

## Review Article

# Genome size evolution in the Archaea

Siri Kellner<sup>1</sup>, Anja Spang<sup>2,3</sup>, Pierre Offre<sup>2</sup>, Gergely J. Szöllösi<sup>4,5</sup>, Celine Petitjean<sup>6</sup> and  Tom A. Williams<sup>6</sup>

<sup>1</sup>School of Earth Sciences, University of Bristol, Bristol BS8 1TQ, U.K.; <sup>2</sup>NIOZ, Royal Netherlands Institute for Sea Research, Department of Marine Microbiology and Biogeochemistry, and Utrecht University, P.O. Box 59, NL-1790 AB Den Burg, The Netherlands; <sup>3</sup>Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, SE-75123, Uppsala, Sweden; <sup>4</sup>MTA-ELTE Lendület Evolutionary Genomics Research Group, 1117 Budapest, Hungary; <sup>5</sup>Department of Biological Physics, Eötvös Loránd University, 1117 Budapest, Hungary; <sup>6</sup>School of Biological Sciences, University of Bristol, Bristol BS8 1TQ, U.K.

**Correspondence:** Celine Petitjean (celine.petitjean@bristol.ac.uk) or Tom A. Williams (tom.a.williams@bristol.ac.uk)



What determines variation in genome size, gene content and genetic diversity at the broadest scales across the tree of life? Much of the existing work contrasts eukaryotes with prokaryotes, the latter represented mainly by Bacteria. But any general theory of genome evolution must also account for the Archaea, a diverse and ecologically important group of prokaryotes that represent one of the primary domains of cellular life. Here, we survey the extant diversity of Bacteria and Archaea, and ask whether the general principles of genome evolution deduced from the study of Bacteria and eukaryotes also apply to the archaeal domain. Although Bacteria and Archaea share a common prokaryotic genome architecture, the extant diversity of Bacteria appears to be much higher than that of Archaea. Compared with Archaea, Bacteria also show much greater genome-level specialisation to specific ecological niches, including parasitism and endosymbiosis. The reasons for these differences in long-term diversification rates are unclear, but might be related to fundamental differences in informational processing machineries and cell biological features that may favour archaeal diversification in harsher or more energy-limited environments. Finally, phylogenomic analyses suggest that the first Archaea were anaerobic autotrophs that evolved on the early Earth.

## Introduction

One of the major challenges in evolutionary genetics is to explain the enormous variation in genome size and gene content across the tree of life in terms of basic evolutionary processes such as mutation, genetic drift and selection. Phylogenomics suggests that the deepest split in the universal tree lies between the two prokaryotic domains, Bacteria and Archaea, with eukaryotes evolving more recently through a symbiosis between the two [1–4]. But despite the evolutionary and ecological importance of the Archaea, they have not been widely considered [5] in evolutionary genetic accounts for the origins of biodiversity. Until recently, few archaeal genomes were available, and it has been difficult to determine whether Bacteria and Archaea share a common prokaryotic evolutionary regime, or if instead there are conserved differences in the evolutionary processes and macroevolutionary trajectories of the two prokaryotic groups. A better understanding of archaeal evolutionary genetics will be essential for making sense of evolution at the broadest scales, and for testing hypotheses about the differing evolutionary trajectories of prokaryotic and eukaryotic cells [6,7].

In recent years, tremendous progress in the use of cultivation-independent sequencing techniques has greatly improved genomic sampling of both Bacteria and Archaea, and has led to the discovery of major new lineages in the tree of life [8–10]. This wealth of new genomic data allows us to revisit previous work on archaeal comparative genomics and to better characterise the features and processes of archaeal genome evolution. We compare the diversity of modern archaeal and bacterial genomes in terms of genome size and organisation, and ask whether the general principles of genome evolution deduced from the study of other lifeforms also apply to the archaeal domain. In Bacteria and eukaryotes, lifestyle has a profound impact on genome evolution, with the genomes of symbionts and parasites often experiencing extensive remodelling and reduction. We evaluate whether archaeal symbionts

Received: 5 July 2018  
Revised: 26 September 2018  
Accepted: 28 September 2018

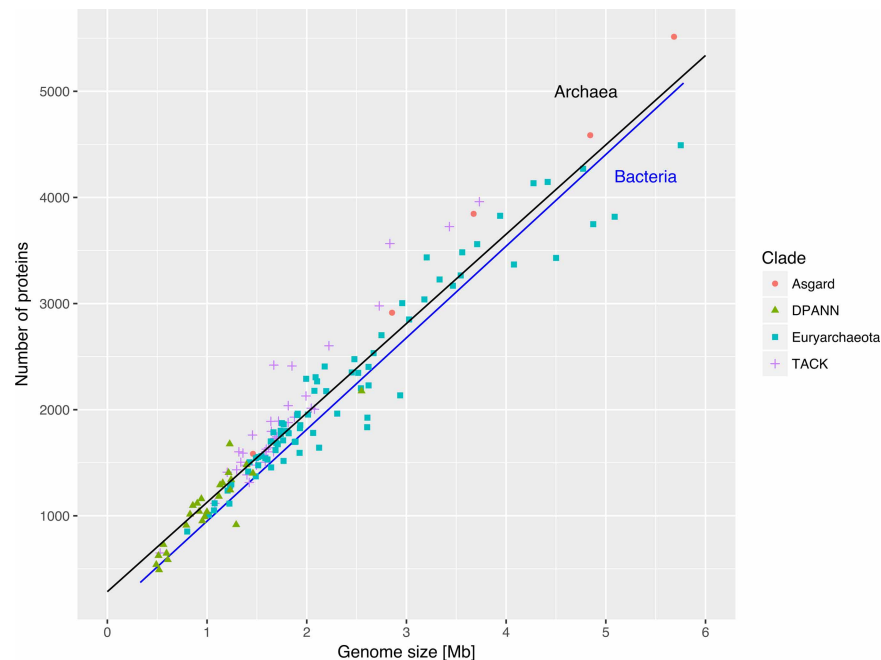
Version of Record published:  
14 November 2018

and parasites, including recently described ultrasmall lineages [9,11,12], evolve in the same way. Finally, we review comparative genomic insights into long-term trends in archaeal genome evolution and the nature of the earliest Archaea.

## A common prokaryotic genome architecture in Bacteria and Archaea

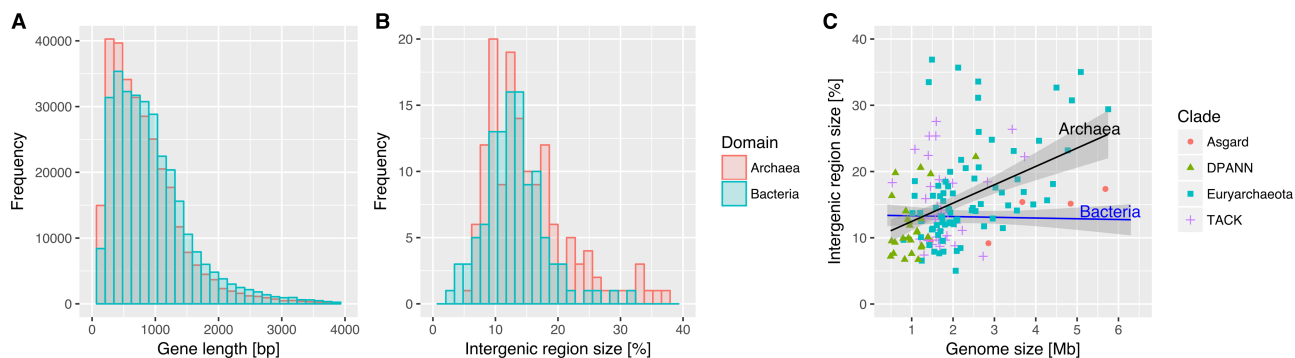
Ten years ago, Koonin and Wolf [5] compared the genome structure and evolution of Bacteria and Archaea using the genomes then available. They concluded that archaeal and bacterial genomes share a common structure, with a main circular chromosome, high gene density, absence of introns and relatively short intergenic spaces [5]. The expanded sample of prokaryotic genomes now available confirms several of Koonin and Wolf's results, including the high gene density and low non-coding content of both archaeal and bacterial genomes (Figures 1 and 2). This common gene-dense prokaryotic genome architecture is thought to arise convergently from the generally high effective population sizes of prokaryotes, which permit efficient selection against the accumulation of nonfunctional or parasitic DNA [13], as well as a bias towards deletions during DNA replication [14,15]. The relatively low absolute numbers of genes encoded by bacterial and archaeal genomes, in contrast, has been proposed to reflect the diminishing adaptive returns of new genes as the number of existing genes in the genome increases [16]. It is interesting to note that genome architectural similarities between Bacteria and Archaea are likely to ultimately reflect similar selective regimes rather than similar molecular biology, because the DNA replication machineries of the two domains are largely non-homologous [17].

Despite the close overall similarity in bacterial and archaeal genome architectures, some differences are also apparent. In Bacteria, the proportion of the genome consisting of intergenic regions is relatively constant across a broad range of genome sizes, while in Archaea — as in eukaryotes [13] — larger genomes contain a greater proportion of intergenic material (Figure 2c). In Archaea [18] and eukaryotes [6], but not Bacteria [16], genome size appears to correlate negatively with the strength of selection. This relaxation of selection is often invoked to explain the proliferation of 'selfish' DNA, such as introns and transposable elements, in the large genomes of multicellular eukaryotes [19]. It is tempting to speculate that a similar process might be at work among the larger archaeal genomes; unfortunately, we still know relatively little about the diversity of selfish DNA in Archaea, and, while characterised families (such as insertion sequences [20]) do vary in abundance among closely related Archaea [21], they do not appear to be more abundant in the larger genomes [18].



**Figure 1. A common gene-dense genome architecture in Bacteria and Archaea.**

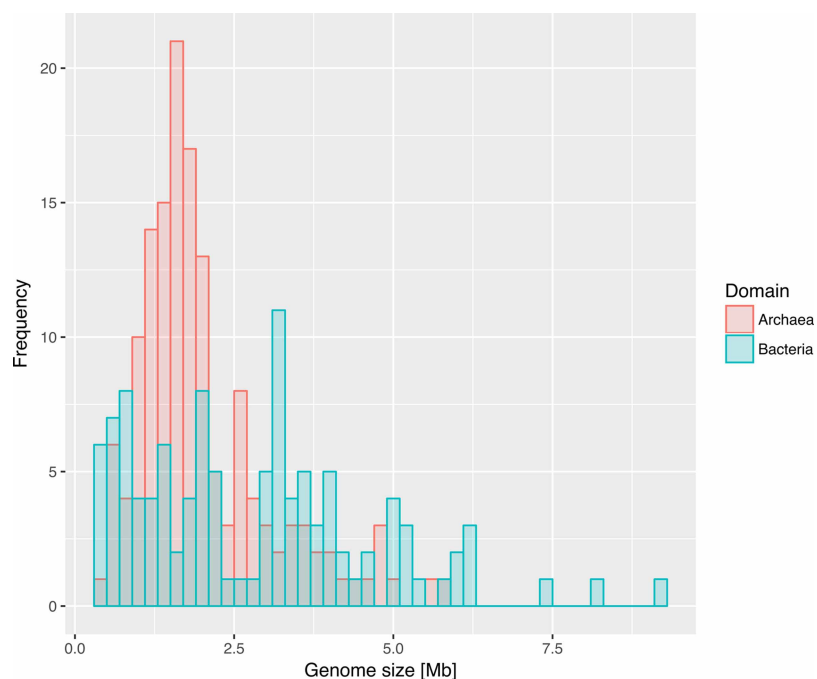
Major archaeal clades are distinguished by shape and colour. The near-linear relationship between genome size and the number of encoded proteins may reflect efficient selection against the accumulation of nonfunctional DNA in prokaryotes [13].



**Figure 2. Genome composition in Bacteria and Archaea.**

Distributions for the lengths of (a) genes and (b) intergenic regions are broadly similar among prokaryotes, with somewhat longer intergenic regions in sampled Archaea. (c) The relationship between genome size and proportion of intergenic material. In Archaea, as in eukaryotes, larger genomes contain a higher proportion of intergenic material ( $P = 2.79 \times 10^{-5}$ , phylogenetic least-squares regression), although this does not appear to be the case for Bacteria.

A second key difference between bacterial and archaeal genomes [5] is the range of genome sizes observed for the two groups. Koonin and Wolf reported that bacterial genome sizes were distributed bimodally, while archaeal genomes were distributed around a single, lower mean. Although Koonin and Wolf argued that this pattern might be explained by a bias towards sequencing parasitic and symbiotic Bacteria, it appears to hold across the much larger range of both bacterial and archaeal diversity now available (Figure 3). The unimodal distribution of genome sizes in Archaea suggests that characterised archaeal symbionts and parasites have not experienced the same degree of reductive genome evolution as in Bacteria; the underlying evolutionary basis



**Figure 3. The distribution of genome sizes among sequenced Archaea (red) and Bacteria (blue).**

Archaea has a unimodal peak at 1.6 Mb, whereas Bacteria show a bimodal peak at ~1.2 and ~3.2 Mb. The size distribution of sampled bacterial genomes has a long tail, extending to at least 14.7 Mb[25], although these outliers do not change the overall distribution. Distributions calculated from a representative sample of Bacteria and Archaea drawn evenly from across the known diversity from recent phylogenomic surveys [8,26].

for these differences remains unclear. Furthermore, the variation of genome size in Archaea appears to be an order of magnitude less than in bacteria[5]; characterised bacterial genome sizes vary ~100-fold, from 0.112 Mb (*Nasuia deltocephalinicola*) to 16.04 Mb (*Minicystis rosea*), while archaeal genomes vary ~10-fold, from 0.49 Mb (the ectosymbiont *Nanoarchaeum equitans*) up to 5.75 Mb (*Methanosarcina acetivorans*).

The smallest genomes from Archaea and Bacteria belong to parasites or symbionts, but free-living members of both groups from nutrient-limited habitats such as open marine waters also possess small genomes. Some of the most abundant ocean bacteria (*Prochlorococcus*, SAR11) are characterised by genomes below 2 Mb [22,23], and the same trend is seen in Archaea: while marine Thaumarchaeota can have genome sizes as low as 1.23 Mb (*Nitrosopelagicus brevis*), characterised relatives from terrestrial environments have genomes ranging up to 3.43 Mb (*Nitrocosmicus oleophilus*). Thus, in marine ecosystems, low nutrient availability may select for minimal genomes and metabolisms [22].

The above patterns are based on a single representative genome for each archaeal ‘species’, but we know that bacterial genome sizes and gene contents can vary substantially over short evolutionary distances; for example, sequenced *E. coli* isolates vary in size by ~20%, from 4.56 to 5.7 Mb [24]. There are few Archaea for which multiple closely related genomes are available; the best-studied case is the crenarchaeon *Sulfolobus islandicus*, which also shows some degree of size variation among the eight completely sequenced isolates (2.47–2.85 Mb). More genomes from closely related Archaea will be needed to evaluate how within-species variation compares between the two prokaryotic groups.

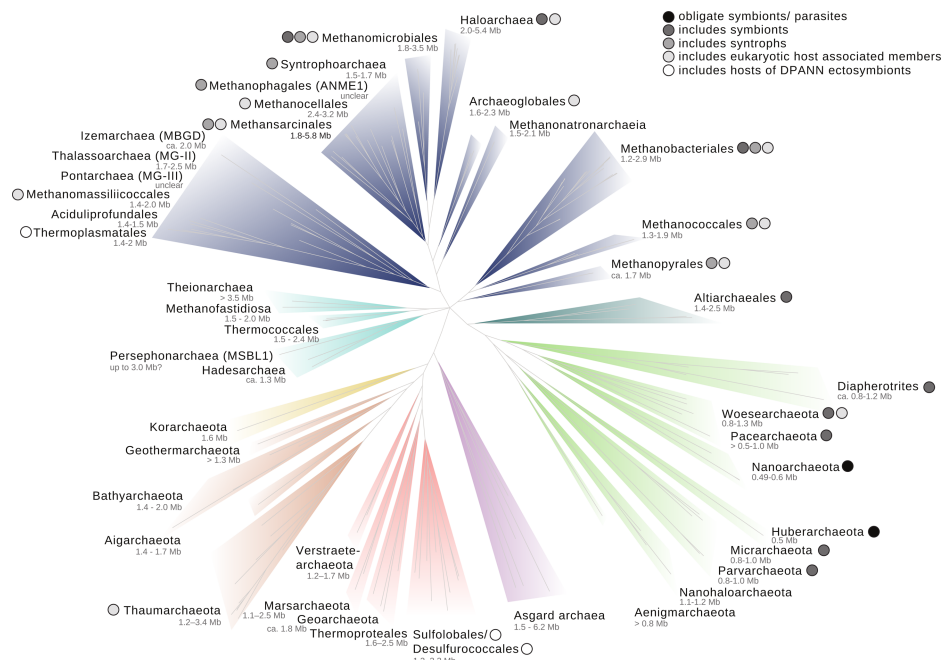
## Genome evolution of host-associated Archaea

Symbioses — mutualistic, commensal and parasitic relationships between organisms [27,28] — are abundant in nature and can have profound consequences for genome evolution. In Bacteria and eukaryotes, the trend is typically toward significant reductive evolution of symbiont genomes — including the loss of genes and pathways needed for a free-living lifestyle [29]. In some cases, this can lead to the complete disintegration and subsequent replacement of the symbiont [30]. However, this extensive reductive evolution is predominantly seen in obligate, vertically transmitted intracellular symbionts: genome sizes of symbionts vary greatly and may even increase when compared with close free-living relatives [31]. This testifies to many different evolutionary trajectories for the size and content of symbiont genomes depending on factors such as the life history of the symbiont and its transmission mode, and may help to make sense of the patterns observed for the genomes of archaeal symbionts.

Very little is known about genome evolution in archaeal symbionts, which is in part due to our limited knowledge of archaeal symbiotic diversity. Many potentially symbiotic, host-associated Archaea have been reported, most notably among members of the Euryarchaeota, Thaumarchaeota and DPANN superphylum [32], suggesting that host-associated lifestyles have evolved repeatedly within the Archaea (Figure 4).

Symbiotic and host-associated Euryarchaeota include methanogens and anaerobic hydrocarbon-oxidising Archaea (ANME and Syntrophoarchaea), as well as halophiles. For example, various methanogens engage in syntrophic interactions with different bacteria and anaerobic fungi [34–36] or form part of the gut microbiome of a large range of animals including humans [37]. Some methanogens and haloarchaea are also ecto- and endosymbionts of diverse anaerobic protists [37–39] where mutualistic interactions were suggested. Several clades within the ammonia-oxidising Thaumarchaeota have consistently been detected in marine sponge and coral microbiomes [40–42] and comprise several putative symbionts — some of which may be transmitted vertically via the larvae [43]. These symbioses appear to be mutualistic or commensal, in which the archaeal symbionts contribute to the detoxification of nitrogen waste products of the host [44,45].

Perhaps the currently most striking and best-understood example of an archaeal host-symbiont system comprises the ectoparasite *N. equitans* and its crenarchaeal host *Ignicoccus hospitalis* [46]. *N. equitans* is dependent on various metabolites from *Ignicoccus* and lowers host proliferation, but does not apparently cause sustained damage [47]. Single cell and metagenomics approaches have recently led to the discovery of various additional clades of ultrasmall genome-reduced Archaea [11,12]. Thus far, phylogenies have suggested that these Archaea may form a monophyletic ‘DPANN’ superphylum which also includes Nanoarchaeota [11,12,48], although phylogenetic artefacts may erroneously group some archaeal lineages within the DPANN [49]. Limited metabolic gene repertoires suggest that DPANN Archaea may be dependent on symbiotic interactions with other organisms [11]. In particular, electron microscopy and co-occurrence analyses have revealed that DPANN members Parva- and Micrarchaeota are commonly found in association with Thermoplasmatales-related hosts, while Huberarchaea may be ectoparasites of Altiarchaea [50–54], which themselves include symbionts [55].



**Figure 4. Distribution of host-associated lineages across the tree of Archaea.**

Host-associated lifestyles have evolved repeatedly within the Archaea. Members of unlabelled groups are assumed to be free-living, as it is currently unknown whether they engage in symbiotic associations. Black numbers denote clades where the genome size ranges are based on complete genomes; grey numbers denote approximate ranges derived from metagenome bins and/or single-celled genomes. The backbone tree is derived from a maximum likelihood analysis (82 concatenated single-copy orthologues, LG + G+F in IQ-Tree [33], some uncertain relationships near the root (for *Theionarchaea*, *Methanofastidiosa*, *Persephonarchaea* and *Thermococcales*) have been collapsed.

Altogether, this indicates that DPANN comprises a largely unexplored diversity of potential novel archaeal ecto- or perhaps even endosymbionts.

### Genomic features of archaeal symbionts?

While systematic and comparative studies of genomic features of archaeal symbionts are lacking, currently available genomes of methanogens, ANME and Syntrophoarchaea cover a relatively broad range of sizes (0.49–5.8 Mb) indicating that the ability of some Archaea to take part in syntrophic interactions is not characteristically associated with a reduced genome and proteome. Much remains to be learned about the molecular and cellular traits involved in the various known syntrophic interactions [35], and although the general features defining archaeal syntrophs remain poorly characterised, recent work has indicated that large multi-heme *c*-type cytochromes mediate syntrophic interactions in Archaea [56].

The most genome-reduced Archaea belong to the DPANN superphylum among which the ectoparasitic Nanoarchaeota and Huberarchaea have the smallest known genomes. For example, the genomes of *N. equitans* [57] and its close relative *Nanopusillus acidilobi* [58] are only 490 and 606 kb in size, respectively, and lack genes for various anabolic and catabolic pathways including a functional ATP synthase [58,59]. In general, members of the DPANN have genomes ranging from ~0.5 to 1.5 Mb in size and many representatives lack genes for central carbon and energy metabolism. In contrast with bacterial endosymbionts, however, the DPANN genomes have a surprisingly high coding density and very few pseudogenes [9] and — despite their sparse metabolic gene repertoires — have retained genes for informational processing machinery ([57], confirmed by a new analysis in Supplementary Table S1). Both DPANN and Asgard archaea encode many uncharacterised proteins, but this enrichment of information processing over metabolism is restricted to DPANN, and might therefore be a hallmark of parasitic or symbiotic lifestyles (Supplementary Table S1). Interestingly, diversity generating retroelements, which contribute to rapid and targeted mutations in specific target genes through reverse transcription, are overrepresented in DPANN genomes [60–62]. While the effects of these elements are

target-specific, their activity may underlie the accelerated evolutionary rates seen for at least some proteins in this group. Despite these initial insights, much has to be learned about DPANN genome evolution and the link between reduced genome sizes and putatively symbiotic lifestyles.

### Are there obligate archaeal endosymbionts?

While an archaeal partner played an important role in the evolution of the eukaryotic cell by the acquisition of a bacterial endosymbiont [1,63,64], other examples of Archaea engaging in obligate relationships with a bacterial or eukaryotic partner are unknown thus far. Although some methanogens form intracellular symbioses with anaerobic protistan hosts, it is not known whether these Archaea are obligate symbionts. Erosion of some of the proteinogenic amino acid biosynthetic pathways in these methanogens might indicate a step toward host dependence [65], but extensive genome reduction has not been observed so far. This is in contrast with the multitude of obligate intracellular bacteria reported from eukaryotic hosts, some of which display extreme genome reduction [66].

### Why are archaea less diverse than bacteria?

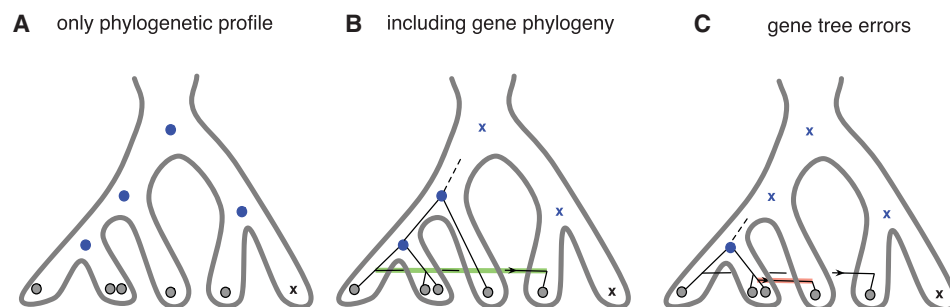
The consensus is that the root of the tree of life lies between the Bacteria and Archaea [67–71], but comprehensive phylogenetic surveys suggest that the diversity of modern Bacteria is much greater than that of the Archaea — as well as of eukaryotes [8,9]. One possibility is that current sampling or sequencing methods provide a biased view of bacterial and archaeal diversity [72]. While this is certainly the case for 16S rRNA-based surveys, it is less of a concern for single cell and in particular for metagenomic approaches, the latter of which target environmental DNA directly [73] and thus circumvent the biases inherent to primer based approaches. Setting aside the potential technical issues, we do not currently have a good explanation for why bacterial and archaeal diversity should be so different, and — since Bacteria and Archaea collectively make up most of life's genetic diversity — this represents a major gap in our understanding of how biodiversity evolves.

One possibility is that the universal root is not between the Bacteria and Archaea. If the root was within the Bacteria, this would provide more time for the accumulation of among-lineage bacterial diversity. Several root positions within Bacteria have been suggested [74–76], but none have received wider support. What little evidence is available from the fossil and geochemical record suggests that Archaea are likely to be quite old, perhaps originating before 3.5 Gya [77,78]. While interpretation of such ancient biomarkers is fraught with difficulty, the antiquity of the Archaea is also supported by recent molecular dating studies combining evidence from gene transfers and relaxed molecular clocks [79,80], with the last archaeal common ancestor (LACA) likely having evolved prior to 3.51 Gya [79,80].

If Bacteria and Archaea are both ancient lineages, then differences in their extant genetic diversity must reflect differences in long-term evolutionary rate, in terms of either mutation rates and selective pressures, or different macroevolutionary histories. Little is known about mutation rates in Archaea, and to the best of our knowledge mutation accumulation data are available for just a single archaeon, the thermophile *Sulfolobus acidocaldarius* [81]. Similar to thermophilic bacteria, the estimated per-base mutation rate in *Sulfolobus* is among the lowest reported for cellular life [82]. Drake [83] proposed that low mutation rates in thermophiles might be selectively advantageous because the average effect of a new mutation is expected to be more deleterious in harsh environments. Conceptually, this shift in the distribution of fitness effects might be thought of as a transition to a rugged fitness landscape in which it is unusually difficult to cross the valleys between the adaptive peaks representing ecotypes or species. Thus, at least for thermophiles, a shift in the distribution of fitness effects might explain both selection for a lower mutation rate and lower long-term rates of diversification [84]. Although most Archaea are not thermophiles, it is tempting to apply this line of reasoning more broadly, because adaptation to harsh conditions of other kinds — particularly energy stress, low energy flux and extremes of pH — have been suggested to be a common feature shared across the archaeal domain [85]. These hypotheses will remain speculative until more data on mutation rates and the distribution of fitness effects in Archaea inhabiting a broad variety of habitats become available.

### Archaeal genome evolution in deep time

The antiquity of the Archaea has led to substantial interest in early archaeal evolution and the nature of LACA, with the aim of providing insight into the metabolisms of the earliest lifeforms and the environments that supported life on the early Earth. As might be expected given the enormous timescales involved, inferences of LACA's genome size and gene content are uncertain, and published estimates vary depending on the



**Figure 5. Ancestral reconstruction using only phylogenetic profiles can lead to artefactually large ancestral gene contents.**

Grey circles denote observed genes in the genomes of present-day organisms; blue circles and crosses denote inferred ancestral presence or absence in ancestral genomes. (a) The phylogenetic profile of this gene family is consistent with presence at all ancestral nodes, with a single loss in the branch leading to one of the modern lineages. (b) The gene family tree indicates that the gene was not present in the common ancestor; instead, it originated later in evolution, but was subsequently transferred into the right-hand side of the species tree. (c) An important caveat of introducing phylogenetic information in the form of gene family trees is that, while it mitigates the systematic bias of profile only methods in overestimating the number of genes in ancestral genomes, it can also lead to an underestimation of the number of genes in ancestral gene contents if errors are present in the gene phylogeny. Here, an incorrect inference of gene transfer places the origin of the gene too recently in the species tree. As discussed in the main text errors in the gene phylogeny can be greatly reduced using species tree aware methods.

reconstruction methods used. Csuros and Miklos [86] used a phylogenetic birth–death model to study the evolution of gene family profiles via gene gain, duplication and loss along a candidate species tree. This method did not use information from the gene phylogenies, but was instead based upon counts of homologous genes on each genome. Analyses under this model suggested that gene loss outnumbered gene gain on most branches of the archaeal tree, so that genomic streamlining from a relatively complex common ancestor was suggested to represent the dominant mode of archaeal genome evolution [86,87].

One potential limitation of profile-based approaches is that, without information from the individual gene trees, there is very limited power to detect horizontal gene transfer unless the gene family has an extremely patchy phylogenetic distribution [88]. As illustrated in Figure 5a,b, this can lead to a systematic overestimation of the number of genes in ancestral genomes [48,86]. This limitation has motivated the development of models that extend the birth–death approach by explicitly considering information from the gene trees, leading to probabilistic gene tree–species tree reconciliation [89]. The main advantage of using species–tree aware gene tree reconstruction methods such as ALE, is that conditional on the species tree being correct, these methods produce dramatically more accurate gene trees [89–92], and correspondingly fewer gene transfer events. This reduction in the number of spurious transfer events caused by errors in the gene tree ameliorates the problem of underestimating the number of genes in ancestral genomes (Figure 5c).

Williams et al. [48] used one such method, ALE [89], to model gene family evolution on the archaeal species tree. In contrast with profile only analyses, the results supported a scenario in which archaeal gene content has gradually increased through time, with *de novo* gene origination, duplication and transfer generally outweighing gene loss. In this analysis, LACA was inferred to have encoded 1090 gene families, rising to ~1500 families among modern Archaea. In this regard, it is important to note that gene tree reconciliation methods (Figure 5) outperform profile only methods, because they are able to distinguish the ancestral gain and subsequent recurrent loss of a gene family from more recent gain followed by gene transfer. As a result, they tend to infer larger rates of transfer and more realistic ancestral genome sizes [48,88]. Thus, the inference of a large ancestral genome in LACA followed by reductive evolution can be explained by the very poor power to detect gene transfer in the absence of evidence from gene tree topologies, and the corresponding systematic inflation of ancestral genome sizes. While some doubt over the ancestral genome size remains, these and other analyses suggest that the Wood–Ljungdahl pathway may have been the earliest carbon fixation pathway in the Archaea [48,93,94], supporting the view that LACA was an anaerobic autotroph.

## Conclusions

Archaeal genome sequencing has lagged behind that for Bacteria, and until the advent of environmental genomics — and the resulting data deluge from abundant but uncultivated microorganisms — it was unclear whether the observed differences between archaeal and bacterial genomes reflected sampling artefacts or biological differences between the domains. Here, we took advantage of the much broader sample of available prokaryotic genome diversity, which allowed us to include genomes representing a broad range of habitats and lifestyles. Our analyses confirm early indications [5] suggesting that, while genome architecture is conserved between Bacteria and Archaea, there appear to be important differences characterising the genomic diversity of the two domains, whether measured in terms of sequence divergence or variation in genome size and coding capacity. The greater genomic malleability of Bacteria is particularly evident for symbionts and parasites: symbiotic Archaea are varied and ecologically important, but — with the important exception of the archaeal host for eukaryote origins — they do not generally experience genome reduction to the same extent as their bacterial counterparts.

The general patterns now seem clear, but we still lack a mechanistic understanding of the evolutionary forces that underlie them. Developing that understanding will require more data from at least two sources. First, we still know very little about the biology and environmental interactions of the many new lineages of Archaea (and indeed Bacteria) that have recently been sequenced using environmental genomics. A more detailed understanding of their lifestyles, and of variation in lifestyle within groups such as the DPANN Archaea, will be critically important in interpreting the broad-scale patterns we have reviewed here. Secondly, testing hypotheses about mutation, selection and diversification will require estimates of the mutation rate and distribution of fitness effects from representative lineages sampled across the archaeal tree, but particularly from mesophilic Archaea. The increasing interest from a broad range of researchers in archaeal genomics and biology, and new techniques for genome-informed cultivation and the study of microbial metabolisms, may now provide the opportunity to begin to explore these questions.

## Summary

- Archaea and Bacteria share common gene-dense prokaryotic genome architecture.
- The range of archaeal genome sizes is much narrower than that of Bacteria. There are many ecologically important archaeal parasites and symbionts, but they are not as extremely reduced as their bacterial counterparts.
- Archaea appear to be as old as Bacteria, but their extant diversity is much lower. We do not know why this is the case.
- The first Archaea were likely anaerobic autotrophs that lived on the early Earth. Their genomes were probably modestly smaller than those of extant Archaea.

## Abbreviations

LACA, last archaeal common ancestor.

## Funding

A.S. is supported by a VR starting grant (2016-03559) from the Swedish Research Council and a WISE fellowship from the NWO-I foundation of the Netherlands Organisation for Scientific Research. G.J.Sz. received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme under grant agreement no. 714774. C. P. is funded by NERC award NE/P000251X/1 to T.A.W., who is supported by a Royal Society University Research Fellowship.

## Acknowledgements

We thank Gareth Coleman for the help with the bacterial proteomes.



## Competing Interests

The Authors declare that there are no competing interests associated with the manuscript.

## References

- 1 Eme, L., Spang, A., Lombard, J., Stairs, C.W. and Ettema, T.J.G. (2017) Archaea and the origin of eukaryotes. *Nat. Rev. Microbiol.* **15**, 711–723 <https://doi.org/10.1038/nrmicro.2017.133>
- 2 Williams, T.A., Foster, P.G., Cox, C.J. and Embley, T.M. (2013) An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* **504**, 231–236 <https://doi.org/10.1038/nature12779>
- 3 Martin, W.F., Garg, S. and Zimorski, V. (2015) Endosymbiotic theories for eukaryote origin. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* **370**, 20140330 <https://doi.org/10.1098/rstb.2014.0330>
- 4 López-García, P. and Moreira, D. (2015) Open questions on the origin of eukaryotes. *Trends Ecol. Evol.* **30**, 697–708 <https://doi.org/10.1016/j.tree.2015.09.005>
- 5 Koonin, E.V. and Wolf, Y.I. (2008) Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* **36**, 6688–6719 <https://doi.org/10.1093/nar/gkn668>
- 6 Lynch, M. and Conery, J.S. (2003) The origins of genome complexity. *Science* **302**, 1401–1404 <https://doi.org/10.1126/science.1089370>
- 7 Lynch, M. and Marinov, G.K. (2017) Membranes, energetics, and evolution across the prokaryote-eukaryote divide. *eLife* **6**. <https://doi.org/10.7554/eLife.20437>
- 8 Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J. et al. (2016) A new view of the tree and life. *Nat. Microbiol.* **1**, 16048 <https://doi.org/10.1038/nmicrobiol.2016.48>
- 9 Castelle, C.J. and Banfield, J.F. (2018) Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell* **172**, 1181–1197 <https://doi.org/10.1016/j.cell.2018.02.016>
- 10 Spang, A., Caceres, E.F. and Ettema, T.J.G. (2017) Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* **357**, eaaf3883 <https://doi.org/10.1126/science.aaf3883>
- 11 Castelle, C.J., Wrighton, K.C., Thomas, B.C., Hug, L.A., Brown, C.T., Wilkins, M.J. et al. (2015) Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* **25**, 690–701 <https://doi.org/10.1016/j.cub.2015.01.014>
- 12 Rinke, C., Schwiertek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.-F. et al. (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 <https://doi.org/10.1038/nature12352>
- 13 Lynch, M. (2006) Streamlining and simplification of microbial genome architecture. *Annu. Rev. Microbiol.* **60**, 327–349 <https://doi.org/10.1146/annurev.micro.60.080805.142300>
- 14 Mira, A., Ochman, H. and Moran, N.A. (2001) Deletional bias and the evolution of bacterial genomes. *Trends Genet.* **17**, 589–596 [https://doi.org/10.1016/S0168-9525\(01\)02447-7](https://doi.org/10.1016/S0168-9525(01)02447-7)
- 15 Kuo, C.-H. and Ochman, H. (2009) Deletional bias across the three domains of life. *Genome Biol. Evol.* **1**, 145–152 <https://doi.org/10.1093/gbe/evp016>
- 16 Sela, I., Wolf, Y.I. and Koonin, E.V. (2016) Theory of prokaryotic genome evolution. *Proc. Natl Acad. Sci. U.S.A.* **113**, 11399–11407 <https://doi.org/10.1073/pnas.1614083113>
- 17 Leipe, D.D., Aravind, L. and Koonin, E.V. (1999) Did DNA replication evolve twice independently? *Nucleic Acids Res.* **27**, 3389–3401 <https://doi.org/10.1093/nar/27.17.3389>
- 18 Lyu, Z., Li, Z.-G., He, F. and Zhang, Z. (2017) An important role for purifying selection in archaeal genome evolution. *mSystems* **2** <https://doi.org/10.1128/mSystems.00112-17>
- 19 Lynch, M., Bobay, L.-M., Catania, F., Gout, J.-F. and Rho, M. (2011) The repatterning of eukaryotic genomes by random genetic drift. *Annu. Rev. Genomics Hum. Genet.* **12**, 347–366 <https://doi.org/10.1146/annurev-genom-082410-101412>
- 20 Filée, J., Siguier, P. and Chandler, M. (2007) Insertion sequence diversity in archaea. *Microbiol. Mol. Biol. Rev.* **71**, 121–157 <https://doi.org/10.1128/MMBR.00031-06>
- 21 Reno, M.L., Held, N.L., Fields, C.J., Burke, P.V. and Whitaker, R.J. (2009) Biogeography of the *Sulfolobus islandicus* pan-genome. *Proc. Natl Acad. Sci. U.S.A.* **106**, 8605–8610 <https://doi.org/10.1073/pnas.0808945106>
- 22 Biller, S.J., Berube, P.M., Lindell, D. and Chisholm, S.W. (2015) Prochlorococcus: the structure and function of collective diversity. *Nat. Rev. Microbiol.* **13**, 13–27
- 23 Giovannoni, S.J. (2017) SAR11 bacteria: the most abundant plankton in the oceans. *Ann. Rev. Mar. Sci.* **9**, 231–255 <https://doi.org/10.1146/annurev-marine-010814-015934>
- 24 Lukjancenko, O., Wassenaar, T.M. and Ussery, D.W. (2010) Comparison of 61 sequenced *Escherichia coli* genomes. *Microb. Ecol.* **60**, 708–720 <https://doi.org/10.1007/s00248-010-9717-3>
- 25 Han, K., Li, Z.-F., Peng, R., Zhu, L.-P., Zhou, T., Wang, L.-G. et al. (2013) Extraordinary expansion of a *Sorangium cellulosum* genome from an alkaline milieu. *Sci. Rep.* **3**, 2101 <https://doi.org/10.1038/srep02101>
- 26 Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B.J., Evans, P.N. et al. (2018) Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* **3**, 253 <https://doi.org/10.1038/s41564-017-0083-5>
- 27 Raina, J.-B., Eme, L., Pollock, F.J., Spang, A., Archibald, J.M. and Williams, T.A. (2018) Symbiosis in the microbial world: from ecology to genome evolution. *Biol. Open* **7**. <https://doi.org/10.1242/bio.032524>
- 28 López-García, P., Eme, L. and Moreira, D. (2017) Symbiosis in eukaryotic evolution. *J. Theor. Biol.* **434**, 20–33 <https://doi.org/10.1016/j.jtbi.2017.02.031>
- 29 Moran, N.A. and Bennett, G.M. (2014) The tiniest tiny genomes. *Annu. Rev. Microbiol.* **68**, 195–215 <https://doi.org/10.1146/annurev-micro-091213-112901>
- 30 Matsuura, Y., Moriyama, M., Łukasik, P., Vanderpool, D., Tanahashi, M., Meng, X.-Y. et al. (2018) Recurrent symbiont recruitment from fungal parasites in cicadas. *Proc. Natl Acad. Sci. U.S.A.* **115**, E5970–E5979 <https://doi.org/10.1073/pnas.1803245115>
- 31 Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiyumi, T., Sasamoto, S. et al. (2002) Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res.* **9**, 189–197 <https://doi.org/10.1093/dnares/9.6.189>

- 32 Moissl-Eichinger, C., Probst, A.J., Birarda, G., Auerbach, A., Koskinen, K., Wolf, P. et al. (2017) Human age and skin physiology shape diversity and abundance of Archaea on skin. *Sci. Rep.* **7**, 4039 <https://doi.org/10.1038/s41598-017-04197-4>
- 33 Nguyen, L.-T., Schmidt, H.A., Von Haeseler, A. and Minh, B.Q. (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 <https://doi.org/10.1093/molbev/msu300>
- 34 Stams, A.J.M. and Plugge, C.M. (2009) Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nat. Rev. Microbiol.* **7**, 568–577 <https://doi.org/10.1038/nrmicro2166>
- 35 Sieber, J.R., McInerney, M.J. and Gunsalus, R.P. (2012) Genomic insights into syntrophy: the paradigm for anaerobic metabolic cooperation. *Annu. Rev. Microbiol.* **66**, 429–452 <https://doi.org/10.1146/annurev-micro-090110-102844>
- 36 Leis, S., Dresch, P., Peintner, U., Fliegerová, K., Sandbichler, A.M., Insam, H. et al. (2014) Finding a robust strain for biomethanation: anaerobic fungi (Neocallimastigomycota) from the Alpine ibex (*Capra ibex*) and their associated methanogens. *Anaerobe* **29**, 34–43 <https://doi.org/10.1016/j.anaerobe.2013.12.002>
- 37 Lange, M., Westermann, P. and Ahring, B.K. (2005) Archaea in protozoa and metazoa. *Appl. Microbiol. Biotechnol.* **66**, 465–474 <https://doi.org/10.1007/s00253-004-1790-4>
- 38 Embley, T.M., Finlay, B.J., Dyal, P.L., Hirt, R.P., Wilkinson, M. and Williams, A.G. (1995) Multiple origins of anaerobic ciliates with hydrogenosomes within the radiation of aerobic ciliates. *Proc. Biol. Sci.* **262**, 87–93 <https://doi.org/10.1098/rspb.1995.0180>
- 39 Filker, S., Kaiser, M., Rosselló-Móra, R., Dunthorn, M., Lax, G. and Stoeck, T. (2014) ‘*Candidatus* Haloectosymbiotes riaformosensis’ (Halobacteriaceae), an archaeal ectosymbiont of the hypersaline ciliate *Platynematum salinarum*. *Syst. Appl. Microbiol.* **37**, 244–251 <https://doi.org/10.1016/j.syapm.2014.01.001>
- 40 Webster, N.S., Watts, J.E. and Hill, R.T. (2001) Detection and phylogenetic analysis of novel crenarchaeote and euryarchaeote 16S ribosomal RNA gene sequences from a Great Barrier Reef sponge. *Mar. Biotechnol.* **3**, 600–608 <https://doi.org/10.1007/s10126-001-0065-7>
- 41 Holmes, B. and Blanch, H. (2007) Genus-specific associations of marine sponges with group I crenarchaeotes. *Mar. Biol.* **150**, 759–772 <https://doi.org/10.1007/s00227-006-0361-x>
- 42 Alves, R.J.E., Minh, B.Q., Urich, T., von Haeseler, A. and Schleper, C. (2018) Unifying the global phylogeny and environmental distribution of ammonia-oxidising archaea based on amoA genes. *Nat. Commun.* **9**, 1517 <https://doi.org/10.1038/s41467-018-03861-1>
- 43 Steger, D., Ettinger-Epstein, P., Whalan, S., Hentschel, U., de Nys, R., Wagner, M. et al. (2008) Diversity and mode of transmission of ammonia-oxidizing archaea in marine sponges. *Environ. Microbiol.* **10**, 1087–1094 <https://doi.org/10.1111/j.1462-2920.2007.01515.x>
- 44 Hentschel, U., Piel, J., Degnan, S.M. and Taylor, M.W. (2012) Genomic insights into the marine sponge microbiome. *Nat. Rev. Microbiol.* **10**, 641–654 <https://doi.org/10.1038/nrmicro2839>
- 45 Radax, R., Hoffmann, F., Rapp, H.T., Leininger, S. and Schleper, C. (2012) Ammonia-oxidizing archaea as main drivers of nitrification in cold-water sponges. *Environ. Microbiol.* **14**, 909–923 <https://doi.org/10.1111/j.1462-2920.2011.02661.x>
- 46 Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C. and Stetter, K.O. (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**, 63–67 <https://doi.org/10.1038/417063a>
- 47 Jahn, U., Gallenberger, M., Paper, W., Junglas, B., Eisenreich, W., Stetter, K.O. et al. (2008) *Nanoarchaeum equitans* and *Ignicoccus hospitalis*: new insights into a unique, intimate association of two archaea. *J. Bacteriol.* **190**, 1743–1750 <https://doi.org/10.1128/JB.01731-07>
- 48 Williams, T.A., Szöllösi, G.J., Spang, A., Foster, P.G., Heaps, S.E., Boussau, B. et al. (2017) Integrative modeling of gene and genome evolution roots the archaeal tree of life. *Proc. Natl Acad. Sci. U.S.A.* **114**, E4602–E4611 <https://doi.org/10.1073/pnas.1618463114>
- 49 Aouad, M., Taib, N., Oudart, A., Lecocq, M., Gouy, M. and Brochier-Armanet, C. (2018) Extreme halophilic archaea derive from two distinct methanogen Class II lineages. *Mol. Phylogenet. Evol.* **127**, 46–54 <https://doi.org/10.1016/j.ympev.2018.04.011>
- 50 Baker, B.J., Comolli, L.R., Dick, G.J., Hauser, L.J., Hyatt, D., Dill, B.D. et al. (2010) Enigmatic, ultrasmall, uncultivated Archaea. *Proc. Natl Acad. Sci. U.S.A.* **107**, 8806–8811 <https://doi.org/10.1073/pnas.0914470107>
- 51 Golyshina, O.V., Toshchakov, S.V., Makarova, K.S., Gavrillov, S.N., Korzhnikov, A.A., La Cono, V. et al. (2017) ‘ARMAN’ archaea depend on association with euryarchaeal host in culture and in situ. *Nat. Commun.* **8**, 60 <https://doi.org/10.1038/s41467-017-00104-7>
- 52 Krause, S., Bremges, A., Münch, P.C., McHardy, A.C. and Gescher, J. (2017) Characterisation of a stable laboratory co-culture of acidophilic nanoorganisms. *Sci. Rep.* **7**, 3289 <https://doi.org/10.1038/s41598-017-03315-6>
- 53 Comolli, L.R., Baker, B.J., Downing, K.H., Siegerist, C.E. and Banfield, J.F. (2009) Three-dimensional analysis of the structure and ecology of a novel, ultra-small archaeon. *ISME J.* **3**, 159–167 <https://doi.org/10.1038/ismej.2008.99>
- 54 Probst, A.J. and Banfield, J.F. (2018) Homologous recombination and transposon propagation shape the population structure of an organism from the deep subsurface with minimal metabolism. *Genome Biol. Evol.* **10**, 1115–1119 <https://doi.org/10.1093/gbe/evy067>
- 55 Probst, A.J., Weinmaier, T., Raymann, K., Perras, A., Emerson, J.B., Rattei, T. et al. (2014) Biology of a widespread uncultivated archaeon that contributes to carbon fixation in the subsurface. *Nat. Commun.* **5**, 1–13 <https://doi.org/10.1038/ncomms6497>
- 56 McGlynn, S.E., Chadwick, G.L., Kempes, C.P. and Orphan, V.J. (2015) Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature* **526**, 531–535 <https://doi.org/10.1038/nature15512>
- 57 Waters, E., Hohn, M.J., Ahel, I., Graham, D.E., Adams, M.D., Barnstead, M. et al. (2003) The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc. Natl Acad. Sci. U.S.A.* **100**, 12984–12988 <https://doi.org/10.1073/pnas.1735403100>
- 58 Wurch, L., Giannone, R.J., Belisle, B.S., Swift, C., Utturkar, S., Hettich, R.L. et al. (2016) Genomics-informed isolation and characterization of a symbiotic Nanoarchaeota system from a terrestrial geothermal environment. *Nat. Commun.* **7**, 12115 <https://doi.org/10.1038/ncomms12115>
- 59 Mohanty, S., Jobichen, C., Chichili, V.P.R., Velázquez-Campoy, A., Low, B.C., Hogue, C.W.V. et al. (2015) Structural basis for a unique ATP synthase core complex from *Nanoarchaeum equitans*. *J. Biol. Chem.* **290**, 27280–27296 <https://doi.org/10.1074/jbc.M115.677492>
- 60 Paul, D., Kumbhare, S.V., Mhatre, S.S., Chowdhury, S.P., Shetty, S.A., Marathe, N.P. et al. (2015) Exploration of microbial diversity and community structure of Lonar Lake: the only hypersaline meteorite crater lake within basalt rock. *Front Microbiol.* **6**, 1553 PMID:26834712
- 61 Paul, B.G., Burstner, D., Castelle, C.J., Handa, S., Arambula, D., Czornyj, E. et al. (2017) Retroelement-guided protein diversification abounds in vast lineages of Bacteria and Archaea. *Nat. Microbiol.* **2**, 17045 <https://doi.org/10.1038/nmicrobiol.2017.45>
- 62 Wu, L., Gingery, M., Abebe, M., Arambula, D., Czornyj, E., Handa, S. et al. (2018) Diversity-generating retroelements: natural variation, classification and evolution inferred from a large-scale genomic survey. *Nucleic Acids Res.* **46**, 11–24 <https://doi.org/10.1093/nar/gkx1150>

- 63 Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E. et al. (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179 <https://doi.org/10.1038/nature14447>
- 64 Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, D., Juzokaitė, L., Vancaester, E. et al. (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 <https://doi.org/10.1038/nature21031>
- 65 Lind, A.E., Lewis, W.H., Spang, A., Guy, L., Embley, T.M. and Ettema, T.J.G. (2018) Genomes of two archaeal endosymbionts show convergent adaptations to an intracellular lifestyle. *ISME J.* <https://doi.org/10.1038/s41396-018-0207-9>
- 66 Keeling, P.J. and McCutcheon, J.P. (2017) Endosymbiosis: the feeling is not mutual. *J. Theor. Biol.* **434**, 75–79 <https://doi.org/10.1016/j.jtbi.2017.06.008>
- 67 Brown, J.R. and Doolittle, W.F. (1995) Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. *Proc. Natl Acad. Sci. U.S.A.* **92**, 2441–2445 <https://doi.org/10.1073/pnas.92.7.2441>
- 68 Iwabe, N., Kumada, K., Hasegawa, M., Osawa, S. and Miyata, T. (1989) Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl Acad. Sci. U.S.A.* **86**, 9355–9359 <https://doi.org/10.1073/pnas.86.23.9355>
- 69 Gogarten, J.P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E.J., Bowman, B.J. et al. (1989) Evolution of the vacuolar H<sup>+</sup>-ATPase: implications for the origin of eukaryotes. *Proc. Natl Acad. Sci. U.S.A.* **86**, 6661–6665 <https://doi.org/10.1073/pnas.86.17.6661>
- 70 Dagan, T., Roettger, M., Bryant, D. and Martin, W. (2010) Genome networks root the tree of life between prokaryotic domains. *Genome Biol. Evol.* **2**, 379–392 <https://doi.org/10.1093/gbe/evq025>
- 71 Zhaxybayeva, O., Lapierre, P. and Gogarten, J.P. (2005) Ancient gene duplications and the root(s) of the tree of life. *Protoplasma* **227**, 53–64 <https://doi.org/10.1007/s00709-005-0135-1>
- 72 Eloe-Fadrosh, E.A., Ivanova, N.N., Woyke, T. and Kyrpides, N.C. (2016) Metagenomics uncovers gaps in amplicon-based detection of microbial diversity. *Nat. Microbiol.* **1**, 15032 PMID:27572438
- 73 Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* **68**, 669–685 <https://doi.org/10.1128/MMBR.68.4.669-685.2004>
- 74 Cavalier-Smith, T. (2006) Rooting the tree of life by transition analyses. *Biol Direct* **1**, 19 <https://doi.org/10.1186/1745-6150-1-19>
- 75 Lake, J.A., Skophammer, R.G., Herbold, C.W. and Servin, J.A. (2009) Genome beginnings: rooting the tree of life. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* **364**, 2177–2185 <https://doi.org/10.1098/rstb.2009.0035>
- 76 Williams, T.A., Heaps, S.E., Cherlin, S., Nye, T.M.W., Boys, R.J., Embley, T.M. et al. (2015) New substitution models for rooting phylogenetic trees. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* **370**, 20140336 <https://doi.org/10.1098/rstb.2014.0336>
- 77 Schopf, J.W., Kitajima, K., Spicuzza, M.J., Kudryavtsev, A.B. and Valley, J.W. (2018) SIMS analyses of the oldest known assemblage of microfossils document their taxon-correlated carbon isotope compositions. *Proc. Natl Acad. Sci. U.S.A.* **115**, 53–58 <https://doi.org/10.1073/pnas.1718063115>
- 78 Ueno, Y., Yamada, K., Yoshida, N., Maruyama, S. and Isozaki, Y. (2006) Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. *Nature* **440**, 516–519 <https://doi.org/10.1038/nature04584>
- 79 Davin, A.A., Tannier, E., Williams, T.A., Boussau, B., Daubin, V. and Szöllösi, G.J. (2018) Gene transfers can date the tree of life. *Nat. Ecol. Evol.* **2**, 904–909 <https://doi.org/10.1038/s41559-018-0525-3>
- 80 Wolfe, J.M. and Fournier, G.P. (2018) Horizontal gene transfer constrains the timing of methanogen evolution. *Nat. Ecol. Evol.* **2**, 897–903 <https://doi.org/10.1038/s41559-018-0513-7>
- 81 Grogan, D.W., Carver, G.T. and Drake, J.W. (2001) Genetic fidelity under harsh conditions: analysis of spontaneous mutation in the thermoacidophilic archaeon *Sulfolobus acidocaldarius*. *Proc. Natl Acad. Sci. U.S.A.* **98**, 7928–7933 <https://doi.org/10.1073/pnas.141113098>
- 82 Lynch, M. (2010) Evolution of the mutation rate. *Trends Genet.* **26**, 345–352 <https://doi.org/10.1016/j.tig.2010.05.003>
- 83 Drake, J.W. (2009) Avoiding dangerous missense: thermophiles display especially low mutation rates. *PLoS Genet.* **5**, e1000520 <https://doi.org/10.1371/journal.pgen.1000520>
- 84 Groussin, M. and Gouy, M. (2011) Adaptation to environmental temperature is a major determinant of molecular evolutionary rates in archaea. *Mol. Biol. Evol.* **28**, 2661–2674 <https://doi.org/10.1093/molbev/msr098>
- 85 Valentine, D.L. (2007) Adaptations to energy stress dictate the ecology and evolution of the Archaea. *Nat. Rev. Microbiol.* **5**, 316–323 <https://doi.org/10.1038/nrmicro1619>
- 86 Csurös, M. and Miklós, I. (2009) Streamlining and large ancestral genomes in Archaea inferred with a phylogenetic birth-and-death model. *Mol. Biol. Evol.* **26**, 2087–2095 <https://doi.org/10.1093/molbev/msp123>
- 87 Wolf, Y.I. and Koonin, E.V. (2013) Genome reduction as the dominant mode of evolution. *Bioessays* **35**, 829–837 <https://doi.org/10.1002/bies.201300037>
- 88 Szöllösi, G.J., Davin, A.A., Tannier, E., Daubin, V. and Boussau, B. (2015) Genome-scale phylogenetic analysis finds extensive gene transfer among fungi. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* **370**, 20140335 <https://doi.org/10.1098/rstb.2014.0335>
- 89 Szöllösi, G.J., Roskiewicz, W., Boussau, B., Tannier, E. and Daubin, V. (2013) Efficient exploration of the space of reconciled gene trees. *Syst. Biol.* **62**, 901–912 <https://doi.org/10.1093/sysbio/syt054>
- 90 Patterson, M., Szöllösi, G., Daubin, V. and Tannier, E. (2013) Lateral gene transfer, rearrangement, reconciliation. *BMC Bioinf.* **14**, S4 <https://doi.org/10.1186/1471-2105-14-S15-S4>
- 91 Groussin, M., Hobbs, J.K., Szöllösi, G.J., Gribaldo, S., Arcus, V.L. and Gouy, M. (2015) Toward more accurate ancestral protein genotype-phenotype reconstructions with the use of species tree-aware gene trees. *Mol. Biol. Evol.* **32**, 13–22 <https://doi.org/10.1093/molbev/msu305>
- 92 Scornavacca, C., Jacox, E. and Szöllösi, G.J. (2015) Joint amalgamation of most parsimonious reconciled gene trees. *Bioinformatics* **31**, 841–848 <https://doi.org/10.1093/bioinformatics/btu728>
- 93 Sousa, F.L. and Martin, W.F. (2014) Biochemical fossils of the ancient transition from geoenergetics to bioenergetics in prokaryotic one carbon compound metabolism. *Biochim. Biophys. Acta* **1837**, 964–981 <https://doi.org/10.1016/j.bbabi.2014.02.001>
- 94 Adam, P.S., Borrel, G. and Gribaldo, S. (2018) Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proc. Natl Acad. Sci. U.S.A.* **115**, E1166–E1173 <https://doi.org/10.1073/pnas.1716667115>