
















# Alternative strategies of nutrient acquisition and energy conservation map to the biogeography of marine ammonia-oxidizing archaea

Wei Qin <sup>1</sup> · Yue Zheng <sup>2,3</sup> · Feng Zhao <sup>2,3</sup> · Yulin Wang <sup>4</sup> · Hidetoshi Urakawa <sup>5</sup> · Willm Martens-Habbena<sup>6</sup> · Haodong Liu<sup>7</sup> · Xiaowu Huang<sup>8</sup> · Xinxu Zhang<sup>9</sup> · Tatsunori Nakagawa<sup>10</sup> · Daniel R. Mende<sup>11</sup> · Annette Bollmann<sup>12</sup> · Baozhan Wang <sup>13</sup> · Yao Zhang<sup>14</sup> · Shady A. Amin <sup>15</sup> · Jeppe L. Nielsen <sup>16</sup> · Koji Mori<sup>17</sup> · Reiji Takahashi<sup>10</sup> · E. Virginia Armbrust <sup>1</sup> · Mari-K.H. Winkler<sup>18</sup> · Edward F. DeLong <sup>11</sup> · Meng Li <sup>9</sup> · Po-Heng Lee<sup>19,8</sup> · Jizhong Zhou <sup>20,21,22</sup> · Chuanlun Zhang<sup>7</sup> · Tong Zhang<sup>4</sup> · David A. Stahl<sup>18</sup> · Anitra E. Ingalls <sup>1</sup>

Received: 15 February 2020 / Revised: 16 June 2020 / Accepted: 25 June 2020 / Published online: 7 July 2020  
© The Author(s) 2020. This article is published with open access

## Abstract

Ammonia-oxidizing archaea (AOA) are among the most abundant and ubiquitous microorganisms in the ocean, exerting primary control on nitrification and nitrogen oxides emission. Although united by a common physiology of chemoautotrophic growth on ammonia, a corresponding high genomic and habitat variability suggests tremendous adaptive capacity. Here, we compared 44 diverse AOA genomes, 37 from species cultivated from samples collected across diverse geographic locations and seven assembled from metagenomic sequences from the mesopelagic to hadopelagic zones of the deep ocean. Comparative analysis identified seven major marine AOA genotypic groups having gene content correlated with their distinctive biogeographies. Phosphorus and ammonia availabilities as well as hydrostatic pressure were identified as selective forces driving marine AOA genotypic and gene content variability in different oceanic regions. Notably, AOA methylphosphonate biosynthetic genes span diverse oceanic provinces, reinforcing their importance for methane production in the ocean. Together, our combined comparative physiological, genomic, and metagenomic analyses provide a comprehensive view of the biogeography of globally abundant AOA and their adaptive radiation into a vast range of marine and terrestrial habitats.

## Introduction

The ammonia-oxidizing archaea (AOA) comprise one of the most abundant and ubiquitous groups of microorganisms in the global biosphere. Studies of their

distribution and diversity based on sequence variation of single-copy core genes, such as those coding for the 16S rRNA and AmoA (the alpha subunit of ammonia monooxygenase), have shown they span phylum-level diversity and constitute up to 30% of microbial plankton in the oceans and up to 5% of microbial populations in soil [1–7]. Genome sequencing of uncultivated AOA also provided early environmental and genomic perspective. For example, the sponge symbiont *Candidatus* (*Ca.*) *Cenarchaeum symbiosum* was the first genome sequenced from AOA, yielding new perspective on the evolution, metabolic potential, and environmental distribution of this ubiquitous archaeal lineage [8]. Their habitat range encompasses the oceanic water column from its surface to hadal depths [3, 9], polar oceans [10], symbiotic systems [11], and terrestrial systems of extremes of pH, temperature, and elevation [12–15]. They are now recognized to exert the primary control on nitrification in oligotrophic environments [16–18]. In addition to having a dominant

---

These authors contributed equally: Wei Qin, Yue Zheng

**Supplementary information** The online version of this article (<https://doi.org/10.1038/s41396-020-0710-7>) contains supplementary material, which is available to authorized users.

- 
- ✉ Wei Qin  
qinwei2010@gmail.com
  - ✉ Feng Zhao  
fzhao@iue.ac.cn
  - ✉ Anitra E. Ingalls  
aingalls@uw.edu

Extended author information available on the last page of the article

role in the global nitrogen cycle, AOA make a significant contribution to carbon fixation through chemosynthetic pathways [19, 20], the production of greenhouse gases nitrous oxide and methane [21, 22], and the provision of the often limiting cofactor cobalamin (vitamin B<sub>12</sub>) in natural systems [23, 24].

Since the isolation of the first AOA species *Nitrosopumilus maritimus* in 2005 [25], over a decade of worldwide cultivation efforts have resulted in nearly 40 AOA strains enriched or isolated from natural and engineered ecosystems. These cultured species represent an environmentally and phylogenetically diverse group of archaeal ammonia oxidizers, spanning nearly 30% 16S rRNA gene divergence and covering four distinct phylogenetic lineages of the phylum *Thaumarchaeota* [26, 27]. Isolates now in culture provide a basis for investigations of the interrelationships between genotype, physiology, and ecology of this biogeochemically significant group, and for further characterizing the supporting biochemistry. For instance, the integrated genomic, physiological, and biochemical characterizations of the model AOA *Nitrosopumilus maritimus* revealed that the remarkable ecological success of AOA in marine environments is associated with their exceptionally high substrate affinity [17], energy-efficient carbon fixation [19], copper-centric respiratory system [28, 29], and unique cell envelop structure [26, 30–33]. In addition, comparative genomic analysis of three *Nitrososphaera* species and four *Ca. Nitrosotalea* species offered insights in the general adaptive strategies of soil AOA to neutral and low pH environments [34, 35].

In order to develop a better understanding of the adaptive radiation of this globally abundant group, we conducted a comprehensive analysis spanning a broad representation of marine, freshwater, soil, and geothermal AOA species. We framed our comparative analyses using genome sequences from isolates and enrichments from distinct habitats available in the literature and from newly generated genome sequences from our laboratories. Gene content of cultured species was then related to gene content and inferred metabolic traits derived from analysis of global ocean metagenomic databases across four oceans and two seas, spanning from epipelagic to hadopelagic zones. Robust association between variation in gene content among oceanic populations and environmental variables revealed the importance of ammonia and phosphorous availability, as well as hydrostatic pressure, in controlling the biogeography of different AOA genotypes. Niche boundaries were in part defined by the repertoire of genes for ammonia and phosphorous acquisition, and mechanisms of ATP generation. In addition, these analyses identified an extensive and variable reservoir of gene functions that appear to confer locally adaptive traits to marine and terrestrial AOA.

## Materials and methods

### Culture maintenance and genome sequencing

All AOA species were maintained in liquid mineral medium, and their genomic DNA was sequenced by Illumina or Pacbio platforms. For details, see Supplementary information.

### Sampling sites and metagenome-assembled genomes (MAGs)

AOA MAGs were recovered from wastewater treatment plants, geothermal environments, and hadopelagic waters. Metagenome binning and taxonomic assignments are described in Supplementary information.

### Comparative genomic analysis

All methods for genomic feature, phylogenomic, core genome, and pan-genome analyses of genomes from cultured AOA, MAGs, and single-cell amplified genomes (SAGs) are described in detail in Supplementary information.

### Evolution experiment and mutation analysis

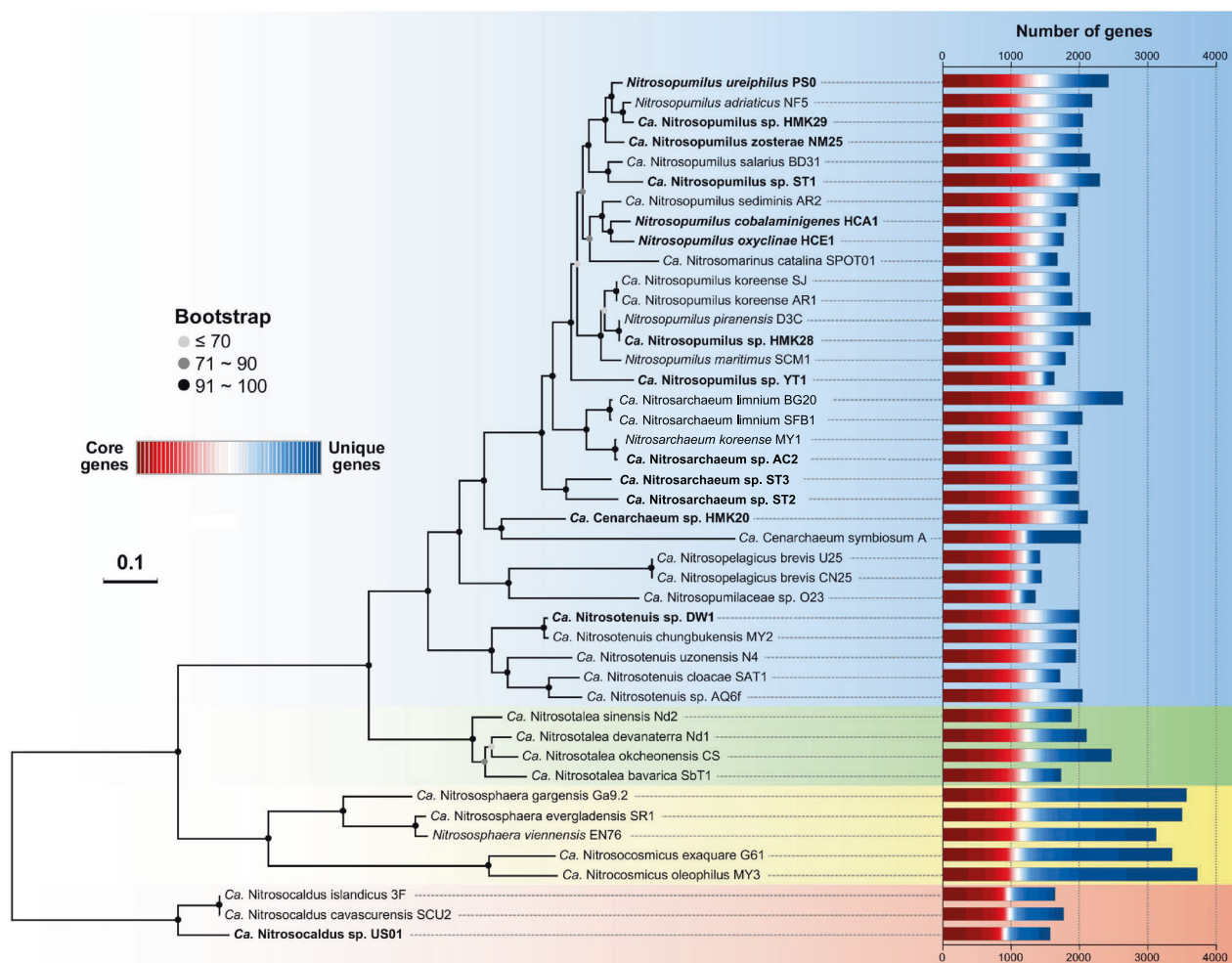
*Nitrosopumilus maritimus* strain SCM1 has been continuously transferred under optimum growth conditions from 2007 to 2018. Cultures were harvested in May 2011 and June 2016 for genome re-sequencing and subsequent mutation analysis. Further details of the methods are given in Supplementary information.

### Distribution and diversity of marine AOA genotypic groups and functional genes in the global ocean

Twenty-five marine AOA species genomes were clustered in seven genotypic groups to represent populations from distinct phylogenetic lineages and ecological habitats. To investigate the overall distribution of these genotypic groups in the global ocean, competitive fragment recruitment was conducted to determine the relative recruitment to available marine AOA species genomes in GOS and *Tara* Oceans metagenomic databases, as well as the metagenomic datasets of the HOT time-series station (125–4000 m), Northeast Pacific Ocean (2000 m), the Yap Trench (5000–5700 m), and the Mariana Trench waters (2000–8000 m). For details, see Supplementary information.

## Results and discussion

A total of 44 genomes sequenced from AOA cultured species or extracted from environmental metagenomes,



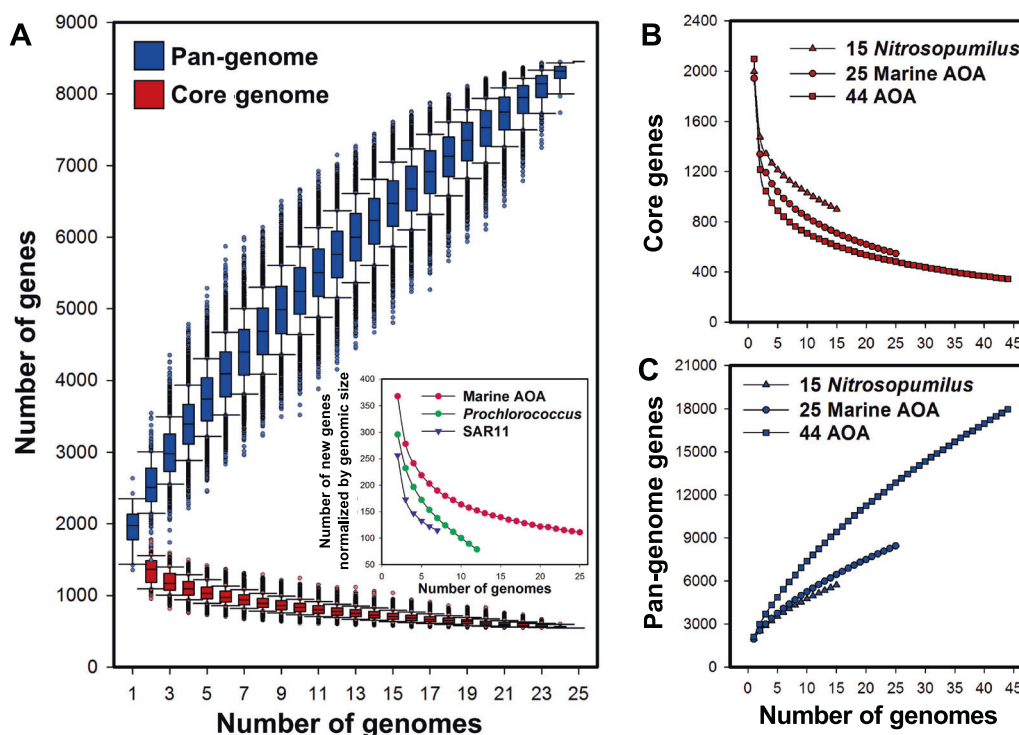
**Fig. 1** Phylogenomic inference of AOA species affiliated to the orders *Nitrosopumilales* (blue), *Ca. Nitrosotaleales* (green), *Nitrososphaerales* (yellow), and *Ca. Nitrosocaldales* (red) based on concatenated sequences of 71 single-copy core genes. The culture genomes and MAGs obtained in this study are highlighted in bold.

Confidence values are on the basis of 100 bootstrap replications. The scale bar represents 10% estimated sequence divergence. Total gene numbers are displayed as histogram plots for each AOA species genome with core genes shown in red and unique genes shown in blue.

representing all known ammonia-oxidizing thaumarchaeotal orders, were analyzed and compared in the present study (Table S1). In addition to the 30 genomes that have been previously reported, seven marine and two terrestrial AOA culture genomes as well as five MAGs from marine, geothermal, and engineered systems were obtained in this study (Table S1). These new genomes originate from diverse geographic locations and habitats, including fresh and coastal marine water columns and sediments, deep ocean trench waters, and hot springs (Fig. S1). To infer the most probable evolutionary relationship among AOA strains and refine the taxonomy of the unresolved thaumarchaeotal taxa, we constructed a maximum likelihood phylogenomic tree using a multiple sequence alignment of 71 single-copy core genes shared among all 44 AOA genomes (Fig. 1 and Table S2). While the topology of this species tree was

broadly congruent with those of the single-gene trees based on 16S rRNA and *amoA* genes, forming four basal order-level lineages within the phylum *Thaumarchaeota* [26, 27, 36, 37], the higher resolution of the phylogenomic analysis clearly differentiated closely related *Nitrosopumilus* strains (Fig. 1). The same speciation phylogeny of AOA was evident from the phylogenomic tree including 21 additional high-quality (completeness >90% and contamination <5%) AOA MAGs and SAGs [38–40] (Fig. S2).

The phylogenetic position of a long-branching taxon, *Ca. Cenarchaeum*, has remained largely ambiguous, because the extensive gene exchange between the only representative of the genus, *Ca. Cenarchaeum symbiosum* [8], and its sponge host confounds the inference of the evolutionary history of the lineage. However, the availability of a free-living *Ca. Cenarchaeum* strain, HMK20, has better defined the



**Fig. 2** AOA core genome and pan-genome analyses. **a** Box-and-whisker plots of core and pan-genome sizes of 25 marine AOA species. The number of genes is plotted as a function of the number of  $n$  strains sequentially added. For  $n$  genomes selected out of 25, there are  $25!/[(n-1)!(25-n)!]$  possible combinations from which to calculate core and pan-genomes. Each possible combination is plotted as a red and a blue dot for core genome and pan-genome analysis, respectively. To compare the pan-genome openness of marine AOA species with

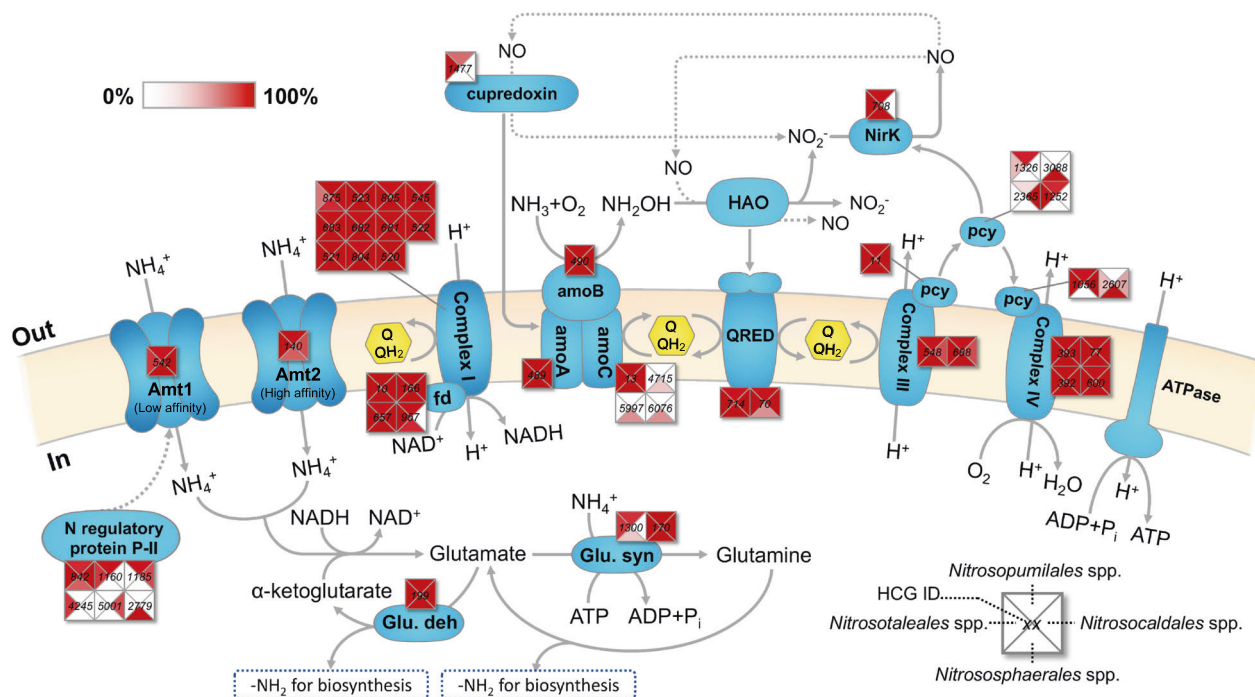
those of *Prochlorococcus* and SAR11 species, the average values of the number of new unique genes per Mb genome was calculated (inset panel) with the sequential addition of each marine AOA (red), *Prochlorococcus* (green), or SAR11 (purple) genomes. In addition to 25 marine AOA species, the average values of the number of core genes (**b**) and pan-genome genes (**c**) are shown for 15 *Nitrosopumilus* species (triangle) and all 44 members of AOA (square), respectively.

phylogenetic affiliation of this lineage. The *Ca.* Cenarchaeum genus appears to have diverged from the same lineage as the *Nitrosopumilus/Nitrosarchaeum* genera (Figs. 1 and S2). Similarly, the uncertain affiliation of *Ca.* Nitrosomarinus catalina SPOT01, a novel marine AOA strain enriched from California coastal waters [41], has been resolved. Although earlier proposed to represent the unique *Ca.* genus Nitrosomarinus, our phylogenomic analysis placed it well within the genus *Nitrosopumilus* (Figs. 1 and S2), and the average nucleotide identity (ANI) values between *Ca.* Nitrosomarinus and *Nitrosopumilus* genera (75–79%) were comparable to many ANI values within *Nitrosopumilus* genus (76–79%) (Fig. S3 and Table S3), suggesting that these two genera are phylogenetically indistinguishable. Given the genus *Nitrosopumilus* was described first [26], we suggest that the *Ca.* genus Nitrosomarinus is a later heterotypic synonym of the genus *Nitrosopumilus*. In addition, two previously defined distinct thermophilic AOA species, *Ca.* Nitrosocaldus islandicus 3F and *Ca.* Nitrosocaldus cavascurensis SCU2 [42, 43], share more than 99.8% similarity across their genomes (Fig. S3 and Table S3). Thus, this pair of genomes only sample

subspecies level diversity, and the denomination should be refined. An unexpected finding was that the MAGs recovered from the hadopelagic waters (>6000 m) of the oceanic trenches (YT1, F8–1, and F8–2) were closely grouped with *Nitrosopumilus* genus and widely separated from other oceanic AOA taxa that comprise *Ca.* Nitrosopelagicus strains and uncultured water column B (WCB)-AOA populations (O23) (Figs. 1 and S2).

Comparative analysis revealed extensive gene content and genome size variation among marine and terrestrial AOA genomes. For example, the gene content and genome sizes of *Ca.* Nitrosocosmicus-like soil AOA species (3395–3758 genes; 2.99–3.43 Mbp) are nearly threefold greater than those of *Ca.* Nitrosopelagicus (WCA) and WCB-like oceanic AOA species (1400–1502 genes; 1.17–1.25 Mbp) (Fig. S4 and Table S1). In addition to their distinct genome sizes, the genome coding density of AOA species varied widely, ranging from 73.7 to 93.9% (Fig. S4). The genome coding density of all AOA species together showed a clear linear decrease with increasing genome size (Fig. S4). Consistent with other oligotrophic marine bacterial genomes, such as SAR11,





**Fig. 3 Reconstruction of the proposed pathways of ammonia oxidation, electron transfer, and ammonia assimilation in AOA species, emphasizing the conservation and uniqueness of pathway enzymes of AOA species.** Alternative archaeal ammonia oxidation models are illustrated based on the current literature on pathway gene identification [29], identification of pathway intermediates [16, 56, 79], isotopic measurements [21, 56], and transcriptional regulation [28, 52]. Squares indicate the COGs of proteins in AOA species genomes, and COG numbers are inside squares. Each square is split into four pieces to

represent AOA species affiliated to the orders *Nitrosopumilales*, *Nitrososphaerales*, *Ca. Nitrosotaleales*, and *Ca. Nitrosocaldales*. Color scale reflects the percent of genomes in each ammonia-oxidizing thaumarchaeotal order with that COG. QRED quinone reductase, HAO putative hydroxylamine oxidoreductase, pcy plastocyanin, fd ferredoxins, Amt1 and Amt2 low-affinity and high-affinity ammonia transporter, respectively, Glu. deh and Glu. syn glutamate dehydrogenase and glutamine synthetase, respectively.

*Prochlorococcus*, and SUP05 (Table S1), the genomes of oceanic AOA species are highly streamlined and relatively gene dense (Fig. S4). In contrast, neutral pH soil AOA species that occupy nutrient-enriched environments tend to have larger genomes, and their genome coding densities are even lower than those of the eutrophic ammonia-oxidizing bacteria species (Fig. S4 and Table S1). The small marine AOA genomes are also associated with generally lower GC content than soil AOA (Fig. S3 and Table S1). Together, these findings indicate that nutrient levels and habitat types both have profound effects on the gene content and organization of AOA genomes.

### The core and pan-genome of the ammonia-oxidizing *Thaumarchaeota*

To provide further quantitative insights into the conserved and flexible gene pools within the ammonia-oxidizing *Thaumarchaeota*, we calculated the core genome that is shared by all AOA species and the pan-genome that represents the global gene repertoire of AOA (Fig. 2). The core genome of ammonia-oxidizing *Thaumarchaeota* comprises

~344 genes (Fig. 2b), including key pathway genes that are involved in the characterized central metabolism of AOA, such as carbon fixation through the 3-hydroxypropionate/4-hydroxybutyrate cycle and cobalamin biosynthesis [19, 23, 24, 38] (Figs. S5 and S6; Table S4). These core gene sets only represent a small fraction of the genomes of AOA species, accounting for ~14–23% of the genes in *Nitrosopumilales* genomes (62–99% ANI), ~9–12% in *Nitrososphaerales* genomes (64–86% ANI), ~14–19% in *Ca. Nitrosotaleales* genomes (78–83% ANI), and ~19–22% in *Ca. Nitrosocaldales* genomes (78% ANI). These low core gene set fractions suggest a large proportion of AOA genomes could be associated with environmental-specific functions that may provide fitness advantages in distinct marine and terrestrial habitats.

The ammonia-oxidizing thaumarchaeotal pan-genome within the available dataset contains a total of ~17,961 genes, and the comparative analysis of 44 AOA species genomes revealed that sampling of their pan-genome is far from saturated (Fig. 2c). The AOA pan-genome possesses a high proportion of putative and hypothetical genes with unknown functions (Fig. S7). The accessory and unique

genes that are assigned to the Clusters of Orthologous Groups (COGs) functional categories amino acid transport and metabolism, transcription, and energy production and conversion are among the most abundant genes that contribute to the global pool of AOA genes (Fig. S7). When considering only the 25 marine AOA species genomes with an ANI range of ~62–99%, the pan-genome consists of nearly 8500 genes (Fig. 2a). An average of ~111 novel unique genes per Mb are expected to be identified with each new marine AOA strain sequenced (Fig. 2a). Open pan-genomes with large genetic repertoires have also been observed for the other two most abundant marine microbial groups, SAR11 and *Prochlorococcus* [44, 45]. Notably, the number of new genes per Mb marine AOA species genome is even greater than those calculated for SAR11 and *Prochlorococcus* species, although they have a comparable degree of genome divergence (~64–97% of ANI) (Fig. 2a and Table S3). These findings further highlight the extensive genomic diversity among the globally abundant archaeal ammonia oxidizers.

### Functional analysis of core and flexible genes in ammonia oxidation and assimilation pathways

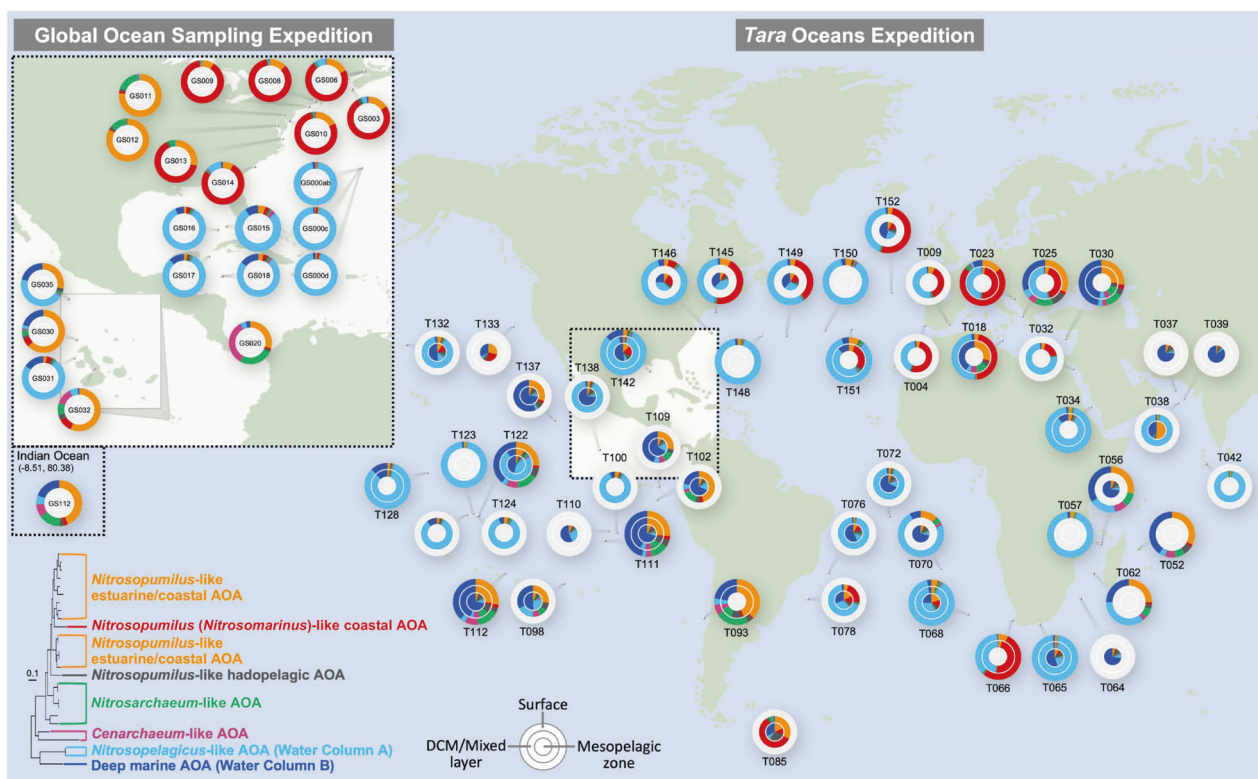
Despite their extraordinary genomic divergence, all AOA species share a common pathway for energy generation by oxidation of ammonia. We found that many essential components of ammonia oxidation, electron transfer, and ammonia assimilation pathways are mostly conserved across all marine and terrestrial AOA genomes, including A and B subunits of ammonia monooxygenase (AMO), quinone reductase, the conventional complex III and IV, ferredoxin and complex I (NADH:ubiquinone oxidoreductase), glutamine synthetase, and glutamate dehydrogenase (Fig. 3 and Table S4). In contrast, other genes inferred to have functions in regulation, stress responses, and ammonia uptake are highly variable among different orders of AOA species (Fig. 3 and Table S4). For instance, only *Nitrososphaerales* species possess extra copies of AmoC homologs hypothesized to provide a chaperone-like function to maintain the structural integrity of AMO holoenzyme under nutrient-depleted conditions [28] (Fig. 3). Given the generally low affinity of *Nitrososphaerales* species for ammonia [46], they may frequently face energy limitation or starvation. Thus, this adaptive genomic feature may allow *Nitrososphaerales*-like AOA to cope with intermittent energy stress in soils. Within the order *Nitrososphaerales*, *Ca. Nitrosocosmicus* species were isolated from the high nutrient environments of fertilized soils and wastewater treatment systems [47–49]. Intriguingly, *Ca. Nitrosocosmicus* species lack the high-affinity ammonium transporter and S-layer proteins associated with the high ammonia affinity of marine AOA species

(Table S4) [28, 32, 50]. The absence of these genes suggests they are adapted to much higher ammonia concentrations than is typical of most natural environments.

Since ammonia is used both as an energy source in ammonia oxidation and a nitrogen source in biosynthesis, sophisticated regulation of these two processes must maintain the metabolic and anabolic balance in AOA cells during adaptation to different redox and nutrient conditions. We found all AOA species encode an extensive gene repertoire for plastocyanin-like proteins that appear integral to respiration and electron transfer reactions associated with ammonia oxidation, as well as PII proteins that are predicted to be involved in regulation of ammonia assimilation (Fig. 3 and Table S4). High expression of many of these genes has been observed in both transcriptomes and proteomes of marine and terrestrial AOA species [28, 34, 51–54], suggesting they serve essential roles in respiratory and biosynthetic activities in AOA. We found large variations in the gene sets encoding plastocyanin proteins and PII proteins among different orders of AOA species (Fig. 3 and Table S4). Thus, these proteins are potentially under strong diversifying selection throughout the evolution of *Thaumarchaeota*, and genomic variation in systems of ammonia oxidation, electron transfer, and nitrogen assimilation likely reflect ecophysiological differences that determine habitat preference.

Nitric oxide (NO) is a central intermediate in the archaeal ammonia oxidation pathway, and the interaction between NO and cobalamin plays an important role in shaping the general ecophysiology of AOA species under stressed conditions [16, 55, 56]. One candidate enzyme for the production of NO in AOA is the putative copper-containing nitrite reductase (NirK). NirK genes are highly expressed during exponential growth of marine and soil AOA cells [28, 34, 51, 52], and the expression of *nirK* genes is tightly regulated by ammonia availability, suggesting NirK serves a key role in AOA ammonia catabolism [28]. However, although *nirK* genes are widely distributed in marine and soil AOA, no NirK homolog has yet been identified in any *Ca. Nitrosocaldales* species from geothermal systems (Table S4) [42, 43]. Since the ammonia oxidation by *Ca. Nitrosocaldales* species is highly sensitive to low concentrations of the NO-scavenging chemical PTIO (2-phenyl-4,4,5,5-tetramethylimidazoline-3-oxide-1-oxyl) [42, 43], NO appears to be an obligate intermediate or reactant in ammonia oxidation by all AOA lineages. Thus, the role of NirK in the pathway of archaeal ammonia oxidation is unresolved.

Apart from the conservation of *nirK* homologs in all known mesophilic AOA, its functional importance was also indicated by the emergence of a *nirK* variant containing a nonsynonymous mutation after extended laboratory cultivation of *N. maritimus*. Since the isolation of *N. maritimus*



**Fig. 4** The global distribution of major marine AOA genotypic groups across the upper oceans. Twenty-five marine AOA species were clustered in seven separate genotypic groups based on their positions in the phylogenomic tree and their distinct geographic origins (Fig. S10). *Tara Oceans* and *GOS* (inset panel) metagenomic reads were recruited to a genomic database containing seven marine AOA genotypic groups to assess the distribution of marine AOA genotypes spanning multiple depth layers of the upper ocean. Only the top hits with a maximum  $E$ -value of  $1e^{-10}$  and a minimum amino acid

identity to reference species genomes of 80% were retained for competitive fragment recruitment analysis. The fraction of metagenomic reads recruited to marine AOA genotypes from the surface water layer, deep chlorophyll maximum (DCM)/mixed layer, and mesopelagic zone are depicted on the outermost ring, middle ring, and inner ring at each *GOS* and *Tara Oceans* sampling station, respectively. The concentric rings are colored as light gray for the station depth layers without metagenomic data collection.

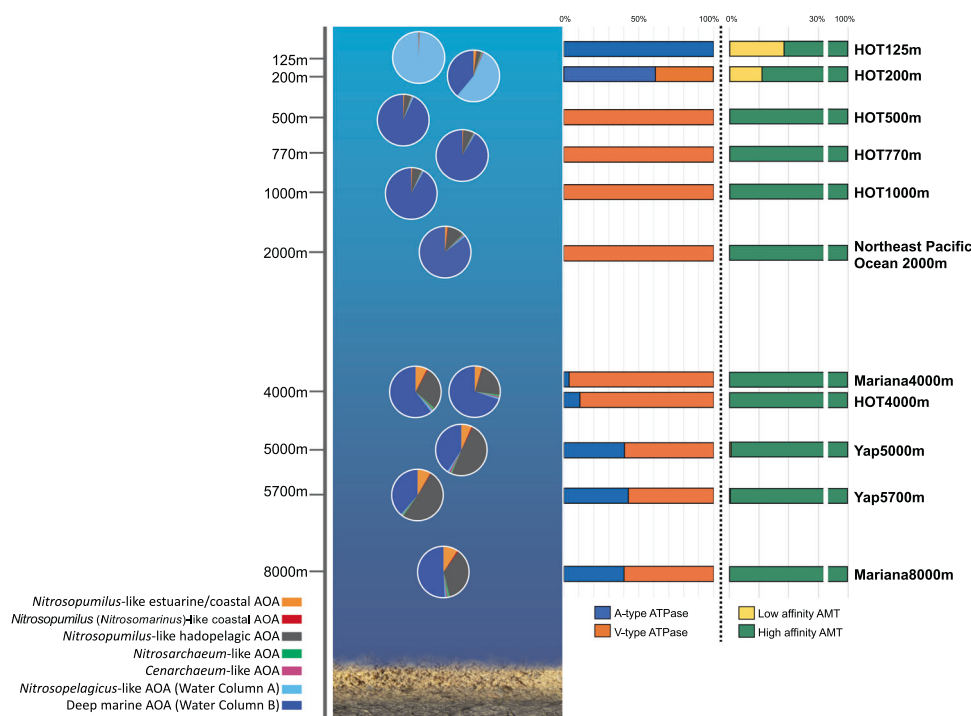
strain SCM1 in 2005, this strain has been continuously transferred under optimum growth conditions over 11 years, representing more than 3000 generations (Fig. S8) [17, 25, 26, 57]. Related to extended laboratory passage, the average generation time of SCM1 culture has decreased from ~2 days in 2007 to ~1 day in 2011 and stabilized at around 1 day by 2018 (Fig. S8). To identify mutations associated with the increased growth rate, we compared the genome sequence of the ancestral SCM1 cultures collected in 2007 with those of the evolved cultures collected in 2011 and 2016 after ~1200 and ~3000 generations of evolution, respectively (Fig. S8). In total, we observed 11 genes that harbored nonsynonymous mutations and seven genes with synonymous mutations across ~3000 generations (Table S5). This leads to a mutation rate estimate of marine AOA in the laboratory of  $4.3 \times 10^{-9}$  per site per generation, which is higher than that reported for *Prochlorococcus* ( $\sim 10^{-10}$  per site per generation) and at the higher end of previous estimates for other bacteria ( $0.8 \times 10^{-10}$ – $9.8 \times 10^{-9}$  per site per generation for 26 species)

[58, 59]. Three of the 11 genes with nonsynonymous mutations are in the ammonia oxidation and assimilation pathways, including genes encoding putative NirK (Nmar\_1667), a low-affinity ammonia transporter (AMT) (Amt1; Nmar\_0588) that mutated within ~1200 generations, and glutamate dehydrogenase (GDH) gene that mutated sometime during ~1800 additional generations of laboratory cultivation (Table S5). All mutations in genes for ammonia oxidation and assimilation are close to fixation, present at >96% frequency, and therefore likely conferred a fitness advantage to *N. maritimus* growing under laboratory culture conditions of relatively high ammonia concentration (Table S5).

Several observations are consistent with a significant functional impact of the mutation. The glycine to serine mutation we observed in NirK is near the copper-binding site of this enzyme (Table S5), and microsensor measurements of NO concentrations showed a higher initial accumulation rate of NO in cultures with the NirK mutation than earlier observed in cultures of lower laboratory passage



**Fig. 5 Depth distribution of marine AOA genotypic groups (pie chart) as well as associated ATPase and ammonia transporter (AMT) types (bar chart) from epipelagic to hadopelagic waters.** Metagenomic reads from the North Pacific HOT time-series station (125–4000 m; HOT cruise 229), Northeast Pacific Ocean (2000 m), the Yap Trench of the western Pacific (5000–5700 m), and the Mariana Trench waters (4000–8000 m) were recruited to a genomic database of major marine AOA genotypic groups to assess the distribution of marine AOA genotypes as well as associated ATPase and AMT types along the whole water column. Genotype designations and the corresponding color schemes are the same as those shown in Fig. 4.



(Fig. S9). This is suggestive of a direct involvement of NirK in NO metabolism. If the current model for archaeal ammonia oxidation is correct [28, 56, 60], a higher rate of enzymatic NO production might alleviate a kinetic limitation imposed by NO availability in the oxidation of ammonia to hydroxylamine, leading to the significantly higher specific growth rates that we observed in evolved cultures (Fig. S8). Similarly, the mutation in the AMT could reflect altered function associated with continuous growth under relatively high concentrations of ammonia not typical of environmental availability.

### Global distribution of the major genotypes of marine AOA

The comparative analysis of 25 marine AOA species genomes allowed us to define seven major genotypic groups (potential functional guilds) of marine AOA that represent the dominant populations in estuarine and coastal areas, surface and deep open oceans, and hadopelagic waters at depths below 6000 m (Fig. S10). In order to analyze the distribution patterns of these genotypes spanning the global upper oceans, we used competitive fragment recruitment to estimate the relative recruitment to the genomes of marine AOA genotypes in global ocean metagenome datasets, including the Global Ocean Sampling (GOS) expedition data and *Tara* Oceans Global expedition data (2009–2013). The GOS expedition contains metagenome datasets recovered from surface water samples, and we found that the

dominant marine AOA population shifted along the inshore–offshore gradient (Fig. 4). *Nitrosopumilus*-like AOA were most abundant in estuarine areas, whereas they were of lower abundance in further offshore waters (Fig. 4). *Nitrosopumilus* (*Ca.* *Nitrosomarinus*)-like AOA were the dominant genotype in coastal areas, while *Ca.* *Nitrosopelagicus*-like AOA represented the major population in open ocean surface waters (Fig. 4). *Tara* Oceans expedition conducted a more extensive global scale metagenome survey at multiple depths across diverse oceanic provinces. The overall distribution and diversity patterns of the marine AOA communities of *Tara* Oceans' surface water samples were broadly similar to those of GOS samples (Fig. 4). Apart from metagenome data collected from shallow waters, the *Tara* Oceans survey extended to mesopelagic depths (250–1000 m), where the WCB genotype was found to dominate marine AOA communities (Fig. 4). It is worth noting that, as has been reported in Monterey Bay surface waters [61], the WCB genotype appeared to be abundant at shallow depths in many upwelling regions (Fig. 4).

In addition to investigating the global surface distribution of marine AOA genotypic groups in the upper oceans, we analyzed the vertical distribution of these lineages from the epipelagic zone of station ALOHA to the hadopelagic waters of Mariana Trench and Yap Trench at depths down to 8000 m (Fig. 5). In the well-stratified water column of station ALOHA, we found a major population shift within a narrow depth interval from 125 m, where *Ca.*



Nitrosopelagicus-like AOA dominated to 500 m, where WCB genotype dominated (Fig. 5). WCB genotype was consistently abundant throughout the dark ocean water column (Fig. 5). In the deep trench waters, a lineage that is closely related to *Nitrosopumilus* genus appeared to be one of the most abundant AOA genotypes (Fig. 5). This genotype was not solely restricted to the hadal zone, and they also constitute a substantial proportion of the total AOA population in energy-depleted bathyal and abyssal zones (Fig. 5). This hadopelagic genotype was well separated from other pelagic AOA lineages (*Ca.* Nitrosopelagicus-like AOA and WCB), but shared more than 98% 16S rRNA gene sequence identity with estuarine and coastal *Nitrosopumilus* species and formed a well-supported monophyletic sister group to the *Nitrosopumilus* genus in the phylogenomic tree (Figs. 1 and S2). Thus, *Nitrosopumilus*-like AOA appear to occupy a wide range of marine habitats and depths that correspond to distinct nutrient regimes, temperatures, and pressures. We hypothesize that genetic adaptations are responsible for their substantial expansion into distinct ecological niches.

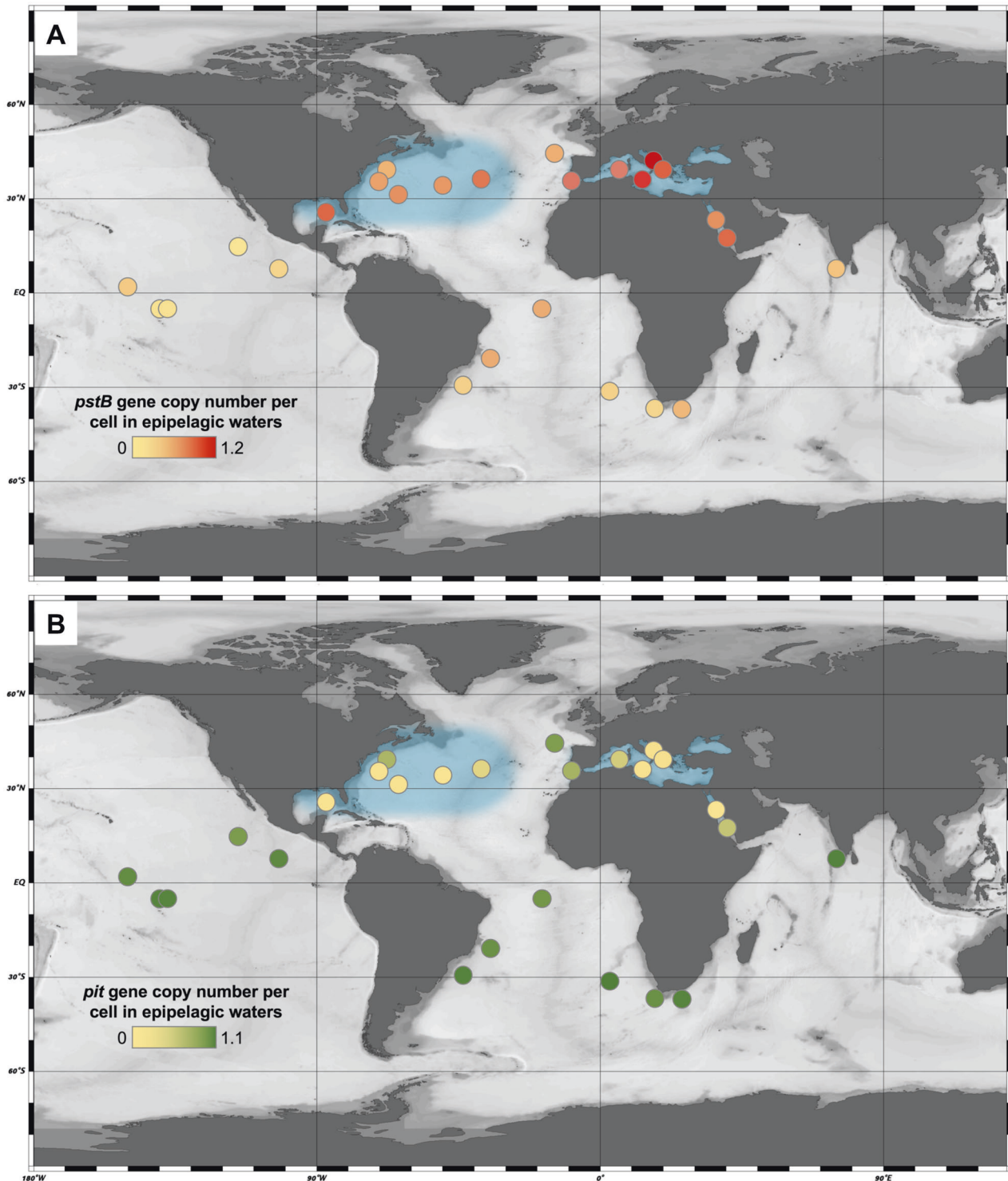
### Genetic diversification associated with niche adaptation in marine AOA

To identify accessory and unique gene contents that are associated with the habitat-specific adaptations as well as abiotic and biotic selective forces that shape genomic heterogeneity among marine AOA populations, we linked the geographic distribution of marine AOA genotypic groups with patterns of variable gene content involved in stress response (Fig. S11 and Supplementary information), nutrient uptake, and metabolic flexibility (Fig. S12 and Supplementary information). A recent comparative analysis of ATPase gene clusters revealed that different marine AOA genotypes contain two distinct types of ATPase with significant variation in subunit composition and structure [62]. *Ca.* Nitrosopelagicus-like AOA encode archaeal-type (A-type) ATPase, while WCB-AOA from the deep ocean possess vacuolar-like (V-type) ATPase [62]. Although the conserved A-type ATPases were found in both closely related *Nitrosopumilus*-like estuarine/coastal AOA and hadopelagic AOA genotypes, an additional V-type ATPase is exclusively present in hadopelagic AOA, suggesting that V-ATPase in *Nitrosopumilus*-like AOA may play a key role in their adaptive expansion to the deep ocean [62].

Consistent with cultured AOA species, environmental marine AOA populations contain two ATPase variants that fall into distinct phylogenetic clusters (Fig. S13A). We examined the distribution of marine AOA A-type and V-type ATPase genes across metagenome datasets from epipelagic to hadopelagic waters. We found that the vertical distribution pattern of these two distinct types of ATPase genes broadly matches

those of the marine AOA genotypes along the water column (Fig. 5). There was an apparent shift in ATPase composition from A-type in the epipelagic zone to V-type in the upper mesopelagic zone (Fig. 5). V-type ATPase predominated between the mesopelagic and bathypelagic zones (500–4000 m) (Fig. 5). Because *Nitrosopumilus*-like hadopelagic AOA species contain both types of ATPase, as the relative abundance of this genotype gradually increased from the bathypelagic to hadopelagic zone, the fraction of A-type ATPase increased, reaching up to 40% of the total ATPase gene reads in metagenomic samples below 5000 m (Fig. 5). Therefore, we extended the original inference of ATPase occurrence in selected marine AOA species, providing a more complete view of the vertical distribution of ATPase gene clusters across metagenomes. The depth partitioning of two distinct types of ATPase revealed by our analysis further supports the ecological role of V-ATPase in adaptation to environmental conditions specific to the deep oceans. Wang et al. (2019) suggest that the acquisition of V-type ATPase in deep marine AOA via horizontal operon transfer confers an adaptive advantage in the deep ocean with elevated hydrostatic pressure, as the proposed function of V-ATPase in pumping excessive cytoplasmic protons at high pressure may serve to maintain the cytosolic pH homeostasis in marine AOA.

Another notable finding of the depth distribution of marine AOA genes is the absence of genes encoding the low-affinity (high  $K_m$ ) AMT in the deep ocean. Although both low- and high-affinity *amt* genes are present in all cultivated marine AOA species [50, 63], the high-quality MAGs recovered from mesopelagic (O23) and hadopelagic waters (YT1, F8–1, and F8–2) lack low-affinity *amt* genes (Table S4). Our metagenomics survey of the entire water column further revealed that low-affinity AOA *amt* genes were restricted to the epipelagic zone and rarely detected at depths below 200 m; the high-affinity AMT was the only type of ammonia uptake system in marine AOA populations inhabiting the ammonia-depleted deep oceans (Fig. 5). In our previous works, we showed that *N. maritimus* sustained high expression levels of the high-affinity *amt* gene at low nanomolar ammonia concentrations or under short-term ammonia starvation, whereas the expression of low-affinity *amt* was depressed under these conditions, indicating that *N. maritimus*, although possessing both types of ammonia uptake systems, selectively retains high-affinity AMT in response to extremely low or even no ammonia supply [28, 50]. Furthermore, our recent findings indicated that the in situ affinity of marine AOA for ammonia increased with depth overall [64]. Consistently, high-affinity AMT would confer a stronger selective advantage over low-affinity AMT for WCB-AOA and *Nitrosopumilus*-like hadopelagic AOA in the deep oceans, supporting their exceptionally high substrate affinities (low  $K_m$  values) at depth where ammonia concentrations and fluxes are extremely low or even undetectable [64].



**Fig. 6** The relative enrichment of high-affinity *pstB* genes and low-affinity *pit* genes in marine AOA populations from the upper water column of diverse oceanic regions. The per-cell gene abundance was estimated based on the relative ratios of *pstB* (a) and *pit* (b) to *amoA* genes recovered from each *Tara* Oceans sampling station by

metagenomic read recruitment, assuming 1 *amoA* gene copy per marine AOA genome. AOA *pstB* gene was enriched in the surface and DCM waters within the low phosphorus regions highlighted by shading, such as the western North Atlantic and the Mediterranean Sea.

We also found that the phylogeny of high-affinity *amt* genes of marine AOA strains tracks habitat, not organismal phylogeny inferred from conserved single-copy genes. The high-affinity *amt* genes of *Nitrosopumilus*-like hadopelagic AOA were not clustered with those encoded by estuarine/coastal *Nitrosopumilus* species, but rather formed a monophyletic group with those of pelagic AOA (Fig. S14). Thus, in addition to V-ATPase, the acquisition of high-affinity AMT from pelagic AOA may also play a significant role in the adaptive expansion of closely related *Nitrosopumilus*-like AOA populations from coastal waters to the oceanic trenches.

The phosphorus utilization capacity in AOA has been poorly characterized relative to their nitrogen utilization and carbon fixation capacities. Many AOA species possess two sets of scavenging systems for orthophosphate, the high-affinity *pst* transporter (*pstABCS*), which is regulated by *phoU*, and low-affinity *pit* transporter, which is regulated by *pit* accessory proteins (Fig. S15). Although a putative phosphonate transporter gene cluster (*phnCDE*) was found in AOA genomes, the lack of an identifiable C-P lyase gene and other described phosphonate hydrolase genes suggests that *phnCDE* encode a transport system for a different substrate in AOA or may be nonfunctional (Fig. S15). The presence of genes encoding polyphosphate utilization (*ppA* and *ppX*) may be responsible for using an intracellular P-reserve under P-limited conditions (Fig. S15).

Although phosphorus concentrations are high in the energy-limited deep oceans, phosphorus can be extremely limited in the upper water column of many oceanic regions [65–67]. Phosphorus deficiency has been increasingly recognized as an important factor that controls community structure and drives genome differentiation of marine microorganisms [68–70]. We used *Tara* Oceans metagenomes to investigate the influence of phosphorus availability on the oceanic distribution of P-acquisition gene content in marine AOA. Notably, the recruitment analysis of AOA P-acquisition genes versus *amoA* genes showed that the frequencies of high-affinity (low  $K_m$ ) and low-affinity (high  $K_m$ ) phosphate transporter genes were strongly associated with phosphate concentrations in the upper ocean. High frequencies of the AOA *pstB* gene that encodes the ATP binding subunit of the high-affinity P transporter were enriched in *Tara* Oceans samples with low phosphate concentrations of less than 200 nM ( $0.54 \pm 0.26$  estimated copy number per cell) (Fig. S16). However, significantly lower frequencies ( $p < 0.01$ ) of AOA *pstB* genes were found in high phosphate samples (0.2–3.3  $\mu$ M) (Fig. S16). In contrast, the low-affinity P transport system (*pit* genes) appears to be part of the core genetic ensemble of marine AOA populations in high phosphate regions, with an average estimated frequency of  $\sim 0.83$  copy number per cell; while frequencies of AOA *pit* genes were significantly

lower in samples with low phosphate concentrations ( $< 200$  nM) relative to high phosphate ( $p < 0.01$ ) (Fig. S16).

No gene encoding the high-affinity P transport system was found in the streamlined genomes of *Ca. Nitrosopelagicus brevis* strains CN25 and U25 recovered from shallow waters of the North Eastern Pacific (Table S4). AOA *pstB* genes were only enriched in upper waters of the oceanic regions with extremely low phosphate concentrations, such as the western North Atlantic and the Mediterranean Sea [65, 66], where marine AOA are expected to constantly experience intense competition for limited phosphate (Fig. 6a). Thus, marine AOA in these oceanic regions appear better equipped to cope with P-limitation than those in high-phosphate regions. Most of the AOA *pstB* genes recovered from these P-depleted oceanic regions formed a monophyletic group that is distinct from those of the cultured AOA species (Fig. S13B). In addition, the frequency of AOA *pit* genes was found to be the lowest in these low phosphate regions (Fig. 6b). The loss of low-affinity P transporter genes is likely a consequence of the transporter having no or limited selective advantage in highly oligotrophic environments [71]. Together, our findings indicate that P availability is a dominant selective force that drives genomic diversification among marine AOA natural populations and revealed the previously underappreciated role of phosphorus in structuring their distribution and ecology.

It has been shown that *Nitrosopumilus maritimus* is capable of synthesizing methylphosphonic acid (MPn), implicated as a major source of methane in the upper oceans [22]. In addition to *N. maritimus*, we found that five estuarine and coastal marine AOA species (strains HMK29, NM25, SPOT01, BD31, and SFB1) and one freshwater AOA species (DW1) possess the complete or nearly complete MPn biosynthetic pathway (Table S4). Methylphosphonic acid synthase (MPnS) represents the key enzyme to synthesize MPn in AOA (Fig. S17). We found thaumarchaeotal *mpnS* genes are widespread in diverse oceanic provinces and enriched in marine AOA populations in deep water habitats relative to shallow water habitats (Fig. S17A, B). Phylogenetic placement of all identified thaumarchaeotal *mpnS* genes and partial fragments revealed that, distinct from the *mpnS* genes encoded in coastal and terrestrial AOA species genomes, most deep marine *mpnS* gene sequences clustered together as new lineages that are likely associated with the uncultured WCB-AOA populations (Fig. S17C). High abundances of *Sulfitobacter* and *Oleiphilus* species that encode C-P lyase for MPn degradation has been found in the deep-water column at Station ALOHA [72], suggesting that phosphonate cycling may be a significant process in the deep ocean. Our findings suggest that MPn is not only a likely source of methane in the upper oceans, but also may be an important source of methane to deep waters. The produced methane may fuel methane-



oxidizing microorganisms in the deep oceans, contributing to the reported decrease in methane concentration and  $^{13}\text{C}$  enrichment of the residual methane at depth [73].

## Conclusions

By bringing together studies of biogeography, cultivation, laboratory evolution, genomics, and physiology, we developed a more holistic understanding of traits associated with the adaptive radiation of AOA into a wide variety of habitats. In particular, by linking the environmental distribution of the major genotypes of marine AOA with their variable gene contents, we identified several variable genes that determine traits associated with ecosystem-specific selection pressures in different oceanic regions. In addition to the previously recognized key environmental variables that control the distribution and diversity of marine AOA populations, such as ammonia concentrations [18, 74], light levels [51, 75], and reactive oxygen species [75–78], our data indicate that phosphate concentrations and hydrostatic pressures drive marine AOA genotypic and gene content variation in the ocean. Likewise, P availability has been identified as a major ecosystem-specific selective pressure that shapes the P-related gene content and gene sequences of another two most abundant marine microbes, *Prochlorococcus* and SAR11 [69]. Our findings reinforce the importance of the acquisition of beneficial nutrient scavenging genes as a common adaptive strategy for marine oligotrophs in nutrient-limited regions of the ocean. Unlike the primary association of *Prochlorococcus* and SAR11 with temperate aquatic environments, AOA are widely distributed—from thermophilic to mesophilic habitats and from terrestrial to marine systems. Our results show that extensive horizontal transfer of genes and entire operons is closely associated with their habitat expansion, likely facilitating their adaptive radiation into a variety of ecological niches, including those spanning a range of temperature, pH, pressure, and nutrient availability.

It has been suggested that AOA play a significant role in shaping biodiversity in marine environments by controlling the forms of fixed nitrogen species available to other microbial assemblages and supplying vitamins to vitamin-dependent populations in the ocean [16, 23]. Our data suggest that interactions between the co-occurring WCB-AOA and bacteria encoding C-P lyase may be important to phosphorus cycling and as a source of methane in the deep ocean. Our comparative genomics and metagenomics analyses should also guide future isolation studies, suggesting that new cultivation strategies, such as high-pressure selection, are possibly required to culture piezotolerant or piezophilic species. In turn, the culture collection of environmental representatives of marine AOA and associated

biota will serve to establish the model systems to investigate how mutualistic or competitive interactions between these dominant taxa and other organisms influence the biogeochemistry of marine and terrestrial ecosystems.

**Acknowledgements** We thank Anthony Bertagnolli, Xiaohua Zhang, and Yanfen Zheng for technical assistance. Jiwei Tian and the crew and scientist party of the R/V Dong Fanghong #2 are thanked for providing the cruise opportunity in which the Mariana Trench F8-1 and F8-2 samples were obtained. This work was supported by the National Natural Science Foundation of China grants 21777155 and 21322703 (to FZ), 41530105 and 91851210 (to CZ), 41907027 (to YZ), U1805242 (to YZ), USA National Science Foundation grants OCE-1046017 and DEB-1664052, and Simons Foundation grants (SCOPE Award ID 329108 to AEI). WQ was supported by Simons Postdoctoral Fellowship in Marine Microbial Ecology (548565). WMH was supported by Florida Agricultural Experiment Station (Hatch project FLA-FTL-005680) and UF IFAS Early Career award (00129069). CZ and HL were also supported by the Shenzhen Key Laboratory of Marine Archaea Geo-Omics, Southern University of Science and Technology (ZDSYS201802081843490). JLN was supported by the Novo Nordisk Foundation (NNF16OC0021818).

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Karner MB, DeLong EF, Karl DM. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature*. 2001;409:507–10.
2. Bates ST, Berg-Lyons D, Caporaso JG, Walters WA, Knight R, Fierer N. Examining the global distribution of dominant archaeal populations in soil. *ISME J*. 2011;5:908–17.
3. Francis CA, Roberts KJ, Beman JM, Santoro AE, Oakley BB. Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. *Proc Natl Acad Sci USA*. 2005;102:14683–8.
4. Biller SJ, Mosier AC, Wells GF, Francis CA. Global biodiversity of aquatic ammonia-oxidizing archaea is partitioned by habitat. *Front Microbiol*. 2012;3:252.
5. Alves RJE, Minh BQ, Urich T, von Haeseler A, Schleper C. Unifying the global phylogeny and environmental distribution of ammonia-oxidising archaea based on amoA genes. *Nat Commun*. 2018;9:1517.
















6. Delong EF. Archaea in coastal marine environments. *Proc Natl Acad Sci USA*. 1992;89:5685–9.
7. Fuhrman JA, McCallum K, Davis AA. Novel major Archaeobacterial group from marine plankton. *Nature*. 1992;356:148–9.
8. Hallam SJ, Mincer TJ, Schleper C, Preston CM, Roberts K, Richardson PM, et al. Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. *PLoS Biol*. 2006;4:520–36.
9. Nunoura T, Takaki Y, Hirai M, Shimamura S, Makabe A, Koide O, et al. Hadal biosphere: insight into the microbial ecosystem in the deepest ocean on Earth. *Proc Natl Acad Sci USA*. 2015;112:1230–6.
10. Delong EF, Wu KY, Prezelin BB, Jovine RVM. High abundance of Archaea in Antarctic marine picoplankton. *Nature*. 1994;371:695–7.
11. Preston CM, Wu KY, Molinski TF, DeLong EF. A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen nov, sp. nov. *Proc Natl Acad Sci USA*. 1996;93:6241–6.
12. Gubry-Rangin C, Hai B, Quince C, Engel M, Thomson BC, James P, et al. Niche specialization of terrestrial archaeal ammonia oxidizers. *Proc Natl Acad Sci USA*. 2011;108:21206–11.
13. de la Torre JR, Walker CB, Ingalls AE, Könneke M, Stahl DA. Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. *Environ Microbiol*. 2008;10:810–8.
14. Zhang CL, Ye Q, Huang ZY, Li WJ, Chen JQ, Song ZQ, et al. Global occurrence of archaeal amoA genes in terrestrial hot springs. *Appl Environ Microbiol*. 2008;74:6417–26.
15. Zhang S, Qin W, Xia X, Xia L, Li S, Zhang L, et al. Ammonia oxidizers in river sediments of the Qinghai-Tibet Plateau and their adaptations to high-elevation conditions. *Water Res*. 2020;173:115589.
16. Martens-Habbena W, Qin W, Horak REA, Urakawa H, Schauer AJ, Moffett JW, et al. The production of nitric oxide by marine ammonia-oxidizing archaea and inhibition of archaeal ammonia oxidation by a nitric oxide scavenger. *Environ Microbiol*. 2015;17:2261–74.
17. Martens-Habbena W, Berube PM, Urakawa H, de la Torre JR, Stahl DA. Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. *Nature*. 2009;461:976–U234.
18. Horak REA, Qin W, Schauer AJ, Armbrust EV, Ingalls AE, Moffett JW, et al. Ammonia oxidation kinetics and temperature sensitivity of a natural marine community dominated by Archaea. *ISME J*. 2013;7:2023–33.
19. Könneke M, Schubert DM, Brown PC, Hugler M, Standfest S, Schwander T, et al. Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO<sub>2</sub> fixation. *Proc Natl Acad Sci USA*. 2014;111:8239–44.
20. Wuchter C, Abbas B, Coolen MJL, Herfort L, van Bleijswijk J, Timmers P, et al. Archaeal nitrification in the ocean. *Proc Natl Acad Sci USA*. 2006;103:12317–22.
21. Santoro AE, Buchwald C, McIlvin MR, Casciotti KL. Isotopic signature of N<sub>2</sub>O produced by marine ammonia-oxidizing archaea. *Science*. 2011;333:1282–5.
22. Metcalf WW, Griffin BM, Cicchillo RM, Gao JT, Janga SC, Cooke HA, et al. Synthesis of methylphosphonic acid by marine microbes: a source for methane in the aerobic ocean. *Science*. 2012;337:1104–7.
23. Heal KR, Qin W, Ribalet F, Bertagnolli AD, Coyote-Maestas W, Hmelo LR, et al. Two distinct pools of B<sub>12</sub> analogs reveal community interdependencies in the ocean. *Proc Natl Acad Sci USA*. 2017;114:364–9.
24. Doxey AC, Kertz DA, Lynch MDJ, Sauder LA, Neufeld JD. Aquatic metagenomes implicate *Thaumarchaeota* in global cobalamin production. *ISME J*. 2015;9:461–71.
25. Könneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA. Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature*. 2005;437:543–6.
26. Qin W, Heal KR, Ramdasi R, Kobelt JN, Martens-Habbena W, Bertagnolli AD, et al. *Nitrosopumilus maritimus* gen. nov., sp. nov., *Nitrosopumilus cobalaminigenes* sp. nov., *Nitrosopumilus oxycyclinae* sp. nov., and *Nitrosopumilus ureiphilus* sp. nov., four marine ammonia-oxidizing archaea of the phylum Thaumarchaeota. *Int J Syst Evol Microbiol*. 2017;67:5067–79.
27. Stieglmeier M, Klingl A, Alves RJE, Rittmann SKMR, Melcher M, Leisch N, et al. *Nitrososphaera viennensis* gen. nov., sp. nov., an aerobic and mesophilic, ammonia-oxidizing archaeon from soil and a member of the archaeal phylum Thaumarchaeota. *Int J Syst Evol Microbiol*. 2014;64:2738–52.
28. Qin W, Amin SA, Lundeen RA, Heal KR, Martens-Habbena W, Turkarslan S, et al. Stress response of a marine ammonia-oxidizing archaeon informs physiological status of environmental populations. *ISME J*. 2018;12:508–19.
29. 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 USA*. 2010;107:8818–23.
30. Schouten S, Hopmans EC, Baas M, Boumann H, Standfest S, Könneke M, et al. Intact membrane lipids of “*Candidatus Nitrosopumilus maritimus*,” a cultivated representative of the cosmopolitan mesophilic group I crenarchaeota. *Appl Environ Microbiol*. 2008;74:2433–40.
31. Hurlley SJ, Elling FJ, Könneke M, Buchwald C, Wankel SD, Santoro AE, et al. Influence of ammonia oxidation rate on thaumarchaeal lipid composition and the TEX<sub>86</sub> temperature proxy. *Proc Natl Acad Sci USA*. 2016;113:7762–7.
32. Li PN, Herrmann J, Tolar BB, Poitevin F, Ramdasi R, Bargar JR, et al. Nutrient transport suggests an evolutionary basis for charged archaeal surface layer proteins. *ISME J*. 2018;12:2389–402.
33. Qin W, Carlson LT, Armbrust EV, Devol AH, Moffett JW, Stahl DA, et al. Confounding effects of oxygen and temperature on the TEX<sub>86</sub> signature of marine Thaumarchaeota. *Proc Natl Acad Sci USA*. 2015;112:10979–84.
34. 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 USA*. 2016;113:7937–46.
35. Herbold CW, Lehtovirta-Morley LE, Jung MY, Jehmlich N, Hausmann B, Han P, et al. Ammonia-oxidizing archaea living at low pH: Insights from comparative genomics. *Environ Microbiol*. 2017;19:4939–52.
36. Jung MY, Islam MA, Gwak JH, Kim JG, Rhee SK. *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;68:3084–95.
37. Bayer B, Vojvoda J, Reinthaler T, Reyes C, Pinto M, Herndl GJ. *Nitrosopumilus adriaticus* sp. nov. and *Nitrosopumilus piranensis* sp. nov., two ammonia-oxidizing archaea from the Adriatic Sea and members of the class *Nitrososphaeria*. *Int J Syst Evol Microbiol*. 2019;7:1892–1902.
38. 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;13:2150–61.
39. Wang Y, Huang JM, Cui GJ, Nunoura T, Takaki Y, Li WL, et al. Genomics insights into ecotype formation of ammonia-oxidizing archaea in the deep ocean. *Environ Microbiol*. 2019;21:716–29.
40. Zou D, Li Y, Kao S-J, Liu H, Li M. Genomic adaptation to eutrophication of ammonia-oxidizing archaea in the Pearl River estuary. *Environ Microbiol*. 2019;21:2320–32.

41. Ahlgren NA, Chen YY, Needham DM, Parada AE, Sachdeva R, Trinh V, et al. Genome and epigenome of a novel marine Thaumarchaeota strain suggest viral infection, phosphorothioation DNA modification and multiple restriction systems. *Environ Microbiol.* 2017;19:2434–52.
42. Abby SS, Melcher M, Kerou M, Krupovic M, Stieglmeier M, Rossel C, et al. *Candidatus Nitrosocaldus cavascurensis*, an ammonia oxidizing, extremely thermophilic archaeon with a highly mobile genome. *Front Microbiol.* 2018;9:28.
43. Daebeler A, Herbold CW, Vierheilig J, Sedlacek CJ, Pjevac P, Albertsen M, et al. Cultivation and genomic analysis of “*Candidatus Nitrosocaldus islandicus*,” an obligately thermophilic, ammonia-oxidizing thaumarchaeon from a hot spring biofilm in Graendalur Valley, Iceland. *Front Microbiol.* 2018;9:193.
44. Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, Rodrigue S, et al. Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet.* 2007;3:2515–28.
45. Grote J, Thrash JC, Huggett MJ, Landry ZC, Carini P, Giovannoni SJ, et al. Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *Mbio.* 2012;3:e00252–12.
46. Kits KD, Sedlacek CJ, Lebedeva EV, Han P, Bulaev A, Pjevac P, et al. Kinetic analysis of a complete nitrifier reveals an oligotrophic lifestyle. *Nature.* 2017;549:269–72.
47. Jung MY, Kim JG, Damste JSS, Rijpstra WIC, Madsen EL, Kim SJ, et al. A hydrophobic ammonia-oxidizing archaeon of the Nitrosocosmicus clade isolated from coal tar-contaminated sediment. *Environ Microbiol Rep.* 2016;8:983–92.
48. Sauder LA, Albertsen M, Engel K, Schwarz J, Nielsen PH, Wagner M, et al. Cultivation and characterization of *Candidatus Nitrosocosmicus exaquare*, an ammonia-oxidizing archaeon from a municipal wastewater treatment system. *ISME J.* 2017;11:1142–57.
49. Lehtovirta-Morley LE, Ross J, Hink L, Weber EB, Gubry-Rangin C, Thion C, et al. Isolation of ‘*Candidatus Nitrosocosmicus franklandus*’, a novel ureolytic soil archaeal ammonia oxidiser with tolerance to high ammonia concentration. *FEMS Microbiol Ecol.* 2016;92:fiw057.
50. Nakagawa T, Stahl DA. Transcriptional response of the archaeal ammonia oxidizer *Nitrosopumilus maritimus* to low and environmentally relevant ammonia concentrations. *Appl Environ Microbiol.* 2013;79:6911–6.
51. Santoro AE, Dupont CL, Richter RA, Craig MT, Carini P, McIlvin MR, et al. Genomic and proteomic characterization of “*Candidatus Nitrosopelagicus brevis*”: an ammonia-oxidizing archaeon from the open ocean. *Proc Natl Acad Sci USA.* 2015;112:1173–8.
52. Carini P, Dupont CL, Santoro AE. Patterns of thaumarchaeal gene expression in culture and diverse marine environments. *Environ Microbiol.* 2018;20:2112–24.
53. Shi YM, Tyson GW, Eppley JM, DeLong EF. Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *ISME J.* 2011;5:999–1013.
54. Hollibaugh JT, Gifford S, Sharma S, Bano N, Moran MA. Metatranscriptomic analysis of ammonia-oxidizing organisms in an estuarine bacterioplankton assemblage. *ISME J.* 2011;5:866–78.
55. Heal KR, Qin W, Amin SA, Devol AH, Moffett JW, Armbrust EV, et al. Accumulation of NO<sub>2</sub>-cobalamin in nutrient-stressed ammonia-oxidizing archaea and in the oxygen deficient zone of the eastern tropical North Pacific. *Environ Microbiol Rep.* 2018;10:453–7.
56. Kozłowski JA, Stieglmeier M, Schleper C, Klotz MG, Stein LY. Pathways and key intermediates required for obligate aerobic ammonia-dependent chemolithotrophy in bacteria and Thaumarchaeota. *ISME J.* 2016;10:1836–45.
57. Qin W, Amin SA, Martens-Habbena W, Walker CB, Urakawa H, Devol AH, et al. Marine ammonia-oxidizing archaeal isolates display obligate mixotrophy and wide ecotypic variation. *Proc Natl Acad Sci USA.* 2014;111:12504–9.
58. Osburne MS, Holmbeck BM, Coe A, Chisholm SW. The spontaneous mutation frequencies of *Prochlorococcus* strains are commensurate with those of other bacteria. *Environ Microbiol Rep.* 2011;3:744–9.
59. Gibson B, Wilson DJ, Feil E, Eyre-Walker A. The distribution of bacterial doubling times in the wild. *Proc R Soc B.* 2018;285:20180789.
60. Stahl DA, de la Torre JR. Physiology and diversity of ammonia-oxidizing archaea. *Annu Rev Microbiol.* 2012;66:83–101.
61. Smith JM, Casciotti KL, Chavez FP, Francis CA. Differential contributions of archaeal ammonia oxidizer ecotypes to nitrification in coastal surface waters. *ISME J.* 2014;8:1704–14.
62. 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;13:3067–79.
63. Offre P, Kerou M, Spang A, Schleper C. Variability of the transporter gene complement in ammonia-oxidizing archaea. *Trends Microbiol.* 2014;22:665–75.
64. Zhang Y, Qin W, Hou L, Zakem EJ, Wan X, Zhao Z, et al. Nitrifier adaptation to low energy flux controls inventory of reduced nitrogen in the dark ocean. *Proc Natl Acad Sci USA.* 2020;117:4823–30.
65. Fanning KA. Nutrient provinces in the sea—concentration ratios, reaction-rate ratios, and ideal covariation. *J Geophys Res-Oceans.* 1992;97:5693–712.
66. Krom MD, Kress N, Brenner S, Gordon LI. Phosphorus limitation of primary productivity in the eastern Mediterranean Sea. *Limnol Oceanogr.* 1991;36:424–32.
67. Karl DM. Microbially mediated transformations of phosphorus in the sea: new views of an old cycle. *Annu Rev Mar Sci.* 2014;6:279–337.
68. Carini P, White AE, Campbell EO, Giovannoni SJ. Methane production by phosphate-starved SAR11 chemoheterotrophic marine bacteria. *Nat Commun.* 2014;5:4346.
69. Coleman ML, Chisholm SW. Ecosystem-specific selection pressures revealed through comparative population genomics. *Proc Natl Acad Sci USA.* 2010;107:18634–9.
70. Martiny AC, Coleman ML, Chisholm SW. Phosphate acquisition genes in *Prochlorococcus* ecotypes: evidence for genome-wide adaptation. *Proc Natl Acad Sci USA.* 2006;103:12552–7.
71. Albalat R, Canestro C. Evolution by gene loss. *Nat Rev Genet.* 2016;17:379–91.
72. Sosa OA, Repeta DJ, Ferron S, Bryant JA, Mende DR, Karl DM, et al. Isolation and characterization of bacteria that degrade phosphonates in marine dissolved organic matter. *Front Microbiol.* 2017;8:1786.
73. Holmes ME, Sansone FJ, Rust TM, Popp BN. Methane production, consumption, and air-sea exchange in the open ocean: an evaluation based on carbon isotopic ratios. *Glob Biogeochem Cy.* 2000;14:1–10.
74. Sintes E, De Corte D, Haberleitner E, Herndl GJ. Geographic distribution of archaeal ammonia oxidizing ecotypes in the Atlantic Ocean. *Front Microbiol.* 2016;7:77.
75. Horak REA, Qin W, Bertagnolli AD, Nelson A, Heal KR, Han H, et al. Relative impacts of light, temperature, and reactive oxygen on thaumarchaeal ammonia oxidation in the North Pacific Ocean. *Limnol Oceanogr.* 2018;63:741–57.
76. Qin W, Meinhardt KA, Moffett JW, Devol AH, Armbrust EV, Ingalls AE, et al. Influence of oxygen availability on the activities

- of ammonia-oxidizing archaea. *Environ Microbiol Rep.* 2017;9: 250–6.
77. Tolar BB, Powers LC, Miller WL, Wallsgrove NJ, Popp BN, Hollibaugh JT. Ammonia oxidation in the ocean can be inhibited by nanomolar concentrations of hydrogen peroxide. *Front Mar Sci.* 2016;3:237.
78. Bayer B, Pelikan C, Bittner MJ, Reinthaler T, Konneke M, Herndl GJ, et al. Proteomic response of three marine ammonia-oxidizing archaea to hydrogen peroxide and their metabolic interactions with a heterotrophic alphaproteobacterium. *Msystems.* 2019;4: e00181–19.
79. Vajrala N, Martens-Habbena W, Sayavedra-Soto LA, Schauer A, Bottomley PJ, Stahl DA, et al. Hydroxylamine as an intermediate in ammonia oxidation by globally abundant marine archaea. *Proc Natl Acad Sci USA.* 2013;110: 1006–11.

## Affiliations

Wei Qin <sup>1</sup> · Yue Zheng <sup>2,3</sup> · Feng Zhao <sup>2,3</sup> · Yulin Wang <sup>4</sup> · Hidetoshi Urakawa <sup>5</sup> · Willm Martens-Habbena<sup>6</sup> · Haodong Liu<sup>7</sup> · Xiaowu Huang<sup>8</sup> · Xinxu Zhang<sup>9</sup> · Tatsunori Nakagawa<sup>10</sup> · Daniel R. Mende<sup>11</sup> · Annette Bollmann<sup>12</sup> · Baozhan Wang <sup>13</sup> · Yao Zhang<sup>14</sup> · Shady A. Amin <sup>15</sup> · Jeppe L. Nielsen <sup>16</sup> · Koji Mori<sup>17</sup> · Reiji Takahashi<sup>10</sup> · E. Virginia Armbrust <sup>1</sup> · Mari-K.H. Winkler<sup>18</sup> · Edward F. DeLong <sup>11</sup> · Meng Li <sup>9</sup> · Po-Heng Lee<sup>19,8</sup> · Jizhong Zhou <sup>20,21,22</sup> · Chuanlun Zhang<sup>7</sup> · Tong Zhang<sup>4</sup> · David A. Stahl<sup>18</sup> · Anitra E. Ingalls <sup>1</sup>

<sup>1</sup> School of Oceanography, University of Washington, Seattle, WA, USA

<sup>2</sup> CAS Key Laboratory of Urban Pollutant Conversion, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen, China

<sup>3</sup> University of Chinese Academy of Sciences, Beijing, China

<sup>4</sup> Environmental Microbiome Engineering and Biotechnology Laboratory, Center for Environmental Engineering Research, Department of Civil Engineering, The University of Hong Kong, Hong Kong, China

<sup>5</sup> Department of Ecology and Environmental Studies, Florida Gulf Coast University, Fort Myers, FL, USA

<sup>6</sup> Fort Lauderdale Research and Education Center, Department of Microbiology and Cell Science, Institute of Food and Agricultural Sciences, University of Florida, Davie, FL, USA

<sup>7</sup> Department of Ocean Science and Engineering, Shenzhen Key Laboratory of Marine Archaea Geo-Omics, Southern University of Science and Technology, Shenzhen, China

<sup>8</sup> Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong, China

<sup>9</sup> Shenzhen Key Laboratory of Marine Microbiome Engineering, Institute for Advanced Study, Shenzhen University, Shenzhen, China

<sup>10</sup> College of Bioresource Sciences, Nihon University, Fujisawa, Kanagawa, Japan

<sup>11</sup> Daniel K. Inouye Center for Microbial Oceanography: Research and Education, University of Hawaii, Honolulu, HI, USA

<sup>12</sup> Department of Microbiology, Miami University, Oxford, OH, USA

<sup>13</sup> Key Lab of Microbiology for Agricultural Environment, Ministry of Agriculture, College of Life Sciences, Nanjing Agricultural University, Nanjing, China

<sup>14</sup> State Key Laboratory of Marine Environmental Sciences and College of Ocean and Earth Sciences, Xiamen University, Xiamen, China

<sup>15</sup> Department of Biology, New York University Abu Dhabi, Abu Dhabi, UAE

<sup>16</sup> Department of Chemistry and Bioscience, Aalborg University, Aalborg, Denmark

<sup>17</sup> NITE Biological Resource Center (NBRC), National Institute of Technology and Evaluation (NITE), Kisarazu, Chiba, Japan

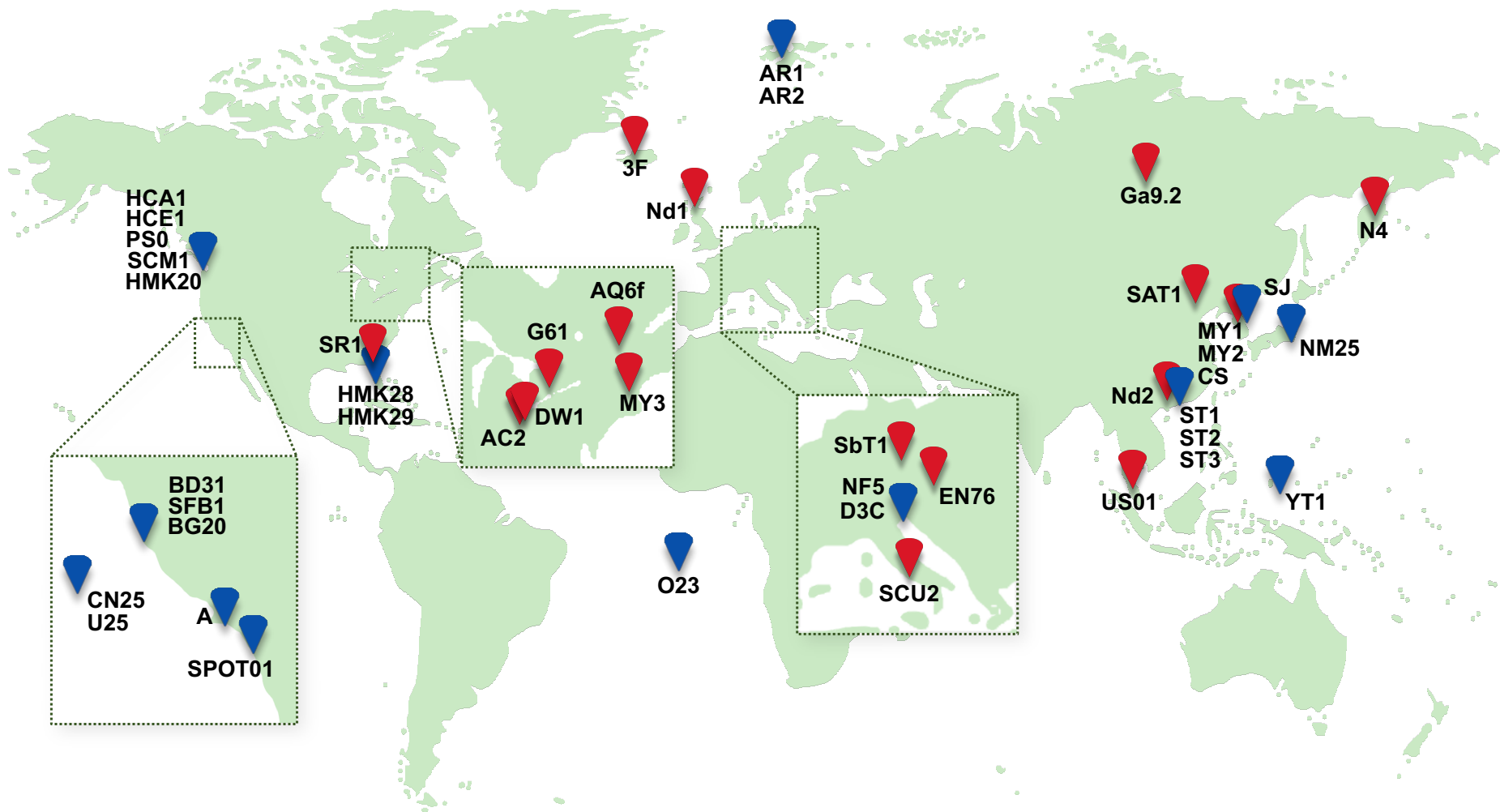
<sup>18</sup> Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, USA

<sup>19</sup> Department of Civil and Environmental Engineering, Imperial College London, London, UK

<sup>20</sup> Institute for Environmental Genomics, Department of Microbiology and Plant Biology, and School of Civil Engineering and Environmental Sciences, University of Oklahoma, Norman, OK, USA

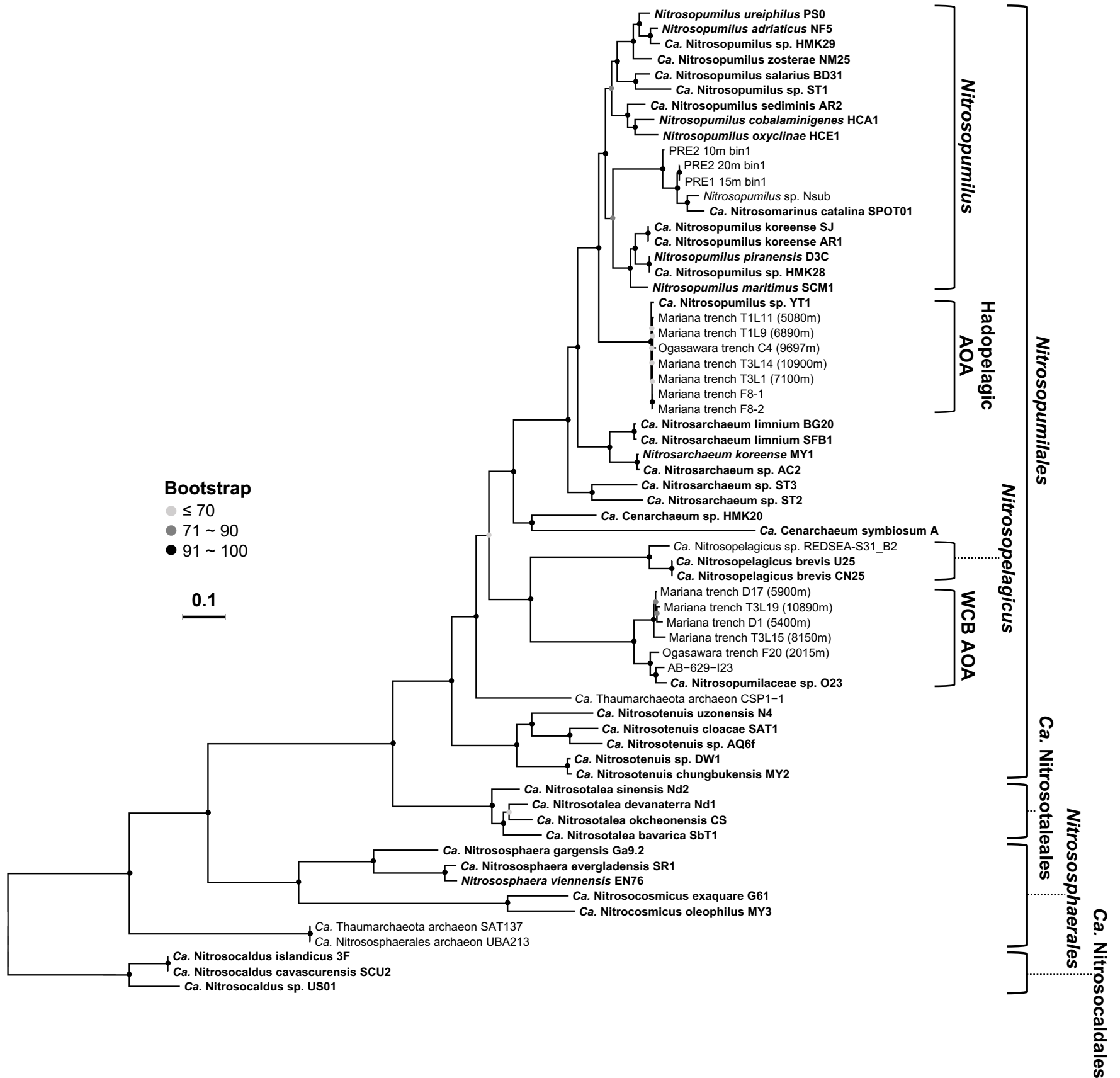
<sup>21</sup> Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

<sup>22</sup> School of Environment, Tsinghua University, Beijing, China

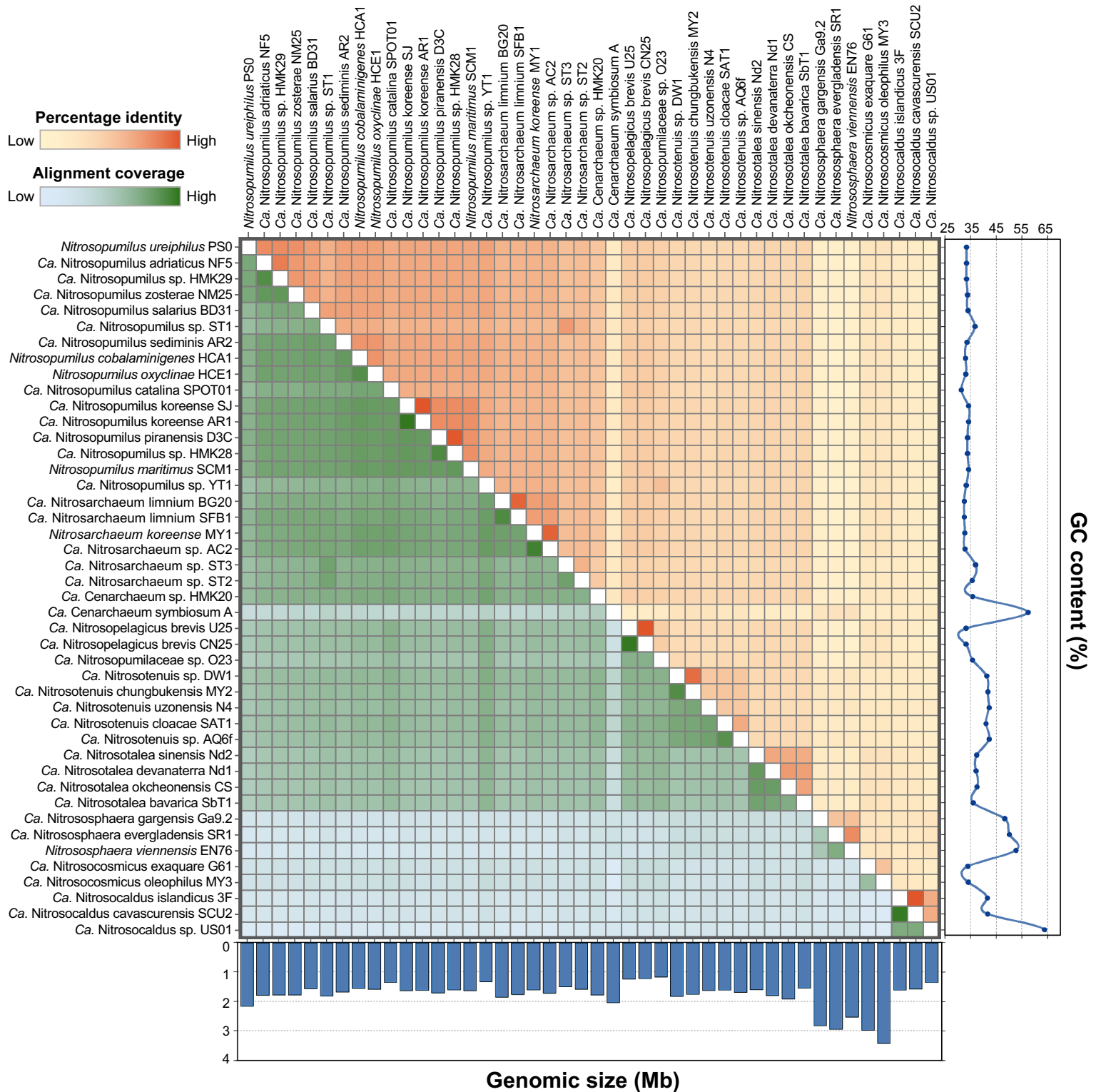


**Figure S1.** The geographic origin of the 44 AOA species described in this study. The blue and red pins indicate the locations of marine and terrestrial AOA species, respectively.

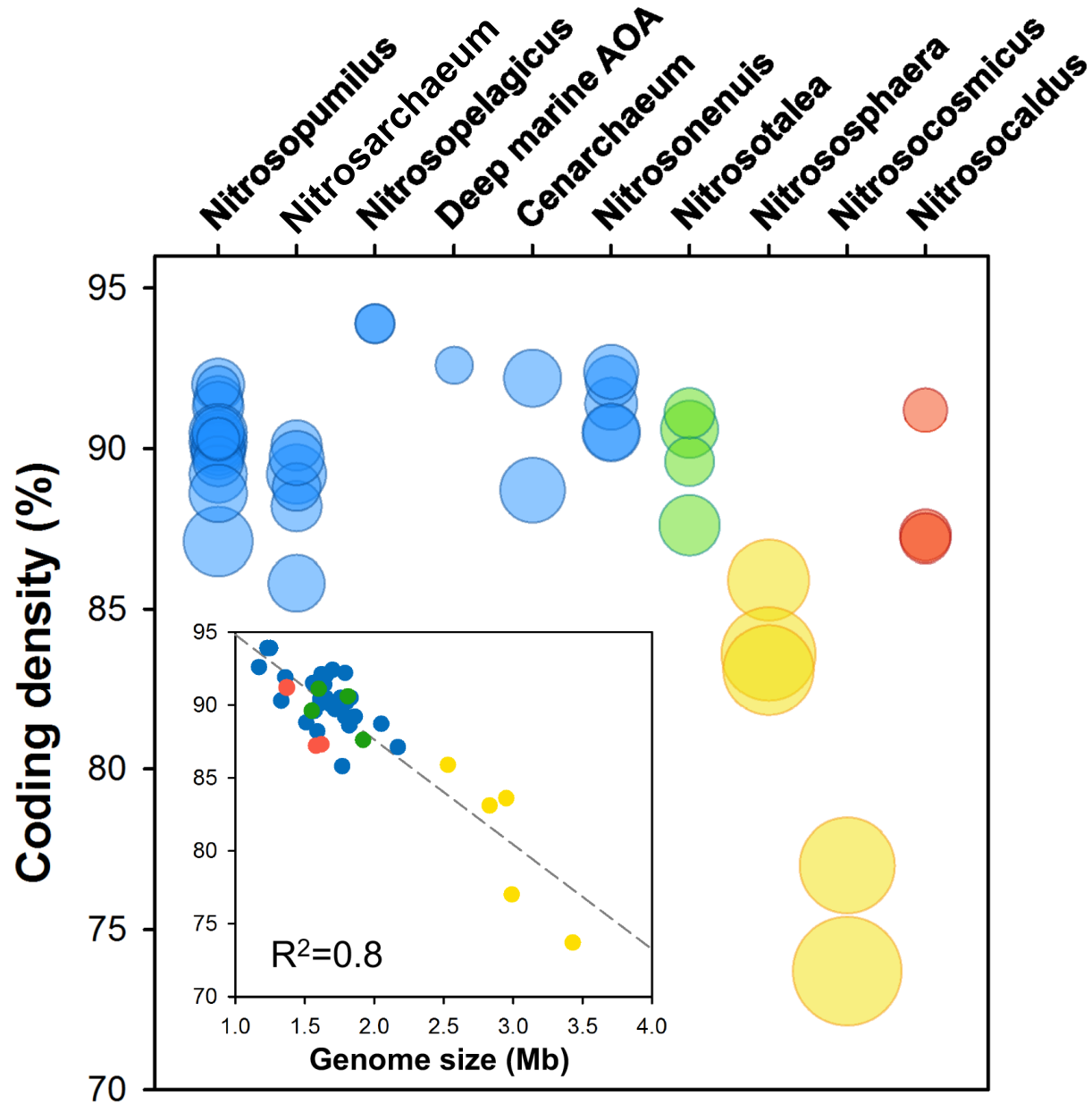




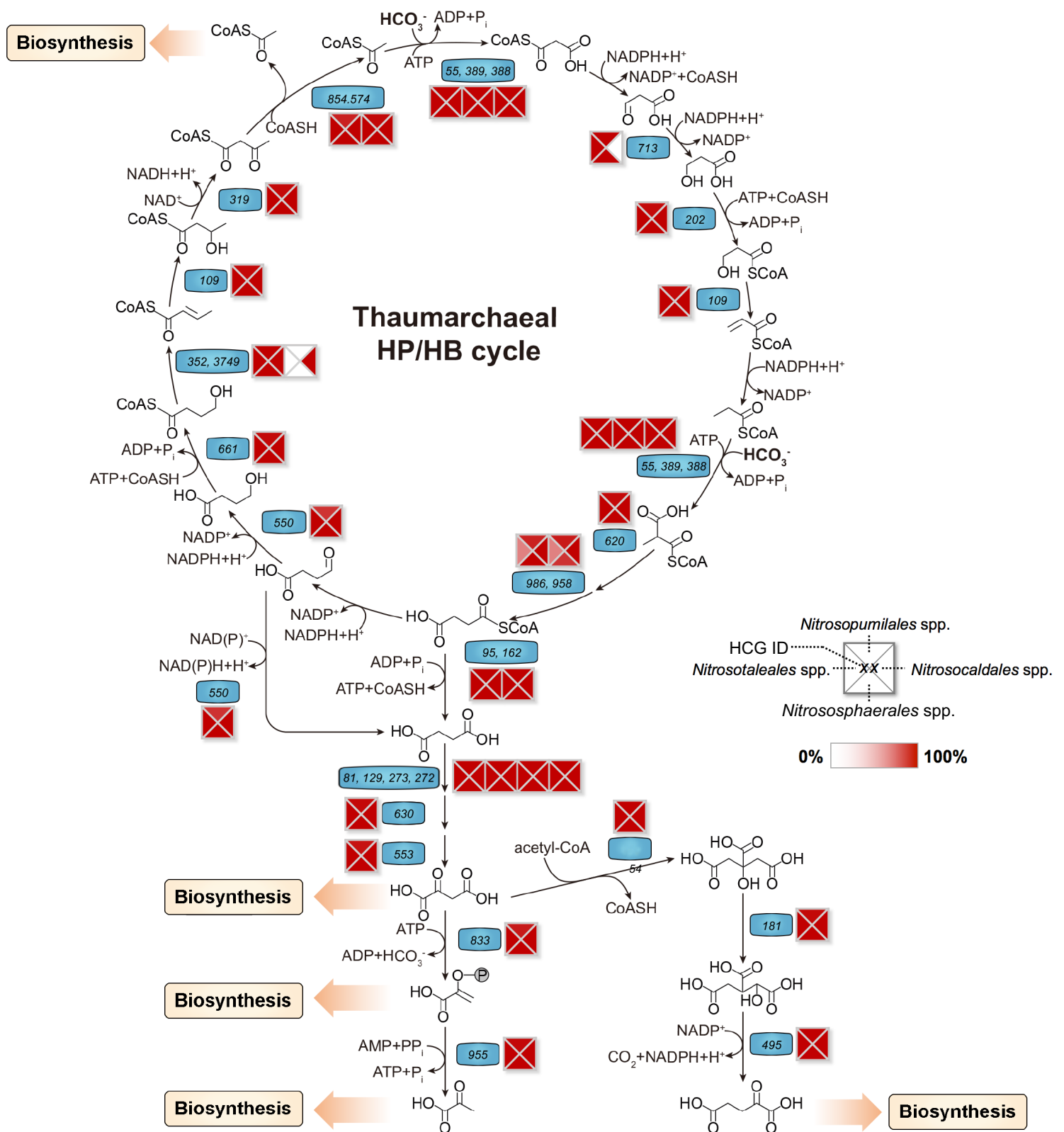
**Figure S2.** Phylogeny of 65 AOA genomes based on concatenated sequences of 71 single-copy core genes (Table S2). The 44 culture genomes and MAGs shown in Figure 1 and used for comparative genomic analysis are highlighted in bold. Confidence values are on the basis of 100 bootstrap replications. The scale bar represents 10% estimated sequence divergence.



**Figure S3.** Heatmap showing the fraction of the total genes in the genome shared between two AOA genomes (bottom left half of the matrix) and the average nucleotide identity (ANI) of the shared genes between the two genomes (upper right half of the matrix). Genome size and GC content (%) are displayed at the bottom and on the side of the matrix, respectively.

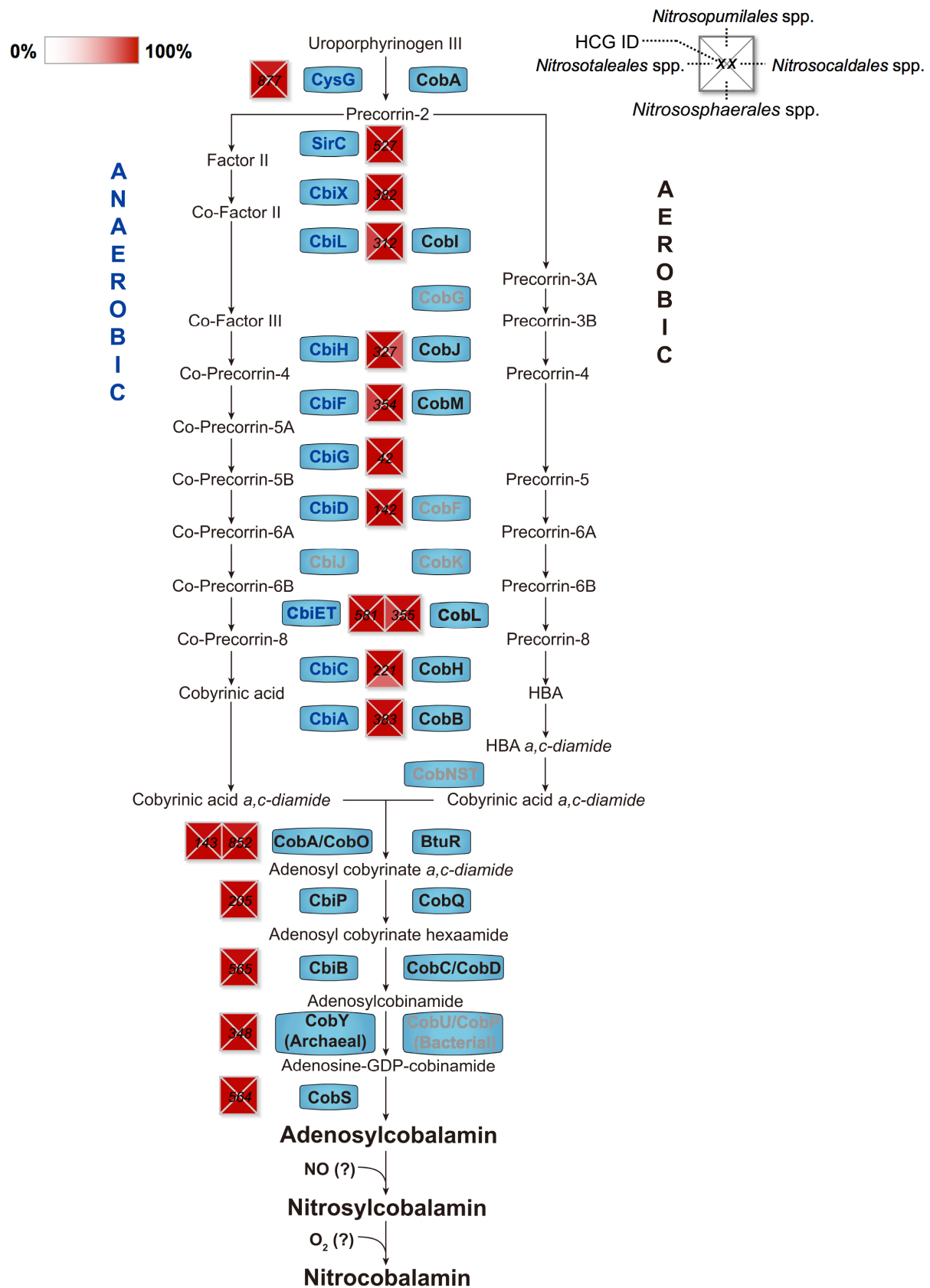


**Figure S4.** Correlation of genome size and genome coding density of AOA members of the orders *Nitrosopumilales* (blue), *Ca. Nitrosotaleales* (green), *Nitrososphaerales* (yellow), and *Ca. Nitrosocaldales* (red). Bubble size indicates the genome size of AOA species. The linear regression line (inset panel) is  $y = -7.1891x + 102$  ( $R^2 = 0.8$ ).

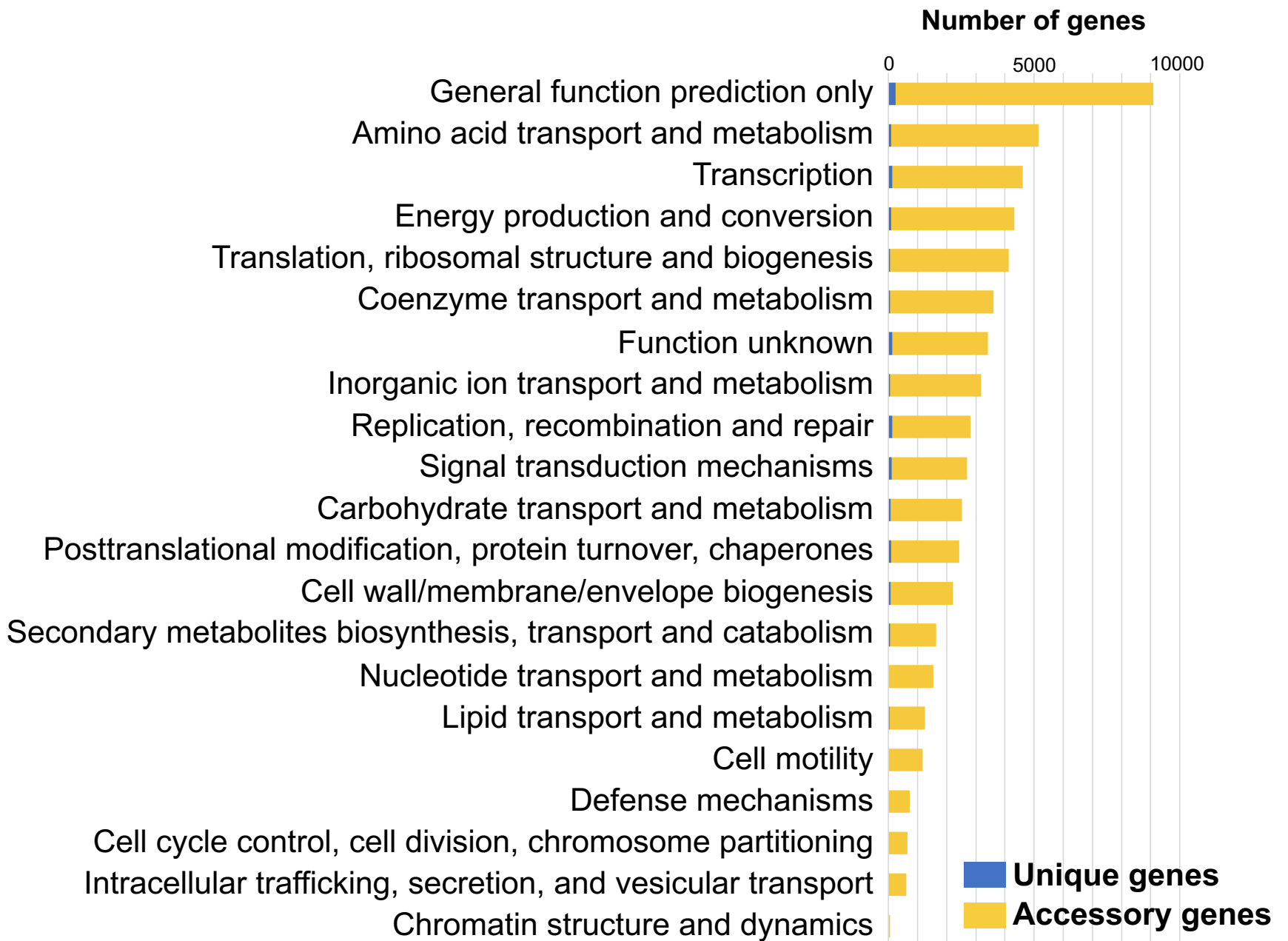


**Figure S5.** Reconstruction of the thaumarchaeotal 3-HP/4-HB carbon fixation pathway emphasizing the conservation and uniqueness of pathway enzymes of AOA species. Squares indicate the COGs of proteins in AOA species genomes, and COG numbers are inside blue boxes. The figure arrangement and color schemes are the same as those shown in Figure 3.

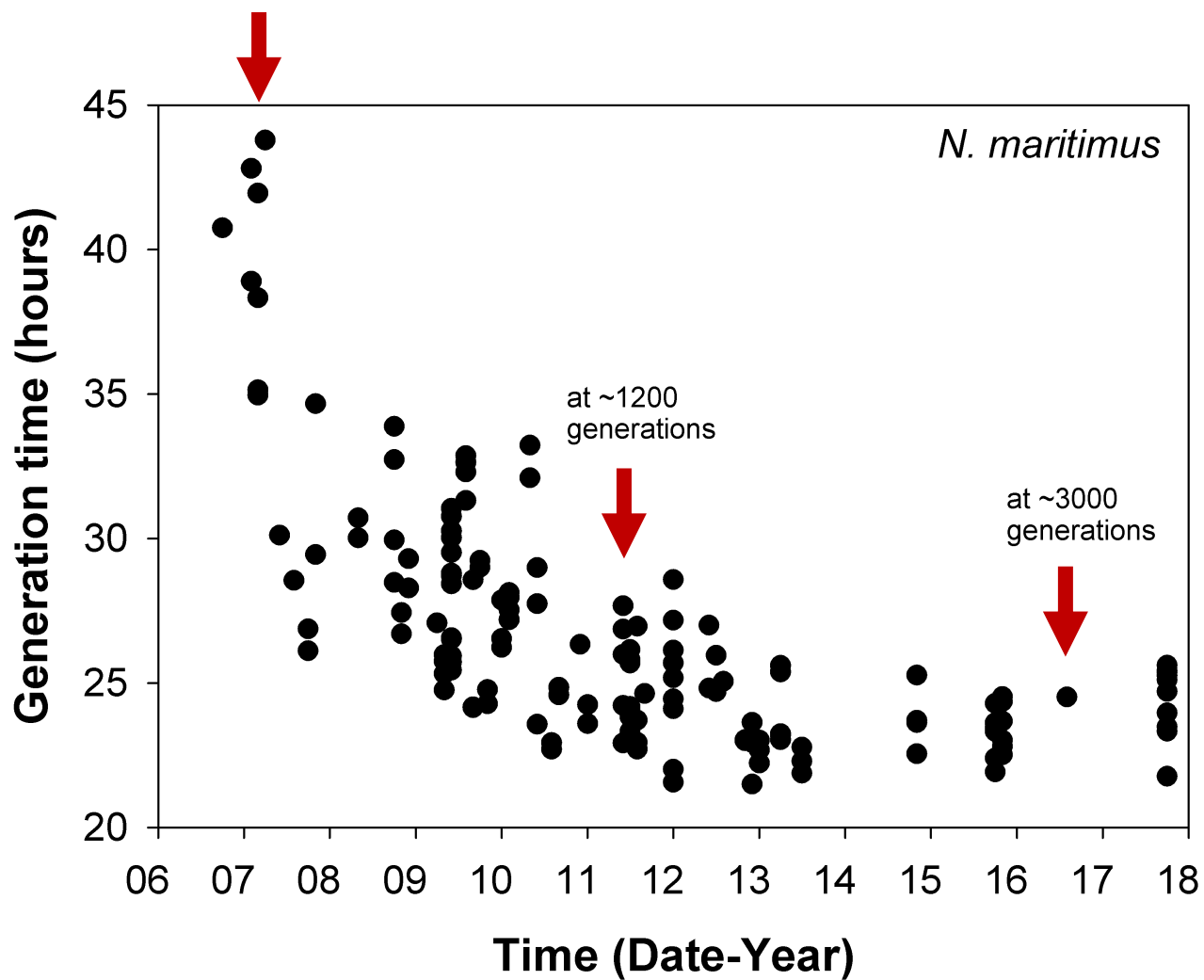




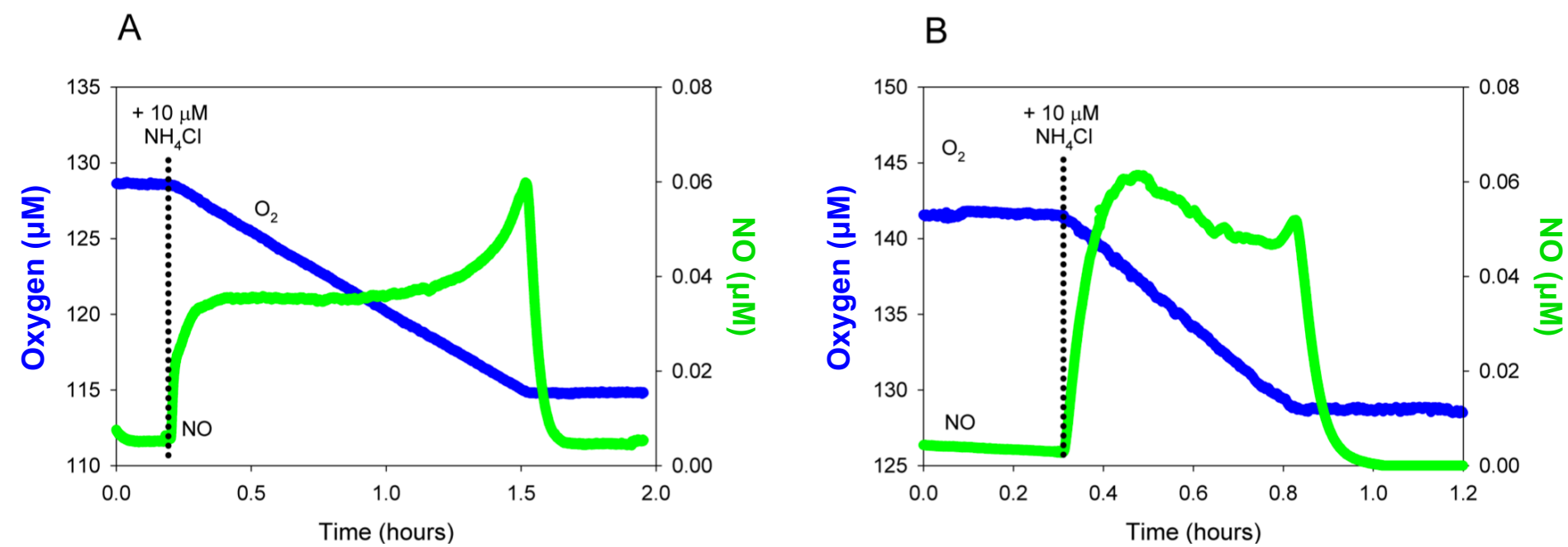
**Figure S6.** Reconstruction of the thaumarchaeotal cobalamin biosynthetic pathway emphasizing the conservation and uniqueness of pathway enzymes of AOA species. The figure arrangement and color schemes are the same as those shown in Figure 3. Gene names in grey correspond to enzymes not yet identified in thaumarchaeotal genomes.



**Figure S7.** Distribution of the COG functional categories assigned to the accessory and unique genes of AOA pan-genome.

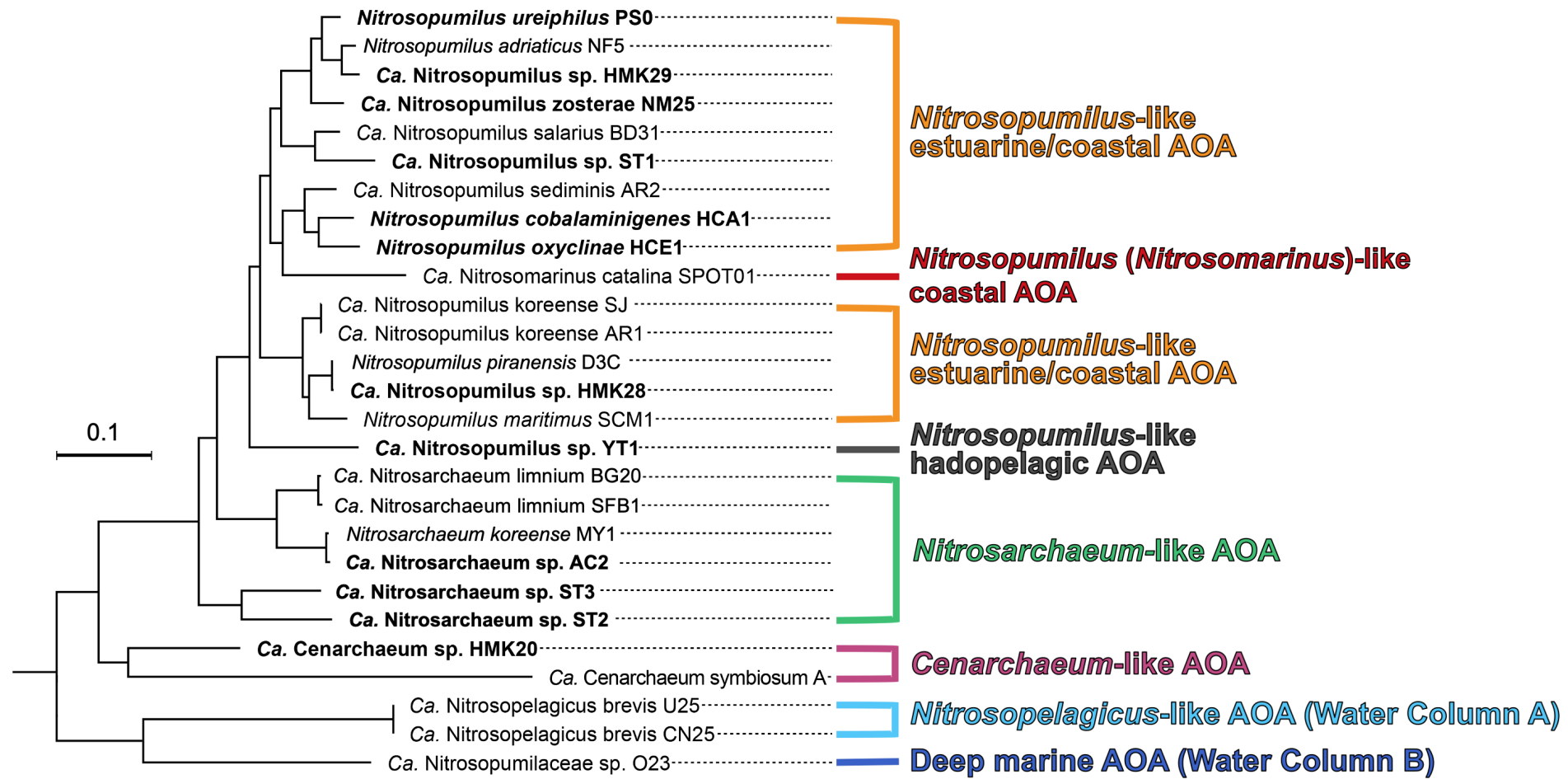


**Figure S8.** Generation time of *Nitrosopumilus maritimus* during adaptive growth over 11 years of continuous transferring. We measured generation time in the exponential phase of growth. Red arrows indicate that the evolved cultures were collected in 2011 and 2016 and mutations were characterized by genome resequencing compared with the ancestral culture genome sequenced in 2007.



**Figure S9.** Oxygen uptake and NO accumulation measurements following ammonia addition to ammonia-depleted ancestral (A) and evolved (B) cultures of *Nitrosopumilus maritimus*.

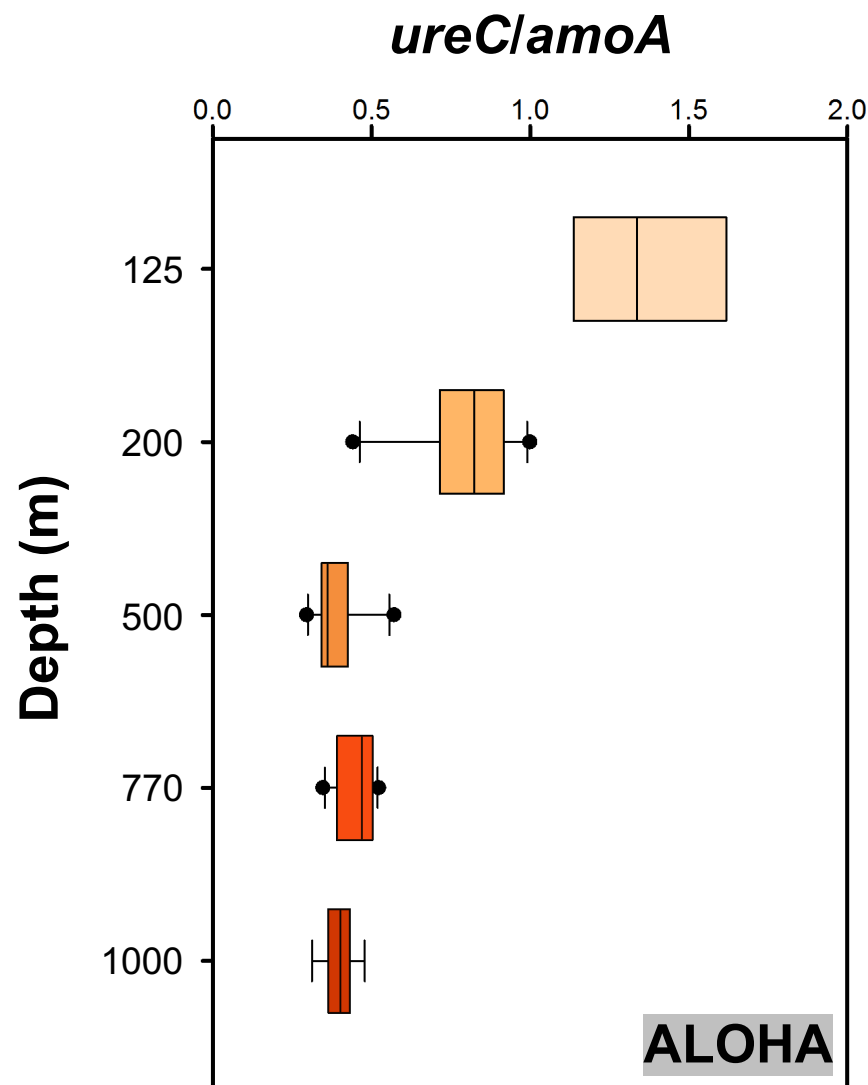




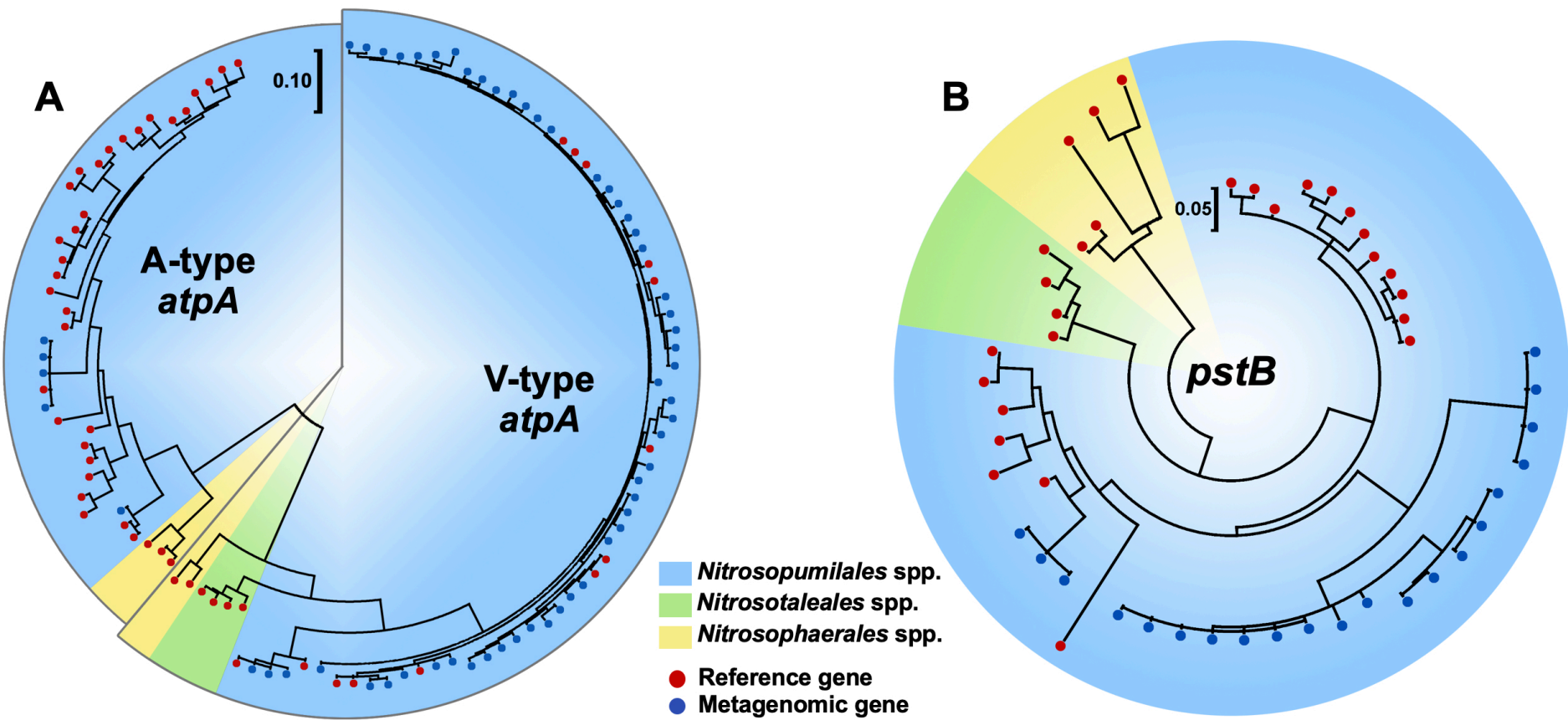
**Figure S10.** Marine AOA genotypic group assignment based on the position of 25 marine AOA species in the phylogenomic tree and their distinct geographic distributions.



**Figure S11.** Map depicting from where AOA ectoine/hydroxyectoine biosynthetic genes were recovered in GOS metagenomes. Red dots and grey dots indicate the GOS sampling stations with and without ectoine/hydroxyectoine biosynthetic genes detection, respectively.



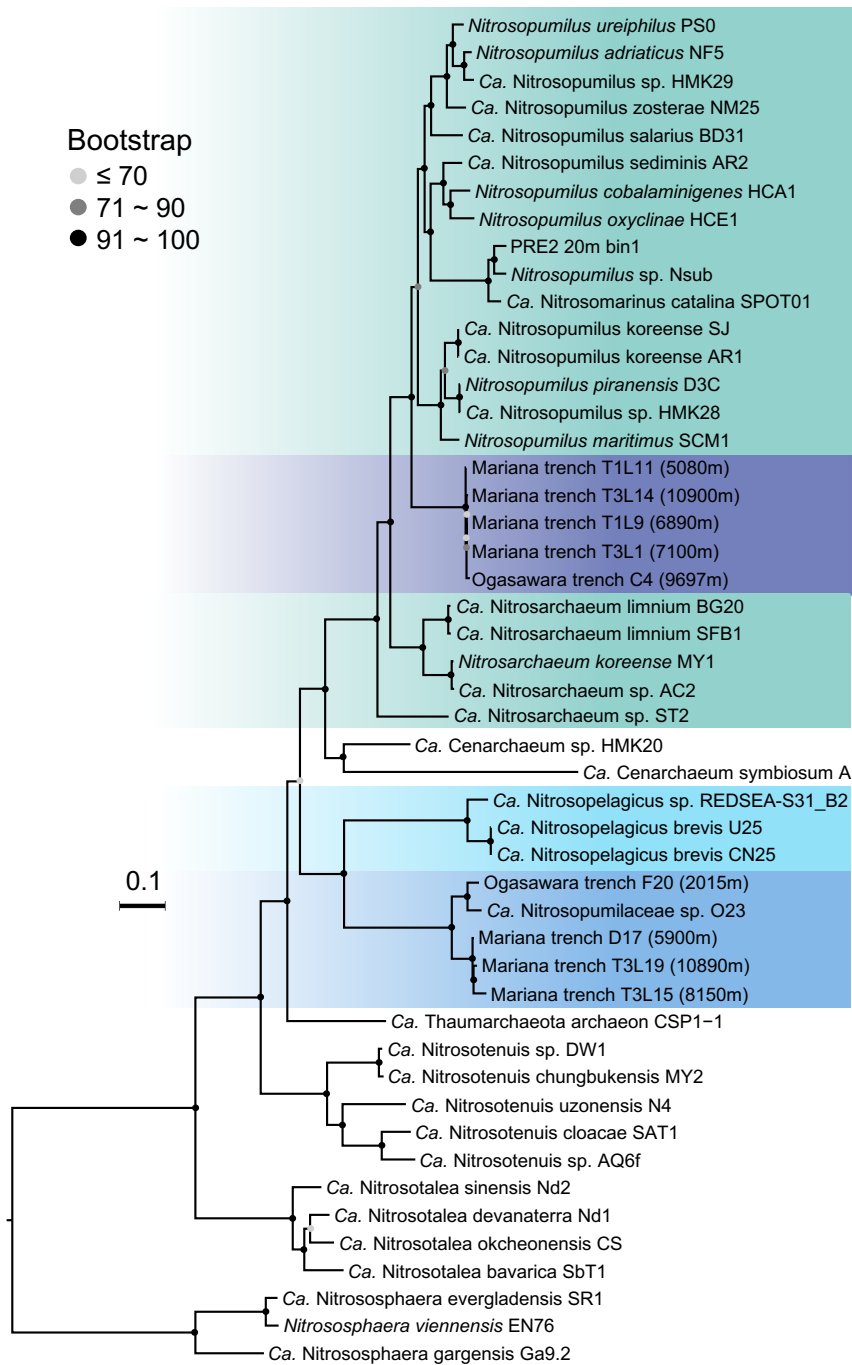
**Figure S12.** Box-and-whisker plots of the estimated relative abundance of *ureC* genes in marine AOA populations throughout the upper ocean water column at Station ALOHA by metagenomic read recruitment. Metagenomic sequencing data were generated from samples obtained at Station ALOHA during HOT cruises 224 225, 227, 229, 231, 232, 233, 234, 236, 237, and 238. The relative ratio of *ureC* to *amoA* genes was calculated as the ratio of gene length-normalized coverage for the reads that with a minimum length of 100 bp, a maximum E-value of  $1 \times e^{-10}$ , and a minimum sequence identity of 80% to marine AOA *ureC* and *amoA* genes found in cultures and environmental clones. The high *ureC:amoA* ratios over 1.0 may reflect different identity thresholds of between sequence divergence of two genes.



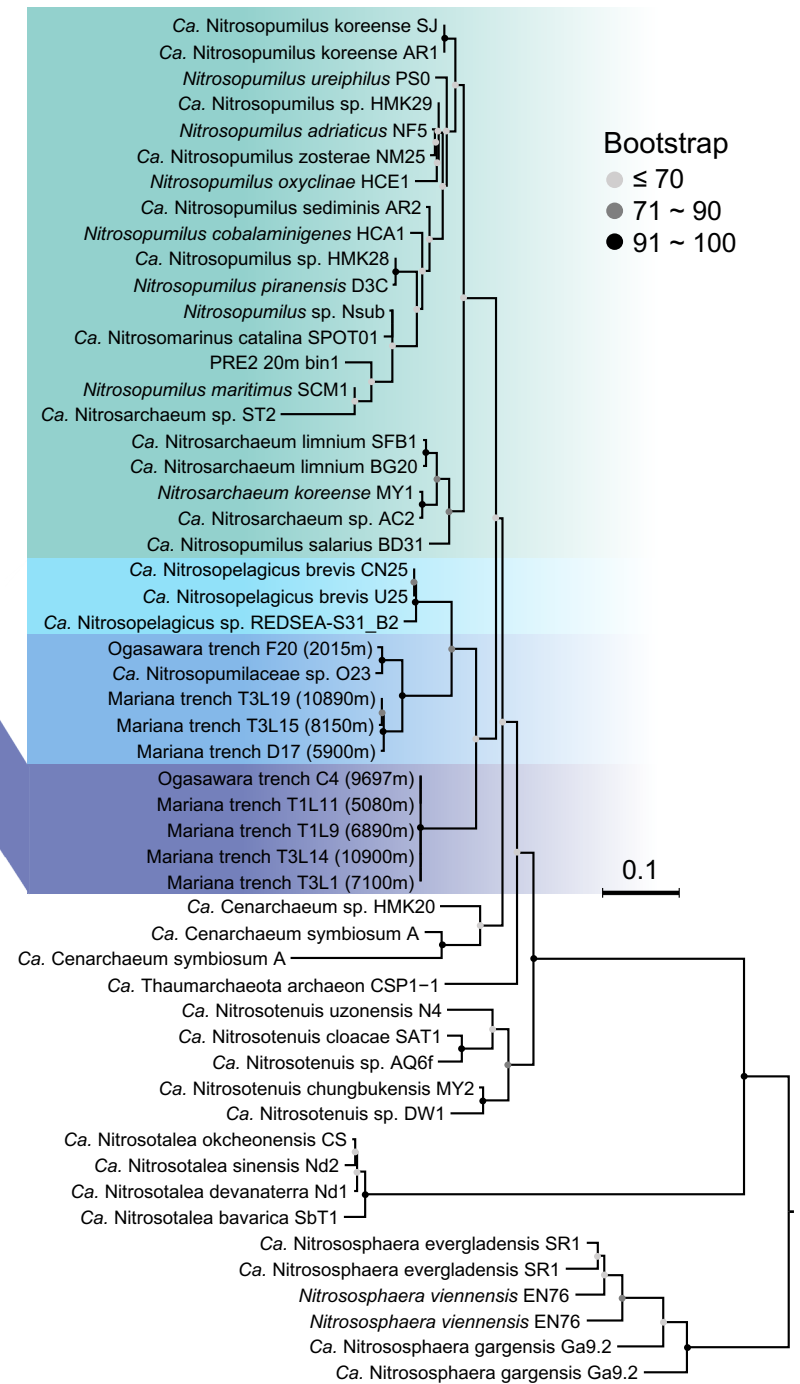
**Figure S13.** Phylogenetic patterns of AOA *atpA* (encoding the alpha subunit of ATPase) and *pstB* genes. The AOA *atpA* (A) and *pstB* (B) genes that were recovered from *Tara* Oceans metagenomes are denoted with blue dots at the branch ends. The *atpA* and *pstB* genes of cultured AOA species are denoted with red dots. The blue, green, and yellow shadings represent the *atpA* and *pstB* genes affiliated to the order *Nitrosopumilales*, *Ca. Nitrosotaleales*, and *Nitrososphaerales*, respectively.



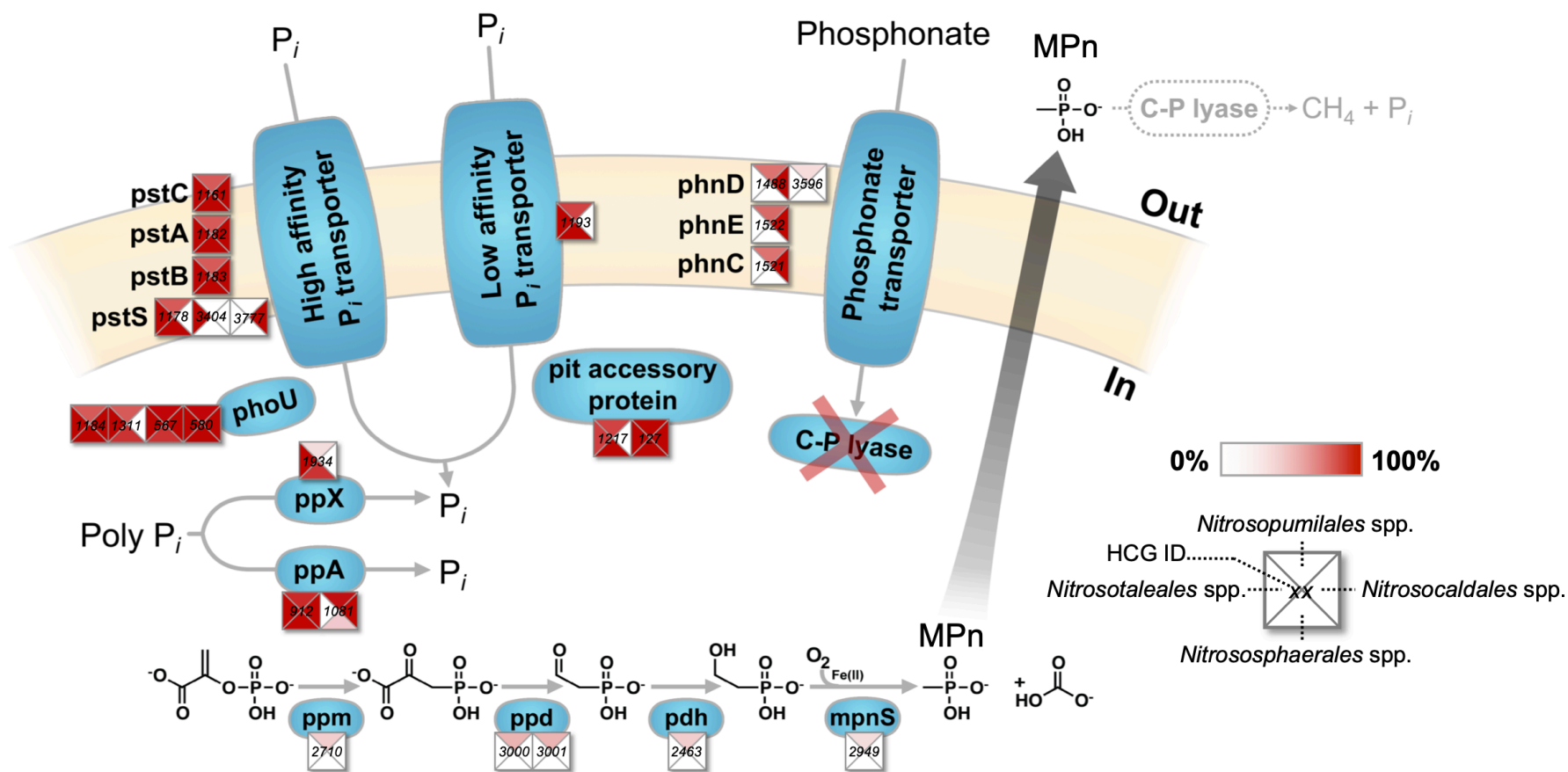
### A) AOA species tree



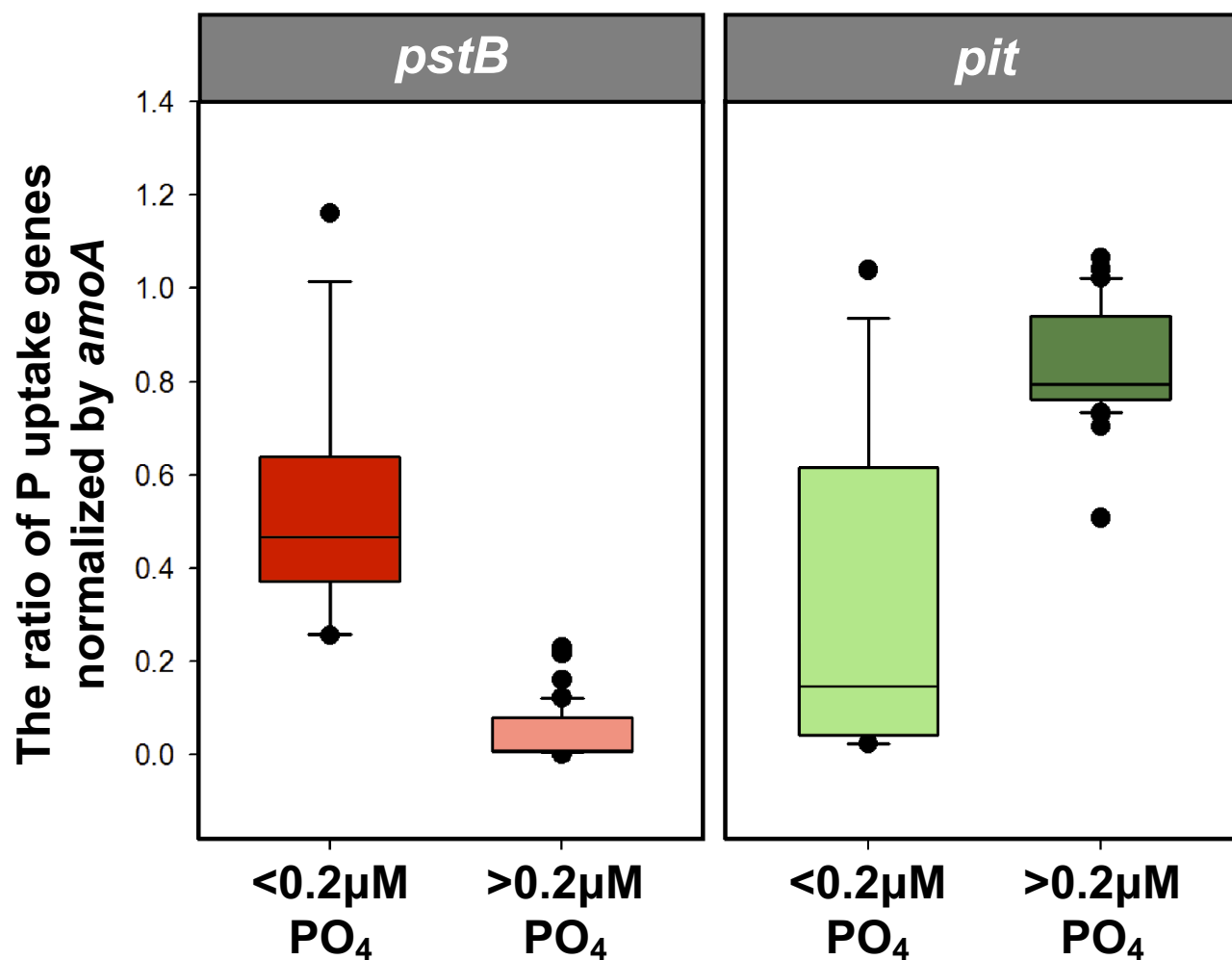
### B) High-affinity ammonia transporter genes



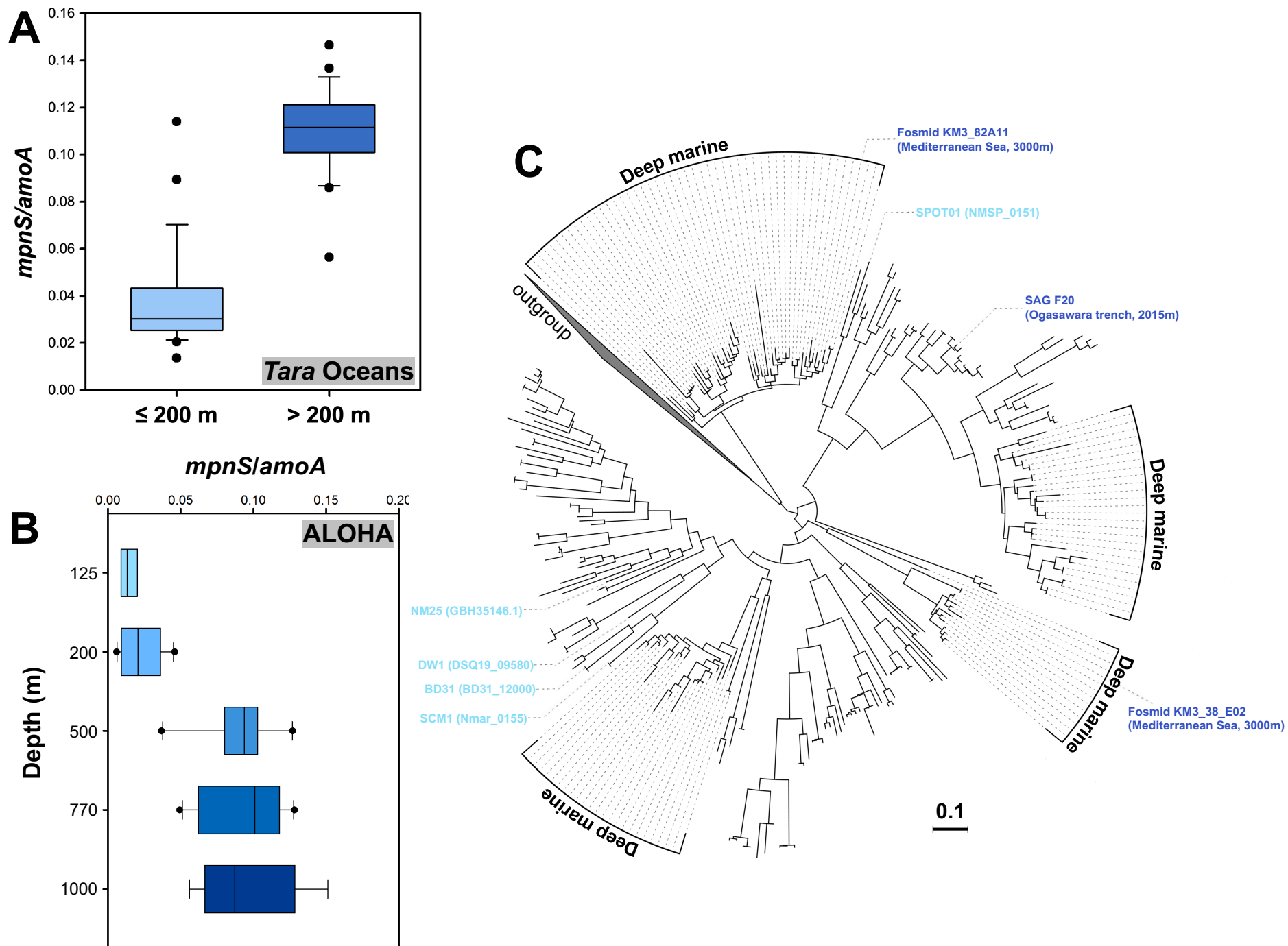
**Figure S14.** Comparative phylogeny of 71 conserved single-copy genes (Table S2) (A) and high-affinity *amt* genes (B) of marine and soil AOA. The green, light blue, dark blue, and purple shadings highlight genomes and *amt* genes affiliated to the *Nitrosopumilus/Nitrosarchaeum* estuarine/coastal AOA, *Nitrosopelagicus*-like AOA (WCA), deep marine AOA (WCB), and *Nitrosopumilus*-like hadopelagic AOA, respectively. Confidence values are on the basis of 100 bootstrap replications. The scale bar represents 0.1 substitutions per nucleotide position.



**Figure S15.** Reconstruction of the proposed pathways of phosphorus acquisition and transformation in AOA emphasizing the conservation and uniqueness of pathway enzymes of AOA species. The figure arrangement and color schemes are the same as those shown in Figure 3. Abbreviations: *pstSCAB*, high-affinity phosphate transport; *phoU*, phosphate uptake regulator; *ppX*, exopolyphosphatase; *ppA*, pyrophosphatase; *phnCDE*, putative phosphonate transport; *ppm*, phosphoenolpyruvate mutase; *ppd*, phosphonopyruvate decarboxylase; *pdh*, phosphonoacetaldehyde dehydrogenase; *mpnS*, methylphosphonic acid synthase; MPn, methylphosphonic acid. Note that no C-P lyase homolog was found in any AOA species. Grey arrow indicates that MPn may be transported out of AOA cells, degraded by other microbes with C-P lyase, which contributes to the methane production in the aerobic ocean.



**Figure S16.** Box-and-whisker plots of the estimated relative abundance of high-affinity (*pstB*) and low-affinity (*pit*) phosphate transporter genes in marine AOA populations from the extremely P-limited (< 200 nM) and relatively P-enriched (0.2 - 3.3  $\mu\text{M}$ ) oceanic regions, respectively, by metagenomic recruitment. The relative ratios of *pstB* and *pit* genes to *amoA* genes were calculated as the ratio of gene length-normalized coverage for the *Tara* Oceans metagenomic sequence reads that with a maximum E-value of  $1 \times e^{-10}$  and a minimum sequence identity of 80% to the *pstB*, *pit*, and *amoA* genes found in marine AOA cultured species and MAGs.



**Figure S17.** Distribution and diversity of AOA *mpnS* genes in the oceans. (A) Box-and-whisker plots of the estimated per-cell *mpnS* gene abundance in marine AOA populations from shallow ( $\leq 200$  m) and deep waters ( $> 200$  m), respectively, at *Tara* Ocean sampling stations by metagenomic recruitment. (B) Box-and-whisker plots of the estimated relative abundance of *mpnS* genes in marine AOA populations throughout the upper ocean water column (125-1000 m) at Station ALOHA by metagenomic read recruitment. Metagenomic sequencing data were generated from samples obtained at Station ALOHA during HOT cruises 224, 225, 227, 229, 231, 232, 233, 234, 236, 237, and 238. The per-cell *mpnS* gene abundance was estimated based on the relative ratios of *mpnS* to *amoA* genes by metagenomic read recruitment, assuming one *amoA* gene copy per marine AOA genome. (C) Phylogeny of *mpnS* genes retrieved from AOA cultures, fosmids, SAGs, *Tara* Oceans and ALOHA metagenomes. Reference sequences of AOA *mpnS* genes from shallow ( $\leq 200$  m) and deep waters (2000-3000 m) are indicated in light and dark blue text, respectively. “Deep marine” clusters represent the *mpnS* sequences recovered from mesopelagic depths (250-1000 m) at *Tara* Oceans and ALOHA stations. The scale bar represents 0.1 substitutions per nucleotide position.



## Supplementary Information

### Results and Discussion

A notable feature of *Nitrosopumilus* physiology is the demonstrated ability of *Nitrosopumilus maritimus* to synthesize ectoine (1,4,5,6-tetrahydro-2-methyl-4-pyrimidinecarboxylic acid) and hydroxyectoine in response to osmotic stress [1]. Our comparative genomics indicated that the complete biosynthetic gene cluster (*hyp-ectABCD*) was not restricted to the genus *Nitrosopumilus*, but also occurred in some species from the genera *Nitrosarchaeum* and *Ca. Cenarchaeum* (Table S4). Previous metagenomics studies reported that ectoine/hydroxyectoine biosynthetic gene clusters were present in halotolerant marine AOA populations from the brine-seawater interface [2]. Our broad metagenomics survey further revealed the widespread distribution of the genetic capacity of ectoine/hydroxyectoine biosynthesis in marine AOA populations spanning estuarine, coastal and open ocean surface waters (Fig. S11). Like other osmolytes, ectoine likely serves additional functions beyond osmoprotection and these functions may be selected for in different environments. For example, in bacteria [3, 4], ectoine may serve to protect AOA cells against UV radiation or cytotoxic stresses in surface waters.

It has been recognized that many cultured and environmental AOA populations contain the complete *ure* gene cluster (urease, urease accessory proteins, and urea transporters) and can use urea as a source of ammonia for energy production and growth [5-9]. Recent field studies reported that the in situ specific affinities for urea-based ammonia oxidation were comparable to those for in situ ammonia-based ammonia oxidation in the ocean, indicating that besides ammonia, urea appears to play an important role in fueling marine nitrification [10]. The Hawaii Ocean Time-Series (HOT) station metagenome datasets were used to assess the prevalence of

urea utilizing genes in natural marine AOA communities by calculating the ratio of reads that mapped to AOA *ureC* genes (encoding the alpha subunit of urease) with the criterion of E-value  $\leq 1 \times e^{-10}$ , reads sequence identity  $\geq 80\%$ , and reads length  $\geq 100$  bp divided by reads that mapped to AOA *amoA* genes (Fig S12). We found that *ureC* genes were distributed throughout the upper ocean water column at the HOT station (125–1,000 m) (Fig. S12). The recruitment ratio of AOA *ureC* genes versus *amoA* genes indicates that ~40–100% of marine AOA contain urea utilizing genes (Fig. S12). Similarly, high *ureC:amoA* ratios were also observed in the equatorial Pacific Ocean (22–55%) [11], the Gulf of Mexico (~10–15%) [8], and the San Pedro Ocean Time-Series (SPOT) site (~60–100%) [12], together highlighting the pervasiveness and importance of urea as an alternative substrate for AOA in the ocean.

## Materials and Methods

### Sample source, culture maintenance and genome sequencing

All AOA species were maintained in liquid mineral medium in the dark without shaking. *Nitrosopumilus ureiphilus* strain PS0, *Nitrosopumilus cobalaminigenes* strain HCA1, and *Nitrosopumilus oxyclinae* strain HCE1 were cultured in HEPES buffered synthetic crenarchaeota medium (SCM) supplemented with 0.2–1 mM NH<sub>4</sub>Cl and 100  $\mu$ M  $\alpha$ -ketoglutaric acid at 25°C as described previously [13]. Cells were harvested in mid-exponential phase via vacuum filtration on 0.22  $\mu$ m Sterivex-GP (Millipore Corporation, Billerica, MA, USA) filters for genome DNA extraction. To minimize DNA shearing, the sucrose lysis method was used for DNA extraction. In brief, the filter was removed from the Sterivex cartridge, aseptically cut into small pieces with flame-sterilized scissors, and transferred to a sterile 15 ml Falcon tube. Two milliliters of fresh sucrose lysis buffer containing 50 mM Tris-HCl at pH 8.0, 40 mM EDTA at pH 8.0, and 0.75 M

sucrose was added to the Falcon tube. Immediate flash freezing of the sample in liquid nitrogen and subsequent thawing at room temperature were conducted to release DNA from AOA cells. After one freeze and thaw cycle, 465  $\mu$ l of 10% SDS and 15  $\mu$ l of 20mg/ml proteinase K were added, and the contents were mixed by vortexing for 2 minutes. The tube was incubated at 37°C for 2 hours with agitation at 30 minutes intervals. The aqueous phase was transferred to a spin column, and DNA was collected using Qiagen Blood & Tissue kit according to the manufacturer's recommendations.

Genomic DNA was sequenced by a combination of the Illumina MiSeq platform with paired-end (PE) reads of 250 bp and the Pacific Biosciences (Pacbio) RSII SMRT sequencing platform using a 20 kb SMRTbell template library. *De novo* assembly of the Pacbio sequence reads was conducted using the Hierarchical Genome Assembly tool in the PacBio SMRT Analysis Software (RS\_HGAP\_Assembly.3) [14]. Illumina sequence reads were trimmed and filtered by Trimmomatic (version 0.36) to remove low-quality reads [15]. The initial Pacbio-based assemblies were error corrected with high-quality Illumina reads using Pilon (version 1.23) [16]. The corrected genomes were annotated through NCBI Prokaryotic Genome Annotation Pipeline [17].

*Ca. Nitrosarchaeum* sp. strain AC2 and *Ca. Nitrosotenuis* sp. strain DW1 were enriched from near-shore sediments of Lake Acton (39°57'N, 84°74'W, USA) and Lake Delaware (40°39'N, 83°05'W, USA), respectively [18]. These two freshwater AOA enrichment cultures were maintained in HEPES buffered freshwater SCM supplemented with 0.5 mM NH<sub>4</sub>Cl at 30°C under dark as described previously [18]. Cells were harvested at mid-exponential phase and

genomic DNA was extracted using a sucrose lysis method as described above. DNA extracted from each culture was sequenced on a SMRT cell using PacBio Sequel platform with insert size of approximately 10 kb. The Pacbio sequence reads were clustered according to the Latent Strain Analysis algorithm [19], and the clustered reads were *de novo* assembled by CANU (version 1.8) [20]. The longest contig obtained from each culture was circularized containing an overlapping region of more than 2,000 bp at both ends of the contig. Two complete genomes were annotated through NCBI Prokaryotic Genome Annotation Pipeline [17].

Pure marine AOA culture *Ca. Nitrosopumilus* sp. strain HMK28 was isolated from Bonita beach seawater (26°19'N, 81°50'W) in Estero, Florida. Enrichment cultures *Ca. Nitrosopumilus* sp. strain HMK29 and *Ca. Cenarchaeum* sp. strain HMK20 were obtained from Pine Island seawater (26°37'N, 82°4'W) in Florida and 10 m depth seawater at Lynch Cove (47°37'N, 123°01'W) in Hood Canal, Washington, respectively. These marine AOA cultures were maintained in seawater-based SCM medium supplemented with 1 mM NH<sub>4</sub>Cl at 25°C in the dark without shaking. Cells were harvested at late-exponential or early stationary phase, and genomic DNA was extracted using the modified phenol-chloroform extraction method as described previously [21]. Extracted DNA was sequenced using Illumina HiSeq 2000 platform with PE of 150 bp. Illumina sequence reads were assembled using CLC's *de novo* assembly algorithm (version 6.0, CLC Bio, Qiagen, Germany), and the resulting contigs were curated by CodonCode Aligner (version 3.7, CodonCode Corp.) as previously described [22]. Three assembled genomes were annotated with the Rapid Annotation using Subsystem Technology (RAST) pipeline [23], and the annotations were checked and corrected by NCBI Prokaryotic Genome Annotation Pipeline [17].



A pure marine AOA culture of *Ca. Nitrosopumilus zosterae* strain NM25 was obtained from the eelgrass sediments in Tanoura Bay of Shimoda, Shizuoka, Japan [24]. Strain NM25 was maintained in HEPES buffered SCM supplemented with 1 mM NH<sub>4</sub>Cl and 100 μM α-ketoglutaric acid [13] at 30°C in the dark with slow stirring. Cells were harvested at late-exponential or early stationary phase via centrifugation, and genomic DNA was extracted from cell pellets with the ISOPLANT II kit (Nippon Gene, Tokyo, Japan) following the manufacturer's instructions. Extracted DNA was sequenced using the Illumina HiSeq1000 platform, and the trimmed reads were assembled using the GS De novo assembler version 2.6 (Roche). The assembled draft genome was annotated with the RAST pipeline [25] and deposited at DDBJ using DDBJ Fast Annotation and Submission Tool (DFAST) [26].

### **Sampling sites and metagenome-assembled genomes**

AOA MAGs ST1, ST2, and ST3 were recovered from activated sludge (AS) samples in the Shatin wastewater treatment plant (Hong Kong, China) that treats saline municipal wastewater. Briefly, three AS samples were collected on October 8<sup>th</sup>, October 10<sup>th</sup>, and October 16<sup>th</sup>, 2013 (ST-1008, ST-1010, and ST-1106). DNA was extracted from these AS samples using FastDNA Spin Kit for Soil (MP Biomedicals, USA) following the manufacturer's instruction and sequenced by Illumina HiSeq 2500 platform with PE of 125 bp and insert size of 300 bp (BGI, China). Quality-filtered (average Q value > 30) metagenome reads of ST-1008, ST-1010, and ST-1016 were co-assembled using CLC's *de novo* assembly algorithm (version 6.04, CLC Bio) with *k*-mer of 35 and contig length > 500 bp. AOA MAGs were recovered using the differential coverage binning method [27]. MAG quality was further improved by using PE tracking to

remove incorrectly binned contigs and recruit the short contigs that were filtered in the preliminary binning process. Scaffolding of contigs in MAGs was performed using SSPACE [15].

Thermophilic AOA MAG US01 was recovered from a hyperthermal hot spring in Ulu Slim, Malaysia. The spring has a near-neutral pH and a maximum temperature at 104°C. Hot spring water and sediment samples were collected in September 2011 at in situ temperature of 90°C. The collected samples were kept at ambient temperature and immediately transferred to the laboratory, where they were stored at 4°C before DNA extraction. Equal volumes of water and sediment samples were used for DNA extraction following the protocol described previously [28]. Extracted DNA was sequenced by Illumina MiSeq with PE of 100 bp and insert size of 300 bp. Quality-filtered (average Q value > 30) metagenome reads were assembled using CLC's *de novo* assembly algorithm (version 6.04, CLC Bio) with *k*-mer of 35 and contig length > 500 bp. A thermophilic AOA MAG was recovered using the differential coverage binning method [27]. Differential coverage of assembled contigs were estimated based on mapping the metagenome reads of sampling site and Sungai Klah hot spring in Malaysia (study accession number: PRJEB7059) to the scaffolds.

Marine AOA MAG YT1 was recovered from 5,000 m water depth in the Yap Trench of the western Pacific (9°52'N, 138°30'E). Briefly, 8 L of seawater was collected by a SeaBird SBE-911 plus CTD in May 2016. Water samples were filtered through a 0.22 µm mesh membrane, and filters were stored at -80°C until further processing. DNA was extracted using a MoBio PowerSoil® DNA Isolation Kit (MO BIO Laboratories, USA) according to the manufacturer's

instructions with a few modifications as previously described [29]. Extracted DNA was sequenced on the Illumina HiSeq X Ten platform with PE of 150 bp. Clean sequence reads were quality trimmed and assembled into contigs using IDBA-UD (Version 1.1.1) with the parameters: -mink 65, -maxk 145, -steps 10 [30]. The ORFs within contigs were predicted by Prodigal with the '-p meta' option (Version 2.6.3) [31]. The initial metagenome binning was performed by MetaBAT (Version 2.12.1) [32] with the modified sensitivity and specificity settings as previously described [33]. A hadopelagic marine AOA MAG was recovered using MetaBAT with the default sensitivity and specificity settings (Version 2.12.1) [32]. Metagenome binning and taxonomic assignments were performed as described elsewhere [33]. Briefly, the initial binning was performed by setting different parameters of sensitivity and specificity in MetaBAT (Version 2.12.1) [32]. Subsequently, all of the bins that retrieved from MetaBAT were pooled for post-dereplication by DAS Tool (Version 1.1) [34].

Hadopelagic marine AOA MAGs F8-1 and F8-2 were recovered from hadopelagic waters in the Challenger Deep region of the Mariana Trench (11°21'N, 142°20'E). In brief, 100 L of seawater for metagenomic analysis was collected from 8,000 m depth by a SeaBird SBE-911 plus CTD in February 2017. Water samples were filtered sequentially through a 3 µm (TSTP, 142 mm, Millipore) and a 0.22 µm (GTTP, 142 mm, Millipore) polycarbonate membrane. Filters were immediately transferred to liquid nitrogen and stored at -80°C until further processing. DNA extracts were obtained with the phenol-chloroform extraction method as previously described [35]. Extracted DNA was sequenced using the Illumina HiSeq X Ten platform with PE of 150 bp and insert size of 350 bp. Clean metagenome reads were quality trimmed and assembled independently using IDBA-UD (Version 1.1.1) with the following parameters: -mink 70, -maxk

100, -steps 10, -pre\_correction [30]. Assembled contigs with length > 10 kb were selected for metagenome binning based on the analysis of tetranucleotide frequencies, GC content, and coverage values as described elsewhere [36].

### **Genomic feature analysis**

Completeness, contamination, and coding density of AOA culture genomes and MAGs were assessed by CheckM [37]. Detailed results can be found in Table S1. Average nucleotide identity (ANI) was estimated by pyani.py (<https://github.com/widdowquinn/pyani>) using BLASTN to align genomic fragments. All predicted genes were searched against the Clusters of Orthologous Group (COG) database [38]. The conserved metabolic pathways shared by  $\geq 95\%$  of the complete or nearly complete AOA genomes were constructed based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) reference pathway map [39]. The presence and absence of each pathway enzyme involved in ten major AOA metabolic pathways, such as ammonia oxidation and electron transfer, carbon fixation, phosphorus utilization, and cobalamin biosynthesis, were assessed within available culture genomes and high-quality MAGs (close to or more than 90% completeness and close to or less than 5% contamination) [40]. Detailed results can be found in Table S4.

### **Core genome and pan-genome analyses**

AOA genome proteins were clustered into homolog cluster groups (HCGs) using OrthoMCL based on all-against-all BLASTP with the thresholds of pairwise coverage of 50% and identity of 50% [41]. Orthologs and paralogs were identified as the reciprocal best similarity pairs that were found between and within species, respectively [42]. Core genome represents the HCG genes

shared by all AOA species genomes, MAGs, and SAGs, and pan-genome represents the genes present in at least one AOA genome. To estimate core genome and pan-genome sizes, the shared/unique gene contents with the number of AOA genomes ( $N$ ) ranging from 2 to 45 were determined ( $N_{\max} = 25$  for marine AOA species;  $N_{\max} = 12$  for *Prochlorococcus* species [43];  $N_{\max} = 7$  for SAR11 species [44]). For each  $N$ , a total of  $C(45, N)$  combinations of genomes were calculated. If the number of combinations was over 5,000, 5,000 random combinations were sampled for core genome and pan-genome analyses.

### **Phylogenomic analysis**

The phylogenomic trees of AOA were constructed based on concatenated alignments of 71 conserved single-copy homologous proteins from 37 complete or nearly complete AOA cultured species genomes and 7 high-quality AOA MAGs. These marker proteins were identified based on the cluster of HCGs, and the alignments were carried out by MAFFT (version 7.221) [45]. The alignments were edited with Gblocks (version 0.91b) to identify conserved regions [46]. These protein sequences were concatenated as a single evolutionary unit. ProtTest (version 3.4) was employed to select the best-fit model of amino acid substitution according to the AIC and BIC values [47]. Subsequently, the maximum likelihood phylogenomic trees were built by RAxML (version 8.0.26) using the LG+I+G+X model on the basis of 100 bootstrap replications [48].

### **Evolution experiment and mutation analysis**

Since the isolation of *Nitrosopumilus maritimus* strain SCM1 [49], it has been continuously transferred from 2007 to 2018 in HEPES buffered SCM supplemented with 1 mM  $\text{NH}_4\text{Cl}$  at



30°C in the dark without shaking. Growth was monitored by microscopy, ammonia consumption, and nitrite accumulation as described previously [5, 13]. Late exponential or early stationary phase cultures were transferred to fresh medium (0.25% inoculum), and the purity of cultures was routinely monitored by microscopic inspection and by the absence of bacterial growth in marine broth medium as described previously [5, 13]. The specific growth rates ( $\mu$ ; h<sup>-1</sup>) were estimated by determining the slope of Ln nitrite concentrations versus time during exponential growth. The generation time (g; h) was calculated as  $g = \ln(2)/\mu$ . Respiratory activity and NO accumulation of *N. maritimus* cultures were measured with O<sub>2</sub> (Unisense AS, Denmark) and NO (amiNO-600, Innovative Instruments, Sarasota, Florida) microsensors, respectively, in custom-built 35 ml of glass vials as described previously [50].

The genome of *N. maritimus* was initially sequenced in 2007 [51]. Cultures were harvested in May 2011 and June 2016 and stored at -80°C for subsequent genome re-sequencing. Genomes of evolved cultures were sequenced to > 50× coverage on Illumina MiSeq platform using sequenced runs of 2 × 250 PE reads. Raw reads were trimmed and filtered by Trimmomatic (version 0.36) to remove low-quality reads [15]. The high-quality reads were mapped to the reference *N. maritimus* published genome using Bowtie2 (version 2.3.4) [52]. The PCR duplicates of sequence reads were identified and removed by Samtools (version 1.6) [53]. Variant calling for SNP and INDEL was searched by GATK (version 3.8) [54] and Samtools [53]. Variant results were then summarized and filtered by GATK [54].

## **Distribution and diversity of marine AOA genotypic groups and functional genes in the global ocean**

Twenty-five marine AOA species genomes were clustered in 7 genotypic groups to represent populations from distinct phylogenetic lineages and ecological habitats. To investigate the overall distribution of these genotypic groups in the upper ocean (< 1,000 m), competitive fragment recruitment was conducted to determine the relative recruitment to available marine AOA species genomes in GOS and *Tara* Oceans metagenomic databases. In addition, the vertical distribution of these genotypic groups from epi- to hadopelagic waters was determined by competitive fragment recruitment analysis using metagenomic datasets of the North Pacific HOT time-series station (125–4,000 m), Northeast Pacific Ocean (2,000 m), the Yap Trench of the western Pacific (5,000–5,700 m), and the Mariana Trench waters (2,000–8,000 m). Briefly, metagenomic sequences were searched via BLASTP (version 2.2.28+) with an E-value of  $\leq 1 \times e^{-10}$  and an identity of  $\geq 80\%$  against an in-house marine AOA species genome database [55]. The recruited amino acid sequences were further classified into seven marine AOA genotypic groups.

To calculate the relative abundance of functional genes in marine AOA natural populations, we first trimmed the raw sequencing reads of *Tara* Oceans and deep ocean metagenomic samples by Trimmomatic (version 0.36) [15]. The trimmed reads were aligned via DIAMOND (version 0.9.24) [56] to the datasets of marine AOA functional genes (including *amoA*, *amt*, *pstB*, *pit*, *mpnS*, and *ureC*) found in available species genomes, MAGs and SAGs. The average fraction of marine AOA populations that possess *pstB*, *pit*, *mpnS*, and *ureC* genes was estimated by calculating the ratio of length-normalized counts for reads that mapped to these accessory functional genes with the criterion of E-value  $\leq 1 \times e^{-10}$ , reads sequence identity  $\geq 80\%$ , and reads length  $\geq 100$  bp divided by reads that mapped to AOA *amoA* genes, assuming each marine

AOA cell contains one copy of the *amoA* gene (Table S4). The counts of metagenome reads that mapped to AOA high-affinity and low-affinity *amt* genes from environmental samples were calculated using HTSeq (version 0.9.1) and normalized by gene length [57].

To investigate the relative distribution of two homologous genes, A-type and V-type AOA *atpA* throughout the water column, we first mapped the trimmed metagenomic reads to the assembled scaffolds of metagenomic samples collected from epi- to hadopelagic waters using Bowtie2 (version 2.3.4) [52]. The metagenomic sequences of A-type and V-type AOA *atpA* genes were searched via BLASTP (version 2.2.28+) [55] with an E-value of  $\leq 1 \times e^{-10}$  and an identity of  $\geq 80\%$  against an in-house marine AOA species genome database containing *atpA* genes found in available AOA species genomes, MAGs and SAGs. The counts of metagenome reads that mapped to AOA A-*atpA* and V-*atpA* genes from environmental samples were calculated using HTSeq (version 0.9.1) and normalized by gene length [57].

To gain insight into the global diversity of marine AOA *mpnS*, *atpA*, and *pstB* genes, we compiled a collection of representative sequences of these genes from the *Tara* Oceans [58] and ALOHA [59] gene catalogs. For this purpose, we compiled *mpnS*, *atpA*, and *pstB* sequences from reference culture genomes, MAGs, and SAGs, aligned them using Clustal Omega [60] and constructed HMM profiles using HMMbuild [61]. These HMM profiles were used to screen the *Tara* Oceans and ALOHA gene catalogs using HMMsearch [61] using E-value of  $\leq 1 \times e^{-10}$ . These candidate genes were then length filtered (only genes  $> 200$  bp length passed this filter), and initial phylogenetic trees including gene sequences from reference genomes were constructed using the ETE toolkit [62] (workflow: standard\_trimmed\_fasttree). Outlier

sequences were manually filtered using the phylogenetic trees as guidance and new trees were generated using the same workflow. The phylogenetic trees combining reference sequences and environmental sequences were then visualized using iTOL [63].

**Data and materials availability:** AOA genomes and metagenomes sequence data are available in the NCBI, JGI, BIGD, or DDBJ databases, and their accession numbers are listed in Table S6. All other data products associated with this study are available from the corresponding authors upon request.

## References

1. Widderich N, Czech L, Elling FJ, Könneke M, Stoveken N, Pittelkow M, et al. Strangers in the archaeal world: osmostress-responsive biosynthesis of ectoine and hydroxyectoine by the marine thaumarchaeon *Nitrosopumilus maritimus*. *Environ Microbiol*. 2016;18:1227-1248.
2. Ngugi DK, Blom J, Alam I, Rashid M, Ba-Alawi W, Zhang GS, 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.
3. Kolp S, Pietsch M, Galinski EA, Gutschow M. Compatible solutes as protectants for zymogens against proteolysis. *Biochim Biophys Acta*. 2006;1764:1234-1242.
4. Hans Jorg K, Georg L, Erwin AG. Industrial production of the cell protectant ectoine: protection mechanisms, processes, and products. *Curr Biotechnol*. 2014;3:10-25.
5. Qin W, Amin SA, Martens-Habbena W, Walker CB, Urakawa H, Devol AH, et al. Marine ammonia-oxidizing archaeal isolates display obligate mixotrophy and wide ecotypic variation. *Proc Natl Acad Sci USA*. 2014;111:12504-12509.
6. Bayer B, Vojvoda J, Offre P, Alves RJE, Elisabeth NH, Garcia JAL, et al. Physiological and genomic characterization of two novel marine thaumarchaeal strains indicates niche differentiation. *ISME J*. 2016;10:1051-1063.
7. Alonso-Saez L, Waller AS, Mende DR, Bakker K, Farnelid H, Yager PL, et al. Role for urea in nitrification by polar marine Archaea. *Proc Natl Acad Sci USA*. 2012;109:17989-17994.
8. Kitzinger K, Padilla CC, Marchant HK, Hach PF, Herbold CW, Kidane AT, et al. Cyanate and urea are substrates for nitrification by Thaumarchaeota in the marine environment. *Nat Microbiol*. 2019;4:234-243.
9. Hallam SJ, Mincer TJ, Schleper C, Preston CM, Roberts K, Richardson PM, et al. Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. *PLoS Biol*. 2006;4:520-536.

10. Xu MN, Li X, Shi D, Zhang Y, Dai M, Huang T, et al. Coupled effect of substrate and light on assimilation and oxidation of regenerated nitrogen in the euphotic ocean. *Limnol Oceanogr.* 2019;64.
11. Santoro AE, Saito MA, Goepfert TJ, Lamborg CH, Dupont CL, DiTullio GR. Thaumarchaeal ecotype distributions across the equatorial Pacific Ocean and their potential roles in nitrification and sinking flux attenuation. *Limnol Oceanogr.* 2017;62:1984-2003.
12. Ahlgren NA, Chen YY, Needham DM, Parada AE, Sachdeva R, Trinh V, et al. Genome and epigenome of a novel marine Thaumarchaeota strain suggest viral infection, phosphorothioation DNA modification and multiple restriction systems. *Environ Microbiol.* 2017;19:2434-2452.
13. Qin W, Heal KR, Ramdasi R, Kobelt JN, Martens-Habbena W, Bertagnolli AD, et al. *Nitrosopumilus maritimus* gen. nov., sp nov., *Nitrosopumilus cobalaminigenes* sp nov., *Nitrosopumilus oxycliniae* sp nov., and *Nitrosopumilus ureiphilus* sp nov., four marine ammonia-oxidizing archaea of the phylum *Thaumarchaeota*. *Int J Syst Evol Microbiol.* 2017;67:5067-5079.
14. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods.* 2013;10:563.
15. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30:2114-2120.
16. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE.* 2014;9.
17. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 2016;44:6614-6624.
18. French E, Kozłowski JA, Mukherjee M, Bullerjahn G, Bollmann A. Ecophysiological characterization of ammonia-oxidizing archaea and bacteria from freshwater. *Appl Environ Microbiol.* 2012;78:5773-5780.
19. Cleary B, Brito IL, Huang K, Gevers D, Shea T, Young S, et al. Detection of low-abundance bacterial strains in metagenomic datasets by eigengene partitioning. *Nat Biotechnol.* 2015;33:1053.
20. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 2017;27:722-736.
21. Urakawa H, Martens-Habbena W, Stahl DA. High abundance of ammonia-oxidizing archaea in coastal waters, determined using a modified DNA extraction method. *Appl Environ Microbiol.* 2010;76:2129-2135.
22. Garcia JC, Urakawa H, Le VQ, Stein LY, Klotz MG, Nielsen JL. Draft genome sequence of *Nitrosospira* sp. strain APG3, a psychrotolerant ammonia-oxidizing bacterium isolated from sandy lake sediment. *Genome Announc.* 2013;1:e00930-00913.
23. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, et al. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* 2005;33:5691-5702.



24. Matsutani N, Nakagawa T, Nakamura K, Takahashi R, Yoshihara K, Tokuyama T. Enrichment of a novel marine ammonia-oxidizing archaeon obtained from sand of an eelgrass zone. *Microbes Environ.* 2011;26:23-29.
25. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST server: Rapid annotations using subsystems technology. *BMC Genom.* 2008;9.
26. Tanizawa Y, Fujisawa T, Nakamura Y. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics.* 2018;34:1037-1039.
27. Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol.* 2013;31:533.
28. Chan CS, Chan KG, Tay YL, Chua YH, Goh KM. Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing. *Front Microbiol.* 2015;6.
29. Zhang XX, Xu W, Liu Y, Cai MW, Luo ZH, Li M. Metagenomics reveals microbial diversity and metabolic potentials of seawater and surface sediment from a hadal biosphere at the Yap Trench. *Front Microbiol.* 2018;9.
30. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics.* 2012;28:1420-1428.
31. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics.* 2010;11.
32. Kang DWD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ.* 2015;3.
33. Liu Y, Zhou ZC, Pan J, Baker BJ, Gu JD, Li M. Comparative genomic inference suggests mixotrophic lifestyle for Thorarchaeota. *ISME J.* 2018;12:1021-1031.
34. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol.* 2018;3:836-843.
35. Liu JW, Zheng YF, Lin HY, Wang XC, Li M, Liu Y, et al. Proliferation of hydrocarbon-degrading microbes at the bottom of the Mariana Trench. *Microbiome.* 2019;7.
36. Haro-Moreno JM, Lopez-Perez M, de la Torre JR, Picazo A, Camacho A, Rodriguez-Valera F. Fine metagenomic profile of the Mediterranean stratified and mixed water columns revealed by assembly and recruitment. *Microbiome.* 2018;6.
37. 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.
38. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 2000;28:33-36.
39. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 2000;28:27-30.
40. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol.* 2017;35:725-731.

41. Li L, Stoeckert CJ, Roos DS. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* 2003;13:2178-2189.
42. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 2002;30:1575-1584.
43. Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, Rodrigue S, et al. Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet.* 2007;3:2515-2528.
44. Grote J, Thrash JC, Huggett MJ, Landry ZC, Carini P, Giovannoni SJ, et al. Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *Mbio.* 2012;3.
45. Katoh K, Standley DM. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30:772-780.
46. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 2007;56:564-577.
47. Abascal F, Zardoya R, Posada D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics.* 2005;21:2104-2105.
48. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30:1312-1313.
49. Könneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA. Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature.* 2005;437:543-546.
50. Martens-Habbena W, Qin W, Horak REA, Urakawa H, Schauer AJ, Moffett JW, et al. The production of nitric oxide by marine ammonia-oxidizing archaea and inhibition of archaeal ammonia oxidation by a nitric oxide scavenger. *Environ Microbiol.* 2015;17:2261-2274.
51. 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 USA.* 2010;107:8818-8823.
52. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9:357-U354.
53. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25:2078-2079.
54. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43:491-498.
55. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST plus : architecture and applications. *BMC Bioinformatics.* 2009;10.
56. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 2015;12:59-60.
57. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015;31:166-169.
58. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. Structure and function of the global ocean microbiome. *Science.* 2015;348.

59. Mende DR, Bryant JA, Aylward FO, Eppley JM, Nielsen T, Karl DM, et al. Environmental drivers of a microbial genomic transition zone in the ocean's interior. *Nat Microbiol.* 2017;2:1367-1373.
60. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li WZ, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 2011;7.
61. Eddy SR. Accelerated Profile HMM Searches. *PLoS Comput Biol.* 2011;7.
62. Huerta-Cepas J, Serra F, Bork P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol.* 2016;33:1635-1638.
63. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 2019;47:W256-W259.

## Supplementary Tables

**Table S1.** Summary information of AOA species genomes and MAGs as well as the genomes of other relevant microbial groups.

**Table S2.** The 71 genes shared among 44 and 65 ammonia-oxidizing thaumarchaeotal genomes that were used to construct the concatenated phylogenomic trees in Figure 2 and Figure S1, respectively.

**Table S3.** Average nucleotide identity (ANI) between two AOA genomes and fraction of genes shared between the two genomes.

**Table S4.** Distribution of major metabolic pathway genes in the analyzed AOA genomes. Numerals indicate the numbers of homologs in each genome. Darker shades of red represent an increasing number of homologs, and white represents the lack of respective COGs in the genome.

**Table S5.** Description of the non-synonymous and synonymous mutations observed during continuous culturing of *Nitrosopumilus maritimus*.

**Table S6.** The list of AOA genomes and metagenomes used in this study. Sequence data are available through accession numbers in NCBI, JGI, BIGD, or DDBJ databases.