Genomes of Thaumarchaeota from deep sea sediments reveal specific adaptations of three independently evolved lineages

Running title: Genomes of Thaumarchaeal lineages from deep sea sediments

Melina Kerou[1*], Rafael I. Ponce-Toledo[1*], Rui Zhao[2#], Sophie S. Abby[1§], Miho Hirai[4], Hidetaka Nomaki[4], Yoshihiro Takaki[4], Takuro Nunoura[3], Steffen L. Jørgensen[2], Christa Schleper[1]

[1] Department of Functional and Evolutionary Ecology, Archaea Biology and Ecogenomics Unit, University of Vienna, Althanstrasse 14, 1090, Vienna, Austria.

[2] Department of Earth Science, K.G. Jebsen Centre for Deep Sea Research, University of Bergen, Bergen, Norway

[3] Research Center for Bioscience and Nanoscience (CeBN), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka, 237-0061, Japan.

[4] Super-cutting-edge Grand and Advanced Research (SUGAR) Program, X-star, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka, 237-0061, Japan.

[#]present address: School of Marine Science and Policy, University of Delaware, Lewes, Delaware 19958, USA

[§] present address: Univ. Grenoble Alps, CNRS, Grenoble INP, TIMC-IMAG, 38000 Grenoble, France.

[*] contributed equally

Abstract

Marine sediments represent a vast habitat for complex microbiomes. Among these, ammonia oxidizing archaea (AOA) of the phylum Thaumarchaeota are one of the most common, yet little explored inhabitants, that seem extraordinarily well adapted to the harsh conditions of the subsurface biosphere. We present 11 metagenome-assembled genomes of the most abundant AOA clades from sediment cores obtained from the Atlantic Mid-Ocean ridge flanks and Pacific abyssal plains. Their phylogenomic placement reveals three independently evolved clades within the order *Ca.* Nitrosopumilales, of which no cultured representative is known yet. In addition to the gene sets for ammonia oxidation and carbon fixation known from other AOA, all genomes encode an extended capacity for the conversion of fermentation products that can be channeled into the central carbon metabolism, as well as uptake of amino acids probably for protein maintenance or as an ammonia source. Two lineages encode an

additional (V-type) ATPase and a large repertoire of gene repair systems that may allow to overcome challenges of high hydrostatic pressure. We suggest that the adaptive radiation of AOA into marine sediments occurred more than once in evolution and resulted in three distinct lineages with particular adaptations to this extremely energy limiting and high-pressure environment.

Introduction

Ammonia oxidizing archaea (AOA) comprise one of the most successful archaeal phyla having colonized almost every imaginable oxic environment of the planet where they emerge as key players in the nitrogen cycle [1–6]. This includes the marine environment where they dominate archaeal communities associated with oxic sediments ranging from shallow estuaries to the open ocean [7–12], and from the surface layers all the way into the deep oceanic crust [13–15]. In these ecosystems they seem to play a critical role in the transformation of nitrogen compounds and control its partitioning into the bottom ocean and the underlying oceanic crust [12, 14–19]).

Studies from the North Atlantic and Pacific show that the composition of the sedimentary AOA population differs drastically from that in the overlying water suggesting distinct ecophysiological potential to colonise sedimentary environments, albeit all were found to belong to the order *Nitrosopumilales* (NP) [20]. Whereas the *amo*A-NP-gamma clade seems to be dominant and omnipresent in these oceans, irrespective of water depths, the *amo*A-NP-alpha clade represents the most abundant ecotypes in deep ocean waters [6, 7, 21–27] (nomenclature based *amo*A gene classification [28]). In contrast, dominant phylotypes in deep-sea sediments belong to the *amo*A-NP-theta and *amo*A-NP-delta clades [11, 12]. In addition, in cases of oligotrophic oceanic regions, these were detected throughout the sediment column and further into the underlying basaltic crust, even at depth where oxygen is below detection [7, 10–12, 14, 15]. In these, sites, they exhibit peaks of abundance and diversity at oxic/anoxic transition zones where increased energy availability is suggested to sustain the higher biomass of nitrifiers [12]. The abundance and distribution of *amo*A-NP-theta and -delta in the energy-starved subsurface suggest that they have adapted and evolved differently than their pelagic counterparts. These clades represent a yet unexplored diversity within *Nitrosopumilales*, and so far have no cultivated or genomic representative [28].

It has been suggested that the AOA common ancestor arose in terrestrial habitats (probably hot springs) where the AOA lineages diversified and then occupied different biomes (e.g. soils, hot springs and freshwater environments) before conquering estuarine and marine shallow water environments and finally, radiating into deeper waters as a result of the oxygenation of the deep ocean during the Neoproterozoic [29, 30] (Abby et al., submitted). AOA are generally well equipped for the manifold

challenges of the oxic deep-sea surface and subsurface environment. They encode the most energy-efficient aerobic carbon fixation pathway [31] making them important primary producers in these environments [32, 33], and their high affinity for ammonia would enable them to utilize this scarce resource [34]. Nevertheless, deep pelagic as well as benthic AOA populations are reported to have the capability for mixotrophy as well, as indicated by uptake of labelled compounds and through the detection of uptake/assimilation genes for organic carbon and nitrogen compounds by shotgun metagenomics [10, 16, 23, 24, 33, 35–37]. Stimulation of autotrophic CO_2 fixation by organic carbon was also shown by isotope labelling studies [32]. In the absence of genomic context however, virtually nothing of the above can be extrapolated to the metabolic potential or adaptations of ecotypes that dominate deep marine sediments, nor can their ecological boundaries be interpreted.

In this study, we address the question of what adaptations enabled specific AOA clades to inhabit bathyal and abyssal (i.e. deep-sea) marine sediments, and the significance of this in the context of thaumarchaeal evolution. To this end, we obtained the first high-quality metagenome-assembled genomes (MAGs) belonging to the so-far uncharacterized *amo*A-NP-theta and *amo*A-NP-delta clades from sediment cores obtained from the Mid-Atlantic Ridge flanks, and from the oligotrophic Pacific Ocean. We also describe two MAGs associated with a novel, deep-branching clade within the *Nitrosopumilales*, which we designate *amo*A-NP-iota (previously NP - *insertae sedis* [28]). The pivotal phylogenetic position of the latter and the distribution of all three clades in phylogenomic trees enables us to shed light on the evolutionary diversification of AOA into marine sediments, which seems much more complex than previously assumed and reveals unique, similar, and also overlapping adaptive strategies in all three clades.

Materials and Methods

Sampling of Atlantic and Pacific sediments

Oligotrophic sediment cores were retrieved from mid-ocean ridge flanks in the Atlantic Ocean: Hole U1383E (22°48.1'N, 46°03.2'W, 4425 m water depth) from North Pond by advanced piston coring during the International Ocean Drilling Program (IODP) Expedition 336 (2011), and GS14-GC08 (71°58.0' N, 0°6.1' E, 2476 m water depth) by gravity coring from the east flank of the central Mohns Ridge (2014) (Fig. 1a). Genomic DNA from four sediment horizons in each core were selected for metagenome sequencing, based on the published porewater geochemical data and 16S rRNA gene profiles [12, 38]. In particular, sediments of 0.1 m (oxic), 10.0 m (oxic), 22.0 m (oxic-anoxic transition zone; OATZ), and 29.5 m (anoxic-oxic transition zone; AOTZ) were selected from U1383E. Sediments of 0.1 meters below the seafloor (mbsf (oxic), 1.0 m (OATZ), 1.6 mbsf (nitrate-ammonium transition

zone), and 2.5 mbsf (Mn-reduction zone) were selected from GS14-GC08. Detailed information about sampling sites, sampling procedure, 16S rRNA gene profiles and porewater analysis was published in [12, 38].

Sediment cores (YK1309-1N and YK1312-12N) from the Pacific abyssal plain were collected using a push corer with a manned submersible *Shinkai6500* during the JAMSTEC cruises YK13-09 (September 2013: 01º15.0'N, 163º14.9'E, 4277 m water depth) and YK13-12 (November 2013: 11º59.9'N, 153º59.9'E, 5920 m water depth) of the *R/V Yokosuka,* respectively. Two sections from each core were selected for shotgun metagenomic sequencing: YK1309-1N-S000 (0-1 cmbsf), YK1309-1N-S300 (30-35 cmbsf), YK1312-12N-S010 (1-2 cmbsf) and YK1312-12N-S200 (20-25 cmbsf).

DNA of all samples was prepared using standard techniques and was sequenced on Illumina Hiseq2500, 16S rRNA gene amplicons were generated and sequenced using standard procedures (see Suppl. Material). Detailed information about sampling sites, sampling procedure, and geochemical analyses are also shown in Suppl. Material.

Assembly and comparative genomics

All sequencing data were processed to remove illumina adapters and low quality reads using Trimmomatic [39] before *de novo* assembly using MEGAHIT [40] (k-mer length of 27-117). Binning of contigs of Pacific metagenomes was performed applying a contig dereplication and binning optimization tool [41] based on the binning output of CONCOCT [41], MetaBAT [42] and MaxBin2 [43] while contigs of Atlantic samples were binned with MaxBin2 [43] followed by a sequence of refinement steps for the Thaumarchaeal bins (see supplementary methods). Completeness and contamination of Pacific and Atlantic bins were evaluated with CheckM ("lineage_wf" parameter) [44]. Assemblies are available on NCBI (accession numbers pending).

A dataset consisting of 163,852 predicted proteins from 85 genomes (11 metagenome-assembled genomes (MAGs) reported here, 31 complete and near-complete AOA genomes and 43 MAGs or single-amplified genomes (SAGs) from NCBI or IMG) was collected for this study (Table S1). Annotation of the MAGs assembled in this study was performed automatically using the Microscope annotation platform from Genoscope [45], followed by extensive manual curation. NCBI annotations were supplemented with arCOG assignments from the archaeal Clusters of Orthologous Genes database (2018 release) [46] using COGsoft [47] (e-value of 10^{-10}). We clustered the protein dataset into protein families based on sequence identity (35 %) and alignment coverage (70 %) using CD-Hit V4.8.1 [48] ("-c 0.35 -aL 0.7 -aS 0.7") (Table S3).

146

147 **Selection of markers and phylogenomic tree**

148

149 The identification of markers to perform the phylogenomic tree reconstruction was based on the
150 phylogenomic workflow proposed by [49] (E-value 10^{-10}) using the archaeal single-copy gene
151 collection [50]. We selected 79 markers (Table S2), present in at least 70 of the 85 genomes used in this
152 study. Each protein family was aligned using MAFFT v7 ("--maxiterate 1000 –localpair") [51] and
153 trimmed with BMGE [52]. The concatenated alignment was used to reconstruct a Maximum likelihood
154 phylogenomic tree in IQTREE (v2.0-rc1) [53] under the LG+C20+F+G model with 1,000 ultrafast
155 bootstrap replicates. For *amo*A phylogeny and detailed methodological procedures, see Supplementary
156 Material.

157

158 **Results and Discussion**

159

160 **Distribution of AOA in deep marine sediments**

161

162 We examined the overall community structure of AOA (all affiliated to the family *Ca.*
163 Nitrosopumilaceae) in these sediments by analyzing 16S rRNA gene amplicon sequencing data
164 generated in this study for the Pacific cores and previously described for the Atlantic cores [12]. AOA
165 communities in sediment horizons deeper than 10 cm were all dominated by the so-called 16S-NP-eta
166 and/or 16S-NP-upsilon and 16S-NP-alpha clades [55], which together correspond to the *amo*A-NP-
167 theta clade (Fig. 1b) [28]. In addition, AOA affiliated to the 16S-NP-epsilon clade (corresponding to
168 the *amo*A-NP-delta clade, Fig 2a) were also repeatedly detected with percentages <25% in the upper
169 portions of these cores (Fig. 1b). Finally, the 16S-NP-lambda clade (now renamed to *amo*A-NP-iota,
170 see below) was also detected as a minor clade in all cores except YK1312-12N, but was notably
171 abundant in the uppermost horizon of GS14-GC08 (29% of the total AOA community, Fig. 1b).

172

173 **Phylogenomic analysis and taxonomic placement reveal MAGs from three independent *amo*A**
174 **clades**

175

176 We obtained a total of 11 AOA metagenome-assembled genomes (MAGs), 9 from Atlantic and 2 from
177 Pacific sediment samples (sequenced horizons are marked by stars in Fig. 1b). Despite high sequencing
178 depths and high AOA abundance in the sample, based on 16S rRNA gene reads in the metagenomes
179 (10 - 16 % in the Pacific cores and 6.8-18.8% in the Atlantic cores), generation of good quality bins
180 was extremely challenging, possibly due to high microdiversity (41 OTUs) within *Ca.* Nitrosopumilales
181 as observed earlier [12]. Eventually we obtained 4 MAGs with >90 % completeness and 3 MAGs with

>80 % completeness, all with contamination levels ≤ 5%, which we consider high quality MAGs in this study, as well as four additional medium quality MAGs (66 to 76% completeness, up to 6.3% contamination level) (Table1). The MAGs genome sizes (0.61 to 1.52 Mb) and GC contents (34.66 ± 0.72%) are in accordance with previous reports of free-living *Ca.* Nitrosopumilales [56].

In order to study the evolution of AOA and place our deep marine sediments-derived MAGs in a phylogenetic context, we reconstructed a maximum-likelihood (ML) phylogenomic tree (Fig. 2a) using 79 concatenated single-copy markers from our entire dataset of 85 complete genomes, MAGs and SAGs representing a broad diversity of habitats (Table S1).

In addition, we performed an *amo*A-based phylogeny as in [28] in order to assign a taxonomical rank and a respective AOA clade to our MAGs (Fig. 2b). Both trees showed similar clustering of MAGs into *Ca.* Nitrosopumilales subclades except for NPMR_NP_delta_1 (see discussion in Supplementary Information) which based on the *amo*A tree clustered within the *amo*A-NP-theta clade but the more robust phylogenomic analysis strongly suggests that it belongs to the *amo*A-NP-delta subclade. The only 16S rRNA gene recovered in MAG NPMR_NP_theta_3 is affiliated with a subclade of 16S-NP-alpha exclusively found in marine sediment (not shown).

Our phylogenomic tree revealed that the 11 AOA MAGs reported here represent the dominant AOA observed in our study (Fig. 1b) and form three well supported monophyletic groups, of which no cultured representative has been reported yet (Fig. 2a). Six MAGs represent the first genomic assemblies from the *amo*A-NP-theta lineage, one of the most dominant AOA groups in marine sediments and also found to be abundant in the crust below [14, 28]. Three MAGs are affiliated to *amo*A-NP-delta, the second most abundant AOA clade in marine sediments, and are the first marine sediment representatives of this clade, which includes a single other MAG (Archaeon CSP1) assembled from river aquifer sediments [57].

Two bins (NPMR_NP_iota_1 and YK1309_NP_iota) clustered together forming a third sediment-dwelling clade, sister to all *Ca.* Nitrosopumilales, which earlier escaped taxonomic assignment as it was only identified based on singular *amo*A sequences and hence had been designated *incertae sedis* [28]. Pairwise average nucleotide identity (ANI) comparisons (Fig. S2), indicate that these two bins share > 70 % ANI with the other NP-MAGs recovered in this study (*amo*A-NP-delta and *amo*A-NP-theta MAGs sharing 73 - 79 % ANI). A comparison of conserved protein families among all NP subclades indicated that this group harbors 320 out of 336 protein families that seem to be part of the *Ca.* Nitrosopumilales core proteome, as opposed to only 260 of these protein families being present in the *Ca.* Nitrosotaleales, the sister lineage to all NP (Fig. 3). Moreover, environmental *amo*A sequences

suggested that this clade might be restricted to deep-sea sediments, an ecological specialization only found in *Ca.* Nitrosopumilales. Taken together, this early-branching clade seems to be a new NP subclade and we propose the designation *amo*A-*NP*-iota (as it forms the ninth NP-clade following the taxonomy of [28]).

Three independent radiations of AOA into marine sediments

While MAGs and SAGs of AOA from bathypelagic (1000-4000m) to abyssopelagic (4000-6000m) and hadopelagic (6000-11000m) environments have been reported previously [23, 24, 58] and shotgun metagenomic analyses of deep-sea sediments have been performed [10, 59, 60], the MAGs reported in this study represent to our knowledge the first high-quality AOA genomes from bathyal and abyssal sediments (> 2000 m depth). Together with our taxon-enriched phylogenomic analyses they shed new light on the ecological transitions and niche differentiation undergone by AOA.

While our phylogenomic tree supports that AOA likely appeared in terrestrial habitats first (Fig. 2a), the sequence and number of colonization events of marine environments seem to be more complex than previously proposed [29]. Importantly, our results suggest that both the deep-water adapted AOA as well as the deep-sediment adapted AOA are polyphyletic. The colonization of deep waters, i.e. pelagic organisms, might have occurred independently at least twice in the evolution of the *Ca.* Nitrosopumilales, once at the origin of *amo*A-NP-alpha clade and the other during the diversification of *amo*A-NP-gamma (Fig. 2a). Interestingly, the *amo*A-NP-gamma clade which is one of the most diverse *Ca.* Nitrosopumilales subclades [28] has undergone particular habitat transitions and niche occupation [28]. Distinct shallow water *amo*A-NP-gamma species have established independently symbiotic associations with sponges [56, 61] while the sub-lineage leading to the soil isolate *Ca.* Nitrosarchaeum koreense [62] could have evolved from an estuarine or shallow water ancestor suggesting a recolonization of land (Fig. 2a).

Regarding the origin of deep sediments-dwelling AOA, the *amo*A-NP-theta lineage branches within mostly marine NP clades (Fig. 2a), suggesting that this lineage might have evolved from a pelagic marine ancestor. However, there is no evolutionary link between any pelagic AOA and the NP-delta clade, as the latter have been mostly retrieved in estuarine and deep marine sediments [57]. The most parsimonious evolutionary scenario would be that this group underwent a direct transition from estuarine sediments to marine sediments during its diversification.

Similarly, the newly proposed clade NP-iota, the earliest branching Nitrosopumilales which has so far exclusively been detected in marine sediments [28] does not seem to be closely related to pelagic

Nitrosopumilales but emerges instead among terrestrial clades (*Ca*. Nitrosotaleales and NP-Eta). Although, it is possible, that pelagic AOA closely related to *amo*A-NP-iota may be detected in further environmental surveys or that respective pelagic lineages got extinct, the *amo*A-NP-iota clade might as well have developed from terrestrial-estuarine organisms, as discussed for *amo*A-NP-delta above.

Comparative genomics of deep-sea sediment AOA

We constructed a total of 33,442 protein families from our taxon-enriched genome dataset representing a wide variety of ecological environments (see Materials and Methods and Table S1). From these, 12,137 have representatives from at least two different genomes. In our analysis, the AOA core proteome comprises 760 protein families present in at least one genome of each of the four major AOA lineages: *Ca*. Nitrosocaldales, *Nitrososphaerales*, *Ca.* Nitrosotaleales and *Ca.* Nitrosopumilales (Fig. 3, Table S3). Thus, our results are similar to previous estimations of the AOA core genome (743 gene families) [63], and slightly lower than our own earlier estimate of 860 gene families (based on only 7 genomes [64]). Only 269 of the core AOA families seem to be AOA-specific (Fig. 3, Table S3). Only 123 out of these 269 families were found to be present in >50 % of the genomes in each of the four AOA orders (a relatively low threshold to account for the incompleteness of MAGs), suggesting a relatively low degree of conservation within these lineages. These results imply great intra-order genomic variability and important differential gene loss among subclades and across genomes during the evolution and diversification of AOA. For instance, despite the fact that Nitrosopumilales have 3091 specific families with proteins encoded in at least two genomes and present in one or more NP-subclades, a subset of solely 24 families were conserved in all 7 NP-subclades (Fig. 3). Considering the very relaxed criteria used, this is an astonishingly small number of conserved families in all 7 NP-subclades.

To identify possible specific adaptations of AOA to deep marine sediments, we searched for families present in at least two of the three marine sediments clades represented by our 11 MAGs (i.e. *amo*A-NP-theta, -delta and -iota), to the exclusion of all the other genomes analyzed in this study (Fig S2, Table S1). 72 families were identified (Fig. 3), of which only 25 % (18 families) could be functionally annotated (Table S3) and were classified into the following categories: information processing systems (7), metabolism (5) and cellular processes (6). Some of these 18 families had functional equivalents in most if not all AOA (e.g. RadA homologs). We additionally found 41 families shared predominantly between NP sub-clades with deep ocean (>1000m) representatives (i.e. *amo*A-NP-alpha, NP-gamma sublineages recovered from the Mariana, Izu-Ogasawara Trenches and the Red Sea [23, 58], NP-theta, NP-iota and NP-delta), to the exclusion of all other NP sub-clades. From these 41 families, 18 have functional annotation: information processing systems (8), metabolism (4) and cellular processes (6).

Families with functional significance specific to marine sediments, such as a putative lactate racemase, or those shared with deep ocean MAGs, are discussed below. Families identified in deep ocean MAGs but not found in the sediment clades are still depicted in Fig. S4 for comparative purposes. Additionally, we investigated the number of clusters shared between deep sediments-derived MAGs and the terrestrial (present in soils and sediments) lineage *Nitrososphaerales*, to the exclusion of all other AOA lineages and NP-subclades. Interestingly they share only one protein family related to coenzyme F_{420}-dependent luciferase-like oxidoreductases.

Metabolic reconstruction of the *amo*A-NP-theta, *amo*A-NP-delta, *amo*A-NP-iota clades

The full annotations for all genes and pathways discussed in the following section can be found in Table S4.

Central energy and carbon metabolism

All three sediment clades (i.e. *amo*A-NP-delta, *amo*A- NP-theta, *amo*A- NP-iota) encode complete sets of genes involved in ammonia oxidation, namely *amo*AXCB in the typical organization observed in other *Nitrosopumilales* (Fig. 5, Table S3) [21, 65, 66]. Missing subunits in certain MAGs seem to be due to genome incompleteness. A nitrite reductase (NirK) homolog is present, as well as multiple blue copper domain proteins putatively functioning as electron carriers. All clades encode a single high-affinity ammonia transporter family protein (Amt), as opposed to two Amt transporters of differing affinities found in other AOA. This could represent an adaptation to an oligotrophic environment [67]. Four out of six NP-theta MAGs and all *amo*A-NP-delta MAGs encode complete or near-complete urease operons (Fig. 4, Table S3). Together with a putative nitrilase (Nit1, conserved in AOA) and a putative omega-amidase (Nit2, present in *amo*A-NP-theta, -delta, -eta, -gamma), these genes indicate expanded substrate utilization capabilities for ammonia (and CO_2) generation by cleaving urea, nitriles and dicarboxylic acid monoamides. Utilization of organic nitrogen compounds is a feature shared with other NP-clades that include deep-sea lineages and previously described for subseafloor AOA (Fig. 4, 5, S4) [10, 23, 24, 33].

All three sediment clades encode full gene sets for electron transfer to O_2 via NADH dehydrogenase (complex I), type bc_1 complex III, and a heme-copper terminal oxidase (complex IV) (Fig. 5, Table S3). No alternative complexes using a different electron acceptor were identified.

All three sediment clades encode the full repertoire conserved among AOA for autotrophic carbon fixation via the 3-hydroxypropionate/4-hydroxybutyrate (3HP/4HB) cycle and carbon metabolism through oxidative TCA and gluconeogenesis up to the formation of glucose-6P via a phosphoglucose

isomerase (not present in NS) homolog in NP-theta, NP-delta (Fig. 5, Table S3) [31, 65, 68]. A malic enzyme, enabling the formation of pyruvate from malate with the concomitant generation of NAD(P)H expands metabolic capacities in NP-theta and NP-iota (also in some other AOA, Fig. 5). As most AOA, all sediment clades have the capacity to synthesize polyhydroxybutyrate (PHB) storage compounds, an obvious advantage in an oligotrophic environment [69].

Complete or near-complete amino acid biosynthesis pathways as well as vitamins (including vitamin B12) are present in all three sediment clades, as in other AOA (Table S3). As observed in NP-alpha representatives [23], the sediment clades use the B12-independent pathway for methionine biosynthesis (*metE*) (Fig. 4). Albeit this being a less catalytically efficient enzyme than the B12-dependent *metH* present in all other AOA, it is nevertheless much less costly energetically [70], and would therefore be an advantage in an energy-limiting environment where maintenance rather than fast growth is the norm [71].

Utilization of exogenous organic compounds

All three sediment lineages seem to be capable of utilizing exogenous organic compounds from fermentation processes such as formate, lactate and 3-aminobutyryl-CoA as a source of carbon, nitrogen and reductive potential. This finding expands the range of organic carbon and nitrogen substrates suggested earlier for deep ocean AOA (previously comprising amino-acids, peptides and compatible solutes) and reinforces their role as key players in nutrient cycling in these biomes [6, 10, 23, 24, 27, 33].

A putative soluble NAD^+-dependent formate dehydrogenase (Fdh), distinct from the iron-sulfur/molybdenum containing Fdh enzymes traditionally found as part of formate-hydrogen lyase systems [72], is found in the NP-theta and NP-iota clades (as well as in certain NS representatives and NP-alpha Fig. 4, 5). However, no additional hydrogenases or known formate transport systems were identified in the marine sediment bins. Our phylogenetic analysis (Fig. S3) indicates that the enzyme is a bona fide NAD^+-dependent Fdh within the superfamily of D-2-hydroxyacid dehydrogenases [73]. This indicates the capacity to use formate for supplementing CO_2 needs while concomitantly supplying reducing equivalents (as in methylotrophs [74, 75]).

A putative 3-aminobutyryl-CoA aminotransferase (Kat, EC 2.6.1.111) and a 3-aminobutyryl-CoA ammonia lyase (Kal, EC 4.3.1.14) were identified in the NP-theta and NP-delta bins, and are also found in some NP-alpha, NP-gamma and NT lineages (Fig. 4, 5). These enzymes participate in lysine fermentation pathway variants in fermentative bacteria [76]. Although the key pathway enzymes are not present in the sediment bins or any other AOA, this intermediate compound (3-aminobutyryl-CoA)

could be scavenged from fermenting microorganisms in the sediment community. Both enzymes can remove ammonia from 3-aminobutyryl-CoA either by transferring it to α-ketoglutarate resulting in the formation of acetoacyl-CoA and glutamate (Kat), or by an elimination reaction that produces crotonyl-CoA and free ammonia (Kal). Both products are intermediates of the 3HB/4HP (CO_2 fixation-) pathway and could be processed accordingly, generating reducing potential in the subsequent steps.

The presence of a putative lactate racemase family protein (LarA), specific to the NP-theta, delta and iota clades (Fig. 4, 5) suggests that lactate is another fermentation product that could be utilized by these lineages. This is one of the very few protein families with a putative function prediction shared specifically between the sediment AOA clades to the exclusion of all other AOA, suggesting an essential role. LarA in lactobacilli catalyzes the interconversion of D- and L-lactate, ensuring an adequate supply of D-lactate which is an important cell wall component conferring resistance to vancomycin [77] (see Supplementary Information). Given the importance of cell envelope maintenance in the adverse conditions of the sediments, it is possible that D-lactate has a similar use in sediment AOA, conferring resistance to exogenous toxic compounds. Alternatively, the lactate dehydrogenase-like malate dehydrogenase homologs found in AOA possess features indicating that they could have a broad substrate specificity, being able to utilize pyruvate in addition to oxaloacetate, and producing the L-stereoisomers of the products (see Supplementary Information and Fig. S5), with the concomitant reduction of NAD^+ [78].

As mentioned above, the only protein family specifically shared among the sediment MAGs and the terrestrial NS lineages, is an F_{420}-dependent luciferase-like oxidoreductase. While the metabolic role of these proteins in AOA in general is still unclear, the ability to degrade recalcitrant carbon via oxygenases in a manner similar to terrestrial organisms has been observed in sediment communities [60, 69], and is proposed to provide an opportunistic advantage for expanded substrate utilization in limiting conditions.

Adaptations to low energy and high pressure environments

Deep-sea sedimentary environments found under the oligotrophic ocean present manifold challenges to microbial life, namely energy limitation, high hydrostatic pressure (HHP), low temperatures (< 4°C) and potential microoxic or anoxic conditions detrimental to aerobic metabolisms [18, 19, 71, 81–84]. Microorganisms respond with global metabolic changes rather than stress responses [85], some of which are found in the deep sediment AOA clades.

Many organisms possessing distinct electron transport, ion gradient generating and ATP synthase complexes that are differentially regulated under HHP [71, 86–88]. Interestingly, both NP-iota MAGs and two out of six NP-theta MAGs encode complete gene clusters for both the A-type ATPase found in neutrophilic AOA and V-type ATPase variant found in acidophilic/acidotolerant/piezotolerant archaea (and AOA) which is homologous to the proton/ion pumping ATPases from eukaryotes and enterococci (Fig. 4, 5) [89, 90]. The remaining NP-theta MAGs encode either the A-type or the V-type ATPase (although this could be attributed to the incompleteness of the MAGs). The V-type ATPase has been suggested to confer physiological advantages in high pressure environments by virtue of its proton-pumping function [89]. This would enable the maintenance of intracellular pH, which is disrupted by the accelerated release of protons from weak acids (such as carbonic acid) under HHP [91]. The presence of both ATPase variants is also observed in abysso/hadopelagic NP-gamma AOA lineages, while the deep marine NP-alpha encode only the V-type ATPase (Fig. 4,5 and Supplementary Information for further discussion) [89]. In contrast, all three NP-delta MAGs encode only the canonical A-type ATPase (Fig. 5), but intriguingly at least two of them seem to contain a partially duplicated NADH dehydrogenase (complex I) operon which could similarly be responsible to alleviate cytoplasm acidification (see Supplementary Information).

The cytoplasmic membrane is severely affected by HHP, which induces a tighter packing of the lipids and a transition to a gel state, resulting in a decrease in fluidity and permeability [92, 93]. The presence of a N-acetylneuraminic acid mutarotase (NanM) in NP-theta, NP-iota, NP-delta MAGs (Fig. 4, shared with a few abyssopelagic/hadal NP-gamma species) indicates the ability to acquire sialic acid [94]. This important component of glycoconjugates found on cell walls has multiple functions including concentrating water on cell surfaces [95] and regulating membrane permeability [96]. It can also enable the regulation of the thickness of the hydration layer surrounding the cell membrane [97], which could prevent system volume change and stabilize membrane protein complexes and membrane structure under pressure [97, 98], while also regulating membrane permeability [96]. Modification of the hydration layer properties has also been identified as a specific adaptation mechanism of the piezophilic archaeon *Thermococcus barophilus* [99].

An ABC-type branched-chain aminoacid transport system of the HAAT family (3.A.1.4) is present in all three sediment clades as well as in one NP-alpha MAG, sponge-associated and few other lineages of the NP-gamma clade, NS and non-AOA Thaumarchaea (Fig. 4, 5, Table S3). The uptake of amino acids has been interpreted earlier as indicative of the possibility of organic carbon utilization via enzymes participating in canonical amino acid biosynthesis pathways and present in all or most AOA (e.g. aspA, ilvA, ilvE, aspC glyA, GDH, ProDH) [10, 23, 37, 61]. Such mixotrophic strategies are also responsible for the enormous ecological success in oligotrophic environments of oceanic cyanobacterial lineages [100]. However, canonical amino acid degradation key enzymes such as amino acid

hydroxylases, the branched-chain α-keto acid dehydrogenase complex or 2-ketoacid:ferredoxin oxidoreductases have not been detected in deep sea or sediment AOA clades, nor are their genomes particularly enriched in proteases (Fig. S4). On the other hand, a metabolic shift from expensive *de novo* biosynthesis of cellular materials (with proteins accounting for 56% of total energy investment in oxic environments) to recycling of exogenous or endogenous resources is observed in HHP-adapted microorganisms. [16, 18, 69, 85, 86, 101, 102]. Therefore, it seems more likely that amino acids are used for recycling, as suggested earlier for AOA by isotope tracer and NanoSIMS experiments with sediment and oceanic crust communities [16, 32, 33, 69], and as observed in the obligate piezophile *T. barophilus* and other facultative piezophiles [102–104]. Moreover, amino acids (mostly glutamate, proline and glutamine) can be accumulated as compatible solutes to ensure the stabilization of macromolecular structures upon pressure or temperature related stress [104, 105]. It cannot be ruled out though that amino acids are also used for replenishing the intracellular ammonia pool, with minimal production (if at all) of reducing equivalents (Fig. 5) (see Supplementary Information for detailed discussion).

The sediment clades and especially NP-theta encode an extended repertoire of enzymes for DNA and protein repair compared to other NP (details in Supplementary Information and Fig. 4, 5, S4). This is an indication of energy investment towards maintenance of cellular components, rectifying damage due to low turnover rates and cellular aging rather than active and fast growth [71, 82, 106]. This strategy together with dormancy is presumed to be responsible for persistence in subseafloor energy limited environments [107].

Osmoregulation

All sediment clades do encode a putative bifunctional CTP:inositol-1-phosphate cytidylyltransferase/di-myo-inositol-1,3′-phosphate-1′-phosphate synthase (ipct/dipps), responsible for the synthesis of the compatible solute di-myo-inositol-1,3'-phosphate (DIP) [108]. The enzyme is also found in deep marine AOA clades [23] (Fig. 4, 5). Biosynthetic genes for this compatible solute have so far only been observed in organisms growing above 55°C, and have been extensively transferred between archaea and bacteria [109], making these AOA clades the first non-thermophilic organisms with the ability to synthesize this inositol derivative. Compatible solutes can confer resistance to various types of stress, so it is possible that this anionic solute has multiple roles in these polyextremophilic organisms [110] especially since no pathways for synthesis/uptake of known osmolytes such as mannosylglycerate, ectoine/hydroxyectoine or glycine/betaine were identified in the NP-theta, NP-iota, NP-delta MAGs (Fig. 4, 5, S3).

Conclusions

Our comparative and phylogenomic analyses using 11 sediment-derived MAGs reported in this study, together with a large collection of AOA genomes with a broad phylogenetic and ecological distribution allowed us to study the evolution, diversification and adaptation mechanisms of AOA into deep marine environments. Based on phylogenomic analyses and different from earlier scenarios [111] we conclude that AOA from deep marine sediments evolved independently within (at least) three lineages. Although it seems that the ancestor of the *amo*A-NP-theta clade was pelagic and descendants of it occupied the deep marine sediments and the oxic subseafloor crust, it is likely that in the case of *amo*A-NP-iota and *amo*A-NP-delta, there was a transition from terrestrial ecosystems/freshwater sediment to marine sediments without having colonized the ocean water column first. Interestingly, all extended capacities and adaptations discussed in this manuscript are found to be combined in lineage *amo*A-NP-theta, which represents the most widely distributed and abundant clade ranging over different marine sediment layers, whereas the other two clades that share some of these features exhibit a more distinct distribution pattern.

All AOA adapted to marine sediments and investigated in this study are able to perform ammonia oxidation in combination with CO_2 fixation like all other described AOA. In addition, all three lineages seem to be capable of utilising exogenous organic fermentation products that they convert into intermediates of their central carbon metabolism, a feature they share with pelagic AOA from the deep ocean and a few other AOA. This, together with the capability of taking up aminoacids, putatively for recycling into proteins or utilization of amine groups, would support growth in this extremely oligotrophic environment and contribute to organic nitrogen and carbon turnover in the sediments. In the absence of any components indicating increased capacity of amino acid degradation in these AOA we argue that recycling of amino acids rather than catabolism as otherwise suggested in [10, 23, 113] represents an advantageous and more plausible strategy for the sedimentary AOA clades. It is also a trait frequently observed in other sedimentary and crustal population groups to overcome the prohibitive energetic costs of *de novo* monomer biosynthesis [69]. Additionally, a broad repertoire of DNA and protein repair enzymes, seem to enable the deep sediment-adapted AOA to counteract the most severe consequences of cellular aging. An important feature shared with HHP-adapted deep marine clades is the presence of two ATPase complexes in *amo*A-NP-theta and *amo*A-NP-iota, with putatively opposing functions that would alleviate the effects of pH imbalance due to HHP, as well as the PMF-destabilizing effects of age-induced membrane leakage. These features shed light onto the mechanisms underlying AOA persistence in the benthic environments beneath the open ocean, from the surface sediments down to the underlying oceanic crust, and further consolidate the central role of these archaea in the global biogeochemical cycles.

Author Contributions

CS, SLJ & TN conceived the study. RZ & SLJ sampled and processed the Atlantic sediments, while TN, HN, MH & YT sampled and processed the Pacific sediments. RZ, SSA & RP assembled the MAGs. RP performed the phylogenetic and comparative genomic analyses. MK annotated and analyzed the genomes. MK, RP, RZ, SSA, SLJ, CS & TN interpreted data. MK, RP & CS wrote the manuscript with contributions from TN, RZ, SLJ & SSA.

Acknowledgements

We thank Philipp Weber for help with visualizations. The LABGeM (CEA/Genoscope & CNRS UMR8030), the France Génomique and French Bioinformatics Institute national infrastructures (funded as part of Investissement d'Avenir program managed by Agence Nationale pour la Recherche, contracts ANR-10-INBS-09 and ANR-11-INBS-0013) are acknowledged for support within the MicroScope annotation platform. We thank onboard scientists, officers and crews of RV *Yokosuka* and the manned submersible *Shinkai 6500* for their help and operation during the YK13-09 and YK13-12 cruises. We wish to acknowledge the entire scientific party and all crewmembers onboard *Joides Resolution* during IODP expedition 336, for their help and expertise. Especially, Co-chief scientists Katrina Edwards and Wolfgang Bach. We also wish to thank chief scientist of the CGB summer cruise 2014 Rolf Berger Pedersen for the sediment coring opportunity and thank Ingeborg Ørkland, Desiree Roerdink, Tamara Bumburger and Ingunn H Thorseth for their help with porewater extraction and nutrient analysis. This project was funded by FWF grant 27017 and ERC AdvGr TACKLE (695192). SLJ was funded by The Trond Mohn starting grant BFS2017REK03 and the Census of Deep Life phase V grant. TN was partially supported by JSPS KAKENHI Grant Number JP19H05684 within JP19H05679 (Post-Koch Ecology).

Competing Interests

The authors declare no competing financial interests.

547 **References**

548

549 1. Schleper C, Nicol GW. Ammonia-oxidising archaea--physiology, ecology and evolution. *Adv*
550 *Microb Physiol* 2010; **57**: 1–41.
551 2. Stahl DA, de la Torre JR. Physiology and diversity of ammonia-oxidizing archaea. *Annu Rev*
552 *Microbiol* 2012; **66**: 83–101.
553 3. Hatzenpichler R. Diversity, physiology, and niche differentiation of ammonia-oxidizing
554 archaea. *Appl Env Microbiol* 2012; **78**: 7501–7510.
555 4. Pester M, Schleper C, Wagner M. The Thaumarchaeota: an emerging view of their phylogeny
556 and ecophysiology. *Curr Opin Microbiol* 2011.
557 5. Lehtovirta-Morley LE. Ammonia oxidation: Ecology, physiology, biochemistry and why they
558 must all come together. *FEMS Microbiol Lett* 2018; **365**.
559 6. Santoro AE, Richter RA, Dupont CL. Planktonic Marine Archaea. *Ann Rev Mar Sci* 2019; **11**:
560 annurev-marine-121916-063141.
561 7. Francis CA, Roberts KJ, Beman JM, Santoro AE, Oakley BB. Ubiquity and diversity of
562 ammonia-oxidizing archaea in water columns and sediments of the ocean. *Proc Natl Acad Sci*
563 *U S A* 2005; **102**: 14683–8.
564 8. Durbin AM, Teske A. Archaea in Organic-Lean and Organic-Rich Marine Subsurface
565 Sediments: An Environmental Gradient Reflected in Distinct Phylogenetic Lineages. *Front*
566 *Microbiol* 2012; **3**.
567 9. Park SJ, Park BJ, Rhee SK. Comparative analysis of archaeal 16S rRNA and amoA genes to
568 estimate the abundance and diversity of ammonia-oxidizing archaea in marine sediments.
569 *Extremophiles* 2008; **12**: 605–615.
570 10. Vuillemin A, Wankel SD, Coskun ÖK, Magritsch T, Vargas S, Estes ER, et al. Archaea
571 dominate oxic subseafloor communities over multimillion-year time scales. *Sci Adv* 2019; **5**:
572 eaaw4108.
573 11. Nunoura T, Nishizawa M, Hirai M, Shimamura S, Harnvoravongchai P, Koide O, et al.
574 Microbial Diversity in Sediments from the Bottom of the Challenger Deep, the Mariana
575 Trench. *Microbes Environ* 2018; **33**: 186–194.
576 12. Zhao R, Hannisdal B, Mogollon JM, Jørgensen SL. Nitrifier abundance and diversity peak at
577 deep redox transition zones. *Sci Rep* 2019; **9**: 1–12.
578 13. Jørgensen SL, Zhao R. Microbial Inventory of Deeply Buried Oceanic Crust from a Young
579 Ridge Flank. *Front Microbiol* 2016; **7**.
580 14. Zhao R, Dahle H, Ramírez GA, Jørgensen SL. Indigenous Ammonia-Oxidizing Archaea in
581 Oxic Subseafloor Oceanic Crust. *mSystems* 2020; **5**.
582 15. Nunoura T, Nishizawa M, Kikuchi T, Tsubouchi T, Hirai M, Koide O, et al. Molecular
583 biological and isotopic biogeochemical prognoses of the nitrification-driven dynamic

584 microbial nitrogen cycle in hadopelagic sediments. *Environ Microbiol* 2013; **15**: 3087–3107.
585 16. Morono Y, Terada T, Nishizawa M, Ito M, Hillion F, Takahata N, et al. Carbon and nitrogen
586 assimilation in deep subseafloor microbial cells. *Proc Natl Acad Sci U S A* 2011; **108**: 18295–
587 300.
588 17. Wheat CG, Fisher AT, McManus J, Hulme SM, Orcutt BN. Cool seafloor hydrothermal
589 springs reveal global geochemical fluxes. *Earth Planet Sci Lett* 2017; **476**: 179–188.
590 18. Orsi WD. Ecology and evolution of seafloor and subseafloor microbial communities. *Nat Rev*
591 *Microbiol* 2018; **16**: 671–683.
592 19. D'Hondt S, Pockalny R, Fulfer VM, Spivack AJ. Subseafloor life and its biogeochemical
593 impacts. *Nat Commun* 2019; **10**: 3519.
594 20. Qin W, Martens-Habbena W, Kobelt JN, Stahl DA. Candidatus Nitrosopumilales . *Bergey's*
595 *Manual of Systematics of Archaea and Bacteria*. 2016.
596 21. Santoro AE, Dupont CL, Richter RA, Craig MT, Carini P, McIlvin MR, et al. Genomic and
597 proteomic characterization of 'Candidatus Nitrosopelagicus brevis': An ammonia-oxidizing
598 archaeon from the open ocean. *Proc Natl Acad Sci U S A* 2015; **112**: 1173–8.
599 22. Nunoura T, Takaki Y, Hirai M, Shimamura S, Makabe A, Koide O, et al. Hadal biosphere:
600 Insight into the microbial ecosystem in the deepest ocean on Earth. *Proc Natl Acad Sci U S A*
601 2015; **112**: E1230–E1236.
602 23. Wang Y, Huang J-M, Cui G-J, Nunoura T, Takaki Y, Li W-L, et al. Genomics insights into
603 ecotype formation of ammonia-oxidizing archaea in the deep ocean. *Environ Microbiol* 2019;
604 **21**: 716–729.
605 24. León-zayas R, Novotny M, Podell S, Shepard CM, Berkenpas E, Nikolenko S, et al. Single
606 Cells within the Puerto Rico Trench Suggest Hadal Adaptation of Microbial Lineages. *AEM*
607 2015; **81**: 8265–8276.
608 25. Sintes E, Bergauer K, De Corte D, Yokokawa T, Herndl GJ. Archaeal amoA gene diversity
609 points to distinct biogeography of ammonia-oxidizing Crenarchaeota in the ocean. *Environ*
610 *Microbiol* 2013; **15**: 1647–58.
611 26. Nunoura T, Hirai M, Yoshida-Takashima Y, Nishizawa M, Kawagucci S, Yokokawa T, et al.
612 Distribution and Niche Separation of Planktonic Microbial Communities in the Water
613 Columns from the Surface to the Hadal Waters of the Japan Trench under the Eutrophic
614 Ocean. *Front Microbiol* 2016; **7**.
615 27. Santoro AE, Saito MA, Goepfert TJ, Lamborg CH, Dupont CL, DiTullio GR. Thaumarchaeal
616 ecotype distributions across the equatorial Pacific Ocean and their potential roles in
617 nitrification and sinking flux attenuation. *Limnol Oceanogr* 2017; **62**: 1984–2003.
618 28. Alves RJE, Minh BQ, Urich T, von Haeseler A, Schleper C. Unifying the global phylogeny
619 and environmental distribution of ammonia-oxidising archaea based on amoA genes. *Nat*
620 *Commun* 2018; **9**: 1517.

29. Ren M, Feng X, Huang Y, Wang H, Hu Z, Clingenpeel S, et al. Phylogenomics suggests oxygen availability as a driving force in Thaumarchaeota evolution. *ISME J* 2019; 1.
30. Stolper DA, Bucholz CE. Neoproterozoic to early Phanerozoic rise in island arc redox state due to deep ocean oxygenation and increased marine sulfate levels. *Proc Natl Acad Sci* 2019; **116**: 8746–8755.
31. Konneke M, Schubert DM, Brown PC, Hugler M, Standfest S, Schwander T, et al. Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO2 fixation. *Proc Natl Acad Sci U S A* 2014; **111**: 8239–8244.
32. Molari M, Manini E, Dell'Anno A. Dark inorganic carbon fixation sustains the functioning of benthic deep-sea ecosystems. *Global Biogeochem Cycles* 2013; **27**: 212–221.
33. Dekas AE, Parada AE, Mayali X, Fuhrman JA, Wollard J, Weber PK, et al. Characterizing Chemoautotrophy and Heterotrophy in Marine Archaea and Bacteria With Single-Cell Multi-isotope NanoSIP. *Front Microbiol* 2019; **10**: 2682.
34. Kits KD, Sedlacek CJ, Lebedeva E V., Han P, Bulaev A, Pjevac P, et al. Kinetic analysis of a complete nitrifier reveals an oligotrophic lifestyle. *Nature* 2017; **549**: pages 269–272.
35. Eloe EA, Fadrosh DW, Novotny M, Zeigler Allen L, Kim M, Lombardo M-J, et al. Going Deeper: Metagenome of a Hadopelagic Microbial Community. *PLoS One* 2011; **6**: e20388.
36. Lloyd KG, Schreiber L, Petersen DG, Kjeldsen KU, Lever M a, Steen AD, et al. Predominant archaea in marine sediments degrade detrital proteins. *Nature* 2013; **496**: 215–8.
37. Li M, Baker BJ, Anantharaman K, Jain S, Breier JA, Dick GJ. Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. *Nat Commun* 2015; **6**: 8933.
38. Zhao R, Mogollón JM, Abby SS, Schleper C, Biddle JF, Roerdink D, et al. In situ growth of anammox bacteria in subseafloor sediments. *bioRxiv* 2019; 729350.
39. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30**: 2114–2120.
40. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 2015; **31**: 1674–1676.
41. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. *Nat Methods* 2014; **11**: 1144–1146.
42. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 2015; **3**: e1165.
43. Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 2016; **32**: 605–607.
44. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome*

Res 2015; **25**: 1043–1055.

45. Vallenet D, Calteau A, Cruveiller S, Gachet M, Lajus A, Josso A, et al. MicroScope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes. *Nucleic Acids Res* 2017; **45**: D517–D528.

46. Makarova KS, Galperin MY, Koonin E V. Comparative genomic analysis of evolutionarily conserved but functionally uncharacterized membrane proteins in archaea: Prediction of novel components of secretion, membrane remodeling and glycosylation systems. *Biochimie* 2015; 1–11.

47. Kristensen DM, Kannan L, Coleman MK, Wolf YI, Sorokin A, Koonin E V., et al. A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. *Bioinformatics* 2010; **26**: 1481–1487.

48. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 2012; **28**: 3150–3152.

49. Graham ED, Heidelberg JF, Tully BJ. Potential for primary productivity in a globally-distributed bacterial phototroph. *ISME J* 2018; **12**: 1861–1866.

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

51. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013; **30**: 772–780.

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

53. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* 2020; **37**: 1530–1534.

54. Hiraoka S, Hirai M, Matsui Y, Makabe A, Minegishi H, Tsuda M, et al. Microbial community and geochemical analyses of trans-trench sediments for understanding the roles of hadal environments. *ISME J* 2020; **14**: 740–756.

55. Durbin AM, Teske A. Sediment-associated microdiversity within the Marine Group I Crenarchaeota. *Environ Microbiol Rep* 2010; **2**: 693–703.

56. Zhang S, Song W, Wemheuer B, Reveillaud J, Webster N, Thomas T. Comparative Genomics Reveals Ecological and Evolutionary Insights into Sponge-Associated Thaumarchaeota. *mSystems* 2019; **4**: e00288-19.

57. Hug LA, Thomas BC, Sharon I, Brown CT, Sharma R, Hettich RL, et al. Critical biogeochemical functions in the subsurface are associated with bacteria from new phyla and little studied lineages. *Environ Microbiol* 2016; **18**: 159–173.

58. Ngugi DK, Blom J, Alam I, Rashid M, Ba-Alawi W, Zhang G, et al. Comparative genomics

reveals adaptations of a halotolerant thaumarchaeon in the interfaces of brine pools in the Red Sea. *ISME J* 2015; **9**: 396–411.

59. Dombrowski N, Teske AP, Baker BJ. Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nat Commun* 2018; **9**: 4999.
60. Tully BJ, Heidelberg JF. Potential Mechanisms for Microbial Energy Acquisition in Oxic Deep Sea Sediments. *Appl Environ Microbiol* 2016; AEM.01023-16.
61. Moeller FU, Webster NS, Herbold CW, Behnam F, Domman D, Albertsen M, et al. Characterization of a thaumarchaeal symbiont that drives incomplete nitrification in the tropical sponge Ianthella basta. *Environ Microbiol* 2019; 1462-2920.14732.
62. Jung M-Y, Islam MA, Gwak J-H, Kim J-G, Rhee S-K. Nitrosarchaeum koreense gen. nov., sp. nov., an aerobic and mesophilic, ammonia-oxidizing archaeon member of the phylum Thaumarchaeota isolated from agricultural soil. *Int J Syst Evol Microbiol* 2018; **68**: 3084–3095.
63. Herbold CW, Lehtovirta-Morley LE, Jung M-Y, Jehmlich N, Hausmann B, Han P, et al. Ammonia-oxidising archaea living at low pH: Insights from comparative genomics. *Environ Microbiol* 2017; **19**: 4939–4952.
64. Kerou M, Offre P, Valledor L, Abby SS, Melcher M, Nagler M, et al. Proteomics and comparative genomics of Nitrososphaera viennensis reveal the core genome and adaptations of archaeal ammonia oxidizers. *Proc Natl Acad Sci U S A* 2016; **113**: E7937–E7946.
65. Walker CB, de la Torre JR, Klotz MG, Urakawa H, Pinel N, Arp DJ, et al. Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc Natl Acad Sci U S A* 2010; **107**: 8818–8823.
66. Bayer B, Vojvoda J, Offre P, Alves RJ, Elisabeth NH, Garcia JA, et al. Physiological and genomic characterization of two novel marine thaumarchaeal strains indicates niche differentiation. *ISME J* 2016; **10**: 1051–1063.
67. Offre P, Kerou M, Spang A, Schleper C. Variability of the transporter gene complement in ammonia-oxidizing archaea. *Trends Microbiol* 2014; **22**: 665–675.
68. Spang A, Poehlein A, Offre P, Zumbragel S, Haider S, Rychlik N, et al. The genome of the ammonia-oxidizing Candidatus Nitrososphaera gargensis: insights into metabolic versatility and environmental adaptations. *Env Microbiol* 2012; **14**: 3122–3145.
69. Li J, Mara P, Schubotz F, Sylvan JB, Burgaud G, Klein F, et al. Recycling and metabolic flexibility dictate life in the lower oceanic crust. *Nature* 2020; **579**: 250–255.
70. Pejchal R, Ludwig ML. Cobalamin-Independent Methionine Synthase (MetE): A Face-to-Face Double Barrel That Evolved by Gene Duplication. *PLoS Biol* 2004; **3**: e31.
71. Jørgensen BB, Marshall IPG. Slow Microbial Life in the Seabed. *Ann Rev Mar Sci* 2016; **8**: 311–332.
72. Crable BR, Plugge CM, McInerney MJ, Stams AJM. Formate Formation and Formate

Conversion in Biological Fuels Production. *Enzyme Res* 2011; **2011**: 1–8.

73. Matelska D, Shabalin IG, Jabłońska J, Domagalski MJ, Kutner J, Ginalski K, et al. Classification, substrate specificity and structural features of D-2-hydroxyacid dehydrogenases: 2HADH knowledgebase. *BMC Evol Biol* 2018; **18**: 199.
74. Tishkov VI, Popov VO. Catalytic mechanism and application of formate dehydrogenase. *Biochem* 2004; **69**: 1252–1267.
75. Overkamp KM, Kötter P, van der Hoek R, Schoondermark-Stolk S, Luttik MAH, van Dijken JP, et al. Functional analysis of structural genes for NAD + -dependent formate dehydrogenase in Saccharomyces cerevisiae. *Yeast* 2002; **19**: 509–520.
76. Perret A, Lechaplais C, Tricot S, Perchat N, Vergne C, Pellé C, et al. A novel acyl-CoA beta-transaminase characterized from a metagenome. *PLoS One* 2011.
77. Desguin B, Soumillion P, Hausinger RP, Hols P. Unexpected complexity in the lactate racemization system of lactic acid bacteria. *FEMS Microbiol Rev* 2017; **41**: S71–S83.
78. Roche J, Girard E, Mas C, Madern D. The archaeal LDH-like malate dehydrogenase from Ignicoccus islandicus displays dual substrate recognition, hidden allostery and a non-canonical tetrameric oligomeric organization. *J Struct Biol* 2019; **208**: 7–17.
79. Greening C, Ahmed FH, Mohamed AE, Lee BM, Pandey G, Warden AC, et al. Physiology, Biochemistry, and Applications of F420 - and Fo -Dependent Redox Reactions. *Microbiol Mol Biol Rev* 2016; **80**: 451–493.
80. Ney B, Ahmed FH, Carere CR, Biswas A, Warden AC, Morales SE, et al. The methanogenic redox cofactor F 420 is widely synthesized by aerobic soil bacteria. *ISME J* 2017; **11**: 125–137.
81. Jørgensen BB, Boetius A. Feast and famine - Microbial life in the deep-sea bed. *Nat Rev Microbiol* 2007; **5**: 770–781.
82. Hoehler TM, Jørgensen BB. Microbial life under extreme energy limitation. *Nat Rev Microbiol* 2013; **11**: 83–94.
83. Orcutt BN, Sylvan JB, Knab NJ, Edwards KJ. Microbial Ecology of the Dark Ocean above, at, and below the Seafloor. *Microbiol Mol Biol Rev* 2011; **75**: 361–422.
84. Morono Y, Terada T, Nishizawa M, Ito M, Hillion F, Takahata N, et al. Carbon and nitrogen assimilation in deep subseafloor microbial cells. *Proc Natl Acad Sci U S A* 2011; **108**: 18295–18300.
85. Lever MA, Rogers KL, Lloyd KG, Overmann J, Schink B, Thauer RK, et al. Life under extreme energy limitation: a synthesis of laboratory- and field-based investigations. *FEMS Microbiol Rev* 2015; **39**: 688–728.
86. Vannier P, Michoud G, Oger P, Marteinsson VP, Jebbar M. Genome expression of Thermococcus barophilus and Thermococcus kodakarensis in response to different hydrostatic pressure conditions. *Res Microbiol* 2015; **166**: 717–725.

87. Ohke Y, Sakoda A, Kato C, Sambongi Y, Kawamoto J, Kurihara T, et al. Regulation of Cytochrome c - and Quinol Oxidases, and Piezotolerance of Their Activities in the Deep-Sea Piezophile Shewanella violacea DSS12 in Response to Growth Conditions. *Biosci Biotechnol Biochem* 2013; **77**: 1522–1528.

88. Campanaro S, De Pascale F, Telatin A, Schiavon R, Bartlett DH, Valle G. The transcriptional landscape of the deep-sea bacterium Photobacterium profundum in both a toxR mutant and its parental strain. *BMC Genomics* 2012; **13**: 567.

89. Wang B, Qin W, Ren Y, Zhou X, Jung M-Y, Han P, et al. Expansion of Thaumarchaeota habitat range is correlated with horizontal transfer of ATPase operons. *ISME J* 2019; 1–13.

90. Grüber G, Manimekalai MSS, Mayer F, Müller V. ATP synthases from archaea: The beauty of a molecular motor. *Biochim Biophys Acta - Bioenerg* 2014; **1837**: 940–952.

91. Abe F, Horikoshi K. Analysis of intracellular pH in the yeast Saccharomyces cerevisiae under elevated hydrostatic pressure: A study in baro- (piezo-) physiology. *Extremophiles* . 1998.

92. Bartlett DH. Pressure effects on in vivo microbial processes. *Biochim Biophys Acta - Protein Struct Mol Enzymol* 2002; **1595**: 367–381.

93. Siliakus MF, van der Oost J, Kengen SWM. Adaptations of archaeal and bacterial membranes to variations in temperature, pH and pressure. *Extremophiles* 2017; **21**: 651–670.

94. Severi E, Müller A, Potts JR, Leech A, Williamson D, Wilson KS, et al. Sialic Acid Mutarotation Is Catalyzed by the Escherichia coli β-Propeller Protein YjhT. *J Biol Chem* 2008; **283**: 4841–4849.

95. Ophir T, Gutnick DL. A Role for Exopolysaccharides in the Protection of Microorganisms from Desiccation. *Appl Environ Microbiol* 1994; **60**: 740–745.

96. Higgins MJ, Polcik M, Fukuma T, Sader JE, Nakayama Y, Jarvis SP. Structured Water Layers Adjacent to Biological Membranes. *Biophys J* 2006; **91**: 2532–2542.

97. Disalvo EA, Lairion F, Martini F, Tymczyszyn E, Frías M, Almaleck H, et al. Structural and functional properties of hydration and confined water in membrane interfaces. *Biochim Biophys Acta - Biomembr* 2008; **1778**: 2655–2670.

98. Disalvo A, de los Angeles Frias M. The Role of Water in the Responsive Properties in Lipid Interphase of Biomimetic Systems. *Liposomes - Advances and Perspectives*. 2019. IntechOpen.

99. Martinez N, Michoud G, Cario A, Ollivier J, Franzetti B, Jebbar M, et al. High protein flexibility and reduced hydration water dynamics are key pressure adaptive strategies in prokaryotes. *Sci Rep* 2016; **6**: 32816.

100. Muñoz-Marín MC, Gómez-Baena G, López-Lozano A, Moreno-Cabezuelo JA, Díez J, García-Fernández JM. Mixotrophy in marine picocyanobacteria: use of organic compounds by Prochlorococcus and Synechococcus. *ISME J* 2020; **14**: 1065–1073.

101. Michoud G, Jebbar M. High hydrostatic pressure adaptive strategies in an obligate piezophile

Pyrococcus yayanosii. *Sci Rep* 2016; **6**: 27289.

102. Cario A, Lormières F, Xiang X, Oger P. High hydrostatic pressure increases amino acid requirements in the piezo-hyperthermophilic archaeon Thermococcus barophilus. *Res Microbiol* 2015; **166**: 710–716.

103. Amrani A, Bergon A, Holota H, Tamburini C, Garel M, Ollivier B, et al. Transcriptomics Reveal Several Gene Expression Patterns in the Piezophile Desulfovibrio hydrothermalis in Response to Hydrostatic Pressure. *PLoS One* 2014; **9**: e106831.

104. Amrani A, van Helden J, Bergon A, Aouane A, Ben Hania W, Tamburini C, et al. Deciphering the adaptation strategies of Desulfovibrio piezophilus to hydrostatic pressure through metabolic and transcriptional analyses. *Environ Microbiol Rep* 2016; **8**: 520–526.

105. Empadinhas N, Da Costa MS. Osmoadaptation mechanisms in prokaryotes: Distribution of compatible solutes. *Int Microbiol* . 2008.

106. Mhatre SS, Kaufmann S, Marshall IPG, Obrochta S, Andrèn T, Jørgensen BB, et al. Microbial biomass turnover times and clues to cellular protein repair in energy-limited deep Baltic Sea sediments. *FEMS Microbiol Ecol* 2019.

107. Bradley JA, Amend JP, LaRowe DE. Survival of the fewest: Microbial dormancy and maintenance in marine sediments through deep time. *Geobiology* 2019; **17**: 43–59.

108. Rodrigues M V., Borges N, Henriques M, Lamosa P, Ventura R, Fernandes C, et al. Bifunctional CTP:Inositol-1-Phosphate Cytidylyltransferase/CDP-Inositol:Inositol-1-Phosphate Transferase, the Key Enzyme for Di-myo-Inositol-Phosphate Synthesis in Several (Hyper)thermophiles. *J Bacteriol* 2007; **189**: 5405–5412.

109. Gonçalves LG, Borges N, Serra F, Fernandes PL, Dopazo H, Santos H. Evolution of the biosynthesis of di-myo-inositol phosphate, a marker of adaptation to hot marine environments. *Environ Microbiol* 2012; **14**: 691–701.

110. Empadinhas N, da Costa MS. Diversity, biological roles and biosynthetic pathways for sugar-glycerate containing compatible solutes in bacteria and archaea. *Environ Microbiol* 2011; **13**: 2056–2077.

111. Ren M, Feng X, Huang Y, Wang H, Hu Z, Clingenpeel S, et al. Phylogenomics suggests oxygen availability as a driving force in Thaumarchaeota evolution. *ISME J* 2019; **13**: 2150–2161.

112. Biller SJ, Mosier AC, Wells GF, Francis CA. Global Biodiversity of Aquatic Ammonia-Oxidizing Archaea is Partitioned by Habitat. *Front Microbiol* 2012; **3**.

113. Li M, Baker BJ, Anantharaman K, Jain S, Breier JA, Dick GJ. Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. *Nat Commun* 2015; **6**: 8933.

114. Kazutaka Katoh DMS. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* 2013; **30**: 772.

115. Katoh K, Standley DM. MAFFT: iterative refinement and additional methods. *Methods Mol Biol* 2014; **1079**: 131–146.

116. Price MN, Dehal PS, Arkin AP. Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 2009.

873 **Figure and Table legends**

874

875 **Table 1. Statistics of deep marine sediments-derived MAGs.**

876 High-quality MAGs (> 80 % completeness and ~ 5% contamination or below) are marked with an
877 asterisk symbol.

878

879 **Fig. 1. Study sites and the community structure of AOA.**

880 **(a)** Global bathymetric map showing the coring locations of the sediment cores used in this study,
881 modified after [19]. The dark blue and light blue regions represent the minimum and maximum areas
882 over which dissolved O_2 is expected to penetrate throughout the sediment from seafloor to basement.
883 **(b)** AOA community structure based on the 16S rRNA gene phylogeny. *Nitrosopumilaceae* 16S rRNA
884 gene OTUs were classified based on their placements in the phylogenetic tree as described in [14]. In
885 the figure key, the corresponding clades of the AOA *amo*A gene are shown in red. They are based on
886 [55] and our phylogenetic analyses (unpublished data). The sediment horizons selected for metagenome
887 sequencing are highlighted by orange stars. Data for NP_U1383E and GS14-GC08 were retrieved from
888 [12].

889

890 **Figure 2 a) Phylogenomic tree of AOA and non-AOA Thaumarchaeota.**

891 The phylogenomic tree was reconstructed based on the concatenated alignment of 79 markers
892 comprising 7,485 amino acid sites using a maximum likelihood approach (see Materials and methods).
893 Yellow circles represent 100 % bootstrap support of nodes. The metagenome-assembled genomes
894 (MAGs) reported in this study are shown in bold. Representative genomes of NP-subclades were added
895 for subclade clarification. Information about the ecological distribution of AOA clades is provided. The
896 scale bar indicates the number of substitutions per amino acid site. In most cases congruence between
897 *amo*A-based and 16S rRNA-based clades was inferred from complete genomes or MAGs where both
898 genes were present [14, 22, 23]. For other cases: 1) *amo*A-NP-iota/16S-NP-lambda pair: the congruence
899 of these two clades is inferred from phylogenetic trees and congruence in relative abundances (Zhao
900 and Jorgensen, unpublished data) of sediment layers with high and changing relative abundances; 2)
901 *amo*A-NP-delta/16S-NP-epsilon,-zeta pair: The congruence of these two clades is based on the MAG
902 of CSP1 (with both genes), the close relationship of 16S-NP-epsilon with CSP1 and the occurrence
903 pattern of this clade in marine sediments (see [12], and data not shown); and 3) *amo*A-NP-theta/16S-
904 NP-alpha,-eta,-upsilon pair: congruence based on [14], own phylogenetic reconstructions (unpublished
905 data) and relative abundances in marine sediment layers. Abbreviations: NA; not assigned.
906 **b) Phylogeny of *amo*A sequences.** The ML phylogenetic tree was reconstructed using 584 nucleotide
907 sites. Colored circles represent MAGs from this study. The star symbol represents the
908 NPMR_NP_delta_1 bin. The incongruity of its phylogenetic placement based on the *amo*A gene and
909 the phylogenomic analysis of concatenated markers is discussed in the text. Greek letters represent the

910 *amo*A-based annotation of AOA subclades as in [28]. The scale bar indicates the number of
911 substitutions per nucleotide site.
912
913 **Figure 3. Comparative analysis of the presence/absence of protein clusters among AOA.**
914 Each bar (x axis) represents that the putative protein family is encoded in the genome (y axis).
915 10, 422 clusters with at least two different genomes are depicted. Non-AOA-specific clusters were
916 excluded from the visualization. Yellow circles represent > 95 % bootstrap support of nodes. Clusters
917 are ordered based on their distribution pattern from the most widespread to the most uncommon: first,
918 across all lineages and then intra-lineage. Groups of clusters have different color codes for better
919 visualization. NP represents Nitrosopumilales; NC, Nitrosocaldades; NS, Nitrososphaerales; NT,
920 Nitrosotaleales.
921
922 **Figure 4. Heatmap depicting the distribution and abundance of genes involved in the main**
923 **functional categories discussed in the text.** Abbreviations: nit2, nitrilase/omega-amidase; ureA,
924 urease subunit gamma; fdh, formate dehydrogenase; larA, lactate racemase; pgi, phosphoglucose
925 isomerase; proDH, proline dehydrogenase; rocA, 1-pyrroline-5-carboxylate dehydrogenase; oat,
926 putative ornithine--oxo-glutarate aminotransferase/class III aminotransferase; kal, 3-aminobutyryl-
927 CoA ammonia lyase; kat, putative 3-aminobutyryl-CoA aminotransferase; gvtTPH, glycine cleavage
928 system proteins T/P/H; metH, methionine synthase II (cobalamin-independent); metE, methionine
929 synthase I (cobalamin-dependent); APC, Amino Acid-Polyamine-Organocation Transporter Family;
930 HAAT, the Hydrophobic Amino Acid Uptake Transporter (HAAT) Family; uvrABC, the Uvr excision
931 repair system endonucleases ABC; udg4/5, Uracil DNA glycosylase family 4/5; mpg,
932 methylpurine/alkyladenine-DNA glycosylase; ogg1, 8-oxoguanine DNA glycosylase; alkA, DNA-3-
933 methyladenine glycosylase; tag, 3-methyladenine DNA glycosylase; pcm, protein-L-isoaspartate
934 carboxylmethyltransferase; nhaP, the Monovalent Cation:Proton Antiporter-1 (CPA1) Family; Trk, the
935 K+ Transporter (Trk) Family; ipct/dipps, bifunctional CTP:inositol-1-phosphate
936 cytidylyltransferase/di-myo-inositol-1,3′-phosphate-1′-phosphate synthase; cspC, cold-shock protein
937 A; cshA, cold-shock DEAD-box protein A; LLM, luciferase-like monooxygenase family protein;
938 nanM, N-acetylneuraminic acid mutarotase; flaK, archaeal preflagellin peptidase FlaK; cheY,
939 chemotaxis response regulator CheY; cheAB, chemotactic sensor histidine kinase cheA &
940 methylesterase cheB; All locus tags and cluster information are in Supplementary tables 3 & 4. An
941 extended version of the heatmap is in Fig S3
942
943 **Figure 5. Metabolic reconstruction of *amo*A-NP-theta, *amo*A-NP-delta and *amo*A-NP-iota AOA.**
944 Schematic reconstruction of the predicted metabolic modules in the sediment MAGs, as discussed in
945 the text. Color code of enzymes/complexes indicates conservation status in AOA. Unless specified by
946 greek letters (θδι), enzymes/modules are present in all sediment clades. Dashed lines indicate

hypothetical reactions. Gray arrow indicates an alternative OFOR reaction. Complexes of the electron transport chain are labelled with roman numerals. Transporters are named according to TCDB classification. Enzymes, gene accession numbers and transporter classes are also listed in Supplementary Table 3. Abbreviations: Amo, ammonia monooxygenase; NirK, nitrite reductase; Nit1, nitrilase; Nit2, nitrilase/omega-amidase; AA, amino acid; Fdh, formate dehydrogenase; Kal, 3-aminobutyryl-CoA ammonia lyase; Kat, putative 3-aminobutyryl-CoA aminotransferase; Lar, lactate racemase; Pcm, protein-L-isoaspartate carboxylmethyltransferase; MCP, methyl-accepting chemotaxis protein; Fla, archaellum; PolD, polymerase family D; PolY, translesion polymerase family Y; UVR, excision repair system; Hef-like, Hef/FANCM/Mph1-like helicase; BER, base-excision repair; Udg4/5, Uracil DNA glycosylase family 4/5; Mpg, methylpurine/alkyladenine-DNA glycosylase; Ogg1, 8-oxoguanine DNA glycosylase; AlkA, DNA-3-methyladenine glycosylase; CspA, cold-shock protein A; CshA, cold-shock DEAD-box protein A; Pcm, protein-L-isoaspartate carboxylmethyltransferase; PHB, polyhydroxybutyrate; MAE, malic enzyme; OFOR, 2-oxoacid:ferredoxin oxidoreductase; PGI, phosphoglucose isomerase; IPS, myo-inositol-1-phosphate synthase; ipct/dipps, bifunctional CTP:inositol-1-phosphate cytidylyltransferase/di-myo-inositol-1,3′-phosphate-1′-phosphate synthase; IMP, DIPP phosphatase; ProDH, proline dehydrogenase; RocA, 1-pyrroline-5-carboxylate dehydrogenase; glutamate dehydrogenase; oat, putative ornithine--oxo-glutarate aminotransferase/class III aminotransferase; aspA, aspartate ammonia-lyase; ilvA, threonine/serine ammonia-lyase; glyA, serine/glycine hydroxymethyltransferase; ilvE, branched-chain-amino-acid transaminase; aspC, aspartate/tyrosine/aromatic aminotransferase; MCO1, multicopper oxidase family 1. NanM, N-acetylneuraminic acid mutarotase; Transporters are named according to TCDB classification (supplementary table 4).

Table 1

Genome Bin	AOA clade	Depth (m)	Depth (m below sea floor)	Contigs (No.)	Genome size (bp)	Proteins (No.)	Longest contig (bp)	N50	GC (%)	Completeness (%)	Contamination (%)	Presence/ absence of 16S rRNA gene
NPMR_NP_delta_1*	NP-delta	4,425	0.1	83	1,173,997	1,467	79,876	28,788	34.00	86.23	5.42	-
NPMR_NP_delta_2	NP-delta	4,425	0.1	270	610,368	870	12,327	3,054	33.87	67.88	0.71	-
NPMR_NP_delta_3	NP-delta	4,425	0.1	143	937,620	1,222	58,541	15,484	34.22	66.29	3.4	-
NPMR_NP_theta_1*	NP-theta	4,425	1	176	1,524,500	2,000	40,454	14,847	35.06	99.51	2.91	-
NPMR_NP_theta_2*	NP-theta	4,425	22	298	1,201,755	1,671	26,499	5,123	34.98	93.69	4.05	-
YK1312_12N_NP_theta*	NP-theta	5,920	0.2	299	1,221,321	1,695	20,455	5,469	34.08	88.83	5.34	-
NPMR_NP_theta_3*	NP-theta	2,476	0.1	552	1,069,514	1,706	14,642	2,738	34.37	83.58	3.87	+
NPMR_NP_theta_4	NP-theta	2,476	0.1	232	1,040,901	1,408	21,324	6,332	33.98	76.24	6.31	-
NPMR_NP_theta_5	NP-theta	4,425	0.1	274	697,980	970	14,641	3,916	35.14	75.01	1.21	-
NPMR_NP_iota_1*	NP-iota	2,476	0.1	68	1,121,898	1,402	75,887	30,338	35.83	98.54	0.97	-
YK1309_1N_NP_iota*	NP-iota	4,277	0.3	227	1,298,549	1,700	40,725	9,447	35.76	96.6	3.24	-

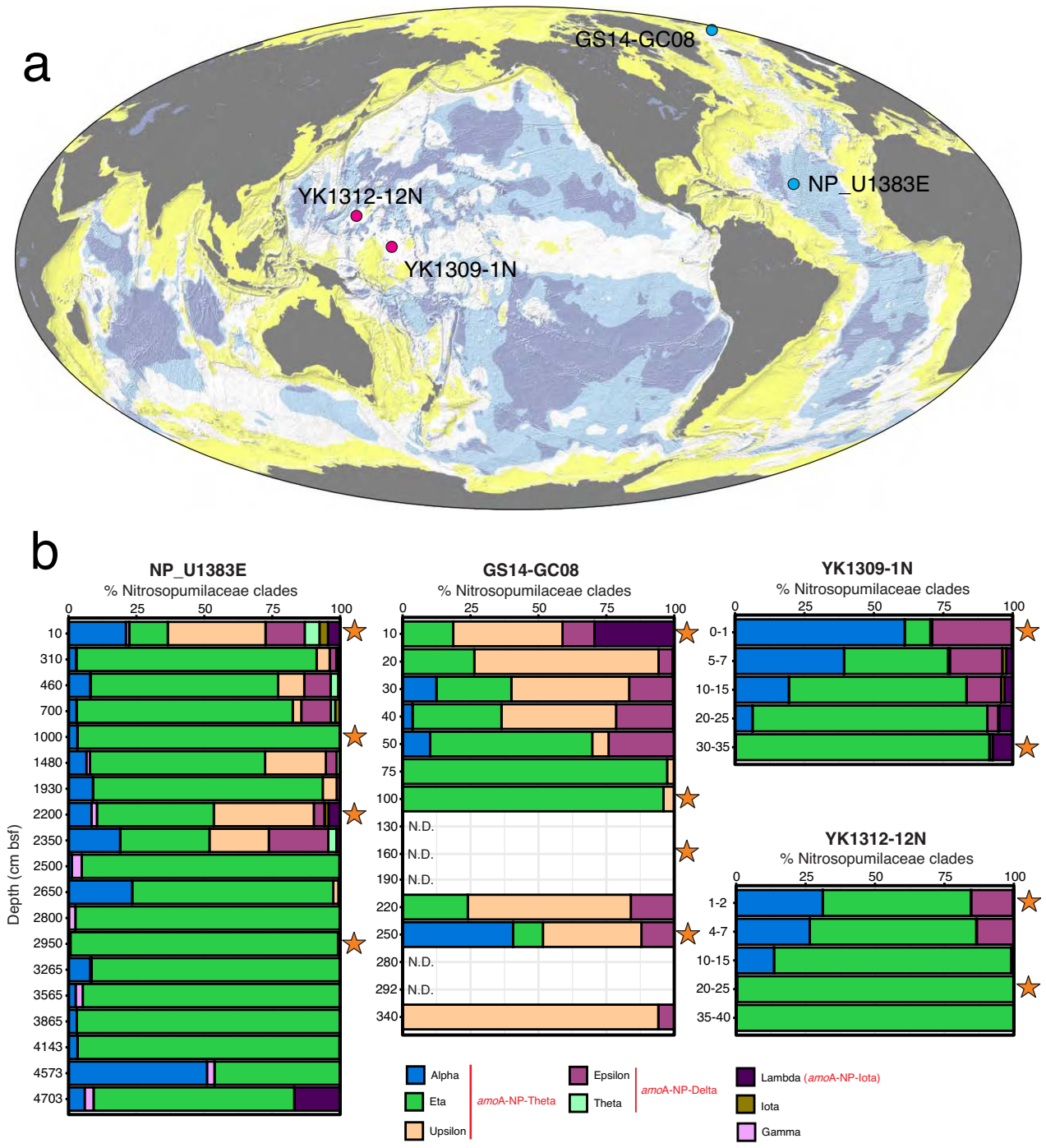
a
GS14-GC08
NP_U1383E
YK1312-12N
YK1309-1N
b
NP_U1383E
% Nitrosopumilaceae clades
GS14-GC08
% Nitrosopumilaceae clades
YK1309-1N
% Nitrosopumilaceae clades
YK1312-12N
% Nitrosopumilaceae clades
Depth (cm bsf)
N.D.
Alpha
Eta
Upsilon
amoA-NP-Theta
Epsilon
Theta
amoA-NP-Delta
Lambda (amoA-NP-Iota)
Iota
Gamma

Figure 1

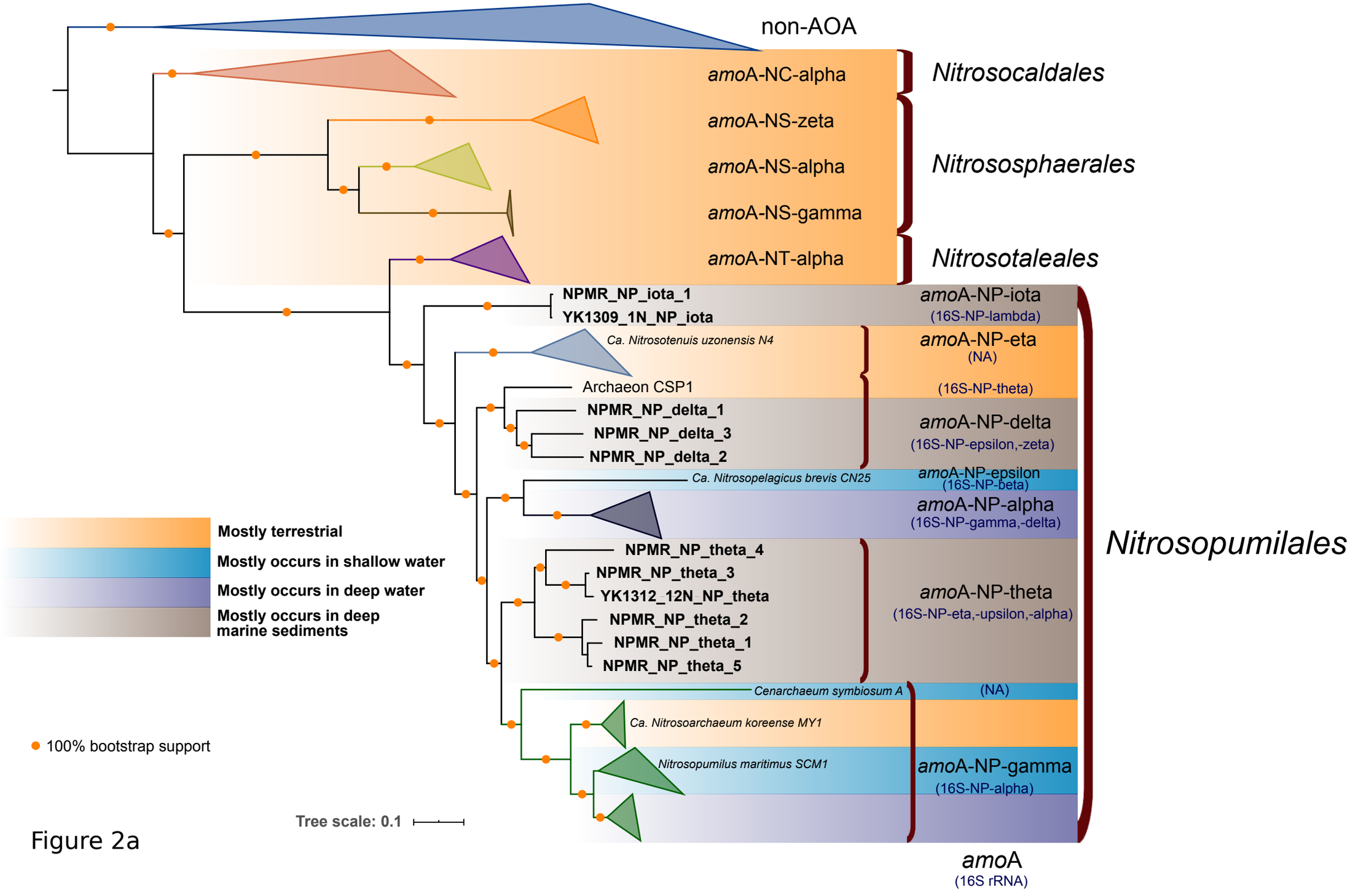
non-AOA
amoA-NC-alpha
amoA-NS-zeta
amoA-NS-alpha
amoA-NS-gamma
amoA-NT-alpha
Nitrosocaldales
Nitrososphaerales
Nitrosotaleales
NPMR_NP_iota_1
YK1309_1N_NP_iota
amoA-NP-iota
(16S-NP-lambda)
Ca. Nitrosotenuis uzonensis N4
amoA-NP-eta
(NA)
Archaeon CSP1
(16S-NP-theta)
NPMR_NP_delta_1
NPMR_NP_delta_3
NPMR_NP_delta_2
amoA-NP-delta
(16S-NP-epsilon,-zeta)
Ca. Nitrosopelagicus brevis CN25
amoA-NP-epsilon
(16S-NP-beta)
amoA-NP-alpha
(16S-NP-gamma,-delta)
Nitrosopumilales
NPMR_NP_theta_4
NPMR_NP_theta_3
YK1312_12N_NP_theta
NPMR_NP_theta_2
NPMR_NP_theta_1
NPMR_NP_theta_5
amoA-NP-theta
(16S-NP-eta,-upsilon,-alpha)
Cenarchaeum symbiosum A
(NA)
Ca. Nitrosoarchaeum koreense MY1
Nitrosopumilus maritimus SCM1
amoA-NP-gamma
(16S-NP-alpha)
Mostly terrestrial
Mostly occurs in shallow water
Mostly occurs in deep water
Mostly occurs in deep marine sediments
100% bootstrap support
Tree scale: 0.1
amoA
(16S rRNA)

Figure 2a

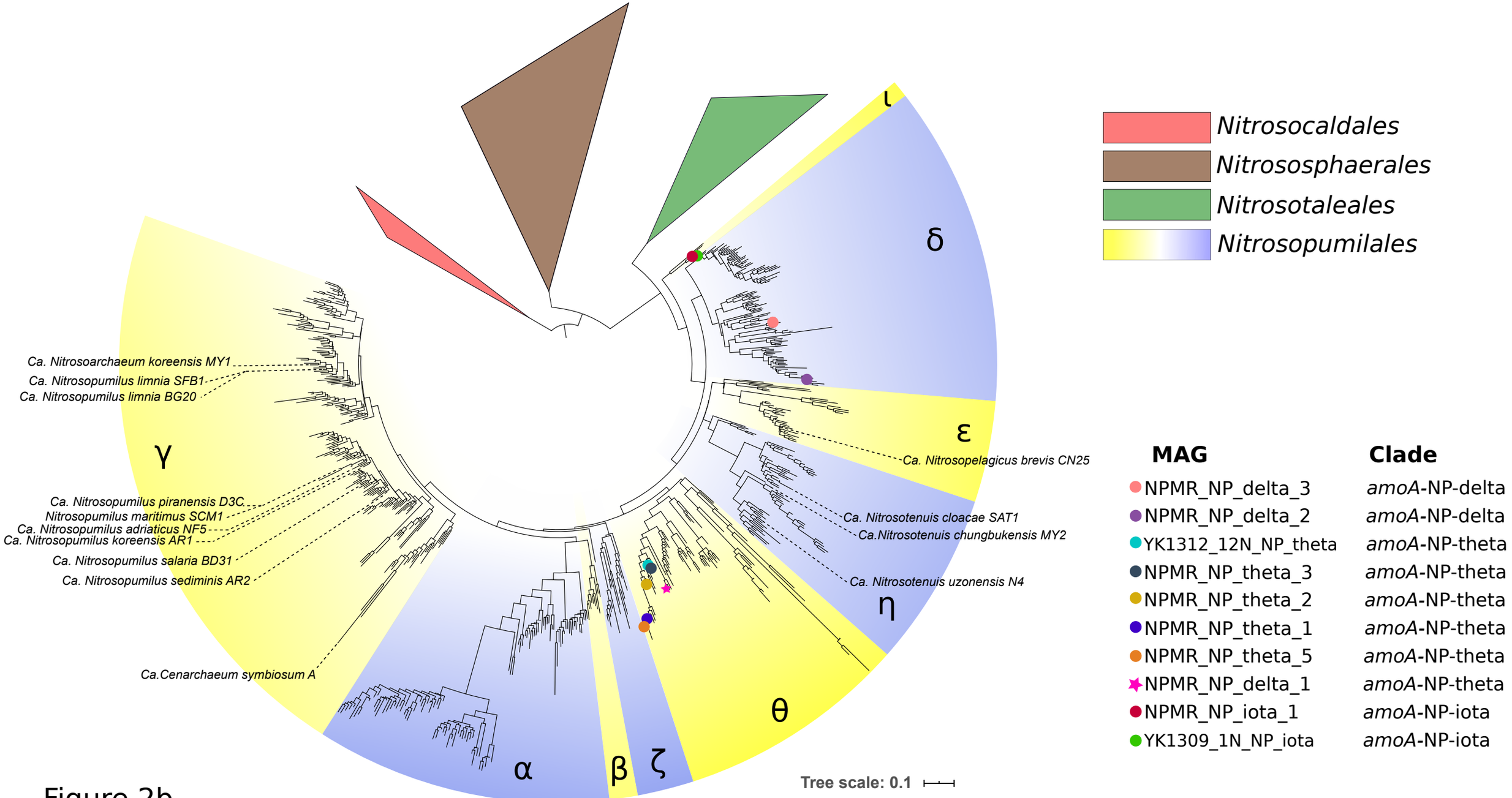
Nitrosocaldales
Nitrososphaerales
Nitrosotaleales
Nitrosopumilales
ι
δ
ε
γ
η
θ
α
β
ζ
Ca. Nitrosoarchaeum koreensis MY1
Ca. Nitrosopumilus limnia SFB1
Ca. Nitrosopumilus limnia BG20
Ca. Nitrosopumilus piranensis D3C
Nitrosopumilus maritimus SCM1
Ca. Nitrosopumilus adriaticus NF5
Ca. Nitrosopumilus koreensis AR1
Ca. Nitrosopumilus salaria BD31
Ca. Nitrosopumilus sediminis AR2
Ca.Cenarchaeum symbiosum A
Ca. Nitrosopelagicus brevis CN25
Ca. Nitrosotenuis cloacae SAT1
Ca.Nitrosotenuis chungbukensis MY2
Ca. Nitrosotenuis uzonensis N4
Tree scale: 0.1
MAG
Clade
NPMR_NP_delta_3
amoA-NP-delta
NPMR_NP_delta_2
amoA-NP-delta
YK1312_12N_NP_theta
amoA-NP-theta
NPMR_NP_theta_3
amoA-NP-theta
NPMR_NP_theta_2
amoA-NP-theta
NPMR_NP_theta_1
amoA-NP-theta
NPMR_NP_theta_5
amoA-NP-theta
NPMR_NP_delta_1
amoA-NP-theta
NPMR_NP_iota_1
amoA-NP-iota
YK1309_1N_NP_iota
amoA-NP-iota

Figure 2b

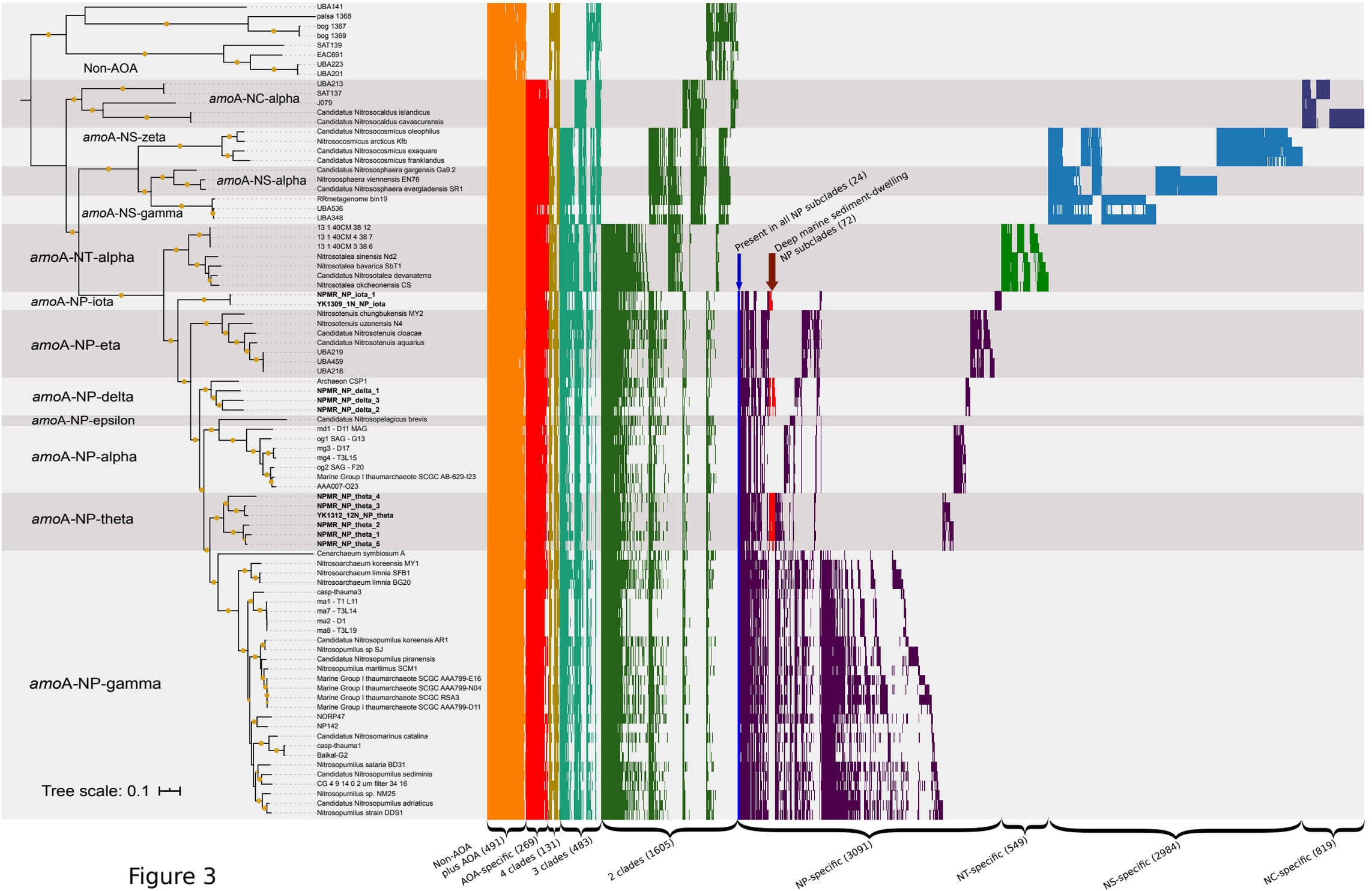
UBA141
palsa 1368
bog 1367
bog 1369
SAT139
EAC691
UBA223
UBA201
Non-AOA
UBA213
SAT137
J079
Candidatus Nitrosocaldus islandicus
Candidatus Nitrosocaldus cavascurensis
amoA-NC-alpha
Candidatus Nitrosocosmicus oleophilus
Nitrososcosmicus arcticus Kfb
Candidatus Nitrosocosmicus exaquare
Candidatus Nitrosocosmicus franklandus
amoA-NS-zeta
Candidatus Nitrososphaera gargensis Ga9.2
Nitrososphaera viennensis EN76
Candidatus Nitrososphaera evergladensis SR1
amoA-NS-alpha
RRmetagenome bin19
UBA536
UBA348
amoA-NS-gamma
13 1 40CM 38 12
13 1 40CM 4 38 7
13 1 40CM 3 38 6
Nitrosotalea sinensis Nd2
Nitrosotalea bavarica SbT1
Candidatus Nitrosotalea devanaterra
Nitrosotalea okcheonensis CS
amoA-NT-alpha
NPMR_NP_iota_1
YK1309_1N_NP_iota
amoA-NP-iota
Nitrosotenuis chungbukensis MY2
Nitrosotenuis uzonensis N4
Candidatus Nitrosotenuis cloacae
Candidatus Nitrosotenuis aquarius
UBA219
UBA459
UBA218
amoA-NP-eta
Archaeon CSP1
NPMR_NP_delta_1
NPMR_NP_delta_3
NPMR_NP_delta_2
amoA-NP-delta
Candidatus Nitrosopelagicus brevis
amoA-NP-epsilon
md1 - D11 MAG
og1 SAG - G13
mg3 - D17
mg4 - T3L15
og2 SAG - F20
Marine Group I thaumarchaeote SCGC AB-629-I23
AAA007-O23
amoA-NP-alpha
NPMR_NP_theta_4
NPMR_NP_theta_3
YK1312_12N_NP_theta
NPMR_NP_theta_2
NPMR_NP_theta_1
NPMR_NP_theta_5
amoA-NP-theta
Cenarchaeum symbiosum A
Nitrosoarchaeum koreensis MY1
Nitrosoarchaeum limnia SFB1
Nitrosoarchaeum limnia BG20
casp-thauma3
ma1 - T1 L11
ma7 - T3L14
ma2 - D1
ma8 - T3L19
Candidatus Nitrosopumilus koreensis AR1
Nitrosopumilus sp SJ
Candidatus Nitrosopumilus piranensis
Nitrosopumilus maritimus SCM1
Marine Group I thaumarchaeote SCGC AAA799-E16
Marine Group I thaumarchaeote SCGC AAA799-N04
Marine Group I thaumarchaeote SCGC RSA3
Marine Group I thaumarchaeote SCGC AAA799-D11
NORP47
NP142
Candidatus Nitrosomarinus catalina
casp-thauma1
Baikal-G2
Nitrosopumilus salaria BD31
Candidatus Nitrosopumilus sediminis
CG 4 9 14 0 2 um filter 34 16
Nitrosopumilus sp. NM25
Candidatus Nitrosopumilus adriaticus
Nitrosopumilus strain DDS1
amoA-NP-gamma
Tree scale: 0.1
Present in all NP subclades (24)
Deep marine sediment-dwelling NP subclades (72)
Non-AOA plus AOA (491)
AOA-specific (269)
4 clades (131)
3 clades (483)
2 clades (1605)
NP-specific (3091)
NT-specific (549)
NS-specific (2984)
NC-specific (819)

Figure 3

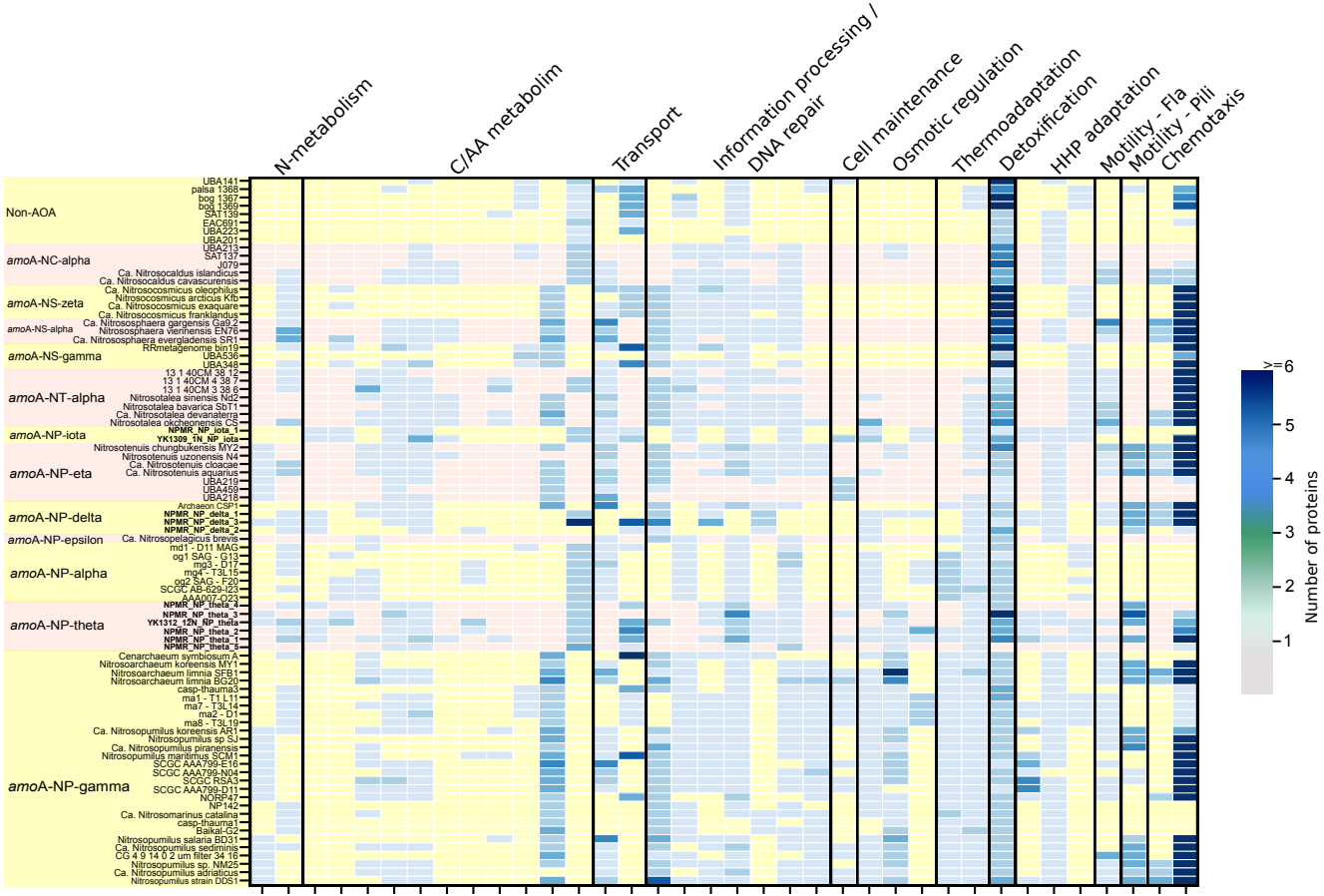
N-metabolism
C/AA metabolism
Transport
Information processing / DNA repair
Cell maintenance
Osmotic regulation
Thermoadaptation
Detoxification
HHP adaptation
Motility - Fla
Motility - Pili
Chemotaxis
Non-AOA
amoA-NC-alpha
amoA-NS-zeta
amoA-NS-alpha
amoA-NS-gamma
amoA-NT-alpha
amoA-NP-iota
amoA-NP-eta
amoA-NP-delta
amoA-NP-epsilon
amoA-NP-alpha
amoA-NP-theta
amoA-NP-gamma
UBA141
palsa 1368
bog 1367
bog 1369
SAT139
FAC891
UBA223
UBA201
UBA213
SAT137
J079
Ca. Nitrosocaldus islandicus
Ca. Nitrosocaldus cavascurensis
Ca. Nitrosocosmicus oleophilus
Nitrosocosmicus arcticus Kfb
Ca. Nitrosocosmicus exaquare
Ca. Nitrosocosmicus franklandus
Ca. Nitrososphaera gargensis Ga9.2
Nitrososphaera viennensis EN76
Ca. Nitrososphaera evergladensis SR1
RRmetagenome_bin19
UBA536
UBA348
13 1 40CM 38 12
13 1 40CM 4 38 7
13 1 40CM 3 38 6
Nitrosotalea sinensis Nd2
Nitrosotalea bavarica SbT1
Ca. Nitrosotalea devanaterra
Nitrosotalea okcheonensis CS
NPMR_NP_iota_1
YK1309_TN_NP_iota
Nitrosotenuis chungbukensis MY2
Nitrosotenuis uzonensis N4
Ca. Nitrosotenuis cloacae
Ca. Nitrosotenuis aquarius
UBA219
UBA458
UBA218
Archaeon CSP1
NPMR_NP_delta_1
NPMR_NP_delta_3
NPMR_NP_delta_2
Ca. Nitrosopelagicus brevis
md1 - D11 MAG
og1 SAG - G13
mg3 - D17
mg4 - T3L15
og2 SAG - F26
SCGC AB-629-I23
AAA007-G23
NPMR_NP_theta_4
NPMR_NP_theta_3
YK1312_TN_NP_theta
NPMR_NP_theta_2
NPMR_NP_theta_1
NPMR_NP_theta_5
Cenarchaeum symbiosum A
Nitrosoarchaeum koreensis MY1
Nitrosoarchaeum limnia SFB1
Nitrosoarchaeum limnia BG20
casp-thauma3
ma1 - T1L11
ma7 - T3L14
ma2 - D1
ma8 - T3L19
Ca. Nitrosopumilus koreensis AR1
Nitrosopumilus sp. SJ
Ca. Nitrosopumilus piranensis
Nitrosopumilus maritimus SCM1
SCGC AAA799-E16
SCGC AAA799-N04
SCGC RSA3
SCGC AAA799-D11
NORP47
NP142
Ca. Nitrosomarinus catalina
casp-thauma1
Baikal-G2
Nitrosopumilus salaria BD31
Ca. Nitrosopumilus sediminis
CG 4 9 14 0 2 um filter 34 16
Nitrosopumilus sp. NM25
Ca. Nitrosopumilus adriaticus
Nitrosopumilus strain DDS1
nit2
ureA
larA
fdh
pgi
putA/proDH
rocA
oat
kal
kat
gcvTPH
metH
metE
APC
HAAT
uvrABC
udg4
udg5
mpg
alkA
ogg1
tag
pcm
nhaP
Trk
ipct/dipps
cspC
cshA
LLM
nanM
A-type_ATPase
V-type_ATPase
flaK
flaI/TadC
cheAB
cheY
Number of proteins
>=6
5
4
3
2
1

Figure 4

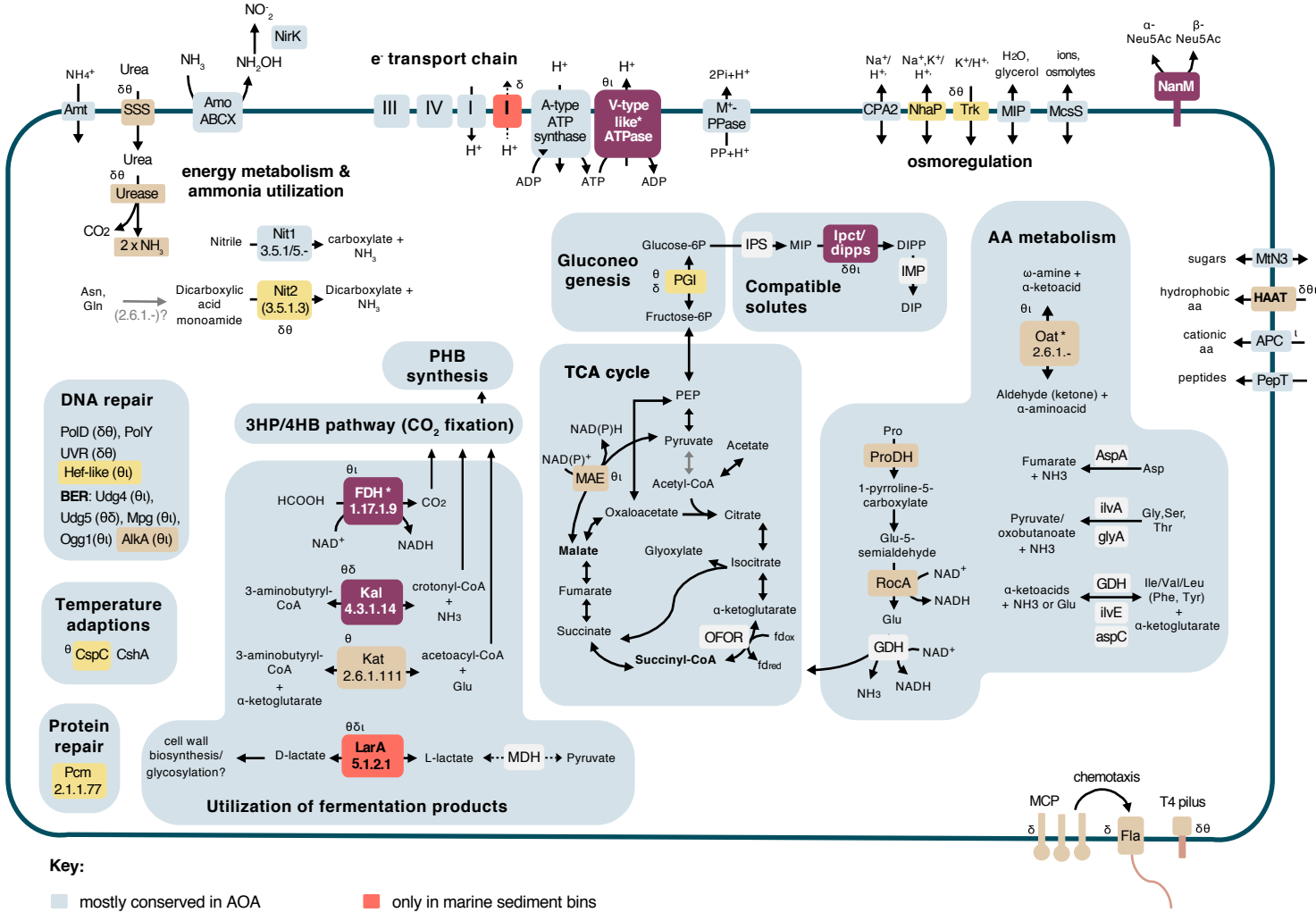
energy metabolism & ammonia utilization
e⁻ transport chain
osmoregulation
Gluconeogenesis
Compatible solutes
AA metabolism
PHB synthesis
3HP/4HB pathway (CO2 fixation)
TCA cycle
DNA repair
Temperature adaptions
Protein repair
Utilization of fermentation products
chemotaxis
Key:
mostly conserved in AOA
in other AOA incl. NS, but not largely conserved
present in NP/NT
only in marine sediment bins
shared only with deep sea NP ecotypes
* present in some NS

Figure 5