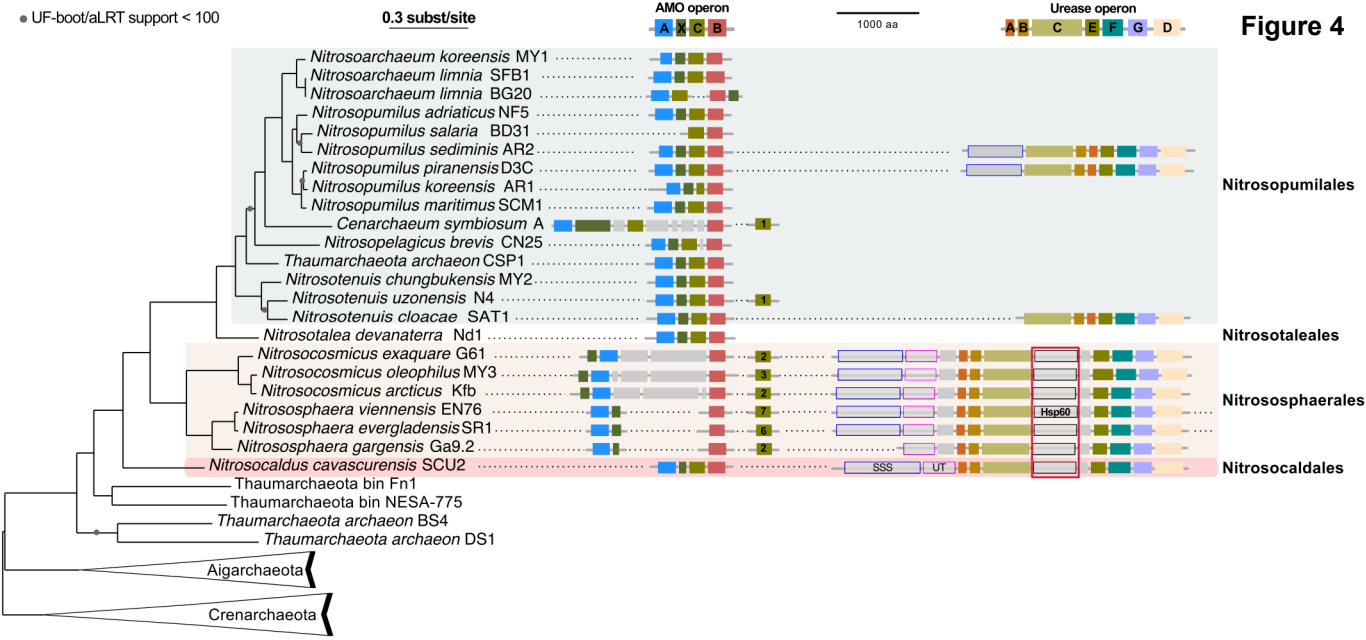
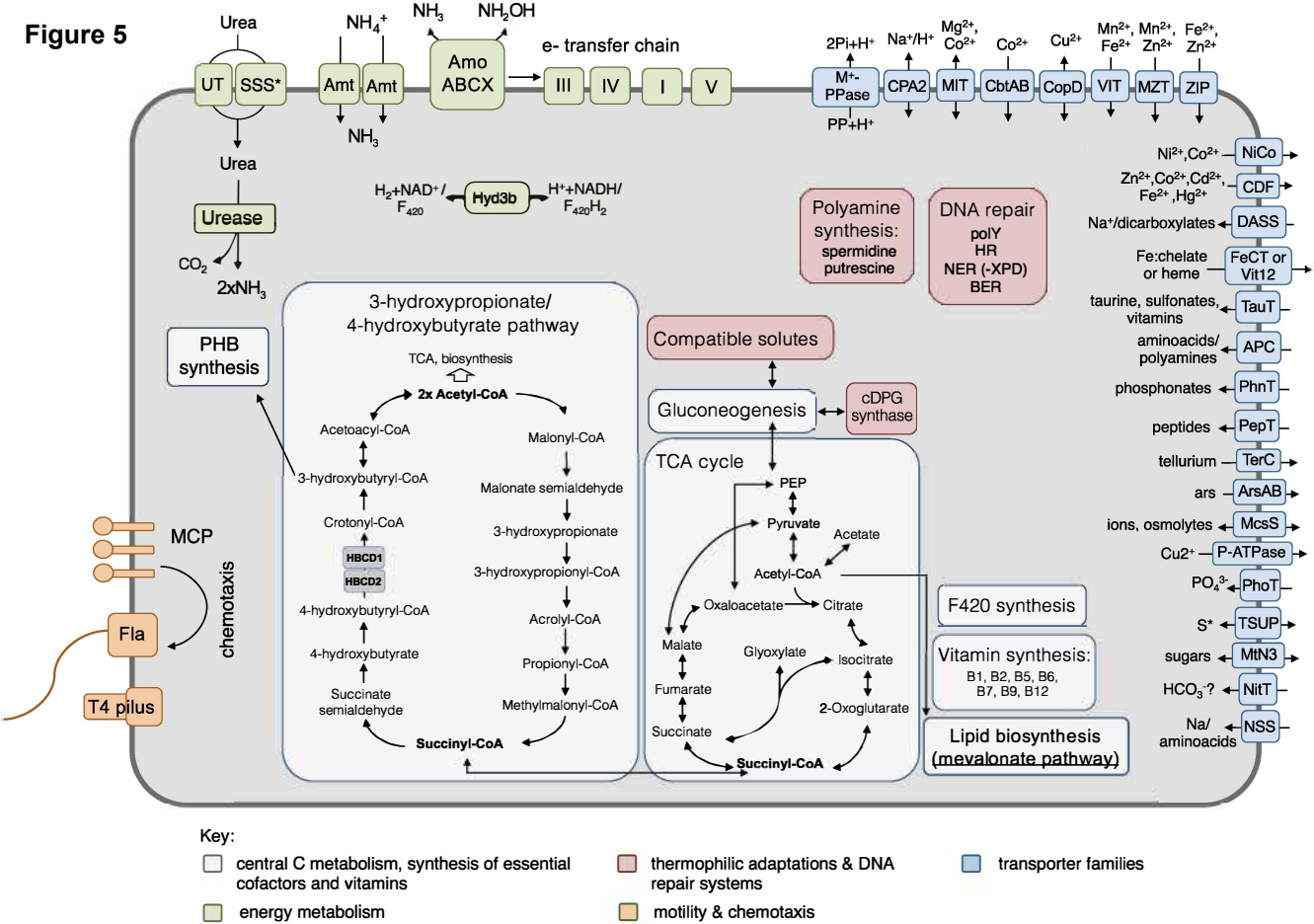

UF-boot/aLRT support < 100
0.3 subst/site
AMO operon
A X C B
1000 aa
Urease operon
A B C E F G D
Nitrosoarchaeum koreensis MY1
Nitrosoarchaeum limnia SFB1
Nitrosoarchaeum limnia BG20
Nitrosopumilus adriaticus NF5
Nitrosopumilus salaria BD31
Nitrosopumilus sediminis AR2
Nitrosopumilus piranensis D3C
Nitrosopumilus koreensis AR1
Nitrosopumilus maritimus SCM1
Cenarchaeum symbiosum A
Nitrosopelagicus brevis CN25
Thaumarchaeota archaeon CSP1
Nitrosotenuis chungbukensis MY2
Nitrosotenuis uzonensis N4
Nitrosotenuis cloacae SAT1
Nitrosotalea devanaterra Nd1
Nitrososocosmicus exaquare G61
Nitrososocosmicus oleophilus MY3
Nitrososocosmicus arcticus Kfb
Nitrososphaera viennensis EN76
Nitrososphaera evergladensis SR1
Nitrososphaera gargensis Ga9.2
Nitrosocaldus cavascurensis SCU2
SSS
UT
Hsp60
Thaumarchaeota bin Fn1
Thaumarchaeota bin NESA-775
Thaumarchaeota archaeon BS4
Thaumarchaeota archaeon DS1
Aigarchaeota
Crenarchaeota
Nitrosopumilales
Nitrosotaleales
Nitrososphaerales
Nitrosocaldales

Figure 4

Figure 5



Key:

- central C metabolism, synthesis of essential cofactors and vitamins
- energy metabolism
- thermophilic adaptations & DNA repair systems
- motility & chemotaxis
- transporter families

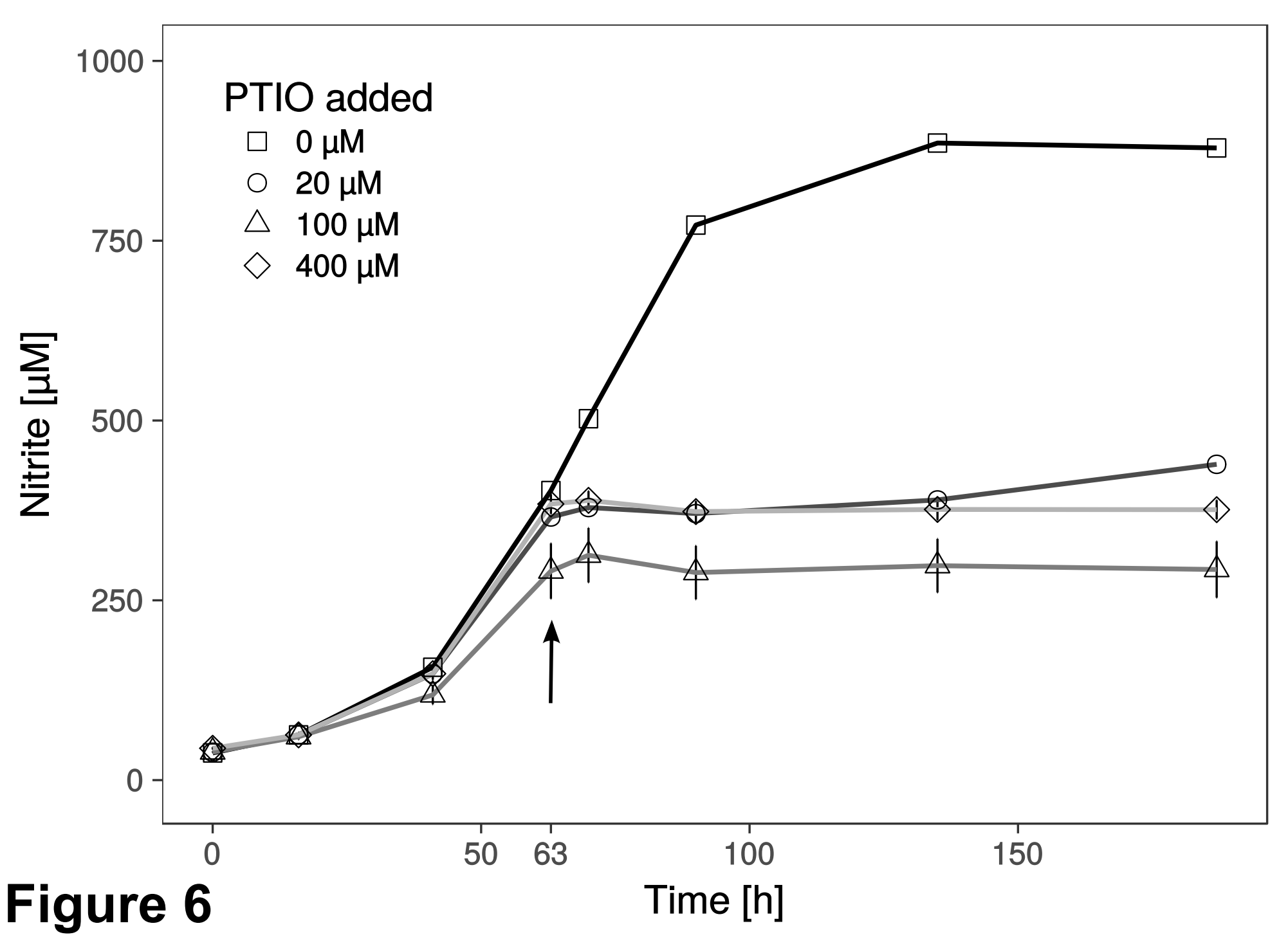
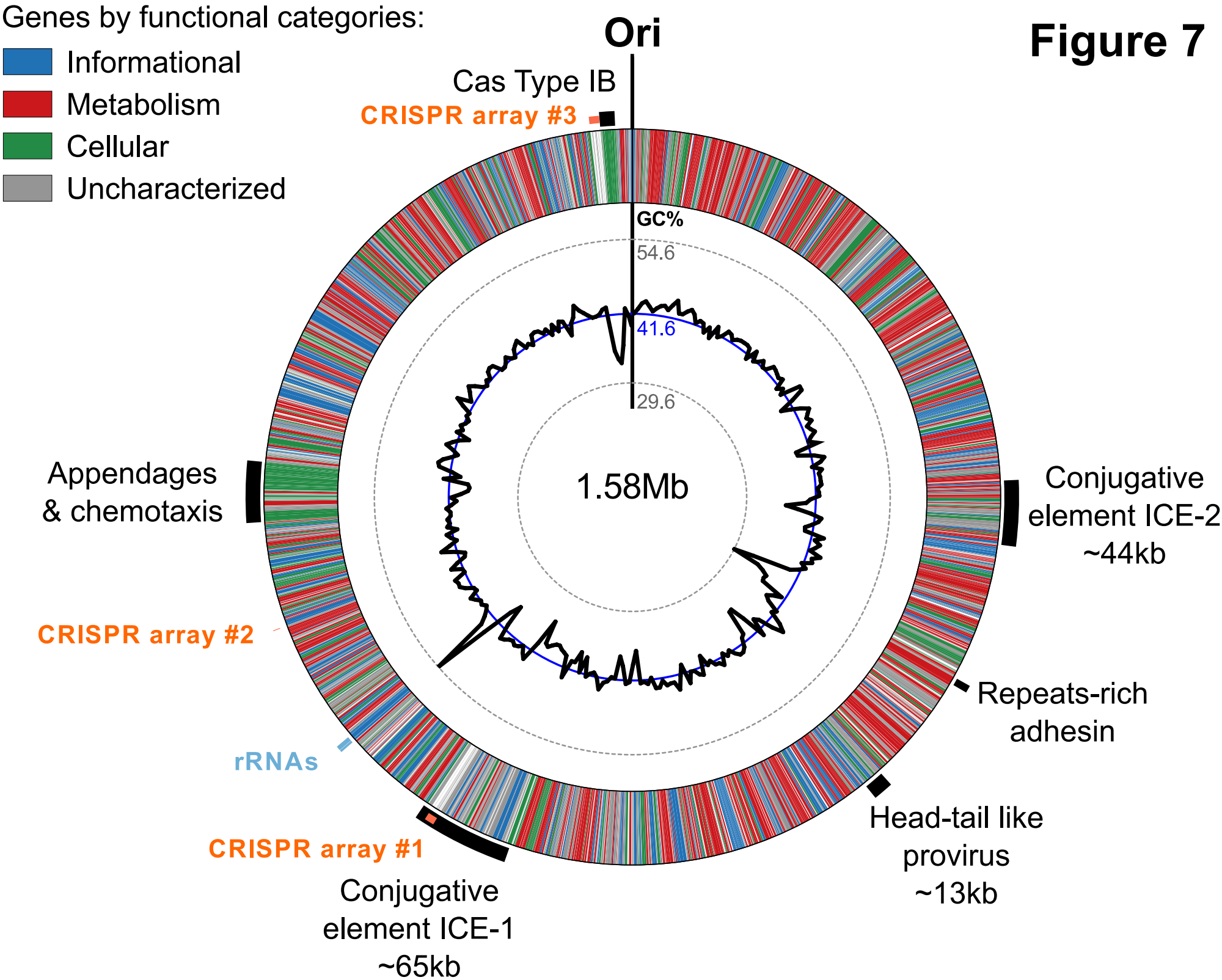

PTIO added
0 µM
20 µM
100 µM
400 µM
Nitrite [µM]
1000
750
500
250
0
0
50
63
100
150
Time [h]

Figure 6

Figure 7


Genes by functional categories:
Informational
Metabolism
Cellular
Uncharacterized
Ori
Cas Type IB
CRISPR array #3
GC%
54.6
41.6
29.6
1.58Mb
Appendages & chemotaxis
Conjugative element ICE-2 ~44kb
CRISPR array #2
Repeats-rich adhesin
rRNAs
Head-tail like provirus ~13kb
CRISPR array #1
Conjugative element ICE-1 ~65kb

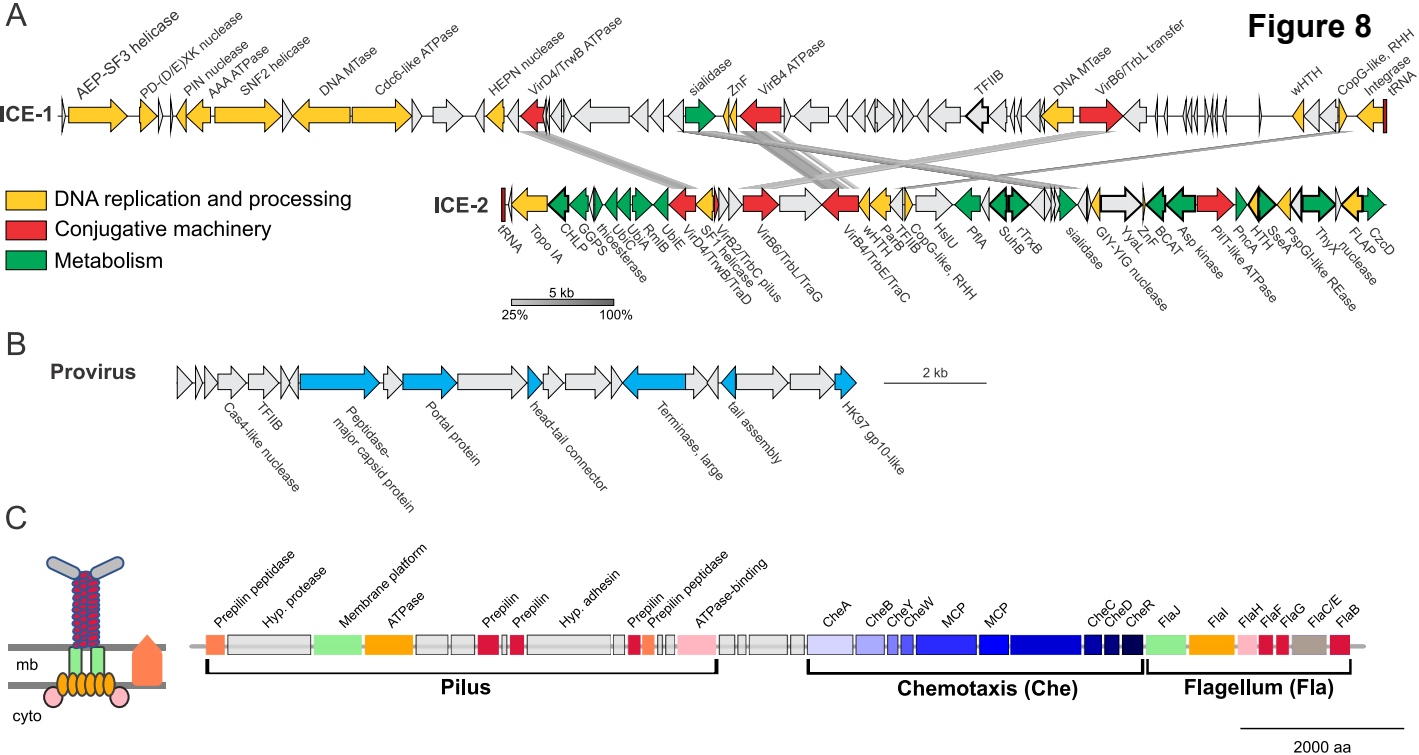


Figure 8
A
ICE-1
AEP-SF3 helicase
PD-(D/E)XK nuclease
PIN nuclease
AAA ATPase
SNF2 helicase
DNA MTase
Cdc6-like ATPase
HEPN nuclease
VirD4/TrwB ATPase
sialidase
ZnF
VirB4 ATPase
TFIIB
DNA MTase
VirB6/TrbL transfer
wHTH
CopG-like, RHH
Integrase
tRNA
DNA replication and processing
Conjugative machinery
Metabolism
ICE-2
tRNA
Topo IA
CHLP
GGPS
UbiA
Thioesterase
UbiA
RnfB
UbiE
VirD4/TrwB/TraD
SFI helicase
VirB2/TrbC pilus
VirB6/TrbL/TraG
VirB4/TrbE/TraC
wHTH
ParB
TFIIB
CopG-like, RHH
HslU
PilA
rTrxB
SuhB
sialidase
GlY-YlG nuclease
YyaL
ZnF
BCAT
Asp kinase
PilT-like ATPase
PhcA
HTH
SseA
PspGI-like REase
ThyX
nuclease
CzcD
FlAP
5 kb
25%
100%
B
Provirus
Cas4-like nuclease
TFIIB
Peptidase-major capsid protein
Portal protein
head-tail connector
Terminase, large
tail assembly
HK97 gp10-like
2 kb
C
mb
cyto
Prepilin peptidase
Hyp. protease
Membrane platform
ATPase
Prepilin
Prepilin
Hyp. adhesin
Prepilin
Prepilin peptidase
ATPase-binding
Pilus
CheA
CheB
CheY
CheW
MCP
MCP
CheC
CheD
CheR
Chemotaxis (Che)
FlaJ
FlaI
FlaH
FlaF
FlaG
FlaC/E
FlaB
Flagellum (Fla)
2000 aa