Organelle transcriptomes: products of a deconstructed genome
Title: Organelle transcriptomes: products of a deconstructed genome

Article Type: 16/5 Genomics

Corresponding Author: Prof. Aleksandra Filipovska, PhD

Corresponding Author's Institution: The University of Western Australia

First Author: Ian D Small, PhD

Order of Authors: Ian D Small, PhD; Oliver Rackham, PhD; Aleksandra Filipovska, PhD
1. Organelle gene expression often differs from that of their prokaryotic ancestors
2. Deep sequencing has advanced the characterization of organelle transcriptomes
3. Organelle transcriptomes have evolved new features to control gene expression
Organelle transcriptomes: products of a deconstructed genome

Ian D. Small\textsuperscript{1}, Oliver Rackham\textsuperscript{2,3} and Aleksandra Filipovska\textsuperscript{2,3}

Genetic drift and mutational pressure have shaped the evolution of mitochondrial and chloroplast genomes, giving rise to mechanisms to regulate their gene expression that are often different from those in their prokaryotic ancestors. Advances in next generation sequencing technologies have enabled highly detailed characterization of organelle transcriptomes and the discovery of new transcripts and mechanisms for controlling gene expression. Here we discuss the common features of organelle transcriptomes that stem from their prokaryotic origin and some of the new innovations that are unique to organelles of multicellular organisms.

Addresses

\textsuperscript{1}Australian Research Council Centre of Excellence in Plant Energy Biology, \textsuperscript{2}Western Australian Institute for Medical Research and Centre for Medical Research and \textsuperscript{3}The School of Chemistry and Biochemistry, The University of Western Australia, Western Australia 6009, Australia.

Corresponding authors: Ian Small (ian.small@uwa.edu.au) and Aleksandra Filipovska (aleksandra.filipovska@waimr.uwa.edu.au)
**Introduction**

Mitochondria and chloroplasts have vital roles in energy production in eukaryotes, however their composition, response to energy demands and gene regulation can vary in different organisms. As descendants of the α-protobacteria and cyanobacteria, most organelle genomes share common features with prokaryotic chromosomes such as a circular structure and often their mode of replication. Although mitochondrial genomes in eukaryotic lineages show considerable diversity in size, structure and gene expression, the common evolutionary pathway that fungal and plant mitochondria shared with animal mitochondria over hundreds of millions of years is still visible in many shared, derived characteristics. Gene content is broadly similar even if genome size is not, and all of these genomes rely on phage-type RNA polymerases for expression. Generally, mitochondrial genomes encode proteins that are components of oxidative phosphorylation (OXPHOS) complexes as well as components necessary for protein translation including rRNAs and tRNAs, and some ribosomal proteins in fungi, plants and protists. Similarly chloroplast genomes encode proteins involved in photosynthesis as well as ribosomal proteins, tRNAs and rRNAs required for their synthesis.

A major feature of the mitochondrial gene expression systems compared with their bacterial ancestors is the tendency towards uncoupling transcription and translation. Organelle mRNAs tend to have much longer half-lives (measured in hours, not minutes) and thus vary in abundance much less quickly than bacterial mRNAs. As an unavoidable consequence of this, transcriptional control of the expression of specific genes is less effective and post-transcriptional control of translation initiation is relatively more effective. Hence, perhaps not surprisingly, organelles appear to
possess very few transcription factors but a variety of RNA-binding proteins with potential regulatory roles. Some large families of RNA-binding proteins appear to be almost entirely specific to organelles, including the pentatricopeptide repeat (PPR) protein and mitochondrial transcription termination factor (mTERF) families [1,2].

Recently next-generation sequencing (NGS) technologies have advanced the study of gene expression in organelles by providing relatively unbiased coverage of entire transcriptomes at single base resolution. Here we discuss the improved detail from NGS that has provided new insights into some of the more intricate processes inside organelles, and in particular the processes that have diverged (or appeared) over the course of evolution.

**The minimalisit mitochondrial transcriptomes of animals**

The animal mitochondrial DNAs (mtDNA) are some of the most compact circular genomes of all eukaryotes [3]. The consequence of a small animal mitochondrial genome is the dramatic reduction in non-coding regions. The genome is transcribed along the entire length of both strands of the mtDNA by the mitochondrial RNA polymerase (POLRMT) producing long polycistronic transcripts [4]. POLRMT is related to the bacteriophage family of polymerases that are different to prokaryotic and cytoplasmic RNA polymerases. Many other components of the mitochondrial transcription machinery in animals are different to those in prokaryotes. In animals the mitochondrial transcription factor A (TFAM) belongs to the group of high mobility group box proteins [4], and is not a subunit of POLRMT like the sigma factors which associate with the prokaryotic RNA polymerases to stimulate the transcription of bacterial (and chloroplast) genomes. Furthermore, TFAM is required
for mtDNA stability and maintenance within nucleoids, in addition to its role in transcription [5]. Mitochondrial transcription initiation can be affected by mitochondrial transcription factors B1 and B2 (TFB1M and TFB2M) that also bind directly to sites within the mtDNA [6], although recently it was found that TFB1M is predominantly required as a methyltransferase for the modification of mitochondrial rRNA [7]. Interestingly the mitochondrial termination factor 1 (MTERF1) that has been thought to regulate heavy strand transcription of the rRNA genes [8] has been found to block transcription of the light-strand polycistronic transcript before it reaches its own promoter preventing transcriptional interference [9], a phenomena observed in bacterial and eukaryotic nuclear transcription [10]. Components of the transcription machinery including POLRMT have been found associated with translation and protein assembly factors perhaps coupling transcription with translation and OXPHOS complex assembly, as a vestige of the close coupling of transcription and translation in bacteria [1,2,11].

The regulation of a thrifty genome

Generally animal mitochondrial genomes lack introns, instead the mitochondrial tRNA genes are arranged between rRNA and mRNA coding genes and act as “punctuation” marks to signal the processing of the polycistronic transcripts by the tRNA processing enzymes [12-15], analogous to certain precursor RNAs in their bacterial ancestors [16]. The tRNAs are processed at the 5’ end by an evolutionary distinct RNase P enzyme in mitochondria of different organisms. The animal mitochondrial RNase P is composed of three proteins that are required for its activity and it lacks a catalytic RNA [13], whereas prokaryotic and nuclear RNase P in animals requires an RNA for the catalytic activity of this enzyme. An RNase P that is
a single protein known as “proteinaceous RNase P” (PRORP1) lacking an RNA component is also found in plant mitochondria and plastids [5,17]. One of the components of the animal RNase P, MRPP3, has a metallonuclease domain and is an ortholog of PRORP1. The contribution of the other two proteins within the metazoan mitochondrial RNase P complex is unclear. NGS data and in vitro studies have shown that one of these two proteins, a tRNA methyltransferase 10 C, can methylate both adenosine and guanosine purines of mitochondrial tRNAs at position 9 [6,14,18]. It remains to be determined why only animal mitochondria have retained tRNA methylation at position 9 from the ancestral host but have adapted it to modify both purines.

Cleavage of the 3′ end of tRNAs is carried out by the mitochondrial RNase Z in eukaryotes. In animals this enzyme is encoded by the ELAC2 gene and its mRNA can produce two distinct proteins by using two different start codons, the longer protein including a mitochondrial targeting sequence while the shorter protein lacks this targeting sequence and is localized to the nucleus [7,19]. Although the mitochondrial RNase Z shares homology with the bacterial RNase Z, the eukaryotic enzyme contains two elaC domains instead of the single one found in bacteria [8,14,19].

Processing of some mRNAs that are not adjacent to tRNA genes requires RNase P, PTCD2 or G-rich sequence binding factor 1 (GRSF1), likely as a result of the tRNA-like secondary structures that form ahead of the 5′ ends of these mRNAs [9,14,20-22], analogous to an observed RNase P cleavage of a bicistronic mRNA not separated by tRNAs in bacteria [10,23].
The regulation of animal mitochondrial transcripts has evolved to cope with its compact genome. For example polyadenylation at the 3′ end of mRNAs is often an indispensable part of the maturation for most mitochondrial mRNAs that use it to complete their termination codons [24]. Polyadenylation of animal mRNAs can be important for RNA stability and translation, although in some cases it can be a signal for degradation of specific transcripts [24]. Specific nucleotides on the rRNAs and tRNAs are modified by methyltransferases such as TFM2B and tRNA modifying enzymes to enable proper folding and recognition by the translation machinery [7,25]. Furthermore, CCA, which is typically genomically encoded at the 3′ end of bacterial tRNAs, is added at the 3′ end of mitochondrial tRNAs by a tRNA nucleotidyltransferase to enable attachment of amino acids and the interactions with translation factors and aminoacyl-tRNA synthetases [25].

Animal mitochondrial translation is unique in a number of respects. Animal mtDNA has an alternative genetic code that differs from that of the nucleus or other organelles found in plants. AUA is used as a start codon instead of coding for isoleucine, UGA as a tryptophan codon instead of a stop codon, and AGA and AGG terminate some open reading frames and were thought to be termination codons instead of coding for arginine [26]. However recently it was shown that the AGA/AGG codons actually contribute to a -1 frameshift that enables the conventional UAG stop codon to be recognized by the mitochondrial release factor a (mtRF1a) within the mitochondrial ribosome and to terminate protein synthesis [27].

Mitochondrial ribosomal RNA has reduced in size considerably during evolution and has been replaced by additional proteins. Many of these additional ribosomal proteins
do not replace the missing RNA helices but instead have unique positions, decorating the exterior of the mitochondrial ribosome [28]. These proteins have likely evolved new functions dictated by the unconventional characteristics of mitochondrial mRNAs. Mitochondrial ribosomes are a patchwork of proteins that share homology with ribosomal proteins from prokaryotes and additional proteins that are unique to mitochondrial ribosomes [28]. In addition to splicing and processing enzymes, there are specific factors which recognize the 5′ UTRs of their cognate mRNAs and regulate the stability or translation of individual mRNAs in yeast [29]. However in animals only the translational activator of cytochrome c oxidase subunit 1 (TACO1) has been identified as an mRNA specific regulator to date [30]. It may be that some of the additional proteins unique to mitochondrial ribosomes in animals have roles in translation of specific transcripts.

Although produced from polycistronic transcripts, the wide variation in mature RNA abundance, processing and polyadenylation [31], indicates that mitochondria have adopted the increased dependence on post-transcriptional regulation typical of the eukaryotic cytoplasm (reviewed in [32]). This variation is particularly evident in different disease states where mitochondrial gene expression is often increased as a compensatory response to decreased energy metabolism. Recently it has become clear that RNA-binding proteins can associate within granules in specific regions inside mitochondria where they have specific and often indispensible roles in RNA metabolism [21,22,33].

**Plant organelle transcriptomes**
In plants, there has been a major additional source of genetic complexity with the incorporation of chloroplasts. The transition from free-living cyanobacteria to organelle proceeded along a similar route to that of the mitochondrial ancestor, with massive transfer of genes to the nucleus and loss of the most of the rest. The resulting genome, like that of mitochondria, codes for only a few tens of key organelle components, plus the expression system needed to make them. As for mitochondria, mRNAs have an extended half-life and translation is largely uncoupled from transcription; although there is some transcriptional control of gene expression [34], the major control points appear to be post-transcriptional [35,36]. There has been considerable exchange between the chloroplast and mitochondrial gene expression systems. They now share RNA polymerases, many RNA-binding protein families, tRNAs and aminoacyl-tRNA synthetases. There are multiple examples of insertion of chloroplast DNA into the mitochondrial genome, leading to functional transfer of tRNAs and RNA processing signals [37,38]. Both organelles carry out essential and extensive RNA editing using homologous factors [39].

Most plant organelle primary transcripts are far longer than the processed mRNAs that are actually translated. Active endo- and exo-nucleases rapidly degrade most non-coding RNAs, divide most polycistronic RNAs into monocistronic RNAs and trim the termini. The machinery that does this appears to be largely inherited from the bacterial ancestors with recognisable RNases E, J, R, Z and polynucleotide phosphorylase all present in plant organelles. Like in bacteria, but unlike in animal mitochondria or the cytosol, the addition of a poly(A) tail to mRNAs generally triggers RNA turnover [40,41]. NGS of partially processed RNAs from wild-type plants and comparison with those from mutants lacking one or more of these
nucleases shows just how crucial these nucleases are to correct formation of translatable RNAs [42].

An important recent discovery was the mapping of many abundant small non-coding RNAs to the exact 5’ and 3’ termini of chloroplast mRNAs [43,44]. In several cases, these small RNAs correspond exactly to the binding sites of known PPR proteins. It seems highly likely that these proteins protect RNA termini by preventing incursion of exonucleases [44], a hypothesis for which there is considerable evidence in the case of PPR10 in maize [45]. The prevalence of these small RNAs suggest that this is the primary mode for determining 5’ and 3’ termini in chloroplasts (44), and probably also operates in plant mitochondria [46]. Loss of these PPR proteins leads to loss of the corresponding mRNA [45,46]. This mechanism for protecting mRNAs (and presumably regulating their abundance) appears to be an invention of plant organelles - it will be interesting to see if analogous mechanisms can be found elsewhere.

**Junk RNA or unsuspected regulatory complexity?**

Recently RNA sequencing of the human mitochondrial transcriptome revealed the presence of small RNAs (sRNA) that comprised ~3% of the whole cell sRNA population [31]. The majority of mitochondrially encoded sRNAs are derived from immediately downstream of the 5’ or 3’ cleavage sites of tRNA genes. The expression of these two classes of sRNAs does not correlate to each other, suggesting that they are generated by independent mechanisms. It is still not clear if these sRNAs have a functional role in gene expression like nuclear sRNAs or if they are footprints of RNA-binding proteins, as it has been observed in chloroplast transcriptomes [43,44]. The high abundance of tRNA-derived sRNAs indicates that they may regulate
mitochondrial gene expression by controlling translation or RNA stability as it was found in prokaryotes and the cytoplasm of eukaryotes [47].

In addition to sRNAs, the animal mitochondrial transcriptome has been shown to include long non-coding RNAs (lncRNAs) encoded by the light strand of the mtDNA, that are regulated by mitochondrial RNA processing proteins such as RNase P, RNase Z and PTCD2 [20,31]. These lncRNAs primarily form intermolecular duplexes, suggesting that they may bind to their complementary mRNAs [20] to either stabilize them or inhibit their translation. Recently the RNA-binding protein GRSF1 was found preferentially associated with MT-ND6 and the lncRNAs ND5 and cyt b where it regulates the stability of these transcripts and consequently their expression [21]. The identification of new levels of complexity, regulation and the abundance of non-coding RNAs, previously not expected of this streamlined transcriptome, has progressed our understanding of the additional levels of posttranscriptonal control in mitochondrial gene expression in animals.

Abundant non-coding RNAs are present in plant organelles where it has long been apparent that even in the large, gene-poor plant mitochondrial genomes, much of the genome is transcribed to some extent. The first NGS studies on chloroplast transcriptomes have confirmed and extended these observations and shown just how much of the intergenic DNA is transcribed and exists as antisense transcripts [42,48,49]. Furthermore abundant antisense transcripts have been detected in the diverse bacterial species [50,51]. Although many of these RNAs are rapidly turned over, some of them accumulate to relatively significant levels. Recently specific endonucleases have been shown to play a role in the clearance of antisense transcripts.
both in plant organelles and in prokaryotes, such as RNase J in chloroplasts [52] and RNase III in bacteria [51]. There is considerable speculation as to whether any of these antisense RNAs have important functions. This would not be without precedent - trans-splicing of the psaA mRNA in Chlamydomonas reinhardtii chloroplasts requires the non-coding tscA RNA [53].

**RNA editing- variety is the spice of life**

An invention characteristic of plant organelles is RNA editing by deamination of specific cytidine residues to uridine. Although many disparate RNA editing systems exist in organelles [54], the editing in plants is virtually unique in relying on a huge family of highly specific PPR proteins, each one inducing editing of one or a few related sites [39]. Recent work on RNA editing has identified a number of other factors required besides the PPR specificity factors [55-58] and there is a reasonable expectation of mechanistically understanding this enigmatic process in the near future. NGS studies of plant organelles are revealing previously unrecognized editing sites and providing much more accurate quantification of editing activity [59]. A major breakthrough has been deciphering how editing factors recognize their targets [60,61] - this allows prediction of which factor might edit which site and will greatly accelerate research on this topic.

**Conclusion**

Advances in NGS enable researchers to obtain a complete inventory of organelle transcriptomes. Recent discoveries show how organelle transcriptomes have been forced to extensively optimize their modes of gene expression to fit their shrinking
host genomes. What remains for the future is to provide a mechanistic understanding of how gene expression is regulated in these intriguing systems.

Acknowledgements

We thank the Australian Research Council and the National Health and Medical Research Council of Australia for the award of our fellowships and grants.
References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

• of special interest

•• of outstanding interest


• This article has redefined the role of MTERF1, showing that it can block transcription of the light-strand polycistronic transcript before it reaches its own promoter preventing transcriptional interference that has been observed in bacterial and eukaryotic nuclear transcription.


- In this article the mammalian mitochondrial RNase P is identified for the first time as an evolutionary distinct protein complex that does not require a catalytic RNA, unlike previously identified nuclear and bacterial RNase P enzymes.


- This article describes the use of next generation technologies to investigate mammalian RNA processing and modification. The authors used NGS data to identify that one of the RNase P components, MRPP1, modifies the 9th nucleotide of mammalian mitochondrial tRNAs.


- This article describes the additional functional role of one of the three proteins that comprise the RNase P complex that is involved in methylation of guanine and adenine residues of mammalian mitochondrial tRNAs, that is not found in organelle tRNAs in other eukaryotes.


   • This article reports the persistence of long non-coding RNAs in mammalian mitochondria generated from the light strand of the mtDNA by tRNA processing enzymes.


   • This is one of the first reports describing the RNA-binding protein GRSF1 and its role in mitochondrial RNA processing including that of long non-coding RNAs.


   • This a parallel report on the role of the RNA-binding protein GRSF1 that also describes the presence of other RNA processing proteins in specific granules within the mitochondrial matrix in cells.


27. Temperley R, Richter R, Dennerlein S, Lightowlers RN, Chrzanowska-
Lightowlers ZM: **Hungry Codons Promote Frameshifting in Human Mitochondrial Ribosomes.** *Science* 2010, **327**:301–301.

- This is an important article which shows that the AGA/AGG codons, previously thought to act as termination codons, contribute to a -1 frameshift that enables the conventional UAG stop codon to be recognized by the mitochondrial release factor and end protein synthesis.


29. Lipinski KA, Kaniak-Golik A, Golik P: **Maintenance and expression of the S. cerevisiae mitochondrial genome--from genetics to evolution and systems biology.** *Biochim Biophys Acta* 2010, **1797**:1086–1098.


- This article presents the first maps of the human mitochondrial transcriptome at single nucleotide resolution obtained using NGS. The maps show wide variation in transcript abundance and new transcripts such as small and long non-coding RNAs revealing previously unexpected complexities from the compact mitochondrial genome.

32. Rackham O, Mercer TR, Filipovska A: **The human mitochondrial transcriptome and the RNA-binding proteins that regulate its expression.** *WIREs RNA* 2012, **3**:675–695.


- The authors in this article identify an evolutionarily conserved mechanism that relays circadian timing information between organelles with distinct genetic systems.


37. Byers E, Bonen L: **Potential role of tRNAs in wheat and Lolium mitochondrial rps7 transcript processing.** *Genome* 2012, **55**:615–621.


40. Chang JH, Tong L: **Mitochondrial poly(A) polymerase and polyadenylation.** *Biochim Biophys Acta* 2012, **1819**:992–997.


42. Hotto AM, Schmitz RJ, Fei Z, Ecker JR, Stern DB: **Unexpected Diversity of Chloroplast Noncoding RNAs as Revealed by Deep Sequencing of the Arabidopsis Transcriptome.** *G3 (Bethesda)* 2011, **1**:559–570.

   • This article presents a detailed survey of chloroplast non-coding RNAs using NGS to reveal that these transcripts are generated by distinct combination of ribonucleases or transcriptional read through.


   •• The authors identify the abundance of short non-coding RNA in chloroplasts which are likely footprints of the PPR RNA-binding family of proteins.


   • The authors in this article identify that the regions protected by RNA-binding proteins are those that are the most processed mRNA termini in chloroplasts.

A detailed study of the maize PPR10 protein identifies its dual role in defining the 5’ ends of chloroplast transcripts and facilitating their translation - providing an archetype for these processes in chloroplasts.


This article identifies an additional factor that is required for RNA editing in mitochondria and chloroplast of plants. This emphasizes importance of additional components of the RNA editing apparatus required for its activity.


•• The authors in this article identified that editing in plants requires additional proteins and some are essential for editing in plant mitochondria. The authors show that these editing factors interact with PPR-containing editing proteins.


• This article demonstrates that NGS of plant organelles is a valuable tool to identify unrecognized editing sites and provides a more accurate measure of editing activity.


•• This is a landmark article describing the RNA-binding code of the large PPR family of proteins that have diverse roles in organelle RNA metabolism.

Figure legends

Figure 1. Similarities and differences between animal and plant mitochondrial transcriptomes. Generally mitochondrial genomes are transcribed by bacteriophage-like polymerases. Transcription of the animal mtDNA is along the entire length of the genome, whereas in plant and other eukaryotic mitochondria transcription is initiated at specific promoters throughout the genome. Mitochondrial gene expression is predominantly regulated post-transcriptionally and RNA-binding proteins play a central role in RNA metabolism and protein synthesis. Animal mitochondrial transcripts do not contain introns, however in plants introns within mitochondrial transcripts are spliced out by the maturases that they encode and essential, imported splicing factors. Processing of polycistronic transcripts where tRNAs or tRNA-like structures are interspersed between rRNA and mRNA sequences is carried out by the mitochondrial RNase P and RNase Z enzymes. In the absence of tRNA sequences, the processing of polycistronic pre-mRNAs to generate mature mRNAs requires other less-specific endonucleases. Maturation of transcripts involves exonucleolytic processing (in plant mitochondria), nucleotide modification, addition of CCA at the 3’ ends of tRNAs and polyadenylation of mRNAs for animal mitochondria. RNA editing is widespread and required in plant mitochondria for different aspects of RNA metabolism. Plant mitochondria import tRNAs from the cytoplasm but this phenomenon has not been observed in animals. The mature transcripts are translated on mitochondrial ribosomes closely associated with the inner mitochondrial membrane. Although polyadenylation of animal mitochondrial transcripts is required for 10 out of the 11 transcripts for their stability and translation, polyadenylation of plant mitochondrial transcripts is required for their degradation.
**Figure 2.** The chloroplast transcriptomes of plants. Chloroplast transcription is initiated at specific promoters throughout the genome and the most active RNA polymerase in chloroplasts is bacterial-like and associates with sigma factors. RNA-binding proteins play a central role in RNA metabolism and protein synthesis of chloroplast gene expression. The splicing of introns within chloroplast transcripts is carried out by the maturases that they encode and essential, post-translationally imported splicing factors. Processing of polycistronic transcripts is similar to that in plant mitochondria, involving RNase P and RNase Z as well as additional endonucleases. Maturation of chloroplast transcripts involves exonucleolytic processing, nucleotide modification, addition of CCA at the 3’ ends of tRNAs and RNA editing of different transcripts, although this process is more predominant in plant mitochondria. Unlike for plant mitochondria, import of tRNAs from the cytoplasm has not been observed in chloroplasts. The mature chloroplast transcripts are translated on ribosomes that closely associated with the thylakoid membrane. Chloroplast transcripts are polyadenylated as part of the degradation process.
This piece of the submission is being sent via mail.