



Cite this article: Mahendrarajah TA, Spang A. 2025 DPANN Archaea and CPR Bacteria: insights into early cellular evolution? *Phil. Trans. R. Soc. B* **380**: 20240096.
<https://doi.org/10.1098/rstb.2024.0096>

Received: 1 November 2024

Accepted: 16 December 2024

One contribution of 21 to a discussion meeting issue 'Chance and purpose in the evolution of biospheres'.

Subject Areas:

evolution, microbiology, palaeontology, genomics

Keywords:

tree of life, archaea, bacteria, last universal common ancestor, CPR, DPANN

Authors for correspondence:

Tara A. Mahendrarajah

e-mail: tara.mahendrarajah@nioz.nl

Anja Spang

e-mail: Anja.Spang@nioz.nl

DPANN Archaea and CPR Bacteria: insights into early cellular evolution?

Tara A. Mahendrarajah¹ and Anja Spang^{1,2}

¹Marine Microbiology and Biogeochemistry, Royal Netherlands Institute for Sea Research, Den Burg, Noord-Holland, The Netherlands

²Evolutionary & Population Biology (IBED-EPB), University of Amsterdam Faculty of Science, Amsterdam, The Netherlands

TAM, 0000-0001-7032-6581

Cultivation-independent techniques have facilitated the discovery of a diversity of thus far uncultivated organisms and led to the identification of new branches within the tree of life. This has reshaped the general view of key evolutionary processes and challenged fundamental understandings of cellular evolution. Two diverse and phylogenetically relevant lineages are the Diapherotrites, Parv-, Aenigma-, Nano- and Nanohaloarchaeota (DPANN) archaea and the Candidate Phyla Radiation (CPR) bacteria, originally proposed to form sister groups of all other Archaea and Bacteria, respectively. Members of the DPANN and CPR have reduced genome and cell sizes and incomplete biosynthetic pathways. In addition, the co-cultivation of the first DPANN and CPR representatives with their respective hosts indicates that they comprise many symbionts. While initial work has suggested that these two prokaryotic groups could have important implications for understanding cellular evolution and the complexity of the ancestors of the Archaea and Bacteria, respectively, more recent data have indicated that they may be derived organisms or have evolved in parallel with their partner organisms. Here, we provide insights into debates on the phylogenetic placement of these radiations and discuss implications for cellular evolution and the last universal common ancestor (LUCA) in light of Earth history.

This article is part of the discussion meeting issue 'Chance and purpose in the evolution of biospheres'.

1. Introduction

Cellular life on Earth is extremely diverse. Aside from eukaryotes that comprise much of the visible biosphere, the vast majority of cellular organisms consists of prokaryotes. Prokaryotes are represented by two primary domains of life, the Archaea and Bacteria [1], members of which have colonized most environments on Earth, thriving under a wide range of conditions (e.g. temperature, pH, pressure, nutrient and trace gas composition). In recent years, advances in environmental sampling and (meta-)genome sequencing have provided a window into the uncultivated biosphere, expanding our understanding of microbial diversity and uncovering newly defined branches in the growing tree of life [2–4]. The tree of life is a network that provides insight into evolutionary relationships between and among extant prokaryotes and eukaryotes from a shared ancestor. Contemporary views of the tree of life highlight the significance of prokaryotes, which have shaped planetary ecosystems for most of Earth's history and underlie the origin of the eukaryotic cell that emerged around 2.0 billion years ago (Ga) [5–11].

All known cellular life shares characteristics that indicate descent from a single ancestor or cellular population, more commonly known as the

last universal common ancestor (LUCA) [12–14]. The nature of LUCA, its metabolic features and timing of divergence into Archaea and Bacteria [1] are foundational to understanding cellular evolution. The recently discovered DPANN (named after the first discovered lineages: Diapherotrites, Parv-, Aenigma-, Nano- and Nanohaloarchaeota) [15,16] and the CPR (i.e. the Candidate Phyla Radiation, also referred to as Patescibacteria) [16–18] have been suggested to be instrumental for addressing unanswered questions surrounding the earliest periods of cellular evolution. Specifically, both the DPANN and CPR lineages were originally positioned as sister clades to all other Archaea and Bacteria, respectively, in phylogenetic analyses (figure 1A) [3,16,21–23], raising the question as to whether members of these groups may more closely resemble LUCA than other archaeal and bacterial lineages [21,24,25].

However, speculations about a simple ancestor inspired by these findings are problematic for several reasons. First of all, most DPANN and CPR representatives with reduced genomes appear to be symbionts that are obligately dependent on a host organism for growth and survival [17,21,26–31]. It is unclear how a cell with such biosynthetic limitations could have sustained a lifestyle independent from (a) host organism(s) in the primordial environment and give rise to all extant life on Earth. In addition, the genomes of symbionts often evolve at faster rates and are characterized by compositionally biased nucleotide compositions, which can contribute to phylogenetic artefacts such as long branch attraction (LBA) [32,33], where fast-evolving taxa are erroneously attracted towards a distant outgroup unless more complex phylogenetic approaches are employed. Thus, it is possible that the early divergence of DPANN and CPR from all other Archaea and Bacteria, respectively, is a result of such artefacts (figure 1A) [20,34–38]. Even if the early divergence of the ancestors of DPANN and CPR were to be correct, it is important to note that they still share a common ancestor with their respective archaeal and bacterial sister-lineage predominantly [39] comprising organisms with lifestyles not obligately dependent on a host organism (i.e. host-independent) (figure 1A). Therefore, tree topology alone cannot be used to infer whether the last bacterial common ancestor (LBCA) and the last archaeal common ancestor (LACA) had a limited metabolic repertoire and a host-dependent lifestyle or whether they were more reminiscent of a *bona fide* bacterium and archaeon, respectively.

In recent years, the increased quality and quantity of genomic reconstructions have greatly expanded the known diversity of members belonging to both the DPANN and CPR lineages. Their integration into careful phylogenetic and ancestral reconstruction analyses is imperative to gain comprehensive insight into cellular evolution and understand the evolutionary mechanisms that give rise to symbiotic lifestyles. Here, we discuss recent phylogenetic discoveries that challenge the initial view of the early divergence of DPANN and CPR for our understanding of early cellular evolution and the divergence of the major domains of life (Archaea and Bacteria).

2. The Candidate Phyla Radiation is a derived sister lineage of the Chloroflexota

The CPR has been suggested to represent a substantial portion of bacterial diversity [3,17,18]. Various phylogenetic reconstructions using different approaches, ranging from traditional concatenation-based methods with a small set of universally conserved ribosomal proteins [3] to supertree methods using a greatly expanded set of marker genes including metabolic genes [23], have initially inferred a placement of this lineage as a sister group to all other bacterial lineages (figure 1A) [3,23]. This finding led to speculations as to whether the small genome-reduced cells, with patchy biosynthetic capabilities, might reflect the cell biology and lifestyle of LBCA and possibly of earlier ancestors along the bacterial stem.

However in recent years, comparative genomic and phylogenetic approaches integrating better-fitting models of evolution and accounting for the various forms of gene transmission including transfers, losses and originations (reviewed in [40]) have challenged the hypothesis of the early divergence of the branch leading to CPR from the rest of the Bacteria, instead resolving them as a sister group to the Chloroflexota within the Terrabacteria (figure 1B) [19,41–43]. For instance, using outgroup-free rooting methods based on gene tree–species tree reconciliations and distinct taxonomic datasets, Coleman *et al.* found a derived position for CPR, with this clade sharing a common ancestor with Chloroflexota and Dormibacterota (figure 1B) [19]. Gene tree–species tree reconciliation approaches [44] furthermore indicated that LBCA was a relatively complex, motile and rod-shaped cell, with a double-membrane and evidence of the presence of several core carbohydrate-processing pathways [19]. This is in agreement with another study that suggested a diderm ancestor of all Bacteria, with a possible loss of the outer membrane along the stem to the CPR/Chloroflexota [43,45]. In turn, current data are best in line with the view that LBCA was a cell similar to extant bacteria, with CPR representing a derived lineage that evolved more recently via genome reduction from a host-independent ancestor.

3. Diapherotrites, Parv-, Aenigma-, Nano- and Nanohaloarchaeota, an early diverging lineage with host-independent lifestyles?

Similar to the CPR, the DPANN archaea [16] comprise organisms with ‘ultrasmall’ cells, reduced genomes and limited metabolic capabilities [15,16]. They are proposed to represent a substantial proportion of archaeal diversity [3], currently comprising 10 archaeal phyla [46]. Initial phylogenetic analyses [3,15,16] identified the DPANN as a monophyletic clade that forms a sister group to all other Archaea within the archaeal tree (figure 1A). A more recent outgroup-free rooting analysis also recovered a root between the DPANN and all other Archaea [22]; however, the number of DPANN genomes included in that analysis was limited and does not reflect the currently known diversity of the group. Indeed, a more recent study integrating a better representation of DPANN could not unequivocally determine the position of the root of the archaeal domain, challenging previously proposed scenarios on the group’s phylogenetic position and monophyly (figure 1B) [20,47]. The placement of the

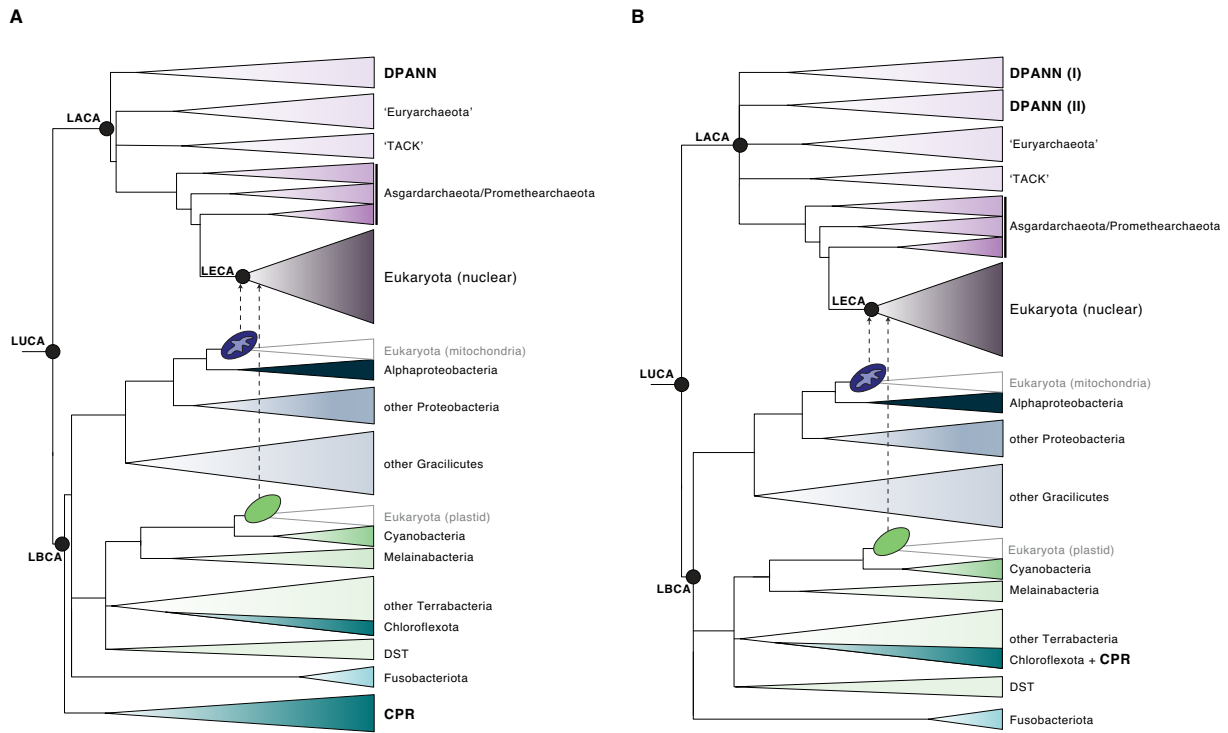


Figure 1. A depiction of the tree of life highlighting shifts in our understanding of the placements of the DPANN archaea and CPR bacteria. The schematic trees are based on recent phylogenetic and molecular dating analyses presented in [19] and [8], and they approximately reflect relative evolutionary distance. (A) Initial phylogenetic inferences positioned the Diapherotrites, Parv-, Aenigma-, Nano-, and Nanohaloarchaeota (DPANN) and the Candidate Phyla Radiation (CPR) clades as the sister clades to all other Archaea and Bacteria, respectively. (B) More recent phylogenetic evidence supports a derived position for the CPR within the Terrabacteria, with the group sharing a common ancestor with the Chloroflexota/Dormibacterota, while the monophyly of DPANN and the root of the archaeal tree remain unresolved. LUCA, LACA, LBCA and LECA refer to the last universal common ancestor, the last archaeal common ancestor, the last bacterial common ancestor and the last eukaryotic common ancestor, respectively. DST refers to the clade comprising Deinococcota, Synergistota and Thermotogota. DPANN (I) refers to DPANN Cluster 1 and DPANN (II) refers to DPANN Cluster 2 [20]. The blue and green oval-shaped cells depict the mitochondrial and plastid endosymbiosis events, respectively.

root of the archaeal phylogeny remains unresolved, complicating our inferences of LACA (figure 1B). Consequently, the debate over the monophyly of the DPANN, their possible derived nature and co-evolution with their hosts remains.

Lineages such as the Nanohaloarchaeota, which have been identified in hypersaline environments associating with the archaeal class Halobacteria [16,28,48], were initially assigned to the Euryarchaeota [34,48,49]. While the Nanohaloarchaeota regularly cluster within the DPANN and exhibit characteristic traits such as extremely small cell sizes and patchy biosynthetic pathways, critiques have argued that they are erroneously attracted toward the DPANN owing to amino acid composition biases in their proteome, which are a consequence of adaptation to hypersaline environments [34,36]. For example, although Feng *et al.* [36] recovered Nanohaloarchaeota as part of the DPANN based on concatenated marker gene tree analyses, they argued that the clustering of ATP synthase subunits of Nanohaloarchaeota with those of Halobacteria indicates a close phylogenetic relationship of these lineages. In contrast, Wang *et al.* [50] suggested that horizontal transfer of the ATP synthase operon from halobacterial hosts into the nanohaloarchaeal symbionts best explains the relationship of their ATP synthase subunits. In our recent work on the early evolution of the ATP synthase across the tree of life, the catalytic headpiece subunit sequences of the Nanohaloarchaeota cluster with those of other DPANN archaea rather than with Halobacteria, while the non-catalytic headpiece subunit sequences cluster with the Halobacteria [8], likely as a result of compositional attraction or host-symbiont gene transfer. In line with this, a recent study integrating careful modelling of evolutionary processes of halophilic archaeal lineages has confirmed the placement of Nanohaloarchaeota with other DPANN [51] and that certain nanohaloarchaeal and halobacterial protein homologues are phylogenetically attracted to each other due to compositional biases resulting from adaptations to an environment with high salt concentrations [8,51].

It is notable that DPANN lineages seem to form at least two distinct sub-clades (Cluster 1 and Cluster 2), with Cluster 2 lineages currently known to comprise phyla with smaller genomes and more limited metabolic potential [20]. In contrast, the Cluster 1 DPANN comprise, in part, taxa that do not seem to have reduced genomes and are instead predicted to have host-independent lifestyles. For instance, certain members of the Diapherotrites have been proposed to have secondarily evolved greater independence, including the ability to synthesize ATP and various amino acids, nucleotides and cofactors, as a consequence of gene acquisitions from putative bacterial partners [52]. Furthermore, members of the Altiarchaeota do not seem to have reduced genomes; for instance, they are predicted to use the Wood-Ljungdahl pathway (WLP) for carbon fixation [53] and may even serve as hosts for other DPANN, such as members of the Huberarchaeota [54,55]. Altogether, it is therefore possible that the ancestor of the DPANN Cluster 1 lineage may not have been a genome-reduced symbiont. Thus, irrespective of the root placement, current data are consistent with a host-independent LACA. Prospective work integrating a better representation of recently discovered DPANN lineages with outgroup-free rooting methods and ancestral reconstructions will, however, be needed to further validate this hypothesis.

4. What have we learned about LUCA and timing of cellular evolution?

Phylogenetic approaches have helped to improve our resolution of the tree of life and the placement of DPANN and CPR, indicating that the ancestors of the Archaea and Bacteria, respectively, were complex organisms similar to extant representatives of these domains [19,22]. This has also allowed us to improve our knowledge of LUCA, the 'elusive' last universal common ancestor(s) of all cellular life [56–58]. Piecing together the characteristics of LUCA as well as its timing relative to Earth history is challenging owing to the sparsity of fossils and available biomarkers for prokaryotes. Consequently, inference of LUCA's metabolism and ecological role relies on phylogenetic and ancestral genome reconstruction approaches that examine gene histories in the context of archaeal and bacterial species phylogenies [14,59–63]. However, owing to the use of varying approaches and datasets to reconstruct the nature of LUCA, these analyses have resulted in distinct perspectives on what LUCA looked like [56]. For instance, some authors have proposed a communal, acellular and/or RNA-based LUCA, while others envisioned LUCA to have been more complex than extant prokaryotes ([14,57,59–61,64–66]; and recently reviewed in [57]). In contrast, some recent studies indicated that LUCA may have been a cellular organism similar to extant prokaryotic life [56,62,63], which would be in agreement with the universality of various cellular features across all organismal life. For example, by analysing genes that are shared by Archaea and Bacteria and broadly distributed amongst extant organisms, Weiss *et al.* proposed gene families that may trace back to LUCA [63]. They inferred LUCA to be a hyperthermophilic chemolithoautotrophic anaerobe that was hydrogen (H₂)-dependent, could fix carbon dioxide (CO₂) via the WLP [67,68] and conserve energy via substrate-level phosphorylation coupled to an ATP synthase [63]. Indeed, the WLP is often considered to represent an ancestral carbon fixation pathway as it allows for the generation of ATP during CO₂ fixation, making it one of the most energetically favourable carbon-fixation pathways [69,70].

However, these findings [63] have been challenged [71–74]. For instance, subsequent work indicated that several of the protein families inferred to have been present in LUCA on the basis of reciprocal domain monophyly may represent false positives, with domain monophyly resulting from undersampling of the protein families [71]. Furthermore, the subsequent phylogenetic inspection of key protein families in Weiss *et al.*'s reconstruction, such as the reverse gyrase and nitrogenase, has challenged the notion of a hyperthermophilic nitrogen-fixing LUCA [71,72,75]. While reciprocal domain monophyly of a protein family was used by Weiss *et al.* as a major criterion for its inference to LUCA [63], the violation of domain monophyly does not exclude the presence of a gene family in LUCA. However, interdomain transfers increase the difficulty of accurately inferring the LUCA proteome.

To overcome this challenge, we recently employed a probabilistic gene tree–species tree reconciliation approach that infers the history of gene duplication, transfer, loss and origination events along a species tree, thereby allowing us to determine the probability of a gene family having been present at any given node in a species tree [40,62]. Our results resolved LUCA's predicted genome size to approximately 2.75 Mb with 2657 genes, similar in complexity to extant Archaea and Bacteria. The inferred LUCA protein set shares consistencies with the inference of Weiss *et al.* [63], suggesting that LUCA may have been an anaerobic hyperthermophilic organism/cellular population that encoded the WLP and may have been able to grow autotrophically [62]. These results may localize LUCA's habitat to environments where H₂ and CO₂ were readily available [76]. However, it is important to emphasize that the WLP is not unique to autotrophs, with its presence being equally compatible with an organo-heterotrophic or mixotrophic metabolism [62,77,78]. Thus, while the improvement of our understanding of the tree of life and the use of novel computational approaches have recently allowed for more systematic reconstructions of the LUCA proteome, more work is needed to further inform deep cellular evolution and resolve contradictory hypotheses [57].

Linking cellular life's evolutionary timeline to past geological and biogeochemical data can provide complementary information on the evolution of microbial metabolism. Molecular clock techniques coupled with fossil and biomarker constraints can be used to timescale evolutionary events. However, current molecular clock methods are limited due to sparse prokaryotic fossil evidence necessary for clock calibration and challenges associated with adequately modelling rate variation through time. Alternative approaches are necessary to improve time estimates for prokaryotic evolution. Recently, we addressed some of these challenges [8] by applying an approach referred to as cross-bracing [10]. Genes resulting from ancient duplications and those inherited via endosymbiosis result in node-equivalency, a concept where multiple nodes within a phylogeny (single gene phylogenies or concatenated species phylogenies) correspond to the same evolutionary event. This temporal equivalence provides additional constraints to molecular dating as it allows for bracing corresponding nodes in the phylogeny to the same age. We made use of phylogenetic marker genes encoding protein machinery of which eukaryotes have inherited at least two paralogous forms from their respective archaeal and bacterial ancestors [79,80], such as ribosomal proteins and the ATP synthase. To improve our timing estimates, we combined this novel implementation of the cross-bracing approach with a relative time constraint [81], requiring that the node corresponding to the last plastid common ancestor be younger than the nodes corresponding to the last nuclear and mitochondrial eukaryotic common ancestors [8]. Using this approach, we estimated that LUCA lived between 4.52 and 4.32 Ga. Interestingly, our results also indicated that LBCA lived in an overlapping period from 4.49 to 4.05 Ga, whereas LACA lived from 3.95 to 3.37 Ga. Additional analyses are required to determine whether the estimated age of LACA is the result of poor taxonomic sampling along the archaeal stem or indicates a bottleneck in early archaeal evolution [8]. These findings are consistent with other recent studies using alternative dating methods that resolved the age of LUCA to approximately 4.52–4.48 Ga [5], 4.33–4.09 Ga [62] and a range of ages between 4.40–4.22 Ga [82] (table 1). Others have obtained younger ages for LUCA, e.g. a range of ages from 4.48 to 3.93 Ga [84] and 4.04–3.98 Ga [83] (table 1). LUCA's age is highly sensitive to, and clashes against, the maximum age calibration applied to the root node (table 1) (see also [5,42]). Furthermore, results may differ depending on whether the late heavy bombardment (LHB) (*ca.* 3.9 Ga) [85], the earliest evidence of liquid water (*ca.* 4.4 Ga) [86,87], the Moon-forming impact (4.52–4.51 Ga) [88] or the earliest evidence of life based on biomarkers and microfossils (4.1–3.5 Ga) [89–91] is applied as soft and/or hard age constraints [5,8,42,62,82–84] (table

Table 1. An overview of estimated age(s) and age ranges of LUCA (last universal common ancestor) and applied root calibrations that were reported in some recent studies mentioned in this opinion article.

study	inferred age of LUCA (Ga)	root node calibration (Ma)	description	references
Betts <i>et al.</i> , 2018	4.52–4.48	4520–3347	Any life on Earth would have emerged after sterilization post Moon-forming impact (max); age of the Strelley Pool formation (min).	[5]
Parsons <i>et al.</i> , 2021	4.04–3.98	3800 (conservative) 4100 (liberal)	Both root node calibrations are based on the earliest evidence of life on Earth; the conservative calibration represents the consensus age, whereas the liberal calibration represents the earliest age reported in the literature. Root ages inferred using the liberal calibration exceeded 4.5 Ga and were not considered in the analysis.	[83]
Mateos <i>et al.</i> , 2023	4.48 4.05 3.93	3800 (conservative) 4100 (liberal)	Both root node calibrations are based on the earliest evidence of life on Earth; the conservative calibration represents the consensus age, whereas the liberal calibration represents the earliest age reported in the literature. Different age inferences using the conservative calibration (reported here) correspond to different molecular clock models. Root ages inferred using the liberal calibration were between 5.3 and 6.0 Ga, exceeding the age of Earth.	[84]
Mahendrarajah <i>et al.</i> , 2023	4.52–4.32	4520–3347	Any life on Earth would have emerged after sterilization post Moon-forming impact (max); age of the Strelley Pool formation (min).	[8]
Boden <i>et al.</i> , 2024	4.40–4.22 4.40–4.34 4.40–4.29	4400–3500	Earliest evidence of liquid water on Earth (max); earliest evidence of life based on microfossils (min). Different age inferences (reported here) correspond to different molecular clock models.	[82]
Moody <i>et al.</i> , 2024	4.33–4.09	4520–3347	Any life on Earth would have emerged after sterilization post Moon-forming impact (max); age of the Strelley Pool formation (min).	[62]

1). It seems that increasing evidence questions the occurrence of a cataclysmic LHB in favour of a more drawn-out period of bombardments [85,92], which may have allowed life to emerge any time after the moon-forming impact and survive periods of bombardment in locally shielded environments. Further improvements in molecular dating approaches combined with new planetary, fossil and biomarker data will allow us to resolve more closely the time when LUCA thrived.

5. Conclusions

Resolving the phylogenetic position of DPANN archaea and CPR bacteria is crucial for our understanding of cellular evolution. Improved phylogenetic methods have revised the placement of the CPR bacteria as a derived sister-group of Chloroflexota, indicating genome streamlining from an ancestor that was not obligately host-dependent [19,37,41,43]. In contrast, while several different analyses have resolved the DPANN to be a sister group of all other Archaea [3,15,16,22,23], their monophyly and their exact root placement within the Archaea remain debated. Resolving the archaeal tree of life will thus be crucial to further improve our understanding of early cellular and archaeal evolution. For example, improved gene tree–species tree aware methods combined with a better sampling of extant archaeal biodiversity [40] would allow the rooting of an expanded archaeal phylogeny in the absence of an outgroup and predict the protein content of the archaeal ancestor. Gene family histories can furthermore help to elucidate the evolution of genome-streamlining (and potentially expansion) patterns across the tree of life. Integrating data from DPANN genomes into evolutionary reconstructions of key metabolic pathways such as the WLP [77], methanogenesis and anaerobic methane/alkane oxidation [93–96] will be key to further link biogeochemistry with cellular evolution. We expect progress in our understanding of early evolution through increased focus on multidisciplinary endeavours that brings together experts from a variety of fields including not only biogeochemistry, geology and chemistry but also synthetic biology and data sciences. For instance, the integration of novel approaches rooted in machine learning, such as structural predictions and phylogenetics, may help overcome some of the limitations of sequence-based reconstructions [40,97]. Finally, resurrecting and experimentally studying ancestral proteins will allow us to better understand their function under simulated early Earth conditions and test predictions [98–100].

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. This article has no additional data.

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. T.A.M.: conceptualization, data curation, writing—original draft, writing—review and editing; A.S.: conceptualization, writing—original draft, writing—review and editing.

Both authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. A.S. has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 947317, ASymbEL), from the Gordon and Betty Moore Foundation (GBMF9741) and the John Templeton Foundation (63451) (the opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation) to support T.A.M..

References

- Williams TA, Foster PG, Cox CJ, Embley TM. 2013 An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* **504**, 231–236. (doi:10.1038/nature12779)
- Eme L, Tamarit D. 2024 Microbial diversity and open questions about the deep tree of life. *Genome Biol. Evol.* **16**, evae053. (doi:10.1093/gbe/evae053)
- Hug LA *et al.* 2016 A new view of the tree of life. *Nat. Microbiol.* **1**, 16048. (doi:10.1038/nmicrobiol.2016.48)
- Spang A, Mahendrarajah TA, Offre P, Stairs CW. 2022 Evolving perspective on the origin and diversification of cellular life and the virosphere. *Genome Biol. Evol.* **14**, evac034. (doi:10.1093/gbe/evac034)
- Betts HC, Puttick MN, Clark JW, Williams TA, Donoghue PCJ, Pisani D. 2018 Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat. Ecol. Evol.* **2**, 1556–1562. (doi:10.1038/s41559-018-0644-x)
- Craig JM, Kumar S, Hedges SB. 2023 The origin of eukaryotes and rise in complexity were synchronous with the rise in oxygen. *Front. Bioinform.* **3**, 1233281. (doi:10.3389/fbinf.2023.1233281)
- Eme L, Sharpe SC, Brown MW, Roger AJ. 2014 On the age of eukaryotes: evaluating evidence from fossils and molecular clocks. *Cold Spring Harb. Perspect. Biol.* **6**, 1–16. (doi:10.1101/cshperspect.a016139)
- Mahendrarajah TA *et al.* 2023 ATP synthase evolution on a cross-braced dated tree of life. *Nat. Commun.* **14**, 7456. (doi:10.1038/s41467-023-42924-w)
- Parfrey LW, Lahr DJG, Knoll AH, Katz LA. 2011 Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc. Natl Acad. Sci. USA* **108**, 13624–13629. (doi:10.1073/pnas.1110633108)
- Shih PM, Matzke NJ. 2013 Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins. *Proc. Natl Acad. Sci. USA* **110**, 12355–12360. (doi:10.1073/pnas.1305813110)
- Strassert JFH, Irisarri I, Williams TA, Burki F. 2021 A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. *Nat. Commun.* **12**, 1879. (doi:10.1038/s41467-021-22044-z)
- Delaye L, Becerra A. 2012 Cenacestator, the last universal common ancestor. *Evolution* **5**, 382–388. (doi:10.1007/s12052-012-0444-8)
- Forterre P, Philippe H. 1999 The last universal common ancestor (LUCA), simple or complex? *Biol. Bull.* **196**, 373–375. (doi:10.2307/1542973)
- Woese CR, Fox GE. 1977 The concept of cellular evolution. *J. Mol. Evol.* **10**, 1–6. (doi:10.1007/BF01796132)
- Castelle CJ *et al.* 2015 Genomic expansion of domain Archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* **25**, 690–701. (doi:10.1016/j.cub.2015.01.014)
- Rinke C *et al.* 2013 Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437. (doi:10.1038/nature12352)
- Brown CT *et al.* 2015 Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208–211. (doi:10.1038/nature14486)
- Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017 Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* **2**, 1533–1542. (doi:10.1038/s41564-017-0012-7)
- Coleman GA, Davin AA, Mahendrarajah TA, Szánthó LL, Spang A, Hugenholtz P, Szöllösi GJ, Williams TA. 2021 A rooted phylogeny resolves early bacterial evolution. *Science* **372**, eabe0511. (doi:10.1126/science.abe0511)
- Dombrowski N, Williams TA, Sun J, Woodcroft BJ, Lee JH, Minh BQ, Rinke C, Spang A. 2020 Undinarchaeota illuminate DPANN phylogeny and the impact of gene transfer on archaeal evolution. *Nat. Commun.* **11**, 3939. (doi:10.1038/s41467-020-17408-w)
- Castelle CJ, Banfield JF. 2018 Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell* **172**, 1181–1197. (doi:10.1016/j.cell.2018.02.016)
- Williams TA, Szöllösi GJ, Spang A, Foster PG, Heaps SE, Boussau B, Ettema TJG, Embley TM. 2017 Integrative modeling of gene and genome evolution roots the archaeal tree of life. *Proc. Natl Acad. Sci. USA* **114**, E4602–E4611. (doi:10.1073/pnas.1618463114)
- Zhu Q *et al.* 2019 Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. *Nat. Commun.* **10**, 5477. (doi:10.1038/s41467-019-13443-4)
- Castelle CJ, Brown CT, Anantharaman K, Probst AJ, Huang RH, Banfield JF. 2018 Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. *Nat. Rev. Microbiol.* **16**, 629–645. (doi:10.1038/s41579-018-0076-2)
- Lineweaver CH. 2020 What do the DPANN Archaea and the CPR bacteria tell us about the last universal common ancestors? In *Extremophiles as astrobiological models* (eds J Seckbach, H Stan-Lotter), pp. 359–367. Hoboken, NJ: Wiley-Scrivener. (doi:10.1002/9781119593096)
- Bor B, McLean JS, Foster KR, Cen L, To TT, Serrato-Guillen A, Dewhurst FE, Shi W, He X. 2018 Rapid evolution of decreased host susceptibility drives a stable relationship between ultrasmall parasite TM7x and its bacterial host. *Proc. Natl Acad. Sci. USA* **115**, 12277–12282. (doi:10.1073/pnas.1810625115)
- Ding S, Hamm JN, Bale N, Sinnighe Damsté JS, Spang A. 2023 Selective lipid recruitment by an archaeal DPANN symbiont from its host. *Nat. Commun.* **15**, 3405. (doi:10.1038/s41467-024-47750-2)
- Hamm JN *et al.* 2019 Unexpected host dependency of Antarctic Nanohaloarchaeota. *Proc. Natl Acad. Sci. USA* **116**, 14661–14670. (doi:10.1073/pnas.1905179116)
- He X *et al.* 2015 Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. *Proc. Natl Acad. Sci. USA* **112**, 244–249. (doi:10.1073/pnas.1419038112)
- Huber H, Hohn MJ, Rachel R, Fuchs T, Wimmer VC, Stetter KO. 2002 A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**, 63–67. (doi:10.1038/417063a)

31. St. John E, Liu Y, Podar M, Stott MB, Meneghin J, Chen Z, Lagutin K, Mitchell K, Reysenbach AL. 2019 A new symbiotic nanoarchaeote (*Candidatus Nanoclepta minutus*) and its host (*Zestosphaera tikiterensis* gen. nov., sp. nov.) from a New Zealand hot spring. *Syst. Appl. Microbiol.* **42**, 94–106. (doi:10.1016/j.syapm.2018.08.005)
32. Bergsten J. 2005 A review of long-branch attraction. *Cladistics* **21**, 163–193. (doi:10.1111/j.1096-0031.2005.00059.x)
33. Philippe H, Zhou Y, Brinkmann H, Rodrigue N, Delsuc F. 2005 Heterotachy and long-branch attraction in phylogenetics. *BMC Evol. Biol.* **5**, 50. (doi:10.1186/1471-2148-5-50)
34. Aouad M, Taib N, Oudart A, Lecocq M, Gouy M, Brochier-Armanet C. 2018 Extreme halophilic archaea derive from two distinct methanogen Class II lineages. *Mol. Phylogenetics Evol.* **127**, 46–54. (doi:10.1016/j.ympev.2018.04.011)
35. Brochier-Armanet C, Forterre P, Gribaldo S. 2011 Phylogeny and evolution of the Archaea: one hundred genomes later. *Curr. Opin. Microbiol.* **14**, 274–281. (doi:10.1016/j.mib.2011.04.015)
36. Feng Y, Neri U, Gosselin S, Louyakis AS, Papke RT, Gophna U, Gogarten JP. 2021 The evolutionary origins of extreme halophilic archaeal lineages. *Genome Biol. Evol.* **13**, evab166. (doi:10.1093/gbe/evab166)
37. Méheust R, Burstein D, Castelle CJ, Banfield JF. 2019 The distinction of CPR bacteria from other bacteria based on protein family content. *Nat. Commun.* **10**, 4173. (doi:10.1038/s41467-019-12171-z)
38. Raymann K, Forterre P, Brochier-Armanet C, Gribaldo S. 2014 Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in Archaea. *Genome Biol. Evol.* **6**, 192–212. (doi:10.1093/gbe/evu004)
39. Kellner S, Spang A, Offre P, Szöllősi GJ, Petitjean C, Williams TA. 2018 Genome size evolution in the Archaea. *Emerg. Top. Life Sci.* **2**, 595–605. (doi:10.1042/etls20180021)
40. Williams TA *et al.* 2024 Phylogenetic reconciliation: making the most of genomes to understand microbial ecology and evolution. *ISME J.* **18**, e129. (doi:10.1093/ismej/wrae129)
41. Martínez-Gutiérrez CA, Aylward FO. 2021 Phylogenetic signal, congruence, and uncertainty across bacteria and Archaea. *Mol. Biol. Evol.* **38**, 5514–5527. (doi:10.1093/molbev/msab254)
42. Moody ERR, Mahendrarajah TA, Dombrowski N, Clark JW, Petitjean C, Offre P, Szöllősi GJ, Spang A, Williams TA. 2022 An estimate of the deepest branches of the tree of life from ancient vertically evolving genes. *eLife* **11**, 66695. (doi:10.7554/eLife.66695)
43. Taib N, Megrian D, Witwinowski J, Adam P, Poppleton D, Borrel G, Beloin C, Gribaldo S. 2020 Genome-wide analysis of the Firmicutes illuminates the diderm/monoderm transition. *Nat. Ecol. Evol.* **4**, 1661–1672. (doi:10.1038/s41559-020-01299-7)
44. Szöllősi GJ, Rosikiewicz W, Boussau B, Tannier E, Daubin V. 2013 Efficient exploration of the space of reconciled gene trees. *Syst. Biol.* **62**, 901–912. (doi:10.1093/sysbio/syt054)
45. Beaud Benyahia B, Taib N, Beloin C, Gribaldo S. 2024 Terrabacteria: redefining bacterial envelope diversity, biogenesis and evolution. *Nat. Rev. Microbiol.* **23**, 41–56. (doi:10.1038/s41579-024-01088-0)
46. Rinke C, Chuvochina M, Mussig AJ, Chaumeil PA, Davin AA, Waite DW, Whitman WB, Parks DH, Hugenholtz P. 2021 A standardized archaeal taxonomy for the Genome Taxonomy Database. *Nat. Microbiol.* **6**, 946–959. (doi:10.1038/s41564-021-00918-8)
47. Aouad M, Flandrois JP, Jauffrit F, Gouy M, Gribaldo S, Brochier-Armanet C. 2022 A divide-and-conquer phylogenomic approach based on character supermatrices resolves early steps in the evolution of the Archaea. *BMC Ecol. Evol.* **22**, 1. (doi:10.1186/s12862-021-01952-0)
48. Narasingarao P, Podell S, Ugalde JA, Brochier-Armanet C, Emerson JB, Brocks JJ, Heidelberg KB, Banfield JF, Allen EE. 2012 *De novo* metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *ISME J.* **6**, 81–93. (doi:10.1038/ismej.2011.78)
49. Petitjean C, Deschamps P, López-García P, Moreira D. 2014 Rooting the domain Archaea by phylogenomic analysis supports the foundation of the new kingdom Proteoarchaeota. *Genome Biol. Evol.* **7**, 191–204. (doi:10.1093/gbe/evu274)
50. Wang B *et al.* 2019 Expansion of *Thaumarchaeota* habitat range is correlated with horizontal transfer of ATPase operons. *ISME J.* **13**, 3067–3079. (doi:10.1038/s41396-019-0493-x)
51. Baker BA *et al.* 2024 Expanded phylogeny of extremely halophilic archaea shows multiple independent adaptations to hypersaline environments. *Nat. Microbiol.* **9**, 964–975. (doi:10.1038/s41564-024-01647-4)
52. Youssef NH, Rinke C, Stepanauskas R, Farag I, Woyke T, Elshahed MS. 2015 Insights into the metabolism, lifestyle and putative evolutionary history of the novel archaeal phylum 'Diapherotrites'. *ISME J.* **9**, 447–460. (doi:10.1038/ismej.2014.141)
53. Probst AJ *et al.* 2014 Biology of a widespread uncultivated archaeon that contributes to carbon fixation in the subsurface. *Nat. Commun.* **5**, 5497. (doi:10.1038/ncomms6497)
54. Probst AJ *et al.* 2018 Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nat. Microbiol.* **3**, 328–336. (doi:10.1038/s41564-017-0098-y)
55. Schwank K, Bornemann TLV, Dombrowski N, Spang A, Banfield JF, Probst AJ. 2019 An archaeal symbiont–host association from the deep terrestrial subsurface. *ISME J.* **13**, 2135–2139. (doi:10.1038/s41396-019-0421-0)
56. Crapitto AJ, Campbell A, Harris AJ, Goldman AD. 2022 A consensus view of the proteome of the last universal common ancestor. *Ecol. Evol.* **12**, e8930. (doi:10.1002/ece3.8930)
57. Forterre P. 2024 The last universal common ancestor of ribosome-encoding organisms: portrait of LUCA. *J. Mol. Evol.* **92**, 550–583. (doi:10.1007/s00239-024-10186-9)
58. Weiss MC, Preiner M, Xavier JC, Zimorski V, Martin WF. 2018 The last universal common ancestor between ancient Earth chemistry and the onset of genetics. *PLoS Genet.* **14**, e1007518. (doi:10.1371/journal.pgen.1007518)
59. Doolittle WF. 2000 The nature of the universal ancestor and the evolution of the proteome. *Curr. Opin. Struct. Biol.* **10**, 355–358. (doi:10.1016/S0959-440X(00)00096-8)
60. Doolittle WF, Brown JR. 1994 Tempo, mode, the progenote, and the universal root. *Proc. Natl Acad. Sci. USA* **91**, 6721–6728. (doi:10.1073/pnas.91.15.6721)
61. Forterre P. 2002 The origin of DNA genomes and DNA replication proteins. *Curr. Opin. Microbiol.* **5**, 525–532. (doi:10.1016/S1369-5274(02)00360-0)
62. Moody ERR *et al.* 2024 The nature of the last universal common ancestor and its impact on the early Earth system. *Nat. Ecol. Evol.* **8**, 1–13. (doi:10.1038/s41559-024-02461-1)
63. Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, Martin WF. 2016 The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* **1**, 16116. (doi:10.1038/nmicrobiol.2016.116)
64. Line MA. 2002 The enigma of the origin of life and its timing. *Microbiology* **148**, 21–27. (doi:10.1099/00221287-148-1-21)
65. Penny D, Poole A. 1999 The nature of the last universal common ancestor. *Curr. Opin. Genet. Dev.* **9**, 672–677. (doi:10.1016/S0959-437X(99)00020-9)
66. Woese C. 1998 The universal ancestor. *Proc. Natl Acad. Sci. USA* **95**, 6854–6859. (doi:10.1073/pnas.95.12.6854)
67. Ragsdale SW, Pierce E. 2008 Acetogenesis and the Wood–Ljungdahl pathway of CO₂ fixation. *Biochim. Et Biophys. Acta* **1784**, 1873–1898. (doi:10.1016/j.bbapap.2008.08.012)
68. Schuchmann K, Müller V. 2014 Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat. Rev. Microbiol.* **12**, 809–821. (doi:10.1038/nrmicro3365)
69. Berg IA. 2011 Ecological aspects of the distribution of different autotrophic CO₂ fixation pathways. *Appl. Environ. Microbiol.* **77**, 1925–1936. (doi:10.1128/aem.02473-10)
70. Fuchs G. 2011 Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? *Annu. Rev. Microbiol.* **65**, 631–658. (doi:10.1146/annurev-micro-090110-102801)

71. Berkemer SJ, McGlynn SE. 2020 A new analysis of Archaea–Bacteria domain separation: variable phylogenetic distance and the tempo of early evolution. *Mol. Biol. Evol.* **37**, 2332–2340. (doi:10.1093/molbev/msaa089)
72. Catchpole RJ, Forterre P. 2019 The evolution of reverse gyrase suggests a nonhyperthermophilic last universal common ancestor. *Mol. Biol. Evol.* **36**, 2737–2747. (doi:10.1093/molbev/msz180)
73. Forterre P. 2022 Carl Woese: still ahead of our time. *mLife* **1**, 359–367. (doi:10.1002/mlf2.12049)
74. Gogarten JP, Deamer D. 2016 Is LUCA a thermophilic progenote? [Review of 'Is LUCA a thermophilic progenote?']. *Nat. Microbiol.* **1**, 16229. (doi:10.1038/nmicrobiol.2016.229)
75. Boyd ES, Anbar AD, Miller S, Hamilton TL, Lavin M, Peters JW. 2011 A late methanogen origin for molybdenum-dependent nitrogenase. *Geobiology* **9**, 221–232. (doi:10.1111/j.1472-4669.2011.00278.x)
76. Martin W, Baross J, Kelley D, Russell MJ. 2008 Hydrothermal vents and the origin of life. *Nat. Rev. Microbiol.* **6**, 805–814. (doi:10.1038/nrmicro1991)
77. Adam PS, Borrel G, Gribaldo S. 2018 Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proc. Natl Acad. Sci. USA* **115**, E1166–E1173. (doi:10.1073/pnas.1716667115)
78. Schuchmann K, Müller V. 2016 Energetics and application of heterotrophy in acetogenic bacteria. *Appl. Environ. Microbiol.* **82**, 4056–4069. (doi:10.1128/aem.00882-16)
79. Donoghue PCJ, Kay C, Spang A, Szöllösi G, Nenarokova A, Moody ERR, Pisani D, Williams TA. 2023 Defining eukaryotes to dissect eukaryogenesis. *Curr. Biol.* **33**, R919–R929. (doi:10.1016/j.cub.2023.07.048)
80. Vosseberg J, van Hooff JJE, Marcet-Houben M, van Vlimmeren A, van Wijk LM, Gabaldón T, Snel B. 2021 Timing the origin of eukaryotic cellular complexity with ancient duplications. *Nat. Ecol. Evol.* **5**, 92–100. (doi:10.1038/s41559-020-01320-z)
81. Szöllösi GJ, Höhna S, Williams TA, Schrempf D, Daubin V, Boussau B. 2022 Relative time constraints improve molecular dating. *Syst. Biol.* **71**, 797–809. (doi:10.1093/sysbio/syab084)
82. Boden JS, Zhong J, Anderson RE, Stüeken EE. 2024 Timing the evolution of phosphorus-cycling enzymes through geological time using phylogenomics. *Nat. Commun.* **15**, 3703. (doi:10.1038/s41467-024-47914-0)
83. Parsons C, Stüeken EE, Rosen CJ, Mateos K, Anderson RE. 2021 Radiation of nitrogen-metabolizing enzymes across the tree of life tracks environmental transitions in Earth history. *Geobiology* **19**, 18–34. (doi:10.1111/gbi.12419)
84. Mateos K, Chappell G, Klos A, Le B, Boden J, Stüeken E, Anderson R. 2023 The evolution and spread of sulfur cycling enzymes reflect the redox state of the early Earth. *Sci. Adv.* **9**, eade4847. (doi:10.1126/sciadv.ade4847)
85. Bottke WF, Norman MD. 2017 The Late Heavy Bombardment. *Annu. Rev. Earth Planet. Sci.* **45**, 619–647. (doi:10.1146/annurev-earth-063016-020131)
86. Mojzsis SJ, Harrison TM, Pidgeon RT. 2001 Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago. *Nature* **409**, 178–181. (doi:10.1038/35051557)
87. Wilde SA, Valley JW, Peck WH, Graham CM. 2001 Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. *Nature* **409**, 175–178. (doi:10.1038/35051550)
88. Hanan BB, Tilton GR. 1987 60025: relict of primitive lunar crust? *Earth Planet. Sci. Lett.* **84**, 15–21. (doi:10.1016/0012-821x(87)90171-3)
89. Bell EA, Boehnke P, Harrison TM, Mao WL. 2015 Potentially biogenic carbon preserved in a 4.1 billion-year-old zircon. *Proc. Natl Acad. Sci. USA* **112**, 14518–14521. (doi:10.1073/pnas.1517557112)
90. Mojzsis SJ, Arrhenius G, McKeegan KD, Harrison TM, Nutman AP, Friend CR. 1996 Evidence for life on Earth before 3,800 million years ago. *Nature* **384**, 55–59. (doi:10.1038/384055a0)
91. Nutman AP, Bennett VC, Friend CRL, Van Kranendonk MJ, Chivas AR. 2016 Rapid emergence of life shown by discovery of 3,700-million-year-old microbial structures. *Nature* **537**, 535–538. (doi:10.1038/nature19355)
92. Mann A. 2018 Bashing holes in the tale of Earth's troubled youth. *Nature* **553**, 393–395. (doi:10.1038/d41586-018-01074-6)
93. Borrel G *et al.* 2019 Wide diversity of methane and short-chain alkane metabolisms in uncultured archaea. *Nat. Microbiol.* **4**, 603–613. (doi:10.1038/s41564-019-0363-3)
94. Evans PN, Boyd JA, Leu AO, Woodcroft BJ, Parks DH, Hugenholtz P, Tyson GW. 2019 An evolving view of methane metabolism in the Archaea. *Nat. Rev. Microbiol.* **17**, 219–232. (doi:10.1038/s41579-018-0136-7)
95. Musat F, Kjeldsen KU, Rotaru AE, Chen SC, Musat N. 2024 Archaea oxidizing alkanes through alkyl-coenzyme M reductases. *Curr. Opin. Microbiol.* **79**, 102486. (doi:10.1016/j.mib.2024.102486)
96. Wegener G, Krukenberg V, Riedel D, Tegetmeyer HE, Boetius A. 2015 Intercellular wiring enables electron transfer between methanotrophic archaea and bacteria. *Nature* **526**, 587–590. (doi:10.1038/nature15733)
97. Garg SG, Hochberg GKA. 2025 A General Substitution Matrix for Structural Phylogenetics. *Mol. Biol. Evol.* **42**, msaf124. (doi:10.1093/molbev/msaf124)
98. García AK, Kaçar B. 2019 How to resurrect ancestral proteins as proxies for ancient biogeochemistry. *Free Radic. Biol. Med.* **140**, 260–269. (doi:10.1016/j.freeradbiomed.2019.03.033)
99. Hochberg GKA, Thornton JW. 2017 Reconstructing ancient proteins to understand the causes of structure and function. *Annu. Rev. Biophys.* **46**, 247–269. (doi:10.1146/annurev-biophys-070816-033631)
100. Schulz L *et al.* 2022 Evolution of increased complexity and specificity at the dawn of form I Rubiscos. *Science* **378**, 155–160. (doi:10.1126/science.abq1416)