Biochimica et Biophysica Acta 1827 (2013) 224–231

Contents lists available at SciVerse ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbabio

Review

Loss, replacement and gain of proteins at the origin of the mitochondria ☆

Martijn A. Huynen *, Isabel Duarte, Radek Szklarczyk

Centre for Molecular and Biomolecular Informatics, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, P.O. Box 9101, 6400 HB Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Received 2 May 2012
Received in revised form 19 July 2012
Accepted 5 August 2012
Available online 10 August 2012

Keywords:
Alphaproteobacteria
Mitochondrial proteome
Non-orthologous gene displacement
Holliday junction resolvase
Endosymbiosis

ABSTRACT

We review what has been inferred about the changes at the level of the proteome that accompanied the evolution of the mitochondrion from an alphaproteobacterium. We regard these changes from an alphaproteobacterial perspective: which proteins were lost during mitochondrial evolution? And, of the proteins that were lost, which ones have been replaced by other, non-orthologous proteins with a similar function? Combining literature-supported replacements with quantitative analyses of mitochondrial proteomics data we infer that most of the loss and replacements that separate current day mitochondria in mammals from alphaproteobacteria took place before the radiation of the eukaryotes. Recent analyses show that also the acquisition of new proteins to the large protein complexes of the oxidative phosphorylation and the mitochondrial ribosome occurred mainly before the divergence of the eukaryotes. These results indicate a significant number of pivotal evolutionary events between the acquisition of the endosymbiont and the radiation of the eukaryotes and therewith support an early acquisition of mitochondria in eukaryotic evolution. Technically, advancements in the reconstruction of the evolutionary trajectories of loss, replacement and gain of mitochondrial proteins depend on using profile-based homology detection methods for sequence analysis. We highlight the mitochondrial Holliday junction resolvase endonuclease, for which such methods have detected new "family members" and in which function differentiation is accompanied by the loss of catalytic residues for the original enzymatic function and the gain of a protein domain for the new function. This article is part of a Special Issue entitled: The evolutionary aspects of bioenergetic systems.

1. Introduction

Even though the "host's" benefit from the endosymbiosis of an alphaproteobacterium that gave rise to the mitochondrion has been hotly debated [1–3], the fact that this primary endosymbiosis is one of the defining events in the origin of the eukaryotes is generally accepted. Here we review what is known about the change that accompanied this endosymbiosis at the level of the proteins present in that bacterium and early organelle. A number of developments have advanced this knowledge: the elucidation of bacterial and mitochondrial protein function, the characterization of mitochondrial proteomes from multiple species as well as those of specific mitochondrial protein complexes, the sequencing of diverse eukaryotic and alphaproteobacterial genomes and the employment of advanced sequence analysis tools to establish homology and orthology. Capitalizing on these developments, the origin of current day mitochondrial proteins and the fate of proteins of alphaproteobacterial origin in eukaryotes have been analyzed a number of times, both in large-scale studies at various levels of phylogenetic resolution [4–7], in detailed analyses about specific pathways, e.g. about cytochrome biogenesis [8], and in reviews [9,10]. To offer a fresh perspective on the proteome evolution of mitochondria in the evolutionary lineage that currently ends in mammals, we analyze it from an alphaproteobacterial point of view. We ask which proteins were lost in this evolution, and of the proteins that were lost, which have been replaced by non-orthologous proteins with a similar function. We show that most of the losses and a large fraction of the replacements that separate current mammalian mitochondria from their alphaproteobacterial ancestor occurred early in mitochondrial evolution, before the radiation of the eukaryotes. Such a dramatic change in the mitochondrial proteome of the eukaryotic common ancestor has also been observed in recent analyses of the gain of proteins in the large mitochondrial protein complexes of the oxidative phosphorylation [11] and in the mitochondrial ribosome [12]. Nevertheless, also more subtle proteome evolution within individual mitochondrial proteins, is continuously being documented. We discuss the mitochondrial Holliday junction resolvase family that shows evolution at the level of its domain structure and at the level of its catalytic residues, illustrating the loss and replacement themes of this review.

1.1. Protein loss

In the research on mitochondrial evolution, there has been much emphasis on systems that are derived from the alphaproteobacterial ancestor, like oxidative phosphorylation and translation, and on new systems that were invented or modified using existing proteins, like

☆ This article is part of a Special Issue entitled: The evolutionary aspects of bioenergetic systems.

* Corresponding author.
E-mail addresses: huynen@cmbi.ru.nl (M.A. Huynen), i.duarte@cmbi.ru.nl (I. Duarte), radek@cmbi.ru.nl (R. Szklarczyk).

0005-2728/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbabio.2012.08.001

protein import [13]. There are, however, also bacterial systems that have been lost. A specific example that was noted with the first sequenced nuclear genome of a eukaryote is the bacterial-type sec machinery that is missing from *Saccharomyces cerevisiae* [14], while other lost systems have been replaced (see below). To obtain a quantitative and comprehensive overview of the proteins that were lost, we need to know the set of proteins that were present in the alphaproteobacterial ancestor. However, a close relative of the alphaproteobacterium that gave rise to the mitochondria is not expected to exist anymore, after the (at least) 1.5 billion years of evolution that separates eukaryotes from bacteria [15]. Furthermore, although mitochondria have often been placed close to the Rickettsiales [16], recent analyses suggest that mitochondria are derived from an ocean dwelling clade from which no complete genomes are yet available [17]. To obtain a set of proteins that were likely present in the free living alphaproteobacterial ancestor of the mitochondria, one can examine which genes are widely distributed among current alphaproteobacteria [18]. We used the COG database as represented in eggNOG [19] to unify genes from the various species into orthologous groups (COGs). We included the COGs that are present in 95% of the 117 alphaproteobacterial genomes in eggNOG version 3.0 as a likely minimal ancestral set. With the exception of 11 genes (that were manually added), this set contains all the protein coding genes encoded in the *Reclinomonas americana* mitochondrial genome, which, in turn, contains all mitochondrial genes of alphaproteobacterial origin [20]. The set of universal alphaproteobacterial genes (Supplementary Table 1) contains in total 370 COGs. Although such a set is biased toward genes involved in information processing and the synthesis of nucleotides and cofactors [18], it does serve as a starting point to examine the proteins lost in the evolution of mitochondria.

Similar to the situation with the alphaproteobacteria, we do not have an ancestral mitochondrial proteome and have to reconstruct it from current day mitochondria. We collected the set of proteins that have been observed in large-scale studies on mitochondrial proteins from mammals [21,22], fungi [23], land plants [24] and ciliates [25], or that are annotated as such in more detailed experiments [26]. We assume that proteins that have not been observed in mitochondria in any of the current mitochondria and that are present in 95% of the current alphaproteobacteria, have likely been lost from mitochondria before the radiation of the eukaryotes. Although we cannot exclude that some of the universal alphaproteobacterial proteins that are absent from eukaryotes arose in the alphaproteobacteria after the divergence of the mitochondria, and were therefore never lost from eukaryotes, such a scenario appears unlikely given that published phylogenies place the mitochondria within the alphaproteobacteria and not at their root, e.g. [16,17,27]. Further reconstruction of loss events within the eukaryotes depends on the branching order of the eukaryotes. This branching order has not been completely resolved [28] and we therefore assume a star-like phylogeny in which the alveolates (including ciliates), the archaeplastidae (including land plants) and opisthokonts (fungi + metazoa) diverged at the same point in time, and in which fungi and the metazoa diverged later from each other. This allows us to recognize, in the evolution of the mitochondrial proteome from an ancestral bacterium to its current state in mammals, four stages: 1) proteins that were present in the bacterial ancestor (reconstructed from alphaproteobacterial genomes and the mitochondrial genome of *R. americana*); 2) proteins from the ancestor present in the last eukaryotic common ancestor (LECA); 3) proteins from the ancestor present in the opisthokonts and 4) proteins from the ancestor present in mammals. Orthologous groups were sorted into functional classes using the COG classification (Fig. 1).

Two elements stand out from this analysis: 1) the loss of alphaproteobacterial proteins from the mitochondrion before the radiation of the eukaryotes has been much more dramatic than subsequent losses along the evolutionary lineage to the mammalian mitochondria. Of the 370 COGS that likely were present in the bacterial ancestor, 161 have not been detected in current day mammalian mitochondria. From these 161, ~71% (115 proteins) have not been detected in any mitochondria, and were thus likely lost from the organelle before the divergence of eukaryotes. 2) The loss of proteins is not evenly distributed among the function classes and displays a rather eclectic conservation of pathways involved in (energy) metabolism and translation, at the expense of replication, repair and recombination of DNA, transcription, secretion, the cell wall, the cell cycle and chromosome segregation. This general trend of protein loss is similar to what was observed by comparing the genes in a single alphaproteobacterial genome (*Caulobacter crescentus*) to the proteins known to be present in mitochondria of mammals and of *S. cerevisiae* [29]. Some of these observations are, in hindsight, not unexpected. The inter membrane space of mitochondria, although derived from the periplasm of bacteria, is quite different from it, e.g. by lacking a cell wall. Indeed the proteins involved in peptidoglycan synthesis, the lipoproteins and the proteins involved in modifications of those lipoproteins have all been lost (Supplementary Table 1). Even a rare periplasmic "holdout" like LACTB that is involved in peptidoglycan metabolism in bacteria, appears to have obtained a different function in the intermembrane space of mammalian mitochondria [30]. The loss of genes for cell envelope biogenesis has also been observed in genomes of endosymbiotic bacteria, see ref [31] for a review. Other parallels with the evolution of endosymbiotic bacteria are loss of transcription regulation, DNA repair and recombination [31]. Not all proteins that were lost from the mitochondrion were lost from the eukaryotic cell completely. A number of the enzymes that were lost from the mitochondrion have, concomitant with the relocation of their genes to the nucleus, actually been targeted to other parts of the eukaryotic cell. A classic case of relocalization is the heme biosynthesis pathway that only partly takes place in mitochondria [32], but also other enzymes like the ones involved in fatty acid elongation are now located in the eukaryotic cell but outside the mitochondria [7].

1.2. Non-orthologous protein replacement

A number of alphaproteobacterial proteins that have been lost, have actually been replaced by other proteins, in a variation of the "non-orthologous gene displacement" that was originally analyzed on a large scale when comparing bacterial genomes [33]. Well known examples are the RNA polymerase that has been replaced with a T3/T7-phage like polymerase [34], the DNA polymerase that has been replaced with a T3/T7-phage DNA polymerase [35], the DsbA/DsbB disulfide relay system in the intermembrane space of alphaproteobacteria that has been replaced with the MIA40/Erv1 system [36], and the recently discovered replacement of the TAT export system with the AAA-ATPase Bcs1 for the export of the folded Rieske protein [37]. Other instances are more subtle, like the replacement of the Glycyl-tRNA synthetase subunits of alphaproteobacterial origin by their homologs of archaeal origin [38]. The exact order of replacement events cannot always confidently be retraced. An example is the replacement in trypanosomatids of the mitochondrial protein translocase of the outer membrane – TOM40 – with a protein of bacterial origin that might actually represent an ancestral state of the mitochondrion [39]. Proposed horizontal gene transfer of genes, like the heme lyase [8], complicates the situation even further. Furthermore, calling a protein replacement in itself is not unequivocal: how similar do the functions of two proteins have to be, to be called a replacement? Can we call the beta subunit of DNA polymerase gamma that is involved in the polymerase's processivity a replacement for the sliding clamp subunit of the bacterial DNA polymerase [40]? To obtain an overview of which bacterial systems were replaced (Table 1) we combined two approaches. First we compared Enzyme Commission numbers (EC number) of proteins in the ancestral set and in current day mitochondria with each other to identify pairs of non-orthologous proteins with the same EC number. Second we performed literature searches for proteins from the ancestral set that were lost, but whose functions can still be found in mitochondria. We then reconstructed the origin of a mitochondrial protein whose function is equivalent to that of a protein that was lost by examining its published phylogenetic

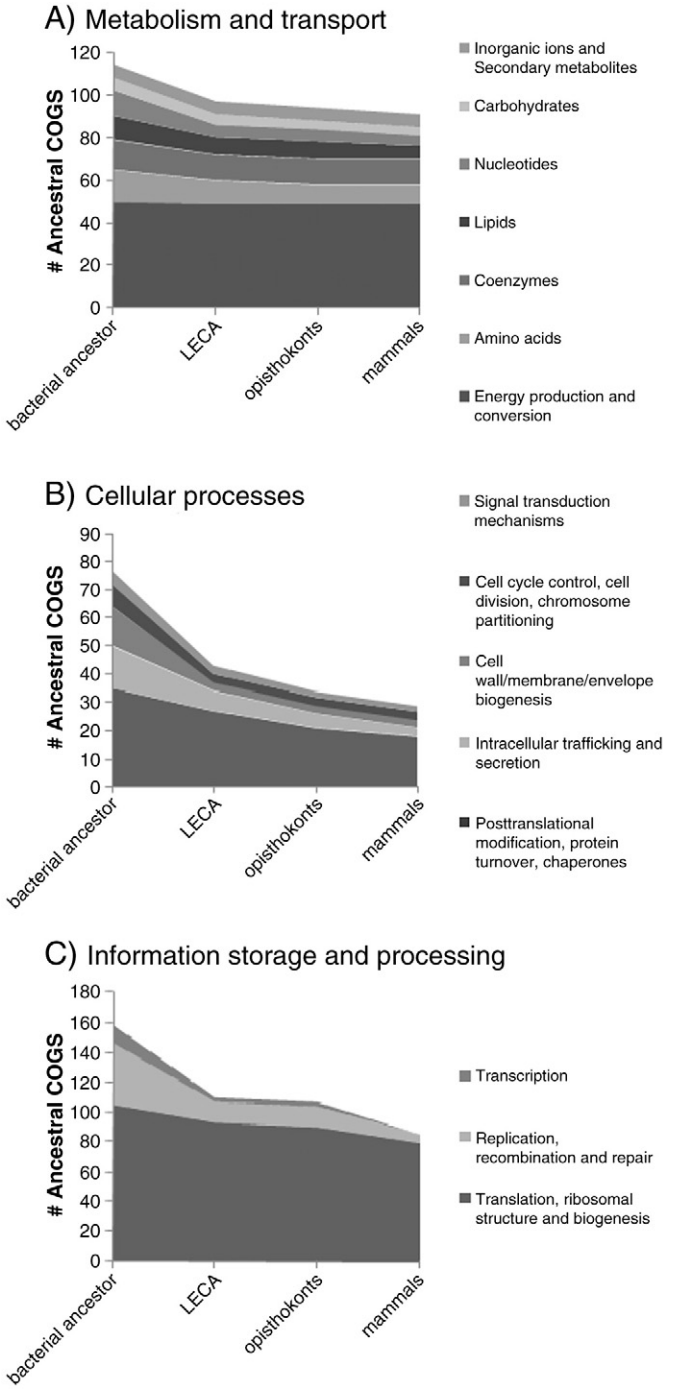

Fig. 1. Reconstruction of the loss of proteins from the alphaproteobacterial ancestor of mitochondria to their current representatives in mammals. The vertical axis shows the minimal number of different orthologous groups of alphaproteobacterial ancestry that have been estimated to be present in the bacterial ancestor of the mitochondria, in the last eukaryotic common ancestor (LECA), in the ancestor of the metazoa and the fungi (opisthokonts) and in current mammalian mitochondria. A) Loss of proteins involved in metabolic processes and the transport of metabolites. B) Loss of proteins involved in cellular processes. C) Loss of proteins involved in information storage and processing. Function classes are based on the COG function classification system [95]. See Supplementary Table 1 for the full list of proteins, their function classification and when they were lost in evolution.

distribution, or, when the latter was not available, by performing phylogenetic analyses ourselves. It should thereby be noted that the gain of the new (replacing) protein does not always coincide with the loss of the old one. Orthologs of the sliding clamp subunit of the bacterial DNA polymerase have not been observed in mitochondria, but the "replacing" beta subunit of DNA polymerase gamma is restricted to metazoa [41], leaving e.g. the fungi without a (known) processivity factor. Such a gap is also visible for the mitochondrial transcription elongation factor TEFM that originated in metazoa [42], replacing a transcription elongation factor that was lost since LECA. The result is that, although eight out of seventeen replacing proteins have originated before LECA (Table 1), the non-orthologous protein replacements are not as strongly biased toward early eukaryotic evolution as is the loss of proteins.

1.3. What drives protein replacement?

A number of alternative explanations have been offered for protein replacement. One explanation is basically an extension of the hydrophobicity hypothesis, originally formulated by von Heijne [43], and most recently advocated by de Grey [44]. The argument is that the replacement of some systems is caused by, on the one hand the pressure to move mitochondrial genes to the nucleus, e.g. to escape the high mutation rate in the mitochondrial genome, and on the other hand, the problematic transport of their integral membrane proteins back into mitochondria. The strongest support for that comes actually from cases where the replacement has not occurred in all eukaryotes, like bacterial cytochrome system 1 that in eukaryotes has partially been replaced with system III [8]. In those species where it has not been replaced, its transmembrane proteins are still encoded in the mitochondrial genome [8]. Similarly, the genes for mitochondrial transmembrane proteins tatA and tatC of the TAT system that was replaced by the Bcs1 system in *S. cerevisiae* [37], are still encoded in the mitochondrial genomes of a wide variety of other species, including land plants, stramenopiles, the closest single cell relative of the metazoa *Monosiga brevicolis* [45] and early branching metazoa like the sponge *Oscarella malakhovi* [46], but not in nuclear genomes. A hydrophobicity explanation also applies to the DsbA/DsbB disulfide relay, of which DsbB is an integral membrane protein. Nevertheless, for the latter system also other explanations have been formulated. These include that MiA40/Erv1 has to work in a more reducing environment than the DsbA/DsbB disulfide relay and that it has a simpler function than DsbA/DsbB [47]. Relative simplicity of the system has also been advocated as an explanation for the phage-like DNA polymerase, DNA helicase and RNA polymerase. If those genes are of prophage origin, the fact that the systems are monomeric might have favored the genes' translocation from the endosymbiont to the nucleus over that of the multimeric bacterial systems [48]. Finally, for cases where the mitochondria share functions with the other parts of the cell, just that some functions were already encoded in the nuclear genome, like amino-acyl tRNA synthetases, or nucleic acid interacting proteins like a DNA ligase or a ribonuclease may have made the version of the gene that is of organellar ancestry superfluous, leading to its loss and replacement, e.g. by dual targeting of the nuclear encoded protein [49] or by gene duplication [50].

1.4. Protein gain in the large complexes

Most mitochondrial protein complexes of alphaproteobacterial origin have expanded with so-called supernumerary subunits. As with mitochondrial proteomes, we can reconstruct likely ancestral states of mitochondrial complexes by comparing current day complexes from multiple species with each other. In the absence of proteomics data of the complexes studied they can also be compared with mitochondrial proteomes and with genomes of eukaryotic species. The main assumption behind such an analysis is that when a protein is part of a (mitochondrial) complex in a species, its orthologs are also part of that complex in other eukaryotic species. There are exceptions to this pattern, like the acyl carrier protein, a fraction of which, in fungi and vertebrates, is associated with complex I but that is not part of complex I in plants [51]. Nevertheless, there is ample support for the conservation of complex-membership between orthologs in large-scale systematic analyses of complex membership between species in general [52], and in comparisons of mitochondrial complexes between species in particular

Table 1
Non-orthologous protein replacement between the alphaproteobacterial ancestor and current mammalian mitochondria. The third column describes the function of a universal alphaproteobacterial protein that has been replaced by a non-orthologous, mitochondrial protein. The rightmost column indicates when the new protein has been inferred to appear in evolution. The reference is, when possible, to experimental data relevant to the phylogenetic distribution of that new protein and therefore its origin, otherwise to phylogenetic analyses supporting that origin. In the absence of a reference, the origin was traced using homology searches followed by phylogenetic analyses. For smpB, a member of the tmRNA system, a candidate replacement has been hypothesized based on 3D modeling but has not yet been experimentally validated [96].

COG	Bacterial gene name	Function	Human gene	EC number	Origin of the new protein
COG1138	ccmE	Cytochrome C biogenesis	HCCS	–	LECA [8]
COG0592	dnaN	DNA polymerase processivity factor	POLG2	–	Metazoa [41]
COG0587	dnaE	DNA polymerase	POLG	2.7.7.7	Opisthokonts [78]
COG0847	dnaQ	3–5 exonuclease	POLG	3.1.13	Opisthokonts [78]
COG1651	dsbA	Protein disulfide isomerase	CHCHD4	5.3.4.1	LECA [103]
COG0492	trxB	Thioredoxin reductase	TXNRD2	1.8.1.9	Metazoa
COG0358	dnaG	DNA primase	POLRMT	2.7.7.-	LECA [104]
COG0751	glyS	Glycyl-tRNA aminoacyltransferase	GARS	6.1.1.14	LECA [17]
COG0752	glyQ				
COG0805	tatC	Protein translocation	BCS1L	–	LECA [105]
COG1826	tatA				
COG0272	ligA	DNA ligase	LIG3	6.5.1.2	Filozoa [106]
COG0202	rpoA	RNA polymerase	POLRMT	2.7.7.6	LECA [104]
COG0164	rnhB	Ribonuclease	RNASEH1	3.1.26.4	Metazoa [107]
COG1158	Rho	Transcription termination	MTERF	–	LECA [108]
COG0782	greA	Transcription elongation	TEFM	–	Metazoa [42]
COG0691	smpB	Recycling stalled ribosomes	mtRF1?	–	Vertebrates [96]
COG0625	Gst	Glutathione-S-transferase	GSTK1	2.5.1.18	Metazoa [109]
COG0305	dnaB	Replicative helicase	TWINKLE	3.6.1.-	LECA [78]

[11,53]. One theme that has become increasingly apparent in recent analyses of the evolution of mitochondrial complexes, is that most of the addition of new proteins to complexes of alphaproteobacterial origin that are currently present in mammals has happened very early in mitochondrial evolution, as these subunits appear to be shared by multiple eukaryotic crown groups. Specifically, studies of complex V [54], of complex I [11,55,56] and of the mitochondrial ribosome [12,57], have "pushed back" the origin of an increasing number of supernumerary subunits and assembly factors before LECA. Some of the supernumerary subunits and assembly factors of complex I have an even earlier origin, as they are encoded in the bacterial genomes [55] and three of the supernumerary subunits of complex I have been shown to be actually part of complex I in the alphaproteobacterium *Paracoccus denitrificans* [58]. To examine whether the addition of new eukaryotic subunits to existing complexes before LECA is a more general trend, we complemented the published analyses on complexes I, V and the mitochondrial ribosome with a tracing of the origin of supernumerary subunits from complex III and complex IV, as well as the assembly factors of complex IV and complex V, and quantified the combined results (Fig. 2). As such analyses do not uniquely rely on proteomics data but also on sequenced genomes, they allow for a higher resolution in reconstructing when in evolution certain proteins of a complex appeared than the protein loss reconstructions in Fig. 1. Consistent with the observations from the literature with respect to mitochondrial protein complex evolution, most of the supernumerary subunits that have been added to the current mammalian mitochondrial protein complexes, have been added before LECA. Furthermore, also the new, eukaryotic proteins that are involved in the assembly of the complexes were already present in LECA. Note however, that this pattern does not necessarily hold for protein complexes in evolutionary lineages other than the one leading to mammals. Specifically in trypanosomatids many new proteins with no detectable homology outside of the trypanosomatids, were found attached to the mitochondrial ribosome [59,60] and to complex V [61], while the latter complex also appears to have an atypical composition in *Chlamydomonas reinhardtii* [54].

1.5. Technical aspects of comparative genome analysis

Technically, the newly discovered homology between supernumerary subunits from the same complex in multiple species and their therewith inferred early evolutionary origin depend on: 1) a better sampling of mitochondrial protein complexes and eukaryotic genomes, allowing the detection of homologs in distantly related taxa and therewith inference of their early presence in eukaryotic evolution. An example of this is the presence of a carbonic anhydrase in complex I of the amoeboid protozoon *Acanthamoeba castellanii* that was subsequently found to be encoded in a wide range of unicellular eukaryotes [62], placing its origin before LECA rather than in the taxonomic branch leading to land plants and algae [63]. 2) The usage of sensitive, profile-to-sequence and profile-to-profile based tools [64] for finding homology relationships. An example being the resolution of the homology relationship between the 9.5-kDa protein in fungi and the B9 subunit in mammals [11], for which the literature had been equivocal before [55,63,65]. Another example is the discovery of the ATPase assembly factor TMEM70 in single celled eukaryotes [66]. 3) The construction of sequence profiles using an increasing number of

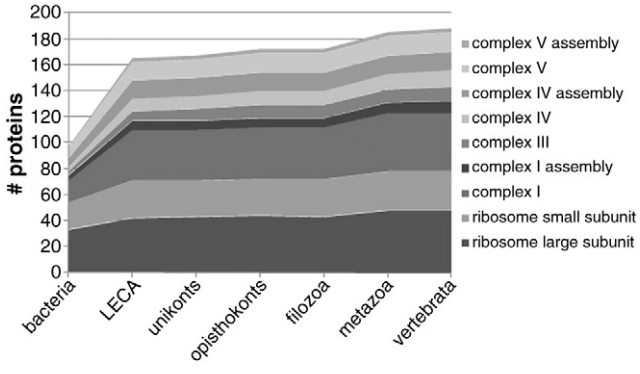

Fig. 2. Expansion of the number of proteins within the large mitochondrial complexes, from the bacterial ancestor of mitochondria to their current composition in mammals. Data are based on proteins in the mitochondrial protein complexes, or when unavailable for specific taxa, on the presence of orthologs in genomes. The data for the mitochondrial ribosome are from [12]. Data for complex I are from [11], and for complex I assembly from [56]. Data for complex IV assembly are from [71]. Data for complex V are based on [54] to which the assembly factors ATP11, ATP12 and TMEM70 were added. For complex III and complex IV the origin of the subunits was established using profile-based homology detection, and where necessary, phylogenetic analyses to establish where a gene duplication occurred. See Table S2 for an overview and for exceptions to the published reconstructions.

sequences from closely related species to obtain profiles that are more representative of the sequence family and therefore allow better detection of homology relationships between those profiles and those from other taxa. The latter has e.g. led to an improvement of the annotation of mitochondrial genomes of ciliates, which have many hypothetical open reading frames. Some of these have recently been annotated as coding for ribosomal proteins [67,68] largely because more ciliate mitochondrial genomes were available to create the sequence profiles. Nevertheless, even with those advances, not all homology relationships within the oxidative phosphorylation complexes can unequivocally be established. A recently determined, mitochondrially encoded ATPase subunit from *Tetrahymena thermophila* [69] is, using a profile–profile homology detection approach [64] not significantly similar to any ATPase subunit outside of the alveolates (data not shown). In a similar vein, a recent evaluation of complex I evolution discussed possible homology relationships between fungal and metazoan subunits that were not statistically significant [11]. Nevertheless, based on manual inspection of the alignments and on the fact that both subunits are member of the same complex, these homology relationships were deemed likely to be true [11]. In such cases one implicitly uses the argument that the E-value that is reported is too high (not significant enough) because it is based on the comparison with a database with e.g. all human proteins, while it should be based on a database that only contains proteins from the specific complex compared between the species. Systematic analyses have shown that such implicit reasoning does indeed detect "true" homologs when comparing protein complexes between species and including insignificant E-values [70]. Even when including only significant E-values, the gain in orthology relations that can be discovered when using profile-based homology detection tools is considerable. A recent comparison of mitochondrial proteomes from fungi and mammals showed that by using profile based methods a 20% percent increase in the number of orthology relationships could be obtained relative to only using pairwise sequence comparisons [71]. Most of the "newly discovered" relationships concern short, rapidly evolving proteins like assembly factors. Out of eleven thus predicted new human complex IV assembly factors, five could be validated by mitochondrial localization and copurification with other Cox (assembly) proteins [71]. These results show that the increase in sensitivity of profile-based homology detection does not come at a price of a decrease in reliability, specifically when it is combined with a best-bidirectional hit criterion as a proxy for orthology. As a case in point, the recently identified human complex IV assembly factor C2orf64 was shown to be orthologous to the fungal assembly factor PET191 [72] while the assembly factor C12orf62 [73] was shown to be orthologous to *S. cerevisiae* Cox14 [71].

1.6. Loss of critical residues and variations in domain composition

Evolutionary analyses based on the presence/absence of genes of an orthologous group miss more subtle changes in the sequences or domain compositions of the proteins themselves. Although such changes tend to be dominated by neutral amino-acid substitutions, they can also be functionally relevant. Examples range from the loss of parts of the α and β subunits of the Phenylalanine tRNA synthetase and fusing of the remaining proteins into a single, monomeric Phenylalanyl-tRNA synthetase [74] that has lost its editing activity, leading to a stable mischarging of $tRNA^{Phe}$ with tyrosine [75]; the loss of the DUF59 domain in the evolution of the FeS complex I assembly factor IND1 from a minD/MRP family protein ancestor in bacteria [76]; to the loss of three helices of complex I subunit ND2 [77] and the degeneration of the primase domain in Twinkle in the metazoa [78]. Loss of a catalytic residue has been shown to be correlated with the loss of catalytic activity in the protein peptide deformylase, HsPDF [79]. Based on this observation the protein was called an evolutionary remnant [79], nevertheless the protein's presence appears to be required for the accumulation of mitochondrial DNA-encoded proteins [80].

The mitochondrial Holliday junction resolvase (HJR) endonuclease family illustrates how the loss of a function on the one hand and the gain of a new function on the other hand correlate with the loss of a residue known to be critical for the first function and the gain of a domain known to be involved in the second function. A mitochondrial HJR endonuclease subunit was first discovered in *S. cerevisiae* [81] and *Schizosaccharomyces pombe* [82]. The similarity of these proteins with their bacterial counterpart (RuvC) was too low to detect the homology by pairwise sequence comparison. Only by profile-based analyses was this homology relationship established [83], which was subsequently confirmed by 3D structure comparison [84]. Recently, C17orf42, a human homolog of the fungal HJRs, was detected using profile-based sequence comparison methods [42]. Nevertheless, in this protein the aspartate at position 8 (numbering based on the positions in *Escherichia coli*'s RuvC) that was shown to be necessary for the catalytic activity of RuvC, but not for Holliday junction binding [85], has been mutated to a valine. The situation is similar to MRS1, a paralog of CCE1 that arose by gene duplication in the fungi. MRS1 was shown to be involved in splicing and in it the same D8 residue is mutated [86]. Experimental analysis of C17orf42 failed to detect endonuclease activity, and rather the protein appeared involved in mitochondrial transcription elongation [42]. Furthermore, instead of an N-terminal SAP domain that is present in the fungal Holliday junction resolvases, the metazoan members of the family have an N-terminal Helix–hairpin–Helix motif. Holliday junction resolvases have an RNase H fold, and the combination of a Helix–hairpin–Helix motif with an RNaseH fold has also been found in the transcription elongation factor Spt6 [87], suggesting convergent evolution of domain composition to obtain a transcription elongation function in mitochondria. We further analyzed the distribution of this protein family among eukaryotes and also uncovered homologs in plants and algae (Fig. 3). Interestingly, the plant proteins contain a mitochondrial targeting signal, and the algae contain a SAP domain, like their fungal homologs. Furthermore, also in plants, a gene duplication has been followed by the loss of one of the catalytic residues. Finally, degeneration of critical residues in the arthropod sequences has even progressed further than in mammals, with also the C-terminal D139 and D142 having been mutated.

2. Concluding remarks

Most of the loss of proteins of bacterial ancestry, a large fraction of the replacement of alphaproteobacterial systems with proteins of a different ancestry, and most of the increase of complexity of mitochondrial protein complexes of bacterial ancestry appear to have occurred early in eukaryotic evolution, before the eukaryotic radiation and concomitantly to the translocation of mitochondrial genes to the nucleus. Why there has been such an increase in the complexity of some of the mitochondrial protein machineries has been widely debated ever since their discovery, e.g. [88] and recently again for complex I [89]. Supernumerary subunits have been proposed to have roles in regulation, protection against ROS, assembly and stability. While the reasons vary from complex to complex and from protein to protein, for complex I there are specific data pertaining to a role of the supernumerary subunits in increasing the stability of the complex [90]. Such an increase of the complexity to increase stability would not be incompatible with a constructive neutral evolution explanation [91]. In that scenario, the addition of new, nuclear encoded subunits would compensate the accumulation of mutations in the older, mitochondrial encoded, subunits. It would explain why complex II, the only completely nuclear encoded oxidative phosphorylation complex, is the only complex without supernumerary subunits. The argument is analogous to the increase of the number of nuclear encoded proteins in the mitochondrial ribosome to compensate for the loss of mitochondrial encoded rRNA [92], although there is little support for the latter argument [93]. Nevertheless, in such a scenario one might expect a gradual increase in the number of subunits per complex, specifically in taxa with small effective population sizes like the metazoa.

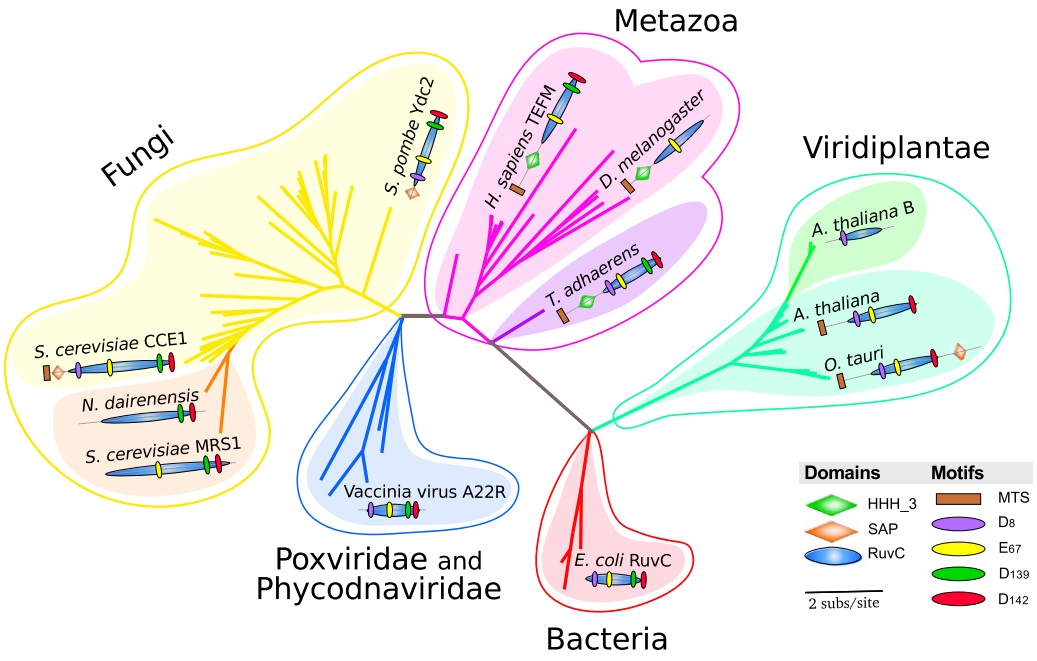

Fig. 3. The evolution of the Holliday junction resolvase endonuclease illustrates the variation in the domain composition and presence of catalytic residues in a single mitochondrial protein family. An original bacterial Holliday junction resolvase is still functional in fungal mitochondria and in eukaryotic viruses like the poxviridae [97]. In those taxa it has retained all the residues known to be critical for its catalytic activity, while in fungi it has gained the DNA/RNA binding SAP domain. Other eukaryotic variants have lost critical residues, have gained or lost domains and have acquired new functions: MRS1 is involved in mitochondrial splicing in *S. cerevisiae*, and it has lost both a catalytic aspartate (D8) and the SAP domain. Transcription elongation factor of mitochondria (TEFM) has also lost this aspartate, has acquired a Helix–hairpin–Helix motif and is involved in transcription elongation. The insects present an even more derived RuvC domain, having lost three of the four catalytic residues. Representatives of the family in plants and algae have also lost one or more catalytic residues and, in one case, have also acquired a SAP domain. Holliday junction resolvase homologous sequences were retrieved using *E. coli*'s RuvC protein sequence (GI: 15802276) as query seed for a PSI-BLAST [98]. The results were manually curated in order to remove redundant sequences and guarantee a broad phylogenetic coverage among the eukaryotes. The dataset was aligned with ClustalW [99] and visually inspected with Jalview [100] to assure the correct alignment of all functionally relevant domains and residues. A Maximum Likelihood phylogeny was computed for the final alignment containing 82 sequences with 970 aligned positions, using RAxML [101]. The evolutionary model, chosen according to the Akaike Information Criterion (AIC) as implemented in ProtTest [102], was an LG plus the empirical frequency of amino acids and a 4 discrete rate categories Gamma distribution (LG + F + 4G). In order to obtain the domain composition of each main phylogenetic branch, a Pfam A domain search was conducted for each sequence using HHPred [64].

Instead, the large majority of additions appear to have occurred early in eukaryotic evolution, suggesting that their addition is not the result of constructive neutrality.

Our quantitative analyses and recapitulations of existing literature support, at least at the level of the proteins present in the mitochondria, an already very advanced last ancestor of the eukaryotes. This pattern is not unique to mitochondria. Also the eukaryotic endomembrane system, of which the evolution can be reconstructed via duplications in key proteins in membrane trafficking, appears to have been in place by the time of the eukaryotic radiation [94]. One may expect that the pre-LECA evolution of the mitochondrial proteome would require a significant amount of time, specifically for the gain and replacement of proteins. This provides quantitative support for the thesis that mitochondria have been added early in eukaryotic evolution and therefore likely to have played a crucial role in its origin rather than being added to an existing eukaryotic cell.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbabio.2012.08.001.

Acknowledgements

We would like to thank the two referees for their comments. RS is supported by CSBR (Centres for Systems Biology Research) from the Netherlands Organisation for Scientific Research (NWO) (CSBR09/013V), ID by the Portuguese Foundation for Science and Technology—FCT (SFRH/BD/32959/2006) and by "BolsasRui Tavares 2010."

References

[1] W. Martin, M. Muller, The hydrogen hypothesis for the first eukaryote, Nature 392 (1998) 37–41.
[2] D. Moreira, P. Lopez-Garcia, Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis, J. Mol. Evol. 47 (1998) 517–530.
[3] S.G. Andersson, C.G. Kurland, Origins of mitochondria and hydrogenosomes, Curr. Opin. Microbiol. 2 (1999) 535–541.
[4] E.M. Marcotte, I. Xenarios, A.M. van Der Bliek, D. Eisenberg, Localizing proteins in the cell from their phylogenetic profiles, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 12115–12120.
[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 (2004) 1643–1660.
[6] C.G. Kurland, S.G. Andersson, Origin and evolution of the mitochondrial proteome, Microbiol. Mol. Biol. Rev. 64 (2000) 786–820.
[7] T. Gabaldon, M.A. Huynen, Reconstruction of the proto-mitochondrial metabolism, Science 301 (2003) 609.
[8] J.W. Allen, A.P. Jackson, D.J. Rigden, A.C. Willis, S.J. Ferguson, M.L. Ginger, Order within a mosaic distribution of mitochondrial c-type cytochrome biogenesis systems? FEBS J. 275 (2008) 2385–2402.
[9] S.G. Andersson, O. Karlberg, B. Canback, C.G. Kurland, On the origin of mitochondria: a genomics perspective, Philos. Trans. R. Soc. Lond. B Biol. Sci. 358 (2003) 165–177 (discussion 177–9).
[10] R. Szklarczyk, M.A. Huynen, Mosaic origin of the mitochondrial proteome, Proteomics 10 (2010) 4012–4024.
[11] P. Cardol, Mitochondrial NADH:ubiquinone oxidoreductase (complex I) in eukaryotes: a highly conserved subunit composition highlighted by mining of protein databases, Biochim. Biophys. Acta 1807 (2011) 1390–1397.
[12] E. Desmond, C. Brochier-Armanet, P. Forterre, S. Gribaldo, On the last common ancestor and early evolution of eukaryotes: reconstructing the history of mitochondrial ribosomes, Res. Microbiol. 162 (2011) 53–70.
[13] V. Hewitt, F. Alcock, T. Lithgow, Minor modifications and major adaptations: the evolution of molecular machines driving mitochondrial protein import, Biochim. Biophys. Acta 1808 (2011) 947–954.
[14] B.S. Glick, G. Von Heijne, *Saccharomyces cerevisiae* mitochondria lack a bacterial-type sec machinery, Protein Sci. 5 (1996) 2651–2652.
[15] N.J. Butterfield, A.H. Knoll, K. Swett, A bangiophyte red alga from the Proterozoic of arctic Canada, Science 250 (1990) 104–107.
[16] K.P. Williams, B.W. Sobral, A.W. Dickerman, A robust species tree for the alphaproteobacteria, J. Bacteriol. 189 (2007) 4578–4586.

[17] B. Brindefalk, T.J. Ettema, J. Viklund, M. Thollesson, S.G. Andersson, A phylometagenomic exploration of oceanic alphaproteobacteria reveals mitochondrial relatives unrelated to the SAR11 clade, PLoS One 6 (2011) e24457.
[18] B. Boussau, E.O. Karlberg, A.C. Frank, B.A. Legault, S.G. Andersson, Computational inference of scenarios for alpha-proteobacterial genome evolution, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 9722–9727.
[19] S. Powell, D. Szklarczyk, K. Trachana, A. Roth, M. Kuhn, J. Muller, R. Arnold, T. Rattei, I. Letunic, T. Doerks, L.J. Jensen, C. von Mering, P. Bork, eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges, Nucleic Acids Res. 40 (2012) D284–D289.
[20] M.W. Gray, G. Burger, B.F. Lang, Mitochondrial evolution, Science 283 (1999) 1476–1481.
[21] D.J. Pagliarini, S.E. Calvo, B. Chang, S.A. Sheth, S.B. Vafai, S.E. Ong, G.A. Walford, C. Sugiana, A. Boneh, W.K. Chen, D.E. Hill, M. Vidal, J.G. Evans, D.R. Thorburn, S.A. Carr, V.K. Mootha, A mitochondrial protein compendium elucidates complex I disease biology, Cell 134 (2008) 112–123.
[22] V.K. Mootha, J. Bunkenborg, J.V. Olsen, M. Hjerrild, J.R. Wisniewski, E. Stahl, M.S. Bolouri, H.N. Ray, S. Sihag, M. Kamal, N. Patterson, E.S. Lander, M. Mann, Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria, Cell 115 (2003) 629–640.
[23] J. Reinders, R.P. Zahedi, N. Pfanner, C. Meisinger, A. Sickmann, Toward the complete yeast mitochondrial proteome: multidimensional separation techniques for mitochondrial proteomics, J. Proteome Res. 5 (2006) 1543–1554.
[24] J.L. Heazlewood, J.S. Tonti-Filippini, A.M. Gout, D.A. Day, J. Whelan, A.H. Millar, Experimental analysis of the *Arabidopsis* mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins, Plant Cell 16 (2004) 241–256.
[25] D.G. Smith, R.M. Gawryluk, D.F. Spencer, R.E. Pearlman, K.W. Siu, M.W. Gray, Exploring the mitochondrial proteome of the ciliate protozoon *Tetrahymena thermophila*: direct analysis by tandem mass spectrometry, J. Mol. Biol. 374 (2007) 837–863.
[26] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, Nat. Genet. 25 (2000) 25–29.
[27] N. Rodriguez-Ezpeleta, T.M. Embley, The SAR11 group of alpha-proteobacteria is not related to the origin of mitochondria, PLoS One 7 (2012) e30520.
[28] V. Hampl, L. Hug, J.W. Leigh, J.B. Dacks, B.F. Lang, A.G. Simpson, A.J. Roger, Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups", Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 3859–3864.
[29] T. Gabaldon, M.A. Huynen, From endosymbiont to host-controlled organelle: the hijacking of mitochondrial protein synthesis and metabolism, PLoS Comput. Biol. 3 (2007) e219.
[30] Z. Polianskyte, N. Peitsaro, A. Dapkunas, J. Liobikas, R. Soliymani, M. Lalowski, O. Speer, J. Seitsonen, S. Butcher, G.M. Cereghetti, M.D. Linder, M. Merckel, J. Thompson, O. Eriksson, LACTB is a filament-forming protein localized in mitochondria, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 18960–18965.
[31] J.P. McCutcheon, N.A. Moran, Extreme genome reduction in symbiotic bacteria, Nat. Rev. Microbiol. 10 (2011) 13–26.
[32] S. Sano, S. Inoue, Y. Tanabe, C. Sumiya, S. Koike, Significance of mitochondria for porphyrin and heme biosynthesis, Science 129 (1959) 275–276.
[33] E.V. Koonin, A.R. Mushegian, P. Bork, Non-orthologous gene displacement, Trends Genet. 12 (1996) 334–336.
[34] B.S. Masters, L.L. Stohl, D.A. Clayton, Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7, Cell 51 (1987) 89–99.
[35] J. Filee, P. Forterre, T. Sen-Lin, J. Laurent, Evolution of DNA polymerase families: evidences for multiple gene exchange between cellular and viral proteins, J. Mol. Evol. 54 (2002) 763–773.
[36] J.M. Herrmann, J. Riemer, The mitochondrial disulfide relay: redox-regulated protein import into the intermembrane space, J. Biol. Chem. 287 (2012) 4426–4433.
[37] N. Wagener, M. Ackermann, S. Funes, W. Neupert, A pathway of protein translocation in mitochondria mediated by the AAA-ATPase Bcs1, Mol. Cell 44 (2011) 191–202.
[38] S. Andersson, Y. Hofmann, A. Nordling, X.Q. Li, S. Nivelius, T.B. Andersson, M. Ingelman-Sundberg, I. Johansson, Characterization and partial purification of the rat and human enzyme systems active in the reduction of N-hydroxymelagatran and benzamidoxime, Drug Metab. Dispos. 33 (2005) 570–578.
[39] M. Pusnik, O. Schmidt, A.J. Perry, S. Oeljeklaus, M. Niemann, B. Warscheid, T. Lithgow, C. Meisinger, A. Schneider, Mitochondrial preprotein translocase of trypanosomatids has a bacterial origin, Curr. Biol. 21 (2011) 1738–1743.
[40] L. Fan, S. Kim, C.L. Farr, K.T. Schaefer, K.M. Randolph, J.A. Tainer, L.S. Kaguni, A novel processive mechanism for DNA synthesis revealed by structure, modeling and mutagenesis of the accessory subunit of human mitochondrial DNA polymerase, J. Mol. Biol. 358 (2006) 1229–1243.
[41] Y.I. Wolf, E.V. Koonin, Origin of an animal mitochondrial DNA polymerase subunit via lineage-specific acquisition of a glycyl-tRNA synthetase from bacteria of the *Thermus-Deinococcus* group, Trends Genet. 17 (2001) 431–433.
[42] M. Minczuk, J. He, A.M. Duch, T.J. Ettema, A. Chlebowski, K. Dzionek, L.G. Nijtmans, M.A. Huynen, I.J. Holt, TEFM (c17orf42) is necessary for transcription of human mtDNA, Nucleic Acids Res. 39 (2011) 4284–4299.
[43] G. von Heijne, Why mitochondria need a genome, FEBS Lett. 198 (1986) 1–4.
[44] A.D. de Grey, Forces maintaining organellar genomes: is any as strong as genetic code disparity or hydrophobicity? Bioessays 27 (2005) 436–446.
[45] B.F. Lang, C. O'Kelly, T. Nerad, M.W. Gray, G. Burger, The closest unicellular relatives of animals, Curr. Biol. 12 (2002) 1773–1778.
[46] E. Gazave, P. Lapebie, E. Renard, J. Vacelet, C. Rocher, A.V. Ereskovsky, D.V. Lavrov, C. Borchiellini, Molecular phylogeny restores the supra-generic subdivision of homoscleromorph sponges (Porifera, Homoscleromorpha), PLoS One 5 (2010) e14290.
[47] J.M. Herrmann, F. Kauff, H.E. Neuhaus, Thiol oxidation in bacteria, mitochondria and chloroplasts: common principles but three unrelated machineries? Biochim. Biophys. Acta 1793 (2009) 71–77.
[48] J. Filee, P. Forterre, Viral proteins functioning in organelles: a cryptic origin? Trends Microbiol. 13 (2005) 510–513.
[49] A.M. Duchene, A. Giritch, B. Hoffmann, V. Cognat, D. Lancelin, N.M. Peeters, M. Zaepfel, L. Marechal-Drouard, I.D. Small, Dual targeting is the rule for organellar aminoacyl-tRNA synthetases in *Arabidopsis thaliana*, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 16484–16489.
[50] R. Szklarczyk, M.A. Huynen, Expansion of the human mitochondrial proteome by intra- and inter-compartmental protein duplication, Genome Biol. 10 (2009) R135.
[51] E.H. Meyer, J.L. Heazlewood, A.H. Millar, Mitochondrial acyl carrier proteins in *Arabidopsis thaliana* are predominantly soluble matrix proteins and none can be confirmed as subunits of respiratory Complex I, Plant Mol. Biol. 64 (2007) 319–327.
[52] L. Fokkens, B. Snel, Cohesive versus flexible evolution of functional modules in eukaryotes, PLoS Comput. Biol. 5 (2009) e1000276.
[53] U. Brandt, Energy converting NADH:quinone oxidoreductase (complex I), Annu. Rev. Biochem. 75 (2006) 69–92.
[54] M. Lapaille, A. Escobar-Ramirez, H. Degand, D. Baurain, E. Rodriguez-Salinas, N. Coosemans, M. Boutry, D. Gonzalez-Halphen, C. Remacle, P. Cardol, Atypical subunit composition of the chlorophycean mitochondrial F1FO-ATP synthase and role of Asa7 protein in stability and oligomycin resistance of the enzyme, Mol. Biol. Evol. 27 (2010) 1630–1644.
[55] T. Gabaldon, D. Rainey, M.A. Huynen, Tracing the evolution of a large protein complex in the eukaryotes, NADH:ubiquinone oxidoreductase (Complex I), J. Mol. Biol. 348 (2005) 857–870.
[56] M.A. Huynen, M. de Hollander, R. Szklarczyk, Mitochondrial proteome evolution and genetic disease, Biochim. Biophys. Acta 1792 (2009) 1122–1129.
[57] P. Smits, J.A. Smeitink, L.P. van den Heuvel, M.A. Huynen, T.J. Ettema, Reconstructing the evolution of the mitochondrial ribosomal proteome, Nucleic Acids Res. 35 (2007) 4686–4703.
[58] C.Y. Yip, M.E. Harbour, K. Jayawardena, I.M. Fearnley, L.A. Sazanov, Evolution of respiratory complex I: "supernumerary" subunits are present in the alpha-proteobacterial enzyme, J. Biol. Chem. 286 (2011) 5023–5033.
[59] D.A. Maslov, M.R. Sharma, E. Butler, A.M. Falick, M. Gingery, R.K. Agrawal, L.L. Spremulli, L. Simpson, Isolation and characterization of mitochondrial ribosomes and ribosomal subunits from Leishmania tarentolae, Mol. Biochem. Parasitol. 148 (2006) 69–78.
[60] A. Zikova, A.K. Panigrahi, R.A. Dalley, N. Acestor, A. Anupama, Y. Ogata, P.J. Myler, K. Stuart, Trypanosoma brucei mitochondrial ribosomes: affinity purification and component identification by mass spectrometry, Mol. Cell. Proteomics 7 (2008) 1286–1296.
[61] A. Zikova, A. Schnaufer, R.A. Dalley, A.K. Panigrahi, K.D. Stuart, The F(0)F(1)-ATP synthase complex contains novel subunits and is essential for procyclic *Trypanosoma brucei*, PLoS Pathog. 5 (2009) e1000436.
[62] R.M. Gawryluk, M.W. Gray, Evidence for an early evolutionary emergence of gamma-type carbonic anhydrases as components of mitochondrial respiratory complex I, BMC Evol. Biol. 10 (2010) 176.
[63] P. Cardol, F. Vanrobaeys, B. Devreese, J. Van Beeumen, R.F. Matagne, C. Remacle, Higher plant-like subunit composition of mitochondrial complex I from *Chlamydomonas reinhardtii*: 31 conserved components among eukaryotes, Biochim. Biophys. Acta 1658 (2004) 212–224.
[64] J. Soding, A. Biegert, A.N. Lupas, The HHpred interactive server for protein homology detection and structure prediction, Nucleic Acids Res. 33 (2005) W244–W248.
[65] H. Heinrich, J.E. Azevedo, S. Werner, Characterization of the 9.5-kDa ubiquinone-binding protein of NADH:ubiquinone oxidoreductase (complex I) from *Neurospora crassa*, Biochemistry 31 (1992) 11420–11424.
[66] A.I. Jonckheere, M. Huigsloot, M. Lammens, J. Jansen, L.P. van den Heuvel, U. Spiekerkoetter, J.C. von Kleist-Retzow, M. Forkink, W.J. Koopman, R. Szklarczyk, M.A. Huynen, J.A. Fransen, J.A. Smeitink, R.J. Rodenburg, Restoration of complex V deficiency caused by a novel deletion in the human TMEM70 gene normalizes mitochondrial morphology, Mitochondrion 11 (2011) 954–963.
[67] R.M. de Graaf, T.A. van Alen, B.E. Dutilh, J.W. Kuiper, H.J. van Zoggel, M.B. Huynh, H.D. Gortz, M.A. Huynen, J.H. Hackstein, The mitochondrial genomes of the ciliates *Euplotes minuta* and *Euplotes crassus*, BMC Genomics 10 (2009) 514.
[68] E.C. Swart, M. Nowacki, J. Shum, H. Stiles, B.P. Higgins, T.G. Doak, K. Schotanus, V.J. Magrini, P. Minx, E.R. Mardis, L.F. Landweber, The *Oxytricha trifallax* mitochondrial genome, Genome Biol. Evol. 4 (2012) 136–154.
[69] P. Balabaskaran Nina, N.V. Dudkina, L.A. Kane, J.E. van Eyk, E.J. Boekema, M.W. Mather, A.B. Vaidya, Highly divergent mitochondrial ATP synthase complexes in *Tetrahymena thermophila*, PLoS Biol. 8 (2010) e1000418.
[70] L. Fokkens, S.M. Botelho, J. Boekhorst, B. Snel, Enrichment of homologs in insignificant BLAST hits by co-complex network alignment, BMC Bioinforma. 11 (2010) 86.
[71] R. Szklarczyk, B.F. Wanschers, T.D. Cuypers, J.J. Esseling, M. Riemersma, M.A. van den Brand, J. Gloerich, E. Lasonder, L.P. van den Heuvel, L.G. Nijtmans, M.A. Huynen, Iterative orthology prediction uncovers new mitochondrial proteins and identifies C12orf62 as the human ortholog of COX14, a protein involved in the assembly of cytochrome c oxidase, Genome Biol. 13 (2012) R12.
[72] M. Huigsloot, L.G. Nijtmans, R. Szklarczyk, M.J. Baars, M.A. van den Brand, M.G. Hendriksfranssen, L.P. van den Heuvel, J.A. Smeitink, M.A. Huynen, R.J. Rodenburg, A

mutation in C2orf64 causes impaired cytochrome c oxidase assembly and mitochondrial cardiomyopathy, Am. J. Hum. Genet. 88 (2011) 488–493.
[73] W. Weraarpachai, F. Sasarman, T. Nishimura, H. Antonicka, K. Aure, A. Rotig, A. Lombes, E.A. Shoubridge, Mutations in C12orf62, a factor that couples COX I synthesis with cytochrome c oxidase assembly, cause fatal neonatal lactic acidosis, Am. J. Hum. Genet. 90 (2012) 142–151.
[74] J.M. Bullard, Y.C. Cai, L.L. Spremulli, Expression and characterization of the human mitochondrial leucyl-tRNA synthetase, Biochim. Biophys. Acta 1490 (2000) 245–258.
[75] H. Roy, J. Ling, J. Alfonzo, M. Ibba, Loss of editing activity during the evolution of mitochondrial phenylalanyl-tRNA synthetase, J. Biol. Chem. 280 (2005) 38186–38192.
[76] K. Bych, S. Kerscher, D.J. Netz, A.J. Pierik, K. Zwicker, M.A. Huynen, R. Lill, U. Brandt, J. Balk, The iron–sulphur protein Ind1 is required for effective complex I assembly, EMBO J. 27 (2008) 1736–1746.
[77] J.A. Birrell, J. Hirst, Truncation of subunit ND2 disrupts the threefold symmetry of the antiporter-like subunits in complex I from higher metazoans, FEBS Lett. 584 (2010) 4247–4252.
[78] T.E. Shutt, M.W. Gray, Twinkle, the mitochondrial replicative DNA helicase, is widespread in the eukaryotic radiation and may also be the mitochondrial DNA primase in most eukaryotes, J. Mol. Evol. 62 (2006) 588–599.
[79] K.T. Nguyen, X. Hu, C. Colton, R. Chakrabarti, M.X. Zhu, D. Pei, Characterization of a human peptide deformylase: implications for antibacterial drug design, Biochemistry 42 (2003) 9952–9958.
[80] S. Escobar-Alvarez, J. Gardner, A. Sheth, G. Manfredi, G. Yang, O. Ouerfelli, M.L. Heaney, D.A. Scheinberg, Inhibition of human peptide deformylase disrupts mitochondrial function, Mol. Cell. Biol. 30 (2010) 5099–5109.
[81] S. Kleff, B. Kemper, R. Sternglanz, Identification and characterization of yeast mutants and the gene for a cruciform cutting endonuclease, EMBO J. 11 (1992) 699–704.
[82] M.F. White, D.M. Lilley, Characterization of a Holliday junction-resolving enzyme from *Schizosaccharomyces pombe*, Mol. Cell. Biol. 17 (1997) 6465–6471.
[83] L. Aravind, K.S. Makarova, E.V. Koonin, Survey and summary: Holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories, Nucleic Acids Res. 28 (2000) 3417–3432.
[84] S. Ceschini, A. Keeley, M.S. McAlister, M. Oram, J. Phelan, L.H. Pearl, I.R. Tsaneva, T.E. Barrett, Crystal structure of the fission yeast mitochondrial Holliday junction resolvase Ydc2, EMBO J. 20 (2001) 6601–6611.
[85] K. Ichiyanagi, H. Iwasaki, T. Hishida, H. Shinagawa, Mutational analysis on structure–function relationship of a Holliday junction specific endonuclease RuvC, Genes Cells 3 (1998) 575–586.
[86] G.S. Bassi, D.M. de Oliveira, M.F. White, K.M. Weeks, Recruitment of intron-encoded and co-opted proteins in splicing of the bI3 group I intron RNA, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 128–133.
[87] D. Close, S.J. Johnson, M.A. Sdano, S.M. McDonald, H. Robinson, T. Formosa, C.P. Hill, Crystal structures of the *S. cerevisiae* Spt6 core and C-terminal tandem SH2 domain, J. Mol. Biol. 408 (2011) 697–713.
[88] Y. Planques, N. Capitanio, G. Capitanio, E. De Nitto, G. Villani, S. Papa, Role of supernumerary subunits in mitochondrial cytochrome c oxidase, FEBS Lett. 258 (1989) 285–288.
[89] J. Hirst, Why does mitochondrial complex I have so many subunits? Biochem. J. 437 (2011) e1–e3.
[90] H. Angerer, K. Zwicker, Z. Wumaier, L. Sokolova, H. Heide, M. Steger, S. Kaiser, E. Nubel, B. Brutschy, M. Radermacher, U. Brandt, V. Zickermann, A scaffold of accessory subunits links the peripheral arm and the distal proton-pumping module of mitochondrial complex I, Biochem. J. 437 (2011) 279–288.
[91] M.W. Gray, J. Lukes, J.M. Archibald, P.J. Keeling, W.F. Doolittle, Cell biology. Irremediable complexity? Science 330 (2010) 920–921.
[92] T.W. O'Brien, Evolution of a protein-rich mitochondrial ribosome: implications for human genetic disease, Gene 286 (2002) 73–79.
[93] M.R. Sharma, E.C. Koc, P.P. Datta, T.M. Booth, L.L. Spremulli, R.K. Agrawal, Structure of the mammalian mitochondrial ribosome reveals an expanded functional role for its component proteins, Cell 115 (2003) 97–108.
[94] J.B. Dacks, A.A. Peden, M.C. Field, Evolution of specificity in the eukaryotic endomembrane system, Int. J. Biochem. Cell Biol. 41 (2009) 330–340.
[95] R.L. Tatusov, M.Y. Galperin, D.A. Natale, E.V. Koonin, The COG database: a tool for genome-scale analysis of protein functions and evolution, Nucleic Acids Res. 28 (2000) 33–36.
[96] M.A. Huynen, I. Duarte, Z.M. Chrzanowska-Lightowlers, S.B. Nabuurs, Structure based hypothesis of a mitochondrial ribosome rescue mechanism, Biol. Direct 7 (2012) 14.
[97] A.D. Garcia, L. Aravind, E.V. Koonin, B. Moss, Bacterial-type DNA Holliday junction resolvases in eukaryotic viruses, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 8926–8931.
[98] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.
[99] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, Clustal W and Clustal X version 2.0, Bioinformatics 23 (2007) 2947–2948.
[100] A.M. Waterhouse, J.B. Procter, D.M. Martin, M. Clamp, G.J. Barton, Jalview Version 2—a multiple sequence alignment editor and analysis workbench, Bioinformatics 25 (2009) 1189–1191.
[101] A. Stamatakis, RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models, Bioinformatics 22 (2006) 2688–2690.
[102] F. Abascal, R. Zardoya, D. Posada, ProtTest: selection of best-fit models of protein evolution, Bioinformatics 21 (2005) 2104–2105.
[103] C. Carrie, E. Giraud, O. Duncan, L. Xu, Y. Wang, S. Huang, R. Clifton, M. Murcha, A. Filipovska, O. Rackham, A. Vrielink, J. Whelan, Conserved and novel functions for *Arabidopsis thaliana* MIA40 in assembly of proteins in mitochondria and peroxisomes, J. Biol. Chem. 285 (2010) 36138–36148.
[104] B. Hedtke, T. Borner, A. Weihe, Mitochondrial and chloroplast phage-type RNA polymerases in *Arabidopsis*, Science 277 (1997) 809–811.
[105] T. Frickey, A.N. Lupas, Phylogenetic analysis of AAA proteins, J. Struct. Biol. 146 (2004) 2–10.
[106] D. Simsek, M. Jasin, DNA ligase III: a spotty presence in eukaryotes, but an essential function where tested, Cell Cycle 10 (2011) 3636–3644.
[107] H. Kochiwa, M. Itaya, M. Tomita, A. Kanai, Stage-specific expression of *Caenorhabditis elegans* ribonuclease H1 enzymes with different substrate specificities and bivalent cation requirements, FEBS J. 273 (2006) 420–429.
[108] E. Babiychuk, K. Vandepoele, J. Wissing, M. Garcia-Diaz, R. De Rycke, H. Akbari, J. Joubes, T. Beeckman, L. Jansch, M. Frentzen, M.C. Van Montagu, S. Kushnir, Plastid gene expression and plant development require a plastidic protein of the mitochondrial transcription termination factor family, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 6674–6679.
[109] E. Petit, X. Michelet, C. Rauch, J. Bertrand-Michel, F. Terce, R. Legouis, F. Morel, Glutathione transferases kappa 1 and kappa 2 localize in peroxisomes and mitochondria, respectively, and are involved in lipid metabolism and respiration in *Caenorhabditis elegans*, FEBS J. 276 (2009) 5030–5040.