# Phylogenomic Evidence for the Presence of a Flagellum and $cbb_3$ Oxidase in the Free-Living Mitochondrial Ancestor

Davide Sassera,†,[1] Nathan Lo,†,[2] Sara Epis,[1] Giuseppe D'Auria,[3] Matteo Montagna,[1] Francesco Comandatore,[1] David Horner,[4] Juli Peretó,[3,5,6] Alberto Maria Luciano,[7] Federica Franciosi,[7] Emanuele Ferri,[8] Elena Crotti,[9] Chiara Bazzocchi,[1] Daniele Daffonchio,[9] Luciano Sacchi,[10] Andres Moya,[3,5,11] Amparo Latorre,[3,5,11] and Claudio Bandi[1,*]

[1]Dipartimento di Patologia Animale, Igiene e Sanità Pubblica Veterinaria, Università degli Studi di Milano, Milano, Italy

[2]School of Biological Sciences, The University of Sydney, Sydney, New South Wales, Australia

[3]Centro Superior de Investigación en Salud Pública, Generalitat Valenciana, Valencia, Spain

[4]Dipartimento di Scienze Biomolecolari e Biotecnologie, Università degli Studi di Milano, Milano, Italy

[5]Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, València, Spain

[6]Departament de Bioquímica i Biologia Molecular, Universitat de València, València, Spain

[7]Dipartimento di Scienze Animali, Università degli Studi di Milano, Milano, Italy

[8]Dipartimento di Biotecnologie e Bioscienze, Università degli Studi di Milano Bicocca, Milano, Italy

[9]Dipartimento di Scienze e Tecnologie Alimentari e Microbiologiche, Università degli Studi di Milano, Milano, Italy

[10]Dipartimento di Biologia Animale, Università degli Studi di Pavia, Pavia, Italy

[11]Departament de Genètica, Universitat de València, València, Spain

†These authors contributed equally to this work.

***Corresponding author:** E-mail: claudio.bandi@unimi.it.

**Associate editor:** Richard Thomas

## Abstract

The initiation of the intracellular symbiosis that would give rise to mitochondria and eukaryotes was a major event in the history of life on earth. Hypotheses to explain eukaryogenesis fall into two broad and competing categories: those proposing that the host was a phagocytotic proto-eukaryote that preyed upon the free-living mitochondrial ancestor (hereafter FMA), and those proposing that the host was an archaebacterium that engaged in syntrophy with the FMA. Of key importance to these hypotheses are whether the FMA was motile or nonmotile, and the atmospheric conditions under which the FMA thrived. Reconstructions of the FMA based on genome content of Rickettsiales representatives—generally considered to be the closest living relatives of mitochondria—indicate that it was nonmotile and aerobic. We have sequenced the genome of *Candidatus* Midichloria mitochondrii, a novel and phylogenetically divergent member of the Rickettsiales. We found that it possesses unique gene sets found in no other Rickettsiales, including 26 genes associated with flagellar assembly, and a $cbb_3$-type cytochrome oxidase. Phylogenomic analyses show that these genes were inherited in a vertical fashion from an ancestral α-proteobacterium, and indicate that the FMA possessed a flagellum, and could undergo oxidative phosphorylation under both aerobic and microoxic conditions. These results indicate that the FMA played a more active and potentially parasitic role in eukaryogenesis than currently appreciated and provide an explanation for how the symbiosis could have evolved under low levels of oxygen.

**Key words:** mitochondrion, symbiosis, eukaryogenesis, *Midichloria mitochondrii*, rickettsiales, phylogenomics.

## Introduction

The symbiosis that would ultimately give rise to mitochondria and their eukaryotic hosts is recognized as one of the major transitions in the history of life on earth (Margulis 1970; Maynard Smith and Szathmary 1997; de Duve 2005; Lane and Martin 2010). All eukaryotes examined thus far have been shown to contain mitochondria or modified versions of this organelle (hydrogenosomes or mitosomes) (Bui et al. 1996; Tovar et al. 2003; Embley and Martin 2006). In the wake of the genomic revolution, it is widely accepted that mitochondria arose only once from a free-living bacterium that took up residence in its host's cytoplasm (Embley and Martin 2006; Kurland et al. 2006; de Duve 2007). Recent phylogenomic studies indicate that this bacterium—the last free-living common ancestor of mitochondria (hereafter FMA)—was closely related to members of the Rickettsiales, within the α-proteobacteria (Fitzpatrick et al. 2006; Williams et al. 2007).

Despite these advances, consensus has yet to be reached over a number of issues concerning eukaryogenesis (Embley and Martin 2006; Poole and Penny 2006). Chief among these are: 1) the nature of the amitochondriate host; 2) how the FMA was engulfed; and 3) whether eukaryogenesis occurred under oxic, microoxic, or anoxic conditions. On the nature of the amitochondriate host,

hypotheses fall into two general categories. The first, more traditional, view is that the amitochondriate host contained most eukaryotic features, such as a nucleus, cytoskeleton, endomembrane system, and the ability to phagocytose (de Duve 2007; Cavalier-Smith 2009). This "proto-eukaryote" is considered by some authors to have been phylogenetically distinct from the lineages leading to extant archaea and eubacteria (Hartman and Fedorov 2002; Kurland et al. 2006; Poole and Penny 2006; Gribaldo et al. 2010). The second group of hypotheses posits the amitochondriate eukaryote host as an archaebacterium (Rivera and Lake 1992; Martin and Muller 1998; Vellai et al. 1998; Cox et al. 2008; Davidov and Jurkevitch 2009). Under these hypotheses, eukaryotic features evolved after the fusion of the FMA and its archaebacterial host.

On the issue of engulfment, models of eukaryogenesis typically assume that the FMA was nonmotile, playing a relatively passive role in engulfment (Andersson and Kurland 1999; Embley and Martin 2006) (see Figures therein). This is in agreement with the absence of flagella in close relatives of mitochondria (Williams et al. 2007), including all examined Rickettsiales (Andersson et al. 1998; Wu et al. 2004; Dunning Hotopp et al. 2006; Cho et al. 2007) and *Pelagibacter ubique* (Giovannoni et al. 2005). Proto-eukaryote models usually assume that the FMA acted as nonmotile prey for its phagocytic host (Cavalier-Smith 2009). This idea is consistent with the common occurrence of phagocytosis of prokaryotes by extant unicellular eukaryotes (Matz and Kjelleberg 2005; Kurland et al. 2006; Poole and Penny 2007). In the case of archaebacterial host models, explaining engulfment of the FMA is challenging due to the absence of phagocytosis in prokaryotes. However, some prokaryotes are known to exist within other prokaryotes (von Dohlen et al. 2001; Davidov and Jurkevitch 2009), indicating that engulfment may occur under some circumstances. An alternative to the general view of passive engulfment of the FMA is the view that the FMA displayed flagellar motility and played a more active role in eukaryogenesis. This idea has been widely overlooked, with a few rare exceptions (Guerrero et al. 1986; Davidov and Jurkevitch 2009).

A further unresolved issue is whether eukaryogenesis occurred under oxic, microoxic, or anoxic conditions. In a number of hypotheses, the FMA is considered to have been aerobic, rescuing, or detoxifying its initially anaerobic host from dramatically rising oxygen levels some 2.2 Gyr ago (Margulis 1970; Andersson and Kurland 1999; Dyall et al. 2004; de Duve 2005). This idea is consistent with the presence of similar respiratory chains in mitochondria and members of the Rickettsiales. A challenge, however, for rescue and "ox-tox" hypotheses is explaining how the FMA would have survived in the microoxic or anaerobic conditions under which its host presumably existed. In contrast with oxygen rescue hypotheses, some archaebacterial host hypotheses propose that the symbiosis originated under anaerobic conditions, for example, between a hydrogen-producing FMA and a hydrogen-utilizing archaeon (Martin and Muller 1998).

Phylogenomic studies of diverse eukaryotes continue to provide important insights into the nature of the eukaryotic ancestor (Dacks and Doolittle 2001; Slamovits and Keeling 2006; Brinkmann and Philippe 2007; Fritz-Laylin et al. 2010). Similarly, the genomes of numerous Rickettsiales and other α-proteobacteria have significantly advanced our understanding of the nature of the FMA (Andersson et al. 1998; Gabaldón and Huynen 2003; Darby et al. 2007). However, unlike the case of the eukaryotes, where representatives of all six main lineages have been sequenced (Fritz-Laylin et al. 2010), genome studies in the Rickettsiales have probably not properly sampled the entire diversity of this bacterial order (Darby et al. 2007). Recent 16S rDNA-based phylogenetic analyses show the Rickettsiales contains a number of novel lineages (Vannini et al. 2010), which are highly divergent from the two well-studied families, the Anaplasmataceae and the Rickettsiaceae. One such lineage contains the recently described *Candidatus* Midichloria mitochondrii (hereafter *M. mitochondrii*), so named because it is the only described bacterium able to enter the mitochondria of any multicellular organism (Lo et al. 2006; Sassera et al. 2006) (fig. 1*A* and *B*). *Midichloria mitochondrii* is an intracellular symbiont of the tick *Ixodes ricinus*; phylogenetically related endosymbionts have been found in other tick species and other invertebrates (Epis et al. 2008; Beninati et al. 2009). In *I. ricinus*, in situ hybridization and electron microscopic studies indicate that the bacterium is primarily restricted to the cells of the ovary (Zhu et al. 1992; Beninati et al. 2004; Sacchi et al. 2004). The role of *M. mitochondrii* in host biology is unclear, but various lines of evidence suggest that it is a facultative mutualist. The prevalence of the symbiont is 100% in female *I. ricinus* ticks collected in the field. This is typical of a mutualist or of a manipulator of the host reproduction. However, the fact that the 100% prevalence is observed throughout the entire geographical distribution of *I. Ricinus*, and the lack of evidence for alterations of the reproduction in this tick species argue against the second hypothesis. Furthermore, the loss of the symbiont in tick colonies maintained in the laboratory indicates that the symbiosis is not obligate, suggesting a facultative mutualism (for a complete description of the known characteristics of *M. mitochondrii*, see Lo et al. 2006; Epis et al. 2008; Sassera et al. 2008).

Here, we report the complete genome sequence of *M. mitochondrii* and the results of phylogenomic comparisons with related bacteria as well as mitochondria. We focus on two novel gene complements present in *M. mitochondrii*, which have key implications for our understanding of eukaryogenesis and for addressing the unresolved issues discussed above.

## Materials and Methods

### Transmission Electron Microscopy

Twenty-five females of the hard tick *I. ricinus* were collected from different hosts (dogs, roe deer, and sheep) in different areas of Italy (counties of Novara, Varese, Parma, Trento, Ascoli Piceno). Ovaries were dissected in saline solution, then fixed, postfixed, dehydrated, and embedded in Epon 812 resin as described (Epis et al. 2008). Semi- and ultrathin

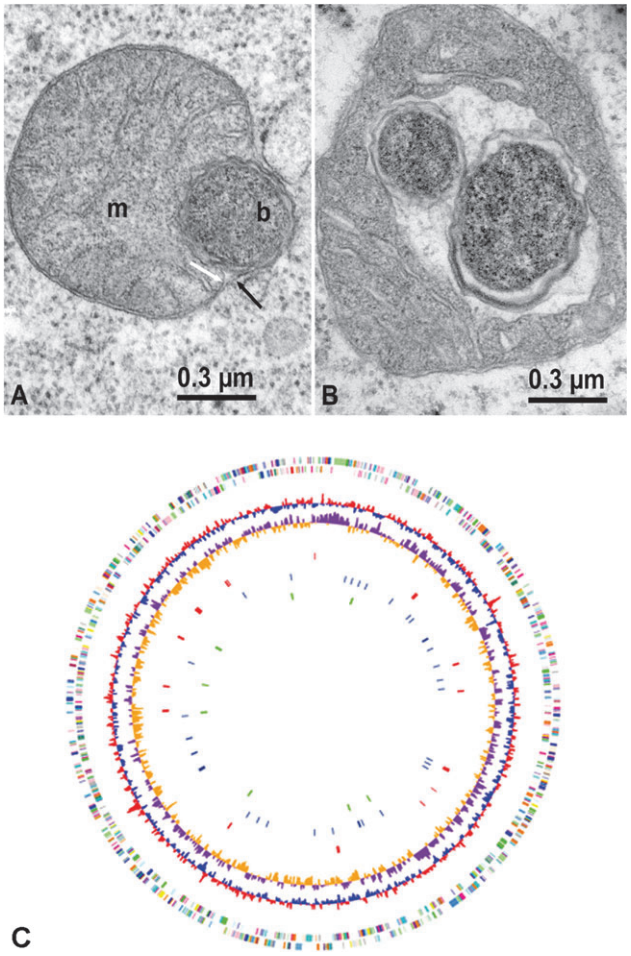


FIG. 1. Transmission electron microscopic images of *Midichloria mitochondrii* in the mitochondria (*A* and *B*) and circular genome map (*C*). (*A*) Initial step of infection showing a bacterium (b) between the outer (black arrow) and inner (white arrow) membranes of the mitochondrion (m). (*B*) A mitochondrion harboring two bacteria; the matrix appears to be partially consumed. (*C*) The two outermost circles show all predicted coding regions respectively on the plus and minus strand, divided by color based on COG functional groups; For a complete color-code explanation, see supplementary figure S5, Supplementary Material online. The third circle shows the GC content, and the fourth circle shows the GC skew. The fifth circle shows the location of the flagellar genes in red, the sixth circle shows the structural RNAs in blue, and the innermost circle shows the putative pseudogenes in green.

sections were stained and examined under optical and transmission electron microscopes.

## DNA Purification and Sequencing

A semiengorged *I. ricinus* adult female was collected from a dog in the county of Varese (Italy), and *M. mitochondrii* DNA was purified and amplified as described previously (Epis et al. 2010). Briefly, following tick washing, the ovary was dissected under a stereomicroscope and single oocytes were mechanically detached. Increasing quantities of collagenase A (Sigma-Aldrich, St. Louis, MO) 20 mg/ml diluted in 50% sterile water and 50% phosphate-buffered saline were added to cause osmotic breakage of the membrane of the oocytes. Twenty pools of 10–12 cytoplasms were collected using a microcapillary, avoiding the nuclei trapped in the remnants of the cytoplasmic membranes. Cytoplasmic preparations were stained with Hoechst 33342, to discard preparations containing undesired nuclei. The remaining 15 cytoplasm preparations were used as templates for multiple displacement amplification (MDA) using Repli-g Mini Kit (Qiagen, Hilden, Germany). MDA products were purified using a QIAEX II Gel Extraction Kit (Qiagen) and used as templates for a previously described set of real-time polymerase chain reactions (PCRS) to assess whether the genome of *M. mitochondrii* was amplified with limited biases and to exclude the possible residual presence of detectable *I. ricinus* nuclear DNA contamination. For a detailed description of primers and PCR conditions, see Epis et al. (2010).

Thirteen preparations showed uniform amplification of the seven *M. mitochondrii* loci, and two of these exhibited no sign of *I. ricinus* nuclear DNA contamination. These two MDA products were pooled and used as template for pyrosequencing and for 1–3 kbp plasmid library construction, to be sequenced using Sanger-based technology. One half GS-FLX Titanium run and one quarter 3 kb paired-ends GS-FLX run were performed, netting a total of 544,141 and 156,393 reads, respectively. Sanger sequences were obtained from both ends of 856 clones.

## Assembly and Annotation

Data resulting from the two pyrosequencing runs and the Sanger sequences were used for assembly using the Mira assembler v2.9 (Chevreux et al. 2004), obtaining 196 large contigs (from 2,005 to 124,307 bp). Contigs were manually checked and joined with gap4 (Staden et al. 2000) and thus reduced to 154 contigs. A scaffold was then generated using Bambus (Pop et al. 2004), and the scaffold was used as a backbone for a second assembly with Mira to obtain 38 large contigs (from 13,459 to 424,616 bp). Although the Mira assembler provides an algorithm for chimera recognition and removal, all contigs were visually examined in gap4, with specific attention to all low coverage regions ($<5\times$). Manual correction was performed to close sequence gaps. Eighteen remaining gaps were then closed by PCR and inverse PCR as described by Hartl and Ochman (1996).

Putative coding regions were identified with Glimmer v3.02 for protein-coding genes (Delcher et al. 1999), tRNAscan-SE for transfer RNAs (Schattner et al. 2005), and BlastN for ribosomal RNAs. All putative protein-coding genes were analyzed in local with blastP and rpsblast algorithms searching the pfam, eggnog, nr, uniprot, and omniome databases. Results from the different databases were compared and manually curated in order to assign gene functions. Forty-seven regions of interest (possible pseudogenes, flagellar genes, $cbb_3$ genes, genes coding for surface proteins) were amplified using specific primers on genomic DNA not subjected to MDA and then sequenced using the Sanger method. Metabolic pathways were reconstructed using the KEGG automated annotation server KAAS (Kanehisa and Goto 2000) and manually checked. Sequences of 112

regions of interest (possible pseudogenes, all flagellar genes, all *cbb3* genes, genes coding for surface proteins, genes containing homopolymers) were reobtained by Sanger sequencing after specific PCR amplification from DNA not subjected to MDA. All of the sequences generated after PCR and Sanger sequencing were identical to the previous consensus, with the exception of six homopolymeric regions that were corrected for the generation of the consensus sequence. Expression of five flagellar genes and three $cbb_3$ genes was investigated by reverse transcription PCR on cDNA obtained from *I. ricinus* adult females and egg batches.

### Phylogenetics

For phylogenomic analysis, an alignment of 88 conserved genes from 72 organisms, previously used in a global $\alpha$-proteobacteria phylogeny (Williams et al. 2007), was kindly provided by K.P. Williams, Sandia National Laboratories, Livermore, California. Orthologous *M. mitochondrii* genes were then added to the aforementioned alignment. For the study of the flagellum phylogeny, two data sets were constructed, one with 14 core flagellar genes and one with 24 conserved bacterial genes (as described in Liu and Ochman 2007), using alignments kindly provided by R. Liu, University of California, Riverside. These alignments were modified via the addition of extra proteobacterial representatives, including *M. mitochondrii* orthologous genes, for a total of 24 $\alpha$-, nine $\beta$-, 13 $\gamma$-, seven $\delta$-, six $\varepsilon$-proteobacteria, and three outgroups. A fourth data set was prepared using 12 highly conserved bacterial proteins that have homologs in mitochondrial genomes (as described in Williams et al. 2007), retrieved from 55 taxa, including 7 outgroups, 8 mitochondria, and 14 members of the Rickettsiales. $cbb_3$ phylogenetic analyses were performed on an alignment of the three genes coding for the fundamental subunits of this complex (*ccoN*, *ccoO*, and *ccoP*) for 35 taxa, with the addition of *M. mitochondrii*.

For all phylogenetic analyses, genes were aligned after addition of the *M. mitochondrii* orthologs using Muscle (Edgar 2004). Alignments were trimmed with Gblocks (Castresana 2000) with optimization of the parameters for each single-gene alignment, based on the number of taxa for which the amino acid sequence was present: $-b1 = (N/2) + 1$; $-b2 = (N/2) + 1$; $-b3 = (N/2)$; $-b4 = 2$; $-b5 =$ h, with $N =$ number of taxa. Each alignment was analyzed with ProtTest 2.1 (Abascal et al. 2005) to choose the most suitable model of evolution. PhyML version 3.0 (Guindon and Gascuel 2003) and MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001) were then used for construction of maximum likelihood (ML) and Bayesian analysis trees, respectively, both on single flagellar genes and on the concatenated gene alignments of the four data sets. PhyML was used to construct ML-based phylogenetic trees with nonparametric boostrap analysis (100 boostrap replicates), with the following options: Le and Gascuel (LG) as the amino acid substitution model; optimized proportions of invariable sites; amino acid frequency estimation counting the number of different amino acids observed in the data; optimized rate variation across sites into eight substitution rate categories; the best of NNI and SPR tree searching operation; and computation of the starting tree with BioNJ. Bayesian analyses were performed on the partitioned concatenated alignment to allow parameter optimization for each gene. All the Markov Chain Monte Carlo analyses were implemented into two runs with four chains each (one hot and three cold). Each analysis had samplefreq=10. We used Whelan and Goldman substitution model + proportions of invariable sites (I) and gamma ($\Gamma$) with eight categories. The convergence of each run was verified with Tracer 1.4 (Drummond and Rambaut 2007).

In order to investigate the influence of base composition on our phylogenetic reconstructions, we performed a Bayesian phylogenetic analysis of $\alpha$-proteobacterial relationships under a model of heterogeneous composition as implemented in the software P4 (Foster 2004). These analyses recovered topologies that were entirely consistent with the homogeneous Bayesian and ML analyses presented in the figures (results not shown).

### Genome Analysis

Amino acid sequences derived from the genomes of *Wolbachia* endosymbiont of *Drosophila melanogaster*, *Rickettsia prowazekii* str. Madrid, *R. conorii*, *R. bellii* OSU 85-389, *Orientia tsutsugamushi* str. Boryong, *Anaplasma phagocytophilum* HZ, and *Ehrlichia chaffensis* were downloaded from ftp://ftp.ncbi.nih.gov/genomes/Bacteria/. Two sets of four genomes were subjected to a five-way comparison with the genome of *M. mitochondrii* by OrthoMCL v1.4 (Li et al. 2003) using default parameters, in order to find orthologous groups conserved between the Rickettsiales. Clusters of Orthologous Groups (COG) functional group assignment was obtained for *M. mitochondrii* and four other Rickettsiales proteins by local rpsblast on the COG database. A graphic representation of the *M. mitochondrii* genome was constructed using the software genomeviz (Ghai et al. 2004), using the COG functional group assignment, and other features extracted from the genome annotation. Metabolic pathways for the synthesis of vitamins and cofactors were manually inferred from the annotation of the *M. mitochondrii* genome. Repeat content in the *M. mitochondrii* genome was assessed using the ugene repeat finder tool and compared with the repeat content of other selected Rickettsiales. Codon usage bias analysis was performed using the local version of CAIcalc (Puigbo et al. 2008) using default parameters, in order to calculate the codon adaptation index (CAI) of all the protein-coding genes of the *M. mitochondrii* genome and to evaluate if the CAI values of $cbb_3$ and flagellar genes are statistically divergent from the CAI expected.

## Results and Discussion

### Characteristics of the *M. mitochondrii* Genome and Its Phylogenetic Relationship to Other $\alpha$-Proteobacteria and Mitochondria

The genome of *M. mitochondrii* consists of a single 1,183,732 bp circular chromosome with a G+C content of 36.6% (fig. 1C). Most of the genome content of *M. mitochondrii* is typical of other members of the Rickettsiales (Andersson et al. 1998; Wu et al. 2004; Dunning Hotopp et al. 2006)

**Table 1.** General Genome Properties of Selected Rickettsiales.

| | Organisms | | | | |
|---|---|---|---|---|---|
| **Properties** | **MID** | **APH** | **ECH** | **RPR** | **WME** |
| **Genome size** | **1,183,732** | **1,471,282** | **1,176,248** | **1,111,523** | **1,267,782** |
| **GC (%)** | **36.6** | **41.6** | **30.1** | **29.1** | **35.2** |
| **ORFs** | **1,245** | **1,369** | **1,115** | **834** | **1,271** |
| **Transfer RNAs** | **35** | **37** | **37** | **33** | **34** |
| **Ribosomal RNAs** | **3** | **3** | **3** | **3** | **3** |
| **Average gene length** | **698** | **775** | **840** | **1,005** | **855** |
| **Coding (%)** | **73.7** | **72.2** | **79.9** | **75.4** | **85.7** |
| **Assigned function** | **634** | **747** | **604** | **523** | **719** |
| **Conserved hypothetical** | **123** | **82** | **111** | **NR** | **123** |
| **Hypothetical** | **391** | **458** | **314** | **208** | **337** |

NOTE.—MID, *Midichloria mitochondrii*; APH, *Anaplasma phagocytophilum* HZ; ECH, *Ehrlichia chaffensis*; RPR, *Rickettsia prowazekii* str. Madrid; and WME, *Wolbachia* endosymbiont of *Drosophila melanogaster*; NR, not reported; ORFs, open reading frames.

(table 1)(supplementary fig. S1, Supplementary Material online). Thus, *M. mitochondrii* possesses a relative scarcity of genes encoding amino acid and nucleotide biosynthesis pathways, compared with free-living α-proteobacterial relatives. Although *M. mitochondrii* has diminished biosynthetic capabilities, like many other Rickettsiales (Dunning Hotopp et al. 2006), it does possess genes for the production of several cofactors, including coenzyme A, biotin, lipoic acid, tetrahydrofolate, panthotenate, heme, and ubiquinone (fig. 2). *Midichloria mitochondrii* may supply host cells with these essential cofactors. *Midichloria mitochondrii* is inferred to have a functional Krebs cycle, gluconeogenesis pathway, and pyruvate dehydrogenase complex, and almost all enzymes required for glycolysis (supplementary fig. S2, Supplementary Material online). *Midichloria mitochondrii* is thus able to synthesize ATP, and the presence of a gene coding for an ATP/ADP translocase indicates that it may also be able to import/export ATP from/to the host.

*Midichloria mitochondrii*'s genome contains a high number of imperfect repeats, possibly due to a past proliferation of mobile elements (supplementary table S1, Supplementary Material online). However, few intact genes associated with mobile and extrachromosomal functions were found. Most identified sequences of this nature (phage capsid proteins, transposases) appear to be truncated pseudogenes. Like other Rickettsiales, *M. mitochondrii* possesses genes encoding type IV and Sec-independent protein secretion systems, ankyrin repeat proteins, and various putative membrane-associated proteins. These genes may be associated with the symbiont's unique ability to invade host mitochondria. Overall, *M. mitochondrii*'s genome does not provide a clear answer to the question of whether it engages in a mutualistic or a parasitic relationship with its tick host. Targeted experimental studies (e.g., examining the fitness of the host with and without the symbiont, and functional genomic studies) are required to address this question.

Phylogenetic analyses of an alignment of 88 conserved genes from 66 α-proteobacterial representatives plus six outgroups confirmed the divergent phylogenetic position of *M. mitochondrii* among the Rickettsiales (fig. 3A). We found high support for *M. mitochondrii* forming a sister group with the Rickettsiaceae (which includes the

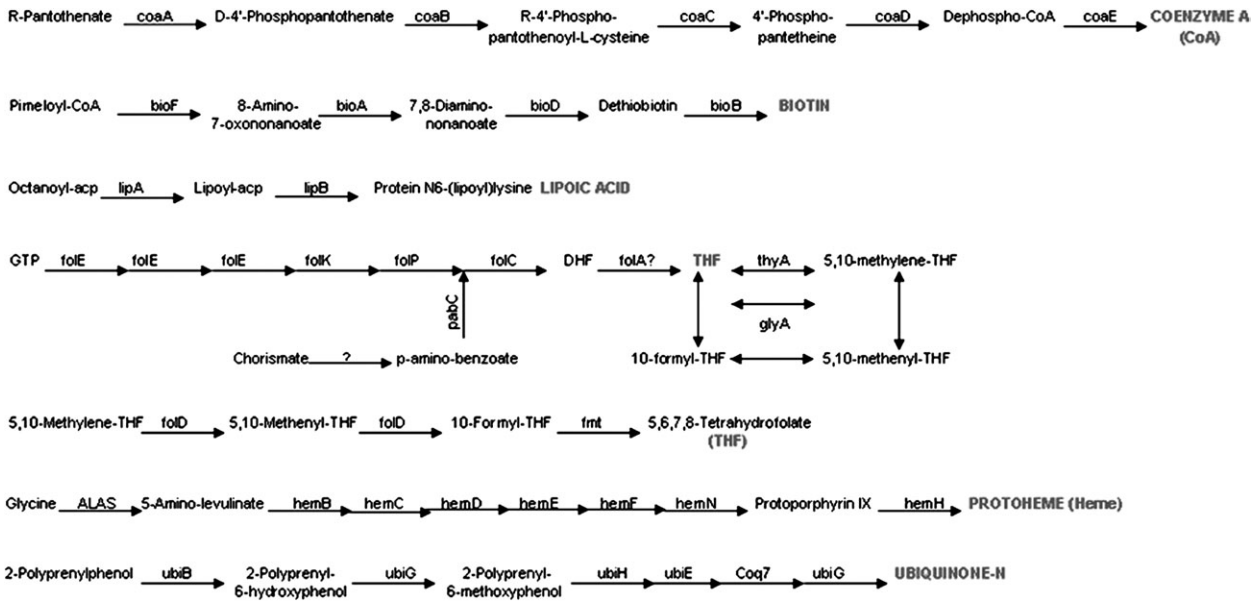


**FIG. 2.** Vitamin and cofactor metabolic pathways present in *Midichloria mitochondrii*, manually inferred from the annotation of the genome.

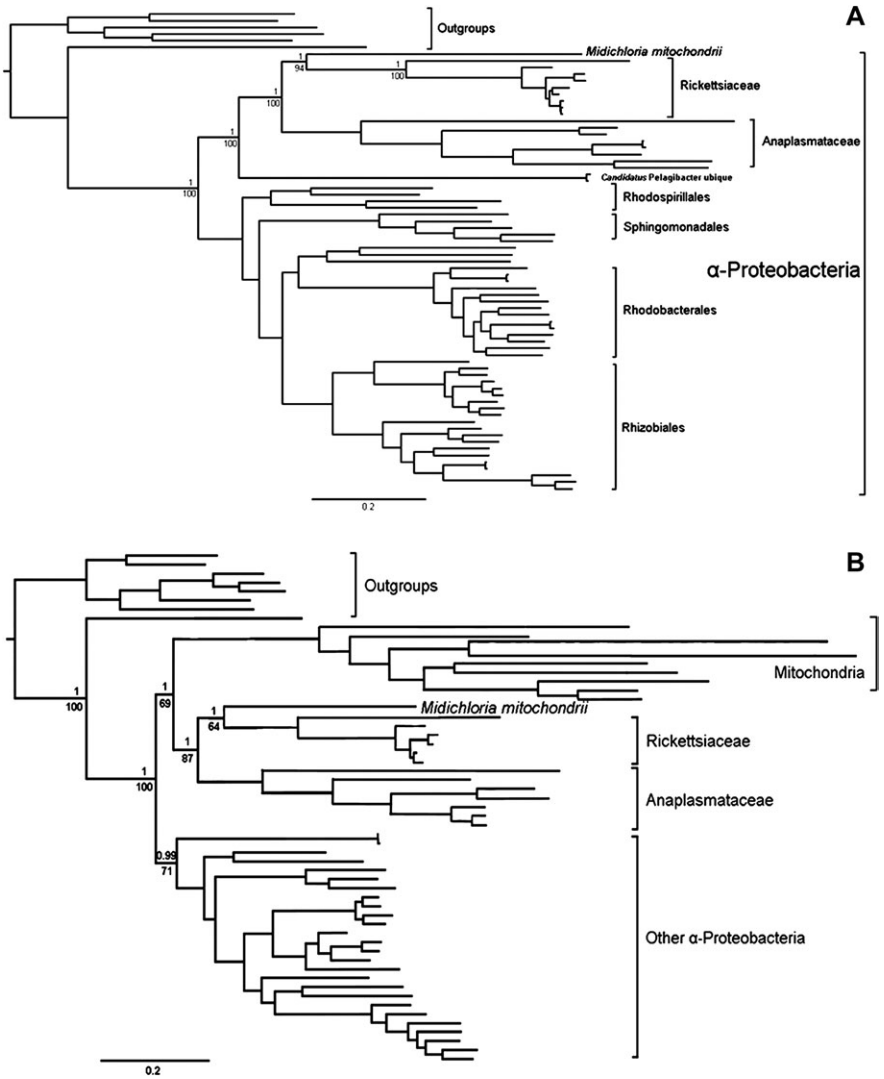


**Fig. 3.** Phylogenetic analysis based on conserved proteins from (*A*) diverse representatives of the α-proteobacteria (88 single-copy proteins) and (*B*) α-proteobacteria and mitochondria (12 single-copy proteins). Concatenated alignments were analyzed using MrBayes (topology shown) and PhyML. Posterior probabilities and PhyML bootstrap supports are respectively shown above and below nodes of interest. In each case, *Midichloria mitochondrii* is placed as the sister group of the Rickettsiaceae (a family within the order Rickettsiales). In *B*, the Rickettsiales are placed as sister group to mitochondria. For simplicity, trees are shown without taxon names; more detailed trees, with names of the taxa, are provided in the supplementary information—supplementary figure S3, Supplementary Material online.

genera *Rickettsia* and *Orientia*). The analyses also confirmed the early branching position of the Rickettsiales relative to other orders among the α-proteobacteria (Dunning Hotopp et al. 2006; Williams et al. 2007). We performed additional phylogenetic analyses to examine the relationship of *M. mitochondrii* to mitochondria. Orthologous genes from *M. mitochondrii* were added to a previous alignment of 12 genes (Williams et al. 2007) present both in large mitochondrial genomes and diverse α-proteobacteria, including 14 representatives of Rickettsiales. *Midichloria mitochondrii* was found as the sister group to the Rickettsiaceae in the 12 gene analysis (fig. 3*B*), equivalent to its position in the 88 gene analysis. The Rickettsiales were placed as the sister group of the mitochondria, which together formed a deep branch within the α-proteobacteria, again in accordance with figure 3*A*.

## Genes Putatively Encoding Flagellar Proteins in the *M. mitochondrii* Genome

Unexpectedly, we found 26 genes in the *M. mitochondrii* genome that putatively encode a flagellum, including all key components, such as the hook, filament, and basal body. Such genes are found in none of the other ~20 Rickettsiales genomes sequenced thus far (Darby et al. 2007), nor are they found in *P. ubique*, which has phylogenetic affinities with the Rickettsiales (Williams et al. 2007). Electron microscopic examinations of *M. mitochondrii* in the ovarian cells of 25 wild-collected ticks revealed no evidence for a flagellum (fig. 1*A* and *B*), confirming the results of previous studies (Zhu et al. 1992; Sacchi et al. 2004). The function of these genes is, therefore, unclear. There is evidence that *M. mitochondrii* is transmitted horizontally among ticks, most likely via their vertebrate hosts (Epis et al. 2008). The flagellum might be assembled in as-yet-unexplored stages of the bacterium's life

cycle. It may also play a role in the invasion of tick mitochondria by *M. mitochondrii*.

Each of the flagellar genes is of a similar length to orthologs in other eubacteria and contains no stop codons. Genomic and reverse transcriptase PCR assays on a subset of these genes (*fliC*, *fliD*, *flgL*, *flgK*, *flgE*) in multiple *I. ricinus* specimens confirmed their uniform presence in *M. mitochondrii* and indicated that they are transcribed (data not shown). Thus, there is no indication that they are pseudogenes. GC content of the 26 flagellar genes found in the *M. mitochondrii* genome ($35.6 \pm 2.3\%$) was not significantly different to the GC content of other *M. mitochondrii* protein-coding genes ($36.8 \pm 4\%$; Mann–Whitney *U* test, $P = 0.137$), indicating that the flagellar genes do not derive from recent lateral gene transfer. Additionally, CAI analysis (Puigbo et al. 2008) indicated no statistically significant difference between the expected CAIs for flagellar genes and all other genes ($P < 0.05$).

We performed phylogenetic analysis to investigate the evolutionary origin of *M. mitochondrii*'s flagellar genes. Based on comparisons of phylogenetic trees inferred from 14 concatenated conserved flagellar genes with those inferred from 24 core eubacterial genes, Liu and Ochman (2007) previously demonstrated that flagellar genes have been inherited in a largely vertical fashion throughout eubacterial diversification. We repeated these analyses on expanded gene data sets, including *M. mitochondrii* and additional proteobacterial representatives. Our analyses revealed the same pattern of vertical inheritance of flagellar genes found by Liu and Ochman (2007). Thus, *M. mitochondrii* is placed as the deepest branch relative to flagellated $\alpha$-proteobacteria, both in trees inferred from flagellar genes (fig. 4*A*) and in trees inferred from core genes (fig. 4*B*). For the 14 flagellar single-gene analyses, all nodes with >75% and 0.8 posterior probability values (PhyML and MrBayes, respectively) agreed with those in the concatenated tree, indicating that the examined flagellar genes have followed a common evolutionary pattern since their origin. To test the robustness of the position of *M. mitochondrii* in the concatenated PhyML tree, we performed the SH (Shimodaira and Hasegawa 1999), two sided and one sided KH (Kishino and Hasegawa 1989; Goldman et al. 2000), and expected likelihood weight (Strimmer 2002) statistical tests implemented in the program TreePuzzle (Schmidt et al. 2002). We forced the *M. mitochondrii* position with either the outgroups, the $\gamma$-proteobacteria, or the $\beta$-proteobacteria. All the performed statistical tests indicated that the original topology was a significantly better explanation of the data than the alternative solutions ($P < 0.05$).

The position of *M. mitochondrii* in figure 4 is in accordance with the position of the Rickettsiales as an early branch of $\alpha$-proteobacteria (fig. 3) and indicates that *M. mitochondrii* has inherited its flagellar genes vertically from the ancestor of the $\alpha$-proteobacteria. Figure 4 also shows that the ancestor of all $\alpha$-proteobacteria possessed a flagellum, as demonstrated previously on the basis of genome content comparisons (Boussau et al. 2004). This result, taken together with the results shown in figure 3, indicates that flagellar genes were present in the ancestor of the Rickettsiales and mitochondria (fig. 5). Flagellar genes are known to be lost upon transition from the free-living state to the intracellular state (Toft and Fares 2008). We infer that such loss occurred during the transition of the FMA to the intracellular state. Independent loss of flagellar genes also occurred in the lineages leading to present day Rickettsiaceae and Anaplasmataceae (fig. 5).

## Genes Putatively Encoding a $cbb_3$ Oxidase in the *M. mitochondrii* Genome

A second set of genes present in *M. mitochondrii*, but in no other examined members of the Rickettsiales, is predicted to encode a cytochrome $cbb_3$ oxidase, and proteins associated with its assembly. The $cbb_3$ oxidases belong to the C-family of heme-copper oxidases (HCOs). C-family HCOs are phylogenetically divergent from the more common A-family HCOs ($aa_3$-type), which are widespread in prokaryotes and mitochondria, and from the B-family HCOs ($bo_3$-type or quinol oxidases), which are present mainly in crenarcheota (Ducluzeau et al. 2008; Buschmann et al. 2010). $cbb_3$ oxidases reduce $O_2$ with lower efficiency compared with A- and B-family HCOs but have a higher affinity for it. This characteristic is used by pathogenic proteobacteria that infect microaerobic host tissues and by symbiotic diazotrophs that are able to simultaneously undertake aerobic respiration and nitrogen fixation (which requires an oxygen-sensitive nitrogenase) (Pitcher and Watmough 2004). In ticks, oogenesis leads to a dramatic increase in oxygen use (Aboul-Nasr and Bassal 1972). The presence of a *cbb3* oxidase may enable *M. mitochondrii* to synthesize ATP at oxygen concentrations that are suboptimal for the mitochondrion. We might thus speculate that *M. mitochondrii* could serve as an additional ATP source for the host cell during oogenesis. Transfer of ATP to the host cell might occur through the ATP/ADP translocase of *M. mitochondrii*, as proteins of this family are known to have a reversible function, based on ATP/ADP concentrations and proton gradient.

The genes *ccoN*, *ccoO*, and *ccoP*, which encode the three fundamental subunits of $cbb_3$, are found together in *M. mitochondrii* in the same order as *Pseudomonas stutzeri* strain Zobell (Buschmann et al. 2010), though the *M. mitochondrii ccoN* gene is split into two open reading frames. Each of the key histidine and methionine residues associated with coordination of heme irons are present (Buschmann et al. 2010). Two additional genes, *coo*I and *ccoG*, reported to be accessory proteins for assembly of the $cbb_3$ complex, are also present in the genome. Reverse transcription PCR assays demonstrated that *ccoN*, *ccoO*, and *ccoP* are transcribed (data not shown). GC content and CAI analysis were performed for $cbb_3$ genes as described above for flagellar genes; no evidence for recent horizontal transfer of these genes was found (GC of $cbb_3$ genes: $39.9 \pm 1.6\%$, Mann–Whitney *U* test, $P = 0.161$; CAI test $P < 0.05$).

The $cbb_3$ oxidases are distributed primarily in the proteobacteria and are believed to have evolved in an early progenitor of this phylum or perhaps earlier (Ducluzeau et al. 2008). All known $\alpha$-proteobacterial $cbb_3$ oxidase were previously shown to form a monophyletic clade among other proteobacterial clades, suggesting that they have

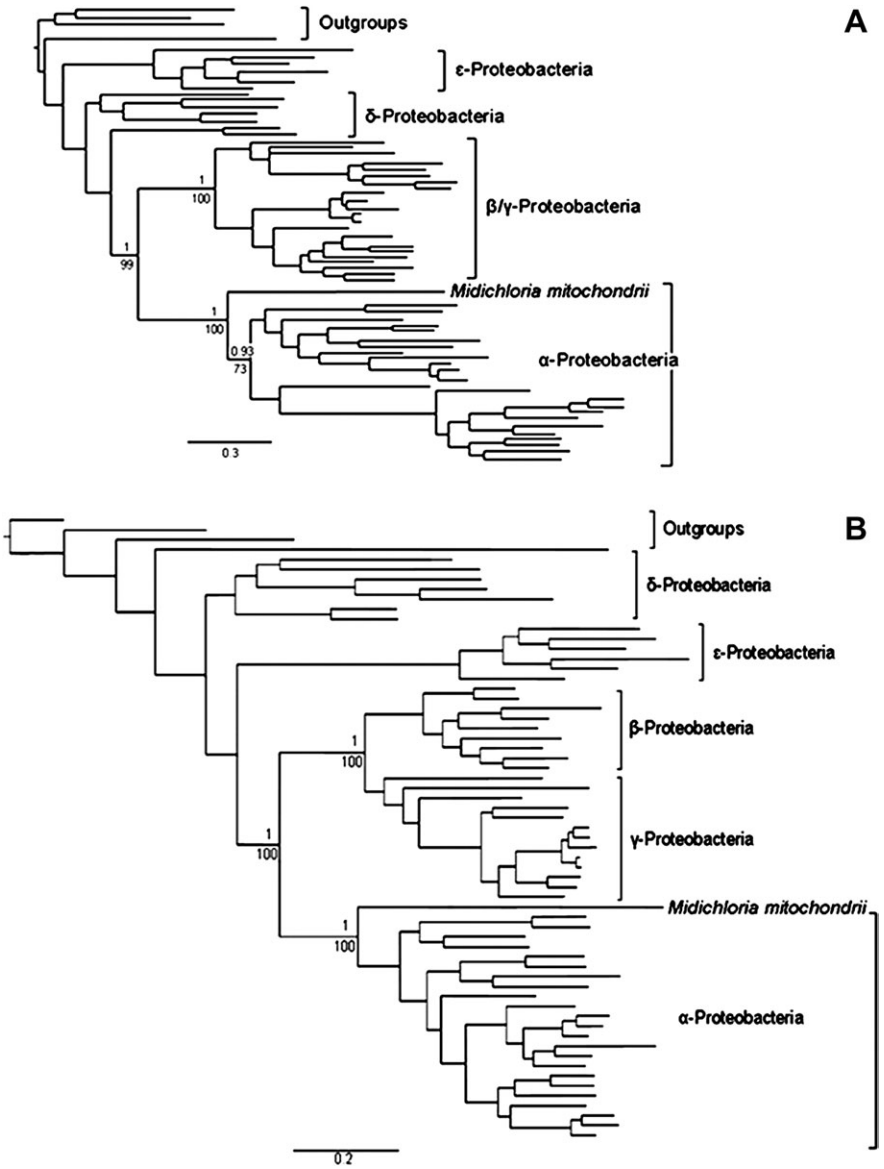


**Fig. 4.** Phylogenetic analysis of 14 conserved flagellar proteins (*A*) and 24 core proteins (*B*) from diverse representatives of the proteobacteria phylum. Concatenated alignments were analyzed using MrBayes (topology shown) and PhyML. Posterior probabilities and PhyML bootstrap supports are respectively shown above and below nodes of interest. *Midichloria mitochondrii* is placed as the deepest branch of the α-proteobacteria in each case. For simplicity, trees are shown without taxon names; more detailed trees are provided in the supplementary information—supplementary figure S4, Supplementary Material online.

been inherited vertically from an α-proteobacterial ancestor. Our phylogenetic analyses of *ccoN*, *ccoO*, and *ccoP* confirmed these findings and showed that the genes from *M. mitochondrii* represent the deepest branch among the α-proteobacteria (fig. 6), in accordance with the results from the flagellar and core gene phylogenies (fig. 4). These results indicate that this gene set, like the flagellar gene set, was inherited vertically from the α-proteobacterial ancestor to *M. mitochondrii*. We therefore infer that the ancestor of Rickettsiales and the ancestor of the FMA possessed a $cbb_3$ oxidase gene set (fig. 5). These genes were later lost during the evolution of mitochondria and in the lineages leading to the Rickettsiaeceae and the Anaplasmataceae.

Most α-proteobacteria that possess $cbb_3$ oxidases also have the more common $aa_3$ oxidase; the presence of both gene sets thus appears to be a primitive trait. Interestingly, although *M. mitochondrii* does not contain a complete $aa_3$ oxidase, it does contain one $aa_3$ oxidase gene fragment with high similarity to $aa_3$ oxidases in other Rickettsiales. Thus, the lineage leading to *M. mitochondrii* may have lost the complete set of $aa_3$ oxidase genes relatively recently. The $aa_3$ oxidases present in the ancestor of the Rickettsiales and the FMA would eventually evolve into the mitochondrial cytochrome oxidase IV complex (fig. 5).

## A Flagellum and $cbb_3$ Oxidase in the FMA: Implications for Eukaryogenesis

Hypotheses of eukaryogenesis have generally assumed the absence of a flagellum in the FMA (Andersson and Kurland

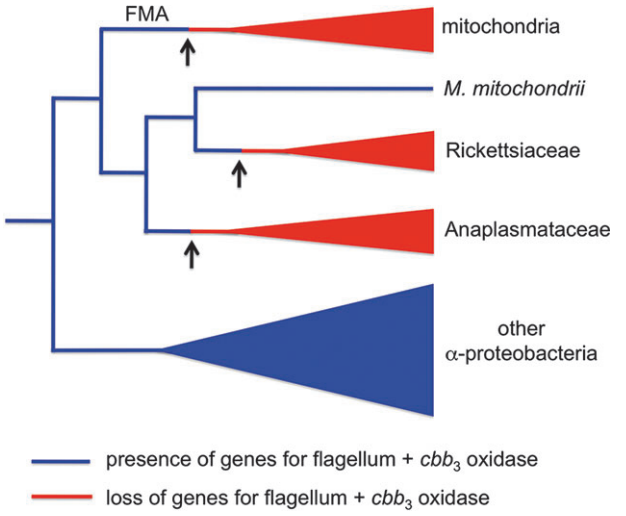


**Fig. 5.** Presence of flagellar and $cbb_3$ oxidase genes in the ancestor of the α-proteobacteria and inferred losses (indicated by arrows and red) in lineages leading to mitochondria, Rickettsiaceae, and Anaplasmataceae. The scheme is based on the results of figures 3, 4, and 6. The FMA is inferred to have possessed flagellar and $cbb_3$ oxidase genes.

1999; Embley and Martin 2006; Cavalier-Smith 2009). In proto-eukaryote host hypotheses, the FMA is considered to have acted as passive prey to a phagocytic predatory host. This idea is appealing due to the numerous extant examples of single-celled eukaryotes that feed on prokaryotes and is also in accord with the universally accepted model of chloroplast evolution via the phagocytosis of a member of the cyanobacteria (whose members all lack flagella) (Douglas 1998). The results presented here represent the first evidence that the FMA possessed a flagellum and thus challenge this assumption. A FMA with flagellar motility is likely to have played more active role in eukaryogenesis than currently appreciated. For example, feeding by some single-celled phagotrophic eukaryotes is dependent upon motility in their prokaryotic prey (Fenchel 1987). Heliozoan predators extend their axopods in all directions and wait for motile prey cells to collide with them. Flagella are also known to play a key role in adhesion and engulfment in many kinds of eukaryotic cells from protist to mammalian (Mahenthiralingam and Speert 1995; Ottemann and Miller 1997; Inglis et al. 2003; Ramos et al. 2004). In addition to the traditional idea that the FMA acted as prey, the possibility that the FMA acted as a parasite of its amitochondriate host should also be considered. Parasitism of protists by motile prokaryotes is a common phenomenon and opens up opportunities for facultative mutualism to evolve.

In the case of archaebacterial host hypotheses, the presence of a flagellum in the FMA may help to explain host entry (in the absence of any phagocytotic mechanism known in prokaryotes). The reemerging "predatory hypothesis" (Davidov and Jurkevitch 2009) for eukaryogenesis is based on the ability of some extant flagellated proteobacteria to penetrate the periplasm or cytoplasm of other bacteria. However, long-term persistence of such bacteria in their "prey" (other than as resting bdellocysts) awaits

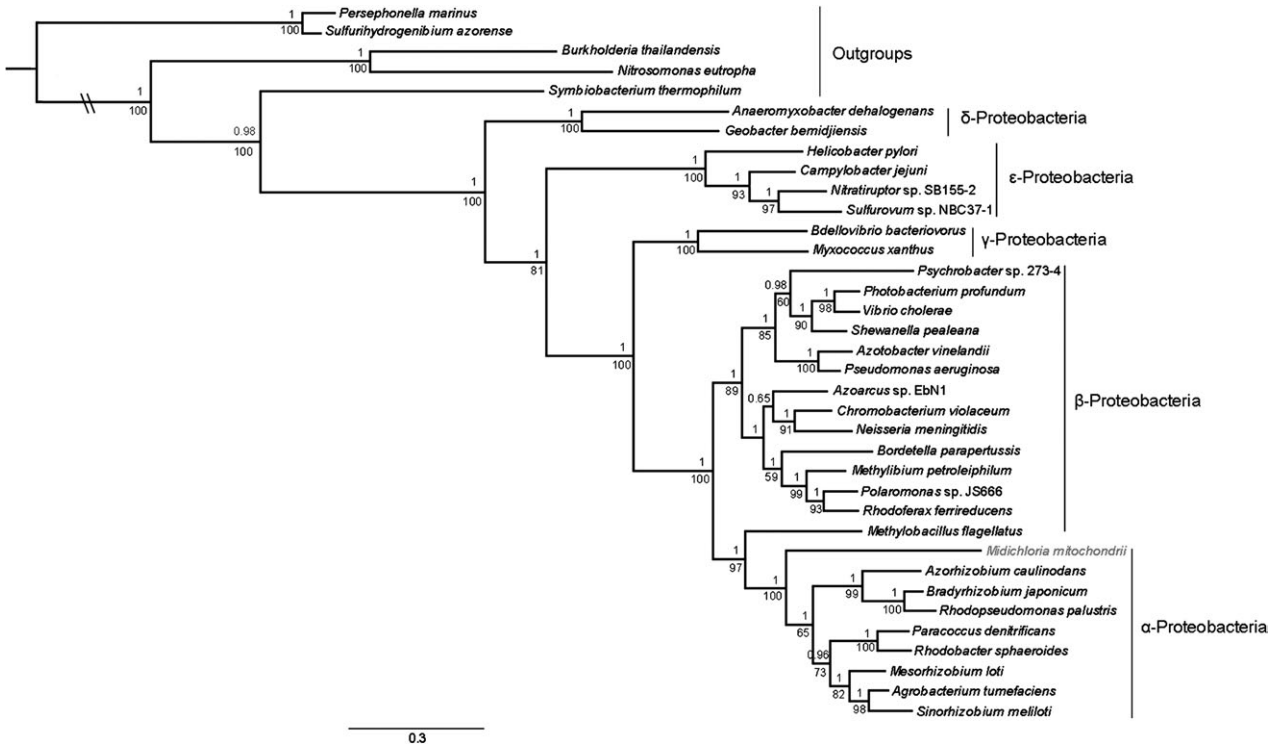


**Fig. 6** Phylogenetic analysis of $cbb_3$ proteins (*ccoN*, *ccoO*, and *ccoP*) from diverse representatives of the proteobacteria phylum. Concatenated alignments of the three proteins were analyzed using MrBayes (topology shown) and PhyML. *Midichloria mitochondrii* is placed as the deepest branch of the α-proteobacteria. Posterior probabilities and PhyML bootstrap supports are respectively shown above and below nodes.

discovery, as does a case of an archaean representative acting as prey or host.

Phylogenomic analysis of $cbb_3$ oxidase genes from *M. mitochondrii* enables to us to infer, for the first time, the presence of a $cbb_3$ oxidase in the FMA. The presence of this enzyme complex in the FMA has important implications for understanding the ecological context in which eukaryogenesis occurred. Oxygen rescue and ox-tox hypotheses propose that the FMA played a key role in allowing their anaerobic amitochondriate eukaryote hosts to adapt to the oxygenation of the earth which occurred some 2.2 Gyr ago (Margulis 1970; Andersson and Kurland 1999; Kurland and Andersson 2000). The presence of a $cbb_3$ oxidase helps to explain how the FMA could have survived in the microoxic conditions under which the interaction began. Following the establishment of the symbiosis and the adaptation of eukaryotes to aerobic conditions, $cbb_3$ oxidase genes would have become less important and eventually lost. The presence of a $cbb_3$ oxidase in the FMA is also relevant to anaerobic eukaryogenesis models (Martin and Muller 1998), in that it helps to explain how the symbiont transitioned to anaerobic conditions.

The origin of the eukaryotic cell is one of the most mysterious and challenging problems facing modern biology. The results herein illustrate the importance of surveying the genomes of diverse relatives of the FMA, many of which probably remain to be discovered. Such genome surveys will allow testing of the hypotheses presented here and may shed further light on the nature of the FMA.

## Supplementary Material

Supplementary table S1 and figures S1–S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

## Acknowledgments

We thank K.P. Williams and R. Liu for providing gene alignments and Mara Mariconti for supporting laboratory activities. Work was supported by Miur Prin to Claudio Genchi, Miur Prin to C.B., and Ispesl to C.B. D.S., L.S., and C.B. thank Lynn Margulis for the stimulating discussions.

## References

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105.

Aboul-Nasr AE, Bassal TTM. 1972. Biochemical and physiological studies of certain ticks (Ixodoidea). Effect of mating, feeding, and oogenesis on oxygen consumption of *Hyalomma* (*H.*) *dromedarii* Koch (Ixodidae). *J Parasitol*. 58:828–831.

Andersson SG, Kurland CG. 1999. Origins of mitochondria and hydrogenosomes. *Curr Opin Microbiol*. 2:535–541.

Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski RM, Naslund AK, Eriksson AS, Winkler HH, Kurland CG. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396:133–140.

Beninati T, Lo N, Sacchi L, Genchi C, Noda H, Bandi C. 2004. A novel α-proteobacterium resides in the mitochondria of ovarian cells of the tick *Ixodes ricinus*. *Appl Environ Microbiol*. 70:2596–2602.

Beninati T, Riegler M, Vilcins I, Sacchi L, McFadyen R, Krockenberger M, Bandi C, O'Neill SL, Lo N. 2009. Absence of the symbiont *Candidatus* Midichloria mitochondrii in the mitochondria of the tick *Ixodes holocyclus*. *FEMS Microbiol Lett*. 299:241–247.

Boussau B, Karlberg EO, Frank AC, Legault BA, Andersson SGE. 2004. Computational inference of scenarios for α-proteobacterial genome evolution. *Proc Natl Acad Sci U S A*. 101:9722–9727.

Brinkmann H, Philippe H. 2007. The diversity of eukaryotes and the root of the eukaryotic tree. In: Back N, Cohen IR, Kritchevsky D, Lajtha A, Paoletti R, editors. Eukaryotic membranes and cytoskeleton. New York: Springer. p 20–37.

Bui ET, Bradley PJ, Johnson PJ. 1996. A common evolutionary origin for mitochondria and hydrogenosomes. *Proc Natl Acad Sci U S A*. 93:9651–9656.

Buschmann S, Warkentin E, Xie H, Langer JD, Ermler U, Michel H. 2010. The structure of $cbb_3$ cytochrome oxidase provides insights into proton pumping. *Science* 329:327–330.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 17:540–552.

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

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WE, Wetter T, Suhai S. 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res*. 14:1147–1159.

Cho NH, Kim HR, Lee JH, et al. (19 co-authors). 2007. The *Orientia tsutsugamushi* genome reveals massive proliferation of conjugative type IV secretion system and host cell interaction genes. *Proc Natl Acad Sci U S A*. 104:7981–7986.

Cox CJ, Foster PG, Hirt RP, Harris SR, Embley TM. 2008. The archaebacterial origin of eukaryotes. *Proc Natl Acad Sci U S A*. 105:20356–20361.

Dacks JB, Doolittle WF. 2001. Reconstructing/deconstructing the earliest eukaryotes: how comparative genomics can help. *Cell* 107:419–425.

Darby AC, Cho NH, Fuxelius HH, Westberg J, Andersson SGE. 2007. Intracellular pathogens go extreme: genome evolution in the Rickettsiales. *Trends Genet*. 23:511–520.

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

de Duve C. 2005. Singularities: landmarks on the pathways of life. Cambridge: Cambridge University Press.

de Duve C. 2007. The origin of eukaryotes: a reappraisal. *Nat Rev Genet*. 8:395–403.

Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res*. 27:4636–4641.

Douglas SE. 1998. Plastid evolution: origins, diversity, trends. *Curr Opin Genet Dev*. 8:655–661.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 7:214.

Ducluzeau AL, Ouchane S, Nitschke W. 2008. The $cbb_3$ oxidases are an ancient innovation of the domain bacteria. *Mol Biol Evol*. 25:1158–1166.

Dunning Hotopp JC, Lin M, Madupu R, Crabtree J, Angiuoli SV, Eisen J, Seshadri R, Ren Q, Wu M, Utterback TR, et al. (40 co-authors). 2006. Comparative genomics of emerging human ehrlichiosis agents. *PLoS Genet*. 2:e21.

Dyall SD, Brown MT, Johnson PJ. 2004. Ancient invasions: from endosymbionts to organelles. *Science* 304:253–257.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 5:113.

Embley TM, Martin W. 2006. Eukaryotic evolution, changes and challenges. *Nature* 440:623–630.

Epis S, Luciano AM, Franciosi F, Bazzocchi C, Crotti E, Pistone D, Bandi C, Sassera D. 2010. A novel method for the isolation of DNA from intracellular bacteria, suitable or genomic studies. *Ann Microbiol*. 60:455–460.

Epis S, Sassera D, Beninati T, Lo N, Beati L, Piesman J, Rinaldi L, McCoy KD, Torina A, Sacchi L, et al. (14 co-authors). 2008. *Midichloria mitochondrii* is widespread in hard ticks (Ixodidae) and resides in the mitochondria of phylogenetically diverse species. *Parasitology* 135:485–494.

Fenchel T. 1987. Ecology of Protozoa. Berlin (Germany): Springer Verlag.

Foster P. 2004. Modeling compositional heterogeneity. *Syst Biol*. 53:485–495.

Fitzpatrick DA, Creevey CJ, McInerney JO. 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.

Fritz-Laylin LK, Prochnik SE, Ginger ML, Dacks JB, Carpenter ML, Field MC, Kuo A, Paredez A, Chapman J, Pham J, et al. (24 co-authors). 2010. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell*. 140:631–642.

Gabaldón T, Huynen MA. 2003. Reconstruction of the proto-mitochondrial metabolism. *Science* 301:609.

Ghai R, Hain T, Chakraborty T. 2004. GenomeViz: visualizing microbial genomes. *BMC Bioinformatics*. 5:198.

Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, Bibbs L, Eads J, Richardson TH, Noordewier M, et al. (14 co-authors). 2005. Genome streamlining in a cosmopolitan oceanic bacterium. *Science*. 309:1242–1245.

Goldman N, Anderson JP, Rodrigo AG. 2000. Likelihood-based tests of topologies in phylogenetics. *Syst Biol*. 49:652–670.

Gribaldo S, Poole AM, Daubin V, Forterre P, Brochier-Armanet CL. 2010. The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? *Nat Rev Microbiol*. 8:743–752.

Guerrero R, Pedros-Alio C, Esteve I, Mas J, Chase D, Margulis L. 1986. Predatory prokaryotes: predation and primary consumption evolved in bacteria. *Proc Natl Acad Sci U S A*. 83:2138–2142.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 52:696–704.

Hartl DL, Ochman H. 1996. Inverse polymerase chain reaction. *Methods Mol Biol*. 58:293–301.

Hartman H, Fedorov A. 2002. The origin of the eukaryotic cell: a genomic investigation. *Proc Natl Acad Sci U S A*. 99:1420–1425.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.

Inglis TJJ, Robertson T, Woods DE, Dutton N, Chang BJ. 2003. Flagellum-mediated adhesion by *Burkholderia pseudomallei* precedes invasion of *Acanthamoeba astronyxis*. *Infect Immun*. 71:2280–2282.

Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 8:27–30.

Kishino H, Hasegawa M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J Mol Evol*. 29:170–179.

Kurland CG, Andersson SGE. 2000. Origin and evolution of the mitochondrial proteome. *Microbiol Mol Biol Rev*. 64:786–820.

Kurland CG, Collins LJ, Penny D. 2006. Genomics and the irreducible nature of eukaryote cells. *Science* 312:1011–1014.

Lane N, Martin W. 2010. The energetics of genome complexity. *Nature* 467:929–934.

Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*. 13:2178–2189.

Liu R, Ochman H. 2007. Stepwise formation of the bacterial flagellar system. *Proc Natl Acad Sci U S A*. 104:7116–7121.

Lo N, Beninati T, Sacchi L, Bandi C. 2006. An α-proteobacterium invades the mitochondria of the tick *Ixodes ricinus*. In: Bourtzis K, Miller TA, editors. Insect symbiosis Vol. 2. Boca Raton (FL): Taylor & Francis. p 25–37.

Mahenthiralingam E, Speert DP. 1995. Nonopsonic phagocytosis of *Pseudomonas aeruginosa* by macrophages and polymorphonuclear leukocytes requires the presence of the bacterial flagellum. *Infect Immun*. 63:4519–4523.

Margulis L. 1970. Origin of eukaryotic cells. New Haven (CT): Yale University Press.

Martin W, Muller M. 1998. The hydrogen hypothesis for the first eukaryote. *Nature* 392:37–41.

Matz C, Kjelleberg S. 2005. Off the hook—how bacteria survive protozoan grazing. *Trends Microbiol*. 13:302–307.

Maynard Smith J, Szathmary E. 1997. The major transitions in evolution. Oxford: Oxford University Press.

Ottemann KM, Miller JF. 1997. Roles for motility in bacterial–host interactions. *Mol Microbiol*. 24:1109–1117.

Pitcher RS, Watmough NJ. 2004. The bacterial cytochrome $cbb_3$ oxidases. *Biochim Biophys Acta (BBA)—Bioenergetics* 1655:388–399.

Poole AM, Penny D. 2006. Evaluating hypotheses for the origin of eukaryotes. *BioEssays* 29:74–84.

Poole AM, Penny D. 2007. Engulfed by speculation. *Nature*. 447:913.

Pop M, Kosack DS, Salzberg SL. 2004. Hierarchical scaffolding with Bambus. *Genome Res*. 14:149–159.

Puigbo P, Bravo IG, Garcia-Vallve S. 2008. CAIcal: a combined set of tools to assess codon usage adaptation. *Biol Direct*. 3:38.

Ramos HC, Rumbo M, Sirard JC. 2004. Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. *Trends Microbiol*. 12:509–517.

Rivera MC, Lake JA. 1992. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* 257:74–76.

Sacchi L, Bigliardi E, Corona S, Beninati T, Lo N, Franceschi A. 2004. A symbiont of the tick *Ixodes ricinus* invades and consumes mitochondria in a mode similar to that of the parasitic bacterium *Bdellovibrio bacteriovorus*. *Tissue Cell*. 36:43–53.

Sassera D, Beninati T, Bandi C, Bouman EA, Sacchi L, Fabbi M, Lo N. 2006. '*Candidatus* Midichloria mitochondrii', an endosymbiont of the tick *Ixodes ricinus* with a unique intramitochondrial lifestyle. *Int J Syst Evol Microbiol*. 56:2535–2540.

Sassera D, Lo N, Bouman E, Epis S, Mortarino M, Bandi C. 2008. *Midichloria* endosymbionts bloom after blood meal of the host, the hard tick *Ixodes ricinus*. *Appl Environ Microbiol*. 74:6138–6140.

Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res*. 33:686–689.

Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*. 18:502–504.

Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol*. 16:1114–1116.

Slamovits C, Keeling P. 2006. A high density of ancient spliceosomal introns in oxymonad excavates. *BMC Evol Biol*. 6:34.

Staden R, Beal KF, Bonfield JK. 2000. The Staden package, 1998. *Methods Mol Biol*. 132:115–130.

Strimmer KR. 2002. Inferring confidence sets of possibly misspecified gene trees. *Proc R Soc Lond B Biol Sci*. 269:137–142.

Toft C, Fares MA. 2008. The evolution of the flagellar assembly pathway in endosymbiotic bacterial genomes. *Mol Biol Evol*. 25:2069–2076.

Tovar J, Leon-Avila G, Sanchez LB, Sutak R, Tachezy J, van der Giezen M, Hernandez M, Muller M, Lucocq JM. 2003. Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. *Nature* 426:172–176.

Vannini C, Ferrantini F, Schleifer KH, Ludwig W, Verni F, Petroni G. 2010. "*Candidatus* anadelfobacter veles" and "*Candidatus* cyrtobacter comes," two new rickettsiales species hosted by the protist ciliate *Euplotes harpa* (Ciliophora, Spirotrichea). *Appl Environ Microbiol*. 76:4047–4054.

Vellai T, Takács K, Vida G. 1998. A new aspect to the origin and evolution of eukaryotes. *J Mol Evol*. 46:499–507.

von Dohlen CD, Kohler S, Alsop ST, McManus WR. 2001. Mealybug beta-proteobacterial endosymbionts contain gamma-proteobacterial symbionts. *Nature*. 412:433–436.

Williams KP, Sobral BW, Dickerman AW. 2007. A robust species tree for the α-proteobacteria. *J Bacteriol*. 189:4578–4586.

Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C, Ahmadinejad N, et al. (30 co-authors). 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol*. 2:E69.

Zhu Z, Aeschlimann A, Gern L. 1992. *Rickettsia*-like microorganisms in the ovarian primordia of molting *Ixodes ricinus* (Acari: Ixodidae) larvae and nymphs. *Ann Parasitol Hum Comp*. 67:99–110.