Spang, A.; Stairs, C.W.; Dombrowski, N.; Eme, E.; Lombard, J.; Caceres, E.F.; Greening, C.; Baker, B.J. & Ettema, T.J.G. (2019). Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. *Nature Microbiology*, 4, 1138–1148

Article to *Nature Microbiology*

Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analysis of Asgard archaeal metabolism

Anja Spang[1,2,*], Courtney W. Stairs[1], Nina Dombrowski[2,3], Laura Eme[1], Jonathan Lombard[1], Eva Fernández Cáceres[1], Chris Greening[4], Brett J. Baker[3] and Thijs J.G. Ettema[1,5*]

[1] Department of Cell- and Molecular Biology, Science for Life Laboratory, Uppsala University, SE-75123, Uppsala, Sweden

[2] NIOZ, Royal Netherlands Institute for Sea Research, Department of Marine Microbiology and Biogeochemistry, and Utrecht University, P.O. Box 59, NL-1790 AB Den Burg, The Netherlands

[3] Department of Marine Science, University of Texas at Austin, Marine Science Institute, Port Aransas, TX-78373, USA

[4] School of Biological Sciences, Monash University, Clayton, Victoria, Australia

[5]Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, Stippeneng 4, 6708WE Wageningen, The Netherlands

* corresponding authors: anja.spang@nioz.nl, thijs.ettema@icm.uu.se

Key words: Archaea, Asgard superphylum, eukaryogenesis, endosymbiosis, syntrophy, evolution, metabolism

Abstract

The origin of eukaryotes represents an unresolved puzzle in evolutionary biology. Current research suggests that eukaryotes evolved from a merger between a host of archaeal descent and an alphaproteobacterial endosymbiont. The discovery of the Asgard archaea, a proposed archaeal superphylum that includes Loki-, Thor-, Odin- and Heimdallarchaeota suggested to comprise the closest archaeal relatives of eukaryotes, has helped elucidating the identity of the putative archaeal host. While Lokiarchaeota are assumed to employ a hydrogen-dependent metabolism, little is known about the metabolic potential of other members of the Asgard superphylum. We infer the central metabolic pathways of Asgard archaea using comparative genomics and phylogenetics to be able to refine current models for the origin of eukaryotes. Our analyses indicate that Thor- and Lokiarchaeota encode proteins necessary for carbon fixation via the Wood-Ljungdahl pathway and for obtaining reducing equivalents from organic substrates. In contrast, *Heimdallarchaeum* LC2 and LC3 genomes encode enzymes potentially enabling the oxidation of organic substrates using nitrate or oxygen as electron acceptors. The gene repertoire of *Heimdallarchaeum* AB125 and *Odinarchaeum* indicates that these organisms can ferment organic substrates and conserve energy by coupling ferredoxin re-oxidation to respiratory proton reduction. Altogether, our genome analyses suggest that Asgard representatives are primarily organoheterotrophs with variable capacity for hydrogen consumption and production. On this basis, we propose the 'reverse flow model', an updated symbiogenetic model for the origin of eukaryotes that involves electron or hydrogen flow from an organoheterotrophic archaeal host to a bacterial symbiont.

Main

The exploration of microbial life inhabiting anoxic environments has changed our perception of microbial diversity, evolution, and ecology and unveiled various previously unknown branches in the tree of life[1], including the proposed Asgard archaea superphylum[2-4]. The analysis of genomes from uncultivated microorganisms has improved our understanding of microbial metabolic diversity and evolution of life on Earth, including the origin of eukaryotes[5]. During the past years, models that posit that eukaryotes evolved from a symbiosis between an alphaproteobacterial endosymbiont and an archaeal host cell have gained increasing support (reviewed in[6-8]). In particular, detailed phylogenomic analyses of Asgard archaea, which currently include Loki-, Thor-, Odin- and Heimdallarchaeota, have suggested that these archaea represent the closest known relatives of eukaryotes[2,4]. Though this view has been challenged[9], additional analyses supported the phylogenetic placement of eukaryotes within the Asgard archaea and reinforced the suggestion that these organisms are key for our understanding of eukaryogenesis[10,11]. The genomes of all members of the Asgard archaea are enriched in genes coding for so-called eukaryotic signature proteins (ESPs)[2,4,12], indicating that the elusive archaeal ancestor of eukaryotes already contained several building blocks for the subsequent evolution of eukaryotic complexity[2,4]. In turn, genome analyses of Lokiarchaeum have reinvigorated discussions about the nature of the archaeal ancestor of eukaryotes and about the evolutionary events that led to the origin of the eukaryotic cell[13-16]. Even though recent data favors hypotheses which underpin a symbiogenetic origin of eukaryotes from only two domains of life[6,8,17,18], the timing of the events leading to eukaryogenesis remain unknown[19].

Here, we perform a comparative analysis of the metabolic potential of the Asgard superphylum, which revealed considerable metabolic versatility in these archaea. In light of our findings, we update previously formulated symbiogenetic hypotheses and present a refined model for the origin of eukaryotes, in which the archaeal host is suggested to be a fermentative organoheterotroph generating reduced compounds that are metabolized by a syntrophic bacterial partner organism.

Results and discussion

Genomic analyses of Asgard archaea reveals different metabolic repertoires

Our reconstruction of the metabolism of the Asgard archaea, based on comparative genome annotation and phylogenetic analyses, revealed that the enzymatic repertoire differs both within and between the metagenome assembled genomes (MAGs) of representatives of the Loki-, Thor-, Odin- and Heimdallarchaeota[4], suggesting that members of these groups are characterized by different physiological lifestyles.

Loki- and Thorarchaeota can likely use organic compounds and hydrogen

The Wood-Ljungdahl pathway (WLP) allows the reduction of carbon dioxide to acetyl-CoA and can be used to support both autotrophic carbon fixation and, when linked to chemiosmotic processes, energy conservation[20-22]. Both Loki- and Thorarchaeota encode all enzymes for a complete (archaeal) WLP (Figure 1, Suppl. Figure 2a, Suppl. Tables 1 and 2, Suppl. Information)[3,13,23]. The presence of the WLP, together with group 3b and group 3c [NiFe]-hydrogenases (Figure 2, Suppl. Table 2, Suppl. Information), indicate that members of the Loki- and Thorarchaeota may have the ability to grow lithoautotrophically using H_2 as an electron donor in agreement with previous suggestions[13]. Based on studies of homologous complexes in methanogens[24], it is possible that the group 3c [NiFe]-hydrogenases of Loki- and Thorarchaeota, which are encoded in gene clusters with soluble heterodisulfide reductase subunits (Suppl. Figure 3), bifurcate electrons from H_2 to ferredoxin and an unidentified heterodisulfide compound. However, it is currently unclear how these organisms could generate a membrane potential through this process given that membrane-bound hydrogenases or other ferredoxin-dependent complexes, such as the Rnf complex, capable of ion translocation[25] could not be identified in the genomes. Unless a currently unidentified enzyme complex can couple H_2 oxidation to membrane potential generation, H_2 may

exclusively be used to support carbon fixation or fermentation in these lineages. Instead, the metabolic repertoire of Loki- and Thorarchaeota reveals the potential to harness electrons from a variety of organic substrates (Suppl. Table 1 and 2, Suppl. Information), including complex carbohydrates, peptides, amino acids, alcohols, fatty acids (Suppl. Figures 2b and 4-13) and hydrocarbons (Suppl. Figure 14). We also identified putative formate dehydrogenases that may support growth on formate as an energy and/or carbon source. While Lokiarchaeota may oxidize these organic substrates using the reverse WLP, the presence of a canonical respiratory chain complex I in Thorarchaeota indicates that members of this latter group have the additional ability to couple organic carbon oxidation to membrane potential generation through vectorial proton translocation. Additionally, it may be speculated that both Loki- and Thorarchaeota can establish a membrane potential through a scalar mechanism from the oxidation of organic compounds by using their putative membrane-bound heterodisulfide reductase (HdrD) for the re-oxidation of quinol species generated by electron transfer flavoproteins (ETF) (Figure 1; Suppl. Information). It is possible that cofactors reduced during organic carbon oxidation may be re-oxidised by hydrogenogenic fermentation using the nicotinamide-dependent group 3b[26] and ferredoxin-dependent group 3c [NiFe]-hydrogenases, which are thought to be reversible under physiological conditions (Figures 1 and 2). Thus, depending on environmental conditions, members of these groups might employ hydrogenogenic or hydrogenotrophic metabolisms. For example, depending on the electron yield of a substrate, these organisms could use the WLP as electron sink similar to heterotrophic acetogenic bacteria[27]. In contrast, smaller organic substrates such as short-chain fatty acids with lower electron yields, could be completely oxidized via the reverse WLP. In case of limited availability of electron acceptors, reduced fermentation products may subsequently be metabolized syntrophically by H_2 or formate-consuming organisms, which keep the partial pressure of these intermediates sufficiently low[28]. Finally, the presence of reductive dehalogenases in Loki- and Thorarchaeota suggests that these organisms are also able to shuttle electrons to organohalide compounds (Suppl. Figure 15; Suppl.

Information). However, given the lack of the classical membrane-anchors of bacterial reductive dehalogenases, the functions of these enzymes in Asgard archaea remain to be investigated.

Odinarchaeum may be a thermophilic fermentative heterotroph

The metabolic repertoire of *Odinarchaeum,* which so far represents the only known thermophilic member of the Asgard archaea[4], seems more limited (Figures 1 and 3; Suppl. Tables 1 and 2). Its genome encodes only a partial tricarboxylic acid cycle and lacks genes for several key enzymes of the WLP as well as beta-oxidation pathway (Suppl. Figure 2a). Yet, it encodes enzymes indicative of the potential to grow on organic substrates, which could be fermented to acetate by a putative ADP-dependent acetyl-CoA synthetase. Furthermore, the *Odinarchaeum* genome contains three distinct homologues of group 4 respiratory H_2-evolving [NiFe] hydrogenases (Figure 2; Suppl. Table 1 and 2, Suppl. Information). Phylogenetic analyses of the large subunit of the [NiFe]-hydrogenase suggest that these hydrogenases form three previously undefined subgroups, together with homologs of other recently obtained archaeal lineages (Figure 2). This supports the notion that the classification of group 4 [NiFe]-hydrogenases needs to be extended[29]. The [NiFe]-hydrogenase gene clusters of *Odinarchaeum* encode NuoL-like subunits (Suppl. Figure 3; Suppl. Material), which are thought to mediate sodium/proton translocation[30], indicating that these hydrogenases are involved in energy conservation. This is reminiscent of hydrogen-evolving group 4d [NiFe]-hydrogenases encoded by members of the Thermococci, which are involved in the fermentation of organic substrates to H_2, acetate and carbon dioxide[30,31]. In these archaea, the transporter-linked membrane-bound group 4 [NiFe]-hydrogenases can couple the thermodynamically favorable transfer of electrons from reduced ferredoxin to protons, to the translocation of ions across the membrane. Subsequently, this ion gradient can be harvested through a Na^+- or H^+-driven ATP synthase[31]. The presence of hydrogen-evolving group 4 [NiFe]-hydrogenases in *Odinarchaeum* and its potential to use organic substrates suggest that this organism can conserve energy by a fermentation process that

simultaneously generates ATP through substrate-level phosphorylation during carbon oxidation and oxidative phosphorylation during ferredoxin re-oxidation, thereby increasing the overall ATP yield.

Heimdallarchaeota may grow heterotrophically by fermentation, or anaerobic and aerobic respiration

Heimdallarchaeota, encode a versatile metabolic repertoire (Figure 1; Suppl. Figure 2a; Suppl. Tables 1 and 2) with the potential to derive reducing equivalents from a range of organic substrates including complex carbohydrates, fatty acids and proteins (Suppl. Figure 2b; Suppl. Information). While they lack most enzymes for the WLP, both the LC2 and LC3 genomes encode various electron transport chain components (Figure 1; Suppl. Tables 1 and 2; Suppl. Information), including an A-type heme-copper oxidase (Suppl. Figure 16a), a bacterial-type nitrate reductase (Nar; Suppl. Figure 16b) and a respiratory chain complex I. These components likely enable the use of oxygen and nitrate as electron acceptors during aerobic and anaerobic respiration, respectively. In agreement with this, a close relative of *Heimdallarchaeum* LC2 was recently identified in oxygen-rich oceanic surface waters[32], while members of the Thor- and Lokiarchaeota so far have only been found in strictly anoxic environments. The genome of this oceanic member of the *Heimdallarchaeota* (PBWW00000000.1) encodes a terminal oxidase that is similar (identity 63%, E-value 0.0) to that of *Heimdallarchaeum* LC2 (OLS29256), indicating that some Heimdallarchaeota may be able to occupy both anoxic and oxic niches. In contrast, *Heimdallarchaeum* AB125 has fewer electron acceptors and lacks proteins homologous to those involved in oxygen or nitrogen cycling. However, similar to *Heimdallarchaeum* LC2 and *Odinarchaeum*, its genome encodes a putative H_2-evolving group 4 [NiFe]-hydrogenase that may allow ferredoxin re-oxidation using H^+ as electron acceptor (Suppl. Information). Finally, all members of the Heimdallarchaeota may be able to use organohalides as electron acceptors (Suppl. Information) and *Heimdallarchaeum* AB125 may in addition be able to use hydrocarbons as substrates (Suppl. Figure 14).

The evolution of the metabolic potential of the Asgard archaea

The comparative genome analyses of Asgard archaea allowed us to infer key metabolic features of the **L**ast **As**gard archaeal **C**ommon **A**ncestor (LAsCA) (Figure 3a). First of all, it is likely that LAsCA had the ability to metabolize organic substrates (Suppl. Information). All Asgard archaea, except for *Odinarchaeum*, have a large variety of genes coding for proteins of all steps of the beta-oxidation pathway (Suppl. Figures 5-13; Suppl. Information). While the phylogenetic history of the respective enzymes is complex and invokes several horizontal gene transfers (HGTs) from different sources (Suppl. Figures 5-13), at least some of these enzymes (Suppl. Figures 6b, 8-10) seem to have evolved vertically within the Asgard archaea or across the archaeal domain, supporting the presence of this pathway in LAsCA. Furthermore, all members of the Asgard archaea have the potential to use additional organic substrates including complex organic compounds (Suppl. Tables 1 and 2).

The presence of the WLP in both Thor- and Lokiarchaeota and the wide occurrence of the key enzyme of the WLP (CO dehydrogenase/Acetyl-CoA synthetase (CODH/ACS)) in all major archaeal phyla[5,33] indicate that this pathway was present in LAsCA. However, at least one gene encoding a subunit of the CODH/ACS may have been subjected to HGT after the radiation of the Asgard archaea (Suppl. Figure 17), Furthermore, few to all genes encoding enzymes of the WLP seem to have been lost in Odinarchaeum and currently known Heimdallarchaeota. While an ADP-dependent acetyl-CoA synthetase (IPR014089), an enzyme responsible for the generation of acetate and ATP by substrate-level phosphorylation, is present in all Asgard lineages, phylogenetic analyses are indicative of frequent and recent HGTs. Thus, we cannot infer the presence of this protein in LAsCA with certainty. Furthermore, while all Asgard archaea encode a homolog of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) family enzymes, which among others include the key enzyme of the Calvin-Benson-Bassham (CBB) cycle (RuBisCO Type I and II) as well as of the related reverse hexulose phosphate pathway[34], phylogenetic analyses of these enzymes and the investigation of key residues indicate that the Asgard archaeal homologs belong to RuBisCO sub-types

192 that are unrelated to the carbon fixation cycles (Suppl. Information, Suppl. Figure 18a) and may have been
193 acquired horizontally (Suppl. Figure 18a). In line with this, phosphoribulokinase, another key enzyme of
194 RuBisCO-based carbon fixation pathways, does not seem to be encoded by Asgard genomes (Suppl.
195 Information). Therefore, there is currently no indication that LAsCA contained carbon fixation strategies
196 other than the WLP.

197 Though the wide distribution of cytosolic group 3b and 3c [NiFe]-hydrogenases in the analysed
198 Asgard archaea may indicate their presence in LAsCA, phylogenetic analyses of the key subunit do not
199 support such inferences (Figures 2 and 3, Suppl. Tables 1 and 2). For instance, most homologs encoded by
200 the different Asgard lineages are not monophyletic and are thus indicative of extensive HGT events
201 throughout the evolution of this archaeal group (Figure 2). Furthermore, based on the patchy distribution
202 of membrane-bound group 4 [NiFe]-hydrogenases and NADH dehydrogenases in the Asgard genomes and
203 results of our phylogenetic analyses, it is unclear whether these enzymes were part of the LAsCA
204 proteome (Figure 3a). Yet, based on current data, it is plausible that the respective ancestors of
205 Heimdallarchaeota harbored a membrane-bound group 4 [NiFe]-hydrogenase.

206 While phylogenetic analyses indicate that putative reductive dehalogenase-like proteins may
207 have been encoded by LAsCA (Suppl. Figure 15), the absence of a terminal nitrate and oxygen reductase
208 in most of the Asgard archaea studied herein, with the exception of *Heimdallarchaeum* LC2 and LC3,
209 indicates that these enzyme complexes were acquired more recently in Heimdallarchaeota (Figure 3a).
210 However, the exact timepoint of the acquisition of the genes encoding Nar and the A-type heme-copper
211 oxidase (Suppl. Figure 16a and b) remains to be determined. While the heimdallarchaeal A-type Cox1
212 homologs branch in a cluster with homologs of *Ca.* Caldiarchaeum subterraneum and Crenarchaeota, we
213 obtained strong support for the monophyly of the eukaryotic and alphaproteobacterial A-type homologs,
214 supporting the view that eukaryotes inherited their terminal oxidases from the bacterial endosymbiont
215 as indicated earlier[35]. Similarly, the monophyly of the bacterial-type nitrate reductase of

Heimdallarchaeum LC2 and LC3 with *Methylomirabilis oxyfera*, *Nitrolancea hollandica* and *Nitrococcus* species, as well as the anaerobic methane-oxidizing euryarchaeote (ANME) *Methanoperedens* sp. BLZ1[36], which contains both a bacterial and an archaeal-type Nar, points towards a late acquisition of this gene cluster in Heimdallarchaeota. Thus, terminal reductases for exogenous electron acceptors may have been absent from the LAsCA proteome.

Altogether, we conclude that LAsCA likely encoded the WLP, which is consistent with the suggested ancestry of this pathway in archaea[33,37]. Furthermore, this organism may have had the potential to grow both lithoautotrophically on H_2 and CO_2 as well as organoheterotrophically using the WLP for growth on organic substrates, including fatty acids and perhaps alkanes or aromatic compounds (Figure 3a). It may also have had the ability of fermentative H_2 production given the presence of the bidirectional group 3b and 3c [NiFe]-hydrogenases in all Asgard phyla. In contrast, the WLP was lost in the currently known Heimdallarchaeota, which instead seem to have acquired membrane-bound electron acceptors, that could enable respiratory growth on organic substrates. Yet, it is currently unclear whether the WLP was still encoded by the shared ancestor of Heimdallarchaeaota and eukaryotes (Figure 3a and b).

The reverse flow model for the emergence of the eukaryotic cell

Various symbiogenetic scenarios for the emergence of the eukaryotic cell have been proposed in the past, e.g.[6-8,16,38]. Among these scenarios, the independently formulated but simultaneously published syntrophic hypothesis[39,40] (Figure 4a) and hydrogen hypothesis[41] (Figure 4b) are perhaps the most articulated and detailed examples. Both of these hypotheses assume a syntrophic interaction based on the transfer of H_2 between a H_2-dependent methanogen and a H_2-producing bacterial partner (Figure 4a and b). Following the initial discovery of *Lokiarchaeum*[2], a modified version of the hydrogen hypothesis was proposed (Figure 4c)[13]. Based on the presence of the WLP in Lokiarchaeota, it was suggested that the archaeal ancestor of eukaryotes likely represented an autotrophic H_2-dependent organism. The analysis

of genomic data of additional Asgard archaea (Figure 1, Suppl. Tables 1 and 2, Suppl. Information) including genomic information of members of the Heimdallarchaeota, which currently seem to represent the closest relatives of eukaryotes among the Asgard archaea (Figure 3b)[2,4] allowed us to refine previous syntrophic scenarios for the origin of eukaryotes.

In our model, referred to as the 'reverse flow model' (Figure 4d), the direction of the syntrophic interaction is suggested to be opposite of what was proposed for the hydrogen and syntrophic hypotheses (Figure 4a and b)[6]; i.e. the archaeal ancestor of eukaryotes would have used fermentative pathways to produce reduced substrates, which were syntrophically metabolized by the facultative anaerobic alphaproteobacterial ancestor of mitochondria. In the following, we outline our scenario from the perspective of the archaeal host and bacterial symbiont:

The archaeal host. The metabolic repertoire of Asgard archaea suggests that the archaeal ancestor of eukaryotes had the potential to use organic substrates including fatty acids and alkanes or aromatic compounds for growth (Figure 3a; Supp. Figures 3-13). The fate of reducing equivalents generated during growth on these organics would differ depending on the metabolic repertoire of this organism, substrate availability and the presence of available electron sinks. In anoxic environments, in which electron acceptors other than CO_2 are often scarce, the syntrophic degradation of organic matter is common[28,42]. Methanogens usually represent the final reducers of biomass in such environments and outcompete autotrophic acetogens. However, most acetogenic bacteria are able to metabolize a wide variety of substrates, often in syntrophy with partner organisms, rather than by using CO_2 as electron acceptor through the WLP[20]. In line with this, we envision a scenario in which the last common ancestor of Heimdallarchaea and eukaryotes degraded small organic compounds in syntrophy with one or more bacterial partners (Figure 4d). For example, the loss of the WLP, a potential electron sink of organoheterotrophically growing acetogens[27], and gain of a membrane-bound hydrogenase (a scenario

observed in Heimdallarchaeota) would have provided a selective pressure for the maintenance of such a relationship. Electrons produced by the archaeal host during the oxidation of organic substrates could have been transferred to the partner in the form of H_2, formate, acetate or via direct electron transfer[43,44]. Alternatively, this ancestor may have coupled the thermodynamically unfavorable anaerobic oxidation of an alkane (e.g. butane) or related compounds through the shuttling of electrons to a bacterial partner (similar to ANME consortia[45]) (Figure 4d). This possibility is inspired by the discovery of an "alkyl-coenzyme M reductase" (encoded by *mcr* genes) in the Helarchaeota[46], which is related to the butane-coenzyme M reductases of Syntrophoarchaea, a group of euryarchaeota that grows in syntrophy with *Ca.* Desulfofervidus auxilii[47]. However, this latter possibility is currently not compatible with the placement of eukaryotes sister to Heimdallarchaeota (Figure 3b), the sparse distribution of these *mcr* genes within the currently sampled Asgard diversity[46] and evidence for HGT of *mcr* gene homologs in other archaeal lineages[5].

The bacterial partner. A recent study indicates that mitochondria derive from a lineage that shares a common ancestry with all known alphaproteobacteria excluding *Magnetococcales*[48]. While ancestral reconstructions are needed to infer the metabolic potential of the alphaproteobacterial ancestor of mitochondria, it is notable that various extant members of this class encode group 1c and 1d [NiFe]-hydrogenases that are known respiratory H_2-uptake hydrogenases and thus allow hydrogenotrophic growth[29]. Some anaerobically functioning mitochondria and related organelles of eukaryotes produce H_2 via a [FeFe]-hydrogenases, the origin of which is still debated[49]. In particular, the vast majority of alphaproteobacteria do not encode an [FeFe]-hydrogenase[29], and the few representatives that do are not from mitochondrial sister groups[48,50]. Furthermore, robust phylogenetic analyses of these proteins fail to support an alphaproteobacterial origin of these proteins in eukaryotes[49] suggesting that the alphaproteobacterial endosymbiont may have oxidized, rather than produced H_2. Thus, hydrogen-evolving [FeFe]-hydrogenases and related enzymes, which are the defining feature of mitochondria-

288 derived organelles in anaerobic eukaryotes, may have been acquired horizontally from bacterial sources
289 other than the alphaproteobacterial endosymbiont subsequent to eukaryogenesis (Figure 4d).
290 Alphaproteobacteria have been shown to engage in syntrophic interactions within microbial consortia[51]
291 and to be able to accept electrons from a variety of donors including from extracellular electrons[52]. We
292 speculate that the alphaproteobacterial ancestor of mitochondria may have served as electron sink for
293 the archaeal partner either under anoxic conditions (e.g. using fumarate or inorganic compounds such as
294 sulfate, or nitrate as electron acceptor) or micro-oxic conditions (using oxygen as electron acceptor).
295
296 Altogether, and in agreement with previous syntrophic models for eukaryogenesis, our scenario
297 is based on the observation that syntrophic interactions are widespread, in particular in anoxic
298 environments[42], and examples include methane or butane oxidizing euryarchaeota, many of which grow
299 in microbial consortia and are obligately dependent on a partner organism[45]. While speculative, we
300 propose that the ESPs encoded by Asgard archaea[2,4] may have aided the transition from a metabolic
301 syntrophy to a more intricate symbiosis. For example, the generation of actin-based cellular protrusions
302 could have led to a tighter interaction between the partners allowing for more efficient metabolic
303 coupling, with the bacterial symbiont(s) eventually being effectively engulfed by the host cell. The origin
304 of the defining features of eukaryotic cells including the nucleus and the relative timing of their emergence
305 remain to be established[53]. However, it is possible that the acquisition of mitochondria was an
306 intermediate event that occurred after the invention of a basic intermembrane-system and actin
307 cytoskeleton but predated the origin of the nucleus[54]. Furthermore, it remains an open question whether
308 membrane lipids could have been of mixed origin during the transition phase[55] and were subsequently
309 replaced by bacterial fatty-acid-type lipids in the outer membrane, following endosymbiont gene
310 transfer[56], gene acquisitions from other sources[57,58] and differential gene loss. Bacterial lipids may have
311 had a selective advantage considering that genes for most metabolic enzymes, including respiratory

chains, were retained from the endosymbiont rather than the host[19] and were likely functioning optimally in bacterial membrane lipids.

Conclusions

Our analyses have revealed that Asgard archaea are metabolically versatile: while Lokiarchaeota may dominantly grow as fermentative organoheterotrophs, Thorarchaeota may be able to derive energy from hydrogen, formate and different organic substrates including fatty acids and hydrocarbons. The presence of a reductive dehalogenase in members of these phyla suggests that they may also be able to use organohalides as electron acceptors during respiration. *Odinarchaeum* seems to represent an organoheterotrophic fermentative organism that can potentially recycle the ferredoxin reduced during fermentation reactions using H^+ as an electron acceptor. In contrast, at least some members of the Heimdallarchaeota appear to have the ability to grow by anaerobic and/or aerobic respiration while others are obligate anaerobes. Based on these analyses, we propose a refined symbiogenetic scenario for the origin of eukaryotes that is based on the transfer of reducing equivalents from the archaeal host to a bacterial partner. We acknowledge that while such a reversed electron flow is most easily reconcilable with the gene set of the herein analyzed Asgard archaeal MAGs, alternative scenarios (Figure 4) cannot be excluded. Subsequent metabolic reconstructions of additional members of this group and of the alphaproteobacteria as well as the functional characterization of ESPs in extant members of the Asgard archaea, together with insights into their cell biology, will be essential to further refine potential symbiotic interactions that may have played a role in the process of eukaryogenesis.

Methods

Annotation.

The metagenomic bins of the different Asgard members discussed in the manuscript have been annotated automatically and were published in[4]. For metabolic reconstructions all proteins were analyzed with IPRscan (version: interproscan-5.22-61.0) to determine protein domain information (IPR domains and PFAMs) as well as KO numbers. Additionally, proteins were queried against NCBI nonredundant (nr) database (February 2017) using Diamond blast[59] and top hits with and without Asgards were extracted. Taxonomy information for the top blast hits was determined using blastdbcmd. Furthermore, arCOGs and COGs were assigned based on the publicly available arCOGs from[4,60]. Proteins were also queried against proteins in the Transporter Classification Database[61] to identify potential homology to proteins associated with transport functions. Finally, the putative large subunits of [NiFe]-hydrogenases present in Asgard genomes (proteins containing PF00374/PF00346 domains) were extracted and classified using HydDB[62], as well as by phylogenetic analyses (see below). The presence of conserved N-and C-terminal C-xx-C motifs characterising the large subunit of [NiFe]-hydrogenases was determined in Jalview. Finally all this data was compiled in a tsv file containing relevant information from all these analyses was generated (Suppl. Table 3). All proteins discussed throughout the manuscript are listed in Suppl. Tables 1-3. Please note that, although the Asgard MAGs are of medium to high quality (Suppl. Figure 1) [63], we cannot currently exclude that the absence of specific genes could be due to genome bin incompleteness.

Prediction of carbohydrate active enzymes and esterases.

Genes encoding for carbohydrate-active enzymes were searched using the Carbohydrate-Active enZYmes (CAZYmes) database using the dbCAN webtool[64]. To search for peptidases, protein sequences were blasted against the MEROPS peptidase database using blastp and an e-value cutoff of $1e^{-20}$ (database downloaded June 2017[65]. Esterases were determined by downloading the HMM profiles from the ESTHER database and searching for positive hits against the Asgard protein sequences using hmmsearch and an e-value cutoff of $1e\text{-}10^{20}$ (downloaded September 2017)[66]. The sequence with the best e-value was

359 selected in case of multiple hits for a given protein-coding sequence. Protein localization of all positive
360 hits was determined using the server-version of PSORT v3.0 (-a option for archaeal sequences)[67].
361 Singletons, hits related to central metabolism (i.e. archaeal proteasome, precursor proteins) and
362 peptidases included in the ESTHER database (to avoid redundancy to the MEROPS database) were not
363 included for further comparisons.
364
365 **Phylogenetic analyses of concatenated ribosomal proteins**
366 In order to reconstruct a species tree including the Heimallarchaeota Chinese Sea genome B2-JM-08 (WGS
367 accession NJBF000000), we used BLASTp to identify orthologues of 56 ribosomal proteins, which we
368 added to the pre-existing alignments taken from[4]. Individual protein datasets were aligned using Mafft-
369 LINSi[68] and ambiguously aligned positions were trimmed using BMGE (-m BLOSUM30)[69]. Maximum
370 likelihood (ML) individual phylogenies were reconstructed using IQ-tree v. 1.5.5[70] under the LG+C20+F+G
371 substitution model with 1000 ultrafast bootstraps and were manually inspected. Trimmed alignments
372 were concatenated into a supermatrix, and an additional dataset was generated by removing DPANN
373 homologues to test the impact of taxon sampling on phylogenetic reconstruction. For each of these
374 concatenated datasets, phylogenies were inferred using ML and Bayesian approaches. ML phylogenies
375 were reconstructed using IQ-tree under the LG+C60+F+G+PMSF model[71] (Figure 3b and Suppl. File 1).
376 Statistical support for branches was calculated using 100 bootstraps replicated under the same model. To
377 test robustness of the placement of eukaryotes, the datasets were subjected to several treatments. For
378 the 'full dataset' and the 'no DPANN dataset' (i.e., 144 and 104 taxa, respectively), we tested the impact
379 of removing the 25% fastest-evolving sites, since it has been shown that in large-scale phylogenetic
380 analyses, these sites are often saturated with multiple substitutions and, as a result of model-
381 misspecification, can generate an artefactual topology [72]. The corresponding ML trees were inferred as
382 described above (Suppl. Files 2 and 3). Bayesian phylogenies were reconstructed with Phylobayes[73] under

the LG+GTR model for the 'no DPANN' dataset): four independent Markov chain Monte Carlo chains were run for ~40 000 generations. After a burn-in of 20%, convergence was achieved for all chains (maxdiff < 0.104). This supermatrix was also recoded into 4 categories in order to ameliorate effects of model misspecification and saturation[74] and the corresponding phylogeny was reconstructed with Phylobayes under the CAT+GTR model. Four independent Markov chain Monte Carlo chains were run for ~160 000 generations. After a burn-in of 20%, convergence was achieved for all chains (maxdiff < 0.103).

Phylogenetic analyses of individual protein sequences

Cytochrome c oxidase, subunit 1 (PF00115, Cox). All sequences in UniProtKB (http://www.uniprot.org) assigned to PF00115 (967,991 sequences) were extracted from Archaea, Bacteria and eukaryotes excluding Opisthokonta (929,952 sequences), from which only reviewed sequences assigned to PF00115 were extracted. This set of sequences includes the distantly related Nitric oxide reductase family. Backbone sequences were filtered by identity using cd-hit[75] (cutoff of 65-70 % for eukaryotes and bacteria and 85% for archaea) and by length (cutoff 400 AA). Alignments were iteratively refined by removing long-branches and poorly aligned sequences after inspection of initial alignments (using hmm-align)[76] and phylogenies (using FastTree, LG model)[77]. In addition, Cox homologs reviewed in swissprot and predicted cytochrome c oxidase, subunit 1 homologs of Asgard archaea (only in Heimdallarchaeota) and some additional archaea not present in uniprot were added to this dataset and datasets were filtered again using cd-hit (cutoff 85%). The final alignment of 443 positions is based on an hmm-alignment in which sequences were aligned to hmm-profile of PF00115 using the --trim flag to only keep the conserved Cox1 domain. Alignments were trimmed using trimAL[78] with the gappyout option and phylogenetic analyses were performed using IQ-tree[70] using under the LG+C20 model with ultrafast bootstraps[79].

406 *[NiFe]-Hydrogenase, large subunit.* Backbone sequences for classifying [NiFe]-hydrogenase were based on
407 HydDB[29]. To reduce sampling size, the backbone dataset was filtered using cd-hit[75] with a sequence
408 identity cutoff of 90% before concatenation with put. large subunit [NiFe]-hydrogenases of Asgard
409 archaea. HeimC3_03700 and HeimC2_23100 were removed from the dataset to prevent potential LBA
410 artifacts as these sequences represented very long branches and have extremely low similarity to known
411 homologs (<30%). They may represent previously uncharacterized groups that should be investigated in
412 the future. All sequences for group 3 and group 4 large subunit [NiFe]-hydrogenase were aligned using
413 Mafft-LINSi[68] and trimmed with BMGE (using an entropy score of 0.55 and 0.65, respectively)[69]. Maximum
414 likelihood phylogenetic analyses were performed using IQ-tree[70] with the best-fit model according to BIC
415 LG+C60+R+F (group 4) and LG+C50+R+F (group 3)). Support values were estimated using SH-like
416 approximate likelihood ratio test[80] and ultrafast bootstraps[79], respectively.
417
418 *RuBisCO.* All proteins of Asgard archaea assigned to orthologues assigned to COG01850 as well as
419 additional homologs of recently sequenced archaeal genomes were extracted and aligned with a
420 representative set of RuBisCO homologs, (including Family I through Family IV) that was based on a
421 backbone dataset kindly provided by Anantharaman[81]. Sequences were aligned using Mafft-LINSi[68] and
422 trimmed with BMGE (entropy score set to 0.55)[69] and the final alignment of 359 aligned positions was
423 subsequently subjected to maximum likelihood analyses using IQ-tree (using the LG+C60+R+F model
424 chosen based on BIC score)[70]. The presence of conserved amino acids was investigated in Jalview based
425 on residues given in[82]. Support values were estimated using SH-like approximate likelihood ratio test[80]
426 and ultrafast bootstraps[79], respectively.
427
428 *Acetyl-CoA synthase/CO dehydrogenase.* Protein sequences with PF03598 encoding the key subunit of
429 Acetyl-CoA synthase/CO dehydrogenase, CdhC (also referred to as CODH/acetyl-CoA synthase beta

subunit or alpha subunit of ACS/CODH) were extracted from UniProtKB (http://www.uniprot.org) (which included Asgard archaeal homologs) and filtered based on a length cutoff of 400 AA (except for the partial homolog of Lokiarchaeaum) and a sequence identity of 95% using cd-hit[75]. Furthermore, sequences lacking taxonomy information (labeled as uncultured) and poorly aligned sequences that likely represent distant paralogs were removed. Final phylogenies are based on Mafft-LINSi alignments[68] upon trimming with TrimAL 50[78], which were subjected to maximum likelihood analyses using IQ-tree[70] with LG+C20+F+R. Bacterial cdhC homologs as well as few archaeal homologs that branch within bacteria encode an N-terminal domain absent in the other canonical archaeal homologs (length ca, 250AA). The N-terminal domain was retained upon trimming with TrimAL 50% and yielded longer alignments (712AA) as compared to BMGE. Support values were estimated using SH-like approximate likelihood ratio test[80] and ultrafast bootstraps[79], respectively.

Beta-oxidation enzymes. Unless otherwise stated, the top 1000 bacterial, archaeal and eukaryotic sequences were retrieved from Genbank non-redundant database (nr) using each Asgard sequence as a query. These datasets were combined and reduced using cd-hit[75] using taxonomy specific sequence identity reduction thresholds for opisthokonts (50%), archaeplastids (50%), bacteria (80%) and archaea (80%). Initial alignments were made with hmmalign using PFAM domains as indicated and trimmed using BMGE[69] (-h 0.6 and -m BLOSUM30). Initial trees were generated with Fasttree[77]. The datasets were further reduced by manual inspection of the trees by removing well supported clades distant to the Asgard sequences and selecting representative taxa. For example, well supported clades composed of dozens of clostridiales were reduced to between 5-10 diverse representatives. This process was repeated iteratively until there were less than 1000 taxa at which point alignments were generated using mafft-LINSi[68] and trees estimated with IQ-tree[70] (LG+C20+G+F) with 1000 ultrafast bootstrap replicates[79].

Acyl-CoA dehydrogenase (ACAD, COG1960) sequences were retrieved from[83] and used for identifying subcategories of ACAD enzymes and combined with the sequences retrieved from nr (described above). Initial alignments were made as described above using hmmalign with profile alignments for PF02771.14 PF00441.22 and PF02770.17 and concatenated. Acetoacetyl-CoA acyltransferases (COG0183) sequences were retrieved from[84] and nr as described above. Initial alignments were generated as described above using the PF00501.26 hmm profile.

Alignments for the following proteins were generated as described above using the indicated PFAM hmm profile: AMP-dependent synthetase and ligase (COG0318; PF00501.26), enoyl-CoA hydratase (COG1024; PF00378.18), 3-hydroxy-acyl-CoA dehydrogenase (COG1250; PF00725.20, PF02737.16). Initial alignments of alternative acyl-CoA dehydrogenase (COG2368) were generated using Mafft-LINSi[68].

Pyruvate-Formate Lyase superfamily. All sequence assigned to IPR004184 (Pyruvate formate lyase domain) were downloaded from UniProtKB (http://www.uniprot.org) and filtered by length (only archaeal and bacterial sequences between 700 AA and 950 AA were kept). Bacterial homologs were additionally filtered by identity using cd-hit[75] with a cutoff of 75%. Note: all homologs of Asgard archaea were kept. Sequences of poor quality as well as poorly aligned sequences were removed upon inspection of initial Mafft-LINSi alignments. The final set of sequences was re-aligned using Mafft-LINSi[68] and trimmed with BMGE[69] (BLOSUM30, entropy: 0.55) leaving 336 sites, which were subjected to Maximum Likelihood phylogenetic analyses using IQ-tree[70] (LG+C20+F+R). Support values were estimated using SH-like approximate likelihood ratio test[80] and ultrafast bootstraps[79], respectively.

Nitrate Reductase subunit alpha (NarG). All sequences with similarity to InterPro domain IPR006468 (Nitrate reductase, alpha subunit) were downloaded from UniProtKB (http://www.uniprot.org). Note: this IPR domain only includes bacterial-type Nitrate reductases since the archaeal-type NarG are only distantly

related (e.g. 24% identity and 56% sequence coverage between the bacterial- and archaeal-type (WP_097297416.1, WP_097297472.1) NarG of *Methanoperedens* sp. BLZ1. Sequences were filtered by length (most sequences between 1300 and 1100 AA, which were thus used as lower and upper thresholds, respectively) and identity (using cd-hit[75] with a threshold of 75%) to decrease the amount of sequences while keeping taxonomic representation. Subsequently, sequences were aligned with Mafft-LINSi[68], and trimmed using BMGE[69] (BLOSUM30, entropy 0.55). The final alignment of 1071 positions was subjected to maximum likelihood phylogenetic analyses using IQ-tree[70] (LG+C20+F+R). Support values were estimated using SH-like approximate likelihood ratio test[80] and ultrafast bootstraps[79], respectively.

Reductive dehalogenase. Archaeal proteins assigned to IPR028894 and sequences from marine metagenomes (UniProtKB; http://www.uniprot.org) as well as all Asgard homologs assigned to the more broadly defined COG01600 (includes both reductive dehalogenases and epoxyqueuosine reductases) were added to a bacterial backbone dataset kindly provided by Laura Hug and published in[85]. The even more specific protein domain IPR012832 defining reductive dehalogenases is present in a subset of these Asgard sequences and in the sequences of four Euryarchaeota (2 sequences from Theionarchaea, Ferroglobus and *Methanohaloarchaeum thermophilum*). This set of sequences was pre-aligned using Mafft-LINSi[68] to identify and remove poorly aligned or partial sequences (*Note, one lokiarchaeal protein as well as the homolog of *Methanohalarchaeum thermophilum* was removed in this way). Subsequently, sequences were realigned using Mafft-LINSi[68], trimmed with BMGE[69] using the BLOSUM30 matrix combined with an entropy of 0.6 leaving 160 sites that subjected to maximum likelihood analyses using IQ-tree[70] (using best-fit model according to BIC determined by IQ-tree: LG+C40+R+F). Support values were estimated using SH-like approximate likelihood ratio test[80] and ultrafast bootstraps[79], respectively. The tree was rooted using midpoint rooting. However, most likely all sequences shaded in light green in Suppl. Figure 15, do not represent *bona fide* reductive dehalogenases, which is supported by the absence of the

501 characteristic domain IPR012832. The substrates that can be degraded by some of the characterized
502 reductive dehalogenases are indicated on the tree based on[86,87].
503 All trees were visualized using FigTree (http://tree.bio.ed.ac.uk/publications/) and modified/annotated
504 with adobe illustrator to increase readability.
505

506 ## Data availability

507 The genomes of the herein analysed Asgard archaea have been made publicly available on NCBI
508 previously[2,4]. Detailed annotations of the metabolic repertoire are provided in Suppl. Tables 1-3
509 accompanying this manuscript. Raw data files are made available via figshare under the following link:
510 https://figshare.com/s/5f153d1dcacadd3b3ed6.
511

512 ## Code availability

513 Small custom scripts used for genome annotation and phylogenetic analyses are made available on
514 figshare and can be accessed under the following link: https://figshare.com/s/5f153d1dcacadd3b3ed6.
515

516 ## Acknowledgements

517 This work was supported by grants of the European Research Council (ERC Starting grant 310039-
518 PUZZLE_CELL to T.J.G.E.), the Swedish Foundation for Strategic Research (SSF-FFL5 to T.J.G.E.), the
519 Swedish Research Council (VR grant 2015-04959 to T.J.G.E. and VR starting grant 2016-03559 to A.S.), the
520 NWO-I foundation of the Netherlands Organisation for Scientific Research (WISE fellowship to A.S.), the
521 European Commission (Marie Curie IEF European grants 625521 to A.S. and 704263 to L.E.), the Wenner-
522 Gren Foundations in Stockholm (UPD2016-0072 to J.L.), the European Molecular Biology Organization
523 (EMBO long-term fellowship ALTF-997-2015 to C.W.S), the Natural Sciences and Engineering Research

Council of Canada (C.W.S), the Australian Research Council (DE170100310 and DP180101762 to C.G) and the National Science Foundation (DEB: Systematics and Biodiversity Sciences; award number 1737298 to B.J.B.).

We would like to thank Kasia Zaremba-Niedzwiedzka and Jimmy Saw for reconstruction some of these genomes and helpful discussions. We would also like to acknowledge Steffen L. Jørgensen, the chief scientist R. B. Pedersen, the scientific party and the entire crew on board the Norwegian research vessel G.O. Sars during the summer 2010 expedition, which allowed us access to samples from Loki's Castle. Finally, we thank Pierre Offre for discussions on metabolic inferences.

Author Contributions

A.S. and T.J.G.E. conceived the study. A.S., C.S., E.F.C., J.L., C. G., B.J.B. and N.D. analysed genomic data. A.S., C.S. and L.E. performed phylogenetic analyses. A.S. and T.J.G.E. wrote the manuscript with input from all authors and A.S., C.S. and N.D wrote the Supplementary Information. All documents were edited and approved by all authors.

Author Information

 Correspondence and requests for materials should be addressed to A.S. (anja.spang@nioz.nl) or T.J.G.E. (thijs.ettema@icm.uu.se)

Competing interests

The authors declare no competing financial interests.

References

1 Hug, L. A. *et al.* A new view of the tree of life. *Nature Microbiology* **1**, doi:Artn 16048 10.1038/Nmicrobiol.2016.48 (2016).

2 Spang, A. *et al.* Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173-+, doi:10.1038/nature14447 (2015).

3 Seitz, K. W., Lazar, C. S., Hinrichs, K. U., Teske, A. P. & Baker, B. J. Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction. *The ISME Journal*, doi:10.1038/ismej.2015.233 (2016).

4 Zaremba-Niedzwiedzka, K. *et al.* Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353-358, doi:10.1038/nature21031 (2017).

5 Spang, A., Caceres, E. F. & Ettema, T. J. G. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* **357**, doi:10.1126/science.aaf3883 10.1126/science.aaf3883. (2017).

6 Lopez-Garcia, P. & Moreira, D. Open Questions on the Origin of Eukaryotes. *Trends Ecol Evol* **30**, 697-708, doi:10.1016/j.tree.2015.09.005 (2015).

7 Guy, L., Saw, J. H. & Ettema, T. J. The archaeal legacy of eukaryotes: a phylogenomic perspective. *Cold Spring Harb Perspect Biol* **6**, a016022, doi:10.1101/cshperspect.a016022 (2014).

8 Martin, W. F., Garg, S. & Zimorski, V. Endosymbiotic theories for eukaryote origin. *Philos Trans R Soc Lond B Biol Sci* **370**, 20140330-20140330, doi:10.1098/rstb.2014.0330 10.1098/rstb.2014.0330. (2015).

9 Da Cunha, V., Gaia, M., Gadelle, D., Nasir, A. & Forterre, P. Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLoS Genet* **13**, e1006810, doi:10.1371/journal.pgen.1006810 (2017).

10 Spang, A. *et al.* Asgard archaea are the closest prokaryotic relatives of eukaryotes. *PLoS Genet* **14**, e1007080-e1007080, doi:10.1371/journal.pgen.1007080 10.1371/journal.pgen.1007080. eCollection 2018 Mar. (2018).

11 Narrowe, A. B. *et al.* Complex evolutionary history of translation Elongation Factor 2 and diphthamide biosynthesis in Archaea and parabasalids. *Genome Biol Evol*, doi:10.1093/gbe/evy154 (2018).

12 Klinger, C. M., Spang, A., Dacks, J. B. & Ettema, T. J. G. Tracing the Archaeal Origins of Eukaryotic Membrane-Trafficking System Building Blocks. *Molecular Biology and Evolution* **33**, 1528-1541, doi:10.1093/molbev/msw034 (2016).

13 Sousa, F. L., Neukirchen, S., Allen, J. F., Lane, N. & Martin, W. F. Lokiarchaeon is hydrogen dependent. *Nature Microbiology* **1**, doi:Artn 16034 10.1038/Nmicrobiol.2016.34 (2016).

14 Martin, W. F., Tielens, A. G. M., Mentel, M., Garg, S. G. & Gould, S. B. The Physiology of Phagocytosis in the Context of Mitochondrial Origin. *Microbiology and Molecular Biology Reviews : MMBR* **81**, doi:10.1128/MMBR.00008-17
10.1128/MMBR.00008-17. Print 2017 Sep. (2017).

15 Zachar, I., Szilagyi, A., Szamado, S. & Szathmary, E. Farming the mitochondrial ancestor as a model of endosymbiotic establishment by natural selection. *Proceedings of the National Academy of Sciences of the United States of America* **115**, E1504-E1510, doi:10.1073/pnas.1718707115
10.1073/pnas.1718707115. Epub 2018 Jan 30. (2018).

16 Speijer, D. Alternating terminal electron-acceptors at the basis of symbiogenesis: How oxygen ignited eukaryotic evolution. *Bioessays* **39**, doi:10.1002/bies.201600174 (2017).

17 Koonin, E. V. Origin of eukaryotes from within archaea, archaeal eukaryome and bursts of gene gain: eukaryogenesis just made easier? *Philos Trans R Soc Lond B Biol Sci* **370**, 20140333-20140333, doi:10.1098/rstb.2014.0333 (2015).

18 Williams, T. A., Foster, P. G., Cox, C. J. & Embley, T. M. An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* **504**, 231-236, doi:10.1038/nature12779 (2013).

19 Lopez-Garcia, P., Eme, L. & Moreira, D. Symbiosis in eukaryotic evolution. *J Theor Biol* **434**, 20-33, doi:10.1016/j.jtbi.2017.02.031
10.1016/j.jtbi.2017.02.031. Epub 2017 Feb 28. (2017).

20 Ragsdale, S. W. & Pierce, E. Acetogenesis and the Wood-Ljungdahl pathway of CO(2) fixation. *Biochim Biophys Acta* **1784**, 1873-1898, doi:10.1016/j.bbapap.2008.08.012 (2008).

21 Schuchmann, K. & Muller, V. Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat Rev Microbiol* **12**, 809-821, doi:10.1038/nrmicro3365 (2014).

22 Adam, P. S., Borrel, G., Brochier-Armanet, C. & Gribaldo, S. The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *ISME J* **11**, 2407-2425, doi:10.1038/ismej.2017.122
10.1038/ismej.2017.122. Epub 2017 Aug 4. (2017).

23 Liu, Y. *et al.* Comparative genomic inference suggests mixotrophic lifestyle for Thorarchaeota. *ISME J* **12**, 1021-1031, doi:10.1038/s41396-018-0060-x (2018).

24 Wagner, A. *et al.* Mechanisms of gene flow in archaea. *Nat Rev Microbiol*, doi:10.1038/nrmicro.2017.41
10.1038/nrmicro.2017.41. (2017).

25 Buckel, W. & Thauer, R. K. Energy conservation via electron bifurcating ferredoxin reduction and proton/Na(+) translocating ferredoxin oxidation. *Biochim Biophys Acta* **1827**, 94-113, doi:10.1016/j.bbabio.2012.07.002
10.1016/j.bbabio.2012.07.002. Epub 2012 Jul 16. (2013).

26 Bryant, F. O. & Adams, M. W. Characterization of hydrogenase from the hyperthermophilic archaebacterium, Pyrococcus furiosus. *J Biol Chem* **264**, 5070-5079 (1989).

27 Schuchmann, K. & Muller, V. Energetics and application of heterotrophy in acetogenic bacteria. *Appl Environ Microbiol*, doi:10.1128/AEM.00882-16 (2016).

28 Stams, A. J. & Plugge, C. M. Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nat Rev Microbiol* **7**, 568-577, doi:10.1038/nrmicro2166 (2009).

29 Greening, C. *et al.* Genomic and metagenomic surveys of hydrogenase distribution indicate H2 is a widely utilised energy source for microbial growth and survival. *The ISME Journal* **10**, 761-777, doi:10.1038/ismej.2015.153
10.1038/ismej.2015.153. Epub 2015 Sep 25. (2016).

30 Yu, H. *et al.* Structure of an Ancient Respiratory System. *Cell* **173**, 1636-1649 e1616, doi:10.1016/j.cell.2018.03.071 (2018).

31 Schut, G. J., Boyd, E. S., Peters, J. W. & Adams, M. W. The modular respiratory complexes involved in hydrogen and sulfur metabolism by heterotrophic hyperthermophilic archaea and their evolutionary implications. *FEMS Microbiol Rev* **37**, 182-203, doi:10.1111/j.1574-6976.2012.00346.x (2013).

32 Tully, B. J., Graham, E. D. & Heidelberg, J. F. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Sci Data* **5**, 170203-170203, doi:10.1038/sdata.2017.203
10.1038/sdata.2017.203. (2018).

33 Adam, P. S., Borrel, G. & Gribaldo, S. Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proceedings of the National Academy of Sciences of the United States of America* **115**, E1166-E1173, doi:10.1073/pnas.1716667115
10.1073/pnas.1716667115. Epub 2018 Jan 22. (2018).

34 Kono, T. *et al.* A RuBisCO-mediated carbon metabolic pathway in methanogenic archaea. *Nat Commun* **8**, 14007-14007, doi:10.1038/ncomms14007
10.1038/ncomms14007. (2017).

35 Lang, B. F., Gray, M. W. & Burger, G. Mitochondrial genome evolution and the origin of eukaryotes. *Annu Rev Genet* **33**, 351-397, doi:10.1146/annurev.genet.33.1.351
10.1146/annurev.genet.33.1.351. (1999).

36 Arshad, A. *et al.* A Metagenomics-Based Metabolic Model of Nitrate-Dependent Anaerobic Oxidation of Methane by Methanoperedens-Like Archaea. *Front Microbiol* **6**, 1423-1423, doi:10.3389/fmicb.2015.01423
10.3389/fmicb.2015.01423. eCollection 2015. (2015).

37 Williams, T. A. *et al.* Integrative modeling of gene and genome evolution roots the archaeal tree of life. PNAS **114(23)**:E4602-E4611 (2017).

38 Zachar, I. & Szathmary, E. Breath-giving cooperation: critical review of origin of mitochondria hypotheses : Major unanswered questions point to the importance of early ecology. *Biol Direct* **12**, 19-19, doi:10.1186/s13062-017-0190-5
10.1186/s13062-017-0190-5. (2017).

652 39 Moreira, D. & Lopez-Garcia, P. Symbiosis between methanogenic archaea and delta-
653 proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J Mol Evol* **47**, 517-530 (1998).

654 40 López-García, P. & Moreira, D. Selective forces for the origin of the eukaryotic nucleus. *BioEssays:*
655 *News and Reviews in Molecular, Cellular and Developmental Biology* **28**, 525-533, doi:10.1002/bies.20413
656 (2006).

657 41 Martin, W. & Muller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37-41,
658 doi:10.1038/32096
659 10.1038/32096. (1998).

660 42 Sieber, J. R., McInerney, M. J. & Gunsalus, R. P. Genomic insights into syntrophy: the paradigm for
661 anaerobic metabolic cooperation. *Annu Rev Microbiol* **66**, 429-452, doi:10.1146/annurev-micro-090110-
662 102844
663 10.1146/annurev-micro-090110-102844. Epub 2012 Jul 9. (2012).

664 43 McGlynn, S. E., Chadwick, G. L., Kempes, C. P. & Orphan, V. J. Single cell activity reveals direct
665 electron transfer in methanotrophic consortia. *Nature* **526**, 531-535, doi:10.1038/nature15512 (2015).

666 44 Wegener, G., Krukenberg, V., Riedel, D., Tegetmeyer, H. E. & Boetius, A. Intercellular wiring
667 enables electron transfer between methanotrophic archaea and bacteria. *Nature* **526**, 587-590,
668 doi:10.1038/nature15733 (2015).

669 45 Knittel, K. & Boetius, A. Anaerobic oxidation of methane: progress with an unknown process.
670 *Annual review of microbiology* **63**, 311-334, doi:10.1146/annurev.micro.61.080706.093130 (2009).

671 46 Seitz, K. W. *et al.* New Asgard archaea capable of anaerobic hydrocarbon cycling. *bioRxiv* (2019).

672 47 Laso-Perez, R. *et al.* Thermophilic archaea activate butane via alkyl-coenzyme M formation.
673 *Nature* **539**, 396-401, doi:10.1038/nature20152 (2016).

674 48 Martijn, J., Vosseberg, J., Guy, L., Offre, P. & Ettema, T. J. G. Deep mitochondrial origin outside the
675 sampled alphaproteobacteria. *Nature* **557**, 101-105, doi:10.1038/s41586-018-0059-5 (2018).

676 49 Leger, M. M., Eme, L., Stairs, C. W. & Roger, A. J. Demystifying Eukaryote Lateral Gene Transfer
677 (Response to Martin 2017 DOI: 10.1002/bies.201700115). *Bioessays* **40**, e1700242,
678 doi:10.1002/bies.201700242 (2018).

679 50 Stairs, C. W. *et al.* Microbial eukaryotes have adapted to hypoxia by horizontal acquisitions of a
680 gene involved in rhodoquinone biosynthesis. *Elife* **7**, doi:10.7554/eLife.34292 (2018).

681 51 Norlund, K. L. *et al.* Microbial architecture of environmental sulfur processes: a novel syntrophic
682 sulfur-metabolizing consortia. *Environ Sci Technol* **43**, 8781-8786, doi:10.1021/es803616k (2009).

683 52 Bose, A., Gardel, E. J., Vidoudez, C., Parra, E. A. & Girguis, P. R. Electron uptake by iron-oxidizing
684 phototrophic bacteria. *Nat Commun* **5**, 3391, doi:10.1038/ncomms4391 (2014).

685 53 Eme, L., Spang, A., Lombard, J., Stairs, C. W. & Ettema, T. J. G. Archaea and the origin of
686 eukaryotes. *Nat Rev Microbiol* **16**, 120-120, doi:10.1038/nrmicro.2017.154

687 10.1038/nrmicro.2017.154. Epub 2017 Nov 27. (2018).

688 54 Ettema, T. J. Evolution: Mitochondria in the second act. *Nature* **531**, 39-40,
689 doi:10.1038/nature16876 (2016).

690 55 Caforio, A. *et al.* Converting Escherichia coli into an archaebacterium with a hybrid heterochiral
691 membrane. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 3704-
692 3709, doi:10.1073/pnas.1721604115
693 10.1073/pnas.1721604115. Epub 2018 Mar 19. (2018).

694 56 Martin, W. *et al.* Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* **393**, 162-
695 165, doi:10.1038/30234
696 10.1038/30234. (1998).

697 57 Pittis, A. A. & Gabaldon, T. Late acquisition of mitochondria by a host with chimaeric prokaryotic
698 ancestry. *Nature* **531**, 101-104, doi:10.1038/nature16941
699 10.1038/nature16941. Epub 2016 Feb 3. (2016).

700 58 Roger, A. J., Munoz-Gomez, S. A. & Kamikawa, R. The Origin and Diversification of Mitochondria.
701 *Curr Biol* **27**, R1177-R1192, doi:10.1016/j.cub.2017.09.015
702 10.1016/j.cub.2017.09.015. (2017).

703 59 Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat*
704 *Methods* **12**, 59-60, doi:10.1038/nmeth.3176
705 10.1038/nmeth.3176. Epub 2014 Nov 17. (2015).

706 60 Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Archaeal Clusters of Orthologous Genes (arCOGs): An
707 Update and Application for Analysis of Shared Features between Thermococcales, Methanococcales, and
708 Methanobacteriales. *Life (Basel)* **5**, 818-840, doi:10.3390/life5010818
709 10.3390/life5010818. (2015).

710 61 Saier Jr, M. H., Tran, C. V. & Barabote, R. D. TCDB: the Transporter Classification Database for
711 membrane transport protein analyses and information. *Nucleic Acids Research* **34**, D181-186,
712 doi:10.1093/nar/gkj001 (2006).

713 62 Sondergaard, D., Pedersen, C. N. & Greening, C. HydDB: A web tool for hydrogenase classification
714 and analysis. *Sci Rep* **6**, 34212-34212, doi:10.1038/srep34212
715 10.1038/srep34212. (2016).

716 63 Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG) and a
717 metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* **35**, 725-731,
718 doi:10.1038/nbt.3893 (2017).

719 64 Yin, Y. *et al.* dbCAN: a web resource for automated carbohydrate-active enzyme annotation.
720 *Nucleic Acids Research* **40**, W445-451, doi:10.1093/nar/gks479
721 10.1093/nar/gks479. Epub 2012 May 29. (2012).

722 65 Rawlings, N. D., Barrett, A. J. & Finn, R. Twenty years of the MEROPS database of proteolytic
723 enzymes, their substrates and inhibitors. *Nucleic Acids Research* **44**, D343-350, doi:10.1093/nar/gkv1118

10.1093/nar/gkv1118. Epub 2015 Nov 2. (2016).

66 Lenfant, N. *et al.* ESTHER, the database of the alpha/beta-hydrolase fold superfamily of proteins: tools to explore diversity of functions. *Nucleic Acids Research* **41**, D423-429, doi:10.1093/nar/gks1154
10.1093/nar/gks1154. Epub 2012 Nov 27. (2013).

67 Yu, N. Y. *et al.* PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**, 1608-1615, doi:10.1093/bioinformatics/btq249
10.1093/bioinformatics/btq249. Epub 2010 May 13. (2010).

68 Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**, 3059-3066 (2002).

69 Criscuolo, A. & Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* **10**, 210-210, doi:10.1186/1471-2148-10-210
10.1186/1471-2148-10-210. (2010).

70 Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268-274, doi:10.1093/molbev/msu300 (2015).

71 Wang, H. C., Minh, B. Q., Susko, E. & Roger, A. J. Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. *Syst Biol* **67**, 216-235, doi:10.1093/sysbio/syx068 (2018).

72 Kamikawa, R. *et al.* Parallel re-modeling of EF-1alpha function: divergent EF-1alpha genes co-occur with EFL genes in diverse distantly related eukaryotes. *BMC Evolutionary Biology* **13**, 131-131, doi:10.1186/1471-2148-13-131
10.1186/1471-2148-13-131. (2013).

73 Lartillot, N., Rodrigue, N., Stubbs, D. & Richer, J. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst Biol* **62**, 611-615, doi:10.1093/sysbio/syt022
10.1093/sysbio/syt022. Epub 2013 Apr 5. (2013).

74 Susko, E. & Roger, A. J. On reduced amino acid alphabets for phylogenetic inference. *Molecular Biology and Evolution* **24**, 2139-2150, doi:10.1093/molbev/msm144 (2007).

75 Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152, doi:10.1093/bioinformatics/bts565
10.1093/bioinformatics/bts565. Epub 2012 Oct 11. (2012).

76 Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput Biol* **7**, e1002195-e1002195, doi:10.1371/journal.pcbi.1002195
10.1371/journal.pcbi.1002195. Epub 2011 Oct 20. (2011).

77 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490-e9490, doi:10.1371/journal.pone.0009490 10.1371/journal.pone.0009490. (2010).

78 Capella-Gutierrez, S. *et al.* trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973, doi:10.1093/bioinformatics/btp348 (2009).

79 Minh, B. Q., Nguyen, M. A. & von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* **30**, 1188-1195, doi:10.1093/molbev/mst024 10.1093/molbev/mst024. Epub 2013 Feb 15. (2013).

80 Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59**, 307-321, doi:10.1093/sysbio/syq010 (2010).

81 Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun* **7**, 13219-13219, doi:10.1038/ncomms13219 10.1038/ncomms13219. (2016).

82 Wrighton, K. C. *et al.* RubisCO of a nucleoside pathway known from Archaea is found in diverse uncultivated phyla in bacteria. *The ISME Journal* **10**, 2702-2714, doi:10.1038/ismej.2016.53 10.1038/ismej.2016.53. Epub 2016 May 3. (2016).

83 Swigonova, Z., Mohsen, A. W. & Vockley, J. Acyl-CoA dehydrogenases: Dynamic history of protein family evolution. *J Mol Evol* **69**, 176-193, doi:10.1007/s00239-009-9263-0 10.1007/s00239-009-9263-0. Epub 2009 Jul 29. (2009).

84 Dibrova, D. V., Galperin, M. Y. & Mulkidjanian, A. Y. Phylogenomic reconstruction of archaeal fatty acid metabolism. *Environ Microbiol* **16**, 907-918, doi:10.1111/1462-2920.12359 (2014).

85 Hug, L. A. *et al.* Overview of organohalide-respiring bacteria and a proposal for a classification system for reductive dehalogenases. *Philos Trans R Soc Lond B Biol Sci* **368**, 20120322-20120322, doi:10.1098/rstb.2012.0322 10.1098/rstb.2012.0322. Print 2013 Apr 19. (2013).

86 Jugder, B. E., Ertan, H., Lee, M., Manefield, M. & Marquis, C. P. Reductive Dehalogenases Come of Age in Biological Destruction of Organohalides. *Trends Biotechnol* **33**, 595-610, doi:10.1016/j.tibtech.2015.07.004 10.1016/j.tibtech.2015.07.004. (2015).

87 Neumann, A., Wohlfarth, G. & Diekert, G. Tetrachloroethene dehalogenase from Dehalospirillum multivorans: cloning, sequencing of the encoding genes, and expression of the pceA gene in Escherichia coli. *J Bacteriol* **180**, 4140-4145 (1998).

88 Vignais, P. M. & Billoud, B. Occurrence, classification, and biological function of hydrogenases: an overview. *Chem Rev* **107**, 4206-4272, doi:10.1021/cr050196r (2007).

796 89 Rochette, N. C., Brochier-Armanet, C. & Gouy, M. Phylogenomic test of the hypotheses for the
797 evolutionary origin of eukaryotes. *Mol Biol Evol* **31**, 832-845, doi:10.1093/molbev/mst272
798 10.1093/molbev/mst272. Epub 2014 Jan 7. (2014).
799

800

801

Figures

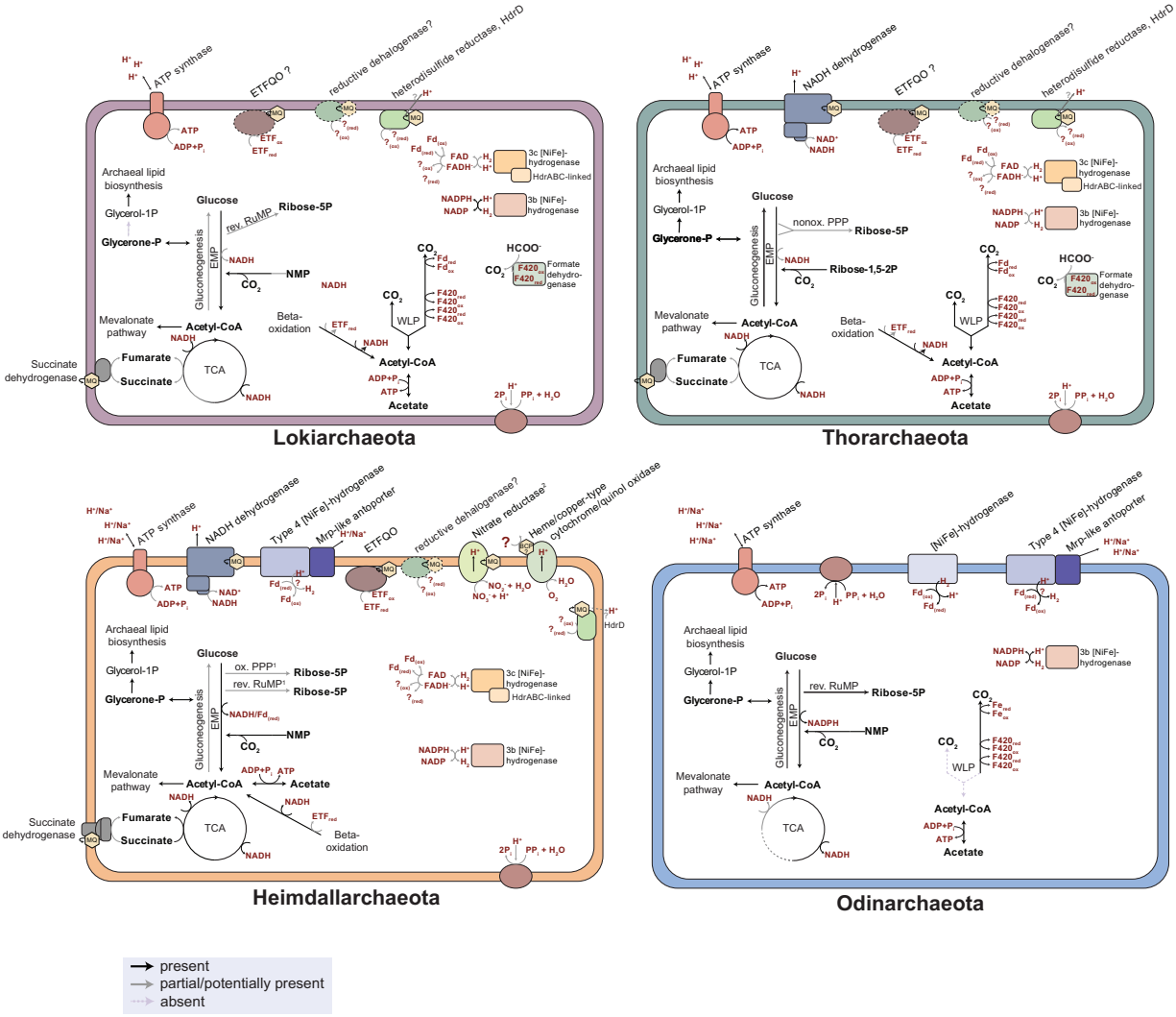
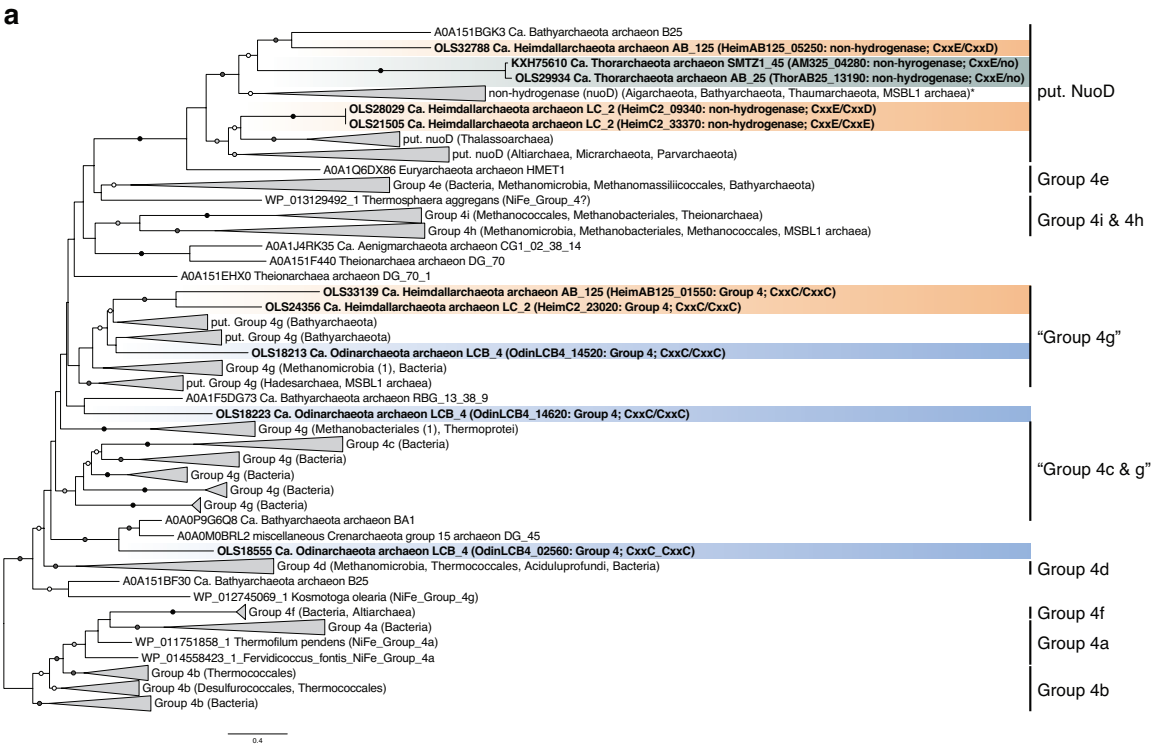


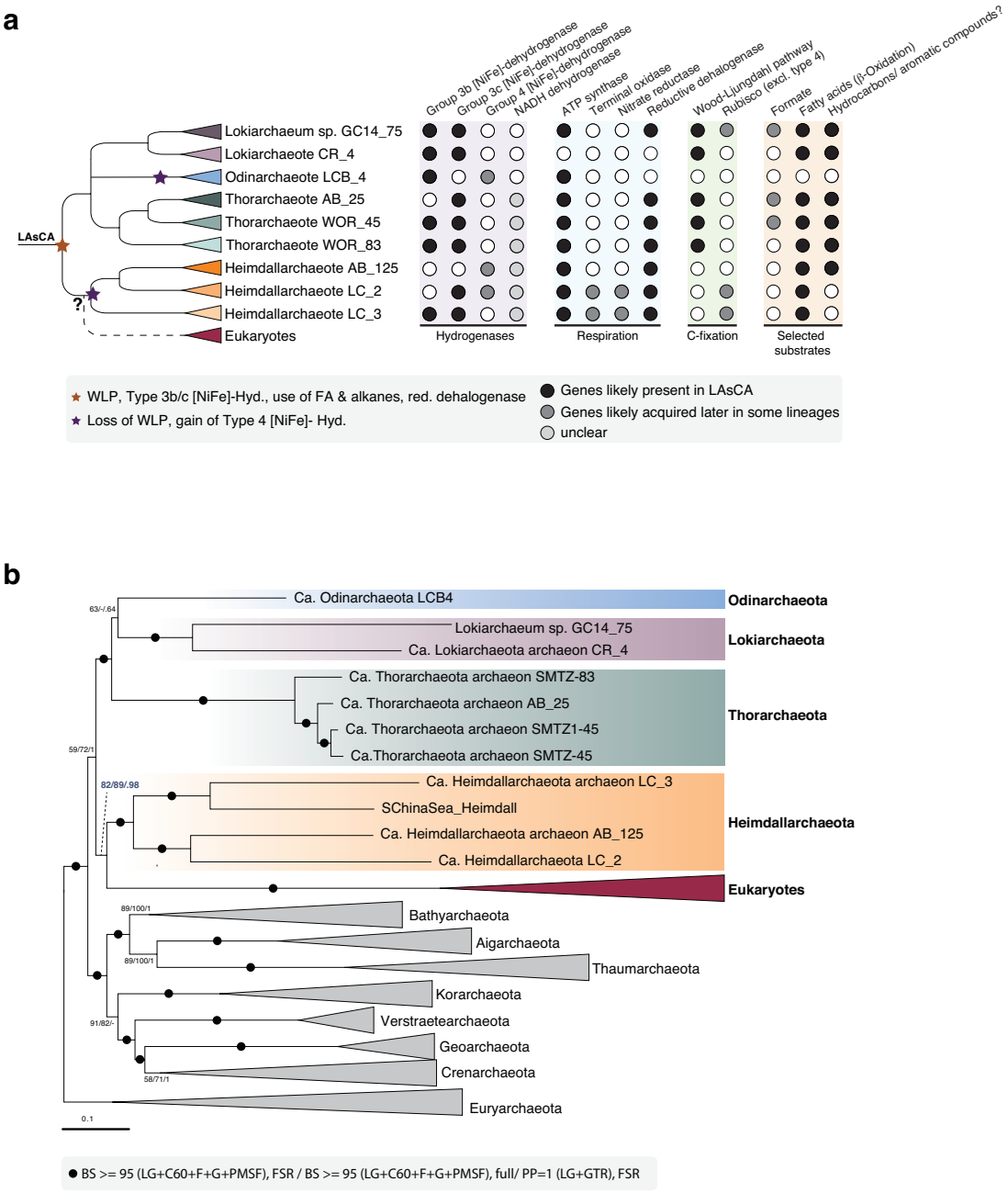
Figure 1. Metabolic potential of different Asgard phyla. Each of the arrows represent functions that were assigned to predicted proteins encoded in the respective genomes (Lokiarchaeota (2 MAGs), Thorarchaeota (3 MAGs), Heimdallarchaeota (3 MAGs), Odinarchaeota (1MAG)). The shading of the arrows depends on the distribution of a specific function across the different members of a groups as indicated in the figure panel. Notes: [1]oxidative pentose phosphate pathway (PPP) in *Heimdallarchaeum* LC2 and LC3 and reductive ribulose monophosphate (RuMP) pathway in *Heimdallarchaeum* AB125; [2]location of active site of nitrate reductase in Asgard homologs likely in the cytoplasm (Suppl.

Information). Color code as indicated in the figure inlet is as follows: arrows in black indicate enzymatic steps encoded by the respective organisms; grey arrows indicate enzymatic steps for which either only putative enzymes could be identified or enzymatic steps that are only encoded in some of the representatives of a given phylum; dashed grey arrows reveal enzymatic steps for which no (candidate) enzymes could be found in any of the representatives of a particular phylum. The presence/function of enzyme complexes surrounded by black dashed lines is tentative.


a
A0A151BGK3 Ca. Bathyarchaeota archaeon B25
OLS32788 Ca. Heimdallarchaeota archaeon AB_125 (HeimAB125_05250: non-hydrogenase; CxxE/CxxD)
KXH75610 Ca. Thorarchaeota archaeon SMTZ1_45 (AM325_04280: non-hyrogenase; CxxE/no)
OLS29934 Ca. Thorarchaeota archaeon AB_25 (ThorAB25_13190: non-hydrogenase; CxxE/no)
non-hydrogenase (nuoD) (Aigarchaeota, Bathyarchaeota, Thaumarchaeota, MSBL1 archaea)*
OLS28029 Ca. Heimdallarchaeota archaeon LC_2 (HeimC2_09340: non-hydrogenase; CxxE/CxxD)
OLS21505 Ca. Heimdallarchaeota archaeon LC_2 (HeimC2_33370: non-hydrogenase; CxxE/CxxE)
put. nuoD (Thalassoarchaea)
put. nuoD (Altiarchaea, Micrarchaeota, Parvarchaeota)
put. NuoD
A0A1Q6DX86 Euryarchaeota archaeon HMET1
Group 4e (Bacteria, Methanomicrobia, Methanomassiliicoccales, Bathyarchaeota)
Group 4e
WP_013129492_1 Thermosphaera aggregans (NiFe_Group_4?)
Group 4i (Methanococcales, Methanobacteriales, Theionarchaea)
Group 4h (Methanomicrobia, Methanobacteriales, Methanococcales, MSBL1 archaea)
Group 4i & 4h
A0A1J4RK35 Ca. Aenigmarchaeota archaeon CG1_02_38_14
A0A151F440 Theionarchaea archaeon DG_70
A0A151EHX0 Theionarchaea archaeon DG_70_1
OLS33139 Ca. Heimdallarchaeota archaeon AB_125 (HeimAB125_01550: Group 4; CxxC/CxxC)
OLS24356 Ca. Heimdallarchaeota archaeon LC_2 (HeimC2_23020: Group 4; CxxC/CxxC)
put. Group 4g (Bathyarchaeota)
put. Group 4g (Bathyarchaeota)
OLS18213 Ca. Odinarchaeota archaeon LCB_4 (OdinLCB4_14520: Group 4; CxxC/CxxC)
Group 4g (Methanomicrobia (1), Bacteria)
put. Group 4g (Hadesarchaea, MSBL1 archaea)
"Group 4g"
A0A1F5DG73 Ca. Bathyarchaeota archaeon RBG_13_38_9
OLS18223 Ca. Odinarchaeota archaeon LCB_4 (OdinLCB4_14620: Group 4; CxxC/CxxC)
Group 4g (Methanobacteriales (1), Thermoprotei)
Group 4c (Bacteria)
Group 4g (Bacteria)
Group 4g (Bacteria)
Group 4g (Bacteria)
Group 4g (Bacteria)
"Group 4c & g"
A0A0P9G6Q8 Ca. Bathyarchaeota archaeon BA1
A0A0M0BRL2 miscellaneous Crenarchaeota group 15 archaeon DG_45
OLS18555 Ca. Odinarchaeota archaeon LCB_4 (OdinLCB4_02560: Group 4; CxxC_CxxC)
Group 4d (Methanomicrobia, Thermococcales, Aciduluprofundi, Bacteria)
Group 4d
A0A151BF30 Ca. Bathyarchaeota archaeon B25
WP_012745069_1 Kosmotoga olearia (NiFe_Group_4g)
Group 4f (Bacteria, Altiarchaea)
Group 4f
Group 4a (Bacteria)
WP_011751858_1 Thermofilum pendens (NiFe_Group_4a)
WP_014558423_1_Fervidicoccus_fontis_NiFe_Group_4a
Group 4a
Group 4b (Thermococcales)
Group 4b (Desulfurococcales, Thermococcales)
Group 4b (Bacteria)
Group 4b
0.4
b
Group 3d (Bacteria)
KKK40681 Lokiarchaeum sp. GC14_75 (Lokiarch_49310: put. Group 3c; CxxC/CxxC)[1]
Group 3d
Group 3c (Bathyarchaeota)
OLS27265 Ca. Heimdallarchaeota archaeon LC_3 (HeimC3_04360: Group 3c; CxxC/CxxC)
Group 3c (Theionarchaea)
Group 3c (Bacteria, Euryarchaeota)
Group 3c (Archaeoglobus, Korarchaeota, Bacteria)
Group 3c (Anaeromyxobacter)
OLS24369 Ca. Heimdallarchaeota archaeon LC_2 (HeimC2_23150: Group 3c; CxxC/CxxC)
Group 3c (MSBL1 archaea)
Group 3c (Bacteria)
WP_015233716_1 Natronobacterium gregoryi (NiFe_Group_3c)
Group 3c (Bacteria)
Group 3c
A0A147JTG5 Hadesarchaea archaeon DG_33_1
Group 3c (Methanocella, MSBL1 archaea, Methanobacteriales, Methanococcales, Methanopyrales)
Group 3c (Methanobacteriales, Methanocellales, Methanomicrobiales)
Group 3c (Bacteria)
Group 3c (Methanococcales)
KXH73312 Ca. Thorarchaeota archaeon SMTZ1_45 (AM325_14150: Group 3c; CxxC/no)
OLS21753 Ca. Thorarchaeota archaeon AB_25 (ThorAB25_26490: Group 3c; CxxC/no)
KXH71847 Ca. Thorarchaeota archaeon SMTZ1_83 (AM324_08070: Group 3c; CxxC/no)
OLS29367 Ca. Thorarchaeota archaeon AB_25 (ThorAB25_16580: Group 3c; CxxC/no)
KKK41105 Lokiarchaeum sp GC14_75 (Lokiarch_45490: Group 3c; no/no*)
Group 3c
Group 3b (Bacteria, Caldiarchaeum, Fervidicoccus)
Group 3b (Bacteria, Bathyarchaeota)
Group 3b (Aciduliprofundum, MSBL1 archaea, Hadesarchaea, DPANN)
Group 3b (Bacteria, MSBL1 archaea, Bathyarchaeota, Theionarchaea)
Group 3b (Bacteria)
OLS27427 Ca. Heimdallarchaeota archaeon LC_3 (HeimC3_04260: Group 3b; CxxC/CxxC)
Group 3b (Thermococcales)
KXH74491 Ca. Thorarchaeota archaeon SMTZ1_45 (AM325_05810: Group 3b; CxxC/CxxC)
KXH73819 Ca. Thorarchaeota archaeon SMTZ1_83 (AM324_16110: Group 3b; CxxC/CxxC)
KKK42945 Lokiarchaeum sp GC14_75 (Lokiarch_31190: Group 3b; no/CxxC*)
Group 3b (Thermococcales)
Group 3b (Bathyarchaeota)
OLS17582 Ca. Odinarchaeota archaeon LCB_4 (OdinLCB4_08000: Group 3b; CxxC/CxxC)
WP_012308976_1 Korarchaeum cryptofilum (NiFe_Group_3b)
Group 3b (Bacteria)
Group 3b (Bacteria)
Group 3b
OLS14703 Ca. Lokiarchaeota archaeon CR_4 (RBG13Loki_1669: Group 3; CxxC/CxxC)[2]
A0A0M0BR92 miscellaneous Crenarchaeota group 15 archaeon DG_45
Group 3b (Bathyarchaeota)
A0A1Q6DUI2 Euryarchaeota archaeon HMET1
Group 3a
Group 3a
0.7
•100/100
•90-99.9/100 or 100/90-99
•90-99.9/90-99
•80-89.9/80-100 or 80-100/80-89

818

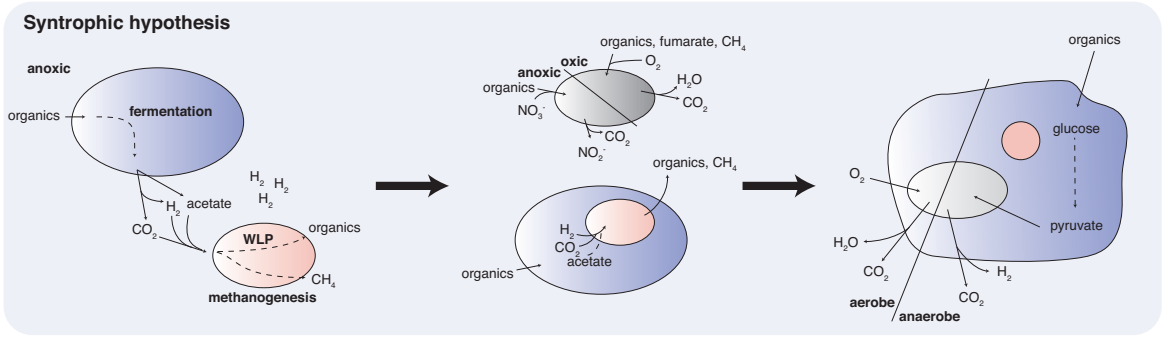
819 **Figure 2. Complex evolutionary history of Group 3 and Group 4 [NiFe]-hydrogenases in Asgard archaea.**
820 Maximum likelihood phylogenetic reconstructions of the large subunit of Group 4 (376 sequences) (a) and
821 Group 3 (543 sequences) (b) [NiFe]-hydrogenases (359 and 319 amino acids, respectively). Asterisk
822 indicate partial genes. Asgard homologs are highlighted in boldface and their annotation includes
823 information as to whether the N- and C-terminal CxxC motifs, which ligate H_2-binding metal centres[88], are
824 conserved. Numbers at branches indicate SH-like approximate likelihood ratio test[80] and ultrafast
825 bootstrap support values[79], respectively. Only bootstrap values above 80 are shown. The phylogeny of
826 large subunit [NiFe]-hydrogenases, group 4 (a) was rooted arbitrarily due to unstable position of the root.
827 The phylogeny of large subunit [NiFe]-hydrogenase Group 3 (b) was rooted between group 3a and all
828 other groups, as determined from a preliminary phylogenetic analysis of large subunit [NiFe]-
829 hydrogenases including all members (group 1-4) (not shown). Scale bars indicate the average number of
830 substitutions per site. [1]Gene cluster analysis of this divergent group 3 [NiFe]-hydrogenase homolog
831 suggests that it represents a Group 3c member. [2] While phylogenetic analysis indicates this this sequence
832 is from a group 3b [NiFe]-hydrogenase, it belongs to a gene cluster with an organization typical of Group
833 3c [NiFe]-hydrogenases. Thus, the annotation of this hydrogenase is currently unclear.
834
835



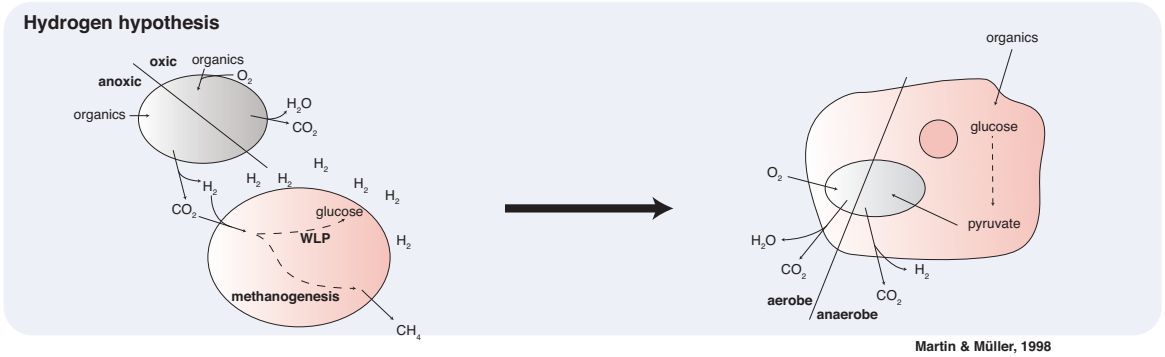
836

837 **Figure 3. Evolution of the Asgard superphylum and selected metabolic features**. Schematic
838 representation of the relationship of the Asgard phyla (9 MAGs) and the distribution of selected metabolic
839 features across the different members of this group (a). Maximum-likelihood phylogenetic analysis of a
840 concatenated set of 56 universal protein markers[5] from 144 representative taxa
841 https://figshare.com/s/5f153d1dcacadd3b3ed6 (b). Numbers at branches indicate bootstrap statistical

support estimated by IQ-TREE (LG+C60+F+G+PMSF model) obtained from 100 replicates on the 'full' and 'fast-sites removed' datasets, and posterior probabilities estimated by Phylobayes (LG-GTR model) on the 'fast-sites removed' dataset (see methods for further detail).

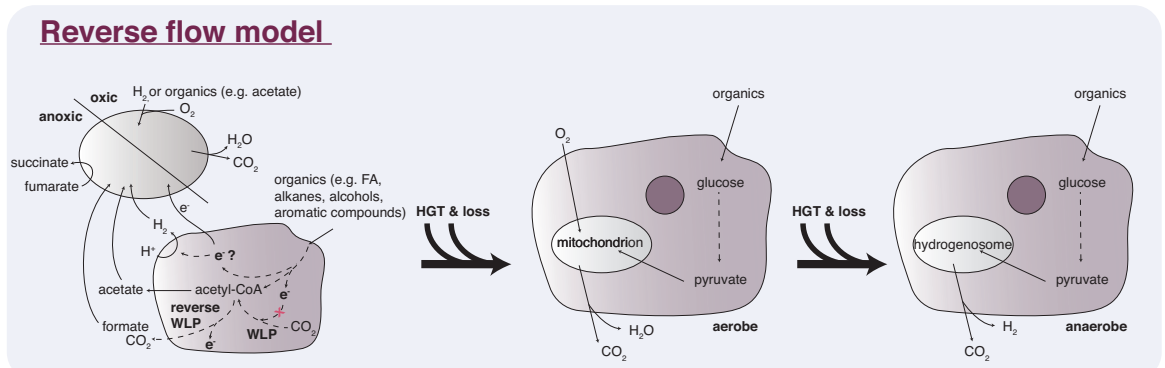


Moreira & Lopez-Garcia, 1998; Lopez-Garcia & Moreira, 2006



Martin & Müller, 1998

Sousa *et al.*, 2016



facultative aerobic alphaproteobacterium

fermentative deltaproteobacterium

autotrophic methanogenic archaeon

autotrophic lokiarchaeum-like archaeon

heterotrophic asgard-like archaeon

850

Figure 4. Evolutionary scenarios for the origin of the eukaryotic cell

(a) Depiction of the syntrophic hypothesis previously proposed by Moreira and Lopez-Garcia, 1998 and 2006[39,40], which invokes two bacterial and one archaeal partner(s) in the origin of the eukaryotic cell; i.e. first a syntrophic relationship was established between a fermentative deltaproteobacterium and a hydrogen-dependent archaeal methanogen, which was incorporated into the cytoplasm of the bacterium through endosymbiosis. Subsequently, a second endosymbiosis event led to the uptake of a facultative aerobic alphaproteobacterium, which was suggested to have oxidized organic compounds and hydrocarbons produced by the host. While this model can explain the origin of the nucleus from the archaeal endosymbiont, it currently lacks support from genomic and phylogenomic analyses[57,89]. (b) depiction of the hydrogen hypothesis originally proposed by Martin and Müller, 1998[41], which suggest that a symbiosis between a strictly autotrophic hydrogen-dependent methanogenic archaeaon and an H_2- and CO_2-producing alphaproteobacterium led to the origin of the eukaryotic cell. (c) depiction of the updated hydrogen hypothesis based on the first analysis of the metabolic repertoire of *Lokiarchaeum*[13], which suggests that the archaeal host was an autotrophic hydrogen-dependent acetogen rather than a methanogen. (d) Our syntrophy model referred to as the 'reverse flow model', which is based on our comparative analysis of the metabolic repertoire encoded by the various members of the Asgard archaea[2,4]. This model suggests that a metabolic syntrophy between anaerobic ancestral Asgard archaea and facultative anaerobic alphaproteobacteriahas has provided the selective force for the establishment of a stable symbiotic interaction that has subsequently led to the origin of the eukaryotic cell,. In this scenario, the archaeal progenitor generated reducing equivalents during growth on small organic substrates (e.g. hydrocarbons and fatty acids) and the bacterial partner utilized these in form of H_2, small reduced inorganic or organic compounds or by direct electron transfer. We acknowledge that HGT from free-living bacteria played an important role in this process and contributed to the gene repertoire of the eukaryotic cell including for the evolution of mitochondria-derived organelles such as hydrogenosomes in

anaerobic protists[58]. Please note, that our metabolic reconstructions do not provide insights into the origin of the nucleus and the relative timing of eukaryogenesis (see main text for details).

Supplementary Information for

Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analysis of Asgard archaeal metabolism

Anja Spang[1,2], Courtney Stairs[1], Nina Dombrowski[3], Laura Eme[1], Jonathan Lombard[1], Eva Fernández Cáceres[1], Chris Greening, Brett J. Baker[3] and Thijs J. G. Ettema[1]

Table of contents

A. Supplementary text

1. Carbon metabolism and the potential for autotrophic carbon fixation

Central Carbon metabolism. All members of the Asgard archaea contain candidate genes encoding enzymes for most steps of glycolysis via the Embden-Meyerhof-Parnas (EMP) pathway as well as gluconeogenesis and the tricarboxylic acid cycle (TCA; Suppl. Tables 1 and 2, Suppl. Figure 2). However, the specific enzymes catalyzing these steps differ among the various lineages (Suppl. Tables 1 and 2). For example, our analysis of enzymes potentially involved in glycolysis revealed that only three members of the Asgard archaea appear to encode enzymes belonging to described hexo- or glucokinase families: the archaeal ADP-dependent phosphofructokinase/glucokinase in the two Heimdallarchaeota LC3 and AB125 (PF04587) and a ROK (repressor, open reading frame, kinase) family enzyme in *Lokiarchaeum* GC14-75 (PF00480) (Suppl. Table 2), which could potentially phosphorylate glucose and thus perform the first step of the EMP pathway. While all genomes encode various proteins with Phosphofructokinase (Pfk) B family carbohydrate kinase (PF00294) or FGGY family of carbohydrate kinase domains (PF02782)[1], it is currently unclear whether some of those could represent new kinase families acting on glucose as substrate. Additionally, *Odinarchaeum* lacks candidate genes for a phosphoglucose isomerase (EC 5.3.1.8/9) and phosphofructokinase (EC 2.7.1.11/ 2.7.1.146). It is also interesting to note that genes for the bifunctional archaeal fructose-1,6-bisphosphate aldolase/phosphatase (FBPA/FBPase), which has been suggested to represent an ancient carbon fixation enzyme in archaea[2], could not be identified in the genomes of Lokiarchaeota and two Heimdallarchaeota (while present in the other Asgard genomes). However, these organisms (except *Lokiarchaeum* CR4) could perhaps catalyze this reaction in two steps using a fructose-bisphosphate aldolase and an archaeal fructose-1,6-bisphosphatase (Suppl. Tables 1 and 2, Suppl. Figure 2).

The TCA is nearly complete in most Asgard genomes (Suppl. Figure 2, Suppl. Tables 1 and 2) except *Odinarchaeum* and perhaps *Lokiarchaeum* CR_4, which lack genes for all subunits of a succinyl-CoA synthetase (EC 6.2.1.5) and for a succinate dehydrogenase/fumarate reductase (EC 1.3.5.1/4). Several genomes lack genes for canonical malate dehydrogenases as well as a membrane-bound malate:quinone-oxidoreductase (IPR006231), which catalyze the oxidation of malate to oxaloacetate (Suppl. Figure 2). However, in those organisms, malate could be converted to pyruvate using malic enzyme (EC 1.1.1.38/40).

Interestingly, the gene repertoire coding for proteins involved in ribose sugar metabolism and the corresponding pathways vary substantially across Asgard lineages (Fig. 1, Suppl. Tables 1 and 2). While Lokiarchaeota (both lineages), *Heimdallarchaeum* AB125 and *Odinarchaeum* seem to use the ribulose-5-phosphate pathway, genomes of Thorarchaeota appear to encode genes for enzymes of the non-oxidative pentose phosphate (PP) pathway and *Heimdallarchaeum* LC2 and LC3 might employ the oxidative PP pathway (Fig. 1, Suppl. Tables 1 and 2). Notably, the *Odinarchaeaum* genome lacks genes for ribose 5-phosphate isomerase (both Type A (IPR004788) and B (IPR003500, IPR004785) enzymes), which reversibly converts D-ribulose 5-phosphate to D-ribose 5-phosphate. Since D-ribose 5-phosphate represents a precursor for nucleotide biosynthesis[3], a homologue of this enzyme, i.e. of

Type A, is encoded by most archaeal genomes independent of whether the respective organisms employ the PP- or ribulose monophosphate (RuMP)-pathway. Prospective genomic analyses of additional members of these archaea is needed to determine whether these patterns reflect different catabolic and anabolic abilities or are a result of genome bin incompleteness.

Wood-Ljungdahl pathway. In accordance with previous analyses[4-6], we could identify genes for all enzymes of the Wood-Ljungdahl pathway (WLP)[7,8] in Loki- and Thorarchaeota (Suppl. Figure 2). In particular, these genomes encode the archaeal methyl-branch and carbamoyl-branch as well as the key enzyme CO dehydrogenase/acetyl-CoA synthase (ACS/CODH), which is comprised of five subunits (CdhA-E) and links the two branches by synthesizing acetyl-CoA from carbon monoxide and methyl-H_4MPT via the methylated corrinoid Fe-S protein CH_3-Co(III)FeSP (Suppl. Figure 2, Suppl. Tables 1 and 2)[8]. In contrast, genomes of Heimdallarchaeota do not seem to encode any of the enzymes involved in this pathway. While the genome of the sole representative of the Odinarchaeota encodes genes for the five steps of the methyl-branch leading to the generation of methyl-H_4MPT, it appears to lack genes for all subunits of ACS/CODH (CdhABCDE). In methanogenic archaea, methyl-H_4MPT not only plays a role in carbon fixation but is also the key intermediate in methanogenesis[9]. As such, methyl-H_4MPT is converted to methyl-CoM and subsequently disproportionated to methane and CO_2 by the tetrahydromethanopterin S-methyltransferase complex (Mtr) and methyl-CoM reductase (Mcr), respectively. However, we could not detect genes coding for the eight subunits of the Mtr complex or the five subunits of Mcr in any of the investigated Asgard archaea (see below). This suggests that Loki- and Thorarchaeota are not able to perform methanogenesis, although the use of methylated substrates cannot be excluded (see below). Instead they may use the WLP for autotrophic carbon fixation and perhaps for energy conservation using inorganic (at least Thorarchaeota) and/or organic substrates (Fig. 1). In addition, the WLP could serve as electron sink and thereby allow energetically efficient fermentative growth on various organic substrates (see below) or function in reverse for the complete oxidation of organic substrates[7,10]. Given the lack of ACS/CODH as well as key enzymes of methanogenesis in Odinarchaeota, the functional role of the enzymes of the methyl-branch in this organism remain unknown. However, it should be noted that we can currently not exclude the possibility that the absence of the respective genes is due to genome incompleteness.

Genes encoding for proteins related to methanogenesis. Surprisingly, genomes of Loki-, Thor- and even Heimdallarchaeota encode various proteins assigned to the uroporphyrinogen decarboxylase (URO-D) protein family (IPR000257) (Suppl. Figure 2, Suppl. Tables 1 and 2), which includes candidate proteins for methylcobalamin:coM methyltransferases. Furthermore, each of these genomes encodes at least one protein assigned to the trimethylamine:corrinoid methyltransferase family and *Thorarchaeum* SMTZ1-45 in addition contains a gene coding for a putative monomethylamine methyltransferase. Methyltransferases belonging to these protein families allow several methanogenic archaea to grow on methylated compounds such as methyl sulfides, methylamines and methanol[11]. Notably, homologues of these enzymes encoded by genomes of methanogenic archaea generally

contain the amino acid pyrrolysine, which is thought to be essential for catalysis[12]. However, these methyltransferase families are extremely diverse and not restricted to organisms that encode pyrrolysine. For instance, a trimethylamine:corrinoid methyltransferase that does not contain pyrrolysine has been shown to function as glycine betaine methyltransferase in the gram-positive *Desulfitobacterium hafniense*[13]. Since Asgard archaea seem to lack genes coding for pyrrolysine, their putative trimethylamine:corrinoid methyltransferase may demethylate substrates other than trimethylamine. Methyl-groups could be transferred to corrinoid proteins, candidates for which (COG05012) are encoded by genomes several Asgard genomes (Suppl. Table 2). Corrinoid proteins are involved in methyl-transfer and act as substrate for the methylcobalamin:coenzyme M methyltransferase (IPR000257) leading to the formation of methyl-CoM in methanogens[11]. Lokiarchaeota and Thorarchaeota not only contain genes coding for various proteins assigned to IPR000257 (corrinoid proteins and methyltransferases), but their genomes also reveal the organization of these genes in gene clusters. This suggests that at least Loki- and Thorarchaeota have the ability to convert specific methylated compounds to methyl-CoM. In methanogenic archaea, methyl-CoM is further converted to methane by the key enzyme of methanogenesis: methyl-CoM reductase (Mcr)[11]. However, in contrast to archaea known to be able to use methylated compounds, all herein studied members of the Asgard archaea lack gene homologues coding for the five subunits of Mcr. Furthermore, their genomes do not encode all eight subunits of the membrane-bound tetrahydromethanopterin S-methyltransferase (MtrA-H), which performs a central energy-conserving and sodium-ion-translocating step during methanogenesis and links the methyl-branch of the WLP to methane generation from methyl-CoM. Yet, genomes of Loki-, Odin- and Thorarchaeota encode the methyltransferase subunit MtrH (Suppl. Table 2), which also occurs in various non-methanogenic archaeal and bacterial lineages[14] and has been suggested to also mediate the transfer of a methyl-group between methylcobalamin and tetrahydrofolate[13]. In contrast, in methanogens the methyl-group is transferred to tetrahydromethanopterin via the corrinoid prosthetic group of MtrA[15]. Interestingly, in the different Thorarchaeota, one to two genes coding for MtrH are located in gene clusters with genes encoding MtrA homologues as well as a protein of unknown function related to methanogenesis marker protein 14 (IPR008303). This gene cluster is reminiscent of the *mtr*XAH operon of *Methanosarcina barkeri*, which was suggested to play a role in the methyl-transfer from methylated substrates to tetrahydromethanopterin[16]. It therefore seems possible that at least in Thorarchaeota, MtrH and MtrA homologues may provide a link between Methyl-CoM and Methyl-H_4MPT and/or Methyl-H4F[a] and allow the use of methylated compounds by these archaea.

3-hydroxypropionate/4-hydroxybutyrate cycle. In addition to genes for the WLP, genomes of Loki- and Thorarchaeota but also Heimdallarchaeota encode several enzymes characteristic of the 3-hydroxypropionate/4-hydroxybutyrate (3H4H) carbon fixation pathway (Suppl. Table 1). However, in *Archaeoglobus lithotrophicus,* whose genome also encodes multiple carbon fixation pathways[17], these proteins are involved in the degradation of fatty

[a] Methyl-H4F represents a key intermediate of the methyl-branch in the bacterial-type WLP, genes of which were also identified in Asgard genomes and may encode enzymes functioning in folate biosynthesis (Suppl. Figure 2).

acids and hydrocarbons. It is likely that these enzymes perform similar functions in the Loki-, Thor- and Heimdallarchaeota. For instance, some of the genes encoding these enzymes co-occur with genes for β-oxidation related proteins and are absent in the genome of *Odinarchaeum*, which also lacks genes coding for β-oxidation-related proteins. Furthermore, in *Lokiarchaeum* GC14_75 and *Heimdallarchaeota* LC2 and LC3, some of these genes are encoded in the same region of the chromosome (Suppl. Table 3). For example, in *Lokiarchaeum* GC14_75, methylmalonyl-CoA mutase (COG2185 and COG1884) can be found within two genes of the short chain dehydrogenase gene (ACAD, COG1960, discussed below). Altogether, this supports the idea that these proteins could be involved in the degradation of fatty acids and hydrocarbons in several Asgard archaea.

Calvin-Benson-Bassham Cycle. Furthermore, genomes of all members of the Asgard archaea encode a protein belonging to the ribulose 1,5-bisphosphate carboxylase family (RuBisCO), the key carbon fixation enzyme of the Calvin-Benson-Bassham (CBB) cycle (Suppl. Figure 18a), which adds CO_2 to ribulose 1,5-bisphosphate (RuBP). Our phylogenetic analyses classify a homologue of each of the analysed Thorarchaeota and Lokiarchaeota as Type-IV Rubisco, which is not known to have canonical RuBisCO activity and may instead act on analogous compounds (Suppl. Figure 18a)[18]. In agreement with this, key residues of the active site[19] are not conserved (Suppl. Figure 18b) in most of these Type-IV sequences. However, genomes of *Heimdallarchaeum* AB125, *Lokiarchaeum* CR4, *Lokiarchaeum* GC14_75 and *Heimdallarchaeum* LC3 encode homologues that affiliate either with Type-III RuBisCO or lack clear affiliation, and retain all essential active site residues (Suppl. Figure 18b, except for some with one mutation that maintains the hydrophobicity of the respective amino acid side chain). Thus, those sequences may act as *bona fide* RuBisCO enzymes. However, further research is needed to elucidate the functional context in which these proteins operate. For example, it is possible that they are involved in an AMP salvage pathway[20] (Suppl. Tables 1 and 2), like in *Thermococcus kodakarensis*, rather than in carbon fixation. This is supported by the presence of candidate genes of the AMP salvage pathway in these genomes, i.e. genes coding for a putative AMP phosphorylase as well as a ribose-1,5-bisphosphate isomerase (Suppl. Tables 1 and 2) as well as by the lack of genes encoding additional key enzymes of the CBB and the related carbon fixation pathway referred to as the reductive hexulose-phosphate (RHP) pathway[21]. For example, proteins with the signature domain for phosphoribulokinase (IPR006082), another key enzyme of RuBisCO-based carbon fixation pathways, are not encoded by Asgard genomes. While, *Heimdallarchaeota* LC2 and LC3 do encode enzymes belonging the broader phosphoribulokinase/uridine kinase family (IPR006083; Suppl. Table 2), these seem to represent uridine kinases rather than phosphoribulokinases as they lack the signal domain of and are only distantly related to known phosphoribulokinases (e.g. the use of known phophoribulokinases of *Synechococcus elongatus* (Q31PL2) and *Methanospirillum hungatei* (Q2FUB5) as query against the heimdallarchaeal genomes did not recover a significant hit). Finally, the RuBisCO homologues of *Odinarchaeum* LCB_4 and *Heimdallarchaeum* LC2 have two mutations within the active site residues (Suppl. Figure 18b) and it is therefore unlikely that the corresponding enzymes have retained the canonical carboxylase activity.

2. Asgard archaea have the potential to use various substrates as carbon source including hydrocarbons

The genomes of all Asgard archaea reveal the potential for the degradation of various substrates including complex carbohydrates as evidenced by various carbon active enzymes, peptidases and esterases as well as amino acids (Suppl. Table 4). Also notable is the large variety of paralogues encoding key enzymes of the β-oxidation pathway in all Asgard archaea except for *Odinarchaeum*, which may enable these organisms to use fatty acids of varying chain lengths as carbon source. In addition, Loki- and Thorarchaeota genomes encode putative archaeal formate dehydrogenases (Suppl. Tables 1 and 2), suggesting that these organisms are able to also use formate as a carbon and electron source, enabling the reduction of cofactor F420. Reduced F420 could subsequently allow carbon fixation via the WLP (Fig. 1). Alternatively, the formate dehydrogenase could be used for the generation of formate from CO_2 during the complete oxidation of organic substrates such as fatty acids via the WLP. This may require growth with syntrophic formate-utilizing partner organisms, that would lower the concentration of formate[22]. Altogether, the potential of using organic substrates for growth is in agreement with a higher relative abundance of Lokiarchaeota (DSAG) in organic-rich sediment layers[23,24].

Carbohydrate active enzymes and peptidases. Each Asgard archaea genome has the genetic potential to degrade complex substrates and together they encode 532 carbohydrate-active enzymes (CAZYmes), 338 peptidases and 128 ESTHERs (ESTerases and alpha/beta-Hydrolase Enzymes and Relatives; Suppl. Table 4). Overall, the genome of *Odinarchaeum* LCB_4 encodes for the fewest CAZYmes (24), peptidases (14) and ESTHERs (1). In contrast, the genomes of Lokiarchaeota encoded for most CAZymes and ESTHERs (~80 and ~40, respectively) and Heimdall- and Thorarchaeota for most of the peptidases (~35). However, the potential to degrade complex substrates is similar among the different Asgard archaea, when compared to their overall proteome size (Suppl. Figure 4). Notably, the main difference is the relative low number of predicted esterases in *Odinarchaeum*, which is in agreement with the lack of β-oxidation genes encoded in its genome. Potential substrates for these enzymes include complex carbohydrates, such as starch and glycogen (GH13, GH57, GH133), cellulose (GH12, GH5) or arabinan (GH93). Additionally, β-galactosidases (GH2), β-glucosidases (GH3) and α-mannosidases (GH38) might be involved in the degradation of oligosaccharides. A subset of these CAZymes were predicted to be extracellular, suggesting the degradation of more complex carbohydrates outside of and subsequent uptake into the cell. For example, genomes of Heimdallarchaeota and Lokiarchaeota encode for potentially secreted α-amylases (GH57 detected in AB125, LC3 and GC14_75) that cleave sugars from starch[25]. Consistent with this finding, we detected carbohydrate transporters in their genomes, including a potential disaccharide transport system in Heimdallarchaeota and Lokiarchaeota (transporter classification database (TCDB) ID: 3.A.1.1.22, 3.A.1.1.30, respectively) that could import sugars cleaved by secreted α-amylases (Suppl. Table 3). This observation as well as the presence of genes for most of the glycolytic pathway in Asgard genomes suggests that these genomes are capable to transform more complex carbohydrates and simple sugars to acetyl-CoA.

Aside from carbohydrate-degradation, genomes of Asgard archaea encode for several peptidase families (Suppl. Table 4). Few of these peptidases are predicted to be secreted and those mostly belong to the alkaline serine peptidase family S08 (detected in most of the genomes, Suppl. Table 4), which is implicated in nutritional substrate degradation[26]. In agreement with this, we detected several peptide transport systems across Asgard genomes. These include oligopeptide (Opp, TCDB IDs 3.A.1.5.12, 3.A.1.5.15 and 3.A.1.5.29) and dipeptide transport systems (Dpp, TCDB ID 3.A.1.5.39) in most of the Asgard genomes. Notably, oligopeptide transporters of the Opp/Dpp family can also be involved in complex carbohydrate transport, such as transport of mannans, rhamnose or xylan in *Thermotoga maritima*[27]. While genes for the degradation of rhamnose or xylans were absent, we frequently detected genes encoding α-mannosidases (GH38) and Opp transporters in the same genomes but not genetically linked on the same contig (*Lokiarchaeum* CR_4, *Thorarchaeum* AB_25, *Thorarchaeum* SMTZ-45 and SMTZ1-45). Additionally, we detected several uptake mechanisms for amino acids, including the amino acid permease YhdG/CtrA (TCDB ID 2.A.3.6.1) in all genomes except *Odinarchae*um, a transport system for branched chain amino acid in Thorarchaeota and Heimdallarchaeota (TCDB IDs 3.A.1.4.1 and 3.A.1.4.10) and basic amino acids in Thorarchaeota (TCDB ID 3.A.1.3.27). Most genomes of Asgard members encode genes involved in the intracellular degradation of peptides include potential oligopeptidases (M03, M13, S09), dipeptidases (M19, M38, S51, T02), aminopeptidases (M24, M29, M42, S33) and carboxypeptidases (M20, M32, S12) (Suppl. Table 4). Our analyses indicate that the key genes encoding proteins for the degradation of various amino acids (e.g. glutamate, alanine, glycine etc.) seem to be present in all lineages (Suppl.Tables 1 and 2). Amino acid degradation products such as pyruvate and oxoglutarate can then enter the TCA and contribute to the generation of reducing equivalents. Finally, the genomes of *Odinarchaeum* as well as Lokiarchaeota encode a complete suite of enzymes of the urea cycle, which may tie into the arginine metabolism (Suppl. Tables 1 and 2).

Consistent with the presence of fatty acid degradation genes in most genomes, we also find several genes that for example code for proteins belonging to the hydrolase family 4, which includes the monoglyceride lipase/lysophospholipase subfamily. Members of this subfamily have a broad substrate range but generally degrade monoacylglycerides to free fatty acids, which could subsequently be fed into the β-oxidation pathway. Consistent with the absence of this pathway in *Odinarchae*um, we did not detect any potential lipases in this genome. Apart from lipid degradation, most genomes except *Odinarchaeum* also encode for haloalkane dehydrogenases, members of the carbohydrate esterase CE10 family (i.e. para-nitrobenzyl esterases) and the lipase family XIII, which can act on acyl chain esters of butyrate or valerate[28].

Fatty acids (FA). One of the characteristics that distinguish archaeal lipids from those of bacteria and eukaryotes is the presence of isoprenoid and not fatty acyl chains[29]. Over the past three decades, a few notable exceptions to this observation have been reported in the Halobacteria[30,31]. Indeed, surveys of archaeal genomes revealed components of both fatty acid synthesis and breakdown[32,33]. In eukaryotes and bacteria, fatty acids are synthesized or oxidized on two different carrier molecules, acyl-carrier protein (ACP) or CoA respectively. However, in most archaea, fatty

acids are likely synthesized and oxidized on CoA[33,34]. While the genomes of all members of the Asgard archaea encode canonical enzymes to synthesize archaeal lipids (with the exception of the lack of genes encoding glycerol-1-phosphate dehydrogenase in Lokiarchaeota (Fig. 1, Suppl. Figure 2)), we identified multiple homologues of fatty acid metabolism proteins in each of the Asgard genomes except Odinarchaeota. This suggests that most of the Asgard archaea are capable of fatty acid degradation and synthesis. Here, we describe the pathway with respect to fatty acid breakdown although we suspect this pathway could also function in fatty acid synthesis, which has been reported elsewhere[32,33] (Suppl. Figure 2).

Long chain fatty acid CoA ligase - COG0318. Fatty acids destined for degradation are first activated with CoA via long chain fatty acid CoA ligase (COG0318) to generate an acyl-CoA. This protein family includes a variety of CoA ligases (e.g., Acyl-CoA synthetase, 2-succinylbenzoate--CoA ligase, medium and long chain fatty acid CoA ligase). With the exception of Odinarchaeota, each of the Asgard genomes encode multiple paralogues of this protein (Suppl. Table 1 and 2). To determine if any of these paralogues could be involved in fatty acid metabolism, we reconstructed the evolutionary history of this protein using sequences retrieved with BLAST against the nr and swissprot databases (see Methods for details, Suppl. Figure 5). We observed that the proteins are distributed across prokaryotic and eukaryotic diversity in multiple paralogues, though poor backbone support makes it difficult to determine the exact function of these proteins. Interestingly, *Heimdallarchaeum* LC2 has a MenE-like homologue (OLS23766) that might be involved in quinone metabolism (discussed below).

Acyl-CoA dehydrogenase (ACAD) COG1960. The resulting acyl-CoA can be oxidized via either an acyl-CoA dehydrogenase (ACAD, COG1960) or 4-hydroxybutyryl-CoA dehydratase (COG2368) to generate an acyl-2-enoyl-CoA. Most organisms have multiple paralogues of ACAD enzymes that are specific for different chain-lengths of fatty acids[35]. We observed multiple copies of COG1960 in each of the non-Odinarchaeota Asgards suggesting these organisms might be capable of metabolizing fatty acids (Suppl. Table 5). To help determine the substrate of the various Asgard ACADs, we performed a large phylogenetic analysis of ACADs across the tree of life with an emphasis on those sequences where the substrate specificity is known as previously reported in[35] (Suppl. Figure 6). Since these gene families consisted of 1000s of members, we subdivided the dataset into three main clades based on a preliminary phylogenetic analysis (see Methods for details). When possible, sub-clade distinctions were made based on representative sequences reported in[35]. The number of distinct paralogues identified in each Asgard genome ranged from 1-6 as indicated by numbers within the circles on Suppl. Figure 2b Clade 1 (Suppl. Figure 6a) consists of two unclassified clades (clade 1a and 1b) with sequences from *Lokiarchaeum* CR4 and all the Thorarcheota and a sub-clade of medium chain acyl-CoA dehydrogenases (clade 1c; Suppl. Figure 6a). Within clade 2, *Heimdallarchaeum* LC2 and LC3 resolved in a clade sister to eukaryotic and bacterial ACAD10/11 sequences (Suppl. Figure 6b). The ACAD10 and 11 proteins in humans have been linked to branched chain (16C) and long-chain (20-26C) metabolism respectively[36], although the function of these ACADs remains largely unknown in other organisms. The Asgards

branch throughout the rest of clade 2, with the majority grouping with other archaeal sequences in the unclassified clade 2c and clade 2d (Suppl. Figure 6b). Clade 3 consists of mostly short chain (clade 3a and 3b) or branched chain (clade 3b and 3e) dehydrogenases (Suppl. Figure 6c). These clade 3 dehydrogenases were found only in the Lokiarchaeota and Heimdallarchaeota genomes but not in Thorarchaeota.

In general, the poor resolution of the tree and patchy distribution of the ACAD genes across the Asgards and Archaea makes it challenging to infer the evolutionary history of these proteins. Some Asgard ACAD sequences represent clear cases of recent events of horizontal gene transfer (HGT) from bacteria to archaea (e.g., Heimdallarchaeota OLS19783 and OLS26460, sub-clade 2a; Heimdallarchaeota OLS24605 sub-clade 2e; Heimdallarchaeota in sub-clade 3c, Heimdallarchaeota OLS26526 in sub-clade 3d; Suppl. Figure 6) while others proteins are found throughout archaea (e.g., sub-clades 2c and 2d), suggesting they could be conserved feature of the Asgards or HGTs from other archaea. The lack of studied representatives within many of the ACAD clades makes it difficult to predict the function of the Asgard sequences. However, we suspect that at least Lokiarchaeota and Heimdallarchaeota are capable of metabolizing short branched and unbranched acyl-CoA chains because they have homologues of clade 3a and 3b enzymes (Suppl. Figure 6c). Despite the absence of well-characterized bacterial or eukaryotic ACAD genes in archaea (e.g., medium chain ACAD, clade 1c; very Long chain ACAD, clade 2a), *Archaeoglobus fulgidus* is able to degrade long-chain fatty acids[37]. This suggests that there might be alternative paralogues that archaea use to metabolize fatty acids that are distinct from their bacterial and eukaryotic counterparts. Prime candidates for such paralogues could be those sequences found in clade 2c or 2d that are enriched in archaeal representatives.

ACAD alternative - aromatic ring hydroxylase (ARH) or 4-hydroxybutyryl-CoA dehydrogenase - COG2368. Dibrova and colleagues[33] recognized that ARH represents an analogue of the ACAD enzymes and only appeared in archaeal genomes with other fatty acid metabolism genes. This prediction is supported by previous studies which demonstrated that ARH has ACAD-like activity in *Clostridium*[38] suggesting that ARH could replace ACAD in fatty acid biosynthesis. We analyzed the evolutionary history of this protein across the tree of life and found that, in general, this gene is rare in archaea (Suppl. Figure 7). Many of the archaeal sequences, including those encoded by genomes of Loki- and Thorarchaeota, cluster together in a divergent clade composed of Cren- and Euryarchaeota suggesting these archaeal paralogues could share a common origin. However, the remainder of the Asgard sequences group mostly with bacteria, suggesting these sequences derive from bacteria-Asgard HGT events.

Enoyl-CoA hydratase - ECH - COG1024. The acyl-2-enoyl-CoA produced by ACAD or the ACAD alternative enzymes discussed above is subsequently hydrated to 3-hydroxyacyl-CoA by the action of enoyl-CoA hydratase (ECH; COG1024). We observed many copies of COG1024 in the Loki-, Thor- and Heimdallarchaeota genomes, and as many as 23 copies in Lokiarchaeum (Suppl. Table 5). Of these 23 Lokiarchaeum ECH homologues, seven were excluded as they were very similar to another copy (due to strain heterogeneity[39]) or too short for further analysis. Like ACAD,

ECH are part of a large superfamily. To refine our inferences, we therefore split the dataset into two distinct clades, ECH1 and ECH2 (Suppl. Figure 8 and 9). In general, support along the backbone of both clade 1 and clade 2 phylogenies is poor. We observed that this gene is present in multiple paralogous copies across prokaryotes and eukaryotes. Each non-Odinarchaeota Asgard genome encoded at least 1 paralogue of ECH as indicated by numbers within the circles on Suppl. Figure 2b. As previously suggested by Dibrova and colleagues, the origin of this gene in Archaea is best explained by multiple independent HGT events from bacteria. For example, Asgard sequences that branch within a clade of only bacteria rather than grouping with sequences of other members of the Asgard archaea likely represent more recent transfer events (e.g., clade 1, OLS26255 *Heimdallarchaeum* LC 3; clade 2, KKK42140, OLS14134, KKK40657, Lokiarchaeota).

3-hydroxyacyl-CoA dehydrogenase - HDH - COG1250. The 3-hydroxyacyl-CoA generated by ECH is oxidized further by 3-hydroxyacyl-CoA dehydrogenase (HDH; COG1250) to generate 3-ketoacyl-CoA. We observed that most of the non-Odinarchaeota Asgards have at least one copy of this gene with Lokiarchaeota having as many as eight copies. Phylogenetic analyses revealed that there are multiple paralogues of this gene across Asgards as indicated by numbers within the circles in Suppl. Figure 2b. Most of the Asgard sequences branch with other archaea suggesting that these proteins could have an archaeal origin (Suppl. Figure 10). However, some of the Lokiarchaeota sequences branch with bacteria (e.g., KKK40272, KKK40637, KKK42255, and KKK41074) and thus could have been acquired via HGT events.

β*-ketothiolase - acetoacetyl-CoA acyltransferase - BKL/AACT - COG0183.* The final step of β-oxidation is catalyzed by β-ketothiolase (BKL, COG0183) which cleaves the 3-ketoacyl-CoA to form acetyl-CoA and an acyl-CoA molecule that is two carbons shorter. This family of enzymes includes the acetoacetyl-CoA transferase (AACT) enzymes which catalyze acetoacetyl-CoA formation from acetyl-CoA as the first step in the mevalonate pathway for the biosynthesis of isoprenoids in archaea. All of the Asgard genomes (except *Heimdallarchaeum* AB125) have at least one copy of BKL/AACT. Initial phylogenetic analysis of this superfamily of proteins revealed two distinct clades of sequences composed of mainly archaea (clade 1; Suppl. Figure 11) or bacteria (clade 2; Suppl. Figure 12). For ease of discussion, we labeled the subclades of clade 1 A-M however the corresponding clade identifiers first described by Dibrova et al. can be found in Suppl. Table 5. The Asgard sequences branch throughout clade 1 with other archaea. In general, the tree topology does not reflect a vertical inheritance pattern for many of the Asgard sequences. In fact, there are no clades that have a representative sequence from each of the Asgard phyla (i.e., Loki-, Thor-, Heimdall- and Odinarchaeota), suggesting that these genes have either been subject to multiple replacements via HGT or were differentially lost among the different Asgard phyla. Within clade 2, many of the Asgard homologues branch with bacteria (e.g., *Heimdallarchaeum* LC3 OLS23724; *Lokiarchaeum* CR4 OLS15559, OLS12682, OLS15498, OLS14603), suggesting they might derive from bacteria-Asgard HGT events. However, at least one clade includes seven representative sequences from Loki-, Thor-, and Heimdallarchaeota phyla (Suppl. Figure 11, bottom clade contains

Heimdallarchaeum LC3 sequence OLS26180) and branches with other archaea, and could therefore represent an ancient feature of the Asgard superphylum.

According to previous studies, some of the archaeal BKL/AACT sequences found in clade 1 (subclades 1B, 1C, 1D and 1H) are encoded next to genes related to mevalonate metabolism[33]. To expand on this observation, we searched the 15 genes upstream and downstream of the Asgard BKL/AACT for genes related to β-oxidation (COGs 0318, 1960, 2368, 1024, 1042 or 1250) and mevalonate metabolism (COG3425) (Suppl. Figure 13). We found that some of the Asgard BKL/AACT sequences in subclades 1A, 1D, 1H, 1J, 1K, and 1L are linked to β-oxidation associated genes while those in 1B, 1E, and 1L are linked to mevalonate associated genes (e.g., HMG-CoA synthase). Within clade 2, only two of the Lokiarchaeota sequences appear to be linked to β-oxidation genes (Suppl. Figure 12). The gene proximity of the BKL/AACT sequences to various elements of β-oxidation and mevalonate metabolism could be used as a proxy for determining in which pathway these proteins participate, although further functional characterization is needed.

A note on fatty acid biosynthesis. Bacteria and eukaryotes typically have different enzymes dedicated to β-oxidation and biosynthesis of fatty acids. For example, the steps catalyzed by ACAD and HDH are in fact catalyzed by members of the short-chain dehydrogenase protein family (COG1028) namely enoyl-ACP reductase and ketoacyl-ACP reductase. We detected over 150 genes within this protein family across the Asgard superphylum, some of which might be dedicated for fatty acid metabolism, however, these pathways have yet to be characterized in archaea (Suppl. Table 5). Another fundamental difference between β-oxidation and fatty acid biosynthesis is the use of CoA versus acyl-carrier protein (ACP). In general, ACPs are rare in archaea and likely represent a bacterial feature[33,34]. Notably, however, we identified homologues of ACP in *Heimdallarchaeum* LC3 and all members of Thorarchaeota (Suppl. Table 5), suggesting that these organisms might be capable of ACP-mediated fatty acid biosynthesis. In the other Asgards and archaea without ACPs, it is possible that the β-oxidation enzymes might be able to synthesize fatty acids using CoA as the carrier molecule as is the presumed mechanism in other archaea[33,34]. Indeed, recent studies using synthetic biology, have successfully engineered the β-oxidation pathways of *E. coli* and yeast to elongate fatty acids[40,41] suggesting that these pathways have the potential to function in both directions.

Hydrocarbons. Notably, all analyzed Loki- and Thorarchaeota as well as *Heimdallarchaeum* AB125 contain several genes coding for pyruvate formate lyase (Pfl) domain (IPR004184) proteins (Suppl. Tables 1 and 2). This protein family includes glycyl radical enzymes, some of which are involved in the anaerobic activation of hydrocarbons through the addition of fumarate[42]. The glycyl is activated by glycyl-activating enzymes (reviewed in[43]), the genes for which are present in all Asgard archaea and occur in a gene cluster with IPR004184-domain containing proteins in *Lokiarchaeum* CR4 and *Heimdallarchaeum* AB125. Our phylogenetic analyses of the Pfl domain proteins show that archaeal homologues including those of Asgard archaea do not represent *bona fide* Pfl (Suppl. Figure 14). In contrast, they affiliate with those proteins comprising known alkylsuccinate and benzylsuccinate synthases[44], choline

trimethylamine lyases[45] or 4-hydroxyphenlyacetate decarboxylases[46], some of which allow the activation of hydrocarbons by fumarate addition. However, the different archaeal lineages comprise various separate clusters in phylogenetic analyses indicating independent gene transfers from different bacterial sources without clear affiliation to specific characterized enzymes. Thus, the function of these enzymes in Asgard archaea is unresolved. Yet, it seems possible, that at least some of these protein homologues could function in the activation of different hydrocarbons. For instance, a Pfl domain protein of *Archaeoglobus fulgidus* related to a homologue of *Thorarchaeum* SMTZ1_83 (Suppl. Figure 14: A0A135VWA9, arrow) has been analyzed previously and was suggested to represent an alkylsuccinate synthase involved in the activation of C16-alkanes by fumarate addition[37]. Furthermore, it was shown that *Thermococcus sibiricus* has the potential to grow on an alkane likely using its PFL-homologue (Suppl. Figure 14, C6A251, arrow). Prospective analyses are required to test the hypothesis of growth on hydrocarbons in the respective Asgard archaea and to elucidate the exact pathway of hydrocarbon degradation in archaea, which may differ from bacteria. For example, in bacteria, malonyl-CoA decarboxylase (IPR007956, IPR035372) is critical for the alkane degradation pathway[47], though we could not detect this protein in archaea. However, several other proteins involved in alkane degradation in bacteria[47] belong to enzyme families that also comprise proteins involved in the 3-hydroxypropionate/4-hydroxybutyrate carbon fixation pathway in some members of the TACK archaea. Candidate genes coding for enzymes of these protein families are present in the genomes of the different Asgard lineages (except for Odinarchaeota; Suppl. Tables 1 and 2) but also in several members of the Archaeoglobales, in which they were suggested to enable alkane degradation[17].

3) Electron transport chains differ among the Asgard lineages

[Ni-Fe]-hydrogenases. All members of the Asgards lack [FeFe]-hydrogenases or the methanogen-specific [Fe]-hydrogenases. However, all Asgard genomes encode various [NiFe]-hydrogenases, which were annotated based on classification of the large subunit using HydDB[48], phylogenetic analyses (Suppl. Table 2, Fig. 2) and operon architecture (Suppl. Figure 3). In particular, the Asgard genomes encode [NiFe]-hydrogenases belonging to Group 3 and Group 4, yet lack representatives of Group 1 and Group 2 [NiFe]-hydrogenases. First of all, both the genomes of Thorarchaeota, *Heimdallarchaeum* LC2 and LC3 as well as of *Lokiarchaeum* GC14_75 encode genes for [NiFe]-Hydrogenases belonging to Group 3c, which are located in gene clusters encoding all or some components of heterodisulfide reductases (Suppl. Figure 3). In methanogens, hydrogenase (MvhADG)-heterodisulfide reductase (HdrABC) complexes mediate electron bifurcation, with electrons from H_2 oxidation being simultaneously transferred to ferredoxin and heterodisulfide[49]. However, the substrate for these enzymes outside the methanogens (which use CoM-S-S-CoB) remains to be determined. While, the Heimdallarchaeota (LC2 and LC3) large subunit 3c homologues have retained the C-XX-C motifs at the C-terminus, this motif is not conserved in most of the thor- and lokiarchaeal homologues. It thus needs to be confirmed that these Group 3c [NiFe]-hydrogenases can bind the [NiFe] centers required to mediate H_2 oxidation[49]. In addition, Loki-, Odin- and Thorarchaeota as well as *Heimdallarchaeum*

LC3 contain gene clusters encoding [NiFe]-hydrogenases belonging to Group 3b (Suppl. Figure 3), which are bidirectional and may either evolve H_2 using electrons from NADPH or recycle H_2 to produce NADPH as reported for members of the Thermococcales[50,51]. While, the Group 3b enzyme complexes of *Pyrococcus furiosus* were shown to also have sulfhydrogenase activity (i.e. to reduce elemental sulfur to H_2S) *in vitro*[52], this does not seem to be their physiological function ([51] and references therein).

Finally, genomes of Heimdall- and Odinarchaeota also encode various membrane-bound Group 4 [NiFe]-hydrogenases, which are only distantly related to known Group 4 subgroups. In fact, the addition of homologues from novel archaeal genomes to our phylogenetic analyses, revealed that Group 4 hydrogenases are even more diverse than anticipated earlier[53] and suggests that future efforts are needed to define novel subgroups. Currently, most of these archaeal Group 4 large subunit homologues, including those present in Asgard archaea, comprise novel clusters within the loosely defined respiratory H_2-evolving 4g subgroup (Fig. 2). While, the exact function is unconfirmed, it has been suggested that the previously described Group 4g hydrogenases can couple ferredoxin oxidation to proton reduction in a simple electron chain that can contribute to the membrane potential and enhance overall energy yield. In order to get more insights into the potential function of group 4 [NiFe]-hydrogenases of Asgard archaea, we investigated the arrangement of the corresponding gene clusters. This revealed that the arrangement of their gene clusters (Suppl. Figure 3) is highly variable, in agreement with the large sequence diversity of the key subunit of these Group 4 [NiFe]-hydrogenases. However, all group 4 [NiFe]-hydrogenases seem to be membrane-anchored and ferredoxin represents the most likely co-substrate. Furthermore, the gene cluster architecture suggests, that at least two Group 4 hydrogenases of Odinarchaeum (OdinLCB4_14620, OdinLCB4_02560) and the Group 4 hydrogenase of *Heimdallarchaeum* LC2 (HeimC2_23020) can mediate sodium/proton translocation similar to *Pyrococcus furiosus* [54] as indicated by the presence of NuoL-like (also Mrp antiporter-like) membrane subunits (Suppl. Figure 3). Currently, the gene cluster that encodes the group 4 [NiFe]-hydrogenase of *Heimdallarchaeum* AB125 does not include genes encoding NuoL-like subunits (Suppl. Figure 3). However, in this regard, it has to be noted that the contig on which this gene cluster is located starts with the hypD gene (OLS33135). Furthermore, *nuoL/M/N-like* genes (e.g. OLS33159, OLS33160, OLS29628) are encoded towards the end of two other genomic contigs of this organism. This indicates that the gene cluster displayed in Suppl. Figure 3 is currently incomplete as a result of genome assembly that has split the hydrogenase gene cluster into separate contigs. Thus and based on the similarity to the large subunit of the Group 4 Hydrogenases of *Heimdallarchaeum* LC2 (Figure 2), it is most likely that the Group 4 hydrogenase of *Heimdallarchaeum* AB125 is also ion conductive and thus involved in energy conservation.

Complex I and alternative NADH:quinone reductases. Only genomes of Heimdallarchaeota and Thorarchaeota seem to encode a multisubunit NADH dehydrogenase (Complex 1) (Suppl. Table S2) including a homolog of NuoD (Fig. 2). However, candidates for an alternative NADH:quinone reductase are present in Loki- and Thorarchaeota. Genomes of members of these phyla encode several proteins with nitronate monooxygenase

domains (PF03060). It was recently reported that this protein family includes a new class of NADH:quinone reductases (PA1024)[55]. Interestingly, at least one PF03060 domain protein of each Loki- and Thorarchaeota is much more similar to the NADH:quinone reductases of *Pseudomonas aeruginosa* PAO1 (PA1024) (ca. 40% AA identity) than to its nitronate monooxygenase (PA4202) (ca. 28% AA identity). This may suggest that PF03060 domain proteins in these Asgard archaea may represent novel types of quinone reductases and could comprise a NADH:quinone reductases, which could for example function in the re-oxidation of NAD(P)H generated during growth on organic substrates.

Reductive dehalogenases. Surprisingly, genomes of all Asgard archaea except for *Odinarchaeum* encode various proteins assigned to epoxyqueuosine reductases (COG01600), some of which have significant similarity to reductive dehalogenase domains (IPR028894) (Suppl. Table 2). In bacteria, reductive dehalogenases serve as terminal reductases enabling the degradation of organohalides[56] using H_2, different organic substrates (e.g. pyruvate and acetate) or even NADPH as electron donors[57]. Organohalides such as halogenated phenols, dioxins, biphenyls, and aliphatic hydrocarbons represent persistent environmental pollutants and bacteria able to engage in dehalogenation of these compounds are of great ecological importance. Reductive dehalogenases are corrinoid Fe-S containing proteins and homologous to epoxyqueuosine reductases (COG01600)[58]. Notably, our phylogenetic analyses of this protein family using a previously described backbone dataset[59] revealed that genomes of Loki-, Thor- and *Heimdallarchaeum* LC2 and LC3 as well as Theionarchaea each encode at least one putative *bona fide* reductive dehalogenase (Suppl. Figure 15, dark green) suggesting the ability to use organohalides as terminal electron acceptors. The function of other homologs belonging to the epoxyqueuosine reductase family and which lack IPR domains of canonical reductive dehalogenases (Suppl. Figure 15, light green box)[58] is currently unclear, yet it may be speculated that some could represent novel electron acceptors with yet unknown the substrates.

In organohalide-respiring bacteria, the reductive dehalogenase genes are generally encoded in an operon with a gene encoding a membrane anchoring protein. However, since homologues of these membrane anchors could not be detected in the Asgard archaea, it is difficult to determine if these reductive dehalogenases are membrane-associated or soluble enzymes. Therefore, it will be important to characterize some of these proteins biochemically to determine whether Asgard archaea can indeed engage in the degradation of specific pollutants and elucidate the substrate spectrum of these proteins. For example, in Loki- and Thorarchaeota, which appear to lack classical terminal reductases, these reductive dehalogenases could allow the re-oxidation of the quinols generated during β-oxidation and ETF.

Terminal oxidase and nitrate reductase in Heimdallarchaeota. Surprisingly, genomes of *Heimdallarchaeote* LC2 and LC3 encode terminal A-type heme-copper oxidases (Suppl. Figure 16a), which may allow the respective organisms to use oxygen as terminal electron acceptor. Furthermore, under anoxic conditions, the presence of a respiratory nitrate reductase (Suppl. Table 1) seems to enable the use of nitrate as terminal acceptor through anaerobic

phosphorylation. Notably, a closer investigation of the alpha subunits of the nitrate reductase (NarG) encoded by *Heimdallarchaeum* LC2 and LC3 genomes, revealed that they belong to the bacterial-type nitrate reductase family (IPR006468) and lack the TAT signal motif common in canonical archaeal-type nitrate reductases which are localized at the extracellular site of the cytoplasmic membrane. Thus, Nar of Heimdallarchaeota is likely located in the cytoplasm similar to the bacterial-type Nar[60-62]. Phylogenetic analyses of NarG revealed that the Heimdallarchaeota homologues form a monophyletic cluster with the bacterial-type nitrate reductase of *Methanoperedens* sp. BLZ1 (Suppl. Figure 16b) as well as with a homologue of *Methylomirabilis oxyfera*, *Nitrolancea hollandica*, *Nitrococcus mobilis* and a *Nitrobacter* sp. species. Most likely, the few archaeal genomes that encode a bacterial version of Nar (Suppl. Figure 16b) have acquired the corresponding genes independently from bacterial sources.

Other potential electron chain components. Besides the presence of terminal reductases and a Complex I in *Heimdallarchaeum* LC2 and LC3, the genomes of these organisms encode a membrane-bound succinate dehydrogenase/fumarate reductase (Complex II), as well as a putative electron transfer flavoprotein quinone:oxidoreductase complex (ETF-QO; composed of FixA, FixB, FixC and FixX), both of which may allow the reduction of quinone-species (Fig. 1; Suppl. Tables 1 and 2). While genomes of Loki- and Thorarchaeota also encode FixA and FixB homologues as well as candidate proteins for FixC (belonging to arCOG00570), the latter did not occur in a gene cluster with FixX homologues. It remains therefore unclear whether Loki- and Thorarchaeota have a complete membrane-associated ETF-QO. In this regard, it is notable that genomes of Loki- and Thorarchaeota encode a large amounts of homologues belonging to subunits of soluble heterodisulfide reductases comprised of HdrA, HdrB and HdrC as well as of the D subunit of membrane-bound heterodisulfide reductases (which also includes HdrE)[9]. However, similar to *Methanomassiliicoccales*[63], genomes of Loki- and Thorarchaeota do not seem to encode homologues of HdrE, and the functional role of their HdrD homologues remains to be determined. For example, some of the loki- and thorarchaeal HdrD family homologues (arCOC00333) may be involved in the re-oxidation of quinol species generated by ETF during growth on organic substrates (e.g. β-oxidation) (Fig. 1). It also remains to be determined whether the succinate dehydrogenases/fumarate reductases of Loki- and Thorarchaeota are membrane-associated and could represent another means for the re-oxidation of quinol species (Fig. 1). Furthermore, genomes of most Asgard archaea, except for *Odinarchaeoum*, encode putative multihaem cytochromes (IPR011031) and/or cytochrome c-552/DMSO reductase-like, haem-binding domains (IPR019020), yet it is unclear whether these proteins play a role in electron transfer in these organisms. In two Thorarchaeota, the multihaeme cytochromes encode additional signature domains for cytochrome c552 (IPR003321), which occurs in formate-dependent cytochrome C nitrite reductases and catalyzes the reduction of nitrite to ammonia (Suppl. Tables 1, and 2)[64]. It remains to be investigated whether these thorarchaeal proteins, which are only distantly related to known nitrite reductases, could have similar functions. Most notably, *Heimdallarchaeum* LC2 and LC3 genomes code for cytochrome b561 domain proteins as well as a multitude of cupredoxin domain proteins, the latter of which may

represent important electron carriers (besides quinones) in these organisms and may shuttle reducing equivalents to their terminal oxidases (Fig. 1, Suppl. Table 2).

Menaquinone biosynthesis. All Asgard archaea genomes contain several genes encoding proteins potentially involved in ubiquinone/menaquinone biosynthesis. The set of genes is most complete for Heimdallarchaeota, supporting the idea that quinone species are components of their electron chains (Suppl. Table 1, Fig. 1). However, *Heimdallarchaeum* LC2 is the only representative of the Asgard archaea whose genome encodes homologues for most enzymes of the menaquinone biosynthesis pathway (Suppl. Tables 1 and 2; including MenB, MenF, MenD, MenE and candidates for MenA and MenG). For example, one of the steps is catalyzed by a member of the acyl-CoA synthetase (AMP-forming)/AMP-acid ligase II family. While genomes of all Asgard archaea encode representatives of this enzyme family (mainly functioning in β-oxidation), the *Heimdallarchaeum* LC2 genome uniquely encodes a homologue, which branches within the MenE clade in phylogenetic analyses and is annotated as a 2-succinylbenzoate-CoA ligase (Suppl. Figure 5). In contrast, the biosynthesis of menaquinones is less clear in *Heimdallarchaeum* LC3 and AB125. While they lack homologues of most canonical menaquinone biosynthesis genes, these genomes encode a homologue of MqnC, an enzyme of an alternative menaquinone biosynthesis pathway[65]. However, even though some of the genes annotated as ubiquinone biosynthesis genes may function in this alternative pathway, *Heimdallarchaeum* LC3 and AB125 seem to lack some enzymes characteristic of this pathway and the exact steps leading to menaquinone are therefore unclear in these latter organisms. Genomes of Loki- and Thorarchaeota encode only few candidate proteins for the biosynthesis of quinone species and the presence of quinones in these organisms needs to be investigated in future studies.

B. References

1 Zhang, Y., Zagnitko, O., Rodionova, I., Osterman, A. & Godzik, A. The FGGY carbohydrate kinase family: insights into the evolution of functional specificities. *PLoS Comput Biol* **7**, e1002318-e1002318, doi:10.1371/journal.pcbi.1002318
10.1371/journal.pcbi.1002318. Epub 2011 Dec 22. (2011).

2 Say, R. F. & Fuchs, G. Fructose 1,6-bisphosphate aldolase/phosphatase may be an ancestral gluconeogenic enzyme. *Nature* **464**, 1077-1081, doi:10.1038/nature08884
10.1038/nature08884. Epub 2010 Mar 28. (2010).

3 Brasen, C., Esser, D., Rauch, B. & Siebers, B. Carbohydrate metabolism in Archaea: current insights into unusual enzymes and pathways and their regulation. *Microbiology and Molecular Biology Reviews* **78**, 89-175, doi:10.1128/MMBR.00041-13
10.1128/MMBR.00041-13. (2014).

4 Sousa, F. L., Neukirchen, S., Allen, J. F., Lane, N. & Martin, W. F. Lokiarchaeon is hydrogen dependent. *Nature Microbiology* **1**, doi:Artn 16034 10.1038/Nmicrobiol.2016.34 (2016).

5 Seitz, K. W., Lazar, C. S., Hinrichs, K. U., Teske, A. P. & Baker, B. J. Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction. *The ISME Journal*, doi:10.1038/ismej.2015.233 (2016).

6 Liu, Y. *et al.* Comparative genomic inference suggests mixotrophic lifestyle for Thorarchaeota. *ISME J* **12**, 1021-1031, doi:10.1038/s41396-018-0060-x (2018).

7 Ragsdale, S. W. & Pierce, E. Acetogenesis and the Wood-Ljungdahl pathway of CO(2) fixation. *Biochim Biophys Acta* **1784**, 1873-1898, doi:10.1016/j.bbapap.2008.08.012 (2008).

8 Adam, P. S., Borrel, G. & Gribaldo, S. Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proceedings of the National Academy of Sciences of the United States of America* **115**, E1166-E1173, doi:10.1073/pnas.1716667115
10.1073/pnas.1716667115. Epub 2018 Jan 22. (2018).

9 Thauer, R. K., Kaster, A. K., Seedorf, H., Buckel, W. & Hedderich, R. Methanogenic archaea: ecologically relevant differences in energy conservation. *Nat Rev Microbiol* **6**, 579-591, doi:10.1038/nrmicro1931
10.1038/nrmicro1931. Epub 2008 Jun 30. (2008).

10 Schuchmann, K. & Muller, V. Energetics and application of heterotrophy in acetogenic bacteria. *Appl Environ Microbiol*, doi:10.1128/AEM.00882-16 (2016).

11 Borrel, G., Adam, P. S. & Gribaldo, S. Methanogenesis and the Wood-Ljungdahl Pathway: An Ancient, Versatile, and Fragile Association. *Genome Biol Evol* **8**, 1706-1711, doi:10.1093/gbe/evw114 (2016).

12 Krzycki, J. A. Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine methyltransferases. *Curr Opin Chem Biol* **8**, 484-491, doi:10.1016/j.cbpa.2004.08.012
10.1016/j.cbpa.2004.08.012. (2004).

13 Ticak, T., Kountz, D. J., Girosky, K. E., Krzycki, J. A. & Ferguson Jr, D. J. A nonpyrrolysine member of the widely distributed trimethylamine methyltransferase family is a glycine betaine methyltransferase. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E4668-4676, doi:10.1073/pnas.1409642111
10.1073/pnas.1409642111. Epub 2014 Oct 13. (2014).

14 Wang, S., Chen, Y., Cao, Q. & Lou, H. Long-Lasting Gene Conversion Shapes the Convergent Evolution of the Critical Methanogenesis Genes. *G3 (Bethesda)* **5**, 2475-2486, doi:10.1534/g3.115.020180
10.1534/g3.115.020180. (2015).

15 Hippler, B. & Thauer, R. K. The energy conserving methyltetrahydromethanopterin:coenzyme M methyltransferase complex from methanogenic archaea: function of the subunit MtrH. *FEBS Lett* **449**, 165-168 (1999).
16 Harms, U. & Thauer, R. K. Identification of the active site histidine in the corrinoid protein MtrA of the energy-conserving methyltransferase complex from Methanobacterium thermoautotrophicum. *Eur J Biochem* **250**, 783-788 (1997).
17 Estelmann, S. *et al.* Carbon dioxide fixation in 'Archaeoglobus lithotrophicus': are there multiple autotrophic pathways? *Fems Microbiology Letters* **319**, 65-72, doi:10.1111/j.1574-6968.2011.02268.x
10.1111/j.1574-6968.2011.02268.x. Epub 2011 Apr 4. (2011).
18 Tabita, F. R., Satagopan, S., Hanson, T. E., Kreel, N. E. & Scott, S. S. Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *J Exp Bot* **59**, 1515-1524, doi:10.1093/jxb/erm361
10.1093/jxb/erm361. Epub 2008 Feb 16. (2008).
19 Solden, L., Lloyd, K. & Wrighton, K. The bright side of microbial dark matter: lessons learned from the uncultivated majority. *Curr Opin Microbiol* **31**, 217-226, doi:10.1016/j.mib.2016.04.020 (2016).
20 Sato, T., Atomi, H. & Imanaka, T. Archaeal type III RuBisCOs function in a pathway for AMP metabolism. *Science* **315**, 1003-1006, doi:10.1126/science.1135999
10.1126/science.1135999. (2007).
21 Kono, T. *et al.* A RuBisCO-mediated carbon metabolic pathway in methanogenic archaea. *Nat Commun* **8**, 14007-14007, doi:10.1038/ncomms14007
10.1038/ncomms14007. (2017).
22 Stams, A. J. & Plugge, C. M. Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nat Rev Microbiol* **7**, 568-577, doi:10.1038/nrmicro2166 (2009).
23 Jorgensen, S. L. *et al.* Correlating microbial community profiles with geochemical data in highly stratified sediments from the Arctic Mid-Ocean Ridge. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E2846-2855, doi:10.1073/pnas.1207574109 (2012).
24 Jorgensen, S. L., Thorseth, I. H., Pedersen, R. B., Baumberger, T. & Schleper, C. Quantitative and phylogenetic study of the Deep Sea Archaeal Group in sediments of the Arctic mid-ocean spreading ridge. *Front Microbiol* **4**, 299-299, doi:10.3389/fmicb.2013.00299 (2013).
25 Janecek, S. & Blesak, K. Sequence-structural features and evolutionary relationships of family GH57 alpha-amylases and their putative alpha-amylase-like homologues. *Protein J* **30**, 429-435, doi:10.1007/s10930-011-9348-7
10.1007/s10930-011-9348-7. (2011).
26 Rawlings, N. D., Barrett, A. J. & Finn, R. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research* **44**, D343-350, doi:10.1093/nar/gkv1118
10.1093/nar/gkv1118. Epub 2015 Nov 2. (2016).
27 Conners, S. B. *et al.* An expression-driven approach to the prediction of carbohydrate transport and utilization regulons in the hyperthermophilic bacterium Thermotoga maritima. *J Bacteriol* **187**, 7267-7282, doi:10.1128/JB.187.21.7267-7282.2005
10.1128/JB.187.21.7267-7282.2005. (2005).
28 Levisson, M., van der Oost, J. & Kengen, S. W. Characterization and structural modeling of a new type of thermostable esterase from Thermotoga maritima. *FEBS J* **274**, 2832-2842, doi:10.1111/j.1742-4658.2007.05817.x
10.1111/j.1742-4658.2007.05817.x. Epub 2007 Apr 27. (2007).

29 Koga, Y., Nishihara, M., Morii, H. & Akagawa-Matsushita, M. Ether polar lipids of methanogenic bacteria: structures, comparative aspects, and biosyntheses. *Microbiol Rev* **57**, 164-182 (1993).
30 Pugh, E. L., Wassef, M. K. & Kates, M. Inhibition of fatty acid synthetase in Halobacterium cutirubrum and Escherichia coli by high salt concentrations. *Canadian journal of biochemistry* **49**, 953-958 (1971).
31 Falb, M. *et al.* in *Extremophiles* (2008).
32 Lombard, J., López-García, P. & Moreira, D. The early evolution of lipid membranes and the three domains of life. *Nature Reviews Microbiology* **10**, 507-515, doi:10.1038/nrmicro2815 (2012).
33 Dibrova, D. V., Galperin, M. Y. & Mulkidjanian, A. Y. Phylogenomic reconstruction of archaeal fatty acid metabolism. *Environ Microbiol* **16**, 907-918, doi:10.1111/1462-2920.12359 (2014).
34 Lombard, J., López-García, P. & Moreira, D. An ACP-independent fatty acid synthesis pathway in archaea: Implications for the origin of phospholipids. *Molecular Biology and Evolution*, doi:10.1093/molbev/mss160 (2012).
35 Swigonova, Z., Mohsen, A. W. & Vockley, J. Acyl-CoA dehydrogenases: Dynamic history of protein family evolution. *J Mol Evol* **69**, 176-193, doi:10.1007/s00239-009-9263-0
10.1007/s00239-009-9263-0. Epub 2009 Jul 29. (2009).
36 He, M. *et al.* Identification and characterization of new long chain acyl-CoA dehydrogenases. *Mol Genet Metab* **102**, 418-429, doi:10.1016/j.ymgme.2010.12.005 (2011).
37 Khelifi, N. *et al.* Anaerobic oxidation of long-chain n-alkanes by the hyperthermophilic sulfate-reducing archaeon, Archaeoglobus fulgidus. *The ISME Journal* **8**, 2153-2166, doi:10.1038/ismej.2014.58
10.1038/ismej.2014.58. Epub 2014 Apr 24. (2014).
38 Müh, U., Çinkaya, I., Albracht, S. P. J. & Buckel, W. 4-Hydroxybutyryl-CoA dehydratase from Clostridium aminobutyricum: Characterization of FAD and iron-sulfur clusters involved in an overall non- redox reaction. *Biochemistry*, doi:10.1021/bi9601363 (1996).
39 Spang, A. *et al.* Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173-+, doi:10.1038/nature14447 (2015).
40 Clomburg, J. M., Vick, J. E., Blankschien, M. D., Rodríguez-Moyá, M. & Gonzalez, R. A synthetic biology approach to engineer a functional reversal of the β-oxidation cycle. *ACS Synthetic Biology*, doi:10.1021/sb3000782 (2012).
41 Lian, J. & Zhao, H. Reversal of the β-oxidation cycle in saccharomyces cerevisiae for production of fuels and chemicals. *ACS Synthetic Biology*, doi:10.1021/sb500243c (2015).
42 Callaghan, A. V. Enzymes involved in the anaerobic oxidation of n-alkanes: from methane to long-chain paraffins. *Front Microbiol* **4**, 89-89, doi:10.3389/fmicb.2013.00089
10.3389/fmicb.2013.00089. eCollection 2013. (2013).
43 Shisler, K. A. & Broderick, J. B. Glycyl radical activating enzymes: structure, mechanism, and substrate interactions. *Arch Biochem Biophys* **546**, 64-71, doi:10.1016/j.abb.2014.01.020
10.1016/j.abb.2014.01.020. Epub 2014 Jan 31. (2014).
44 Leuthner, B. *et al.* Biochemical and genetic characterization of benzylsuccinate synthase from Thauera aromatica: a new glycyl radical enzyme catalysing the first step in anaerobic toluene metabolism. *Mol Microbiol* **28**, 615-628 (1998).
45 Craciun, S. & Balskus, E. P. Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 21307-21312, doi:10.1073/pnas.1215689109
10.1073/pnas.1215689109. Epub 2012 Nov 14. (2012).
46 Selmer, T. & Andrei, P. I. p-Hydroxyphenylacetate decarboxylase from Clostridium difficile. A novel glycyl radical enzyme catalysing the formation of p-cresol. *Eur J Biochem* **268**, 1363-1372 (2001).

47 Jarling, R. *et al.* Stereochemical investigations reveal the mechanism of the bacterial activation of n-alkanes without oxygen. *Angew Chem Int Ed Engl* **51**, 1334-1338, doi:10.1002/anie.201106055
10.1002/anie.201106055. Epub 2011 Nov 30. (2012).
48 Sondergaard, D., Pedersen, C. N. & Greening, C. HydDB: A web tool for hydrogenase classification and analysis. *Sci Rep* **6**, 34212-34212, doi:10.1038/srep34212
10.1038/srep34212. (2016).
49 Buckel, W. & Thauer, R. K. Energy conservation via electron bifurcating ferredoxin reduction and proton/Na(+) translocating ferredoxin oxidation. *Biochim Biophys Acta* **1827**, 94-113, doi:10.1016/j.bbabio.2012.07.002
10.1016/j.bbabio.2012.07.002. Epub 2012 Jul 16. (2013).
50 Kanai, T. *et al.* Distinct physiological roles of the three [NiFe]-hydrogenase orthologs in the hyperthermophilic archaeon Thermococcus kodakarensis. *J Bacteriol* **193**, 3109-3116, doi:10.1128/JB.01072-10
10.1128/JB.01072-10. Epub 2011 Apr 22. (2011).
51 Jenney Jr, F. E. & Adams, M. W. Hydrogenases of the model hyperthermophiles. *Ann N Y Acad Sci* **1125**, 252-266, doi:10.1196/annals.1419.013
10.1196/annals.1419.013. (2008).
52 Ma, K., Weiss, R. & Adams, M. W. Characterization of hydrogenase II from the hyperthermophilic archaeon Pyrococcus furiosus and assessment of its role in sulfur reduction. *J Bacteriol* **182**, 1864-1871 (2000).
53 Greening, C. *et al.* Genomic and metagenomic surveys of hydrogenase distribution indicate H2 is a widely utilised energy source for microbial growth and survival. *The ISME Journal* **10**, 761-777, doi:10.1038/ismej.2015.153
10.1038/ismej.2015.153. Epub 2015 Sep 25. (2016).
54 Yu, H. *et al.* Structure of an Ancient Respiratory System. *Cell* **173**, 1636-1649 e1616, doi:10.1016/j.cell.2018.03.071 (2018).
55 Ball, J., Salvi, F. & Gadda, G. Functional Annotation of a Presumed Nitronate Monoxygenase Reveals a New Class of NADH:Quinone Reductases. *J Biol Chem* **291**, 21160-21170, doi:10.1074/jbc.M116.739151
10.1074/jbc.M116.739151. Epub 2016 Aug 8. (2016).
56 Bommer, M. *et al.* Structural basis for organohalide respiration. *Science* **346**, 455-458, doi:10.1126/science.1258118
10.1126/science.1258118. Epub 2014 Oct 2. (2014).
57 Jugder, B. E., Ertan, H., Lee, M., Manefield, M. & Marquis, C. P. Reductive Dehalogenases Come of Age in Biological Destruction of Organohalides. *Trends Biotechnol* **33**, 595-610, doi:10.1016/j.tibtech.2015.07.004
10.1016/j.tibtech.2015.07.004. (2015).
58 Miles, Z. D., McCarty, R. M., Molnar, G. & Bandarian, V. Discovery of epoxyqueuosine (oQ) reductase reveals parallels between halorespiration and tRNA modification. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 7368-7372, doi:10.1073/pnas.1018636108
10.1073/pnas.1018636108. Epub 2011 Apr 18. (2011).
59 Hug, L. A. *et al.* Overview of organohalide-respiring bacteria and a proposal for a classification system for reductive dehalogenases. *Philos Trans R Soc Lond B Biol Sci* **368**, 20120322-20120322, doi:10.1098/rstb.2012.0322
10.1098/rstb.2012.0322. Print 2013 Apr 19. (2013).

60 Arshad, A. *et al.* A Metagenomics-Based Metabolic Model of Nitrate-Dependent Anaerobic Oxidation of Methane by Methanoperedens-Like Archaea. *Front Microbiol* **6**, 1423-1423, doi:10.3389/fmicb.2015.01423
10.3389/fmicb.2015.01423. eCollection 2015. (2015).
61 Bertero, M. G. *et al.* Insights into the respiratory electron transfer pathway from the structure of nitrate reductase A. *Nat Struct Biol* **10**, 681-687, doi:10.1038/nsb969
10.1038/nsb969. Epub 2003 Aug 10. (2003).
62 Martinez-Espinosa, R. M. *et al.* Look on the positive side! The orientation, identification and bioenergetics of 'Archaeal' membrane-bound nitrate reductases. *Fems Microbiology Letters* **276**, 129-139, doi:10.1111/j.1574-6968.2007.00887.x
10.1111/j.1574-6968.2007.00887.x. Epub 2007 Sep 20. (2007).
63 Kroninger, L., Berger, S., Welte, C. & Deppenmeier, U. Evidence for the involvement of two heterodisulfide reductases in the energy-conserving system of Methanomassiliicoccus luminyensis. *FEBS J* **283**, 472-483, doi:10.1111/febs.13594 (2016).
64 Einsle, O. *et al.* Structure of cytochrome c nitrite reductase. *Nature* **400**, 476-480, doi:10.1038/22802
10.1038/22802. (1999).
65 Hiratsuka, T. *et al.* An alternative menaquinone biosynthetic pathway operating in microorganisms. *Science* **321**, 1670-1673, doi:10.1126/science.1160446
10.1126/science.1160446. (2008).
66 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* **25**, 1043-1055, doi:10.1101/gr.186072.114
10.1101/gr.186072.114. Epub 2015 May 14. (2015).
67 Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* **35**, 725-731, doi:10.1038/nbt.3893 (2017).
68 Neumann, A., Wohlfarth, G. & Diekert, G. Tetrachloroethene dehalogenase from Dehalospirillum multivorans: cloning, sequencing of the encoding genes, and expression of the pceA gene in Escherichia coli. *J Bacteriol* **180**, 4140-4145 (1998).
69 Wrighton, K. C. *et al.* RubisCO of a nucleoside pathway known from Archaea is found in diverse uncultivated phyla in bacteria. *The ISME Journal* **10**, 2702-2714, doi:10.1038/ismej.2016.53
10.1038/ismej.2016.53. Epub 2016 May 3. (2016).

C Supplementary Tables and Figures

Suppl. Table 1: Overview of presence/absence of discussed enzymes in Asgard lineages

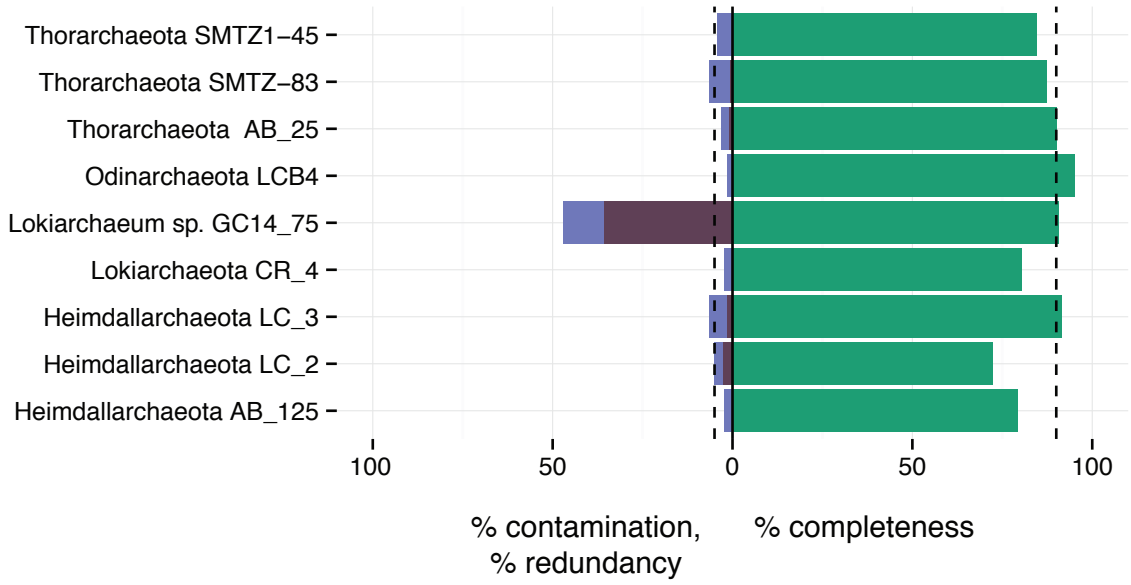
Suppl. Table 2: Annotations for proteins, which serve as candidate enzymes potentially involved in the various metabolic pathways discussed throughout this manuscript.

Suppl. Table 3: Automatic annotation of all genes

Suppl. Table 4: Carbohydrate active enzymes, peptidases, esterases and information on extracellular localization

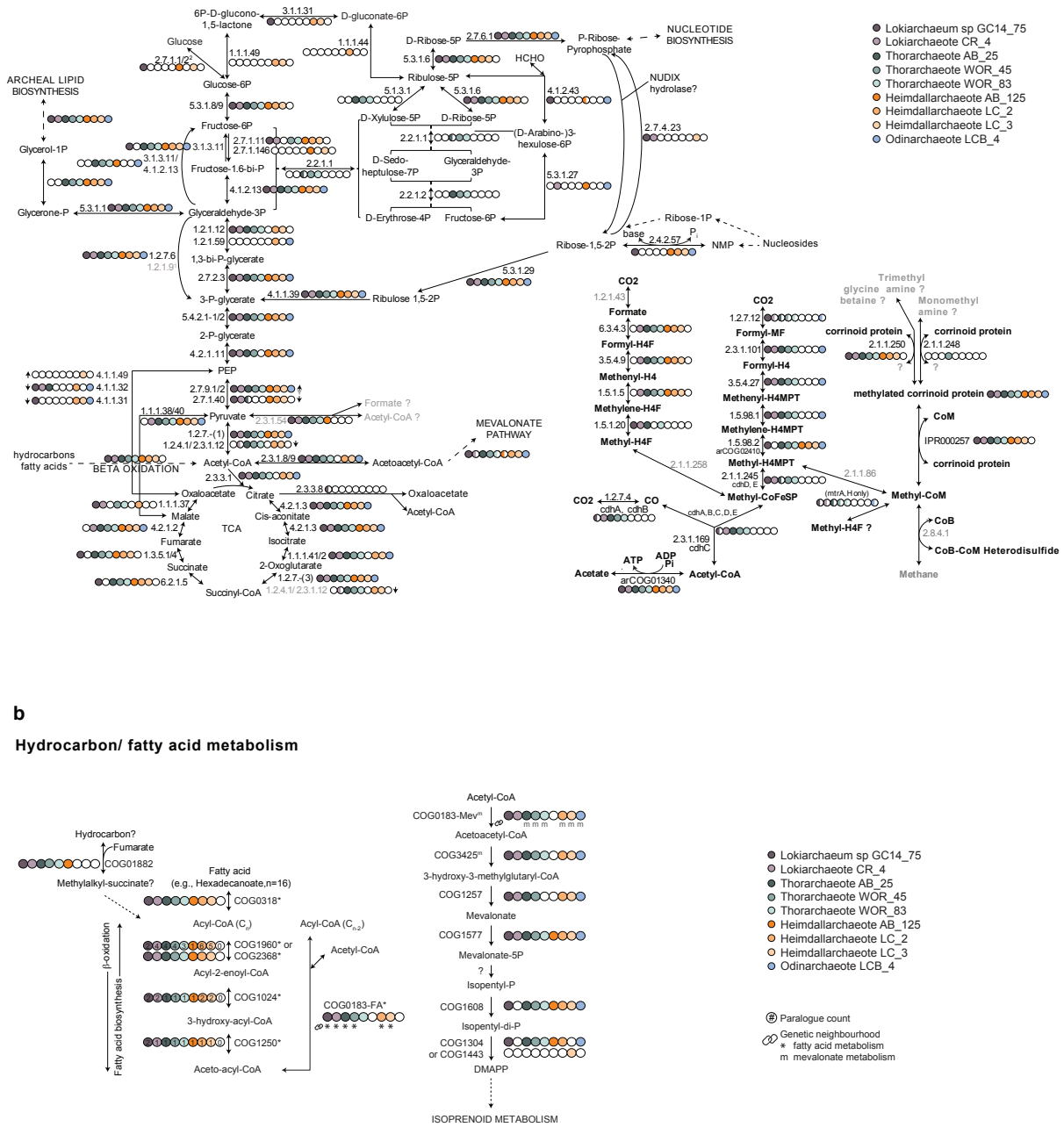
Suppl. Table 5: Annotation of beta-oxidation genes encoded by Asgard genomes per protein family/phylogeny.

Please note that all Suppl. Tables are provided in two separate excel files referred to as "S-Tables_1-4.xlsx" (includes Suppl. Tables 1-4) and "S-Table5.xlsx" (includes Suppl. Table 5; with each sheet corresponding to one particular protein family of enzymes involved in β-oxidation)



Suppl. Figure 1: Illustration of the quality of metagenome-assembled genomes (MAGs) of Asgard archaea

Barplot visualising the estimated completeness and contamination/redundancy values obtained with checkM[66] for the nine Asgard MAGs analysed in this work. The dashed line indicates the thresholds for contamination and completeness cutoffs representing high-quality MAGs according to the recently established metagenomic standards[67]. Two of the MAGs are of high quality according this criteria (Odinarchaeote LCB4 and Thorarchaeote AB_25), while the other MAGs represent medium quality bins. Purple bars represent the fraction of the contamination values (light blue bars) that is due to strain heterogeneity (redundant markers that share at least 90% of amino acid identity). As previously described in Spang et al. 2015[39], the Lokiarchaeum sp. GC14_75 MAG contains sequences from closely related strains, which explains the high contaminations/redundancy values observed in this MAG.



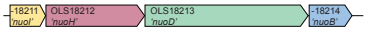
Suppl. Figure 2: Metabolic map of central carbon and lipid metabolism of Asgard archaea

For each enzyme, the corresponding enzyme commission (E.C.) or COG are provided and can be cross referenced in Suppl. Table 1 and 2. Coloured circles indicate the genomes in which homologues were detected, Lokiarchaeota (shades of purple), Thorarchaeota (shades of green), Heimdallarchaeota (shades of orange) and Odinarchaeota

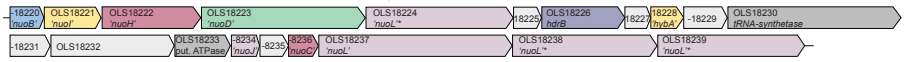
(blue). a, Central carbon metabolism in Asgard archaea. b, Proposed CoA-coupled β-oxidation and fatty acid biosynthesis pathways of the Asgard archaea. For β-oxidation (counter-clockwise), fatty acids (e.g., Hexadecanoate) are activated with a long chain fatty acid CoA ligase (COG0318) generating an acyl-CoA species which is oxidized by a chain-length specific acyl-CoA dehydrogenase (COG1960) or 4-hydroxybutyryl-CoA dehydratase (COG2368). The resulting acyl-2-enoyl-CoA is further oxidized by enoyl-CoA hydratase (COG1024) and 3-hydroxy-acyl-CoA dehydrogenase (COG1250) to generate aceto-acyl-CoA which is ultimately cleaved by a B-ketothiolase/acetyl-CoA acetyltransferase (COG0183-FA) to generate acetyl-CoA and a Cn-2 acyl-CoA chain. Asterisks indicate that at least one B-ketothiolase/acetyl-CoA orthologue in the represented genome was encoded within 15 genes of one or more genes related to fatty acid metabolism (COG0318, COG1960, COG2368, COG1024, or COG1250). For complex protein families, the number of paralogues identified are shown in each circle. Fatty acid biosynthesis is depicted as the reverse of β-oxidation (clockwise) using CoA[33] and not acyl-carrier protein as an acyl carrier. **c,** Proposed mevalonate biosynthesis pathway. Acetyl-CoA is consecutively acetylated by the actions of B-ketothiolase/acetyl-CoA acetyltransferase (COG0183-Mev) and 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) synthase (COG3425) to produce HMG-CoA which in turn is converted to mevalonate by HMG-CoA reductase (COG1257). Mevalonate is phosphorylated by mevalonate kinase (COG1577) and converted to isopentyl-phosphate by an unknown enzyme. Isopentyl-phosphate is phosphorylated by isopentyl-phosphate kinase (COG1608) to isopentyl-di-phosphate which is isomerized to dimethylallyl pyrophosphate (DMAPP) by isopentyl-diphosphate isomerase 1 (COG1443) or 2 (COG1304). 'm' indicates that the COG0183 orthologue was encoded within 15 genes of the HMG-CoA synthase (COG3425). Greyed out text indicates: presence unclear, putative.

Group 4 [NiFe]-hydrogenase

Ca. Odinarchaeota archaeon LCB_4 – Group 4 [NiFe]-hydrogenase (OdinLCB4_14520)



Ca. Odinarchaeota archaeon LCB_4 – Group 4 [NiFe]-hydrogenase (OdinLCB4_14620)



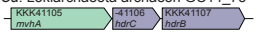
Ca. Odinarchaeota archaeon LCB_4 – Group 4 [NiFe]-hydrogenase (OdinLCB4_02560)

Ca. Heimdallarchaeota archaeon AB_125 – Group 4 [NiFe]-hydrogenase (HeimAB125_01550)

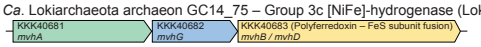
Ca. Heimdallarchaeota archaeon LC_2 – Group 4 [NiFe]-hydrogenase (HeimC2_23020)

Group 3c [NiFe]-hydrogenase

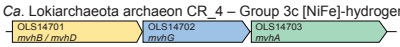
Ca. Lokiarchaeota archaeon GC14_75 – partial Group 3c [NiFe]-hydrogenase (Lokiarch_45490)



Ca. Lokiarchaeota archaeon GC14_75 – Group 3c [NiFe]-hydrogenase (Lokiarch_49310)

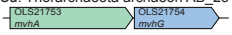


Ca. Lokiarchaeota archaeon CR_4 – Group 3c [NiFe]-hydrogenase (RBG13Loki_1669)

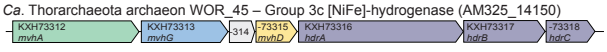


Ca. Thorarchaeota archaeon AB_25 – Group 3c [NiFe]-hydrogenase (ThorAB25_16580)

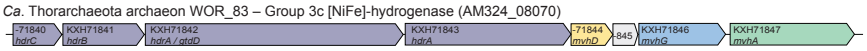
Ca. Thorarchaeota archaeon AB_25 – Group 3c [NiFe]-hydrogenase (ThorAB25_26490)



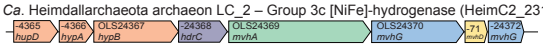
Ca. Thorarchaeota archaeon WOR_45 – Group 3c [NiFe]-hydrogenase (AM325_14150)



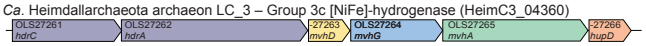
Ca. Thorarchaeota archaeon WOR_83 – Group 3c [NiFe]-hydrogenase (AM324_08070)



Ca. Heimdallarchaeota archaeon LC_2 – Group 3c [NiFe]-hydrogenase (HeimC2_23150)



Ca. Heimdallarchaeota archaeon LC_3 – Group 3c [NiFe]-hydrogenase (HeimC3_04360)

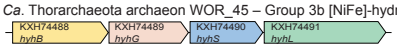


Group 3b [NiFe]-hydrogenase

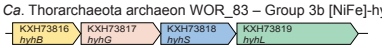
Ca. Lokiarchaeota archaeon GC14_75 – partial Group 3b [NiFe]-hydrogenase (Lokiarch_31190)

Ca. Odinarchaeota archaeon LCB_4 – Group 3b [NiFe]-hydrogenase (OdinLCB4_08000)

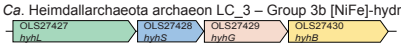
Ca. Thorarchaeota archaeon WOR_45 – Group 3b [NiFe]-hydrogenase (AM325_05810)

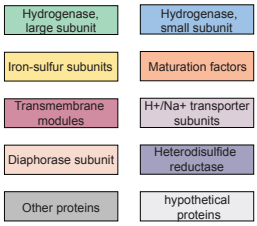


Ca. Thorarchaeota archaeon WOR_83 – Group 3b [NiFe]-hydrogenase (AM324_16110)

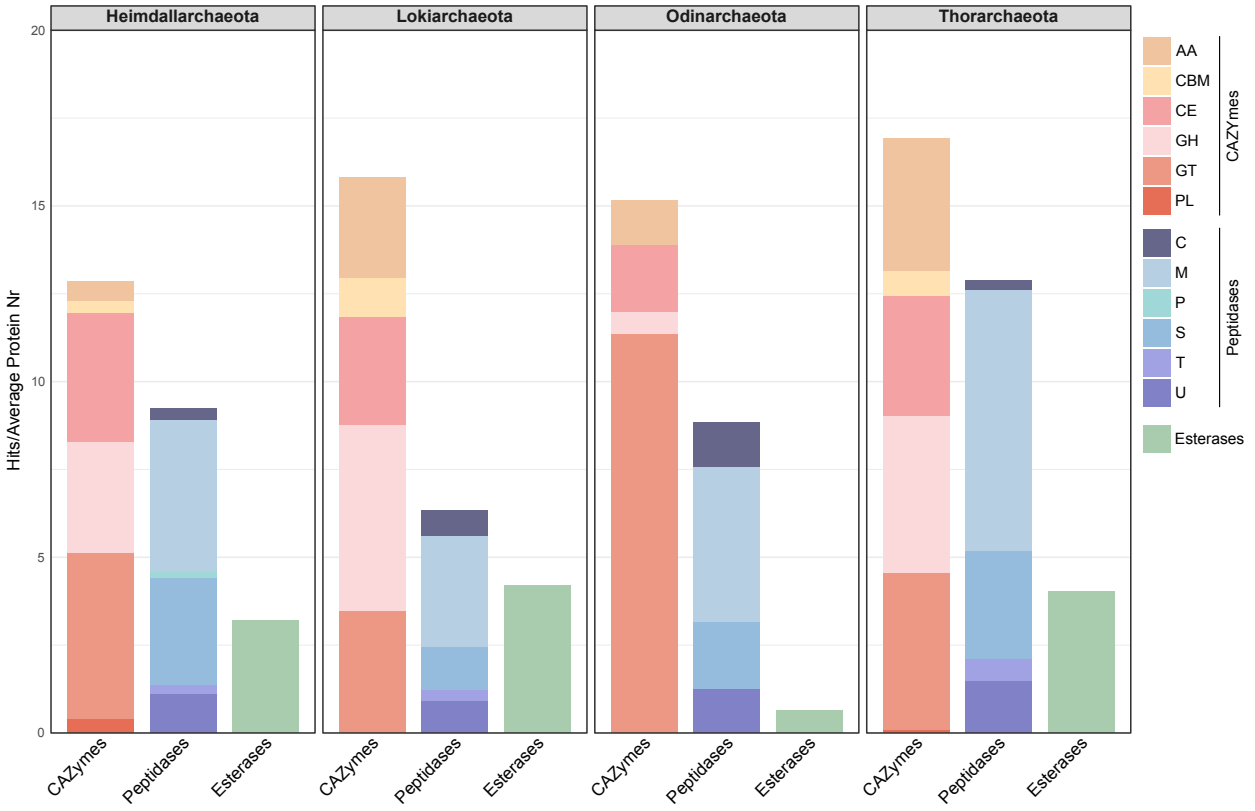


Ca. Heimdallarchaeota archaeon LC_3 – Group 3b [NiFe]-hydrogenase (HeimC3_04260)



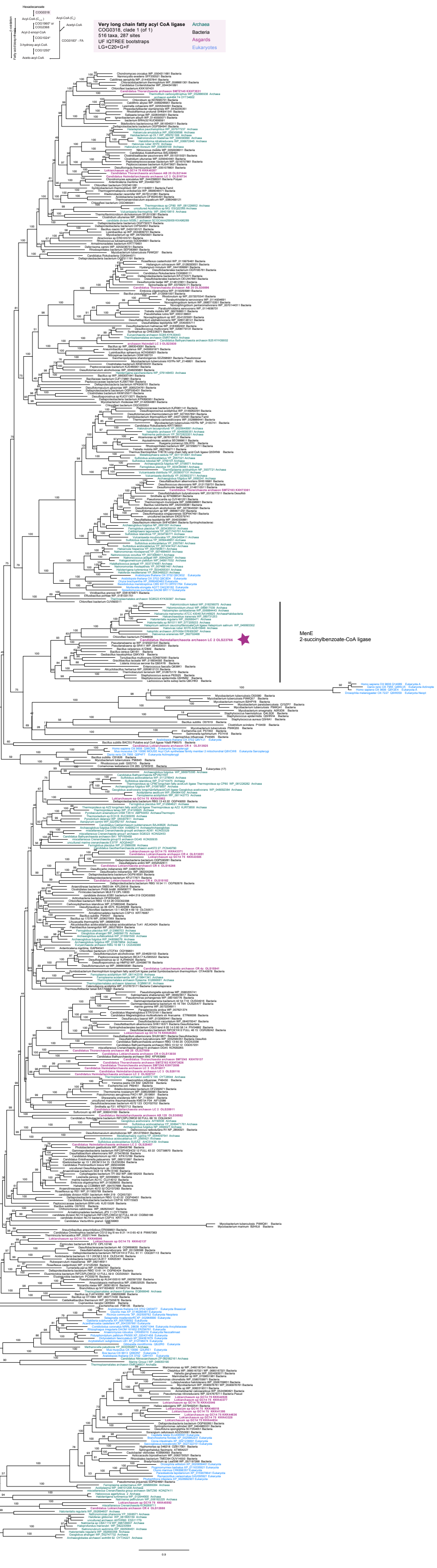


Suppl. Figure 3: Structure of [NiFe]-hydrogenase gene clusters. Genes were classified based on their top hits to the Conserved Domain Database (CDD) webserver and comparison of their sequences with those of hydrogenase operons represented in the hydrogenase database (HydDB). All genes are drawn to scale. Short gene names are given in italic and are shown in quotes whenever they are distantly related to known proteins. For group 4 [NiFe]-hydrogenases, the homologous genes to Nuo (Complex I; NADH-quinone oxidoreductase) are shown, with the more divergent *nuoL* homologs asterisked. Abbreviations are as follows: hdr, heterodisulfide reductase; hyh, hyperthermophile hydrogenase (group 3b [NiFe]-hydrogenase); mvh, methyl-viologen-reducing hydrogenase (group 3c [NiFe]-hydrogenase); hyp, Hydrogenase maturation factor; nuo, NADH dehydrogenase. [1]Note that this gene cluster is located at the end of the contig and thus incomplete. *NuoL*-like subunits are encoded on other short contigs, which are part of the *Heimdallarchaeum* AB125 genome. Thus, the encoded Group 4 [NiFe]-hydrogenase of this organism is most likely also ion conductive. Please note that also other gene clusters could be incomplete due to their location at the end of a contig.



Suppl. Figure 4: Abundance of carbohydrate-active enzymes, peptidases and esterases in the Asgard genomes

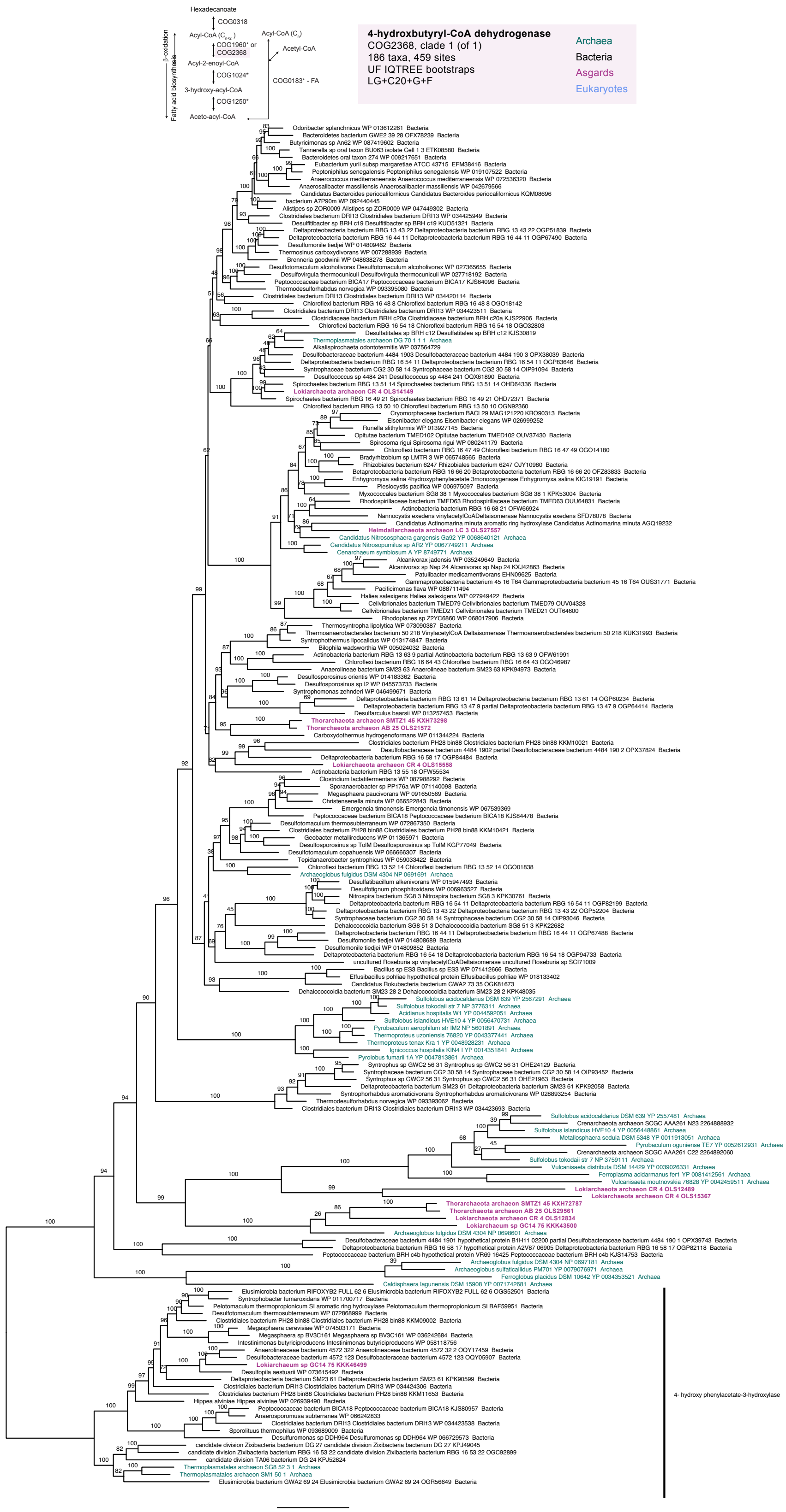
Carbohydrate-active enzymes (CAZYmes), peptidases and esterases were identified in Asgard genomes using the dbCAN webserver, MEROPs and ESTER databases respectively (see Method section for details). Count data was averaged across the number of genomes represented by each phylum and normalized by the average number of proteins detected in each phylum. AA, auxiliary activities; CBM, carbohydrate-binding module; CE, carbohydrate esterase; GH, glycoside hydrolase; GT, glycosyltransferases; PL, polysaccharide lyase. A, aspartic peptidase; C, cysteine peptidase; M, metallopeptidase; P, mixed peptidase; S, serine peptidases; T, mixed catalytic type; U, Unknown Catalytic Type. Count data is summarized in Suppl. Table 4.


Hexadecanoate
COG0318
Acyl-CoA (C_{n+2})
COG1960* or
COG2368
Acyl-2-enoyl-CoA
COG1024*
3-hydroxy-acyl-CoA
COG1250*
Aceto-acyl-CoA
Acyl-CoA (C_n)
Acetyl-CoA
COG0183* - FA
β-oxidation
Fatty acid biosynthesis
Very long chain fatty acyl CoA ligase
COG0318, clade 1 (of 1)
516 taxa, 287 sites
UF IQTREE bootstraps
LG+C20+G+F
Archaea
Bacteria
Asgards
Eukaryotes
MenE
2-succinylbenzoate-CoA ligase

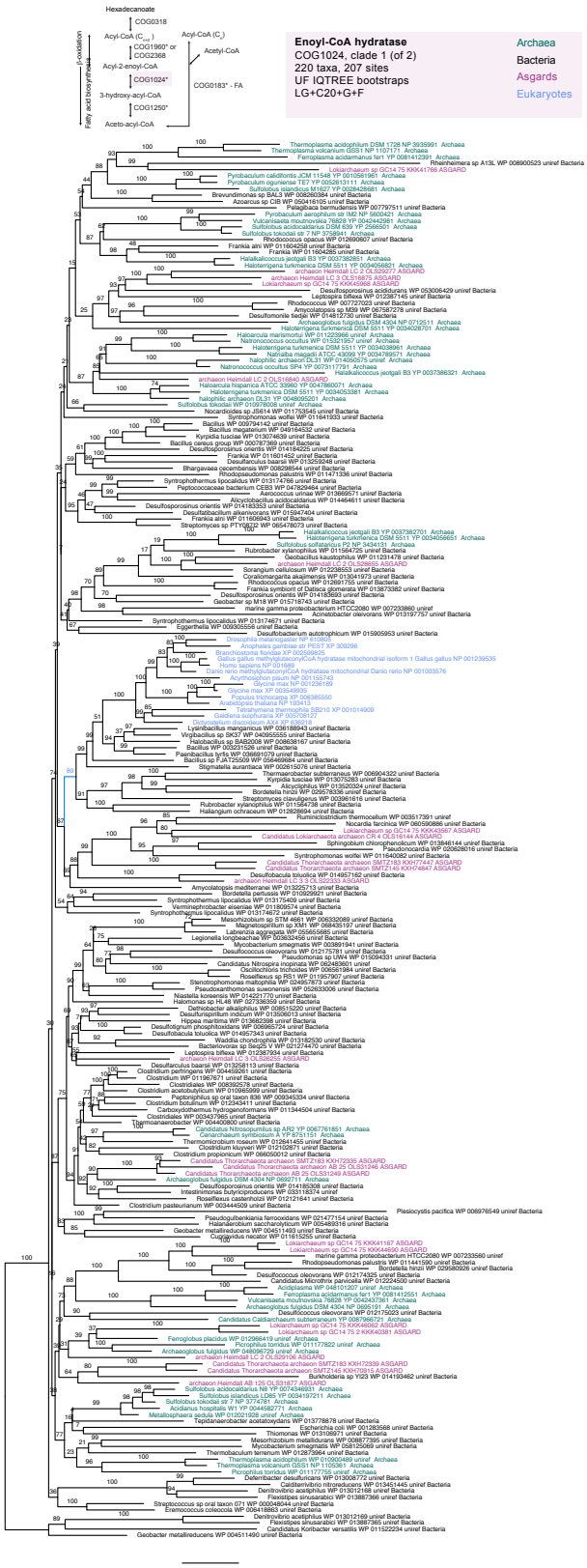
Suppl. Figure 5: Phylogenetic analysis of long chain fatty acid CoA ligase (COG0318). An unrooted phylogenetic tree was generated using IQ-tree LG+C20+F on an alignment of 516 taxa and 287 sites. Bacteria, Archaea, eukaryotes and Asgard archaea are colored in black, green, blue and purple respectively. Heimdallarchaeota MenE-like homologue is indicated with a star. Raw data files are available via figshare (see Data availability for more details).

Suppl. Figure 6: Asgard archaea have a versatile repertoire of Acyl-CoA dehydrogenases (COG1960). The acyl-CoA dehydrogenase family was subdivided into three large clades (a-c) based on preliminary phylogenetic analysis of sequences retrieved from Swigonova et al. 2009 and the best BLAST hits for each ASGARD sequences. Each unrooted tree was estimated using IQ-tree LG+C20+F for a) Clade 1 (117 taxa, 365 sites), b) Clade 2 (542 taxa, 311 sites), and c) Clade 3 (419 taxa, 300 sites) independently. Bacteria, Archaea, eukaryotes and Asgard archaea are colored in black, green, blue and purple, respectively. Bipartition support values are depicted as open or closed circles for ultra-fast bootstrap values greater than 70 and 90, respectively. A summary of the clade distribution among Asgard

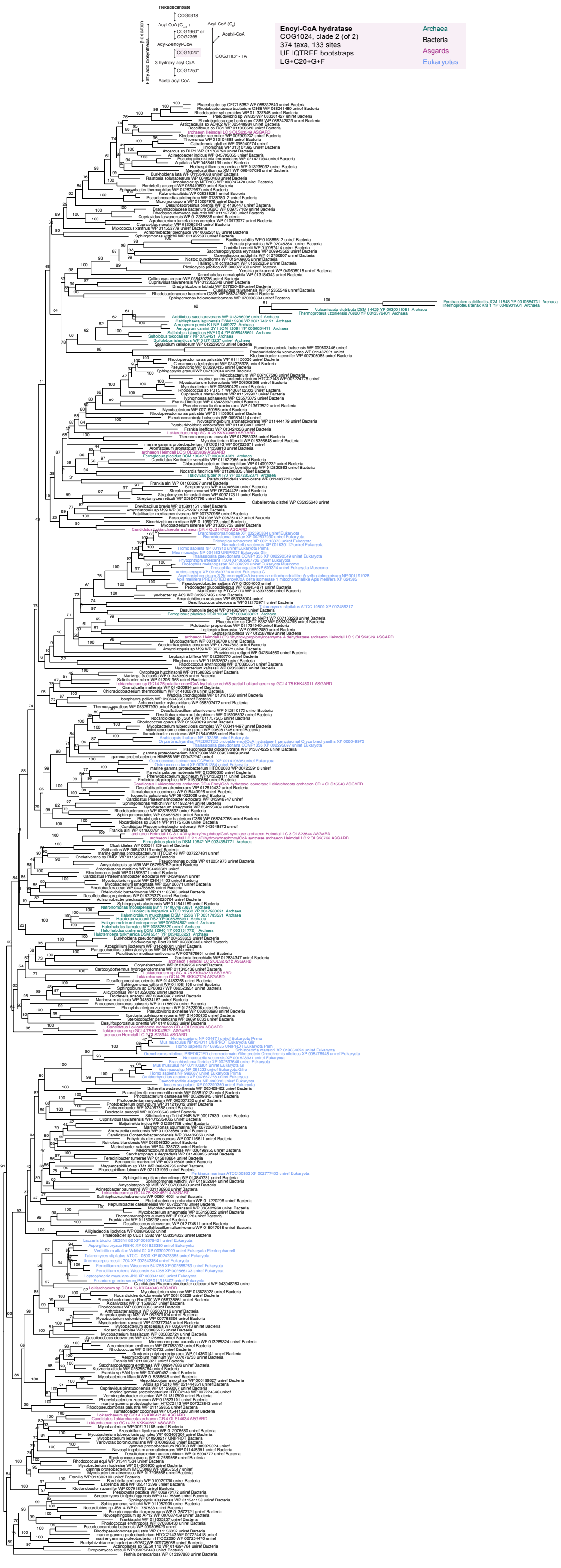
genomes is shown in d. d) Clade summaries for each Asgard genome. Raw data files are available via figshare (see Data availability for more details).


4-hydroxbutyryl-CoA dehydrogenase
COG2368, clade 1 (of 1)
186 taxa, 459 sites
UF IQTREE bootstraps
LG+C20+G+F
Archaea
Bacteria
Asgards
Eukaryotes
Hexadecanoate
COG0318
Acyl-CoA
COG1960* or COG2368
Acyl-2-enoyl-CoA
COG1024*
3-hydroxy-acyl-CoA
COG1250*
Aceto-acyl-CoA
Acetyl-CoA
COG0183* - FA
β-oxidation
Fatty acid biosynthesis
4- hydroxy phenylacetate-3-hydroxylase
0.5

Suppl. Figure 7: Phylogenetic analysis of 4-hydroxy-butyryl-CoA dehydrogenases (COG2368). Phylogenetic tree was estimated using IQ-tree LG+C20+F on an alignment of 186 taxa and 459 sites and is unrooted. Bacteria, Archaea, eukaryotes and Asgards are colored in black, green, blue and purple, respectively. Raw data files are available via figshare (see Data availability for more details).


Hexadecanoate
COG0318
Acyl-CoA (C_{n+2})
COG1960* or COG2368
Acyl-2-enoyl-CoA
COG1024*
3-hydroxy-acyl-CoA
COG1250*
Aceto-acyl-CoA
Acyl-CoA (C_n)
Acetyl-CoA
COG0183* - FA
β-oxidation
Fatty acid biosynthesis
Enoyl-CoA hydratase
COG1024, clade 1 (of 2)
220 taxa, 207 sites
UF IQTREE bootstraps
LG+C20+G+F
Archaea
Bacteria
Asgards
Eukaryotes
0.5

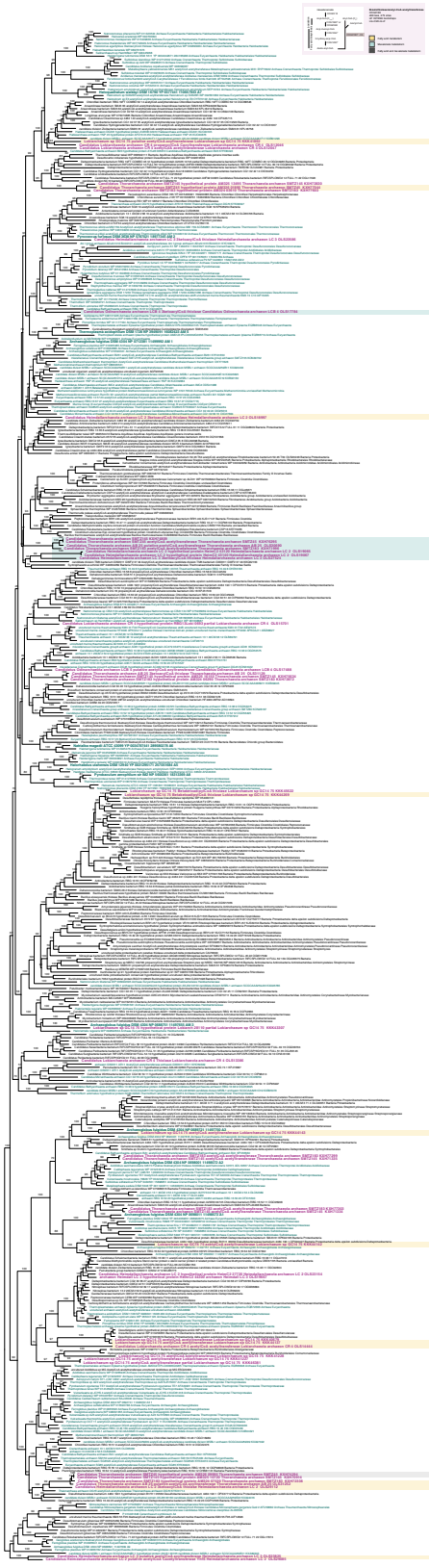
Suppl. Figure 8: Phylogenetic analysis of Enoyl-CoA reductase clade 1 (COG1024). Phylogenetic tree was estimated using IQ-tree LG+C20+F on an alignment of 220 taxa and 207 sites and is unrooted. Bacteria, Archaea, eukaryotes and Asgards are colored in black, green, blue and purple respectively. Raw data files are available via figshare (see Data availability for more details).


Hexadecanoate
COG0318
Acyl-CoA (C_{n+2})
COG1960* or COG2368
Acyl-2-enoyl-CoA
COG1024*
3-hydroxy-acyl-CoA
COG1250*
Aceto-acyl-CoA
Acyl-CoA (C_n)
Acetyl-CoA
COG0183* - FA
β-oxidation
Fatty acid biosynthesis
Enoyl-CoA hydratase
COG1024, clade 2 (of 2)
374 taxa, 133 sites
UF IQTREE bootstraps
LG+C20+G+F
Archaea
Bacteria
Asgards
Eukaryotes
2.0

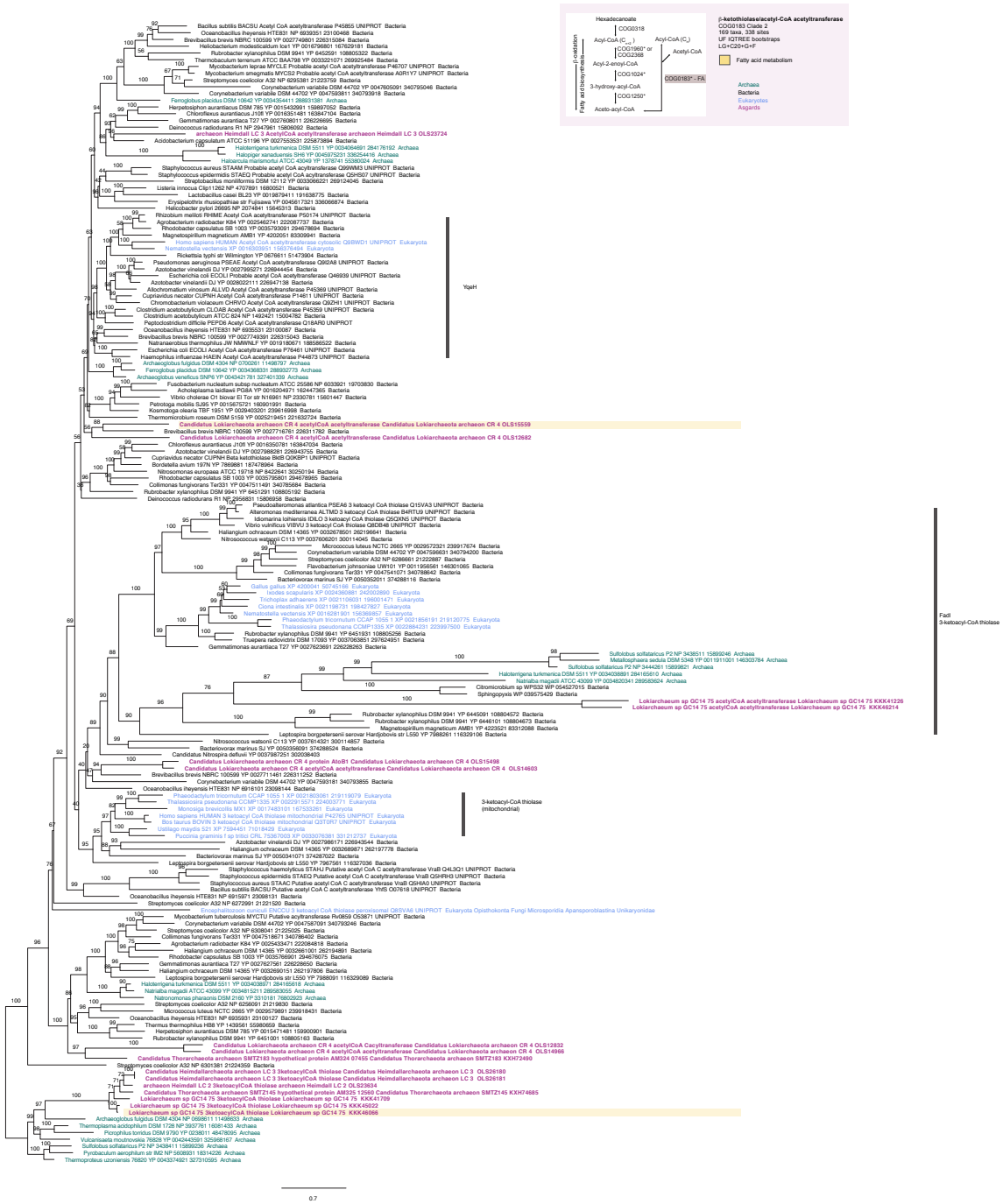
Suppl. Figure 9: Phylogenetic analysis of Enoyl-CoA reductase clade 2 (COG1024). Phylogenetic tree was estimated using IQ-tree LG+C20+F on an alignment of 374 taxa and 133 sites and is unrooted. Bacteria, Archaea, eukaryotes and Asgards are colored in black, green, blue and purple respectively. Raw data files are available via figshare (see Data availability for more details).


Hexadecanoate
Acyl-CoA (C_)
Acyl-CoA (C_)
Acetyl-CoA
Acyl-enoyl-CoA
3-hydroxy-acyl-CoA
Acetoacetyl-CoA
FA elongation cycle (4 steps)
3-hydroxyacyl-CoA dehydrogenase
Archaea
Bacteria
Eukaryotes
Asgards
0.4

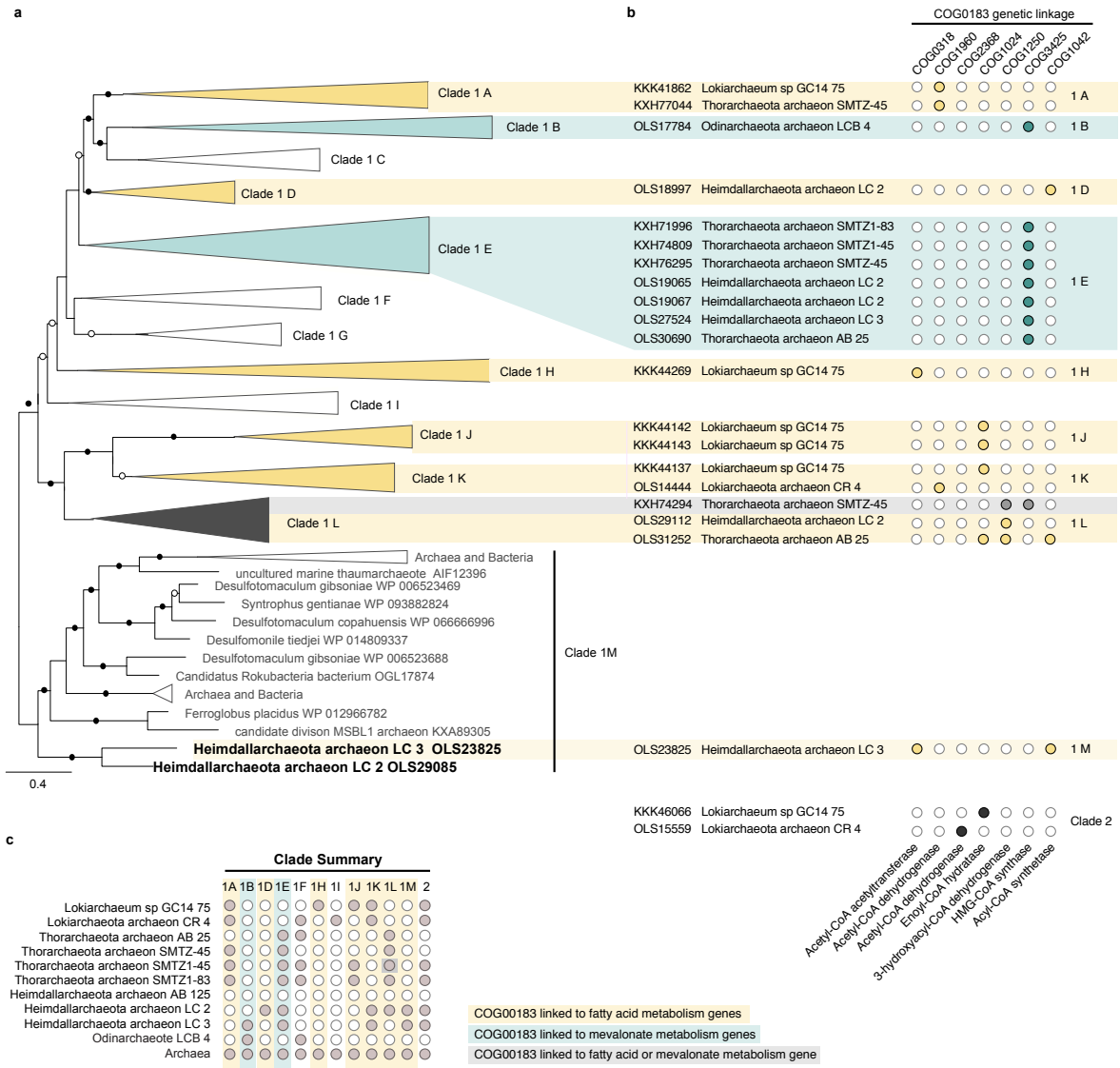
Suppl. Figure 10: Phylogenetic analysis of 3-hydroxy-acyl-CoA dehydrogease (COG1250). Phylogenetic tree was estimated using IQ-tree LG+C20+F on an alignment of 288 taxa and 240 sites and is unrooted. Bacteria, Archaea, eukaryotes and Asgards are colored in black, green, blue and purple respectively. Raw data files are available via figshare (see Data availability for more details).


Clade 1A
Clade 1B
Clade 1C
Clade 1D
Clade 1E
Clade 1F
Clade 1G
Clade 1H
Clade 1I
Clade 1J
Clade 1K
Clade 1L
Clade 1M

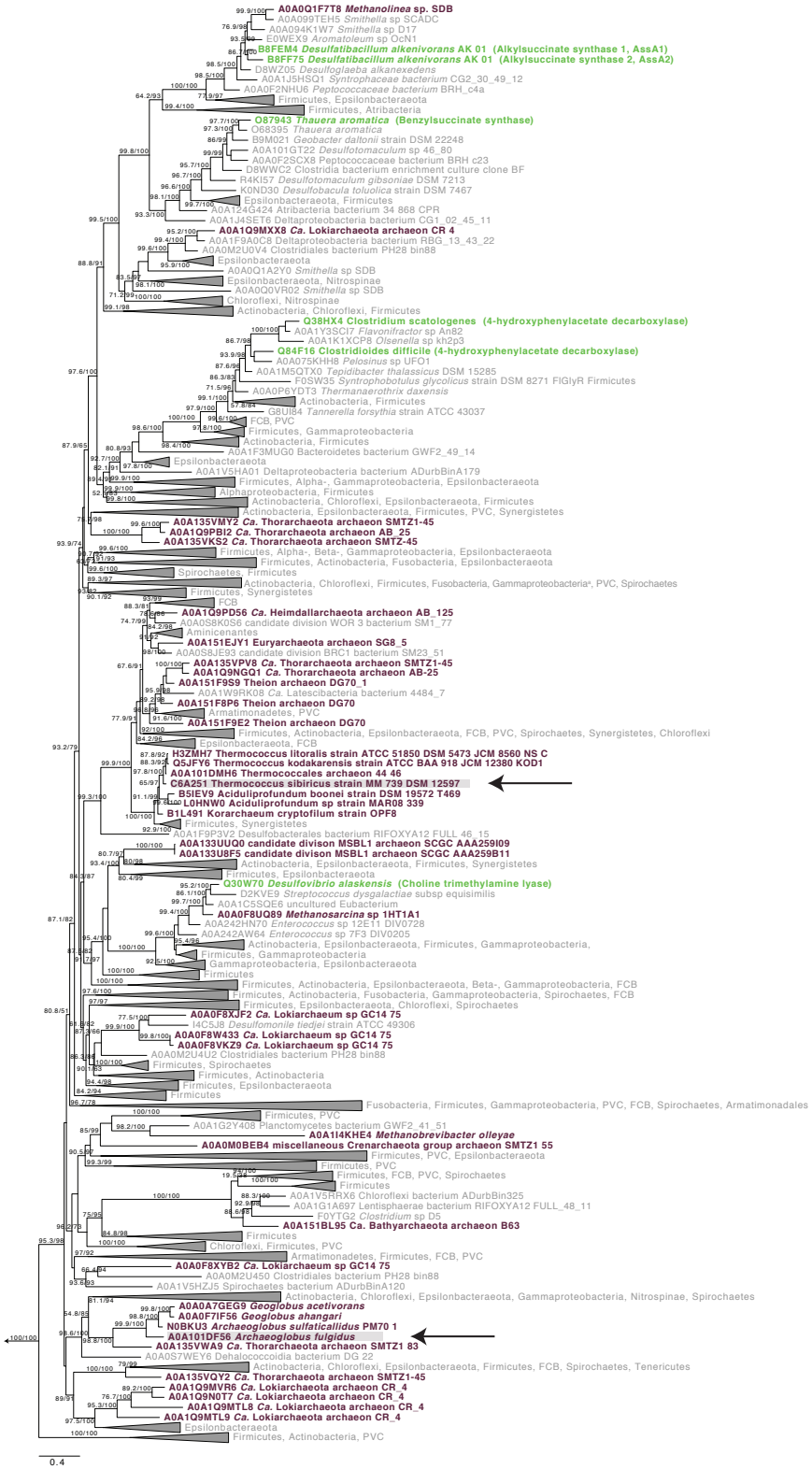
Suppl. Figure 11: Phylogenetic analysis of clade 1 of B-ketothiolase/acetyl-CoA acetyltransferase (COG0183). Phylogenetic tree was estimated using IQ-tree LG+C20+F on an alignment of 466 taxa and 279 sites. Bacteria, Archaea, and Asgards are colored in black, green, and purple respectively. Gene sequences from Asgard that are genetically linked to fatty acid metabolism or mevalonate metabolism or both fatty acid and mevalonate metabolism are shown in yellow, green and grey respectively summarized in Suppl. Figure 13. Raw data files are available via figshare (see Data availability for more details).


Hexadecanoate
COG0318
Acyl-CoA (C16)
COG1960* or COG2368
Acyl-CoA (C4)
Acetyl-CoA
Acyl-2-enoyl-CoA
COG1024*
COG0183* - FA
3-hydroxy-acyl-CoA
COG1250*
Aceto-acyl-CoA
β-oxidation
Fatty acid biosynthesis
β-ketothiolase/acetyl-CoA acetyltransferase
COG0183 Clade 2
169 taxa, 338 sites
UF IQTREE bootstraps
LG+C20+G+F
Fatty acid metabolism
Archaea
Bacteria
Eukaryotes
Asgards
YqeH
FadI
3-ketoacyl-CoA thiolase
3-ketoacyl-CoA thiolase (mitochondrial)
0.7

Suppl. Figure 12: Phylogenetic analysis of clade 2 of B-ketothiolase/acetyl-CoA acetyltransferase (COG0183). Phylogenetic tree was estimated using IQ-tree LG+C20+F on an alignment of 169 taxa and 338 sites. Bacteria, Archaea, and Asgards are coloured in black, green, and purple respectively. Gene sequences from Asgard that are genetically linked to fatty acid metabolism yellow and summarized in Suppl. Figure 13. Raw data files are available via figshare (see Data availability for more details).

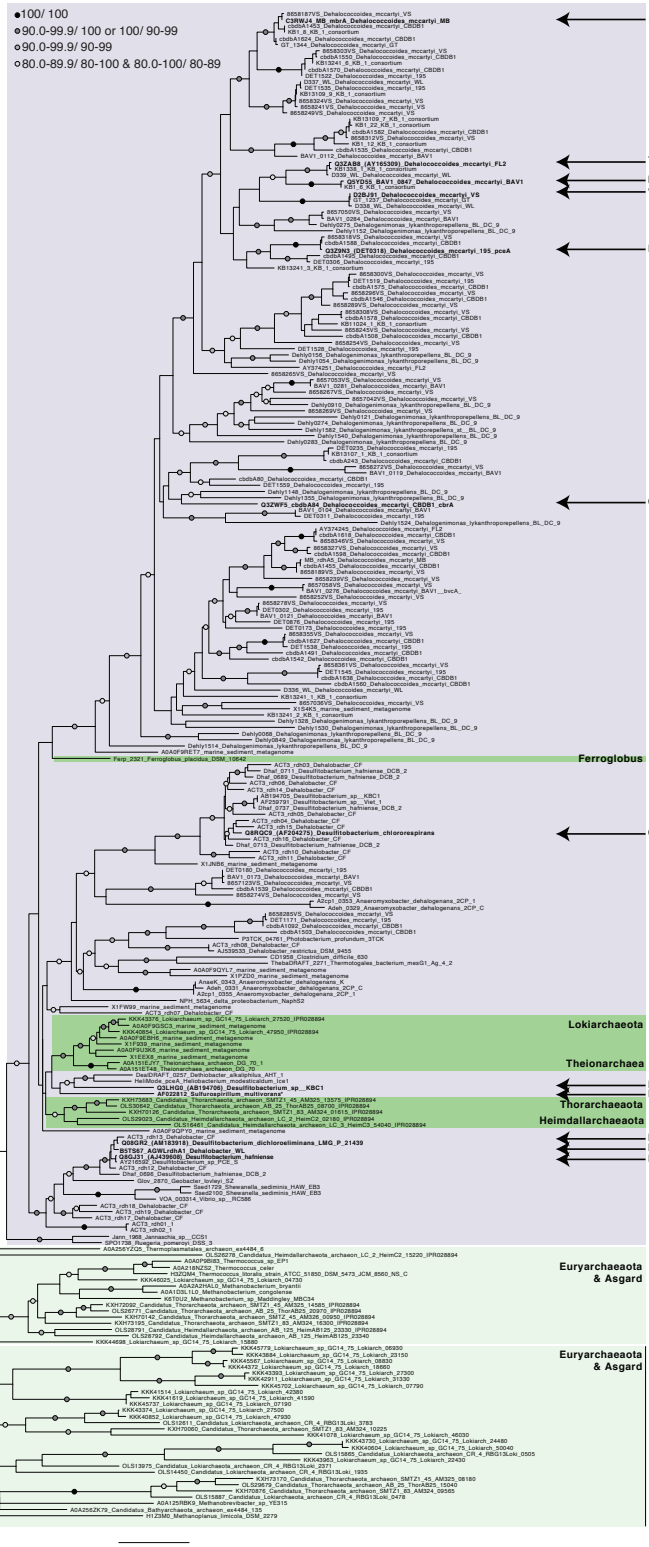


Suppl. Figure 13: Phylogenetic and gene linkage analysis of B-ketothiolase/acetyl-CoA acetyltransferase (COG0183). a, The COG0183 protein family was divided into two clades based on previous reports by[33] where clade 1 is composed of mostly archaeal sequences previously implicated in fatty acid and mevalonate metabolism. A schematic representation of clade 1 based on a tree constructed with IQ-tree LG+C20+F on an alignment of 466 taxa and 279 sites and unrooted. Bipartition support values are depicted as open or closed circles for ultra-fast bootstrap values greater than 70 and 90 respectively. Full phylogeny can be found in Suppl. Figs 10, 11. b, Gene sequences from Asgard archaea that are genetically linked (i.e., within 15 genes of the B-ketothiolase/acetyl-CoA acetyltransferase gene) to the indicated fatty acid metabolism or mevalonate metabolism or both fatty acid and mevalonate metabolism are shown in yellow, green and grey respectively. c, Distribution of Asgard archaeal sequences in each subclade. Raw data files are available via figshare (see Data availability for more details).

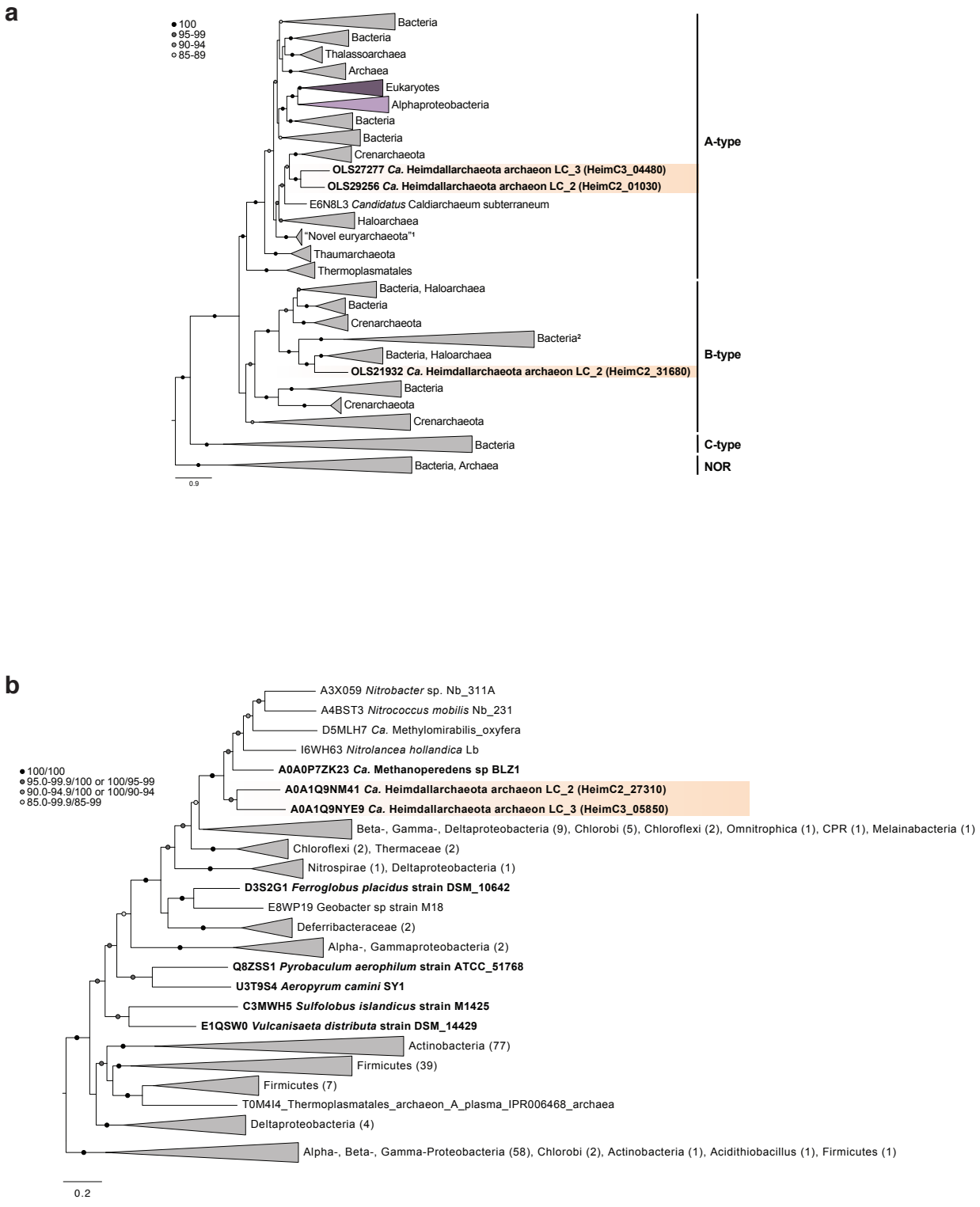

A0A0Q1F7T8 Methanolinea sp. SDB
A0A099TEH5 Smithella sp SCADC
A0A094K1W7 Smithella sp D17
E0WEX9 Aromatoleum sp OcN1
B8FEM4 Desulfatibacillum alkenivorans AK 01 (Alkylsuccinate synthase 1, AssA1)
B8FF75 Desulfatibacillum alkenivorans AK 01 (Alkylsuccinate synthase 2, AssA2)
D8WZ05 Desulfoglaeba alkanexedens
A0A1J5HSQ1 Syntrophaceae bacterium CG2_30_49_12
A0A0F2NHU6 Peptococcaceae bacterium BRH_c4a
Firmicutes, Epsilonbacteraeota
Firmicutes, Atribacteria
O87943 Thauera aromatica (Benzylsuccinate synthase)
O68395 Thauera aromatica
B9M021 Geobacter daltonii strain DSM 22248
A0A101GT22 Desulfotomaculum sp 46_80
A0A0F2SCX8 Peptococcaceae bacterium BRH c23
D8WWC2 Clostridia bacterium enrichment culture clone BF
R4KI57 Desulfotomaculum gibsoniae DSM 7213
K0ND30 Desulfobacula toluolica strain DSM 7467
Epsilonbacteraeota, Firmicutes
A0A124G424 Atribacteria bacterium 34 868 CPR
A0A1J4SET6 Deltaproteobacteria bacterium CG1_02_45_11
A0A1Q9MXX8 Ca. Lokiarchaeota archaeon CR 4
A0A1F9A0C8 Deltaproteobacteria bacterium RBG_13_43_22
A0A0M2U0V4 Clostridiales bacterium PH28 bin88
Epsilonbacteraeota
A0A0Q1A2Y0 Smithella sp SDB
Epsilonbacteraeota, Nitrospinae
A0A0Q0VR02 Smithella sp SDB
Chloroflexi, Nitrospinae
Actinobacteria, Chloroflexi, Firmicutes
Q38HX4 Clostridium scatologenes (4-hydroxyphenylacetate decarboxylase)
A0A1Y3SCI7 Flavonifractor sp An82
A0A1K1XCP8 Olsenella sp kh2p3
Q84F16 Clostridioides difficile (4-hydroxyphenylacetate decarboxylase)
A0A075KHH8 Pelosinus sp UFO1
A0A1M5QTX0 Tepidibacter thalassicus DSM 15285
F0SW35 Syntrophobotulus glycolicus strain DSM 8271 FlGlyR Firmicutes
A0A0P6YDT3 Thermanaerothrix daxensis
Actinobacteria, Firmicutes
G8UI84 Tannerella forsythia strain ATCC 43037
FCB, PVC
Firmicutes, Gammaproteobacteria
Actinobacteria, Firmicutes
A0A1F3MUG0 Bacteroidetes bacterium GWF2_49_14
Epsilonbacteraeota
A0A1V5HA01 Deltaproteobacteria bacterium ADurbBinA179
Firmicutes, Alpha-, Gammaproteobacteria, Epsilonbacteraeota
Alphaproteobacteria, Firmicutes
Actinobacteria, Chloroflexi, Epsilonbacteraeota, Firmicutes
Actinobacteria, Epsilonbacteraeota, Firmicutes, PVC, Synergistetes
A0A135VMY2 Ca. Thorarchaeota archaeon SMTZ1-45
A0A1Q9PBI2 Ca. Thorarchaeota archaeon AB_25
A0A135VKS2 Ca. Thorarchaeota archaeon SMTZ-45
Firmicutes, Alpha-, Beta-, Gammaproteobacteria, Epsilonbacteraeota
Firmicutes, Actinobacteria, Fusobacteria, Epsilonbacteraeota
Spirochaetes, Firmicutes
Actinobacteria, Chloroflexi, Firmicutes, Fusobacteria, Gammaproteobacteria[a], PVC, Spirochaetes
Firmicutes, Synergistetes
FCB
A0A1Q9PD56 Ca. Heimdallarchaeota archaeon AB_125
A0A0S8K0S6 candidate division WOR 3 bacterium SM1_77
Aminicenantes
A0A151EJY1 Euryarchaeota archaeon SG8_5
A0A0S8JE93 candidate division BRC1 bacterium SM23_51
A0A135VPV8 Ca. Thorarchaeota archaeon SMTZ1-45
A0A1Q9NGQ1 Ca. Thorarchaeota archaeon AB-25
A0A151F9S9 Theion archaeon DG70_1
A0A1W9RK08 Ca. Latescibacteria bacterium 4484_7
A0A151F8P6 Theion archaeon DG70
Armatimonadetes, PVC
A0A151F9E2 Theion archaeon DG70
Firmicutes, Actinobacteria, Epsilonbacteraeota, FCB, PVC, Spirochaetes, Synergistetes, Chloroflexi
Epsilonbacteraeota, FCB
H3ZMH7 Thermococcus litoralis strain ATCC 51850 DSM 5473 JCM 8560 NS C
Q5JFY6 Thermococcus kodakarensis strain ATCC BAA 918 JCM 12380 KOD1
A0A101DMH6 Thermococcales archaeon 44 46
C6A251 Thermococcus sibiricus strain MM 739 DSM 12597
B5IEV9 Aciduliprofundum boonei strain DSM 19572 T469
L0HNW0 Aciduliprofundum sp strain MAR08 339
B1L491 Korarchaeum cryptofilum strain OPF8
Firmicutes, Synergistetes
A0A1F9P3V2 Desulfobacterales bacterium RIFOXYA12 FULL 46_15
A0A133UUQ0 candidate division MSBL1 archaeon SCGC AAA259I09
A0A133U8F5 candidate division MSBL1 archaeon SCGC AAA259B11
Actinobacteria, Epsilonbacteraeota, Firmicutes, Synergistetes
Firmicutes, Epsilonbacteraeota
Q30W70 Desulfovibrio alaskensis (Choline trimethylamine lyase)
D2KVE9 Streptococcus dysgalactiae subsp equisimilis
A0A1C5SQE6 uncultured Eubacterium
A0A0F8UQ89 Methanosarcina sp 1HT1A1
A0A242HN70 Enterococcus sp 12E11 DIV0728
A0A242AW64 Enterococcus sp 7F3 DIV0205
Actinobacteria, Epsilonbacteraeota, Firmicutes, Gammaproteobacteria,
Firmicutes, Gammaproteobacteria
Gammaproteobacteria, Epsilonbacteraeota
Firmicutes
Firmicutes, Actinobacteria, Epsilonbacteraeota, Beta-, Gammaproteobacteria, FCB
Firmicutes, Actinobacteria, Fusobacteria, Gammaproteobacteria, Spirochaetes, FCB
Firmicutes, Epsilonbacteraeota, Chloroflexi, Spirochaetes
A0A0F8XJF2 Ca. Lokiarchaeum sp GC14 75
I4C5J8 Desulfomonile tiedjei strain ATCC 49306
A0A0F8W433 Ca. Lokiarchaeum sp GC14 75
A0A0F8VKZ9 Ca. Lokiarchaeum sp GC14 75
A0A0M2U4U2 Clostridiales bacterium PH28 bin88
Firmicutes, Spirochaetes
Firmicutes, Actinobacteria
Firmicutes, Epsilonbacteraeota
Firmicutes
Fusobacteria, Firmicutes, Gammaproteobacteria, PVC, FCB, Spirochaetes, Armatimonadales
Firmicutes, PVC
A0A1G2Y408 Planctomycetes bacterium GWF2_41_51
A0A1I4KHE4 Methanobrevibacter olleyae
A0A0M0BEB4 miscellaneous Crenarchaeota group archaeon SMTZ1 55
Firmicutes, PVC, Epsilonbacteraeota
Firmicutes, PVC
Firmicutes, FCB, PVC, Spirochaetes
Firmicutes
A0A1V5HRX6 Chloroflexi bacterium ADurbBin325
A0A1G1A697 Lentisphaerae bacterium RIFOXYA12 FULL_48_11
F0YTG2 Clostridium sp D5
A0A151BL95 Ca. Bathyarchaeota archaeon B63
Firmicutes
Chloroflexi, Firmicutes, PVC
Armatimonadetes, Firmicutes, FCB, PVC
A0A0F8XYB2 Ca. Lokiarchaeum sp GC14 75
A0A0M2U450 Clostridiales bacterium PH28 bin88
A0A1V5HZJ5 Spirochaetes bacterium ADurbBinA120
Actinobacteria, Chloroflexi, Epsilonbacteraeota, Gammaproteobacteria, Nitrospinae, Spirochaetes
A0A0A7GEG9 Geoglobus acetivorans
A0A0F7IF56 Geoglobus ahangari
N0BKU3 Archaeoglobus sulfaticallidus PM70 1
A0A101DF56 Archaeoglobus fulgidus
A0A135VWA9 Ca. Thorarchaeota archaeon SMTZ1 83
A0A0S7WEY6 Dehalococcoidia bacterium DG 22
Actinobacteria, Chloroflexi, Epsilonbacteraeota, Firmicutes, FCB, Spirochaetes, Tenericutes
A0A135VQY2 Ca. Thorarchaeota archaeon SMTZ1-45
A0A1Q9MVR6 Ca. Lokiarchaeota archaeon CR_4
A0A1Q9N0T7 Ca. Lokiarchaeota archaeon CR_4
A0A1Q9MTL8 Ca. Lokiarchaeota archaeon CR_4
A0A1Q9MTL9 Ca. Lokiarchaeota archaeon CR_4
Epsilonbacteraeota
Firmicutes, Actinobacteria, PVC
0.4

Suppl. Figure 14: Maximum likelihood phylogenetic analysis of pyruvate-formate lyase superfamily proteins. Phylogenetic tree was estimated using IQ-tree LG+C20+R+F and is based on a protein alignment of 336 aligned sites and includes 1012 sequences. Support values (SH-like approximate likelihood ratio test and ultrafast bootstraps) are not shown whenever one of the values was below 50. Scale bar indicates number of substitutions per site. Arrows point to proteins of organisms that are known to utilize hydrocarbons. Raw data files are available via figshare (see Data availability for more details).

Organohalides:
TCB, trichlorobenzene;
TeCB, tetrachlorobenzene;
PCE, perchloroethylene;
TCE, trichloroethylene;
DCE, dichloroethene;
DCP, dichlorophenol;
VC, vinyl chloride;
DCA, dichloroethane;
3Cl-4-HBA, 3-chloro-4-hydroxybenzoate.

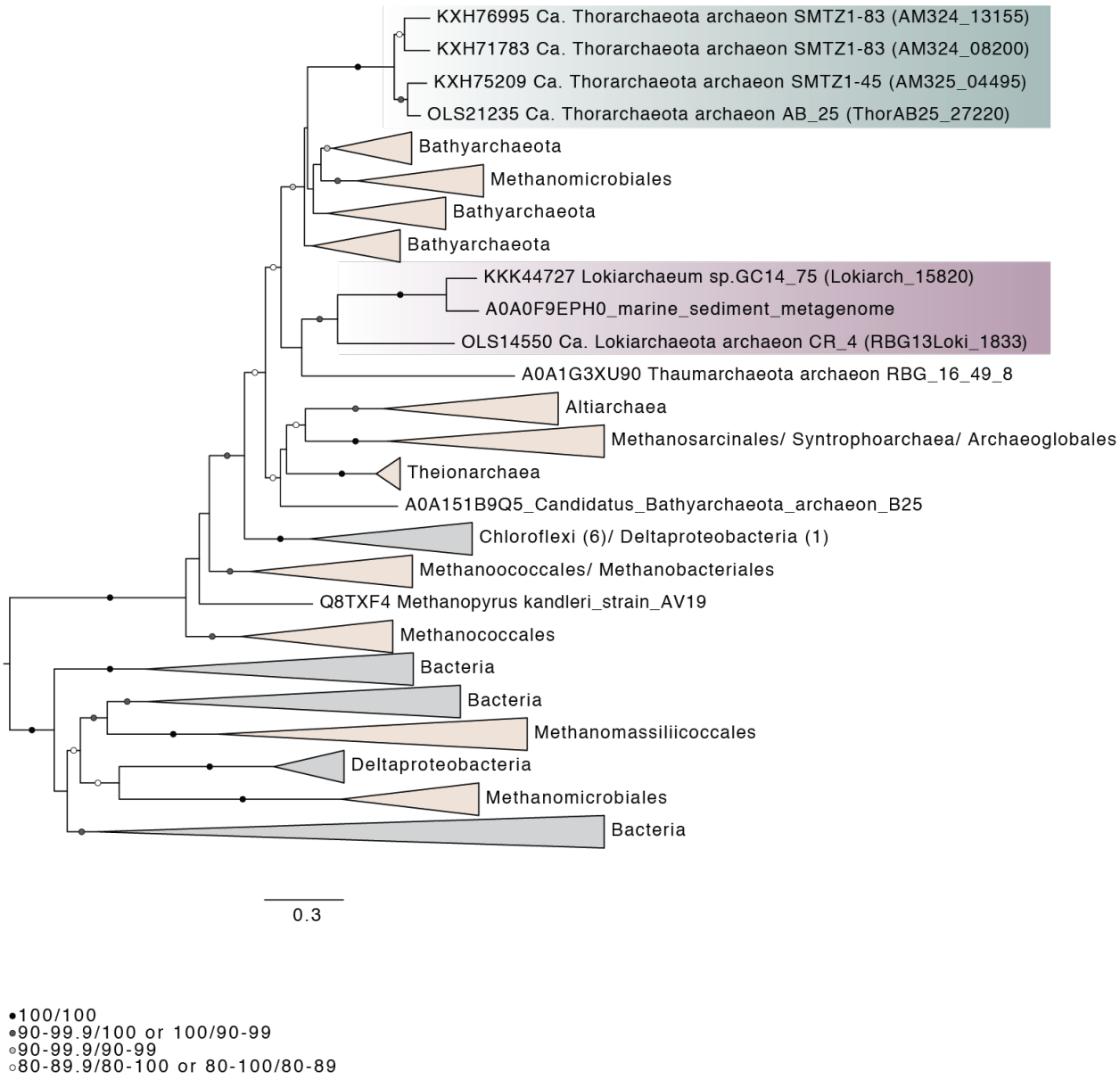


Suppl. Figure 15: Maximum likelihood phylogenetic analysis of reductive dehalogenases. Phylogenetic tree was estimated using IQ-tree LG+C40+R+F and is based on a protein alignment of 160 aligned sites and includes 243 sequences. All support values (SH-like approximate likelihood ratio test and ultrafast bootstraps) below 80 as well as support values of closely related sequences were removed to increase readability and are represented by filled circles according to the color code in the figure legend. Tree was rooted using midpoint rooting. Scale bar indicates the number of substitutions per site. Archaeal clades are shaded in green, bacteria are shaded in violet. Dark green shading refers to archaeal sequences that might represent *bona fide* reductive dehalogenases while the function of sequences highlighted in light green is less clear. The current outgroup represents homologous sequences of Asgard archaea assigned to COG1600 but lacking significant hit to reductive dehalogenase domains (IPR028894). However, notice that some archaeal sequences which encode IPR028894 cluster with these divergent Asgard sequences. The substrates of some know reductive dehalogenases are indicated based on[57,68] and the backbone dataset is based on sequences kindly provided by Laura Hug and described in[59]. Abr.: Organohalides: TCB, trichlorobenzene; TeCB, tetrachlorobenzene; PCE, perchloroethylene; TCE, trichloroethylene; DCE, dichloroethene; DCP, dichlorophenol; VC, vinyl chloride; DCA, dichloroethane; 3Cl-4-HBA, 3-chloro-4-hydroxybenzoate. *This sequence is 99% similar to W6EQP0, for which a X-ray crystal structure has been obtained[56]. Raw data files are available via figshare (see Data availability for more details).



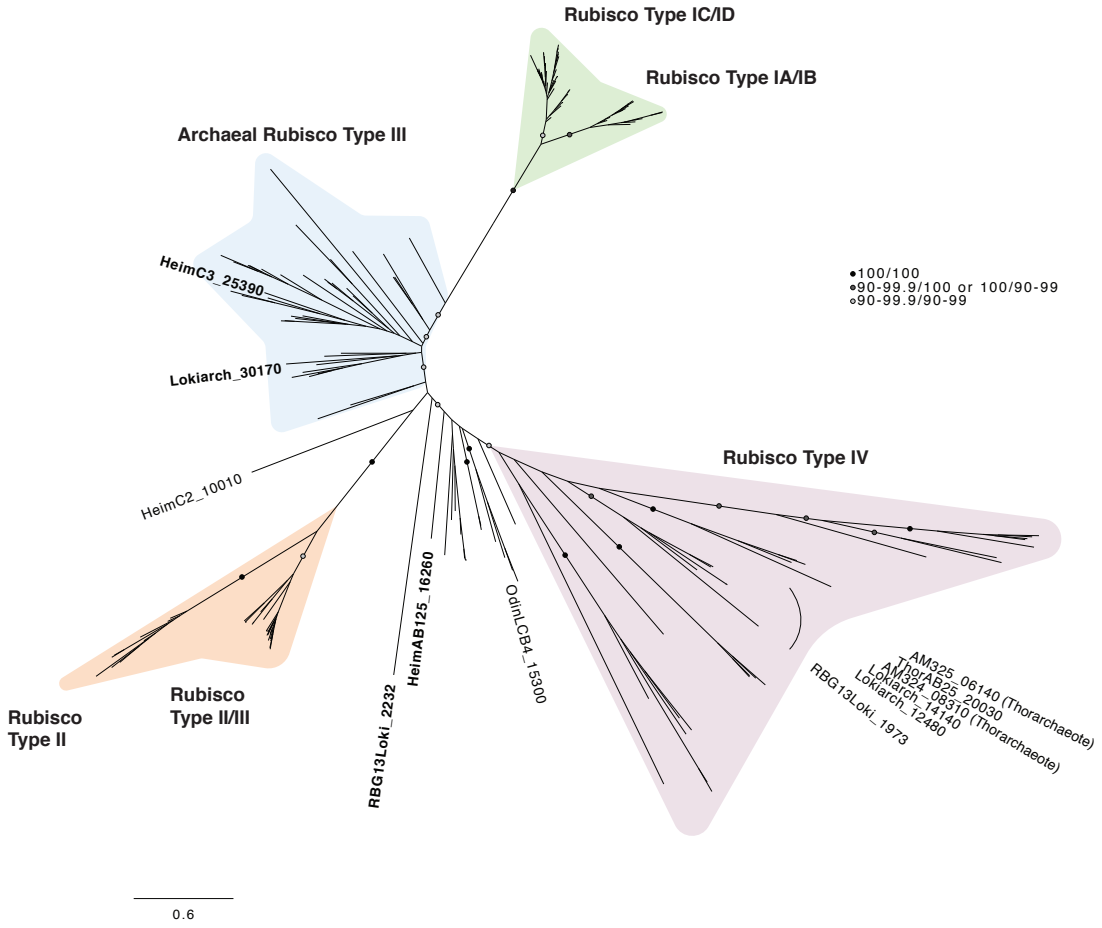
Suppl. Figure 16: Maximum likelihood phylogenetic analysis of Heme-copper-type cytochrome c oxidase, subunit 1 and nitrate reductase, alpha subunit. a, Phylogenetic tree of Heme-copper-type cytochrome c oxidase, subunit 1

was estimated using IQ-tree LG+C20 and is based on a protein alignment of 443 aligned sites and includes 1078 sequences. Only support values (ultrafast bootstraps) above 85 are shown and are represented by filled circles according to the color code in the figure legend. The tree was rooted using the distantly related nitric-oxide reductase (NOR). b, Phylogenetic tree of the alpha-subunit of the bacterial nitrate reductase family (IPR006468) was estimated using IQ-tree LLG+C20+F+R and is based on a protein alignment of 1071 aligned sites and including 233 sequences. Only support values (SH-like approximate likelihood ratio test and ultrafast bootstraps) above 80 are represented by filled circles according to the color code in the Fig. legend. The tree was rooted arbitrarily. a,b Scale bars indicate number of substitutions per site. Raw data files are available via figshare (see Data availability for more details).



Suppl. Figure 17: Maximum likelihood phylogenetic analysis of key subunit of acetyl-CoA synthase/CO dehydrogenase. Phylogenetic tree was estimated for protein sequences with PF03598 encoding the key subunit of Acetyl-CoA synthase/CO dehydrogenase, cdhC (also referred to as CODH/acetyl-CoA synthase beta subunit or alpha subunit of ACS/CODH) using IQ-tree with the LG+C20+F+R. The protein alignment consisted of 712 aligned sites and includes 442 sequences. Only support values (SH-like approximate likelihood ratio test and ultrafast bootstraps) above 80 are represented by filled circles according to the color code in the figure legend. Scale bar indicates number of substitutions per site. Raw data files are available via figshare (see Data availability for more details).

a



b

		57	52	120	172	174	176	195	196	198	200	201	291	292	324	331	376	377	400	401
Type IB	YP_170840 *Synechococcus elongatus* PCC_6301	E	T	N	K	K	G	D	F	K	D	E	H	R	H	K	S	G	G	G
Type II	P50922 *Rhodobacter capsulatus* ATCC11166	E	T	N	K	K	G	D	F	K	D	E	H	R	H	K	S	G	G	G
Type II/III	YP_566926 *Methanococcoides burtonii* DSM_6242	E	T	N	K	K	G	D	F	K	D	E	H	R	H	K	S	G	G	S
Type III	AAB99239 *Methanocaldococcus jannaschii* DSM_2661	E	T	N	K	K	G	D	L	K	D	E	H	R	H	K	S	G	G	G
	BAD86479 *Thermococcus kodakaraensis* KOD1	E	T	N	K	K	G	D	Y	K	D	E	H	R	H	K	S	G	G	G
	OLS23198.1 *Ca.* Heimdallarchaeota archaeon LC_3 (HeimC3_25390)	E	T	N	K	K	G	D	V	K	D	E	H	R	H	K	S	G	G	G
	KKK43062.1 *Ca.* Lokiarchaeum sp. GC14_75 (Lokiarch_30170)	E	T	N	K	K	G	D	L	K	D	E	H	R	H	K	S	G	G	G
	OLS27808.1 *Ca.* Heimdallarchaeota archaeon LC_2 (HeimC2_10010)	E	T	N	K	K	G	T	W	K	D	E	H	R	H	K	S	G	G	G
	OLS31478.1 *Ca.* Heimdallarchaeota archaeon AB125 (HeimAB125_16260)	E	T	N	K	K	G	D	F	K	D	E	H	R	H	K	S	G	G	G
	OLS14054.1 *Ca.* Lokiarchaeota archaeon CR_4 (RBG13Loki_2232)	E	T	N	K	K	G	D	V	K	D	E	H	R	H	K	S	G	G	G
	OLS18291.1 *Ca.* Odinarchaeota archaeon LCB_4 (OdinLCB4_15300)	E	T	N	K	K	G	T	S	K	D	E	H	R	H	K	S	G	G	G
	OLS14417.1 *Ca.* Lokiarchaeota archaeon CR_4 (RBG13Loki_1973)	E	T	N	K	K	G	D	I	K	D	E	H	T	L	K	S	G	G	G
	KKK44923.1 *Ca.* Lokiarchaeum sp. GC14_75 (Lokiarch_14140)	E	T	N	K	C	G	D	V	K	D	E	H	M	V	K	S	G	G	G
	KKK45122.1 *Ca.* Lokiarchaeum sp. GC14_75 (Lokiarch_12480)	E	T	N	K	C	G	D	V	K	D	E	H	M	V	K	S	G	G	G
Type IV	KXH74397.1 *Ca.* Thorarchaeota archaeon SMTZ1-45 (AM325_06140)	E	T	N	K	C	Y	D	I	K	D	E	H	M	V	K	S	G	G	A
	KXH71633.1 *Ca.* Thorarchaeota archaeon SMTZ1-83 (AM324_08310)	E	T	N	K	C	Y	D	I	K	D	E	H	M	V	K	S	G	G	A
	OLS27625.1 *Ca.* Thorarchaeota archaeon AB_25 (ThorAB25_20030)*	E	T	N	K	C	Y	D	I	-	-	-	-	-	-	-	-	-	-	-
	CAB13232 *Bacillus subtilis subsp subtilis* str. 168	G	S	K	K	V	G	D	L	K	D	E	H	P	L	S	S	A	G	G

*partial

Suppl. Figure 18: Maximum likelihood phylogenetic analysis of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) and analysis of essential sites. a, Phylogenetic tree was estimated using IQ-tree LG+C60+R+F and is based on a protein alignment of 359 aligned sites and includes 171 sequences. All support values (SH-like approximate likelihood ratio test and ultrafast bootstraps) below 90 as well as support values of closely related sequences were removed to increase readability and are represented by filled circles according to the color code in the figure legend. Scale bar indicates number of substitutions per site. b, Alignment positions of active site residues[69] present in Asgard homologues as compared to representative species for each known RuBisCO-type. Amino acids, which are given as single-letter IUPAC code, are numbered according to *Synechococcus elongatus* PCC 6301 reference sequence. Raw data files are available via figshare (see Data availability for more details).

Suppl. File 1: Phylogenetic analyses of universal marker proteins. Maximum-likelihood analysis based on a concatenated set of 56 universal protein markers for archaeal representatives of all major clades (144 taxa, 6453 sites) using IQ-TREE under the LG+C60+F+G+PMSF model. The tree is rooted on the DPANN. Numbers at branches indicate bootstrap statistical support (100 replicates). Raw data files are available via figshare (see Data availability for more details).

(Heimdallarchaeote_AB_125:0.5119457884,((((((((((Candidatus_Bathyarchaeota_archaeon_B25:0.4531130314,Candidatus_Bathyarchaeota_archaeon_B24:0.4731823306)72:0.0400240000,miscellaneous_Crenarchaeota_group_archaeon_SMTZ_80:0.6020819648)67:0.0237290000,((miscellaneous_Crenarchaeota_group_6_archaeon_AD8_1:0.4186633848,(Candidatus_Bathyarchaeota_archaeon_B26_2:0.2162053929,(Candidatus_Bathyarchaeota_archaeon_BA2:0.1488715737,Candidatus_Bathyarchaeota_archaeon_BA1:0.1587608596)100:0.1023070000)75:0.0474330000)100:0.1125940000,((miscellaneous_Crenarchaeota_group_15_archaeon_DG_45:0.2178359671,Thaumarchaeota_archaeon_SCGC_AB_539_E09:0.2870999698)100:0.1023590000,Candidatus_Bathyarchaeota_archaeon_B23:0.2427957722)100:0.3052240000)100:0.0651900000)98:0.0659720000,(((((Cenarchaeum_symbiosum:0.2528176005,Nitrosopumilus_maritimus:0.1505108344)100:0.0882820000,Candidatus_Nitrosotenuis_cloacae:0.1798435586)100:0.0812630000,Candidatus_Nitrosotalea_devanaterra:0.1815256419)100:0.2647440000,(Ca_Nitrososphaera:0.0788246241,Nitrososphaera_viennensis_EN76:0.1004829283)100:0.2441430000)100:0.5402010000,((Aigarchaeota_archaeon_JGI_0000106_J15:0.3605721576,((Aigarchaeota_archaeon_OS1:0.0367179819,Aigarchaeota_archaeon_SCGC_AAA471_B22:0.0617268900)100:0.2609830000,Aigarchaeota_archaeon_SCGC_AAA471_E14:0.3729145924)100:0.1671580000)90:0.0451010000,Candidatus_Caldiarchaeum_subterraneum:0.4441103811)100:0.3178320000)98:0.0696250000)100:0.0619260000,(((Geoarchaeon_NAG1:0.2720989573,Geo_AAA471:0.2534452452)100:0.5434200000,(((Sulfolobus_acidocalcarius:0.2319193431,Metallosphaera_cuprina:0.2972266276)100:0.3617150000,(((Pyrolobus_fumarii:0.2051900286,Aeropyrum_pernix:0.4018732523)84:0.0565330000,Ignicoccus_hospitalis:0.3559141863)98:0.0452870000,Desulfurococcus_kamchatkensis:0.3921172805)100:0.0606660000)100:0.1281290000,(((Pyrobaculum_aerophilum:0.1815051956,Thermoproteus_uzoni:0.1880619751)100:0.1580240000,(Caldivirga_maquilingensis:0.3822967131,Vulcanisaeta_distri:0.2599948501)100:0.0723910000)100:0.2475740000,Thermofilum_pendens:0.4584292309)100:0.1184660000)84:0.0511610000)94:0.0320470000,((Candidatus_Methanosuratus_petracarbonis_V5:0.0090453197,Candidatus_Methanosuratus_petracarbonis_V4:0.0000026919)100:0.1482610000,((Candidatus_Methanomethylicus_mesodigestum_V1:0.0046958387,Candidatus_Methanomethylicus_mesodigestum_V2:0.0005984168)100:0.0441760000,Candidatus_Methanomethylicus_oleusabulum_V3:0.0368734041)100:0.1757180000)100:0.5020500000)99:0.0618830000)68:0.0321740000,((Kor1:0.5365366568,(Kor4_v2.filtered:0.3036017157,(Kor2:0.2238184972,Ca_Korarchaeum:0.2962272840)100:0.0724230000)100:0.1627930000)100:0.0883790000,Kor3:0.5793397151)100:0.2823680000)91:0.0414900000,(((((((((archaeon_GW2011_AR11:0.4966087548,archaeon_GW2011_AR3:0.4550070139)84:0.0782380000,(archaeon_GW2011_AR4:0.5813326123,archaeon_GW2011_AR15:0.5035291483)62:0.0530790000)69:0.0369600000,archaeon_GW2011_AR9:0.7132641404)100:0.1073650000,((archaeon_GW2011_AR20:0.4855603253,Nanoarchaeota_archaeon_SCGC_AAA011_G17:0.4730819495)70:0.0793580000,(archaeon_GW2011_AR17:0.6073917730,archaeon_GW2011_AR18:0.6214375403)59:0.0442480000)100:0.0620860000)99:0.0789890000,(((archaeon_GW2011_AR13:0.3102680013,archaeon_GW2011_AR19:0.3025266702)42:0.0354230000,Nanoarchaeota_archaeon_SCGC_AAA011_D5:0.2702275719)100:0.5334460000,archaeon_GW2011_AR6:0.6534922299)100:0.3003840000)100:0.0955910000,(Nanoarchaeum_equitans_Kin4_M:0.5414425351,Candidatus_Nanopusillus_acidilobi:0.7000242570)100:0.3461100000)94:0.0483270000,Candidatus_Parvarchaeum_acidiphilum_ARMAN_4:1.4189863069)100:0.0545450000,(((((Nanohaloarchaea_archaeon_SG9:0.1948546983,(Nanohaloarchaea_archaeon_PL_Br10_U2g27:0.1570703854,(Nanohaloarchaea_archaeon_B1_Br10_U2g29:0.1299062234,Candidatus_Haloredivivus_sp_G17:0.1448038422)69:0.0368570000)100:0.1222030000)89:0.0362240000,Candidatus_Nanosalina_sp_J07AB43:0.2904755147)96:0.0364810000,(Nanohaloarchaea_archaeon_PL_Br10_U2g16:0.1879147201,Nanohaloarchaea_archaeon_B1_Br10_U2g1:0.1747657297)100:0.0640860000)96:0.0780700000,Candidatus_Nanosalinarum_sp_J07AB56:0.4245938503)100:0.8025340000,(archaeon_GW2011_AR5:0.7784679340,((Candidatus_Aenigmarchaeota_archaeon_JGI_0000106_F11:0.3480567321,Candidatus_Micrarchaeota_archaeon_RBG_16_49_10:0.4891597952)1

00:0.2345570000,(Candidatus_Aenigmarchaeum_subterraneum_SCGC_AAA011_O16:0.4175100808,Candidatus_Micrarchaeota_archaeon_RBG_16_36_9:0.3305634076)100:0.3070420000)87:0.0558810000)100:0.1104580000)100:0.0636050000)96:0.0381950000,(((((archaeon_GW2011_AR21:0.3689632658,Diapherotrites_archaeon_SCGC_AAA011_K09_6G:0.4743490602)73:0.0589080000,Diapherotrites_archaeon_SCGC_AAA011_N19_5G:0.5635857648)42:0.0606430000,archaeon_GW2011_AR10:0.4108879722)46:0.0932110000,Nano_AAA011:0.4849340888)100:0.4424210000,Candidatus_Micrarchaeum_acidiphilum_ARMAN_2:1.3613198185)97:0.0834780000)100:0.0494900000,(((Candidatus_Altiarchaeales_archaeon_IMC4:0.4270153522,Candidatus_Altiarchaeales_archaeon_WOR_SM1_SCG:0.3334887534)97:0.0733080000,Candidatus_Altiarchaeales_archaeon_WOR_SM1_86_2:0.3783174389)100:0.0978190000,(Candidatus_Altiarchaeum_sp_CG2_30_32_3053:0.1037902554,Altiarchaeum_SM1_MSI:0.1029317431)100:0.7956100000)100:0.2627390000)100:0.0383750000,(((((Hadesarchaea_archaeon_YNP_45:0.2456273360,Hadesarchaea_archaeon_DG_33:0.1787536205)99:0.0645250000,(Hadesarchaea_archaeon_DG_33_1:0.2448048479,Hadesarchaea_archaeon_YNP_N21:0.2000892364)70:0.0312690000)100:0.3482460000,(((Arc_I_group_archaeon_B15fssc0709_Meth_Bin003:0.0448513845,Arc_I_group_archaeon_U1lsi0528_Bin089:0.1083845831)100:0.4109600000,Thermoplasmatales_archaeon_DG_70:0.4393983001)100:0.1433250000,(Pyrococcus_furiosus:0.0807378113,Thermococcus_kodakarensis:0.0883115265)100:0.3189490000)99:0.0461040000)79:0.0271940000,((Methanococcus_maripaludis:0.3305886271,Methanocaldococcus_jannaschii:0.1743835270)100:0.2712160000,(((Methanosphaera_stadtmanae:0.2717423870,Methanobacterium_AL:0.1476723137)100:0.0955320000,Methanothermobacter_thermauto:0.1126648235)100:0.1433530000,Methanothermus_ferv:0.1821868526)100:0.2012670000)87:0.0430680000)94:0.0305370000,(((((Methanosaeta_thermophila:0.3907974225,(Methanosarcina_acetivorans:0.3460832542,Candidatus_Methanoperedens_nitroreducens:0.3492219008)100:0.0618860000)99:0.0514910000,(Methanocella_paludicola:0.4338084092,((Methanocorpusculum_labreanum:0.3378774428,(Methanoculleus_marisnigri:0.2102787828,Methanoplanus_petrolearius:0.2290565846)98:0.0573700000)100:0.2484840000,(Haloferax_volcanii:0.1741280583,Haloarcula_marismortui:0.1836330649)100:0.5455300000)86:0.0553610000)45:0.0320800000)63:0.0456490000,(uncultured_archaeon_ANME_1:0.4963718545,(Candidatus_Syntrophoarchaeum_caldarius:0.1397866544,Candidatus_Syntrophoarchaeum_butanivorans:0.0980292217)100:0.2650430000)100:0.0776360000)100:0.0971060000,(Ferroglobus_placidus:0.1241508911,Archaeoglobus_fulgidus:0.1685411238)100:0.3091740000)100:0.0799120000,(((Thermoplasma_acidoph:0.2832671403,Ferroplasma_acidarmanus:0.3767134251)100:0.4018670000,Aciduliprofundum_boonei:0.2991713160)100:0.1175000000,(Methanomassiliicoccus:0.4740189710,(Thermoplasmatales_archaeon_SM1_50:0.2641616620,Thermoplasmatales_archaeon_SG8_52_3:0.2039809914)100:0.2867050000)91:0.0846780000)100:0.1363960000)100:0.0738590000)96:0.0268940000)100:0.1050430000)78:0.0342760000,((((Thorarchaeote_AB_25:0.0529313888,(Thor45:0.0204144184,Candidatus_Thorarchaeota_archaeon_SMTZ_45:0.0327720733)100:0.0398360000)100:0.0750040000,Thor83:0.1320739741)100:0.5338300000,Odinarchaeote_LCB4:0.5164121338)94:0.0433970000,(Lokiarchaeum_GC14_75:0.6968888177,Lokiarchaeote_CR_4:0.5709798690)100:0.2487610000)100:0.0504240000)90:0.0310000000,(((((Plasmodium_falciparum:0.3641438771,Tetrahymena_thermophila:0.5012707949)82:0.0567880000,((((Chlamydomonas_reinhardtii:0.1853510417,Arabidopsis_tha:0.1649083532)100:0.0845610000,(Bigelowiella_natans:0.3482763535,Tpseudonana:0.3345170553)87:0.0440040000)76:0.0255950000,(Saccharomyces_cerev:0.3515848858,Homo_sapiens:0.1965011570)99:0.0527320000)51:0.0275410000,(Dictyostelium_dis:0.3371247249,Thecamonas_trahens:0.3295523016)55:0.0424430000)84:0.0365720000)66:0.0352300000,(Naegleria_gruberi:0.4297535442,Leishmania_infantum:0.5017477514)77:0.0569910000)68:0.0588530000,Entamoeba_histolytica:0.4846889392)100:0.1100600000,Trichomonas_vaginalis:0.6764020263)100:0.9057540000)100:0.0804370000,(Heimdallarchaeote_LC_3:0.6004704979,SChinaSea_Heimdall:0.4926646175)100:0.2158920000)100:0.1665880000,Heimdallarchaeote_LC_2:0.7012944200);

Suppl. File 2: Phylogenetic analyses of universal marker proteins after removal of fast evolving sites. Maximum-likelihood analysis is based on a concatenated set of 56 universal protein markers, after removal of the fastest-evolving sites (144 taxa, 5020 sites) using IQ-TREE under the LG+C60+F+G+PMSF model. Numbers at branches indicate bootstrap statistical support (100 replicates). Raw data files are available via figshare (see Data availability for more details).

(Heimdallarchaeote_AB_125:0.2928752461,(((((((Candidatus_Bathyarchaeota_archaeon_B25:0.2576332972,Candidatus_Bathyarchaeota_archaeon_B24:0.2526796273)96:0.019728,miscellaneous_Crenarchaeota_group_archaeon_SMTZ_80:0.3263807535)95:0.017056,(((miscellaneous_Crenarchaeota_group_6_archaeon_AD8_1:0.2367516196,(Candidatus_Bathyarchaeota_archaeon_BA2:0.0861171418,Candidatus_Bathyarchaeota_archaeon_BA1:0.0974642542)100:0.047737)87:0.023556,Candidatus_Bathyarchaeota_archaeon_B26_2:0.1161076513)100:0.071130,((miscellaneous_Crenarchaeota_group_15_archaeon_DG_45:0.1238104185,Thaumarchaeota_archaeon_SCGC_AB_539_E09:0.1504660449)100:0.064200,Candidatus_Bathyarchaeota_archaeon_B23:0.1406423645)100:0.168348)100:0.033258)96:0.032607,(((((Cenarchaeum_symbiosum:0.1423967759,Nitrosopumilus_maritimus:0.0880727544)100:0.052961,Candidatus_Nitrosotenuis_cloacae:0.1066986373)100:0.053142,Candidatus_Nitrosotalea_devanaterra:0.1108124325)100:0.154594,(Ca_Nitrososphaera:0.0414908913,Nitrososphaera_viennensis_EN76:0.0603424390)100:0.142028)100:0.298809,((Aigarchaeota_archaeon_JGI_0000106_J15:0.1885502863,((Aigarchaeota_archaeon_OS1:0.0262797011,Aigarchaeota_archaeon_SCGC_AAA471_B22:0.0432531308)100:0.143105,Aigarchaeota_archaeon_SCGC_AAA471_E14:0.2030318584)100:0.092678)97:0.025985,Candidatus_Caldiarchaeum_subterraneum:0.2432666605)100:0.195677)95:0.038464)100:0.038334,(((Kor1:0.2881651482,(Kor4_v2.filtered:0.1742037378,(Kor2:0.1284213180,Ca_Korarchaeum:0.1586226874)100:0.038116)100:0.101010)100:0.057113,Kor3:0.3103890041)100:0.167677,(((Geoarchaeon_NAG1:0.1550803276,Geo_AAA471:0.1430709836)100:0.300106,(((Sulfolobus_acidocalcarius:0.1334921306,Metallosphaera_cuprina:0.1644045996)100:0.204357,(((Pyrolobus_fumarii:0.1178680447,Aeropyrum_pernix:0.2278035065)97:0.030335,Ignicoccus_hospitalis:0.2079668745)100:0.024511,Desulfurococcus_kamchatkensis:0.2350887965)100:0.036425)100:0.080812,(((Pyrobaculum_aerophilum:0.1043569862,Thermoproteus_uzoni:0.1135277736)100:0.093661,(Caldivirga_maquilingensis:0.2147770111,Vulcanisaeta_distri:0.1517204874)100:0.039938)100:0.143774,Thermofilum_pendens:0.2432647327)100:0.067185)95:0.026303)99:0.015357,((Candidatus_Methanosuratus_petracarbonis_V5:0.0063215991,Candidatus_Methanosuratus_petracarbonis_V4:0.0000025546)100:0.076680,((Candidatus_Methanomethylicus_mesodigestum_V1:0.0028693107,Candidatus_Methanomethylicus_mesodigestum_V2:0.0001875353)100:0.027948,Candidatus_Methanomethylicus_oleusabulum_V3:0.0177341150)100:0.095345)100:0.270462)99:0.029900)83:0.019305)98:0.029100,(((((Thorarchaeote_AB_25:0.0297522209,(Thor45:0.0124677315,Candidatus_Thorarchaeota_archaeon_SMTZ_45:0.0197536180)100:0.022932)100:0.037425,Thor83:0.0795752878)100:0.288900,Odinarchaeote_LCB4:0.2786545193)88:0.018743,(Lokiarchaeum_GC14_75:0.4027761540,Lokiarchaeote_CR_4:0.3345673018)100:0.139449)100:0.025326,(((((Plasmodium_falciparum:0.2085223709,Tetrahymena_thermophila:0.3194920166)73:0.034773,(((Chlamydomonas_reinhardtii:0.1102354254,Arabidopsis_tha:0.0993637014)100:0.055215,(Bigelowiella_natans:0.2053383662,Tpseudonana:0.2066663550)99:0.028938)79:0.016541,((Saccharomyces_cerev:0.2156302829,Homo_sapiens:0.1224174531)94:0.027732,(Dictyostelium_dis:0.2083541297,Thecamonas_trahens:0.2008817613)62:0.024938)61:0.014422)96:0.024404)68:0.020668,(Naegleria_gruberi:0.2690815171,Leishmania_infantum:0.3024806769)78:0.030179)74:0.039396,Entamoeba_histolytica:0.2968605001)100:0.078008,Trichomonas_vaginalis:0.4194799783)100:0.524230)42:0.010032)67:0.016736,((((((((((archaeon_GW2011_AR11:0.2704708130,archaeon_GW2011_AR3:0.2570247121)96:0.040935,(archaeon_GW2011_AR4:0.3186664594,archaeon_GW2011_AR15:0.2948153160)92:0.029914)100:0.021840,archaeon_GW2011_AR9:0.4048077455)100:0.065613,((archaeon_GW2011_AR20:0.2871980917,Nanoarchaeota_archaeon_SCGC_AAA011_G17:0.2795197136)79:0.046001,(archaeon_GW2011_AR17:0.3455077894,archaeon_GW2011_AR18:0.3572807861)79:0.029573)100:0.029174)100:0.047009,((archaeon_GW2011_AR13:0.1784562909,(Nanoarchaeota_archaeon_SCGC_AAA011_D5:0.1652680619,archaeon_GW2011_AR19:0.1846220301)98:0.027684)100:0.316135,archaeon_GW2011_AR6:0.3757921568)100:0.178003)100:0.060528,(Nanoarchaeum_equitans_Kin4_M:0.3136124966,Candidatus_Nanopusillus_acidilobi:0.

4411257580)100:0.202730)100:0.029227,Candidatus_Parvarchaeum_acidiphilum_ARMAN_4:0.8461739376)100:0.034398,((((Nanohaloarchaea_archaeon_SG9:0.1132825771,(((Nanohaloarchaea_archaeon_PL_Br10_U2g27:0.0945141476,Nanohaloarchaea_archaeon_B1_Br10_U2g29:0.0838380440)99:0.026215,Candidatus_Haloredivivus_sp_G17:0.0857715585)100:0.076165,Candidatus_Nanosalina_sp_J07AB43:0.1706675622)100:0.023863)100:0.033432,(Nanohaloarchaea_archaeon_PL_Br10_U2g16:0.1121395047,Nanohaloarchaea_archaeon_B1_Br10_U2g1:0.1105100764)100:0.042800)100:0.045492,Candidatus_Nanosalinarum_sp_J07AB56:0.2737061602)100:0.451827,(archaeon_GW2011_AR5:0.4375157604,((Candidatus_Aenigmarchaeota_archaeon_JGI_0000106_F11:0.2039825969,Candidatus_Micrarchaeota_archaeon_RBG_16_49_10:0.2799267749)100:0.142718,(Candidatus_Aenigmarchaeum_subterraneum_SCGC_AAA011_O16:0.2597734948,Candidatus_Micrarchaeota_archaeon_RBG_16_36_9:0.2019659819)100:0.156027)99:0.035202)100:0.067332)100:0.042257)99:0.024786,((((archaeon_GW2011_AR21:0.2311695189,Diapherotrites_archaeon_SCGC_AAA011_K09_6G:0.2739890648)100:0.049886,(Diapherotrites_archaeon_SCGC_AAA011_N19_5G:0.3172921306,archaeon_GW2011_AR10:0.1888879368)85:0.037809)83:0.053895,Nano_AAA011:0.2673741530)100:0.263781,Candidatus_Micrarchaeum_acidiphilum_ARMAN_2:0.8154579369)100:0.053567)100:0.028880,(((Candidatus_Altiarchaeales_archaeon_IMC4:0.2342178374,Candidatus_Altiarchaeales_archaeon_WOR_SM1_SCG:0.1744397512)100:0.045033,Candidatus_Altiarchaeales_archaeon_WOR_SM1_86_2:0.2147362419)100:0.057120,(Candidatus_Altiarchaeum_sp_CG2_30_32_3053:0.0737814278,Altiarchaeum_SM1_MSI:0.0662605690)100:0.460538)100:0.150970)100:0.024629,(((((Hadesarchaea_archaeon_YNP_45:0.1432494873,Hadesarchaea_archaeon_DG_33:0.1088459762)95:0.036844,(Hadesarchaea_archaeon_DG_33_1:0.1417478473,Hadesarchaea_archaeon_YNP_N21:0.1134198622)93:0.017617)100:0.191723,(((Arc_I_group_archaeon_B15fssc0709_Meth_Bin003:0.0258586073,Arc_I_group_archaeon_U1lsi0528_Bin089:0.0781824935)100:0.213623,Thermoplasmatales_archaeon_DG_70:0.2300603898)100:0.082953,(Pyrococcus_furiosus:0.0457492873,Thermococcus_kodakarensis:0.0513814250)100:0.172184)100:0.028674)91:0.015003,((Methanococcus_maripaludis:0.1861893675,Methanocaldococcus_jannaschii:0.0968532236)100:0.147919,(((Methanosphaera_stadtmanae:0.1416960487,Methanobacterium_AL:0.0750576720)100:0.049283,Methanothermobacter_thermauto:0.0649599757)100:0.084065,Methanothermus_ferv:0.0986154825)100:0.110005)100:0.026342)100:0.014888,(((((Methanosaeta_thermophila:0.2028158404,(Methanosarcina_acetivorans:0.1957996843,Candidatus_Methanoperedens_nitroreducens:0.1845454941)100:0.030342)100:0.023981,(Methanocella_paludicola:0.2316270532,((Methanocorpusculum_labreanum:0.1799391639,(Methanoculleus_marisnigri:0.1103367772,Methanoplanus_petrolearius:0.1249728363)100:0.035637)100:0.133049,(Haloferax_volcanii:0.0986221119,Haloarcula_marismortui:0.1061408156)100:0.297410)100:0.030385)94:0.018797)93:0.024075,(uncultured_archaeon_ANME_1:0.2729942273,(Candidatus_Syntrophoarchaeum_caldarius:0.0710397049,Candidatus_Syntrophoarchaeum_butanivorans:0.0516405097)100:0.145365)100:0.033684)100:0.052239,(Ferroglobus_placidus:0.0706377564,Archaeoglobus_fulgidus:0.0940691233)100:0.166343)100:0.046167,(((Thermoplasma_acidoph:0.1594383260,Ferroplasma_acidarmanus:0.2068978205)100:0.220409,Aciduliprofundum_boonei:0.1534174837)100:0.062302,(Methanomassiliicoccus:0.2404226722,(Thermoplasmatales_archaeon_SM1_50:0.1508279051,Thermoplasmatales_archaeon_SG8_52_3:0.1149976575)100:0.159494)88:0.040828)100:0.071061)100:0.043855)99:0.018153)100:0.059263)100:0.063213,(Heimdallarchaeote_LC_3:0.3338784722,SChinaSea_Heimdall:0.2708131853)100:0.122024)100:0.093803,Heimdallarchaeote_LC_2:0.3898869217);

Suppl. File 3: Phylogenetic analyses of universal marker proteins after removal of DPANN archaea and fast evolving sites. Maximum-likelihood analysis was based on a concatenated set of 56 universal protein markers, after removal of the fastest-evolving sites, and excluding DPANN representatives (104 taxa, 5008 sites) using IQ-TREE under the LG+C60+F+G+PMSF model. Numbers at branches indicate bootstrap statistical support (100 replicates). Raw data files are available via figshare (see Data availability for more details).

(Heimdallarchaeote_AB_125:0.2753135815,(((((((((Candidatus_Bathyarchaeota_archaeon_B25:0.2336842229,Candidatus_Bathyarchaeota_archaeon_B24:0.2293072611)67:0.0160270000,miscellaneous_Crenarchaeota_group_archaeon_SMTZ_80:0.2994076157)74:0.0127260000,(((miscellaneous_Crenarchaeota_group_6_archaeon_AD8_1:0.2128375546,(Candidatus_Bathyarchaeota_archaeon_BA2:0.0791803903,Candidatus_Bathyarchaeota_archaeon_BA1:0.0856296015)100:0.0381420000)73:0.0221990000,Candidatus_Bathyarchaeota_archaeon_B26_2:0.1032920367)100:0.0643980000,((miscellaneous_Crenarchaeota_group_15_archaeon_DG_45:0.1119863372,Thaumarchaeota_archaeon_SCGC_AB_539_E09:0.1421864268)100:0.0560090000,Candidatus_Bathyarchaeota_archaeon_B23:0.1302738407)100:0.1560470000)100:0.0275500000)89:0.0293320000,(((((Cenarchaeum_symbiosum:0.1256062741,Nitrosopumilus_maritimus:0.0818520869)100:0.0505150000,Candidatus_Nitrosotenuis_cloacae:0.0941423095)100:0.0464360000,Candidatus_Nitrosotalea_devanaterra:0.1025858591)100:0.1463980000,(Ca_Nitrososphaera:0.0380180376,Nitrososphaera_viennensis_EN76:0.0535685788)100:0.1298000000)100:0.2821570000,((Aigarchaeota_archaeon_JGI_0000106_J15:0.1742193317,((Aigarchaeota_archaeon_OS1:0.0239533738,Aigarchaeota_archaeon_SCGC_AAA471_B22:0.0405684018)100:0.1294120000,Aigarchaeota_archaeon_SCGC_AAA471_E14:0.1869180340)100:0.0825500000)77:0.0228150000,Candidatus_Caldiarchaeum_subterraneum:0.2206140559)100:0.1817890000)89:0.0414090000)100:0.0343000000,(((Kor1:0.2651649861,(Kor4_v2.filtered:0.1575735627,(Kor2:0.1131258394,Ca_Korarchaeum:0.1404138858)100:0.0341560000)100:0.0928520000)100:0.0533280000,Kor3:0.2874064018)100:0.1528820000,(((Geoarchaeon_NAG1:0.1461450081,Geo_AAA471:0.1233876504)100:0.2902970000,(((Sulfolobus_acidocalcarius:0.1193094710,Metallosphaera_cuprina:0.1480239482)100:0.1867750000,(((Pyrolobus_fumarii:0.1042088633,Aeropyrum_pernix:0.2039007103)55:0.0248560000,Ignicoccus_hospitalis:0.1875035844)97:0.0215690000,Desulfurococcus_kamchatkensis:0.2113263888)99:0.0298950000)100:0.0732740000,(((Pyrobaculum_aerophilum:0.0949735492,Thermoproteus_uzoni:0.0994199202)100:0.0872270000,(Caldivirga_maquilingensis:0.1901766946,Vulcanisaeta_distri:0.1405652332)97:0.0358030000)100:0.1291230000,Thermofilum_pendens:0.2183540751)100:0.0622840000)58:0.0222840000)99:0.0148690000,((Candidatus_Methanosuratus_petracarbonis_V5:0.0054809690,Candidatus_Methanosuratus_petracarbonis_V4:0.0000025071)100:0.0713590000,((Candidatus_Methanomethylicus_mesodigestum_V1:0.0030118134,Candidatus_Methanomethylicus_mesodigestum_V2:0.0001516040)100:0.0244820000,Candidatus_Methanomethylicus_oleusabulum_V3:0.0163362919)100:0.0853660000)100:0.2470540000)95:0.0255920000)91:0.0168890000)96:0.0286730000,((((((Hadesarchaea_archaeon_YNP_45:0.1329479445,Hadesarchaea_archaeon_DG_33:0.0944241647)95:0.0303230000,Hadesarchaea_archaeon_DG_33_1:0.1319438479)57:0.0181990000,Hadesarchaea_archaeon_YNP_N21:0.1006370334)100:0.1755270000,((Methanococcus_maripaludis:0.1688536747,Methanocaldococcus_jannaschii:0.0900005940)100:0.1317540000,(((Methanosphaera_stadtmanae:0.1280112983,Methanobacterium_AL:0.0668465289)100:0.0462100000,Methanothermobacter_thermauto:0.0557823216)100:0.0769810000,Methanothermus_ferv:0.0877605990)100:0.1014160000)97:0.0197300000)53:0.0071820000,(((Arc_I_group_archaeon_B15fssc0709_Meth_Bin003:0.0232475234,Arc_I_group_archaeon_U1lsi0528_Bin089:0.0716052256)100:0.2018200000,Thermoplasmatales_archaeon_DG_70:0.2031773496)100:0.0804710000,(Pyrococcus_furiosus:0.0417240096,Thermococcus_kodakarensis:0.0458043458)100:0.1542150000)100:0.0297100000)84:0.0130200000,(((((Methanosaeta_thermophila:0.1834352438,(Methanosarcina_acetivorans:0.1801134713,Candidatus_Methanoperedens_nitroreducens:0.1705921165)90:0.0237020000)99:0.0214860000,(Methanocella_paludicola:0.2120856965,((Methanocorpusculum_labreanum:0.1658945587,(Methanoculleus_marisnigri:0.0959918219,Methanoplanus_petrolearius:0.1132098362)97:0.0292970000)100:0.1210410000,(Haloferax_volcanii:0.0895252802,Haloarcula_marismortui:0.0983743921)100:0.2730120000)95:0.0282950000)92:0.0181260000)96:0.0223730000,(uncultured_archaeon_ANME_1:0.2502074410,(Candidatus_Syntrophoarchaeum_caldarius:0.0652268066,Candidatus_Syntrophoarchaeum_buta

nivorans:0.0457084893)100:0.1295230000)98:0.0294570000)100:0.0477320000,(Ferroglobus_placidus:0.0624801670,Archaeoglobus_fulgidus:0.0835863236)100:0.1508200000)100:0.0397200000,(((Thermoplasma_acidoph:0.1478047480,Ferroplasma_acidarmanus:0.1871859881)100:0.2056740000,Aciduliprofundum_boonei:0.1391464850)100:0.0582040000,(Methanomassiliicoccus:0.2232765614,(Thermoplasmatales_archaeon_SM1_50:0.1345921651,Thermoplasmatales_archaeon_SG8_52_3:0.1101539943)100:0.1414890000)88:0.0409040000)100:0.0701090000)100:0.0369440000)100:0.0719980000)59:0.0121050000,(((Thorarchaeote_AB_25:0.0234233586,(Thor45:0.0107773770,Candidatus_Thorarchaeota_archaeon_SMTZ_45:0.0188485947)97:0.0200180000)100:0.0343690000,Thor83:0.0709674518)100:0.2760890000,(Odinarchaeote_LCB4:0.2535515008,(Lokiarchaeum_GC14_75:0.3908516406,Lokiarchaeote_CR_4:0.3151145991)100:0.1123350000)63:0.0096020000)99:0.0238650000)82:0.0165220000,((((Plasmodium_falciparum:0.1909199990,Tetrahymena_thermophila:0.2960092071)67:0.0273520000,((((Chlamydomonas_reinhardtii:0.1015493084,Arabidopsis_tha:0.0856689814)100:0.0444510000,(Bigelowiella_natans:0.1836013735,Tpseudonana:0.1914315382)88:0.0253360000)71:0.0161450000,((Saccharomyces_cerev:0.1956095640,Homo_sapiens:0.1085922128)83:0.0228620000,(Dictyostelium_dis:0.1917619524,Thecamonas_trahens:0.1844618962)48:0.0192750000)41:0.0124220000)61:0.0187360000,(Entamoeba_histolytica:0.2783573085,Leishmania_infantum:0.2891016899)65:0.0345450000)42:0.0132550000)52:0.0226570000,Naegleria_gruberi:0.2508036160)100:0.0697630000,Trichomonas_vaginalis:0.4057911732)100:0.5014970000)100:0.0401890000,(Heimdallarchaeote_LC_3:0.3151418784,SChinaSea_Heimdall:0.2473933463)100:0.1156070000)100:0.0862280000,Heimdallarchaeote_LC_2:0.3630172997);