Ancestral Reconstructions Decipher Major Adaptations of Ammonia-Oxidizing Archaea upon Radiation into Moderate Terrestrial and Marine Environments

Sophie S. Abby,[a,b] Melina Kerou,[a] Christa Schleper[a]

[a]University of Vienna, Archaea Biology and Ecogenomics Unit, Department of Functional and Evolutionary Ecology, Vienna, Austria
[b]University Grenoble Alpes, CNRS, Grenoble INP, TIMC-IMAG, Grenoble, France

Sophie S. Abby and Melina Kerou contributed equally to this work. Author order was determined alphabetically.

ABSTRACT Unlike all other archaeal lineages, ammonia-oxidizing archaea (AOA) of the phylum *Thaumarchaeota* are widespread and abundant in all moderate and oxic environments on Earth. The evolutionary adaptations that led to such unprecedented ecological success of a microbial clade characterized by highly conserved energy and carbon metabolisms have, however, remained underexplored. Here, we reconstructed the genomic content and growth temperature of the ancestor of all AOA, as well as the ancestors of the marine and soil lineages, based on 39 available complete or nearly complete genomes of AOA. Our evolutionary scenario depicts an extremely thermophilic, autotrophic, aerobic ancestor from which three independent lineages of a marine and two terrestrial groups radiated into moderate environments. Their emergence was paralleled by (i) a continuous acquisition of an extensive collection of stress tolerance genes mostly involved in redox maintenance and oxygen detoxification, (ii) an expansion of regulatory capacities in transcription and central metabolic functions, and (iii) an extended repertoire of cell appendages and modifications related to adherence and interactions with the environment. Our analysis provides insights into the evolutionary transitions and key processes that enabled the conquest of the diverse environments in which contemporary AOA are found.

KEYWORDS type IV pili, catalase, aconitase, *Archaea*, ammonia oxidizing archaea, reactive oxygen species, evolution, stress adaptation

Ammonia-oxidizing archaea of the phylum *Thaumarchaeota* have a very broad ecological distribution on Earth and are key players in the global nitrogen cycle. They are generally found in studies of aerobic environments as a stable part of the microbial community, including in soils, ocean waters, marine sediments, freshwater environments, hot springs, plants, and animals, including humans (1–3). This distribution is particularly impressive considering that even though the domain *Archaea* includes a large variety of energy and central carbon metabolisms (4), no other lineage is known so far to be represented this widely in oxic environments. Other widespread functional guilds of archaea are mostly confined to anoxic environments and include methanogenic archaea (classes I and II) and potentially also the so-far-uncultivated archaea of the novel lineages *Bathyarchaeota* and *Verstraetearchaeota*, as well as possibly some lineages of the Asgard and DPANN superphyla (5–9).

In contrast to methanogens, AOA form a monophyletic lineage now classified as *Nitrososphaeria* within the phylum *Thaumarchaeota* (10). Based on environmental studies and genomic data, this taxon comprises thermophilic species growing at around 70°C (11–13), but most cultivated organisms stem from ocean waters and soils

Address correspondence to Christa Schleper, christa.schleper@univie.ac.at.

(1–3, 14). Phylogenetic and phylogenomic studies have shown that the thermophilic group of "*Candidatus* Nitrosocaldales" from hot springs forms a sister group to all AOA adapted to lower temperatures (11–13). The latter are split into two major lineages, one of which encompasses the order *Nitrososphaerales*, with cultivated (and noncultivated) representatives residing mostly in soils (15). The second major clade is further divided into the orders "*Candidatus* Nitrosotaleales," the members of which are mostly found in acidic soils, and the large group of "*Candidatus* Nitrosopumilales," the members of which are mostly found in marine environments (16, 17). Several cultivated species of AOA have greatly contributed to a better understanding of their physiology and have revealed that AOA, unlike their bacterial counterparts, are adapted to far lower ammonia concentrations (18), which might explain their ecological success in so many oligotrophic environments. However, a recent meta-analysis based on more than 30,000 *amoA* genes deposited in public databases revealed that the cultivated strains and enrichments represent only 7 out of 19 total AOA clades, indicating that the (eco)physiological potential of AOA could be far greater than assumed (14). Nevertheless, genomes of cultivated species from diverse environments and a large number of fully or partially assembled (meta)genomes of AOA are now available and reveal a very well-conserved core of energy and carbon metabolism (19, 20), with ammonia oxidation catalyzed by an ammonia monooxygenase (AMO) and CO_2 fixed via an extremely efficient version of the 3-hydroxypropionate/4-hydroxybutyrate pathway (HP/HB) (21). This finding confirms the general assumption that AOA play an important role in the global nitrogen cycle, as ammonia oxidation represents the first step of nitrification, which is an essential process to eventually enable the conversion of reactive nitrogen species via denitrification into the inert gaseous compound N_2, which is then released to the atmosphere. Their broad occurrence and well-conserved central metabolism, as well as their monophyly, render AOA an excellent model for elucidating the evolution and adaptations of a microbial lineage that is very likely to have originated in hot springs (11, 12, 22–25) and from there successfully radiated, probably between 1 and 2.1 billion years ago (26, 27), into most if not all moderate oxic ecosystems on Earth.

In this work, we have investigated in a robust phylogenetic framework 39 complete (or nearly complete) genomes from cultivated or environmental strains of AOA that stem from ocean waters, soils, estuarine, sediments, wastewater, a marine sponge, and hot springs. This enabled us to reconstruct the optimal growth temperatures and genome contents of the last common ancestor of all ammonia-oxidizing archaea, as well as the ancestors of major AOA lineages, to reveal the paths of adaptations that gave rise to radiations into oxic marine and terrestrial environments, respectively.

RESULTS

A molecular thermometer suggests a thermophilic ancestor for AOA and subsequent parallel adaptations to mesophily. We compiled a 76-genome data set comprising publicly available complete genomes and high-quality metagenome-assembled genomes (MAGs) or single-cell assembled genomes (SAGs) (see Materials and Methods for selection criteria) from 39 AOA, 13 non-ammonia-oxidizing *Thaumarchaeota*, 8 *Aigarchaeota*, 5 *Bathyarchaeota*, and 11 representative genomes of orders within *Crenarchaeota* to serve as outgroups. We built a species tree from the concatenation of a selection of 33 protein families that were found in at least 74 out of the 76 genomes to serve as a backbone tree for our ancestral reconstructions of growth temperature and genome content (see below) (see Materials and Methods and Fig. 1 and Table S1A in the supplemental material).

Our extended tree recovered the major clades of AOA as described previously with a smaller data set (10), separating all AOA into the following four orders: "*Ca.* Nitrosopumilales" from mostly marine environments and sediments and with rather small genomes (<2 Mbp; Fig. 2), "*Ca.* Nitrosotaleales" from acidic soils, *Nitrososphaerales* with representatives exclusively from soils and sediments and exhibiting genomes of double the size those in all other groups (around 3 Mbp), and "*Ca.* Nitrosocaldales" with organisms exclusively from hot environments (Fig. 1).
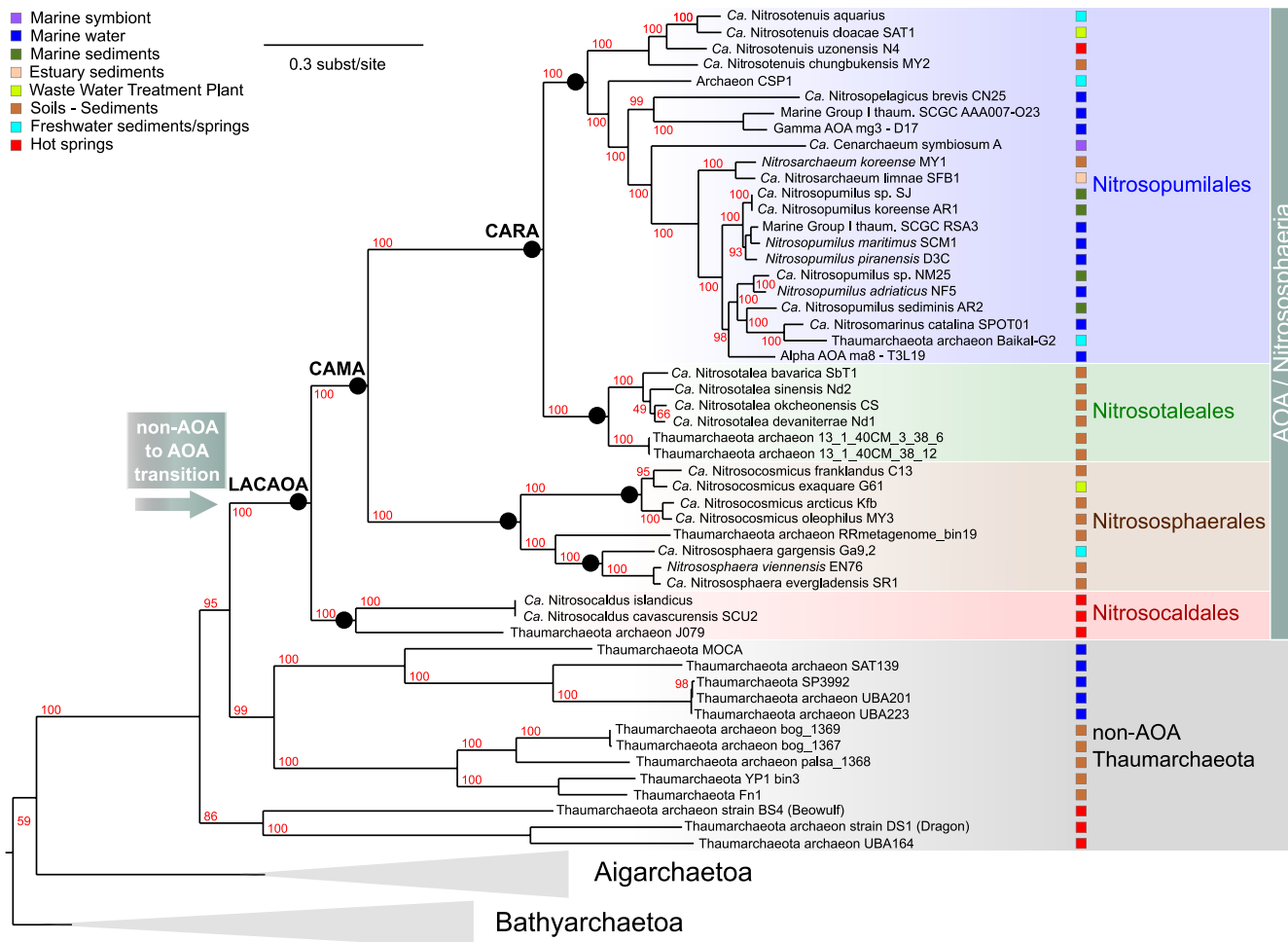
FIG 1 Phylogenetic tree (maximum likelihood) of the 76 genomes analyzed, from the concatenation of 33 conserved protein families that were present in 74 out of our 76-genome data set (see Table S1A in the supplemental material). The *Crenarchaeota* outgroup (11 genomes) is not displayed here, and *Bathyarchaeota* (5 genomes) and *Aigarchaeota* (8 genomes) are collapsed. Ammonia-oxidizing archaea (AOA) lineages and non-AOA *Thaumarchaeota* are represented by differently colored shades, and their isolation sources are displayed with colored boxes. The tree was built with IQ-TREE using the LG+C60+F model. Supports at nodes are ultrafast bootstrap supports (117).

In order to approach the optimal growth temperatures (OGTs) of the ancestor of all AOA and of ancestors of terrestrial and marine lineages, we used a "molecular thermometer" to perform a linear regression that enabled us to predict the OGTs from the predicted ancestral 16S rRNA G+C stem compositions as in (29, 30) (see Materials and Methods, Table S1C, Fig. S1, and Text S1). Not all genomes could be represented, as 20 of them were missing assigned 16S rRNA genes (Table S1A). The OGTs of AOA decreased along time, from around 76°C for the last common ancestor of AOA (LACAOA), to 67°C for the common ancestor of the mesophilic clades of AOA (CAMA), to 64°C and 52°C for the ancestors of *Nitrososphaerales* and "*Ca.* Nitrosopumilales" plus "*Ca.* Nitrosotaleales" (the ancestor of the latter two is here called "CARA," for common ancestor of rod-shaped AOA), respectively (Fig. 2). It thus seems that the ancestors of the three different major orders of AOA were all thermophilic, and parallel adaptations to lower temperatures occurred over time in these lineages. Interestingly, the order *Nitrososphaerales*, containing moderately thermophilic species like "*Ca.* Nitrososphaera gargensis" (OGT = 46°C) and *Nitrososphaera viennensis* (OGT = 42°C), seemed to have preserved a relatively high G+C content in their 16S rRNA, especially in the genus *Nitrososphaera* (predicted ancestral OGT of 64°C to 68°C), which might result from a recent or still ongoing adaptation to lower temperatures from thermophilic ancestors. Based on this analysis, "*Ca.* Nitrosopumilales" have most drastically adapted to lower
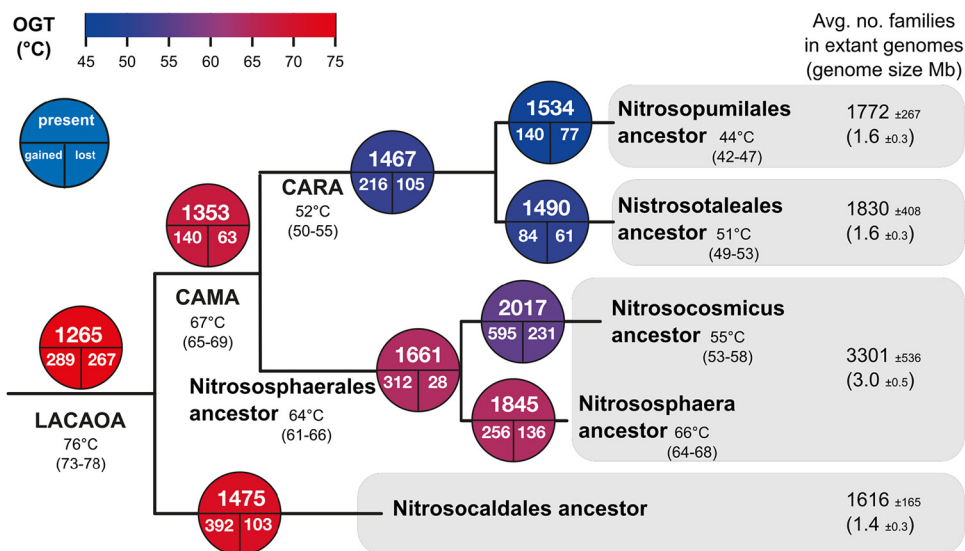
FIG 2 Reconstruction of optimal growth temperatures (OGTs) and genome repertoires of AOA ancestors. Numbers of families inferred as having a probability higher than 0.5 to be present, gained, or lost are displayed in circles at each node symbolizing the corresponding ancestral genomes and their dynamics (see main text). Ancestral OGTs are displayed at the corresponding nodes (top number) together with their confidence interval (number range below). As there were no 16S rRNA sequences available for the J079 genomic bin, we could not infer an OGT for the ancestor of "*Candidatus* Nitrosocaldales." The colors of the bubbles represent the transition from hot (red) to colder (blue) growth temperatures. For each candidate order of AOA, the number of protein families found on average in their respective genomes is indicated, as well as the average genome size (along with standard deviations). LACAOA, last common ancestor of AOA; CAMA, common ancestor of mesophilic AOA; CARA, common ancestor of rod-shaped AOA.

temperatures, which is in line with their adaptation to colder, marine environments. Since we infer an ancestor for all AOA with an OGT around 76°C (73°C to 78°C), we conclude that the ability to oxidize ammonia for energy acquisition in archaea was most probably acquired in thermophilic or even hyperthermophilic organisms, as has been suggested since their discovery and through subsequent findings of deeply branching thermophilic lineages within *Thaumarchaeota* and AOA (11–13, 22–25).

Genome dynamics and reconstruction of ancestral AOA genomes reveal distinct evolutionary paths from thermophilic to moderate AOA lineages. In order to investigate the evolutionary path of metabolic and adaptive features of AOA, we reconstructed the gene repertoire of LACAOA and of subsequent ancestors by analyzing the distribution of families of homologous proteins from our 76-genome data set and inferring their gains and losses along the species tree with a birth-and-death model for gene family size evolution implemented in the Count program (31) (see Table S2 in the supplemental material). Our analysis enabled us to reconstruct key ancestral states on the evolutionary path of AOA (Fig. 1 and 2), such as that of LACAOA; that of CAMA, the common ancestor of mesophilic AOA (after the divergence of "*Ca.* Nitrosocaldales"); that of CARA, the common ancestor of rod-shaped AOA (the common ancestor of "*Ca.* Nitrosotaleales" and the mostly marine "*Ca.* Nitrosopumilales"); and the ancestor of *Nitrososphaerales* (encompassing mostly terrestrial lineages), as well as the ancestors of the genera "*Candidatus* Nitrosocosmicus" and *Nitrososphaera*.

We infer that at least 1,265 protein families were present in LACAOA, the ancestor of AOA. Along the transition from non-AOA to AOA, 289 families were inferred to be gained and 267 lost (Fig. 1 and 2 and Tables S2 and S3 in the supplemental material). The huge metabolic transition provoked by the adoption of an ammonia-oxidizing metabolism corresponded to a slightly increased gain rate for LACAOA (i.e., number of gain events divided by branch length) compared to that of other *Thaumarchaeota* ancestors (see Fig. S2 in the supplemental material; Wilcoxon signed-rank test, *P* value ~ 0.001), yet the general gain rate trend was consistent with an overall

continuous flux of genes throughout AOA evolution rather than a massive influx of genes that would have marked the onset of this lineage (Text S1 and Fig. S2).

Despite the streamlined genomes of contemporary "*Ca.* Nitrosopumilales" (1.6 Mbp on average) compared to those of *Nitrososphaerales* (3.0 Mbp on average), their ancestors had very similar predicted family contents (1,534 and 1,661, respectively) (Fig. 2 and Table S2). Therefore, most of the differences in genome sizes can be attributed to the massive gains that occurred rather continuously within *Nitrososphaerales*, with 312 gains by the ancestor of *Nitrososphaerales* and 595 and 256 gains in the ancestors of the "*Ca.* Nitrosocosmicus" clade and *Nitrososphaera* clade, respectively. This parallel increase in the genome sizes of the two "soil" clades led to distinct genomic repertoires and physiologies (see following sections). Our inferences of ancestral genome sizes differed from those recently obtained by Hua et al. (22), resulting in apparent distinct genome dynamics with larger ancestral genomes and more downstream losses in the latter. The Dollo parsimony used in that study to infer gains and losses with Count was earlier deemed inappropriate for studying prokaryotic genomes where lateral gene transfers are common (32). Therefore, this difference in Count settings is likely to account for the differences observed between the studies. Furthermore, the probabilistic setting we used in this study, where gains and losses rates were estimated for each family separately, is likely to be more realistic.

In the following sections, we detail the evolutionary path followed by AOA toward adaptations to moderate environments, starting by describing the ancestor of all AOA. In order to confirm the evolutionary scenarios inferred by the Count program, we used comprehensive phylogenetic reconstructions for approximately 80 crucial families discussed in the rest of this study (see Table S4 in the supplemental material). This allowed us to correct—where needed—the evolutionary scenarios inferred by the program Count, and in some cases to make assumptions on the donating lineage (see Materials and Methods, Data Set S1, and Table S4).

The last common ancestor of AOA was a thermophilic, autotrophic aerobic organism. (i) Metabolic acquisitions by LACAOA. As expected, LACAOA acquired the three genes for the ammonia monooxygenase complex (*amoABC*) and the fourth candidate subunit (*amoX*) (33), the urease gene set, and two urea transporters (SSS and UT types) (12, 22, 34) and expanded its set of preexisting ammonia (Amt) transporters (from one to two, as in extant AOA) (Fig. 3 and Table S3 in the supplemental material). Although the evolutionary history of the archaeal AMO is not trivial to resolve, recent phylogenetic analyses indicated that it is more closely related to actinobacterial hydrocarbon monooxygenases than to the bacterial AMO (14). It should also be noted that none of the non-AOA thaumarchaeal genomes to date encode any genes related to ammonia oxidation or urea utilization. However, one has to note that the pathway has not been fully elucidated yet. Meanwhile, a number of cupredoxin domain families were acquired, equipping it with the unique copper-reliant biochemistry encountered in contemporary AOA. The heavy reliance on copper necessitated the acquisition of copper uptake systems, such as the CopC/CopD family proteins (35). Curiously, glutamate synthase (GOGAT), which in most archaea participates in the high-affinity ammonia assimilation pathway, is inferred to be lost at this stage, leaving the glutamate dehydrogenase (usually exhibiting lower affinity [36]) as the primary ammonia assimilation route in contemporary AOA. This could be regarded as an adaptation to an energy- and carbon-limited lifestyle, as less carbon and ATP are consumed during NH_3 assimilation. It is also the preferred strategy under these conditions in some bacteria (37).

Similarly, at least two genes for crucial steps in the 3-hydroxypropionate/4-hydroxybutyrate (HP/HB) CO_2 fixation pathway conserved in all AOA were acquired by LACAOA. These were an ADP-forming 3-hydroxypropionyl-coenzyme A (CoA) synthetase (*hpcs*), responsible for the energy efficiency of the cycle (21), possibly acquired from *Euryarchaeota* (Data Set S1), and methylmalonyl-CoA epimerase (*mce*). A third component of the pathway, the ADP-forming 4-hydroxybutyryl-CoA synthetase, was replaced at the stage of mesophilic AOA (i.e., in CAMA). LACAOA might also have newly
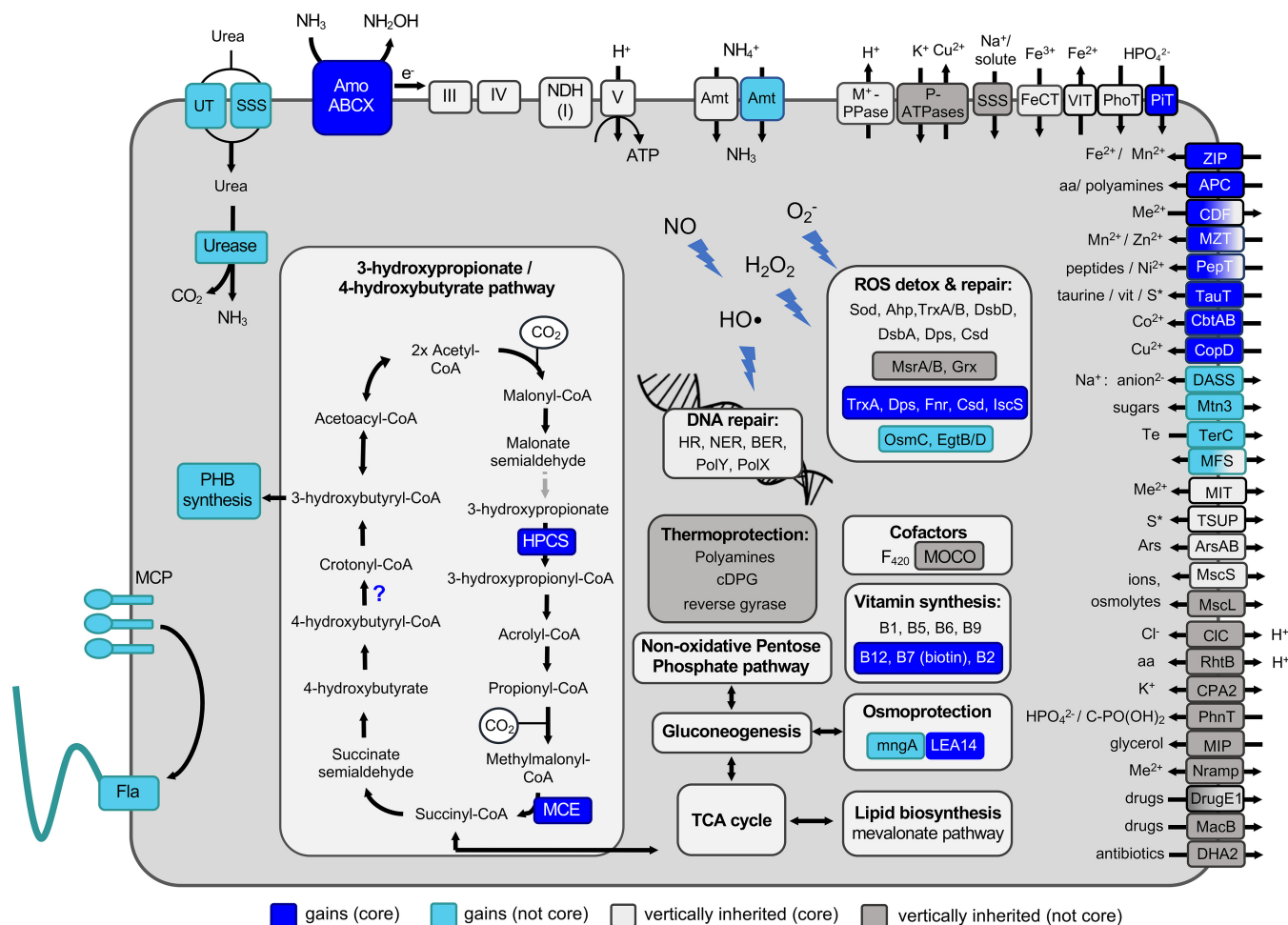
FIG 3 Reconstruction of main metabolic and physiological features and transport capabilities of the last common ancestor of AOA (LACAOA). Metabolic modules inferred to be vertically inherited are in light gray if present in all currently known extant AOA or in light blue if lost in certain lineages (not core). Newly acquired capacities that became core components for all extant AOA are depicted in blue, or in turquoise if they were more scattered throughout extant AOA genomes (not core). Gray arrows indicate reactions for which a candidate enzyme was not identified in LACAOA. Due to the inherent difficulty in annotation and substrate prediction, transporters were considered part of the core genome if found in all ancestors examined in this study or if a functionally equivalent family exchange took place in the ancestors, but their distribution in extant genomes was not taken into account. Gradient boxes are used in the case of multiple families belonging to different categories. Transporters are named according to the TCDB classification (see Table S3 in the supplemental material) (129), with indicative substrates and directionality of transport. A question mark indicates an unclear family history; see text for details. S*, organo-sulfur compounds/sulfite/sulfate; Me^{2+}, divalent cations; aa, amino acids; AcnA, AcnA/IRP group aconitase; MCP, methyl-accepting chemotaxis proteins; Fla, archaeal flagellum (or "archaellum").

acquired the O_2-tolerant enzyme 4-hydroxybutyryl-CoA dehydratase (see also Fig. S2 in reference 12). Overall, the carbon fixation pathway found in extant AOA seems to be the result of the integration of primarily vertically inherited and some more recently acquired enzymes.

The key enzyme for production of polyhydroxyalkanoates (PHAs), PHA synthase, was also acquired by LACAOA (Table S2, Table S3, and Data Set S1), equipping AOA with the ability to produce carbon storage compounds typically synthesized in response to unbalanced growth conditions (38). Additionally, LACAOA acquired the families associated with the biosynthesis of the vitamin B cofactors cobalamin (B_{12}), biotin (B_7), and riboflavin (B_2), as pointed out in earlier studies (27, 39).

(ii) Genes for aerobic autotrophic growth were already present in LACAOA. The gene families inferred to be present already and not gained by LACAOA (labeled with light and dark gray in Fig. 3) illustrate the genetic background of the archaeon which first acquired the metabolism of ammonia oxidation (Tables S2 and S3). The ancestor of LACAOA was capable of using O_2 as an electron acceptor, as suggested by the

presence of a heme-copper oxidase (routinely used to infer an aerobic metabolism [22, 40]) and the absence of an alternative electron acceptor. The inference of an aerobic ancestor stands in contrast to a recent study (27) that concluded LACAOA arose from an anaerobic ancestor, based on the assumption that all lineages basal to AOA are anaerobic (i.e., non-AOA *Thaumarchaeota*, *Aigarchaeota* and *Bathyarchaeota*). However, most *Aigarchaeota* and at least two non-AOA *Thaumarchaeota* (BS4 and pSL12) included in our data set were predicted to be facultative aerobes—the latter being recently further supported by a study reporting metagenome-assembled genomes from the pSL12 lineage with potential for aerobic respiration (22, 41–44). In our ancestral reconstructions, genes involved in anaerobic metabolisms were inferred as gains in the respective lineages, such as the *codH* subunits in the ancestor of *Bathyarchaeota* or nitrate reductase and adenylylsulfate reductase families in *Thaumarchaeota* BS4 and DS1. Additionally, the presence of most enzyme families of the HP/HB carbon fixation pathway (including the key enzymes acetyl-CoA/propionyl-CoA carboxylase and 4-hydroxybutyryl-CoA dehydratase) indicate the potential for autotrophic growth. They are present in all known AOA and shared with a number of aerobic *Crenarchaeota* and *Aigarchaeota*, albeit with distinct differences (see above and references 21, 45, and 46).

Concurrently, we observe a loss of gene families associated with heterotrophic growth, particularly those involved in glycolysis and connecting pools of C_3 and C_4 compounds. Among them are two phosphofructokinases, the 2-oxoacid dehydrogenase complex (OADHC), and an $(abcd)_2$-type 2-oxoacid:ferredoxin oxidoreductase (OFOR), while an $(ab)_2$-type OFOR exclusively found in aerobes is retained in LACAOA and is present in all extant AOA (see Text S1 and Data Set S1). Losses and gene family contractions (loss of gene copies in an ancestrally multicopy family) were also observed in families involved in amino acid catabolism, such as the glycine cleavage system and amino acid and sugar transporter families (see Text S1 for details).

Overall, our analysis depicts an autotrophic aerobic ancestor, which accompanied the switch to ammonia oxidation with an apparent minimization of the ability to grow heterotrophically. Whether this ability was completely lost in the emerging AOA lineages remains an open question, as many genomic studies have indicated the potential for mixotrophic growth and many isolates benefit from the addition of organics (although this effect was later attributed to ROS detoxification [47]). Nevertheless, assimilation of organic carbon has so far not been demonstrated in AOA.

Radiations to moderate environments were paralleled by independent adaptations in three major functional categories. In accordance with an inferred adaptation to lower temperatures, several traits involved in the thermophilic lifestyle of LACAOA were lost in subsequent evolutionary steps, namely, reverse gyrase, the hallmark enzyme of organisms living at extremely high temperatures (48, 49), and the thermoprotectant cyclic 2,3-diphosphoglycerate (cDPG) (50) synthetase were lost in CAMA, while the thermoprotectant mannosyl-3-phosphoglycerate synthase (MPGS) was lost in CARA (Tables S2 and S3 and Fig. 4). In contrast, a homologue of the "cold shock protein" CspA (51) was found to be acquired from bacteria in an ancestor of "*Ca*. Nitrosopumilales" (with the exception of the soil-residing genus "*Candidatus* Nitrosotenuis") (Fig. 4 and Fig. S3 and Table S2 in the supplemental material, corroborating the generally lower OGT of the marine lineages, as inferred by our analysis and existing pure cultures.

When surveying the functions of families acquired and lost along the evolution of mesophilic AOA, we observed that the majority of acquisitions in all examined ancestors, ranging from 18 to 34% of all functionally annotated families (and a few key losses), could be classified into the following three major categories, discussed further below: (i) adaptations to various kinds of abiotic stress, particularly oxygen-related stress, (ii) specific metabolic regulations and general increase of the regulatory potential, and (iii) extension of capacities to engage in complex interactions with the environment (Fig. 4, Fig. S3, Table S3, and Text S1).
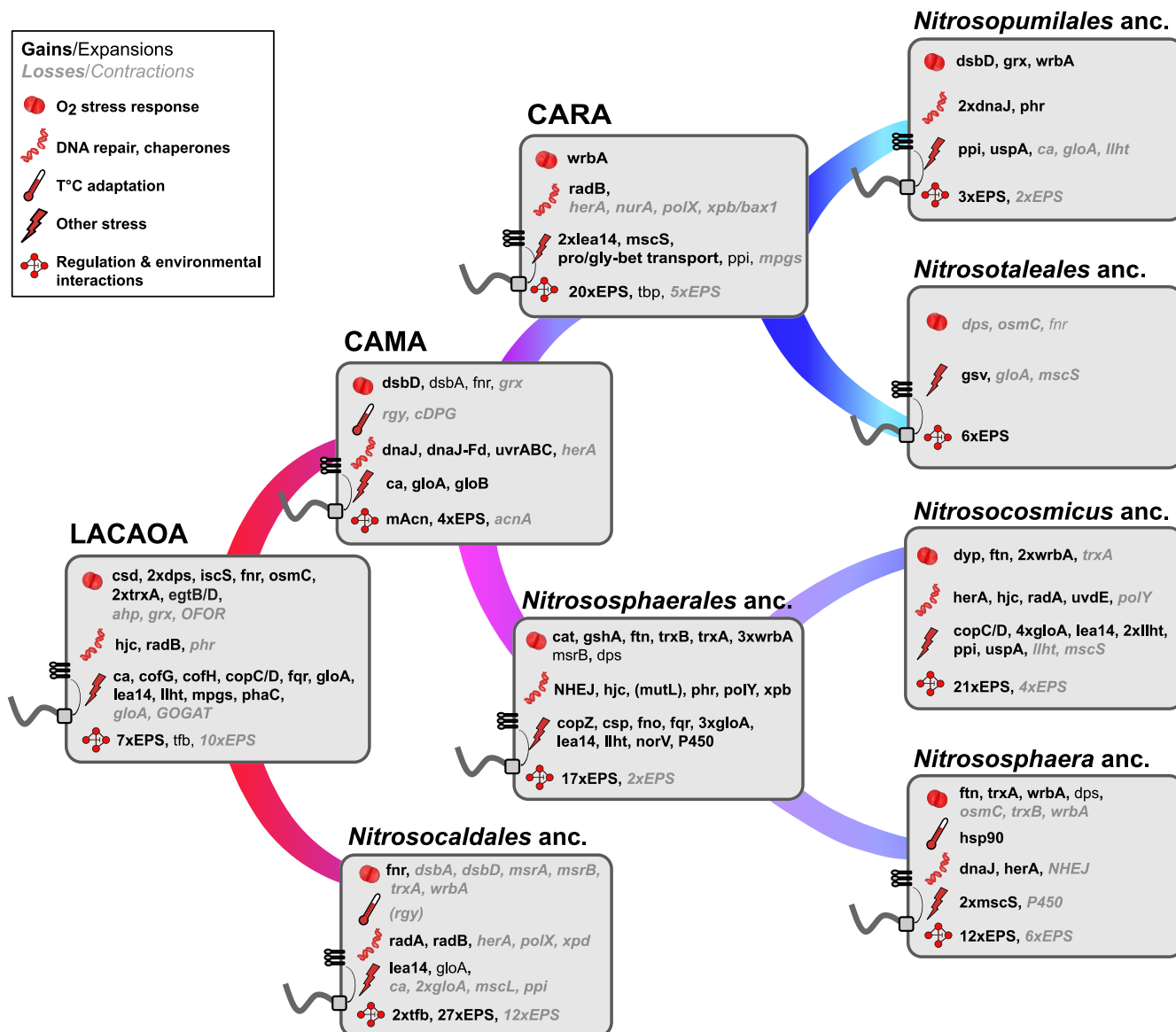
FIG 4 Summary of crucial gene gains and losses in ancestors of AOA and subsequently evolved lineages. The evolutionary events (gains/expansions in bold/normal black type, respectively, and losses/contractions in bold italic/italic gray type, respectively) related to each category are displayed next to the corresponding symbol. Ancestral cells inferred to have been flagellated harbor a schematic flagellum and chemotaxis receptors. For discussion of functions see the main text and Table S3 and Text S1 in the supplemental material). Gene names and other abbreviations: acnA, AcnA/IRP type aconitase; mAcn, mitochondrial-like aconitase; ahp, alkyl hydroperoxide reductase; ca, carbonic anhydrase; cat, catalase; cDPG, cyclic 2,3-diphosphoglycerate synthetase; cofG/H, FO synthase subunits 1/2; copZ, copper binding protein; copC/D, copper resistance family proteins; csd/iscS, cysteine desulfurase families; csp, four-helix bundle copper storage protein; dps, DNA protection during starvation family protein; dsbA/D, disulfide bond oxidoreductase A/D; dyp, dyp-type peroxidase; EPS, extracellular polymeric substances synthesis families; fqr, F420H(2)-dependent quinone reductase; fno, F420H2:NADP oxidoreductase; fnr, flavodoxin reductase (ferredoxin-NADPH reductase); fprA/norV, flavorubredoxin; ftn, ferritin-like protein; gloA/B, glyoxalase I/II; GOGAT, glutamate synthase; grx, glutaredoxin; gshA, glutamate-cysteine ligase; gsv, gas vesicle formation genes; herA, HR helicase; hjc, Holliday junction resolvase; lea14-like, LEA14-like desiccation related protein; llht, luciferase-like hydride transferase; mpgs, mannosyl-3-phosphoglycerate synthase; mscL/S, large- and small-conductance mechanosensitive channel, respectively; msrA/B, methionine sulfoxide reductase A/B; mut-L, putative DNA mismatch repair enzyme MutL-like; NHEJ, nonhomologous end joining; nurA, HR nuclease; OFOR, 2-oxoacid:ferredoxin oxidoreductase; OsmC, peroxiredoxin; P450, cytochrome p450-domain protein; phaC/E, poly(R)-hydroxyalkanoic acid synthase subunits C/E; phr, photolyase; ppi, peptidylprolyl isomerase; polY/X, DNA repair polymerases Y and X; ProP (proline/glycine-betaine):(H+/Na+) symporter; radA/B, DNA repair and recombination proteins; rgy, reverse gyrase; tbp, TATA box binding protein; tfb, transcription initiation factor TFIIB homolog; trxA, thioredoxin; trxB, thioredoxin reductase; wrbA, multimeric flavodoxin WrbA; dnaJ/dnaK/grpE/dnaJ-ferredoxin, chaperones; uspA, universal stress protein family A; uvrABC/uvdE, UV radiation repair excinuclease ABC system/UV damage endonuclease uvdE; xpb/bax1, NER helicase/nuclease pair; xpd, NER helicase.

Stress adaptations were crucial to colonize "moderate environments." (i) Strategies against oxidative and nitrosative stress. Although LACAOA, being an aerobe, was already equipped with a set of proteins dedicated to reactive oxygen species (ROS) detoxification, redox homeostasis and ROS damage repair (Fig. 3 and Tables S2 and S3),

a number of additional genes involved in these processes were frequently and continuously acquired along all lines of AOA evolution. Some of these specific acquisitions also shaped in distinct ways the genomic repertoire of entire clades, determining their ecological boundaries.

Thiol oxidoreductases of the thioredoxin (Trx) and disulfide bond reductase (Dsb) protein families, involved in reducing oxidized cysteine pairs (52) were continuously acquired throughout all AOA lineages (Fig. 4, Fig. S3, and Text S1), while the methionine sulfoxide reductase B (MsrB) family, which is responsible for the reduction of oxidized methionines, was expanded only in *Nitrososphaerales*. Thioredoxin reductase and MsrA families were present in LACAOA and were inferred to be vertically inherited. Dsb and Msr family proteins may also be involved in detoxification of reactive nitrogen species (RNS), such as NO (53, 54), produced by AOA (as in other ammonia-oxidizing bacteria) as an intermediate of ammonia oxidation (55–57).

Interestingly, only the ancestor of *Nitrososphaerales* acquired a manganese catalase, apparently from *Terrabacteria* (as suggested by phylogenetic analyses; see Table S5 and Data Set S1), which was lost in *N. viennensis*), whereas "*Ca.* Nitrosopumilales," "*Ca.* Nitrosocaldales," and "*Ca.* Nitrosotaleales" all lack this enzyme, which is considered to be a hallmark enzyme for aerobic metabolisms (Fig. 5 and Fig. S3). Although this absence remains puzzling, in bacteria catalases are the primary scavengers only at high H_2O_2 concentrations, while the activity of alkyl hydroperoxide reductase (Ahp, an enzyme present in LACAOA and vertically inherited in all extant AOA) is sufficient during logarithmic growth (58). Nevertheless, most AOA species lacking catalase have been shown to be dependent on, or stimulated by, the presence of an external H_2O_2 scavenger, such as catalase, dimethylthiourea, or α-ketoacids (47, 59), or were/are dependent on cocultures with bacteria (60, 61). Stress reduction is a known factor that shapes microbial communities (62), and these interdependencies are often the reason behind the low success rates in isolating and characterizing novel species, as is the case with AOA.

The role of low-molecular-weight thiols (LMWT) in intracellular redox control and oxidative stress defense is undisputed (63). While the most well-known LMWT, glutathione, has not been detected in archaea, they have nevertheless been shown to accumulate a variety of analogs such as thiosulfate and coenzyme A (64). The key genes (*egtB* and *egtD*) for the synthesis of the LMTW ergothioneine (EGT), a histidine betaine derivative with a thiol group synthesized by fungi and bacteria, were gained in LACAOA (Fig. 4 and Table S2). In other archaeal phyla, they are only found in a few representatives of *Methanosarcinales* and *Thermoplasmatales*. A donor could not be identified, but the origin of the genes possibly lies within the *Actinobacteria* (65).

Another notable acquisition by the ancestor of *Nitrososphaerales* is the protein family encoding a gamma-glutamylcysteine ligase (*gshA*), the evolutionary origin of which could not be unambiguously traced. This enzyme catalyzes the formation of gamma-glutamylcysteine from glutamate and cysteine, typically a precursor of glutathione biosynthesis in bacteria and eukaryotes (66). Accumulation of this compound was shown in *Haloarchaea* (67, 68), making *Nitrososphaerales* the second archaeal group to follow this strategy. The second enzyme responsible for glutathione biosynthesis, GshB, is missing in both *Haloarchaea* and *Thaumarchaeota*, while candidates for a putative bis-gamma-glutamylcystine reductase can be found among the pyridine nucleotide-disulfide reductase family homologs present in all AOA.

(ii) Iron, copper, and redox homeostasis. Systems that control the levels of unincorporated iron and copper in the cell are upregulated upon oxidative stress in archaea and bacteria (69) (Text S1). Both Fe^{2+} and Cu(I) can react with H_2O_2 and generate hydroxyl radicals (HO·) via the Fenton and Haber-Weiss reactions, respectively (70), which can subsequently cause severe cellular damage, including DNA lesions. Dps family proteins (DNA-binding proteins from starved cells), known archaeal antioxidants that consume both substrates of the Fenton reaction and physically shield the DNA (69, 71), were acquired by LACAOA and expanded in *Nitrososphaerales*. Dps and ferritin-like
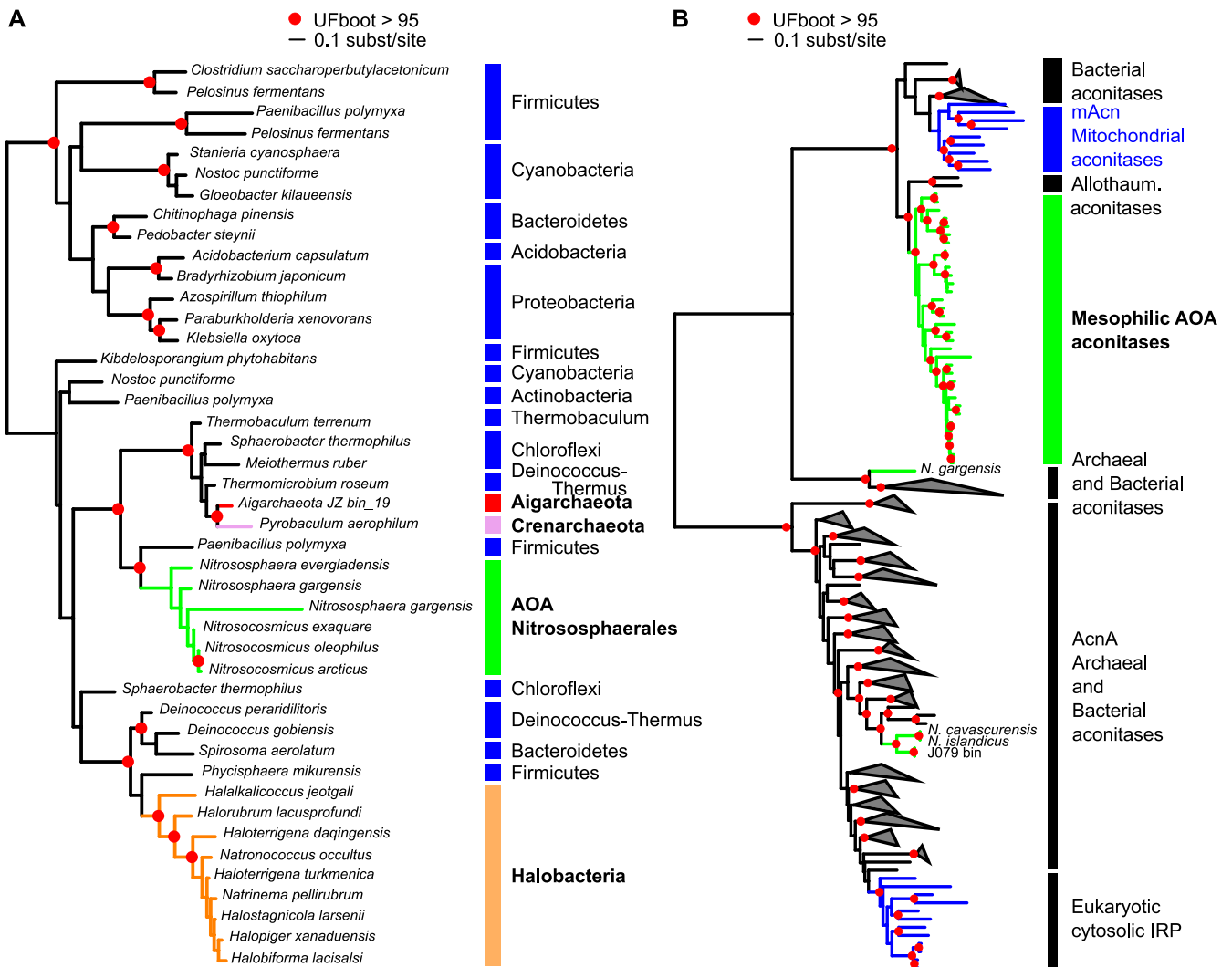
FIG 5 (A) Phylogenetic tree of the catalase family. A maximum-likelihood phylogenetic tree was obtained for the corresponding protein family, FAM002068 (Tables S2 and S3). AOA sequences are shown in green and bacterial ones are in black. Archaeal clades are indicated in bold font. (B) Phylogenetic tree of the aconitase family. A maximum-likelihood phylogenetic tree was obtained for the corresponding protein family, FAM001301. AOA sequences are shown in green and eukaryotic ones in blue. Gray triangles correspond to collapsed groups of archaeal and bacterial aconitases. In both panels, branches with UFBoot support above 95% are indicated by a red circle.

superfamily proteins were also gained as part of the parallel expansion of the soil lineages in the "*Ca.* Nitrosocosmicus" and *Nitrososphaera* ancestors (Fig. 4 and Fig. S3). Cu(I) can also cause the displacement of iron in enzymatic Fe-S clusters (72). Interestingly, a Cys-rich four-helix bundle copper storage protein (Csp), recently identified in methanotrophs (73), was acquired in the ancestor of *Nitrososphaerales*, probably from bacteria (Data Set S1).

Enzymes involved in maintaining redox homeostasis, an essential function to prevent or alleviate oxidative stress, were gained throughout the evolution of AOA. In particular, a ferredoxin:NADP(H) oxidoreductase (FNR) family was gained by LACAOA and expanded in CAMA, while WrbA flavodoxins were acquired by CARA, the ancestor of "*Ca.* Nitrosopumilales," the ancestor of *Nitrososphaerales*, and the subsequent terrestrial lineages (Fig. 4, Table S3, Fig. S3, and Text S1).

(iii) Frequent exchanges of pathways involved in desiccation, protein stability, and osmotic stress relief occurred during the colonization of osmotically variable environments. LACAOA acquired the key enzyme mannosyl-3-phosphoglycerate synthase (MPGS) (Data Set S1), responsible for the first of the two-step synthesis pathway

of the compatible solute mannosylglycerate (MG), whose role beyond osmotic stress involves maintenance of protein structure and activity during freezing, desiccation, ROS detox, and thermal denaturation (74, 75). Interestingly, the pathway for MG synthesis was lost in CARA. The emerging, mostly marine lineages instead acquired the ability to synthesize or take up ectoine/hydroxyectoine (gained by the ancestor of the *Nitrosopumilus* genus [76]) and proline/glycine-betaine (gained in CARA) during the adaptation to osmotically variable (e.g., estuarine) and high salinity (e.g., open ocean) environments (see Fig. 4, Text S1, and Fig. S3). Additional small-conductance mechanosensitive channel families (MscS) enabling the rapid outflux of solutes in response to excessive turgor caused by hypoosmotic shock (77) were acquired in CARA and subsequent lineages, while the large-conductance MscL was lost in marine lineages adapted to high salinity, as is frequently the case in marine-dwelling bacteria (77) (Table S3).

LEA14-like protein families containing a conserved WHy (water stress and hyper-sensitive response) domain (78) were acquired in LACAOA, in CARA, and repeatedly during AOA diversification, ending up with multiple copies in extant genomes (Fig. 4, Fig. S3, Tables S2 and S3, and Data Set S1), thus equipping AOA with proteins that act as "molecular shields," preventing denaturation and inactivation of cellular components under conditions of osmotic imbalance.

(iv) Gain of common and rare DNA repair systems in mesophilic AOA. Oxidative DNA lesions are hypothesized to be the primary source of DNA damage in oxic habitats, and LACAOA was already equipped with the basic components of most DNA repair systems. Interestingly, the bacterial-type UvrABC system common in mesophilic and thermophilic bacteria (79) but not found in the hyperthermophilic AOA ancestors or in non-AOA *Thaumarchaeota* was acquired by CAMA, possibly from bacteria or by multiple transfers from bacteria first to other archaeal clades (methanogens, halophiles, and the lineages comprising the TACK superphylum, i.e., Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota) and then to AOA (Data Set S1). Genetic exchanges continued to shape the DNA repair arsenal of the different thaumarchaeal lineages (Fig. 4, Fig. S3, and Text S1). In particular, an impressive inflow of DNA repair families took place into the ancestor of *Nitrososphaerales* with the acquisition of a complete nonhomologous end joining (NHEJ) complex from bacteria (Data Set S1), while a UvdE endonuclease family involved in the alternate UV damage repair pathway (UVDR) was acquired in "*Ca.* Nitrosocosmicus." The former is extremely rare in archaea (80), with a full complex described so far only from *Methanocella paludicola* (81).

Expansion of transcriptional and redox-based metabolic regulation. An intriguing enzyme family exchange took place in CAMA, the last common ancestor of all mesophilic AOA, in which the AcnA/IRP-type aconitase, present in "*Ca.* Nitrosocaldus" genomes, other archaea, the eukaryotic cytosol, and most bacteria, was exchanged for a mitochondrial-like aconitase (mAcn) present in eukaryotic mitochondria, *Bacteroides* (as previously observed [82]), and only a few other lineages of bacteria (*Spirochaetes*, *Fibrobacteres*, *Deltaproteobacteria*, and *Nitrospirae*; Fig. 5 and Data Set S1). The key to this exchange may lie in the differential sensitivity and recovery rate of these Acn types to oxidative inactivation of their Fe-S clusters. While the AcnA/IRP group aconitases are more resistant to oxidative damage and are implicated in oxidative stress responses, their repair is slower, as it might require complete cluster disassembly (83). On the other hand, the eukaryotic mAcn and bacterial AcnB aconitases, although distantly related, are easily inactivated by ROS but can be readily reactivated when iron homeostasis is restored (83, 84). This makes them the enzymes of choice for aerobic respiration in eukaryotes and some bacteria, as it provides a sophisticated redox regulation mechanism of the central carbon metabolism (85–87). Interestingly, two non-AOA *Thaumarchaeota* genome bins from oceans, likely to be from mesophilic organisms as well, also harbor this mitochondrial-like version of the aconitase, while a replacement has taken place in the moderate thermophile "*Ca.* N. gargensis" (Fig. 5). To our knowledge, this is the first observation of a mitochondrial-type aconitase in the *Archaea*, which can be

interpreted as a clearly beneficial, perhaps crucial, adaptation to an increasingly oxic environment, although this remains to be experimentally validated.

The two basic transcription factors of *Archaea* (TFB and TBP), which are homologs of the eukaryotic TBP and TFIID, are found in extended numbers in AOA (as noted earlier for smaller data sets [88, 89]) (Fig. S3). Our ancestral reconstructions, supported by phylogenetic trees, indicate that LACAOA already encoded multiple TFB copies (probably around three), and extant mesophilic AOA genomes encode 4 to 11 copies. Interestingly, TBP expansions occurred in CARA, with extant genomes encoding 2 to 5 copies (Fig. 4, Table S2, and Fig. S3). These expansions enabled multiple potential combinations of TFBs and TBPs, resulting in the emergence of global regulatory networks and rapid physiological adaptation in changing environments, as noted earlier for halophilic archaea and AOA (90–92).

Interactions with the environment: extracellular structures and cell wall modifications. A variety of enzymes involved in acetamidosugar biosynthesis, exopolysaccharide (EPS) production, cell envelope biogenesis, and adhesion were gained by every ancestor we reconstructed, with an apparent enhancement of the process in *Nitrososphaerales* (Fig. 4 and Tables S2 and S3), reflecting their demonstrated ability to form biofilms (19, 93). Formation of single- or multispecies biofilms is an understudied but very successful ecological adaptation in *Archaea* (reviewed in reference 94), as these structures not only offer protection against environmental stress and nutrient limitation (95, 96), but can also provide favorable conditions for direct nutrient or electron exchange that facilitate biogeochemical cycling.

Cellular structures play crucial roles in biofilm formation, motility, and in mediating various forms of environmental and cell-cell interactions. Our analysis indicates that at least four different types of archaeal type IV pilus (T4P) have been gained along AOA diversification (Fig. 6 and Text S1), and were sometimes subsequently lost, leading to a complex distribution pattern (Fig. S4). The AOA ancestor already possessed at least one T4P, as indicated by the families inferred to be present and acquired (T4P biosynthesis and chemotaxis genes). AOA possess the archaeal flagellum ("archaellum") (97), a pilus of unknown function related to Ups and Bas pili (involved in cell-cell contact and DNA repair and sugar metabolism, respectively, in *Sulfolobales*) (98, 99), as well as two other adhesion-like pili that could play a role in biofilm formation or biotic interactions (Fig. 6) (12, 100). Interestingly, "*Ca.* Nitrosocosmicus spp." did not harbor any T4P, unlike the soil-associated sister group *Nitrososphaera*, suggesting a different adaptive strategy to terrestrial environments, as pointed out by a number of differential gains/losses in the two lineages (see above). The precise functions of these four pili remain to be elucidated, but their abundance and diversity indicate frequent and varied environmental interactions and exchanges across AOA. Such a level of diversity in the pili repertoire of AOA even exceeds the diversity observed in *Sulfolobales*, and it is only paralleled by the diversity we find here in *Bathyarchaeota* (Fig. 6).

DISCUSSION

Our analysis highlights the many obstacles that AOA might have endured in the course of their diversification and adaptation from hot to moderate temperature environments.

Overall, a continuous acquisition of crucial components, many of which are implicated in coping with stress derived from oxygen exposure, seems to have been the key to the ecological success in the different lineages of AOA. We assume that the higher solubility of oxygen at lower temperatures (which is about 4-fold when shifting from 80°C to 20°C), together with a metabolic optimization resulting in higher production of ROS and possibly reactive nitrogen species through ammonia oxidation, might have necessitated these extensive adaptations. In summary, we observe a higher proportion of families implicated in oxygen tolerance compared to thermophilic sister lineages in our data set (*Aigarchaeota*, *Crenarchaeota*, and some non-AOA *Thaumarchaeota*) (Fig. S3 and Table S3 in the supplemental material). In addition, moderate habitats are discontinuous and unstable in terms of nutrient availability, hydration content, and
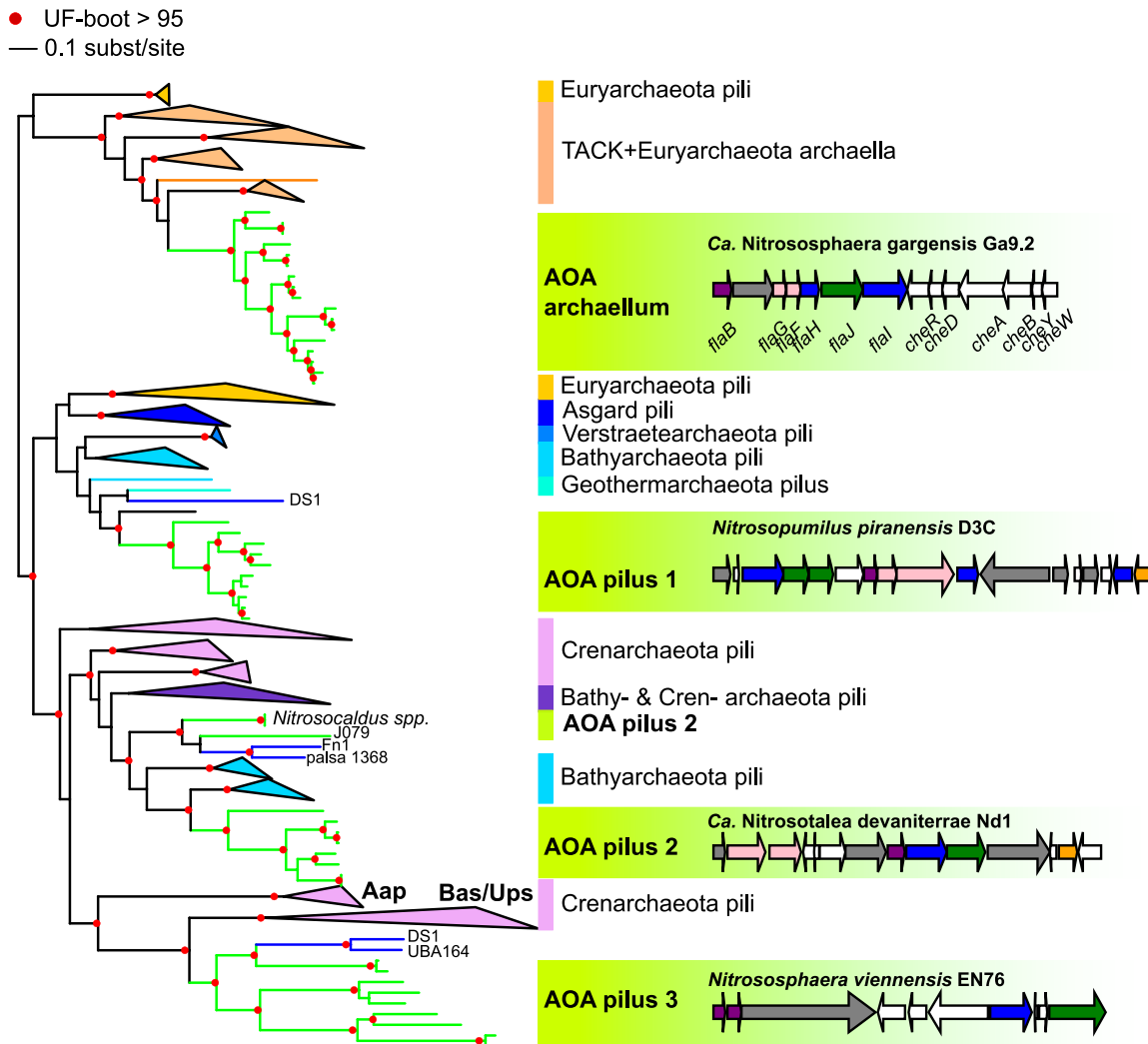
FIG 6 Phylogenetic tree of the type IV pilus ATPase family and genetic structure of AOA pili. A maximum-likelihood phylogenetic tree was obtained for the T4P ATPase in AOA (FAM001536) and representative genomes of other *Archaea*. Four distinct types of T4P are found in AOA (green branches). Clades other than AOA and non-AOA *Thaumarchaeota* (blue branches) were collapsed and colored by taxonomy. Annotations are based on the position of experimentally validated pili (archaeal flagellum/archaellum, Aap, Ups, and Bas pili); otherwise, the generic term "pilus" was used. The genetic architecture of the different types of T4P found in AOA is displayed in green boxes in front of their corresponding position in the ATPase tree. Branches with UFBoot support above 95% are indicated by a red circle. See also Fig. S4 in the supplemental material for the precise genomic distribution within AOA of these different pili.

radiation exposure, in drastically different ways than a hot spring. This is reflected in regulatory adaptations in nonthermophilic AOA, including the acquisition of the regulatable mitochondrial-type aconitase in the common mesophilic ancestor (CAMA), that is found in every mesophilic AOA to date, as well as a considerable expansion of basic transcription factors in all AOA lineages. Third, AOA had to cope increasingly with competition at lower temperatures resulting from higher overall microbial diversity and abundance, which might have triggered their diversification of the outer cell wall and the acquisition of different kinds of type IV pili in most lineages.

The specific differences in adaptations along the evolutionary lineages of marine AOA ("*Ca.* Nitrosopumilales") and the two divergently evolved terrestrial lineages ("*Ca.* Nitrosocosmicus" and *Nitrososphaera*) reflect their current major ecological distribution. While the ancestors of "*Ca.* Nitrosopumilales" acquired strategies to cope with higher osmotic pressures and lower temperatures in the ocean, the ancestor of the terrestrial group (*Nitrososphaerales*) acquired a range of families to cope with oxygen, fluctuating nutrient availability, and other stressors, as well as with DNA damage, while it diversified its cell wall modifications and EPS-forming capacities.

The path of AOA into moderate environments looks similar to that of halophilic *Archaea*, where continuous acquisitions of bacterial genes (to *Haloarchaea*) seem to have contributed to a metabolic switch from anaerobic methanogenic autotrophy to heterotrophy and aerobic respiration, as well as to osmotic adaptations (91, 101–103). These gradual adaptations contradict the earlier idea of massive lateral gene transfers at the onset of major archaeal clades (91, 101–103) (see also below). Furthermore, we see clear parallels between halophilic *Archaea* and AOA with respect to their stress adaptations, especially regarding oxidative stress (this study, and references 91, 101, and 104) and both lineages also have expanded sets of basic transcription factors, indicating the possibility of building complex regulatory networks (92).

Besides *Thaumarchaeota* and *Halobacteria*, "*Candidatus* Posidoniales" (formerly marine group II archaea) and class II methanogens (*Methanomicrobia*) thrive in oxic habitats (40, 105, 106). Although they are strictly anaerobic, the latter show particular adaptations in the methanogenesis pathway that result in decreased ROS production, together with an enrichment in protein families involved in ROS detoxification and DNA/protein repair (106), i.e., similar to adaptations outlined here for AOA but to a smaller extent.

While we agree that exposure to oxygen was a driving evolutionary pressure, we depict here a more complex scenario for adaptation of AOA to moderate environments than that described in a recent study (27). We show that adaptation to higher levels of oxygen occurred in addition to adaptations to various types of stress and that these occurred gradually, involving both innovations and preexisting ancestral sets of genes. In contrast to Ren et al. (27), we infer an already aerobic ancestor for all AOA, which is in agreement with sister lineages to AOA being at least facultative aerobes (*Bathyarchaeota*, *Aigarchaeota*, and non-AOA *Thaumarchaeota*). This is also in line with a more recent diversification of AOA than that suggested by the use of molecular clock in Ren et al. (27). The diversification of mesophilic AOA (CAMA) has indeed been estimated to begin no earlier than 950 million years based on the acquisition of the fused version of DnaJ-ferredoxin genes from the ancestor of Viridiplantae (some time between 750 and 950 million years ago) (26), an acquisition which we could confirm with our comprehensive genomic data set (Fig. 4, DnaJ-Fd). LACAOA, and at least major mesophilic lineages of AOA, might thus have diversified well after the Great Oxidation Event 2.3 billion years ago.

Although the phylogenetic trees we obtained could support Count scenarios and clarify at what point a given family was acquired, in many cases they could not help to identify the donor lineage. This is consistent with another recent study (27), and this inability to precisely identify donors persists in spite of the huge increase in taxon sampling since the first studies of lateral gene transfers in *Thaumarchaeota* (23, 107, 108). It may not only be caused by limited taxon sampling but could also result from shifts in evolutionary rates of the acquired families and the resulting difficulties in solving the gene phylogenies. It is, however, clear that a mix of preadaptations (vertical inheritance) and transfers of both bacterial and archaeal genes together fell into place to allow the successful radiation of ammonia-oxidizing archaea into so many different environments. The complexity of this "genetic cocktail" could explain why the ecological success seen for AOA appears to be unique in the domain *Archaea*.

MATERIALS AND METHODS

Genome data set. We gathered complete genomes of AOA, and enriched the data set with complete or metagenome-/single-cell-assembled genomes of high quality. We used CheckM (version 1.0.7) (109) to assess the completeness and contamination of the genomes gathered ("lineage_wf" parameter) and used a >80% completeness and <5% contamination threshold to include MAGs or SAGs. However, these criteria were sometimes relaxed in order to broaden the phylogenetic diversity included (see Table S1A in the supplemental material). In total, we selected 76 genomes for the analysis, including 39 AOA, 13 non-AOA *Thaumarchaeota*, 8 *Aigarchaeota*, 5 *Bathyarchaeota*, and 11 *Crenarchaeota* (Table S1A). When no annotations were available for the selected genomes, the Prokka suite (v1.12) (110) was used to annotate genes and proteins.

Protein family construction. Protein families were built for the set of genomes selected. A BLAST search of "all sequences against all" was performed, and the sequences representing hits with an *E* value

below 10^{-4} were used as input for the SiLiX (v1.2.11) and HiFiX (v1.0.6) programs to cluster the sets of similar sequences into families (111, 112). Two sequences were clustered in the same SiLiX family when they shared at least 35% identity and their BLAST alignment covered at least 70% of the two sequence lengths. The HiFiX program was then used on the SiLiX families for refinement. We obtained 37,517 familie, indicated throughout the text and supplementary tables as FAMXXXXXX, of which 12,367 had more than one sequence.

Phylogenomic tree inference. Universal archaeal families described by Rinke et al. (113) were annotated in our genome data set using HMMER (v3.1b2) (114) as described previously (12). The best hit was selected in each genome, and a subset of 33 informational protein families that were present in 74 out of 76 genomes of our data set was selected. The sequences from each family were extracted and aligned using MAFFT (linsi algorithm, v7.313) (115), and the alignments were filtered using BMGE (BLOSUM30, v1.12) (116). The filtered alignments were then concatenated, a maximum-likelihood tree was inferred with IQ-TREE (v1.6.11) using the LG+C60+F model of sequence evolution, and 1,000 ultrafast bootstraps were performed (117).

The 33 families, given with Pfam identifiers, were as follows: PF00177 ribosomal protein S7p/S5e, PF00189 ribosomal protein S3, PF00203 ribosomal protein S19, PF00237 ribosomal protein L22p/L17e, PF00238 ribosomal protein L14p/L23e, PF00252 ribosomal protein L16p/L10e, PF00276 ribosomal protein L23, PF00347 ribosomal protein L6, PF00366 ribosomal protein S17, PF00380 ribosomal protein S9/S16, PF00410 ribosomal protein S8, PF00411 ribosomal protein S11, PF00416 ribosomal protein S13/S18, PF00466 ribosomal protein L10, PF00572 ribosomal protein L13, PF00573 ribosomal protein L4/L1 family, PF00673 ribosomal L5P family, PF00687 ribosomal protein L1p/L10e family, PF00831 ribosomal L29 protein, PF01090 ribosomal protein S19e, PF01157 ribosomal protein L21e, PF01200 ribosomal protein S28e, PF01201 ribosomal protein S8e, PF01280 ribosomal protein L19e, PF01667 ribosomal protein S27, PF01000 RNA polymerase Rpb3/RpoA insert domain, PF01191 RNA polymerase Rpb5, C-terminal domain, PF01192 RNA polymerase Rpb6, PF01912 eIF-6 family, PF09173 initiation factor eIF2 gamma, PF03439 early transcription elongation factor of RNA pol II, PF04406 type IIB DNA topoisomerase, and PF11987 translation-initiation factor 2.

Inference of ancestral optimal growth temperature. A data set of 16S rRNA sequences and corresponding OGTs was extracted from Eme et al. (25) and from the literature (Table S1C). The stem position predictions were obtained from the RNA STRAND database (v2.0) for all the sequences in our data set that were available (20 rRNA molecules). The 16S rRNA sequences that were available (56 out of 76 genomes; see Table S1A) were all aligned together with RNA STRAND sequences with SSU-ALIGN (v0.1.1) (118), and the consensus of the positions found in stems were conservatively selected to be part of stem regions (878 positions). The GC% of the predicted stems was computed for each sequence. A linear regression between the stem GC% and the OGT was computed. A nonhomogeneous model was used in bppml (BppSuite, v2.4.1) (119) to estimate the evolutionary model of the 16S rRNA sequences along the reference tree, as in (29). The estimated parameters were then used in the program BppAncestor to reconstruct 100 replicates of ancestral sequences along the reference tree. The OGT of the ancestor of extant AOA was inferred using the linear regression coefficients and the bootstrap replicates to estimate a confidence interval, as described in Groussin and Gouy (29).

Count analysis and ancestral genomes reconstruction. A matrix of occurrences of each protein family in extant genomes was created for all the genomes under analysis. This matrix was used as an input for the Count program (v10.04) (31) along with the reference phylogeny to compute (i) the rates of gains and losses along the tree independently for each family given a phylogenetic birth-and-death model (gain-loss-duplication model, default parameters) and (ii) the posterior probabilities for each family to have been gained and lost given the inferred rates in each branch of the tree. R (v3.5.1) and Python (v2.7) scripts were used to analyze and extract the resulting data. Evolutionary events inferred on each branch of the reference tree (family gains, losses, expansions: number of family members increases; contractions: number of family members decreases) were selected when their probability was above a fixed threshold of 0.5. This threshold was also used to define the presence and multicopy state of each family within each node. The sets of families inferred to be present at a given node constituted the ancestral genome of the branches emerging from this node. We therefore could extract the sets of evolutionary events (gains, losses, expansions, and contractions) before a lineage diversification, together with the families inferred to have been present in the ancestor of the lineage (see Table S2 in the supplemental material).

Annotation of protein families and quantification of functional categories. Aside from the annotations provided with the genomes, we obtained annotations using Pfam (v28.0 PFAM-A, "cut-tc" option, best independent E-value (i-Evalue) match selected with HMMER) (120), TIGRFAM (v15, "cut-tc" option, best i-Evalue match) (121) using the corresponding HMM protein profiles and the HMMER program version 3.1b2 (114). Archaeal clusters of orthologous genes (arCOGs) (2014 version) (122) and COGs (2014 version) (123) were assigned using the COGnitor scripts, with an E-value threshold set to 10^{-10}. The MacSyFinder program (v1.0.5) (124) was also used to annotate type IV pili in genomes using models and HMM profiles described earlier (12). Manual inspection of the protein families resulted in grouping into custom functional categories, which were then quantified in extant genomes as in Table S3.

Phylogenetic analyses of gained genes. We first gathered a representative data set of 281 genomes from bacteria, archaea, and eukaryotes (see Table S1B). This data set, composed of complete genomes (NCBI RefSeq database), was selected to cover a vast range of organismal diversity and was enriched to cover the diversity of organisms from the TACK superphylum by adding 68 highly complete metagenomic bins (Table S1B).

For protein families of interest, we retrieved all the sequences classed as part of the family in our 76 genomes data set and used them to query the representative data set of genomes using BLAST (v2.6.0+) (125). The first 250 sequences corresponding to the best-score hits with an E value lower than 10^{-10} were selected. The corresponding sequences were extracted, dereplicated with UCLUST ("-cluster_fast," 100% identity level, USEARCH v10.0.240) (126), and aligned to the sequences from the original protein family using the MAFFT program (linsi algorithm, v7.397) (115). The alignments were then filtered using the BMGE program (BLOSUM30) (116). The filtered alignments were used to build phylogenetic trees by maximum-likelihood (IQ-TTREE, "-m TESTNEW," best evolutionary model selected by Bayesian information criterion, 1,000 ultrafast bootstrap and approximate likelihood ratio test [aLRT] replicates) (117). Upon manual inspection of the trees, a few sequences were excluded during protein family phylogeny reconstructions when the corresponding branches had a length in the tree of >1 substitutions/site. In those few cases, alignments and trees were reconstructed without the corresponding sequences. All resulting trees are provided in Data Set S1, and a list of the generated trees can be found in Table S4 in the supplemental material.

Generation of figures. Drawings of trees and genes were generated with iTOl (v4) (127) and GeneSpy (v1.1) (128), respectively.

Data availability. The genomes analyzed in this study are publicly available in NCBI or IMG/JGI databases, and all accession numbers are given in Table S1A. All data generated in this study are provided in Tables S1 to S4, and generated phylogenetic trees are in Data Set S1.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

TEXT S1, DOCX file, 0.1 MB.
FIG S1, PDF file, 0.1 MB.
FIG S2, PDF file, 0.5 MB.
FIG S3, PDF file, 1.5 MB.
FIG S4, PDF file, 0.5 MB.
TABLE S1, XLSX file, 0.1 MB.
TABLE S2, XLSX file, 2.7 MB.
TABLE S3, XLSX file, 1.2 MB.
TABLE S4, XLSX file, 0.01 MB.
DATA SET S1, PDF file, 2.3 MB.

ACKNOWLEDGMENTS

REFERENCES

1. Schleper C, Nicol GW. 2010. Ammonia-oxidising archaea—physiology, ecology and evolution. Adv Microb Physiol 57:1–41. https://doi.org/10.1016/B978-0-12-381045-8.00001-1.

2. Stahl DA, de la Torre JR. 2012. Physiology and diversity of ammonia-oxidizing archaea. Annu Rev Microbiol 66:83–101. https://doi.org/10.1146/annurev-micro-092611-150128.

3. Hatzenpichler R. 2012. Diversity, physiology, and niche differentiation of ammonia-oxidizing archaea. Appl Environ Microbiol 78:7501–7510. https://doi.org/10.1128/AEM.01960-12.

4. Offre P, Spang A, Schleper C. 2013. Archaea in biogeochemical cycles. Annu Rev Microbiol 67:437–457. https://doi.org/10.1146/annurev-micro-092412-155614.

5. Manoharan L, Kozlowski JA, Murdoch RW, Löffler FE, Sousa FL, Schleper C. 2019. Metagenomes from coastal marine sediments give insights into the ecological role and cellular features of *Loki*- and *Thorarchaeota*. mBio 10:e02039-19. https://doi.org/10.1128/mBio.02039-19.

6. Dombrowski N, Lee J-H, Williams TA, Offre P, Spang A. 2019. Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. FEMS Microbiol Lett366:fnz008. https://doi.org/10.1093/femsle/fnz008.

7. Kubo K, Lloyd KG, F Biddle J, Amann R, Teske A, Knittel K. 2012. Archaea of the miscellaneous crenarchaeotal group are abundant, diverse and widespread in marine sediments. ISME J 6:1949–1965. https://doi.org/10.1038/ismej.2012.37.

8. Pan J, Zhou Z, Béjà O, Cai M, Yang Y, Liu Y, Gu J-D, Li M. 2020. Genomic and transcriptomic evidence of light-sensing, porphyrin biosynthesis, Calvin-Benson-Bassham cycle, and urea production in *Bathyarchaeota*. Microbiome 8:43. https://doi.org/10.1186/s40168-020-00820-1.

9. Borrel G, Adam PS, McKay LJ, Chen L-X, Sierra-García IN, Sieber CMK, Letourneur Q, Ghozlane A, Andersen GL, Li W-J, Hallam SJ, Muyzer G, de Oliveira VM, Inskeep WP, Banfield JF, Gribaldo S. 2019. Wide diversity of methane and short-chain alkane metabolisms in uncultured archaea. Nat Microbiol 4:603–613. https://doi.org/10.1038/s41564-019-0363-3.

10. Kerou M, Alves RJE, Schleper C. 2016. *Nitrososphaeria*. *In* Bergey's manual of systematics of archaea and bacteria. John Wiley & Sons, Hoboken, NJ. https://doi.org/10.1002/9781118960608.cbm00055.

11. de la Torre JR, Walker CB, Ingalls AE, Könneke M, Stahl DA. 2008. Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. Environ Microbiol 10:810–818. https://doi.org/10.1111/j.1462-2920.2007.01506.x.

12. Abby SS, Melcher M, Kerou M, Krupovic M, Stieglmeier M, Rossel C, Pfeifer K, Schleper C. 2018. *Candidatus* Nitrosocaldus cavascurensis, an ammonia oxidizing, extremely thermophilic archaeon with a highly mobile genome. Front Microbiol 9:28. https://doi.org/10.3389/fmicb.2018.00028.

13. Daebeler A, Herbold CW, Vierheilig J, Sedlacek CJ, Pjevac P, Albertsen M, Kirkegaard RH, de la Torre JR, Daims H, Wagner M. 2018. 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 9:193. https://doi.org/10.3389/fmicb.2018.00193.

14. Alves RJE, Minh BQ, Urich T, von Haeseler A, Schleper C. 2018. Unifying the global phylogeny and environmental distribution of ammonia-oxidising archaea based on *amoA* genes. Nat Commun 9:1517. https://doi.org/10.1038/s41467-018-03861-1.

15. Kerou M, Alves RJE, Schleper C. 2018. Nitrososphaerales. *In* Bergey's manual of systematics of archaea and bacteria. John Wiley & Sons, Hoboken, NJ. https://doi.org/10.1002/9781118960608.obm00124.

16. Prosser JI, Nicol GW. 2016. *Candidatus* Nitrosotaleales. *In* Bergey's manual of systematics of archaea and bacteria. John Wiley & Sons, Hoboken, NJ. https://doi.org/10.1002/9781118960608.obm00123.

17. Qin W, Martens-Habbena W, Kobelt JN, Stahl DA. 2016. *Candidatus* Nitrosopumilales. *In* Bergey's manual of systematics of archaea and bacteria. John Wiley & Sons, Hoboken, NJ. https://doi.org/10.1002/9781118960608.obm00122.

18. Kits KD, Sedlacek CJ, Lebedeva EV, Han P, Bulaev A, Pjevac P, Daebeler A, Romano S, Albertsen M, Stein LY, Daims H, Wagner M. 2017. Kinetic analysis of a complete nitrifier reveals an oligotrophic lifestyle. Nature 549:269–272. https://doi.org/10.1038/nature23679.

19. Kerou M, Offre P, Valledor L, Abby SS, Melcher M, Nagler M, Weckwerth W, Schleper C. 2016. Proteomics and comparative genomics of *Nitrososphaera viennensis* reveal the core genome and adaptations of archaeal ammonia oxidizers. Proc Natl Acad Sci U S A 113:E7937–E7946. https://doi.org/10.1073/pnas.1601212113.

20. Herbold CW, Lehtovirta-Morley LE, Jung M-Y, Jehmlich N, Hausmann B, Han P, Loy A, Pester M, Sayavedra-Soto LA, Rhee S-K, Prosser JI, Nicol GW, Wagner M, Gubry-Rangin C. 2017. Ammonia-oxidising archaea living at low pH: insights from comparative genomics. Environ Microbiol 19:4939–4952. https://doi.org/10.1111/1462-2920.13971.

21. Könneke M, Schubert DM, Brown PC, Hügler M, Standfest S, Schwander T, Schada von Borzyskowski L, Erb TJ, Stahl DA, Berg IA. 2014. Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO$_2$ fixation. Proc Natl Acad Sci U S A 111:8239–8244. https://doi.org/10.1073/pnas.1402028111.

22. Hua Z-S, Qu Y-N, Zhu Q, Zhou E-M, Qi Y-L, Yin Y-R, Rao Y-Z, Tian Y, Li L-X, Liu L, Castelle CJ, Hedlund BP, Shu W-S, Knight R, Li W-J. 2018. Genomic inference of the metabolism and evolution of the archaeal phylum *Aigarchaeota*. Nat Commun 9:2832. https://doi.org/10.1038/s41467-018-05284-4.

23. Lopez-Garcia P, Brochier C, Moreira D, Rodriguez-Valera F. 2004. Comparative analysis of a genome fragment of an uncultivated mesopelagic crenarchaeote reveals multiple horizontal gene transfers. Environ Microbiol 6:19–34. https://doi.org/10.1046/j.1462-2920.2003.00533.x.

24. Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P. 2008. Mesophilic *Crenarchaeota*: proposal for a third archaeal phylum, the *Thaumarchaeota*. Nat Rev Microbiol 6:245–252. https://doi.org/10.1038/nrmicro1852.

25. Eme L, Reigstad LJ, Spang A, Lanzén A, Weinmaier T, Rattei T, Schleper C, Brochier-Armanet C. 2013. Metagenomics of Kamchatkan hot spring filaments reveal two new major (hyper)thermophilic lineages related to *Thaumarchaeota*. Res Microbiol 164:425–438. https://doi.org/10.1016/j.resmic.2013.02.006.

26. Petitjean C, Moreira D, López-García P, Brochier-Armanet C. 2012. Horizontal gene transfer of a chloroplast DnaJ-Fer protein to *Thaumarchaeota* and the evolutionary history of the DnaK chaperone system in *Archaea*. BMC Evol Biol 12:226. https://doi.org/10.1186/1471-2148-12-226.

27. Ren M, Feng X, Huang Y, Wang H, Hu Z, Clingenpeel S, Swan BK, Fonseca MM, Posada D, Stepanauskas R, Hollibaugh JT, Foster PG, Woyke T, Luo H. 2019. Phylogenomics suggests oxygen availability as a driving force in *Thaumarchaeota* evolution. ISME J 13:2150–2161. https://doi.org/10.1038/s41396-019-0418-8.

28. Reference deleted.

29. Groussin M, Gouy M. 2011. Adaptation to environmental temperature is a major determinant of molecular evolutionary rates in archaea. Mol Biol Evol 28:2661–2674. https://doi.org/10.1093/molbev/msr098.

30. Galtier N, Lobry JR. 1997. Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. J Mol Evol 44:632–636. https://doi.org/10.1007/pl00006186.

31. Csuros M. 2010. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. Bioinformatics 26:1910–1912. https://doi.org/10.1093/bioinformatics/btq315.

32. Rogozin IB, Wolf YI, Babenko VN, Koonin EV. 2007. Dollo parsimony and the reconstruction of genome evolution. *In* Albert VA (ed), Parsimony, phylogeny, and genomics. Oxford University Press, Oxford, UK.

33. Bartossek R, Spang A, Weidler G, Lanzen A, Schleper C. 2012. Metagenomic analysis of ammonia-oxidizing archaea affiliated with the soil group. Front Microbiol 3:208. https://doi.org/10.3389/fmicb.2012.00208.

34. Carini P, Dupont CL, Santoro AE. 2018. Patterns of thaumarchaeal gene expression in culture and diverse marine environments. Environ Microbiol 20:2112–2124. https://doi.org/10.1111/1462-2920.14107.

35. Lawton TJ, Kenney GE, Hurley JD, Rosenzweig AC. 2016. The CopC family: structural and bioinformatic insights into a diverse group of periplasmic copper binding proteins. Biochemistry 55:2278–2290. https://doi.org/10.1021/acs.biochem.6b00175.

36. Reitzer L. 2003. Nitrogen assimilation and global regulation in *Escherichia coli*. Annu Rev Microbiol 57:155–176. https://doi.org/10.1146/annurev.micro.57.030502.090820.

37. van Heeswijk WC, Westerhoff HV, Boogerd FC. 2013. Nitrogen assimilation in *Escherichia coli*: putting molecular data into a systems perspective. Microbiol Mol Biol Rev 77:628–695. https://doi.org/10.1128/MMBR.00025-13.

38. Schlegel HG, Gottschalk G, von Bartha R. 1961. Formation and utilization of poly-beta-hydroxybutyric acid by Knallgas bacteria (*Hydrogenomonas*). Nature 191:463–465. https://doi.org/10.1038/191463a0.

39. Doxey AC, Kurtz DA, Lynch MD, Sauder LA, Neufeld JD. 2015. Aquatic metagenomes implicate *Thaumarchaeota* in global cobalamin production. ISME J 9:461–471. https://doi.org/10.1038/ismej.2014.142.

40. Adam PS, Borrel G, Brochier-Armanet C, Gribaldo S. 2017. The growing tree of *Archaea*: new perspectives on their diversity, evolution and ecology. ISME J 11:2407–2419. https://doi.org/10.1038/ismej.2017.122.

41. Beam JP, Jay ZJ, Schmid MC, Rusch DB, Romine MF, Jennings R. d M, Kozubal MA, Tringe SG, Wagner M, Inskeep WP. 2016. Ecophysiology of an uncultivated lineage of *Aigarchaeota* from an oxic, hot spring filamentous 'streamer' community. ISME J 10:210–224. https://doi.org/10.1038/ismej.2015.83.

42. Takemoto H, Arai W, Takemoto K, Uchiyama I, Taniguchi T. 2015. Functional classification of uncultured "*Candidatus* caldiarchaeum subterraneum" using the Maple system. PLoS One 10:e0132994. https://doi.org/10.1371/journal.pone.0132994.

43. Aylward FO, Santoro AE. 2020. Heterotrophic *Thaumarchaea* with small genomes are widespread in the dark ocean. mSystems 5:e00415-20. https://doi.org/10.1128/mSystems.00415-20.

44. Reji L, Francis CA. 2020. Metagenome-assembled genomes reveal unique metabolic adaptations of a basal marine *Thaumarchaeota* lineage. ISME J 14:2105–2111. https://doi.org/10.1038/s41396-020-0675-6.

45. Berg IA, Kockelkorn D, Ramos-Vera WH, Say RF, Zarzycki J, Hügler M, Alber BE, Fuchs G. 2010. Autotrophic carbon fixation in archaea. Nat Rev Microbiol 8:447–460. https://doi.org/10.1038/nrmicro2365.

46. Beam JP, Jay ZJ, Kozubal MA, Inskeep WP. 2014. Niche specialization of novel *Thaumarchaeota* to oxic and hypoxic acidic geothermal springs of Yellowstone National Park. ISME J 8:938–951. https://doi.org/10.1038/ismej.2013.193.

47. Kim J-G, Park S-J, Sinninghe Damsté JS, Schouten S, Rijpstra WIC, Jung M-Y, Kim S-J, Gwak J-H, Hong H, Si O-J, Lee S, Madsen EL, Rhee S-K. 2016. Hydrogen peroxide detoxification is a key mechanism for growth of ammonia-oxidizing archaea. Proc Natl Acad Sci U S A 113:7888–7893. https://doi.org/10.1073/pnas.1605501113.

48. Bouthier de la Tour C, Portemer C, Nadal M, Stetter KO, Forterre P,

Duguet M. 1990. Reverse gyrase, a hallmark of the hyperthermophilic archaebacteria. J Bacteriol 172:6803–6808. https://doi.org/10.1128/jb .172.12.6803-6808.1990.

49. Forterre P. 2002. A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein. Trends Genet 18:236–237. https://doi.org/10.1016/S0168-9525(02)02650-1.

50. Shima S, Herault DA, Berkessel A, Thauer RK. 1998. Activation and thermostabilization effects of cyclic 2, 3-diphosphoglycerate on enzymes from the hyperthermophilic Methanopyrus kandleri. Arch Microbiol 170:469–472. https://doi.org/10.1007/s002030050669.

51. López-García P, Zivanovic Y, Deschamps P, Moreira D. 2015. Bacterial gene import and mesophilic adaptation in archaea. Nat Rev Microbiol 13:447–456. https://doi.org/10.1038/nrmicro3485.

52. Imlay JA. 2013. The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. Nat Rev Microbiol 11:443–454. https://doi.org/10.1038/nrmicro3032.

53. Lee WL, Gold B, Darby C, Brot N, Jiang X, de Carvalho LPS, Wellner D, St John G, Jacobs WR, Nathan C. 2009. Mycobacterium tuberculosis expresses methionine sulfoxide reductases A and B that protect from killing by nitrite and hypochlorite. Mol Microbiol 71:583–593. https:// doi.org/10.1111/j.1365-2958.2008.06548.x.

54. Sliskovic I, Raturi A, Mutus B. 2005. Characterization of the S-denitrosation activity of protein disulfide isomerase. J Biol Chem 280: 8733–8741. https://doi.org/10.1074/jbc.M408080200.

55. Stein LY, Klotz MG. 2016. The nitrogen cycle. Curr Biol 26:R94–R98. https://doi.org/10.1016/j.cub.2015.12.021.

56. Kozlowski JA, Stieglmeier M, Schleper C, Klotz MG, Stein LY. 2016. Pathways and key intermediates required for obligate aerobic ammonia-dependent chemolithotrophy in bacteria and Thaumarchaeota. ISME J 10:1836–1845. https://doi.org/10.1038/ismej.2016.2.

57. Caranto JD, Lancaster KM. 2017. Nitric oxide is an obligate bacterial nitrification intermediate produced by hydroxylamine oxidoreductase. Proc Natl Acad Sci U S A 114:8217–8222. https://doi.org/10.1073/pnas .1704504114.

58. Seaver LC, Imlay JA, Loewen P. 2001. Alkyl hydroperoxide reductase is the primary scavenger of endogenous hydrogen peroxide in Escherichia coli. J Bacteriol 183:7173–7181. https://doi.org/10.1128/JB.183.24 .7173-7181.2001.

59. Qin W, Amin SA, Martens-Habbena W, Walker CB, Urakawa H, Devol AH, Ingalls AE, Moffett JW, Armbrust EV, Stahl DA. 2014. Marine ammonia-oxidizing archaeal isolates display obligate mixotrophy and wide ecotypic variation. Proc Natl Acad Sci U S A 111:12504–12509. https:// doi.org/10.1073/pnas.1324115111.

60. Tourna M, Stieglmeier M, Spang A, Könneke M, Schintlmeister A, Urich T, Engel M, Schloter M, Wagner M, Richter A, Schleper C. 2011. Nitrososphaera viennensis, an ammonia oxidizing archaeon from soil. Proc Natl Acad Sci U S A 108:8420–8425. https://doi.org/10.1073/pnas .1013488108.

61. Bayer B, Pelikan C, Bittner MJ, Reinthaler T, Könneke M, Herndl GJ, Offre P. 2019. Proteomic response of three marine ammonia-oxidizing archaea to hydrogen peroxide and their metabolic interactions with a heterotrophic alphaproteobacterium. mSystems 4:e00181-19. https:// doi.org/10.1128/mSystems.00181-19.

62. Morris JJ, Johnson ZI, Szul MJ, Keller M, Zinser ER. 2011. Dependence of the cyanobacterium Prochlorococcus on hydrogen peroxide scavenging microbes for growth at the ocean's surface. PLoS One 6:e16805. https:// doi.org/10.1371/journal.pone.0016805.

63. Fahey RC. 2013. Glutathione analogs in prokaryotes. Biochim Biophys Acta 1830:3182–3198. https://doi.org/10.1016/j.bbagen.2012.10.006.

64. Rawat M, Maupin-Furlow JA. 2020. Redox and thiols in archaea. Antioxidants (Basel) 9:381. https://doi.org/10.3390/antiox9050381.

65. Jones GW, Doyle S, Fitzpatrick DA. 2014. The evolutionary history of the genes involved in the biosynthesis of the antioxidant ergothioneine. Gene 549:161–170. https://doi.org/10.1016/j.gene.2014.07 .065.

66. Copley SD, Dhillon JK. 2002. Lateral gene transfer and parallel evolution in the history of glutathione biosynthesis genes. Genome Biol 3:research0025. https://doi.org/10.1186/gb-2002-3-5-research0025.

67. Malki L, Yanku M, Borovok I, Cohen G, Mevarech M, Aharonowitz Y. 2009. Identification and characterization of gshA, a gene encoding the glutamate-cysteine ligase in the halophilic archaeon Haloferax volcanii. J Bacteriol 191:5196–5204. https://doi.org/10.1128/JB.00297-09.

68. Sundquist AR, Fahey RC. 1989. The function of gamma-glutamyl-

cysteine and bis-gamma-glutamylcystine reductase in Halobacterium halobium. J Biol Chem 264:719–725.

69. Wiedenheft B, Mosolf J, Willits D, Yeager M, Dryden KA, Young M, Douglas T. 2005. An archaeal antioxidant: characterization of a Dps-like protein from Sulfolobus solfataricus. Proc Natl Acad Sci U S A 102: 10551–10556. https://doi.org/10.1073/pnas.0501497102.

70. Doerrer LH. 2018. Cu in biology: unleashed by O2 and now irreplaceable. Inorganica Chim Acta 481:4–24. https://doi.org/10.1016/j.ica.2017 .11.051.

71. Calhoun LN, Kwon YM. 2011. Structure, function and regulation of the DNA-binding protein Dps and its role in acid and oxidative stress resistance in Escherichia coli: a review. J Appl Microbiol 110:375–386. https://doi.org/10.1111/j.1365-2672.2010.04890.x.

72. Macomber L, Imlay JA. 2009. The iron-sulfur clusters of dehydratases are primary intracellular targets of copper toxicity. Proc Natl Acad Sci U S A 106:8344–8349. https://doi.org/10.1073/pnas.0812808106.

73. Vita N, Platsaki S, Baslé A, Allen SJ, Paterson NG, Crombie AT, Murrell JC, Waldron KJ, Dennison C. 2015. A four-helix bundle stores copper for methane oxidation. Nature 525:140–143. https://doi.org/10.1038/ nature14854.

74. Empadinhas N, da Costa MS. 2011. Diversity, biological roles and biosynthetic pathways for sugar-glycerate containing compatible solutes in bacteria and archaea. Environ Microbiol 13:2056–2077. https:// doi.org/10.1111/j.1462-2920.2010.02390.x.

75. Borges N, Jorge CD, Gonçalves LG, Gonçalves S, Matias PM, Santos H. 2014. Mannosylglycerate: structural analysis of biosynthesis and evolutionary history. Extremophiles 18:835–852. https://doi.org/10.1007/ s00792-014-0661-x.

76. Widderich N, Czech L, Elling FJ, Könneke M, Stöveken N, Pittelkow M, Riclea R, Dickschat JS, Heider J, Bremer E. 2016. Strangers in the archaeal world: osmostress-responsive biosynthesis of ectoine and hydroxyectoine by the marine thaumarchaeon Nitrosopumilus maritimus. Environ Microbiol 18:1227–1248. https://doi.org/10.1111/1462-2920 .13156.

77. Booth IR, Miller S, Müller A, Lehtovirta-Morley L. 2015. The evolution of bacterial mechanosensitive channels. Cell Calcium 57:140–150. https:// doi.org/10.1016/j.ceca.2014.12.011.

78. Mertens J, Aliyu H, Cowan DA. 2018. LEA proteins and the evolution of the WHy domain. Appl Environ Microbiol 84:e00539-18. https://doi.org/ 10.1128/AEM.00539-18.

79. Morita R, Nakane S, Shimada A, Inoue M, Iino H, Wakamatsu T, Fukui K, Nakagawa N, Masui R, Kuramitsu S. 2010. Molecular mechanisms of the whole DNA repair system: a comparison of bacterial and eukaryotic systems. J Nucleic Acids 2010:179594–179532. https://doi.org/10.4061/ 2010/179594.

80. Aravind L, Koonin E. 2001. Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. Genome Res 11:1365–1374. https://doi.org/10.1101/gr.181001.

81. Bartlett EJ, Brissett NC, Doherty AJ. 2013. Ribonucleolytic resection is required for repair of strand displaced nonhomologous end-joining intermediates. Proc Natl Acad Sci U S A 110:E1984–E1991. https://doi .org/10.1073/pnas.1302616110.

82. Baughn AD, Malamy MH. 2002. A mitochondrial-like aconitase in the bacterium Bacteroides fragilis: implications for the evolution of the mitochondrial Krebs cycle. Proc Natl Acad Sci U S A 99:4662–4667. https://doi.org/10.1073/pnas.052710199.

83. Varghese S, Tang Y, Imlay JA. 2003. Contrasting sensitivities of Escherichia coli aconitases A and B to oxidation and iron depletion. J Bacteriol 185:221–230. https://doi.org/10.1128/jb.185.1.221-230.2003.

84. Gardner PR. 2002. Aconitase: sensitive target and measure of superoxide. Methods Enzymol 349:9–23. https://doi.org/10.1016/S0076 -6879(02)49317-2.

85. Walden WE. 2002. From bacteria to mitochondria: aconitase yields surprises. Proc Natl Acad Sci U S A 99:4138–4140. https://doi.org/10 .1073/pnas.082108799.

86. Tang Y, Quail MA, Artymiuk PJ, Guest JR, Green J. 2002. Escherichia coli aconitases and oxidative stress: post-transcriptional regulation of sodA expression. Microbiology (Reading) 148:1027–1037. https://doi.org/10 .1099/00221287-148-4-1027.

87. Scandroglio F, Tórtora V, Radi R, Castro L. 2014. Metabolic control analysis of mitochondrial aconitase: influence over respiration and mitochondrial superoxide and hydrogen peroxide production. Free Radic Res 48:684–693. https://doi.org/10.3109/10715762.2014.900175.

88. Walker CB, de la Torre JR, Klotz MG, Urakawa H, Pinel N, Arp DJ, Brochier-Armanet C, Chain PSG, Chan PP, Gollabgir A, Hemp J, Hügler M, Karr EA, Könneke M, Shin M, Lawton TJ, Lowe T, Martens-Habbena W, Sayavedra-Soto LA, Lang D, Sievert SM, Rosenzweig AC, Manning G, Stahl DA. 2010. *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. Proc Natl Acad Sci U S A 107:8818–8823. https://doi.org/10.1073/pnas.0913533107.

89. Spang A, Poehlein A, Offre P, Zumbrägel S, Haider S, Rychlik N, Nowka B, Schmeisser C, Lebedeva EV, Rattei T, Böhm C, Schmid M, Galushko A, Hatzenpichler R, Weinmaier T, Daniel R, Schleper C, Spieck E, Streit W, Wagner M. 2012. The genome of the ammonia-oxidizing *Candidatus* Nitrososphaera gargensis: insights into metabolic versatility and environmental adaptations. Environ Microbiol 14:3122–3145. https://doi.org/10.1111/j.1462-2920.2012.02893.x.

90. Turkarslan S, Reiss DJ, Gibbins G, Su WL, Pan M, Bare JC, Plaisier CL, Baliga NS. 2011. Niche adaptation by expansion and reprogramming of general transcription factors. Mol Syst Biol 7:554–554. https://doi.org/10.1038/msb.2011.87.

91. Becker EA, Seitzer PM, Tritt A, Larsen D, Krusor M, Yao AI, Wu D, Madern D, Eisen JA, Darling AE, Facciotti MT. 2014. Phylogenetically driven sequencing of extremely halophilic archaea reveals strategies for static and dynamic osmo-response. PLoS Genet 10:e1004784. https://doi.org/10.1371/journal.pgen.1004784.

92. Martinez-Pastor M, Tonner PD, Darnell CL, Schmid AK. 2017. Transcriptional regulation in archaea: from individual genes to global regulatory networks. Annu Rev Genet 51:143–170. https://doi.org/10.1146/annurev-genet-120116-023413.

93. Jung M-Y, et al. 2016. A hydrophobic ammonia-oxidizing archaeon of the *Nitrosocosmicus* clade isolated from coal tar-contaminated sediment. Environ Microbiol Rep 11:E7937–E7946.

94. van Wolferen M, Orell A, Albers S-V. 2018. Archaeal biofilm formation. Nat Rev Microbiol 16:699–713. https://doi.org/10.1038/s41579-018-0058-4.

95. Dang H, Lovell CR. 2016. Microbial surface colonization and biofilm development in marine environments. Microbiol Mol Biol Rev 80:91–138. https://doi.org/10.1128/MMBR.00037-15.

96. Gambino M, Cappitelli F. 2016. Mini-review: biofilm responses to oxidative stress. Biofouling 32:167–178. https://doi.org/10.1080/08927014.2015.1134515.

97. Albers S-V, Jarrell KF. 2018. The archaellum: an update on the unique archaeal motility structure. Trends Microbiol 26:351–362. https://doi.org/10.1016/j.tim.2018.01.004.

98. Ajon M, Fröls S, van Wolferen M, Stoecker K, Teichmann D, Driessen AJM, Grogan DW, Albers S-V, Schleper C. 2011. UV-inducible DNA exchange in hyperthermophilic archaea mediated by type IV pili. Mol Microbiol 82:807–817. https://doi.org/10.1111/j.1365-2958.2011.07861.x.

99. Zolghadr B, Klingl A, Rachel R, Driessen A, Albers S. 2011. The bindosome is a structural component of the *Sulfolobus solfataricus* cell envelope. Extremophiles 15:235–244. https://doi.org/10.1007/s00792-010-0353-0.

100. Makarova KS, Koonin EV, Albers SV. 2016. Diversity and evolution of type IV pili systems in archaea. Front Microbiol 7:667. https://doi.org/10.3389/fmicb.2016.00667.

101. Nelson-Sathi S, Dagan T, Landan G, Janssen A, Steel M, McInerney JO, Deppenmeier U, Martin WF. 2012. Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of *Haloarchaea*. Proc Natl Acad Sci U S A 109:20537–20542. https://doi.org/10.1073/pnas.1209119109.

102. Nelson-Sathi S, Sousa FL, Roettger M, Lozada-Chávez N, Thiergart T, Janssen A, Bryant D, Landan G, Schönheit P, Siebers B, McInerney JO, Martin WF. 2015. Origins of major archaeal clades correspond to gene acquisitions from bacteria. Nature 517:77–15. https://doi.org/10.1038/nature13805.

103. Groussin M, Boussau B, Szöllősi G, Eme L, Gouy M, Brochier-Armanet C, Daubin V. 2016. Gene acquisitions from bacteria at the origins of major archaeal clades are vastly overestimated. Mol Biol Evol 33:305–310. https://doi.org/10.1093/molbev/msv249.

104. Méheust R, Watson AK, Lapointe F-J, Papke RT, Lopez P, Bapteste E. 2018. Hundreds of novel composite genes and chimeric genes with bacterial origins contributed to haloarchaeal evolution. Genome Biol 19:1–12. https://doi.org/10.1186/s13059-018-1454-9.

105. Rinke C, Rubino F, Messer LF, Youssef N, Parks DH, Chuvochina M,

Brown M, Jeffries T, Tyson GW, Seymour JR, Hugenholtz P. 2019. A phylogenomic and ecological analysis of the globally abundant marine group II archaea (*Ca.* Poseidoniales ord. nov.). ISME J 13:663–675. https://doi.org/10.1038/s41396-018-0282-y.

106. Lyu Z, Lu Y. 2018. Metabolic shift at the class level sheds light on adaptation of methanogens to oxidative environments. ISME J 12:411–423. https://doi.org/10.1038/ismej.2017.173.

107. Brochier-Armanet C, Deschamps P, López-García P, Zivanovic Y, Rodríguez-Valera F, Moreira D. 2011. Complete-fosmid and fosmid-end sequences reveal frequent horizontal gene transfers in marine uncultured planktonic archaea. ISME J 5:1291–1302. https://doi.org/10.1038/ismej.2011.16.

108. Deschamps P, Zivanovic Y, Moreira D, Rodriguez-Valera F, López-García P. 2014. Pangenome evidence for extensive interdomain horizontal transfer affecting lineage core and shell genes in uncultured planktonic *Thaumarchaeota* and *Euryarchaeota*. Genome Biol Evol 6:1549–1563. https://doi.org/10.1093/gbe/evu127.

109. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.

110. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

111. Miele V, Penel S, Duret L. 2011. Ultra-fast sequence clustering from similarity networks with SiLiX. BMC Bioinformatics 12:116. https://doi.org/10.1186/1471-2105-12-116.

112. Miele V, Penel S, Daubin V, Picard F, Kahn D, Duret L. 2012. High-quality sequence clustering guided by network topology and multiple alignment likelihood. Bioinformatics 28:1078–1085. https://doi.org/10.1093/bioinformatics/bts098.

113. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu W-T, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T. 2013. Insights into the phylogeny and coding potential of microbial dark matter. Nature 499:431–437. https://doi.org/10.1038/nature12352.

114. Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol 7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195.

115. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066. https://doi.org/10.1093/nar/gkf436.

116. Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol 10:210. https://doi.org/10.1186/1471-2148-10-210.

117. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300.

118. Nawrocki EP. 2009. Structural RNA homology search and alignment using covariance models. PhD thesis. Washington University in St. Louis, St. Louis, MO.

119. Gueguen L, et al. 2013. Bio++: efficient extensible libraries and tools for computational molecular evolution. Mol Biol Evol 30:1745–1750. https://doi.org/10.1093/molbev/mst097.

120. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. 2014. Pfam: the protein families database. Nucleic Acids Res 42:D222–D230. https://doi.org/10.1093/nar/gkt1223.

121. Haft DH, Selengut JD, White O. 2003. The TIGRFAMs database of protein families. Nucleic Acids Res 31:371–373. https://doi.org/10.1093/nar/gkg128.

122. Makarova KS, Wolf YI, Koonin EV. 2015. Archaeal clusters of orthologous genes (arCOGs): an update and application for analysis of shared features between *Thermococcales*, *Methanococcales*, and *Methanobacteriales*. Life (Basel) 5:818–840. https://doi.org/10.3390/life5010818.

123. Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res 43:D261–D269. https://doi.org/10.1093/nar/gku1223.

124. Abby SS, Neron B, Menager H, Touchon M, Rocha EP. 2014. MacSyFinder: a program to mine genomes for molecular systems with

an application to CRISPR-Cas systems. PLoS One 9:e110726. https://doi.org/10.1371/journal.pone.0110726.

125. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389.

126. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461. https://doi.org/10.1093/bioinformatics/btq461.

127. Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23:127–128. https://doi.org/10.1093/bioinformatics/btl529.

128. Garcia PS, Jauffrit F, Grangeasse C, Brochier-Armanet C. 2019. GeneSpy, a user-friendly and flexible genomic context visualizer. Bioinformatics 35:329–331. https://doi.org/10.1093/bioinformatics/bty459.

129. Saier MH, Jr, Reddy VS, Tamang DG, Vastermark A. 2014. The transporter classification database. Nucleic Acids Res 42:D251–D258. https://doi.org/10.1093/nar/gkt1097.