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*]

1. Department of Marine Microbiology and Biogeochemistry, NIOZ, Royal Netherlands Institute for Sea Research, 1790 AB Den Burg, The Netherlands
2. Department of Evolutionary & Population Biology, Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, 1090 GE Amsterdam, The Netherlands
3. Faculty of Biology, Department of Microbiology, University of Innsbruck, 6020 Innsbruck, Austria
4. Chinese Academy of Sciences Key Laboratory of Urban Pollutant Conversion, Department of Environmental Science and Engineering, University of Science and Technology of China, 230026 Hefei, China
5. Model-Based Evolutionary Genomics Unit, Okinawa Institute of Science and Technology Graduate University, 904-0495 Okinawa, Japan
6. Institute of Evolution, HUN-REN Center for Ecological Research, H-1121 Budapest, Hungary
7. Laboratory of Microbiology, Wageningen University & Research, Wageningen, The Netherlands
8. Bristol Palaeobiology Group, School of Biological Sciences, University of Bristol, BS8 1TQ, Bristol, UK

*Corresponding authors: Anja Spang (anja.spang@nioz.nl) and Tom A. Williams (tom.a.williams@bristol.ac.uk)

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 characterising 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 two 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 Pangiuarchaeales 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 modelled. *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 based on H_2-syntrophy. Thus, *Panguiarchaeum* may contain the first obligate symbionts within Asgard archaea.

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 (Spang et al. 2015; Zaremba-Niedzwiedzka et al. 2017; Liu et al. 2021; Wu et al. 2022; Valentin-Alvarado et al. 2024) encode eukaryotic signature proteins (ESPs) (Vosseberg et al. 2024), which are homologs of eukaryotic proteins that underlie cellular complexity. Experimental evidence suggests that at least some ESPs of Asgard archaea have functions similar to those in eukaryotes (Akıl and Robinson 2018; Akıl 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 at least some members form symbiotic, (that is, 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-Garcia 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; 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 (formerly Thaumarchaeota, Aigarchaeota, and Crenarchaeota) and Korarchaeota (Elkins et al. 2008). Early analyses supported a clade of “TACK” Archaea (Guy and Ettema 2011) (initially comprising the phyla Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Koarchaeota now classified as class-level lineages within the Thermoproteota) as the sister lineage to the 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 Thermoproteota and Asgard clade (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) and the Panguiarchaeales (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, either as 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 Asgard 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 is more common when these taxa are excluded (Eme et al. 2023; Valentin-Alvarado et al. 2024), depending on the dataset and phylogenetic method. Similarly, the recently-reported Panguiarchaeales, 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 site-homogeneous model (Qu et al. 2023). But conflicting placement as Asgard archaea was observed using 53 marker genes from GTDB (release 207v2) (Dombrowski et al. 2020; Rinke et al. 2021). Panguiarchaeales were proposed to represent anaerobic amino acid fermenters with a symbiotic lifestyle (Qu et al. 2023) because comparative genomic analyses indicated that they lack certain genes for archaeal lipid, purine and amino acid biosynthesis pathways reminiscent of DPANN archaea (Qu et al. 2023). However, the authors 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 did not significantly differ in size from other Korarchaeota with which they were proposed 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 Panguiarchaeales 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 spectre 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 Njordarchaeles 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 MAGs quality of key representatives of the Pangui- and Njordarchaeales and applied a variety of comparative genomics, phylogenomics, and network approaches to determine the phylogenetic placement of Njordarchaeales and Panguiarchaeales 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 Panguiarchaeales 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, *Panguiarchaeaceae* 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 *Panguiarchaeum* and members of the Thermoproteota. Overall, this study shows that MAG contamination is not the cause for 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

Panguiarchaeum 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 Panguiarchaeales and Njordarchaeales (13 MAGs, **see Supplementary Data 1**), 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), totalling 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 1.1**, **Supplementary Data 2-4**, **Supplementary Figures 1-8**). 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 4**) and selected the top 50% of these markers corresponding to 43 proteins(**Supplementary Data 4**) 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 summarised in **Supplementary Data 2-4**.

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 (Qu et al. 2023; Zhang et al. 2025) or Asgard archaea (Eme et al. 2023; Valentin-Alvarado et al. 2024) (**Figure 1A**). To distinguish between these hypotheses (i.e. Korarchaeota versus Asgard archaea topologies (**Figure 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 normalise the taxonomic ranks (Rinke et al. 2021; Tamarit et al. 2024) (see Methods and **Supplementary Figure 1** 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 previous findings which suggested that Panguiarchaea 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 two family-level lineages, the *Panguiarchaeaceae* and *Njordarchaeaceae* (**Figure 1BC, Supplementary Figure 9,** See also **Supplementary Information 1.2**). Furthermore, analyses using the best-fitting substitution models robustly supported the placement of the Njordarchaeia within the Asgard archaea (**Figure 1BC**, **Supplementary Figure 1**). 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 Figure 1** and **Supplementary Figure 10A-C**), consistent with the hypothesis that this inferred relationship might be a phylogenetic artefact (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, in previous case studies, 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).

Modelling site-wise compositional heterogeneity. Site-wise compositional variation results from selective constraints on protein function: typically, only a small subset of the twenty 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). As they do not account for site-specific selective constraints, simple models under-estimate the probability of convergent evolution to the same amino acid in independent, distantly-related lineages. This means they can sometimes mis-interpret instances of molecular convergence as evidence that these taxa are closely related, an artefact 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 (**Figure 1D**). This analysis revealed that the support for the Korarchaeota tree in simpler/underfitting model (**Figure 1E, assessed using model adequacy test** (Giacomelli et al. 2025)), LG+G+F/LG+C60+G+F, was greater at the more compositionally constrained sites, i.e., effective number of amino acids (K_{eff}) ≤ 9, whereas with the best-fitting model and second best-fitting model, LG+CAT-PMSF and Poisson+CAT-PMSF (**Figure 1E, assessed usingmodel adequacy test** (Giacomelli et al. 2025)), both constrained and compositionally variable sites supported the Asgard archaea tree (**Figure 1D)**. Additionally, under the best-fitting model LG+CAT-PMSF and second best-fitting model Poisson+CAT-PMSF, Asgard archaea tree were consistenly recovered (**Supplementary Figures 10-13, Figure 1E**), and the Korarchaeota tree is confidently rejected by the approximately unbiased (AU) test (Shimodaira 2002) ($P < 0.05$) (**Supplementary Data 5**). These analyses indicate that failure to model site-wise compositional heterogeneity contributes to the incorrect

269 placement of Njordarchaeia with Korarchaeota in analyses with poorly-fitting models
270 such as LG+G+F and LG+C60+G+F (**Figure 1E**).
271
272 *Modelling branch-wise compositional heterogeneity.* In addition to site-wise variation,
273 sequence composition can also vary among lineages with, for example, distantly-
274 related thermophiles converging on greater use of thermostable amino acids (Ile, Val,
275 Tyr, Trp, Arg, Glu, Leu (IVYWREL (Zeldovich et al. 2007)) or charged amino acids:
276 Asp, Glu, Lys, Arg (DEKR) (Cambillau and Claverie 2000; Szilágyi and Závodszky
277 2000)). Njordarchaeia and Korarchaeota were both predicted to be
278 (hyper-)thermophiles, and the placement of Njordarchaeia with Korarchaeota was
279 suggested to be a result of compositional biases (Eme et al. 2023). In line with a
280 previous study (Eme et al. 2023), a comparison of proteome composition across
281 Archaea indicated that they indeed have similar amino acid compositions
282 (**Supplementary Figure 14AB**). We subsequently investigated whether
283 compositional attraction between the two lineages might also contribute to the
284 recovery of the Korarchaeota tree under simple models. First, we performed a site-
285 filtering analysis. We ranked sites by compositional bias, iteratively removing the sites
286 that made the greatest individual contribution to across-branch compositional
287 heterogeneity (Dombrowski et al. 2020; Baker et al. 2024) (**see Methods**), and re-
288 evaluated support for the competing topologies (**Figure 1F**). Removal of the top 20%
289 of most biased sites shifted support from the Korarchaeota topology to the Asgard
290 archaea topology under LG+C60+G+F, while under the better-fitting and custom site-
291 heterogeneous model, LG+EDM0256LCLR+G+F, support was unchanged, with all
292 analyses favouring the Asgard archaea topology. We interpret these results as
293 evidence that compositional attraction between thermophilic branches favours the
294 Korarchaeota topology in the full dataset, but whether this non-phylogenetic signal
295 drives the final result depends on other aspects of the analysis. While neither the
296 LG+C60+G+F model nor the LG+EDM0256LCLR+G+F model account for across-
297 branch compositional variation, C60 fares worse than EDM00256 in modelling across-
298 site variation (**Supplementary Figure 11;** Bayesian information criterion (BIC) of
299 LG+C60+G+F: 6668927.72, LG+EDM0256LCLR+G+F: 6532626.81). We therefore
300 suggest that support for the Korarchaeota topology from the 303 taxa set in the C60
301 analysis reflects a combination of phylogenetic artifacts from both unmodelled branch-
302 wise compositional attraction and inadequately-modelled site-wise compositional

303 variation (that is, long branch attraction), with support shifting to the Asgard archaea
304 topology when one of these sources of error is removed, either by filtering out highly
305 biased sites or by better modelling of site-wise compositions with
306 LG+EDM0256LCLR+G+F.
307
308 Second, we identify RYPEIWKVL and QSNTDGC, as overrepresented and
309 underrepresented amino acids, respectively, in the concatenation (Zeldovich et al.
310 2007) (**Supplementary Figure 14CD**). Subsequently, we performed phylogenetic
311 analysis using a substitution model that accounts for changing compositions across
312 the tree, i.e. GFMix (Muñoz-Gómez et al. 2022). The
313 LG+EDM00256LCLR+G+F+GFMix (RYPEIWKVL/QSNTDGC and RYPEI/QSNT)
314 and LG+C60+G+F+GFmix (RYPEIWKVL/QSNTDGC and RYPEI/QSNT) model
315 provided strong support for the Asgard archaea tree over the Korarchaeota tree
316 (**Supplementary Data 6**), again consistent with the hypothesis that compositional
317 attraction due to a shared environmental adaptation contributes to the Korarchaeota
318 tree in analyses with models that do not account for branch heterogeneity (Eme et al.
319 2023).
320
321 *Is the inferred phylogenetic position of Panguiarchaeceae and Njordarchaeceae*
322 *affected by metagenomic contamination?*
323
324 The phylogenetic position of *Njordarchaeia* has been the subject of debate based on
325 different marker sets. A recent analysis (Dombrowski et al. 2020; Zhang et al. 2025)
326 argued that these MAGs may be chimeric, comprising sequences derived from both
327 Thermoproteota and Asgard archaea, which contributes to conflicting phylogenetic
328 placements observed in different marker sets. It has been pointed out that MAGs often
329 contains various degree of contamination (Parks et al. 2015; Shaiber and Eren 2019;
330 Chen et al. 2020), and the MAGs analysed in this study are no exception as assessed
331 by CheckM and differential coverage (**Figure 2; Supplementary Data 1**).
332 Nevertheless, the key question is whether the marker genes used to place these
333 lineages are affected by contamination. Using the same analytical approach, we
334 inspected the distribution of marker genes, and within-bin coverage variation of
335 Panguiarchaeceae and Njordarchaeceae MAGs across different metagenomes from
336 hydrothermal vents in the Guaymas Basin (Dombrowski et al. 2018) and a hot spring

337 in Yunnan, China (Qu et al. 2023). CheckM analysis indicates that all representative
338 Njordarchaeia MAGs analysed here (**taxon set 303, Supplementary Data 1**) are at
339 least medium quality (Bowers et al. 2017) with only a moderate level of contamination
340 (<10%), with 53% (7 MAGs) having low level of contamination (<5%) (**Supplementary**
341 **Data 1**). MAGs with/above medium quality are the majority in different databases such
342 as the Genomes from Earth's Microbiomes catalogs (Nayfach et al. 2021), and
343 reported in various studies (Dombrowski et al. 2020; Martijn et al. 2020; Zhang et al.
344 2025). To investigate whether our marker gene phylogenies are nevertheless affected
345 by this low level of contaminations, we chose to examine 4 MAGs from our focal taxon
346 set that derive from hydrothermal vents in Guaymas basin and a hot spring from China
347 (estimated contamination: 0.93% to 7.01%; **Figure 2**) that were also examined in a
348 previous study (Zhang et al. 2025). While our analyses confirmed the presence of
349 small potentially contaminating contigs (**Figure 2ABCD**), all marker genes mapped
350 back to a single majority cluster corresponding to the target population, even in the
351 MAG estimated with the highest level of contamination (**Figure 2C**). This suggests that
352 contamination, though present at a low/moderate level in some of these MAGs, does
353 not affect the marker genes used, and so does not underpin the phylogenetic conflicts.
354 Instead, conflicting support appears to depend on the substitution model used. For
355 example, analyses of two marker genes situated within a non-contaminated contig
356 (**Figure 2ABC, highlighted in red box**) under the poorly-fitting LG+F+G and
357 LG+C60+F+G models differed in their support for the Asgard and Korarchaeota
358 topologies, whereas analysis of the same marker genes under the best-fitting
359 LG+CAT-PMSF model provides moderate support for Asgard archaea topology
360 (**Figure 2ABC**). We also observed in our dataset that LG+C60+G+F did not
361 significantly distinguish these two topologies (AU tests (Shimodaira 2002), $P < 0.05$;
362 **Supplementary Data 5)** nor describe site-heterogeneity adequately **(Figure 1E**).
363 However, as model fit improved, consistent support for the Asgard topology emerged
364 (**Supplementary Data 5, Figure 1DEF**).
365
366 While we saw no evidence of contamination in our analyses, we did see some variation
367 in support for the Korarchaeota or Asgard topologies on a per-gene basis when the
368 data were analysed using different models. To quantify disagreements among marker
369 genes, we used AU tests (Shimodaira 2002) to determine, for each marker gene and
370 substitution model, whether either the Korarchaeota tree or Asgard archaea tree could

371 be rejected (at P < 0.05). Across all models tested, 27 of 43 marker genes did not reject
372 either hypothesis, suggesting that individual marker genes often do not contain
373 sufficient information to discriminate among phylogenetic hypotheses (**Figure 2E**). For
374 the remaining marker genes, support on a per-gene basis shifted from the
375 Korarchaeota to the Asgard archaea topology as model fit (assessed using model
376 adequacy test (Giacomelli et al. 2025)) improved. Under the worst-fitting LG+G+F
377 model, 5 genes rejected the Korarchaeota topology and 11 rejected the Asgard
378 topology at the significance level (**Figure 2E**). Under the site-heterogenous
379 LG+C60+G+F model, 5 genes rejected Korarchaeota and 9 rejected Asgard archaea.
380 Under the best-fitting (LG+CAT-PMSF) and second best-fitting (Poisson+CAT-PMSF)
381 models, 9 and 12 genes, respectively, rejected the Korarchaeota topology, while 4
382 rejected the Asgard topology (**Figure 2E**). The improvement in congruence across
383 marker genes and the shift in support to the Asgard topology as model fit improves
384 suggests that the apparent strong but inconsistent support observed in analyses under
385 the poorly-fitting LG+G+F and LG+C60+G+F models is artefactual, resulting from the
386 failure of these models to adequately account for non-historical signal in sequence data
387 (**Figure 1E**; **Supplementary Information 1.3-1.4**).
388
389
390 *Assignment of Njordarchaeia to Asgard archaea is supported by shared eukaryotic*
391 *signature proteins (ESPs) and protein presence-absence profiles*
392
393 The phylogenetic analyses described above provided strong support for the placement
394 of *Panguiarchaeaceae* within Njorarchaeales as a sister to *Njordarchaeaceae* within
395 the Asgard archaea (**Figure 1B-F**). Consistent with that placement, and in line with
396 (Eme et al. 2023), we observed, we observed that *Panguiarchaeaceae* and
397 *Njordarchaeaceae* encode a number of eukaryotic signature proteins (ESPs) shared
398 with other Asgard archaea **(Figure 3, Supplementary Data 7**), based on analyses
399 using the Asgard Cluster of Orthologues (AsCOGs) database (Liu et al. 2021). Of
400 particular interest, we identified homologs of ribosomal L28e/Mak16 (PF01778) in
401 *Panguiarchaeaceae* MAGs, whose distribution in published archaeal genomes is
402 otherwise restricted to other Njordarchaeia, the Hodarchaeales, and Wukongarchaeia
403 (Eme et al. 2023). Phylogenetic analysis of ribosomal L28e/Mak16 revealed a topology
404 consistent with the species tree **(Supplementary Figure 15, Supplementary**

Information 2.2.1), 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 (**Supplementary Figures 16-17, Supplementary Information 2.2.2**). Homologs of proteins involved in the N-glycosylation process were also identified in *Panguiarchaeaceae* as well as *Njordarchaeaceae* MAGs (Eme et al. 2023), and the phylogenetic analyses support their affiliation with Asgard archaeal homologs (**Supplementary Figures 18-19**, **Supplementary Information 2.2.3**). Phylogenetic analyses of Njordarchaeia SNF7 proteins also support them as Asgard archaeal homologs (**Supplementary Figure 20**, **Supplementary Information 2.2.4**).

Finally, protein presence-absence profiles revealed a pattern of transcription, translation and replication distinct from Korarchaeota (**Supplementary Figures 21-22**). Interestingly however, the profile for other COG categories was more similar to that of Korarchaeota than to other Asgard archaea (**Supplementary Figures 21-22**), likely reflecting adaptation to 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 *Njordarcheaceae* 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 synthesise all organic compounds needed for growth (**Supplementary Figure 23**). Of note, members of the *Panguiarchaeum* show a reduced metabolic potential, in particular in purine and lipid biosynthesis pathways, compared to Njordarchaeacea and Korarchaeota (**Supplementary Figure 23-27, Supplementary Information 3.1-3.2**).

438 *Catabolism and energy conservation*. The catabolic potential of the *Panguiarchaeum*
439 clade (i.e. upon divergence from GCA_029856635.1) is more limited than that of
440 *Njordarchaeaceae* (**Figure 4**, **Supplementary Data 8-10, Supplementary Figure 23**).
441 In particular, while both clades encode a nearly complete Embden-Meyerhof-Parnas
442 (EMP) pathway, members of *Njordarchaeaceae* encode more genes for enzymes of
443 the tricarboxylic acid cycle (TCA) and beta-oxidation pathway. Furthermore, members
444 of both clades might have the potential to grow on organic substrates, which could be
445 fermented to acetate by a putative acetate-CoA ligase (ACD) with concomitant
446 production of ATP (Glasemacher et al. 1997). Specifically, *Panguiarchaeum*
447 representatives revealed a potential to harness electrons from organic substrates such
448 as amino acids, peptides, 2-ketoacids (e.g., pyruvate, oxoglutarate, and indole-
449 pyruvate) (Qu et al. 2023), and aldehydes (**Supplementary Data 8**, **Supplementary**
450 **Information 3.2**). Additionally, we identified homologues of the small and large
451 subunits of [NiFe]-hydrogenases belonging to group 3 and group 4 in *Panguiarchaeum*
452 MAGs. Phylogenetic analyses of the large subunit of the [NiFe] hydrogenase indicated
453 that panguiarchaeal homologs belong to subgroup 4g (Qu et al. 2023), forming a clade
454 with Odinarchaeia sequences (**Supplementary Figure 28A**). Group 4g [NiFe]-
455 hydrogenases of *Panguiarchaeum* encode a Nuo-L-like subunit, which might be
456 involved in sodium/ion translocation (Yu et al. 2018) (**Supplementary Figure 28B**).
457 This suggests that *Panguiarchaeum* may conserve energy by coupling the oxidation
458 of organic substrates via reduced ferredoxin to the formation of H_2 to generate a proton
459 gradient for ATP synthase. Members of *Panguiarchaeum* also contain various group
460 3 [NiFe]-hydrogenases assigned to subgroup 3c and 3b (Qu et al. 2023), which may
461 be involved in the reoxidation of cofactors during carbon oxidation and the production
462 of H_2, a reversible process (**Supplementary Figure 29, Supplementary Data 11**).
463 Together, this indicates that representatives of the *Panguiarchaeum* clade can
464 conserve energy through substrate-level phosphorylation as well as via the ATP
465 synthase using a proton gradient generated by the oxidation of ferredoxin similar to
466 Odinarchaeia and Thermococci (Spang et al. 2019; Qu et al. 2023).
467
468 *Anabolism*. The biosynthetic potential of members of the *Panguiarchaeum* clade
469 seemed more restricted than that of *Njordarchaeaceae* (**Supplementary Figure 23**).
470 For instance, all three *Panguiarchaeum* MAGs lack various genes encoding enzymes
471 for *de novo* purine biosynthesis (from phosphoribosyl pyrophosphate to inosine

472 monophosphate) (Qu et al. 2023), such as glutamine phosphoribosylpyrophosphate
473 amidotransferase (PurF) and Phosphoribosylaminoimidazole (AIR) synthetase (PurM)
474 indicating a possible dependence on external purine sources (Brown et al. 2011).
475 Notably, *Panguiarchaeum* clade MAGs do not encode enzymes of the mevalonate
476 pathway (MVAP) for isoprenoid biosynthesis and fatty acid biosynthesis, as indicated
477 previously (Qu et al. 2023). However, the presence of CDP-archaeol synthase (CarS),
478 phosphatidylglycerophosphate synthase (PgsA), phosphatidylglycerophosphate
479 synthase (PssA) and cardiolipin synthase (Cls), suggests that *Panguiarchaeum* have
480 the potential to attach polar head groups to di-o-geranylgeranylglyceryl phosphate
481 (DGGGP). The *Panguiarchaeum* MAGs also encode a gene for
482 digeranylgeranylglycerophospholipid reductase (GGR), which could catalyse the
483 hydrogenation/saturation of geranylgeranyl chains of DGGGP (Nishimura and Eguchi
484 2006). Phylogenetic analyses of CarS showed that the homologs of *Panguiarchaeum*
485 branched sister to Heimdallarchaeia with moderate support (UFBOOT/SH-aLRT:
486 88.8/83; **Supplementary Figure 30**). Similarly, although the PgsA homologs of
487 *Panguiarchaeum* branched sister to a clade comprising one Thermococcus and two
488 Heimdallarchaeia sequences with weak support (UFBOOT/SH-aLRT: 50.5/83), they
489 were nested within a cluster of predominantly Asgard archaeal sequences
490 (**Supplementary Figure 31**), Thus, these genes were likely inherited vertically from
491 their common ancestor with Heimdallarchaeia (**Supplementary Figure 30-32**).

492 ***Panguiarchaeum* has evolved by genome streamlining from a larger ancestor**
493 **with Heimdallarchaeia**

494 Reconciliation analyses have been used to infer the ancestral genome content of
495 various archaeal lineages (Williams et al. 2017; Martijn et al. 2020; Huang et al. 2021;
496 Baker et al. 2024; Williams et al. 2024), including Asgard archaea (Eme et al. 2023),
497 shedding light on ancestral gene repertoires and gene content evolution. To
498 reconstruct the evolution of a putative symbiotic lifestyle of *Panguiarchaeum*, we
499 reconciled single gene trees of 6077 arCOG gene families with the inferred species
500 tree (from streamlined set 2 with 303 taxa, **Figure 1B**) to determine the pattern and
501 extent of reductive genome evolution in the *Panguiarchaeum* clade and its sister
502 lineages.

503 Our analysis suggests an ongoing decrease in gene content through time in the
504 evolution of Njordarchaeia, similar to a previous analysis (Eme et al. 2023), from
505 around 1951 arCOG families in the common ancestor with Wukong-Heimdallarchaeia
506 to 911-1545 (mean: 1186.5) arCOG families among extant Njordarchaeia
507 (**Supplementary Data 12**). Based on the relationship between arCOG family
508 members and the genome size of extant taxa (Pearson's correlation coefficient: 0.89,
509 P-value: 1.96x1e-110), we inferred the genome size of the last common ancestor of
510 Wukong-Njord-Heimdallarchaeia to be 2.85 Mbp (95% confidence interval, 2.57 to
511 3.06, **Supplementary Data 12**). A notable reduction in genome size is inferred at the
512 branch leading to the Njordarchaeia (from 0.53 to 0.84 Mbp; node id: 571-563) after
513 divergence from their common ancestor with Wukong-Heimdallarchaeia (**Figure 5A**,
514 **Supplementary Data 12**). The arCOG families lost along this branch include
515 components of amino acid biosynthesis, de novo purine biosynthesis (from ribose-5-
516 phosphate to inosine monophosphate (IMP)), Wood-Ljungdahl pathway genes and
517 cofactors (methanofuran) biosynthesis (**Figure 5B**, **Supplementary Data 13**).
518 Additional genes, including those encoding various enzymes involved in archaeal lipid
519 biosynthesis and beta-oxidation, appear to have been lost along the branch leading to
520 the *Panguiarchaeum* clade after its divergence from GCA_029856635.1 and the
521 common ancestor with *Njordarchaeaceae*, which resulted in an inferred genome
522 reduction of 0.13 Mbp to 0.17 Mbp. *Panguiarchaeum* and *Njordarchaeaceae* were
523 inferred to encode fewer ESPs than the other Asgard archaea (**Figure 3**), and the loss
524 on the *Njordarchaeaceae* and *Panguiarchaeaceae* stem is predicted to include the
525 reduction in ESP repertoire as well. For instance, we inferred the loss of a
526 Roadblock/LC7 domain-containing protein (arCOG02605_01) on the
527 *Njordarchaeaceae* branch; the loss of Ras family GTPase (arCOG05343_01) and Yip1
528 domain family (arCOG02054_01) on the *Panguiarchaeaceae* branch; the loss of VPS4
529 (arCOG01307_01) on the *Panguiarchaeum* branch (i.e. upon divergence from
530 GCA_029856635.1) (**Supplementary Data 13**); and the absence of Snf7 family
531 proteins in the extant *Panguiarchaeum* MAGs indicated their loss on the
532 *Panguiarchaeum* branch (**Supplementary Figure 20**).

533 Njordarchaeia encodes reverse gyrase, a topoisomerase typically encoded by
534 hyperthermophiles (Forterre 2002). Consistent with previous work (Eme et al. 2023)
535 and proposal of a thermophilic ancestry of Asgard archaea (Lu et al. 2024), we inferred

536 the presence of reverse gyrase (Presence probability, PP = 0.96) in the last common
537 ancestor of Asgard archaea. This gene family was, however, subsequently lost in
538 Heimdallarchaeia (**Supplementary Data 13**) but not in Wukongarchaeia and
539 Njordarchaeia. Interestingly, phylogenetic and reconciliation analyses suggest a later
540 re-acquisition of a reverse gyrase in the *Panguiarchaeum* clade (**Supplementary**
541 **Figures 33-34**; **Supplementary Information 3.1.2**). The MAG/genome size of
542 putative hyperthermophilic Asgard archaea encoding reverse gyrase (mean: 2.08 Mbp,
543 n=18) is significantly smaller than that of Asgard archaea representatives without
544 reverse gyrase (mean: 3.54 Mbp, n=45) (p=5.669x10-7, two-tailed Wilcoxon rank sum
545 test). Genome reduction has previously been reported in cases of thermophilic
546 adaptation, perhaps due to a selective advantage for smaller cell volumes (and so
547 genome) at high temperatures (Sabath et al. 2013; Pierpont et al. 2024). We therefore
548 hypothesise that the initial reductive evolution of Njordarchaeia might have been
549 associated with their adaptation to high-temperature environments. The further
550 genome reduction in *Panguiarchaeum*, including the loss of essential genes for
551 biosynthetic pathways, might be associated with a transition to a host-associated
552 lifestyle (McCutcheon and Moran 2011).

553 ***Panguiarchaeum* co-occurs with Thermoproteota taxa**

554

555 To investigate the inferred dependence of *Panguiarchaeaum* on other community
556 members, we assessed the extent to which members of this group co-occur with other
557 microbes using network inferences. First, we tested the inference of our co-occurrence
558 networks with the known host-symbiont pair *Nanoarchaeum equitans* and *Ignicoccus*
559 *hospitalis* (Huber et al. 2002), and the proposed symbiosis between *Huberarchaeum*
560 *crystalense* and *Altiarchaeum hamiconexum* (Schwank et al. 2019). We did not detect
561 any associations between the former, mainly due to insufficient sampling of *N.*
562 *equitans* which is present in a single sample only (**Supplementary Figure 35-36,**
563 **Supplementary Information 4**). However, for *Huberarchaeum-Altiarchaeum,* we
564 found a high correlation (Spearman's rho = 0.931) and a robust detection frequency
565 (100%) of this suggested interaction starting at a sample size of 40 (**Supplementary**
566 **Figure 36F**; **see Methods**). Next, we identified 120 samples containing taxa
567 annotated as family *Panguiarchaeaceae* in GTDB (Rinke et al. 2021), a lineage
568 equivalent to the order Njordarchaeales proposed in this manuscript, from 248,559

NCBI metagenomes community profiles in the 'Sandpiper' website (https://sandpiper.qut.edu.au). 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 14-15**), respectively, in agreement with abundance-based co-occurrence patterns (**Figure 6)**. Within the order Njordarchaeales, there was only one OTU (OTU_804; **Supplementary Data 14**) assigned to the genus *Panguiarchaeum*, that 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 order, i.e. Thermoproteota (15 of 29), Thermoprotei_A (9 of 15), and Sulfolobales (4 of 9), and *Desulfurococcaceae* (3 of 4) (**Supplementary Data 15**). These results indicate 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. Thus, the previous placement of the *Panguiarchaeum* clade as a sister group to Korarchaeota appears to be due to phylogenetic artefacts (Qu et al. 2023) including the use of problematic marker gene families in concatenations and insufficient modelling of compositional biases and site heterogeneity similar to what

603 has been shown for Njordarchaeales (Eme et al. 2023). In contrast, our analyses did
604 not provide support for the recent suggestion that phylogenetic placement of
605 Njordarchaeia has been affected by metagenomic contamination (Zhang et al. 2025).
606 While some Njordarchaeaia MAGs do show evidence of low/moderate levels of
607 contamination (**Figure 2**), we found no evidence that marker genes were affected in
608 any case. Our analyses instead suggest that disagreement among marker genes as
609 to the placement of Njordarchaeia is a result of poor-fitting phylogenetic models, not
610 MAG chimerism. These results emphasise the importance of evolutionary model fit in
611 analyses of prokaryotic evolution, particularly for datasets containing pervasive
612 compositional heterogeneity.
613
614 The placement of *Panguiarchaeum* within the Asgard archaea is supported by the
615 presence of various ESPs shared with *Njordarchaeaceae*. However, metabolic
616 reconstructions reveal that members of *Panguiarchaeum* have various auxotrophies
617 and encode fewer ESPs than other Asgards. This is indicative of genome reduction
618 potentially associated with a host-associated lifestyle, with co-occurrence analyses
619 pointing towards an archaeal host. Notably, *Ca*. P. syntrophicum, the first cultivated
620 representative of the Asgard archaea (i.e. a member of the
621 Lokiarchaeia/Prometheoarchaeia), appears to syntrophically grow on amino acids with
622 partner organisms such as *Halodesulfovibrio* and *Methanogenium*[13]. However, this
623 symbiotic interaction does not seem to be obligate and has not resulted in reductive
624 genome evolution. Our inferences of the energy metabolism of *Panguiarchaeum,* and
625 perhaps Njordarchaeaceae as well, which seems to resemble that of Odinarchaeia
626 and *Thermococcus*, indicates the potential for syntrophic growth (Schut et al. 2013;
627 Topçuoğlu et al. 2016; Yu et al. 2018; Spang et al. 2019). For example,
628 *Panguiarchaeum* encodes 3 different [NiFe]-hydrogenases, including a putative
629 membrane-bound group 4g [NiFe]-hydrogenase, and might use electrons harnessed
630 from small organic acids/simple carbohydrates to produce H_2. While speculative in the
631 absence of experimental data, it may be hypothesised that the hydrogen could support
632 the growth of a H_2-ultilising archaeon, perhaps, in return for essential cellular building
633 blocks such as lipids and vitamins (Ver Eecke et al. 2012; Topçuoğlu et al. 2016;
634 Imachi et al. 2020; Yu et al. 2024).
635

636 Likewise, it is interesting to note that our ancestral reconstructions infer a complex
637 ancestor of Asgard archaea as was observed previously (Eme et al. 2023). Further,
638 our analyses predict that the Asgard archaeal ancestor has encoded a Wood-Lundgal
639 pathway (WLP) (**Figure 5, Supplementary Data 13**), i.e. a pathway that can serve as
640 electron sink supporting both autotrophic and heterotrophic growth modes (Ragsdale
641 and Pierce 2008; Schuchmann and Müller 2014; Schuchmann and Müller 2016).
642 However, this pathway seems to have been lost in various members of the Asgard
643 archaea (Spang et al. 2019; Liu et al. 2021) (**Supplementary Data 10-13**); specifically,
644 our analyses indicate that genes for proteins involved in the WLP were lost on the
645 branch leading to the Njordarchaeia after their split from Heimdall- and
646 Wukongarchaeia (**Figure 5**). The absence of the WLP as electron sink may have
647 strengthened the dependency of members of this group on syntrophic H_2-ultilising
648 partners, such as Thermoproteota (Huber et al. 2000; Sakai, Nakamura, et al. 2022;
649 Leung et al. 2024) that could serve as external electron sinks. It is tempting to
650 speculate that such a syntrophic lifestyle may have permitted the loss of genes
651 involved in various biosynthetic pathways in the ancestor of the *Panguiarchaeum*
652 clade (**Figure 4,5**), increasing the dependency of its members on partner organisms.
653
654 This scenario on the evolution of a potentially obligate symbiont lineage among the
655 Asgard archaea points to interesting parallels with the evolutionary trajectory leading
656 to the origin of eukaryotes. Many current models on eukaryogenesis hypothesise
657 syntrophy as a key driver for the establishment of an intricate symbiotic relationship
658 between the archaeal ancestor of eukaryotes (Martin and Müller 1998; Spang et al.
659 2019; Imachi et al. 2020; López-García and Moreira 2020; Vosseberg et al. 2024), i.e.
660 a likely sister lineage of the Hodarchaeales (Williams et al. 2020), and the
661 alphaproteobacterial partner that evolved into the mitochondrion (Lane and Martin
662 2010; Poole and Gribaldo 2014; Ettema 2016; Martijn et al. 2018; Muñoz-Gómez et al.
663 2022). Similar to *Panguiarchaeum*, the eukaryotic host lineage also appears to have
664 lost ancestral archaeal genes during eukaryogenesis, i.e. the transition from the First
665 to the Last Eukaryotic Common Ancestor. For example, glycolysis and iron-sulphur
666 cluster biosynthesis in eukaryotes are implemented by pathways with patchwork
667 ancestries that include both archaeal, alphaproteobacterial and other bacterial
668 contributions (Freibert et al. 2017), while the archaeal TCA cycle was replaced by that
669 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; Vosseberg et al. 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 hypothesise 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 characterisation.

Materials and Methods

Marker gene inspection and ranking analysis

Marker gene inspection: To accurately place Njordarchaeales and Panguiarchaeals 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 cluster of orthologs (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 four other previous studies (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 1325 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 were reported in some archaea (Werner 2007), we used a custom HMM

704 profile of COG0085 and COG0086 to detect “split” subunits and concatenate these
705 subunits before alignment (https://doi.org/10.5281/zenodo.15477107).
706
707 We then aligned sequences assigned to these 112 marker gene families using mafft
708 v7.453 (Katoh and Standley 2013) and removed poorly aligned amino acid sites with
709 BMGE 1.12 (Criscuolo and Gribaldo 2010) (settings: -m BLOSUM30 -h 0.55). The
710 maximum likelihood trees for these marker genes were inferred using IQ-tree v2.1.2
711 (Minh et al. 2020) (-m LG+G -B 1000). We decorated the tips of these marker gene
712 trees with annotations from KEGG orthology (Aramaki et al. 2020), PFAM (Bateman
713 et al. 2004), COG (Galperin et al. 2021) and archeal COG (arCOG) (Makarova et al.
714 2015) information. We manually inspected all single protein phylogenies to identify
715 marker gene families that did not meet archaeal monophyly and excluded those
716 families from subsequent analyses. Furthermore, we identified and removed
717 paralogues and sequences that experienced horizontal gene transfer (HGT) events
718 from bacteria to archaea. In total, this yielded 100 COG marker gene families.
719
720 We performed a second round of inspection and curation based on corresponding
721 arCOG families of these 100 COG families except COG0085 and COG0086 (See
722 above). In this round, we only included the marker gene families if homologs were
723 present in at least 60% of archaeal taxa and were duplicated in less than 20% of the
724 genomes of major archaeal lineages (e.g., Asgard archaea, Thermoproteota and
725 Halobacteriota), which resulted in 90 unique families with arCOG identifiers.
726 Sequences were realigned with mafft-linsi v7.453, and poorly aligned sites were
727 removed using BMGE as described above. The initial ML trees were inferred using IQ-
728 tree v 2.1.2 (Minh et al. 2020) (-m LG+G -B 1000). These trees were then used as
729 guide tree to perform phylogenetic inference based on the best-fitting model
730 (Kalyaanamoorthy et al. 2017) (settings: -mset LG -madd
731 LG+C10,LG+C20,LG+C30,LG+C40,LG+C50,LG+C60,LG+C10+R+F,LG+C20+R+F,
732 LG+C30+R+F,LG+C40+R+F,LG+C50+R+F,LG+C60+R+F --score-diff all) with 1000
733 ultrafast bootstraps. We inspected the single gene trees to identify and remove
734 sequences which seem to represent paralogues and/or were affected by long-branch
735 attraction (LBA) artefacts or HGT.
736

737 **Ranking analysis:** We used a previously developed marker gene ranking procedure
738 to rank each of the 90 markers based on the extent to which they recovered
739 established archaeal phylum- or order-level lineages (Dombrowski et al. 2020; Moody
740 et al. 2022). In brief, the number of splits (i.e., the occurrence of a certain taxon failing
741 to group within its expected taxonomic clade, **Supplementary Data 16**) was counted
742 across all bootstrap trees for each marker gene. We then defined the highest ranking
743 (top 50%, 43 markers) marker sets based on the number of splits per phylogenetic
744 cluster and the total number of splits normalised by the number of genomes within
745 each tree.
746
747 **Concatenation, site filtration and species tree inference**
748 **Concatenated gene trees using taxon set 1: i.e. 966 taxa:** The top 50% marker
749 gene sequences were aligned with mafft-linsi v7.453 (Katoh and Standley 2013) and
750 poorly aligned amino acid sites were removed with BMGE 1.12 (Criscuolo and
751 Gribaldo 2010) (settings: -m BLOSUM30 -h 0.55). Genomes with more than 50% gaps
752 in the alignment were further removed. The trimmed sequences of these 45 markers
753 were then subsequently concatenated into a supermatrix comprising 966 taxa. Two
754 additional supermatrixes were constructed by either excluding DNA topoisomerase VI
755 subunit A (arCOG04143) alone or together with DNA topoisomerase VI subunit B
756 (arCOG01165) for the following reasons (See also **Supplementary Information 1.1**):
757 1. The mixing of orthologues of arCOG04143 provides spurious support for the
758 relationship between Njordarchaeia and Asgard archaea.
759 2. The interaction between DNA topoisomerase VI subunits might lead to co-
760 evolution of some amino acid residues.
761
762 Concatenated gene trees both with and without DNA topoisomerase VI subunits were
763 inferred using IQ-Tree v2.1.2 (Minh et al. 2020) under LG+C60+G+F model with
764 posterior mean sites frequency approximation (Wang et al. 2018) with guide trees
765 inferred under LG+G+F model.
766
767 **Concatenated gene trees using taxon set 2: i.e. 303 taxa:** To resolve the
768 relationship between Njordarchaeales, Panguiarchaeales, Korarchaeota and Asgard
769 archaea, we next excluded divergent sequences from DPANN MAGs/genomes and
770 downsampled the Thermoproteota (previously TACK), and Halobacteriota,

771 Thermoplasmatota, Hydrothermarchaeota, Methanobacteriota, Methanobacteriota_A
772 and Methanobacteriota_B lineages (previously "Euryarchaea") based on the predicted
773 genome completeness and contamination (completeness-5 X contamination) (Parks
774 et al. 2017) to create a dataset of 303 taxa by one-per-
775 Thermoproteota/Asgardarchaeota-family and one-per-"Euryarchaea"-order citeria.
776 Marker gene sequences were realigned with mafft-linsi v7.453 (Katoh and Standley
777 2013) and trimmed with BMGE 1.12 (Criscuolo and Gribaldo 2010) (settings: -m
778 BLOSUM30 -h 0.55). We concatenated three supermatrices by including all top 50%
779 ranked markers and excluding arCOG04143 alone or with arCOG01165. Trees were
780 inferred under the LG+G+F and LG+C60+G+F model, as well as
781 LG+EDM0256LCLR+G+F model using IQ-Tree v2.1.2 (Minh et al. 2020) and IQ-Tree
782 v2.2.2.7 (Minh et al. 2020) respectively (settings: -B 1000 -alrt 1000, see also 71 taxa).
783 Focusing on the supermatrix of 303 archaeal taxa, excluding DNA topoisomerase
784 subunits, we progressively removed 10% to 50% of fastest evolving sites according to
785 the empirical Bayesian rate under LG+C60+G+F model (settings –rate), and 5%, 10%
786 to 50% most heterogeneous sites using Alignment_pruner.pl
787 (https://github.com/novigit/davinciCode/blob/master/perl). Trees were inferred under
788 either LG+C60+G+F or LG+EDM0256LCLR+G+F models (settings: -B 1000 -alrt
789 1000).
790
791 **Concatenated gene trees using taxon set 3: i.e. 71 taxa:** A smaller dataset of 71
792 taxa was created based on the 303 taxa set (set 2) for Bayesian phylogenetic analyses.
793 In particular, we kept one genome of "Euryarchaea" for each GTDB R207 class as an
794 outgroup and one genome of thermoproteotal (TACK) lineages for each GTDB R207
795 order considering predicted genome completeness and contamination (completeness-
796 5xcontamniation) (Parks et al. 2017). Marker gene sequences were realigned with
797 mafft-linsi v7.453 (Katoh and Standley 2013) and trimmed with BMGE 1.12 (Criscuolo
798 and Gribaldo 2010) (settings: -m BLOSUM30 -h 0.55). We concatenated only one
799 supermatrix without both DNA topoisomerase subunits (arCOG04143 and
800 arCOG01165). ML trees were inferred under LG+G+F and LG+C60+G+F model using
801 IQ-Tree v2.1.2 (Minh et al. 2020) (settings: -B 1000 -alrt 1000).
802
803 Additionally, we performed a comprehensive phylogenetic analysis using a recently
804 developed CAT-PMSF pipeline (Szánthó et al. 2023). In particular, we used three

805 different guide trees representing different hypotheses regarding the relationship of
806 Njordarchaeia to Korarchaeota and Asgard archaea, namely, Korarchaeota topology:
807 Njordarchaeia-sister-to-Korarchaeota; Asgard archaea topology: Njordarchaeia-
808 sister-to-Wukong-Heimdallarchaeia and Asgard archaea sister topology:
809 Njordarchaeia-sister-to-Asgard-Archaea to test if used guide trees affect topology
810 inference. We then used these 3 different starting topologies in a subsequent bayesian
811 analysis with the CAT model in combination with either Poisson or LG
812 exchangeabilities and a 4 discrete categories gamma rate model. We ran four different
813 Markov chains for each topology representing different hypotheses under the
814 Poisson+CAT+G model in Phylobayes v 1.8 (Lartillot et al. 2013). Additionally, four
815 different Markov chains were run for both Asgard archaea and Korarchaeota topology
816 under the LG+CAT+G model. These chains were run until the effective sample size of
817 all parameters was above 100 and/or the relative divergence of all parameters was
818 below 0.3, or visual inspection with Tracer 1.7 (Rambaut et al. 2018) indicated
819 convergence (**Supplementary Data 17**). Next, we extracted posterior mean site-
820 specific stationary distributions of amino acids for each chain using readpb_mpi
821 (Lartillot et al. 2013), which were converted to site-specific frequency profile using the
822 python script convert-site-dist.py (https://github.com/drenal/cat-pmsf-
823 paper/blob/main/scripts/convert-site-dists.py) (Szánthó et al. 2023). The site-specific
824 frequency profile was then used to infer ML trees using IQ-Tree v2.1.2 (settings: -
825 fs .sitefreq -B 1000 -alrt 1000).
826
827 We next applied the EDCluster algorithm (Schrempf et al. 2020) on one of the
828 LG+CAT+G4 chains that yielded the highest likelihood under LG+CAT-PMSF model
829 to infer EDM with various components (8, 16, 32, 64, 128 and 256) using log centred
830 log-ratio transformation (LCLR). ML trees were then inferred under different
831 components of EDM models using IQ-Tree v2.2.2.7 (e.g., settings: -m
832 LG+EDM0256LCLR+G+F -B 1000 -alrt 1000) and universal distribution mixture
833 models (UDM) with 128, 256 and 512 components (Schrempf et al. 2020). The model
834 fit between Poisson+CAT-PMSF, LG+CAT-PMSF, LG+G+F and LG+C60+G+F were
835 assessed using a recently developed parametric bootstrap method (Giacomelli et al.
836 2025), which compares the model adequacy in describing across-site compositional
837 heterogeneity.
838

839 **Compositional constrained analyses**

840 Constrained searches were performed for Asgard archaea (-sister) and Korarchaeota
841 topology based on the unconstrained ML tree under the same model using IQ-Tree
842 v2.1.2 (Minh et al. 2020) via the -g and -wslr flags. AU tests were performed using
843 consel v 1.20 (Shimodaira 2001) with 10000 bootstrap replicates. Site-wise log-
844 likelihood differences were calculated for Asgard archaea and Korarchaeota topology
845 under LG+G+F, LG+C60+G+F, Poisson+CAT-PMSF and LG+CAT-PMSF models to
846 investigate the attraction of Njordarchaeia towards Korarchaeota. The effective amino
847 acids (K_{eff}) values of mixture models and site-specific frequency profile were calculated
848 using the convert-site-dists-to-k_eff.py script (Schrempf et al. 2020; Szánthó et al.
849 2023). The site-wise log-likelihood differences were then grouped into different bins
850 according to partitions, evolving rates and the effective amino acids.

851

852 **RED values scaling and classification of ranks**

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

861

862 **Amino acid composition**

863 We used a custom Python script to estimate the relative frequency of each amino acid
864 in our 71 taxa dataset for the concatenation without DNA topoisomerase subunits
865 (https://doi.org/10.5281/zenodo.15477107). A principal component analysis (PCA)
866 was performed using Scikit-learn (Pedregosa et al. 2011), and two axes were plotted
867 along with eigenvectors using Seaborn (Waskom 2021). We identified
868 RYPEI(WKVL)/QSNT(DGC) as thermophilic enriched/depleted amino acid residues in
869 the concatenation using a binomial test (Baker et al. 2024) and modified the GFmix
870 (Muñoz-Gómez et al. 2022) model to account for the variation of the ratio
871 RYPEI(WKVL)/QSNT(DGC) at every branch. The likelihood of different tree topologies
872 under the GFmix-RYPEI(WKVL)/QSNT(DGC) model was then calculated with

LG+C60+G+F and LG+EDM0256LCLR+G+F, with weights of mixtures 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.15477107). 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 (Galperin et al. 2021) (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_nonredudant 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: --iprlookup --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 summarised for Njordarchaeales MAGs. Additionally, 1325 bacterial genomes/MAGs and 137 eukaryotic genomes and/or largely complete transcriptomes (Santana-Molina et al. 2025) (**Supplementary Data 18**) were searched against COG,

907 arCOG, PFAM and KO databases using the same pipeline. Njordarchaeia MAGs
908 annotation results were summarised in **Supplementary Data 7-11 and 19-21**.
909
910 **Metabolic comparison**
911 Metabolic comparisons were based on the annotation results described above. We
912 counted the occurrence of each gene across each MAG/genome and reported the
913 presence and absence pattern using Pandas (The pandas development team 2024)
914 in Python v3.9.0 (Van Rossum and Drake 2009). This data was then summarised per
915 GTDB R207 class level whenever possible. The Occurrence of ESPs was compared
916 using the results from asCOG searches. The presence and absence pattern of genes
917 across each database (COG_DB and ARCOG_DB) was analysed with PCA and t-
918 distributed Stochastic Neighbor Embedding (t-SNE) using scikit-learn (Pedregosa et
919 al. 2011).
920
921 **Contamination assessment of representative Njordarchaeia MAGs and its**
922 **impact on phylogenetic analyses**
923 Metagenomes of the representative MAGs and similar samples were downloaded from
924 public database (Dombrowski et al. 2018; Qu et al. 2023). These metagenomes
925 include samples from hydrothermal vents in Guaymas Basin (Dombrowski et al. 2018)
926 and a hot spring in Yunan, China (Qu et al. 2023). We used the bowtie2 v2.5.3
927 (Langmead and Salzberg 2012) to map reads to the MAGs, and samtools v1.18 to
928 process the mapped profiles. We used Anvi'o v8 (Eren et al. 2021) to perform
929 hierarchical clustering based on sequence composition and coverage patterns across
930 different metagenomes. Clusters were identified based on the similarity of coverage
931 profile across different samples while taking into account sample similarity. Marker
932 genes used in our phylogenetic analyses and homologs from marker sets in a recent
933 study (Zhang et al. 2025) were mapped back to the contigs using sequences selected
934 for phylogenetic analyses and annotations described above.
935
936 **Single gene tree inferences for ancestral reconstruction**
937 The predicted protein sequence of 303 archaeal genomes/MAGs was searched
938 against the arCOG database using hmmsearch v3.3.2 (Finn et al. 2011) (settings: --
939 tblout –domtblout –notextw). We used the best hits (the lowest evalue and highest
940 bitscore, cutoff: e-value < 1E-3) to identify and split potential fused proteins.

941 Specifically, we first subtracted the position of the first domain hit from the full protein
942 using bedtools subtract (Quinlan and Hall 2010) (v2.26.0). Secondly, we investigated
943 the presence of a secondary domain hit that is assigned to a different arCOGs than
944 the first domain. Subsequently, we repeated these two steps until all the proteins were
945 investigated for up to four domains. Finally, the individual domains from the original
946 protein were extracted using the positional domain information from hmmsearch
947 output using bedtools getfasta function. To keep track of unsplit and split domains from
948 the original protein, we assigned 'a0' notation for unsplit proteins and 'a1' to 'a4' for
949 split domains (primarily to quarterly) in the generated trees.
950
951 We then combined all protein sequences and removed sequences with ambiguous
952 amino acids using a custom Python script (remove_seq_with_specific_char.py,
953 https://doi.org/10.5281/zenodo.15477107). To improve the phylogenetic signal in the
954 single gene trees, we carefully performed three rounds of alignments before the final
955 phylogenetic inference. First, we aligned sequences using mafft or mafft-linsi, for
956 sequences with greater than or less than 1000 sequences, respectively. Alignments
957 were then trimmed using trimAl 1.2rev59 (Capella-Gutiérrez et al. 2009) with –
958 gappyout flag. Trimmed sequences with greater than or equal to 50% gaps were
959 removed using a custom Python script (faa_drop.py,
960 https://doi.org/10.5281/zenodo.15477107). Second, we realigned the arCOG gene
961 families (with more than or equal to 4 sequences) using MAFFT as described above
962 and trimmed the alignment with BMGE (v1.12; settings: -m BLOSUM30 -b 2 -h 0.55)
963 (Criscuolo and Gribaldo 2010). We then inferred a phylogenetic tree using IQ-Tree
964 v2.1.2 (settings: -m LG+G) (Minh et al. 2020). To alleviate the effect of long-branching
965 sequences on phylogenetic inferences, we identified sequences/clusters of
966 sequences on long branches of the inferred tree using a custom script
967 (cut_gene_tree.py, https://doi.org/10.5281/zenodo.15477107) with a branch cutoff of
968 two (Davín et al. 2025). Single sequences were discarded, and clusters with more than
969 four sequences were separated into different gene families. We appended the notation
970 "_ [0-3]" to keep track of the separation of sequences into a new cluster. Finally, we
971 aligned and trimmed the sequences as described in the second step. For each family,
972 we selected the best-fit model using the model test with IQ-Tree v2.1.2 (settings: -m
973 MF -mset LG -madd LG+C10, LG+C10+G, LG+C10+R, LG+C10+F, LG+C10+R+F,
974 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 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 6077 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 (Szöllősi et al. 2015) program, we first reconciled the gene trees against species tree using the default parameters, which assumes 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 (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 maximising the total reconciliation likelihood over all gene families in that category using two Python scripts (setup_OR_estimation.py and O_R_optimisation.py, https://doi.org/10.5281/zenodo.15477107). 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 13**. 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 (1000 times).

1009 **Phylogenetic analyses of DNA topoisomerase subunit A and DNA**
1010 **topoisomerase B**
1011 The homologs of DNA topoisomerase subunit A and B, i.e. COG1389 and COG1697,
1012 were identified and retrieved from 966 archaeal, 1325 bacterial and 137 eukaryotic
1013 genomes/MAGs or largely complete transcriptomes. Sequences of each gene family
1014 were first aligned with mafft-linsi v7.453, and poorly aligned alignment sites were
1015 trimmed with BMGE v1.2 (Criscuolo and Gribaldo 2010) (Settings: -m BLOSUM30 -h
1016 0.55). The inference of phylogenetic trees for individual maker gene families was
1017 conducted as follows: an initial ML tree was inferred under LG+G with IQ-Tree v 2.1.2
1018 (Minh et al. 2020). Single gene trees were manually inspected to remove putative LBA
1019 artefacts, and the remaining sequences for each gene were realigned and trimmed as
1020 described above; finally, an ML tree was inferred for each gene under
1021 LG+C60+G+F+PMSF (Wang et al. 2018) model (settings: -B 1000 -alrt 1000) using
1022 guide tree inferred from LG+F+G.
1023
1024 **Phylogenetic analyses of Actin homologs**
1025 The actin homologs were identified from 966 archaeal, 1325 bacterial and 137
1026 eukaryotic genomes/MAGs or largely complete transcriptomes using IPR00400. The
1027 cell shape determining protein MreB homologs, a distant homolog to actin, in 966
1028 archaeal genomes were identified using COG1077 and dereplicated using cd-hit v4.7
1029 (-c 0.75) (Fu et al. 2012). These sequences were combined and aligned with mafft-
1030 linsi v7.453 (Katoh and Standley 2013). Poorly aligned sequences were trimmed using
1031 BMGE (Criscuolo and Gribaldo 2010) (settings: -m BLOSUM30 -h 0.55). An initial ML
1032 tree was inferred under LG+G using IQ-Tree v 2.1.2 (Minh et al. 2020), and sequences
1033 showing potential instances of LBA artefacts were removed; the remaining sequences
1034 were then realigned and trimmed as described above; finally, an ML tree was then
1035 inferred under LG+C60+G+F+PMSF model (settings: -B 1000 -alrt 1000) using a guide
1036 tree inferred from LG+F+G. 20% of the most heterogeneous sites were removed using
1037 alignment_pruner.pl and an ML tree was inferred in the same way under
1038 LG+C60+G+F+PMSF using a guide tree inferred from LG+F+G
1039
1040 **Phylogenetic analyses of ribosomal protein L28**
1041 The ribosomal protein L28/Mak16 homologs (i.e. PF01778) were retrieved from 966
1042 archaeal, 1325 bacterial and 137 eukaryotic genomes/MAGs or largely complete

1043 transcriptomes (evalue: 0.0001). Sequences were first aligned with mafft-linsi v7.453,
1044 and gappy columns were removed using trimal 1.2rev59 (Capella-Gutiérrez et al. 2009)
1045 with -gappyout flag. An initial ML tree was inferred under LG+G using IQ-Tree v2.1.2
1046 (Minh et al. 2020), and sequences showing potential instances of LBA artefacts were
1047 removed; remaining sequences were then realigned and trimmed as described above;
1048 finally, an ML tree was inferred under LG+C60+G+F+PMSF model (settings: -B 1000
1049 -alrt 1000) using the guide tree inferred from LG+G.
1050
1051
1052 **Phylogenetic analyses of oligosaccharyltransferase 3/6 like (OST3/6 like) and**
1053 **translocon-associated protein (TRAP) complexes subunit beta**
1054
1055 The OST3/6-like homologs and TRAP beta were retrieved from 966 archaeal, 1325
1056 bacterial and 137 eukaryotic genomes/MAGs or largely complete transcriptomes using
1057 PF04756 and PF05753 respectively, from the annotation results described above.
1058 These sequences were dereplicated using cd-hit v4.7 (Fu et al. 2012) (-c 0.7).
1059 Sequences of each gene family were first aligned with mafft-linsi v7.453, and gappy
1060 columns were removed using trimal 1.2rev59 (Capella-Gutiérrez et al. 2009) with -
1061 gappyout flag. An initial ML tree was inferred under LG+G using IQ-Tree v2.1.2 (Minh
1062 et al. 2020) for each gene family, and sequences showing potential instances of LBA
1063 artefacts were removed; remaining sequences for each gene family were then
1064 realigned and trimmed as described above; finally an ML tree was then inferred for
1065 each gene under LG+C60+G+F+PMSF model (settings: -B 1000 -alrt 1000) using
1066 guide tree inferred from LG+G.
1067
1068 **Phylogenetic analyses of the Snf7 family**
1069 The Snf7 family homologs were identified and retrieved from 966 archaeal and 137
1070 eukaryotic genomes/MAGs based on the presence of the domain IPR005024.
1071 Eukaryotic sequences were dereplicated using cd-hit v4.7 (Capella-Gutiérrez et al.
1072 2009) (-c 0.7). We created two datasets for this family, i.e. a dataset with only archaeal
1073 sequences and another with archaeal and eukaryotic sequences. Sequences of each
1074 dataset were first aligned with mafft-linsi v7.453 (Katoh and Standley 2013), and poorly
1075 aligned sites were removed using BMGE v1.1.2 (Criscuolo and Gribaldo 2010)
1076 (settings: -m BLOSUM30 -h 0.55). The inference of phylogenetic trees was conducted

1077 as follows: an initial ML tree was inferred under LG+G using IQ-Tree v 2.1.2 (Minh et
1078 al. 2020), and sequences showing potential instances of LBA artefacts were removed;
1079 remaining sequences were then realigned and trimmed as described above; finally,
1080 an ML tree was then inferred funder the LG+C60+G+F+PMSF model (settings: -B
1081 1000 -alrt 1000) using a guide tree inferred from LG+G. We also removed the 20%
1082 most heterogeneous sites from the alignment with both archaeal and eukaryotic
1083 sequences included using Alignment_pruner.pl
1084 (https://github.com/novigit/davinciCode/blob/master/perl) and inferred an ML tree
1085 using the same approach.
1086
1087
1088 **Phylogenetic analyses of CDP-archaeol synthase (CarS), archaetidylinositol**
1089 **phosphate synthase (PgsA) and phosphatidylserine synthase (PssA)**
1090 Sequences identified as CarS (K19664), PgsA (K17884), and PssA (K17103) were
1091 retrieved from 966 archaeal, 1325 bacterial and 137 eukaryotic genomes/MAGs or
1092 largely complete transcriptomes, respectively, from the annotation results described
1093 above. Sequences of each gene family were first aligned with mafft-linsi v7.453 (Katoh
1094 and Standley 2013), and gappy columns were removed using trimal 1.2rev59
1095 (Capella-Gutiérrez et al. 2009) with the -gappyout flag. The inference of phylogenetic
1096 tree for individual gene families was conducted as follows: an initial ML tree was
1097 inferred under LG+G using IQ-Tree v2.1.2 (Minh et al. 2020) for each gene family, and
1098 sequences showing potential instance of LBA artefacts were removed; remaining
1099 sequences for each gene were then realigned and trimmed as described above; finally
1100 an ML tree was then inferred for each gene under LG+C60+G+F+PMSF model
1101 (settings: -B 1000 -alrt 1000) using guide tree inferred from LG+G.
1102
1103 **Phylogenetic analyses of [NiFe] group hydrogenase, large subunit**
1104 Backbone sequences for phylogenetic analyses of [NiFe] group 3 and 4 hydrogenases
1105 were obtained from a previous study (Spang et al. 2019). Njordarchaeia homologs
1106 were added to the backbone sequences and aligned with mafft-linsi v7.453. Poorly
1107 aligned sites were removed using BMGE v1.1.2 (Criscuolo and Gribaldo
1108 2010)(settings: -m BLOSUM30 -h 0.55). A phylogenetic tree was inferred under the
1109 LG+C60+G+F+PMSF model guide trees inferred from the LG+G+F model using IQ-

1110 Tree v2.1.2 (Minh et al. 2020) (settings: -B 1000 -alrt 1000). Phylogenetic tree was
1111 visualised using ggtree 3.2.0 (Xu et al. 2022).
1112
1113 **Network inferences**
1114 The input dataset for network prediction was obtained from OTU community profiles
1115 derived from 248,559 NCBI metagenomes (Woodcroft et al. 2024) from the 'Sandpiper'
1116 website (https://sandpiper.qut.edu.au), which were annotated with the GTDB R214
1117 taxonomy (Parks et al. 2022). First, we used the known association between
1118 *Huberarchaeum* and *Altiarchaeum* to confirm that our network association approach
1119 is able to detect symbiotic interactions. We selected all samples that contain taxa
1120 annotated as *Huberarchaeum* (67 samples), respectively. We transformed the
1121 coverage of the taxa in these samples to counts by multiplying with 100 and formatted
1122 the data into a table listing files in rows and all OTUs in columns. To reduce the scarcity
1123 of the dataset, we filtered for OTUs present in 10%, 20%, 30%, 40%, 50%, 60%, 70%,
1124 80%, 90% of the samples. To estimate the minimum number of samples required for
1125 robust, true positive prediction of *Huberarchaeum-Altiarchaeum* associations, we drew
1126 random samples and applied two different filter criteria. Numbers of 10, 20, 30, 40, 50,
1127 and 60 samples were drawn; the random data sets were filtered for OTUs detected in
1128 at least 50% and 80% of all random samples, respectively. For each condition, ten
1129 networks were inferred. Statistics were run across all ten networks within one condition.
1130 These networks were inferred using SpiecEasi v1.1.3 (Kurtz et al. 2015). Across all
1131 networks, we checked for associations between *Huberarchaeum* and *Altiarchaeum*. In
1132 all networks inferred (>30 samples), the expected association was robustly detected.
1133 However, below a number of 30 samples, the association was detected only in 50%
1134 of the samples. This clearly indicates a minimum number of > 30 samples (equaling
1135 40% of samples) that is necessary for reliable inference of association.
1136 We also attempted to confirm the known association between *Nanoarchaeum* and its
1137 host *Ignicoccus* (order Sulfolobales). Our dataset contained several OTUs, assigned
1138 to the genus *Nanoarchaeum* (otu_658, otu_3077) and the genus *Ignicoccus* (otu_3007,
1139 otu_678, otu_3062, otu_3063, otu_3922, otu_4367). However, some of the taxa were
1140 present in only very few samples (**Supplementary Figure 35,36)**, and subsequently
1141 in the inferred network, no association between any of these taxa was predicted.
1142 Interestingly, some OTUs showed strong, significant negative correlations
1143 (**Supplementary Figure 35,36**), suggesting a high specificity of the symbiotic

association on 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 order Njordarchaeales proposed in this manuscript (**Figure 1BC**), 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 R 2024).

Data availability

All genomic data/transcriptomic data analysed in this study are available at NCBI (Supplementary Data 16, 18) and are deposited in the data repository (https://doi.org/10.5281/zenodo.15477107). 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.15477107) under the following license CC BY 4.0.

Code availability

Workflows for annotations, phylogenies, reconciliation, and custom scripts (bash/python) to analyse 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.15477107).

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 947317, ASymbEL to A.S. and grant agreement No. 714774, GENECLOCKS to G.J.S.). Furthermore, this work was supported by the Gordon and Betty Moore Foundation (GBMF9741 to T.A.W., A.S., and G.J.S.) 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. We also want to thank Tara Mahendrarajah for valuable feedback on the manuscript.

Author contributions

W.H. T.A.W. and A.S. conceived the study; W.H. analysed and annotated genomes and generated accompanying data tables; W.H. and A.S. analysed 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.E., and A.S. interpreted data; W.H., T.W., and A.S wrote and all authors edited and approved the manuscript.

Competing interests

The authors declare no competing interests.

1208
1209
1210
1211
1212
1213
1214
1215 **Figures and legends**
1216
1217 **Figure 1**

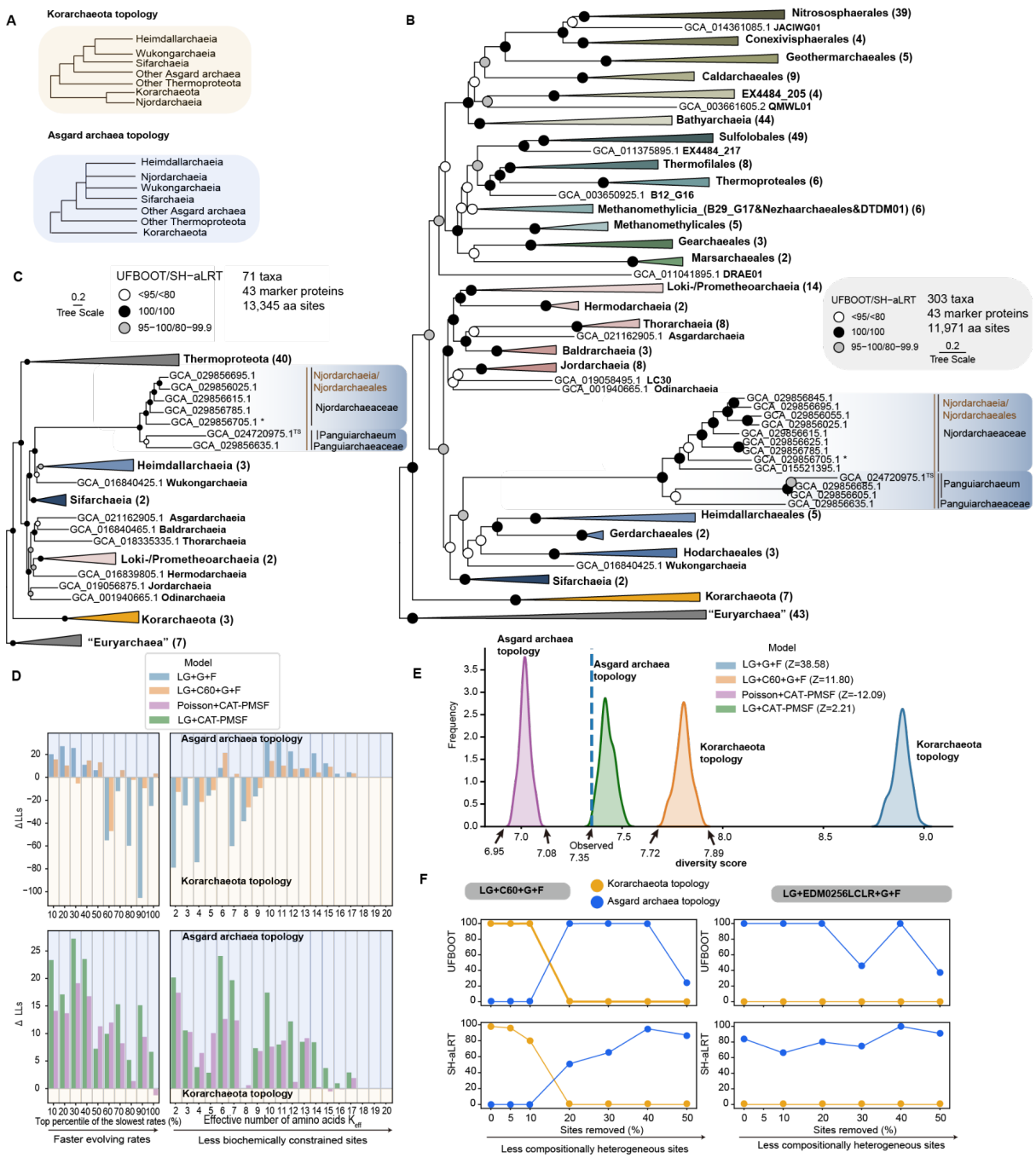


Figure 1. The phylogenetic placement of Njordarchaeia. A. Schematic of the phylogenetic hypotheses tested in this study. Korarchaeota topology: Njordarchaeia branch sister to Korarchaeota. Asgard archaea topology: Njordarchaeia branch within Asgard archaea. **B**. Maximum-likelihood phylogenetic analysis 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:

1228 average substitutions per site. The number of taxa in each collapsed clade is shown
1229 by the number in parenthesis next to the clade name. GCA_029856705.1* is the
1230 closest genome to the first proposed type material of *Njordarchaeum guaymasis* (See
1231 Supplementary Discussion). The GCA_024720975.1^{TS} is the proposed type material
1232 for *P.* symposium. The tree is rooted by "Euryarchaea" (including Halobacteriota,
1233 Thermoplasmatota, Hydrothermarchaeota, Methanobacteriota, Methanobacteriota_A
1234 and Methanobacteriota_B). **C**. Maximum-likelihood phylogenetic analysis of the
1235 concatenated 50% top-ranked marker proteins (n=43) and streamlined set3 (71 taxa)
1236 using LG+CAT-PMSF model. Scale bar: average substitutions per site. The number
1237 of taxa in each collapsed clade is shown by the number in parenthesis next to the
1238 clade name. **D**. Evaluation of phylogenetic signals for Asgard archaea and
1239 Korarchaeota topology under LG+G+F, LG+C60+G+F, Poisson+CAT-PMSF and
1240 LG+CAT-PMSF models, listed in an order of increasing model fit. Sites underlying the
1241 signals were binned by substitution rates (based on LG+C60+G+F) and effective
1242 amino acids per site (based on LG+CAT-PMSF). **E.** Distribution of amino acid diversity
1243 scores for the simulated alignments under LG+G+F, LG+C60+G+F, one
1244 representative Poisson+CAT-PMSF and LG+CAT-PMSF. The dashed line indicates
1245 the amino acid diversity score from the real dataset. *Z*-scores quantifying the
1246 difference between the compositional heterogeneity of the real dataset and that of the
1247 simulated datasets from each model are labelled. Note that *Z*-scores of LG+C60+G+F
1248 and Poisson+CAT-PMSF are similar but right tail of Poisson+CAT-PMSF is closer to
1249 real dataset, indicating a better fit. **F.** Impact of progressive removal of the most
1250 compositionally biased sites ranked by chi-square score on the statistical support of
1251 Asgard archaea and Korarchaeota topology. The support values, UFBOOT and SH-
1252 aLRT, were estimated on the 50% top-ranked markers of streamlined set2 (303 taxa)
1253 based on LG+C60+G+F and LG+EDM0256LCLR+G+F model.
1254

1255

1256

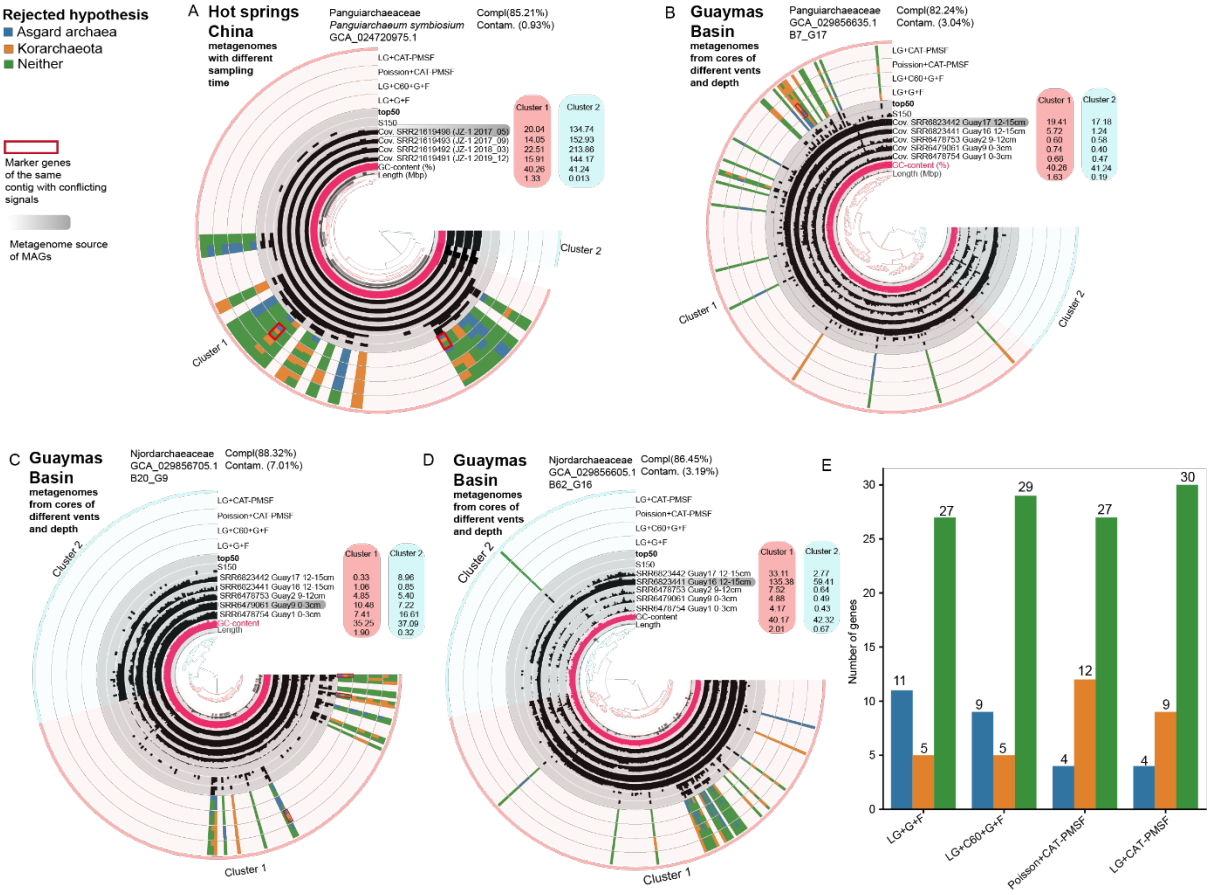


Figure 2. Hierarchical clustering of contigs in four MAGs from Panguiarchaeaceae (A, B) and Njordarchaeaceae (C, D) based on sequence composition and depth of coverage across different metagenomes. Note that the tip of hierarchical clustering tree represents splits of long contigs at a size of 20000 bp. This analysis demonstrates metagenomic contamination is not a source of conflicting phylogenetic signals. The ring indicating coverage across the source metagenomes is highlighted in a gray box. 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 red boxes. **E**. Barplot showing the number of genes that confidently reject Asgard archaea, Korarchaeota or neither of the two topologies under different models.

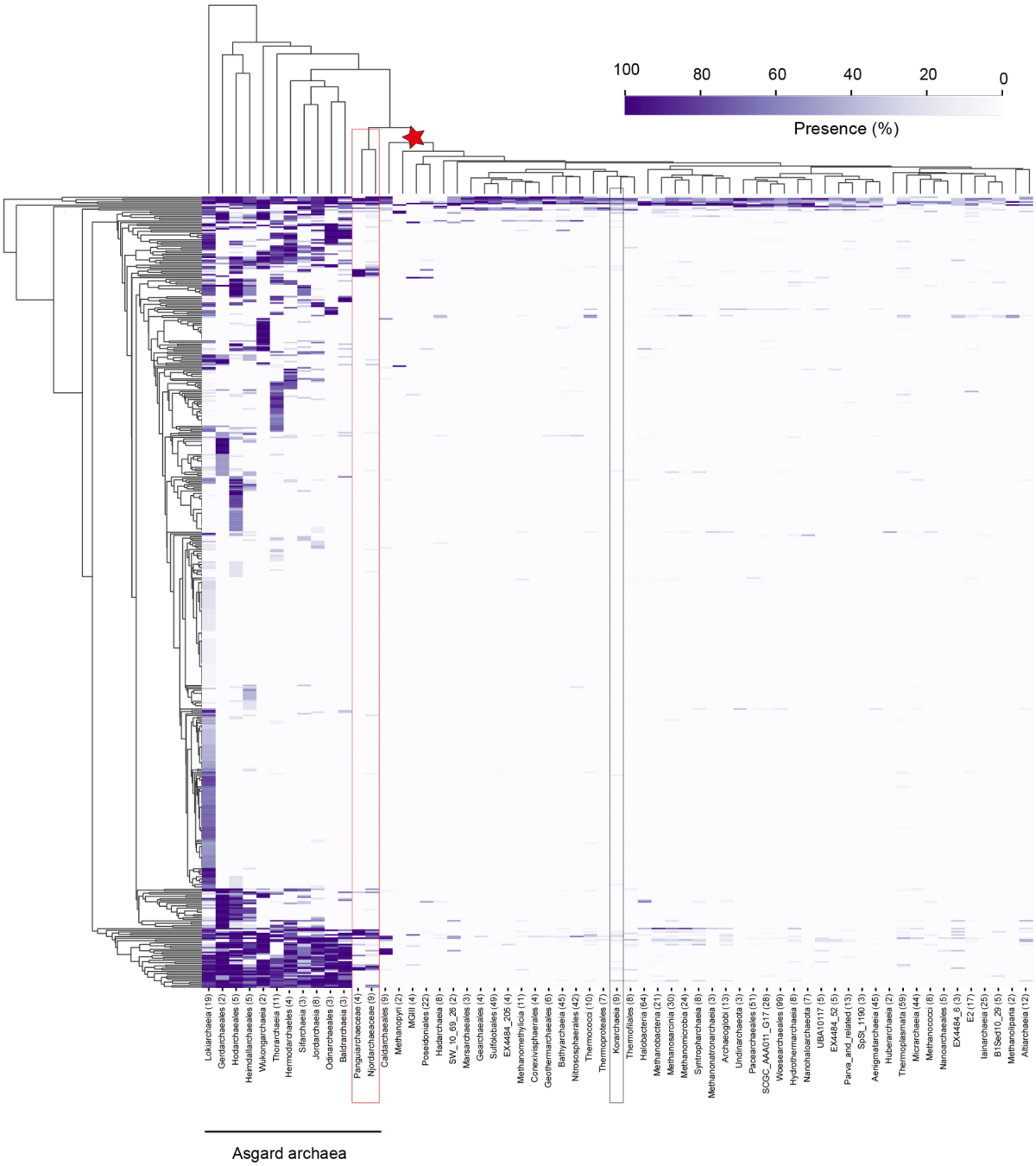


Figure 3 Presence of Asgard cluster of orthologues (AsCOGs) across major archaeal lineages. Panguiarchaeceae and Njordarchaeceae shared a number of AsCOGs homologs with other Asgard archaea albeit reduced, which is distinct from Korarchaeota. AsCOG presence was determined across 966 archaeal genomes, and the occurrence of AsCOGs within any taxonomic clade of interest (based on the genome taxonomy database) was recorded in percentage. Clusters with less than two genomes were not included. The number of genomes for each cluster is shown in parentheses.

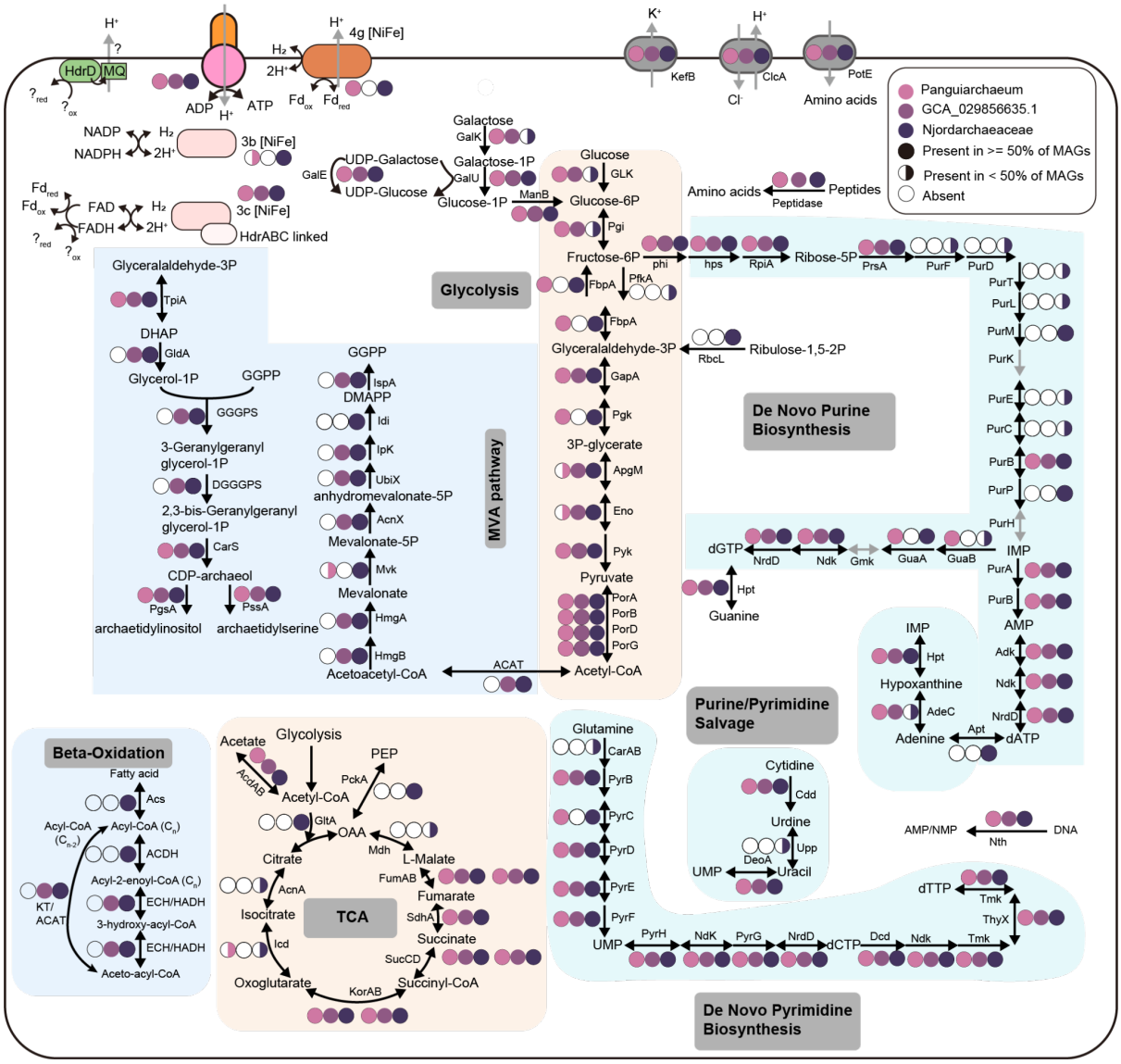


Figure 4. Metabolic characteristics of Njordarchaeia. Overall, Njordarchaeia are thermophilic fermentative heterotrophs, with core pathways of varying completeness suggesting some core metabolites must be obtained from other organisms in the environment. Full circles indicate that a gene is present in 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 MAGs. Open circles denote genes absent from all Njordarchaeia MAGs. A detailed list of genes encoded by Njordarchaeia can be found in Supplementary Data 7,8,9.

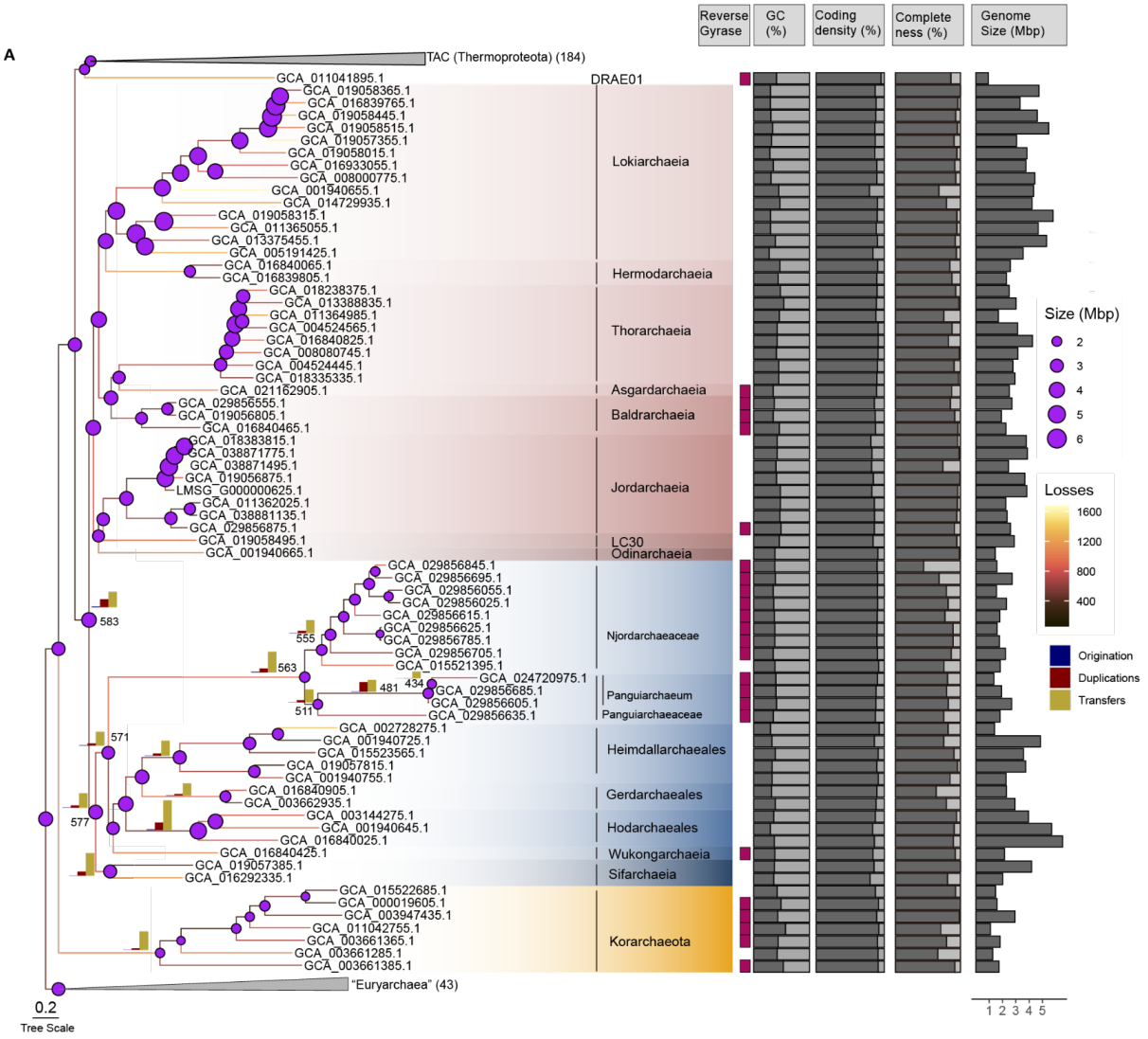
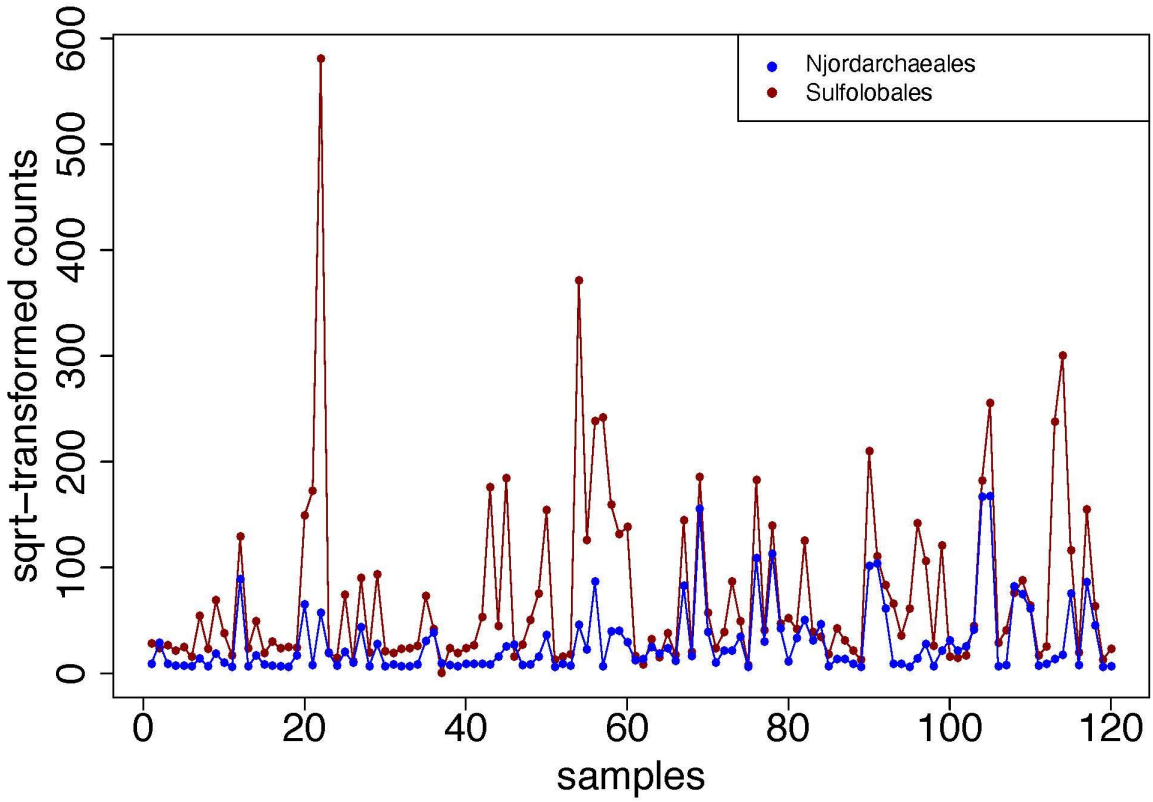


Figure 5. Gene-tree-aware ancestral reconstruction of *Panguiarchaeum* lineages. Njordarchaeia, in particular *Panguiarchaeum*, might have experienced genome reduction from their common ancestor with a larger ancestor Wukongarchaeia and Heimdallarchaeia (0.53-0.84 Mpb on the stem: node 571 to node 563). A. Inferred species of asgard archaeal, thermoproteotal and korarchaeotal lineages from Figure

1299 1B. The colours on the branches represent the number of gene losses, and the bar
1300 plots on the branches of interest show the number of originations, duplications and
1301 transfers. The size of circles at nodes represents the inferred ancestral genome size.
1302 The number of taxa in each collapsed clade is shown by the number in parenthesis
1303 next to the clade name. B. The presence probability of genes of interest at the
1304 ancestral nodes leading to *Panguiarchaeum*. Node numbers of ancestral nodes are
1305 shown and referred to in Panel A. Full circles indicate a presence probability (PP) ≥
1306 0.75, half circles show a PP< 0.75 but ≥ 0.5, while PP<0.5 is shown in open circles
1307 (See Supplementary Data 12).

1308
1309
1310
1311
1312



1313

1314 **Figure 6. Abundance-based co-occurrence of Njordarchaeales and Sulfolobales.**
1315 For plotting, the counts of both orders Njordarchaeales (including Panguiarcheaceae)
1316 and Sulfolobales, were summed in each sample and plotted. For visualisation
1317 purposes, the counts were square root transformed.

1318

1319

1320

1321

1322

1323

1324

1325

1326 **Reference**:

1327 Akıl C, Robinson RC. 2018. Genomes of Asgard archaea encode profilins that
1328 regulate actin. *Nature* 562:439–443.

1329 Akıl C, Tran LT, Orhant-Prioux M, Baskaran Y, Manser E, Blanchoin L, Robinson
1330 RC. 2020. Insights into the evolution of regulated actin dynamics via
1331 characterization of primitive gelsolin/cofilin proteins from Asgard archaea. *Proc.*
1332 *Natl. Acad. Sci. U. S. A.* 117:19904–19913.

1333 Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997.
1334 Gapped BLAST and PSI-BLAST: a new generation of protein database search
1335 programs. *Nucleic Acids Res.* 25:3389–3402.

1336 Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H.
1337 2020. KofamKOALA: KEGG Ortholog assignment based on profile HMM and
1338 adaptive score threshold. *Bioinformatics* 36:2251–2252.

1339 Baker BA, Gutiérrez-Preciado A, Rodríguez del Río Á, McCarthy CGP, López-García
1340 P, Huerta-Cepas J, Susko E, Roger AJ, Eme L, Moreira D. 2024. Expanded
1341 phylogeny of extremely halophilic archaea shows multiple independent
1342 adaptations to hypersaline environments. *Nature Microbiology* 9:964–975.

1343 Baker BJ, De Anda V, Seitz KW, Dombrowski N, Santoro AE, Lloyd KG. 2020.
1344 Diversity, ecology and evolution of Archaea. *Nat Microbiol* 5:887–900.

1345 Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A,
1346 Marshall M, Moxon S, Sonnhammer ELL, et al. 2004. The Pfam protein families
1347 database. *Nucleic Acids Res.* 32:D138–D141.

1348 Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK,
1349 Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, et al. 2017. Minimum
1350 information about a single amplified genome (MISAG) and a metagenome-
1351 assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 35:725–
1352 731.

1353 Brown AM, Hoopes SL, White RH, Sarisky CA. 2011. Purine biosynthesis in
1354 archaea: variations on a theme. *Biol Direct* 6:63.

1355 Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using
1356 DIAMOND. *Nat. Methods* 12:59–60.

1357 Cambillau C, Claverie JM. 2000. Structural and genomic correlates of
1358 hyperthermostability. *J Biol Chem* 275:32383–32386.

1359 Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for
1360 automated alignment trimming in large-scale phylogenetic analyses.
1361 *Bioinformatics* 25:1972–1973.

1362 Castelle CJ, Brown CT, Anantharaman K, Probst AJ, Huang RH, Banfield JF. 2018.
1363 Biosynthetic capacity, metabolic variety and unusual biology in the CPR and

DPANN radiations. *Nat. Rev. Microbiol.* 16:629–645.

Chen L-X, Anantharaman K, Shaiber A, Eren AM, Banfield JF. 2020. Accurate and complete genomes from metagenomes. *Genome Res* 30:315–333.

Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. 2023. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat. Methods* 20:1203–1212.

Coleman GA, Davín AA, Mahendrarajah TA, Szánthó LL, Spang A, Hugenholtz P, Szöllősi GJ, Williams TA. 2021. A rooted phylogeny resolves early bacterial evolution. *Science* 372:eabe0511.

Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* 10:210.

Davín AA, Woodcroft BJ, Soo RM, Morel B, Murali R, Schrempf D, Clark JW, Álvarez-Carretero S, Boussau B, Moody ERR, et al. 2025. A geological timescale for bacterial evolution and oxygen adaptation. *Science* 388:eadp1853.

Dombrowski N, Teske AP, Baker BJ. 2018. Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nat Commun* 9:4999.

Dombrowski N, Williams TA, Sun J, Woodcroft BJ, Lee J-H, Minh BQ, Rinke C, Spang A. 2020. Undinarchaeota illuminate DPANN phylogeny and the impact of gene transfer on archaeal evolution. *Nat. Commun.* 11:3939.

Elkins JG, Podar M, Graham DE, Makarova KS, Wolf Y, Randau L, Hedlund BP, Brochier-Armanet C, Kunin V, Anderson I, et al. 2008. A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc Natl Acad Sci U S A* 105:8102–8107.

Eme L, Spang A, Lombard J, Stairs CW, Ettema TJG. 2017. Archaea and the origin of eukaryotes. *Nat Rev Microbiol* 15:711–723.

Eme L, Tamarit D, Caceres EF, Stairs CW, De Anda V, Schön ME, Seitz KW, Dombrowski N, Lewis WH, Homa F, et al. 2023. Inference and reconstruction of the heimdallarchaeial ancestry of eukaryotes. *Nature* 618:992–999.

Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, et al. 2021. Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol* 6:3–6.

Ettema TJG. 2016. Evolution: Mitochondria in the second act. *Nature* 531:39–40.

Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39:W29–W37.

Forterre P. 2002. A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein. *Trends Genet* 18:236–237.

1402 Freibert S-A, Goldberg AV, Hacker C, Molik S, Dean P, Williams TA, Nakjang S,
1403 Long S, Sendra K, Bill E, et al. 2017. Evolutionary conservation and in vitro
1404 reconstitution of microsporidian iron–sulfur cluster biosynthesis. *Nature*
1405 *Communications* 8:1–12.

1406 Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-
1407 generation sequencing data. *Bioinformatics* 28:3150–3152.

1408 Galperin MY, Wolf YI, Makarova KS, Vera Alvarez R, Landsman D, Koonin EV.
1409 2021. COG database update: focus on microbial diversity, model organisms,
1410 and widespread pathogens. *Nucleic Acids Res.* 49:D274–D281.

1411 Giacomelli M, Vecchi M, Guidetti R, Rebecchi L, Donoghue PCJ, Lozano-Fernandez
1412 J, Pisani D. 2025. CAT-Posterior Mean Site Frequencies Improves Phylogenetic
1413 Modeling Under Maximum Likelihood and Resolves Tardigrada as the Sister of
1414 Arthropoda Plus Onychophora. *Genome Biol Evol* 17:evae273.

1415 Glasemacher J, Bock AK, Schmid R, Schønheit P. 1997. Purification and properties
1416 of acetyl-CoA synthetase (ADP-forming), an archaeal enzyme of acetate
1417 formation and ATP synthesis, from the hyperthermophile Pyrococcus furiosus.
1418 *Eur J Biochem* 244:561–567.

1419 Guy L, Ettema TJG. 2011. The archaeal “TACK” superphylum and the origin of
1420 eukaryotes. *Trends Microbiol.* 19:580–587.

1421 Guy L, Saw JH, Ettema TJG. 2014. The archaeal legacy of eukaryotes: a
1422 phylogenomic perspective. *Cold Spring Harb Perspect Biol* 6:a016022.

1423 Hatano T, Palani S, Papatziamou D, Salzer R, Souza DP, Tamarit D, Makwana M,
1424 Potter A, Haig A, Xu W, et al. 2022. Asgard archaea shed light on the
1425 evolutionary origins of the eukaryotic ubiquitin-ESCRT machinery. *Nat.*
1426 *Commun.* 13:3398.

1427 Huang W-C, Liu Y, Zhang X, Zhang C-J, Zou D, Zheng S, Xu W, Luo Z, Liu F, Li M.
1428 2021. Comparative genomic analysis reveals metabolic flexibility of
1429 Woesearchaeota. *Nature Communications* 12:1–14.

1430 Huber H, Burggraf S, Mayer T, Wyschkony I, Rachel R, Stetter KO. 2000. Ignicoccus
1431 gen. nov., a novel genus of hyperthermophilic, chemolithoautotrophic Archaea,
1432 represented by two new species, Ignicoccus islandicus sp nov and Ignicoccus
1433 pacificus sp nov. and Ignicoccus pacificus sp. nov. *Int J Syst Evol Microbiol* 50
1434 Pt 6:2093–2100.

1435 Huber H, Hohn MJ, Rachel R, Fuchs T, Wimmer VC, Stetter KO. 2002. A new
1436 phylum of Archaea represented by a nanosized hyperthermophilic symbiont.
1437 *Nature* 417:63–67.

1438 Hurtig F, Burgers TCQ, Cezanne A, Jiang X, Mol FN, Traparić J, Pulschen AA,
1439 Nierhaus T, Tarrason-Risa G, Harker-Kirschneck L, et al. 2023. The patterned
1440 assembly and stepwise Vps4-mediated disassembly of composite ESCRT-III
1441 polymers drives archaeal cell division. *Sci Adv* 9:eade5224.

1442 Imachi H, Nobu MK, Kato S, Takaki Y, Miyazaki M, Miyata M, Ogawara M, Saito Y,
1443 Sakai S, Tahara YO, et al. 2024. Promethearchaeum syntrophicum gen. nov.,
1444 sp. nov., an anaerobic, obligately syntrophic archaeon, the first isolate of the
1445 lineage "Asgard" archaea, and proposal of the new archaeal phylum
1446 Promethearchaeota phyl. nov. and kingdom Promethearchaeati regn. nov. *Int J*
1447 *Syst Evol Microbiol* 74:006435.

1448 Imachi H, Nobu MK, Nakahara N, Morono Y, Ogawara M, Takaki Y, Takano Y,
1449 Uematsu K, Ikuta T, Ito M, et al. 2020. Isolation of an archaeon at the
1450 prokaryote-eukaryote interface. *Nature* 577:519–525.

1451 Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J,
1452 Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function
1453 classification. *Bioinformatics* 30:1236–1240.

1454 Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017.
1455 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat.*
1456 *Methods* 14:587–589.

1457 Kapli P, Flouri T, Telford MJ. 2021. Systematic errors in phylogenetic trees. *Current*
1458 *Biology* 31:R59–R64.

1459 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version
1460 7: improvements in performance and usability. *Mol. Biol. Evol.* 30:772–780.

1461 Kato S, Ogasawara A, Itoh T, Sakai HD, Shimizu M, Yuki M, Kaneko M, Takashina
1462 T, Ohkuma M. 2022. Nanobdella aerobiophila gen. nov., sp. nov., a
1463 thermoacidophilic, obligate ectosymbiotic archaeon, and proposal of
1464 Nanobdellaceae fam. nov., Nanobdellales ord. nov. and Nanobdellia class. nov.
1465 *Int J Syst Evol Microbiol* 72:005489.

1466 Koonin EV, Yutin N. 2014. The dispersed archaeal eukaryome and the complex
1467 archaeal ancestor of eukaryotes. *Cold Spring Harb Perspect Biol* 6:a016188.

1468 Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. 2015. Sparse
1469 and Compositionally Robust Inference of Microbial Ecological Networks. *PLOS*
1470 *Computational Biology* 11:e1004226.

1471 Lane N, Martin W. 2010. The energetics of genome complexity. *Nature* 467:929–
1472 934.

1473 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat*
1474 *Methods* 9:357–359.

1475 Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site
1476 heterogeneities in the amino-acid replacement process. *Mol Biol Evol* 21:1095–
1477 1109.

1478 Lartillot N, Rodrigue N, Stubbs D, Richer J. 2013. PhyloBayes MPI: phylogenetic
1479 reconstruction with infinite mixtures of profiles in a parallel environment. *Syst.*
1480 *Biol.* 62:611–615.

1481 Leung PM, Grinter R, Tudor-Matthew E, Lingford JP, Jimenez L, Lee H-C, Milton M,
1482 Hanchapola I, Tanuwidjaya E, Kropp A, et al. 2024. Trace gas oxidation
1483 sustains energy needs of a thermophilic archaeon at suboptimal temperatures.
1484 *Nat Commun* 15:3219.

1485 Liu Y, Li M. 2022. The unstable evolutionary position of Korarchaeota and its
1486 relationship with other TACK and Asgard archaea. *mLife* 1:218–222.

1487 Liu Y, Makarova KS, Huang W-C, Wolf YI, Nikolskaya AN, Zhang X, Cai M, Zhang
1488 C-J, Xu W, Luo Z, et al. 2021. Expanded diversity of Asgard archaea and their
1489 relationships with eukaryotes. *Nature* 593:553–557.

1490 Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The
1491 carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.*
1492 42:D490–D495.

1493 López-García P, Moreira D. 2015. Open Questions on the Origin of Eukaryotes.
1494 *Trends Ecol Evol* 30:697–708.

1495 López-García P, Moreira D. 2020. The Syntrophy hypothesis for the origin of
1496 eukaryotes revisited. *Nat Microbiol* 5:655–667.

1497 Lu Z, Fu T, Li T, Liu Y, Zhang S, Li J, Dai J, Koonin EV, Li G, Chu H, et al. 2020.
1498 Coevolution of Eukaryote-like Vps4 and ESCRT-III Subunits in the Asgard
1499 Archaea. *mBio* 11:e00417–e00420.

1500 Lu Z, Xia R, Zhang S, Pan J, Liu Y, Wolf YI, Koonin EV, Li M. 2024. Evolution of
1501 optimal growth temperature in Asgard archaea inferred from the temperature
1502 dependence of GDP binding to EF-1A. *Nature Communications* 15:1–7.

1503 Makarova KS, Wolf YI, Koonin EV. 2015. Archaeal Clusters of Orthologous Genes
1504 (arCOGs): An Update and Application for Analysis of Shared Features between
1505 Thermococcales, Methanococcales, and Methanobacteriales. *Life* 5:818–840.

1506 Martijn J, Schön ME, Lind AE, Vosseberg J, Williams TA, Spang A, Ettema TJG.
1507 2020. Hikarchaeia demonstrate an intermediate stage in the methanogen-to-
1508 halophile transition. *Nat Commun* 11:5490.

1509 Martijn J, Vosseberg J, Guy L, Offre P, Ettema TJG. 2018. Deep mitochondrial origin
1510 outside the sampled alphaproteobacteria. *Nature* 557:101–105.

1511 Martin W, Müller M. 1998. The hydrogen hypothesis for the first eukaryote. *Nature*
1512 392:37–41.

1513 McCutcheon JP, Moran NA. 2011. Extreme genome reduction in symbiotic bacteria.
1514 *Nature Reviews Microbiology* 10:13–26.

1515 Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A,
1516 Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for
1517 Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* 37:1530–1534.

1518 Moody ERR, Mahendrarajah TA, Dombrowski N, Clark JW, Petitjean C, Offre P,

Szöllősi GJ, Spang A, Williams TA. 2022. An estimate of the deepest branches of the tree of life from ancient vertically evolving genes. *Elife* 11:e66695.

Moreira D, Lopez-Garcia P. 1998. Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J Mol Evol* 47:517–530.

Muñoz-Gómez SA, Susko E, Williamson K, Eme L, Slamovits CH, Moreira D, López-García P, Roger AJ. 2022. Site-and-branch-heterogeneous analyses of an expanded dataset favour mitochondria as sister to known Alphaproteobacteria. *Nat Ecol Evol* 6:253–262.

Munson-McGee JH, Field EK, Bateson M, Rooney C, Stepanauskas R, Young MJ. 2015. Nanoarchaeota, Their Sulfolobales Host, and Nanoarchaeota Virus Distribution across Yellowstone National Park Hot Springs. *Appl Environ Microbiol* 81:7860–7868.

Nayfach S, Roux S, Seshadri R, Udwary D, Varghese N, Schulz F, Wu D, Paez-Espino D, Chen I-M, Huntemann M, et al. 2021. A genomic catalog of Earth's microbiomes. *Nat Biotechnol* 39:499–509.

Nishimura Y, Eguchi T. 2006. Biosynthesis of archaeal membrane lipids: digeranylgeranylglycerophospholipid reductase of the thermoacidophilic archaeon Thermoplasma acidophilum. *J. Biochem.* 139:1073–1081.

Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. 2022. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* 50:D785–D794.

Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 36:996–1004.

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055.

Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2:1533–1542.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Müller A, Nothman J, Louppe G, et al. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12:2825–2830.

Pierpont CL, Baroch JJ, Church MJ, Miller SR. 2024. Idiosyncratic genome evolution of the thermophilic cyanobacterium Synechococcus at the limits of phototrophy. *ISME J* 18:wrae184.

Pittis AA, Gabaldón T. 2016. Late acquisition of mitochondria by a host with

chimaeric prokaryotic ancestry. *Nature* 531:101–104.

Podar M, Makarova KS, Graham DE, Wolf YI, Koonin EV, Reysenbach A-L. 2013. Insights into archaeal evolution and symbiosis from the genomes of a nanoarchaeon and its inferred crenarchaeal host from Obsidian Pool, Yellowstone National Park. *Biol Direct* 8:9.

Poole AM, Gribaldo S. 2014. Eukaryotic origins: How and when was the mitochondrion acquired? *Cold Spring Harb Perspect Biol* 6:a015990.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.

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. 2023. Panguiarchaeum symbiosum, a potential hyperthermophilic symbiont in the TACK superphylum. *Cell Rep.* 42:112158.

Ragsdale SW, Pierce E. 2008. Acetogenesis and the Wood-Ljungdahl pathway of CO(2) fixation. *Biochim Biophys Acta* 1784:1873–1898.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* 67:901–904.

Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. 2018. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* 46:D624–D632.

R Core Team R. 2024. R: A language and environment for statistical computing.

Rinke C, Chuvochina M, Mussig AJ, Chaumeil P-A, Davín AA, Waite DW, Whitman WB, Parks DH, Hugenholtz P. 2021. A standardized archaeal taxonomy for the Genome Taxonomy Database. *Nat Microbiol* 6:946–959.

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S, Swan BK, Gies EA, et al. 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499:431–437.

Roger AJ, Muñoz-Gómez SA, Kamikawa R. 2017. The Origin and Diversification of Mitochondria. *Curr Biol* 27:R1177–R1192.

Sabath N, Ferrada E, Barve A, Wagner A. 2013. Growth temperature and genome size in bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation. *Genome Biol Evol* 5:966–977.

Saier MH, Reddy VS, Moreno-Hagelsieb G, Hendargo KJ, Zhang Y, Iddamsetty V, Lam KJK, Tian N, Russum S, Wang J, et al. 2021. The Transporter Classification Database (TCDB): 2021 update. *Nucleic Acids Res.* 49:D461–D467.

Sakai HD, Nakamura K, Kurosawa N. 2022. Stygiolobus caldivivus sp. nov., a

1597 facultatively anaerobic hyperthermophilic archaeon isolated from the Unzen hot
1598 spring in Japan. *Int J Syst Evol Microbiol* 72:005486.

1599 Sakai HD, Nur N, Kato S, Yuki M, Shimizu M, Itoh T, Ohkuma M, Suwanto A,
1600 Kurosawa N. 2022. Insight into the symbiotic lifestyle of DPANN archaea
1601 revealed by cultivation and genome analyses. *Proc Natl Acad Sci U S A*
1602 119:e2115449119.

1603 Santana-Molina C, Williams TA, Snel B, Spang A. 2025. Chimeric origins and
1604 dynamic evolution of central carbon metabolism in eukaryotes. *Nat Ecol Evol*
1605 9:613–627.

1606 Schrempf D, Lartillot N, Szöllősi G. 2020. Scalable Empirical Mixture Models That
1607 Account for Across-Site Compositional Heterogeneity. *Mol. Biol. Evol.* 37:3616–
1608 3631.

1609 Schuchmann K, Müller V. 2014. Autotrophy at the thermodynamic limit of life: a
1610 model for energy conservation in acetogenic bacteria. *Nat Rev Microbiol*
1611 12:809–821.

1612 Schuchmann K, Müller V. 2016. Energetics and Application of Heterotrophy in
1613 Acetogenic Bacteria. *Appl Environ Microbiol* 82:4056–4069.

1614 Schut GJ, Boyd ES, Peters JW, Adams MWW. 2013. The modular respiratory
1615 complexes involved in hydrogen and sulfur metabolism by heterotrophic
1616 hyperthermophilic archaea and their evolutionary implications. *FEMS Microbiol*
1617 *Rev* 37:182–203.

1618 Schwank K, Bornemann TLV, Dombrowski N, Spang A, Banfield JF, Probst AJ.
1619 2019. An archaeal symbiont-host association from the deep terrestrial
1620 subsurface. *ISME J* 13:2135–2139.

1621 Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*
1622 30:2068–2069.

1623 Shaiber A, Eren AM. 2019. Composite Metagenome-Assembled Genomes Reduce
1624 the Quality of Public Genome Repositories. *mBio* 10:e00725–19.

1625 Shimodaira H. 2001. Multiple comparisons of log-likelihoods and combining
1626 nonnested models with applications to phylogenetic tree selection. *Commun.*
1627 *Stat. Theory Methods* 30:1751–1772.

1628 Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection.
1629 *Syst Biol* 51:492–508.

1630 Søndergaard D, Pedersen CNS, Greening C. 2016. HydDB: A web tool for
1631 hydrogenase classification and analysis. *Sci. Rep.* 6:34212.

1632 Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van
1633 Eijk R, Schleper C, Guy L, Ettema TJG. 2015. Complex archaea that bridge the
1634 gap between prokaryotes and eukaryotes. *Nature* 521:173–179.

1635 Spang A, Stairs CW, Dombrowski N, Eme L, Lombard J, Caceres EF, Greening C,
1636 Baker BJ, Ettema TJG. 2019. Proposal of the reverse flow model for the origin
1637 of the eukaryotic cell based on comparative analyses of Asgard archaeal
1638 metabolism. *Nat Microbiol* 4:1138–1148.

1639 St John E, Liu Y, Podar M, Stott MB, Meneghin J, Chen Z, Lagutin K, Mitchell K,
1640 Reysenbach A-L. 2019. A new symbiotic nanoarchaeote (Candidatus
1641 Nanoclepta minutus) and its host (Zestosphaera tikiterensis gen. nov., sp. nov.)
1642 from a New Zealand hot spring. *Syst Appl Microbiol* 42:94–106.

1643 Szánthó LL, Lartillot N, Szöllősi GJ, Schrempf D. 2023. Compositionally Constrained
1644 Sites Drive Long-Branch Attraction. *Syst. Biol.* 72:767–780.

1645 Szilágyi A, Závodszky P. 2000. Structural differences between mesophilic,
1646 moderately thermophilic and extremely thermophilic protein subunits: results of
1647 a comprehensive survey. *Structure* 8:493–504.

1648 Szöllősi GJ, Davín AA, Tannier E, Daubin V, Boussau B. 2015. Genome-scale
1649 phylogenetic analysis finds extensive gene transfer among fungi. *Philos Trans R*
1650 *Soc Lond B Biol Sci* 370:20140335.

1651 Tahon G, Köstlbacher S, Pelve EA, Baker BJ, Saw JH, Eme L, Tamarit D, Schön
1652 ME, Ettema TJG. 2023. Phylogenomics and ancestral reconstruction of
1653 Korarchaeota reveals genomic adaptation to habitat switching. *bioRxiv*
1654 [Internet]. Available from: http://dx.doi.org/10.1101/2023.09.28.559970

1655 Tamarit D, Köstlbacher S, Appler KE, Panagiotou K, De Anda V, Rinke C, Baker BJ,
1656 Ettema TJG. 2024. Description of Asgardarchaeum abyssi gen. nov. spec. nov.,
1657 a novel species within the class Asgardarchaeia and phylum Asgardarchaeota
1658 in accordance with the SeqCode. *Syst Appl Microbiol* 47:126525.

1659 The pandas development team. 2024. pandas-dev/pandas: Pandas. Zenodo
1660 Available from: https://zenodo.org/doi/10.5281/zenodo.3509134

1661 Topçuoğlu BD, Stewart LC, Morrison HG, Butterfield DA, Huber JA, Holden JF.
1662 2016. Hydrogen limitation and syntrophic growth among natural assemblages of
1663 thermophilic methanogens at deep-sea hydrothermal vents. *Front. Microbiol.*
1664 7:1240.

1665 Tria FDK, Brueckner J, Skejo J, Xavier JC, Kapust N, Knopp M, Wimmer JLE,
1666 Nagies FSP, Zimorski V, Gould SB, et al. 2021. Gene Duplications Trace
1667 Mitochondria to the Onset of Eukaryote Complexity. *Genome Biol Evol*
1668 13:evab055.

1669 Valentin-Alvarado LE, Appler KE, De Anda V, Schoelmerich MC, West-Roberts J,
1670 Kivenson V, Crits-Christoph A, Ly L, Sachdeva R, Greening C, et al. 2024.
1671 Asgard archaea modulate potential methanogenesis substrates in wetland soil.
1672 *Nat. Commun.* 15:6384.

1673 Van Rossum G, Drake FL. 2009. Python 3 Reference Manual: (Python
1674 Documentation Manual Part 2). CreateSpace

Ver Eecke HC, Butterfield DA, Huber JA, Lilley MD, Olson EJ, Roe KK, Evans LJ, Merkel AY, Cantin HV, Holden JF. 2012. Hydrogen-limited growth of hyperthermophilic methanogens at deep-sea hydrothermal vents. *Proc Natl Acad Sci U S A* 109:13674–13679.

Vosseberg J, van Hooff JJE, Köstlbacher S, Panagiotou K, Tamarit D, Ettema TJG. 2024. The emerging view on the origin and early evolution of eukaryotic cells. *Nature* 633:295–305.

Vosseberg J, van Hooff JJE, Marcet-Houben M, van Vlimmeren A, van Wijk LM, Gabaldón T, Snel B. 2021. Timing the origin of eukaryotic cellular complexity with ancient duplications. *Nat Ecol Evol* 5:92–100.

Wang H-C, Minh BQ, Susko E, Roger AJ. 2018. Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. *Syst Biol* 67:216–235.

Waskom M. 2021. seaborn: statistical data visualization. *J. Open Source Softw.* 6:3021.

Werner F. 2007. Structure and function of archaeal RNA polymerases. *Mol. Microbiol.* 65:1395–1404.

Williams TA, Cox CJ, Foster PG, Szöllősi GJ, Embley TM. 2020. Phylogenomics provides robust support for a two-domains tree of life. *Nat Ecol Evol* 4:138–147.

Williams TA, Davin AA, Szánthó LL, Stamatakis A, Wahl NA, Woodcroft BJ, Soo RM, Eme L, Sheridan PO, Gubry-Rangin C, et al. 2024. Phylogenetic reconciliation: making the most of genomes to understand microbial ecology and evolution. *ISME J* 18:wrae129.

Williams TA, Schrempf D, Szöllősi GJ, Cox CJ, Foster PG, Embley TM. 2021. Inferring the Deep Past from Molecular Data. *Genome Biol Evol* 13:evab067.

Williams TA, Szöllősi GJ, Spang A, Foster PG, Heaps SE, Boussau B, Ettema TJG, Embley TM. 2017. Integrative modeling of gene and genome evolution roots the archaeal tree of life. *Proc. Natl. Acad. Sci. U. S. A.* 114:E4602–E4611.

Woodcroft BJ, Aroney STN, Zhao R, Cunningham M, Mitchell JAM, Blackall L, Tyson GW. 2024. SingleM and Sandpiper: Robust microbial taxonomic profiles from metagenomic data. *bioRxiv* [Internet]:2024.01.30.578060. Available from: http://dx.doi.org/10.1101/2024.01.30.578060

Wu F, Speth DR, Philosof A, Crémière A, Narayanan A, Barco RA, Connon SA, Amend JP, Antoshechkin IA, Orphan VJ. 2022. Unique mobile elements and scalable gene flow at the prokaryote-eukaryote boundary revealed by circularized Asgard archaea genomes. *Nat Microbiol* 7:200–212.

Wurch L, Giannone RJ, Belisle BS, Swift C, Utturkar S, Hettich RL, Reysenbach A-L, Podar M. 2016. Genomics-informed isolation and characterization of a symbiotic Nanoarchaeota system from a terrestrial geothermal environment. *Nat Commun* 7:12115.

1715 Xie R, Wang Y, Huang D, Hou J, Li L, Hu H, Zhao X, Wang F. 2022. Expanding
1716 Asgard members in the domain of Archaea sheds new light on the origin of
1717 eukaryotes. *Sci. China Life Sci.* 65:818–829.

1718 Xu S, Li L, Luo X, Chen M, Tang W, Zhan L, Dai Z, Lam TT, Guan Y, Yu G. 2022. : A
1719 serialized data object for visualization of a phylogenetic tree and annotation
1720 data. *Imeta* 1:e56.

1721 Yu H, Wu C-H, Schut GJ, Haja DK, Zhao G, Peters JW, Adams MWW, Li H. 2018.
1722 Structure of an Ancient Respiratory System. *Cell* 173:1636–1649.e16.

1723 Yu T, Fu L, Wang Y, Dong Y, Chen Y, Wegener G, Cheng L, Wang F. 2024.
1724 Thermophilic Hadarchaeota grow on long-chain alkanes in syntrophy with
1725 methanogens. *Nat Commun* 15:6560.

1726 Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L,
1727 Vancaester E, Seitz KW, Anantharaman K, Starnawski P, Kjeldsen KU, et al.
1728 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity.
1729 *Nature* 541:353–358.

1730 Zeldovich KB, Berezovsky IN, Shakhnovich EI. 2007. Protein and DNA sequence
1731 determinants of thermophilic adaptation. *PLoS Comput Biol* 3:e5.

1732 Zhang J, Feng X, Li M, Liu Y, Liu M, Hou L-J, Dong H-P. 2025. Deep origin of
1733 eukaryotes outside Heimdallarchaeia within Asgardarchaeota. *Nature*:1–9.

1734