



BIOCHEMISTRY

Multiple approaches of cellular metabolism define the bacterial ancestry of mitochondria

Otto Geiger¹, Alejandro Sanchez-Flores², Jonathan Padilla-Gomez¹, Mauro Degli Esposti^{1*}

We breathe at the molecular level when mitochondria in our cells consume oxygen to extract energy from nutrients. Mitochondria are characteristic cellular organelles that derive from aerobic bacteria and carry out oxidative phosphorylation and other key metabolic pathways in eukaryotic cells. The precise bacterial origin of mitochondria and, consequently, the ancestry of the aerobic metabolism of our cells remain controversial despite the vast genomic information that is now available. Here, we use multiple approaches to define the most likely living relatives of the ancestral bacteria from which mitochondria originated. These bacteria live in marine environments and exhibit the highest frequency of aerobic traits and genes for the metabolism of fundamental lipids that are present in the membranes of eukaryotes, sphingolipids, and cardiolipin.

Copyright © 2023 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

INTRODUCTION

Unveiling the origins of mitochondria continues to challenge science. While there is broad consensus that mitochondria first evolved 1600 million to 1800 million years ago, the critical question of from which bacteria they originated remains unanswered (1–13). Previous research has primarily relied on phylogenetic inference to identify the possible bacterial ancestors of mitochondria, hereafter called protomitochondria. However, this approach has produced inconsistent and varying results depending on the phylogenetic approach, taxonomic sampling, and corrections used to reduce artifacts (1–8, 12, 13). Many different alphaproteobacteria have been proposed to be close to protomitochondria [see (3, 5, 7, 8, 11, 13) and references therein]. The inconclusiveness of available evidence suggests that phylogenetic trees may not be adequate for identifying the extant bacteria that are closest to protomitochondria. This is likely due to the vast amount of time passed since the original symbiotic event, which has diluted and dispersed the phylogenetic signal of contemporary bacterial proteins with respect to their mitochondrial homologs (2, 6). Differential loss of ancestral genes may additionally contribute to the complexity of defining the bacterial ancestors of mitochondria (6, 13). Moreover, the debate about whether the bacterial ancestor of mitochondria was an obligate or facultative aerobe [see (10) for a recent review] further complicates the evaluation of the metabolic ancestry of protomitochondria. Thus, new and robust evidence is needed to unveil the origins of mitochondria (6, 10, 12).

To provide such evidence, we introduce here alternative approaches with different sources of biases (table S1), covering aerobic and anaerobic metabolic traits shared by bacteria and mitochondria (6–11). One of such traits contemplates the enzymes involved in the metabolism of cardiolipin, a typical prokaryote phospholipid that is present only in the mitochondrial membranes of eukaryotic cells (14). The guiding principle of our strategy is that the creation of the first eukaryotic cell involved genomic transmission of metabolic traits from a bacterium that could have surviving descendants today. Although the transmission has been a rare, if not

singular event (1, 3, 6, 10), it might have left vestigial traces in the genome of some of those descendants—similar to “missing link” features found in other major transitions in evolution. An example of evolutionary traces of this kind is the synteny of two genes of cytochrome *c* oxidase (COX or complex IV) (11, 15), the mitochondrial enzyme which ultimately consumes oxygen in our cells (Fig. 1 and fig. S1). Seven genes for COX subunits and accessory proteins form a conserved genomic cluster (operon) that is characteristic of alphaproteobacteria (11). Four of these genes are encoded in mitochondrial DNA (mtDNA) of early-branching unicellular eukaryotes (11, 15–17), while two others are present in mitochondrial complex IV of the protist *Tetrahymena* (Fig. 1) (18). Moreover, the gene for the assembly protein Cox11 (Cox11_CtaG) always precedes the gene for cytochrome oxidase subunit III (COX3; Fig. 1), forming collinearity that is conserved in the mtDNA of some protists (fig. S1). The Cox11-COX3 synteny can thus be considered a genomic relic of the aerobic ancestry of protomitochondria, providing a selection criterion for putative bacterial relatives of mitochondria (11). Its absence in the genome of many bacteria, including the Rickettsiales often considered relatives of mitochondria (4, 7, 15), would exclude such prokaryotes from the ancestry of protomitochondria. Here, we present diverse new approaches that confirm this exclusion and indicate that the ancestor of protomitochondria was likely related to marine alphaproteobacteria never considered before for the evolution of mitochondria.

RESULTS AND DISCUSSION

Unveiling a new synteny in complex III genes

We searched for other genomic traces with equivalent discriminating power as the Cox11-COX3 synteny, focusing on the possible genomic association of two proteins that are structurally and functionally interconnected in complex III: the mitochondrial processing protease (MPP) and the Rieske iron-sulfur protein (ISP; Fig. 1). Mitochondrial complex III (ubiquinol:cytochrome *c* reductase) derives from the bacterial cytochrome *bc*₁ complex, which is encoded by the *petABC* operon now represented by the cytochrome *b* gene in mtDNA [Fig. 1, cf. (15, 19)]. In plants and protists, two MPP proteins form a large domain of complex III structure, not

¹Center for Genomic Sciences, UNAM Campus de Morelos, Cuernavaca, México.

²Unidad Universitaria de Secuenciación Masiva y Bioinformática, Institute of Biotechnology, UNAM, Cuernavaca, México.

*Corresponding author. Email: mdegli@ccg.unam.mx, mauro1italia@gmail.com

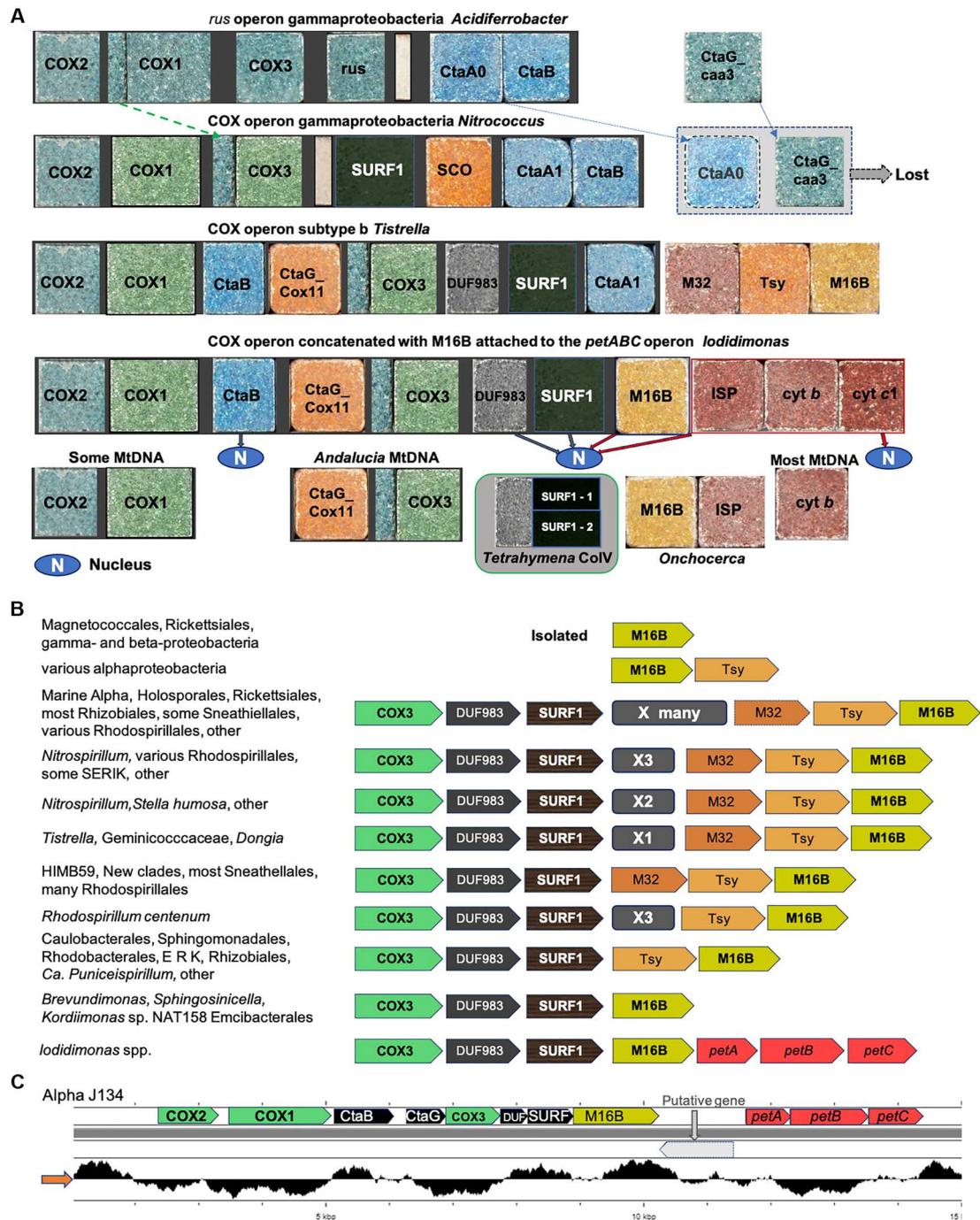


Fig. 1. Evolution of the cytochrome oxidase gene cluster from proteobacteria to mitochondria. (A) The figure illustrates the gene clusters (operons) of cytochrome c oxidase (complex IV) rendered with Roman mosaic tiles to indicate the mosaic nature of the whole mitochondrial proteome (3). The archetype *rus* COX operon (58) includes an ancestral form of both heme A synthase (CtaA0) and transmembrane CtaG (CtaG_caa3). These genes are retained in the genome of nitrifying *Nitrococcus* (second row in the illustration) but have been subsequently lost (58). The transition from the *rus* operon to the COX operon of *Nitrococcus* included the acquisition of two additional transmembrane helices (TM) at its N terminus (18, 20, 22, 58). These extra TM may derive from those present at the N terminus of ancestral COX1 (58), as indicated by the dashed green arrow on the top of the illustration. The gene for the M16.06 group of M16 zinc peptidases (23) is frequently associated with the end of alphaproteobacterial COX operons, often with the gene for threonine synthase (Tsy). Potential homologs of bacterial DUF983 proteins and two SURF1 isoforms are part of *Tetrahymena* Complex IV (18). The last row of the illustration includes also a protein from the filarial *Onchocerca* containing the collinear fusion of M16B with iron-sulfur protein (ISP) (accession OZC11663) encoded in the nuclear DNA of eukaryotes (N symbol). Similar fused proteins have been found in two other nematodes. **(B)** Genomic positioning for the M16B genes in various bacteria. **(C)** Reassembled metagenomic contig829 for alpha J134 (61), a metagenomic-assembled genome (MAG) that clusters with *Iodidimonas* in phylogenetic trees (fig. S6). The red arrow on the left indicates the %GC profile of the reassembled DNAs obtained with the program Proksee (<https://proksee.ca/>). An equivalent gene cluster is present in the MAG, Rhodothalassiacae bacterium KatS3mg119, which has been sequenced with long reads (75) and clusters with Iodidimonadales (see Materials and Methods).

Table 1. List of the aerobic traits considered for the analysis in Fig. 2. AOX, alternative oxidase.

Protein and defining trait	Premium for synteny*	Penalty for absence	Considered earlier
Cyt C cytochrome c			No, this work
COX1 COX – complex IV	+2 with COX2	–1	Yes
COX2 COX – complex IV		–1	Yes
COX3 COX – complex IV	+2 with COX11		Yes
Cox11 COX assembly			Yes
Cox15 type2, heme A synthesis			Yes
SURF1 COX assembly			Yes
SCO COX assembly	Numeral as its multiple genes†		Yes
ZF-CHCC precursor subunit Vb			No, this work
M16B precursor of MPP	+2 with ISP	0 if not close to COX	No, this work
Cbp3 bc1 assembly			No, this work
ISP bc1 – complex III			Yes
CytB bc1 – complex III			Yes
CytC1 bc1 – complex III			Yes
CcmF cyt C biosynthesis	+1 with two other Ccm genes		Yes
CcmE cyt C biosynthesis			No, this work
CcmA cyt C biosynthesis			No, this work
AOX alternative oxidase	Numeral as its multiple genes†		Yes
M16A fused zinc peptidases	Capped numeral for multiple genes‡		No, this work
Cys near C terminus COX1			No, this work

*Numerical premium given to collinear synteny, as described in the Materials and Methods. †A few alphaproteobacterial genomes have multiple genes for SCO proteins, while some eukaryotes have multiple AOX genes. ‡Several eukaryotes have up to eight M16 (A, B, and C subfamily) peptidases, but their maximal numeral was capped at five as described in Materials and Methods.

present in bacterial bc_1 complex (18, 20, 21). In animal mitochondria, MPP derivatives called core proteins (CPs) have the same structural organization (19, 22), while the MPP heterodimer is a separate soluble enzyme (20). We have confirmed two genes for MPP proteins but hardly any for CP proteins in the genomes of Rhodophyta, Discoba, and other single-cell eukaryotes (table S2). Hence, it is highly likely that MPP proteins were constitutive components of the first mitochondrial complex III, as seen in *Tetrahymena* (18). Notably, CPs retain the function of processing the presequence of ISP (19, 22), thereby underlying the intimate connection between MPP and ISP. Following the finding of a filarial protein corresponding to the fusion of MPP with ISP (Fig. 1A), we systematically searched the genomes of currently available bacteria for the contiguity of genes encoding the bacterial homologs of MPP and ISP.

The bacterial homolog and likely precursor of MPP are a zinc peptidase belonging to a specific group of the M16B subfamily (23). While the gene for M16B is isolated in gammaproteobacteria and early-branching alphaproteobacteria such as Rickettsiales, it is often associated with that of threonine synthase (Tsy) in other alphaproteobacteria. These two genes appear to progressively migrate close to the COX operon in the genome of various alphaproteobacteria, becoming attached to the gene of Surfeit locus protein 1 (SURF1) ending the same operon in a number of taxa (Fig. 1). This genomic contiguity is intermixed with the gene of carboxypeptidase M32 (abbreviated as M32 in Fig. 1) in Rhodospirillales. Only in Iodidimonadales, however, the M16B genes are found close to, or directly associated with, the *petABC* operon of the bc_1 complex (Fig. 1, A and C). This direct association is present in the genomes of *Iodidimonas* spp. (Fig. 1), as well as in related alphaproteobacteria Q-1 (table S3). Such taxa belong to the order Iodidimonadales that is part of the group of Sneathiellales, Emericellales, Rhodothalassiales, Iodidimonadales, and Kordiimonadales (SERIK) (2). Detailed searches of currently available genomes failed to retrieve genomic associations equivalent to those found in cultivated taxa of Iodidimonadales (Fig. 1, A and B), except for metagenomic-assembled genomes (MAGs) that cluster with Iodidimonadales as alphaproteobacteria bacterium J134 (alpha J134; Fig. 1C).

Distribution of aerobic traits in mitochondria and alphaproteobacterial lineages

The rare M16B-ISP synteny (Fig. 1) would represent a novel trace of the metabolic ancestry of protomitochondria or may derive from some unusual genomic streamlining. To discriminate between these possibilities, we followed different approaches, focusing first on the central part of the respiratory chain pivoting on cytochrome *c* (24). This part of the respiratory chain defines the aerobic metabolism of mitochondria, which must have been crucial in the environmental adaptation of the first eukaryotes (3, 10); it is switched off in eukaryotes adapted to anoxia [see (10, 25) and references therein]. Three operons contribute to cytochrome *c* biogenesis and function in bacteria, and the majority of their genes, together with a few isolated genes for the assembly of the respiratory complexes such as Cox15 (heme A synthase), are present in eukaryotes (24). Overall, the number of the shared genes for the central part of the respiratory chain is 25, including soluble cytochrome *c* (see Material and Methods for details). We considered each of these genes as an individual trait contributing to the aerobic metabolism of

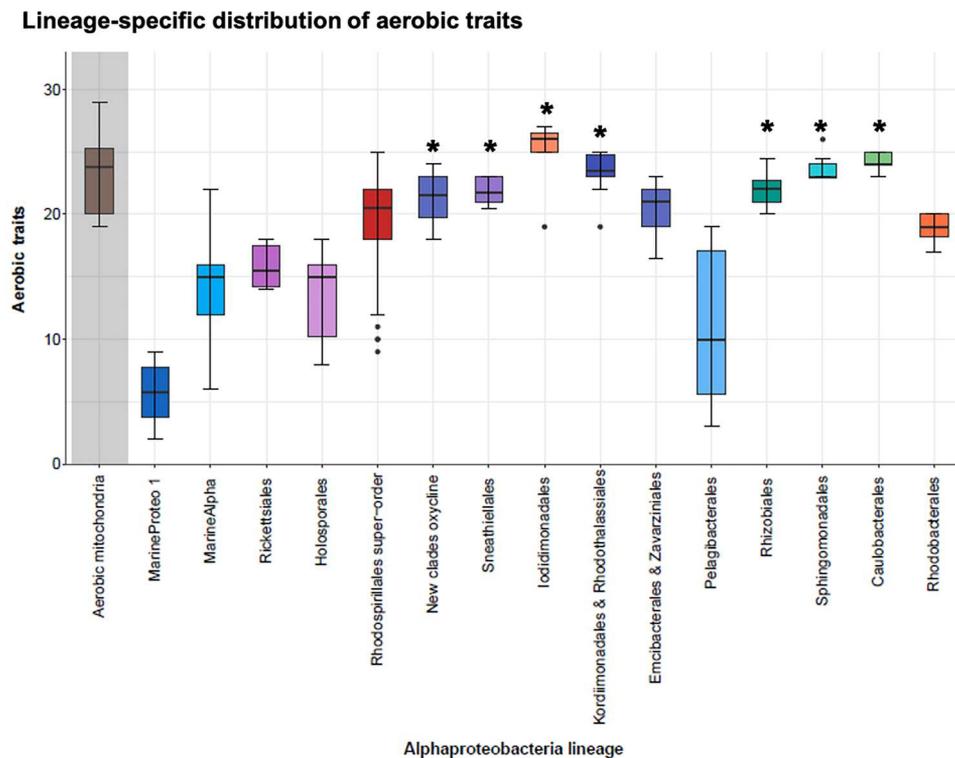


Fig. 2. Distribution of aerobic traits scores along alphaproteobacterial lineages and mitochondria. The figure presents the lineage-dependent plot of the cumulative scores for the set of 20 aerobic traits listed in Table 1. The mean values for each lineage follow the branching order of such alphaproteobacteria lineages along the x axis (except for the mitochondria on the left). Asterisks indicate the distribution values that are not significantly different from that of aerobic mitochondria ($P > 0.1$ with the 99% confidence t test; table S4). Note that the lineage of MarineProteo1 lacks the A1 type COX1–3 proteins that are shared by alphaproteobacteria and mitochondria. The figure presents the lineage-dependent plot of the cumulative scores for the set of 20 aerobic traits listed in Table 1. The mean values for each lineage follow the branching order of such alphaproteobacteria lineages along the x axis (except for the mitochondria on the left).

bacteria and mitochondria and computed their cumulative distribution in different combinations (Table 1 and fig. S2A). We then settled on a representative set of 20 traits including the bacterial Zn-finger precursor of Cox5B (26) and a cysteine signature in a conserved C-terminal region of catalytic COX subunit 1 (COX1) (27). The latter trait was chosen for its common presence in alphaproteobacteria with reduced genomes including Rickettsiales (table S3). We also considered the genes for M16A peptidases, which have multiple homologs in eukaryotes (Table 1) (28). These and other traits have not been considered before in relation to eukaryogenesis (1–8, 15), as indicated in Table 1.

Figure 2 presents the lineage-specific distribution (table S3) of the chosen set of 20 aerobic traits in quantitative terms, showing an apparent peak around Iodidimonadales. This peak depends only partially on the premium given to the presence of the M16B-ISP synteny and is present in all other combinations of aerobic traits (fig. S2). The cumulative aerobic traits score for Iodidimonadales does not substantially differ from aerobic mitochondria (Fig. 2 and table S4). Sneathiellales and related new clades of marine taxa (2), Kordiimonadales, Rhizobiales, Sphingomonadales, and Caulobacterales also showed a distribution of aerobic traits comparable with that of mitochondria (Fig. 2 and table S4). In contrast, the lineages of MarineProteo1, MarineAlpha, Rickettsiales, Holosporales, Pelagibacterales, and Rhodobacterales have significantly lower

cumulative scores than aerobic mitochondria (Fig. 2), suggesting their exclusion from the aerobic ancestry of protomitochondria. However, it is difficult to discriminate which alphaproteobacterial lineage may display the best match of the aerobic metabolism of mitochondria. Hence, different approaches are needed to further filter alphaproteobacteria lineages for identifying the most likely bacterial ancestor of protomitochondria (Tables 1 and 2, see also table S1).

Distribution of genes for ceramide and kynurenine biosynthesis in alphaproteobacteria

Our next approach was based on a completely different metabolic pathway that has hardly been considered before in regard to eukaryogenesis: the biosynthesis of ceramide-based lipids, sphingolipids. Sphingolipids constitute a vast class of membrane lipids that are ubiquitous in eukaryotes but scarcely present in bacteria (29, 30). To produce the precursor of ceramide, bacteria often require a four-gene operon ending with the gene for an α -oxoamine synthase catalyzing the key step of ceramide biosynthesis: serine palmitoyl-transferase (SPT) [Fig. 3A, cf. (29–31)]. We found the *spt* gene encoding SPT in Iodidimonadales and other members of the SERIK group, as well as in some Rhodospirillales and a few Rhizobiales (Fig. 3 and table S5). Our phylogenetic analysis indicates that *Odyssella* sp. NEW MAG-112 may have the earliest enzyme for ceramide biosynthesis of all alphaproteobacteria (Fig. 3B).

Table 2. Discriminatory criteria to evaluate which alphaproteobacteria may be close to protomitochondria. Taxa were selected from those that presented with cumulative scores of aerobic traits above 20, thus overlapping the distribution values of aerobic mitochondria (Fig. 2), both types of cardiolipin synthase (Fig. 3E) and a collinearity bloc of ribosomal proteins (RP) encoded in the mtDNA of Jakobida (15). “Yes” indicates that a taxon has passed the discriminatory criterion indicated in a given column. Taxa are listed at the genus level when other species of the same genus passed equivalent criteria. Rhodothal. MAG indicates Rhodothalassiaceae bacterium KatS3mg119 (75), an early-branching Iodidimonadales.

Alphaproteobacteria taxon	Both Cls types	RP synteny*	NO INDELS [†]	SPT & KynU	≥3 anaerobic traits	Top hits MPPbeta	M16B-ISP synteny
<i>Tistrella</i>		Yes	Yes				
<i>Elioraea rosea</i>	Yes	Yes					
<i>Nitrospirillum amazon.</i>				Yes	Yes		
<i>Caenispirillum</i>	Yes	Yes			Yes		
Rhodospirillaceae SP28	Yes		Yes			Yes	
80m_m2_115 new clade		Yes			Yes		
<i>Oceanibacterium</i>			Yes	Yes		Yes	
oxycline co234_bin8		Yes	Yes		Yes	Yes	
Alpha J067 [‡]	Yes				Yes		
Rhodothal. MAG§	Yes	Yes	Yes	Yes			Yes
Alpha J134 [‡]	Yes		Yes			Yes	Yes
<i>Iodidimonas muriae</i>	Yes	Yes	Yes	Yes	Yes		Yes
<i>Kordiimonas pumila</i>		Yes		Yes	Yes		
<i>Odyssella</i> sp. [§]	Yes			Yes	Yes		
<i>Camelimonas</i>	Yes	Yes	Yes				
<i>Tepidicaulis marinus</i>	Yes	Yes					
<i>Hansschlegelia quercus</i>	Yes	Yes					
<i>Microvirga arabica</i>	Yes		Yes				
<i>Methylotella</i>	Yes		Yes				
<i>Aestuariivirga litoralis</i>	Yes			Yes			

*Major collinearity bloc of RP proteins and RNA polymerase shared with Jakobida mtDNA (15). †No conserved INDELS in COX3 and ISP as in mitochondrial homologs (11). ‡MAG largely incomplete, even after reassembling (see Materials and Methods). §Taxa with cumulative aerobic traits below the second quartile of mitochondria.

Alphaproteobacterial SPT appears to be the ancestor of both isoforms of eukaryotic SPT (Fig. 3B), as well as the SPT of nitrifying taxa such as *Nitrococcus* (fig. S3A). Such nitrifying bacteria have intracytoplasmic membranes resembling mitochondrial cristae, as in alphaproteobacterial methanotrophs [Methylocystaceae (32, 33)] that also have SPT (fig. S3A). Ceramide is well known to modulate the curvature and shape of lipid bilayers (34); therefore, it may be crucial for the formation of bacterial intracytoplasmic membranes. The genomic distribution of *spt* and its partner genes shows a maximum within the SERIK group, besides the expected high frequency in Sphingomonadales (Fig. 3C). The limited presence of the same genes in different lineages probably derives from events of lateral gene transfer (LGT) since they are present only in a few taxa that have SPT proteins clustering with those of other alphaproteobacteria, as in the case of *Caulobacter* (fig. S3A). We also found that SPT distribution often matches that of the kynureninase gene *kynU* (table S5), another pyridoxal 5'-phosphate-dependent enzyme defining the kynurenine pathway for NAD(P)⁺ [nicotinamide adenine dinucleotide (phosphate)] biosynthesis in bacteria (35). An equivalent pathway is required for de novo synthesis of rhodoquinone (RQ) in nematodes and other eukaryotes, which

use this quinone in their adaptation to low levels of oxygen (36) (see below). The *kynU* gene is generally associated with *kynA* encoding tryptophan dioxygenase, the upstream enzyme of the kynurenine pathway (35, 36). This genomic association is present in several taxa of the SERIK group but absent in Caulobacterales (table S5). Hence, the SPT-kynureninase combination produces a stringent criterion for discriminating alphaproteobacterial lineages from the ancestry of protomitochondria, which must have had both traits now present in a variety of eukaryotes.

Distribution of the enzymes for the metabolism of cardiolipin

Cardiolipin or diphosphatidylglycerol, as it is traditionally called in microbiology literature, is a dimeric phospholipid that provided one of the first hallmarks for the bacterial origin of mitochondria (14, 37). In eukaryotic cells, it is synthesized in mitochondria and remodeled with the participation of extramitochondrial enzymes but normally resides in the inner mitochondrial membrane (38, 39). In prokaryotes, cardiolipin and its various derivatives are constituents of the cytoplasmic membrane, often fulfilling essential roles in viability (40). Despite the well-known bacterial origin of

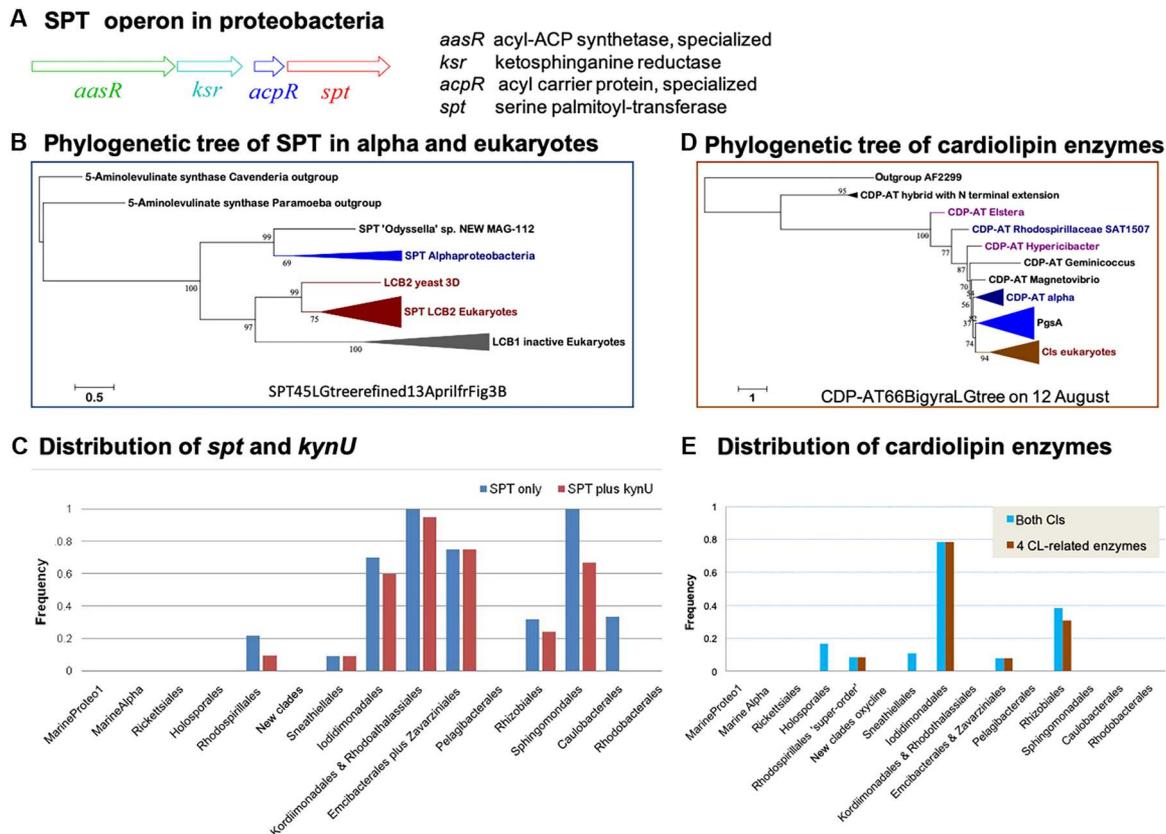


Fig. 3. Genomics and phylogenetic trees of bacterial and eukaryotic SPT, KynU, and cardiolipin enzymes. (A) Representation of the four-gene operon comprising SPT found in proteobacteria, for example, *Caulobacter* (29, 31). (B) Phylogenetic maximum likelihood (ML) tree of SPT from alphaproteobacteria and various eukaryotes, which have two isoforms (30): the catalytic LCB2 and the inactive LCB1. Paralog proteins of 5-aminolevulinatase synthase are used as outgroups providing the root of the tree. The alpha MAG originally named *Odysseella* sp. NEW MAG-112 (GCA_016792765.1) is not a member of the Holosporales but of the order α _Bin65 according to GTDB taxonomy (65). It is included here in the lineage of Epsilonbacterales and Zavarziniales for similar INDELS profiles (tables S3 and S4) and therefore labeled as “Odysseella.” Numbers indicate the strength of the nodes in percentage values of ultrafast bootstraps. (C) Frequency distribution of the combination of the *spt* and *kynU* genes, irrespective of their surrounding genes, along the alphaproteobacterial lineages. Note that the majority of the eukaryotic taxa used for the analysis in Fig. 2 do have SPT proteins, while the distribution of the *kynU* gene is scattered (36). (D) Phylogenetic ML tree of various CDP-alcohol Transferases (CDP-AT) proteins. (E) Frequency distribution of the enzyme for cardiolipin metabolism, cumulatively indicated as CL-related enzymes. Blue histograms represent the distribution data of the presence of both types of cardiolipin synthase (ClS) in the same genome, while the brownish histograms represent the distribution data of the same ClS proteins plus that of PgsA and relatives of Cls1, cardiolipin-specific lipase of yeast (table S5) (39).

cardiolipin, the evolution of its metabolism is quite complex and thus challenging because of the presence of multiple pathways for cardiolipin biosynthesis in different organisms (14, 37, 40, 41). The so-called “bacterial” pathway using an enzyme related to phospholipase D (40), ClsA or Cls_pld, is also present in Kinetoplastida (42) and other eukaryotes (14, 37). Conversely, the typical “mitochondrial” cardiolipin synthase belongs to the superfamily of cytidine 5'-diphosphate–alcohol transferases [CDP-AT; cf. (43)]. This enzyme is widespread among bacteria too (14) and is structurally very similar to another member of the same superfamily, PgsA, catalyzing the biosynthesis of phosphatidylglycerol-phosphate (PGP; Fig. 3D and fig. S4A) (43, 44).

The taxonomic distribution of the CDP-AT type of cardiolipin synthase, Cls-AT, is not homogenous among Opisthokonts, the eukaryotic group spanning Amoebozoa and metazoa (45). While metazoans and fungi have it, other members of the super-group such as Amoebozoa have the Cls_pld type instead (14, 37, 41). Conversely, we found that the genome of *Techamonas*, a unicellular

biflagellate basal to the super-group of Opisthokonts (45), contains both types of cardiolipin synthase—KNC46788 for Cls-AT and KNC55721 for Cls_pld (fig. S4, B and D). A similar dual presence of the different cardiolipin synthases has been previously reported for two species of Stramenopiles belonging to the *phylum* Bigyra of the Stramenopiles, Alveolates and Rhizaria (SAR) super-group (41). We found two other species of Bigyra that have both types of cardiolipin synthase, i.e., *Hordea fermentalgiana* and *Bicosoecida* sp. CB-2014. The sequence of the Cls_pld type of the latter organism does not have a conserved Lys in the first HisLysAsp (HKD) motif of catalytic residues in phospholipase D enzymes (fig. S4). Mutagenesis of this invariant Lys produces a total loss of activity (46); consequently, the Cls_pld enzyme of *Bicosoecida* sp. CB-2014 is likely inactive. The genome of *Bicosoecida* sp. CB-2014 may thus represent an evolutionary transition from the ancestral presence of both different types of cardiolipin synthases and the current dominance of the Cls-AT type among Stramenopiles and many other eukaryotes (41).

The simplest explanation for the dual presence of different enzymes for cardiolipin biosynthesis in some eukaryotes, and the scattered distribution of either type in different eukaryotic groups (14, 41), would be that the bacterial ancestor of protomitochondria had both types of cardiolipin synthase too. Differential loss from the ancestral state having both types of enzymes would then rationalize the complex distribution among eukaryotes. In support of this possibility, we found the dual presence of the different cardiolipin synthases in the genome of some alphaproteobacteria, with maximal frequency in the lineages of Iodidimonadales and Rhizobiales (Fig. 3E and table S6). While MAGs clustering with Iodidimonadales such as alpha J067 have a complete Cls-AT and a standard Cls_pld (table S6), *Iodidimonas* spp. has a different subtype of CDP-AT proteins with long N-terminal extensions that may function as the Cls-AT of other taxa, despite its overall similarity with

the distant CDP-AT AF2299 (43). The Iodidimonadales proteins with N-terminal extension may represent an ancestral relative of the Cls-AT type of cardiolipin synthase since they form an early-branching clade in the phylogeny of both the bacterial and mitochondrial enzymes, as well as PgsA (Fig. 3D). They display local sequence similarities with conserved signatures of Cls-AT that are much higher than with other members of the CDP-AT superfamily (see Materials and Methods for details). Proteins closely related to those of Iodidimonadales are present in a few Rhodospirillales, for example, *Niveispirillum*, and various species of *Caulobacter*, though cardiolipin lipids seem to be absent in these taxa, including the most studied *Caulobacter vibrioides* (formerly *Caulobacter crescentus*) (47). Consistent with biochemical data, our genomic analysis indicates that enzymes of cardiolipin metabolism besides PgsA are essentially absent in the Caulobacterales lineage (Fig. 3E and table S6).

Distribution of anaerobic traits in a selection of alphaproteobacteria

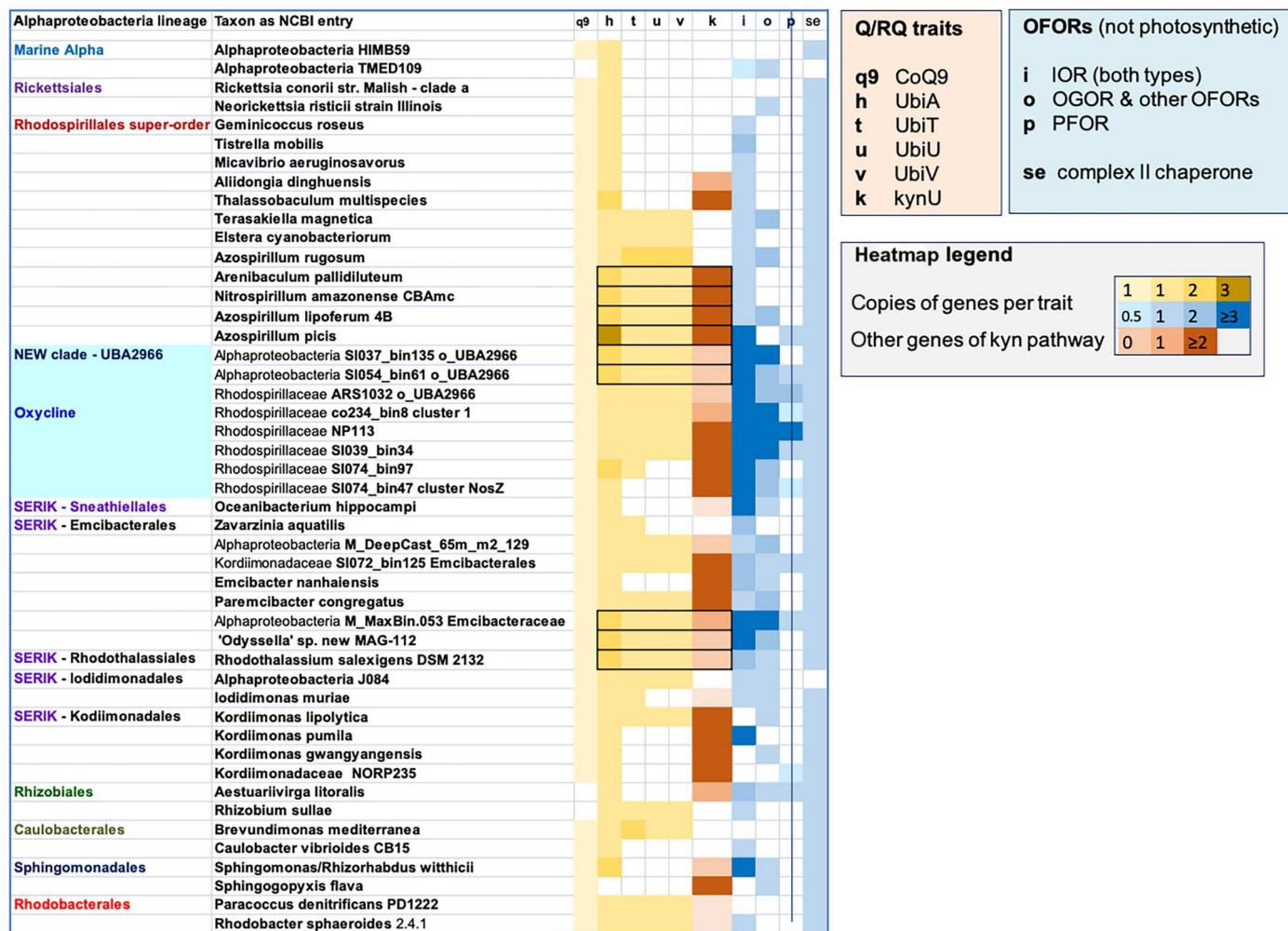


Fig. 4. Distribution of proteins for anaerobic traits in a selection of alphaproteobacteria. At least two representatives for each of the lineages of alphaproteobacteria considered here (Fig. 2) are presented in a compacted phylogenetic sequence from top to bottom. Several taxa are considered for Rhodospirillales and the SERIK group, as well as the lineage of new clades (2), because they present the highest concentration of OFORs traits (fig. S5), as well as of those for the anaerobic production of Q (2, 51, 52). "0.5" indicates partial proteins. Traits boxed in black squares indicate the potential for de novo biosynthesis of RQ (see text). The abbreviations for the various traits are described in the legend boxes on the right. The traits of CoQ9 (q9) on the far left and of the chaperone of complex II (se) on the far right are reference common genes shared with mitochondria.

Considering the above evidence, it is likely that the *Cls_pld* type of cardiolipin synthase in *Iodidimonas* (table S6) predominantly produces the small levels of cardiolipin reported in this taxon (48). Note that the content in the inner mitochondrial membrane of several eukaryotes can amount to up to 20% of total lipids (49), a value much higher than that usually encountered in alphaproteobacteria. Energetically, the reactions catalyzed by the *Cls_pld* type or the *Cls_AT* are quite different. Whereas, in the *Cls_pld*-catalyzed reaction, no standard free energy is liberated, the standard free energy for the *Cls_AT*-catalyzed reaction is quite negative due to the hydrolysis of CDP—diacylglycerol to cytidine monophosphate (14, 37). The liberation of such free energy shifts the equilibrium of the *Cls-AT* reaction toward the cardiolipin product, thereby achieving high concentrations of cardiolipin in the near absence of the phosphatidylglycerol precursor, which is much more concentrated in bacterial than in mitochondrial membranes (37, 40, 44). For this reason, *Cls-AT* type might have been selected to guarantee cardiolipin levels under conditions of low availability of phosphatidylglycerol as in the marine environments where *Iodidimonas* spp. live (48). Notably, *Bigyra* taxa that also have both types of cardiolipin synthase live in similar marine environments (41), as the earliest eukaryotes presumably did (10). These considerations would rationalize why the presence of both types of cardiolipin synthase appears to be the ancestral state for cardiolipin metabolism in eukaryotes (41).

Analysis of anaerobic traits along alphaproteobacteria lineages

The previous analysis of *kynU* distribution has suggested that de novo synthesis of RQ may occur also in alphaproteobacteria. This would be an important novelty relevant to the bacterial ancestry of protomitochondria because RQ is one rare trait of anaerobic metabolism that is shared by mitochondria and alphaproteobacteria (25, 36, 50). Members of the Azospirillaceae family that have *kynU* and related genes of the kynurenine pathway (table S5) also have two or more genes encoding different forms of the prenyl-transferase (UbiA) protein catalyzing a critical step in ubiquinone (Q) biosynthesis (Figure 4). This is an infrequent occurrence that echoes a distinctive feature of de novo RQ biosynthesis in *Caenorhabditis elegans*: the presence of a second UbiA protein that specifically transfers the isoprenoid tail to the ring precursor derived from the kynurenine pathway (36). Therefore, the presence of multiple genes for UbiA proteins in alphaproteobacteria that likely have the kynurenine pathway (boxed in Fig. 4) suggests that such bacteria may synthesize RQ by a de novo system equivalent to that of *C. elegans*. We additionally found that the genomes of several alphaproteobacteria also have the *ubiTUV* genes for the anaerobic biosynthesis of Q (Fig. 4) (51, 52). These are *Nitrospirillum amazonense*, *Arenibaculum*, most members of the *Azospirillum lipoferum* clade, *Rhodothalassium* spp., and four MAGs including *Odyssella* sp. NEW MAG-112 encountered before (Fig. 3B). Notably, the proteins encoded by *ubiU* and *ubiV* bind a 4Fe4S cluster promoting the oxygen-independent hydroxylation of the Q ring under anaerobic conditions (52). Under normal oxygen conditions, this hydroxylation is catalyzed by flavin hydroxylases such as UbiH (51). We have found paralogs of UbiU in Chlorophyta and other protists, often in fused proteins containing a similar, UbiV-like domain (see Materials and Methods for further details). It is thus possible that the *ubiU-ubiV* synteny has been transmitted by an alphaproteobacterial progenitor to primordial eukaryotes,

thereby rendering these genes additional markers for the anaerobic metabolism shared by facultatively aerobic bacteria and eukaryotes (2, 25, 51, 52).

Next, we considered the distribution of 2-oxoacid:ferredoxin oxidoreductases (OFORs) as established traits for anaerobic metabolism (7, 25). Five different types of OFORs (53) are found in the genomes of alphaproteobacteria, the long indolepyruvate:ferredoxin oxidoreductase being the most common (Fig. 4 and fig. S5A). The cumulative distribution of OFORs genes is uneven along the various lineages, with maximal concentration in the new clades (Fig. 4 and fig. S5). Eukaryotes adapted to anoxic conditions predominantly have the pyruvate:ferredoxin oxidoreductase and the oxoglutarate:ferredoxin oxidoreductase in hydrogenosomes or other mitochondria-related organelles (MRO) (25). The genes for these proteins are concentrated in the phylogenetic space spanning the new clades and the SERIK group (2), which also shows high scores for aerobic traits (Fig. 4 and fig. S5). Hence, the distribution of the metabolic traits considered so far is not random among extant alphaproteobacteria, converging on a central phylogenetic region that may have the highest probability to be close to the ancestors of protomitochondria. To investigate this possibility, we used various discriminatory criteria derived from additional approaches (Table 2), which were selected from a large set of traits (Table 3) on the basis of discriminatory power (mainly derived from a relatively narrow distribution among alphaproteobacteria compared with a common presence in eukaryotes), monophyletic clustering of eukaryotic proteins and independence from other approaches.

Selection of candidate bacteria and relationships with phylogenetic evidence

Table 2 lists alphaproteobacteria selected after an initial screening based on the presence of both types of cardiolipin synthase (Fig. 3E and table S6), which were further evaluated with discriminatory criteria derived from different approaches, for example, the collinearity in a bloc of ribosomal genes as in the mtDNA of *Jakobida* (15). Among the 20 bacteria that passed at least two of such criteria, *Iodidimonas* and related taxa passed the highest number, suggesting that these taxa may have a higher probability to match the metabolic profile of protomitochondria than other alphaproteobacteria (Table 2). Detailed genomic analysis of isolated genes and various operons indicated that this finding was not correlated to peculiar features of gene insertion, transfer, or duplication in Iodidimonadales versus other alphaproteobacteria. Therefore, the results in Table 2 sustain the novel possibility that the origin of protomitochondria was close to the phylogenetic space encompassing extant Iodidimonadales. This contrasts with the recent proposal that protomitochondria might have originated from a lineage outside core alphaproteobacteria (1, 8, 54, 55). We believe that this discrepancy fundamentally derives from the complexities of sophisticated phylogenetic analysis (12).

The proposal that protomitochondria may be a sister group of core alphaproteobacteria is based on the strict interpretation of maximum likelihood (ML) phylogenies (1, 2, 8, 54), with the assumption that Magnetococci used for rooting the phylogenetic trees are not part of the alphaproteobacteria class (1, 8, 54). This phylogenetic placement of Magnetococci remains controversial (5, 8, 12, 56) and therefore affects the branching order of ML trees, which are known to produce different basal bifurcations when additional deep-branching sequences are included (57). In

Table 3. Other metabolic systems and traits analyzed in this work and their characteristics. NADH, reduced form of nicotinamide adenine dinucleotide; TCA, Tricarboxylic acid cycle; ATP, adenosine 5'-triphosphate; FOF1, FOF1 ATP synthase; MICOS, Mitochondrial contact sites and cristae organizing system; SQMO, Squalene Monooxygenase; MRO, Mitochondria Related Organelle.

Traits	Metabolic system	Reference	Taxonomic distribution and phylogenetic patterns for alphaproteobacteria (alpha) vs. eukaryotes	<i>Iodidimonas</i>
≥16	NADH-quinone oxidoreductase, complex I; TCA cycle also under anaerobiosis	(2, 8)	Already analyzed in previous works. The largest membrane subunits have strong phylogenetic signal but suffer from compositional artifacts due to their hydrophobicity. Other subunits have variable signals, with some too short for providing valuable phylogenies. Constitutively present in basically all alpha*	Yes
5	Succinate dehydrogenase, complex II; TCA cycle also under anaerobiosis	This work and (8)	Already analyzed in previous works. The two largest catalytic subunits have good phylogenetic signal, while the two membrane subunits are short and poorly conserved. Constitutively present in basically all alpha*	Yes
≥12	FOF1 ATP synthase	(8, 80)	Already analyzed in previous works. The largest catalytic subunits have strong phylogenetic signal, while most membrane subunits are short and poorly conserved. Constitutively present in all alpha*	Yes
1	ETF-Q dehydrogenase	This work	Monophyletic eukaryotic clade. Constitutively present in basically all alpha*.	Yes
1	Sulfite oxidase	This work	Monophyletic eukaryotic clade but localized in peroxisomes in Viridiplanta. Widely distributed among alphaproteobacteria.	Yes
4	[Fe-Fe]hydrogenase and its assembly factors - anaerobic metabolism	(25, 82)	The mature hydrogenase localizes in MRO and also the cytoplasm of various anaerobic eukaryotes, which might have acquired the traits via LGT from different bacteria. Very limited distribution in alpha.	No
2	NirBD for nitrate assimilation	(11)	The pathway is not localized in mitochondria. Predominantly present in Opisthokonts forming polyphyletic eukaryotic clades with intermixed bacteria.	No
1	Mic60	This work and (8, 33)	Already analyzed in previous works. It may have a different function in bacteria (part of the hem operon) and mitochondria (part of the MICOS system regulating cristae). Poor conservation of bacterial proteins.	Yes
2	Ftsy and Fhp for membrane sorting	This work and (83)	Only Fhp produces a monophyletic eukaryotic clade. Distribution is limited to a small set of eukaryotes.	Yes
≥8	Aerobic biosynthesis of ubiquinone	This work and (51)	Proteins are poorly conserved and some have yet unrecognized eukaryotic counterparts. Both polyphyletic and monophyletic eukaryotic clades. Constitutively present in all alpha* without parasitic lifestyle.	Yes
≥4	Biosynthesis of sterols	This work and (10)	The pathway is not localized in mitochondria. SQMO catalyzes the first oxygen-dependent step but produces polyphyletic eukaryotic clades with intermixed bacteria. Moreover, this and other enzymes for sterol biosynthesis are scarcely present in alpha.	Partially
≥4	Biosynthesis of phospholipids: PC PI and PS	This work and (40)	The pathways are not localized in mitochondria and include members of the CDP-AT superfamily related to enzymes evaluated here for cardiolipin metabolism. Poor conservation between alpha and eukaryotes. Scattered distribution among alpha.	Partially
1	CTP synthase	(80)	Nonmonophyletic clade when most eukaryotes are considered. Constitutively present in most alpha*.	Yes
4	Nuclear encoded respiratory proteins, similarity by BLAST	This work and (58)	The eukaryotic proteins generally produce monophyletic clades. In this work, BLAST searches were computed for some proteins as described in Materials and Methods. The results of MPPbeta are presented in fig. S2C.	Yes

*In principle, traits that are constitutively present in most alphaproteobacteria cannot provide discriminatory criteria for selecting relatives of protomitochondria, contrary to the traits in Tables 1 and 2.

our experience, the addition of proteins from members of the MarineProteo1 lineage to taxonomically broad alignments of complex I subunits produced ML trees in which these proteins formed a sister clade to that of core alphaproteobacteria (2). When mitochondrial proteins were also added, they formed a clade branching between the basal clade of MarineProteo1 and that of core proteobacteria (2), thus reproducing previous results (1, 8). Without the proteins from MarineProteo1, the same alignments often generated ML trees in which the mitochondrial clade was sister to that of

Rickettsiales (2, 7), most likely due to compositional bias and other artifacts (8). This situation is common to many proteins defining the traits examined here, while several of such proteins are generally not present in the genomes of MarineProteo1 taxa (Figs. 2 and 3 and tables 3 to 5). Published phylogenies with MarineProteo1 were constructed without proper representation of proteins defining aerobic traits (1, 54), reflecting instead the dominant signal of complex I proteins. This distortion was not adequately balanced by adopting a much larger set of protein markers (8) (see Material and

Methods for details). Hence, the critical issue of an appropriate choice of protein markers and taxonomic sampling has remained unsettled.

We addressed this issue further by using phylogenetic trees reconstructed with the COX3 protein, which has a relatively good phylogenetic signal (2, 11) and does not show mitochondria segregating in a sister clade to core alphaproteobacteria (fig. S6). Moreover, the molecular history of COX3 involved the N terminal fusion of a domain containing two TM that might derive from ancestral COX1 proteins [Fig. 1A cf. (58)]. This fusion left a large insert between the second and third TM of the 7TM COX3 of various bacteria (59), which is common among the proteins of gammaproteobacteria such as *Nitrococcus*. However, this insert was lost in the COX3 proteins of basal alphaproteobacteria such as *Caenispirillum* and other members of the Rhodospirillaceae family (56), which have a simple variant of the COX operon without CtaB and SURF1 [Fig. 5 cf. (11)]. Subsequently, COX proteins acquired one to three different inserts—or more appropriately INsertions and DELETion (INDEL)s (59)—along the evolution of alphaproteobacteria, around similar positions but with different sequence length and signatures with respect to those of gammaproteobacteria (fig. S6). These differences indicate independent acquisition of INDELs. Mitochondrial COX3 does not have any of these INDELs, and therefore, it is close to the ancestral COX3 of basal alphaproteobacteria. Nevertheless, the crown of the mitochondrial clade of COX3 has a stem length comparable to that of diverse alphaproteobacterial lineages. For example, the relative distance of the crown encompassing Rhodothalassiales, Iodidimonadales, and Kordiimonadales (RIK) is 1.08 ± 0.08 with respect to crown mitochondria— $n = 15$, hence not significantly different (Fig. 5 and fig. S6).

The branching pattern just described changed completely when phylogenetic trees were reconstructed with the COX3 proteins of the alphaproteobacteria taxa in Table 2 (Fig. 5), which have been selected by criteria completely different from those used in recent papers (1, 6–8). Now, the mitochondrial clade is the latest branching and often clusters with proteins of members of the SERIK and new clades. *Caenispirillum* COX3 always forms the basal branch, generally followed by that of *Tistrella*, which has a different variant of COX operon (Figs. 1 and 5). The tree in Fig. 5, therefore, offers an alternative interpretation for the possible evolution of COX3, an essential component of the aerobic metabolism shared by bacteria and mitochondria. The mitochondrial proteins likely evolved after separation from the ancestor of the alphaproteobacterial lineages of SERIK, Rhizobiales and Rhodospirillales, but maintained the absence of INDELs characteristic of ancestral COX3 such as that of *Caenispirillum*. Among the above lineages, only Iodidimonadales retain a COX3 without INDELs, even if the branching pattern is not resolved enough to indicate a direct connection with basal branches. Nevertheless, in trees such as shown in Fig. 5, the crown distance of mitochondrial COX3 is significantly longer than that of Iodidimonadales (relative mean distance 0.82 ± 0.05 , $n = 9$, $P < 0.01$; inset in Fig. 5). This evidence supports the possibility that the ancestor of extant Iodidimonadales was at least contemporary with that of protomitochondria. Although devoid of catalytic centers, COX3 plays a fundamental role in regulating the oxygen affinity of mitochondrial complex IV (11, 60). Therefore, the molecular history of this protein subunit bears relevance to the evolution of aerobic metabolism in eukaryotes.

Together, our results indicate that very few alphaproteobacteria would be selected by combining various discriminatory criteria (Table 2) for identifying the possible ancestors of protomitochondria. *Iodidimonas* spp. and relatives have a superior probability than other possible candidates because they pass more discriminatory criteria than other bacteria (Table 2). Considering our results combined with phylogenetic data (Fig. 5 and fig. S6), we hypothesize that deep-branching MAGs of the Iodidimonadales such as alpha J134 [Fig. 1C, cf. (61)] may define the phylogenetic space from which protomitochondria originated. Alpha J134 and other MAGs clustering with Iodidimonadales live in ecological niches comparable to those present in Proterozoic oceans (10, 61). In turn, they are phylogenetically close to various MAGs thriving in marine zones with oxygen gradient (2, 61)—the kind of environment that may be nearest to that pervading Proterozoic oceans when protomitochondria evolved (10, 25). Therefore, our data and insights dovetail with the emerging picture that a facultative aerobe was the likely bacterial ancestor of protomitochondria (10).

MATERIALS AND METHODS

The aim of the project leading to this paper was to unveil the metabolic ancestry of protomitochondria (<https://osf.io/t9qze/>, accessed on 7 April 2023). To achieve this, we have followed diverse approaches that have different sources of bias (table S1) (62) and provide alternative data for the study of the bacterial origin of mitochondria mainly because they focus on shared metabolic traits. The fundamental approach used previously was phylogenetic inference (1–8) based on increasingly wider sets of proteins shared by alphaproteobacteria and mitochondria (1, 8). The contribution of complex I protein subunits was predominant in the set of “24 alphamitos COG” introduced by Ettema and coworkers (1, 54) since 36% of all proteins and more than 50% of the amino acids in the concatenated alignments belonged to complex I subunits. We have not considered the protein subunits of complex I (reduced form of nicotinamide adenine dinucleotide–ubiquinone oxidoreductase) nor those of complex II (succinate dehydrogenase) because these enzyme complexes are present in basically all alphaproteobacteria (Table 3), including those that do not have a central metabolism equivalent to that of aerobic eukaryotes (11, 53). We have excluded other traits that are shared by eukaryotes and alphaproteobacteria for similar and other reasons, listed in Table 3. The primary focus of our work was on aerobic metabolism, which has been undervalued recently because the protein traits defining this metabolism have been systematically underrepresented: 11.1% of all proteins in the expanded list of (8) and 16.7% in the 24 alphamitos COG set (1).

Choice of aerobic traits and their quantitative analysis

The cytochrome part of the respiratory chain forms the core of the aerobic metabolism of mitochondria (Fig. 1) and in bacteria is contributed by 25 proteins that are mostly shared with eukaryotes: 10 in the COX operon (Fig. 1A) and separate genes for Cox15, Synthesis of Cytochrome c Oxidase (SCO), and Cox5B (Table 1); 9 for the Ccm system (system I) of cytochrome c maturation (Ccm) (63); 6 for complex III, including the Cbp3 chaperone (24) and the two MPP-related CPs (20); and 1 for the soluble cytochrome c that functions as an electron acceptor for complex III and electron donor for complex IV. Several homologs of this soluble cytochrome are often

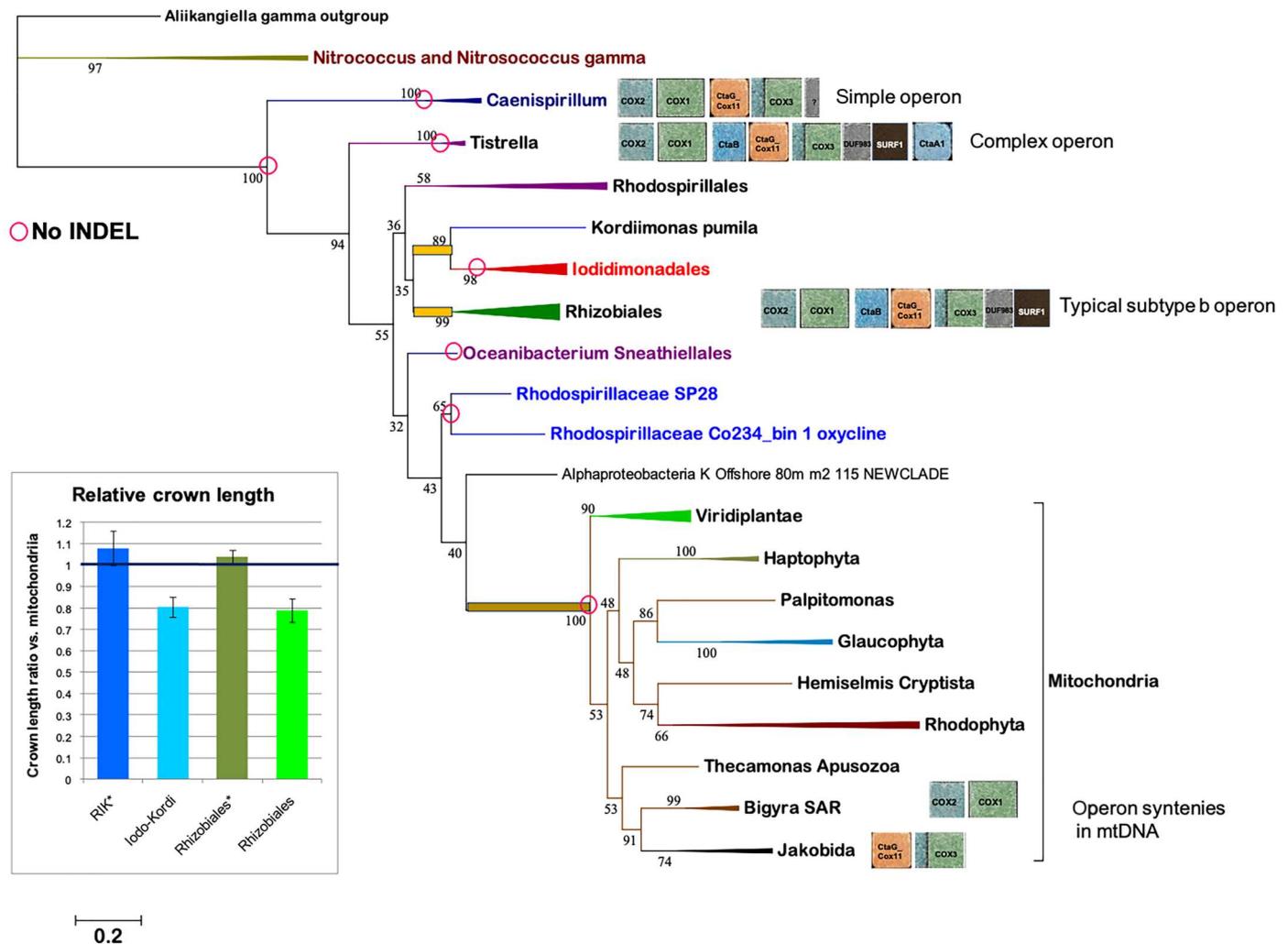


Fig. 5. Phylogenetic tree of COX3 proteins from the alphaproteobacteria of Table 2. The figure shows a representative of several ML trees obtained with an alignment of 46 COX3 proteins and different models of programs (IQ-Tree, Phy-ML3, and MEGA5). The manually curated alignments included three outgroup sequences from gammaproteobacteria such as *Nitrococcus* (Fig. 1A), 24 from alphaproteobacteria (most of those in Table 2, plus *Caenispirillum salinarum* and *Tistrella bauzanensis* to strengthen the clades of *Caenispirillum* and *Tistrella*, respectively), and also a representative of the new clade (2) and 19 mitochondrial sequences from the eukaryotic taxa listed in table S3. In all cases, the alignments had 317 amino acid sites. See fig. S6 for a much larger ML tree including most taxa examined here. COX operon variants of various alphaproteobacteria are inserted on the right of the respective clade, rendered as in Fig. 1A. The stem of the crown group of Iodidimonadales plus Kordiimonas (Iodo-Kordi in the inset of the left) and that of Rhizobiales are indicated by the orange bar. The length of these stems has been normalized to that of crown mitochondria (brownish bar) to evaluate the relative distance to crown mitochondria, proportional to divergence times estimated with Bayesian inference (78, 79); the mean values plus SD of the relative distance data are presented in the inset, with darker colored histograms representing data from $n = 15$ ML trees containing many more alphaproteobacterial COX3 proteins as in fig. S6. The pink circles annotate the branches with COX3 proteins devoid of the INDELS that are scattered among alphaproteobacteria (59), as shown in fig. S6.

present in bacterial genomes; here, we systematically selected the proteins showing the highest homology to mitochondrial cytochrome *c* of *Andalucia* and other early-branching eukaryotes. We found no eukaryotic homologs for two proteins of the Ccm system, i.e., CcmD and CcmI, while CcmG is too similar to various redox proteins for identifying specific eukaryotic homologs. Moreover, CcmB is extremely similar to CcmA and therefore could not be considered as a separate protein trait, as in the case of MPPalpha versus MPPbeta of complex III. Last, we excluded CtaB/Cox10 because it is not encoded in any mtDNA nor is part of plant or protist complex IV. The total number of remaining proteins

considered as separate aerobic traits was around 20, to which we added: alternative oxidase (AOX), the only terminal oxidase other than COX in eukaryotes (64); multiple forms M16A and M16C peptidases that frequently occur in eukaryotic genomes [(28, 65); www.ebi.ac.uk/merops/cgi-bin/famsum?family=M16, last accessed on 7 April 2023]; a Cys signature toward the C terminus of COX1 lying in a conserved region at the negative side of the membrane, which is common in taxa with reduced genomes such as Rickettsiales, more prone to differential gene loss than other lineages (6). We then constructed a preliminary presence/absence table for all the above traits in our selection of alphaproteobacteria and eukaryotes,

detailed below, assigning a value of 1 for complete and 0.5 for incomplete sequences. A -1 penalty was given for the absence of COX1 and COX2 of A1 type COX, the type of catalytic subunits of mitochondrial complex IV which must be present in possible ancestors of mitochondria (11). At difference with previous presence-absence analyses (2, 7, 11, 15, 16, 54), we further assigned a premium score of 2 for the following syntenic associations: (i) COX1-COX2, forming a collinearity in the mtDNA of *Andalucia* (15) and marine members of the TSAR super-group (66)—they are also fused together in the mtDNA of Amoebozoa (11); (ii) Cox11 and COX3, which are syntenic in the mtDNA of *Andalucia* (15) and other early-branching eukaryotes (fig. S1A); and (iii) M16B and ISP, which form a rare synteny (Fig. 1). We assigned a value of 1 to the genes for M16B.016 peptidases closest to eukaryotic MPP (23)—all part of the sister clade of alphaproteobacteria proteins (65) which were identified by specific signatures in the aligned sequences—when such genes were associated to the COX operon, either directly or intermixed with Tsy or M32 (Fig. 1). A value of 0.5 was given for genomic associations separated by three or less other genes from that of SURF1 (Fig. 1B). In addition, we assigned a premium of 1 whenever the gene for CcmF, the critical heme lyase of System I for cytochrome *c* biogenesis (63), was surrounded by at least two other Ccm genes, which generally did not correspond to those considered as separate traits here (Table 1). Multiple genes for the same trait within a genome were generally quantified with a numeral equivalent to such genes. Bacterial SCO and eukaryotic AOX showed such relatively uncommon situations (table S3). Conversely, multiple genes for M16A and M16C peptidases frequently occur in eukaryotic genomes (28). In such cases, we capped the maximal numeral at 5 (Table 1). In sum, the cumulative score of the various aerobic traits that we have analyzed here provides a multilayered and exhaustive evaluation of the aerobic metabolism of bacteria and eukaryotes that has never been considered before. To verify how the overall profile of the cumulative aerobic traits was influenced by the total number of different combinations of such traits, we carried out comparative plots of the median values. An example of these comparisons is presented in fig. S2A, showing a similar overall profile along the alphaproteobacterial lineages we have selected to represent the whole class (see below). Using all 23 traits produced a more flattened distribution of values than in our previous analysis of 18 traits (fig. S2A), reported in the BioRxiv precursor version of this paper (59). We then selected the intermediate set of 20 traits listed in Table 1 to represent the quantitative distribution of aerobic metabolism (Fig. 2).

Taxonomic sampling of mitochondria and grouping of alphaproteobacterial lineages

We applied our approaches to quantify the distribution of aerobic and other metabolic traits by making a thoughtful selection of all currently available genomes of alphaproteobacteria (including Magnetococcales), lithotrophic gammaproteobacteria, and eukaryotes with aerobic mitochondria. We excluded other bacteria because they do not have the subtype of COX operon that is characteristic of alphaproteobacteria and, in part, of mitochondria from early-branching eukaryotes [Fig. 1, cf. (11)]. Moreover, the MPP proteins of eukaryotes have been recently reported to be closely related to the M16B proteins of alphaproteobacteria (65), confirming our previous results (59). The selection of the genomes for our analyses was based on extensive efforts to thoroughly deal with the

issue of taxonomic sampling, most critical for evaluating the bacterial origin of mitochondria (1–8, 32, 54). The initial efforts focused on eukaryotic taxa that have aerobic mitochondria and can be considered early branching, in the sense that they occupy basal branches in the various super-groups of eukaryotes (45) or have a large number of protein-coding genes in their mtDNA (15, 67, 68). The latter feature was particularly important in our selection because of the limited number of available mtDNA genomes that code at least one Ccm gene for cytochrome *c* biogenesis, a critical precondition for undertaking a balanced comparison in the aerobic traits of alphaproteobacteria and mitochondria. More than 95% of alphaproteobacterial genomes have several genes of the Ccm operon, which has been subsequently vertically inherited by eukaryotes (15, 16, 63). Four or less of such genes are coded in the mtDNA of the following eukaryotes: plants, Jakobida (15, 17), *Diphylleia* (16), *Palpitomons* (68), and Cryptista such as *Microheliella*, *Ancoracysta*, Malawimonadida, and Cyanidiales among Rhodophyta (69). With the exception of Ciliophora such as *Tetrahymena* (70), other major super-groups of eukaryotes have System III for cytochrome *c* biogenesis (63, 71), consisting of a single heme lyase without bacterial precedents. We therefore considered only a few taxa, including *Techamonas*, that have the eukaryotic innovation of system III (50, 53) and none having its possible precursor, system V restricted to Euglenozoa (71). After detailed analysis, in part, used for other aspects of this work, we also excluded Rhodelphida (69) because their genome is incomplete and does not encode the Ccm system. Considering the limited number of nuclear genomes that are currently available for Discoba and Archaeplastidia other than plants, we settled to analyze a set of 20 eukaryotes predominantly having Ccm proteins (85%) to represent aerobic mitochondria (table S3). Permutations of some taxa of this set with other eukaryotes, either having Ccm proteins or lacking them as Bigyra (SAR), did not fundamentally alter the distribution of the aerobic traits in mitochondria and their cumulative score, which maintained a median value around 24 (Fig. 2).

Next, we endeavored to produce a thorough selection of alphaproteobacteria representing the major lineages of the class with their specific metabolic traits. Recognizing that it is indispensable to consider the spectrum of alphaproteobacteria diversity to evaluate their possible relationships with protomitochondria, we built the import repository of 314 genomes (publicly available at <https://osf.io/t9qze/>) representing all the major lineages of the class (2) plus the group of MarineProteo1 (1). The genomes were systematically reannotated for all the genes encoding proteins involved in aerobic metabolism and other metabolic pathways examined. Iterative Position-Specific Iterated BLAST (PSI-BLAST) searches of representative proteins (2, 58) were carried out to evaluate their completeness and re-annotation congruity. Our previous taxonomic analysis (2) guided the selection of 15 separate lineages of alphaproteobacteria, spanning early-branching Rickettsiales to late-branching Rhodobacterales. Intermediate lineages also corresponded to orders in the current taxonomy of alphaproteobacteria (www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=28211, accessed on 11 April 2022), which essentially derives from a recent reclassification (56). However, we considered as separate lineages two groups that most likely belong to the Rhodospirillales superorder (2) due to phylogenomic and other observations, as follows. HIMB59 and related marine MAGs were clustered together with the lineage called “MarineAlpha,” which

includes other marine bacteria with similarly AT-rich genomes, in particular those classified under the TMED109 and TMED127 orders in GTDB taxonomy (72). Holosporales, which generally cluster with clades of Rhodospirillales after correcting for their AT-rich genomes (7, 8), was considered an intermediate lineage between Rickettsiales and Rhodospirillales, a placement sustained by their low cumulative scores of aerobic traits (Fig. 2). We used around 60 taxa to encompass the wide genomic diversity of the Rhodospirillales superorder, including at least two taxa for each major subdivision (2). We routinely kept the recently identified new clades of marine MAGs as a separate lineage from related Sneathiellales (including also Minwuiiales) because of the diversity in their bioenergetic traits (2). We also kept the lineage of Iodidimonadales separate from other SERIK groups (2) essentially because of their unique synteny of M16B-ISP (Fig. 1 and fig. S1). To increase the genomic diversity of Iodidimonadales, we included two alphaproteobacterial MAGs found in an anoxic hydrothermal niche (61), proteins of which cluster together with those of Iodidimonas in phylogenetic trees (for example, COX3 in fig. S6). We have also re-evaluated the raw metagenomic data from (61) to expand the genome completeness of alpha J134, as described below. Conversely, we merged the order of Rhodothalassiales with Kordiimonadales, as well as that of Zavarziniales with Emcibacterales, by considering common features such as the number of INDELS in respiratory proteins (59). Representatives of all major families were included to encompass the genomic diversity of Rhizobiales (2, 56). In contrast, only the deepest branching and major genera were included to represent the limited taxonomic diversity of the lineages of Pelagibacterales, Sphingomonadales, Caulobacterales, and Rhodobacterales. The Pelagibacterales lineage was placed just before the Rhizobiales lineage in accordance with recent phylogenetic analyses (2, 56). We did not consider the orders of Parvularculales and Maricaulales because they contain late-branching taxa that often cluster with representatives of the established orders of Caulobacterales or Rhodobacterales (2, 56). However, we considered members of the Micropepsales (72, 73) in some analyses. We found that a small group of MAGs listed among Micropepsales in GTDB Release 07-RS207 (8 April 2022) (https://gtdb.ecogenomic.org/searches?s=gt&q=o__Micropepsales, accessed on 18 August 2022) clustered within the clade of Rhizobiales, having similar sets of aerobic traits (fig. S2C). In all cases, the composition of the 15 alphaproteobacterial lineages was balanced to provide a good match in the relative frequency of bioenergetic traits with respect to the complete set of available genomes, as previously reported for NosZ (2). We verified that substituting several taxa listed in table S3 with other taxa of the most diverse lineages, Rhodospirillales and Rhizobiales, hardly changed the median cumulative score of aerobic traits. Although we favored the inclusion of cultivated taxa with a complete genome, in several cases, this was not possible because the lineage included only MAGs, as in the case of MarineProteo1 (1, 8). In such lineages, we considered MAGs with the most complete genomes, as deduced from the GTDB database or our own evaluations conducted as described earlier (2). MarineProteo1_Bin1 was kept despite its poor coverage because of its prototypic position in the MarineProteo1 lineage (1, 8). In sum, the lineage-dependent plots of various traits we present in this work reflect divergence time along the x axis, essentially following the consistent branching order of currently known alphaproteobacteria (2).

Analysis of genomic data

We used the genomic region from *Iodidimonas muriae* whole-genome assembly [National Center for Biotechnology Information (NCBI) accession number: BMOV01000000, contig BMOV01000004] as a reference to find clusters of genes equivalent to those of Fig. 1A in the metagenomic data of (61), reported under NCBI Bioproject PRJNA392119. First, we manually inspected the reported genome of *I. muriae* (74) to verify the correctness of its annotation, detecting one artifact from the annotation process (accession GGO11109.1), while another gene within the COX operon (accession GGO11122.1) was reannotated as a DUF983 domain-containing protein. Three MAGs found to cluster with Iodidimonadales (Fig. 5 and fig. S6) were downloaded: RFID00000000.1 (J067), RFKS00000000.1 (J134), and RFIU00000000.1 (J084). Using the available raw data, each metagenome was reassembled using SPADes v3.14.1 with default parameters. The obtained assembly was used as a database to search for the genes of the clustered COX-M16B-bc1 operons of *I. muriae* (Fig. 1A), using the NCBI tBLASTN suite with default parameters. The selected contigs thus obtained were used for manual inspection and curation using the genome browser Artemis v18.1.0. Among the contigs examined, contig RFKS01000109 of alpha J134 was found to contain the sequence of genes starting from partial COX1, essentially as previously reported (61). To improve and fully reconstruct the genomic region of interest, a metagenomic reassembly was performed using the Sequence Read Archive (SRA) toolkit with the fasterq-dump program and --split-files flag to obtain the original fastq files in four reported samples (SRR7905022, SRR7905023, SRR7905024, and SRR7905025) (61). From the resulting new assembly, we extracted contig 829 which was 33.2 kb long and contained the complete series of the *I. muriae* genomic cluster, plus a DNA insertion after the M16B gene that might encode for a hypothetical protein (Fig. 1C). We could not firmly validate this possibility since contig RFKS01000109 included a different DNA sequence after the M16B gene. However, we later found that the recently reported MAG, Rhodothalassiaceae bacterium KatS3mg119 (75), presents the same gene sequence uncovered for J134 (Fig. 1C), with the insertion of an aryl esterase gene between M16B peptidase and ISP. Notably, the proteins of this MAG closely cluster with those of J067, which is early branching among Iodidimonadales. The genome assembly of J067 and related J084 (61) was more fragmented than that of J134; even if partial gene clusters of the COX-bc1 operons were present in some fragments, the complete synteny with the *I. muriae* cluster (Fig. 1A) could not be confirmed.

Distribution of the traits for ceramide and kynurenine biosynthesis

The second approach that we used followed a conventional absence-presence analysis (table S5). We first conducted PSI-BLAST searches of putative homologs of *Caulobacter* SPT (29, 31) in all the genomes of our in-house repository, plus those of closely related taxa. An E value cutoff of 5×10^{-68} was generally sufficient to differentiate genuine SPT homologs from related α -oxoamine enzymes such as 5-aminolevulinate synthase. The genomic regions containing the gene encoding the identified SPT homologs were then carefully inspected to verify the presence of the four-gene operon that is common in ceramide-synthesizing bacteria (Fig. 3A) and the eventual vicinity of other genes required for sphingolipid biosynthesis (31). Each protein of the operon was then analyzed in detailed

alignments to identify its biochemical nature (table S5). Various bacterial SPT proteins were then used in PSI-BLAST searches extended to eukaryotes, focusing on the species selected for the comparative analysis of the aerobic traits (table S5). We considered both the active and inactive isoforms of eukaryotic SPT, which likely derive from a process of duplication and differentiation of the single *spt* gene of bacteria (Fig. 3B) (30). An equivalent strategy was used for identifying bacterial and eukaryotic homologs of *Pseudomonas* kynureninase, encoded by the *kynU* gene which is often adjacent to *kynA* and other genes for components of the kynurenine pathway (35, 36).

Analysis of enzymes involved cardiolipin metabolism

Cardiolipin or diphosphatidylglycerol is a typical membrane lipid of prokaryotes that is present almost exclusively in the inner membrane of mitochondria (14, 37–42). In this work, we have tackled the complicated issue of the evolutionary pathways of cardiolipin biosynthesis and metabolism (14, 37, 40). The issue is complicated because there are two fundamentally different enzymes that directly synthesize cardiolipin, which have a scattered distribution in prokaryotes and eukaryotes (14, 37, 41). We compiled a detailed analysis of the distribution of enzymes for cardiolipin biosynthesis in alphaproteobacteria (fig. S5C and table S6). After finding some alphaproteobacteria taxa that have both types of cardiolipin synthase, Cls-AT and Cls_pld (Fig. 3E and table S6), we realized that our results would support a previous proposal (41): the ancestral state of early eukaryotes might have contained both types of the synthase. We thus undertook a detailed analysis of the following enzymes involved in cardiolipin metabolism: (i) PGP synthase, PgsA (44); (ii) cardiolipin synthase of the CDP-AT type, Cls-AT; (iii) cardiolipin synthase of the PLD type, Cls_pld, and; (iv) cardiolipin-specific phospholipase of the cld1 type, typical of yeast (39) but with clear homologs in alphaproteobacteria, as verified by phylogenetic analysis. Given the variety of CDP-AT proteins present in some alphaproteobacterial genomes, especially in MAGs belonging to the Rhodospirillales “superorder,” and considering the limited differences between PgsA and Cls-AT (37), we undertook detailed sequence analysis using manually curated alignments of all these proteins, taking into consideration known three-dimensional features from available structures (43, 44). We were able to distinguish homologs to PgsA by the presence of a conserved Arg residue that is involved in PGP binding—R108 in the structure of *Staphylococcus* PgsA (44), indicated by the red arrow in the alignment block of fig. S4B. Cls-AT proteins from either bacteria or mitochondria substitute this Arg with hydrophobic amino acids (fig. S4B), while proteins of other families of the CDP-AT superfamily have different sequence features in the same protein region (44). On the basis of these molecular differences, we could consider the CDP-AT proteins with elongated N terminus found in Iodidimonadales (table S5) as possible relatives of Cls-AT. Phylogenetic analysis further confirmed this probable relatedness (Fig. 3D), indicating that such CDP-AT proteins with elongated N termini may represent an ancestral form of A family CDP-AT including bacterial and mitochondrial Cls-AT (44). We cross-checked the genomic distribution of cardiolipin-synthesizing enzymes with the documented presence of cardiolipin in bacteria, obtaining a good correspondence between the absence of such genes (table S5) and the lack of cardiolipin in *Caulobacter* spp. (47). Similar good correlations

were obtained with alphaproteobacterial lineages such as Rickettsiales (37), for which no biochemical evidence of cardiolipin exists.

Analysis of anaerobic traits

We have expanded the analysis of the kynurenine pathway to evaluate whether such a pathway might be involved in de novo biosynthesis of RQ (36) in alphaproteobacteria. We cross-checked this analysis with the evaluation of the taxonomic distribution of the *rqaA* gene found in *Rhodospirillum rubrum*, which enables the conversion of RQ from preformed Q (50). We found *rqaA* distribution to be mutually exclusive with the combination of two genes for UbiA plus the *kynU* gene adjacent to other genes for the kynurenine pathway (see text, cf. Fig. 4). This mutual exclusivity has been verified in the genomes of the genus *Rhodoferax* known to contain RQ (76). For example, *Rhodoferax fermentans* has *rqaA* (protein accession WP_078364963) but not *kynU*, whereas *Rhodoferax sediminis* has the *kynU* gene for a functional kynureninase (protein accession WP_142820272) but no *rqaA* homolog. The presence of the *ubiTUV* triad of genes required for the anaerobic biosynthesis of Q (52) was identified by the genomic collinearity of these genes and protein alignments defining the distinctive signatures of the coded proteins (2), including the conserved Cys ligands of the [4Fe4S] cluster (51, 52). BLAST searches extended to eukaryotes identified diverse paralogs with the same U32 peptidase domain as UbiU (52), including composite proteins showing the fusion of this domain with the DUF3656 domain as in the protein coded by *E. coli* *rlhA* (52, 77), often in combination with a similar, UbiV-like domain fused at the C terminus. Most frequently, such proteins were found among members of the SAR super-group (45) and Chlorophyta; however, similar proteins are present in Magnetococcales and other deep-branching alphaproteobacteria. Eukaryotic protein hits found among invertebrates were identified instead as likely bacterial contaminants on the basis of dedicated BLAST searches. Identification of the different types of OFORs was undertaken following the molecular analysis described earlier (53). Their presence was quantified assigning a value of 1 to each monomeric or heterodimeric type and of 0.5 whenever a protein was incomplete. OFORs involved in the biosynthesis of photosynthetic pigments were ignored.

Phylogenetic and molecular analysis of proteins

Phylogenetic analysis was conducted with large alignments of diverse protein sequences (2, 58). Such alignments were manually implemented after rounds of automated alignment with the MUSCLE program within the MEGA software versions 5 and X (2, 58). Short gaps that were specific for one or a few proteins were deleted, while the N and C termini were minimally trimmed to preserve potential signatures. Phylogenetic analysis was undertaken with ML inference using the program IQ-Tree and Bayesian inference using the BEAST program or neighbor-joining with the MEGA5 program as previously described (58). The Le and Gascuel (LG) and Whelan And Goldman (WAG) models were most frequently used for tree reconstruction. Calculation of the relative stem length of crown groups in phylogenetic trees was undertaken by a modification of the methods reported recently (78, 79), normalizing the distance values from the root branch to the value of the mitochondrial clade = 1. See the legend of Fig. 5 for further details.

Similarity analysis with top alphaproteobacterial hits in BLAST searches

To provide an independent quantitative approach to selecting alphaproteobacteria taxa for the vicinity of their respiratory proteins to mitochondrial homologs, we undertook systematic searches with PSI-BLAST in the NCBI database (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp>, accessed on 24 August 2022). The searches were extended to all alphaproteobacteria, including unclassified and *incertae sedis*. The top five hits, which provided a good balance between specificity and limited noise in the similarity data (80), were tabulated and the frequency of multiple hits was computed for each bacterial taxon. This analysis focused on the nuclear-encoded proteins of the COX-M16B/MPP-bc1 gene cluster (Fig. 1A). We found that the results of MPPbeta queries showed the largest presence of hits (43%) among the alphaproteobacterial taxa previously selected for other analyses (table S3). They are presented in fig. S2C (orange symbols, each representing an individual taxon). The alphaproteobacteria in this figure included some Micropepsales (73) with discrete frequencies of hits.

Statistical analysis

Statistical analysis was conducted with the 99% confidence interval of the *t* test (81). In particular, we conducted independent *t* tests to verify the two-tailed hypothesis that the cumulative scores of the 20 aerobic traits (Table 1) for each lineage significantly differed from mitochondria. To guard against inflated type 1 errors (i.e., false positives) from multiple testing, these analyses were conducted at the more stringent 1% significance level ($\alpha = 0.01$). All analyses were run in R (version 4.1.0) using the stats package.

Supplementary Materials

This PDF file includes:

Supplementary Text

Figs. S1 to S6

Tables S1 to S6

References

REFERENCES AND NOTES

1. J. Martijn, J. Vosseberg, L. Guy, P. Offre, T. J. G. Ettema, Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature* **557**, 101–105 (2018).
2. M. A. Cevallos, M. Degli Esposti, New alphaproteobacteria thrive in the depths of the ocean with oxygen gradient. *Microorganisms* **10**, 455 (2022).
3. M. W. Gray, G. Burger, B. F. Lang, Mitochondrial evolution. *Science* **283**, 1476–1481 (1999).
4. M. P. Ferla, J. C. Thrash, S. J. Giovannoni, W. M. Patrick, New rRNA gene-based phylogenies of the *Alphaproteobacteria* provide perspective on major groups, mitochondrial ancestry and phylogenetic instability. *PLOS ONE* **8**, e83383 (2013).
5. C. Esser, N. Ahmadinejad, C. Wiegand, C. Rotte, F. Sebastiani, G. Gelius-Dietrich, K. Henze, E. Kretschmann, E. Richly, D. Leister, D. Bryant, M. A. Steel, P. J. Lockhart, D. Penny, W. Martin, A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol. Biol. Evol.* **21**, 1643–1660 (2004).
6. F. S. P. Nagies, J. Brueckner, F. D. K. Tria, W. F. Martin, A spectrum of verticality across genes. *PLOS Genet.* **16**, e1009200 (2020).
7. L. Fan, D. F. Wu, V. Goremykin, J. Xiao, Y. B. Xu, S. Garg, C. L. Zhang, W. F. Martin, R. X. Zhu, Phylogenetic analyses with systematic taxon sampling show that mitochondria branch within Alphaproteobacteria. *Nat. Ecol. Evol.* **4**, 1213–1219 (2020).
8. S. A. Muñoz-Gómez, E. Susko, K. Williamson, L. Eme, C. H. Slamovits, D. Moreira, P. López-García, A. J. Roger, Site-and-branch-heterogeneous analyses of an expanded dataset favour mitochondria as sister to known Alphaproteobacteria. *Nat. Ecol. Evol.* **6**, 253–262 (2022).
9. H. Imachi, M. K. Nobu, N. Nakahara, Y. Morono, M. Ogawara, Y. Takaki, Y. Takano, K. Uematsu, T. Ikuta, M. Ito, Y. Matsui, M. Miyazaki, K. Murata, Y. Saito, S. Sakai, C. Song, E. Tasumi, Y. Yamanaka, T. Yamaguchi, Y. Kamagata, H. Tamaki, K. Takai, Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* **577**, 519–525 (2020).
10. D. B. Mills, R. A. Boyle, S. J. Daines, E. A. Sperling, D. Pisani, I. C. J. Donoghue, T. M. Lenton, Eukaryogenesis and oxygen in Earth history. *Nat. Ecol. Evol.* **6**, 520–532 (2022).
11. M. Degli Esposti, B. Chouaia, F. Comandatore, E. Crotti, D. Sasser, P. M.-J. Lievens, D. Daffonchio, C. Bandi, Evolution of mitochondria reconstructed from the energy metabolism of living bacteria. *PLOS ONE* **9**, e96566 (2014).
12. L. Fan, D. Wu, V. Goremykin, K. Trost, M. Knopp, C. Zhang, W. F. Martin, R. Zhu, Reply to: Phylogenetic affiliation of mitochondria with Alpha-II and Rickettsiales is an artefact. *Nat. Ecol. Evol.* **6**, 1832–1835 (2022).
13. T. Thiergart, G. Landan, M. Schenk, T. Dagan, W. F. Martin, An evolutionary network of genes present in the eukaryote common ancestor polls genomes on eukaryotic and mitochondrial origin. *Genome Biol. Evol.* **4**, 466–485 (2012).
14. H.-F. Tian, J.-M. Feng, J.-F. Wen, The evolution of cardiolipin biosynthesis and maturation pathways and its implications for the evolution of eukaryotes. *BMC Evol. Biol.* **12**, 32 (2012).
15. G. Burger, M. W. Gray, L. Forget, B. F. Lang, Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. *Genome Biol. Evol.* **5**, 418–438 (2013).
16. R. Kamikawa, T. Shiratori, K.-I. Ishida, H. Miyashita, A. J. Roger, Group II intron-mediated trans-splicing in the Gene-Rich Mitochondrial Genome of an Enigmatic Eukaryote, *Diphyllia rotans*. *Genome Biol. Evol.* **8**, 458–466 (2016).
17. A. Yabuki, Y. Gyaltsen, A. A. Heiss, K. Fujikura, E. Kim, *Ophirina amphinema* n. gen., n. sp., a new deeply branching discobid with phylogenetic affinity to Jakobids. *Sci. Rep.* **8**, 16219 (2018).
18. L. Zhou, M. Maldonado, A. Padavannil, F. Guo, J. A. Letts, Structures of *Tetrahymena*'s respiratory chain reveal the diversity of eukaryotic core metabolism. *Science* **376**, 831–839 (2022).
19. Z. L. Zhang, L. S. Huang, V. M. Shulmeister, Y.-I. Chi, K. K. Kim, L.-W. Hung, A. R. Crofts, E. A. Berry, S.-H. Kim, Electron transfer by domain movement in cytochrome *bc₁*. *Nature* **392**, 677–684 (1998).
20. H. P. Braun, U. K. Schmitz, Are the 'core' proteins of the mitochondrial *bc₁* complex evolutionary relics of a processing protease? *Trends Biochem. Sci.* **20**, 171–175 (1995).
21. M. Maldonado, F. Guo, J. A. Letts, Atomic structures of respiratory complex III₂, complex IV, and supercomplex III₂-IV from vascular plants. *eLife* **10**, e62047 (2021).
22. S. Iwata, J. W. Lee, K. Okada, J. K. Lee, M. Iwata, B. Rasmussen, T. A. Link, S. Ramaswamy, B. K. Jap, Complete structure of the 11-subunit bovine mitochondrial cytochrome *bc₁* complex. *Science* **281**, 64–71 (1998).
23. A. E. Aleshin, S. Gramatikova, G. L. Hura, A. Bobkov, A. Y. Strongin, B. Stec, J. A. Tainer, R. C. Liddington, J. W. Smith, Crystal and solution structures of a prokaryotic M16B peptidase: An open and shut case. *Structure* **17**, 1465–1475 (2009).
24. M. H. Barros, G. P. McStay, Modular biogenesis of mitochondrial respiratory complexes. *Mitochondrion* **50**, 94–114 (2020).
25. M. Müller, M. Mentel, J. J. van Hellemond, K. Henze, C. Woehle, S. B. Gould, R. Y. Yu, M. van der Giezen, A. G. M. Tielens, W. F. Martin, Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol. Mol. Biol. Rev.* **76**, 444–495 (2012).
26. N. Kubo, S.-I. Arimura, N. Tsutsumi, K.-I. Kadowaki, M. Hirai, Isolation and characterization of the pea cytochrome *c* oxidase Vb gene. *Genome* **49**, 1481–1489 (2006).
27. R. García-Villegas, Y. Camacho-Villasana, M. Á. Shingú-Vázquez, A. Cabrera-Orefice, S. Uribe-Carvajal, T. D. Fox, X. Pérez-Martínez, The Cox1 C-terminal domain is a central regulator of cytochrome *c* oxidase biogenesis in yeast mitochondria. *J. Biol. Chem.* **292**, 10912–10925 (2017).
28. R. Grinter, I. D. Hay, J. N. Song, J. W. Wang, D. Teng, V. Dhanasekaran, J. J. Wilksch, M. R. Davies, D. Littler, S. A. Beckham, I. R. Henderson, R. A. Strugnell, G. Dougan, T. Lithgow, FusC, a member of the M16 protease family acquired by bacteria for iron piracy against plants. *PLOS Biol.* **16**, e2006026 (2018).
29. R. J. Olea-Ozuna, S. Poggio, EdBergström, E. Quiroz-Rocha, D. A. García-Soriano, D. X. Sahonero-Canavesi, J. Padilla-Gómez, L. Martínez-Aguilar, I. M. López-Lara, J. Thomas-Oates, O. Geiger, Five structural genes required for ceramide synthesis in *Caulobacter* and for bacterial survival. *Environ. Microbiol.* **23**, 143–159 (2021).
30. H. Ikushiro, M. M. Islam, H. Tojo, H. Hayashi, Molecular characterization of membrane-associated soluble serine palmitoyltransferases from *Sphingobacterium multivorum* and *Bdellovibrio stolpii*. *J. Bacteriol.* **189**, 5749–5761 (2007).
31. J. Padilla-Gómez, R. J. Olea-Ozuna, S. Contreras-Martínez, O. Morales-Tarré, D. A. García-Soriano, D. X. Sahonero-Canavesi, S. Poggio, S. Encarnación-Guevara, I. M. López-Lara, O. Geiger, Specialized acyl carrier protein used by serine palmitoyltransferase to synthesize sphingolipids in *Rhodobacteria*. *Front. Microbiol.* **13**, 961041 (2022).
32. M. Degli Esposti, Bioenergetic evolution in proteobacteria and mitochondria. *Genome Biol. Evol.* **6**, 3238–3251 (2014).
33. S. A. Muñoz-Gómez, J. G. Wideman, A. J. Roger, C. H. Slamovits, The origin of mitochondrial cristae from alphaproteobacteria. *Mol. Biol. Evol.* **34**, 943–956 (2017).

34. A. Alonso, F. M. Goñi, The physical properties of ceramides in membranes. *Annu. Rev. Biophys.* **47**, 633–654 (2018).
35. R. S. Phillips, Structure, mechanism, and substrate specificity of kynureninase. *Biochim. Biophys. Acta* **1814**, 1481–1488 (2011).
36. G. Salinas, D. N. Langelaa, J. N. Shepherd, Rhodoquinone in bacteria and animals: Two distinct pathways for biosynthesis of this key electron transporter used in anaerobic bioenergetics. *Biochim. Biophys. Acta Bioenerg.* **1861**, 148278 (2020).
37. L. A. Luévano-Martínez, A. L. Duncan, Origin and diversification of the cardiolipin biosynthetic pathway in the Eukarya domain. *Biochem. Soc. Trans.* **48**, 1035–1046 (2020).
38. M. Degli Esposti, Lipids, cardiolipin and apoptosis: A greasy licence to kill. *Cell Death Differ.* **9**, 234–236 (2002).
39. T. Klecker, B. Westermann, Pathways shaping the mitochondrial inner membrane. *Open Biol.* **11**, 210238 (2021).
40. I. M. Lopez-Lara, O. Geiger, Bacterial lipid diversity. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* **1862**, 1287–1299 (2017).
41. F. Noguchi, G. Tanifuji, M. W. Brown, K. Fujikura, K. Takishita, Complex evolution of two types of cardiolipin synthase in the eukaryotic lineage stramenopiles. *Mol. Phylogenet. Evol.* **101**, 133–141 (2016).
42. M. Serricchio, P. Butikofer, An essential bacterial-type cardiolipin synthase mediates cardiolipin formation in a eukaryote. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E954–E961 (2012).
43. G. Sciarra, O. B. Clarke, D. Tomasek, B. Kloss, S. Tabuso, R. Byfield, R. Cohn, S. Banerjee, K. R. Rajashankar, V. Slavkovic, J. H. Graziano, L. Shapiro, F. Mancina, Structural basis for catalysis in a CDP-alcohol phosphotransferase. *Nat. Commun.* **5**, 4068 (2014).
44. B. W. Yang, H. B. Yao, D. F. Li, Z. F. Liu, The phosphatidylglycerol phosphate synthase PgsA utilizes a trifurcated amphipathic cavity for catalysis at the membrane-cytosol interface. *Curr. Res. Struct. Biol.* **3**, 312–323 (2021).
45. F. Burki, A. J. Roger, M. W. Brown, A. G. B. Simpson, The new tree of eukaryotes. *Trends Ecol. Evol.* **35**, 43–55 (2020).
46. T. C. Sung, R. L. Roper, Y. Zhang, S. A. Rudge, R. Temel, S. M. Hammond, A. J. Morris, B. Moss, J. Engebrecht, M. A. Frohman, Mutagenesis of phospholipase D defines a superfamily including a trans-Golgi viral protein required for poxvirus pathogenicity. *EMBO J.* **16**, 4519–4530 (1997).
47. A. J. Desiervo, A. D. Homola, Analysis of *Caulobacter crescentus* lipids. *J. Bacteriol.* **143**, 1215–1222 (1980).
48. T. Iino, M. Ohkuma, Y. Kamagata, S. Amachi, Iodidimonas muriae gen. nov., sp. nov., an aerobic iodide-oxidizing bacterium isolated from brine of a natural gas and iodine recovery facility, and proposals of Iodidimonadaceae fam. nov., Iodidimonadales ord. nov., Emcibacteraceae fam. nov. and Emcibacterales ord. nov. *Int. J. Syst. Evol. Microb.* **66**, 5016–5022 (2016).
49. W. B. Ball, J. K. Neff, V. M. Gohil, The role of nonbilayer phospholipids in mitochondrial structure and function. *FEBS Lett.* **592**, 1273–1290 (2018).
50. C. W. Stairs, L. Eme, S. A. Munoz-Gomez, A. Cohen, G. Dellaire, J. N. Shepherd, J. P. Fawcett, A. J. Roger, Microbial eukaryotes have adapted to hypoxia by horizontal acquisitions of a gene involved in rhodoquinone biosynthesis. *eLife* **7**, e34292 (2018).
51. S. S. Abby, K. Kazemzadeh, C. Vraginau, L. Pelosi, F. Pierrel, Advances in bacterial pathways for the biosynthesis of ubiquinone. *BBA-Bioenergetics* **1861**, 148259 (2020).
52. L. Pelosi, C. D. T. Vo, S. S. Abby, L. Loiseau, B. Rascalou, M. H. Chehade, B. Faivre, M. Gousse, C. Chenal, N. Touati, L. Binet, D. Cornu, C. D. Fyfe, M. Fontecave, F. Barras, M. Lombard, F. Pierrel, Ubiquinone biosynthesis over the entire O₂ range: Characterization of a conserved O₂-independent pathway. *MBio* **10**, e01319-19 (2019).
53. P. Y. T. Chen, B. Li, C. L. Drennan, S. J. Elliott, A reverse TCA cycle 2-oxoacid: Ferredoxin oxidoreductase that makes C-C bonds from CO₂. *Joule* **3**, 595–611 (2019).
54. M. E. Schön, V. V. Zlatogursky, R. P. Singh, C. Poirier, S. Wilken, V. Mathur, J. F. H. Strasser, J. Pinhasi, A. Z. Worden, P. J. Keeling, T. J. G. Ettema, J. G. Wideman, F. Burki, Single cell genomics reveals plastid-lacking Picozoa are close relatives of red algae. *Nat. Commun.* **12**, 6651 (2021).
55. V. I. Verhoeve, J. J. Gillespie, Origin of rickettsial host dependency unravelled. *Nat. Microbiol.* **7**, 1110–1111 (2022).
56. A. Hördt, M. G. Lopez, J. P. Meier-Kolthoff, M. Schleuning, L. M. Weinhold, B. J. Tindall, S. Gronow, N. C. Kyrpides, T. Woyke, M. Goker, Analysis of 1,000+ type-strain genomes substantially improves taxonomic classification of *Alphaproteobacteria*. *Front. Microbiol.* **11**, 468 (2020).
57. S. Stefanovic, D. W. Rice, J. D. Palmer, Long branch attraction, taxon sampling, and the earliest angiosperms: Amborella or monocots? *BMC Evol. Biol.* **4**, 35 (2004).
58. M. Degli Esposti, A. Moya-Beltran, R. Quatrini, L. Hederstedt, Respiratory heme A-containing oxidases originated in the ancestors of iron-oxidizing bacteria. *Front. Microbiol.* **12**, 664216 (2021).
59. M. Degli Esposti, O. Geiger, A. F. Sanchez-Flores, M. Degli Esposti, On the bacterial ancestry of mitochondria: New insights with triangulated approaches. *bioRxiv* (2022).
60. K. Shinzawa-Itoh, H. Aoyama, K. Muramoto, H. Terada, T. Kurauchi, Y. Tadehara, A. Yamasaki, T. Sugimura, S. Kurono, K. Tsujimoto, T. Mizushima, E. Yamashita, T. Tsukihara, S. Yoshikawa, Structures and physiological roles of 13 integral lipids of bovine heart cytochrome c oxidase. *EMBO J.* **26**, 1713–1725 (2007).
61. L. M. Ward, A. Idei, M. Nakagawa, Y. Ueno, W. W. Fischer, S. E. McGlynn, Geochemical and metagenomic characterization of Jinata onsen, a proterozoic-analog hot spring, reveals novel microbial diversity including iron-tolerant phototrophs and thermophilic lithotrophs. *Microbes. Environ.* **34**, 278–292 (2019).
62. D. A. Lawlor, K. Tilling, G. D. Smith, Triangulation in aetiological epidemiology. *Int. J. Epidemiol.* **45**, 1866–1886 (2017).
63. A. F. Verissimo, F. Dalldal, Cytochrome c biogenesis system I: An intricate process catalyzed by a maturase supercomplex? *BBA-Bioenergetics* **1837**, 989–998 (2014).
64. A. Atteia, R. van Lis, J. J. van Hellemond, A. G. M. Tielen, W. Martin, K. Henze, Identification of prokaryotic homologues indicates an endosymbiotic origin for the alternative oxidases of mitochondria (AOX) and chloroplasts (PTOX). *Gene* **330**, 143–148 (2004).
65. C. Garrido, F. A. Wollman, I. Lafontaine, The evolutionary history of peptidases involved in the processing of organelle-targeting peptides. *Genome Biol. Evol.* **14**, evac101 (2022).
66. J. G. Wideman, A. Monier, R. Rodriguez-Martinez, G. Leonard, E. Cook, C. Poirier, F. Maguire, D. S. Milner, N. A. T. Irwin, K. Moore, A. E. Santoro, P. J. Keeling, A. Z. Worden, T. A. Richards, Unexpected mitochondrial genome diversity revealed by targeted single-cell genomics of heterotrophic flagellated protists. *Nat. Microbiol.* **5**, 154–165 (2020).
67. Y. Nishimura, K. Kume, K. Sonehara, G. Tanifuji, T. Shiratori, K. Ishida, T. Hashimoto, Y. Inagaki, M. Ohkuma, Mitochondrial genomes of hemiarma marina and leucocryptos marina revised the evolution of cytochrome c maturation in cryptista. *Front. Ecol. Evol.* **8**, 140 (2020).
68. A. Yabuki, Y. Inagaki, K. Ishida, Palpitomonas bilix gen. et sp. nov.: A novel deep-branching heterotroph possibly related to archaeplastida or hacrobia. *Protist* **161**, 523–538 (2010).
69. R. M. R. Gawryluk, D. V. Tikhonenkov, E. Hehenberger, F. Husnik, A. P. Mlynikov, P. J. Keeling, Non-photosynthetic predators are sister to red algae. *Nature* **572**, 240–243 (2019).
70. C. F. Brunk, L. C. Lee, A. B. Tran, J. L. Li, Complete sequence of the mitochondrial genome of Tetrahymena thermophila and comparative methods for identifying highly divergent genes. *Nucleic Acids Res.* **31**, 1673–1682 (2003).
71. A. Belbelazi, R. Neish, M. Carr, J. C. Mottram, M. L. Ginger, Divergent cytochrome c maturation system in kinetoplastid protists. *MBio* **12**, e00166-21 (2021).
72. P. A. Chaumeil, A. J. Mussig, P. Hugenholtz, D. H. Parks, GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* **36**, 1925–1927 (2020).
73. A. B. Harbison, L. E. Price, M. D. Flythe, S. L. Brauer, Micropepsis pineolensis gen. nov., sp. nov., a mildly acidophilic alphaproteobacterium isolated from a poor fen, and proposal of Micropepsaceae fam. nov. within Micropepsales ord. nov. *Int. J. Syst. Evol. Microb.* **67**, 839–844 (2017).
74. L. H. Wu, J. C. Ma, The Global Catalogue of Microorganisms (GCM) 10K type strain sequencing project: Providing services to taxonomists for standard genome sequencing and annotation. *Int. J. Syst. Evol. Microb.* **69**, 895–898 (2019).
75. S. Kato, S. Masuda, A. Shibata, K. Shirasu, M. Ohkuma, Insights into ecological roles of uncultivated bacteria in Katase hot spring sediment from long-read metagenomics. *Front. Microbiol.* **13**, 1045931 (2022).
76. H. Miyadera, A. Hiraishi, H. Miyoshi, K. Sakamoto, R. Mineki, K. Murayama, K. V. P. Nagashima, K. Matsuura, S. Kojima, K. Kita, Complex II from phototrophic purple bacterium Rhodoferrax fermentans displays rhodoquinol-fumarate reductase activity. *Eur. J. Biochem.* **270**, 1863–1874 (2003).
77. S. Kimura, Y. Sakai, K. Ishiguro, T. Suzuki, Biogenesis and iron-dependency of ribosomal RNA hydroxylation. *Nucleic Acids Res.* **45**, 12974–12986 (2017).
78. J. F. H. Strasser, I. Irisarri, T. A. Williams, F. Burki, Author correction: A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. *Nat. Commun.* **12**, 3574 (2021).
79. S. S. Wang, H. W. Luo, Dating alphaproteobacteria evolution with eukaryotic fossils. *Nat. Commun.* **12**, 3324 (2021).
80. M. Degli Esposti, L. Lozano, E. Martinez-Romero, Current phylogeny of Rhodospirillaceae: A multi-approach study. *Mol. Phylogenet. Evol.* **139**, 106546 (2019).
81. B. Gerald, A brief review of independent, dependent and one sample t-test. *Int. J. Appl. Math. Theor. Phys.* **4**, 50–54 (2018).
82. M. Degli Esposti, D. Cortez, L. Lozano, S. Rasmussen, H. B. Nielsen, E. M. Romero, Alpha proteobacterial ancestry of the [Fe-Fe]-hydrogenases in anaerobic eukaryotes. *Biol. Direct* **11**, 34 (2016).
83. J. Pyrih, T. Panek, I. M. Durante, V. Raskova, K. Cimrhanzlova, E. Kriegova, A. D. Tsaousis, M. Elias, J. Lukes, Vestiges of the bacterial signal recognition particle-based protein targeting in mitochondria. *Mol. Biol. Evol.* **38**, 3170–3187 (2021).

84. M. Degli Esposti, M. Mentel, W. Martin, F. L. Sousa, Oxygen reductases in alphaproteo-bacterial genomes: Physiological evolution from low to high oxygen environments. *Front. Microbiol.* **10**, 499 (2019).

Acknowledgments: We are indebted to Michelle Degli Esposti for invaluable contributions of statistical analyses and the OSL repository. We thank D. Gonzalez-Halphen, E. Martinez-Romero, M. Angel Cevallos, M. Maldonado, S. Amachi, C. Martinez-Gutierrez, L. Hederstedt, and B. Martin for discussion and feedback. We also thank L. Lozano for technical assistance. **Funding:** This work was supported by Dirección General de Asuntos del Personal Académico (IN202223)/ Universidad Nacional Autónoma de México (to O.G. and J.P.-G.) and Consejo Nacional de Ciencia y Tecnología de México (118 in Investigación en Fronteras de la Ciencia) (to O.G.).

Author contributions: The contributions by the authors were as follows. Conceptualization: M.D.E., O.G., and A.S.-F. Methodology: M.D.E., J.P.-G., and A.S.-F. Investigation: M.D.E., O.G., J.P.-G., and A.S.-F. Visualization: M.D.E., J.P.-G., O.G., and A.S.-F. Supervision: O.G. and M.D.E. Writing—original draft: M.D.E., O.G., and A.S.-F. Writing—review and editing: M.D.E., O.G., J.P.-G., and

A.S.-F. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Tables S3 to S6 have been deposited in Dryad (doi:10.5061/dryad.9p8cz8wn0) and also in the repository OSF (<https://osf.io/t9qze/>). Additional materials deposited in the Dryad repository include the following: Import folder, which includes all the raw .xls files to import and search using the R code "Search Function"; R script "Search Function", which searches all the .xls files in the Import folder; and Contig829.fasta with the DNA sequence of the reassembled metagenomic DNA for alpha J134. See Materials and Methods for further information on these materials.

Submitted 7 February 2023

Accepted 11 May 2023

Published 9 August 2023

10.1126/sciadv.adh0066

Multiple approaches of cellular metabolism define the bacterial ancestry of mitochondria

Otto Geiger, Alejandro Sanchez-Flores, Jonathan Padilla-Gomez, and Mauro Degli Esposti

Sci. Adv., **9** (32), eadh0066.
DOI: 10.1126/sciadv.adh0066

View the article online

<https://www.science.org/doi/10.1126/sciadv.adh0066>

Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)