Functional Analysis of Nuclear-Encoded Mitochondrial Gene Duplicates with Testis-Biased Expression
in *Drosophila melanogaster*

by

Mohammadmehdi Eslamieh

DISSERTATION

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at

The University of Texas at Arlington

August 2019

Arlington, TX

Supervising Committee:

Esther Betrán, Supervising Professor

Jeff Demuth

Laura D Mydlarz

Matthew K. Fujita

Todd Castoe

i

DEDICATION

To the two beautiful angels in my life:

my wife, Mahnaz,

&

our daughter, Ariana

ACKNOWLEDGEMENTS

Date: 31 July 2019

ABSTRACT

Functional Analysis of Nuclear-Encoded Mitochondrial Gene Duplicates with Testis-Biased Expression in *Drosophila melanogaster*

Mohammadmehdi Eslamieh, Ph.D.

The University of Texas at Arlington, 2019

Supervising Professor: Esther Betrán

Most of the genes encoding proteins that function in the mitochondria are located in the nucleus and are called nuclear-encoded mitochondrial genes (N-mt genes). In *Drosophila melanogaster*, most of the tissue-specific duplicated N-mt genes have acquired testis-biased expression. These genes appear to be older and relocated more often than other duplicated genes and have energy-related functions. These patterns reveal strong selection for the retention of new genes for male germline mitochondrial functions. Maternal inheritance of mitochondria and male-male competition for fertilization are two common forces that might drive changes in mitochondrial functions. We analyzed N-mt genes in the human genome to test the presence of this pattern in other species. We find that about 18% of human N-mt genes fall into gene families, but unlike in *Drosophila*, only 28% of the N-mt duplicates have tissue-biased expression and only 36% of these have testis-biased expression. To study the function of the *Drosophila* testis-biased duplicates, we knocked down 39 N-mt duplicated genes with testis-biased expression in *D. melanogaster*. Multiple RNAi lines and a few TRiP-CRISPR knockout lines were used to study the fertility effects of these

iv

genes. Forty-nine percent of the RNAi knockdowns in germline showed male sterility and seventy-two percent of TRiP-CRISPR knockout lines and all few previously published knockout studies confirmed these results. We have failed to find any evidence of the parental gene being able to compensate for the N-mt duplicated gene in the fertility effects. These effects along with the different pattern of expression that leads to the presence of these genes in the sperm proteome/mitochondria, their protein divergence, and positive selection inferred for some of these genes suggest that the new duplicate genes have a different function than their parental genes during spermatogenesis and/or in sperm. We also produced a knockout mutant for *Cytochrome c oxidase 4-like* (*COX4L*) in *D. melanogaster* to understand the function of this new gene in detail. The knockout strain for this gene was generated using CRISPR-Cas9 technology. The knockout of *COX4L* produce completely sterile males unlike the knockdown that produced partial male sterility and reveals that knockdown results might have milder phenotypes than knockouts as expected. A lack of sperm individualization is observed in those males. The male infertility is rescued by driving *COX4L*-HA in the germline. In addition, ectopic expression of *COX4L* in soma caused embryonic lethality as expected, if *COX4L* has a specialized male germline function. Overexpression of *COX4L* in male germline leads to a reduction in male fertility revealing that fine tuning of the expression of this gene is needed. Several stainings revealed that the mitochondria of *COX4L*-KO are not functional and the ROS production in the testes was higher than in controls which might explain the male sterility of these flies. This prominent phenotype along with having energy-related functions, testis-biased expression, and also being present in the *Drosophila* sperm proteome database supports the idea that males might use different mitochondria in their germline and/or sperm.

Table of contents:

Chapter 1

Few Nuclear-Encoded Mitochondrial Gene Duplicates Contribute to Male Germline-Specific Functions in Humans

Mohammadmehdi Eslamieh, Anna Williford, and Esther Betrán*

Department of Biology, University of Texas at Arlington

*Corresponding author: E-mail: betran@uta.edu.

Few Nuclear-Encoded Mitochondrial Gene Duplicates Contribute to Male Germline-Specific Functions in Humans

Mohammadmehdi Eslamieh, Anna Williford, and Esther Betrán*

Department of Biology, University of Texas at Arlington

*Corresponding author: E-mail: betran@uta.edu.

Abstract

Most of the genes encoding proteins that function in the mitochondria are located in the nucleus and are called nuclear-encoded mitochondrial genes, or N-mt genes. In *Drosophila melanogaster*, about 23% of N-mt genes fall into gene families, and all duplicates with tissue-biased expression (76%) are testis biased. These genes are enriched for energy-related functions and tend to be older than other duplicated genes in the genome. These patterns reveal strong selection for the retention of new genes for male germline mitochondrial functions. The two main forces that are likely to drive changes in mitochondrial functions are maternal inheritance of mitochondria and male–male competition for fertilization. Both are common among animals, suggesting similar N-mt gene duplication patterns in different species. To test this, we analyzed N-mt genes in the human genome. We find that about 18% of human N-mt genes fall into gene families, but unlike in *Drosophila*, only 28% of the N-mt duplicates have tissue-biased expression and only 36% of these have testis-biased expression. In addition, human testis-biased duplicated genes are younger than other duplicated genes in the genome and have diverse functions. These contrasting patterns between species might reflect either differences in selective pressures for germline energy-related or other mitochondrial functions during spermatogenesis and fertilization, or differences in the response to similar pressures.

Key words: nuclear-encoded mitochondrial genes, gene duplication, male-biased expression, human.

Introduction

Mitochondria are organelles that arose from proteobacteria through endosymbiosis (Sagan 1967; Yang et al. 1985) and have undergone drastic reduction in genome size during evolution (e.g., human mtDNA is only ~16.5 Kb; Berg and Kurland 2000). Currently, most of the genes encoding mitochondrial proteins are located in the nucleus (i.e., nuclear-encoded mitochondrial or N-mt genes). In humans, there are around 1,500 N-mt protein-coding genes that interact with 37 genes encoded by mtDNA (mt genes), 13 of which are protein coding. Together they fulfill all mitochondrial functions, including replication, transcription, translation, transport, mitochondria-nucleus communication and energy production. Only a small fraction of the N-mt genes (10–20%) have a strong support for a-proteobacterial origin whereas the rest originated from prokaryotic (unclear lineage) or eukaryotic organisms' genomes, and acquired mitochondrial function at some point in time (Gray 2015).

Interestingly, some N-mt genes are duplicates of other N-mt genes and have acquired tissue-biased expression (Emerson et al. 2004; Bai et al. 2007; Porcelli et al. 2007; Wolff et al. 2014). This could be seen as unexpected given that the mitochondrial functions are required in nearly every cell type. An extreme example of this comes from the analysis of the *Drosophila melanogaster* (*D. melanogaster*) genome where all duplicates with tissue-biased expression (76%) are testis-biased (Gallach et al. 2010 and this work, see Results). In addition, these genes are 1) older than the average age of duplicated genes from the entire genome, 2) involved in interchromosomal duplication, including an excess of retrogene duplications among N-mt duplicates compared with all duplicates in the genome, 3) overrepresented on autosomes, and 4) enriched for energy-related functions (catabolic and carbohydrate metabolic processes; Gallach et al. 2010). These features reveal strong selection for the retention of these duplicates in the *Drosophila* genome (Gallach et al. 2010).

We proposed that these duplicates might have evolved in response to intralocus sexual antagonism generated by the appearance of a new male-beneficial and female-harming allele at a single-copy locus. Subsequent duplication of the

male-beneficial allele accompanied by the acquisition of male-biased expression could then lead to the resolution of this intra-locus sexual antagonism (Gallach et al. 2010; Gallach and Betran 2011). Given that males do not pass mitochondria to their offspring and are under strong male–male competition to fertilize females' eggs, selection might favor mitochondria that are specialized in high-energy production despite the potential for increased generation of reactive oxygen species and mutations associated with this increase (Waris and Ahsan 2006; Murphy 2009; Gallach et al. 2010; Gallach and Betran 2011).

However, several other nonexclusive hypotheses could explain the patterns observed in *Drosophila*. New duplicates could have been retained because of their dose effects or have been fixed by chance with subsequent specialization for male germline or sperm functions (See Innan and Kondrashov 2010 for a review of gene duplication models). It has also been suggested that some of these duplicates might compensate for mtDNA male-harming mutations that do not hurt females and cannot be selected against (Rogell et al. 2014). Since mitochondria are inherited only from the mother, mutations in the mitochondrial genome can get fixed even if they are harmful to males (Partridge and Hurst 1998; Rand et al. 2006; Montooth et al. 2010). Mutations that are harmful only to males are likely to affect testis functions (Innocenti et al. 2011; Rogell et al. 2014) and to occur in energy-related genes if selection for energy-related functions is different in males (Gallach et al. 2010; Gallach and Betran 2011).

Under any of the above hypotheses, we argue that selection to improve male germline energy-related mitochondrial function in the presence of male–male competition and/or in the presence of male-deleterious mitochondrial mutations should be common among species. If recruiting N-mt duplicates is the most efficient response to these selective pressures, we would expect to find large number of N-mt duplicates with male germline energy functions in other species. To check this expectation, we analyzed the entire set of N-mt genes in the human genome. Our analyses show that human N-mt duplicates have an entirely different pattern: duplicates show tissue-biased expression in many tissues (not just in testis as in the case of *Drosophila*), and they fall into various functional categories (not enriched in energy-related functions as in the case of *Drosophila*). These results suggest that either the selective pressures to improve/maintain energy-related mitochondrial function in human testes are not as strong as in *Drosophila* or the response to these pressures has been different in the human lineage.

Materials and Methods

Data Collection and Statistical Analysis

We used the online data mining tool BIOMART (http://www.biomart.org, last accessed March 10, 2017; Smedley et al. 2009) to mine the ENSEMBL genome database and extract nuclear genes with mitochondrion annotation (N-mt genes) as specified by the Gene Ontology ID GO: 0005739. We retrieved and used information for human (GRCh38.p3) and *D. melanogaster* (BDGP6) genomes from Ensemble Genes 81 Database including chromosomal position, transcript count, exon count, and protein sequence. Only the longest coding sequence (CDS) of each N-mt gene was used in the following analyses. The final data set included a total of 19,766 genes with 1,640 N-mt genes from human genome and a set of 13,900 genes with 583 N-mt genes from *D. melanogaster* genome. BlastP searches (Altschul et al. 1997) were performed for every protein against every other protein. Genes were clustered into families using a cutoff levels of 50% identity over 50% of the protein length using Markov Cluster Algorithm (http://micans.org/mcl, last accessed March 12, 2017; Enright et al. 2002; Dongen and Abreu-Goodger 2012). This approach was also used to characterize gene families for complete human and *D. melanogaster* genomes. Scripts used to perform these steps are provided in Supplementary Material.

The parent and child relationships were assigned using the GenTree database (http://gentree.ioz.ac.cn, last accessed March 17, 2017) which uses the comparative genomic approaches to infer in what branch the particular gene originated. For the human genome, there are 14 different branches (0–13; branch zero is the oldest branch) where genes assigned dating back to the split with Zebrafish (Zhang et al. 2010b). However, there are seven different branches (0–6; branch zero is the oldest branch here too) in *Drosophila* (Zhang et al. 2010a). When a gene in a gene family is younger than the other closer related copy, this gene is labeled as child and the other as parent. We were not able to assign the child–parent relationship for the genes that originated in the same branch. We consider gene duplication to be tandem if the two genes are on the same chromosome, do not have any type of overlap and the genes are adjacent. We did not assign child–parent relationship for tandem duplications. Retrogenes were identified by looking at the intron number of parental and duplicated genes. Retrogenes were gene copies that had no introns while their parental genes had at least one intron.

The R Stats package (R Development Core Team 2013) and gplots package (Neuwirth 2007; Warnes et al. 2013) were used to perform the statistical analyses and draw heatmap, respectively (http://www.r-project.org, last accessed March 17, 2017). Files needed to draw the heatmaps are provided in the supplementary material.

Expression Data

For *D. melanogaster*, we used the expression data from FlyAtlas (http://flyatlas.org, last accessed March 18, 2017; Chintapalli et al. 2007) and the approach of Gallach et al. (2010) where a gene was considered tissue biased if the

expression level of that gene in a given tissue was higher than the average expression level of that gene in the whole fly (i.e., upregulated in a given tissue) and downregulated in the rest of the adult tissues following the nomenclature from FlyAtlas (Chintapalli et al. 2007).

The expression data for human tissues was downloaded from Human Protein Atlas website (http://proteinatlas.org, last accessed March 18, 2017; Uhlén et al. 2015) and assigned to each gene in the human data set. The assignations of Uhlén and colleagues (Uhlén et al. 2015) were used to assign tissue- biased expression to our gene set. Uhlén and colleagues de- fined five different tissue expression categories. A gene is "Tissue Enriched", if the mRNA level of that gene is at least 5-fold higher than in all other tissues. If mRNA level in a par- ticular tissue is at least 5-fold higher than the average levels in all tissues, that gene is called "Tissue Enhanced" and if the mRNA levels is at least 5-fold higher in a group of 2–7 tissues, that gene is "Group Enriched". In addition, if the gene is expressed in all tissues and the mRNA level of that gene is <1, that gene is placed in "Housekeeping" category. Genes that did not belong to neither of these categories were con- sidered "Mixed". We considered that a gene has tissue- biased expression if it belongs to Tissue Enriched, Group Enriched or Tissue Enhanced categories.

Gene Ontology Analysis

The gene ontology (GO) analysis was performed using the FatiGO software (Al-Shahrour et al. 2004). All N-mt genes in families were analyzed against the background set of the entire single N-mt genes using the two unranked lists of genes anal- ysis mode. GOslim, GO overview, is reported for this analysis.

Results and Discussion

Duplication of N-Mt Genes

Strong selective pressures appear to be at work in the fly genome to retain new N-mt gene duplicates with testis- biased expression (Gallach et al. 2010). We argue that these pressures should be at work in other genomes and below, we describe features of N-mt duplications in the human genome. To make sure that the results of this study are comparable to those of *Drosophila*, we analyzed the human genome and reanalyzed the *D. melanogaster* genome with the same pipe- line. This pipeline introduces minor modifications to the one used by Gallach and colleagues (Gallach et al. 2010; See Materials and Methods). The online ENSEMBL genome data- base was used via the BIOMART data-mining tool. We re- trieved N-mt protein-coding genes for the human and the *D. melanogaster* genomes; 1,640 and 583 genes, respectively (See Materials and Methods for more details and supplementary tables 1–4, Supplementary Material online). Similarities between proteins were assessed using BlastP and proteins were clustered into families using a minimum of 50%

amino acid identity and a Markov Cluster Algorithm as de- scribed in the Materials and Methods. To see if the patterns of N-mt duplications differ from those observed for gene dupli- cations in the whole genome, we applied the same approach to characterize gene families for the complete human and *D. melanogaster* genomes (table 1 and supplementary tables 5– 8, Supplementary Material online). The reanalysis of *Drosophila* genes produced results that were consistent with the previous study (Gallach et al. 2010; See table 1).

We first compared N-mt duplications with duplications in the whole genome. We find that 23% of N-mt genes and 17% of all genes in the genome cluster into gene families in *Drosophila*. This trend is reversed in humans, where 18% of N-mt genes and 40.5% of all genes cluster into gene families. These com- parisons show that N-mt genes have not been duplicated/ retained as extensively in the human genome as in *Drosophila*. We used information about the lineage of gene origination as described by Zhang and colleagues (Zhang et al. 2010b) to date the duplication events and to assign parent–child rela- tionship to each duplicated pair (See Materials and Methods for details and supplementary tables 1, 3, 5, and 7, Supplementary Material online for assignments). We found that on average, N-mt duplications are older than duplications in the whole genome both in the human and *Drosophila* genomes (table 1). However, this effect is much more striking in flies where 78% of the N-mt duplicates and 49% of the duplicates in the whole genomes are older than 63 My (time of *Drosophila* genus diversification). In humans, 88% of N-mt duplicates and 64% of duplicates in the whole genome are older than 104.7 My (cut off that separates mammalian dupli- cates from older ones; table 1). This result confirms that *Drosophila* tends to retain its N-mt duplicates for much longer than other duplicates whereas, in the human genome, we observe a trend in the same direction but not as strong.

In *Drosophila*, N-mt duplicated genes are often found on a different chromosome compared with the parental genes (54%; table 1), but this percentage is much lower for other gene duplications in the genome (22%; table 1). This reloca- tion pattern has been proposed to be a feature that facilitates the acquisition of testis expression and increases N-mt gene retention (Gallach et al. 2010). However, in the human ge- nome N-mt duplicates are relocated as often as nuclear gene duplicates (66% vs. 63%; table 1). These relocation patterns could be explained by the age of duplications since young duplicates tend to be found on their parental chromosomes. On average, N-mt intrachromosomal duplications are younger (have larger branch number; (Zhang et al. 2010b) than inter- chromosomal duplications. In humans, average branch num- bers are 1.5 for old versus 3.0 for young duplicated genes. In flies, average branch numbers are 0.23 for old versus 0.38 for young duplicated genes. In flies, this reveals selection for the retention of interchromosomal duplications because chromo- somal rearrangements with time are very rare (i.e., chromo- somal arms known as Muller elements have not suffer major

Table 1

Characteristics of the N-mt Gene Duplicates and Gene Duplicates in the Whole Genome

	Human[a]			Drosophila melanogaster[a]		
Features Compared	N-mt Genes (%)	Whole Genome Gene Set (%)	Inference	N-mt Genes (%)	Whole Genome Gene Set (%)	Inference
Gene duplication						
Genes in gene families	298 (18.2)	8005 (40.5)	N-mt genes have been duplicated less often than nuclear genes ($P < 2.2e$-16).	132 (22.6)	2335 (16.8)	N-mt genes have been duplicated more of- ten than nuclear genes (P ¼ 0.0004).
Duplication events[b]	167	5375		75	1504	
Retrogenes	12 (7.2)	96 (1.8)	RNA-mediated duplica- tions are more prev- alent for N-mt genes (P ¼ 9.03e-05).	26 (34.7)	92 (6.1)	RNA-mediated duplica- tions are more prev- alent for N-mt genes (P ¼2.613e-12).
Duplication age[c]			Duplications in the whole genome are significantly younger than N-mt duplica- tions (P ¼ 5.201e-07).			Duplications in the whole genome are significantly older than N-mt duplica- tions (P ¼ 0.0006).
Older	86 (87.8)	1124 (64.2)		29 (78.4)	430 (48.9)	
Younger	12 (12.2)	626 (35.8)		8 (21.6)	449 (51.1)	
Relocation pattern[d]			N-mt duplicates were not significantly more relocated than nuclear gene dupli- cates (P ¼ 0.4933).			N-mt duplicates have been significantly more relocated than nuclear gene dupli- cates (P ¼4.119e-05).
Same chromosome	40 (33.6)	686 (37.0)		17 (45.9)	972 (77.5)	
Different chromosomes	79 (66.4)	1167 (63.0)		20 (54.1)	282 (22.5)	

N$_{OTE}$.—All of the P values are based on Fisher's exact tests. All of D. melanogaster inferences are consistent with previous observations (Gallach et al. 2010).

[a]The total number of N-mt genes in the human genome is 1,640. The total number of genes in the genome for the genome version used (Ensembl Genes 80, GRCH38.p3) was 19,766. The total number of N-mt genes in the D. melanogaster genome is 583. The total number of genes in the genome for the genome version used (Ensembl Genes 80, BDGP6) was 13,900.

[b]Duplications events were inferred from the number of events needed to explain the number of genes in that gene family. For example, two genes in a gene family requires only one duplication event but three genes requires two duplication events.

[c]An arbitrary 104.7 My cut off (i.e., mammalian duplications vs. older duplication events) was used here for human genome analyses. However, for Drosophila, a 63 My cut off (time of Drosophila genus diversification) was used. Age is from Gentree database (http://gentree.ioz.ac.cn, last accessed March 17, 2017; Zhang et al. 2010).

[d]Only genes with inferred child and parent (See Materials and Methods for more details) were used here. Because we couldn't assign child–parent relationship to the tandem duplications, we didn't consider those in this analysis.

rearrangements in the genus; Gallach et al. 2010; Powell 1997). However, interchromosomal rearrangements are frequent between autosomal chromosomes in vertebrates (Murphy et al. 2005; Kemkemer et al. 2009), and that can lead to an apparent excess of relocated genes for the older gene set. So, the interpretation of this result in the human genome is less straightforward as it could be due to chromosomal rearrangements after gene duplication or initial interchromosomal duplication being retained by selection.

We also studied retrogenes (i.e., gene duplications produced through an mRNA intermediate) and found that human N-mt genes are duplicated/retained through RNA-mediated duplication more often than in the whole genome, but not as often as in Drosophila (7% vs. 2% in human and 35% vs. 6% in Drosophila; table 1). In Drosophila, this ~5-fold excess of retrogenes within N-mt duplicated genes has been interpreted as evidence for relocation as a mechanism that facilitates testis-biased expression and long-term retention (Gallach et al. 2010). In the human genome, we do not see as many N-mt retrogenes and whereas their transcription is testis biased, they are mostly young (See more details and discussion below).

We conclude that N-mt gene duplication in the human genome has not been as extensive as in Drosophila, and these duplicates have not been retained for long periods of time as in flies. In addition, although many N-mt gene duplicates in the human genome relocate to a different chromosome, it is unclear if this feature was selected right after duplication or is a product of chromosomal rearrangements over time.

Expression Analysis of the N-Mt Genes

We studied the expression of the new N-mt genes compared with the parental genes whenever we could assign

child–parent relationship. For *Drosophila*, we used microarray expression data from FlyAtlas (http://flyatlas.org, last accessed March 18, 2017; Chintapalli et al. 2007; See Materials and Methods and supplementary table 3, Supplementary Material online). To study gene expression of N-mt gene duplicates in the human genome, we used expression data (RNAseq, FPKM) from the Human Protein Atlas website (http://proteinatlas.org, last accessed March 18, 2017; Uhlén et al. 2015); See Materials and Methods and supplementary table 1, Supplementary Material online).

In *Drosophila*, we were able to assign child–parent relation- ship (See Materials and Methods and supplementary table 3, Supplementary Material online) for only a fraction (37/75) of the duplication events. Figure 1 shows gene expression profile of new N-mt genes for adult tissues. We find that 76% of N-mt duplicates (28/37) in *D. melanogaster* can be classified as having tissue-biased expression (See Materials and Methods). All of the genes with tissue-biased expression have testis-biased expression (figs. 1 and 2). The rest of the duplicates are either expressed broadly in different tissues (7/37; fig. 1) or do not have expression data (2/37). The pattern of testis-biased expression holds for gene families even in the instances where we cannot assign the child–parent relationship as we observe that 79% of all the gene families have a testis-biased member (supplementary fig. 1, Supplementary Material online).

In the human genome, we were able to assign child–parent relationship (See Materials and Methods and supplementary table 1, Supplementary Material online) for only 99 of the 167 duplication events. Figure 1 shows gene expression profile of new N-mt genes. We observe that only 28 (i.e., 28% of these duplications; 28/99) can be classified as tissue-biased while the rest are either housekeeping genes (59.8%) or have mixed/undetected expression in the human genome (11.9%). Among all adult tissues, testis is the tissue with the most tissue-biased N-mt gene duplications (10 out of 28, 35.7%) followed by liver (5 out of 28, 17.8%) and skeletal muscle (4 out of 28, 14.3%; fig. 2). If we consider testis and prostate as male tissues and ovary, fallopian tube and endo- metrium as female tissues, we observed more genes with male-biased expression than with female-biased expression among N-mt duplicates (11 genes vs. 1 gene, respectively) but not significantly more than single copy N-mt genes male or female biased (23 genes vs. 6 genes, respectively; Fisher's exact test; $P = 0.6515$). The testis-biased new genes are often duplicates of housekeeping/mixed parental genes (80%; 8/10; supplementary table 1, Supplementary Material online). These features reveal stronger selection for N-mt testis-biased duplicates in *Drosophila* than in the human lineage where only 10% of the N-mt duplicates are testis biased compared with 76% in flies. They also reveal stronger selection for N-mt tissue-biased duplicates for male functions than female functions.

Our analysis of gene duplication lineage of origin/age shows that human testis-biased new N-mt genes are younger than other tissue-biased N-mt new genes. On average, they originated in more recent branches (supplementary table 1, Supplementary Material online) and are often retrogenes. We observe 12 retrogenes among all new N-mt genes and, interestingly, 7 of them (54%) have acquired testis-biased expression after duplication, which is higher than testis expression of retroposed copies in the human genome in general (~40%; Carelli et al. 2016). So, in the human genome, we observe very few testis-biased new N-mt genes, but when relocated by means of retroposition they acquire testis-biased expression as in *Drosophila*. These testis-biased retogenes tend to be on average younger than any other retrogenes (Human average branch number: 5.6 vs. 3) and also younger than the rest of N-mt duplicates (Human branch number: 5.6 vs. 1.4). In the human genome, some new N-mt genes show biased expression in other tissues such as liver or skeletal muscle revealing tissue specialization of mitochondrial function that does not seem to exist in flies. New N-mt genes with tissue-biased expression in flies are only biased for testis expression out of 17 different adult tissues (fig. 1). This tissue specialization in humans should lead to additional levels of interactions and coevolution between nuclear and mitochondrial genes (Wolff et al. 2014).

Functional Enrichments of N-Mt Duplicate Genes in the Human Genome

Since N-mt duplicates in *Drosophila* are enriched for energy-related functions (Gallach et al. 2010), we studied the functional enrichments of N-mt duplicate genes set (i.e., genes that are in gene families) in the human genome. We tested the GO enrichment of N-mt duplicated genes against the single copy N-mt genes as was previously done in flies. Based on GOSlim GOA database, transport GO terms such as transmembrane transport (GO:0055085; 28% vs. 8%; $P = 4.5e-6$) are overrepresented among duplicated genes. Also, carbohydrate metabolic process (GO: 0005975; 15% vs. 4%; $P = 0.0014$) and kinase activity (GO: 0016301; 14% vs. 4%; $P = 0.0040$) are overrepresented among the duplicated genes. RNA binding (GO: 0003723; 4% vs. 15%; $P = 0.0420$) is the only underrepresented term in the GOSlim (See the supplementary table 9, Supplementary Material online for all other GOs enrichments). Since testis-biased genes are a small fraction of all duplicates, we explore the GO enrichment among testis-biased duplicates, but there was not any statistically significant GO term. There is, however, not a lot of power to detect effects in this small gene set (10 genes).

N-mt testis duplicates in human genome have diverse functions, but energy- and catabolism-related functions as well as carbohydrate metabolism are well represented. These functions include an ataxin (ATXN3L; deubiquitinating enzyme), a succinyl-CoA:3-ketoacid coenzyme A transferase 2 (OXCT2; protein involved in energy metabolism of ketone bodies), a glycerol kinase (GK2; protein involved in carbohydrate metabolism), two cytochrome c oxidase subunit duplicates (COX6B2 and COX7B2), a ferritin (FTMT; involved in iron metabolism in
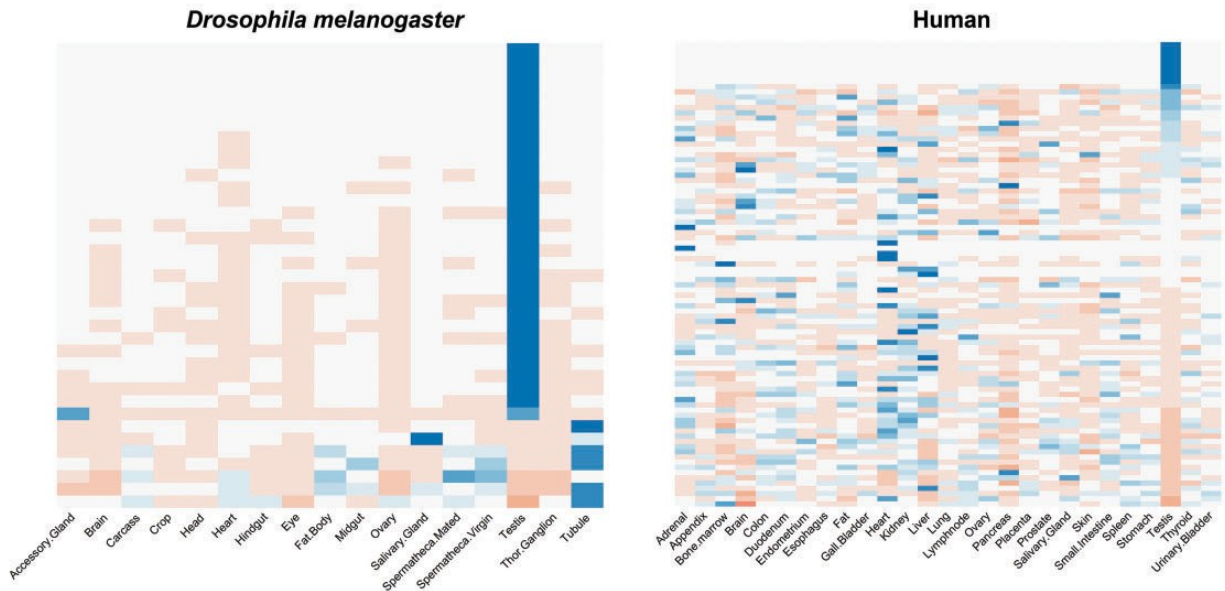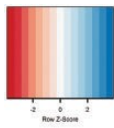
F<small>IG</small>. 1.—Heatmap of gene expression for N-mt duplicated genes from *Drosophila* (left) and human (right). Only new duplicated N-mt genes (i.e., genes for which the child status was confirmed) are shown on the Y axis ranked by their Z score for testis. Only adult tissues are shown on the X axis (27 out of 32 tissues for human, and 17 out of 27 tissues for *D. melanogaster*). Raw expression data was transformed into Z scores to show the extent of tissue-biased expression. Expression levels higher/lower than the mean expression across all tissues have positive/negative Z scores as indicated by blue/pink colors.
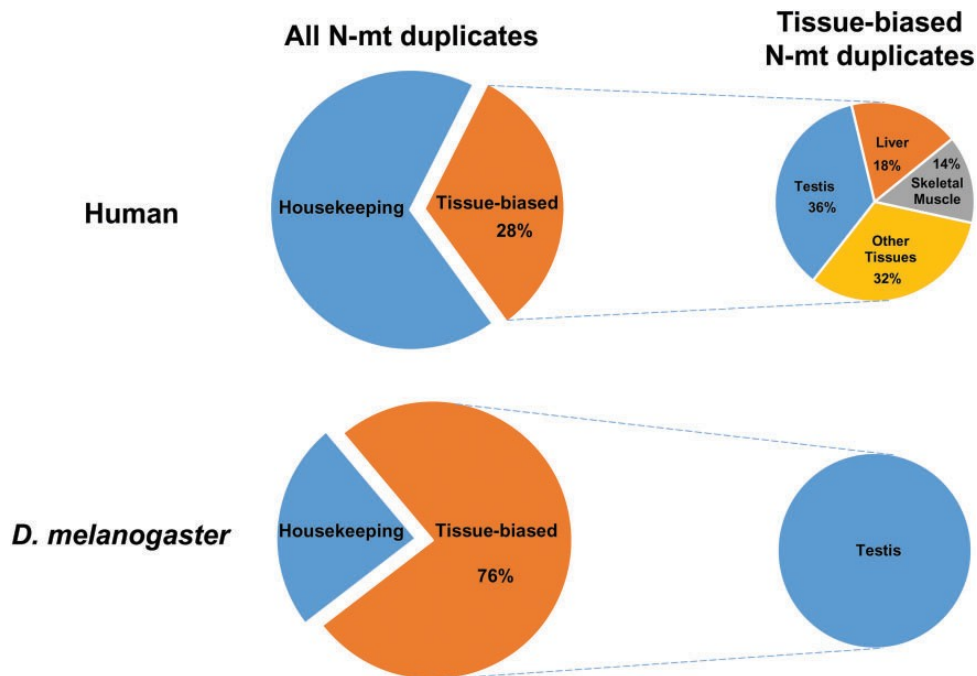
F<small>IG</small>. 2.—Expression pattern of N-mt duplicated genes in human and *D. melanogaster*. Tissue-biased N-mt duplicated genes in human show expression bias in multiple tissues while all tissue-biased genes in *D. melanogaster* are testis biased.

the mitochondria), a protein phosphatase (DUSP21), a pyruvate dehydrogenase (PDHA2), ADP/ATP carrier and a cationic and neutral amino acid carriers. Some of these genes are well known to be testis specific (Dahl et al. 1990; Levi et al. 2001; Tanaka et al. 2001; Hood et al. 2002; Dolce et al. 2005; Fornuskova et al. 2010) and some are known to localize in sperm mitochondria (Tanaka et al. 2001). So, some of these duplicates are likely enhancing energy production in testis and/or sperm.

We also studied if human testis-biased N-mt new genes are enriched for N-mt genes that interact with mt genes. We did this as a first step to test the idea that testis-biased N-mt duplicates are selected to compensate for male-deleterious mutations in mtDNA (Rand et al. 2004; Rogell et al. 2014). If this is the case, N-mt genes that duplicate might comprise genes that encode for proteins which interact with mt genes. There are 13 protein-coding mt genes and they are part of four out of five OXPHOS complexes (complexes I, III, IV, and V; Bar-Yaacov et al. 2012). We used the online Mitominer database for human (http://mitominer.mrc-mbu.cam.ac.uk, last accessed May 15, 2017; Tripoli et al. 2005; Smith and Robinson 2016) and MitoDrome for *D. melanogaster* (Tripoli et al. 2005) and found only 2 genes out of 10 testis-biased new genes in humans and 7 out of 28 testis-biased new genes in flies are duplicates of parental genes in the OXPHOS complexes. The level of interaction of new genes with OXPHOS complexes is not significantly different from testis and non-testis new genes (2/10 vs. 10/89; Fisher's exact test, P ¼ 0.3481) in the human genome nor in *Drosophila* (7/28 vs. 5/9; Fisher's exact test, P ¼ 0.1161). There might be, however, many ways in which testis-biased N-mt duplicates could be selected to compensate for male deleterious mutations in mtDNA (e.g., potentially any gene that increases the flux through complexes), and a negative result might not be enough to reject the hypothesis.

Concluding Remarks

Here, we studied N-mt gene duplications in the human genome to understand if the extent of their contribution to male germline mitochondria-specific functions is similar to that in *Drosophila*. In *Drosophila*, 76% of N-mt duplicates have tissue-biased expression, all of which are expressed in testis. In contrast, 28% of human N-mt duplicates have tissue- biased expression, and 36% of these tissue-biased duplicates have testis-biased expression. Expression of other human N- mt duplicates is biased in multiple other tissues, including liver and skeletal muscle (fig. 2). In addition, human tissue-biased N-mt duplicates have diverse functional annotations and are not enriched with genes with energy-related functions as is the case in *Drosophila*. This result suggests that selective pressures that shape tissue-biased mitochondrial function are different between species. Alternatively, the two species might respond differently to similar selective pressures.

Fertilization and male–male competition for fertilization is expected to drive changes in male germline mitochondria that would improve sperm quality or quantity (Cummins 2009; Pizzari and Parker 2009) These forces are probably stronger in *Drosophila*, due to larger effective population size and stronger sperm competition (Price et al. 1999; Simmons et al. 2004; Pacey 2009; Pizzari and Parker 2009), and may account for the observed differences in recruitment of testis-biased N-mt duplicates. On the other hand, sperm tail motility is powered differently in flies and humans: mammals do not rely on oxidative phosphorylation as much as on glycolysis and have mitochondria only in the midpiece of the sperm tail (Cummins 2009) and not along the sperm tail like *Drosophila* (Noguchi et al. 2011). In mammalian sperm, sev- eral testis-biased gene duplicates of glycolytic enzymes appear to have unique characteristics to localize along the sperm tail to provide energy for its motility (Krisfalusi et al. 2006). Multiple glycolytic enzymes are tightly bound to the fibrous sheath of mouse spermatozoa (Krisfalusi et al. 2006). These include aldolase A isozymes that are generated by alternative splicing and retrotransposition, and duplicated genes of glyceraldehyde-3-phosphate dehydrogenase, and lactate dehydrogenase A (Krisfalusi et al. 2006; Vemuganti et al. 2007). In addition, mouse *Phosphoglycerate kinase 2* (*Pgk2*) is a duplicate of *Pgk1* and is essential for sperm motility and male fertility (Danshina et al. 2010), Thus, mammals appear to improve sperm motility through recruitment of new testis-biased glycolysis genes rather than male-biased N-mt duplicates. In *Drosophila*, several testis-biased duplicates for glycolytic enzymes (*CG5432, CG7069, CG17645, CG32849, CG7024*, and *CG9961*) have been identified, but they have not been found in the sperm proteome (Dorus et al. 2006; Wasbrough et al. 2010) and if there is a role for these genes in sperm motility, it is still unknown. This is not the case for N-mt male-biased duplicates where 17 out of 28 gene products have been detected in the sperm proteome (Dorus et al. 2006; Wasbrough et al. 2010).

Additional species should be studied to understand if en- ergy generation pathways in sperm shape the recruitment of testis-biased N-mt duplicates. For now, the reported pattern indicates that N-mt duplicates contribute to testis-specific MT function to a greater extent in *Drosophila* than in humans. A recent study in *Drosophila* demonstrated that the replacement of mitochondria with mitochondria from a separate lineage can have extreme male fertility consequences (Innocenti et al. 2011). This effect is attributed to the accumulation of male-harming mutations in mitochondrial DNA such that introduction of foreign mtDNA disturbs coevolved nuclear-mitochondrial networks responsible for male fertility. If the number of N-mt testis-biased genes reflects the extent to which male germline mitochondria differ from the mitochondria of somatic cells, we would predict a weaker effect of mitochondria replacement on male fertility in humans compared with *Drosophila*.

1. Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

2. Acknowledgments

3. Literature Cited

Al-Shahrour F, D'ıaz-Uriarte R, Dopazo J. 2004. FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. Bioinformatics 20(4):578–580.

Altschul SF. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25(17):3389–3402.

Bai Y, Casola C, Feschotte C, Betran E. 2007. Comparative genomics reveals a constant rate of origination and convergent acquisition of functional retrogenes in Drosophila. Genome Biol. 8(1):R11.

Bar-Yaacov D, Blumberg A, Mishmar D. 2012. Mitochondrial-nuclear co-evolution and its effects on OXPHOS activity and regulation. Biochim Biophys Acta 1819(9–10):1107–1111.

Berg OG, Kurland CG. 2000. Why mitochondrial genes are most often found in nuclei. Mol Biol Evol. 17(6):951–961.

Carelli FN, et al. 2016. The life history of retrocopies illuminates the evolution of new mammalian genes. Genome Res. 26(3):301–314.

Chintapalli VR, Wang J, Dow JA. 2007. Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nat Genet. 39(6):715–720.

Cummins J. 2009. Sperm motility and energetics. In: Birkhead TR, Hosken DJ, Pitnick S, editors. Sperm biology: an evolutionary perspective. Oxford, UK: Elsevier. p. 185–206.

Dahl HH, Brown RM, Hutchison WM, Maragos C, Brown GK. 1990. A testis-specific form of the human pyruvate dehydrogenase E1 alpha subunit is coded for by an intronless gene on chromosome 4. Genomics 8(2):225–232.

Danshina PV, et al. 2010. Phosphoglycerate kinase 2 (PGK2) is essential for sperm function and male fertility in mice. Biol Reprod. 82(1):136–145. Dolce V, Scarcia P, Iacopetta D, Palmieri F. 2005. A fourth ADP/ATP carrier isoform in man: identification, bacterial expression, functional characterization and tissue distribution. FEBS Lett. 579(3):633–637.

Dongen S, Abreu-Goodger C. 2012. Using MCL to extract clusters from networks. Methods Mol Biol. 804:281–295.

Dorus S, et al. 2006. Genomic and functional evolution of the *Drosophila melanogaster* sperm proteome. Nat Genet. 38(12):1440–1445.

Emerson JJ, Kaessmann H, Betran E, Long M. 2004. Extensive gene traffic on the mammalian x chromosome. Science 303(5657):537–540.

Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30(7):1575–1584.

Fornuskova D, Stiburek L, Wenchich L, Vinsova K, Hansikova H, Zeman J. 2010. Novel insights into the assembly and function of human nuclear-encoded cytochrome c oxidase subunits 4, 5a, 6a, 7a and 7b. Biochem J. 428(3):363–374.

Gallach M, Betran E. 2011. Gene duplication might resolve intralocus sexual conflict. Trends Ecol Evol. 26(11):558–559.

Gallach M, Chandrasekaran C, Betran E. 2010. Analyses of nuclearly-encoded mitochondrial genes suggest gene duplication as a mecha- nism for resolving intralocus sexual antagonistic conflict in Drosophila. Genome Biol Evol. 2:835–850.

Gray MW. 2015. Mosaic nature of the mitochondrial proteome: Implications for the origin and evolution of mitochondria. Proc Natl Acad Sci U S A. 112(33):10133–10138.

Hood KL, Tobin JF, Yoon C. 2002. Identification and characterization of two novel low-molecular-weight dual specificity phosphatases. Biochem Biophys Res Commun. 298(4):545–551.

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classi- fying and distinguishing between models. Nat Rev Genet. 11(2):97–108.

Innocenti P, Morrow EH, Dowling DK. 2011. Experimental evidence sup- ports a sex-specific selective sieve in mitochondrial genome evolution. Science 332(6031):845–848.

Kemkemer C, et al. 2009. Gene synteny comparisons between different vertebrates provide new insights into breakage and fusion events dur- ing mammalian karyotype evolution. BMC Evol Biol. 9:84.

Krisfalusi M, Miki K, Magyar PL, O'Brien DA. 2006. Multiple glycolytic enzymes are tightly bound to the fibrous sheath of mouse spermato- zoa. Biol Reprod. 75(2):270–278.

Levi S, et al. 2001. A human mitochondrial ferritin encoded by an intron-less gene. J Biol. Chem. 276(27):24437–24440.

Montooth KL, Meiklejohn CD, Abt D, Randa DM. 2010. Mitochondrial-nuclear epistasis affects fitness within species but does not contribute to fixed incompatibilities between species of Drosophila. Evolution 64(12):3364–3379.

Murphy MP. 2009. How mitochondria produce reactive oxygen species. Biochem J. 417(1):1–13.

Murphy WJ, et al. 2005. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science 309(5734):613–617.

Neuwirth E. 2007. RColorBrewer: ColorBrewer palettes. R package ver- sion 1.1–2. https://CRAN.R-project.org/package=RColorBrewer; last accessed August, 10, 2017)

Noguchi T, Koizumi M, Hayashi S. 2011. Sustained elongation of sperm tail promoted by local remodeling of giant mitochondria in Drosophila. Curr Biol. 21(10):805–814.

Pacey AA. 2009. 15 - Sperm, human fertility and society A2 - Birkhead, Tim R. In: Hosken DJ, Pitnick S, editors. Sperm biology. London: Academic Press. p. 565–597.

Partridge L, Hurst LD. 1998. Sex and conflict. Science 281(5385):2003–2008.

Pizzari T, Parker GA. 2009. 6 - Sperm competition and sperm phenotype A2 - Birkhead, Tim R. In: Hosken DJ, Pitnick S, editors. Sperm biology. London: Academic Press. p. 207–245.

Porcelli D, Barsanti P, Pesole G, Caggese C. 2007. The nuclear OXPHOS genes in insecta: a common evolutionary origin, a common cis-regulatory motif, a common destiny for gene duplicates. BMC Evol Biol. 7:215.

Powell JR. 1997. Progress and prospects in evolutionary biology: the Drosophila model. New York: Oxford University Press.

Price CS, Dyer KA, Coyne JA. 1999. Sperm competition between Drosophila males involves both displacement and incapacitation. Nature 400(6743):449–452.

R Development Core Team. 2013. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Rand DM, Fry A, Sheldahl L. 2006. Nuclear-mitochondrial epistasis and Drosophila aging: introgression of *Drosophila simulans* mtDNA modi- fies longevity in *D. melanogaster* nuclear backgrounds. Genetics 172:329–341.

Rand DM, Haney RA, Fry AJ. 2004. Cytonuclear coevolution: the genomics of cooperation. Trends Ecol Evol. 19(12):645–653.

Rogell B, Dean R, Lemos B, Dowling D. 2014. Mito-nuclear interactions as drivers of gene movement on and off the X-chromosome. BMC Genomics 15(1):1–9.

Sagan L. 1967. On the origin of mitosing cells. J Theor. Biol. 14(3): 255–274.

Simmons LW, Firman RC, Rhodes G, Peters M. 2004. Human sperm com- petition: testis size, sperm production and rates of extrapair copula- tions. Anim Behav. 68(2):297–302.

Smedley D, et al. 2009. BioMart: biological queries made easy. BMC Genomics 10:22.

Smith AC, Robinson AJ. 2016. MitoMiner v3.1, an update on the mitochon- drial proteomics database. Nucleic Acids Res. 44(D1):D1258–D1261.

Tanaka H, Kohroki J, Iguchi N, Onishi M, Nishimune Y. 2001. Cloning and characterization of a human orthologue of tetis-specific succinyl CoA: 3-oxo acid CoA tranferase (Scot-t) cDNA. Mol Hum Reprod. 8(1):16–23.

Tripoli G, D'Elia D, Barsanti P, Caggese C. 2005. Comparison of the oxi- dative phosphorylation (OXPHOS) nuclear genes in the genomes of *Drosophila melanogaster*, *Drosophila pseudoobscura* and *Anopheles gambiae*. Genome Biol. 6(2):R11.

Uhlén M, et al. 2015. Tissue-based map of the human proteome. Science 347(6220):1260419.

Vemuganti SA, et al. 2007. Three male germline-specific aldolase A iso- zymes are generated by alternative splicing and retrotransposition. Dev Biol. 309(1):18–31.

Waris G, Ahsan H. 2006. Reactive oxygen species: role in the devel- opment of cancer and various chronic conditions. J Carcinogen. 5:14.

Warnes GR, et al. 2013. gplots: Various R programming tools for plotting data. R package version 2.12.1. In: Available online at: http://CRAN.R- project. org/package¼ gplots, last accessed March 17, 2017.

Wasbrough ER, et al. 2010. The *Drosophila melanogaster* sperm proteome-II (DmSP-II). J Proteomics 73(11):2171–2185.

Wolff JN, Ladoukakis ED, Enrʼıquez JA, Dowling DK. 2014. Mitonuclear interactions: evolutionary consequences over multi- ple biological scales. Philos Trans R Soc Lond B Biol Sci. 369(1646):20130443.

Yang D, Oyaizu Y, Oyaizu H, Olsen GJ, Woese CR. 1985. Mitochondrial origins. Proc Natl Acad Sci U S A. 82(13):4443–4447.

Zhang YE, Vibranovski MD, Krinsky BH, Long M. 2010. Age-dependent chromosomal distribution of male-biased genes in Drosophila. Genome Res. 20(11):1526–1533.

Zhang YE, Vibranovski MD, Landback P, Marais GAB, Long M, Barton NH. 2010. Chromosomal redistribution of male-biased genes in mamma- lian evolution with two bursts of gene gain on the X chromosome. PLoS Biol. 8(10):e1000494.

Associate editor: Belinda Chang

Chapter 2

Functional Analysis of Nuclear-Encoded Mitochondrial Gene Duplicates with Testis-Biased Expression in *Drosophila melanogaster*

Abstract

Most of the tissue-specific duplicated nuclear-encoded mitochondrial genes (N-mt genes) in *Drosophila melanogaster* have acquired testis-biased expression. These genes appear to be older and relocated more often than other duplicated genes and have energy-related functions. Since males are under intense male-male competition pressures to fertilize females' eggs and do not pass the mitochondria to the offspring, selection might favor the replacement of N-mt genes in mitochondria during spermatogenesis or in sperm with high-energy production paralogs even though they might also result in higher reactive oxygen species production. Mitochondria elongate all along the sperm tail and are structurally very different to somatic mitochondria after the fusion of all mitochondria of the haploid cell into two huge derivatives, and there might be a need for a different set of N-mt genes in this differentiated mitochondria. Given these potential functions of the testis-specific paralogs, we predict that they have a role in male fertility. To test this hypothesis, we knocked down 39 N-mt duplicated genes with testis-biased expression in *D. melanogaster*. Multiple RNAi lines and few TRiP-KO lines were used to study the fertility effects of these genes. RNAi driven with the germline driver *Bam-Gal4* for these N-mt duplicated genes often produced male sterility. Seventy-two percent of TRiP-KO lines confirmed the RNAi results. We failed to find evidence of the parental gene compensating for the N-mt duplicated gene the fertility effects even when they overlap in expression. These effects along with the different pattern of expression that leads to the presence of these genes in the sperm proteome/mitochondria, their protein divergence, positive selection inferred for some of these genes and likely additional posttranslational processing required for some of these genes suggest that the new duplicate genes have different function than their parental genes during spermatogenesis and/or in sperm.

Keywords: nuclear-encoded mitochondrial genes, RNAi, gene duplication, TRiP-CRISPR, male-biased expression.

Introduction

The vast majority of cellular energy is provided by mitochondria (Wallace 2005; Wolff, et al. 2014). The eukaryotic cells are dependent on the mitochondria, and this cellular organelle cannot survive independently making this a symbiotic partnership that started 1.5 billion years ago between an eubacterium and an archea cell (Martin, et al. 2015; Martijn, et al. 2018). As a consequence of this origin, the mitochondrion is different from other organelles in the eukaryotic cell, with the exception of the chloroplast, because it has its own genome (mtDNA) which likely originated from an alpha-proteobacteria (Lane and Martin 2010; Gray 2012) but see (Martijn, et al. 2018)). Nowadays, mtDNA harbors only a subset of the original alpha-proteobacterial genes and regulatory regions (i.e., 37 intronless genes in the *Drosophila melanogaster* mtDNA) which are essential for mitochondrial activity and a control region which is important for mitochondrial replication and transcription (Garesse and Kaguni 2005). Around 800 genes are encoded by the nuclear DNA that are also required for mitochondrial functions (nuclear-encoded mitochondrial genes or N-mt genes; Calvo and Mootha 2010). The N-mt genes are transcribed in the nucleus, translated in the cytoplasm, and finally imported into the mitochondrion via several import pathways. So far, five different pathways (presequence pathway, carrier pathway, β-barrel pathway, mitochondrial import and assembly (MIA) machinery pathway, and mitochondrial import (MIM) pathway) have been discovered (Wiedemann and Pfanner 2017). So, N-mt genes are responsible for most of the mitochondrial functions, but the interactions between mitochondrial genes and N-mt genes are also fundamental for some mitochondrial functions such as oxidative phosphorylation system (OXPHOS) and mitochondrial-specific protein translational machinery (Levin, et al. 2014) which both genomes encode components of.

Gene duplication involves the formation of a copy of a gene, which could either maintain parental function or acquire a new function (Ohno, et al. 1968). The analysis of N-mt duplicated genes in *D. melanogaster* has shown a very striking pattern (i.e., 54% of N-mt duplicates show tissue-biased expression, and all N-mt duplicated genes with tissue-biased expression are testis-biased genes, 39 genes, while the parental genes are highly expressed in every tissue; Gallach, et al. 2010). This could be seen as unexpected, given that the mitochondrial functions are needed in every cell type. These N-mt duplicated genes are present in all 12 sequenced *Drosophila* species and are older than the duplicated genes from the entire genome (Gallach, et al. 2010). They are also enriched for energy-related functions and overrepresented on autosomes. All together, these features support that these genes have been under strong selection for their location, function and pattern of expression in the *Drosophila* genome (Gallach, et al. 2010). The N-mt duplicated genes have also been studied in human where N-mt duplicated genes did not show the unique expression pattern observed in *Drosophila*. In human, few N-mt duplicated genes express specifically in testis and there is bias for duplicates to have high expression level in liver and muscle cells (Eslamieh, et al. 2017).

Several hypotheses have been proposed about the evolution of these duplicated genes. They might evolve in response to strong selection for different mitochondria during spermatogenesis and/or in sperm that could have led to intralocus sexual antagonism, i.e., the appearance of a new male-beneficial and female-harming allele initially at a single-copy locus. Based on this hypothesis, duplication of the male-beneficial allele accompanied by the acquisition of male-biased expression could resolve this intralocus sexual antagonism (Gallach, et al. 2010; Gallach and Betran 2011b). In particular, since males do not pass the mitochondria to the offspring, the mitochondria in males might specialize for high-energy production despite the potential for high

reactive oxygen species production and associated mutations (Gallach, et al. 2010; Gallach and Betran 2011b). Alternatively, since mitochondria are inherited only from the mother (maternal inheritance) mitochondrial genome mutations can potentially fix even if they are male-harming mutations (Partridge and Hurst 1998; Rand, et al. 2006; Montooth, et al. 2010) and some of these duplicates might have evolved to compensate for mtDNA male-harming mutations that do not hurt females and cannot be selected against (Rogell, et al. 2014). However, these duplicates could also have been retained because they partition a complex pattern of expression, due to their dosage effects or fixed by chance and eventually undergo specialization for male germline or sperm functions (Innan and Kondrashov 2010).

Interestingly, many of these N-mt duplicated genes with testis-biased expression are found in the *Drosophila* sperm proteome (DSP; Wasbrough, et al. 2010; 24/39; 61.5%) while only a few N-mt parental genes (9/39; 23%) are presented in that database. This observation, the protein divergence (on average parental proteins have 62% identity at the protein level to the duplicated gene), and the positive selection that has been reported for some of these duplicated genes (Pröschel, et al. 2006) suggest that the N-mt duplicated genes with testis-biased expression might have a distinct function from their parental gene in sperm with an important role in male fertility. The duplication of the N-mt parental gene could have allowed the new gene to optimize for a testis-specific function without pleiotropic effects that these changes would have caused for the ubiquitous function of the parental gene.

Although mitochondria are essential organelles for both somatic and germ cell development, mitochondrial shape and localization change immensely during *Drosophila* spermatogenesis (Tokuyasu, et al. 1972). The number and the shape of mitochondria are comparable to the somatic cells mitochondrial in the early stages of spermatogenesis as they are

distinct and isolated in the cytoplasm. After meiosis II in mammals, the axoneme is covered with ring-shaped mitochondria, which form the mid-piece of the mature sperm (Olson and Winfrey 1990). However, in many insects including *D. melanogaster*, mitochondria in spermatids undergo a series of transformations, such as aggregation, fusion, and elongation. With the help of a testis-specific gene, *Fuzzy onions* (*Fzo*), the cytoplasmic mitochondria fuse and form two giant derivatives (Tokuyasu, et al. 1972; Hales and Fuller 1997). These two aggregates wrap around each other and join into a spherical aggregate known as Nebenkern (Pratt 1968). This unique giant structure is adjacent to the nucleus and differentiates into two mitochondrial derivatives which elongate along with the axoneme to form the sperm tail (Tokuyasu, et al. 1972). Microtubules support many of these changes in cytoplasm and interestingly mitochondria play a surprising role in elongation of sperm tail in *Drosophila*. Mitochondria serve as an organizer of cytoskeletal dynamics for shaping sperm morphology where they provide a structural platform for microtubule reorganization to support the robust elongation at the tip of *Drosophila* sperm tail (Noguchi, et al. 2011). So, mitochondrial derivatives, actin bundles, cytoplasmic microtubules, and the axoneme are four major structures that contribute to sperm tail elongation. A growing body of research on male-sterile mutants of testis-biased N-mt duplicates shows different mitochondrial defects which suggest that these genes have a fundamental role in *Drosophila* spermatogenesis (Hales and Fuller 1997; Aldridge, et al. 2007; Riparbelli and Callaini 2007; Noguchi, et al. 2011; Sawyer, et al. 2017). Therefore, most N-mt duplicated genes could play an important role in mitochondrial shaping and performance during spermatogenesis. They also could provide more energy for the long sperm tail in *D. melanogaster*.

The availability of numerous genomic tools for *D. melanogaster* makes this species an excellent model organism to investigate the functions of these genes to try to understand the

evolutionary forces that lead to the retention of these genes in the genome. As a first step to study the function of the N-mt duplicated genes with testis-biased expression, we knocked down some of these duplicates using RNAi driven using the Gal4-UAS system (Brand and Perrimon 1993). We also knocked out some of these N-mt duplicated genes in germline and soma with tissue-specific CRISPR strategy (in vivo CRISPR technique) using TRiP-CRISPR lines and specific Gal4-Cas9 drivers (Housden, et al. 2014). Based on the unique expression pattern of these genes and the essentiality of functional mitochondria for sperm tail elongation and motility (Noguchi, et al. 2011), we hypothesize the N-mt duplicated genes are essential for male fertility and the knockdown of these genes will only lead to sperm mobility and/or spermatogenesis defects. To test this hypothesis, these genes were knocked down with a germline driver, *Bam-Gal4*. Also, *Actin5c-Gal4*, a broad *Gal4* which drives in soma and germline, was used to confirm the knockdown results in the germline.

Materials and Methods

Finding N-mt duplicated genes with testis-biased expression

Previous studies have reported a list of the N-mt duplicated genes with testis-biased expression in *Drosophila* (Gallach, et al. 2010; Eslamieh, et al. 2017). We have selected the candidates for this study carefully based on two criteria; first, the gene should be reported in both of the previous studies and second, RNAi stock/s should be available for that gene. Considering these two criteria, we have studied 39 N-mt duplicated genes with testis-biased expression in *D. melanogaster*.

The Gal4-UAS system to knockdown genes in *Drosophila*

To understand the function of N-mt duplicated genes with testis-biased expression, we have taken advantage of the *UAS-Gal4* system to knockdown these genes. The yeast transcriptional activator, Gal4, could be used in transgenic flies under a fly tissue-specific regulatory region to regulate gene expression of another transgene in a tissue-specific manner as long as the second transgene carries the yeast upstream activating sequence (*UAS*; Brand and Perrimon 1993). We have crossed different transgenic flies carrying various *Gal4* drivers with another strain carrying the UASt-RNAi transgene. Each RNAi transgene contains a hairpin that will be processed into small interference RNAs and will degrade the mRNA of the target gene (i.e., knockdown the expression of the gene).

Transgenic flies with UAS-RNAi constructs (RNAi lines) were obtained from VDRC (Dietzl, et al. 2007). The detailed information about these VDRC RNAi lines is provided in a set of Supplementary tables 1-5 that include tables for the testis-biased genes separated by mitochondria compartment. In *D. melanogaster*, there are two VDRC RNAi libraries (GD and KK) targeting different gene regions providing control for the off-target effects. Both libraries were used when they were available. It is worth mentioning that another critical difference between the GD and KK libraries is the insertion site. In the GD library, the insertions are P-element based transgenes with random insertion sites, whereas the KK library contains phiC31-based transgenes with a single, defined insertion site (40D3; (Dietzl, et al. 2007). A second landing site (30B3) was found for KK RNAi lines (Green, et al. 2014; Vissers, et al. 2016). It has been shown that the insertion in the annotated site (present in ~ 25% of KK lines) generates false positive phenotypes because the site is in the 5'UTR of the *tiptop* gene. We checked all KK RNAi lines by PCR (Vissers, et al. 2016) prior to perform knockdown and six KK lines were dropped from this study

because they had an insertion in the annotated landing site. The *Actin5c-Gal4* and *Tub-Gal4* drivers were obtained from the Bloomington Drosophila Stock Center (BDSC stock # 4414 and 5138, respectively). The *Bam-Gal4* and *nos-Gal4* stocks were received as a gift from Michael Buszczak laboratory at UT Southwestern Medical Center. We have used different control stocks for KK and GD knockdowns. For the GD library knockdowns, we used an isogenic host strain w^{1118}, which is the strain researchers injected the library into (VDRC ID: 60000). Analogously, the host line *y,w[1118];P{attP,y[+],w[3`]}* (VDRC ID: 60100) served as a control for KK library knockdowns.

Flies were raised on standard cornmeal/malt medium at room temperature (25°C). All crosses were performed at room temperature except the crosses that were expressing UAS transgenes under a Gal4 driver that were performed also at two additional temperatures: 27°C and 29°C.

The TRiP-CRISPR to knockout genes in *Drosophila*

Previous studies found that the high efficiency of the CRISPR-Cas9 system to create double-strand breaks in *Drosophila* makes this species a great model organism for editing genome or generating mutations in vivo in particular tissues (Bassett, et al. 2013; Gratz, Cummings, Nguyen, Hamm, Donohue, Harrison, Wildonger and O'Connor-Giles 2013; Kondo and Ueda 2013). Each TRiP-KO (Transgenic RNAi Project (TRiP) Knockout) line expresses a unique single CRISPR guide RNA (sgRNA) which directs a double-strand break in the target gene in the presence of the Cas9 protein. The cross between TRiP-KO lines and a line with a tissue-specific-Gal4 driving a UAS-Cas9 construct (Gal4>Cas9 driver) brings together the Cas9 complex and sgRNA in the specific tissue and knocks out the gene in those specific cells as both Cas9 and the sgRNA are under UAS control. For all the ordered lines, the constructs had been injected into *y v;*

attP40 (genotype: y[1] v[1]; P{y[+t7.7]=CaryP}attP40) stock to create TRiP-CRISPR lines. The

TRiP-CRISPR lines were crossed to two different Gal4>Cas9 drivers. The *Act5C-Gal4>Cas9*

(w[*];P{w[+mC]=Act5C-GAL4}25FO1/CyO;P{y[+t7.7]w[+mC]=UAS-Cas9.P2}attP2/ TM6B,

Tb,Hu) which has a strong ubiquitous expression of Cas9 was used to ubiquitously knockout a

target gene. To knockout the gene of interest in the male and female germline, *nos-Gal4>Cas9*

driver (BDSC ID: 67083; *w[*]; P{w[+mC]=GAL4-nos.NGT}40; P{y[+t7.7] w[+mC]=UAS-*

Cas9 .P2 }attP2) was used. The fertility test was performed on F1 flies. The TRiP-CRISPR lines

were available only for eight N-mt duplicated genes with testis-biased expression (seven TRiP-

KO and one TRiP-OE lines) and were obtained from DRSC/TRiP functional genomics resources

(Supplementary table 6).

Viability and fertility tests on RNAi knockdowns

As mentioned above, all RNAi transgenic lines listed here were obtained from the Vienna

Drosophila Resource Center (VDRC; Supplementary tables 1-5; (Dietzl, et al. 2007). The *Actin5c-*

Gal4 flies (a ubiquitous driver; Ito, et al. 1997) was crossed to the RNAi lines to study the effects

of lowering gene expression/knockdown of the genes broadly. We performed a viability test for

the offspring of that kind of cross. For this test, virgin males and females were collected and kept

for three days to make sure they were mature and then two males were crossed with three females.

On day 5, these original flies were dumped out from the vial, and then the number of offspring

were counted on day 15. All viability crosses were performed at three different temperatures,

including 29°C at which *Gal4* has shown the maximum efficiency (Duffy 2002a) but can be

stressful for the flies, specifically during spermatogenesis (Ben-David, et al. 2015) and two other

temperatures (25°C and 27°C that have lower *Gal4* efficiency).

The *Bam-Gal4* driver (a germline driver; Chen and McKearin 2003) was crossed with the RNAi lines to study the knockdown in male and female germline (fertility test). For this test, one F1 virgin male and two virgin females were kept in the vial for five days, and then the progeny was counted on day 15 for the male fertility analyses. Also, two F1 virgin females were kept with one virgin male for five days and the progeny was counted on day 15 for the female fertility analyses. All the crosses for the fertility test were performed at two different temperatures: 27°C and 25°C. Same stocks (VDRC ID: 6000 and 6100 for GD and KK library, respectively) were used as a control for both viability and fertility test. Data were analyzed with R Stats package (R DevelopmentCoreTeam 2013; http://www.r-project.org) by t-test.

Detecting mitochondrial localization signals

N-mt genes were selected based on the Gene Ontology term for mitochondria (GO: 0005739). A way of confirming that these genes actually are imported to the mitochondria is to check their mitochondrial-targeting signal. Most of the protein with mitochondrial presequences are localized in the mitochondria matrix while a few are localizing in the inner membrane and inter-membrane space (Neupert and Herrmann 2007). TIM and TOM protein complexes detect proteins with presequences and import them through the outer and the inner membranes, respectively (Neupert and Herrmann 2007; Chacinska, et al. 2009; Schmidt, et al. 2010). Several mitochondrial targeting prediction tools such as MitoProt, MitoFates and TPpred2 are available to evaluate the probability of mitochondrial targeting of a protein and predict the mitochondrial localization by analyzing the primary amino acid sequence of the proteins (Claros 1995; Savojardo, et al. 2014; Fukasawa, et al. 2015). We have used the most accurate online mitochondrial prediction tool, Mitofate (http://mitf.cbrc.jp/MitoFates), which uses a specific

algorithm (support vector machine; SVM) to detect cleavable N-terminal mitochondrial-targeting signals (presequences) and the cleavage sites on the amino acid sequences (Fukasawa, et al. 2015). The probability of importing the protein of interest to mitochondria (mitochondria localization) is reported based on the amino acid composition and net positive charge of the amino acid sequences (Supplementary tables 1-5).

Ectopic expression and overexpression of N-mt duplicated genes

The FlyORF stocks (Bischof, et al. 2013) were used to ectopically express and overexpress genes in soma and germline. These stocks have a transgene of a particular gene under the UAS regulatory region. Two different ubiquitous Gal4 drivers (*Actin5c-Gal4* and *Tub-Gal4*) and two germline Gal4 drivers (*Bam-Gal4* and *nos-Gal4*) were used to drive expression of genes in the soma and germline, respectively. The FlyORF stocks were obtained from Zurich ORFeom Project Center and were created by authors using the site-specific ΦC31 integrase and inserted on the same landing site (3R, attP-86Fb; Bischof, et al. 2013). So, in crosses of these two lines (*UAS-Gal4* and FlyORF), the gene of interest is expressed ectopically in vivo (i.e., in tissues where the gene is not expressed) or it is overexpressed in the tissue where it is expressed. Reciprocal crosses with at least three replicates were performed for all experiments. FlyORF stocks were only available for three testis-biased N-mt duplicated genes (*COX4L*, *CG7514*, and *CG1907*). For *Tim17b2* (*CG15257*)*,* the TRiP overexpression (OE) stock was received as a gift from DRSC/TRiP functional genomics resources. The TRiP-OE fly stock expresses two sgRNA which can target the upstream of the transcriptional start site of the gene of interest. When this stock crossed with a specific Gal4 driver (Gal4>dCas9-VPR), the gene of interest is overexpressed in the cellular domain where Gal4 is expressed (Lin, et al. 2015). dCas9 (Cas9 Endonuclease Dead) is a mutant

form of Cas9 that lacks its endonuclease activity but can bind to a specific gene with the help of the gRNAs. The dCas9 activation systems can increase the expression of genes of interest by employing transcriptional activators. dCas9-VPR fusion acts as an activator because it interacts with three transcription factors (VP64, p65, and Rta) targeting the gene of interest (Chavez, et al. 2015). It has been shown that dCas9-VPR is an efficient approach to study the overexpression of genes in vivo (Chavez, et al. 2016).

The overexpression of *Tim17b2* in soma and germline was performed with *tubulin* > VPR (w[*]; P{UAS-3XFLAG-dCas9-VPR}attp40; tub-Gal4/ SM5;TM6B Tb[1]) and *nos* > *VPR* (w[*];P{w[+mC]=GAL4-nos.NGT}40/CyO;P{UAS-3XFLAG-dCas9-VPR}attp2/TM6B,Tb[1]) drivers, respectively.

Presence of N-mt parental genes and N-mt duplicated genes in the sperm proteome

We checked the presence of the protein of the N-mt genes analyzed in this work in the *Drosophila* sperm proteome (DSP; Wasbrough, et al. 2010), to test if this presence is consistent with the fertility effects observed in the RNAi knockdowns and TRiP-KOs. So, we have a count of the presence/absence of the parental gene and N-mt duplicated genes in the DSP and fertility/infertility of the N-mt genes and will try to understand if the effects are milder if the parental gene is present and in some way is able to replace the N-mt duplicated gene.

Results

Distribution of testis-specific duplicates across mitochondria compartments

N-mt duplicated genes with testis-biased expression encode proteins for different mitochondrial compartment such as OXPHOS complexes, mitochondrial membranes' proteins

such as TIM, or TOM and membrane transporter proteins, TCA cycle, redox activity, and protein folding. Here, we have studied the function of 39 testis-biased N–mt duplicate genes as shown in Figure 1. Compared to the other compartments, OXPHOS complexes have the highest number (12/39; 31%) of N-mt duplicated genes with testis-biased expression. Since some OXPHOS and mitochondrial-specific protein translational machinery are encoded in the mt-DNA, the presence of duplicates for the OXPHOS N-mt proteins might appear to provide support for the male-harming mtDNA mutation compensation hypothesis in which the gene duplications would mostly observe in the compartment that has interactions between subunits that are encoded both by mitochondrial DNA (mtDNA), and nuclear DNA (nDNA). However, the lack of duplicates of genes interacting with the mitochondrial-specific replication, transcription and translation machinery, the duplications of genes with functions in other compartments of the mitochondria and the lack of enrichment for N-mt genes that interact directly with mt genes (Eslamieh, et al. 2017) do not support that hypothesis.

Mitochondrial localization signals in N-mt duplicated genes and their parental genes

The high probability of mitochondrial localization signals (MLS) is expected for N-mt proteins of OXPHOS complexes, TCA cycle, and other activities inside mitochondria because they have to be imported to the mitochondria to perform their function in the mitochondrial matrix. Conversely, a weak probability of MLS is expected for the mitochondrial membrane and solute carrier N-mt proteins because they do not need to be imported to the mitochondrial matrix. Consistent with our expectation of high MLS for some N-mt proteins, we found 52% (15/29) of N-mt duplicated genes with testis-biased expression have a strong mitochondrial localization signal (Tables 1-5) and high probability of being imported to the mitochondria (arbitrary cutoff =

0.50). Also, consistent with our expectation of a week MLS for N-mt proteins that don't need to be imported to the matrix, we found 100% (10/10) of N-mt genes of mitochondrial membranes and solute carriers showed week MLS signal. The unexpected substantially weaker MLS for some N-mt proteins does not mean that they are not imported to the mitochondria as they might have unknown recognition motifs to translocate to the mitochondria or have different kinds of internal targeting signals. Also, the MLS analysis of parental N-mt genes with expected high MLS probability (e.g., OXPHOS proteins and TCA cycle proteins) showed that 69% of the N-mt parental protein have a high probability of translocating to the mitochondria. For the expected low MLS proteins (i.e., mitochondrial membranes proteins), all the N-mt parental proteins showed low MLS probability.

We also checked the cleavage sites of the duplicated gene and parental gene in a single-family. We found an extra cleavage site, i.e., *Intermediate cleaving peptidase of 55 kDa* (*Icp55*) cleavage site, in five families where the parental gene has one *MPP* (*Mitochondrial Processing Peptidase*) cleavage site while the duplicated gene gained an extra cleavage site (*CG10664*, *CG1154*, *CG14209*, *CG1633*, *CG18347* have MPP cleavage site while *CG10396*, *CG5389*, *CG2612*, *CG6888*, and *CG12201* have an extra cleavage site). The Icp55 is an enzyme which removes one amino acid residue after the *MPP* cleavage site (Naamati, et al. 2009). This removal has been suggested to be important for the stability of the mitochondrial proteome (Vogtle, et al. 2009).

Knockdown of N-mt duplicated genes in the OXPHOS complexes

N-mt duplicated genes with testis-biased expression in mitochondrial oxidative phosphorylation (OXPHOS) are good candidates to test different hypotheses about the evolution

of new N-mt genes. In the electron transport chain (ETC), the electrons travel from one OXPHOS complex to the next which creates a concentration gradient of protons across the inner membrane that will be the source of energy to make ATP at the ATP synthase complex (Jonckheere, et al. 2012). One of the most important by-products of these cellular reactions is reactive oxygen species (ROS) such as free radical molecules ($O_2-\bullet$, $HO_2\bullet$, $\bullet OH$) and non-free radical oxygen molecules such as H_2O_2 and O_2 (Grisham 1992). The leakage of the electrons in the ETC is the primary source of the ROS production (Finkel 2012). Therefore, OXPHOS genes do not only have an important role in ATP production through the OXPHOS pathway but it is this energy production pathway that is one of the primary sources of ROS production in the cell. Among five OXPHOS complexes, four are composed of subunits encoded by mtDNA (13 subunits) and subunits encoded by nDNA (77 subunits; Table 1). Complex II is the only complex that does not have any mitochondria-encoded subunit (Scarpulla 2008). The number of mtDNA and nDNA genes that are encoded for each complex is shown in Table 1. There are 12 OXPHOS N-mt duplicated genes with testis-biased expression functionally studied here. Interestingly, all of these genes are presented in the DSP (Wasbrough, et al. 2010).

Complex I (NADH dehydrogenase) consists of 7 subunits encoded by mitochondrial DNA (mtDNA) and 35 subunits encoded by nDNA which makes this complex the largest complex of the ETC inside the mitochondria inner membrane. This complex transfers electrons from NADH to the respiratory chain. Four N-mt duplicated genes with testis-biased expression are encoding for subunits of this OXPHOS complex. Surprisingly, knocking down of *ND-24L* and *ND-51L1* with *Actin5c-Gal4* driver showed lethality effects while the other two genes of complex I did not show any viability effect (*ND-51L2,* and *ND-49L*; Table 2). Consistent with our results, knockdown *ND-24L* in somatic muscle with *mef2-Gal4* (drives expression in mesoderm and somatic muscle cells

in larvae stage) was reported lethal (Schnorrer, et al. 2010). However, driving *ND-24L* with other Gal4 drivers such as *elav-Gal4* (drives expression in all postmitotic neurons; Neely, et al. 2010), *pnr-Gal4* (drives expression at all stages in a broad band centred on the dorsal midline; (Mummery-Widmer, et al. 2009), and *tinCΔ4-Gal4* (drives expression in cardiac tissue from early developmental stages through adulthood; Neely, et al. 2010) did not show any viability effect. No viability effect was reported for knockdowns of *ND-51L2* and *ND-49L* with *pnr-Gal4*, *tinCΔ4-Gal4*, and *mef2-Gal4* drivers (Mummery-Widmer, et al. 2009; Neely, et al. 2010; Schnorrer, et al. 2010). Our results also showed that knocking down of complex I N-mt duplicated genes in germline with *Bam-Gal4 driver* had no fertility effect in males or females (Table 2).

Complex II (Succinate dehydrogenase) has four subunits and is the only complex that is solely encoded by nDNA genes. The function of this complex is also unique. This complex is at the intersection of OXPHOS and TCA cycle pathways which makes a functional link between these two essential processes (Cardaci, et al. 2015; Lussey-Lepoutre, et al. 2015). Electrons transfer from succinate clusters to ubiquinone (UbQ) in this complex (Cecchini 2003; Miyadera, et al. 2003). Two testis-biased N-mt duplicated genes are encoding subunits for this complex (*SdhAL* and *SdhBL*). Knockdown of *SdhBL* showed lethality effect with a soma driver, *Actin5c-Gal4* while *SdhAL* knockdown showed semi-lethality (Table 2). Previous knockdown studies with *pnr-Gal4* and *tinCΔ4-Gal4* drivers have not shown any viability effect for both genes (Mummery-Widmer, et al. 2009; Neely, et al. 2010; Schnorrer, et al. 2010). No fertility effect was observed when these genes were knocked down in the germline with *Bam-Gal4* driver (Table 2).

Complex III (Coenzyme Q - cytochrome c reductase/Cytochrome b) catalyzes the electrons transfer from ubiquinol to cytochrome *c* and also is important in protons translocation across the mitochondrial membrane (Yu, et al. 2013). Two testis-biased N-mt duplicated genes encode

complex III subunits (*UQCR-14L* and *Cyt-c1L*). The RNAi knockdowns of both genes showed lethality when driven in the soma (Table 2). Consistent with this result, RNAi knockdown of *UQCR-14L* showed larval lethality with *da-Gal4* (another strong ubiquitous driver; Copeland, et al. 2009). Also, a 13% increase in a life span of females was reported when the RNAi of *UQCR-14L* was driven with *tub-GS-Gal4* (a ubiquitous driver). *UQCR-14L* has also been suggested to play a role in complex III assembly (Copeland, et al. 2009). While our results showed male sterility for both genes with *Bam-Gal4* driver (Table 2), the previous study just reported semi-female sterility for *UQCR-14L* using *tub-GS-Gal4* (Copeland, et al. 2009). The male sterility phenotype of *Cyt-c1L* was confirmed with TRiP-KO line of this gene when it was driven with *nos-Gal4>Cas9*. However, the semi-lethality of *Cyt-c1L* RNAi knockdown was not confirmed by TRiP-KO of this gene when it was driven with *Act5C-Gal4>Cas9* as this line didn't show any viability effect.

Complex IV (cytochrome c oxidase) is a multisubunit complex encoded by mtDNA (3 subunits) and nDNA (10 subunits). Mitochondrial protein-coding genes (*CO1*, *CO2*, *CO3*) form the functional core of this complex (Capaldi 1990). This complex catalyzes the reduction of oxygen to water and also serves as a major site for OXPHOS regulation (Kadenbach, et al. 2000). *Cyt-c-d* and *COX4L* are two N-mt genes with a testis-biased expression which encode subunits for this complex. Our results showed that knocking down of *Cyt-c-d* in soma caused viability effects (Table 2). Consistent with these results, previous studies did not find any viability effect for *COX4L*, when RNAi of this gene was expressed under different Gal4 drivers (Neely, et al. 2010). However, inconsistent with these results, a *Cyt-c-d* mutant (P element insertion mutant; loss-of-function allele; Cyt-c-d[EP2049]) was viable (Arama, et al. 2003). Both genes showed male sterility effect when we knocked them down in the germline (Table 2). In agreement with these results,

Cyt-c-d mutant (P-element insertion mutant; *Cyt-c-d^{EP2049}*) showed male infertility with defects in the spermatid individualization stage (Arama, et al. 2006). Moreover, another mutant of this gene (P element insertion; *Cyt-c-d^{bln1}*) showed that *Cyt-c-d* is important for caspase activation of cysts and therefore is required for sperm differentiation in *Drosophila* (Arama, et al. 2003). Defects before individualization step in the testes of this mutant were also observed where the major and minor mitochondrial derivatives developed aberrantly and axonemal microtubules failed to carry out polyglycylation in this mutants (Huh, et al. 2004).

Two subunits encoded by mtDNA interact with 14 subunits encoded by nDNA to phosphorylate ADP to ATP in Complex V (ATP synthase; Jonckheere, et al. 2012). *ATPsynbetaL* and *ATPsynCF6L* are two N-mt duplicated genes with testis-biased expression that encode subunits for Complex V. Our results did not find any viability effect when we knocked down these genes in soma driving RNAi with *Actin5c-Gal4* (Table 2). Both genes showed males sterility effect when we knocked them down in the germline with *Bam-Gal4* driver (Table 2). Similar to these results, previous studies showed male sterility phenotype for both genes (Lindsley, et al. 2013; Sawyer, et al. 2017). In particular, two-point mutations in *ATPsynbetaL* showed the same male sterility phenotypes (Lindsley, et al. 2013). It is worth mentioning that the female fertility was also performed for all of the genes for OXPHOS complexes and no changes in female fertility were observed when we knocked down these genes in germline with *Bam-Gal4*.

Knockdown of N-mt duplicated genes of mitochondrial membrane transport

Mitochondrial membranes are another unique feature of this organelle. The mitochondrion is a double-membraned organelle in eukaryotic cells. The outer membrane is made up double phospholipid layers which separate inside of mitochondrial from the rest of the cell. The basic

structure of the outer membrane is similar to the eukaryotic cell membrane as they both have a similar function; regulates the transportation between two sides of the membrane. Like the cell membrane, some transport proteins are integrated into the inner membrane and are called porins. They are permeable to molecules of about ten kilodaltons or less such as ions, nutrient molecules, or ATP. These proteins form some channels in the phospholipid bilayer which are used to transport smaller proteins by diffusion. For larger proteins, the outer membrane has a unique protein complex called translocase of the outer membrane (TOM). Tom70, Tom22, and Tom20 are receptors of preprotein on the surface of TOM complex while Tom40, Tom5, Tom6, and Tom7 create a conducting channel for the import (Hill, et al. 1998; Künkele, et al. 1998). Most of the mitochondrial proteins are made in the cytoplasm and need to enter into mitochondria to perform the function. When TOM receptors receive a precursor with the signal, the TOM40 form a channel which can be used to translocate the precursor through the TOM complex with the help of TOM subunits' binding sites (Hill, et al. 1998; Chacinska, et al. 2005). Therefore, the TOM complex facilitates proteins import through the outer membrane and into the inter-membrane space.

The import through the inner membrane is more complex than the outer membrane and it is only permeable to oxygen, carbon dioxide, and water. All the ETC complexes, the ATP synthetase complex, and transport proteins are located in the mitochondrial inner membrane. The inner membrane folds (cristae) increase the surface area of the structure which makes more room for all the above structures. TIM23 complex and TIM22 complex are two inner membrane complexes that cooperate with TOM complex to import the preproteins into the matrix (Berthold, et al. 1995; Sirrenberg, et al. 1996; Koehler, et al. 1998). Each TIM complex has a different specificity to precursor protein substrates. Proteins with a positive charge and a matrix-targeting signal import to the matrix by TIM23 complex whereas integral inner membrane proteins which

do not have matrix-target signal import by TIM22 complex (Ryan and Jensen 1995; Sirrenberg, et al. 1996; Bauer, et al. 2000). TIM17 which is an essential component of the TIM23 complex (Kubrich, et al. 1994) has two duplicated genes with testis-biased expression. The inner membrane has an additional translocase, OXA1, to mediate the insertion of precursors from the matrix into the inner membrane. Both mitochondrial-encode and nuclear-encoded proteins use this translocase to enter the mitochondria (Hell, et al. 1998). The two mitochondrial membranes create two compartments: inter-membrane space and matrix. The intermembrane space is the region between the inner and outer membranes, which is the place of primary mitochondrial function, oxidative phosphorylation. The matrix is the main place for the citric acid cycle, and it contains all the enzymes for this cycle along with dissolved oxygen, water, carbon dioxide, etc. Intermembrane folds help the matrix components to have access to the membrane in a relatively short time.

Three N-mt duplicated genes with testis-biased expression are encoded for mitochondrial outer membrane proteins (*tomboy20*, *CG4701*, and *tomboy40*). Knockdown of *CG4701* and *tomboy40* in soma with *Actin5c-Gal4* showed lethality effect. Consistent with our results, previous studies showed inviability effect for *tomboy20* using other Gal4 drivers (Mummery-Widmer, et al. 2009; Schnorrer, et al. 2010). Also, similar to our results, driving RNAi of *tomboy40*, with *E22C-Gal4* (a ubiquitous embryonic driver) and *sqh-Gal4* (a ubiquitous driver) caused lethality (Tsubouchi, et al. 2012). Knockdown of *tomboy40* and *CG4701* in germline with *Bam-Gal4* driver caused male infertility with the defect in the individualization stage while no effect was observed for *tomboy20* in the fertility test (Table 3).

Five testis-biased N-mt duplicated genes are encoded for mitochondrial inner membrane (*Tim17a1*, *Tim17b1*, *GC2*, *CG18418*, and *CG7514*). *Tim17a1* and *Tim17b1* knockdowns in soma did not show any viability effect. However, *Tim17b1* knockdown showed male sterility with *Bam-*

Gal4 driver. Consistent with our results, previous studies showed that driving expression of both genes with different Gal4 drivers in soma such as *mef2-Gal4*, *elav-Gal4*, *pnr-Gal4* and *tinCΔ4-Gal4* did not show any viability effect (Mummery-Widmer, et al. 2009; Neely, et al. 2010; Schnorrer, et al. 2010).

Previous studies did not find any viability effect for individuals knockdown of *CG2* using different Gal4 drivers (Mummery-Widmer, et al. 2009; Neely, et al. 2010; Schnorrer, et al. 2010). However, our results showed lethality and no fertility effect when we knocked down this gene in soma and germline with *Actin5-c-Gal4* and *Bam-Gal4*, respectively. The knockdown of *CG18418* in soma showed no viability effect (Table 3) which is consistent with other studies (Mummery-Widmer, et al. 2009; Neely, et al. 2010). Moreover, knockdown of this gene in germline did not result in any fertility effect (Table 3). The *CG7514* mutant (piggyBac insertion) did not show any viability or fertility effect in a previous study (Thibault, et al. 2004) which is similar to our results where no viability or fertility phenotype was observed in knockdown of *CG7514* in soma and germline (Table 3). Consistently, knockouts of *CG18418* and *CG7514* with the Trip-KO lines with *Act5C-Gal4>Cas9* and *nos-Gal4>Cas9* didn't show any viability and fertility effects, respectively (Supplementary table 6).

We did not observe any viability or fertility effects when we knocked down *CG2616* and *CG9582* in soma and germline. The viability effect of *CG2616* has been studied before and contradictory results have been reported in different studies. Partial lethality with *elav-Gal4* (Neely, et al. 2010) and complete viability with other Gal4 drivers (Mummery-Widmer, et al. 2009; Neely, et al. 2010; Schnorrer, et al. 2010) were reported before. The *CG9582* mutant (piggyBac insertion; *CG9582^{f00857}*) did not show any viability or fertility effect in the previous

study (Thibault, et al. 2004), but male semi-sterility was observed when we knocked down *CG9582* in germline with *Bam-Gal4* (Table 3).

Knockdown of N-mt duplicates with testis-biased expression in the tricarboxylic acid cycle

TCA cycle (also known as Krebs cycle or citric acid cycle) is a series of chemical reactions in which stored energy released through Acetyl-CoA oxidation into ATP and CO_2 (Symposia, et al. 1987). In eukaryotic cells, this cycle forms a key part of aerobic respiration which takes place inside the mitochondrial matrix. In each TCA cycle, three NADH, one FADH2, and one ATP molecule are released as energy-containing compounds (Lieberman and Marks 2009). Five testis-biased N-mt duplicated genes are encoded for the TCA cycle (*CG10749, CG10748, CG14740, mAcon2*, and *Fum3*). *CG10748* and *CG10749* are in the same family but showed different viability effect when we knocked them down in the soma with *Actin5c-Gal4* (Table 4). Our results also showed that knockdown of *Fum3* in soma is lethal while knocking down the other genes (*CG14740* and *mAcon2*) did not show any viability effect (Table 4). In previous studies, no viability effect was observed for knockdowns of all TCA cycle N-mt duplicated genes under different Gal4 drivers (Mummery-Widmer, et al. 2009; Neely, et al. 2010; Schnorrer, et al. 2010). Our results also showed that the two sister genes (*CG10749* and *CG10748*) showed different fertility effects. Only knockdowns of *CG10749* and *Fum3* in germline with *Bam-Gal4* driver caused semi-male sterility while knocking down the other genes did not show any fertility effects (Table 4).

Knockdown of N-mt duplicates with testis-biased expression in cell redox homeostasis

One of the most important by-products of cellular reactions is Reactive Oxygen Species (ROS). The term ROS covers both free radical such as $O_2^-\bullet$, $HO_2\bullet$, $\bullet OH$ and non-free radical oxygen species such as H_2O_2, $1O_2$ (Lu, et al. 2010). Although mitochondria are considered to be a major source of ROS production in the cell (Finkel 2012), low levels of ROS could be formed in peroxisomes, the cytochrome P450 system and some inflammatory cells like neutrophils and eosinophils (Kang 2002). The cell is equipped with the antioxidant defense system which helps to restore the redox hemostasis by detoxifying ROS products (Chaudiere and Ferrari-Iliou 1999; Halliwell and Gutteridge 2007). Redox proteins such as thioredoxins (Trxs), peroxiredoxins (Prxs), and glutaredoxins have been found to play crucial roles in the antioxidant defense's system (Du, et al. 2013). All mitochondrial antioxidant enzymes are encoded in the nucleus and must be imported into the mitochondria (Handy and Loscalzo 2012). Expressing the genes that encoded these enzymes would increase the level of the endogenous cell antioxidant (Thomas and Kalyanaraman 1997). The modulation of ROS level is crucial for the cell because different levels of ROS induce different biological responses. While high levels of ROS can cause oxidative stress and leads to many diseases (Lyras, et al. 1997; Sayre, et al. 2001; Jenner 2003), a growing body of evidence suggests that low level of ROS is important in transducing signals and acts as a signaling molecule in many cellular processes such as cell proliferation (Murrell, et al. 1990), gene expression (Schreck, et al. 1991), and apoptosis (Kim, et al. 2002). Therefore, the delicate balance between ROS beneficial and harmful effects is critical for cell and is achieved by mechanisms called "redox regulation". This process protects cells against oxidative/antioxidative stresses and maintains "redox homeostasis" by controlling the cell's redox state (Dröge 2002). Redox hemostasis is achieved when there is a balance between ROS production and elimination. Thus,

the redox status of the cell is not only governed by ROS inducers but also by ROS scavengers (Bouayed and Bohn 2010). It is worth mentioning that the cell keeps the redox balance slightly to the favor of ROS production which means that the cell can tolerate mild oxidative damage (Dröge 2002). Like mammals, *D. melanogaster* lives in an oxidative environment and (Sacktor 1970) with the exemption of glutathione reductase, has all main antioxidant defense enzymes. In the absence of glutathione reductase, thioredoxin system (TrxR) is a major player in glutathione metabolism and recycling GSH (Kanzok, et al. 2001). Mitochondria and cytosolic isoforms of thioredoxin reductase are encoded by a single gene (*TrxR-1*) in *D. melanogaster* (Missirlis, et al. 2002), whereas only one isoform of thioredoxin peroxidase is localized in both compartments (Radyuk, et al. 2003).

Six N-mt duplicated genes with testis-biased expression are classified into three groups in this compartment. The oxi reduction group consists of two genes (*P5CDh2* and *CG32026*). Both genes did not show any viability or fertility effect when they were knocked down in soma and germline, respectively (Table5). The second group, thioredoxin reductase activity, also consists of two genes (*CG6888* and *Trxr-2*) in which neither of them showed viability nor fertility effects (Table5). Similar to these results, TRiP-KO of *Trxr-2* didn't show any viability or fertility effect with *Act5C-Gal4>Cas9* and *nos-Gal4>Cas9*, respectively. Response to DNA damage is the last group with two genes (*CG8517* and *Rpt3R*). Both genes did not show any viability effect, but only *Rpt3R* showed male sterility when was knocked down in germline with *Bam-Gal4* (Table 5). *CG33092* and *CG9475* have not been studied before. Consistent with our results *CG32026, CG11401, CG6888,* and *CG8517* did not show any viability effect when they were knocked down with different Gal4 drivers (Mummery-Widmer, et al. 2009; Neely, et al. 2010; Schnorrer, et al.

2010). Also, similar to our results, a piggyback mutant of *CG6888* (*CG6888^{f03548}*) was reported to be viable and fertile (Thibault, et al. 2004).

Knockdown of N-mt duplicates with testis-biased expression with other mitochondrial activities

Mitochondrial import most of their proteins from the cytoplasm. These proteins could not translocate into mitochondria as folded proteins. Therefore they must be imported as unfolded proteins and then refold inside mitochondria into three-dimensional structures to be functional (Neupert 1997). It has been shown that some proteins fold with the help of *Hsp60* and *cpn10* (mitochondrial homologs of the bacterial chaperones *GroEL* and *GroES*) while other proteins can be folded without any help (Cheng, et al. 1989; Rospert, et al. 1996). *Heat shock protein 60* (*Hsp60*; Chaperonin family) is a very conserved group of chaperones which are famous for their important role in protein folding and organelle localization (Ranson, et al. 1998; Hartl and Hayer-Hartl 2002; Arya, et al. 2007). This family has also been found to involve in some nonchaperonic functions such as apoptosis and cell signaling. (Zhang and Kalderon 2000; Arya, et al. 2007). In *D. melanogaster*, Hsp60 chaperonins, a group which catalyzes the ATP-dependent protein folding, has six genes (*CG9920*, *CG11267*, *Hsp60A, Hsp60B, Hsp60C* and *Hsp60D*). Of these, *CG9920*, *Hsp60B,* and *Hsp60C* are duplicated genes with testis-biased expression. *Hsp60A* is a general chaperonin and is expressed ubiquitously (Perezgasga, et al. 1999). *Hsp60B* is a testis-biased gene and is essential for spermatogenesis (Timakov and Zhang 2001). The other member of this family, *Hsp60C,* has also been shown to be required for early stages in spermatogenesis (Sarkar and Lakhotia 2005). *Hsp60D* is the last member of this family which is essential for modulation of

apoptosis in *D. melanogaster* (Arya and Lakhotia 2008). *CG11267* is another gene in this chaperonins family which also is a parent of *CG9920* and shows ubiquitous expression too.

There are four N-mt duplicated genes with testis-biased expression in protein folding group (*CG9920*, *Hsp60C*, *Hsp60B*, and *Hsc70-2*). Semi-lethality was observed when we knocked down *Hsp60C* and *CG992* in soma with *Actin5c-Gal4* driver (Table 6). Also, three genes (*CG9920*, *Hsp60C,* and *Hsp60B*) showed male sterility while no fertility effect was seen for knockdown of *Hsc70-2* in the germline with *Bam-Gal4* driver (Table 6). The two genes (*CG5265* and *Scsα2*) in the acetyltransferase activity group didn't show any viability or fertility effect in the RNAi knockdown (Table 6). The knockout of *Hsc70-2*and *CG5265* with TRiP-KO lines were in line with the RNAi knockdown results where no viability effect was observed. However, male fertility was decreased in the TRiP-KO of both genes. The previous study confirmed these results where it has shown that knockdown of *CG9920* was lethal with using *mef2-Gal4* driver (Schnorrer, et al. 2010) but it was viable for other Gal4 drivers (Mummery-Widmer, et al. 2009; Neely, et al. 2010). A mutant of *Hsp60C* (Delta2-3 transposase; *Hsp60C[1]*) was also showed a semi-lethality phenotype. Consistent with our results, this mutant also showed complete male sterility. There were not any individualized and motile sperms in the *Hsp60C[1]* testis (Sarkar and Lakhotia 2005). Similar to our results, a mutant of *Hsp60B* (P element insertion; *Hsp60B[06619]*) showed male sterility (Spradling, et al. 1999). Previous studies could not find any viability effect for the knockdown of *Hsc70-2*, *CG5265*, *Scsα2* with different Gal4 drivers (Mummery-Widmer, et al. 2009; Neely, et al. 2010; Schnorrer, et al. 2010).

Two different spermatogenesis defects in knockdown of N-mt duplicated genes in germline

We dissected the testis of infertile males to understand where the defect(s) are occurring during spermatogenesis. In all of the infertile males of knockdowns, the seminal vesicles were empty (Supplementary file 1). Two different kinds of phenotypes/defects were observed in these knockdowns. The knockdown of 13 genes (*UQCR-14L*, *Cyt-c-d*, *COX4L*, *ATPsynCF6L*, *CG470*, *tomboy40*, *CG9582*, *CG10749*, *Fum*, *Rpt3R*, *CG9920*, *Hsp60C,* and *Hsp60B*) showed empty seminal vesicles with a lot of sperms bundles pile up in the testis right before the seminal vesicle. This suggests that the defects affect the individualization step of spermatogenesis, i.e., the sperms are not individualized and do not enter the seminal vesicle (Figure 2). The other defect was observed in a different stage, earlier in spermatogenesis as the elongated bundles were missing in the knockdowns of three genes (*Cyt-c1L*, *ATPsynbetaL*, and *Tim17b1*; Figure 2). Previous mutants of N-mt duplicated genes showed similar spermatogenesis defects to our RNAi knockdowns. Study of *Cyt-c-d* mutant showed that this gene has a role during spermatid individualization in *Drosophila melanogaster* (Arama, et al. 2003). Similar to our RNAi knockdown observations *ATPsynCF6L* mutant shows failure in mitochondrial elongation and individualization in spermatid cysts which caused male sterility (Sawyer, et al. 2017). The mutant of *Hsp60B* and *Hsp60C* also showed an individualization defect during spermatogenesis which is similar to our observations (Spradling, et al. 1999; Sarkar and Lakhotia 2005). However, while the RNAi knockdown of *ATPsynbetaL* showed an early-stage defect in spermatogenesis, the mutants of *ATPsynbetaL* showed the classic male sterility phenotype in which the spermatids fully elongated but fail in the individualization stage (Wakimoto, et al. 2004).

Ectopic expression and overexpression of four N-mt duplicated genes with testis-biased expression

Because the N-mt genes with testis-biased expression have quite differentiated proteins and have been under positive selection in some instances (Pröschel, et al. 2006) and also are highly expressed only in testis, we would predict that the ectopic expression of these genes can help us to understand if they have retained the parental gene function or they have antagonistic effects. The ectopic expression of a new N-mt duplicated gene with male-biased expression is expected to be deleterious or lethal in soma, if it has change function structurally or is involved in higher ROS production. In contrast, overexpression of the parental gene in the soma might not show any viability effect. Moreover, it is also expected that the expression of the testis-biased N-mt duplicated gene in the female germline would cause deleterious effects and infertility. The FlyORF stocks were used for ectopic expression of three genes (*COX4L*, *CG1907*, and *CG7514*) in the soma and overexpression in germline. The overexpression of *COX4L* in soma with *Actin5c-Gal4* and *Tub-Gal4* drivers caused lethality. Overexpression of this gene in germline with *Bam-Gal4* and *nos-Gal4* reduced male fertility, which suggested that a fine-tuning of expression of *COX4L* is necessary for male fertility. The availability of stocks for two genes of a gene family and help us to study the overexpression of parental (*CG1907*) and duplicated gene (*CG7514*). The overexpression of both genes in soma with *Actin5c-Gal4* and *Tub-Gal4* drivers caused lethality while no effect was observed when the overexpression of the genes performed in germline with *Bam-Gal4* and *nos-Gal4* drivers. These results were against our hypothesis, where we expected to see the lethality effect only for the duplicated gene. The observed lethality might occur as a consequence of the fine-tuning of the expression of a gene that might be essential for the proper function of the gene (i.e., a dose-effect for both genes; Prelich 2012). In addition, the

overexpression of a *Tim17b2*, an N-mt duplicated gene with testis-biased expression in the cross of a TRiP-OE stock and *Actin5c-Cas9* caused lethality while the same cross with *nos-Cas9* didn't affect the fertility of the flies.

Discussion

Testis-biased N-mt duplicated genes often show infertility in knockdown/knockout males

Here, we have studied the function of 39 N-mt duplicated genes with testis-biased expression by knockdown and TRiP-KO in the germline. Of N-mt duplicated genes 41% (16/39) showed male infertility when we knocked them down in the germline with *Bam-Gal4* driver. In addition, the TRiP-KO of seven genes showed either no fertility effect, semi-male sterility or complete male sterility phenotype. Seventy-two percent (5/7) of TRiP-KO showed similar phenotypes in viability (lack of it) and fertility test to the RNAi knockdown of those genes (Supplementary table 6). In both RNAi knockdowns and TRiP-KO experiments, female fertility was unaffected. Six genes (*Cyt-c-d*, *ATPsynbetaL*, *ATPsynCF6L*, *CG7514*, *Hsp60C*, and *Hsp60B*) showed male infertility in knockdowns and all of them have been confirmed by six independent null mutants from previous studies (Spradling, et al. 1999; Thibault, et al. 2004; Sarkar and Lakhotia 2005; Arama, et al. 2006; Lindsley, et al. 2013; Sawyer, et al. 2017). These results reveals that RNAi is a good initial approach to understand fertility effects for testis-biased N-mt duplicated genes. However, as mentioned above, the RNAi knockdown of *ATPsynbetaL* showed an early stage defect in spermatogenesis, but the mutants of *ATPsynbetaL* showed individualization failure (Wakimoto, et al. 2004). RNAi off-target effects might explain this difference.

For *CG18418* and *CG7514* no male fertility effect was observed with RNAi but it was observed with TRiP-KO. So, the lack of RNAi effects should be taken with caution. The efficiency

of the knockdown and knockout for those genes should be studied by RT-PCR and PCR, respectively and compared. The lack of male fertility effect for other RNAi experiments might be explained by the narrow expression of the driver in spermatogenesis. These genes might be expressed later in spermatogenesis where *Bam-Gal4* is not driving. Alternatively, there might be mild effects that are difficult to detect.

Association between the presence of protein in DSP and the fertility effect of N-mt knockdowns

Of the 39 N-mt duplicated genes only 59% (23/39) are found in the DSP. This abundance of N-mt genes encoding sperm proteins and the important role of mitochondria in spermatogenesis and fertility could explain why these genes have unique expression pattern as they are essential for sperm biogenesis (sperm motility) and male fertility. Interestingly, twenty percent (19.8%; 20/101) of *D. melanogaster* retrogenes (RNA mediated duplicated genes) encode components of the sperm and seventy-five percent of those (15/20) have been shown to have mitochondrial gene ontology: OXPHOS, TCA cycle, etc., (Wasbrough, et al. 2010). So retrotransposition has greatly contributed to sperm mitochondria function. The close link between mitochondria processes and flagellum development (microtubule structure development) in *Drosophila* sperm suggests that this linkage in essential for spermatid tail development and the disruption of this connection leads to defects in flagellum development (Noguchi, et al. 2011). So, based on these results and the association of nebenkern and microtubules during sperm flagellum elongation, we propose that the duplication of N-mt genes might be a response to this association as longer sperms have different metabolic needs (need more energy for motility). This could also explain why these N-mt duplicated genes have retained in the genome for a long time. Different duplication rate of N-mt genes between

Drosophila could also be explained by the incredible diversification of flagellum length in these taxa. Because of this we wanted to understand if the presence of the protein in the DSP database might be a good proxy for predicting the knockdown fertility results. In this case, we expected to have fertility effect when the gene is presented in the DSP. To test this hypothesis, we counted the number of genes that showed sterility phenotype and also presented in DSP. Only 34% (8/23) of the N-mt duplicated genes in DSP showed sterility effect while of the genes not present in the DSP 50% showed sterility (8/16) which suggests that the presence of N-mt duplicated genes in DSP database does not predict the fertility test results ($P>0.05$ in a Fisher's exact test). This also suggests that the later genes might be important in different stages of spermatogenesis as they do not end up in the sperm proteome.

Compensation of N-mt parental genes for fertility effects of N-mt duplicates

In the knockdown of N-mt duplicated genes, the presence of the parental gene might be enough to prevent the male infertility phenotype in the germline knockdowns (fertility test). To test for this compensatory effect, we looked at duplicated genes that were present in DSP and compared the fertility effects between those that had the parental gene present in the DSP and also not present in the DSP. Around 42% (3/7) of genes showed no fertility effect when the duplicated and parental gene are both present in DSP. While the parental is not consistently compensating for the duplicated genes in the fertility test (rescue the fertility effect of knockdowns), in general, there is no association between the fertility effect and the presence of the parental gene in the DSP (Ratio of male sterility to no fertility effect when parental gene is present in DSP = 4/5; Ratio of male sterility to no fertility effect when parental gene is absence in DSP = 5/4; Fisher exact test; P-value = 0.6231).

Some testis-biased N-mt duplicated genes show viability effects in RNAi knockdowns

Driving RNAi with *Actin5c-Gal4* in soma produced lethality for 36% (14/39) of the genes. This is unexpected because these genes are highly expressed in the adults' testes. Some factors not related to the gene function may influence knockdown results. The RNAi system could be stressing the cells and have some unexpected results in the soma even if the gene is not expressed in those tissues. However, we only observe lethality and fertility effects for some genes and not others which suggest that these observations might be a real effect of driving RNAi for those genes. Off-target and background effects could produce phenotypes that are not due to the genes under study when we knock them down using the Gal4-RNAi system. Confirming the existence of some of these effects, the lethality observed in some knockdowns is not confirmed by our TRiP-KO experiments. Among seven genes for which we had TRiP-KO lines, two showed lethality with some RNAi but not with the somatic drive of TRiP-KO.

However, many of these male-biased genes also appear to be transcribed in imaginal discs and fat body. The moderate to high expression of some these genes in larvae L3 imaginal disc and pupae fat body expression (Supplementary table 7) might contribute to explain the lethality effects of these knockdowns as we found a significant association between the high expression of these genes in larvae imaginal disc and pupae fat body and lethality effect of RNAi knockdown. On average, genes with the RNAi lethality effect have statistically higher larva imaginal disc expression than non-lethal genes (t-test; P-value = 0.0101). The average fat body expression in pupae stage was also statistically higher for the genes with RNAi lethality effect in comparison to non-lethal genes (t-test; P-value = 0.0151). So, the very high expression of *Cyt-c-d*, *SdhBL* and moderate expression of *SdhAL*, *ND-24L*, *Cyt-c1L*, and *ND-51L1* in the imaginal disc of larvae L3

wandering stage and the potential function of these genes in those tissues could explain the lethality phenotype of RNAi knockdowns and support that these genes are essential for larvae stage viability. So, these unexpected viability results need to be confirmed by null mutants as there is not any knockout mutant available for these particular genes yet.

In the past, RNAi has been used to study gene function for young duplicate genes in *Drosophila* (Chen, et al. 2010). While lethality was observed for a big fraction of the genes, this phenotype has been confirmed for some genes (Ross, et al. 2013; Lee, et al. 2019) but not for others (Kondo, et al. 2017). The reason behind this inconsistency is still unclear but off-target effects might explain some of these inconsistencies. As in the N-mt genes studied here, knockouts should be produced for all these genes to understand their functions.

Ectopic expression and overexpression of testis-biased N-mt duplicated genes

Ectopic/overexpression of parental and new N-mt genes can be useful to study the gene function (Bischof, et al. 2013). Also, we hypothesize that N-mt new genes might have evolved to produce more energy for sperm despite higher ROS production and ectopic expression in the soma would be expected to produce more ROS molecules (that would be interesting to study), ageing or lethality. Conversely, the overexpression of the parental genes in the soma would not be expected to produce any harmful effects. Likewise, overexpression of testis-specific N-mt genes in the female germline is expected to show bigger deleterious effects than the overexpression of the parental gene. However, the ectopic expression of the parental gene in testis is expected to be deleterious because of its absence in the DSP. Consistent with our expectation, the overexpression of N-mt duplicated genes in soma caused lethality. However, in contrast to our expectation, the overexpression of N-mt parental genes in soma also caused lethality. We infer that there are dose

effects for those genes. The overexpression of one N-mt duplicate in male germline reduced male fertility while no change was observed for the other N-mt new gene and its parental gene. The female fertility and viability have not been changed in the overexpression of N-mt duplicated genes and parental gene. The lethality effect of the parental and duplicated gene preclude the comparison that we intended (i.e., the comparison of ROS production levels between overexpression of parental and ectopic expression of the new gene in the soma) to understand if the new genes produce more energy despite producing more ROS.

How different are parental genes from new genes?

So far, the study of N-mt duplicated genes with testis-biased expression suggested that the new genes are different from their parental genes. The proteins of the new genes are on average 62% identical to the parental genes and evolve faster than them (Gallach, et al. 2010). Some of the new genes (*CG18418* and *CG6255*) are known to have been under positive selection (Pröschel, et al. 2006) which reveals that they might have been selected for different functions in the male germline. The presence of the parental gene in the DSP does not compensate for the absence of the new gene at least in the knockdown fertility tests. We find that some of the new genes gained an extra MLS signal. This additional cleavage site has been suggested to be important for protein stability within mitochondria by cleaving one amino acid from the MPP-generated intermediate N-terminus (Vogtle, et al. 2009). This suggests that parental and duplicate have different functions.

It is therefore likely that many N-mt duplicate genes were retained in the genome for their fertility and sperm functions. Their recruitment in the genome might coincide with the lineage in which mitochondria started to be used for the production of energy in the sperm tail. Age of the duplications should be assessed to understand this. Antagonistic effects in other tissues for N-mt

duplicate genes should be tested to understand why those genes are so highly testis biased. As the mtDNA is actively eliminated from the developing sperm in *Drosophila* (Patel 2017; Yu, et al. 2017) and it does not pass to the next generation, the ROS production might be one of the sources of antagonism. However, the presence of dose effects when we overexpress parental or ectopically express the new gene may impede those experiments as we have discussed above. The exact function of the duplicated genes will be discovered when the null mutants are generated for these genes and the fertility and viability tests are performed. The rescue of the null mutant phenotypes (parental and duplicate nulls) with the expression of the duplicate or parental genes, respectively, will also help us to elucidate if both genes are different or have the same function (i.e., they can replace each other).

Figures
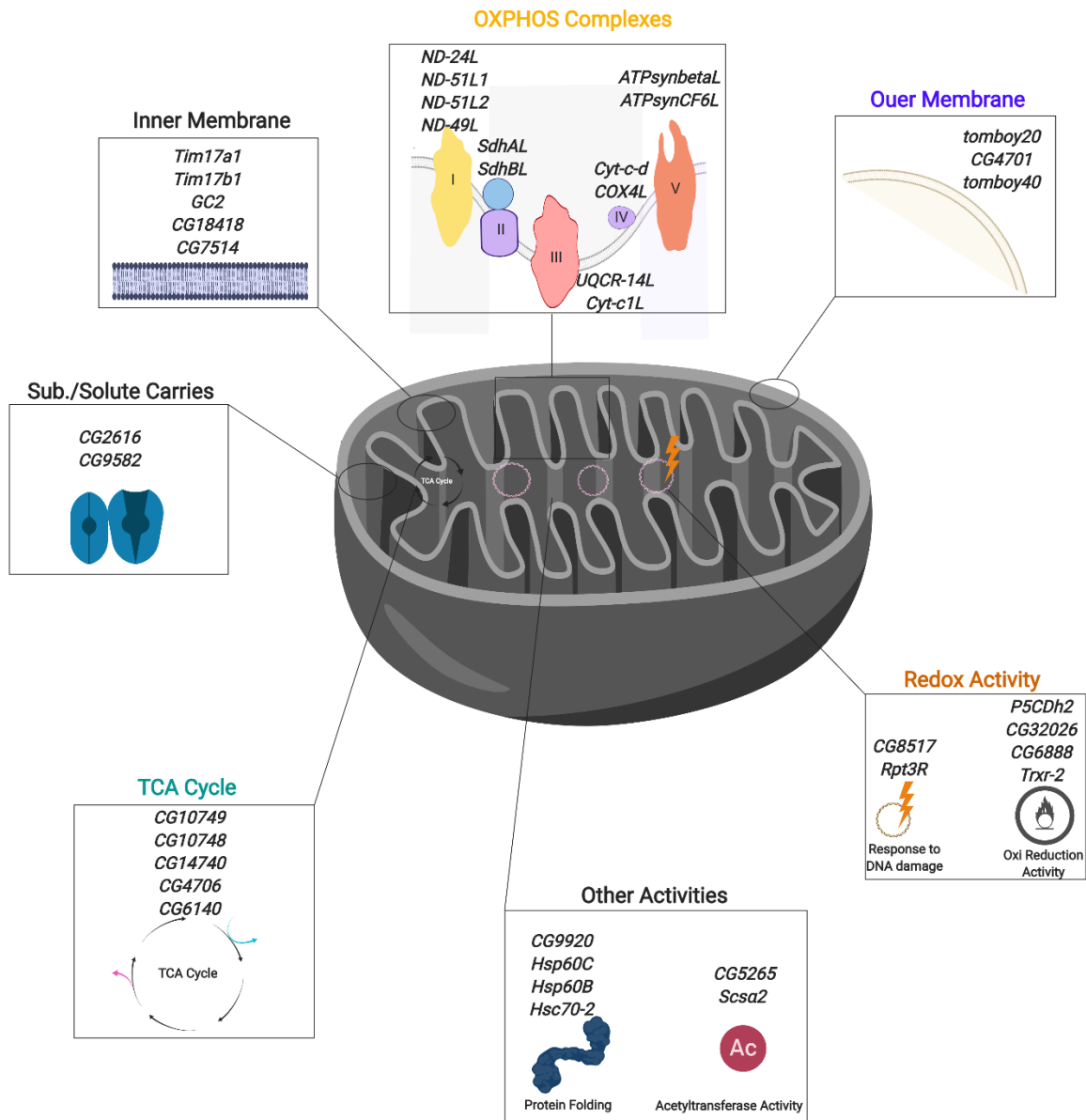
Figure 1. Distribution of N-mt duplicated genes with testis-biased expression studied in this work in different mitochondrial compartments in *D. melanogaster*
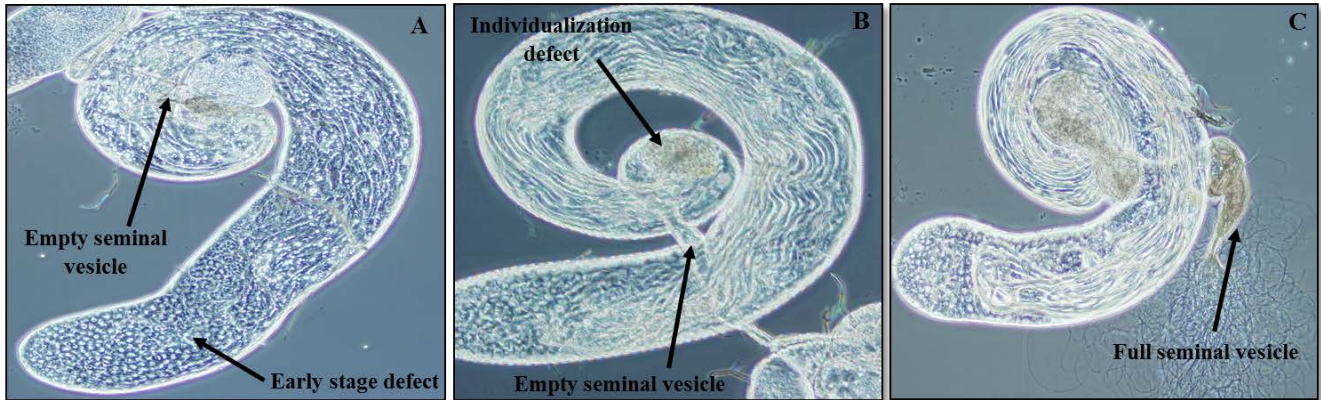
Figure 2. Dissected testis of *D. melanogaster* in the knockdowns with *Bam-Gal4* driver and *w^1118^*. (A) Testis of the knockdown for *Tim17b1* with a defect in the early stage of spermatogenesis is shown here as an example of this kind of phenotype. Notice that the seminal vesicle is empty in this knockdown. (B) Testis of the knockdown for *COX4L* with a defect in the individualization stage of spermatogenesis is shown here as an example of this kind of phenotype. The seminal vesicle of this knockdown is also empty. (C) Testis of wild type fly, *w^1118^* is shown here. Spermatogenesis is normal and the seminal vesicle is full in this fly.

Table 1- Number of genes and subunits for each OXPHOS complex in *D. melanogaster*

Mitochondrial Compartment	Complex I	Complex II	Complex III	Complex IV	Complex V
Subunits encoded by mtDNA	7	0	1	3	2
Subunits encoded by nDNA	35	6	12	10	14
N-mt duplicated genes with testis-biased expression	4	2	2	2	2

Table 2. Knockdown of N-mt duplicates with testis-biased expression in different OXPHOS complexes with *Actin5c-Gal4* and *Bam-Gal4* at three different temperatures and TRiP-KO

Mitochondrial Compartment	Gene CG Number	Gene Name	Mitochondria Localization Probability	Viability Test (*Actin5c-Gal4*)	Fertility Test (*Bam-Gal4*)	Defect in Spermatogenesis
OXPHOS Complex I	CG6485†ɸ3	ND-24L	0.96	Lethal§	No Effect§	-
	CG8102†ɸ	ND-51L2	0.74	No Effect§	No Effect§	-
	CG11423†ɸ	ND-51L1	0.42	Semi-lethal	No Effect	-
	CG11913†ɸ	ND-49L	0.09	No Effect§	No Effect§	-
OXPHOS Complex II	CG5718†ɸ3	SdhAL	0.99	Semi-lethal*§	No Effect*§	-
	CG7349ɸ	SdhBL	0.97	Lethal§	No Effect§	-
OXPHOS Complex III	CG17856†ɸ	UQCR-14L	0.1	Lethal*§	Male sterility*§	Individualization
	CG14508ɸ	Cyt-c1L	0.61	Semi-lethal§	Male sterility§¥	Early Stage
OXPHOS Complex IV	CG13263ɸ	Cyt-c-d	0.02	Semi-lethal*§	Male sterility*§	Individualization
	CG10396†ɸ	COX4L	0.90	No Effect*§	Semi-male sterility*§	Individualization
OXPHOS Complex V	CG5389†ɸ3	ATPsynbetaL	0.22	No Effect	Male sterility	Early Stage
	CG12027ɸ3	ATPsynCF6L	0.96	No Effect§	Semi-male sterility§	Individualization

† Without off-target gene; * Consistent results in both KK and GD libraries; § Consistent results at all temperatures; ɸ New gene presents in the DSP; 3 Parental gene presents in the DSP; ¥ Consistent results with TRiP-KO test.

Table 3. Knockdown of N-mt duplicates with testis-biased expression in mitochondrial membranes with *Actin5c-Gal4* and *Bam-Gal4* at three different temperatures and TRiP-KO

Mitochondrial Compartment	Gene	Gene Name	Mitochondria Localization Probability	Viability Test (*Actin5c-Gal4*)	Fertility Test (*Bam-Gal4*)	Defect in Spermatogenesis
Outer Membrane	*CG14690*[†3]	*tomboy20*	0.03	No Effect	No Effect	-
	CG4701	*CG4701*	0.00	Lethal [§]	Semi-male sterility	Individualization
	CG8330	*tomboy40*	0.00	Lethal	Male sterility[§¥]	Individualization
Inner membrane	*CG10090*[†ϕ]	*Tim17a1*	0.03	No Effect[§]	No Effect[§]	-
	CG1158[†]	*Tim17b1*	0.01	No Effect [§]	Male sterility[§]	Early stage
	CG12201[ϕ]	*GC2*	0.01	Lethal	No Effect[§]	-
	CG18418	*CG18418*	0.03	No Effect[§¥]	No Effect[§]	-
	CG7514[ϕ]	*CG7514*	0.02	No Effect[§¥]	No Effect[§]	-
Mitochondrial Substrate/Solute Carrier	*CG2616*	*CG2616*	0	No Effect[§]	No Effect[§]	
	CG9582[†]	*CG9582*	0.04	No Effect[§]	Semi-male sterility	Individualization

† Without off-target gene; * Consistent results in both KK and GD libraries; § Consistent results at all temperatures; ϕ New gene presents in DSP; ³ Parental gene presents in DSP; ¥ Consistent results with TRiP-KO test.

Table 4. Knockdown of N-mt duplicates with testes-biased expression in TCA cycle with *Actin5c-Gal4* and *Bam-Gal4* at three different temperatures and TRiP-KO

Mitochondrial Compartment	Gene	Gene Name	Mitochondria Localization Probability	Viability Test (*Actin5c-Gal4*)	Fertility Test (*Bam-Gal4*)	Defect in Spermatogenesis
TCA Cycle	*CG10749*[†*]	*CG10749*	0.78	Lethal [§]	Semi-male sterility[§]	Individualization
	CG10748[†]	*CG10748*	0.53	No Effect[§]	No Effect[§]	-
	CG14740[ϕ]	*CG14740*	0.28	No Effect[§]	No Effect[§]	-
	CG4706[†*3]	*mAcon2*	0.53	No Effect [§]	No Effect[§]	-
	CG6140[*3]	*Fum3*	0	Lethal[§]	Semi-male sterility[§]	Individualization

† Without off-target gene; * Consistent results in both KK and GD libraries; § Consistent results at all temperatures; ϕ New gene presents in DSP; ³ Parental gene presents in DSP; ¥ Consistent results with TRiP-KO test.

Table 5. Knockdown of N-mt duplicates with the testis-biased expression for redox activity with *Actin5c-Gal4* and *Bam-Gal4* at three different temperatures and TRiP-KO

Mitochondrial Compartment	Gene	Gene Name	Mitochondria Localization Probability	Viability Test (*Actin5c-Gal4*)	Fertility Test (*Bam-Gal4*)	Defect in Spermatogenesis
Oxi Reduction	*CG33092*[3]	*P5CDh2*	0.81	No Effect[§]	No Effect[§]	-
	CG32026[†φ]	*CG32026*	0.38	No Effect[§]	No Effect[§]	-
Thioredoxin Reductase Activity	*CG6888*[†φ]	*CG6888*	0.01	No Effect[§]	No Effect[§]	-
	CG11401[†]	*Trxr-2*	0.98	No Effect[§¥]	No Effect[§]	-
Response to DNA Damage	*CG8517*[†]	*CG8517*	0.35	No Effect[§]	No Effect[§]	-
	CG9475	*Rpt3R*	0	No Effect[§]	Male sterility[§]	Individualization

† Without off-target gene; * Consistent results in both KK and GD libraries; § Consistent results at all temperatures; φ New gene presents in DSP; [3] Parental gene presents in DSP; ¥ Consistent results with TRiP-KO test.

Table 6. Knockdown of N-mt duplicates with testis-biased expression in other mitochondrial activity with *Actin5c-Gal4* and *Bam-Gal4* at three different temperatures and TRiP-KO

Mitochondrial Compartment	Gene	Gene Name	Mitochondria Localization Probability	Viability Test (*Actin5c-Gal4*)	Fertility Test (*Bam-Gal4*)	Defect in Spermatogenesis
Protein Folding	*CG9920†*	*CG9920*	0.15	Semi Lethal	Male sterility§	Individualization
	CG7235†	*Hsp60C*	0.99	Semi Lethal	Male sterility§	Individualization
	CG2830	*Hsp60B*	0.99	No Effect	Male sterility§	Individualization
	CG7756	*Hsc70-2*	0.01	No Effect¥	No Effect§	-
Acetyltransferase Activity	*CG5265ϕ3*	*CG5265*	0.20	No Effect§¥	No Effect§	-
	CG6255ϕ	*Scsα2*	0.95	No Effect§	No Effect§	-

† Without off-target gene; * Consistent results in both KK and GD libraries; § Consistent results at all temperatures; ϕ New gene presents in DSP; 3 Parental gene presents in DSP; ¥ Consistent results with TRiP-KO test.

Supplementary File 1

Supplementary Figure 1. Testis of W^{1118} (serves as control) reared at 27 °C with the full seminal vesicle
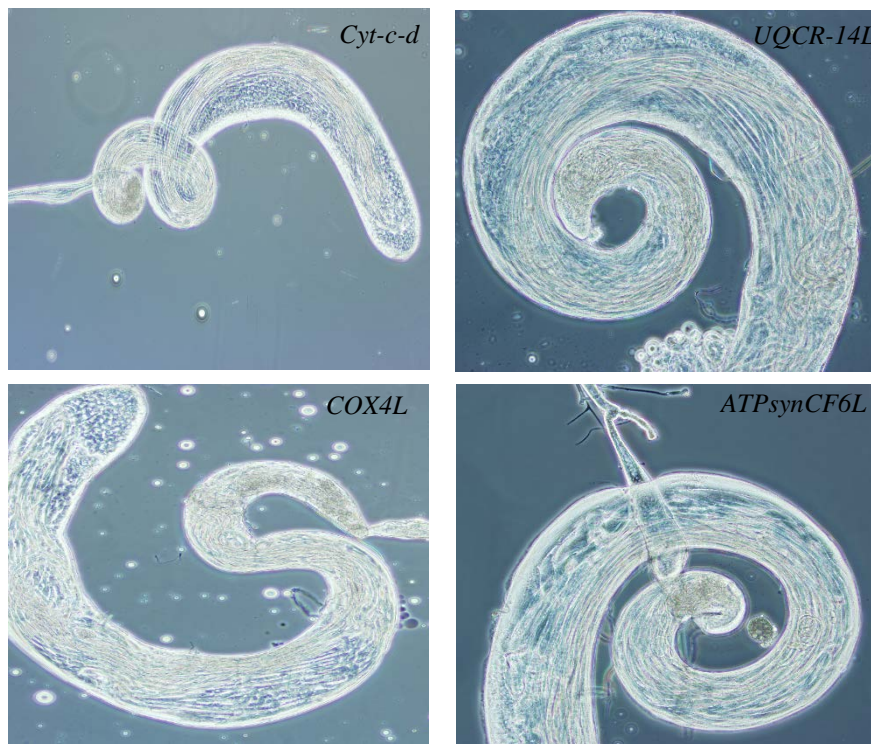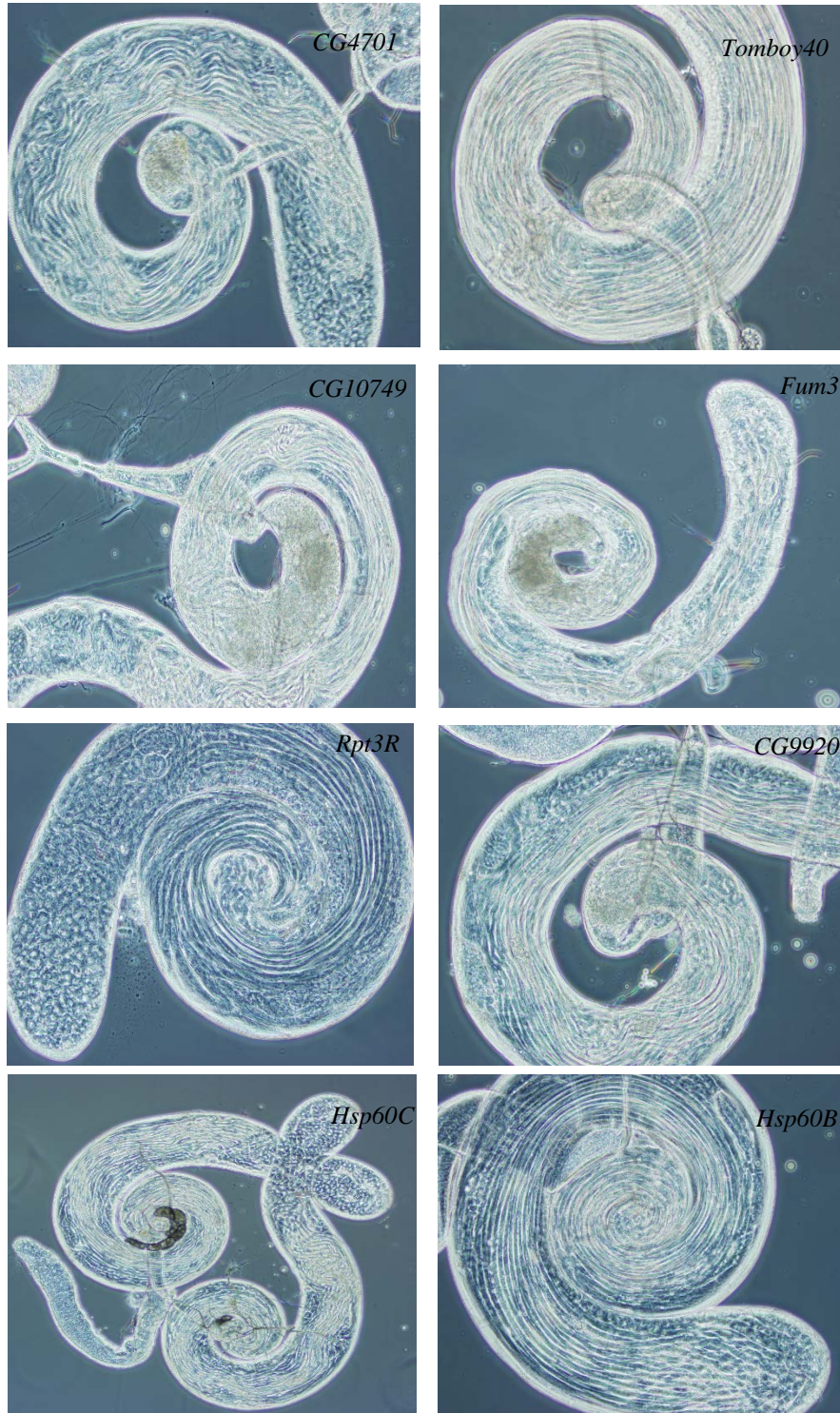
Supplementary Figure 2. Testis of RNAi knockdowns drove with *Bam-Gal4* at 27 °C which has an empty seminal vesicle as the defect was observed in the individualization stage of spermatogenesis
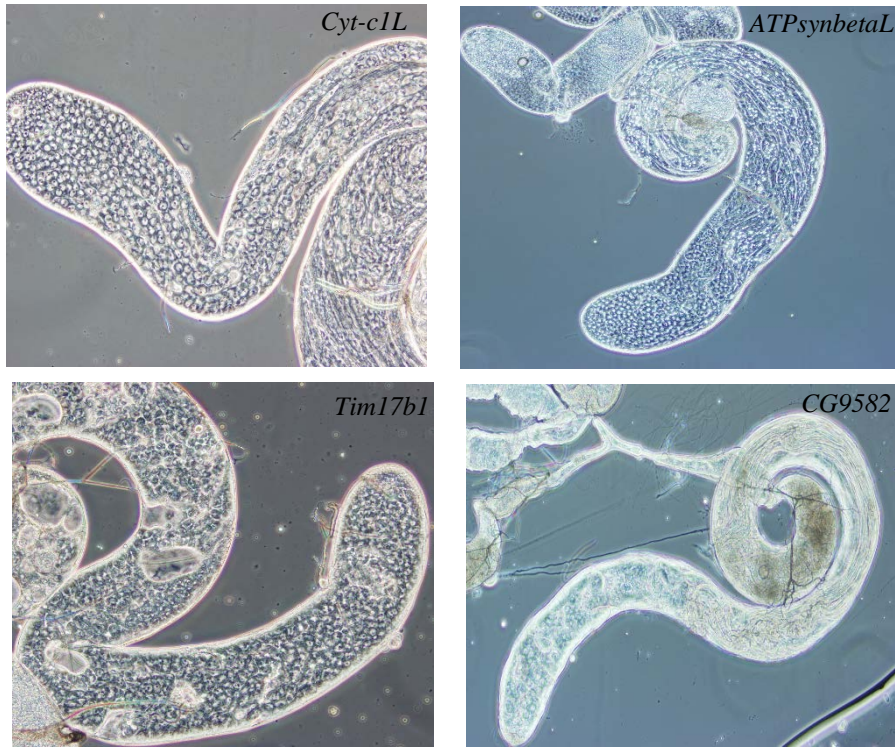
Supplementary Figure 3. Testis of RNAi knockdowns drove with *Bam-Gal4* at 27 °C with the defect was found in the early stages of spermatogenesis.

Supplementary Table 1. Mitochondrial localization probability, DSP status, and knockdown results of N-mt duplicates with the testis-biased expression for OXPHOS complexes

Compartment		Gene	# of libraries	Present in the DSP	Off-Target	Mt Localization probability-Mitofate	Fertility Test (*Bam-Gal4*)*	Parent Present in the DSP	Parent	Mt Localization probability of Parent-Mitofate	Defect Stage in Spermatogenesis	Viability Test (*Actin5c-Gal4*) 27c*
OXPHOS Complex	Complex I	CG6485	2	✓	0	0.96	No Effect	✓	CG5703	0.88	-	Lethal
		CG8102	1	✓	0	0.74	No Effect	X	CG9140	0.45	-	No Effect
		CG11423	1	✓	0	0.42	No Effect	X	CG9140	0.45	-	Semi Lethal
		CG11913	1	✓	0	0.09	No Effect	X	CG1970	0.95	-	No Effect
	Complex II	CG5718	2	✓	0	0.985	No Effect	✓	CG17246	0.99	-	Semi Lethal
		CG7349	2	✓	0	0.97	No Effect	X	CG3283	0.99	-	Lethal
	Complex III	CG17856	2	✓	0	0.1	Male sterility	X	CG3560	0.1	Individualization	Lethal
		CG14508	1	✓	2	0.605	Male sterility	X	CG4769	0.96	Early Stage	Semi Lethal
	Complex IV	CG13263	2	✓	1	0.012	Male sterility	X	CG17903	0.1	Individualization	Lethal
		CG10396	2	✓	0	0.901	Semi-male sterility	X	CG10664	0.98	Individualization	No Effect
	Complex V	CG5389	2	✓	1	0.219	Male sterility	✓	CG11154	0.5	Early Stage	No Effect
		CG12027	1	✓	2	0.96	Semi-male sterility	✓	CG4412	0.88	Individualization	No Effect

* All detailed fertility and viability knockdown results for N-mt duplicated genes are provided in a Supplementary Results Excel file.

Supplementary Table 2. Mitochondrial localization probability, DSP status, and knockdown results of N-mt duplicates with the testis-biased expression for mitochondrial membranes

Compartment		Gene	# of libraries	Present in the DSP	Off-Target	Mt Pro. Localization probability-Mitofate	Fertility Test (*Bam-Gal4*)*	Parent Present in the DSP	Parent	Mt Pro. Localization probability of Parent-Mitofate	Defect Stage in Spermatogenesis	Viability Test (*Actin5c-Gal4*) 27c*
Membrane	Outer Membrane	CG14690	2	X	0	0.033	No Effect	✓	CG7654	0.1	-	No Effect
		CG4701	1	X	0	0.003	Semi-male sterility	X	CG5395	0.1	Individualization	Lethal
		CG8330	2	X	0	0.003	Male sterility	X	CG12157	0	Individualization	Lethal
	Inner Membrane	CG10090	2	✓	0	0.033	No Effect	X	CG40451	0	-	No Effect
		CG1158	2	X	0	0.005	Male sterility	X	CG40451	0	Early stage	No Effect
		CG12201	2	✓	0	0.014	No Effect	X	CG18347	0	-	Lethal
		CG18418	2	X	0	0.03	No Effect	X	CG1907	0	-	No Effect
		CG7514	1	✓	0	0.017	No Effect	X	CG1907	0	-	No Effect
	Mitochondrial Substrate/Solute Carrier	CG2616	2	X	0	0	No Effect	X	CG14209	0	-	No Effect
		CG9582	2	X	0	0.037	Semi-male sterility	X	CG5254	0.1	-	No Effect

* All detailed fertility and viability knockdown results for N-mt duplicated genes are provided in a Supplementary Results Excel file.

Supplementary Table 3. Mitochondrial localization probability, DSP status, and knockdown results of N-mt duplicates with the testis-biased expression for TCA cycle

Compartment		Gene	# of libraries	Present in the DSP	Off-Target	Mt Pro. Localization probability-Mitofate	Fertility Test (*Bam-Gal4*)*	Parent Present in the DSP	Parent	Mt Pro. Localization probability of Parent-Mitofate	Defect Stage in Spermatogenesis	Viability Test (*Actin5c-Gal4*) 27c*
TCA Cycle	TCA Cycle	*CG10749*	2	✓	0	0.774	Semi-male sterility	X	*CG7998*	0.3	Individualization	Lethal
		CG10748	1	X	0	0.53	-	X	*CG7998*	0.3	-	No Effect
		CG14740	1	✓	1	0.278	No Effect	X	*CG3861*	0.8	-	No Effect
		CG4706	2	✓	0	0.528	No Effect	✓	*CG9244*	0.7	-	No Effect
		CG6140	2	✓	0	0	Semi-male sterility	✓	*CG4094*	0.5	Individualization	Lethal

* All detailed fertility and viability knockdown results for N-mt duplicated genes are provided in a Supplementary Results Excel file.

Supplementary Table 4. Mitochondrial localization probability, DSP status, and knockdown results of N-mt duplicates with the testis-biased expression for redox activities

Compartment		Gene	# of libraries	Present in the DSP	Off-Target	Mt Pro. Localization probability-Mitofate	Fertility Test (*Bam-Gal4*)*	Parent Present in the DSP	Parent	Mt Pro. Localization probability of Parent-Mitofate	Defect Stage in Spermatogenesis	Viability Test (*Actin5c-Gal4*) 27c *
ROS & Damage Protection	Oxi Reduction	*CG33092*	2	X	1	0.81	No Effect	✓	*CG7145*	0.85	-	No Effect
		CG32026	1	✓	0	0.38	No Effect	X	*CG12233*	0.79	-	No Effect
	Thioredoxin Reductase Activity	*CG6888*	2	✓	0	0.009	No Effect	X	*CG5826*	0.009	-	No Effect
		CG11401	1	X	0	0.978	No Effect	X	*CG2151*	0.55	-	No Effect
	Response to DNA Damage	*CG8517*	2	X	0	0.345	No Effect	X	*CG8993*	0.98	-	No Effect
		CG9475	1	X	1	0	Male sterility	X	*CG16916*	0	Individualization	No Effect

* All detailed fertility and viability knockdown results for N-mt duplicated genes are provided in a Supplementary Results Excel file.

Supplementary Table 5. Mitochondrial localization probability, DSP status, and knockdown results of N-mt duplicates with the testis-biased expression for others

Compartment	Gene	# of libraries	Present in the DSP	Off-Target	Mt Pro. Localization probability-Mitofate	Fertility Test (*Bam-Gal4*)*	Parent Present in the DSP	Parent	Mt Pro. Localization probability of Parent-Mitofate	Defect Stage in Spermatogenesis	Viability Test (*Actin5c-Gal4*) 27c*
	CG9920	2	X	0	0.104	Male sterility	X	*CG11267*	*0.68*	Individualization	Semi Lethal
Protein Folding	*CG7235*	*1*	X	0	0.996	Male sterility	X	CG12101	0.98	Individualization	Semi Lethal
	CG2830	2	✓	1	0.996	Male sterility	X	CG12101	0.98	Individualization	No Effect
	CG7756	2	X	1	0.007	No Effect	X	CG4147	0	-	No Effect
Acetyltransferase Activity	*CG5265*	2	✓	0	0.196	No Effect	✓	CG1041	0.27	-	No Effect
succinate-CoA ligase	*CG6255*	1	✓	0	0.947	No Effect	X	CG1065	0.84	-	No Effect

Others (row group label)

* All detailed fertility and viability knockdown results for N-mt duplicated genes are provided in a Supplementary Results Excel file.

Supplementary Table 6. Knockout and overexpression of N-mt duplicates genes with the testis-biased expression with TRiP-CRISPR

Gene	Gene Name	TRiP-CRISPR	Compartment	Ch.	Fertility test *	Viability test*	Consistence to RNAi fertility test	Consistence to RNAi viability test
CG8330	*tomboy40*	TRiP-KO	Outer membrane	2R	No effect	No effect	X	X
CG11401	*Trxr-2*	TRiP-KO	Thioredoxin reductase activity	3L	No effect	No effect	✓	✓
CG18418	*CG18418*	TRiP-KO	Inner membrane	3L	No effect	No effect	✓	✓
CG14508	*Cyt-c1L*	TRiP-KO	Complex III	3R	Male sterility	No effect	✓	X
CG7514	*CG7514*	TRiP-KO	Inner membrane	3L	No effect	No effect	✓	✓
CG6255	*CG6255*	TRiP-KO	Succinate-CoA ligase	3R	No effect	No effect	✓	✓
CG7756	*Hsc70-2*	TRiP-KO	Protein folding	3R	Semi-male sterility	No effect	X	✓
CG15257	*Tim17b2*	TRiP-OE	Inner membrane	2L	-	No effect	-	-

* All detailed fertility and viability knockdown results for N-mt duplicated genes are provided in a Supplementary Results Excel file.

Supplementary Table 7. RNAseq expression of N-mt duplicated genes for different tissues

	FBGN	Viability Test	imaginal disc, larvae L3 wandering	fat body, white prepupae	fat body, pupae P8	testis, mated 4-day male	accessory gland, mated 4-day male
Complex I	FBgn0036706	Lethal	25	3	7	204	14
	FBgn0034007	-	9	2	0	36	4
	ND-51L1	Semi-lethal	56	10	3	80	17
	FBgn0039331	-	61	10	14	130	22
Complex II	FBgn0030975	Lethal	212	32	18	226	80
	FBgn0036222	lethal/semi	16	5	7	188	15
Complex III	UQCR-14L	Lethal	27	9	55	239	19
	FBgn0039651	Semi-lethal	37	5	9	85	11
Complex IV	FBgn0033020	-	93	25	40	461	51
	FBgn0086907	Lethal	224	47	70	723	158
Complex V	FBgn0035585	-	57	17	14	98	22
	FBgn0036568	-	51	8	1	97	14
Outer Membrane	tomboy20	-	58	18	27	273	34
	FBgn0028868	Lethal	28	5	5	96	13
	CG8330	Lethal	28	7	8	233	22
Inner Membrane	FBgn0038018	-	10	1	6	34	9
	FBgn0037310	-	56	5	9	148	10
	CG12201	-	24	11	18	130	32
	FBgn0035568	-	25	6	0	172	14
	FBgn0035567	-	3	1	0	25	4
Mitochondrial Substarte	FBgn0037512	-	14	2	3	41	5
	FBgn0032090	-	4	2	4	28	2
TCA Cycle	FBgn0036327	-	17	2	0	23	2
	FBgn0036328	lethal/semi	20	5	11	97	17
	FBgn0037988	-	32	6	11	130	22
	FBgn0037862	-	24	5	4	42	12
	FBgn0036162	Semi-lethal	40	8	19	286	68
Oxi Reduction	FBgn0053092	-	25	6	10	215	23
	FBgn0052026	-	59	15	15	122	51
Thrioredoxin Reducatse	FBgn0036490	-	36	9	12	326	21
	FBgn0037170	-	8	1	2	51	4
Response to DNA Damage	FBgn0034472	-	80	16	24	316	47
	FBgn0037742	-	35	5	7	45	7
Protein folding	FBgn0038486	-	17	4	4	78	17
	FBgn0038708	-	37	8	21	330	27
	FBgn0011244	-	74	16	29	356	52
	FBgn0001217	-	77	16	34	279	49
	FBgn0031728	Semi-lethal	239	53	80	940	171
	FBgn0038200	Semi-lethal	393	62	103	762	153

Chapter 3

COX4-like, a Nuclear-Encoded Mitochondrial Gene Duplicate,

Essential for Male Fertility in *Drosophila melanogaster*

Abstract

A study of nuclear-encoded mitochondrial genes (N-mt genes) in *Drosophila melanogaster* showed that there is a unique expression pattern for newly duplicated N-mt genes. Many of these new N-mt genes have testis-biased expression. Some of these genes have been shown to be essential in spermatogenesis. Here, we study the function of an additional duplicate N-mt gene with testis-biased expression, *Cytochrome c oxidase 4-like* (*COX4L*), in *D. melanogaster* to contribute to the understanding of their function and reasons for their retention in the genome. *COX4L* is a duplicate of *Cytochrome c oxidase 4* (*COX4*) of OXPHOS complex IV. While the *COX4* gene has been found in all the eukaryotes including single-cell eukaryotes like yeast, we show that the *COX4L* gene is only presented in *Brachycera* suborder of *Diptera*. So, both genes are presented in all *Drosophila* species but have a different pattern of expression. *COX4* is expressed highly in every tissue while *COX4L* is highly expressed in testis. To understand the function of this new gene, we knocked it down in *D. melanogaster* germline using two different RNAi lines driven by the *Bam-Gal4* driver. We also created a knockout strain for this gene using CRISPR-Cas9 technology. Knockdowns and the knockout of *COX4L* produce partial sterility and completely sterility in males, respectively. A lack of sperm individualization is observed in those males. The male infertility is rescued by driving *COX4L*-HA in the germline. In addition, ectopic expression of *COX4L* in soma causes embryonic lethality and overexpression in germline lead to a reduction in male fertility. The mitochondria of *COX4L*-KO are not functional and the ROS production in the testes was extremely high which might explain the male sterility of these flies. This prominent phenotype along with having energy-related functions, testis-biased expression, and also being present in the *Drosophila* sperm proteome database suggests that males might use different mitochondria in their germline and/or sperm. We should try to rescues the *COX4L* mutant phenotype with the parental gene. If selection has favored a different, higher energy-producing mitochondria function in male germline than in female germline and soma, the parental gene might not completely rescue the phenotype.

Keywords: nuclear-encoded mitochondrial gene, gene duplication, RNAi, *COX4L*, spermatogenesis, *Drosophila melanogaster*

Introduction

Mitochondria not only produce a big fraction of the cellular energy in eukaryotic cells but are also involved in a diverse set of cellular functions. They have been shown to be involved in metabolism (Wai and Langer 2016), immune regulation (Weinberg, et al. 2015), and cell death (Wallace 2005). Based on phylogenetic studies, alphaproteobacteria had been suggested as the mitochondria ancestor (Gray 2012). However, a recent study showed that mitochondria might have evolved from a proteobacterial lineage which branched off before the divergence of alphaproteobacteria (Martijn, et al. 2018). Mitochondria are one of the two organelles in the eukaryotic cells that have their DNA (mtDNA; chloroplast being the other). During ~1.5 billion years of evolution, the mitochondrial genome (mtDNA) has experienced many changes. Some of the ancestral genes (proteobacterial genes) have been lost because mitochondria did not need them in their new role (i.e., as a part of a eukaryotic cell). Some other genes that encode mitochondrial proteins have been transferred to the nucleus (i.e., nuclear-encoded mitochondrial genes or N-mt genes). Moreover, some of the nuclear genes have evolved to gain mitochondrial functions like novel accessory proteins in the OXPHOS system (Björkholm, et al. 2015). However, only a few genes remained in the mitochondrial genome (37 genes) which encode for 13 protein-coding genes, two rRNAs genes, and 22 tRNAs genes in most metazoan mitochondrial genomes (Gray 2012). These changes caused a severe reduction in mtDNA size as this organelle has a very small genome now (e.g., *Drosophila melanogaster* (*D. melanogaster*) mtDNA is only ~19.5 Kb; (Lewis, et al. 1995). As a consequence, mitochondria import most of their proteins from outside. The N-mt genes are transcribed in the nucleus and translated in the cytoplasm and ultimately the mitochondrial

67

proteins enter to the mitochondria by five transport pathways (Wiedemann and Pfanner 2017). Because the mitochondrial proteins are encoded by two different genomes, the interactions between N-mt proteins (encoded by nuclear DNA) and 13 proteins encoded by mitochondrial DNA are critical for all mitochondrial functions (Bar-Yaacov, et al. 2012; Friedman and Nunnari 2014). The only prominent protein-protein interactions between subunits encoded by these two genomes appeared in the mitochondrial OXPHOS complexes.

In the *D. melanogaster* genome, there are 786 N-mt protein-coding genes (http://www. biomart.org; Ensemble Gene96, BDGP 6.22; Smedley, et al. 2009). It has been shown that some (24%) of the N-mt genes belong to gene families and many of these duplicated genes (54%) acquired tissue-specific expression in this species (Gallach, et al. 2010). Intriguingly, all N-mt duplicated genes in *D. melanogaster* with tissue-specific expression are testis-specific genes (Gallach, et al. 2010). The unique expression pattern of these new genes is unexpected given that every cell in the body needs mitochondrial functions. New N-mt genes are duplicates of the genes that encode for different mitochondrial compartments such as OXPHOS complexes, TCA cycle, mitochondrial membranes, redox activity and protein folding (Gallach, et al. 2010; Eslamieh, et al. 2017). These genes are enriched for energy functions and assumed or observed to replace each other during spermatogenesis or in sperm (Wasbrough, et al. 2010).

The unique expression pattern and high enrichment for energy functions of these new genes lead to several nonexclusive hypotheses about the duplication of these genes. They might have evolved in response to male mtDNA harming mutations, to resolve intralocus sexual conflict at the parental gene, and to partition the pattern of expression or to have more

of the same protein (Gallach and Betran 2011b; Rogell, et al. 2014). Some of these new genes'

proteins are presented in the *Drosophila* Sperm Proteome (DSP; Wasbrough, et al. 2010) and

some of these new genes have been found to be essential during spermatogenesis (Timakov

and Zhang 2001; Sarkar and Lakhotia 2005; Lindsley, et al. 2013; Sawyer, et al. 2017). In

fact, 71% (17/24) of the N-mt duplicated with testis-biased expression are presented in DSP

while their parents are not. In addition, some of N-mt duplicated gene have been found to be

under positive selection (Pröschel, et al. 2006). It has been shown that the parental gene can

rescue the phenotype of the duplicated gene if the two genes have a similar function (Aoidi,

et al. 2016) but failed to rescue if both genes have different functions (Zhang, et al. 2007;

Venken, et al. 2010). These facts suggest that many of these genes have important/specialized

roles in spermatogenesis or/and sperm functions.

In *D. melanogaster*, there are more N-mt duplicated genes with testis-specific

expression for OXPHOS complexes subunits (which are essential for ATP production) than

for any other mitochondrial compartment (12/39; 31%). Cytochrome c Oxidase

(mitochondrial respiratory complex IV) is the last complex in the mitochondrial electron

transport chain and also one of the major regulation sites for oxidative phosphorylation

(Kadenbach, et al. 2000). The 13 subunits of this complex are encoded by two genomes. The

three biggest subunits (COXI, COXII, and COXIII) are homologous to their corresponding

subunits in prokaryotes (Capaldi 1990) and are encoded by the mtDNA. The rest of 10

subunits and other cytochrome c oxidase-specific regulatory proteins are encoded by the

nuclear genome (N-mt genes which only exist in eukaryotes; Tsukihara, et al. 1996;

Barrientos, et al. 2002; Saraste 2009). These N-mt subunits have been proposed to modify the

catalytic activity and stability of the holoprotein at complex IV (Kadenbach, et al. 2000; Fornuskova, et al. 2010). *Cytochrome c Oxidase 4* (*COX4*, *CG10664*) is one of the complex IV N-mt genes which has been duplicated through DNA duplication and the new copy, *Cytochrome c Oxidase 4-Like* (*COX4L*, *CG10396*), is believed to still encode a subunit in that complex. *COX4L* was known to be an old duplicate (at least 63 Myr old; time of *Drosophila* genus diversification) and present in all *Drosophila* species (Zhang, et al. 2010). *COX4* is a very conserved gene which was found in all eukaryotes and even eukaryotic single-celled including yeast but it is not seen in bacteria as they lack this subunit. In *D. melanogaster*, *COX4* is expressed highly in every tissue and is considered to be a non-biased gene (Sex Biased Ratio (SBR)= 0.58; Campos, et al. 2018). *COX4L*, however, is highly expressed in testis and is considered a male-biased gene (SBR= 9). Since only *COX4L* is in the sperm proteome (Wasbrough, et al. 2010), this gene might replace the parental gene in sperm mitochondria. Transcriptional studies of testes with GeneChip and RNA-Seq analyses have shown that maximum expression of *COX4L* occurs at the proximal region of testis where the expression is significantly higher in meiosis than post-meiosis and mitosis during spermatogenesis (Vibranovski, et al. 2009; Vedelek, et al. 2018). The parental gene (*COX4*) is highly conserved among eukaryotes and has been studied widely before. For example, a reduction in COX activity, impaired ATP production, and elevated ROS production was reported in the patient with mutations at *COX4I1* gene in human (Abu-Libdeh, et al. 2017). In addition, the knockdown of *COX4* in *D. melanogaster* showed a reduction in the rate of mitochondrial respiration, walking speed with *arm-Gal4* (drives ubiquitous expression in

embryos and larvae) and complete lethality with either of *da-Gal4* or *Tub-Gal4* (drives ubiquitous expression at all developmental stages) drivers (Klichko, et al. 2014).

Here, we study the phylogenetic distribution and function of *COX4L* in *D. melanogaster* to provide some insights into the evolutionary pressures that have led to the retention of this gene after duplication. Both knocking down *COX4L* expression in the germline and knocking out the gene from the genome suggest that this gene is essential for male fertility. This prominent phenotype along with having energy-related functions, testis-biased expression, and presence in *Drosophila* sperm proteome database where the *COX4* is absent suggests that males might use different mitochondria in their germline and selection might favor different, higher energy-producing mitochondria in the male germline than in the female germline and the soma. We show that the *COX4L* gene is older than previously known and present in the Brachycera suborder of Diptera.

Materials and Methods

COX4L RNAi knockdown, viability and fertility tests

Flies were raised on standard cornmeal/malt medium at room temperature (25°C). All crosses were performed at room temperature except the crosses that were set up to express a UAS transgene under a Gal4 driver for RNAi which were run at two other temperatures: 27°C and 29°C. This was done because the Gal4-UAS system works more efficiently at higher temperature (Duffy 2002b). Transgenic flies with UAS-RNAi constructs (i.e., RNAi lines) of KK and GD libraries were obtained from the Vienna *Drosophila* Resource Center (VDRC; Dietzl, et al. 2007). The GD library insertions are P-element based transgenes with random

insertion sites, whereas the KK library contains phiC31-based transgenes with a single, defined insertion site (Dietzl, et al. 2007). Information for all the lines used is provided in Supplementary table 1. The *Actin5c-Gal4* driver (a ubiquitous driver) was crossed to the RNAi lines to study the knockdown in every tissue (viability test) and was obtained from the Bloomington *Drosophila* Stock Center (stock # 4414). The *Bam-Gal4* driver (a germline driver) was crossed with RNAi lines to study the knockdown in male and female germlines (fertility test; Chen and McKearin 2003) and was received as a gift from Michael Buszczak laboratory at UT Southwestern Medical Center. The original strains that were used to make the KK and GD libraries were obtained from VDRC and used as knockdown controls. These are the isogenic strain w^{1118} (VDRC ID 60000) for the GD line and the *y,w[1118];P{attP,y[+],w[3`]}* strain (VDRC ID 60100) for the KK line. Reciprocal crosses with at least three replicates were performed for all experiments. For the viability test, virgin males and females were collected and kept for three days to make sure they were mature and then two males were crossed with three females. On day five, flies were dumped out from the vials and then the number of offspring were counted on day 15. Reciprocal crosses with at least three replicates were performed for each cross. All viability crosses were made at three different temperatures: 25°C, 27°C, and 29 °C. For the fertility test, one virgin male and two virgin females were kept in the vial for five days and then the number of progeny was counted on day 15. All the crosses were made at two different temperatures: 25°C and 27°C. The higher temperatures were not used as they can be stressful for the flies, specifically during spermatogenesis (Ben-David, et al. 2015). For fertility test, the same VDRC stocks (VDRC ID: 6000 and 6100) were used as control and reciprocal crosses with at least three replicates

72

were performed. For both tests, data were analyzed with R Stats package (http://www.r-project.org; R DevelopmentCoreTeam 2013) by t-test.

Generating the *COX4L* knockout flies, viability and fertility tests

We used the CRISPR-Cas9 technology to generate *COX4L* knockout (*COX4L*-KO) flies in collaboration with Michael Buszczak laboratory. Two guide RNAs (gRNAs) were designed using the online platform http://tools.flycrispr.molbio.wisc.edu/targetFinder (Gratz, et al. 2014) and synthesized by IDT, Inc. Then, each gRNA was annealed, phosphorylated and ligated into the BbsI sites of pU6-BbsI-chiRNA plasmid (Addgene # 45946) separately producing to plasmids to express the guide RNAs in germline upon embryo injection (Gratz, Cummings, Nguyen, Hamm, Donohue, Harrison, Wildonger and O'Connor-Giles 2013). In addition, two homologous arms were designed with the same tool to be assembled in the donor vector cloned into pHD-DsRed-attP flanking the eye driven DsRed cassette and designed to replace *COX4L* (Addgene # 80898; Gratz, et al. 2014). NEBuilder HiFi DNA Assembly (NEB, Inc.) was used to assemble the homologous arms flanking the DsRed cassette as shown in Figure 1A. The two gRNAs plasmids and the donor vector were co-injected into preblastoderm embryos of nos-Cas9 attp2 by Rainbow Transgenic Flies, Inc. (Camarillo, CA). The final concentration of injected plasmids for pHD-DsRed-attP donor vector and each of the pU6-BbsI-chiRNA containing the guide RNAs was 250 ng/μl and 20ng/μl, respectively. The gRNAs and donor vector design is provided in the Supplementary table 2. Flies were collected from injected embryos and crossed with *w^{1118}* flies. The progeny of this cross was crossed with a second chromosome balancer stock (CyO/sna[Sco]; BDSC ID: 2555) and the

73

offspring was screened for fluorescent glowing eyes which confirms the replacement of the gene by the eye-driven DsRed gene. The absence of a PCR band in the homozygote knockout individuals performed with two primers inside *COX4L* (Supplementary table 2) confirmed that *COX4L* has been removed from the genome completely (Figure 1B).

The viability test was performed for the *COX4L*-KO in which the homozygote virgin males and females were mated for 5 days at 25°C and were dumped out from the vial at day 5. The number of offspring were counted on day 15. For the fertility test on the *COX4L*-KO, one homozygote virgin male and two homozygote virgin females were kept in the vial for five days at 25°C, and then the progeny was counted on day 15. Data were analyzed with R Stats package (R DevelopmentCoreTeam 2013; http://www.r-project.org) by t-test.

Rescue of the *COX4L*-KO with a *COX4L* transgene

The FlyORF stock of *COX4L* was obtained from Zurich ORFeom Project Center (Fly Line ID: F002652) and was used to rescue the *COX4L*-KO phenotype. FlyORF stocks have been created using the site-specific ΦC31 integrase and insertion of the transgenes into an identical integration site on the right arm of the third chromosome (attP-86Fb) to insert ORFs under UAS in the genome (Bischof, et al. 2013). These UAS-ORFs are under the UAS regulatory region and can be expressed in vivo using the Gal4-UAS system. *COX4L* UAS-ORF line is a valuable stock that can be used either for ectopic expression, overexpression or expression of a gene in the KO mutant to rescue the loss of function effects. We drove *COX4L* UAS-ORF with both *Bam-* and *nos-Gal4* drivers to rescue *COX4L*-KO phenotype.

74

Ectopic expression and overexpression of *COX4L*

The *COX4L* UAS-ORF line was also used to study the consequences of ectopic expression of *COX4L* in soma and overexpression of *COX4L* in the germline. Two different ubiquitous Gal4 drivers (*Actin5c-Gal4* and *Tub-Gal4*) were used for ectopic expression of this gene in the soma and two germline Gal4 drivers (*Bam-Gal4* and *nos-Gal4*) were used for overexpression of this gene in the germline. Three replicates and reciprocal crosses were performed at two different temperatures (25°C and 27°C) for all crosses. To breakdown the lethality effect of *COX4L* overexpression, *COX4L* UAS-ORF flies were crossed with *Actin5c-Gal4* flies and were placed in chambers with a plate containing agar mixed with molasses and a yeast paste in the middle of the plate. The embryos were collected every 3 hours and the first instar larvae were collected every 24 hours. All embryos and larvae were transferred to a vial containing media (three vials per sample). Vials were kept at 25°C and the adults were counted after 15 days.

Drosophila testes staining

Interestingly the *COX4L* UAS-ORF (Fly Line ID: F002652) introduced above has a 3xHA tag at the C-terminus and was used to study COX4L localization. This stock was crossed to both *Bam-Gal4* and *nos-Gal4* drivers at 27°C and the virgin males were collected. The testes of one-day-old virgin males were dissected in 1X PBS within 20 minutes and fixed and stained as previously described (White-Cooper 2004). In short, the testes were fixed in

75

4% PFA in PBS for 10 minutes and then thoroughly washed with PBT (PBS + 0.1% Triton X) for 30 minutes. Primary antibody (Anti-HA; Cat # C29F4, Cell Signaling Technology, Inc.) was added to the fixed tissues in a concentration of 1:100 in PBT and stored at room temperature for 2 hours. After testes were thoroughly washed, the secondary antibody (Anti-Rabbit; Cat # A11008, Invitrogen, Thermo Fisher Scientific) was added at a concentration of 1:200 in PBT and stored at room temperature for 2 hours. The testes were washed in 1X PBS three times and then mounted on slides for imaging. DAPI (Slowfade Gold antifade reagent with DAPI; Ref # S36939, Invitrogen, Thermo Fisher Scientific) was used to stain the nucleus. For mitochondria staining, MitoTracker™ Deep Red FM (Cat # M22426, Invitrogen, Thermo Fisher Scientific) is a mitochondrial potential-dependent dye and was used to stain mitochondria (Gilmore and Wilson 1999). MitoTracker® Red CM-H2XRos (Cat# M7513, Cell signalling Technology) in the concentration of 500 nM in DMSO was used to stain the level of ROS production. This Mitotracker is a ROS-sensitive dye which can monitor the oxidation status of the cell (Chen and Gee 2000). In addition, TMRE (tetramethylrhodamine, ethyl ester; Cat # T669, Invitrogen, Thermo Fisher Scientific) was used to stain active mitochondria in the concentration of 100 nM. The mitochondrial membrane potential can be measured by TMRE stating and this dye also can be used to find functional/active mitochondria. The TMRE is positive dye which can stained healthy/functional mitochondria with negative potential (Crowley, et al. 2016). The fluorescent microscope at UT Arlington (Olympus BX51) and the confocal microscope at UNT Health Science Center (Zeiss, LSM510META) were used for imaging. The quantification of images for comparison was performed in ImageJ (Schneider, et al. 2012) and ZEISS software.

Computational analysis

Previous analyses of N-mt duplicated genes revealed that *COX4* and *COX4L* are in the same gene family (Gallach, et al. 2010; Eslamieh, et al. 2017) as they have more than 50% identity at the protein level (51.4%). To explore if both genes might have the same function, we performed analyses to characterize the presence of mitochondrial localization signal (MLS) for both genes. We used an online webserver, MitoFates (Fukasawa, et al. 2015) which analyzes the 100 amino acids from the N-terminus of any given peptide and reports the probability of mitochondrial pre-sequence, a cleavable localization signal with its position. To compare protein structure of both genes, domains of both proteins were predicted by CDD (Marchler-Bauer, et al. 2017) and Phyre was employed to characterize the tertiary structure of predicted domains (Kelley, et al. 2015). The BioGRID (Oughtred, et al. 2019) was used predict protein-protein interactions for both *COX4* and *COX4L* in *D. melanogaster*. The dN/dS ratio was calculated using Codeml program in PAMLX (Xu and Yang 2013). In addition, we attempted to evaluate the predicted protein-protein partners by calculating evolutionary rate covariation (ERC) using the ERC Analysis Web server from Pittsburg University (https://csb.pitt.edu/erc_analysis; Clark, et al. 2012, 2013; Findlay, et al. 2014). The ERC analyses were performed by using the top genes search. In this analysis, *COX4L* gene was compared to the rest of N-mt duplicated genes and the highest statistically significant ERC values were retrieved.

Phylogenetic analysis

The COX4 and COX4L sequences were retrieved for Diptera from Ensembl database (https://www.ensembl.org) and aligned with MUSCLE (Edgar 2004). Maximum likelihood gene tree (Chor and Tuller 2005) was constructed using PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference.

Results

COX4L is a well-conserved sperm protein differentiated from _COX4_

COX4L (on chr2R) originated at least 63 million years ago via DNA duplication from _COX4_ (on chr2L) as it is present in all 12 sequenced _Drosophila_ species (Clark, et al. 2007). Protein alignment of these two genes shows that despite their long evolutionary history, only 102 amino acids of _COX4L_ were different after the origination. _COX4L_ is a DNA mediated duplication (Figure 1A). We used the BLASTP (Altschul, et al. 1997) and the available Diptera genomes in Ensembl to date the _COX4L_ duplication. We find it in all _Drosophila_ species, _Musca domestica_, _Glossina morsitans_ and _Lucilia cuprina_ but not in mosquito or farther related genomes. This will make _COX4L_ between 208 and 245 Myr old (Figure 2A; Wiegmann, et al. 2003). See alignment of the _COX4L_ and _COX4_ proteins from these species and _COX4_ proteins in additional species (Supplementary file 1). A tree using those sequences and the maximum likelihood gene tree reconstruction method (See Materials and Methods and Figure 2A) shows that the two genes cluster separately as their proteins are quite different because _COX4L_ has been evolving fast (See below). Only _COX4L_ protein was found in the DSP (Wasbrough, et al. 2010) which suggests that this gene is important for sperm function

and might have a different function than its parental gene (*COX4*). The two genes also have different evolutionary rate (dN/dS$_{COX4}$ = 0.05; dN/dS$_{COX4L}$ = 0.2508 where *COX4L* is evolving significantly faster than *COX4* (Maximum likelihood ratio test; P-value = 0.001).

The conserved domain analysis of proteins showed that both proteins have seven polypeptide binding sites (subunit IV/I, subunit IV/II, subunit IV/IIb, subunit IV/IIIb, subunit IV/Va, subunit IV/Vb, and subunit IV/VIc) with one chemical binding site (putative ATP/ADP binding site). The tertiary structures of the two proteins don't look very different (Figure 1C) which suggests that the two genes might have similar functions in different tissues.

Mitochondrial localization signal analysis (See Materials and Methods) predicted that both genes have a high probability over 0.9 of being imported into the mitochondria. *COX4* has one cleavage site (mitochondrial processing peptidase (MPP) cleavage site) which is important for cleaving off the presequences once the protein is inside the matrix (Hawlitschek, et al. 1988). However, *COX4L* appears to have an extra cleavage site, i.e., Intermediate cleaving peptidase of 55 kDa (Icp55) cleavage site. This additional cleavage site has been suggested to be important for protein stability within mitochondria by cleaving one amino acid from the MPP-generated intermediate N-terminus (Vogtle, et al. 2009). The physical protein interaction analysis with BioGRID (Oughtred, et al. 2019), an online biological interaction repository, shows very different interaction patterns (Supplementary table 4). COX4 interacts with many other OXPHOS complexes subunits particularly with complex V proteins while COX4L does not seem to have interaction with any OXPHOS complexes

subunits in *D. melanogaster*. Physical interactions of COX4L with CG30089 and CG11163 (Zinc transporter) have been reported previously (Giot, et al. 2003).

The evolutionary rate covariance (ERC) can be measured across a phylogeny and can help us find genes that directly interact and coevolve (i.e., have similar evolutionary histories). Typically, a high ERC value between two genes suggests that they are working in a common pathway or related function (Clark, et al. 2012, 2013). Therefore, previously unknown functional connection between genes can be found with the ERC value (Findlay, et al. 2014). ERC analysis of COX4L with other N-mt genes in families (parental and duplicated genes) shows that COX4L has higher ERC values with other N-mt duplicated genes than its parental gene (0.6582 COX4L average *vs.* 0.348 COX4 average). The same analysis for COX4 shows that COX4 has higher ERC value with other N-mt parental genes than COX4L (0.664 COX4 average *vs.* 0.348 COX4L average). These data suggest that COX4 and COX4L have different evolutionary rate covariance and new N-mt genes, in general, might work together/replace parental during spermatogenesis and in sperm. Altogether, these data suggest that *COX4L* is a well-conserved duplicated sperm protein among *Brachycera* suborder which is also likely functionally differentiated from its parent (*COX4*).

COX4L protein localizes inside the mitochondria

COX4L protein was detected in the *Drosophila* sperm proteome but the *COX4L* UAS-ORF stock can be used to detect its precise localization. Taking advantage of the HA tag present in *COX4L* UAS-ORF stock, we were able to show the protein is localizing in the

mitochondria (Figure 3A). The protein was abundant all over the sperm tails, where the mitochondria are the dominant organelle.

Knockdown of *COX4L*

Two different RNAi libraries (KK and GD) were used to knock down the expression of *COX4L* in the soma and germline. To study the effect of this gene on viability, the UAS lines were crossed with *Actin5c-Gal4* line and the number of progeny was counted (See more details in Materials and Methods). The results between UAS libraries and different temperature (25°C, 27°C and 29°C) were consistent with each other, where no significant differences were observed between UAS-Gal4 crosses and controls (P > 0.05 in all comparisons; Supplementary table 3) at any of the temperatures. These results suggest that *COX4L* is not needed for viability.

To study the effect of this gene on fertility, the RNAi lines (GD and KK library) were driven with *Bam-Gal4* (drives expression in male and female germline cells). This Gal4 driver is expressing in the germarium, cyst cells, spermatogonia, cystoblasts, and cystocytes (Chen and McKearin 2003). The UAS lines were crossed with *Bam-Gal4* line, and the virgins were collected and crossed with w^{1118} virgin flies. The progeny of the last crosses were counted for the experimental and control crosses (See more details in Materials and Methods). Knockdown of *COX4L* in the germline causes semi-sterility in males while it causes a significant increase in female fertility. The fertility crosses weren't performed at 29°C because this temperature has been shown to have a detrimental effect on male fertility (Ben-David, et al. 2015). The results were consistent between the two knockdown libraries and also across

both temperatures (25°C and 27°C) and the reciprocal crosses (32% reduction in male's fertility and 26% increase in female's fertility; Supplementary table 5). To further investigate male infertility in knockdown flies, the testes of knockdown males were dissected and studied. All early steps of spermatogenesis seemed normal, but the seminal vesicles of knockdown males were empty (Figure 4A). These observations suggest that a defect in the sperm individualization step causes male sterility in *COX4L* knockdowns. The results of knockdowns in soma and germline mostly followed our hypothesis based on the unique expression pattern of this gene where only male infertility was expected.

Knockout of *COX4L* results in male sterility

To confirm the knockdown results, we decided to create a null mutant for *COX4L* and study the function of this gene in more detail. Taking advantages of CRISPR-Cas9 method the whole CDS of *COX4L* was removed from the genome (*COX4L*-KO mutant). The gene was replaced by an eye-driven DsRed gene (See Materials and Methods for more details). The removal of the gene was confirmed by PCR and sequencing (Figure 1B). Viability and fertility assays were performed on six independent knockout lines (Data not shown). Results are shown only for one line and no significant change in viability was observed between the *COX4L*-KO and the controls (P> 0.05). However, when we performed the fertility assay on *COX4L*-KO flies, the male flies were completely sterile (Figure 5A). Compare to the w^{1118}, female fertility of *COX4L*-KO increased when we performed the fertility assay (P< 0.05; Figure 5B).

Male sterility in *COX4L*-KO mutant is due to late spermatogenesis defects

In order to understand the function of *COX4L* during spermatogenesis, we tested the morphology of *COX4L* mutant testes. We found, consistent with the KD phenotype introduced above, that the seminal vesicles in *COX4L*-KO flies were empty and again inferred that there is a problem in the individualization step in spermatogenesis (Figure 4B and D).

Rescue the *COX4L*-KO phenotype

To further confirm that *COX4L* is responsible for the infertility phenotype, we tried to rescue the phenotype using the *COX4L* FlyORF line. We were able to rescue the *COX4L*-KO phenotype driving the *COX4L* FlyORF line under *Bam-Gal4* and *nos-Gal4* drivers (Figure 5 A and B). See also supplementary information for the cross design used for this rescue (Supplementary figure 1). The male fertility was completely rescued in these crosses and no viability/fertility effects were observed when the rescue was performed on females (Figure 5 A and B).

Overexpression of *COX4L*

The *COX4L* FlyORF line was also used to study the overexpression of this gene in soma and germline. Overexpression of *COX4L* in soma caused complete lethality when it was driven with two different ubiquitous drivers: *Actin5c-Gal4* and *Tub-Gal4*. The overexpression of *COX4L* with germline Gal4 drivers (*Bam-Gal4* and *nos-Gal4*) didn't show any viability effect. However, overexpression of *COX4L* in germline showed fertility reduction in males compared to the control group (Figure 5). Interestingly, female fertility didn't change when

we overexpressed *COX4L* in the germline. These results suggest that a fine-tuning expression of *COX4L* is necessary for male fertility. Further investigation of *COX4L* overexpression with ubiquitous drivers revealed that the viability effect of *COX4L* overexpression was not happening in the pupa stage as all the pupas were empty (no pupa arrest). To find the lethality stage, we collected the embryos and larvae from overexpression crosses with ubiquitous drivers. The viability effect happened before larvae stage as no significant difference was observed between overexpression crosses and controls for larvae viability (Supplementary table 5). The overexpression of *COX4L* in the soma in collected embryo suggested that the lethality effect is happening at early embryonic stage (Supplementary figure 1).

An increase of ROS production in *COX4L-KO* flies

Staining mitochondria of *COX4L*-KO testes with MitoTracker® Red CM-H2XRos shows that the ROS production increased dramatically in that tissue as the high intensity of the red color in sperm bundles is observed in the *COX4L*-KO compared to the controls (Figure 6 A-B). This increase in ROS production in the knockouts could explain the male sterility of these flies as the oxidation status of the cell has changed in the absence of *COX4L*. In addition, staining mitochondria of *COX4L-KO* testes with TMRE shows the mitochondrial membrane potential of *COX4L*-KO testes is decreased in these testes compare to the controls (Figure 7 A and B). This suggests that the mitochondria in the knockout are not functional which could be results of the electron leakage and higher ROS production in *COX4L-KO*.

Discussion

Here, we studied a duplicate of *COX4* with testis-specific expression in *D. melanogaster*. We revealed that the duplication is present in other flies in addition to *Drosophila* including *Musca domestica*, *Glossina morsitans*, and *Lucilia cuprina*. This observation makes this gene much older than previously known but its origin does not appear to coincide with the advent of giant mitochondria along the sperm tail in these flies as it is likely a trait of many insects and arthropods (Noguchi et al. 2012).

COX4L is evolving at different rate than *COX4* and has different inferred interactions but it likely replaces *COX4* at least in sperm as only *COX4L* has been found in the *Drosophila* Sperm Proteome (Wasbrough, et al. 2010) and both show a high probability of targeting the mitochondria. In addition, *COX4L* shows an extra cleavage site known to evolve to increase protein stability in mitochondria (Vogtle, et al. 2009).

Knockdowns in germline and knockout of *COX4L* cause partial and complete infertility in males, respectively. The partial infertility of the knockdown could be explained by the efficiency of the RNAi or the UAS-Gal4 system which lead to the presence of enough mRNA to show some fertility. The complete rescue of the *COX4L*-KO with *COX4L* FlyORF line confirms that the male infertility is due to the absence of this gene. Although we do not have direct evidence, this phenotype is consistent with the absence of COX4L in sperm but the presence of COX4 in other cell types. Absent of COX4 function has been shown to reduce ATP production in other cell types (Abu-Libdeh, et al. 2017). The spermatogenesis, an energy-demanding process, might proceed without COX4L until the individualization step but failed after that step because of the ATP reduction. It seems that COX4L is important for

complex IV functionality and the absence of this protein causes leakage of electrons from this complex. These free-electrons could increase the level of ROS in the mitochondria. This enhancement in the level of ROS production can decrease the mitochondrial membrane potential which depolarized mitochondrial membrane and makes it less negative and nonfunctional. Further analyses should reveal if elongation is completed in *COX4L*-KO as it seems an energy-demanding step as well and at what point COX4L replaces COX4 in the mitochondria during spermatogenesis.

So why would we see the evolution of the replacement of COX4 by COX4L in sperm? The fact that this duplicate has energy-related functions suggests that males might use different mitochondria in their germline and selection might have favored different, higher energy-producing mitochondria in male germline than in female germline and soma if there is a cost, e.g. ROS production (Gallach, et al. 2010). Selection of specialized mitochondria has been reported before where a distinct germline division of mitochondria function and structure was seen between males and females (de Paula, et al. 2013). Mitochondria is metabolically different in males and females. Mitochondria of female gametes (oocytes) are small with the suppression of DNA transcription, electron transport, and free radical production. Conversely, mitochondria of male gametes (sperm) is metabolically active in which transcribe mitochondrial genes for respiratory electron and also produce free radicals (de Paula, et al. 2013; de Paula Wilson, et al. 2013) that might produce mtDNA damage but not be selected against because they are not passing mitochondria to their offspring. All together, these results support the hypothesis that *COX4-L* plays a role in higher/specialized energy production for sperm function.

We drove *COX4L* FlyORF in the soma to study ROS production and metabolism in somatic cells expressing this gene and test the hypothesis proposed above for the retention of *COX4L*. However, we see lethality when we do that. This could be due to a perturbation in the dose affecting assembly of this OXPHOS complex. This could be tested by driving overexpressing the parental in the some but no *COX4* FlyORF line exists. We are, however, in the process of getting that line. Once it is available, we should also be able to test if the parental gene can rescue the *COX4L*-KO or not and test if *COX4* and *COX4L* functions are different as currently proposed. At this point, it seems that they might have evolved in response to selection for specialization and potentially to resolve intralocus sexual conflict (Gallach and Betran 2011a) but not to have more of the same protein or partition the pattern of the expression.

Acknowledgements
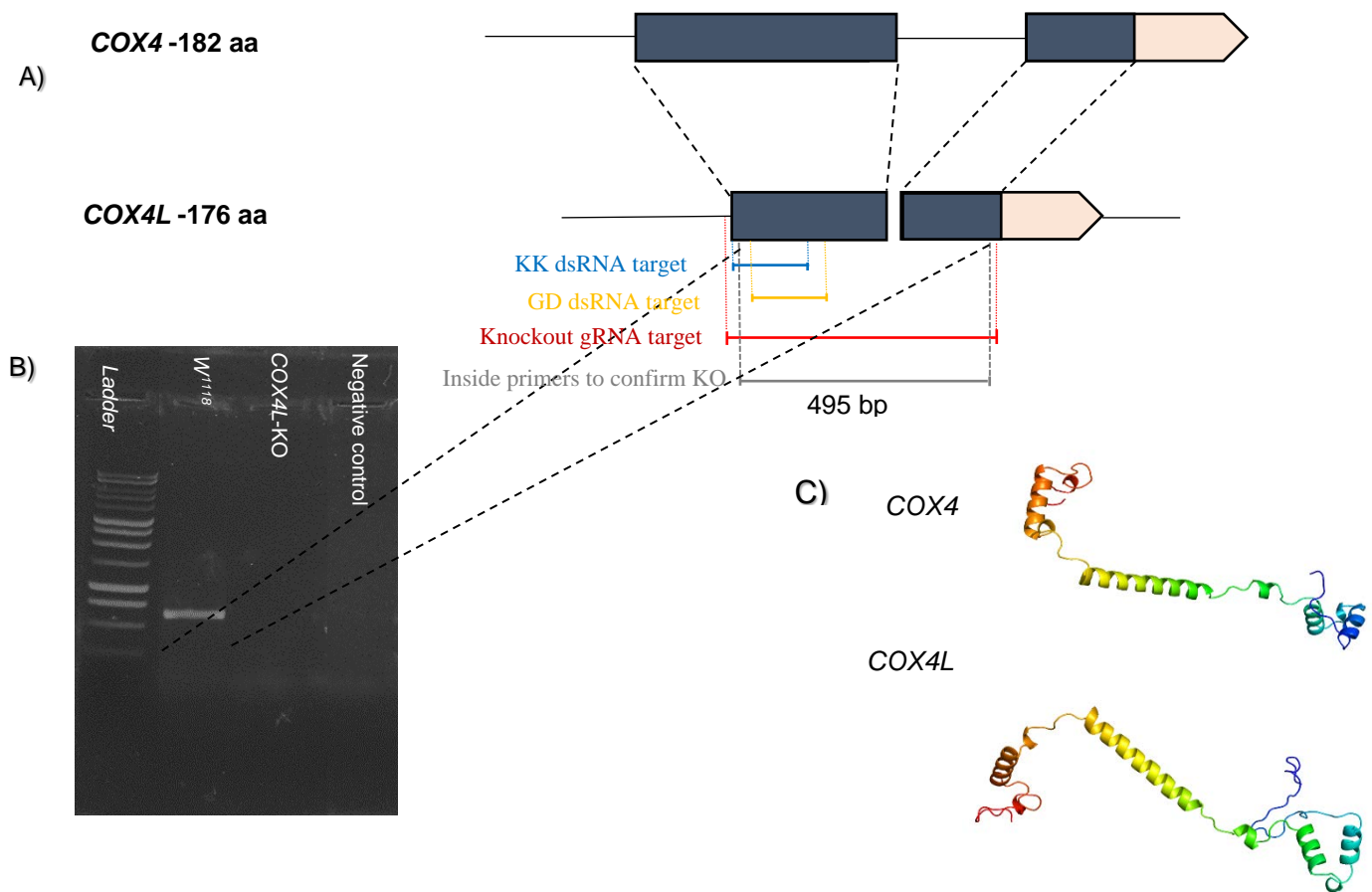
Figures

Figure 1. (A) *COX4* and *COX4L* gene structures. *COX4L* is a DNA-mediated duplication of *COX4* which is only 6 aa shorter than its parental protein. The regions of dsRNA expressed in the RNAi knockdowns using the KK and GD lines are shown. The region deleted and replaced by eye-driven DsRed after the sgRNAs are recognized by Cas9 and repaired is shown. The location of the primers for PCR within *COX4L* are depicted. (B) The lack of a PCR for *COX4L*-KO homozygotes is shown here and confirms the removal of *COX4L* form genome. (C) Tertiary structures of *COX4* and *COX4L* generated with Phyre.

A)
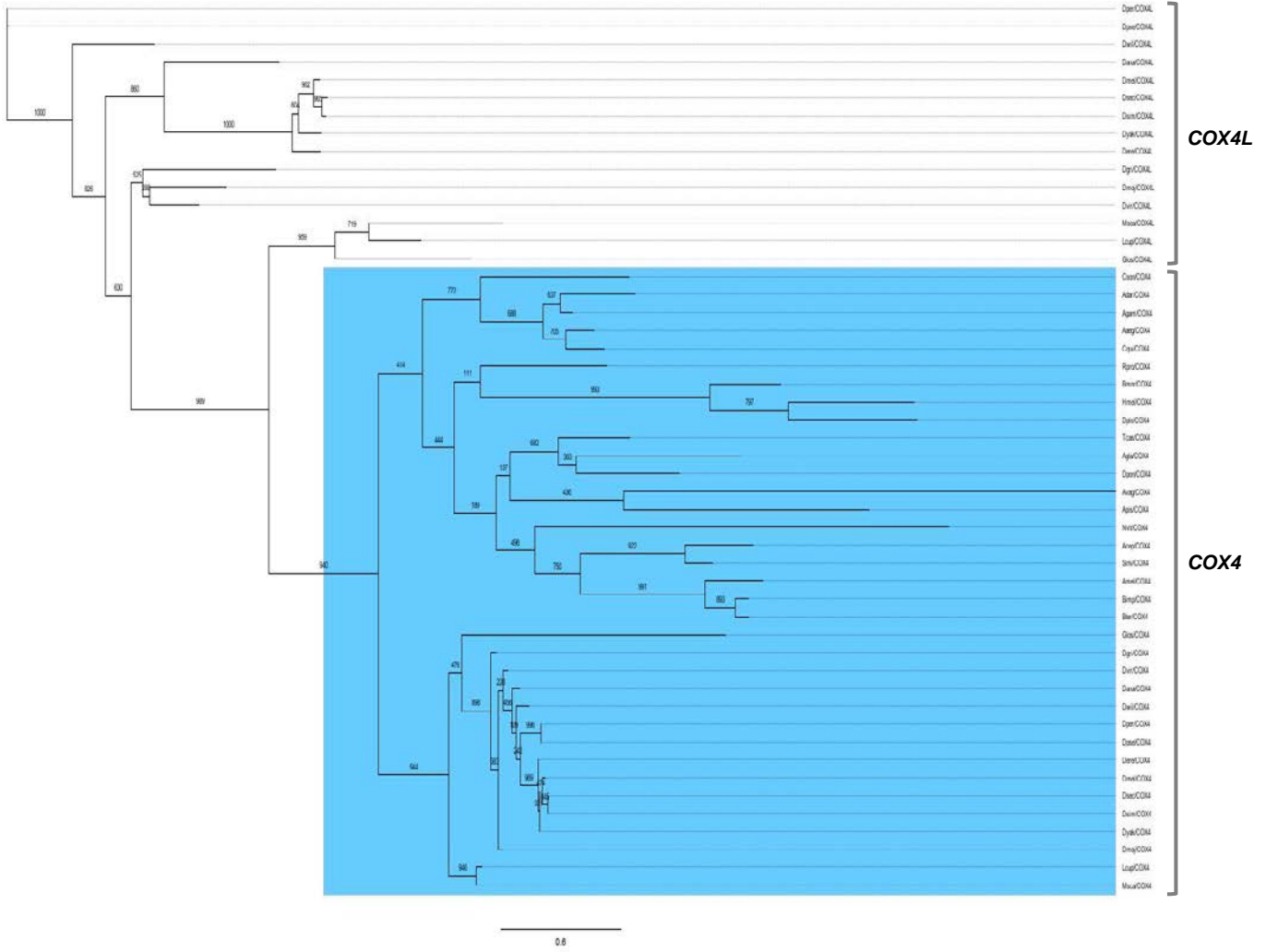
B)
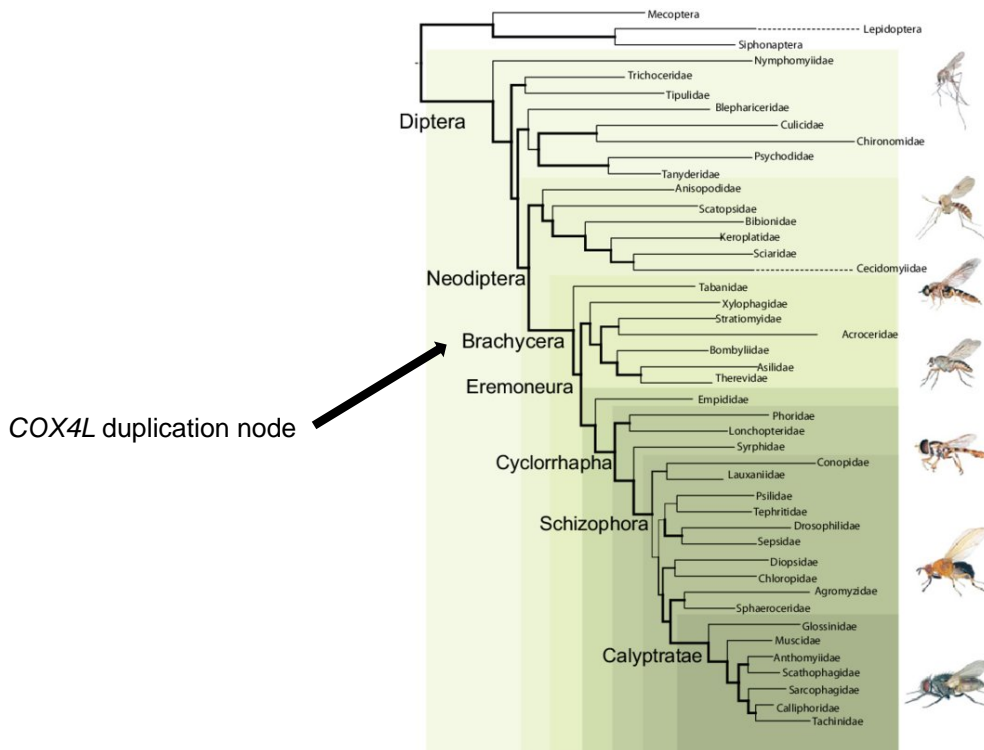
Figure 2. (A) The maximum likelihood gene tree of COX4 and COX4L using amino acid sequences of the genes in Diptera is shown. The Le-Gascuel 2008 model (Le and Gascuel 2008) was used. Bootstrap values are shown at the nodes refer to 1000 trials on PhyML performed using the webserver at www.phylogeny.fr. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analyses were conducted in PhyML online webserver (PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference). COX4L sequences cluster separately from COX4 although they do not branch from the branch where the duplication occurred due to COX4L fast evolution. (B) The duplication node of COX4L was shown in Dipteran tree. Based on the distribution of COX4L in the phylogenetic tree, COX4L probably duplicated in the lineage leading to *Brachycera*, a suborder of *Diptera*.
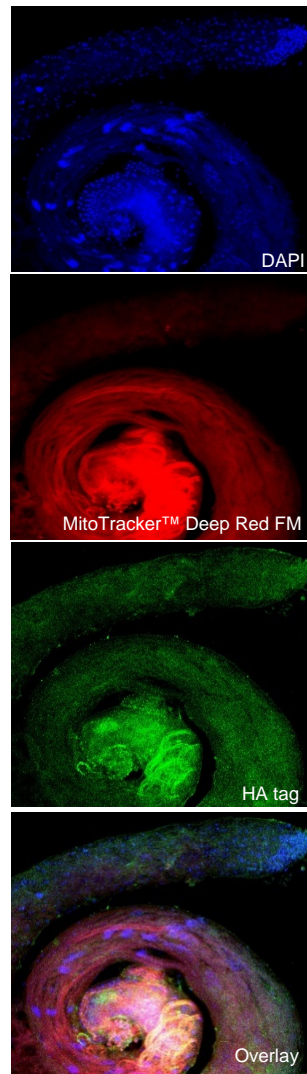
Figure 3. Confocal images are shown. Localization of COX4L in testis after driving *COX4L* FlyORF with HA tag with *Bam-Gal4* drivers is shown. Nucleus, mitochondria, and COX4-L protein are shown stained with DAPI, Mitotracker red, and green fluorescent secondary antibody, respectively.
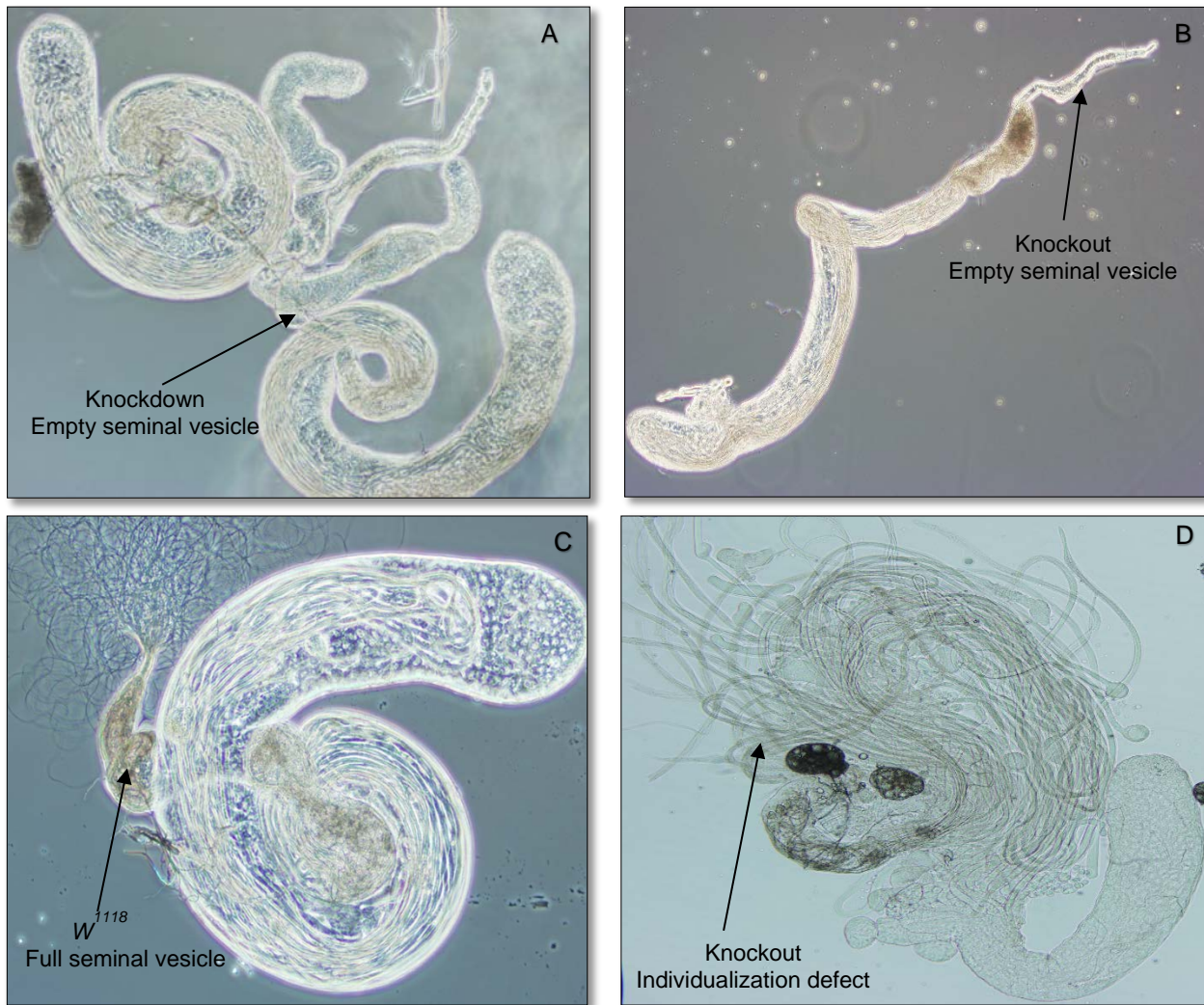
Figure 4. Dissected testes of *COX4L-KD* with *Bam-Gal4* and *COX4L*-KO. (A) The dissected testes of *COX4L* UAS from KK library with *Bam-Gal4* are shown. Arrow points at the empty seminal vesicle. (B) The dissected testes of *COX4L*-KO with an empty seminal vesicle are shown. (C) w^{1118} testis is shown here. All the spermatogenesis stage are normal and the seminal vesicle is full. Some mature sperms are moving around the raptured seminal vesicle. (D) Dissected testes of *COX4L*-KO males are shown here. The defect was observed in the individualization step of spermatogenesis where no mature sperm could be found. The sperms couldn't enter to the seminal vesicle and accumulate before seminal vesicle because of the defect in the individualization step.
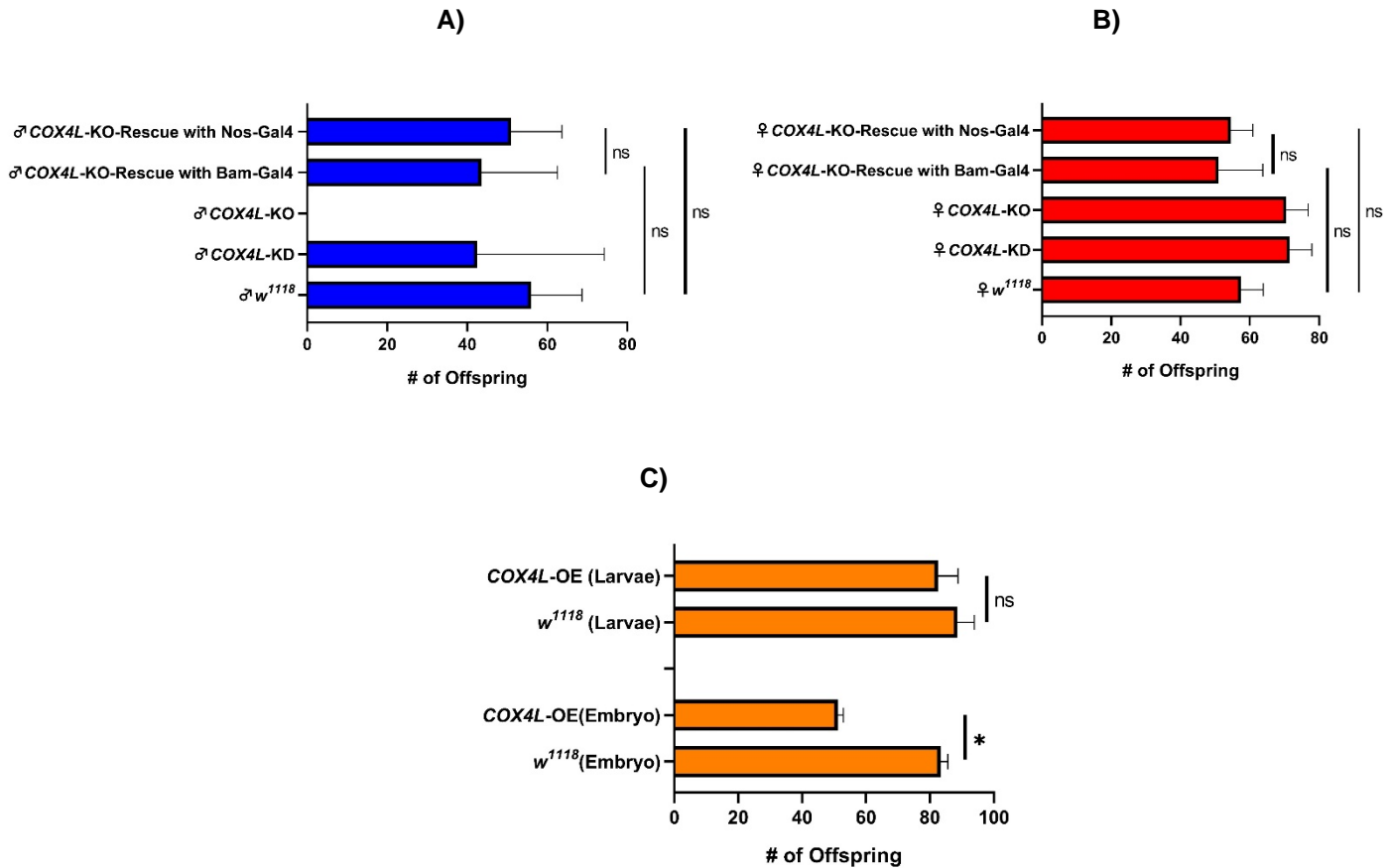
Figure 5. Fertility and viability test for *COX4L* in males and females. (A) Fertility study of COX4L in males is shown here. The knock down of *COX4L* with *Bam-Gal4* was semi-fertile while the *COX4L*-KO male are compete sterile. The *COX4L-KO* fertility phenotype was completely rescued by *Bam-Gal4* and *nos-Gal4* drivers. The overexpression of *COX4L* in soma reduce viability of the males which suggest that a fine-tune expression of this gene. (B) Fertility study of COX4L in females is shown here. The knock down of *COX4L* with *Bam-Gal4* and *COX4L*-KO females significantly performed higher than controls. The overexpression of *COX4L* in soma didn't show any viability effect in females. (C) Study the overexpression of *COX4L* in soma at larvae and embryo stages. The *COX4L* was overexpressed in soma with *Actin-5c* and *Tub* Gal4 drivers and the larvae and embryos were collected. The significant difference in emergence rate of overexpressed embryos which suggests that the lethality of overexpression occurs in the embryo stage.

94
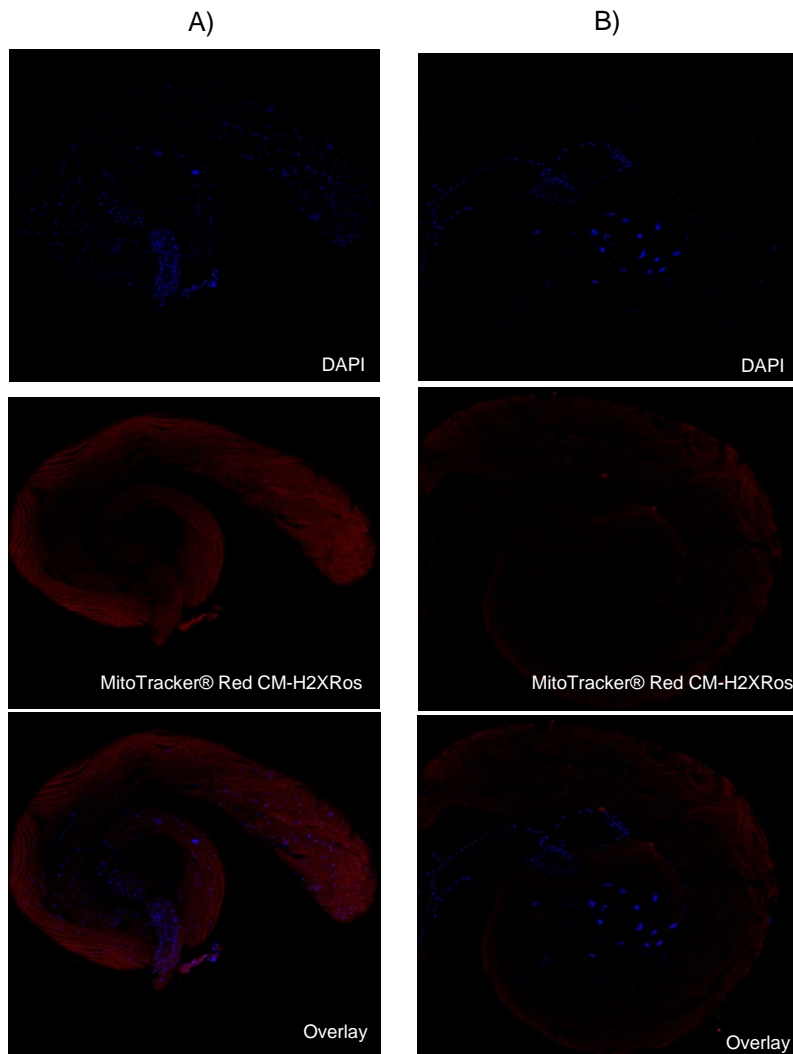
Figure 6. Confocal images of *COX4L*-KO and w^{1118} testes are shown here. The testes were dissected in PBS and then were stained for nucleus and the mitochondria with DAPI and MitoTracker Red CM-H2XRos, respectively. (A) *COX4L*-KO testis is shown here. (B) w^{1118} testis was used as a negative control for this comparison. The intensity of the red color was used to measure the amount of ROS production. The higher intensity of red color in the *COX4L*-KO suggests that this tissue have higher ROS production than the w^{1118} testis.
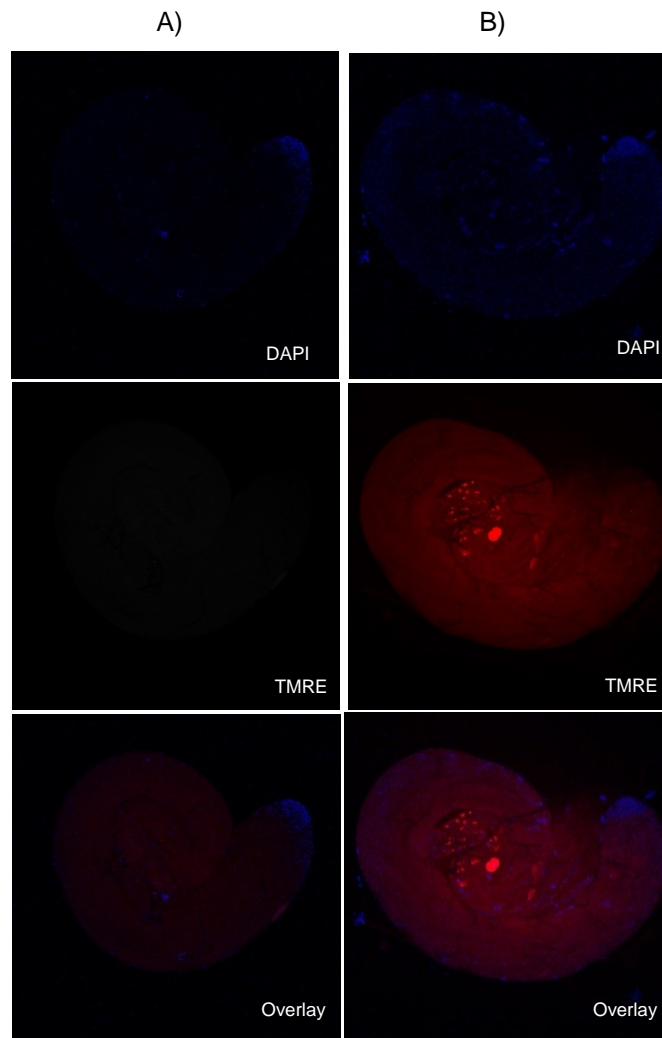
Figure 7. Confocal images of *COX4L*-KO and w^{1118} testes are shown in A and B. The testes were dissected in PBS and then were stained for nucleus and the mitochondria with DAPI and TMRE, respectively. (A) *COX4L*-KO testis is shown here. (B) w^{1118} testis was used as a control for this comparison. Only functional mitochondria are staining with TMRE. So the higher intensity of red color in the w^{1118} than *COX4L*-KO suggests that the mitochondria in the knockout are not functional.

Supplementary Tables

Supplementary Table 1. The detailed information about lines used in this study

Line name	Genotype	Stock Center
Actin5C-Gal4	*y[1] w[*]; P{w[+mC]=Act5C-GAL4}25FO1/CyO, y[+]*	BDSC
COX4L; CG10396	*P[GD4141]v1482*	VDRC
COX4L; CG10396	*P[KK102531]v106700*	VDRC
Nos Cas9 attp2	*(y,sc,v; +/+; nos-Cas9)*	Rainbow Transgenic Flies, Inc.
W1118- 60100	*y,w[1118];P{attP,y[+],w[3`]}*	VDRC
Tub-Gal4	*y[1] w[*]; P{w[+mC]=tubP-GAL4}LL7/TM3, Sb[1] Ser[1]*	BDSC

Supplementary Table 2. Primers and gene blocks designed for *COX4L* knockout and

confirm the *COX4L* knockout

COX4L guide 1 sense	CTTCGTGTAATTATGCGCAAGCACT
COX4L guide 1 antisense	AAACAGTGCTTGCGCATAATTACAC
COX4L homologous arm 1 EcoRI	GGGTGTCGCCCTTCGCTGAAGCAGGTGGAATgcttttcgct ggccatagctctcgtggcccttgcagccactctggttcgtgccaatatagatgataacct gccaattaacactgagggacggattgattttgatttcggcttcgaagaggcaatttgggg aaaaaaccgtaaaaataggaaatatctaaagaatagctcgacaattttttcacaaattaca aatttaattattaaataattatttggaaagttttaaaattttttgtttcggaattgtttttatttttgt gtttttttttttcatgatatttacttttaagagattggcaaatgcttcattctctaagtagagcga gattgtctttaatgtcttatattttctaaagtatagctttttttaaaattcttaagggtgggccaa caatgttattgcgatttaaaaattttttgaaaaaagtcaactagttgattcttaaacttttatcaa aatttcagatattgaaaactggacgtgggcaaaaaaaataattattgggcaaacagttct agatttcaaaaattcgattttttccgaacccagcttctttgagctgacatgacagccatttta aaaatgtttgttttttttttttgtgacaaaaaatttgatcttcataattttttgccacgccttaaaca atttttaagaagaagtaaaattttcagactatcttagtgctcaacgaagagtgcaattcaga acttaaaaagtacatctagtttgtagataaggaaactgtcatatttttttttttgtattcaacaaa cagactagagaaatttcattttcattcgacacgagcaacacaactgtcgaatttccggatg aagtaaaaaacaaaaaattgaaaagcgagtataaaataaaatacactcaaggtacagtt

	acgaccaagtAATTCTTGCATGCTAGCGGCCGCGGACATAT
COX4L guide 2 sense	CTTCGTTCTCTCGGTAGCACCATT
COX4L guide 2 antisense	AAACAATGGTGCTACCGAGAGAAC
COX4L homologous arm 2 XhoI	TGCATAAGGCGCGCCTAGGCCTTCTGCAGCggtgctaccgagagaacaagtggaagtagcacatcaacacgatgattttccgtaaactatgtacagaaacgtaactagcaaaatacaattcaacagcaaagtcgcttgccatttgatcgttacgtgctgaatcgggcaaatacccaatatctatagattttttgtgtctctagctgtgtaactcgactatagcatttcctcccgtttgaaattagggtgtgtatgtaaattctcagacacaacttaatttagtgtaattttagtccacgatagatatgttaagcattgaaatcgtgtcctgtgttcctttgactagtacacgtacactgcgcgtcatcagattagcgcctccctgtatgccaccgtttcatcttatgatctgtattttccattgcacgaaaatctatcaatgttattgtttttttgttcactgatattccctctctctttgagaataaaaaagaggttgagagaagaacagttatctcttttattctgctttgtgttaactttggcgcaaaattgaacacgtgtttttgcatcatcagattagcgcccccagttttgaatatcgtccataatttgaaaggtaggacaacaaaattttattaaaaacaaggaatcttatagaaaaaactataattgtggcaataaccgtgtcgttaccgcggcagattgcaggtccatacttcgaattgcttccaattcccacgactcctccgccaaaattcgagaaaaaagtggtataagcgcgagcaaatcgacgattccaagggtgacttagaaagctaaacatttaaagcacaacgaactaacGCTCGAGGCTCTTCCGTCAATCGAGTTCAAG
Forward primer to confirm the *COX4L* knockout	GTACACTGCCGTCGAAATGAG
Reverse primer to confirm the *COX4L* knockout	GGGTTCATTTGCAGCTGGATG

Supplementary Table 3. Viability results of *COX4L* knockdown in soma with *Actin5c-Gal4*

driver at 25°C, 27°C, and 29°C

Crosses	Average number of progeny at 25 °C	Average number of progeny at 27 °C	Average number of progeny at 29 °C
♀COX4L-KK x ♂Actin5c-Gal4	50.0	91.3	101.3
♂COX4L-KK x ♀Actin5c-Gal4	51.3	86.7	100.7
♀COX4L-KK x ♂w1118	44.3	80.7	85.7
♂COX4L-KK x ♀w1118	49.0	65.0	87.3
♀Actin5c-Gal4 x ♂w1118	46.0	78.7	79.3
♂Actin5c-Gal4 x ♀w1118	56.0	80.0	82.0
♀COX4L-GD x ♂Actin5c-Gal4	38.0	80.0	64.7
♂COX4L-GD x ♀Actin5c-Gal4	38.7	59.3	72.7
♀COX4L-GD x ♂w1118	38.3	85.7	66.3
♂COX4L-GD x ♀w1118	43.3	70.7	58.3
♀Actin5c-Gal4 x ♂w1118	36.7	67.3	62.0
♂Actin5c-Gal4 x ♀w1118	44.0	72.7	67.3

Supplementary Table 4. Physical protein interaction analysis of *COX4L* with BioGRID

#	Node1	Node2	Node1 annotation	Node2 annotation	Score
1	*CG10396*	*CG11043*	Cytochrome-c oxidase activity	Cytochrome-c oxidase activity	0.997
2	*CG10396*	*CG30354*	Cytochrome-c oxidase activity	Ubiquinol-cytochrome-c reductase activity	0.976
3	*CG10396*	*CG3731*	Cytochrome-c oxidase activity	Ubiquinol-cytochrome-c reductase activity; metalloendopeptidase activity	0.987
4	*CG10396*	*CG4169*	Cytochrome-c oxidase activity	Ubiquinol-cytochrome-c reductase activity	0.989
5	*CG10396*	*CG7580*	Cytochrome-c oxidase activity	Ubiquinol-cytochrome-c reductase activity; ubiquinone binding	0.969
6	*CG10396*	*CoVa*	Cytochrome-c oxidase activity	Cytochrome-c oxidase activity	0.991
7	*CG10396*	*CoVb*	Cytochrome-c oxidase activity	Cytochrome-c oxidase activity	0.997
8	*CG10396*	*RFeSP*	Cytochrome-c oxidase activity	Rieske iron-sulfur protein	0.989
9	*CG10396*	*mt:CoI*	Cytochrome-c oxidase activity	Cytochrome-c oxidase activity	0.983
10	*CG10396*	*mt:CoII*	Cytochrome-c oxidase activity	Cytochrome-c oxidase activity	0.967

Supplementary Table 5. Fertility results of *COX4L* knockdown in soma with *Bam-Gal4*

driver at 25°C, and 27°C

Crosses	Average number of progeny at 25 °C	Average number of progeny at 27 °C
♀ [♀KK x ♂Gal4](27c) x ♂W1118(25c)	69.3	80.3
♀[♀KK x ♂W1118] (27c) x♂W1118(25c)	57.0	68.0
♂ [♀KK x ♂Gal4](27c) x ♀W1118(25c)	44.7	51.7
♂[♀KK x ♂W1118] (27c) x♀W1118(25c)	72.7	71.0
♀[♂GD x ♀Gal4] (27c) X ♂W1118(25c)	69.0	80.3
♀[♂GD x ♀W1118](27c) X ♂W1118(25c)	60.3	65.7
♂[♂GD x ♀Gal4] (27c) X ♀W1118(25c)	47.0	59.7
♂[♂GD x ♀W1118] (27c) X ♀W1118(25c)	62.3	74.0

Supplementary Figures

Supplementary figure1- Rescue the *COX4L-KO* phenotype with driving COX4L-ORF with *bam* and *nos* Gal4 drivers
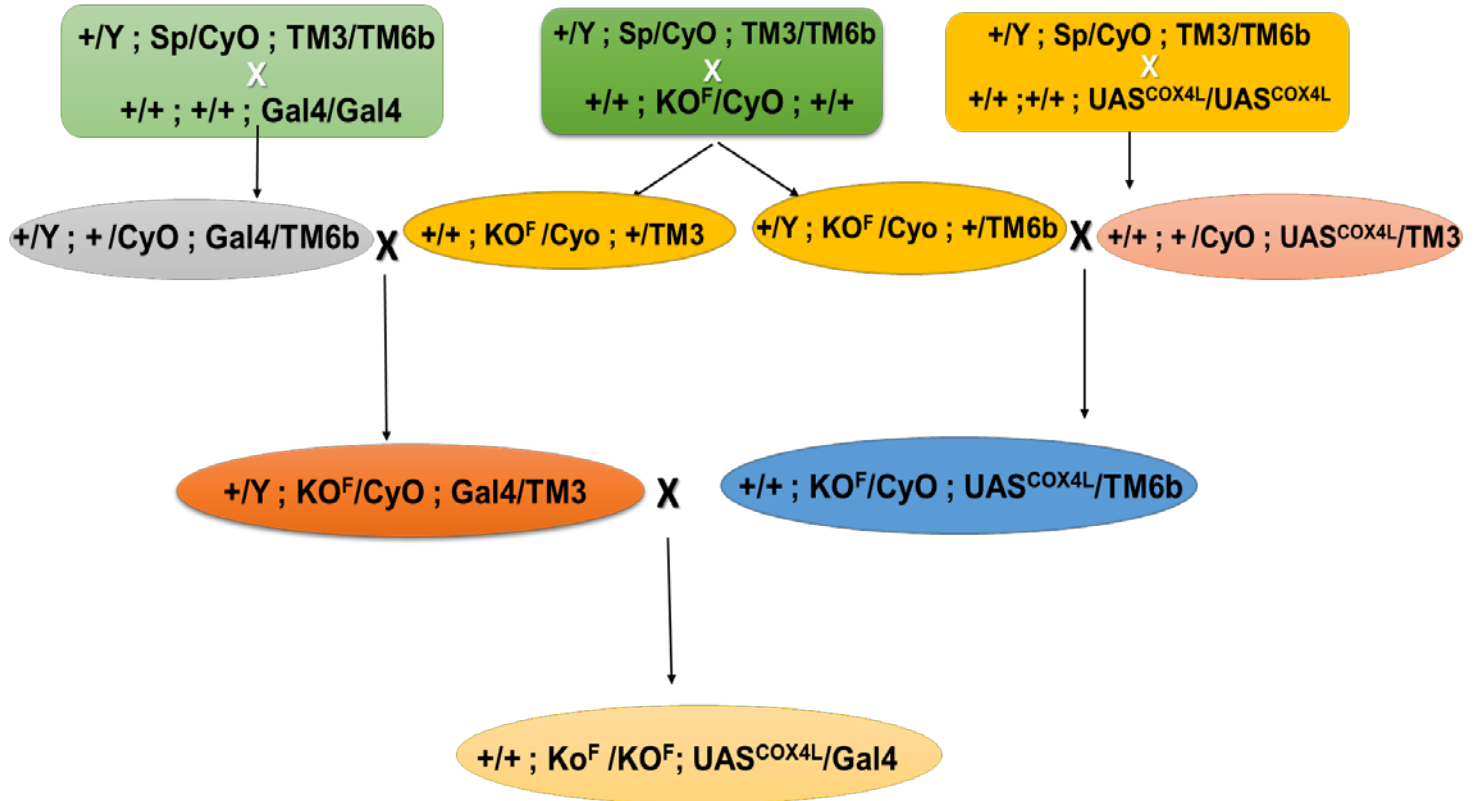
Supplementary Files

Supplementary File 1. Protein alignment of COX4 and COX4L with Clustal Omega

```
                        cov    pid 161        .              .        2         .         .        . 240
 1 Avag/COX4   100.0% 100.0%  GIVFAYDGWRRIGYTFDKGDSWNDEKLKKQLQYHIAARQGPMTGLSSKWDYETGLFFFYFLSFTCIFLF---------
 2 Hmel/COX4    67.9%  26.3%  -----------------------LRRMLELRVNPIDGLSSKWDYDNDNWKRIFSCKVVNLRISFYFTLATIGA
 3 Bmor/COX4    65.6%  36.3%  FPFWAMIFNHYYVVE-PLPESLSKESQKAQLRRMLELRVNPIDGLSSKWDYDNDRWKVDYFYEALDPSRKLRNRHQRSHG
 4 Dple/COX4    83.3%  24.1%  FTFWAAMFYHHYVVE-PLPSTFSKESQKAQLRRMLELRVNPIDGISSLWDYDNDKWLVAIVLAFIGMSLANPVPSGLGGL
 5 Nvit/COX4    90.5%  25.8%  LSLWIHLFIHHKVYV-KWPESFSTENRIACYYRQVDLNMQPFSGIYRPPGAELRIHAKNKSMLKKAHEDETN--------
 6 Glos/COX4    74.2%  24.3%  HGHLDS----------------------LHEFVRIR----
 7 Agla/COX4    79.2%  36.1%  VALWVYYGIKICTLN-LVPDTFSPENREAQLRRILDLQINPIOGIASKWDYCKDDWKK--------------------
 8 Dpon/COX4    79.6%  39.3%  CALWVIYGLKAFVYP-PLPSSFSPENRAAQLRRILDLQINPIOGISSKWDYEKDDWKK--------------------
 9 Tcas/COX4    78.7%  41.1%  IALWGFYGMKTFVYA-PLPESFNEENRRAQLRRIIDLQINPVOGLSSKWDYEKDDWKK--------------------
10 Amel/COX4    74.2%  37.9%  FSLLTYILLKIYAFP-PLPESFNLENRLAQLERMKLLQVNPIAGISSKN--------------------
11 Bimp/COX4    75.6%  37.9%  FSLWVVMFLKTFAYP-PLPESFNLENRLAQLERMKILEVNPITGISAQK--------------------
12 Bter/COX4    80.1%  37.5%  FSLWVVMFLKTFAYP-PLPESFNLENRLAQLERMKILEVNPISGISAKKINLICLCLKT--------------------
13 Acep/COX4    72.4%  31.6%  AAIWLFLYFKEFSNP-PLPETFSMERRLAQLDRMKKLDMNPIDGLCARK--------------------
14 Sinv/COX4    74.2%  34.8%  SGIWLYLYFKTFAYP-PLPETFTLERRLAQLDRMKKLDMNPIEGWVGSRKNLN--------------------
15 Apis/COX4    86.9%  38.3%  IAVGSFVGVRKFLTSTADDPTLSLEYRQACKRMIDLRVDPIDGLSSWMDYEKNTWKS--------------------
16 Cson/COX4    77.4%  36.9%  LALLVSLWMAAFVYT-DHPVTLTEERKKAQLKRMLDLNVNPIHGISSKWDYENNRWK--------------------
17 Aaeg/COX4    78.7%  38.7%  MSLATMLLMKAFVYD-DIPVTFDDEHQKAQLKRMLDLGVGNIIGLISSKWDYDNNKWK--------------------
18 Cqui/COX4    79.6%  37.0%  MSITCMLLMKAFVYE-KIPETFDDEHQKAQLKRMLDLGVGPVNGLSSKWDYDNNKWK--------------------
19 Adar/COX4    90.0%  30.6%  MSLAGMLLLKAFVYE-PIPETFDEEHQKAQLKRMLDLNINPIHGVSSKWDYDNNKWK--------------------
20 Agam/COX4    78.7%  34.3%  MSLIGMMLLKAFVYE-PIPETFDEEHQKAQLKRMLDLNINPIHGVSSKWDYDNNKWK--------------------
21 Rpro/COX4    78.7%  40.0%  LAVWLYMFFKLFALYNPTLPGTFDDAHQKAQLKRMLELKMNPIITGLSSKWDYEKNDWKK--------------------
22 Lcup/COX4    81.9%  36.9%  LALWVAIFMNTYVYD-EMPVTFDEEHQKAQLKRMLDLEMNPVTGISSKWDYENKRWK--------------------
23 Msca/COX4    81.9%  36.9%  LALWVAIFMNTYVYD-EMPVTFDEEHQKAQLKRMLDLEMNPVTGISSKWDYENKRWK--------------------
24 Dper/COX4    80.5%  35.5%  GAIWIAILMNLFVYD-ELPVTFDEEHQKAQLKRIIDLEMNPVIGLTSKWDYETKQWKN--------------------
25 Dpse/COX4    80.5%  35.5%  GAIWIAILMNLFVYD-ELPVTFDEEHQKAQLKRIIDLEMNPVGLTSKWDYETKQWKN--------------------
26 Dsec/COX4    80.5%  36.1%  AAIWVAVLMNLFVYD-ELPVTFDEEHQKAQLQRIIDLEINPVIGLTSKWDYENKKWKN--------------------
27 Dsim/COX4    80.5%  36.1%  AAIWVAVLMNLFVYD-ELPVTFDEEHQKAQLQRIIDLEINPVIGLTSKWDYENKKWKN--------------------
28 Dmel/COX4    80.5%  36.1%  AAIWVAVLMNLFVYD-ELPVTFDEEHQKAQLQRIIDLEINPVIGLTSKWDYENKKWKN--------------------
29 Dere/COX4    80.5%  36.1%  AAIWVAVLMNLFVYD-ELPVTFDEEHQKAQLQRIIDLEINPVIGLTSKWDYENKKWKN--------------------
30 Dyak/COX4    80.5%  36.1%  AAIWVAVLMNLFVYD-ELPVTFDEEHQKAQLQRIIDLEINPVIGLTSKWDYENKKWKN--------------------
31 Dgri/COX4    80.5%  35.0%  AAIWIAVLMNLFVYD-ELPVTFDEEHQKAQLKRIIDLEINPVGVTSLWDYENKKWKN--------------------
32 Dmoj/COX4    80.5%  33.9%  AAIWIAVLMNLFVYD-ELPVTFDEEHQKAQLKRMIDLEINPVIGLTSKWDYENNKWKN--------------------
33 Dvir/COX4    80.5%  35.5%  AAIWIAVLMNLFVYD-ELPITFDEEHQKAQLKRMIDLEINPVIGLTSKWDYENKKWKN--------------------
34 Dana/COX4    80.5%  36.1%  AAIWVAVLMNLFVYD-ELPVTFDEEHQKAQLKRIIDLEINPVIGLTSKWDYENKKWKN--------------------
35 Dwil/COX4    80.5%  37.2%  AAIWIAVLMNLFVYD-ELPITFDEEHQKAQLKRIIDLEINPVGLTSKWDYENNKWKN--------------------
   consensus/100%                                             .h...h.ht.......
   consensus/90%                hshhhhhhhphhsh...hP.ohs.Epp.AQLpRhhtLphsPlpGlsuph...........
   consensus/80%                hulhhhhhhphasYs..hP.gFs.EpphAQLcRhl-LphNPlpGloS-WDY-pp.ah....
   consensus/70%                hulWshhhhphFlys..lP.TFscEppcAQLcRhl-LclNPlsGloSKWDY-pppWK....
```
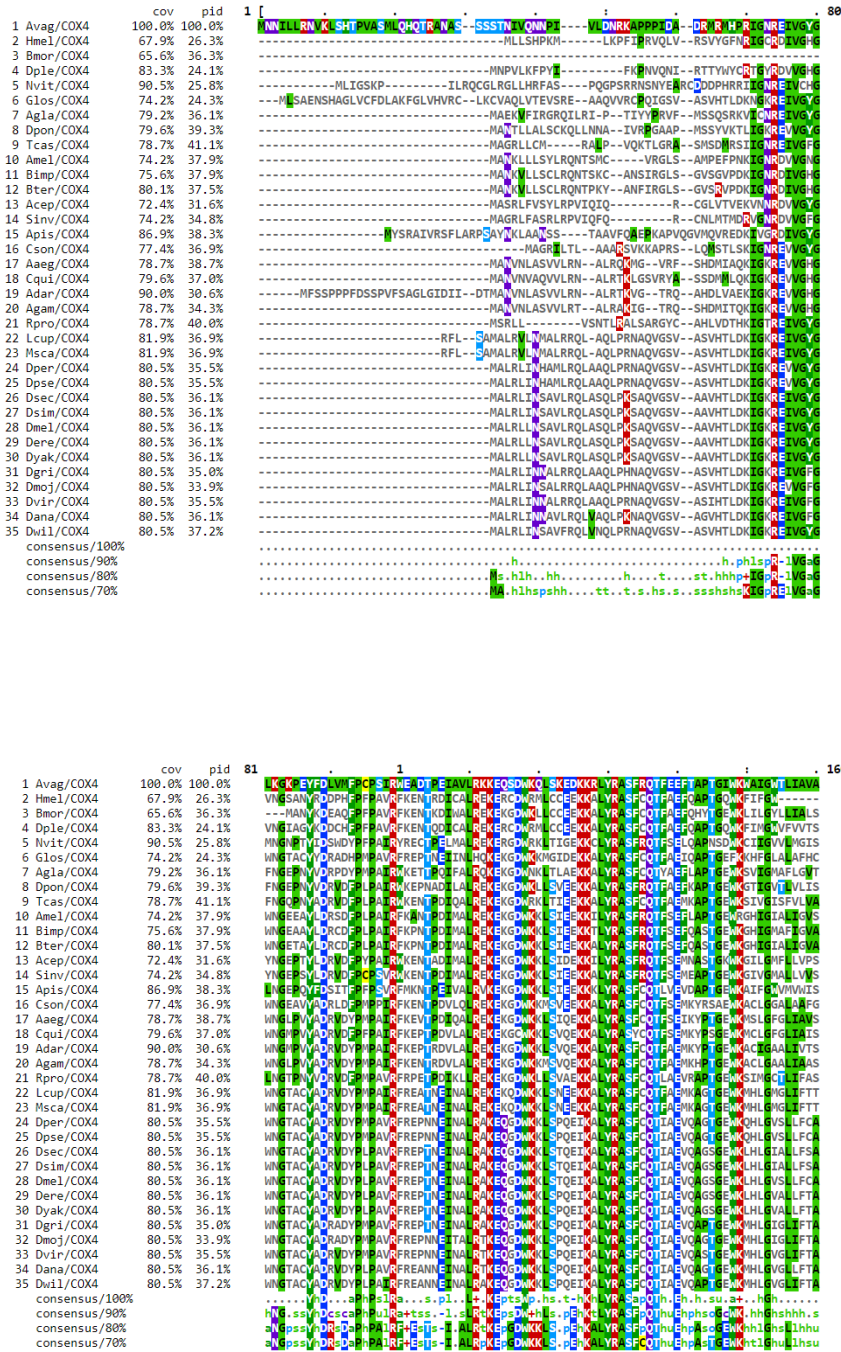
```
                        cov    pid 241         :         .         .         .         . ] 292
 1 Avag/COX4   100.0% 100.0%  --------------------------------------------------
 2 Hmel/COX4    67.9%  26.3%  IGVARAKYLFSAPQPPTSLTSFLFY-------------------------
 3 Bmor/COX4    65.6%  36.3%  --------------------------------------------------
 4 Dple/COX4    83.3%  24.1%  MEHEQIMRSMNDKEWQYEGIPIELPLRRSETVKKTTPKKKITSKGFVYVSFE
 5 Nvit/COX4    90.5%  25.8%  --------------------------------------------------
 6 Glos/COX4    74.2%  24.3%  --------------------------------------------------
 7 Agla/COX4    79.2%  36.1%  --------------------------------------------------
 8 Dpon/COX4    79.6%  39.3%  --------------------------------------------------
 9 Tcas/COX4    78.7%  41.1%  --------------------------------------------------
10 Amel/COX4    74.2%  37.9%  --------------------------------------------------
11 Bimp/COX4    75.6%  37.9%  --------------------------------------------------
12 Bter/COX4    80.1%  37.5%  --------------------------------------------------
13 Acep/COX4    72.4%  31.6%  --------------------------------------------------
14 Sinv/COX4    74.2%  34.8%  --------------------------------------------------
15 Apis/COX4    86.9%  38.3%  --------------------------------------------------
16 Cson/COX4    77.4%  36.9%  --------------------------------------------------
17 Aaeg/COX4    78.7%  38.7%  --------------------------------------------------
18 Cqui/COX4    79.6%  37.0%  --------------------------------------------------
19 Adar/COX4    90.0%  30.6%  --------------------------------------------------
20 Agam/COX4    78.7%  34.3%  --------------------------------------------------
21 Rpro/COX4    78.7%  40.0%  --------------------------------------------------
22 Lcup/COX4    81.9%  36.9%  --------------------------------------------------
23 Msca/COX4    81.9%  36.9%  --------------------------------------------------
24 Dper/COX4    80.5%  35.5%  --------------------------------------------------
25 Dpse/COX4    80.5%  35.5%  --------------------------------------------------
26 Dsec/COX4    80.5%  36.1%  --------------------------------------------------
27 Dsim/COX4    80.5%  36.1%  --------------------------------------------------
28 Dmel/COX4    80.5%  36.1%  --------------------------------------------------
29 Dere/COX4    80.5%  36.1%  --------------------------------------------------
30 Dyak/COX4    80.5%  36.1%  --------------------------------------------------
31 Dgri/COX4    80.5%  35.0%  --------------------------------------------------
32 Dmoj/COX4    80.5%  33.9%  --------------------------------------------------
33 Dvir/COX4    80.5%  35.5%  --------------------------------------------------
34 Dana/COX4    80.5%  36.1%  --------------------------------------------------
35 Dwil/COX4    80.5%  37.2%  --------------------------------------------------
   consensus/100%              ..................................................
   consensus/90%               ..................................................
   consensus/80%               ..................................................
   consensus/70%               ..................................................
```

Chapter 4

Mode of evolution and coevolution between N-mt genes in

Drosophila melanogaster

Abstract

Most of the mitochondrial proteins are encoded by nuclear-encoded mitochondrial genes but only a few of them interact directly with the mitochondria-encoded proteins. The interactions between these proteins are crucial to maintaining mitochondrial functions as the disruption of these interactions can cause deleterious effects. However, the mitochondria genomes evolve faster than the nuclear genome because they have a higher mutation rate, which could interrupt this crosstalk. To maintain the crosstalk between the two genomes, the N-mt genes might experience compensatory evolution in response to the mutations of the fast-evolving mitochondria genome. Also, in *Drosophila melanogaster*, there are many testis-specific duplicates that might even respond to mitochondria genome male-harming mutations that the female cannot select against. As a first approximation to understand if those evolutionary processes are operating in *Drosophila*, we estimated the rates of evolution and coevolution in N-mt genes and test the positive selection in testis-specific N-mt genes. We also test for a higher population differentiation (high Fst) of N-mt genes between populations than for other genes. Some of these expectations are challenged in *Drosophila* by the recent introgressions of the mitochondria caused by *Wolbachia* spreading. We find that testis-specific N-mt duplicated genes have a higher rate of evolution and are more often under positive selection than the parental genes. There is strong coevolution of testis-specific N-mt duplicated genes with each other and the same is true for their parental genes supporting that the different classes of genes fucntion together but separately from each other in two differentiated mitochondria types. The high coevolution between testis-specific N-mt duplicated genes could also be due to their fast evolution. Our initial Fst analyses of testis-specific N-mt duplicated genes support neither consistently higher nor lower differentiation (expectation if N-mt duplicated genes are co-introgressed with the mitochondria during *Wolbachia* invasion) of any of these genes between populations.

Introduction

The presence of two genomes (mtDNA and nDNA) in eukaryotic cells is a consequence of the acquisition of the mitochondrion into the modern eukaryotes' ancestor (symbiotic theory) which has been suggested to have been crucial in facilitating the evolution of eukaryotic life (Archibald 2015). During the course of eukaryotic evolution, many of mitochondrial genes translocate to the nucleus and only a few genes remain encoded by mtDNA in most modern eukaryotes (Wolstenholme 1992; Bjorkholm, et al. 2015) which play critical roles in mitochondrial functions (Rand, et al. 2004; Wolff, et al. 2014). However, the mt-encoded proteins are not solely enough for mitochondrial functions, and this organelle needs to import many proteins from outside. The nuclear-encoded mitochondrial genes (N-mt) are transcribed in the nucleus and translated in the cytoplasm. Then, the mature protein is imported into the mitochondria to fulfilled mitochondrial functions either with interactions with mt-encoded products (e.g., OXPHOS complexes) or without interactions (e.g., mitochondrial membranes; (Pagliarini, et al. 2008; Bar-Yaacov, et al. 2012). As a result, extensive molecular interactions exist between the components that are encoded by mtDNA and nDNA (Rand, et al. 2004; Gray 2012; Hill 2015). However, mtDNA evolves faster than nDNA (Brown, et al. 1979) which can interrupt this crosstalk and even cause deleterious effects (Hudson, et al. 2005; Reinhardt, et al. 2013; Dowling 2014). It has been shown that tight coevolution between interacting subunits that are encoded by nDNA and mtDNA is crucial to maintaining mitochondrial functions (Grossman, et al. 2004; Bar-Yaacov, et al. 2012; Dowling 2014) and play a pivotal role in population hybrid breakdown in some taxa (Burton, et al. 2006; Lee, et al. 2008; Barreto and Burton 2013; but also see Montooth, et al.

2010; Meiklejohn, et al. 2013), speciation events (Gershoni, et al. 2009), disease susceptibility (Hudson, et al. 2005; Potluri, et al. 2009), and longevity (Zhu, et al. 2014). In addition, several studies have shown the importance of mito-nuclear coevolution to physical interactions between mtDNA and nuclear DNA-encoded factors in OXPHOS complexes (Rand, et al. 2004; Meiklejohn, et al. 2007; Gershoni, et al. 2010). The interactions between gene products from two different genomes raise some interesting questions about how nuclear and mitochondrial genomes coevolve in the light of different mutation and substitution rates (Rand, et al. 2004; Hill 2015). Even though mitochondria originated from an endosymbiotic integration of α-proteobacteria (Sagan 1967; Pittis and Gabaldon 2016), eukaryotes have very divergent mitochondrial substitution rate (Lynch, et al. 2006; Baer, et al. 2007; Sloan, et al. 2009; Havird and Sloan 2016). In contrast, bilaterian animals and plants have very similar substitution rates in their nuclear genomes (Lynch 2010). Among the insects, very high rates of substitution and rearrangements of gene order have been reported for mitochondrial genome of Hymenoptera and Psocodea (Yoshizawa and Johnson 2013; Li, et al. 2015; Song, et al. 2016) while Diptera shows much lower rates (Oliveira, et al. 2008; Song, et al. 2016; Yan, et al. 2019). For example, the synonymous substitution rate of the mitochondria genome is greater 40 times in *Nasonia* (parasitoid wasp; Oliveira, et al. 2008) and nine times in *D. melanogaster* than the average nuclear gene (Moriyama and Powell 1997). To maintain the crosstalk between these two genomes, the N-mt genes might have a compensatory evolution in response to the mutations of fast-evolving mitochondria genome.

Several other features of mitochondria contribute to its mode of evolution. The effective population size of mitochondria is small because they have a unique inheritance

pattern (maternal inheritance) and mildly deleterious mutations might accumulate in mitochondrial genome simply by genetic drift. Furthermore, the mitochondrial genome doesn't recombine (Howell 1997; Arbogast 2015; but see Rokas, et al. 2003; Kraytsberg, et al. 2004) which could also help the accumulation of mildly deleterious mutation in mitochondrial genome by hitchhiking during a selective sweep (Meiklejohn, et al. 2007). In addition, since mitochondria are inherited only from the mother (maternal inheritance), the mitochondrial genome is not under direct selection in males. Therefore, the population frequency of mitochondrial mutations that are deleterious only to the males could increase and even become fixed (Frank and Hurst 1996; Partridge and Hurst 1998; Rand, et al. 2006; Montooth, et al. 2010). This phenomenon is known as "mother's curse" (Gemmell, et al. 2004; Hill 2015). Some of the testis-specific N-mt duplicates might evolve to compensate for mtDNA male-harming mutations that do not hurt females and cannot be selected against (Rogell, et al. 2014) and coevolve with mtDNA differently. So, to maintain the crosstalk between these two genomes, the N-mt genes might show compensatory evolution in response to fast-evolving mitochondria genome leading to the mtDNA and nDNA coevolution to maintain mitochondrial integrity and nDNA-mtDNA interactions (Ehrlich and Raven 1964; Janzen 1980). The unidirectional inheritance of mitochondria could cause the unidirectional responses of N-mt genes to the fixation of mitochondrial mutations which eventually impose selective pressure on the N-mt interactor genes.

Since nuclear genome has large effective population size and can recombine, it is expected that compensatory nuclear mutations will alleviate the effects of any deleterious mutations in the mitochondrial genome (Oliveira, et al. 2008; Barreto and Burton 2013; Sloan,

et al. 2014; Barreto, et al. 2018). The compensatory mutations in nDNA which select for additional adaptive mitochondrial substitutions could also create mitochondria-nucleus feedback (Rand, et al. 2004). Based on this hypothesis, additional functional changes could compensate for functional mutations in the mtDNA. These extra mutations could happen either within the mtDNA or by coevolution between genes encoded by the mtDNA and N-mt genes. The more difference between the mutation rate of interactors, the stronger selective pressure on epistasis between them (Levin, et al. 2014). The coevolution could be different for broadly expressed N-mt genes and testis-specific N-mt genes. As mentioned above, some of the testis-specific N-mt duplicate genes might evolve to compensate for male-deleterious substitutions. These effects might be so strong to be detected as linkage disequilibrium between populations and incompatibilities in the introgression of mitochondria have been observed (Innocenti, et al. 2011; Meiklejohn, et al. 2013; Darras and Aron 2015). Consistent with the coevolution hypothesis, studies of N-mt proteins in lineages with different mitochondrial evolutionary rates (human, an angiosperm genus (*Silene*), and insects showed that substitution rates in N-mt genes are higher in lineages that have faster mitochondria evolutionary rate (Sloan, et al. 2014; Havird, et al. 2015; Li, et al. 2017). For these reasons, interactions between mitochondrial and N-mt proteins could promote the coevolution of two genomes.

Like other animals, *D. melanogaster* has circular, double-stranded, compacted mtDNA of 19,517 bp (Lewis, et al. 1995). However, mtDNA size in *D. melanogaster* varies among different populations, mostly due to differences in the length of the control region (Fauron and Wolstenholme 1976; Fauron and Wolstenholme 1980; Montooth, et al. 2009).

Although *Drosophila* mitochondrial DNA genome has the same gene content as vertebrates, the gene order and distribution on both DNA strands are different. *Drosophila* mtDNA contains 37 genes encoding 13 polypeptides, 22 tRNAs and two rRNAs. In *Drosophila*, the control region is rich with A+T, which could be from 1 to 5 Kb in size among different subgroups. Two different types of repeats (type I and II) could be found in this region. The origin of replication (OR) is in this region, too (Garesse and Kaguni 2005).

Studying the nonrandom association between alleles at a nuclear locus with a haplotype at mtDNA locus which is often called mito-nuclear linkage disequilibrium (Mito-nuclear LD; (Asmussen, et al. 1987; Schnabel and Asmussen 1989) is one way to study the coevolution between mtDNA and nDNA. However, mito-nuclear LD can be influenced by population structure (i.e., an admixture of two species or divergent populations; Nei and Li 1973), gene flow (positive or negative mito-nuclear selection (Schnabel and Asmussen 1992) and non-random mating/inbreeding (Asmussen, et al. 1989; Slatkin 2008). There was only one study that calculated the mito-nuclear LD in a sample of human populations, and a significant LD was reported, but the associations were generally weak, across the genome and explained by the mixed population structure in humans (Sloan, et al. 2015). However, this approach (calculating LD between N-mt genes and mitochondrial genes between populations) in *D. melanogaster* would not be expected to show the interaction between the two genomes because of recent *Wolbachia* introgressions (Pool, et al. 2012; Richardson, et al. 2012). *Wolbachia* has been introgressed from one population to another in *Drosophila melanogaster* several times recently due to the cytoplasmic incompatibility they cause. Cytoplasmic incompatibility leads to infected females being compatible with all males but uninfected

females being compatible only with uninfected males lowering the fitness of uninfected females. Several mitochondria introgressions which have swept away mtDNA variation between different populations of *D. melanogaster* has been reported. For this reason, the coevolution between N-mt genes and mt genes might not be expected anymore. However, the introgression of the mt genome that accompanied the introgression of *Wolbachia* might also bring the introgression of N-mt genes that work better with that mt genome. So, we might expect little differentiation between populations for some N-mt genes (i.e., the genes that show the highest strength of coevolution with the mt genome). Interestingly, co-introgression of mt genome and an N-mt gene has been documented before in *Drosophila* (Beck, et al. 2015). Alternatively, adaptation to the new mt genome might occur (i.e., beneficial replacements) as different populations adapt during the sweep. So, higher or lower differentiation between populations (i.e., Fst) for N-mt genes might be expected depending on the gene or population.

Another evolutionary parameter has been used to investigate the potential interactions between N-mt and mitochondrial genes and the coevolution between two genomes. Evolutionary rate correlation (ERC; evolutionary rate covariance; evolutionary rate coevolution) is a sequence analysis method which can be used to find genes with similar evolutionary histories. ERC can also identify proteins with physical interaction, and discover new proteins in pathways, particularly for the fast-evolving genes where coevolution play an important role (de Juan, et al. 2013; Findlay, et al. 2014; Clark and Wolfe 2015). Functionally related genes which share the same history would have a higher ERC value (de Juan, et al. 2013). Recently, a strong positive ERC was reported between mitochondrial encoded proteins

and N-mt proteins (OXPHOS and rRNA) for 64 holometabolous insect taxa. Interestingly, the mitochondrial-nuclear rate correlation was stronger for mitochondrial encoded proteins that are in contact with N-mt proteins (Yan, et al. 2019). Unfortunately, this study did not calculate the ERC between mitochondrial proteins and N-mt duplicated proteins because the duplicates were not presented in all 64 holometabolous insect taxa studied. So, ERC can provide information about the coevolution of N-mt genes that interact with each other.

Here, we study the rate and mode of evolution of N-mt genes and their coevolution with each other to understand if the genes are under selection potentially to compensate for mtDNA fast evolution and if they are working together. In addition, in *Drosophila*, testis-specific N-mt proteins have evolved fast after gene duplication (Gallach, et al. 2010; Eslamieh, et al. 2017). However, they have been proposed to be under relaxed selection (Havird and McConie 2019). We study if they have been under positive selection and potentially compensating for mtDNA changes or male-harming mutations using divergence and polymorphism data that should have more power to detect a small fraction of adaptive changes than divergence only. We have also studied population differentiation (Fst) of testis-biased N-mt duplicated genes among 30 different *D. melanogaster* populations all over the world. Based on the new expectations after the *Wolbachia* introgressions. We expect that the N-mt genes will have either less or more divergence between different populations than other genes in the genome as a consequence of introgression or adaptation, respectively.

We find that testis-specific N-mt duplicated genes have a high rate of evolution and are often under positive selection. There is strong coevolution of testis-specific N-mt duplicated genes with each other but could be due to their fast evolution. The initial population

differentiation analyses of testis-specific N-mt duplicated genes do not support neither higher nor lower differentiation for any of these genes between populations.

Materials and Methods

The list of the N-mt genes in families and N-mt single genes in *D. melanogaster* were obtained from a previous study of these genes (Eslamieh, et al. 2017). The availability of previously sequenced genomes, analyses and databases for *D. melanogaster* and close related species allows us to perform an in-depth analysis of N-mt gene evolution and their interactions for these species. The evolutionary rate and rate covariance were used to study the rate and mode of evolution for N-mt parental genes, testis-specific N-mt duplicate and single-copy genes. The Ka/Ks ratio of genes were retrieved from (Li, et al. 2010). We also retrieved and performed McDonald–Kreitman (MK) test for the testis-specific N-mt duplicate genes. Results were retrieved from a population study of *D. simulans* that was compared to *D. melanogaster* (Begun, et al. 2007) for 20 N-mt duplicated genes and their parental genes. Because two genes had very close to the significant p-value, we performed the MK test by adding more sequences from *D. melanogaster*, *D. simulans*, and *D. yakuba*. In short, the CDS sequences of *D. melanogaster* were retrieved for Zambia population from Popfly Database (https://popfly.uab.cat; Hervas, et al. 2017). For *D. simulans* and *D. yakuba* gene sequences were retrieved from the previous study of polymorphism and divergence in this species (Begun, et al. 2007). The sequences then aligned with MEGA-X (Kumar, et al. 2018) and the MK test was performed with DNAsp (Rozas, et al. 2017).

Values for the evolutionary rate covariation (ERC) between genes in *Drosophila* (calculated from 12 *Drosophila* sequenced genomes) were retrieved from the University of Pittsburg online database (https://csb.pitt.edu/erc_analysis/index.php; Findlay, et al. 2014) for different sets of genes comparisons. The statistics for the different comparisons were also calculated.

F_{ST} values were retrieved from Popfly database (https://popfly.uab.cat; Hervas, et al. 2017) between nine *Drosophila* populations (Table 1). To focus only on the coding region of the genes (exons) and to avoid the intron in the Fst analysis, only the retro N-mt duplicated genes with testes-biased expression (21 genes) were studied in these analyses, and the F_{ST} was obtained for the smallest window size (1kb). The F_{ST} of these genes were compared to a control set (21 retrogenes). For the control, a random set of retrogenes with testes-biased expression was selected form RetrogenDB (http://yeti.amu.edu.pl/retrogenedb; Rosikiewicz, et al. 2017). To have a fair comparison and minimize the changes between the N-mt set and random set, the distribution of random genes on the different chromosome and the length of the random genes were carefully considered when the control set was selected. Therefore, the two sets are comparable.

Results and Discussion

Some N-mt duplicated genes are under positive selection

The Ka / Ks ratio (i.e., nonsynonymous substitutions per nonsynonymous site divided by synonymous substitutions per synonymous site ratio) analysis of N-mt duplicated and parental genes show that all of the N-mt duplicates and their parental genes are under purifying

selection (Supplementary table 8) However, the average ratio for N-mt duplicates was significantly higher than parental genes (0.089 vs. 0.049, respectively; P-value= 0.0188) which suggest that the duplicates are evolving at least two times faster than their parental genes. In addition, we compared the ratio between N-mt genes and two other random sets of genes. The Ka/Ks of N-mt duplicates wasn't significantly higher than the random set of the genes (0.089 vs. 0.069). In general, testis-specific genes are evolving faster than other genes (Meiklejohn, et al. 2003). Therefore we compare the N-mt duplicated genes to the set of random genes that highly expressed in testis. The Ka/Ks of N-mt genes wasn't significantly different than testis-specific genes (0.089 vs 0.11; P-value = 0.156).

We then explored more detailed analyses using polymorphism and divergence. The results of the MK test on N-mt duplicated genes and their parental genes suggest that at least three N-mt duplicated genes (*CG33791*, *CG14740*, and *CG11913*) are under positive selection after retrieving data from previous analyses while only one of the parental genes is under positive selection (C*G8323*; Supplementary table 5). We added polymorphism data from *D. yakuba*. For *CG6255*, the P value of MK test became significant after we added polymorphism data from *D. yakuba* in the comparison with *D. melanogaster* (Dn/Ds= 42/57; Pn/Ps = 5/20; MK P-value = 0.0414).

Higher ERC between N-mt duplicated genes than other gene sets

The ERC values of 20 testis-specific N-mt duplicated genes, their parental genes, and single N-mt genes were retrieved from the ERC web server (https://csb.pitt.edu/erc_analysis) and analyzed separately (Supplementary tables 2-4). The mean ERC-value between testis-

specific N-mt duplicated genes was higher than their parental genes (0.114 *vs*. 0.062; Table 2) and single N-mt genes (0.114 *vs* 0.021; Table 2). However, those differences were not significantly different (P-value=0.5150 and 0.1635, respectively). This high evolutionary rate covariation suggests that N-mt genes have the same evolutionary history and either interact or work on the same pathways. The ERC analyses of a random sample of testis-specific genes revealed, however, that the high ERC rate can be a consequence of high rate of evolution of this kind of genes (ERC_{testis}=0.192). The ERC values between testis-specific N-mt duplicated genes and between N-mt parental genes were significantly higher than observed by chance between a random set of genes in the genome with 100,000 permutations (Supplementary table 2). The highest ERC was observed between *CG14740*, a TCA cycle duplicate and a duplicate for a subunit of complex IV, *CG10396* (+0.829; Figure 1). The analysis of ERC values between duplicated genes and parental in the same family showed that the ERC value between parental gene and duplicated gene is often negative. This suggests that they evolve at a different rate and might have a different function. On average, the duplicated gene showed higher ERC with the other duplicates than their parental gene when we analyzed them together. The highest ERC in this comparison was between a transporter gene, *Colt* and *Sdha* in complex II (+ 0.865; Figure 2). The ERC value between single genes was the lowest ERC in this analysis as the single N-mt gene have diverse functions (Figure 3). Also, we have checked the highest ERC value for each parental and N-mt genes in the entire genome and found that 77.5% (31/40) of N-mt genes have the highest ERC value in the genome with single N-mt genes (Supplementary table 1).

Fst for N-mt genes between different populations

The introgression of *Wolbachia* between *Drosophila* populations might have swept away mtDNA variation between different populations of *D. melanogaster* and any signal of coevolution between N-mt genes and mt genes within this species. For this reason, we might not expect coevolution between N-mt genes and mt genes in this species any longer. However, the introgression of the mt genome that accompanied the introgression of *Wolbachia* might also bring the introgression of N-mt genes that work better with that mt genome or adaptation as described in the Introduction. As an initial approximation to test these expectations we study the differentiation between N-mt genes in *Drosophila* between populations. In particular, we study Fst (Weir and Cockerham 1984), a measure of population differentiation, for testis-specific N-mt retroduplicated genes (21 genes) and their parental. We compare the level of differentiation in these genes to a random set of testis-specific retrogenes of similar lengths and genomic location (Supplementary table 6). We study retrogenes to make sure most of the signal comes from the gene itself and not from differences in intron length between genes. The average can be compared and it is expected to be lower if consistent introgression has occurred for the testis-specific N-mt retrogenes compared to the random set of retrogenes. The average can be higher if consistent adaptation has occurred for the testis-specific N-mt retrogenes compared to the random set of retrogenes. If both have taken place depending on the gene, the variance for Fst values between genes will higher for testis-specific N-mt retrogenes than for the random set of retrogenes.

All pairwise comparisons between nine *D. melanogaster* populations showed that there are no significant differences for Fst between testis-specific N-mt retrogenes and a

random set of retrogenes (Figure 4). Therefore, there is no support for higher or lower differentiation the N-mt duplicated genes between these populations.

Tables

Table 1. Populations of *D. melanogaster* used in Fst analysis

ID	Continent	Number of Samples	Populations included (# samples)
1	America	10	RAL (10)
2	Equatorial Africa	30	EA (10), EF (10), RG (10)
3	Europe / North Africa	20	EG (10), FR (10)
4	Southern Africa	30	SD (10), SP (10), ZI (10)

Table 2. ERC compassions between N-mt genes.

	Number of the genes	Mean ERC value	Median ERC value	Probability of observed mean by chance 100,000 permutations
N-mt Duplicated Genes	20	0.114	0.151	0.00255
N-mt between New and parental Genes	20	0.062	0.056	0.04232
N-mt Single Genes	409	0.021	0.004	0.26359

Figures
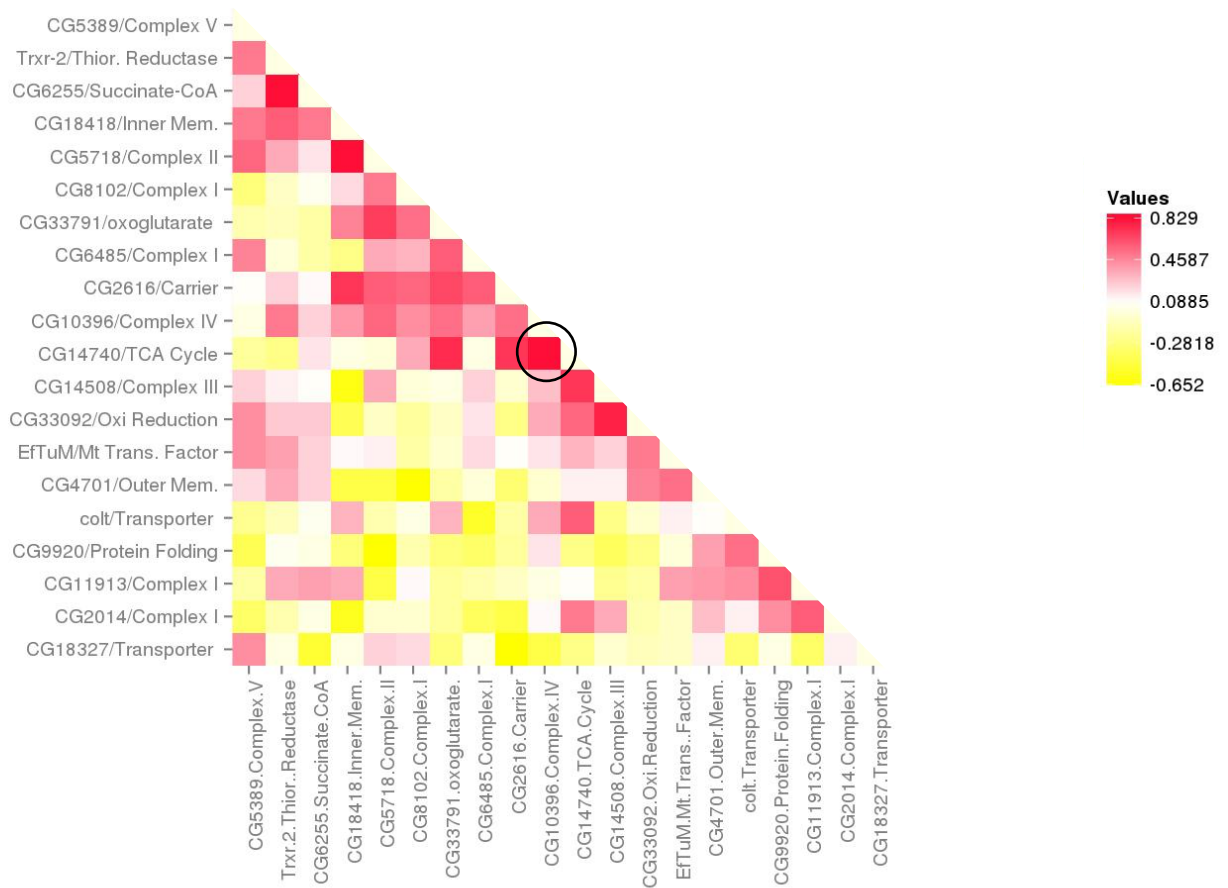
Figure 1. ERC Heatmap of N-mt duplicated genes. The pairwise comparison of ERC values between any two genes is shown here. The circle shows the highest ERC value in this set between *COX4L* and *CG14740*.
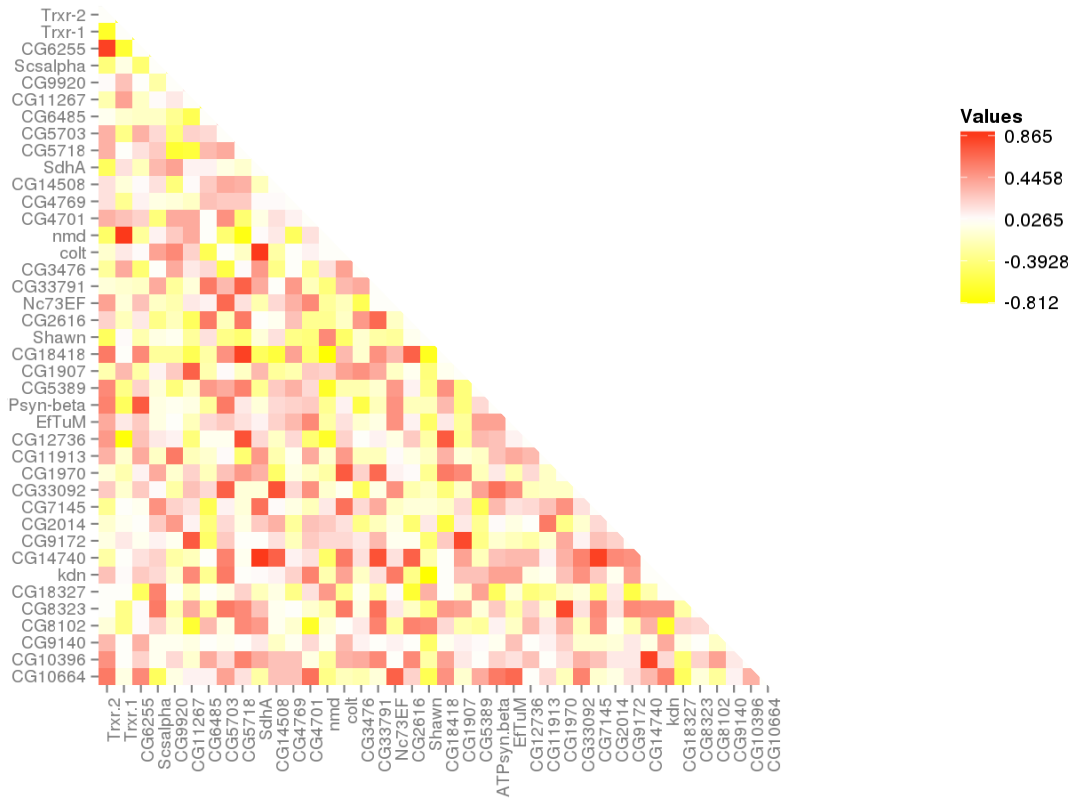
Figure 2. ERC Heatmap of N-mt duplicated genes and their parental genes. The pairwise comparison of ERC values between any two genes is shown here. The parental gene and the duplicated genes are shown after each other. The circle shows the highest ERC value in this set between *Colt* and *SdhA*.
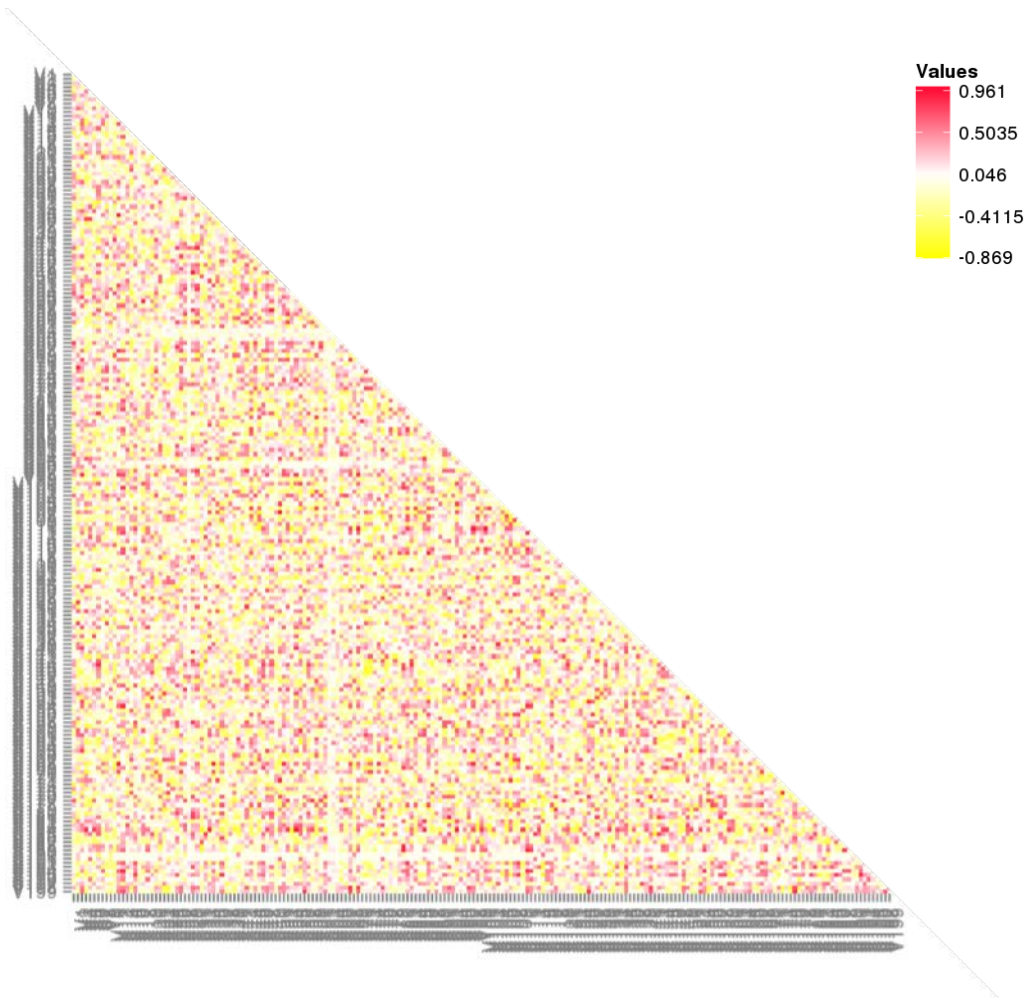
Figure 3. ERC Heatmap of N-mt single genes. The pairwise comparison of ERC values between any two genes is shown here.
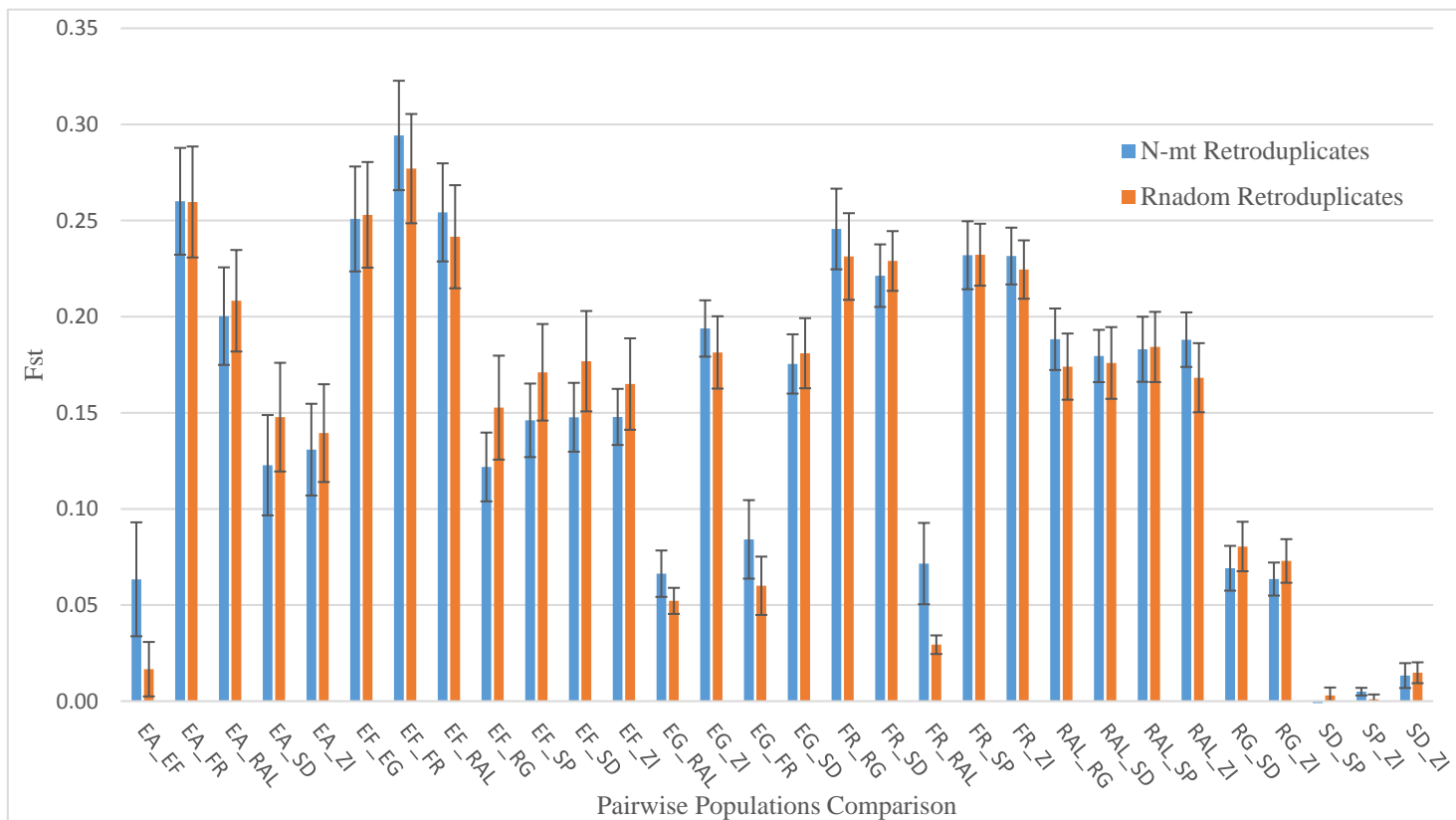
Figure 4. Pairwise comparison of Fst between N-mt retroduplicates and a random set of retrogenes. No significant differences was observed for any of these sets genes when they were compared between two populations.

Supplementary Tables

Supplementary Table 1. Highest ERC value between duplicated, parental, and single N-mt genes in *D. melanogaster*

Family	Gene Name	Child/Duplicate	ERC between New and Parent genes	Highest ERC with all new and Parent genes	ERC Value	Highest ERC with all new, Parent, and single genes	ERC Value	Highest ERC with entire genome
Fam1	Trxr-2	New	-0.658	CG6255	0.827	CG6255		CG6255
	Trxr-1	Parent		nmd	0.859	nmd		nmd
Fam2	CG6255	New	-0.433	Trxr-2	0.827	Trxr-2		Trxr-2
	Scsalpha1	Parent		CG18327	0.532	CG14077	0.793	CG14077
Fam3	CG9920	New	0.127	CG11913	0.593	CG6673	0.666	CG6673
	CG11267	Parent		CG9172	0.705	CG9172		CG9172
Fam4	CG6485	New	0.17	CG33791	0.579	CG18624	0.826	CG18624
	CG5703	Parent		CG33092	0.676	blw	0.784	blw
Fam5	CG5718	New	-0.121	CG18418	0.822	mRpL38	0.878	mRpL38
	SdhA	Parent		colt	0.865	CG7311	0.906	CG7311
Fam6	CG14508	New	0.038	CG33092	0.747	Menl-1	0.801	Menl-1
	CG4769	Parent		CG33791	0.438	Coprox	0.632	Coprox
Fam7	CG4701	New	0.16	CG10664	0.612	CG10761	0.923	CG10761
	nmd	Parent		Trxr-1	0.859	Trxr-1		Trxr-1
Fam8	colt	New	0.41	SdhA	0.865	SdhA		SdhA
	CG3476	Parent		CG1907	0.473	CG5976	0.669	CG5976
Fam9	CG33791	New	0.012	CG14740	0.734	CG4995	0.872	CG4995
	Nc73EF	Parent		CG10664	0.664	blw	0.697	blw
Fam10	CG2616	New	0.008	CG18418	0.692	Menl-1	0.872	Menl-1
	Shawn	Parent		CG8102	0.514	CG7382	0.67	CG7382
Fam11	CG18418	New	-0.031	CG5718	0.822	CG4942	0.88	CG4942
	CG1907	Parent		CG9172	0.775	CG9172		CG9172
Fam12	CG5389	New	0.189	CG15390	0.538	CG15390	0.831	CG15390
	ATPsynbeta	Parent		CG6255	0.712	Ucp4B	0.74	Ucp4B
Fam13	CG2014	New	0.197	CG11913	0.566	CG10920	0.748	CG10920
	CG9172	Parent		CG1907	0.755	CG1907		CG1907
Fam14	CG14740	New	ND	SdhA	0.86	CG7311	0.912	CG7311
	kdn	Parent		CG5703	0.595	blw	0.741	blw
Fam15	mEFTu1	New	0.093	CG10664	0.652	Mtch	0.704	Mtch
	CG12736	Parent		CG5718	0.718	CG5718		CG5718
Fam16	CG11913	New	-0.152	CG9920	0.593	Ucp4B	0.866	Ucp4B
	CG1970	Parent		CG8323	0.782	CG4995	0.816	CG4995
Fam17	CG18327	New	-0.253	Scsalpha	0.56	CG8004	0.776	CG8004
	CG8323	Parent		CG1970	0.782	CG6914	0.819	CG6914
Fam18	CG8102	New	-0.174	CG2616	0.54	CG4169	0.846	CG4169
	CG9140	Parent		kdn	0.364	CG3902	0.759	CG3902
Fam19	CG10396	New	0.348	CG14740	0.852	Menl-1	0.932	Menl-1
	CG10664	Parent		Nc73EF	0.664	CG11110	0.767	CG11110
Fam20	CG33092	New	0.130	CG14508	0.747	CG4908	0.792	CG4908
	CG7145	Parent		CG14740	0.809	CG12400	0.82	CG12400

Supplementary Table 2. ERC value between duplicated N-mt genes in *D. melanogaster*

	CG5389	Trxr-2	CG6255	CG18418	CG5718	CG8102	CG33791	CG6485	CG2616	CG10396	CG14740	CG14508	CG33092	EfTuM	CG4701	colt	CG9920	CG11913	CG2014	CG18327
CG5389	N/A	0.501	0.225	0.493	0.538	-0.296	-0.13	0.445	0.061	0	-0.195	0.23	0.419	0.415	0.184	-0.227	-0.4	-0.154	-0.329	0.415
Trxr-2	0.501	N/A	0.827	0.581	0.341	-0.083	-0.098	-0.008	0.223	0.475	-0.266	0.138	0.258	0.382	0.337	-0.111	0.04	0.332	-0.127	0
CG6255	0.225	0.827	N/A	0.502	0.149	0.04	-0.172	-0.17	0.103	0.215	0.15	0.059	0.237	0.235	0.222	0.033	0.014	0.362	0.005	-0.494
CG18418	0.493	0.581	0.502	N/A	0.822	0.193	0.469	-0.245	0.692	0.4	ND	-0.584	-0.398	0.1	-0.418	0.32	-0.292	0.336	-0.55	ND
CG5718	0.538	0.341	0.149	0.822	N/A	0.5	0.665	0.327	0.589	0.554	-0.013	0.342	-0.076	0.145	-0.442	-0.148	-0.635	-0.443	-0.044	0.209
CG8102	-0.296	-0.083	0.04	0.193	0.5	N/A	0.531	0.309	0.54	0.424	0.355	-0.025	-0.205	-0.152	-0.652	0.01	-0.12	0.103	-0.037	0.203
CG33791	-0.13	-0.098	-0.172	0.469	0.665	0.531	N/A	0.579	0.644	0.527	0.734	0.002	-0.062	-0.043	-0.175	0.306	-0.294	-0.199	-0.184	-0.287
CG6485	0.445	-0.008	-0.17	-0.245	0.327	0.309	0.579	N/A	0.573	0.371	ND	0.228	0.163	0.187	-0.005	-0.512	-0.341	-0.12	-0.38	ND
CG2616	0.061	0.223	0.103	0.692	0.589	0.54	0.644	0.573	N/A	0.526	0.688	-0.033	-0.239	0.078	-0.321	-0.174	-0.18	-0.073	-0.442	-0.63
CG10396	0	0.475	0.215	0.4	0.554	0.424	0.527	0.371	0.526	N/A	0.829	0.275	0.343	0.155	-0.056	0.352	0.166	0.023	0.11	-0.438
CG14740	-0.195	-0.266	0.15	ND	-0.013	0.355	0.734	ND	0.688	0.829	N/A	0.684	0.554	0.315	0.147	0.578	-0.266	0.064	0.497	-0.24
CG14508	0.23	0.138	0.059	-0.584	0.342	-0.025	0.002	0.228	-0.033	0.275	0.684	N/A	0.747	0.225	0.134	-0.252	-0.376	-0.213	0.337	-0.056
CG33092	0.419	0.258	0.237	-0.398	-0.076	-0.205	-0.062	0.163	-0.239	0.343	0.554	0.747	N/A	0.478	0.466	-0.035	-0.264	-0.152	-0.141	-0.09
EfTuM	0.415	0.382	0.235	0.1	0.145	-0.152	-0.043	0.187	0.078	0.155	0.315	0.225	0.478	N/A	0.505	0.142	-0.004	0.378	-0.063	-0.086
CG4701	0.184	0.337	0.222	-0.418	-0.442	-0.652	-0.175	-0.005	-0.321	-0.056	0.147	0.134	0.466	0.505	N/A	0.079	0.362	0.391	0.288	0.143
colt	-0.227	-0.111	0.033	0.32	-0.148	0.01	0.306	-0.512	-0.174	0.352	0.578	-0.252	-0.035	0.142	0.079	N/A	0.521	0.438	0.144	-0.322
CG9920	-0.4	0.04	0.014	-0.292	-0.635	-0.12	-0.294	-0.341	-0.18	0.166	-0.266	-0.376	-0.264	-0.004	0.362	0.521	N/A	0.593	0.435	0.01
CG11913	-0.154	0.332	0.362	0.336	-0.443	0.103	-0.199	-0.12	-0.073	0.023	0.064	-0.213	-0.152	0.378	0.391	0.438	0.593	N/A	0.566	-0.332
CG2014	-0.329	-0.127	0.005	-0.55	-0.044	-0.037	-0.184	-0.38	-0.442	0.11	0.497	0.337	-0.141	-0.063	0.288	0.144	0.435	0.566	N/A	0.119
CG18327	0.415	0	-0.494	ND	0.209	0.203	-0.287	ND	-0.63	-0.438	-0.24	-0.056	-0.09	-0.086	0.143	-0.322	0.01	-0.332	0.119	N/A

Supplementary Table 3. ERC value between duplicated and parental N-mt genes in *D. melanogaster*

	Tnxr-2	Tnxr-1	CG6255	Scsalpha	CG9920	CG11267	CG6485	CG5703	CG5718	SdhA	CG14508	CG4769	CG4701	nmd	colt	CG3476	CG33791	Nc73EF	CG2616	Shawn	CG18418	CG1907	CG5389	ATPsyn-be	EfTuM	CG12736	CG11913	CG1970	CG33092	CG7145	CG2014	CG9172	CG14740	kdn	CG18327	CG8323	CG8102	CG9140	CG10396	CG10664
Tnxr-2	N/A	-0.66	0.827	-0.392	0.04	-0.227	-0.008	0.347	0.341	-0.48	0.138	0.131	0.337	-0.45	-0.11	-0.285	-0.098	0.427	0.223	-0.501	0.581	-0.233	0.501	0.546	0.382	0.438	0.332	-0.081	0.258	-0.329	-0.127	-0.069	-0.266	0.283	0	-0.007	-0.083	0.32	0.475	0.59
Tnxr-1	-0.66	N/A	-0.638	-0.06	0.263	0.408	-0.13	-0.374	ND	0.152	-0.087	-0.328	0.27	0.809	0.127	0.366	-0.138	-0.082	-0.188	0.048	ND	0.315	-0.374	-0.484	0.12	-0.752	-0.129	-0.225	-0.127	0.015	-0.04	0.118	ND	0.042	ND	-0.355	-0.358	0.021	0	-0.063
CG6255	0.827	-0.64	N/A	-0.433	0.014	-0.166	-0.17	0.334	0.149	-0.19	0.059	0.061	0.222	-0.29	0.033	-0.441	-0.172	0.291	0.103	-0.24	0.502	-0.263	0.225	0.712	0.235	0.265	0.362	0.081	0.237	-0.012	0.005	-0.008	0.15	0.237	-0.494	0.042	0.04	0.354	0.215	0.517
Scsalpha	-0.39	-0.06	-0.433	N/A	-0.243	0.052	-0.171	0.177	0.231	0.328	0.151	-0.055	-0.386	0.073	0.404	0.007	0.37	-0.147	-0.356	-0.052	-0.29	0.071	-0.099	-0.074	-0.058	0.114	-0.159	0.366	0.056	0.485	0.237	0.078	0.207	0.183	0.58	0.595	0.221	-0.037	-0.015	-0.5
CG9920	0.04	0.263	0.014	-0.243	N/A	0.127	-0.341	-0.392	-0.635	0.398	-0.376	-0.101	0.362	0.256	0.521	0.365	-0.294	-0.209	-0.18	-0.012	-0.292	0.236	-0.4	-0.024	-0.004	0.091	0.593	-0.14	-0.264	0.209	0.435	-0.027	-0.266	-0.13	0.01	-0.233	-0.12	-0.039	0.166	-0.046
CG11267	-0.23	0.408	-0.166	0.052	0.127	N/A	-0.542	0.215	-0.599	0.078	0.032	-0.147	0.382	0.373	0.217	0.115	-0.089	0.1	-0.493	-0.302	-0.525	0.684	-0.154	-0.064	-0.197	-0.397	-0.185	0.209	0.294	0.159	0.079	0.706	-0.201	0.501	-0.201	0.385	-0.639	-0.059	-0.125	0.089
CG6485	-0.01	-0.13	-0.17	-0.171	-0.341	-0.542	N/A	0.17	0.327	0.082	0.228	0.289	-0.005	0	-0.51	0.182	0.579	0.085	0.573	0.158	-0.245	-0.235	0.445	-0.279	0.187	-0.01	-0.12	-0.202	0.163	-0.529	-0.38	-0.473	ND	-0.32	ND	-0.283	0.309	0.009	0.371	0.091
CG5703	0.347	-0.37	0.334	0.177	-0.392	0.215	0.17	N/A	0.394	-0.06	0.379	0.258	0.477	-0.41	0.021	-0.571	0.309	0.635	-0.194	-0.352	0.463	0.051	0.393	0.576	0.249	-0.024	0.036	0.289	0.676	0.075	0.177	0.232	0.543	0.595	-0.071	0.576	-0.055	-0.022	0.185	0.403
CG5718	0.341	ND	0.149	0.231	-0.635	-0.599	0.327	0.394	N/A	-0.12	0.342	0.235	-0.442	-0.73	-0.15	0.056	0.665	0.138	0.589	-0.418	0.822	-0.155	0.538	0.12	0.145	0.716	-0.443	0.453	-0.076	-0.044	0.007	-0.013	0	0.209	0.5	0.5	-0.007	0.554	0.151
SdhA	-0.48	0.152	-0.188	0.328	0.398	0.078	0.082	-0.055	-0.121	N/A	-0.203	0.038	0.034	0.861	0.444	-0.366	0.002	0.194	-0.033	0.137	-0.508	0.304	-0.332	-0.106	0.071	0.189	0.453	0.351	-0.104	0.626	0.247	0.018	-0.86	0.029	-0.181	0.268	0.455	0.123	0.362	-0.246
CG14508	0.138	-0.09	0.059	0.151	-0.376	0.032	0.228	0.379	0.342	-0.2	N/A	0.038	0.134	0.098	-0.366	0.002	0.194	-0.033	0.137	-0.584	-0.281	0.23	-0.154	-0.213	-0.449	0.747	0.034	0.337	0.136	0.684	0.091	-0.056	-0.014	-0.025	-0.149	0.275	0.285			
CG4769	0.131	-0.33	0.061	-0.055	-0.101	-0.147	0.289	0.258	0.235	0.032	0.038	N/A	0.065	-0.48	0.022	0.013	0.438	0.343	0.277	-0.357	0.419	-0.211	0.342	0.225	0.315	0.181	-0.017	-0.042	0.192	-0.048	-0.291	-0.196	0.08	0.202	-0.37	0.022	-0.113	-0.032	0.265	0.274
CG4701	0.337	0.27	0.222	-0.386	0.362	0.382	-0.005	0.477	-0.442	-0.07	0.134	0.065	N/A	0.16	0.079	-0.337	-0.175	0.504	-0.321	-0.404	-0.418	0.25	0.184	0.252	0.505	-0.432	0.391	-0.172	0.466	-0.177	0.288	0.24	0.147	0.542	0.143	-0.111	-0.652	0.167	-0.056	0.612
nmd	-0.45	0.809	-0.287	0.073	0.256	0.373	0	-0.414	-0.726	0.034	0.098	-0.483	0.16	N/A	0.002	0.129	-0.409	-0.301	-0.329	0.502	-0.812	0.206	-0.67	-0.329	-0.267	-0.672	-0.132	-0.266	-0.119	0.116	0.23	0.181	-0.453	-0.24	0.439	-0.239	-0.102	-0.009	-0.308	-0.343
colt	-0.11	0.127	0.033	0.404	0.521	0.217	-0.512	0.021	-0.148	0.861	-0.252	0.022	0.079	0.002	N/A	0.41	0.306	-0.181	-0.174	-0.469	0.32	0.419	-0.227	0.08	0.142	0.206	0.438	0.699	-0.035	0.623	0.144	0.186	0.578	0.402	-0.322	0.57	0.01	0.315	0.352	-0.186
CG3476	-0.29	0.366	-0.441	0.007	0.365	0.115	0.182	-0.571	0.056	0.444	-0.366	0.013	-0.337	0.129	0.41	N/A	0.371	-0.514	0.446	-0.114	-0.12	0.473	-0.218	-0.57	-0.13	0.007	-0.202	0.227	-0.499	0.174	-0.37	0.121	0.16	-0.16	-0.014	-0.07	0.096	0.079	0.368	-0.23
CG33791	-0.1	-0.14	-0.172	0.37	-0.294	-0.089	0.579	0.309	0.685	0.389	0.002	0.438	-0.175	-0.41	0.306	0.371	N/A	0.012	0.644	-0.267	0.469	0.384	-0.13	-0.197	-0.043	0.064	-0.199	0.634	-0.062	0.39	-0.184	0.325	0.734	0.157	-0.287	0.603	0.531	-0.003	0.527	-0.08
Nc73EF	0.427	-0.08	0.291	-0.147	-0.209	0.1	0.085	0.635	0.138	-0.24	0.194	0.343	0.504	-0.3	-0.18	-0.514	0.012	N/A	-0.168	-0.298	0.311	-0.173	0.444	0.488	0.48	-0.058	0.048	0.061	0.467	-0.157	-0.045	-0.024	0.133	0.59	0.054	0.127	-0.285	0.071	0.021	0.664
CG2616	0.223	-0.19	0.103	-0.356	-0.18	-0.493	0.573	-0.194	0.589	0.026	-0.033	0.277	-0.321	-0.33	-0.17	0.446	0.644	-0.168	N/A	-0.008	0.692	-0.014	0.061	-0.196	0.078	0.243	-0.073	0.058	-0.09	-0.442	-0.032	0.688	-0.37	-0.63	-0.173	0.54	0.019	0.526	0.21	
Shawn	-0.5	0.048	-0.24	-0.052	-0.012	-0.302	0.158	-0.352	-0.418	-0.11	0.137	-0.357	-0.404	0.502	-0.114	-0.267	-0.298	-0.008	-0.685	N/A	-0.031	0.493	-0.365	-0.377	-0.465	0.1	-0.17	-0.311	-0.15	-0.102	0.125	-0.032	-0.239	-0.15	0.357	-0.391	0.514	-0.453	-0.25	-0.463
CG18418	0.581	ND	0.502	-0.29	-0.292	-0.525	-0.245	0.463	0.822	-0.51	-0.584	0.419	-0.418	-0.81	0.32	-0.12	0.469	-0.685	0.692	-0.685	N/A	-0.377	0.493	0.307	0.1	0.702	0.336	0.575	-0.398	-0.17	-0.55	0.255	ND	ND	0.448	0.193	-0.011	0.4	0.496	
CG1907	-0.23	0.315	-0.263	0.071	0.236	0.684	-0.235	0.051	-0.155	0.304	-0.281	-0.211	0.25	0.206	0.419	0.473	0.384	-0.173	-0.014	-0.377	-0.031	N/A	-0.377	-0.394	-0.284	-0.319	-0.077	0.501	-0.173	0.182	0.095	0.775	0.401	0.293	-0.11	0.419	-0.231	0.079	0.106	-0.13
CG5389	0.501	-0.37	0.225	-0.099	-0.4	-0.154	0.445	0.393	0.538	-0.33	0.23	0.342	0.184	-0.67	-0.23	-0.218	-0.13	0.444	0.061	-0.365	0.493	-0.377	N/A	0.189	0.415	0.313	-0.154	-0.195	0.419	-0.521	-0.329	-0.351	-0.195	0.294	0.415	-0.103	-0.296	0.046	0	0.395
ATPsyn-be	0.546	-0.48	0.712	-0.074	-0.024	-0.064	-0.279	0.576	0.12	-0.11	0.178	0.225	0.252	-0.33	0.08	-0.57	-0.197	0.488	-0.196	-0.187	0.307	-0.394	0.189	N/A	0.423	0.268	0.283	0.126	0.602	0.313	0.043	-0.11	0.265	0.402	-0.491	0.235	0.062	-0.021	0.212	0.568
EfTuM	0.382	0.12	0.235	-0.058	-0.004	-0.197	0.187	0.249	0.145	0.071	0.225	0.315	0.505	-0.27	0.142	-0.13	-0.043	0.48	0.078	-0.465	0.1	-0.284	0.415	0.423	N/A	0.093	0.378	-0.229	0.478	0.142	-0.063	-0.194	0.315	0.411	-0.086	-0.248	-0.152	0.272	0.155	0.652
CG12736	0.438	-0.75	0.265	0.114	0.091	-0.397	-0.01	-0.024	0.746	0.189	-0.154	0.181	-0.432	-0.67	0.206	0.007	0.064	-0.058	0.243	-0.264	0.702	-0.319	0.313	0.268	0.093	N/A	0.324	0.177	-0.184	0.186	0.012	-0.248	0.328	-0.06	-0.404	0.204	0.317	0.01	0.266	0.009
CG11913	0.332	-0.13	0.362	-0.159	0.593	-0.185	-0.12	0.036	-0.443	0.453	-0.213	-0.017	0.391	-0.13	0.438	-0.202	-0.199	0.048	-0.073	-0.17	0.336	-0.077	-0.154	0.283	0.378	0.324	N/A	-0.152	-0.152	0.266	0.566	-0.122	0.064	-0.03	-0.332	-0.111	0.103	0.091	0.023	0.176
CG1970	-0.08	-0.23	0.081	0.366	-0.14	0.209	-0.202	0.289	0.453	0.361	-0.449	-0.042	-0.172	-0.27	0.699	0.227	0.634	0.061	0.058	-0.311	0.575	0.501	-0.195	0.126	-0.229	0.177	-0.152	N/A	-0.162	0.48	-0.348	0.384	0.356	0.288	-0.153	0.782	0.354	0.216	0.156	-0.177
CG33092	0.258	-0.13	0.237	0.056	-0.264	0.294	0.163	0.676	-0.076	-0.1	0.747	0.192	0.466	-0.12	-0.04	-0.499	-0.062	0.467	-0.239	-0.15	-0.398	-0.173	0.419	0.602	0.478	-0.184	-0.152	-0.162	N/A	0.13	-0.141	0.08	0.554	0.478	-0.09	0.145	-0.205	-0.139	0.343	0.553
CG7145	-0.33	0.015	-0.012	0.485	0.209	0.159	-0.529	0.075	-0.089	0.626	0.034	-0.048	-0.177	0.116	0.623	0.174	0.39	-0.157	-0.09	-0.102	-0.17	0.182	-0.521	0.313	0.142	0.186	0.266	0.48	0.13	N/A	0.191	0.283	0.909	0.233	-0.403	0.505	0.474	-0.134	0.288	-0.11
CG2014	-0.13	-0.04	0.005	0.237	0.435	0.079	-0.38	0.177	-0.044	0.247	0.337	-0.291	0.288	0.23	0.144	-0.37	-0.184	-0.045	-0.442	0.125	-0.55	0.095	-0.329	0.043	-0.063	0.012	0.566	-0.348	-0.141	0.191	N/A	0.197	0.497	-0.01	0.119	0.149	-0.037	-0.106	0.11	-0.073
CG9172	-0.07	0.118	-0.008	0.078	-0.027	0.766	-0.473	0.232	0.007	0.018	0.136	-0.196	0.24	0.181	0.186	0.121	0.325	-0.124	0.255	-0.032	-0.42	0.255	-0.351	-0.11	-0.194	-0.248	-0.122	0.364	0.08	0.283	0.197	N/A	0.481	0.403	-0.558	0.528	-0.228	0.103	0.124	0.131
CG14740	-0.27	ND	0.15	0.207	-0.266	-0.201	ND	0.543	-0.013	0.86	0.684	0.08	0.147	-0.45	0.578	0.16	0.734	0.133	0.688	-0.239	ND	0.401	-0.195	0.265	0.315	0.328	0.064	0.356	0.554	0.808	0.497	0.481	N/A	-0.24	0.477	0.355	-0.05	0.829	0.303	
kdn	0.283	0.042	0.237	0.183	-0.125	0.501	-0.323	0.595	0	0.029	0.091	0.202	0.542	-0.24	0.402	-0.157	0.157	0.59	-0.368	-0.801	ND	0.293	0.294	0.402	0.411	-0.059	-0.025	0.288	0.478	0.233	-0.013	0.403	ND	N/A	ND	0.489	-0.658	0.364	0.185	0.542
CG18327	0	ND	-0.494	0.56	0.01	-0.201	ND	-0.071	0.209	-0.18	-0.056	-0.37	0.143	0.439	-0.32	-0.014	-0.287	0.054	-0.63	0.357	ND	-0.11	0.415	-0.491	-0.086	-0.404	-0.332	-0.153	-0.09	-0.403	0.119	-0.558	-0.24	ND	N/A	-0.253	0.203	-0.056	-0.438	-0.501
CG8323	-0.01	-0.36	0.042	0.595	-0.233	0.385	-0.283	0.576	0.5	0.268	-0.014	0.022	-0.111	-0.24	0.57	-0.07	0.603	0.127	-0.173	-0.391	0.448	0.419	-0.103	0.235	-0.248	0.204	-0.111	0.782	0.145	0.505	0.149	0.528	0.477	0.489	-0.253	N/A	0.177	0	0.22	-0.099
CG8102	-0.08	-0.36	0.04	0.221	-0.12	-0.639	0.309	-0.055	0.5	0.455	-0.025	-0.113	-0.652	-0.1	0.01	0.096	0.531	-0.285	0.54	0.514	0.193	-0.231	-0.296	0.062	-0.152	0.317	0.103	0.354	-0.205	0.474	-0.037	-0.228	0.355	-0.66	0.203	0.177	N/A	-0.174	0.424	-0.289
CG9140	0.32	0.021	0.354	-0.037	-0.039	-0.059	0.009	-0.022	-0.007	0.123	-0.149	-0.032	0.167	-0.01	0.315	0.079	-0.003	0.071	0.019	-0.453	-0.011	0.079	0.046	-0.021	0.272	0.01	0.091	0.216	-0.139	-0.134	-0.106	0.103	-0.05	0.364	-0.056	0	-0.174	N/A	0.111	0.065
CG10396	0.475	0	0.215	-0.015	0.166	-0.125	0.371	0.185	0.554	0.362	0.275	0.265	-0.056	-0.31	0.352	0.368	0.527	0.021	0.526	-0.25	0.4	0.106	0	0.212	0.155	0.266	0.023	0.156	0.343	0.288	0.11	0.124	0.829	0.185	-0.438	0.22	0.424	0.111	N/A	0.348
CG10664	0.59	-0.06	0.517	-0.5	-0.046	0.089	0.091	0.403	0.151	-0.25	0.285	0.274	0.612	-0.34	-0.19	-0.23	-0.08	0.664	0.21	-0.516	0.496	-0.13	0.395	0.568	0.652	0.009	0.176	-0.177	0.553	-0.11	-0.073	0.131	0.303	0.542	-0.501	-0.099	-0.289	0.065	0.348	N/A

Supplementary Table 4. ERC value between single N-mt genes in *D. melanogaster*

Supplementary Table 5. McDonald–Kreitman test for N-mt duplicated and parental genes between *D. melanogaster*, *D. yakuba*, and *D. simulans*.

N-mt Duplicates		Dn/Ds	Pn/Ps	Unpolarized MK Test P-value	N-mt Parents		Dn/Ds	Pn/Ps	Unpolarized MK Test P-value
FBgn0037170	*Trxr-2*	-	-	Not in that study	FBgn0020653	*Trxr-1*	12/23	1/8	0.173
FBgn0038708	*CG6255*	10/32	1/19	0.0663	FBgn0004888	*Scsalpha1*	-	-	No Polymorphism Data available
FBgn0038200	*CG9920*	-	-	No Polymorphism Data available	FBgn0036334	*CG11267*	-	-	No Polymorphism Data available
FBgn0036706	*CG6485*	4/12	6/14	0.758	FBgn0030853	*CG5703*	-	-	No Polymorphism Data available
FBgn0036222	*CG5718*	7/40	3/49	0.121	FBgn0261439	*SdhA*	-	-	No Polymorphism Data available
FBgn0039651	*CG14508*	10/18	3/18	0.0862	FBgn0035600	*CG4769*	-	-	No Polymorphism Data available
FBgn0028868	*CG4701*	12/29	3/12	0.371	FBgn0005322	*nmd*	8/12	2/7	0.311
FBgn0019830	*colt*	-	-	No Polymorphism Data available	FBgn0031881	*CG3476*	-	-	No Polymorphism Data available
FBgn0035240	*CG33791*	49/69	3/43	0.00000364	FBgn0010352	*Nc73EF*	-	-	No Polymorphism Data available
FBgn0037512	*CG2616*	14/32	0/8	0.0739	FBgn0031039	*Shawn*	-	-	No Polymorphism Data available
FBgn0035568	*CG18418*	8/29	8/13	0.95	FBgn0039674	*CG1907*	-	-	No Polymorphism Data available
FBgn0036568	*CG5389*	-	-	No Polymorphism Data available	FBgn0010217	*ATPsynbeta*	-	-	Not in that study
FBgn0039669	*CG2014*	-	-	No Polymorphism Data available	FBgn0030718	*CG9172*	-	-	No Polymorphism Data available
FBgn0037988	*CG14740*	15/31	0/20	0.00191	FBgn0261955	*kdn*	-	-	No Polymorphism Data available
FBgn0024556	*mEFTu1*	-	-	No Polymorphism Data available	FBgn0033184	*CG12736*	6/26	3/20	0.429
FBgn0039331	*CG11913*	8/37	1/34	0.0364	FBgn0039909	*CG1970*	-	-	Not in that study
FBgn0033904	*CG18327*	-	-	No Polymorphism Data available	FBgn0033903	*CG8323*	6/20	0/18	0.0326
FBgn0034007	*CG8102*	9/32	1/32	0.0532	FBgn0031771	*CG9140*	-	-	No Polymorphism Data available
FBgn0033020	*COX4L*	-	-	Not in that study	FBgn0032833	*COX4*	5/7	2/5	0.474
FBgn0053092	*CG33092*	-	-	Not in that study	FBgn0037138	*CG7145*	6/46	1/11	0.609

Supplementary Table 6. Fst analysis of N-mt retrogenes between 9 different populations of *D. melanogaster*.

CG ID	EA_EF	EA_FR	EA_RAL	EA_SD	EA_ZI	EF_EG	EF_FR	EF_RAL	EF_RG	EF_SP	EF_SD	EF_ZI	EG_RAL	EG_ZI	EG_FR	EG_SD	FR_RG	FR_SD	FR_RAL	FR_SP	FR_ZI	RAL_RG	RAL_SD	RAL_SP	RAL_ZI	RG_SD	RG_ZI	SD_SP	SP_ZI	SD_ZI
CG10090	0.36	0.52	0.49	0.22	0.21	0.55	0.63	0.59	0.14	0.29	0.25	0.22	-0.03	0.26	-0.03	0.22	0.41	0.30	0.02	0.40	0.34	0.38	0.28	0.38	0.32	0.09	0.10	0.00	0.00	-0.02
CG10748	0.04	0.24	0.17	0.15	0.21	0.21	0.35	0.28	0.12	0.17	0.19	0.20	0.00	0.26	0.03	0.11	0.20	0.26	0.47	0.20	0.34	0.11	0.19	0.12	0.32	0.04	0.10	0.01	0.00	-0.02
CG10749	-0.04	0.21	0.18	0.10	0.13	0.19	0.21	0.18	0.04	0.11	0.10	0.13	0.03	0.13	0.04	0.12	0.11	0.17	0.03	0.16	0.20	0.10	0.14	0.11	0.17	0.03	0.05	0.01	0.02	0.00
CG11401	0.00	0.21	0.17	0.06	0.07	0.23	0.21	0.19	0.16	0.07	0.08	0.09	0.06	0.17	0.07	0.18	0.21	0.16	0.02	0.18	0.14	0.17	0.14	0.15	0.12	0.09	0.06	-0.03	0.01	0.01
CG11913	-0.08	0.07	0.01	0.12	0.04	0.05	0.15	0.13	0.25	0.22	0.28	0.19	0.05	0.18	0.08	0.27	0.26	0.30	0.08	0.24	0.20	0.18	0.23	0.16	0.12	0.09	0.02	-0.02	0.02	0.05
CG17856	-0.05	0.13	0.11	0.05	0.23	0.10	0.12	0.10	0.07	0.18	0.09	0.23	0.07	0.35	0.08	0.19	0.18	0.14	0.01	0.23	0.26	0.14	0.12	0.20	0.23	0.04	0.11	0.04	0.01	0.11
CG18418	0.05	0.27	0.26	0.08	0.09	0.26	0.33	0.34	0.18	0.13	0.15	0.13	0.03	0.20	0.03	0.17	0.15	0.21	0.00	0.23	0.25	0.15	0.21	0.23	0.24	0.21	0.04	0.00	-0.01	0.08
CG4706	0.25	0.10	0.09	0.19	0.12	0.29	0.15	0.30	0.32	0.33	0.31	0.24	0.06	0.12	0.06	0.15	0.19	0.21	0.09	0.20	0.14	0.17	0.20	0.16	0.18	0.02	0.07	-0.01	0.02	0.02
CG5265	0.04	0.34	0.24	0.17	0.14	0.43	0.50	0.42	0.29	0.31	0.31	0.28	0.02	0.12	0.05	0.09	0.21	0.10	0.03	0.16	0.17	0.18	0.08	0.13	0.13	0.02	0.02	-0.01	0.01	0.01
CG5718	0.00	0.22	0.15	0.13	0.09	0.29	0.29	0.23	0.10	0.19	0.16	0.13	0.14	0.24	0.20	0.30	0.22	0.24	0.04	0.23	0.18	0.14	0.18	0.17	0.12	0.10	0.06	-0.01	0.02	0.02
CG6255	0.00	0.26	0.34	0.14	0.20	0.05	0.23	0.29	0.05	0.09	0.10	0.14	0.15	0.16	0.08	0.14	0.25	0.22	0.02	0.16	0.19	0.32	0.29	0.23	0.27	0.03	0.08	-0.02	0.00	0.00
CG6485	-0.04	0.45	0.33	0.16	0.19	0.37	0.44	0.33	0.08	0.17	0.16	0.19	0.04	0.15	0.06	0.17	0.42	0.28	0.03	0.29	0.25	0.29	0.19	0.20	0.17	0.17	0.17	0.01	0.00	0.01
CG7514	0.06	0.14	0.19	0.01	0.01	0.25	0.25	0.29	0.11	0.15	0.16	0.17	0.03	0.21	0.03	0.20	0.13	0.18	0.04	0.18	0.20	0.18	0.23	0.23	0.25	0.04	0.04	0.00	0.00	0.00
CG8330	0.47	0.56	0.45	0.56	0.50	0.26	0.18	0.08	0.08	0.02	0.07	0.09	0.16	0.30	0.14	0.25	0.29	0.13	0.09	0.16	0.23	0.13	0.07	0.04	0.09	0.16	0.10	0.01	0.01	0.05
CG9920	0.07	0.32	0.12	-0.03	0.02	0.34	0.45	0.32	0.03	0.03	0.04	0.06	0.19	0.20	0.42	0.14	0.37	0.32	0.09	0.43	0.36	0.24	0.18	0.32	0.23	0.01	0.02	0.01	0.01	-0.02
CG14508	-0.02	0.21	0.15	0.10	0.09	0.27	0.27	0.21	0.16	0.19	0.22	0.21	0.07	0.22	0.05	0.31	0.18	0.24	0.06	0.19	0.19	0.12	0.20	0.12	0.13	0.04	0.01	0.00	-0.01	0.05
CG1409	0.09	0.15	0.10	0.06	0.08	0.32	0.33	0.31	0.10	0.08	0.09	0.11	0.03	0.21	-0.01	0.09	0.21	0.16	0.04	0.20	0.21	0.21	0.14	0.19	0.19	0.04	0.05	0.00	0.00	0.01
CG3057	0.01	0.23	0.19	0.16	0.16	0.14	0.20	0.16	0.04	0.13	0.09	0.09	0.11	0.07	0.13	0.04	0.14	0.19	0.10	0.25	0.20	0.10	0.14	0.20	0.15	0.01	0.02	0.00	0.00	-0.01
CG4701	0.05	0.26	0.10	0.01	0.03	0.34	0.41	0.24	0.15	0.09	0.11	0.11	0.08	0.25	0.08	0.22	0.39	0.28	0.10	0.24	0.32	0.19	0.15	0.10	0.17	0.08	0.08	-0.02	-0.01	0.01
CG6050	0.01	0.32	0.19	0.03	0.02	0.08	0.24	0.13	0.05	0.05	0.08	0.03	0.05	0.13	0.13	0.21	0.39	0.40	0.10	0.39	0.33	0.26	0.29	0.27	0.21	0.08	0.06	-0.01	0.01	0.01
CG1724	0.06	0.26	0.20	0.11	0.11	0.24	0.26	0.22	0.04	0.07	0.08	0.08	0.06	0.15	0.04	0.13	0.24	0.16	0.06	0.17	0.17	0.19	0.15	0.15	0.15	0.08	0.07	-0.01	0.00	0.00

Supplementary Table 7. Fst analysis of random retrogenes between 9 different populations of *D. melanogaster*

CG ID	EA_EF	EA_FR	EA_RAL	EA_SD	EA_ZI	EF_EG	EF_FR	EF_RAL	EF_RG	EF_SP	EF_SD	EF_ZI	EG_RAL	EG_ZI	EG_FR	EG_SD	FR_RG	FR_SD	FR_RAL	FR_SP	FR_ZI	RAL_RG	RAL_SD	RAL_SP	RAL_ZI	RG_SD	RG_ZI	SD_SP	SP_ZI	SD_ZI
CG18607	-0.06	0.25	0.17	0.22	0.20	0.20	0.12	0.19	0.14	0.23	0.21	0.19	0.03	0.20	0.08	0.18	0.23	0.26	0.02	0.30	0.28	0.18	0.22	0.26	0.08	0.08	0.08	-0.02	-0.01	-0.01
CG7496	-0.01	0.14	0.10	0.03	0.03	0.25	0.20	0.17	0.05	0.04	0.08	0.04	0.05	0.28	0.02	0.25	0.17	0.20	0.01	0.24	0.22	0.15	0.18	0.21	0.20	0.01	0.01	0.01	-0.01	0.04
CG4621	-0.02	0.43	0.31	0.18	0.18	0.25	0.50	0.38	0.24	0.21	0.20	0.21	0.05	0.10	0.14	0.11	0.50	0.36	0.04	0.39	0.31	0.35	0.22	0.26	0.17	0.07	0.10	0.02	0.01	0.10
CG13946	-0.02	0.25	0.11	0.10	0.08	0.14	0.17	0.19	0.07	0.12	0.16	0.14	0.02	0.09	0.06	0.04	0.19	0.16	0.05	0.20	0.22	0.09	0.07	0.09	0.11	0.04	0.01	0.00	0.00	0.02
CG3213	0.02	0.18	0.12	0.03	0.05	0.14	0.22	0.16	0.05	0.10	0.09	0.09	0.06	0.07	0.11	0.05	0.20	0.18	0.02	0.18	0.19	0.13	0.11	0.12	0.12	0.02	0.03	0.00	0.01	0.00
CG3610	0.01	0.30	0.30	0.13	0.09	0.36	0.41	0.40	0.17	0.20	0.19	0.18	0.01	0.23	-0.01	0.23	0.25	0.27	0.02	0.28	0.28	0.24	0.26	0.26	0.26	0.11	0.09	0.04	0.01	0.03
CG7094	0.00	0.05	0.12	0.17	0.14	0.05	0.07	0.13	0.14	0.09	0.13	0.10	0.07	0.09	0.03	0.10	0.12	0.13	0.08	0.08	0.11	0.03	0.08	0.05	0.09	0.07	0.08	-0.03	-0.01	0.00
CG12192	0.04	0.30	0.23	0.09	0.09	0.09	0.24	0.18	0.05	0.07	0.04	0.05	0.04	0.12	0.09	0.11	0.29	0.26	0.02	0.28	0.27	0.22	0.19	0.21	0.20	0.03	0.03	-0.01	0.01	-0.01
CG6873	0.02	0.48	0.49	0.33	0.34	0.51	0.47	0.48	0.21	0.36	0.35	0.35	0.05	0.29	0.02	0.31	0.13	0.29	0.02	0.28	0.27	0.14	0.33	0.33	0.32	0.20	0.19	-0.01	-0.01	0.00
CG6036	0.00	0.15	0.11	0.07	0.05	0.12	0.16	0.11	0.58	0.07	0.09	0.05	0.05	0.08	0.03	0.13	0.25	0.15	0.03	0.16	0.11	0.20	0.11	0.10	0.08	0.15	0.11	0.00	0.01	0.02
CG32090	-0.01	0.18	0.11	0.06	0.09	0.23	0.18	0.11	0.08	0.06	0.05	0.07	0.06	0.27	0.03	0.24	0.22	0.16	0.01	0.18	0.19	0.15	0.09	0.10	0.11	0.06	0.04	0.00	0.00	0.02
CG1287	-0.06	0.18	0.20	0.47	0.37	0.09	0.20	0.21	0.14	0.45	0.49	0.39	0.04	0.15	0.02	0.23	0.09	0.29	0.01	0.21	0.19	0.08	0.26	0.24	0.18	0.17	0.09	0.04	0.01	0.04
CG17645	0.12	0.41	0.29	0.23	0.20	0.23	0.20	0.11	0.10	0.06	0.09	0.05	0.07	0.20	0.03	0.17	0.33	0.17	0.05	0.21	0.19	0.22	0.09	0.13	0.10	0.13	0.13	0.04	0.00	0.04
CG10839	-0.09	0.21	0.13	0.23	0.26	0.36	0.29	0.22	0.15	0.29	0.29	0.31	0.13	0.24	0.31	0.18	0.13	0.21	0.06	0.22	0.25	0.06	0.12	0.14	0.17	0.17	0.20	-0.01	-0.01	0.01
CG32087	0.01	0.20	0.15	0.09	0.07	0.26	0.19	0.14	0.04	0.09	0.12	0.10	0.11	0.17	0.07	0.22	0.15	0.16	0.01	0.14	0.13	0.10	0.11	0.10	0.09	0.05	0.04	0.00	-0.01	0.01
CG7423	0.01	0.25	0.27	0.10	0.11	0.36	0.33	0.35	0.11	0.15	0.17	0.18	0.03	0.40	0.02	0.37	0.28	0.36	0.02	0.33	0.38	0.29	0.38	0.35	0.40	0.03	0.04	-0.01	0.00	0.00
CG12362	0.05	0.43	0.24	0.04	0.07	0.24	0.40	0.20	0.07	0.06	0.07	0.06	0.00	0.18	0.03	0.20	0.45	0.33	0.08	0.33	0.32	0.25	0.16	0.15	0.15	0.06	0.06	-0.02	0.00	0.01
CG8584	0.09	0.57	0.51	0.36	0.36	0.34	0.45	0.39	0.20	0.24	0.26	0.24	0.05	0.15	0.05	0.14	0.28	0.20	0.01	0.22	0.20	0.22	0.14	0.16	0.14	0.15	0.08	0.02	-0.02	0.00
CG7804	0.20	0.17	0.15	0.03	0.02	0.48	0.48	0.47	0.30	0.29	0.32	0.29	0.01	0.16	0.02	0.17	0.20	0.22	0.02	0.23	0.20	0.17	0.16	0.20	0.17	0.06	0.04	0.01	0.00	0.00
CG9722	-0.01	0.18	0.13	-0.04	-0.02	0.31	0.29	0.25	0.07	0.12	0.04	0.10	0.06	0.24	0.03	0.26	0.20	0.20	0.03	0.22	0.19	0.18	0.18	0.20	0.16	0.01	0.04	0.01	0.03	0.01
CG3219	0.04	0.14	0.13	0.18	0.16	0.28	0.25	0.24	0.25	0.29	0.30	0.28	0.10	0.11	0.08	0.11	0.20	0.23	0.02	0.20	0.22	0.20	0.23	0.20	0.22	0.01	0.03	-0.01	-0.01	0.00

Supplementary Table 8. Ka/Ks analysis of N-mt duplicated, parental, and random set of genes in *D. melanogaster*

N-mt Duplicates		Parent		Testis-specific Random Genes		Random	
Ensembl Gene ID	**Ka/Ks**	**Ensembl Gene ID**	**Ka/Ks**	**Ensembl Gene ID**	**Ka/Ks**	**Ensembl Gene ID**	**Ka/Ks**
FBgn0036706	0.079011	FBgn0016041	0.0742	FBgn0004171	0.009089	FBgn0035340	0.111
FBgn0034007	0.045163	FBgn0020653	0.039345	FBgn0040371	0.012284	FBgn0035806	0.103751
FBgn0034251	-	FBgn0004888	-	FBgn0001217	0.013785	FBgn0036545	0.028976
FBgn0039331	0.033713	FBgn0036334	0.016598	FBgn0038630	0.018544	FBgn0083943	0.030798
FBgn0030975	0.098958	FBgn0030853	0.042865	FBgn0045827	0.037911	FBgn0031545	0.023798
FBgn0036222	0.024092	FBgn0030733	0.03527	FBgn0038598	0.042042	FBgn00455271	0.023316
FBgn0039576	-	FBgn0261439	0.027322	FBgn0011273	0.058561	FBgn0032650	0.049168
FBgn0039651	0.059965	FBgn0014391	0.123178	FBgn0033610	0.065176	FBgn0036148	0.042
FBgn0033020	-	FBgn0037440	-	FBgn0052436	0.022227	FBgn0033489	0.045168
FBgn0086907	0.053196	FBgn0035600	0.042786	FBgn0031476	0.034784	FBgn0037321	0.031238
FBgn0035585	0.120275	FBgn0005322	0.09444	FBgn0036437	0.046965	FBgn0032598	0.012354
FBgn0036568	0.049948	FBgn0031881	0.079396	FBgn0069354	0.081171	FBgn0037506	0.011054
FBgn0037828	-	FBgn0037242	0.081909	FBgn0035240	0.037286	FBgn0011270	0.02314
FBgn0028868	0.105122	FBgn0038387	0.091232	FBgn0083943	0.06211	FBgn0028858	0.113617
FBgn0033074	-	FBgn0031039	0.080343	FBgn0035197	0.052404	FBgn0032598	0.05645
FBgn0038018	0.169107	FBgn0010352	0.013461	FBgn0034132	0.024052	FBgn00312598	0.071285
FBgn0037310	-	FBgn0039674	0.033179	FBgn0039228	0.055425	FBgn0036082	0.002035
FBgn0037970	-	FBgn0040383	0.0155	FBgn0029659	0.088333	FBgn0033322	0.067885
FBgn0035568	-	FBgn0030718	0.024241	FBgn0062517	0.094901	FBgn0069854	0.069885
FBgn0035567	0.055494	FBgn0010100	0.0225	FBgn0035776	0.103634	FBgn0012589	0.062153
FBgn0037512	0.091935	FBgn0033184	0.042619	FBgn0031347	0.057895	FBgn0034824	0.121275
FBgn0032090	0.191583	FBgn0261955	-	FBgn0036687	0.051442		
FBgn0036327	0.152113	FBgn0050489	0.048678	FBgn0035491	0.31257		
FBgn0036328	0.153718	FBgn0260008	-	FBgn0259794	0.067903		
FBgn0037988	-	FBgn0039909	-	FBgn0037939	0.07918		
FBgn0037862	0.02865	FBgn0262559	0.04311	FBgn0034824	0.121275		
FBgn0036162	-	FBgn0033903	0.051909	FBgn0038208	0.063357		
FBgn0053092	0.124675	FBgn0031771	0.019184	FBgn0051913	0.138361		
FBgn0052026	0.100338			FBgn0038655	0.123059		
FBgn0036490	-			FBgn0034850	0.111556		
FBgn0037170	0.020043			FBgn0039398	0.090479		
FBgn0034472				FBgn0028943	0.36514		
FBgn0037742	0.051082			FBgn0038248	0.093938		
FBgn0038486	-			FBgn0031410	0.244737		
FBgn0038708	-			FBgn0036125	0.10379		
FBgn0011244	-			FBgn0033442	0.147933		
FBgn0001217	0.013785			FBgn0037985	0.248014		
FBgn0031728	-			FBgn0033953	0.150253		
FBgn0038200	0.02252			FBgn0031372	0.307033		

References

Abu-Libdeh B, Douiev L, Amro S, Shahrour M, Ta-Shma A, Miller C, Elpeleg O, Saada A. 2017. Mutation in the COX4I1 gene is associated with short stature, poor weight gain and increased chromosomal breaks, simulating Fanconi anemia. Eur J Hum Genet 25:1142-1146.

Aldridge AC, Benson L, Siegenthaler MM, Whigham BT, Stowers RS, Hales KG. 2007. Roles for Drp1, a Dynamin-related Protein, and Milton, a Kinesin-associated Protein, in Mitochondrial Segregation, Unfurling, and Elongation During Drosophila Spermatogenesis. Fly 1:38-46.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic. Acids. Res. 25:3389-3402.

Aoidi R, Maltais A, Charron J. 2016. Functional redundancy of the kinases MEK1 and MEK2: Rescue of the Mek1 mutant phenotype by Mek2 knock-in reveals a protein threshold effect. Science Signaling 9:ra9-ra9.

Arama E, Agapite J, Steller H. 2003. Caspase activity and a specific cytochrome C are required for sperm differentiation in Drosophila. Dev Cell 4:687-697.

Arama E, Bader M, Srivastava M, Bergmann A, Steller H. 2006. The two Drosophila cytochrome C proteins can function in both respiration and caspase activation. The EMBO Journal 25:232-243.

Arbogast BS. 2015. Phylogeography: The History and Formation of Species. American Zoologist 41:134-135.

Archibald JM. 2015. Endosymbiosis and Eukaryotic Cell Evolution. Curr Biol 25:R911-921.

Arya R, Lakhotia SC. 2008. Hsp60D is essential for caspase-mediated induced apoptosis in Drosophila melanogaster. Cell Stress and Chaperones 13:509-526.

Arya R, Mallik M, Lakhotia SC. 2007. Heat shock genes — integrating cell survival and death. Journal of Biosciences 32:595-610.

Asmussen MA, Arnold J, Avise JC. 1987. Definition and Properties of Disequilibrium Statistics for Associations between Nuclear and Cytoplasmic Genotypes. Genetics 115:755-768.

Asmussen MA, Arnold J, Avise JC. 1989. The effects of assortative mating and migration on cytonuclear associations in hybrid zones. Genetics 122:923.

Baer CF, Miyamoto MM, Denver DR. 2007. Mutation rate variation in multicellular eukaryotes: causes and consequences. Nat Rev Genet 8:619-631.

Bar-Yaacov D, Blumberg A, Mishmar D. 2012. Mitochondrial-nuclear co-evolution and its effects on OXPHOS activity and regulation. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1819:1107-1111.

Barreto FS, Burton RS. 2013. Evidence for compensatory evolution of ribosomal proteins in response to rapid divergence of mitochondrial rRNA. Mol Biol Evol 30:310-314.

Barreto FS, Watson ET, Lima TG, Willett CS, Edmands S, Li W, Burton RS. 2018. Genomic signatures of mitonuclear coevolution across populations of Tigriopus californicus. Nature Ecology & Evolution 2:1250-1257.

Barrientos A, Barros MH, Valnot I, Rotig A, Rustin P, Tzagoloff A. 2002. Cytochrome oxidase in health and disease. Gene 286:53-63.

Bassett AR, Tibbit C, Ponting CP, Liu JL. 2013. Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. Cell Rep 4:220-228.

Bauer MF, Hofmann S, Neupert W, Brunner M. 2000. Protein translocation into mitochondria: the role of TIM complexes. Trends Cell Biol 10:25-31.

Beck EA, Thompson AC, Sharbrough J, Brud E, Llopart A. 2015. Gene flow between Drosophila yakuba and Drosophila santomea in subunit V of cytochrome c oxidase: A potential case of cytonuclear cointrogression. Evolution; international journal of organic evolution 69:1973-1986.

Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh YP, Hahn MW, Nista PM, Jones CD, Kern AD, Dewey CN, et al. 2007. Population genomics: whole-genome analysis of polymorphism and divergence in Drosophila simulans. PLoS Biol 5:e310.

Ben-David G, Miller E, Steinhauer J. 2015. Drosophila spermatid individualization is sensitive to temperature and fatty acid metabolism. Spermatogenesis 5:e1006089.

Berthold J, Bauer MF, Schneider HC, Klaus C, Dietmeier K, Neupert W, Brunner M. 1995. The MIM complex mediates preprotein translocation across the mitochondrial inner membrane and couples it to the mt-Hsp70/ATP driving system. Cell 81:1085-1093.

Bischof J, Bjorklund M, Furger E, Schertel C, Taipale J, Basler K. 2013. A versatile platform for creating a comprehensive UAS-ORFeome library in Drosophila. Development 140:2434-2442.

Bjorkholm P, Harish A, Hagstrom E, Ernst AM, Andersson SG. 2015. Mitochondrial genomes are retained by selective constraints on protein targeting. Proc Natl Acad Sci U S A 112:10154-10161.

Björkholm P, Harish A, Hagström E, Ernst AM, Andersson SGE. 2015. Mitochondrial genomes are retained by selective constraints on protein targeting. Proceedings of the National Academy of Sciences 112:10154-10161.

Bouayed J, Bohn T. 2010. Exogenous antioxidants--Double-edged swords in cellular redox state: Health beneficial effects at physiologic doses versus deleterious effects at high doses. Oxid Med Cell Longev 3:228-237.

Brand AH, Perrimon N. 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118:401-415.

Brown WM, George M, Jr., Wilson AC. 1979. Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci U S A 76:1967-1971.

Burton RS, Ellison CK, Harrison JS. 2006. The sorry state of F2 hybrids: consequences of rapid mitochondrial DNA evolution in allopatric populations. Am Nat 168 Suppl 6:S14-24.

Calvo SE, Mootha VK. 2010. The mitochondrial proteome and human disease. Annu Rev Genomics Hum Genet 11:25-44.

Campos JL, Johnston KJA, Charlesworth B. 2018. The Effects of Sex-Biased Gene Expression and X-Linkage on Rates of Sequence Evolution in Drosophila. Molecular Biology and Evolution 35:655-665.

Capaldi RA. 1990. Structure and function of cytochrome c oxidase. Annu Rev Biochem 59:569-596.

Cardaci S, Zheng L, MacKay G, van den Broek NJ, MacKenzie ED, Nixon C, Stevenson D, Tumanov S, Bulusu V, Kamphorst JJ, et al. 2015. Pyruvate carboxylation enables growth of SDH-deficient cells by supporting aspartate biosynthesis. Nat Cell Biol 17:1317-1326.

Cecchini G. 2003. Function and structure of complex II of the respiratory chain. Annu Rev Biochem 72:77-109.

Chacinska A, Koehler CM, Milenkovic D, Lithgow T, Pfanner N. 2009. Importing mitochondrial proteins: machineries and mechanisms. Cell 138:628-644.

Chacinska A, Lind M, Frazier AE, Dudek J, Meisinger C, Geissler A, Sickmann A, Meyer HE, Truscott KN, Guiard B, et al. 2005. Mitochondrial presequence translocase: switching between TOM tethering and motor recruitment involves Tim21 and Tim17. Cell 120:817-829.

Chaudiere J, Ferrari-Iliou R. 1999. Intracellular antioxidants: from chemical to biochemical mechanisms. Food Chem Toxicol 37:949-962.

Chavez A, Scheiman J, Vora S, Pruitt BW, Tuttle M, E PRI, Lin S, Kiani S, Guzman CD, Wiegand DJ, et al. 2015. Highly efficient Cas9-mediated transcriptional programming. Nat Methods 12:326-328.

Chavez A, Tuttle M, Pruitt BW, Ewen-Campen B, Chari R, Ter-Ovanesyan D, Haque SJ, Cecchi RJ, Kowal EJK, Buchthal J, et al. 2016. Comparison of Cas9 activators in multiple species. Nat Methods 13:563-567.

Chen CS, Gee KR. 2000. Redox-dependent trafficking of 2,3,4,5, 6-pentafluorodihydrotetramethylrosamine, a novel fluorogenic indicator of cellular oxidative activity. Free Radic Biol Med 28:1266-1278.

Chen D, McKearin DM. 2003. A discrete transcriptional silencer in the bam gene determines asymmetric division of the Drosophila germline stem cell. Development 130:1159-1170.

Chen S, Zhang YE, Long M. 2010. New genes in Drosophila quickly become essential. Science 330:1682-1685.

Cheng MY, Hartl FU, Martin J, Pollock RA, Kalousek F, Neuper W, Hallberg EM, Hallberg RL, Horwich AL. 1989. Mitochondrial heat-shock protein hsp60 is essential for assembly of proteins imported into yeast mitochondria. Nature 337:620-625.

Chor B, Tuller T. 2005. Maximum likelihood of evolutionary trees: hardness and approximation. Bioinformatics 21 Suppl 1:i97-106.

Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN, et al. 2007. Evolution of genes and genomes on the Drosophila phylogeny. Nature 450:203-218.

Clark NL, Alani E, Aquadro CF. 2013. Evolutionary rate covariation in meiotic proteins results from fluctuating evolutionary pressure in yeasts and mammals. Genetics 193:529-538.

Clark NL, Alani E, Aquadro CF. 2012. Evolutionary rate covariation reveals shared functionality and coexpression of genes. Genome Res 22:714-720.

Clark NL, Wolfe NW. 2015. ERC analysis: web-based inference of gene function via evolutionary rate covariation. Bioinformatics 31:3835-3837.

Claros MG. 1995. MitoProt, a Macintosh application for studying mitochondrial proteins. Comput Appl Biosci 11:441-447.

Copeland JM, Cho J, Lo T, Jr., Hur JH, Bahadorani S, Arabyan T, Rabie J, Soh J, Walker DW. 2009. Extension of Drosophila life span by RNAi of the mitochondrial respiratory chain. Curr Biol 19:1591-1598.

Crowley LC, Christensen ME, Waterhouse NJ. 2016. Measuring Mitochondrial Transmembrane Potential by TMRE Staining. Cold Spring Harb Protoc 2016.

Darras H, Aron S. 2015. Introgression of mitochondrial DNA among lineages in a hybridogenetic ant. Biology Letters 11.

de Juan D, Pazos F, Valencia A. 2013. Emerging methods in protein co-evolution. Nat Rev Genet 14:249-261.

de Paula WBM, Agip A-NA, Missirlis F, Ashworth R, Vizcay-Barrena G, Lucas CH, Allen JF. 2013. Female and Male Gamete Mitochondria Are Distinct and Complementary in Transcription, Structure, and Genome Function. Genome Biology and Evolution 5:1969-1977.

de Paula Wilson BM, Lucas Cathy H, Agip Ahmed-Noor A, Vizcay-Barrena G, Allen John F. 2013. Energy, ageing, fidelity and sex: oocyte mitochondrial DNA as a protected genetic template. Philosophical Transactions of the Royal Society B: Biological Sciences 368:20120263.

Dietzl G, Chen D, Schnorrer F, Su K, Barinova Y, Fellner M, Gasser B, Kinsey K, Oppel S, Scheiblauer S, et al. 2007. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature 448:151-156.

Dowling DK. 2014. Evolutionary perspectives on the links between mitochondrial genotype and disease phenotype. Biochim Biophys Acta 1840:1393-1403.

Dröge W. 2002. Free Radicals in the Physiological Control of Cell Function. Physiological Reviews 82:47-95.

Du Y, Zhang H, Zhang X, Lu J, Holmgren A. 2013. Thioredoxin 1 is inactivated due to oxidation induced by peroxiredoxin under oxidative stress and reactivated by the glutaredoxin system. J Biol Chem 288:32241-32247.

Duffy JB. 2002a. GAL4 system in Drosophila: a fly geneticist's Swiss army knife. Genesis 34:1-15.

Duffy JB. 2002b. GAL4 system in drosophila: A fly geneticist's swiss army knife. genesis 34:1-15.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-1797.

Ehrlich PR, Raven PH. 1964. Butterflies and Plants: A Study in Coevolution. Evolution 18:586-608.

Eslamieh M, Williford A, Betrán E. 2017. Few Nuclear-Encoded Mitochondrial Gene Duplicates Contribute to Male Germline-Specific Functions in Humans. Genome Biology and Evolution 9:2782-2790.

Fauron CM, Wolstenholme DR. 1980. Extensive diversity among Drosophila species with respect to nucleotide sequences within the adenine + thymine-rich region of mitochondrial DNA molecules. Nucleic Acids Res 8:2439-2452.

Fauron CM, Wolstenholme DR. 1976. Structural heterogeneity of mitochondrial DNA molecules within the genus Drosophila. Proceedings of the National Academy of Sciences 73:3623-3627.

Findlay GD, Sitnik JL, Wang W, Aquadro CF, Clark NL, Wolfner MF. 2014. Evolutionary rate covariation identifies new members of a protein network required for Drosophila melanogaster female post-mating responses. PLoS Genet 10:e1004108.

Finkel T. 2012. Signal transduction by mitochondrial oxidants. J Biol Chem 287:4434-4440.

Fornuskova D, Stiburek L, Wenchich L, Vinsova K, Hansikova H, Zeman J. 2010. Novel insights into the assembly and function of human nuclear-encoded cytochrome oxidase subunits 4, 5a, 6a, 7a and 7b. Biochemical Journal 428:363-374.

Frank SA, Hurst LD. 1996. Mitochondria and male disease. Nature 383:224.

Friedman JR, Nunnari J. 2014. Mitochondrial form and function. Nature 505:335-343.

Fukasawa Y, Tsuji J, Fu SC, Tomii K, Horton P, Imai K. 2015. MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. Mol Cell Proteomics 14:1113-1126.

Gallach M, Betran E. 2011a. Gene duplication might resolve intralocus sexual conflict. Trends in Ecology and Evolution 26:558-559.

Gallach M, Betran E. 2011b. Intralocus sexual conflict resolved through gene duplication. Trends Ecol Evol 26:222-228.

Gallach M, Chandrasekaran C, Betran E. 2010. Analyses of nuclearly-encoded mitochondrial genes suggest gene duplication as a mechanism for resolving intralocus sexual antagonistic conflict in Drosophila. Genome Biology and Evolution 2:835-850.

Garesse R, Kaguni LS. 2005. A Drosophila model of mitochondrial DNA replication: proteins, genes and regulation. IUBMB Life 57:555-561.

Gemmell NJ, Metcalf VJ, Allendorf FW. 2004. Mother's curse: the effect of mtDNA on individual fitness and population viability. Trends Ecol Evol 19:238-244.

Gershoni M, Fuchs A, Shani N, Fridman Y, Corral-Debrinski M, Aharoni A, Frishman D, Mishmar D. 2010. Coevolution predicts direct interactions between mtDNA-encoded and nDNA-encoded subunits of oxidative phosphorylation complex i. J Mol Biol 404:158-171.

Gershoni M, Templeton AR, Mishmar D. 2009. Mitochondrial bioenergetics as a major motive force of speciation. BioEssays 31:642-650.

Gilmore K, Wilson M. 1999. The use of chloromethyl-X-rosamine (Mitotracker Red) to measure loss of mitochondrial membrane potential in apoptotic cells is incompatible with cell fixation. Cytometry 36:355-358.

Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, et al. 2003. A protein interaction map of Drosophila melanogaster. Science 302:1727-1736.

Gratz SJ, Cummings AM, Nguyen JN, Hamm DC, Donohue LK, Harrison MM, Wildonger J, O'Connor-Giles KM. 2013. Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics 194:1029-1035.

Gratz SJ, Ukken FP, Rubinstein CD, Thiede G, Donohue LK, Cummings AM, O'Connor-Giles KM. 2014. Highly Specific and Efficient CRISPR/Cas9-Catalyzed Homology-Directed Repair in Drosophila. Genetics 196:961-971.

Gray MW. 2012. Mitochondrial Evolution. Cold Spring Harb Perspect Biol 4:a011403.

Green EW, Fedele G, Giorgini F, Kyriacou CP. 2014. A Drosophila RNAi collection is subject to dominant phenotypic effects. Nat Methods 11:222-223.

Grisham MB. 1992. Reactive Metabolites of Oxygen and Nitrogen in Biology and Medicine: R.G. Landes Company.

Grossman LI, Wildman DE, Schmidt TR, Goodman M. 2004. Accelerated evolution of the electron transport chain in anthropoid primates. Trends Genet 20:578-585.

Hales K, Fuller M. 1997. Developmentally regulated mitochondrial fusion mediated by a conserved, novel, predicted GTPase. Cell 90:121-129.

Halliwell B, Gutteridge J. 2007. Free radicals in biology and medicine. New York: Oxford University Press.

Handy DE, Loscalzo J. 2012. Redox regulation of mitochondrial function. Antioxid Redox Signal 16:1323-1367.

Hartl FU, Hayer-Hartl M. 2002. Molecular Chaperones in the Cytosol: from Nascent Chain to Folded Protein. Science 295:1852-1858.

Havird JC, McConie HJ. 2019. Sexually Antagonistic Mitonuclear Coevolution in Duplicate Oxidative Phosphorylation Genes. Integrative and Comparative Biology.

Havird JC, Sloan DB. 2016. The Roles of Mutation, Selection, and Expression in Determining Relative Rates of Evolution in Mitochondrial versus Nuclear Genomes. Mol Biol Evol 33:3042-3053.

Havird JC, Whitehill NS, Snow CD, Sloan DB. 2015. Conservative and compensatory evolution in oxidative phosphorylation complexes of angiosperms with highly divergent rates of mitochondrial genome evolution. Evolution 69:3069-3081.

Hawlitschek G, Schneider H, Schmidt B, Tropschug M, Hartl FU, Neupert W. 1988. Mitochondrial protein import: identification of processing peptidase and of PEP, a processing enhancing protein. Cell 53:795-806.

Hell K, Herrmann JM, Pratje E, Neupert W, Stuart RA. 1998. Oxa1p, an essential component of the N-tail protein export machinery in mitochondria. Proceedings of the National Academy of Sciences 95:2250.

Hervas S, Sanz E, Casillas S, Pool JE, Barbadilla A. 2017. PopFly: the Drosophila population genomics browser. Bioinformatics 33:2779-2780.

Hill GE. 2015. Mitonuclear Ecology. Mol Biol Evol 32:1917-1927.

Hill K, Model K, Ryan MT, Dietmeier K, Martin F, Wagner R, Pfanner N. 1998. Tom40 forms the hydrophilic channel of the mitochondrial import pore for preproteins. Nature 395:516.

Housden BE, Lin S, Perrimon N. 2014. Cas9-based genome editing in Drosophila. Methods Enzymol 546.

Howell N. 1997. mtDNA recombination: what do in vitro data mean? American journal of human genetics 61:19-22.

Hudson G, Keers S, Yu-Wai-Man P, Griffiths P, Huoponen K, Savontaus ML, Nikoskelainen E, Zeviani M, Carrara F, Horvath R, et al. 2005. Identification of an X-chromosomal locus and haplotype modulating the phenotype of a mitochondrial DNA disorder. Am J Hum Genet 77:1086-1091.

Huh JR, Vernooy SY, Yu H, Yan N, Shi Y, Guo M, Hay BA. 2004. Multiple apoptotic caspase cascades are required in nonapoptotic roles for Drosophila spermatid individualization. PLoS Biol 2:E15.

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet 11:97-108.

Innocenti P, Morrow EH, Dowling DK. 2011. Experimental evidence supports a sex-specific selective sieve in mitochondrial genome evolution. Science 332:845-848.

Ito K, Awano W, Suzuki K, Hiromi Y, Yamamoto D. 1997. The Drosophila mushroom body is a quadruple structure of clonal units each of which contains a virtually identical set of neurones and glial cells. Development 124:761-771.

Janzen DH. 1980. WHEN IS IT COEVOLUTION? Evolution 34:611-612.

Jenner P. 2003. Oxidative stress in Parkinson's disease. Ann Neurol 53:S26-36; discussion S36-28.

Jonckheere AI, Smeitink JAM, Rodenburg RJT. 2012. Mitochondrial ATP synthase: architecture, function and pathology. Journal of Inherited Metabolic Disease 35:211-225.

Kadenbach B, Huttemann M, Arnold S, Lee I, Bender E. 2000. Mitochondrial energy metabolism is regulated via nuclear-coded subunits of cytochrome c oxidase. Free Radic Biol Med 29:211-221.

Kang DH. 2002. Oxidative stress, DNA damage, and breast cancer. AACN Clin Issues 13:540-549.

Kanzok SM, Fechner A, Bauer H, Ulschmid JK, Muller HM, Botella-Munoz J, Schneuwly S, Schirmer R, Becker K. 2001. Substitution of the thioredoxin system for glutathione reductase in Drosophila melanogaster. Science 291:643-646.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845-858.

Kim KM, Kim PK, Kwon YG, Bai SK, Nam WD, Kim YM. 2002. Regulation of apoptosis by nitrosative stress. J Biochem Mol Biol 35:127-133.

Klichko V, Sohal Barbara H, Radyuk Svetlana N, Orr William C, Sohal Rajindar S. 2014. Decrease in cytochrome *c* oxidase reserve capacity diminishes robustness of *Drosophila melanogaster* and shortens lifespan. Biochemical Journal 459:127-135.

Koehler CM, Jarosch E, Tokatlidis K, Schmid K, Schweyen RJ, Schatz G. 1998. Import of Mitochondrial Carriers Mediated by Essential Proteins of the Intermembrane Space. Science 279:369.

Kondo S, Ueda R. 2013. Highly improved gene targeting by germline-specific Cas9 expression in Drosophila. Genetics 195:715-721.

Kondo S, Vedanayagam J, Mohammed J, Eizadshenass S, Kan L, Pang N, Aradhya R, Siepel A, Steinhauer J, Lai EC. 2017. New genes often acquire male-specific functions but rarely become essential in Drosophila. Genes Dev 31:1841-1846.

Kraytsberg Y, Schwartz M, Brown TA, Ebralidse K, Kunz WS, Clayton DA, Vissing J, Khrapko K. 2004. Recombination of Human Mitochondrial DNA. Science 304:981.

Kubrich M, Keil P, Rassow J, Dekker PJ, Blom J, Meijer M, Pfanner N. 1994. The polytopic mitochondrial inner membrane proteins MIM17 and MIM23 operate at the same preprotein import site. FEBS Lett 349:222-228.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35:1547-1549.

Künkele K-P, Heins S, Dembowski M, Nargang FE, Benz R, Thieffry M, Walz J, Lill R, Nussberger S, Neupert W. 1998. The Preprotein Translocation Channel of the Outer Membrane of Mitochondria. Cell 93:1009-1019.

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

Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. Mol Biol Evol 25:1307-1320.

Lee HY, Chou JY, Cheong L, Chang NH, Yang SY, Leu JY. 2008. Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. Cell 135:1065-1073.

Lee YCG, Ventura IM, Rice GR, Chen DY, Colmenares SU, Long M. 2019. Rapid evolution of gained essential developmental functions of a young gene via interactions with other essential genes. Mol Biol Evol.

Levin L, Blumberg A, Barshad G, Mishmar D. 2014. Mito-nuclear co-evolution: the positive and negative sides of functional ancient mutations. Front Genet 5:448.

Lewis OL, Farr CL, Kaguni LS. 1995. Drosophila melanogaster mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons. Insect Molecular Biology 4:263-278.

Li H, Shao R, Song N, Song F, Jiang P, Li Z, Cai W. 2015. Higher-level phylogeny of paraneopteran insects inferred from mitochondrial genome sequences. Sci Rep 5:8527.

Li S, Shih C-H, Kohn MH. 2010. Functional and evolutionary correlates of gene constellations in the Drosophila melanogaster genome that deviate from the stereotypical gene architecture. BMC genomics 11:322-322.

Li Y, Zhang R, Liu S, Donath A, Peters RS, Ware J, Misof B, Niehuis O, Pfrender ME, Zhou X. 2017. The molecular evolutionary dynamics of oxidative phosphorylation (OXPHOS) genes in Hymenoptera. BMC Evol Biol 17:269.

Lieberman M, Marks AD. 2009. Marks' Basic Medical Biochemistry: A Clinical Approach: Wolters Kluwer Health/Lippincott Williams & Wilkins.

Lin S, Ewen-Campen B, Ni X, Housden BE, Perrimon N. 2015. In Vivo Transcriptional Activation Using CRISPR/Cas9 in Drosophila. Genetics 201.

Lindsley DL, Roote J, Kennison JA. 2013. Anent the genomics of spermatogenesis in Drosophila melanogaster. PLoS One 8:e55915.

Lu JM, Lin PH, Yao Q, Chen C. 2010. Chemical and molecular mechanisms of antioxidants: experimental approaches and model systems. J Cell Mol Med 14:840-860.

Lussey-Lepoutre C, Hollinshead KE, Ludwig C, Menara M, Morin A, Castro-Vega LJ, Parker SJ, Janin M, Martinelli C, Ottolenghi C, et al. 2015. Loss of succinate dehydrogenase activity results in dependency on pyruvate carboxylation for cellular anabolism. Nat Commun 6:8784.

Lynch M. 2010. Evolution of the mutation rate. Trends Genet 26:345-352.

Lynch M, Koskella B, Schaack S. 2006. Mutation Pressure and the Evolution of Organelle Genomic Architecture

10.1126/science.1118884. Science 311:1727-1730.

Lyras L, Cairns NJ, Jenner A, Jenner P, Halliwell B. 1997. An assessment of oxidative damage to proteins, lipids, and DNA in brain from patients with Alzheimer's disease. J Neurochem 68:2061-2069.

Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res 45:D200-d203.

Martijn J, Vosseberg J, Guy L, Offre P, Ettema TJG. 2018. Deep mitochondrial origin outside the sampled alphaproteobacteria. Nature 557:101-105.

Martin WF, Garg S, Zimorski V. 2015. Endosymbiotic theories for eukaryote origin. Philos Trans R Soc Lond B Biol Sci 370:20140330.

Meiklejohn CD, Holmbeck MA, Siddiq MA, Abt DN, Rand DM, Montooth KL. 2013. An Incompatibility between a mitochondrial tRNA and its nuclear-encoded tRNA synthetase compromises development and fitness in Drosophila. PLoS Genet 9:e1003238.

Meiklejohn CD, Montooth KL, Rand DM. 2007. Positive and negative selection on the mitochondrial genome. Trends in Genetics 23:259-263.

Meiklejohn CD, Parsch J, Ranz JM, Hartl DL. 2003. Rapid evolution of male-biased gene expression in Drosophila. Proc Natl Acad Sci U S A 100:9894-9899.

Missirlis F, Ulschmid JK, Hirosawa-Takamori M, Gronke S, Schafer U, Becker K, Phillips JP, Jackle H. 2002. Mitochondrial and cytoplasmic thioredoxin reductase variants encoded by a single Drosophila gene are both essential for viability. J Biol Chem 277:11521-11526.

Miyadera H, Shiomi K, Ui H, Yamaguchi Y, Masuma R, Tomoda H, Miyoshi H, Osanai A, Kita K, Omura S. 2003. Atpenins, potent and specific inhibitors of mitochondrial complex II (succinate-ubiquinone oxidoreductase). Proc Natl Acad Sci U S A 100:473-477.

Montooth KL, Abt DN, Hofmann JW, Rand DM. 2009. Comparative genomics of Drosophila mtDNA: Novel features of conservation and change across functional domains and lineages. Journal of molecular evolution 69:94-114.

Montooth KL, Meiklejohn CD, Abt D, Randa DM. 2010. Mitochondrial-nuclear epistasis affects fitness within species but does not contribute to fixed incompatibilities between species of Drosophila. Evolution; international journal of organic evolution 64:3364-3379.

Moriyama EN, Powell JR. (Moriyama1997 co-authors). 1997. Synonymous substitution rates in Drosophila: Mitochondrial versus nuclear genes. Journal of Molecular Evolution 45:378-391.

Mummery-Widmer JL, Yamazaki M, Stoeger T, Novatchkova M, Bhalerao S, Chen D, Dietzl G, Dickson BJ, Knoblich JA. 2009. Genome-wide analysis of Notch signalling in Drosophila by transgenic RNAi. Nature 458:987-992.

Murrell GA, Francis MJ, Bromley L. 1990. Modulation of fibroblast proliferation by oxygen free radicals. Biochemical Journal 265:659-665.

Naamati A, Regev-Rudzki N, Galperin S, Lill R, Pines O. 2009. Dual targeting of Nfs1 and discovery of its novel processing enzyme, Icp55. The Journal of biological chemistry 284:30200-30208.

Neely GG, Kuba K, Cammarato A, Isobe K, Amann S, Zhang L, Murata M, Elmen L, Gupta V, Arora S, et al. 2010. A global in vivo Drosophila RNAi screen identifies NOT3 as a conserved regulator of heart function. Cell 141:142-153.

Nei M, Li W-H. 1973. Linkage disequilibrium in subdivided populations. Genetics 75:213.

Neupert W. 1997. Protein import into mitochondria. Annu Rev Biochem 66:863-917.

Neupert W, Herrmann JM. 2007. Translocation of proteins into mitochondria. Annu Rev Biochem 76:723-749.

Noguchi T, Koizumi M, Hayashi S. 2011. Sustained Elongation of Sperm Tail Promoted by Local Remodeling of Giant Mitochondria in Drosophila. Current Biology 21:805-814.

Ohno S, Wolf U, Atkin NB. 1968. Evolution from fish to mammals by gene duplication. Hereditas 59:169-187.

Oliveira DC, Raychoudhury R, Lavrov DV, Werren JH. 2008. Rapidly evolving mitochondrial genome and directional selection in mitochondrial genes in the parasitic wasp nasonia (hymenoptera: pteromalidae). Mol Biol Evol 25:2167-2180.

Olson GE, Winfrey VP. 1990. Mitochondria-cytoskeleton interactions in the sperm midpiece. J Struct Biol 103:13-22.

Oughtred R, Stark C, Breitkreutz BJ, Rust J, Boucher L, Chang C, Kolas N, O'Donnell L, Leung G, McAdam R, et al. 2019. The BioGRID interaction database: 2019 update. Nucleic Acids Res 47:D529-d541.

Pagliarini DJ, Calvo SE, Chang B, Sheth SA, Vafai SB, Ong SE, Walford GA, Sugiana C, Boneh A, Chen WK, et al. 2008. A mitochondrial protein compendium elucidates complex I disease biology. Cell 134:112-123.

Partridge L, Hurst LD. 1998. Sex and Conflict. Science 281:2003-2008.

Patel MR. 2017. Inheritance: Male mtDNA Just Can't Catch a Break. Curr Biol 27:R264-r266.

Perezgasga L, Segovia L, Zurita M. 1999. Molecular characterization of the 5' control region and of two lethal alleles affecting the hsp60 gene in Drosophila melanogaster. FEBS Lett 456:269-273.

Pittis AA, Gabaldon T. 2016. Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. Nature 531:101-104.

Pool JE, Corbett-Detig RB, Sugino RP, Stevens KA, Cardeno CM, Crepeau MW, Duchen P, Emerson JJ, Saelao P, Begun DJ, et al. 2012. Population Genomics of sub-saharan Drosophila melanogaster: African diversity and non-African admixture. PLoS Genet 8:e1003080.

Potluri P, Davila A, Ruiz-Pesini E, Mishmar D, O'Hearn S, Hancock S, Simon M, Scheffler IE, Wallace DC, Procaccio V. 2009. A novel NDUFA1 mutation leads to a progressive mitochondrial complex I-specific neurodegenerative disease. Mol Genet Metab 96:189-195.

Pratt SA. 1968. An electron microscope study of nebenkern formation and differentiation in spermatids of Murgantia histrionica (Hemiptera, Pentatomidae). Journal of Morphology 126:31-65.

Prelich G. 2012. Gene overexpression: uses, mechanisms, and interpretation. Genetics 190:841-854.

Pröschel M, Zhang Z, Parsch J. 2006. Widespread adaptive evolution of Drosophila genes with sex-biased expression. Genetics 174:893-900.

R DevelopmentCoreTeam. 2013. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Radyuk SN, Sohal RS, Orr WC. 2003. Thioredoxin peroxidases can foster cytoprotection or cell death in response to different stressors: over- and under-expression of thioredoxin peroxidase in Drosophila cells. Biochemical Journal 371:743-752.

Rand DM, Fry A, Sheldahl L. 2006. Nuclear-mitochondrial epistasis and drosophila aging: introgression of Drosophila simulans mtDNA modifies longevity in D. melanogaster nuclear backgrounds. Genetics 172:329-341.

Rand DM, Haney RA, Fry AJ. 2004. Cytonuclear coevolution: the genomics of cooperation. Trends Ecol Evol 19:645-653.

Ranson NA, White HE, Saibil HR. 1998. Chaperonins. Biochemical Journal 333:233.

Reinhardt K, Dowling DK, Morrow EH. 2013. Mitochondrial Replacement, Evolution, and the Clinic. Science 341:1345-1346.

Richardson MF, Weinert LA, Welch JJ, Linheiro RS, Magwire MM, Jiggins FM, Bergman CM. 2012. Population genomics of the Wolbachia endosymbiont in Drosophila melanogaster. PLoS Genet 8:e1003129.

Riparbelli MG, Callaini G. 2007. The Drosophila parkin homologue is required for normal mitochondrial dynamics during spermiogenesis. Dev Biol 303:108-120.

Rogell B, Dean R, Lemos B, Dowling D. 2014. Mito-nuclear interactions as drivers of gene movement on and off the X-chromosome. BMC genomics 15:1-9.

Rokas A, Ladoukakis E, Zouros E. 2003. Animal mitochondrial DNA recombination revisited. Trends Ecol Evol 18:411-417.

Rosikiewicz W, Kabza M, Kosinski JG, Ciomborowska-Basheer J, Kubiak MR, Makalowska I. 2017. RetrogeneDB-a database of plant and animal retrocopies. Database : the journal of biological databases and curation 2017:bax038.

Rospert S, Looser R, Dubaquie Y, Matouschek A, Glick BS, Schatz G. 1996. Hsp60-independent protein folding in the matrix of yeast mitochondria. The EMBO Journal 15:764-774.

Ross BD, Rosin L, Thomae AW, Hiatt MA, Vermaak D, de la Cruz AF, Imhof A, Mellone BG, Malik HS. 2013. Stepwise evolution of essential centromere function in a Drosophila neogene. Science 340:1211-1214.

Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sanchez-Gracia A. 2017. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Mol Biol Evol 34:3299-3302.

Ryan KR, Jensen RE. 1995. Protein translocation across mitochondrial membranes: what a long, strange trip it is. Cell 83:517-519.

Sacktor B. 1970. Regulation of Intermediary Metabolism, with Special Reference to the Control Mechanisms in Insect Flight Muscle. In: J.W.L. Beament JET, Wigglesworth VB, editors. Advances in Insect Physiology: Academic Press. p. 267-347.

Sagan L. 1967. On the origin of mitosing cells. In. Journal of theoretical biology. p. 255-274.

Saraste M. 2009. Structural features of cytochrome oxidase. Quarterly Reviews of Biophysics 23:331-366.

Sarkar S, Lakhotia SC. 2005. The Hsp60C gene in the 25F cytogenetic region in Drosophila melanogaster is essential for tracheal development and fertility. J Genet 84:265-281.

Savojardo C, Martelli PL, Fariselli P, Casadio R. 2014. TPpred2: improving the prediction of mitochondrial targeting peptide cleavage sites by exploiting sequence motifs. Bioinformatics 30:2973-2974.

Sawyer EM, Brunner EC, Hwang Y, Ivey LE, Brown O, Bannon M, Akrobetu D, Sheaffer KE, Morgan O, Field CO, et al. (Sawyer2017 co-authors). 2017. Testis-specific ATP synthase peripheral stalk subunits required for tissue-specific mitochondrial morphogenesis in Drosophila. BMC Cell Biology 18:16.

Sayre LM, Smith MA, Perry G. 2001. Chemistry and biochemistry of oxidative stress in neurodegenerative disease. Curr Med Chem 8:721-738.

Scarpulla RC. 2008. Transcriptional Paradigms in Mammalian Mitochondrial Biogenesis and Function. Physiological Reviews 88:611.

Schmidt O, Pfanner N, Meisinger C. 2010. Mitochondrial protein import: from proteomics to functional mechanisms. Nat Rev Mol Cell Biol 11:655-667.

Schnabel A, Asmussen MA. 1992. Comparative Effects of Pollen and Seed Migration on the Cytonuclear Structure of Plant Populations. II. Paternal Cytoplasmic Inheritance. Genetics 132:253-267.

Schnabel A, Asmussen MA. 1989. Definition and properties of disequilibria within nuclear-mitochondrial-chloroplast and other nuclear-dicytoplasmic systems. Genetics 123:199-215.

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671.

Schnorrer F, Schonbauer C, Langer CC, Dietzl G, Novatchkova M, Schernhuber K, Fellner M, Azaryan A, Radolf M, Stark A, et al. 2010. Systematic genetic analysis of muscle morphogenesis and function in Drosophila. Nature 464:287-291.

Schreck R, Rieber P, Baeuerle PA. 1991. Reactive oxygen intermediates as apparently widely used messengers in the activation of the NF-kappa B transcription factor and HIV-1. Embo j 10:2247-2258.

Sirrenberg C, Bauer MF, Guiard B, Neupert W, Brunner M. 1996. Import of carrier proteins into the mitochondrial inner membrane mediated by Tim22. Nature 384:582-585.

Slatkin M. 2008. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. Nat Rev Genet 9:477-485.

Sloan DB, Fields PD, Havird JC. 2015. Mitonuclear linkage disequilibrium in human populations. Proceedings of the Royal Society B: Biological Sciences 282.

Sloan DB, Oxelman B, Rautenberg A, Taylor DR. 2009. Phylogenetic analysis of mitochondrial substitution rate variation in the angiosperm tribe Sileneae. BMC Evol Biol 9:260.

Sloan DB, Triant DA, Wu M, Taylor DR. 2014. Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. Mol Biol Evol 31:673-682.

Smedley D, Haider S, Ballester B, Holland R, London D, Thorisson G, Kasprzyk A. 2009. BioMart – biological queries made easy. BMC genomics 10:22.

Song F, Li H, Jiang P, Zhou X, Liu J, Sun C, Vogler AP, Cai W. 2016. Capturing the Phylogeny of Holometabola with Mitochondrial Genome Data and Bayesian Site-Heterogeneous Mixture Models. Genome Biol Evol 8:1411-1426.

Spradling AC, Stern D, Beaton A, Rhem EJ, Laverty T, Mozden N, Misra S, Rubin GM. 1999. The Berkeley Drosophila Genome Project gene disruption project: Single P-element insertions mutating 25% of vital Drosophila genes. Genetics 153:135-177.

Suski JM, Lebiedzinska M, Bonora M, Pinton P, Duszynski J, Wieckowski MR. 2012. Relation between mitochondrial membrane potential and ROS formation. Methods Mol Biol 810:183-205.

Symposia BS, Kay J, Weitzman PDJ. 1987. Krebs' Citric Acid Cycle: Half a Century and Still Turning : Biochemical Society Symposium No. 54 Held at University of Leicester, April 1987: Biochemical Society.

Thibault ST, Singer MA, Miyazaki WY, Milash B, Dompe NA, Singh CM, Buchholz R, Demsky M, Fawcett R, Francis-Lang HL, et al. 2004. A complementary transposon tool kit for Drosophila melanogaster using P and piggyBac. Nature Genetics 36:283.

Thomas CE, Kalyanaraman B. 1997. Oxygen Radical and the Disease Process. Netherlands: Harvard Academic Publishers.

Timakov B, Zhang P. 2001. The hsp60B gene of Drosophila melanogaster is essential for the spermatid individualization process. Cell Stress Chaperones 6:71-77.

Tokuyasu KT, Peacock WJ, Hardy RW. 1972. Dynamics of spermiogenesis in Drosophila melanogaster. I. Individualization process. Z Zellforsch Mikrosk Anat 124:479-506.

Tsubouchi A, Caldwell JC, Tracey WD. 2012. Dendritic filopodia, Ripped Pocket, NOMPC, and NMDARs contribute to the sense of touch in Drosophila larvae. Curr Biol 22:2124-2134.

Tsukihara T, Aoyama H, Yamashita E, Tomizaki T, Yamaguchi H, Shinzawa-Itoh K, Nakashima R, Yaono R, Yoshikawa S. 1996. The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 A. Science 272:1136-1144.

Vedelek V, Bodai L, Grézal G, Kovács B, Boros IM, Laurinyecz B, Sinka R. 2018. Analysis of Drosophila melanogaster testis transcriptome. BMC Genomics 19:697.

Venken KJT, Popodi E, Holtzman SL, Schulze KL, Park S, Carlson JW, Hoskins RA, Bellen HJ, Kaufman TC. 2010. A molecularly defined duplication set for the X chromosome of Drosophila melanogaster. Genetics 186:1111-1125.

Vibranovski MD, Lopes HF, Karr TL, Long M. 2009. Stage-Specific Expression Profiling of Drosophila Spermatogenesis Suggests that Meiotic Sex Chromosome Inactivation Drives Genomic Relocation of Testis-Expressed Genes. PLoS Genetics 5:e1000731.

Vissers JHA, Manning SA, Kulkarni A, Harvey KF. 2016. A Drosophila RNAi library modulates Hippo pathway-dependent tissue growth. Nat Commun 7:10368.

Vogtle FN, Wortelkamp S, Zahedi RP, Becker D, Leidhold C, Gevaert K, Kellermann J, Voos W, Sickmann A, Pfanner N, et al. 2009. Global analysis of the mitochondrial N-proteome identifies a processing peptidase critical for protein stability. Cell 139:428-439.

Wai T, Langer T. 2016. Mitochondrial Dynamics and Metabolic Regulation. Trends Endocrinol Metab 27:105-117.

Wakimoto BT, Lindsley DL, Herrera C. 2004. Toward a comprehensive genetic analysis of male fertility in Drosophila melanogaster. Genetics 167:207-216.

Wallace DC. 2005. A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. Annu Rev Genet 39:359-407.

Wasbrough ER, Dorus S, Hester S, Howard-Murkin J, Lilley K, Wilkin E, Polpitiya A, Petritis K, Karr TL. 2010. The Drosophila melanogaster sperm proteome-II (DmSP-II). J Proteomics 73:2171-2185.

Weinberg SE, Sena LA, Chandel NS. 2015. Mitochondria in the regulation of innate and adaptive immunity. Immunity 42:406-417.

Weir BS, Cockerham CC. 1984. Estimating F-Statistics for the Analysis of Population Structure. Evolution; international journal of organic evolution 38:1358-1370.

White-Cooper H. 2004. Spermatogenesis: analysis of meiosis and morphogenesis. Methods Mol Biol 247:45-75.

Wiedemann N, Pfanner N. 2017. Mitochondrial Machineries for Protein Import and Assembly. Annu Rev Biochem 86:685-714.

Wiegmann BM, Yeates DK, Thorne JL, Kishino H. 2003. Time flies, a new molecular time-scale for brachyceran fly evolution without a clock. Syst Biol 52:745-756.

Wolff JN, Ladoukakis ED, Enríquez JA, Dowling DK. 2014. Mitonuclear interactions: evolutionary consequences over multiple biological scales. Philosophical Transactions of the Royal Society of London B: Biological Sciences 369.

Wolstenholme DR. 1992. Animal Mitochondrial DNA: Structure and Evolution. In: Wolstenholme DR, Jeon KW, editors. International Review of Cytology: Academic Press. p. 173-216.

Xu B, Yang Z. 2013. pamlX: A Graphical User Interface for PAML. Mol Biol Evol 30:2723-2724.

Yan Z, Ye G, Werren JH. 2019. Evolutionary rate correlation between mitochondrial-encoded and mitochondria-associated nuclear-encoded proteins in insects. Mol Biol Evol.

Yoshizawa K, Johnson KP. 2013. Changes in base composition bias of nuclear and mitochondrial genes in lice (Insecta: Psocodea). Genetica 141:491-499.

Yu C-A, Yu L, Xia D. 2013. Ubiquinol-Cytochrome c Oxidoreductase (Complex III). In: Roberts GCK, editor. Encyclopedia of Biophysics. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 2679-2684.

Yu Z, O'Farrell PH, Yakubovich N, DeLuca SZ. 2017. The Mitochondrial DNA Polymerase Promotes Elimination of Paternal Mitochondrial Genomes. Curr Biol 27:1033-1039.

Zhang J, Park SI, Artime MC, Summy JM, Shah AN, Bomser JA, Dorfleutner A, Flynn DC, Gallick GE. 2007. AFAP-110 is overexpressed in prostate cancer and contributes to tumorigenic growth by regulating focal contacts. The Journal of clinical investigation 117:2962-2973.

Zhang Y, Kalderon D. 2000. Regulation of cell proliferation and patterning in Drosophila oogenesis by Hedgehog signaling. Development 127:2165-2176.

Zhang YE, Vibranovski MD, Krinsky BH, Long M. 2010. Age-dependent chromosomal distribution of male-biased genes in Drosophila. Genome Res 20:1526-1533.

Zhu CT, Ingelmo P, Rand DM. 2014. GxGxE for lifespan in Drosophila: mitochondrial, nuclear, and dietary interactions that modify longevity. PLOS Genet 10:e1004354.