ANNUAL REVIEWS

Annual Review of Microbiology

Origin and Early Evolution of the Eukaryotic Cell

Toni Gabaldón[1,2,3]

[1]Barcelona Supercomputing Centre (BCS-CNS), 08034 Barcelona, Spain; email: toni.gabaldon.bcn@gmail.com

[2]Institute for Research in Biomedicine (IRB), The Barcelona Institute of Science and Technology (BIST), 08028 Barcelona, Spain

[3]Catalan Institution for Research and Advanced Studies (ICREA), 08010 Barcelona, Spain

Annu. Rev. Microbiol. 2021. 75:631–47

First published as a Review in Advance on August 3, 2021

The *Annual Review of Microbiology* is online at micro.annualreviews.org

https://doi.org/10.1146/annurev-micro-090817-062213

Keywords

eukaryogenesis, mitochondria, endosymbiosis, eukaryotic evolution, LECA

Abstract

The origin of eukaryotes has been defined as the major evolutionary transition since the origin of life itself. Most hallmark traits of eukaryotes, such as their intricate intracellular organization, can be traced back to a putative common ancestor that predated the broad diversity of extant eukaryotes. However, little is known about the nature and relative order of events that occurred in the path from preexisting prokaryotes to this already sophisticated ancestor. The origin of mitochondria from the endosymbiosis of an alphaproteobacterium is one of the few robustly established events to which most hypotheses on the origin of eukaryotes are anchored, but the debate is still open regarding the time of this acquisition, the nature of the host, and the ecological and metabolic interactions between the symbiotic partners. After the acquisition of mitochondria, eukaryotes underwent a fast radiation into several major clades whose phylogenetic relationships have been largely elusive. Recent progress in the comparative analyses of a growing number of genomes is shedding light on the early events of eukaryotic evolution as well as on the root and branching patterns of the tree of eukaryotes. Here I discuss current knowledge and debates on the origin and early evolution of eukaryotes. I focus particularly on how phylogenomic analyses have challenged some of the early assumptions about eukaryotic evolution, including the widespread idea that mitochondrial symbiosis in an archaeal host was the earliest event in eukaryogenesis.

Contents

INTRODUCTION

The term eukaryote derives from Greek and means true nucleus. The nucleus is indeed one of the most defining features of eukaryotes, a synapomorphy of the group. This is not, however, the only trait setting eukaryotes apart from archaea and bacteria—two distinct groups collectively referred to as prokaryotes (but see 13). Beyond the nucleus, which harbors linear chromosomes and has characteristic membrane pores, there are many other features that are present in all or almost all eukaryotes and absent from all or nearly all prokaryotes. These include, among others, an intricate endomembrane system differentiated into subcompartments such as the endoplasmic reticulum and the Golgi apparatus, mitochondria, and a complex cytoskeleton able to perform endocytosis and phagocytosis. It was the realization of this sharp divide in cellular complexity that made Roger Stanier (68, p. 85) and his colleagues define this transition as the "greatest single evolutionary discontinuity to be found in the present day world." How eukaryotes may have arisen from simpler cellular forms remains a central question in evolutionary biology. It has been the focus of many studies, and not fewer heated debates. Historically the question has been approached from the perspectives of cellular biology and microbial diversity, which brought about fundamental ideas such as the potential of symbiotic processes to derive complex organisms from the stable merging of simpler ones (29, 59). Later, the advent of molecular phylogenetics provided a most needed tool to confirm or refute expectations from earlier proposed scenarios and a new type of empirical data to consider when proposing new ones. For instance, phylogenetic analysis of mitochondrial DNA confirmed the bacterial origin of these organelles, established the monophyletic nature of all mitochondria, and found that the most likely ancestor was a proteobacterium (30). More recently, the genomic revolution brought about by high-throughput sequencing has provided researchers with an unprecedented wealth of molecular data from virtually every extant organism within our reach. Indeed sequencing allows us not only to access the entire genetic complement of organisms that we can bring to the laboratory but also to reconstruct genomes from unculturable organisms, including those that can be studied only as an integral part of complex communities in barely accessible environments (26, 63, 74). Genomic information can be used to infer the features and capacities of organisms and their ancestors (23, 40). When, over the last few years, this large and fresh wave of genomic data was scrutinized through the evolutionary lens—i.e., by the use of phylogenomics (38)—new patterns emerged that have deeply shaken the field. The new data have

not only illuminated previously unexplored branches of the tree of life and provided higher-power resolution to our reconstruction of the tree of eukaryotes but have also enabled us to zoom in on the individual evolutionary histories of genes, tracing them back to their earliest origins. These incredible advances notwithstanding, the currently available empirical data are still fragmented and insufficient to unequivocally resolve many of open questions (**Figure 1**). Here, I provide an overview of recent findings and remaining questions, putting my focus on how genome sequencing and phylogenomic analysis have shaken some preconceptions that were once well-established.

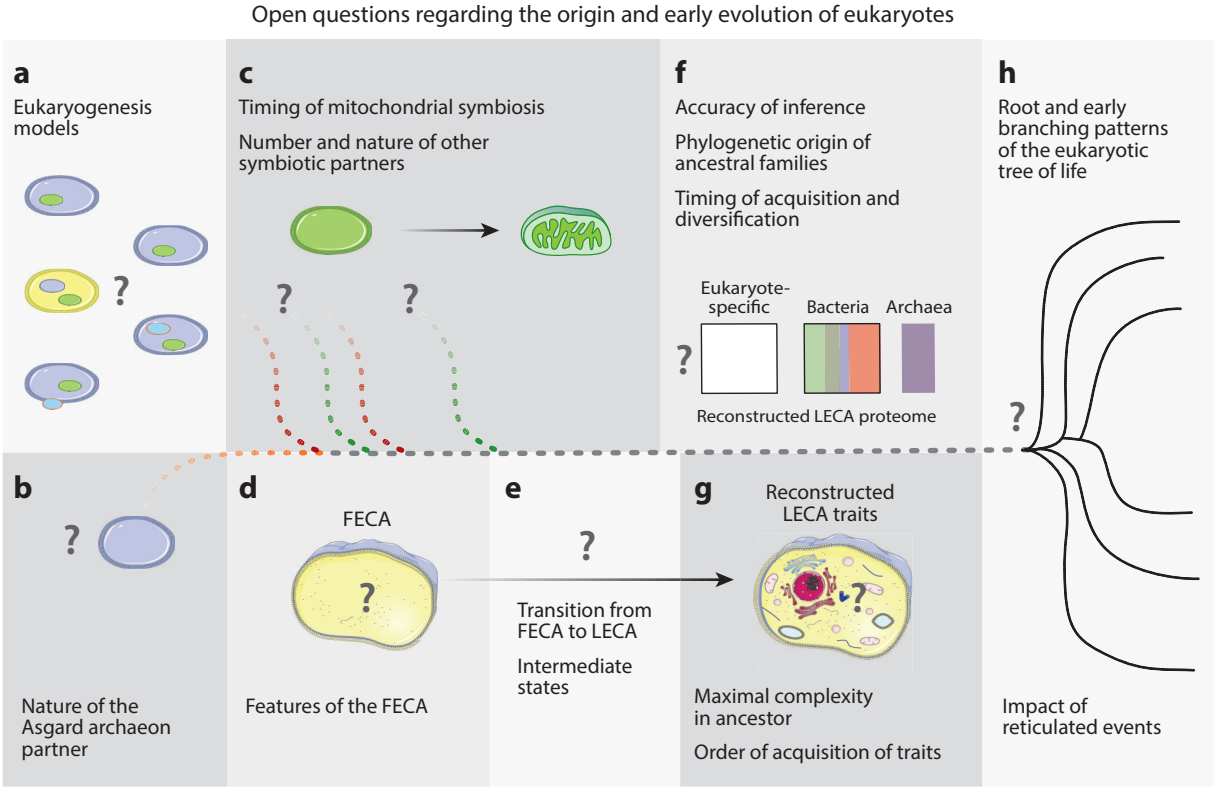

Figure 1

Open questions regarding the origin and early evolution of eukaryotes. (*a*) Several not fully exclusive models of symbioses involved in eukaryogenesis have been proposed that have different metabolic, phylogenetic, or structural implications. (*b*) Although much progress has been achieved in recent years regarding the nature of the archaeal ancestor of eukaryotes, many questions remain regarding its structural, genomic, and metabolic traits. (*c*) All models include symbiosis with an alphaproteobacterial ancestor that later became the mitochondrion, yet the exact metabolic nature and the timing of acquisition of this ancestor are unknown. Similarly, the number, timing, and nature of other putative symbiotic partners contributing to the eukaryotic lineage are debated. (*d*) The features of the FECA are far from established. (*e*) The time span of the transition from FECA to LECA is unknown, and so is the nature of the intermediates. (*f*) Our knowledge of the LECA relies on the presence of gene families or traits, inferred based on the analysis of extant eukaryotes. We do not know the accuracy of these reconstructions. Phylogenetic analysis of gene families present in the LECA shows an excess of bacterial proteins and a multiplicity of phylogenetic adscriptions to diverse bacterial groups. The origin of a significant fraction of eukaryote-specific proteins is unclear. We know very little about when the different proteomic components of the LECA originated, where they were acquired, and how they diversified through duplication events. (*g*) Most complex eukaryotic traits are inferred to have been present in the LECA, but we do not know in which order they appeared. (*h*) Despite much progress, uncertainty surrounds the position of the eukaryotic root and the relative branching order of the major lineages. Importantly, we do not know the extent and potential impact of reticulated processes (horizontal gene transfer, hybridization, etc.) during the early diversification of eukaryotes. Abbreviations: FECA, first eukaryotic common ancestor; LECA, last eukaryotic common ancestor. Cell images adapted from Servier Medical Art. They are available for reuse under the CC-BY 3.0 Unported license.

EUKARYOTIC DIVERSITY AND THE ELUSIVE TREE OF EUKARYOTES

Eukaryotes comprise a broad assemblage of lineages, most of which are solely unicellular organisms. Although sequencing efforts are still biased toward plant, animal, and fungal lineages (54), emerging environmental metagenomics and targeted efforts are slowly but steadily closing the gap in our knowledge about eukaryotic diversity, albeit still only from a genomic perspective. The new breadth of data, coupled with technical developments in phylogenetic methods, has increased the accuracy and resolution with which we can reconstruct the tree of eukaryotes. The major uncertainties concerning the origin of eukaryotes are the early branching patterns that separate the major eukaryotic clades, the placement of the root, and the length of the lineage separating eukaryotes from their prokaryotic ancestors. The deeper taxonomy of eukaryotes is not yet fully resolved, but most authors agree on the existence of at least five major eukaryotic clades, which leave out some taxa of unclear adscription (36) (**Figure 2**): (*a*) Opisthokonta include fungi, metazoans, and their closest relatives. (*b*) Amoebozoa encompass many lineages of amoeboid unicellular morphology and a variety of lifestyles. (*c*) Archaeplastida comprise unicellular and multicellular lineages with a cellulose cell wall, and their ancestor acquired photosynthetic capabilities by engulfing a cyanobacterial endosymbiont. This group includes land plants as well as green and red algal lineages. (*d*) The Stramenopila, Alveolata, Rhizaria (SAR) group is a diverse ensemble of unicellular lineages—but also some multicellular algae like kelps—with diverse morphologies and lifestyles. (*e*) Finally, earlier versions of the eukaryotic tree of life usually included a fifth supergroup, Excavata, whose monophyly has always been controversial. The latest versions of the tree tend to dismantle this contentious supergroup into at least two groups of unicellular organisms: Metamonada, including anaerobic flagellated lineages harboring mitochondrial relict organelles; and Discoba, comprising aerobic flagellates. This classification leaves Malawimonada, previously in Excavata, a small orphan lineage. Many other lineages of unclear adscription form additional minor groups (**Figure 2**). The relative relationships among the major supergroups and between them and the many orphan lineages are not robustly established. Difficulty in resolving early branching patterns may be due to saturation of the signal, early reticulations (e.g., horizontal gene transfer), a fast radiation, or a combination of these factors.

The root of the eukaryotic tree has been elusive, with alternative placements supported by different data sets and methods. Although controversy remains (36), two of the most recently supported scenarios place the root as separating Discoba from the rest of the supergroups (32) or as separating a clade with Opisthokonta and Amoebozoa from the rest of eukaryotes (10). Based on the latter scenario, the eukaryotic supergroups are often classified into broader groups: Amorphea (Opisthokonta and Amoebozoa) and Diaphoretickes (the rest of the supergroups) (1). Another important aspect of this phylogeny is the length of the stem that subtends the root of the eukaryotic tree and separates them from their closest bacterial relatives (38). As I elaborate in the section titled A Relative Time Line of Events, in most phylogenetic reconstructions the length subtending the last eukaryotic common ancestor (LECA) and connecting to its closest ancestors is relatively great—comparable to the length from the LECA to extant relatives. I have earlier referred to the tree of eukaryotes as the "palm tree of eukaryotes" (18, p. 1191) to stress the relatively long stem coupled to a subsequent rapid radiation. The most recent extreme of the stem, just preceding the radiation of the major supergroups, is represented by the LECA. Somewhere earlier in this stem is the first eukaryotic common ancestor (FECA), conceptualized as the common ancestor of all eukaryotes, including those that might have gone extinct. Key unresolved questions are how much time this stem represents and what the order of events in the evolutionary transition from the FECA to the LECA was.

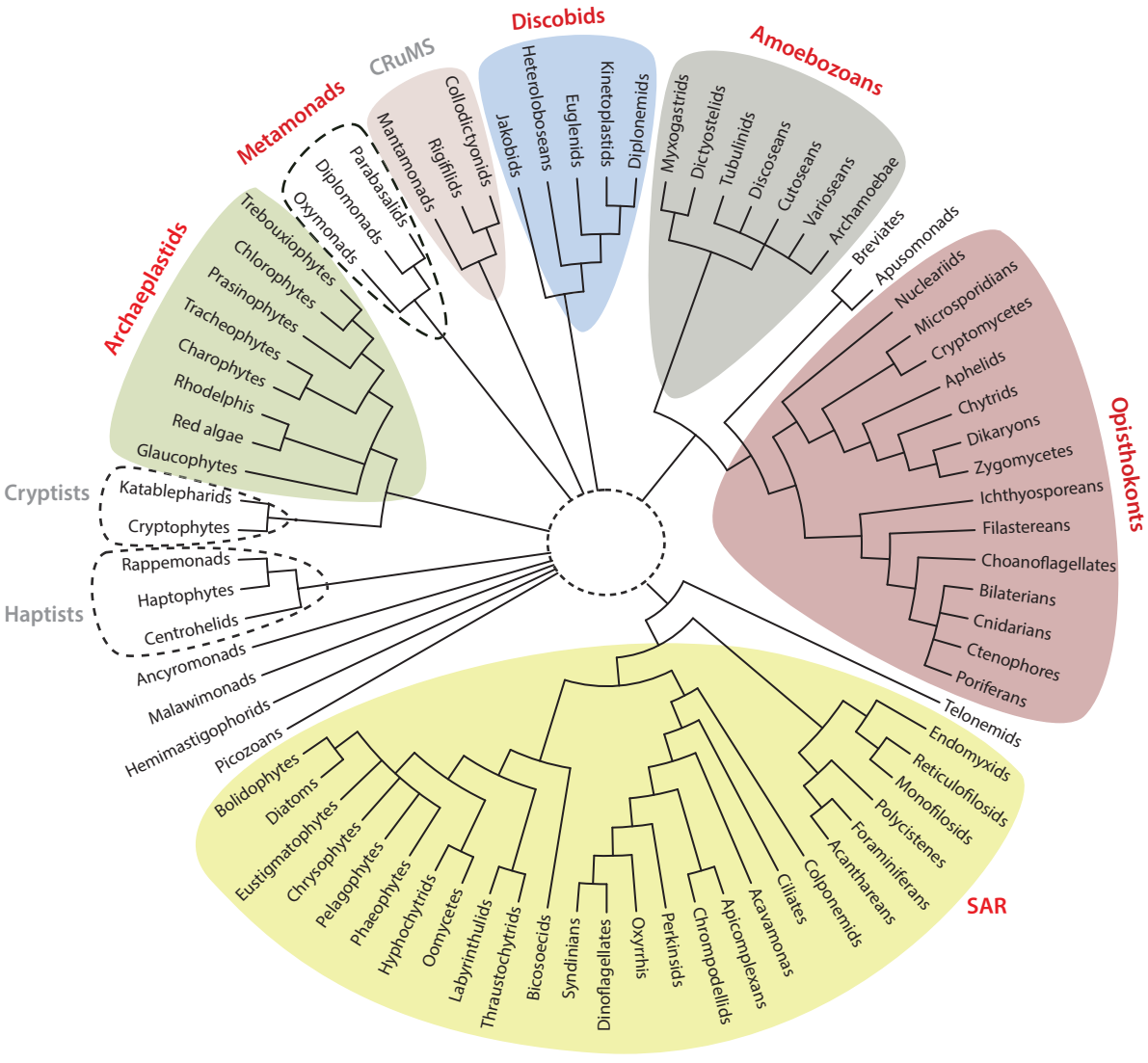

Figure 2

The eukaryotic tree of life. The phylogenetic relationships among the main eukaryotic lineages are schematically represented in this dendrogram, as defined by Reference 36. The names of the six main eukaryotic supergroups are red, and the lineages they comprise are circumscribed. Dashed borders surround assemblages of various eukaryotic lineages that are not included within the six major supergroups; the names of these smaller groups are gray. Multifurcations at various nodes indicate unresolved or disputed relationships. The root is represented as a dashed circle with several multifurcations to underscore its unresolved nature. Abbreviations: CRuMs, collodictyonids, rigifilids, and *Mantamonas*; SAR, Stramenopila, Alveolata, Rhizaria.

THE COMPLEXITY OF RECONSTRUCTED EUKARYOTIC ANCESTRAL TRAITS

A catalog of gene families and features that were likely present in the common ancestor can be inferred from a comparison of genomes and traits of extant species within an evolutionary framework (23). This approach has extensively been used to reconstruct the ancestral proteome and

the putative features of the LECA (7, 40). Irrespective of the position of the root, and of the early branching patterns of the eukaryotic tree (see above), most studies converge on a similar depiction of the LECA. The first aspect that emerges from these analyses is that the LECA proteome had a chimeric nature, comprising proteins originating from archaea or bacteria as well as a subset of proteins for which prokaryotic homologs cannot be identified (38). In addition, all reconstructions suggest that the bacterial component of the LECA had more proteins than the archaeal one, and that it had a diversity of putative phylogenetic origins beyond alphaproteobacteria, which suggests bacterial contributions other than a single alphaproteobacterial endosymbiont (18). When the functions of the reconstructed proteome are scrutinized to reconstruct traits likely present in the LECA, the results depict an already complex ancestor not much different from the most complex extant unicellular eukaryotes. Traits of the LECA that can be reconstructed include flagella; mitochondria; peroxisomes; a complex actin-tubulin-based cytoskeleton; and an endomembrane system already differentiated into endoplasmic reticulum, Golgi apparatuses, and other vesicles. The LECA not only was structurally complex but also had already developed intricate biological processes such as apoptosis, intron splicing, and meiosis. This reconstructed complexity raises two very fundamental questions. The first is how this complexity may have been built from simpler life-forms. The second question relates to whether the LECA really does represent a single entity (i.e., a species). Although most texts informally describe the LECA as a cell, views of the LECA range from a consortium of several lineages to a population from which we have reconstructed a pangenome (55). Although these alternatives would make a larger assemblage of ancestral genes possible, they only transpose the problem of cellular complexity to that of population or consortium complexity. Additionally, in my opinion, part of the inferred complexity of the LECA may result from constraints imposed by the widespread assumptions of vertical inheritance applied in current methods for ancestral reconstruction. Any such method will be easily confounded by early lateral gene transfers that occurred close to the roots of the major supergroups. Allow just some degree of early transfer, and the picture would change. Traits inferred to be present in the LECA would still be ancient, predating the divergence of most eukaryotic groups, but these might have appeared in different lineages and might have been brought together only after the LECA. Complexity has emerged as an ancestral feature in almost every single reconstruction of diverse eukaryotic ancestors, from yeasts (61) to animals (12), and it is reasonable to ask whether this pervasive pattern emerges partially from constraints imposed by our methodological approach. Findings in recent years have increasingly pointed to a role of lateral gene transfer in the evolution of many lineages of unicellular eukaryotes (19), and it is reasonable to assume that the eukaryotic ancestors were able to exchange genes, particularly at the start of their diversification, when they were more closely related.

THE CLOSEST ARCHAEAL RELATIVES OF EUKARYOTES

As we will see below, current hypotheses about eukaryotic origins generally propose the involvement of at least two partners in the process leading to the LECA: a bacterial endosymbiont that subsequently became the mitochondrion, and a host cell. The identity of the host has been investigated through the use of molecular phylogenies. In earlier trees based on ribosomal RNA, archaea, bacteria, and eukaryotes formed three separate clades named primary domains. A rooted version of this three-domain tree placed Archaea and Eukarya as sister clades, suggesting that eukaryotes were very distantly related to archaea and not more related to any specific group. More recently, phylogenetic analyses using more sophisticated models and expanded gene data sets have provided increasing support for an alternative tree topology in which the eukaryotic clade branches within Archaea, rather than next to it (73). The implications of this new two-domain topology were many,

including the possibility that better insights about eukaryotic ancestors could be gained by identifying the closest archaeal relatives of eukaryotes. The discovery via environmental genomics sampling of a new group of archaea, the Asgard archaea, that branched closer to eukaryotes than other archaea did was a major leap forward in our understanding of the origins of eukaryotes (63, 75). In recent years phylogenies with additional sequences have reinforced the relationship between eukaryotes and Asgard archaea and have pointed to a more specific Asgard group, the *Heimdallarchaeota* (72). Consistent with the close relationship between Asgard archaea and eukaryotes, the reconstructed genomes of Asgard archaea were found to encode many protein families previously thought to be eukaryote-specific, including proteins involved in cytoskeleton formation, membrane trafficking, and vesicle formation and transport. This genetic tool kit was suggestive of an intricate cellular complexity, perhaps intermediate between that of other archaea and that of eukaryotes. After much effort, some Asgard archaea have been brought to the laboratory, the first being the "*Candidatus* Prometheoarchaeum syntrophicum" strain MK-D1 (34). This is a major advance that enables deep characterization at the cellular and ultrastructural levels. Initial analyses of MK-D1 revealed no organelle-like intracellular structures but showed evidence of extracellular vesicles and elongated cellular protrusions. This strain was enriched in cultures containing the sulfur-reducing deltaproteobacterium *Halodesulfovibrio* and the methanogenic archaeon *Methanogenium*, which were inferred to consume hydrogen produced by the Asgard archaeon. However, Imachi et al. (34) were able to remove the deltaproteobacterium partner from the consortium and replace the methanogenic archaeon with another methanogenic archaeon, *Methanobacterium*, belonging to a different taxonomic class. This indicates that H_2-producing amino acid degradation and partner dependence may be common features across the Asgard archaea but that the syntrophic partner is not necessarily fixed and can be exchanged. MK-D1 belongs to *Lokiarchaeota*, and it is likely that similar deep characterizations of *Heimdallarchaeota* cells present in the initial enriched cultures will shed additional light on the closest relatives of eukaryotes. As we will see below, these recent discoveries have prompted the reformulation of hypothetical scenarios for the origin of eukaryotes in the last few years.

MITOCHONDRIAL ORIGINS AND EVOLUTION

Most scenarios for the origin and early evolution of eukaryotes grant the mitochondrial endosymbiosis a preeminent role (see below). Virtually all eukaryotic organisms analyzed to date possess mitochondria or mitochondrion-derived organelles (58), with extremely rare exceptions resulting from secondary loss (35). Therefore, the presence of this organelle in the LECA is not debated. As discussed above, early molecular phylogenetic analyses identified alphaproteobacteria as the closest relatives of mitochondria (30). Over the years, several potential lineages within *Alphaproteobacteria* have been proposed as sister clades of mitochondria (15, 58), with current scenarios suggesting that mitochondria derive from an early branching alphaproteobacterial lineage or a sister lineage of alphaproteobacteria (47), although the issue is not settled (11). If mitochondrial phylogenetic origins are not specific or mitochondria are similarly distant from different alphaproteobacterial clades, this means we cannot easily reconstruct the traits of the mitochondrial ancestor based only on observations from a particular alphaproteobacterial group. Features of the mitochondrial ancestor have been a matter of much debate, and a plethora of alternative scenarios have been proposed to explain its stable association with the host. Most proposals entail syntrophic scenarios pivoting around a single pathway or molecular exchange. For instance, scenarios envisioning the ancestral endosymbiont as an aerobically respiring heterotroph may stress secretion of ATP (59), removal of fermentation waste (67), a methane sink (43), or elimination of toxic oxygen from within (41). Scenarios depicting the endosymbiont as a

photosynthetic, facultatively aerobic, and biochemically versatile bacterium still stress a particular pathway, such as the release of hydrogen during fermentation (49) or the production of organic photosynthate (5). Ancestral genome reconstruction can help to narrow down such possibilities, and the gene repertoire of the proto-mitochondrion has been inferred based on phylogenetically tracing alphaproteobacterium derived genes present in eukaryotes (20, 22). These reconstructions yielded two important results. First, only a limited fraction of mitochondrial proteins that are currently present in this organelle (about 10–16%) can be traced back to an alphaproteobacterial ancestor. Second, most of the alphaproteobacterium-derived proteins were localized elsewhere in the cell, including the cytoplasm and peroxisomes (21). These findings imply that the alphaproteobacterial heritage of eukaryotes extends well beyond the extant mitochondrial organelle, with ramifications extending to almost every corner of the cell. This multitude of conserved pathways attests to a multifaceted benefit of the early mitochondrial symbiosis rather than merely a simple exchange of a few molecules (15, 20).

Whatever the initial trade between the host and the mitochondrial ancestor, the early endosymbiont underwent numerous changes until becoming an organelle. Perhaps the most relevant change was the development of a protein-transport system. This allowed the relocalization to the mitochondrion of nuclear-encoded proteins of alphaproteobacterial origin as well as proteins of different evolutionary origins (21, 22, 58). Other developments included streamlining of the mitochondrial genome as genes were lost or transferred to the nuclear genome and integration into the signaling and regulatory systems of the host. Most such transformations were already well established in the LECA, which likely possessed a mature organelle rather than a domesticated symbiont. This raises many questions regarding how long the transition from endosymbiont to organelle lasted and how this development relates to other developments elsewhere in the cell (discussed below). The metabolism of the mitochondrion also evolved. Mitochondrial energy metabolism, including oxidative phosphorylation, is one of the mitochondrial metabolic processes that are most widespread and enriched in alphaproteobacterium-derived proteins, suggesting this function was selected early on (22). However, energy metabolism pathways have been streamlined or lost independently multiple times in eukaryotic lineages adapted to anaerobiosis (58). Perhaps the most widespread metabolic feature of mitochondria, including anaerobic ones, is the pathway for Fe-S cluster assembly. This alphaproteobacterium-derived mitochondrial pathway synthesizes Fe-S clusters, which are essential components of many proteins. Although an Fe-S cluster assembly pathway exists in the cytoplasm, it requires an unknown sulfur-containing factor produced by the mitochondrial pathway. Due to its essential function, the mitochondrial Fe-S cluster assembly pathway is widespread, being an integral part even of highly reduced forms of mitochondria. Indeed, the acquisition of a bacterium-derived cytosolic version of this pathway through horizontal gene transfer likely led to the final degradation steps that caused the complete loss of mitochondria in truly amitochondriate eukaryotic species (35).

THE ENDOMEMBRANE SYSTEM: NUCLEUS, ENDOPLASMIC RETICULUM, AND PEROXISOMES

LECA reconstructions suggest a complex endomembrane system, comprising a nucleus, endoplasmic reticulum, Golgi apparatuses, and peroxisomes, and an endosomal system (7, 24, 40). Although it seems clear that other endomembrane compartments were initially derived from the nuclear envelope, the origin of the nucleus itself is poorly understood. Multiple hypotheses on the origin of the nucleus propose scenarios ranging from invaginations of the plasma membrane of an archaeon (2) to endosymbiosis of an archaeal cell within a bacterium (53) to de novo

synthesis of a new intracellular membrane as a response to mitochondrial endosymbiosis (28, 48). One central aspect of the debate regarding the origin of the nucleus is the chemical nature of the phospholipids forming the membrane. Archaeal membranes are more rigid and typically have glycerol-1-phosphate (G1P)-based, ether-linked isoprenoid phospholipids, while eukaryotes and bacteria share more flexible and permeable membranes composed of glycerol-3-phosphate (G3P)-based, usually ester-linked fatty acid phospholipids (44). Thus, models envisioning eukaryotic membranes originating from an archaeal host must invoke a transition in the chemical nature of membranes. Although recent experiments have shown that engineered bacteria can incorporate up to 30% of G1P-based phospholipids in their membranes (3), full G3P-to-G1P transition in a cell has never been accomplished experimentally, nor has it been observed in extant archaeal lineages.

Other endomembrane compartments would have progressively developed into a complex system of interconnected compartments and vesicles. The endomembrane trafficking system relies on specific types of proteins that define organelle identity and location and specific transport of vesicles and their cargoes. These proteins make up large superfamilies with paralogous clades that include a broad range of eukaryotes, with each clade typically functioning in a specific compartment. Most phylogenetic reconstructions of these families support a pre-LECA diversification of those broadly distributed paralogous clades, with subsequent lineage-specific duplications and losses resulting in intricate phylogenetic patterns (51). The organelle paralogy hypothesis posits that ancestral organelles differentiated into distinct compartments following duplications of organelle-specific proteins (51). According to this view, pre-LECA duplication of membrane trafficking systems likely reflects a stepwise diversification of the endomembrane system. This, together with the widespread presence of the different types of endomembrane compartments, suggests that the LECA possessed a complex compartmentalized endomembrane system and a fairly sophisticated protein-targeting and vesicle transport system.

Peroxisomes, now broadly recognized as an integral part of the endomembrane system, were once thought to be of endosymbiotic origin (8). However, later phylogenomic analyses and experimental results indicated that peroxisomes are evolutionary offshoots of the endomembrane system (14, 25). Also consistent with the organelle paralogy hypothesis, the origin of peroxisomes from the endomembrane system can be traced back to the duplication of an entire protein complex, the endoplasmic reticulum–associated degradation (ERAD) system, which works by translocating misfolded proteins from the ER lumen to drive their degradation by cytosolic proteasomes (25, 60). The paralogous complex in the peroxisome serves as a protein-transport system mediating import of peroxisome-targeted proteins into the peroxisomal matrix. Peroxisomal proteomes are very diverse across lineages, but they generally include enzymes for long-chain fatty acid oxidation and enzymes for reactive oxygen species detoxification, such as catalase (14). The presence of some widespread peroxisomal proteins of alphaproteobacterial descent that are involved in fatty acid oxidation prompted the proposal that peroxisomes originated to accommodate certain mitochondrial pathways and thus reduce oxidative damage in mitochondria (65, 66). An alternative metabolic scenario considers that several ancestral peroxisomal proteins, including one producing hydrogen peroxide, are not of alphaproteobacterial origin. In this scenario, fatty acid metabolic pathways producing oxidative damage were progressively isolated in endomembrane compartments, eventually giving rise to peroxisomes (16, 17). This latter scenario naturally integrates developmental and evolutionary connections between the peroxisome and the ER and is able to accommodate subsequent acquisition of proteins from different sources, including the mitochondrion. Furthermore, it is agnostic with regard to the relative timing of the origin of peroxisomes and mitochondria.

COMPLEXITY BEYOND MEMBRANES: THE ORIGINS OF OTHER EUKARYOTIC TRAITS

Beyond membrane-bound organelles, the ancestral eukaryote is inferred to have already possessed several complex molecular machineries, such as the cytoskeleton and the flagellum, and biological processes that are present only in eukaryotes, such as meiosis. Most central for theories on the origins of eukaryotes is the rise of the cytoskeleton and the related ability to perform phagocytosis. Whether the host of the mitochondrial endosymbiont had this ability is a recurrent and hotly debated topic (9, 50). Although the cytoskeleton was long considered a eukaryote-specific feature, it is now clear that both archaea and bacteria possess dynamic cytoskeletons (71). However, the eukaryotic cytoskeleton is much more intricate, comprising three types of cytoskeletal polymers (actin filaments, intermediate filaments, and microtubules) and three families of motor proteins (myosins, kinesins, and dyneins), and there is no simple relationship of homology between the cytoskeletons of prokaryotes and those of eukaryotes (57). Recent phylogenetic analyses support an archaeal origin of the main eukaryotic cytoskeleton proteins, suggesting that the archaeal ancestor of eukaryotes possessed an actin-based cytoskeleton including branched actin filaments, which would have allowed it to produce actin-supported membrane protrusions (39, 57, 71), as observed in Asgard archaea (34). Phylogenomic analyses also reveal a burst of gene duplications predating the LECA that increased the complexity of the cytoskeleton to levels similar to that of extant eukaryotes (39, 57). However, these results do not reveal whether the mitochondrion was engulfed by membrane protrusions of an archaeon-like host, by a more developed phagocytosis process in a more evolved proto-eukaryote, or by other means.

Early hypotheses proposed that flagella originated from an endosymbiotic, spirochete-like bacterium, based on morphological similarities (59), but this idea never received empirical support and has been abandoned. Eukaryotic, archaeal, and bacterial flagella are only superficially similar. They have fundamental ultrastructural differences and are clearly nonhomologous (37). Reconstructions of the evolution of eukaryotic flagellar structures trace them to the LECA, where flagella were already fully developed (52), but how this complex machinery originated before the LECA remains poorly understood. Meiosis is another widespread feature of eukaryotes that is absent from any known prokaryote. Although archaea and bacteria are also capable of genetic exchange and recombination, the nature of these processes in eukaryotes is unique: Sexual reproduction involves reassortment and recombination of chromosomes, controlled fusion of haploid gametes, and diploid-to-haploid reduction via meiosis. That a sexual cycle existed in the LECA is no longer a matter of debate. Sex and genes exclusively involved in meiosis have been observed in all major groups of eukaryotes. Recent hypotheses posit that mechanisms for cell fusion and ploidy reduction appeared before the meiotic cycle (27). However, it is uncertain how meiosis emerged from preexisting DNA repair machineries.

We could enumerate other typical eukaryotic complex features: the spliceosome, the nuclear pore complex, the degradosome, the mitochondrial import machinery, and so on. Tales of the origins of other complex eukaryotic traits can often be summarized in a similar way: They were already highly developed in the LECA and fundamentally distinct from analogous systems (if any) present in prokaryotes. Sometimes we can trace homology relationships of some components to archaeal or bacterial counterparts, or to both, but we are always able to find fundamental eukaryotic innovations that resulted in extra layers of complexity. All must have been present in a complex form in the LECA, and all must have been absent or very simple in the FECA. The emerging picture is not one but a plethora of parallel developments of structures in the transition from FECA to LECA. This may seem overwhelming, and the degree of the challenge in envisioning such a complex ancestor very much depends on how much time this transition involved. Were all these developments truly parallel? Did all proceed as a quick succession of

radical events, or did complexity increase slowly in a stepping-stone-like manner through less complex intermediates that are long extinct?

A RELATIVE TIME LINE OF EVENTS

Most scenarios for the origin of eukaryotes comprise a few steps, without elaborate descriptions of how the intermediate transitions occurred. This is understandable given the scarcity of elements to judge what intermediate steps might have been involved and in what order they might have occurred. One common approach to the time line of events has been to judge the plausibility of intermediate scenarios. If, for instance, the existence of trait 1 is dependent on trait 2, then we can infer that development of trait 2 predated development of trait 1. Although apparently simple, this type of reasoning is extremely complicated with regard to ancestral eukaryotic evolution. Let us take two sharply debated examples that have provided no conclusive answers. In the first case, the apparent requirement for a fully developed ability to phagocytose an endosymbiont seemed to support the existence of a fully developed cytoskeleton as a prerequisite for mitochondrial endosymbiosis (4, 9). However, this inference was weakened by the realization that some bacteria can live within other bacteria (69). Eukaryote-like phagocytosis is therefore not necessary for endosymbiosis. In the second case, the proposition that significant genome and cellular complexity could be achieved only by the energy boost from mitochondria argued that mitochondrial endosymbiosis was a necessary early step in eukaryogenesis (42). However, this argument has been seriously questioned because of the energetics assumptions (6, 46) and because there are complex eukaryotes devoid of ATP-producing mitochondria (31). Nevertheless, that mitochondrial endosymbiosis could have occurred before the evolution of phagocytosis does not mean that it did; and conversely, that mitochondria are not necessary for building cellular complexity does not mean that complexity predated mitochondrial endosymbiosis.

Phylogenomics is an alternative approach to assess possible time lines. Phylogenetic reconstruction provides information on topology (i.e., how clades are related to each other in a phylogeny) and on branch length, which represents the number of substitutions per site along a lineage. As mentioned above, the stem subtending the eukaryotic clade is long. Importantly, this length depends on what out-group is used, with archaeal out-groups providing longer stems than bacterial out-groups, and with Asgard archaeal and alphaproteobacterial out-groups providing shorter stems compared to their archaeal and bacterial counterparts (18). Interpretation of branch lengths is complex because they are determined by both elapsed time and evolutionary rate, and rates can vary across protein families and clades. Thus, one can interpret variations in stem length by stressing differences in evolutionary rates of the clades or proteins compared. This interpretation implies a scenario in which evolutionary rates shifted differently between proteins of different phylogenetic origin but similarly among proteins of the same phylogenetic origin, despite their functional and structural heterogeneity. Alternatively, one can interpret these differences assuming that different out-groups provide different branch lengths, simply because they diverged from eukaryotes at different times, so that the archaeal contributions to eukaryotes occurred before bacterial ones. That topological analyses place Asgard archaea and alphaproteobacteria as the closest archaeal and bacterial sister clades of eukaryotes is also congruent with shorter divergence times with these out-groups compared to other clades of the same domain. This approach was the basis of a phylogenomic analysis correlating branch length with phylogenetic ancestry in individual proteins of the LECA proteome (56). This study revealed that alphaproteobacterium-derived genes typically had shorter stems compared to genes with different bacterial origins, regardless of their molecular function or current cellular localization. This suggests that mitochondria arose relatively late in a host that was already chimeric, harboring many other proteins of different bacterial

origins. More recently, this approach has been used to compare relative branch lengths among pre-LECA duplications, establishing a putative time line for the diversification of the different organelles (70).

REFRAMING SCENARIOS FOR THE ORIGIN OF EUKARYOTES

Many findings over the last few years have greatly shaken the field. This has led to the proposal of novel scenarios or the reframing of existing ones to better reconcile them with the data. With no intention of being exhaustive, I outline here the most relevant novel ideas and readjustments of previous models. In doing so I focus on the following central aspects: the metabolic relationships of the mitochondrial endosymbiont and its host, the mechanism of engulfment, the type and number of partners involved, and, finally, the relative time line of events.

Most scenarios posit a metabolic symbiosis (syntrophy) of the alphaproteobacterial ancestor of mitochondria and an archaeal partner, usually acting as endosymbiont and host, respectively. Although the alphaproteobacterial nature of the mitochondrial ancestor has been known for decades, precision regarding the phylogenetic adscription of the archaeal ancestor has remained elusive until very recently. The identification of Asgard archaea as the closest archaeal relatives of eukaryotes has helped to narrow down the metabolic potential of the ancestral archaeal partner. As a result, new metabolic models and readjustments of previous ones consider the archaeal partner to be an Asgard archaeon (**Figure 3**). New models have emerged based on observed and inferred properties of Asgard archaea, and two of the first proposed metabolic scenarios, the hydrogen hypothesis (49) and the syntrophy hypothesis (53), have been reframed.

The original hydrogen hypothesis postulated a methanogenic archaeal host using hydrogen and carbon dioxide from a hydrogen-producing endosymbiotic alphaproteobacterium (49). Based on the inferred metabolism of some Asgard archaea, the notion of a methanogenic host was abandoned in favor of a Wood-Ljungdahl pathway for carbon fixation, but it was still believed to be dependent on a putative hydrogen-producing mitochondrial ancestor (62). According to a more recent proposal, the reverse flow model, the hydrogen (or other reducing equivalents) were transferred from an anaerobic heterotrophic archaeon to the alphaproteobacterium (64).

The earliest version of the syntrophy hypothesis proposed a tripartite symbiosis based on the hydrogen exchange between a methanogenic archaeon and a fermentative, sulfate-reducing myxobacterium (*Deltaproteobacteria*) and with the subsequent participation of a facultatively aerobic, methanotrophic alphaproteobacterium (53). This hypothesis has now been substantially revised, with the archaeal partner producing hydrogen and degrading small organics, a sulfate-reducing deltaproteobacterial host using that hydrogen, and a facultative aerobic, sulfide-oxidizing alphaproteobacterium possibly being capable of anoxygenic photosynthesis (45). The recently proposed entangle-engulf-enslave (E3) model also puts forward a tripartite symbiosis in microaerophilic environments, between an Asgard archaeon catabolizing amino acids to short-chain fatty acids and hydrogen, a sulfate-reducing bacterium, and an alphaproteobacterium (34). Models involving more than two partners better explain the presence of a large fraction of bacterial proteins with non-alphaproteobacterial origins, and their possible earlier acquisition (56). However, these models still envision a maximum of two bacterial partners. The original serial endosymbiosis theory postulated several symbioses (59) but explicitly proposed only two bacterial partners predating the LECA, each leaving a footprint in the form of a different organelle (flagellum and mitochondrion).

To reconcile the phylogenetic diversity of LECA proteins and the observed relationship between phylogenetic origin and branch length (56), I recently proposed the premitochondrial symbioses hypothesis, which postulates that other symbiotic partners preceded the mitochondrial

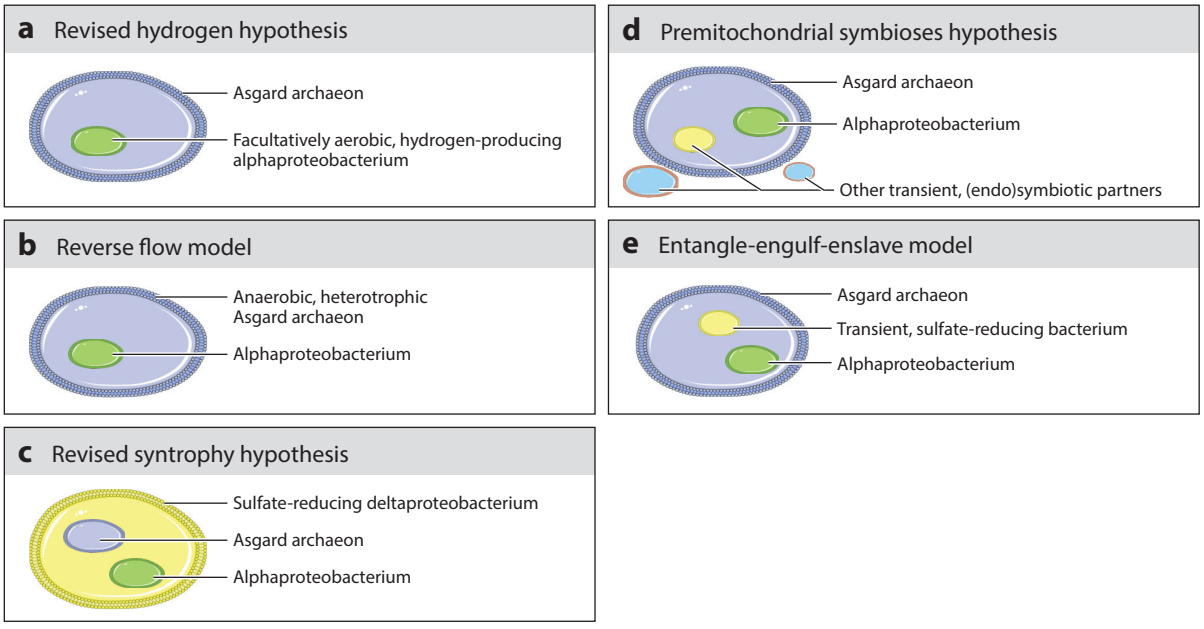

Figure 3

Recently proposed or updated models of eukaryogenesis. (*a*) The revised hydrogen hypothesis posits an Asgard archaeon host using the Wood-Ljungdahl pathway for carbon fixation and using hydrogen produced by a facultatively aerobic alphaproteobacterium (62). (*b*) The reverse flow model proposes the host was an anaerobic heterotrophic archaeon that produced hydrogen (or other reducing equivalents) that was used by an endosymbiotic alphaproteobacterium (64). (*c*) According to the revised syntrophy hypothesis, the archaeal partner produced hydrogen and degraded small organics; a sulfate-reducing deltaproteobacterial host used that hydrogen; and a facultatively aerobic, sulfide-oxidizing alphaproteobacterium was possibly able to perform anoxygenic photosynthesis (45). (*d*) The premitochondrial symbioses hypothesis postulates that other symbiotic partners preceded mitochondrial endosymbiosis (18). These earlier symbioses were not necessarily endosymbiotic and did not leave new organelles but transferred genes to the eukaryotic lineages. The alphaproteobacterial ancestor was one of the last symbionts, and the one that was ultimately fixed as an organelle. (*e*) The recently proposed entangle-engulf-enslave model posits a tripartite symbiosis between an Asgard archaeon catabolizing amino acids to short-chain fatty acids and hydrogen, a sulfate-reducing bacterium, and an alphaproteobacterium (34). Cell icons reproduced from images obtained from Servier Medical Art under a Creative Commons Attribution 3.0 Unported license.

endosymbiosis (18). Contrary to the serial endosymbiotic theory, the proposed earlier symbioses were not necessarily endosymbiotic and did not leave new organelles. Rather, it is proposed that earlier symbioses were mediated by different metabolic associations and were transient but left genes in the partner that later became the LECA. The alphaproteobacterial ancestor was one of the last symbionts, and the one that was ultimately fixed as an organelle, a transition possibly facilitated by previous developments resulting from interactions with alternative partners. This idea fits with current knowledge of Asgard archaea, which can establish symbiotic associations with more than one partner, with this partner being interchangeable with others providing similar metabolic goods (34). It is also reminiscent of the usually transient and serial nature of many symbionts of modern organisms (33). In the premitochondrial symbioses model the mitochondrion was one of the latest partners, being definitely established as an obligate endosymbiont and subsequently evolving into the mitochondrion. The premitochondrial symbioses hypothesis is agnostic about the metabolic association of the mitochondrial ancestor and its host and is thus compatible with some of the proposed metabolic models, as long as they do not posit the ancestor of mitochondria as the first symbiont and they can accommodate additional, earlier partners.

Regardless of the number of partners and the metabolic basis of the symbiosis, the different models may elaborate proposed mechanisms for the evolution of ancestral eukaryotic traits, as well as their time lines. Early eukaryogenesis models such as the archaeozoon hypothesis depicted a proto-eukaryotic ancestor different from archaea and bacteria and having most typical eukaryotic traits except mitochondria (4). These models were abandoned with the realization that putative amitochondriate eukaryotes were actually derived from mitochondriate ancestors. Some current models still assume the mitochondrial ancestor had a relatively complex host, but one that was archaeal (Asgard) in nature, with more or less developed cytoskeleton and endomembrane systems and the ability to phagocytose. Other recent scenarios assume alternative forms of engulfment that were different from modern eukaryote-like phagocytosis but still required complex cytoskeletal organization. According to these models, including the inside-out model (2) and the E3 model (34), mitochondria were also acquired relatively late, by means of membrane extrusions that progressively surrounded the alphaproteobacteria and eventually integrated them. All models involve an integration of archaeal and bacterial cells and thus imply a transition in phospholipids. Most models envision archaeal hosts and a progressive replacement of phospholipids in their membranes after symbiosis with the mitochondrial ancestor. The syntrophy hypothesis, in contrast, does not require a membrane transition, as it proposes that the host for the endosymbiosis was a deltaproteobacterium. In this model, the nucleus originated early from a host-derived compartment containing an endosymbiotic archaeon, with the archaeal membrane eventually disappearing.

CONCLUDING REMARKS

Although the origin of eukaryotes remains mostly unresolved, many recent developments have led to progressively more concrete scenarios and empirical assessment of their feasibility. In the last decade, most developments have come not only from our unprecedented ability to explore genomes of extant organisms and analyze them from an evolutionary perspective (phylogenomics) but also from direct observation of and experimentation with diverse organisms and ecosystems. Given the scarcity and fragmented nature of available data, I argue that only a multidisciplinary approach that encompasses phylogenomics, microbial diversity and ecology, biochemistry, and cell biology and that effectively combines (*a*) observations from natural systems, (*b*) computational inference of past events, and (*c*) direct experimentation will allow us to move forward. However, as with many other fundamental questions about the deep past, we must be prepared to confront the idea that our knowledge will always be incomplete. Inherent to the evolutionary process is the permanent extinction of most lineages and the continuous tinkering of sequences and traits, both of which inevitably limit, blur, and even distort the signals that we read from extant organisms. Our reconstructed scenarios, although assuaging our thirst for knowledge, will almost certainly be found wrong in most details, only approximately true in some aspects, and always subject to reform. In my view, most models likely underestimate the time required for eukaryogenesis, the number of partners involved, and the gradual nature of the process. This is a result of our search for simple models—following Occam's razor—and the compressed nature of the information we obtain. However, phylogenomics results suggest that the formation of the LECA took a long time, that diverse phylogenetic inputs—beyond an alphaproteobacterium and an Asgard archaeon—contributed to the LECA proteome, and that these inputs were likely integrated at different times. Moreover, nature shows us that evolution usually does not join two distant points with a straight line and that adaptation and innovation often progress gradually, involving trial and error, tinkering, and exaptation. All of this argues for the need for more complex models involving more time, more partners, and more gradual steps.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The author acknowledges support from the Spanish Ministry of Science and Innovation (PGC2018-099921-B-I00), cofounded by the European Regional Development Fund (ERDF); from the University and Research Grants Management Agency (AGAUR) SGR423; from the European Union's Horizon 2020 research and innovation program (ERC-2016-724173); from the Gordon and Betty Moore Foundation (GBMF9742); and from the Instituto de Salud Carlos III (INB grant PT17/0009/0023).

LITERATURE CITED

1. Adl SM, Simpson AGB, Lane CE, Lukeš J, Bass D, et al. 2012. The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* 59(5):429–514
2. Baum DA, Baum B. 2014. An inside-out origin for the eukaryotic cell. *BMC Biol.* 12:76
3. Caforio A, Siliakus MF, Exterkate M, Jain S, Jumde VR, et al. 2018. Converting *Escherichia coli* into an archaebacterium with a hybrid heterochiral membrane. *PNAS* 115(14):3704–9
4. Cavalier-Smith T. 1987. Eukaryotes with no mitochondria. *Nature* 326(6111):332–33
5. Cavalier-Smith T. 2006. Origin of mitochondria by intracellular enslavement of a photosynthetic purple bacterium. *Proc. R. Soc. B* 273(1596):1943–52
6. Chiyomaru K, Takemoto K. 2020. Revisiting the hypothesis of an energetic barrier to genome complexity between eukaryotes and prokaryotes. *R. Soc. Open Sci.* 7(2):191859
7. Dacks JB, Field MC, Buick R, Eme L, Gribaldo S, et al. 2016. The changing view of eukaryogenesis—fossils, cells, lineages and how they all come together. *J. Cell Sci.* 129(20):3695–703
8. de Duve C. 1969. Evolution of the peroxisome. *Ann. N. Y. Acad. Sci.* 168(2):369–81
9. de Duve C. 2007. The origin of eukaryotes: a reappraisal. *Nat. Rev. Genet.* 8(5):395–403
10. Derelle R, Torruella G, Klimeš V, Brinkmann H, Kim E, et al. 2015. Bacterial proteins pinpoint a single eukaryotic root. *PNAS* 112(7):E693–99
11. Fan L, Wu D, Goremykin V, Xiao J, Xu Y, et al. 2020. Phylogenetic analyses with systematic taxon sampling show that mitochondria branch within Alphaproteobacteria. *Nat. Ecol. Evol.* 4(9):1213–19
12. Fernández R, Gabaldón T. 2020. Gene gain and loss across the metazoan tree of life. *Nat. Ecol. Evol.* 4(4):524–33
13. Doolittle WF. 2014. How natural a kind is "Eukaryote?" *Cold Spring Harb. Perspect. Biol.* 6(6):a015974
14. Gabaldón T. 2010. Peroxisome diversity and evolution. *Philos. Trans. R. Soc. London B* 365(1541):765–73
15. Gabaldón T. 2012. Mitochondrial origins. In *Organelle Genetics: Evolution of Organelle Genomes and Gene Expression*, ed. CE Bullerwell, pp. 3–18. Berlin: Springer-Verlag
16. Gabaldón T. 2014. Evolutionary considerations on the origin of peroxisomes from the endoplasmic reticulum, and their relationships with mitochondria. *Cell. Mol. Life Sci.* 71(13):2379–82
17. Gabaldón T. 2014. A metabolic scenario for the evolutionary origin of peroxisomes from the endomembranous system. *Cell. Mol. Life Sci.* 71(13):2373–76
18. Gabaldón T. 2018. Relative timing of mitochondrial endosymbiosis and the "pre-mitochondrial symbioses" hypothesis. *IUBMB Life*. 70(12):1188–96
19. Gabaldón T. 2020. Patterns and impacts of nonvertical evolution in eukaryotes: a paradigm shift. *Ann. N. Y. Acad. Sci.* 1476(1):78–92
20. Gabaldón T, Huynen MA. 2003. Reconstruction of the proto-mitochondrial metabolism. *Science* 301(5633):609
21. Gabaldón T, Huynen MA. 2004. Shaping the mitochondrial proteome. *Biochim. Biophys. Acta Bioenerg.* 1659(2–3):212–20

22. Gabaldón T, Huynen MA. 2007. From endosymbiont to host-controlled organelle: the hijacking of mitochondrial protein synthesis and metabolism. *PLOS Comput. Biol.* 3(11):2209–18
23. Gabaldón T, Huynen MA. 2008. Reconstruction of ancestral proteomes. In *Ancestral Sequence Reconstruction*, ed. D Liberles, pp. 128–38. Oxford, UK: Oxford Univ. Press
24. Gabaldón T, Pittis AA. 2015. Origin and evolution of metabolic sub-cellular compartmentalization in eukaryotes. *Biochimie* 119:262–68
25. Gabaldón T, Snel B, van Zimmeren F, Hemrika W, Tabak H, Huynen MA. 2006. Origin and evolution of the peroxisomal proteome. *Biol. Direct.* 1:8
26. Garza DR, Dutilh BE. 2015. From cultured to uncultured genome sequences: metagenomics and modeling microbial ecosystems. *Cell. Mol. Life Sci.* 72(22):4287–308
27. Goodenough U, Heitman J. 2014. Origins of eukaryotic sexual reproduction. *Cold Spring Harb. Perspect. Biol.* 6(3):a016154
28. Gould SB, Garg SG, Martin WF. 2016. Bacterial vesicle secretion and the evolutionary origin of the eukaryotic endomembrane system. *Trends Microbiol.* 24(7):525–34
29. Gray MW. 2017. Lynn Margulis and the endosymbiont hypothesis: 50 years later. *Mol. Biol. Cell* 28(10):1285–87
30. Gray MW, Burger G, Lang BF. 1999. Mitochondrial evolution. *Science* 283(5407):1476–81
31. Hampl V, Čepička I, Eliáš M. 2019. Was the mitochondrion necessary to start eukaryogenesis? *Trends Microbiol.* 27(2):96–104
32. He D, Fiz-Palacios O, Fu CJ, Tsai CC, Baldauf SL. 2014. An alternative root for the eukaryote tree of life. *Curr. Biol.* 24(4):465–70
33. Husnik F, Keeling PJ. 2019. The fate of obligate endosymbionts: reduction, integration, or extinction. *Curr. Opin. Genet. Dev.* 58–59:1–8
34. Imachi H, Nobu MK, Nakahara N, Morono Y, Ogawara M, et al. 2020. Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* 577(7791):519–25
35. Karnkowska A, Vacek V, Zubáčová Z, Treitli SC, Petrželková R, et al. 2016. A eukaryote without a mitochondrial organelle. *Curr. Biol.* 26(10):1274–84
36. Keeling PJ, Burki F. 2019. Progress towards the tree of eukaryotes. *Curr. Biol.* 29(16):R808–17
37. Khan S, Scholey JM. 2018. Assembly, functions and evolution of archaella, flagella and cilia. *Curr. Biol.* 28(6):R278–92
38. Koonin EV. 2010. The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biol.* 11(5):209
39. Koonin EV. 2015. Origin of eukaryotes from within archaea, archaeal eukaryome and bursts of gene gain: eukaryogenesis just made easier? *Philos. Trans. R. Soc. B* 370(1678):20140333
40. Koumandou VL, Wickstead B, Ginger ML, van der Giezen M, Dacks JB, Field MC. 2013. Molecular paleontology and complexity in the last eukaryotic common ancestor. *Crit. Rev. Biochem. Mol. Biol.* 48(4):373–96
41. Kurland C, Andersson S. 2000. Origin and evolution of the mitochondrial proteome. *Microbiol. Mol. Biol. Rev.* 64(4):786–820
42. Lane N, Martin W. 2010. The energetics of genome complexity. *Nature* 467(7318):929–34
43. Lara E, Moreira D, López-García P. 2010. The environmental clade LKM11 and *Rozella* form the deepest branching clade of fungi. *Protist* 161(1):116–21
44. Lombard J, López-García P, Moreira D. 2012. The early evolution of lipid membranes and the three domains of life. *Nat. Rev. Microbiol.* 10(7):507–15
45. López-García P, Moreira D. 2020. The Syntrophy hypothesis for the origin of eukaryotes revisited. *Nat. Microbiol.* 5(5):655–67
46. Lynch M, Marinov GK. 2015. The bioenergetic costs of a gene. *PNAS* 112(51):15690–95
47. Martijn J, Vosseberg J, Guy L, Offre P, Ettema TJG. 2018. Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature* 557(7703):101–5
48. Martin W, Koonin EV. 2006. Introns and the origin of nucleus-cytosol compartmentalization. *Nature* 440(7080):41–45
49. Martin W, Müller M. 1998. The hydrogen hypothesis for the first eukaryote. *Nature* 392(6671):37–41

50. Martin WF, Tielens AGM, Mentel M, Garg SG, Gould SB. 2017. The physiology of phagocytosis in the context of mitochondrial origin. *Microbiol. Mol. Biol. Rev.* 81(3):e00008-17
51. Mast FD, Barlow LD, Rachubinski RA, Dacks JB. 2014. Evolutionary mechanisms for establishing eukaryotic cellular complexity. *Trends Cell Biol.* 24(7):435–42
52. Mitchell DR. 2007. The evolution of eukaryotic cilia and flagella as motile and sensory organelles. *Adv. Exp. Med. Biol.* 607:130–40
53. Moreira D, López-García P. 1998. Symbiosis between methanogenic archaea and δ-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J. Mol. Evol.* 47(5):517–30
54. Mukherjee S, Stamatis D, Bertsch J, Ovchinnikova G, Katta HY, et al. 2019. Genomes OnLine database (GOLD) v.7: updates and new features. *Nucleic Acids Res.* 47(D1):D649–59
55. O'Malley MA, Leger MM, Wideman JG, Ruiz-Trillo I. 2019. Concepts of the last eukaryotic common ancestor. *Nat. Ecol. Evol.* 3(3):338–44
56. Pittis AA, Gabaldón T. 2016. Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature* 531(7592):101–4
57. Pollard TD, Goldman RD. 2018. Overview of the cytoskeleton from an evolutionary perspective. *Cold Spring Harb. Perspect. Biol.* 10(7):a030288
58. Roger AJ, Muñoz-Gómez SA, Kamikawa R. 2017. The origin and diversification of mitochondria. *Curr. Biol.* 27(21):R1177–92
59. Sagan L. 1967. On the origin of mitosing cells. *J. Theor. Biol.* 14(3):255–74
60. Schlüter A, Fourcade S, Ripp R, Mandel JL, Poch O, Pujol A. 2006. The evolutionary origin of peroxisomes: an ER-peroxisome connection. *Mol. Biol. Evol.* 23(4):838–45
61. Shen XX, Opulente DA, Kominek J, Zhou X, Steenwyk JL, et al. 2018. Tempo and mode of genome evolution in the budding yeast subphylum. *Cell* 175(6):1533–45.e20
62. Sousa FL, Neukirchen S, Allen JF, Lane N, Martin WF. 2016. Lokiarchaeon is hydrogen dependent. *Nat. Microbiol.* 1:16034
63. Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, et al. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521(7551):173–79
64. Spang A, Stairs CW, Dombrowski N, Eme L, Lombard J, et al. 2019. Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. *Nat. Microbiol.* 4(7):1138–48
65. Speijer D. 2011. Oxygen radicals shaping evolution: why fatty acid catabolism leads to peroxisomes while neurons do without it. *BioEssays* 33(2):88–94
66. Speijer D. 2014. Reconsidering ideas regarding the evolution of peroxisomes: the case for a mitochondrial connection. *Cell. Mol. Life Sci.* 71(13):2377–78
67. Speijer D. 2017. Alternating terminal electron-acceptors at the basis of symbiogenesis: how oxygen ignited eukaryotic evolution. *BioEssays* 39(2):1600174
68. Stanier R, Doudoroff M, Adelberg E. 1963. *The Microbial World*. Englewood Cliffs, NJ: Prentice Hall. 2nd ed.
69. Von Dohlen CD, Kohler S, Alsop ST, McManus WR. 2001. Mealybug β-proteobacterial endosymbionts contain γ-proteobacterial symbionts. *Nature* 412(6845):433–36
70. Vosseberg J, van Hooff JJE, Marcet-Houben M, van Vlimmeren A, van Wijk LM, et al. 2021. Timing the origin of eukaryotic cellular complexity with ancient duplications. *Nat. Ecol. Evol.* 5:92–100
71. Wickstead B, Gull K. 2011. The evolution of the cytoskeleton. *J. Cell Biol.* 194(4):513–25
72. Williams TA, Cox CJ, Foster PG, Szöllősi GJ, Embley TM. 2020. Phylogenomics provides robust support for a two-domains tree of life. *Nat. Ecol. Evol.* 4(1):138–47
73. Williams TA, Foster PG, Nye TMW, Cox CJ, Embley TM. 2012. A congruent phylogenomic signal places eukaryotes within the Archaea. *Proc. R. Soc. B* 279(1749):4870–79
74. Yu FB, Blainey PC, Schulz F, Woyke T, Horowitz MA, Quake SR. 2017. Microfluidic-based mini-metagenomics enables discovery of novel microbial lineages from complex environmental samples. *eLife* 6:e26580
75. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L, et al. 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541(7637):353–58