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 and a new Asgard lineage Njordarchaeota was supposed as the known closest branch with the eukaryotic nuclear host lineage. Metabolic reconstruction suggests that the Asgard archaea described here have potential to fix inorganic carbon via the Wood-Ljungdahl pathway and degrade organic matters except Njordarchaeota, which may possess a heterotrophic lifestyle with capability of peptides and amino acids utilization. Additionally, the Ack/Pta pathway for homoacetogenesis and *de novo* anaerobic cobalamin biosynthesis pathway were found in Balderarchaeota and Tyrarchaeota, respectively. 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: Asgard archaea, eukaryotic origin, metagenome

42 **Introduction**

43 The origin of eukaryotes is considered as a critical biological evolutionary event on Earth
44 (Embley and Martin 2006, López-García and Moreira 2015, Rochette et al., 2014,
45 Lopez-Garcia and Moreira 2006). The common ancestor of eukaryotes is generally believed
46 to have evolved from a symbiotic process (López-García and Moreira 1999, Martin et al.,
47 2015) in which one endosymbiotic bacterium within the Proteobacteria phylum evolved into a
48 mitochondrion (Esser et al., 2004, Moreira and López-García 1998) and one endosymbiotic
49 host cell became the cell nucleus (Spang et al., 2015, Zaremba-Niedzwiedzka et al., 2017a,
50 Spang et al., 2019a). The identity of the host cell ancestor has been vigorously debated, and
51 two hypotheses regarding 2- or 3-domain trees of life have been raised (Williams et al., 2013,
52 Woese et al., 1990). However, increasing evidence provided by phylogenomic analyses
53 (Zaremba-Niedzwiedzka et al., 2017a, Williams et al., 2020), as well as the presence of
54 eukaryotic signature proteins (ESPs) (Hartman and Fedorov 2002) in the Asgard archaea, has
55 supported the idea that eukaryotic cells originated from an archaeal host cell related to
56 archaeal Asgard superphylum (Spang et al., 2015, Zaremba-Niedzwiedzka et al., 2017a,
57 Bulzu et al., 2019). The identification of Lokiarchaeota in the Loki's Castle hydrothermal
58 vent field provided pivotal genomic and phylogenetic evidence that eukaryotes originated
59 within the domain Archaea, supporting a 2-domain tree of life, which is consistent with the
60 eocyte hypothesis (Spang et al., 2015). Further discovery and proposal of the Asgard
61 superphylum have provided new insights into the transition of archaea to eukaryotes and into
62 the origin of eukaryotic cell complexity (Zaremba-Niedzwiedzka et al., 2017a). Within the
63 Asgard superphylum, Heimdallarchaeota had been identified to be the closest Asgard archaeal

64 lineage to the eukaryotic branch on the phylogenetic tree on the basis of carefully selected
65 conserved protein sequences (Zaremba-Niedzwiedzka et al., 2017a, Williams et al., 2020).
66 Recently, Imachi *et al.* cultivated one Asgard archaeon, *Candidatus* Prometheoarchaeum
67 syntrophicum strain MK-D1, in the laboratory and observed, for the first time, the
68 intertwining of this archaeon with bacterial cells via extracellular protrusions under a
69 transmission electron microscope (Imachi et al., 2020). The idea of an archaeal origin of
70 eukaryotes and a 2-domain tree of life has recently become increasingly favorable (Williams
71 et al., 2020, Akıl et al., 2020).The Asgard archaea are described as mixotrophic or
72 heterotrophic (Spang et al., 2019a, Liu et al., 2018a) and are ubiquitously distributed in
73 various environments, such as hydrothermal vents (Spang et al., 2015,
74 Zaremba-Niedzwiedzka et al., 2017a); lake, river and marine sediments (Seitz et al., 2016);
75 microbial mats (Wong et al., 2018); and mangroves (Liu et al., 2018a). These organisms
76 potentially play important roles in global geochemical cycling (MacLeod et al., 2019);
77 nevertheless, our understanding of the evolution of Asgard archaea, the archaea-eukaryote
78 transition, and the ecological and geochemical roles of these evolutionarily important archaea
79 remains incomplete. This lack of understanding is largely due to the limited number of
80 high-quality genomes of Asgard archaea, which are considered highly diverse as revealed by
81 16S rRNA gene surveys (MacLeod et al., 2019, Zhang et al., 2020); yet only a small fraction
82 have representative genomes. In this study, we described five previously unknown
83 phylum-level Asgard archaeal group, greatly expanding the Asgard genomic diversity within
84 the domain of Archaea and shed new light on the origin of eukaryotes.

85 **Results and Discussion**

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

87 In total, 11 metagenomic datasets were used in this study, including two samples from
88 hydrothermal sediment of Guaymas Basin, six samples from Tengchong hot spring sediment,
89 as well as 3 metagenomic datasets from the publicly available National Center for
90 Biotechnology Information (NCBI) Sequence Read Archive (SRA) database (Permission
91 granted, Supplementary Table 1). After subsequent assembling, binning and classification as
92 described in the Methods section, 128 Asgard metagenome-assembled genomes (MAGs)
93 were obtained and in-depth phylogenomic analyses were performed with 37 concatenated
94 conserved proteins (Wang et al., 2019) under LG+C60+F+G4 model to confirm the
95 placement of these MAGs on phylogenomic tree. The analysis revealed that, in addition to the
96 previously described Loki-, Thor-, Odin-, Heimdall-, Hela- and Hermodarchaeota clades
97 (Zaremba-Niedzwiedzka et al., 2017b, Zhang et al., 2021), there are five additional
98 monophyletic branching clades (Fig. 1a), here tentatively named Tyr-, Sigyn-, Freyr-, Njord-
99 and Balderarchaeota after the Asgard gods in the Norse mythology (Tyr, the god of war;
100 Sigyn, the god of victory; Freyr, the god of peace; Njord, the god of seas; and Balder, the god
101 of light). The MAGs of these new Asgard lineages were recovered from different
102 environments: Njordarchaeota and Freyrarchaeota were derived from hydrothermal sediment;
103 Tyrarchaeota were found in estuary sediments; Sigynarchaeota were reconstructed from hot
104 spring sediments and Balderarchaeota were retrieved from hot spring and hydrothermal
105 sediments. Additionally, a clade of Hermodarchaeota was identified in high temperature
106 habitats (~85℃), similar with Odinarchaeota, which were considered as the only thermophilic
107 member of Asgard archaea to date (Zaremba-Niedzwiedzka et al., 2017a). Near-complete

108 MAGs ranging in size from 2.1 to 5.5 Mb with completeness ranging from 87.38 to 97.20%
109 were constructed for representatives of each new Asgard clade (Supplementary Table 2). To
110 further assess their distinctiveness compared to the Asgard members already defined, we
111 calculated the average nucleotide identity (ANI) (Supplementary Fig. 1) and average amino
112 acid identity (AAI) (Supplementary Fig. 2) between them and other Asgard MAGs. The AAI
113 values showed that all the MAGs of new lineages discovered here share a low AAI with the
114 known Asgard archaea (<50%) and fall within the phylum-level classification range
115 (40%~52%) (Luo et al., 2014), providing additional support for the uniqueness of these new
116 Asgard lineages.

117 To determine the phylogenetic positions of these new Asgard lineages in relation to
118 eukaryotes, we performed comprehensive phylogenetic analyses using 21 conserved marker
119 proteins carefully selected by Williams *et al*. (Williams et al., 2020). The taxa included in
120 these analyses were also selected on the basis of the instructions of Williams *et al.* (Williams
121 et al., 2020): a representative taxon set was constructed comprising 85 archaeal genomes (53
122 within Asgard), 19 eukaryotic genomes, and 36 bacterial genomes (Supplementary Table 3).
123 Single-gene datasets were also carefully inspected with single protein trees to avoid potential
124 phylogenetic artifacts resulting from HGT, long branch attraction (LBA) and BLASTp
125 inspection was performed to remove eukaryotic genes from the mitochondria or plastids.
126 After concatenating, the maximum likelihood tree under the LG+C60+F+G4 model for 21
127 conserved marker proteins was inferred. The phylogenetic analysis of 21 marker genes
128 suggested that eukaryotes are the sister group of Njordarchaeota with high support (bootstrap

129 support (BS) = 93, Fig. 1b). The phylogenomic analyses provide strong support for a
130 2-domain tree with Njordarchaeota as the closest potential relatives to eukaryotes.

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

132 The potential ESPs were identified from the newly reconstructed Asgard MAGs (Fig. 2).
133 Consistent with previous reports (Spang et al., 2015, Zaremba-Niedzwiedzka et al., 2017a,
134 Bulzu et al., 2019, Seitz et al., 2019), different key subunits of informational processing
135 machinery were found. For example, topoisomerase IB protein-encoding genes were
136 identified in Balderarchaeota and Hermodarchaeota, while all MAGs of Balderarchaeota were
137 found to encode a RNA polymerase subunit G. Homologues of eukaryotic ribosomal protein
138 L22e were identified in Freyr-, Hermodarchaeota and Tyrarchaeota. The new clades of
139 Asgard archaea were also found to contain genes related to cell division and the cytoskeleton,
140 but tubulin-encoding genes were not detected. With regard to actin-related proteins, two to
141 three related subunits were detected in Tyr- and Balderarchaeota, whereas profilin domain
142 protein-encoding genes were identified in all Asgard lineages described here.

143 Ubiquitin-based signaling is an important cellular process in eukaryotes (Raiborg and
144 Stenmark 2009). Previous studies have reported the presence of the related protein domains in
145 Loki-, Odin-, Hel- and Heimdallarchaeota but not in Thorarchaeota (Spang et al., 2015, Seitz
146 et al., 2019, Grau-Bové et al., 2015). Here, we identified ubiquitin system-related
147 protein-encoding genes in nearly all newly assembled MAGs including several
148 ubiquitin-related domains, zinc fingers, ubiquitin-activating enzyme (E1),
149 ubiquitin-conjugating protein (E2), and UFM1-protein ligase 1 (E3), indicating that the
150 ubiquitin system is widespread in Asgard archaea.

151 The endosomal sorting complex required for transport (ESCRT) machinery consisting of
152 complexes I-III and associated subunits (Spang et al., 2015, Leung et al., 2008, Field and
153 Dacks 2009) were identified in the newly recovered MAGs. Genes coding for Vps28
154 domain-containing proteins previously found in Loki-, Odin-, Hel- and Heimdallarchaeota
155 (Zaremba-Niedzwiedzka et al., 2017b, Seitz et al., 2019) were also identified in Tyr-, Freyr-,
156 Balder- and Hermodarchaeota but were absent in Sigynarchaeota and Njordarchaeota. The
157 Sigynarchaeota MAGs also lack genes for both EAP30 domain- and steadiness box
158 domain-containing proteins. Notably, all new Asgard MAGs contain cyclin-like
159 protein-encoding genes, whereas ESCRT complexes I-III were only identified in
160 Freyrarchaeota.

161 ESPs with intracellular trafficking and secretion functions were also identified in MAGs
162 here. However, only Tyrarchaeota contains genes coding for proteins with homology to
163 TRAPP-domain protein and Sec23/24-type protein-encoding gene was only found in
164 Balderarchaeota. All MAGs reported in present study possess genes coding RLC7 roadblock
165 domain protein. Genes coding for both the N- and C-termini of arrestins were found in
166 Hermodarchaeota; the organization resembles one previously reported in the genome of
167 Lokiarchaeote CR-4, in which the C- and N-terminal domain proteins are separated from each
168 other by one gene (Zaremba-Niedzwiedzka et al., 2017b).

169 We also analyzed the oligosaccharyltransferase (OST) complex in the reconstructed
170 MAGs, and the results showed that OST complex-encoding genes were present in the MAGs
171 of all five Asgard clades. Ribophorin I homolog-encoding genes were found in the Hermod-,
172 Freyr-, Sigyn- and Tyrarchaeota MAGs. Homologs of OST3/6, which have been

173 demonstrated to influence yeast glycosylation efficiency (Schulz et al., 2009), were also
174 identified in several Asgard MAGs, while STT32 subunit protein-encoding genes were
175 detected all MAGs, consistent with previous reports on Loki-, Odin-, Thor-, Hel- and
176 Heimdallarchaeota (Zaremba-Niedzwiedzka et al., 2017b).

177 In addition to reported ESPs, we identified a potential ESP belonging to mu/sigma
178 subunit of AP (adaptor protein) complex-encoding genes in Balder- and Freyrarchaeota
179 MAGs (Fig. 2). Homologues of mu/sigma subunit of AP complex contain IPR022775 domain.
180 AP complexes are classified into AP-1, AP-2, AP-3, AP-4 and AP-5 and all AP complexes
181 are heterotetramers consisting of two large subunits (adaptins), one medium-sized subunit
182 (mu) and one small-sized subunit (sigma) (Park and Guo 2014, Hirst et al., 2011). AP
183 complexes play a vital role in mediating intracellular membrane trafficking (Tan and Gleeson
184 2019). Taken together, the identification of potential ESP provides further insight into the
185 origins of eukaryotic cellular complexity.

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

187 Balderarchaeota MAGs contain genes coding for complete glycolysis via
188 Embden-Meyerhof-Parnas (EMP) pathway, major steps of the tricarboxylic acid (TCA) cycle
189 and β-oxidation pathway, suggesting that members of Balderarchaeota may have potential to
190 metabolize organic compounds including carbohydrates and fatty acids (Fig. 3). The
191 ADP-dependent acetyl-CoA synthetase (ACD) for acetogenesis, which is widely found in
192 archaea (Lazar et al., 2016), was identified in Balderarchaeota MAGs. Meanwhile, phosphate
193 acetyltransferase (Pta) and acetate kinase (Ack) were also found in all Balderarchaeota MAGs
194 (Fig. 3, Supplementary Table 5). Although the *pta* gene was found in Sigynarchaeota as well,

195 all of their MAGs lack the *ack* gene. The Pta/Ack pathway for acetate production, which is
196 common in bacteria, was so far only found in Bathyarchaeota and the methanogenic genus
197 Methanosarcina in archaea (He et al., 2016, Rother and Metcalf 2004) and it is the first case
198 that genes coding for Pta and Ack were discovered in the Asgard archaea. The archaeal
199 *pta/ack* genes were considered HGT from bacteria donors. For example, the genes in
200 Methanosarcina were postulated to acquire from a cellulolytic Clostridia group (Fournier and
201 Gogarten 2008) whereas the *pta/ack* genes donor of Bathyarchaeota was still unclear, possibly
202 transferring independently from unknown clades of Bacteria (He et al., 2016). For
203 Balderarchaeota, the phylogenetic analysis of the *ack* gene sequences revealed that *ack* genes
204 of Balderarchaeota form a monophylogenetic branch (Supplementary Fig. 3), but the
205 phylogenetic tree of *pta* genes shows that the Balderarchaeota clade are within
206 Bathyarchaeota branch (Supplementary Fig. 4), indicating that *pta* genes in Balderarchaeota
207 and Bathyarchaeota probably evolved from the same bacterial donor. Taken together, the
208 Pta/Ack pathway in Balderarchaeota may have acquired from different bacterial donors by
209 two separate HGT events. Additionally, genes coding for the nitrite reductase (NADH) large
210 subunit (*nir*B) were detected in MAGs of Balderarchaeota, implying a potential nitrite
211 reduction capability.

212 The Wood Ljungdahl (WL) pathway enables organisms to reduce two molecules of CO_2
213 to form acetyl-CoA, and then to acetate to produce ATP in anaerobic conditions and Archaea
214 normally utilize tetrahydromethanopterin (THMPT) as C_1 carrier. Complete gene sets
215 encoding for the THMPT-WL pathway were found in Freyrarchaeota, suggesting the ability to
216 reduce CO_2 through autotrophic acetogenesis (Fig. 3). Tetrahydrofolate (THF) could be as C1

217 carrier as well, which was generally found in acetogenic bacteria. Sigynarchaeota MAGs
218 contain genes for both types of the WL pathway but 5,10-methylenetetrahydromethanopterin
219 reductase (Mer) which converts 5-methyltetrahydromethanopterin to
220 5,10-methylenetetrahydromethanopterin was missing in all MAGs identified here, suggesting
221 that Sigynarchaeota probably use the THF-WL pathway for acetate production (Fig. 3,
222 Supplementary Table 5). Sigynarchaeota contain not only all genes responsible for glycolysis
223 but also abundant genes coding for extracellular carbohydrate-degrading enzymes, including
224 α-amylase, cellulase, α-mannosidases and β-glucosidases (Supplementary Table 4), indicating
225 that archaea in Sigynarchaeota have the capacity to degrade complex carbohydrates.
226 Sigynarchaeota MAGs contain neither NADH dehydrogenase nor type 4 [NiFe] hydrogenase
227 (Supplementary Table 5), nevertheless, they probably use membrane-bound heterodisulfide
228 reductase (Hdr) to generate proton motive force as described in their sister-lineage
229 Lokiarchaeota (Spang et al., 2019a).

230 Among genes associated with glycolysis, the specific enzymes catalyzing these steps is
231 different in the three lineages. In the first step of glycolysis, for example, Sigynarchaeota use
232 ATP-dependent ROK (repressor, open reading frame, kinase) family enzymes while
233 Balderarchaeota utilizes ADP-dependent glucokinase. Likewise, Sigynarchaeota encode
234 ATP-dependent phosphofructokinase (PfkB) but fructose 6-phosphate (F6P) to fructose
235 1,6-bisphosphate (F1,6P) was catalyzed by ADP-dependent phosphofructokinase (ADP-PFK)
236 in Balderarchaeota.

237 We also compared differences of metabolic characteristic between the newly identified
238 Hermodarchaeota members and Odinarchaeota since both of them were recovered from

239 high-temperature environments. Hermodarchaeota encodes all enzymes for the complete
240 THMPT-WL pathway, while Odinarchaeota lack several key genes of THMPT-WL pathway.
241 Furthermore, only group 3 [NiFe]-hydrogenases were found in Hermodarchaeota genomes,
242 lacking group 4 [NiFe]- hydrogenases which were identified in Odinarchaeota (Spang et al.,
243 2019a). The presence of THMPT-WL pathway and group 3 [NiFe]-hydrogenases in
244 Hermodarchaeota indicate that they could grow lithoautotrophically by using H_2 as an
245 electron donor. DnaK-DnaJ-GrpE chaperone system, one of the characteristics of
246 hyperthermophilic archaea (Richter et al., 2010), were found in Hermodarchaeota and
247 Odinarchaeota but genes coding for reverse gyrase were absent in both of them.

248 In Tyrarchaeota MAGs, glycolysis and TCA pathways are not complete, however, the
249 presence of genes coding for THMPT-WL pathway and group 3 [NiFe] hydrogenases implies
250 its potential to harness energy from H_2 oxidation, possibly for lithoautotrophic growth,
251 depending on environmental conditions, as suggested for Lokiarchaeota and Thorarchaeota
252 (Spang et al., 2019b, Liu et al., 2018b). Moreover, *de novo* anaerobic cobalamin (vitamin B_{12})
253 biosynthesis pathway was found in Tyrarchaeota (Fig. 3, Supplementary Table 5), suggesting
254 that Tyrarchaeota harbor the potential of cobalamin synthesis. In nature, only limited
255 members of bacteria and archaea possess capacity of *de novo* cobalamin synthesis using one
256 of two alternative pathways: aerobic or anaerobic pathway (Fang et al., 2018). Within
257 Archaea, some members within Euryarchaeota, Thaumarchaeota, Crenarchaeota and
258 Bathyarchaeota have been reported possessing cobalamin synthesizing pathway (Doxey et al.,
259 2015, Pan et al., 2020), whereas only Tyrarchaeota seems to have this ability in the Asgard
260 archaea reported so far.

261 Njordarchaeota MAGs contain limited genes coding for major carbon metabolic
262 repertoire. Key genes coding for EMP pathway, TCA, the WL pathway and β-oxidation
263 pathway were missing (Fig. 3, Supplementary Table 5). But genes coding for amino acids
264 utilization were found in Njordarchaeota, including aminotransferases and
265 2-oxoacid:ferredoxin oxidoreductases (the former catalyzes the interconversion of amino
266 acids and 2-oxoacids and the latter oxidates 2-oxoacids to acyl-CoA), indicating that
267 Njordarchaeota has potential to metabolize amino acids. The amino acid carboxylate is
268 transferred to CO_2 and reducing ferredoxin during 2-oxoacid oxidation, which further could
269 be oxidized into formate and H_2 by formate dehydrogenase (Fdh) and [NiFe] hydrogenases,
270 respectively (Imachi et al., 2020). [NiFe]-hydrogenases were detected in Njordarchaeota but
271 lacking Fdh, indicating that Njordarchaeota might have a fermentative life style by produce
272 acetate or H_2 via degrading amino acids. Nevertheless, cultivation experiments are necessary
273 to verify all these predictions described here.

274 Several models for the origin of eukaryotes have been proposed, based on a metabolic
275 symbiosis between one archaeon and bacterial partner, which was fostered by the discovery of
276 natural syntrophy between *Candidatus* Prometheoarchaeum syntrophicum (that can degrade
277 amino acids to H_2 or formate) and Deltaproteobacteria (that can utilize H_2 or Formate and
278 provide amino acids or vitamin B_{12} to partner). Compared with other members of Asgard
279 archaea, Njordarchaeota possess limited pathway for carbon metabolism (Supplementary
280 Table 6), implying that they prone to grow in symbiosis with other organisms to adapt
281 complicated and volatile environment. Combining the phylogenetic affiliation with
282 eukaryotes and metabolic characteristic of Njordarchaeota, we speculate that the archaeal

283 ancestor of eukaryotes probably has potential to degrade amino acids to produce acetate or H_2
284 which can further benefit to the bacterial partner, although other additional lifestyles could not
285 be excluded.

286 **Conclusion**

287 Undoubtedly, the origin of eukaryotes is one of the most important evolutionary events. The
288 discovery of Asgard archaea has boosted the eocyte hypothesis that eukaryotes derive from
289 within archaea because the Asgard archaea possess two remarkable features: robust
290 evolutionary affinity with eukaryotes and various ESPs existence. In the present study, five
291 novel Asgard lineages were discovered based on phylogenetic analyses and AAI value
292 comparison, which significantly expand the phylogenetic and metabolic diversity of the
293 Asgard archaea. Our analyses support a 2-domain tree of life and demonstrate that the
294 eukaryotes lineage cluster with Njordarchaeota alone, suggesting that Njordarchaeota lineage
295 is the closest relatives to eukaryotes than Heimdallarchaeota. Metabolic characteristic of
296 Njordarchaeota shows different carbon metabolic pathways from Heimdallarchaeota that were
297 considered living in anoxic or oxic environment using various organic substrates such as
298 carbohydrates and fatty acids (Spang et al., 2019a, Bulzu et al., 2019, Cai et al., 2020),
299 whereas Njordarchaeota lack both complete glycolysis and WL pathway and the most
300 possibility of metabolism type is using amino acids in anoxic niches. This finding does not
301 contradict to the hypothesis by Spang *et al.* inferring metabolic feature of the archaeal
302 ancestor of eukaryotes (Spang et al., 2019a) that it used organic substrates to produce acetate,
303 formate, H_2, which might be beneficial for symbiosis. In general, the characterization of
304 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_236" consisting of 2.0 Mbps in 104 contigs with an estimated completeness of 91.59%, an estimated contamination of 6.54% 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"

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

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

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

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

345 ***Candidatus* Njordarchaeum guaymasis** (guayma'sis N.L. neut. adj. pertaining to Guaymas
346 Basin, located in the Gulf of California, México). This uncultured lineage is represented by
347 the genome "GB_154" consisting of 2.1 Mbps in 191 contigs with an estimated completeness
348 of 87.38%, an estimated contamination of 4.67% and 20 tRNAs The MAG recovered from
349 Guaymas Basin sediment.

350 ***Candidatus* Tyrarchaeaceae** (Tyr.ar.chae.ace'ae. N.L. neut. n. *Tyrarchaeum*, Candidatus
351 generic name; -aceae ending to denote the family; N.L. fem. pl. n. *Tyrarchaeaceae*, the
352 *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.

374 ***Candidatus* Freyrarchaeaceae** (Freyr.ar.chae.ace'ae. N.L. neut. n. *Freyrarchaeum*,

375 Candidatus generic name; -aceae ending to denote the family; N.L. fem. pl. n.

376 *Freyrarchaeaceae*, the *Freyrarchaeum* family).

377 The family is described based on 37 concatenated conserved marker genes phylogeny. The

378 description is the same as that of its sole genus and species. Type genus is *Candidatus*

379 Freyrarchaeum.

380 ***Candidatus* Freyrarchaeales** (Freyr.ar.chae.a'les. N.L. neut. n. *Freyrarchaeum*, Candidatus

381 generic name; -ales ending to denote the order; N.L. fem. pl. n. *Freyrarchaeales*, the

382 *Freyrarchaeum* order).

383 The order is described based on 37 concatenated conserved marker genes phylogeny. The

384 description is the same as that of its sole genus and species. Type genus is *Candidatus*

385 Freyrarchaeum.

386 ***Candidatus* Freyrarchaeia** (Freyr.ar.chae'i.a. N.L. neut. n. *Freyrarchaeum*, Candidatus

387 generic name; -ia ending to denote the class; N.L. fem. pl. n. *Freyrarchaeia*, the

388 *Freyrarchaeum* class).

389 The class is described based on 37 concatenated conserved marker genes phylogeny. The

390 description is the same as that of its sole genus and species. Type genus is *Candidatus*

391 Freyrarchaeum.

392 ***Candidatus* Freyrarchaeota** (Freyr.ar.chae.o'ta. N.L. neut. n. *Freyrarchaeum*, Candidatus

393 generic name; -ota ending to denote the phylum; N.L. fem. pl. n. *Freyrarchaeota*, the

394 *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).

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* Balderarchaeum.

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

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

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

474 **Methods**

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

481 **Data collection.** Asgard archaea are distributed mainly in different environmental sediments,
482 including estuary sediments (Seitz et al., 2016), mangrove sediments (Liu et al., 2018a),
483 hydrothermal sediments (Dombrowski et al., 2018), hot spring sediments
484 (Zaremba-Niedzwiedzka et al., 2017a), marine sediments (Spang et al., 2015), and freshwater
485 sediments (Narrowe et al., 2018). According to the environmental distributions of Asgard
486 archaea, metagenomic data were collected and downloaded from the SRA database
487 (https://www.ncbi.nlm.nih.gov/sra/).

488 **Metagenomic assembly and genomic binning.** The raw reads were trimmed using
489 Trimmomatic (v.0.38) (Bolger et al., 2014) to remove adapters and low-quality reads. After
490 trimming, the reads of each sample were *de novo* assembled using Megahit (v.1.2.5) (Li et al.,
491 2015) with a k-step of 6. Samples from the same location or similar environments were
492 assembled together. Contigs were binned separately using MetaBAT (v.2.12.1) (Kang et al.,
493 2019), MaxBin (v.2.2.7) (Wu et al., 2016), and Concoct (v.1.1.0) (Alneberg et al., 2014) with
494 the default parameters, and the initial taxonomic classification of each MAGs was performed
495 using GTDB-Tk (v.1.2.0) (Chaumeil et al., 2019) to extract Asgard MAGs. The completeness

496 and contamination of Asgard MAGs were evaluated with the CheckM lineage_wf workflow
497 (v.1.0.12) (Parks et al., 2015). Finally, Asgard MAGs with completeness above 50% and
498 contamination below 10% were selected for further analyses, and Prodigal (v.2.6.1) (Hyatt et
499 al., 2010) was used to predict protein-coding genes for these selected Asgard MAGs.

500 **Phylogenetic analyses of Asgard MAGs.** To determine the exact phylogenetic affiliations in
501 the Asgard superphylum, 37 conserved marker genes were selected as described in the
502 literature (Wang et al., 2019, Jay et al., 2018). Homologs of the 37 conserved marker proteins
503 were identified using Diamond (v.2.0.4) (Buchfink et al., 2015). Each dataset of marker
504 proteins was aligned with MAFFT-L-INS-i (v.7.313) (Katoh and Standley 2013) and trimmed
505 by trimAl (v.1.4.22) (Capella-Gutiérrez et al., 2009) with the "automated1" option.
506 Maximum-likelihood phylogenies for the 37 conserved marker proteins was built using
507 IQ-TREE (v.2.0.5) (Nguyen et al., 2015) under the model "LG+F+G4+C60". The support
508 values were calculated using 1000 ultrafast bootstraps.

509 **Phylogenetic tree of life.** To confirm the phylogenetic affiliations of eukaryotes and the novel
510 Asgard lineages, 21 taxonomic marker genes shared among three domains selected by
511 Williams (Williams et al., 2020) were used for phylogenetic analyses. Single-gene trees were
512 inferred for all the markers using IQ-TREE with the LG+G4+F model to exclude
513 phylogenetic artefacts such as long branch attraction, HGT (eukaryotic genes falling into
514 bacterial clade or scattering in archaeal clade were considered as HGT) and a BLASTp
515 inspection was further performed to identify all eukaryotic genes that originated from the
516 nuclear genome and to remove genes of mitochondrial or chloroplastic origin. The
517 maximum-likelihood tree was built with IQ-TREE under the LG+C60+F+G4 model with

518 1000 ultrafast bootstraps.

519 **Identification of ESPs.** All predicted proteins encoded by the MAGs of the five novel Asgard
520 lineages and Hermodarchaeota were analyzed using InterProScan (Jones et al., 2014)
521 (v.5.47-82.0) with default parameters to annotate protein domains and were assigned to
522 archaeal clusters of orthologous genes (arCOGs) (Makarova et al., 2015) by eggnog-mapper
523 (v. 2.0.1b) (Huerta-Cepas et al., 2019) with default settings.. The lists of InterPro accession
524 numbers (IPRs) and arCOG identifiers previously published by Zaremba-Niedzwiedzka *et al.*
525 (Zaremba-Niedzwiedzka et al., 2017a) and Bulzu *et al.* (Bulzu et al., 2019) were used to
526 identify potential ESPs. Some key words related to eukaryote-specific processes or cell
527 structures were used to search annotation information of interProScan to identify potential
528 ESPs previously not reported in Asgard archaea. Several candidate ESPs were further
529 examined using HHpred (Zimmermann et al., 2018) with default parameters.

530 **Metabolic reconstruction.** The proteome of each MAG reported in the present study was
531 uploaded to the KEGG Automatic Annotation Server (KAAS) (Moriya et al., 2007) and run
532 with several settings: the GHOSTX, Prokaryotes and Bidirectional Best Hit (BBH) settings.
533 Additionally, proteins were queried against the nonredundant (NR) protein database
534 (downloaded from NCBI on February 2020) using the Diamond (v.2.0.4) BLASTp search
535 (e-value cutoff <1e−5). Metabolic pathways were reconstructed based on combination of the
536 NR annotations, protein domain information and KEGG Ontology (KO) numbers.

537 The dbCAN2 (Zhang et al., 2018) web server was used to identify carbohydrate-degrading
538 enzymes with the default settings. The putative large subunits of [NiFe] hydrogenases were
539 identified by querying against a local database based on HydDB (Søndergaard et al., 2016)

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) (Rawlings et al., 2016) searched for peptidases by Diamond (v.2.0.4) with an E-value cutoff of 1×10^{-20}, and PSORT (v.3.0.2) was used to identify protein localization (Horton et al., 2007).

Calculation of ANI and average AAI. The ANI and AAI values were calculated using OrthoANI (v.1.2) (Lee et al., 2016) and CompareM (https://github.com/dparks1134/CompareM), respectively, with the default parameters.

Acknowledgments

We are grateful for Dr. Tom A. William for his suggestions regarding phylogenetic analysis. We thank Brett Baker for allowing us use their metagenomic data freely. These sequence data were produced by the US Department of Energy Joint Genome Institute http://www.jgi.doe.gov/ in collaboration with the user community.

Data availability

The genomes of Asgard archaea generated in this study have been made available at the eLibrary of Microbial Systematics and Genomics (eLMSG; https://www.biosino.org/elmsg/index) under accession numbers LMSG_G000000610.1 - LMSG_G000000611.1 for Njordarchaeota, LMSG_G000000612.1 - LMSG_G000000613.1 for Tyrarchaeota, LMSG_G000000614.1 - LMSG_G000000616.1 and LMSG_G000000628.1 for Sigynarchaeota, LMSG_G000000617.1 - LMSG_G000000618.1 for Hermodarchaeota, LMSG_G000000622.1 for Freyrarchaeota, LMSG_G000000625.1 - LMSG_G000000627.1

for Sigynarchaeota.

The initial phylogenetic trees have been deposited at figshare and can be accessed at the following link:

https://figshare.com/s/c20d9eccb7e4591b429c.

Funding

This work was supported by the Natural Science Foundation of China (Grant No. 91751205, 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.

575 **References**

576 Embley T M, Martin W. 2006. Eukaryotic evolution, changes and challenges. *Nature*, **440**, 623-630.

577 López-García P, Moreira D. 2015. Open questions on the origin of eukaryotes. *Trends in ecology &*
578 *evolution*, **30**, 697-708.

579 Rochette N C, Brochier-Armanet C, Gouy M. 2014. Phylogenomic test of the hypotheses for the
580 evolutionary origin of eukaryotes. *Molecular biology and evolution*, **31**, 832-845.

581 Lopez-Garcia P, Moreira D. 2006. Selective forces for the origin of the eukaryotic nucleus. *Bioessays*,
582 **28**, 525-533. doi,10.1002/bies.20413.

583 López-García P, Moreira D. 1999. Metabolic symbiosis at the origin of eukaryotes. *Trends in*
584 *biochemical sciences*, **24**, 88-93.

585 Martin W F, Garg S, Zimorski V. 2015. Endosymbiotic theories for eukaryote origin. *Philosophical*
586 *Transactions of the Royal Society B: Biological Sciences*, **370**, 20140330.

587 Esser C, Ahmadinejad N, Wiegand C, Rotte C, Sebastiani F, Gelius-Dietrich G, Henze K, Kretschmann
588 E, Richly E, Leister D. 2004. A genome phylogeny for mitochondria among α-proteobacteria
589 and a predominantly eubacterial ancestry of yeast nuclear genes. *Molecular biology and*
590 *evolution*, **21**, 1643-1660.

591 Moreira D, López-García P. 1998. Symbiosis between methanogenic archaea and δ-proteobacteria as
592 the origin of eukaryotes: the syntrophic hypothesis. *Journal of Molecular Evolution*, **47**,
593 517-530.

594 Spang A, Saw J H, Jorgensen S L, Zaremba-Niedzwiedzka K, Martijn J, Lind A E, van Eijk R, Schleper
595 C, Guy L, Ettema T J G. 2015. Complex archaea that bridge the gap between prokaryotes and
596 eukaryotes. *Nature*, **521**, 173-179. doi,10.1038/nature14447.

597 Zaremba-Niedzwiedzka K, Caceres E F, Saw J H, Bäckström D, Juzokaite L, Vancaester E, Seitz K W,
598 Anantharaman K, Starnawski P, Kjeldsen K U. 2017a. Asgard archaea illuminate the origin of
599 eukaryotic cellular complexity. *Nature*, **541**, 353.

600 Spang A, Stairs C W, Dombrowski N, Eme L, Lombard J, Caceres E F, Greening C, Baker B J, Ettema
601 T J. 2019a. Proposal of the reverse flow model for the origin of the eukaryotic cell based on
602 comparative analyses of Asgard archaeal metabolism. *Nature microbiology*, 1.

603 Williams T A, Foster P G, Cox C J, Embley T M. 2013. An archaeal origin of eukaryotes supports only
604 two primary domains of life. *Nature*, **504**, 231-236.

605 Woese C R, Kandler O, Wheelis M L. 1990. Towards a natural system of organisms: proposal for the
606 domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences*,
607 **87**, 4576-4579.
608 Williams T A, Cox C J, Foster P G, Szollosi G J, Embley T M. 2020. Phylogenomics provides robust
609 support for a two-domains tree of life. *Nat Ecol Evol*, **4**, 138-147.
610 doi,10.1038/s41559-019-1040-x.
611 Hartman H, Fedorov A. 2002. The origin of the eukaryotic cell: a genomic investigation. *Proceedings*
612 *of the National Academy of Sciences*, **99**, 1420-1425.
613 Bulzu P-A, Andrei A-Ş, Salcher M M, Mehrshad M, Inoue K, Kandori H, Beja O, Ghai R, Banciu H L.
614 2019. Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche. *Nature*
615 *microbiology*, **4**, 1129-1137.
616 Liu Y, Zhou Z, Pan J, Baker B J, Gu J-D, Li M. 2018a. Comparative genomic inference suggests
617 mixotrophic lifestyle for Thorarchaeota. *The ISME journal*, **12**, 1021-1031.
618 Seitz K W, Lazar C S, Hinrichs K-U, Teske A P, Baker B J. 2016. Genomic reconstruction of a novel,
619 deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur
620 reduction. *The ISME journal*, **10**, 1696-1705.
621 Wong H L, White R A, Visscher P T, Charlesworth J C, Vázquez-Campos X, Burns B P. 2018.
622 Disentangling the drivers of functional complexity at the metagenomic level in Shark Bay
623 microbial mat microbiomes. *The ISME journal*, **12**, 2619-2639.
624 MacLeod F, Kindler G S, Wong H L, Chen R, Burns B P. 2019. Asgard archaea: diversity, function, and
625 evolutionary implications in a range of microbiomes. *AIMS microbiology*, **5**, 48.
626 Imachi H, Nobu M K, Nakahara N, Morono Y, Ogawara M, Takaki Y, Takano Y, Uematsu K, Ikuta T,
627 Ito M, Matsui Y, Miyazaki M, Murata K, Saito Y, Sakai S, Song C, Tasumi E, Yamanaka Y,
628 Yamaguchi T, Kamagata Y, Tamaki H, Takai K. 2020. Isolation of an archaeon at the
629 prokaryote-eukaryote interface. *Nature*, **577**, 519-525. doi,10.1038/s41586-019-1916-6.
630 Akıl C, Tran L T, Orhant-Prioux M, Baskaran Y, Manser E, Blanchoin L, Robinson R C. 2020. Insights
631 into the evolution of regulated actin dynamics via characterization of primitive gelsolin/cofilin
632 proteins from Asgard archaea. *Proceedings of the National Academy of Sciences*.
633 Zhang R-Y, Zou B, Yan Y-W, Jeon C O, Li M, Cai M, Quan Z-X. 2020. Design of targeted primers
634 based on 16S rRNA sequences in meta-transcriptomic datasets and identification of a novel

635 taxonomic group in the Asgard archaea. *BMC microbiology*, **20**, 25.

636 Wang Y, Wegener G, Hou J, Wang F, Xiao X. 2019. Expanding anaerobic alkane metabolism in the
637 domain of Archaea. *Nat Microbiol*, **4**, 595-602. doi,10.1038/s41564-019-0364-2.

638 Zaremba-Niedzwiedzka K, Caceres E F, Saw J H, Backstrom D, Juzokaite L, Vancaester E, Seitz K W,
639 Anantharaman K, Starnawski P, Kjeldsen K U, Stott M B, Nunoura T, Banfield J F, Schramm
640 A, Baker B J, Spang A, Ettema T J. 2017b. Asgard archaea illuminate the origin of eukaryotic
641 cellular complexity. *Nature*, **541**, 353-358. doi,10.1038/nature21031.

642 Zhang J W, Dong H P, Hou L J, Liu Y, Ou Y F, Zheng Y L, Han P, Liang X, Yin G Y, Wu D M, Liu M,
643 Li M. 2021. Newly discovered Asgard archaea Hermodarchaeota potentially degrade alkanes
644 and aromatics via alkyl/benzyl-succinate synthase and benzoyl-CoA pathway. *ISME J*.
645 doi,10.1038/s41396-020-00890-x.

646 Luo C, Rodriguez-r L M, Konstantinidis K T. 2014. MyTaxa: an advanced taxonomic classifier for
647 genomic and metagenomic sequences. *Nucleic acids research*, **42**, e73-e73.

648 Seitz K W, Dombrowski N, Eme L, Spang A, Lombard J, Sieber J R, Teske A P, Ettema T J G, Baker B
649 J. 2019. Asgard archaea capable of anaerobic hydrocarbon cycling. *Nat Commun*, **10**, 1822.
650 doi,10.1038/s41467-019-09364-x.

651 Raiborg C, Stenmark H. 2009. The ESCRT machinery in endosomal sorting of ubiquitylated membrane
652 proteins. *Nature*, **458**, 445-452.

653 Grau-Bové X, Sebé-Pedrós A, Ruiz-Trillo I. 2015. The Eukaryotic Ancestor Had a Complex Ubiquitin
654 Signaling System of Archaeal Origin. *Molecular Biology and Evolution*, **32**, 726-739.
655 doi,10.1093/molbev/msu334.

656 Leung K F, Dacks J B, Field M C. 2008. Evolution of the multivesicular body ESCRT machinery;
657 retention across the eukaryotic lineage. *Traffic*, **9**, 1698-1716.
658 doi,10.1111/j.1600-0854.2008.00797.x.

659 Field M C, Dacks J B. 2009. First and last ancestors: reconstructing evolution of the endomembrane
660 system with ESCRTs, vesicle coat proteins, and nuclear pore complexes. *Curr Opin Cell Biol*,
661 **21**, 4-13. doi,10.1016/j.ceb.2008.12.004.

662 Schulz B L, Stirnimann C U, Grimshaw J P, Brozzo M S, Fritsch F, Mohorko E, Capitani G,
663 Glockshuber R, Grutter M G, Aebi M. 2009. Oxidoreductase activity of
664 oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation

efficiency. *Proc Natl Acad Sci U S A*, **106**, 11061-11066. doi,10.1073/pnas.0812515106.

Park S Y, Guo X. 2014. Adaptor protein complexes and intracellular transport. *Bioscience reports*, **34**.

Hirst J, Barlow L D, Francisco G C, Sahlender D A, Seaman M N, Dacks J B, Robinson M S. 2011. The fifth adaptor protein complex. *PLoS Biol*, **9**, e1001170.

Tan J Z A, Gleeson P A. 2019. Cargo sorting at the trans-Golgi network for shunting into specific transport routes: role of Arf small G proteins and adaptor complexes. *Cells*, **8**, 531.

Lazar C S, Baker B J, Seitz K, Hyde A S, Dick G J, Hinrichs K U, Teske A P. 2016. Genomic evidence for distinct carbon substrate preferences and ecological niches of B athyarchaeota in estuarine sediments. *Environmental microbiology*, **18**, 1200-1211.

He Y, Li M, Perumal V, Feng X, Fang J, Xie J, Sievert S, Wang F. 2016. Genomic and enzymatic evidence for acetogenesis among multiple lineages of the archaeal phylum Bathyarchaeota widespread in marine sediments. *Nature microbiology*, **1**, 1-9.

Rother M, Metcalf W W. 2004. Anaerobic growth of Methanosarcina acetivorans C2A on carbon monoxide: an unusual way of life for a methanogenic archaeon. *Proceedings of the National Academy of Sciences*, **101**, 16929-16934.

Fournier G P, Gogarten J P. 2008. Evolution of acetoclastic methanogenesis in Methanosarcina via horizontal gene transfer from cellulolytic Clostridia. *Journal of bacteriology*, **190**, 1124-1127.

Sousa F L, Martin W F. 2014. Biochemical fossils of the ancient transition from geoenergetics to bioenergetics in prokaryotic one carbon compound metabolism. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, **1837**, 964-981.

Ma K, Schicho R N, Kelly R M, Adams M W. 1993. Hydrogenase of the hyperthermophile Pyrococcus furiosus is an elemental sulfur reductase or sulfhydrogenase: evidence for a sulfur-reducing hydrogenase ancestor. *Proc Natl Acad Sci U S A*, **90**, 5341-5344. doi,10.1073/pnas.90.11.5341.

Richter K, Haslbeck M, Buchner J. 2010. The heat shock response: life on the verge of death. *Molecular Cell*, **40**, 253-266.

Spang A, Stairs C W, Dombrowski N, Eme L, Lombard J, Caceres E F, Greening C, Baker B J, Ettema T J G. 2019b. Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. *Nat Microbiol*, **4**, 1138-1148. doi,10.1038/s41564-019-0406-9.

695 Liu Y, Zhou Z, Pan J, Baker B J, Gu J D, Li M. 2018b. Comparative genomic inference suggests
696 mixotrophic lifestyle for Thorarchaeota. *ISME J*, **12**, 1021-1031.
697 doi,10.1038/s41396-018-0060-x.
698 Fang H, Li D, Kang J, Jiang P, Sun J, Zhang D. 2018. Metabolic engineering of Escherichia coli for de
699 novo biosynthesis of vitamin B 12. *Nature communications*, **9**, 1-12.
700 Doxey A C, Kurtz D A, Lynch M D, Sauder L A, Neufeld J D. 2015. Aquatic metagenomes implicate
701 Thaumarchaeota in global cobalamin production. *The ISME journal*, **9**, 461-471.
702 Pan J, Zhou Z, Béjà O, Cai M, Yang Y, Liu Y, Gu J-D, Li M. 2020. Genomic and transcriptomic
703 evidence of light-sensing, porphyrin biosynthesis, Calvin-Benson-Bassham cycle, and urea
704 production in Bathyarchaeota. *Microbiome*, **8**, 1-12.
705 Cai M, Liu Y, Yin X, Zhou Z, Friedrich M W, Richter-Heitmann T, Nimzyk R, Kulkarni A, Wang X, Li
706 W. 2020. Diverse Asgard archaea including the novel phylum Gerdarchaeota participate in
707 organic matter degradation. *Science China Life Sciences*, 1-12.
708 Feng X, Wang Y, Zubin R, Wang F. 2019. Core metabolic features and hot origin of Bathyarchaeota.
709 *Engineering*, **5**, 498-504.
710 Dombrowski N, Teske A P, Baker B J. 2018. Expansive microbial metabolic versatility and biodiversity
711 in dynamic Guaymas Basin hydrothermal sediments. *Nature communications*, **9**, 1-13.
712 Narrowe A B, Spang A, Stairs C W, Caceres E F, Baker B J, Miller C S, Ettema T J. 2018. Complex
713 evolutionary history of translation Elongation Factor 2 and diphthamide biosynthesis in
714 Archaea and parabasalids. *Genome biology and evolution*, **10**, 2380-2393.
715 Bolger A M, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data.
716 *Bioinformatics*, **30**, 2114-2120.
717 Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for
718 large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, **31**,
719 1674-1676.
720 Kang D D, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning
721 algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*,
722 **7**, e7359.
723 Wu Y-W, Simmons B A, Singer S W. 2016. MaxBin 2.0: an automated binning algorithm to recover
724 genomes from multiple metagenomic datasets. *Bioinformatics*, **32**, 605-607.

725 Alneberg J, Bjarnason B S, De Bruijn I, Schirmer M, Quick J, Ijaz U Z, Lahti L, Loman N J, Andersson
726 A F, Quince C. 2014. Binning metagenomic contigs by coverage and composition. *Nature*
727 *methods*, **11**, 1144-1146.
728 Chaumeil P A, Mussig A J, Hugenholtz P, Parks D H. 2019. GTDB-Tk: a toolkit to classify genomes
729 with the Genome Taxonomy Database. *Bioinformatics*. doi,10.1093/bioinformatics/btz848.
730 Parks D H, Imelfort M, Skennerton C T, Hugenholtz P, Tyson G W. 2015. CheckM: assessing the
731 quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome*
732 *research*, **25**, 1043-1055.
733 Hyatt D, Chen G-L, LoCascio P F, Land M L, Larimer F W, Hauser L J. 2010. Prodigal: prokaryotic
734 gene recognition and translation initiation site identification. *BMC bioinformatics*, **11**, 119.
735 Jay Z J, Beam J P, Dlakić M, Rusch D B, Kozubal M A, Inskeep W P. 2018. Marsarchaeota are an
736 aerobic archaeal lineage abundant in geothermal iron oxide microbial mats. *Nature*
737 *microbiology*, **3**, 732-740.
738 Buchfink B, Xie C, Huson D H. 2015. Fast and sensitive protein alignment using DIAMOND. *Nature*
739 *methods*, **12**, 59-60.
740 Katoh K, Standley D M. 2013. MAFFT multiple sequence alignment software version 7: improvements
741 in performance and usability. *Molecular biology and evolution*, **30**, 772-780.
742 Capella-Gutiérrez S, Silla-Martínez J M, Gabaldón T. 2009. trimAl: a tool for automated alignment
743 trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972-1973.
744 Nguyen L-T, Schmidt H A, Von Haeseler A, Minh B Q. 2015. IQ-TREE: a fast and effective stochastic
745 algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*,
746 **32**, 268-274.
747 Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka
748 G. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics*, **30**,
749 1236-1240.
750 Makarova K S, Wolf Y I, Koonin E V. 2015. Archaeal clusters of orthologous genes (arCOGs): an
751 update and application for analysis of shared features between Thermococcales,
752 Methanococcales, and Methanobacteriales. *Life*, **5**, 818-840.
753 Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund S K, Cook H, Mende D R,
754 Letunic I, Rattei T, Jensen L J. 2019. eggNOG 5.0: a hierarchical, functionally and

755 phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses.
756 *Nucleic acids research*, **47**, D309-D314.
757 Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas A N,
758 Alva V. 2018. A completely reimplemented MPI bioinformatics toolkit with a new HHpred
759 server at its core. *Journal of molecular biology*, **430**, 2237-2243.
760 Moriya Y, Itoh M, Okuda S, Yoshizawa A C, Kanehisa M. 2007. KAAS: an automatic genome
761 annotation and pathway reconstruction server. *Nucleic acids research*, **35**, W182-W185.
762 Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk P K, Xu Y, Yin Y. 2018. dbCAN2: a meta
763 server for automated carbohydrate-active enzyme annotation. *Nucleic acids research*, **46**,
764 W95-W101.
765 Søndergaard D, Pedersen C N, Greening C. 2016. HydDB: a web tool for hydrogenase classification
766 and analysis. *Scientific reports*, **6**, 1-8.
767 Rawlings N D, Barrett A J, Finn R. 2016. Twenty years of the MEROPS database of proteolytic
768 enzymes, their substrates and inhibitors. *Nucleic acids research*, **44**, D343-D350.
769 Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier C, Nakai K. 2007. WoLF PSORT:
770 protein localization predictor. *Nucleic acids research*, **35**, W585-W587.
771 Lee I, Kim Y O, Park S-C, Chun J. 2016. OrthoANI: an improved algorithm and software for
772 calculating average nucleotide identity. *International journal of systematic and evolutionary*
773 *microbiology*, **66**, 1100-1103.
774

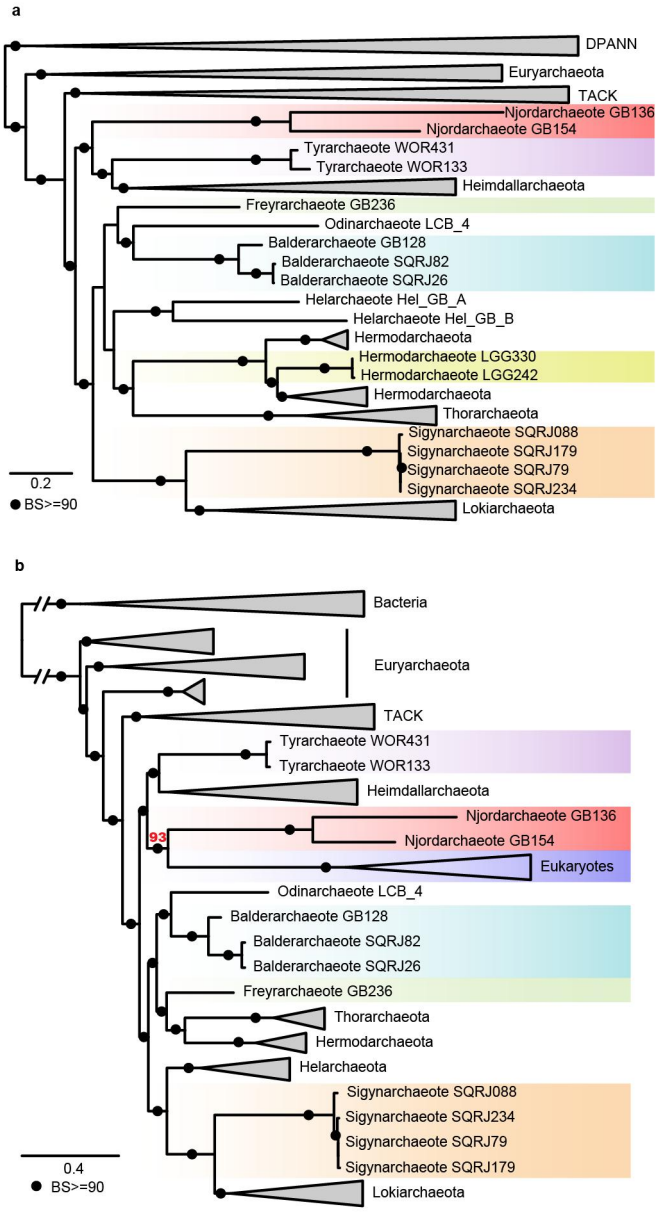

Figure 1. Phylogenetic analyses of Asgard archaea. a, Maximum-likelihood tree of 37 concatenated marker proteins inferred with the LG+F+C60+G4 model in IQ-TREE rooted in DPANN, 19 representatives of DPANN, 55 representatives of Euryarchaeota, 29 representatives of TACK and 68 genomes (including five new lineages) of Asgard were used to infer phylogenetic tree. **b**, Maximum likelihood inference of 21 concatenated conserved protein sequences under the LG+F+C60+G4 model rooted in bacteria, 85 archaeal genomes (53 within Asgard), 19 eukaryotic genomes, and 36 bacterial genomes were used to infer phylogenetic tree. The bootstrap support values above 90 were shown with black filled circles.

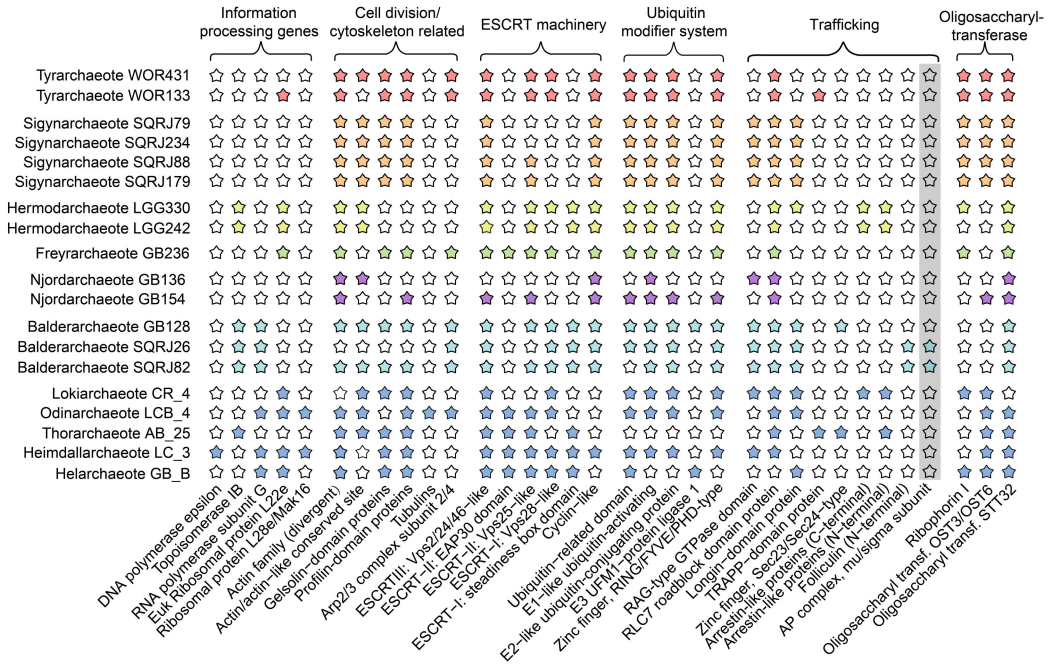

Figure 2. Comparison of the distributions of ESPs in the five new Asgard lineages and other representative Asgard clades. Colored stars indicate the presence of ESPs, whereas empty stars indicate the absence of ESPs. The grey box highlights ESP identified in this study. The new ESP, mu/sigma subunit of AP complex, was detected in Balderarchaeote SQRJ26 and Balderarchaeote SQRJ82.

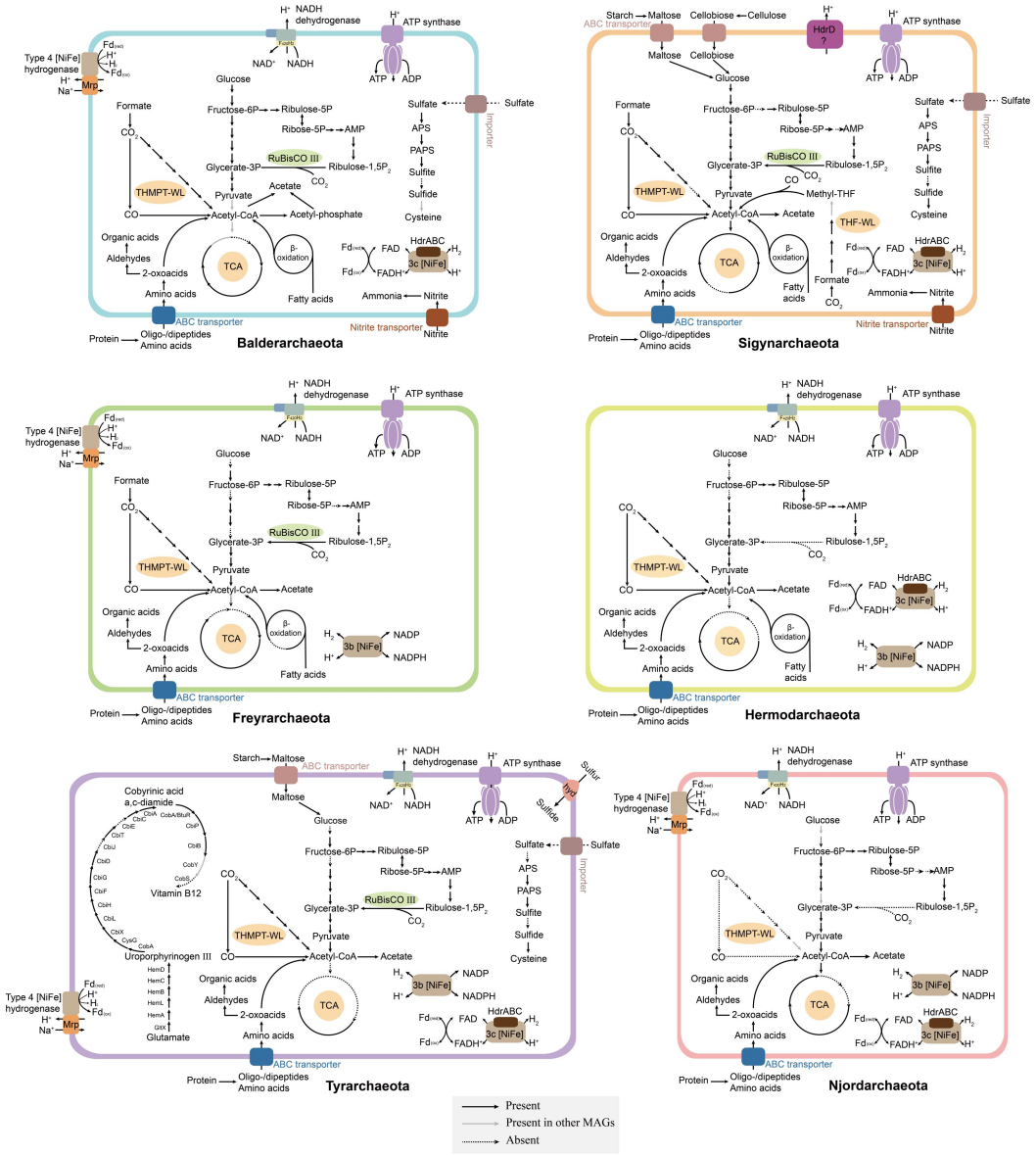

Figure 3. Inferred metabolic pathways of the five new Asgard lineages and the new clade of Hermodarchaeota based on genes identified using the KEGG database and the NCBI NR protein database. A black line indicates that a component/process is present in representative MAGs, a grey line indicates that 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, GB236; Njordarchaeota, GB154; Sigynarchaeota, SQRJ79; and Tyrarchaeota, WOR431; Hermodarchaeota, LGG330. Details about the genes are provided in Supplementary Table 5. 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.