Current Biology
Review

The Origin and Diversification of Mitochondria

Andrew J. Roger[1,*], Sergio A. Muñoz-Gómez[1], and Ryoma Kamikawa[2]
[1]Centre for Comparative Genomics and Evolutionary Bioinformatics, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada
[2]Graduate School of Human and Environmental Studies, Graduate School of Global Environmental Studies, Kyoto University, Japan
*Correspondence: Andrew.Roger@Dal.Ca
https://doi.org/10.1016/j.cub.2017.09.015

Mitochondria are best known for their role in the generation of ATP by aerobic respiration. Yet, research in the past half century has shown that they perform a much larger suite of functions and that these functions can vary substantially among diverse eukaryotic lineages. Despite this diversity, all mitochondria derive from a common ancestral organelle that originated from the integration of an endosymbiotic alphaproteobacterium into a host cell related to Asgard Archaea. The transition from endosymbiotic bacterium to permanent organelle entailed a massive number of evolutionary changes including the origins of hundreds of new genes and a protein import system, insertion of membrane transporters, integration of metabolism and reproduction, genome reduction, endosymbiotic gene transfer, lateral gene transfer and the retargeting of proteins. These changes occurred incrementally as the endosymbiont and the host became integrated. Although many insights into this transition have been gained, controversy persists regarding the nature of the original endosymbiont, its initial interactions with the host and the timing of its integration relative to the origin of other features of eukaryote cells. Since the establishment of the organelle, proteins have been gained, lost, transferred and retargeted as mitochondria have specialized into the spectrum of functional types seen across the eukaryotic tree of life.

Introduction

Mitochondria are essential double-membrane bound subcellular compartments that are best known as the 'powerhouses' that supply eukaryotes with energy in the form of ATP to serve their cellular needs. We are taught in introductory biology courses that mitochondria are the site of aerobic respiration, a complex biochemical process by which pyruvate is oxidized to CO_2, generating reduced cofactors that drive the electron transport chain (ETC) to chemiosmotically fuel ATP synthesis. The final electron acceptor for this process is oxygen, which is why the majority of eukaryotes require oxygen to survive. Yet the last half a century of research into the mitochondria of a number of model system eukaryotes has revealed that these organelles do far more than just aerobic respiration. Indeed, mitochondrial proteomes typically consist of greater than 1,000 proteins that function in a wide variety of critically important biochemical processes including protein synthesis, amino acid and nucleotide metabolism, fatty-acid catabolism, lipid, quinone and steroid biosynthesis, iron-sulfur (Fe/S) cluster biogenesis, apoptosis, and ion homeostasis, to name a few [1–5].

As our understanding of mitochondrial function in model systems has expanded, so too has our knowledge regarding the origins of mitochondria and their diversity in structure, metabolism and function across the eukaryote tree. In 1967, Lynn Margulis (then Lynn Sagan) famously published *On the Origin of Mitosing Cells* [6] in which she proposed that eukaryotic organelles including mitochondria and chloroplasts evolved from endosymbiotic bacteria, as had been proposed by others in the early 20th century [7]. Although her ideas were initially controversial, phylogenetic analyses of genes and proteins of these two organelles in the late 1970s [8,9] and early 1980s [10] confirmed that their prokaryotic provenance was distinct from the eukaryotic nuclear lineage. Since then, orders of magnitude more data have become available through the advent of high-throughput sequencing and proteomics technologies. The availability of hundreds of thousands of whole genome sequences has the potential to clarify the deepest relationships in the tree of life. Phylogenomic analyses have shown that the 'host' lineage of eukaryotes is most closely related to a newly discovered group of Archaea, known as the Asgards [11,12]. Modern analyses also confirm that the mitochondrial endosymbiont was indeed related to alphaproteobacteria [13], although controversy still persists as to which lineages within this group are their closest relatives [14]. Genomic and cell biology investigations of diverse protistan and multicellular lineages have further revealed that all known living eukaryotes descend from a mitochondrion-containing ancestor (the last eukaryote common ancestor — LECA) that had most of the genetic and cellular features of modern eukaryotes (Figure 1) [15–17].

Knowing that the mitochondrial compartment was once an endosymbiotic bacterium raises many fascinating questions. For example, what was the nature of the original symbiotic interaction between the alphaproteobacterial endosymbiont and the proto-eukaryotic host? How did the endosymbiont 'integrate' structurally, physiologically and genetically into the host and how did they coordinate their biogenesis and reproduction? What role did the mitochondrial symbiont have in the origin of the eukaryotic cell itself and how early did the symbiosis happen in the prokaryote-to-eukaryote transition? These questions have all been subjects of active research, theorizing and debate in the past few decades with critical discoveries coming from the disparate fields of biochemistry, molecular and cellular biology, genomics, microbiology and evolutionary biology.

Since the establishment of the integrated mitochondrial organelle, evolutionary divergence in mitochondrial form and

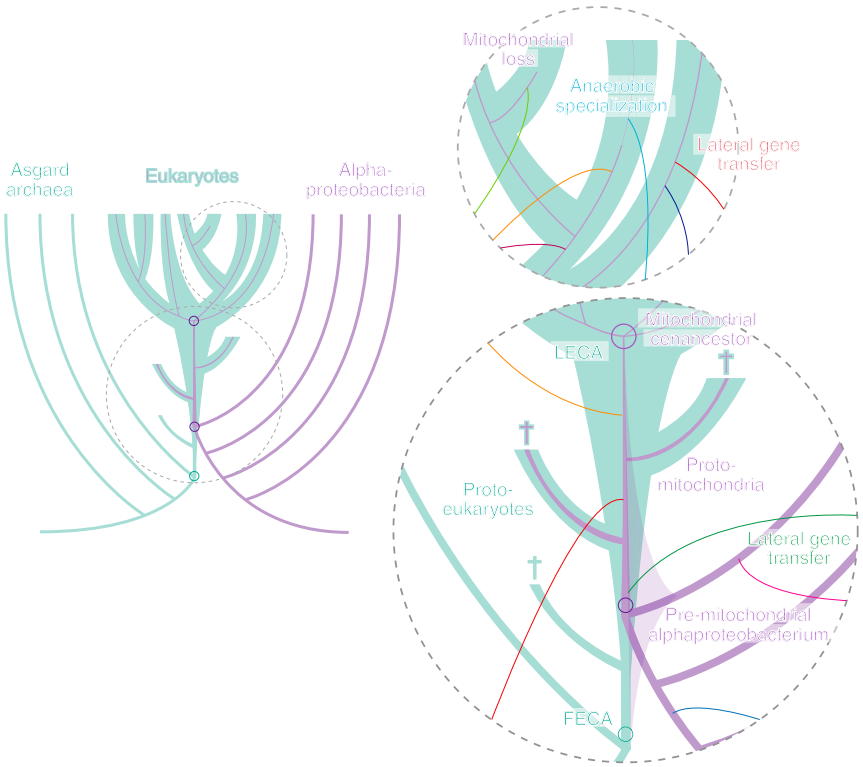


Figure 1. The origin and evolution of mitochondria and eukaryotes.
Mitochondria evolved from an endosymbiotic alphaproteobacterium (purple) within an archaeal-derived host cell that was most closely related to Asgard archaea (green). The earliest ancestor of mitochondria (that is not also an ancestor of an extant alphaproteobacterium) is the pre-mitochondrial alphaproteobacterium. Proto-mitochondria evolved from this first alphaproteobacterial endosymbiont, and comprise all transitional forms of mitochondria before the mitochondrial cenancestor, the mitochondrion in the last eukaryotic common ancestor (LECA). The timing of the mitochondrial endosymbiosis is uncertain (indicated by a purple shadow along the proto-eukaryotic stem) but postdates the first eukaryote common ancestor (FECA) and predates LECA. As far as we know, transitional 'proto-eukaryotes' between FECA and LECA went extinct (indicated by crosses). The complexity of the proto-eukaryotic genome and proteome gradually increased during eukaryogenesis (increasingly wider green branches), but the mitochondrial endosymbiont's genome and proteome were reduced, as the organelle incorporated proteins of host and foreign origin (progressively thinner purple branches for the mitochondrial endosymbiont contribution, with thin coloured branches indicating lateral gene transfers). Adaptations of mitochondria to anaerobiosis and outright loss of mitochondria (upper right circle) were facilitated by lateral gene transfer events.

function has continued along branches of the eukaryote tree of life. Studies of the mitochondria of diverse unicellular, multicellular, photosynthetic and anaerobic eukaryotes have overturned the essentialist textbook view of mitochondria as a single 'type' of organelle; mitochondrial genomes and proteomes differ substantially across eukaryotic diversity [18]. In lineages of eukaryotes adapted to low oxygen conditions, mitochondria have been drastically reduced, functionally altered and, in one case, completely lost [19]. Gene duplication, re-targeting of proteins to and from mitochondria, secondary gene loss and lateral gene transfer (LGT) have all played important roles in generating this diversity in mitochondria, although the relative importance of these mechanisms is debated [20–23].

Here we review the evolutionary origins and diversity of mitochondria across the eukaryotic tree and discuss the mechanisms and evolutionary forces that have shaped these diverse organelles. We highlight points of consensus and areas of controversy that new data will be especially helpful to resolve.

The Nature of the Pre-Mitochondrial Endosymbiont

In the following discussion, we refer to the first common ancestor of mitochondria as the 'pre-mitochondrial alphaproteobacterium' (Figure 1). As discussed below, it is unclear whether or not this organism was, itself, an endosymbiont or free-living. In contrast, 'proto-mitochondria' will refer to all intermediate or transitional forms that evolved, and diversified, from the pre-mitochondrial alphaproteobacterium on the lineage leading to the last eukaryote common ancestor (LECA). It was during this phase of evolution that the drastic reduction from an endosymbiotic alphaproteobacterium to a fully integrated organelle took place. This transition entailed thousands of changes to the original symbiont- and host-derived genomes and compartments. The mitochondrion of LECA, herein referred to as the 'mitochondrial cenancestor', was a fully integrated organelle in the eukaryotic cell, capable of aerobic respiration as well as dozens of the other biochemical functions performed by modern aerobic mitochondria [18,24,25].

The properties of the pre-mitochondrial alphaproteobacterium are important to clarify the nature of the initial symbiosis (Figure 1). We can gain insight into this question by reconstructing the last common ancestor between extant alphaproteobacteria and the mitochondrial lineage (e.g., see [24]). But to do this, we need to confidently place the mitochondrial lineage in a phylogenetic context. Unfortunately, it has been difficult to reliably determine the precise alphaproteobacterial lineage that is most closely related to mitochondria (Box 1). Phylogenetic analyses of large sets of genes often show mitochondria as emerging either within, or as a sister group to, the Rickettsiales (e.g. [13,24]), a group containing exclusively intracellular parasitic and endosymbiotic bacteria (Figure 2). This position may not be reliable because genes of mitochondrial origin are often highly divergent and could be artefactually grouping in phylogenies with the similarly divergent genes of Rickettsiales (Box 1). Other analyses have found affinities of mitochondrial sequences to various other alphaproteobacterial groups [26,27], or suggest that mitochondria may form a very deep independent branch in the alphaproteobacterial tree (as found in some analyses in [14]) (Figure 2).

A great many phenotypes have been imagined for the first mitochondrial endosymbiont. These proposals have been made by more or less taking into consideration: (a) the modern

Box 1. Difficulties in inferring the phylogenetic placement of mitochondria.

Phylogenetic analyses of the first full-length small subunit rRNA sequences from many mitochondria [10] clearly indicated their alphaproteobacterial origins and distinctness from the nucleocytoplasmic lineage. As data for single genes accumulated, the resulting trees frequently recovered a specific affinity between intracellular parasitic alphaproteobacteria like *Rickettsia* (order Rickettsiales) and mitochondria [110]. More recently, the mitochondria–Rickettsiales relationship has been frequently recovered using concatenations of genes from complete genomes [42,111–114]. Building trees from multi-gene datasets aims to make statistically sound inferences by increasing the amount of data analyzed simultaneously. When genome data from the order Pelagibacterales (also known as SAR11; Figure 2) were incorporated into phylogenetic analyses, they branched either as sister to a mitochondria–Rickettsiales clade [114] or as the immediate sister group to mitochondria [115,116]. Later analyses suggested these phylogenetic hypotheses were actually artefacts caused by the convergent similarities between mitochondrial genomes and the streamlined genomes of Pelagibacterales [14,117,118]. The most recent analyses of large concatenated gene data sets have relied on both mitochondrion-encoded and nucleus-encoded mitochondrial genes [13,24]. These analyses, which include novel and slower-evolving genomes from protistan endosymbionts from the family Holosporaceae, recover mitochondria emerging as a group within the Rickettsiales, being sister to the Rickettsiaceae, Anasplamataceae and Midichloriaceae [13,24].

The genomes of mitochondria, Pelagibacterales and Rickettsiales share several features. All three have evolved rapidly and in a reductive fashion by losing many genes, and have ended up being heavily compositionally biased (i.e. enriched in A+T and proteins enriched in amino acids specified by A+T-rich codons) relative to all other alphaproteobacteria [14]. This led to concerns about their possible artefactual phylogenetic attraction in trees (i.e. the so-called 'long-branch attraction' artefact coupled with convergent compositional biases), especially in those made from multiple proteins that together might exacerbate systematic errors in phylogenetic inference [119]. When the model of protein evolution does not adequately capture the complexities of sequence evolution, the wrong tree topologies can end up being strongly supported. This systematic error is especially problematic when there are extremely fast-evolving sequences separated from related sequences by short branches on phylogenetic trees [119,120]. This means that the consistent affiliation of mitochondria to Rickettsiales and/or Pelagibacterales in multi-gene trees could be the outcome of overall convergent evolution at the genome level, and does not reflect historical relationships [14].

As an alternative to concatenated analyses, large sets of individual mitochondrial genes have been analyzed to determine which specific alphaproteobacterial taxa are closest relatives to mitochondrial homologs in terms of sequence similarity or in single gene phylogenies [3,26,121,122]. Network analysis is another alternative way to assess phylogenetic affinities [123]. Networks can be inferred from either sets of single gene phylogenies or directly from concatenated multi-gene datasets. Instead of producing a strictly bifurcating tree, they output a network that displays some of the conflicting phylogenetic signals. Single gene and network analyses both show that there is a heterogeneous phylogenetic signal amongst mitochondrial genes. For example, it was shown that many mitochondrial genes tend to be phylogenetically closer, or are more similar, to genes of *Rhodospirillum* (Rhodospirillales) [27] or *Ochrobactrum* (Rhizobiales) [26], respectively, rather than to genes from Rickettsiales. This apparent mixed phylogenetic signal has been interpreted as either: (a) statistical noise because of little signal (due to ancient divergences), (b) artefactual signals and/or (c) a chimeric ancestry (i.e. gene acquisition in alphaproteobacteria prior to, or since, the mitochondrial symbiosis). Unfortunately, single gene analyses not only suffer from the systematic errors afflicting concatenated analyses, but are also particularly prone to stochastic error and lack of signal because of the limited data available for each gene. It is unclear if the apparent heterogeneity in signal amongst genes reflects true incongruence amongst markers, versus systematic or random error.

Ultimately the solution likely lies in using probabilistic gene tree-species tree reconciliation analyses of large sets of symbiont-derived genes that allow 'vertical' phylogenetic signals to be estimated even in the presence of lateral gene transfers (LGTs), gene duplications and loss events [124]. However, these methods are still in early phases of development and they do not yet employ the most sophisticated sequence evolution models. Therefore, concatenated gene analyses using sophisticated phylogenetic models are currently still the best option, especially if genes are carefully selected to have slower evolutionary rates and low compositional bias and to be robust to LGT. Most importantly, the inclusion of data from 'environmental' alphaproteobacteria recently identified through metagenomic studies (e.g., [44]) not only has the potential to improve taxonomic sampling, but could also yield new candidates for the closest relatives of mitochondria.

capabilities of mitochondria, (b) the phylogenetic affiliations of mitochondria, (c) the (hypothetical) nature of the host cell, (d) the nature of the interaction between host and endosymbiont, and (e) the environmental setting of the early phases of the mitochondrial endosymbiosis. As there is great uncertainty about all these criteria, all of the resulting scenarios are highly speculative.

Several proposals envision an aerobic respiring heterotroph. The endosymbiont would have provided an advantage to the host, usually envisaged to be an anaerobic fermenter, by secreting ATP [28], removing fermentation waste [29,30], serving as a methane sink [31,32] or removing toxic oxygen from within [33]. Alternatively, the first endosymbiotic mitochondrial ancestor was suggested to have been a biochemically versatile facultatively aerobic photosynthetic bacterium. It would have been useful to its host by allowing it to either: move to aerobic niches [34], oxidize sulfide produced by host respiration [35], release hydrogen from fermentation [36], or by secreting organic photosynthate [37,38]. As an alternative to single-resource-based syntrophy scenarios, some have argued that the complexity of the biochemical properties of the mitochondrial cenancestor indicates that the metabolic association between endosymbiont and host was multifaceted [39]. Others have

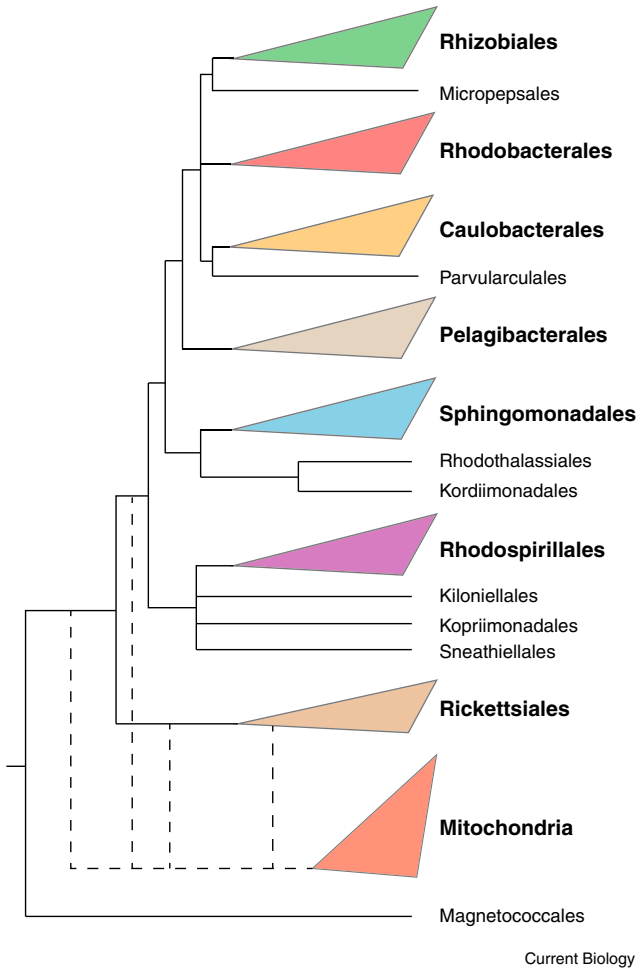


Figure 2. The phylogenetic position of mitochondria among alphaproteobacteria remains contentious.
The class Alphaproteobacteria encompasses well-defined diverse orders: the Rhizobiales, Rhodobacterales, Caulobacterales, Sphingomonadales, Pelagibacterales (SAR11), Rhodospirillales, and Rickettsiales [5,35] (the order Magnetococcales [36] is a distant sister to all other alphaproteobacteria). Some recently proposed candidate orders with sole or few representatives are also depicted (see [129]). The mitochondrial lineage could be placed at the base of Alphaproteobacteria, as sister to all 'free-living' alphaprotebacteria (e.g., [14]) as sister to the Rickettsiales or within the Rickettsiales (e.g., [13]); all positions are shown with dashed lines. Alphaproteobacteria are incredibly diverse. The Rhizobiales include plant-associated nitrogen-fixing rhizobia, facultative intracellular parasites as well as methanotrophs. The order Rhodobacterales encompasses purple non-sulfur bacteria, as well as abundant aerobic oceanic phototrophs and diverse heterotrophs. Some of the most abundant bacteria in the ocean are the small heterotrophic pelagibacterales. The Rickettsiales is composed exclusively of obligately intracellular endosymbionts or parasites. Phototrophs are found among the Rhizobiales, Rhodobacterales, Caulobacterales, Sphingomonadales, and Rhodospirillales.

proposed non-mutualistic scenarios in which the pre-mitochondrial alphaproteobacterium was a bacterial periplasmic predator (like *Bdellovibrio* or *Micavibrio*) that aggressively invaded its host [40,41], or an intracellular parasite of eukaryotes like some members of the Rickettsiales.

The phylogenetic placement of mitochondria, although still controversial, can help to constrain these speculations (Figure 2 and Box 1). Hypotheses in which mitochondria branch as sister to, or within, Rickettsiales have suggested the mitochondrial symbiont was initially a facultative intracellular energy parasite that invaded ancestral proto-eukaryotic cells [24]. These ideas also suggest it was a flagellated cell with a repertoire of terminal cytochromes adapted to hypoxia; a bacterium perhaps not too different from holosporacean protist endosymbionts [13,24,42]. However, if mitochondria emerge from a deeper position within alphaproteobacteria [14], it is then more likely that the pre-mitochondrial alphaproteobacterium was free-living. If so, it may have been a facultative anoxygenic photosynthesizer [38], as this physiology might have been ancestral to all alphaproteobacteria [43].

All of these proposals are based on what is currently known about alphaproteobacterial diversity. Metagenomic studies of aquatic and terrestrial environments worldwide are revealing a vast diversity of novel prokaryotic lineages, dwarfing what was previously known [44]. Explorations of the diversity and phylogeny of novel environmental alphaproteobacteria and the characterization of their physiologies could greatly enhance our understanding of the evolution of physiological traits within the group, allowing us to better pinpoint the nature of the pre-mitochondrial alphaproteobacterium.

The Transition From an Endosymbiotic Alphaproteobacterium to an Organelle

Despite uncertainty over the nature of the initial endosymbiosis, it is clear that mitochondria were ultimately retained in large part because of their capacity to efficiently generate ATP through aerobic respiration. The capability to generate ATP by fully oxidizing organic 'food', carbohydrates, amino acids and lipids, through aerobic respiration may well have been a new physiological property brought to the host by the mitochondrial endosymbiont. The transformation of a bacterium into an organelle was then effectively a process of integration with the host as the endosymbiont lost its autonomy and eventually became specialized as an aerobically respiring ATP-producing organelle with additional roles in a multitude of metabolic and biosynthetic pathways.

The transition from an autonomous endosymbiotic alphaproteobacterium to the mitochondrial cenancestor entailed many evolutionary changes including: (1) insertion of small molecule transporters/carriers into the endosymbiont inner membrane, (2) origin and elaboration of the protein-import machinery, (3) genome reduction through loss of redundant or unnecessary genes, (4) endosymbiotic gene transfer (EGT) to the nucleus, (5) modification of the endosymbiont cell envelope, (6) integration of biochemical pathways and systems between host and symbiont, (7) origin of an organelle division mechanism that was coordinated with the host-derived cell cycle, (8) specialization of cristae, (9) evolution of contact sites between proto-mitochondria and the endomembrane system, (10) retargeting of proteins of diverse origins (and localizations) to mitochondria, and (11) evolution of anchors between mitochondria and the cytoskeleton. Given the number and complexity of the changes involved, the evolutionary transformation of the pre-mitochondrial alphaproteobacterium into the mitochondrial cenancestor was *necessarily* incremental and produced many transitional forms long extinct (Figure 1). Below we discuss several of these evolutionary changes in more detail before discussing the origins of the mitochondrial proteome as a whole.

Evolutionary narratives often focus on the origin of the mitochondrial protein import machinery and the mitochondrial envelope as the target of the first and most important changes in organellogenesis. One view proposes that the mitochondrial protein import machinery first evolved to insert solute carriers into the proto-mitochondrial inner membrane [38,45]. These carriers are responsible for the movement of small molecules such as anionic metabolites, amino acids, nucleotides (ATP, ADP) and inorganic ions across the mitochondrial inner membrane. Many inner membrane transporters are members of the eukaryotic 'mitochondrial carrier family' (MCFs) that may have evolved from a single ancestral carrier that was inserted into the proto-mitochondrial inner membrane [38,46,47]. The original carrier has often been proposed to be an ADP/ATP translocator that allowed the host to tap the symbiont's ATP supply [46–48] (but see [38] for an alternative proposal). Another general scenario instead suggests that the protein import machinery first evolved to insert host proteins into the proto-mitochondrial outer membrane [49]. If the latter were true, the selective advantage for the evolution of the protein import machinery may have been to gain control over the biogenesis of the proto-mitochondrial endosymbiont envelope.

In any case, as part of the integration of the nascent organelle, the proto-eukaryote had to control the symbiont-derived compartment's growth and division. Mitochondrial biogenesis fundamentally depends on protein import, and the incorporation of proteins and lipids into mitochondrial membranes. In modern mitochondria, nucleus-encoded mitochondrial proteins are targeted to mitochondria after they have been synthesized by cytosolic ribosomes [50,51]. Many mitochondrial proteins (~60% in some model system eukaryotes) have 10–100 amino acid amino-terminal positively charged amphipathic alpha-helical presequences (mitochondrial targeting sequences) that are essential for their import; the remainder use presequence-independent import pathways [51]. Most mitochondrial proteins are imported through the Translocase of the Outer mitochondrial Membrane (TOM) complex into the intermembrane space and are either: (1) inserted into the outer membrane by the Sorting and Assembly Machinery (SAM) complex, (2) folded and oxidized by the Mitochondrial Import and Assembly (MIA) machinery to remain in the intermembrane space, (3) transferred to the TIM23–PAM complex for the further translocation across the mitochondrial inner membrane into the mitochondrial matrix, or (4) directly inserted into the inner membrane by the TIM22 insertase [51]. Proteins that are further translocated into the mitochondrial matrix with the help of the TIM23–PAM complexes can remain in the matrix or be inserted into the inner membrane by the OXA complex. Inner membrane proteins encoded by the mitochondrial genome are also inserted by the OXA complex (reviewed in [51]). Amino-terminal mitochondrial targeting presequences are typically cleaved off imported proteins by a mitochondrial processing peptidase (MPP).

The mitochondrial protein import machinery evolved through modification of the existing alphaproteobacterial protein export and membrane protein insertion systems. The SAM, TIM23–PAM, OXA and MPP complexes (or subunits of them) have alphaproteobacterial homologs [50,52]. Most of these complexes acquired additional subunits with specific roles during proto-mitochondrial evolution. The origins of the TOM complex are less clear, although the TOM40 pore and the related VDAC outer membrane proteins may have evolved from a bacterial outer membrane β-barrel protein [50]. Although the functions of the SAM and OXA complexes are similar to their ancestral alphaproteobacterial roles, the TOM and TIM23–PAM complexes appear to have acquired their current functions during organellogenesis [50]. After the origin of the protein import machinery (or at least a rudimentary form of this system), many endosymbiont-encoded genes whose products were essential to organellar function could be transferred to the eukaryotic nucleus.

The import of lipids into mitochondrial membranes (and the assembly of outer membrane β-barrel proteins) depends on contact sites between the endomembrane system — mainly the endoplasmic reticulum (ER) — and mitochondria [53,54]. Once lipid precursors have been transferred from the ER to mitochondria, they can be modified into specialized lipids (e.g., phosphatidylethanolamine and cardiolipin) by mitochondria; some of these lipids can be transferred back to the ER and the rest of the cell [53]. The contact sites that mediate lipid transfer between the ER and mitochondria are made by the four-subunit ERMES complex, which functions both as a tether and a lipid transfer complex [53,54]. ERMES subunits have no known prokaryotic homologs, but it was likely present in LECA, although it has been lost in major eukaryotic lineages like metazoans and green plants [55]. The loss of ERMES is likely compensated for by alternative tethers between the endomembrane system and mitochondria like EMC or vCLAMPs [53,55].

In addition to organelle biogenesis, the proto-eukaryote had to ensure the segregation of the mitochondrial compartments into daughter cells during cell division, and to control mitochondrial distribution throughout the cell. The pre-mitochondrial alphaproteobacterium divided using a contractile Z ring composed of polymerized FtsZ protein. The location of this ring was controlled by a mutually antagonistic system of Min proteins, and, based on other bacterial division systems, a suite of proteins anchoring the ring to the cell membrane. Of these components, the mitochondrial cenancestor retained at least the FtsZ protein (which underwent duplication prior to LECA), and the three Min proteins [56]. In extant eukaryotes, these proteins are all encoded in the nucleus. In addition, it had acquired an external dynamin ring that aided constriction at the mitochondrial mid-point [57]. Eukaryotic dynamins are also involved in vesicle fission; they may have diversified from an ancestral bacterial gene either present in the proto-eukaryote host (pre-endosymbiosis), inherited from the pre-mitochondrial alphaproteobacterium, or acquired by LGT during proto-eukaryotic evolution. The ER and actin cytoskeleton may also have been recruited during stem proto-eukaryotic evolution to aid mitochondrial division externally [58–60]. Some eukaryote groups, such as opisthokonts and plants, further lost all remnants of the alphaproteobacterial division system [56], and now rely only on an external dynamin division ring and, likely, unknown internal factors [61].

Mitochondrial cristae (inner membrane invaginations specialized for aerobic respiration) also evolved during organellogenesis. Cristae increase the surface area for housing large numbers of respiratory complexes, improving the efficiency of aerobic respiration. Two main factors are responsible for the development of cristae. First, multimers of ATP synthase complexes arranged along crista membranes bend them, and are thus largely

responsible for cristae morphology. This dimerization and cristae-bending capability of the ATP synthase (as a consequence of subunits *e* and *g*) appears to have occurred in the proto-mitochondrial phase of evolution, as these subunits are absent in alphaproteobacteria [62]. The second factor is the Mitochondrial Contact Site and Cristae Organizing System (MICOS). This multi-protein complex creates small 'necks' (crista junctions) that compartmentalize cristae and anchor them to the mitochondrial envelope. The core Mic60 subunit of MICOS has an alphaproteobacterial homolog, but the entire complex expanded in subunit composition prior to the cenancestral mitochondrion [63,64]. The pre-endosymbiotic origin of core components of MICOS suggests that the respiratory cristae of mitochondria could have evolved from the bioenergetic membrane invaginations known amongst alphaproteobacteria [43].

In modern eukaryotes, metabolic pathways and biosynthetic systems in mitochondria and the cytosol are coordinated through membrane transporters and redox/metabolite shuttles. One such biosynthetic system, the iron-sulfur (Fe/S) cluster biogenesis machinery, is a notable example of the proto-eukaryote nucleocytoplasm having become obligately dependent on the symbiont-derived system. Almost all modern eukaryotes share a conserved 'iron-sulfur cluster' (ISC) system in the mitochondrial matrix comprising 18 proteins [65], the majority of which have alphaproteobacterial origins [66]. The ISC system not only serves to synthesize Fe/S clusters and attach them to mitochondrial apoproteins, but it is also essential for the synthesis of Fe/S clusters in cytosolic and nuclear Fe/S proteins involved in key pathways (e.g. ribosome assembly and function, nuclear DNA replication and repair) [65]. The CIA system, responsible for cytosolic and nuclear Fe/S cluster biogenesis, depends on an unknown sulfur-containing factor produced by the ISC system that is transported across the inner mitochondrial membrane by Atm1, an ABC transporter of alphaproteobacterial origin [66]. As a result of its critical role, ISC is the only known mitochondrial biosynthetic pathway that is essential in yeast [65] and is a highly conserved system across eukaryotic diversity [66].

Most of the foregoing systems evolved during the integration phase of proto-mitochondrial evolution and have, at their cores, alphaproteobacterial molecular systems. In most cases, however, these original systems were greatly 'complexified' by the addition of protein subunits, while their functions were dramatically altered during organellogenesis. This pattern of mixed evolutionary origins applies more generally to entire mitochondrial proteomes.

Origins of Mitochondrial Proteomes

Mitochondrial proteomes are inherently chimeric [18,25,39,67] and differ substantially in protein content amongst eukaryotic groups [25,68,69]. Although they typically consist of ~1,000 proteins [1–5], the number of proteins of endosymbiotic origin in mitochondrial proteomes is surprisingly low. Latest estimates suggest that only 10–20% of proteins in mitochondria show alphaproteobacterial affinity [24,68]. An additional 20–30% of mitochondrial proteomes have phylogenetic affiliations more generally to proteobacteria [25,68]. Many of these could be proteins of true mitochondrial endosymbiotic origin that have lost their alphaproteobacterial signature, although some could easily be independent LGTs. A large number of mitochondrial proteins (~40% of mitochondrial proteomes) have no known prokaryotic or viral homologs [25]. Many of the genes encoding these proteins originated in the proto-mitochondrial phase before the cenancestral mitochondrion and the diversification of modern eukaryotes [18], and their specific origins are unclear. However, a significant proportion of mitochondrial proteins with no detectable prokaryotic homologs are 'lineage-specific' and likely evolved in specific eukaryote groups after LECA [18,25]. The remaining proteome fraction (~15%) has prokaryotic, non-proteobacterial affinities [25]. This fraction likely includes genes that had been laterally transferred to the pre-mitochondrial endosymbiont before endosymbiosis, genes of archaeal ancestry in the proto-eukaryotic host and genes that had been laterally transferred from bacteria or viruses to the proto-eukaryote nucleus before, or after, the initial mitochondrial endosymbiosis. Regardless, during organellogenesis the proto-mitochondrial compartment progressively lost its alphaproteobacterial identity through loss of genes as well as the acquisition and creation of new genes in the proto-eukaryotic genome whose products were targeted to the organelle [39,67,68,70].

Many alphaproteobacterial-derived proteins (encoded in both mitochondrial and nuclear genomes) serve direct or indirect roles in aerobic respiration [24,68]. These proteins take part in (1) the ETC that conserves energy through chemiosmosis to make ATP, (2) the mitochondrial ribosome that supports the translation of genes encoded in the mitochondrial genome, many of which encode ETC components, (3) the Krebs cycle that feeds reduced cofactors (NADH and $FADH_2$) to the respiratory chain, (4) the oxidative decarboxylation of pyruvate that feeds acetyl-CoA into the Krebs cycle, (5) the β-oxidation pathway for fatty acids that provides NADH for the respiratory chain and acetyl-CoA to the Krebs cycle, (6) the biosynthesis of cofactors (Fe/S clusters, heme and biotin) that are required for the assembly of many proteins of the respiratory complexes and other mitochondrial enzymes, and (7) the biosynthesis of cardiolipin and ubiquinone, which are essential for the proper function of the respiratory chain. Interestingly, for the foregoing systems involving multi-protein complexes, the central core subunits are of alphaproteobacterial descent (e.g., respiratory complexes, ribosomes, translocons, and the MICOS complex) [24,68]. Eukaryotic-origin proteins frequently surround the alphaproteobacterial cores of these multi-protein complexes. A large proportion of mitochondrial proteins of eukaryotic origin also function in the mitochondrial inner and outer membranes (e.g., protein import, metabolite transport, organelle division, etc.), close to the interface between symbiont-derived and host-derived compartments [25,68], reflecting adaptations related to host–symbiont integration and coordination.

The pre-mitochondrial alphaproteobacterium appears to have contributed more genes to eukaryotes than just those whose products have specific mitochondrial functions. Some of the genes transferred to the proto-eukaryote 'nuclear genome' from proto-mitochondria now have roles elsewhere in the cell. The latest analyses indicate that 30–40% of nuclear genes of clear alphaproteobacterial origin are not functioning in mitochondria of that organism [24]. Many of these are mitochondrion-targeted proteins in some organisms, but targeted to the

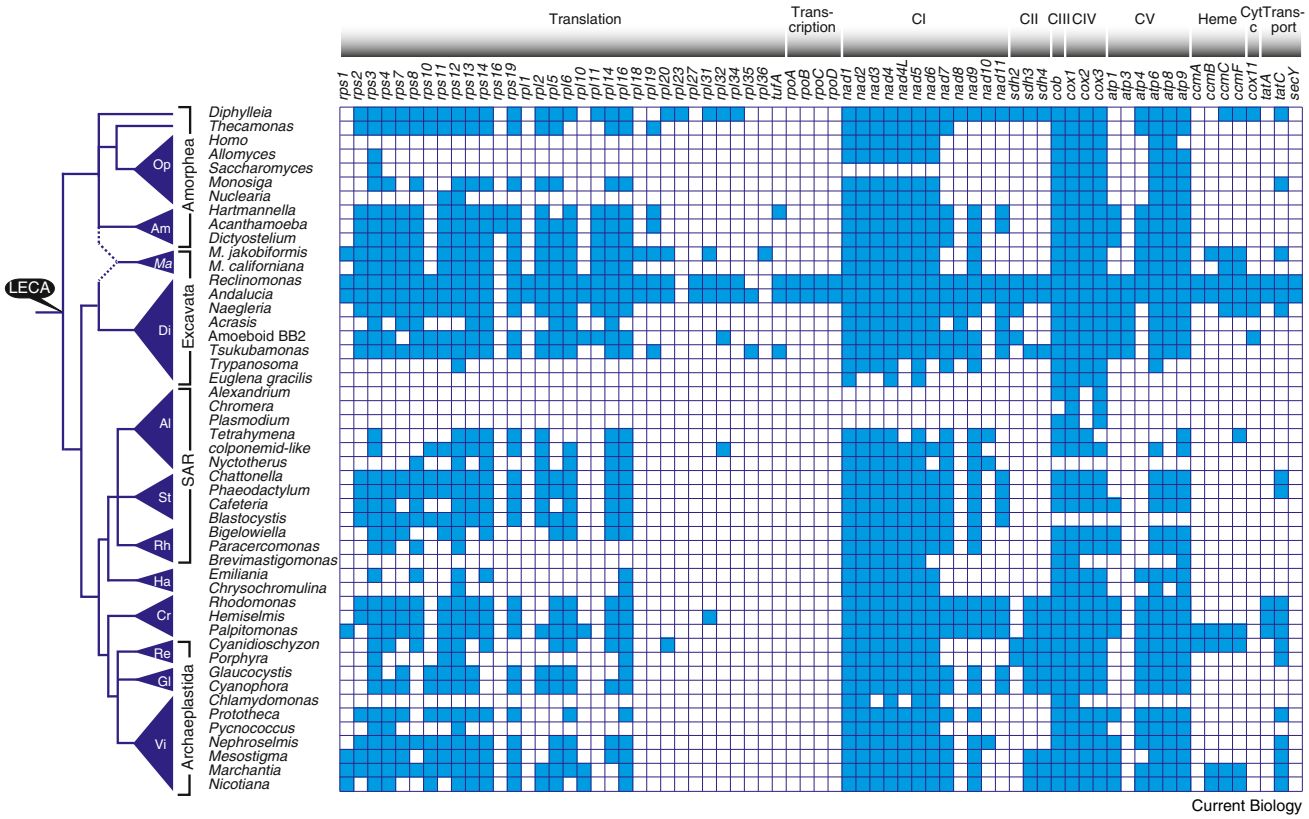


Figure 3. Protein-coding gene content in mitochondrial genomes across eukaryotic diversity.
Presence or absence of genes in mitochondrial genomes of various eukaryotes is shown by closed and open boxes, respectively. Phylogenetic relationships of eukaryotes are depicted, as summarized by [87,130]. Abbreviations are Ma: *Malawimonas*; Op: Opisthokonta; Am: Amoebozoa; Di: Discoba; Al: Alveolata; St: Stramenopiles; Rh: Rhizaria; Cr: Cryptophyceae; Ha: Haptophyta; Re: Red algae (Rhodophyceae); Gl: Glaucophyta; Vi: Viridiplantae; CI–CV: electron transport chain complexes I–V.

cytosol or other organelles (e.g. peroxisomes [69]) in other eukaryotes and were likely in the mitochondrial cenancestral proteome [18]. Others seem to have assumed non-mitochondrial roles prior to LECA. Examples of the latter include enzymes involved in sterol and glycosphingolipid biosynthesis in the ER [24].

Evolution of Mitochondrial Genomes

Mitochondrial genomes are vastly reduced in gene content and simplified compared to the genomes of their alphaproteobacterial relatives. The content of complete genomes of alphaproteobacteria sequenced so far range from 800 to 8000 genes [71] with their common ancestor having ~3000 [72]. In sharp contrast, comparisons of diverse mitochondrial genomes suggest that 69 different conserved protein-coding genes and a full set of tRNA and ribosomal RNA genes were present in the genome of the mitochondrial cenancestor (Figure 3) [71]. Although the mitochondrial cenancestral genome likely encoded a few more proteins, it is still a miniscule fraction of modern mitochondrial proteomes.

From the foregoing it should be clear that hundreds (if not thousands) of genes were lost from the endosymbiont genome during the 'proto-mitochondrial' phase of evolution. The reductive evolutionary process likely started once the mitochondrial symbiont was no longer capable of replicating outside of the host cell. The confinement to host cells reduced the symbiont population size leading to the increased fixation of slightly deleterious mutations [73,74]. Inevitably, this resulted in an increase in rates of sequence evolution and increased A+T nucleotide composition and led to loss of non-essential genes. Similar reductive trajectories are well documented for genomes of insect endosymbionts [74], obligate intracellular parasites [73] and the cyanobacterial symbiont in *Paulinella chromatophora* [75]. Mitochondrial genomes could eventually reduce even further, as the advent of the protein import system allowed many essential genes for symbiont function to be transferred to the 'host' genome. A prerequisite for the subsequent loss of these genes from the organellar genome was the faithful targeting of the host-encoded copies to the organelle.

Mitochondrial genome reduction and EGT have continued since the mitochondrial cenancestor on diverse branches of the eukaryote tree. The largest mitochondrial gene contents are found amongst the jakobid flagellates whose mitochondrial genomes encode up to 66 identifiable protein genes [71]. Curiously, mitochondrial genomes of other eukaryotic lineages generally have gene repertoires that are subsets of those found in jakobids, suggesting that reductive evolution has slowed in these protists. Indeed, jakobid mitochondrial genomes are

Box 2. Why do mitochondria retain genomes?

The fact that so many genes from the mitochondrial symbiont were transferred to the nuclear genome raises the question of why mitochondria retain a genome at all. Of all the explanations for organellar gene retention proposed, two hypotheses seem most credible given current data [125].

The Co-location for Redox Regulation (CoRR) hypothesis [126] posits that genes for particular electron transport chain (ETC) components must be retained in the mitochondrial compartment to allow for rapid organelle-specific gene expression regulation of components of the ETC. The argument is that single organelles (out of potentially many in a cell) must be able to rapidly adjust the expression of ETC genes in response to sensing its redox state. Without a rapid regulatory response to the redox state of a particular organelle, the ETC would cease to function efficiently and generate damaging reactive oxygen species, providing a selective advantage for retention of the genes on the organellar chromosome. A second explanation [127] holds that many of the proteins that are frequently retained on organellar genomes cannot be expressed in the cytosol because they are large hydrophobic proteins with multiple transmembrane domains that would preferentially be mistargeted to the ER. Specifically, when 12 ETC components frequently encoded on mtDNA are expressed in the nucleus, they have been shown to localize to the ER in human cells, even when fused with a canonical mitochondrial targeting peptide [127]. Mislocalization provides a likely barrier to their evolutionary transfer to the nucleus.

Both of the foregoing hypotheses are supported by observations that mitochondrial genome-encoded ETC components, especially those that are most centrally located in complexes, are amongst the least likely to be transferred to the nucleus (Figure 3) [128]. It should also be noted that these two hypotheses are not mutually exclusive; the products of genes that are most often retained could *both* be difficult to retarget to mitochondria *and* be critical regulatory points for ETC function. Finally, neither hypothesis addresses the fact that the mitochondrial small and large subunit rRNA genes are always retained on organellar chromosomes. This could be in part because large structured RNAs are difficult to translocate across membranes. Because of the potential for mis-assembly, there may also be selection against having rRNAs expressed in the cell compartments where other rRNAs (host-derived) and ribosomal proteins (host and symbiont) are expressed.

unique amongst eukaryotes in encoding a bacterial-type multi-subunit RNA polymerase (all other eukaryotes have a nucleus-encoded phage-type mitochondrial RNA polymerase [76]) and the ancestral alphaproteobacterial SecY protein translocator [71].

Most eukaryotes possess several dozen genes on their mitochondrial genomes. The core genes on mitochondrial genomes conserved across many eukaryotes encode ETC components (e.g. subunits of complexes I, III, IV and V) and translation (tRNAs and rRNAs). Other genes, such as those encoding ribosomal proteins, complex II, heme maturation enzymes, cytochrome *c* oxidase assembly proteins and the translation elongation factor *tufA*, are much more patchily distributed (Figure 3). The surprisingly large differences in mitochondrial gene contents across eukaryotic diversity are the result of multiple events of EGT in different lineages, which sometimes relocate the same genes, in parallel, to the nucleus [77]. The most extreme reduction in gene content is found in the enigmatic coral-associated photosynthetic protist *Chromera velia*, a relative of apicomplexan parasites. The *Chromera* mitochondrial genome consists of heterogeneous linear molecules that apparently encode only two ETC component proteins and fragmented rRNAs [78].

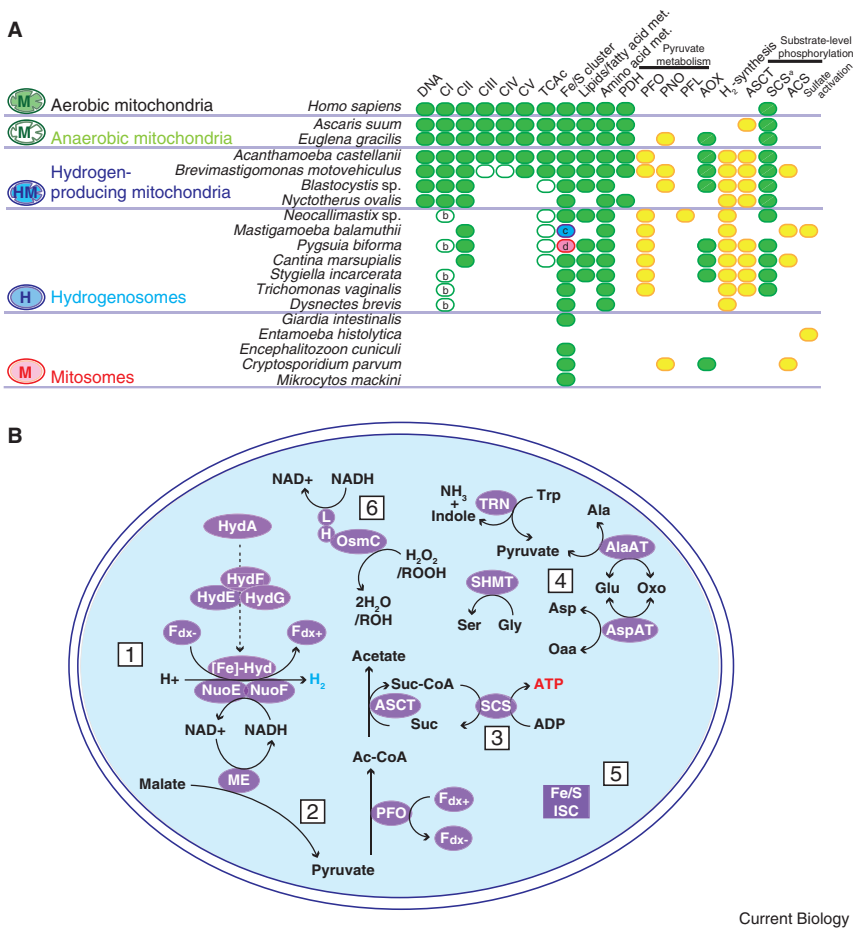
In some eukaryotic lineages, mitochondrial genomes have gone 'wild'. Although many mitochondrial genomes are single circular mapping chromosomes, there are stranger forms including multiple tiny linear chromosomes, mini-circles or, as in some land plants, giant chromosomes full of laterally acquired DNA (reviewed in [74,79,80]). In many cases standard genetic coding 'rules' have changed as well, with multiple independent appearances of *trans*-splicing and RNA editing systems required to make proper coding transcripts and changes to the genetic code itself [80]. Although different evolutionary mechanisms may be implicated in the various peculiarities of mitochondria of different eukaryotic lineages, much of this diversity likely stems from neutral changes in genomes as a consequence of a reduced population size, high mutation rates and small coding requirements.

The reasons for the retention of mitochondrial genomes are debated. The potential for mistargeting of ETC components and/or the need for efficient redox regulation may explain why some genes cannot be relocated to the nucleus (Box 2). In any case, some anaerobic protists *have* completely lost their mitochondrial genomes. Most such organisms retain organelles that have lost oxidative phosphorylation, but still carry out some ancestral mitochondrial functions as well as newly acquired biochemical capacities.

Diversity of Mitochondrial Functions in Anaerobic Eukaryotes

Many distinct eukaryotic lineages have adapted to living in low-oxygen conditions in aquatic and terrestrial environments or in animal gastrointestinal tracts. These conditions pose a problem for aerobic respiration, and, as a result, many of these organisms have evolved mitochondria with reduced, or no, cristae that function anaerobically. Müller and colleagues [81] have classified mitochondrion-related organelles (MROs) into five types based on their energy metabolism: aerobic mitochondria, anaerobic mitochondria, hydrogen-producing mitochondria, hydrogenosomes, and mitosomes (Figure 4). However, newly discovered MROs in free-living protists indicate that these organelles represent more of a functional continuum than a discrete set of classes. Below we briefly review the diversity in anaerobically functioning mitochondria across the eukaryotic tree (Figures 4 and 5).

Facultatively anaerobic mitochondria occur in eukaryotes whose life-cycles either have an anaerobic phase (e.g. the parasitic worm *Ascaris*) or that transiently experience hypoxia



Current Biology

Figure 4. Energy metabolic functions of mitochondria of aerobic and anaerobic eukaryotes.

(A) Distribution of enzymes of aerobic and anaerobic energy metabolism in mitochondria (or MROs) of selected eukaryotes (see [21,81,85,87] and references therein). Presence and absence of corresponding enzymes in eukaryotes is shown by presence and absence of ellipses, respectively. Closed green ellipses indicate proteins typically found in aerobic mitochondria (but sometimes also in anaerobic mitochondria/MROs), open green ellipses indicate incomplete (or degenerate) sets of proteins for aerobic pathways, and closed yellow ellipses represent proteins for anaerobic metabolism. a: also involved in tricarboxylic acid cycle (TCAc), b: only NuoE and NuoF, c: NIF system (blue), d: SUF system (magenta). PDH: pyruvate dehydrogenase complex, PFO: pyruvate:ferredoxin oxidoreductase, PNO: pyruvate:NADPH oxidoreductase, PFL: pyruvate formate lyase, AOX: alternative oxidase, AcK: acetate kinase, SCS: succinyl-CoA synthase, and ASCT: acetate:succinyl-CoA transferase families 1b and 1c. (B) Reconstructed metabolic pathways in the hydrogenosomes of *Trichomonas vaginalis* [81,131]. 1. H_2-synthesis, 2. Pyruvate metabolism, 3. Substrate-level phosphorylation, 4. Amino acid metabolism, 5. Fe-S cluster assembly, 6. Detoxification. NuoE and F: 24 and 51 kDa of mitochondria NADH:ubiquinone oxidoreductase, respectively, Fe-Hyd: Fe-hydrogenase, HydE/F/G: Fe-hydrogenase maturases, Fdx: ferredoxin, ME: malic enzyme, SHMT: serine hydroxymethyltransferase, H/L: glycine cleavage system proteins H/L, OsmC: osmotically inducible protein, GDH: glutamate dehydrogenase, AlaAT: Alanine aminotransferase, AspAT: aspartate aminotransferase, Fe/S ISC: iron-sulfur cluster assembly ISC system.

(e.g. intertidal animals and the protist *Euglena gracilis*). Under high oxygen conditions their mitochondria respire aerobically but under low oxygen conditions, they shift to an ETC that uses an endogenously produced electron acceptor, such as fumarate, instead of oxygen (reviewed in detail by [81]). Many of these mitochondria carry out malate dismutation, a branched pathway with several unique features not found in obligately aerobic mitochondria including: firstly, a low redox potential quinone, rhodoquinone (instead of ubiquinone) that can be reduced by complex I, but that donates electrons for fumarate reduction by a complex II-related fumarate reductase; and, secondly, an acetate:succinate-CoA transferase (ASCT) enzyme that generates succinyl-CoA directly from acetyl-CoA, bypassing several enzymes of the Krebs cycle that are inhibited under reducing conditions. Under anaerobic conditions, malate dismutation produces ATP both by substrate-level phosphorylation by succinyl-CoA synthetase (SCS) and chemiosmotically by complex V (F_1F_o-ATPase).

The MROs of *Nyctotherus ovalis* [82] and *Blastocystis* sp. [83] are likely examples of hydrogen-producing mitochondria, although H2 production has yet to be demonstrated for the *Blastocystis* organelles. They have genomes and encode proton-pumping complex I subunits, but lack genes encoding complexes III to V of the ETC in both their organellar and nuclear genomes, suggesting that they do not make ATP chemiosmotically. Instead, they synthesize ATP by substrate-level phosphorylation via anaerobic pyruvate metabolism and H_2 production (Figure 4). Briefly, this pathway involves an oxygen-sensitive pyruvate:ferredoxin oxidoreductase (PFO) that oxidatively decarboxylates pyruvate to acetyl-CoA (and CO_2), an ASCT to generate acetate, and succinyl-CoA (from acetyl-CoA and succinate), and the SCS to generate ATP and succinate from succinyl-CoA. The electrons from the PFO-catalyzed reaction are transferred to a ferredoxin, which then passes them to an [Fe]-hydrogenase, ultimately reducing protons to H_2. The retention of an ETC, the lack of O_2 as the terminal electron acceptor and the presence of the anaerobic H_2-producing pathway is argued to be diagnostic of hydrogen-producing mitochondria [81]. However, it was recently shown that the amoeba *Acanthamoeba castellanii* has all the machinery for aerobic mitochondrial respiration but also possesses the full set of anaerobic pyruvate metabolizing/H_2-producing enzymes [84], blurring the distinction between anaerobic mitochondria and hydrogen-producing mitochondria. Interestingly, the MRO of the anaerobic cercozoan flagellate *Brevimastigomonas motovehiculus* also has the anaerobic pyruvate metabolism/H_2-producing pathway, but is in the early stages of losing its capacity for oxidative phosphorylation, as the genes encoding subunits of complexes III, IV and V are either degenerating, missing or fragmented [85] (Figure 4A).

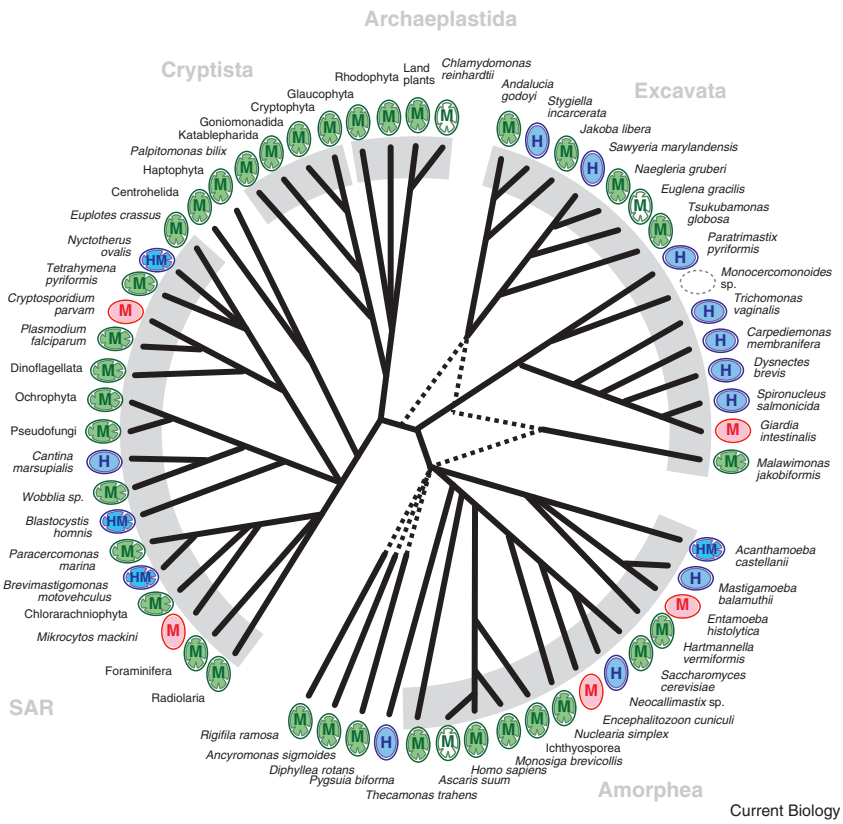


Figure 5. The phylogenetic distribution of mitochondria and mitochondrion-related organelles (MROs) in eukaryotes.
Phylogenetic relationships of eukaryotes are depicted as summarized in [87,130]. Each type of MRO (see Figure 4) is depicted next to the species name, except for *Monocercomonoides* sp., which lacks mitochondria completely [19].

Hydrogenosomes and mitosomes are much more reduced forms of MROs, and both kinds of organelles lack genomes [21,81]. Hydrogenosomes, such as those of the parasite *Trichomonas vaginalis*, generate ATP exclusively by substrate-level phosphorylation via anaerobic pyruvate metabolism and H_2 synthesis, and completely lack an ETC. This metabolic type of organelle appears to have evolved multiple times as hydrogenosomes occur in disparate parasitic and free-living lineages across the eukaryote tree, although they do vary substantially in functional capacity (Figure 4). Mitosomes, in contrast, do not produce ATP or H_2 at all. The main function of many mitosomes (e.g. those of *Giardia intestinalis* [86] and some microsporidia [66]) is Fe/S cluster assembly (Figure 4A). Recently, it was shown that the free-living anaerobic excavate flagellate *Dysnectes brevis* possesses a new type of H_2-producing organelle that is probably incapable of ATP synthesis [87]. Given *Dysnectes*' close phylogenetic relationship to *Giardia* and *Trichomonas* (Figure 5), its MRO provides a critical snapshot into how the *Giardia* mitosome and the *Trichomonas* hydrogenosome evolved reductively from a common ancestral organelle.

As alluded to above, the ISC system for Fe/S cluster biogenesis is an essential system conserved in most forms of mitochondria [66]. Interestingly, however, it has been lost in three separate lineages of anaerobic protists. The amoebic dysentery parasite *Entamoeba histolytica* and its free-living flagellate relative *Mastigamoeba balamuthi* lack any traces of the ISC system, but instead appear to make their Fe/S clusters with nucleus-encoded proteins related to the 'nitrogen-fixation' (NIF) Fe/S cluster system that was acquired by LGT from an epsilonproteobacterium [88,89]. Whereas *Entamoeba*'s NIF system appears to be cytosolic, *M. balamuthi* has duplicated genes encoding NIF subunits, with one set functioning in its hydrogenosomes and the other in the cytosol [89]. The breviate flagellate *Pygsuia biforma* is another free-living anaerobic flagellate that has replaced most of its ISC system; it has a simple archaeal-type 'sulfur mobilization' (SUF) Fe/S system that it acquired by LGT [90]. In a similar pattern to the NIFs of *M. balamuthi*, *P. biforma* has duplicated isoforms of SUF, one of which now functions in its MRO and the other in the cytosol. The final example is of the gut-inhabiting flagellate *Monocercomonoides* sp. that laterally acquired genes for a bacterial-type SUF system that now functions in its cytosol. The acquisition of this system not only allowed *Monocercomonoides* to lose its ISC machinery, but it also seems to have led to the outright loss of the mitochondrial compartment (its nuclear genome completely lacks any traces of mitochondrial genes) [19]. This is the only known case of complete loss of the mitochondrial compartment in an autonomous eukaryotic cell (Figure 5).

In summary, the adaptation of mitochondria to function in low oxygen conditions has occurred multiple times independently [21]. It likely proceeds in a stepwise fashion with mitochondria first gaining the ability to produce ATP both aerobically and anaerobically. Then, in adaptation to permanent hypoxia, there is progressive loss of components of the ETC and oxidative phosphorylation, and increasing reliance on anaerobic ATP production via substrate-level phosphorylation coupled with hydrogen production. The final stage of reductive evolution appears to be loss of ATP production exemplified by the mitosomes of parasites. Complete loss of mitochondria is only possible if the ISC system, essential for mitochondrial and cellular Fe/S cluster biogenesis, is replaced by a cytosolic system acquired by LGT [19].

Origins of Proteins Involved in Anaerobic Metabolism and MROs

To understand how anaerobically functioning mitochondria and MROs have evolved multiple times independently from canonical aerobic mitochondria (Figure 5), the origins of enzymes of anaerobic metabolism are especially relevant. So far, two scenarios have been proposed. The 'ancestral anaerobic mitochondria' view suggests that genes for anaerobic enzymes were

critical in cementing the initial mitochondrial–archaeal symbiosis and have been vertically inherited from a facultatively anaerobic pre-mitochondrial alphaproteobacterium [36,81,91]. In the alternative 'LGT origins' scenario, genes encoding enzymes of anaerobic metabolism were acquired more recently by various lineages of eukaryotes adapting to anaerobiosis in multiple instances of prokaryote-to-eukaryote and eukaryote-to-eukaryote LGT [21,22,92,93].

Deciding between these hypotheses is made difficult by the fact that these genes have been frequently transferred between bacterial groups and their phylogenies are sometimes poorly resolved [92]. Nevertheless, there are several reasons to prefer the LGT origins scenario. None of the enzymes exclusive to anaerobic metabolism in eukaryotes has clear phylogenetic affinities to alphaproteobacterial homologs [94]. A recent report that claimed an alphaproteobacterial origin for eukaryotic [Fe]-hydrogenase [95] is invalidated by the failure to include relevant non-alphaproteobacterial homologs in the analysis (see [94] for a more comprehensive analysis). For several of these enzymes, including [Fe]-hydrogenase and ASCT, eukaryotic homologs group in phylogenetic trees into multiple distinct subfamilies, each of which is most closely related to enzymes from different bacterial taxa. For these multiple eukaryotic isoforms to be ancestral to mitochondria, the premitochondrial alphaproteobacterium would have to have encoded many paralogous copies of the same enzyme, only to have them differentially lost in most descendant eukaryotic lineages, an unlikely scenario. In the trees of anaerobic enzymes in which eukaryote homologs do form a clade (e.g. PFO, pyruvate formate lyase [96] and hydrogenase maturases [94]), the phylogenies significantly conflict with known eukaryote relationships, suggestive of eukaryote-to-eukaryote LGT. Finally, anaerobic pyruvate metabolism/H_2-producing enzymes and their associated maturases are rare in alphaproteobacteria (currently only ~50 out of ~1,400 complete alphaproteobacterial genomes available in GenBank possess homologs) and are very patchily distributed, consistent with recent acquisition of these enzymes by LGT within subgroups of alphaproteobacteria [21,92]. Thus, they seem unlikely to have been ancestral to the pre-mitochondrial alphaproteobacterium.

In any case, some caution in interpretation is warranted as the 'donors' of these genes to eukaryotes are very difficult to assess; this is because the relationships amongst bacterial homologs are often phylogenetically scrambled as a result of frequent LGT [94]. Furthermore, it is possible that a 'mixture' of the two scenarios explaining their origins may be correct. For example, *some* of these genes could have been ancestral to eukaryotes, acquired prior to LECA by LGT from a bacterium unrelated to the mitochondrial endosymbiont, whereas others may be more recent acquisitions. Currently, however, there is no clear link between the origin of these anaerobic pyruvate metabolism/H_2-producing enzymes and the mitochondrial endosymbiont.

Mitochondria and the Origin of Eukaryotic Cells

Since the mitochondrial endosymbiosis took place prior to LECA, the relative timing of the major events of the prokaryote-to-eukaryote transition — for example the origins of the endomembrane system, nucleus, cytoskeleton, mitosis, sex/meiosis, some ~3,000 gene families, and mitochondria — cannot be easily resolved [17]. In the last few decades there has been a proliferation of hypotheses postulating different timings of the mitochondrial symbiosis that vary in their implications for the origin of other eukaryote cellular features. Two general kinds of hypotheses have received most attention. At one extreme, 'mitochondria-early' hypotheses suggest the mitochondrial symbiosis was amongst the first events in eukaryogenesis and even triggered the process [91]. Alternatively, 'mitochondria-late' hypotheses hold that the mitochondrial symbiosis occurred after many other features of modern eukaryotes had already evolved, either autogenously [38,45,97] or through additional symbioses [98]

The most widely publicized mitochondrion-early hypothesis, the 'Hydrogen Hypothesis' [36], holds that the initial mitochondrial symbiosis involved a facultatively anaerobic alphaproteobacterium in syntrophy with a hydrogen-consuming anaerobic autotrophic archaeon. Under low oxygen conditions, the alphaproteobacterium would ferment and generate H_2 and acetate that would be consumed by the archaeon. At higher oxygen tensions, the alphaproteobacterium would respire aerobically. Under this scenario, as the two partners became more metabolically integrated, the archaeon came to enclose the alphaproteobacterium and, ultimately, the latter became specialized as a bioenergetic organelle. Energetic scaling arguments have also been advanced to support mitochondria-early scenarios [99]. Briefly, the loss of energetically expensive non-essential genes in the mitochondrial endosymbiont genome coupled with increased surface area of bioenergetic membranes for aerobic respiration have been argued to provide a surplus of ATP that was required for the subsequent evolutionary experimentation that generated complex eukaryotic proteomes and cells [99].

Most 'mitochondria late' views, in contrast, hold that the eukaryotic endomembrane system and cytoskeleton evolved autogenously in the proto-eukaryote lineage primarily because of the advantages afforded by predatory feeding *via* phagocytosis [45,97]. After (or during) the origin of phagocytosis, the mitochondrial ancestor was engulfed and, after escaping the phagosomal membrane, became an endosymbiont in much the same way as the plastid likely evolved, and many modern endosymbioses within protists are established. Ultimately, the host tapped the symbiont's aerobic respiratory capacity to produce ATP, as the latter was integrated into the cell as an organelle.

Many arguments have been marshaled for and against these competing scenarios — too many to comprehensively review here (see [17,45,91,97,98]). However, several recent discoveries have shifted the frame of reference in which these hypotheses can be evaluated. Partial genomes from an entirely new major group of Archaea, the 'Asgard' taxa (related to the TACK archaea), have been recovered from metagenomic surveys of various marine and hot spring sediments [11,12]. Phylogenomic analyses of concatenated proteins place the Asgards as the closest known relatives of the eukaryote nucleocytoplasmic lineage. Furthermore, Asgard metagenomes encode more 'eukaryote signature proteins' (ESPs) than any previously discovered archaeal group. These include homologs of proteins involved, in eukaryotes, in the endomembrane system, vesicle trafficking and the cytoskeleton. Although the function of these proteins is currently unknown, it is possible that some or all of the Asgards (and perhaps their TACK relatives) have some kind of rudimentary endomembrane system and simple cytoskeleton [12]. Indeed, a recent paper demonstrates the existence of

intracellular membranous compartments, vesicles and filaments in the TACK archaeon *Ignicoccus hospitalis* [100]. If a primitive endomembrane system and cytoskeleton — or, at least, a number of proteins underpinning these structures — had already evolved in the common ancestor of eukaryotes and Asgards, then clearly their origin had nothing to do with the mitochondrial endosymbiont *per se*. However, it has been argued that the Asgard *Lokiarchaeum* sp. is an H_2-dependent autotroph, consistent with the predictions of the Hydrogen Hypothesis [101]. The situation is more complicated, however, as the Thorarchaeota (another Asgard group) appear to be heterotrophs that metabolize proteins scavenged from their environments [102], a physiology suggested for the proto-eukaryote host in Cavalier-Smith's autogenous account of eukaryogenesis [45]. In any case, the physiologies of present day Asgards appear to be quite variable [103]. As they are more than a billion years removed from their common ancestor, reconstructing the metabolism of the Asgard-eukaryote ancestor is expected to be difficult, especially given that metabolic genes are frequently laterally transferred.

An alternative way of assessing the timing of the mitochondrial symbiosis was recently proposed by Pittis and Gabaldón [104]. These authors assembled a subset of proteins inferred to be in LECA that have homologs in prokaryotes. By reconstructing phylogenies of each of these ~1,100 proteins, they evaluated the length of the 'stem' branch separating eukaryotic and prokaryotic sequences (branch lengths = evolutionary rate × time since divergence). After normalization for the gene-specific evolutionary rate, they found that the genes of archaeal origin tended to be the most distant from eukaryotes whereas genes of alphaproteobacterial origin (likely of mitochondrial provenance) were, on average, closest to eukaryotes. They also found a number of genes of bacterial origin where different bacterial groups were sisters to eukaryotes: these had an intermediate 'stem length' distribution between those of archaeal and alphaproteobacterial origin. The authors suggested the differences in average normalized stem lengths over sets of genes of a given origin should reflect their time since acquisition and therefore that the mitochondrial symbiosis took place more recently than the divergence of the eukaryote host lineage from Archaea, supporting 'mitochondria-late' scenarios [104]. This conclusion was challenged by Martin and colleagues [105] and Pittis and Gabaldón have responded [106]. Martin and colleagues questioned the statistical methods used and the quality of the data sets assembled; they argue, once more stringent filters are applied, the differences in stem-lengths between the bacterial non-alphaproteobacterial origin proteins and the alphaproteobacterial origin proteins are not statistically significant. Curiously, they do not show that the same is true for the archaeal versus alphaproteobacterial comparisons. In any case, they further suggest that Pittis and Gabaldón's conclusions rely on the false assumption of a molecular clock (i.e. that evolutionary rate is constant over the tree) for each protein in the analysis. Yet, this objection fails to acknowledge the possibility that even if *no* single gene evolved in a clock-like manner, increases and decreases in rates across lineages and proteins could cancel so that the *average* of stem lengths of a large protein set may be roughly clocklike. The main concern then becomes the possibility of *systematic* rate increases or decreases that have occurred over many proteins of a specific origin as a result of shifts in function (i.e. in mitochondria and/or in the nucleocytoplasmic lineage) during eukaryogenesis. Such a phenomenon could generate the observed distributions of stem-lengths even in the absence of a large time interval between the divergence of Archaea from eukaryotes and the mitochondrial symbiosis.

Another point of contention concerns the number and nature of the genes that the mitochondrial symbiont contributed to LECA. Pittis and Gabaldón used a strict criterion that proteins of mitochondrial origin should show the characteristic alphaproteobacterial phylogenetic affinity [104]. Of the ~1,100 putative LECA proteins, they only find a small fraction (~80) that show this pattern. The remainder of LECA proteins with bacterial affinities were allied to different specific bacterial clades or with 'mixed bacterial' groups. Pittis and Gabaldón suggest that these non-alphaproteobacterial affinity genes could derive from additional bacterial symbioses, or separate 'one-off' LGT events, that occurred prior to the mitochondrial symbiosis. In sharp contrast, Ku and colleagues [23] in a separate analysis attributed all LECA proteins of bacterial origin in eukaryotes to the mitochondrial symbiont, suggesting that LGT of the genes into alphaproteobacteria (before and after the symbiosis) have obscured the origins of the eukaryote homologs. This argument rests on the claim that, except for massive numbers of EGT from mitochondria and plastids, LGT is vanishingly rare in eukaryotes [23]. Therefore, all bacterial-origin proteins in LECA must ultimately trace their origins to the pre-mitochondrial alphaproteobacterial genome. Yet, as discussed above, there are abundant clear examples of prokaryote-to-eukaryote and eukaryote-to-eukaryote LGTs that do not originate from endosymbiont-derived organelles [74,75,107,108]. Since LGT *has* occurred during eukaryote evolution and likely also affected the proto-eukaryotic lineage, it is more reasonable to assume that the phylogenetic affinities that are recovered in the phylogenetic trees, when clearly resolved, are the best estimates of the true donor lineage of the genes in question. This points to a much greater role for LGT in eukaryogenesis than previously appreciated.

The implications of the Asgard lineages, the timing of the mitochondrial symbiosis and the phylogenetic origins of eukaryotic genes are still under debate, so it is difficult to draw firm conclusions. However, we can at least conclude that the mitochondrial symbiosis postdates the origin of the ESPs found in Asgard and TACK archaea. Given the apparent genetic contribution of the alphaproteobacterial endosymbiont to eukaryotic systems in addition to mitochondria [24,68], it seems that the integration of mitochondria had some role in the origin of other eukaryotic cellular and molecular features prior to LECA. Therefore, the mitochondrial endosymbiosis was likely neither the first, nor the last, event in eukaryogenesis.

Conclusions

The endosymbiotic origin of mitochondria was of major importance to eukaryotic evolution, but it was not a single saltational event, as it is sometimes portrayed. Under *any* hypothesis of mitochondrial origins, the endosymbiotic alphaproteobacterium-to-organelle transition involved thousands of evolutionary steps each of which involved 'intermediate' proto-eukaryotes with proto-mitochondria, all of which may be extinct. Similarly, unitary accounts of the nature of the endosymbiotic association may be over-simplified, as different stages in the process quite

likely had different symbiotic characteristics [109]. For example, the symbiont may have started out utilizing host metabolite resources as a mild parasite or the host and symbiont could have been syntrophic partners, but then, once the host had tapped the symbiont ATP supply, the association may have shifted to enslavement. Regardless of how these initial stages of the association played out, the autonomy of both the mitochondrial symbiont and the host cell were ultimately eroded through the progressive integration of both cells. This merger was made possible, in large part, by the origin of the mitochondrial protein import apparatus that allowed host and symbiont compartments to mix genes and proteomes. Although the precise environmental context and nature of the symbiotic association is not known for certain, it is clear that all proto-mitochondrion-containing proto-eukaryotes must have lived in close proximity to oxygen. Part, or all, of these organisms' lifecycles *must* have required aerobic respiration, as these mitochondrial pathways are carried out by proteins with clear phylogenetic affinities to the pre-mitochondrial alphaproteobacterium [24,25].

If we are lucky, metagenomic explorations of microbial biodiversity in under-studied environments may turn up microbes that are more closely related to eukaryotes than the known Asgard archaea, new lineages of alphaproteobacteria with specific affinities to mitochondria, or, even better (and perhaps less likely), offshoots from the proto-mitochondrial/proto-eukaryote phase of eukaryogenesis. If so, then much needed light will be shed on the nature and timing of the mitochondrial endosymbiosis and its impact on eukaryogenesis. However, it is quite possible that most, if not all, of these 'intermediate' lineages of organisms are extinct. To make progress on these questions, then, we will have to rely on the development of better phylogenomic methods, improved sampling of genomic diversity of bacteria and eukaryotes and experimental investigations into mitochondrial functions.

Although some aspects of mitochondrial evolution may end up being unknowable with any degree of certainty, others are more tractable. For example, much progress has been, and is still being, made in understanding how mitochondrial biogenesis, division and metabolism are coordinated with the rest of the eukaryotic cell. Furthermore, as the functions and properties of mitochondria are being explored in more non-model system eukaryotes from all parts of the eukaryote tree, we are learning not only about what the common 'core functions' of mitochondria are and their origins [18,24,25], but also their evolutionary plasticity [21,69,79,81]. Significant lineage-specific shifts in mitochondrial proteome composition and function have occurred in adaptation to anaerobiosis in more than a dozen different eukaryotic lineages [21,81,93]. The key roles of lateral gene transfer, loss, gene duplication and functional divergence in the retailoring of mitochondrial function are only now becoming clear. The fields of mitochondrial biology, evolution and microbial biodiversity are beginning to merge, with great promise for expanding our understanding of this fundamental organelle.

ACKNOWLEDGMENTS

We thank Laura Eme, Michelle Leger, Maureen O'Malley and Michael Gray for critical review of the text and figures and stimulating discussion. A.J.R. also thanks Ford Doolittle, John Archibald, Nick Lane, Alastair Simpson, T. Martin Embley and Miklós Müller for insights they have shared. We apologize to those whose important work on these topics could not be cited due to space constraints. Research on mitochondrial evolution and function in the laboratory of A.J.R. is supported by a Discovery grant from the Natural Sciences and Engineering Research Council (2016-06792) and an Operating grant from the Canadian Institutes of Health Research (MOP-142349). R.K. is supported by grants from the Japan Society for the Promotion of Science (JSPS; no.s 15H05606 and 15K14591). S.A.M.-G. is supported by a Killam Pre-doctoral Scholarship.

REFERENCES

1. Calvo, S.E., and Mootha, V.K. (2010). The mitochondrial proteome and human disease. Annu. Rev. Genomics Hum. Genet. *11*, 25–44.
2. Gawryluk, R.M.R., Chisholm, K.A., Pinto, D.M., and Gray, M.W. (2014). Compositional complexity of the mitochondrial proteome of a unicellular eukaryote (*Acanthamoeba castellanii*, supergroup Amoebozoa) rivals that of animals, fungi, and plants. J. Proteomics *109*, 400–416.
3. Atteia, A., Adrait, A., Brugière, S., Tardif, M., van Lis, R., Deusch, O., Dagan, T., Kuhn, L., Gontero, B., Martin, W., *et al*. (2009). A proteomic survey of *Chlamydomonas reinhardtii* mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of the alphaproteobacterial mitochondrial ancestor. Mol. Biol. Evol. *26*, 1533–1548.
4. Panigrahi, A.K., Ogata, Y., Zíková, A., Anupama, A., Dalley, R.A., Acestor, N., Myler, P.J., and Stuart, K.D. (2009). A comprehensive analysis of *Trypanosoma brucei* mitochondrial proteome. Proteomics *9*, 434–450.
5. Lee, C.P., Taylor, N.L., and Millar, A.H. (2013). Recent advances in the composition and heterogeneity of the *Arabidopsis* mitochondrial proteome. Front. Plant Sci. *4*, 4.
6. Sagan, L. (1967). On the origin of mitosing cells. J. Theor. Biol. *14*, 225–274.
7. Archibald, J. (2014). One Plus One Equals One (Oxford: Oxford University Press).
8. Bonen, L., Cunningham, R.S., Gray, M.W., and Doolittle, W.F. (1977). Wheat embryo mitochondrial 18S ribosomal RNA: evidence for its prokaryotic nature. Nucleic Acids Res. *4*, 663–671.
9. Schwartz, R.M., and Dayhoff, M.O. (1978). Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts. Science *199*, 395–403.
10. Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G.J., and Woese, C.R. (1985). Mitochondrial origins. Proc. Natl. Acad. Sci. USA *82*, 4443–4447.
11. Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., van Eijk, R., Schleper, C., Guy, L., and Ettema, T.J.G. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature *521*, 173–179.
12. Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K.W., Anantharaman, K., Starnawski, P., Kjeldsen, K.U., *et al*. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature *541*, 353–358.
13. Wang, Z., and Wu, M. (2015). An integrated phylogenomic approach toward pinpointing the origin of mitochondria. Sci. Rep. *5*, 7949.
14. Rodríguez-Ezpeleta, N., and Embley, T.M. (2012). The SAR11 group of alpha-proteobacteria is not related to the origin of mitochondria. PLoS One *7*, e30520.
15. Roger, A.J. (1999). Reconstructing early events in eukaryotic evolution. Am. Nat. *154*, S146–S163.
16. Embley, T.M., and Martin, W. (2006). Eukaryotic evolution, changes and challenges. Nature *440*, 623–630.
17. Dacks, J.B., Field, M.C., Buick, R., Eme, L., Gribaldo, S., Roger, A.J., Brochier-Armanet, C., and Devos, D.P. (2016). The changing view of eukaryogenesis - fossils, cells, lineages and how they all come together. J. Cell Sci. *129*, 3695–3703.
18. Gray, M.W. (2015). Mosaic nature of the mitochondrial proteome: Implications for the origin and evolution of mitochondria. Proc. Natl. Acad. Sci. USA *112*, 10133–10138.

19. Karnkowska, A., Vacek, V., Zubáčová, Z., Treitli, S.C., Petrželková, R., Eme, L., Novák, L., Žárský, V., Barlow, L.D., Herman, E.K., *et al*. (2016). A eukaryote without a mitochondrial organelle. Curr. Biol. *26*, 1274–1284.

20. Gabaldón, T., and Huynen, M.A. (2004). Shaping the mitochondrial proteome. Biochim. Biophys. Acta *1659*, 212–220.

21. Stairs, C.W., Leger, M.M., and Roger, A.J. (2015). Diversity and origins of anaerobic metabolism in mitochondria and related organelles. Philos. Trans. R. Soc. B *370*, 20140326.

22. Nývltová, E., Stairs, C.W., Hrdý, I., Rídl, J., Mach, J., Pačes, J., Roger, A.J., and Tachezy, J. (2015). Lateral gene transfer and gene duplication played a key role in the evolution of *Mastigamoeba balamuthi* hydrogenosomes. Mol. Biol. Evol. *32*, 1039–1055.

23. Ku, C., Nelson-Sathi, S., Roettger, M., Sousa, F.L., Lockhart, P.J., Bryant, D., Hazkani-Covo, E., McInerney, J.O., Landan, G., and Martin, W.F. (2015). Endosymbiotic origin and differential loss of eukaryotic genes. Nature *524*, 427–432.

24. Wang, Z., and Wu, M. (2014). Phylogenomic reconstruction indicates mitochondrial ancestor was an energy parasite. PLoS One *9*, e110685.

25. Szklarczyk, R., and Huynen, M.A. (2010). Mosaic origin of the mitochondrial proteome. Proteomics *10*, 4012–4024.

26. Abhishek, A., Bavishi, A., Bavishi, A., and Choudhary, M. (2011). Bacterial genome chimaerism and the origin of mitochondria. Can. J. Microbiol. *57*, 49–61.

27. Thiergart, T., Landan, G., Schenk, M., Dagan, T., and Martin, W.F. (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.

28. Margulis, L. (1970). Origin of Eukaryotic Cells (New Haven, CT: Yale University Press).

29. Whatley, J.M., John, P., and Whatley, F.R. (1979). From extracellular to intracellular: the establishment of mitochondria and chloroplasts. Proc. R. Soc. Lond. B Biol. Sci. *204*, 165–187.

30. Speijer, D. (2017). Alternating terminal electron-acceptors at the basis of symbiogenesis: How oxygen ignited eukaryotic evolution. BioEssays *39*, https://doi.org/10.1002/bies.201600174.

31. Moreira, D., and López-García, P. (1998). Symbiosis between methanogenic archaea and δ-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. J. Mol. Evol. *47*, 517–530.

32. López-García, P., and Moreira, D. (2015). Open questions on the origin of eukaryotes. Trends Ecol. Evol. *30*, 697–708.

33. Andersson, S.G., and Kurland, C.G. (1999). Origins of mitochondria and hydrogenosomes. Curr. Opin. Microbiol. *2*, 535–541.

34. Woese, C.R. (1977). Endosymbionts and mitochondrial origins. J. Mol. Evol. *10*, 93–96.

35. Searcy, D.G. (1992). Origins of mitochondria and chloroplasts from sulphur-based symbioses. In The Origin and Evolution of the Cell, H. Hartman, and K. Matsuno, eds. (Singapore: World Scientific), pp. 47–78.

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

37. Fenchel, T., and Bernard, C. (1993). Endosymbiotic purple non-sulphur bacteria in an anaerobic ciliated protozoon. FEMS Microbiol. Lett. *110*, 21–25.

38. Cavalier-Smith, T. (2006). Origin of mitochondria by intracellular enslavement of a photosynthetic purple bacterium. Proc. R. Soc. Lond. B Biol. Sci. *273*, 1943–1952.

39. Gabaldón, T., and Huynen, M.A. (2003). Reconstruction of the proto-mitochondrial metabolism. Science *301*, 609–609.

40. Davidov, Y., Huchon, D., Koval, S.F., and Jurkevitch, E. (2006). A new alpha-proteobacterial clade of *Bdellovibrio*-like predators: implications for the mitochondrial endosymbiotic theory. Environ. Microbiol. *8*, 2179–2188.

41. Davidov, Y., and Jurkevitch, E. (2009). Predation between prokaryotes and the origin of eukaryotes. BioEssays *31*, 748–757.

42. Sassera, D., Lo, N., Epis, S., D'Auria, G., Montagna, M., Comandatore, F., Horner, D., Peretó, J., Luciano, A.M., Franciosi, F., *et al*. (2011). Phylogenomic evidence for the presence of a flagellum and cbb3 oxidase in the free-living mitochondrial ancestor. Mol. Biol. Evol. *28*, 3285–3296.

43. Muñoz-Gómez, S.A., Wideman, J.G., Roger, A.J., and Slamovits, C.H. (2017). The origin of mitochondrial cristae from alphaproteobacteria. Mol. Biol. Evol. *34*, 943–956.

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

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

46. Andersson, G.E., Karlberg, O., Canbäck, B., and Kurland, C.G. (2003). On the origin of mitochondria: a genomics perspective. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *358*, 165–179.

47. Amiri, H., Karlberg, O., and Andersson, S.G.E. (2003). Deep origin of plastid/parasite ATP/ADP translocases. J. Mol. Evol. *56*, 137–150.

48. John, P., Whatley, F.R., and John, P. (1975). *Paracoccus denitrificans* and the evolutionary origin of the mitochondrion. Nature *254*, 495–498.

49. Gross, J., and Bhattacharya, D. (2011). Endosymbiont or host: who drove mitochondrial and plastid evolution? Biol. Direct *6*, 12.

50. Hewitt, V., Alcock, F., and Lithgow, T. (2011). Minor modifications and major adaptations: the evolution of molecular machines driving mitochondrial protein import. Biochim. Biophys. Acta *1808*, 947–954.

51. Wiedemann, N., and Pfanner, N. (2017). Mitochondrial machineries for protein import and assembly. Annu. Rev. Biochem. *86*, 685–714.

52. Hewitt, V., Lithgow, T., and Waller, R.F. (2014). Modifications and innovations in the evolution of mitochondrial protein import pathways. In Endosymbiosis (Vienna: Springer), pp. 19–35.

53. Dimmer, K.S., and Rapaport, D. (2017). Mitochondrial contact sites as platforms for phospholipid exchange. Biochim. Biophys. Acta *1862*, 69–80.

54. Ellenrieder, L., Rampelt, H., and Becker, T. (2017). Connection of protein transport and organelle contact sites in mitochondria. J. Mol. Biol. *429*, 2148–2160.

55. Wideman, J.G., and Muñoz-Gómez, S.A. (2016). The evolution of ERMIONE in mitochondrial biogenesis and lipid homeostasis: An evolutionary view from comparative cell biology. Biochim. Biophys. Acta *1861*, 900–912.

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

57. Purkanti, R., and Thattai, M. (2015). Ancient dynamin segments capture early stages of host-mitochondrial integration. Proc. Natl. Acad. Sci. USA *112*, 2800–2805.

58. Friedman, J.R., Lackner, L.L., West, M., DiBenedetto, J.R., Nunnari, J., and Voeltz, G.K. (2011). ER tubules mark sites of mitochondrial division. Science *334*, 358–362.

59. Murley, A., Lackner, L.L., Osman, C., West, M., Voeltz, G.K., Walter, P., and Nunnari, J. (2013). ER-associated mitochondrial division links the distribution of mitochondria and mitochondrial DNA in yeast. eLife *2*, e00422.

60. Wideman, J.G., Gawryluk, R.M.R., Gray, M.W., and Dacks, J.B. (2013). The ancient and widespread nature of the ER-mitochondria encounter structure. Mol. Biol. Evol. *30*, 2044–2049.

61. Smirnova, E., Griparic, L., Shurland, D.L., and van der Bliek, A.M. (2001). Dynamin-related protein Drp1 is required for mitochondrial division in mammalian cells. Mol. Biol. Cell *12*, 2245–2256.

62. Kühlbrandt, W. (2015). Structure and function of mitochondrial membrane protein complexes. BMC Biol. *13*, 89.

63. Muñoz-Gómez, S.A., Slamovits, C.H., Dacks, J.B., Baier, K.A., Spencer, K.D., and Wideman, J.G. (2015). Ancient homology of the mitochondrial contact site and cristae organizing system points to an endosymbiotic origin of mitochondrial cristae. Curr. Biol. *25*, 1489–1495.

64. Huynen, M.A., Mühlmeister, M., Gotthardt, K., Guerrero-Castillo, S., and Brandt, U. (2016). Evolution and structural organization of the mitochondrial contact site (MICOS) complex and the mitochondrial intermembrane space bridging (MIB) complex. Biochim. Biophys. Acta *1863*, 91–101.

65. Braymer, J.J., and Lill, R. (2017). Iron-sulfur cluster biogenesis and trafficking in mitochondria. J. Biol. Chem. *292*, 12754–12763.

66. Freibert, S.-A., Goldberg, A.V., Hacker, C., Molik, S., Dean, P., Williams, T.A., Nakjang, S., Long, S., Sendra, K., Bill, E., *et al*. (2017). Evolutionary conservation and in vitro reconstitution of microsporidian iron-sulfur cluster biosynthesis. Nat. Commun *8*, 13932.

67. Karlberg, O., Canbäck, B., Kurland, C.G., and Andersson, S.G. (2000). The dual origin of the yeast mitochondrial proteome. Yeast *17*, 170–187.

68. Gabaldón, T., and Huynen, M.A. (2007). From endosymbiont to host-controlled organelle: the hijacking of mitochondrial protein synthesis and metabolism. PLoS Comput. Biol. *3*, e219.

69. Gabaldón, T., and Pittis, A.A. (2015). Origin and evolution of metabolic sub-cellular compartmentalization in eukaryotes. Biochimie *119*, 262–268.

70. Kurland, C.G., and Andersson, S.G. (2000). Origin and evolution of the mitochondrial proteome. Microbiol. Mol. Biol. Rev. *64*, 786–820.

71. Burger, G., Gray, M.W., Forget, L., and Lang, B.F. (2013). Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. Genome Biol. Evol. *5*, 418–438.

72. Boussau, B., Karlberg, E.O., Frank, A.C., Legault, B.-A., and Andersson, S.G.E. (2004). Computational inference of scenarios for α-proteobacterial genome evolution. Proc. Natl. Acad. Sci. USA *101*, 9722–9727.

73. McCutcheon, J.P., and Moran, N.A. (2011). Extreme genome reduction in symbiotic bacteria. Nat. Rev. Microbiol. *10*, 13–26.

74. McCutcheon, J.P. (2016). From microbiology to cell biology: when an intracellular bacterium becomes part of its host cell. Curr. Opin. Cell Biol. *41*, 132–136.

75. Nowack, E.C.M., Price, D.C., Bhattacharya, D., Singer, A., Melkonian, M., and Grossman, A.R. (2016). Gene transfers from diverse bacteria compensate for reductive genome evolution in the chromatophore of *Paulinella chromatophora*. Proc. Natl. Acad. Sci. USA *113*, 12214–12219.

76. Shutt, T.E., and Gray, M.W. (2006). Bacteriophage origins of mitochondrial replication and transcription proteins. Trends Genet. *22*, 90–95.

77. Maier, U.-G., Zauner, S., Woehle, C., Bolte, K., Hempel, F., Allen, J.F., and Martin, W.F. (2013). Massively convergent evolution for ribosomal protein gene content in plastid and mitochondrial genomes. Genome Biol. Evol. *5*, 2318–2329.

78. Flegontov, P., Michálek, J., Janouškovec, J., Lai, D.-H., Jirků, M., Hajdušková, E., Tomčala, A., Otto, T.D., Keeling, P.J., Pain, A., *et al*. (2015). Divergent mitochondrial respiratory chains in phototrophic relatives of apicomplexan parasites. Mol. Biol. Evol. *32*, 1115–1131.

79. Zíková, A., Hampl, V., Paris, Z., Týč, J., and Lukeš, J. (2016). Aerobic mitochondria of parasitic protists: Diverse genomes and complex functions. Mol. Biochem. Parasitol *209*, 46–57.

80. Smith, D.R., and Keeling, P.J. (2016). Protists and the wild, wild west of gene expression: new frontiers, lawlessness, and misfits. Annu. Rev. Microbiol. *70*, 161–178.

81. Müller, M., Mentel, M., van Hellemond, J.J., Henze, K., Woehle, C., Gould, S.B., Yu, R.-Y., van der Giezen, M., Tielens, A.G.M., and Martin, W.F. (2012). Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. Microbiol. Mol. Biol. Rev. *76*, 444–495.

82. Graaf, D.M.R., Ricard, G., Alen, V.A.T., Duarte, I., Dutilh, B.E., Burgtorf, C., Kuiper, J.W.P., Staay, V.D., *et al*. (2011). The organellar genome and metabolic potential of the hydrogen-producing mitochondrion of *Nyctotherus ovalis*. Mol. Biol. Evol. *28*, 2379–2391.

83. Stechmann, A., Hamblin, K., Pérez-Brocal, V., Gaston, D., Richmond, G.S., van der Giezen, M., Clark, C.G., and Roger, A.J. (2008). Organelles in *Blastocystis* that blur the distinction between mitochondria and hydrogenosomes. Curr. Biol. *18*, 580–585.

84. Leger, M.M., Gawryluk, R.M.R., Gray, M.W., and Roger, A.J. (2013). Evidence for a hydrogenosomal-type anaerobic ATP generation pathway in *Acanthamoeba castellanii*. PLoS One *8*, e69532.

85. Gawryluk, R.M.R., Kamikawa, R., Stairs, C.W., Silberman, J.D., Brown, M.W., and Roger, A.J. (2016). The earliest stages of mitochondrial adaptation to low oxygen revealed in a novel rhizarian. Curr. Biol. *26*, 2729–2738.

86. Tovar, J., León-Avila, G., Sánchez, L.B., Sutak, R., Tachezy, J., van der Giezen, M., Hernández, M., Müller, M., and Lucocq, J.M. (2003). Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. Nature *426*, 172–176.

87. Leger, M.M., Kolisko, M., Kamikawa, R., Stairs, C.W., Kume, K., Čepička, I., Silberman, J.D., Andersson, J.O., Xu, F., Yabuki, A., *et al*. (2017). Organelles that illuminate the origins of *Trichomonas* hydrogenosomes and Giardia mitosomes. Nat. Ecol. Evol. *1*, 92.

88. Ali, V., Shigeta, Y., Tokumoto, U., Takahashi, Y., and Nozaki, T. (2004). An intestinal parasitic protist, *Entamoeba histolytica*, possesses a non-redundant nitrogen fixation-like system for iron-sulfur cluster assembly under anaerobic conditions. J. Biol. Chem. *279*, 16863–16874.

89. Nývltová, E., Šuták, R., Harant, K., Šedinová, M., Hrdy, I., Paces, J., Vlček, Č., and Tachezy, J. (2013). NIF-type iron-sulfur cluster assembly system is duplicated and distributed in the mitochondria and cytosol of *Mastigamoeba balamuthi*. Proc. Natl. Acad. Sci. USA *110*, 7371–7376.

90. Stairs, C.W., Eme, L., Brown, M.W., Mutsaers, C., Susko, E., Dellaire, G., Soanes, D.M., van der Giezen, M., and Roger, A.J. (2014). A SUF Fe-S cluster biogenesis system in the mitochondrion-related organelles of the anaerobic protist *Pygsuia*. Curr. Biol. *24*, 1176–1186.

91. Martin, W.F., Garg, S., and Zimorski, V. (2015). Endosymbiotic theories for eukaryote origin. Philos. Trans. R. Soc. B *370*, 20140330.

92. Hug, L.A., Stechmann, A., and Roger, A.J. (2010). Phylogenetic distributions and histories of proteins involved in anaerobic pyruvate metabolism in eukaryotes. Mol. Biol. Evol. *27*, 311–324.

93. Embley, T.M. (2006). Multiple secondary origins of the anaerobic lifestyle in eukaryotes. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *361*, 1055–1067.

94. Leger, M.M., Eme, L., Hug, L.A., and Roger, A.J. (2016). Novel hydrogenosomes in the microaerophilic jakobid *Stygiella incarcerata*. Mol. Biol. Evol. *33*, 2318–2336.

95. Degli Esposti, M., Cortez, D., Lozano, L., Rasmussen, S., Nielsen, H.B., and Martinez Romero, E. (2016). Alpha proteobacterial ancestry of the [Fe-Fe]-hydrogenases in anaerobic eukaryotes. Biol. Direct *11*, 34.

96. Stairs, C.W., Roger, A.J., and Hampl, V. (2011). Eukaryotic pyruvate formate lyase and its activating enzyme were acquired laterally from a firmicute. Mol. Biol. Evol. *28*, 2087–2099.

97. Martijn, J., and Ettema, T.J.G. (2013). From archaeon to eukaryote: the evolutionary dark ages of the eukaryotic cell. Biochem. Soc. Trans. *41*, 451–457.

98. López-García, P., Eme, L., and Moreira, D. (2017). Symbiosis in eukaryotic evolution. J. Theor. Biol., In Press. https://doi.org/10.1016/j.jtbi.2017.02.031.

99. Lane, N., and Martin, W.F. (2015). The energetics of genome complexity. Nature *467*, 929–934.

100. Heimerl, T., Flechsler, J., Pickl, C., Heinz, V., Salecker, B., Zweck, J., Wanner, G., Geimer, S., Samson, R.Y., Bell, S.D., *et al*. (2017). A complex endomembrane system in the archaeon *Ignicoccus hospitalis* tapped by *Nanoarchaeum equitans*. Front. Microbiol. *8*, 1072.

101. Sousa, F.L., Neukirchen, S., Allen, J.F., Lane, N., and Martin, W.F. (2016). Lokiarchaeon is hydrogen dependent. Nat. Microbiol. *1*, 16034.

102. Seitz, K.W., Lazar, C.S., Hinrichs, K.-U., Teske, A.P., and Baker, B.J. (2016). Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction. ISME J. *10*, 1696–1705.

103. Spang, A., Caceres, E.F., and Ettema, T.J.G. (2017). Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. Science *357*, pii: eaaf3883.

104. Pittis, A.A., and Gabaldón, T. (2016). Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. Nature *531*, 101–104.

105. Martin, W.F., Roettger, M., Ku, C., Garg, S.G., Nelson-Sathi, S., and Landan, G. (2017). Late mitochondrial origin is an artefact. Genome Biol. Evol., In Press. https://doi.org/10.1093/gbe/evx027.

106. Pittis, A.A., and Gabaldon, T. (2016). On phylogenetic branch lengths distribution and the late acquistion of mitochondria. bioRxiv, 64873.

107. Hirt, R.P., Alsmark, C., and Embley, T.M. (2015). Lateral gene transfers and the origins of the eukaryote proteome: a view from microbial parasites. Curr. Opin. Microbiol. *23*, 155–162.

108. Richards, T.A., and Talbot, N.J. (2013). Horizontal gene transfer in osmotrophs: playing with public goods. Nat. Rev. Microbiol. *11*, 720–727.

109. Keeling, P.J., and McCutcheon, J.P. (2017). Endosymbiosis: The feeling is not mutual. J. Theor. Biol., In Press. https://doi.org/10.1016/j.jtbi.2017.06.008.

110. Viale, A.M., and Arakaki, A.K. (1994). The chaperone connection to the origins of the eukaryotic organelles. FEBS Lett. *341*, 146–151.

111. Andersson, S.G.E., Zomorodipour, A., Andersson, J.O., Sicheritz-Pontén, T., Alsmark, U.C.M., Podowski, R.M., Nässlund, A.K., Eriksson, A.-S., Winkler, H.H., and Kurland, C.G. (1998). The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature *396*, 133–140.

112. Wu, M., Sun, L.V., Vamathevan, J., Riegler, M., Deboy, R., Brownlie, J.C., McGraw, E.A., Martin, W., Esser, C., Ahmadinejad, N., *et al*. (2004). Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. PLoS Biol. *2*, E69.

113. Fitzpatrick, D.A., Creevey, C.J., and McInerney, J.O. (2006). Genome phylogenies indicate a meaningful α-proteobacterial phylogeny and support a grouping of the mitochondria with the Rickettsiales. Mol. Biol. Evol. *23*, 74–85.

114. Williams, K.P., Sobral, B.W., and Dickerman, A.W. (2007). A robust species tree for the alphaproteobacteria. J. Bacteriol *189*, 4578–4586.

115. Thrash, J.C., Boyd, A., Huggett, M.J., Grote, J., Carini, P., Yoder, R.J., Robbertse, B., Spatafora, J.W., Rappé, M.S., and Giovannoni, S.J. (2011). Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade. Sci. Rep *1*, 13.

116. Georgiades, K., Madoui, M.-A., Le, P., Robert, C., and Raoult, D. (2011). Phylogenomic analysis of Odyssella thessalonicensis fortifies the common origin of Rickettsiales, *Pelagibacter ubique* and *Reclimonas americana* mitochondrion. PLoS One *6*, e24857.

117. Brindefalk, B., Ettema, T.J.G., Viklund, J., Thollesson, M., and Andersson, S.G.E. (2011). A phylometagenomic exploration of oceanic alphaproteobacteria reveals mitochondrial relatives unrelated to the SAR11 clade. PLoS One *6*, e24457.

118. Viklund, J., Martijn, J., Ettema, T.J.G., and Andersson, S.G.E. (2013). Comparative and phylogenomic evidence that the alphaproteobacterium HIMB59 Is not a member of the oceanic SAR11 clade. PLoS One *8*, e78858.

119. Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T.J., Manuel, M., Wörheide, G., and Baurain, D. (2011). Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. *9*, e1000602.

120. Yang, Z., and Rannala, B. (2012). Molecular phylogenetics: principles and practice. Nat. Rev. Genet. *13*, 303–314.

121. Esser, C., Ahmadinejad, N., Wiegand, C., Rotte, C., Sebastiani, F., Gelius-Dietrich, G., Henze, K., Kretschmann, E., Richly, E., Leister, D., *et al*. (2004). A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. Mol. Biol. Evol. *21*, 1643–1660.

122. Georgiades, K., and Raoult, D. (2011). The rhizome of *Reclinomonas americana, Homo sapiens, Pediculus humanus* and *Saccharomyces cerevisiae* mitochondria. Biol. Direct *6*, 55.

123. Huson, D.H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. *23*, 254–267.

124. Szöllősi, G.J., Tannier, E., Daubin, V., and Boussau, B. (2015). The inference of gene trees with species trees. Syst. Biol. *64*, e42–62.

125. Daley, D.O., and Whelan, J. (2005). Why genes persist in organelle genomes. Genome Biol. *6*, 110.

126. Allen, J.F. (2015). Why chloroplasts and mitochondria retain their own genomes and genetic systems: Colocation for redox regulation of gene expression. Proc. Natl. Acad. Sci. USA *112*, 10231–10238.

127. Björkholm, P., Ernst, A.M., Hagström, E., and Andersson, S.G.E. (2017). Why mitochondria need a genome revisited. FEBS Lett. *591*, 65–75.

128. Johnston, I.G., and Williams, B.P. (2016). Evolutionary inference across eukaryotes identifies specific pressures favoring mitochondrial gene retention. Cell Syst *2*, 101–111.

129. Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., and Thompson, F., eds. (2014). The Prokaryotes - Alphaproteobacteria and Betaproteobacteria (Berlin, Heidelberg: Springer).

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

131. Nývltová, E., Smutná, T., Tachezy, J., and Hrdý, I. (2016). OsmC and incomplete glycine decarboxylase complex mediate reductive detoxification of peroxides in hydrogenosomes of *Trichomonas vaginalis*. Mol. Biochem. Parasitol. *206*, 29–38.