© 2016. Published by The Company of Biologists Ltd | Journal of Cell Science (2016) 0, 1-9 doi:10.1242/jcs.178566

COMMENTARY

The changing view of eukaryogenesis – fossils, cells, lineages and how they all come together

Joel B. Dacks[1], Mark C. Field[2], Roger Buick[3], Laura Eme[4], Simonetta Gribaldo[5], Andrew J. Roger[4], Céline Brochier-Armanet[6] and Damien P. Devos[7,*]

ABSTRACT

Eukaryogenesis – the emergence of eukaryotic cells – represents a pivotal evolutionary event. With a fundamentally more complex cellular plan compared to prokaryotes, eukaryotes are major contributors to most aspects of life on Earth. For decades, we have understood that eukaryotic origins lie within both the Archaea domain and α-Proteobacteria. However, it is much less clear when, and from which precise ancestors, eukaryotes originated, or the order of emergence of distinctive eukaryotic cellular features. Many competing models for eukaryogenesis have been proposed, but until recently, the absence of discriminatory data meant that a consensus was elusive. Recent advances in paleogeology, phylogenetics, cell biology and microbial diversity, particularly the discovery of the 'Candidatus Lokiarcheaota' phylum, are now providing new insights into these aspects of eukaryogenesis. The new data have allowed finessing the time frame during which the events of eukaryogenesis occurred, a more precise identification of the contributing lineages and their likely biological features. The new data have allowed finessing of the time frame during which the events of eukaryogenesis occurred, a more precise identification of the contributing lineages and clarification of their probable biological features.

KEY WORDS: Eukaryogenesis, Evolution, Archaea, Molecular fossil, Molecular dating, Last eukaryotic common ancestor, Chemical fossil, Endosymbiosis, Archaeogenesis, First eukaryotic common ancestor

Introduction

Eukaryogenesis, or the process by which eukaryotes originated, had a revolutionary impact on the subsequent history of life, including the evolution of complex multicellular organisms. Consequently, determining the players, timing and dynamics of eukaryogenesis is key to understanding the origins of major drivers of the global ecosystem and their subsequent development. It is also crucial to identify those common mechanisms embedded within eukaryotic cells and to discriminate these from features that constitute lineage-specific and/or niche-adaptive mechanisms. This knowledge is of importance both for analysing the origins of cellular organelles and other unique features of eukaryotic cells, and is essential for understanding of the mechanisms underlying diseases and/or pathogenesis.

The transition between prokaryotic and eukaryotic cellular architectures is typically simplified and represented in textbooks as essentially a singular rapid event. However, given the vast number of modifications to cellular systems that needed to be accommodated, eukaryogenesis must have been a multi-step process. Unravelling the individual steps of eukaryogenesis is a tractable problem when sufficient discriminatory data, allowing the acceptance or rejection of specific models, are available. However, although many aspects of eukaryogenesis are generally agreed upon, elucidating the precise mechanisms, the order of many of these events and the identity of the specific cellular lineages involved remains problematic, in large part owing to the absence of intermediate forms – i.e. a prokaryote transitioning to become a eukaryote. Multiple competing hypotheses, which are all reasonable and supported by diverse lines of evidence, have been proposed (López-García and Moreira, 2015; Martin et al., 2015).

Recently, multiple advances in geology, phylogenetics, comparative genomics, cell biology and the charting of microbial diversity have provided data that allow us to discriminate between those scenarios that are most likely and those that are less so. Here, we discuss how these contributions have placed boundaries on the timing of eukaryogenic events, how they have resolved the crucial inter-relationship between eukaryotes and archaea (including revealing the closest present-day archaeal relatives of eukaryotes) and how they have clarified the origins of key eukaryotic traits.

Eukaryogenesis – what is agreed upon and what are the major questions?

As many aspects of eukaryogenesis remain hotly contested, it is crucial to establish generally agreed terms of reference and outline those aspects solidly supported by evidence and, thus, essentially universally accepted. Eukaryogenesis is the entire process by which the defining traits of eukaryotic cells arose in the lineages that eventually gave rise to all present-day eukaryotes (Fig. 1). Lineages, in plural, is key here, as one crucial, uncontroversial aspect of eukaryogenesis is that extant eukaryotes have a chimeric origin; one fraction of eukaryotic genes is of bacterial ancestry and a second shares a common origin with archaea (Rivera et al., 1998). In addition, there is a third group of genes that represents strictly eukaryotic innovations and, thus, includes a substantial proportion of the genes required to build and define many intracellular compartments. Although some of these features are likely of prokaryotic ancestry, there is a considerable expansion of the associated gene families, and their elaboration makes them true eukaryotic innovations (Klinger et al., 2016).

[1]Department of Cell Biology, University of Alberta, Edmonton, Canada T6G 2H7. [2]School of Life Sciences, University of Dundee, Dundee DD1 5EH, UK. [3]Department of Earth and Space Science and Astrobiology Program, University of Washington, Seattle, WA 98195-1310, USA. [4]Centre for Comparative Genomics and Evolutionary Bioinformatics, Department of Biochemistry and Molecular Biology, Dalhousie University, 5850 College Street, Halifax, Nova Scotia, Canada B3H 4R2. [5]Institut Pasteur, Département de Microbiologie, Unité de Biologie Moleculaire du Gene chez les Extremophiles, rue du Dr Roux, Paris 75015, France. [6]Université de Lyon, Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, 43 Boulevard du 11 Novembre 1918, Villeurbanne F-69622, France. [7]Centro Andaluz de Biología del Desarrollo (CABD), Universidad Pablo de Olavide, Seville 41013, Spain.

*Author for correspondence (damienpdevos@gmail.com)

D.P.D., 0000-0002-9453-4753

Journal of Cell Science (2016) 0, 1-9 doi:10.1242/jcs.178566

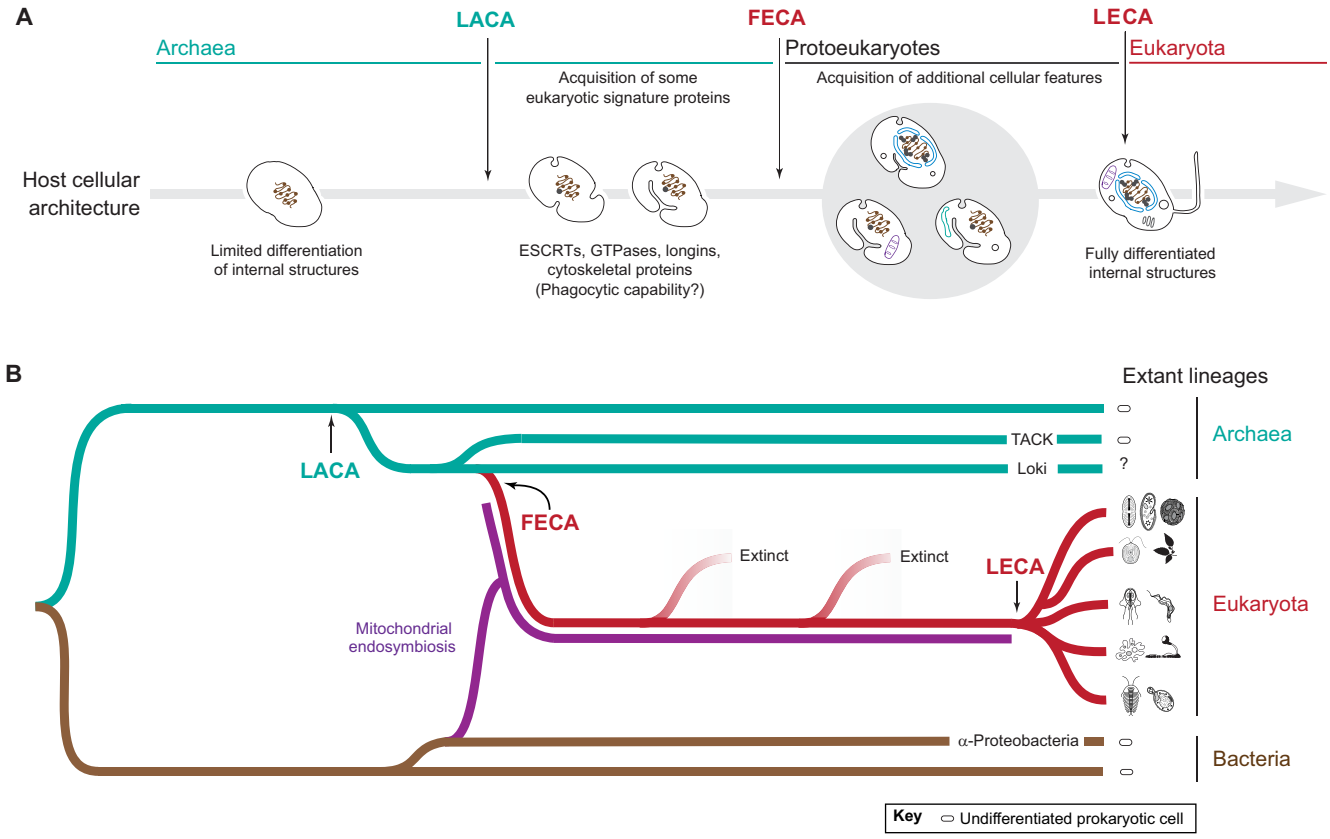

Fig. 1. Events leading to eukaryogenesis. (A) The major events involved in eukaryogenesis, from a cellular architecture and organelle perspective are shown here, with the earliest to the left. Noted at the top are two transitional periods – the first occurs between the ancestor of the closest archaeal lineage to eukaryotes (last archaeal common ancestor; LACA) and the first eukaryotic common ancestor (FECA). The second transition occurs between the FECA and the development of the full complement of traits common to the last common ancestor of all eukaryotes (LECA). The possible cellular architecture of these ancestors is shown based on attributes that have been defined as being present in the archaeal and α-proteobacterial lineages contributing to eukaryotes, such as the emergence of more complex internal cellular structures, crucially the nucleus (in blue), mitochondria (purple) and a primordial endomembrane organelle (teal). Their order of acquisition is unclear; hence, three possibilities are shown with the nucleus, mitochondria or primordial endomembrane organelle differentiating first. Permutations of concurrent trait origins are not illustrated for simplicity nor are additional cellular features. It remains unclear how early before the LECA that all the features of modern eukaryotes emerged. Possible ancestral homologues of various molecular systems have recently been identified within the TACK lineages of Archaea and were, therefore, present before the FECA, including cytoskeletal proteins and building blocks of the membrane trafficking system (GTPases and longins). (B) Relationships between archaeal, bacterial and eukaryotic lineages, illustrating key symbioses, speciation events and landmarks. Bacteria are shown in brown, Archaea in teal and Eukaryota in red. The earliest times are to the left, and the present day and extant lineages to the right. Selected extant lineages are shown as representative diagrams, with the exception of the Lokiarchaeum (Loki) described recently and closely related to eukaryotes and TACK, but for which no cellular data are available. Key early events are: differentiation of archaea from the last archaeal common ancestor (LACA, teal), differentiation of the first eukaryotic common ancestor (FECA, red) from the ancestor of the archaeal lineage that is closer to eukaryotes and acquisition of the mitochondrion from an α-protobacterial donor (purple). Later events include: extinction of transitional eukaryotic forms and radiation from their last eukaryotic common ancestor (LECA, red). The branching position of Lokiarchaeum as the sister to eukaryotes remains to be confirmed at present.

Teasing apart eukaryogenesis involves conceptually overlapping aspects (Fig. 1). A cell biology aspect relates to the origins and order of acquisition of those features that would convert a recognisably prokaryotic organism into one that possessed one or more cellular traits, such that it would become, by common meaning ‘eukaryotic’ (Fig. 1A). Phylogenetic and paleontological aspects relate to the underlying timeline for these acquisitions. How long did this process take and what are some of the useful defined landmarks on the timeline, which can serve as points before or after which certain traits can be inferred to have evolved

Most intuitive of these landmarks is the last eukaryotic common ancestor (LECA) (Makarova et al., 2005). By definition, this is the most recent common ancestral eukaryotic cell (or possibly an interbreeding population) from which all modern eukaryotes are derived. The LECA can be reconstructed by comparative ultrastructural and genomic studies across the current diversity of eukaryotic cells (Dacks and Doolittle, 2001). Such studies have shown that LECA was already complex in many aspects, such as the presence of protein families for compartmentalisation of the cytoplasm (Koonin, 2010; Koumandou et al., 2013). Equally important, but more nebulous, is the first eukaryotic common ancestor (FECA) – i.e. the first ancestor of the eukaryotic lineage. This is the latest branching point at which the eukaryote lineage and its closest extant relatives separated. FECA and LECA represent the oldest and youngest boundaries between which eukaryogenesis itself occurred (Fig. 1B).

With these landmarks established, the fundamental issue can be distilled to the following question: from which ancestors did these key eukaryotic traits originate and in what order did they emerge? The list of traits to be accounted for includes the nucleus, the cytoskeleton, the mitochondrion and the endomembrane system – i.e. the endoplasmic reticulum (ER), Golgi and endosomes. Additional traits associated with genome functions (e.g. linear chromosomes with telomeres, spliceosomal introns, a

large genome size and RNA splicing), as well as associated cellular features and capabilities (e.g. lipid biosynthesis, mitosis, meiosis) must also be accounted for. Some of these are not uniquely eukaryotic, and there are abundant examples of similar prokaryotic features or even molecular analogues (Devos et al., 2014). For example, some prokaryotes can be as large as typical eukaryotic cells (Schulz et al., 2009), and some eukaryotes can be as small as prokaryotes [e.g. picoeukaryotes (Simon et al., 2015)]. Similarly, prokaryotes can have small linear chromosomes (e.g. *Borrelia burgdorferi*) or very large genomes (e.g. *Sorangium cellulosum* has a genome of >13 Mb). Moreover, many prokaryotes have internal membranes and structures that strongly resemble organelles, such as the anammoxosome in '*Candidatus* Brocadia anammoxidans' (Neumann et al., 2014). The key challenge is to distinguish functional analogy from true homology and, hence, the convergent evolution of features from homologies arising through direct descent (McInerney et al., 2011).

A clear way forward is to identify the prokaryotic origins of these individual eukaryotic traits. The evidence supporting mitochondria as being derived from an α-proteobacterium was originally deeply contentious, but once this was accepted, it provided a powerful explanation for the origins of certain key eukaryotic traits. For instance, if the genes that encode the proteins underlying a trait – i.e. in this case, mitochondrial function – are demonstrated to be either only present in α-Proteobacteria or only in taxa most closely related to them, a mitochondrial ancestry for this trait can then be inferred (Fig. 2, trait 1). Evidence of precisely this sort can be found for aerobic mitochondrial energy-generating pathways, and for mitochondrial systems such as mitochondrial protein translocation and the iron–sulfur (Fe-S) cluster pathway (for example, see Müller et al., 2012). Because many of the genes of α-proteobacterial origin are encoded in the eukaryotic nucleus, the origin of a fraction of the bacteria-derived eukaryotic gene set is explained by endosymbiotic gene transfer.

Spurred by progress in understanding the origin of mitochondrial traits, the role of mitochondria in eukaryogenesis has been at the centre of debate (O'Malley, 2010). The main arguments essentially revolve around details of mitochondrial provenance, the timing of acquisition and whether the advent of a mitochondrion was a necessary prerequisite for the origin of eukaryotes (Lane and Martin, 2010). It is still unclear if acquisition of the mitochondrial precursor ultimately led to a significant advantage in terms of the energy available to the nascent eukaryote. However, the key questions are: firstly, whether this was an early event, and secondly, whether or not it was necessary in order to meet the energetic demands of the more complex eukaryotic architecture (Lynch and Marinov, 2015). Alternatively, some (or most) of the eukaryotic machinery – specifically membrane transport systems, including those involved in phagocytosis – could have evolved earlier and facilitated acquisition of the α-proteobacterium (Koumandou et al., 2013).

Although the connection between eukaryotes and α-proteobacteria, through the origin of mitochondria, is now uncontested, the evolutionary relationship between eukaryotes and Archaea has been much less clear (Gribaldo et al., 2010). The discovery of the Archaea is rightly considered one of the most important advances in modern evolutionary biology (Woese and Fox, 1977) and raised the possibility of several alternative scenarios with regard to the relationship between Archaea and eukaryotes. The distinction between the Archaea and Bacteria domains was confirmed by rooting the three-domains of life using pairs of anciently duplicated genes (Gogarten et al., 1989; Gribaldo and

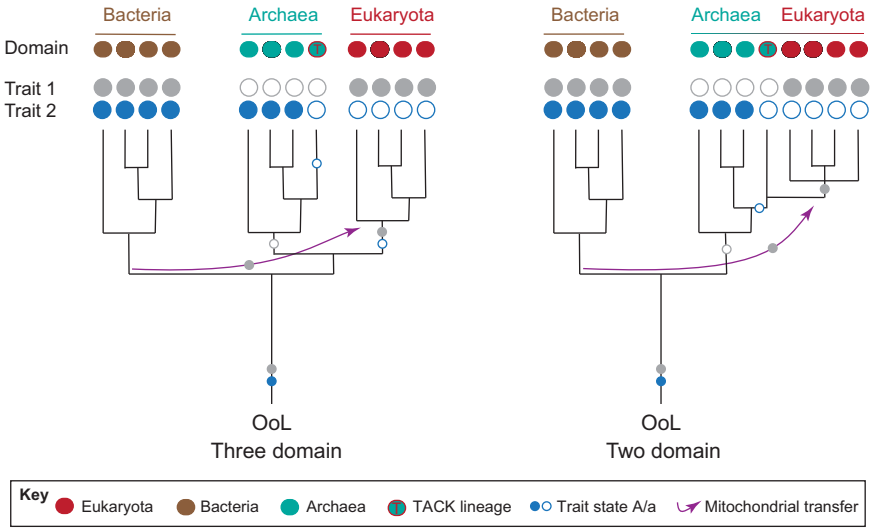

Fig. 2. Inference of feature evolution based on the phylogenetic relationship of Eukaryota and Archaea. The classic view of the tree of life places the eukaryotes (red), bacteria (brown) and archaea (teal) as distinct domains, with the eukaryotes and archaea more closely related to each other than either is to the bacteria (three-domain tree of life). More recent evidence quite convincingly supports the notion of the emergence of eukaryotes from within the archaea, and, specifically, from a lineage close to the TACK archaea (teal, red outline) (two-domain tree of life). This has a profound impact on the origins of eukaryotes and how specific traits (blue and grey) arose. For example, if bacteria and eukaryotes share a given state for a specific trait (grey circles) that is distinct from the Archaea (open grey circles), then the distribution can be explained in the frame of the three-domain model by a trait state change at the root of the Archaea, switching from closed to open grey. However, the same distribution is explained as two independent changes under the two-domain model. In both models, the trait state change can also be explained by a bacteria contribution to eukaryogenesis, potentially through mitochondrial endosymbiosis (purple arrow). If a given trait (open versus closed blue circles) is shared by eukaryotes and some, but not all Archaea (e.g. the Lokiarchaeota), the two-domain model allows for a simple explanation of the distribution in the common ancestor of eukaryotes and their immediate ancestral archaeal sisters. Under the three-domain model, such a trait distribution is most simply explained by two independent acquisitions. Notably, the chemistry of lipids comprising the cell membrane was, until recently, thought to follow a pattern like that shown for trait 1, but it now appears to follow the pattern of trait 2 (Villanueva et al., 2016). Importantly, support for the two-domain versus the three-domain model relies on phylogenetic information, and not on trait distribution, with recent evidence favouring the two-domain scenario. Ool, origin of life.

Journal of Cell Science (2016) 0, 1-9 doi:10.1242/jcs.178566

Cammarano, 1998; Zhaxybayeva et al., 2005). This suggested that Archaea are sisters to Eukaryota, meaning that they separately evolved from a common ancestor, as opposed to one from within the other. However, alternative scenarios have been proposed, such as that there are only two primary domains (i.e. Bacteria and Archaea) and that eukaryotes arose from within the Archaea (Fig. 2) (López-García and Moreira, 2015; Martin et al., 2015). Resolving the issue between two- or three-domains is crucial to understanding how, and from which origin, traits that are characteristically described as being of eukaryotic origin. The notion of Archaea and Eukaryota as sister lineages implies that any trait they share would have been inherited from their common ancestor, which in the three-domain phylogenetic tree, is the last archaea–eukaryote common ancestor. By contrast, if eukaryotes are embedded within Archaea, they must be a sister lineage to a specific present-day archaeal lineage, with the consequence that the traits of the archaeal contributor to eukaryogenesis could be determined through analysis of the closest archaeal sisters of eukaryotes (Fig. 2). Therefore, understanding the identity and biology of the closest archaeal relatives of eukaryotes could allow a more refined reconstruction of the traits that the archaeal lineage contributed to eukaryogenesis. Recent advances in paleontology and molecular phylogenetics place time constraints on the timeline for eukaryogenesis and essentially resolve the two- versus three-domain question, potentially pinpointing the archaeal lineage that gave rise to eukaryotes.

Relative ages of bacteria, archaea and eukaryotes – the fossil record

Constraining the timing of early events in the evolution of Archaea and Eukarya requires consideration of the fossil record. It is generally believed that early cells were small and, owing to their great age, are unlikely to be well-preserved. Nonetheless, two types of microbial fossils are present in rocks billions of years old – physical remnants, such as microfossils and stromatolites, and chemical residues, such as biomarker molecules and isotopic fractionations.

Life on Earth is an ancient phenomenon, but the available evidence cannot give a clear answer regarding the exact nature of the oldest biological forms. It is, however, widely accepted that the earliest life forms were prokaryotic. Indeed, the oldest identified bacterial fossils are 3.48 billion years (giga-annum, Ga) old (Fig. 3) (Shen et al., 2001, 2009; Ueno et al., 2008). Sulfate crystals from the Dresser Formation in Australia contain microscopic sulfides that show the large negative sulfur isotopic fractionation that is characteristic of dissimilatory sulfate reduction (Shen et al., 2001). Although several groups of Archaea and Bacteria perform this metabolism, only Bacteria are known to do so at low temperatures. As the source sulfate came from crystals that were originally gypsum (comprising calcium sulfate dihydrate), which is only formed at low temperatures, the sulfate reducers must therefore have been Bacteria. This places an early age limit on the separation of the Archaea–Eukarya lineage from Bacteria. Archaea have simple cell structures and, thus, are unlikely to have left unambiguously identifiable physical fossils; indeed, their fossil record consists solely of chemical evidence. The oldest geological evidence of Archaea comes from isotopically light methane found in fluid inclusions in a chert-quartz vein cutting 3.49-Ga-old rocks at North Pole in Western Australia (Ueno et al., 2006). However, the origin of the veins at this locality is debated; some argue that the veins are neptunian dykes (i.e. filled from above) (Buick, 1984, 1988), whereas others consider them as hydrothermal feeder dykes that emanated from below (Nijman et al., 1998; Van Kranendonk, 2006); in the latter case they could be remobilizing methane from older underlying rocks. Regardless, the age of the inclusions is constrained by the Dresser Formation that lies above them and has been reliably dated at 3.48 Ga (Van Kranendonk et al., 2008). At ∼2.7 Ga, sedimentary rocks from the Fortescue Group in the geographic area contain organic matter with extremely light carbon isotope ratios depleted down to −60‰ (with 60 parts per thousand less of the light-stable isotope of carbon than in a standard). As these low values occur in diverse environments ranging from deep marine (Eigenbrode and Freeman, 2006) to alkaline lakes (Stüeken et al., 2015), and as methanogenesis is the only process known to produce such consistently depleted isotope values, this shows that by the late Archean, archaeal methanogens were ecologically important members of microbial communities across many habitats.

In contrast to Archaea, the complex cell structure of Eukaryota makes it easier to identify their microfossils by their intricate surface ornamentation, complex wall ultrastructure, excystment splits or spiny protrusions that extend from their cell surface (Buick, 2010; Javaux et al., 2003). Moreover, eukaryotes produce complex sterols that can be preserved under mild metamorphic conditions in the form of steranes, which are non-functional and fully saturated sterol

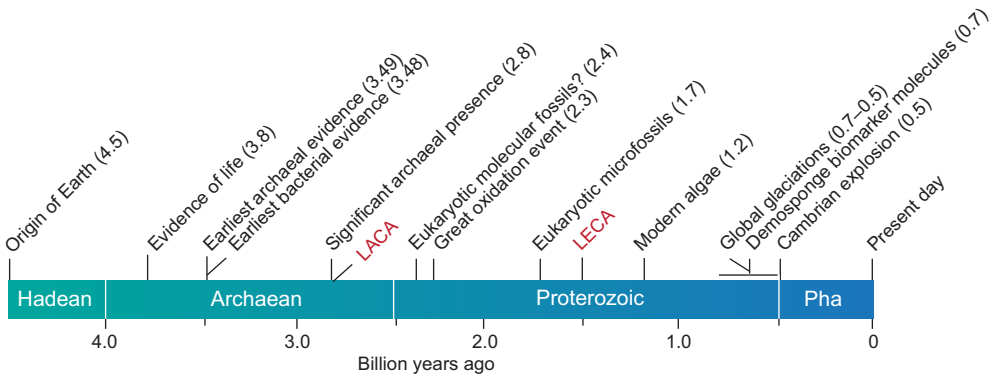

Fig. 3. Timeline of life on Earth. Important events and dates for the earliest available evidence are shown. The main bar is divided into the important recognised geological epochs from the origin of Earth to the modern era, the Phaenozoic era (Pha). The scale at the bottom is expressed in years before the present, and approximate dates for events are given in billions of years (Ga) before the present. Estimated times when the last archaeal common ancestor (LACA) and the last eukaryotic common ancestor (LECA) existed are highlighted in red. The position at which the first eukaryotic common ancestor (FECA) emerged remains difficult to place. It is also unclear if the earliest eukaryotic microfossils represent lineages within the transition period between the FECA and LECA; current estimates suggest that they could indeed predate the LECA, but future evidence might overturn this view.

derivatives that retain their carbon skeleton (Summons and Walter, 1990). It had been proposed that steranes from drill-core samples of ∼2.7-Ga-old shales from the Fortescue Group represent the oldest fossil evidence of eukaryotes (Brocks et al., 1999), potentially making them as old as Archaea and lending support to the theory of Archaea and Eukarya as sister taxa. However, more recent isotopic studies of the same rocks showed that the $\delta^{13}C$ values of the soluble hydrocarbons were inconsistent with co-exisiting kerogen and pyrobitumen, suggesting that the steranes were younger contaminants (Rasmussen et al., 2008). This has now been confirmed by an analysis of a new drill-core that was obtained using ultraclean drilling and sampling techniques from a site alongside the original core that yielded the highest sterane concentrations (French et al., 2015). Despite using many different analytical approaches, none of the new samples yielded any steranes whatsoever, thereby strongly indicating that those found in previous studies were indeed contamination introduced during drilling or sampling. Thus, 2.7 Ga can no longer be considered as the age of the oldest evidence of eukaryotes. Several additional rock formations aged between 2.4 and 1.4 Ga have been suggested to contain indigenous molecular eukaryotic fossils (Pawlowska et al., 2013), but none of these have been fully accepted by the wider scientific community. Thus, the oldest unambiguous molecular fossils of eukaryotes are from the 0.7- to 0.63-Ga Huqf Supergroup of Oman in the form of 24-isopropylcholestanes derived from Demosponges (Love and Summons, 2015; Love et al., 2009).

Eukaryotic microfossils are somewhat less controversial. Several species of acritarchs (organic vesicular microfossils) found in the ∼1.7-Ga-old Changcheng Group in China (Yan and Liu, 1993) and the 1.65-Ga-old Mallapunyah Formation in Australia (Javaux et al., 2004) have complex surface ornamentation and probable excystment structures, which are persuasive evidence of a eukaryotic origin (Knoll et al., 2006). All of the purported older instances of eukaryotic body fossils lack such strong structural evidence, so microfossils indicate that ∼1.7 Ga is a robust earliest date for the appearance of eukaryotes. The oldest member of an identifiable extant group is the fossil *Bangiomorpha*, a rhodophyte alga from the ∼1.2-Ga Hunting Formation of Canada (Butterfield, 2000).

Relative ages – dating based on phylogenetics

A second approach to determining when eukaryotes arose is molecular dating, which allows divergence times to be estimated from genetic distances. Originally, these approaches relied on the assumption of a strict molecular clock, which postulated a constant rate of evolution over the entire phylogenetic tree and proposes that differences between homologous proteins of different species are proportional to their divergence time (Zuckerkandl and Pauling, 1965). However, variation in substitution rates has been widely documented, and 'relaxed' molecular clock methods have been developed that take into account that the rate of sequence evolution might vary across different branches (Ho and Phillips, 2009; Lepage et al., 2007; Welch and Bromham, 2005).

To estimate divergence times by using molecular clock approaches, the phylogenetic tree is calibrated with several known dates associated with the available paleobiological data. For ancient evolutionary events, calibrations are commonly based on the fossil record and, to a lesser extent, on biomarkers, as described above. Tree calibration also requires a robust phylogenetic tree. Luckily, the broad relationships between the main groups of eukaryotes have been better resolved in the past few years. Importantly, a number of lineages that were assigned as early, based on ribosomal (r)RNA trees, and were thought to retain 'primitive' characteristics, are now considered as highly derived, fast-evolving members of multiple lineages (Roger and Hug, 2006). This means that these lineages cannot be considered as proxies for the biology of earlier (and likely extinct) eukaryotes. Instead, the most recent view considers at least five major eukaryotic superphyla or supergroups, with a relatively well-resolved backbone in most clades (Adl et al., 2012; Burki et al., 2016) (Fig. 1B).

As our understanding of eukaryote phylogeny has improved, fossil-calibrated molecular-clock-based methods have been applied to date important diversification events (Berney and Pawlowski, 2006; Douzery et al., 2004; Hedges and Kumar, 2004; Hedges et al., 2001; Parfrey et al., 2011). However, these have yielded vastly different estimates. These discrepancies can be explained by a myriad of sources of variability and error due to various factors. Firstly, although the resolution in the tree of eukaryotes appears to be steadily improving (Burki et al., 2016), there is still uncertainty in the location of the root (Derelle et al., 2015; He et al., 2014). Secondly, controversy in assigning some of the Proterozoic fossils [i.e. from 2500 to 542 million years (megaannum, Ma) ago] to extant eukaryote groups suggests that molecular clock analyses rely heavily on extrapolation from the younger, but richer, Phanerozoic (less than 542 Ma ago) record. Thirdly, there are also inherent biases and uncertainties associated with assigning fossil calibrations to nodes in molecular phylogenies. These factors, combined with the variability in estimates and credible intervals yielded by different molecular clock model assumptions, have led to the wide ranges of estimated ages for LECA and the eukaryote supergroups that have been published in the last decade. The most recent analyses provide estimates for the age of LECA in the range of 1000 to 1600 Ma (Eme et al., 2014). Despite the uncertainty about the precise ages, these analyses define a relatively short time interval of ∼300 million years between the age of LECA and the emergence of all eukaryotic supergroups, which is consistent with rapid diversification events.

The relationship between Archaea and Eukaryota

Having an established timeframe by which eukaryogenesis took place puts the question of archaeal and eukaryotic relationships into focus, with major implications for the origins of eukaryotic cellular traits. Increased genome sequence data from a larger fraction of archaeal diversity, combined with improved phylogenetic methods have substantially changed our views of archaeal evolution (Brochier-Armanet et al., 2011). The traditional separation of the Archaea into Crenarchaeota and Euryarchaeota, as suggested by rRNA-based phylogeny and other criteria (Woese et al., 1990), has been blurred by the identification of new phyla, such as Thaumarchaeota and Korarchaeota, which possess a combination of both crenarchaeotal and euryarchaeotal features (Elkins et al., 2008). At the same time, technological progresses in obtaining genomic data from new and understudied microbial lineages, without the need to cultivate or isolate them, are progressively bringing to life a large fraction of microbial diversity colloquially known as 'microbial dark matter' (Rinke et al., 2013). Concerning the Archaea, this highlights a puzzling assemblage of uncultured lineages represented by very small cells and reduced genomes, which might reflect symbiotic or parasitic lifestyles (Castelle et al., 2015). It has been suggested that these lineages could form a new and deep-branching candidate archaeal phylum (DPANN) (Brown et al., 2015; Williams and Embley, 2014). However, as the clustering of fast-evolving lineages in molecular phylogenies is a well-known artefact (Gribaldo and Philippe, 2002), the branching of the DPANN in the archaeal tree, or even their very existence, is another issue that requires resolution.

Journal of Cell Science (2016) 0, 1-9 doi:10.1242/jcs.178566

There is now considerable evidence that eukaryotes emerged from within the Archaea, supporting the 'two domains of life' model (Fig. 2). Importantly, multiple approaches to reconstruct the relationship between eukaryotes and Archaea have also supported this view (Cox et al., 2008; Raymann et al., 2015). Furthermore, continued exploration of archaeal diversity has revealed the presence of homologues of components of typical eukaryotic features, in particular those related to the cytoskeleton (e.g. actin, tubulin), cytokinesis and/or membrane-remodelling systems (e.g. the ESCRT complex) (Makarova et al., 2010). Because the genes encoding these features are present in a 'patchwork' pattern across archaeal taxa, they have been referred to as a 'dispersed archaeal eukaryome' (Koonin and Yutin, 2014). This also suggests that the last archaeal common ancestor (LACA) might have been more complex than its known present-day descendants (Brochier-Armanet et al., 2011; Wolf et al., 2012). These features are particularly prominent in a clade uniting Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota, also called TACK (Guy and Ettema, 2011).

Excitingly, recent reconstructions suggest that the lineage that eventually led to the LECA can be more specifically pinpointed to being within this clade. Using metagenomics, Spang et al. have recently obtained the first genomic data from uncultured members of the deep-sea archaeal group (DSAG) lineage, which is related to the TACK superphylum, and proposed a new phylum called 'Candidatus Lokiarchaeota' (Spang et al., 2015). Their data suggest that Lokiarchaeota are the archaeal lineage closest to Eukaryota, and revealed the presence of a large set of eukaryotic signature proteins that previously had only been seen in diverse TACK members, and, once more, the presence of proteins related to the eukaryotic cytoskeleton components (Klinger et al., 2016; Spang et al., 2015).

Although the placement of Lokiarchaeota with respect to eukaryotes will need to be confirmed by further genomic data from this clade, these findings open great opportunities for investigating the specific origins of many cellular features deemed eukaryotic. In particular, priority is now given to isolating a member of Lokiarchaeota and to understanding the role of these eukaryotic-like characters in an archaeal cellular setting, as well as to further investigating that region of the archaeal tree (Fig. 2).

Towards revealing the prokaryotic origins of eukaryotic cell biology

The advances described above have produced a much clearer idea of where to look for the prokaryotic antecedents of eukaryotic cell biology. Progress in defining the sets of proteins that underpin such features in eukaryotic organelles and cellular systems brings the search for prokaryotic antecedents into the realm of a tractable bioinformatics problem (Koumandou et al., 2013 and references therein.

An obvious starting place is the origin of the nuclear envelope. In modern eukaryotes, the nuclear envelope is contiguous with the ER and punctuated by nuclear pores, which transport proteins and RNA between the cytoplasm and nuclear matrix. Importantly, the nuclear envelope comprises two membranes – the inner and outer membranes – which have distinct compositions, with the outer nuclear envelope membrane compositionally similar to the ER. This double membrane configuration is strong evidence that the nuclear envelope arose as a subdomain of an endomembrane compartment that also gave rise to the ER but with the function to enclose the genetic material, such that the inner nuclear envelope differentiated into a platform for chromosome organisation.

Additional evidence relating to the evolution of the nucleus comes from analysis of molecular machinery associated with the nuclear pore complex (NPC). A large macromolecular structure, the NPC is the gateway regulating all exchanges between the nucleoplasm and the cytoplasm. Many proteins that form the conserved core of the NPC are the result of gene duplication events, and a simpler NPC was probably present in pre-LECA species. Although some differences have recently been described for NPCs from divergent taxa (Obado et al., 2016), the key point is that the basic structure is both highly conserved and structurally related to a complement of proteins that share a domain structure of a β-propeller followed by an α-solenoid, collectively termed 'protocoatomers'. Because protocoatomers are components of the nuclear pore, the intraflagellar transport (IFT) machinery and protein coat complexes involved in the formation of transport vesicles and membrane deformation (Devos et al., 2014), the evolution of the nucleus, flagella and organelles of the endomembrane system are thus connected. Comparative genomics has shown that the LECA possessed a large set of NPC, IFT and coat proteins (Devos et al., 2004; Neumann et al., 2010; Schlacht and Dacks, 2015). Consequently, the expansion of this family must have taken place during transition from FECA to LECA, although the origin of the family and its expansion are likely to have begun before establishment of the FECA.

Although the protocoatomer family can be used to assess the deep evolution of the endomembrane system, it is not the only such protein family. Much of the machinery that defines organelle identity comprises interacting paralogous gene families, specifically SNAREs, GTPases and longins along with protocoatomers. It has been proposed that an iterative model of paralogue expansion and co-evolution of these interacting paralogues producing exclusive organelle and pathway-specific versions can explain the generation of new cellular compartments (Dacks and Field, 2007; Dacks et al., 2008). Again, comparative genomic and phylogenetic studies demonstrate that expansion of these families in the LECA had reached a level of sophistication that rivals that seen in many modern eukaryotes (Koumandou et al., 2013; Schlacht et al., 2014). The order of this expansion, however, represents another set of transitional events from FECA to LECA that remain to be resolved.

Analysis of these same paralogous families has provided links into the prokaryotic ancestry of the endomembrane system. The Lokiarchaeota phylum yields tantalizing insights as it features extensive complements of GTPases and the first reported presence of longin domains in a non-eukaryotic genome (Klinger et al., 2016). Longin domains are present in conserved eukaryotic protein superfamilies and had previously been believed to be exclusive to eukaryotes. However, despite the presence of these proteins in Archaea, there are no direct orthologues of the Rab and ARF GTPase sub-families that have specific cellular functions in eukaryotes (Klinger et al., 2016). The Lokiarchaeota composite genome also contains the most extensive group of archaeal homologues of the ESCRT machinery yet described, which has a conserved role in cytokinesis, in both archaeal and eukaryotic cells (Makarova et al., 2010), as well as functions in late endosomal trafficking in eukaryotes.

Another aspect that differentiates eukaryotic from prokaryotic cells is the cytoskeleton. Although bacteria use distant homologues of actin (MreB) to maintain cell shape, elaboration of a complex intracellular cytoskeleton is another event that clearly took place during the FECA-to-LECA transition. Similar to the case of the endomembrane system, we cannot currently reconstruct the route of evolution of these cytoskeleton proteins during the transition, and in

Journal of Cell Science • Advance article

this context, accurate phylogenetics is crucial. Tubulin-like proteins are present in the Verrucomicrobia (Pilhofer et al., 2007). There are many distinct actin-like families in archaea, for example Crenactin, restricted to the Korarchaeota, Aigarchaeota, Lokiarchaeota and some Crenarchaeota (Spang et al., 2015), and there are some cytoskeletal protein orthologues specific to Lokiarcheota (Klinger et al., 2016; Spang et al., 2015). Demonstration of the TACK archaea as the lineages potentially closest to the one that gave rise to the Eukarya has profound implications for how we view such evidence. Perhaps the most exciting implication for the presence of orthologues for both membrane-trafficking machinery and cytoskeleton proteins in these archaeal lineages is that FECA might have already possessed the genetic potential to develop phagocytosis. The origin of this trait was a key step in eukaryogenesis (Koonin, 2015; Poole and Gribaldo, 2014).

Of course, not all eukaryotic cell biology traces back to Archaea. The most widespread eukaryotic organelle with bacterial ancestry is the mitochondrion. An extensive set of organellar functions are clearly of direct ancestry from α-proteobacteria, including organelle maintenance and replication (Leger et al., 2015), much of aerobic energy metabolism (Müller et al., 2012, as well as others) and even some recent surprises about organelle morphology, such as the mitochondrial contact site (MICOS) complex that is responsible for cristae formation (Muñoz-Gómez et al., 2016). The α-proteobacterial contribution can also be seen in other processes, such as in Fe-S cluster formation (Barberà et al., 2010), β-oxidation of fatty acids (Bolte et al., 2015), the glycine cleavage system (Nývltová et al., 2015) and hemebiosynthesis (Cenci et al., 2016).

However, the issue is less clear when considering bacterial contributions, the origins of which are not strongly supported as being from α-proteobacteria. The only non-eukaryotic organisms possessing proteins with the 'protocoatomer' domain organisation are planctomycetes and their relatives (Santarella-Mellwig et al., 2010). There is currently no phylogenetic evidence supporting that these proteins are homologues to eukaryotic proteins, and thus, whether these represent convergent analogues is an open but tantalizing question. A recent large-scale analysis of genomic data has assessed the eukaryotic proteins that are likely to be of prokaryotic origin (Pittis and Gabaldón, 2016). Importantly, this study identified a category of genes of apparent bacterial origin but that was not clearly of α-proteobacterial origin, consistent with previous data (Rochette et al., 2014). By analysing the phylogenetic signal of these proteins, the study concluded a non-α-proteobacterial contribution to eukaryogenesis before mitochondrial endosymbiosis. Although the full implications of this study remain to be assessed by the field, it hints that information regarding the process of eukaryogenesis might be found in one or more bacteria other than that which constituted the mitochondrial ancestor (Pittis and Gabaldón, 2016).

One cellular feature that has puzzled evolutionary biologists aiming to resolve eukaryogenesis is lipid biosynthesis. Bacteria and eukaryotes both possess membranes composed of fatty acyl chains linked through ester bonds to a glycerol 3-phosphate backbone, whereas archaea possess ether-linked isoprene chain lipids on a glycerol 1-phosphate backbone. This 'lipid divide' was more easily explained when Archaea and eukaryotes were thought to be sister taxa (three-domain model) but requires a more complicated explanation under the currently supported hypothesis of eukaryotes arising from within Archaea (Lombard et al., 2012; López-García and Moreira, 2006). Scenarios to explain this have included involvement of a third prokaryotic contributor (Forterre, 2011) and contribution from the α-proteobacterium at the origin of mitochondria (Martin and Koonin, 2006). Very recent analyses provide evidence for a third option, that the 'lipid divide' is not as clear as it appears. Enzymes for production of fatty acyl chain lipids on a glycerol 3-phosphate backbone have been identified in a variety of archaeal genomes, including the Lokiarchaeota (Villanueva et al., 2016). Although these *in silico* predictions need to be confirmed with experimental characterisation, they raise the possibility that the FECA might have already possessed eukaryotic-type lipid membranes.

A remaining unresolved event in the period preceding LECA is where to place the points at which the nucleus and the mitochondrion were acquired. Energy considerations have led many in the field to favour the endosymbiotic event between an α-proteobacteria and the early host as a crucial early eukaryogenesis step (Lane and Martin, 2010, 2015). This is based on suggestions that the energy required to elaborate the complex eukaryotic cell is simply too costly to be sustained by amitochondrial metabolism. However, this notion can be countered by the very existence of eukaryotes that can support their complex cells in the absence of energy from mitochondria (Müller et al., 2012, and others) and the recent discovery that the oxymonad *Monocercomonoides* sp. has completely lost the organelle (Karnkowska et al., 2016). Moreover, recent energetics calculations have questioned the need for an extensive energy boost for expansion of the genome and proteome that accompanied eukaryogenesis (Lynch and Marinov, 2015). Taken together with the absence of a clear correlation between cell size and possession of a mitochondrion, as well as the increasing array of clear homologues of eukaryotic cellular components present in the host lineage, the possibility that the mitochondrion was acquired later must remain in consideration.

Conclusions

The view of eukaryogenesis as a biological 'quantum leap' that resulted in the rapid emergence of a vastly more sophisticated cell type has largely been overturned. Despite the remaining difficulties in elucidating the precise sequence of events, the discovery of features in Archaea formerly thought to be specific to eukaryotes, together with the vast number of changes needed in order to lead to the development of the LECA from a prokaryotic cell, have challenged our thinking. We still lack a clear concept of the timescale of the FECA-to-LECA transition, but we now have a better understanding of when the LECA must have arisen. Furthermore, the boundary between prokaryotes and eukaryotes has been blurred with the recognition of eukaryotic protein homologues that are involved in Archaea cellular processes, such as cytoskeleton formation, cell division and membrane trafficking. Based on these recent insights, a gradual climb towards eukaryotic cellular complexity and sophistication has emerged as being the driving force of eukaryogenesis.

The eukaryogenesis field has moved into a new phase. Previously, the best that could be achieved was to generate elegant, but ultimately highly speculative, models of eukaryogenesis. Resolving the order of transitions is challenging but remains one of the crucial questions concerning the steps of eukaryogenesis (Poole and Gribaldo, 2014). With the recent improvements in analysis of genomic data as well as the identification of novel microbial lineages, a more robust evidence-based scientific path can be forged.

Acknowledgements

This publication arose through discussions at a workshop run by The Company of Biologists at Wiston House, Steyning, West Sussex, UK in March 2015 (http://workshops.biologists.com/workshop-archive/workshop-march-2015/) with C.B.-A.

and D.P.D. as organizers. C.B.-A. and D.P.D. would like to thank The Company of Biologists and Journal of Cell Science for sponsoring the meeting and, in particular Nicky Le Blond for help with organisational aspects. A significant portion of this manuscript was also developed during the Program in Evolutionary Cell Biology, a symposium funded and hosted by the Kavli Institute for Theoretical Physics, supported in part by the National Science Foundation [grant number NSF PHY11-25915]. We thank Shelley Sazer and Mike Lynch for comments on the manuscript and suggestions.

Competing interests

The authors declare no competing or financial interests.

Funding

D.P.D. is supported by the Consejería de Economía, Innovación, Ciencia y Empleo, Junta de Andalucía C2A program; and the Spanish Ministerio de Economía y Competitividad [grant number BFU2013-40866-P]. J.B.D. is supported as the Canada Research Chair (Tier II; Government of Canada) in Evolutionary Cell Biology.

References

Adl, S. M., Simpson, A. G. B., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., Brown, M. W., Burki, F., Dunthorn, M., Hampl, V. et al. (2012). The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* **59**, 429-514.

Barberà, M. J., Ruiz-Trillo, I., Tufts, J. Y. A., Bery, A., Silberman, J. D. and Roger, A. J. (2010). Sawyeria marylandensis (Heterolobosea) has a hydrogenosome with novel metabolic properties. *Eukaryot. Cell* **9**, 1913-1924.

Berney, C. and Pawlowski, J. (2006). A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. *Proc. R. Soc. B Biol. Sci.* **273**, 1867-1872.

Bolte, K., Rensing, S. A. and Maier, U.-G. (2015). The evolution of eukaryotic cells from the perspective of peroxisomes: phylogenetic analyses of peroxisomal beta-oxidation enzymes support mitochondria-first models of eukaryotic cell evolution. *Bioessays* **37**, 195-203.

Brochier-Armanet, C., Forterre, P. and Gribaldo, S. (2011). Phylogeny and evolution of the Archaea: one hundred genomes later. *Curr. Opin. Microbiol.* **14**, 274-281.

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

Brown, C. T., Hug, L. A., Thomas, B. C., Sharon, I., Castelle, C. J., Singh, A., Wilkins, M. J., Wrighton, K. C., Williams, K. H. and Banfield, J. F. (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208-211.

Buick, R. (1984). Carbonaceous filaments from North Pole, Western Australia: are they fossil bacteria in Archaean stromatolites? *Precambrian Res.* **24**, 157-172.

Buick, R. (1988). Carbonaceous filaments from North Pole, Western Australia: are they fossil bacteria in archaean stromatolites? A reply. *Precambrian Res.* **39**, 311-317.

Buick, R. (2010). Early life: ancient acritarchs. *Nature* **463**, 885-886.

Burki, F., Kaplan, M., Tikhonenkov, D. V., Zlatogursky, V., Minh, B. Q., Radaykina, L. V., Smirnov, A., Mylnikov, A. P. and Keeling, P. J. (2016). Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc. R. Soc. B Biol. Sci.* **283**. doi: 10.1098/rspb.2015.2802.

Butterfield, N. J. (2000). Bangiomorpha pubescensn. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology* **26**, 386.

Castelle, C. J., Wrighton, K. C., Thomas, B. C., Hug, L. A., Brown, C. T., Wilkins, M. J., Frischkorn, K. R., Tringe, S. G., Singh, A., Markillie, L. M. et al. (2015). Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* **25**, 690-701.

Cenci, U., Moog, D., Curtis, B. A., Tanifuji, G., Eme, L., Lukeš, J. and Archibald, J. M. (2016). Heme pathway evolution in kinetoplastid protists. *BMC Evol. Biol.* **16**, e1004007.

Cox, C. J., Foster, P. G., Hirt, R. P., Harris, S. R. and Embley, T. M. (2008). The archaebacterial origin of eukaryotes. *Proc. Natl. Acad. Sci. USA* **105**, 20356-20361.

Dacks, J. B. and Doolittle, W. F. (2001). Reconstructing/deconstructing the earliest eukaryotes: how comparative genomics can help. *Cell* **107**, 419-425.

Dacks, J. B. and Field, M. C. (2007). Evolution of the eukaryotic membrane-trafficking system: origin, tempo and mode. *J. Cell Sci.* **120**, 2977-2985.

Dacks, J. B., Poon, P. P. and Field, M. C. (2008). Phylogeny of endocytic components yields insight into the process of nonendosymbiotic organelle evolution. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 588-593.

Derelle, R., Torruella, G., Klimeš, V., Brinkmann, H., Kim, E., Vlček, Č., Lang, B. F. and Eliáš, M. (2015). Bacterial proteins pinpoint a single eukaryotic root. *Proc. Natl. Acad. Sci. USA* **112**, E693-E699.

Devos, D., Dokudovskaya, S., Alber, F., Williams, R., Chait, B. T., Sali, A. and Rout, M. P. (2004). Components of coated vesicles and nuclear pore complexes share a common molecular architecture. *PLoS Biol.* **2**, e380.

Devos, D. P., Gräf, R. and Field, M. C. (2014). Evolution of the nucleus. *Curr. Opin. Cell Biol.* **28**, 8-15.

Douzery, E. J. P., Snell, E. A., Bapteste, E., Delsuc, F. and Philippe, H. (2004). The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proc. Natl. Acad. Sci. USA* **101**, 15386-15391.

Eigenbrode, J. L. and Freeman, K. H. (2006). Late Archean rise of aerobic microbial ecosystems. *Proc. Natl. Acad. Sci. USA* **103**, 15759-15764.

Elkins, J. G., Podar, M., Graham, D. E., Makarova, K. S., Wolf, Y., Randau, L., Hedlund, B. P., Brochier-Armanet, C., Kunin, V., Anderson, I. et al. (2008). A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc. Natl. Acad. Sci. USA* **105**, 8102-8107.

Eme, L., Sharpe, S. C., Brown, M. W. and Roger, A. J. (2014). On the age of eukaryotes: evaluating evidence from fossils and molecular clocks. *Cold Spring Harb. Perspect. Biol.* **6**, a016139.

Forterre, P. (2011). A new fusion hypothesis for the origin of Eukarya: better than previous ones, but probably also wrong. *Res. Microbiol.* **162**, 77-91.

French, K. L., Hallmann, C., Hope, J. M., Schoon, P. L., Zumberge, J. A., Hoshino, Y., Peters, C. A., George, S. C., Love, G. D., Brocks, J. J. et al. (2015). Reappraisal of hydrocarbon biomarkers in Archean rocks. *Proc. Natl. Acad. Sci. USA* **112**, 5915-5920.

Gogarten, J. P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E. J., Bowman, B. J., Manolson, M. F., Poole, R. J., Date, T. and Oshima, T. (1989). Evolution of the vacuolar H+-ATPase: implications for the origin of eukaryotes. *Proc. Natl. Acad. Sci. USA* **86**, 6661-6665.

Gribaldo, S. and Cammarano, P. (1998). The root of the universal tree of life inferred from anciently duplicated genes encoding components of the protein-targeting machinery. *J. Mol. Evol.* **47**, 508-516.

Gribaldo, S. and Philippe, H. (2002). Ancient phylogenetic relationships. *Theor. Popul. Biol.* **61**, 391-408.

Gribaldo, S., Poole, A. M., Daubin, V., Forterre, P. and 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.

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

He, D., Fiz-Palacios, O., Fu, C.-J., Fehling, J., Tsai, C.-C. and Baldauf, S. L. (2014). An alternative root for the eukaryote tree of life. *Curr. Biol.* **24**, 465-470.

Hedges, S. B. and Kumar, S. (2004). Precision of molecular time estimates. *Trends Genet.* **20**, 242-247.

Hedges, S. B., Chen, H., Kumar, S., Wang, D. Y., Thompson, A. S. and Watanabe, H. (2001). A genomic timescale for the origin of eukaryotes. *BMC Evol. Biol.* **1**, 4.

Ho, S. Y. W. and Phillips, M. J. (2009). Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Syst. Biol.* **58**, 367-380.

Javaux, E. J., Knoll, A. H. and Walter, M. (2003). Recognizing and interpreting the fossils of early eukaryotes. *Orig. Life Evol. Biosph.* **33**, 75-94.

Javaux, E. J., Knoll, A. H. and Walter, M. R. (2004). TEM evidence for eukaryotic diversity in mid-Proterozoic oceans. *Geobiology* **2**, 121-132.

Karnkowska, A., Vacek, V., Zubáčová, Z., Treitli, S. C., Petrželková, R., Eme, L., Novák, L., Žárský, V., Barlow, L. D., Herman, E. K. et al. (2016). A Eukaryote without a Mitochondrial Organelle. *Curr. Biol.* **26**, 1274-1284.

Klinger, C. M., Spang, A., Dacks, J. B. and Ettema, T. J. G. (2016). Tracing the archaeal origins of eukaryotic membrane-trafficking system building blocks. *Mol. Biol. Evol.* **33**, 1528-1541.

Knoll, A. H., Javaux, E. J., Hewitt, D. and Cohen, P. (2006). Eukaryotic organisms in Proterozoic oceans. *Philos. Trans. R. Soc. B. Biol. Sci.* **361**, 1023-1038.

Koonin, E. V. (2010). Preview. The incredible expanding ancestor of eukaryotes. *Cell* **140**, 606-608.

Koonin, E. V. (2015). Origin of eukaryotes from within archaea, archaeal eukaryome and bursts of gene gain: eukaryogenesis just made easier? *Philos. Trans. R. Soc. B. Biol. Sci.* **370**, 20140333.

Koonin, E. V. and Yutin, N. (2014). The dispersed archaeal eukaryome and the complex archaeal ancestor of eukaryotes. *Cold Spring Harb. Perspect. Biol.* **6**, a016188.

Koumandou, V. L., Wickstead, B., Ginger, M. L., van der Giezen, M., Dacks, J. B. and Field, M. C. (2013). Molecular paleontology and complexity in the last eukaryotic common ancestor. *Crit. Rev. Biochem. Mol. Biol.* **48**, 373-396.

Lane, N. and Martin, W. (2010). The energetics of genome complexity. *Nature* **467**, 929-934.

Lane, N. and Martin, W. F. (2015). Eukaryotes really are special, and mitochondria are why. *Proc. Natl. Acad. Sci. USA* **112**, E4823.

Leger, M. M., Petrů, M., Žárský, V., Eme, L., Vlček, Č., Harding, T., Lang, B. F., Eliáš, M., Doležal, P. and Roger, A. J. (2015). An ancestral bacterial division system is widespread in eukaryotic mitochondria. *Proc. Natl. Acad. Sci. USA* **112**, 10239-10246.

Lepage, T., Bryant, D., Philippe, H. and Lartillot, N. (2007). A general comparison of relaxed molecular clock models. *Mol. Biol. Evol.* **24**, 2669-2680.

Lombard, J., López-García, P. and Moreira, D. (2012). The early evolution of lipid membranes and the three domains of life. *Nat. Rev. Microbiol.* **10**, 507-515.
López-García, P. and Moreira, D. (2006). Selective forces for the origin of the eukaryotic nucleus. *Bioessays* **28**, 525-533.
López-García, P. and Moreira, D. (2015). Open questions on the origin of eukaryotes. *Trends Ecol. Evol.* **30**, 697-708.
Love, G. D. and Summons, R. E. (2015). The molecular record of Cryogenian sponges - a response to Antcliffe (2013). *Palaeontology* **58**, 1131-1136.
Love, G. D., Grosjean, E., Stalvies, C., Fike, D. A., Grotzinger, J. P., Bradley, A. S., Kelly, A. E., Bhatia, M., Meredith, W., Snape, C. E. et al. (2009). Fossil steroids record the appearance of Demospongiae during the Cryogenian period. *Nature* **457**, 718-721.
Lynch, M. and Marinov, G. K. (2015). The bioenergetic costs of a gene. *Proc. Natl. Acad. Sci. USA* **112**, 15690-15695.
Makarova, K. S., Wolf, Y. I., Mekhedov, S. L., Mirkin, B. G. and Koonin, E. V. (2005). Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res.* **33**, 4626-4638.
Makarova, K. S., Yutin, N., Bell, S. D. and Koonin, E. V. (2010). Evolution of diverse cell division and vesicle formation systems in Archaea. *Nat. Rev. Microbiol.* **8**, 731-741.
Martin, W. and Koonin, E. V. (2006). Introns and the origin of nucleus–cytosol compartmentalization. *Nature* **440**, 41-45.
Martin, W. F., Garg, S. and Zimorski, V. (2015). Endosymbiotic theories for eukaryote origin. *Philos. Trans. R. Soc. B. Biol. Sci.* **370**, 20140330.
Mclnerney, J. O., Martin, W. F., Koonin, E. V., Allen, J. F., Galperin, M. Y., Lane, N., Archibald, J. M. and Embley, T. M. (2011). Planctomycetes and eukaryotes: a case of analogy not homology. *Bioessays* **33**, 810-817.
Müller, M., Mentel, M., van Hellemond, J. J., Henze, K., Woehle, C., Gould, S. B., Yu, R.-Y., Giezen, M., Tielens, A. G. M. and Martin, W. F. (2012). Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol. Mol. Biol. Rev.* **76**, 444-495.
Muñoz-Gómez, S. A., Slamovits, C. H., Dacks, J. B. and Wideman, J. G. (2016). The evolution of MICOS: ancestral and derived functions and interactions. *Commun. Integr. Biol.* **8**, e1094593.
Neumann, N., Lundin, D. and Poole, A. M. (2010). Comparative genomic evidence for a complete nuclear pore complex in the last eukaryotic common ancestor. *PLoS ONE* **5**, e13241.
Neumann, S., Wessels, H. J. C. T., Rijpstra, W. I. C., Sinninghe Damsté, J. S., Kartal, B., Jetten, M. S. M. and van Niftrik, L. (2014). Isolation and characterization of a prokaryotic cell organelle from the anammox bacterium K uenenia stuttgartiensis. *Mol. Microbiol.* **94**, 794-802.
Nijman, W., de Bruijne, K. H. and Valkering, M. E. (1998). Growth fault control of Early Archaean cherts, barite mounds and chert-barite veins, North Pole Dome, Eastern Pilbara, Western Australia. *Precambrian Res.* **88**, 25-52.
Nývltová, E., Stairs, C. W., Hrdý, I., Rídl, J., Mach, J., Pačes, J., Roger, A. J. and Tachezy, J. (2015). Lateral gene transfer and gene duplication played a key role in the evolution of Mastigamoeba balamuthi hydrogenosomes. *Mol. Biol. Evol.* **32**, 1039-1055.
Obado, S. O., Brillantes, M., Uryu, K., Zhang, W., Ketaren, N. E., Chait, B. T., Field, M. C. and Rout, M. P. (2016). Interactome mapping reveals the evolutionary history of the nuclear pore complex. *PLoS Biol.* **14**, e1002365.
O'Malley, M. A. (2010). The first eukaryote cell: an unfinished history of contestation. *Stud. Hist. Philos. Biol. Biomed. Sci.* **41**, 212-224.
Parfrey, L. W., Lahr, D. J. G., Knoll, A. H. and Katz, L. A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc. Natl. Acad. Sci. USA* **108**, 13624-13629.
Pawlowska, M. M., Butterfield, N. J. and Brocks, J. J. (2013). Lipid taphonomy in the Proterozoic and the effect of microbial mats on biomarker preservation. *Geology* **41**, 103-106.
Pilhofer, M., Rosati, G., Ludwig, W., Schleifer, K.-H. and Petroni, G. (2007). Coexistence of tubulins and ftsZ in different Prosthecobacter species. *Mol. Biol. Evol.* **24**, 1439-1442.
Pittis, A. A. and Gabaldón, T. (2016). Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature* **531**, 101-104.
Poole, A. M. and Gribaldo, S. (2014). Eukaryotic origins: how and when was the mitochondrion acquired? *Cold Spring Harb. Perspect. Biol.* **6**, a015990.
Rasmussen, B., Fletcher, I. R., Brocks, J. J. and Kilburn, M. R. (2008). Reassessing the first appearance of eukaryotes and cyanobacteria. *Nature* **455**, 1101-1104.
Raymann, K., Brochier-Armanet, C. and 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.
Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J.-F., Darling, A., Malfatti, S., Swan, B. K., Gies, E. A. et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431-437.
Rivera, M. C., Jain, R., Moore, J. E. and Lake, J. A. (1998). Genomic evidence for two functionally distinct gene classes. *Proc. Natl. Acad. Sci. USA* **95**, 6239-6244.
Rochette, N. C., Brochier-Armanet, C. and Gouy, M. (2014). Phylogenomic test of the hypotheses for the evolutionary origin of eukaryotes. *Mol. Biol. Evol.* **31**, 832-845.
Roger, A. J. and Hug, L. A. (2006). The origin and diversification of eukaryotes: problems with molecular phylogenetics and molecular clock estimation. *Philos. Trans. R. Soc. B Biol. Sci.* **361**, 1039-1054.
Santarella-Mellwig, R., Franke, J., Jaedicke, A., Gorjanacz, M., Bauer, U., Budd, A., Mattaj, I. W. and Devos, D. P. (2010). The compartmentalized bacteria of the planctomycetes-verrucomicrobia-chlamydiae superphylum have membrane coat-like proteins. *PLoS Biol.* **8**, e1000281.
Schlacht, A. and Dacks, J. B. (2015). Unexpected ancient paralogs and an evolutionary model for the COPII coat complex. *Genome Biol. Evol.* **7**, 1098-1109.
Schlacht, A., Herman, E. K., Klute, M. J., Field, M. C. and Dacks, J. B. (2014). Missing pieces of an ancient puzzle: evolution of the eukaryotic membrane-trafficking system. *Cold Spring Harb. Perspect. Biol.* **6**, a016048.
Schulz, I., Baumann, O., Samereier, M., Zoglmeier, C. and Gräf, R. (2009). Dictyostelium Sun1 is a dynamic membrane protein of both nuclear membranes and required for centrosomal association with clustered centromeres. *Eur. J. Cell Biol.* **88**, 621-638.
Shen, Y., Buick, R. and Canfield, D. E. (2001). Isotopic evidence for microbial sulphate reduction in the early Archaean era. *Nature* **410**, 77-81.
Shen, Y., Farquhar, J., Masterson, A., Kaufman, A. J. and Buick, R. (2009). Evaluating the role of microbial sulfate reduction in the early Archean using quadruple isotope systematics. *Earth Planet. Sci. Lett.* **279**, 383-391.
Simon, M., Jardillier, L., Deschamps, P., Moreira, D., Restoux, G., Bertolino, P. and López-García, P. (2015). Complex communities of small protists and unexpected occurrence of typical marine lineages in shallow freshwater systems. *Environ. Microbiol.* **17**, 3610-3627.
Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., van Eijk, R., Schleper, C., Guy, L. and Ettema, T. J. G. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173-179.
Stüeken, E. E., Buick, R., Guy, B. M. and Koehler, M. C. (2015). Isotopic evidence for biological nitrogen fixation by molybdenum-nitrogenase from 3.2 Gyr. *Nature* **520**, 666-669.
Summons, R. E. and Walter, M. (1990). Molecular fossils and microfossils of prokaryotes and protists from Proterozoic sediments. *Am. J. Sci.* **290**, 212-244.
Ueno, Y., Yamada, K., Yoshida, N., Maruyama, S. and Isozaki, Y. (2006). Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. *Nature* **440**, 516-519.
Ueno, Y., Ono, S., Rumble, D. and Maruyama, S. (2008). Quadruple sulfur isotope analysis of ca. 3.5 Ga Dresser Formation: new evidence for microbial sulfate reduction in the early Archean. *Geochim. Cosmochim. Acta* **72**, 5675-5691.
Van Kranendonk, M. (2006). Volcanic degassing, hydrothermal circulation and the flourishing of early life on Earth: a review of the evidence from c. 3490-3240 Ma rocks of the Pilbara Supergroup, Pilbara Craton, Western Australia. *Earth Sci. Rev.* **74**, 197-240.
Van Kranendonk, M., Philippot, P., Lepot, K., Bodorkos, S. and Pirajno, F. (2008). Geological setting of Earth's oldest fossils in the ca. 3.5 Ga Dresser Formation, Pilbara Craton, Western Australia. *Precambrian Res.* **167**, 93-124.
Villanueva, L., Schouten, S. and Damsté, J. S. (2016). Phylogenomic analysis of lipid biosynthetic genes of Archaea shed light on the "lipid divide". *Environ. Microbiol.* doi: 10.1111/1462-2920.13361.
Welch, J. J. and Bromham, L. (2005). Molecular dating when rates vary. *Trends Ecol. Evol.* **20**, 320-327.
Williams, T. A. and Embley, T. M. (2014). Archaeal "dark matter" and the origin of eukaryotes. *Genome Biol. Evol.* **6**, 474-481.
Woese, C. R. and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. USA* **74**, 5088-5090.
Woese, C. R., Kandler, O. and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* **87**, 4576-4579.
Wolf, Y. I., Makarova, K. S., Yutin, N. and Koonin, E. V. (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.
Yan, Y. and Liu, Z. L. (1993). Significance of eukaryotic organisms in the microfossil flora of Changcheng system. *Acta Micropalaeontol. Sin.* **10**, 167-180.
Zhaxybayeva, O., Lapierre, P. and Gogarten, J. P. (2005). Ancient gene duplications and the root(s) of the tree of life. *Protoplasma* **227**, 53-64.
Zuckerkandl, E. and Pauling, L. (1965). Molecules as documents of evolutionary history. *J. Theor. Biol.* **8**, 357-366.