Bulletin of Regional Natural History
Formerly Bollettino della Società dei Naturalisti in Napoli

The unlikely cell: origins and diversification of eukaryotes

Roberto Ligrone

DOI https://doi.org/10.6093/2724-4393/9380

***Correspondence:**
roberto.ligrone@gmail.com
https://orcid.org/0000-0003-4857-8162

Affiliation:
Dipartimento di Scienze e Tecnologie Ambientali, Biologiche e Farmaceutiche, Università della Campania Luigi Vanvitelli, Italy

Conflict of Interest: The author declares that he has no conflict of interest.

Financial Disclosure Statement: The Author declares that no specific funding was received for this work.

Submitted: 24 Apr 2022
Accepted: 24 Jun 2022

Associate Editor: Antonino Pollio

Abstract

Extant eukaryotes are a monophyletic lineage sharing a set of unique cellular and molecular traits inherited from a last common ancestor (LECA). There are no known intermediates between the eukaryotic and prokaryotic cellular organization. In contrast, the eukaryote pangenome has a chimeric structure combining eukaryote-specific genes and genes with homologs in bacteria and archaea, with bacterial genes only in part acquired via the mitochondrial symbiosis. At odd with the long-held view of a sister relationship between the archaea and eukaryotes, more recent phylogenomic work places the eukaryotes within the archaea and the root of the tree of life between the archaea and bacteria, thus supporting a 2-Domain tree of life. Challenging the traditional endosymbiotic scenario of eukaryogenesis, this novel phylogenetic paradigm has prompted hypotheses of archaeal-bacterial symbiosis with emphasis on the timing and impact of mitochondrial evolution (mitochondrion-first vs. mitochondrion-later models). Phylogenomic analysis has resolved the extant eukaryotic diversity into two major clades, the Amorphea and Diaphoretickes, leaving out several minor taxa listed as *incertae sedis*. The root of the eukaryote tree is still undefined. The chloroplast primarily evolved in the unicellular ancestor of Archaeplastida (Plantae) from a cyanobacterial endosymbiont, and was then transferred horizontally to other eukaryotic lineages by further events of endosymbiosis. Molecular-clock analysis integrated with paleontological evidence dates the appearance of eukaryotes to at least 1.6 billion years ago. LECA is consistently dated to about 1.2 billion years ago, and extant lineages from about 1 billion years ago onwards. The

diffusion of eukaryotes in the Neoproterozoic and Phanerozoic enhanced global primary production by orders of magnitude, led to the development of ecosystems of unprecedented complexity, and drove the planetary shift to a highly oxygenated condition.

Keywords: Eukaryogenesis, Eukaryotes, Phylogeny, Tree of life

Riassunto

Gli eucarioti sono un gruppo monofiletico con tratti cellulari e molecolari unici, ereditati da un progenitore comune. Non sono note forme intermedie fra l'organizzazione cellulare procariotica e quella eucariotica. Sorprendentemente, tuttavia, il pangenoma degli eucarioti è chimerico, combinando geni unicamente eucariotici con geni omologhi a geni di archei e batteri, questi ultimi solo in parte acquisiti attraverso la simbiosi mitocondriale. In contrasto con la filogenesi tradizionale che tratta archei ed eucarioti come cladi gemelli derivati da un comune progenitore, recenti analisi filogenomiche risolvono gli eucarioti come un clade interno agli archei e pongono la radice dell'albero della vita fra archei e batteri, riducendo i Domini della vita da tre a due. Questo paradigma filogenetico ha stimolato la formulazione di nuovi modelli di eucariogenesi centrati su ipotesi di simbiosi tra batteri e archei, con enfasi sul momento di apparizione e l'impatto evolutivo del mitocondrio. L'analisi filogenomica ripartisce gli eucarioti esistenti in due grandi cladi, gli *Amorphea* e i *Diaphoretickes*, lasciando fuori vari taxa minori annotati come *incertae sedis*. La posizione della radice nell'albero degli eucarioti rimane incerta. Il cloroplasto si è primariamente evoluto nel progenitore degli Archaeplastida (Plantae) da un cianobatterio endosimbiotico, ed è stato poi trasmesso orizzontalmente ad altre linee di eucarioti. Studi paleontologici e molecolari stimano che gli eucarioti siano apparsi almeno 1,6 miliardi di anni fa. L'ultimo comune progenitore degli eucarioti moderni verosimilmente apparve intorno a 1,2 miliardi di anni fa, mentre i cladi esistenti avrebbero iniziato a divergere circa 1,0 miliardo di anni fa. La diffusione degli eucarioti nel Neoproterozoico e poi nel Fanerozoico ha accresciuto la produttività primaria globale di ordini di grandezza, ha enormemente espanso la complessità degli ecosistemi, e ha spinto il pianeta verso uno stato altamente ossigenato.

Parole chiave: Albero della Vita, Eucarioti, Eucariogenesi, Filogenesi

How to cite

R. Ligrone. (2022). The unlikely cell: origins and diversification of eukaryotes. Bulletin of Regional Natural History (BORNH), Bollettino della Società dei Naturalisti in Napoli. Vol.2, n. 2, pp. 1- 43. ISSN: 2724-4393.

List of abbreviations

DG1P: D-glycerol-1-phosphate
ER: endoplasmic reticulum
eTOL: eukaryotic tree of life
FECA: first eukaryotic common ancestor
GY: billion years
GYA: billion years ago
LACA: last archaeal common ancestor
LECA: last eukaryotic common ancestor
LG3P: L-glycerol-3-phosphate
MYA: million years ago
TOL: tree of life

1. Introduction

The origins of eukaryotes is one of the hottest topics in modern biology. The separation between organisms with nucleate and anucleate cells was first recognized by Haeckel (1866). The terms *prokaryotes* and *eukaryotes* for the two types of cells were informally introduced by the French protozoologist Edouard Chatton (Chatton 1925) and rediscovered by microbiologists Roger Stanier and C.B. van Neil thirty-seven years later (Stanier & van Niel 1962). Thenceforth the prokaryote-eukaryote dichotomy was universally accepted as the primary divide in the biological world until rRNA phylogenetics revitalized microbial systematics in the 1970s. Ribosomal RNA phylogeny and congruent biochemical divergences replaced the previous bifurcation of life with the three domains of Eubacteria, Archaebacteria and Eucarya (Woese & Fox 1976) (Fig. 1A). When rRNA-based phylogeny revealed that the Archaebacteria are more closely related to the Eucarya that the Eubacteria, Woese et al. (1990) proposed the terms Bacteria and Archaea to replace Eubacteria and Archaebacteria. The spelling of Eucarya is etymologically incorrect and the term has later been replaced with Eukarya or Eukaryota. In this paper we will normally use the informal terms bacteria, archaea, and eukaryotes.

The phylogenetic interrelationships of the three domains have been problematic from the beginning. Bacteria and archaea are both prokaryotes, and there is no known transitional form between the prokaryotic and eukaryotic cellular organization. Nevertheles, research in the late 1980s revealed important molecular signatures shared by the archaea and eukaryotes, pointing to a sister relationship of the two domains (Fig. 1B). In parallel, phylogenomic work sorted the expanding diversity of known archaeal taxa into two sister clades, the Euryarchaeota and Chrenarchaeota (Woese et al. 1990). The molecular topology of membrane lipids is perhaps the greater obstacle to molecular evidence of a closer relationship of eukaryotes with the archaea than with bacteria. Membrane lipids in archaea are made of branched isoprenoids chains ether-bound to D-glycerol-1-phosphate, whereas in bacteria and

eukaryotes they consist of linear aliphatic chains ester-bound to L-glycerol-3-phosphate (De Rosa et al. 1986; Lombard et al. 2012; Balleza et al. 2014). Although isoprenoid compounds also occur in bacteria and eukaryotes, and play important roles in the biochemistry of these organisms, only the archaea use isoprenoid chains to make membrane phospholipids. Mainstream models on the origins of eukaryotes offer a range of divergent explanations to this conundrum. Besides membranes, the archaea have other unique traits that keep them apart from bacteria and eukaryotes. These include unique metabolic patways such as methanogenesis, unique enzymes such as specific DNA topoisomerases and DNA polymerases, and unique cell surface structures (Gribaldo et al. 2010).

A second issue that raises phylogenetic problems is the chimeric nature of the eukaryotic pangenome (Glossary). Besides genes unique to eukaryotes, this encompasses genes with homologues in archaea and bacteria. Genomic studies initially suggested that archaeal and bacterial genes accounted for over 50% of the eukaryotic pangenome (Koonin 2010). More recent work based on more stringent criteria has confirmed chimerism but has dramatically lowered the estimation. A survey of 14 eukaryotes representative of the main eukaryotic lineages, 52 bacteria, and 52 archaea reported an average of about 4.8% of genes with bacterial homology, 2.1% of genes with archaeal homology, 2% of genes with ambiguous attribution, and 91% of eukaryote-specific genes (Alvarez-Ponce et al. 2013). An even wider survey, encompassing 19,050,992 protein sequences from 5,655 bacterial and 212 archaeal genomes and 3,420,731 protein sequences from 150 eukaryotic genomes, suggests that the eukaryote pangenome encompasses only about 1% of protein-encoding genes with prokaryotic homologues, with bacterial genes slightly prevailing over archaeal genes (Brueckner & Martin 2020).

Starting in the late 1970s, evidence of the prokaryotic nature of mitochondrial and plastid DNA demonstrated beyond any possible doubt that these organelles have an endosymbiotic origin, thus offering an explanation for the bacterial component of the eukaryote pangenome. Yet, genes of alfaproteobacterial ancestry (the bacterial group probably encompassing the progenitor of the mitochondrion) account for only a fraction of nuclear genes with bacterial affinity in non-photosynthetic eukaryotes (Koonin 2010; Vosseberg et al. 2020). The occurrence of a sizable number of bacterial genes not of proteobacterial origin, yet involved in mitochondrial maintenance, suggests that symbioses with other types of bacteria preceded the evolution of the mitochondrion (Roger 2017). An alternative explanation is that non-proteobacterial genes were incorporated by horizontal gene transfer in the genome of the mitochondrial progenitor before endosymbiosis (Ku et al. 2015a).

Early evidence suggesting that the archaea and eukaryotes are sister groups, as well as more recent phylogenetic trees in which the eukaryotes stem from within the archaea explain the occurrence of “archaeal” genes in the eukaryote pangenome in terms of vertical inheritance. Less straightforward is explaining why eukaryotic genes with bacterial affinity mainly control metabolic functions (“house-keeping” or “operational” genes), whereas those with archaeal affinity are mostly involved in informational processes, viz. DNA replication, transcription, and repair (Thiegardt et al. 2012).

Eukaryotes are defined by a vast suite of unique traits (Table 1). Despite traditional focus on the nucleus, the most distinguishing feature of eukaryotes is probably phagotrophy, the ability of hunting and

Table 1. Major traits shared by extant eukaryotes barring secondary losses. The list only includes traits that presumably evolved in a common ancestor before extant eukaryote lineages diverged, thus excluding later additions such as the chloroplast or intermediate filaments. Adapted from Cavalier-Smith (2009).

1.Actin and actin-related proteins (Arps) functioning as actin nucleators
2.Myosin, microfilament mediated intracellular transport and ameboid movement
3.α-, β-, γ-tubulin, the last functioning in microtubule nucleation
4.Molecular motors associated with microtubules (dyneins e kinesins)
5.Phagocytosis
6.Mitochondria
7.Genome amplification by addition of extensive non-coding sequences
8.Protoplasmic volume controlled by a single genome (energid) larger than in prokaryotes
9.Endomembrane system and biochemical machinery for vesicular transport
10.DNA enclosed in nuclear envelope: transcription spatially and temporally separate from translation
11.Chromatin (nucleosomes)
12.Linear chromosomes with telomeres synthesized by telomerases
13.Nucleolus
14.80 S ribosomes nearly twice as massive as archaeal or bacterial ribosomes
15.Mitosis, centromeres, kinetochores
16.Cell division mediated by actin, not FtsZ
17.Meiosis and synaptonemal complex
18.Sexual reproduction
19.Flagella with an inner skeleton of 9+2 microtubule pairs
20.Peroxisomes
21.Sphingolipids
22.Phosphatidylinositol
23.Sterol synthesis
24.Calmodulin
25.Ubiquitin and polyubiquitin labelling system
26.26S proteasomes with 19S regulatory subunit
27.Spliceosomal introns and spliceosomes
28.mRNA capping by 7-methyl-guanosine-triphosphate at 5' end
29.Three RNA polymerases (I, II, III) with distinct functions
30.RNA-interference machinery
31.Cell cycle resetting by anaphase proteolysis

ingesting other cells (Leander 2020). The evolution of phagotrophy was made possible by a cytoskeleton controlled by nucleating factors and molecular motors, which is by itself another distinctive property of the eukaryotic cell (Theriot 2013).

The universal occurrence of such a large set of unique traits, most of which extremely complex, indicates that they were inherited from a common ancestor, dubbed LECA (Poole & Newmann 2011; Schlacht et al. 2014; Koreny & Field 2016).

Glossary

Asgard: a novel archaeal lineage recognized by phylogenetic analysis of metagenomic sequences and resolved in some trees as the sister group to eukaryotes.

Clade: a taxonomic group encompassing an ancestor and all its descendants, to the exclusion of any other organism.

C-paradox: genome size does not correlate with organismal complexity in eukaryotes. The terms C-DNA or C-value indicate the amount of DNA in a haploid genome.

Effective population size: the number of individuals in an idealized population necessary for the behaviour of a specified parameter in simulation experiments to mirror the pattern observed in real populations.

ESCRT: acronym from *Endosomal Sorting Complexes Required for Transport*. A machinery made up of cytosolic protein complexes known as ESCRT-0, ESCRT-I, ESCRT-II, and ESCRT-III, which controls membrane bending/budding in eukaryotic cells. ESCRT-III controls membrane constriction and therefore participates in cell division. Cell division in bacteria depends on the FtsZ-based system, whereas the archaea use homologues of either the bacterial or eukaryotic system, with some taxa seemingly having a third novel system containing an actin-like protein (Makarova et al. 2010).

Metagenomics: sequence analysis of genetic material recovered directly from environmental samples. The metagenomic approach permits the identification of novel taxa without the need for isolation and cultivation, revealing that the vast majority of microbial biodiversity has been missed by cultivation-based methods.

Monophyletic group: see clade

Pangenome. The full complement of genes present in a taxonomic group, including genes not shared by all individuals. The pangenome concept does not consider allelic variants of the same gene. When referred to taxonomic groups above the species level, for example a phylum, kingdom or domain, the pangenome encompasses all homologous gene variants present in the whole spectrum of organisms belonging to that group (see Ligrone 2021 for further details).

Paraphyletic group: a taxonomic group encompassing an ancestor and only a part of its descendants

Proteome: the entire set of proteins that is expressed by a genome, cell, tissue, or organism at a certain time and under certain conditions.

Protists: unicellular eukaryotes

Sister groups: two lineages diverged from a common ancestor.

Syntrophy: an obligate mutualistic association between different types of organisms, in which the growth of each partner depends on the metabolic activity of the other(s). Syntrophy plays a major role in microbial ecological interactions

The evolutionary process that led to the emergence of modern eukaryotes, known as *eukaryogenesis*, spanned the interval between the appearance of the First Eukaryotic Common Ancestor (FECA) and LECA. It is generally agreed that the eukaryotic grade of cellular organization must have arisen from a prokaryotic grade, and that endosymbiosis, the bringing together of distinct cells one inside the other, has had a central role in eukaryogenesis. Yet, despite enormous progress in molecular and cellular research, fifty years after the first sequence data were analysed there is still little consensus about the pathway of eukaryogenesis. In contrast, the expansion of genome sequencing in the last two decades has fostered dramatic progress in the reconstruction of the eukaryote tree of life. Molecular phylogeny has sorted extant eukaryote diversity into two major branches, the Amorphea and the Diaphoretickes, leaving out a few minor lineages of uncertain position (Adl et al. 2018). Mirroring the absence of intermediates in cellular organization,

sequence analysis has so far failed to pinpoint the root of the eukaryote tree, viz. the basalmost extant eukaryotic lineage. Likewise, the position of several branches both within and outside major groups remains doubtful.

Hundreds of papers have been written on eukaryogenesis just after the turn of the century, and novel work continues to appear almost weekly. To date, however, no proposed scenario appears to be fully consistent with all the available data. The present paper reviews competing hypotheses of eukaryogenesis and the current insight into the evolutionary history of modern eukaryotes.

2. Three Domains of life, or only two?

After the discovery of the archaea as a prokaryotic lineage distinct from bacteria, phylogenomic work led to the recognition of two groups with the taxonomic rank of kingdoms, the Crenarchaeota and the Euryarchaeota, the first encompassing only hyperthermophilic forms, the latter with a diversity of phenotypes including hyperthermophilic, mesophilic, methanogenic, and halophilic forms (Woese et al. 1990).

At the end of the past century, sequence analysis of protein translation elongation factors in archaea, bacteria and eukaryotes produced a novel tree in which eukaryotic sequences branched from within the Crenarchaeota. This revived the "eocyte" hypothesis presented by James Lake in 1984 and originally based on the observation that the ribosomes of Crenarchaeota and eukaryotes were more similar in shape to each other than to ribosomes in the bacteria or Euryarchaeota (Lake et al. 1984). The eocyte model posits that the eukaryotes emerged from within the Crenarcheota (dubbed eocytes, i.e. ancient cells), implying that the Archaea is a paraphyletic group (Glossary). A critical examination of the evidence for and against the eocyte hypothesis supported topologies consistent with the eocyte scenario (Cox et al. 2008). In contrast, phylogenomic work using a large data set (Yutin et al. 2008) did not confirm any special affinity of eukaryotes with the Chrenarchaeota or the Euryarchaota, suggesting instead that the eukaryotes originated from an archaeal branch outside the archaeal diversity known at that time.

In the following decade, metagenomic analysis (Glossary) of samples from a diversity of sites worldwide revealed a vast assemblage of uncultured archaeal lineages that were accommodated in three novel phyla named Thaumarchaeota, Aigarchaeota and Korarchaeota. Phylogenomic analysis placed together these novel phyla and the Crenarchaeota in the "TACK" superphylum (Guy and Ettema 2011). Single-cell genomics, a novel method consisting in the amplification and sequencing of DNA extracted from single cells, supported the TACK group and revealed other novel archaeal lineages with extremely small cellular sizes. These clustered into a single clade with the rank of superphylum, named "DPANN" from the initials of the first groups discovered, the Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota (Rinke et al. 2013) (Fig. 1C). A study based on the classical ribosomal protein data set (32 proteins) plus other 38 conserved proteins placed the root of the archaeal tree between the Euryarchaeota (including the Nanoarchaeota) and the other known archaeal taxa, which nested together in a large clade named Proteoarchaeota (Petitjean et al. 2015).

Almost simultaneously, metagenomic analysis of marine sediments from Loki Castle, a site near the Mid-Atlantic Ridge in the Arctic Ocean, revealed a novel group of

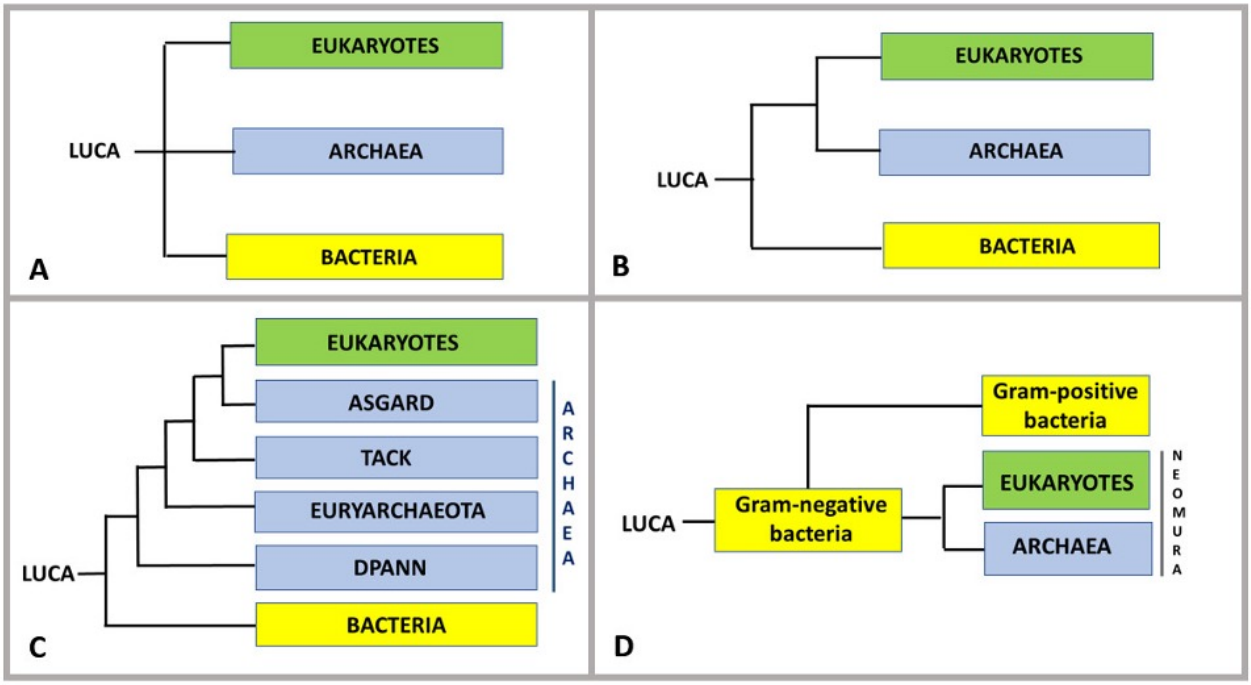

Figure 1: (A) Early version of the Tree of Life, with three Domains diverging from the LUCA (Woese and Fox 1976). (B) The archaea and eukaryotes are sister groups and form together a sister clade to the bacteria (Woese 1990; Ciccarelli et al. 2006). (C) Current version of the eocyte model: the eukaryote lineage diverged from within the archaea, in a sister position to the Asgard. (D) Neomuran model (Cavalier-Smith & Chao 2020): the archaea and eukaryotes are sister groups and diverged relatively late from within gram-negative bacteria. The eocyte model implies that the Archaea is paraphyletic; the neomuran model implies that the Bacteria is paraphyletic. See Glossary for a definition of the terms used.

psychrophilic (cold-adapted) archaea that were dubbed Lokiarchaeota (Spang et al. 2015; Klinger et al. 2016). Slightly later, metagenomic analysis of aquatic sediments from numerous sites worldwide brought to light other lineages named Odinarchaeota, Thorarchaeota and Heimdallarchaeota. These clustered together in a novel clade dubbed the Asgard, which was resolved as the sister group to TACK in most trees (Zaremba-Niedzwiedzka et al. 2017). The Asgard genome encompasses sequences homologous with eukaryotic genes including the replication initiation complex, ubiquitin, histones, actin, tubulin, and ESCRT-III. In addition, the Asgard genome encodes proteins showing clear homology with actin-related proteins Arp 2 and 3, and with the large family of small GTPases that in eukaryotes play key functions in the regulation of the cytoskeleton, cell motility, compartment identity, and intracellular vesicle traffic (Eme et al. 2017; Spang et al. 2015, 2018; Zaremba-Niedzwiedzka et al. 2017). Sequence homologs of important eukaryotic genes also occur in TACK archaea not belonging to the Asgard lineage, including genes for three ribosomal proteins, two sub-units of RNA polymerase, and the transcription factor Elf1 (Saw et al. 2015). The TACK, DPANN and Asgard superphyla continue expanding for the inclusion of other newly discovered taxa (Baker et al. 2020).

Imachi et al. (2020) for the first time isolated and cultured an Asgard archaeon from deep-sea methane-seep sediment of the Nankai Trough (Japan). Provisionally named Candidatus Prometheoarchaeum syntrophicum, this is an extremely slow-

growing anaerobic microorganism that lives degrading amino acids and producing hydrogen and/or formate by-products. Its cells are cocci with an average diameter of 550 nm, lack internal differentiation, and produce superficial membrane-bound blebs and branched projections. *P. syntrophicum* associates in nature with the sulfate-reducing bacterium Halodesulfovibrio and/or the methane-producing archaeon Methanogenium, which oxidize hydrogen and formate using sulfate or carbon dioxide as electron acceptors, respectively. This is an instance of syntrophy (Glossary) in which the removal of hydrogen/formate by-products is essential for the oxidation of organic substrates by the archaeon to be energetically convenient. In addition, P. syntrophicum obtains a diversity of vitamins from its syntrophic associates (Imachi et al. 2020).

The inclusion of TACK archaea in deep-branch phylogenetics consistently produced a two-Domain tree of life with the eukaryotes nested within the TACK (Spang et al. 2015; Hugh et al. 2016; Zaremba-Niedzwiedzka et al. 2017). After the discovery of the Asgard lineage, phylogenetic analyses including Asgard sequences produced a 2-Domain tree of life (TOL) with the eukaryotes sister to the Asgard and the origin between the archaea and bacteria (Spang et al. 2018; Williams et al. 2020) (Fig. 1C). Besides confirming a 2-Domain TOL, Raymann et al. (2015) placed the archaeal root in the Euryarchaeota, that were resolved as a paraphyletic group. At odd with these results, phylogenetic work based on RNA polymerase large subunit strongly supported a 3-Domain tree of life (i.e., with the eukaryotes branching outside the archaea) and resolved the Asgards as sister to the Euryarchaeota, not the TACK (da Cunha et al. 2018). The same study challenged the tree topology featuring the Asgards as sister to eukaryotes as an artifact due to sequence contamination, the inclusion of fast-evolving lineages in the datasets, and/or a wrong choice of phylogenetic markers. Criticism over the eocyte hypothesis and its implications was also expressed by Cavalier-Smith (2014) and Forterre 2013). Both argued that the occurrence of homologs of distinctive eukaryotic genes in the Asgards and TACK was not to be taken as an incontrovertible validation of the eocyte hypothesis, because these genes might have been inherited from a common ancestor of Archaea and Eukaryotes and conserved in the Asgards and TACK but lost in the rest of Archaea. Skophammer et al. (2007) compiled several reasons to argue that the archaea are derived from bacilli (Firmicutes), notably the fact that several enzymes involved in the biosynthesis of archaeal membrane lipids also occur in these bacteria. Phylogenetic work by Valas & Bourne (2011) also supported an origin of archaea from Firmicutes. In contrast, Cavalier-Smith and Chao (2020) argued that the archaea and eukaryotes are sister groups diverged from a gram-negative bacterial ancestor (Fig. 1D).

Similarities in the DNA replication machinery of eukaryotes and archaea suggest that the archaeal/eukaryotic common ancestor, independently of its basal or derived position in the phyletic tree (cf. Fig. 1C and D), possessed a DNA replication apparatus that was as complex in its main features as in modern eukaryotes (Lindås & Bernander 2013; Makarova & Koonin 2013). Interestingly, the archaea present a surprising dicothomy in their cell division system. A part of them (roughly the Euryarchaeota and some of the TACK) employ the FtsZ system for cell division, as in bacteria. Others lack FtsZ and use the newly discovered Cdv (cell division) machinery that is homologous with the eukaryotic ESCRT III protein family for cell division (Glossary). The occurrence and scattered distribution of the

FtsZ and Cdv division systems in extant archaea suggests that both machineries were present in the last common ancestor of the group (LACA) and were differentially lost in multiple lineages (Koonin 2015). The archaea display an analogous dichotomy also in the mechanism of sexual recombination. Members of the Chrenarchaeota employ a system of DNA import similar to but not homologous with bacterial conjugation (van Volleren et al. 2016). In contrast, sexual recombination in members of the Euryarchaeota involves cellular fusion and subsequent chromosome segregation (Naor & Gophna 2013; Shalev et al. 2017). No information is currently available on the molecular basis of this mechanism to evaluate possible homologies with eukaryotic sexual reproduction. Because the genes encoding eukaryote-like traits are present in a patchwork pattern across archaeal taxa, they are referred to as the "dispersed archaeal eukaryome" (Koonin & Yutin, 2014). The patchy distribution of essential molecular machineries in extant archaea suggests that LACA was more complex than its known present-day descendants.

3. Cell sizes, gene costs, and cellular energetics

One of the most prominent traits of the eukaryotic cell is phagotrophy, the ability to engulf prey within membrane-bound vesicles and digest it intracellularly. This requires cells larger than or at least as large as the prey. Indeed, eukaryotic cells are on average at least one order of magnitude larger that bacteria and archaea, although there are extremes at either end of the size range in all three lineages. Some bacteria, for example *Thiomargarita* or *Epulopiscium*, are so large as to be visible to the naked eye (Fig. 2), whereas the photosynthetic eukaryotic protist *Ostreococcus* measures only 0.8 µm in diameter.

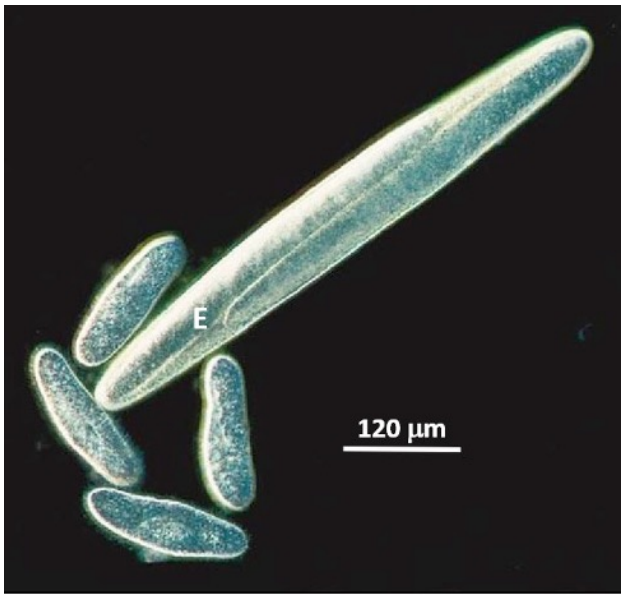

Figure 2: Phase-contrast light micrograph of the bacterium *Epulopiscium fishelsoni* (E). This giant gram-positive bacterium lives as a symbiont in the intestine of surgeonfish. The smaller cells close by are parameciums living in the same habitat. The *Epulopiscium* cell shown in the picture contains several large endospores ready to be liberated by dissolution of the mother cell envelope. *Epulopiscium* may have grown so large to avoid predation from ciliates. Courtesy of Esther Angert, Cornell University, USA.

In terms of metabolic activity, the bacteria produce more energy per unit cell mass than eukaryotes. Nevertheless, the amount of energy available for each gene or for Mb (million base pairs) of DNA is on average much larger in eukaryotes than bacteria. According to Lane & Martin (2010), this depends on the fact that larger cellular sizes in eukaryotes bring about an isometric increase in the mass of mitochondria, hence in the amount of energy produced per unit cell mass. During the evolution of the mitochondrion, most genes necessary for mitochondrial maintenance were transferred to the nucleus, where they are generally present as single copies, although the cytoplasm usually contains multiple mitochondria. This converted the mitochondrion into a highly efficient energy-

producing machine whilst reducing its genome size to a minimum. Lane & Martin (2010) argue that the evolution of the mitochondrion suppressed the selection pressure against genome expansion that strongly affects prokaryote evolution, thus fostering a dramatic increase in the size of eukaryote genomes. According to this reconstruction, the possibility to maintain and express an increasing number of genes was pivotal to the evolution of phagocytosis and other distinctive eukaryotic traits. In this scenario, therefore, the acquisition of the mitochondrion was the very event that triggered eukaryogenesis.

Lynch & Marinov (2015) showed that the cost of a gene in terms of duplication (sDNA), transcription (sRNA) and translation (sPRO) increases by one to two orders of magnitude in the sequence sDNA<sRNA<sPRO, and that the total gene cost declines with cell volume in *both* eukaryotes and prokaryotes. By applying well known notions of population genetics, Lynch & Marinov (2015) showed that sequences with neutral or even weakly disadvantageous phenotypic effects can be detected and eliminated by purifying selection only if their energy cost is higher than genetic drift, defined as 1/Ne for a haploid population (where Ne is the effective population size, see Glossary). Because eukaryotic populations are orders of magnitudes smaller that bacterial populations, they ordinarily experience higher genetic drift; consequently, natural selection severely limits genome expansion in bacterial populations, but not so in eukaryotes (Fig. 3). These conclusions suppress the need to invoke an energetic barrier to the evolution of cellular complexity. According to Lynch and Marinov, eukaryotes are prone to colonization by novel genes just because of larger cellular sizes and smaller effective population sizes. These conclusions conflict with Lane & Martin's (2010) hypothesis that the mitochondrion was an essential prerequisite for genome-size expansion in eukaryotes. A second important implication of Lynch and Marinov's work is that the cost of DNA duplication (sDNA) in eukaryotes is generally too low to be detectable by selection, suggesting that the incorporation of substantial amounts of DNA in the genome of large eukaryotes is neutral from a bioenergetic perspective, provided it is not translated. This may help explain the abundance of non-coding DNA in eukaryotic genomes, a trait that most likely had a profound impact on eukaryote evolution (Section 6).

Large cellular sizes and relatively small population sizes are most likely related traits: in any ecosystem, larger organisms necessarily are less numerous that smaller ones, whether uni- or multicellular. If larger cellular sizes reduce the metabolic burden independently of the type of cellular organization, why did eukaryotes evolve large cells whereas prokaryotes remained small? A possible answer is that eukaryotes did so under selection pressure to facilitate engulfment of other cells by phagocytosis. With increasing cellular sizes, diffusion rate probably became a significant limiting factor. Eukaryotic cells faced the problem by evolving unique mechanisms of intracellular transport based on *actin* and *tubulin*. The extraordinary versatility of the eukaryotic cytoskeleton does not depend on special properties of eukaryotic actin and tubulin but rather on a diversity of accessory proteins, notably nucleation factors and molecular motors (Wickstead & Gull 2011).

Tubulin and actin homologues exist in both bacteria and archaea (Cabeen & Jacobs-Wagner 2010; Spang et al. 2015). Prokaryotes have motors that act on DNA and RNA but lack cytoskeleton-associated motors or nucleators. Prokaryotic cytoskeletal proteins, in fact, polymerize spontaneously. Because the evolution of

cytoskeleton-associated motors or nucleators does not seem a particularly unlikely event, Theriot (2013) suggests that their absence in prokaryotes reflects a deeper divergence in the functional architecture of the prokaryotic and eukaryotic cell. To control the assemblage of cytoskeletal scaffolds in space and time, prokaryotes do not use nucleators bur merely stabilize or de-stabilize spontaneously polymerizing filaments. The involvement of nucleators and motors imparts the eukaryotic cytoskeleton radically different properties. Prokaryotic and eukaryotic cytoskeletal structures both follow the kinetic pattern known as "dynamic instability", characterized by simultaneous assembly and disassembly of filaments. Yet,

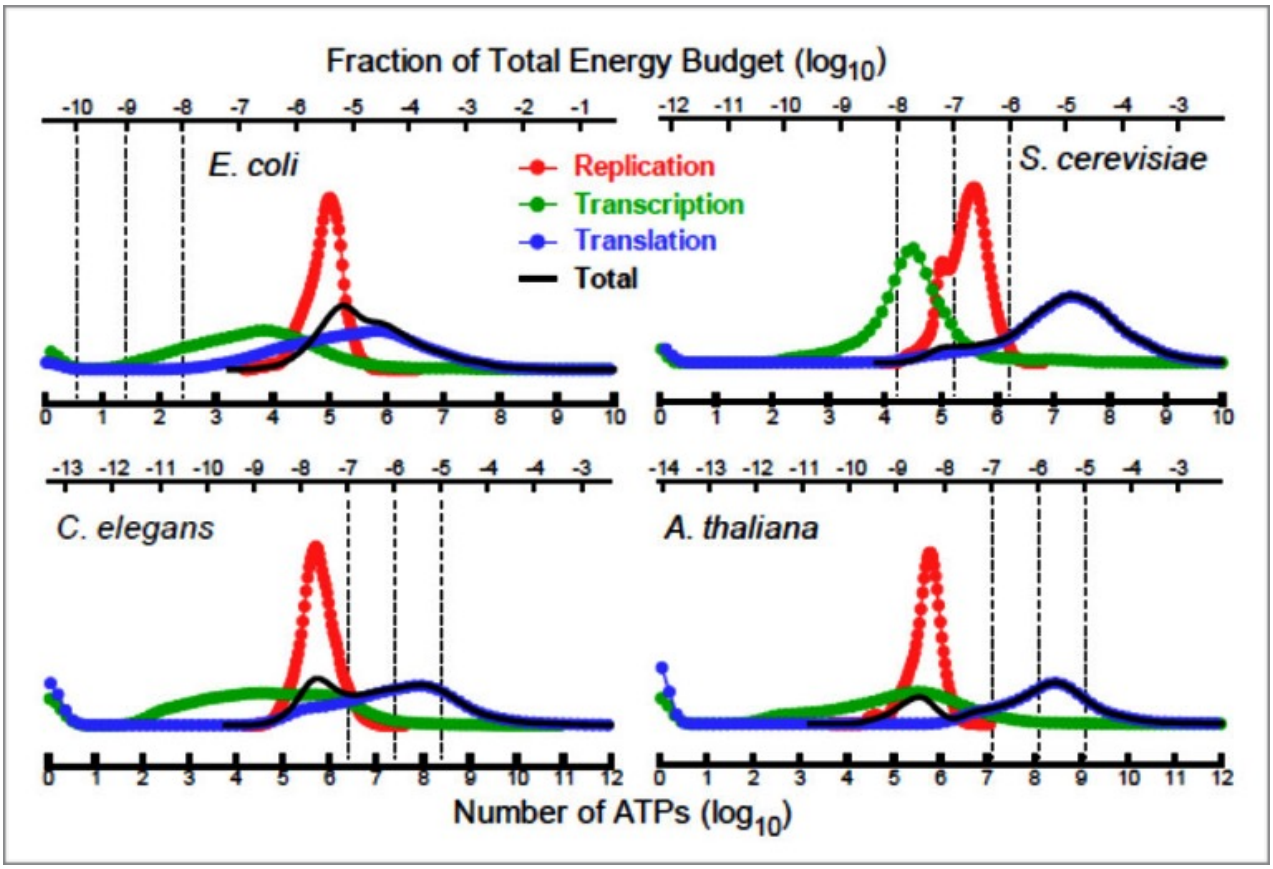

Figure 3: Distribution of energy costs for the full sets of annotated genes in the bacterium *Escherichia coli* and three eukaryotic species (the yeast *Saccharomyces cerevisiae*, the nematode worm *Caenorhabditis elegans*, and the higher plant *Arabidopsis thaliana*). In each diagram, the bottom axis shows the absolute costs in ATP units, the upper axis shows the corresponding costs as the fraction of the cell's lifetime energy budget. The dashed vertical lines denote levels below which the energy cost is expected to be too low to be opposed by selection (in the absence of any additional advantages from the specific gene); for genes to the left of a particular vertical bar (with logarithmic value x on the upper axis), the energetic cost becomes neutral if the effective population size (Ne) is $>10^{-x}$. For example, an energetic cost from 10^{-10} of the total energy budget would not be perceived by natural selection in a population of *Escherichia coli* above 10^{10} cells. The critical x value shifts in the range 10 to 8 for *E. coli*, 8 to 6 for *Saccharomyces*, and 7 to 5 for *Caenorhabditis* and *Arabidopsis*. From Linch & Marinov (2015) under the terms of the Creative Commons CC BY license.

prokaryotic filaments grow and shrink at both ends, whereas eukaryotic filaments are polarized and can add or lose subunits only at the free end (also known as the *plus* end) because the opposite one (*minus* end) is bound to the nucleating complex. An important consequence of this apparently minor difference is that nucleators impart directionality to the eukaryotic cytoskeleton, enabling it to perceive, produce and transmit spatial information. Theriot (2013) suggests that differences between the prokaryotic and eukaryotic cytoskeleton may ultimately depend on the way spatial information is generated and transmitted in the two cell types. Bacterial cells typically have a single chromosome bound to the cell envelope in a spatial arrangement that is faithfully reproduced at each cell division (Toro & Shapiro 2010). As a result, genes maintain a specific cellular location across cell generations; this provides spatial information that the bacterial cell can use whenever necessary, for example during cell growth and cell division, without the participation of cytoskeletal structures. Small sizes permit bacteria to rely on diffusion whereas eukaryotes need compartmentation and active intracellular transport. Large bacterial cells such as *Thiomargarita* or *Epulopiscium* solve diffusion problems by making thousands copies of the chromosome scattered along the cell membrane (Schulz-Vogt et al. 2007). As discussed in Section 6, a second fundamental difference among eukaryotes and prokaryotes probably lies in quantitative regulation of gene expression.

4. The shift from the classic endosymbiotic model to archaeal/bacterial syntrophic consortia

Besides eukaryote phylogenetic position, either sister to or branched from within the archaea, a second controversial issue in modern biology is the succession of events that produced the eukaryotic cell. A multitude of hypotheses have been proposed, none of which has gained unanimous consensus whilst most had to be discharged in the light of novel discoveries. We will briefly examine the traditional endosymbiotic model and some of the most recent alternative models, referring to former reviews for a more thorough historical exploration of the topic (Poole & Gribaldo 2014; Lake 2015; Martin et al. 2015; Archibald 2015a; Dacks et al. 2016; Silar 2016; Speijer 2020a,b).

The classical endosymbiotic model has been the standard textbook account of eukaryogenesis throughout the 1980s and 1990s. It posits that a prokaryotic ancestor evolved phagocytosis and other fundamental eukaryotic traits and subsequently acquired the mitochondrion by endosymbiosis (Fig. 4).

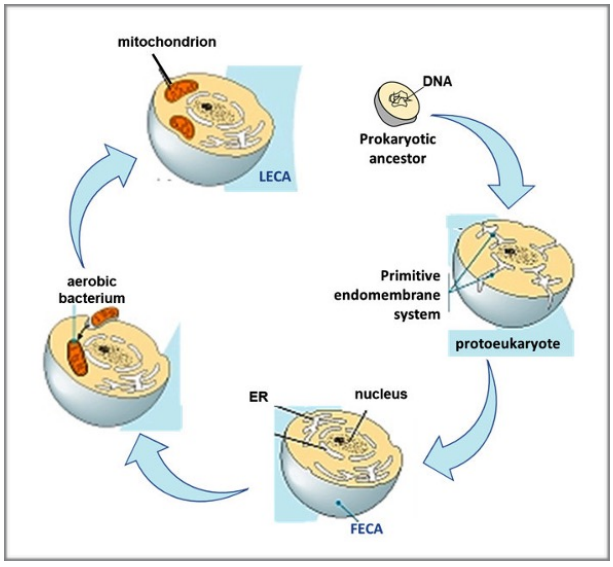

Figure 4: The classical endosymbiotic model posits that eukaryotes lacking mitochondria gradually evolved from a prokaryotic ancestor. FECA is supposed to have already evolved fundamental eukaryotic traits including a phagocytotic machinery. The acquisition of the mitochondrion gave rise to LECA and its modern descendants.

The model received support from the discovery of eukaryotic lineages, such as the Archamoebae and Metamonada, which lacked mitochondria. These eukaryotes were dubbed "archezoa" and supposed to be the surviving descendants of a primitive eukaryotic lineage preceding the mitochondrial symbiosis (Cavalier-Smith 1989). This model went into crisis at the turn of the century with the discovery that the archezoa derived from a mitochondriate ancestor and had artefactually clustered together at the base of the eukaryote tree (Keeling 1998).

The recognition that there is no known eukaryote primarily lacking the mitochondrion is not by itself irreconcilable with the traditional endosymbiotic model. The notion, however, rapidly fostered novel scenarios of eukaryogenesis that describe the eukaryotic cell as the result of a consortium of prokaryotic organisms. The simplest, more parsimonious post-archezoan model that accounts for both the universal presence of mitochondria in eukaryotes and the chimeral composition of the eukaryote genome posits engulfment of an alfaproteobacterium by an archaeal host in a syntrophyc (Glossary) ecological context. Several variants of this model have been proposed since the end of the past century, generally differing from each other only in details. The most recent version, dubbed the *reverse flow model* (Spang et al. 2019) is rooted in phylogenetic analysis pointing at Asgards as the closest archaeal relatives of eukaryotes. It assumes that the eukaryotic cell originated from an endosymbiotic association of an anaerobic Asgard host with a facultative anaerobic alphaproteobacterium. Based on metabolic properties of Asgards inferred from genomic analysis, the reverse flow model assumes that the archaeal host oxidized small organic substrates such as hydrocarbons and fatty acids using a reversed Wood–Ljungdahl pathway and released reducing equivalents in the form of hydrogen or small organic compounds. These in turn were oxidized by a syntrophic alfaproteobacterium that transferred the electrons to oxygen by means of a [NiFe]-hydrogenase. The model owns the name to the fact that it postulates a flow of reducing equivalents from the archaeon to the alfaproteobacterium, in contrast to earlier syntrophic models assuming a flow in the opposite direction. To explain the transition from archaeal to bacterial membrane lipids, the model posits that the necessary gene set was transferred from the alfaproteobacterial endosymbiont to the archaeal host, possibly via the mechanism proposed by Gould et al. (2016) (Section 5). The discovery by Bulzu et al. (2019) that the Heimdallarchaeia (the closest archaeal lineage to eukaryotes to date) probably have a microoxic niche, supports the metabolic landscape of the reverse flow model centered on interaction between a member of the Asgard and an oxygen-dependent alfaproteobacterium.

An alternative model, initially put forward in 1998 and recently revised by its proponents, is the *HS syntrophic model* (Lòpez-Garcìa & Moreira 2020). The proposed ecological scenario was Early Proterozoic microbial mats encompassing an oxygen-rich superficial layer of cyanobacteria, an underlying transitional zone hosting versatile sulfide-oxidizing alfaproteobacteria, and a deeper zone with a low reduction potential inhabited by Asgard archaea and sulfate-reducing deltaproteobacteria (Fig. 5A). In this ecological context, the archaea anaerobically oxidized simple organic compounds leached from the cyanobacterial layer, producing hydrogen that in turn was oxidized by deltaproteobacteria using sulfate as an electron acceptor. The resulting sulfide by-product was aerobically reconverted into sulfate by alfaproteobacteria, thus maintaining a cyclic

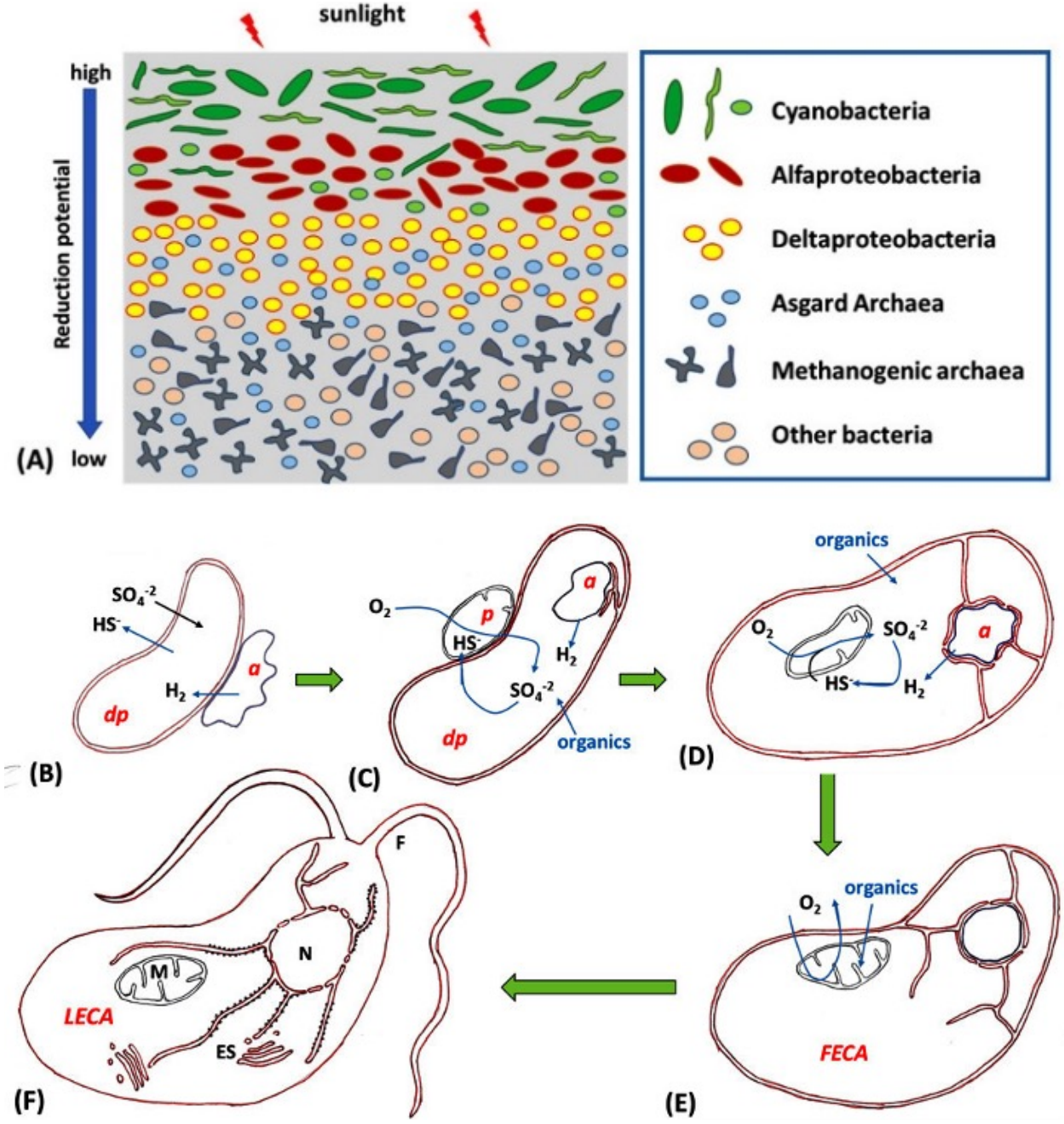

Figure 5: A. The HS syntrophic model by Lòpez-Garcìa & Moreira (2020) posits that eukaryotes originated in a superficial microbial mat hosting a diversity of bacterial and archaeal species distributed along a gradient of reduction potential. **B.** The first step was a syntrophic association between a sulfate-reducing deltaproteobacterium (*dp*) and an Asgard archaeon (*a*). **C.** The deltaproteobacterium engulfed the archaeon and the resulting consortium migrated to an upward level of the microbial mat, where it established a second syntrophic interaction with aerobic proteobacteria (*p*) that re-oxidized sulfide ions to sulfate. **D.** A proteobacterium (p) was engulfed and retained as a second endosymbiont within the deltaproteobacterial host. In parallel, the archaeal endosymbiont was enclosed within a membranous sheet deriving from the inner membrane of the bacterial host; this might have facilitated the transport of organic substrates from the outside to the archaeon. **E**. With the stabilization and genomic integration of the two endosymbionts, the tripartite consortium evolved into the first eukaryote common ancestor (FECA), which featured primitive versions of the nucleus, mitochondrion, and endomembrane system. **F.** The transition from FECA to the last common ancestor of extant eukaryotes (LECA) involved full integration of the symbionts and the completion/addition of traits shared by extant eukaryotes, notably a nucleus (N), an endomembrane system (ES), the mitochondrion (M), a specialized cytoskeleton, and a couple of 9+2 microtubular flagella (F). The model postulates that the inner membrane of the host lost topological continuity with the endomembrane system and disappeared, its role being transferred to the outer membrane. The prokaryote-to-eukaryote transition involved a steady increase in cellular and genome sizes.

flow of matter largely supported by cyanobacterial photosynthetic activity. The HS syntrophic model suggests that this syntrophic association evolved into a tripartite symbiotic consortium when, in two successive steps, an archaeon and an alphaproteobacterium became endosymbionts within the deltaproteobacterium, the first giving rise to the nucleus and losing the archaeal membrane, the latter becoming the mitochondrion (Fig. 5B-F). Assuming that the host was a bacterium, not an archaeon, the HS syntrophic model does not need to postulate a membrane transition. To solve ecological incompatibility between obligate anaerobic Asgard archaea and aerobic proteobacteria, the HS syntrophic model suggests that the deltaproteobacterium first acquired the archaeal endosymbiont in the anoxic layer, then the resulting consortium migrated upwards along the reducing potential gradient and established the second endosymbiosis. The proponents view the first eukaryotic common ancestor (FECA) neither as an archaeon nor as a bacterium, but as the first tripartite symbiotic consortium. Fig 5F depicts LECA as a protist with two dissimilar flagella, instead of a single flagellum as assumed by Lòpez-Garcìa & Moreira (2020), this probably being the basal condition in extant eukaryotes (Cavalier-Smith 2014; Derelle et al. 2015).

A weak point of the HS syntrophy model is in the assumption that, after giving rise to the endomembrane system including the nuclear envelope, the inner membrane of the deltaproteobacterial host disappeared, its role as the cell membrane being taken over by the outer membrane. This appears to be topologically and operatively unlikely, because the outer membrane of gram-negative bacteria is highly specialized and profoundly different from the inner membrane in composition and functions. Whereas the central role of the inner membrane is in the control of molecular exchanges with the environment including the chemiosmotic mechanism, the outer membrane essentially works as a barrier to prevens harmful molecules including antibiotics from entering the cell (Nikaido 2003; May & Grabowicz 2018). In addition, the outer membrane has a major role in cellular stiffening (Rojas et al. 2018). Because of this, the outer membrane could hardly take the role of the cell membrane. In addition, the transition proposed by Lòpez-Garcìa and Moreira would require the concomitant loss of the peptidoglycan layer (localized in the periplasmic space between the two membranes in gram-negative bacteria), because its retention would most likely be incompatible with the outer membrane functioning as the cell membrane. Thus, the loss of the outer membrane appears to be a more likely alternative, with a thin peptidoglycan envelope persisting to provide mechanical support until the symbiotic consortium developed a replacement. The only major change required in this scenario would be the establishment of topological discontinuity between the inner membrane and ER, with the co-translational protein insertion pathway and lipid-synthesizing complexes becoming restricted to ER as in modern eukaryotes. Independently of the model adopted, this was a necessary step in eukaryogenesis, probably associated with the evolution of the Sec61 protein translocon complex from the SecYEC prokaryotic counterpart (Cavalier-Smith 2014). Concurrently, the F-ATP synthase (Junge & Nelson 2015) originally present in the cell membrane was lost and its function taken over by the mitochondrial homologue.

The reverse flow model considers the evolution of the mitochondrial endosymbiont as the very event that triggered eukaryogenesis, whereas the HS syntrophic model posits that the acquisition

of the mitochondrial endosymbiont post-dated the primary endosymbiosis that gave rise to the nucleus and triggered the development of a phagocytotic machinery. The excess of bacterial versus archaeal genes in the eukaryote pangenome (Bruekner & Martin 2020) appears to favour the HS syntrophic model, although gene acquisition from putative bacterial endosymbionts other than the mitochondrial one (Roger et al. 2017) might also account for the slight overabundance of bacterial genes.

The apparent absence of phagocytosis in bacteria and archaea has long been considered a major problem with models assuming endosymbiotic association of prokaryotic cells. Some models (see, for example, Martijn & Ettema 2013) have accepted a symbiogenetic eukaryote origin from an archaeal host only under the premise that this had already evolved an endomembrane system, a eukaryotic cytoskeleton and phagocytosis prior to the engulfment of the mitochondrial ancestor. This reiterates the narrative of the classic endosymbiotic model with the only difference that an archaeal ancestor takes the place of the generic "prokaryote" ancestor in Fig. 4. The recent report of a planctomycete bacterium, 'Candidatus Uab amorphum', which is able to engulf and digest intracellularly other bacteria and small eukaryotic cells through a phagocytosis-like, mechanism proves that bacterial phagocytosis does exist, albeit based on different molecular grounds than eukaryotic phagocytosis (Shiratori et al. 2019, Fig. 6).

Universal eukaryotic protein families of alphaproteobacterial ancestry and of mitochondrial localization retain greater similarity to their homologues in free-living prokaryotic relatives compared to other eukaryotic proteins with different prokaryotic origin (Pittis & Gabaldón 2016). This is interpreted as evidence that mitochondrial endosymbiosis was a late step in eukaryogenesis, thus conflicting with "mitochondrion-first" models but in line with the HS syntrophic model. Crucially, alphaproteobacterial genes account for a minor fraction of eukaryotic genes of bacterial ancestry. Moreover, whereas eukaryotic genes of alpha-proteobacterial origin mostly relate to mitochondrial functions, bacterial genes of non-alphaproteobacterial ancestry are involved in other essential eukaryotic traits such as the endomembrane system, suggesting that they entered the (proto)eukaryotic genome prior to mitochondrial symbiosis (Pittis & Gabaldón 2016; Gabaldón 2018).

5. The endosymbiotic model revisited

Profound affinities in all aspects of cellular physiology and molecular machinery of archaea and eukaryotes (Table 2) clearly reflect a shared evolutionary history. Cavalier-Smith (2002, 2006, 2014) has proposed a variant of the classic endosymbiotic model based on the assumption that the eukaryotes and archaea are sister groups derived from a bacterial ancestor. Cavalier-Smith's latest favourite candidate for the role of neomuran ancestor is a member of the Planctobacteria (Cavalier-Smith & Chao 2020), a choice in line with the recent discovery of a phagocytosis-like machinery in a member of this group (Shiratori et al. (2019). The event that gave origin to the putative novel lineage was the replacement of the outer membrane and the thin murein (peptidoglycan) layer in the gram-negative planctomycete ancestor with a more flexible envelope of N-glycoproteins linked to the cell membrane (Fig. 7). To emphasize the evolutionary relevance of this change, Cavalier-Smith (2002) named the

eukaryote-archaea clade *Neomura*, from Greek *neos* (new) and Latin *murus* (wall). The selection pressure underpinning the transition from a peptidoglycan to a N-glycoprotein cellular exoskeleton might have been adaptation to mildly acidic, moderately hot conditions that weakened murein integrity, or competition with microorganisms producing inhibitors of murein synthesis.

The neomuran model asserts that the bacteria are substantially older than and ancestral to the neomura, and that eukaryotes and archaea vertically inherited shared bacterial traits from their bacterial ancestor. The scenario posits that the neomura replaced active DNA supercoiling by DNA gyrase with passive DNA supercoiling by histones before the divergence of eukaryotes and archaea. The novel system of DNA coiling allegedly forced

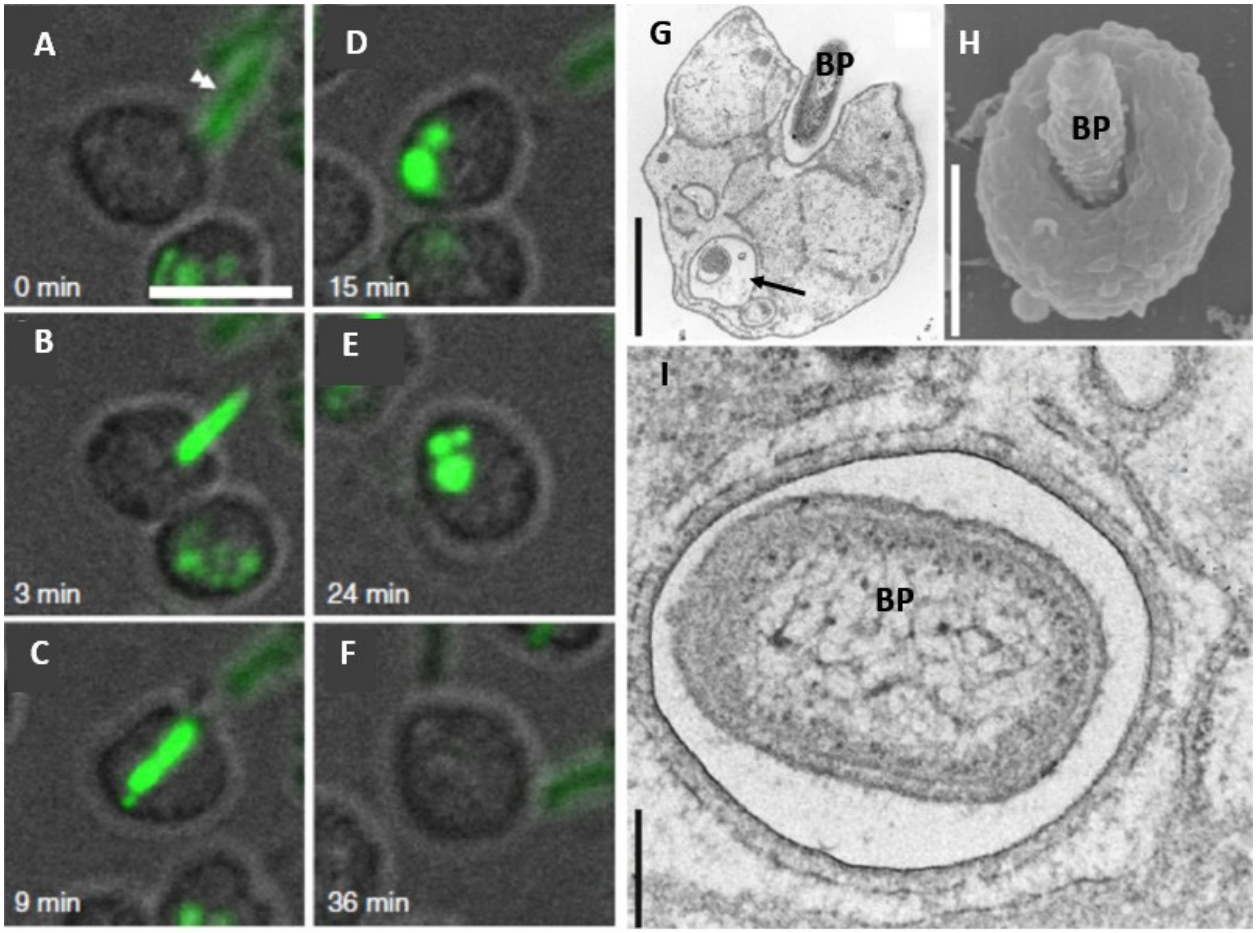

Figure 6: Phagocytosis requires relatively large cellular sizes to accommodate the prey. The Planctomycete Candidatus Uab amorphum has larger cells than typical bacteria and is able to engulf bacterial prey such as *Escherichia coli*. **A-F**: selected images of time-lapse video showing engulfment and digestion of *Escherichia coli* cells labelled with a green fluorochrome. Note the disappearance of the label in F, due to intracellular digestion and lysis of the engulfed cell. **G.** Transmission electron micrograph of 'Candidatus Uab amorphum' showing engulfment of a bacterial prey (BP) and remnants of a digested bacterium (arrow). **H.** Scanning electron micrograph showing the same. I. Transmission electron micrograph of a bacterial prey (BP) sequestered in an intracellular vacuole after engulfment. Scale bars: 5 µm (A-F), 500 nm (G, H), 200 nm (I). From Shiratori et al. (2019) under the terms of the Creative Commons CC BY license.

Table 2: Major traits shared by archaea and eukaryotes.

1. Integral N-glycoproteins exposed on the outer side of the cell membrane.
2. Murein (also known as peptidoglycan, a fundamental component of the bacterial cell wall) is lacking. An analogue of murein (pseudomurein) not containing muramic acid evolved secondarily in the Methanobacteriales, an archaeal lineage belonging to the Euryarchaeota.
3. Proteins are inserted in, or translocated across membranes only co-translationally, with the participation of a signal-recognition complex (SRP) containing a 7S RNA and a translation-arrest domain that delays the extension of the polypeptide chain until the ribosome/nascent protein complex binds to a SRP receptor anchored in the target membrane. The bacterial SRP complex lacks both the 7S RNA and the delay mechanism.
4. Histones: H1, H2a, H2b, H3 and H4 in the eukaryotes, homologs of H3 and H4 reported in the Euryarchaeota.
5. DNA polymerase of the B type *i.e.* inhibited by aphidicoline.
6. TATA boxes (repeated sequences of adenine-thymine) initiate transcription (absent in genes transcribed by RNA polymerases I and III in eukaryotes). Sigma factors absent (essential in bacteria to initiate transcription).
7. Several unique DNA-repair enzymes.
8. Similarities in ribosomal RNA and proteins; ribosomes insensitive to chloramphenicol; peptidyl transferase sensitive to anisomycin.
9. CCA 3′ terminus of tRNA added post-translationally, not gene-encoded.
10. Protein synthesis initiates with methionine, not N-formyl methionine.
11. Multiple origins for chromosome replication.

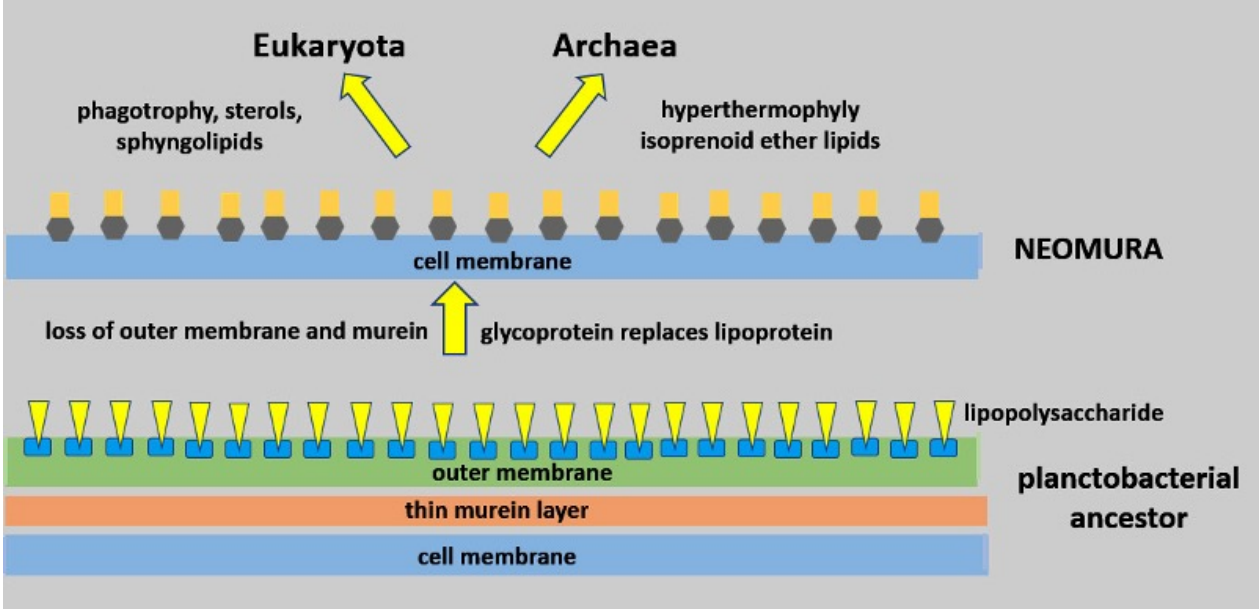

Figure 7: The neomuran model posits that the eukaryotes and archaea are sister groups and collectively form the Neomura clade. According to the latest version (Cavalier-Smith & Chao 2020), the common ancestor of Neomura was a planctobacterium that replaced the outer membrane and the murein layer with a flexible envelope of N-glycoprotein bound to the cell membrane.

drastic changes in the ancestral machinery of DNA replication, repair, and transcription, which were inherited en-bloc by eukaryotes and archaea from the common neomuran ancestor (Table 2). Uniquely eukaryotic and archaeal characters evolved after the two lines separated from their last common ancestor, with the eukaryotes emerging as phagotrophic predators, and the archaea ancestrally adapting to hyperthermal acidic environments. By following divergent evolutionary pathways, eukaryotes and archaea developed the set of unique traits that distinguish the two lineages. As an example, the neomuran model posits that the putatively ancestral bacterial flagellum was lost during the neomuran transition, and novel locomotory organelles (the eukaryotic flagellum or undulipodium, and the archaeal flagellum or archaellum) evolved independently in each lineage. It has been proposed that the unifying trait of archaea is adaptation to chronic energy stress (Valentine 2007). This enables the archaea to outcompete the bacteria in habitats that are consistently extreme, whereas the bacteria do better in habitats with fluctuating conditions.

Cavalier-Smith agrees that the acquisition of the mitochondrion greatly enhanced the amount of energy accessible to eukaryotes, yet he maintains that the decisive trigger for eukaryogenesis was the evolution of an endomembrane system and phagocytosis, which in his view predated the mitochondrion, nucleus and mitosis.

The neomuran model is at odd with phylogenomic evidence of eukaryotes nested within the archaea. Moreover, Cavalier-Smith's assumption that the archaea are sister to eukaryotes and therefore a relatively recent lineage contrasts with isotopic evidence of methanogenic archaea and methane-metabolizing proteobacteria in Early Archaean sediments over 3,46-billion-year-old (Ueno et al 2006; Schopf et al. 2018). However, whereas mitochondrion-first models focus almost exclusively on phylogeny, paying little attention to cellular issues, the neomuran model integrates a vast body of molecular and cellular data into a well-structured and finely detailed narrative that seemingly accounts for all known facts except mainstream molecular phylogeny.

Refusing hypotheses of archeal-bacterial "fusion", Forterre (2013) proposed an evolutionary model almost completely convergent with the scenario by Cavalier-Smith (2002, 2014). According to Forterre's model, the archaea and eukaryotes originated from a common ancestor, the former following a reductive evolutionary pattern, the latter proceeding towards increasing complexity and eventually acquiring the mitochondrion by endosymbiosis. The most significant difference between the two models was in the hypothesis by Forterre that the archaea and eukaryotes did not arise from a bacterial lineage as proposed by Cavalier-Smith, but directly diverged from LUCA as the sister clade to the bacteria. In addition, Forterre posits that viral vectors played a major role in the emergence of all three kingdoms.

Valas & Gourne (2011) presented an evolutionary analysis largely in line with the neomuran model, with antibiotic warfare underpinning the evolution of Neomura from Firmicutes (endospore-forming gram-positive bacteria). This study, however, was agnostic about whether the archaeal lineage is monophyletic (hence sister to eukaryotes) or paraphyletic.

Whether the host cell that started the mitochondrial symbiosis was a complex archaeon, a primitive eukaryote or a bacterium, however, is not merely a matter of semantics. The membrane-bound compartments that are defining features of all eukaryotic cells are dynamic entities, constantly exchanging material with one another via vesicles while maintaining their

unique identities (Dacks & Field 2018). Supporters of the eocyte scenario suggest that the putative archaeal host acquired the ability to make acyl-ester lipids by gene transfer from the alfa-proteobacterial symbiont during its conversion into a mitochondrion, in response to the necessity to harmonize membrane biochemistry in the two partners (Dey et al. 2016). Based on the observation that gram-negative bacteria and mitochondria are able to release vesicles from their outer membrane, it has been suggested that a similar process initiated the evolution of the endomembrane system during eukaryogenesis (Gould et al. 2016). A transition from archaeal to bacterial membrane chirality is theoretically possible because, in contrast with former belief, hybrid membranes made with archaeal and bacterial lipids are stable and functional (Shimada & Yamagishi 2011). Moreover, an engineered bacterium expressing the archaeal lipid biosynthetic pathway and producing hybrid membranes was perfectly viable (Caforio et al. 2018). The membrane transition hypothesis has received support from the discovery that the Lokiarchaeota and several uncultured Euryarchaeota lack the gene to synthesize G1P and, consequently, the capacity to make archaeal membrane lipids. Yet, these archaea possess the genetic potential for the synthesis of chimeric membrane lipids, namely di- or tetraether-linked isoprenoid lipids with G3P stereochemistry, or lipids with one ether-linked isoprenoid chain at position sn-1 of a G3P backbone and one ester-bound fatty acid at position sn-2 (Villanueva et al. 2016).

6. Why a nucleus?

The nucleus is the trait that gives eukaryotes their name. Far from being just a DNA bag, the nucleus is a highly specialized organelle whose properties are central to the functioning of the eukaryotic cell. The nuclear envelope consists of two membranes with distinct compositions, the outer one being more like ER membranes. This configuration is strong evidence that the nuclear envelope arose as a subdomain of the same endomembrane compartment that gave rise to the ER, but with the function to enclose the genetic material (Dacks et al. 2016). The nuclear envelope bears specialized pores made of over thirty different kinds of proteins (nucleoporins), which control the traffic of molecules and keep gene duplication and transcription (intranuclear) separate from translation (cytoplasmic). This permits mRNA to be processed and eventually released into the cytoplasm in the form ready for translation. During the interphase, the chromosomes are bound to the nuclear envelope, each occupying a discrete territory (Speicher & Carter 2005).

Nucleoporins are structurally related to a family of proteins termed protocoatomers. The basic structure of the nuclear pores and about twenty nucleoporins are conserved across all eukaryotic taxa, and were probably inherited from LECA. In contrast, other components display surprising diversity between lineages, suggesting that they evolved after the divergence from LECA (Makarov et al. 2021). Because protocoatomers are components of the nuclear pores as well as of the intraflagellar transport machinery and protein coat complexes involved in membrane budding, the origins of the nucleus, flagella and organelles of the endomembrane system are probably inter-linked (Dacks et al. 2016).

Why and when the eukaryotic cell evolved a nucleus is a matter of speculation. By analogy with the mitochondrion and chloroplast, the double-membrane structure of the nuclear envelope has prompted hypotheses of a symbiotic origin (Poole & Penny 2006; Martin et al. 2015), now dismissed in favour of autogenous scenarios

(Jekely 2008; Cavalier-Smith 2010a). A novel version of the symbiotic hypothesis was proposed by Lòpez-Garcìa & Moreira (2020) in their HS syntrophic model (Section 4), which posits that the nuclear material (chromatin) derived from an archaeal endosymbiont, whereas the nuclear envelope derived from the endomembrane system of a proto-eukaryote host (Fig. 5).

Affinity between proteins of the nuclear pore scaffold with type I and type II coatomer families involved in vesicle traffic between the ER and Golgi suggests that the nucleus evolved in cells that already possessed at least a primitive form of the endomembrane system (Field and Rout 2019). In bacteria, the cell envelope has a central role in chromosome spatial organization, replication and segregation, and a cell wall-bound FtsZ complex controls cell division (Makarova et al. 2010; Toro & Shapiro 2010; Stouf et al. 2013). Cavalier-Smith (2014) suggested that the nuclear envelope evolved early in eukaryogenesis to replace the cell envelope as a support for DNA when the ancestral eukaryotes specialized as phagotrophs and lost the original cell envelope. In this scenario, the ESCRT complex ("*Endosomal Sorting Complexes Required for Transport*") and a novel cytoskeletal system (the mitotic spindle) replaced the ancestral prokaryotic cytokinetic apparatus. Chromosome segregation in eukaryotes involves the interaction of spindle microtubules with kinetochores (KT), large multiprotein complexes assembled at specialized chromatin sites. There is no evidence for a common descent of known bacterial chromosome segregation systems and the eukaryotic KT, the latter having probably evolved by duplication and neofunctionalization of proteins or protein domains involved in other processes such as ubiquitination, transcription, and flagellar and vesicular transport (Tromer et al. 2019). Garg et al. (2016) argue that the microtubule-dependent eukaryotic mechanism of chromosome segregation was too expensive in terms of energy to be affordable before the evolution of the mitochondrion, which they champion as the initial event of eukaryogenesis.

If the necessity to free the cell envelope from the chromosome burden may account for the internalization of the genetic system and the evolution of a dedicated cytoskeletal system, it remains to explain why the genetic system was enclosed within the nuclear envelope. One of the hypotheses put forward is that the nuclear envelope was necessary to prevent harmful ribosome chimerism after the transfer of ribosomal protein genes from the evolving mitochondrion to the host genome (Jekely 2008). Without a nuclear envelope, host rRNA might bind to bacterial ribosomal proteins (many of which have sequence homologies with eukaryotic counterparts), producing faulty ribosomes. The nuclear envelope permits the host rRNA to be retained in the nucleus, where it correctly associates with host ribosomal proteins (synthesized in the cytosol and translocated to the nucleus across nuclear pores), whereas mitochondrial ribosomal proteins are synthesized in the cytosol and translocated to the mitochondrion by specialized translocons.

A second hypothesis independently put forward by two research groups (Martin and Koonin 2006; Lòpez-Garcìa & Moreira 2006) also associates the origin of the nucleus with the mitochondrial symbiosis but proposes that the nuclear envelope was a response to the invasion of the host genome by group II introns, self-replicating sequences from the mitochondrial symbiont. These sequences replicated extensively and inserted randomly within host genes, producing *spliceosomal introns*, non-coding sequences that presently account for a significant fraction of total genome in eukaryotes (Gregory 2005; Elliot

& Gregory 2015). In the absence of a repair mechanism, the insertion of foreign sequences within host genes would seriously disrupt genetic information. The evolution of the nuclear envelope introduced a spatial and temporal separation between transcription and translation, thus permitting the removal of introns from transcripts before these could bind to ribosomes. Known as *RNA splicing*, this important operation is performed by the *spliceosomes*, complexes of small nuclear ribonucleoproteins (Irimia & Roy 2014). In prokaryotes there are no spliceosomes, and RNA splicing is rare and mostly affects non-coding RNAs (Cavalier Smith 2014).

An important difference between eukaryotes and prokaryotes is in ribosomal sizes, with a mass of over 4 million Da in the former and about 2.5 million in the latter. The extra mass is due to rRNA expansion and numerous additional ribosomal proteins (Melkinov et al. 2012). Despite this, eukaryotic ribosomes display no evident advantage over their prokaryotic counterparts in terms of translation accuracy. Because of the abundance of ribosomes in eukaryotic cells, their extra mass accounts for several percent of total RNA and protein contents, implying a significant metabolic burden. The significance of the larger sizes of eukaryotic ribosomes is unclear. It has been suggested that the extra mass is necessary to control translocation, as separate subunits, across the nuclear envelope. The fact, however, that ribosomes in complex multicellular eukaryotes (e.g., the vertebrates) are significantly larger than in simpler eukaryotes such as the unicellular yeast *Saccharomyces* suggests that the additional components participate in other functions besides ribosomal biogenesis, for example the docking, trafficking, and chaperoning of neo-synthesized proteins (Bernier et al. 2018).

The mechanism of cell division in prokaryotes generally requires the genome to be vehiculated in a single chromosome with a single replication origin (Toro & Shapiro 2010; Egan & Volmer 2013). In contrast, mitosis is perfectly compatible with genomes split into several chromosomes, each with more than a single replication origin (McIntosh 2016). Mitosis, therefore, permitted early eukaryotes to expand their genomes without impairing the precision of chromosome segregation or the rate of replication. The main mechanism of genome expansion in eukaryotes is genome duplication followed by loss of function of redundant sequences (Van de peer et al. 2009), with a minor contribution from lateral gene transfer (Ku et al. 2015b). The nucleus-to-cytoplasm volume ratio, or *caryoplasmic ratio*, tends to approach a value of 0.1 in metabolically active cells including multinucleate cells. In other words, cells with large cytoplasmic volumes (setting aside vacuoles and other storage compartments) have large nuclei and vice versa, independently of the cellular type or taxonomy. Moreover, cellular and nuclear sizes are linearly correlated with genome size but not with gene number. Although being orders of magnitude larger that prokaryotic genomes, eukaryotic genomes mostly consist of non-coding DNA (Gregory 2005; Elliot & Gregory 2015). The *skeletal DNA hypothesis* by Cavalier-Smith (2005) proposes that non-coding DNA is an essential component of the nuclear matrix, a molecular scaffold needed for the spatial arrangement of chromosomes in the interphase nucleus. Cavalier-Smith suggests that the nuclear matrix also provides the chemical environment needed for gene transcription and transcript maturation. In this perspective, the more abundant is non-coding DNA, the more active is gene expression, thus explaining the so-called C-paradox (Glossary; Gregory 2005; Elliot &

Gregory 2015). According to the skeletal DNA hypothesis, by incorporating massive amounts of non-coding DNA in their genome, eukaryotes could amplify gene expression and make large cells without proportionally increasing the number of gene copies. In contrast, bacteria can attain large sizes only by making multiple copies of their genome (Section 3). Because of relatively low effective population sizes and high genetic drift, the cost of DNA duplication in eukaryotes is so low as to permit incorporation of large chunks of non-coding DNA without incurring in negative selection (Lynch & Marinov 2015).

Character state reconstruction suggests that the multinucleate condition is ancestral in eukaryotes, probably occurring in their last common ancestor (Skejo et al. 2021; also see Section 7). In the scenario proposed, the multinucleate condition was essential in eukaryogenesis because it compensated for errors in chromosome segregation from rudimentary mitosis and favoured genetic integration of the nuclear and mitochondrial genomes.

7. What do we currently know about LECA?

LECA is the ancestral state that gave rise to all extant eukaryotes. LECA is distinct from FECA, defined as the oldest ancestor of eukaryotes that is not also an ancestor of an extant archaeal lineage. FECA is thus the ancestor of all eukaryotes that ever existed, whether extant or extinct, whereas LECA is the ancestor only of extant known eukaryotes plus extinct post-LECA lineages (O'Malley et al. 2019). These definitions are agnostic about whether the mitochondrion first appeared in LECA or evolved at an earlier stage between LECA and FECA.

Because plastid phylogeny does not trace back to LECA (Sections 7 and 8), LECA could not have been a photosynthetic, autotrophic eukaryote, it was a heterotroph. It is widely held that LECA was a phagotrophic predator of other unicellular organisms, mainly bacteria (Cavalier-Smith 2006, 2009; Leander 2020). Mills (2020) challenged this view, suggesting that LECA only possessed a rudimentary phagocytotic machinery (supposed to have been vertically inherited from a putative archaeal ancestor), and that phagocytosis independently evolved several times in crown (derived) eukaryotic lineages. In sharp contrast with this inference, phylogenetic analysis of over 2,000 Arf GTPases (a subfamily of small GTPases with key roles in secretory, endocytic and membrane-recycling pathways in eukaryotic cells) has produced evidence of the presence of a large complement of Arf genes in LECA and of differential simplification of the endomembrane system in eukaryote evolution (Vargova et al. 2021). Indeed, LECA compartmental complexity probably exceeded many extant eukaryotes, for example a unicellular yeast such as *Saccharomyces cerevisiae*.

O'Malley et al. (2019) describe LECA as a population rather than an organism and argue that considerable genomic diversity might have developed before extant lineages diverged. Conceptualizing LECA as a population with a large pangenome implies that major extant eukaryotic lineages, although all ultimately originated from LECA, might derive from partially differentiated subpopulations. In addition, the original LECA population might have hosted a diversity of prokaryotic symbionts, much as in presently living amoebas, ciliates or metamonads. O'Malley et al. (2019) suggest that putative symbionts were independently lost in derived lineages but left different sets of genes in the host genomes. Ancestral genetic heterogeneity possibly favoured early eukaryote diversification and might explain difficulty at identifying the root of the eukaryote tree (Section 9). It is to be noted,

however, that the notion of LECA as a population with a large pangenome distributed across numerous strains is at odd with evidence suggesting that LECA had meiotic sex, a condition incompatible with a large pangenome (Ligrone 2021).

Genome sequencing and notation of *Naegleria gruberi*, a member of the Discoba (Section 9), has revealed that about 4100 genes over a total of 15,727 also occur in at least one of the other major eukaryotic lineages. This is evidence that (at least) 4100 genes present in the eukaryote pangenome were directly inherited from LECA (Fritz-Laylin et al. 2010; Koumandou et al. 2013). Thiegardt et al. (2012) identified 571 genes that were present in LECA and have sequence homology with either bacterial or archaeal genes, the former being mostly involved in metabolic functions, the latter in informational functions. More recent work comparing 209 eukaryotic and 3457 prokaryotic proteomes (Vosseberg et al. 2020) has produced a wealth of further information:

-The LECA genome contained around 12,753 genes, therefore approaching the genome size of a typical extant eukaryote.

-The evolution of pre-LECA genomes was dominated by massive duplication of a relatively small set of genes, which heavily expanded certain protein families whilst leaving others unchanged. Multiple duplications produced the large protein families that control the cytoskeleton, the endomembrane system and nuclear functions, whereas little or no duplication was inferred for protein families involved in metabolic pathways.

-Among protein families with prokaryotic homologues, those with archaeal homologues showed the higher duplication levels, whereas those with alfaproteobacterial homologues the lowest.

-Once again, the Asgards were found to be the closest archaeal relatives to eukaryotes, with the Heimdallarchaeota showing the greater affinity within the Asgards.

-An estimate of the relative time of emergence of protein families suggests that the increase in cellular complexity before the mitochondrial acquisition was mainly related with the evolution of cytoskeletal, intracellular trafficking and nucleolar components.

-Protein families of archaeal affinity are more ancient than those with bacterial affinity, and among the archaea, Asgard proteins are the most recent acquisitions, as are proteobacterial proteins relative to the rest of bacterial proteins.

The work by Vossemberg et al. (2020) suggests that gene families shared with Asgard archaea and involved in the genesis of the endomembrane system and the cytoskeleton duplicated early during eukaryote evolution, thus supporting a mitochondrion-intermediate scenario of eukaryogenesis. In contrast, almost concomitant work by Tria et al. (2021) shows that, among gene duplications traceable to LECA (because occurring in at least two major eukaryotic lineages), those involving genes of bacterial ancestry are much more numerous than duplication of genes of archaeal origin or eukaryotic-specific genes. This is consistent with an early acquisition of the mitochondrion (i.e., preceding the emergence of LECA) and suggests that the host cell was not more complex than a typical prokaryote. Martin et al. (2017) strongly champion the notion that phagocytosis could only have evolved after the mitochondrion. Arguing that phagocytosis is incompatible with the persistence of a chemiosmotic machinery in the cell membrane, they estimate that a phagotrophic organism lacking chemiosmosis should ingest about 34 times its body weight in prokaryotic prey to obtain enough ATP to support one cell division. In the scenario inferred by Martin et al. (2017),

the acquisition of the mitochondrion by a mechanism different from canonical phagocytosis paved the way to the evolution of phagocytosis and other energy-costly eukaryotic traits. In line with this hypothesis, most amitochondrial protists (e.g., the Microsporidia, Metamonada, and the stramenopile *Blastocystis*) are obligate parasites or symbionts of other organisms and live as osmotrophs. Several protists are known, however, which have retained a phagotrophic lifestyle despite having completely lost the mitochondrion, for example *Monocercomonoides exilis*, a bacterivorous oxymonad living as a putative commensal in the intestine of caviomorph rodents (Hampl et al. 2019). Many anaerobic eukaryotes, on the other hand, possess mitochondrion-derived hydrogenosomes that retain a chemiosmotic machinery, whereas others have anaerobic mitochondria that use compounds other than oxygen as the final electron acceptors (Müller et al. 2012).

Although several instances are known of eukaryotes with a single flagellum, the uniflagellate condition (unikonty) is probably derived, two dissimilar flagella being more likely the ancestral condition inherited from LECA (Cavalier-Smith 2014; Derelle et al. 2015).

The widespread occurrence of meiotic genes (i.e., genes involved in meiosis) in extant eukaryotes, including lineages in which sexual reproduction has never been observed, is considered evidence that LECA had meiotic sex (Hofstatter & Lahr 2019). Ancestral character state reconstruction for representatives of a wide set of eukaryotic taxa suggests that LECA, besides being mitochondriate and meiotic, was multinucleate with a predominance of the haploid condition (Skejo et al. 2021). This study suggests that a multinucleate condition was an essential pre-requisite for proto-eukaryotes to survive the transition from a prokaryotic mechanism of chromosome segregation to mitosis (Section 6).

8. The eukaryote tree of life

The five-kingdom classification by Robert Whittaker (1969) has been the reference system for the second half of the past century. It recognized four eukaryotic kingdoms, three of which were clearly circumscribed (Plants, Animals and Fungi), and the fourth (Protists) encompassed all eukaryotic organisms that could not be assigned to any of the other kingdoms. Whereas Plants (restricted to land plants, or Embryophyta), Animals and Fungi were correctly treated as natural (monophyletic) groups, the scientific community was aware from the very beginning that Protists were not an evolutionarily cohesive entity. Electron microscopy rapidly made it clear that classic protist morphological categories such as flagellates and testate or naked amoebae are phylogenetically linked with multicellular forms distributed across the eukaryotic tree of life. Margulis et al. (1990) reserved the term protists for microscopic organisms and proposed the more inclusive kingdom Protoctista for large multicellular eukaryotes that could not be assigned to Plants, Animals or Fungi. The term protists is still currently used in Whittaker's sense, therefore covering a huge diversity of uni- and multicellular forms ranging from heterotrophic protozoa to large brown algae (Archibald et al. 2017), whereas the term Protoctista has been almost completely abandoned.

Starting from the 1990s, molecular phylogeny has brought about tremendous advances in our understanding of the diversity and phylogeny of eukaryotes, revealing numerous novel protist lineages and novel diversity in major known lineages. Inferring phylogenetic inter-relationships is by itself a worthwhile objective, but it is now

clear that an accurate and comprehensive tree of life is a fundamental tool for organizing biological information, putting forward hypotheses and planning research.

From the presence or absence of a dihydrofolate reductase-thymidylate synthase (DHFR-TS) gene fusion and specific myosin gene families, Stechmann & Cavalier-Smith (2003) divided the eukaryotes into two major lineages called "Unikonts" and "Bikonts". These denominations reflected the belief that Unikonts were primarily uniflagellate and some of them secondarily evolved bi- and multiflagellate forms, whereas the Bikonts were primarily biflagellate. This assumption has later proved wrong, being based on misinterpretation of the flagellar cycle in the myxogastrid slime mould *Physarum polycephalum* (Roger & Simpson 2009). As already observed, it is now generally agreed that extant eukaryotes originated from a biflagellate ancestor (Derelle et al. 2015). Therefore, the term Unikonts is clearly misleading and the term Bikonts should embrace all extant eukaryotes, thus being no longer useful. The gene fusion trait also turned out to be questionable because the Apusomonada, a group now associated with former unikonts, presents a bikont-like gene fusion (Roger & Simpson 2009).

Phylogenetic analysis based on larger datasets has confirmed the existence of a basal dichotomy in the eukaryote tree, with two major clades or "eukaryotic domains" encompassing most known taxa, plus several minor lineages of uncertain position but not forming together a separate clade (Adl et al. 2018). The two major clades were called Amorphea and Diaphoretickes (Fig. 8). The term Amorphea (shape-less) refers to the prevalence of ameboid cellular forms in the group, whereas the term Diaphoretickes (diverse) points to the vast diversity of forms included in the group.

The Amorphea, currently defined as the "least inclusive" (namely the smallest) clade encompassing *Homo sapiens*, *Neurospora crassa* and *Dictyostelium discoideum* (Adl et al. 2018), embraces most of the organisms previously referred to as unikonts, plus the Breviata and Apusomonada. Within the Amorphea, the animals (Metazoa) and fungi plus a few related protists lineages form the Opistokonta clade, which in turn forms the Obazoa clade with the Breviata and Apusomonada (Fig. 8). The second major amorphean clade is the Amoebozoa, encompassing a number of ameboid/flagellate protists (Fig. 8). The likely sister group to the Amoeboza is the CRuMs, a novel proposed clade named as an acronym of its constituent members: collodictyonids (syn. diphylleids) + Rigifilida + *Mantamonas*. These are free-living heterotrophic protozoa that cluster together in sequence analysis although exhibiting widely divergent cellular organizations (Burki et al. 2020).

The Diaphoretickes encompasses over half of extant eukaryotic diversity. Molecular phylogeny distinguishes four major diaphoretic lineages: the Archaeplastida (or Plantae), Cryptista, Haptista, and TSARs. The Archaeplastida comprises eukaryotes with a primary plastid, namely a plastid directly derived from a cyanobacterial endosymbiont (Section 8). These are the Chloroplastida or Viridiplantae (green algae and land plants), Rhodophyta (red algae) and Glaucophyta. The Cryptista include the Cryptophytes (microalgae with a secondary chloroplast) and some heterotrophic relatives (katablepharids and the recently discovered *Palpitomonas*). The Haptista comprises the Haptophytes, a large group of planktonic microalgae with a secondary chloroplast, and the Centrohelida, non-flagellate heterotrophic protozoa with radial projections that capture food and allow mobile forms to move about. Within the TSARs, the vast SAR (Stramenopiles-

Alveolata-Rhizaria) clade encompasses several major groups of microbial algae (e.g., diatoms, dinoflagellates, xantophytes), large seaweeds (kelps), ecologically important free-living protozoa (ciliates, foraminiferans, radiolarians), and many protozoan parasites (apicomplexans, oomycetes). The likely sister group of SARs is the enigmatic taxon *Telonemia*, with only two described species (Burki et al. 2020).

The larger eukaryotic assemblage that does not fit within either the Amorphea or Diaphoretikes is the Excavata, a protist group that owns the name to a ventral feeding groove present in some members (e.g., the Jacobida) and used for the capture of food. In the original definition by Stechmann and Cavalier-Smith (2003), the Excavata were placed within the Bikonta. The phylogenetic relationships and systematics of the Excavata have been the object of extensive investigation but are still incompletely defined. Phylogenetics and phylogenomics have brought to light two monophyletic subgroups, the Discoba and Metamonada, but have not consistently placed them together as a single clade (Simpson et al. 2017). Because of the lack of clear interrelationships in phylogenetic analyses, which is an indication of paraphyly, Adl et al. (2018) chose to report these lineages under the collective informal name "Excavates" rather than Excavata. The Discoba include the Euglenozoa (euglenids + kinetoplastids) and Heterolobosea, both characterized by dish-like mitochondrial cristae and therefore collectively named Discicristata, and the Jakobida, which have tubular cristae. The Metamonada are amitochondriate anaerobic protozoa, most living as symbionts or parasites of animals. The group includes the retortamonads, diplomonads (e.g., *Giardia*), oximonads, and parabasalids (e.g., *Trichomonas*).

In addition to the lineages mentioned above, there are several species-poor taxa for which phylogenomic analyses have so far failed to provide a convincing phylogenetic placement. Sometimes referred to as the "orphan" clades, these encompass the Ancoracysta, Picozoa, Malawimonadida, and ancyromonads (= planomonads), all free-living protozoa (Fig. 8).

The position of the root (origin) of the eTOL remains elusive. Hypotheses placing the root between the Opisthokonta and all other Eukaryotes (Katz & Grant 2015), between the Unikonta and Bikonta (Stechmann & Cavalier-Smith 2003; Derelle et al. 2015), or between the Excavata and the rest of eukaryotes (He et al. 2014) are inconsistent with the current eukaryote tree (Fig. 8). Cavalier-Smith & Chao (2020) place the root between the Discoba (which they call "Eozoa") and the rest of eukaryotes, which they name "neokaryotes".

9. Photosynthetic eukaryotes

Oxygenic photosynthesis evolved in cyanobacteria at least 2.7 GYA and was transferred to eukaryotes by endosymbiosis. The host was probably a freshwater biflagellate phagotrophic protist, whereas the closest extant relative of the cyanobacterial endosymbiont is probably *Gloeomargarita lithophora*, a member of an early-branched cyanobacterial lineage (Ponce-Toledo et al. 2017). Endowed with a photosynthetic machinery, the eukaryote host adopted an autotrophic lifestyle. Its descendants encased their cells in a cellulosic cell wall, lost phagotrophy (but see Maruyama et al. 2013), and generated the large clade named Archaeplastida (literally "ancient chloroplasts") or Plantae (plants). The archaeplastid lineage encompasses the Glaucophyta, Rhodophyta (red algae) and Viridiplantae. The order of divergence of the three clades is uncertain. Molecular evidence and the persistence of a discrete peptidoglycan layer in the chloroplast

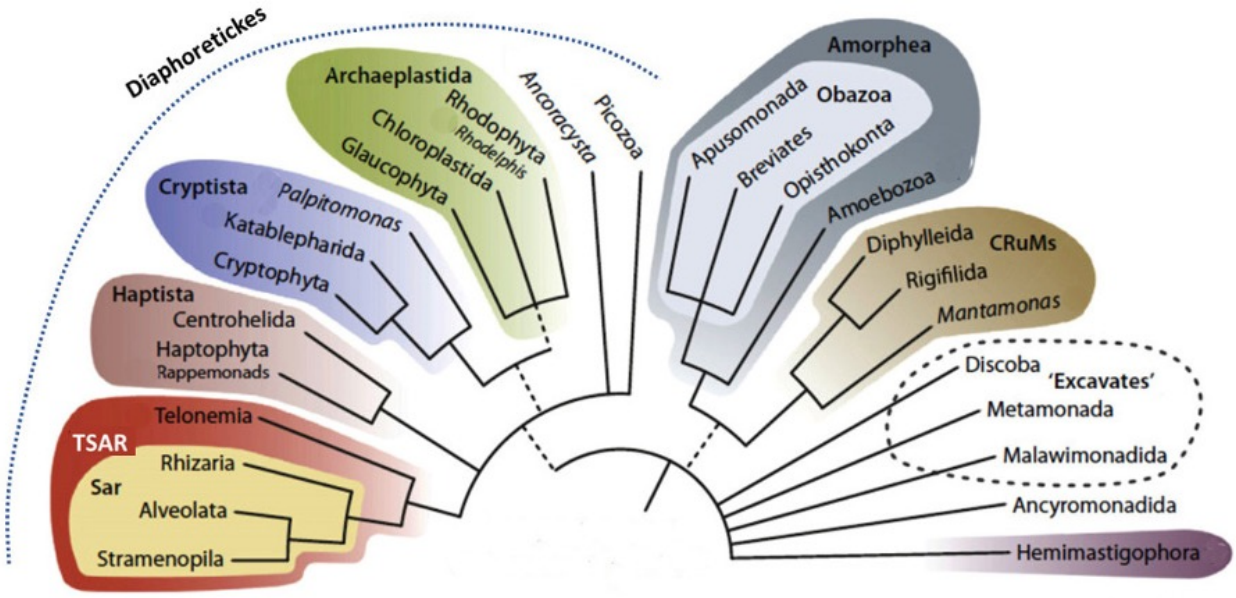

Figure 8: The eukaryote tree based on a consensus of recent phylogenomic studies. Most known eukaryotes fall within two major Domains named the Amorphea and Diaphoretickes. A third domain, reported under the acronym CruMs, has been resolved as the sister group to the Amorphea. Besides, the current eukaryote tree features three monophyletic lineages encompassing most excavate protists (Discoba, Metamonada and Malawimonadida) and two additional lineages (Ancyromonadida and Hemimastigophora) whose interrelationships and relative position are still undefined. Phylogenetic analysis of environmental sequences not associated with any defined organism has shown that almost all eukaryotic sequences can be assigned to known major groups. Broken lines reflect uncertainties about the monophyly of certain groups. Figure reproduced from Burki et al. (2020) under Creative Commons CC-BY license and modified by the author.

suggests that the Glaucophyta diverged first, whereas the Rhodophyta and Viridiplantae are probably sister groups in a separate clade (Li et al. 2014).

The archaeplastid chloroplast is a primary chloroplast because it directly descends from an enslaved cyanobacterium. The primary chloroplast features an inner and outer bounding membrane that are homologous with the inner and outer membrane of the cyanobacterial endosymbiont, respectively. The peptidoglycan layer that in cyanobacteria lies between the inner and outer membrane, is no longer visible in the chloroplast of plants except the Glaucophyta. Interestingly, there is evidence of a peptidoglycan layer enveloping the chloroplast in the moss *Physcomitrella patens*, which is essential for the organelle division (Hirano et al. 2016). The evolution of the chloroplast has been extensively reviewed (Gould et al. 2008; Howe et al. 2008; Keeling 2010, 2013; Dorrell & Howe 2012; Ligrone 2019) and will be considered here only in broad outlines.

Molecular evidence clearly shows that the chloroplast is monophyletic, namely it evolved only once and was vertically transmitted within the Archaeplastida and horizontally transmitted to other eukaryote lineages (Stiller et al. 2014). The only known exception to this is *Paulinella chromatophora*,

a filose amoeba harbouring a cyanobacterial endosymbiont related to the *Prochloron* lineage, which is at an advanced stage of conversion into a novel type of chloroplast (Nowack 2014). Molecular-clock analysis dated the origin of the *Paulinella* consortium to about 60 million years ago, whereas the archaeplastid chloroplast is much more ancient (Section 9).

The conversion of a cyanobacterial endosymbiont into a chloroplast followed much the same pathway as for the mitochondrion. In particular, the process involved the transfer of a substantial number of genes from the endosymbiont to the host nucleus. The maintenance of a modern chloroplast requires over one thousand genes, most of which are in the nucleus. The genome of a modern chloroplast, also known as the *plastome*, encompasses about 100 to 250 genes according to the taxon, and is almost entirely of cyanobacterial origin. Nuclear chloroplast genes include genes of cyanobacterial ancestry (about 50% of the total) and genes of eukaryotic origin, plus a minor stock of genes possibly acquired by horizontal gene transfer from other bacteria and secondarily deployed for

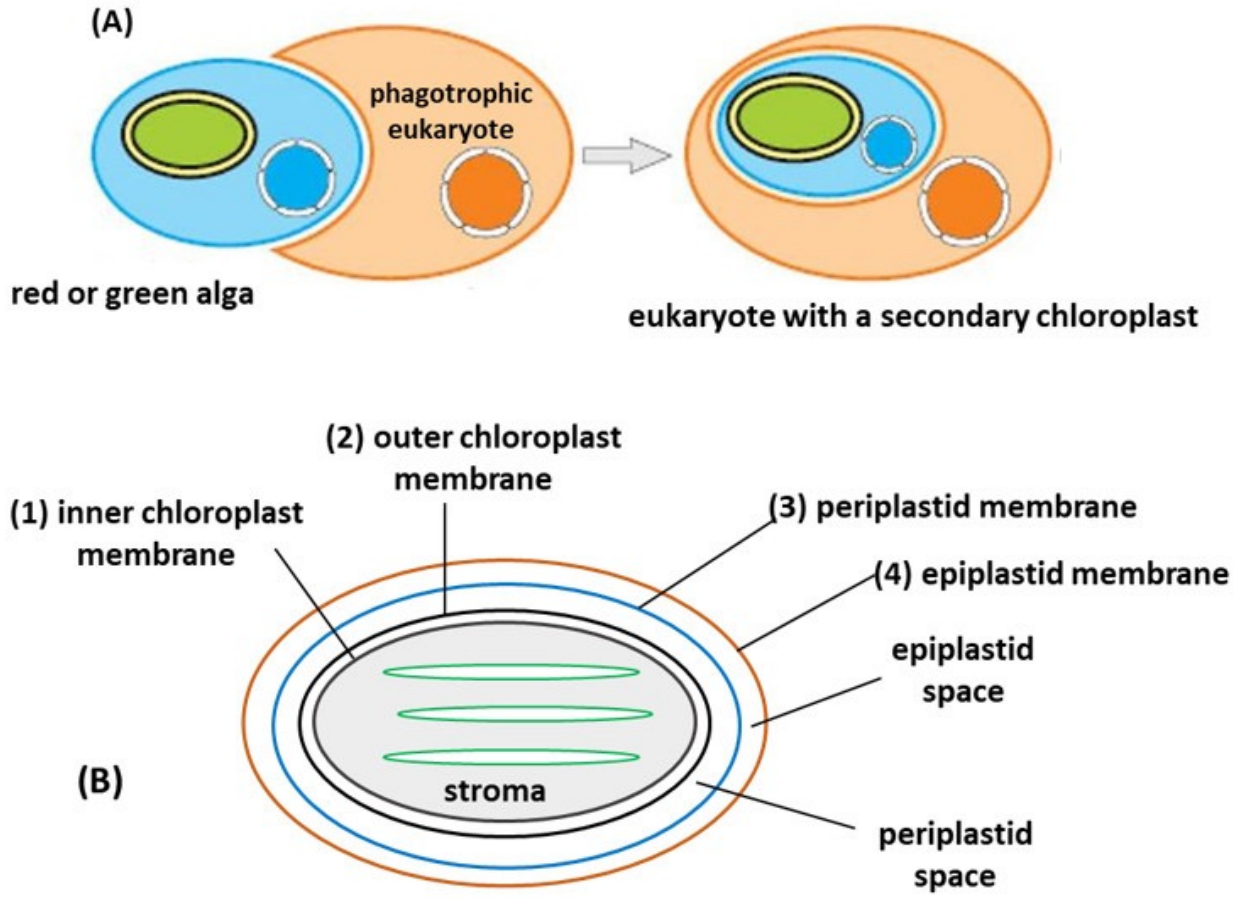

Figure 9: (A) Secondary chloroplasts arise from enslaved photosynthetic eukaryotes. **(B)** Besides the two original enveloping membranes (1 and 2), secondary chloroplasts usually have two extra bounding membranes named epi-and periplastid membrane. The periplastid space in the secondary chloroplasts of Chlorarachniophytes and Cryptophytes contains a rudimentary nucleus, known as the *nucleomorph*, which derives from the nucleus of the eukaryotic endosymbiont. Whereas primary chloroplasts lie in the cytoplasm, secondary chloroplast are topologically in the lumen of the endomembrane system of the secondary host. Redrawn and modified by the author from Ligrone 2019.

chloroplast maintenance (Archibald 2015b). In parallel with genomic reassortment, the symbiotic consortium developed the complex biochemical machinery necessary to transfer proteins encoded by nuclear genes and synthesized in the cytoplasm to the different topological compartments of the nascent chloroplast. This machinery encompasses proteins in part of cyanobacterial origin and in part from the host (Jarvis 2008; Thomson et al. 2020). In addition, functional integration of the symbionts required the evolution of carriers for the transport of small molecules such as phosphate, ATP, sugars and sugar-phosphates, amino acids, and ions across the chloroplast envelope, the majority of which are of eukaryotic origin (Karkar et al. 2015).

The core of the photosynthetic machinery of the cyanobacterial endosymbiont including chlorophyll *a*-based reaction centres was retained almost unchanged in the chloroplast. In contrast, accessory light-harvesting complexes underwent major changes underpinning adaptation of eukaryotic hosts to novel photic niches. Notably, the red algae evolved three novel phycobiliproteins, R-phycocyanin and R- and B-phycoerythrin, whilst retaining cyanobacterial allophycocyanin and C-phycoerythrin, whereas the Viridiplantae evolved chlorophyll b and completely lost the phycobiliproteins (Lee 2018). The family of enzymes responsible for cellulose synthesis in Archaeplastida most likely derives from *CesA*, a cyanobacterial gene inherited through the chloroplast (Nobles & Brown 2004). The ancestral storage polysaccharide of eukaryotes is glycogen, a multibranched polymer made of α-(1-4) glucan chains with α-(1-6) bonds at branching points. The Archaeplastida replaced ancestral glycogen with starch, a less densely branched α-(1-4) glucan, by deploying debranching isoamylases of cyanobacterial origin. Starch stored in the cytoplasm as observed in extant red algae and glaucophytes is thought to be the ancestral condition; starch synthesis was transferred to the chloroplast stroma in the Viridiplantae (Ball et al 2011).

Secondary chloroplasts arise from eukaryotic endosymbionts, and in most cases have two additional enveloping membranes besides the two original membranes (Fig. 9).

The secondary chloroplasts of the Euglenozoa and Chlorarachniophytes independently evolved from two unicellular chlorophytes, probably both belonging to the Ulvophyceae-Chlorophyceae-Trebouxiophyceae (UCT) clade (Rogers et al. 2007). Molecular evidence demonstrates that the secondary chloroplasts of Cryptophytes, Haptophytes, Stramenopiles and Dinophyta have a monophyletic origin from a red algal endosymbiont (Stiller et al. 2014). Besides plastome sequence homologies, these plastids contain chlorophyll *c* (with the exception of Chrompodellida, which probably secondarily lost it), and are collectively known as "brown chloroplasts" (Stiller et al. 2014). The demonstration of the monophyletic origin of brown chloroplasts lends support to the hypothesis originally put forward by Cavalier-Smith (1999) that the eukaryotes with brown chloroplasts form a monophyletic group, the "Chromoalveolates", derived from a photosynthetic ancestor. This hypothesis implies that non-photosynthetic chromoalveolates secondarily lost photosynthesis, as is the case of Apicomplexa, or the whole plastid as is the case of Oomycetes. Phylogenomic analysis of plastid sequences strongly supports the Cryptophyta, Alveolata, Stramenopila and Haptophyta as a clade, baptised as the CASH clade. Nevertheless, no significant support for CASH monophyly was obtained from mitochondrial or nuclear sequences

(Archibald 2015b). Moreover, phylogenomic analysis of mitochondrial and nuclear sequences consistently places non-photosynthetic stramenopiles in a basal position relative to their photosynthetic relatives (Simpson et al. 2017). Last, the current eTOL features the Cryptista and Haptista outside the Stramenopia-Alveolata-Rhizaria (SAR) clade (Fig. 8), thus suggesting Chromoalveolata paraphyly. Because of difficulty in reconciling the chromoalveolate scenario with molecular phylogenies, serial symbiosis hypotheses are gaining consensus (Bodył et al. 2009; Keeling 2013; Stiller et al. 2014; Bodył 2018). The serial scenario posits that the CASH chloroplast evolved once from a red algal symbiont and was then transmitted horizontally to other lineages through multiple endosymbiosis events. A corollary to this scenario is that the SAR, Cryptista and Haptista clades independently arose from a non-photosynthetic progenitor.

10. Eukaryote time scaling

In line with unresolved difficulty at reconstructing cellular transitions underpinning eukaryogenesis, dating the evolutionary history of eukaryotes has proven a challenging goal. Basically, there are three approaches to the issue. The first is paleontology, the search of fossilized eukaryotes in sedimentary rocks. The second is paleogeochemistry, the search for chemical markers of eukaryotic life in the geologic record. The third is molecular dating, the practice of inferring divergence times using molecular data from extant organisms.

The primary difficulty met by paleontology in the effort of dating eukaryote evolutionary history is the recognition of fossils as eukaryotic and, subsequently, their attribution to a specific lineage. Eukaryotes have great propensity to evolve multicellular forms, yet early eukaryotes were certainly unicellular, and still now unicellular forms account for most eukaryotic diversity. Morphological traits such as multi-layered cell walls or surface ornamentations, which in extant eukaryotes are under the control of the cytoskeleton and endomembrane system, may be regarded as eukaryotic signatures (Knoll 2014). This criterion leads to interpreting as eukaryotes large microfossils (100-300 m) with concentrically striated walls (acritarchs) such as *Tappania* and *Valeria*, dated to about 1.6 GYA (Javaux & Lepot 2018). A macroscopic multicellular organization is recognized as an indicative trait of eukaryotes. A likely candidate is *Bangiomorpha pubescens*, a filamentous fossil tentatively interpreted as a red alga akin to present-living *Bangia*. *Bangiomorpha* was originally dated to about 1.2 GYA (Butterfield et al. 1990) and subsequently post-dated to about 1.0 GYA (Gibson et al. 2018). More recently, decimetre-scale thalloid fossils have been reported from the 1,560-Myr-old Gaoyuzhuang Formation, North China, and suggested to be benthic photosynthetic eukaryotes (Zhu et al. 2015).

Despite paleontological evidence suggesting an earlier origin, undisputed eukaryotic fossils antedating 800 MYA are rare and poorly diversified, becoming more abundant and diverse between 800 and 720 MYA, as documented by the detection of numerous novel "species" of vase-shaped protists and siliceous microfossils (Knoll 2014). The palaeontological record from this interval also encompasses likely representatives of extant eukaryotic clades such as red and green algae, heterokonts, amoebozoans, cercozoans and fungi. After a long pause concomitant with the coldest phase of Neoproterozoic glaciations, a multitude of eukaryotic microfossils reappeared from 660 MYA. A worldwide shift to a higher Zn/C ratio in marine sediments may be evidence that about 800 MYA eukaryotic phytoplankton replaced

cyanobacteria as the main primary producers (Isson et al. 2018).

Steranes, a class of 4-cyclic compounds derived from spontaneous degradation of steroids or sterols, are considered eukaryotic markers. Brocks et al. (1999) reported steranes in 2.7-billion-year-old sedimentary rocks. Subsequent investigation showed that these molecules were probably contaminants from above-lying, more recent layers (Rasmussen et al. 2008). Further study reported steranes in Transvaal Supergroup sediments dated 2.67 to 2.46 GYA, yet no significant amounts of steranes were found in cores extracted from the 2.6 GY-old Pilbara Craton in Australia (see review by Knoll 2014). Cholesterol and other sterols have a crucial role in the control of membrane fluidity in modern eukaryotes (Subczynski et al. 2017). It is worth noting that sterol biosynthesis requires molecular oxygen (Desmond and Gribaldo 2009), yet paleogeochemical data indicate that the Archaean atmosphere and oceans have been essentially anoxic until about 2.4 GYA (Bekker 2014). This is, therefore, the maximum age constraint for the evolution of the mitochondrion. Steranes start being to be regularly present in sedimentary rocks after 800 MYA, documenting the onset of Neoproterozoic eukaryotic diversification. Notably, the occurrence of C_{28} e C_{29} sterans in rocks aged about 750 MY suggests that by that time green and red algae had become major primary producers on continental shelves (Kodner et al. 2008).

The molecular-clock approach aims at dating the divergence between branches of a phylogenetic tree by estimating the rate of mutation in homologous sequences. Because the mutation rate varies with taxonomy and other variables, molecular-clock analysis needs calibration from fossils. The procedure works well with relatively recent clades but becomes increasingly haphazard for deep cladogenesis. Molecular-clock analysis of eukaryotes has produced results ranging over extremely wide intervals, often dating deep events long before the fossil record. This is in part understandable, because the paleontological record provides a minimum, not a maximum age constraint. Nevertheless, divergences of several hundred million years cast doubts on the soundness of the approach. Besides problems with fossil misinterpretation, saturation of sequence changes, clock calibration, or analytical methods, molecular-clock analysis of deep eukaryote evolution also faces problems from authors' own biases and reference to questionable phylogenetic framework. As a matter of fact, estimates for the first eukaryote (FECA) appearance range from over 3.4 GYA (Betts et al. 2018) to about 2.7 GYA (Hedges et al. 2004), both dates largely predating the first undisputed paleontological record of eukaryotic life. LECA was dated to about 2.0 GYA (Hedges et al. 2004), 1.866-1.679 GYA (Parfrey et al. 2010; Betts et al. 2018), 1.9-1.0 GYA (Eme et al. 2014), 1.3-1.0 GYA (Chernikova et al. 2011), or definitely around 1.2 GYA (Douzery et al. 2004; Shih & Matzeb 2013). A relatively late LECA is consistent with estimates placing the divergence of major modern eukaryotic lineages between 1.0 and 0.8 GYA (Douzery et al. 2004; Chernikova et al. 2011; Eme et al. 2014). Shih & Matzeb (2013) date the evolution of the mitochondrion to about 1.2 GYA, thus linking this event to the appearance of LECA. The same study dates the evolution of the primary chloroplast, hence of the Archaeplastida, to 0.857-1.055 GYA, a range just compatible with the current dating of *Bangiomorpha* (Gibson et al. 2018).

Conclusions

The origin of the eukaryotic cell remains an open issue because none of the models

presented so far accounts for the whole body of evidence available.

The biological success of eukaryotes is due to multiple factors. The first was phagocytosis coupled with the mitochondrion, an organelle specialized in energy production. Whether phagocytosis was an early trait of eukaryotes, as strongly argued by Cavalier-Smith (2009, 2014), or post-dated the acquisition of the mitochondrion (Martin et al. 2017; Spang et al. 2019), it is incontestable that eukaryotes primarily evolved as predatory phagotrophs. Cellular predation does exist among bacteria, and predatory bacteria display a diversity of hunting and killing strategies including a phagocytosis-like mechanism (Pérez et al. 2016; Shiratori et al. 2019). Eukaryotes, however, perfected the predatory strategy to a level of complexity that has no equal in the prokaryotic world. Indeed, most eukaryotic traits such as a multi-compartmented endomembrane system, a novel cell division machinery, and the 9+2 flagellum appear to be directly linked to phagocytosis, whereas mitosis probably evolved as an indirect consequence (Cavalier-Smith 2009, 2014).

Although a microtubule-based primitive mitosis probably appeared early in eukaryogenesis, the nuclear envelope and its nuclear-pore machinery might have evolved after the mitochondrion. On the other hand, the mitochondrion could only evolve after the Great Oxygenation Event (GOE), dated to about 2.4 GYA. Geochemical evidence suggests that oxygen rose to about 0.8% or even more between 2.4 and 2.2 GYA, but then plummeted to about 0.02% and remained around this level up to 0.8 GYA (Lyons et al 2014; Cole et al. 2016). An oxygen concentration of 0.02% is much below the Pasteur point (0.3%) at which facultative aerobic organisms can switch from fermentation to aerobic respiration. An oxygen concentration around 0.02% during most of the Proterozoic, therefore, appears to contrast with molecular-clock dating of LECA before 1.0 GYA. An independent study based on the distribution of redox-sensitive trace metals points to an oxygen level as high as 0.8% around 1.4 GYA (Zhang et al. 2016), which is in line with estimates of LECA appearance around 1.3 GYA. The persistence of relatively low oxygen levels during most of the Proterozoic may account for the widespread distribution of anaerobic biochemistry in extant eukaryotes (Mentel & Martin 2008; Stairs et al. 2015).

The second key to eukaryote success was photosynthesis, horizontally acquired through a second event of endosymbiosis after the mitochondrion. The diversification of Archaeplastida and the spread of the chloroplast across a wide spectrum of eukaryote diversity involved major changes in complementary photosynthetic pigments, reflecting adaptation to specific photic niches. The rise of photosynthetic eukaryotes increased global productivity by orders of magnitude and was probably the main driver of the Late Proterozoic planetary shift to a highly oxygenated state, which in turn paved the way to the Phanerozoic burst of complex life (Geider et al. 2001; Falkowski et al 2004; Brocks et al. 2017).

The third major factor behind the biological success of eukaryotes was their ability to evolve a complex multicellular organization, which appeared independently multiple times (Niklas & Newman 2013). Niche construction activity (Laland et al. 2017) by multicellular eukaryotes in the Phanerozoic (538.8 MYA to the present) expanded ecosystem complexity to a level never attained in the preceding three billion years, creating unprecedented opportunities for biological evolution.

References

Adl SM et al (2018). Revisions to the nomenclature, classification and diversity

of eukaryotes. Journal of Eukaryotic Microbiology. Doi: 10.1111/jeu.12691

Alvarez-Ponce D et al (2013). Gene similarity networks provide tools for understanding eukaryote origins and evolution. Proceedings of the National Academy of Sciences USA 110: E1594-E1603; https://doi.org/10.1073/pnas.1211371110

Archibald JM (2015a). Endosymbiosis and eukaryotic cell evolution. Current Biology 25: R911-R921. http://dx.doi.org/10.1016/j.cub.2015.07.055

Archibald JM (2015b). Genomic perspectives on the birth and spread of plastids. Proceedings of the National Academy of Sciences USA 112: 10147-10153.

Archibald JM et al, eds. (2017). Handbook of the Protists. Springer International Publishing. Doi: 10.1007/978-3-319-32669-6_45-1

Baker BJ et al (2020). Diversity, ecology and evolution of Archaea. Nature Microbiology 5: 887-900.

Ball S et al (2011). The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. Journal of Experimental Botany 62: 1775-1801

Balleza D et al (2014). Ether- versus ester-linked phospholipid bilayers containing either linear or branched apolar chains. Biophysical Journal 107: 1364-1374.

Bekker A (2014). Great Oxygenation Event. Encyclopedia of Astrobiology. Doi 10.1007/978-3-642-27833-4_1752-4

Bernier CR et al (2018). Translation: the universal structural core of life. Molecular Biology and Evolution 35: 2065-2076. Doi: 10.1093/molbev/msy101

Betts HC et al (2018). Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin, Nature Ecology & Evolution 2: 1556-1562.

Bodył A (2018). Did some red alga-derived plastids evolve via kleptoplastidy? A hypothesis. Biological Reviews of the Cambridge Philosophical Society 93: 201-222. Doi: 10.1111/brv.12340

Bodył A, Stiller JW, Mackiewicz P (2009) .Chromalveolate plastids: direct descent or multiple endosymbioses? Trends Ecol. Evol. 24: 119-121

Brocks JJ et al (2017). The rise of algae in Cryogenian oceans and the emergence of animals. Nature 548: 578-58.

Brueckner J, Martin WF (2020). Bacterial genes outnumber archaeal genes in eukaryotic genomes. Genome Biology and Evolution 12: 282-292.

Bulzu PA et al. (2019). Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche. Nature Microbiology 4: 1129-1137.

Butterfield NJ, Knoll A., Swett K (1990). A bangiophyte red alga from the Proterozoic of arctic Canada. Science 250: 104-107.

Cabeen MT, Jacobs-Wagner C (2010). The bacterial cytoskeleton. Annual Review of Genetics 44: 365-392

Cavalier-Smith T (1989). Archaebacteria and Archaezoa. Nature 339: 100-101.

Cavalier-Smith T (2005). Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion. Annals of Botany 95: 147-175

Cavalier-Smith T (2006). Cell evolution and Earth history: stasis and revolution. Philosophical Transactions of the Royal Society B 361: 969–1006.

Cavalier-Smith T (2009). Predation and eukaryote cell origins: a coevolutionary perspective. International Journal of Biochemistry & Cell Biology 41: 307-322.

Cavalier-Smith T (2010a). Origin of the cell nucleus, mitosis and sex: roles of intracellular coevolution. Biology Direct 5:

7. http://www.biology-direct.com/content/5/1/7

Cavalier-Smith T (2010b). Kingdoms Protists and Chromista and the eozoan root of eukaryotes. Biology Letters 6: 342-345

Cavalier-Smith T (2014). The neomuran revolution and phagotrophic origin of eukaryotes and cilia in the light of intracellular coevolution and a revised tree of life. Cold Spring Harbor Perspectives in Biology 6:a016006. Doi: 0.1101/cshperspect.a016006

Cavalier-Smith T, Chao E E-Y (2020). Multidomain ribosomal protein trees and the planctobacterial origin of neomura (eukaryotes, archaebacteria). Protoplasma 257: 621-753.

Chatton E (1925). *Pansporella perplexa*. Réflexions sur la biologie et la phylogénie des protozoaires. Annales des Sciences Naturelles Zoologie. 10e serie, VII: 1-84.

Chernikova D et al (2011). A late origin of the extant eukaryotic diversity: divergence time estimates using rare genomic changes. Biology Direct 6: 26. http://www.biology-direct.com/content/6/1/26

Cole DB et al (2016). A shale-hosted Cr isotope record of low atmospheric oxygen during the Proterozoic. Geology 44: 555-558.

Cox CJ et al (2008). The archaebacterial origin of eukaryotes. Proceedings of the National Academy of Sciences USA 105: 20356-20361.

Dacks JB et al (2016). The changing view of eukaryogenesis - fossils, cells, lineages and how they all come together. Journal of Cell Science 129: 3695-3703. doi: 10.1242/jcs.178566.

Dacks JB, Field MC (2018). Evolutionary origins and specialisation of membrane transport. Current Opinion in Cell Biology 53: 70-76. Doi: 10.1016/j.ceb.2018.06.001.

Desmond E, Gribaldo S (2009). Phylogenomics of sterol synthesis: insights into the origin, evolution, and diversity of a key eukaryotic feature. Genome Biology and Evolution 1: 364-381.

Derelle R et al (2015). Bacterial proteins pinpoint a single eukaryotic root. Proceedings of the National Academy of Sciences USA 112: E693-E699. Doi: 10.1073/pnas.1420657112.

De Rosa M, Gambacorta A, Gliozzi A (1986). Structure, biosynthesis, and physicochemical properties of archaebacterial lipids. Microbiological Reviews. 50: 70-80.

Dey G, Thattai M, Baum B (2016). On the archaeal origins of eukaryotes and the challenges of inferring phenotype from genotype. Trends in Cell Biology 26: 476-485.

Dorrell RG, Howe CJ (2012). What makes a chloroplast? Reconstructing the establishment of photosynthetic symbioses. Journal of Cell Science 125: 1865-1875.

Egan AJ, Vollmer W (2013). The physiology of bacterial cell division. Annals of the New York Academy of Sciences 1277: 8-28.

Elliott TA, Gregory TR (2015). What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. Philosophical Transactions of the Royal Society B 370: 20140331 http://dx.doi.org/10.1098/rstb.2014.0331.

Eme L et al (2017). Archaea and the origin of eukaryotes. Nature Microbiology 15: 711-723.

Falkowski PG et al (2004) The evolution of modern eukaryotic phytoplankton. Science 305: 354-360.

Field MC, Rout MP (2019). Pore timing: the evolutionary origins of the nucleus and nuclear pore complex. F1000Research 8: 369, https://doi.org/10.12688/f1000research.16402.1.

Fritz-Laylin LK et al (2010). The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. Cell 140: 631-642.

Forterre P (2013). The common ancestor of Archaea and Eukarya was not an archaeon. Hindawi Publishing Corporation, Archaea Volume, Article ID 372396, http://dx.doi.org/10.1155/2013/372396.

Gabaldón T (2018). Relative timing of mitochondrial endosymbiosis and the "pre-mitochondrial symbioses" hypothesis. IUBMB Life, https://doi.org/10.1002/iub.1950.

Geider RJ et al (2001). Primary productivity of planet earth: biological determinants and physical constraints in terrestrial and aquatic habitats. Global Change Biology 7: 849-882.

Gibson TM et al (2018). Precise age of *Bangiomorpha pubescens* dates the origin of eukaryotic photosynthesis. Geology 46: 135-138.

Glansdorff N, Xu Y, Labedan B (2008). The Last Universal Common Ancestor: emergence, constitution, and genetic legacy of an elusive forerunner. Biology Direct 3: 29. Doi: 10.1186/1745-6150-3-29.

Gould SB, Garg SG, Martin WF (2016). Bacterial vesicle secretion and the evolutionary origin of the eukaryotic endomembrane system. Trends in Microbiology 24: 525-534. Doi: https://doi.org/10.1016/j.tim.2016.03.005.

Gould SB, Waller RF, McFadden GI (2008). Plastid evolution. Annual Review of Plant Biology 59: 491–517.

Guy L, Ettema TJ (2011). The archaeal 'TACK' superphylum and the origin of eukaryotes. Trends in Microbiology 19: 580-587.

Haeckel E (1866). Generelle Morphologie der Organismen. Reimer, Berlin.

Hampl V, Čepička I, Eliáš M (2019). Was the mitochondrion necessary to start eukaryogenesis? Trends in Microbiology 27: 96-104.

He D et al (2014). An alternative root for the eukaryote tree of life. Current Biology 24: 465-470. http://dx.doi.org/10.1016/j.cub.2014.01.036.

Hirano T et al (2016). Moss chloroplasts are surrounded by a peptidoglycan wall containing D-amino acids. The Plant Cell 28: 1521-1532.

Hofstatter PG, Lahr DJ (2019). All eukaryotes are sexual, unless proven otherwise: many so-called asexuals present meiotic machinery and might be able to have sex. Bioessays 41: e1800246.

Howe CJ et al (2008). The origin of chloroplasts. Philosophical Transactions of the Royal Society B 363: 2675-2685.

Hug LA et al (2016) .A new view of the tree of life. Nature Microbiology 1: 16048. https://doi.org/10.1038/nmicrobiol.2016.48.

Imachi H et al (2020). Isolation of an archaeon at the prokaryote-eukaryote interface. Nature 577: 519-525.

Irimia M, Roy SW (2014). Origin of spliceosomal introns and alternative splicing. Cold Spring Harbor Perspectives in Biology 6: a016071.

Javaux EJ, Lepot K (2018). The Paleoproterozoic fossil record: Implications for the evolution of the biosphere during Earth's middle-age. Earth-Science Reviews 176: 68-86.

Jékely G. (2008). Origin of the nucleus and Ran-dependent transport to safeguard ribosome biogenesis in a chimeric cell. Biology Direct 3: 31. Doi: 10.1186/1745-6150-3-31.

Jaquemet A et al (2009). Archaeal tetraether bipolar lipids: Structures, functions and applications. Biochimie 91: 711-717.

Jarvis P (2008). Targeting of nucleus-encoded proteins to chloroplasts in plants. New Phytologist 179: 257-285.

Junge W, Nelson N (2015). ATP synthase. Annual Review of Biochemistry 84: 631-657.

Karkar S et al (2015). Metabolic connectivity as a driver of host and endosymbiont integration. Proceedings of the National Academy of Sciences USA 112: 10208-10215.

Katz LA, Grant JR (2015). Taxon-rich phylogenomic analyses resolve the eukaryotic tree of life and reveal the power of subsampling by sites. Systematic Biology 64: 406-415.

Keeling PJ (1998). A kingdom's progress: Archezoa and the origin of eukaryotes. Bioessays 20: 87-95.

Keeling PJ (2010). The endosymbiotic origin, diversification and fate of plastids. Philosophical Transactions of the Royal Society B 365: 729-748.

Keeling PJ (2013). The number,speed, and impact of plastid endosymbioses in eukaryotic evolution. Annual Review of Plant Biology 64: 583-607.

Klinger CM et al (2016). Tracing the archaeal origins of eukaryotic membrane-trafficking system building blocks. Molecular Biology and Evolution 33: 1528-1541.

Knoll AH (2014). Paleontology of eukaryotes. Cold Spring Harbor Perspectives in Biology; 6: a016121. Doi: 10.1101/cshperspect.a016121.

Kodner RB et al (2008). Sterols in red and green algae: quantification, phylogeny and relevance for the interpretation of geologic steranes. Geobiology 6: 411-420.

Koonin EV (2010). The origin and early evolution of eukaryotes in the light of phylogenomics. Genome Biology 11: 209-221.

Koonin EV (2015). Origin of eukaryotes from within archaea, archaeal eukaryome and bursts of gene gain: eukaryogenesis just made easier? Philosophical Transactions of the Royal Society B 370: 20140333 http://dx.doi.org/10.1098/rstb.2014.0333

Koonin EV, Yutin N (2014). The dispersed archaeal eukaryome and the complex archaeal ancestor of eukaryotes. Cold Spring Harbor Perspectives in Biology; 6:a016188.Doi:10.1101/cshperspect.a016188.

Koreny L, Field MC (2016). Ancient eukaryotic origin and evolutionary plasticity of nuclear lamina. Genome Biology and Evolution 8: 2663-2671.

Koumandou VL et al (2013). Molecular paleontology and complexity in the last eukaryotic common ancestor. Critical Reviews in Biochemistry and Molecular Biology 48: 373-396.

Ku et al (2015a). Endosymbiotic gene transfer from prokaryotic pangenomes: inherited chimerism in eukaryotes. Proceedings of the National Academy of Sciences USA 112: 10139-10146.

Ku C et al (2015b). Endosymbiotic origin and differential loss of eukaryotic genes. Nature 524: 427-432.

Lake JA et al (1984). Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. Proceedings of the National Academy of Sciences 81: 3786-3790.

Lake JA (2015). Eukaryotic origins. Philosophical Transactions of the Royal Society B 370: 20140321. http://dx.doi.org/10.1098/rstb.2014.0321.

Laland K, Odling-Smee J, Endler J (2017). Niche construction, sources of selection and trait coevolution. Interface Focus 7: 20160147. http://dx.doi.org/10.1098/rsfs.2016.0147.

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

Leander BS (2020) Predatory protists. Current Biology 30: R451-R520. Doi: 10.1016/j.cub.2020.03.052.

Lee RE (2018). Phycology. Cambridge University Press. ISBN: 9781316407219.

Li B et al (2014). Compositional biases among synonymous substitutions cause conflict between gene and protein trees for plastid origins. Molecular Biological Evolution 31: 1697-1709.

Ligrone R (2019). The chloroplast and photosynthetic eukaryotes. In: Biological innovations that built the world. A four-billion-year journey through life and Earth history, p. 269-310. Springer.

Ligrone R (2021). Death, sex, and immortality. Bulletin of Regional Natural History 1: 1-19. Doi: https://doi.org/10.6093/2724-4393/8315.

Lombard J, Moreira D (2011). Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. Molecular Biology and Evolution 28: 87-99.

Lombard J, López-García P, Moreira D (2012). The early evolution of lipid membranes and the three domains of life. Nature Reviews Microbiology 10: 507-515.

Lòpez-Garcìa P, Moreira D (2006). Selective forces for the origin of the eukaryotic nucleus. BioEssays 28: 525-533.

Lòpez-Garcìa P, Moreira D (2020). The syntrophy hypothesis for the origin of eukaryotes revisited. Nature Microbiology, https://doi.org/10.1038/s41564-020-0710-4.

Lynch M, Marinov GK (2015). The bioenergetic costs of a gene. Proceedings of the National Academy of Sciences USA 112: 15690-15695.

Lyons TW, Reinhard CT, Planavsky NJ (2014a) .The rise of oxygen in Earth's early ocean and atmosphere. Nature 506: 307-315.

Makarov AA et al (2021). Evolution and diversification of the nuclear pore complex. Biochemical Society Transactions. Doi: org/10.1042/BST20200570.

Makarova KS et al (2005). Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. Nucleic Acids Research 33: 4626-4638.

Makarova KS et al (2010). Evolution of diverse cell division and vesicle formation systems in Archaea. Nature Reviews Microbiology 8: 731-741.

Makarova KS, Koonin EV (2013). Archaeology of eukaryotic DNA replication. Cold Spring Harbor Perspectives in Biology 5:a012963. Doi: 10.1101/cshperspect.a012963.

Margulis L, Corliss JO, Melkonian M, Chapman DJ, eds (1990). Handbook of Protoctista. Jones & Bartlett Learning, Sudbury.

Martin W, Garg S, Zimorski V (2015). Endosymbiotic theories for eukaryote origin. Philosophical Transactions of the Royal Society B 370: 20140330. http://dx.doi.org/10.1098/rstb.2014.0330.

Martin W, Koonin EV (2006). Introns and the origin of nucleus-cytosol compartmentalization. Nature 440: 41-45.

Martin WF (2017). Symbiogenesis, gradualism, and mitochondrial energy in eukaryote evolution. Periodicum Biologorum 119: 141-158.

Martin WF et al (2017). The physiology of phagocytosis in the context of mitochondrial origin. Microbiology and Molecular Biology Reviews 81: e00008-17. https://doi.org/10.1128/MMBR.00008-17.

Martijn J et al (2018). Deep mitochondrial origin outside the sampled alphaproteobacteria. Nature 557: 101-105.

Maruyama S et al (2013). A modern descendant of early green algal phagotrophs. Current Biology 23: 1081-1084.

May KL, Grabovicz M (2018). The bacterial outer membrane is an evolving antibiotic barrier. Proceedings of the National Academy of Sciences USA 115: 8852-8854.

McIntosh JR (2016). Mitosis. Cold Spring Harbor Perspectives in Biology 8: a023218 doi: 10.1101/cshperspect.a023218.

Melnikov S et al (2012). One core, two shells: bacterial and eukaryotic ribosomes. Nature Structural & Molecular Biology 19: 560-567.

Mentel M, Martin W (2008). Energy metabolism among eukaryotic anaerobes in light of Proterozoic ocean chemistry. Philosophical Transactions of the Royal Society B 363: 2717-2729.

Mills DB. 2020. The origin of phagocytosis in Earth history. Interface Focus 10: 20200019. http://dx.doi.org/10.1098/rsfs.2020.0019.

Müller M et al (2012). Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. Microbiology and Molecular Biology Reviews 76: 444-495.

Naor A, Gophna R (2013). Cell fusion and hybrids in Archaea. Bioengineered 4:126-129.

Ng SY et al (2008). Cell surface structures of archaea. Journal of Bacteriology 190: 6039-6047.

Nikaido H (2003). Molecular basis of bacterial outer membrane permeability revisited. Microbiology and Molecular Biology Reviews 67: 593-656.

Niklas KJ, Newman SA (2013) The origins of multicellular organisms. Evolution & Development 15: 41-52.

Nobles DR, Brown RM (2004). The pivotal role of cyanobacteria in the evolution of cellulose synthases and cellulose synthase-like proteins. Cellulose 11: 437-448.

Nowack ECM (2014). *Paulinella chromatophora* - rethinking the transition from endosymbiont to organelle. Acta Societatis Botanicorum Poloniae 83: 387-397.

O'Malley M et al (2019). Concepts of the last eukaryotic common ancestor. Nature Ecology & Evolution 3: 338-344.

Pérez J et al (2016). Bacterial predation: 75 years and counting! Environmental Microbiology 18: 766-779.

Pittis AA, Gabaldón T (2016). Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. Nature 531: 101-104

Ponce-Toledo RI et al (2017). An Early-branching freshwater cyanobacterium at the origin of plastids. Current Biology 27: 386-391.

Poole AM, Neumann N (2011). Reconciling an archaeal origin of eukaryotes with engulfment: a biologically plausible update of the eocyte hypothesis. Research in Microbiology 162: 71-76.

Poole AM, Gribaldo S (2014). Eukaryotic origins: how and when was the mitochondrion acquired? Cold Spring Harbor Perspectives in Biology 6:a015990. Doi: 10.1101/cshperspect.a015990.

Raymann K, Brochier-Armanet C, Gribaldo S (2015). The two-domain tree of life is linked to a new root for the Archaea. Proceedings of the National Academy of Sciences 112: 6670-6675.

Rinke W et al (2013). Insights into the phylogeny and coding potential of microbial dark matter. Nature 499: 431-437.

Roger AJ, Muñoz-Gómez SA, Kamikawa R (2017). The origin and diversification of mitochondria. Current Biology 27: R1177-R1192.

Roger AJ, Simpson AGB (2009). Evolution: revisiting the root of the eukaryote tree. Current Biology 19: R165-R166.

Rogers MB et al (2007). The complete chloroplast genome of the Chlorarachniophyte *Bigelowiella natans*: evidence for independent origins of chlorarachniophyte and euglenid

secondary endosymbionts. Molecular Biological Evolution 24: 54-62.

Rojas ER et al (2018). The outer membrane is an essential load-bearing element in Gram-negative bacteria. Nature 559: 617-621.

Saw JH et al (2015). Exploring microbial dark matter to resolve the deep archaeal ancestry of eukaryotes. Philosophical Transactions of the Royal Society B 370: 20140328 http://dx.doi.org/10.1098/rstb.2014.0328

Schlacht A et al (2014). Missing pieces of an ancient puzzle: evolution of the eukaryotic membrane-trafficking system. Cold Spring Harbor Perspectives in Biology. Doi: 10.1101/cshperspect.a016048.

Schopf IJ et al (2018). SIMS analyses of the oldest known assemblage of microfossils document their taxon-correlated carbon isotope compositions. 115: 53-58.

Schulz-Vogt HN, Angert ER, Garcia-Pichel F (2007). Giant bacteria. Encyclopedia of Life Sciences. Doi: 10.1002/9780470015902.a0020371.pub2

Shalev Y et al (2017). Cell surface glycosylation is required for efficient mating of *Haloferax volcanii*. Frontiers in Microbiology 8:1253. Doi: 10.3389/fmicb.2017.01253.

Shih PM, Matzkeb NJ (2013). Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins. Proceedings of the National Academy of Sciences USA 110: 12355-12360.

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

Shiratori T (2019). Phagocytosis-like cell engulfment by a planctomycete bacterium. Nature Communications 10: 5529. Doi: 10.1038/s41467-019-13499-2.

Silar P (2016). Protistes eucaryotes: origine, evolution et biologie des microbes eucaryotes. https://hal.archives-ouvertes.fr/hal-01263138.

Simpson AGB, Slamovits CH, Archibald JM (2017). Protist diversity and eukaryote phylogeny. In (Archibald JM et al. eds.) Handbook of the Protists, p 1-21. Springer International Publishing AG, Cham.

Skejo J (2021). Evidence for a syncytial origin of eukaryotes from ancestral state reconstruction. Genome Biology and Evolution 13. Doi:10.1093/gbe/evab096.

Skophammer RG et al (2007). Evidence for a gram-positive eubacterial root of the tree of life. Molecular Biology and Evolution 24: 1761-1768.

Spang A et al (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature 521: 173-184.

Spang, A. et al (2018). Asgard archaea are the closest prokaryotic relatives of eukaryotes. PLoS Genetics 14, e1007080, Doi: 10.1371/journal.pgen.1007080.

Spang A et al (2019). Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. Nature Microbiology 4: 1138–1148.

Spejer D (2020a). Debating eukaryogenesis - Part 1: Does eukaryogenesis presuppose symbiosis before uptake? Bioessays 42:e1900157. Doi: 10.1002/bies.201900157.

Spejer D (2020b). Debating eukaryogenesis - Part 2: How anachronistic reasoning can lure us into inventing intermediates. Bioessays 42:e1900153. Doi: 10.1002/bies.201900153.

Stairs CW, Leger MM, Roger AJ (2015). Diversity and origins of anaerobic metabolism in mitochondria and related organelles. Philosophical Transactions Royal Society. B 370: 20140326. http://dx.doi.org/10.1098/rstb.2014.0326.

Stanier RY, van Niel CB (1962). The concept of a bacterium. Archiv für Mikrobiologie 42:17-35.

Stechmann A, Cavalier-Smith T (2002) .Rooting the eukaryote tree by using a derived gene fusion. Science 297: 89-91.

Stiller JW et al (2014). The evolution of photosynthesis in chromist algae through serial endosymbiosis. Nature Communications 5: 5764. Doi: 10.1038/ncomms6764

Stouf M, Meile JC, Cornet F (2013). FtsK actively segregates sister chromosomes in *Escherichia coli*. Proceedings of the National Academy of Sciences 110: 11157-11162.

Subczynski WK et al (2017). High cholesterol/low cholesterol: effects in biological membranes. A review. Cell Biochemistry & Biophysics 75: 369-385.

Theriot JA (2013). Why are bacteria different from eukaryotes? BMC Biology 11: 119 http://www.biomedcentral.com/1741-7007/11/119.

Thiergart T et al (2012). An evolutionary network of genes present in the eukaryote common ancestor polls genomes on eukaryotic and mitochondrial origin. Genome Biology and Evolution 4: 466-485.

Thomson S et al (2020). Protein import into chloroplasts and its regulation by the ubiquitin-proteasome system. Biochemical Society Transactions 48: 71-82.

Toro E, Shapiro L (2010). Bacterial chromosome organization and segregation. Cold Spring Harbor Perspectives in Biology 2: a000349. Doi: 10.1101/cshperspect.a000349.

Tria FDK et al (2021). Gene duplications trace mitochondria to the onset of eukaryote complexity. Genome Biology and Evolution 13(5). Doi: 10.1093/gbe/evab055.

Tromer EC et al (2019). Mosaic origin of the eukaryotic kinetochore. Proceedings of the National Academy of Sciences 116: 12873-12882.

Ueno Y et al (2006). Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. Nature 440: 516-519.

Valas RE, Bourne PE (2011). The origin of a derived superkingdom: how a gram-positive bacterium crossed the desert to become an archaeon. Biology Direct 6: 16. Doi: 10.1186/1745-6150-6-16.

Valentine DL (2007). Adaptations to energy stress dictate the ecology and evolution of the Archaea. Nature Reviews Microbiology 5: 316-323.

Van de Peer Y, Maere S, Meyer A (2009). The evolutionary significance of ancient genome duplications. Nature Reviews Genetics 10: 725-732.

van Wolferen M et al (2015). The archaeal Ced system imports DNA. Proceedings of the National Academy of Sciences of USA 113: 2496-2501.

Villanueva L, Schouten S, Sinninghe Damsté JS (2016). Phylogenomic analysis of lipid biosynthetic genes of Archaea shed light on the 'lipid divide'. Environmental Microbiology 19: 54-69.

Vosseberg J et al (2020). Timing the origin of eukaryotic cellular complexity with ancient duplications. Nature Ecology & Evolution. https://doi.org/10.1038/s41559-020-01320-z

Yutin N et al (2012). Phylogenomics of prokaryotic ribosomal proteins. PLoS ONE 7: e36972. Doi:10.1371/ journal.pone.0036972

Wickstead B, Gull K (2011). The evolution of the cytoskeleton. Journal of Cell Biology 194: 513-525.

Williams TA et al (2020). Phylogenomics provides robust support for a two-Domain tree of life. Nature Ecology & Evolution 4: 138-147.

Whittaker RH (1969). New concepts of kingdoms of organisms. Science 163: 150-160.

Woese CR, Fox GE (1976). The phylogenetic structure of the prokaryotic Domains: the primary kingdoms. Proceedings of the National Academy of Sciences 74: 5088-5090.

Woese CR, Kandler O, Wheelis ML (1990). Towards a natural system of organisms: proposal for the Domains Archaea, Bacteria, and Eucarya. Proceedings of the National Academy of Sciences USA 87: 4576-4579.

Bulletin of Regional Natural History (BORNH) ISSN 2724-4393.

Zaremba-Niedzwiedzka K et al (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature 541: 353-358.

Zhang S et al (2016) Sufficient oxygen for animal respiration 1,400 million years ago. Proceedings of the National Academy of Sciences USA 113: 1731-1736.

Zhu S et al (2015). Decimetre-scale multicellular eukaryotes from the 1.56-billion-year-old Gaoyuzhuang Formation in North China. Nature Communications 7:11500. Doi: 10.1038/ncomms11500.