

# Expression profiling of *Drosophila* mitochondrial genes via deep mRNA sequencing

Tatiana Teixeira Torres<sup>1,\*</sup>, Marlies Dolezal<sup>1</sup>, Christian Schlötterer<sup>1</sup> and Birgit Ottenwälder<sup>2</sup>

<sup>1</sup>Institut für Populationsgenetik, Veterinärmedizinische Universität Wien, Vienna, Austria and <sup>2</sup>Eurofins Medigenomix GmbH, Martinsried, Germany

Received July 13, 2009; Revised and Accepted September 24, 2009

## ABSTRACT

Mitochondria play an essential role in several cellular processes. Nevertheless, very little is known about patterns of gene expression of genes encoded by the mitochondrial DNA (mtDNA). In this study, we used next-generation sequencing (NGS) for transcription profiling of genes encoded in the mitochondrial genome of *Drosophila melanogaster* and *D. pseudoobscura*. The analysis of males and females in both species indicated that the expression pattern was conserved between the two species, but differed significantly between both sexes. Interestingly, mRNA levels were not only different among genes encoded by separate transcription units, but also showed significant differences among genes located in the same transcription unit. Hence, mRNA abundance of genes encoded by mtDNA seems to be heavily modulated by post-transcriptional regulation. Finally, we also identified several transcripts with a noncanonical structure, suggesting that processing of mitochondrial transcripts may be more complex than previously assumed.

## INTRODUCTION

Mitochondrial function is vital for a handful of cellular processes such as ATP production, oxidation of fatty acids, biosynthesis of amino acids and signal transduction. Proteins involved in such processes are encoded in the nuclear genome and also in its own organellar genome. Animal mitochondrial DNA (mtDNA) is a compact molecule typically encoding 13 protein-coding genes (PCGs) involved in oxidative phosphorylation and a subset of the translation machinery components, namely, 2 ribosomal RNA subunits (rRNAs) and 22 transfer RNAs (tRNAs) (1). There is only one major noncoding

region in the mtDNA, the control region, also known as the displacement loop (D loop) or A+T-rich region (in insects). This portion of the mtDNA has been shown to contain the replication origin (2) and promoters for transcription initiation (3).

The animal mitochondrial genome has some features that are different from those of the nuclear genome. Owing to the compact structure of the mtDNA, polypeptide, tRNA and rRNA genes are smaller than the cytosolic and prokaryotic counterparts (4), there are overlapping genes, some termination codons are not encoded in mtDNA sequence, there are no intronic sequences and almost no intergenic sequence and transcripts lack untranslated regions (UTRs) completely (1,3). These structural features of the mtDNA determine its unique transcriptional system. RNA synthesis in *Drosophila* mtDNA starts at five different transcription initiation sites, two on the heavy (H) strand and three on the light (L) strand. The tRNA sequences of the five polycistronic transcripts acquire the cloverleaf structure and act as signals for the cleavage of these primary transcripts. This model of RNA processing is known as the tRNA punctuation model (5). Cleaved mRNAs correspond to the mature transcripts. They are mono or bicistronic with two overlapping PCGs, and carry 50- to 60-bp poly(A) tails (6,7). The mature mitochondrial mRNAs start directly at the initiation codon or have an extremely short untranslated 5'-end (1–3 nt). Furthermore, they do not contain recognizable ribosome binding sites and are, apparently, translated from the first initiation codon at the 5'-end (6,7).

For more than two decades, significant effort has gone into random sequencing of cDNA clones for gene discovery and annotation of genomes. Now, the development of next-generation sequencing (NGS) technologies has permitted an increase in throughput over traditional sequencing methods. Parallel sequencing of short cDNA fragments has been demonstrated as an excellent tool to generate genome-wide sequence information as well as levels of gene expression (8). Although the analysis

\*To whom correspondence should be addressed. Tel: +55 19 35211141; Fax: +55 19 35211089; Email: ttorres@unicamp.br

Present address:

Tatiana Teixeira Torres, Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, CP 6010, 13083-875 Campinas, SP, Brazil.

of expressed sequence tags (ESTs) is employed primarily to gather information about nuclear transcripts, a remarkably high proportion of sampled cDNAs corresponds to mitochondrial transcripts (9–11). Such ESTs are usually neglected, but they can be extremely useful to collect information on mitochondrial transcription in organisms for which the mitochondrial genomic sequence is already known, or even to obtain data on mitochondrial gene organization and expression in organisms in which mtDNA sequence data are unavailable (12).

Here, we report the analysis of ESTs generated by 454 sequencing technology (FLX) for *Drosophila melanogaster* and *D. pseudoobscura* mitochondrial genomes. We show that NGS could provide important insights into the expression pattern of genes encoded by mtDNA and their conservation between species. Furthermore, we discovered some rare transcripts that might play a role in mitochondrial function.

## MATERIALS AND METHODS

### *Drosophila* strains

Experimental flies of each species were obtained from reciprocal crosses between the genome strain and a second inbred strain, all obtained from the Tucson Stock Center (stock numbers 14021-0231.36 and 14021-0231.15 for *D. melanogaster*; 14011-0121.94 and 14011-0121.88 for *D. pseudoobscura*). Flies were grown at 20°C in standard cornmeal medium. Newly emerged males and females were collected separately and allowed to age from 3 to 7 days.

### Library preparation and sequencing

Methods for shotgun library preparation and 454 sequencing were previously described (8,13). Briefly, total RNA was extracted with Trizol (Invitrogen) from 30 flies (three replicates of 10 flies each). RNA samples were treated with DNase I (10 U/50 mg of total RNA) and the absence of contaminating genomic DNA was confirmed by PCR.

First-strand cDNA was generated from ~5 mg of total RNA using the RevertAid™ H Minus First Strand cDNA Synthesis Kit (Fermentas) according to manufacturer's instructions. The synthesis was carried out using a biotinylated oligo(dT) fused to the 454 sequencing primer B [5'-Biotin-GCCTTGCCAGCCCGCTCAG(T)<sub>17</sub>V-3', where V stands for any base but T]. Double-stranded cDNA was synthesized by addition of 30 U of *Escherichia coli* DNA polymerase I, and 1 U of *E. coli* ribonuclease H to the first-strand synthesis reaction, following Fermentas's suggested protocol for second-strand cDNA synthesis.

Approximately 5 mg of double-stranded cDNA was nebulized using 3 bar nitrogen for 1 min. The 3'-nebulized fragments were recovered using M-270 Streptavidin beads (Dyna), blunt-ended with T4 DNA polymerase and ligated to the double-stranded linker A.

To reduce the technical error, all four libraries (two sexes × two species) were prepared in duplicates and pooled prior to sequencing. Sequencing was performed

on a Genome Sequencer FLX Instrument (Roche Diagnostic) following standard protocols (14).

### Assembly and annotation of *D. pseudoobscura* mtDNA

*Drosophila melanogaster* mitochondrial genes (release 5.6) were BLASTed against *D. pseudoobscura* whole genome shotgun (WGS) sequences (release 2.2), both available from FlyBase (<http://flybase.bio.indiana.edu/>). Sequences that resulted in significant hits against *D. melanogaster* mitochondrial genes were assembled using the Cap3 program (15) and carefully checked using EagleView (16) to generate the whole mtDNA sequence of *D. pseudoobscura* (Supplementary Table S1).

Two contigs were generated and there were two gaps relative to the *D. melanogaster* mitochondrial genome. The first gap corresponded to ~550 bp of *cox1* gene and the second to five genes, *nd1*, tRNA<sup>Leu</sup>, *rrnL*, tRNA<sup>Val</sup> and *rrnS* (~3300 bp). The *D. pseudoobscura* mtDNA (with exception of the control region) was completed with three mitochondrial sequences deposited in GenBank (accession numbers EU493633, EU494363 and EU494484) and with contigs from the *de novo* assembly of *D. pseudoobscura* 454 reads. One of the WGS contigs contained 648 bp of a noncoding sequence that is most likely a portion of the control region. This portion was not included in the final assembly. The 454 reads were mapped to the assembly, and inconsistencies such as premature stop codons, long insertions and deletions were corrected to generate the final assembly.

The PCGs, rRNA and tRNA genes were identified and annotated using DOGMA (17).

### EST to mtDNA and transcript mapping

Alignment of the 454 ESTs with annotated mitochondrial transcripts was performed with PanGEA (18). We used the 5.5 release of transcript annotations of *D. melanogaster* as a reference dataset, correcting the transcripts for the genes *atp6*, *atp8*, *nd4* and *nd4L* that were not annotated as bi-cistronic transcripts. For *D. pseudoobscura*, we used the assembled and annotated mitochondrial genome. Mapping of the 454 ESTs was also performed against the whole mtDNA of both species using PanGEA (18). Only reads with at least 95% identity with a sequence in the database (genome or transcripts) over at least 60% of their length were considered.

For the mitochondrial genome-wide discovery of chimeric reads, we BLASTed all mitochondrial reads against the reference database containing full-length mitochondrial transcripts. We implemented an algorithm to parse BLAST results (Perl script available upon request). First, we checked for intra-molecule chimeras by checking high-scoring segment pairs (HSPs) with different read orientation or different sequence order. We also checked inter-molecule chimeras by searching for sequences mapping to more than one transcript. For the BLAST searches, we required a minimum of 50% of the read to be involved in the first HSP, with at least 95% of identity.

To identify chimeric transcripts shared among different libraries, intra-molecule chimeras were clustered based on at least 80% sequence similarity and at least 90% length coverage using BLASTclust (I. Dondoshansky and Y. Wolf, unpublished software, available at <http://www.ncbi.nlm.nih.gov/BLAST/docs/blastclust.html>). As we have observed the same chimeric EST in multiple-sequence reads from the same library as well as in independent libraries from different species, we considered this to be sufficient evidence for presence of the chimeric molecules, and additional reverse transcription-polymerase chain reaction (RT-PCR) experiments were not performed.

### Statistical analyses

We used the number of ESTs mapping to a given transcript as measure of the expression level for that gene. Expression levels were normalized by dividing the EST counts for each transcript by the total number of reads in a library. In a second step, these relative counts were log transformed to meet assumptions for the linear models applied. The linear models included sex, species, gene and all possible two-way interactions between these fixed effects. Least square means were obtained with the *lsmeans* option of *proc GLM* in the software package SAS 9.1.3 (SAS Institute Inc., Cary, NC, USA). *P*-values were Bonferroni corrected for multiple testing (19). To answer the question whether patterns of gene expression differ among genes within the same transcriptional unit as well as among genes interacting in the same protein complex, we fitted all pairwise contrasts for genes within a unit. *P*-values were again Bonferroni corrected (19). The level of significance was set to  $P < 0.05$ . All statistical analyses were performed with SAS 9.1.3 (SAS Institute Inc., Cary, NC, USA).

### Data availability

The EST sequences and the gene expression profiles have been deposited in the GEO repository under the accession number GSE16651 and the *D. pseudoobscura* mitochondrial genome has been deposited in GenBank under the accession number FJ899745.

## RESULTS

### Assembly and annotation of *D. pseudoobscura* mtDNA

To ensure a reliable mapping and identification of ESTs prior to the analyses of the 454 reads, we assembled the full mitochondrial genome of *D. pseudoobscura* (with the exception of the control region). The mtDNA of *D. pseudoobscura* had the identical gene composition and gene order as *D. melanogaster* and the average sequence divergence between these two species ranged from 6.9% to 14.3% for PCGs and it was 5.1 and 5.9% for the large and small rRNA genes (Supplementary Table S2).

The published *D. pseudoobscura* nuclear genome was generated by WGS sequencing of DNA prepared from embryonic nuclei and therefore mitochondrial sequences were not expected in the library (20,21). Nonetheless, unknown singletons were found in the genome sequence

**Table 1.** Summary statistics for 454 ESTs sequencing and mapping to *D. melanogaster* and *D. pseudoobscura* mitochondrial genomes

	Total	Mapped to mitochondria <sup>a</sup>	Mt rRNA	Protein coding
<i>Drosophila melanogaster</i>				
Male library	464 253	76 316	47 714	28 602
Female library	68 919	7873	5236	2637
<i>Drosophila pseudoobscura</i>				
Male library	37 994	7962	5707	2255
Female library	41 265	6515	4943	1572

<sup>a</sup>ESTs that had at least 95% identity with the mitochondrial genome at least 60% of their length.

that contained blocks of complete genes with high sequence and structure conservation relative to *D. melanogaster* and *D. yakuba* mtDNA. In total, six unknown singletons and three unknown groups produced significant hits to two or more *D. melanogaster* mitochondrial genes.

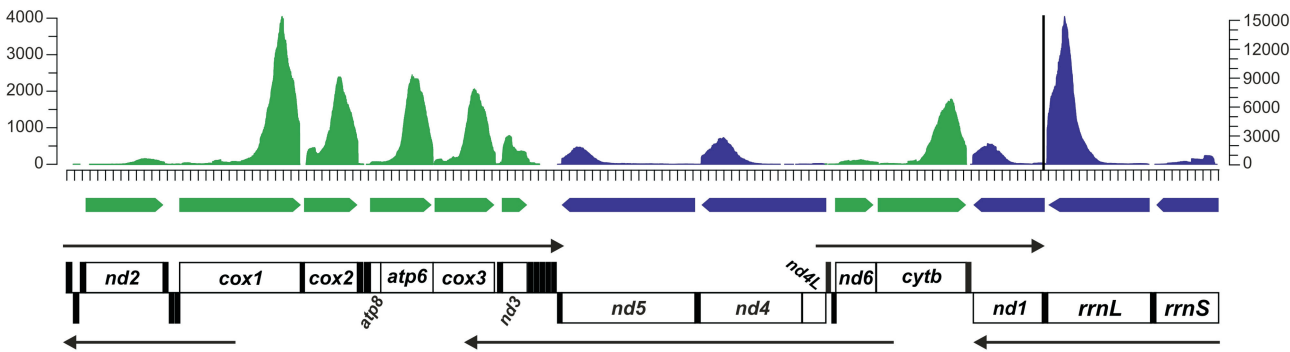
### Mapping of mitochondrial 454 ESTs

We previously showed that NGS of 3'-ESTs provides an excellent tool for expression profiling (8). We used this strategy to obtain ESTs from *D. melanogaster* and *D. pseudoobscura*. We took advantage of the massive number of ESTs from mtDNA, to explore aspects related to the transcription of these genes.

The reads identified as mitochondrial-derived ESTs were mapped to the mitochondrial genome (Table 1, Figure 1). Due to the selection of 3'-fragments, associated with the under-representation of long transcripts intrinsic to the 454 sequencing method (8), the distribution of ESTs was skewed towards the 3'-end of all transcripts and the complete 5'-end was not recovered for all genes (Figure 1). ESTs were recovered for all PCGs and for the two ribosomal genes in both species. A number of ESTs showed a poly(A) stretch, starting from 1 to 18 bp downstream of the stop codon of a PCG (Supplementary Table S3), but in *cox2* and *nd5* (for *D. melanogaster* and *D. pseudoobscura*), *nd4* (*D. melanogaster*) and *nd2* (*D. pseudoobscura*), the poly(A) tail completes the TAA stop codon from T or TA codons in the mtDNA. When the primary transcripts containing these genes are processed, the removal of an adjacent tRNA gene leaves an inframe T or TA. Post-transcriptional polyadenylation is the most likely mechanism for the creation of a complete TAA stop codon for these genes (5). For the *nd1* gene, in both species, the analysis of the polyadenylated ESTs also allowed the identification of two new molecules, probably due to indels in the stop codon on the mtDNA sequence. In one of these molecules, an insertion changed the stop codon from TAG to TAA. There was also a rare transcript (<6 % of the *nd1* transcripts sampled in this site) in which a single nucleotide deletion destroyed the stop codon.

Two cases of overlapping genes, *nd4* + *nd4L* and *atp6* + *atp8*, expressed as transcripts with no poly(A) tails downstream of the first open reading frame (ORF) confirmed the presence of the two mature bi-cistronic





**Figure 1.** Mapping of mtESTs from PCGs (left of the solid black line) and from the ribosomal subunits (rRNAs) (right of the solid black line) of *D. melanogaster* to the mitochondrial genome. PCGs and rRNA genes are represented by the labeled boxes and tRNAs by the black unlabeled boxes. Green and blue boxes represent the mature transcripts of genes in the heavy and light strand, respectively. The arrows represent the transcription units described for *Drosophila* (35) and their direction of transcription. Coverage per nucleotide position of the mapped 454 reads is shown by the histograms (in green and blue).

transcripts. The reading frames of these genes overlap by 1 (*nd4* + *nd4L*) or 7 nt (*atp6* + *atp8*) in both species. Therefore, we mapped 11 different transcripts corresponding to the 13 PCGs and two different transcripts corresponding to the ribosomal genes.

Expression profiling of ribosomal genes

The number of reads that mapped to each transcript is proportional to the transcript abundance, so it can be used as a measure of the absolute level of expression of that transcript (8). The gene expression measure as the absolute EST counts and the expression relative to the library size for each mitochondrial gene in each library is displayed in Supplementary Table S4, and Table 2, respectively. Of all mtESTs from our nonnormalized libraries, ~60% originate from the large mitochondrial ribosomal subunit in *D. melanogaster* and >70% in *D. pseudoobscura*.

In humans, the termination factor mTERF binds mtDNA immediately downstream of the 3'-end of the ribosomal gene cluster and it has been shown that the ligation of the mTERF is responsible for an attenuation/termination event that causes the steady-state level of rRNAs to be higher than that of the downstream mRNAs (22). In *Drosophila*, the mitochondrial termination factor DmTTF is thought to influence the abundance of mitochondrial transcripts in a similar manner, regulating the expression levels of the mitochondrial transcripts (23). The abundance of both ribosomal subunits should be equal, since both genes are co-transcribed as a single primary transcript. Nevertheless, we observed a pronounced difference in sequence reads for both ribosomal subunits. The large subunit of the mitochondrial rRNA, as well as the protein-coding mitochondrial genes, have the typical, long poly(A) stretch, while the short subunit has a very short poly(A) tail (approximately 6-nt long). Given that the length of the poly(A) tail differs substantially among the mature transcripts of both genes, we reasoned that the exceptionally short poly(A) tail of the small subunit prevented an efficient cDNA synthesis using poly(T) oligos.

**Table 2.** Expression profiles of *D. melanogaster* and *D. pseudoobscura* mitochondrial PCGs and rRNA genes

	<i>Drosophila melanogaster</i>		<i>Drosophila pseudoobscura</i>		<i>Drosophila melanogaster</i> Berthier et al. (35) <sup>a</sup>
	Female	Male	Female	Male	
<i>atp6/atp8</i> <sup>b</sup>	7.9	9.1	10.0	14.3	4.0
<i>cytb</i>	4.8	7.7	3.7	4.5	2.5
<i>cox1</i>	9.7	16.4	6.8	13.0	5.5
<i>cox2</i>	5.8	9.1	4.6	7.0	5.0
<i>cox3</i>	4.5	8.5	5.1	9.6	7.5
<i>nd1</i>	1.4	2.3	1.9	1.8	1.5
<i>nd2</i>	0.3	0.6	1.6	1.3	1.5
<i>nd3</i>	0.9	1.9	0.8	2.2	1.5
<i>nd4/nd4L</i> <sup>b</sup>	2.0	3.4	1.6	3.3	1.5
<i>nd5</i>	0.8	2.1	1.7	1.8	1.5
<i>nd6</i>	0.1	0.6	0.2	0.6	0.0
<i>rrnL</i>	70.1	97.2	119.5	148.6	10.0
<i>rrnS</i>	5.8	5.6	0.3	1.6	10.0

Expression profiles are presented as relative proportion of ESTs mapped to each gene to total reads obtained.

<sup>a</sup>Steady-state levels of mtRNAs (35).

<sup>b</sup>Bi-cistronic transcripts.

Consequently, we excluded the two ribosomal genes from our further analyses.

Expression profiling of PCGs

Our overall model was highly significant ( $F = 34.87$ ,  $P < 0.0001$ ) and explained 99% of the variation in gene expression. As expected, most of the variation is among individual genes ( $F = 101.44$ ,  $P < 0.0001$ ). We found a significant effect of sex but not species (sex,  $F = 66.89$ ,  $P < 0.0001$ ; species,  $F = 3.05$ ,  $P = 0.1116$ ) on overall gene expression, with males showing a higher transcriptional activity than females. Among all possible two-way interactions, only the interaction between genes and species was significant ( $F = 3.88$ ,  $P = 0.0216$ ). The interaction between sex and species can be considered as a correction for overall differences between the libraries including technical error and was not significant. This is very promising, as the lack of duplicates or replicates did

**Table 3.** Pairwise comparison of the levels of gene expression between genes in the same transcriptional unit

Transcriptional unit	Gene pair	Mean square	F-value	P-value	Corrected P-value
1	<i>atp6/atp8</i> versus <i>cox1</i>	0.00231415	0.24	0.6366	1.00000
1	<i>atp6/atp8</i> versus <i>cox2</i>	0.07637781	7.84	0.0188	0.30111
1	<i>atp6/atp8</i> versus <i>cox3</i>	0.06836339	7.01	0.0244	0.39015
1	<i>atp6/atp8</i> versus <i>nd2</i>	2.46272246	252.65	<0.0001	<0.0001
1	<i>atp6/atp8</i> versus <i>nd3</i>	1.56069902	160.11	<0.0001	<0.0001
1	<i>cox1</i> versus <i>cox2</i>	0.10528143	10.80	0.0082	0.13118
1	<i>cox1</i> versus <i>cox3</i>	0.09583333	9.83	0.0106	0.16944
1	<i>cox1</i> versus <i>nd2</i>	2.61602162	268.38	<0.0001	<0.0001
1	<i>cox1</i> versus <i>nd3</i>	1.68320797	172.68	<0.0001	<0.0001
1	<i>cox2</i> versus <i>cox3</i>	0.00022205	0.02	0.8830	1.00000
1	<i>cox2</i> versus <i>nd2</i>	1.67169645	171.50	<0.0001	<0.0001
1	<i>cox2</i> versus <i>nd3</i>	0.94656174	97.11	<0.0001	0.00003
1	<i>cox3</i> versus <i>nd2</i>	1.71045183	175.48	<0.0001	<0.0001
1	<i>cox3</i> versus <i>nd3</i>	0.97577940	100.11	<0.0001	0.00003
1	<i>nd2</i> versus <i>nd3</i>	0.10241726	10.51	0.0088	0.15041
2	<i>cytb</i> versus <i>nd6</i>	3.4034830	311.91	<0.0001	<0.0001
3	<i>nd4/nd4L</i> versus <i>nd5</i>	0.08906566	9.14	0.0128	0.20534

P-values were Bonferroni corrected to account for multiple testing (19).

**Table 4.** Pairwise comparison of the levels of gene expression between genes interacting the same protein complex

Complex	Gene pair	Mean square	F-value	P-value	Corrected P-value
cox	<i>cox1</i> versus <i>cox2</i>	0.10528143	10.80	0.0082	0.14758
cox	<i>cox1</i> versus <i>cox3</i>	0.09583333	9.83	0.0106	0.19062
cox	<i>cox2</i> versus <i>cox3</i>	0.00022205	0.02	0.8830	1.00000
nd	<i>nd1</i> versus <i>nd2</i>	0.26957847	27.66	0.0004	0.00664
nd	<i>nd1</i> versus <i>nd3</i>	0.03967367	4.07	0.0713	1.00000
nd	<i>nd1</i> versus <i>nd4/nd4L</i>	0.03292959	3.38	0.0959	1.00000
nd	<i>nd1</i> versus <i>nd5</i>	0.01368282	1.40	0.2635	1.00000
nd	<i>nd1</i> versus <i>nd6</i>	1.26766942	130.05	<0.0001	0.00001
nd	<i>nd2</i> versus <i>nd3</i>	0.10241726	10.51	0.0088	0.15925
nd	<i>nd2</i> versus <i>nd4/nd4L</i>	0.49094487	50.37	<0.0001	0.00060
nd	<i>nd2</i> versus <i>nd5</i>	0.16179370	16.60	0.0022	0.04023
nd	<i>nd2</i> versus <i>nd6</i>	0.36808343	37.76	0.0001	0.00196
nd	<i>nd3</i> versus <i>nd4/nd4L</i>	0.14489261	14.86	0.0032	0.05731
nd	<i>nd3</i> versus <i>nd5</i>	0.00675831	0.69	0.4245	1.00000
nd	<i>nd3</i> versus <i>nd6</i>	0.85882064	88.11	<0.0001	0.00005
nd	<i>nd4/nd4L</i> versus <i>nd5</i>	0.08906566	9.14	0.0128	0.23101
nd	<i>nd4/nd4L</i> versus <i>nd6</i>	1.70922516	175.35	<0.0001	<0.0001
nd	<i>nd5</i> versus <i>nd6</i>	1.01794919	104.43	<0.0001	0.00002

P-values were Bonferroni corrected to account for multiple testing (19).

not allow the estimation of technical error of the four libraries sequenced with 454 technology.

Expression levels were significantly different for some of the genes within a transcriptional unit or protein complex (Tables 3 and 4). However, we observed that PCGs within a transcriptional unit forming a protein complex are consistently not significantly different from each other.

### Identification of novel transcripts

Alignments of the large subunit of the rRNA (*rrnL*) reads to reference genes indicated that some ESTs resulted from rearrangements of the original *rrnL* mRNA. In these rearranged ESTs, a sequence inversion has occurred, causing one of the segments to be inverted from its original orientation in the mitochondrial genome (Figure 2). The inverted segment was flanked by a

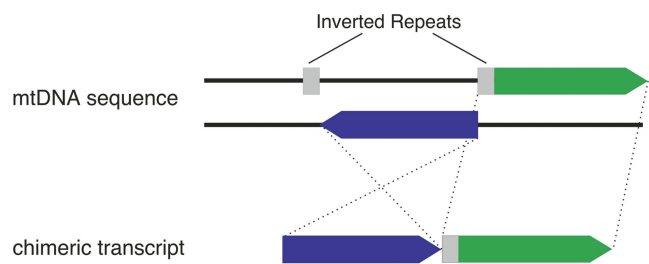
perfect inverted repeat (IR) of 6 nt. These chimeric transcripts could have been artifacts generated during cDNA synthesis. However, similar transcripts were found as naturally occurring molecules in mouse (24) and human cells (25) and in a different mitochondrial region, *atp6*, in porcine brain cells (26). Therefore, we investigated the occurrence of such structures in the whole mitochondrial dataset.

Chimeric transcripts were found for all mitochondrial genes (Table 5). The structure of the chimeras for all other genes resembled that of *rrnL*; they were all inverted segments joined by a small region (5–11 nt) that occur as a perfect repeat in the mitochondrial genome sequence. Because of the short length of the 454 reads, it was not possible to infer the full structure of the chimeras. Therefore, we clustered the chimeric reads from the four libraries to identify unique transcripts (Table 5). Each

cluster obtained represents a different type of chimeric transcript. The number of different chimeric structures ranged from one in the *rrnS* gene and 64 in the *rrnL* gene. Interestingly, we observed the same chimeric ESTs in multiple libraries, suggesting that these sequences reflect real structures that are conserved between species (Supplementary Table S5). Nevertheless, for individual chimeric transcripts no clear pattern emerged, probably due to the low number of observations.

DISCUSSION

Basic mechanisms of mtDNA transcription have been investigated in the last few decades, but the recent advances of the sequencing technologies permit now an accurate quantification of transcripts. In this study, we took advantage of NGS and tested to what extent gene expression differs between sexes and species.



**Figure 2.** Structure of chimeric transcripts. Chimeras consisted of a 3'-subregion (green arrow) joined to an inverted segment (blue arrow) by a small region (4–11 nt) that occur as a perfect repeat in the mitochondrial genomic sequence (grey boxes).

Mitochondrial origin of the *D. pseudoobscura* WGS contigs and 454 reads

Nuclear insertions of mitochondrial-like sequences (numts) are widespread in genomes of vertebrates and invertebrates (27). Numts present a challenge for studying short mitochondrial sequences derived from PCR amplification or shotgun libraries prepared from total genomic DNA due to the possible inclusion of paralogous nuclear sequences (28,29).

Numts rarely occur in the *Drosophila* genome (27,28). A previous analysis of the complete genome of *D. melanogaster* revealed that only approximately 500 bp of mtDNA was transferred to the nuclear genome (28). The scarcity of numts in the *D. melanogaster* genome supports the view that there are substantial selective constraints in *Drosophila* noncoding DNA. The *Drosophila* genome is highly compact. Very few pseudogenes can be found in the *D. melanogaster* genome (30) and they are lost at a very high rate, demonstrating a strong deletion bias (31).

Some additional evidence supports the mitochondrial origin of *D. pseudoobscura* WGS contigs. First, the assembled singletons contained syntenic blocks of complete genes relative to *D. melanogaster* and *D. yakuba* mtDNA. Secondly, the PCGs used the invertebrate mitochondrial genetic code, and the tRNAs found among the genes had a typical mitochondrial cloverleaf structure (32). Finally, the estimation of the average divergence showed that the 454 EST data and the assembled mtDNA genome were highly similar (Supplementary Table S6).

The 454 ESTs analyzed here derive from the sequencing of cDNA libraries constructed from the total poly(A)+ RNA fraction and there is no evidence for the expression of these nuclear insertions. Numts are found in noncoding regions of the nuclear genome and most of them are

**Table 5.** Number of intra-molecular chimeric transcripts

	Chimeric reads in all libraries <sup>a</sup>	Number of clusters <sup>b</sup>	Singletons <sup>c</sup>	<i>D. melanogaster</i>		<i>D. pseudoobscura</i>	
				Male Counts (%) <sup>d</sup>	Female Counts (%) <sup>d</sup>	Male Counts (%) <sup>d</sup>	Female Counts (%) <sup>d</sup>
<i>atp6/atp8</i>	85	37	19	64 (1.51)	9 (1.65)	11 (2.03)	1 (0.24)
<i>cytb</i>	60	25	16	45 (1.26)	10 (3.00)	5 (2.91)	0 (0)
<i>cox1</i>	87	27	15	69 (0.91)	14 (2.09)	0 (0)	4 (1.42)
<i>cox2</i>	48	25	11	37 (0.87)	5 (1.26)	3 (1.12)	3 (1.58)
<i>cox3</i>	37	21	14	27 (0.69)	7 (2.28)	2 (0.55)	1 (0.48)
<i>nd1</i>	4	4	4	4 (0.38)	0 (0)	0 (0)	0 (0)
<i>nd2</i>	11	3	1	11 (3.78)	0 (0)	0 (0)	0 (0)
<i>nd3</i>	7	5	4	4 (0.46)	0 (0)	0 (0)	3 (8.57)
<i>nd4/nd4L</i>	12	12	12	12 (0.76)	0 (0)	0 (0)	0 (0)
<i>nd5</i>	17	9	7	15 (1.56)	0 (0)	1 (1.47)	1 (1.39)
<i>nd6</i>	2	2	2	1 (0.39)	0 (0)	1 (4.76)	0 (0)
<i>rrnL</i>	426	64	40	356 (0.79)	19 (0.39)	29 (0.51)	22 (0.45)
<i>rrnS</i>	1	1	1	0 (0)	1 (0.25)	0 (0)	0 (0)
Total	797	235	146	645 (0.85)	65 (0.83)	52 (0.65)	35 (0.54)

<sup>a</sup>Total number of chimeric reads in the four libraries.  
<sup>b</sup>Each cluster represents a different type of chimeric transcript.  
<sup>c</sup>Number of clusters with only one read (number of molecules sequenced only once).  
<sup>d</sup>Percentage of the total counts for each gene.

only gene fragments (33). They are under different mutational and selective constraints compared with mitochondrial genes (29) and the identity of the ESTs obtained here to the corresponding mtDNA sequence is >95%. The nuclear origin of these ESTs is very unlikely, since we would have to assume a recent gene transfer from mitochondria to nucleus, because of their high similarity to mtDNA and a very high expression level of the fragments transferred.

### Differential gene expression of mitochondrial genes

We detected a substantial heterogeneity in gene expression among the PCGs of the mtDNA. First, we found significant differences among genes from the NADH dehydrogenase (complex I, Table 4). Similarly, we also observed significant differences among genes encoded by one of the four transcription units. The uncoordinated expression of individual transcripts is unexpected, since most of the mitochondrial genes are transcribed together as large transcriptional units. This heterogeneity in gene expression may be caused by different stability of the transcripts and post-transcriptional mechanisms acting on them. Another likely post-transcriptional mechanism involves the processing of the transcriptional units to generate mature transcripts by the tRNA punctuation model (5). In order to produce the mature transcripts, the tRNA sequences are cleaved sequentially in a 3'-to 5'-direction (34). The maturation of mitochondrial transcripts also involves the removal of noncoding nucleotides in the 5'-end (35). These additional processing to generate mature transcripts possibly play an important role in keeping the steady-state levels of mitochondrial transcripts than the regulation of transcription.

In order to validate our gene expression data, we compared our results to a previous study (35), which determined the relative abundance of the mtDNA transcripts by northern blots, which were probed by specific *D. melanogaster* mtDNA fragments. We found a good agreement of our quantification with the previous results of Berthier *et al.* (35). The Pearson correlation coefficients between the expression data ranged from 0.55 in the female *D. pseudoobscura* library to 0.62 in the female *D. melanogaster* library (Table 6). In our approach, we had an under-representation of *rrnS* transcripts due to the short length of their poly(A) tails. Excluding this gene

from the comparison, the correlation coefficients in the pairwise comparisons increased to 0.73–0.78 (in the female *D. pseudoobscura* and in the male *D. melanogaster* libraries, respectively). The differences between the two expression measures could be due to different sensitivities of the employed techniques; the dynamic range was much larger in our experiment compared with the northern approach allowing for a better resolution of the expression profiles found by Berthier *et al.* (35). For instance, the same level of expression had been assigned to all mitochondrial transcripts in the *nd* complex, except to *nd6* gene that was not detected. We have also observed that the least expressed gene in the *nd* complex was *nd6*, but we were able to detect its expression as well as subtle differences in gene expression among the genes in *nd* complex that were not evident in the former study. The high correlation between our RNA-Seq study, which relies on the effective capture of the mitochondrial RNAs by oligo dT primers, shows that variation in poly(A) tail length among the RNA molecules and thus different recovery cannot explain the observed heterogeneity among the mitochondrial RNAs. The only exception to this is the small RNA subunit. Consistent with this observation, previous studies have characterized the mitochondrial transcripts and found that, except for the small rRNA subunit, all the transcripts have poly(A) tails with length of 40–50 nt (34,36). Nevertheless, experimental validation of the homogeneity in poly(A) tail length by sequencing is not possible, as the 454 sequencing technology is not well suited for measuring homopolymers length (14).

The high concordance of the two studies confirmed the observed differences in the levels of gene expression among the mitochondrial genes. Genes in mtDNA are clustered to gain efficiency by coordinating activities and regulation. As RNA synthesis is costly, it may be expected that the organization of the genes into five different transcription units has been optimized to minimize the waste of energy. Assuming similar translation efficiencies among the mitochondrial genes, the level of gene expression should reflect the stoichiometry of the subunits of the enzymatic complexes. Hence, to minimize RNA waste, it is expected that subunits with the same stoichiometry should be preferentially encoded on the same transcript. Loss or disruption of such proximity may decrease metabolic flux and have deleterious consequences in individual fitness.

We find support for our hypothesis for the cytochrome c oxidase complex in which subunits 1, 2 and 3 are encoded in the same transcriptional unit and needed at a stoichiometry of 1:1:1 in many species [enzyme 1.9.3.1 in (37)]. Unfortunately, no information on the stoichiometry of the polypeptide components is available for the NADH dehydrogenase complex, and for the ATP synthase hydrophobic portion  $F_0$  that is formed, among others, by the subunits encoded by the *atp6* and *atp8* genes (38).

Another interesting aspect of our investigation was the comparison of the overall gene expression in males and females in both species, *D. melanogaster* and *D. pseudoobscura*. This analysis indicated that the

**Table 6.** Correlation coefficients between the expression profiles of *D. melanogaster* and *D. pseudoobscura* obtained here and the levels of mRNA estimated by Berthier *et al.* (35) by northern hybridization of fragments of the mitochondrial genome to the RNA fraction isolated from mitochondria

	All transcripts	Without <i>rrnS</i>	Only PCG
<i>Drosophila melanogaster</i>			
Male library	0.61	0.78	0.81
Female library	0.62	0.76	0.75
<i>Drosophila pseudoobscura</i>			
Male library	0.57	0.75	0.81
Female library	0.55	0.73	0.72



expression pattern was conserved among genes in the two species, but differed significantly between both sexes. This expression difference is a very interesting observation and may reflect differences in mitochondrial bioenergetics between males and females (39).

### Novel mitochondrial transcripts

Several studies have reported the occurrence of chimeric transcripts. In most of the cases, these transcripts were described as technical artifacts, such as end-to-end joining of noncontiguous cDNA sequences and template switching by the reverse transcriptase during cDNA synthesis (40,41). However, several classes of natural chimeric transcripts play important biological roles such as the regulation of translation efficiency (42) and increasing protein diversity (43). These hybrid transcripts are generated *in vivo* by known mechanisms, mRNA trans-splicing and processing of polycistronic transcription units (42,44). Detailed analysis of such transcripts focused on the evaluation of hybrids between heterologous mRNAs (44), but single-gene chimeras are usually overlooked in the analyses.

While we have no empirical proof that the chimeric transcripts found in this study are no technical artifacts, some lines of evidence argue against technical artifacts. End-to-end joining during library construction would randomly ligate molecules and we would expect not only intra-molecule chimeras but also ligation of independently transcribed mRNAs. In particular, due to the highest abundance of the *rrnL* transcript, it would be expected that this transcript would be frequently ligated to other transcripts forming intermolecule chimeras, but this was only observed in two cases (out of 797). Furthermore, we observed the same chimeric EST in multiple-sequence reads from the same library as well as in independent libraries.

Template switching by the reverse transcriptase can artificially delete portions of cDNAs that can be wrongly interpreted as an alternative transcript (40). Most of the chimeric transcripts found here were inverted sequences with one of the segments being flanked by a 5–10 bp perfect IR spaced by 15–20 bp in the mitochondrial genomic sequence (Figure 2). A homology-dependent template switching could have generated these hybrid cDNAs artificially. During first-strand cDNA synthesis, poly(A)<sup>+</sup> RNAs are primed by an oligo(dT) and when the reverse transcriptase reaches the 5'-end of the junction site in the mRNA, the nascent cDNA switches the template to the IR in the antisense mRNA that will now be used as a new template. For the cDNA synthesis, we used the RevertAid<sup>TM</sup> H Minus M-MuLV Reverse Transcriptase (Fermentas), which lacks RNase H activity. It has been shown that homologous recombination between two distinct RNA templates promoted by the reverse transcriptase enzyme requires the involvement of RNase H activity (45–47). Also, contrary to the *in vitro* template switching hypothesis is the fact that we have not found independently transcribed mRNAs containing identical sequences forming hybrids between them.

Two different groups have provided strong evidence for the presence of such mitochondrial chimeric (also called fusion) transcripts in porcine brain (26), mouse cells (24) and human proliferating cells (25,48). Michel and co-workers (26) identified a chimeric transcript in the *atp6* region that consisted of inverted sequences joined by 7 nt (TTACTAT) that is also present in the canonical transcript as an IR. The authors demonstrated that the unusual mitochondrial transcripts were naturally occurring mitochondrial RNAs and not a technical artifact performing S1 nuclease protection assays with total RNA from the sample analyzed. Interestingly, they have found that this transcript was differentially expressed in fetal and adult brain samples. Villegas *et al.* (24) also found a noncoding mitochondrial transcript (ncmtRNA) with a similar structure in mouse cells, but involving the large subunit of the ribosomal RNA. An equivalent transcript was detected later in human proliferating cells (25). This human transcript had an inverted region of 815 nt linked to the 5'-end of the *rrnL*. Several lines of evidence showed that this transcript is indeed synthesized in mitochondria and its expression was correlated with the replicative state of the human cells. In a recent study, Burzio *et al.* (48) described two additional ncmtRNAs with a very similar structure to the transcript identified by Villegas *et al.* (25). They had a very interesting finding regarding the pattern of expression of the three human ncmtRNAs. Normal proliferating cells express all three ncmtRNAs, while neither is expressed in nondividing cells. In tumor cells, however, only the former one is expressed; the two new ncmtRNAs (48) are down-regulated. Burzio *et al.* (48) hypothesized that the identified transcripts are involved in the regulation of the cell cycle.

Our study showed that chimeric transcripts could be detected for almost all mitochondrial genes. While we have no empirical proof for the presence of these chimeric transcripts, evidence provided elsewhere showed that similar structures are present in mitochondria of different organisms. Owing to the high content of A + T in insect mitochondrial genomes, short perfect repeats occur very frequently favoring the *in vivo* generation of chimeric RNA by site-specific recombination. In plant mitochondria, several studies have reported a massive and continuous production of cryptic transcripts [(49) and references therein]. These presumably nonfunctional RNAs are produced because of the complex structure of the mitochondrial genome combined with its relaxed control of transcription (49). On the other hand, these transcripts could be involved in the regulation of gene expression of mitochondrial genes. As in Burzio *et al.* (48), they might play a role in the regulation of the cell cycle. As our samples were prepared from whole adult bodies, we cannot distinguish expression profiles of proliferating and nonproliferating cells.

A number of *cis*-acting noncoding RNAs (ncRNAs) are known to contain functional information [see (50) for a review], often affecting the translation efficiency or mRNA stability. These *cis*-acting ncRNAs are usually transcribed from UTRs flanking the affected gene, from intronic or intergenic regions. Owing to their compact structure,



animal mtDNA lacks UTRs and intronic regions. The ncmtRNAs may have evolved in animal mtDNA as a backup mechanism ensuring the fine-tuning of the gene expression. Further investigations are necessary to analyze the origin of the chimeric transcripts, the mechanism of generation and their possible biological function.

## ACCESSION NUMBERS

GSE16651, GenBank accession number FJ899745.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We are thankful to the CS lab members for helpful comments and suggestions and to S. Glinka for 454 sequencing.

## FUNDING

Fonds zur Förderung der wissenschaftlichen Forschung (FWF) grants (P19832, L403 to C.S.); fellowship of the Brazilian National Council for Scientific and Technological Development (CNPq 200512/2006-4 to T.T.T.). Funding for open access charge: Förderung der wissenschaftlichen Forschung.

*Conflict of interest statement.* None declared.

## REFERENCES

- Boore, J.L. (1999) Animal mitochondrial genomes. *Nucleic Acids Res.*, **27**, 1767–1780.
- Goddard, J.M. and Wolstenholme, D.R. (1980) Origin and direction of replication in mitochondrial DNA molecules from the genus *Drosophila*. *Nucleic Acids Res.*, **8**, 741–757.
- Taanman, J.W. (1999) The mitochondrial genome: structure, transcription, translation and replication. *Biochim. Biophys. Acta*, **1410**, 103–123.
- Wolstenholme, D.R. (1992) Animal mitochondrial DNA: structure and evolution. *Int. Rev. Cytol.*, **141**, 173–216.
- Ojala, D., Montoya, J. and Attardi, G. (1981) tRNA punctuation model of RNA processing in human mitochondria. *Nature*, **290**, 470–474.
- Hirsch, M. and Penman, S. (1973) Mitochondrial polyadenylic acid-containing RNA: localization and characterization. *J. Mol. Biol.*, **80**, 379–391.
- Hirsch, M. and Penman, S. (1974) The messenger-like properties of the poly(A) plus RNA in mammalian mitochondria. *Cell*, **3**, 335–339.
- Torres, T.T., Metta, M., Ottenwälder, B. and Schlötterer, C. (2008) Gene expression profiling by massively parallel sequencing. *Genome Res.*, **18**, 172–177.
- Laveder, P., De Pitta, C., Toppo, S., Valle, G. and Lanfranchi, G. (2002) A two-step strategy for constructing specifically self-subtracted cDNA libraries. *Nucleic Acids Res.*, **30**, e38.
- Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M. and Gilad, Y. (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.*, **18**, 1509–1517.
- Hillier, L.D., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chisoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W. *et al.* (1996) Generation and analysis of 280,000 human expressed sequence tags. *Genome Res.*, **6**, 807–828.
- Gissi, C. and Pesole, G. (2003) Transcript mapping and genome annotation of ascidian mtDNA using EST data. *Genome Res.*, **13**, 2203–2212.
- Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. (1995) Serial analysis of gene expression. *Science*, **270**, 484–487.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z.T. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.
- Huang, X. and Madan, A. (1999) CAP3: a DNA sequence assembly program. *Genome Res.*, **9**, 868–877.
- Huang, W. and Marth, G. (2008) EagleView: a genome assembly viewer for next-generation sequencing technologies. *Genome Res.*, **18**, 1538–1543.
- Wyman, S.K., Jansen, R.K. and Boore, J.L. (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*, **20**, 3252–3255.
- Kofler, R., Torres, T.T., Lelley, T. and Schlötterer, C. (2009) PanGEA: identification of allele specific gene expression using the 454 technology. *BMC Bioinformatics*, **10**, 143.
- Rice, W.R. (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.
- Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., Markow, T.A., Kaufman, T.C., Kellis, M., Gelbart, W., Iyer, V.N. *et al.* (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*, **450**, 203–218.
- Richards, S., Liu, Y., Bettencourt, B.R., Hradecky, P., Letovsky, S., Nielsen, R., Thornton, K., Hubisz, M.J., Chen, R., Meisel, R.P. *et al.* (2005) Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. *Genome Res.*, **15**, 1–18.
- Martin, M., Cho, J., Cesare, A., Griffith, J. and Attardi, G. (2005) Termination factor-mediated DNA loop between termination and initiation sites drives mitochondrial rRNA synthesis. *Cell*, **123**, 1227–1240.
- Roberti, M., Bruni, F., Polosa, P.L., Gadaleta, M.N. and Cantatore, P. (2003) DmTTF, a novel mitochondrial transcription termination factor that recognises two sequences of *Drosophila melanogaster* mitochondrial DNA. *Nucleic Acids Res.*, **31**, 1597–1604.
- Villegas, J., Zarraga, A.M., Muller, I., Montecinos, L., Werner, E., Brito, M., Meneses, A.M. and Burzio, L.O. (2000) A novel chimeric mitochondrial RNA localized in the nucleus of mouse sperm. *DNA Cell Biol.*, **19**, 579–588.
- Villegas, J., Burzio, V., Villota, C., Landerer, E., Martinez, R., Santander, M., Pinto, R., Vera, M.I., Boccardo, E., Villa, L.L. *et al.* (2007) Expression of a novel non-coding mitochondrial RNA in human proliferating cells. *Nucleic Acids Res.*, **35**, 7336–7347.
- Michel, U., Stringaris, A.K., Nau, R. and Rieckmann, P. (2000) Differential expression of sense and antisense transcripts of the mitochondrial DNA region coding for ATPase 6 in fetal and adult porcine brain: identification of novel unusually assembled mitochondrial RNAs. *Biochem. Biophys. Res. Commun.*, **271**, 170–180.
- Richly, E. and Leister, D. (2004) NUMTs in sequenced eukaryotic genomes. *Mol. Biol. Evol.*, **21**, 1081–1084.
- Bensasson, D., Zhang, D., Hartl, D.L. and Hewitt, G.M. (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol. Evol.*, **16**, 314–321.
- Zhang, D.X. and Hewitt, G.M. (1996) Highly conserved nuclear copies of the mitochondrial control region in the desert locust *Schistocerca gregaria*: some implications for population studies. *Mol. Ecol.*, **5**, 295–300.
- Harrison, P.M., Milburn, D., Zhang, Z., Bertone, P. and Gerstein, M. (2003) Identification of pseudogenes in the *Drosophila melanogaster* genome. *Nucleic Acids Res.*, **31**, 1033–1037.
- Petrov, D.A. and Hartl, D.L. (2000) Pseudogene evolution and natural selection for a compact genome. *J. Hered.*, **91**, 221–227.
- Helm, M., Brule, H., Friede, D., Giege, R., Putz, D. and Florentz, C. (2000) Search for characteristic structural features of mammalian mitochondrial tRNAs. *RNA*, **6**, 1356–1379.

33. Blanchard, J.L. and Schmidt, G.W. (1996) Mitochondrial DNA migration events in yeast and humans: integration by a common end-joining mechanism and alternative perspectives on nucleotide substitution patterns. *Mol. Biol. Evol.*, **13**, 893.
34. Stewart, J.B. and Beckenbach, A.T. (2009) Characterization of mature mitochondrial transcripts in *Drosophila*, and the implications for the tRNA punctuation model in arthropods. *Gene*, **445**, 49–57.
35. Berthier, F., Renaud, M., Alziari, S. and Durand, R. (1986) RNA mapping on *Drosophila* mitochondrial DNA: precursors and template strands. *Nucleic Acids Res.*, **14**, 4519–4533.
36. Benkel, B.F., Duschesnay, P., Boer, P.H., Genest, Y. and Hickey, D.A. (1988) Mitochondrial large ribosomal RNA: an abundant polyadenylated sequence in *Drosophila*. *Nucleic Acids Res.*, **16**, 9880.
37. Schomburg, D., Chang, A. and Schomburg, I. (2007) (eds), “Class 1 oxidoreductases X EC 1.9–1.13”, *Springer Handbook of Enzymes*, 2nd edn. Springer, New York.
38. Yoshida, M., Muneyuki, E. and Hisabori, T. (2001) ATP synthase – a marvellous rotary engine of the cell. *Nat. Rev. Mol. Cell Biol.*, **2**, 669–677.
39. Ballard, J.W., Melvin, R.G., Miller, J.T. and Katewa, S.D. (2007) Sex differences in survival and mitochondrial bioenergetics during aging in *Drosophila*. *Aging Cell*, **6**, 699–708.
40. Cocquet, J., Chong, A., Zhang, G. and Veitia, R.A. (2006) Reverse transcriptase template switching and false alternative transcripts. *Genomics*, **88**, 127–131.
41. Roy, S.W. and Irimia, M. (2008) When good transcripts go bad: artifactual RT-PCR ‘splicing’ and genome analysis. *Bioessays*, **30**, 601–605.
42. Nilsen, T.W. (1993) Trans-splicing of nematode premessenger RNA. *Annu. Rev. Microbiol.*, **47**, 413–440.
43. Bonen, L. (1993) Trans-splicing of pre-mRNA in plants, animals, and protists. *FASEB J.*, **7**, 40–46.
44. Romani, A., Guerra, E., Trerotola, M. and Alberti, S. (2003) Detection and analysis of spliced chimeric mRNAs in sequence databanks. *Nucleic Acids Res.*, **31**, e17.
45. Negroni, M., Ricchetti, M., Nouvel, P. and Buc, H. (1995) Homologous recombination promoted by reverse transcriptase during copying of two distinct RNA templates. *Proc. Natl Acad. Sci. USA*, **92**, 6971–6975.
46. Peliska, J.A. and Benkovic, S.J. (1992) Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. *Science*, **258**, 1112–1118.
47. Zhu, S., Li, W. and Cao, Z. (2001) A naturally occurring non-coding fusion transcript derived from scorpion venom gland: implication for the regulation of scorpion toxin gene expression. *FEBS Lett.*, **508**, 241–244.
48. Burzio, V.A., Villota, C., Villegas, J., Landerer, E., Boccardo, E., Villa, L.L., Martinez, R., Lopez, C., Gaete, F., Toro, V. *et al.* (2009) Expression of a family of noncoding mitochondrial RNAs distinguishes normal from cancer cells. *Proc. Natl Acad. Sci. USA*, **106**, 9430–9434.
49. Holec, S., Lange, H., Canaday, J. and Gagliardi, D. (2008) Coping with cryptic and defective transcripts in plant mitochondria. *Biochim. Biophys. Acta*, **1779**, 566–573.
50. Mattick, J.S. (2009) The genetic signatures of noncoding RNAs. *PLoS Genet.*, **5**, e1000459.