# Expanding Asgard members in the domain of Archaea shed new light on the origin of eukaryotes

Ruize Xie[1,#], Yinzhao Wang[2,#], Danyue Huang[1], Jialin Hou[2], Liuyang Li[2], Haining Hu[2], Xiaoxiao Zhao[2], Fengping Wang[1,2,3]*

[1]School of Oceanography, Shanghai Jiao Tong University, Shanghai 200030, China

[2]State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

[3]Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, Guangdong, China

# These authors contributed equally to this paper

*Corresponding author:

Fengping Wang

School of Oceanography, Shanghai Jiao Tong University

800 Dongchuan Road, Minhang District, Shanghai 200240, China

fengpingw@sjtu.edu.cn

**Abstract**

The hypothesis that eukaryotes originated from within the domain Archaea has been strongly supported by recent phylogenomic analyses placing Heimdallarchaeota from the Asgard superphylum as the closest known archaeal sister-group to eukaryotes. At present, only six phyla are described in the Asgard superphylum, which limits our understanding of the relationship between eukaryotes and archaea, as well as the evolution and ecological functions of the Asgard archaea. Here, we describe five previously unknown phylum-level Asgard archaeal lineages, tentatively named Tyr-, Sigyn-, Freyr-, Njord- and Balderarchaeota. Comprehensive phylogenomic analyses further supported the origin of eukaryotes within Archaea to form a 2-domain tree of life and a new Asgard lineage Njordarchaeota was identified as the potential closest branch with the eukaryotic nuclear host lineage rather than Heimdallarchaeota that were previously considered as the closest archaeal relatives of eukaryotes. Metabolic reconstruction of Njordarchaeota suggests a heterotrophic lifestyle, with potential capability of peptides and amino acids utilization. This study largely expands the Asgard superphylum, provides additional evidences to support the 2-domain life tree and sheds new light on the evolution of eukaryotes.

Keywords: archaea, Asgard, eukaryotic origin

40 **Introduction**

41 The origin of eukaryotes is considered as a critical biological evolutionary event on Earth[1, 2, 3,
42 4]. The common ancestor of eukaryotes is generally believed to have evolved from a symbiotic
43 process[5, 6] in which one endosymbiotic bacterium within the Proteobacteria phylum evolved
44 into a mitochondrion [7, 8] and one endosymbiotic host cell became the cell nucleus[9, 10, 11]. The
45 identity of the host cell ancestor has been vigorously debated, and two hypotheses regarding
46 2- or 3-domain trees of life have been raised[12, 13]. However, increasing evidence provided by
47 phylogenomic analyses[10, 14], as well as the presence of eukaryotic signature proteins (ESPs)[15]
48 in the Asgard archaea, has supported the idea that eukaryotic cells originated in the domain
49 Archaea, particularly in the archaeal Asgard superphylum[9, 10, 16]. The Asgard archaea are
50 described as mixotrophic or heterotrophic[11, 17] and are ubiquitously distributed in various
51 environments, such as hydrothermal vents[9, 10]; lake, river and marine sediments[18]; microbial
52 mats[19]; and mangroves[17]. These organisms potentially play important roles in global
53 geochemical cycling[20]. The identification of Lokiarchaeota in the Loki's Castle hydrothermal
54 vent field provided pivotal genomic and phylogenetic evidence that eukaryotes originated
55 within the domain Archaea, supporting a 2-domain tree of life, which is consistent with the
56 eocyte hypothesis[9]. Further discovery and proposal of the Asgard superphylum have provided
57 new insights into the transition of archaea to eukaryotes and into the origin of eukaryotic cell
58 complexity[10]. Within the Asgard superphylum, Heimdallarchaeota had been identified to be
59 the closest Asgard archaeal lineage to the eukaryotic branch on the phylogenetic tree on the
60 basis of carefully selected conserved protein sequences[10, 14]. Recently, Imachi *et al.* cultivated
61 one Asgard archaeon, *Candidatus* Prometheoarchaeum syntrophicum strain MK-D1, in the

62 laboratory and observed, for the first time, the intertwining of this archaeon with bacterial
63 cells via extracellular protrusions under a transmission electron microscope[21]. The idea of an
64 archaeal origin of eukaryotes and a 2-domain tree of life has recently become increasingly
65 favorable[14, 22]; nevertheless, our understanding of the evolution of Asgard archaea, the
66 archaea-eukaryote transition, and the ecological and geochemical roles of these evolutionarily
67 important archaea remains incomplete. This lack of understanding is largely due to the limited
68 number of high-quality genomes of Asgard archaea, which are considered highly diverse as
69 revealed by 16S rRNA gene surveys[20, 23]; yet only a small fraction have representative
70 genomes. In this study, we assembled five previously unknown phylum-level Asgard archaeal
71 group, greatly expanded the Asgard genomic diversity within the domain of Archaea and shed
72 new light on the origin of eukaryotes.

73 **Results**

74 **Expanded Asgard archaea support 2-domain tree of life**

75 In total, 17 metagenomic datasets were used in this study, including two samples from
76 hydrothermal sediment of Guaymas Basin, six samples from Tengchong hot spring sediment,
77 as well as 9 metagenomic datasets from the publicly available National Center for
78 Biotechnology Information (NCBI) Sequence Read Archive (SRA) database (Permission
79 granted, Supplementary Table 1). After subsequently assembling, binning and classification
80 as described in the Methods section, 128 Asgard metagenome-assembled genomes (MAGs)
81 were obtained and in-depth phylogenomic analyses were performed with 37 concatenated
82 conserved proteins[24] under LG+C60+F+G4 model to confirm the placement of these MAGs
83 on phylogenomic tree. The analysis revealed that, in addition to the previously described

84 Loki-, Thor-, Odin-, Heimdall-, Hela- and Hermodarchaeota clades[25], there are five additional
85 monophyletic branching clades (Fig. 1), here tentatively named Tyr-, Sigyn-, Freyr-, Njord-
86 and Balderarchaeota after the Asgard gods in the Norse mythology (Tyr, the god of war;
87 Sigyn, the god of victory; Freyr, the god of peace; Njord, the god of seas; and Balder, the god
88 of light). The MAGs of these new Asgard lineages were recovered from different
89 environments: Njordarchaeota and Freyrarchaeota were derived from hydrothermal sediment;
90 Tyrarchaeota were found in estuary sediments; Sigynarchaeota were reconstructed from hot
91 spring sediments and Balderarchaeota were retrieved from hot spring and hydrothermal
92 sediments. Additionally, a clade of Hermodarchaeota was identified in high temperature
93 habitats (~85℃), similar with Odinarchaeota, which were considered as the only thermophilic
94 member of Asgard archaea to date[10]. Near-complete MAGs ranging in size from 2.1 to 5.5
95 Mb with completeness ranging from 87.38 to 97.20% were constructed for representatives of
96 each new Asgard clade (Supplementary Table 2). To further assess their distinctiveness
97 compared to the Asgard members already defined, we calculated the average nucleotide
98 identity (ANI) (Supplementary Fig. 1) and average amino acid identity (AAI) (Supplementary
99 Fig. 2) between them and other Asgard MAGs. The AAI values showed that all the MAGs of
100 new lineages discovered here share a low AAI with the known Asgard archaea (<50%) and
101 fall within the phylum-level classification range (40%~52%)[26], providing additional support
102 for the uniqueness of these new Asgard lineages.

103 To determine the phylogenetic positions of these new Asgard lineages in relation to
104 eukaryotes, we performed comprehensive phylogenetic analyses using 21 conserved marker
105 proteins carefully selected by Williams *et al.*[14] and 54 archaeal-eukaryotic ribosomal

106 proteins[10]. The taxa included in these analyses were also selected on the basis of the
107 instructions of Williams *et al.*[14]: a representative taxon set was constructed comprising 85
108 archaeal genomes (53 within Asgard), 19 eukaryotic genomes, and 36 bacterial genomes
109 (Supplementary Table 3). To avoid potential phylogenetic artifacts resulting from horizontal
110 gene transfer (HGT, including inter-archaeal horizontal gene transfers arHGTs), long branch
111 attraction (LBA) and eukaryotic genes from the mitochondria or plastids, single-gene datasets
112 were carefully inspected with single protein trees and BLASTp inspection was performed. We
113 then concatenated two gene sets, then inferred maximum likelihood trees under the
114 LG+C60+F+G4 model for 21 conserved marker proteins and 54 archaeal-eukaryotic
115 ribosomal proteins respectively. The phylogenetic analysis of 21 marker genes showed that
116 eukaryotes are the sister group of Njordarchaeota rather than Heimdallarchaeota which was
117 previously described the sister lineage to eukaryotes[11, 14] with high support (bootstrap support
118 (BS) = 93, Fig. 2a). The phylogenetic analysis of 54 ribosomal proteins also indicates that the
119 eukaryotes lineage forms a monophyletic cluster with Tyr-, Heimdall- and Njordarchaeota,
120 while Njordarchaeota is the deepest lineage close to eukaryotes (BS = 61, Fig. 2b). In
121 summary, the phylogenomic analyses provide strong support for a 2-domain tree with
122 Njordarchaeota as the closest potential relatives to eukaryotes.

123 **ESP-encoding genes widely shared by the Asgard archaea**

124 The potential ESPs were identified from the newly reconstructed Asgard MAGs (Fig. 3).
125 Consistent with previous reports[9, 10, 16, 27], different key subunits of informational processing
126 machinery were found. For example, topoisomerase IB protein-encoding genes were
127 identified in Freyr-, Balder- and Hermodarchaeota, while all MAGs of Balderarchaeota and

128 Freyrarchaeote GB11 were found to encode a RNA polymerase subunit G. Homologues of
129 eukaryotic ribosomal protein L22e were identified in Freyr-, Hermodarchaeota and
130 Tyrarchaeota. The new clades of Asgard archaea were also found to contain genes related to
131 cell division and the cytoskeleton, but tubulin-encoding genes were not detected. With regard
132 to actin-related proteins, two to three related subunits were detected in Tyr- and
133 Balderarchaeota, whereas profilin domain protein-encoding genes were identified in all
134 Asgard lineages described here.

135 Ubiquitin-based signaling is an important cellular process in eukaryotes[28]. Previous
136 studies have reported the presence of the related protein domains in Loki-, Odin-, Hel- and
137 Heimdallarchaeota but not in Thorarchaeota[9, 27, 29]. Here, we identified ubiquitin
138 system-related protein-encoding genes in nearly all newly assembled MAGs including several
139 ubiquitin-related domains, zinc fingers, ubiquitin-activating enzyme (E1),
140 ubiquitin-conjugating protein (E2), and UFM1-protein ligase 1 (E3), indicating that the
141 ubiquitin system is widespread in Asgard archaea.

142 The endosomal sorting complex required for transport (ESCRT) machinery consisting of
143 complexes I-III and associated subunits[9, 30, 31] were identified in the newly recovered MAGs.
144 Genes coding for Vps28 domain-containing proteins previously found in Loki-, Odin-, Hel-
145 and Heimdallarchaeota[25, 27] were also identified in Tyr-, Freyr-, Balder- and Hermodarchaeota
146 but were absent in Sigynarchaeota and Njordarchaeota. The Sigynarchaeota MAGs also lack
147 genes for both EAP30 domain- and steadiness box domain-containing proteins. Notably, all
148 new Asgard MAGs contain cyclin-like protein-encoding genes, whereas ESCRT complexes
149 I-III were only identified in Freyrarchaeota.

150 ESPs with intracellular trafficking and secretion functions were also identified in MAGs
151 here. However, only Tyrarchaeota contains genes coding for proteins with homology to
152 TRAPP-domain protein and Sec23/24-type protein-encoding gene was only found in
153 Balderarchaeota. All MAGs reported in present study possess genes coding RLC7 roadblock
154 domain protein. Genes coding for both the N- and C-termini of arrestins were found in
155 Freyrarchaeota and Hermodarchaeota; the organization resembles one previously reported in
156 the genome of Lokiarchaeote CR-4, in which the C- and N-terminal domain proteins are
157 separated from each other by one gene[25].

158 We also analyzed the oligosaccharyltransferase (OST) complex in the reconstructed
159 MAGs, and the results showed that OST complex-encoding genes were present in the MAGs
160 of all five Asgard clades. Ribophorin I homolog-encoding genes were found in the Hermod-,
161 Freyr-, Sigyn- and Tyrarchaeota MAGs. Homologs of OST3/6, which have been
162 demonstrated to influence yeast glycosylation efficiency[32], were also identified in several
163 Asgard MAGs, while STT32 subunit protein-encoding genes were detected all MAGs,
164 consistent with previous reports on Loki-, Odin-, Thor-, Hel- and Heimdallarchaeota[25].

165 In addition to reported ESPs, we identified a potential ESP belonging to mu/sigma
166 subunit of AP (adaptor protein) complex-encoding genes in Balder- and Freyrarchaeota
167 MAGs (Fig. 3). Homologues of mu/sigma subunit of AP complex contain IPR022775 domain.
168 AP complexes are classified into AP-1, AP-2, AP-3, AP-4 and AP-5 and all AP complexes
169 are heterotetramers consisting of two large subunits (adaptins), one medium-sized subunit
170 (mu) and one small-sized subunit (sigma)[33, 34]. AP complexes play a vital role in mediating

171 intracellular membrane trafficking[35]. Taken together, the identification of potential ESP
172 provides further insight into the origins of eukaryotic cellular complexity.

173 **Metabolic reconstructions of the five new Asgard lineages**

174 Balderarchaeota MAGs contain genes coding for complete glycolysis via
175 Embden-Meyerhof-Parnas (EMP) pathway, major steps of the tricarboxylic acid (TCA) cycle
176 and β-oxidation pathway, suggesting that members of Balderarchaeota may have potential to
177 metabolize organic compounds including carbohydrates and fatty acids (Fig. 4). The Wood
178 Ljungdahl (WL) pathway enables organisms to reduce two molecules of $CO_2$ to form
179 acetyl-CoA, and then to acetate to produce ATP. The ADP-dependent acetyl-CoA synthetase
180 (ACD) for acetogenesis, which is widely found in archaea[36], was identified in
181 Balderarchaeota MAGs. Meanwhile, phosphate acetyltransferase (Pta) and acetate kinase
182 (Ack) were also found in all Balderarchaeota MAGs (Fig. 4, Supplementary Table 5).
183 Although the *pta* gene was found in Sigynarchaeota and Freyrarchaeota as well, all of their
184 MAGs lack the *ack* gene. The Pta/Ack pathway for acetate production, which is common in
185 bacteria, was so far only found in Bathyarchaeota and the methanogenic genus
186 *Methanosarcina* in archaea[37, 38] and it is the first case that genes coding for Pta and Ack were
187 discovered in the Asgard archaea. The archaeal *pta*/*ack* genes were considered HGT from
188 bacteria donors. For example, the genes in *Methanosarcina* were postulated to acquire from a
189 cellulolytic Clostridia group[39] whereas the *pta*/*ack* genes donor of Bathyarchaeota was still
190 unclear, possibly one unknown clade of Bacteria[37]. For Balderarchaeota, the phylogenetic
191 analysis of the *ack* gene sequences revealed that *ack* genes of Balderarchaeotal branch closely
192 to a bacteria lineage Petrotoga (Supplementary Fig. 3), indicating that *ack* genes of

193 Balderarchaeota probably acquired from Petrotoga. While the phylogenetic tree of *pta* genes
194 shows that the Balderarchaeota clade are within Firmicutes branch (Supplementary Fig. 4).
195 Taken together, the Pta/Ack pathway in Balderarchaeota may have acquired from different
196 bacterial donor by two separate HGT events. Additionally, genes coding for the nitrite
197 reductase (NADH) large subunit (*nir*B) were detected in MAGs of Balderarchaeota, implying
198 a potential nitrite reduction capability.

199 Sigynarchaeota contain not only all genes responsible for glycolysis but also abundant
200 genes coding for extracellular carbohydrate-degrading enzymes, including α-amylase,
201 cellulase, α-mannosidases and β-glucosidases (Supplementary Table 4), indicating that
202 archaea in Sigynarchaeota have the capacity to degrade complex carbohydrates. There are two
203 types of the WL pathway using different enzymes as $C_1$ carrier, one using
204 tetrahydromethanopterin (THMPT), the other using tetrahydrofolate (THF). Archaea normally
205 utilize the THMPT-WL pathway while acetogenic bacteria generally utilize the THF-WL
206 pathway[40]. Sigynarchaeota MAGs contain genes for both types of the WL pathway but
207 5,10-methylenetetrahydromethanopterin reductase (Mer) which converts
208 5-methyltetrahydromethanopterin to 5,10-methylenetetrahydromethanopterin was missing in
209 all MAGs identified here, suggesting that Sigynarchaeota probably use the THF-WL pathway
210 for acetate production (Fig. 4, Supplementary Table 5). Sigynarchaeota MAGs contain neither
211 NADH dehydrogenase nor type 4 [NiFe] hydrogenase (Supplementary Table 5), nevertheless,
212 they probably use membrane-bound heterodisulfide reductase (Hdr) to generate proton motive
213 force as described in their sister-lineage Lokiarchaeota[11].

214 Except for genes relevant to carbon cycling (including the WL pathway, EMP pathway
215 and β-oxidation), Freyrarchaeota MAGs contain more genes involved in nitrogen and sulfur
216 cycling compared with the other newly discovered Asgard clades here. With regard to
217 nitrogen metabolism, Freyrarchaeota were found to contain genes coding for the potential
218 nitrogen fixation-catalyzing subunit of nitrogenase (*nif*H) and nitrogenase cofactors. With
219 regard to sulfur metabolism, the key enzymes functioning in the assimilatory sulfate reduction
220 pathway and sulfate import were identified in members of Freyrarchaeota, suggesting that this
221 clade has potential to assimilate sulfate. Moreover, complete subunits of
222 sulfhydrogenase-encoding genes (*hyd*ABGD) were found in Freyrarchaeota MAGs. This
223 bifunctional hydrogenase has been verified in the hyperthermophilic *Pyrococcus furiosus*,
224 which can either remove reductants produced during fermentation by utilizing protons or use
225 polysulfides as electron acceptors[41]. Among genes associated with glycolysis, the specific
226 enzymes catalyzing these steps is different in the three lineages. In the first step of glycolysis,
227 for example, Freyrarchaeota and Sigynarchaeota use ATP-dependent ROK (repressor, open
228 reading frame, kinase) family enzymes while Balderarchaeota utilizes ADP-dependent
229 glucokinase. Likewise, Freyrarchaeota and Sigynarchaeota encode ATP-dependent
230 phosphofructokinase (PfkB) but fructose 6-phosphate (F6P) to fructose 1,6-bisphosphate
231 (F1,6P) was catalyzed by ADP-dependent phosphofructokinase (ADP-PFK) in
232 Balderarchaeota.

233 We also compared differences of metabolic characteristic between the newly identified
234 Hermodarchaeota members and Odinarchaeota since both of them were recovered from
235 high-temperature environments. Hermodarchaeota encodes all enzymes for the complete

236 THMPT-WL pathway, while Odinarchaeota lack several key genes of THMPT-WL pathway.
237 Furthermore, only group 3 [NiFe]-hydrogenases were found in Hermodarchaeota genomes,
238 lacking group 4 [NiFe]- hydrogenases which were widely identified in Odinarchaeota. The
239 presence of THMPT-WL pathway and group 3 [NiFe]-hydrogenases in Hermodarchaeota
240 indicate that they could grow lithoautotrophically by using $H_2$ as an electron donor.
241 DnaK-DnaJ-GrpE chaperone system, one of the characteristics of hyperthermophilic
242 archaea[42], were found in Hermodarchaeota and Odinarchaeota but genes coding for reverse
243 gyrase were absent in both of them.

244 In Tyrarchaeota MAGs, glycolysis and TCA pathways are not complete, however, the
245 presence of genes coding for THMPT-WL pathway and group 3 [NiFe] hydrogenases implies
246 its potential to harness energy from $H_2$ oxidation, possibly for lithoautotrophic growth,
247 depending on environmental conditions, as suggested for Lokiarchaeota and Thorarchaeota[43,
248 44]. Moreover, *de novo* anaerobic cobalamin (vitamin $B_{12}$) biosynthesis pathway was found in
249 Tyrarchaeota (Fig. 4, Supplementary Table 5), suggesting that Tyrarchaeota harbor the
250 potential of cobalamin synthesis. In nature, only limited members of bacteria and archaea
251 possess capacity of *de novo* cobalamin synthesis using one of two alternative pathways:
252 aerobic or anaerobic pathway[45]. Within Archaea, some members within Euryarchaeota,
253 Thaumarchaeota, Crenarchaeota and Bathyarchaeota have been reported possessing
254 cobalamin synthesizing pathway[46, 47], whereas only Tyrarchaeota seems to have this ability
255 in the Asgard archaea reported so far.

256 Njordarchaeota MAGs contain limited genes coding for major carbon metabolic
257 repertoire. Key genes coding for EMP pathway, TCA, the WL pathway and β-oxidation

258 pathway were missing (Fig. 4, Supplementary Table 5). But genes coding for amino acids
259 utilization were found in Njordarchaeota, including aminotransferases and
260 2-oxoacid:ferredoxin oxidoreductases (the former catalyzes the interconversion of amino
261 acids and 2-oxoacids and the latter oxidates 2-oxoacids to acyl-CoA), indicating that
262 Njordarchaeota has potential to metabolize amino acids. The amino acid carboxylate is
263 transferred to $CO_2$ and reducing ferredoxin during 2-oxoacid oxidation, which further could
264 be oxidized into formate and $H_2$ by formate dehydrogenase (Fdh) and [NiFe] hydrogenases,
265 respectively[21]. Only [NiFe]-hydrogenases were detected in Njordarchaeota but lacking Fdh.
266 Together, Njordarchaeota might have a fermentative life style by produce acetate or $H_2$ while
267 degrading amino acids, or it may be living in a symbiotic way. Nevertheless, cultivation
268 experiments are required to verify all these predictions described here.

269 Various models for the origin of eukaryotes have been proposed, based on a metabolic
270 symbiosis between one archaeon and bacterial partner, which was fostered by the discovery
271 of natural syntrophy between *Candidatus* Prometheoarchaeum syntrophicum (that can
272 degrade amino acids to $H_2$ or formate) and Deltaproteobacteria (that can utilize $H_2$ or Formate
273 and provide amino acids or vitamin $B_{12}$ to partner). Compared with other members of Asgard
274 archaea, Njordarchaeota possess limited pathway for carbon metabolism (Supplementary
275 Table 6), implying that they prone to grow in symbiosis with other organisms to adapt
276 complicated and volatile environment. Combining the phylogenetic affiliation with
277 eukaryotes and metabolic characteristic of Njordarchaeota, we speculate that the archaeal
278 ancestor of eukaryotes probably has potential to degrade amino acids to produce acetate or $H_2$
279 which can further benefit to the bacterial partner, although other additional lifestyles could

280 not be excluded. This “auxotrophy” life style may provide a selective force that enables
281 Njordarchaeota or even deeper branched Asgard lineages to eukaryotes stably symbiose with
282 bacterial partner, which further facilitates integration of symbiogenetic consortium, and
283 eventually evolved to original eukaryotes ancestor.

284 **Conclusion**

285 Undoubtedly, the origin of eukaryotes is one of the most important evolutionary events. The
286 discovery of Asgard archaea has boosted the eocyte hypothesis that eukaryotes derive from
287 within archaea because the Asgard archaea possess two remarkable features: robust
288 evolutionary affinity with eukaryotes and various ESPs existence. In the present study, five
289 novel Asgard lineages were discovered based on phylogenetic analyses and AAI value
290 comparison, which significantly expand the phylogenetic and metabolic diversity of the
291 Asgard archaea. Our analyses strongly support a 2-domain tree of life and clearly demonstrate
292 that the eukaryotes lineage cluster with either Njordarchaeota alone or Njord-, Tyr- and
293 Heimdallarchaeota, suggesting that Njordarchaeota lineage is the closest relatives or a deeper
294 branching lineage to eukaryotes than Heimdallarchaeota. Metabolic characteristic of
295 Njordarchaeota shows different carbon metabolic pathways from Heimdallarchaeota that were
296 considered living in aerobic or anerobic environment using various organic substrates such as
297 carbohydrates and fatty acids[11, 16, 48], whereas Njordarchaeota lack both complete glycolysis
298 and WL pathway and the most possibility of metabolism type is using amino acids in anoxic
299 niches. This finding does not contradict to the hypothesis by Spang *et al.* inferring metabolic
300 feature of the archaeal ancestor of eukaryotes[11] that it used organic substrates to produce
301 acetate, formate, $H_2$, which might be beneficial for symbiosis. In general, the characterization

of additional genomes and continuous efforts to cultivate Asgard archaea will provide additional insights into the evolution of archaea and their potential evolution into eukaryotes. Such insights will enable greater understanding of the ecological and geochemical roles of archaea in Earth's history.

**Propose type of new taxa**

***Candidatus* Tyrarchaeum** (Tyr.ar.chae'um. N.L. neut. n. *archaeum* archaeon; N.L. neut. n. *Tyrarchaeum* an archaeon named after Tyr, the god of war in North mythology). Type species: Candidatus Tyrarchaeum oakense.

***Candidatus* Tyrarchaeum oakense** (oak'ense N.L. neut. adj. pertaining to white oak river, North Carolina in the United States). This uncultured lineage is represented by the genome "WOR_431" consisting of 2.4 Mbps in 246 contigs with an estimated completeness of 91.56%, an estimated contamination of 2.95% and 20 tRNAs The MAG recovered from white oak river sediment.

***Candidatus* Freyrarchaeum** (Freyr.ar.chae'um. N.L. neut. n. *archaeum* archaeon; N.L. neut. n. *Freyrarchaeum* an archaeon named after Freyr, the god of peace in North mythology). Type species: Candidatus Freyrarchaeum guaymasis.

***Candidatus* Freyrarchaeum guaymasis** (guayma'sis N.L. neut. adj. pertaining to Guaymas Basin, located in the Gulf of California, México). This uncultured lineage is represented by the genome "GB_11" consisting of 2.3 Mbps in 71 contigs with an estimated completeness of 93.93%, an estimated contamination of 4.67% and 19 tRNAs The MAG recovered from Guaymas Basin sediment.

***Candidatus* Sigynarchaeum** (Sigyn.ar.chae'um. N.L. neut. n. *archaeum* archaeon; N.L. neut. n. *Sigynarchaeum* an archaeon named after Sigyn, the god of victory in North mythology). Type species: Candidatus Sigynarchaeum springense.

***Candidatus* Sigynarchaeum springense** (spring'ense N.L. neut. adj. pertaining to hot spring, Tengchong, China). This uncultured lineage is represented by the genome "SQRJ_234"

329 consisting of 5.9 Mbps in 269 contigs with an estimated completeness of 91.59%, an
330 estimated contamination of 4.67% and 21 tRNAs The MAG recovered from hot spring
331 sediment.

332 ***Candidatus* Balderarchaeum** (Balder.ar.chae'um. N.L. neut. n. *archaeum* archaeon; N.L.
333 neut. n. *Balderarchaeum* an archaeon named after Balder, the god of light in North
334 mythology). Type species: Candidatus Balderarchaeum guaymasis.

335 ***Candidatus* Balderarchaeum guaymasis** (guayma'sis N.L. neut. adj. pertaining to Guaymas
336 Basin, located in the Gulf of California, México). This uncultured lineage is represented by
337 the genome "GB_128" consisting of 3.8 Mbps in 131 contigs with an estimated completeness
338 of 97.2%, an estimated contamination of 4.21% and 20 tRNAs The MAG recovered from
339 Guaymas Basin sediment.

340 ***Candidatus* Njordarchaeum** (Njord.ar.chae'um. N.L. neut. n. *archaeum* archaeon; N.L. neut.
341 n. *Njordarchaeum* an archaeon named after Njord, the god of seas in North mythology). Type
342 species: Candidatus Njordarchaeum guaymasis.

343 ***Candidatus* Njordarchaeum guaymasis** (guayma'sis N.L. neut. adj. pertaining to Guaymas
344 Basin, located in the Gulf of California, México). This uncultured lineage is represented by
345 the genome "GB_154" consisting of 2.1 Mbps in 191 contigs with an estimated completeness
346 of 87.38%, an estimated contamination of 6.23% and 20 tRNAs The MAG recovered from
347 Guaymas Basin sediment.

348 ***Candidatus* Tyrarchaeaceae** (Tyr.ar.chae.ace'ae. N.L. neut. n. *Tyrarchaeum*, Candidatus
349 generic name; -aceae ending to denote the family; N.L. fem. pl. n. *Tyrarchaeaceae*, the
350 *Tyrarchaeum* family).

The family is described based on 37 concatenated conserved marker genes phylogeny. The description is the same as that of its sole genus and species. Type genus is *Candidatus* Tyrarchaeum.

***Candidatus* Tyrarchaeales** (Tyr.ar.chae.a'les. N.L. neut. n. *Tyrarchaeum*, Candidatus generic name; -ales ending to denote the order; N.L. fem. pl. n. *Tyrarchaeales*, the *Tyrarchaeum* order).

The order is described based on 37 concatenated conserved marker genes phylogeny. The description is the same as that of its sole genus and species. Type genus is *Candidatus* Tyrarchaeum.

***Candidatus* Tyrarchaeia** (Tyr.ar.chae'i.a. N.L. neut. n. *Tyrarchaeum*, Candidatus generic name; -ia ending to denote the class; N.L. fem. pl. n. *Tyrarchaeia*, the *Tyrarchaeum* class).

The class is described based on 37 concatenated conserved marker genes phylogeny. The description is the same as that of its sole genus and species. Type genus is *Candidatus* Tyrarchaeum.

***Candidatus* Tyrarchaeota** (Tyr.ar.chae.o'ta. N.L. neut. n. *Tyrarchaeum*, Candidatus generic name; -ota ending to denote the phylum; N.L. fem. pl. n. *Tyrarchaeota*, the *Tyrarchaeum* phylum).

The phylum is described based on 37 concatenated conserved marker genes phylogeny. The description is the same as that of its sole genus and species. Type genus is *Candidatus* Tyrarchaeum.

***Candidatus* Freyrarchaeaceae** (Freyr.ar.chae.ace'ae. N.L. neut. n. *Freyrarchaeum*, Candidatus generic name; -aceae ending to denote the family; N.L. fem. pl. n. *Freyrarchaeaceae*, the *Freyrarchaeum* family).

The family is described based on 37 concatenated conserved marker genes phylogeny. The description is the same as that of its sole genus and species. Type genus is *Candidatus* Freyrarchaeum.

***Candidatus* Freyrarchaeales** (Freyr.ar.chae.a'les. N.L. neut. n. *Freyrarchaeum*, Candidatus generic name; -ales ending to denote the order; N.L. fem. pl. n. *Freyrarchaeales*, the *Freyrarchaeum* order).

The order is described based on 37 concatenated conserved marker genes phylogeny. The description is the same as that of its sole genus and species. Type genus is *Candidatus* Freyrarchaeum.

***Candidatus* Freyrarchaeia** (Freyr.ar.chae'i.a. N.L. neut. n. *Freyrarchaeum*, Candidatus generic name; -ia ending to denote the class; N.L. fem. pl. n. *Freyrarchaeia*, the *Freyrarchaeum* class).

The class is described based on 37 concatenated conserved marker genes phylogeny. The description is the same as that of its sole genus and species. Type genus is *Candidatus* Freyrarchaeum.

***Candidatus* Freyrarchaeota** (Freyr.ar.chae.o'ta. N.L. neut. n. *Freyrarchaeum*, Candidatus generic name; -ota ending to denote the phylum; N.L. fem. pl. n. *Freyrarchaeota*, the *Freyrarchaeum* phylum).

The phylum is described based on 37 concatenated conserved marker genes phylogeny. The description is the same as that of its sole genus and species. Type genus is *Candidatus* Freyrarchaeum.

***Candidatus* Sigynarchaeaceae** (Sigyn.ar.chae.ace'ae. N.L. neut. n. *Sigynarchaeum*, Candidatus generic name; -aceae ending to denote the family; N.L. fem. pl. n. *Sigynarchaeaceae*, the *Sigynarchaeum* family).

The family is described based on 37 concatenated conserved marker genes phylogeny. The description is the same as that of its sole genus and species. Type genus is *Candidatus* Sigynarchaeum.

***Candidatus* Sigynarchaeales** (Sigyn.ar.chae.a'les. N.L. neut. n. *Sigynarchaeum*, Candidatus generic name; -ales ending to denote the order; N.L. fem. pl. n. *Sigynarchaeales*, the *Sigynarchaeum* order).

The order is described based on 37 concatenated conserved marker genes phylogeny. The description is the same as that of its sole genus and species. Type genus is *Candidatus* Sigynarchaeum.

***Candidatus* Sigynarchaeia** (Sigyn.ar.chae'i.a. N.L. neut. n. *Sigynarchaeum*, Candidatus generic name; -ia ending to denote the class; N.L. fem. pl. n. *Sigynarchaeia*, the *Sigynarchaeum* class).

The class is described based on 37 concatenated conserved marker genes phylogeny. The description is the same as that of its sole genus and species. Type genus is *Candidatus* Sigynarchaeum.

***Candidatus* Sigynarchaeota** (Sigyn.ar.chae.o'ta. N.L. neut. n. *Sigynarchaeum*, Candidatus generic name; -ota ending to denote the phylum; N.L. fem. pl. n. *Sigynarchaeota*, the *Sigynarchaeum* phylum).

The phylum is described based on 37 concatenated conserved marker genes phylogeny. The description is the same as that of its sole genus and species. Type genus is *Candidatus* Sigynarchaeum.

***Candidatus* Balderarchaeaceae** (Balder.ar.chae.ace'ae. N.L. neut. n. *Balderarchaeum*, Candidatus generic name; -aceae ending to denote the family; N.L. fem. pl. n. *Balderarchaeaceae*, the *Balderarchaeum* family).

The family is described based on 37 concatenated conserved marker genes phylogeny. The description is the same as that of its sole genus and species. Type genus is *Candidatus* Balderarchaeum.

***Candidatus* Balderarchaeales** (Balder.ar.chae.a'les. N.L. neut. n. *Balderarchaeum*, Candidatus generic name; -ales ending to denote the order; N.L. fem. pl. n. *Balderarchaeales*, the *Balderarchaeum* order).

The order is described based on 37 concatenated conserved marker genes phylogeny. The description is the same as that of its sole genus and species. Type genus is *Candidatus* Balderarchaeum.

***Candidatus* Balderarchaeia** (Balder.ar.chae'i.a. N.L. neut. n. *Balderarchaeum*, Candidatus generic name; -ia ending to denote the class; N.L. fem. pl. n. *Balderarchaeia*, the *Balderarchaeum* class).

435 The class is described based on 37 concatenated conserved marker genes phylogeny. The
436 description is the same as that of its sole genus and species. Type genus is *Candidatus*
437 Balderarchaeum.

438 ***Candidatus* Balderarchaeota** (Balder.ar.chae.o'ta. N.L. neut. n. *Balderarchaeum*,
439 Candidatus generic name; -ota ending to denote the phylum; N.L. fem. pl. n. *Balderarchaeota*,
440 the *Balderarchaeum* phylum).

441 The phylum is described based on 37 concatenated conserved marker genes phylogeny. The
442 description is the same as that of its sole genus and species. Type genus is *Candidatus*
443 Balderarchaeum.

444 ***Candidatus* Njordarchaeaceae** (Njord.ar.chae.ace'ae. N.L. neut. n. *Njordarchaeum*,
445 Candidatus generic name; -aceae ending to denote the family; N.L. fem. pl. n.
446 *Njordarchaeaceae*, the *Njordarchaeum* family).

447 The family is described based on 37 concatenated conserved marker genes phylogeny. The
448 description is the same as that of its sole genus and species. Type genus is *Candidatus*
449 Njordarchaeum.

450 ***Candidatus* Njordarchaeales** (Njord.ar.chae.a'les. N.L. neut. n. *Njordarchaeum*, Candidatus
451 generic name; -ales ending to denote the order; N.L. fem. pl. n. *Njordarchaeales*, the
452 *Njordarchaeum* order).

453 The order is described based on 37 concatenated conserved marker genes phylogeny. The
454 description is the same as that of its sole genus and species. Type genus is *Candidatus*
455 Njordarchaeum.

***Candidatus* Njordarchaeia** (Njord.ar.chae'i.a. N.L. neut. n. *Njordarchaeum*, Candidatus generic name; -ia ending to denote the class; N.L. fem. pl. n. *Njordarchaeia*, the *Njordarchaeum* class).

The class is described based on 37 concatenated conserved marker genes phylogeny. The description is the same as that of its sole genus and species. Type genus is *Candidatus* Njordarchaeum.

***Candidatus* Njordarchaeota** (Njord.ar.chae.o'ta. N.L. neut. n. *Njordarchaeum*, Candidatus generic name; -ota ending to denote the phylum; N.L. fem. pl. n. *Njordarchaeota*, the *Njordarchaeum* phylum).

The phylum is described based on 37 concatenated conserved marker genes phylogeny. The description is the same as that of its sole genus and species. Type genus is *Candidatus* Njordarchaeum.

469 **Methods**

470 **Sampling and processing.** Detailed methods for collection, DNA extraction, and
471 metagenome sequencing of Guaymas Basin samples has been described in previous study[49].
472 Six sediment samples of hot spring were taken from Tengchong, Yunnan, China on September,
473 2019 (24.95°N, 98.44°E). DNA was extracted from 10 g of each sample by using PowerSoil
474 DNA Isolation Kit (Mo Bio). Metagenomic sequence data for the six samples were generated
475 using Illumina HiSeq 2500 instruments.

476 **Data collection.** Asgard archaea are distributed mainly in different environmental sediments,
477 including estuary sediments[18], mangrove sediments[17], hydrothermal sediments[50], hot spring
478 sediments[10], marine sediments[9], and freshwater sediments[51]. According to the environmental
479 distributions of Asgard archaea, metagenomic data were collected and downloaded from the
480 SRA database (https://www.ncbi.nlm.nih.gov/sra/).

481 **Metagenomic assembly and genomic binning.** The raw reads were trimmed using
482 Trimmomatic (v.0.38)[52] to remove adapters and low-quality reads. After trimming, the reads
483 of each sample were *de novo* assembled using Megahit (v.1.2.5)[53] with a k-step of 6. Samples
484 from the same location or similar environments were assembled together. Contigs were
485 binned separately using MetaBAT (v.2.12.1)[54], MaxBin (v.2.2.7)[55], and Concoct (v.1.1.0)[56]
486 with the default parameters, and the initial taxonomic classification of each MAGs was
487 performed using GTDB-Tk (v.1.2.0)[57] to extract Asgard MAGs. The completeness and
488 contamination of Asgard MAGs were evaluated with the CheckM lineage_wf workflow
489 (v.1.0.12)[58]. Finally, Asgard MAGs with completeness above 50% and contamination below
490 10% were selected for further analyses, and Prodigal (v.2.6.1)[59] was used to predict

491 protein-coding genes for these selected Asgard MAGs.

492 **Phylogenetic analyses of Asgard MAGs.** To determine the exact phylogenetic affiliations in
493 the Asgard superphylum, 37 conserved marker genes were selected as described in the
494 literature[24, 60]. Homologs of the 37 conserved marker proteins were identified using Diamond
495 (v.2.0.4)[61]. Each dataset of marker proteins was aligned with MAFFT-L-INS-i (v.7.313)[62] and
496 trimmed by trimAl (v.1.4.22)[63] with the "automated1" option. Maximum-likelihood
497 phylogenies for the 37 conserved marker proteins was built using IQ-TREE (v.2.0.5)[64] under
498 the model "LG+F+G4+C60". The support values were calculated using 1000 ultrafast
499 bootstraps.

500 **Phylogenetic tree of life.** To confirm the phylogenetic affiliations of eukaryotes and the novel
501 Asgard lineages, 21 taxonomic marker genes shared among three domains selected by
502 Williams[14] and 54 ribosomal proteins shared between archaea and eukaryotes[10] were used for
503 phylogenetic analyses. Single-gene trees were inferred for all the markers using IQ-TREE
504 with the LG+G4+F model to exclude phylogenetic artefacts such as long branch attraction,
505 HGT (eukaryotic genes falling into bacterial clade or scattering in archaeal clade were
506 considered as HGT) and a BLASTp inspection was further performed to identify all
507 eukaryotic genes that originated from the nuclear genome and to remove genes of
508 mitochondrial or chloroplastic origin. The maximum-likelihood tree was built with IQ-TREE
509 under the LG+C60+F+G4 model with 1000 ultrafast bootstraps.

510 **Identification of ESPs.** All predicted proteins encoded by the MAGs of the five novel Asgard
511 lineages and Hermodarchaeota were analyzed using InterProScan[65] (v.5.47-82.0) with default
512 parameters to annotate protein domains and were assigned to archaeal clusters of orthologous

genes (arCOGs)[66] by eggnog-mapper (v. 2.0.1b)[67] with default settings.. The lists of InterPro accession numbers (IPRs) and arCOG identifiers previously published by Zaremba-Niedzwiedzka *et al.*[10] and Bulzu *et al.*[16] were used to identify potential ESPs. Some key words related to eukaryote-specific processes or cell structures were used to search annotation information of interProScan to identify potential ESPs previously not reported in Asgard archaea. Several candidate ESPs were further examined using HHpred[68] with default parameters.

**Metabolic reconstruction.** The proteome of each MAG reported in the present study was uploaded to the KEGG Automatic Annotation Server (KAAS)[69] and run with several settings: the GHOSTX, Prokaryotes and Bidirectional Best Hit (BBH) settings. Additionally, proteins were queried against the nonredundant (NR) protein database (downloaded from NCBI on February 2020) using the Diamond (v.2.0.4) BLASTp search (e-value cutoff $<1e-5$). Metabolic pathways were reconstructed based on combination of the NR annotations, protein domain information and KEGG Ontology (KO) numbers.

The dbCAN2[70] web server was used to identify carbohydrate-degrading enzymes with the default settings. The putative large subunits of [NiFe] hydrogenases were identified by querying against a local database based on HydDB[71] using Diamond (v.2.0.4) with an E-value cutoff of $1e^{-20}$ and sequences containing CxxC motifs in both N-terminal and C-terminal were considered as hydrogenases. Additionally, a local MEROPS database (downloaded September 2020)[72] searched for peptidases by Diamond (v.2.0.4) with an E-value cutoff of $1\times10^{-20}$, and PSORT (v.3.0.2) was used to identify protein localization[73].

534 **Calculation of ANI and average AAI.** The ANI and AAI values were calculated using
535 OrthoANI (v.1.2)[74] and CompareM (https://github.com/dparks1134/CompareM), respectively,
536 with the default parameters.
537

538 **Acknowledgments**

539 We are grateful for Dr. Tom A. William for his suggestions regarding phylogenetic analysis.
540 We thank Brett Baker and Nina Dombrowski for allowing us use their metagenomic data
541 freely. These sequence data were produced by the US Department of Energy Joint Genome
542 Institute http://www.jgi.doe.gov/ in collaboration with the user community and the datasets
543 used in the current study along with the contributors' names are listed in Supplemental Table
544 2.

545 **Data availability**

546 The genomes of Asgard archaea generated in this study have been made available at the
547 eLibrary of Microbial Systematics and Genomics (eLMSG;
548 https://www.biosino.org/elmsg/index) under accession numbers
549 LMSG_G000000610.1-LMSG_G000000628.1.
550 The initial phylogenetic trees have been deposited at figshare and can be accessed at the
551 following link:
552 https://figshare.com/s/c20d9eccb7e4591b429c.

553 **Funding**

554 This work was supported by the Natural Science Foundation of China (Grant No. 91751205,
555 41525011), the National Key Research and Development Project of China (Grant No.

2016YFA0601102), the Senior User Project of RV KEXUE (KEXUE2019GZ06).

**Author contributions**

R.Z.X., Y.Z.W. and F.P.W. conceived the study. R.Z.X., Y.Z.W., D.Y.H., H.J.L., H.N.H., L.Y.L. and X.X.Z. analyzed the data. R.Z.X., Y.Z.W. and F.P.W. wrote the paper.

**Compliance and ethics**

The author(s) declare that they have no conflicts of interest.

562 **References**


563 1. Embley TM, Martin W. Eukaryotic evolution, changes and challenges. *Nature* **440**, 623-630
564 (2006).

565 2. López-García P, Moreira D. Open questions on the origin of eukaryotes. *Trends in ecology &*
566 *evolution* **30**, 697-708 (2015).

567 3. Rochette NC, Brochier-Armanet C, Gouy M. Phylogenomic test of the hypotheses for the
568 evolutionary origin of eukaryotes. *Molecular biology and evolution* **31**, 832-845 (2014).

569 4. Lopez-Garcia P, Moreira D. Selective forces for the origin of the eukaryotic nucleus.
570 *Bioessays* **28**, 525-533 (2006).

571 5. López-García P, Moreira D. Metabolic symbiosis at the origin of eukaryotes. *Trends in*
572 *biochemical sciences* **24**, 88-93 (1999).

573 6. Martin WF, Garg S, Zimorski V. Endosymbiotic theories for eukaryote origin. *Philosophical*
574 *Transactions of the Royal Society B: Biological Sciences* **370**, 20140330 (2015).

575 7. Esser C*, et al.* A genome phylogeny for mitochondria among α-proteobacteria and a
576 predominantly eubacterial ancestry of yeast nuclear genes. *Molecular Biology and Evolution*
577 **21**, 1643-1660 (2004).

578 8. Moreira D, López-García P. Symbiosis between methanogenic archaea and δ-proteobacteria as
579 the origin of eukaryotes: the syntrophic hypothesis. *Journal of Molecular Evolution* **47**,
580 517-530 (1998).

581 9. Spang A*, et al.* Complex archaea that bridge the gap between prokaryotes and eukaryotes.
582 *Nature* **521**, 173-179 (2015).

583 10. Zaremba-Niedzwiedzka K*, et al.* Asgard archaea illuminate the origin of eukaryotic cellular
584 complexity. *Nature* **541**, 353 (2017).

585 11. Spang A*, et al.* Proposal of the reverse flow model for the origin of the eukaryotic cell based
586 on comparative analyses of Asgard archaeal metabolism. *Nature microbiology*, 1 (2019).

587 12. Williams TA, Foster PG, Cox CJ, Embley TM. An archaeal origin of eukaryotes supports only
588 two primary domains of life. *Nature* **504**, 231-236 (2013).

589 13. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the
590 domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences* **87**,
591 4576-4579 (1990).

592 14. Williams TA, Cox CJ, Foster PG, Szollosi GJ, Embley TM. Phylogenomics provides robust
593 support for a two-domains tree of life. *Nat Ecol Evol* **4**, 138-147 (2020).
594 15. Hartman H, Fedorov A. The origin of the eukaryotic cell: a genomic investigation.
595 *Proceedings of the National Academy of Sciences* **99**, 1420-1425 (2002).
596 16. Bulzu P-A*, et al.* Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche.
597 *Nature microbiology* **4**, 1129-1137 (2019).
598 17. Liu Y, Zhou Z, Pan J, Baker BJ, Gu J-D, Li M. Comparative genomic inference suggests
599 mixotrophic lifestyle for Thorarchaeota. *The ISME journal* **12**, 1021-1031 (2018).
600 18. Seitz KW, Lazar CS, Hinrichs K-U, Teske AP, Baker BJ. Genomic reconstruction of a novel,
601 deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur
602 reduction. *The ISME journal* **10**, 1696-1705 (2016).
603 19. Wong HL, White RA, Visscher PT, Charlesworth JC, Vázquez-Campos X, Burns BP.
604 Disentangling the drivers of functional complexity at the metagenomic level in Shark Bay
605 microbial mat microbiomes. *The ISME journal* **12**, 2619-2639 (2018).
606 20. MacLeod F, Kindler GS, Wong HL, Chen R, Burns BP. Asgard archaea: diversity, function,
607 and evolutionary implications in a range of microbiomes. *AIMS microbiology* **5**, 48 (2019).
608 21. Imachi H*, et al.* Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* **577**,
609 519-525 (2020).
610 22. Akıl C*, et al.* Insights into the evolution of regulated actin dynamics via characterization of
611 primitive gelsolin/cofilin proteins from Asgard archaea. *Proceedings of the National Academy*
612 *of Sciences* **117(33)**, 19904-19913 (2020).
613 23. Zhang R-Y*, et al.* Design of targeted primers based on 16S rRNA sequences in
614 meta-transcriptomic datasets and identification of a novel taxonomic group in the Asgard
615 archaea. *BMC microbiology* **20**, 25 (2020).
616 24. Wang Y, Wegener G, Hou J, Wang F, Xiao X. Expanding anaerobic alkane metabolism in the
617 domain of Archaea. *Nat Microbiol* **4**, 595-602 (2019).
618 25. Zaremba-Niedzwiedzka K*, et al.* Asgard archaea illuminate the origin of eukaryotic cellular
619 complexity. *Nature* **541**, 353-358 (2017).
620 26. Luo C, Rodriguez-r LM, Konstantinidis KT. MyTaxa: an advanced taxonomic classifier for
621 genomic and metagenomic sequences. *Nucleic acids research* **42**, e73-e73 (2014).

622 27. Seitz KW, *et al.* Asgard archaea capable of anaerobic hydrocarbon cycling. *Nat Commun* **10**,
623 1822 (2019).
624 28. Raiborg C, Stenmark H. The ESCRT machinery in endosomal sorting of ubiquitylated
625 membrane proteins. *Nature* **458**, 445-452 (2009).
626 29. Grau-Bové X, Sebé-Pedrós A, Ruiz-Trillo I. The Eukaryotic Ancestor Had a Complex
627 Ubiquitin Signaling System of Archaeal Origin. *Molecular Biology and Evolution* **32**, 726-739
628 (2015).
629 30. Leung KF, Dacks JB, Field MC. Evolution of the multivesicular body ESCRT machinery;
630 retention across the eukaryotic lineage. *Traffic* **9**, 1698-1716 (2008).
631 31. Field MC, Dacks JB. First and last ancestors: reconstructing evolution of the endomembrane
632 system with ESCRTs, vesicle coat proteins, and nuclear pore complexes. *Curr Opin Cell Biol*
633 **21**, 4-13 (2009).
634 32. Schulz BL, *et al.* Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and
635 Ost6p defines site-specific glycosylation efficiency. *Proc Natl Acad Sci U S A* **106**,
636 11061-11066 (2009).
637 33. Park SY, Guo X. Adaptor protein complexes and intracellular transport. *Bioscience reports* **34**,
638 (2014).
639 34. Hirst J, *et al.* The fifth adaptor protein complex. *PLoS Biol* **9**, e1001170 (2011).
640 35. Tan JZA, Gleeson PA. Cargo sorting at the trans-Golgi network for shunting into specific
641 transport routes: role of Arf small G proteins and adaptor complexes. *Cells* **8**, 531 (2019).
642 36. Lazar CS, *et al.* Genomic evidence for distinct carbon substrate preferences and ecological
643 niches of B athyarchaeota in estuarine sediments. *Environmental Microbiology* **18**, 1200-1211
644 (2016).
645 37. He Y, *et al.* Genomic and enzymatic evidence for acetogenesis among multiple lineages of the
646 archaeal phylum Bathyarchaeota widespread in marine sediments. *Nature microbiology* **1**, 1-9
647 (2016).
648 38. Rother M, Metcalf WW. Anaerobic growth of Methanosarcina acetivorans C2A on carbon
649 monoxide: an unusual way of life for a methanogenic archaeon. *Proceedings of the National*
650 *Academy of Sciences* **101**, 16929-16934 (2004).
651 39. Fournier GP, Gogarten JP. Evolution of acetoclastic methanogenesis in Methanosarcina via

652 horizontal gene transfer from cellulolytic Clostridia. *Journal of bacteriology* **190**, 1124-1127
653 (2008).
654 40. Sousa FL, Martin WF. Biochemical fossils of the ancient transition from geoenergetics to
655 bioenergetics in prokaryotic one carbon compound metabolism. *Biochimica et Biophysica*
656 *Acta (BBA)-Bioenergetics* **1837**, 964-981 (2014).
657 41. Ma K, Schicho RN, Kelly RM, Adams MW. Hydrogenase of the hyperthermophile
658 Pyrococcus furiosus is an elemental sulfur reductase or sulfhydrogenase: evidence for a
659 sulfur-reducing hydrogenase ancestor. *Proc Natl Acad Sci U S A* **90**, 5341-5344 (1993).
660 42. Richter K, Haslbeck M, Buchner J. The heat shock response: life on the verge of death.
661 *Molecular cell* **40**, 253-266 (2010).
662 43. Spang A*, et al.* Proposal of the reverse flow model for the origin of the eukaryotic cell based
663 on comparative analyses of Asgard archaeal metabolism. *Nat Microbiol* **4**, 1138-1148 (2019).
664 44. Liu Y, Zhou Z, Pan J, Baker BJ, Gu JD, Li M. Comparative genomic inference suggests
665 mixotrophic lifestyle for Thorarchaeota. *ISME J* **12**, 1021-1031 (2018).
666 45. Fang H, Li D, Kang J, Jiang P, Sun J, Zhang D. Metabolic engineering of Escherichia coli for
667 de novo biosynthesis of vitamin B 12. *Nature communications* **9**, 1-12 (2018).
668 46. Doxey AC, Kurtz DA, Lynch MD, Sauder LA, Neufeld JD. Aquatic metagenomes implicate
669 Thaumarchaeota in global cobalamin production. *The ISME journal* **9**, 461-471 (2015).
670 47. Pan J*, et al.* Genomic and transcriptomic evidence of light-sensing, porphyrin biosynthesis,
671 Calvin-Benson-Bassham cycle, and urea production in Bathyarchaeota. *Microbiome* **8**, 1-12
672 (2020).
673 48. Cai M*, et al.* Diverse Asgard archaea including the novel phylum Gerdarchaeota participate in
674 organic matter degradation. *Science China Life Sciences*, 1-12 (2020).
675 49. Feng X, Wang Y, Zubin R, Wang F. Core metabolic features and hot origin of Bathyarchaeota.
676 *Engineering* **5**, 498-504 (2019).
677 50. Dombrowski N, Teske AP, Baker BJ. Expansive microbial metabolic versatility and
678 biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nature communications* **9**,
679 1-13 (2018).
680 51. Narrowe AB*, et al.* Complex evolutionary history of translation Elongation Factor 2 and
681 diphthamide biosynthesis in Archaea and parabasalids. *Genome biology and evolution* **10**,

682 2380-2393 (2018).
683 52. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
684 *Bioinformatics* **30**, 2114-2120 (2014).
685 53. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution
686 for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*
687 **31**, 1674-1676 (2015).
688 54. Kang DD*, et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome
689 reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
690 55. Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover
691 genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605-607 (2016).
692 56. Alneberg J*, et al.* Binning metagenomic contigs by coverage and composition. *Nature methods*
693 **11**, 1144-1146 (2014).
694 57. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes
695 with the Genome Taxonomy Database. *Bioinformatics*, 1925-1927 (2019).
696 58. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the
697 quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome*
698 *research* **25**, 1043-1055 (2015).
699 59. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic
700 gene recognition and translation initiation site identification. *BMC bioinformatics* **11**, 119
701 (2010).
702 60. Jay ZJ, Beam JP, Dlakić M, Rusch DB, Kozubal MA, Inskeep WP. Marsarchaeota are an
703 aerobic archaeal lineage abundant in geothermal iron oxide microbial mats. *Nature*
704 *microbiology* **3**, 732-740 (2018).
705 61. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND.
706 *Nature methods* **12**, 59-60 (2015).
707 62. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
708 improvements in performance and usability. *Molecular biology and evolution* **30**, 772-780
709 (2013).
710 63. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment
711 trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009).

712 64. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective
713 stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and*
714 *evolution* **32**, 268-274 (2015).
715 65. Jones P*, et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**,
716 1236-1240 (2014).
717 66. Makarova KS, Wolf YI, Koonin EV. Archaeal clusters of orthologous genes (arCOGs): an
718 update and application for analysis of shared features between Thermococcales,
719 Methanococcales, and Methanobacteriales. *Life* **5**, 818-840 (2015).
720 67. Huerta-Cepas J*, et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically
721 annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic acids*
722 *research* **47**, D309-D314 (2019).
723 68. Zimmermann L*, et al.* A completely reimplemented MPI bioinformatics toolkit with a new
724 HHpred server at its core. *Journal of molecular biology* **430**, 2237-2243 (2018).
725 69. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome
726 annotation and pathway reconstruction server. *Nucleic acids research* **35**, W182-W185 (2007).
727 70. Zhang H*, et al.* dbCAN2: a meta server for automated carbohydrate-active enzyme annotation.
728 *Nucleic Acids Research* **46**, W95-W101 (2018).
729 71. Søndergaard D, Pedersen CN, Greening C. HydDB: a web tool for hydrogenase classification
730 and analysis. *Scientific reports* **6**, 1-8 (2016).
731 72. Rawlings ND, Barrett AJ, Finn R. Twenty years of the MEROPS database of proteolytic
732 enzymes, their substrates and inhibitors. *Nucleic acids research* **44**, D343-D350 (2016).
733 73. Horton P*, et al.* WoLF PSORT: protein localization predictor. *Nucleic acids research* **35**,
734 W585-W587 (2007).
735 74. Lee I, Kim YO, Park S-C, Chun J. OrthoANI: an improved algorithm and software for
736 calculating average nucleotide identity. *International journal of systematic and evolutionary*
737 *microbiology* **66**, 1100-1103 (2016).

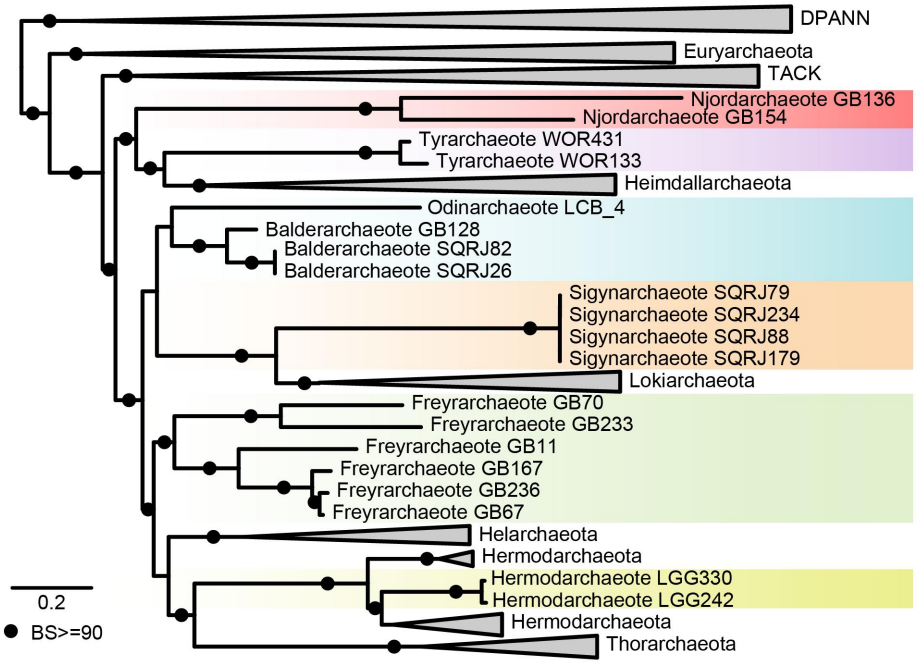


**Figure 1. Phylogenetic tree of the Asgard archaea using DPANN as an outgroup.** Maximum-likelihood tree of 37 concatenated marker proteins inferred with the LG+F+C60+G4 model in IQ-TREE; The bootstrap support values above 90 were shown with black filled circles. 19 representatives of DPANN, 55 representatives of Euryarchaeota, 29 representatives of TACK and 68 genomes (including five new lineages) of Asgard were used to infer phylogenetic tree.

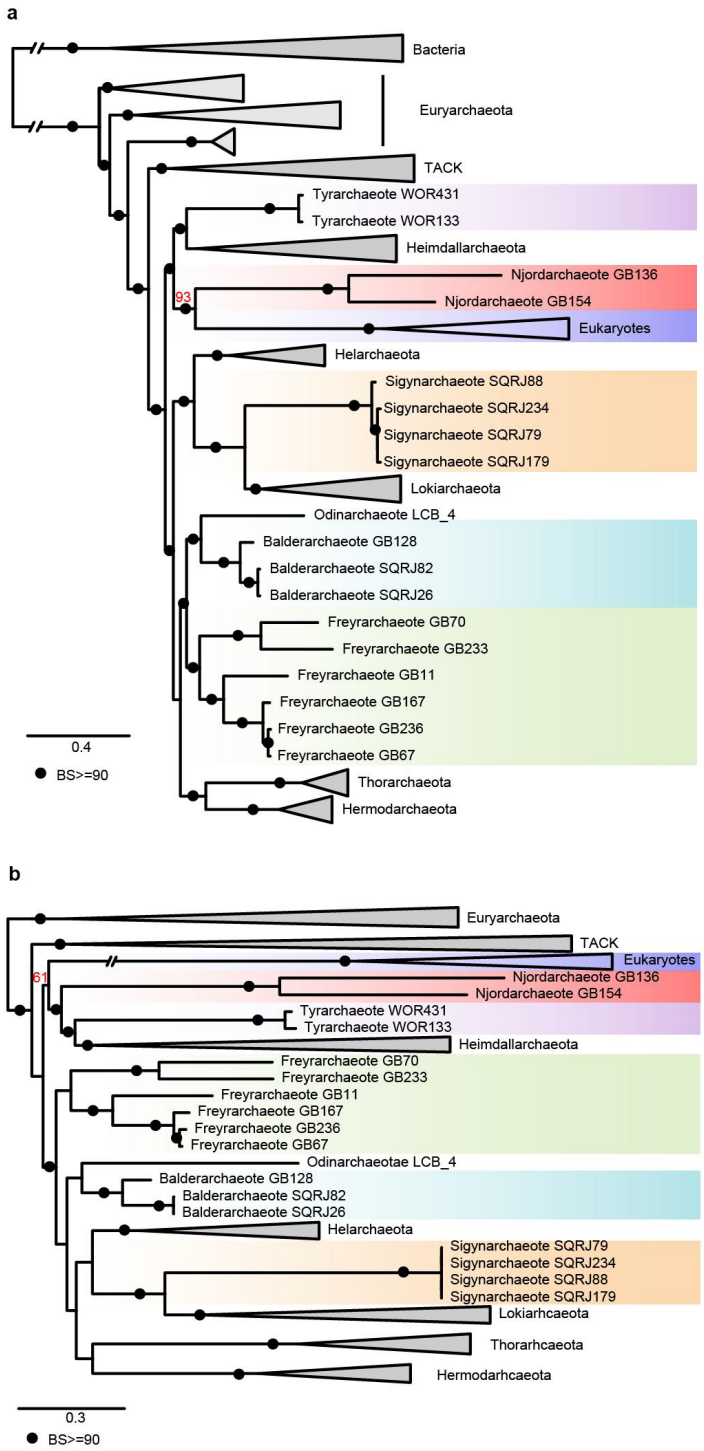


**Figure 2. Phylogenetic affiliations of bacteria, archaea and eukaryotes. a**, Maximum likelihood inference of 21 concatenated conserved protein sequences under the LG+F+C60+G4 model rooted in bacteria; **b**, Maximum-likelihood analysis of 54 archaeal-eukaryotic ribosomal proteins under the LG+F+C60+G4 model rooted in Euryarchaeota.

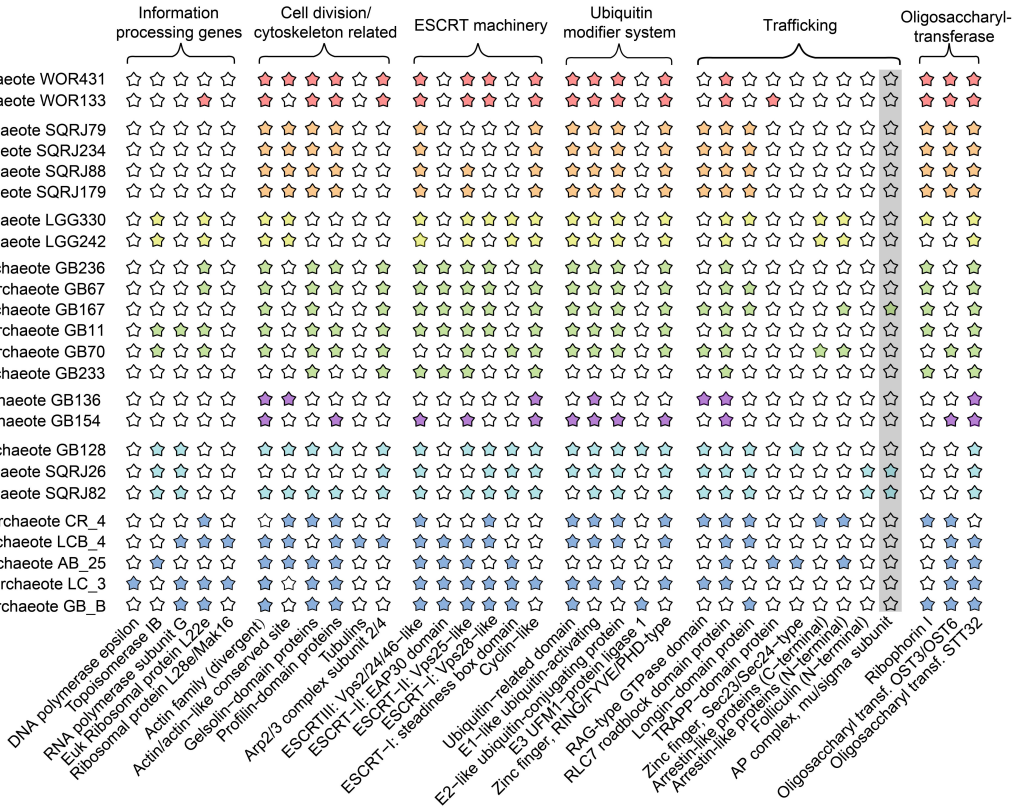


**Figure 3. Comparison of the distributions of ESPs in the five new Asgard lineages and other representative Asgard clades**. Colored stars indicate the presence of ESPs, whereas empty stars indicate the absence of ESPs. The grep box highlights ESP identified in this study. The new ESP, mu/sigma subunit of AP complex, was detected in Balderarchaeote SQRJ26, Balderarchaeote SQRJ82 and Freyrarchaeote GB167.

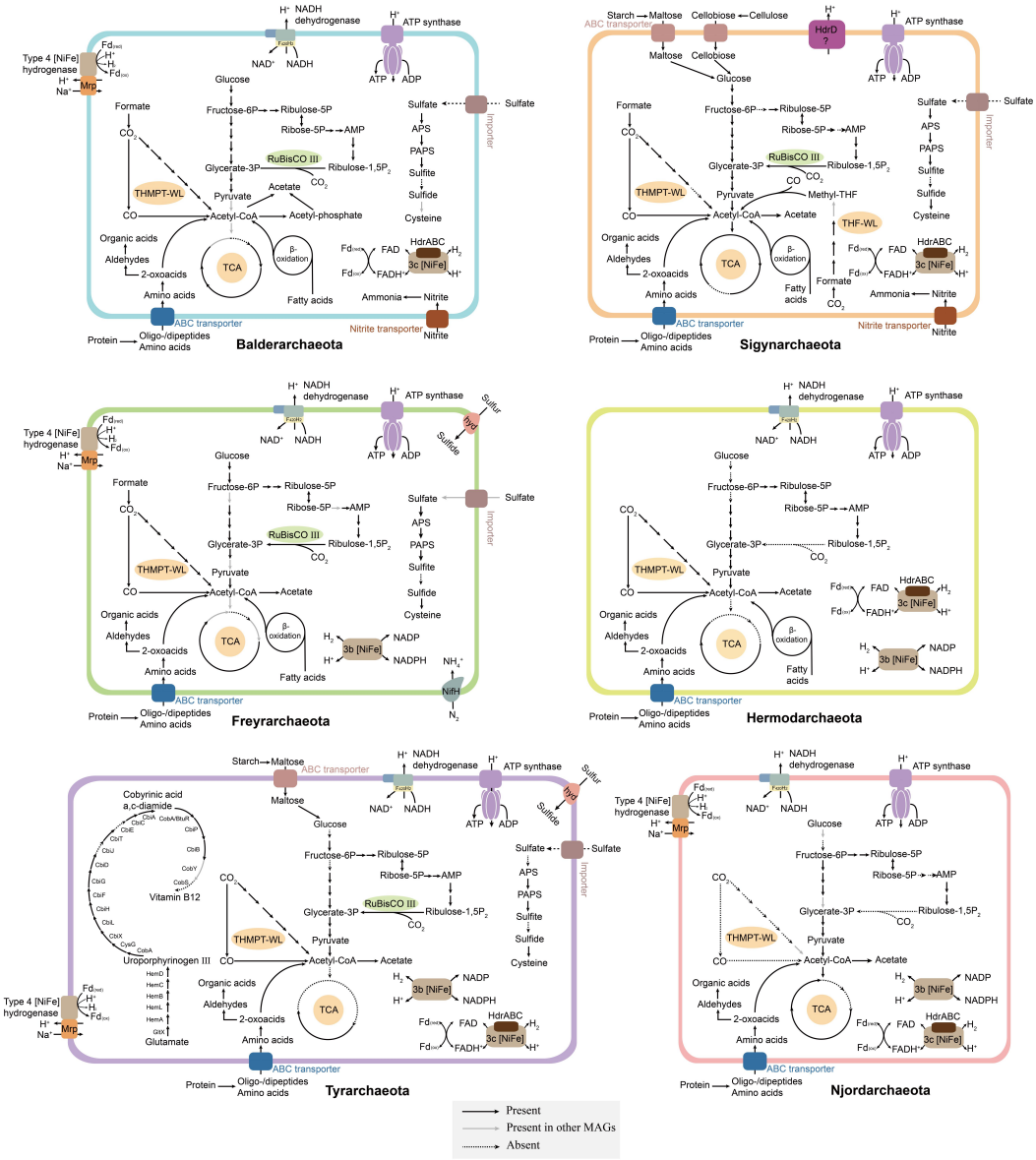


**Figure 4. Inferred metabolic pathways of the five new Asgard lineages and the new clade of Hermodarchaeota based on genes identified using the KEGG database and the NCBI NR protein database**. A black line indicates that a component/process is present in representative MAGs, a grey line indicates that a component/process is present in other MAGs, and a dashed line indicates that a certain pathway or enzyme is absent from all genomes. The representatives of the different lineages are as follows: Balderarchaeota, SQRJ26; Freyrarchaeota, GB11; Njordarchaeota, GB154; Sigynarchaeota, SQRJ79; and Tyrarchaeota, WOR431; Hermodarchaeota, LGG330. Details about the genes are provided in Supplementary Table 6. Hdr, heterodisulfide reductase; TCA, tricarboxylic acid cycle; THMPT-WL, tetrahydromethanopterin Wood-Ljungdahl pathway; THF-WL, tetrahydrofolate Wood-Ljungdahl; Mrp, Mrp $Na^+/H^+$ antiporters; hyd, sulfhydrogenase; AMP, AMP phosphorylase.