Does Mitochondrial DNA Evolution in Metazoa Drive the Origin of New Mitochondrial Proteins?

S. L. van Esveld\textsuperscript{1,2} M. A. Huynen\textsuperscript{1}\textsuperscript{*}

\textsuperscript{1}Radboud Center for Mitochondrial Medicine, Radboudumc, Nijmegen, The Netherlands
\textsuperscript{2}Centre for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboudumc, Nijmegen, The Netherlands

Abstract

Most eukaryotic cells contain mitochondria with a genome that evolved from their \( \alpha \)-proteobacterial ancestor. In the course of eukaryotic evolution, the mitochondrial genome underwent a dramatic reduction in size, caused by the loss and translocation of genes. This required adjustments in mitochondrial gene expression mechanisms and resulted in a complex collaborative system of mitochondrially encoded transfer RNAs and ribosomal RNAs with nuclear encoded proteins to express the mitochondrial encoded oxidative phosphorylation (OXPHOS) proteins. In this review, we examine mitochondrial gene expression from an evolutionary point of view: to what extent can we correlate changes in the mitochondrial genome in the evolutionary lineage leading to human with the origin of new nuclear encoded proteins. We dated the evolutionary origin of mitochondrial proteins that interact with mitochondrial DNA or its RNA and/or protein products in a systematic manner and compared them with documented changes in the mitochondrial DNA. We find anecdotal but accumulating evidence that metazoan RNA-interacting proteins arose in conjunction with changes of the mitochondrial DNA. We find no substantial evidence for such compensatory evolution in new OXPHOS proteins, which appear to be constrained by the ability to form supercomplexes.

\( \text{© 2018 IUBMB Life, 70(12):1240–1250, 2018} \)

Keywords: mitochondrial gene expression; oxidative phosphorylation; mtDNA; mtRNA; animal evolution; mitochondrial genome; mitochondrial proteome

INTRODUCTION

Almost all eukaryotes contain one or multiple mitochondria per cell. A proper functioning organelle is essential for survival and health of the organisms. This is illustrated by several human mitochondrial diseases, like Leigh-syndrome (1), and by targeting parasites’ mitochondria with drugs, like Atovaquone that targets mitochondrial complex III in \textit{Plasmodium} species (2).

All mitochondrial organelles originated from a single endosymbiotic event of an \( \alpha \)-proteobacterium (3) whose phylogenetic position remains a subject of discussion (4). During the course of evolution, the host cell and the bacterial cell adjusted to each other, resulting in the large variety of eukaryotic cells and mitochondria-related organelles known today. This adjustment involved degradation, or if you will, streamlining of the \( \alpha \)-proteobacterial genome, which in human is reduced to only 13 protein coding genes. Some genes were lost, while others were transferred to the host genome (5). As a result of this ongoing process in the many evolutionary lineages, both the mitochondrial genome and the mitochondrial proteome show a large variety between species (6, 7), and only a minor fraction of the current mitochondrial proteome can confidently be traced back to \( \alpha \)-Proteobacteria (8).

The proteins that most often remain encoded by the mitochondrial genome are hydrophobic subunits of oxidative phosphorylation (OXPHOS) complexes (9, 10). Combined these
complexes oxidize NADH and FADH, and use the released energy to produce ATP (11), which is an important player in for example muscle contraction, intercellular signaling, and DNA/RNA synthesis (12). In humans, four out of five respiratory chain complexes contain mitochondrially encoded subunits, namely complex I, III, IV, and V. The expression of these genes is completely dependent on proteins encoded by the nuclear genome, while the required ribosomal RNA (rRNAs) and transfer RNAs (tRNAs) are encoded by the mitochondrial genome (13, 14). Besides proteins necessary for mitochondrial DNA (mtDNA) expression, OXPHOS also depends on a set of nuclear encoded OXPHOS subunits, of which the number increased quite dramatically between the symbiosis event and the eukaryotic radiation (15). Given the many mitochondrial and nuclear encoded OXPHOS subunits, a proper communication between mitochondrial and nuclear gene expression is required.

Human mtDNA is a circular molecule of ~16.6 kb containing one large non-coding region, few intergenic nucleotides, and 13 OXPHOS, 22 tRNA, and 2 rRNA genes (13, 14). The DNA is bound by nuclear encoded proteins to form nucleoids, which occur as discrete punctae throughout the mitochondrial network in immunofluorescent images (16–18). Initial experiments aimed to identify the proteins that are part of these structures assumed a static nucleoid composition (19). The pool of proteins found to be associated with nucleoids however turned out to be dynamic, making the identification task difficult (19). Replicating nucleoids are mainly membrane bound (19) and according to in vitro experiment require the association of at least three proteins, namely Twinkle, POLG, and mtSSB (20), but in vivo the number is probably larger.

When associated with transcription proteins (e.g., POLRMT, TFB2M, TFAM, TEFM) mtDNA is transcribed into two long polycistronic transcripts. The light strand transcript contains one protein coding gene and eight tRNA genes, while the heavy strand contains 12 protein coding genes, 14 tRNA genes, and two ribosomal RNA genes (13, 14). Most of the rRNA and protein coding genes are surrounded by tRNA genes. According to the tRNA punctuation gene order model, the rRNA gene order helps processing of the polycistronic transcript into separate RNA molecules (21). Already within the polycistronic transcript the tRNAs are thought to form a secondary and tertiary structure, allowing nuclease cleavage at the 3‘ end by ELAC2 and at the 5‘ and by the RNase P complex (22–24). Subsequently, transcripts further mature by the addition of modifications and/or polyA tails to obtain a functional pool of RNAs. The rRNA molecules form the basis of the mitochondrial ribosome that can translate the 13 OXPHOS genes. Part of the cleavage of the polycistronic transcript and the modifications are thought to occur in so-called mitochondrial RNA granules, which like nucleoids consist of a pool of nuclear encoded proteins and instead of DNA contain newly synthesized mtRNA molecules (e.g., Refs. (25, 26).

Our understanding of the mitochondrial gene expression in vertebrates is far from complete. For example, not all genes are surrounded by tRNA genes, so how are these exactly separated from the polycistronic transcript? What is the purpose of organizing RNA molecules in RNA granules? And how does translation stop in genes that do not have a stop codon, even after the addition of a polyA tail or correction via a ribosomal frameshift (27).

The tRNA punctuation gene order likely evolved relatively recently in evolution, as it can only be found in Bilateria (28). Metazoan lineages that originated before the Bilateria, like sponges and jellyfish, have a different gene order, varying amounts of non-coding DNA and varying sets of tRNAs, some of which are incomplete (29). Besides the variations in the order and number of genes in metazoan mitochondria, there are also variations in the presence of introns, nucleotide substitution rates, genetic code, GC-skew, RNA editing, translational frame shifts, and genome size and shape (linear or circular) (29, 30). This variation makes it interesting to examine the mitochondrial gene expression proteome from an evolutionary point of view. Are changes in the mtDNA paralleled by changes in the mitochondrial proteome that is involved in mitochondrial gene expression? Analyzing this might help us understand the origin of mitochondrial gene expression mechanisms in human. After all, nothing in biology makes sense, except in the light of evolution (31). To be sure, co-evolution between the mitochondrial genome and the nuclear encoded gene expression program has been documented before, e.g. by the co-evolution of mitochondrial encoded tRNAs with nuclear encoded aminoacyl-tRNA synthetases (32) or by the co-evolution of the presence of stop-codons and their release factors (27), while the question about co-evolution of the mitochondrial encoded rRNAs with the nuclear encoded mitochondrial ribosomal proteins has been raised multiple times (33–35). Here, we specifically focused on the origin of new nuclear encoded, mitochondrial proteins in the Metazoa in conjunction with changes in the mtDNA in that taxon. We excluded the rRNA and the ribosomal proteome as those have already been well studied (33–35).

The human mitochondrial genome resembles the typical mtDNA organization observed across many bilaterian species (29, 30, 36, 37). This bilaterian conserved organization and the large differences observed when comparing all metazoan mitochondrial genomes with each other argue that most major changes in mtDNA in the lineage leading to human occurred early in animal evolution (29). Changes in animal mtDNA organization appear to correlate with two main transitions in animal evolution: the origin of multicellularity and the origin of the Bilateria (28). Since there is such a huge diversity in metazoan mtDNA molecules, and many metazoan nuclear genomes have been sequenced it is especially interesting to study changes in mitochondrial gene expression within this group.

As described earlier, mtDNA evolution in metazoa has been well documented (29, 30, 36–38), both with respect to the changes that occurred as with respect to when those changes occurred. To correlate that evolution to changes in the mitochondrial proteome, we also need the exact date of the origin of mitochondrial proteins. We used TreeFam Release 9 (39) and subsequent manual curation of individual protein families to

VAN ESVELD AND HUYNEN
systematically determine the age of mitochondrial proteins. We focused on changes after the split of the metazoa from other Opisthokonts, like fungi and the unicellular ancestors of the metazoa like the choanoflagellates. We specifically examined proteins that in humans interact with mtDNA or its encoded products to map and potentially understand co-evolution of the mitochondrial genome and the nuclear encoded mitochondrial proteome.

MITOCHONDRIAL PROTEINS GAINED IN THE METAZOA

Proteins can either be evolutionary innovations in mitochondria within a specific taxonomic clade (i) when they do not have homologs outside of that clade, like the complex 1 protein NDUFCl that can only be detected in vertebrates (Euteleostomi); (ii) when they arose from a gene duplication in that clade and their “ancestral” protein was not mitochondrial, like GRSF1 that arose from a gene duplication at the root of the vertebrates from a non-mitochondrial protein family (25); or (iii) when they are duplications of proteins that were already mitochondrial, so called intra-compartmental protein duplications (40) like MTRF1 that arose from a duplication of MTRF1L at the root of the vertebrates (27). Nevertheless, this last class was excluded because we aimed to detect truly novel mitochondrial functions, while duplications of mitochondrial proteins are often associated with subfunctionalization, e.g. via tissue-specific gene expression (40). We used TreeFam Release 9 (39) as a first filter to determine which human mitochondrial proteins [MITOCARTA2.0 (41)] originated in the Metazoa. The TreeFam database contains phylegetic trees of genes from 104 animal genomes (101 are Bilateria, of which 56 are vertebrates) and five outgroup genomes (two choanoflagellates, two ascomycetes, and one plant) (39). Each tree gets an identification number (e.g., TF315274), is available on the TreeFam webpage (http://www.treefam.org/) and consists of a gene family that descended from one gene that first occurred within the Metazoa or was present in the last common ancestor of all Metazoa (42). These trees thus allowed us to determine which mitochondrial genes originate in Metazoa by either occurring as a novel gene within the Metazoa (no outgroup species in the tree) or occurring due to a duplication event (multiple human genes in the tree).

The obtained list of metazoan proteins was manually curated in three steps. The first step was based on function descriptions to select for proteins that are known to interact with mtDNA, mtRNA, or mitochondrial encoded protein(s). The second step was done to confirm the age of the genes that occur in novel metazoan families, using literature (15, 38, 43–45). PSI-blast (46), and J ACKHMMR (47). The reason for the second curation step is that the BLAST and HMMER based methods used by TreeFam to annotate genes to a family are not sensitive enough to find all orthologs and that TreeFam does not contain all published genomes. It therefore might identify a protein family as specific to the Metazoa that is actually older than that.

Novel metazoan genes were excluded based on PSI-blast or JACKHMMR when a 1:1 ortholog was found in a non-metazoan species, or when a mitochondrial paralog was found that resulted from a pre-metazoan duplication. A third curation step was performed by examining the TreeFam trees of duplicated genes to only include mitochondrial proteins that duplicated from a non-mitochondrial ancestor protein and not vice versa. The exceptions that we retained in the list are new OXPHOS proteins and assembly factors resulting from the duplication of an already mitochondrial protein, like DMAC2 (ATP5SL) a complex I assembly factor that resulted from a duplication of ATP5S in the Bilateria [TF315274 (39)]. Also, TOP1MT, a protein with a likely mitochondrial and nuclear location before the duplication, was retained. After these manual curation steps, we obtained a list of 33 proteins (Table 1) that interact with mtDNA, mtRNA, or mitochondrial encoded protein(s). These can be subdivided into three classes: 16 mtRNA interacting, 3 mtDNA interacting, and 14 OXPHOS interacting proteins. We examined the possible correlation between the gain of these proteins and changes in the mtDNA in detail.

DNA AND RNA INTERACTING PROTEINS

TRMT10C and tRNA Folding

The protein TRMT10C/MRPP1 is of bilaterian origin, likely originating from TRMT10B or TRMT10A (Fig. 1). The protein forms a subcomplex with HSD17B10 (MRPP2) and methylates the N1-methylguanine (m1G6) and N1-methyladenine at position 9 (m1A9) in tRNAs (48). Methylation at position nine prevents this position from base pairing in tRNA(Lys), shifting the equilibrium secondary structure to one that allows the typical L-shaped and functional tertiary structure (49). In general, bilaterian mitochondrial tRNAs appear methylated at position nine (50).

The secondary structure of tRNA(Ser) of Bilateria appears highly derived from the canonical cloverleaf structure, with one of the prominent features being a deletion in the D-arm (also named D-loop) (51). As non-bilaterian animals like the demosponges, cnidarians, and Placozoa have conventional tRNAs, this deletion likely originated at the root of the Bilateria (52). It should however be noted that a large amount of variation in the size of the D-arm has been observed in (non-bilaterian) glass sponges (52). It would be interesting to determine whether TRMT10C is also required for proper folding of tRNA(Ser) in Bilateria and whether non-bilaterian Metazoa, which do not have TRMT10C, have a non-methylated position nine in their mitochondrial tRNAs.

RNA Processing Proteins and Gene Order Rearrangement

Most bilaterian animals differ from other animals since they have a mtDNA molecule containing rRNA and mtRNA genes that are almost all surrounded by tRNA genes (29). This typical bilaterian gene order formed the basis of the tRNA punctuation model. To obtain separate transcripts from the polycistronic
transcript, tRNA genes are cleaved out, thereby releasing separate rRNAs and mRNAs (21).

As this bilaterian processing is dependent on the position of the tRNA genes, the gene order rearrangements likely influenced RNA processing of the polycistronic transcript. We therefore examined RNA interacting proteins that originate at the Bilateria: TRUB2 that duplicated from TRUB1 (TF320759), MTPAP that duplicated from TUT1 (TF354308), TRMT10C that duplicated from TRMT10A/B (Fig. 1), the FASTK family of proteins that underwent multiple duplications (TF324885, TF331796, TF352874, TF352875), and the novel protein DHX30 (TF352030) (39).

TRUB2 together with RPUSD3 contribute to the conversion of uridine to pseudouridine in mitochondrial mRNAs, specifically of COXI and COXIII. Depletion of TRUB2 results in a decrease in mitochondrial protein synthesis, without changing transcript abundance or stability (55). DHX30 plays an important role in the assembly of the large ribosomal subunit and knockdown of the protein results in a severe decrease in mitochondrial protein synthesis, especially of ND5, ND6, COXI, and COXII, while messenger transcript levels were only moderately affected (56). TRUB2 and DHX30 are thus not directly involved in processing of the tRNA punctuated polycistronic transcripts and not linked to the gene order rearrangement, but rather influence the translation efficiency without changing the abundance of mitochondrial mRNAs. It is not apparent why they are specific to the Bilateria.

The bilaterian protein TRMT10C, besides its function in tRNA methylation, is a component of human protein-only RNase

TABLE 1  
Metazoan origin genes which interact with mitochondrial encoded DNA, RNA, or protein

| Gene | Description | Date\(^a\) | Interaction |
|------|-------------|----------|-------------|
| NDUFB1 | NADH:ubiquinone oxidoreductase subunit B1 | Bilateria I | I |
| NDUFB6 | NADH:ubiquinone oxidoreductase subunit B6 | Bilateria I | I |
| TMEM126B | transmembrane protein 126B | Bilateria I | I |
| NDUFC1 | NADH:ubiquinone oxidoreductase subunit C1 | Euteleostomi I | I |
| DMAC2 | distal membrane arm assembly complex 2 | Bilateria I | I |
| ECSIT | ECSIT signalling integrator | Metazoa I | I |
| ACAD9 | acyl-CoA dehydrogenase family member 9 | Metazoa I | I |
| TTC19 | tetrameric repeat domain 19 | Eumetazoa III | III |
| COX7B | cytochrome c oxidase subunit 7B | Euteleostomi IV | IV |
| COX7B2 | cytochrome c oxidase subunit 7B2 | Euteleostomi IV | IV |
| COX8A | cytochrome c oxidase subunit 8A | Bilateria IV | IV |
| COX8C | cytochrome c oxidase subunit 8C | Bilateria IV | IV |
| COX7A2L | cytochrome c oxidase subunit 7A2 like | Vertebrata IV | IV |
| ATP5S | ATP synthase, H+ transporting, mitochondrial Fo complex subunit s (factor B) | Metazoa V | V |
| POLG2 | DNA polymerase gamma 2, accessory subunit | Metazoa mtDNA | mtDNA |
| TOP1MT | DNA topoisomerase I mitochondrial | Euteleostomi mtDNA | mtDNA |
| NT5M | 5’-3’-nucleotidase, mitochondrial | Bilateria mtRNA | mtRNA |
| TRMT1 | tRNA methyltransferase 1 | Metazoa mtRNA | mtRNA |
| MTPAP | mitochondrial poly(A) polymerase | Bilateria mtRNA | mtRNA |
| HARS2 | histidyl-tRNA synthetase 2, mitochondrial | Mammalia mtRNA | mtRNA |
| AARS2 | alanyl-tRNA synthetase 2, mitochondrial | Metazoa mtRNA | mtRNA |
| GRSF1 | G-rich RNA sequence binding factor 1 | Euteleostomi mtRNA | mtRNA |
| DHX30 | DEH-box helicase 30 | Bilateria mtRNA | mtRNA |
| FASTK | Fas activated serine/threonine kinase | Vertebrata mtRNA | mtRNA |
| FASTKD1 | FAST kinase domains 1 | Bilateria mtRNA | mtRNA |
| FASTKD2 | FAST kinase domains 2 | Bilateria mtRNA | mtRNA |
| FASTKD3 | FAST kinase domains 3 | Bilateria mtRNA | mtRNA |
| FASTKD4 | FAST kinase domains 4 (TBRG4) | Bilateria mtRNA | mtRNA |
| FASTKD5 | FAST kinase domains 5 | Euteleostomi mtRNA | mtRNA |
| TRUB2 | TruB pseudouridine synthase family member 2 | Bilateria mtRNA | mtRNA |
| TEFM | transcription elongation factor, mitochondrial | Metazoa mtRNA | mtRNA |
| PABPC5 | poly(A) binding protein cytoplasmic 5 | Eutheria mtRNA | mtRNA |
| TRMT10C | tRNA methyltransferase 10C, mitochondrial RNase P subunit | Bilateria mtRNA | mtRNA |

\(^a\) Date of duplication event when this involved a non-mitochondrial and a mitochondrial gene, else date of gene family in TreeFam checked or corrected with manual curation.


increase the stability of mitochondrially encoded complex I. Deadenylation of mitochondrial transcripts was shown to be a common occurrence in bacteria. Polyadenylation primes the transcripts for degradation, with the addition of a poly(A) tail stabilizing the transcripts. However, in bac-

tain are not present in the non-bilaterian species. In complete UAA stop codons for the switch in the G-C skew is possibly an inversion of the non-bilaterian genome organization, while a negative G-C skew is present in the mitochondrial genome of non-bilaterian species. Crystallography of the homologous tRNA(Val) and tRNA(Ala) in the mitochondrial subunits, while it decreases the stability of complex IV subunits (57, 58). The role of polyadenylation in the stability of mitochondrial transcripts is therefore not completely understood. In addition to this role in (de)stabilization, it has been suggested that polyadenylation in humans also creates the otherwise absent conventional stop codon UAA for ND1, ND2, ND3, ND4, COXIII, ATP6, and CytB (13, 14). Incomplete UAA stop codons are not present in the non-bilaterian species Trichoplax adhaerens, Mnemiopsis leidyi, and Amphimedon compressa, but can be found in the bilaterian species Drosophila melanogaster, Branchiostoma floridae, and Homo sapiens, and in some Cnidaria species (present in Nematostella sp. JVK-2006, while absent in Hydra sinensis), which are the closest relatives of Bilateria. We speculate that MTPAP corrected the truncation of the UAA stop codon via polyadenylation. The polyA binding protein PABPC5, which originates in the vertebrates (TF300458, (39)), has been shown to co-immunoprecipitate with MTPAP (59), the role of this interaction is still unclear.

The FASTK protein family, whose domain organization, which combines two regions conserved in eukaryotic Fas-activated serine/threonine (FAST) kinases with an RNA binding RAP domain, is unique to Metazoa (45). The FastA family has expanded to six members (FASTK and FASTKD1–5) in humans. Low levels of sequence identity between the different members of this family preclude the making of a reliable phylogeny that includes all human family members. Nevertheless, in neighbour joining cluster analysis of the FASTK protein family (data not shown), all bilaterian members of this family clustered together and were monophyletic relative to the non-bilaterian members of this family, suggesting that all members of this family in human arose from duplications at the root of the Bilateria (FASTKD1/2/3/4) or later in evolution (FASTK and FASTKD5 in vertebrates) (TF324885, TF331796, TF352874, TF352875 (39)). Of the bilaterian FASTK protein family members, FASTK4 is most similar to the family members in the non-Bilateria, but we lack formal criteria to define which FASTK proteins are new to mitochondria. Nevertheless, it is interesting to note that the role of FASTKD1, FASTKD2, and FASTKD3 in stability or processing of miRNA coincides with the origin of the bilaterian genome organization.

GRSF1 and G-C Skew Switch
The heavy-strand of the vertebrates’ mitochondrial genome contains more guanine bases than cytosine bases, having a so-called positive G-C skew. Asymmetry in the mitochondrial replication, in which one strand is single stranded during a longer period of time than the other, is thought to cause this mitochondrial genome G-C skew as is also observed in mitochondria of cancer cells (60). The mitochondrial G-C skew is variable among the non-vertebrate Metazoa, but as the most closely related species to the vertebrates, the non-vertebrate chordates Branchiostoma floridae and Ciona intestinalis have negative G-C skews (G-C/ (G+C) is −0.15 and −0.11 respectively), it is likely that the ancestor of the vertebrates had a negative G-C skew. The reason for the switch in the G-C skew is possibly an inversion of
the control region at the root of the vertebrates, because correlations between switches in G-C skew and inversions of the control region have been observed within the vertebrates. Five fish species have been reported to contain an inverted, negative G-C skew (30, 36). These species all contain unusual control regions and in four out of the five species the region(s) were inverted compared to closely related species (30, 36). This suggests that upon inversion of the control region, the mitochondrial replication mechanism inverted its polarity, which over time led to an inverted G-C skew (30, 36).

G-rich sequence factor 1 (GRSF1) originates through a duplication event from nuclear HNRNPH3 at the root of the vertebrates [TF316157 (39)] and has gained a mitochondrial targeting signal (25). The new RNA-interacting protein GRSF1 is known to bind G-rich regions, raising the question whether there is a connection between the vertebrate GC-skew switch and the arrival of GRSF1 (25).

GRSF1 mainly interacts with G-rich light strand lncND5, ND6, and lncCytB transcripts (25). It binds sequences with the motif AGGGD in which D stands for A, U, or G (25, 61–63). ND6 contains, with four motifs, the largest number of motifs within a protein coding gene (25). The complete human light strand transcript (GenBank ID: AP008819.1) contains 159 AGGGD motifs (versus only 19 on the heavy strand transcript), which is actually 50 more than expected based on individual nucleotide frequencies, suggesting that the enrichment of AGGGD motifs is not solely a side effect of the G-C skew.

High density of guanine in RNA might lead to RNA G-quadruplex (G4) structure formation. These structures have been shown to play essential roles in cytosolic RNA metabolism, including pre-mRNA splicing, polyadenylation, mRNA targeting, and translation (64). A switch in the G-C skew changes the RNA G4 structure formation and might hamper mtRNA processing. In human there are 65 possible G4 structures predicted in lncCytB, four in ND6 and ten in lncND5, corresponding with the main RNA interaction partners of GRSF1, while none were predicted for the heavy-strand transcripts (25, 65). It is possible that the arrival of GRSF1 compensated G4 formation by binding the G-rich regions and thereby allowing proper processing of the RNA. Indeed absence of GRSF1 was shown to lead to abnormal cleavage of the polycistronic transcript (26). While this article was under review it was reported that GRSF1 facilitates degradation of non-coding RNAs via G4 melting (66) and the arrival of GRSF1 was proposed to be an adaptation to mitochondrial genome changes by enabling control of G4 containing transcripts (66). It would be interesting to examine whether GRSF1 was lost from the fish species with an inverted G-C skew, but no genomic or EST data are available from these species.

**TOP1MT and Inversion of Control Region**

As discussed earlier, the replication control region likely inverted at the root of vertebrates. We found one new mtDNA interacting protein that originates via a duplication from a likely dual targeted TOP1 at the root of the vertebrates, namely TOP1MT [TF105281 (39). This mitochondrial topoisomerase type IB is important for mtDNA replication because it removes the tensions resulting from mtDNA replication (67).

Yeast contains only one TOP1 gene that is suggested to act both in the nucleus and in the mitochondrion, while vertebrates contain two copies that appear to be adjusted to the compartment in which they act (TOP1 neutral pH, TOP1MT pH around eight) (67, 68). Targeting TOP1 to the mitochondrion and TOP1MT to the nucleus showed that the proteins function differently (68).

There is indeed proof that TOP1MT interacts with the control region. Absence of TOP1MT effects non-coding RNAs of the Dloop region, 7S DNA (69) and negative supercoiling (70). In addition, trapping of TOP1MT with camptothecin showed a cluster of TOP1mt sites confined to a 150 bp region downstream from the site at which replication is prematurely terminated. This generates a 650 bp region (7S DNA) that forms the mitochondrial D-loop (71). Analysis of the stable cleavage complexes formed by expression of a mutant TOP1MT with ChIP-on-chip assays showed that TOP1MT binds to the noncoding regulatory region, accumulating especially at the two origins O1L and O2L and in the ribosomal genes (72). Taken together this raises the question whether the inversion of the control region required differentiation between a nuclear and mitochondrial TOP1 to retain normal replication.

**OXPHOS PROTEINS**

**New complex I subunits**

The genes encoding the OXPHOS subunits are, like the rest of the mitochondrial genome in mammals, under high mutation pressure and like tRNAs and the rRNAs (28) have lost parts of their nucleotides in Bilateria (73). Comparison of the mammalian complex I structure from Bos taurus with Thermus thermophilus showed loss of a C-terminal helix from ND1, three N-terminal helices and a β-sheet hairpin from ND2, the C-terminal half of trans-membrane helix 14 of ND4, and the truncation of the β-sheet hairpin from ND4 (74). Besides these losses an insertion in ND6 was observed, resulting in a displacement of transmembrane helix four within the structure (74).

Mitochondrial genome comparisons support that the losses in ND2 and ND4 occurred at the root of Bilateria, while the other changes either precede the split of the Metazoa from other Opisthokonts or cannot be determined with certainty. The ND1 C-terminal helix was lost before the Metazoa as the corresponding sequence is also absent from the Yarrowia lipolytica structure (75) and from sequences of non-bilaterian Metazoa. The large deletion observed in ND2 occurred at the root of Bilateria (37). We examined whether ND4 and ND5, paralogs of ND2, have also lost part of their sequence or structure at the origin of the Bilateria. Both the β-sheet hairpin and part of the C-terminal helix of ND4 are lost at the root of the Bilateria as they can confidently be predicted in the non-bilaterian Metazoa.
using an alignment of non-bilaterian ND4 sequences and quick2D (76). For ND5 no losses were reported previously and also our results are inconclusive. Although a β-sheet hairpin is present in the \textit{T. thermophilus} and \textit{Y. lipolytica} complex I structures (75, 77) and absent from mammalian \textit{Ovis aries} complex I (78), a β-sheet hairpin, although smaller compared to \textit{T. thermophilus}, is present in the human complex I structure (79). The insertion in ND6 is highly variable between orthologs both with respect to length and sequence identity, with e.g. the non-bilaterian metazoan \textit{T. adhaerens} having an extra ~57 amino acids in this region, while also within the mammals the length varies with 12 amino acids.

Compared to α-proteobacteria, animals contain 28 additional complex I subunits. Most of these are added early in evolution, predating the aforementioned losses in old membrane proteins observed in animals (15). Of the three subunits that originate in Metazoa, NDUFC1 occurs as a new protein in the vertebrates. The other two metazoan subunits, NDUFB1 and NDUFB6, do originate as novel proteins at the Bilateria, and their origin therefore coincides with the losses in mtDNA sequences. Besides these, we also included NDUFA10 in Fig. 2 as its association with complex I has thus far only been observed in Metazoa even though it does have orthologs outside the Metazoa, like deoxyguanosine kinase in \textit{Dictyostelium discoideum}. Within the human structure (79), NDUFA10, NDUFB6, and NDUFC1 are however not located particularly close to the lost amino acid chains (Fig. 2A), we therefore have no evidence that the gain of these subunits somehow compensates the losses in mtDNA in the Bilateria. For NDUFB1 and ND4 are both part of a 230 kDa P0-module assembly intermediate (80) and the transmembrane helix of NDUFB1 interacts with ND4. NDUFB1 is however not specifically close to the unstructured former β-sheet hairpin of ND4 nor to its lost part of the C-terminal helix (Fig. 2A). With respect to the origin of new complex I assembly factors in the Metazoa (ACAD9, TMEM126B, ECSIT, DMAC2), it is interesting to note that, like the new subunits, they interact all with the mitochondrial encoded membrane arm proteins.

Note that the three metazoan new subunits that are part of the membrane arm and that are located in the membrane, are situated at one side of complex I (79), opposite the side that interacts with complex III during supercomplex formation (Fig. 2B). Also, complex I subunit NDUFA10 and newly arrived complex IV subunits COX7B, COX7B2, COX8B, and COX8C are on the outside of the supercomplex structure (Fig. 2B). COX7A2L, also a new Complex IV protein, likely sits at the interface between complex III and IV, as it is known to be involved in supercomplex formation by promoting the stability of the supercomplex III2 + IV (81), replacing COX7A in the complex IV (82). This suggests that evolutionarily, addition of new
OXPHOS subunits has been constrained by supercomplex formation.

COEVOLUTION OF THE MITOCHONDRIAL PROTEOME WITH THE MITOCHONDRIAL GENOME

The mitochondrial genome has undergone a number of substantial changes in the Metazoa, ranging from the organization of the genes to the loss of genes and the loss of parts of genes (28, 29, 37). Here we have analyzed whether these changes coincide with the gain of proteins that potentially could compensate for these changes, an idea that has been dubbed the “prosthetic hypothesis” (83). According to this theory the reduction of mitochondrial rRNA would be compensated by the arrival of new ribosomal proteins, but previous analysis have not supported this because (i) the main expansion of the mitochondrial ribosomal protein content far predates that of the massive loss of mtRNA (15) and (ii) the new ribosomal proteins are located on the outside of the ribosome and do not replace the mtRNA (33). Also, for OXPHOS protein complexes the main expansion occurred early in eukaryotic evolution, before the radiation of the eukaryotes, while losses from mitochondrial encoded proteins occurred later in time (73). Furthermore, as we have shown here, the new OXPHOS subunits in the Metazoa are in general, although part of the membrane arm, not located specifically close to where the deletions in the mitochondrial encoded subunits occurred. There is thus no compelling evidence for the prosthetic hypothesis at the level of complete proteins, in contrast to the level of individual amino acid changes where evidence for co-evolution has been documented (84–86).

At the level of protein function, results are more encouraging. A correlation between the mitochondrial genome and the nuclear encoded proteome is supported by the absence of a RNase P gene from mtDNA of Holozoa (Metazoa and their single cell relatives) in combination with the presence of mitochondrial protein-only RNase P protein KIAA0391 (38), while in Fungi the
species that contain a RNase P gene in their mtDNA, do not contain organellar protein-only RNase P (38). In Metazoa, we could temporally and functionally link the origin of some mitochondrial DNA/rRNA interacting proteins, like TRMT10C, MTPAP, and GRSF1, to mtDNA changes (Fig. 3). It is tentative to argue that this co-evolution is driven by the mtDNA, given the relatively high rate at which it accumulates mutations in Metazoa (88), but evolutionary arguments are often rooted in wishful thinking. To make a more convincing argument we would have to observe such co-evolution in other parallel evolutionary lineages with a high accumulation of slightly deleterious mutations. Furthermore, even though we mapped the co-evolution of the mitochondrial genome and the origin of new mitochondrial proteins by narrowing down events in both using phylogenetic analyses, we still face a chicken and egg problem: we do not know whether the origin of a new, nuclear encoded mitochondrial protein allowed the change in the mitochondrial genome, or whether a change in the mitochondrial genome provided the selective advantage for the maintenance of a new mitochondrial protein. The former scenario is effectively constructive neutral evolution (CNE) (89): a new protein that initially has no selective advantage allows the accumulation of mutations that would be damaging without that protein and after the accumulation of those mutations the protein does confer a selective advantage. In a CNE scenario, one can hardly argue that mitochondrial genome evolution drives the origin of new proteins. Irrespective of the exact order of events, in Table 1 there are also proteins, like POLG2 and TEFM, which have a known function associated with mtDNA or mtRNA, but the reason why only Metazoa need these new proteins remains unclear. At the very least, the coincidence of the arrival of new mitochondrial proteins with changes in the mtDNA can provide hypotheses about the functions of those proteins and therewith drive experimental tests.

ACKNOWLEDGEMENTS

Authors thank Dennis Lavrov for his constructive comments on the earlier versions of the manuscript. In addition, authors also thank the referees, among which Jeremy Wideman, for their comments. This work was supported by the Radboud Institute for Molecular Life Sciences; Radboudumc (R0002792).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Sofou, K., De Coo, I. F., Isohanni, P., Ostergaard, E., Naess, K., et al. (2014) A multicenter study on Leigh syndrome: disease course and predictors of survival. Orphanet J. Rare Dis. 9, 52.

[2] Fry, M., and Pudney, M. (1992) Site of action of the antimalarial hydroxy-naphthoquinone, 2-[trans-4-(4-chlorophenyl) cyclohexy1]-3-hydroxy-1,4-naphthoquinone (566C80). Biochem. Pharmacol. 43, 1545–1553.

[3] Gray, M. W., Burger, G., and Lang, B. F. (1999) Mitochondrial evolution. Science 283, 1476–1481.

[4] Martijn, J., Visser, J., Guy, L., Offre, P., and Ettema, T. J. G. (2018) Deep mitochondrial origin outside the sampled alphaproteobacteria. Nature 557, 101–105.

[5] Kurland, C. G., and Andersson, S. G. (2000) Origin and evolution of the mitochondrial proteome. Microbiol. Mol. Biol. Rev. 64, 786–820.

[6] Gray, M. W., Lang, B. F., and Burger, G. (2004) Mitochondria of protists. Annu. Rev. Genet. 38, 477–524.

[7] Smith, D. R., and Keeling, P. J. (2015) Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. Proc. Natl. Acad. Sci. U.S.A. 112, 10177–10184.

[8] Karlberg, O., Canbäck, B., Kurland, C. G., and Andersson, S. G. E. (2000) The dual origin of the yeast mitochondrial proteome. Yeast (Chichester, England) 17, 170–187.

[9] Race, H. L., Herrmann, R. G., and Martin, W. (1999) Why have organelle retained genomes? Trends Genet. 15, 364–370.

[10] Bjorkholm, P., Harish, A., Hagstrom, E., Ernst, A. M., and Andersson, S. G. (2000) Origin and evolution of the mitochondrial proteome. Trends Genet. 15, 364.

[11] Chaban, Y., Boekema, E. J., and Dudkina, N. V. (2014) Structures of mitochondrial oxidative phosphorylation supercomplexes and mechanisms for their stabilisation. Biochim. Biophys. Acta 1837, 418–426.

[12] Imamura, H., Huynh Nhat, K. P., Togawa, H., Salto, K., Iino, R., et al. (2009) Visualization of ATP levels inside single living cells with fluorescence resonance energy transfer-based genetically encoded indicators. Proc. Natl. Acad. Sci. U.S.A. 106, 15651–15656.

[13] Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., et al. (1981) Sequence and organization of the human mitochondrial genome. Nature 290, 457–465.

[14] Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M., et al. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. 23, 147.

[15] Huynen, M. A., Duarte, I., and Szklarczyk, R. (2013) Loss, replacement and gain of proteins at the origin of the mitochondria. Biochim. Biophys. Acta 1827, 224–231.

[16] Satoh, M., and Kuroiwa, T. (1991) Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. Exp. Cell Res. 196, 137–140.

[17] Spielbrink, J. N., Li, F. Y., Tiranti, V., Nikali, K., Yuan, Q. P., et al. (2001) Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria. Nat. Genet. 26, 223–231.

[18] Alam, T. I., Kanki, T., Muta, T., Ukaji, K., Abe, Y., et al. (2003) Human mitochondrial DNA is packaged with TFAM. Nucleic Acids Res. 31, 1640–1645.

[19] Rajala, N., Gerhold, J. M., Martinsson, P., Klymov, A., and Spielbrink, J. N. (2014) Replication factors transiently associate with mtDNA at the mitochondrial inner membrane to facilitate replication. Nucleic Acids Res. 42, 952–967.

[20] Korhonen, J. A., Pham, X. H., Pellegrini, M., and Falkenberg, M. (2004) Reconstitution of a minimal mtDNA replisome in vitro. EMBO J. 23, 2423–2429.

[21] Ojala, D., Montoya, J., and Attardi, G. (1981) tRNA gain of proteins at the origin of the mitochondria. Biochim. Biophys. Acta 1827, 224–231.

[22] Holzmann, J., Frank, P., Löffler, E., Bennett, K. L., Gerner, C., et al. (2008) RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. Cell 135, 462–474.

[23] Brzezniak, L. K., Bijata, M., Szczesny, R. J., and Stepien, P. P. (2011) Involve- ment of human ELAC2 gene product in 3' end processing of mitochondrial tRNAs. RNA Biol. 8, 616–626.

[24] Sanchez, M. I., Mercer, T. R., Davies, S. M., Shearwood, A. M., Nygard, K. K., et al. (2011) RNA processing in human mitochondria. Cell Cycle (Georgetown, Tex) 10, 2904–2916.

[25] Antonicka, H., Sasarman, F., Nishimura, T., Pape, V., and Shoubridge Eric, A. (2013) The mitochondrial RNA-binding protein GRSF1 localizes to RNA granules and is required for posttranscriptional mitochondrial gene expression. Cell Metab. 17, 386–398.

[26] Jourdain, A., Koppen, M., Wydro, M., Rodley Chris, D., Lightowlers Robert, N., et al. (2013) GRSF1 regulates RNA processing in mitochondrial RNA granules. Cell Metab. 17, 399–410.
[27] Duarte, I., Nabuurs, S. B., Magnuo, R., and Huynen, M. (2012) Evolution and diver- 
sification of the organellar release factor family. Mol. Biol. Evol. 29, 3487–3512.
[28] Lavrov, D. V. (2007) Key transitions in animal evolution: a mitochondrial DNA 
perspective. Integr. Comp. Biol. 47, 734–743.
[29] Lavrov, D. V., and Pett, W. (2016) Animal mitochondrial DNA as we do not 
know it: mt-genome organization and evolution in nonbilaterian lineages. 
Genome Biol. Evol. 8, 2896–2913.
[30] Fonseca, M. M., Harris, D. J., and Posada, D. (2014) The inversion of the con- 
Control region in three mitogenomes provides further evidence for an asymmet-
ric model of vertebrate mtDNA replication. PLoS One 9, e106654.
[31] Dobzhansky, T. (1973) Nothing in biology makes sense except in the light of 
Dobzhansky. Mol. Biol. Evol. 30, 310–314.
[32] Greber, B. J., and Ban, N. (2016) Structure and function of the mitochondrial 
rnase. Annu. Rev. Biochem. 85, 103–132.
[33] Fonseca, M. M., Posada, D., and Harris, D. J. (2008) Inverted replication of 
vertebrate mitochondria. Mol. Biol. Evol. 25, 805–808.
[34] Birrell, J. A., and Hirst, J. (2010) Truncation of subunit ND2 disrupts the 
threefold symmetry of the antiporter-like subunits in complex I from higher 
metazoans. FEBS Lett. 584, 4247–4252.
[35] Schreiber, F., Patricio, M., Muffato, M., Pignatelli, M., and Bateman, A. (2014) 
TreeFam v9: a new website, more species and orthology-on-the-fly. Nucleic 
Acids Res. 42, D922–D925.
[36] Szklarzyk, R., and Huynen, M. A. (2009) Expansion of the human mitochon-
Driale proteome by intra- and inter-compartmental protein duplication. 
Genome Biol. 10, R135.
[37] Calvo, S. E., Clauser, K. R., and Mootha, V. K. (2015) MitoCarta2.0: an updated 
inventory of mammalian mitochondrial proteins. Nucleic Acids Res. 44(D1), 
D1251–D1257.
[38] Li, H., Coghlan, A., Ruan, J., Coin, L. J., Hériché, J.-K., et al. (2006) TreeFam: 
a curated database of phylogenetic trees of animal gene families. Nucleic 
Acids Res. 34, D572–D580.
[39] Szklarzyk, R., Wanschers, B. F., Cuypera, T. D., Esseling, J. J., Riemersma, M., 
et al. (2012) 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, R12.
[40] Elurbe, D. M., and Huynen, M. A. (2016) The origin of the supernumerary 
subunits and assembly factors of complex I: a treasure trove of pathway evo-
lution. Biochim. Biophys. Acta 1857, 971–979.
[41] Jourdain, A. A., Popow, J., de la Fuente, M. A., Martinou, J. C., Anderson, P., 
et al. (2017) The FASTK family of proteins: emerging regulators of mitochon-
drial RNA biology. Nucleic Acids Res. 45, 10941–10947.
[42] Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein 
database search programs. Nucleic Acids Res. 25, 3389.
[43] Schreiber, F., Patricio, M., Muffato, M., Pignatelli, M., and Bateman, A. (2014) 
TreeFam v9: a new website, more species and orthology-on-the-fly. Nucleic 
Acids Res. 42, D922–D925.
[44] Kazak, L., Reyes, A., Duncan, A. L., Rorchbach, J., Wood, S. R., et al. (2013) 
Alternative translation initiation augments the human mitochondrial prote-
one. Nucleic Acids Res. 41, 2354–2369.
[45] Ju, Y. S., Alexandrov, L. B., Gerstung, M., Martincorena, I., Nik-Zainal, S., 
et al.; ICGC Breast Cancer Group, ICGC Chronic Myeloid Disorders Group, 
ICGC Prostate Cancer Group. (2014) Origins and functional consequences of 
somatic mitochondrial DNA mutations in human cancer. eLife 3, 1–28.
[46] Kash, J. C., Cunningham, D. M., Smit, M. W., Park, Y., Fritz, D., et al. (2002) 
Selective translation of eukaryotic mRNAs: functional molecular analysis of 
GRSF-1, a positive regulator of influenza virus protein synthesis. J. Virol. 76, 
10417–10426.
[47] Schaub, M. C., Lopez, S. R., and Caputi, M. (2007) Members of the heteroge-
neous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splic-
ing substrate by promoting formation of ATP-dependent spliceosomal 
complexes. J. Biol. Chem. 282, 13617–13626.
[48] Ufer, C., Wang, C. C., Fahling, M., Schiebel, H., Thiele, B. J., et al. (2008) 
Translational regulation of glutathione peroxidase 4 expression through 
guanine-rich sequence-binding factor 1 is essential for embryonic brain 
development. Genes Dev. 22, 1838–1850.
[49] Song, J., Perreault, J.-P., Topisirovic, I., and Richard, S. (2016) RNA G- 
quadruplexes and their potential regulatory roles in translation. Translation 
4, e1244031.
[50] Menendez, C., Frees, S., and Bagga, P. S. (2012) GQRS-H Predictor: a web 
server for predicting homologous quadruplex forming G-rich sequence 
motifs in nucleotide sequences. Nucleic Acids Res. 40, W96–W103.
[51] Pietras, Z., Wojcik, M. A., Borowski, L. S., Szewczyk, M., Kuliński, T. M., 
et al. (2018) Dedicated surveillance mechanism controls G-quadruplex form-
ing non-coding RNAs in human mitochondria. Nat. Commun. 9, 2558.
[52] Zhang, H., Barceló, J. M., Lee, B., Kohlhagen, G., Zimonjic, D. B., et al. (2001) 
Human mitochondrial topoisomerase I. Proc. Natl. Acad. Sci. U.S.A. 98, 
10608–10613.
[53] Dalla Rosa, I., Goffart, S., Wurm, M., Wiek, C., Essmann, F., et al. (2008) Adap-
tation of topoisomerase I paralogs to nuclear and mitochondrial DNA. 
Nucleic Acids Res. 37, 6414–6428.
[54] Dalla Rosa, I., Zhang, H., Khiati, S., Wu, X., and Pommier, Y. (2017) Transcrip-
tion profiling suggests that mitochondrial topoisomerase IB acts as a topo-
logical barrier and regulator of mitochondrial DNA transcription. J. Biol. 
Chem. 292, 20162–20172.

Van Esveld and Huynen 1249
