Mitochondrial Evolution

Michael W. Gray

Centre for Comparative Genomics and Evolutionary Bioinformatics, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3M 4R2, Canada

Correspondence: m.w.gray@dal.ca

Viewed through the lens of the genome it contains, the mitochondrion is of unquestioned bacterial ancestry, originating from within the bacterial phylum α-Proteobacteria (*Alphaproteobacteria*). Accordingly, the endosymbiont hypothesis—the idea that the mitochondrion evolved from a bacterial progenitor via symbiosis within an essentially eukaryotic host cell—has assumed the status of a theory. Yet mitochondrial genome evolution has taken radically different pathways in diverse eukaryotic lineages, and the organelle itself is increasingly viewed as a genetic and functional mosaic, with the bulk of the mitochondrial proteome having an evolutionary origin outside *Alphaproteobacteria*. New data continue to reshape our views regarding mitochondrial evolution, particularly raising the question of whether the mitochondrion originated after the eukaryotic cell arose, as assumed in the classical endosymbiont hypothesis, or whether this organelle had its beginning at the same time as the cell containing it.

In 1970, Lynn Margulis published *Origin of Eukaryotic Cells*, an influential book that effectively revived the long-standing but mostly moribund idea that mitochondria and plastids (chloroplasts) evolved from free-living bacteria via symbiosis within a eukaryotic host cell (Margulis 1970). The discovery in the 1960s of DNA within these organelles together with the recognition that they contain a translation system distinct from that of the cytosol were two of the observations that Margulis marshaled in support of the endosymbiont hypothesis of organelle origins. Indeed, throughout her career, Margulis forcefully argued that symbiosis is a potent but largely unrecognized and unappreciated force in evolution (Margulis 1981). Technological developments in DNA cloning and sequencing in the 1970s and 1980s opened the way to the detailed characterization of mitochondrial genomes and genes, and the generation of key molecular data that were instrumental in affirming a bacterial origin of the mitochondrial and plastid genomes, allowing researchers to pinpoint the extant bacterial phyla to which these two organelles are most closely related. Over the past several decades, numerous reviews have documented in detail the biochemical and molecular and cell biological data bearing on the endosymbiont hypothesis of organelle origins (Gray 1982, 1983, 1989a,b, 1992, 1993, 1999; Gray and Doolittle 1982; Wallace 1982; Cavalier-Smith 1987b, 1992; Gray and Spencer 1996; Andersson and Kurland 1999; Gray et al. 1999, 2001, 2004; Lang et al. 1999; Andersson et al. 2003; Burger et al. 2003a; Bullerwell and Gray 2004). Various endosymbiotic

M.W. Gray

models proposed over the years have been comprehensively critiqued (Martin et al. 2001), while the debates surrounding the endosymbiont hypothesis have been recounted in an engaging perspective that traces the development of ideas regarding organelle origins (Sapp 1994). Within a historical context, the present article emphasizes more recent data and insights that are relevant to continuing questions regarding how mitochondria originated and have since evolved.

WHAT DO GENETIC, GENOMIC, AND PHYLOGENOMIC DATA TELL US ABOUT THE ORIGIN OF MITOCHONDRIA?

The genetic function of mitochondrial DNA (mtDNA) was first fully revealed by complete sequencing of the ∼16-kb mitochondrial genome from several mammalian species (Anderson et al. 1981, 1982; Bibb et al. 1981). This work established the paradigm that mtDNA encodes a small number (13 in mammals) of protein subunits of the mitochondrial electron transport chain and ATP synthase, as well as the ribosomal RNA (rRNA) and transfer RNA (tRNA) components of a mitochondrial translation system. The paradigm was revised when investigations of mtDNA from non-animal species showed quite extraordinary variation in size, physical form, coding capacity, organizational patterns, and modes of expression across the eukaryotic domain (Gray et al. 1998). In some non-animal taxa, additional proteins are encoded in mtDNA, notably extra respiratory proteins and ribosomal proteins (Table 1). On the other hand, some mitochondrial genomes have been reduced drastically in size, losing many of the protein genes encoded in animal mtDNA as well as some or all mtDNA-encoded tRNA genes. At ∼6 kb in size, the mitochondrial genome of *Plasmodium falciparum* (human malaria parasite) and related apicomplexans is the smallest known, harboring only three protein genes, highly fragmented and rearranged small subunit (SSU) and large subunit (LSU) rRNA genes, and no tRNA genes (Feagin 2000). In marked contrast, within land plants, mtDNA has expanded substantially in size (>200 kb) if not in coding

capacity, with the largest known mitochondrial genome in this lineage (∼11,000 kb) exceeding the size of some bacterial and even some nuclear genomes (Sloan et al. 2012). What was evident even early on is that none of the initial mtDNAs investigated by detailed sequencing, including animal mtDNAs, looks anything like a typical bacterial genome in the way in which genes are organized and expressed.

Ribosomal RNA genes are among the few genes universally encoded by mtDNA across eukaryotes, and the corresponding rRNA sequences were exploited early on in phylogenetic reconstructions aimed at elucidating their evolutionary origin. Although mitochondrial large subunit (LSU) and small subunit (SSU) rRNAs are in general very different in size and secondary structure compared with their cytosolic and prokaryotic counterparts, they retain a sufficient degree of primary sequence and secondary structure correspondence that they can be incorporated into the aligned sequence databases on which these phylogenetic reconstructions are based (Gray et al. 1984, 1989; Cedergren et al. 1988). These initial phylogenetic trees showed that mitochondrial rRNA sequences, and presumably the genomes encoding them, emanate from within the domain Bacteria (eubacteria) and not from within Archaea (archaebacteria) or Eucarya (eukaryotes). Plant mitochondrial rRNAs are especially slowly evolving and bacteria-like (Schnare and Gray 1982; Spencer et al. 1984) and were instrumental in pinpointing the α-class of Proteobacteria (*Alphaproteobacteria*) as the specific bacterial lineage from within which they originated (Yang et al. 1985). Subsequent phylogenetic reconstructions using a much larger number of both mitochondrial and bacterial rRNA sequences have consistently confirmed this affiliation and, moreover, pointed to the Rickettsiales, one of six or more orders within *Alphaproteobacteria* (Williams et al. 2007)—and comprising obligate parasites such as *Rickettsia*, *Anaplasma*, and *Ehrlichia*—as especially close relatives of mitochondria (Gray and Spencer 1996; Gray 1998).

These phylogenetic conclusions were considerably bolstered by the sequencing in 1997 of a minimally derived (ancestral) mitochon-
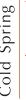
Table 1. Mitochondrial DNA-encoded genes and their functions

(1) Coupled electron transport–oxidative phosphorylation (ATP synthesis)	
Complex I (NADH:ubiquinone oxidoreductase)	**nad1**, **2**, **3**, **4**, **4L**, **5**, **6**, *7, 8, 9, 10, 11*
Complex II (succinate:ubiquinone oxidoreductase)	*sdh 2, 3, 4*
Complex III (ubiquinol:cytochrome *c* oxidoreductase)	**cob**
Complex IV (cytochrome *c*:O$_2$ oxidoreductase)	**cox1**, **2**, **3**
Complex V (F$_1$F$_0$ ATP synthase)	*atp1, 3,* **4**,[a] **6**, **8**,[b] **9**
(2) Translation	
Ribosomal RNAs	**rnl** (LSU), **rns** (SSU), *rrr5* (5S)
Ribosomal proteins	
Small subunit (SSU)	*rps1, 2, 3, 4, 7, 8, 10, 11, 12, 13, 14, 19*
Large subunit (LSU)	*rpl1, 2, 5, 6, 10, 11, 14, 16, 18, 19, 20, 27, 31, 32, 34*
Transfer RNAs	**trnA**, **C**, . . . **W**, **Y**
Elongation factor	*tufA*
tm RNA (unstalling of translation)	*ssrA*
(3) Transcription	
Core RNA polymerase	*rpoA, B, C*
Sigma factor	*rpoD*
(4) RNA processing	
RNase P RNA (5′ tRNA processing)	*rnpB*
(5) Protein import	
ABC transporter	*ccmA* (*yejV*), *ccmB* (*yejW*)
Heme delivery	*ccmC* (*yejU*)
SecY-type transporter	*secY*
Sec-independent transporter	*tatA* (*mttA*),[c] *tatC* (*mttB*)
(6) Protein maturation	
Cytochrome oxidase assembly	*cox11*
Heme *c* maturation	*ccmF* (*yejR*)

A subset of the genes is encoded in the mtDNA of various eukaryotes; for example, those in bold are found in human and other mammalian mitochondrial genomes. Only the most ancestral (gene-rich) mtDNAs, for example, those of jakobid flagellates such as *Reclinomonas americana*, encode all or almost all of this gene set. Rare genes of uncertain origin and function include *dpo* (plasmid-derived DNA polymerase), *rpo* (plasmid-derived RNA polymerase), *rtl* (reverse transcriptase), *mutS* (DNA repair), and *dam* (methyltransferase). See Gray et al. (2004) for details.

[a]Also referred to as *orf25* and *ymf39* (Burger et al. 2003b).
[b]Also referred to as *orfB* and *ymf19* (Gray et al. 1998).
[c]Jacob et al. (2004).

drial genome from *Reclinomonas americana*, a member of an obscure group of protists (eukaryotic microbes) termed jakobid flagellates. The ∼69-kb, circular-mapping mtDNA of *R. americana* not only contains more genes than any other characterized mitochondrial genome (including genes specifying 5S rRNA and RNase P RNA) (Table 1), but it also displays evident bacterial characteristics not seen in combination in any other mtDNA, such as operon-like gene clusters, highly bacteria-like rRNA and tRNA secondary structures, and putative Shine–Dalgarno motifs upstream of

protein-coding genes. Indeed, *R. americana* mtDNA—so different from that of animal, plant, fungal, and many other protist mt-DNAs—has been dubbed "a eubacterial genome in miniature," so striking is its resemblance to a typical bacterial genome (Lang et al. 1997).

At the same time, sequencing of the first Rickettsiales genome, that of *Rickettsia prowazekii*, showed it to be a markedly reduced bacterial genome, superficially resembling mitochondria in its dependence on a host cell (Andersson et al. 1998). However, the genomes of the mitochondrion and members of Rickettsiales are

M.W. Gray

clearly the products of independent evolutionary reduction, implying that mitochondria were not derived directly from a Rickettsiales taxon (Andersson et al. 1998; Gray 1998); rather, these two groups share a more remote common ancestor. Although the mitochondria–Rickettsiales connection has been a consistent phylogenetic finding (Viale and Arakaki 1994; Gupta 1995; Sicheritz-Pontén et al. 1998; Lang et al. 1999), it is still not certain whether the two are sister groups, or whether mitochondria actually branch within Rickettsiales, which is composed of two distinct families, Rickettsiaceae and Anaplasmataceae (Williams et al. 2007). In several studies, mitochondria appear to be more closely related to the former family, containing various *Rickettsia* species, than to the latter, comprising the genera *Anaplasma*, *Ehrlichia*, and *Wolbachia* (Karlin and Brocchieri 2000; Emelyanov 2001a,b, 2003a,b).

The specific affiliation of mitochondria and Rickettsiales in phylogenetic trees, although robust, has been questioned (e.g., Esser et al. 2004) on the grounds that this inferred relationship might be a phylogenetic artifact due to the high rate of sequence divergence and elevated A+T content of the genomes of Rickettsiales taxa and mitochondria—in other words, a long-branch-attraction (LBA) artifact. For that reason, there has been considerable interest in expanding the database of both mitochondrial and bacterial sequences and in applying a more comprehensive phylogenomics approach (based on many genes) to the reconstruction of phylogenetic trees, in combination with multiple and more sophisticated algorithms, rigorous statistical evaluation methods, and more realistic evolutionary models.

Of particular interest in broadening taxon sampling has been the identification of free-living members of *Alphaproteobacteria* that may be specifically affiliated with the parasitic group Rickettsiales (Williams et al. 2007). These previously unknown α-proteobacterial lineages came to light through the Global Oceanic Survey, metagenomic data from which revealed that oceanic α-Proteobacteria are abundant, with one particular clade (termed SAR11) comprising 30%–40% of total oceanic cell counts. The 1.3-Mb genome of one SAR11 member, *Candidatus Pelagibacter ubique*, is the smallest known genome, encoding the fewest genes, of any free-living bacterium (Giovannoni et al. 2005). Several recent reports have suggested that the SAR11 clade including *Ca. P. ubique* shares a common ancestor with mitochondria, the two together forming a sister group to Rickettsiales (Georgiades et al. 2011; Thrash et al. 2011). Other investigators reject this affiliation, concluding instead that *Ca. P. ubique* is most closely related to soil and aquatic α-Proteobacteria with large genomes (Viklund et al. 2012) and arguing that a rare oceanic group of α-Proteobacteria, termed Oceanic Mitochondria Affiliated Clade (OMAC), represents the closest free-living relatives to mitochondria (Brindefalk et al. 2011). Still other groups have suggested that free-living members of α-proteobacterial orders other than Rickettsiales should be considered as possible sources of the mitochondrial progenitor (Esser et al. 2004; Atteia et al. 2009).

These differing conclusions emphasize the challenges inherent in these sorts of analyses, which have to contend with various types of systematic error (for review, see Thrash et al. 2011), biased taxon sampling, and the highly restricted gene content of mitochondrial genomes. At present, although we know a great deal regarding the mitochondrial family tree, we have to admit that the identity of the immediate next of kin remains elusive.

Phylogenetic and phylogenomic reconstructions strongly and consistently support a monophyletic mitochondrial assemblage, and therefore a single origin of mitochondria. Two other pieces of genomic evidence support the view that extant mitochondrial genomes share a single common ancestor. First, the genes encoded by mitochondrial genomes are, with few exceptions, a subset of the genes encoded by the most gene-rich mtDNA, that of *R. americana*. Because the mitochondrial genome is considered to be a highly reduced version of a much larger α-proteobacterial progenitor genome, it is extremely unlikely that genome reduction in independently acquired bacterial symbionts would separately converge on the same small repertoire of respiratory-chain and ribosomal-

protein genes. Second, in plant and many ancestral protist mtDNAs, ribosomal protein genes are clustered in the same transcriptional order in which they appear in the corresponding bacterial operons. However, some genes present in the bacterial operons are missing from the corresponding mitochondrial gene clusters, either having been moved elsewhere in the mitochondrial genome or to the nuclear genome, as a result of endosymbiotic gene transfer (EGT). The same gene deletions are seen in the homologous mitochondrial ribosomal gene clusters where these occur throughout eukaryotes, arguing that these deletions were already present in the common ancestor of extant mitochondrial genomes (Gray et al. 1999). This conclusion is reinforced by consideration of various nucleus-encoded mitochondrial proteins that are components of the mitochondrial proteome (see below).

HOW DID THE MITOCHONDRIAL SYMBIOSIS HAPPEN?

Endosymbiotic models for the origin of mitochondria (for review, see Martin et al. 2001) are basically variations on two fundamentally different themes that have been referred to, respectively, as the "archezoan scenario" and the "symbiogenesis scenario" (Koonin 2010). In the archezoan scenario, "The host of the proto-mitochondrial endosymbiont was a hypothetical primitive amitochondrial eukaryote, termed archezoan" (Koonin 2010). In contrast, in the symbiogenesis scenario, "A single endosymbiotic event involving the uptake of an α-Proteobacterium by an archaeal cell led to the generation of the mitochondria," followed subsequently "by the evolution of the nucleus and compartmentalization of the eukaryotic cell" (Koonin 2010). The archezoan scenario most closely approximates the classical endosymbiont hypothesis of mitochondrial origin (Margulis 1970; Doolittle 1980), whereas the hydrogen hypothesis (Martin and Müller 1998) is an example of the symbiogenesis scenario. A fundamental difference between these two scenarios is whether the α-Proteobacterial endosymbiosis that gave rise to the proto-mitochondrion happened at the same

time as and was integral to the formation of the eukaryotic cell (symbiogenesis scenario) or occurred subsequent to the formation of a primitive, amitochondriate cell that served as host and that was already essentially eukaryotic (archezoan scenario).

Archezoan Scenario

A major boost to the classical endosymbiont hypothesis came from early phylogenetic reconstructions, based initially on SSU rRNA sequences, that showed several eukaryotic lineages branching deeply within the eukaryotic domain of the resulting trees. These early-branching lineages, collectively termed "archezoa," consisted of protists such as microsporidians, diplomonads, and parabasalids, living as parasites in anaerobic environments and characterized by the absence of recognizable mitochondria (Cavalier-Smith 1987a, 1989). Members of archezoa, assumed to be primitively amitochondriate (i.e., never having had mitochondria in their evolutionary history: the so-called "archezoa hypothesis"), could be seen as modern representatives of the sort of ancestral eukaryote that might have played host to an α-Proteobacterial symbiont in a classical endosymbiosis scheme.

Two findings led to the ultimate abandonment of the archezoa hypothesis. First, more rigorous phylogenetic reconstructions combined with other sorts of data convincingly showed that the early-branching position of archezoan taxa in the eukaryotic tree is a methodological artifact, due to a relatively high rate of sequence divergence of the archezoan sequences used in the analysis. This property gives rise to an LBA effect that incorrectly positions these taxa at the base of the eukaryotic clade, closest to the outgroup (prokaryotic) sequences used to root the tree. Microsporidia, for example, were eventually recognized as secondarily degenerate fungi rather than as primitive, early-branching eukaryotes (Hirt et al. 1999; Keeling et al. 2000). Indeed, the current eukaryotic tree is more accurately characterized as a bush, with no one lineage clearly the earliest diverging (Keeling et al. 2005; Koonin 2010). Second, in every apparently amitochondrial lineage that

M.W. Gray

has been carefully investigated, evolutionary remnants of mitochondria ("mitochondrion-related organelles," or MROs) have been identified (see below). Thus, we currently know of no extant eukaryotic lineages that are convincingly amitochondrial and that therefore might have been primitively amitochondriate (Embley and Hirt 1998). This conclusion does not mean that such lineages do not exist. We may simply not have discovered them yet, or they may have existed at some point in evolutionary history but have now all become extinct. Nevertheless, rejection of the archezoa hypothesis on phylogenetic grounds coupled with the apparent absence of any known amitochondriate eukaryotic lineages considerably weakens the case for an acquisition of mitochondria via the classical endosymbiont route.

Symbiogenesis Scenario

An alternative view, that the host cell for the mitochondrial endosymbiosis was a prokaryote—and specifically an archaeon, not a eukaryote—has recently gained in prominence (Koonin 2010). The "hydrogen hypothesis" (Martin and Müller 1998) is perhaps the best-known example of a symbiogenesis scenario. This scheme proposes that eukaryotes arose

> . . .through symbiotic association of an anaerobic, strictly hydrogen-dependent, strictly autotrophic archaebacterium (the host) with a eubacterium (the symbiont) that was able to respire, but generated molecular hydrogen as a waste product of anaerobic heterotrophic metabolism. The host's dependence upon molecular hydrogen produced by the symbiont is put forward as the selective principle that forged the common ancestor of eukaryotic cells. (Martin and Müller 1998)

Thus, the hydrogen hypothesis "posits that the origins of the heterotrophic organelle (the symbiont) and the origins of the eukaryotic lineage are identical" (Martin and Müller 1998). A corollary of the hydrogen hypothesis and other symbiogenesis scenarios is that the complexity of the eukaryotic cell and its defining features emerged *after* the mitochondrial symbiosis, rather than *before*.

Several arguments can be advanced against a symbiogenesis scenario for the origin of mitochondria (Koonin 2010). Endocytosis (a eukaryotic hallmark) has long been considered an essential function for incorporating a bacterial symbiont, although it is the case that bacterial endosymbioses (e.g., γ-Proteobacteria inside β-Proteobacteria) have been documented (von Dohlen et al. 2001; Thao et al. 2002). In addition, assuming an α-Proteobacterial symbiont as the mitochondrial progenitor in a partnership that simultaneously gave rise to this organelle and the rest of the eukaryotic cell, one might expect that any eubacterial genetic signal in the nuclear genome would be overwhelmingly α-Proteobacterial. However, although an α-Proteobacterial signal does, in fact, predominate (Pisani et al. 2007), in any given eukaryotic lineage collectively more bacterial-type genes appear to derive from diverse non-α-Proteobacterial lineages or fail to affiliate robustly with any specific bacterial phylum. Nevertheless, it is possible that ancestral lineages contributing to a bacterial–archaeal symbiogenesis might have possessed genomes already "scrambled" to a certain extent by horizontal gene transfer (HGT).

A prominent feature of the hydrogen hypothesis is its claim to account simultaneously for the origins of both aerobic and anaerobic energy metabolism in eukaryotes, the assumption being that both pathways were contained in and contributed to the hybrid cell by the α-Proteobacterial partner. It is supposed that the two pathways would have been differentially expressed in the free-living bacterial symbiont when it encountered the appropriate environmental conditions. The hypothesis further posits that genes for aerobic respiration were lost in those eukaryotic lineages in which the mitochondrion was converted to an MRO, some types of which (e.g., hydrogenosome) function in anaerobic energy metabolism (see below). The hydrogen hypothesis predicts that genes of anaerobic energy metabolism in MROs should have been inherited vertically throughout eukaryotes from a common ancestor, clustering as a monophyletic lineage together with α-Proteobacteria in phylogenetic reconstructions. However, a rigorous study of the

Cite this article as *Cold Spring Harb Perspect Biol* 2012;4:a011403

phylogenetic distributions and histories of proteins involved in anaerobic pyruvate metabolism in eukaryotes has not provided support for this prediction (Hug et al. 2010). Rather, MROs and the enzymatic machinery they contain for anaerobic energy metabolism appear to reflect a high degree of independent and convergent evolution (see below).

Very recently, Lane and Martin (2010) have argued, from a consideration of the energetics of genome complexity, that because eukaryotes encode and express a substantially larger number of proteins than do prokaryotes, this increased expression demands a level of cellular energy that only the mitochondrion is able to satisfy. Accordingly, these investigators view the mitochondrion as the *sine qua non* to eukaryotic genomic and cellular complexity, concluding, rather definitively, that "the host for mitochondria was a prokaryote."

On balance, it would seem that a symbiogenesis scenario (bacterial endosymbiont in an archaeal host) better accommodates the accumulated data relevant to the question of mitochondrion origin than does an archezoan scenario (bacterial endosymbiont in an amitochondriate but essentially eukaryotic host). However, as emphasized above, the latter scenario cannot be entirely discounted at this point. Each scenario raises its own set of issues that are difficult to rationalize without resorting to ad hoc explanations. In the end, each faces the conundrum that there is no straightforward and compelling way to discern how similar the genomes of the proposed prokaryotic ancestors of the eukaryotic cell were to their modern-day descendants.

MITOCHONDRION-RELATED ORGANELLES (MROs)

An extreme form of mitochondrial genome reduction is found in two types of MROs, hydrogenosomes and mitosomes, which entirely lack mtDNA. These two MRO types are distinguished by the fact that hydrogenosomes retain ATP-generating capacity, whereas mitosomes do not. The hydrogenosome, a double-membrane-bound organelle originally discovered in

the parabasalid *Trichomonas vaginalis* (Lindmark and Müller 1973), was subsequently identified in various other anaerobic eukaryotes. The *T. vaginalis* hydrogenosome not only lacks a genome but has an incomplete tricarboxylic acid cycle and electron transport chain and no cytochromes. ATP is generated from pyruvate via a substrate-level pathway comprising a characteristic suite of enzymes, including pyruvate: ferredoxin oxidoreductase and an iron–iron hydrogenase. The organelle takes its name from the fact that molecular hydrogen (H_2) is one of the end products of this pathway. The anaerobic metabolism performed by the hydrogenosome initially suggested that the organelle might have originated through an endosymbiosis with an anaerobic bacterium such as *Clostridium* (Whatley et al. 1979). However, subsequent studies have revealed that the *T. vaginalis* hydrogenosome contains several proteins typical of mitochondria, including chaperonins (Bui et al. 1996), the NADH dehydrogenase module of electron transport complex I (Hrdy et al. 2004), and components of the ISC biosynthesis pathway, the mitochondrial machinery for synthesis of iron–sulfur (Fe–S) clusters (Sutak et al. 2004). These results strongly support the view that the *T. vaginalis* hydrogenosome is a relict mitochondrion.

A second group of double-membrane-bound MROs, in this case unable to generate ATP, has been discovered in a number of anaerobic, parasitic protists that were initially considered to be amitochondriate, including the amoebozoons *Entamoeba histolytica* (Clark and Roger 1995; Mai et al. 1999; Tovar et al. 1999) and *Mastigamoeba balamuthi* (Gill et al. 2007), the microsporidians *Trachipleistophora hominis* (Williams et al. 2002) and *Encephalitozoon cuniculi* (Goldberg et al. 2008; Tsaousis et al. 2008), and the diplomonad *Giardia lamblia* (Tovar et al. 2003). The MROs in these protists are collectively termed "mitosomes" (Embley et al. 2003; Embley 2006; Hjort et al. 2010). Here again, identification of typical mitochondrial proteins in this MRO argues that mitosomes, like hydrogenosomes, are evolutionary derivatives of conventional mitochondria, but are even more highly reduced than hydrogenosomes (see Hjort

M.W. Gray

et al. 2010 for a detailed listing and discussion of relevant data). Very limited metabolic capacity is retained by the mitosome, of which Fe−S cluster formation stands out. This pathway is ubiquitous in both conventional mitochondria and MROs, suggesting that Fe−S cluster formation, rather than oxidative phosphorylation, may be the *raison d'être* of the mitochondrion and its evolutionary derivatives (Tovar et al. 2003).

The discovery of what appear to be transitional forms of MRO in certain anaerobic eukaryotes has recently blurred the distinction between mitochondria, on the one hand, and hydrogenosomes and mitosomes, on the other (Gray 2011). These novel MROs, like conventional hydrogenosomes, are able to generate H_2 via an ATP-producing, hydrogenase-mediated pathway; however, they retain a genome, albeit lacking several typical mtDNA-encoded genes, notably ones specifying components of respiratory complexes III, IV, and V (see Table 1). Thus, these MROs lack the ability to generate ATP via coupled electron transport and oxidative phosphorylation; however, they support a considerably more complex biochemistry than either hydrogenosomes or mitosomes. MROs of this type have been identified in the anaerobic ciliate *Nyctotherus ovalis* (Boxma et al. 2005; de Graaf et al. 2011) and the anaerobic stramenopiles *Blastocystis* sp. (Pérez-Brocal and Clark 2008; Stechmann et al. 2008; Wawrzyniak et al. 2008) and *Proteromonas lacertae* (Pérez-Brocal et al. 2010), relatives of brown algae, diatoms, and oomycetes. In these cases, the MRO genome specifies components of an organellar translation system (rRNAs, tRNAs, ribosomal proteins) as well as subunits of electron transport complexes I and II, suggesting the presence of a partial electron transport chain.

The sporadic phylogenetic distribution of anaerobic MRO-containing eukaryotes and their interspersion with aerobic taxa within particular lineages strongly indicate that MROs have originated independently, multiple times, throughout the eukaryotic domain (Roger and Silberman 2002; Embley et al. 2003; Embley 2006; Hjort et al. 2010). Consequently, many of the characteristics shared between MROs

(e.g., between those of *Nyctotherus* and *Blastocystis*) are almost certainly due to convergent evolution rather than vertical inheritance. The continued study of variously evolved MROs will be important for understanding both the pathways and mechanisms involved in the evolutionary conversion of a conventional mitochondrion to an MRO. At the same time, we will gain a better appreciation of the evolutionary flexibility of mitochondria, a theme considered below in the discussion of mitochondrial proteome evolution.

ORIGIN AND EVOLUTION OF THE MITOCHONDRIAL PROTEOME

Only 27 mitochondrial proteins are encoded by the gene-rich *R. americana* mitochondrial genome, and only three by the most gene-poor mtDNA, that of apicomplexan parasites such as *P. falciparum*. The vast majority of mitochondrial proteins are encoded in the nucleus, synthesized in the cytosol, and imported back into the organelle. Overall, the number of mitochondrial proteins is estimated to range from several hundred in *P. falciparum* to more than 3000 in vertebrate animals (Richly et al. 2003). Relocation of functional genes from the mitochondrion to the nucleus via EGT has contributed significantly to the nucleus-encoded cohort (Gray 1999; Adams and Palmer 2003; Timmis et al. 2004). Occasionally, such genes are transferred in pieces (Adams et al. 2001; Pérez-Martínez et al. 2001; Waller and Keeling 2006; Gawryluk and Gray 2009; Morales et al. 2009), sometimes with a portion remaining in the mitochondrial genome (Oudot et al. 1999; Kück et al. 2000; Nedelcu et al. 2000; Gawryluk and Gray 2010a). Notably, only a minority of the protein products of these relocated genes are imported back into the organelle; most now function in other subcellular compartments (Gabaldón and Huynen 2007). Extensive retailoring of the mitochondrial proteome in the course of evolution has also involved addition of novel proteins and new functions, as discussed below.

Identification of the proteins contributed to the eukaryotic cell by the proto-mitochondrial endosymbiont has been approached through

Cite this article as *Cold Spring Harb Perspect Biol* 2012;4:a011403

comparisons of the gene contents of sequenced α-Proteobacterial and eukaryotic genomes. This strategy has identified at least 840 orthologous groups that bear a clear α-Proteobacterial signature—that is, a close and specific evolutionary relationship to α-Proteobacterial homologs, without evidence of recent HGT (Gabaldón and Huynen 2003, 2007; Szklarczyk and Huynen 2010). Comparisons among α-Proteobacterial genomes suggest that the free-living bacterial ancestor of mitochondria contained about 3000–5000 genes (Boussau et al. 2004), with an upper bound of approximately 1700 ancestral clusters of orthologous genes in the proto-mitochondrial genome (Szklarczyk and Huynen 2010). These estimates indicate that about 1000–3000 genes were lost in the transition from bacterial symbiont to proto-organelle. Significantly, of the more than 800 human genes that display an α-Proteobacterial signature, only about 200 are found in the human mitochondrial proteome, indicating that the proto-mitochondrial contribution to eukaryotic cell evolution and function extends beyond the mitochondrion itself, to encompass other cellular compartments, as well.

In the proto-mitochondrion, the complete electron transport chain and complete pathway for β-oxidation of fatty acids (providing NADH and $FADH_2$ to the former) were likely present, indicating that the mitochondrial endosymbiont had an aerobic metabolism. Also prominently represented were pathways for the synthesis of lipids, biotin, heme, and Fe−S clusters, as well as an abundance of cation transporters. In all, the reconstructed metabolism suggests that the proto-mitochondrion was capable of at least facultative aerobic respiration (Szklarczyk and Huynen 2010). More than half of what remains of this proto-mitochondrial metabolism in modern mitochondria comprises proteins that function in oxidative phosphorylation and translation, including posttranslational modifications, a clear trend toward specialization in energy metabolism. In short, mitochondrial proteome evolution has been characterized by loss of many original functions, retargeting of others to different cellular locations, and wholesale addition of host-derived proteins—a veritable "hijacking of mitochondrial protein synthesis and metabolism" by the eukaryotic cell (Gabaldón and Huynen 2007).

Initial delineation of the mitochondrial proteome relied on various bioinformatics approaches, including homology searches as well as techniques developed to identify amino-terminal mitochondrial targeting signals (MTSs) (e.g., Claros and Vincens 1996; Emanuelsson et al. 2000). Not all imported mitochondrial proteins feature MTSs identifiable in this way, and the accuracy and sensitivity of the search algorithms may be compromised where protein sequences are highly divergent (Richly et al. 2003). An alternative technique that has proven useful is high-throughput immunolocalization of tagged gene products to determine their subcellular localization. Based on these approaches, the yeast mitochondrial proteome has been estimated to comprise about 400–800 proteins, or between ∼7% and ∼13% of the total yeast proteome of approximately 6100 proteins (Karlberg et al. 2000; Marcotte et al. 2000; Kumar et al. 2002).

Considering the clear α-Proteobacterial signature of the mitochondrial genome, we might have anticipated that the mitochondrial proteome would consist largely of α-Proteobacteria-like proteins plus proteins exclusive to the eukaryotic domain, "invented" during the evolutionary transition from bacterial endosymbiont to organelle. Somewhat surprisingly, this expectation has not been met. In surveys of the yeast mitochondrial proteome (Karlberg et al. 2000; Marcotte et al. 2000), only ∼10%–15%—a much smaller proportion than might have been anticipated—proved to originate clearly from the α-Proteobacterial lineage. A larger, generically "prokaryotic" proportion (∼40%–50%) consists of proteins that appear to originate outside α-Proteobacteria but are not necessarily affiliated with any specific bacterial or archaeal clade. Another, "eukaryotic" fraction (∼20%–30%) contains proteins having no obvious homologs in either Archaea or Bacteria. A final, "unique" subset (∼20%) comprises seemingly yeast-specific proteins having no identifiable homologs in other eukaryotes or in prokaryotes. These results indicate that the

M.W. Gray

yeast mitochondrial proteome has multiple evolutionary origins and a complex evolutionary history (Kurland and Andersson 2000; Gray et al. 2001), a conclusion now firmly established for the mitochondria of other eukaryotes (Gabaldón and Huynen 2004; Szklarczyk and Huynen 2010). Bacteriophage-like proteins (notably the mitochondrial RNA polymerase in most eukaryotes) make an additional small contribution to the evolutionary mix (Shutt and Gray 2006).

Comparative mitochondrial genomics, based on complete sequencing of mtDNAs, has told us much regarding the nature of the ancestral mitochondrial genome and about patterns and mechanisms of mitochondrial genome evolution (Gray et al. 1998; Gray 1999). In the same way, comparative mitochondrial proteomics, based on a combination of bioinformatics data mining and direct analysis of whole mitochondria or submitochondrial fractions and complexes (Dreger 2003; Yan et al. 2009), is proving to be an equally incisive approach for unraveling the evolution of the mitochondrial proteome. Investigations using mass spectrometry (MS) have confirmed and extended the initial, bioinformatics-based findings that indicated a mosaic evolutionary origin of the mitochondrial proteome.

MS provides a powerful adjunct to bioinformatics-based approaches in its capacity to reveal novel mitochondrial proteins that have no identifiable sequence homologs and that lack MTSs. At least 13 novel, "ciliate-specific" proteins identified during MS analysis of mitochondria from *Tetrahymena pyriformis* (Smith et al. 2007) were subsequently found as components of the purified mitochondrial F_1F_0-ATP synthase (complex V) of this protist (Nina et al. 2010). These observations highlight an emerging theme of taxon-specific retooling of mitochondrial complexes such as electron transport chain assemblies and ribosomes, only the core α-Proteobacteria-like components of which appear to have been contributed by the proto-mitochondrion. A feature of this retailoring is the addition of novel proteins of generally unknown function, sometimes accompanied by loss of otherwise conserved components. An

example is the ATP synthase (complex V) of the chlorophycean green alga, *Chlamydomonas reinhardtii*, in which nine novel subunits of unknown evolutionary origin replace eight subunits that are otherwise conserved in the ATP synthase of other non-chlorophycean green algae, as well as in plants, animals, and fungi (Lapaille et al. 2010).

Other mitochondrial respiratory complexes, such as complex I (CI; NADH:ubiquinone oxidoreductase), clearly exemplify this retailoring theme. Bacterial CI comprises 14 subunits, all of which are found in the corresponding mammalian complex, with seven of these subunits encoded in mammalian mtDNA (Table 1). A further 18 subunits in mammalian CI are assumed to be eukaryote-specific additions already present in the last eukaryotic common ancestor, because they are ubiquitous throughout eukaryotes but are not found in bacteria (but see below). Thirteen other mammalian CI subunits appeared to display a narrow phylogenetic distribution, being identified initially only in metazoan animals (Brandt 2006).

In attempting to define the ancestral state of selected mitochondrial components and pathways, the importance of taking a comprehensive phylogenetic approach (examining mitochondrial proteomes from as wide a range of eukaryotes as possible) needs to be recognized. As noted above, mitochondrial CI has an additional 18 "eukaryote-specific" subunits that are considered to have been incorporated at the earliest stages of mitochondrial CI evolution (Gabaldón et al. 2005; Brandt 2006). In plants (Heazlewood et al. 2003; Parisi et al. 2004) and green algae (Cardol et al. 2004), mitochondrial CI additionally contains proteins with high similarity to γ-type carbonic anhydrases (γCAs). Comparative studies over a relatively narrow phylogenetic range initially suggested that these proteins represented specific additions in the plant lineage (Parisi et al. 2004). However, a more recent study focusing on protists—eukaryotic microbes, wherein lies most of the evolutionary diversity within the domain Eucarya—has revealed a much broader distribution of mitochondrial γCAs, either shown or presumed to be associated with mitochondrial

Cite this article as *Cold Spring Harb Perspect Biol* 2012;4:a011403

CI (Gawryluk and Gray 2010b), than previously suspected. It appears likely that γCAs were ancestral components of mitochondrial CI and that they were subsequently lost from CI specifically in the evolutionary line leading to animals and fungi (opisthokonts), rather than added to CI specifically in the line leading to plants and algae. These results emphasize the importance of comprehensive taxon coverage in drawing conclusions regarding mitochondrial proteome evolution. In fact, wider taxon sampling has recently demonstrated that almost all of the 18 supposedly metazoan-specific CI subunits were likely already present in CI of the last eukaryotic common ancestor (Cardol 2011; Gawryluk et al. 2012). In the same way, a more phylogenetically broad analysis of homologous bacterial complexes appears warranted, judging by the recent demonstration that CI from the α-Proteobacterium *Paracoccus dentrificans* contains three of the "eukaryote-specific" CI proteins that had long been considered to be absent from bacterial CI (Yip et al. 2011). (In this regard, it is notable that 35 years ago [before the advent of sequence-based phylogenetic reconstructions], John and Whatley [1975] pointed out that *P. denitrificans* "resembles a mitochondrion more closely than do other bacteria, in that it effectively assembles in a single organism those features of the mitochondrial respiratory chain and oxidative phosphorylation which are otherwise distributed at random among most other aerobic bacteria.")

Another mitochondrial complex that shows a profound degree of evolutionary retailoring is the mitochondrial ribosome. In many eukaryotes (e.g., most animals), the mitochondrial LSU and SSU rRNAs have become much smaller than their bacterial counterparts, while at the same time many new ribosomal proteins have been added, so that what was originally an RNA-rich complex is now a protein-rich one (O'Brien 2002). The most extreme example of this ribosome retailoring is seen in the kinetoplastid protozoa, such as *Trypanosoma brucei* (Ziková et al. 2008) and *Leishmania tarentolae* (Sharma et al. 2009). The *T. brucei* mitochondrial ribosome, for example, contains 56 SSU and 77 LSU proteins, compared with the 21 SSU and 34 LSU proteins present in the *Escherichia coli* ribosome. The novel mitoribosomal proteins identified in these analyses do not have detectable homologs outside of the kinetoplastid protozoa and display only a low degree of sequence conservation within this lineage. These observations reinforce the importance of direct MS analyses of isolated mitochondrial complexes in order to accurately determine their composition, given that so many of these components appear to be new, lineage-specific inventions.

Other studies have shown that the ancestral mitochondrial ribosome in the last eukaryotic common ancestor was already much larger than its bacterial ancestor, containing some 19 additional eukaryote-specific proteins (Smits et al. 2007; Desmond et al. 2011). The fact that these novel mitoribosomal proteins are found in all of the currently recognized eukaryotic supergroups is yet another strong argument in favor of a monophyletic origin of contemporary mitochondria, a conclusion in this case based on eukaryote-specific rather than prokaryote-specific features.

CONCLUDING REMARKS

Comparative mitochondrial genomics has offered us a glimpse of what the ancestral mitochondrial genome was like, and what genes it contained. A monophyletic origin of the mitochondrial genome from within α-Proteobacteria is strongly supported, with Rickettsiales most often identified as the α-Proteobacterial order most closely related to mitochondria. This phylogenetic-based approach has revealed that mitochondrial genome evolution has been characterized by massive expansion in some lineages and extreme reduction and compaction in others, with endosymbiotic gene transfer relocating much of the initial genetic information in the proto-mitochondrial genome to the nucleus. Comparative mitochondrial proteomics has provided evidence of a mosaic evolutionary origin of the protein complement of this organelle, with a much smaller proportion of the proteome than might have been anticipated clearly deriving from an α-Proteobacterial progenitor.

M.W. Gray

Mitochondri-specific proteins, pathways, and functions emerging within the eukaryotic lineage subsequent to the α-Proteobacterial endosymbiosis are being identified. Retailoring of key mitochondrial complexes relative to their α-Proteobacterial antecedents is seen to have occurred through addition of novel protein components, often in a narrow, lineage-specific manner. We may anticipate that the unabated deluge of genomic and proteomic data will confront us with currently unappreciated aspects of mitochondrial structure and function across the broad range of eukaryotes, with new data and insights continually reshaping and refining our ideas regarding mitochondrial evolution.

ACKNOWLEDGMENTS

REFERENCES

Adams KL, Palmer JD. 2003. Evolution of mitochondrial gene content: Gene loss and transfer to the nucleus. *Mol Phylogenet Evol* **29:** 380–395.

Adams KL, Ong HC, Palmer JD. 2001. Mitochondrial gene transfer in pieces: Fission of the ribosomal protein gene *rpl2* and partial or complete gene transfer to the nucleus. *Mol Biol Evol* **18:** 2289–2297.

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, et al. 1981. Sequence and organization of the human mitochondrial genome. *Nature* **290:** 457–465.

Anderson S, de Bruijn MHL, Coulson AR, Eperon IC, Sanger F, Young IG. 1982. Complete sequence of bovine mitochondrial DNA conserved features of the mammalian mitochondrial genome. *J Mol Biol* **156:** 683–717.

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

Andersson SGE, Zomorodipour A, Andersson JO, Sicheritz-Pontén T, Alsmark UCM, Podowski RM, Näslund AK, Eriksson A-S, Winkler HH, Kurland CG. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396:** 133–140.

Andersson SG, Karlberg O, Canbäck B, Kurland CG. 2003. On the origin of mitochondria: A genomics perspective. *Philos Trans R Soc Lond B Biol Sci* **358:** 177–179.

Atteia A, Adrait A, Brugière S, Tardif M, van Lis R, Deusch O, Dagan T, Kuhn L, Gontero B, Martin W, et al. 2009. A proteomic survey of *Chlamydomonas reinhardtii* mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of the α-proteobacterial mitochondrial ancestor. *Mol Biol Evol* **26:** 1533–1548.

Bibb MJ, Van Etten RA, Wright CT, Walberg MW, Clayton DA. 1981. Sequence and gene organization of mouse mitochondrial DNA. *Cell* **26:** 167–180.

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

Boxma B, de Graaf RM, van der Staay GW, van Alen TA, Ricard G, Gabaldon T, van Hoek AH, Moon-van der Staay SY, Koopman WJ, van Hellemond JJ, et al. 2005. An anaerobic mitochondrion that produces hydrogen. *Nature* **434:** 74–79.

Brandt U. 2006. Energy converting NADH:quinone oxidoreductase (complex I). *Annu Rev Biochem* **75:** 69–92.

Brindefalk B, Ettema TJG, Viklund J, Thollesson M, Andersson SGE. 2011. A phylometagenomic exploration of oceanic Alphaproteobacteria reveals mitochondrial relatives unrelated to the SAR11 clade. *PLoS ONE* **6:** e24457.

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

Bullerwell CE, Gray MW. 2004. Evolution of the mitochondrial genome: Protist connections to animals, fungi and plants. *Curr Opin Microbiol* **7:** 528–534.

Burger G, Gray MW, Lang BF. 2003a. Mitochondrial genomes: Anything goes. *Trends Genet* **19:** 709–716.

Burger G, Lang BF, Braun H-P, Marx S. 2003b. The enigmatic mitochondrial ORF *ymf39* codes for ATP synthase chain b. *Nucleic Acids Res* **31:** 2353–2360.

Cardol P. 2011. Mitochondrial NADH:ubiquinone oxidoreductase (complex I) in eukaryotes: A highly conserved subunit composition highlighted by mining of protein databases. *Biochim Biophys Acta* **1807:** 1390–1397.

Cardol P, Vanrobaeys F, Devreese B, Van Beeumen J, Matagne RF, Remacle C. 2004. Higher plant-like subunit composition of mitochondrial complex I from *Chlamydomonas reinhardtii*: 31 conserved components among eukaryotes. *Biochim Biophys Acta* **1658:** 212–224.

Cavalier-Smith T. 1987a. Eukaryotes with no mitochondria. *Nature* **326:** 332–333.

Cavalier-Smith T. 1987b. The simultaneous symbiotic origin of mitochondria, chloroplasts, and microbodies. *Ann NY Acad Sci* **503:** 55–71.

Cavalier-Smith T. 1989. Archaebacteria and Archezoa. *Nature* **339:** 100–101.

Cavalier-Smith T. 1992. The number of symbiotic origins of organelles. *Biosystems* **28:** 107–108.

Cedergren R, Gray MW, Abel Y, Sankoff D. 1988. The evolutionary relationships among known life forms. *J Mol Evol* **28:** 98–112.

Clark AG, Roger AJ. 1995. Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica*. *Proc Natl Acad Sci* **92:** 6518–6521.

Claros MG, Vincens P. 1996. Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem* **241:** 779–786.

de Graaf RM, Ricard G, van Alen TA, Duarte I, Dutilh BE, Burgtorf C, Kuiper JWP, van der Staay GWM, Tielens AGM, Huynen MA, et al. 2011. The organellar genome and metabolic potential of the hydrogen-producing mitochondrion of *Nyctotherus ovalis*. *Mol Biol Evol* **28:** 2379–2391.

Desmond E, Brochier-Armanet C, Forterre P, Gribaldo S. 2011. On the last common ancestor and early evolution of eukaryotes: Reconstructing the history of mitochondrial ribosomes. *Res Microbiol* **162:** 53–70.

Doolittle WF. 1980. Revolutionary concepts in evolutionary biology. *Trends Biochem Sci* **5:** 146–149.

Dreger M. 2003. Proteome analysis at the level of subcellular structures. *Eur J Biochem* **270:** 589–599.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* **300:** 1005–1016.

Embley TM. 2006. Multiple secondary origins of the anaerobic lifestyle in eukaryotes. *Philos Trans R Soc Lond B Biol Sci* **361:** 1055–1067.

Embley TM, Hirt RP. 1998. Early branching eukaryotes? *Curr Opin Genet Dev* **8:** 624–629.

Embley TM, van der Giezen M, Horner DS, Dyal PL, Bell S, Foster PG. 2003. Hydrogenosomes, mitochondria and early eukaryotic evolution. *IUBMB Life* **55:** 387–395.

Emelyanov VV. 2001a. Evolutionary relationship of Rickettsiae and mitochondria. *FEBS Lett* **501:** 11–18.

Emelyanov VV. 2001b. Rickettsiaceae, Rickettsia-like endosymbionts, and the origin of mitochondria. *Biosci Rep* **21:** 1–17.

Emelyanov VV. 2003a. Common evolutionary origin of mitochondrial and rickettsial respiratory chains. *Arch Biochem Biophys* **420:** 130–141.

Emelyanov VV. 2003b. Mitochondrial connection to the origin of the eukaryotic cell. *Eur J Biochem* **270:** 1599–1618.

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

Feagin JE. 2000. Mitochondrial genome diversity in parasites. *Int J Parasitol* **30:** 371–390.

Gabaldón T, Huynen MA. 2003. Reconstruction of the proto-mitochondrial metabolism. *Science* **302:** 609.

Gabaldón T, Huynen MA. 2004. Shaping the mitochondrial proteome. *FEBS Lett* **1659:** 212–220.

Gabaldón T, Huynen MA. 2007. From endosymbiont to host-controlled organelle: The hijacking of mitochondrial protein synthesis and metabolism. *PLoS Comput Biol* **3:** e219.

Gabaldón T, Rainey D, Huynen MA. 2005. Tracing the evolution of a large protein complex in the eukaryotes, NADH:ubiquinone oxidoreductase (Complex I). *J Mol Biol* **348:** 857–870.

Gawryluk RMR, Gray MW. 2009. A split and rearranged nuclear gene encoding the iron-sulfur subunit of mitochondrial succinate dehydrogenase in Euglenozoa. *BMC Res Notes* **2:** 16.

Gawryluk RMR, Gray MW. 2010a. An ancient fission of mitochondrial *cox1*. *Mol Biol Evol* **27:** 7–10.

Gawryluk RMR, Gray MW. 2010b. Evidence for an early evolutionary emergence of γ-type carbonic anhydrases as components of mitochondrial respiratory complex I. *BMC Evol Biol* **348:** 857–870.

Gawryluk RMR, Chisholm KA, Pinto DM, Gray MW. 2012. Composition of the mitochondrial electron transport chain in *Acanthamoeba castellanii*: Structural and evolutionary insights. *Biochim Biophys Acta* doi: org/10.1016/j.bbabio.2012.06.005.

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

Gill EE, Diaz-Triviño S, Barberà MJ, Silberman JD, Stechmann A, Gaston D, Tamas I, Roger AJ. 2007. Novel mitochondrion-related organelles in the anaerobic amoeba *Mastigamoeba balamuthi*. *Mol Microbiol* **66:** 1306–1320.

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

Goldberg AV, Molik S, Tsaousis AD, Neumann K, Kuhnke G, Delbac F, Vivares CP, Hirt RP, Lill R, Embley TM. 2008. Localization and functionality of microsporidian iron–sulphur cluster assembly proteins. *Nature* **452:** 624–629.

Gray MW. 1982. Mitochondrial genome diversity and the evolution of mitochondrial DNA. *Can J Biochem* **60:** 157–171.

Gray MW. 1983. The bacterial ancestry of plastids and mitochondria. *BioScience* **33:** 693–699.

Gray MW. 1989a. The evolutionary origins of organelles. *Trends Genet* **5:** 294–299.

Gray MW. 1989b. Origin and evolution of mitochondrial DNA. *Annu Rev Cell Biol* **5:** 25–50.

Gray MW. 1992. The endosymbiont hypothesis revisited. *Int Rev Cytol* **141:** 233–357.

Gray MW. 1993. Origin and evolution of organelle genomes. *Curr Opin Genet Dev* **3:** 884–890.

Gray MW. 1998. Rickettsia, typhus and the mitochondrial connection. *Nature* **396:** 109–110.

Gray MW. 1999. Evolution of organellar genomes. *Curr Opin Genet Dev* **9:** 678–687.

Gray MW. 2011. The incredible shrinking organelle. *EMBO Rep* **12:** 873.

Gray MW, Doolittle WF. 1982. Has the endosymbiont hypothesis been proven? *Microbiol Rev* **46:** 1–42.

Gray MW, Spencer DF. 1996. Organellar evolution. In *Evolution of microbial life* (ed. Roberts D, et al.), pp. 109–126. Cambridge University Press, Cambridge.

Gray MW, Sankoff D, Cedergren RJ. 1984. On the evolutionary descent of organisms and organelles: A global phylogeny based on a highly conserved structural core in small subunit ribosomal RNA. *Nucleic Acids Res* **12:** 5837–5852.

Gray MW, Cedergren R, Abel Y, Sankoff D. 1989. On the evolutionary origin of the plant mitochondrion and its genome. *Proc Natl Acad Sci* **86:** 2267–2271.

Gray MW, Lang BF, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Brossard N, Delage E, Littlejohn TG, et al. 1998. Genome structure and gene content in protist mitochondrial DNAs. *Nucleic Acids Res* **26:** 865–878.

Gray MW, Burger G, Lang BF. 1999. Mitochondrial evolution. *Science* **283:** 1476–1481.

Gray MW, Burger G, Lang BF. 2001. The origin and early evolution of mitochondria. *Genome Biol* **2:** 1018.1–1018.5.

Gray MW, Lang BF, Burger G. 2004. Mitochondria of protists. *Annu Rev Genet* **38:** 477–524.

Gupta RS. 1995. Evolution of the chaperonin families (Hsp60, Hsp10 and Tcp-1) of proteins and the origin of eukaryotic cells. *Mol Microbiol* **15:** 1–11.

Heazlewood JL, Howell KA, Millar AH. 2003. Mitochondrial complex I from *Arabidopsis* and rice: Orthologs of mammalian and fungal components coupled with plant-specific subunits. *Biochim Biophys Acta* **1604:** 159–169.

Hirt RP, Logsdon JM Jr, Healy B, Dorey MW, Doolittle WF, Embley TM. 1999. Microsporidia are related to Fungi: Evidence from the largest subunit of RNA polymerase II and other proteins. *Proc Natl Acad Sci* **96:** 580–585.

Hjort K, Goldberg AV, Tsaousis AD, Hirt RP, Embley TM. 2010. Diversity and reductive evolution of mitochondria among microbial eukaryotes. *Philos Trans R Soc Lond B Biol Sci* **365:** 713–727.

Hrdy I, Hirt RP, Dolezal P, Bardonová L, Foster PG, Tachezy J, Embley TM. 2004. *Trichomonas* hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. *Nature* **432:** 618–622.

Hug LA, Stechmann A, Roger AJ. 2010. Phylogenetic distributions and histories of proteins involved in anaerobic pyruvate metabolism in eukaryotes. *Mol Biol Evol* **27:** 311–324.

Jacob Y, Seif E, Paquet P-O, Lang BF. 2004. Loss of the mRNA-like region in mitochondrial tmRNAs of jakobids. *RNA* **10:** 605–614.

John P, Whatley FR. 1975. *Paracoccus denitrificans* and the evolutionary origin of the mitochondrion. *Nature* **254:** 495–498.

Karlberg O, Canbäck B, Kurland CG, Andersson SGE. 2000. The dual origin of the yeast mitochondrial proteome. *Yeast* **17:** 170–187.

Karlin S, Brocchieri L. 2000. Heat shock protein 60 sequence comparisons: Duplications, lateral transfer, and mitochondrial evolution. *Proc Natl Acad Sci* **97:** 11348–11353.

Keeling PJ, Luker MA, Palmer JD. 2000. Evidence from β-tubulin phylogeny that microsporidia evolved from within the fungi. *Mol Biol Evol* **17:** 23–31.

Keeling PJ, Burger G, Durnford DG, Lang BF, Lee RW, Pearlman RE, Roger AJ, Gray MW. 2005. The tree of eukaryotes. *Trends Ecol Evol* **20:** 670–676.

Koonin EV. 2010. The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biol* **11:** 209.

Kück U, Jekosch K, Holzamer P. 2000. DNA sequence analysis of the complete mitochondrial genome of the green alga *Scenedesmus obliquus*: Evidence for UAG being a leucine and UCA being a non-sense codon. *Gene* **253:** 13–18.

Kumar A, Agarwal S, Heyman JA, Matson S, Heidtman M, Piccirillo S, Umansky L, Drawid A, Jansen R, Liu Y, et al. 2002. Subcellular localization of the yeast proteome. *Genes Dev* **16:** 707–719.

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

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

Lang BF, Burger G, O'Kelly CJ, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Gray MW. 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* **387:** 493–497.

Lang BF, Gray MW, Burger G. 1999. Mitochondrial genome evolution and the origin of eukaryotes. *Annu Rev Genet* **33:** 351–397.

Lapaille M, Escobar-Ramírez A, Degand H, Baurain D, Rodríguez-Salinas E, Coosemans N, Boutry M, Gonzalez-Halphen D, Remacle C, Cardol P. 2010. Atypical subunit composition of the chlorophycean mitochondrial F_1F_O-ATP synthase and role of Asa7 protein in stability and oligomycin resistance of the enzyme. *Mol Biol Evol* **27:** 1630–1644.

Lindmark DG, Müller M. 1973. Hydrogenosome, a cytoplasmic organelle of the anaerobic flagellate *Tritrichomonas foetus*, and its role in pyruvate metabolism. *J Biol Chem* **248:** 7724–7728.

Mai Z, Ghosh S, Frisardi M, Rosenthal B, Rogers R, Samuelson J. 1999. Hsp60 is targeted to a cryptic mitochondrion-derived organelle ("crypton") in the microaerophilic protozoan parasite *Entamoeba histolytica*. *Mol Cell Biol* **19:** 2198–2205.

Marcotte EM, Xenarios I, van Der Bliek AM, Eisenberg D. 2000. Localizing proteins in the cell from their phylogenetic profiles. *Proc Natl Acad Sci* **97:** 12115–12120.

Margulis L. 1970. *Origin of eukaryotic cells*. Yale University Press, New Haven, CT.

Margulis L. 1981. *Symbiosis in cell evolution*. Freeman, San Francisco.

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

Martin W, Hoffmeister M, Rotte C, Henze K. 2001. An overview of endosymbiotic models for the origins of eukaryotes, their ATP-producing organelles (mitochondria and hydrogenosomes), and their heterotrophic lifestyle. *Biol Chem* **382:** 1521–1539.

Morales J, Mogi T, Mineki S, Takashima E, Mineki R, Hirawake H, Sakamoto K, Ōmura S, Kita K. 2009. Novel mitochondrial complex II isolated from *Trypanosoma cruzi* is composed of 12 peptides including a heterodimeric Ip subunit. *J Biol Chem* **284:** 7255–7263.

Nedelcu AM, Lee RW, Lemieux C, Gray MW, Burger G. 2000. The complete mitochondrial DNA sequence of *Scenedesmus obliquus* reflects an intermediate stage in the evolution of the green algal mitochondrial genome. *Genome Res* **10:** 819–831.

Nina PB, Dudkina NV, Kane LA, van Eyk JE, Boekema EJ, Mather MW, Vaidya AB. 2010. Highly divergent mitochondrial ATP synthase complexes in *Tetrahymena thermophila*. *PLoS Biol* **8:** e1000418.

O'Brien TW. 2002. Evolution of a protein-rich mitochondrial ribosome: Implications for human genetic disease. *Gene* **286:** 73–79.

Oudot M-P, Kloareg B, Loiseaux-de Goër S. 1999. The mitochondrial *Pylaiella littoralis nad11* gene contains only the N-terminal FeS-binding domain. *Gene* **235:** 131–137.

Parisi G, Perales M, Fornasari MS, Colaneri A, Gonzalez-Schain N, Gomez-Casati D, Zimmermann S, Brennicke A, Araya A, Ferry JG, et al. 2004. Gamma carbonic anhydrases in plant mitochondria. *Plant Mol Biol* **55:** 193–207.

Pérez-Brocal V, Clark CG. 2008. Analysis of two genomes from the mitochondrion-like organelle of the intestinal parasite *Blastocystis*: Complete sequences, gene content and genome organization. *Mol Biol Evol* **25:** 2475–2482.

Pérez-Brocal V, Shahar-Golan R, Clark CG. 2010. A linear molecule with two large inverted repeats: The mitochondrial genome of the stramenopile *Proteromonas lacertae*. *Genome Biol Evol* **2:** 257–266.

Pérez-Martínez X, Antaramian A, Vázquez-Acevedo M, Funes S, Tolkunova E, d'Alayer J, Claros MG, Davidson E, King MP, González-Halphen D. 2001. Subunit II of cytochrome *c* oxidase in chlamydomonad algae is a heterodimer encoded by two independent nuclear genes. *J Biol Chem* **276:** 11302–11309.

Pisani D, Cotton JA, McInerney JO. 2007. Supertrees disentangle the chimerical origin of eukaryotic genomes. *Mol Biol Evol* **24:** 1752–1760.

Richly E, Chinnery PF, Leister D. 2003. Evolutionary diversification of mitochondrial proteomes: Implications for human disease. *Trends Genet* **19:** 356–362.

Roger AJ, Silberman JD. 2002. Cell evolution: Mitochondria in hiding. *Nature* **418:** 827–829.

Sapp J. 1994. *Evolution by association. A history of symbiosis*. Oxford University Press, New York.

Schnare MN, Gray MW. 1982. 3′-Terminal sequence of wheat mitochondrial 18S ribosomal RNA: Further evidence of a eubacterial evolutionary origin. *Nucleic Acids Res* **10:** 3921–3932.

Sharma MR, Booth TM, Simpson L, Maslov DA, Agrawal RK. 2009. Structure of a mitochondrial ribosome with minimal RNA. *Proc Natl Acad Sci* **106:** 9637–9642.

Shutt TE, Gray MW. 2006. Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends Genet* **22:** 90–95.

Sicheritz-Pontén T, Kurland CG, Andersson SGE. 1998. A phylogenetic analysis of the cytochrome *b* and cytochrome *c* oxidase I genes supports an origin of mitochondria from within the Rickettsiaceae. *Biochim Biophys Acta* **1365:** 545–551.

Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR. 2012. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biol* **10:** e1001241.

Smith DG, Gawryluk RM, Spencer DF, Pearlman RE, Siu KW, Gray MW. 2007. Exploring the mitochondrial proteome of the ciliate protozoon *Tetrahymena thermophila*: Direct analysis by tandem mass spectrometry. *J Mol Biol* **374:** 837–863.

Smits P, Smeitink JAM, van den Heuvel LP, Huynen MA, Ettema TJG. 2007. Reconstructing the evolution of the mitochondrial ribosomal proteome. *Nucleic Acids Res* **35:** 4686–4703.

Spencer DF, Schnare MN, Gray MW. 1984. Pronounced structural similarities between the small subunit ribosomal RNA genes of wheat mitochondria and *Escherichia coli*. *Proc Natl Acad Sci* **81:** 493–497.

Stechmann A, Hamblin K, Perez-Brocal V, Gaston D, Richmond GS, van der Giezen M, Clark CG, Roger AJ. 2008. Organelles in *Blastocystis* that blur the distinction between mitochondria and hydrogenosomes. *Curr Biol* **18:** 580–585.

Sutak R, Dolezal P, Fiumera HL, Hrdy I, Dancis A, Delgadillo-Correa M, Johnson PJ, Müller M, Tachezy J. 2004. Mitochondrial-type assembly of FeS centers in the hydrogenosomes of the amitochondriate eukaryote *Trichomonas vaginalis*. *Proc Natl Acad Sci* **101:** 10368–10373.

Szklarczyk R, Huynen M. 2010. Mosaic origin of the mitochondrial proteome. *Proteomics* **10:** 4012–4024.

Thao ML, Gullan PJ, Baumann P. 2002. Secondary (γ-Proteobacteria) endosymbionts infect the primary (β-Proteobacteria) endosymbionts of mealybugs multiple times and coevolve with their hosts. *Appl Environ Microbiol* **68:** 3190–3197.

Thrash JC, Boyd A, Huggett MJ, Grote J, Carini P, Yoder RJ, Robbertse B, Spatafora JW, Rappe MS, Giovannoni SJ. 2011. Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade. *Sci Rep* **1:** 13.

Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet* **5:** 123–135.

Tovar J, Fischer A, Clark CG. 1999. The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. *Mol Microbiol* **32:** 1013–1021.

Tovar J, León-Avila G, Sánchez LB, Sutak R, Tachezy J, van Der Giezen M, Hernández M, Müller M, Lucocq JM. 2003. Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. *Nature* **426:** 172–176.

M.W. Gray

Tsaousis AD, Kunji ERS, Goldberg AV, Lucocq JM, Hirt RP, Embley Tm. 2008. A novel route for ATP acquisition by the remnant mitochondria of *Encephalitozoon cuniculi*. *Nature* **453:** 553–557.

Viale AM, Arakaki AK. 1994. The chaperone connection to the origins of the eukaryotic organelles. *FEBS Lett* **341:** 146–151.

Viklund J, Ettema TJG, Andersson SGE. 2012. Independent genome reduction and phylogenetic reclassification of the oceanic SAR11 clade. *Mol Biol Evol* **29:** 599–615.

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

Wallace DC. 1982. Structure and evolution of organelle genomes. *Microbiol Rev* **46:** 208–240.

Waller RF, Keeling PJ. 2006. Alveolate and chlorophycean mitochondrial *cox2* genes split twice independently. *Gene* **383:** 33–37.

Wawrzyniak I, Roussel M, Diogon M, Couloux A, Texier C, Tan KS, Vivarès CP, Delbac F, Wincker P, El Alaoui H. 2008. Complete circular DNA in the mitochondria-like organelles of *Blastocystis hominis*. *Int J Parasitol* **38:** 1377–1382.

Whatley JM, John P, Whatley FR. 1979. From extracellular to intracellular: The establishment of mitochondria and chloroplasts. *Proc R Soc Lond B Biol Sci* **204:** 165–187.

Williams BA, Hirt RP, Lucocq JM, Embley TM. 2002. A mitochondrial remnant in the microsporidian *Trachipleistophora hominis*. *Nature* **418:** 865–869.

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

Yan W, Aebersold R, Raines EW. 2009. Evolution of organelle-associated protein profiling. *J Proteomics* **72:** 4–11.

Yang D, Oyaizu Y, Oyaizu H, Olsen GJ, Woese CR. 1985. Mitochondrial origins. *Proc Natl Acad Sci* **82:** 4443–4447.

Yip C-y, Harbour ME, Jayawardena K, Fearnley IM, Sazanov LA. 2011. Evolution of respiratory complex I. "Supernumerary" subunits are present in the α-proteobacterial enzyme. *J Biol Chem* **286:** 5023–5033.

Ziková A, Panigrahi AK, Dalley RA, Acestor N, Anupama A, Ogata Y, Myler PJ, Stuart K. 2008. *Trypanosoma brucei* mitochondrial ribosomes, affinity purification and component identification by mass spectrometry. *Mol Cell Proteomics* **7:** 1286–1296.

Mitochondrial Evolution

Michael W. Gray

Subject Collection Mitochondria

For additional articles in this collection, see http://cshperspectives.cshlp.org/cgi/collection/