

Review



Cite this article: Koonin EV. 2015 Origin of eukaryotes from within archaea, archaeal eukaryome and bursts of gene gain: eukaryogenesis just made easier? *Phil. Trans. R. Soc. B* **370**: 20140333.
<http://dx.doi.org/10.1098/rstb.2014.0333>

Accepted: 9 June 2015

One contribution of 17 to a theme issue 'Eukaryotic origins: progress and challenges'.

Subject Areas:

evolution

Keywords:

endosymbiosis, phagocytosis, cytoskeleton, horizontal gene transfer, archaea

Author for correspondence:

Eugene V. Koonin
e-mail: koonin@ncbi.nlm.nih.gov

Origin of eukaryotes from within archaea, archaeal eukaryome and bursts of gene gain: eukaryogenesis just made easier?

Eugene V. Koonin

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

The origin of eukaryotes is a fundamental, forbidding evolutionary puzzle. Comparative genomic analysis clearly shows that the last eukaryotic common ancestor (LECA) possessed most of the signature complex features of modern eukaryotic cells, in particular the mitochondria, the endomembrane system including the nucleus, an advanced cytoskeleton and the ubiquitin network. Numerous duplications of ancestral genes, e.g. DNA polymerases, RNA polymerases and proteasome subunits, also can be traced back to the LECA. Thus, the LECA was not a primitive organism and its emergence must have resulted from extensive evolution towards cellular complexity. However, the scenario of eukaryogenesis, and in particular the relationship between endosymbiosis and the origin of eukaryotes, is far from being clear. Four recent developments provide new clues to the likely routes of eukaryogenesis. First, evolutionary reconstructions suggest complex ancestors for most of the major groups of archaea, with the subsequent evolution dominated by gene loss. Second, homologues of signature eukaryotic proteins, such as actin and tubulin that form the core of the cytoskeleton or the ubiquitin system, have been detected in diverse archaea. The discovery of this 'dispersed eukaryome' implies that the archaeal ancestor of eukaryotes was a complex cell that might have been capable of a primitive form of phagocytosis and thus conducive to endosymbiont capture. Third, phylogenomic analyses converge on the origin of most eukaryotic genes of archaeal descent from within the archaeal evolutionary tree, specifically, the TACK superphylum. Fourth, evidence has been presented that the origin of the major archaeal phyla involved massive acquisition of bacterial genes. Taken together, these findings make the symbiogenetic scenario for the origin of eukaryotes considerably more plausible and the origin of the organizational complexity of eukaryotic cells more readily explainable than they appeared until recently.

1. Introduction

The origin of eukaryotes is one of the hardest and most intriguing problems in the study of the evolution of life, and arguably, in the whole of biology. Compared to archaea and bacteria (collectively, prokaryotes), eukaryotic cells are three to four orders of magnitude larger in volume and display a qualitatively higher level of complexity of intracellular organization [1–3]. Unlike the great majority of prokaryotes, eukaryotic cells possess an extended system of intracellular membranes that includes the eponymous eukaryotic organelle, the nucleus, and fully compartmentalizes the intracellular space. In eukaryotic cells, proteins, nucleic acids and small molecules are distributed by specific trafficking mechanisms rather than by free diffusion as is largely the case in bacteria and archaea [4,5]. Thus, eukaryotic cells function on different physical principles compared to prokaryotic cells, which is directly due to their (comparatively) enormous size.

The gulf between the cellular organizations of eukaryotes and prokaryotes is all the more striking because no intermediates have been found. Comparative

analysis of eukaryotic cells and genomes confidently maps highly advanced functional systems and macromolecular complexes to the last eukaryotic common ancestor (LECA). The actin and tubulin cytoskeletons, the nuclear pore, the spliceosome, the proteasome and the ubiquitin signalling system are only a few of the striking examples of the organizational complexity that seems to be a 'birthright' of eukaryotic cells [6–10]. The formidable problem that these fundamental complex features present to evolutionary biologists makes Darwin's famous account of the evolution of the eye look like a simple, straightforward case. Indeed, so intimidating is the challenge of eukaryogenesis that the infamous notion of 'irreducible complexity' has sneaked into serious scientific debate [11], albeit followed by a swift refutation [12].

Molecular phylogenetics and phylogenomics revealed fundamental aspects of the origin of eukaryotes. The 'standard model' of molecular evolution, derived primarily from the classic phylogenetic analysis of 16S RNA by Woese and co-workers and supported by subsequent phylogenetic analyses of universal genes, identifies eukaryotes as the sister group of archaea, to the exclusion of bacteria [13–16]. Within the eukaryotic part of the tree, early phylogenetic studies have placed into the root position several groups of unicellular organisms, primarily parasites, that unlike the majority of eukaryotes, lack mitochondria. These organisms have been construed as 'archezoa', i.e. the primary amitochondrial eukaryotes that were thought to have hosted the proto-mitochondrial endosymbiont [17–20].

However, advances of comparative genomics jointly with discoveries of cell biology have put the archezoan scenario of eukaryogenesis into serious doubt. First, it has been shown that all the purported archezoa possess organelles, such as hydrogenosomes and mitosomes, that appeared to be derivatives of the mitochondria. These mitochondria-like organelles typically lack genomes but contain proteins encoded by genes of apparent bacterial origin that encode homologous mitochondrial proteins in other eukaryotes [21,22]. Combined, the structural and phylogenetic observations leave no reasonable doubt that hydrogenosomes and mitosomes indeed evolved from the mitochondria. Accordingly, no primary amitochondrial eukaryotes are currently known, suggesting that the primary α -proteobacterial endosymbiosis antedated the LECA [22–24]. Compatible with this conclusion, subsequent, refined phylogenetic studies have placed the former 'archezoa' within different groups of eukaryotes indicating that their initial position at the root was an artefact caused by their fast evolution, most probably causally linked to the parasitic lifestyle [25–27]. These parallel developments left the archezoan scenario without concrete support but have not altogether eliminated its attractiveness. An adjustment to the archezoan scenario simply posited that the archezoa was an extinct group that had been driven out of existence by the more efficient mitochondrial eukaryotes [28,29]. A concept predicated on an extinct group of organisms that is unlikely to have left behind any fossils and is refractory to evolutionary reconstruction due to the presence of mitochondria (or vestiges thereof) in all eukaryotes is quite difficult to refute but can hardly get much traction without any concrete evidence of the existence of archezoa.

The radical alternative to the elusive archezoa is offered by symbiogenetic scenarios of eukaryogenesis according to which archezoa, i.e. primary amitochondrial eukaryotes, have never existed, and the eukaryotic cell is the product of a symbiosis

between two prokaryotes [1,2,12,23,30,31]. Comparative genomic analysis clearly demonstrates that eukaryotes possess two distinct sets of genes, one of which shows phylogenetic affinity with homologues from archaea, whereas the other one includes genes affiliated with bacterial homologues (apart from these two classes, there are many eukaryotic genes of uncertain provenance). The eukaryotic genes of apparent archaeal descent encode, primarily, proteins involved in information processing (translation, transcription, replication, repair), whereas the genes of inferred bacterial origin encode mostly proteins with 'operational' functions such as metabolic enzymes, components of membranes and other cellular structures and others [32–35]. Notably, altogether, the number of eukaryotic protein-coding genes of bacterial origin exceeds the number of 'archaeal' proteins about threefold. Thus, although many highly conserved, universal genes of eukaryotes indeed appear to be of archaeal origin, the archaeo-eukaryotic affinity certainly does not tell the entire story of eukaryogenesis, not even most of that story if judged by the proportions of genes of apparent archaeal and bacterial descent.

Although several symbiogenetic scenarios that differ in terms of the proposed partners and even the number of symbiotic events involved have been proposed, the simplest, parsimonious one that accounts for both the ancestral presence of mitochondria in eukaryotes and the hybrid composition of the eukaryotic gene complement involves engulfment of an α -proteobacterium by an archaeon [12,30,33,36]. Under this scenario, a chain of events has been proposed that leads from the endosymbiosis to the emergence of eukaryotic innovations such as the endomembrane system, including the nucleus and the cytoskeleton [36,37]. Subsequently, argument has been developed that the energy demand of a eukaryotic cell that is orders of magnitude higher than that of a typical prokaryotic cell cannot be met by means other than utilization of multiple 'power stations' such as the mitochondria [1,2,31].

A major problem faced by this scenario (and symbiogenetic scenarios in general) is the mechanistic difficulty of the engulfment of one prokaryotic cell by another [20,28,29,38]. Although bacterial endosymbionts of certain proteobacteria have been described [39,40], such a relationship appears to be a rarity. By contrast, in many unicellular eukaryotes, such as amoeba, engulfment of bacterial cells is routine due to the phagotrophic lifestyle of these organisms [20]. The apparent absence of phagocytosis in archaea and bacteria prompted the reasoning that the host of the proto-mitochondrial endosymbiont was a primitive phagotrophic eukaryote, which implies the presence of an advanced endomembrane system and cytoskeleton [20,28,29,38]. Thus, argument from cell biology seemed to justify rescuing the archezoan scenario, the lack of positive evidence notwithstanding.

However, comparative analysis of the increasingly diverse collection of archaeal and bacterial genomes has yielded multiple lines of evidence that might change the notion of the implausibility of an archaeo-bacterial endosymbiosis. In this article, I discuss the results of genome evolution reconstructions that imply complex ancestral archaeal forms and the discovery of the dispersed archaeal 'eukaryome'. The eukaryome consists of multiple genes identified in different archaea that encode key components of the cytoskeleton, the cell division apparatus, the ubiquitin system and other signature eukaryotic cellular systems. A complementary line of recent developments shows that massive acquisition of bacterial genes probably occurred on multiple occasions in the course of the evolution of archaea.

Taken together, these findings seem to be making the scenario of archaeo-bacterial symbiosis considerably more plausible than it appeared even recently.

2. Burgeoning archaeal diversity, complex archaeal ancestor and origin of eukaryotes from within the archaea

As pointed out above, the 'standard model' phylogeny of Woese and co-workers clearly identifies archaea and eukaryotes as sister groups [13–15]. However, an alternative phylogeny inferred from trees of the same 16S rRNAs using a different phylogenetic method and compatible also with the results of ribosome structure comparison appeared shortly after the publication of the three-domain tree of life [41,42]. That alternative topology led to the eocyte hypothesis under which eukaryotes evolved from within the archaea and are a sister group to the 'eocytes', i.e. the archaeal phylum that is currently known as Crenarchaeota [41–44]. Support for the eocyte hypothesis has been subsequently reported from comparative analysis of ribosomal protein sequences that did not involve phylogeny reconstruction [45] and from a novel phylogenomic approach [33]. A later, sophisticated phylogenetic analysis of multiple conserved genes that employed a technique eliminating fast-evolving alignment columns and has been reported to minimize the effect of common artefacts of phylogenetic analysis, such as long-branch attraction, has supported the affinity of eukaryotes with Crenarchaeota [46]. The alternative topologies including the standard three-domain phylogeny with archaea and eukaryotes as sister groups have been deemed to result from phylogenetic artefacts that are overcome by this approach.

The eocytes (Crenarchaeota) are not the only group of archaea that has been proposed for the role of the archaeal ancestor of eukaryotes. Evolutionary scenarios based on different versions of metabolic cooperation between the archaeal and bacterial partners of the primary endosymbiosis, such as the hydrogen hypothesis [30] and the syntrophic hypothesis [47], implied origin of the 'archaeal' genes of eukaryotes from euryarchaea, and in particular, methanogens. The methanogen affinity for eukaryotes has been claimed also from some phylogenetic analyses [48,49]. Yet other phylogenetic studies have produced results compatible with the standard model, placing eukaryotes outside the known archaeal diversity [50–53]. Biological considerations, such as the greater role of RNA in a variety of processes in eukaryotic cells (splicing, defence, regulation of gene activity and more), have even led to the idea that eukaryotes were the first cellular life forms [54–56].

Most of the inferences of archaeo-eukaryotic relationship studies mentioned above were based either on phylogenetic analysis of a single, universal gene, such as 16S rRNA, or on concatenated sequences of several universal proteins (e.g. ribosomal proteins), or on non-sequence characters such as gene repertoires (phyletic patterns) and domain architectures of multidomain proteins. Obviously, extensive sequencing of genomes from all three domains of cellular life opens the door for comprehensive phylogenomic analyses. The first such extensive phylogenomic exercise involved analysis of nearly 6000 gene sets from 185 archaeal, bacterial and eukaryotic genomes and employed a supertree approach to combine information from the multiple trees; the

results suggested affinity between eukaryotes and the Thermoplasmatales branch of Euryarchaeota, albeit with limited statistical support [34]. *Thermoplasma* or a related, wall-less archaeon also has been proposed as a plausible ancestor of eukaryotes, on the basis of biochemical and cytological considerations, long before the phylogenomic analysis became feasible [57–61].

Another comprehensive phylogenetic analysis of individual eukaryotic genes of apparent archaeal origin has suggested an origin outside of the known archaeal diversity for most of these genes but also identified many genes with a crenarchaeal (eocyte) affinity and a smaller number of genes with a euryarchaeal affinity [62]. In this study, the possibility has been brought up that the discrepancies between the tree topologies for different genes did not necessarily result from phylogenetic artefacts, but instead could reflect joint presence of genes currently identified in different archaea in the genome of the ancestral form that became the host of the proto-mitochondrial endosymbiont.

The diversity of the outcomes of phylogenetic analysis, with the origin of eukaryotes scattered around the archaeal diversity, has led to considerable frustration and suggested that a 'phylogenomic impasse' has been reached, owing to the inadequacy of the available phylogenetic methods for disambiguating deep relationships [63]. However, the landscape of archaeal phylogenomics started to radically change when genome analysis of several poorly characterized archaea suggested the existence of several new phyla, in particular Korarchaeota [64] and Thaumarchaeota, the latter encompassing mesophilic archaea previously included within the Crenarchaeota [65]. Subsequently, it has been shown that Thaumarchaeota are a widespread microbial group of major geochemical importance that includes, in particular, the key ammonia oxidizers in marine and soil habitats [66–68]. For Korarchaeota, there is still a single complete genome but metagenomic studies suggest substantial diversity in various marine and terrestrial habitats [69,70]. Genome analysis of the uncultivated archaeon *Candidatus Caldiarchaeum subterraneum* has suggested yet another archaeal phylum, dubbed Aigarchaeota [71]. The latest, extensive metagenomic and single-cell genomics studies have led to a veritable 'bonanza' of putative new archaeal phyla [67,72–74] (figure 1).

Independent phylogenomic analyses of multiple conserved genes consistently support a deeply rooted archaeal 'TACK' superphylum that originally encompassed Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota [80–84], but according to the latest comprehensive phylogenetic study, additionally contains two novel phyla, Bathyarchaeota and Geoarchaeota (however, a subsequent re-analysis has suggested inclusion of Geoarchaeota into Crenarchaeota, thus denying this group the status of a new phylum (figure 1)). This new phylogeny also includes another putative superphylum designated DPANN that combines Nanoarchaeota and other archaeal groups with small genomes.

A recent detailed phylogenomic study that included an expanded set of nearly universal phylogenetic markers and improved phylogenetic methods has confidently placed the root of the archaeal tree between the Euryarchaeota, including Nanoarchaeota and other fast-evolving groups, and the rest of the archaeal phyla that comprise the TACK superphylum, or proposed new kingdom Proteoarchaeota [75,76] (figure 1). However, other recent phylogenomic analysis using different techniques and datasets variously place the root between the DPANN superphylum and the rest of the Archaea [85] or

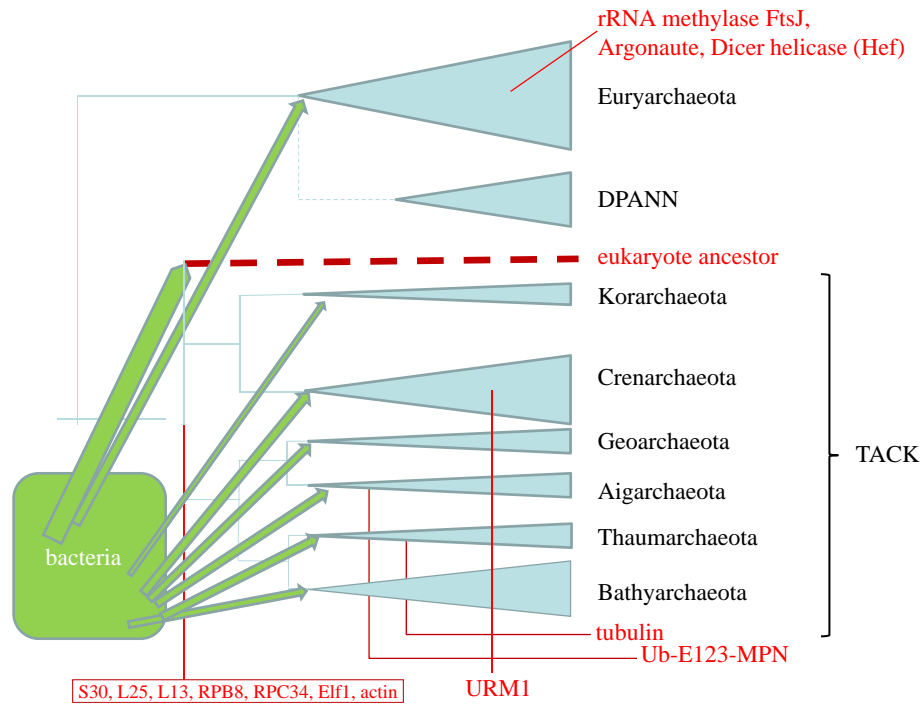


Figure 1. A schematic evolutionary tree of the archaea, the likely origin of eukaryotes and the distribution of eukaryome components. The tree topology is from [73] except that the DPANN branch was moved to the base of the Euryarchaeota according to [75–77]. The size of the triangles very roughly shows the diversity of the respective groups. The hypothetical lineages of eukaryote ancestors are tentatively shown as a deep branch within the TACK superphylum. The inferred origins of some key eukaryotic genes and functional systems [78] are indicated by red lines; S30, L25 and L13 are ribosomal proteins; Ub-E123-MPN denotes the ubiquitin system where E123 are the respective subunits of the ubiquitin ligase and MPN is the deubiquitinase. The green arrows from bacteria denote the gene flow associated with the origin of the major groups of archaea [79]. The thick arrow pointing at the putative ancestor of eukaryotes denotes the massive gene flow from the endosymbiont.

within the Euryarchaeota [86,87]. Thus, the archaeal root position remains an open problem. Given that in order to establish the root position for the Archaea the use of a representative sample of bacterial homologues as an outgroup is essential, attempts to solve this problem involve the deepest relationships between cellular life forms for which detection of an unequivocal signal is inherently difficult.

The monophyly of the TACK superphylum is further buttressed by the reconstruction of the evolution of the archaeal gene repertoire that revealed probably massive gene gain at the base of the TACK branch (figure 1) [84]. This reconstruction, which extended previous analyses [88,89], reveals a remarkable trend in the evolution of archaea that is likely to reflect a general pattern of genome evolution. This general tendency consists in the dominance of genome reduction at the most common course of evolution that is punctuated with episodes of explosive genome expansion [90]. These reconstructions indicate that each of the major archaeal lineages underwent some degree of genome reduction in the course of its evolution and that the gene complement of the last archaeal common ancestor (LACA) was at least as complex as that of most of the extant archaea. Reductive evolution associated with genome streamlining appears to be an extremely general evolutionary phenomenon characteristic of successful groups that reach large effective population sizes and evolve under strong selective pressure [90]. However, evolution of the archaea might be specifically conducive to genome reduction as part of adaptation to high stress conditions [54,91].

Identification of new archaeal phyla and the putative TACK superphylum stimulated further phylogenomic effort aimed at

the elucidation of the archaeal ancestry of eukaryotes. Two independent, detailed phylogenetic analyses of rRNA and universal protein-coding genes employing state-of-the-art phylogenetic methods have shown significant support for the monophyly of eukaryotes with the TACK superphylum but not with any specific lineage within the TACK [81,92,93]. An alternative analysis has placed eukaryotes within the TACK superphylum, as a sister group to Thaumarchaeota [94]. The controversy over the phylogenetic position of eukaryotes, or more precisely, universal eukaryotic genes encoding translation system components, vis-a-vis archaea has not been put to rest by these analyses. Thus, a phylogenomic analysis that focused on the archaeal ‘dark matter’, i.e. sequences from numerous uncultivated organisms, has supported the standard model topology, i.e. eukaryotes outside the archaea [72]. However, a re-analysis that employed a better fitting phylogenetic model and excluded some eukaryotic genes of mitochondrial and chloroplast origin that appeared to have crept into the dark matter study has once again recovered the eukaryote–TACK affinity [85].

Importantly, an evolutionary affinity between eukaryotes and the TACK superphylum is compatible with a series of findings that are independent of phylogenetic methodology. It has been shown that several ancestral genes are shared exclusively by eukaryotes and archaea of the TACK superphylum, in contrast to a smaller number of such shared derived characters between eukaryotes and Euryarchaeota (figure 1) [78]. Notably, the shared derived characters of eukaryotes and the TACK superphylum include several components of the core information processing system, including three ribosomal proteins [83], the RNA polymerase subunits

RPB8 [95] and RPC34 [96] and the transcription factor Elf1 [97]. Genes in this category are relatively rarely transferred horizontally, so the shared ones are likely to come from the common ancestor of the respective groups, in this case, eukaryotes and the TACK superphylum archaea.

Taking together all the relevant lines of evidence, it appears that, although claiming a definitive solution could be premature, a consensus is shaping up on the specific origin of the archaeal heritage of eukaryotes. Most of the genes that eukaryotes inherited from archaea appear to originate from the TACK superphylum, although there are interesting exceptions to this pattern as discussed below (figure 1). Thus, the results of increasingly robust phylogenomic analyses appear to be best compatible with a model of two primary domains of cellular life, Archaea and Bacteria, with eukaryotes emerging from within the Archaea, as opposed to the standard three-domain model [93]. This conclusion does not rule out the possibility that the eukaryotic ancestor that evolved from a common root with or from within the TACK and eventually hosted the proto-mitochondrial endosymbiont was an 'archezoan' but appears to fit more smoothly into a scenario of the host being a bona fide archaeon.

3. Massive gene gain: apparent common denominator in the origin of new archaeal phyla

As pointed out above, reconstructions of the evolution of the archaeal gene complement imply episodic gene gain, conceivably associated with the emergence of major groups, followed by gradual gene loss leading to genome streamlining in multiple lineages [84,90]. Recently, this view of archaeal evolution has received strong support from focused studies of acquisition of bacterial genes by archaea. Ever since the first genomes of mesophilic archaea were reported, it has become clear that these organisms contain numerous genes of apparent bacterial origin, many more than archaeal thermophiles [98–100]. Subsequent phylogenomic study of Halobacteria have revealed a striking pattern: over 1000 bacterial genes apparently have been acquired by a methanogenic ancestor of Halobacteria and recruited for the characteristic halobacterial metabolic pathways [101]. Thus, this massive capture of bacterial genes seems to have led to the emergence of a major group of archaea. Comprehensive phylogenomic analysis of all available archaeal and bacterial genomes has expanded these observations by showing that capture of multiple bacterial genes is characteristic of 13 major groups of archaea [79]. Moreover, topologies of the phylogenetic trees appear to be best compatible with massive acquisition of the bacterial genes at the base of each archaeal branch as opposed to piecemeal acquisition along the branch (figure 1).

These findings have obvious and striking implications for the origin of eukaryotes. Acquisitions of numerous bacterial genes that amount to genomic chimaerism and lead to substantial remodeling of cell physiology and emergence of groups with new lifestyles appears to be a recurrent rather than unique event in evolution, at least in archaea. Could it be that most if not all major groups of archaea emerged from botched endosymbiotic events? Should that be the case, eukaryogenesis only differs in that the endosymbiont survived, retaining part of its physical and genetic identity.

4. The scattered archaeal eukaryome

The comparative genomic observations discussed above seem to increase the plausibility of an archaeal host for the mitochondrial endosymbiont and further indicate that evolutionary events leading to massive acquisition of bacterial genes were relatively common in archaeal evolution. Yet, the main obstacle faced by the symbiogenetic scenarios of eukaryogenesis, namely the mechanistic difficulty of engulfment of one prokaryotic cell by another, has remained as formidable as ever. As long as the chance of an archaeon engulfing a bacterium is considered to be vanishingly low, the symbiogenetic scenarios can be dismissed as unrealistic [28,29].

However, the latest findings of comparative genomics cast this thorny issue in a different light. It has become clear that, apart from the core of the universal information processing systems, probably archaeal ancestors of signature eukaryotic genes and entire functional systems that have to do with the intracellular architecture are often found in diverse groups of archaea. We denote this set of genes the 'eukaryome' to emphasize their specific importance for the biology of eukaryotic cells [78].

Unexpectedly, for a substantial number of ancestral eukaryotic genes, homologues have been detected in only one group of archaea. These lineage-specific ancestral genes are scattered across the entire archaeal domain but are most common in the TACK superphylum and in particular in *Ca. Caldiarchaeum subterraneum*, so far the only representative of Aigarchaeota with a complete genome [78]. Below I discuss the most striking cases of the dispersal of the eukaryome components among archaea.

(a) The cytoskeleton

The indispensable structural framework of all eukaryotic cells is the advanced, elaborate cytoskeleton that consists of two major types of elements, namely actin-based filaments and tubulin-based microtubules [102,103]. The cytoskeleton is central to the discussion of the origin of eukaryotes, in particular because actin filaments play the key role in phagocytosis, the process that is considered critical for the engulfment of the proto-mitochondrial endosymbiont by its host, whatever the exact nature of the latter [38]. Until recently, bacteria and archaea have been thought to encode only distant homologues of actin and tubulin, the proteins of the MreB/FtsA and FtsZ families, respectively, that perform essential functions in the septation of bacterial and some archaeal cells [104,105]. The sequence similarity between the bacterial and archaeal proteins of these families and the eukaryotic cytoskeleton components is rather low, so that homology has been considered firmly established only through structural comparisons [106–109].

Recently, analysis of the expanding archaeal genome collection has changed this perspective. Proteins with high sequence similarity and unambiguous phylogenetic affinity to eukaryotic actins have been discovered in the crenarchaeal order of Thermoproteales, *Korarchaeum* and *Ca. Caldiarchaeum subterraneum*, with the implication that actin was already present in the last common ancestor of the TACK superphylum [110,111]. Following these findings of comparative genomics, it has been shown that archaeal actin homologues, named crenactins, actually form helical filaments resembling typical eukaryotic actin filaments [112–114].

Sequence analysis of the crenactins has indicated that these proteins contained several inserts that are present in eukaryotic actin-related proteins (ARPs) but not in actins themselves [110]. Once two crenactin structures have been solved, these inserts

have been shown to form extended loops [115,116]. The corresponding loops in the ARPs are required for the formation of branched filaments which are involved in phagocytosis [117]. This similarity led to the proposition that crenactin filaments could endow some members of the TACK superphylum with at least a rudimentary phagocytic capacity [110].

Highly conserved orthologues of tubulins, named artubulins, so far have been discovered only in the genomes of several ammonium-oxidizing Thaumarchaeota of the genera *Nitrosoarchaeum* and *Nitrosotenus* [118,119]. Although in this case horizontal gene transfer (HGT) from eukaryotes to archaea could not be technically ruled out, phylogenetic analysis results appear to be best compatible with an ancestral status of the artubulins with respect to the eukaryotic tubulins [118]. The structures and functions of the artubulins remain to be characterized but it appears highly likely that they form a microtubule-type cytoskeleton.

Thus, in a remarkable departure from recent common beliefs, both major forms of the eukaryotic cytoskeleton seem to belong within the archaeal heritage of eukaryotes. However, the apparent ancestral forms have been detected in widely different groups of extant archaea.

(b) Cell division and membrane remodelling systems

Cell division obviously is central to all cellular life forms. Nevertheless, the cell division mechanisms substantially differ between bacteria and at least some archaea, on the one hand, and eukaryotes, on the other hand. In bacteria, division is coupled to chromosome replication, with the progeny DNA molecules being pumped into the daughter cells concomitant with replication. Division is then completed by septation, with the formation of the septum driven by the Z-ring that consists of the FtsZ protein, a GTPase that is a distant homologue of eukaryotic tubulins [120,121]. In addition to nearly all bacteria, the FtsZ-centred division machinery is encoded in the genomes of most of the Euryarchaeota and Thaumarchaeota as well as *Korarchaeum*, with the implication that the division mechanisms of these archaea are similar to the bacterial one [111].

Unexpectedly, a distinct division system homologous to the eukaryotic ESCRT-III membrane remodelling complex has been discovered in the crenarchaeon *Sulfolobus acidocaldarius* [122–125] and subsequently identified with comparative genomic methods in all members of two of the three crenarchaeal orders, *Sulfolobales* and *Desulfurococcales*, as well as some Thaumarchaeota and Euryarchaeota [111]. Subsequently, it has been shown that the ESCRT-III-like complex is the primary cell division system in the thaumarchaeon *Nitrosopumilus maritimus* [126]. Furthermore, one of the ESCRT-III protein homologues, CdvA of *Sulfolobus acidocaldarius*, has been shown to form helical filaments that mediate membrane scission during cell division [127]. This finding reveals the distinct form of cytoskeleton that is required for division in ESCRT-III-encoding archaea.

The presence of the FtsZ-based and ESCRT-III-like division systems in a broad variety of diverse archaea implies that both machineries were present in the LACA, with subsequent differential losses in multiple lineages. The Crenarchaeota in the order Thermoproteales lack both of these division systems and thus must employ a distinct third one that most probably relies on the crenactin cytoskeleton [111,113]. Given that the origin of crenactin can be mapped to the base of the TACK superphylum (figure 1 and see above), one comes to the striking conclusion

that the common ancestor of the TACK most probably possessed all three cell division systems that are scattered among the extant archaea (although the likelihood of this conclusion depends on the position of the archaeal root). Discovery of additional distinct variants of the cell division apparatus in archaea appears plausible. At least one available archaeal genome, that of *Picrophilus torridus* (order Thermoplasmatales), lacks all three division machineries discussed above and hence is expected to employ a novel mechanism [111].

(c) The ubiquitin signalling system

The ubiquitin (Ub) system is the central signalling and regulatory network of the eukaryotic cell. This extremely complex machinery regulates protein degradation, topogenesis and function in all eukaryotes through modification of proteins by conjugation with various forms of (poly)Ub and its paralogues [128–130]. For many years, the Ub system had been considered a key eukaryotic innovation that seemed to have evolved by the exaptation route, i.e. recruitment of prokaryotic enzymes involved in thiamine and molybdopterin coenzyme biosynthesis for a novel function [131,132]. Subsequently, it has been shown that, with the participation of a homologue of the E1 subunit of eukaryotic Ub ligases, some of the archaeal Ub homologues are conjugated with proteins and target them for degradation [133–135]. However, the proteins involved in these processes are distant homologues of Ub and E1, so this discovery did not necessarily shed light directly on the origin of the eukaryotic Ub systems; in particular, the origin of the E2 and E3 subunits of the Ub ligases remained elusive.

Analysis of the genome of *Ca. Caldiarchaeum subterraneum*, the founding member of the putative phylum Aigarchaeota, has transformed the entire story of the origin of the Ub system by revealing a predicted operon encoding a Ub-like protein and homologs of all three Ub ligase subunits along with a key deubiquitinating enzyme [71]. Operons with a similar organization of Ub-related genes have been identified also in several bacteria suggestive of horizontal dissemination of these operons among prokaryotes [78]. In contrast to the distant homologues of Ub system components that have been detected previously in other archaea, the homologues from *Ca. Caldiarchaeum subterraneum* show high sequence similarity to eukaryotic counterparts, and phylogenetic analysis of the E1 subunit of Ub ligase and the deubiquitinating enzyme MPN (the two largest and most conserved proteins in the Ub system that accordingly are conducive to phylogenetic analysis) place *Ca. Caldiarchaeum subterraneum* in the midst of eukaryotes [78]. Subsequent comparative genomic analysis has led to the identification of similar predicted operons in multiple genomes of Aigarchaeota that have been assembled from metagenomic sequences [136]. The possibility of acquisition of the Ub system by archaea from eukaryotes via HGT can be ruled out given the operonic organization of the archaeal genes. Thus, to date, Aigarchaeota encode the best candidate for the ancestor of the eukaryotic Ub system.

Other homologues of the Ub system components are scattered among archaea [136,137]. In particular, archaeal orthologues of the distant Ub homologue Urm1, which is conserved in all eukaryotes and performs a dual function as a sulfur carrier in tRNA thiolation and in protein modification [138,139], are detectable only among the members of the crenarchaeal order Sulfolobales [137] (figure 1). Thus, within the broadly defined Ub system, at least two distinct

archaeal ancestors of essential eukaryotic functional modules have been detected.

(d) The RNA interference system (RNAi)

The RNAi is a hallmark eukaryotic functional system that is involved in the defence against viruses and transposons and in multiple pathways of gene expression regulation [140–142]. Phylogenomic reconstructions suggest that LECA already possessed a diversified RNAi system [140–143]. The diverse RNAi mechanisms are centred around two key protein families, the Dicers and the Argonautes. The Dicers combine helicase and RNase activities and are primarily responsible for the processing of small interfering (si) RNAs and microRNAs [144–147]. The Argonautes are nucleases of the RNase H superfamily, some of which directly attack the RNA targets of RNAi (slicers) whereas others bind microRNAs and guide it to the target with cleaving of the latter [145,148,149].

The Dicers are signature eukaryotic proteins that have no direct counterparts in bacteria or archaea and encompass a fusion of a helicase domain and two RNase III domains that are unique to eukaryotes. The dsRNA-specific RNase III is nearly ubiquitous in bacteria and apparently has been acquired by some mesophilic Euryarchaeota via HGT [150]. By contrast, the helicase domain of Dicer appears to have evolved from the euryarchaeal helicase–nuclease Hef that is involved in DNA replication and repair [141,151].

In contrast to Dicers, the Argonautes have numerous homologues in bacteria and archaea, primarily Euryarchaeota, and phylogenetic analysis clearly points to a euryarchaeal origin of the eukaryotic Argonaute family [152,153]. Comparative genomic analysis of the gene neighbourhoods of the archaeal and bacterial Argonautes has led to the hypothesis that these proteins are involved in RNA- or DNA-dependent defence against foreign nucleic acids, similar to their eukaryotic homologues [152]. Subsequently, the defence function of Argonautes in bacteria was shown through the demonstration that these proteins employ RNA or DNA guide molecules to target and cleave foreign DNA [154–156].

A third key component of the eukaryotic RNAi is an RNA-dependent RNA polymerase (lost in some eukaryotic lineages including vertebrates) that serves as an amplifier of siRNAs. Homologues and possible ancestors of this polymerase have been identified in some bacteriophages where they are most likely to function in transcription, as DNA-dependent RNA polymerases [141,157].

Thus, the RNAi system, a signature eukaryotic defence and regulatory network, appears to have evolved from the archaeal Argonaute-centred defence machinery through the accretion of additional components. The ancestral archaeal system remains to be thoroughly characterized, and it cannot be ruled out that functional analogues and possibly even direct homologues of Dicers will eventually be discovered in some archaea.

5. The archaeal ancestor of eukaryotes: a complex, ancient group within the TACK superphylum with a prototype phagocytic ability?

Arguably, the greatest difficulty faced by the endosymbiotic scenarios of eukaryogenesis is the apparent implausibility

(or at least extreme rarity) of the engulfment of one prokaryotic cell by another. The recent advances of comparative genomics, complemented by the progress in the cell biology of archaea, seem to be closing this gap. Combined with the quantitative findings of genome evolution reconstructions on extensive differential gene loss in most archaeal lineages, the discovery of the ‘dispersed’ archaeal eukaryome implies a highly complex archaeal ancestor of eukaryotes [78,86]. Conceivably, this ancestral form possessed advanced cellular organization and certain ‘eukaryote-like’ functional capacities provided by the ancestral versions of various eukaryotic functional systems that are represented in different lineages of extant archaea (figure 1). The critical point is that the hypothetical eukaryotic ancestor probably possessed a cytoskeleton that consisted of both actin filaments and tubulin microtubules and could provide for a primitive phagocytic capacity [82,110]. Furthermore, the likely presence of multiple cell division systems, such as the FtsZ-based machinery and ESCRT-III [111], in the archaeal ancestor of eukaryotes implies that one of these, perhaps the latter, was involved in processes distinct from division proper, such as membrane remodelling, that could contribute to phagocytosis. Indeed, eukaryotic ESCRT complexes are implicated in phagocytosis-related processes, in particular autophagy [158].

Most probably, the archaeal ancestor of eukaryotes was a wall-less mesophile that coexisted with diverse bacteria, so even a limited capacity for phagocytosis would greatly facilitate the capture of prospective endosymbionts. Extant mesophilic archaea, such as *Methanosarcinales* or *Halobacteria*, clearly have acquired numerous genes via HGT [84,98,100,101]. Moreover, the latest comparative genomic results suggest that massive acquisition of bacterial genes underlay the emergence of most if not all major archaeal phyla [79]. Conceivably, in the archaeal ancestor of eukaryotes, this gene gaining capacity was enhanced by the primitive phagocytosis, through transient engulfment of other archaea and bacteria. This ‘protophagocytic’ lifestyle would probably cause acquisition of genes from diverse bacterial sources, not the proto-mitochondrial endosymbiont alone, which could in part account for the weakness of the α -proteobacterial signal among the eukaryotic genes of apparent bacterial descent [159,160].

The results of phylogenomic analysis outlined above strongly suggest that the ancestor of eukaryotes was a deep branch within the TACK superphylum [81,92], possibly a distinct phylum, in addition to the currently recognized Crenarchaeota, Thaumarchaeota, Korarchaeota, Aigarchaeota, Bathyarchaeota and (possibly) Geoarchaeota (figure 1). Given that evolutionary reconstructions indicate that evolution of most of the major groups of archaea was dominated by genome reduction and streamlining [84,90], the ancestor of eukaryotes could have been a highly complex ancient archaeon. Therefore, it appears plausible that a still unidentified group of extant archaea within the TACK superphylum is a streamlined descendant of the eukaryotic ancestor (figure 1). The rapidly progressing metagenomics and especially single-cell genomic sequencing clearly have the potential to uncover this elusive ancestral group in the case that its archaeal descendants indeed have survived to this day.

If the archaeal ancestor of eukaryotes (the host of the proto-mitochondrial endosymbiont) was a complex organism with some signature features of eukaryotic cells, such as the cytoskeleton, the question emerges whether this ancestral form was an archaeon or a primitive, amitochondrial

eukaryote, i.e. an archezoan. The answer hinges on the definition or rather the salient features that ‘make an organism a eukaryote’. The cytoskeleton, membrane remodelling systems such as ESCRT-III, the Ub system and RNAi certainly are eukaryotic signatures. Yet, the defining traits of eukaryotes appear to be the large size of the eukaryotic cells coupled with the presence of the elaborate endomembrane system which includes the nucleus endowed with nuclear pores and the splicing machinery that is linked to the nucleocytoplasmic trafficking. Quantitative argument has been developed that the eukaryotic cellular organization is unsustainable without multiple energy-converting organelles, such as mitochondria [1]. Furthermore, coherent scenarios have been proposed for the origin of the endomembranes, the nucleus and the spliceosome-mediated splicing in the wake of endosymbiosis [36,37,47]. The combination of these findings and inferences strongly suggests that the host of the mitochondrial endosymbiont was an archaeon although perhaps a highly complex one.

6. Conclusion

Four groups of recent observations increase the plausibility of the symbiogenetic scenario for the origin of eukaryotes. The first line of evidence comes from the reconstructions of archaeal genome evolution which imply complex ancestral forms, with the subsequent evolution in most lineages dominated by gene loss. The related and perhaps most important clues come from the observations on the archaeal eukaryome that is scattered among diverse extant archaea. The putative complex archaeal ancestor of eukaryotes could have encoded most if not all components of the eukaryome within the same genome, possibly endowing this ancestral archaeon with certain eukaryote-like functionalities such as the ability to efficiently engulf other cells (a primitive version of phagocytosis). The third line of evidence consists of the increasingly confident demonstrations of the origin of the core eukaryotic genes from within the archaea, or more specifically, from a deeply branching group within the TACK superphylum. Given that all known extant members of this superphylum are typical archaea and not archezoa, these findings appear to favour an archaeal host for the proto-mitochondrial endosymbiont. Finally, the indications that massive acquisition of bacterial genes most probably triggered the emergence of the major groups of archaea put the origin of

eukaryotes into a more general evolutionary context. These discoveries make the origin of eukaryotes appear less dramatically different from the origin of other groups of organisms than is generally perceived. Horizontal transfer of numerous genes appeared to have been central in each case. The key difference is that in eukaryotes the source of the foreign genes, i.e. the endosymbiont, survived as an organelle, precipitating the radical restructuring of the cell. Given the likely origin of eukaryotes from within the archaeal diversity and the observations on the dispersed eukaryome, there seems to be high promise of new evolutionary insights coming from metagenomics and single-cell genomics. The discovery of archaeal descendants of the elusive host of the mitochondrial endosymbiont cannot be ruled out.

Addendum

Shortly after this manuscript was submitted, a game-changing discovery bearing on the archaeal ancestry of eukaryotes has been published [161,162]. Deep metagenomic sequencing uncovered a remarkable group of archaea from marine sludge that combined the two key properties expected of the eukaryotic ancestor. First, one of these novel organisms, tentatively classified as a new phylum Lokiarchaeota (already affectionately known as Loki), represents a sister group to eukaryotes, and the Loki–eukaryote branch is confidently lodged deep within the TACK superphylum. Second, the genome of Loki recapitulates with an uncanny precision the reconstructed gene repertoire of the putative archaeal ancestor of eukaryotes that is outlined above. In particular, Loki encode crenactins, homologues of eukaryotic gelsolins, the ESCRT-III complex, an expanded family of small Ras-like GTPases and the complete ubiquitin system. This gene repertoire translates into a confident prediction of a complex cytoskeleton and membrane remodelling systems and is compatible with a rudimentary phagocytic capability that has been predicted for the archaeal ancestor of eukaryotes. Further exploration of the genomes and hopefully the actual biology of the Loki are likely to dramatically enhance our understanding of eukaryogenesis.

Competing interests. I declare I have no competing interests.

Funding. The author’s research is supported by intramural funds of the US Department of Health and Human Services (to the National Library of Medicine, NIH).

References

- Lane N, Martin W. 2010 The energetics of genome complexity. *Nature* **467**, 929–934. (doi:10.1038/nature09486)
- Lane N. 2011 Energetics and genetics across the prokaryote–eukaryote divide. *Biol. Direct* **6**, 35. (doi:10.1186/1745-6150-6-35)
- Lane N. 2015 *The vital question: energy, evolution, and the origins of complex life*. London, UK: W.W. Norton & Company.
- Hudder A, Nathanson L, Deutscher MP. 2003 Organization of mammalian cytoplasm. *Mol. Cell Biol.* **23**, 9318–9326. (doi:10.1128/MCB.23.24.9318-9326.2003)
- Guigas G, Kalla C, Weiss M. 2007 The degree of macromolecular crowding in the cytoplasm and nucleoplasm of mammalian cells is conserved. *FEBS Lett.* **581**, 5094–5098. (doi:10.1016/j.febslet.2007.09.054)
- Mans BJ, Anantharaman V, Aravind L, Koonin EV. 2004 Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle* **3**, 1612–1637. (doi:10.4161/cc.3.12.1345)
- Collins L, Penny D. 2005 Complex spliceosomal organization ancestral to extant eukaryotes. *Mol. Biol. Evol.* **22**, 1053–1066. (doi:10.1093/molbev/msi091)
- Baptiste E, Charlebois RL, MacLeod D, Brochier C. 2005 The two tempos of nuclear pore complex evolution: highly adapting proteins in an ancient frozen structure. *Genome Biol.* **6**, R85. (doi:10.1186/gb-2005-6-10-r85)
- Makarova KS, Wolf YI, Mekhedov SL, Mirkin BG, Koonin EV. 2005 Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res.* **33**, 4626–4638. (doi:10.1093/nar/gki775)
- Schlacht A, Herman EK, Klute MJ, Field MC, Dacks JB. 2014 Missing pieces of an ancient puzzle: evolution of the eukaryotic membrane-trafficking

- system. *Cold Spring Harb. Perspect. Biol.* **6**, a016048. (doi:10.1101/cshperspect.a016048)
11. Kurland CG, Collins LJ, Penny D. 2006 Genomics and the irreducible nature of eukaryote cells. *Science* **312**, 1011–1014. (doi:10.1126/science.1121674)
 12. Martin W, Dagan T, Koonin EV, Dipippo JL, Gogarten JP, Lake JA. 2007 The evolution of eukaryotes. *Science* **316**, 542–543; author reply 542–543. (doi:10.1126/science.316.5824.542c)
 13. Woese CR, Kandler O, Wheelis ML. 1990 Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl Acad. Sci. USA* **87**, 4576–4579. (doi:10.1073/pnas.87.12.4576)
 14. Pace NR. 1997 A molecular view of microbial diversity and the biosphere. *Science* **276**, 734–740. (doi:10.1126/science.276.5313.734)
 15. Pace NR. 2006 Time for a change. *Nature* **441**, 289. (doi:10.1038/441289a)
 16. Pace NR. 2009 Mapping the tree of life: progress and prospects. *Microbiol. Mol. Biol. Rev.* **73**, 565–576. (doi:10.1128/MMBR.00033-09)
 17. Cavalier-Smith T. 1993 Kingdom protozoa and its 18 phyla. *Microbiol. Rev.* **57**, 953–994.
 18. Patterson DJ. 1999 The diversity of eukaryotes. *Am. Nat.* **154**, S96–S124. (doi:10.1086/303287)
 19. Roger AJ. 1999 Reconstructing early events in eukaryotic evolution. *Am. Nat.* **154**, S146–S163. (doi:10.1086/303290)
 20. Cavalier-Smith T. 2002 The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int. J. Syst. Evol. Microbiol.* **52**, 297–354.
 21. Gray MW, Lang BF, Burger G. 2004 Mitochondria of protists. *Annu. Rev. Genet.* **38**, 477–524. (doi:10.1146/annurev.genet.37.110801.142526)
 22. van der Giezen M. 2009 Hydrogenosomes and mitosomes: conservation and evolution of functions. *J. Eukaryot. Microbiol.* **56**, 221–231. (doi:10.1111/j.1550-7408.2009.00407.x)
 23. Embley TM, Martin W. 2006 Eukaryotic evolution, changes and challenges. *Nature* **440**, 623–630. (doi:10.1038/nature04546)
 24. Shiflett AM, Johnson PJ. 2010 Mitochondrion-related organelles in eukaryotic protists. *Annu. Rev. Microbiol.* **64**, 409–429. (doi:10.1146/annurev.micro.62.081307.162826)
 25. Katz LA. 1998 Changing perspectives on the origin of eukaryotes. *Trends Ecol Evol.* **13**, 493–497. (doi:10.1016/S0169-5347(98)01490-6)
 26. Gray MW, Burger G, Lang BF. 1999 Mitochondrial evolution. *Science* **283**, 1476–1481. (doi:10.1126/science.283.5407.1476)
 27. Lang BF, Gray MW, Burger G. 1999 Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* **33**, 351–397. (doi:10.1146/annurev.genet.33.1.351)
 28. Poole A, Penny D. 2007 Eukaryote evolution: engulfed by speculation. *Nature* **447**, 913. (doi:10.1038/447913a)
 29. Poole AM, Penny D. 2007 Evaluating hypotheses for the origin of eukaryotes. *Bioessays* **29**, 74–84. (doi:10.1002/bies.20516)
 30. Martin W, Muller M. 1998 The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41. (doi:10.1038/32096)
 31. Lane N, Martin WF. 2012 The origin of membrane bioenergetics. *Cell* **151**, 1406–1416. (doi:10.1016/j.cell.2012.11.050)
 32. Esser C *et al.* 2004 A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol. Biol. Evol.* **21**, 1643–1660. (doi:10.1093/molbev/msh160)
 33. Rivera MC, Lake JA. 2004 The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* **431**, 152–155. (doi:10.1038/nature02848)
 34. Pisani D, Cotton JA, McInerney JO. 2007 Supertrees disentangle the chimerical origin of eukaryotic genomes. *Mol. Biol. Evol.* **24**, 1752–1760. (doi:10.1093/molbev/msm095)
 35. Thiergart T, Landan G, Schenk M, Dagan T, Martin WF. 2012 An evolutionary network of genes present in the eukaryote common ancestor polls genomes on eukaryotic and mitochondrial origin. *Genome Biol. Evol.* **4**, 466–485. (doi:10.1093/gbe/evs018)
 36. Martin W, Koonin EV. 2006 Introns and the origin of nucleus–cytosol compartmentation. *Nature* **440**, 41–45. (doi:10.1038/nature04531)
 37. Koonin EV. 2006 The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biol. Direct* **1**, 22. (doi:10.1186/1745-6150-1-22)
 38. Cavalier-Smith T. 2009 Predation and eukaryote cell origins: a coevolutionary perspective. *Int. J. Biochem. Cell Biol.* **41**, 307–322. (doi:10.1016/j.biocel.2008.10.002)
 39. von Dohlen CD, Kohler S, Alsop ST, McManus WR. 2001 Mealybug beta-proteobacterial endosymbionts contain gamma-proteobacterial symbionts. *Nature* **412**, 433–436. (doi:10.1038/35086563)
 40. McCutcheon JP, Moran NA. 2012 Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**, 13–26. (doi:10.1038/nrmicro2670)
 41. Lake JA. 1988 Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. *Nature* **331**, 184–186. (doi:10.1038/331184a0)
 42. Lake JA, Henderson E, Oakes M, Clark MW. 1984 Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *Proc. Natl Acad. Sci. USA* **81**, 3786–3790. (doi:10.1073/pnas.81.12.3786)
 43. Lake JA. 1998 Optimally recovering rate variation information from genomes and sequences: pattern filtering. *Mol. Biol. Evol.* **15**, 1224–1231. (doi:10.1093/oxfordjournals.molbev.a026030)
 44. Rivera MC, Lake JA. 1992 Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* **257**, 74–76. (doi:10.1126/science.1621096)
 45. Vishwanath P, Favaretto P, Hartman H, Mohr SC, Smith TF. 2004 Ribosomal protein-sequence block structure suggests complex prokaryotic evolution with implications for the origin of eukaryotes. *Mol. Phylogenet. Evol.* **33**, 615–625. (doi:10.1016/j.ympev.2004.07.003)
 46. Cox CJ, Foster PG, Hirt RP, Harris SR, Embley TM. 2008 The archaeobacterial origin of eukaryotes. *Proc. Natl Acad. Sci. USA* **105**, 20356–20361. (doi:10.1073/pnas.0810647105)
 47. Lopez-Garcia P, Moreira D. 2006 Selective forces for the origin of the eukaryotic nucleus. *Bioessays* **28**, 525–533. (doi:10.1002/bies.20413)
 48. Horiike T, Hamada K, Miyata D, Shinozawa T. 2004 The origin of eukaryotes is suggested as the symbiosis of pyrococcus into gamma-proteobacteria by phylogenetic tree based on gene content. *J. Mol. Evol.* **59**, 606–619. (doi:10.1007/s00239-004-2652-5)
 49. Moreira D, Lopez-Garcia P. 1998 Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J. Mol. Evol.* **47**, 517–530. (doi:10.1007/PL00006408)
 50. Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P. 2006 Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**, 1283–1287. (doi:10.1126/science.1123061)
 51. Fukami-Kobayashi K, Minezaki Y, Tateno Y, Nishikawa K. 2007 A tree of life based on protein domain organizations. *Mol. Biol. Evol.* **24**, 1181–1189. (doi:10.1093/molbev/msm034)
 52. Hedges SB, Chen H, Kumar S, Wang DY, Thompson AS, Watanabe H. 2001 A genomic timescale for the origin of eukaryotes. *BMC Evol. Biol.* **1**, 4. (doi:10.1186/1471-2148-1-4)
 53. Tekaia F, Yeramian E. 2005 Genome trees from conservation profiles. *PLoS Comput. Biol.* **1**, e75. (doi:10.1371/journal.pcbi.0010075)
 54. Forterre P. 1995 Thermoreduction, a hypothesis for the origin of prokaryotes. *C. R. Acad. Sci. III* **318**, 415–422.
 55. Poole A, Jeffares D, Penny D. 1999 Early evolution: prokaryotes, the new kids on the block. *Bioessays* **21**, 880–889. (doi:10.1002/(SICI)1521-1878(199910)21:10<880::AID-BIES11>3.0.CO;2-P)
 56. Collins LJ, Kurland CG, Biggs P, Penny D. 2009 The modern RNP world of eukaryotes. *J. Hered.* **100**, 597–604. (doi:10.1093/jhered/esp064)
 57. Hixon WG, Searcy DG. 1993 Cytoskeleton in the archaeobacterium *Thermoplasma acidophilum*? Viscosity increase in soluble extracts. *Biosystems* **29**, 151–160. (doi:10.1016/0303-2647(93)90091-P)
 58. Margulis L. 1996 Archaeal–eubacterial mergers in the origin of Eukarya: phylogenetic classification of life. *Proc. Natl Acad. Sci. USA* **93**, 1071–1076. (doi:10.1073/pnas.93.3.1071)
 59. Margulis L, Dolan MF, Guerrero R. 2000 The chimeric eukaryote: origin of the nucleus from the karyomastigont in amitochondriate protists. *Proc. Natl Acad. Sci. USA* **97**, 6954–6959. (doi:10.1073/pnas.97.13.6954)
 60. Margulis L, Stolz JF. 1984 Cell symbiosis theory: status and implications for the fossil record. *Adv. Space Res.* **4**, 195–201. (doi:10.1016/0273-1177(84)90562-3)
 61. Searcy DG, Stein DB, Green GR. 1978 Phylogenetic affinities between eukaryotic cells and a thermophilic mycoplasma. *Biosystems* **10**, 19–28. (doi:10.1016/0303-2647(78)90024-2)

62. Yutin N, Makarova KS, Mekhedov SL, Wolf YI, Koonin EV. 2008 The deep archaeal roots of eukaryotes. *Mol. Biol. Evol.* **25**, 1619–1630. (doi:10.1093/molbev/msn108)
63. Gribaldo S, Poole AM, Daubin V, Forterre P, Brochier-Armanet C. 2010 The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? *Nat. Rev. Microbiol.* **8**, 743–752. (doi:10.1038/nrmicro2426)
64. Elkins JG *et al.* 2008 A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc. Natl Acad. Sci. USA* **105**, 8102–8107. (doi:10.1073/pnas.0801980105)
65. Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P. 2008 Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat. Rev. Microbiol.* **6**, 245–252. (doi:10.1038/nrmicro1852)
66. Pester M, Schleper C, Wagner M. 2011 The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. *Curr. Opin. Microbiol.* **14**, 300–306. (doi:10.1016/j.mib.2011.04.007)
67. Lloyd KG *et al.* 2013 Predominant archaea in marine sediments degrade detrital proteins. *Nature* **496**, 215–218. (doi:10.1038/nature12033)
68. Brochier-Armanet C, Gribaldo S, Forterre P. 2012 Spotlight on the Thaumarchaeota. *ISME J.* **6**, 227–230. (doi:10.1038/ismej.2011.145)
69. Reigstad LJ, Jorgensen SL, Schleper C. 2010 Diversity and abundance of Korarchaeota in terrestrial hot springs of Iceland and Kamchatka. *ISME J.* **4**, 346–356. (doi:10.1038/ismej.2009.126)
70. Miller-Coleman RL, Dodsworth JA, Ross CA, Shock EL, Williams AJ, Hartnett HE, McDonald AI, Havig JR, Hedlund BP. 2012 Korarchaeota diversity, biogeography, and abundance in Yellowstone and Great Basin hot springs and ecological niche modeling based on machine learning. *PLoS ONE* **7**, e35964. (doi:10.1371/journal.pone.0035964)
71. Nunoura T *et al.* 2011 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. (doi:10.1093/nar/gkq1228)
72. Rinke C *et al.* 2013 Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437. (doi:10.1038/nature12352)
73. Castelle CJ *et al.* 2015 Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* **25**, 690–701. (doi:10.1016/j.cub.2015.01.014)
74. Eme L, Doolittle WF. 2015 Microbial diversity: a bonanza of phyla. *Curr. Biol.* **25**, R227–R230. (doi:10.1016/j.cub.2014.12.044)
75. Petitjean C, Deschamps P, Lopez-Garcia P, Moreira D. 2015 Rooting the domain archaea by phylogenomic analysis supports the foundation of the new kingdom Proteoarchaeota. *Genome Biol. Evol.* **7**, 191–204. (doi:10.1093/gbe/evu274)
76. Petitjean C, Deschamps P, Lopez-Garcia P, Moreira D, Brochier-Armanet C. 2015 Extending the conserved phylogenetic core of archaea disentangles the evolution of the third domain of life. *Mol. Biol. Evol.* **32**, 1242–1254. (doi:10.1093/molbev/msv015)
77. Podar M, Makarova KS, Graham DE, Wolf YI, Koonin EV, Reysenbach AL. 2013 Insights into archaeal evolution and symbiosis from the genomes of a nanoarchaeon and its inferred crenarchaeal host from Obsidian Pool, Yellowstone National Park. *Biol. Direct* **8**, 9. (doi:10.1186/1745-6150-8-9)
78. Koonin EV, Yutin N. 2014 The dispersed archaeal eukaryome and the complex archaeal ancestor of eukaryotes. *Cold Spring Harb. Perspect. Biol.* **6**, a016188. (doi:10.1101/cshperspect.a016188)
79. Nelson-Sathi S *et al.* 2014 Origins of major archaeal clades correspond to gene acquisitions from bacteria. *Nature* **517**, 77–80. (doi:10.1038/nature13805)
80. Guy L, Ettema TJ. 2011 The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol.* **19**, 580–587. (doi:10.1016/j.tim.2011.09.002)
81. Williams TA, Foster PG, Nye TM, Cox CJ, Embley TM. 2012 A congruent phylogenomic signal places eukaryotes within the Archaea. *Proc. R. Soc. B* **279**, 4870–4879. (doi:10.1098/rspb.2012.1795)
82. Martijn J, Ettema TJ. 2013 From archaeon to eukaryote: the evolutionary dark ages of the eukaryotic cell. *Biochem. Soc. Trans.* **41**, 451–457. (doi:10.1042/BST20120292)
83. Yutin N, Puigbo P, Koonin EV, Wolf YI. 2012 Phylogenomics of prokaryotic ribosomal proteins. *PLoS ONE* **7**, e36972. (doi:10.1371/journal.pone.0036972)
84. Wolf YI, Makarova KS, Yutin N, Koonin EV. 2012 Updated clusters of orthologous genes for Archaea: a complex ancestor of the Archaea and the byways of horizontal gene transfer. *Biol. Direct* **7**, 46. (doi:10.1186/1745-6150-7-46)
85. Williams TA, Embley TM. 2014 Archaeal 'dark matter' and the origin of eukaryotes. *Genome Biol. Evol.* **6**, 474–481. (doi:10.1093/gbe/evu031)
86. Guy L, Saw JH, Ettema TJ. 2014 The archaeal legacy of eukaryotes: a phylogenomic perspective. *Cold Spring Harb. Perspect. Biol.* **6**, a016022. (doi:10.1101/cshperspect.a016022)
87. Raymann K, Brochier-Armanet C, Gribaldo S. 2015 The two-domain tree of life is linked to a new root for the Archaea. *Proc. Natl Acad. Sci. USA* **112**, 6670–6675. (doi:10.1073/pnas.1420858112)
88. Makarova KS, Sorokin AV, Novichkov PS, Wolf YI, Koonin EV. 2007 Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea. *Biol. Direct* **2**, 33. (doi:10.1186/1745-6150-2-33)
89. Csuros M, Miklos I. 2009 Streamlining and large ancestral genomes in Archaea inferred with a phylogenetic birth-and-death model. *Mol. Biol. Evol.* **26**, 2087–2095. (doi:10.1093/molbev/msp123)
90. Wolf YI, Koonin EV. 2013 Genome reduction as the dominant mode of evolution. *BioEssays* **35**, 829–837. (doi:10.1002/bies.201300037)
91. Boussau B, Blanquart S, Necsulea A, Lartillot N, Gouy M. 2008 Parallel adaptations to high temperatures in the Archaeal eon. *Nature* **456**, 942–945. (doi:10.1038/nature07393)
92. Lasek-Nesselquist E, Gogarten JP. 2013 The effects of model choice and mitigating bias on the ribosomal tree of life. *Mol. Phylogenet. Evol.* **69**, 17–38. (doi:10.1016/j.ympev.2013.05.006)
93. Williams TA, Foster PG, Cox CJ, Embley TM. 2013 An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* **504**, 231–236. (doi:10.1038/nature12779)
94. Kelly S, Wickstead B, Gull K. 2011 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. (doi:10.1098/rspb.2010.1427)
95. Koonin EV, Makarova KS, Elkins JG. 2007 Orthologs of the small RPB8 subunit of the eukaryotic RNA polymerases are conserved in hyperthermophilic Crenarchaeota and 'Korarchaeota'. *Biol. Direct* **2**, 38. (doi:10.1186/1745-6150-2-38)
96. Blombach F, Makarova KS, Marrero J, Siebers B, Koonin EV, van der Oost J. 2009 Identification of an ortholog of the eukaryotic RNA polymerase III subunit RPC34 in Crenarchaeota and Thaumarchaeota suggests specialization of RNA polymerases for coding and non-coding RNAs in Archaea. *Biol. Direct* **4**, 39. (doi:10.1186/1745-6150-4-39)
97. Daniels JP, Kelly S, Wickstead B, Gull K. 2009 Identification of a crenarchaeal orthologue of Elf1: implications for chromatin and transcription in Archaea. *Biol. Direct* **4**, 24. (doi:10.1186/1745-6150-4-24)
98. Koonin EV, Makarova KS, Aravind L. 2001 Horizontal gene transfer in prokaryotes—quantification and classification. *Annu. Rev. Microbiol.* **55**, 709–742.
99. Kennedy SP, Ng WV, Salzberg SL, Hood L, DasSarma S. 2001 Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence. *Genome Res.* **11**, 1641–1650. (doi:10.1101/gr.190201)
100. Deppenmeier U *et al.* 2002 The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. *J. Mol. Microbiol. Biotechnol.* **4**, 453–461.
101. Nelson-Sathi S, Dagan T, Landan G, Janssen A, Steel M, McInerney JO, Deppenmeier U, Martin WF. 2012 Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea. *Proc. Natl Acad. Sci. USA* **109**, 20 537–20 542. (doi:10.1073/pnas.1209119109)
102. Lowe J, Amos LA. 2009 Evolution of cytomotive filaments: the cytoskeleton from prokaryotes to eukaryotes. *Int. J. Biochem. Cell Biol.* **41**, 323–329. (doi:10.1016/j.biocel.2008.08.010)
103. Aylett CH, Lowe J, Amos LA. 2011 New insights into the mechanisms of cytomotive actin and tubulin filaments. *Int. Rev. Cell Mol. Biol.* **292**, 1–71. (doi:10.1016/B978-0-12-386033-0.00001-3)
104. Cabeen MT, Jacobs-Wagner C. 2010 The bacterial cytoskeleton. *Annu. Rev. Genet.* **44**, 365–392. (doi:10.1146/annurev-genet-102108-134845)

105. Celler K, Koning RI, Koster AJ, van Wezel GP. 2013 Multidimensional view of the bacterial cytoskeleton. *J. Bacteriol.* **195**, 1627–1636. (doi:10.1128/JB.02194-12)
106. Bork P, Sander C, Valencia A. 1992 An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins. *Proc. Natl Acad. Sci. USA* **89**, 7290–7294. (doi:10.1073/pnas.89.16.7290)
107. Kabsch W, Holmes KC. 1995 The actin fold. *FASEB J.* **9**, 167–174.
108. Lowe J, Amos LA. 1998 Crystal structure of the bacterial cell-division protein FtsZ. *Nature* **391**, 203–206. (doi:10.1038/34472)
109. Nogales E, Downing KH, Amos LA, Lowe J. 1998 Tubulin and FtsZ form a distinct family of GTPases. *Nat. Struct. Biol.* **5**, 451–458. (doi:10.1038/nsb0698-451)
110. Yutin N, Wolf MY, Wolf YI, Koonin EV. 2009 The origins of phagocytosis and eukaryogenesis. *Biol. Direct* **4**, 9. (doi:10.1186/1745-6150-4-9)
111. Makarova KS, Yutin N, Bell SD, Koonin EV. 2010 Evolution of diverse cell division and vesicle formation systems in Archaea. *Nat. Rev. Microbiol.* **8**, 731–741. (doi:10.1038/nrmicro2406)
112. Ettema TJ, Lindas AC, Bernander R. 2011 An actin-based cytoskeleton in archaea. *Mol. Microbiol.* **80**, 1052–1061. (doi:10.1111/j.1365-2958.2011.07635.x)
113. Bernander R, Lind AE, Ettema TJ. 2011 An archaeal origin for the actin cytoskeleton: implications for eukaryogenesis. *Commun. Integr. Biol.* **4**, 664–667. (doi:10.4161/cib.16974)
114. Lindas AC, Bernander R. 2013 The cell cycle of archaea. *Nat. Rev. Microbiol.* **11**, 627–638. (doi:10.1038/nrmicro3077)
115. Lindas AC, Chruszcz M, Bernander R, Valegard K. 2014 Structure of crenactin, an archaeal actin homologue active at 90 degrees C. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 492–500. (doi:10.1107/S1399004714000935)
116. Izore T, Duman R, Kureisaite-Ciziene D, Lowe J. 2014 Crenactin from *Pyrobaculum calidifontis* is closely related to actin in structure and forms steep helical filaments. *FEBS Lett.* **588**, 776–782. (doi:10.1016/j.febslet.2014.01.029)
117. Goley ED, Welch MD. 2006 The ARP2/3 complex: an actin nucleator comes of age. *Nat. Rev. Mol. Cell Biol.* **7**, 713–726. (doi:10.1038/nrm2026)
118. Yutin N, Koonin EV. 2012 Archaeal origin of tubulin. *Biol. Direct* **7**, 10. (doi:10.1186/1745-6150-7-10)
119. Lebedeva EV *et al.* 2013 Enrichment and genome sequence of the group I.1a ammonia-oxidizing Archaeon ‘*Ca. Nitrosotenuis uzonensis*’ representing a clade globally distributed in thermal habitats. *PLoS ONE* **8**, e80835. (doi:10.1371/journal.pone.0080835)
120. Margolin W. 2009 Sculpting the bacterial cell. *Curr. Biol.* **19**, R812–R822. (doi:10.1016/j.cub.2009.06.033)
121. Adams DW, Errington J. 2009 Bacterial cell division: assembly, maintenance and disassembly of the Z ring. *Nat. Rev. Microbiol.* **7**, 642–653. (doi:10.1038/nrmicro2198)
122. Samson RY, Obita T, Freund SM, Williams RL, Bell SD. 2008 A role for the ESCRT system in cell division in archaea. *Science* **322**, 1710–1713. (doi:10.1126/science.1165322)
123. Lindas AC, Karlsson EA, Lindgren MT, Ettema TJ, Bernander R. 2008 A unique cell division machinery in the Archaea. *Proc. Natl Acad. Sci. USA* **105**, 18 942–18 946. (doi:10.1073/pnas.0809467105)
124. Samson RY, Bell SD. 2009 Ancient ESCRTs and the evolution of binary fission. *Trends Microbiol.* **17**, 507–513. (doi:10.1016/j.tim.2009.08.003)
125. Ettema TJ, Bernander R. 2009 Cell division and the ESCRT complex: A surprise from the archaea. *Commun. Integr. Biol.* **2**, 86–88. (doi:10.4161/cib.7523)
126. Pelve EA, Lindas AC, Martens-Habbena W, de la Torre JR, Stahl DA, Bernander R. 2011 Cdv-based cell division and cell cycle organization in the thaumarchaeon *Nitrosopumilus maritimus*. *Mol. Microbiol.* **82**, 555–566. (doi:10.1111/j.1365-2958.2011.07834.x)
127. Dobro MJ, Samson RY, Yu Z, McCullough J, Ding HJ, Chong PL, Bell SD, Jensen GJ. 2013 Electron cryotomography of ESCRT assemblies and dividing *Sulfolobus* cells suggests that spiraling filaments are involved in membrane scission. *Mol. Biol. Cell* **24**, 2319–2327. (doi:10.1091/mbc.E12-11-0785)
128. Burroughs AM, Iyer LM, Aravind L. 2012 Structure and evolution of ubiquitin and ubiquitin-related domains. *Methods Mol. Biol.* **832**, 15–63. (doi:10.1007/978-1-61779-474-2_2)
129. Burroughs AM, Iyer LM, Aravind L. 2012 The natural history of ubiquitin and ubiquitin-related domains. *Front. Biosci.* **17**, 1433–1460. (doi:10.2741/3996)
130. van der Veen AG, Ploegh HL. 2012 Ubiquitin-like proteins. *Annu. Rev. Biochem.* **81**, 323–357. (doi:10.1146/annurev-biochem-093010-153308)
131. Hochstrasser M. 2000 Evolution and function of ubiquitin-like protein-conjugation systems. *Nat. Cell Biol.* **2**, E153–E157. (doi:10.1038/35019643)
132. Maupin-Furlow JA. 2014 Prokaryotic ubiquitin-like protein modification. *Annu. Rev. Microbiol.* **68**, 155–175. (doi:10.1146/annurev-micro-091313-103447)
133. Humbard MA, Miranda HV, Lim JM, Krause DJ, Pritz JR, Zhou G, Chen S, Wells L, Maupin-Furlow JA. 2010 Ubiquitin-like small archaeal modifier proteins (SAMPs) in *Haloferax volcanii*. *Nature* **463**, 54–60. (doi:10.1038/nature08659)
134. Miranda HV, Nembhard N, Su D, Hepowit N, Krause DJ, Pritz JR, Phillips C, Soll D, Maupin-Furlow JA. 2011 E1- and ubiquitin-like proteins provide a direct link between protein conjugation and sulfur transfer in archaea. *Proc. Natl Acad. Sci. USA* **108**, 4417–4422. (doi:10.1073/pnas.1018151108)
135. Maupin-Furlow JA. 2013 Ubiquitin-like proteins and their roles in archaea. *Trends Microbiol.* **21**, 31–38. (doi:10.1016/j.tim.2012.09.006)
136. Grau-Bove X, Sebe-Pedros A, Ruiz-Trillo I. 2015 The eukaryotic ancestor had a complex ubiquitin signaling system of archaeal origin. *Mol. Biol. Evol.* **32**, 726–739. (doi:10.1093/molbev/msu334)
137. Makarova KS, Koonin EV. 2010 Archaeal ubiquitin-like proteins: functional versatility and putative ancestral involvement in tRNA modification revealed by comparative genomic analysis. *Archaea* **2010**, 710303. (doi:10.1155/2010/710303)
138. Wang F, Liu M, Qiu R, Ji C. 2011 The dual role of ubiquitin-like protein Urm1 as a protein modifier and sulfur carrier. *Protein Cell* **2**, 612–619. (doi:10.1007/s13238-011-1074-6)
139. Van der Veen AG, Schorpp K, Schlieker C, Buti L, Damon JR, Spooner E, Ploegh HL, Jentsch S. 2011 Role of the ubiquitin-like protein Urm1 as a noncanonical lysine-directed protein modifier. *Proc. Natl Acad. Sci. USA* **108**, 1763–1770. (doi:10.1073/pnas.1014402108)
140. Cerutti H, Casas-Mollano JA. 2006 On the origin and functions of RNA-mediated silencing: from protists to man. *Curr. Genet.* **50**, 81–99. (doi:10.1007/s00294-006-0078-x)
141. Shabalina SA, Koonin EV. 2008 Origins and evolution of eukaryotic RNA interference. *Trends Ecol. Evol.* **23**, 578–587. (doi:10.1016/j.tree.2008.06.005)
142. Carthew RW, Sontheimer EJ. 2009 Origins and mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–655. (doi:10.1016/j.cell.2009.01.035)
143. Burroughs AM, Ando Y, Aravind L. 2013 New perspectives on the diversification of the RNA interference system: insights from comparative genomics and small RNA sequencing. *Wiley Interdiscip. Rev. RNA* **5**, 141–182. (doi:10.1002/wrna.1210)
144. Tijsterman M, Plasterk RH. 2004 Dicers at RISC; the mechanism of RNAi. *Cell* **117**, 1–3. (doi:10.1016/S0092-8674(04)00293-4)
145. Jaskiewicz L, Filipowicz W. 2008 Role of Dicer in posttranscriptional RNA silencing. *Curr. Top. Microbiol. Immunol.* **320**, 77–97.
146. Kidwell MA, Chan JM, Doudna JA. 2014 Evolutionarily conserved roles of the Dicer helicase domain in regulating RNA interference processing. *J. Biol. Chem.* **289**, 28 352–28 362. (doi:10.1074/jbc.M114.589051)
147. Gao Z, Wang M, Blair D, Zheng Y, Dou Y. 2014 Phylogenetic analysis of the endoribonuclease Dicer family. *PLoS ONE* **9**, e95350. (doi:10.1371/journal.pone.0095350)
148. Tolia NH, Joshua-Tor L. 2007 Slicer and the Argonautes. *Nat. Chem. Biol.* **3**, 36–43. (doi:10.1038/nchembio848)
149. Hock J, Meister G. 2008 The Argonaute protein family. *Genome Biol.* **9**, 210. (doi:10.1186/gb-2008-9-2-210)
150. Court DL, Gan J, Liang YH, Shaw GX, Tropea JE, Costantino N, Waugh DS, Ji X. 2013 RNase III: Genetics and function; structure and mechanism. *Annu. Rev. Genet.* **47**, 405–431. (doi:10.1146/annurev-genet-110711-155618)
151. Lestini R, Duan Z, Allers T. 2010 The archaeal Xpf/Mus81/FANCM homolog Hef and the Holliday junction resolvase Hjc define alternative pathways that are essential for cell viability in *Haloferax volcanii*. *DNA Repair (Amst)* **9**, 994–1002. (doi:10.1016/j.dnarep.2010.06.012)

152. Makarova KS, Wolf YI, van der Oost J, Koonin EV. 2009 Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol. Direct* **4**, 29. (doi:10.1186/1745-6150-4-29)
153. Swarts DC, Makarova K, Wang Y, Nakanishi K, Ketting RF, Koonin EV, Patel DJ, van der Oost J. 2014 The evolutionary journey of Argonaute proteins. *Nat. Struct. Mol. Biol.* **21**, 743–753. (doi:10.1038/nsmb.2879)
154. Olovnikov I, Chan K, Sachidanandam R, Newman DK, Aravin AA. 2013 Bacterial Argonaute samples the transcriptome to identify foreign DNA. *Mol. Cell* **51**, 594–605. (doi:10.1016/j.molcel.2013.08.014)
155. Swarts DC *et al.* 2014 DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* **507**, 258–261. (doi:10.1038/nature12971)
156. Hur JK, Olovnikov I, Aravin AA. 2014 Prokaryotic Argonautes defend genomes against invasive DNA. *Trends Biochem. Sci.* **39**, 257–259. (doi:10.1016/j.tibs.2014.04.006)
157. Iyer LM, Koonin EV, Aravind L. 2003 Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. *BMC Struct. Biol.* **3**, 1. (doi:10.1186/1472-6807-3-1)
158. Hurley JH. 2010 The ESCRT complexes. *Crit. Rev. Biochem. Mol. Biol.* **45**, 463–487. (doi:10.3109/10409238.2010.502516)
159. Esser C, Martin W, Dagan T. 2007 The origin of mitochondria in light of a fluid prokaryotic chromosome model. *Biol. Lett.* **3**, 180–184. (doi:10.1098/rsbl.2006.0582)
160. Ku C, Nelson-Sathi S, Roettger M, Garg S, Hazkani-Covo E, Martin WF. In press. Endosymbiotic gene transfer from prokaryotic pangenomes: inherited chimerism in eukaryotes. *Proc. Natl Acad. Sci. USA.* (doi:10.1073/pnas.1421385112)
161. Spang A *et al.* 2015 Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179. (doi:10.1038/nature14447)
162. Embley TM, Williams TA. 2015 Evolution: steps on the road to eukaryotes. *Nature* **521**, 169–170. (doi:10.1038/nature14522)