Review

The emerging view on the origin and early evolution of eukaryotic cells

https://doi.org/10.1038/s41586-024-07677-6

Received: 2 November 2023

Accepted: 5 June 2024

Published online: 11 September 2024

Check for updates

Julian Vosseberg[1], Jolien J. E. van Hooff[1], Stephan Köstlbacher[1], Kassiani Panagiotou[1], Daniel Tamarit[2] & Thijs J. G. Ettema[1]✉

The origin of the eukaryotic cell, with its compartmentalized nature and generally large size compared with bacterial and archaeal cells, represents a cornerstone event in the evolution of complex life on Earth. In a process referred to as eukaryogenesis, the eukaryotic cell is believed to have evolved between approximately 1.8 and 2.7 billion years ago from its archaeal ancestors, with a symbiosis with a bacterial (proto-mitochondrial) partner being a key event. In the tree of life, the branch separating the first from the last common ancestor of all eukaryotes is long and lacks evolutionary intermediates. As a result, the timing and driving forces of the emergence of complex eukaryotic features remain poorly understood. During the past decade, environmental and comparative genomic studies have revealed vital details about the identity and nature of the host cell and the proto-mitochondrial endosymbiont, enabling a critical reappraisal of hypotheses underlying the symbiotic origin of the eukaryotic cell. Here we outline our current understanding of the key players and events underlying the emergence of cellular complexity during the prokaryote-to-eukaryote transition and discuss potential avenues of future research that might provide new insights into the enigmatic origin of the eukaryotic cell.

Eukaryotic cells display multiple characteristics that distinguish them from cells of prokaryotes, including a nucleus, the feature that gives eukaryotes their name, other membrane-bound organelles, an elaborate trafficking system and a dynamic cytoskeleton. This, in addition to their comparatively large cell size and large gene content, has resulted in the view that eukaryotes represent more complex life forms compared to prokaryotes. The seminal work of Carl Woese and colleagues during the last quarter of the previous century unveiled two distinct prokaryotic domains: Bacteria and Archaea[1–3]. Biochemical and genetic studies of some of the few archaeal lineages that could be cultured revealed that archaea and eukaryotes share components of information-processing machineries, such as those responsible for transcription[4,5] and genome replication[6]. Eukaryotes are now widely believed to have originated from both the archaeal and bacterial domain[7,8]. This has resulted in the view that only the prokaryotic domains, Archaea and Bacteria, should be regarded as primary domains of life[9]. Eukaryotes, rather, evolved from an archaea-related host cell and a bacteria-related endosymbiont, whose descendants are still present in modern eukaryotic cells in the form of mitochondria.

The evolutionary transition from prokaryote to eukaryote, the process referred to as eukaryogenesis (Box 1), is inferred to have taken place mostly in the Paleoproterozoic era[10]. Recent estimates from molecular dating analyses suggest that this transition took hundreds of millions of years between 2.7 and 1.8 billion years ago[10,11], although such estimates are under debate[12]. The study of this transition is complicated by the extinction of intermediate lineages, which has contributed to the long branch of the eukaryotic clade, its stem, in the tree of life[10]. The fossil and biomarker record of proposed stem eukaryotes is limited, and its interpretation is debated[13–15]. This, combined with the fact that the cellular complexity of the last eukaryotic common ancestor (LECA) has been inferred to resemble that of modern eukaryotes[16] and that the eukaryotic cell evolved only once, makes eukaryogenesis an evolutionary conundrum. Because the eukaryotic features that would define a stem eukaryote as a eukaryote-grade organism[17] remain contentious, here we use the taxonomic definition of eukaryotes[17] and define eukaryogenesis as the period between the first eukaryotic common ancestor (FECA) and LECA, spanning the entire eukaryotic stem lineage (Fig. 1a). Different definitions of FECA have been used in the literature. Here we use the original definition[18] of FECA as the first stem eukaryote after the separation between the eukaryotic lineage and its closest archaeal sister group. FECA was therefore not unlike its archaeal relatives and—merely owing to historical contingency—only had eukaryotes as extant descendants. Similar to the archaea-related FECA, the mitochondrial stem lineage starts with the first mitochondrial common ancestor (FMCA). It is important to realize that FECA and FMCA are entities based on inferences of genomic data[19] and that their nature and phylogenetic position in the tree of life can change upon the discovery of even closer sister groups of eukaryotes and mitochondria[20]. The same would apply to LECA in case new deep-branching eukaryotic lineages were to be identified. The inferred nature of LECA illustrates the fundamental gap in cellular complexity between prokaryotes and eukaryotes that was bridged during eukaryogenesis[21]. Here we review recent progress in the field of eukaryogenesis and highlight future research avenues that could help finding new pieces of this enigmatic evolutionary puzzle.

[1]Laboratory of Microbiology, Wageningen University & Research, Wageningen, the Netherlands. [2]Theoretical Biology and Bioinformatics, Department of Biology, Faculty of Science, Utrecht University, Utrecht, the Netherlands. ✉e-mail: thijs.ettema@wur.nl

Box 1

Glossary

Alphaproteobacteria. A bacterial class comprising numerous orders, including Rickettsiales, Pelagibacterales, Rhizobiales and Rhodobacterales.

Archezoa. A proposed paraphyletic group of alleged amitochondriate eukaryotes that would have originated from the first splits in the eukaryotic crown group. Later studies have confidently shown that these eukaryotes do not represent an ancestral amitochondriate state as they have evolved from mitochondria-bearing groups and in fact contain mitochondrion-related organelles themselves.

Asgard archaea. An archaeal phylum (formal name Asgardarchaeota[162]) comprising Lokiarchaeia, Thorarchaeia, Odinarchaeia, Heimdallarchaeia and several additional classes that have been proposed recently.

Endosymbiosis. Symbiosis in which one of the partners, the endosymbiont, lives inside the cell of the other symbiont, the host.

Eukaryogenesis. The transition from prokaryote-grade to eukaryote-grade organisms, during which the eukaryotic features evolved; in essence the period of time between FECA and LECA.

Eukaryotic signature proteins (ESPs). Eukaryotic proteins involved in conserved eukaryotic functions that, at least previously, were considered to be specific to eukaryotes—that is, they have no prokaryotic homologues.

First eukaryotic common ancestor (FECA). The oldest common ancestor of eukaryotes that only has eukaryotes as its extant descendants.

First mitochondrial common ancestor (FMCA). The oldest common ancestor of mitochondria that only has mitochondria as its extant descendants; also referred to as Alphaproteobacteria-derived FECA[17] or pre-mitochondrial alphaproteobacterium[66].

Horizontal gene transfer. Gene exchange between two organisms that is not due to vertical inheritance.

Last eukaryotic common ancestor (LECA). The most recent common ancestor of all present-day eukaryotes.

Paleoproterozoic era. The oldest era of the Proterozoic eon, spanning from 2.5 to 1.6 billion years ago, during which eukaryotes are likely to have originated.

Phylogenomics. Analyses to reconstruct the evolution of genomes, typically by assessing the evolutionary histories of multiple genes at the same time.

Proto-mitochondria. Transitional forms of mitochondria living in (endo)symbiosis with stem eukaryotes.

Stem eukaryotes. Extinct species that are more closely related to the eukaryotic crown group (that is, LECA and its descendants) than present-day Asgard archaea; the oldest stem eukaryote is FECA.

Supermatrix. A large multiple sequence alignment obtained from the concatenation of single-marker alignments of protein or DNA sequences, used to establish evolutionary relationships between taxa.

Syntrophy. Interdependent, mutualistic metabolic cross-feeding of at least two partners.

Two-domains (2D) tree of life. A tree topology in which eukaryotes are nested within archaea, originally proposed in the eocyte hypothesis[27]. In the alternative three-domains (3D) tree of life, eukaryotes are a sister group of archaea.

Eukaryotes in the changing tree of life

Owing to the recent shifts of the eukaryotic branch in the tree of life, our perception of the starting point of eukaryogenesis—FECA—has changed considerably. Early phylogenetic analyses of small subunit ribosomal RNA gene sequences[3,22–24] and various individual proteins[25,26] generated conflicting scenarios for the origin of eukaryotes. Throughout the past three decades, two main hypotheses shaped views on eukaryogenesis. In the initially dominant three-domain (3D) tree of life, eukaryotes were placed as the sister group of all archaea[3]. The main competing scenario was the 'eocyte tree'—also known as the two-domain (2D) tree of life—in which eukaryotes were most closely related to an archaeal group referred to as eocytes[27,28]. Because the eocyte tree pre-dated most of the currently known archaeal diversity, these eocytes comprised only members of the Crenarchaea.

Unlike studies based on phylogenetic analyses of single genes, modern efforts to resolve the position of eukaryotes in the tree of life utilize a phylogenomic approach, and aim to simultaneously assess the evolutionary histories of multiple genes. This is most frequently done by concatenating gene alignments into a single supermatrix, under the assumption that their inferred shared evolutionary histories signify vertical inheritance. In practice, this supermatrix comprises alignments of proteins that are present across all organisms, typically proteins that are involved in essential processes such as protein synthesis (for example, ribosomal proteins). Early phylogenomic studies were limited by the low diversity of available genomic data[29–32]. This was particularly true for the sequence data from archaeal taxa, which were obtained only through painstaking cultivation efforts and then-costly sequencing technologies. Additionally, phylogenetic methodologies were only starting to accommodate both the copious amounts of input data provided by concatenated alignments and the complex phylogenetic signals. As a result, such analyses often still suffered from potential phylogenetic artefacts (Box 2). Yet, the implementation of more realistic models of sequence evolution started a trend of increasing support for 2D trees[31,32]. As the rapid development of metagenomic sequencing steadily revealed genome sequences from uncharacterized microbial groups, more studies provided support for the 2D topology, including the monophyly of eukaryotes with the TACK superphylum, which initially comprised Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota[9,33–38].

The second half of the 2010s marked the discovery of a novel archaeal group related to TACK archaea—the Asgard archaea[39–41]—which have had a substantial impact on the eukaryogenesis field. Asgard archaea were found to represent the closest extant relatives of eukaryotes, providing further support for the 2D tree of life[7,8,39,41–50] (Fig. 1b). However, the exact position of eukaryotes with respect to Asgard archaea remained unclear. Although eukaryotes were placed as a sister lineage to Asgard archaea in several studies[45,47,50], studies utilizing the most sophisticated phylogenetic analyses obtained support for an affiliation between eukaryotes and a specific group within the Asgard archaea: the Heimdallarchaeia[7,8]. These latter analyses examined multiple gene sets and methods that were explicitly aimed to disentangle the deep, vertical phylogenetic signal from other sequence patterns (Box 2). The most exhaustive study to date in terms of dataset size and methodology indicated that the heimdallarchaeial order Hodarchaeales represents the closest archaeal relatives of eukaryotes[8]. Future efforts to accurately position eukaryotes with respect to Asgard archaea will corroborate these findings or find support for alternative phylogenetic scenarios, and will be key for inferring the genomic properties of FECA, and thus for increasing our understanding of eukaryogenesis.

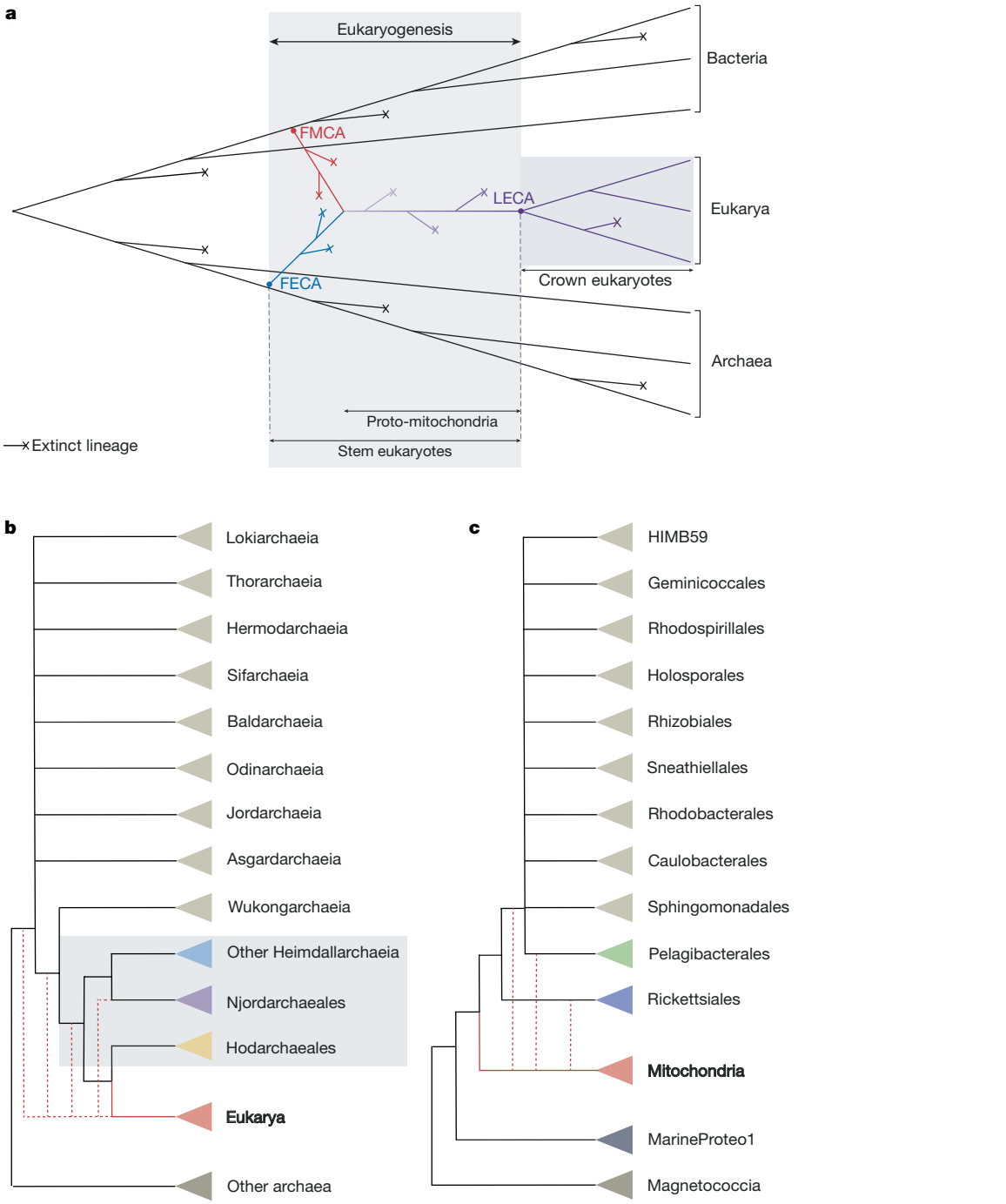

Fig. 1 | The phylogenetic origins of eukaryotes. a, The tree of life, showing the chimeric origin of the eukaryotic lineage and the period of eukaryogenesis. In this illustration, the eukaryotic lineage diverges slightly earlier from Asgard archaea compared with Alphaproteobacteria, as suggested recently[11]. FECA, first eukaryotic common ancestor; FMCA, first mitochondrial common ancestor; LECA, last eukaryotic common ancestor. **b**, Phylogenetic placement of eukaryotes with respect to Asgard archaea. The currently sampled diversity of Asgard classes is shown, in addition to relevant orders in the Heimdallarchaeia (highlighted in blue). The position of eukaryotes that is supported by the most sophisticated analyses to date[7,8], the Hodarchaeales–sister topology, is shown with a solid red line. Other recovered positions, as sister group of Njordarchaeales[8,49], all Heimdallarchaeia[7,8,41–43,46], the Heimdallarchaeia-Wukongarchaeia clade[45] or all Asgard archaea[47,50], are displayed with dashed red lines. **c**, Phylogenetic placement of eukaryotes (mitochondria) with respect to Alphaproteobacteria and Magnetococcia. A selection of alphaproteobacterial orders is shown. The position of mitochondria that is recovered by the most sophisticated analyses performed so far, the Alphaproteobacteria–sister topology[77,79,80], is shown with a solid red line. Other recovered positions, as sister group of Rickettsiales[68,69,72,73,75,76,78], Pelagibacterales[70,71] or a clade comprising a wide range of orders excluding Rickettsiales[73,74], are displayed with dashed red lines.

The nature of FECA

The nature of FECA is illuminated as we study genomic and cellular features of present-day Asgard archaea and interpret these given the resolved evolutionary relationships. Several recent studies have significantly expanded the genomic diversity of Asgard archaea, revealing a phylum comprising at least ten candidate classes[8] (Fig. 1b). Asgard genomes display a large variability in genome size and gene numbers, but in particular Hodarchaeales and Lokiarchaeia genomes are considerably larger compared with genomes of other archaeal groups[8].

Box 2

Challenges in placing eukaryotes in the tree of life

Resolving deep branches in the tree of life is notoriously challenging owing to the low and sometimes distorted phylogenetic signal for such ancient events. A principle of phylogenomics is to utilize multiple gene markers to increase the number of available informative sites. However, a variety of factors obscure phylogenetic signals that are indicative of vertical inheritance, potentially causing competing non-vertical signals and incongruent trees if not properly assessed[163–166].

The large evolutionary divergence of the groups involved—possibly dating back to more than two billion years ago[10,11]—represents a major challenge in placing eukaryotes in the tree of life. Over such extended time periods, multiple substitutions can occur per position and the resulting mutational saturation is difficult to properly accommodate in models of sequence evolution. This can lead to the erroneous grouping of long branches, a well-known artefact in phylogenetic analyses. Several lineages relevant for eukaryogenesis have such long branches, including the eukaryotic stem lineage and several Asgard archaeal (for example, Njordarchaeales) and alphaproteobacterial clades (for example, Rickettsiales and Pelagibacterales). Ideally, this problem can be mitigated by obtaining sequence data from taxa that break these long branches[8,39,41,45,77,80]. Alternatively, it is possible to reduce branch lengths by removing fast-evolving sites, sequences or even entire taxonomic groups from the analysis. Another strategy, if done with care[167], is to recode amino acids into a small number of saturation-prone categories, consequently only analysing the less frequent substitutions between these categories[168].

Various factors can cause conflicting signals in phylogenetic analyses. One significant source of error is the failure to correctly infer orthologous sequences—that is, homologous sequences that trace back to a speciation event in their common ancestor. In these cases, conflicting, non-vertical signals could originate from horizontally transferred genes, distant gene duplications or contaminating genes in genome assemblies. Using robust orthology inference pipelines and evaluating phylogenetic trees from individual gene markers to identify these conflicts is therefore essential. Another—often overlooked—source of potential conflicting signals represents convergent evolution resulting from compositional bias towards certain amino acids in protein sequences. Examples include substitutions associated with AT richness and adaptations to high growth temperatures, affecting mitochondria and certain alphaproteobacterial groups[73,77,80,169] (such as Rickettsiales and Pelagibacterales) and Asgard archaeal taxa[8] (such as Njordarchaeales), respectively. Convergent evolution of sites may erroneously be interpreted as close relatedness and, as a result, taxa with similar compositional bias can be artefactually grouped together. These compositional artefacts can be ameliorated by removing strongly biased sites or taxa and by utilizing models of sequence evolution accommodating compositional heterogeneity across sites or branches[80].

Evaluating the effect of the factors described above is far from straightforward and often requires comparing methodological approaches, taxon sampling strategies and marker gene sets. This includes carrying out multiple rounds of inference with different datasets and parameters, and testing the adequacy and incongruence of the obtained results. Although more realistic, complex evolutionary models can improve topology estimation[7,8,80], these models tend to be too computationally demanding for large datasets. Strategies that involve reducing dataset sizes[7,8] or fixing tree topologies[80] enable exploration under these more powerful phylogenetic approaches.

Although initially suspected as potential contaminants from eukaryotic genomes[51], it is now well-established that Asgard archaeal genomes encode a diverse subset of proteins previously deemed unique to eukaryotes. Many of these eukaryotic signature proteins (ESPs) represent homologues of eukaryotic proteins that are involved in processes underlying eukaryotic cellular complexity, such as dynamic cytoskeleton formation, membrane remodelling and vesicular trafficking[8,39,41,45]. Several families of ESPs, such as the small GTPase, endosomal sorting complex required for transport-III (ESCRT-III) and actin families, have expanded in certain Asgard clades[8,39,52,53], illustrating that in parallel to eukaryotic genome evolution, gene duplication events have contributed to shaping Asgard archaeal genome content. Furthermore, many ESPs have been found to represent multidomain proteins with diverse domain architectures, suggesting an important role of domain shuffling during Asgard archaeal genome evolution[45]. Recent analyses of expanded and more diverse sets of Asgard archaeal genomes have revealed substantial numbers of previously undetected ESPs[8,45]. Besides their pivotal roles in Asgard archaeal cell biology (discussed below), the inferred presence of ESPs in FECA indicates that FECA was genetically primed for the emergence of cellular complexity during eukaryogenesis[54].

Recently developed gene-tree-aware reconciliation methods enable the inference of evolutionary genome dynamics and gene content of ancestors from present-day genomes[55,56]. Using this approach, a recent study inferred larger numbers of genes for most Asgard archaeal ancestors compared with other archaeal ancestors, and increased gene duplication rates in Heimdallarchaeia and Lokiarchaeia[8]. On the basis of the inferred origin of the eukaryotic lineage between the last common ancestors of Heimdallarchaeia and the Hodarchaeales, inferences of FECA were drawn from the reconstructed gene content of these ancestors. Notably, multiple copies of several cytoskeletal and membrane-trafficking proteins were inferred in FECA in this way[8]. Various studies have revealed patchy distributions of ESPs across Asgard archaeal genomes, in particular of homologues of eukaryotic proteins involved in vesicular trafficking and endosomal sorting[8,41], highlighted by homologues of adaptor proteins and subunits of the retromer, Golgi-associated retrograde protein (GARP), homotypic fusion and protein sorting (HOPS) and class C core vacuole/endosome tethering (CORVET) complexes, respectively. These patchily distributed ESPs were likely to be already present in the Asgard archaeal lineage leading to FECA and lost or transferred multiple times among Asgard archaea, although these ESPs could also have been acquired by stem eukaryotes via horizontal gene transfer[48].

Although the metagenomic exploration of Asgard archaea and subsequent analysis of their gene content provided new views on the nature of FECA and the process of eukaryogenesis, the functional roles of ESPs in Asgard archaeal cell biology are largely unknown. As these ESPs have evolved independently of eukaryotes in diverse Asgard archaeal groups for billions of years, in silico analyses alone are insufficient to elucidate the extent of functional equivalence to their eukaryotic counterparts. Asgard archaea are not easily cultured, let alone genetically tractable, and therefore studies on Asgard archaeal proteins and complexes depend on heterologous expression in suitable

host organisms followed by in vitro characterization. Multiple ESPs have been characterized using this approach in recent years.

Most Asgard archaea encode actin homologues[39,52] and actin-binding proteins such as profilin and gelsolin[39,41], which together led to speculations of a dynamic actin cytoskeleton in contemporary Asgard archaea. The molecular characterization of Asgard archaeal profilin and gelsolin have indeed demonstrated that they not only adopt a similar fold as their eukaryotic counterparts but can also perform key roles in modulating actin polymerization[57–59]. Aside from actin, another structural component of the cytoskeleton found in Asgard archaeal genomes—although only identified in Odinarchaeia so far—is tubulin[41]. However, in vitro these Asgard tubulin homologues form curved protofilaments that spiral around other protofilaments to form a tubule, more resembling the dynamics of the prokaryotic homologue FtsZ rather than eukaryotic tubulin[60].

The ESCRT system is one of the major eukaryotic systems involved in membrane remodelling and endosomal sorting. Eukaryotes transport ubiquitylated cargo with the concordant action of the highly conserved ESCRT-I, ESCRT-II, ESCRT-III subcomplexes, and the ATPase Vps4[61]. Homologues of ESCRT-III and Vps4, which execute the last step of membrane fission, can be found in several archaea. However, only Asgard archaea have been shown to possess additional components related to ESCRT-I and ESCRT-II subcomplexes[8,39,41,62]. Recent experiments suggest that Asgard archaeal ESCRT-I and ESCRT-II homologues are indeed involved in ubiquitin-directed recruitment of ESCRT-III, as in eukaryotes[62]. Soluble N-ethylmaleimide-sensitive factor attachment protein receptor (SNARE) proteins are key proteins in the critical last step of vesicular transport—the fusion of the vesicle with its target membrane. Homologues of SNARE proteins have been identified in Heimdallarchaeia and were shown to form stable interactions with eukaryotic SNARE proteins in vitro[63].

On the basis of in silico analysis of gene content complemented with in vitro studies, a global picture begins to emerge of Asgard archaeal cell biology that involves certain complex eukaryotic-like features, such as a dynamic cytoskeleton and vesicular trafficking functions. The cultivation of two closely related Lokiarchaeia representatives, '*Candidatus* Prometheoarchaeum syntrophicum' and '*Candidatus* Lokiarchaeum ossiferum', provided an important first glimpse into the physiology and cell biology of extant Asgard archaea[44,50]. Reminiscent of proposed models for eukaryogenesis, these lokiarchaeial lineages live in anaerobic, syntrophic communities with bacteria and other archaea that utilize their fermentation products. High-resolution microscopy analyses revealed that their cells form small cocci with complex, sometimes branched, membrane protrusions. While lacking robust evidence for vesicular trafficking, ultrastructure analysis revealed an elaborate actin cytoskeleton within the cell body and protrusions[50]. Together, these findings sparked speculation of an involvement of such protrusions in the potential interaction with, and engulfment of, syntrophic bacteria during eukaryogenesis[44,50]. However, such speculations should be made with caution. First, Lokiarchaeia are only distantly related to the most likely sister clade of eukaryotes. Second, their genomes encode a distinct subset of ESPs, and they have diverged from the last common ancestor of all Asgard archaea for probably more than two billion years[8,10,11]. This makes it unlikely that all their cell biological features are representative of FECA. Given their phylogenetic and genomic diversity (Fig. 1b), it is anticipated that successful cultivation efforts of additional Asgard archaeal lineages will probably reveal diverse physiological and cell biological features. For example, recent in situ imaging efforts of lokiarchaeial and heimdallarchaeial cells extracted from marine sediments revealed spatial separation of genomic DNA and ribosomes, a feature thus far not observed in cultured Lokiarchaeia[64]. Ultimately, an amalgamation of computational analyses and experimental work, including ultrastructure imaging of diverse Asgard archaea isolates and experimental characterization of ESPs, will be needed to infer the nature of FECA and to refine current eukaryogenesis models.

The endosymbiotic origin of mitochondria

Mitochondria are membrane-bound organelles that as well as having a central role in chemical energy generation, are involved in various processes in eukaryotic cells, including fatty acid and sterol biosynthesis, Fe/S cluster biosynthesis, apoptosis and calcium homeostasis. Given their central roles in eukaryotic cell biology and the fact that all known eukaryotes have, or once had, mitochondria, the origin of mitochondria (referred to as proto-mitochondria in stem eukaryotes) has been a crucial event during eukaryogenesis. It is widely established that—having evolved from once free-living bacteria—mitochondria have an endosymbiotic origin[65,66]. Yet, the identity of the proto-mitochondrial endosymbiont and the timing and nature of the symbiosis, including the mechanism of acquisition, have been subject to ongoing debate.

To infer details about the nature of proto-mitochondrial symbiosis, it is required to trace the identity and physiology of the FMCA. In contrast to the relatively recent discovery of Asgard archaea[39,41], the evolutionary relationship of mitochondria with Alphaproteobacteria has been recognized for several decades[67]. Nonetheless, conflicting results have been obtained throughout the years regarding the phylogenetic position of mitochondria in the alphaproteobacterial tree[68–82], although most past studies reported a common ancestry with Rickettsiales (Fig. 1c). On the basis of this 'Rickettsiales–sister' topology, FMCA was even inferred to have been an energy parasite[75]. As for the identification of the archaea-related host lineage, obtaining a stable phylogenetic placement of mitochondria has proved difficult owing to high rates of sequence evolution and compositional bias of mitochondria and several alphaproteobacterial clades (Box 2). Recent studies that aimed to adequately address compositional heterogeneity and high sequence divergence have recovered a deep phylogenetic position of mitochondria outside Alphaproteobacteria, the 'Alphaproteobacteria–sister' topology[77,80,81]. Although still challenged[78,82], the Alphaproteobacteria–sister relationship may be further corroborated upon the discovery of alphaproteobacterial relatives from new metagenomic surveys. These findings could help to draw a clearer picture of the nature of the FMCA, which is currently hindered by the deep position of mitochondria as sister group of Alphaproteobacteria, a phylogenetically diverse bacterial class displaying a wide variety of lifestyles[83].

As a result of the changing inferences about the proposed identity and nature of both the host and endosymbiont, various hypotheses have been proposed over the past decades to explain the initial proto-mitochondrial endosymbiosis (reviewed in, for example, ref. 84). Here we revisit syntrophic hypotheses that have been proposed or revised upon the discovery of Asgard archaea. In the revised hydrogen hypothesis[85,86] and reverse flow hypothesis[42] the symbiosis entailed the exchange of hydrogen and/or monocarboxylic organic acids between the archaea-related host and the proto-mitochondrial endosymbiont (Fig. 2a). Where in the hydrogen hypothesis the syntrophic interaction involved interspecies hydrogen transfer from the proto-mitochondrial to the archaea-derived symbiont, the direction of electron and hydrogen transfer was opposite in the reverse flow model. In the revised syntrophy hypothesis[87,88] and entangle–engulf–endogenize (E3) hypothesis[44], a third symbiotic partner was involved, a sulfate-reducing (deltaproteo)bacterium. Whereas in the syntrophy hypothesis this bacterium was the host that first took up the archaea-derived symbiont and later the proto-mitochondrial ancestor, the role of the sulfate-reducing bacterium in the E3 model was limited to hydrogen scavenging in the initial tripartite symbiosis (Fig. 2a).

On the basis of a recent study that placed eukaryotes as a sister group of the Hodarchaeales, FECA was inferred to have been a mesophilic heterotroph[8]. Given the presence of several oxygen-dependent enzymes in Heimdallarchaeia, including terminal oxidases, FECA may have been able to occupy both anoxic and (micro)oxic niches and perform aerobic respiration[8,42,89]. At a minimum, the notion that all Asgard archaea should be regarded as strict anaerobes is incorrect.

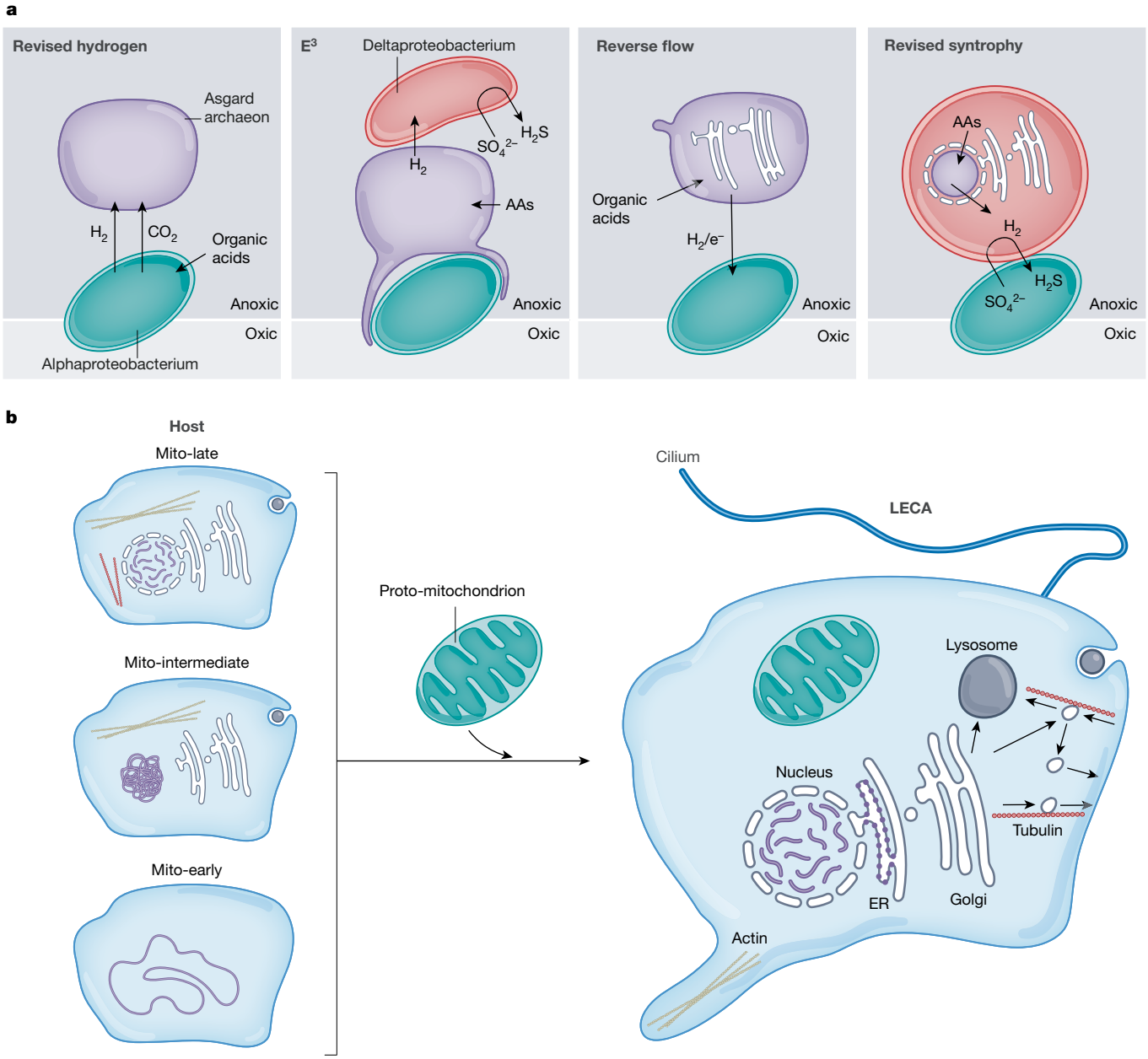

Fig. 2 | Ancestral reconstructions and potential interactions between the prokaryotic partners during eukaryogenesis. a, Symbiotic hypotheses for eukaryogenesis. The players involved are coloured on the basis of their phylogenetic affiliation. AAs, amino acids. **b**, Hypotheses for the timing of proto-mitochondrial symbiosis with respect to the cellular complexity of the host and the transition to LECA. Differences between mito-late, mito-intermediate and mito-early hypotheses regarding the presence of an endomembrane system, cytoskeleton and phagocytic capabilities in the host are shown. By contrast, the cellular complexity of LECA with, among other features, a nucleus, endoplasmic reticulum (ER), Golgi apparatus, lysosome, cytoplasmic vesicles, mitochondria, actin-based and tubulin-based cytoskeleton, cilium and phagocytosis is widely agreed upon.

It opens the possibility that the proto-mitochondrial and archaea-related symbionts interacted under (micro)oxic conditions, in line with current data supporting that eukaryogenesis occurred during a period of variable surface-water oxygenation of the Proterozoic era[90]. For reasons outlined above, the lifestyle of FMCA is less clear. According to most recent data, FMCA was most probably a free-living, facultative anaerobic proteobacterium that may have had the capacity to perform anoxygenic photosynthesis[91]. Several studies have reconstructed a diverse metabolic repertoire in FMCA, including lipid, amino acid and energy metabolism[92,93]. Yet, whereas the aerobic respiration machinery in mitochondria can be confidently traced back to FMCA[66,75,80,92,93], this is not the case for the enzymes involved in anaerobic metabolism, which are patchily distributed across diverse eukaryotes and generally do not show an alphaproteobacterial ancestry[66,94,95]. The continuing presence of oxygen-dependent metabolism in proto-mitochondria makes prolonged anoxic conditions during eukaryogenesis unlikely[96] (but see ref. 90), suggesting that, at least after the proto-mitochondrial endosymbiosis, stem eukaryotes predominantly inhabited (micro) oxic environments.

Several mechanisms have been proposed for the proto-mitochondrial acquisition by the host cell. The best known is phagocytosis, in which the proto-mitochondrial symbiont was engulfed and retained as some indigested or farmed prey[97–99]. Alternatively, a slower engulfment process has been proposed in which host protrusions gradually entangled or entrapped the proto-mitochondrial symbiont[44,100]. Recent studies[101,102] have questioned the generally accepted phagocytic

nature of LECA[16,103]. However, the inferred age of Legionellales, a clade of otherwise obligate intracellular bacteria infecting eukaryotes via the phagocytic route, supports a pre-LECA origin of phagocytosis[104]. Despite arguments that phagocytosis is only energetically favourable in the presence of mitochondria and requires the full complexity of a eukaryotic cell[105], mitochondria-lacking (amitochondriate) eukaryotes are capable of phagocytosis[106] and at least one prokaryote, the planctomycete bacterium '*Candidatus* Uabimicrobium amorphum', can engulf other cells by a phagocytosis-like mechanism[107]. The latter example is a notable case of convergent evolution since no homologues of eukaryotic phagocytosis-related genes have been detected in its genome, with the notable exception of an actin homologue, which was possibly acquired from an Asgard archaeon[52,107]. Thus, even though Asgard archaea do not seem to encode most genes that eukaryotes use for phagocytosis[108], it currently cannot be excluded that the host cell utilized membrane-deformation-dependent mechanisms to capture or entrap the proto-mitochondrion.

The timing and role of endosymbiosis

In addition to the identity and nature of the endosymbiont and host, the relative timing of the symbiosis is a hotly debated topic. Depending on the level of cellular complexity of the host cell, different scenarios have been proposed (Fig. 2b). According to the mito-early scenario, the host that took up the proto-mitochondrial symbiont resembled a typical archaeon (that is, it lacked complex cellular features), and the symbiosis was the key event that triggered the emergence of the complex features characteristic of eukaryotic cells. The original hydrogen hypothesis[85] is an archetypical mito-early example. By contrast, mito-late scenarios posit that the host cell was an amitochondriate cell reminiscent of present-day eukaryotic cells in terms of complexity. The Archezoa hypothesis[109], based on the alleged existence of several deep-branching amitochondriate protist groups, represents a mito-late model. However, owing to the discovery of mitochondria-derived genes, suggesting secondary loss of mitochondria in these protists, and their phylogenetic re-classification, support for this hypothesis has become negligible[110]. On the basis of the presence of ESPs in Asgard archaeal genomes potentially involved in eukaryote-like membrane biology, a third, mito-intermediate scenario was proposed in which the archaea-related host displayed a certain—albeit limited—degree of eukaryote-like cellular complexity prior to the proto-mitochondrial endosymbiosis[111]. Although initial characterization of the first cultured Asgard archaea revealed intriguing cellular biological features, so far there is a lack of compelling evidence for intracellular complexity in Asgard archaea, leaving the debate open for the moment.

Mito-early proponents have argued that the absence of prokaryotes that evolved eukaryote-like complexity is paramount evidence for a critical role of mitochondria. The proto-mitochondria with internal respiratory membranes controlled by their own genomes have been proposed to have released stem eukaryotes from bioenergetic constraints, enabling them to become larger[112,113]. An increase in proto-mitochondrial copy number, coupled to a reduction in size of its genomes, would have enabled the nuclear genome to expand by the surplus of available energy. This reasoning, and the underlying bioenergetic assumptions, have been challenged by others and has sparked an ongoing debate[106,114–121]. For example, recent studies have argued that prokaryotes should, in principle, be able to achieve eukaryote-like cell volumes and genome sizes[106,121]. Indeed, larger cell volumes are possible without mitochondria, as observed in giant bacteria[122] and the recently discovered phagocytic planctomycete bacterium with a considerable cell size[107] (up to 10 μm). Furthermore, prokaryotic cellular compartmentation is illustrated by organelles that are surrounded by a lipid bilayer, such as thylakoids, magnetosomes and anammoxosomes[123], and other internal membrane structures[122,124–126]. Thus, the evolution of complex eukaryote-like features does not necessarily require mitochondria.

Several studies have used phylogenetic analyses to assess the relative timing of genetic influxes during eukaryogenesis[54,127]. A considerable influx of bacterial genes was inferred before the proto-mitochondrial endosymbiosis[127], which has sparked renewed interest in the serial endosymbiosis theory, as these early bacterial contributions might have been the result of pre-mitochondrial (endo)symbioses[128]. A later study that additionally analysed gene duplications inferred that, besides this early influx of bacterial genes, a first wave of gene duplications expanded the genome of stem eukaryotes during early stages of eukaryogenesis, enabling the emergence of a more complex eukaryote-like cell with a dynamic cytoskeleton and membrane trafficking[54]. Some of these early duplicated genes have been inferred to contain introns, suggesting the early presence of a proto-nuclear structure separating transcription from translation[129]. A second wave of gene duplications following the proto-mitochondrial acquisition has been suggested to have resulted in the evolution of a more sophisticated endomembrane system and more intricate systems for cell signalling and gene expression[54]. Further development and validation of these timing methods could improve the obtained time estimates of pre-LECA events[11,17,130,131].

Together, the above arguments, combined with the observed genetic potential for cellular complexity in Asgard archaea, suggest that stem eukaryotes had evolved some cellular complexity prior to, and independent of, the acquisition of the proto-mitochondrial symbiont, lending support for mito-intermediate scenarios of eukaryogenesis.

Genomic innovation from FECA to LECA

The innate complex nature of modern eukaryotic cells is shaped by an extensive protein repertoire encoded by their large genomes. In contrast to earlier studies indicating a relatively limited size[18,132], the LECA genome was inferred to have encoded well over 10,000 genes in a recent study[54]. The inferred gene content of the heimdallarchaeial ancestor (around 4,000 genes[8]) from which the eukaryotic lineage is likely to have emerged, though relatively large compared to that of other archaea, is still considerably smaller than that of LECA. The abundance of genomic novelty acquired by stem eukaryotes is the result of several evolutionary mechanisms that jointly contributed to the emergence of eukaryotic complexity (Fig. 3). The best characterized mechanism is the influx of genes via the proto-mitochondrial endosymbiosis. The endosymbiont lost most of its gene content, but genes involved in protein synthesis, Fe/S cluster biosynthesis and metabolism (in particular aerobic respiration[93,133]) were retained and often transferred to the nuclear genome in a process known as endosymbiotic gene transfer[134]. The evolved dependency of stem eukaryotes on proto-mitochondrial Fe/S cluster biosynthesis is especially remarkable, as a complete loss of the organelle has only occurred when mitochondrial Fe/S cluster biosynthesis was replaced by a cytosolic system acquired via horizontal gene transfer[135]. Whereas some of these proto-mitochondrial genes retained their original function within the mitochondrion, others were repurposed to operate in other pathways or organelles, such as in the peroxisome[136].

FECA and the proto-mitochondrial symbiont are unlikely to have been the only contributing lineages to the genome content of LECA: the eukaryotic stem lineage was also imbued with genes from other prokaryotes[54,127,137] and viruses[138] (but see ref. 139). Nonetheless, we lack a complete picture of the identity of the donating lineages, as well as how and when stem eukaryotes acquired these genes. This is most probably owing to pervasive horizontal gene transfer among prokaryotes and insufficient phylogenetic signal that is needed to trace the evolutionary history of these proteins. A prime example is represented by the genes involved in eukaryotic fatty acid biosynthesis. Whereas eukaryotic membranes contain bacterial-type, fatty acid-based lipids, FECA is inferred to have had archaeal, isoprenoid-based lipids. It has been

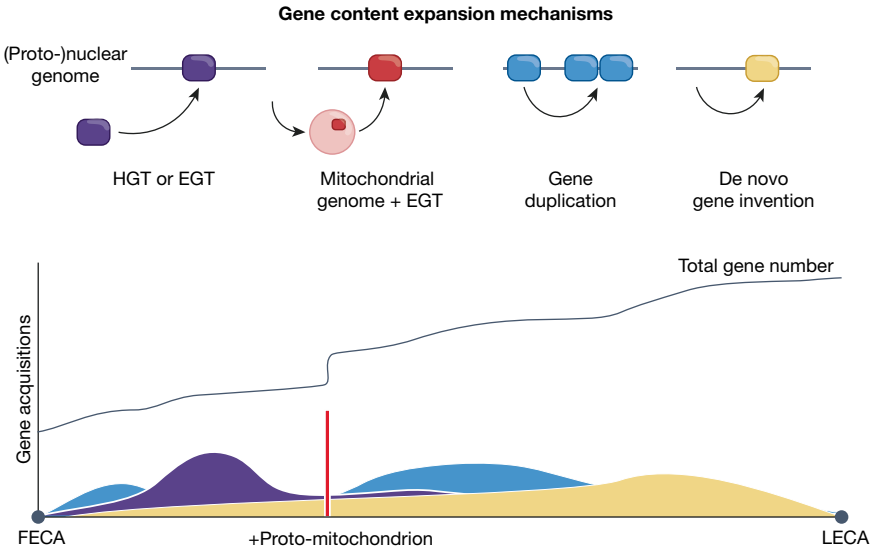

Fig. 3 | Genome evolution during eukaryogenesis. The relative contribution and timing of different mechanisms that expanded the genome between FECA and LECA are illustrated, loosely based on the branch length-based estimates from single gene trees[54,127]. They comprise the contributions to both the nuclear and mitochondrial genomes. The red bar indicates the gain of proto-mitochondrial genes upon endosymbiosis. The timing estimates of de novo gene inventions correspond to a minimal estimate of their age based on the earliest duplications they underwent. The black line illustrates the total number of genes in the composite genome, including only those genes that were still present in LECA. The loss of proto-mitochondrial and FECA genes before LECA cannot be timed. EGT, endosymbiotic gene transfer; HGT, horizontal gene transfer.

proposed that eukaryotes obtained their fatty acid lipid biosynthesis genes from the mitochondrial ancestor[140], yet none of these genes show a clear alphaproteobacterial provenance[141]. More generally, given the magnitude and predominantly pre-mitochondrial origin of genes that were acquired during eukaryogenesis[54,127], it seems unlikely that most of these genes were ultimately derived from the proto-mitochondrial endosymbiosis.

In addition to horizontal and endosymbiotic acquisition, genome expansion during eukaryogenesis has long been recognized to have encompassed many gene duplication events, accounting for almost a doubling of the genome content[18,54]. Genes of Asgard archaeal origin were shown to have duplicated particularly frequently[54]. These genes often function in processes that are regularly affected by gene duplications, such as cell-shape formation and intracellular trafficking[54]. Gene duplications increased the complexity of eukaryotic cellular machineries, such as observed in the proteasome[142], spliceosome[143], kinetochore[144] or microtubules[145]. Such complexity is sometimes associated with functional novelty, but it may also result from non-adaptive mutations making proteins interdependent in a ratchet-like, irreversible process[146]. Additionally, many proto-eukaryotic paralogues operate in different eukaryotic machineries or processes, such as in different compartments of the endomembrane system, where duplication and subsequent divergence of multiple gene families may have contributed to the emergence of different organelles[147,148].

Truly novel eukaryotic genes that originated de novo through gene genesis from non-coding DNA are arguably the most enigmatic. Previous work indicated that a substantial fraction—22% to 40%—of LECA genes[54,132] do not exhibit discernible homology to prokaryotic sequences. Some of these seemingly de novo LECA genes may in fact have prokaryotic homologues—for example, because these homologues are not represented in currently available genomes or because they are too divergent from the eukaryotic sequences to detect them as homologues. Using recently developed protein structure prediction and comparison algorithms[149,150], homology with prokaryotic proteins may eventually be detected, which will narrow down the extent of true eukaryotic novelty.

In addition to gene-level innovation, much of the genetic innovation in eukaryotes occurred within genes, giving rise to novel protein products—for example, through domain rearrangements and duplications, extreme sequence divergence and the acquisition of coiled-coil and intrinsically disordered regions[18,54,151–153]. Such innovations may have been facilitated by the emergence of meiosis[154–156] and the establishment of spliceosomal introns[157,158]. Although the overall scale and impact of these innovations and their underlying mechanisms during eukaryogenesis are not yet well understood, they are known to have had important roles in other evolutionary transitions, such as the emergence of animals[159,160]. Furthermore, a specific type of domain rearrangement, between domains of archaeal and of bacterial origin, has been proposed to have enhanced the evolution of eukaryotic information-processing machineries, which might have contributed to extensive crosstalk and more efficient functional interactions between proteins of different evolutionary origins[161]. In sum, it is becoming increasingly clear that eukaryotic stem lineages contained an expanded genomic content and were genetically diverse and dynamic. Future studies will reveal whether some of the processes underlying the emergence of genomic novelty and complexity were already established in Asgard lineages before FECA[8,45].

Conclusions and future perspective

The metagenomic exploration of Earth's prokaryotic diversity, combined with the development of powerful computational approaches to analyse large amounts of genome data, has deeply impacted the field of eukaryogenesis in the past decade. The discovery of an extended diversity of archaeal, including Asgard archaea, and bacterial lineages provided several new insights and hypotheses about the identity and nature of the symbionts involved, as well as the symbiotic interaction that drove them together. Of particular interest are efforts to bring members of the Asgard archaea into culture. The recent successful cultivation of two Lokiarchaeia-related lineages revealed a first glimpse of their unique cellular ultrastructure, with membrane protrusions facilitated by a eukaryote-like actin cytoskeleton. Future cultivation efforts of a phylogenetically diverse set of Asgard lineages, including those more closely related to eukaryotes, such as the Hodarchaeales, might reveal additional clues about the cellular and physiological features of FECA. Given the significant time gap of more than two billion years

estimated between FECA and present-day Asgard lineages[10,11], however, inferring details about the process of eukaryogenesis should be guided by an evolutionary framework. Several computational studies utilizing such a strategy have started to provide an increased resolution about the order and timing of the key events that shaped the complex nature of eukaryotic cells[8,54,127]. These approaches enable multiple aspects of eukaryogenesis hypotheses to be scrutinized, although the lack of distinct nodes in phylogenetic trees for the host and proto-mitochondrial endosymbiont complicates their direct testing[17]. The development of new, sophisticated tools to analyse the growing amounts of molecular sequence and protein structure data more efficiently will undoubtedly help to move the eukaryogenesis field forward. Concerted efforts of studies implementing such tools, combined with those utilizing cultivation-based approaches, will certainly reveal new, exciting pieces of the eukaryogenesis puzzle in the coming decade.

1. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl Acad. Sci. USA* **74**, 5088–5090 (1977).
This seminal paper was the first to recognize archaea—then called archaebacteria—as a separate prokaryotic group from bacteria.
2. Woese, C. R. Bacterial evolution. *Microbiol. Rev.* **51**, 221–271 (1987).
3. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains archaea, bacteria, and eucarya. *Proc. Natl Acad. Sci. USA* **87**, 4576–4579 (1990).
4. Huet, J., Schnabel, R., Sentenac, A. & Zillig, W. Archaebacteria and eukaryotes possess DNA-dependent RNA polymerases of a common type. *EMBO J.* **2**, 1291–1294 (1983).
5. Ouzounis, C. & Sander, C. TFIIB, an evolutionary link between the transcription machineries of archaebacteria and eukaryotes. *Cell* **71**, 189–190 (1992).
6. Myllykallio, H. et al. Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science* **288**, 2212–2215 (2000).
7. Williams, T. A., Cox, C. J., Foster, P. G., Szöllősi, G. J. & Embley, T. M. Phylogenomics provides robust support for a two-domains tree of life. *Nat. Ecol. Evol.* **4**, 138–147 (2020).
Using better-fitting models and additional in-depth analyses, this study scrutinized previous studies that reported 3D trees, resulting in robust 2D trees that show a close relationship between Heimdallarchaeia and eukaryotes.
8. Eme, L. et al. Inference and reconstruction of the heimdallarchaeial ancestry of eukaryotes. *Nature* **618**, 992–999 (2023).
This study presented the expanding diversity of Asgard archaea, the Hodarchaeales–sister relationship of eukaryotes based on elaborate phylogenomics, the presence of additional ESPs in Asgard genomes and the reconstructed gene content of Asgard ancestral nodes.
9. 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).
10. Betts, H. C. et al. Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat. Ecol. Evol.* **2**, 1556–1562 (2018).
11. Mahendrarajah, T. A. et al. ATP synthase evolution on a cross-braced dated tree of life. *Nat. Commun.* **14**, 7456 (2023).
12. Eme, L., Sharpe, S. C., Brown, M. W. & Roger, A. J. On the age of eukaryotes: evaluating evidence from fossils and molecular clocks. *Cold Spring Harb. Perspect. Biol.* **6**, a016139 (2014).
13. Cohen, P. A. & Kodner, R. B. The earliest history of eukaryotic life: uncovering an evolutionary story through the integration of biological and geological data. *Trends Ecol. Evol.* **37**, 246–256 (2022).
14. Brocks, J. J. et al. Lost world of complex life and the late rise of the eukaryotic crown. *Nature* **618**, 767–773 (2023).
15. Porter, S. M. & Riedman, L. A. Frameworks for interpreting the early fossil record of eukaryotes. *Annu. Rev. Microbiol.* **77**, 173–191 (2023).
16. Koumandou, V. L. et al. Molecular paleontology and complexity in the last eukaryotic common ancestor. *Crit. Rev. Biochem. Mol. Biol.* **48**, 373–396 (2013).
17. Donoghue, P. C. J. et al. Defining eukaryotes to dissect eukaryogenesis. *Curr. Biol.* **33**, R919–R929 (2023).
18. Makarova, K. S., Wolf, Y. I., Mekhedov, S. L., Mirkin, B. G. & Koonin, E. V. Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res.* **33**, 4626–4638 (2005).
This paper provided a first systematic estimate of the number of gene acquisitions, duplications and inventions during eukaryogenesis based on the homology between eukaryotic clusters of orthologues and between eukaryotic and prokaryotic gene clusters.
19. O'Malley, M. A., Leger, M. M., Wideman, J. G. & Ruiz-Trillo, I. Concepts of the last eukaryotic common ancestor. *Nat. Ecol. Evol.* **3**, 338–344 (2019).
20. 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).
21. Dacks, J. B. et al. The changing view of eukaryogenesis—fossils, cells, lineages and how they all come together. *J. Cell Sci.* **129**, 3695–3703 (2016).
22. Woese, C. R. & Olsen, G. J. Archaebacterial phylogeny: perspectives on the Urkingdoms. *Syst. Appl. Microbiol.* **7**, 161–177 (1986).
23. Lake, J. A. Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. *Nature* **331**, 184–186 (1988).
24. Gouy, M. & Li, W.-H. Phylogenetic analysis based on rRNA sequences supports the archaebacterial rather than the eocyte tree. *Nature* **339**, 145–147 (1989).
25. Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. & Miyata, T. Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl Acad. Sci. USA* **86**, 9355–9359 (1989).
26. Baldauf, S. L., Palmer, J. D. & Doolittle, W. F. The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. *Proc. Natl Acad. Sci. USA* **93**, 7749–7754 (1996).
27. Lake, J. A., Henderson, E., Oakes, M. & Clark, M. W. Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *Proc. Natl Acad. Sci. USA* **81**, 3786–3790 (1984).
On the basis of ribosome structures, the authors of this study postulated the eocyte hypothesis, in which eukaryotes are most closely related to a specific group of archaea (the 2D tree).
28. Rivera, M. C. & Lake, J. A. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* **257**, 74–76 (1992).
29. Brown, J. R., Douady, C. J., Italia, M. J., Marshall, W. E. & Stanhope, M. J. Universal trees based on large combined protein sequence data sets. *Nat. Genet.* **28**, 281–285 (2001).
30. Ciccarelli, F. D. et al. Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**, 1283–1287 (2006).
31. Cox, C. J., Foster, P. G., Hirt, R. P., Harris, S. R. & Embley, T. M. The archaebacterial origin of eukaryotes. *Proc. Natl Acad. Sci. USA* **105**, 20356–20361 (2008).
Using phylogenetic models that take compositional changes into account, the 2D tree was robustly recovered for the first time in this phylogenomics study.
32. Foster, P. G., Cox, C. J. & Embley, T. M. The primary divisions of life: a phylogenomic approach employing composition-heterogeneous methods. *Phil. Trans. R. Soc. B* **364**, 2197–2207 (2009).
33. Guy, L. & Ettema, T. J. G. The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol.* **19**, 580–587 (2011).
34. Kelly, S., Wickstead, B. & Gull, K. Archaeal phylogenomics provides evidence in support of a methanogenic origin of the archaea and a thaumarchaeal origin for the eukaryotes. *Proc. R. Soc. B* **278**, 1009–1018 (2011).
35. Lasek-Nesselquist, E. & Gogarten, J. P. The effects of model choice and mitigating bias on the ribosomal tree of life. *Mol. Phylogenetics Evol.* **69**, 17–38 (2013).
36. Guy, L., Saw, J. H. & Ettema, T. J. G. The archaeal legacy of eukaryotes: a phylogenomic perspective. *Cold Spring Harb. Perspect. Biol.* **6**, a016022 (2014).
37. Williams, T. A. & Embley, T. M. Archaeal "dark matter" and the origin of eukaryotes. *Genome Biol. Evol.* **6**, 474–481 (2014).
38. Raymann, K., Brochier-Armanet, C. & Gribaldo, S. The two-domain tree of life is linked to a new root for the Archaea. *Proc. Natl Acad. Sci. USA* **112**, 6670–6675 (2015).
39. Spang, A. et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179 (2015).
This paper described the discovery of the first Asgard archaeon, *Lokiarchaeum*, and showed both its close relationship with eukaryotes and the presence of multiple new ESPs in its genome.
40. 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).
41. Zaremba-Niedzwiedzka, K. et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).
42. Spang, A. et al. Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. *Nat. Microbiol.* **4**, 1138–1148 (2019).
43. Seitz, K. W. et al. Asgard archaea capable of anaerobic hydrocarbon cycling. *Nat. Commun.* **10**, 1822 (2019).
44. Imachi, H. et al. Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature* **577**, 519–525 (2020).
This study presented the first cultured Asgard archaeon, the lokiarchaeon *Candidatus* P. syntrophicum, showing remarkable cell physiology (see also ref. 50).
45. Liu, Y. et al. Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature* **593**, 553–557 (2021).
46. Sun, J. et al. Recoding of stop codons expands the metabolic potential of two novel Asgardarchaeota lineages. *ISME Commun.* **1**, 30 (2021).
47. Aouad, M. et al. A divide-and-conquer phylogenomic approach based on character supermatrices resolves early steps in the evolution of the archaea. *BMC Ecol. Evo.* **22**, 1 (2022).
48. Wu, F. et al. Unique mobile elements and scalable gene flow at the prokaryote–eukaryote boundary revealed by circularized Asgard archaea genomes. *Nat. Microbiol.* **7**, 200–212 (2022).
49. Xie, R. et al. Expanding Asgard members in the domain of archaea sheds new light on the origin of eukaryotes. *Sci. China Life Sci.* **65**, 818–829 (2022).
50. Rodrigues-Oliveira, T. et al. Actin cytoskeleton and complex cell architecture in an Asgard archaeon. *Nature* **613**, 332–339 (2023).
51. 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).
52. Stairs, C. W. & Ettema, T. J. G. The archaeal roots of the eukaryotic dynamic actin cytoskeleton. *Curr. Biol.* **30**, R521–R526 (2020).
53. 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).
54. Vosseberg, J. et al. Timing the origin of eukaryotic cellular complexity with ancient duplications. *Nat. Ecol. Evol.* **5**, 92–100 (2021).
This paper reconstructed the numerous gene duplications that occurred during eukaryogenesis from phylogenetic trees and inferred their relative timing, also in comparison with gene transfer events, using the branch lengths approach adapted from ref. 127.
55. Szöllősi, G. J., Rosikiewicz, W., Boussau, B., Tannier, E. & Daubin, V. Efficient exploration of the space of reconciled gene trees. *Syst. Biol.* **62**, 901–912 (2013).

56. Williams, T. A. et al. Parameter estimation and species tree rooting using ALE and GeneRax. *Genome Biol. Evol.* **15**, evad134 (2023).
57. Akıl, C. & Robinson, R. C. Genomes of Asgard archaea encode profilins that regulate actin. *Nature* **562**, 439–443 (2018).
This article is the first of a series of biochemical papers investigating the molecular function of Asgard ESPs by expressing them in heterologous systems, in this case focusing on the interaction between Asgard profilin and eukaryotic actin.
58. Akıl, C. et al. Insights into the evolution of regulated actin dynamics via characterization of primitive gelsolin/cofilin proteins from Asgard archaea. *Proc. Natl Acad. Sci. USA* **117**, 19904–19913 (2020).
59. Survery, S. et al. Heimdallarchaea encodes profilin with eukaryotic-like actin regulation and polyproline binding. *Commun. Biol.* **4**, 1024 (2021).
60. Akıl, C. et al. Structure and dynamics of Odinarchaeota tubulin and the implications for eukaryotic microtubule evolution. *Sci. Adv.* **8**, eabm2225 (2022).
61. Leung, K. F., Dacks, J. B. & Field, M. C. Evolution of the multivesicular body ESCRT machinery; retention across the eukaryotic lineage. *Traffic* **9**, 1698–1716 (2008).
62. Hatano, T. et al. Asgard archaea shed light on the evolutionary origins of the eukaryotic ubiquitin–ESCRT machinery. *Nat. Commun.* **13**, 3398 (2022).
63. Neveu, E., Khalifeh, D., Salamin, N. & Fasshauer, D. Prototypic SNARE proteins are encoded in the genomes of Heimdallarchaeota, potentially bridging the gap between the prokaryotes and eukaryotes. *Curr. Biol.* **30**, 2468–2480 (2020).
64. Avcı, B. et al. Spatial separation of ribosomes and DNA in Asgard archaeal cells. *ISME J.* **16**, 606–610 (2022).
65. Gray, M. W., Burger, G. & Lang, B. F. Mitochondrial evolution. *Science* **283**, 1476–1481 (1999).
66. Roger, A. J., Muñoz-Gómez, S. A. & Kamikawa, R. The origin and diversification of mitochondria. *Curr. Biol.* **27**, R1177–R1192 (2017).
67. Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G. J. & Woese, C. R. Mitochondrial origins. *Proc. Natl Acad. Sci. USA* **82**, 4443–4447 (1985).
68. Fitzpatrick, D. A., Creevey, C. J. & McInerney, J. O. Genome phylogenies indicate a meaningful α-proteobacterial phylogeny and support a grouping of the mitochondria with the Rickettsiales. *Mol. Biol. Evol.* **23**, 74–85 (2006).
69. Williams, K. P., Sobral, B. W. & Dickerman, A. W. A robust species tree for the Alphaproteobacteria. *J. Bacteriol.* **189**, 4578–4586 (2007).
70. Thrash, J. C. et al. Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade. *Sci. Rep.* **1**, 13 (2011).
71. Georgiades, K., Madoui, M.-A., Le, P., Robert, C. & Raoult, D. Phylogenomic analysis of *Odyssella thessalonicensis* fortifies the common origin of Rickettsiales, *Pelagibacter ubique* and *Reclimonas americana* mitochondrion. *PLoS ONE* **6**, e24857 (2011).
72. Sassera, D. et al. Phylogenomic evidence for the presence of a flagellum and cbb_3 oxidase in the free-living mitochondrial ancestor. *Mol. Biol. Evol.* **28**, 3285–3296 (2011).
73. Rodríguez-Ezpeleta, N. & Embley, T. M. The SAR11 group of alpha-proteobacteria is not related to the origin of mitochondria. *PLoS ONE* **7**, e30520 (2012).
74. Viklund, J., Martijn, J., Ettema, T. J. G. & Andersson, S. G. E. Comparative and phylogenomic evidence that the alphaproteobacterium HIMB59 is not a member of the oceanic SAR11 clade. *PLoS ONE* **8**, e78858 (2013).
75. Wang, Z. & Wu, M. Phylogenomic reconstruction indicates mitochondrial ancestor was an energy parasite. *PLoS ONE* **9**, e110685 (2014).
76. Wang, Z. & Wu, M. An integrated phylogenomic approach toward pinpointing the origin of mitochondria. *Sci. Rep.* **5**, 7949 (2015).
77. Martijn, J., Vosseberg, J., Guy, L., Offre, P. & Ettema, T. J. G. Deep mitochondrial origin outside the sampled Alphaproteobacteria. *Nature* **557**, 101–105 (2018).
This study recovered several novel marine alphaproteobacterial groups and performed careful phylogenomic analyses to address long-branch and compositional artefacts, revealing the novel Alphaproteobacteria–sister position of mitochondria.
78. Fan, L. et al. Phylogenetic analyses with systematic taxon sampling show that mitochondria branch within Alphaproteobacteria. *Nat. Ecol. Evol.* **4**, 1213–1219 (2020).
79. Wang, S. & Luo, H. Dating Alphaproteobacteria evolution with eukaryotic fossils. *Nat. Commun.* **12**, 3324 (2021).
80. Muñoz-Gómez, S. A. et al. Site-and-branch-heterogeneous analyses of an expanded dataset favour mitochondria as sister to known Alphaproteobacteria. *Nat. Ecol. Evol.* **6**, 253–262 (2022).
This study corroborated the Alphaproteobacteria–sister relationship of mitochondria using a newly developed model that accounts for compositional heterogeneity across sites and branches.
81. Martijn, J., Vosseberg, J., Guy, L., Offre, P. & Ettema, T. J. G. Phylogenetic affiliation of mitochondria with Alpha-II and Rickettsiales is an artefact. *Nat. Ecol. Evol.* **6**, 1829–1831 (2022).
82. Fan, L. et al. Reply to: Phylogenetic affiliation of mitochondria with Alpha-II and Rickettsiales is an artefact. *Nat. Ecol. Evol.* **6**, 1832–1835 (2022).
83. Ettema, T. J. G. & Andersson, S. G. E. The α-proteobacteria: the Darwin finches of the bacterial world. *Biol. Lett.* **5**, 429–432 (2009).
84. Martin, W. F., Garg, S. & Zimorski, V. Endosymbiotic theories for eukaryote origin. *Phil. Trans. R. Soc. B* **370**, 20140330 (2015).
85. Martin, W. & Müller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41 (1998).
86. Sousa, F. L., Neukirchen, S., Allen, J. F., Lane, N. & Martin, W. F. Lokiarchaeon is hydrogen dependent. *Nat. Microbiol.* **1**, 16034 (2016).
87. Moreira, D. & López-García, P. Symbiosis between methanogenic archaea and δ-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J. Mol. Evol.* **47**, 517–530 (1998).
88. López-García, P. & Moreira, D. The syntrophy hypothesis for the origin of eukaryotes revisited. *Nat. Microbiol.* **5**, 655–667 (2020).
89. Bulzu, P.-A. et al. Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche. *Nat. Microbiol.* **4**, 1129–1137 (2019).
90. Mills, D. B. et al. Eukaryogenesis and oxygen in Earth history. *Nat. Ecol. Evol.* **6**, 520–532 (2022).
91. Muñoz-Gómez, S. A., Wideman, J. G., Roger, A. J. & Slamovits, C. H. The origin of mitochondrial cristae from Alphaproteobacteria. *Mol. Biol. Evol.* **34**, 943–956 (2017).
92. Gabaldón, T. & Huynen, M. A. Reconstruction of the proto-mitochondrial metabolism. *Science* **301**, 609–609 (2003).
93. Gabaldón, T. & Huynen, M. A. From endosymbiont to host-controlled organelle: the hijacking of mitochondrial protein synthesis and metabolism. *PLoS Comput. Biol.* **3**, e219 (2007).
94. Stairs, C. W., Leger, M. M. & Roger, A. J. Diversity and origins of anaerobic metabolism in mitochondria and related organelles. *Phil. Trans. R. Soc. B* **370**, 20140326 (2015).
95. Stairs, C. W. et al. Chlamydial contribution to anaerobic metabolism during eukaryotic evolution. *Sci. Adv.* **6**, eabb7258 (2020).
96. Speijer, D. Alternating terminal electron-acceptors at the basis of symbiogenesis: How oxygen ignited eukaryotic evolution. *BioEssays* **39**, 1600174 (2017).
97. Cavalier-Smith, T. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int. J. Syst. Evol. Microbiol.* **52**, 297–354 (2002).
98. Martijn, J. & Ettema, T. J. G. From archaeon to eukaryote: the evolutionary dark ages of the eukaryotic cell. *Biochem. Soc. Trans.* **41**, 451–457 (2013).
99. Zachar, I., Szilágyi, A., Számadó, S. & Szathmáry, E. Farming the mitochondrial ancestor as a model of endosymbiotic establishment by natural selection. *Proc. Natl Acad. Sci. USA* **115**, E1504–E1510 (2018).
100. Baum, D. A. & Baum, B. An inside-out origin for the eukaryotic cell. *BMC Biol.* **12**, 76 (2014).
101. Mills, D. B. The origin of phagocytosis in Earth history. *Interface Focus* **10**, 20200019 (2020).
102. Bremer, N., Tria, F. D. K., Skejo, J., Garg, S. G. & Martin, W. F. Ancestral state reconstructions trace mitochondria but not phagocytosis to the last eukaryotic common ancestor. *Genome Biol. Evol.* **14**, evac079 (2022).
103. Yutin, N., Wolf, M. Y., Wolf, Y. I. & Koonin, E. V. The origins of phagocytosis and eukaryogenesis. *Biol. Direct* **4**, 9 (2009).
104. Hugoson, E., Guliaev, A., Ammunét, T. & Guy, L. Host adaptation in *Legionellales* Is 1.9 Ga, coincident with eukaryogenesis. *Mol. Biol. Evol.* **39**, msac037 (2022).
105. 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–e00017 (2017).
106. Hampl, V., Čepička, I. & Eliáš, M. Was the mitochondrion necessary to start eukaryogenesis? *Trends Microbiol.* **27**, 96–104 (2019).
107. Shiratori, T., Suzuki, S., Kakizawa, Y. & Ishida, K. Phagocytosis-like cell engulfment by a planctomycete bacterium. *Nat. Commun.* **10**, 5529 (2019).
108. Burns, J. A., Pittis, A. A. & Kim, E. Gene-based predictive models of trophic modes suggest Asgard archaea are not phagocytotic. *Nat. Ecol. Evol.* **2**, 697–704 (2018).
109. Cavalier-Smith, T. Archaebacteria and archezoa. *Nature* **339**, 100–101 (1989).
110. Embley, T. M. & Hirt, R. P. Early branching eukaryotes? *Curr. Opin. Genet. Dev.* **8**, 624–629 (1998).
111. Ettema, T. J. G. Evolution: mitochondria in the second act. *Nature* **531**, 39–40 (2016).
112. Lane, N. & Martin, W. The energetics of genome complexity. *Nature* **467**, 929–934 (2010).
113. Lane, N. Energetics and genetics across the prokaryote–eukaryote divide. *Biol. Direct* **6**, 35 (2011).
114. Booth, A. & Doolittle, W. F. Eukaryogenesis, how special really? *Proc. Natl Acad. Sci. USA* **112**, 10278–10285 (2015).
115. Lynch, M. & Marinov, G. K. The bioenergetic costs of a gene. *Proc. Natl Acad. Sci. USA* **112**, 15690–15695 (2015).
116. Koonin, E. V. Energetics and population genetics at the root of eukaryotic cellular and genomic complexity. *Proc. Natl Acad. Sci. USA* **112**, 15777–15778 (2015).
117. Lynch, M. & Marinov, G. K. Membranes, energetics, and evolution across the prokaryote–eukaryote divide. *eLife* **6**, e20437 (2017).
118. Lane, N. Serial endosymbiosis or singular event at the origin of eukaryotes? *J. Theor. Biol.* **434**, 58–67 (2017).
119. Chiyomaru, K. & Takemoto, K. Revisiting the hypothesis of an energetic barrier to genome complexity between eukaryotes and prokaryotes. *R. Soc. Open Sci.* **7**, 191859 (2020).
120. Lane, N. How energy flow shapes cell evolution. *Curr. Biol.* **30**, R471–R476 (2020).
121. Schavemaker, P. E. & Muñoz-Gómez, S. A. The role of mitochondrial energetics in the origin and diversification of eukaryotes. *Nat. Ecol. Evol.* **6**, 1307–1317 (2022).
122. Volland, J.-M. et al. A centimeter-long bacterium with DNA contained in metabolically active, membrane-bound organelles. *Science* **376**, 1453–1458 (2022).
123. Greening, C. & Lithgow, T. Formation and function of bacterial organelles. *Nat. Rev. Microbiol.* **18**, 677–689 (2020).
124. Küper, U., Meyer, C., Müller, V., Rachel, R. & Huber, H. Energized outer membrane and spatial separation of metabolic processes in the hyperthermophilic Archaeon Ignicoccus hospitalis. *Proc. Natl Acad. Sci. USA* **107**, 3152–3156 (2010).
125. Wiegand, S., Jogler, M. & Jogler, C. On the maverick planctomycetes. *FEMS Microbiol. Rev.* **42**, 739–760 (2018).
126. Katayama, T. et al. Isolation of a member of the candidate phylum 'Atribacteria' reveals a unique cell membrane structure. *Nat. Commun.* **11**, 6381 (2020).
127. Pittis, A. A. & Gabaldón, T. Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature* **531**, 101–104 (2016).
This study presented a novel approach to use phylogenetic branch lengths to infer the relative timing of gene acquisitions during eukaryogenesis, pointing to rampant bacterial gene flow to stem eukaryotes prior to the proto-mitochondrial acquisition.
128. Gabaldón, T. Relative timing of mitochondrial endosymbiosis and the "pre-mitochondrial symbioses" hypothesis. *IUBMB Life* **70**, 1188–1196 (2018).
129. Vosseberg, J., Schinkel, M., Gremmen, S. & Snel, B. The spread of the first introns in proto-eukaryotic paralogs. *Commun. Biol.* **5**, 476 (2022).
130. Susko, E., Steel, M. & Roger, A. J. Conditions under which distributions of edge length ratios on phylogenetic trees can be used to order evolutionary events. *J. Theor. Biol.* **526**, 110788 (2021).
131. Tricou, T., Tannier, E. & de Vienne, D. M. Ghost lineages can invalidate or even reverse findings regarding gene flow. *PLoS Biol.* **20**, e3001776 (2022).

132. Fritz-Laylin, L. K. et al. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell* **140**, 631–642 (2010).
133. Huynen, M. A., Duarte, I. & Szklarczyk, R. Loss, replacement and gain of proteins at the origin of the mitochondria. *Biochim. Biophys. Acta* **1827**, 224–231 (2013).
134. Timmis, J. N., Ayliffe, M. A., Huang, C. Y. & Martin, W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5**, 123–135 (2004).
135. Karnkowska, A. et al. A eukaryote without a mitochondrial organelle. *Curr. Biol.* **26**, 1274–1284 (2016).
136. Gabaldón, T. et al. Origin and evolution of the peroxisomal proteome. *Biol. Direct* **1**, 8 (2006).
137. 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).
138. Irwin, N. A. T., Pittis, A. A., Richards, T. A. & Keeling, P. J. Systematic evaluation of horizontal gene transfer between eukaryotes and viruses. *Nat. Microbiol.* **7**, 327–336 (2022).
139. Ku, C. et al. Endosymbiotic origin and differential loss of eukaryotic genes. *Nature* **524**, 427–432 (2015).
140. Gould, S. B., Garg, S. G. & Martin, W. F. Bacterial vesicle secretion and the evolutionary origin of the eukaryotic endomembrane system. *Trends Microbiol.* **24**, 525–534 (2016).
141. Coleman, G. A., Pancost, R. D. & Williams, T. A. Investigating the origins of membrane phospholipid biosynthesis genes using outgroup-free rooting. *Genome Biol. Evol.* **11**, 883–898 (2019).
142. Volker, C. & Lupas, A. N. in *The Proteasome–Ubiquitin Protein Degradation Pathway* (eds Zwickl, P. & Baumeister, W.) 1–22 (Springer, 2002).
143. Vosseberg, J., Stolker, D., von der Dunk, S. H. A. & Snel, B. Integrating phylogenetics with intron positions illuminates the origin of the complex spliceosome. *Mol. Biol. Evol.* **40**, msad011 (2023).
144. Tromer, E. C., Hooff, J. J. E., van, Kops, G. J. P. L. & Snel, B. Mosaic origin of the eukaryotic kinetochore. *Proc. Natl Acad. Sci. USA* **116**, 12873–12882 (2019).
145. Findeisen, P. et al. Six subgroups and extensive recent duplications characterize the evolution of the eukaryotic tubulin protein family. *Genome Biol. Evol.* **6**, 2274–2288 (2014).
146. Muñoz-Gómez, S. A., Bilolikar, G., Wideman, J. G. & Geiler-Samerotte, K. Constructive neutral evolution 20 years later. *J. Mol. Evol.* **89**, 172–182 (2021).
147. Dacks, J. B. & Field, M. C. Evolution of the eukaryotic membrane-trafficking system: origin, tempo and mode. *J. Cell Sci.* **120**, 2977–2985 (2007).
148. Dacks, J. B. & Field, M. C. Evolutionary origins and specialisation of membrane transport. *Curr. Opin. Cell Biol.* **53**, 70–76 (2018).
149. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
150. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
151. Ekman, D., Björklund, Å. K., Frey-Skött, J. & Elofsson, A. Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions. *J. Mol. Biol.* **348**, 231–243 (2005).
152. Liu, J. & Rost, B. Comparing function and structure between entire proteomes. *Protein Sci.* **10**, 1970–1979 (2001).
153. Xue, B., Dunker, A. K. & Uversky, V. N. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J. Biomol. Struct. Dyn.* **30**, 137–149 (2012).
154. Colnaghi, M., Lane, N. & Pomiankowski, A. Genome expansion in early eukaryotes drove the transition from lateral gene transfer to meiotic sex. *eLife* **9**, e58873 (2020).
155. van Dijk, B., Bertels, F., Stolk, L., Takeuchi, N. & Rainey, P. B. Transposable elements promote the evolution of genome streamlining. *Phil. Trans. R. Soc. B* **377**, 20200477 (2022).
156. Colnaghi, M., Lane, N. & Pomiankowski, A. Repeat sequences limit the effectiveness of lateral gene transfer and favored the evolution of meiotic sex in early eukaryotes. *Proc. Natl Acad. Sci. USA* **119**, e2205041119 (2022).
157. Gilbert, W. Why genes in pieces? *Nature* **271**, 501–501 (1978).
158. Liu, M. & Grigoriev, A. Protein domains correlate strongly with exons in multiple eukaryotic genomes – evidence of exon shuffling? *Trends Genet.* **20**, 399–403 (2004).
159. Grau-Bové, X. et al. Dynamics of genomic innovation in the unicellular ancestry of animals. *eLife* **6**, e26036 (2017).
160. Ocaña-Pallarès, E. et al. Divergent genomic trajectories predate the origin of animals and fungi. *Nature* **609**, 747–753 (2022).
161. Méheust, R. et al. Formation of chimeric genes with essential functions at the origin of eukaryotes. *BMC Biol.* **16**, 30 (2018).
162. Tamarit, D. et al. Description of *Asgardarchaeum abyssi* gen. nov. spec. nov., a novel species within the class *Asgardarchaeia* and phylum *Asgardarchaeota* in accordance with the SeqCode. *Syst. Appl. Microbiol.* **47**, 126525 (2024).
163. Delsuc, F., Brinkmann, H. & Philippe, H. Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* **6**, 361–375 (2005).
164. Kapli, P., Yang, Z. & Telford, M. J. Phylogenetic tree building in the genomic age. *Nat. Rev. Genet.* **21**, 428–444 (2020).
165. Steenwyk, J. L., Li, Y., Zhou, X., Shen, X.-X. & Rokas, A. Incongruence in the phylogenomics era. *Nat. Rev. Genet.* **24**, 834–850 (2023).
166. Fleming, J. F., Valero-Gracia, A. & Struck, T. H. Identifying and addressing methodological incongruence in phylogenomics: a review. *Evol. Appl.* **16**, 1087–1104 (2023).
167. Foster, P. G. et al. Recoding amino acids to a reduced alphabet may increase or decrease phylogenetic accuracy. *Syst. Biol.* **72**, 723–737 (2023).
168. Susko, E. & Roger, A. J. On reduced amino acid alphabets for phylogenetic inference. *Mol. Biol. Evol.* **24**, 2139–2150 (2007).
169. Viklund, J., Ettema, T. J. G. & Andersson, S. G. E. Independent genome reduction and phylogenetic reclassification of the oceanic SAR11 clade. *Mol. Biol. Evol.* **29**, 599–615 (2012).

Acknowledgements The authors thank B. Snel for helpful advice. This work was supported by the Dutch Research Council (VI.C.192.016 to T.J.G.E. and VI.Veni.212.099 to J.J.E.v.H.), the European Research Council (ERC Consolidator grant 817834 to T.J.G.E.), Volkswagen Foundation ('Life' grant 96725 to T.J.G.E.) and the Moore-Simons Project on the Origin of the Eukaryotic Cell (Simons Foundation 73592LPI to T.J.G.E.).

Author contributions All authors contributed to the writing of this manuscript.

Competing interests The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Thijs J. G. Ettema.

Peer review information *Nature* thanks Toni Gabaldón and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.