Archaea and the origin of eukaryotes

Laura Eme, Anja Spang, Jonathan Lombard, Courtney W. Stairs and Thijs J. G. Ettema

Abstract | Woese and Fox's 1977 paper on the discovery of the Archaea triggered a revolution in the field of evolutionary biology by showing that life was divided into not only prokaryotes and eukaryotes. Rather, they revealed that prokaryotes comprise two distinct types of organisms, the Bacteria and the Archaea. In subsequent years, molecular phylogenetic analyses indicated that eukaryotes and the Archaea represent sister groups in the tree of life. During the genomic era, it became evident that eukaryotic cells possess a mixture of archaeal and bacterial features in addition to eukaryotic-specific features. Although it has been generally accepted for some time that mitochondria descend from endosymbiotic alphaproteobacteria, the precise evolutionary relationship between eukaryotes and archaea has continued to be a subject of debate. In this Review, we outline a brief history of the changing shape of the tree of life and examine how the recent discovery of a myriad of diverse archaeal lineages has changed our understanding of the evolutionary relationships between the three domains of life and the origin of eukaryotes. Furthermore, we revisit central questions regarding the process of eukaryogenesis and discuss what can currently be inferred about the evolutionary transition from the first to the last eukaryotic common ancestor.

Sister groups
Two descendants that split from the same node; the descendants are each other's closest relative.

Monophyletic groups
A monophyletic group is a group of organisms that forms a clade, which consists of all the descendants of a common ancestor.

Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Box 596, Uppsala SE-75123, Sweden.

Correspondence to T.J.G.E.
thijs.ettema@icm.uu.se

doi:10.1038/nrmicro.2017.133
Published online 10 Nov 2017; corrected online 27 Nov 2017

The pioneering work by Carl Woese and colleagues revealed that all cellular life could be divided into three major evolutionary lines (also called domains): the Eukarya (or eukaryotes), the Bacteria and the Archaea[1,2] (FIG. 1). In the late 1980s, phylogenetic trees that were constructed on the basis of ancient gene duplications provided the first strong evidence that eukaryotes and archaea were sister groups[2,3]. This tree topology is generally referred to as the three-domains tree of life (FIG. 1a). The discovery of several molecular features that are shared between only archaea and eukaryotes appeared consistent with this rooting of the tree of life. For example, archaeal RNA polymerases were found to be more complex than their bacterial counterparts, and their subunit composition was found to resemble that of eukaryotes[4,5]. However, the proposed evolutionary relationship between monophyletic groups of the Eukarya and the Archaea has been challenged. The three-dimensional structures of ribosomes and a shared amino acid insertion in a conserved region of the elongation factor 1α homologues indicated that eukaryotes are a sister group of the 'eocyte archaea' (that is, Crenarchaeota)[6,7]. In other words, these analyses supported the idea that eukaryotes emerged from within the Archaea, which would be in favour of a two-domains tree of life. In this evolutionary scenario, Archaea and Bacteria represent the only primary domains of life, and eukaryotes later emerged from lineages within these groups (FIG. 1b).

In this Review, we discuss how culture-independent genomics has transformed our understanding of archaeal diversity and how this has influenced our understanding of the topology of the tree of life. Specifically, we discuss how the discovery of novel archaeal superphyla combined with improved molecular phylogenetic approaches has led to unprecedented insights into eukaryogenesis — the processes by which eukaryotic cells evolved from prokaryotic precursors. We outline the main questions that need to be addressed to understand the process of eukaryogenesis, provide details on how archaeal research has allowed us to start answering some of these questions and highlight future research priorities.

Uncovering archaeal diversity

In the 1990s, the advent of DNA sequencing in molecular phylogenetics enabled the analysis of various genes (or markers) to investigate the evolutionary relationships between major branches in the tree of life. Initially, these analyses were limited to a small number of genes that had been sequenced from a limited number of cultivated organisms, and for a long time, the Crenarchaeota and the Euryarchaeota were the only recognized archaeal phyla. These analyses, together with the simpler phylogenetic methods and evolutionary models that were available at the time, yielded conflicting results, and the relationships among the domains of life remained

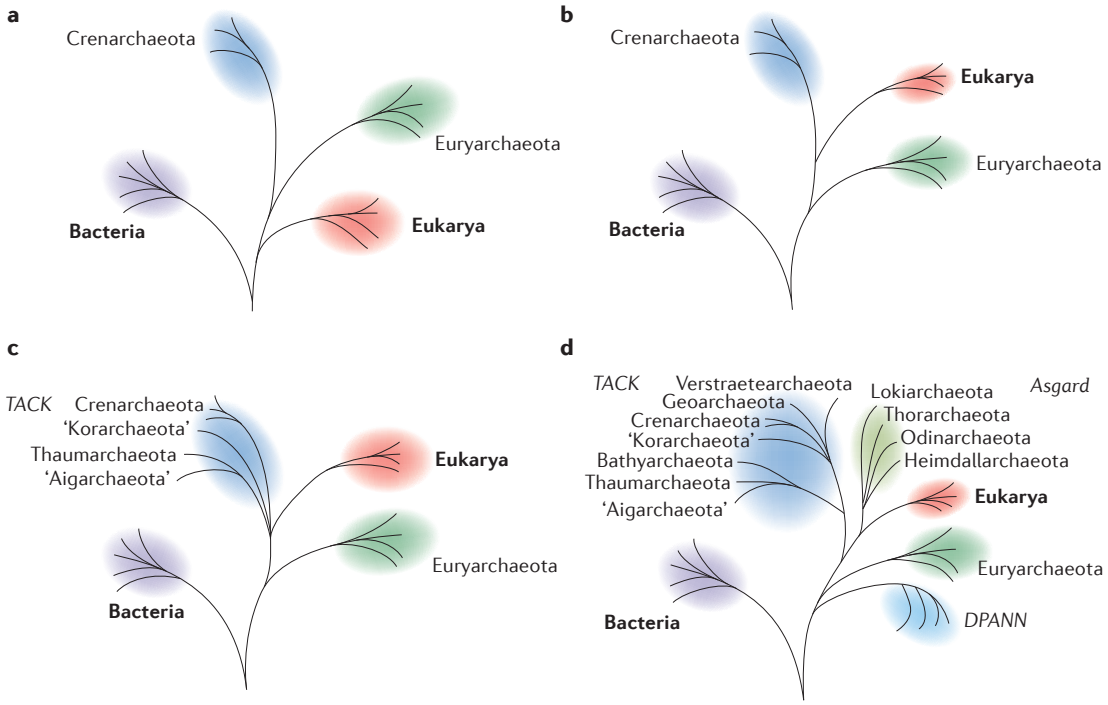

Figure 1 | **Evolution of the tree of life.** A schematic representation of our understanding of the relationships between eukaryotes and archaea over the past 40 years. In a three-domains tree of life (part **a**), Archaea and Eukarya each represent a monophyletic group and share a unique common ancestor to the exclusion of Bacteria[2,3]. By contrast, a two-domain tree of life was proposed early on, and the topology of this tree has undergone several changes. When the Archaea was thought to consist of the Euryarchaeota and the Crenarchaeota phyla only (part **b**), several kinds of molecular evidence supported the close relationship of eukaryotes and Crenarchaeota[6,7]. In the early 2010s, eukaryotes were found to branch within, or as sister to, the Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota (TACK) superphylum[25,28–32] (part **c**). Phylogenomic analyses that included members of the Asgard superphylum strongly suggested that eukaryotes originated from within the Asgard archaea or that they represented a sister group to them[47,52] (part **d**). Bacteria and Eukarya are indicated in light purple and red, respectively, whereas green and blue represent archaeal lineages. Domain and superphylum rank-level names are bolded or italicized, respectively. Names of proposed phyla are in quotation marks. Note that the Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota (DPANN) lineage is represented as a monophyletic lineage, although this is a topic of debate[155,156].

Eukaryogenesis
The whole sequence of evolutionary events occurring between the first eukaryotic common ancestor (FECA) and the last eukaryotic common ancestor (LECA) explaining the process by which eukaryotic cells evolved from prokaryotic ancestors.

Metagenomics
The sequencing of genetic material extracted directly from environmental samples.

Genome-resolved metagenomics
The assembly of complete or draft genomes exclusively from metagenomic sequencing data.

DPANN
A proposed archaeal superphylum comprising Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanohaloarchaeota and Nanoarchaeota. More recently, it was suggested that additional candidate phyla such as Woesearchaeota, Pacearchaeota, Micrarchaeota and possibly Altiarchaeales are part of this group.

Hydrothermal vents
Areas of the sea floor from which geothermally heated water issues.

the subject of intense debate[8–14]. However, during the past two decades, advances in DNA sequencing technologies and cultivation-independent genomic approaches have provided genomic data from new archaeal lineages that are distantly related to the cultivated Crenarchaeota and Euryarchaeota. Their discovery has substantially extended our knowledge of the diversity, evolution, metabolic capabilities and ecological impact of archaea. These genomic approaches first allowed the genome sequencing of representatives of several phylum-level archaeal lineages: the proposed phylum Korarchaeota[15–18] and, more recently, the ammonia-oxidizing Thaumarchaeota[19–23] and their sister phylum, the proposed phylum Aigarchaeota[24]. Together, Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota (TACK) were found to form a monophyletic group referred to as the TACK superphylum[25,26] (or the Proteoarchaeota[27]; FIG. 2). The use of phylogenomic approaches combined with improved models of sequence evolution and broader taxon sampling (BOX 1) yielded support for the two-domains tree of life[25,28–34], in which eukaryotes initially appeared to branch from within the Archaea, specifically within or as sister group to the TACK superphylum (FIG. 1c).

During recent years, newly developed metagenomics approaches, such as genome-resolved metagenomics, have revealed additional phyla belonging to the TACK superphylum, including the Bathyarchaeota[35–37], the Geoarchaeota[38,39] and the Verstraetearchaeota[40] (FIG. 2). Other studies have revealed a variety of archaeal lineages whose members often have small cells and small genomes, and have been proposed to form a deep-branching superphylum called DPANN (Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota)[41–44]. In addition to these proposed new archaeal phyla, several euryarchaeal lineages have also been identified[45,46].

Another newly discovered phylum-level lineage is the Lokiarchaeota[47]. These organisms were identified from marine sediments that were sampled near Loki's Castle (a field of five hydrothermal vents that are located in the middle of the Atlantic Ocean between Greenland and Norway)[48]. Initial sequencing surveys of these samples revealed that ~10% of the 16S ribosomal RNA (rRNA) genes belonged to the Deep Sea Archaeal Group (or Marine Benthic Group B). This clade had been previously hypothesized to represent an unexplored lineage

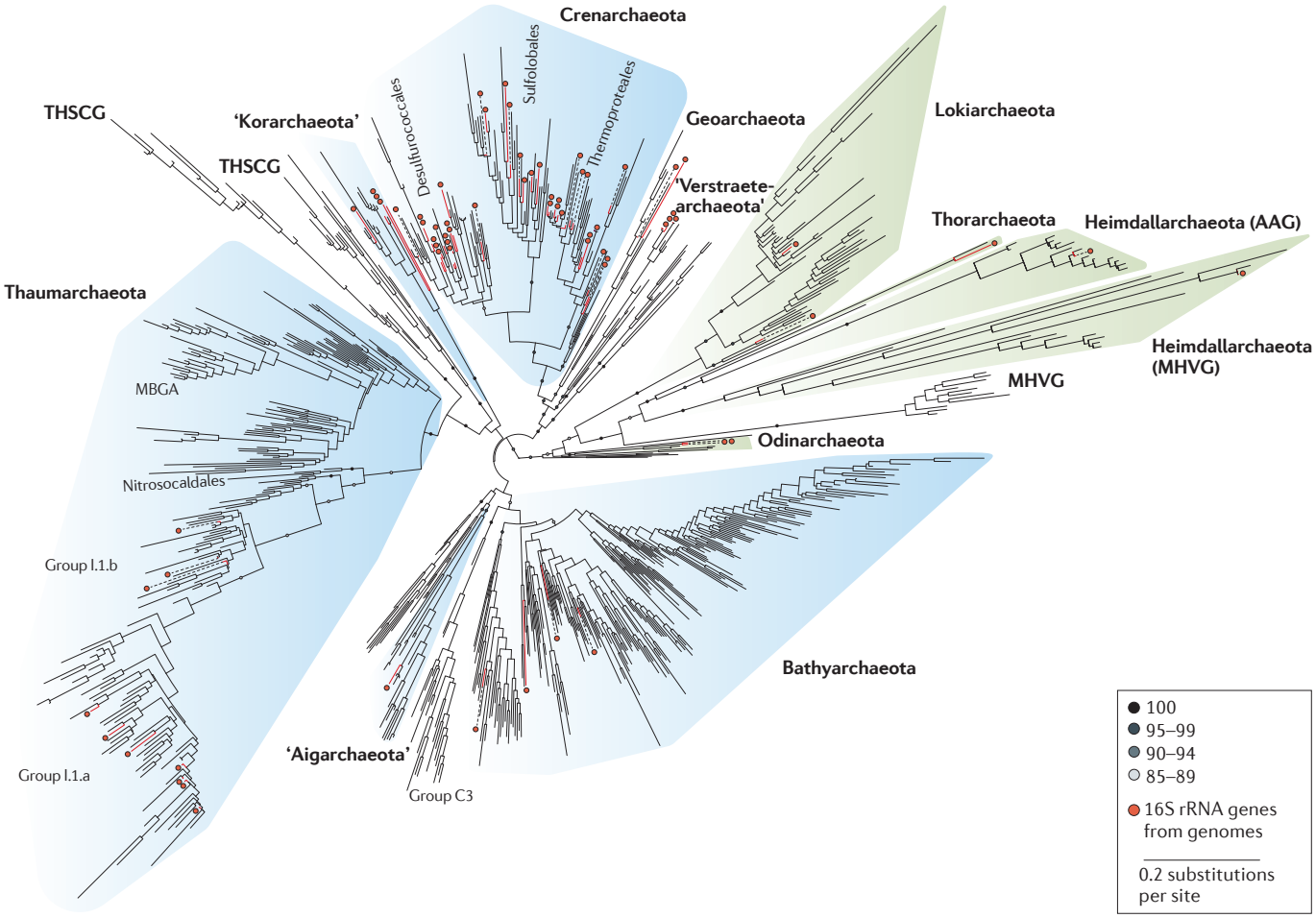

Figure 2 | **Genomic exploration of TACK and Asgard archaeal diversity.** The tree represents our current understanding of the diversity of Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota (TACK) archaea (in blue) and Asgard archaea (in green) based on 16S ribosomal RNA (rRNA) gene sequencing surveys. Clades for which genomic data are available are indicated by red branches and dots. All other lineages are known from 16S rRNA gene sequences only. The phylogeny was reconstructed from 16S rRNA genes of TACK and Asgard archaea (1,471 positions) in a maximum-likelihood framework by use of IQ-TREE[157] and the GTR + F0 + G model of sequence evolution. All 16S rRNA sequences >1,000 bp that were assigned to TACK and Asgard lineages were retrieved from the Silva rRNA database and clustered at 95% identity with the Cluster Database at High Identity with Tolerance (cd-hit) software[158]. Ultrafast bootstrap statistical support values for branches corresponding to major clades are indicated by circles in different shades of grey according to the figure label. The scale bar indicates the number of substitutions per site. Names of proposed phyla are in quotation marks. AAG, Ancient Archaeal Group; MBGA, Marine Benthic Group A; MHVG, Marine Hydrothermal Vent Group; THSCG, Terrestrial Hot Spring Crenarchaeotic Group. Modified with permission from REF. 45, Science/AAAS.

of the TACK archaea[25], making it a lineage of potential importance regarding the origin of eukaryotes. After further metagenomic analyses, a near-complete genome was reconstructed for *Lokiarchaeum* sp. GC14_75, and partial genomes were reconstructed for two low-abundance and distantly related organisms (Loki2 and Loki3).

The discovery of the Lokiarchaeota has provided additional evidence for the two-domains tree of life because at the time of discovery, it was shown to contain the closest relatives of eukaryotes in phylogenomic analyses. Moreover, members of Lokiarchaeota were found to carry genes encoding proteins that were assumed to be specific to eukaryotes[47] (often called eukaryotic signature proteins, or ESPs[49]). A recent study has challenged this view of the tree of life, alternatively suggesting that Lokiarchaeota represent a deep-branching Euryarchaeota-related lineage[50], although this work was itself questioned[51]. However, in support of the two-domains tree of life scenario, another recent study identified eight additional representatives from three novel phyla that are related to Lokiarchaeota (that is, the Odinarchaeota, Thorarchaeota and Heimdallarchaeota)[52]. Collectively, these new archaeal lineages represent the Asgard superphylum, and phylogenomic analyses have confirmed that they are closely related to eukaryotes[47,52] (FIG. 1d). Furthermore, in-depth analyses of the Asgard archaea genomes confirmed the presence of ESPs, which belong to protein families that represent building blocks that are fundamental for eukaryotic cellular complexity.

Eukaryotic signature proteins
(ESPs). Proteins involved in key eukaryotic processes and conserved across most eukaryotic diversity.

Box 1 | **The difficulty of inferring ancient evolutionary relationships**

The past two decades have seen substantial changes in molecular phylogenetic approaches. Of importance has been the transition from analysing individual genes to reconstructing phylogenies that are based on 'supermatrices' of concatenated gene or protein sequences, up to several thousand. However, when trying to disentangle ancient evolutionary events, such as the relationships between the three domains of life, one is confronted with several persistent problems. First, the phylogenetic signal reflecting deep relationships is usually weak because it erodes over time: the more time has passed since the event of interest, the more nucleotide substitutions can occur multiple times at a given site in the sequence. Second, tree reconstruction artefacts (for example, due to functional shifts, compositional bias and evolutionary rate variation across sites and branches[139]) are usually greater at this evolutionary timescale. Although the former can, in theory, be overcome through the analysis of a large number of markers, in reality, studies aiming to resolve the relationships between Archaea, Bacteria and Eukarya have been constrained to using a fairly small set of genes that are conserved in the three domains of life. Several enhancements in phylogenomic analyses have helped to overcome this 'phylogenomic impasse' (REF. 9). For example, phylogenetic models, implemented in either maximum-likelihood or Bayesian frameworks, have been improved to better fit the evolution of molecular sequence data. These improvements include taking into account the heterogeneity of composition across sites[140] or across time[34,141]. Another conceptual change proposed to resolve the tree of life has been to analyse not only universal markers (such as ribosomal proteins) but also different sets of carefully selected non-universal gene sets in order to resolve specific parts of the tree (for example, the position of the root of archaea)[32,142]. The recent increase in the characterization of novel microbial genomes[43,44,47,52,143–145] has led to major breakthroughs in resolving deep splits in the tree of life. Improved taxonomic sampling minimizes phylogenetic reconstruction artefacts through a better estimation of the substitution process, which is crucial for resolving difficult phylogenetic questions[139].

Although the discovery of these novel archaeal lineages has provided important insights into the origin of eukaryotes, it has also raised new questions. In particular, the exact position of eukaryotes in the tree of life with respect to the Asgard archaea remains unresolved to date. To identify the precise position of eukaryotes within the tree of life, obtaining genomic data from broader taxonomic archaeal lineages is crucial. Excitingly, such efforts might be achievable, because phylogenetic analyses of environmental 16S rRNA gene sequences indicate that the currently available genomic data for Asgard archaea represent only a small fraction of the existing diversity (FIG. 2).

The archaeal ancestry of eukaryotes

Eukaryogenesis: the matter of definitions. An evolutionary link between Asgard archaea and eukaryotes has numerous implications for our understanding of the process of eukaryogenesis. However, the literature on this subject can sometimes be confusing owing to the use of terms that are not always explicitly defined and may have different meanings. The term, eukaryogenesis, is generally used to refer broadly to the evolutionary events that occurred during the emergence of the eukaryotic cell from its prokaryotic ancestor(s). In this Review, we refer to eukaryogenesis as the whole sequence of evolutionary events that occurred between the existence of the first eukaryotic common ancestor (FECA) and the existence of the last eukaryotic common ancestor (LECA). Both of these organisms, or perhaps populations, are defined solely by phylogenetic criteria. LECA is the last common ancestor of all extant eukaryotes, that is, the youngest ancestor of all living eukaryotes (FIG. 3). It is important to note that even though all eukaryotes living today descend from LECA, not all descendants of LECA will have survived to date (FIG. 3). In fact, it is well recognized that several mass extinctions have affected eukaryotic biodiversity long after the existence of LECA[53]. FECA is defined as the oldest ancestor whose only living descendants are eukaryotes. Similarly, it is possible (and most likely) that many other now-extinct lineages descended from FECA, which did not have any eukaryotic features (FIG. 3). FECA thus represents the first generation after the divergence from its closest prokaryotic relatives that do not have any descendants that exist today other than eukaryotes. As these organisms are phylogenetic entities, we do not need to know what kind of cellular, genetic and metabolic features they had in order to be able to identify their precise position in the tree of life. However, it is important to keep in mind that LECA and FECA are moving targets; as we discover a broader diversity of organisms, both the branching points (nodes) in the tree of life corresponding to LECA and FECA and our inferences about their biological properties will change (FIG. 3). The discovery of the Asgard archaea is an example of this. Before their discovery, some evolutionary scenarios suggested that eukaryotes were a sister group to the TACK archaea, and FECA was virtually indistinguishable from the last common ancestor of the TACK archaea and eukaryotes (FIG. 3). Now that the Asgard archaea are known to be the closest archaeal relatives of eukaryotes that have been identified to date, FECA is inferred to be essentially identical to the last common ancestor of Asgard archaea and eukaryotes. This suggests that the evolutionary gap that exists between FECA and LECA will decrease only if we discover closer archaeal relatives of eukaryotes and/or 'deeper-branching' eukaryotic lineages (FIG. 3). Consequently, the series of evolutionary events that are thought to have occurred during eukaryogenesis (that is, in the time between FECA and LECA) will also have to be revised accordingly. It is important to keep in mind that as we discover new lineages that branch between the current FECA and LECA, and whose phenotypes may not resemble model archaea and eukaryotes, decisions will have to be made regarding the classification of new organisms to determine whether they will be considered archaea or eukaryotes (FIG. 3).

In this framework, investigating eukaryogenesis involves answering the following questions: which features were present in LECA? When did these features emerge? How many of these features originated autogenously between FECA and LECA and how many were inherited from their prokaryotic ancestor(s)? Which prokaryotic partners participated in eukaryogenesis? What does this allow us to infer about FECA? In the following sections, we address these questions in detail.

What do we currently know about LECA? Phylogenomic and comparative genomic analyses have led to the hypothesis that LECA, estimated to have lived

First eukaryotic common ancestor
(FECA). The most ancient organism whose only living descendants are present-day eukaryotes.

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

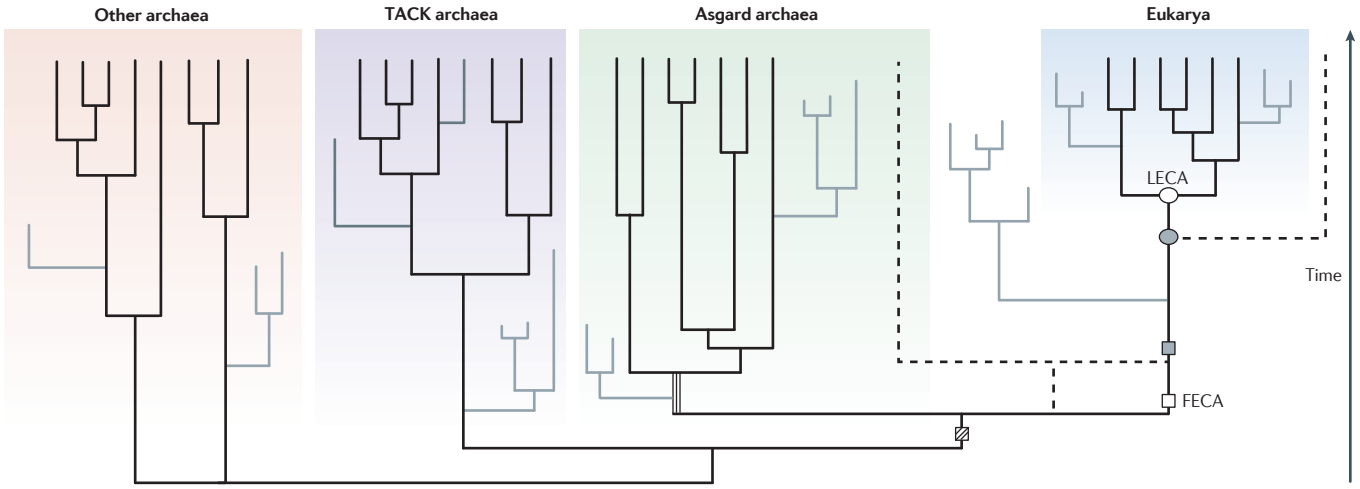

Figure 3 | **Key time points in eukaryogenesis.** Several evolutionary time points relevant to the discussions about eukaryogenesis are represented on a schematic unrooted tree of archaea and eukaryotes. The white oval and square represent the last eukaryotic common ancestor (LECA) and the first eukaryotic common ancestor (FECA), respectively. The grey oval highlights that LECA would represent a more ancient organism if an early diverging eukaryotic lineage was discovered (right-hand dashed line), whereas the grey square shows that FECA can be pushed forward in time only if we discover an archaeal lineage that is closer to eukaryotes (left-hand dashed line). The hatched square signifies where FECA was placed before the discovery of Asgard archaea. Grey branches depict extinct lineages. The triple line at the base of Asgard archaea illustrates that their monophyly is still unclear and that eukaryotes might branch within them. Note that lineages diverging between FECA and LECA (extinct or extant) will arbitrarily be considered an archaeon or a deeply diverging eukaryote on the basis of phenotypic criteria that remain to be universally accepted; this is illustrated by the lineages without a coloured background.

~1–1.9 billion years ago[54], already was a fully fledged eukaryote and possessed a large number of features that are uniquely found in modern eukaryotes[55,56]. This organism possessed the eponymous feature of the eukaryotes: the nucleus, including nuclear pores and associated complexes, and nuclear lamina[55,57]. This nucleus is believed to have contained linear chromosomes with telomeres, encoding ~4,000 genes[58,59] containing spliceosomal introns[60]. LECA is thought to have possessed complex gene expression regulatory mechanisms, including an RNAi system[59] and small non-coding RNAs[61], and chromatin that existed in different states dependent on histone packaging[55]. Transcription was uncoupled from translation and involved extensive RNA processing (including intron splicing, capping and polyadenylation)[62]. This ancestor also had an elaborate protein regulation and recycling system composed of a proteasome and a ubiquitin signalling system[63].

The cellular environment of LECA was highly compartmentalized through the presence of a sophisticated endomembrane system[55,64,65] composed of the endoplasmic reticulum, the Golgi apparatus, endosomes, lysosomes and peroxisomes; this organism possessed exocytic and endocytic pathways (including phagocytosis). A modern actin-based and tubulin-based cytoskeleton and associated molecular motor proteins[55] enabled intracellular trafficking, cell motility and a complex cell cycle[66,67], including meiosis[68]. Additionally, LECA was likely able to synthesize phospholipids composed of glycerol 3-phosphate and fatty acids[69], as well as the sterols[70] and sphingolipids[71] characteristic of present-day eukaryotes (BOX 2). LECA was also the host of aerobic, or facultatively aerobic, mitochondria[72] that descended from a once free-living alphaproteobacterium[73–75]. It is clear that these organelles had a respiratory electron transport chain that generated ATP in the presence of oxygen. However, the presence or absence of anaerobic metabolism in LECA is still intensely debated[76,77].

Despite the consensus in the field on the presence of these features in LECA, it is important to keep in mind that many of these inferences are based solely on the presence of (some of) these components in diverse eukaryotic lineages and the assumption that they were vertically inherited. It is worth noting that several of these features are absent in various extant, and often distantly related, eukaryotic lineages. This implies that present-day eukaryotes share very few universal characteristics but instead contain a collection of defining features that are not, or are rarely, found in prokaryotes[78,79]; each eukaryotic organism harbours a large subset of typically eukaryotic features but rarely all of these features. By inferring the existence of the entirety of these sometimes sparsely distributed characteristics in LECA, we are contemplating an ancestor that appears to be substantially more complex than most of its descendants; this implies that there were important changes in evolutionary trends before and after LECA (that is, complexification and simplification, respectively)[78].

In light of this, several additional factors need to be taken into consideration. First, and as previously suggested[78], the role of horizontal gene transfer (HGT) in the evolution of eukaryotes is likely to be more influential than what is currently assumed, especially HGT between unicellular eukaryotes[80–83]. HGT allows several distant

Telomeres
Repetitive nucleotide sequences located at the ends of the linear chromosomes of most eukaryotic organisms.

Spliceosomal introns
Introns in the nuclear protein-coding genes of eukaryotes that are removed by spliceosomes.

Proteasome
A large protein complex responsible for regulated degradation of proteins as part of the ubiquitin system found in all eukaryotes.

Horizontal gene transfer
(HGT). Exchange of genetic material between cells and/or organisms; sometimes called lateral gene transfer. This contrasts with the vertical inheritance of DNA from parent to offspring.

Box 2 | **The membrane anomaly**

All cells are bound by lipid membranes, of which phospholipids are the main components. However, archaeal phospholipids are made up of isoprenoids and *sn*-glycerol-1-phosphate (G1P) instead of the conventional fatty acid-based and *sn*-glycerol-3-phosphate (G3P)-based phospholipids that are found in bacteria and eukaryotes[69]. Although this apparent lipid divide had a historical role in the acknowledgement of Archaea as a separate domain of life[1,146], it has now become one of the main challenges when considering different eukaryogenesis scenarios: if the eukaryotic host lineage originated from within archaea, how did the bacterial-like eukaryotic phospholipids evolve[147]? This is not just a matter of lipid chemistry; it also has important implications in the evolution of many essential membrane proteins that had to adapt to a completely new membrane environment throughout the lipid transition[106].

Some eukaryogenesis scenarios suggest that the phospholipid transfer from the bacterial ancestor of mitochondria to the host facilitated the lipid shift, either before or after the endosymbiotic event[148,149]. When the first Lokiarchaeum metagenome became available, it was shown to lack the homologues of both the G1P and G3P dehydrogenases while containing putative mechanisms to metabolize G3P, isoprenoids and fatty acids. Therefore, lokiarchaeal phospholipids were initially suggested to have a mixed intermediate composition[150]. Although appealing, this hypothesis is weakened by the discovery of genes that encode G1P dehydrogenase homologues in genomes of other Asgard archaea.

However, we are still limited by the scarcity of information on lipid biosynthesis pathways in archaea[151]. For instance, several archaea have been postulated to have the capacity to synthesize fatty acids[152,153] and G3P as components of phospholipids[150,154], but this has never been confirmed. Further experimental efforts to test these hypotheses, combined with further comparative genomic analyses as more archaeal genomes become available, will undoubtedly provide us with new opportunities to tackle these questions.

descendant lineages to exhibit similar features that were not necessarily present in their last common ancestor; ignoring this mechanism leads to an inevitable overestimation of the gene content in LECA. Yet, although it has certainly led to the acquisition of a number of individual enzymes and their accessory proteins, it is less likely that HGT has heavily influenced the evolutionary history of complex molecular systems[84]. Second, although unicellular organisms represent most of the eukaryotic diversity, genome sequence data from those major microbial eukaryotic groups remain sparse[85]. Where these data do exist, they are heavily biased towards parasitic organisms with reduced genomes. This can lead to the misconception that genomic streamlining is the dominant mode of evolution in eukaryotes[86]. Finally, as lineages diverge and adapt to different niches, ancestral genes that are no longer essential for viability of the organism will be lost, whereas other genes that were not present in their ancestor(s) will be gained. When accounting for these different evolutionary mechanisms during the analysis of genomic data, one does not need to invoke a dramatic change in gene number or mode of evolution before and after LECA.

Which features of LECA were inherited from prokaryotic ancestors? Establishing which features of LECA were inherited from prokaryotes and the identity of these prokaryotic ancestors is crucial to understanding eukaryogenesis. Several phylogeny-based analyses have tried to define the subset of LECA genes with a prokaryotic origin[87–90]. Although these studies vary widely in the estimated number of LECA gene families of prokaryotic descent (~550–1,100), they still suggest that only a minority of the genes that were present in LECA (~4,000)[58,59] were inherited from prokaryotic ancestors. Among these, the main phylogenetic signal can be traced back more often to bacteria (56–71%)[88,90] than to archaea (18–37%)[87,90]. Genes with a bacterial ancestry are overwhelmingly linked to metabolic processes, whereas archaeal genes tend to be involved in information processing[87–90]. However, tracing the identity of the bacterial or archaeal donor lineages has been more challenging. Consistent with the alphaproteobacterial origin of mitochondria[73–75], the most common bacterial sister lineage to eukaryotic genes in phylogenies is that of alphaproteobacteria. Nonetheless, this represents only a minority of the total number of eukaryotic genes of prokaryotic ancestry (5.6–8.5%)[88,89]. By contrast, phylogenetic studies of other gene families have revealed a sister relationship between various specific bacterial lineages and eukaryotes. There are several explanations for this. First, single-gene phylogenies can display artefactual topologies in phylogenetic trees owing to several causes, most notably the limited evolutionary signal that can be extracted from their characteristically short sequences (BOX 1). In fact, when branch statistical support for any given tree topology is taken into account[88], only a few gene phylogenies strongly support a sister relationship between eukaryotes and non-alphaproteobacterial bacteria. Second, it has been argued that LECA genes of bacterial origin were acquired with the mitochondrial ancestor and that the phylogenetic signal linking many of these genes to various non-alphaproteobacterial ancestors was the result of continuous HGT between prokaryotic lineages[89] (FIG. 4). By contrast, another study suggests that genes of origins other than alphaproteobacteria were acquired before the mitochondrial endosymbiosis, reflecting either several symbiotic associations or gradual waves of HGT into the host genome[90] (FIG. 4). Similarly, in all these studies[87–90], genes of archaeal descent in eukaryotes grouped with Euryarchaeota as often as they did with TACK archaea lineages[91]. Although this may be because of the low resolution of relationships between sequences that is typically found in single-gene trees (BOX 1), it is also likely that the absence of genomic data from Asgard archaea in these analyses prevented the detection of a clear signal from the host lineage. Likewise, given the past and current debates over the identity of the closest alphaproteobacterial relative of mitochondria[92–101], it is conceivable that this (possibly deep-branching) lineage has yet to be discovered and that its absence from current phylogenetic reconstructions prevents the precise determination of genes of mitochondrial origin.

Furthermore, it should also be considered that some LECA genes of putative bacterial origin were already present in the archaeal ancestor, following ancient HGTs from bacteria. Supporting this idea, genomes of Asgard archaea appear to contain a large number of genes of bacterial origin[47,52] (FIG. 4). Finally, some of these bacterial genes that are sparsely distributed across the tree of eukaryotes might have been absent in LECA. Instead, they may have been acquired more recently by HGT

Genomic streamlining
A form of genome evolution that occurs through size reduction and simplification in terms of gene content; particularly common among parasitic organisms.

Phylogenetic signal
Information contained in homologous molecular sequences used to reconstruct the historical relationships between the sequences.

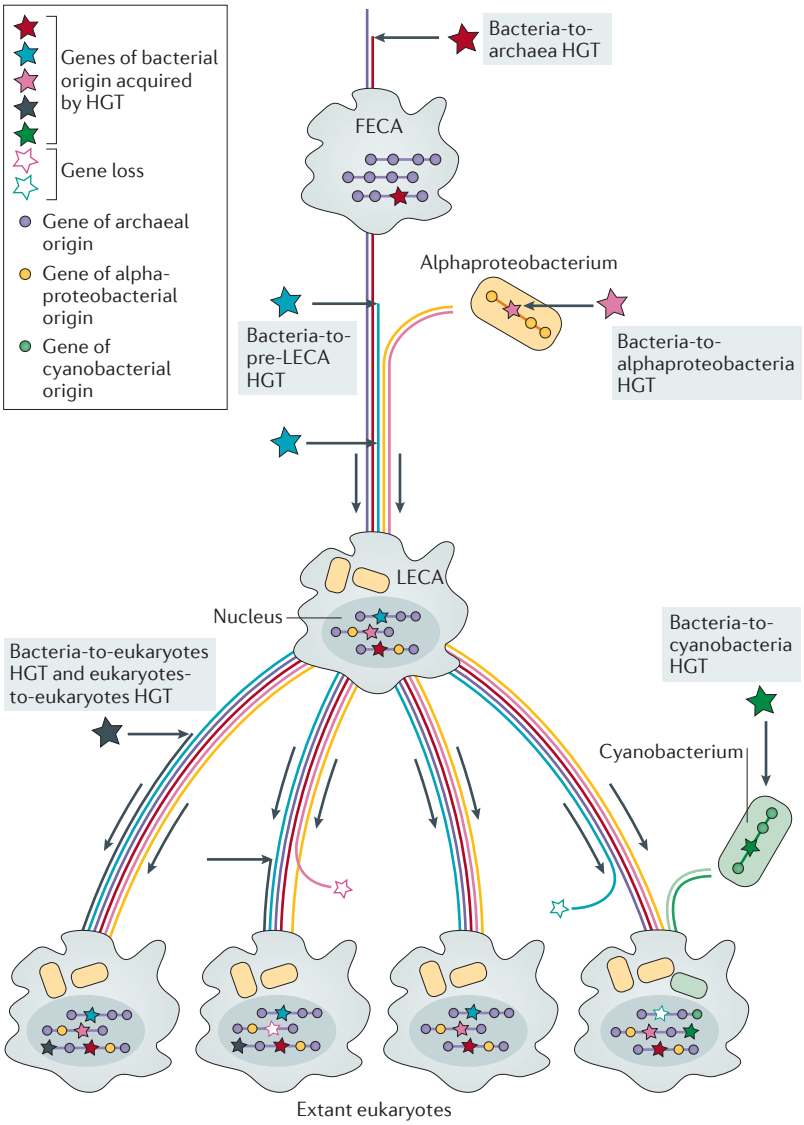

Figure 4 | **Evolutionary scenarios for the origins of bacterial genes in eukaryotes.** Within eukaryotic genomes, most genes of bacterial origin cannot be traced back to alphaproteobacterial (mitochondrial, yellow) or cyanobacterial (plastid, green) ancestors. These genes could derive from horizontal gene transfer (HGT) into the alphaproteobacterial or cyanobacterial ancestors of organelles (pink and dark green stars, respectively) before their endosymbiosis. Other bacterial genes could have been transferred into the eukaryotic lineage before or after the mitochondrial endosymbiosis (blue stars) or after the diversification of eukaryotes (black stars) and propagated among eukaryotes by eukaryote-to-eukaryote HGT. Some bacterial genes were probably acquired before the first eukaryotic common ancestor (FECA) and were already present in the archaeal ancestor of eukaryotes (dark red stars). Outgoing lines indicate lineage-specific gene losses after the existence of the last eukaryotic common ancestor (LECA). These have likely affected genes of all origins; for clarity, only examples affecting genes of bacterial origin are shown. These various evolutionary events are not mutually exclusive, and a debate lies in their relative importance (see main text). As most genes of endosymbiotic origins have been transferred to the nucleus before LECA, they are depicted in the nucleus only; organelles are shown without DNA for simplicity. Circles represent genes of archaeal, alphaproteobacterial and cyanobacterial origins. Filled stars represent laterally acquired genes. Empty stars represent losses of laterally acquired genes.

from a bacterial donor to a eukaryotic lineage and subsequently transferred between eukaryotic organisms. For example, this has been one of the proposed evolutionary scenarios regarding the origin of genes that are involved in anaerobic energy metabolism, which are found in many eukaryotic microorganisms that experience hypoxia[76] (FIG. 4).

Regardless of the relative contribution of each of these evolutionary scenarios to the observed phylogenetic patterns, it is important to distinguish cellular ancestors from genetic ancestors. Whereas the latter can describe any prokaryotic lineage that contributed to some of the genetic material of LECA by HGT, the former refers to a prokaryotic lineage whose only existing descendants are eukaryotes. Given current evidence, there are only two presumed cellular ancestors of eukaryotes: the archaeal lineage whose descendant was the putative host of the mitochondrial endosymbiosis and the alphaproteobacterial ancestor of mitochondria. Traditionally, the evolutionary events that occurred during eukaryogenesis are mapped on a line running from the archaeal host to LECA. Although this may appear to be arbitrary, it reflects the idea that many eukaryotic features could have developed in the host lineage before the acquisition of the mitochondrial endosymbiont (see below).

The relative timing of the emergence of LECA features. Many hypotheses have been proposed to explain the origin of eukaryotes[72,102–105]. With the discovery of Asgard archaea and the additional evidence that the eukaryotic lineage emerged from within Archaea[47,52], a number of these models can now be excluded[106,107]. Nonetheless, accepting that Bacteria and Archaea were the two primary domains of life does not solve the debate over the relative timing of emergence of the features that were present in LECA. In particular, controversy persists in the field regarding the timing and the influence of mitochondrial endosymbiosis, that is, about how much time passed and which evolutionary innovations occurred between the divergence from archaea and the mitochondrial acquisition. Autogenous or 'mito-late' models[108] assume that endosymbiosis occurred in a proto-eukaryotic host, that is, an organism that already possessed many eukaryotic cellular features (for example, the nucleus, a dynamic cytoskeleton and a sophisticated endomembrane system); these features are often considered to be essential prerequisites in such models to explain the engulfment of the alphaproteobacterial ancestor of mitochondria through phagocytosis. By contrast, other hypotheses[109,110], generally referred to as symbiogenic or 'mito-early' models, postulate that mitochondrial endosymbiosis took place very early during eukaryogenesis and has been the driving force that has led to the evolutionary innovations between FECA and LECA. The absence of amitochondriate organisms among living eukaryotes (with one currently known exception, resulting from a secondary loss[111]) is often used as an argument against mito-late hypotheses. It has been argued that this is evidence for mitochondrial endosymbiosis being the first and the most important event that drove eukaryogenesis. However, this hypothesis fails to account for the fact that no living eukaryote has been observed to possess an evolutionary intermediate form of any of the multitude of complex

features that emerged between FECA and LECA. It is difficult to conceive that the entirety of eukaryotic cellular complexification and the origination of thousands of genes took place so rapidly that no speciation events occurred between FECA and LECA (FIG. 3). A more likely explanation is that these evolutionary intermediates did exist but did not leave any living descendants. In addition, eukaryotic microorganisms have not been studied to the same extent as multicellular eukaryotes, and, although perhaps unlikely, we cannot exclude the possibility that deep-branching eukaryotes with intermediate forms of some cellular features do exist but have thus far escaped identification.

Owing to the absence of known intermediate lineages that lack some features of LECA, inferring the relative timing of emergence of these characteristics has proven difficult. A recent study aimed to estimate the relative age of the mitochondrial endosymbiosis event by utilizing phylogenetic distances between eukaryotic proteins and their prokaryotic homologues as a proxy for divergence times[90]. The authors observed shorter evolutionary distances between alphaproteobacteria-derived proteins and their bacterial counterparts than between LECA proteins of other bacterial and archaeal origins; this was interpreted as evidence for a moderately late acquisition of mitochondria by a host that already contained bacteria-derived and archaea-derived protein families. The results of this study have fuelled considerable debate[112,113]. Even if these conclusions are correct[114], we are left with few insights into the timing of the emergence of all the genes and associated features that have no recognizable prokaryotic homologues, which represent the vast majority of the gene content of LECA.

The nature of the last common ancestor of Asgard archaea and eukaryotes. Although determining the order of events in eukaryogenesis is difficult, defining the gene content of the last common ancestor of Asgard archaea and eukaryotes is more feasible. When attempting to reconstruct the gene content of this common ancestor, it is crucial to keep in mind that extant Asgard archaea diverged from this ancestor as long ago as extant eukaryotes did. All the features of the Asgard–eukaryotic common ancestor are therefore unlikely to have been conserved in all its living archaeal descendants[115]. In light of this, the patchy distribution of ESPs across different archaeal lineages (FIG. 5) is perhaps less surprising.

The first features that were traced back to the archaeal ancestor of eukaryotes included components of the DNA replication, transcription and translation machineries, as well as those of the proteasome, exosome and a ubiquitin modifier system[116]. Subsequent analyses of TACK and Asgard archaea genomes have revealed additional components of these systems. The presence of homologues of the eukaryotic ribosomal proteins L22e and L28e in the genomes of Asgard archaea[47,52] suggests that these archaea carry genes for the most eukaryote-like ribosomes that have been identified thus far. In addition, a homologue of a putative DNA polymerase-ε was identified in Heimdallarchaeota[52]. Interestingly, homologues of some eukaryotic RNA polymerase components are uniquely found in only some TACK and Asgard lineages[22,24,52,117]; they have likely been lost in other lineages after divergence from their last common ancestor (FIG. 5).

Although the proteasome has long been thought to have an archaeal origin[118], homologous components of the eukaryotic ubiquitin modifier system were discovered only recently in archaea. *Bona fide* ubiquitins and potential ubiquitin-interacting proteins such as E1 ubiquitin-activating enzymes were first identified in '*Candidatus* Caldiarchaeum subterraneum', a member of the proposed phylum Aigarchaeota, whose genome was found to encode a functional eukaryotic-like E1–ubiquitin-conjugating enzyme (E2)–(RING) ubiquitin ligase (E3) ubiquitylation cascade[24,119]. Ubiquitin-related modifier 1, which is involved in protein modification in eukaryotes, has been described in only a few Crenarchaeota[120], but preliminary analyses indicate that this protein is present in Bathyarchaeota and Asgard archaea as well (A.S. and T.J.G.E., unpublished observations). These observations are consistent with the discovery of all major components of the ubiquitin modifier system in most representatives of these lineages[47,52] (FIG. 5). In Asgard archaea, these proteins are encoded by conserved gene clusters, which in some organisms are adjacent to genes encoding components of the endosomal complex required for transport (ESCRT) machinery (see below). Altogether, these findings indicate that FECA already harboured a sophisticated ubiquitin modifier system, which may have had a function in protein targeting and degradation, perhaps by using a pathway mediated by an ESCRT-like machinery[121].

Actin and tubulin represent the basic units of microfilaments and microtubules, key components of the eukaryotic cytoskeleton, and they both belong to large protein families. Members of these superfamilies are present in almost all prokaryotic cells; however, the precise origin of the eukaryotic cytoskeleton (either from Bacteria or Archaea) could not be inferred until recently[122]. Additional members of these superfamilies, more closely related to their eukaryotic counterparts, have now been identified in archaea (FIG. 5). Distant homologues of actins (that is, crenactins) were shown to be present in Korarchaeota, Crenarchaeota and Aigarchaeota and were found to assemble into actin-like filaments[123–126]. Subsequently, analyses of the genomes of some members of the Thaumarchaeota revealed distant tubulin homologues (that is, ar-tubulins)[127]. Detailed analyses of Asgard archaea genomes helped to identify the origin of several eukaryotic cytoskeletal components[47,52]. All Asgard archaea possess various *bona fide* actin-related proteins, which are more closely related to eukaryotic actins than to archaeal homologues. Furthermore, these genomes encode homologues of eukaryotic profilin and gelsolin domain proteins, which are key regulators of actin filament dynamics. Finally, Odinarchaeota carry genes for tubulin homologues that are considerably more similar to eukaryotic tubulins than previously described archaeal

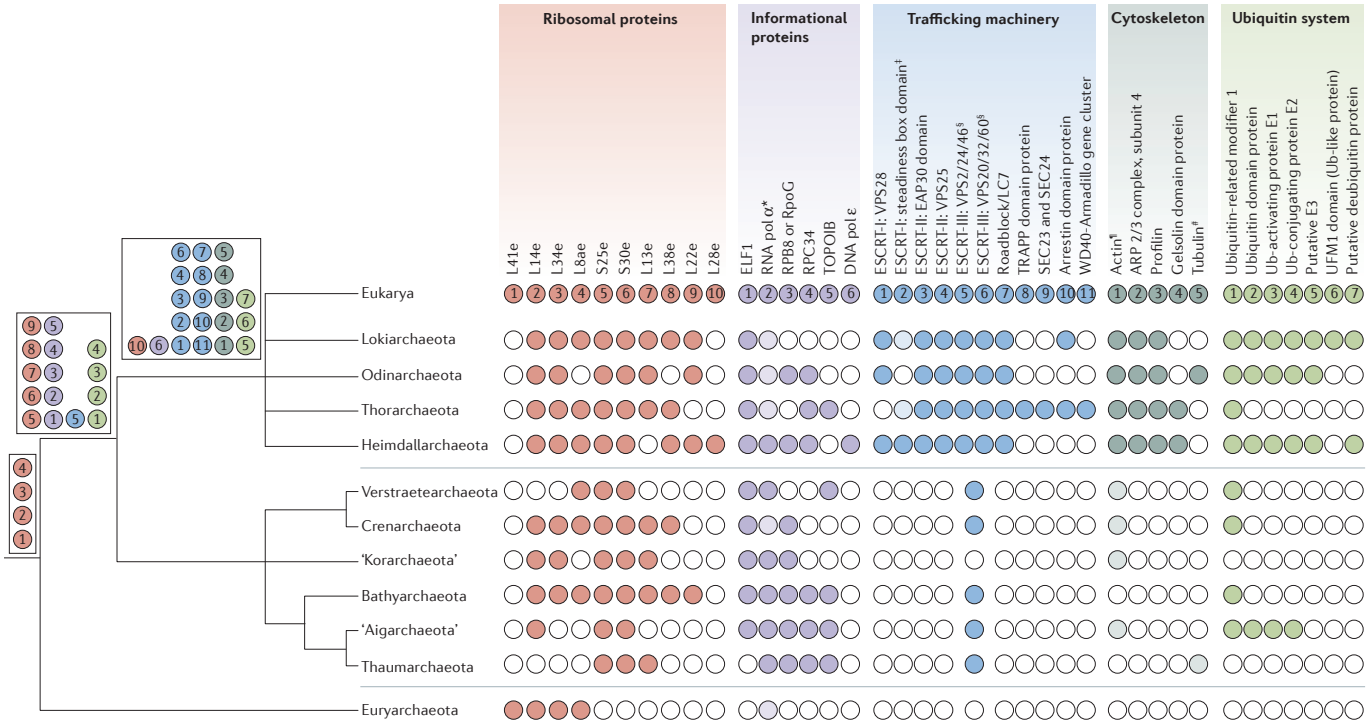

Figure 5 | **Origins of eukaryotic signature proteins present in FECA.** This figure indicates the presence of homologues of eukaryotic signature proteins (ESPs) in various archaeal lineages (filled circles), as well as their inferred time of emergence along the schematic tree of Archaea. The origin of each ESP is indicated on a schematic tree of life (left-hand side). Owing to the involvement of ESPs in often complex structures, and sometimes even in informational processes, our inferences assume that they were vertically inherited[84]. This would be in line with recent analyses suggesting that during archaeal evolution, the majority of genetic transmission events appear to be vertical rather than horizontal[142]. Eukaryotes and Asgard archaea are shown as a multifurcation (the occurrence of several branches splitting in the tree from the same point) because their relationships are not yet clearly resolved. Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota (DPANN) archaea are not represented because of the current uncertainty regarding their phylogenetic placement. Names of proposed phyla are in quotation marks. *Light shading indicates the presence of RNA polymerase A homologues encoded by two separate genes, whereas dark shading indicates single subunit homologues, similar to what is found in eukaryotes. ‡Putative homologues are lightly shaded. §Vps2/24/46 and Vps20/32/60 represent paralogous protein families belonging to the endosomal complex required for transport (ESCRT)-III. In contrast to other archaea, Asgard archaea genomes encode homologues of both families. ¶Light shades indicate distant actin homologues (that is, crenactin). #Light shades indicate distant tubulin homologues (that is, ar-tubulins). Ub, ubiquitin. Modified and expanded with permission from REFS 47,52, Macmillan Publishers Limited.

homologues[52]. Altogether, these results suggest that FECA already carried some of the key building blocks of the eukaryotic cytoskeleton.

Genomic analyses have also led to the suggestion of an archaeal origin of the eukaryotic trafficking machinery[128]. Initial insights came from studies of the crenarchaeote *Sulfolobus acidocaldarius*[129,130] that showed that a distant homologue of eukaryotic ESCRT-III subunit proteins (that is, an SNF7-domain protein) together with a Vps4-like ATPase were involved in cytokinesis in these archaea. Later, an ESCRT-based cell division system was also identified in Thaumarchaeota[131], Aigarchaeota and Bathyarchaeota[132] (FIG. 5). Nonetheless, the differences in complexity and components between the archaeal and eukaryotic systems were found to be considerable; therefore, the existence of components of the eukaryotic trafficking machinery in FECA remained unclear. However, this view changed with the discovery of Asgard archaea. Comparative analyses of Asgard archaea genomes revealed the presence of the major components of all three ESCRT complexes, ESCRT-I, ESCRT-II and ESCRT-III[47,52] (FIG. 5). The Asgard archaeal ESCRT components are encoded in gene clusters, which seem to be conserved in all known members[52]. Perhaps even more surprising was the discovery of an unexpectedly large set of small GTPases in all Asgard archaea in addition to the presence of proteins that contain the longin domain and the roadblock/LC7 domain, which could represent GTPase-interacting proteins[47,52,133]. This is remarkable because some of these elements are important regulators of the active eukaryotic transport machineries and are rarely or never found in other prokaryotic genomes. The genomes of Thorarchaeota were also found to encode homologues of eukaryotic proteins that are involved in vesicle budding and trafficking, such as transport

protein particle complex proteins[52], which are involved in the tethering of vesicles to target membranes[134,135]. In addition, these genomes encode homologues of the Sec23 and Sec24 subunits of the coat protein complex II (COPII), which mediates anterograde transport from the endoplasmic reticulum to the Golgi apparatus[135]. Surprisingly, all Thorarchaeota genomes that have been analysed to date possess neighbouring genes encoding proteins with predicted β-propeller or α-solenoid secondary structures. In eukaryotes, these domains are fused and are typically found in many vesicle coat complexes. This suggests that a protocoatomer gene repertoire was present in FECA[52].

Altogether, comparative genomics of archaea, and in particular of members of the TACK and Asgard superphyla, has furthered our understanding of the origin and evolution of fundamental ESPs and provided crucial insights into the nature of FECA. The discovery of homologues of ESPs in these archaea suggests that the genetic basis for some aspects of eukaryotic cellular complexity emerged early on, before mitochondrial endosymbiosis. Identifying the likely origin of these genes during archaeal evolution (FIG. 5) could help to determine the relative timing of the emergence of specific cellular systems that are found in eukaryotes, before the origin of the eukaryotic lineage.

However, functional studies on these components in archaea are much needed, not only for better understanding the biology of these organisms but also for elucidating their potential biological roles in FECA. The latter will be a challenging task because these functions are likely to have changed over ~2 billion years of evolutionary time, both in archaeal and in eukaryotic lineages.

Future reconstructions of the entire FECA gene repertoire will entail exploring sets of genes with different evolutionary fates. Indeed, this ancestor contained genes that were maintained both in extant Asgard archaea and eukaryotic genomes; the previously discussed ESPs are part of this set. However, FECA also contained genes that have been lost during eukaryogenesis and are now found only in archaea and genes that are present only in extant eukaryotes and have been lost during 'asgardogenesis'. Finally, the eukaryote–Asgard common ancestor must have had a number of genes that were lost in both lineages. Efforts have so far been focused on the investigation of genes that have been maintained in both lineages until today; in the long term, eukaryogenesis research will require an in-depth examination of the other subsets of genes. This will lead to a greater understanding of the nature of this ancestor and could improve our understanding of the origin of eukaryotes.

Conclusions and future perspectives

The discovery of archaea 40 years ago led to long-lasting debates on their evolutionary relationships with eukaryotes. Recent methodological improvements that have enabled the genomes of uncultivated archaea to be sequenced and explored have brought us much closer to understanding the origin and early evolution of eukaryotes.

Nevertheless, many questions regarding the process of eukaryogenesis remain unresolved. Even though many ESPs have recently been traced back to FECA, the evolutionary gap between this archaeal-like ancestor and any *bona fide* eukaryote is vast. The order of events and the evolutionary forces that led to the increase in cellular complexity, the emergence of at least ~3,000 genes and the loss of many typical archaeal features (for example, archaeal type lipids; see BOX 2) continue to be obscure. In addition, the origin of bacterial genes in eukaryotes and the role of HGT during the evolution of LECA from FECA remain unclear.

Future research priorities should focus on obtaining both genomic and functional data from the many underexplored lineages of deep-branching eukaryotic microorganisms and from Asgard archaea. Indeed, exploring archaeal diversity further might reveal even closer relatives of eukaryotes, uncovering a FECA that is younger than previously thought and with even more eukaryotic features. Similarly, it is possible that unknown deeper-branching 'eukaryotes' that display intermediate forms of certain eukaryotic cellular features exist; this would imply the existence of a simpler and older LECA. Overall, exploring the genomic and cellular diversity of archaea and eukaryotes will be crucial to revealing the exact breadth and nature of the evolutionary gap between prokaryotes and eukaryotes. Improved taxonomic sampling among archaea will also be essential to confidently resolve the phylogenetic placement of eukaryotes in the tree of life. Additionally, it will clarify the mechanisms and patterns underlying genome evolution in archaea and how these have influenced the evolutionary history of ESPs along the archaeal tree of life through gene loss and gain, and HGT.

Concomitantly, the cultivation of members of the Asgard archaea will be crucial for elucidating the cellular biology of these archaea and for functionally characterizing ESPs in these organisms. This, in turn, will improve our understanding of the cellular nature of FECA and the functional shifts that occurred during eukaryogenesis. However, this will certainly be challenging because most of these archaea have so far been found in anoxic environments that contain low levels of nutrients, suggesting that these organisms have long generation times. In addition, Asgard archaea are difficult to isolate and cultivate because they usually represent low-abundant members of the microbial communities in which they are found. Extending our understanding of the metabolism of the various Asgard archaea[136–138] will be an important research priority. The reconstruction and comparative analysis of the metabolic potential of members of these groups not only will be essential to guide strategies for cultivation but will also help us to understand the metabolism of the pre-mitochondrial ancestor of eukaryotes and to further formulate hypotheses about the nature of the syntrophic interactions between the mitochondrial endosymbiont and the host.

Evidently, exciting times are ahead of us, and many research avenues remain to be explored to lead to a more complete picture of the origin of eukaryotes and their early evolution from prokaryotic ancestors.

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 paper represents one of the most important studies in microbiology of the past century, providing the first evidence that cellular life was composed of three distinct types of organisms — later called Archaea, Bacteria and Eukarya.
2. 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).
3. 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).
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. Zillig, W. *et al.* The phylogenetic relations of DNA-dependent RNA polymerases of archaebacteria, eukaryotes, and eubacteria. *Can. J. Microbiol.* **35**, 73–80 (1989).
6. 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).
This study, based on analyses of ribosome structures, proposes that eukaryotes might have evolved from within the archaeal domain of life.
7. Rivera, M. C. & Lake, J. A. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* **257**, 74–76 (1992).
8. Brown, J. R. & Doolittle, W. F. Archaea and the prokaryote-to-eukaryote transition. *Microbiol. Mol. Biol. Rev.* **61**, 456–502 (1997).
9. Gribaldo, S., Poole, A. M., Daubin, V., Forterre, P. & Brochier-Armanet, C. The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? *Nat. Rev. Microbiol.* **8**, 743–752 (2010).
10. Lake, J. A. Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. *Nature* **331**, 184–186 (1988).
11. Tourasse, N. J. & Gouy, M. Accounting for evolutionary rate variation among sequence sites consistently changes universal phylogenies deduced from rRNA and protein-coding genes. *Mol. Phylogenet. Evol.* **13**, 159–168 (1999).
12. 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).
13. Gouy, M. & Li, W. H. Phylogenetic analysis based on rRNA sequences supports the archaebacterial rather than the eocyte tree. *Nature* **339**, 145–147 (1989).
14. Cammarano, P., Creti, R., Sanangelantoni, A. M. & Palm, P. The archaea monophyly issue: a phylogeny of translational elongation factor G(2) sequences inferred from an optimized selection of alignment positions. *J. Mol. Evol.* **49**, 524–537 (1999).
15. Barns, S. M., Fundyga, R. E., Jeffries, M. W. & Pace, N. R. Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc. Natl Acad. Sci. USA* **91**, 1609–1613 (1994).
16. Barns, S. M., Delwiche, C. F., Palmer, J. D. & Pace, N. R. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc. Natl Acad. Sci. USA* **93**, 9188–9193 (1996).
17. Elkins, J. G. *et al.* A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc. Natl Acad. Sci. USA* **105**, 8102–8107 (2008).
18. Reigstad, L. J., Jorgensen, S. L. & Schleper, C. Diversity and abundance of Korarchaeota in terrestrial hot springs of Iceland and Kamchatka. *ISME J.* **4**, 346–356 (2010).
19. Könneke, M. *et al.* Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**, 543–546 (2005).
20. Brochier-Armanet, C., Boussau, B., Gribaldo, S. & Forterre, P. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat. Rev. Microbiol.* **6**, 245–252 (2008).
This paper proposes the existence of the first archaeal phylum outside of the Crenarchaeota and Euryarchaeota, namely, the Thaumarchaeota.
21. Walker, C. B. *et al.* Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc. Natl Acad. Sci. USA* **107**, 8818–8823 (2010).
22. Spang, A. *et al.* Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. *Trends Microbiol.* **18**, 331–340 (2010).
23. Eme, L. *et al.* Metagenomics of Kamchatkan hot spring filaments reveal two new major (hyper)thermophilic lineages related to Thaumarchaeota. *Res. Microbiol.* **164**, 425–438 (2013).
24. Nunoura, T. *et al.* Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res.* **39**, 3204–3223 (2011).
This study represents the first genome reconstructed purely from metagenomic data and uncovered the existence of homologues of the eukaryotic ubiquitin system in archaea.
25. Guy, L. & Ettema, T. J. G. The archaeal "TACK" superphylum and the origin of eukaryotes. *Trends Microbiol.* **19**, 580–587 (2011).
This opinion piece is the first extensive review of the evidence for a scenario in which the archaeal parent of eukaryotes emerged from within the TACK superphylum.
26. 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).
27. Petitjean, C., Deschamps, P., López-García, P. & Moreira, D. Rooting the domain archaea by phylogenomic analysis supports the foundation of the new kingdom Proteoarchaeota. *Genome Biol. Evol.* **7**, 191–204 (2014).
28. 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. Biol. Sci.* **278**, 1009–1018 (2011).
29. 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).
30. Lasek-Nesselquist, E. & Gogarten, J. P. The effects of model choice and mitigating bias on the ribosomal tree of life. *Mol. Phylogenet. Evol.* **69**, 17–38 (2013).
31. Williams, T. A. & Embley, T. M. Archaeal "dark matter" and the origin of eukaryotes. *Genome Biol. Evol.* **6**, 474–481 (2014).
32. 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).
This paper proposes an innovative strategy to increase the number of phylogenetic markers usable to investigate the tree of life and suggests a new position for the root of the tree of the Archaea.
33. 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).
This study reports the first convincing evidence for a two-domains tree of life by use of phylogenomic approaches that employed advanced evolutionary models.
34. Foster, P. G., Cox, C. J. & Embley, T. M. The primary divisions of life: a phylogenomic approach employing composition-heterogeneous methods. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 2197–2207 (2009).
35. Evans, P. N. *et al.* Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* **350**, 434–438 (2015).
36. He, Y. *et al.* Genomic and enzymatic evidence for acetogenesis among multiple lineages of the archaeal phylum Bathyarchaeota widespread in marine sediments. *Nat. Microbiol.* **1**, 16035 (2016).
37. Lazar, C. S. *et al.* Genomic evidence for distinct carbon substrate preferences and ecological niches of Bathyarchaeota in estuarine sediments. *Environ. Microbiol.* **18**, 1200–1211 (2016).
38. Kozubal, M. A. *et al.* Geoarchaeota: a new candidate phylum in the Archaea from high-temperature acidic iron mats in Yellowstone National Park. *ISME J.* **7**, 622–634 (2013).
39. Guy, L., Spang, A., Saw, J. H. & Ettema, T. J. G. "Geoarchaeote NAG1" is a deeply rooting lineage of the archaeal order Thermoproteales rather than a new phylum. *ISME J.* **8**, 1353–1357 (2014).
40. Vanwonterghem, I. *et al.* Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nat. Microbiol.* **1**, 16170 (2016).
41. Baker, B. J. *et al.* Enigmatic, ultrasmall, uncultivated Archaea. *Proc. Natl Acad. Sci. USA* **107**, 8806–8811 (2010).
42. Rinke, C. *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).
43. Castelle, C. J. *et al.* Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* **25**, 690–701 (2015).
44. Eme, L. & Doolittle, W. F. Microbial diversity: a bonanza of phyla. *Curr. Biol.* **25**, R227–R230 (2015).
45. 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).
46. 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.* http://dx.doi.org/10.1038/ismej.2017.122 (2017).
47. Spang, A. *et al.* Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179 (2015).
This paper describes the discovery of Lokiarchaeota and provides evidence that they form a monophyletic group with eukaryotes and that their genomes encode an expanded repertoire of ESPs.
48. Pedersen, R. B. *et al.* Discovery of a black smoker vent field and vent fauna at the Arctic Mid-Ocean Ridge. *Nat. Commun.* **1**, 126 (2010).
49. Hartman, H. & Fedorov, A. The origin of the eukaryotic cell: a genomic investigation. *Proc. Natl Acad. Sci. USA* **99**, 1420–1425 (2002).
50. 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).
51. Spang, S. *et al.* Asgard archaea are the closest prokaryotic relatives of eukaryotes. *PLoS Genet.* (in the press) (2017).
52. Zaremba-Niedzwiedzka, K. *et al.* Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).
This work describes the Asgard superphylum and expands on the known repertoire of ESPs in archaea.
53. Knoll, A. H., Bambach, R. K., Canfield, D. E. & Grotzinger, J. P. Comparative Earth history and Late Permian mass extinction. *Science* **273**, 452–457 (1996).
54. 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).
55. Koumandou, V. L. *et al.* Molecular paleontology and complexity in the last eukaryotic common ancestor. *Crit. Rev. Biochem. Mol. Biol.* **48**, 373–396 (2013).
This publication represents a comprehensive review of the inferred features of LECA.
56. Koonin, E. V. The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biol.* **11**, 209 (2010).
57. Koreny, L. & Field, M. C. Ancient eukaryotic origin and evolutionary plasticity of nuclear lamina. *Genome Biol. Evol.* **8**, 2663–2671 (2016).
58. 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).
59. Koonin, E. V. Preview. The incredible expanding ancestor of eukaryotes. *Cell* **140**, 606–608 (2010).
60. Martin, W. & Koonin, E. V. Introns and the origin of nucleus-cytosol compartmentalization. *Nature* **440**, 41–45 (2006).
61. Shabalina, S. A. & Koonin, E. V. Origins and evolution of eukaryotic RNA interference. *Trends Ecol. Evol.* **23**, 578–587 (2008).
62. Collins, L. & Penny, D. Complex spliceosomal organization ancestral to extant eukaryotes. *Mol. Biol. Evol.* **22**, 1053–1066 (2005).
63. Grau-Bové, X., Sebé-Pedrós, A. & Ruiz-Trillo, I. The eukaryotic ancestor had a complex ubiquitin signaling system of archaeal origin. *Mol. Biol. Evol.* **32**, 726–739 (2015).

64. Field, M. C. & Dacks, J. B. First and last ancestors: reconstructing evolution of the endomembrane system with ESCRTs, vesicle coat proteins, and nuclear pore complexes. *Curr. Opin. Cell Biol.* **21**, 4–13 (2009).
65. Schlacht, A., Herman, E. K., Klute, M. J., Field, M. C. & Dacks, J. B. Missing pieces of an ancient puzzle: evolution of the eukaryotic membrane-trafficking system. *Cold Spring Harb. Perspect. Biol.* **6**, a016048 (2014).
66. Eme, L., Moreira, D., Talla, E. & Brochier-Armanet, C. A complex cell division machinery was present in the last common ancestor of eukaryotes. *PLoS ONE* **4**, e5021 (2009).
67. Eme, L., Trilles, A., Moreira, D. & Brochier-Armanet, C. The phylogenomic analysis of the anaphase promoting complex and its targets points to complex and modern-like control of the cell cycle in the last common ancestor of eukaryotes. *BMC Evol. Biol.* **11**, 265 (2011).
68. Speijer, D., Lukeš, J. & Eliáš, M. Sex is a ubiquitous, ancient, and inherent attribute of eukaryotic life. *Proc. Natl Acad. Sci. USA* **112**, 8827–8834 (2015).
69. Lykidis, A. Comparative genomics and evolution of eukaryotic phospholipid biosynthesis. *Prog. Lipid Res.* **46**, 171–199 (2007).
70. Desmond, E. & Gribaldo, S. Phylogenomics of sterol synthesis: insights into the origin, evolution, and diversity of a key eukaryotic feature. *Genome Biol. Evol.* **1**, 364–381 (2009).
71. Hannich, J. T., Umebayashi, K. & Riezman, H. Distribution and functions of sterols and sphingolipids. *Cold Spring Harb. Perspect. Biol.* **3**, a004762 (2011).
72. Embley, T. M. & Martin, W. Eukaryotic evolution, changes and challenges. *Nature* **440**, 623–630 (2006).
73. Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G. J. & Woese, C. R. Mitochondrial origins. *Proc. Natl Acad. Sci. USA* **82**, 4443–4447 (1985).
74. Gray, M. W. Organelle origins and ribosomal RNA. *Biochem. Cell Biol.* **66**, 325–348 (1988).
75. Gray, M. W. Mitochondrial evolution. *Cold Spring Harb. Perspect. Biol.* **4**, a011403 (2012).
76. Stairs, C. W., Leger, M. M. & Roger, A. J. Diversity and origins of anaerobic metabolism in mitochondria and related organelles. *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 20140326 (2015).
77. Müller, M. *et al.* Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol. Mol. Biol. Rev.* **76**, 444–495 (2012).
78. Doolittle, W. F. How natural a kind is "eukaryote?". *Cold Spring Harb. Perspect. Biol.* **6**, a015974 (2014).
79. Boyd, R. Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philos. Stud.* **61**, 127–148 (1991).
80. Andersson, J. O. Gene transfer and diversification of microbial eukaryotes. *Annu. Rev. Microbiol.* **63**, 177–193 (2009).
81. Soanes, D. & Richards, T. A. Horizontal gene transfer in eukaryotic plant pathogens. *Annu. Rev. Phytopathol.* **52**, 583–614 (2014).
82. Eme, L., Gentekaki, E., Curtis, B., Archibald, J. M. & Roger, A. J. Lateral gene transfer in the adaptation of the anaerobic parasite blastocystis to the gut. *Curr. Biol.* **27**, 807–820 (2017).
83. Alsmark, C. *et al.* Patterns of prokaryotic lateral gene transfers affecting parasitic microbial eukaryotes. *Genome Biol.* **14**, R19 (2013).
84. Jain, R., Rivera, M. C. & Lake, J. A. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl Acad. Sci. USA* **96**, 3801–3806 (1999).
85. Sibbald, S. J. & Archibald, J. M. More protist genomes needed. *Nat. Ecol. Evol.* **1**, 145 (2017).
86. Bapteste, E. & Gribaldo, S. The genome reduction hypothesis and the phylogeny of eukaryotes. *Trends Genet.* **19**, 696–700 (2003).
87. Thiergart, T., Landan, G., Schenk, M., Dagan, T. & Martin, W. F. An evolutionary network of genes present in the eukaryote common ancestor polls genomes on eukaryotic and mitochondrial origin. *Genome Biol. Evol.* **4**, 466–485 (2012).
88. 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).
This paper proposes a thorough systematic analysis of the phylogenetic relationships between ancestral eukaryotic genes and archaeal and bacterial genes.
89. Ku, C. *et al.* Endosymbiotic origin and differential loss of eukaryotic genes. *Nature* **524**, 427–432 (2015).
90. Pittis, A. A. & Gabaldón, T. Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature* **531**, 101–104 (2016).
This paper represents the first formal testing of the timing of acquisition of the mitochondrion by use of comparisons of phylogenetic distances between eukaryotic proteins and their closest prokaryotic relatives.
91. Yutin, N., Makarova, K. S., Mekhedov, S. L., Wolf, Y. I. & Koonin, E. V. The deep archaeal roots of eukaryotes. *Mol. Biol. Evol.* **25**, 1619–1630 (2008).
92. Esser, C. *et al.* A genome phylogeny for mitochondria among α-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol. Biol. Evol.* **21**, 1643–1660 (2004).
93. Davidov, Y., Huchon, D., Koval, S. F. & Jurkevitch, E. A new α-proteobacterial clade of Bdellovibrio-like predators: implications for the mitochondrial endosymbiotic theory. *Environ. Microbiol.* **8**, 2179–2188 (2006).
94. 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).
95. Williams, K. P., Sobral, B. W. & Dickerman, A. W. A robust species tree for the alphaproteobacteria. *J. Bacteriol.* **189**, 4578–4586 (2007).
96. Wu, M. *et al.* Phylogenomics of the reproductive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol.* **2**, E69 (2004).
97. 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).
98. 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).
99. Thrash, J. C. *et al.* Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade. *Sci. Rep.* **1**, 13 (2011).
100. Brindefalk, B., Ettema, T. J. G., Viklund, J., Thollesson, M. & Andersson, S. G. E. A phylometagenomic exploration of oceanic alphaproteobacteria reveals mitochondrial relatives unrelated to the SAR11 clade. *PLoS ONE* **6**, e24457 (2011).
101. Wang, Z. & Wu, M. An integrated phylogenomic approach toward pinpointing the origin of mitochondria. *Sci. Rep.* **5**, 7949 (2015).
102. Poole, A. M. & Penny, D. Evaluating hypotheses for the origin of eukaryotes. *Bioessays* **29**, 74–84 (2007).
103. Keeling, P. J. The impact of history on our perception of evolutionary events: endosymbiosis and the origin of eukaryotic complexity. *Cold Spring Harb. Perspect. Biol.* **6**, a016196 (2014).
104. McInerney, J. O., O'Connell, M. J. & Pisani, D. The hybrid nature of the Eukaryota and a consilient view of life on Earth. *Nat. Rev. Microbiol.* **12**, 449–455 (2014).
105. Moreira, D. & Lopez-Garcia, P. Symbiosis between methanogenic archaea and δ-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J. Mol. Evol.* **47**, 517–530 (1998).
This publication details the syntrophy hypothesis, which proposes a detailed mechanism suggesting that eukaryotes evolved from a two-step symbiosis.
106. López-García, P. & Moreira, D. Open questions on the origin of eukaryotes. *Trends Ecol. Evol.* **30**, 697–708 (2015).
Among other topics, this review discusses the necessity to determine the mechanistic and selective forces explaining the origin of key eukaryotic features, such as the nucleus or the bacterial-like eukaryotic membrane system.
107. 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).
108. Cavalier-Smith, T. Molecular phylogeny. Archaebacteria and Archezoa. *Nature* **339**, 100–101 (1989).
109. Martin, W. & Müller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41 (1998).
This study proposes one of the first and most elaborate models of a symbiogenetic origin of eukaryotes.
110. Searcy, D. G. in *The Origin and Evolution of the Cell* (eds Hartmann, H. *et al.*) 47–78 (World Scientific, 1992).
111. Karnkowska, A. *et al.* A eukaryote without a mitochondrial organelle. *Curr. Biol.* **26**, 1274–1284 (2016).
112. Martin, W. F. *et al.* Late mitochondrial origin is an artefact. *Genome Biol. Evol.* **9**, 373–379 (2017).
113. Pittis, A. A. & Gabaldon, T. On phylogenetic branch lengths distribution and the late acquistion of mitochondria. Preprint at https://www.biorxiv.org/content/early/2016/07/20/064873.article-info (2016).
114. Ettema, T. J. G. Evolution: mitochondria in the second act. *Nature* **531**, 39–40 (2016).
115. Wolf, Y. I., Makarova, K. S., Yutin, N. & Koonin, E. V. Updated clusters of orthologous genes for Archaea: a complex ancestor of the Archaea and the byways of horizontal gene transfer. *Biol. Direct* **7**, 46 (2012).
116. Koonin, E. V. & Yutin, N. The dispersed archaeal eukaryome and the complex archaeal ancestor of eukaryotes. *Cold Spring Harb. Perspect. Biol.* **6**, a016188 (2014).
117. Koonin, E. V., Makarova, K. S. & Elkins, J. G. Orthologs of the small RPB8 subunit of the eukaryotic RNA polymerases are conserved in hyperthermophilic Crenarchaeota and "Korarchaeota". *Biol. Direct* **2**, 38 (2007).
118. Zwickl, P. *et al.* Primary structure of the Thermoplasma proteasome and its implications for the structure, function, and evolution of the multicatalytic proteinase. *Biochemistry* **31**, 964–972 (1992).
119. James, R. H. *et al.* Functional reconstruction of a eukaryotic-likeE1/E2/(RING) E3 ubiquitylation cascade from an uncultured archaeon. *Nat Commun.* (in the press).
120. Makarova, K. S. & Koonin, E. V. Archaeal ubiquitin-like proteins: functional versatility and putative ancestral involvement in tRNA modification revealed by comparative genomic analysis. *Archaea* **2010**, 710303 (2010).
121. Raiborg, C. & Stenmark, H. The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins. *Nature* **458**, 445–452 (2009).
122. Jékely, G. Origin and evolution of the self-organizing cytoskeleton in the network of eukaryotic organelles. *Cold Spring Harb. Perspect. Biol.* **6**, a016030 (2014).
123. Ettema, T. J. G., Lindås, A.-C. & Bernander, R. An actin-based cytoskeleton in archaea. *Mol. Microbiol.* **80**, 1052–1061 (2011).
124. Lindås, A. C., Chruszcz, M., Bernander, R. & Valegård, K. Structure of crenactin, an archaeal actin homologue active at 90 °C. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 492–500 (2014).
125. Izoré, T., Duman, R., Kureisaite-Ciziene, D. & Löwe, J. Crenactin from *Pyrobaculum calidifontis* is closely related to actin in structure and forms steep helical filaments. *FEBS Lett.* **588**, 776–782 (2014).
126. Yutin, N., Wolf, M. Y., Wolf, Y. I. & Koonin, E. V. The origins of phagocytosis and eukaryogenesis. *Biol. Direct* **4**, 9 (2009).
127. Yutin, N. & Koonin, E. V. Archaeal origin of tubulin. *Biol. Direct* **7**, 10 (2012).
128. Makarova, K. S., Yutin, N., Bell, S. D. & Koonin, E. V. Evolution of diverse cell division and vesicle formation systems in Archaea. *Nat. Rev. Microbiol.* **8**, 731–741 (2010).
129. Lindås, A.-C., Karlsson, E. A., Lindgren, M. T., Ettema, T. J. G. & Bernander, R. A unique cell division machinery in the Archaea. *Proc. Natl Acad. Sci. USA* **105**, 18942–18946 (2008).
130. Samson, R. Y., Obita, T., Freund, S. M., Williams, R. L. & Bell, S. D. A role for the ESCRT system in cell division in archaea. *Science* **322**, 1710–1713 (2008).
131. Pelve, E. A. *et al.* Cdv-based cell division and cell cycle organization in the thaumarchaeon *Nitrosopumilus maritimus*. *Mol. Microbiol.* **82**, 555–566 (2011).
132. Saw, J. H. *et al.* Exploring microbial dark matter to resolve the deep archaeal ancestry of eukaryotes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20140328 (2015).
133. 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).
134. Brandizzi, F. & Barlowe, C. Organization of the ER-Golgi interface for membrane traffic control. *Nat. Rev. Mol. Cell Biol.* **14**, 382–392 (2013).
135. Rout, M. P. & Field, M. C. The evolution of organellar coat complexes and organization of the eukaryotic cell. *Annu. Rev. Biochem.* **86**, 637–657 (2017).
This is an extensive review on the origin and early evolution of the eukaryotic endomembrane system.

136. Sousa, F. L., Neukirchen, S., Allen, J. F., Lane, N. & Martin, W. F. Lokiarchaeon is hydrogen dependent. *Nat. Microbiol.* **1**, 16034 (2016).
137. 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).
138. Mariotti, M. *et al. Lokiarchaeota* marks the transition between the archaeal and eukaryotic selenocysteine encoding systems. *Mol. Biol. Evol.* **33**, 2441–2453 (2016).
139. Philippe, H. *et al.* Pitfalls in supermatrix phylogenomics. *Eur. J. Taxon.* **283**, 1–25 (2017).
140. Lartillot, N. & Philippe, H. A. Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21**, 1095–1109 (2004).
141. Foster, P. G. Modeling compositional heterogeneity. *Syst. Biol.* **53**, 485–495 (2004).
142. 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).
This work investigates archaeal gene family evolution to find the root of the archaeal tree and to infer the metabolism of the archaeal ancestors.
143. Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* **7**, 13219 (2016).
144. Brown, C. T. *et al.* Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208–211 (2015).
145. Hug, L. A. *et al.* A new view of the tree of life. *Nat. Microbiol.* **1**, 16048 (2016).
146. Woese, C. R., Magrum, L. J. & Fox, G. E. Archaebacteria. *J. Mol. Evol.* **11**, 245–251 (1978).
147. Lombard, J., López-García, P. & Moreira, D. The early evolution of lipid membranes and the three domains of life. *Nat. Rev. Microbiol.* **10**, 507–515 (2012).
148. Baum, D. A. & Baum, B. An inside-out origin for the eukaryotic cell. *BMC Biol.* **12**, 76 (2014).
149. 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).
150. Villanueva, L., Schouten, S. & Damsté, J. S. Phylogenomic analysis of lipid biosynthetic genes of Archaea shed light on the 'lipid divide'. *Environ. Microbiol.* **19**, 54–69 (2017).
151. Villanueva, L., Damsté, J. S. & Schouten, S. A re-evaluation of the archaeal membrane lipid biosynthetic pathway. *Nat. Rev. Microbiol.* **12**, 438–448 (2014).
152. Lombard, J., López-García, P. & Moreira, D. An ACP-independent fatty acid synthesis pathway in archaea: implications for the origin of phospholipids. *Mol. Biol. Evol.* **29**, 3261–3265 (2012).
153. Dibrova, D. V., Galperin, M. Y. & Mulkidjanian, A. Y. Phylogenomic reconstruction of archaeal fatty acid metabolism. *Environ. Microbiol.* **16**, 907–918 (2014).
154. Yokobori, S.-I., Nakajima, Y., Akanuma, S. & Yamagishi, A. Birth of archaeal cells: molecular phylogenetic analyses of G1P dehydrogenase, G3P dehydrogenases, and glycerol kinase suggest derived features of archaeal membranes having G1P polar lipids. *Archaea* **2016**, 1802675 (2016).
155. Brochier-Armanet, C., Forterre, P. & Gribaldo, S. Phylogeny and evolution of the Archaea: one hundred genomes later. *Curr. Opin. Microbiol.* **14**, 274–281 (2011).
156. Narasingarao, P. *et al. De novo* metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *ISME J.* **6**, 81–93 (2012).
157. 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).
158. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).

Acknowledgements
The authors wish to thank A. Roger and W. F. Doolittle for fruitful discussions. L.E. is funded by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 704263. J.L. is supported by a postdoctoral fellowship for foreign researchers from the Wenner-Gren Foundations in Stockholm (UPD2016-0072). C.W.S. is supported by a European Molecular Biology Organization long-term fellowship (ALTF-997-2015) and the Natural Sciences and Engineering Research Council of Canada postdoctoral research fellowship (PDF-487174-2016). This work is supported by grants of the European Research Council (ERC Starting Grant 310039-PUZZLE_CELL), the Swedish Foundation for Strategic Research (SSF-FFL5) and the Swedish Research Council (VR grant 2015–04959), awarded to T.J.G.E.

Author contributions
L.E., A.S., J.L., C.W.S. and T.J.G.E. researched data for the article. L.E., A.S., J.L., C.W.S. and T.J.G.E. substantially contributed to the discussion of content. L.E., A.S., J.L. and T.J.G.E. wrote the article. L.E., A.S., J.L., C.W.S. and T.J.G.E. reviewed and edited the manuscript before submission.

Competing interests statement
The authors declare no competing interests.

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

CORRIGENDUM

Archaea and the origin of eukaryotes

Laura Eme, Anja Spang, Jonathan Lombard, Courtney W. Stairs and Thijs J. G. Ettema

Nature Reviews Microbiology **15**, 711–723 (2017)

On pages 714–715 of this article, in the first paragraph of the section *What do we currently know about LECA?*, the sentence "Phylogenomic and comparative genomic analyses have led to the hypothesis that LECA, estimated to have lived ~1–1.9 million years ago[54], already was a fully fledged eukaryote and possessed a large number of features that are uniquely found in modern eukaryotes[55,56]." should have read "Phylogenomic and comparative genomic analyses have led to the hypothesis that LECA, estimated to have lived ~1–1.9 billion years ago[54], already was a fully fledged eukaryote and possessed a large number of features that are uniquely found in modern eukaryotes[55,56]." This has been corrected in the online version of the article. The authors apologize to the readers for any misunderstanding caused.