## REVIEW SUMMARY

### MICROBIAL GENOMICS

# Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life

Anja Spang, Eva F. Caceres, Thijs J. G. Ettema*

**BACKGROUND:** Archaea have been recognized as a major domain of life besides Bacteria and Eukarya for about 40 years. Much of the pioneering research on archaeal organisms was dedicated to studying cellular machineries governing basal cellular processes such as transcription and translation and on elucidating physiological characteristics of the few archaeal strains that could be cultured. Although environmental surveys had started to reveal a plethora of uncultured archaeal lineages during the 1990s, in-depth knowledge about their diversity, ecology, and evolution remained limited. Recent developments in the field of metagenomics and single-cell genomics, however, have allowed the reconstruction of archaeal genomes directly from environmental samples without prior cultivation. Altogether, the use of these cultivation-independent approaches has led to the discovery of a multitude of previously unidentified archaeal lineages, some of which represent completely new branches in the archaeal tree. In this Review, we provide an overview of the currently recognized archaeal diversity, summarize new findings on the metabolic potential of recently described archaeal lineages, and discuss these data in light of archaeal evolution.

**ADVANCES:** Up until about a decade ago, all archaea were assigned to one of two major clades: the Crenarchaeota, which mostly comprise extreme thermophiles, or the Euryarchaeota, which mainly included methanogens and halophiles. Since then, our knowledge on archaeal diversity has increased rapidly. Now, Archaea include at least four major supergroups, the Euryarchaeota and the TACK, Asgard, and DPANN archaea, all of which comprise several different, potentially phylum-rank clades (see the figure). The lineages of these groups are not restricted to extreme habitats, as was once thought common for archaeal species; rather, archaea are widespread and occur in all thinkable environments on Earth, where they can make up substantial portions of the microbial biomass.

In line with their vast diversity, comparative genomics analyses reveal that Archaea are metabolically versatile and are characterized by different lifestyles. Recently discovered archaeal lineages include mesophiles and (hyper-)thermophiles, anaerobes and aerobes, autotrophs and heterotrophs, a large diversity of putative archaeal symbionts, as well as previously unknown acetogens and different groups of methanogens (see the figure). In fact, both the genetic potential to use the ancient Wood-Ljungdahl carbon fixation pathway and indications for methanogenic traits as well as the ability to anaerobically oxidize methane and other short hydrocarbons have now been found in various lineages outside the Euryarchaeota—the phylum comprising traditional methanogens and methane oxidizers. So far, little is known about the actual physiology of these archaea in their environmental niches or about their potential syntrophic relationships with other organisms, but recent findings highlight the importance and wide occurrence of these metabolic regimes in a wide diversity of archaea from anaerobic environments. Furthermore, these findings support hypotheses that suggest that all extant archaea evolved from an anaerobic autotrophic ancestor that used the Wood-Ljungdahl pathway and may have been able to obtain energy through methanogenesis.

Last, the investigation of informational processing and cellular machineries have revealed that genomes of Asgard archaea, which affiliate with eukaryotes in the tree of life (see the figure), encode proteins that they only share with eukaryotes. Excitingly, these proteins are functionally enriched for membrane bending, vesicular biogenesis, and trafficking activities, suggesting that eukaryotes evolved from an archaeal host that contained some key components that governed the emergence of eukaryotic cellular complexity after endosymbiosis. It will be exciting to determine the function of these proteins, which may be involved in species-species interactions in extant members of the Asgard archaea.

**OUTLOOK:** The wealth of novel archaeal genomic data represents a treasure trove for generating testable hypotheses on the metabolic and cellular features of these archaea and will thus help to unveil their thus far poorly characterized biology. Recent findings emphasize the importance of investigating members of the archaeal domain of life in order to obtain a more comprehensive view of microbial ecology, symbiosis, and metabolic interdependencies involving archaeal partners, and of evolution of life on Earth in regard to the deep roots of archaea as well as our microbial ancestry. ∎
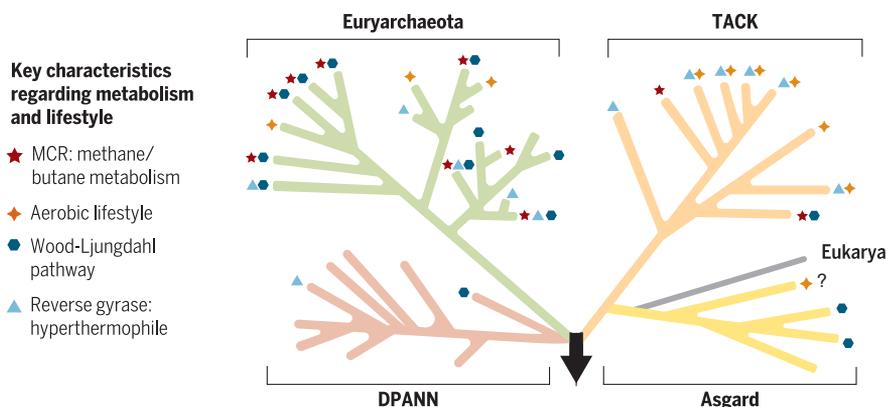
**The archaeal tree of life.** Schematic phylogenetic tree including major archaeal groups for which genomic data are currently available. The currently recognized groups include the Euryarchaeota (green branches) and the TACK, Asgard, and DPANN archaea (tan, yellow, and pink branches, respectively). Eukaryotes (gray branch) are suggested to have emerged from the Asgard archaea upon endosymbiosis with an alphaproteobacterial partner (the mitochondrial endosymbiont). For each archaeal lineage, key characteristics regarding metabolism and lifestyle are depicted.

Key characteristics regarding metabolism and lifestyle:
- ★ MCR: methane/butane metabolism
- ✦ Aerobic lifestyle
- ⬢ Wood-Ljungdahl pathway
- ▲ Reverse gyrase: hyperthermophile

## REVIEW

# Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life

Anja Spang, Eva F. Caceres, Thijs J. G. Ettema*

About 40 years ago, Archaea were recognized as a major prokaryotic domain of life besides Bacteria. Recently, cultivation-independent sequencing methods have produced a wealth of genomic data for previously unidentified archaeal lineages, several of which appear to represent newly revealed branches in the tree of life. Analyses of some recently obtained genomes have uncovered previously unknown metabolic traits and provided insights into the evolution of archaea and their relationship to eukaryotes. On the basis of our current understanding, much archaeal diversity still defies genomic exploration. Efforts to obtain and study genomes and enrichment cultures of uncultivated microbial lineages will likely further expand our knowledge about archaeal phylogenetic and metabolic diversity and their cell biology and ecological function.

This year will mark the 40th anniversary of the publication of Woese and Fox's article in which Archaea were described as a primary division of life, on a par with Bacteria and Eukarya (*1*). Before this landmark study, all life had been assigned to either of two major groups of organisms, the eukaryotes or prokaryotes. Whereas eukaryotes contain internal compartments and a nucleus, prokaryotes included organisms without intracellular structures. On the basis of comparisons of the small subunit of the ribosomal RNA of a wide range of eukaryotes and prokaryotes, Woese and Fox documented that the latter was composed of two fundamentally different types of organisms and suggested a tripartite tree of life that included three major domains: the Bacteria, Archaea, and Eukarya (*2*). However, this proposal was not immediately accepted by the scientific community. First, it clashed with the then-accepted eukaryote-prokaryote dichotomy. Second, it was based on results obtained by molecular sequence data alone, which contrasted with established approaches to infer taxonomy by using morphological traits (*3*). The construction of phylogenetic trees was still in its infancy and primarily focused on protein sequence comparisons (*4*). Last, most of the identified Archaea were sampled from extreme environments [reviewed in (*5*)], leading to the presumption that Archaea just represented an extreme group of bacteria with minor evolutionary relevance.

However, further evidence for the distinctiveness of Archaea has subsequently been obtained. For example, archaeal cell walls lack peptidoglycan and are thus fundamentally different from those of Bacteria (*6*). Furthermore, archaeal DNA-

dependent RNA polymerase was found to resemble that of eukaryotes (*7*). Also, it was found that in contrast to bacterial membranes, archaeal membranes comprised isoprene-based lipids attached to glycerol-1-phosphate via ether bonds (*8*). All these findings yielded mounting support for the uniqueness of Archaea, whose independent evolutionary status was widely accepted in 1996, when the first archaeal genome was published (*9*).

Yet, the study of the diversity and ecological relevance of the archaea was hampered by the lack of archaeal organisms in culture. This changed with the establishment of methods for the recovery of small-subunit ribosomal RNA (rRNA) sequence data directly from environmental samples, which revolutionized our understanding of microbial diversity and revealed that Archaea are present in most environments on Earth [reviewed in (*10*)]. Clearly, the previous notion that members of the Archaea are predominantly extremophiles had to be abandoned. More recently, the rise of cultivation-independent genomics approaches, such as metagenomics and single-cell genomics, has enabled the generation of genomic data from the uncultivated microbial majority (Fig. 1) (*11, 12*). Genome-resolved metagenomics has enabled the reconstruction of near complete microbial genomes (also "genomic bins") from complex environmental samples, including many previously unknown archaeal genomes (*13, 14*). Here, we review how the use of these cultivation-independent genomics approaches has expanded our knowledge on the diversity, ecology, and evolution of Archaea.

## The expanding archaeal tree

The recent discovery of many archaeal lineages has changed the shape of the phylogenetic tree of Archaea and expanded our knowledge of their diversity. For over a decade (1990–2002), Euryarchaeota and Crenarchaeota were the only recognized ar-

chaeal phyla (Fig. 2A). Therefore, the majority of 16*S* rRNA gene sequences detected in environmental surveys were initially assigned to either of these two phyla (*2*). Between 2002 and 2011, however, several new archaeal phyla were proposed on the basis of phylogenetic and genomic analyses: Korarchaeota, Nanoarchaeota, and the ammonia-oxidizing Thaumarchaeota (Fig. 2B) (*15–17*). The archaeal candidate phylum Aigarchaeota was proposed in 2011 (*18*) and comprises together with Thaum-, Cren-, and Korarchaeota the archaeal "TACK" superphylum (or "Proteoarchaeota") (Fig. 2B) (*19, 20*) . During the past 5 years, our knowledge of archaeal diversity has increased further (Fig. 2C). Now, TACK archaea comprise three additional archaeal lineages of high taxonomic rank: Geoarchaeota (*21, 22*), Bathyarchaeota (*23–25*), and Verstraetearchaeota (*26*). Furthermore, two new archaeal superphyla—the Asgard superphylum (*13, 27, 28*) as well as the DPANN superphylum—were proposed (Fig. 2C and Box 1) (*14, 29*). Currently described members of the Asgard archaea belong to the phyla Loki-, Thor-, Odin-, and Heimdallarchaeota. DPANN archaea comprise at least nine different archaeal phylum-level clades (*14*), including the previously characterized Nanoarchaeota (*16*). Yet, it is debated whether all proposed DPANN lineages are monophyletic and form a deep-branching archaeal superphylum (*20, 30*) and whether Altiarchaea are part of the DPANN (*31, 32*). Last, various lineages affiliating with Euryarchaeota have been discovered (Fig. 1) (*33*), and genomic data has now been obtained for several of these (Figs. 2C and 3).

The current taxonomy of recognized archaeal groups is inconsistent because of the lack of clear genome-based guidelines for the classification of microorganisms (*34*). Future efforts that use well-defined criteria based on phylogenomic and comparative genomic frameworks will help to establish a robust taxonomy of the archaeal domain of life.

## Metabolic potential of previously unidentified archaeal lineages

The availability of genomic data of uncultivated archaea has resulted in groundbreaking discoveries that provide a new perspective on the diversity of metabolic traits in Archaea. In particular, our knowledge on the role of archaea in the cycling of nitrogen and carbon has been extended (*35*). For instance, pioneering metagenomics studies revealed that Thaumarchaeota are key players of the global nitrogen cycle (Box 2) (*36, 37*). More recently, genome analyses of diverse uncultivated archaeal lineages (Fig. 1) [reviewed in (*38*)] have provided a first glimpse into the metabolic diversity of archaea and revealed among others the presence of genes involved in methanogenesis and/or anaerobic oxidation of methane (AOM) and the Wood-Ljungdahl carbon fixation pathway (WLP) (Box 2).

### ANME and Syntrophoarchaea

ANME lineages (Box 2) grow on methane through reverse methanogenesis and comprise at least three distinct paraphyletic lineages (*39*). They have been found in a wide range of anoxic habitats—including

Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, SE-75123 Uppsala, Sweden.
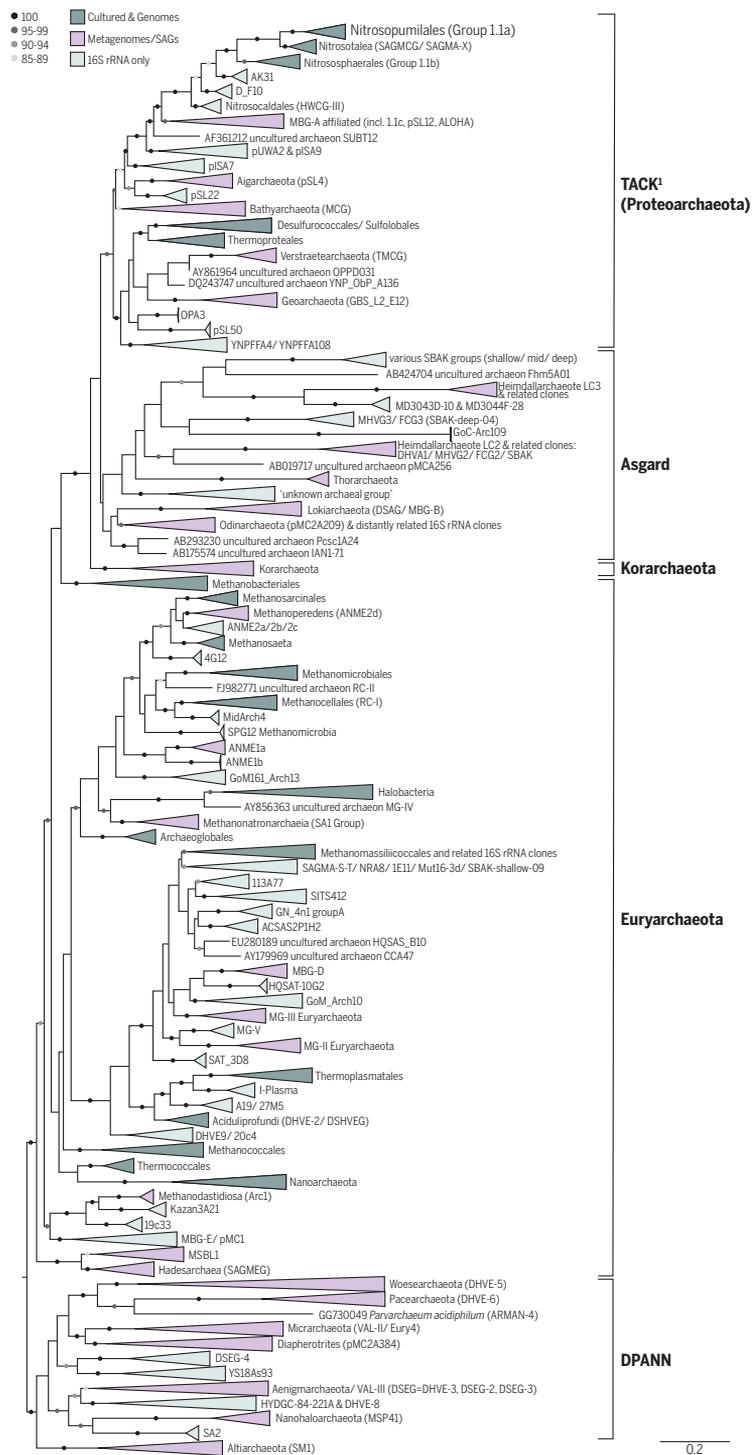*Corresponding author. Email: thijs.ettema@icm.uu.se

**Fig. 1. 16S rRNA tree revealing the archaeal diversity.** What archaea are out there? This phylogenetic tree based on an updated 16S rRNA gene data set (*33*), shows the currently recognized archaeal diversity and emphasizes our ignorance regarding the physiology, ecology, and evolution of the archaeal domain of life. Whereas genomes (shaded in dark teal) and metagenomes (shaded in pink) have recently been obtained for many of these archaeal groups, various archaeal lineages are only represented based on 16S rRNA sequence data (light teal). Cultured representatives only exist for very few archaeal groups, and information regarding metabolic and physiological properties is also sparse. Sequences were retrieved from the National Center for Biotechnology Information nr database, aligned with MAFFT-LINSi (*96*), trimmed with trimal (*97*), and analyzed by using IQ-TREE (*98*). Branch support values above 85 are indicated by dots. Scale bar indicates the number of substitutions per site.

methane seeps, hydrothermal vents, and marine water columns—but are particularly widespread in the sulfate-methane transition zone (SMTZ), which marks the transition between upper sulfate-rich and deeper methane-rich sediment layers. Because AOM yields very little energy, ANME often form syntrophic consortia with bacterial partners that serve as electron sinks and allow the coupling of methane oxidation to the reduction of sulfate (*40*), nitrate (*41*), or manganese and iron (*42*). Although the coupling of AOM to sulfate reduction may be mediated by direct electron transfer (*43*, *44*), members of the ANME-2d (such as *Methanoperedens*) transfer nitrite to a bacterial partner during AOM coupled to nitrate reduction (*45*). Whereas ANME are slow-growing and challenging to cultivate, cultivation-independent approaches have provided insights into their metabolic potential (*46–49*). The enrichment and subsequent genome sequencing of organisms closely related to ANME1 archaea, referred to as *Syntrophoarchaea* (Fig. 3), has shown that these archaea oxidize butane rather than methane (*50*). Butane oxidation is enabled through the channeling of electrons to sulfate-reducing partner bacteria and involves genes coding for the key enzyme of methanogenesis/methane oxidation (butyl coenzyme M reductase) (Fig. 4 and Box 2). This indicates that coenzyme M reductases may have a broader substrate spectrum than was assumed previously and that the mere presence of *mcr* genes in genomes of uncultivated organisms is not sufficient to infer its ability to engage in methane transformations.

### Thermoplasmatales-related lineages

The Thermoplasmatales are characterized by an acidophilic and thermophilic lifestyle and chemolitho- or chemoorganoheterotrophic growth. Environmental surveys have revealed various archaeal groups distantly related to Thermoplasmatales (Fig. 1) and together comprising the Thermoplasmata. These previously unidentified members occur in a wide range of environments, including acid mine drainage systems, anoxic marine sediments, black smokers, and hydrothermal vent fields, but also in intestinal tracts of animals and the marine water column, where they sometimes represent abundant community members (*33*). In agreement with their taxonomic diversity and widespread distribution, the various Thermoplasmata sublineages represent a metabolically versatile group of Archaea that includes both anaerobic heterotrophs, organisms with the ability to respire oxygen, and previously unknown methanogens (Fig. 3). Recently identified Thermoplasmata lineages include the aerobic (photo-)heterotrophic Thalassoarchaea and MG-III archaea from oceanic waters (*51*, *52*), the anaerobic heterotrophic MBG-D archaea from sediment environments (*53*, *54*), and the Methanomassiliicoccales (Fig. 3) (*55*). In contrast to other Thermoplasmata, members of the Methanomassiliicoccales encode the key-enzyme for methanogenesis (Fig. 4) and generate methane from the hydrogen-dependent conversion of methanol, methylated amines, and dimethyl sulfide (*55*, *56*).
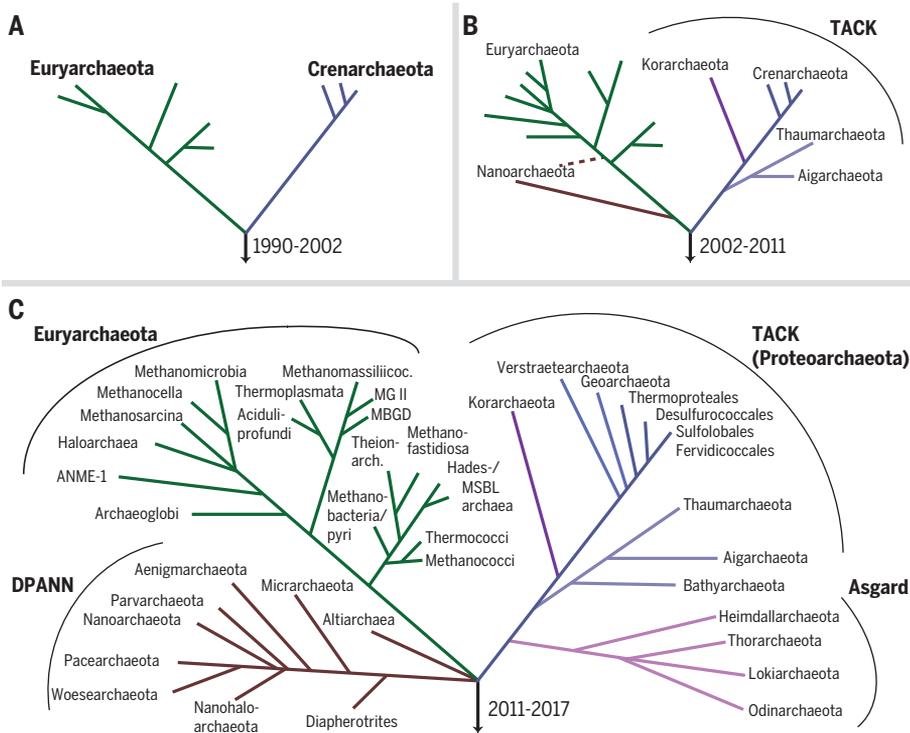
**Fig. 2. The expanding archaeal diversity.** Schematic depiction of how the shape of the archaeal tree of life has expanded over the years, revealing the major impact of cultivation-independent genomics during the past 5 years. (**A**) Between 1990 and 2002, only two archaeal phyla were known (Euryarchaeota and Crenarchaeota). (**B**) Additional phyla were identified between 2002 and 2011, but the phylum-status of Nanoarchaeota was still controversial. (**C**) Since 2011, various additional archaeal lineages of high taxonomic rank have been discovered, propelled by advances in sequencing technologies and the use of metagenomic and single-cell genomic techniques.

### Deeply branching Euryarchaeota

Several recent studies have reported genomic data from deeply branching Euryarchaeota. One of these, the Hadesarchaea, represents a group of Archaea previously referred to as SAGMEG (Box 1) that thrives in subsurface environments, including marine and terrestrial sediments (33). Reconstructed genomes of these archaea revealed that they may be capable of anaerobic CO oxidation but lack key genes for methanogenesis and the WLP (Fig. 3) (57). MSBL1 (Mediterranean Sea Brine Lakes 1) archaea are closely related to Hadesarchaea, and single-cell genomes from brine pools of the Red Sea indicate that these archaea are sugar-fermenters capable of carbon fixation via the WLP or the reductive tricarboxylic acid cycle (58). Similar to Hadesarchaea, the MSBL1 archaea lack genes for methanogenesis (Fig. 3) (58).

MSBL1 and Hadesarchaea are sister groups to WSA2/Arc1, a candidate methanogen clade referred to as Methanofastidiosa that is distributed in sediments, contaminated groundwater, and bioreactors (Fig. 3) (59). Metabolic reconstructions from draft genomes of eight Methanofastidiosa lineages provided support for their potential for methanogenesis through methylated thiol reduction. However, these archaea lack key genes for the WLP and have the ability to use more conventional sub-strates (Box 2), indicating a previously unknown but restricted methanogenic metabolism (Fig. 3) (59).

Last, Lazar et al. described the discovery of the Theionarchaea (Fig. 3), which appear to be widespread in various anaerobic sediments and subsurface environments (54) and are characterized by the presence of the WLP and the absence of genes for the Mcr complex (Fig. 3). Altogether, these findings highlight the patchy distribution of genes for methanogenesis and the WLP in different archaeal lineages (Fig. 3) (56).

### Haloarchaea

Halobacteria (referred to as Haloarchaea hereafter) represent an archaeal clade characterized by extreme salt tolerance that has been suggested to be derived from methanogenic ancestors (60, 61). Massive amounts of gene influx from bacterial sources may have led to their drastic change in lifestyle (61) [but see also (62)]. Despite that most Haloarchaea are heterotrophs and have the ability to respire oxygen (Fig. 3), some members of this group display a facultative anaerobic lifestyle and can use alternative electron acceptors (such as nitrate or dimethyl sulfoxide) during anaerobic respiration or switch to fermentative growth [(63) and references therein]. The discovery of a strictly anaerobic haloarchaeon growing on acetate or pyruvate with the concomitant reduction of sulfur (63)

indicates that within Haloarchaea, metabolic diversity may be higher than presumed before.

Furthermore, enrichment cultures and genomes of SA1 archaea revealed that this group represents methanogenic halophilic archaea that seem to branch basally to Haloarchaea (Fig. 1) (64). These archaea, referred to as Methanonatronarchaeia, have a methyl-reducing heterotrophic methanogenic lifestyle similar to that of Methanomassiliicoccales (64).

### Previously unidentified members of the TACK superphylum

Marine subsurface environments are often dominated by members of Bathyarchaeota (formerly MCG), which comprises an extremely diverse archaeal clade (>17 different sublineages) that forms a sister-group to Thaum- and Aigarchaeota (Fig. 3) (25, 65, 66). Single-cell genomes (53) and genomic bins (23–25) from different subgroups of this phylum have revealed a highly variable metabolic potential. Interestingly, most members of this phylum appear to encode proteins with a potential role in acetate production using the archaeal version of the WLP (24, 25). Some subgroups appear able to ferment organic substrates and perform homoacetogenesis, which was previously thought to be restricted to Bacteria (24). Furthermore, analyses of genomic data indicates that some of these Archaea have the ability to grow heterotrophically on substrates such as peptides, cellulose, chitin, aromatic compounds (24, 25, 53, 66), and fatty acids (23) and therefore may be able to switch between heterotrophic and autotrophic lifestyles. Bathyarchaeota genomes reconstructed from a metagenome from coal-bed methane wells revealed the presence of the WLP pathway as well as genes encoding Mcr subunits and genes for methanogenesis from methyl sulfides, methanol, and methylated amines suggestive of methylotrophic methanogenesis (23). However, it remains to be determined whether these Bathyarchaeota are indeed capable of methanogensis or AOM, or have the ability to anaerobically oxidize alternative substrates such as butane in Syntrophoarchaea (50).

Verstraetearchaeota (formerly TMCG) (Fig. 1), whose members occur in anoxic environments with high methane fluxes, comprises a recently described archaeal phylum affiliating with the TACK superphylum (Fig. 3) (26). Remarkably, all five metagenomic bins obtained for this lineage encode a complete Mcr complex and the genetic repertoire to perform methanogenesis from methanol, methanethiol, and methylamine (26). Hence, the metabolic potential of these Archaea is similar to that of Methanofastidiosa and Methanomassiliicoccales, which is in agreement with the monophyletic clustering of McrAB of these groups (Fig. 4).

### Asgard archaea

Asgard archaea currently comprise the Lokiarchaeota (27), Thorarchaeota (28), Heimdallarchaeota, and Odinarchaeota (13) and have unveiled insights into eukaryogenesis. Lokiarchaeota, formerly known as MBG-B/DSAG, often represent abundant

**Box 1. Glossary box.**

AAG: Archaea of the ancient archaeal group. Heimdallarchaeota include lineages previously assigned to AAG.

Amo: Ammonia monooxygenase (Box 2).

ANME: A paraphyletic group of anaerobic methane-oxidizing euryarchaeota (Box 2).

AOA: Ammonia-oxidizing archaea (Box 2).

AOM: Anaerobic oxidation of methane (Box 2).

ARMAN: Archaeal Richmond Mine acidophilic nanoorganisms (Fig. 1).

Autotrophy: Ability of an organism to produce complex organic compounds from inorganic compounds such as $CO_2$.

Chemolithotrophy: Ability to conserve energy from the oxidation of inorganic compounds.

Chemoorganotrophy: Use of organic substrates for energy conservation.

Clade: Group of organisms that includes a common ancestor and all its descendants.

Contigs: Contiguous stretches of DNA of different length obtained upon assembly of sequence reads. These generally represent subregions of the genomic DNA of an organism.

DHVE: Refers to deep-sea hydrothermal vent euryarchaeota (Fig. 1).

DPANN: A proposed archaeal superphylum comprising Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanohaloarchaeota, and Nanoarchaeota. More recently, additional candidate phyla such as Woesearchaeota, Pacearchaeota, and Micrarchaeota and possibly the Altiarchaeota were suggested to be part of this group.

DSEG: Refers to deep-sea euryarchaeotal group (Fig. 1).

Environmental surveys: Profiling of microbial community composition via isolation of total nucleic acids from environmental samples and subsequent sequencing, amplification, and analysis of 16S and 18S rRNA genes. Amplification is based on polymerase chain reactions by using primer sequences that target conserved regions of 16S/18S rRNA genes.

Genomic bin: Composite genome comprising metagenomic contigs that are inferred to be derived from the same genome or species; also referred to as metagenome-assembled genome (MAG).

Halophile: An organism that is dependent on high sodium chloride concentrations for growth and survival. Adaptation to salt content varies between organisms (ranges from 0.3 to more than 5 M).

Heterotrophy: The requirement of complex organic compounds as source for cellular carbon.

Horizontal gene transfer (HGT): The lateral acquisition of genetic material from cells other than the parent cell, as opposed to vertical transmission from parent to offspring.

Hyperthermophile: An organism that is adapted to temperatures above 80°C.

LACA: Last archaeal common ancestor.

MBG-B: Archaea of the marine benthic group B (also DSAG), now known as Lokiarchaeota (Fig. 1).

MBG-D: Archaea of the marine benthic group D (Fig. 1).

Mcr: Methyl−coenzyme M reductase (Box 2).

MG-II and MG-III: Two distinct marine groups belonging to the Euryarchaeota (Fig. 1). MG-II archaea are now referred to as Thalassoarchaea.

MCG: Archaea of the Miscellaneous Crenarchaeotal Group, now referred to as Bathyarchaeota (Fig. 1).

Metagenomic binning: The clustering of metagenomics contigs into taxonomic bins based on similarities in nucleotide composition profiles and/or read coverage.

MSBL1: Archaea of the Mediterranean Sea Brine Lakes 1 group (Fig. 1).

Mesophile: An organism that thrives best at moderate temperatures between 20° and 45°C.

Metagenomics: The generation of DNA sequence data directly from environmental samples.

Methane seep: Locations at the seafloor that are characterized by the release of hydrocarbons dominated by methane.

Methylotrophy: A metabolic strategy in which organisms use organic compounds without C−C bonds as electron donors and carbon sources.

Monophyletic group: A set of organisms, which forms a clade and thus shares a common ancestor and includes all of its descendants.

Paraphyletic group: A set of organisms, which forms a clade and shares a common ancestor but does not include all of its descendant lineages.

SA1: Refers to a group of euryarchaeota, originally detected in brine-seawater interface samples from Shaban Deep, Red Sea.

TACK: A proposed archaeal superphylum comprising Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota, in addition to the tentative phyla Geoarchaeota, Bathyarchaeota, and Verstraetearchaeota. TACK is also referred to as Proteoarchaeota.

Thermophile: Organisms that thrive best at temperatures between 45° and 80°C.

TMCG: Archaea from the terrestrial miscellaneous crenarchaeota group, now also referred to as Verstraetearchaeota (Fig. 1).

SAGMEG: Refers to the South-African Gold Mine Miscellaneous Euryarchaeal Group, now named Hadesarchaea (Fig. 1).

Syntrophy: A metabolic process in which the degradation of a substrate gets energetically favorable through the cooperation of two different organisms—a product generated by one partner is consumed by the other partner.

WLP: Wood-Ljungdahl carbon fixation pathway (Box 2).

WSA2/Arc1: A group of archaea first identified in a clone library ("WSA") from a methanogenic sulfate-reducing core (Fig. 1) and later detected in an anaerobic sludge digester and described as Arc1; now referred to as Methanofastidiosa.

community members in deep-sea sediments (Fig. 1) (33, 67). In contrast, Thorarchaeota and Heimdallarchaeota (Fig. 1) seem to represent rare members of sediment microbial communities (13). Currently identified Odinarchaeota were found in sediments of hot springs or hydrothermal vent systems with temperatures ranging from about 60° to 70°C. Both Thor- and Lokiarchaeota encode

key enzymes of the WLP and may be capable of acetogenesis (28, 68). Furthermore, these genomes show the potential for heterotrophic growth, suggesting that they are also able to use the WLP as electron acceptor during fermentation of organic substrates (28). Interestingly, the genetic repertoire for the WLP appears to be absent from genomic bins of Odin- and Heimdallarchaeota

(Fig. 3). Prospective analyses of the metabolic features of Asgard archaea will certainly reveal a more detailed picture of their functional potential and ecological role.

### DPANN superphylum and Altiarchaea

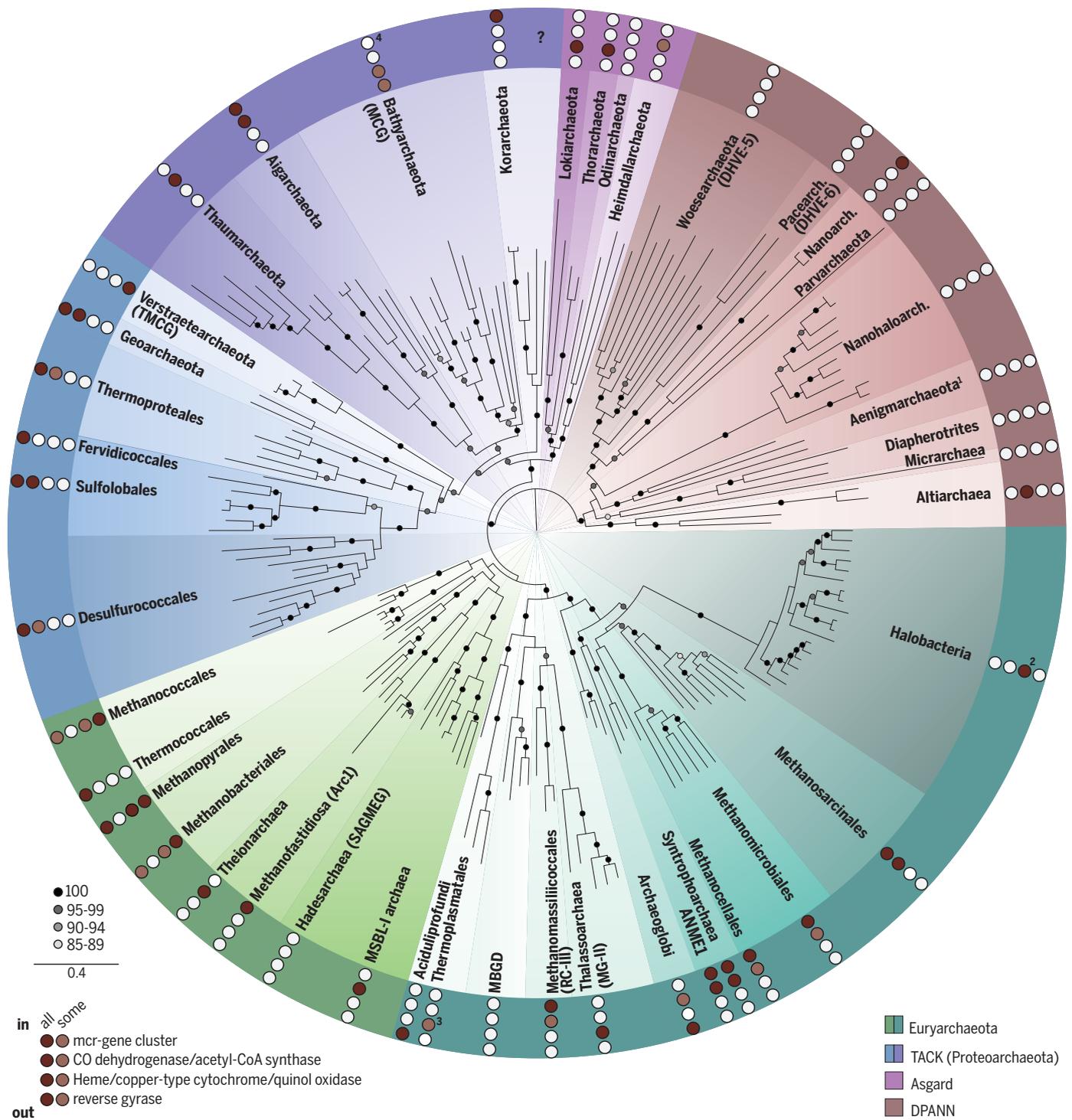The DPANN superphylum represents a phylogenetically diverse archaeal superphylum that

**Fig. 3. The genomic tree of Archaea and discussed key characteristics.** This phylogenetic tree shows the current diversity of archaea for which genomes (including single-cell and metagenome assembled genomes) have been obtained. Furthermore, on the outer ring, this figure highlights key characteristics of Archaea (discussed in this Review) that are important for understanding the evolution of the physiology and metabolism in this domain of life. The tree is based on a maximum-likelihood analysis [IQ-TREE (98) with LG+C60+F, 1000 ultrafast bootstraps] of a concatenated alignment of at least 11 of 15 ribosomal proteins (14) with 2611 positions [MAFFT-LINSi (96), trimal (97)]. The final tree was rooted with the DPANN archaea. Additionally, (1) a strictly anaerobic haloarchaeon has recently been enriched; (2) Thermoplasmatales grow aerobically, but some members lack heme/copper-type cytochrome/quinol oxidases (99); and (3) one representative of Bathyarchaeota encodes reverse gyrase. Branch support values above 90 are indicated by dots. Scale bar indicates number of substitutions per site.
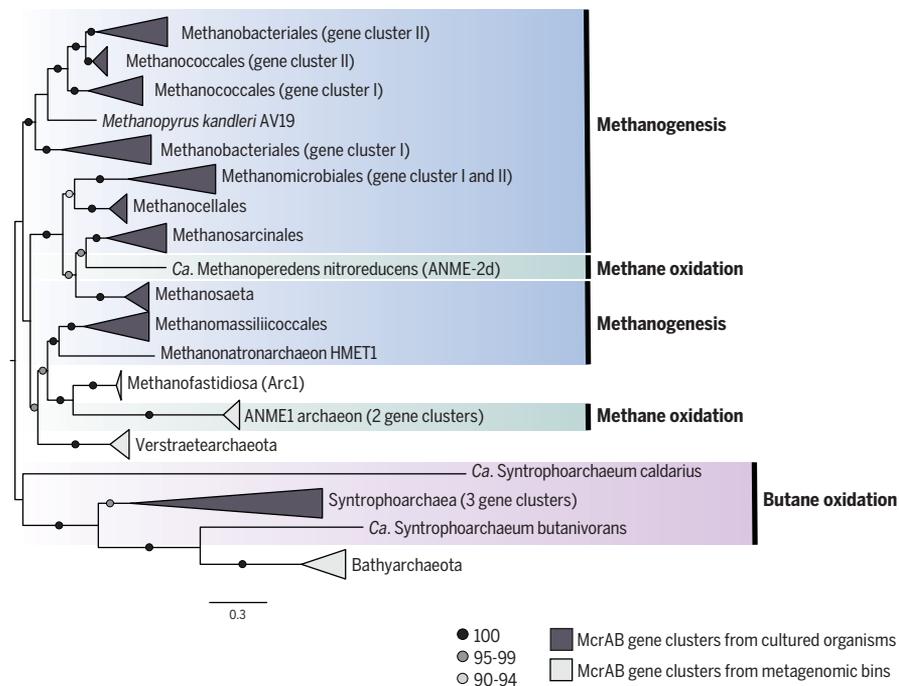
**Fig. 4. The complex evolutionary history of methyl–coenzyme M reductase (Mcr).** This phylogenetic tree depicts the diversity of the key enzyme (Mcr) for methanogenesis (taxa shaded in blue) and anaerobic methane oxidation (taxa shaded in green), which was recently also implicated in the oxidation of butane in Syntrophoarchaea (taxa shaded in lavender). It is based on maximum-likelihood analyses [IQ-TREE (*98*) with LG+C60+F, 1000 ultrafast bootstraps], which were performed on a concatenated McrA and B alignment of 981 positions [MAFFT-LINSi (*96*) combined with trimal (*97*)]. MrcA and B subunits derived from the same gene cluster were concatenated. Branch support values above 85 are indicated with dots. Scale bar indicates number of substitutions per site.

includes environmental lineages such as DHVE-5, DHVE-6, and DSEG (Fig. 1) [(*33*) and references therein]. DPANN archaea occur in diverse habitats, ranging from hypersaline lakes, marine and lake waters, sediments, and acid mine drainage to hot springs (*14, 16, 29, 31, 32, 69–74*). Because of the small cell size of many members of this group, DPANN archaea have been overlooked by conventional analyses, and cultivated representatives are only available for Nanoarchaeota, which grow attached to a crenarchaeal host cell (*16, 69*). "ARMAN" archaea—the Parvarchaeota and Micrarchaeota—were among the first DPANN lineages to be explored at the genomic level (Fig. 3) (*73*). Although these organisms may be host-associated, their genomes revealed the potential to perform central steps in carbon metabolism, lipid degradation, and oxidative phosphorylation. Analyses of several near-complete genomes of Woese-, Pace-, and Aenigmarchaeota as well as Diapherotrites revealed that these have a reduced genomic repertoire and lack major anabolic and catabolic pathways (*14, 29*). These findings indicate that some of these archaea have a symbiotic and/or host-dependent lifestyle, whereas others may be capable of a saccharolytic or fermentative growth (*14*). Members of the Nanohaloarchaeota are found in hypersaline lakes and also seem to be free-living and may have an aerobic heterotrophic or even photoheterotrophic lifestyle (*70, 71*) or grow anaerobically by fermentation (*72*). Last, the investigation of genomes of biofilm-

forming Altiarchaea suggested that these archaea are able to grow autotrophically on carbon dioxide and potentially also on acetate, formate, or carbon monoxide (*31, 32*). Although Altiarchaea are nonmethanogenic, it was shown that their genomes encode a modified archaeal WLP (*31*).

## Uncovering the evolution and deep roots of the archaea

### The role of horizontal gene transfer in archaeal evolution

Horizontal gene transfer (HGT) represents the major source of gene gain and thus innovation in archaeal and bacterial genomes [reviewed in (*75*)]. HGT seems to have had a variable impact on the evolution of different archaeal lineages being particularly prevalent in Haloarchaea, Thaumarchaeota, MG II/III, Sulfolobales, Methanosarcinales, and Methanomassiliicoccus (*30, 61, 76–79*). Gene acquisitions from bacteria have contributed to the adaptation of archaea to new environments and lifestyles, aiding the diversification of this domain of life (*75, 77, 80*). To what extent these gene acquisitions occurred over a short period of time or represent a continuous process is currently under debate (*62, 77, 79*). HGT rates seem to be asymmetrical, with bacteria-to-archaea transfers appearing to be at least five times more frequent than vice versa (*77, 78*). This observed asymmetry may be caused by the bacterial overrepresentation over archaea in most environments or by different

mechanistic, selective, and adaptive biases (*77, 80*). However, methodological biases such as an underrepresentation of genomic data from archaea as compared with bacteria can currently not be excluded. Improvements of methods to detect HGT as well as the inclusion of genomic data from previously unknown archaeal lineages will likely reveal a clearer picture of the prevalence and directionality of HGT between Archaea and Bacteria.

### The root of the archaeal tree

The "coming of age" of the archaeal tree (Fig. 3) represents an opportunity to readdress long-standing questions about the early evolution of the Archaea and the nature of the last archaeal common ancestor (LACA). In this regard, it is essential to confidently place the root position in the archaeal tree and determine the phylogenetic position of DPANN archaea. Because the use of a distant outgroup to root phylogenetic trees can cause long-branch attraction artifacts, fast-evolving or compositionally biased sequences, such as those of DPANN lineages, could artificially emerge as a monophyletic basal group [(*30, 81*) and references therein].

Recent studies have reported varying positions of the root. For instance, the reconstruction of phylogenetic trees by using both conserved protein markers as well as 38 new markers placed the root between Euryarchaeota and the TACK superphylum (*20*). However, a two-step approach (based on only two domains of life at a time)

**Box 2. Definitions and terminology for archaeal key metabolic strategies discussed in this review.**

Metagenomic approaches have revealed that archaea are important players in the nitrogen and carbon cycles because they are capable of ammonia oxidation, methanogenesis, and methane oxidation. In the following, we introduce the key enzymes and terminology associated with these metabolic regimes as well as the Wood-Ljungdahl carbon fixation pathway (WLP) in archaea.

Aerobic ammonia oxidation is the oxidation of ammonia to nitrite and represents the first step of nitrification, a fundamental part of the nitrogen cycle [reviewed in (37)]. The key enzyme for the oxidation of ammonia is ammonia monooxygenase (Amo). A specific group of archaea—the ammonia-oxidizing archaea (AOA), which belong to the Thaumarchaeota and several groups of ammonia-oxidizing bacteria (AOB)—encode amo genes and are capable of aerobic ammonia oxidation.

Methanogenesis and anaerobic oxidation of methane (AOM) is the consumption or production, respectively, of methane through simultaneous energy conservation. These processes are globally important and specific to archaea (35, 100). Although most methanogens can produce methane by the reduction of carbon dioxide with hydrogen and formate, some can use alternative methylated substrates. However, about two-thirds of all methane is generated through acetoclastic methanogenesis, in which methane is generated from the splitting of acetate (100). Current estimates of AOM flux indicate that an estimated 7 to 25% of the global methane emissions may be consumed by "Anaerobic Methane Oxidizing Euryarchaeota" (ANME) through a reversal of the methanogenesis pathway, which also includes the key enzymatic complex referred to as methyl–coenzyme M reductase (Mcr) (39). Therefore, methanogens and ANME share several key metabolic enzymes, which may trace back to the earliest nodes in the tree of archaea (84).

The WLP [or acetyl–coenzyme A (CoA) pathway] represents a means for the fixation of two molecules of carbon dioxide to acetyl-CoA and/or acetate with the simultaneous generation of energy (101). Two molecules of $CO_2$ are converted to a cofactor-bound methyl group and CO and subsequently combined to generate acetyl-CoA by the key enzyme complex carbon monoxide dehydrogenase/acetyl-CoA synthase (ACS/CODH) (102). The WLP has been suggested to represent the oldest carbon fixation pathway on Earth (103) and is used for the generation of acetate in acetogenic bacteria (101). In methanogenic archaea, the archaeal version of the WLP is used for autotrophic growth and methanogenesis by using acetate as substrate.

suggested that the root of the archaeal tree was placed inside Euryarchaeota, between a clade comprising Thermococcales, Methanococcales, Methanobacteriales, Methanopyrales, and the TACK superphylum and a clade comprising the rest of the Euryarchaeota (81). Yet, both of these approaches relied on using a distant outgroup (Bacteria), which may have caused phylogenetic artifacts. Sophisticated Bayesian models to root the archaeal tree without the need for an outgroup retrieved a root between TACK archaea and a clade comprising Euryarchaeota/DPANN (82). However, the data set analyzed in this study was relatively small and comprised solely 16S and 23S rRNA gene sequences, which are known to be compositionally biased (83). A consensus unrooted archaeal multigene supertree, rooted with a newly developed model of genome evolution, placed the root between the DPANN superphylum and all other Archaea and supported the monophyly of Euryarchaeota (30). Future analyses using alternative phylogenomic approaches and more comprehensive data sets will likely provide a more definite answer on this matter.

### The nature of the last common ancestor of the archaea

Reconstructions of ancestral gene sets have indicated that LACA may have been an anaerobic chemolithoautotroph that used the WLP for

carbon fixation and may have been able to obtain energy through methanogenesis [(30, 81, 84–86) and references therein]. The recent discovery of the key genes for the WLP (ACS/CODH) and methanogenesis (Mcr) in archaeal lineages outside of the Euryarchaeota (Fig. 3 and Box 2) and the patchy distribution of terminal oxidases in currently available genomes (Fig. 3) supports this hypothesis [(56) and references therein]. Mcr is now known to be encoded by various Euryarchaeota lineages as well as by members of TACK, such as Bathyarchaeota and Verstraetearchaeota (Fig. 3) (23, 26). However, phylogenetic analyses of Mcr indicate that protein sequences may reflect functional similarity rather than phylogenetic relatedness of the respective species and implicate a complex evolutionary history of this gene cluster involving HGT and gene losses (Fig. 4) (50, 56, 64). For example, the bathyarchaeal Mcr homologs are most closely related to those present in Syntrophoarchaea (50), whereas those of Verstraetearchaeota, Methanomassiliicoccales, Methanonatronarchaeia, Methanofastidiosa, and ANME1 seem to share a common ancestry, despite the distant evolutionary relationship of these lineages (Fig. 4) (23, 26, 64). Therefore, Mcr homologs in TACK may have been acquired horizontally from euryarchaeal lineages rather than representing an ancestral feature. In-depth analyses of the complete machinery of methanogenesis

across different archaeal lineages, a solid rooted tree, and a robust placement of DPANN are needed to determine the likelihood for the presence of methanogensis in LACA (30, 80). The WLP has now been detected in all major archaeal superphyla, including TACK and Asgard, Euryarchaeota, and putative deep-branching members of the DPANN (Altiarchaeales) (Fig. 3). Similar to Archaeoglobales, many of the lineages encoding the archaeal version of the WLP lack mcr genes (Figs. 3 and 4), indicating that the WLP is not necessarily linked to methane metabolism [reviewed in (56)]. The broad distribution of WLP-related genes in archaea is in agreement with hypotheses that suggest its presence in LACA (30, 84, 86).

Inferences of optimal growth temperatures have suggested that LACA was a thermophilic or hyperthermophilic organism (30, 87) and may have encoded reverse gyrase, a genetic marker for hyperthermophily (Fig. 3) (80, 85). The adaptation to moderate environments but also to aerobic lifestyles likely occurred independently in different archaeal lineages and were aided by HGT from Bacteria (30, 80, 86). However, these hypotheses need to be evaluated further by taking into account alternative positions of the root and previously unidentified, potentially mesophilic archaeal groups such as the Bathyarchaeota, Verstraetearchaeota, DPANN, and Asgard archaea.

### The archaeal ancestry of eukaryotes

The expansion of genomic data from archaea has forced us to rethink the evolutionary transitions of life, particularly with respect to the origin of eukaryotes. Archaea have featured prominently in various eukaryogenic hypotheses, in particular in symbiogenic scenarios that suggest that eukaryotes evolved from a merger of an archaeal host cell and an alphaproteobacterial endosymbiont [reviewed in (88)]. Early on, Archaea and eukaryotes were shown to share a common ancestry to the exclusion of Bacteria ("Three Domains tree") (Fig. 5A) (2, 89). However, better genomic taxonsampling combined with the use of more sophisticated phylogenomic tools [(19, 90) and references therein] yielded trees in which eukaryotes branched from within the archaeal domain of life ("Two Domains tree") (Fig. 5B) (91). These analyses recovered a monophyletic clade comprising eukaryotes and the archaeal TACK superphylum, a topology that was in agreement with the presence of several eukaryotic signature proteins in TACK archaea (Fig. 5D) (19). Subsequently, the discovery of Asgard archaea has provided further evidence for symbiogenic scenarios and the origin of eukaryotes from an archaeal host (13, 27). First of all, phylogenomic analyses of marker gene data sets place Asgard as a sister group of eukaryotes (Fig. 5C) (13). Furthermore, genome analyses of these archaea revealed that they encode a multitude of proteins previously regarded as specific to eukaryotes (13, 27), including several components involved in membrane bending and vesicleformation processes (Fig. 5D) (92, 93). Intriguingly, Thorarchaeota were shown to encode yet additional components of the eukaryotic membrane trafficking machinery, including proteins with Sec23/24 and
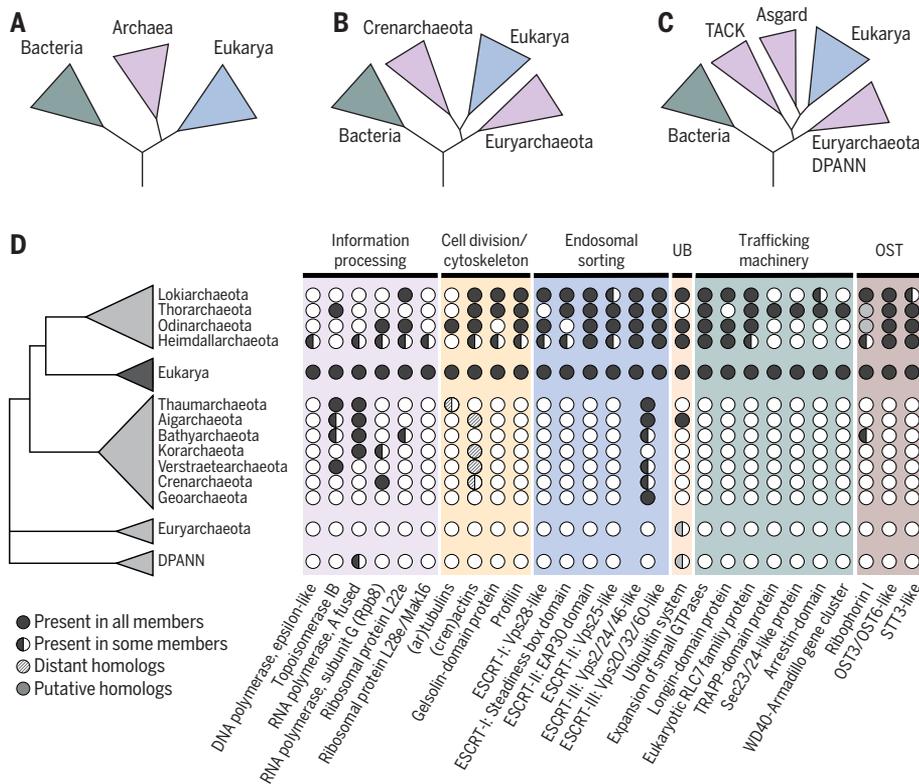
**Fig. 5. Archaea and the changing tree of life.** (**A** to **C**) Schematic depiction of the relationship of Archaea, Bacteria, and Eukaryotes according to the (A) Three and (B) Two Domains topology, updated with (C) the Asgard archaea. The Three Domains tree suggested that Archaea, Bacteria, and Eukaryotes represent primary domains of life. In contrast, a Two Domains topology [(B) and (C)] is more compatible with the view that eukaryotes represent a secondary domain of life derived from a merger of two prokaryotes: an archaeal host and an alphaproteobacterial symbiont. (**D**) A schematic tree of the archaea and their relationship with the eukaryotes emphasizing the placement of eukaryotes with Asgard archaea and showing the distribution of exclusive sequential patterns (ESPs) in the different lineages. The presence or absence pattern of ESPs is based on previous studies and was updated by taking into account publicly available metagenome-assembled genomes.

with TRAPP domains (Fig. 5D). The identification of gene clusters consisting of β-propeller and α-solenoid domain proteins represents the most important discovery (*13*) because these protein domains are the basic building blocks of eukaryotic coat proteins essential for vesicle biogenesis and transport [reviewed in (*94*)]. In light of these findings, it seems plausible that these key components in the archaeal ancestor of eukaryotes were involved in mediating the relationship with a syntrophic or symbiotic partner and provided the building blocks for the subsequent evolution of the trafficking machinery and compartmentalized nature of eukaryotic cells.

## Summary and future perspectives

Advances in cultivation-independent genomics approaches have started to reveal the unprecedented genomic diversity of the archaeal domain of life. Novel genomic data has provided insights into the metabolic repertoire and evolution of members of the Archaea and their relationship with Eukaryotes. Recent estimates indicate that about half of the existing archaeal diversity still defies genomic characterization (Fig. 1) (*95*), suggesting

that the coming years will witness additional major discoveries. Many of the unexplored archaeal lineages are hiding in little-investigated and low-biomass environments and may represent low abundant members of microbial communities. Continued efforts in genome sequencing and culturing of these archaea will be crucial to study the role that archaea play in their natural habitats, their interactions with other organisms, and their function in biogeochemical cycles and Earth's ecosystems. Last, these data will provide a treasure trove for analyses to identify the lifestyle and metabolic potential of the archaeal ancestor and hopefully to reconstruct life's earliest evolutionary trajectories.

## REFERENCES AND NOTES

1. C. R. Woese, G. E. Fox, Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5088–5090 (1977). doi: 10.1073/pnas.74.11.5088; pmid: 270744
2. C. R. Woese, O. Kandler, M. L. Wheelis, Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria and Eucarya. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4576–4579 (1990). doi: 10.1073/pnas.87.12.4576; pmid: 2112744
3. R. Y. Stanier, C. B. Van Niel, The concept of a bacterium. *Arch. Mikrobiol.* **42**, 17–35 (1962). doi: 10.1007/BF00425185; pmid: 13916221
4. W. M. Fitch, E. Margoliash, Construction of phylogenetic trees. *Science* **155**, 279–284 (1967). doi: 10.1126/science.155.3760.279; pmid: 5334057
5. K. O. Stetter, History of discovery of the first hyperthermophiles. *Extremophiles* **10**, 357–362 (2006). doi: 10.1007/s00792-006-0012-7; pmid: 16941067
6. O. Kandler, H. Hippe, Lack of peptidoglycan in the cell walls of *Methanosarcina barkeri. Arch. Microbiol.* **113**, 57–60 (1977). doi: 10.1007/BF00428580; pmid: 889387
7. W. Zillig, K. O. Stetter, D. Janeković, DNA-dependent RNA polymerase from the archaebacterium *Sulfolobus acidocaldarius. Eur. J. Biochem.* **96**, 597–604 (1979). doi: 10.1111/j.1432-1033.1979.tb13074.x; pmid: 380989
8. T. A. Langworthy, P. F. Smith, W. R. Mayberry, Lipids of *Thermoplasma acidophilum. J. Bacteriol.* **112**, 1193–1200 (1972). pmid: 4344918
9. C. J. Bult et al., Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii. Science* **273**, 1058–1073 (1996). doi: 10.1126/science.273.5278.1058; pmid: 8688087
10. N. R. Pace, Mapping the tree of life: Progress and prospects. *Microbiol. Mol. Biol. Rev.* **73**, 565–576 (2009). doi: 10.1128/MMBR.00033-09; pmid: 19946133
11. R. Stepanauskas, Single cell genomics: An individual look at microbes. *Curr. Opin. Microbiol.* **15**, 613–620 (2012). doi: 10.1016/j.mib.2012.09.001; pmid: 23026140
12. N. Sangwan, F. Xia, J. A. Gilbert, Recovering complete and draft population genomes from metagenome datasets. *Microbiome* **4**, 8 (2016). doi: 10.1186/s40168-016-0154-5; pmid: 26951112
13. K. Zaremba-Niedzwiedzka et al., Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017). doi: 10.1038/nature21031; pmid: 28077874
14. C. J. Castelle et al., Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* **25**, 690–701 (2015). doi: 10.1016/j.cub.2015.01.014; pmid: 25702576
15. J. G. Elkins et al., A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 8102–8107 (2008). doi: 10.1073/pnas.0801980105; pmid: 18535141
16. H. Huber et al., A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**, 63–67 (2002). doi: 10.1038/417063a; pmid: 11986665
17. C. Brochier-Armanet, B. Boussau, S. Gribaldo, P. Forterre, Mesophilic Crenarchaeota: Proposal for a third archaeal phylum, the Thaumarchaeota. *Nat. Rev. Microbiol.* **6**, 245–252 (2008). doi: 10.1038/nrmicro1852; pmid: 18274537
18. T. Nunoura et al., Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res.* **39**, 3204–3223 (2011). doi: 10.1093/nar/gkq1228; pmid: 21169198
19. L. Guy, T. J. Ettema, The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol.* **19**, 580–587 (2011). doi: 10.1016/j.tim.2011.09.002; pmid: 22018741
20. C. Petitjean, P. Deschamps, P. López-García, D. Moreira, Rooting the domain archaea by phylogenomic analysis supports the foundation of the new kingdom Proteoarchaeota. *Genome Biol. Evol.* **7**, 191–204 (2014). doi: 10.1093/gbe/evu274; pmid: 25527841
21. M. A. Kozubal et al., Geoarchaeota: A new candidate phylum in the Archaea from high-temperature acidic iron mats in Yellowstone National Park. *ISME J.* **7**, 622–634 (2013). doi: 10.1038/ismej.2012.132; pmid: 23151644
22. L. Guy, A. Spang, J. H. Saw, T. J. Ettema, 'Geoarchaeote NAG1' is a deeply rooting lineage of the archaeal order Thermoproteales rather than a new phylum. *ISME J.* **8**, 1353–1357 (2014). doi: 10.1038/ismej.2014.6; pmid: 24522265
23. P. N. Evans et al., Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* **350**, 434–438 (2015). doi: 10.1126/science.aac7745; pmid: 26494757
24. Y. He et al., Genomic and enzymatic evidence for acetogenesis among multiple lineages of the archaeal phylum Bathyarchaeota widespread in marine sediments. *Nat. Microbiol.* **1**, 16035 (2016). doi: 10.1038/nmicrobiol.2016.35; pmid: 27572832
25. C. S. Lazar et al., Genomic evidence for distinct carbon substrate preferences and ecological niches of

Bathyarchaeota in estuarine sediments. *Environ. Microbiol.*
**18**, 1200–1211 (2016). doi: 10.1111/1462-2920.13142;
pmid: 26626228

26. I. Vanwonterghem *et al.*, Methylotrophic methanogenesis
discovered in the archaeal phylum Verstraetearchaeota.
*Nat. Microbiol.* **1**, 16170 (2016). doi: 10.1038/
nmicrobiol.2016.170; pmid: 27694807

27. A. Spang *et al.*, Complex archaea that bridge the gap
between prokaryotes and eukaryotes. *Nature* **521**, 173–179
(2015). doi: 10.1038/nature14447; pmid: 25945739

28. K. W. Seitz, C. S. Lazar, K. U. Hinrichs, A. P. Teske, B. J. Baker,
Genomic reconstruction of a novel, deeply branched
sediment archaeal phylum with pathways for acetogenesis
and sulfur reduction. *ISME J.* **10**, 1696–1705 (2016).
doi: 10.1038/ismej.2015.233; pmid: 26824177

29. C. Rinke *et al.*, Insights into the phylogeny and coding
potential of microbial dark matter. *Nature* **499**, 431–437
(2013). doi: 10.1038/nature12352; pmid: 23851394

30. T. A. Williams *et al.*, Integrative modeling of gene and genome
evolution roots the archaeal tree of life. *Proc. Natl. Acad.
Sci. U.S.A.* **114**, E4602–E4611 (2017). doi: 10.1073/
pnas.1618463114; pmid: 28533395

31. A. J. Probst *et al.*, Biology of a widespread uncultivated
archaeon that contributes to carbon fixation in the
subsurface. *Nat. Commun.* **5**, 5497 (2014). doi: 10.1038/
ncomms6497; pmid: 25425419

32. J. T. Bird, B. J. Baker, A. J. Probst, M. Podar, K. G. Lloyd,
Culture independent genomic comparisons reveal
environmental adaptations for Altiarchaeales. *Front.
Microbiol.* **7**, 1221 (2016). doi: 10.3389/fmicb.2016.01221;
pmid: 27547202

33. A. M. Durbin, A. Teske, Archaea in organic-lean and
organic-rich marine subsurface sediments: An environmental
gradient reflected in distinct phylogenetic lineages.
*Front. Microbiol.* **3**, 168 (2012). doi: 10.3389/
fmicb.2012.00168; pmid: 22666218

34. S. Gribaldo, C. Brochier-Armanet, Time for order in microbial
systematics. *Trends Microbiol.* **20**, 209–210 (2012).
doi: 10.1016/j.tim.2012.02.006; pmid: 22440793

35. P. Offre, A. Spang, C. Schleper, Archaea in biogeochemical
cycles. *Annu. Rev. Microbiol.* **67**, 437–457 (2013).
doi: 10.1146/annurev-micro-092412-155614; pmid: 23808334

36. M. Pester, C. Schleper, M. Wagner, The Thaumarchaeota:
An emerging view of their phylogeny and ecophysiology.
*Curr. Opin. Microbiol.* **14**, 300–306 (2011). doi: 10.1016/
j.mib.2011.04.007; pmid: 21546306

37. D. A. Stahl, J. R. de la Torre, Physiology and diversity of
ammonia-oxidizing archaea. *Annu. Rev. Microbiol.* **66**, 83–101
(2012). doi: 10.1146/annurev-micro-092611-150128;
pmid: 22994489

38. A. Teske, K. B. Sørensen, Uncultured archaea in deep marine
subsurface sediments: Have we caught them all? *ISME J.* **2**,
3–18 (2008). doi: 10.1038/ismej.2007.90; pmid: 18180743

39. K. Knittel, A. Boetius, Anaerobic oxidation of methane: Progress
with an unknown process. *Annu. Rev. Microbiol.* **63**, 311–334
(2009). doi: 10.1146/annurev.micro.61.080706.093130;
pmid: 19575572

40. T. M. Hoehler, M. J. Alperin, D. B. Albert, C. S. Martens, Field
and laboratory studies of methane oxidation in an anoxic
marine sediment—Evidence for a methanogen-sulfate
reducer consortium. *Global Biogeochem. Cycles* **8**, 451–463
(1994). doi: 10.1029/94GB01800

41. A. A. Raghoebarsing *et al.*, A microbial consortium couples
anaerobic methane oxidation to denitrification. *Nature*
**440**, 918–921 (2006). doi: 10.1038/nature04617;
pmid: 16612380

42. E. J. Beal, C. H. House, V. J. Orphan, Manganese- and iron-
dependent marine methane oxidation. *Science* **325**, 184–187
(2009). doi: 10.1126/science.1169984; pmid: 19589998

43. S. E. McGlynn, G. L. Chadwick, C. P. Kempes, V. J. Orphan,
Single cell activity reveals direct electron transfer in
methanotrophic consortia. *Nature* **526**, 531–535 (2015).
doi: 10.1038/nature15512; pmid: 26375009

44. G. Wegener, V. Krukenberg, D. Riedel, H. E. Tegetmeyer,
A. Boetius, Intercellular wiring enables electron transfer
between methanotrophic archaea and bacteria. *Nature*
**526**, 587–590 (2015). doi: 10.1038/nature15733;
pmid: 26490622

45. M. F. Haroon *et al.*, Anaerobic oxidation of methane coupled to
nitrate reduction in a novel archaeal lineage. *Nature* **500**,
567–570 (2013). doi: 10.1038/nature12375; pmid: 23892779

46. S. J. Hallam *et al.*, Reverse methanogenesis: Testing the
hypothesis with environmental genomics. *Science* **305**,

1457–1462 (2004). doi: 10.1126/science.1100025;
pmid: 15353801

47. A. Pernthaler *et al.*, Diverse syntrophic partnerships from
deep-sea methane vents revealed by direct cell capture
and metagenomics. *Proc. Natl. Acad. Sci. U.S.A.* **105**,
7052–7057 (2008). doi: 10.1073/pnas.0711303105;
pmid: 18467493

48. A. E. Dekas, R. S. Poretsky, V. J. Orphan, Deep-sea archaea
fix and share nitrogen in methane-consuming microbial
consortia. *Science* **326**, 422–426 (2009). doi: 10.1126/
science.1178223; pmid: 19833965

49. A. Meyerdierks *et al.*, Metagenome and mRNA expression
analyses of anaerobic methanotrophic archaea of the
ANME-1 group. *Environ. Microbiol.* **12**, 422–439 (2010).
doi: 10.1111/j.1462-2920.2009.02083.x; pmid: 19878267

50. R. Laso-Pérez *et al.*, Thermophilic archaea activate butane via
alkyl-coenzyme M formation. *Nature* **539**, 396–401 (2016).
doi: 10.1038/nature20152; pmid: 27749816

51. V. Iverson *et al.*, Untangling genomes from metagenomes:
Revealing an uncultured class of marine Euryarchaeota.
*Science* **335**, 587–590 (2012). doi: 10.1126/science.1212665;
pmid: 22301318

52. M. Li *et al.*, Genomic and transcriptomic evidence for
scavenging of diverse organic compounds by widespread
deep-sea archaea. *Nat. Commun.* **6**, 8933 (2015).
doi: 10.1038/ncomms9933; pmid: 26573375

53. K. G. Lloyd *et al.*, Predominant archaea in marine sediments
degrade detrital proteins. *Nature* **496**, 215–218 (2013).
doi: 10.1038/nature12033; pmid: 23535597

54. C. S. Lazar, B. J. Baker, K. W. Seitz, A. P. Teske, Genomic
reconstruction of multiple lineages of uncultured benthic
archaea suggests distinct biogeochemical roles and
ecological niches. *ISME J.* **11**, 1118–1129 (2017). doi: 10.1038/
ismej.2016.189; pmid: 28085154

55. G. Borrel *et al.*, Phylogenomic data support a seventh
order of Methylotrophic methanogens and provide insights
into the evolution of Methanogenesis. *Genome Biol. Evol.* **5**,
1769–1780 (2013). doi: 10.1093/gbe/evt128;
pmid: 23985970

56. G. Borrel, P. S. Adam, S. Gribaldo, Methanogenesis and the
Wood-Ljungdahl pathway: An ancient, versatile, and fragile
association. *Genome Biol. Evol.* **8**, 1706–1711 (2016).
doi: 10.1093/gbe/evw114; pmid: 27189979

57. B. J. Baker *et al.*, Genomic inference of the metabolism of
cosmopolitan subsurface Archaea, Hadesarchaea. *Nat.
Microbiol.* **1**, 16002 (2016). doi: 10.1038/nmicrobiol.2016.2;
pmid: 27572167

58. R. Mwirichia *et al.*, Metabolic traits of an uncultured archaeal
lineage—MSBL1—from brine pools of the Red Sea. *Sci. Rep.*
**6**, 19181 (2016). doi: 10.1038/srep19181; pmid: 26758088

59. M. K. Nobu, T. Narihiro, K. Kuroda, R. Mei, W. T. Liu, Chasing
the elusive Euryarchaeota class WSA2: Genomes reveal a
uniquely fastidious methyl-reducing methanogen. *ISME J.* **10**,
2478–2487 (2016). doi: 10.1038/ismej.2016.33;
pmid: 26943620

60. A. Oren, Taxonomy of halophilic Archaea: Current status
and future challenges. *Extremophiles* **18**, 825–834 (2014).
doi: 10.1007/s00792-014-0654-9; pmid: 25102811

61. S. Nelson-Sathi *et al.*, 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
(2012). doi: 10.1073/pnas.1209119109; pmid: 23184964

62. E. A. Becker *et al.*, Phylogenetically driven sequencing of
extremely halophilic archaea reveals strategies for static and
dynamic osmo-response. *PLOS Genet* **10**, e1004784 (2014).

63. D. Y. Sorokin *et al.*, Elemental sulfur and acetate can support
life of a novel strictly anaerobic haloarchaeon. *ISME J.* **10**,
240–252 (2016). doi: 10.1038/ismej.2015.79;
pmid: 25978546

64. D. Y. Sorokin *et al.*, Discovery of extremely halophilic,
methyl-reducing euryarchaea provides insights into the
evolutionary origin of methanogenesis. *Nat. Microbiol.* **2**,
17081 (2017). doi: 10.1038/nmicrobiol.2017.81;
pmid: 28555626

65. K. Kubo *et al.*, Archaea of the Miscellaneous Crenarchaeotal
Group are abundant, diverse and widespread in marine
sediments. *ISME J.* **6**, 1949–1965 (2012). doi: 10.1038/
ismej.2012.37; pmid: 22551871

66. J. Meng *et al.*, Genetic and functional properties of
uncultivated MCG archaea assessed by metagenome and
gene expression analyses. *ISME J.* **8**, 650–659 (2014).
doi: 10.1038/ismej.2013.174; pmid: 24108328

67. S. L. Jorgensen *et al.*, Correlating microbial community
profiles with geochemical data in highly stratified sediments

from the Arctic Mid-Ocean Ridge. *Proc. Natl. Acad. Sci. U.S.A.*
**109**, E2846–E2855 (2012). doi: 10.1073/pnas.1207574109;
pmid: 23027979

68. F. L. Sousa, S. Neukirchen, J. F. Allen, N. Lane, W. F. Martin,
Lokiarchaeon is hydrogen dependent. *Nat. Microbiol.* **1**,
16034 (2016). doi: 10.1038/nmicrobiol.2016.34;
pmid: 27572645

69. M. Podar *et al.*, Insights into archaeal evolution and symbiosis
from the genomes of a nanoarchaeon and its inferred
crenarchaeal host from Obsidian Pool, Yellowstone National
Park. *Biol. Direct* **8**, 9 (2013). doi: 10.1186/1745-6150-8-9;
pmid: 23607440

70. R. Ghai *et al.*, New abundant microbial groups in aquatic
hypersaline environments. *Sci. Rep.* **1**, 135 (2011).
doi: 10.1038/srep00135; pmid: 22355652

71. P. Narasingarao *et al.*, De novo metagenomic assembly
reveals abundant novel major lineage of Archaea in
hypersaline microbial communities. *ISME J.* **6**, 81–93 (2012).
doi: 10.1038/ismej.2011.78; pmid: 21716304

72. C. D. Vavourakis *et al.*, Metagenomic insights into the
uncultured diversity and physiology of microbes in four
hypersaline Soda Lake brines. *Front. Microbiol.* **7**, 211 (2016).
doi: 10.3389/fmicb.2016.00211; pmid: 26941731

73. B. J. Baker *et al.*, Enigmatic, ultrasmall, uncultivated Archaea.
*Proc. Natl. Acad. Sci. U.S.A.* **107**, 8806–8811 (2010).
doi: 10.1073/pnas.0914470107; pmid: 20421484

74. N. H. Youssef *et al.*, Insights into the metabolism, lifestyle
and putative evolutionary history of the novel archaeal
phylum 'Diapherotrites'. *ISME J.* **9**, 447–460 (2015).
doi: 10.1038/ismej.2014.141; pmid: 25083301

75. E. V. Koonin, Horizontal gene transfer: Essentiality and
evolvability in prokaryotes, and roles in evolutionary
transitions. *F1000Res.* **5**, 1805 (2016). doi: 10.12688/
f1000research.8737.1; pmid: 27508073

76. Y. I. Wolf, K. S. Makarova, N. Yutin, E. V. Koonin, Updated
clusters of orthologous genes for Archaea: A complex
ancestor of the Archaea and the byways of horizontal
gene transfer. *Biol. Direct* **7**, 46 (2012). doi: 10.1186/
1745-6150-7-46; pmid: 23241446

77. S. Nelson-Sathi *et al.*, Origins of major archaeal clades
correspond to gene acquisitions from bacteria.
*Nature* **517**, 77–80 (2015). doi: 10.1038/nature13805;
pmid: 25317564

78. P. Deschamps, Y. Zivanovic, D. Moreira, F. Rodriguez-Valera,
P. López-García, 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 (2014).
doi: 10.1093/gbe/evu127; pmid: 24923324

79. M. Groussin *et al.*, Gene acquisitions from bacteria at the
origins of major archaeal clades are vastly overestimated.
*Mol. Biol. Evol.* **33**, 305–310 (2016). doi: 10.1093/molbev/
msv249; pmid: 26541173

80. P. López-García, Y. Zivanovic, P. Deschamps, D. Moreira,
Bacterial gene import and mesophilic adaptation in archaea.
*Nat. Rev. Microbiol.* **13**, 447–456 (2015). doi: 10.1038/
nrmicro3485; pmid: 26075362

81. K. Raymann, C. Brochier-Armanet, S. Gribaldo, The two-
domain tree of life is linked to a new root for the Archaea.
*Proc. Natl. Acad. Sci. U.S.A.* **112**, 6670–6675 (2015).
doi: 10.1073/pnas.1420858112; pmid: 25964353

82. T. A. Williams *et al.*, New substitution models for rooting
phylogenetic trees. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*
**370**, 20140336 (2015). doi: 10.1098/rstb.2014.0336;
pmid: 26323766

83. N. Galtier, J. R. Lobry, Relationships between genomic G+C
content, RNA secondary structures, and optimal growth
temperature in prokaryotes. *J. Mol. Evol.* **44**, 632–636
(1997). doi: 10.1007/PL00006186; pmid: 9169555

84. F. L. Sousa *et al.*, Early bioenergetic evolution. *Philos. Trans.
R. Soc. Lond. B Biol. Sci.* **368**, 20130088 (2013).
doi: 10.1098/rstb.2013.0088; pmid: 23754820

85. K. S. Makarova, A. V. Sorokin, P. S. Novichkov, Y. I. Wolf,
E. V. Koonin, Clusters of orthologous genes for 41 archaeal
genomes and implications for evolutionary genomics
of archaea. *Biol. Direct* **2**, 33 (2007). doi: 10.1186/1745-6150-
2-33; pmid: 18042280

86. C. E. Blank, Phylogenomic dating—The relative antiquity of
archaeal metabolic and physiological traits. *Astrobiology* **9**,
193–219 (2009). doi: 10.1089/ast.2008.0248;
pmid: 19371161

87. M. Groussin, M. Gouy, Adaptation to environmental
temperature is a major determinant of molecular

evolutionary rates in archaea. *Mol. Biol. Evol.* **28**, 2661–2674 (2011). doi: 10.1093/molbev/msr098; pmid: 21498602

88. P. López-García, D. Moreira, Open questions on the origin of eukaryotes. *Trends Ecol. Evol.* **30**, 697–708 (2015). doi: 10.1016/j.tree.2015.09.005; pmid: 26455774

89. N. Iwabe, K. Kuma, M. Hasegawa, S. Osawa, T. Miyata, Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9355–9359 (1989). doi: 10.1073/pnas.86.23.9355; pmid: 2531898

90. T. A. Williams, P. G. Foster, C. J. Cox, T. M. Embley, An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* **504**, 231–236 (2013). doi: 10.1038/nature12779; pmid: 24336283

91. J. A. Lake, E. Henderson, M. Oakes, M. W. Clark, Eocytes: A new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3786–3790 (1984). doi: 10.1073/pnas.81.12.3786; pmid: 6587394

92. C. M. Klinger, A. Spang, J. B. Dacks, T. J. G. Ettema, Tracing the archaeal origins of eukaryotic membrane-trafficking system building blocks. *Mol. Biol. Evol.* **33**, 1528–1541 (2016). doi: 10.1093/molbev/msw034; pmid: 26893300

93. J. Surkont, J. B. Pereira-Leal, Are there Rab GTPases in Archaea? *Mol. Biol. Evol.* **33**, 1833–1842 (2016). doi: 10.1093/molbev/msw061; pmid: 27034425

94. M. P. Rout, M. C. Field, The evolution of organellar coat complexes and organization of the eukaryotic cell. *Annu. Rev. Biochem.* **86**, 637–657 (2017). doi: 10.1146/annurev-biochem-061516-044643; pmid: 28471691

95. E. A. Eloe-Fadrosh, N. N. Ivanova, T. Woyke, N. C. Kyrpides, Metagenomics uncovers gaps in amplicon-based detection of microbial diversity. *Nat. Microbiol.* **1**, 15032 (2016). doi: 10.1038/nmicrobiol.2015.32; pmid: 27572438

96. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002). doi: 10.1093/nar/gkf436; pmid: 12136088

97. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009). doi: 10.1093/bioinformatics/btp348; pmid: 19505945

98. L. T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015). doi: 10.1093/molbev/msu300; pmid: 25371430

99. V. B. Borisov, R. B. Gennis, J. Hemp, M. I. Verkhovsky, The cytochrome bd respiratory oxygen reductases. *Biochim. Biophys. Acta* **1807**, 1398–1413 (2011). doi: 10.1016/j.bbabio.2011.06.016; pmid: 21756872

100. J. G. Ferry, How to make a living by exhaling methane. *Annu. Rev. Microbiol.* **64**, 453–473 (2010). doi: 10.1146/annurev.micro.112408.134051; pmid: 20528692

101. S. W. Ragsdale, E. Pierce, Acetogenesis and the Wood-Ljungdahl pathway of $CO_2$ fixation. *Biochim. Biophys. Acta* **1784**, 1873–1898 (2008). doi: 10.1016/j.bbapap.2008.08.012; pmid: 18801467

102. K. Schuchmann, V. Müller, Autotrophy at the thermodynamic limit of life: A model for energy conservation in acetogenic bacteria. *Nat. Rev. Microbiol.* **12**, 809–821 (2014). doi: 10.1038/nrmicro3365; pmid: 25383604

103. G. Fuchs, Alternative pathways of carbon dioxide fixation: Insights into the early evolution of life? *Annu. Rev. Microbiol.* **65**, 631–658 (2011). doi: 10.1146/annurev-micro-090110-102801; pmid: 21740227

# Science

## Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life

Anja Spang, Eva F. Caceres and Thijs J. G. Ettema

**Archaeal diversity and evolution**
Archaea are prokaryotes that make up a third branch of the tree of life. Knowledge of archaeal biological diversity and their role in evolution has rapidly expanded in the past decade. Despite the discovery of previously unknown groups and lineages, few lineages have been well studied. Spang *et al.* review the diversity of Archaea and their genomes, metabolomes, and history, which clarifies the biology and placement of recently discovered archaeal lineages.
*Science*, this issue p. eaaf3883

| | |
|---|---|
| **ARTICLE TOOLS** | http://science.sciencemag.org/content/357/6351/eaaf3883 |
| **REFERENCES** | This article cites 103 articles, 22 of which you can access for free<br>http://science.sciencemag.org/content/357/6351/eaaf3883#BIBL |
| **PERMISSIONS** | http://www.sciencemag.org/help/reprints-and-permissions |

Use of this article is subject to the Terms of Service