

Genome Biol. 2001; 2(6): reviews1018.1–reviews1018.5.  
Published online 2001 June 5.  
[Copyright](#) © 2001 BioMed Central Ltd

## The origin and early evolution of mitochondria

Michael W Gray,<sup>✉1</sup> Gertraud Burger,<sup>2</sup> and B Franz Lang<sup>2</sup>

<sup>1</sup>Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada

<sup>2</sup>Département de biochimie, Université de Montreal, Montreal, Quebec H3C 3J7, Canada

Correspondence: Michael W Gray. E-mail: [M.W.Gray@dal.ca](mailto:M.W.Gray@dal.ca)

Corresponding author.

Michael W Gray: [M.W.Gray@dal.ca](mailto:M.W.Gray@dal.ca)

### Abstract

Complete sequences of numerous mitochondrial, many prokaryotic, and several nuclear genomes are now available. These data **confirm that the mitochondrial genome originated from a eubacterial (specifically a-proteobacterial) ancestor** but raise questions about the evolutionary antecedents of the mitochondrial proteome.

Recent debates about eukaryotic cell evolution have been closely connected to the issue of how mitochondria originated and have evolved [[1,2,3,4,5,6,7](#)]. These debates have posed such questions as the following: Did the mitochondrion arise at the same time as, or subsequent to, the rest of the eukaryotic cell? Did it originate under initially anaerobic or aerobic conditions? What is the evolutionary relationship between mitochondria and hydrogenosomes (H<sub>2</sub>-generating and ATP-producing organelles that are found in eukaryotes lacking mitochondria)? Is the amitochondrial condition in these organisms a secondary adaptation or is it evolutionarily primitive - or, in other words, did any organisms diverge from the main line of eukaryotic evolution before the advent of mitochondria? Whereas the issue of how the eukaryotic cell arose remains controversial [[8,9](#)], current genomic data do allow us to make a number of reasonably compelling inferences about how mitochondria themselves originated and have since evolved.

### A mitochondrial genomics perspective

Much evidence supports the conclusion that the mitochondrial genome originated from within the (eu)bacterial [[8,9,10](#)], not the archaeal [[11](#)], domain of life. Specifically, among extant bacterial phyla, the a-proteobacteria are the closest identified relatives of mitochondria, as indicated, for example, by phylogenetic analyses of both protein-coding genes [[8,9](#)] and ribosomal RNA (rRNA) genes [[12](#)] specified by mitochondrial DNA (mtDNA).

Over the past two decades, many complete mitochondrial genome sequences have been determined, and several recent surveys have summarized various aspects of mitochondrial genome structure, gene content, organization and expression [[13,14,15,16](#)]. Two comprehensive mitochondrial genome-sequencing programs have particularly targeted

mtDNA in protists [17] and fungi [18]. A number of specific and general insights into mitochondrial genome evolution follow from these data. The first is that ATP production, coupled to electron transport, and translation of mitochondrial proteins represent the essence of mitochondrial function: these functions are common to all mitochondrial genomes and can be traced unambiguously and directly to an  $\alpha$ -proteobacterial ancestor. The mitochondrial genome encodes essential components for both of these processes [8,9].

The second insight is that the most ancestral (least derived), most bacterium-like and most gene-rich mitochondrial genome yet described is the 69,034 base pair (bp) mtDNA of the protist *Reclinomonas americana*, a jakobid flagellate [19] (jakobids are a group of putatively early diverging protozoa that share ultrastructural features with certain amitochondrial protists). By comparison, some other protist mtDNAs, most fungal, and all animal mtDNAs are highly derived, having diverged away from the ancestral pattern exemplified by *R. americana* mtDNA.

Sequencing has also shown that mitochondrial genomes have, to variable extents, undergone a streamlining process ("reductive evolution" [20]), leading to a marked loss of coding capacity compared to that of their closest eubacterial relatives. Mitochondrial gene content varies widely, from a high of 67 protein-coding genes in *R. americana* mtDNA to only three in the mitochondrial genome of apicomplexans [8,9], a group of strictly parasitic protists (specific relatives of dinoflagellates) including such organisms as *Plasmodium falciparum*, the causative agent of malaria. Differential gene content in mtDNAs is attributable primarily to mitochondrion-to-nucleus gene transfer [8,9,10,21,22] (which is demonstrably an on-going process in certain lineages, notably flowering plants [23]). Mitochondrial DNA may also lose genes whose functions are substituted for by unrelated genes encoded in the nucleus. A notable example is the replacement of an original multi-subunit bacteria-like RNA polymerase (inherited from the proto-mitochondrial ancestor and still encoded in certain jakobid - but no other - mitochondrial genomes) by a single-subunit bacteriophage T3/T7-like RNA polymerase, which directs mitochondrial transcription in virtually all eukaryotes [24]. Conversely, there may be complete loss of particular mitochondrial genes (and hence the corresponding functions) without functional complementation by nuclear genes. The complex I (*nad*) genes of the respiratory chain are one example of such loss. In the yeast *Saccharomyces cerevisiae*, neither the mitochondrial nor the nuclear genome contains classical complex I genes [25]; their disappearance from yeast mtDNA results in the absence of the first coupling site in the yeast electron-transport chain.

Furthermore, genome sequencing shows that the mitochondrial genome (and therefore mitochondria *per se*) arose only once in evolution. Several observations support this contention [8,9,10]. First, in any particular mitochondrial genome (with few exceptions [26]), genes that have an assigned function are a subset of those found in *R. americana* mtDNA. Second, in a number of cases, mitochondrial protein-coding clusters retain the gene order of their bacterial homologs, but these clusters exhibit mitochondrion-specific deletions that are most parsimoniously explained as having occurred in a common ancestor of mitochondrial genomes, subsequent to its divergence from the bacterial ancestor. Third, mitochondria form a monophyletic assemblage to the exclusion of bacterial species in phylogenetic reconstructions using concatenated protein sequences [8,9,25,27,28] as well in small-subunit rRNA trees [12].

A final insight from mitochondrial genome sequencing is the emergence of striking parallels in phylogenetic trees separately reconstructed from genes encoded by nuclear DNA [7] and mtDNA [8,9]. In both cases, certain clades (such as animals plus fungi or red plus green algae) have become robust, although connections among these clades and other eukaryotic

species or groups cannot yet be precisely resolved. These emerging parallels support the view that mitochondrial and nuclear genomes have evolved in concert throughout much, if not most, of the evolutionary history of the domain Eukarya.

## A prokaryotic genomics perspective

Among the many complete bacterial genome sequences that are now available, that of *Rickettsia prowazekii* [27] (1,111,523 bp) stands out as the 'most mitochondrial'. Comparison of this sequence with the mitochondrial genome sequence of *Reclinomonas americana* (the 'most bacterial' of sequenced mtDNAs) solidifies the conclusion drawn from other kinds of data that the mitochondrial genome has arisen from within a subdivision of the  $\alpha$ -proteobacteria that contains *Rickettsia* and certain other obligate intracellular parasites. Yet this comparison also highlights a number of important distinctions.

First, although the *R. americana* mitochondrial and *R. prowazekii* DNAs are both "stunning examples of highly derived genomes" [27], it is clear that they are the products of independent processes of reductive evolution, as are the genomes of many other bacterial pathogens. In particular, no shared derived traits (such as gene order) are apparent that specifically link mitochondrial and *R. prowazekii* genomes to the exclusion of other bacterial genomes. Rather, the two genome types must have shared a common free-living ancestor that presumably had a much larger gene content, with separate processes of genome reduction ensuing in the two descendant lineages [8,12].

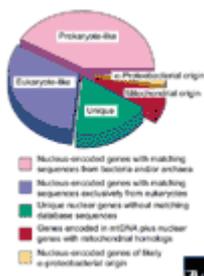
A second consideration is that although mitochondria and *R. prowazekii* exhibit very similar functional profiles with respect to ATP production (reflecting the common evolutionary origin of their electron transport chains), associated aspects of ATP utilization are quite different. For example, whereas mitochondria export ATP to the cytosol, *Rickettsia* uses the ATP it produces, and even imports ATP from the host during early stages in its development [29]. The membrane-associated ADP/ATP translocases in *Rickettsia* and mitochondria are not specifically related, evidently having arisen independently during the intracellular adaptation of parasite and organelle, after their divergence from a last common ancestor. In fact, many of the metabolic similarities between *Rickettsia* and mitochondria (for example, the absence of glycolytic enzymes) probably reflect convergent evolution rather than vertical inheritance [12,27,30].

Finally, because the *Rickettsia* genome sequence is so highly reduced and the organism itself is an obligate intracellular parasite, this particular genome sequence does not readily address questions about the original gene complement that the mitochondrial ancestor would have possessed when it was still a free-living  $\alpha$ -proteobacterium. For this reason, it will be essential to have complete sequences for a variety of the larger genomes of free-living  $\alpha$ -proteobacteria. The first such complete sequence, that of *Caulobacter crescentus* (4,016,942 bp), has just been published [31]. Comparison of this sequence with those of other, substantially different  $\alpha$ -proteobacterial genomes (such as the 8.7 megabase (Mb) genome of *Bradyrhizobium japonicum* and the genomes of photosynthetic  $\alpha$ -proteobacteria such as *Rhodobacter*) will undoubtedly provide a clearer picture of the metabolic versatility with which the proto-mitochondrion might have been endowed.

## A view from the nucleus

The availability of complete sequences for several nuclear genomes has prompted studies to probe the evolutionary origin(s) of the mitochondrial proteome: the collection of proteins that make up the mitochondrion and are involved in mitochondrial biogenesis. In *S. cerevisiae*, some 423 proteins (393 specified by the nuclear genome) have been annotated as putatively encoding mitochondrial proteins [32,33]. Karlberg *et al.* [34] employed similarity searches and phylogenetic reconstructions to examine the evolutionary affiliation of these proteins. In a separate study, Marcotte *et al.* [35] used a computational genetics approach [36] to assign yeast proteins to particular subcellular compartments on the basis of the phylogenetic distribution of their homologs. By this approach, Marcotte *et al.* [35] estimated that there are about 630 mitochondrial proteins in yeast (10% of its coding information).

Although differing in detail, both of these studies [34,35] come to similar general conclusions about the origin of the yeast mitochondrial proteome. In particular, the two studies - which both consist fundamentally of similarity searches - identify three categories of yeast mitochondrial proteins (Figure 1): 'prokaryote-specific' (50-60% of the total), 'eukaryote-specific' (20-30%) and 'organism-specific', or 'unique' (about 20%). Prokaryote-specific mitochondrial proteins are defined as those that have counterparts in prokaryotic genomes; eukaryote-specific mitochondrial proteins have counterparts in other eukaryotic genomes but not in prokaryotic genomes; and organism-specific mitochondrial proteins are ones so far unique to *S. cerevisiae*. In addition, both studies point out that this classification correlates with the known or inferred functions of the proteins in each category: prokaryote-specific mitochondrial proteins predominantly perform roles in biosynthesis, bioenergetics and protein synthesis, whereas eukaryote-specific mitochondrial proteins function mainly as membrane components and in regulation and transport.



**Figure 1**

Division of the yeast mitochondrial proteome into different categories according to inferred evolutionary origin. The estimated proportions of yeast mitochondrial proteins in the various classes are taken from [34].

What do we make of these provocative observations? The presence of a large fraction of prokaryote-specific components in the mitochondrial proteome is not at all unexpected, given the demonstrated eubacterial origin of the mitochondrial genome. But although it has been suggested that the approximately 215 [34] or 370 [35] prokaryote-specific yeast mitochondrial genes provide "an estimate of the number of genes contributed by the ancestral mitochondrial genome" [35], this value should be viewed with caution, for three reasons. Firstly, a large proportion of the 'prokaryote-specific' mitochondrial proteins (about half according to Karlberg *et al.* [34]) have counterparts in eukaryotes as well as in bacteria and archaea; some or even many of these could well have thus been present in the universal common ancestor of all life forms and, therefore, were conceivably already present in whatever organism contributed the nuclear genome at the time of the mitochondrial endosymbiosis. Secondly, only a minority (38) of the prokaryote-specific, nucleus-encoded mitochondrial proteins of yeast can readily be placed with the  $\alpha$ -proteobacteria on the basis of phylogenetic reconstruction [34]. Thirdly, only about two thirds (24) of these  $\alpha$ -proteobacterial genes have

homologs in one or more characterized mitochondrial genomes [34]. The remaining 14 genes are claimed to be "strong candidates for ancient gene transfers from a-proteobacteria to nuclear genomes" [34]. Because no mtDNA-encoded homologs of these genes are currently known, however, the formal possibility exists that some of them (for instance, those encoding mitochondrial heat-shock proteins) have arisen by lateral gene transfer at a separate time from the mitochondrial endosymbiosis [37]. Strictly speaking, we can only be certain of the 64 protein-coding genes of assigned function in *R. americana* mtDNA [19] as deriving directly from the mitochondrial endosymbiont.

Perhaps the most intriguing aspect of these two studies is the eukaryote-specific fraction of the yeast mitochondrial proteome and the implication that "a large number of novel mitochondrial genes were recruited from the nuclear genome to complement the remaining genes from the bacterial ancestor" [34]. Certainly, there are functions (one likely candidate being protein import, mediated by the TOM and TIM protein translocases) that must have been acquired by mitochondria subsequent to the initial endosymbiosis event and that were instrumental in transforming the proto-mitochondrion into an integrated cell organelle. Here again, however, some caution is warranted in the interpretation of these observations, because fairly stringent BLAST cutoffs ( $E < 10^{-10}$  in [34] and  $E < 10^{-6}$  in [35]) were used in the similarity searches conducted in these analyses. These searches are thus 'best-case scenarios', in which only homologs retaining relatively high levels of sequence similarity would have been detected. Many transferred endosymbiont genes may simply have diverged too far in sequence to be identified as prokaryotic, let alone specifically a-proteobacterial. This may be particularly true for yeast, which is an evolutionarily derived organism with a dramatically reduced set of genes, and in which the identification of even mtDNA-encoded genes is not always straightforward [14]. For example, a gene encoding ribosomal protein S3 in *S. cerevisiae* mtDNA was only identified recently through the analysis of sophisticated multiple alignments that included sequences from a large number of less highly derived ascomycetes and lower fungi [38].

Inference of homology requires rigorous phylogenetic analyses [39] and a large database of sequences with an appropriate phylogenetic distribution [25]. Further genomic data and genome comparisons will no doubt refine our assessment of how much of the original proto-mitochondrial gene complement was lost, as opposed to being transferred to the nuclear genome, and how much of the mitochondrial proteome represents genuinely recruited functions that evolved within the eukaryotic cell after its formation. The data and insights generated by Karlberg *et al.* [34] and Marcotte *et al.* [35] will certainly stimulate additional detailed analysis of the mitochondrial proteome in other organisms. While it is easy to understand why yeast was the organism of choice for these initial explorations, we would argue that we very much need genomic data from a range of other eukaryotes to address questions about the origin of the mitochondrial proteome. Particularly appealing are those protists in which a minimally derived and gene-rich mitochondrial genome may signal a comparably ancestral nuclear genome in which transferred mitochondrial genes can be more readily and confidently identified.

## References

1. López-García P, Moreira D. Metabolic symbiosis at the origin of eukaryotes. Trends Biochem Sci. 1999;24:88–93. [PubMed]
2. Andersson SGE, Kurland CG. Origins of mitochondria and hydrogenosomes. Curr Opin Microbiol. 1999;2:535–541. [PubMed]

3. Dyall SD, Johnson PJ. Origins of hydrogenosomes and mitochondria: evolution and organelle biogenesis. *Curr Opin Microbiol.* 2000;3:404–411. [[PubMed](#)]
4. Rotte C, Henze K, Müller M, Martin W. Origins of hydrogenosomes and mitochondria. *Curr Opin Microbiol.* 2000;3:481–486. [[PubMed](#)]
5. Embley TM, Hirt RP. Early branching eukaryotes? *Curr Opin Genet Dev.* 1998;8:624–629. [[PubMed](#)]
6. Roger AJ. Reconstructing early events in eukaryotic evolution. *Am Nat.* 1999;154:S146–S163. [[PubMed](#)]
7. Philippe H, Germot A, Moreira D. The new phylogeny of eukaryotes. *Curr Opin Genet Dev.* 2000;10:596–601. [[PubMed](#)]
8. Gray MW, Burger G, Lang BF. Mitochondrial evolution. *Science.* 1999;283:1476–1481. doi: 10.1126/science.283.5407.1476. [[PubMed](#)]
9. Lang BF, Gray MW, Burger G. Mitochondrial genome evolution and the origin of eukaryotes. *Annu Rev Genet.* 1999;33:351–397. [[PubMed](#)]
10. Gray MW. Evolution of organellar genomes. *Curr Opin Genet Dev.* 1999;9:678–687. [[PubMed](#)]
11. Karlin S, Brocchieri L, Mrazek J, Campbell AM, Spormann AM. A chimeric prokaryotic ancestry of mitochondria and primitive eukaryotes. *Proc Natl Acad Sci USA.* 1999;96:9190–9195. [[PubMed](#)]
12. Gray MW. *Rickettsia*, typhus and the mitochondrial connection. *Nature.* 1998;396:109–110. [[PubMed](#)]
13. Paquin B, Laforest MJ, Forget L, Roewer I, Wang Z, Longcore J, Lang BF. The fungal mitochondrial genome project: evolution of fungal mitochondrial genomes and their gene expression. *Curr Genet.* 1997;31:380–395. doi: 10.1007/s002940050220. [[PubMed](#)]
14. Gray MW, Lang BF, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Brossard N, Delage E, Littlejohn TG, et al. Genome structure and gene content in protist mitochondrial DNAs. *Nucleic Acids Res.* 1998;26:865–878. [[PubMed](#)]
15. Lang BF, Seif E, Gray MW, O'Kelly CJ, Burger G. A comparative genomics approach to the evolution of eukaryotes and their mitochondria. *J Eukaryot Microbiol.* 1999;46:320–326. [[PubMed](#)]
16. Boore JL. Animal mitochondrial genomes. *Nucleic Acids Res.* 1999;27:1767–1780. [[PubMed](#)]
17. The Organelle Genome Megasequencing Program (OGMP).  
<http://megasun.bch.umontreal.ca/ogmpproj.html>
18. The Fungal Mitochondrial Genome Project (FMGP).  
<http://megasun.bch.umontreal.ca/People/lang/FMGP/FMGP.html>
19. Lang BF, Burger G, O'Kelly CJ, Cedergren R, Golding BG, Lemieux C, Sankoff D, Turmel M, Gray MW. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature.* 1997;387:493–497. [[PubMed](#)]
20. Andersson SGE, Kurland CG. Reductive evolution of resident genomes. *Trends Microbiol.* 1998;6:263–268. [[PubMed](#)]
21. Martin W, Herrmann RG. Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physiol.* 1998;118:9–17. [[PubMed](#)]
22. Berg OG, Kurland CG. Why mitochondrial genes are most often found in nuclei. *Mol Biol Evol.* 2000;17:951–961. [[PubMed](#)]
23. Adams KL, Daley DO, Qiu YL, Whelan J, Palmer JD. Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature.* 2000;408:354–357. doi: 10.1038/35042567. [[PubMed](#)]
24. Gray MW, Lang BF. Transcription in chloroplasts and mitochondria: a tale of two polymerases. *Trends Microbiol.* 1998;6:1–3. [[PubMed](#)]

25. Kurland CG, Andersson SGE. Origin and evolution of the mitochondrial proteome. *Microbiol Mol Biol Rev.* 2000;64:786–820. [[PubMed](#)]
26. Pont-Kingdon G, Okada NA, Macfarlane JL, Beagley CT, Watkins-Sims CD, Cavalier-Smith T, Clark-Walker GD, Wolstenholme DR. Mitochondrial DNA of the coral *Sarcophyton glaucum* contains a gene for a homologue of bacterial MutS: a possible case of gene transfer from the nucleus to the mitochondrion. *J Mol Evol.* 1998;46:419–431. [[PubMed](#)]
27. Andersson SGE, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski RM, Naslund AK, Eriksson AS, Winkler HH, Kurland CG. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature.* 1998;396:133–140. [[PubMed](#)]
28. Burger G, Saint-Louis D, Gray MW, Lang BF. Complete sequence of the mitochondrial DNA of the red alga *Porphyra purpurea* : cyanobacterial introns and shared ancestry of red and green algae. *Plant Cell.* 1999;11:1675–1694. [[PubMed](#)]
29. Andersson SGE. Bioenergetics of the obligate intracellular parasite *Rickettsia prowazekii*. *Biochim Biophys Acta.* 1998;1365:105–111. [[PubMed](#)]
30. Müller M, Martin W. The genome of *Rickettsia prowazekii* and some thoughts on the origin of mitochondria and hydrogenosomes. *BioEssays.* 1999;21:377–381. doi: 10.1002/(SICI)1521-1878(199905)21:5<377::AID-BIES4>3.3.CO;2-N. [[PubMed](#)]
31. Nierman WC, Feldblyum TV, Laub MT, Paulsen IT, Nelson KE, Eisen J, Heidelberg JF, Alley MR, Ohta N, Maddock JR, et al. Complete genome sequence of *Caulobacter crescentus*. *Proc Natl Acad Sci USA.* 2001;98:4136–4141. [[PubMed](#)]
32. Hodges PE, McKee AH, Davis BP, Payne WE, Garrels JI. The yeast proteome database (YPD): a model for the organization and presentation of genome-wide functional data. *Nucleic Acids Res.* 1999;27:69–73. [[PubMed](#)]
33. The Yeast Proteome Database. <http://www.proteome.com/YPDhome.html>
34. Karlberg O, Canbäk B, Kurland CG, Andersson SGE. The dual origin of the yeast mitochondrial proteome. *Yeast.* 2000;17:170–187. doi: 10.1002/1097-0061(20000930)17:3<170::AID-YEA25>3.0.CO;2-V. [[PubMed](#)]
35. Marcotte EM, Xenarios I, van der Blik AM, Eisenberg D. Localizing proteins in the cell from their phylogenetic profiles. *Proc Natl Acad Sci USA.* 2000;97:12115–12120. [[PubMed](#)]
36. Marcotte EM. Computational genetics: finding protein function by nonhomology methods. *Curr Opin Struct Biol.* 2000;10:359–365. [[PubMed](#)]
37. Karlin S, Brocchieri L. Heat shock protein 60 sequence comparisons: duplications, lateral transfer, and mitochondrial evolution. *Proc Natl Acad Sci USA.* 2000;97:11348–11353. [[PubMed](#)]
38. Bullerwell CE, Burger G, Lang BF. A novel motif for identifying Rps3 homologs in fungal mitochondrial genomes. *Trends Biochem Sci.* 2000;25:363–365. [[PubMed](#)]
39. Ribeiro S, Golding GB. The mosaic nature of the eukaryotic nucleus. *Mol Biol Evol.* 1998;15:779–788. [[PubMed](#)]