nature microbiology

REVIEW ARTICLE

https://doi.org/10.1038/s41564-020-0715-z

Diversity, ecology and evolution of Archaea

Brett J. Baker[1]✉, Valerie De Anda[1], Kiley W. Seitz[1], Nina Dombrowski[1], Alyson E. Santoro[2] and Karen G. Lloyd[3]

Compared to bacteria, our knowledge of archaeal biology is limited. Historically, microbiologists have mostly relied on culturing and single-gene diversity surveys to understand Archaea in nature. However, only six of the 27 currently proposed archaeal phyla have cultured representatives. Advances in genomic sequencing and computational approaches are revolutionizing our understanding of Archaea. The recovery of genomes belonging to uncultured groups from the environment has resulted in the description of several new phyla, many of which are globally distributed and are among the predominant organisms on the planet. In this Review, we discuss how these genomes, together with long-term enrichment studies and elegant in situ measurements, are providing insights into the metabolic capabilities of the Archaea. We also debate how such studies reveal how important Archaea are in mediating an array of ecological processes, including global carbon and nutrient cycles, and how this increase in archaeal diversity has expanded our view of the tree of life and early archaeal evolution, and has provided new insights into the origin of eukaryotes.

Initially, all single-celled, non-eukaryotic microorganisms were classified as the 'Prokaryota'[1] and, later, the 'Monera'[2]. These early classifications lumped Bacteria and Archaea into a single group based primarily on morphology. The development of nucleic acid-based comparative and phylogenetic analyses by Carl Woese and colleagues provided the first glimpses into the evolutionary relatedness of microorganisms[2,3]. These methods showed that life should be divided into three domains: Eukarya, Eubacteria and Archaebacteria (with the last two being later renamed Bacteria and Archaea)[4–6]. A closer examination of Archaea revealed many genetic and biochemical similarities with Eukarya, which led Woese to controversially propose Archaea to be more closely related to Eukarya than to Bacteria[7]. Some of the features lacking in Bacteria and shared among Archaea and Eukarya include the presence of histones, complex RNA polymerases and methionine translation initiation[8]. As we will review here, the list of similarities between these domains is growing with the discovery of new archaeal taxa. Archaea are as distinct from Bacteria as eukaryotes are; each of the three domains have their own unique genetic, biochemical and cellular characteristics. In fact, Archaea may have more in common with eukaryotes than Bacteria, such as genomic structure (for example, introns, histones and multiple origins of replication), transcriptional and translational machinery, and their lack of peptidoglycan in cell walls.

Particular innovations in DNA amplification, such as up-scaling of the polymerase chain reaction (PCR) and lowering the cost of sequencing, enabled the generation of large numbers of the 16S ribosomal RNA (rRNA) gene sequences from the environment. Ribosomal genes are evolutionarily conserved in all known lifeforms and are therefore ideal marker genes that can be used for phylogenetic and taxonomic identification of distantly related organisms. This led to the discovery of a vast, uncultured microbial diversity in nature and the realization that our view of microbiology is often limited by our inability to cultivate microorganisms[9–16]. Initially, archaeal rRNA genes fit into two primary phyla, the Euryarchaeota and Crenarchaeota[10,17,18]. However, broader sampling of environments began to reveal deep evolutionary branches that did not fall within these two phyla, such as the thermophilic Korarchaeota[19]. Even within the Crenarchaeota and Euryarchaeota, 16S rRNA gene sequencing revealed many deeply branching groups[19]. This uncultivated diversity is commonly referred to as 'microbial dark matter'.

In the past ten years, several technological advances in DNA sequencing and computational approaches have enabled the rapid reconstruction of genomes from nature, including the description of 20 phyla within the Archaea[20]. Specifically, these advances include high-throughput sequencing, metagenomic assembly and binning (that is, the grouping of DNA fragments into metagenome-assembled genomes; MAGs), as well as the ability to amplify and sequence DNA from a single microbial cell[21–25]. The accelerated reconstruction of archaeal genomes has led to the discovery of several new taxonomic groups, in addition to those that had been detected in rRNA diversity surveys. More importantly, the frequent recovery of these lineages illustrates how common, yet unknown, these organisms are. These genomes have enabled a more robust comparison of diversity and evolutionary histories via multi-protein phylogenomics, as well as a more detailed analysis of the metabolism of uncultivated Archaea (Box 1)[26,27]. This rate of discovery has shown no signs of slowing; in fact, by the time this review is published, there will surely be additional phyla added to the tree of life. Looking at the number of available archaeal genomes, it becomes clear that we have a long way to go before we have a comprehensive sampling of many of the lineages (Fig. 1). There is a large positive bias in the number of genomes that have been obtained from phyla with cultured representatives. Nevertheless, this broader taxonomic sampling of archaeal diversity has allowed the clustering of several archaeal phyla into larger taxonomic groups that are now referred to as superphyla (Table 1). Currently, there are three named superphyla recognized: Asgard, DPANN and TACK. DPANN and TACK Archaea were originally named based on the phyla present in them: DPANN referring to the Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanohaloarchaeota and Nanoarchaeota; and TACK to the Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota[28,29] (Table 1). These superphyla now include several additional phyla which have been discovered since they were first named, which we will detail in this review. At present one phylum, the Euryarchaeota, does not fall within a superphylum.

[1]Department of Marine Science, Marine Science Institute, University of Texas Austin, Port Aransas, TX, USA. [2]Ecology, Evolution and Marine Biology, University of California, Santa Barbara, CA, USA. [3]Microbiology Department, University of Tennessee, Knoxville, TN, USA. ✉e-mail: brett_baker@utexas.edu

Box 1 | Resolving biodiversity and evolutionary histories with phylogenomics

Innovations in sequencing technologies and computational tools have made it possible to reconstruct large numbers of genomes from nature. Notably, among these tools are improved metagenomic assemblers[21,171] and more accurate binning approaches[24,172–175]. These approaches result in partial to complete genomes, with many of the missing regions explained by difficulties in assembling short-read genomic data (for example, repeats) rather than signifying truly incomplete genomes. This has led to an explosion of new archaeal genomes and has provided a framework to build more character-rich and robust phylogenetic analyses. 16S rRNA gene phylogenies alone have proven inadequate in resolving branching patterns among phyla[176], thus multi-loci concatenated trees needed to be employed. It has become common to use a reliable subset of ribosomal proteins, and other conserved proteins such as tRNA synthetases, to delineate phyla and determine relatedness of phyla within superphyla[27,88,177,178]. These proteins have been chosen for several reasons: (1) for their limited lateral gene transfer[179], (2) for their consistency in the clustering of phyla with rRNA gene phylogenies and (3) for their broad distribution throughout the tree of life. To date, three phylogenomic clusters of archaeal phyla (superphyla) have been described using ribosomal protein phylogenies; Asgard, TACK and DPANN. As detailed in this Review, a total of 27 phyla have been proposed based on concatenated protein phylogenies. In addition to classifying diversity, this approach has enabled the resolution of deep-branching topology, for example the relatedness of Archaea to eukaryotes[148,151,179]. To update the archaeal tree of life we obtained a comprehensive set of 3,599 uncultured and cultured archaeal genomes and generated robust phylogenomic analyses using a set of conserved ribosomal marker proteins (see Fig. 4).

In addition to resolving our view of the diversity of the Archaea, the recovery of these genomes has enhanced our understanding of the metabolic and ecological roles of the Archaea. Previously, our understanding of archaeal physiology was limited to cultured representatives. However, uncovering additional genomes from uncultivated Archaea has revealed a variety of new ecological roles for Archaea, such as in carbon and nutrient cycling (Figs. 2,3), whose characterization is being aided by cultivation studies and environmental measurements. A greater genomic sampling of the tree of life has also provided insights into the evolutionary histories of the Archaea and their role in the origin of complex cellular life, as discussed below in more detail.

Euryarchaeota

The first published archaeal isolate of the order Methanobacteriales was obtained from cattle rumen in 1958 (refs. [4,30]), and a member of the class Thermoplasmata was isolated in 1970 by Thomas Brock's laboratory[31], several years before Archaea were first described by Carl Woese, George Fox and others[4,32]. Soon after, several other representatives of methanogens were obtained[33]. Now, the Euryarchaeota contain the greatest number and diversity of cultured lineages. In addition to methanogens, which comprise several named classes (Methanobacteria, Methanococci, Methanomicrobia, Methanonatronarchaeia and Methanopyri), members from several other classes have been cultured, for example from halophilic (that is, Halobacteria)[34], acidophilic (Thermoplasmata) or thermophilic environments (Archaeoglobi, Methanopyri and Thermococci)[35]. The discovery of extremophiles belonging to the classes Thermoplasmata, Thermococci and Archaeoglobi has been particularly instrumental in advancing our understanding of the biology of this phylum as they have provided numerous cultured representatives for physiological studies and genomic analyses[36]. Also, the discovery of these Archaea has advanced our understanding of the biological adaptations of living in extreme conditions. For example, a species of Methanopyri (*Methanopyrus kandleri*) has been shown to maintain growth in temperatures up to 122 °C and high pressures of 20 MPa[37], which has extended the upper temperature limit of life on the planet. In recent years, in situ geochemical and 'omic' studies ('omic' refers to characterization at genomic, transcriptomic and proteomic levels) have revealed that Euryarchaeota are not just involved in methane production, but also anaerobic methane oxidation[38]. Furthermore, it is becoming apparent that these Archaea are also involved in the anaerobic oxidation of other short-chain hydrocarbons[39], suggesting that these microbes have much more varied roles in biogeochemical cycles than previously thought.

New branches of Euryarchaeota life. While Archaea are often considered rare members of the microbial biosphere, uncultured euryarchaeotes have been shown to be dominant members of several ecosystems. For example, Marine Group II and III (MG-II and MG-III; now referred to as Poseidoniales and Pontarchaea) Archaea occasionally constitute a large fraction (at times up to 40%) of marine microbial communities[40–42]. Another example is the South-African Gold Mine Miscellaneous Euryarchaeal Group (SAGMEG; now named Hadesarchaea), which are widespread in both terrestrial and subseafloor environments[43]. Culture-independent studies have shown how important these and other prominent lineages are for global biogeochemical cycling, highlighting how the recovery of SAGMEG genomes and other deep-branching classes have expanded our understanding of biodiversity and ecology of this phylum.

Several genomes have been obtained belonging to distinct lineages within the Euryarchaeota. For example, genomes from a lineage designated the Marine Benthic Group E (MBG-E)[44] were recovered from a subseafloor igneous basement. These genomes have been proposed to be a separate phylum, now named Hydrothermarchaeota[44,45]. However, their placement within the archaeal tree of life remains debated and different phylogenetic analyses suggest they are monophyletic with the Euryarchaeota (Fig. 4) or even branch closely to the DPANN Archaea[46]. Other Euryarchaeota lineages, such as the Hadesarchaea and Theionarchaea, have been suggested to be distinct phyla rather than classes[47]. Nonetheless, these deep-branching groups possess metabolic commonalities with other Euryarchaeota including the presence of F420-hydrogenases, CO monooxygenases and the Wood–Ljungdahl pathway[43,48]. Also, previous phylogenies for Euryarchaeota have proposed the inclusion of the Altiarchaeales[49], which have recently been shown to constitute the phylum Altiarchaeota instead[50] and are debated to be affiliated with the DPANN Archaea (see DPANN superphylum section below). Many other uncultured branches based on rRNA genes exist within the Euryarchaeota, especially from oceanic waters and sediments, and will likely soon have genomic data as well[51].

Ecological roles of the Euryarchaeota. Euryarchaeotes are classically known as methanogens, however, some Euryarchaeota lineages are also involved in syntrophic anaerobic methane oxidation; these lineages are referred to as anaerobic methane oxidizers (ANME)[52]. ANME Archaea use the core pathway for methane production in the reverse direction[53] which depends on a bacterial partner, usually a sulfate reducer[52], denitrifier[54] or even a nitrate reducer[55], to make it energetically favourable. A close physical association of these syntrophic partners in nature makes it possible to couple these redox reactions, exchange electrons and obtain energy[56,57].

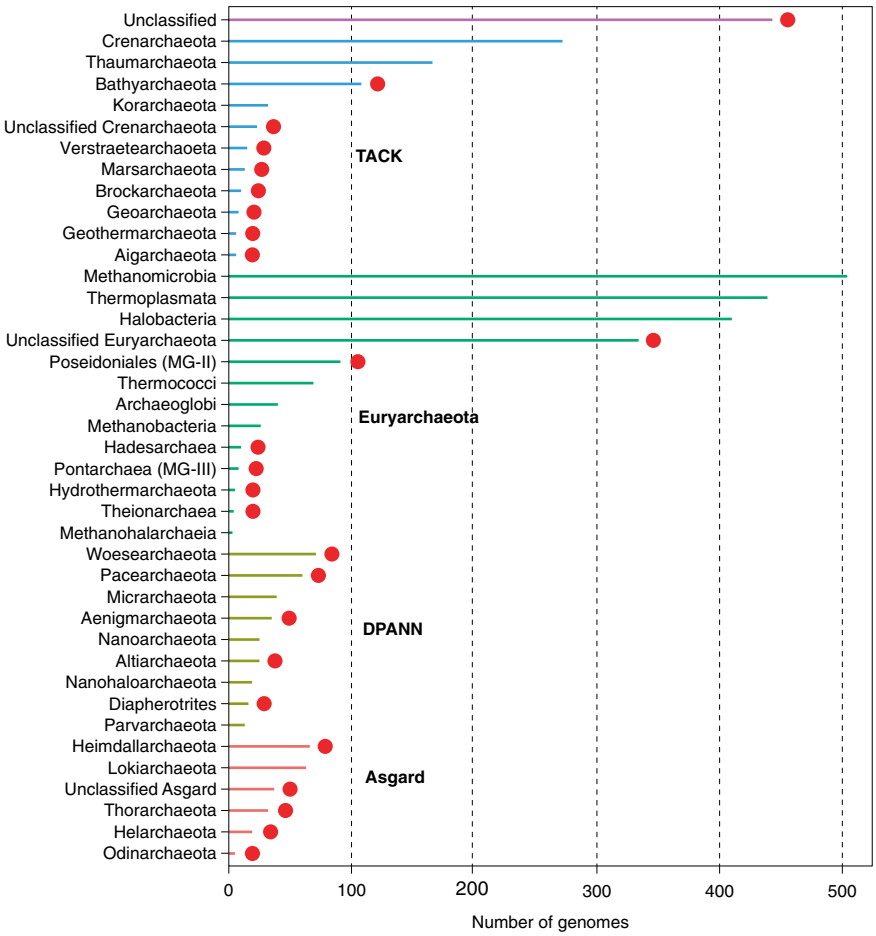

Fig. 1 | Archaeal genomic diversity. Genomic diversity that has been obtained among different archaeal taxonomic groups so far, based on a database of 3,599 archaeal genomes (also used in the phylogeny in Fig. 4; a complete list of these genomes is provided in the Supplementary Data). Metagenome-, single- and isolate-assembled genomes were included. For each individual genome, taxonomy was obtained from their National Center for Biotechnology Information taxonomy ID. Those with red dots are groups which have no cultivated members. Taxa bars are coloured based on the phyla and superphyla, labelled on the right, to which they belong.

In addition to their ability to metabolize methane, recent studies have suggested that some euryarchaeotes may also degrade other hydrocarbons. For example, two genomes belonging to the uncultured lineage referred to as the GoM-Arc87 clade were found to contain gene encoding pathways similar to those used for methanogenesis, including a homologue of the key enzyme methyl-coenzyme M reductase (MCR)[58]. Enrichments containing GoM-Arc87 and Bacteria of the HotSeep-1 cluster were revealed to be involved in syntrophic anaerobic oxidation of butane[58]. GoM-Arc87 was therefore re-named *Ca.* Syntrophoarchaeum. This was the first experimental evidence suggesting that MCR proteins have a broader substrate specificity than methane (Fig. 5). Interestingly, the genomic reconstruction of uncultured GoM-Arc1 archaea, commonly found in hypersaline methane seeps[59,60], showed that these microorganisms have *mcr* genes that are phylogenetically distinct from those involved in butane or methane oxidation[61] (Fig. 5). Metabolic reconstruction of GoM-Arc1 revealed that they lack the butane (butyryl-coenzyme A) oxidation pathway and contain a novel, methanogenesis-like pathway. Based on these findings, it was hypothesized that they are involved in the oxidation of another hydrocarbon, likely ethane[61,62], which was later confirmed by enriching these Archaea in a ten-year culturing effort[63]. Recently, it has been shown that a single organism, '*Candidatus* Methanoliparia', contains both methane- and non-methane alkane-types of *mcr* genes[58]. This suggests that these Archaea are capable of producing methane from alkanes via disproportionation.

Apart from their roles in the transformation of hydrocarbons, euryarchaeotes are also commonly involved in sulfur, nitrogen and iron cycling[64]. For example, two new classes within the Euryarchaeota with roles in nitrite and sulfur reduction have been recently described. These comprise Hadesarchaea (formerly SAGMEG[43,65]) and Theionarchaea (formerly designated Z7ME43)[48]. The Hadesarchaea were first identified in the deep terrestrial subsurface and have since been detected in a variety of subseafloor environments[66–69]. Hadesarchaea have relatively streamlined genomes and, based on gene content, may be coupling carbon monoxide and H_2 oxidation to nitrite reduction. A sister lineage to Hadesarchaea, formerly Mediterranean Sea Brine Lakes Group 1 (MSBL-1), has been genomically characterized from brine pools from the Red Sea and proposed to be named Persephonarchaea[20]. Theionarchaea genomes were reconstructed from estuary sediments and have been implicated in intermediate sulfur (S^0, polysulfide and thiosulfate) reduction[20,48].

There have also been several recent examples of different lineages being involved in the degradation of organic matter (such as proteins), suggesting that they occupy a variety of niches in nature. The first near-complete archaeal genomes to be reconstructed from metagenomic datasets were members of the Thermoplasmatales

Table 1 | List of proposed archaeal phyla and their original designations

Superphylum	Phylum	Previous designation	References
	Euryarcheota	Numerous	7
	Hydrothermarchaeota	MBG-E	45
Asgard	Lokiarchaeota	MBG-B and DSAG	148
Asgard	Thorarchaeota	Previously undescribed	48
Asgard	Odinarchaeota	Previously undescribed	151
Asgard	Heimdallarchaeota	AAG and MHVG	151
Asgard	Helarcheota	Previously undescribed	152
DPANN	Micrarchaeota	ARMAN-1 and -2	88,86
DPANN	Parvarchaeota	ARMAN-4 and -5	88,86
DPANN	Pacearcheota	DHVEG-6	88
DPANN	Aenigmarchaeota	DSEG	29
DPANN	Diapherotites	pMC2A384	29
DPANN	Woesearchaeota	Previously undescribed	88
DPANN	Altiarchaeota	SM1	49
DPANN	Nanoarchaeota	Previously undescribed	49,83
DPANN	Nanohaloarchaeota	Previously undescribed	29,87
DPANN	Huberarchaeota	Previously undescribed	49
TACK	Korarchaeota	Previously undescribed	19
TACK	Verstraetearchaeota	TMCG	45,100
TACK	Nezhaarchaeota	Previously undescribed	39
TACK	Crenarchaeota	Many	7
TACK	Aigarchaeota	HWCG-I	169
TACK	Thaumarchaeota	MG-I, MBG-A	97
TACK	Bathyarchaeota	MCG	135
TACK	Geothermarchaeota	THSCG	45
TACK	Geoarcheota	NAG1	107
TACK	Marsarchaeota	NAG2	170

AAG, ancient archaeal group; DHVEG, deep-sea hydrothermal vent group; DSEG, deep-sea euryarchaeotal group; SBAR, Santa Barbara Archaea; SM1, Sippenauer Moor; HWCG, hot water crenarchaeotic group.

order (referred to as A-I plasma), obtained from acid mine drainage[22]. Their marine counterpart, *Aciduliprofundum boonei*, had been known only by 16S rRNA gene sequences as the abundant Deep-Sea Hydrothermal Vent Euryarchaeota 2 (DHVE2; refs. [24,70]) group, until it was isolated and shown to be an obligately thermoacidophilic sulfur- or iron-reducing heterotroph capable of degrading proteins for energy[71]. The MBG-D[44], also called DHVE1 (ref. [71]), often appear in marine sediments including methane seeps[72,73] as well as some anoxic terrestrial environments[74]. Names proposed for this group include Izemarchaea[20] and Thermoprofundales[75]. Partial genomes for MBG-D have been obtained from single-cell genomics and were shown to contain a high content of predicted extracellular peptidases whose activity can be directly assayed in anoxic marine sediments[73].

MG-II and MG-III Archaea are commonly found in oceanic waters worldwide[10,40,76,77] and have recently been designated Poseidoniales[29] and Pontarchaea[20,78]. Genomes for MG-II and MG-III have been obtained from various metagenomic reconstructions and are suggested to be involved in the recycling of high molecular weight organic compounds[79–81]. Some MG-II organisms contain rhodopsins which are predicted to use light to boost energy yield or facilitate substrate transport, and are also capable of protein degradation[41,82].

DPANN superphylum

While cultivating *Ignicoccus hospitalis* from hydrothermal vents, Karl Stetter's laboratory found small (~400 nm) archaeal cells attached to *Ignicoccus* that were named *Nanoarchaeum equitans*[83]. Phylogenetic analyses using concatenated ribosomal proteins revealed this species belongs to a new phylum: Nanoarchaeota[83,84]. The genome of *N. equitans* is one of the smallest known genomes (490 kb) and lacks many essential genes including lipid, cofactor, amino acid and nucleotide biosynthesis genes[84], suggesting that *N. equitans* is an obligate symbiont and is dependent on its host for essential molecules.

Several years later, during one of the first community level analyses of metagenomic assemblies from an acid mine drainage site, DNA fragments containing novel 16S rRNA genes from other small Archaea were recovered[85]. Filtration enrichment and microscopic identification of these cells revealed that they were also much smaller than other Archaea (<500 nm) and similar to those of Nanoarchaeota. Thus, they were referred to as acidophilic Richmond Mine archaeal nanoorganisms (ARMAN). Interestingly, it was shown that the common PCR primers used for diversity surveys overlooked the ARMAN lineages[85]. Using filtration to enrich for ARMAN and one of the first employments of metagenomic binning, three near-complete genomes of ~1 Mb in size[85,86] were obtained. This small genome size and other features similar to the ones found in *N. equitans* suggest they too are symbionts in some capacity.

Recently, ARMAN and Nanoarchaeota have been shown to be part of a diverse superphylum called DPANN[29]. At the time, DPANN included Diapherotites (formerly pMC2A384), Parvarchaeota

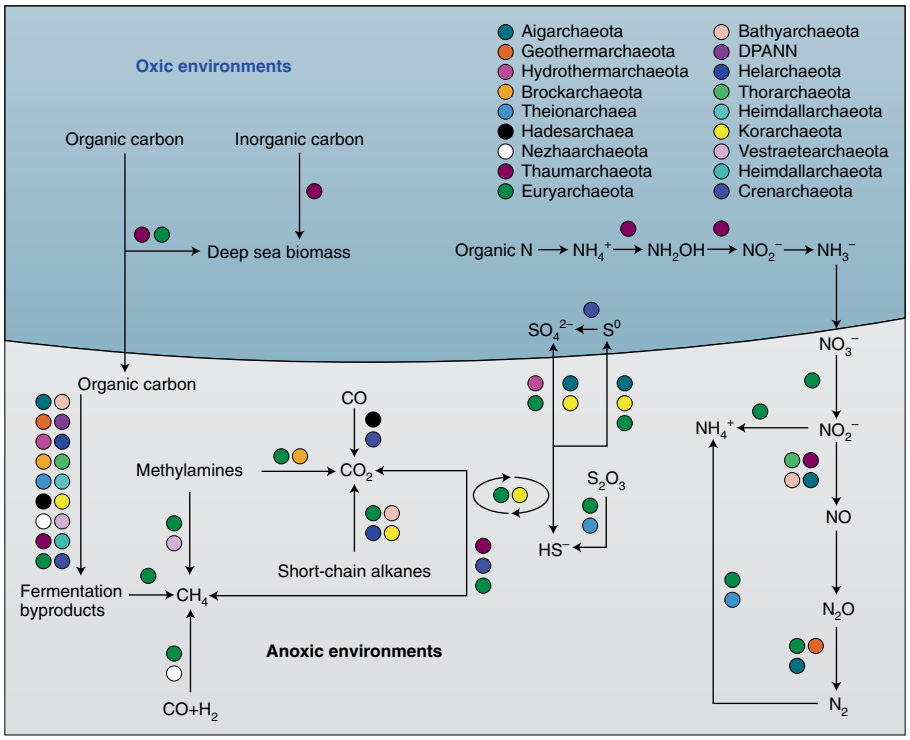

Fig. 2 | Archaeal ecological roles. Overview of key biogeochemical roles in carbon, nitrogen and sulfur cycles attributed to different archaeal lineages in both oxic and anoxic environments. As detailed in the text, some of these processes have been confirmed via cultivation and supported by environmental measurements, while others are based solely on the presence of metabolic pathways.

(ARMAN-4 and -5), Aenigmarchaeota, Nanohaloarchaeota and Nanoarchaeota[29]. Two Nanohaloarchaeota genomes (~1.2 Mb in size) were first reconstructed from saline Lake Tyrell and their cells were estimated to be ~600 nm in diameter[87]. A broader genomic sampling of DPANN from groundwater has considerably expanded the known diversity of the DPANN superphylum to include Woesearchaeota, Pacearchaeota[88], Huberarchaeota[49,89] and several uncharacterized archaeal phyla (UAP1 and UAP2 (ref. [90])). Increasing the phylogenetic diversity also revealed that the ARMAN are in fact two distinct phyla, Micrarchaeota (groups 1–3) and Parvarchaeota (groups 4 and 5), thus adding Micrarchaeota as a new phylum to the DPANN Archaea[88]. Altiarchaeota (formerly named Sippenauer Moor (SM1) Euryarchaeota) were first assigned to the Euryarchaeota phylum[50]. Initially, however, trees did not include DPANN Archaea and updated phylogenies now suggest that Altiarchaeota is a phylum that branches with the DPANN, even though their exact placement is still debated[49,91]. Similarly, the branching positions of the DPANN Archaea affect the discussion around the root of the archaeal tree (Box 2).

Ecological roles of DPANN Archaea. Archaea belonging to the DPANN superphylum had previously been overlooked by diversity surveys—partly due to their unique rRNA gene composition[85]—until they were fortuitously observed in cultures of *I. hospitalis* and genomic reconstructions from natural samples[83]. When first identified, it was thought that they were restricted to extreme acidic and hydrothermal environments; however, we now know that they are ubiquitous in nature[29,88]. The presence of DPANN in diverse environments suggests they play important ecological roles which are just beginning to be realized. Rapid genomic expansion of this superphylum has revealed that limited genetic and metabolic capabilities are a shared feature of most DPANN Archaea and that they therefore must rely on interactions with other Archaea to obtain essential biomolecules. Even though the DPANN are metabolically constrained, some of them have the genomic capacity to utilize organic compounds such as carbon and lipids, via glycolysis, beta oxidation and other pathways[86,88].

Reduced genomes and symbiotic lifestyles. There are several unique features that have been identified in DPANN genomes that are thought to be signatures of symbiotic relationships. In addition to having small genomes and lacking a variety of biosynthetic pathways, they have unique introns in their 16S rRNA and transfer RNA genes, and relatively short average gene lengths, which may be a result of genomic reduction[86]. Microscopic examination of ARMAN in biofilms revealed that they are occasionally connected with other Archaea belonging to Thermoplasmatales[86]. A stable co-culture containing Micrarchaeota and Thermoplasmatales have been obtained, opening the door to advancing our understanding of these interactions[92]. Metatranscriptomic analyses of this co-culture suggest that the ARMAN primarily use amino acids. Additionally, several other interaction partners have been identified in the DPANN. For example, *Ca*. Huberarchaeum crystalense potentially interact with members of the Altiarchaeales based on co-varying cell abundance profiles and microscopic imaging[50,89]. Two strains of the Nanohaloarchaeota belonging to *Ca*. Nanohaloarchaeum antarcticus were cultured and found to be reliant on *Halorubum lacusprofundi* for survival in Antarctic hypersaline environments[93].

In natural biofilms, Thermoplasmatales cells form long appendages that penetrate the cell walls of ARMAN; the purpose of the connection is not known[86]. Interestingly, ARMAN have also been shown to contain internal tubular membrane-bound structures[94], suggesting that they may have some form of compartmentalization. Recently, an endomembrane system has also been identified in *I. hospitalis* that is in direct contact with *N. equitans*[95]. In addition to appendages forming connections between ARMAN and other Archaea, Altiarchaeota create unique barbed wire appendages with tripartate grappling

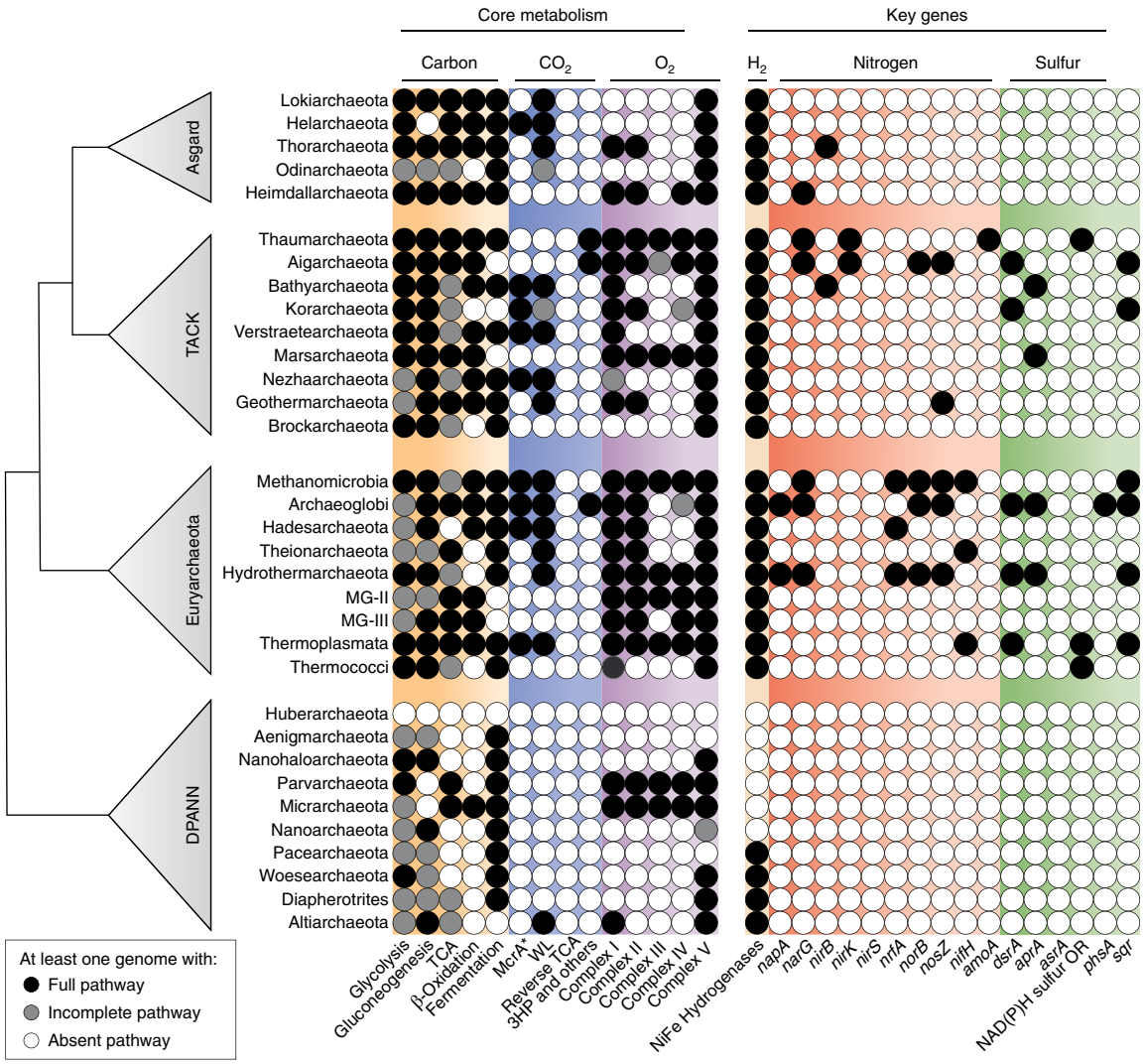

Fig. 3 | Archaea metabolic potential. Physiological capabilities of archaeal lineages based on gene content, with distribution of metabolic pathways in archaeal phyla that have been described to date. In many instances among uncultured lineages the processes have not been confirmed and are based solely on the presence of pathways in the genomes. The asterisk includes canonical methane as well as alkane-utilizing McrA proteins. TCA, tricarboxylic acid cycle; WL, Wood–Ljungdahl pathway; 3HL, 3-hydroxypropionate.

hooks that appear to be involved in attachment for biofilm formation[50] and are missing in marine sediment clades[91]. These observations demonstrate that there is much to be learned about the different modes of interactions between DPANN and other Archaea.

Notably, it has also been shown that ARMAN in mine biofilms have an abundance of infecting viruses and many have two morphological types of viruses attached to their cells[94]. Intriguingly, the ARMAN cells that were most commonly infected were those that were in physical contact with other Archaea. This suggests that there may be unresolved viral defensive interactions between ARMAN and their host Archaea, Thermoplasmatales. Also, it has recently been shown that uncultured Micrarchaeota and Parvarchaeota contain clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein 9 (CAS9) viral defence systems, which is the first example of this system in the Archaea and may lead to important biotechnological applications[96].

TACK superphylum

Thermophilic and acidophilic Crenarchaeota are among the most readily cultivated Archaea, making this phylum one of the most extensively studied to date. Nevertheless, 16S rRNA gene diversity surveys of natural communities revealed several uncultivated, deeply branching lineages related to the Crenarchaeota[10,19]. These included MG-I Archaea, Miscellaneous Crenarchaeota Group (MCG) Archaea, MBG-A and MBG-E, and Terrestrial Miscellaneous Crenarchaeota Group (TMCG). These groups were commonly referred to as 'mesophilic Crenarchaeota' due to their broader distribution in moderate temperature sediments and ocean water. Their genomes have enabled a more thorough phylogeny and have placed some lineages into new phyla; for example, MG-I is now referred to as Thaumarchaeota[97,98].

In addition to Crenarchaeota and Thaumarchaeota, genomes belonging to Korarchaeota[99] and the recently designated Aigarchaeota have been shown to be monophyletic, and these four phyla have started to be collectively referred to as the TACK superphylum[28,99]. Genomes of additional TACK lineages include the phyla Bathyarchaeota (formerly MCG), Hydrothermarchaeota (formerly MBG-E)[45], Geothermarchaeota (formerly the Terrestrial Hot Spring Crenarchaeota Group; THSCG)[45] and Verstraetearchaeota (formerly TMCG)[100] phyla.

Crenarchaeota. Crenarchaeota contain some of the first cultured species of the archaea from the *Sulfolobus* genus that were isolated

Fig. 4 | An updated archaeal tree of life. Phylogeny generated using 36 conserved marker proteins (mostly ribosomal proteins; extracted using PhyloSift[75]). Names of proposed phyla and uncultured classes of Euryarchaeota are provided. The branches with red dots have no cultured representatives. The phylogeny was generated with 3,549 archaeal genomes (that contain >50% of the marker proteins; Supplementary Data) and 40 bacterial genomes as outgroup using IQtree v.1.6.11 with a best fit LG+F+R10 model selected using the Bayesian Information Criterion (BIC). Bootstraps are based on 1,000 replicated trees. The alignment of 3,589 genomes was generated using MAFFT (v.1.4.0, algorithm autoselection) with a BLOSUM62 scoring and contains 4,962 characters after masking gaps present in at least 50% of the taxa.

from hot springs in Yellowstone National Park[81]. Subsequently, many more cultures have been obtained entirely from high temperature terrestrial and marine locations. Cultures of the Crenarchaeum *Pyrolobus fumarii* are hyperthermophiles capable of growing in temperatures up to 113 °C[101]. Crenarchaea were initially classified to be a distinct phylum from other archaea based on 16S rRNA gene phylogenies and the absence of histones, however, later studies found histones in some Crenarchaeotes[102].

Many Crenarchaea are anaerobic heterotrophs, utilizing proteins and sugars, while others are sulfur (oxidation and reduction)-cycling chemolithoautotrophs. The most extensively studied Crenarchaeotes are *Sulfolobus* spp. which were originally isolated

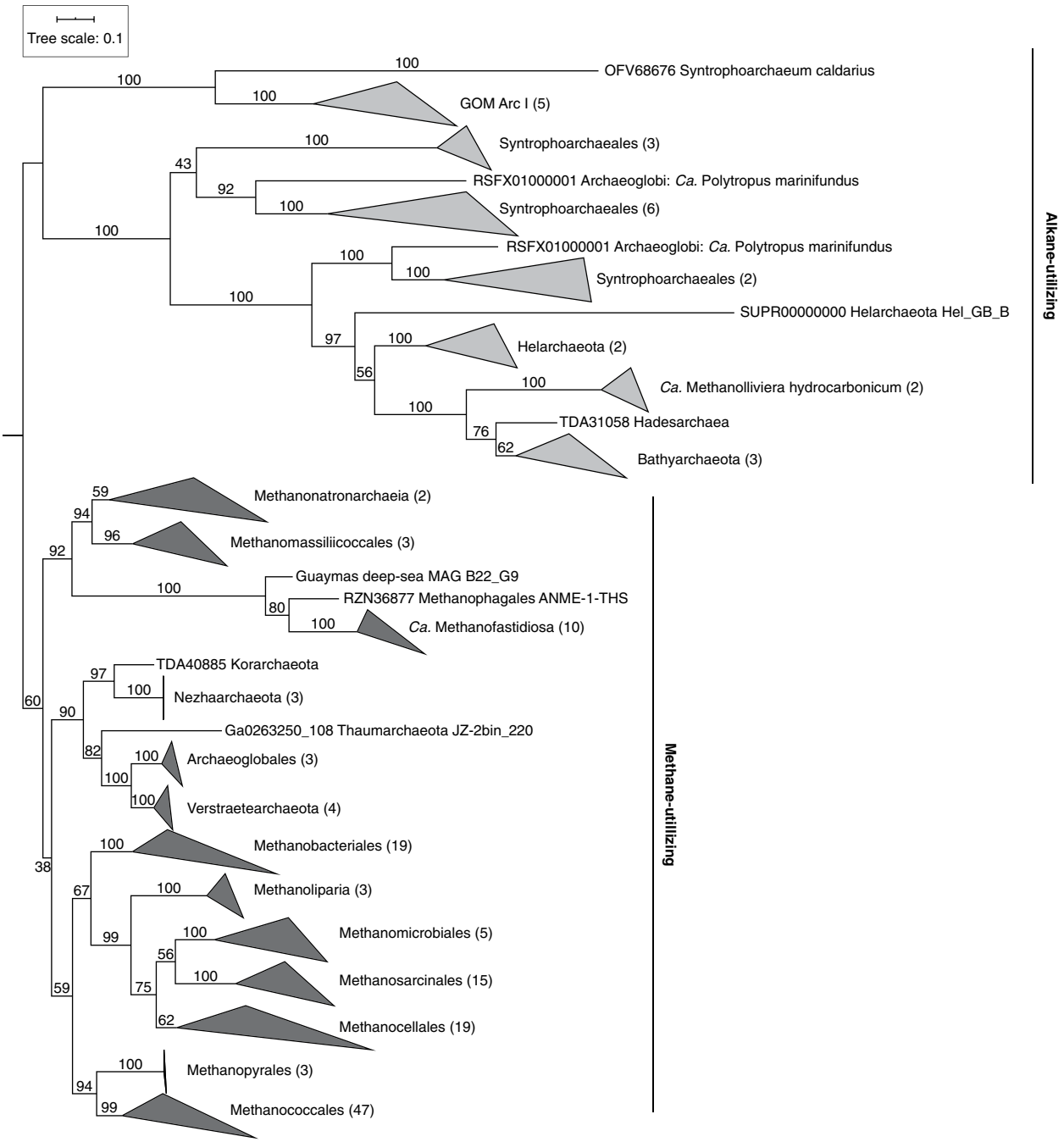

Fig. 5 | Diversity of archaeal MCR proteins. Phylogeny of MCR subunit A (McrA) proteins from several archaeal lineages. Generated using RAxML method (v.8.0.0, RAxMLHPC-PTHREADS-AVX with a gamma substitution model) using 164 protein sequences. The auto-bootstrapping option was used to bootstrap values which are displayed on the nodes. Two hundred iterations were performed. The numbers in parentheses are the numbers of sequences in those collapsed clades.

from sulfur-rich acidic hot springs in Yellowstone National Park by Thomas Brock and co-workers[103] and have become model Archaea for studies of hyperthermophiles and viral interactions[104]. The majority of these lineages were cultured from sulfur-rich, hot environments and thus they are primarily thermophiles and hyperthermophiles[105,106]. The cultured representatives within the Sulfolobales comprise the main lineages within the Crenarchaeota, with the exception being those belonging to Novel Archaeal Group 1 (NAG1) that were reconstructed from Yellowstone National Park hot springs[107]. Initially, NAG1 was proposed to be a new archaeal phylum that was named Geoarchaeota; however, a more thorough phylogenomic examination revealed they are actually deeply branched in the Thermoproteales[108].

Thaumarchaeota. MG-I were originally thought to be members of the Crenarchaeota, however, phylogenomic analyses revealed they form a distinct phylum named Thaumarchaeota[97,98]. Metagenomic characterization of the Sargasso Sea water column yielded a DNA fragment belonging to MG-I that contained ammonia monooxygenase genes, suggesting that they were involved in ammonia oxidation to nitrite (the first step of nitrification[23]), a suggestion supported by metagenomic fragments from soil 'crenarchaea'[109]. A near-complete

Box 2 | Resolving basal lineages of Archaea

Finding the root of the archaeal tree is essential to answer burgeoning questions about the metabolism and evolution of early cells. Despite this, the root is far from resolved and several positions have been proposed, for example between the Euryarchaeota and the TACK archaea[180]. However, newer studies place the root between DPANN Archaea and all other Archaea[181], which is often confirmed in trees that choose Bacteria as an outgroup[27,88]. Challenging this placement is the fact that host association often coincides with genome reduction and an elevated rate of sequence evolution[182]. For example, in depth analyses of the *N. equitans* genome have shown biases in codon and amino acid usage[183] and potentially higher rates of horizontal gene transfer with their hosts[183,184]. These characteristics could create artefacts in phylogenetic analyses, such as long-branch attraction, and are argued to artificially group DPANN together as a monophyletic, deep-branching clade. Whether fast evolutionary rates and compositional biases are characteristics of all DPANN Archaea still has to be determined, and more robust phylogenomic analyses need to be performed to resolve the true branching positions of the DPANN. Are their basal positions in the tree of life accurate or artefacts of similar symbiotic lifestyles?

genome of a symbiotic member of this group, *Cenarchaeum symbiosum*, which inhabits marine sponges, revealed that they also contain a complete pathway for carbon fixation[110], now shown to be among the most energetically efficient carbon fixation pathways[111]. In 2005, the first culture of MG-I was obtained[112] and the genome of this isolate, *Nitrosopumilus maritimus*, was later sequenced[113]. This revealed that some MG-I are able to oxidize ammonia at the low concentrations found in the open ocean[114]. Furthermore, Thaumarchaeota are able to obtain ammonia from urea and cyanate[115,116]. Therefore, Thaumarchaeota have important links to climate change, as their activity has been linked to the production of the greenhouse gas nitrous oxide (N_2O)[117]. They have also been shown to synthesize methylphosphonate, a potential substrate for aerobic methane production in the nutrient-limited open ocean[118]. Interestingly, some basal lineages containing *mcr* genes but lacking ammonia oxidation and aerobic pathways have been recovered from hot springs[119]. Ammonia-oxidizing Thaumarchaeota are thought to be among the most numerically abundant Archaea on the planet, as they constitute a large proportion of the deep ocean[120] as well as soils[121]. Though they apparently share a common core chemolithoautotrophic metabolism, they can be found in an astonishing range of environmental conditions—from fresh water to salinities over 160 ppt[122], from pH 3.5 to pH 8.7 (ref. [123]) and from the Arctic to hyperthermal environments up to 74 °C[124]. They have been found to dominate oxic marine sediments such as those underneath oligotrophic oceanic gyres[125], although few studies have been conducted on such sediments. Even individual isolates are capable of growth across large ranges of pH and temperature[126]. Major questions still remain about the exact biochemistry of thaumarchaeal ammonia oxidation, including the enzymes involved in oxidation of the intermediate hydroxylamine (NH_2OH) and the role of nitric oxide (NO) as an intermediate[127,128]. The fate of these intermediates likely plays an important role in the production of N_2O during ammonia oxidation, particularly in low oxygen environments[129–131]. Several multicopper oxidases that may also function in nitrogen cycling are synthesized by thaumarchaea, including a putative nitrite reductase (NirK), but these await assignment of precise functional roles[113].

Bathyarchaeota. It has been estimated that half of the microbial cells in the world's oceans inhabit sediments[132]. Although they are not as well-studied as seawater, marine sediments have been shown to contain a considerable diversity of uncultured archaeal lineages[51,68,132]. The MCG often dominate archaeal DNA libraries in nearshore as well as oceanic subseafloor environments[68,125,133]. The first functional genetic information for MCG was based on a fosmid clone library which suggests this group encodes for a bacteriochlorophyll synthesis gene[134]. However, later work suggested this may not be a general feature of MCG, which is unsurprising given that they are most commonly found in deep sediments. However, this does suggest that some members of this group may be capable of light-driven metabolism.

Detailed metabolic reconstructions became possible when a partial genomic reconstruction of MCG was obtained via single-cell genomics[68,73], and MCG was later determined to be a distinct phylum now named Bathyarchaeota[135]. What is astonishing about the Bathyarchaeota is that they are ubiquitous, abundant and highly diverse in anoxic marine, anoxic freshwater and high temperature hot spring locations[68,73,136–138]. Since they are often numerically dominant among the Archaea[133] it can be argued that they, like the Thaumarchaeota, are among the most abundant Archaea on the planet.

Genomes of several subgroups of Bathyarchaeota have been recovered via metagenomics from marine and freshwater sediments around the world[133,139,140]. There is evidence that Bathyarchaeota remineralize detrital proteins, possibly coupling protein degradation to hydrogen production[73,139]. Inference of their physiology based on genomes recovered from deep terrestrial coal beds suggests that they are involved in protein and cellulose degradation as well as CO_2 fixation. A lipid analysis of the MCG in marine sediments supports their ability to fix CO_2 (ref. [141]). Additionally, some lineages contain pathways for methanogenesis as well as phylogenetically distinct *mcr* genes[138,142]. This discovery was the first instance of *mcr* genes occurring outside the Euryarchaeota. Given the recent finding that equally novel Mcr proteins are involved in butane oxidation[58], it is possible that Bathyarchaeota are capable of oxidizing short-chain hydrocarbons other than methane. Recently, enrichments of group 8 Bathyarchaeota were obtained with lignin and inorganic carbon (bicarbonate) as the carbon source, suggesting that they are capable of utilizing recalcitrant organic matter[58,143]. The high diversity of sub-clades within the Bathyarchaeota suggests that new physiologies will continue to be discovered within this uncultured phylum.

The expanding TACK tree of life. Metagenomic reconstructions from anaerobic cellulose digesters and deep terrestrial coal beds resulted in the recovery of the uncultured TMCG[58,100,143]. Phylogenomic analyses of these genomes, named *Ca.* Methanomethylicus spp., revealed they are a distinct phylum named Verstraetearchaeota. Interestingly, Methanomethylicus have pathways involved in methylotrophic methane production, providing further support for the intriguing idea that capabilities for archaeal hydrocarbon cycling exist outside of the Euryarchaeota. Enrichments of *Korarchaeum cryptofilum* were obtained from the Obsidian Pool hot spring (located in Yellowstone National Park), enabling the genomic reconstruction of the first member of the Korarchaeota phylum[99]. The number and diversity of Korarchaeota genomes has recently been increased via metagenomic analyses of deep-sea hydrothermal sediments[26]. Moreover, several genomes of Korarchaeota have been obtained that contain *mcr* genes, which are thought to be involved in methane cycling coupled to sulfate reduction[144]. Another novel phylum is Aigarchaeota, from which *Caldiarchaeum subterraneum* was the first genome reconstructed, which showed the phylum is primarily associated with oxic hot spring communities[145]. Based on gene content they are likely able to utilize an array of extracellular polymers and thus may be important in cycling dissolved organic carbon[146].

Asgard superphylum

"…in the melting waters life appeared: the likeness of a person bigger than words, huger than any giant there will be or has ever been."[147]

Another widespread archaeal group detected in marine sediment diversity surveys from around the world is the Deep Sea Archaeal Group (DSAG[67]), also called the MBG-B[44]. Well known solely from 16S rRNA gene sequences, the first genome of this group was reconstructed from sediments near the Loki's Castle hydrothermal vent field in the North Atlantic Ocean and was subsequently named Lokiarchaeota[148]. Interestingly, phylogenomic analyses of Lokiarchaeota suggest that it is monophyletic with Eukaryotes, meaning that they are related in phylogenomic trees and that these Archaea and eukaryotes share a common ancestor. The placement of eukaryotes within the Archaea indicates that the first eukaryotic cells were derived from an archaeal ancestor. Supporting this further, Lokiarchaeota contain a variety of genes that encode for proteins with homology to eukaryotic actin, the endosomal sorting complexes required for transport (ESCRT)-I and III complexes, and the ubiquitin modifier system[148]. Recently, primordial type eukaryote-like profilins, which regulate actin cytoskeleton, have also been identified in Lokiarchaeota[149]. A few of these eukaryotic signature proteins (ESPs), such as components of the ESCRT-III complexes, have also been found in other phyla, particularly in TACK Archaea; however, only with a very patchy distribution and never as complete as for Lokiarchaeota. Therefore, it was suggested that Lokiarchaeota might be descendants of the progenitor archaea that lead to eukaryotes.

Following the discovery of the Lokiarchaeota, genomes belonging to a new phylum related to Lokiarchaeota were recovered from estuary sediments and named Thorarchaeota[148,150]. Genomes that are similar to Lokiarchaeota and Thorarchaeota were recently recovered from other anaerobic sediments including hot springs and groundwater[151], revealing two additional phyla termed Heimdallarchaeota and Odinarchaeota. Like other Archaea, these organisms were also overlooked in traditional community studies due to mismatches with commonly used PCR primers. To continue the Norse god theme, the superphylum was named Asgard[151]. Recently, genomes comprising another phylum named Helarchaeota were recovered from sediments associated with deep-sea hydrothermal vents[152]. Based on community metagenomic analyses of the sediments where Asgard have been identified thus far, they are low abundance (<1%) populations. Like Lokiarchaea, these new lineages contain a variety of ESPs involved in cytoskeleton formation, transport, translation, transcription and degradation pathways[151]. These include, but are not limited to, ubiquitin-activation enzymes and ESCRT complex proteins which are involved in trafficking mechanisms such as proteasome formation and membrane budding[153,154]. Other ESPs include cytoskeletal-associated actin and actin homologues, eukaryotic-specific ribosomal proteins and oligosaccharyl-transferases used in protein modification and secretion systems[155–157].

Along with the identification of these ESPs, the Asgard superphylum provided insights into the evolution of the bi-lipid structure that makes up the cellular membranes. Traditionally, an identifying characteristic of the archaeal domain is the presence of ether bonds in their lipid membranes instead of the ester bonding seen in both Bacteria and Eukaryotes[3]. This difference is referred to as the 'lipid divide' and has prompted much discussion on the evolution of modern eukaryotic cells and the likelihood of ester-linked membranes evolving separately for Bacteria and eukaryotic cells[158–160]. However, recent analyses of members of the Lokiarchaeota and other archaeal phyla have shown pathways that could be used in the production of Bacteria- and Eukarya-type ester-linked lipid membranes and supports the hypothesis that an Asgard ancestor gave rise to the first eukaryotic cell[161]. However, the exact functions of these Asgard proteins remains unresolved.

What does this tell us about the evolutionary history of Archaea and their role in early eukaryotic evolution? These findings indicate that an ancestor of the Asgard Archaea was the archaeal host of the symbiotic interaction with a proteobacterium, which later became mitochondria and formed early eukaryotic cells. Therefore, examining the ecological and metabolic capabilities of Asgard will likely provide insights into the nature of this symbiotic interaction. Comparative analyses of metabolic pathways of Asgard genomes has revealed significant metabolic differences between the phyla[162]. Recently, a co-culture of a Lokiarchaeon (named *Ca*. Prometheoarchaeum syntrophicum MK-D1), and a sulfate-reducing deltaproteobacterium (*Halodesulfovibrio*) has been obtained[163]. As predicted from the genomes, stable-isotope experiments confirmed that there is a syntrophic exchange of formate and hydrogen. Interestingly, microscopy of this enrichment culture indicates that *Ca*. P. syntrophicum forms new types of appendages that appear to entangle their partners, perhaps providing a mechanism for the entrapment of the alphaproteobacterial ancestral mitochondrion. Further investigations into the mechanisms of these unique physical interactions will provide new insights into eukaryogenesis. Heimdallarchaeota, another lineage more closely related to eukaryotes, are likely heterotrophs that may also rely on the exchange of hydrogen, electrons and/or simple carbon compounds. It is clear that syntrophy is not limited to the Lokiarchaeota and Heimdallarchaeota; for example, anaerobic alkane oxidation (via MCR) appears to be present in Helarchaeota[152], suggesting that many Asgard Archaea are involved in syntrophic interactions. Resolving the mechanisms of these interactions will enhance our understanding of the biological events that led to the origin of eukaryotes.

Outlook

The first Archaea identified were methanogens or were from thermophilic environments. As a result, our understanding of the metabolic roles of Archaea was limited to methane production (methanogenesis) and sulfur respiration[4]. Furthermore, genomic representation of the Archaea was limited to cultured representatives within the Euryarchaeota and Crenarchaeota. However, recent technological advances and long-term enrichment studies unveiled the functional potential of the uncultured biosphere, challenging our assumptions about which taxa have certain functions, and how well particular gene groups can be linked to a function. Recently, a method has been proposed to quantitatively merge phylogenomics with taxonomy called the Genome Taxonomy Database (GTDB), allowing placements of novel genomes into a taxonomic structure[164]. This is based on scores of relative evolutionary distances between the new genome and its nearest relatives, thus providing a possibility of unifying taxonomic hierarchies. Despite the adoption of the GTDB into commonly used taxonomic classification databases, the placement of archaeal phyla should be viewed as preliminary until a much deeper sampling of genomes is available. The limited nature of genomic sampling of the Archaea at the moment means that current taxonomic assignments may be significantly altered with greater sampling in the future.

Although the last decade witnessed a substantial expansion of archaeal diversity, a broader genomic exploration of Archaea in nature is still needed. For example, there are several novel phylum level groups that have been detected in 16S rRNA sequences but have not been genomically sampled (such as some related to Asgard, Marine Hydrothermal Vent Groups 1 and 2, pMC2A209, MG-IV and others). As is the case with many genomes recovered from metagenomes, there are certainly others that may not even have been detected due to biases in rRNA surveys[85]. Furthermore, eight phyla (Helarchaeota, Huberarchaeota, Nanoarchaeota, Nanohaloarchaeota, Nezaarchaeota, Odinarchaeota, Thorarchaeota and Woesearchaeota) that had not been known or characterized by rRNA sequences alone have now been identified via genomic

Box 3 | Origins of archaeal metabolism

Recent studies have started to expand the known diversity of metabolic strategies that can be used by Archaea from different phyla across multiple environments. This more comprehensive diversity sampling of archaeal genomes has also begun to enable the tracking of evolutionary histories of metabolic pathways among Archaea. Modelling of ancestral lineages is achieved by estimating gene gain, loss and transfer among lineages. This approach enables the prediction of physiological capabilities of ancestral lineages. For example, this approach has been applied to the Aigarchaeota and indicates that they have undergone considerable gene loss, an acquisition of key genes from Bacteria (genes involved in dissimilatory sulfite reduction and CO oxidation) and that their ancestors were thermophilic aerobes[185]. A broader modelling of archaea indicates that the last archaeal common ancestor was likely anaerobic and capable of CO_2 reduction to acetate[181]. Methyl-CoM reductases and the H_4MPT branch of the Wood–Ljungdahl pathway for hydrocarbon utilization (including methanogenesis), once thought to be exclusive to Euryarchaeota, are now known to be broadly distributed throughout the Archaea and are found both in the TACK and Asgard Archaea[39,186,187]. This broader taxonomic distribution indicates that methanogenesis is an ancient metabolism, likely present in the universal common ancestor of both Archaea and Bacteria[188,189]. In fact, carbon isotopes of kerogen indicate methanogenesis to be between 2.8 billion–2.6 billion years old[190]. However, a more recent evaluation indicates that this pathway has likely existed since the last common ancestor of Archaea, and the genes present in Bacteria likely originated in the Euryarchaeota[186]. Furthermore, the sporadic distribution of these genes among Archaea may be the result of ancient horizontal transfers. Overall, it is becoming clear that methane cycling has long played important roles in Earth's history. A greater sampling of archaeal genomes, including those from undescribed lineages (phyla), will enhance the predictive capabilities of ancestral modelling.

recovery[85,151,152]. Some of these phyla had not been detected due to PCR primers not targeting them, while others had been recovered but with little representation and thus were not given a phylogenetic designation. This suggests that rRNA diversity surveys have overlooked, or not characterized, a substantial amount of biodiversity that is present in nature. In recent years, the reconstruction of genomes (via metagenomic assembly and binning) from whole community libraries has significantly accelerated the number of available genomes from uncultivated lineages (Fig. 4). This will continue to be an active pursuit in advancing our understanding of archaeal biology and such increased taxonomical sampling is also likely to provide new insights into archaeal evolution (Box 3).

In recent years, there have been many cases where pathways for key ecological processes have been identified in genomes belonging to novel lineages (Fig. 3). The presence of pathways does not always equate to activity and it can therefore be difficult to accurately assign function to novel genes. Therefore, there is a need for more studies that provide in situ physiological measurements, including linking functions measured in bulk samples to those found in genomic reconstructions. Several approaches have been developed to quantify community level activities including DNA stable isotope probing (DNA-SIP), metatranscriptomics, metabolomics and metaproteomics. Genomics, metatranscriptomics, metabolomics and enzymatic assays can provide powerful support to functions predicted by gene information[165]. Another new technique that is similar to microautoradiography[85,166] (where cells taking up radioactive substrates can be visualized) called bioorthogonal noncanonical amino acid tagging (BONCAT) may reveal active cells by tracking the incorporation of synthetic amino acids into newly synthesized proteins[167]. Since the active cells are fluorescent, it is possible to couple microscopy with rRNA-targeted fluorescent in situ hybridization (FISH) or physically sort cells via fluorescent-activated cell sorting (BONCAT–FACS). This approach presents a potentially high-throughput technique to determine phenotype (growth characteristics and substrate usage) and link it to genotype at the community level. A limitation of this technique could be that it may be limited to organisms that can synthesize new proteins relatively quickly. However, coupling community-level genomic reconstruction and activity assays may provide linkages to diversity and mechanisms (metabolic pathways) of microbial mediation of processes at a broad taxonomic level.

Assigning functions to novel genes is another great hurdle for resolving the ecological roles of new archaeal branches. Some success has come from the heterologous expression of archaeal genes amplified directly from environmental genomic material[168], suggesting that much more can be learned from a wider application of these techniques. Ultimately, further developments in assessing archaeal physiology will enhance our understanding of biogeochemical cycling and enable us to more accurately model the flow of carbon and energy through microbial ecosystems. Our accelerating ability to obtain genomes and an ever-advancing toolkit for tracking activity of natural communities will continue to enlighten archaeal ecology and evolution. Given how much our view has advanced in the last few years, there are certainly new frontiers to be charted among new branches in the Archaea.

Received: 7 October 2019; Accepted: 30 March 2020;
Published online: 4 May 2020

References

1. Chatton, E. *Titres et travaux scientifiques (1906–1937) de Edouard Chatton*. (Sète: Impr. E. Sottan, 1938).
2. Whittaker, R. H. New concepts of kingdoms of organisms. *Science* **163**, 150–160 (1969).
3. Woese, C. R. et al. Conservation of primary structure in 16S ribosomal RNA. *Nature* **254**, 83–86 (1975).
4. Balch, W. E., Magrum, L. J., Fox, G. E., Wolfe, R. S. & Woese, C. R. An ancient divergence among the bacteria. *J. Mol. Evol.* **9**, 305–311 (1977).
5. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl Acad. Sci. USA* **74**, 5088–5090 (1977).
6. Woese, C. R., Magrum, L. J. & Fox, G. E. Archaebacteria. *J. Mol. Evol.* **11**, 245–251 (1978).
7. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl Acad. Sci. USA* **87**, 4576–4579 (1990).
8. Doolittle, W. F. & Logsdon, J. M. Jr. Archaeal genomics: do archaea have a mixed heritage? *Curr. Biol.* **8**, R209–R211 (1998).
9. MacGregor, B. J., Moser, D. P., Alm, E. W., Nealson, K. H. & Stahl, D. A. Crenarchaeota in Lake Michigan sediment. *Appl. Environ. Microbiol.* **63**, 1178–1181 (1997).
10. DeLong, E. F. Archaea in coastal marine environments. *Proc. Natl Acad. Sci. USA* **89**, 5685–5689 (1992).
11. Olsen, G. Microbial ecology and evolution: a ribosomal RNA approach. *Annu. Rev. Microbiol.* **40**, 337–365 (1986).
12. Rappé, M. S. & Giovannoni, S. J. The uncultured microbial majority. *Annu. Rev. Microbiol.* **57**, 369–394 (2003).
13. Reysenbach, A. L., Giver, L. J., Wickham, G. S. & Pace, N. R. Differential amplification of rRNA genes by polymerase chain reaction. *Appl. Environ. Microbiol.* **58**, 3417–3418 (1992).
14. Hugenholtz, P., Goebel, B. M. & Pace, N. R. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* **180**, 4765–4774 (1998).
15. Pace, N. R. A molecular view of microbial diversity and the biosphere. *Science* **276**, 734–740 (1997).
16. Baker, B. J. & Dick, G. J. Omic approaches in microbial ecology: charting the unknown. *Microbe* **8**, 353–359 (2013).
17. Fuhrman, J. A., McCallum, K. & Davis, A. A. Novel major archaebacterial group from marine plankton. *Nature* **356**, 148–149 (1992).
18. Barns, S. M., Fundgya, R. E., Jeffries, M. W. & Pace, N. R. Remarkable archaeal diversity detected in a Yellowstone national park hot spring environment. *Proc. Natl Sci. USA* **91**, 1609–1613 (1994).

19. Barns, S. M., Delwiche, C. F., Palmer, J. D. & Pace, N. R. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc. Natl Acad. Sci. USA* **93**, 9188–9193 (1996).
20. Adam, P. S., Borrel, G., Brochier-Armanet, C. & Gribaldo, S. The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *ISME J.* **11**, 2407–2425 (2017).
21. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).
22. Tyson, G. W. et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37–43 (2004).
23. Venter, J. C. et al. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
24. Dick, G. J. et al. Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* **10**, R85 (2009).
25. Stepanauskas, R. & Sieracki, M. E. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc. Natl Acad. Sci. USA* **104**, 9052–9057 (2007).
26. Dombrowski, N., Teske, A. P. & Baker, B. J. Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nat. Commun.* **9**, 4999 (2018).
27. Hug, L. A. et al. A new view of the tree of life. *Nat. Microbiol.* **1**, 16048 (2016).
28. Guy, L. & Ettema, T. J. G. The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol.* **19**, 580–587 (2011).
29. Rinke, C. et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).
30. Smith, P. H. & Hungate, R. E. Isolation and characterization of *Methanobacterium ruminantium* n. sp. *J. Bacteriol.* **75**, 713–718 (1958).
31. Brock, T. D. & Darland, G. K. Limits of microbial existence: temperature and pH. *Science* **169**, 1316–1318 (1970).
32. Fox, G. E., Magrum, L. J., Balch, W. E., Wolfe, R. S. & Woese, C. R. Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proc. Natl Acad. Sci. USA* **74**, 4537–4541 (1977).
33. Wolfe, R. S. Microbial formation of methane. *Adv. Microb. Physiol.* **6**, 107–146 (1971).
34. Larsen, H. Halophilic and halotolerant microorganisms-an overview and historical perspective. *FEMS Microbiol. Lett.* **2**, 3–7 (1986).
35. Andrei, A.-Ş., Banciu, H. L. & Oren, A. Living with salt: metabolic and phylogenetic diversity of archaea inhabiting saline ecosystems. *FEMS Microbiol. Lett.* **330**, 1–9 (2012).
36. Klenk, H. P. et al. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390**, 364–370 (1997).
37. Takai, K. et al. Cell proliferation at 122 °C and isotopically heavy CH_4 production by a hyperthermophilic methanogen under high-pressure cultivation. *Proc. Natl Acad. Sci. USA* **105**, 10949–10954 (2008).
38. Orphan, V. J., House, C. H., Hinrichs, K.-U., McKeegan, K. D. & DeLong, E. F. Multiple archaeal groups mediate methane oxidation in anoxic cold seep sediments. *Proc. Natl Acad. Sci. USA* **99**, 7663–7668 (2002).
39. Wang, Y., Wegener, G., Hou, J., Wang, F. & Xiao, X. Expanding anaerobic alkane metabolism in the domain of Archaea. *Nat. Microbiol.* **4**, 595–602 (2019).
40. Vetriani, C., Reysenbach, A. L. & Doré, J. Recovery and phylogenetic analysis of archaeal rRNA sequences from continental shelf sediments. *FEMS Microbiol. Lett.* **161**, 83–88 (1998).
41. Iverson, V. et al. Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science* **335**, 587–590 (2012).
42. Needham, D. M. & Fuhrman, J. A. Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nat. Microbiol.* **1**, 16005 (2016).
43. Baker, B. J. et al. Genomic inference of the metabolism of cosmopolitan subsurface Archaea, Hadesarchaea. *Nat. Microbiol.* **1**, 16002 (2016).
44. Vetriani, C., Jannasch, H. W., MacGregor, B. J., Stahl, D. A. & Reysenbach, A. L. Population structure and phylogenetic characterization of marine benthic Archaea in deep-sea sediments. *Appl. Environ. Microbiol.* **65**, 4375–4384 (1999).
45. Jungbluth, S. P., Amend, J. P. & Rappé, M. S. Metagenome sequencing and 98 microbial genomes from Juan de Fuca Ridge flank subsurface fluids. *Sci. Data* **4**, 170037 (2017).
46. Carr, S. A. et al. Carboxydotrophy potential of uncultivated Hydrothermarchaeota from the subseafloor crustal biosphere. *ISME J.* **13**, 1457–1468 (2019).
47. Chuvochina, M. et al. The importance of designating type material for uncultured taxa. *Syst. Appl. Microbiol.* **42**, 15–21 (2019).
48. Lazar, C. S., Baker, B. J., Seitz, K. W. & Teske, A. P. Genomic reconstruction of multiple lineages of uncultured benthic archaea suggests distinct biogeochemical roles and ecological niches. *ISME J.* **11**, 1058 (2017).
49. Probst, A. J. et al. Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nat. Microbiol.* **3**, 328–336 (2018).
50. Probst, A. J. et al. Biology of a widespread uncultivated archaeon that contributes to carbon fixation in the subsurface. *Nat. Commun.* **5**, 5497 (2014).
51. Teske, A. & Sørensen, K. B. Uncultured archaea in deep marine subsurface sediments: have we caught them all? *ISME J.* **2**, 3–18 (2008).
52. Orphan, V. J., House, C. H., Hinrichs, K. U., McKeegan, K. D. & DeLong, E. F. Methane-consuming archaea revealed by directly coupled isotopic and phylogenetic analysis. *Science* **293**, 484–487 (2001).
53. Hallam, S. J. et al. Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* **305**, 1457–1462 (2004).
54. Raghoebarsing, A. A. et al. A microbial consortium couples anaerobic methane oxidation to denitrification. *Nature* **440**, 918–921 (2006).
55. Haroon, M. F. et al. Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* **500**, 567–570 (2013).
56. McGlynn, S. E., Chadwick, G. L., Kempes, C. P. & Orphan, V. J. Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature* **526**, 531–535 (2015).
57. Scheller, S., Yu, H., Chadwick, G. L., McGlynn, S. E. & Orphan, V. J. Artificial electron acceptors decouple archaeal methane oxidation from sulfate reduction. *Science* **351**, 703–707 (2016).
58. Laso-Pérez, R. et al. Thermophilic archaea activate butane via alkyl-coenzyme M formation. *Nature* **539**, 396–401 (2016).
59. Lloyd, K. G., Lapham, L. & Teske, A. An anaerobic methane-oxidizing community of ANME-1b archaea in hypersaline Gulf of Mexico sediments. *Appl. Environ. Microbiol.* **72**, 7218–7230 (2006).
60. Maignien, L. et al. Anaerobic oxidation of methane in hypersaline cold seep sediments. *FEMS Microbiol. Ecol.* **83**, 214–231 (2013).
61. Dombrowski, N., Seitz, K. W., Teske, A. P. & Baker, B. J. Genomic insights into potential interdependencies in microbial hydrocarbon and nutrient cycling in hydrothermal sediments. *Microbiome* **5**, 106 (2017).
62. Borrel, G. et al. Wide diversity of methane and short-chain alkane metabolisms in uncultured archaea. *Nat. Microbiol.* **4**, 603–613 (2019).
63. Chen, S.-C. et al. Anaerobic oxidation of ethane by archaea from a marine hydrocarbon seep. *Nature* **568**, 108–111 (2019).
64. Offre, P., Spang, A. & Schleper, C. Archaea in biogeochemical cycles. *Annu. Rev. Microbiol.* **67**, 437–457 (2013).
65. Mwirichia, R. et al. Metabolic traits of an uncultured archaeal lineage -MSBL1- from brine pools of the Red Sea. *Sci. Rep.* **6**, 19181 (2016).
66. Takai, K., Moser, D. P., DeFlaun, M., Onstott, T. C. & Fredrickson, J. K. Archaeal diversity in waters from deep South African gold mines. *Appl. Environ. Microbiol.* **67**, 5750–5760 (2001).
67. Inagaki, F. et al. Microbial communities associated with geological horizons in coastal subseafloor sediments from the sea of okhotsk. *Appl. Environ. Microbiol.* **69**, 7224–7235 (2003).
68. Parkes, R. J. et al. Deep sub-seafloor prokaryotes stimulated at interfaces over geological time. *Nature* **436**, 390–394 (2005).
69. Biddle, J. F. et al. Heterotrophic Archaea dominate sedimentary subsurface ecosystems off Peru. *Proc. Natl Acad. Sci. USA* **103**, 3846–3851 (2006).
70. Takai, K. & Horikoshi, K. Genetic diversity of archaea in deep-sea hydrothermal vent environments. *Genetics* **152**, 1285–1297 (1999).
71. Reysenbach, A.-L. et al. A ubiquitous thermoacidophilic archaeon from deep-sea hydrothermal vents. *Nature* **442**, 444–447 (2006).
72. Knittel, K., Lösekann, T., Boetius, A., Kort, R. & Amann, R. Diversity and distribution of methanotrophic archaea at cold seeps. *Appl. Environ. Microbiol.* **71**, 467–479 (2005).
73. Lloyd, K. G. et al. Predominant archaea in marine sediments degrade detrital proteins. *Nature* **496**, 215–218 (2013).
74. Dettling, M. D., Yavitt, J. B., Cadillo-Quiroz, H., Sun, C. & Zinder, S. H. Soil–methanogen interactions in two peatlands (Bog, Fen) in central New York State. *Geomicrobiol. J.* **24**, 247–259 (2007).
75. Zhou, Z. et al. Genomic and transcriptomic insights into the ecology and metabolism of benthic archaeal cosmopolitan, Thermoprofundales (MBG-D archaea). *ISME J.* **13**, 885–901 (2019).
76. Massana, R., DeLong, E. F. & Pedrós-Alió, C. A few cosmopolitan phylotypes dominate planktonic archaeal assemblages in widely different oceanic provinces. *Appl. Environ. Microbiol.* **66**, 1777–1787 (2000).
77. Fuhrman, J. A. & Davis, A. A. Widespread Archaea and novel Bacteria from the deep sea as shown by 16S rRNA gene sequences. *Mar. Ecol. Prog. Ser.* **150**, 275–285 (1997).
78. Martin-Cuadrado, A.-B. et al. A new class of marine Euryarchaeota group II from the Mediterranean deep chlorophyll maximum. *ISME J.* **9**, 1619–1634 (2015).
79. Li, M. et al. Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. *Nat. Commun.* **6**, 8933 (2015).

80. Tully, B. J. Metabolic diversity within the globally abundant Marine Group II Euryarchaea offers insight into ecological patterns. *Nat. Commun.* **10**, 271 (2019).
81. Rinke, C. et al. A phylogenomic and ecological analysis of the globally abundant Marine Group II archaea (*Ca.* Poseidoniales ord. nov.). *ISME J.* **13**, 663–675 (2019).
82. Orsi, W. D. et al. Diverse, uncultivated bacteria and archaea underlying the cycling of dissolved protein in the ocean. *ISME J.* **10**, 2158–2173 (2016).
83. Huber, H. et al. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**, 63–67 (2002).
84. Waters, E. et al. The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc. Natl Acad. Sci. USA* **100**, 12984–12988 (2003).
85. Baker, B. J. et al. Lineages of acidophilic archaea revealed by community genomic analysis. *Science* **314**, 1933–1935 (2006).
86. Baker, B. J. et al. Enigmatic, ultrasmall, uncultivated Archaea. *Proc. Natl Acad. Sci. USA* **107**, 8806–8811 (2010).
87. Narasingarao, P. et al. De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *ISME J.* **6**, 81–93 (2012).
88. Castelle, C. J. et al. Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* **25**, 690–701 (2015).
89. Schwank, K. et al. An archaeal symbiont-host association from the deep terrestrial subsurface. *ISME J.* **13**, 2135–2139 (2019).
90. Parks, D. H. et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* **2**, 1533–1542 (2017).
91. Bird, J. T., Baker, B. J., Probst, A. J., Podar, M. & Lloyd, K. G. Culture independent genomic comparisons reveal environmental adaptations for Altiarchaeales. *Front. Microbiol.* **7**, 1221 (2016).
92. Krause, S., Bremges, A., Münch, P. C., McHardy, A. C. & Gescher, J. Characterisation of a stable laboratory co-culture of acidophilic nanoorganisms. *Sci. Rep.* **7**, 3289 (2017).
93. Hamm, J. N. et al. Unexpected host dependency of Antarctic Nanohaloarchaeota. *Proc. Natl Acad. Sci. USA* **116**, 14661–14670 (2019).
94. Comolli, L. R., Baker, B. J., Downing, K. H., Siegerist, C. E. & Banfield, J. F. Three-dimensional analysis of the structure and ecology of a novel, ultra-small archaeon. *ISME J.* **3**, 159–167 (2009).
95. Heimerl, T. et al. A complex endomembrane system in the Archaeon *Ignicoccus hospitalis* tapped by *Nanoarchaeum equitans*. *Front. Microbiol.* **8**, 1072 (2017).
96. Burstein, D. et al. New CRISPR–Cas systems from uncultivated microbes. *Nature* **542**, 237–241 (2017).
97. Brochier-Armanet, C., Boussau, B., Gribaldo, S. & Forterre, P. Mesophilic crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat. Rev. Microbiol.* **6**, 245–252 (2008).
98. Spang, A. et al. Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. *Trends Microbiol.* **18**, 331–340 (2010).
99. Elkins, J. G. et al. A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc. Natl Acad. Sci. USA* **105**, 8102–8107 (2008).
100. Vanwonterghem, I. et al. Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nat. Microbiol.* **1**, 16170 (2016).
101. Blöchl, E. et al. *Pyrolobus fumarii*, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 °C. *Extremophiles* **1**, 14–21 (1997).
102. Cubonová, L., Sandman, K., Hallam, S. J., Delong, E. F. & Reeve, J. N. Histones in crenarchaea. *J. Bacteriol.* **187**, 5482–5485 (2005).
103. Brock, T. D., Brock, K. M., Belly, R. T. & Weiss, R. L. *Sulfolobus*: A new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Archiv. Mikrobiol.* **84**, 54–68 (1972).
104. Zhang, C., Phillips, A. P. R., Wipfler, R. L., Olsen, G. J. & Whitaker, R. J. The essential genome of the crenarchaeal model *Sulfolobus islandicus*. *Nat. Commun.* **9**, 4908 (2018).
105. Zillig, W. et al. The Archaebacterium *Thermofilum pendens* represents, a novel genus of the thermophilic, anaerobic sulfur respiring Thermoproteales. *Syst. Appl. Microbiol.* **4**, 79–87 (1983).
106. Nakagawa, S. *Aeropyrum camini* sp. nov., a strictly aerobic, hyperthermophilic archaeon from a deep-sea hydrothermal vent chimney. *Int. J. Syst. Evol. Microbiol.* **54**, 329–335 (2004).
107. Kozubal, M. A. 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).
108. Guy, L., Spang, A., Saw, J. H. & Ettema, T. J. G. 'Geoarchaeote NAG1' is a deeply rooting lineage of the archaeal order Thermoproteales rather than a new phylum. *ISME J.* **8**, 1353–1357 (2014).
109. Schleper, C., Jurgens, G. & Jonuscheit, M. Genomic studies of uncultivated archaea. *Nat. Rev. Microbiol.* **3**, 479–488 (2005).
110. Hallam, S. J. et al. Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc. Natl Acad. Sci. USA* **103**, 18296–18301 (2006).
111. Konneke, M. et al. Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO_2 fixation. *Proc. Natl Acad. Sci. USA* **111**, 8239–8244 (2014).
112. Könneke, M. et al. Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**, 543–546 (2005).
113. Walker, C. B. et al. *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc. Natl Acad. Sci. USA* **107**, 8818–8823 (2010).
114. Martens-Habbena, W., Berube, P. M., Urakawa, H., de la Torre, J. R. & Stahl, D. A. Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. *Nature* **461**, 976–979 (2009).
115. Baker, B. J., Lesniewski, R. A. & Dick, G. J. Genome-enabled transcriptomics reveals archaeal populations that drive nitrification in a deep-sea hydrothermal plume. *ISME J.* **6**, 2269–2279 (2012).
116. Palatinszky, M. et al. Cyanate as an energy source for nitrifiers. *Nature* **524**, 105–108 (2015).
117. Santoro, A. E., Buchwald, C., McIlvin, M. R. & Casciotti, K. L. Isotopic signature of N_2O produced by marine ammonia-oxidizing Archaea. *Science* **333**, 1282–1285 (2011).
118. Metcalf, W. W. et al. Synthesis of methylphosphonic acid by marine microbes: a source for methane in the aerobic ocean. *Science* **337**, 1104–1107 (2012).
119. Hua, Z.-S. et al. Insights into the ecological roles and evolution of methyl–coenzyme M reductase-containing hot spring Archaea. *Nat. Commun.* **10**, 4574 (2019).
120. Karner, M. B., DeLong, E. F. & Karl, D. M. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**, 507–510 (2001).
121. Leininger, S. et al. Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442**, 806–809 (2006).
122. Pachiadaki, M. G., Yakimov, M. M., LaCono, V., Leadbetter, E. & Edgcomb, V. Unveiling microbial activities along the halocline of Thetis, a deep-sea hypersaline anoxic basin. *ISME J.* **8**, 2478–2489 (2014).
123. Gubry-Rangin, C. et al. Coupling of diversification and pH adaptation during the evolution of terrestrial Thaumarchaeota. *Proc. Natl Acad. Sci. USA* **112**, 9370–9375 (2015).
124. de la Torre, J. R., Walker, C. B., Ingalls, A. E., Könneke, M. & Stahl, D. A. Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. *Environ. Microbiol.* **10**, 810–818 (2008).
125. Durbin, A. M. & Teske, A. Microbial diversity and stratification of South Pacific abyssal marine sediments. *Environ. Microbiol.* **13**, 3219–3234 (2011).
126. Qin, W., Martens-Habbena, W., Kobelt, J. N. & Stahl, D. A. in *Bergey's Manual of Systematics of Archaea and Bacteria* (eds Whitman, W. B. et al.) 1–2 (Wiley, 2016).
127. Stahl, D. A. & de la Torre, J. R. Physiology and diversity of ammonia-oxidizing Archaea. *Annu. Rev. Microbiol.* **66**, 83–101 (2012).
128. Vajrala, N. et al. Hydroxylamine as an intermediate in ammonia oxidation by globally abundant marine archaea. *Proc. Natl Acad. Sci. USA* **110**, 1006–1011 (2013).
129. Stieglmeier, M. et al. Aerobic nitrous oxide production through N-nitrosating hybrid formation in ammonia-oxidizing archaea. *ISME J.* **8**, 1135–1146 (2014).
130. Kozlowski, J. A., Stieglmeier, M., Schleper, C., Klotz, M. G. & Stein, L. Y. Pathways and key intermediates required for obligate aerobic ammonia-dependent chemolithotrophy in bacteria and Thaumarchaeota. *ISME J.* **10**, 1836–1845 (2016).
131. Liu, S. et al. Abiotic conversion of extracellular NH_2OH contributes to N_2O emission during ammonia oxidation. *Environ. Sci. Technol.* **51**, 13122–13132 (2017).
132. Kallmeyer, J., Pockalny, R., Adhikari, R. R., Smith, D. C. & D'Hondt, S. Global distribution of microbial abundance and biomass in subseafloor sediment. *Proc. Natl Acad. Sci. USA* **109**, 16213–16216 (2012).
133. Kubo, K. et al. Archaea of the Miscellaneous Crenarchaeotal Group are abundant, diverse and widespread in marine sediments. *ISME J.* **6**, 1949–1965 (2012).
134. Meng, J. et al. An uncultivated crenarchaeota contains functional bacteriochlorophyll a synthase. *ISME J.* **3**, 106–116 (2009).
135. Meng, J. et al. Genetic and functional properties of uncultivated MCG archaea assessed by metagenome and gene expression analyses. *ISME J.* **8**, 650–659 (2014).
136. Ochsenreiter, T., Selezi, D., Quaiser, A., Bonch-Osmolovskaya, L. & Schleper, C. Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR. *Environ. Microbiol.* **5**, 787–797 (2003).
137. Jurgens, G. Identification of novel Archaea in bacterioplankton of a boreal forest lake by phylogenetic analysis and fluorescent in situ hybridization. *FEMS Microbiol. Ecol.* **34**, 45–56 (2000).

138. McKay, L. J., Hatzenpichler, R., Inskeep, W. P. & Fields, M. W. Occurrence and expression of novel methyl-coenzyme M reductase gene (*mcr*A) variants in hot spring sediments. *Sci. Rep.* **7**, 7252 (2017).
139. Lazar, C. S. et al. Genomic evidence for distinct carbon substrate preferences and ecological niches of Bathyarchaeota in estuarine sediments. *Environ. Microbiol.* **18**, 1200–1211 (2016).
140. He, Y. 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).
141. Meador, T. B. et al. The archaeal lipidome in estuarine sediment dominated by members of the Miscellaneous Crenarchaeotal Group. *Environ. Microbiol.* **17**, 2441–2458 (2015).
142. Evans, P. N. et al. Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* **350**, 434–438 (2015).
143. Yu, T. et al. Growth of sedimentary Bathyarchaeota on lignin as an energy source. *Proc. Natl Acad. Sci. USA* **115**, 6022–6027 (2018).
144. McKay, L. J. et al. Co-occurring genomic capacity for anaerobic methane and dissimilatory sulfur metabolisms discovered in the Korarchaeota. *Nat. Microbiol.* **4**, 614–622 (2019).
145. Nunoura, T. 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).
146. Beam, J. P. et al. Ecophysiology of an uncultivated lineage of Aigarchaeota from an oxic, hot spring filamentous 'streamer' community. *ISME J.* **10**, 210–224 (2016).
147. Gaiman, N. *Norse Mythology* (Bloomsbury Publishing, 2017).
148. Spang, A. et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179 (2015).
149. Akıl, C. & Robinson, R. C. Genomes of Asgard archaea encode profilins that regulate actin. *Nature* **562**, 439–443 (2018).
150. Seitz, K. W., Lazar, C. S., Hinrichs, K.-U., Teske, A. P. & Baker, B. J. Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction. *ISME J.* **10**, 1696–1705 (2016).
151. Zaremba-Niedzwiedzka, K. et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).
152. Seitz, K. W. et al. Asgard archaea capable of anaerobic hydrocarbon cycling. *Nat. Commun.* **10**, 1822 (2019).
153. Hochstrasser, M. Origin and function of ubiquitin-like proteins. *Nature* **458**, 422–429 (2009).
154. Hurley, J. H. ESCRTs are everywhere. *EMBO J.* **34**, 2398–2407 (2015).
155. Ettema, T. J. G., Lindås, A.-C. & Bernander, R. An actin-based cytoskeleton in archaea. *Mol. Microbiol.* **80**, 1052–1061 (2011).
156. Koonin, E. V. & Yutin, N. The dispersed archaeal eukaryome and the complex archaeal ancestor of eukaryotes. *Cold Spring Harb. Persp. Biol.* **6**, a016188 (2014).
157. Dalziel, M., Crispin, M., Scanlan, C. N., Zitzmann, N. & Dwek, R. A. Emerging principles for the therapeutic exploitation of glycosylation. *Science* **343**, 1235681 (2014).
158. Koga, Y., Kyuragi, T., Nishihara, M. & Sone, N. Did archaeal and bacterial cells arise independently from noncellular precursors? A hypothesis stating that the advent of membrane phospholipid with enantiomeric glycerophosphate backbones caused the separation of the two lines of descent. *J. Mol. Evol.* **47**, 631–631 (1998).
159. Sojo, V., Pomiankowski, A. & Lane, N. A bioenergetic basis for membrane divergence in archaea and bacteria. *PLoS Biol.* **12**, e1001926 (2014).
160. Gray, M. W. & Doolittle, W. F. Has the endosymbiont hypothesis been proven? *Microbiol. Rev.* **46**, 1–42 (1982).
161. Villanueva, L., Schouten, S. & Sinninghe Damsté, J. S. Phylogenomic analysis of lipid biosynthetic genes of Archaea shed light on the 'lipid divide'. *Environ. Microbiol.* **19**, 54–69 (2017).
162. Spang, A. et al. Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. *Nat. Microbiol.* **4**, 1138–1148 (2019).
163. Imachi, H. et al. Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature* **577**, 519–525 (2020).
164. Parks, D. H. et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
165. Bird, J. T. et al. Uncultured microbial phyla suggest mechanisms for multi-thousand-year subsistence in Baltic Sea sediments. *mBio* **10**, e02376–18 (2019).
166. Nielsen, J. L. & Nielsen, P. H. in *Handbook of Hydrocarbon and Lipid Microbiology* (ed. Timmis, K. N.) 4093–4102 (2010).
167. Hatzenpichler, R. et al. Visualizing in situ translational activity for identifying and sorting slow-growing archaeal-bacterial consortia. *Proc. Natl Acad. Sci. USA* **113**, E4069–E4078 (2016).
168. Michalska, K. et al. New aminopeptidase from 'microbial dark matter' archaeon. *FASEB J.* **29**, 4071–4079 (2015).
169. Martijn, J. & Ettema, T. J. G. From archaeon to eukaryote: the evolutionary dark ages of the eukaryotic cell. *Biochem. Soc. Trans.* **41**, 451–457 (2013).
170. Jay, Z. J. et al. Marsarchaeota are an aerobic archaeal lineage abundant in geothermal iron oxide microbial mats. *Nat. Microbiol.* **3**, 732–740 (2018).
171. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
172. Alneberg, J. et al. Binning metagenomic contigs by coverage and composition. *Nat. Methods* **11**, 1144–1146 (2014).
173. Eren, A. M. et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**, e1319 (2015).
174. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).
175. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
176. Pace, N. R. Mapping the tree of life: progress and prospects. *Microbiol. Mol. Biol. Rev.* **73**, 565–576 (2009).
177. Baker, B. J., Lazar, C. S., Teske, A. P. & Dick, G. J. Genomic resolution of linkages in carbon, nitrogen, and sulfur cycling among widespread estuary sediment bacteria. *Microbiome* **3**, 14 (2015).
178. Brown, C. T. et al. Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208–211 (2015).
179. Sorek, R. et al. Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* **318**, 1449–1452 (2007).
180. Petitjean, C., Deschamps, P., López-García, P. & Moreira, D. Rooting the domain archaea by phylogenomic analysis supports the foundation of the new kingdom Proteoarchaeota. *Genome Biol. Evol.* **7**, 191–204 (2014).
181. Williams, T. A. et al. Integrative modeling of gene and genome evolution roots the archaeal tree of life. *Proc. Natl Acad. Sci. USA* **114**, E4602–E4611 (2017).
182. Toft, C. & Andersson, S. G. E. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat. Rev. Genet.* **11**, 465–475 (2010).
183. Das, S., Paul, S., Bag, S. K. & Dutta, C. Analysis of *Nanoarchaeum equitans* genome and proteome composition: indications for hyperthermophilic and parasitic adaptation. *BMC Genomics* **7**, 186 (2006).
184. Podar, M. et al. A genomic analysis of the archaeal system *Ignicoccus hospitalis–Nanoarchaeum equitans*. *Genome Biol.* **9**, R158 (2008).
185. Hua, Z.-S. et al. Genomic inference of the metabolism and evolution of the archaeal phylum Aigarchaeota. *Nat. Commun.* **9**, 2832 (2018).
186. Adam, P. S., Borrel, G. & Gribaldo, S. An archaeal origin of the Wood–Ljungdahl HMPT branch and the emergence of bacterial methylotrophy. *Nat. Microbiol.* **4**, 2155–2163 (2019).
187. Borrel, G., Adam, P. S. & Gribaldo, S. Methanogenesis and the Wood–Ljungdahl pathway: an ancient, versatile, and fragile association. *Genome Biol. Evol.* **8**, 1706–1711 (2016).
188. Chistoserdova, L. The enigmatic planctomycetes may hold a key to the origins of methanogenesis and methylotrophy. *Mol. Biol. Evol.* **21**, 1234–1241 (2004).
189. Chistoserdova, L. Wide distribution of genes for tetrahydromethanopterin/methanofuran-linked C1 transfer reactions argues for their presence in the common ancestor of bacteria and archaea. *Front. Microbiol.* **7**, 1425 (2016).
190. Ueno, Y., Yamada, K., Yoshida, N., Maruyama, S. & Isozaki, Y. Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. *Nature* **440**, 516–519 (2006).

Acknowledgements

This review was supported by the National Science Foundation Systematics and Biodiversity Sciences award no. 1737298 (to B.J.B.), grant no. NSF-OCE-1431598 (to K.G.L.) and the Simons Foundation grant no. 404586 (to K.G.L.).

Author contributions

B.J.B. and A.E.S. wrote this review and all authors provided input. B.J.B., K.W.S. and V.D.A. generated the phylogeny. B.J.B., K.G.L., V.D.A. and N.D. generated the figures.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-020-0715-z.

Correspondence should be addressed to B.J.B.

Reprints and permissions information is available at www.nature.com/reprints.

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