

Phylogenomic Analyses Reveal that *Panguiarchaeum* Is a Clade of Genome-Reduced Asgard Archaea Within the Njordarchaeia

Wen-Cong Huang ^{1,2}, Maraike Probst ³, Zheng-Shuang Hua ⁴, Lénárd L. Szánthó ⁵, Gergely J. Szöllősi ^{5,6}, Thijs J.G. Ettema ⁷, Christian Rinke ³, Tom A. Williams ^{8,*}, Anja Spang ^{1,2,*}

¹Department of Marine Microbiology and Biogeochemistry, NIOZ, Royal Netherlands Institute for Sea Research, Den Burg 1790 AB, The Netherlands

²Department of Evolutionary & Population Biology, Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, Amsterdam 1090 GE, The Netherlands

³Faculty of Biology, Department of Microbiology, University of Innsbruck, Innsbruck 6020, Austria

⁴Chinese Academy of Sciences Key Laboratory of Urban Pollutant Conversion, Department of Environmental Science and Engineering, University of Science and Technology of China, Hefei 230026, China

⁵Model-Based Evolutionary Genomics Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa 904-0495, Japan

⁶Institute of Evolution, HUN-REN Center for Ecological Research, Budapest 1121, Hungary

⁷Laboratory of Microbiology, Wageningen University & Research, Wageningen, The Netherlands

⁸Department of Life Sciences, University of Bath, Bath BA2 7AX, UK

*Corresponding authors: E-mails: anja.spang@nioz.nl; tw2316@bath.ac.uk.

Associate editor: Purificación Lopez-García

Abstract

The Asgard archaea are a diverse archaeal phylum important for our understanding of cellular evolution because they include the lineage that gave rise to eukaryotes. Recent phylogenomic work has focused on characterizing the diversity of Asgard archaea in an effort to identify the closest extant relatives of eukaryotes. However, resolving archaeal phylogeny is challenging, and the positions of 2 recently described lineages—Njordarchaeales and Panguiarchaeales—are uncertain, in ways that directly bear on hypotheses of early evolution. In initial phylogenetic analyses, these lineages branched either with Asgards or with the distantly related Korarchaeota, and it has been suggested that their genomes may be affected by metagenomic contamination. Resolving this debate is important because these clades include genome-reduced lineages that may help inform our understanding of the evolution of symbiosis within Asgard archaea. Here, we performed phylogenetic analyses revealing that the Njordarchaeales and Panguiarchaeales constitute the new class Njordarchaeia within Asgard archaea. We found no evidence of metagenomic contamination affecting phylogenetic analyses. Njordarchaeia exhibit hallmarks of adaptations to (hyper-)thermophilic lifestyles, including biased sequence compositions that can induce phylogenetic artifacts unless adequately modeled. *Panguiarchaeum* is metabolically distinct from its relatives, with reduced metabolic potential and various auxotrophies. Phylogenetic reconciliation recovers a complex common ancestor of Asgard archaea that encoded the Wood–Ljungdahl pathway. The subsequent loss of this pathway during the reductive evolution of *Panguiarchaeum* may have been associated with the switch to a symbiotic lifestyle, potentially based on H₂-syntrophy. Thus, *Panguiarchaeum* may contain the first obligate symbionts within Asgard archaea besides the lineage leading to eukaryotes.

Keywords: Asgard archaea, phylogenetics, symbiosis, evolution, *Panguiarchaeum*

Introduction

Asgard archaea are a diverse phylum from which the archaeal ancestor of eukaryotes emerged (Spang et al. 2015; Zaremba-Niedzwiedzka et al. 2017; Liu et al. 2021; Zhang et al. 2025). The genomes of Asgard archaea encode eukaryotic signature proteins (ESPs) (Spang et al. 2015; Zaremba-Niedzwiedzka et al. 2017; Liu et al. 2021; Wu et al. 2022; Valentin-Alvarado et al. 2024; Vosseberg et al. 2024), which are homologs of eukaryotic proteins that underlie their cellular complexity. Experimental evidence suggests that at least some ESFs of Asgard archaea have functions similar to those in eukaryotes (Akil and Robinson 2018; Akil et al. 2020; Lu et al. 2020; Hatano et al. 2022; Hurtig et al. 2023). Genomic

predictions and the cultivation of the first representatives of the Asgard archaea suggest that some members form symbiotic (i.e. syntrophic), interactions with other cells (Spang et al. 2019; Imachi et al. 2020). Indeed, eukaryotes are thought to have evolved from one such interaction (Martin and Müller 1998; Moreira and Lopez-García 1998; Lane and Martin 2010; Guy et al. 2014; Koonin and Yutin 2014; Poole and Gribaldo 2014; López-García and Moreira 2015; Ettema 2016; Roger et al. 2017; Vosseberg et al. 2024). The phylogeny and genomic features of Asgard archaea and related groups therefore provide an important framework for understanding the origin of eukaryotic cells and the evolution of prokaryotic symbioses (Spang et al. 2015; Zaremba-Niedzwiedzka et al. 2017;

Received: May 22, 2025. Revised: July 10, 2025. Accepted: July 21, 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Williams et al. 2020; Liu et al. 2021; Zhang et al. 2025). A major current challenge lies in resolving the evolutionary relationships among Asgard archaea and their relatives, including Thermoproteota, which were originally referred to as TACK archaea (Guy and Ettema 2011) and comprise the former Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota (Elkins et al. 2008). Early analyses supported a sister-relationship between all Thermoproteota (including the Koarchaeota) and Asgard archaea (Spang et al. 2015; Zaremba-Niedzwiedzka et al. 2017; Baker et al. 2020). However, more recent phylogenetic analyses taking advantage of improved genomic sampling have instead placed the Korarchaeota at the base of a monophyletic clade comprising Thermoproteota and Asgard archaea (Tahon et al. 2023).

Two additional lineages have recently been described that appear to be part of a broader Asgard/Thermoproteota/Korarchaeota clade, although their phylogenetic affinities remain uncertain: the Njordarchaeales (Liu and Li 2022; Xie et al. 2022; Eme et al. 2023) and the Pangiarchaeales (Qu et al. 2023). Njordarchaeales have been recovered either as sister to Korarchaeota (Liu and Li 2022)—with whom they share a thermophilic lifestyle—or, intriguingly, within Asgard archaea, as either sister to the eukaryotic nuclear lineage (Xie et al. 2022) or an order within the Heimdallarchaeia (Eme et al. 2023; Valentin-Alvarado et al. 2024). Placement outside the Asgard archaea is observed when the hyperthermophilic Korarchaeota are included (Liu and Li 2022; Eme et al. 2023; Qu et al. 2023), while their inclusion within Asgard archaea is more common when these taxa are excluded (Eme et al. 2023; Valentin-Alvarado et al. 2024), depending on the dataset and phylogenetic method used. Similarly, the recently reported Pangiarchaeales, composed of 10 metagenome-assembled genomes (MAGs) derived from terrestrial geothermal metagenomes, all belonging to the same species, were suggested to branch with Korarchaeota based on four distinct marker sets using a site-homogeneous substitution model (Qu et al. 2023). However, a conflicting placement with Asgard archaea was observed using 53 marker genes from GTDB (release 207v2). Pangiarchaeales were proposed to represent anaerobic amino acid fermenters with a symbiotic lifestyle (Qu et al. 2023) because they lack certain genes for archaeal lipid, purine, and amino acid biosynthesis pathways reminiscent of DPANN archaea (Qu et al. 2023). However, the authors of that study could not identify direct evidence for recent reductive genome evolution since key genes for DNA mismatch repair mechanisms and homologous recombination were present in the investigated MAGs, which furthermore did not significantly differ in the genome size from other Korarchaeota with which they were proposed to be affiliated (Qu et al. 2023). Reduced metabolic capabilities are hallmarks of the DPANN archaea, which represent diverse phylum-level lineages composed of putative symbionts (Rinke et al. 2013; Castelle et al. 2018), but are otherwise uncommon among archaea. Thus, resolving the evolutionary relationships within the Asgard/Thermoproteota/Korarchaeota clade, and the positions of Njordarchaeales and Pangiarchaeales within them, is key to understanding the evolutionary trajectories of genome reduction and symbiotic associations in Archaea as well as for placing the origin of eukaryotic cells in the broader context of archaeal evolution. An additional challenge in this regard is the specter of metagenomic contamination, with a recent study (Zhang et al. 2025) suggesting that at least some Njordarchaeales MAGs might be chimeras, containing a mixture of sequences derived from both Asgard archaea and Korarchaeota. If correct, this

would provide an alternative explanation for the difficulty of placing Njordarchaeales in phylogenies, because combining marker genes with different evolutionary histories would undermine a core assumption of concatenated phylogenetic analyses.

Here, we carefully assembled an archaeal taxa set, inspected the MAGs quality of key representatives of the Pangi- and Njordarchaeales, and applied a variety of comparative genomics, phylogenomics, and network approaches to determine the phylogenetic placement of Njordarchaeales and Pangiarchaeales within the archaeal tree of life and to assess the genomic potential and proposed symbiotic lifestyle of the latter. Our analyses reveal that the previously described Njordarchaeales and Pangiarchaeales orders are order- or family-level lineages which together form the new class Njordarchaeia within the Asgard archaea rather than the Korarchaeota. We found no evidence for metagenomic contamination affecting marker gene sequences; instead, our analyses show that the difficulty in placing Njordarchaeia in the archaeal tree is due to pervasive site-wise and branch-wise compositional heterogeneity in the sequence data that are challenging to model adequately, particularly with simpler substitution models. While all Njordarchaeia encode a similar set of ESPs, Pangiarchaeaceae MAGs show indications of reductive genome evolution and auxotrophies, suggesting that they may be dependent on partner organisms or other community members to supplement their growth. Indeed, our co-occurrence analyses reveal a higher-than-expected frequency of association between *Pangiarchaeum* and members of the Thermoproteota. Overall, this study shows that MAG contamination is not the cause of the conflicting phylogenetic signal observed for members of the Njordarchaeia and instead highlights the need to use phylogenetic models that provide an adequate fit to the data when inferring the archaeal phylogeny. It further provides the first example of a putative symbiotic lineage within the Asgard archaea previously assigned to Korarchaeota that has experienced gene loss rather than genome expansion.

Results

Pangiarchaeum Genomes Form a Sister Group to Njordarchaeaceae and Branch Within Asgard Archaea

Assembling a Set of Vertically Evolving Archaeal Marker Genes

We used a range of phylogenetic analyses to investigate the evolutionary relationships among Pangiarchaeales and Njordarchaeales (13 MAGs, see [supplementary data S1, Supplementary Material](#) online), with Korarchaeota (Tahon et al. 2023) and Asgard archaea, officially referred to as Asgardarchaeota based on SeqCode (Tamarit et al. 2024) and as Promethearchaeota (Imachi et al. 2024) based on International Code of Nomenclature of Prokaryotes. Our starting point was the set of vertically evolved marker genes that have been used to investigate archaeal phylogeny and Njordarchaeia placement in several previous analyses, including ribosomal proteins as well as genes with alternative functions (Dombrowski et al. 2020; Williams et al. 2020; Rinke et al. 2021; Moody et al. 2022), totaling 120 distinct homologous Archaeal Clusters of Orthologous Genes (arCOG) gene families. We then performed a series of analyses to select the most reliable marker genes for concatenation, filtering out families that have been affected by gene replacement transfers from Bacteria (9 families), that were found in less than 60% of taxa (4 markers), or which have a complex history of gene duplication, transfer and loss within Archaea, and so are

unreliable markers for inferring vertical evolution (11 families, including some subunits of DNA topoisomerase—see [supplementary information S1.1, Supplementary Material](#) online; [supplementary data S2 to S4, Supplementary Material](#) online; [supplementary figs. S1 to S8, Supplementary Material](#) online). We ranked the remaining markers based on split score quantifying the degree to which they recovered accepted monophyletic archaeal clades ([Dombrowski et al. 2020](#); [Liu and Li 2022](#); [Moody et al. 2022](#) [see Methods: Ranking analysis; [supplementary data S4, Supplementary Material](#) online]) and selected the top 50% of these markers corresponding to 43 proteins ([supplementary data S4, Supplementary Material](#) online) for phylogenetic analyses. The overlap between this marker set and those used in previous studies ([Dombrowski et al. 2020](#); [Williams et al. 2020](#); [Liu et al. 2021](#); [Moody et al. 2022](#); [Eme et al. 2023](#); [Qu et al. 2023](#); [Valentin-Alvarado et al. 2024](#); [Zhang et al. 2025](#)), for both the initial and final curated set of markers, is summarized in [supplementary data S2 to S4, Supplementary Material](#) online.

Phylogenetic Analyses that Model Site- and Branch-Compositional Heterogeneity Place *Panguiarchaeaceae* and *Njordarchaeaceae* Within Asgard Archaea

Previous analyses have suggested that *Panguiarchaeales* and/or *Njordarchaeales* may branch either with *Korarchaeota* ([Liu and Li 2022](#); [Qu et al. 2023](#); [Zhang et al. 2025](#)) or Asgard archaea ([Eme et al. 2023](#); [Valentin-Alvarado et al. 2024](#)) ([Fig. 1a](#)). To distinguish between these hypotheses (i.e. *Korarchaeota* vs. Asgard archaea topologies ([Fig. 1a](#))) and assess the taxonomic ranks of these clades, we performed phylogenetic analyses on the 43-gene concatenation using distinct taxon sets as well as relative evolutionary distance (RED) scaling to normalize the taxonomic ranks ([Rinke et al. 2021](#); [Tamarit et al. 2024](#)) (see Methods and [supplementary fig. S1, Supplementary Material](#) online for method overview). All analyses provided strong support for the monophyly of the previously proposed *Njordarchaeales* and *Panguiarchaeales* clades, together forming a putative class-level taxon, i.e. the *Njordarchaeia*. This contrasts with previous findings which suggested that *Panguiarchaeae* are part of the *Korarchaeota* ([Qu et al. 2023](#)) while *Njordarchaeales* were assigned to the Asgard archaea ([Eme et al. 2023](#); [Valentin-Alvarado et al. 2024](#)). Specifically, our results reveal that the *Njordarchaeia* comprises the *Njordarchaeales* order with 2 family-level lineages, the *Panguiarchaeaceae* and *Njordarchaeaceae* ([Fig. 1b and c](#); [supplementary fig. S9, Supplementary Material](#) online, see also [supplementary information S1.2, Supplementary Material](#) online). Furthermore, analyses using the best-fitting substitution models robustly supported the placement of the *Njordarchaeia* within the Asgard archaea ([Fig. 1b and c](#); [supplementary fig. S1, Supplementary Material](#) online). However, we did observe the *Korarchaeota* topology, i.e. the placement of all *Njordarchaeia* with *Korarchaeota*, in several analyses using simpler substitution and empirical distribution mixture (EDM) models with few components (LG + G + F, LG + EDM0004LCLR + G + F, LG + EDM0008LCLR + G + F; [supplementary fig. S1, Supplementary Material](#) online and [supplementary fig. S10a–c, Supplementary Material](#) online), consistent with the hypothesis that this inferred relationship might be a phylogenetic artifact ([Eme et al. 2023](#)).

Current phylogenetic methods are not consummate, and even the best models are mere approximations of the

evolutionary processes that gave rise to the observed data. However, previous work has identified two aspects of the evolutionary process that are not well-captured by simple models, which have had a substantial impact on phylogenetic inference ([Williams et al. 2020](#); [Muñoz-Gómez et al. 2022](#)). These are variations in amino acid preferences across the alignment sites (due to varying protein biochemical constraints) and across the tree branches (due to varying environmental pressures and mutational biases).

Modeling Site-Wise Compositional Heterogeneity

Site-wise compositional variation results from selective constraints on protein function: typically, only a small subset of the 20 possible amino acids are permissible at a given position in the protein due to the requirement for particular types of side chains to maintain protein structure and function ([Lartillot and Philippe 2004](#)). While being a pervasive feature of molecular sequence data, this site-wise variation is not accounted for in simple substitution models such as LG + G, which assume that all sites have the same equilibrium amino acid frequencies ([Kapli et al. 2021](#); [Williams et al. 2021](#)). Thus, simple models under-estimate the probability of convergent evolution to the same amino acid in independent, distantly related lineages. This means they can sometimes misinterpret instances of molecular convergence as evidence that these taxa are closely related, an artifact that is expected to be strongest at the most selectively constrained positions. To evaluate whether under-estimation of convergent evolution might drive the support for the cladehood of *Njordarchaeia* and *Korarchaeota* in analyses with simpler models, we grouped amino acid sites by compositional constraint (see Methods) and investigated support for the two trees from each class of sites ([Fig. 1d](#)). This analysis revealed that the support for the *Korarchaeota* tree in the simpler/underfitting models ([Fig. 1e](#), assessed using model adequacy test [[Giacomelli et al. 2025](#)]), LG + G + F and LG + C60 + G + F, was greater at the more compositionally constrained sites, i.e. effective number of amino acids ($K_{\text{eff}} \leq 9$), whereas with the best-fitting model and second best-fitting model, LG + CAT-PMSF and Poisson + CAT-PMSF ([Fig. 1e](#), assessed using model adequacy test [[Giacomelli et al. 2025](#)]), both constrained and compositionally variable sites supported the Asgard archaea tree ([Fig. 1d](#)). Additionally, under the best-fitting model LG + CAT-PMSF and the second best-fitting model Poisson + CAT-PMSF, the Asgard archaea tree was consistently recovered ([supplementary figs. S10 to S13, Supplementary Material](#) online, [Fig. 1e](#)), and the *Korarchaeota* tree was confidently rejected by an approximately unbiased (AU) test ([Shimodaira 2002](#)) ($P < 0.05$) ([supplementary data S5, Supplementary Material](#) online). These analyses indicate that failure to model site-wise compositional heterogeneity contributes to the incorrect placement of *Njordarchaeia* with *Korarchaeota* in analyses with poorly fitting models such as LG + G + F and LG + C60 + G + F ([Fig. 1e](#)).

Modeling Branch-Wise Compositional Heterogeneity

In addition to site-wise variation, sequence composition can also vary among lineages with, for example, distantly related thermophiles converging on greater use of thermostable amino acids (i.e. Ile, Val, Tyr, Trp, Arg, Glu, Leu (IVYWREL)) ([Zeldovich et al. 2007](#)) or charged amino acids (i.e. Asp, Glu, Lys, Arg (DEKR)) ([Cambillau and Claverie 2000](#); [Szilágyi and Závodszy 2000](#)). *Njordarchaeia* and

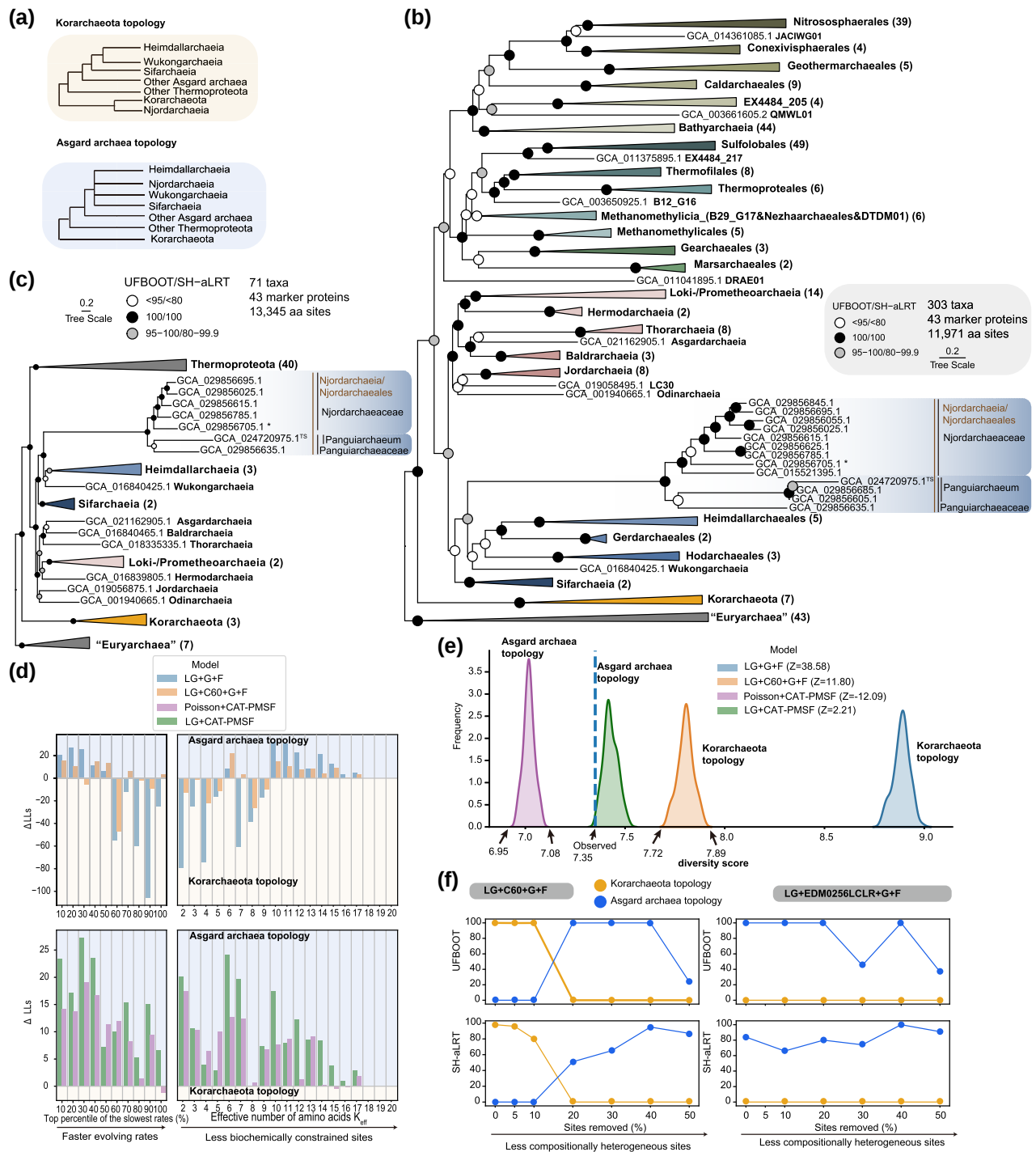


Fig. 1. The phylogenetic placement of Njordarchaeia. a) Schematic of the phylogenetic hypotheses tested in this study. Korarchaeota topology: Njordarchaeia branch sister to Korarchaeota (note that Korarchaeota was recovered as a sister group to a clade comprising both Thermoproteota and Asgard archaea in this study). Asgard archaea topology: Njordarchaeia branch within Asgard archaea. b) Maximum-likelihood phylogenetic analysis of the concatenated 50% top-ranked marker proteins ($n = 43$) and streamlined set2 (303 taxa) using a custom site-heterogeneous substitution model (LG + EDM0256LCLR + G + F). Scale bar: average substitutions per site. The number of taxa in each collapsed clade is shown by the number in parentheses next to the clade name. GCA_029856705.1* is the closest genome to the first proposed type material of *Njordarchaeum guaymasii* (see [supplementary discussion, Supplementary Material](#) online). The GCA_024720975.1^{TS} is the proposed type material for *Panguiarchaeum symbiosum*. The tree is rooted by “Euryarchaea” (including Halobacteriota, Thermoplasmata, Hydrothermarchaeota, Methanobacteriota, Methanobacteriota_A, and Methanobacteriota_B). c) Maximum-likelihood phylogenetic analysis of the concatenated 50% top-ranked marker proteins ($n = 43$) and streamlined set3 (71 taxa) using the LG + CAT-PMSF model. Scale bar: average substitutions per site. The number of taxa in each collapsed clade is shown by the number in parentheses next to the clade name. d) Evaluation of phylogenetic signals for Asgard archaea and Korarchaeota topology under LG + G + F, LG + C60 + G + F, Poisson + CAT-PMSF, and LG + CAT-PMSF models using the streamlined set 3 (71 taxa), listed in an order of increasing model fit. Sites underlying the signals were binned by substitution rates (based on LG + C60 + G + F) and effective amino acids per site (based on LG + CAT-PMSF). e) Distribution of amino acid diversity scores for the simulated alignments under LG + G + F, LG + C60 + G + F, one representative Poisson + CAT-PMSF, and LG + CAT-PMSF simulation using the streamlined set 3 (71 taxa). The dashed line indicates the amino acid diversity score from the real dataset. Z-scores quantifying the difference between the compositional heterogeneity of the real dataset and that of the simulated datasets from each model are listed in brackets for each model. Note that Z-scores of LG + C60 + G + F and Poisson + CAT-PMSF are similar, but the right tail of Poisson + CAT-PMSF is closer to the real dataset, indicating a better fit. f) Impact of progressive removal of the most compositionally biased sites ranked by chi-square score on the statistical support of Asgard archaea and Korarchaeota topology. The support values, UFBOOT and SH-aLRT, were estimated on the 50% top-ranked markers of streamlined set2 (303 taxa) based on LG + C60 + G + F and LG + EDM0256LCLR + G + F models.

Korarchaeota were both predicted to be (hyper-)thermophiles, and the placement of Njordarchaeia with Korarchaeota was suggested to be a result of compositional biases (Eme et al. 2023). In line with this previous study (Eme et al. 2023), a comparison of proteome composition across Archaea indicated that they indeed have similar amino acid compositions (supplementary fig. S14a and b, Supplementary Material online). We subsequently investigated whether compositional attraction between the two lineages might also contribute to the recovery of the Korarchaeota tree under simple models. First, we performed a site-filtering analysis. We ranked sites by compositional bias, iteratively removing the sites that made the greatest individual contribution to across-branch-compositional heterogeneity (Dombrowski et al. 2020; Baker et al. 2024) (see Methods), and re-evaluated support for the competing topologies (Fig. 1f). Removal of the top 20% of most biased sites shifted support from the Korarchaeota topology to the Asgard archaea topology under LG + C60 + G + F, while under the better-fitting and custom site-heterogeneous model, LG + EDM0256LCLR + G + F, support was unchanged, with all analyses favoring the Asgard archaea topology. We interpret these results as evidence that compositional attraction between thermophilic branches favors the Korarchaeota topology in the full dataset, but whether this non-phylogenetic signal drives the final result depends on other aspects of the analysis. While neither the LG + C60 + G + F model nor the LG + EDM0256LCLR + G + F model account for across-branch compositional variation, C60 fares worse than EDM0256 in modeling across-site variation (supplementary fig. S11, Supplementary Material online; Bayesian information criterion (BIC) of LG + C60 + G + F: 6668927.72, LG + EDM0256LCLR + G + F: 6532626.81). We therefore suggest that support for the Korarchaeota topology from the 303 taxa set in the C60 analysis reflects a combination of phylogenetic artifacts from both unmodeled branch-wise compositional attraction and inadequately modeled site-wise compositional variation (i.e. long-branch attraction), with support shifting to the Asgard archaea topology when one of these sources of error is removed, either by filtering out highly biased sites or by better modeling of site-wise compositions with LG + EDM0256LCLR + G + F.

Second, we identify RYPEIWKVL and QSNTDGC, as over-represented and underrepresented amino acids in putative (hyper-)thermophiles, respectively, in the concatenation (Zeldovich et al. 2007) (see Methods; supplementary fig. S14c and d, Supplementary Material online). We thus performed phylogenetic analysis using a substitution model that accounts for changing compositions across the tree, i.e. GFMix (Muñoz-Gómez et al. 2022). The LG + EDM0256LCLR + G + F + GFMix (RYPEIWKVL/QSNTDGC and RYPEI/QSNT) and LG + C60 + G + F + GFMix (RYPEIWKVL/QSNTDGC and RYPEI/QSNT) model provided strong support for the Asgard archaea tree over the Korarchaeota tree (supplementary data S6, Supplementary Material online), again consistent with the hypothesis that compositional attraction due to a shared environmental adaptation contributes to the Korarchaeota tree in analyses with models that do not account for branch heterogeneity (Eme et al. 2023).

Is the Inferred Phylogenetic Position of Panguiarchaeaceae and Njordarchaeaceae Affected by Metagenomic Contamination?

The phylogenetic position of Njordarchaeia has been the subject of debate based on different marker sets. A recent analysis

(Zhang et al. 2025) argued that these MAGs may be chimeric, comprising sequences derived from both Thermoproteota and Asgard archaea contributing to conflicting phylogenetic placements observed in different marker sets. It has been pointed out that MAGs often contain various degrees of contamination (Parks et al. 2015; Shaiber and Eren 2019; Chen et al. 2020), and the MAGs analyzed in this study are no exception, as assessed by CheckM and differential coverage (Fig. 2; supplementary data S1, Supplementary Material online). Nevertheless, the key question is whether the marker genes used to place these lineages are affected by contamination. Using the same analytical approach, we inspected the distribution of marker genes and within-bin coverage variation of Panguiarchaeaceae and Njordarchaeaceae MAGs across different metagenomes from hydrothermal vents in the Guaymas Basin (Dombrowski et al. 2018) and a hot spring in Yunnan, China (Qu et al. 2023). CheckM analysis indicates that all representative Njordarchaeia MAGs analyzed here (taxon set 303; supplementary data S1, Supplementary Material online) are of at least medium quality (Bowers et al. 2017) with only a moderate level of contamination (<10%), with 53% (7 MAGs) having low level of contamination (<5%) (supplementary data S1, Supplementary Material online). MAGs with above medium quality are the majority in different databases such as the Genomes from Earth's Microbiomes catalogs (Nayfach et al. 2021), and reported in various studies (Dombrowski et al. 2020; Martijn et al. 2020; Zhang et al. 2025). To investigate whether our marker gene phylogenies are affected by this low level of contamination, we chose to examine 4 MAGs from our focal taxon set that were derived from hydrothermal vents in Guaymas basin and a hot spring from China (estimated contamination: 0.93% to 7.01%; Fig. 2). Three of these Njordarchaeaceae MAGs were also examined in a previous study (Zhang et al. 2025). While our analyses confirmed the presence of small potentially contaminating contigs (Fig. 2a–d), all our marker genes mapped back to a single majority cluster corresponding to the target population, even in the MAG estimated with the highest level of contamination (Fig. 2c). Additionally, njordarchaeial homologs formed a monophyletic group in all single gene phylogenies of the markers used in our concatenation (supplementary data S4, Supplementary Material online). This suggests that contamination, though present at low to moderate levels in some of these MAGs, does not affect the marker genes used, and so does not underpin the phylogenetic conflicts. Instead, conflicting support appears to depend on the substitution model used. For example, analyses of two marker genes situated within a non-contaminated contig (Fig. 2a–c, highlighted in thick outlined box) under the poorly fitting LG + F + G and LG + C60 + F + G models differed in their support for the Asgard and Korarchaeota topologies, whereas analysis of the same marker genes under the best-fitting LG + CAT-PMSF model provides moderate support for the Asgard archaea topology (Fig. 2a–c). We also observed in our dataset that LG + C60 + G + F did not significantly distinguish these two topologies (AU tests [Shimodaira 2002] $P < 0.05$; supplementary data S5, Supplementary Material online) nor describe site-heterogeneity adequately (Fig. 1e). However, as model fit improved, consistent support for the Asgard topology emerged (supplementary data S5, Supplementary Material online, Fig. 1d–f).

While we saw no evidence of contamination in our analyses, we did see some variation in support for the Korarchaeota or Asgard topologies on a per-gene basis when the data were

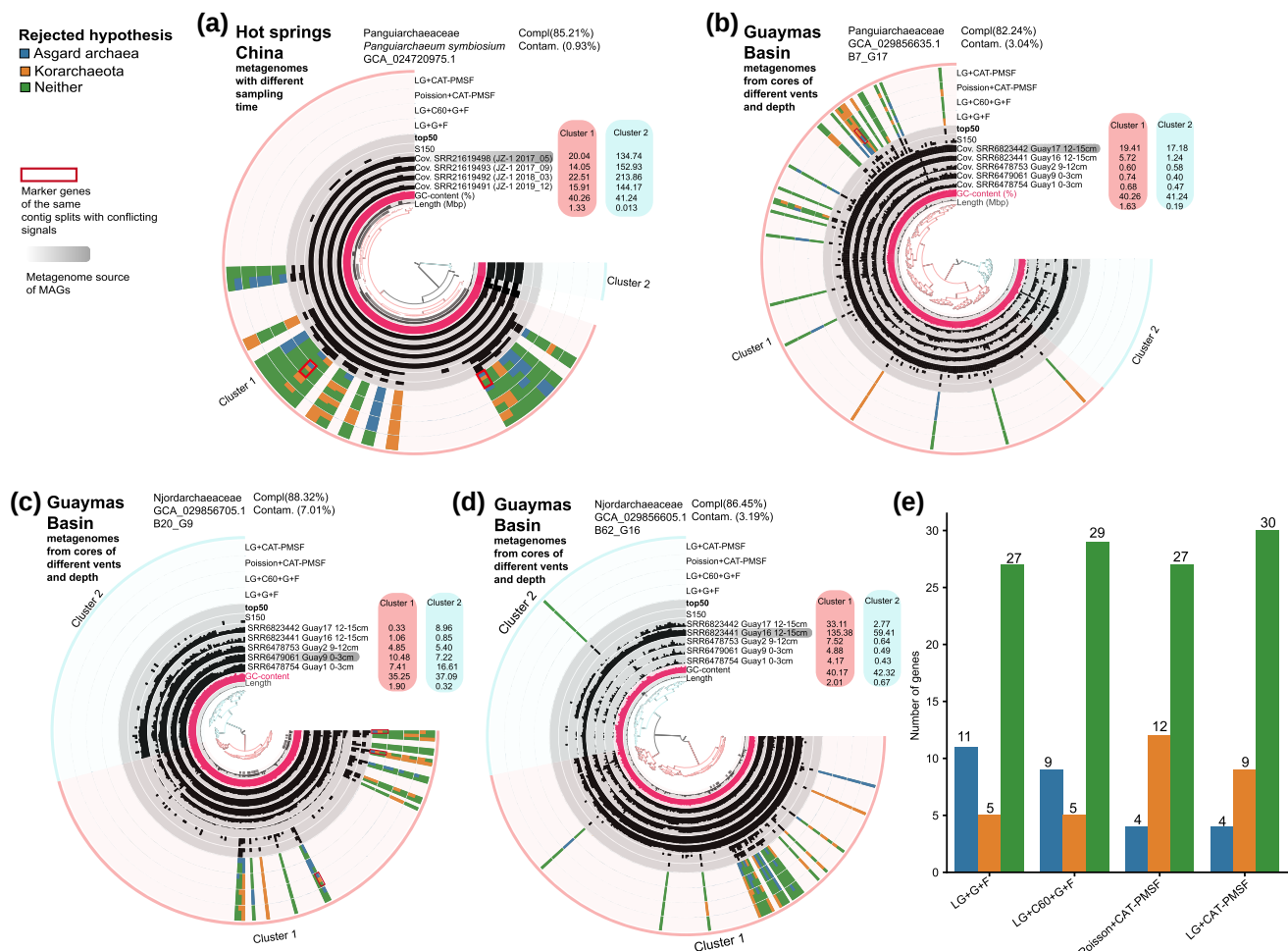


Fig. 2. Hierarchical clustering of contigs in 4 MAGs from Pangiarchaeaceae (a, b) and Njordarchaeaceae (c, d) based on sequence composition and depth of coverage across different metagenomes. Note that the tip of the hierarchical clustering tree represents splits of long contigs at a size of 20,000 bp. This analysis demonstrates that metagenomic contamination is not a source of conflicting phylogenetic signals. The ring indicating coverage across the source metagenomes is highlighted with a shaded area. The locations of marker gene sequences used in this study (ring labeled top50) and marker gene homologs from a previous study (ring labeled S150) are shown as bar plots, with bar height corresponding to the number of genes. Notably, our marker gene sequences are located on clusters of contigs that share similar coverage profiles across metagenomes, as are the majority of the S150 homologs. Support for Asgard archaea and Korarchaeota topologies is shown in a stacked bar plot, with genes on the same contig that show conflicting topological support highlighted in thick outlined boxes. e) Barplot showing the number of genes that confidently reject Asgard archaea, Korarchaeota, or neither of the two topologies under different models.

analyzed using different models. To quantify disagreements among marker genes, we used AU tests (Shimodaira 2002) to determine, for each marker gene and substitution model, whether either the Korarchaeota tree or the Asgard archaea tree could be rejected (at $P < 0.05$). Across all models tested, 27 of 43 marker genes did not reject either hypothesis, suggesting that individual marker genes often do not contain sufficient information to discriminate among phylogenetic hypotheses (Fig. 2e). For the remaining marker genes, support on a per-gene basis shifted from the Korarchaeota to the Asgard archaea topology as model fit (assessed using the model adequacy test [Giacomelli et al. 2025]) improved. Under the worst-fitting LG+G+F model, 5 genes rejected the Korarchaeota topology and 11 rejected the Asgard topology at the significance level (Fig. 2e). Under the site-heterogeneous LG+C60+G+F model, 5 genes rejected Korarchaeota and 9 rejected Asgard archaea. In contrast, under the best-fitting (LG+CAT-PMSF) and second best-fitting (Poisson + CAT-PMSF) models, 9 and 12 genes, respectively, rejected the Korarchaeota topology, while 4 rejected the Asgard topology (Fig. 2e). The improvement in congruence

across marker genes and the shift in support to the Asgard topology as model fit improves suggests that the apparent strong but inconsistent support observed in analyses under the poorly fitting LG+G+F and LG+C60+G+F models is artifactual, resulting from the failure of these models to adequately account for non-historical signal in sequence data (Fig. 1e; supplementary information S1.3 to S1.4, Supplementary Material online).

Assignment of Njordarchaeia to Asgard Archaea Is Supported by Shared ESPs and Protein Presence–Absence Profiles

The phylogenetic analyses described above provided strong support for the placement of Pangiarchaeaceae within Njordarchaeales as a sister to Njordarchaeaceae within the Asgard archaea (Fig. 1b–f). Consistent with that placement, and in line with (Eme et al. 2023), we observed that Pangiarchaeaceae and Njordarchaeaceae encode a number of eukaryotic signature proteins (ESPs) shared with other Asgard archaea (Fig. 3; supplementary data S7, Supplementary Material online), based on analyses using the Asgard Cluster

of Orthologues (AsCOGs) database (Liu et al. 2021). Of particular interest, we identified homologs of ribosomal L28e/Mak16 (PF01778) in Panguiarchaeaceae MAGs, whose distribution in published archaeal genomes is otherwise restricted to other Njordarchaeia, the Hodarchaeales, and Wukongarchaeia (Eme et al. 2023). Phylogenetic analysis of ribosomal L28e/Mak16 revealed a topology consistent with the species tree (supplementary fig. S15, Supplementary Material online; supplementary information S2.2.1, Supplementary Material online), with the Njordarchaeia sequences branching sister to Wukongarchaeia and other Heimdallarchaeia. Furthermore, Panguiarchaeaceae MAGs encode an actin homolog that branches with the conserved lokiactin clade present in all asgard archaeal lineages when compositionally biased sites were removed (supplementary figs. S16 and S17, Supplementary Material online; supplementary information S2.2.2, Supplementary Material online). Homologs of proteins involved in the N-glycosylation process were also identified in both Panguiarchaeaceae and Njordarchaeaceae MAGs (Eme et al. 2023), and the phylogenetic analyses support their affiliation with Asgard archaea homologs (supplementary figs. S18 and S19, Supplementary Material online; supplementary information S2.2.3, Supplementary Material online). Furthermore, phylogenetic analyses of Njordarchaeia SNF7 proteins support them as Asgard archaea homologs (supplementary fig. S20, Supplementary Material online; supplementary information S2.2.4, Supplementary Material online).

Finally, protein presence–absence profiles revealed a pattern of transcription, translation, and replication distinct from Korarchaeota (supplementary figs. S21 and S22, Supplementary Material online). Interestingly, however, the profiles for other cluster of orthologs (COG) categories were more similar to that of Korarchaeota than to other Asgard archaea (supplementary figs. S21 and S22, Supplementary Material online), likely reflecting adaptation to a high-temperature environment.

Incomplete Biosynthetic Pathways Indicate that *Panguiarchaeum* May Depend on Other Organisms

Comparative genome analyses and the inference of metabolic potentials for Njordarchaeia MAGs (i.e. the Panguiarchaeaceae clade, including *Panguiarchaeum* and GCA_029856635.1, and the Njordarchaeaceae clade) suggest that these organisms are thermophilic fermentative heterotrophs in agreement with previous work (Qu et al. 2023) that have varying levels of auxotrophies; that is, they cannot synthesize all organic compounds needed for growth (supplementary fig. S23, Supplementary Material online). Of note, members of the *Panguiarchaeum* show a reduced metabolic potential, in particular in purine and lipid biosynthesis pathways, compared to Njordarchaeaceae (supplementary figs. S23 to S27, Supplementary Material online; supplementary information S3.1 to S3.2, Supplementary Material online).

Catabolism and Energy Conservation

The catabolic potential of the *Panguiarchaeum* clade (i.e. upon divergence from GCA_029856635.1) is more limited than that of Njordarchaeaceae (Fig. 4; supplementary data S8 to S10, Supplementary Material online; supplementary fig. S23, Supplementary Material online). In particular, while both clades encode a nearly complete Embden–Meyerhof–Parnas (EMP) pathway, members of Njordarchaeaceae encode more genes for enzymes of the tricarboxylic acid cycle (TCA) and beta-oxidation pathway. Furthermore, members

of both clades might have the potential to grow on organic substrates, which could be fermented to acetate by a putative acetate-CoA ligase (ACD) with concomitant production of ATP (Glasemacher et al. 1997). Specifically, *Panguiarchaeum* representatives revealed a potential to harness electrons from organic substrates such as amino acids, peptides, 2-ketoacids (e.g. pyruvate, oxoglutarate, and indole-pyruvate) (Qu et al. 2023), and aldehydes (supplementary data S8, Supplementary Material online; supplementary information S3.2, Supplementary Material online). Additionally, we identified homologs of the small and large subunits of [NiFe]-hydrogenases belonging to group 3 and group 4 in *Panguiarchaeum* MAGs. Phylogenetic analyses of the large subunit of the [NiFe]-hydrogenase indicated that panguiarchaeal homologs belong to subgroup 4g (Qu et al. 2023), forming a clade with Odinarchaeia sequences (supplementary fig. S28a, Supplementary Material online). Group 4g [NiFe]-hydrogenases of *Panguiarchaeum* encode a Nuo-L-like subunit, which might be involved in sodium/ion translocation (Yu et al. 2018) (supplementary fig. S28b, Supplementary Material online). This suggests that *Panguiarchaeum* may conserve energy by coupling the oxidation of organic substrates via reduced ferredoxin to the formation of H₂ to generate a proton gradient for ATP synthase. Members of *Panguiarchaeum* also contain various group 3 [NiFe]-hydrogenases assigned to subgroups 3c and 3b (Qu et al. 2023), which may be involved in the reoxidation of co-factors during carbon oxidation and the production of H₂, a reversible process (supplementary fig. S29, Supplementary Material online; supplementary data S11, Supplementary Material online). Together, this indicates that representatives of the *Panguiarchaeum* clade may be able to conserve energy through substrate-level phosphorylation as well as via the ATP synthase using a proton gradient generated by the oxidation of ferredoxin, similar to what has been hypothesized for Odinarchaeia and Thermococci (Spang et al. 2019).

Anabolism

The biosynthetic potential of members of the *Panguiarchaeum* clade seemed more restricted than that of Njordarchaeaceae (supplementary fig. S23, Supplementary Material online). For instance, all 3 *Panguiarchaeum* MAGs lack various genes encoding enzymes for de novo purine biosynthesis (from phosphoribosyl pyrophosphate to inosine monophosphate) (Qu et al. 2023), such as glutamine phosphoribosylpyrophosphate amidotransferase (PurF) and phosphoribosylaminoimidazole (AIR) synthetase (PurM), indicating a possible dependence on external purine sources (Brown et al. 2011). Notably, *Panguiarchaeum* clade MAGs do not encode enzymes of the mevalonate pathway (MVAP) for isoprenoid biosynthesis and fatty acid biosynthesis, as indicated previously (Qu et al. 2023). However, the presence of CDP-archaeol synthase (CarS), phosphatidylglycerophosphate synthase (PgsA), phosphatidylglycerophosphate synthase (PssA), and cardiolipin synthase (Cls) suggests that *Panguiarchaeum* have the potential to attach polar head groups to di-o-geranylgeranyl glyceryl phosphate (DGGGP). The *Panguiarchaeum* MAGs also encode a gene for digeranylgeranyl glycerophospholipid reductase (GGR), which could catalyze the hydrogenation/saturation of geranylgeranyl chains of DGGGP (Nishimura and Eguchi 2006). Phylogenetic analyses of CarS showed that the homologs of *Panguiarchaeum* branched sister to Heimdallarchaeia with moderate support (UFBOOT/SH-aLRT: 88.8/83; supplementary fig. S30, Supplementary Material online). Similarly, although the PgsA homologs of *Panguiarchaeum*

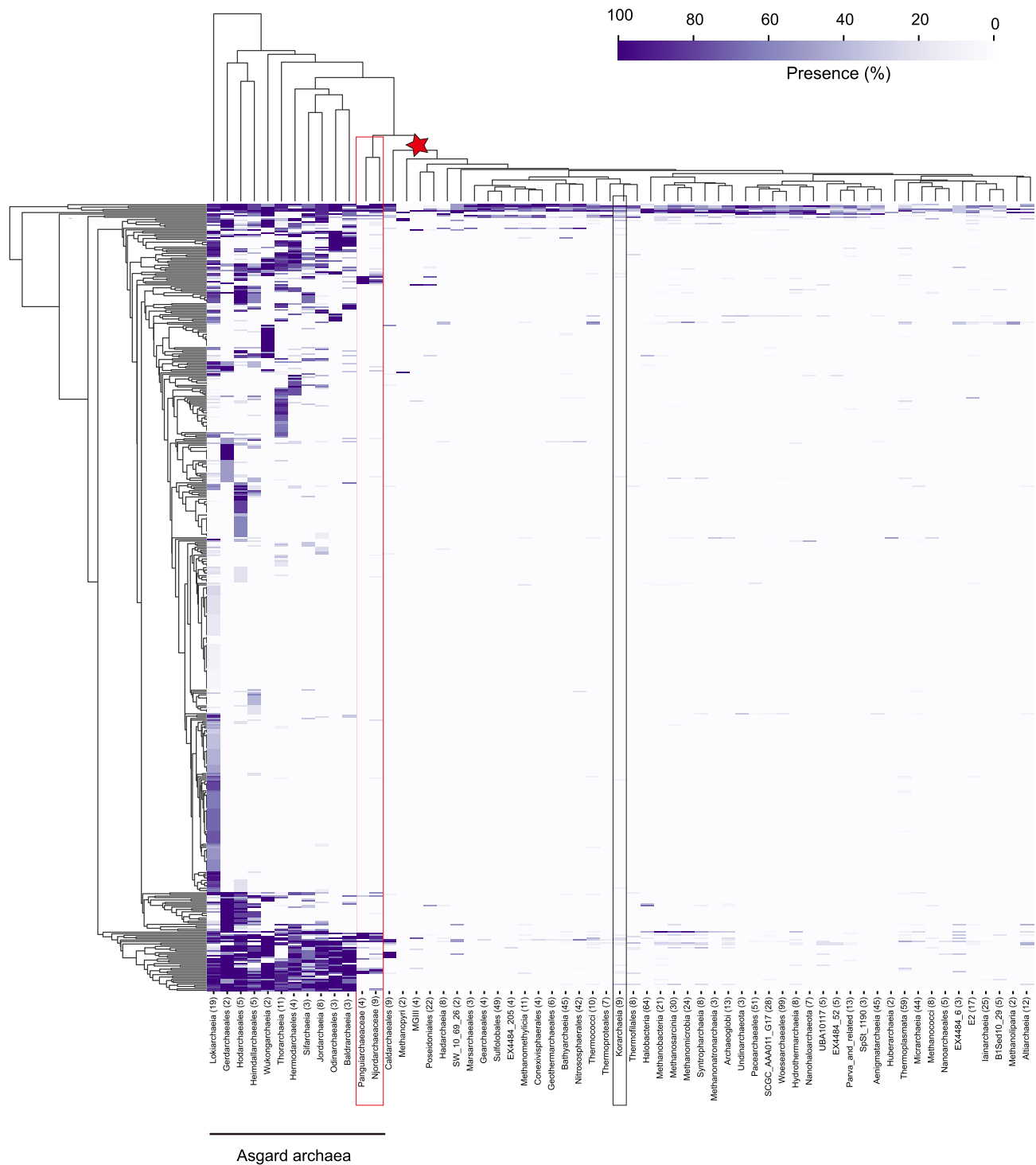


Fig. 3. Presence of Asgard cluster of orthologues (AsCOGs) across major archaeal lineages. Pangiarchaeaceae and Njordarchaeaceae shared a number of AsCOGs homologs with other Asgard archaea, in contrast to Korarchaeota. AsCOG presence was determined across 966 archaeal genomes, and the occurrence of AsCOGs within any taxonomic clade of interest (based on the genome taxonomy database) was recorded as a percentage. Clusters with fewer than 2 genomes were not included. The number of genomes for each cluster is shown in parentheses.

branched sister to a clade comprising one *Thermococcus* and two Heimdallarchaeia sequences with weak support (UFBOOT/SH-aLRT: 50.5/83), they were nested within a cluster of predominantly Asgard archaea sequences (supplementary fig. S31, Supplementary Material online). Thus, these genes were likely inherited vertically from their common ancestor with Heimdallarchaeia (supplementary fig. S30 to S32, Supplementary Material online).

Pangiarchaeum Has Evolved by Genome Streamlining From a Larger Ancestor With Heimdallarchaeia

Reconciliation analyses have been used to infer the ancestral genome content of various archaeal lineages (Williams et al. 2017, 2024; Martijn et al. 2020; Huang et al. 2021; Baker et al. 2024), including Asgard archaea (Eme et al. 2023), shedding light on ancestral gene repertoires and gene content

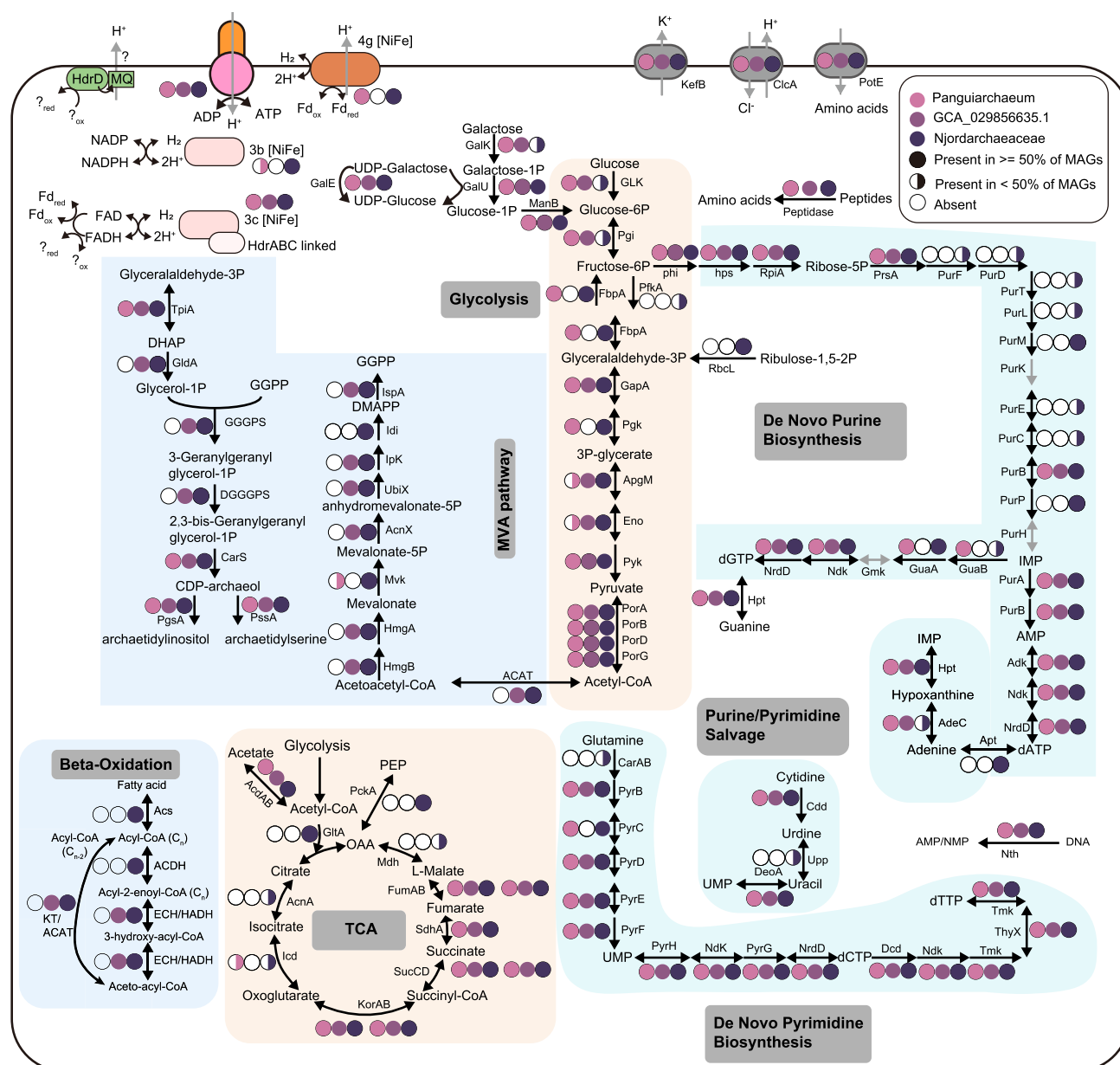


Fig. 4. Metabolic characteristics of Njordarchaeia. Overall, Njordarchaeia are thermophilic fermentative heterotrophs, with core pathways of varying completeness, suggesting some core metabolites must be obtained from other organisms in the environment. Full circles indicate that a gene is present in all or more than 50% of the MAGs, while half circles indicate that a gene is found in at least one, but less than half, of the MAGs. Open circles denote genes absent from all MAGs of a group. A detailed list of genes encoded by Njordarchaeia can be found in [supplementary data S7, S8, and S9, Supplementary Material](#) online.

evolution. To reconstruct the evolution of a putative symbiotic lifestyle of *Panguiarchaeum*, we reconciled single gene trees of 6077 arCOG gene families with the inferred species tree (from streamlined set 2 with 303 taxa, [Fig. 1b](#)) to determine the pattern and extent of reductive genome evolution in the *Panguiarchaeum* clade and its sister lineages.

Our analysis suggests an ongoing decrease in gene content through time in the evolution of Njordarchaeia, similar to a previous analysis ([Eme et al. 2023](#)), from around 1951 arCOG families in the common ancestor with Wukong–Heimdallarchaeia to 911 to 1545 (mean: 1186.5) arCOG families among extant Njordarchaeia ([supplementary data S12, Supplementary Material](#) online). Based on the relationship between arCOG family members and the genome size of extant taxa (Pearson’s correlation coefficient: 0.89, P -value: 1.96×10^{-110}), we

inferred the genome size of the last common ancestor of Wukong–Njord–Heimdallarchaeia to be 2.85 Mbp (95% confidence interval, 2.57 to 3.06; [supplementary data S12, Supplementary Material](#) online). A notable reduction in genome size is inferred at the branch leading to the Njordarchaeia (95% confidence interval: between 0.53 and 0.84 Mbp; node id: 571 to 563) after divergence from their common ancestor with Wukong–Heimdallarchaeia ([Fig. 5a; supplementary data S12, Supplementary Material](#) online). The arCOG families lost along this branch include components of amino acid biosynthesis, de novo purine biosynthesis (from ribose-5-phosphate to inosine monophosphate (IMP)), Wood–Ljungdahl pathway genes and cofactors (methanofuran) biosynthesis ([Fig. 5b; supplementary data S13, Supplementary Material](#) online). Additional genes, including those encoding various enzymes



reduction between 0.13 and 0.17 Mbp (95% confidence interval). *Panguiarchaeum* and Njordarchaeaceae were inferred to encode fewer ESPs than the other Asgard archaea (Fig. 3), and the loss on the Njordarchaeaceae and Panguiarchaeaceae stem is predicted to include the reduction in ESP repertoire

as well. For instance, we inferred the loss of a Roadblock/LC7 domain-containing protein (arCOG02605_01) on the Njordarchaeaceae branch; the loss of Ras family GTPase (arCOG05343_01) and Yip1 domain family (arCOG 02054_01) on the Panguiarchaeaceae branch; the loss of VPS4 (arCOG01307_01) on the *Panguiarchaeum* branch (i.e. upon divergence from GCA_029856635.1) (supplementary data S13, Supplementary Material online); and the absence of Snf7 family proteins in the extant *Panguiarchaeum* MAGs indicated their loss on the *Panguiarchaeum* branch (supplementary fig. S20, Supplementary Material online).

Njordarchaeia encodes reverse gyrase, a topoisomerase typically encoded by hyperthermophiles (Forterre 2002). Consistent with previous work (Eme et al. 2023) and the proposal of a thermophilic ancestry of Asgard archaea (Lu et al. 2024), we inferred the presence of reverse gyrase (presence probability, PP=0.96) in the last common ancestor of Asgard archaea. This gene family was, however, subsequently lost in Heimdallarchaeia (supplementary data S13, Supplementary Material online) but not in Wukongarchaeia and Njordarchaeia. Interestingly, phylogenetic and reconciliation analyses suggest a later putative re-acquisition of a reverse gyrase in the *Panguiarchaeum* clade (supplementary figs. S33 to S34, Supplementary Material online; supplementary information S3.1.2, Supplementary Material online). The MAG/genome size of putative hyperthermophilic Asgard archaea encoding reverse gyrase (mean: 2.08 Mbp, $n=18$) is significantly smaller than that of Asgard archaea representatives without reverse gyrase (mean: 3.54 Mbp, $n=45$; $P=5.669 \times 10^{-7}$, 2-tailed Wilcoxon rank sum test). Genome reduction has previously been reported in cases of thermophilic adaptation, perhaps due to a selective advantage for smaller cell volumes (and so genome) at high temperatures (Sabath et al. 2013; Pierpont et al. 2024). We therefore hypothesize that the initial reductive evolution of Njordarchaeia might have been associated with their adaptation to high-temperature environments. The further genome reduction in *Panguiarchaeum*, including the loss of essential genes for biosynthetic pathways, might be associated with a transition to a host-associated lifestyle (McCutcheon and Moran 2011).

Panguiarchaeum Co-Occurs With Thermoproteota Taxa

To investigate the inferred dependence of *Panguiarchaeum* on other community members, we assessed the extent to which members of this group co-occur with other microbes using network inferences. First, we tested the inference of our co-occurrence networks with the known host-symbiont pair *Nanoarchaeum equitans* and *Ignicoccus hospitalis* (Huber et al. 2002), and the proposed symbiosis between *Huberarchaeum crystalense* and *Altiarchaeum hamiconexum* (Schwank et al. 2019). We did not detect any associations between the former, mainly due to insufficient sampling of *N. equitans* which is present in a single sample only (supplementary figs. S35 and S36, Supplementary Material online; supplementary information S4, Supplementary Material online). However, for *Huberarchaeum*-*Altiarchaeum*, we found a high correlation (Spearman's rho = 0.931) and a robust detection frequency (100%) of this suggested interaction starting at a sample size of 40 (supplementary fig. S36f, Supplementary Material online; see Methods). Next, we identified 120 samples containing taxa annotated as family 'Panguiarchaeaceae' in GTDB (Rinke et al. 2021), a

lineage equivalent to the class Njordarchaeia and order Njordarchaeales proposed in this manuscript, from 248,559 NCBI metagenomes community profiles in the 'Sandpiper' website (<https://sandpiper.qut.edu.au>) (Woodcroft et al. 2025). Community profiles of these 120 samples were subjected to various data transformations (see Methods) and network inferences. Across all original networks, i.e. independent of filtering criteria applied before network inference, Njordarchaeales taxa were associated with Thermoproteota taxa. The frequency of association between these taxa was up to ~4 times higher than that between Njordarchaeales and any other taxa or random number controls. On phylum, class, and order level, Njordarchaeales were associated with Thermoproteota (73 out of 256), Thermoprotei_A (33 of 73), and Sulfolobales (23 of 33) (supplementary data S14 to S15, Supplementary Material online), respectively, in agreement with abundance-based co-occurrence patterns (Fig. 6). Within the Njordarchaeales, there was only one OTU (OTU_804; supplementary data S14, Supplementary Material online) assigned to the genus *Panguiarchaeum*, which found association partners in the network. This genus had 29 associations in total, whereby the taxonomic annotation of its partners was similar to the associations detected for the group as a whole, i.e. Thermoproteota (15 of 29), Thermoprotei_A (9 of 15), and Sulfolobales (4 of 9), and Desulfurococcaceae (3 of 4) (supplementary data S15, Supplementary Material online). These results suggest that members of the Njordarchaeales, in particular the genus *Panguiarchaeum*, might interact with members of the phylum Thermoproteota, in particular with Sulfolobales, an order comprising hosts of thermophilic Nanoarchaeota (Huber et al. 2002; Podar et al. 2013; Munson-McGee et al. 2015; Wurch et al. 2016; St John et al. 2019; Kato et al. 2022; Sakai, Nur et al. 2022).

Discussion

Reconstructing the tree that best depicts the evolutionary relationship between major archaeal lineages is important for our understanding of archaeal genome evolution, the evolution of archaeal symbioses, and eukaryogenesis. Our comprehensive phylogenetic and comparative genomic analyses reveal that *Panguiarchaeum*, together with GCA_029856635.1, forms a family within the Njordarchaeia as part of the Asgard archaea.

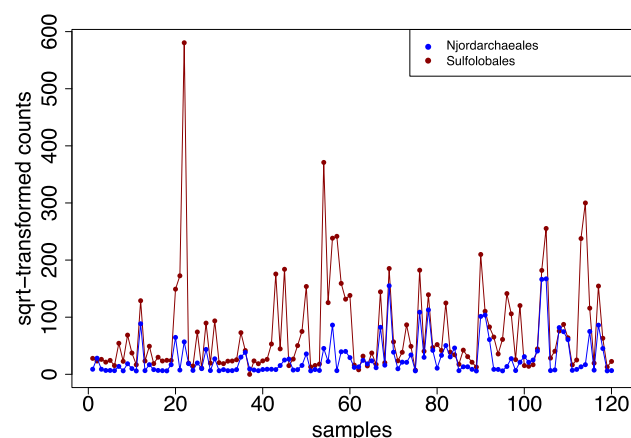


Fig. 6. Abundance-based co-occurrence of Njordarchaeales and Sulfolobales. For plotting, the counts of both orders Njordarchaeales (including Panguiarchaeaceae) and Sulfolobales, were summed in each sample and plotted. For visualization purposes, the counts were square root transformed.

Thus, the previous placement of the *Panguiarchaeum* clade as a sister group to Korarchaeota appears to be due to phylogenetic artifacts (Qu et al. 2023), including the use of problematic marker gene families in concatenations and insufficient modeling of compositional biases and site heterogeneity similar to what has been shown for Njordarchaeales (Eme et al. 2023). In contrast, our analyses did not provide support for the recent suggestion that phylogenetic placement of Njordarchaeia has been affected by metagenomic contamination (Zhang et al. 2025). While some Njordarchaeia MAGs do show evidence of low to moderate levels of contamination (Fig. 2), we found no evidence that marker genes were affected in any case. Our analyses instead suggest that disagreement among marker genes as to the placement of Njordarchaeia is a result of poor-fitting phylogenetic models, not MAG chimerism. These results emphasise the importance of evolutionary model fit in analyses of prokaryotic evolution, particularly for datasets containing pervasive compositional heterogeneity.

The placement of *Panguiarchaeum* within the Asgard archaea is supported by the presence of various ESPs shared with Njordarchaeaceae. However, metabolic reconstructions reveal that members of *Panguiarchaeum* have various auxotrophies and encode fewer ESPs than other Asgard archaea. This is indicative of genome reduction potentially associated with a host-associated lifestyle, with co-occurrence analyses pointing toward putative associations with other archaea. Notably, *Ca. P. syntrophicum*, the first cultivated representative of the Asgard archaea (i.e. a member of the Lokiarchaeia/Prometheoarchaeia), appears to syntrophically grow on amino acids with partner organisms such as *Halodesulfovibrio* and *Methanogenium* (Imachi et al. 2020). However, this symbiotic interaction does not seem to be obligate and has not resulted in reductive genome evolution. Our inferences of the energy metabolism of *Panguiarchaeum*, and perhaps Njordarchaeaceae as well, which seems to resemble that of Odinararchaeia and *Thermococcus*, indicate the potential for syntrophic growth (Schut et al. 2013; Topçuoğlu et al. 2016; Yu et al. 2018; Spang et al. 2019). For example, *Panguiarchaeum* encodes 3 different [NiFe]-hydrogenases, including a putative membrane-bound group 4g [NiFe]-hydrogenase, and might use electrons harnessed from small organic acids/simple carbohydrates to produce H₂. While speculative in the absence of experimental data, it may be hypothesized that the hydrogen could support the growth of an H₂-utilizing archaeon, perhaps, in return for essential cellular building blocks such as lipids and vitamins (Ver Eecke et al. 2012; Topçuoğlu et al. 2016; Imachi et al. 2020; Yu et al. 2024).

Our ancestral reconstructions infer a complex ancestor of Asgard archaea, as was observed previously (Eme et al. 2023). Further, our analyses predict that the Asgard archaeal ancestor has encoded a Wood–Lundgal pathway (WLP) (Fig. 5; supplementary data S13, Supplementary Material online), i.e. a pathway that can serve as an electron sink supporting both autotrophic and heterotrophic growth modes (Ragsdale and Pierce 2008; Schuchmann and Müller 2014, 2016). However, this pathway seems to have been lost in various members of the Asgard archaea (Spang et al. 2019; Liu et al. 2021) (supplementary data S10 to S13, Supplementary Material online); specifically, our analyses indicate that genes for proteins involved in the WLP were lost on the branch leading to the Njordarchaeia after their split from Heimdall- and Wukongarchaeia (Fig. 5). The absence of the WLP as an electron sink may have strengthened the dependency of members of this group on syntrophic H₂-utilizing partners, such as

Thermoproteota (Huber et al. 2000; Sakai, Nakamura et al. 2022; Leung et al. 2024) that could serve as external electron sinks. It is tempting to speculate that such a syntrophic lifestyle may have permitted the loss of genes involved in various biosynthetic pathways in the ancestor of the *Panguiarchaeum* clade (Figs. 4 and 5), increasing the dependency of its members on partner organisms.

This scenario on the evolution of a potentially obligate symbiont lineage among the Asgard archaea points to interesting parallels with the evolutionary trajectory leading to the origin of eukaryotes. Many current models on eukaryogenesis hypothesize syntrophy as a key driver for the establishment of an intricate symbiotic relationship between the archaeal ancestor of eukaryotes (Martin and Müller 1998; Spang et al. 2019; Imachi et al. 2020; López-García and Moreira 2020; Vosseberg et al. 2024), i.e. a likely sister lineage of the Hodarchaeales (Williams et al. 2020; Eme et al. 2023), or of a broader Wukong/Heimdall clade (Liu et al. 2021; Zhang et al. 2025), and the alphaproteobacterial partner that evolved into the mitochondrion (Lane and Martin 2010; Poole and Grihaldo 2014; Ettema 2016; Martijn et al. 2018; Muñoz-Gómez et al. 2022). Similar to *Panguiarchaeum*, the eukaryotic host lineage also appears to have lost ancestral archaeal genes during eukaryogenesis, i.e. the transition from the First to the Last Eukaryotic Common Ancestor. For example, glycolysis and iron-sulfur cluster biosynthesis in eukaryotes are implemented by pathways with patchwork ancestries that include both archaeal, alphaproteobacterial, and other bacterial contributions (Freibert et al. 2017), while the archaeal TCA cycle was replaced by that of the mitochondrion in eukaryotes (Santana-Molina et al. 2025). Similarly, the evolution of *Panguiarchaeum* was shaped by the loss of genes involved in the TCA cycle and several biosynthetic pathways. However, reductive evolution in *Panguiarchaeum* has also led to the loss of several genes encoding ESPs found in its close relatives. In contrast, the eukaryotic lineage expanded an already large protein repertoire and its ESPs via gene duplication and genome expansion, as well as acquisitions from mitochondrial and other bacterial genes along its stem (Pittis and Gabaldón 2016; Eme et al. 2017; Tria et al. 2021; Vosseberg et al. 2021, 2024). Recent analyses have demonstrated that Asgard archaeal genomes often contain substantial proportions of horizontally acquired genes of bacterial origin (Wu et al. 2022), and we hypothesize that part of the difference in evolutionary trajectories of *Panguiarchaeum* and eukaryotes may result from the nature of the symbiotic partner, either archaeal (*Panguiarchaeum*) or bacterial (eukaryotes), as well as environmental parameters (e.g. temperature). To address these hypotheses, prospective efforts would benefit from a better sampling of symbiotic members of the Asgard archaea, including both syntrophic and/or genome-reduced representatives, combined with their co-cultivation with partner organisms and physiological characterization.

Materials and Methods

Marker Gene Inspection and Ranking Analysis

Marker Gene Inspection

To accurately place Njordarchaeales and Panguiarchaeales in the archaeal tree of life, we first downloaded the NM57 A64 dataset from Eme et al. 2023 (Eme et al. 2023) and assigned a COG family to each sequence using hmmsearch (Finn et al. 2011) v3.3.2 (settings: -E 1e-5). The best hit for each

protein sequence was selected based on the lowest e-value and highest score, and the COG families representing the majority of sequences in each marker gene were selected to annotate each marker. These COG families were then compared to 185 marker gene families used in 3 other previous studies (Dombrowski et al. 2020; Williams et al. 2020; Moody et al. 2022) to create a non-redundant marker set comprising 112 unique COG families for further inspection. Potential orthologous sequences assigned to these 112 unique families were identified in 966 archaeal and 1,325 bacterial genomes/MAGs using *hmmsearch* v3.3.2 (settings: -E 1e-5) and used for subsequent phylogenetic analyses. Since gene fission of DNA-directed RNA polymerase subunits A and B was reported in some archaea (Werner 2007), we used a custom Hidden Markov Model (HMM) profile of COG0085 and COG0086 to detect “split” subunits and concatenate these subunits before alignment (<https://doi.org/10.5281/zenodo.15848218>).

We then aligned sequences assigned to these 112 marker gene families using *mafft* v7.453 (Katoh and Standley 2013) and removed poorly aligned amino acid sites with *BMGE* 1.12 (Criscuolo and Grigaldo 2010) (settings: -m BLOSUM30 -h 0.55). The maximum likelihood trees for these marker genes were inferred using *IQ-tree* v2.1.2 (Minh et al. 2020) (-m LG + G -B 1000). We decorated the tips of these marker gene trees with annotations from Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology (Aramaki et al. 2020), Protein FAMilies database (PFAM) (Bateman et al. 2004), COG (Galperin et al. 2021), and archeal COG (arCOG) (Makarova et al. 2015) information. We manually inspected all single protein phylogenies to identify marker gene families that did not meet archaeal monophyly and excluded those families from subsequent analyses. Furthermore, we identified and removed paralogues and sequences that experienced horizontal gene transfer (HGT) events from bacteria to archaea. In total, this yielded 100 COG marker gene families.

We performed a second round of inspection and curation based on corresponding arCOG families of these 100 COG families except COG0085 and COG0086 (see above). In this round, we only included the marker gene families if homologs were present in at least 60% of archaeal taxa and were duplicated in less than 20% of the genomes of major archaeal lineages (e.g. Asgard archaea, Thermoproteota, and Halobacteriota), which resulted in 90 unique families with arCOG identifiers. Sequences were realigned with *mafft-linsi* v7.453 (Katoh and Standley 2013), and poorly aligned sites were removed using *BMGE* as described above. The initial ML trees were inferred using *IQ-tree* v 2.1.2 (Minh et al. 2020) (-m LG + G -B 1000). These trees were then used as guide tree to perform phylogenetic inference based on the best-fitting model (Kalyanamoorthy et al. 2017) (settings: -mset LG -madd LG + C10, LG + C20, LG + C30, LG + C40, LG + C50, LG + C60, LG + C10 + R + F, LG + C20 + R + F, LG + C30 + R + F, LG + C40 + R + F, LG + C50 + R + F, LG + C60 + R + F -score-diff all) with 1,000 ultrafast bootstraps. We inspected the single gene trees to identify and remove sequences that included paralogues and/or were affected by long-branch attraction (LBA) artifacts or HGT.

Ranking Analysis

We used a previously developed marker gene ranking procedure to rank each of the 90 markers based on the extent to which they recovered established archaeal phylum- or order-level lineages (Dombrowski et al. 2020; Moody et al. 2022). In brief, the number of splits (i.e. the occurrence of a certain taxon failing to group within its expected taxonomic clade;

supplementary data S16, Supplementary Material online) was counted across all bootstrap trees for each marker gene. We then defined the highest ranking (top 50%, 43 markers) marker sets based on the number of splits per phylogenetic cluster and the total number of splits normalized by the number of genomes within each tree.

Concatenation, Site Filtration, and Species Tree Inference

Concatenated Gene Trees Using Taxon Set 1: i.e. 966 Taxa

The top 50% marker gene sequences were aligned with *mafft-linsi* v7.453 (Katoh and Standley 2013), and poorly aligned amino acid sites were removed with *BMGE* 1.12 (Criscuolo and Grigaldo 2010) (settings: -m BLOSUM30 -h 0.55). Genomes with more than 50% gaps in the alignment were further removed. The trimmed sequences of these 45 markers were then subsequently concatenated into a supermatrix comprising 966 taxa. Two additional supermatrices were constructed by either excluding DNA topoisomerase VI subunit A (arCOG04143) alone or together with DNA topoisomerase VI subunit B (arCOG01165) for the following reasons (see also supplementary information S1.1, Supplementary Material online):

1. The mixing of orthologues of arCOG04143 provides spurious support for the relationship between Njordarchaeia and Asgard archaea.
2. The interaction between DNA topoisomerase VI subunits might lead to the co-evolution of some amino acid residues.

Concatenated gene trees both with and without DNA topoisomerase VI subunits were inferred using *IQ-Tree* v2.1.2 (Minh et al. 2020) under the LG + C60 + G + F model with posterior mean sites frequency approximation (Wang et al. 2018) with guide trees inferred under LG + G + F model.

Concatenated Gene Trees Using Taxon Set 2: i.e. 303 taxa

To resolve the relationship between Njordarchaeales, Panguiarchaeales, Korarchaeota, and Asgard archaea, we next excluded divergent sequences from DPANN MAGs/genomes and downsampled the Thermoproteota (previously TACK), and Halobacteriota, Thermoplasmata, Hydrothermarchaeota, Methanobacteriota, Methanobacteriota_A, and Methanobacteriota_B lineages (previously “Euryarchaea”) based on the predicted genome completeness and contamination (completeness – 5 × contamination) (Parks et al. 2017) to create a dataset of 303 taxa by one-per-Thermoproteota/Asgardarchaeota-family and one-per-“Euryarchaea”-order criteria. Marker gene sequences were realigned with *mafft-linsi* v7.453 (Katoh and Standley 2013) and trimmed with *BMGE* 1.12 (Criscuolo and Grigaldo 2010) (settings: -m BLOSUM30 -h 0.55). We concatenated 3 supermatrices by including all top 50% ranked markers and excluding arCOG04143 alone or with arCOG01165. Trees were inferred under the LG + G + F and LG + C60 + G + F model, as well as LG + EDM0256LCLR + G + F model using *IQ-Tree* v2.1.2 (Minh et al. 2020) and *IQ-Tree* v2.2.2.7 (Minh et al. 2020), respectively (settings: -B 1,000 -alrt 1000, see also 71 taxa). Focusing on the supermatrix of 303 archaeal taxa, excluding DNA topoisomerase subunits, we progressively removed 10% to 50% of fastest evolving sites according to the empirical Bayesian rate under LG + C60 + G + F model (settings -rate), and 5%, 10% to

50% most heterogeneous sites using Alignment_pruner.pl (<https://github.com/novigitt/davinciCode/blob/master/perl>). Trees were inferred under either LG + C60 + G + F or LG + EDM0256LCLR + G + F models (settings: -B 1000 -alrt 1000).

Concatenated Gene Trees Using Taxon Set 3: i.e. 71 Taxa

A smaller dataset of 71 taxa was created based on the 303 taxa set (set 2) for Bayesian phylogenetic analyses. In particular, we kept one genome of “Euryarchaea” for each GTDB R207 class as an outgroup and one genome of thermoproteotal (TACK) lineages for each GTDB R207 (i.e. release 207) order, considering predicted genome completeness and contamination (completeness – 5 × contamination) (Parks et al. 2017). Marker gene sequences were realigned with mafft-linsi v7.453 (Katoh and Standley 2013) and trimmed with BMGE 1.12 (Criscuolo and Grigaldo 2010) (settings: -m BLOSUM30 -h 0.55). We concatenated only one supermatrix excluding both DNA topoisomerase subunits (arCOG04143 and arCOG01165). ML trees were inferred under LG + G + F and LG + C60 + G + F models using IQ-Tree v2.1.2 (Minh et al. 2020) (settings: -B 1000 -alrt 1000).

Additionally, we performed a comprehensive phylogenetic analysis using a recently developed CAT-PMSF pipeline (Szánthó et al. 2023). In particular, we used 3 different guide trees representing different hypotheses regarding the relationship of Njordarchaea to Korarchaeota and Asgard archaea, namely, Korarchaeota topology: Njordarchaea-sister-to-Korarchaeota; Asgard archaea topology: Njordarchaea-sister-to-Wukong-Heimdallarchaea and Asgard archaea sister topology: Njordarchaea-sister-to-Asgard-Archaea, to test if the used guide trees affect topology inference. We then used these 3 different starting topologies in a subsequent Bayesian analysis with the CAT model in combination with either Poisson or LG exchangeabilities and a 4 discrete categories gamma rate model. We ran 4 different Markov chains for each topology representing different hypotheses under the Poisson + CAT + G model in Phylobayes v 1.8 (Lartillot et al. 2013). Additionally, 4 different Markov chains were run for both Asgard archaea and Korarchaeota topology under the LG + CAT + G model. These chains were run until the effective sample size of all parameters was above 100 and/or the relative divergence of all parameters was below 0.3, or visual inspection with Tracer 1.7 (Rambaut et al. 2018) indicated convergence (supplementary data S17, Supplementary Material online). Next, we extracted posterior mean site-specific stationary distributions of amino acids for each chain using readpb_mpi (Lartillot et al. 2013), which were converted to site-specific frequency profiles using the python script convert-site-dist.py (<https://github.com/drenal/cat-pmsf-paper/blob/main/scripts/convert-site-dists.py>) (Szánthó et al. 2023). The site-specific frequency profile was then used to infer ML trees using IQ-Tree v2.1.2 (settings: -fs .sitefreq -B 1000 -alrt 1000).

We next applied the EDCluster algorithm (Schrempf et al. 2020) on one of the LG + CAT + G4 chains that yielded the highest likelihood under LG + CAT-PMSF model to infer EDM with various components (8, 16, 32, 64, 128, and 256) using log-centered log-ratio transformation (LCLR). ML trees were then inferred under different components of EDM models using IQ-Tree v2.2.2.7 (e.g. settings: -m LG + EDM0256LCLR + G + F -B 1000 -alrt 1000) and universal distribution mixture models (UDM) with 128, 256 and 512 components (Schrempf et al. 2020). The model fit between Poisson + CAT-PMSF, LG + CAT-PMSF, LG + G + F, and LG + C60 + G + F was assessed using a recently developed

parametric bootstrap method (Giacomelli et al. 2025), which compares the model adequacy in describing across-site compositional heterogeneity.

Compositional Constrained Analyses

Constrained searches were performed for Asgard archaea (-sister) and Korarchaeota topology based on the unconstrained ML tree under the same model using IQ-Tree v2.1.2 (Minh et al. 2020) via the -g and -wslr flags for both streamlined set 2 (303 taxa) and 3 (71 taxa). AU tests were performed using consel v 1.20 (Shimodaira 2001) with 10,000 bootstrap replicates. Site-wise log-likelihood differences were calculated for Asgard archaea and Korarchaeota topology under LG + G + F, LG + C60 + G + F, Poisson + CAT-PMSF, and LG + CAT-PMSF models to investigate the attraction of Njordarchaea toward Korarchaeota using the streamlined set 3 (71 taxa). The effective amino acids (K_{eff}) values of mixture models and site-specific frequency profiles were calculated using the convert-site-dists-to-k_eff.py script (Schrempf et al. 2020; Szánthó et al. 2023). The site-wise log-likelihood differences were then grouped into different bins according to partitions, evolving rates, and the effective amino acids.

RED Values Scaling and Classification of Ranks

Relative evolutionary distance (RED) metric on the species trees was calculated with methods described in the PhyloRank v1.1.12 tool (Parks et al. 2018). Class, family, and order rank nodes in the 303 archaeal sets were extracted from our phylogenetic tree based on LG + EDM0256LCLR + G + F model, as well as from the GTDB R220 reference tree (Rinke et al. 2021). Linear regression analysis of the RED values of our phylogenetic tree and the GTDB R220 tree was performed using the Scipy package. The taxonomic boundaries used in GTDB R220 were scaled onto our tree by the fitted slope and intercept.

Amino Acid Composition

We used a custom Python script to estimate the relative frequency of each amino acid in our 71 taxa dataset for the concatenation without DNA topoisomerase subunits (<https://doi.org/10.5281/zenodo.15848218>). A principal component analysis (PCA) was performed using Scikit-learn (Pedregosa et al. 2011), and 2 axes were plotted along with eigenvectors using Seaborn (Waskom 2021). Njordarchaea and Korarchaeota were found to share a similar composition along axis 2, which seemed to be associated with the presence of reverse gyrase, an indicator of a thermophilic lifestyle. Hence, using gene composition separated by axis 2, we identified RYPEI(WKVL)/QSNT(DGC) as thermophilic-enriched/depleted amino acid residues in the concatenation using a binomial test (Baker et al. 2024) and modified the GFmix (Muñoz-Gómez et al. 2022) model to account for the variation of the ratio RYPEI(WKVL)/QSNT(DGC) at every branch. The likelihood of different tree topologies under the GFmix-RYPEI(WKVL)/QSNT(DGC) model was then calculated with LG + C60 + G + F and LG + EDM0256LCLR + G + F, with weights of the mixture model, branch length, and alpha shape parameters estimated using IQ-Tree v2.1.2 (Minh et al. 2020).

Gene Calling and Annotations

We annotated 966 archaeal genomes/MAGs using an in-house annotation pipeline (<https://doi.org/10.5281/zenodo.15848218>). In brief: Gene calling was performed using Prokka (Seemann 2014) (v1.14.6, settings: -kingdom

Archaea –addgenes –increment 10 –compliant –centre UU –norrna –notrna). The generated protein files were searched against COG database (Galperin et al. 2021) (NCBI_COGs_Oct2020.hmm), arCOG database (Makarova et al. 2015) (All_Arcogs_2018.hmm), PFAM (Bateman et al. 2004) (Release 34.0), TIGRFAM (Release 15.0), KEGG orthology (Aramaki et al. 2020) (KO) profile (downloaded April, 2019), the Carbohydrate-Active enZymes (CAZy) database (Lombard et al. 2014) (downloaded from dbCAN2 in September 2019), the Transporter Classification Database (Saier et al. 2021) (TCDB; downloaded in November 2018), the hydrogenase database (Søndergaard et al. 2016) (HydDB; downloaded in November 2018), the MERPOS database (Rawlings et al. 2018) (Release 12.4) and NCBI_nonredundant database (NCBI_nr; downloaded in November 2018). Additionally, we used Interproscan version 5.61.93.0 to scan for protein domains (Jones et al. 2014) (settings: –iplookup –goterms).

COG (settings: –E 1e–5), arCOG (settings: –E 1e–5), PFAM (settings: –E 1e–5), TIGRFAM (settings: –E 1e–5), KO (settings: –E 1e–5), and CAZy (settings: –E 1e–10) identifiers were assigned using hmmsearch v3.3.2. The TCDB, HydDB, and MERPOS database were searched using BLASTp v2.12.0 (Altschul et al. 1997) (settings: –outfmt 6, –evalue 1e–20), and NCBI_nr database was searched using DIAMOND v2.0.6 (Buchfink et al. 2015) (settings: –more-sensitive –e-value 1e–5 –seq 50 –no-self-hits –taxonmap prot.accession2taxid.gz). Asgard COG (Liu et al. 2021) (asCOG) database was searched with psi-blast v2.12.0 (settings: –show_gis –outfmt 6 –dbsize 100000000 –comp_based_stats F –seg no –evalue 1e–20). The best hit of each protein for all other database searches was selected based on the lowest e-value and highest score, and results were summarized for Njordarchaeales MAGs. Additionally, 1,325 bacterial genomes/MAGs and 137 eukaryotic genomes and/or largely complete transcriptomes (Santana-Molina et al. 2025) (supplementary data S18, Supplementary Material online) were searched against COG, arCOG, PFAM, and KO databases using the same pipeline. These annotations were also the basis for identifying homologs for selected single gene trees (see below). Njordarchaeia MAGs annotation results were summarized in supplementary data S7 to S11 and S19 to S21, Supplementary Material online.

Metabolic Comparison

Metabolic comparisons were based on the annotation results described above. We counted the occurrence of each gene across each MAG/genome and reported the presence and absence pattern using Pandas (The Pandas Development Team 2024) in Python v3.9.0 (Van Rossum and Drake 2009). These data were then summarized per GTDB R207 class level whenever possible. The occurrence of ESPs was compared using the results from asCOG searches. The presence and absence pattern of genes across each database (COG_DB and ARCOG_DB) was analyzed with PCA and t-distributed Stochastic Neighbor Embedding (t-SNE) using scikit-learn (Pedregosa et al. 2011).

Contamination Assessment of Representative Njordarchaeia MAGs and Its Impact on Phylogenetic Analyses

Previously published metagenomes (see Fig. 2) of the representative MAGs were downloaded from the NCBI sequence read

archive (SRA) database (Dombrowski et al. 2018; Qu et al. 2023). These metagenomes include samples from hydrothermal vents in Guaymas Basin (Dombrowski et al. 2018) and a hot spring in Yunan, China (Qu et al. 2023). We used bowtie2 v2.5.3 (Langmead and Salzberg 2012) to map reads to the MAGs, and Samtools v1.18 to process the mapped profiles. Anvi'o v8 (Eren et al. 2021) was used to perform hierarchical clustering based on sequence composition and coverage patterns across different metagenomes. Clusters were identified based on the similarity of coverage profile across different samples while taking into account sample similarity. Marker genes used in our phylogenetic analyses and homologs from marker sets in a recent study (Zhang et al. 2025) were mapped back to the contigs using sequences selected for phylogenetic analyses and annotations described above.

Single Gene Tree Inferences for Ancestral Reconstruction

The predicted protein sequence of 303 archaeal genomes/MAGs was searched against the arCOG database using hmmsearch v3.3.2 (Finn et al. 2011) (settings: –tblout –domtblout –notextw). We used the best hits (the lowest e-value and highest bitscore, cutoff: e-value < 1E–3) to identify and split potential fused proteins. Specifically, we first subtracted the position of the first domain hit from the full protein using bedtools subtract (Quinlan and Hall 2010) (v2.26.0). Second, we investigated the presence of a secondary domain hit that is assigned to a different arCOGs than the first domain. Subsequently, we repeated these two steps until all the proteins were investigated for up to 4 domains. Finally, the individual domains from the original protein were extracted using the positional domain information from hmmsearch output using bedtools getfasta function. To keep track of unsplit and split domains from the original protein, we assigned “a0” notation for unsplit proteins and “a1” to “a4” for split domains (primarily to quarterly) in the generated trees.

We then combined all protein sequences and removed sequences with ambiguous amino acids (i.e. X) using a custom Python script (remove_seq_with_specific_char.py, <https://doi.org/10.5281/zenodo.15848218>). To improve the phylogenetic signal in the single gene trees, we carefully performed 3 rounds of alignments before the final phylogenetic inference. First, we aligned sequences using mafft or mafft-linsi (Katoh and Standley 2013), for sequences with greater than or less than 1,000 sequences, respectively. Alignments were then trimmed using trimAl 1.2rev59 (Capella-Gutiérrez et al. 2009) with –gappyout flag. Trimmed sequences with greater than or equal to 50% gaps were removed using a custom Python script (faa_drop.py, <https://doi.org/10.5281/zenodo.15848218>). Second, we realigned the arCOG gene families (with more than or equal to 4 sequences) using MAFFT as described above and trimmed the alignment with BMGE (v1.12; settings: –m BLOSUM30 –b 2 –h 0.55) (Criscuolo and Gribaldo 2010). We then inferred a phylogenetic tree using IQ-Tree v2.1.2 (settings: –m LG + G) (Minh et al. 2020). To alleviate the effect of long-branching sequences on phylogenetic inferences, we identified sequences/clusters of sequences on long branches of the inferred tree using a custom script (cut_gene_tree.py, <https://doi.org/10.5281/zenodo.15848218>) with a branch cutoff of 2 (Davín et al. 2025). Single sequences were discarded, and clusters with more than 4 sequences were separated into different gene families. We appended the notation “_ [0-3]” to keep track of the separation of sequences into a

new cluster. Finally, we aligned and trimmed the sequences as described in the second step. For each family, we selected the best-fit model using the model test with IQ-Tree v2.1.2 (settings: -m MF -mset LG -madd LG + C10, LG + C10 + G, LG + C10 + R, LG + C10 + F, LG + C10 + R + F, LG + C10 + G + F, LG + C20, LG + C20 + G, LG + C20 + F, LG + C20 + G + F, LG + C20 + R, LG + C20 + R + F, LG + C30, LG + C30 + G, LG + C30 + R, LG + C30 + F, LG + C30 + R + F, LG + C30 + G + F -score-diff all -T 2). For the selected best-fitting model with profile mixtures, we used the posterior mean site frequency (PMSF) approximation (Wang et al. 2018) to infer the bootstrap tree distribution (settings: -s trimmed_aln.faa -m BEST-FITTING MODEL -T 2 -wbtl -B 1000 -pers 0.2 -nstop 500) using a guide tree inferred from the LG + G model (settings: -T 2) (Minh et al. 2020). For the best-fitting non-mixture models, bootstrap tree distribution was inferred using IQ-Tree v2.1.2 with the settings: -T 1 -wbtl -B 1000 -pers 0.2 -nstop 500. In total, we generated 6,077 single gene tree distributions in the analysis for ancestral reconciliations.

Gene Tree and Species Tree Reconciliations

We used the amalgamated likelihood estimation (ALE, v1.0) approach to reconcile single gene trees against the species tree. First, we generated the ALE objects from the bootstrap trees that approximate tree uncertainty using ALEobserve. Using genome completeness estimation from CheckM v2 (Chklovski et al. 2023) and ALEml_undated (Szöllösi et al. 2015) program, we first reconciled the gene trees against the species tree using the default parameters, which assume origination at any node of the species tree was uniform. Based on the reconciliation results from this, we next used an additional reconciliation approach described by Coleman et al. (2021) that assumes the origination probabilities at the root of the species tree (O_R) are different to those for all other internal nodes. We first inferred the O_R for each of the 21 arCOG functional categories by maximizing the total reconciliation likelihood over all gene families in that category using 2 Python scripts (setup_OR_estimation.py and O_R_optimisation.py, <https://doi.org/10.5281/zenodo.15848218>). Subsequently, we used the category-specific probabilities for origination at the root to reconcile the single gene trees and species trees and estimated the presence probabilities/copies of each family in the internal nodes of the tree. The presence probabilities of both approaches for each family for the nodes of interest are provided in [supplementary data S13, Supplementary Material](#) online. Ancestral genome size was predicted using locally weighted smoothing regression based on the relationship between the genome size of extant taxa and their arCOG gene family number. The uncertainty of the prediction of genome size was taken into account using the bootstrap resampling approach (1,000 times).

Phylogenetic Analyses of DNA Topoisomerase Subunit A and DNA Topoisomerase B

The homologs of DNA topoisomerase subunit A and B, i.e. COG1389 and COG1697, were identified and retrieved from 966 archaeal, 1,325 bacterial, and 137 eukaryotic genomes/MAGs or largely complete transcriptomes based on the output from the annotation results described above (see section on [Gene Calling and Annotations](#)). Sequences of each gene family were first aligned with mafft-linsi v7.453 (Katoh and Standley 2013), and poorly aligned alignment sites were trimmed with BMGE v1.2 (Criscuolo and Gribaldo 2010) (Settings: -m BLOSUM30 -h 0.55). The inference of

phylogenetic trees for individual maker gene families was conducted as follows: an initial ML tree was inferred under LG + G with IQ-Tree v 2.1.2 (Minh et al. 2020). Single gene trees were manually inspected to remove putative LBA artifacts, and the remaining sequences for each gene were realigned and trimmed as described above; finally, an ML tree was inferred for each gene under LG + C60 + G + F + PMSF (Wang et al. 2018) model (settings: -B 1000 -alrt 1000) using guide tree inferred from LG + F + G.

Phylogenetic Analyses of Actin Homologs

The actin homologs were identified from 966 archaeal, 1,325 bacterial, and 137 eukaryotic genomes/MAGs or largely complete transcriptomes by extracting all proteins with the domain IPR00400 as predicted by our annotation pipeline (see section on [Gene Calling and Annotations](#)). The cell shape-determining protein MreB homologs, a distant homolog to actin, in 966 archaeal genomes were identified using COG1077 and dereplicated using cd-hit v4.7 (-c 0.75) (Fu et al. 2012). These sequences were combined and aligned with mafft-linsi v7.453 (Katoh and Standley 2013). Poorly aligned sequences were trimmed using BMGE (Criscuolo and Gribaldo 2010) (settings: -m BLOSUM30 -h 0.55). An initial ML tree was inferred under LG + G using IQ-Tree v 2.1.2 (Minh et al. 2020), and sequences showing potential instances of LBA artifacts were removed; the remaining sequences were then realigned and trimmed as described above; finally, an ML tree was then inferred under LG + C60 + G + F + PMSF model (settings: -B 1,000 -alrt 1000) using a guide tree inferred from LG + F + G. 20% of the most heterogeneous sites were removed using alignment_pruner.pl and an ML tree was inferred in the same way under LG + C60 + G + F + PMSF using a guide tree inferred from LG + F + G.

Phylogenetic Analyses of Ribosomal Protein L28

The ribosomal protein L28/Mak16 homologs (i.e. PF01778) were retrieved from 966 archaeal, 1,325 bacterial, and 137 eukaryotic genomes/MAGs or largely complete transcriptomes (e-value: 0.0001) by using the PFAM profiles as a query to search gene sequences of these taxa with hmmsearch v3.3.2. Sequences assigned to PF01778 were extracted. Please note that we slightly relaxed our e-value cutoff here compared to our annotation pipeline due to low sequence conservation of members of this family. Sequences were first aligned with mafft-linsi v7.453 (Katoh and Standley 2013), and gappy columns were removed using trimal 1.2rev59 (Capella-Gutiérrez et al. 2009) with -gappypout flag. An initial ML tree was inferred under LG + G using IQ-Tree v2.1.2 (Minh et al. 2020), and sequences showing potential instances of LBA artifacts were removed; remaining sequences were then realigned and trimmed as described above; finally, an ML tree was inferred under LG + C60 + G + F + PMSF model (settings: -B 1000 -alrt 1000) using the guide tree inferred from LG + G.

Phylogenetic Analyses of Oligosaccharyltransferase 3/6 Like (OST3/6 Like) and Translocon-associated Protein (TRAP) Complexes Subunit Beta

The OST3/6-like homologs and TRAP beta were retrieved from our 966 archaeal, 1,325 bacterial, and 137 eukaryotic genomes/MAGs or largely complete transcriptomes by extracting all proteins with PF04756 and PF05753 domains as inferred by our annotation pipeline, respectively (see section on [Gene Calling and Annotations](#)). These sequences were dereplicated using cd-hit v4.7 (Fu et al. 2012) (-c 0.7). Sequences

of each gene family were first aligned with mafft-linsi v7.453 (Katoh and Standley 2013), and gappy columns were removed using trimal 1.2rev59 (Capella-Gutiérrez et al. 2009) with -gappypout flag. An initial ML tree was inferred under LG + G using IQ-Tree v2.1.2 (Minh et al. 2020) for each gene family, and sequences showing potential instances of LBA artifacts were removed; remaining sequences for each gene family were then realigned and trimmed as described above; finally an ML tree was then inferred for each gene under LG + C60 + G + F + PMSF model (settings: -B 1000 -alrt 1000) using guide tree inferred from LG + G.

Phylogenetic Analyses of the Snf7 Family

The Snf7 family homologs were identified and retrieved from 966 archaeal and 137 eukaryotic genomes/MAGs by extracting proteins with the presence of the domain IPR005024 as inferred using our annotation pipeline (see section on [Gene Calling and Annotations](#)). Eukaryotic sequences were dereplicated using cd-hit v4.7 (Capella-Gutiérrez et al. 2009) (-c 0.7). We created 2 datasets for this family, i.e. a dataset with only archaeal sequences and another with archaeal and eukaryotic sequences. Sequences of each dataset were first aligned with mafft-linsi v7.453 (Katoh and Standley 2013), and poorly aligned sites were removed using BMGE v1.1.2 (Criscuolo and Gribaldo 2010) (settings: -m BLOSUM30 -h 0.55). The inference of phylogenetic trees was conducted as follows: an initial ML tree was inferred under LG + G using IQ-Tree v2.1.2 (Minh et al. 2020), and sequences showing potential instances of LBA artifacts were removed; remaining sequences were then realigned and trimmed as described above; finally, an ML tree was then inferred under the LG + C60 + G + F + PMSF model (settings: -B 1000 -alrt 1000) using a guide tree inferred from LG + G. We also removed the 20% most heterogeneous sites from the alignment with both archaeal and eukaryotic sequences included using Alignment_pruner.pl (<https://github.com/novigit/davinciCode/blob/master/per1>) and inferred an ML tree using the same approach.

Phylogenetic Analyses of CDP-archaeol Synthase (CarS), Archaetidylinositol Phosphate Synthase (PgsA), and Phosphatidylserine Synthase (PssA)

Sequences identified as CarS (K19664), PgsA (K17884), and PssA (K17103) were retrieved from 966 archaeal, 1,325 bacterial, and 137 eukaryotic genomes/MAGs or largely complete transcriptomes, respectively, as inferred from our annotation pipeline (see section on [Gene Calling and Annotations](#)). Sequences of each gene family were first aligned with mafft-linsi v7.453 (Katoh and Standley 2013), and gappy columns were removed using trimal 1.2rev59 (Capella-Gutiérrez et al. 2009) with the -gappypout flag. The inference of phylogenetic tree for individual gene families was conducted as follows: an initial ML tree was inferred under LG + G using IQ-Tree v2.1.2 (Minh et al. 2020) for each gene family, and sequences showing potential instance of LBA artifacts were removed; remaining sequences for each gene were then realigned and trimmed as described above; finally an ML tree was then inferred for each gene under LG + C60 + G + F + PMSF model (settings: -B 1000 -alrt 1000) using guide tree inferred from LG + G.

Phylogenetic Analyses of [NiFe]-hydrogenase, Large Subunit

Backbone sequences for phylogenetic analyses of [NiFe] group 3 and 4 hydrogenases were obtained from a previous

study (Spang et al. 2019). Njordarchaeal homologs were identified based on the output of the hydrogenase database (Søndergaard et al. 2016) in our annotation pipeline (see section on [Gene Calling and Annotations](#)) and further refined using the online classifier (<http://services.birc.au.dk/hyddb>). Njordarchaeal homologs were added to the backbone sequences and aligned with mafft-linsi v7.453 (Katoh and Standley 2013). Poorly aligned sites were removed using BMGE v1.1.2 (Criscuolo and Gribaldo 2010) (settings: -m BLOSUM30 -h 0.55). A phylogenetic tree was inferred under the LG + C60 + G + F + PMSF model guide trees inferred from the LG + G + F model using IQ-Tree v2.1.2 (Minh et al. 2020) (settings: -B 1000 -alrt 1000). The phylogenetic tree was visualized using ggtree 3.2.0 (Xu et al. 2022).

Network Inferences

The input dataset for network prediction was obtained from OTU community profiles derived from 248,559 NCBI metagenomes from the “Sandpiper” website (<https://sandpiper.qut.edu.au>) (Woodcroft et al. 2025), which were annotated with the GTDB R214 taxonomy (Parks et al. 2022). First, we used the known association between *Huberarchaeum* and *Altiarchaeum* to confirm that our network association approach is able to detect symbiotic interactions. We selected all samples that contain taxa annotated as *Huberarchaeum* (67 samples), respectively. We transformed the coverage of the taxa in these samples to counts by multiplying by 100 and formatted the data into a table listing files in rows and all OTUs in columns. To reduce the scarcity of the dataset, we filtered for OTUs present in 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90% of the samples. To estimate the minimum number of samples required for robust, true positive prediction of *Huberarchaeum*-*Altiarchaeum* associations, we drew random samples and applied two different filter criteria. Numbers of 10, 20, 30, 40, 50, and 60 samples were drawn; the random data sets were filtered for OTUs detected in at least 50% and 80% of all random samples, respectively. For each condition, ten networks were inferred. Statistics were run across all ten networks within one condition. These networks were inferred using SpiecEasi v1.1.3 (Kurtz et al. 2015). Across all networks, we checked for associations between *Huberarchaeum* and *Altiarchaeum*. In all networks inferred (>30 samples), the expected association was robustly detected. However, below a number of 30 samples, the association was detected only in 50% of the samples. This clearly indicates a minimum number of >30 samples (equaling 40% of samples) that is necessary for reliable inference of association.

We also attempted to confirm the known association between *Nanoarchaeum* and its host *Ignicoccus* (order Sulfolobales). Our dataset contained several OTUs, assigned to the genus *Nanoarchaeum* (otu_658, otu_3077) and the genus *Ignicoccus* (otu_3007, otu_678, otu_3062, otu_3063, otu_3922, otu_4367). However, some of the taxa were present in only very few samples (supplementary figs. S35 and S36, Supplementary Material online), and subsequently in the inferred network, no association between any of these taxa was predicted. Interestingly, some OTUs showed strong, significant negative correlations (supplementary figs. S35 and S36, Supplementary Material online), suggesting a high specificity of the symbiotic association on the OTU level, although larger datasets are needed to confirm this assumption.

Next, we selected all samples that contain taxa annotated as “*Panguiarchaeaceae*” in GTDB R220 (Rinke et al. 2021;

Parks et al. 2022), a taxonomic rank that is equivalent to the class Njordarchaeia/order Njordarchaeales proposed in this manuscript (Fig. 1b and c), resulting in a subset of 120 samples. To reduce the scarcity of the dataset, we filtered the dataset before network inference. We used 5 different occupancy thresholds based on shoulder theory and commonly applied criteria to prepare 5 different datasets from which separate networks were inferred. OTUs had to be present in (1) at least 24 samples as well as in at least (2) 40%, (3) 50%, (4) 60%, and (5) 70% of the entire dataset to be included in the network. Across all networks, we checked for associations involving OTUs of our target lineage in a comparative manner. To exclude random associations involving our target lineage, we prepared random networks and selected only those associations that were unlikely to be found randomly. For each filter criterion, 999 random networks were drawn based on the original network inferred from the filtered data. Then, we compared the taxonomic distribution of OTUs associated with OTUs annotated as Njordarchaeales (equivalent to “Panguiarchaeaceae” in GTDB R220) between the random networks and the original network. Association frequencies were compared based on phylum, class, order, and family level. Next, we explored associations at the genus level. A single OTU within the genus *Panguiarchaeum* was detected in higher numbers across samples, that is OTU_804 (43 samples), for which we investigated association partners. All analyses were performed in R v4.4.1 (R Core Team 2024).

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

We also want to thank Tara Mahendrarajah for the valuable feedback on the manuscript.

Author Contributions

W.H., T.A.W., and A.S. conceived the study; W.H. analyzed and annotated genomes and generated accompanying data tables; W.H. and A.S. analyzed genomic data; W.H., T.A.W., and A.S. performed and analyzed phylogenetic analyses; C.R. and M.P. performed co-proportionality analysis; W.H., Z.H., T.W., C.R., M.P., G.J.S., L.S., T.J.G.E., and A.S. interpreted data; W.H., T.W., and A.S. wrote and all authors edited and approved the manuscript.

Funding

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement No. 947317, ASymbEL to A.S., grant agreement No. 714774, GENECLOCKS to G.J.S., and grant agreement No. 101142180, DARKROOTS to T.J.G.E.). Furthermore, this work was supported by the Gordon and Betty Moore Foundation (GBMF9741 to T.A.W., A.S., and G.J.S.), the Dutch Research Council (grant agreement No. VI.C.192.016 to T.J.G.E.) and the Simons Foundation with the Moore-Simons Project on the Origin of the Eukaryotic Cell, 735929LPI (<https://doi.org/10.46714/735929LPI>) (to A.S. and coworkers). Our research is funded by the John Templeton Foundation (63451 to L.Sz., G.J.Sz., T.A.W.,

and A.S.; the opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

Conflict of Interest

The authors declare no competing interests.

Data Availability

All genomic data/transcriptomic data analyzed in this study are available at NCBI (supplementary data S16 and S18, Supplementary Material online) and are deposited in the data repository (<https://doi.org/10.5281/zenodo.15848218>). Data generated in this study including single gene and concatenated phylogenies (i.e. sequence files, alignments, and treefiles) have also been deposited in our data repository at Zenodo (<https://doi.org/10.5281/zenodo.15848218>) under the following license CC BY 4.0. Workflows for annotations, phylogenies, reconciliation, and custom scripts (bash/python) to analyze and parse annotation data, generate single gene trees, and perform ancestral gene content reconstruction are deposited in the repository (<https://doi.org/10.5281/zenodo.15848218>).

References

- Akil C, Robinson RC. Genomes of Asgard archaea encode profilins that regulate actin. *Nature*. 2018;562(7727):439–443. <https://doi.org/10.1038/s41586-018-0548-6>.
- Akil C, Tran LT, Orhant-Prioux M, Baskaran Y, Manser E, Blanchoin L, Robinson RC. Insights into the evolution of regulated actin dynamics via characterization of primitive gelsolin/cofilin proteins from Asgard archaea. *Proc Natl Acad Sci U S A*. 2020;117(33):19904–19913. <https://doi.org/10.1073/pnas.2009167117>.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389–3402. <https://doi.org/10.1093/nar/25.17.3389>.
- Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics*. 2020;36(7):2251–2252. <https://doi.org/10.1093/bioinformatics/btz859>.
- Baker BA, Gutiérrez-Preciado A, Rodríguez del Río Á, McCarthy CGP, López-García P, Huerta-Cepas J, Susko E, Roger AJ, Eme L, Moreira D. Expanded phylogeny of extremely halophilic archaea shows multiple independent adaptations to hypersaline environments. *Nat Microbiol*. 2024;9(4):964–975. <https://doi.org/10.1038/s41564-024-01647-4>.
- Baker BJ, De Anda V, Seitz KW, Dombrowski N, Santoro AE, Lloyd KG. Diversity, ecology and evolution of Archaea. *Nat Microbiol*. 2020;5(7):887–900. <https://doi.org/10.1038/s41564-020-0715-z>.
- Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer ELL, et al. The Pfam protein families database. *Nucleic Acids Res*. 2004;32(90001):138D–D141. <https://doi.org/10.1093/nar/gkh121>.
- Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloef-Fadrosch EA, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol*. 2017;35(8):725–731. <https://doi.org/10.1038/nbt.3893>.
- Brown AM, Hoopes SL, White RH, Sarisky CA. Purine biosynthesis in archaea: variations on a theme. *Biol Direct*. 2011;6(1):63. <https://doi.org/10.1186/1745-6150-6-63>.
- Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12(1):59–60. <https://doi.org/10.1038/nmeth.3176>.

- Cambillau C, Claverie JM. Structural and genomic correlates of hyperthermostability. *J Biol Chem*. 2000;275(42):32383–32386. <https://doi.org/10.1074/jbc.C000497200>.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25(15):1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>.
- Castelle CJ, Brown CT, Anantharaman K, Probst AJ, Huang RH, Banfield JF. Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. *Nat Rev Microbiol*. 2018;16(10):629–645. <https://doi.org/10.1038/s41579-018-0076-2>.
- Chen L-X, Anantharaman K, Shaiber A, Eren AM, Banfield JF. Accurate and complete genomes from metagenomes. *Genome Res*. 2020;30(3):315–333. <https://doi.org/10.1101/gr.258640.119>.
- Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. Checkm2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods*. 2023;20(8):1203–1212. <https://doi.org/10.1038/s41592-023-01940-w>.
- Coleman GA, Davin AA, Mahendrarajah TA, Szánthó LL, Spang A, Hugenholtz P, Szöllösi GJ, Williams TA. A rooted phylogeny resolves early bacterial evolution. *Science*. 2021;372(6542):eabe0511. <https://doi.org/10.1126/science.abe0511>.
- Criscuolo A, Gribaldo S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol*. 2010;10(1):210. <https://doi.org/10.1186/1471-2148-10-210>.
- Davin AA, Woodcroft BJ, Soo RM, Morel B, Murali R, Schrepf D, Clark JW, Álvarez-Carretero S, Boussau B, Moody ERR, et al. A geological timescale for bacterial evolution and oxygen adaptation. *Science*. 2025;388(6742):eadp1853. <https://doi.org/10.1126/science.adp1853>.
- Dombrowski N, Teske AP, Baker BJ. Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nat Commun*. 2018;9(1):4999. <https://doi.org/10.1038/s41467-018-07418-0>.
- Dombrowski N, Williams TA, Sun J, Woodcroft BJ, Lee J-H, Minh BQ, Rinke C, Spang A. Undinarchaeota illuminate DPANN phylogeny and the impact of gene transfer on archaeal evolution. *Nat Commun*. 2020;11(1):3939. <https://doi.org/10.1038/s41467-020-17408-w>.
- Elkins JG, Podar M, Graham DE, Makarova KS, Wolf Y, Randau L, Hedlund BP, Brochier-Armanet C, Kunin V, Anderson I, et al. A korarchaeal genome reveals insights into the evolution of the archaea. *Proc Natl Acad Sci U S A*. 2008;105(23):8102–8107. <https://doi.org/10.1073/pnas.0801980105>.
- Eme L, Spang A, Lombard J, Stairs CW, Ettema TJG. Archaea and the origin of eukaryotes. *Nat Rev Microbiol*. 2017;15(12):711–723. <https://doi.org/10.1038/nrmicro.2017.133>.
- Eme L, Tamarit D, Caceres EF, Stairs CW, De Anda V, Schön ME, Seitz KW, Dombrowski N, Lewis WH, Homa F, et al. Inference and reconstruction of the heimdallarchaeal ancestry of eukaryotes. *Nature*. 2023;618(7967):992999. <https://doi.org/10.1038/s41586-023-06186-2>.
- Eren AM, Kiehl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, et al. Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol*. 2021;6(1):3–6. <https://doi.org/10.1038/s41564-020-00834-3>.
- Ettema TJG. Evolution: mitochondria in the second act. *Nature*. 2016;531(7592):39–40. <https://doi.org/10.1038/nature16876>.
- Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 2011;39(Suppl):W29–W37. <https://doi.org/10.1093/nar/gkr367>.
- Forterre P. A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein. *Trends Genet*. 2002;18(5):236–237. [https://doi.org/10.1016/S0168-9525\(02\)02650-1](https://doi.org/10.1016/S0168-9525(02)02650-1).
- Freibert S-A, Goldberg AV, Hacker C, Molik S, Dean P, Williams TA, Nakjang S, Long S, Sendra K, Bill E, et al. Evolutionary conservation and in vitro reconstitution of microsporidian iron–sulfur cluster biosynthesis. *Nat Commun*. 2017;8(1):1–12. <https://doi.org/10.1038/ncomms13932>.
- Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012;28(23):3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>.
- Galperin MY, Wolf YI, Makarova KS, Vera Alvarez R, Landsman D, Koonin EV. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res*. 2021;49(D1):D274–D281. <https://doi.org/10.1093/nar/gkaa1018>.
- Giacomelli M, Vecchi M, Guidetti R, Rebecchi L, Donoghue PCJ, Lozano-Fernandez J, Pisani D. CAT-posterior mean site frequencies improves phylogenetic modeling under maximum likelihood and resolves tardigrada as the sister of Arthropoda plus onychophora. *Genome Biol Evol*. 2025;17(1):evae273. <https://doi.org/10.1093/gbe/evae273>.
- Glasemacher J, Bock AK, Schmid R, Schönheit P. Purification and properties of acetyl-CoA synthetase (ADP-forming), an archaeal enzyme of acetate formation and ATP synthesis, from the hyperthermophile *Pyrococcus furiosus*. *Eur J Biochem*. 1997;244(2):561–567. <https://doi.org/10.1111/j.1432-1033.1997.00561.x>.
- Guy L, Ettema TJG. The archaeal “TACK” superphylum and the origin of eukaryotes. *Trends Microbiol*. 2011;19(12):580–587. <https://doi.org/10.1016/j.tim.2011.09.002>.
- Guy L, Saw JH, Ettema TJG. The archaeal legacy of eukaryotes: a phylogenomic perspective. *Cold Spring Harb Perspect Biol*. 2014;6(10):a016022. <https://doi.org/10.1101/cshperspect.a016022>.
- Hatano T, Palani S, Papatziomou D, Salzer R, Souza DP, Tamarit D, Makwana M, Potter A, Haig A, Xu W, et al. Asgard archaea shed light on the evolutionary origins of the eukaryotic ubiquitin-ESCRT machinery. *Nat Commun*. 2022;13(1):3398. <https://doi.org/10.1038/s41467-022-30656-2>.
- The Pandas Development Team. 2024. Pandas-dev/pandas: Pandas. <https://doi.org/10.5281/zenodo.3509134>.
- Huang W-C, Liu Y, Zhang X, Zhang C-J, Zou D, Zheng S, Xu W, Luo Z, Liu F, Li M. Comparative genomic analysis reveals metabolic flexibility of Woesearchaeota. *Nat Commun*. 2021;12(1):1–14. <https://doi.org/10.1038/s41467-021-25565-9>.
- Huber H, Burggraf S, Mayer T, Wyszchony I, Rachel R, Stetter KO. *Ignicoccus* gen. nov., a novel genus of hyperthermophilic, chemolithoautotrophic Archaea, represented by two new species, *Ignicoccus islandicus* sp nov and *Ignicoccus pacificus* sp nov. and *Ignicoccus pacificus* sp. nov. *Int J Syst Evol Microbiol*. 2000;50(6):2093–2100. <https://doi.org/10.1099/00207713-50-6-2093>.
- Huber H, Hohn MJ, Rachel R, Fuchs T, Wimmer VC, Stetter KO. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature*. 2002;417(6884):63–67. <https://doi.org/10.1038/417063a>.
- Hurtig F, Burgers TCQ, Cezanne A, Jiang X, Mol FN, Traparić J, Pulschen AA, Nierhaus T, Tarrason-Risa G, Harker-Kirschneck L, et al. The patterned assembly and stepwise Vps4-mediated disassembly of composite ESCRT-III polymers drives archaeal cell division. *Sci Adv*. 2023;9(11):eade5224. <https://doi.org/10.1126/sciadv.ade5224>.
- Imachi H, Nobu MK, Kato S, Takaki Y, Miyazaki M, Miyata M, Ogawara M, Saito Y, Sakai S, Tahara YO, et al. *Promethearchaeum syntrophicum* gen. nov., sp. nov., an anaerobic, obligately syntrophic archaeon, the first isolate of the lineage “Asgard” archaea, and proposal of the new archaeal phylum Promethearchaeota phyl. nov. and kingdom Promethearchaeati regn. nov. *Int J Syst Evol Microbiol*. 2024;74(7):006435. <https://doi.org/10.1099/ijsem.0.006435>.
- Imachi H, Nobu MK, Nakahara N, Morono Y, Ogawara M, Takaki Y, Takano Y, Uematsu K, Ikuta T, Ito M, et al. Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature*. 2020;577(7791):519–525. <https://doi.org/10.1038/s41586-019-1916-6>.
- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30(9):1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14(6):587–589. <https://doi.org/10.1038/nmeth.4285>.

- Kapli P, Flouri T, Telford MJ. Systematic errors in phylogenetic trees. *Curr Biol*. 2021;31(2):R59–R64. <https://doi.org/10.1016/j.cub.2020.11.043>.
- Kato S, Ogasawara A, Itoh T, Sakai HD, Shimizu M, Yuki M, Kaneko M, Takashina T, Ohkuma M. *Nanobdella aerobiophila* gen. nov., sp. nov., a thermoacidophilic, obligate ectosymbiotic archaeon, and proposal of Nanobdellaceae fam. nov., Nanobdellales ord. nov. and Nanobdellia class. nov. *Int J Syst Evol Microbiol*. 2022;72(8):005489. <https://doi.org/10.1099/ijsem.0.005489>.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–780. <https://doi.org/10.1093/molbev/mst010>.
- Koonin EV, Yutin N. The dispersed archaeal eukaryome and the complex archaeal ancestor of eukaryotes. *Cold Spring Harb Perspect Biol*. 2014;6(4):a016188. <https://doi.org/10.1101/cshperspect.a016188>.
- Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput Biol*. 2015;11(5):e1004226. <https://doi.org/10.1371/journal.pcbi.1004226>.
- Lane N, Martin W. The energetics of genome complexity. *Nature*. 2010;467(7318):929–934. <https://doi.org/10.1038/nature09486>.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357–359. <https://doi.org/10.1038/nmeth.1923>.
- Lartillot N, Philippe H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol*. 2004;21(6):1095–1109. <https://doi.org/10.1093/molbev/msh112>.
- Lartillot N, Rodrigue N, Stubbs D, Richer J. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst Biol*. 2013;62(4):611–615. <https://doi.org/10.1093/sysbio/syt022>.
- Leung PM, Grinter R, Tudor-Matthew E, Lingford JP, Jimenez L, Lee H-C, Milton M, Hanchapola I, Tanuwidjaya E, Kropp A, et al. Trace gas oxidation sustains energy needs of a thermophilic archaeon at suboptimal temperatures. *Nat Commun*. 2024;15(1):3219. <https://doi.org/10.1038/s41467-024-47324-2>.
- Liu Y, Li M. The unstable evolutionary position of korarchaeota and its relationship with other TACK and Asgard archaea. *mLife*. 2022;1(2):218–222. <https://doi.org/10.1002/mlf2.12020>.
- Liu Y, Makarova KS, Huang W-C, Wolf YI, Nikolskaya AN, Zhang X, Cai M, Zhang C-J, Xu W, Luo Z, et al. Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature*. 2021;593(7860):553–557. <https://doi.org/10.1038/s41586-021-03494-3>.
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*. 2014;42(D1):D490–D495. <https://doi.org/10.1093/nar/gkt1178>.
- López-García P, Moreira D. Open questions on the origin of eukaryotes. *Trends Ecol Evol*. 2015;30(11):697–708. <https://doi.org/10.1016/j.tree.2015.09.005>.
- López-García P, Moreira D. The syntrophy hypothesis for the origin of eukaryotes revisited. *Nat Microbiol*. 2020;5(5):655–667. <https://doi.org/10.1038/s41564-020-0710-4>.
- Lu Z, Fu T, Li T, Liu Y, Zhang S, Li J, Dai J, Koonin EV, Li G, Chu H, et al. Coevolution of eukaryote-like Vps4 and ESCRT-III subunits in the Asgard archaea. *mBio*. 2020;11(3):e00417–e00420. <https://doi.org/10.1128/mBio.00417-20>.
- Lu Z, Xia R, Zhang S, Pan J, Liu Y, Wolf YI, Koonin EV, Li M. Evolution of optimal growth temperature in Asgard archaea inferred from the temperature dependence of GDP binding to EF-1A. *Nat Commun*. 2024;15(1):1–7. <https://doi.org/10.1038/s41467-023-43650-z>.
- Makarova KS, Wolf YI, Koonin EV. Archaeal clusters of orthologous genes (arCOGs): an update and application for analysis of shared features between Thermococcales, Methanococcales, and Methanobacteriales. *Life*. 2015;5(1):818–840. <https://doi.org/10.3390/life5010818>.
- Martijn J, Schön ME, Lind AE, Vosseberg J, Williams TA, Spang A, Ettema TJG. Hikarchaeia demonstrate an intermediate stage in the methanogen-to-halophile transition. *Nat Commun*. 2020;11(1):5490. <https://doi.org/10.1038/s41467-020-19200-2>.
- Martijn J, Vosseberg J, Guy L, Offre P, Ettema TJG. Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature*. 2018;557(7703):101–105. <https://doi.org/10.1038/s41586-018-0059-5>.
- Martin W, Müller M. The hydrogen hypothesis for the first eukaryote. *Nature*. 1998;392(6671):37–41. <https://doi.org/10.1038/32096>.
- McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol*. 2011;10(1):13–26. <https://doi.org/10.1038/nrmicro2670>.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020;37(5):1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
- Moody ERR, Mahendrarajah TA, Dombrowski N, Clark JW, Petitjean C, Offre P, Szöllösi GJ, Spang A, Williams TA. An estimate of the deepest branches of the tree of life from ancient vertically evolving genes. *Elife*. 2022;11:e66695. <https://doi.org/10.7554/eLife.66695>.
- Moreira D, Lopez-Garcia P. Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J Mol Evol*. 1998;47(5):517–530. <https://doi.org/10.1007/PL00006408>.
- Muñoz-Gómez SA, Susko E, Williamson K, Eme L, Slamovits CH, Moreira D, López-García P, Roger AJ. Site-and-branch-heterogeneous analyses of an expanded dataset favour mitochondria as sister to known Alphaproteobacteria. *Nat Ecol Evol*. 2022;6(3):253–262. <https://doi.org/10.1038/s41559-021-01638-2>.
- Munson-McGee JH, Field EK, Bateson M, Rooney C, Stepanauskas R, Young MJ. Nanoarchaeota, their sulfobiales host, and nanoarchaeota virus distribution across yellowstone national park hot springs. *Appl Environ Microbiol*. 2015;81(22):7860–7868. <https://doi.org/10.1128/AEM.01539-15>.
- Nayfach S, Roux S, Seshadri R, Udvardy D, Varghese N, Schulz F, Wu D, Paez-Espino D, Chen I-M, Huntemann M, et al. A genomic catalog of Earth's microbiomes. *Nat Biotechnol*. 2021;39(4):499–509. <https://doi.org/10.1038/s41587-020-0718-6>.
- Nishimura Y, Eguchi T. Biosynthesis of archaeal membrane lipids: diglyceranylethylglycerophospholipid reductase of the thermoacidophilic archaeon *Thermoplasma acidophilum*. *J Biochem*. 2006;139(6):1073–1081. <https://doi.org/10.1093/jb/mvj118>.
- Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res*. 2022;50(D1):D785–D794. <https://doi.org/10.1093/nar/gkab776>.
- Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol*. 2018;36(10):996–1004. <https://doi.org/10.1038/nbt.4229>.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. Checkm: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. 2015;25(7):1043–1055. <https://doi.org/10.1101/gr.186072.114>.
- Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol*. 2017;2(11):1533–1542. <https://doi.org/10.1038/s41564-017-0012-7>.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Müller A, Nothman J, Louppe G, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825–2830. <https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html>.
- Pierpont CL, Baroch JJ, Church MJ, Miller SR. Idiosyncratic genome evolution of the thermophilic cyanobacterium *Synechococcus* at the limits of phototrophy. *ISME J*. 2024;18(1):wrae184. <https://doi.org/10.1093/ismejo/wrae184>.
- Pittis AA, Gabaldón T. Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature*. 2016;531(7592):101–104. <https://doi.org/10.1038/nature16941>.
- Podar M, Makarova KS, Graham DE, Wolf YI, Koonin EV, Reysenbach A-L. Insights into archaeal evolution and symbiosis from the

- genomes of a nanoarchaeon and its inferred crenarchaeal host from Obsidian Pool, Yellowstone National Park. *Biol Direct*. 2013; 8(1):9. <https://doi.org/10.1186/1745-6150-8-9>.
- Poole AM, Gribaldo S. Eukaryotic origins: how and when was the mitochondrion acquired? *Cold Spring Harb Perspect Biol*. 2014;6(12):a015990. <https://doi.org/10.1101/cshperspect.a015990>.
- Qu Y-N, Rao Y-Z, Qi Y-L, Li Y-X, Li A, Palmer M, Hedlund BP, Shu W-S, Evans PN, Nie G-X, *et al*. Panguiarchaeum symbiosum, a potential hyperthermophilic symbiont in the TACK superphylum. *Cell Rep*. 2023;42(3):112158. <https://doi.org/10.1016/j.celrep.2023.112158>.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
- R Core Team. *R: a language and environment for statistical computing*. Vienna (Austria): R Foundation for Statistical Computing; 2024. <https://www.R-project.org/>.
- Ragsdale SW, Pierce E. Acetogenesis and the Wood–Ljungdahl pathway of CO(2) fixation. *Biochim Biophys Acta*. 2008;1784(12):1873–1898. <https://doi.org/10.1016/j.bbapap.2008.08.012>.
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst Biol*. 2018;67(5):901–904. <https://doi.org/10.1093/sysbio/syy032>.
- Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res*. 2018;46(D1):D624–D632. <https://doi.org/10.1093/nar/gkx1134>.
- Rinke C, Chuvochina M, Mussig AJ, Chaumeil P-A, Davin AA, Waite DW, Whitman WB, Parks DH, Hugenholtz P. A standardized archaeal taxonomy for the Genome Taxonomy Database. *Nat Microbiol*. 2021;6(7):946–959. <https://doi.org/10.1038/s41564-021-00918-8>.
- Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S, Swan BK, Gies EA, *et al*. Insights into the phylogeny and coding potential of microbial dark matter. *Nature*. 2013;499(7459):431–437. <https://doi.org/10.1038/nature12352>.
- Roger AJ, Muñoz-Gómez SA, Kamikawa R. The origin and diversification of mitochondria. *Curr Biol*. 2017;27(21):R1177–R1192. <https://doi.org/10.1016/j.cub.2017.09.015>.
- Sabath N, Ferrada E, Barve A, Wagner A. Growth temperature and genome size in bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation. *Genome Biol Evol*. 2013;5(5):966–977. <https://doi.org/10.1093/gbe/evt050>.
- Saier MH, Reddy VS, Moreno-Hagelsieb G, Hendargo KJ, Zhang Y, Iddamsetty V, Lam KJK, Tian N, Russum S, Wang J, *et al*. The transporter classification database (TCDB): 2021 update. *Nucleic Acids Res*. 2021;49(D1):D461–D467. <https://doi.org/10.1093/nar/gkaa1004>.
- Sakai HD, Nakamura K, Kurosawa N. *Stygiolobus caldivivus* sp. nov., a facultatively anaerobic hyperthermophilic archaeon isolated from the Unzen hot spring in Japan. *Int J Syst Evol Microbiol*. 2022;72(8):005486. <https://doi.org/10.1099/ijsem.0.005486>.
- Sakai HD, Nur N, Kato S, Yuki M, Shimizu M, Itoh T, Ohkuma M, Suwanto A, Kurosawa N. Insight into the symbiotic lifestyle of DPANN archaea revealed by cultivation and genome analyses. *Proc Natl Acad Sci U S A*. 2022;119(3):e2115449119. <https://doi.org/10.1073/pnas.2115449119>.
- Santana-Molina C, Williams TA, Snel B, Spang A. Chimeric origins and dynamic evolution of central carbon metabolism in eukaryotes. *Nat Ecol Evol*. 2025;9(4):613–627. <https://doi.org/10.1038/s41559-025-02648-0>.
- Schrempf D, Lartillot N, Szöllösi G. Scalable empirical mixture models that account for across-site compositional heterogeneity. *Mol Biol Evol*. 2020;37(12):3616–3631. <https://doi.org/10.1093/molbev/msaa145>.
- Schuchmann K, Müller V. Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat Rev Microbiol*. 2014;12(12):809–821. <https://doi.org/10.1038/nrmicro3365>.
- Schuchmann K, Müller V. Energetics and application of heterotrophy in acetogenic Bacteria. *Appl Environ Microbiol*. 2016;82(14):4056–4069. <https://doi.org/10.1128/AEM.00882-16>.
- Schut GJ, Boyd ES, Peters JW, Adams MW. The modular respiratory complexes involved in hydrogen and sulfur metabolism by heterotrophic hyperthermophilic archaea and their evolutionary implications. *FEMS Microbiol Rev*. 2013;37(2):182–203. <https://doi.org/10.1111/j.1574-6976.2012.00346.x>.
- Schwank K, Bornemann TLV, Dombrowski N, Spang A, Banfield JF, Probst AJ. An archaeal symbiont-host association from the deep terrestrial subsurface. *ISME J*. 2019;13(8):2135–2139. <https://doi.org/10.1038/s41396-019-0421-0>.
- Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30(14):2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
- Shaiber A, Eren AM. Composite metagenome-assembled genomes reduce the quality of public genome repositories. *mBio*. 2019;10(3):e00725-19. <https://doi.org/10.1128/mBio.00725-19>.
- Shimodaira H. Multiple comparisons of log-likelihoods and combining nonnested models with applications to phylogenetic tree selection. *Commun Stat Theory Methods*. 2001;30(8-9):1751–1772. <https://doi.org/10.1081/STA-100105696>.
- Shimodaira H. An approximately unbiased test of phylogenetic tree selection. *Syst Biol*. 2002;51(3):492–508. <https://doi.org/10.1080/10635150290069913>.
- Søndergaard D, Pedersen CNS, Greening C. HydDB: a web tool for hydrogenase classification and analysis. *Sci Rep*. 2016;6(1):34212. <https://doi.org/10.1038/srep34212>.
- Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van Eijk R, Schleper C, Guy L, Ettema TJG. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*. 2015;521(7551):173–179. <https://doi.org/10.1038/nature14447>.
- Spang A, Stairs CW, Dombrowski N, Eme L, Lombard J, Caceres EF, Greening C, Baker BJ, Ettema TJG. Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. *Nat Microbiol*. 2019;4(7):1138–1148. <https://doi.org/10.1038/s41564-019-0406-9>.
- St John E, Liu Y, Podar M, Stott MB, Meneghin J, Chen Z, Lagutin K, Mitchell K, Reysenbach A-L. A new symbiotic nanoarchaeote (*Candidatus Nanocleptia minutus*) and its host (*Zestosphaera tikiterensis* gen. nov., sp. nov.) from a New Zealand hot spring. *Syst Appl Microbiol*. 2019;42(1):94–106. <https://doi.org/10.1016/j.syapm.2018.08.005>.
- Szánthó LL, Lartillot N, Szöllösi GJ, Schrempf D. Compositionally constrained sites drive long-branch attraction. *Syst Biol*. 2023;72(4):767–780. <https://doi.org/10.1093/sysbio/syad013>.
- Szilágyi A, Závodszy P. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure*. 2000;8(5):493–504. [https://doi.org/10.1016/S0969-2126\(00\)00133-7](https://doi.org/10.1016/S0969-2126(00)00133-7).
- Szöllösi GJ, Davin AA, Tannier E, Daubin V, Boussau B. Genome-scale phylogenetic analysis finds extensive gene transfer among fungi. *Philos Trans R Soc Lond B Biol Sci*. 2015;370(1678):20140335. <https://doi.org/10.1098/rstb.2014.0335>.
- Tahon G, Köstlbacher S, Pelve EA, Baker BJ, Saw JH, Eme L, Tamarit D, Schön ME, Ettema TJG. Phylogenomics and ancestral reconstruction of Korarchaeota reveals genomic adaptation to habitat switching. *bioRxiv*. <https://doi.org/10.1101/2023.09.28.559970>, 2 October 2023, preprint: not peer reviewed.
- Tamarit D, Köstlbacher S, Applier KE, Panagiotou K, De Anda V, Rinke C, Baker BJ, Ettema TJG. Description of *Asgardarchaeum abyssi* gen. nov. spec. nov., a novel species within the class Asgardarchaeia and phylum Asgardarchaeota in accordance with the SeqCode. *Syst Appl Microbiol*. 2024;47(4):126525. <https://doi.org/10.1016/j.syapm.2024.126525>.
- Topçuoğlu BD, Stewart LC, Morrison HG, Butterfield DA, Huber JA, Holden JF. Hydrogen limitation and syntrophic growth among natural assemblages of thermophilic methanogens at deep-sea hydrothermal vents. *Front Microbiol*. 2016;7:1240. <https://doi.org/10.3389/fmicb.2016.01240>.
- Tria FDK, Brueckner J, Skejo J, Xavier JC, Kapust N, Knopp M, Wimmer JLE, Nagies FSP, Zimorski V, Gould SB, *et al*. Gene

- duplications trace mitochondria to the onset of eukaryote complexity. *Genome Biol Evol.* 2021;13(5):evab055. <https://doi.org/10.1093/gbe/evab055>.
- Valentin-Alvarado LE, Appler KE, De Anda V, Schoelmerich MC, West-Roberts J, Kivenson V, Crits-Christoph A, Ly L, Sachdeva R, Greening C, *et al.* Asgard archaea modulate potential methanogenesis substrates in wetland soil. *Nat Commun.* 2024;15(1):6384. <https://doi.org/10.1038/s41467-024-49872-z>.
- Van Rossum G, Drake FL. *Python 3 reference manual*. Scotts Valley (CA): CreateSpace; 2009.
- Ver Eecke HC, Butterfield DA, Huber JA, Lilley MD, Olson EJ, Roe KK, Evans LJ, Merkel AY, Cantin HV, Holden JF. Hydrogen-limited growth of hyperthermophilic methanogens at deep-sea hydrothermal vents. *Proc Natl Acad Sci U S A.* 2012;109(34):13674–13679. <https://doi.org/10.1073/pnas.1206632109>.
- Vosseberg J, van Hooff JJE, Köstlbacher S, Panagiotou K, Tamarit D, Ettema TJG. The emerging view on the origin and early evolution of eukaryotic cells. *Nature.* 2024;633(8029):295–305. <https://doi.org/10.1038/s41586-024-07677-6>.
- Vosseberg J, van Hooff JJE, Marcet-Houben M, van Vlimmeren A, van Wijk LM, Gabaldón T, Snel B. Timing the origin of eukaryotic cellular complexity with ancient duplications. *Nat Ecol Evol.* 2021;5(1):92–100. <https://doi.org/10.1038/s41559-020-01320-z>.
- Wang H-C, Minh BQ, Susko E, Roger AJ. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst Biol.* 2018;67(2):216–235. <https://doi.org/10.1093/sysbio/syx068>.
- Waskom M. Seaborn: statistical data visualization. *J Open Source Softw.* 2021;6(60):3021. <https://doi.org/10.21105/joss.03021>.
- Werner F. Structure and function of archaeal RNA polymerases. *Mol Microbiol.* 2007;65(6):1395–1404. <https://doi.org/10.1111/j.1365-2958.2007.05876.x>.
- Williams TA, Cox CJ, Foster PG, Szöllösi GJ, Embley TM. Phylogenomics provides robust support for a two-domains tree of life. *Nat Ecol Evol.* 2020;4(1):138–147. <https://doi.org/10.1038/s41559-019-1040-x>.
- Williams TA, Davin AA, Szánthó LL, Stamatakis A, Wahl NA, Woodcroft BJ, Soo RM, Eme L, Sheridan PO, Gubry-Rangin C, *et al.* Phylogenetic reconciliation: making the most of genomes to understand microbial ecology and evolution. *ISME J.* 2024;18(1):wrae129. <https://doi.org/10.1093/ismejo/wrae129>.
- Williams TA, Schrempf D, Szöllösi GJ, Cox CJ, Foster PG, Embley TM. Inferring the deep past from molecular data. *Genome Biol Evol.* 2021;13(5):evab067. <https://doi.org/10.1093/gbe/evab067>.
- Williams TA, Szöllösi GJ, Spang A, Foster PG, Heaps SE, Boussau B, Ettema TJG, Embley TM. Integrative modeling of gene and genome evolution roots the archaeal tree of life. *Proc Natl Acad Sci U S A.* 2017;114(23):E4602–E4611. <https://doi.org/10.1073/pnas.1618463114>.
- Woodcroft BJ, Aroney STN, Zhao R, Cunningham M, Mitchell JAM, Blackall L, Tyson GW. Comprehensive taxonomic identification of microbial species in metagenomic data using SingleM and Sandpiper. *Nat Biotechnol.* 2025. <https://doi.org/10.1038/s41587-025-02738-1>, 31 January 2024, preprint: not peer reviewed.
- Wu F, Speth DR, Filosof A, Crémère A, Narayanan A, Barco RA, Connors SA, Amend JP, Antoshchuk IA, Orphan VJ. Unique mobile elements and scalable gene flow at the prokaryote-eukaryote boundary revealed by circularized Asgard archaea genomes. *Nat Microbiol.* 2022;7(2):200–212. <https://doi.org/10.1038/s41564-021-01039-y>.
- Wurch L, Giannone RJ, Belisle BS, Swift C, Utturkar S, Hettich RL, Reysenbach A-L, Podar M. Genomics-informed isolation and characterization of a symbiotic Nanoarchaeota system from a terrestrial geothermal environment. *Nat Commun.* 2016;7(1):12115. <https://doi.org/10.1038/ncomms12115>.
- Xie R, Wang Y, Huang D, Hou J, Li L, Hu H, Zhao X, Wang F. Expanding Asgard members in the domain of archaea sheds new light on the origin of eukaryotes. *Sci China Life Sci.* 2022;65(4):818–829. <https://doi.org/10.1007/s11427-021-1969-6>.
- Xu S, Li L, Luo X, Chen M, Tang W, Zhan L, Dai Z, Lam TT, Guan Y, Yu G. A serialized data object for visualization of a phylogenetic tree and annotation data. *Imeta.* 2022;1(4):e56. <https://doi.org/10.1002/imt2.56>.
- Yu H, Wu C-H, Schut GJ, Haja DK, Zhao G, Peters JW, Adams MWW, Li H. Structure of an ancient respiratory system. *Cell.* 2018;173(7):1636–1649.e16. <https://doi.org/10.1016/j.cell.2018.03.071>.
- Yu T, Fu L, Wang Y, Dong Y, Chen Y, Wegener G, Cheng L, Wang F. Thermophilic Hadarchaeota grow on long-chain alkanes in syntrophy with methanogens. *Nat Commun.* 2024;15(1):6560. <https://doi.org/10.1038/s41467-024-50883-z>.
- Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L, Vancaester E, Seitz KW, Anantharaman K, Starnawski P, Kjeldsen KU, *et al.* Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature.* 2017;541(7637):353–358. <https://doi.org/10.1038/nature21031>.
- Zeldovich KB, Berezovsky IN, Shakhnovich EI. Protein and DNA sequence determinants of thermophilic adaptation. *PLoS Comput Biol.* 2007;3(1):e5. <https://doi.org/10.1371/journal.pcbi.0030005>.
- Zhang J, Feng X, Li M, Liu Y, Liu M, Hou L-J, Dong H-P. Deep origin of eukaryotes outside Heimdallarchaea within Asgardarchaeota. *Nature.* 2025;642:990–998. <https://doi.org/10.1038/s41586-025-08955-7>.