PERSPECTIVES

The hybrid nature of the Eukaryota and a consilient view of life on Earth

James O. McInerney, Mary J. O'Connell and Davide Pisani

Abstract | The origin of the eukaryotic cell, which is known as eukaryogenesis, has puzzled scientists for more than 100 years, and many hypotheses have been proposed. Recent analyses of new data enable the safe elimination of some of these hypotheses, whereas support for other hypotheses has increased. In this Opinion article, we evaluate the available theories for their compatibility with empirical observations and conclude that cellular life consists of two primary, paraphyletic prokaryotic groups and one secondary, monophyletic group that has symbiogenic origins — the eukaryotes.

In 1905, Konstantin Mereschkowski postulated a symbiotic origin of plastids and animal cells[1], and in 1923, Ivan Wallin argued for a bacterial origin of the mitochondrion[2]. Since then, dozens of scenarios have been put forward to explain eukaryogenesis[3–10]. How then should the different lines of evidence be viewed and how should their contributions to the arguments be weighted?

Although many readers will be familiar with Karl Popper[11], who advocated falsifiability — that is, generating, testing and rejecting hypotheses — as the framework for generating knowledge, the framework by which we understand evolution to be a strong theory is that of consilience[12]. Consilience is said to be the 'jumping together' of facts or, in other words, the analysis of whether there is agreement across different classes of evidence, thereby amounting to a single unifying theory[12]. For example, evolution is a strong theory because there is consilience between empirical observations from genetics, palaeontology, chemistry, physics and behaviour, and together, these observations support an overall theory of evolution.

In this Opinion article, we argue that hypothesis testing and inductive reasoning are consilient with a symbiogenic hypothesis for the origin of the Eukaryota. We address eukaryogenesis from four perspectives — molecular phylogenetics, palaeontology,

bioenergetics, and modern cell biology and biochemistry — each of which has contributed important and surprisingly congruent insights relating to this argument. It is the consilience of these lines of evidence that leads us to conclude that eukaryotes are not a primary lineage of life; rather, they are a relatively late innovation, are perhaps younger than any of the two prokaryotic groups and emerged from both of them. Indeed, the 'eukaryotes-late' hypotheses imply that there was a merger between an archaebacterium and a eubacterium.

A recent report focusing on the evidence that supports a paraphyletic organization of the Archaebacteria[13] has indicated that life is divided into two domains. In this Opinion article, we explore the evidence for eukaryotic origins for all parts of the eukaryotic cell.

Background: the proposed models

The origin of eukaryotes is intimately intertwined with the study of endosymbiosis, particularly with the origin of the mitochondrion. Various endosymbiotic hypotheses for eukaryogenesis have been suggested, and specific prokaryotic groups have been proposed as the sister groups of the mitochondrion and the plastid[4,7,8,14–17]. The origin of the mitochondrion is key to distinguishing between alternative hypotheses regarding the origins of eukaryotes. Both endosymbiotic[4,14] and autogenous[18,19] models

of eukaryogenesis have been proposed. There is substantial molecular support for endosymbiosis[4,14], and the mechanism (that is, selective advantage) for this model has also been explored[6,9,20–22]. By contrast, autogenous models, which imply that eukaryotes arose by stepwise mutations without the external acquisition of genetic material, require an excessive amount of *ad hoc* explanation[23]. In addition, as it has been found that amitochondriate eukaryotes, such as microsporidia[24] and *Giardia* spp.[25], are not primitively amitochondriate but are instead derived from mitochondrion-bearing ancestors, the evidence suggests that acquisition of the mitochondrion occurred early in eukaryotic evolution and certainly predates the origin of all known extant eukaryotes.

The details of the process of mitochondrial acquisition are still debated. Much of our understanding of the nature of life on Earth and the deepest philosophical ramifications of this knowledge depend almost entirely on a precise understanding of the origin of the mitochondrion[4,14,22,26]. One scenario that is commonly seen in textbooks[27] describes Eukaryota as a lineage that separated from the Archaebacteria before the diversification of the Archaebacteria. In this scenario, mitochondrial origins merely correspond to the introgression[28] of a symbiont into an otherwise well-defined eukaryotic lineage. This is known as the three-domains hypothesis[27] (FIG. 1a). The three-domains tree recognizes Eubacteria, Archaebacteria and Eukaryota as three primary, monophyletic lineages of ancient origins. The rooting of this tree, which was achieved using deeply diverging paralogues[29], identified Eukaryota as the sister lineage to the Archaebacteria. An alternative rooting of the three-domains tree places the Eukaryota as the sister lineage to a monophyletic prokaryotic group (which consists of the Archaebacteria and the Eubacteria)[5,30]. This is known as the 'eukaryotes-early' hypothesis (FIG. 1b). In this scenario, an ancient complex eukaryotic cell underwent various reduction events to give rise to prokaryotes[5]; no recent phylogenetic analysis has provided convincing support for this hypothesis. Other alternatives to the three-domains tree include a phagotrophic hypothesis[18], hypotheses that suggest an

a | Three-domains hypothesis

b | Eukaryotes-early hypothesis

c | Eocyte hypothesis

d | Ring of life hypothesis
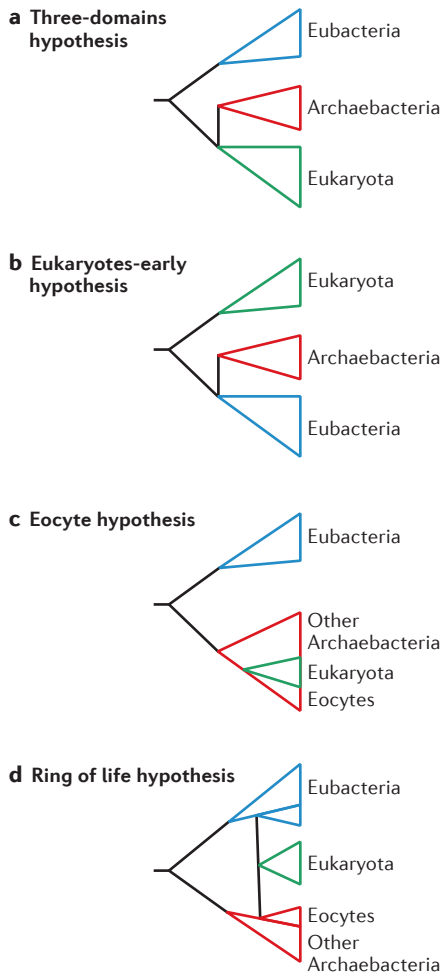
Figure 1 | **Competing scenarios for the origin of the eukaryotes and the highest-level structure for describing the diversity of cellular life. a** | The 'three-domains' phylogenetic tree, which is also known as the three-domains tree of life. **b** | The 'eukaryotes-early' hypothesis implies that prokaryotes are derived from a eukaryotic ancestor. **c** | The Eocyte hypothesis postulates that eukaryotes are derived from a lineage of the Archaebacteria that post-dates the diversification of the Archaebacteria. **d** | The 'ring of life' hypothesis postulates that eukaryotes arose from two lineages of prokaryotes, one of which was archaebacterial and one of which was eubacterial, and consequently, it is the only monophyletic group, although with symbiogenic origins. The ring of life hypothesis also implies that eukaryotes are late to arise and are a secondary grouping of life, whereas the two prokaryote groups are primary groupings.

involvement of viruses[31] and hypotheses that suggest a role for Planctomycetes as intermediate forms between prokaryotes and eukaryotes[19]. No convincing evidence supports these hypotheses, and the involvement of Planctomycetes has been refuted[23].

Other autogenous theories include the gradual evolution of eukaryotes from within the Archaebacteria[32] or from the Eubacteria, with some lateral gene transfer (LGT) from the Archaebacteria[33].

Current, viable alternatives to autogeneous models are represented by a set of hypotheses in which the eukaryotic lineage was established as a consequence of a merger in which two (or more) prokaryotes either fused or entered a process of symbiosis[7–9,15,20,21,34,35]. These hypotheses differ in the number of mergers they assume, the phylogenetic affinities of the mergers and the order in which the mergers occurred. Nevertheless, they all imply that the Eukaryota are not a primary lineage of life, and hence — in contrast to the three-domains hypothesis — they do not entail the existence of a proto-eukaryotic lineage as old as the Archaebacteria.

It is important to note that not all hypotheses that propose the existence of only two primary lineages of life are equally well supported. Although many homologues are shared by Eukaryota and Archaebacteria, the observed contribution to eukaryotic genomes from specifically hydrogen-dependent archaebacteria is limited, which weakens support for the hydrogen hypothesis[20] and for sulphur-syntrophy hypotheses[21]. Furthermore, the contributions from Delta-proteobacteria[7] and Spirochaetes[15,35] are not significant in number[17] and are probably limited to individual LGTs. This effectively argues against the serial endosymbiotic theory (SET) of Margulis[15,35] and the syntrophy hypothesis of López–García and Moreira[7]. Considering that the ancestor of the mitochondrion was a eubacterium and that it must have engaged in pre- and post-symbiotic LGTs[17], comparative genomic evidence can be interpreted, parsimoniously and confidently, as showing that only two prokaryotes had a role in the origin of the eukaryotes — an archaebacterium and an alphaproteobacterium[8].

A particular perspective on the origin of eukaryotes proposed by Lake *et al.*[36], which is known as the Eocyte hypothesis, suggests that the Eukaryota emerged from within the Archaebacteria. A broader hypothesis of greater consequence, which is known as the ring of life hypothesis, was more recently proposed by Rivera and Lake[9]. This hypothesis postulates that eukaryotes emerged from an already diversified archaebacterial group — the Eocytes (FIG. 1c,d) — that merged with a eubacterium. The authors do not explicitly say whether this eubacterium was distinct from the ancestor of

the mitochondrion or identical to it. In its original definition[36], Eocyte simply referred to those organisms that are a sister group to the Eukaryota within the Archaebacteria. The ring of life hypothesis argues that the Eukaryota emerged from within the Archaebacteria[8,9,34,37] at the same time as from within the Eubacteria and that the Eocytes are the archaebacterial sister group of the Eukaryota.

Regardless of the placement of the root of life, whether it is within the Eubacteria[38] or between the Eubacteria and Archaebacteria[29,39], the ring of life and the three-domains tree are mutually incompatible. Indeed, the ring of life type of model has a profound implication for how we view life, as it suggests that the Eukaryota are not a primary lineage but are a secondary lineage of life and that the Archaebacteria and Eubacteria are not monophyletic (FIG. 2). The first step towards understanding the origins of eukaryotes is therefore to clarify whether the current evidence supports the Eocyte–eukaryote grouping, which is compatible with the ring of life, or the three-domains grouping, which is not.

Evidence from phylogenetic analyses

The phylogenetic relationships of eukaryotes have been analysed using both genome-scale data sets[3,8,39] and smaller sets of genes (varying from one to approximately 85 genes, and often encoding ribosomal proteins) that are assumed to be 'genealogy-defining' (REFS 27,37,40,41). Woese originally proposed ribosomal RNA genes to be the ideal molecules for inferring phylogenies[27], and later studies broadened the list of preferred genes. However, a recent analysis has shown that even these genealogy-defining genes undergo a constant rate of horizontal transfer[42], and the use of such a small collection of genes has been criticized for only being able to recover a "tree of one percent" (REF. 43). Of particular interest are the methods of analysis that are used in different studies and the different perspectives that are given by different approaches.

Many studies have addressed the phylogenetic relationships among all cellular life forms, but, as these studies often disagree with each other[44,45], some authors have claimed that we have reached a 'phylogenetic impasse' (REF. 46). However, a critical assessment of these studies shows that there is no impasse at all. Genome-scale data sets have shown that eukaryotic genes manifest in only three important placements: one that groups eukaryotes with the Cyanobacteria (which indicates the symbiotic origin of

the plastid), one that associates the eukaryotes with the Alphaproteobacteria (which indicates the symbiotic origin of the mitochondrion) and one that nests the eukaryotes within the Archaebacteria (which indicates the archaebacterial origin of the host cell)[8]. The three-domains tree has been generally — although not universally — recovered using data sets of pre-selected genealogy-defining genes[40,44,45]. In addition, the three-domains tree is frequently recovered when long concatenations of genes that are affiliated with the Archaebacteria are analysed using methods that are more easily swayed by tree-reconstruction artefacts[16,42,47]. Eukaryotes are subtended by a long branch (as can be seen in small subunit ribosomal RNA (SSU rRNA) trees), indicating a rapid evolutionary rate in the stem eukaryotic lineage, and can be attracted by the highly divergent Eubacteria[30] via long-branch attraction. Initial analyses of nuclear SSU rRNA genes supported the three-domains tree[48], as did analyses of the eukaryotic transcription and translation machinery[44]. By contrast, analysis of ribosome structural features[36], an insertion in the gene encoding elongation factor Tu (EF-Tu)[49] and further analyses of nuclear SSU rRNA data[45] provided support for the Eocyte hypothesis.

In phylogenetics, the substitution model is a major determinant of tree topology[50]. Initial analyses of SSU and large subunit (LSU) rRNA genes used simplistic substitution models (or no model at all) and provided support for the three-domains tree using neighbour joining, maximum parsimony and evolutionary parsimony[44,48]. This cemented the three-domains tree as the accepted interpretation of the data. However, these studies — even when they used a model — unrealistically assumed the homogeneity of both nucleotide composition and substitution rate across different positions in their alignments and lineages. Later applications of heterogeneous models that focused on modelling among-site rate variation (ASRV) and compositional and rate variation across sites and along the tree[45,51,52] obtained very different results and instead supported an Eocyte topology. The use of a node discrete compositional heterogeneity model (NDCH model)[53], the CAT mixture model[54] and the correction for ASRV applied to a genealogy-defining gene set, as well as to a novel data set of 51 proteins, by contrast, supported the extended-Eocyte tree (FIG. 1c) over the three-domains tree[51,52] (FIG. 1a). Here, extended-Eocyte tree refers to the inclusion of new lineages of archaebacteria[55] that were unknown at the

time that the Eocyte hypothesis was initially formulated. Posterior-predictive simulations showed that heterogeneous models fit the data significantly better than homogeneous alternatives, which strengthens the support for the extended-Eocyte topology[51]. Further development of these models included the incorporation of node discrete rate heterogeneity (NDRH)[52], accommodating across-site compositional heterogeneity and across-tree compositional and rate-matrix heterogeneity. Interestingly, a taxonomically extended data set that included a further seven crenarchaeotes (in addition to the three that had already been published by Cox *et al.*[51]) was analysed using homogeneous models and maximum parsimony and provided support for the three-domains tree[52]. However, Bayes-factor analyses showed a significant improvement of the fit of the model to the data when additional parameters that were needed to account for the covarion and heterogeneous processes across the data were used. Using these better-fitting models, the Eocyte tree was recovered[52]. These studies are compelling, as they concomitantly show that, as the fit of the model to the data is improved, a shift from the three-domains to the Eocyte tree is observed and that, when simplistic approaches that are more likely to be swayed

---

Glossary

Among-site rate variation
A term used to describe the fact that different nucleotide or amino acid positions in a molecular sequence can change at different rates. This is usually caused by variation in selection pressure, although it may also be caused by variation in mutation rate.

CAT mixture model
A phylogenetic model that assumes that alignment positions might evolve according to different processes. The number of processes, the equilibrium frequencies of amino acids and the assignation of sites to categories are all free parameters of this model.

Eocyte hypothesis
A hypothesis relating to nucleocytosolic genes in eukaryotes that evolutionary history has placed within the Archaebacteria in a sister-group relationship with 'Eocyte' archaebacteria. This Eocyte hypothesis implies that the Archaebacteria are not a monophyletic taxon and that eukaryotes arose from within the Archaebacteria.

Evolutionary parsimony
(Also known as 'Lakes invariants'). A method of resolving a four-taxon problem. A set of equations calculate three invariants — one for each topology. For the correct phylogenetic tree, the invariant is non-zero, whereas the invariant is zero for the other two topologies.

Heterogeneous models
Phylogenetic models that allow for different evolutionary processes in different parts of the data and evolutionary

history. Heterogeneity can be allowed in evolutionary rates and sequence composition.

Homogeneous model
A phylogenetic model that implies that there has been a constant rate of evolution, a constant sequence composition, or both, for the duration of the evolutionary history of the sequences that are under consideration.

Maximum parsimony
A method that prefers phylogenetic trees that minimize the number of substitutions required to explain the observed distribution of character states in a data set.

Neighbour joining
A fast clustering approach to infer phylogenetic trees based on distance matrices that have been derived from alignments.

Node discrete compositional heterogeneity model
A model that allows different branches of a phylogenetic tree to evolve using different sequence compositions. Composition vectors are distributed throughout the tree and their placement is calculated as part of the optimization process.

Nucleomorph
A vestigial eukaryotic nucleus thought to be descended from algae that were themselves engulfed by other eukaryotes. Therefore, nucleomorphs represent secondary endosymbiotic events.

Ring of life hypothesis
A proposal that eukaryotic genomes are composed of genes that have two separate sources — one from within the Eubacteria and one from within the Archaebacteria — effectively creating a ring of life, not a tree of life.

Sequence similarity network
A network that consists of nodes that can either represent genes or genomes; the edges that connect these nodes are statements of homology. Therefore, the basic unit of a gene similarity network is a pair of nodes connected by an edge; however, these networks can be very large, consisting of connected components that embed thousands, or even millions, of nodes.

Symbiogenic
A term used to describe the merging of two separate organisms to form one new organism.

TACK group
A group of Archaebacteria that consists of four smaller groups (equivalent to phyla); these are the Thaumarchaeota, the Aigarchaeota, the Crenarchaeota and the Korarchaeota.

Three-domains hypothesis
A hypothesis that depicts the Eukaryota, the Eubacteria and the Archaebacteria as three monophyletic groups, generating a three-domains tree of life. This topology is generally recovered using nucleocytosolic informational proteins and homogeneous models of sequence evolution.
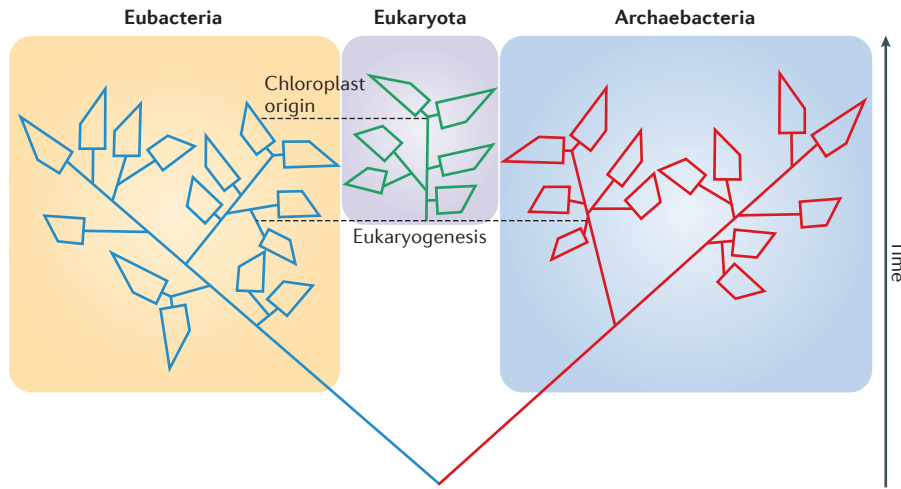

Figure 2 | **The ring of life hypothesis.** Schematic representation of the flow of genetic material from the two major prokaryotic groups into the base of the eukaryotes and the separate flow of genetic material from cyanobacteria into plastid-containing eukaryotes.

by tree-reconstruction artefacts are used, the three-domains tree is recovered. The maximum parsimony analysis is an elegant counter experiment that undermines the credibility of the three-domains tree considerably.

Contrary to the opinion of Gribaldo et al.[46], we conclude that we have not reached a phylogenetic impasse, and instead, the extended-Eocyte tree has constantly been gaining support, whereas support for the three-domains tree is diminishing. The Eocyte taxon is currently considered to most probably be synonymous with the TACK group[55], although new prokaryotes are constantly being discovered[16] and the true Eocyte group might still be unknown. Irrespective of this, it is clear that the extended-Eocyte topology for the relationship between Eukaryota and Archaebacteria should be the current standard, rather than the three-domains tree. The question therefore arises: how should we test the extended-Eocyte hypothesis? From a phylogenetic perspective, the answer is obvious. Given the relevance of dense taxonomic sampling and the importance of using well-fitting models to counter systematic biases and ensure phylogenetic accuracy, further studies should aim at extending sampling both within and outside the Eukaryota and analysing these extended data sets using well-fitting models. Interestingly, the recent study of Rinke and collaborators[16] substantially extended taxonomic sampling in the context of an analysis of 38 protein-coding genes and rRNA sequences, and they found support for the three-domains tree. However, poorly fitting models were used in the

analyses. ASRV was modelled using the CAT approximation[56], which should not be confused with the well-fitting, site-heterogeneous CAT mixture model that is implemented in the PhyloBayes software[54]. In addition, the outdated JTT model of amino acid substitution was used for the protein data set. Furthermore, a mixture of nuclear, mitochondrial and plastid sequences was used to represent eukaryotes[47]. Accordingly, the support for the three-domains tree that was generated by this study should be considered dubious. Indeed, Williams and Embley have recently shown that, when these problems are accounted for, the data set of Rinke and collaborators[16] does not support the three-domains hypothesis and instead supports the extended-Eocyte tree[47].

Specific relationships between eukaryotes and archaebacteria have not solely been inferred by modelling of the phylogenetic history of informational proteins. Phylogenetic analysis of membrane remodelling and cell division proteins across all domains of life has also revealed a complex pattern of homologues, with a eubacterial-like *ftsZ* homologue being absent in the Crenarchaeota, but present in most other archaebacteria. Concomitantly, the Sulfolobales and the Desulfurococcales use the ESCRT-III system for cell division, which is homologous to the system that is used by eukaryotes[57]. These similarities in gene content between the Eukaryota and the Crenarchaeota and between the Euryarchaeota and the Eubacteria may reflect lineage history, although LGT or other complications cannot be ruled out.

Evidence from the fossil record

The Earth is about 4.5 billion years old, and the evidence for life on Earth dates back almost to this date (~3.86 billion years ago). The earliest evidence for life is found in rocks from the Isua supergroup in Greenland and is based on inclusions that are enriched in light isotopic carbon[58]. The earliest possible evidence for cells in the fossil record is from the Apex Chert in the Pilbara formation of Western Australia (which is ~3.4 billion years old)[59]. To which extant organism lineage these fossils relate (if any) is not known, but what can be taken for granted is that they do not have any of the morphological features that characterize the eukaryotes. Indeed, the first confirmed eukaryotic fossil (*Shuiyousphaeridium macroreticulatum*) is from the Ruyang group, China, and is dated at ~1.6–1.8 billion years old[60]. The presence of eukaryotic biomarkers (specifically, steranes, which are presumed to be fossil derivatives of eukaryotic-specific sterols) in 2.7 billion-year-old bitumens from the Pilbara formation[61], Western Australia, had been interpreted as evidence that extends the fossil history of the eukaryotes to almost 3 billion years ago. If correct, this would make the eukaryotes almost as old as the prokaryotes. However, this evidence was suggested to be invalid by further studies showing that steranes could have percolated in the 2.7 billion-year-old bitumens from younger strata[62,63].

The three-domains tree implies that the eukaryotic lineage is at least as old as the Archaebacteria (which is inferred to be its sister group). If the three-domains tree is rooted in such a way as to represent the eukaryotes-early hypothesis, then the eukaryotic lineage is even older than the prokaryotes. Such early-stem eukaryotes — if they existed — were unlikely to possess all of the morphological features that are seen in extant (that is, crown) eukaryotes. They could certainly not have possessed chloroplasts and mitochondria (given that the ancestors of these organelles did not yet exist). Hence, it is unlikely that 3.4 billion-year-old fossil eukaryotes (if they exist) would be distinguishable from prokaryotes. However, it would be expected that the autogenous features that are found in eukaryotic cells (for example, the endoplasmic reticulum or pseudopods) should be progressively acquired in the proto-eukaryote lineage and that fossils that are gradually more similar to present-day eukaryotes should be found when moving towards the present. We have not seen this pattern in the fossil record. However, hypotheses in

which eukaryotes are not seen as a primary lineage of life (that is, originating from the merger of the Eubacteria and the Archaebacteria) imply that eukaryotes must be younger than both primary prokaryotic lineages and — more importantly — younger than the Alphaproteobacteria. Not much is currently known about the date of origin of the different prokaryotic phyla. However, geobiologists are constantly looking for reliable, lineage-specific, prokaryotic biomarkers in Precambrian rocks, and important advances are being made[64] in the definition of the minimal ages of specific prokaryotic lineages. Potential fossil biomarkers are known for some Gammaproteobacteria (Okenane) and Alphaproteobacteria (2α-Methylhopanes)[64], and, on the basis of these biomarkers, the oldest possible evidence for the Proteobacteria is currently found in ~1.64 billion-year-old Australian rocks.

The first eukaryotic fossil is approximately 1.6–1.8 billion years old, and molecular clock studies that estimated the age of the extant eukaryotic diversity have systematically inferred that this lineage radiated 1.7 billion years ago[65] or less[66]. Overall, these results are inconsistent with the three-domains hypothesis and instead suggest a late emergence of eukaryotes — a finding that is more consistent with ring of life hypotheses.

Evidence from bioenergetics
Lane and Martin[6] have shown that a very important problem for the transition from a relatively simple single-celled or colonial prokaryote to a complex eukaryote is the requirement for energy. Only eukaryotic cells can become complex, and Lane and Martin assert that this can only happen when a cellular power generator that is based on transmembrane chemiosmotic energy coupling can provide the requisite energy. In other words, it is not possible to become a eukaryote without a mitochondrion. The argument is that power production per gram of cells is higher in prokaryotes, such as *Escherichia coli*, than in eukaryotes, such as *Euglena* spp. However, when viewed in terms of power per haploid genome (bearing in mind that *E. coli* is orders-of-magnitude smaller than *Euglena* spp.), a protozoan has approximately 5,000 times more metabolic power than a single bacterium. Taking the massively expanded genomes of protozoans into consideration, prokaryotes and eukaryotes have similar amounts of metabolic energy per Mb of genome. So, why did prokaryotes not evolve into eukaryotes more than once? The reason is that it is impossible to generate this amount of energy in a single cell unless that

cell has a powerhouse of energy production with a large membrane surface — that is, a mitochondrion. In addition, it is necessary for the mitochondrion to lose most genes, so that they are not synthesized in thousands of copies when they are not needed. This requirement would have created a massive selective advantage for those lineages of early eukaryotes in which endosymbiotic gene transfer (EGT) took place[67].

Lynch and Conery[68] reported the counter-intuitive result that the likelihood that a newly arisen gene will survive the accumulation of mutations increases with decreasing effective population size. If the first eukaryotic cells arose as a consequence of cellular or genome merging, then it is at least possible that this created a substantial decrease in population size, which would then have enabled the accumulation of all kinds of rapid evolutionary events. In addition, Lynch and Conery showed that features such as introns within protein-coding genes are extremely unlikely to establish in small genomes. Although this is a distinct argument, it is compatible with, and augments, the assertion by Lane and Martin that larger genomes can only be found in organisms with a power source that is based on transmembrane chemiosmotic potential. The smaller effective population sizes that are found in a bottleneck, and also in those organisms with genomes larger than 10 Mb, make the radical reorganization of a stem eukaryotic genome a very likely possibility. These calculations make a eukaryotes-early hypothesis or a proto-eukaryote hypothesis very unlikely.

Evidence from modern cell biology
If a merger scenario for eukaryogenesis is probable, then we might expect this event to be detectable in various features of the cellular and molecular biology of extant eukaryotes. In a study of the best understood eukaryotic genome (that of the yeast *Saccharomyces cerevisiae*), significant differences were found between archaebacteria-affiliated and eubacteria-affiliated genes in terms of expression level, likelihood of being lethal-upon-deletion, duplicability and position in protein-interaction networks (PINs)[34]. The genes that have archaebacterial affiliation are more highly expressed than those that have eubacterial affiliation, are more likely to be lethal if they are deleted or disrupted and are more central in PINs (using both closeness-centrality and betweenness-centrality measures), whereas the eubacteria-affiliated genes are more highly duplicated on average. When the data

were split into two groups — that is, into informational genes and operational genes[69] — the informational genes were more likely to be affiliated with the Archaebacteria. However, when the data were examined across both informational and operational genes with respect to lethality, the archaebacterial homologues were more likely to be lethal-upon-deletion in both cases. Gene provenance and essentiality are therefore linked for both classes of genes. These results have also been shown to be true for the human genome[70]. Archaebacteria-affiliated genes are more highly and broadly expressed in humans, they evolve more slowly than eubacterial genes and are more central in human PINs. By contrast, eubacterial genes have duplicated much more than archaebacterial genes, and the proportion of eubacteria-affiliated genes in humans exceeds that in yeast. In terms of human disease, eubacterial genes are significantly more likely to be involved in Mendelian diseases than archaebacterial genes. This is because archaebacterial genes are more important for human cellular functioning, and mutations are more likely to be lethal and less likely to feature in diseases of living individuals[70].

In an analysis of genome-content evolution across a broad range of cellular life forms, recently developed sequence similarity network approaches were used. These networks included eukaryotes, archaebacteria and eubacteria and revealed the presence of 28 connected components with a four-element chain-like structure[3]. In these connected components, the basic structure consists of eukaryotic genes connected to archaebacterial genes, which are connected to eubacterial genes, which are connected to other eukaryotic genes. It is impossible to fit the evolution of these 28 chain-like structures onto a tree, but they can be easily accommodated by a ring. In an analysis of gene copy-number variation, when duplications or losses occur, they occur in the eubacteria-affiliated components of those genomes to a far greater extent than in the archaebacteria-affiliated components. In terms of ratio, tiny nucleomorph genomes tend to have a greater proportion of archaebacterial genes than of eubacterial genes, whereas vertebrates and plants, which have large genomes, tend to have a much greater proportion of eubacterial genes. That is, the archaebacterial component of the eukaryotic cell seems to be less changeable in terms of gene number. These facts argue again that eukaryotes are a merger of two lineages, which contributed

different numbers of genes that differ in their importance to the eukaryote.

Is a 'merger' hypothesis even plausible?

The absence of evidence that prokaryotic cells undergo fusion has been used as a criterion to suggest that hypotheses that are based on two primary lineages of life are not tenable[71]. However, recent studies[72,73] show that some archaebacteria can easily engage in cell fusions, which result in the generation of heterodiploid genomes. In addition to cell fusion, chromosomal recombination has been observed for *Haloferax volcanii* and *Haloferax mediterannei*[72,73]. These experiments show that cellular fusions involving prokaryotic cells can and do occur. A further argument that was used against the merger hypothesis was that archaebacterial membranes contain glycerol-1-phosphate lipids, whereas eukaryotic and eubacterial membranes contain glycerol-3-phosphate lipids. In this context, it was suggested that evolution via intermediates was 'selectively disfavoured' (REF. 74), hence the host in the eukaryogenetic process could not have been an archaebacterium. However, it has recently been shown that heterochiral hybrid membranes that consist of both types of lipids can be synthesized and that they are stable[75]. Furthermore, an analysis of the yeast genome has shown that there is no difference in the percentage of archaebacterial and eubacterial genes encoding plasma membrane components[3].

It has been argued that phagocytosis is a eukaryote-specific trait. However, careful phylogenetic and structural analyses of eukaryotic actin-family proteins and archaebacterial actin-like proteins suggests that ancestral actin-like proteins were capable of forming branched-filament structures and networks, which were likely to have conferred the ability to form the equivalent of eukaryotic lamellipodia or filopodia that could facilitate the archaebacterial engulfment of another prokaryote[76].

The question then remains whether it is possible for a cell to remain robust and fit following the introduction of an enormous number of completely new genes. A recent report has identified the flow of approximately 1,000 eubacterial genes into an archaebacterial cell at the origin of the Haloarchaea[77]. In addition, it has been suggested that the origin of double-membrane eubacteria occurred as a result of a cellular merger[78]. More generally, 'highways' of gene sharing, which have carried many genes into recipient genomes, have been uncovered[79]. The demonstration of such events shows that large-scale

mergers[28] have occurred and that introgression is a fundamental feature of evolution.

Although these results, when taken in isolation, do not prove that cellular fusion (consistent with a ring of life hypothesis) occurred at the origin of eukaryotes, the data clearly show that such an event was possible.

Conclusions

There is a considerable amount of consilience[12] between the phylogenetic, cell biology, population biology, biochemical and paleontological evidence in favour of a scenario in which a merger of just two prokaryotes — one eubacterial and one archaebacterial — formed the eukaryotic cell relatively late in the history of life. The eubacterial ancestor subsequently became the mitochondrion. Many of its genes were transferred to the nucleus as a consequence of selective pressures against the replication of large amounts of DNA, the protein products of which were not needed in such large quantities. Unfortunately, the consideration that this was the relationship of a 'host' and a 'symbiont' has led to the denigration of the role of the eubacterial component of the eukaryotic cell in favour of a focus on the archaebacterial component. However, in most eukaryotes (including animals), the eubacterial gene components are numerically greater and span all major functions, including both informational and operational functions. Therefore, a proper view of eukaryotic origins needs to take account of both ancestries of eukaryotic genes. In addition, given the importance of anaerobic energy metabolism in the evolution of eukaryotes[80], future work is needed to clarify the nature of this merger.

Eukaryotes are indeed 'irreducibly complex' — it is very difficult to envision a situation in which a complex eukaryotic ancestor gave rise to two streamlined prokaryotes with substantially different kinds of genes. In easily identifiable situations of reduced genomes, we see that the same kinds, not different kinds, of genes are lost, and these are usually metabolic genes[81]. Eukaryotes-early hypotheses imply a pattern of gene loss that is simply not seen anywhere else, and these hypotheses can safely be excluded owing to the high number of unusual *ad hoc* explanations they would require. In addition, autogenous eukaryogenesis seems to be very unlikely — particularly from the perspectives of energetics and phylogenetics. A key influence of viral genes has been proposed[82,83], but the yeast nucleus is not enriched in genes that have viral homologues, which might be expected if viruses played an important part in eukaryogenesis[3].

What would falsify the ring of life hypotheses that are advocated in this Opinion article, causing their reappraisal? First, the finding of a primitively amitochondriate eukaryote would falsify this idea. That is, the finding of a eukaryote that branches deeply as a sister group of mitochondrion-containing eukaryotes, inferred using well-fitting phylogenetic models, in which no evidence can be found that this organism possess even one mitochondria-derived gene. Second, the discovery of fossil eukaryotes that are older than the Alphaproteobacteria would disprove this hypothesis. Unfortunately, the precise age of the Proteobacteria still needs to be clarified, but there is considerable scope for modern geobiology to clarify this point in the future[64]. Still, it is clear that uncontroversial fossil evidence of eukaryotes that are close to 3 billion years old would substantially weaken any ring of life hypothesis.

The chimeric nature of the eukaryotes has profound implications for our understanding of the nature of life on Earth. If the Eukaryota have hybrid origins, then strictly speaking, both the extant Eubacteria and Archaebacteria are paraphyletic, whereas the Eukaryota, as a lineage, is monophyletic, despite its symbiogenic ancestry. Gribaldo *et al.*[46] characterized the Eocyte hypothesis (and hence the ring of life hypothesis) as a two-domain hypothesis. This characterization is inherently flawed owing to the chimeric nature of the eukaryotes. Given that the current evidence suggests that there is a single highest-level category of life with two (ancient) primary groups — the Eubacteria and the Archaebacteria — followed by a secondary group of more recent origins — the Eukaryota, we advocate a 'domain–free' view of life on Earth. It is time to get used to depicting life on this planet as a graph with rings.

James O. McInerney is at the Bioinformatics and Molecular Evolution Unit, Department of Biology, National University of Ireland, Maynooth, County Kildare, Ireland.

Mary J. O'Connell is at the Bioinformatics and Molecular Evolution Group, School of Biotechnology, Dublin City University, Glasnevin, Dublin 9, Ireland.

Davide Pisani is at the School of Biological Sciences and School of Earth Sciences, University of Bristol, Woodland Road, Bristol BS8 1UG, UK.

Correspondence to J.O.M.
e-mail: james.o.mcinerney@nuim.ie

1. Martin, W. & Kowallik, K. Annotated English translation of Mereschkowsky's 1905 paper 'Über Natur und Ursprung der Chromatophoren imvPflanzenreiche'. *Eur. J. Phycol.* **34**, 287–295 (1999).
2. Wallin, I. E. The mitochondria problem. *Am. Naturalist* **57**, 255–261 (1923).

3. Alvarez-Ponce, D., Bapteste, E., Lopez, P. & McInerney, J. O. Gene similarity networks provide new tools for understanding eukaryote origins and evolution. *Proc. Natl Acad. Sci. USA* **110**, E1594–E1603 (2013).

4. 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).

5. Kurland, C. G., Collins, L. J. & Penny, D. Genomics and the irreducible nature of eukaryote cells. *Science* **312**, 1011–1014 (2006).

6. Lane, N. & Martin, W. The energetics of genome complexity. *Nature* **467**, 929–934 (2010).

7. López-García, P. & Moreira, D. Selective forces for the origin of the eukaryotic nucleus. *Bioessays* **28**, 525–533 (2006).

8. Pisani, D., Cotton, J. A. & McInerney, J. O. Supertrees disentangle the chimerical origin of eukaryotic genomes. *Mol. Biol. Evol.* **24**, 1752–1760 (2007).

9. Rivera, M. C. & Lake, J. A. The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* **431**, 152–155 (2004).

10. Stechmann, A. & Cavalier-Smith, T. The root of the eukaryote tree pinpointed. *Curr. Biol.* **13**, 665–666 (2003).

11. Popper, K. R. *The Logic of Scientific Discovery* (Routledge, 1959).

12. Whewell, W. *The Philosophy Of Inductive Sciences, Founded Upon Their History* (John W. Parker, 1840).

13. 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).

14. 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).

15. Margulis, L. Archaeal–eubacterial mergers in the origin of Eukarya: phylogenetic classification of life. *Proc. Natl Acad. Sci. USA* **93**, 1071–1076 (1996).

16. Rinke, C., Schwientek, P., Sczyrba, A. & Ivanova, N. N. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).

17. 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).

18. Cavalier-Smith, T. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int. J. Syst. Evol. Microbiol.* **52**, 297–354 (2002).

19. Devos, D. P. & Reynaud, E. G. Evolution. Intermediate steps. *Science* **330**, 1187–1188 (2010).

20. Martin, W. & Müller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41 (1998).

21. Searcy, D. G. Metabolic integration during the evolutionary origin of mitochondria. *Cell Res.* **13**, 229–238 (2003).

22. Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G. J. & Woese, C. R. Mitochondrial origins. *Proc. Natl Acad. Sci. USA* **82**, 4443–4447 (1985).

23. McInerney, J. O. *et al.* Planctomycetes and eukaryotes: a case of analogy not homology. *Bioessays* **33**, 810–817 (2011).

24. Hirt, R. P. *et al.* Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc. Natl Acad. Sci. USA* **96**, 580–585 (1999).

25. Roger, A. J. *et al.* A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. *Proc. Natl Acad. Sci. USA* **95**, 229–234 (1998).

26. Rodriguez-Ezpeleta, N. & Embley, T. M. The SAR11 group of α-proteobacteria is not related to the origin of mitochondria. *PLOS ONE* **7**, e30520 (2012).

27. Woese, C. R. Bacterial evolution. *Microbiol. Rev.* **51**, 221 (1987).

28. Bapteste, E. *et al.* Evolutionary analyses of non-genealogical bonds produced by introgressive descent. *Proc. Natl Acad. Sci. USA* **109**, 18266–18272 (2012).

29. Gogarten, J. P. *et al.* Evolution of the vacuolar H+-ATPase: implications for the origin of eukaryotes. *Proc. Natl Acad. Sci. USA* **86**, 6661–6665 (1989).

30. Brinkmann, H. & Philippe, H. Archaea sister group of Bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. *Mol. Biol. Evol.* **16**, 817–825 (1999).

31. Forterre, P. The origin of DNA genomes and DNA replication proteins. *Curr. Opin. Microbiol.* **5**, 525–532 (2002).

32. Van Valen, L. M. & Maiorana, V. C. The archaebacteria and eukaryotic origins. *Nature* **287**, 248–250 (1980).

33. de Duve, C. The origin of eukaryotes: a reappraisal. *Nature Rev. Genet.* **8**, 395–403 (2007).

34. Cotton, J. A. & McInerney, J. O. Eukaryotic genes of archaebacterial origin are more important than the more numerous eubacterial genes, irrespective of function. *Proc. Natl Acad. Sci. USA* **107**, 17252–17255 (2010).

35. Margulis, L., Bermudes, D. & Obar, R. Symbiosis in evolution: status of the hypothesis of the spirochete origin of undulipodia. *Orig. Life Evol. Biosph.* **16**, 319 (1986).

36. 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).

37. Williams, T. A., Foster, P. G., Nye, T. M. W., Cox, C. J. & Embley, T. M. A congruent phylogenomic signal places eukaryotes within the Archaea. *Proc. Biol. Sci.* **279**, 4870–4879 (2012).

38. Lake, J. A., Servin, J. A., Herbold, C. W. & Skophammer, R. G. Evidence for a new root of the tree of life. *Systemat. Biol.* **57**, 835–843 (2008).

39. Dagan, T., Roettger, M., Bryant, D. & Martin, W. Genome networks root the tree of life between prokaryotic domains. *Genome Biol. Evol.* **2**, 379–392 (2010).

40. Ciccarelli, F. D. *et al.* Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**, 1283–1287 (2006).

41. 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).

42. Creevey, C. J., Doerks, T., Fitzpatrick, D. A., Raes, J. & Bork, P. Universally distributed single-copy genes indicate a constant rate of horizontal transfer. *PLOS ONE* **6**, e22099 (2011).

43. Dagan, T. & Martin, W. The tree of one percent. *Genome Biol.* **7**, 118 (2006).

44. Gouy, M. & Li, W. H. Phylogenetic analysis based on rRNA sequences supports the archaebacterial rather than the eocyte tree. *Nature* **339**, 145–147 (1989).

45. 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).

46. 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? *Nature Rev. Microbiol.* **8**, 743–752 (2010).

47. Williams, T. A. & Embley, T. M. Archaeal "dark matter" and the origin of eukaryotes. *Genome Biol. Evol.* **6**, 474–481 (2014).

48. 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 (1990).

49. 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).

50. Keane, T. M., Creevey, C. J., Pentony, M. M., Naughton, T. J. & McLnerney, J. O. Assessment of methods for amino acid matrix selection and their use on empirical data shows that *ad hoc* assumptions for choice of matrix are not justified. *BMC Evol. Biol.* **6**, 29 (2006).

51. Cox, C., Foster, P., Hirt, R. & Harris, S. The archaebacterial origin of eukaryotes. *Proc. Natl Acad. Sci. USA* **105**, 20356–20361 (2008).

52. Foster, P. G., Cox, C. J. & Embley, T. M. The primary divisions of life: a phylogenomic approach employing composition-heterogeneous methods. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **364**, 2197–2207 (2009).

53. Foster, P. G. Modeling compositional heterogeneity. *Syst. Biol.* **53**, 485–495 (2004).

54. Lartillot, N., Lepage, T. & Blanquart, S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288 (2009).

55. Guy, L. & Ettema, T. J. The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol.* **19**, 580–587 (2011).

56. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).

57. Makarova, K. S., Yutin, N., Bell, S. D. & Koonin, E. V. Evolution of diverse cell division and vesicle formation systems in Archaea. *Nature Rev. Microbiol.* **8**, 731–741 (2010).

58. Mojzsis, S. J. *et al.* Evidence for life on Earth before 3,800 million years ago. *Nature* **384**, 55–59 (1996).

59. Wacey, D., Kilburn, M. R., Saunders, M., Cliff, J. & Brasier, M. D. Microfossils of sulphur-metabolizing cells in 3.4-billion-year-old rocks of Western Australia. *Nature Geosci.* **4**, 698–702 (2011).

60. Knoll, A. H., Javaux, E. J., Hewitt, D. & Cohen, P. Eukaryotic organisms in Proterozoic oceans. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 1023 (2006).

61. Brocks, J. J., Logan, G. A., Buick, R. & Summons, R. E. Archean molecular fossils and the early rise of eukaryotes. *Science* **285**, 1033–1036 (1999).

62. Knoll, A. H. Paleobiological perspectives on early eukaryotic evolution. *Cold Spring Harb. Perspect. Biol.* **6**, a016113.

63. Rasmussen, B., Fletcher, I. R., Brocks, J. J. & Kilburn, M. R. Reassessing the first appearance of eukaryotes and cyanobacteria. *Nature* **455**, 1101–1104 (2008).

64. Brocks, J. J. & Banfield, J. Unravelling ancient microbial history with community proteogenomics and lipid geochemistry. *Nature Rev. Microbiol.* **7**, 601–609 (2009).

65. Parfrey, L. W., Lahr, D. J. G., Knoll, A. H. & Katz, L. A. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc. Natl Acad. Sci. USA* **108**, 13624–13629 (2011).

66. Shih, P. M. & Matzke, N. J. Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins. *Proc. Natl Acad. Sci. USA* **110**, 12355–12360 (2013).

67. Timmis, J. N., Ayliffe, M. A., Huang, C. Y. & Martin, W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Rev. Genet.* **5**, 123–135 (2004).

68. Lynch, M. & Conery, J. S. The origins of genome complexity. *Science* **302**, 1401–1404 (2003).

69. Jain, R., Rivera, M. C. & Lake, J. A. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl Acad. Sci. USA* **96**, 3801 (1999).

70. Alvarez-Ponce, D. & McInerney, J. O. The human genome retains relics of its prokaryotic ancestry: human genes of archaebacterial and eubacterial origin exhibit remarkable differences. *Genome Biol. Evol.* **3**, 782–790 (2011).

71. Poole, A. M. & Penny, D. Evaluating hypotheses for the origin of eukaryotes. *Bioessays* **29**, 74–84 (2007).

72. Naor, A. & Gophna, U. Cell fusion and hybrids in Archaea: prospects for genome shuffling and accelerated strain development for biotechnology. *Bioengineered* **4**, 126–129 (2013).

73. Naor, A., Lapierre, P., Mevarech, M., Papke, R. T. & Gophna, U. Low species barriers in halophilic archaea and the formation of recombinant hybrids. *Curr. Biol.* **22**, 1444–1448 (2012).

74. Wachtershauser, G. From pre-cells to Eukarya — a tale of two lipids. *Mol. Microbiol.* **47**, 13–22 (2003).

75. Shimada, H. & Yamagishi, A. Stability of heterochiral hybrid membrane made of bacterial sn-G3P lipids and archaeal sn-G1P lipids. *Biochemistry* **50**, 4114–4120 (2011).

76. Yutin, N., Wolf, M. Y., Wolf, Y. I. & Koonin, E. V. The origins of phagocytosis and eukaryogenesis. *Biol. Direct* **4**, 9 (2009).

77. Nelson-Sathi, S. *et al.* Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea. *Proc. Natl Acad. Sci. USA* **109**, 20537–20542 (2012).

78. Lake, J. A. Evidence for an early prokaryotic endosymbiosis. *Nature* **460**, 967–971 (2009).

79. Beiko, R. G., Harlow, T. J. & Ragan, M. A. Highways of gene sharing in prokaryotes. *Proc. Natl Acad. Sci. USA* **102**, 14332 (2005).

80. Muller, M. *et al.* Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol. Mol. Biol. Rev.* **76**, 444–495 (2012).

81. Moran, N. A. Microbial minimalism: genome reduction in bacterial pathogens. *Cell* **108**, 583–586 (2002).

82. Forterre, P. Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. *Proc. Natl Acad. Sci. USA* **103**, 3669–3674 (2006).

83. Forterre, P. The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.* **117**, 5–16 (2006).