ARTICLES
https://doi.org/10.1038/s41564-019-0406-9
nature microbiology

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

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

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 Lokiarchaeota, Thorarchaeota, Odinarchaeota and Heimdallarchaeota suggested to comprise the closest archaeal relatives of eukaryotes, has helped to elucidate the identity of the putative archaeal host. Whereas 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 Thorarchaeota and Lokiarchaeota encode proteins necessary for carbon fixation via the Wood–Ljungdahl pathway and for obtaining reducing equivalents from organic substrates. By 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 reoxidation 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.

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 refs. [6–8]). In particular, detailed phylogenomic analyses of Asgard archaea, which currently include Lokiarchaeota, Thorarchaeota, Odinarchaeota and Heimdallarchaeota, have suggested that these archaea represent the closest known relatives of eukaryotes[2,4]. Although 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 encoding so-called eukaryotic signature proteins[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 favour hypotheses that underpin a symbiogenetic origin of eukaryotes from only two domains of life[6,8,17,18], the timing of the events leading to eukaryogenesis remains 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 reveal 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 Lokiarchaeota, Thorarchaeota, Odinarchaeota and Heimdallarchaeota[4], suggesting that members of these groups are characterized by different physiological lifestyles.

[1]Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [2]NIOZ, Royal Netherlands Institute for Sea Research, Department of Marine Microbiology and Biogeochemistry, and Utrecht University, AB Den Burg, The Netherlands. [3]Department of Marine Science, University of Texas at Austin, Marine Science Institute, Port Aransas, TX, USA. [4]School of Biological Sciences, Monash University, Clayton, Victoria, Australia. [5]Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, Wageningen, The Netherlands. *e-mail: anja.spang@nioz.nl; thijs.ettema@wur.nl

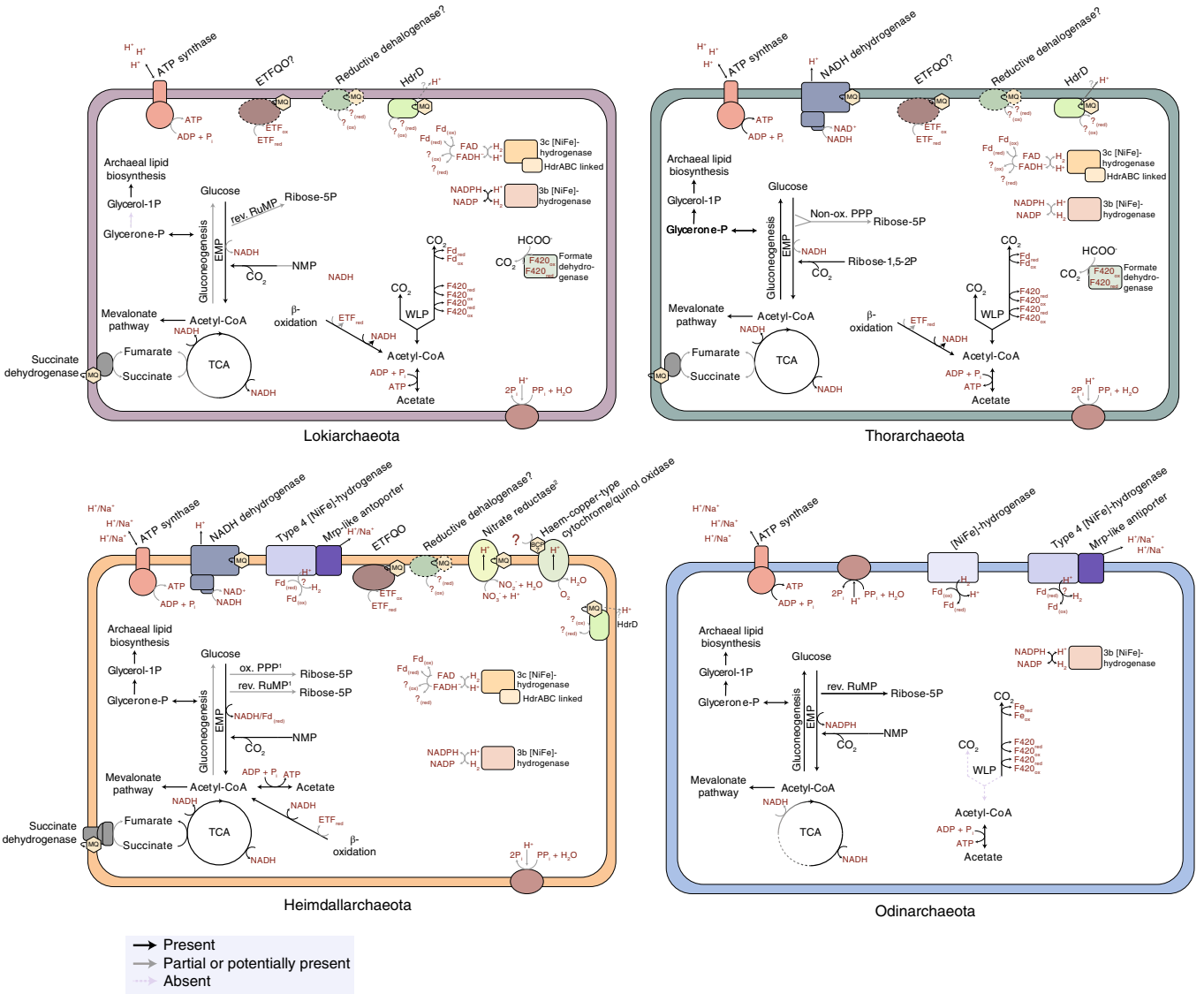

Fig. 1 | Metabolic potential of different Asgard phyla. Each of the arrows represent functions that were assigned to predicted proteins encoded in the respective genomes (Lokiarchaeota (two MAGs), Thorarchaeota (three MAGs), Heimdallarchaeota (three MAGs) and Odinarchaeota (one MAG)). [1]The oxidative pentose phosphate pathway (ox. PPP) is encoded by *Heimdallarchaeum* LC2 and LC3, and the reductive ribulose monophosphate (rev. RuMP) pathway by *Heimdallarchaeum* AB125. [2]The location of the active site of nitrate reductase in Asgard homologues is probably in the cytoplasm (Supplementary Information). The colour code as indicated in the figure key is as follows: the black arrows indicate enzymatic steps encoded by the respective organisms; the solid 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; and the 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 or function of enzyme complexes surrounded by the black dashed lines is tentative. ETF, electron transfer flavoprotein; ETFQO, ETF-ubiquinone oxidoreductase; Fd, ferredoxin; HdrD, heterodisulfide reductase; ox, oxidation; P, phosphate; P_i, inorganic phosphate; PP_i, inorganic pyrophosphate; red, redox; TCA; tricarboxylic acid cycle.

Lokiarchaeota and Thorarchaeota can probably use organic compounds and hydrogen. The Wood–Ljungdahl pathway (WLP) allows the reduction of carbon dioxide to acetyl-coenzyme A (CoA) and can be used to support both autotrophic carbon fixation and, when linked to chemiosmotic processes, energy conservation[20–22]. Both Lokiarchaeota and Thorarchaeota encode all enzymes for a complete (archaeal) WLP[3,13,23] (Fig. 1, Supplementary Fig. 2a, Supplementary Tables 1 and 2 and Supplementary Information). The presence of the WLP, together with group 3b and group 3c [NiFe]-hydrogenases (Fig. 2, Supplementary Table 2 and Supplementary Information), indicate that members of the Lokiarchaeota 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 Lokiarchaeota and Thorarchaeota, which are encoded in gene clusters with soluble heterodisulfide reductase subunits (Supplementary Fig. 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, that are 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,

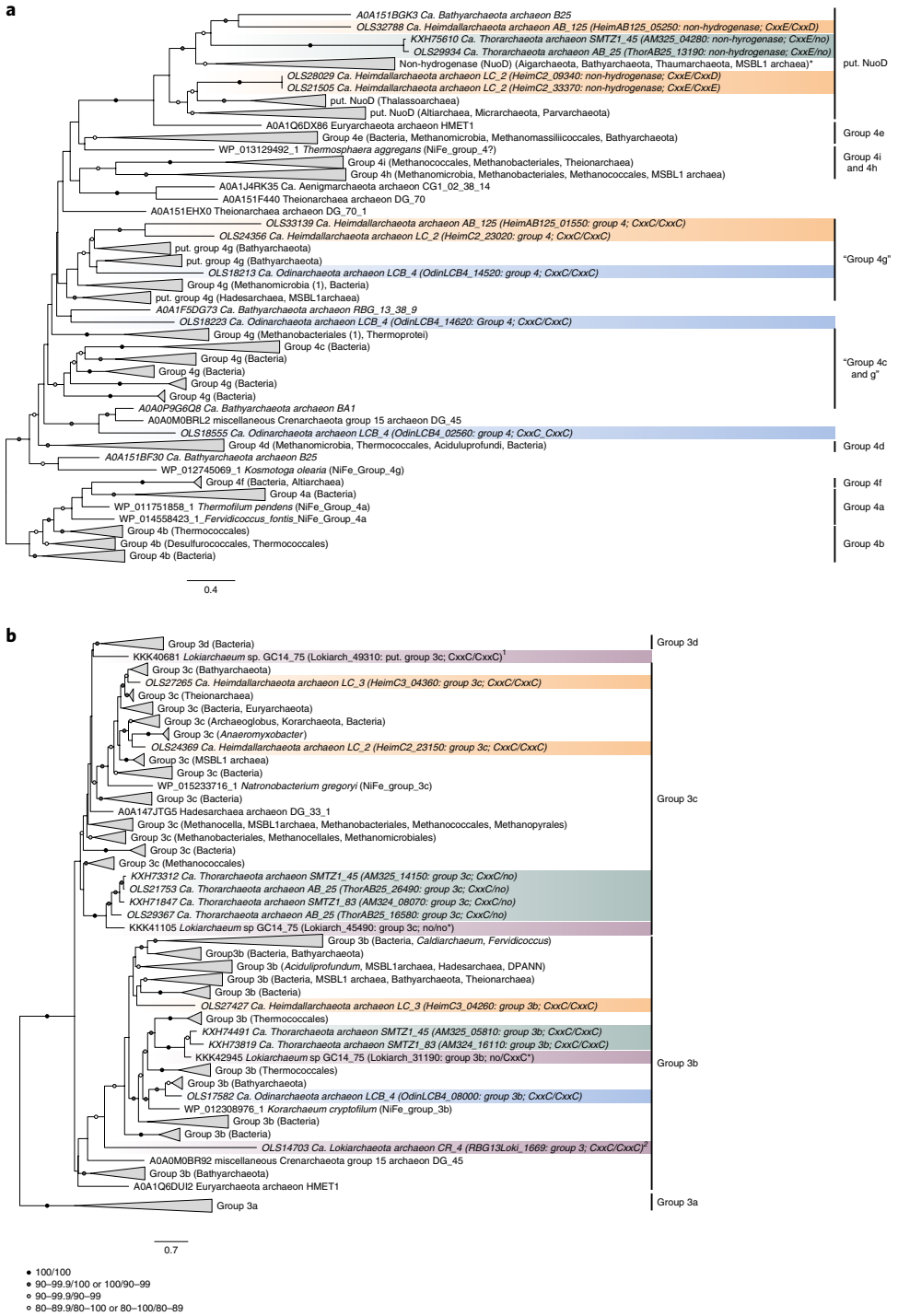

Fig. 2 | Complex evolutionary history of group 3 and group 4 [NiFe]-hydrogenases in Asgard archaea. a,**b**, Maximum-likelihood phylogenetic reconstructions of the large subunit of group 4 (376 sequences) (**a**) and group 3 (543 sequences) (**b**) [NiFe]-hydrogenases (359 and 319 amino acids, respectively). The asterisks indicate partial genes. Asgard homologues are colour-coded as follows: Heimdallarchaeota in orange, Odinarchaeum in blue, Thorarchaeota in green and Lokiarchaeota in purple. Their annotation includes information as to whether the N-terminal and C-terminal CxxC motifs, which ligate H_2-binding metal centres[88], are conserved. The numbers at the branches indicate the SH-like approximate-likelihood ratio test[80] and ultrafast bootstrap support values[79], respectively. Only bootstrap values above 80 are shown. The phylogeny of the large subunit of group 4 [NiFe]-hydrogenases (**a**) was rooted arbitrarily due to the unstable position of the root. The phylogeny of the large subunit of group 3 [NiFe]-hydrogenases (**b**) was rooted between group 3a and all other groups, as determined from a preliminary phylogenetic analysis of the large subunit of [NiFe]-hydrogenases including all members (groups 1–4) (not shown). The scale bars indicate the average number of substitutions per site. [1]Gene cluster analysis of this divergent group 3 [NiFe]-hydrogenase homologue suggests that it represents a group 3c member (Supplementary Fig. 3). Although phylogenetic analysis indicates that this sequence is from a group 3b [NiFe]-hydrogenase, it belongs to a gene cluster with an organization typical of group 3c [NiFe]-hydrogenases. Thus, the annotation of this hydrogenase is currently unclear. put. NuoD, putative NADH-quinone oxidoreductase subunit D.

the metabolic repertoire of Lokiarchaeota and Thorarchaeota reveals the potential to harness electrons from various organic substrates (Supplementary Tables 1 and 2 and Supplementary Information), including complex carbohydrates, peptides, amino acids, alcohols, fatty acids (Supplementary Figs. 2b and 4–13) and hydrocarbons (Supplementary Fig. 14). We also identified putative formate dehydrogenases that may support growth on formate as an energy and/or carbon source. Although 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. In addition, it may be speculated that both Lokiarchaeota 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 for the reoxidation of quinol species generated by electron transfer flavoproteins (Fig. 1 and Supplementary Information). It is possible that cofactors that are reduced during organic carbon oxidation may be reoxidized 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 (Figs. 1 and 2). Thus, depending on the 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 an electron sink similar to heterotrophic acetogenic bacteria[27]. By 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 Lokiarchaeota and Thorarchaeota suggests that these organisms are also able to shuttle electrons to organohalide compounds (Supplementary Fig. 15 and Supplementary 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 (Figs. 1 and 3 and Supplementary Tables 1 and 2). Its genome encodes only a partial tricarboxylic acid cycle and lacks genes for several key enzymes of the WLP and the β-oxidation pathway (Supplementary Fig. 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 (Fig. 2, Supplementary Tables 1 and 2 and Supplementary Information). Phylogenetic analyses of the large subunit of the [NiFe]-hydrogenase suggest that these hydrogenases form three previously undefined subgroups, together with homologues of other recently obtained archaeal lineages (Fig. 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 (Supplementary Fig. 3 and Supplementary Information), 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 favourable 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^+-driven 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 reoxidation, thereby increasing the overall ATP yield.

Heimdallarchaeota may grow heterotrophically by fermentation or by anaerobic and aerobic respiration. Heimdallarchaeota encode a versatile metabolic repertoire (Fig. 1, Supplementary Fig. 2a and Supplementary Tables 1 and 2) with the potential to derive reducing equivalents from a range of organic substrates, including complex carbohydrates, fatty acids and proteins (Supplementary Fig. 2b and Supplementary Information). Although they lack most enzymes for the WLP, both the LC2 and LC3 genomes encode various electron transport chain components (Fig. 1, Supplementary Tables 1 and 2 and Supplementary Information), including an A-type haem-copper oxidase (Supplementary Fig. 16a), a bacterial-type nitrate reductase (Supplementary Fig. 16b) and a respiratory chain complex I. These components probably 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], whereas members of the Thorarchaeota 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%) to that of *Heimdallarchaeum* LC2 (OLS29256), indicating that some Heimdallarchaeota may be able to occupy both anoxic and oxic niches. By 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 reoxidation using H^+ as the electron acceptor (Supplementary Information). Finally, all members of the Heimdallarchaeota may be able to use organohalides as electron acceptors (Supplementary Information) and *Heimdallarchaeum* AB125 may in addition be able to use hydrocarbons as substrates (Supplementary Fig. 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 last Asgard archaeal common ancestor (LAsCA) (Fig. 3a). First, it is likely that LAsCA had the ability to metabolize organic substrates (Supplementary Information). All Asgard archaea, except for *Odinarchaeum*, have a large variety of genes encoding proteins of all steps of the β-oxidation pathway (Supplementary Figs. 5–13 and Supplementary Information). Although the phylogenetic history of the respective enzymes is complex and invokes several horizontal gene transfers (HGTs) from different sources (Supplementary Figs. 5–13 and Supplementary Table 5), at least some of these enzymes (Supplementary Figs. 6b and 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 (Supplementary Fig. 4 and Supplementary Table 4).

The presence of the WLP in both Thorarchaeota and Lokiarchaeota and the wide occurrence of the key enzyme of the

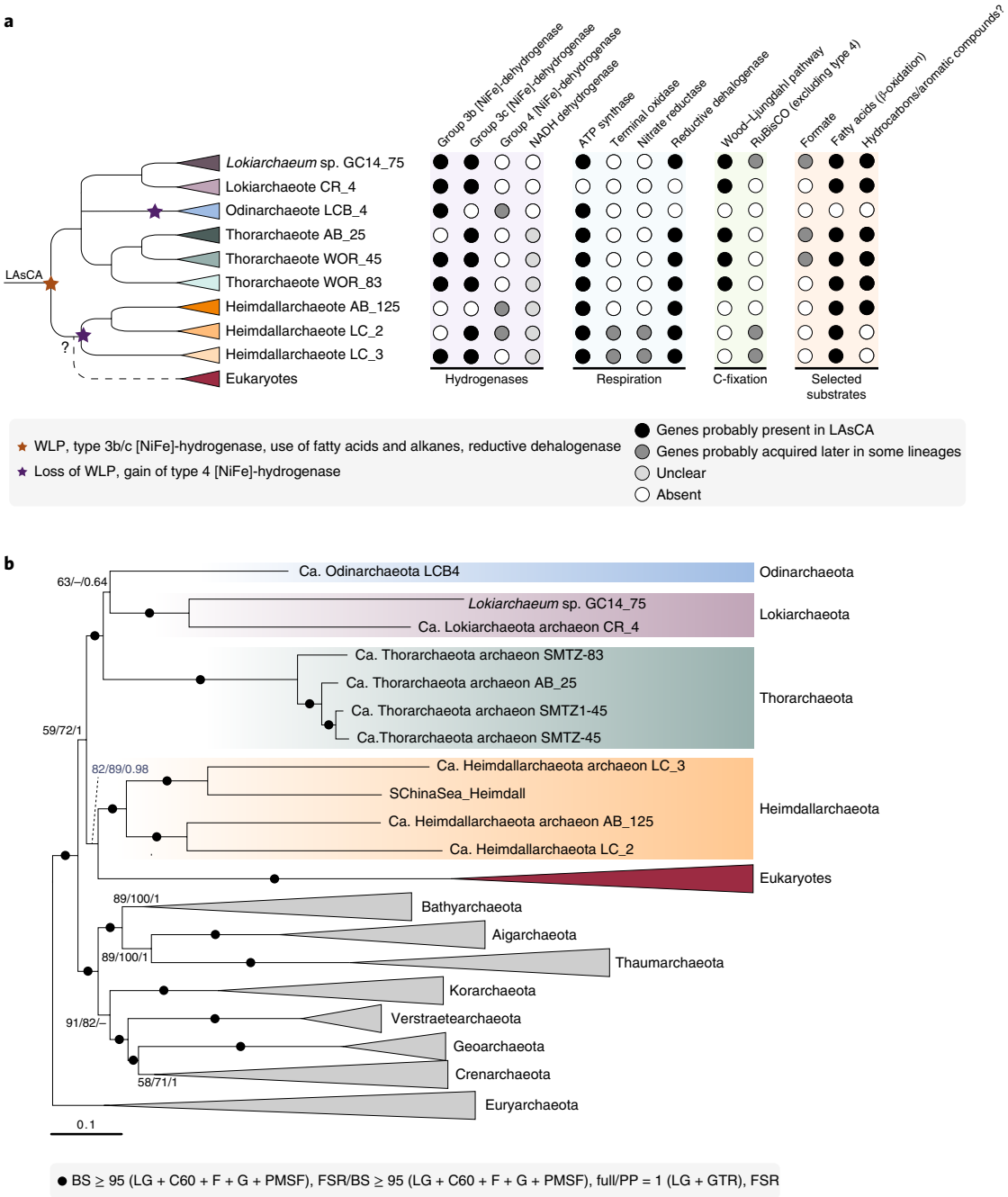

Fig. 3 | Evolution of the Asgard superphylum and selected metabolic features. **a**, Schematic representation of the relationship of the Asgard phyla (nine MAGs) and the distribution of selected metabolic features across the different members of this group. **b**, Maximum-likelihood phylogenetic analysis of a concatenated set of 56 universal protein markers[4] from 144 representative taxa (https://figshare.com/s/5f153d1dcacadd3b3ed6). The numbers at the branches indicate bootstrap (BS) statistical support estimated by IQ-TREE (LG + C60 + F + G + PMSF model) obtained from 100 replicates on the 'full' and 'fast-sites removed' (FSR) data sets, and posterior probabilities (PP) estimated by Phylobayes (LG-GTR model) on the fast-sites removed data set (see Methods for further details).

WLP (carbon monoxide dehydrogenase/acetyl-CoA synthetase) 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 carbon monoxide dehydrogenase/acetyl-CoA synthetase may have been subjected to HGT after the radiation of the Asgard archaea (Supplementary Fig. 17). Furthermore, few to all genes encoding enzymes of the WLP seem to have been lost in *Odinarchaeum* and currently known Heimdallarchaeota, respectively. Although 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, although all Asgard archaea encode a homologue of the ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) family enzymes, which among others include the key enzyme of the Calvin–Benson–Bassham cycle (RuBisCO types 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

homologues belong to RuBisCO subtypes that are unrelated to the carbon fixation cycles and may have been acquired horizontally (Supplementary Information and Supplementary Fig. 18a). In line with this, phosphoribulokinase, another key enzyme of RuBisCO-based carbon fixation pathways, does not seem to be encoded by Asgard genomes (Supplementary Information). Thus, there is currently no indication that LAsCA contained carbon fixation strategies other than the WLP.

Although the wide distribution of cytosolic group 3b and 3c [NiFe]-hydrogenases in the analysed Asgard archaea may indicate their presence in LAsCA, phylogenetic analyses of the key subunit do not support such inferences (Figs. 2 and 3 and Supplementary Tables 1 and 2). For instance, most homologues encoded by the different Asgard lineages are not monophyletic, indicating extensive HGT events throughout the evolution of this archaeal group (Fig. 2). Furthermore, based on the patchy distribution of membrane-bound group 4 [NiFe]-hydrogenases and NADH dehydrogenases in the Asgard genomes and the results of our phylogenetic analyses, it is unclear whether these enzymes were part of the LAsCA proteome (Fig. 3a). Yet, based on the current data, it is plausible that the respective ancestors of Heimdallarchaeota harboured a membrane-bound group 4 [NiFe]-hydrogenase.

Although phylogenetic analyses indicate that putative reductive dehalogenase-like proteins may have been encoded by LAsCA (Supplementary Fig. 15), the absence of a terminal nitrate and oxygen reductase in most of the Asgard archaea studied herein, with the exception of *Heimdallarchaeum* LC2 and LC3, indicates that these enzyme complexes were acquired more recently in Heimdallarchaeota (Fig. 3a). However, the exact timepoint of the acquisition of the genes encoding the A-type haem-copper oxidase and nitrate reductase (Supplementary Fig. 16a,b) remains to be determined. Although the heimdallarchaeal A-type Cox1 homologues branches in a cluster with homologues of *Ca*. Caldiarchaeum subterraneum and Crenarchaeota, we obtained strong support for the monophyly of the eukaryotic and alphaproteobacterial A-type homologues, supporting the view that eukaryotes inherited their terminal oxidases from the bacterial endosymbiont 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 *Methanoperedens* sp. BLZ1 (ref. [36]), which contains both a bacterial and an archaeal-type nitrate reductase, 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 probably 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 (Fig. 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. By 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 Heimdallarchaeota and eukaryotes (Fig. 3a,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, for example refs. [6–8,16,38]. Among these scenarios, the independently formulated but simultaneously published syntrophic hypothesis[39,40] (Fig. 4a) and hydrogen hypothesis[41] (Fig. 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 (Fig. 4a,b). Following the initial discovery of *Lokiarchaeum*[2], a modified version of the hydrogen hypothesis was proposed[13] (Fig. 4c). Based on the presence of the WLP in Lokiarchaeota, it was suggested that the archaeal ancestor of eukaryotes probably represented an autotrophic H_2-dependent organism. The analysis of genomic data of additional Asgard archaea (Fig. 1, Supplementary Tables 1 and 2 and Supplementary Information) including genomic information of members of the Heimdallarchaeota, which currently seems to represent the closest relatives of eukaryotes among the Asgard archaea[2,4] (Fig. 3b), allowed us to refine previous syntrophic scenarios for the origin of eukaryotes.

In our model, referred to as the 'reverse flow model' (Fig. 4d), the direction of the syntrophic interaction is suggested to be opposite of what was proposed for the hydrogen and syntrophic hypotheses[6] (Fig. 4a,b); that is, 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 (Fig. 3a and Supplementary Figs. 4–14). 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 an 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 (Fig. 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 unfavourable anaerobic oxidation of an alkane (for example, butane) or related compounds through the shuttling of electrons to a bacterial partner (similar to anaerobic methane-oxidizing euryarchaeote consortia[45]) (Fig. 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 (Fig. 3b), the sparse distribution of these *mcr* genes within the currently sampled Asgard diversity[46] and the evidence for HGT of *mcr* gene homologues 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

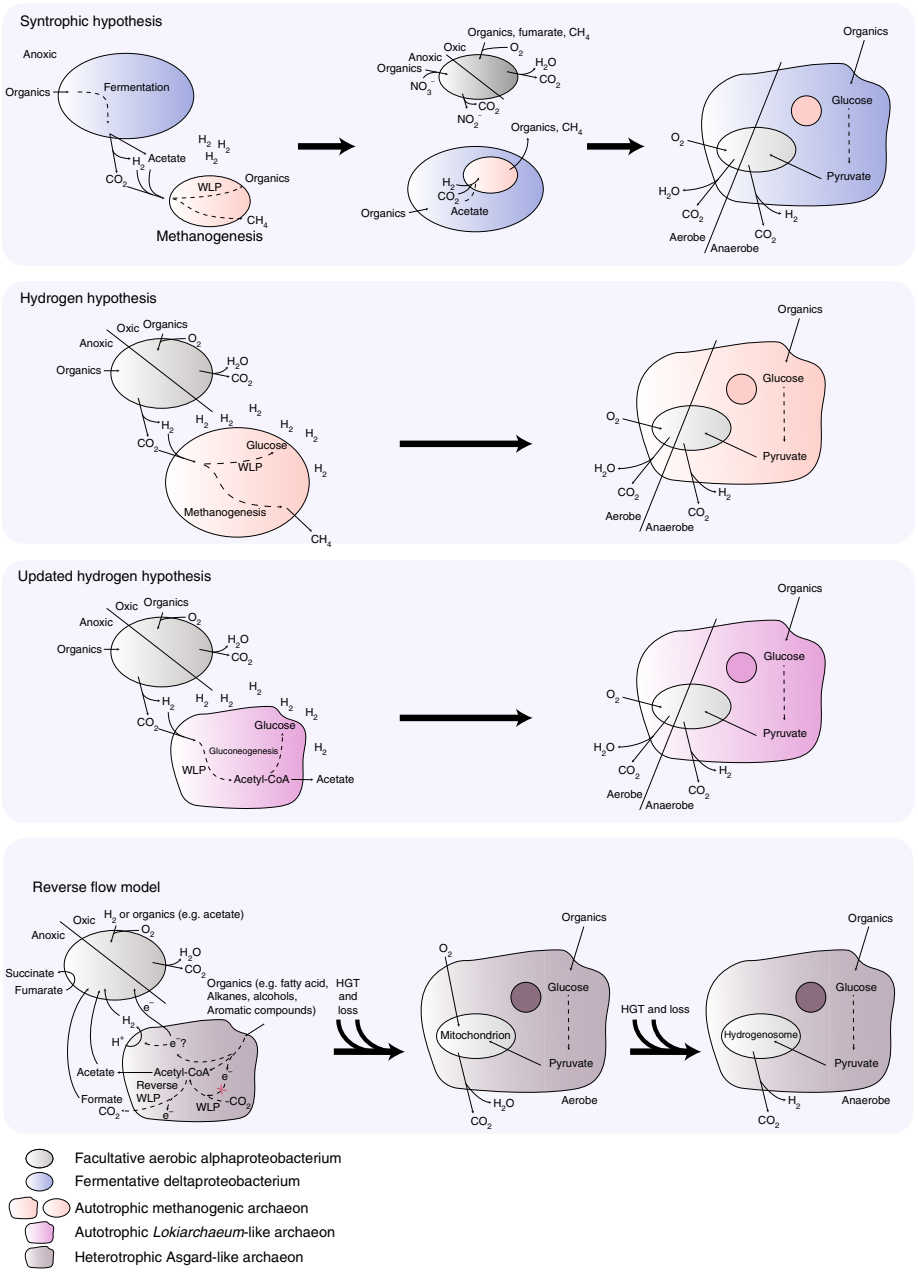

Fig. 4 | Evolutionary scenarios for the origin of the eukaryotic cell. a, Depiction of the syntrophic hypothesis previously proposed by Moreira and Lopez-Garcia in 1998 and 2006 (refs. [39,40]), which invokes two bacterial and one archaeal partner(s) in the origin of the eukaryotic cell; that is, first, 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. Although 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 in 1998 (ref. [41]), which suggests that a symbiosis between a strictly autotrophic hydrogen-dependent methanogenic archaeon and an H_2-producing 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', 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 alphaproteobacteria 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 (for example, hydrocarbons and fatty acids) and the bacterial partner utilized these in the 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-related 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 the main text for details). The dashed arrows represent metabolic pathways involved in the conversion of the respective metabolites.

alphaproteobacteria excluding Magnetococcales[48]. Although 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 therefore allow hydrogenotrophic growth[29]. Some anaerobically functioning mitochondria and related organelles of eukaryotes produce H_2 via [FeFe]-hydrogenases, the origin of which is still debated[49]. In particular, the vast majority of alphaproteobacteria do not encode a [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-related organelles in anaerobic eukaryotes, may have been acquired horizontally from bacterial sources other than the alphaproteobacterial endosymbiont subsequent to eukaryogenesis (Fig. 4d). Alphaproteobacteria have been shown to engage in syntrophic interactions within microbial consortia[51] and to be able to accept electrons from various donors, including from extracellular electrons[52]. We speculate that the alphaproteobacterial ancestor of mitochondria may have served as an electron sink for the archaeal partner either under anoxic conditions (for example, using fumarate or inorganic compounds such as sulfate, or nitrate as an electron acceptor) or micro-oxic conditions (using oxygen as an electron acceptor).

Altogether, and in agreement with previous syntrophic models for eukaryogenesis, our scenario is based on the observation that syntrophic interactions are widespread, in particular, in anoxic environments[42], and examples include methane-oxidizing or butane-oxidizing Euryarchaeota, many of which grow in microbial consortia and are obligately dependent on a partner organism[45]. Although speculative, we propose that the eukaryotic signature proteins encoded by Asgard archaea[2,4] may have aided the transition from a metabolic syntrophy to a more intricate symbiosis. For example, the generation of actin-based cellular protrusions could have led to a tighter interaction between the partners, allowing for more efficient metabolic coupling, with the bacterial symbiont(s) eventually being effectively engulfed by the host cell. The origin of the defining features of eukaryotic cells including the nucleus and the relative timing of their emergence remain to be established[53]. However, it is possible that the acquisition of mitochondria was an intermediate event that occurred after the invention of a basic intermembrane system and actin cytoskeleton but predated the origin of the nucleus[54]. Furthermore, it remains an open question whether membrane lipids could have been of mixed origin during the transition phase[55] and were subsequently replaced by bacterial fatty-acid-type lipids in the outer membrane, following endosymbiont gene transfer[56], gene acquisitions from other sources[57,58] and differential gene loss. Bacterial lipids may have had a selective advantage considering that genes for most metabolic enzymes, including respiratory chains, were retained from the endosymbiont rather than from the host[19] and were probably functioning optimally in bacterial membrane lipids.

Conclusions

Our analyses have revealed that Asgard archaea are metabolically versatile: whereas 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. By contrast, at least some members of the Heimdallarchaeota seem to have the ability to grow by anaerobic and/or aerobic respiration, whereas others are obligate anaerobes. On the basis of these analyses, we propose a refined symbiogenetic scenario for the origin of eukaryotes that is based on the transfer of reducing equivalents from an archaeal host to a bacterial partner. We acknowledge that, although such a reversed electron flow is most easily reconcilable with the gene set of the herein analysed Asgard archaeal MAGs, alternative scenarios (Fig. 4) cannot be excluded. Subsequent metabolic reconstructions of additional members of this group and of the alphaproteobacteria, as well as the functional characterization of eukaryotic signature proteins 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 paper have been annotated automatically and were published in ref. [4]. For metabolic reconstructions, all proteins were analysed with IPRscan (version: interproscan-5.22–61.0) to determine protein domain information (IPR domains and PFAMs) as well as KO (KEGG Orthology) numbers. In addition, proteins were queried against the NCBI non-redundant (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, archaeal clusters of orthologous genes (arCOGs) and clusters of orthologous genes (COGs) were assigned based on the publicly available arCOGs from refs. [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 amino-terminal and carboxy-terminal CxxC motifs characterizing the large subunit of [NiFe]-hydrogenases was determined in Jalview. Finally, all of these data were compiled in a tsv file containing relevant information from all of these analyses (Supplementary Table 3). All proteins discussed throughout the paper are listed in Supplementary Tables 1–5. Please note that, although the Asgard MAGs are of medium-to-high quality[63] (Supplementary Fig. 1), 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 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 cut-off of 1×10^{-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 cut-off of 1×10^{-10} (ref. [20]) (downloaded September 2017)[66]. The sequence with the best *e*-value was selected in case of multiple hits for a given protein-coding sequence. Protein localization of all positive hits was determined using the server version of PSORT v3.0 (-a option for archaeal sequences)[67]. Singletons, hits related to central metabolism (that is, archaeal proteasome and precursor proteins) and peptidases included in the ESTHER database (to avoid redundancy to the MEROPS database) were not included for further comparisons.

Phylogenetic analyses of concatenated ribosomal proteins. To reconstruct a species tree including the Heimallarchaeota Chinese Sea genome B2-JM-08 (WGS accession NJBF000000), we used BLASTp to identify orthologues of 56 ribosomal proteins, which we added to the pre-existing alignments taken from ref. [4]. Individual protein data sets were aligned using Mafft-LINSi[68], and ambiguously aligned positions were trimmed using block mapping and gathering with entropy (BMGE) (-m BLOSUM30)[69]. Maximum-likelihood individual phylogenies were reconstructed using IQ-tree v.1.5.5 (ref. [70]) under the LG + C20 + F + G substitution model with 1,000 ultrafast bootstraps and were manually inspected. Trimmed alignments were concatenated into a supermatrix, and an additional data set was generated by removing DPANN homologues to test the effect of taxon sampling on phylogenetic reconstruction. For each of these concatenated data sets, phylogenies were inferred using maximum-likelihood and Bayesian approaches. Maximum-likelihood phylogenies were reconstructed using IQ-tree under the LG + C60 + F + G + PMSF model[71] (Fig. 3b and Supplementary File 1). Statistical support for branches was calculated using 100 bootstraps replicated under the same model. To test the robustness of the

placement of eukaryotes, the data sets were subjected to several treatments. For the 'full data set' and the 'no DPANN data set' (that is, 144 and 104 taxa, respectively), we tested the effect of removing the 25% fastest-evolving sites, as it has been shown that, in large-scale phylogenetic analyses, these sites are often saturated with multiple substitutions and, as a result of model misspecification, can generate an artefactual topology[72]. The corresponding maximum-likelihood trees were inferred as described above (Supplementary Files 2 and 3). Bayesian phylogenies were reconstructed with Phylobayes[73] under the LG + GTR model for the no DPANN data set): 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 four categories to ameliorate the 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] (cut-off of 65–70% for eukaryotes and bacteria and 85% for archaea) and by length (cut-off of 400 amino acids). 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 homologues reviewed in swissprot and predicted cytochrome *c* oxidase, subunit 1 homologues of Asgard archaea (only in Heimdallarchaeota) and some additional archaea not present in UniProt were added to this data set, and data sets were filtered again using cd-hit (cut-off of 85%). The final alignment of 443 positions is based on an hmm-alignment in which sequences were aligned to the 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 the LG + C20 model with ultrafast bootstraps[79].

[NiFe]-hydrogenase, large subunit. Backbone sequences for classifying [NiFe]-hydrogenase were based on HydDB[29]. To reduce sampling size, the backbone data set was filtered using cd-hit[75] with a sequence identity cut-off of 90% before adding put. large subunit [NiFe]-hydrogenases of Asgard archaea. HeimC3_03700 and HeimC2_23100 were removed from the data set to prevent potential long branch attraction (LBA) artefacts, as these sequences represented very long branches and have extremely low similarity to known homologues (<30%). They may represent previously uncharacterized groups that should be investigated in the future. All sequences for group 3 and group 4 large subunit [NiFe]-hydrogenases were aligned using Mafft-LINSi[68] and trimmed with BMGE (using an entropy score of 0.55 and 0.65, respectively)[69]. Maximum-likelihood phylogenetic analyses were performed using IQ-tree[70] with the best-fit model according to Bayesian information criterion (BIC); for example, LG + C60 + R + F (group 4) and LG + C50 + R + F (group 3)). Support values were estimated using the SH-like approximate-likelihood ratio test[80] and ultrafast bootstraps[79], respectively.

RuBisCO. All proteins of Asgard archaea assigned to COG01850 as well as additional homologues of recently sequenced archaeal genomes were extracted and aligned with a representative set of RuBisCO homologues (including family I to family IV) that was based on a backbone data set kindly provided by K. Anantharaman (University of Wisconsin–Madison)[81]. Sequences were aligned using Mafft-LINSi[68] and trimmed with BMGE (entropy score set to 0.55)[69], and the final alignment of 359 aligned positions was subsequently subjected to maximum-likelihood analyses using IQ-tree (using the LG + C60 + R + F model chosen based on the BIC score)[70]. The presence of conserved amino acids was investigated in Jalview based on residues given in ref. [82]. Support values were estimated using the SH-like approximate-likelihood ratio test[80] and ultrafast bootstraps[79], respectively.

Acetyl-CoA synthase/carbon monoxide dehydrogenase. Protein sequences with PF03598 encoding the key subunit of acetyl-CoA synthase/carbon monoxide dehydrogenase, CdhC (also referred to as acetyl-CoA synthase/carbon monoxide dehydrogenase β-subunit or α-subunit) were extracted from UniProtKB (http://www.uniprot.org) (which included Asgard archaeal homologues) and filtered based on a length cut-off of 400 amino acids (except for the partial homologue of *Lokiarchaeaum*) and a sequence identity of 95% using cd-hit[75]. Furthermore, sequences lacking taxonomy information (labelled as uncultured) and poorly aligned sequences that probably represent distant paralogues were removed. Final phylogenies are based on Mafft-LINSi alignments[68] upon trimming with TrimAL 50% (ref. [78]), which were subjected to maximum-likelihood analyses using IQ-tree[70] with LG + C20 + F + R. Bacterial CdhC homologues as well as few archaeal homologues that branch within bacteria encode an N-terminal domain absent in the other canonical archaeal homologues (length of ~250 amino acids). The N-terminal domain was retained upon trimming with TrimAL 50% and yielded longer alignments (712 amino acids) than BMGE. Support values were estimated using the SH-like approximate-likelihood ratio test[80] and ultrafast bootstraps[79], respectively.

β-*Oxidation enzymes.* Unless otherwise stated, the top 1,000 bacterial, archaeal and eukaryotic sequences were retrieved from the Genbank non-redundant database using each Asgard sequence as a query. These data sets 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 hmm-align using PFAM domains as indicated and trimmed using BMGE[69] (-h 0.6 and -m BLOSUM30). Initial trees were generated with Fasttree[77]. The data sets 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 five and ten diverse representatives. This process was repeated iteratively until there were less than 1,000 taxa, at which point alignments were generated using Mafft-LINSi[68] and trees were estimated with IQ-tree[70] (LG + C20 + G + F) with 1,000 ultrafast bootstrap replicates[79].

Acyl-CoA dehydrogenase (COG1960) sequences were retrieved from ref. [83] and used for identifying subcategories of acyl-CoA dehydrogenase enzymes and combined with the sequences retrieved from the non-redundant database (described above). Initial alignments were made as described above using hmm-align with profile alignments for PF02771.14, PF00441.22 and PF02770.17 and concatenated. Acetoacetyl-CoA acyltransferase (COG0183) sequences were retrieved from ref. [84] and the non-redundant database 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) and 3-hydroxy-acyl-CoA dehydrogenase (COG1250; PF00725.20 and PF02737.16). Initial alignments of alternative acyl-CoA dehydrogenase (COG2368) were generated using Mafft-LINSi[68].

Pyruvate-formate lyase superfamily. All sequences assigned to IPR004184 (the pyruvate-formate lyase domain) were downloaded from UniProtKB (http://www.uniprot.org) and filtered by length (only archaeal and bacterial sequences between 700 and 950 amino acids were kept). Bacterial homologues were additionally filtered by identity using cd-hit[75] with a cut-off of 75%. Note that all homologues of Asgard archaea were kept. Sequences of poor quality as well as poorly aligned sequences were removed on inspection of initial Mafft-LINSi alignments. The final set of sequences was realigned 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 the SH-like approximate-likelihood ratio test[80] and ultrafast bootstraps[79], respectively.

Nitrate reductase subunit-α. All sequences with similarity to InterPro domain IPR006468 (nitrate reductase subunit-α) were downloaded from UniProtKB (http://www.uniprot.org). Note that this IPR domain only includes bacterial-type nitrate reductases as the archaeal-type nitrate reductase subunit-α are only distantly related (for example, 24% identity and 56% sequence coverage between the bacterial-type and the archaeal-type (WP_097297416.1 and WP_097297472.1) nitrate reductase subunit-α of *Methanoperedens* sp. BLZ1. Sequences were filtered by length (most sequences were between 1,300 and 1,100 amino acids, which were therefore used as upper and lower 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 1,071 positions was subjected to maximum-likelihood phylogenetic analyses using IQ-tree[70] (LG + C20 + F + R). Support values were estimated using the 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 homologues assigned to the more broadly defined COG01600 (includes both reductive dehalogenases and epoxyqueuosine reductases) were added to a bacterial backbone data set kindly provided by L. Hug (University of Waterloo) and published in ref. [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 (two sequences from Theionarchaea, Ferroglobus and *Methanohalarchaeum thermophilum*). This set of sequences was prealigned using Mafft-LINSi[68] to identify and remove poorly aligned or partial sequences (note that one lokiarchaeal protein as well as the homologue of *M. thermophilum* were removed in this way). Subsequently, sequences were realigned using Mafft-LINSi[68] and trimmed with BMGE[69] using the BLOSUM30 matrix combined with an entropy of 0.6, leaving 160 sites that were subjected to maximum-likelihood

analyses using IQ-tree[70] (using the best-fit model according to BIC determined by IQ-tree: LG + C40 + R + F). Support values were estimated using the 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 Supplementary Fig. 15 do not represent bona fide reductive dehalogenases, which is supported by the absence of the characteristic domain IPR012832. The substrates that can be degraded by some of the characterized reductive dehalogenases are indicated on the tree based on refs. [86,87].

All trees were visualized using FigTree (http://tree.bio.ed.ac.uk/publications/) and modified or annotated with Adobe Illustrator to increase readability.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Code availability

The small custom scripts used for genome annotation and phylogenetic analyses are made available on figshare and can be accessed at the following link: https://figshare.com/s/5f153d1dcacadd3b3ed6.

Data availability

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

Received: 18 September 2018; Accepted: 8 February 2019;
Published online: 1 April 2019

References

1. Hug, L. A. et al. A new view of the tree of life. *Nat. Microbiol.* **1**, 16048 (2016).
2. Spang, A. et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179 (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. *ISME J.* **10**, 1696–1705 (2016).
4. Zaremba-Niedzwiedzka, K. et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (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**, eaaf3883 (2017).
6. Lopez-Garcia, P. & Moreira, D. Open questions on the origin of eukaryotes. *Trends Ecol. Evol.* **30**, 697–708 (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 (2014).
8. Martin, W. F., Garg, S. & Zimorski, V. Endosymbiotic theories for eukaryote origin. *Phil. Trans. R. Soc. B* **370**, 20140330 (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 (2017).
10. Spang, A. et al. Asgard archaea are the closest prokaryotic relatives of eukaryotes. *PLoS Genet.* **14**, e1007080 (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.* **10**, 2380–2393 (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. *Mol. Biol. Evol.* **33**, 1528–1541 (2016).
13. Sousa, F. L., Neukirchen, S., Allen, J. F., Lane, N. & Martin, W. F. Lokiarchaeon is hydrogen dependent. *Nat. Microbiol.* **4**, 16034 (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. *Microbiol. Mol. Biol. Rev.* **81**, e00008-17 (2017).
15. Zachar, I., Szilagyi, A., Szamado, S. & Szathmary, E. Farming the mitochondrial ancestor as a model of endosymbiotic establishment by natural selection. *Proc. Natl Acad. Sci. USA* **115**, E1504–E1510 (2018).
16. Speijer, D. Alternating terminal electron-acceptors at the basis of symbiogenesis: how oxygen ignited eukaryotic evolution. *Bioessays* **39**, 1600174 (2017).
17. Koonin, E. V. Origin of eukaryotes from within archaea, archaeal eukaryome and bursts of gene gain: eukaryogenesis just made easier? *Phil. Trans. R. Soc. B* **370**, 20140333 (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 (2013).
19. Lopez-Garcia, P., Eme, L. & Moreira, D. Symbiosis in eukaryotic evolution. *J. Theor. Biol.* **434**, 20–33 (2017).
20. Ragsdale, S. W. & Pierce, E. Acetogenesis and the Wood–Ljungdahl pathway of CO_2 fixation. *Biochim. Biophys. Acta* **1784**, 1873–1898 (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 (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 (2017).
23. Liu, Y. et al. Comparative genomic inference suggests mixotrophic lifestyle for Thorarchaeota. *ISME J.* **12**, 1021–1031 (2018).
24. Wagner, A. et al. Mechanisms of gene flow in archaea. *Nat. Rev. Microbiol.* **15**, 492–501 (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 (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.* **82**, 4056–4069 (2016).
28. Stams, A. J. & Plugge, C. M. Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nat. Rev. Microbiol.* **7**, 568–577 (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. *ISME J.* **10**, 761–777 (2016).
30. Yu, H. et al. Structure of an ancient respiratory system. *Cell* **173**, 1636–1649. e16 (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 (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 (2018).
33. Adam, P. S., Borrel, G. & Gribaldo, S. Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proc. Natl Acad. Sci. USA* **115**, E1166–E1173 (2018).
34. Kono, T. et al. A RuBisCO-mediated carbon metabolic pathway in methanogenic archaea. *Nat. Commun.* **8**, 14007 (2017).
35. Lang, B. F., Gray, M. W. & Burger, G. Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* **33**, 351–397 (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 (2015).
37. Williams, T. A. et al. Integrative modeling of gene and genome evolution roots the archaeal tree of life. *Proc. Natl Acad. Sci. USA* **114**, 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 (2017).
39. Moreira, D. & Lopez-Garcia, P. Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J. Mol. Evol.* **47**, 517–530 (1998).
40. López-García, P. & Moreira, D. Selective forces for the origin of the eukaryotic nucleus. *Bioessays* **28**, 525–533 (2006).
41. Martin, W. & Muller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41 (1998).
42. Sieber, J. R., McInerney, M. J. & Gunsalus, R. P. Genomic insights into syntrophy: the paradigm for anaerobic metabolic cooperation. *Annu. Rev. Microbiol.* **66**, 429–452 (2012).
43. McGlynn, S. E., Chadwick, G. L., Kempes, C. P. & Orphan, V. J. Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature* **526**, 531–535 (2015).
44. Wegener, G., Krukenberg, V., Riedel, D., Tegetmeyer, H. E. & Boetius, A. Intercellular wiring enables electron transfer between methanotrophic archaea and bacteria. *Nature* **526**, 587–590 (2015).
45. Knittel, K. & Boetius, A. Anaerobic oxidation of methane: progress with an unknown process. *Annu. Rev. Microbiol.* **63**, 311–334 (2009).
46. Seitz, K. W. et al. New Asgard archaea capable of anaerobic hydrocarbon cycling. Preprint at https://www.biorxiv.org/content/10.1101/527697v2 (2019).
47. Laso-Perez, R. et al. Thermophilic archaea activate butane via alkyl-coenzyme M formation. *Nature* **539**, 396–401 (2016).
48. Martijn, J., Vosseberg, J., Guy, L., Offre, P. & Ettema, T. J. G. Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature* **557**, 101–105 (2018).
49. Leger, M. M., Eme, L., Stairs, C. W. & Roger, A. J. Demystifying eukaryote lateral gene transfer (response to Martin 2017 DOI: 10.1002/bies.201700115). *Bioessays* **40**, e1700242 (2018).
50. Stairs, C. W. et al. Microbial eukaryotes have adapted to hypoxia by horizontal acquisitions of a gene involved in rhodoquinone biosynthesis. *eLife* **7**, e34292 (2018).

51. Norlund, K. L. et al. Microbial architecture of environmental sulfur processes: a novel syntrophic sulfur-metabolizing consortia. *Environ. Sci. Technol.* **43**, 8781–8786 (2009).
52. Bose, A., Gardel, E. J., Vidoudez, C., Parra, E. A. & Girguis, P. R. Electron uptake by iron-oxidizing phototrophic bacteria. *Nat. Commun.* **5**, 3391 (2014).
53. Eme, L., Spang, A., Lombard, J., Stairs, C. W. & Ettema, T. J. G. Archaea and the origin of eukaryotes. *Nat. Rev. Microbiol.* **15**, 711–723 (2017).
54. Ettema, T. J. Evolution: mitochondria in the second act. *Nature* **531**, 39–40 (2016).
55. Caforio, A. et al. Converting *Escherichia coli* into an archaebacterium with a hybrid heterochiral membrane. *Proc. Natl Acad. Sci. USA* **115**, 3704–3709 (2018).
56. Martin, W. et al. Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* **393**, 162–165 (1998).
57. Pittis, A. A. & Gabaldon, T. Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature* **531**, 101–104 (2016).
58. Roger, A. J., Munoz-Gomez, S. A. & Kamikawa, R. The origin and diversification of mitochondria. *Curr. Biol.* **27**, R1177–R1192 (2017).
59. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
60. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Archaeal Clusters of Orthologous Genes (arCOGs): an update and application for analysis of shared features between Thermococcales, Methanococcales, and Methanobacteriales. *Life (Basel)* **5**, 818–840 (2015).
61. Saier, M. H. Jr, Tran, C. V. & Barabote, R. D. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res.* **34**, D181–D186 (2006).
62. Sondergaard, D., Pedersen, C. N. & Greening, C. HydDB: a web tool for hydrogenase classification and analysis. *Sci. Rep.* **6**, 34212 (2016).
63. 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 (2017).
64. Yin, Y. et al. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **40**, W445–W451 (2012).
65. Rawlings, N. D., Barrett, A. J. & Finn, R. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **44**, D343–D350 (2016).
66. Lenfant, N. et al. ESTHER, the database of the α/β-hydrolase fold superfamily of proteins: tools to explore diversity of functions. *Nucleic Acids Res.* **41**, D423–D429 (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 (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 Res.* **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 (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 (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 (2018).
72. Kamikawa, R. et al. Parallel re-modeling of EF-1α function: divergent EF-1α genes co-occur with EFL genes in diverse distantly related eukaryotes. *BMC Evol. Biol.* **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 (2013).
74. Susko, E. & Roger, A. J. On reduced amino acid alphabets for phylogenetic inference. *Mol. Biol. Evol.* **24**, 2139–2150 (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 (2012).
76. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
77. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
78. Capella-Gutierrez, S. et al. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
79. Minh, B. Q., Nguyen, M. A. & von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).
80. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
81. Anantharaman, K. et al. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* **7**, 13219 (2016).
82. Wrighton, K. C. et al. RuBisCO of a nucleoside pathway known from Archaea is found in diverse uncultivated phyla in bacteria. *ISME J.* **10**, 2702–2714 (2016).
83. Swigonova, Z., Mohsen, A. W. & Vockley, J. Acyl-CoA dehydrogenases: dynamic history of protein family evolution. *J. Mol. Evol.* **69**, 176–193 (2009).
84. Dibrova, D. V., Galperin, M. Y. & Mulkidjanian, A. Y. Phylogenomic reconstruction of archaeal fatty acid metabolism. *Environ. Microbiol.* **16**, 907–918 (2014).
85. Hug, L. A. et al. Overview of organohalide-respiring bacteria and a proposal for a classification system for reductive dehalogenases. *Phil. Trans. R. Soc. B* **368**, 20120322 (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 (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 (2007).
89. Rochette, N. C., Brochier-Armanet, C. & Gouy, M. Phylogenomic test of the hypotheses for the evolutionary origin of eukaryotes. *Mol. Biol. Evol.* **31**, 832–845 (2014).

Acknowledgements

This work was supported by grants from the European Research Council (ERC starting grant 310039-PUZZLE_CELL to T.J.G.E.), the Swedish Foundation for Strategic Research (SSF-FFL5 to T.J.G.E.), the Swedish Research Council (VR grant 2015-04959 to T.J.G.E. and VR starting grant 2016-03559 to A.S.), the NWO-I Foundation of the Netherlands Organisation for Scientific Research (WISE fellowship to A.S.), the European Commission (Marie Curie IEF European grants 625521 to A.S. and 704263 to L.E.), the Wenner-Gren Foundations in Stockholm (2016-0072 to J.L.), the European Molecular Biology Organization (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 thank K. Zaremba-Niedzwiedzka and J. Saw for reconstruction of some of these genomes and helpful discussions. We also acknowledge S. 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 P. Offre for discussions on metabolic inferences.

Author contributions

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

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-019-0406-9.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.S. or T.J.G.E.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Corresponding author(s): Anja Spang (anja.spang@nioz.nl)
Thijs Ettema (thijs.ettema@icm.uu.se)

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed
☒	☐ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☒	☐ Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	The genomes of the 10 Asgard representatives were downloaded from NCBI.
Data analysis	Databases used to analyze data and or against which data was compared: - nr (ftp://ftp.ncbi.nlm.nih.gov/) - UniProtKB (https://www.uniprot.org/) - HydDB (https://services.birc.au.dk/hyddb/) - Carbohydrate-Active enZYmes (CAZYmes) database (http://www.cazy.org/) - MEROPS peptidase database (https://www.ebi.ac.uk/merops/) - esterases were predicted using HMM-profiles from the ESTHER database (https://www.re3data.org/repository/r3d100010542) - transporter database (http://www.tcdb.org/) - Metacyc Metabolic Pathway Database (https://metacyc.org/) - KEGG (https://www.genome.jp/kegg/) - COGs and arCOGs (ftp://ftp.ncbi.nih.gov/pub/wolf/COGs/arCOG) Data was analyzed using the following published softwares: - dbCAN webtool75 - blastdbcmd version 2.6.0+ - Interproscan-5.22-61.0 - HMMer vs. 3.1b2

- PSORT v3.0
- Mafft-LINSi v7.305b
- BMGE-1.12
- IQ-TREE v. 1.5.5
- Phylobayes MPI 1.7
- Psort v3.0
- R v3.3.0 (including plyr v 1.8.4 and ggplot2 v3.0.0)
- DIAMOND v0.9.9.110
- CD-HIT version 4.6
- FastTree version 2.1.9
- TrimAL v1.4
- Jalview 2.10.2b2
- Seaview n.d.
- FigTree v1.4.2

Data Visualization:
- Adobe Illustrator CC 2015 (19.2.0)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data and code vailability
The genomes of the herein analysed Asgard archaea have been made publicly available on NCBI previously. Detailed annotations of the metabolic repertoire are provided in Suppl. Tables 1-3 accompanying this manuscript. Raw data files and custom scripts are made available via figshare under the following link: https://figshare.com/s/5f153d1dcacadd3b3ed6.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We performed comparative genomics and phylogenomic analyses of published (publicly available) genomes of the Asgard archaea to infer their metabolic potential. Based on this analysis we update hypotheses on the origin of the eukaryotic cell.
Research sample	We analysed the genomes of Asgard archaea available at NCBI in January 2017, i.e. 2 published Lokiarchaeota (strains GC14_75 and CR4), 4 published Thorarchaeota (strains SMTZ-45, SMTZ1-83, SMTZ1-45 and AB25), 3 published Heimdallarchaeota (strains LC3, AB125 and LC2) and one published Odinarchaeote (strain LCB_4). Files of the metagenome assembled genomes of these Asgard archaea were retrieved from the ftp server of NCBI (ftp://ftp.ncbi.nih.gov/genomes/genbank/archaea/).
Sampling strategy	We did not obtain any new samples, as these genomes have already been reconstructed in previous and published work.
Data collection	The genome data described in the present manuscript was downloaded from the ftp server of NCBI and originally deposited along with following references: Spang, A. et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature 521, 173-+, doi:10.1038/nature14447 (2015). 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). Zaremba-Niedzwiedzka, K. et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature 541, 353-358, doi:10.1038/nature21031 (2017).
Timing and spatial scale	we herein analyzed genomes of Asgard archaea deposited at NCBI in January 2017. Timing and spacial scale were not studied.

Data exclusions	no data was excluded from our analysis
Reproducibility	we have in detail described our methodological approaches for genome annotations as well as phylogenetic analyses to ensure reproducibility. We calculated branch support values for the different clusters obtained in our phylogenies that provide an indication as to how stable a given monophyletic group is. Considering this, our phylogenetic analyses are reproducible (when the same phylogenetic strategy is applied such as the same model of evolution etc). Genome annotations are based on inferences from various different databases. While this provides some level of confidence, annotations may have to be updated when new information is generated and deposited in databases.
Randomization	Since central tendencies and deviations between different treatments and controls have not been studied here, randomization and blinding are not applicable
Blinding	Since central tendencies and deviations between different treatments and controls have not been studied here, randomization and blinding are not applicable

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging