æ

3. THE MAJOR LINES OF DESCENT

12 January 2023

Once the foundations for cellular life were established, various paths towards diversification were set in motion. Populations that are physically isolated from each other for sufficiently long periods naturally accumulate independent mutations, in some cases promoted by natural selection for optimal phenotypes in local environments, and in others simply by random genetic drift. When the genomes of isolated lineages diverge to a sufficient degree, the internally coadapted gene complexes will be mutually incompatible, preventing the production of viable downstream hybrids. Such genetic isolation, which constitutes the speciation process, ensures the survival and evolution of independent lineages on their own merits, as reflected in the millions of species now inhabiting the planet.

Understanding the genealogical relationships (phylogeny) of existing lineages is critical to biology, as it provides a historical overview of what evolution has been able to accomplish, thereby facilitating the development of hypotheses for how evolution occurs. When related species share the same trait, we can often be fairly certain that their common ancestor (at the basal node of the clade) also carried the trait. Because sister taxa evolve from common ancestors, their differences also provide insight into the kinds of changes that are possible from a shared beginning. For example, gains and losses of traits can be inferred when single lineages deviate from their surrounding relatives. Information on many pairs of taxa can then begin to reveal commonalities among traits and parallel paths of evolution.

Analyses like these require well-resolved phylogenies, built from observations on traits other than those under investigation to avoid being circular. The modern age of whole-genome sequencing has brought us to the limits of such information. Nevertheless, despite the millions of informative nucleotide sites now known in thousands of species, many of the earliest branching patterns in the Tree of Life remain ambiguous. By outlining what we do know about the relationships among the major lineages of life, this chapter sets the table for more in-depth comparative analyses in subsequent chapters.

First, we will examine the degree of phylogenetic affinity between the two major organizational grades of cellular life, the prokaryotes and eukaryotes. Although the two groups are generally distinguished by the absence/presence of a nuclear envelope and other membrane-bound organelles, this morphological distinction turns out to be misleading with respect to genealogical relationships. Not only are there two distantly related groups of prokaryotes, the bacteria and the archaea, but the eukaryotic lineage emerged from the latter, rather than being genealogically distinct from both.

Second, the lineage with the greatest morphological diversification, the eukaryotes, will be considered briefly. The main point here is that a very large suite of intracellular embellishments became established prior to the divergence of the major eukaryotic lineages, leaving no intermediate-state traces (at least as so far discovered). The establishment of the last eukaryotic common ancestor (LECA) was then followed fairly quickly by the emergence of major subclades with their own unique features.

Third, analogs of many of the shared (so-called universal) eukaryotic traits can be found in one or more prokaryotic lineages. Although these are not always orthologous in origin, this does indicate that their emergence was not strictly dependent on eukaryogenesis. However, the stem eukaryote distinguished itself in assembling a unique mixture of features into a single lineage. The conditions that might have driven the subsequent big bang of eukaryotic diversity remain unclear, and may not have been ecological in nature. Rather, the radical shift in the genetic system of the eukaryotic cell may have been the primary enabler of species diversification.

The Primary Domains of Life

Just as pairs of individuals within a population are related to various degrees in a pedigree sense, species relationships can be described in the form of a phylogenetic tree. Sibling species reside on adjacent branches separated by a single node (branch point), with pairs of species with lower affinities residing on more distant branches. Historically, the field of taxonomy sought to classify organisms by their physical appearances, but owing to the possibility of convergent phenotypic evolution, such an approach is fraught with interpretative problems. Information at the nucleotide level has an explicit genetic interpretation and is less ambiguous. Thus, almost all attempts to infer phylogenetic relationships are now based on observations on DNA-sequence divergence among extant species (Felsenstein 2004).

Although this is a highly technical field involving computationally demanding algorithms to obtain genealogical relationships that are most compatible with the data, the conceptual basis for these analyses is straight-forward. Because DNA naturally acquires nucleotide substitutions and rearrangements by mutation, some of which are nearly neutral (Chapter 4), related species experience DNA-sequence divergence over evolutionary time. The simple fact that species with higher levels of sequence similarity tend to be more closely related forms the logical basis for virtually all statistical methods for estimating phylogenetic trees and dating evolutionary events.

With the advent of molecular-genetic methods, insights into the broad form of the Tree of Life began to emerge in the 1970s. Up to this point, based on the obvious morphological void between prokaryotes and eukaryotes, the former had been viewed as one large, monophyletic group, ill-defined internally but assumed to be deeply separated from the eukaryotes (Sapp 2005). However, noting that the genomes of all organisms encode for ribosomal RNAs (which comprise the catalytic hearts of ribosomes), Woese and Fox (1977) reasoned that a higher degree of resolution could be obtained by comparative analysis of such sequences. They quickly discovered a deep phylogenetic furrow within the prokaryotes, implying the existence of two major lineages, seemingly as distinct from each other as they are from eukaryotes. These two prokaryotic groups came to be known as the archaea (often called archaebacteria) and the bacteria (sometimes called eubacteria) (Woese et al. 1990).

This division of life into three major groups raised several questions about the base of the Tree of Life. Are the bacterial, archaeal, and eukaryotic groups all monophyletic, each with independent common ancestors (a three-domains of life model), or are one or more clades embedded within another (a one- or two-domains model)? Assuming the groups are monophyletic, are eukaryotes more closely related to archaea or bacteria, or do they have affinities with both? Can the possibility that eukaryotes are ancestral to prokaryotes be formally ruled out?

The key to answering these questions is a correctly rooted phylogenetic tree denoting the location of the most recent common ancestor from which all species in the tree ultimately descend. This hypothetical taxon is often referred to as LUCA (for Last Universal Common Ancestor) (Figure 3.1), whereas the last common ancestors for bacteria, archaea, and eukaryotes are designated LBCA, LACA, and LECA. The first common ancestor for a lineage (e.g., FUCA, FBCA, FACA, and FECA) denotes the most remote point on the branch leading to the last common ancestor not containing any other major clade. Traits that are shared by all members of a clade were almost certainly present in the last common ancestor of the clade, but one cannot rule out an earlier origin on the branch extending from the first common ancestor. For example, a feature shared by all bacteria may have arisen anywhere along the FBCA-LBCA branch.

Two problems conspire to make the ascertainment of the relationships between the three major groups a difficult enterprise. First, although placing a root on a phylogeny is usually a simple matter of including in the analysis a compelling outgroup (i.e., a bird for a mammalian phylogeny), this is not an option when the entire Tree of Life is being considered. Second, the amount of molecular divergence among the few hundred genes shared by all three ancient groups is so vast that the signal of genealogical relationships has been greatly diluted by the accumulation of multiple nucleotide substitutions per site.

This being said, a consensus seems to have emerged on the deepest branches of the Tree of Life. Virtually all analyses indicate that the bacterial lineage is monophyletic and separate from the lineage containing the archaea and eukaryotes (e.g., Raymann et al. 2015; Coleman et al. 2021). Phylogenetic analysis implies that LBCA was a sophisticated cell, with a cell wall sandwiched between two membranes, a capacity for flagellar swimming and chemotaxis, and a CRISPR-Cas system for warding off invasive DNA elements. As the archaeal lineage also appears to be monophyletic (Williams et al. 2017), this leaves the positioning of eukaryotes as the main issue. In principle, eukaryotes could simply join as a separate monophyletic clade at a single node outside of bacteria and archaea. Alternatively, one of the groups might emerge as a sublineage within the other. The first pattern would be consistent with the three-domain model postulated by Woese and colleagues. The second condition would imply a two-domain scenario in which eukaryotes are simply a derived lineage within one of the prokaryotic groups or vice versa.

Resolving this issue has been challenging, owing to complications beyond the statistical problems outlined above. Most notable are: 1) the occurrence of substantial horizontal gene transfer among lineages early in the history of life (Doolittle et al. 2003); and 2) the additional massive transfer of bacterial genes to their eukaryotic host cells following the endosymbiotic establishment of the mitochondrion (derived from a bacterium) in the basal eukaryote (Chapter 23). Gene relocations blur the deep branches on the Tree of Life, as different genes have different phylogenetic histories. Nonetheless, most large-scale analyses now seem to support the eocyte hypothesis of Lake et al. (1984), which postulates eukaryotes as being most closely related to one particular archaeal group (Cox et al. 2008; Guy and Ettema 2011; Kelly et al. 2011; Thiergart et al. 2012; Williams et al. 2012, 2013, 2020; Raymann et al. 2015). This hypothesis essentially eliminates the possibility that eukaryotes are the primordial cellular lineage, rejects the three-domains view, and implicates a member of the archaea as the ultimate source of the eukaryotic nuclear genome.

It has been argued that the closest living relatives to eukaryotes reside within a lineage called the Asgard archaea, known mainly from the sequencing of environmental samples from deep-sea sediments (Spang et al. 2015; Hug et al. 2016; Zaremba-Niedzwiedzka et al. 2017; Tahon et al. 2021), although greater phylogenetic affinity with an alternative archaeal lineage cannot be ruled out entirely (Liu et al. 2021). The genome contents for members of the Asgard archaea imply the presence of actin- and tubulin-related (cytoskeletal) proteins as well as components associated with vesicle trafficking and membrane remodeling, all of which are classical attributes of eukaryotic cells (Ettema et al. 2011; Yutin and Koonin 2012; Akil and Robinson 2018; Liu et al. 2021). The only member of the Asgard archaea that is cultivatable in the lab so far produces long protuberances and seems to depend on a symbiotic relationship with another member of the archaea, but does not exhibit complex internal cell structure (Imachi et al. 2020).

Notably, eukaryotic proteins involved in information processing (e.g., transcription and translation) tend to be more similar to those in archaea than bacteria, as expected if the nuclear genome is derived from a member of the archaea. In contrast, proteins involved in house-keeping functions (e.g., metabolism) tend to most closely resemble those in bacteria (Brown and Doolittle 1997; Rivera et al. 1998; Leipe et al. 1999; Brown et al. 2001; Horiike et al. 2001), many of which may be derived from the colonizing bacterium that became the mitochondrion.

One concern with the two-domains hypothesis relates to the types of phospholipids deployed in the cell membranes of the different major lineages (Figure 3.2). All cells are enveloped by phospholipid bilayers, with the individual molecules comprised of a glycerol-phosphate sandwiched between a head group and two hydrocarbon chains (Chapter 15). However, whereas glycerol-1-phosphate (G1P) is bound to methyl-branched isoprenoid chains by ether linkages in archaea, glycerol-3-phosphate (G3P) is bound to straight fatty-acid chains by ester linkages in bacteria and eukaryotes (Boucher et al. 2004; Peretó et al. 2004). Likewise, the ability to produce membrane steroids appears to be restricted to bacteria and eukaryotes, and absent from archaea (Hoshino and Gaucher 2021). This affiliation of membrane composition in bacteria and eukaryotes is clearly inconsistent with the topology of the Tree of Life suggested above, unless LACA and its early descendants had membranes containing a mixture of both types of lipids (Lombard et al. 2012).

The latter idea has some support. The dehydrogenase enzymes that make G1P and G3P are found in all major lineages, raising the possibility of a nonspecific glycerol-phosphate dehydrogenase in LUCA. In addition, some bacteria and eukaryotes have phospholipids with ether linkers; some archaea have fatty acids; and isoprenoids are universally distributed, although they are synthesized by different pathways in the three major groups (Lange et al. 2000; Lombard and Moreira 2011; Villanueva et al. 2021). It has been argued that a mixed population of lipid molecules will reduce membrane stability, but there are doubts about this idea (Shimada and Yamagishi 2011), and indeed, *E. coli* has been engineered to contain up to 30% archaeal lipids with little negative effects on growth rate (Caforio et al. 2018). Thus, it is plausible that LUCA had a membrane consisting of a mixture of the molecules found in modern-day prokaryotic lineages, with alternative mechanisms for catalyzing pure populations of G1P or G3P molecules evolving independently in isolated lineages (Koga et al. 1998; Martin and Russell 2003; Wächtershäuser 2003).

Finally, it should be noted that the two-domains model makes the implicit assumption that the root of the entire Tree of Life falls between the bacterial and archaeal domains. A more formal way of evaluating the problem uses genes that duplicated prior to the divergence of the main domains, as each member of such a gene pair can serve to root the phylogeny of the other. In the ideal scenario, both trees resulting from such reciprocal rooting would yield the same topology. To exploit this strategy, Gogarten et al. (1989) used anciently duplicated subunits of ATP synthase (Foundations 2.1) to show that archaea and eukaryotes consistently group together to the exclusion of bacteria. The same result has been obtained with several other pairs of ancient duplicate genes (Iwabe et al. 1989; Brown and Doolittle 1995; Baldauf et al. 1996; Lawson et al. 1996; Gribaldo and Cammarano 1998; Zhaxybayeva et al. 2005).

Although the emerging view is that the Tree of Life is rooted as illustrated in Figure 3.1, there is still some dissent on the matter (Philippe and Forterre 1999; Cavalier-Smith and Chao 2020). Devos (2021) goes so far as to advocate a one-domain model in which bacteria are ancestral to all of life, with one bacterial lineage (putatively related to the planctomycetes, which have internal membranes and in some cases are capable of phagocytosis) spawning a common ancestor to archaea and eukaryotes (LAECA). Under this hypothesis, eukaryotes and archaea are each monophyletic sister taxa derived from a common ancestor, with the former embarking down a pathway of increasing complexity and the latter evolving to greater simplicity. Although this one-domain model does not enjoy phylogenetic support based on gene sequence data, given the limited power to confidently reveal relationships at the base of the Tree of Life, there are reasons to still be cautious in embracing the two-domain model as an established fact.

Times of Origin

The preceding description of the basic topology of the main trunks of the Tree of Life leaves unresolved the times of origin of various lineages, i.e., the temporal positions of the first and last common ancestors of the key clades. The gold standard for such estimates is a fossil record. However, only a small fraction of species leave such traces, and even in the best of circumstances, the vagaries of geological activity generally result in substantial gaps and uncertain time horizons in the fossil record. Although there is a well-established fossil record for many groups of land plants and animals, few unicellular organisms are fossilizable, and a wide range of abiotic events can leave traces that can be nearly indistinguishable from those induced by real cells (Javaux 2019). Today's smallest bacteria have diameters $< 1\,\mu$m, and the earliest cells were likely even smaller, further reducing the likelihood of detection. Rock formations older than 3.5 BY (billion years) are extremely rare, further restricting the opportunities of directly inferring the earliest stages of evolution.

The development of methods for detecting organic material in ancient rocks expands the potential for inferring life's presence (Brasier et al. 2015), although as noted in Chapter 2, numerous geological mechanisms can yield organic molecules in the absence of any biology. Given the inference that the complex processes of photosynthesis and methanogenesis were present by 3.4 BYA (billions of years ago) (Ueno et al. 2006; Javaux 2019), this further implies the establishment by this time of many of the metabolic/molecular processes from which all subsequent cellular lineages were built. Thus, it is not far-fetched to suggest that cells were present as early as 4.0 BYA, and some indirect evidence for biological activity as early as 4.1 BYA has been suggested (Bell et al. 2015).

The first evidence of eukaryotic cells appears in shale deposits containing putative molecular biomarkers of membrane components from ~ 2.7 BYA (Brocks et al. 1999), with the first presumptive algal fossils dating to ~ 2.1 BYA (Han and Runnegar 1992). Many other fossils of unicellular eukaryotes with complex surface ornamentations date to 1.5 to 1.7 BYA (Shixing and Huineng 1995; Javaux et al. 2001; Knoll 2004). However, complex multicellularity remained absent for at least another billion years. A dramatic shift occurred ~ 550 MYA (million years ago), when all of the major groups of multicellular animals appear suddenly in the fossil record in what is popularly known as the Cambrian Explosion (keeping in mind that "sudden" from a paleontological perspective can exceed 100 MY). The most visible biota on today's Earth, the jawed vertebrates and land plants, emerged only ~ 440 and ~ 400 MYA, respectively.

Of course, the time of first appearance of a group in the fossil record must postdate the actual time of origin. To work around this problem, attempts have been made to estimate key early divergence points in the Tree of Life using molecular clocks for protein-coding sequences calibrated with more recent fossils from well-understood taxonomic groups. Although numerous assumptions underlie these analyses, the current prognosis is an initial point of divergence of the eukaryotic branch from its archaeal ancestor ~ 1.9 BYA, demarcating the position of FECA (the First Eukaryotic Common Ancestor) , with LECA (the Last Eukaryotic Common Ancestor, at the base of the tree of diverging eukaryotic lineages) dating to ~ 1.0 to 1.7 BYA (Parfrey et al. 2011; Shih and Matzke 2013; Eme et al. 2014). These dates are roughly compatible with the fossil-record data noted above.

If this interpretation is correct, the first two billion years or so of life's timeline was written entirely by prokaryotes, with $> 80\%$ of biological history involving a world containing only single-celled organisms. Significant surprises may still be in store, as genome sequences from environmental samples continue to reveal new microbial lineages (Hug et al. 2016; Zaremba-Niedzwiedzka et al. 2017; Liu et al. 2021; Tahon et al. 2021).

The Emergence of Eukaryotes

Evolutionary cell biology is equally concerned with prokaryotes and eukaryotes. However, given the disproportionate attention give to yeast, plant, and animal cells, which represent only a sliver of the massive expansion of morphological complexity in eukaryotes, a brief excursion on the unity and diversity of the main lineages of the latter group is warranted.

Based on evidence from comparative genomics, there is little question that FECA was a chimera between members of the archaea and bacteria, but how such a liaison came about is less clear. As discussed above, key features of the nuclear genome (genes involved in replication and translation, in particular) were derived from an archaeum, but this leaves open a number of possible scenarios (López-García and Moreira 2020; Martin et al. 2015). In one view, FECA was an archaeal cell that acquired bacterial genes from an endosymbiotic bacterium, which became the mitochondrion. Although the latter point is well-established (Chapter 23), an alternative view is that the original host cell was a bacterium (which became the source of internal membranes) harboring an endosymbiotic archaeum (which became the nucleus), with the mitchondrion joining secondarily (López-García and Moreira 1999, 2020). Whether these alternatives, or any other suggested models, can ever be definitively resolved remains unclear, but numerous aspects of the constitution of LECA are more certain.

The stem eukaryote. Provided that a group of species is monophyletic, as seems to be the case for eukaryotes, we can generally be confident that any feature that is shared across all members of the clade must have been present in its most recent common ancestor (in this case LECA). Based on the logic that highly complex cellular traits are unlikely to have arisen independently in multiple lineages, comparative biology tells us that LECA was a flagellated heterotroph, capable of phagocytosis, with quite complex internal structure, and distinguished from prokaryotes in dozens of other ways at the level of cell structure, intracellular processes, gene structure, and genome organization (Cavalier-Smith 2009; Koumandou et al. 2013). The order in which these features emerged on the path from FECA to LECA are less clear, and will likely remain so unless basal lineages lacking subsets of such traits are discovered. This raises significant challenges for determining the key innovations that might have precipitated the evolutionary cascade of events known as eukaryogenesis. The following provides just a brief overview of the primary changes, with fuller details appearing in subsequent chapters.

The most celebrated eukaryotic attributes are physical ones. Most notably, a nuclear envelope allows a spatial separation between gene transcription within the nucleus and translation of messenger RNAs in the cytoplasm. Unique cytoskeletal structures based on actin and tubulin provide physical support for a variety of cellular functions. These include: platforms for membrane bending essential for vesicle formation, food engulfment by phagocytosis, and osmotic regulation by contractile vacuoles; scaffolds for the ordered transport of chromosomes during cell division; and highways for molecular motors engaged in transporting vesicles and powering flagella. Molecular motors are eukaryotic inventions, and the eukaryotic flagellum is completely different from that deployed in bacteria. Finally, internal membrane-based structures such as the endoplasmic reticulum and the golgi provide sites for molecular processing unique to eukaryotes.

A key eukaryotic organelle is the mitochondrion, which became established at some point between FECA and LECA and is one of the only eukaryotic features whose origin is known. Unlike other organelles, mitochondria contain genomes whose sequences reveal alphaproteobacterial ancestry, with the original colonist eventually becoming an obligate endosymbiont now known as the powerhouse of eukaryotic cells (Chapter 23). Prior to the establishment of the mitochondrion, ATP synthase (Chapter 2), resided on the cell membrane (as it does in all of today's prokaryotes), but in eukaryotes ATP synthase is sequestered to internal mitochondrial membranes. Some have argued that this relocation provided a solution to the reduced surface:volume ratio in larger cells, essentially generating a bioenergetics revolution necessary for the establishment of all other things eukaryotic (Lane 2002, 2015; Lane and Martin 2010). Under this view, colonization of the mitochondrion would have been the causal event in eukaryogenesis, and therefore the first key innovation to appear on the branch from FECA to LECA.

However, despite its superficial attractiveness, an association alone does not indicate the direction of causality, and the idea that the establishment of the mitochondrion spawned a quantum leap in bioenergetic capacity is inconsistent with numerous lines of evidence outlined in subsequent chapters. Nonetheless, once established the mitochondrion generated numerous secondary effects to accommodate its use. For example, substantial transfer of mitochondrial genes to the nuclear genome occurred prior to LECA. Many of these transferred genes generate products that must be sent back to the mitochondrion, in some cases providing components to protein complexes that also contain mitochondrially encoded subunits. This necessitates reliable mechanisms for coordinating the activities of organelle and nuclear genomes and targeting the transport of proteins to their appropriate destinations.

The transition to eukaryotes was also accompanied by major alterations in the mode of genome replication and transmission ((Lynch 2007; Chapter 10). Almost all bacterial genomes consist of single circular chromosomes that replicate bidirectionally in two continuous streams from a single origin of replication, with the daughter genomes moving to opposite ends of the parental cell by fairly simple mechanisms. In contrast, the nuclear genomes of eukaryotes consist of multiple linear chromosomes, spooled around protein complexes called histones, with multiple origins of replication and ends capped by repetitive arrays of short motifs called telomeres.

Eukaryotic cell division requires an organized set of events, known as mitosis, by which multiple chromosomes duplicate simultaneously, with complete offspring sets then being dragged to opposite poles along a microtubule-based spindle apparatus. Moreover, eukaryotes have another specialized form of genome replication called meiosis, which has no counterpart in prokaryotes. During this process, homologous pairs of chromosomes (one haploid set from each parent) line up in parallel in a diploid cell, where they reciprocally exchange material by recombination, ultimately producing four haploid daughter cells (with single copies of each chromosome). The fusion of two such haploid cells reconstitutes the diploid form, completing the sexual life cycle.

The mode of transcript processing also underwent considerable modification in the stem eukaryote (Lynch 2007). Most, if not all, prokaryotic genomes contain operons (cassettes of cotranscribed and often functionally related genes). Such multigene transcripts constitute a significant challenge for the membrane-bound genomes of eukaryotes, as the entire units have to be either exported from the nucleus in their entirety or pre-processed into single-gene messages prior to export. The few known cases of eukaryotic operons (e.g., in nematodes and euglenoids) do, in fact, involve such processing, along with the *trans*-splicing of a small leader sequence to the front end of each individual transcript, a process that is unknown in prokaryotes.

Finally, the nuclear envelope provided a genomic environment that promoted the emergence of more complex gene structure, most notably the colonization of genes by intragenic spacers called introns. Because introns are transcribed along with their surrounding exons, this genes-in-pieces architecture imposes another significant challenge for information processing – introns must be precisely excised and exons spliced back together (*cis*-splicing) prior to the export of mature mRNAs through the nuclear pore to the cytoplasm. Splicing is carried out by a complex molecular machine unique to eukaryotes, the spliceosome, consisting of five small RNA subunits and more than 100 proteins. In striking contrast, nearly all prokaryotic genes consist of a single uninterrupted coding region, and in the very few instances where this is not the case, the introns are self-splicing.

These are just a few of the many features unique to the eukaryotic lineage, the main point being that an enormous remodeling of cell biology occurred on the lineage from FECA to LECA. Notably, however, parallels of many "eukaryotic-specific" attributes can be found in isolated prokaryotic lineages, so one need not invoke *de novo* invention. For example, as already noted, many proteins previously thought to be restricted to eukaryotes are now known to have orthologous relatives in the Asgard archaea. In addition, organelles of a wide variety of types bounded by lipid or protein membranes are known for several members of the bacteria and archaea (Grant et al. 2018; Greening and Lithgow 2020), the planctomycetes in particular, with at least one such lineage being capable of phagocytosis (the engulfment and digestion of other cells) (Boedeker et al. 2017; Shiratori et al. 2019). These types of observations, along with other indirect inferences (Pittis and Gabaldón 2016), clearly indicate that many of the embellishments of eukaryotic cells did not have to await the origin of the mitochondrion as an energy support system.

What remains unclear is how so many odd features of prokaryotic cells came to be colocalized in the same FECA-LECA lineage. Although one might argue that FECA was a highly polymorphic species, with different individuals harboring subsets of traits (O'Malley et al. 2019), it is difficult to conceive of individuals with different constellations of complex traits still being reproductively compatible. Any such exchange would have had to occur prior to the emergence of meiosis, which requires sequence homology between pairing chromosomes.

Thus, the early steps of eukaryogenesis remain a mystery. We do not know the events that triggered eukaryogenesis, nor do we know the extent to which the peculiar features that arose did so via the encouragement of natural selection. Some modifications, such as intron colonization, may have emerged in population settings that enabled mildly deleterious mutations to accumulate passively by mutation pressure alone. Once established, however, the vast set of changes bestowed upon LECA provided the substrate for the evolutionary explosion in cell architectural diversity that is the hallmark of eukaryotes.

The eukaryotic radiation. As with investigations of the prokaryote-eukaryote divide, progress on revealing phylogenetic relationships among the major eukaryotic groups has largely relied on comparative gene-sequence analysis. However, even with whole-genome analyses, a variety of issues still conspires to cloud our understanding of the phylogeny of eukaryotes. These include idiosyncratic changes in rates of evolution, divergent nucleotide compositions across lineages, possibilities of early horizontal gene transfer, gene duplications, and inadequate taxon sampling. Two things are agreed upon. First, the primary eukaryotic lineages are deeply branching in time, with the major groups upon which most biological research is performed (metazoans, fungi, and plants) constituting only a small fraction of eukaryotic phylogenetic diversity. Second, although these three favored sets of study organisms are sometimes viewed as members of a "crown group" of eukaryotes or "higher forms" of life, they do not even comprise a monophyletic lineage.

An attempt to summarize what is known about eukaryotic phylogeny is presented in Figure 3.3, with two caveats. First, this description is by no means complete, as it contains only the groups that will be encountered in the following chapters. Even if all of the major known groups of eukaryotes were included, the story would be an abstract at best, as agnostic searches for molecular sequences from environmental samples suggest that many novel lineages of microbial eukaryotes, never before visualized, reside in our midst (Dawson and Pace 2002). Second, the phylogenetic relationships of many of the main eukaryotic groups remain unresolved. Depending on the authors, between five and eight monophyletic supergroups are recognized (e.g., Baldauf et al. 2000; Richards and Cavalier-Smith 2005; He et al. 2014; Derelle et al. 2015; Katz and Grant 2015; Ren et al. 2016; Burki et al. 2020; Wideman et al. 2020), and these will likely change to some degree as further data emerge.

In one view, the vast majority of eukaryotes fall into two major morphological groups based on the ancestral number of flagella being one or two (Cavalier-Smith 1998). The first of these, the unikonts, are united by the general presence of cells with a single flagellum at some stage of the life cycle (Cavalier-Smith 1998; Steenkamp et al. 2006; Paps et al. 2013). The unikonts contain the opisthokont group, an assemblage of metazoans, choanoflagellates, and fungi (top of Figure 3.3), as well as the amoebozoan group, comprised of the lobose amoeba and the slime molds (Bapteste et al. 2002). Along with a few biflagellate lineages, the unikonts appear to be separated from the remaining supergroups (all of which are biflagellate, and referred to as bikonts) at the root of the eukaryotic tree (Derelle et al. 2015). This motivates the suggestion that LECA was a biflagellate.

The large bikont assemblage contains the remaining supergroups, whose interrelationships remain unresolved. One of these groups, the archaeplastida, encompasses the chloroplast-bearing green plants (including the green algae), red algae (rhodophytes), and glaucophyte algae. The excavate supergroup contains the euglenozoa, which unites the euglenoids (e.g., *Euglena*) with the parasitic kinetoplastids (e.g., the trypanosomes *Trypanosoma* and *Leishmania*), as well as several other groups of flagellates.

Another large supergroup is dubbed the SAR clade, based on its primary component lineages, the stramenopiles, alveolates, and rhizarians. The diverse stramenopile subclade contains the diatoms, brown algae, and oomycetes, whereas the alveolates (united by the presence of alveoli, a system of sacs underlying the cell surface) encompass the ciliates (e.g., *Paramecium* and *Tetrahymena*), the dinoflagellates (a group of aquatic flagellates), and the obligately parasitic apicomplexans (including the malarial parasite *Plasmodium*) (Fast et al. 2002). The rhizaria consist of cercozoans, foraminiferans, and radiolarians, most of which are amoeboid and produce external skeletons (Nikolaev et al. 2004).

Monophyly of the entire bikont group has drawn support from a unique fusion between two key genes (dihydrofolate reductase and thymidylate synthase), which are encoded separately in all unikonts and prokaryotes (Stechmann and Cavalier-Smith 2002). However, some exceptions have been found within the bikonts (Burki 2014), which might represent secondary reversions. In addition, the amitochondriate diplomonad (including *Giardia*) and trichomonad lineages, appear not to contain either gene and so cannot be assigned phylogenetic positions on this basis, although they may be members of the excavate supergroup.

A eukaryotic big bang? Given that the deep lines of descent between bacteria, archaea, and eukaryotes have been resolved with far less data, the inability to fully decipher the more recent relationships among the main lines of eukaryotes is unlikely to be a matter of a shortage of genomic material. The relatively short internal branches of the eukaryotic tree, which imply a rapid early radiation of such groups, is the major issue – shorter branches between related groups leads to lower discriminating power. Thus, the bushy form of the eukaryotic tree has inspired a "big-bang" hypothesis suggesting that most of the major lineages became established in a period of 10 to 100 million years (Philippe et al. 2000; Cavalier-Smith 2002; Koonin 2007). If this idea is correct, the arguments presented above, along with other molecular estimates of the age of LECA, would suggest a radiation set down in a window roughly between 1.7 and 2.0 BYA (Wang et al. 1999; Yoon et al. 2004; Parfrey et al. 2011; Eme et al. 2014).

What might have precipitated such an active phase of lineage isolation? Most attempts at explaining evolutionary radiations resort to ecological arguments, either invoking a dramatic change in the environment or the chance appearance of an evolutionary novelty allowing the exploitation of new ecological niches, e.g., predation as a new way of living (Knoll 2014). However, a species radiation requires more than ecological opportunity. There must also be genetic isolating mechanisms to keep lineages distinct. Ultimately, opportunities for speciation require that populations be isolated for long enough periods to allow the accumulation of sufficient mutational changes that viability and/or fertility will be compromised by parental-genome incompatibilities that arise within hybrids.

Post-reproductive isolating barriers can arise by many different mechanisms (Coyne and Orr 2004), but microchromosomal rearrangements in which genes relocate from one chromosome to another are of particular relevance to the early eukaryotic radiation, which experienced two novel forms of genomic upheaval. Consider first the primordial mitochondrion. Most prokaryotic genomes contain a few thousand genes, while mitochondrial genomes contain no more than a few dozen. Thus, it is clear that hundreds of organelle-to-nuclear gene transfers occurred early in the establishment of mitochondria, although many were probably simply lost (Chapter 23). Because mitochondrial genomes are haploid and generally inherited uniparentally, a relocation of an essential mitochondrial gene to the nuclear genome would create an imbalance in hybrid progeny resulting from a cross with any lineage having the ancestral (non-rearranged) type (Figure 3.4). Regardless of the direction of the cross, both types of diploid hybrids would be presence/absence heterozygotes for the nuclear gene. In addition, one would also harbor a mitochondrial genome devoid of the gene. As a consequence, half of the gametes produced by the latter individual would lack the gene entirely (and half of those produced by the other hybrid type would acquire a double dose of the gene).

Although a single genomic transfer of this sort does not produce complete reproductive isolation, just a few independent transfers have a powerful effect. Imagine an incipient pair of species experiencing n independent organelle-to-nuclear gene transfers in each lineage. Assuming independent assortment of the nuclear genes during meiosis, then the fraction of F_1 gametes entirely lacking in a functional gene at one or more loci is $1 - 0.5^n$, which is 0.969 for $n = 5$. Thus, when one considers the hundreds of organelle-to-nuclear gene transfers that may have occurred soon after the colonization of the primordial mitochondrion, and probably extended over several million years, such gene traffic would have played a significant role in the passive development of isolating barriers among the earliest eukaryotes. Note that such microchromosomal rearrangements only yield reproductive isolating barriers in species with multiple chromosomes and sexual reproduction, as both are necessary for the independent segregation of unlinked loci. As noted above, both speciation-facilitating features were among the novelties that emerged on the branch from FECA to LECA.

A second mechanism of gene relocation relevant to the eukaryotic radiation involves nuclear gene-duplication events, which can passively lead to rearrangements when the original copy is silenced and a descendant copy is preserved on a separate chromosome (Chapter 6). Such events are of interest here because, as discussed in subsequent chapters, there was a massive amount of gene duplication at the base of eukaryotes, possibly a result of one or two complete genome duplications (Chapter 24). Such activities left their imprint on a wide variety of cellular features, including mitosis and meiosis (Ramesh et al. 2005; Malik et al. 2007; Liu et al. 2015; Onesti and MacNeill 2013), the cytoskeleton (Goodson and Hawse 2002; Dutcher 2003; McKean et al. 2001) and the flagellum (van Dam et al. 2013), proteasomes (Bouzat et al. 2000) and chaperones (Fares and Wolfe 2003), the nuclear-pore complex (Alber et al. 2007), and other organelles (Hirst et al. 2011; Schledzewski et al. 1999; Mast et al. 2014).

Thus, the indirect consequences of two of the defining cytological attributes of the stem eukaryote, a genome-bearing mitochondrion and meiotic recombination, along with rampant duplication in the nuclear genome, may have played a central role in the passive and relatively rapid emergence of the basal eukaryotic lineages. Although ecological divergence need not have played any initiating role in such processes, the resultant reproductive isolation would have allowed such lineages to descend down independent evolutionary pathways driven by adaptation to local environmental settings.

Summary

- Although life has classically been divided into eukaryotes and prokaryotes, molecular analyses indicate that these are not meaningful phylogenetic labels. Instead, there appears to be two domains of life, with the two prokaryotic groups (bacteria and archaea) appearing on opposite sides of the root of the Tree of Life, and eukaryotes being the most recent newcomer, emerging from a member of the archaea.

- Prokaryotes were established on Earth ~ 4 billion years ago (BYA), with eukaryotes appearing ~ 3 BYA. Although many eukaryotic lineages may have coexisted during this early period, only one (called LECA) eventually gave rise to today's eukaryotes, forming the base of the tree of extant lineages ~ 2 BYA. Each of these time points has a level of uncertainty of a few hundred million years.

- From the standpoint of morphological diversification, the emergence of eukaryotes marked a dramatic phase in Earth's history. With dozens of eukaryote-specific changes having become established prior to LECA, this keystone species was extraordinarily unique in terms of cellular and genomic architecture. However, the order in which these features arose remains unknown, and many of them are difficult to explain with adaptive arguments.

- Once established, LECA gave rise to an explosive radiation of the major eukaryotic groups on a relatively short time scale. This rapid episode of lineage isolation may have had little to do with ecological factors, instead being an inevitable consequence of two pre-LECA genomic upheavals – the origin of the mitochondrion and a period of rampant nuclear gene duplication. Combined with the evolution of sex and independently segregating chromosomes, these changes would have led to the passive accumulation of microchromosomal rearrangements and reproductive isolation in ways that would have been inoperable in prior lineages of prokaryotes.

Literature Cited

Akil, C., and R. C. Robinson. 2018. Genomes of Asgard archaea encode profilins that regulate actin. Nature 562: 439-443.

Alber, F., S. Dokudovskaya, L. M. Veenhoff, W. Zhang, J. Kipper, D. Devos, A. Suprapto, O. Karni-Schmidt, R. , B. T. Chait, et al. 2007. The molecular architecture of the nuclear pore complex. Nature 450: 695-701.

Baldauf, S. L., J. D. Palmer, and W. F. Doolittle. 1996. The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. Proc. Natl. Acad. Sci. USA 93: 7749-7754.

Baldauf, S. L., A. J. Roger, I. Wenk-Siefert, and W. F. Doolittle. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 290: 972-977.

Bapteste, E., H. Brinkmann, J. A. Lee, D. V. Moore, C. W. Sensen, P. Gordon, L. Durufle, T. Gaasterland, P. Lopez, M. Muller, and H. Philippe. 2002. The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium, Entamoeba,* and *Mastigamoeba.* Proc. Natl. Acad. Sci. USA 99: 1414-1419.

Bell, E. A., P. Boehnke, T. M. Harrison, and W. L. Mao. 2015. Potentially biogenic carbon preserved in a 4.1 billion-year-old zircon. Proc. Natl. Acad. Sci. USA 112: 14518-14521.

Boedeker, C., M. Schüler, G. Reintjes, O. Jeske, M. C. van Teeseling, M. Jogler, P. Rast, D. Borchert, D. P. Devos, M. Kucklick, et al. 2017. Determining the bacterial cell biology of Planctomycetes. Nat. Commun. 8: 14853.

Boucher, Y., M. Kamekura, and W. F. Doolittle. 2004. Origins and evolution of isoprenoid lipid biosynthesis in archaea. Mol. Microbiol. 52: 515-527.

Bouzat, J. L., L. K. McNeil, H. M. Robertson, L. F. Solter, J. E. Nixon, J. E. Beever, H. R. Gaskins, G. Olsen, S. Subramaniam, M. L. Sogin, et al. 2000. Phylogenomic analysis of the alpha proteasome gene family from early diverging eukaryotes. J. Mol. Evol. 51: 532-543.

Brasier, M. D., J. Antcliffe, M. Saunders, and D. Wacey. 2015. Changing the picture of Earth's earliest fossils (3.5-1.9 Ga) with new approaches and new discoveries. Proc. Natl. Acad. Sci. USA 112: 4859-4864.

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

Brown, J. R., and W. F. Doolittle. 1995. Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. Proc. Natl. Acad. Sci. USA 92: 2441-2445.

Brown, J. R., and W. F. Doolittle. 1997. Archaea and the prokaryote-to-eukaryote transition. Microbiol. Mol. Biol. Rev. 61: 456-502.

Brown, J. R., C. J. Douady, M. J. Italia, W. E. Marshall, and M. J. Stanhope. 2001. Universal trees based on large combined protein sequence data sets. Nat. Genet. 28: 281-285.

Burki, F. 2014. The eukaryotic tree of life from a global phylogenomic perspective. Cold Spring Harb. Perspect. Biol. 6: a016147.

Burki, F., A. J. Roger, M. W. Brown, and A. G. B. Simpson. 2020. The new tree of eukaryotes. Trends Ecol. Evol. 35: 43-55.

Caforio, A., M. F. Siliakus, M. Exterkate, S. Jain, V. R. Jumde, R. L. H. Andringa, S. W. M.

Kengen, A. J. Minnaard, A. J. M. Driessen, and J. van der Oost 2018. Converting *Escherichia coli* into an archaebacterium with a hybrid heterochiral membrane. Proc. Natl. Acad. Sci. USA 115: 3704-3709.

Cavalier-Smith, T. 1998. A revised six-kingdom system of life. Biol. Rev. 73: 203-266.

Cavalier-Smith, T. 2002. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. Internat. J. Syst. Evol. Microbiol. 52: 297-354.

Cavalier-Smith, T. 2009. Predation and eukaryote cell origins: a coevolutionary perspective. Int. J. Biochem. Cell Biol. 41: 307-322.

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

Coleman, G. A., A. A. Davin, T. A. Mahendrarajah, L. L. Szánthó, A. Spang, P. Hugenholtz, G. J. Szöllosi, and T. A. Williams. 2021. A rooted phylogeny resolves early bacterial evolution. Science 372: eabe0511.

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

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Assoc., Inc., Sunderland, MA.

Dawson, S. C., and N. R. Pace. 2002. Novel kingdom-level eukaryotic diversity in anoxic environments. Proc. Natl. Acad. Sci. USA 99: 8324-8329.

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

Devos, D. P. 2021. Reconciling Asgardarchaeota phylogenetic proximity to eukaryotes and planctomycetes cellular features in the evolution of life. Mol. Bio. Evol. 38: 3531-3542.

Doolittle, W. F., Y. Boucher, C. L. Nesbo, C. J. Douady, J. O. Andersson, and A. J. Roger. 2003. How big is the iceberg of which organellar genes in nuclear genomes are but the tip? Phil. Trans. Roy. Soc. Lond. B Biol. Sci. 358: 39-57.

Dutcher, S. K. 2003. Long-lost relatives reappear: identification of new members of the tubulin superfamily. Curr. Opin. Microbiol. 6: 634-640.

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

Ettema, T. J., A. C. Lindås, and R. Bernander. 2011. An actin-based cytoskeleton in archaea. Mol. Microbiol. 80: 1052-1061.

Fares, M. A., and K. H. Wolfe. 2003. Positive selection and subfunctionalization of duplicated CCT chaperonin subunits. Mol. Biol. Evol. 20: 1588-1597.

Fast, N. M., L. Xue, S. Bingham, and P. J. Keeling. 2002. Re-examining alveolate evolution using multiple protein molecular phylogenies. J. Eukaryot. Microbiol. 49: 30-37.

Felsenstein, J. 2004. Inferring Phylogenies. Sinauer Assocs., Inc. Sunderland, MA.

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

Goodson, H. V., and W. F. Hawse. 2002. Molecular evolution of the actin family. J. Cell Sci. 115: 2619-2622.

Grant, C. R., J. Wan, and A. Komeili. 2018. Organelle formation in bacteria and archaea. Annu. Rev. Cell Dev. Biol. 34: 217-238.

Greening, C., and T. Lithgow. 2020. Formation and function of bacterial organelles. Nat. Rev. Microbiol. 18: 677-689.

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

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

Han, T. M., and B. Runnegar. 1992. Megascopic eukaryotic algae from the 2.1-billion-year-old Negaunee iron-formation, Michigan. Science 257: 232-235.

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

Hirst, J., L. D. Barlow, G. C. Francisco, D. A. Sahlender, M. N. Seaman, J. B. Dacks, and M. S. Robinson. 2011. The fifth adaptor protein complex. PLoS Biol. 9: e1001170.

Horiike, T., K. Hamada, S. Kanaya, and T. Shinozawa. 2001. Origin of eukaryotic cell nuclei by symbiosis of Archaea in Bacteria is revealed by homology-hit analysis. Nat. Cell Biol. 3: 210-214.

Hoshino, Y., and E. A. Gaucher. 2021. Evolution of bacterial steroid biosynthesis and its impact on eukaryogenesis. Proc. Natl. Acad. Sci. USA 118: e2101276118.

Hug, L. A., B. J. Baker, K. Anantharaman, C. T. Brown, A. J. Probst, C. J. Castelle, C. N. Butterfield, A. W. Hernsdorf, Y. Amano, K. Ise, et al. 2016. A new view of the tree of life. Nat. Microbiol. 1: 16048.

Imachi, H., M. K. Nobu, N. Nakahara, Y. Morono, M. Ogawara, Y. Takaki, Y. Takano, K. Uematsu, T. Ikuta, M. Ito, et al. 2020. Isolation of an archaeon at the prokaryote-eukaryote interface. Nature 577: 519-525.

Iwabe, N., K. Kuma, M. Hasegawa, S. Osawa, and T. Miyata. 1989. Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc. Natl. Acad. Sci. USA 86: 9355-9359.

Javaux, E. J. 2019. Challenges in evidencing the earliest traces of life. Nature 572: 451-460.

Javaux, E. J., A. H. Knoll, and M. R. Walter. 2001. Morphological and ecological complexity in early eukaryotic ecosystems. Nature 412: 66-69.

Katz, L. A., and J. R. Grant. 2015. Taxon-rich phylogenomic analyses resolve the eukaryotic tree of life and reveal the power of subsampling by sites. Syst. Biol. 64: 406-415.

Kelly, S., B. Wickstead, and K. Gull. 2011. Archaeal phylogenomics provides evidence in support of a methanogenic origin of the Archaea and a thaumarchaeal origin for the eukaryotes. Proc. Biol. Sci. 278: 1009-1018.

Knoll, A. H. 2004. Life on a Young Planet: the First Three Billion Years of Evolution on Earth.

Princeton University Press, Princeton, NJ.

Knoll, A. H. 2014. Paleobiological perspectives on early eukaryotic evolution. Cold Spring Harb. Perspect. Biol. 6: a016121.

Koga, Y., T. Kyuragi, M. Nishihara, and N. Sone. 1998. Did archaeal and bacterial cells arise independently from noncellular precursors? A hypothesis stating that the advent of membrane phospholipid with enantiomeric glycerophosphate backbones caused the separation of the two lines of descent. J. Mol. Evol. 46: 54-63.

Koonin, E. V. 2007. The biological big bang model for the major transitions in evolution. Biol. Direct 2: 21.

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

Lake, J. A., E. Henderson, M. Oakes, and M. W. Clark. 1984. Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. Proc. Natl. Acad. Sci. USA 81: 3786-3790.

Lane, N. 2002. Power, Sex, Suicide: Mitochondria and the Meaning of Life. Oxford Univ. Press, Oxford, UK.

Lane, N. 2015. The Vital Question. W. W. Norton & Co., Inc. New York, NY.

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

Lange, B. M., T. Rujan, W. Martin, and R. Croteau. 2000. Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. Proc. Natl. Acad. Sci. USA 97: 13172-13177.

Lawson, F. S., R. L. Charlebois, and J. A. Dillon. 1996. Phylogenetic analysis of carbamoylphosphate synthetase genes: complex evolutionary history includes an internal duplication within a gene which can root the tree of life. Mol. Biol. Evol. 13: 970-977.

Leipe, D. D., L. Aravind, and E. V. Koonin. 1999. Did DNA replication evolve twice independently? Nucleic Acids Res. 27: 3389-3401.

Liu, Y., K. S. Makarova, W. C. Huang, Y. I. Wolf, A. N. Nikolskaya, X. Zhang, M. Cai, C. J. Zhang, W. Xu, Z. Luo, et al. 2021. Expanded diversity of Asgard archaea and their relationships with eukaryotes. Nature 593: 553-557.

Liu, Y., J. Pei, N. Grishin, and W. J. Snell. 2015. The cytoplasmic domain of the gamete membrane fusion protein HAP2 targets the protein to the fusion site in *Chlamydomonas* and regulates the fusion reaction. Development 142: 962-971.

Lombard, J., P. López-García, and D. Moreira. 2012. The early evolution of lipid membranes and the three domains of life. Nat. Rev. Microbiol. 10: 507-515.

Lombard, J., and D. Moreira. 2011. Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. Mol. Biol. Evol. 28: 87-99.

López-García, P., and D. Moreira. 1999. Metabolic symbiosis at the origin of eukaryotes. Trends Biochem. Sci. 24: 88-93.

López-García, P., and D. Moreira. 2020. The Syntrophy hypothesis for the origin of eukaryotes revisited. Nat. Microbiol. 5: 655-667.

Lynch, M. 2007. The Origins of Genome Architecture. Sinauer Assocs., Inc., Sunderland, MA.

Lynch, M., and A. Force. 2000. Gene duplication and the origin of interspecific genomic incompatibility. Amer. Natur. 156: 590-605.

Malik, S. B., M. A. Ramesh, A. M. Hulstrand, and J. M. Logsdon, Jr. 2007. Protist homologs of the meiotic Spo11 gene and topoisomerase VI reveal an evolutionary history of gene duplication and lineage-specific loss. Mol. Biol. Evol. 24: 2827-2841.

Martin, W. F., S. Garg, and V. Zimorski. 2015. Endosymbiotic theories for eukaryote origin. Philos. Trans. R. Soc. Lond. B Biol. Sci. 370: 20140330.

Martin, W., and M. J. Russell. 2003. On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. Philos. Trans. R. Soc. Lond. B Biol. Sci. 358: 59-83.

Mast, F. D., L. D. Barlow, R. A. Rachubinski, and J. B. Dacks. 2014. Evolutionary mechanisms for establishing eukaryotic cellular complexity. Trends Cell Biol. 24: 435-442.

McKean, P. G., S. Vaughan, and K. Gull. 2001. The extended tubulin superfamily. J. Cell Sci. 114: 2723-2733.

Nikolaev, S. I., C. Berney, J. F. Fahrni, I. Bolivar, S. Polet, V. V. Aleshin, and N. B. Petrov. 2004. Molecular phylogenetic analysis places *Percolomonas cosmopolitus* within Heterolobosea: evolutionary implications. J. Eukaryot. Microbiol. 51: 575-581.

O'Malley, M. A., M. M. Leger, J. G. Wideman, and I. Ruiz-Trillo. 2019. Concepts of the last eukaryotic common ancestor. Nat. Ecol. Evol. 3: 338-344.

Onesti, S., and S. A. MacNeill. 2013. Structure and evolutionary origins of the CMG complex. Chromosoma 122: 47-53.

Paps, J., L. A. Medina-Chacón, W. Marshall, H. Suga, and I. Ruiz-Trillo. 2013. Molecular phylogeny of unikonts: new insights into the position of apusomonads and ancyromonads and the internal relationships of opisthokonts. Protist 164: 2-12.

Parfrey, L. W., D. J. Lahr, A. H. Knoll, and L. A. Katz. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc. Natl. Acad. Sci. USA 108: 13624-13629.

Peretó, J., P. López-Garcia, and D. Moreira. 2004. Ancestral lipid biosynthesis and early membrane evolution. Trends Biochem. Sci. 29: 469-477.

Philippe, H., and P. Forterre. 1999. The rooting of the universal tree of life is not reliable. J. Mol. Evol. 49: 509-523.

Philippe, H., P. Lopez, H. Brinkmann, K. Budin, A. Germot, J. Laurent, D. Moreira, M. Müller, and H. Le Guyader. 2000. Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. Proc. Roy. Soc. Lond. B 267: 1213-1221.

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

Ramesh, M. A., S. B. Malik, J. M. Logsdon, Jr. 2005. A phylogenomic inventory of meiotic genes: evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. Curr. Biol. 15: 185-191.

Raymann, K., C. Brochier-Armanet, and S. Gribaldo. 2015. The two-domain tree of life is linked to a new root for the Archaea. Proc. Natl. Acad. Sci. USA 112: 6670-6675.

Ren, R., Y. Sun, Y. Zhao, D. Geiser, H. Ma, and X. Zhou. 2016. Phylogenetic resolution of deep eukaryotic and fungal relationships using highly conserved low-copy nuclear genes. Genome Biol. Evol. 8: 2683-2701.

Richards, T. A., and T. Cavalier-Smith. 2005. Myosin domain evolution and the primary divergence of eukaryotes. Nature 436: 1113-1118.

Rivera, M. C., R. Jain, J. E. Moore, and J. A. Lake. 1998. Genomic evidence for two functionally distinct gene classes. Proc. Natl. Acad. Sci. USA 95: 6239-6244.

Sapp, J. 2005. The prokaryote-eukaryote dichotomy: meanings and mythology. Microbiol. Mol. Biol. Rev. 69: 292-305.

Schledzewski, K., H. Brinkmann, and R. R. Mendel. 1999. Phylogenetic analysis of components of the eukaryotic vesicle transport system reveals a common origin of adaptor protein complexes 1, 2, and 3 and the F subcomplex of the coatomer COPI. J. Mol. Evol. 48: 770-778.

Shih, P. M., and N. J. Matzke. 2013. Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins. Proc. Natl. Acad. Sci. USA 110: 12355-1260.

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

Shiratori, T., S. Suzuki, Y. Kakizawa, and K. I. Ishida. 2019. Phagocytosis-like cell engulfment by a planctomycete bacterium. Nat. Commun. 10: 5529.

Shixing, Z., and C. Huineng. 1995. Megascopic multicellular organisms form the 1700-million-year-old Tuanshanzi formation in the Jixian area, north China. Science 270: 620-622.

Spang, A., J. H. Saw, S. L. Jørgensen, K. Zaremba-Niedzwiedzka, J. Martijn, A. E. Lind, R. van Eijk, C. Schleper, L. Guy, amd T. J. G. Ettema. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature 521: 173-179.

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

Steenkamp, E. T., J. Wright, and S. L. Baldauf. 2006. The protistan origins of animals and fungi. Mol. Biol. Evol. 23: 93-106.

Tahon, G., P. Geesink, and T. J. G. Ettema. 2021. Expanding archaeal diversity and phylogeny: past, present, and future. Annu. Rev. Microbiol. 75: 359-381.

Thiergart, T., G. Landan, M. Schenk, T. Dagan, and W. F. Martin. 2012. An evolutionary network of genes present in the eukaryote common ancestor polls genomes on eukaryotic and mitochondrial origin. Genome Biol. Evol. 4: 466-485.

Ueno, Y., K. Yamada, N. Yoshida, S. Maruyama, and Y. Isozaki. 2006. Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. Nature 440: 516-519.

van Dam, T. J., M. J. Townsend, M. Turk, A. Schlessinger, A. Sali, M. C. Field, and M. A. Huynen. 2013. Evolution of modular intraflagellar transport from a coatomer-like progenitor. Proc. Natl. Acad. Sci. USA 110: 6943-6948.

Villanueva, L., F. A. B. von Meijenfeldt, A. B. Westbye, S. Yadav, E. C. Hopmans, B. E. Dutilh, and J. S. S. Damsté. 2021. Bridging the divide: bacteria synthesizing archaeal membrane lipids. ISME J. 15: 168-182.

Wächtershäuser, G. 2003. From pre-cells to Eukarya – a tale of two lipids. Mol. Microbiol. 47: 13-22.

Wang, D. Y., S. Kumar, and S. B. Hedges. 1999. Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. Proc. Roy. Soc. Lond. B 266: 163-171.

Wideman, J. G., A. Monier, R. Rodríguez-Martínez, G. Leonard, E. Cook, C. Poirier, F. Maguire, D. S. Milner, N. A. T. Irwin, et al. 2020. Unexpected mitochondrial genome diversity revealed by targeted single-cell genomics of heterotrophic flagellated protists. Nat. Microbiol. 5: 154-165.

Williams, T. A., C. J. Cox, P. G. Foster, G. J. Szöllösi, and T. M. Embley. 2020. Phylogenomics provides robust support for a two-domains tree of life. Nat. Ecol. Evol. 4: 138-147.

Williams, T. A., P. G. Foster, C. J. Cox, and T. M. Embley. 2013. An archaeal origin of eukaryotes supports only two primary domains of life. Nature 504: 231-236.

Williams, T. A., P. G. Foster, T. M. Nye, C. J. Cox, and T. M. Embley. 2012. A congruent phylogenomic signal places eukaryotes within the Archaea. Proc. Biol. Sci. 279: 4870-4879.

Williams, T. A., G. J. Szöllösi, A. Spang, P. G. Foster, S. E. Heaps, B. Boussau, T. J. G. Ettema, and T. M. Embley. 2017. Integrative modeling of gene and genome evolution roots the archaeal tree of life. Proc. Natl. Acad. Sci. USA 114: E4602-E4611.

Woese, C. R., and G. E. Fox. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. USA 74: 5088-5090.

Woese, C. R., O. Kandler, and M. L. Wheelis. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. USA 87: 4576-4579.

Yoon, H. S., J. D. Hackett, C. Ciniglia, G. Pinto, and D. Bhattacharya. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Mol. Biol. Evol. 21: 809-818.

Yutin, N., and E. V. Koonin. 2012. Archaeal origin of tubulin. Biol. Direct 7: 10.

Zaremba-Niedzwiedzka, K., E. F. Caceres, J. H. Saw, D. Bäckström, L. Juzokaite, E. Vancaester, K. W. Seitz, K. Anantharaman, P. Starnawski, K. U. Kjeldsen, et al. 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature 541: 353-358.

Zhaxybayeva, O., P. Lapierre, and J. P. Gogarten. 2005. Ancient gene duplications and the root(s) of the tree of life. Protoplasma 227: 53-64.

Figure 3.1. An idealized view of the two-domains view of the Tree of Life. The Last Universal Common Ancestor (LUCA) diverged into the bacterial and archaeal lineages, with eukaryotes then emerging out of the archaeal clade. Horizontal lines with blunt ends denote extinct lineages, and the relative temporal positions of lineage origins are not meant to be taken literally. The vertical blue line denotes the origin of the mitochondrion via endosymbiosis of a colonizing bacterium. Abbreviations used in acronyms for ancestors: F = first; L = last (or most recent); U = universal; B = bacterial; A = archaeal; E = eukaryote.

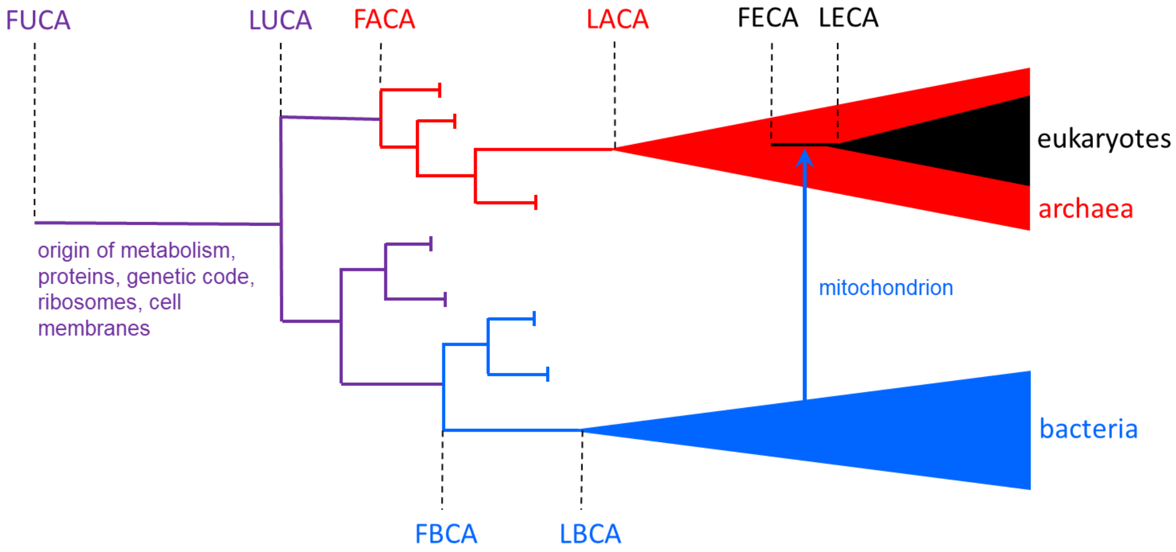

Figure 3.2. Alternative forms of phospholipids deployed by the three major lineages of life. The jagged lines represent chains of carbon atoms.

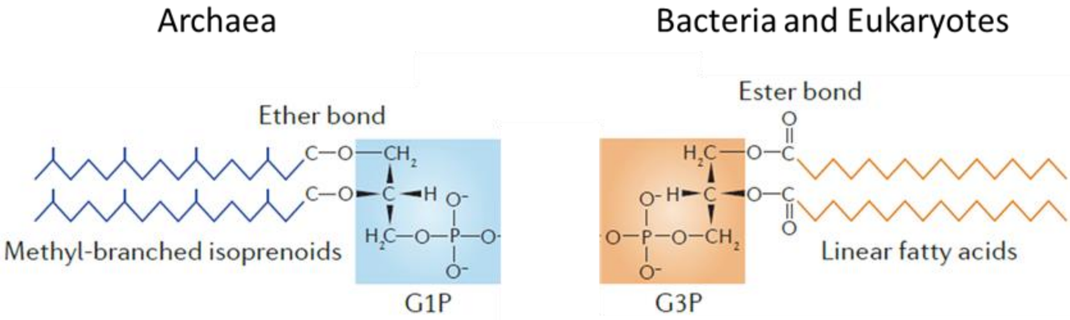

Figure 3.3. An approximate phylogenetic tree for some of the major eukaryotic "supergroups," generalized from the references in the text. The branch lengths are not proportional to time, although all external branches are expected to be in excess of 700 million years in length. Grey lines at the base of the tree denote areas of uncertainty.

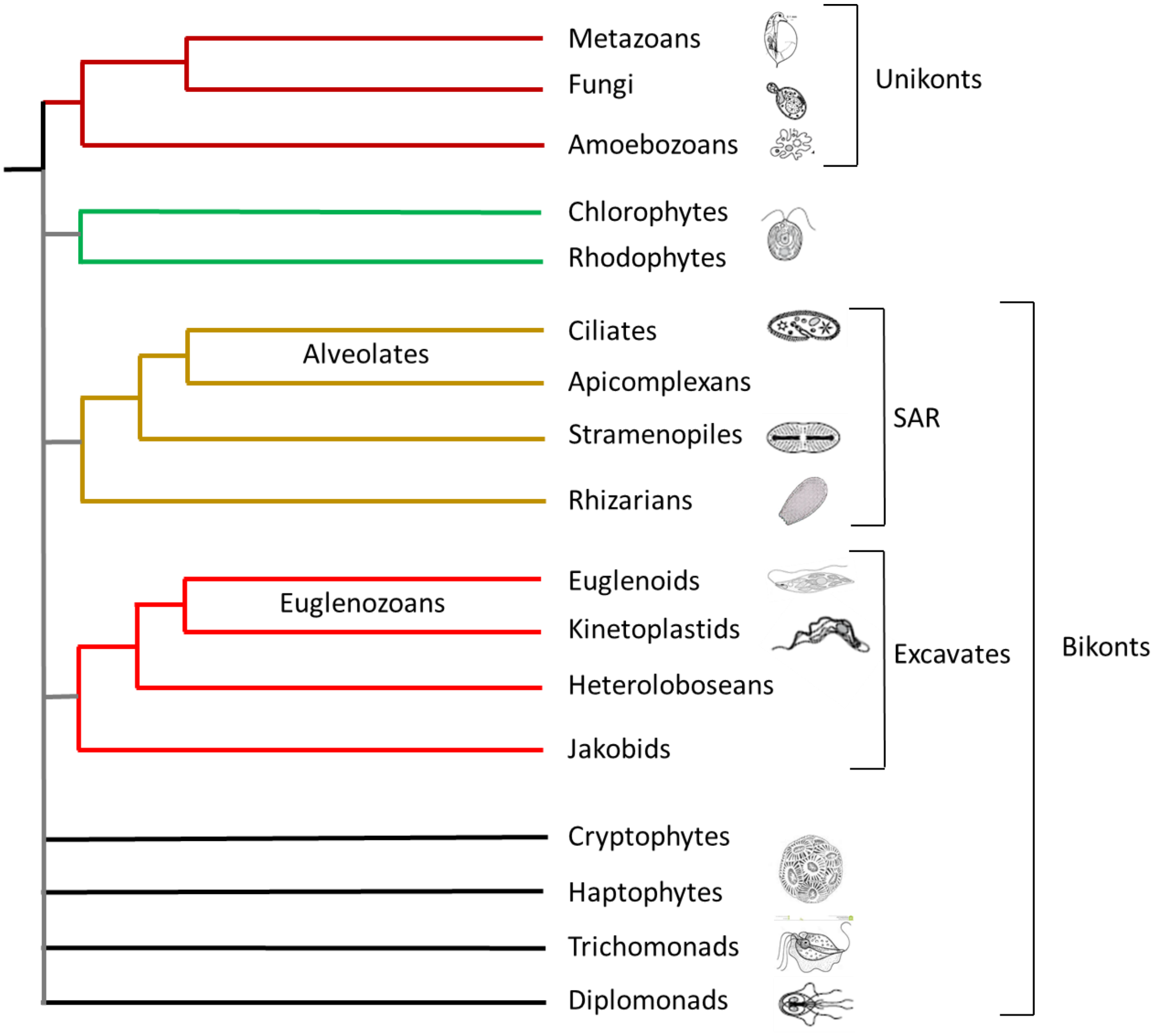

Figure 3.4. The development of reproductive incompatibility following the relocation of a mitochondrial gene. A diploid phase of nuclear genome is shown. Rectangles and circles denote autosomal and organelle gene copies respectively, with open symbols indicating gene absence. Following a geographic isolating event, the incipient species on the left experiences an organelle-gene transfer to the nucleus. Subsequent hybridization yields presence/absence heterozygotes at the autosomal locus, with the status of the uniparentally inherited mitochondrial genome depending on maternal identity. As a consequence of Mendelian segregation of the diploid autosomal locus gene imbalance arises following meiosis: half of the gametes of the individual in the lower left will lack the gene entirely, and half of those on the right will have a double dose of the gene. After Lynch and Force (2000).